



HAL
open science

Robust Estimation of Pareto-Type Tail Index through an Exponential Regression Model

Richard Minkah, Tertius de Wet, Abhik Ghosh

► **To cite this version:**

Richard Minkah, Tertius de Wet, Abhik Ghosh. Robust Estimation of Pareto-Type Tail Index through an Exponential Regression Model. 2019. hal-02116753v2

HAL Id: hal-02116753

<https://hal.science/hal-02116753v2>

Preprint submitted on 14 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust Estimation of Pareto-Type Tail Index through an Exponential Regression Model

Richard Minkah *

Department of Statistics and Actuarial Science, University of Ghana, Ghana

Tertius de Wet

Department of Statistics and Actuarial Science, Stellenbosch University, South Africa

Abhik Ghosh

Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India

Abstract

In this paper, we introduce a robust estimator of the tail index of a Pareto-type distribution. The estimator is obtained through the use of the minimum density power divergence with an exponential regression model for log-spacings of top order statistics. The proposed estimator is compared to existing minimum density power divergence estimators of the tail index based on fitting an extended Pareto distribution and exponential regression model on log-ratio of spacings of order statistics. We derive the influence function and gross error sensitivity of the proposed estimator of the tail index to study its robustness properties. In addition, a simulation study is conducted to assess the performance of the estimators under different contaminated samples from different distributions. The results show that our proposed estimator of the tail index has better mean square errors and is less sensitive to an increase in the number of top order statistics. In addition, the estimation of the exponential regression model yields estimates of second-order parameters that can be used for estimation of extreme events such as quantiles and exceedance probabilities. The proposed estimator is illustrated with a practical dataset on insurance claims.

Keywords: Tail Index, robust estimation, maximum likelihood, exponential regression model, extended Pareto distribution; minimum density power divergence

*Correspondence: rminkah@ug.edu.gh

The first named author (RM) would like to thank the UG-Carnegie NGAA and UG-BANGA Africa projects for providing financial support. The research of the second named author (TdW) was supported by the South African NRF under grant number 115038.

1 Introduction

Extreme value theory (EVT) has become an important tool in many disciplines for the estimation of rare events that are related to environmental science, hydrology, insurance and finance, among others. The process of extreme value analysis involves fitting an extreme value distribution, characterised by a tail index, which measures the tail heaviness of the distribution function. The most common method for estimating the parameters of an extreme value distribution in an extreme value analysis is the maximum likelihood. Also in the semi-parametric framework, the hill estimator (Hill, 1975) remains the most popular among a series of estimators. However, these estimators do not take into account possible deviations from assumed extreme value models. This may arise as a result of possible outliers in the data that may (or may not) have been recorded in error. In such a dataset, the estimators mentioned above are known to be sensitive to such outlying observations which affect the quality of the estimates of model parameters. In addition, small errors in the estimation of model parameters such as the tail index can cause significant errors in the estimation of extreme events such as high quantiles and exceedance probabilities (see e.g. Brazauskas and Serfling, 2000).

Robust statistics presents a better method for addressing outliers and deviations from assumed parametric models. In the context of extreme value analysis, its usage may appear to be contradictory. However, it has been shown that employing robust statistical ideas in extreme value theory, improves the quality and precision of estimates (Dell'Aquila and Embrechts, 2006). Among the early applications of robust estimators include the Optimal Biased Reduced Estimator (OBRE) of the parameters of the GEV distribution (Dupius and Field, 1998), generalised mean and trimmed mean type estimators (Brazauskas and Serfling, 2000, 2001), method of medians for the generalised Pareto distribution (Peng and Welsh, 2001), and an integrated squared error approach on partial density component estimation of the parameters of the generalised Pareto distribution (Vandewalle et al., 2007).

Furthermore, Juarez and Schucany (2004) seem to be first authors to employ the minimum density power divergence (MDPD) of Basu et al. (1998) for the robust estimation of parameters of an extreme value distribution. Since then, this divergence measure has become the most sought after divergence measure for robust estimation of parameters of extreme value distributions. Kim and Lee (2008), Dierckx et al. (2013), Goegebeur et al. (2014), Dierckx et al. (2018) have made use of the MDPD in estimating the tail index and quantiles from Pareto-type distributions. Recently, Ghosh (2017) proposed a robust MDPD estimator for real-valued tail index. This estimator is a robust generalisation of the estimator proposed by Matthys and Beirlant (2003) and the author addresses the non-identical distributions of the exponential regression model using the approach in Ghosh and Basu (2013). Also, Dierckx et al. (2013) employ the MDPD concept on an extended Pareto distribution for relative excesses over a high threshold. This distribution has second-order properties that are suitable for bias reduction such as in quantile estimation (Dierckx et al., 2018).

In the present paper, we propose a robust estimator for tail index of Pareto-type using the MDPD idea on an exponential regression model. Our estimator is a robust generalisation of the estimator in Beirlant et al. (1999), and hence, it is different from the estimator in Ghosh (2017). Again the use of this exponential regression model leads to estimates of other second-order parameters that can be used to obtain bias-reduced estimators of extreme events such as

quantiles and exceedance probabilities.

The rest of the paper is organised as follows. In Section 2, we present the robust estimation methods of the tail index, beginning with an introduction to extreme value theory. The robustness properties of the proposed estimators are studied using influence function and gross error sensitivity analyses in Section 3. In section 4, the proposed estimator of the the Pareto-type tail index is compared with two existing estimators in the literature via a simulation study. Section 5 presents an illustration of the proposed estimator on the estimation of the tail index of a practical data set from the insurance industry. We provide concluding remarks in Section 6.

2 Estimation Method

Let X_1, X_2, \dots, X_n be a sample of independent and identically distributed observations from some process with underlying distribution F . Also let $X_{1,n} \leq X_{2,n} \leq \dots, X_{n,n}$ be the sample order statistics associated with the sample. In order to infer on extreme events in the far tails or beyond the data, one approach is to study the behaviour of the sample maximum, $X_{n,n}$. The well-known Fisher and Tippett (1928) and Gnedenko (1943) theorem provides that a suitably normalised maximum, $X_{n,n} = \max\{X_1, \dots, X_n\}$, converges in distribution to a non-degenerate limit as $n \rightarrow \infty$. Such a limit distribution was shown to be of the so-called generalised extreme value (GEV) distribution. Formally, if normalising sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ exist, then

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = G_\gamma(x) \quad (1)$$

with

$$G_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), & 1 + \gamma x > 0, & \text{if } \gamma \neq 0 \\ \exp(\exp(-x)), & x \in \mathbb{R}, & \text{if } \gamma = 0. \end{cases} \quad (2)$$

If data from a distribution function, F , satisfies (2), F is said to belong to the domain of attraction of G_γ and is denoted by $F \in D(G_\gamma)$. Here, γ is the shape (tail index) and it measures the tail heaviness of the underlying distribution, F . In particular, the distribution belongs to the Pareto domain of attraction for $\gamma > 0$, Gumbel domain of attraction for $\gamma = 0$, and the Weibull domain of attraction for $\gamma < 0$ with a finite right endpoint. The goal of extreme value analysis is mainly to obtain estimates of high quantiles, exceedance probabilities and return periods. However, each of the latter depends on the extreme value index, γ , which measures the tail heaviness of the underlying distribution. Therefore, the estimation of γ remains an important research area in EVT.

Another approach to obtaining the tail index relies on the Balkema and de Haan (1974) and Pickands III (1975) theorem, which states that the distribution is in the max-domain of attraction of the GEV distribution if and only if the distribution of excesses over high thresholds is asymptotically the generalised Pareto (GPD). An application of this theorem in Davison and Smith (1990) gave rise to the so-called Peaks-Over-Threshold (POT) methodology in extreme value analysis.

Among the early and popular estimators for estimating the parameters in (2) include, the maximum likelihood method, probability weighted moments and elemental percentile. In addition, other semi-parametric estimators exist such as Hill (Hill, 1975), moment (Dekkers et al., 1989),

exponential regression (Beirlant et al., 1999, 2009). However, in most instances the parametric distribution, GEV or GPD, may not model all the data well. In addition, small deviations from the assumed model may cause considerable effect on estimation of parameters and thereby affect the estimation of extreme events such as high quantiles and exceedance probabilities. Robust estimation aims at providing estimates that are stable or consistent within the neighbourhood of the assumed model and can provide an assessment for the fit of the data to the model. If an extreme observation is down weighted, then inferences on the GEV or GPD is potentially flawed. Two options available are to base inferences on the part that is well fitted by the extreme value distributions, GEV and GPD, or obtain a desirable model where the weights are consistent with the bulk of the data.

In this paper, we consider the estimation of $\gamma > 0$, i.e. the Pareto domain of attraction. Such domain has survival function,

$$1 - F(x) = x^{-1/\gamma} \ell_F(x) \quad (3)$$

or tail quantile function

$$U(x) = Q(1 - 1/x) = x^\gamma \ell_U(x) \quad (4)$$

with Q the quantile function of F . Here, ℓ_F and ℓ_U are slowly varying functions given for $t > 0$ defined as,

$$\lim_{x \rightarrow \infty} \frac{\ell_F(xt)}{\ell_F(x)} = 1, \quad (5)$$

and similarly for ℓ_U .

In the next two sub-sections, we present the two methods used in estimation of the tail index of a distribution function. In the third sub-section, we discuss the robust method of estimation of the tail index using the minimum density power divergence method of Basu et al. (1998).

2.1 Extended Pareto Model

From (3), the conditional survival function of the relative excesses $P\left(\frac{X}{u} > x \mid X > u\right)$ converges to $x^{-1/\gamma}$ for $x > 1$. Using the k upper order statistics and the Pareto-type behaviour, an estimate of γ is obtained as the slope of the Pareto quantile plot. Also, the maximum likelihood estimate of γ is the usual Hill estimator (Hill, 1975) given by

$$\hat{\gamma}^H = \frac{1}{k} \sum_{j=1}^k j (\log X_{n-j+1,n} - \log X_{n-j,n}). \quad (6)$$

This estimator has been studied extensively in the literature because of its attractive properties. However, it is known to have large bias and sensitive to outliers.

In view of this, Dierckx et al. (2013) employs the second-order condition of Beirlant et al. (2009) on the rate of convergence of (5) to improve on the bias of this estimator. Denote by RV_β , a class of functions regularly varying at infinity, with index β satisfying

$$\lim_{x \rightarrow \infty} \frac{L_F(xt)}{L_F(x)} = t^\beta, \quad (7)$$

with (7) reducing to (5) if $\beta = 0$. The second-order condition needed to obtain the survival function of the extended Pareto distribution is given by:

Condition 1: Suppose $\gamma > 0$ and $\tau < 0$ are constants, the distribution function F is said to satisfy the second-order condition if $x^{1/\gamma}(1 - F(x)) \rightarrow C \in (0, \infty)$ as $x \rightarrow \infty$ and the function δ defined via

$$1 - F(x) = Cx^{-1/\gamma}(1 + \gamma^{-1}\delta(x))$$

is ultimately non-zero, of constant sign and $|\delta| \in RV_\tau$ (Dierckx et al., 2013, pg. 71).

Equivalently, from *Condition 1*, the tail quantile function U , satisfies $y^{-\gamma}U(y) \rightarrow C^\gamma$ as $y \rightarrow \infty$. Also define a function a implicitly as

$$U(y) = C^\gamma y^\gamma (1 + a(y)) \quad (8)$$

with $a(y) = \delta(C^\gamma y^\gamma)(1 + o(1))$ as $y \rightarrow \infty$. Thus, $|a| \in RV_\rho$ where $\rho = \gamma\tau$. The second-order condition was then used to obtain an extended Pareto distribution with survival function given by

$$1 - G(y) = y^{-1/\gamma}(1 + \delta - \delta y^\tau)^{-1/\gamma} \mathbb{I}\{y > 1\} \quad (9)$$

and a density function

$$g(y) = \frac{1}{\gamma} y^{-1/\gamma-1} [1 + \delta(1 - y^\tau)]^{-1/\gamma-1} [1 + \delta(1 - (1 + \tau)y^\tau)], \quad y > 1, \quad (10)$$

where, $\gamma > 0$, $\tau < 0$ and $\delta \in \max\{-1, 1/\tau\}$. In practice, (9) is fitted to relative excesses over a threshold, X_{n-k} , denoted $Y_i = X_{n-j+1}/X_{n-k}$, $j = 1, 2, \dots, k$. The parameters $\gamma > 0$, $\tau < 0$ and δ can be estimated through maximum likelihood (Beirlant et al., 2009).

