



**HAL**  
open science

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

Richard Minkah, Tertius de Wet

► **To cite this version:**

Richard Minkah, Tertius de Wet. Robust Estimation of Pareto-Type Tail Index through an Exponential Regression Model. 2019. hal-02116753v1

**HAL Id: hal-02116753**

**<https://hal.science/hal-02116753v1>**

Preprint submitted on 9 May 2019 (v1), last revised 14 Apr 2020 (v2)

**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  
and

Tertius de Wet

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

## 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 an existing estimator for Pareto-type tail index based on fitting an extended Pareto distribution with the minimum density power divergence. A simulation study is conducted to assess the performance of the estimators under different contaminated samples from different distributions. The results show that the proposed estimator has better mean square errors and less sensitivity 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 estimators are 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

## 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,

---

\*Correspondence: [rminkah@ug.edu.gh](mailto: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.

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 and affects 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 the context of extreme value analysis, its usage may appear to be contradictory. However, it has be 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) seems 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) employs 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 a introduction to extreme value theory. In section 3, the two estimators of the the Pareto-type tail index are compared via a simulation study. Section 4 presents an illustration of the two estimators on the estimation of the tail index of a practical data set from the insurance industry. We provide concluding remarks in Section 5.

## 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 beyond the range of the data, we need to study the behaviour of the sample maximum,  $X_{n,n}$ . The well-known Fisher and Tippett (1928) and Gnedenko (1943) theorem provides that for a suitable 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 extreme value type. Formally, if the sequence of constants  $a_n > 0$  and  $b_n \in \mathbb{R}$  exist, then we can find

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

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 a distribution function,  $F$ , satisfies (2), it is said to belong to the domain of attraction of  $G_\gamma$  and is denoted by  $F \in D(G_\gamma)$ . Here,  $\gamma$  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 right endpoint. If the condition (2) is satisfied, then  $F$  is said to belong to the domain of attraction of  $G_\gamma$ , written as  $F \in D(G_\gamma)$ . The goal of extreme value analysis is mainly to obtain estimates of high quantiles, exceedance probabilities and return periods. However, each of these parameters depend on the extreme value index,  $\gamma$ , which measures the tail heaviness of the underlying distribution. Therefore, the estimation of  $\gamma$  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 an extreme value distribution if and only if the distribution of excesses over high thresholds is asymptotically generalised Pareto (GP). 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 GP, 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 EVD is potentially flawed. Two options available are to base inferences on the part that are well fitted by the EVD 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

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

with  $Q$  the quantile function of  $F$ . Here,  $\ell_F$  and  $\ell_U$  are a 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  $\ell_U$ .

In this section, we present the two methods used in the estimation of tail index of a distribution function. After, this we discuss the robust method of estimation using the density power divergence method of Basu et al. (1998).

## 2.1 Extended Pareto Model

From (3), the conditional distribution of the relative excesses  $P\left(\frac{X}{u} > x | X > u\right)$  converges to  $x^{-1/\gamma}$  for  $x > 1$ . Using this Pareto-type behaviour, an estimate of  $\gamma$  is obtained as the slope of the Pareto quantile plot. Also, the maximum likelihood estimate of  $\gamma$  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 with 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  $RV_\beta$ , a regularly varying function at infinity with index  $\beta$  satisfying

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

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  $\delta$  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  $\delta$  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  $\ell_U$  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) obtains 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 exist 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 an approximate representation of the log-spacings of the order statistics in (6),

$$Z_j = j \cdot (\log X_{n-j+1,n} - \log X_{n-j,n}), \quad 1 \leq j \leq k \leq n, \quad (14)$$

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

$$Z_j \sim \left( \gamma + b_{n,k} \left( \frac{j}{k+1} \right)^{-\rho} \right) E_j \quad (15)$$

where each  $E_j$  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 of (15) were obtained by the 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  $\gamma$  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  $g$  and  $h$ , 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 density power 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 robust parameter,  $\alpha$ , 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 distribution function  $F$  of which  $\theta$  is an unknown parameter of interest. The minimum density power divergence (MDPD) estimator of  $\theta$  is obtained by minimising the divergence between the data and the model density