2.2 The Exponential Regression Model

Consider again X_1, X_2, \dots, X_n i.i.d. random variables with common underlying distribution F and associated quantile function Q . Then for the Pareto-type tails i.e. $\gamma > 0$, the survival function is given by (3). Similarly, the associated tail quantile function U can be written in terms of the associated slowly varying function ℓ_U as in 4. From (4), the order statistics $X_{1,n}, X_{2,n}, \dots, X_{n,n}$ can be represented jointly as

$$\log X_{n-j+1,n} \stackrel{d}{\sim} \gamma \log U_{j,n}^{-1} + \log \ell(U_{j,n}^{-1}) \quad (11)$$

where $U_{j,n}^{-1}, j = 1, 2, \dots, n$ denote the order statistics of the standard uniform distribution, $U(0, 1)$. From (11), Beirlant et al. (1999) obtain an approximate representation for

$$\frac{\log X_{n-j+1,n}}{\log X_{n-k,n}} \stackrel{d}{\sim} \gamma \log \frac{U_{k+1,n}^{-1}}{U_{j,n}^{-1}} + \log \ell \left(\frac{U_{j,n}^{-1}}{U_{k+1,n}^{-1}} \right) \quad (12)$$

$k \in \{2, 3, \dots, n-1\}$. The authors state that a more accurate representation is obtained from (12) by implementing a slow variation with remainder condition on the rate of convergence to the limit in (5). This is given as Condition 2:

Condition 2: There exists a real constant $\rho \leq 0$ and a rate function b satisfying $b(x) \rightarrow 0$ as $x \rightarrow \infty$ such that for all $u \geq 1$,

$$\lim_{x \rightarrow \infty} \log \frac{\ell(ux)}{\ell(x)} = b(x)\kappa_\rho(u) \quad (13)$$

with $\kappa_\rho(u) = \int_1^u v^{\rho-1} dv$.

Under *Condition 2*, Beirlant et al. (1999) show that the weighted log-spacings of the order statistics in (6),

$$Z_i = i. (\log X_{n-i+1,n} - \log X_{n-i,n}), \quad 1 \leq i \leq k \leq n, \quad (14)$$

are approximately exponentially distributed. Specifically, they obtain a regression model given by,

$$Z_i \sim \left(\gamma + b_{n,k} \left(\frac{i}{k+1} \right)^{-\rho} \right) E_i \quad (15)$$

where each E_i is a standard exponential random variable, and $b_{n,k} \rightarrow 0$ as $k, n \rightarrow \infty$, and $\rho < 0$ are second-order parameters. The parameters in (15) were estimated by maximum likelihood in Beirlant et al. (1999) and shown to be better at reducing bias than the traditional estimators such as Hill (1975). Also, when $b_{n,k} = 0$, in (15), the resulting maximum likelihood estimator is exactly the Hill estimator (Hill, 1975).

In this paper, we propose estimating the parameters robustly using the density power divergence method of Basu et al. (1998). Our proposal is different from Ghosh (2017), in three ways. Firstly, whereas we use the distribution of log-spacings of order statistics, Ghosh (2017) uses the distribution of log ratio of order statistics. Secondly, our proposal is strictly for the Pareto domain i.e. $\gamma > 0$ as against $\gamma \in \mathbb{R}$. Lastly, the estimation of γ yields estimates of other second-order parameters that can be used in the reduced-biased estimators such as for quantiles and exceedance probabilities.

2.3 Robust Estimation through the Minimum Density Power Divergence

Consider two density functions f and g , the minimum density power divergence between f and g , introduced by Basu et al. (1998) has been used extensively to provide robust estimators and in recent years has received attention in extreme value analysis (see e.g. Dierckx et al., 2013, 2018; Kim and Lee, 2008). The popularity of the minimum divergence power density function stems from its implicit usage of the empirical density function of the data. In this method, weighted likelihood estimation equations are developed and observations that are outliers in relation to the model distribution are down-weighted by a robustness parameter, α , of the model density.

The density power divergence between any two density functions f and g , is defined as

$$d_\gamma(f, g) = \begin{cases} \int [f^{1+\alpha} - (1 + \frac{1}{\alpha}) f^\alpha g + \frac{1}{\alpha} g^{1+\alpha}], & \text{if } \alpha > 0 \\ \int f \log(g/f), & \text{if } \alpha = 0. \end{cases} \quad (16)$$

Here, the case of $\alpha = 0$, was obtained by taking the limit $\alpha \rightarrow 0$ of the first case $\alpha > 0$ and the resulting divergence is the Kulback-Leibler divergence.

Consider the *i.i.d* sample X_1, \dots, X_n from a model distribution function F of which θ is an

unknown parameter of interest. The minimum density power divergence (MDPD) estimator of θ is obtained by minimising the divergence between the data and the model density

$$d_{\theta}(f, g) = \begin{cases} f^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{j=1}^n f^{\alpha}(X_j), & \text{if } \alpha > 0 \\ -\frac{1}{n} \sum_{j=1}^n \log f(X_j), & \text{if } \alpha = 0. \end{cases} \quad (17)$$

The MDPD estimator of the parameters of the extended Pareto distribution, (10), are obtained from the following system of equations

$$\int_1^{\infty} f^{\alpha}(x) \frac{\partial f(x)}{\partial \gamma} dx - \left(1 + \frac{1}{\alpha}\right) \frac{1}{k} \sum_{j=1}^k f^{\alpha-1} \frac{\partial f(X_j)}{\partial \gamma} = 0 \quad (18)$$

and

$$\int_1^{\infty} f^{\alpha}(x) \frac{\partial f(x)}{\partial \delta} dx - \left(1 + \frac{1}{\alpha}\right) \frac{1}{k} \sum_{j=1}^k f^{\alpha-1} \frac{\partial f(X_j)}{\partial \delta} = 0. \quad (19)$$

The estimating equations (18) and (19) depend on the unknown parameter τ , which is obtained in Dierckx et al. (2013) using the reparametrisation, $\tau = \rho/\gamma$. The asymptotic normality of these estimators are shown in that paper.

In the case of the exponential regression model, described in Section 2.2, the weighted log-spacings of order statistics, Z_i , $i = 1, \dots, k-1$, in (14) each has distribution function F_{θ_i} and corresponding density functions f_{θ_i} . Although the Z_i 's are independent having approximate density f_{θ_i} , an exponential distribution, they are not identically distributed. Note that $\theta_i = \gamma + b_{n,k} (i/k + 1)^{-\rho}$, and hence, it is linear function of γ and non-linear functions of the other parameters, $b_{n,k}$ and ρ .

The minimum density power estimator for the parameters γ , $b_{n,k}$ and ρ , can be obtained by following Ghosh and Basu (2013) and Ghosh (2017) viz. by minimisation of the function

$$H_k(\theta) = \frac{1}{k-1} \sum_{i=1}^{k-1} \left[\int \left[f_{\theta_i}^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) f_{\theta_i}^{\alpha} \hat{g}_i \right] \right], \quad (20)$$

where \hat{g}_i is a non-parametric estimator of g_i from the observed sample. Since there is only one observation for each density, g_i , Ghosh (2017) states that the best possible nonparametric estimator \hat{g}_i of g_i , is given by the non-degenerate distribution at Y_i . Then, rewriting (20) using the exponential density, we obtain as in Ghosh (2017),

$$H_k(\gamma, b, \rho) = \frac{1}{k-1} \sum_{i=1}^{k-1} \left[\frac{1}{(1+\alpha)\theta_i^{\alpha}} - \frac{1+\alpha}{\alpha\theta_i^{\alpha}} \exp\left(-\frac{\alpha z_i}{\theta_i}\right) \right] \quad (21)$$

where $\theta_i = \gamma + b \left(\frac{i}{k+1}\right)^{-\rho}$ with $b = b_{n,k}$. The parameters $\eta = (\gamma, b, \rho)$ can then be obtained by minimising the objective function (21). Alternatively, these estimators can be obtained by solving the estimating equations $\frac{\partial H_k}{\partial \eta} = \mathbf{0}$, given by

$$\frac{1}{k-1} \sum_{i=1}^{k-1} \frac{(1+\alpha)}{\theta_i^{\alpha+2}} \cdot \left[\frac{\alpha\theta_i}{(1+\alpha)^2} + (z_i - \theta_i) e^{-\frac{\alpha z_i}{\theta_i}} \right] \frac{\partial \theta_i}{\partial \eta} = \mathbf{0}. \quad (22)$$

Taking the derivatives in (22) give

$$\frac{1}{k-1} \sum_{i=1}^{k-1} \mathbf{J}_\alpha \left(\frac{i}{k+1}, \boldsymbol{\eta} \right) \cdot \left[\frac{\alpha \theta_i}{(1+\alpha)^2} + (z_i - \theta_i) e^{-\frac{\alpha z_i}{\theta_i}} \right] = \mathbf{0}, \quad (23)$$

where, for any $\mu \in (0, 1)$, we define $\mathbf{J}_\alpha(\mu, \boldsymbol{\eta}) = (J_{1,\alpha}(\mu, \boldsymbol{\eta}), J_{2,\alpha}(\mu, \boldsymbol{\eta}), J_{3,\alpha}(\mu, \boldsymbol{\eta}))'$ with

$$J_{1,\alpha}(\mu, \boldsymbol{\eta}) = \frac{(1+\alpha)}{(\gamma + b\mu^{-\rho})^{\alpha+2}}, \quad (24)$$

$$J_{2,\alpha}(\mu, \boldsymbol{\eta}) = \frac{(1+\alpha)\mu^{-\rho}}{(\gamma + b\mu^{-\rho})^{\alpha+2}}, \quad (25)$$

$$J_{3,\alpha}(\mu, \boldsymbol{\eta}) = -\frac{(1+\alpha)b\mu^{-\rho} \log(\mu)}{(\gamma + b\mu^{-\rho})^{\alpha+2}}. \quad (26)$$

3 Robustness of the Proposed Estimators

3.1 Influence Function Analysis

Hampel (1968) provides a classical tool for measuring robustness known as influence functions. This function gives a first-order approximation of the asymptotic bias of any estimator under contamination by an outlying observation. In practice, we seek a bounded influence function and hence, such an estimator's bias will not increase indefinitely under contamination by very far outlying point.

In order to derive the influence function for our proposed estimator, we first need to define it in terms of statistical functionals. Let Z_i have a true distribution G_i with density g_i for each $i = 1, 2, \dots, k-1$. Denote $\mathbf{G} = (G_1, G_2, \dots, G_{k-1})'$ and $\hat{\mathbf{G}} = (\hat{G}_1, \hat{G}_2, \dots, \hat{G}_{k-1})'$ where \hat{G}_i is the empirical distribution function corresponding to G_i for $i = 1, 2, \dots, k-1$. Then it is easy to see that our minimum density power estimator of $\boldsymbol{\eta} = (\gamma, b, \rho)$ is given by $\hat{\boldsymbol{\eta}}_\alpha = T_\alpha(\hat{\mathbf{G}})$ where $T_\alpha(\hat{\mathbf{G}})$ is the corresponding statistical functional defined as the solution of the following population estimation equation,

$$\frac{1}{k-1} \sum_{i=1}^{k-1} \mathbf{J}_\alpha \left(\frac{i}{k+1}, \boldsymbol{\eta} \right) \cdot \left[\frac{\alpha \theta_i}{(1+\alpha)^2} + \int (z - \theta_i) e^{-\frac{\alpha z}{\theta_i}} g_i(z) dz \right] = 0. \quad (27)$$

Now, following Ghosh and Basu (2013) and Ghosh (2017), we may assume contamination to be in any one or more (even all) g_i 's. For simplicity, let us first assume the contamination is only in g_{i_0} for some fixed $i_0 \in \{1, 2, \dots, k-1\}$ and the corresponding contaminated density and distribution functions are $g_{i_0, \varepsilon} = (1 - \varepsilon)g_{i_0} + \varepsilon \delta_{t_{i_0}}$ and $G_{i_0, \varepsilon} = (1 - \varepsilon)G_{i_0} + \varepsilon \Delta_{t_{i_0}}$ respectively, where $\delta_{t_{i_0}}$ and $\Delta_{t_{i_0}}$ are the density and distribution functions of a degenerate distribution at the contamination point t_{i_0} and ε is the contamination proportion. Then, $\boldsymbol{\eta}_\varepsilon = T_\alpha(G_1, \dots, G_{i_0-1}, G_{i_0, \varepsilon}, G_{i_0+1}, \dots, G_{k-1})$ satisfy the estimating equation (27) with g_{i_0} replaced with $g_{i_0, \varepsilon}$ and $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta}_\varepsilon$ (so that θ_i is replaced by the corresponding value, $\theta_{i, \varepsilon}$, computed from $\boldsymbol{\eta}_\varepsilon$ for all $i = 1, 2, \dots, k-1$), i.e. we have

$$\begin{aligned} & \frac{1}{k-1} \sum_{i=1}^{k-1} \mathbf{J}_\alpha \left(\frac{i}{k+1}, \eta_\varepsilon \right) \frac{\alpha \theta_{i,\varepsilon}}{(1+\alpha)^2} + \frac{1}{k-1} \sum_{i=1, i \neq i_0}^{k-1} \mathbf{J}_\alpha \left(\frac{i}{k+1}, \eta_\varepsilon \right) \int (z - \theta_{i,\varepsilon}) e^{-\frac{\alpha z}{\theta_{i,\varepsilon}}} g_i(z) dz + \\ & \mathbf{J}_\alpha \left(\frac{i_0}{k+1}, \eta_\varepsilon \right) \int (z - \theta_{i_0,\varepsilon}) e^{-\frac{\alpha z}{\theta_{i_0,\varepsilon}}} g_{i_0,\varepsilon}(z) dz = 0. \end{aligned} \quad (28)$$

Differentiating (28) with respect to ε at $\varepsilon = 0$ and evaluating terms, we get the required partial influence function of T_α under contamination only at the i_0 -th density as given by

$$IF(t_{i_0}, T_\alpha, G) = \left. \frac{\delta \eta_\varepsilon}{\delta \varepsilon} \right|_{\varepsilon=0} = \frac{\Psi_n^{-1}(G)}{k-1} \mathbf{J}_\alpha \left(\frac{i_0}{k+1}, \eta^g \right) \left[\frac{\alpha \theta_{i_0}^g}{(1+\alpha)^2} + (t_{i_0} - \theta_{i_0}^g) e^{-\frac{\alpha t_{i_0}}{\theta_{i_0}^g}} \right] \quad (29)$$

where $\eta^g = T_\alpha(G)$, $\theta_{i_0}^g$ is the corresponding value of θ_{i_0} obtained from η^g and $\Psi_n(G)$ is defined from equations (3.3) and (3.5) of Ghosh and Basu (2013). In our case, we can simplify the form of $\Psi_n(G)$ for general distribution functions G_i s as given by

$$\begin{aligned} \Psi_n^{-1}(G) = & \frac{-1}{k-1} \sum_{i=1}^{k-1} \left[\left\{ M_1 \left(\frac{i}{k+1}, \eta^g \right) - \frac{(\alpha+2)}{\theta_i^g} M_2 \left(\frac{i}{k+1}, \eta^g \right) \right\} \frac{\alpha \theta_i^g}{(1+\alpha)} + \right. \\ & \left. \left\{ M_1 \left(\frac{i}{k+1}, \eta^g \right) - \frac{(\alpha+2)}{\theta_i^g} M_2 \left(\frac{i}{k+1}, \eta^g \right) \right\} \int (z - \theta_i^g) e^{-\frac{\alpha z}{\theta_i^g}} g_i(z) dz + \right. \\ & M_2 \left(\frac{i}{k+1}, \eta^g \right) \left\{ \frac{\alpha}{1+\alpha} - \int e^{-\frac{\alpha z}{\theta_i^g}} g_i(z) dz + \right. \\ & \left. \left. \frac{\alpha}{(\theta_i^g)^2} \int z(z - \theta_i^g) e^{-\frac{\alpha z}{\theta_i^g}} g_i(z) dz \right\} \right], \end{aligned} \quad (30)$$

where

$$M_1(\mu, \eta) = \frac{(1+\alpha)}{(\gamma + b\mu^{-\rho})^{\alpha+2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\log(\mu) \cdot \mu^{-\rho} \\ 0 & -\mu^{-\rho} \log(\mu) & b\mu^{-\rho} (\log(\mu))^2 \end{bmatrix},$$

and

$$M_2(\mu, \eta) = \frac{(1+\alpha)}{(\gamma + b\mu^{-\rho})^{\alpha+2}} \begin{bmatrix} 1 & \mu^{-\rho} & -b\mu^{-\rho} \log(\mu) \\ \mu^{-\rho} & \mu^{-2\rho} & -b\mu^{-2\rho} \log(\mu) \\ -b\mu^{-\rho} \log(\mu) & -b\mu^{-2\rho} \log(\mu) & b^2 \mu^{-2\rho} (\log(\mu))^2 \end{bmatrix}.$$

To illustrate the influence function for our estimators, let us simplify the influence function for the case where exponential regression model, (15), is valid. In that case, let $G_i \equiv F_{\theta_i^0}$, the

exponential distribution with mean, θ_i^0 , computed from (true) parameter value $\eta^0 = (\gamma^0, b^0, \rho^0)$ for all $i = 1, 2, \dots, k-1$. In addition, denoting $\mathbf{F}_{\theta_i^0} = (F_{\theta_1^0}, F_{\theta_2^0}, \dots, F_{\theta_{k-1}^0})$, we have

$$T_\alpha(F_{\theta^0}) = \eta^0,$$

and

$$\begin{aligned} \Psi_n^0 &= \Psi_n(\mathbf{F}_{\theta_i^0}) \\ &= \frac{1}{k-1} \sum_{i=1}^{k-1} \left[\frac{(\alpha^4 + 2\alpha^3 + 2\alpha^2 + 1)}{(1+\alpha)^3} M_2\left(\frac{i}{k+1}, \eta^0\right) - \frac{\alpha^2 \theta_i^0}{(1+\alpha)^2} M_1\left(\frac{i}{k+1}, \eta^0\right) \right]. \end{aligned}$$

Then the influence function has the simplified form

$$IF(t_{i_0}, T_\alpha, \mathbf{F}_{\theta_i^0}) = \frac{(\Psi_n^0)^{-1}}{k-1} \mathbf{J}_\alpha\left(\frac{i_0}{k+1}, \eta^0\right) \left[\frac{\alpha \theta_{i_0}^0}{(1+\alpha)^2} + (t_{i_0} - \theta_{i_0}^0) e^{-\frac{\alpha t_{i_0}}{\theta_{i_0}^0}} \right]. \quad (31)$$

Considering the range of values for $\alpha > 0$, the influence function for our estimators is bounded. However, when $\alpha = 0$, the expression (31) simplifies to

$$IF(t_{i_0}, T_\alpha, \mathbf{F}_{\theta_i^0}) = \frac{(\Psi_n^0)^{-1}}{k-1} \mathbf{J}_0\left(\frac{i_0}{k+1}, \eta^0\right) (t_{i_0} - \theta_{i_0}^0). \quad (32)$$

Here, it is easy to see from (32) that the influence function is linear in terms of the contamination point t_{i_0} and hence it is unbounded. Therefore, we conclude that our estimators for the parameters γ , b and ρ of (15) obtained with $\alpha > 0$ are robust with respect to contamination at any or all of the Z_j 's compared with the maximum likelihood estimators obtained in Beirlant et al. (1999) corresponding to $\alpha = 0$.

The fixed-sample influence functions are shown in Figure 1 for various values of γ and at different contamination points, t_{i_0} . Clearly, the influence function is unbounded for $\alpha = 0$: it has a linear increasing function as the contamination point is more extreme. However, in the same Figures, the boundedness of the estimators can easily be seen as it becomes flatter when a point of contamination is farther detached from the bulk of the data.

In a similar way, if there is contamination in all densities as $g_{i,\varepsilon} = (1-\varepsilon)g_i + \varepsilon\Delta_{t_i}$ at contaminant point t_i , for $i = 1, 2, \dots, k-1$, with the corresponding distribution function being $G_{i,\varepsilon} = (1-\varepsilon)G_i + \varepsilon\Lambda_{t_i}$, then we define $\eta_\varepsilon = T_\alpha(G_1, G_2, \dots, G_{k-1})$ and proceed as before to obtain the corresponding influence function. When the model assumptions are correct, i.e. $G_i = F_{\theta_i^0}$ for all i , the total influence function of our proposed estimate of η at the contaminated points $\mathbf{t} = (t_1, t_2, \dots, t_{k-1})'$ has the form

$$IF_0(\mathbf{t}, T_\alpha, \mathbf{F}_{\theta_i^0}) = \frac{(\Psi_n^0)^{-1}}{k-1} \sum_{i=1}^{k-1} \mathbf{J}_\alpha\left(\frac{i}{k+1}, \eta^0\right) \left[\frac{\alpha \theta_i^0}{(1+\alpha)^2} + (t_i - \theta_i^0) e^{-\frac{\alpha t_i}{\theta_i^0}} \right]. \quad (33)$$

Again, clearly it can be seen from (33), that the influence function is bounded for all $\alpha > 0$. However, in the case where $\alpha = 0$, the influence function exhibits unboundedness as it is linear in terms of t_i , $i = 1, 2, \dots, k-1$.

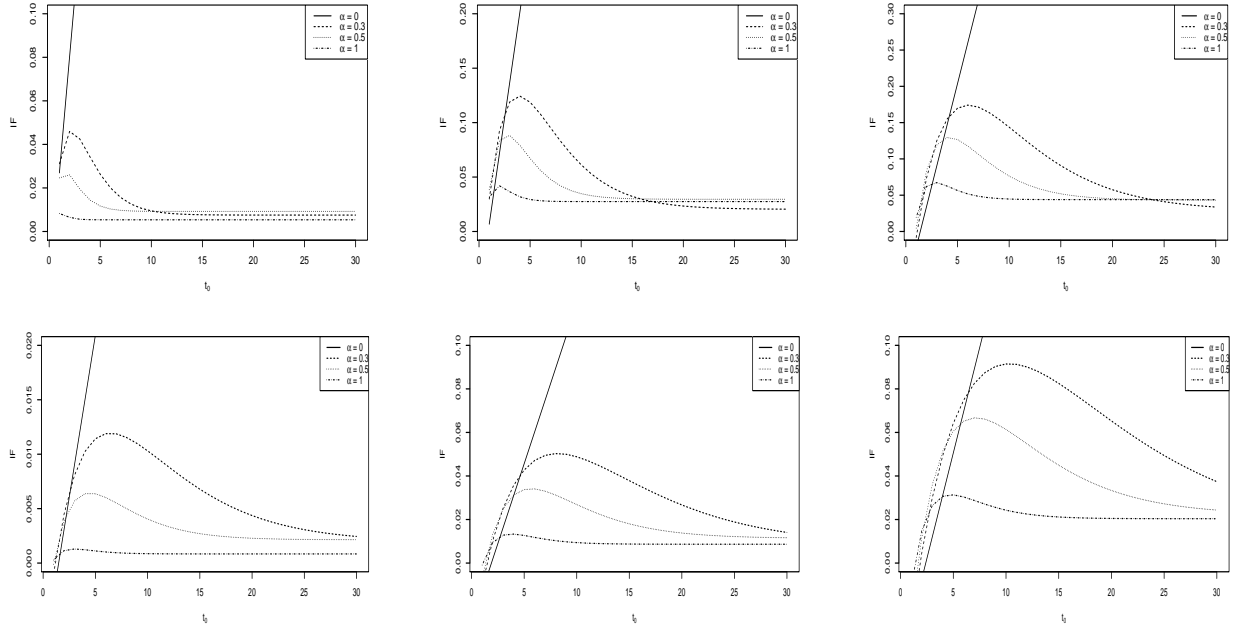


Figure 1: Influence function of T_α over contamination point with $k = 100$, $b = 2$, $\rho = -1$, Top panel: $i_0 = 20$; and bottom panel: $i_0 = 70$. Leftmost column: $\gamma = 0.1$; middlemost column: $\gamma = 0.5$; and rightmost column: $\gamma = 1$.