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

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) depends on the unknown parameter  $\tau$ , 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 log-spacings of order statistics,  $Z_j$ ,  $j = 1, \dots, k-1$ , in (14) each has distribution function  $F_{\theta_j}$  and corresponding density functions  $f_{\theta_j}$ . Although the  $Y_j$ s are independent and follows  $f_{\theta_j}$  which are exponentially distributed, they are not identical. Note that  $\theta_j = \gamma + b_{n,k}(j/k + 1)^{-\rho}$ , and hence, it is linear function of  $\gamma$  and non-linear functions of the other parameters,  $b_{n,k}$  and  $\rho$ .

The minimum density power estimator for the parameters  $\gamma$ ,  $b_{n,k}$  and  $\rho$ , can be obtained by following Ghosh and Basu (2013) and Ghosh (2017) where a minimisation of the function

$$G_k(\theta) = \frac{1}{k-1} \sum_{i=1}^k \left[ \int \left[ f_{\theta_j}^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) f_{\theta_j}^{\alpha} \hat{g}_j \right] \right] \quad (20)$$

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

$$H_k(\gamma, b_{n,k}, \rho) = \frac{1}{k-1} \sum_{i=1}^{k-1} \left[ \frac{1}{(1+\alpha)\theta_j^{\alpha}} - \frac{1+\alpha}{\alpha\theta_j^{\alpha}} \exp\left(-\frac{\alpha Z_j}{\theta_j}\right) \right] \quad (21)$$

where  $\theta_j = \gamma + b_{n,k} \left(\frac{j}{k+1}\right)^{-\rho}$ . The parameters of the (15) can be obtained by minimising the distance of the objective function (21). Alternatively, the estimators of (15) can be obtained by solving the estimating equation  $\frac{\delta G_k}{\delta \theta_j} = 0$ . For example, in the case of  $\gamma$ , differentiating (21) with respect to  $\gamma$ , yields a simplified form,

$$\frac{1}{k-1} \sum_{i=1}^{k-1} \frac{1+\alpha}{\theta_j^{\alpha+2}} \left[ \frac{\alpha\theta_j}{(1+\alpha)^2} + (Z_j - \theta_j) \exp\left(-\frac{\alpha Z_j}{\theta_j}\right) \right] = 0 \quad (22)$$

### 3 Simulation Study

In this section, we compare the performance of the two robust methods of estimating Pareto-type tail index through a simulation study. We consider four distributions in the Fréchet-Pareto domain of attraction namely the Fréchet, Pareto, Burr and the log gamma. The distribution function,  $1 - F$ , tail index,  $\gamma$ , and the second-order parameter,  $\rho$  of the distributions used in the simulation study are presented in Table 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  $\alpha$ , at 0.1, 0.5 and 1 representing increasing levels of robustness.



Table 1: Distributions and their tail indexes

Distribution	$1 - F(x)$	$\gamma$	$\rho$
Burr	$(\eta / (\eta + x^\tau))^\lambda, x > 0, \eta, \tau, \lambda > 0$	$\frac{1}{\tau\lambda}$	$-\frac{1}{\lambda}$
Fréchet	$1 - \exp(-x^{-\beta}), x > 0, \beta > 0$	$1/\beta$	-1
Log gamma	$\int_x^\infty \frac{\lambda^\beta}{\Gamma(\beta)} u^{-(\lambda+1)} (\log u)^{\beta-1} du, x > 1, \beta, \lambda > 0.$	$\frac{1}{\lambda}$	
Pareto	$x^{-\beta}, x > 0, \beta > 0$	$\frac{1}{\beta}$	-1
Uniform	$\frac{x-a}{b-a}, a \leq x \leq b$		$-b$

### 3.1 Performance Measures under Contamination from the same Distribution

We study the behaviour of the proposed estimator and the MDPD estimator of the tail index in the case where contamination of the base distribution comes from the same distribution but with different parameters. The results of the simulation studies for the Burr distribution contaminated by another Burr but with different tail index are presented in Figures 1-3.

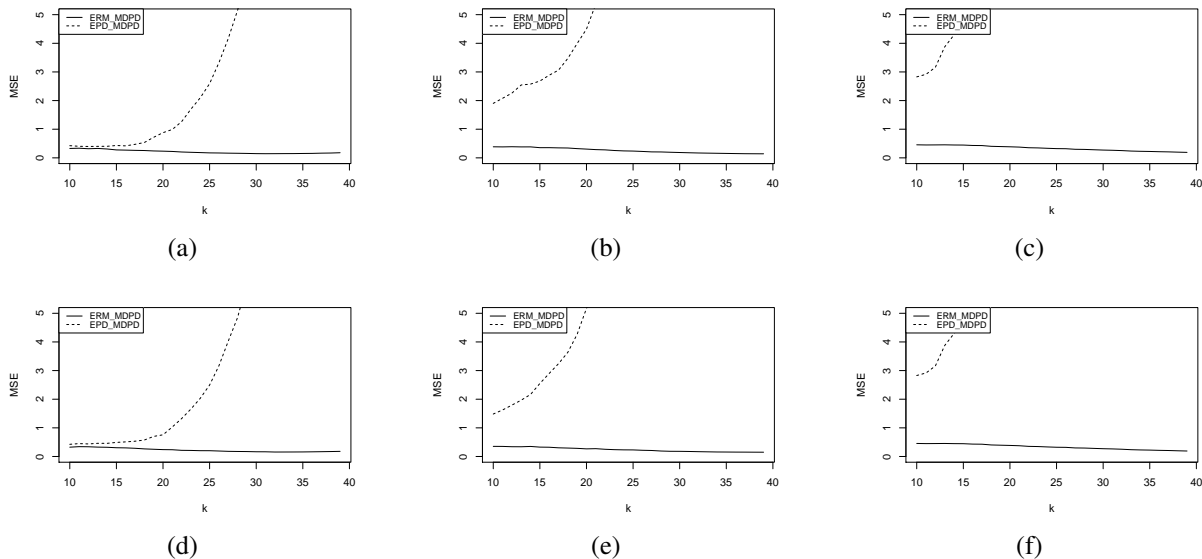


Figure 1: Burr distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

In the case of smaller sample size,  $n = 50$ , the proposed ERM\_MDPD estimator shows clear improvement on MSE than the EPD\_MDPD across the three robust tuning parameters as well as the percentage of contamination. Similar performance can be seen for the other sample sizes, especially as the number of top order statistics increases.