3.2 Gross Error Sensitivity

The Gross-Error Sensitivity (Hampel, 1968) measures the maximum possible values of the bias of an estimator under small infinitesimal contamination. Thus, the gross-error sensitivity is a supremum of the influence function and can be defined as

$$S(T_\alpha, \mathbf{G}) = \sup_t \{IF(t, T_\alpha, \mathbf{F})\}. \quad (34)$$

Therefore, in considering the effect of k and α on the robustness of our estimator T_α , the smaller the value of $S(T_\alpha, \mathbf{G})$, the more robust the estimator is to these parameters. In the case of contamination at a point Y_{i_0} only, the gross-error sensitivity of the proposed estimator T_α is given

by

$$S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0}) = \sup_{t_{i_0}} \left\| IF(t_{i_0}, T_\alpha, \mathbf{F}_{\theta_i^0}) \right\|$$

$$= \begin{cases} \frac{\theta_{i_0} [\alpha^2 + (1 + \alpha)^2 e^{-(1+\alpha)}]}{\alpha(1 + \alpha)^2(k - 1)} \sqrt{J_\alpha\left(\frac{i_0}{k+1}, \eta^0\right)^t (\Psi_n^0)^{-2} J_\alpha\left(\frac{i_0}{k+1}, \eta^0\right)}, & \text{if } \alpha > 0 \\ \infty, & \text{if } \alpha = 0. \end{cases}$$

Figure 2 presents the values of the sensitivity measures $S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0})$ over the parameter α for selected values of k . It can be seen that the value of $S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0})$ decreases with increasing values of α . In addition, the $S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0})$ decreases with increasing k .

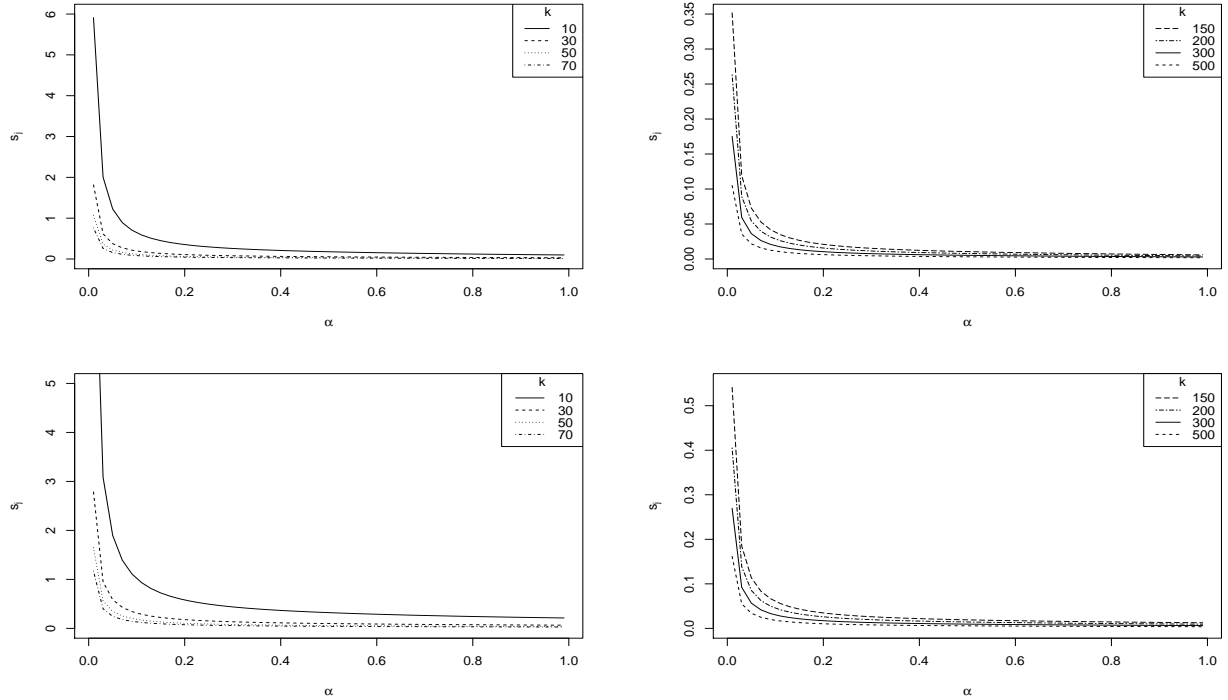


Figure 2: Gross Error Sensitivity $S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0})$ over the tuning parameters α and k for estimating $\gamma = 0.5$ for contamination direction $i_0 = k/2$ (top panel) and $i_0 = k/10$ (bottom panel).

Thus, $S_{i_0}(T_\alpha, \mathbf{F}_{\theta_i^0})$ decreases as α and k increase. Furthermore, the sensitivity values decrease sharply for $\alpha < 0.2$. However, it has smaller and near constant values for $\alpha > 0.2$. These imply that our proposed estimators show strong robustness properties for increasing values of α and k .

Similarly, the sensitivity for contamination in at least two observations can be obtained but has been omitted here for ease of presentation.

4 Simulation Study

In this section, we compare the performance of our proposed estimator with the equivalent minimum density power divergence estimators of the Pareto-type tail index in the literature. Specifically, the proposed exponential regression model estimator based on log-spacings of order statistics, ERM_M, the Dierckx et al. (2013) estimator obtained from fitting an extended Pareto distribution to relative excesses, EPD_D, and the Ghosh (2017) exponential regression model estimator based on log-ratio of order statistics, ERM_G.

4.1 Simulation Design

We consider three distributions in the Fréchet-Pareto domain of attraction namely the Fréchet, Pareto, and Burr. The distribution function, $1 - F$, tail index, γ , and the second-order parameter, ρ , of the distributions used in the simulation study are presented in Table 1.

Table 1: Distributions and their tail indexes

Distribution	$1 - F(x)$	γ	ρ
Burr	$(\eta / (\eta + x^\tau))^\lambda, x > 0, \eta, \tau, \lambda > 0$	$1 / (\tau\lambda)$	$-1/\lambda$
Fréchet	$1 - \exp(-x^{-\beta}), x > 0, \beta > 0$	$1/\beta$	-1
Pareto	$x^{-\beta}, x > 0, \beta > 0$	$1/\beta$	-1

For each distribution F , we generated samples from a mixture contaminated model: $(1 - \varepsilon)F + \varepsilon G$ where G is a nuisance distribution. Specifically, G is chosen in two ways: from the same distribution as F but with different parameters and a different distribution from F . In each case, we assess the robustness of the estimators under different contamination scenarios with $\varepsilon = 0.05$ and $\varepsilon = 0.15$. Furthermore, to assess the effect of the robustness parameter, we take three values of α , at 0.1, 0.5 and 1 representing levels for increased robustness.

4.2 Discussion of Simulation Results

This section discusses the behaviour of the proposed estimator and the two existing estimators of the tail index in the case where contamination of the base distribution comes from the same distribution but with different parameters. Here the contaminating distribution's parameter is chosen such that the observations are generally distinct from the bulk of the data. The results of the simulation studies for the Burr distribution contaminated by another Burr but with different tail index are presented in Figures 3-5 for the Mean Square Error (MSE) and Figures 6-8 for the bias.

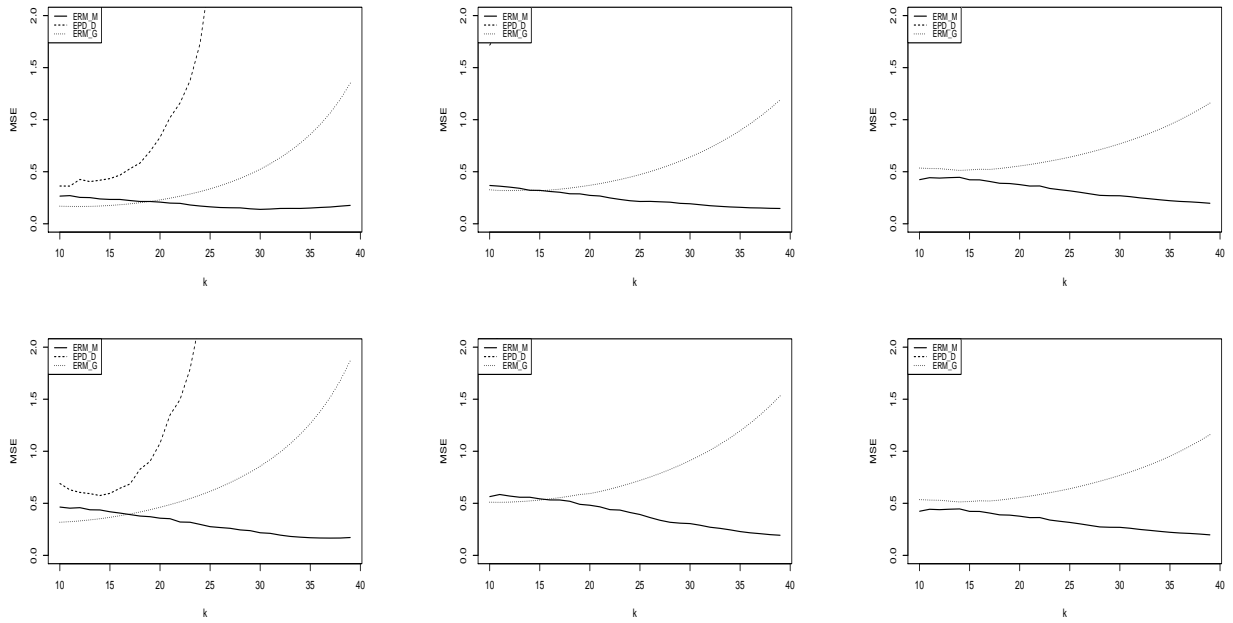


Figure 3: Burr distribution with $n = 50$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

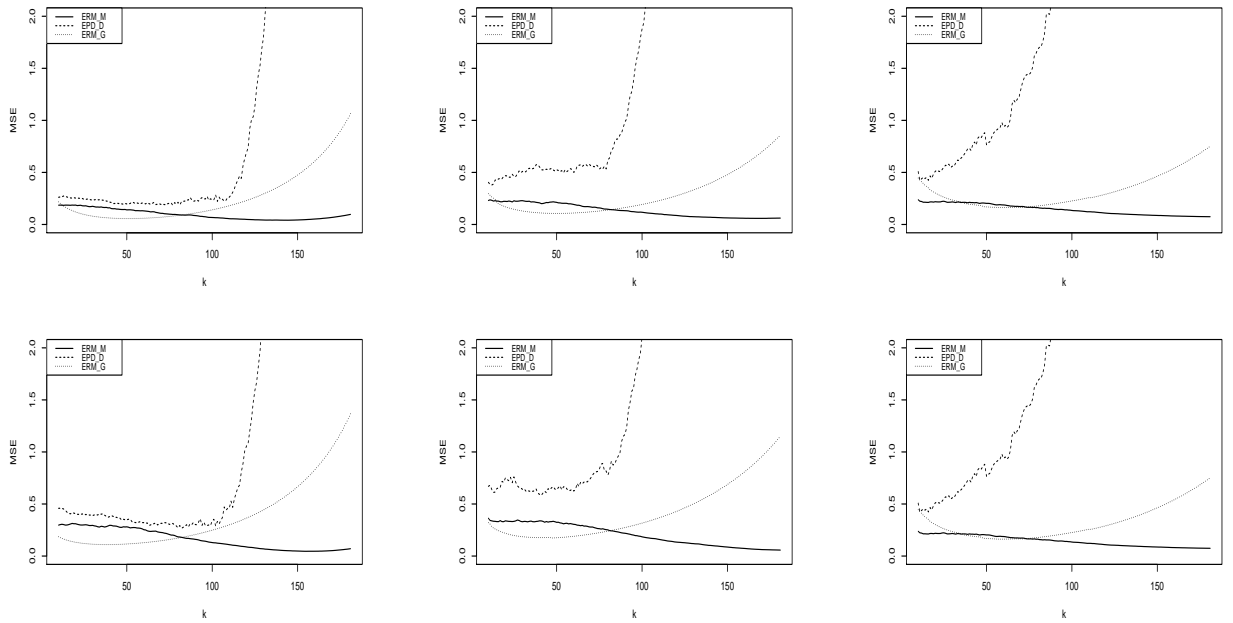


Figure 4: Burr distribution with $n = 200$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