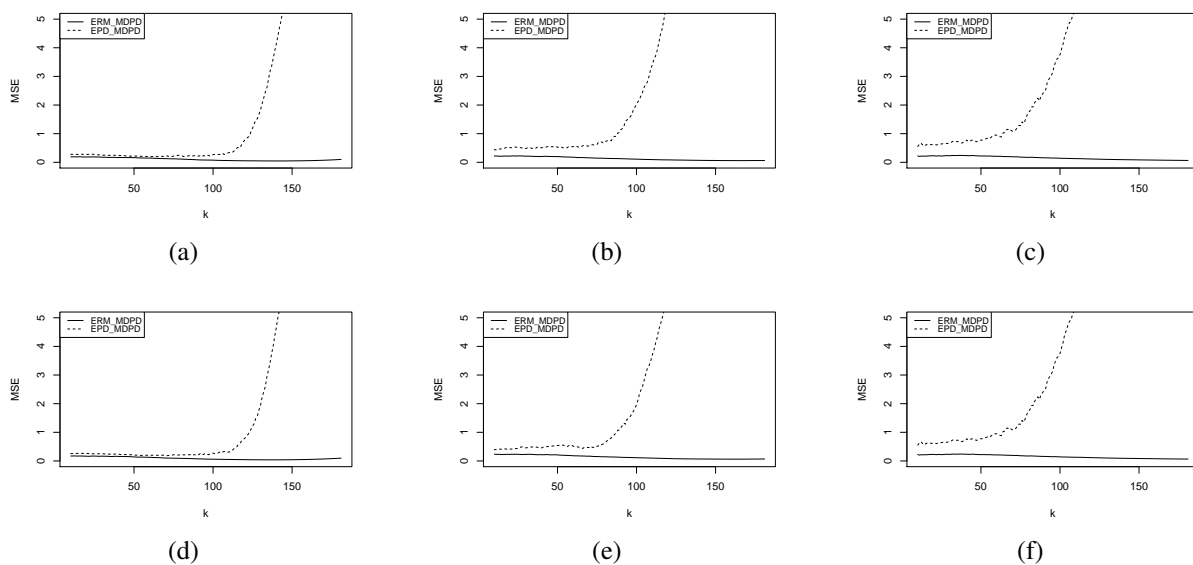


Figure 2: Burr distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

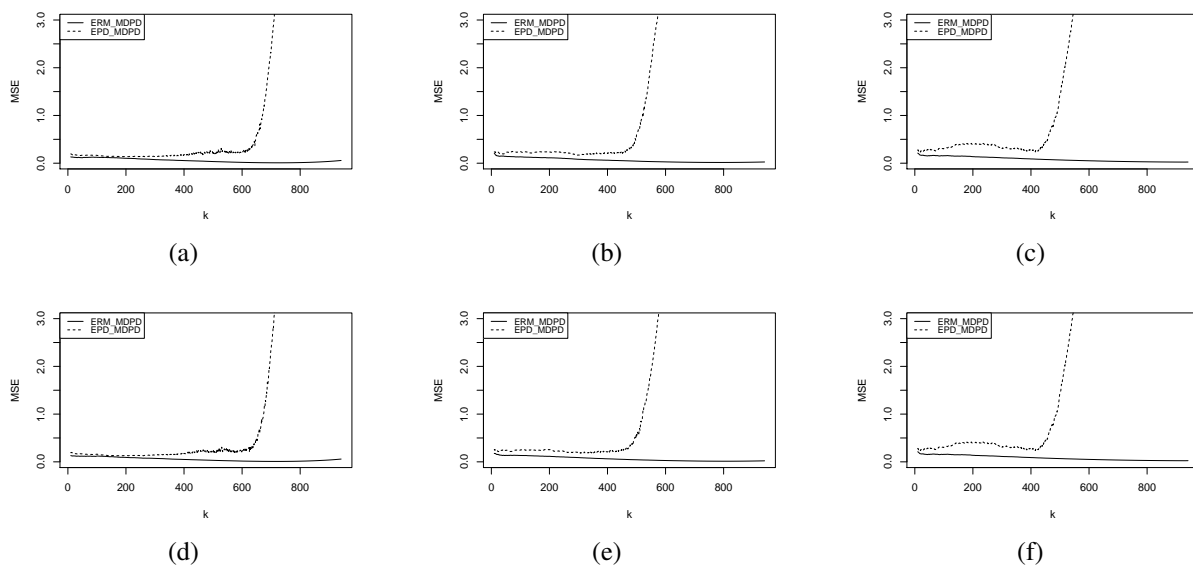


Figure 3: Burr distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

The results for the same scenario above but with Fréchet, Log gamma and Pareto distributions are deferred to appendices A.1, A.2 and A.3 respectively.

Firstly, for the Fréchet distribution, ERM\_MDPD generally outperforms EPD\_MDPD in terms of MSE. However, for smaller sample size, and smaller robust tuning parameter  $\alpha \leq 0.5$ , EPD\_MDPD is preferred as it records lower MSE values. Secondly, in the case of the log

gamma distribution, ERM\_MDPD estimator outperforms the EPD\_MDPD except for few cases of the robust parameter,  $\alpha = 0.1$ . Lastly, ERM\_MDPD is preferred in the estimation of the tail index from the Pareto distribution as it has smaller MSE values for all cases considered especially when  $k \rightarrow n$ .