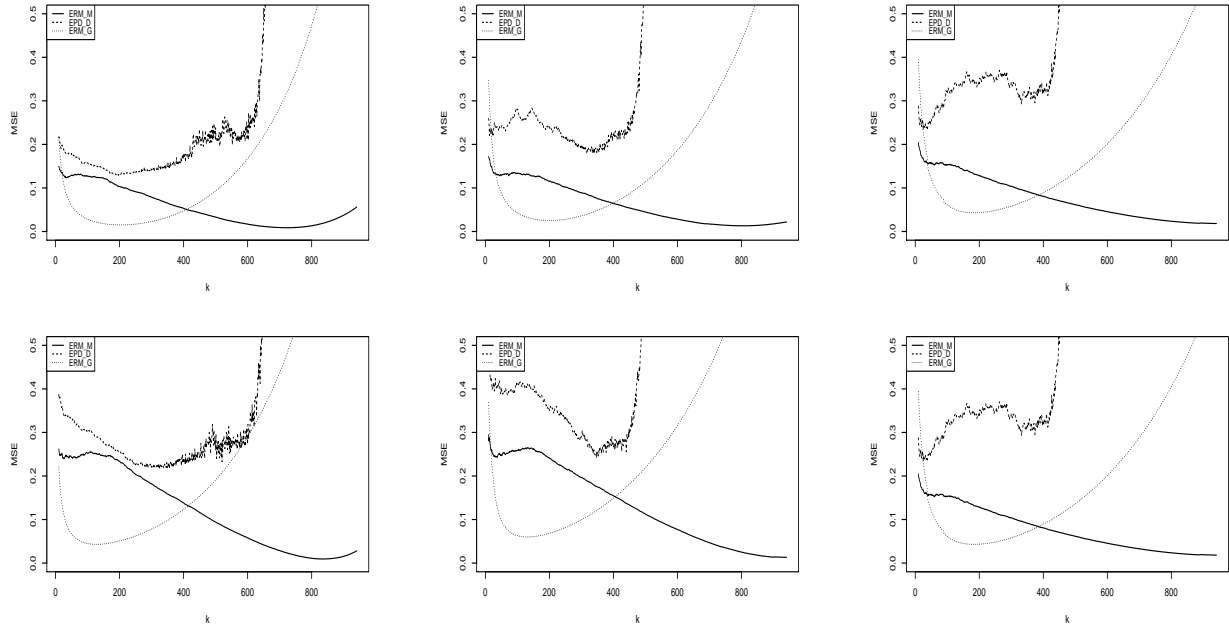


Figure 5: Burr distribution with $n = 1000$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

From these Figures, the proposed ERM_M estimator shows clear improvement on MSE and bias than the ERM_G and EPD_D across the three robust tuning parameters as well as the percentage of contamination. However, for smaller values of k , the ERM_G estimator seems to provide better MSEs than the ERM_M. This can be explained as the ERM_G estimator does not involve second-order parameters and hence should in theory have less variation.

In addition, the performance of the estimators in the case of samples generated from other distributions (i.e. Fréchet and Pareto) are presented in Appendices A.1 and A.2. In the case of the Fréchet distribution, the proposed ERM_M and EPD_G estimators are by far better than the ERM_G. The two estimators, ERM_M and EPD_G, have approximately equal performance under MSE, with the ERM_M slightly on top for larger values of k . However, in terms of bias, the ERM_M is the preferred estimator as it has the least values across the sample sizes, percentage of contamination and the robust parameters.

Similar performance can be seen of the ERM_M estimator in the case of samples generated from the Pareto distribution. It has smaller values of the MSE in most cases. However, unlike the other distributions, the ERM_G estimator is quite competitive and can be considered as an appropriate estimator of the tail index than the EPD_G with bias and MSE values comparable to the ERM_M estimator.

Therefore, the simulation results indicate that across the different distributions and factors such as percentage of contamination and robustness levels, the ERM_M is found to be generally a better alternative to the EPD_D and the ERM_G estimators.

Lastly, the R codes for the computation of the tail index, implementation of the influence function analysis and sensitivity measures can be found at <https://github.com/rminkah/>

RobustTailIndex.

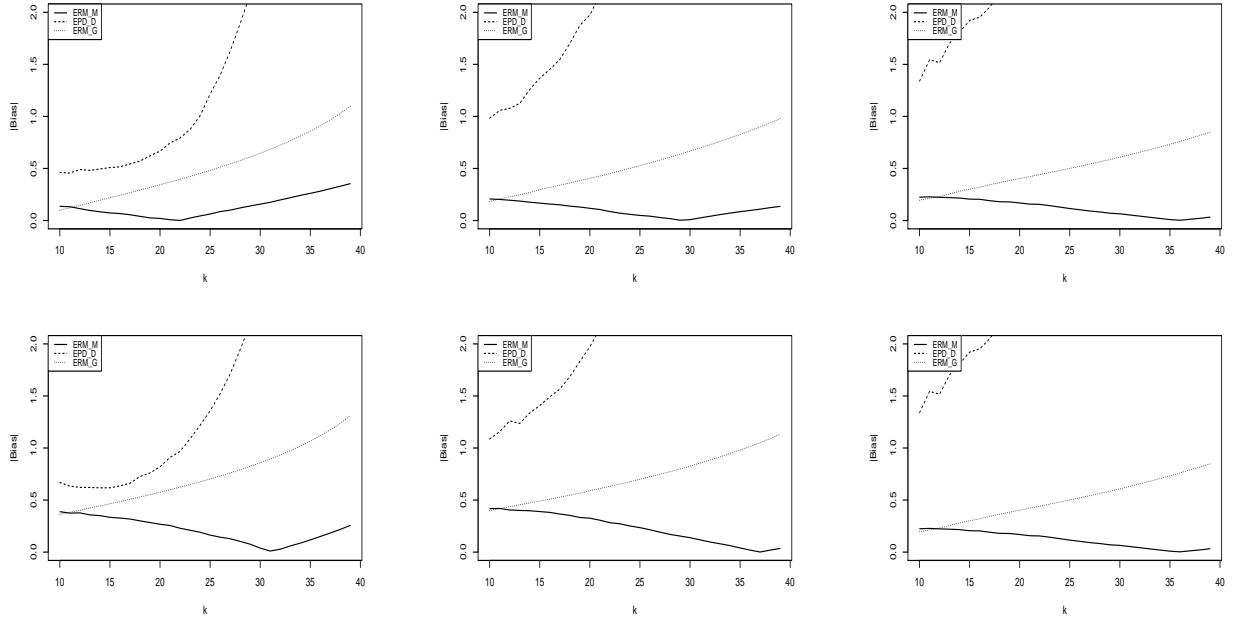


Figure 6: Burr distribution with $n = 50$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

5 Application

In this section, we estimate the tail index of the Society of Actuaries (SOA) Group Medical Insurance data studied in Beirlant et al. (2004, Chapters 1 and 5) and can be found at <https://lstat.kuleuven.be/Wiley/Data/soa.txt>. However, all the estimators used were non-robust including maximum likelihood estimator based on perturbed Pareto distribution and exponential regression model. In this section, we illustrate the application of the proposed robust minimum density power divergence estimator of the tail index based on log-spacings of order statistics discussed in the previous section in estimating the tail index of the SOA data.

The plot of the data in Figure 9 shows that two particular large claims seem to be detached from the bulk of the data. Such outliers have implication on traditional methods of estimation of the parameters of the GP distribution such as maximum likelihood and probability weighted moments. Using different robust tuning parameters, we compute the tail index as a function of the number of top order statistics, k . The result reveals that our proposed estimator, ERM_M, is mostly stable along the path of k compared to the robust estimator of Dierckx et al. (2013) based on the extended Pareto distribution, EPD_D. Also, in conformity with the behaviour of robust estimators, the variation in the estimates increases with increasing α . Therefore, ERM_M provides a better alternative robust estimator for the tail index in the Pareto domain as illustrated with the SOA data.

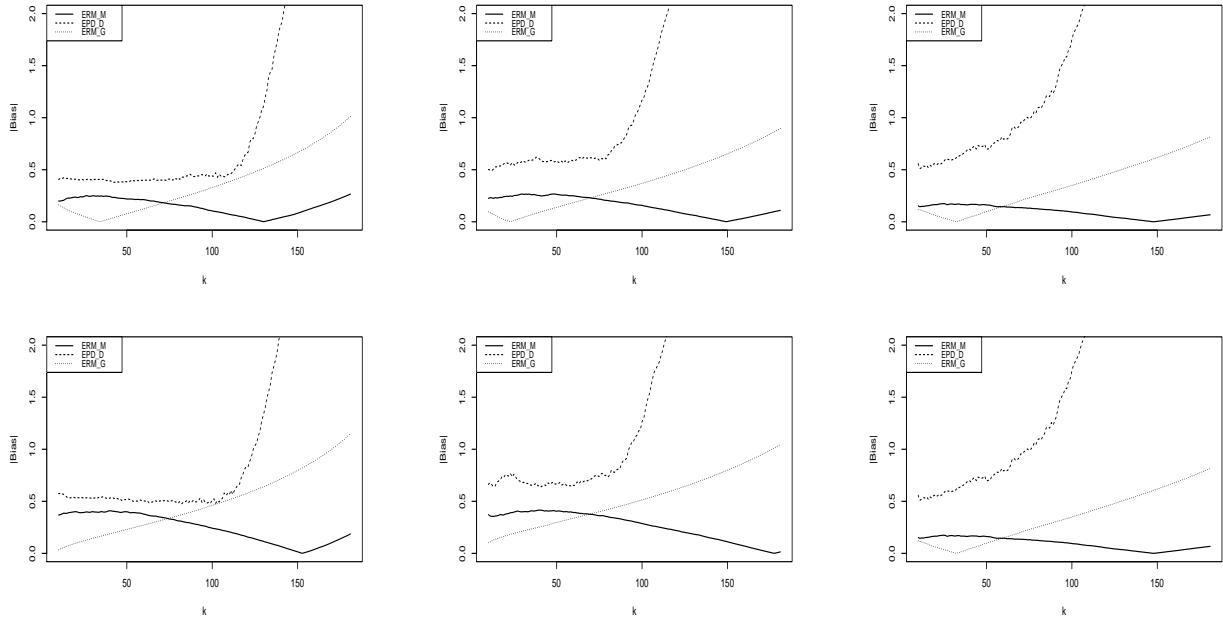


Figure 7: Burr distribution with $n = 200$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

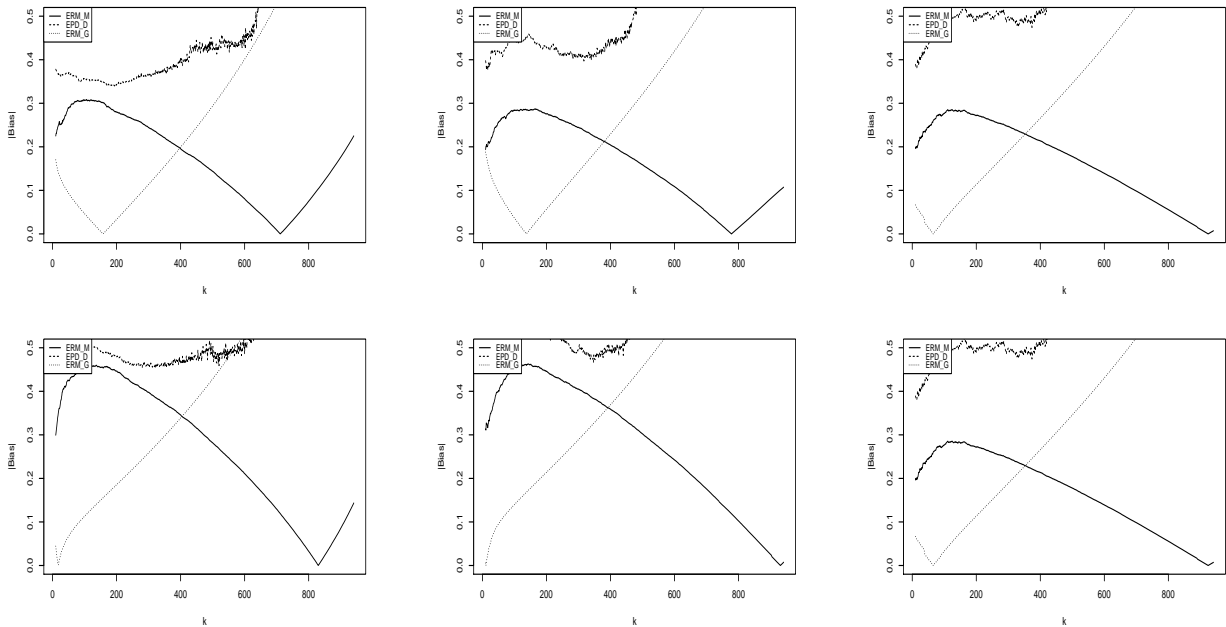


Figure 8: Burr distribution with $n = 1000$ and $\gamma = 0.5$. Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