Therefore, the simulation results show that across the different distributions and the factors such as percentage of contamination and robustness levels, the ERM\_MDPD is found to be generally a better alternative to the EPD\_MDPD estimator.

### 3.2 Performance Measures under Contamination from different Distribution

In this section, we present the results of the simulation study for  $F$  and  $G$  chosen from a different domains of attraction. Here, we present the results for the estimation of the tail index of a Burr distribution with  $\eta = 1$ ,  $\tau = 0.5$  and  $\lambda = 4$ . Therefore, the tail index is 0.5.

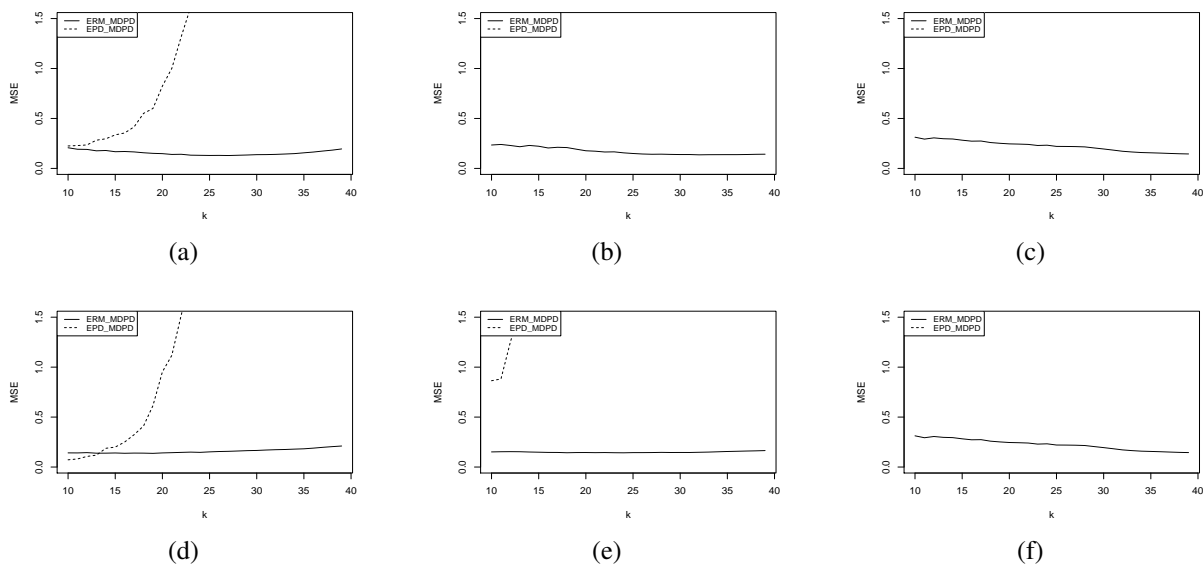


Figure 4: Burr distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

Considering the results from Figures 4-6, we observe that the proposed MDPD estimator based on the exponential Regression Model (EPD\_MDPD), has generally smaller MSE and stable across the values for  $k$ , compared to the existing MDPD estimator from extended Pareto distribution of Dierckx et al. (2013). The other distributions Fréchet, Log gamma, and Pareto have been presented in Appendices B.1, B.2 and B.3 respectively. In each case, the parameters have been chosen such that the tail index is equal to  $\gamma = 0.5$ .

Firstly, for Fréchet distribution, the ERM\_MDPD estimator is generally the best as it has smaller MSEs except for the case of smaller sample size,  $n = 50$ , and  $\alpha \leq 0.5$ . Secondly, the case of the log gamma distribution, the performance of the ERM\_MDPD is better in terms of its stability with increasing  $k$  and level of contamination. Similar conclusions can be reached for the performance of the estimators in the case of the Pareto distribution.