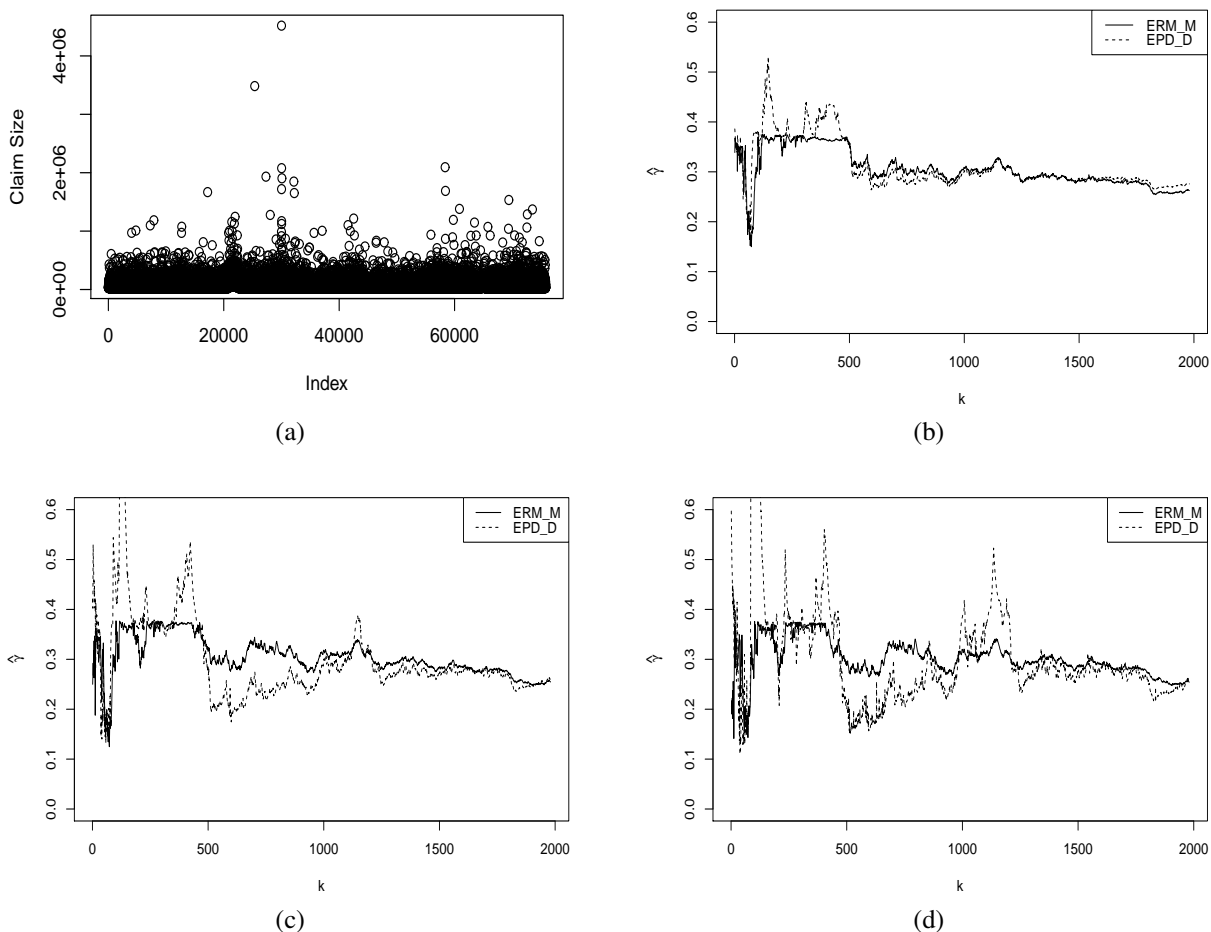


Figure 9: SOA data: (a) scatter plot (b) γ estimates with $\alpha = 0.1$ (c) γ estimates with $\alpha = 0.5$ and (d) γ estimates with $\alpha = 1$

6 Conclusion

In this paper, we proposed a robust estimator of tail index using the minimum density power divergence through an exponential regression model. The estimator is valid for the Pareto domain of attraction, i.e. heavy-tailed distributions. The robustness aspect of this estimator was studied analytically by deriving its influence function and gross error sensitivity functions. In addition, the finite sample properties of the estimator was studied through a simulation study together with a similar estimator using minimum density power divergence but on an extended Pareto distribution fitted to relative excesses and an exponential regression model estimator based on log-spacings of order statistics. The results of the simulation study show that the proposed minimum density power estimator based on an exponential regression model on log-spacings of order statistics generally has better performance in terms of mean square errors and bias than the existing estimators. In addition, the proposed robust estimator of the tail index is less sensitive to the number of top order statistics. Lastly, the proposed estimator is illustrated on real-data set

on insurance claims. The theoretical properties of the proposed estimators, estimation of high quantiles and exceedance probabilities are the subjects for future research.

Acknowledgements

The first named author (RM) was supported in the manuscript preparation by the University of Ghana Building a New Generation of Academics in Africa (BANGA-Africa) Project with funding from the Carnegie Corporation of New York. The research of the second named author (TdW) was supported by the South African NRF under grant number 115038.

References

- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5):792–804.
- Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.
- Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2:177–200.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, England.
- Beirlant, J., Joossens, E., and Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139(3):2800–2815.
- Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal*, 4:12–27.
- Brazauskas, V. and Serfling, R. (2001). Small sample performance of robust estimators of tail parameters for Pareto and exponential models. *Journal of Statistical Computation and Simulation*, 70(1):1–19.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52(3):393–442.
- Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.
- Dell’Aquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets Portfolio Management*, 20:103–118.
- Dierckx, G., Goegebeur, Y., and Guillou, A. (2013). An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *Journal of Multivariate Analysis*, 121:70–86.

- Dierckx, G., Goegebeur, Y., and Guillou, A. (2018). Local robust estimation of Pareto-type tails with random right censoring. *HAL archives*, (01829618).
- Dupius, D. and Field, C. (1998). Robust estimation of of extremes. *The Canadian Journal of Statistics*, 26(2):199–215.
- Fisher, R. and Tippett, L. (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:80–190.
- Ghosh, A. (2017). Divergence based robust estimation of the tail index through an exponential regression model. *Statistical Methods & Applications*, 26(2):181–213.
- Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal Statistics*, 7:2420–2456.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44(3):423–453.
- Goegebeur, Y., Guillou, A., and Verster, A. (2014). Robust and asymptotically unbiased estimation of extreme quantiles for heavy tailed distributions. *Statistics and Probability Letters*, 87(1):108–114.
- Hampel, F. R. (1968). The influence curve and its role in robust estimation. *Journal of American Statistical Association*, 69:383–393.
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174.
- Juarez, S. F. and Schucany, W. R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, 7:237–251.
- Kim, M. and Lee, S. (2008). Estimation of a tail index based on minimum density power divergence. *Journal of Multivariate Analysis*, 99(10):2453–2471.
- Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 13:853–880.
- Peng, L. and Welsh, A. (2001). Robust estimation of the generalized Pareto distribution. *Extremes*, 4:53–65.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The annals of Statistics*, 3(1):119–131.
- Vandewalle, B., Beirlant, J., Christmann, A., and Hubert, M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics & Data Analysis*, 51:6252–6268.

A Simulation Results for other Distributions

For each figure in the sections of the appendix, the following description applies: Topmost row: $\varepsilon = 0.05$; and bottommost row: $\varepsilon = 0.15$. Leftmost column: $\alpha = 0.1$; middlemost column: $\alpha = 0.5$; and rightmost column: $\alpha = 1$.

A.1 Fréchet Distribution

A.1.1 Mean Square Error

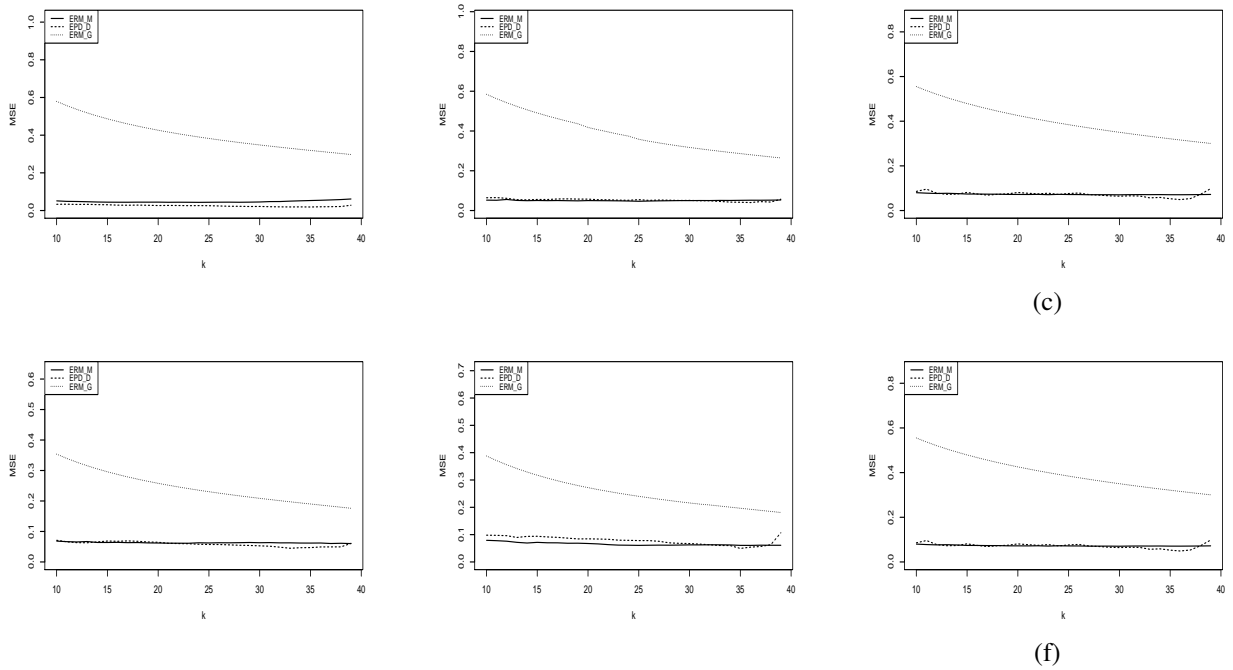
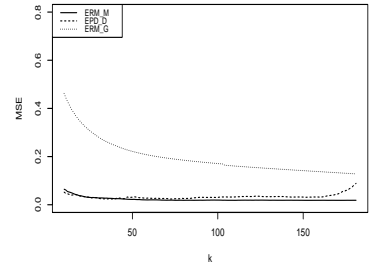
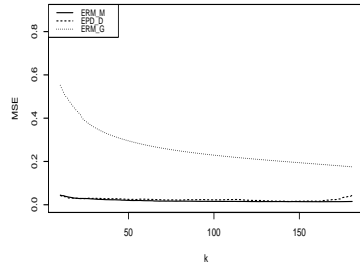
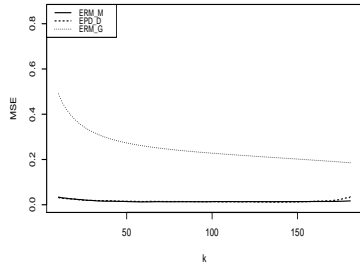
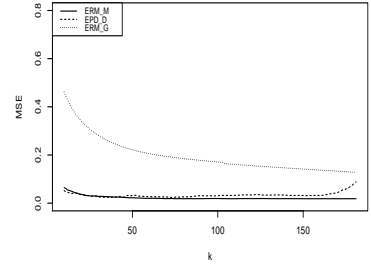
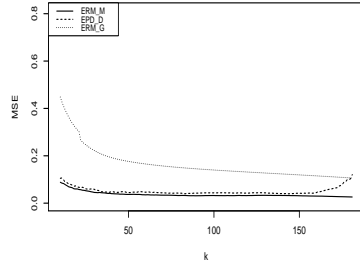
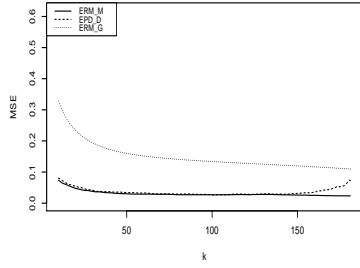


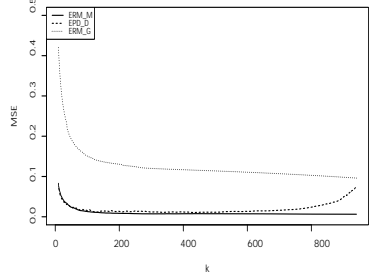
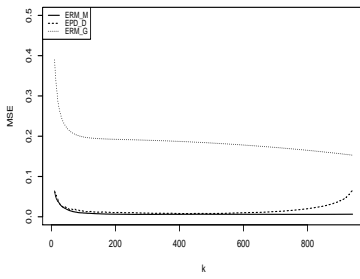
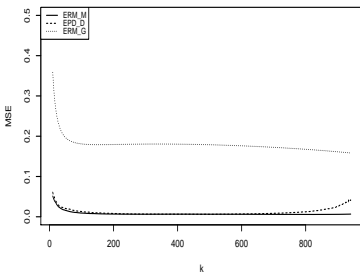
Figure 10: Fréchet distribution with $n = 50$ and $\gamma = 0.5$.



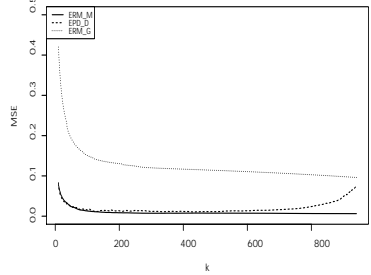
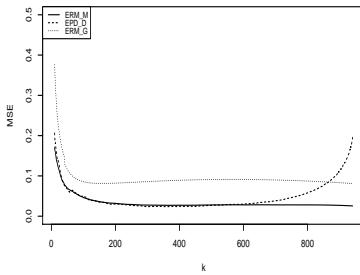
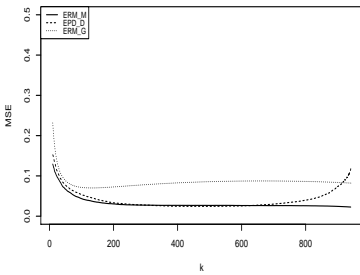
(c)



(f)

Figure 11: Fréchet distribution with $n = 200$ and $\gamma = 0.5$.

(c)



(f)

Figure 12: Fréchet distribution with $n = 1000$ and $\gamma = 0.5$.

A.1.2 Bias

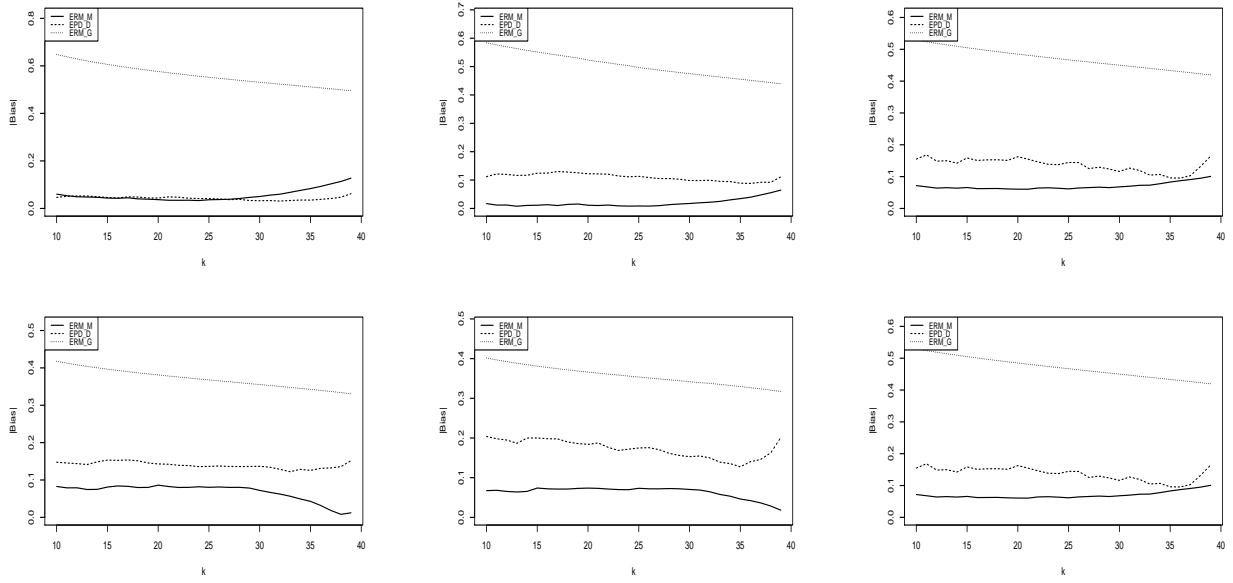


Figure 13: Fréchet distribution with $n = 50$ and $\gamma = 0.5$.

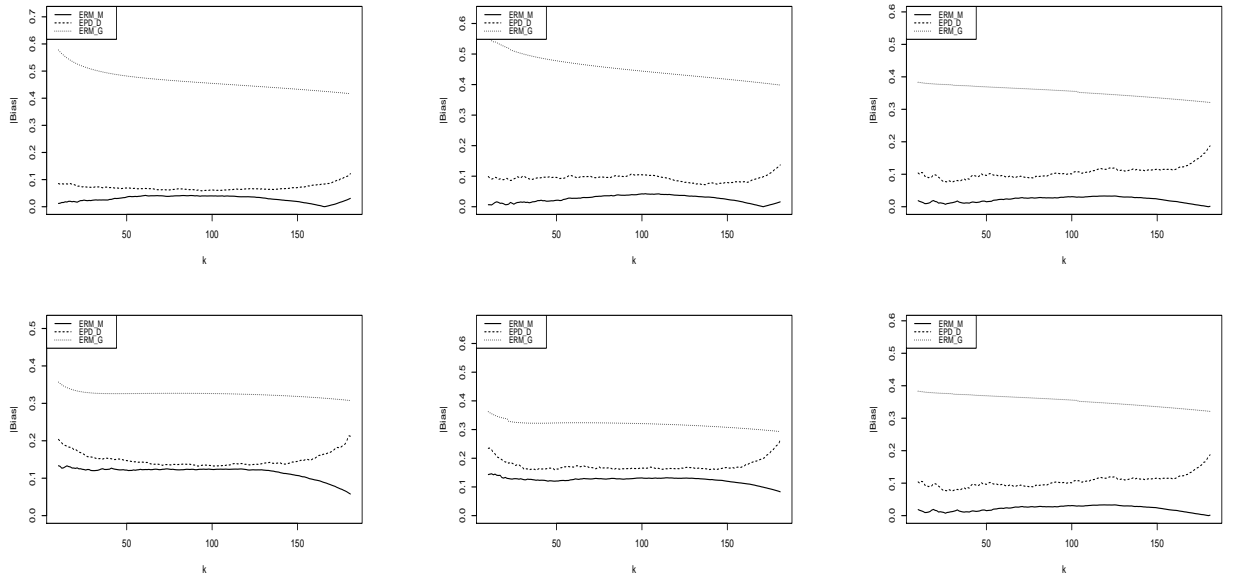


Figure 14: Fréchet distribution with $n = 200$ and $\gamma = 0.5$.

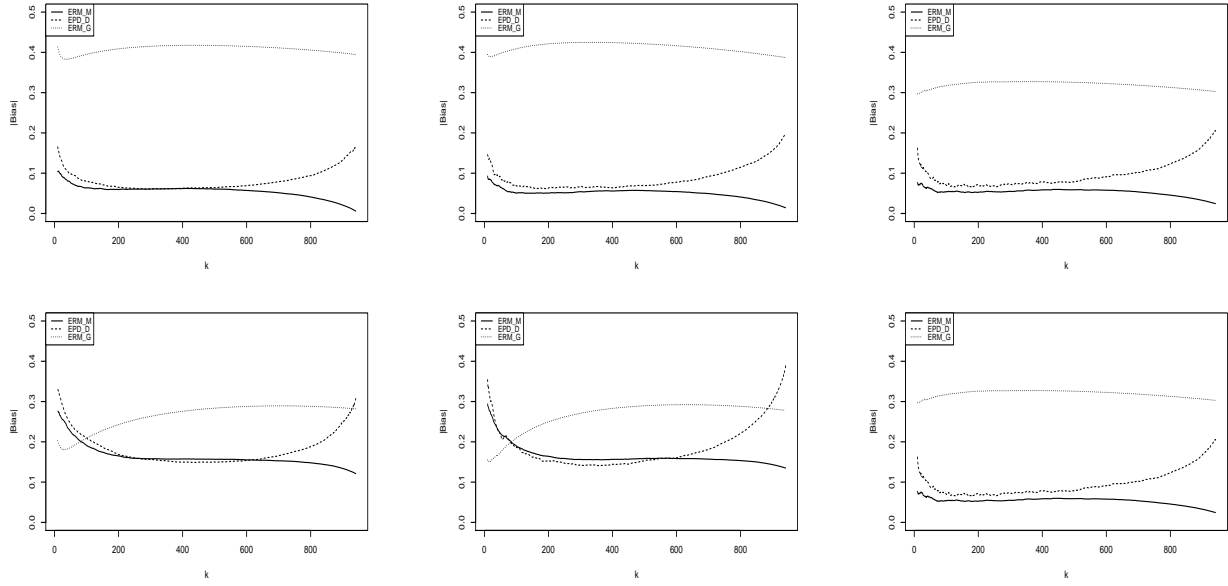


Figure 15: Fréchet distribution with $n = 1000$ and $\gamma = 0.5$.

A.2 Pareto Distribution

A.2.1 Mean Square Errors

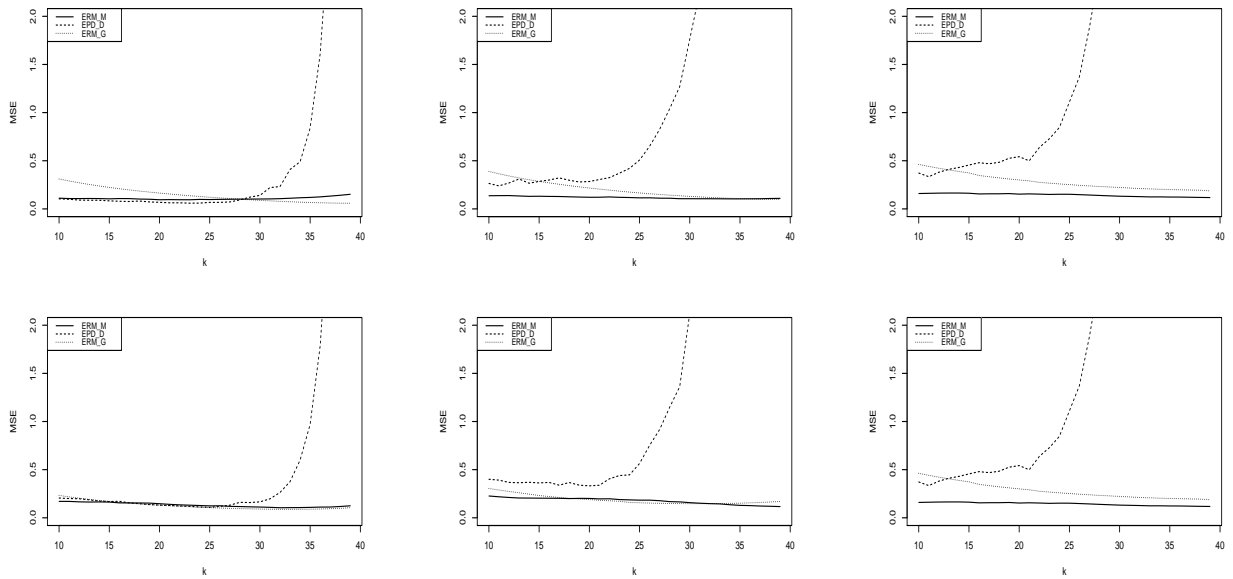


Figure 16: Pareto distribution with $n = 50$ and $\gamma = 0.5$.