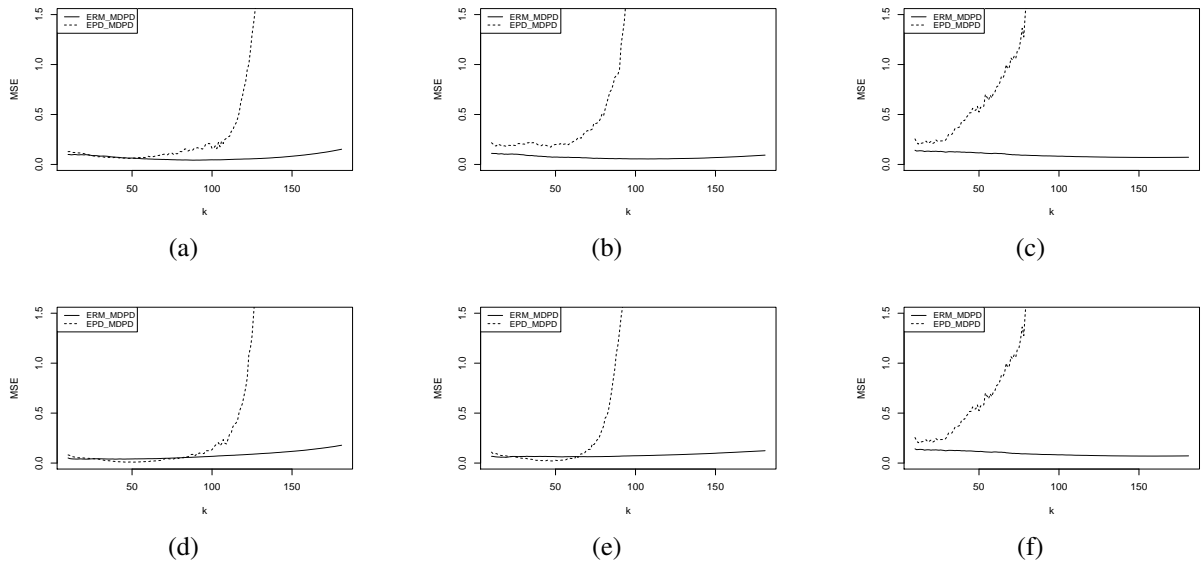


Figure 5: Burr distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

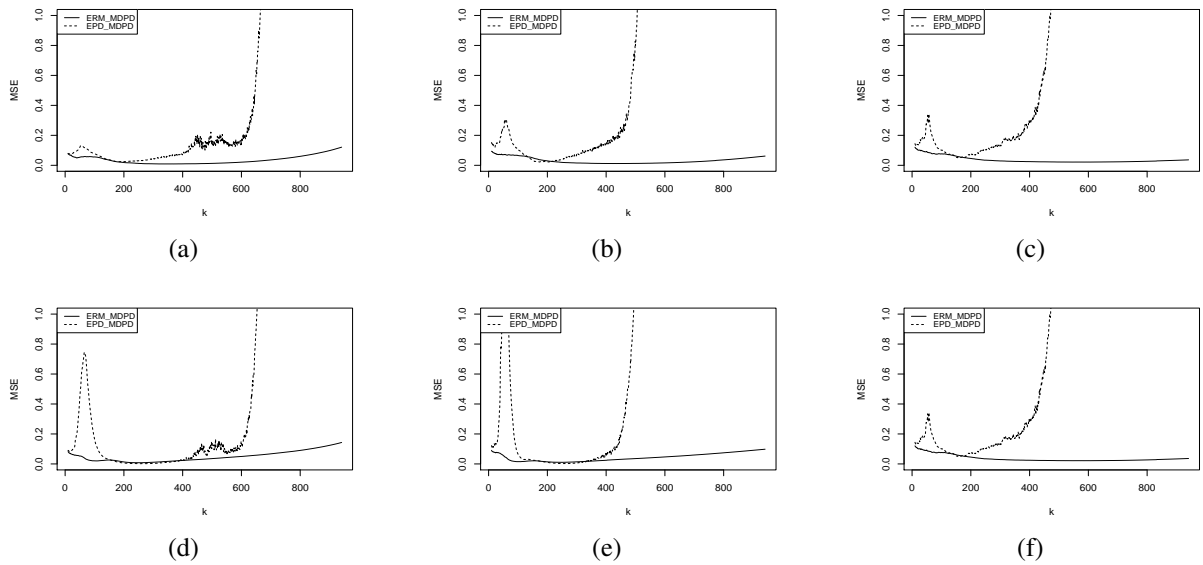


Figure 6: Burr distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

## 4 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 an perturbed Pareto distribution and exponential regression model. In this section, we illustrate the application of the robust estimators discussed in the previous section in estimating the tail index of the SOA data.