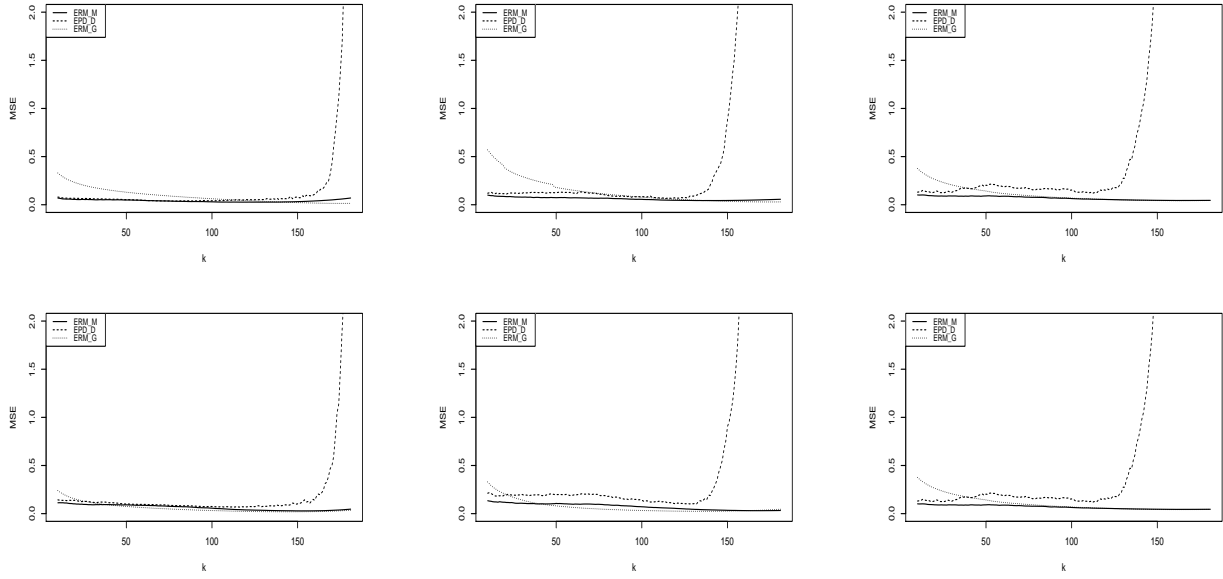


Figure 17: Pareto distribution with $n = 200$ and $\gamma = 0.5$.

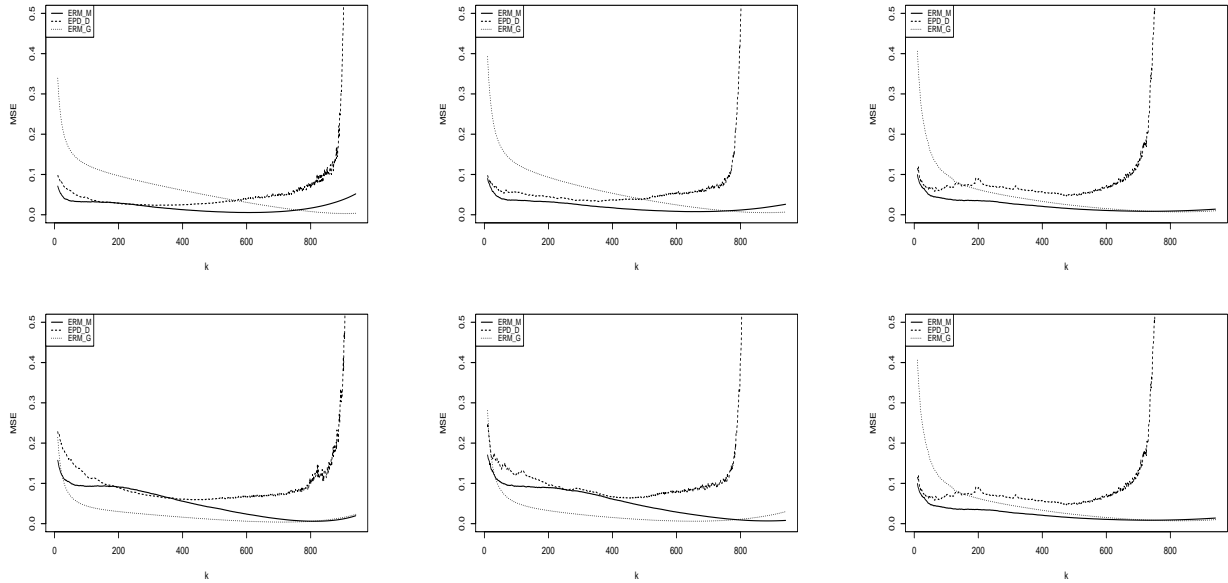


Figure 18: Pareto distribution with $n = 1000$ and $\gamma = 0.5$.

A.2.2 Bias

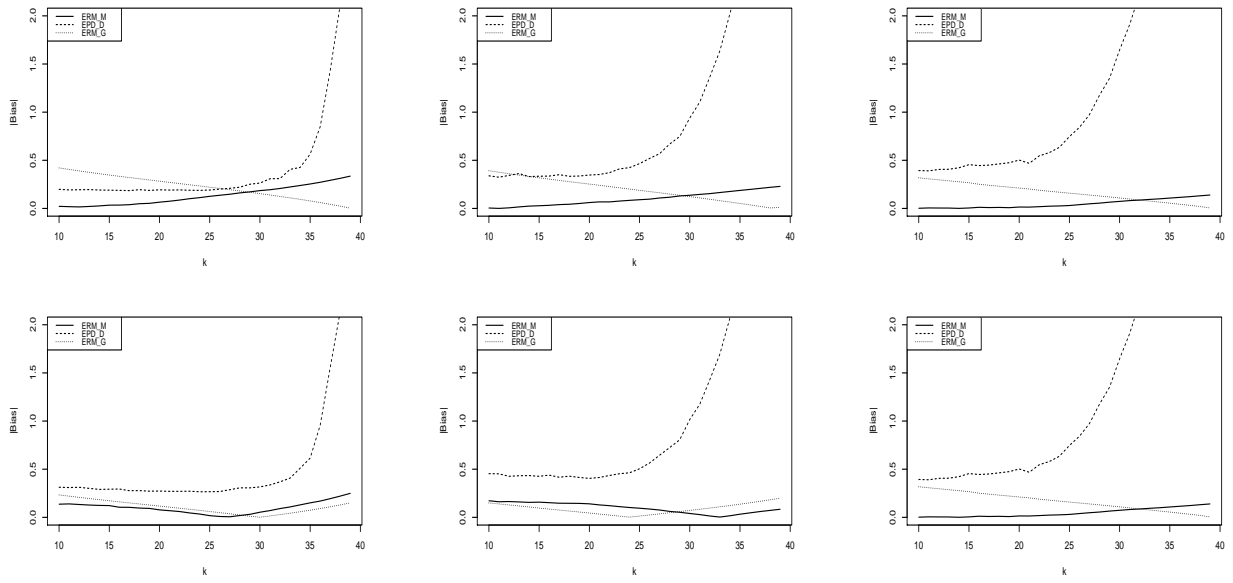


Figure 19: Pareto distribution with $n = 50$ and $\gamma = 0.5$.

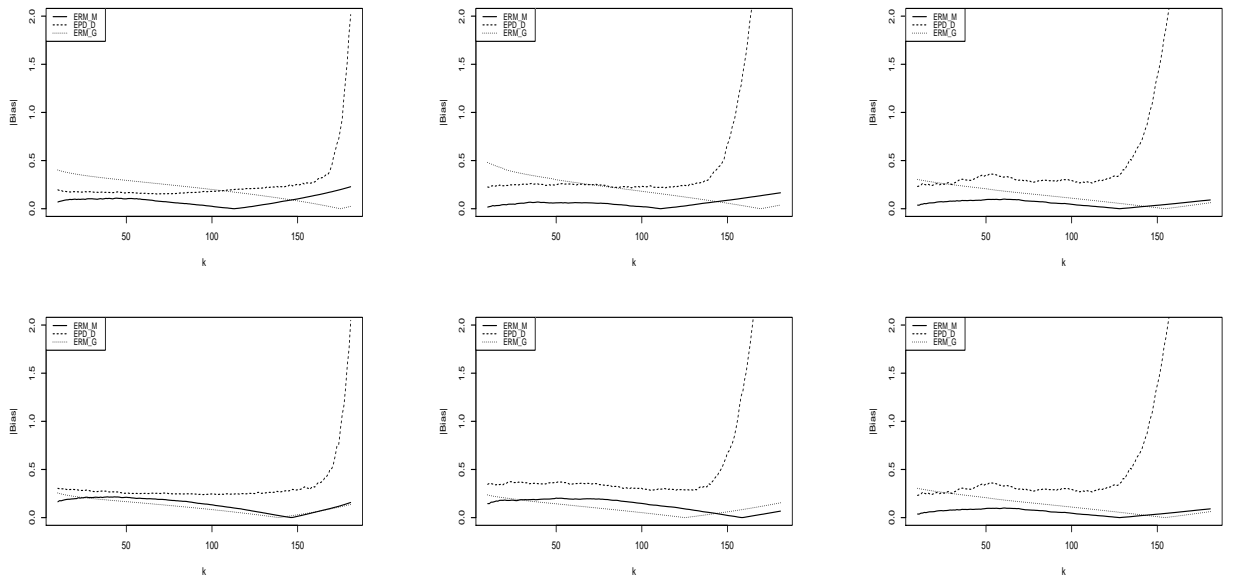


Figure 20: Pareto distribution with $n = 200$ and $\gamma = 0.5$.

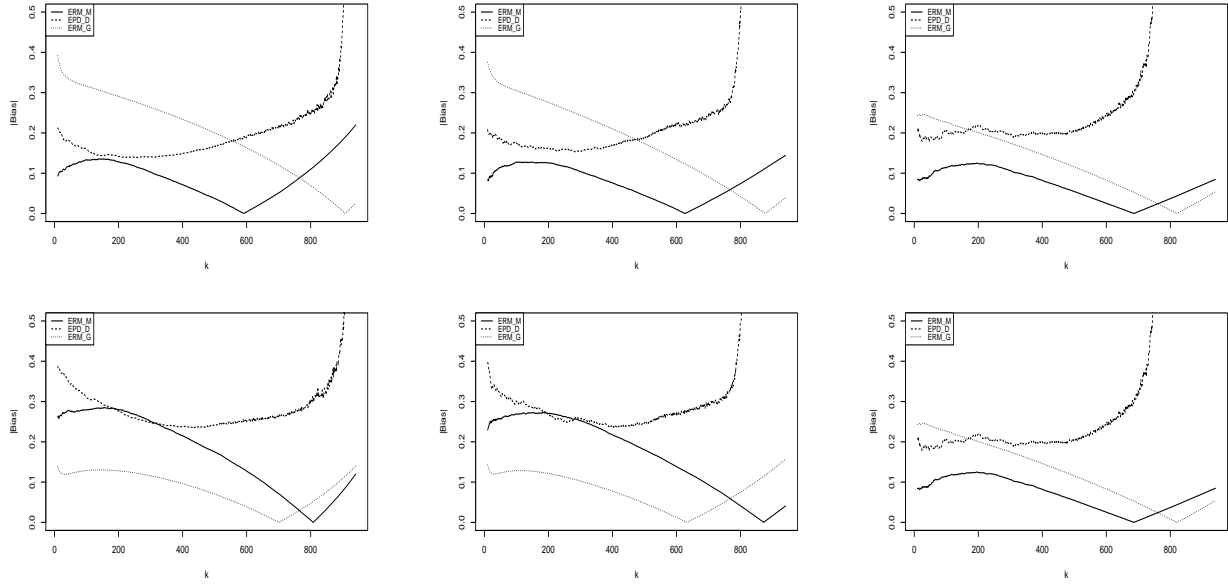


Figure 21: Pareto distribution with $n = 1000$ and $\gamma = 0.5$.