The plot of the data in Figure 7 shows that two particular large claims seems to be detached from the bulk of the data. Such outliers have implications on traditional method of estimation of the parameters of the GP distribution such as maximum likelihood. Using different robust tuning parameters, we compute the tail index as a function of the number of top order statistics,  $k$ . The result shows that the our proposed estimator, ERM\_MDPD, is mostly stable along the path of  $k$  compared to the robust estimator of based on the extended Pareto distribution, EPD\_MDPD. Also, in conformity with the behaviour of robust estimators, the variation in the estimates increases with increasing  $\alpha$ . Therefore, ERM\_MDPD provides a better alternative robust estimator for the tail index in the Pareto domain as illustrated with the SOA data.

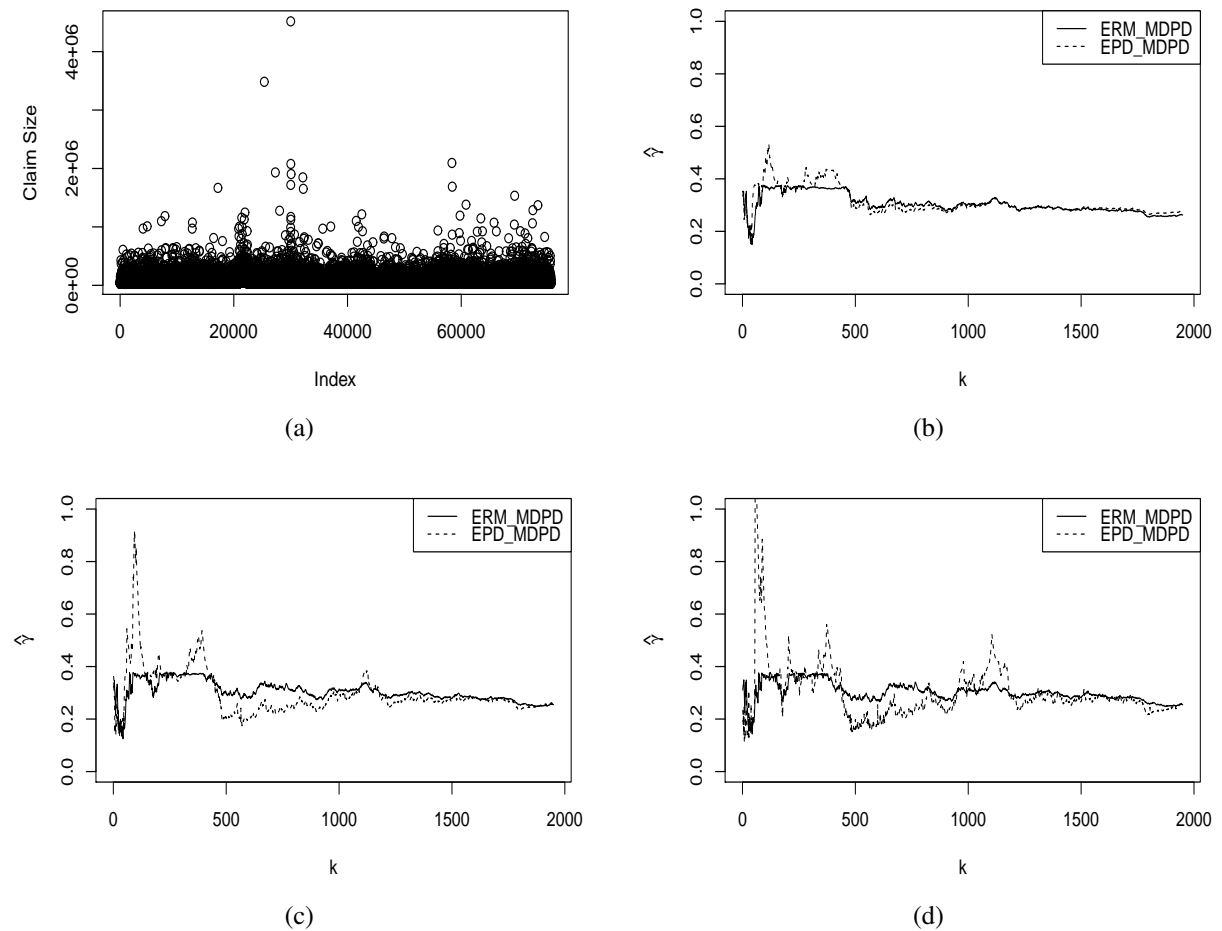


Figure 7: SOA data: (a) scatter plot (b)  $\gamma$  estimates with  $\alpha = 0.1$  (c)  $\gamma$  estimates with  $\alpha = 0.5$  and (d)  $\gamma$  estimates with  $\alpha = 1$

## 5 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 through a simulation study together with a similar estimator using minimum density power divergence but on an extended Pareto distribution fitted to relative excesses. The results of the simulation study shows that the proposed minimum density power estimator based on an exponential regression model generally has a better performance than that of the extended Pareto distribution. In addition, the proposed robust estimator of the tail index is less sensitive to the number of top order statistics. The theoretical and the robustness properties of the proposed estimator is a subject for future research.

## 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 Contamination with same distribution

## A.1 Fréchet Distribution

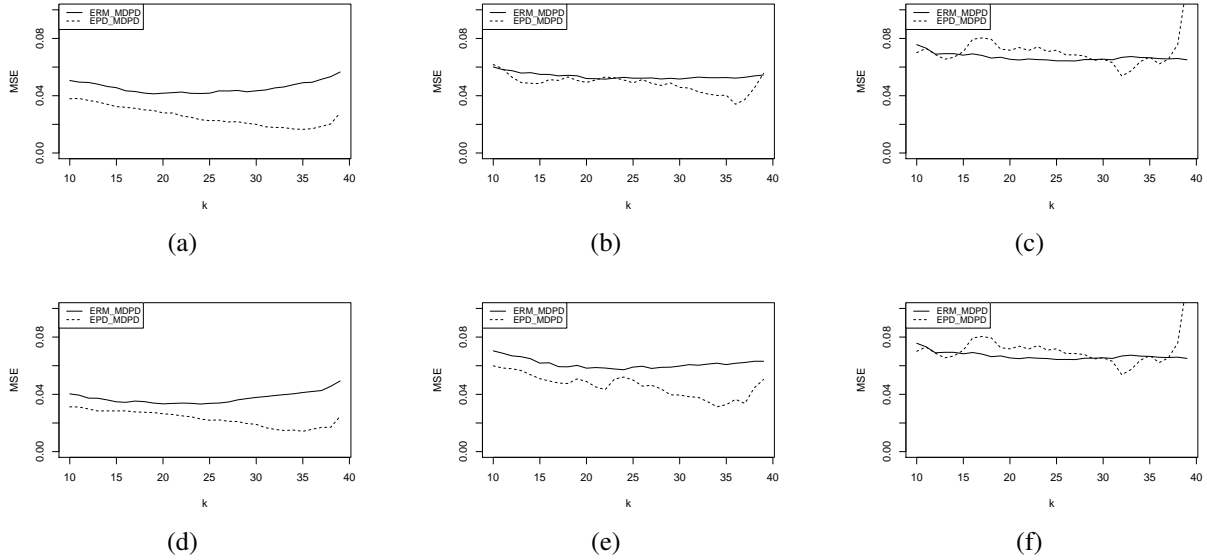


Figure 8: Fréchet distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

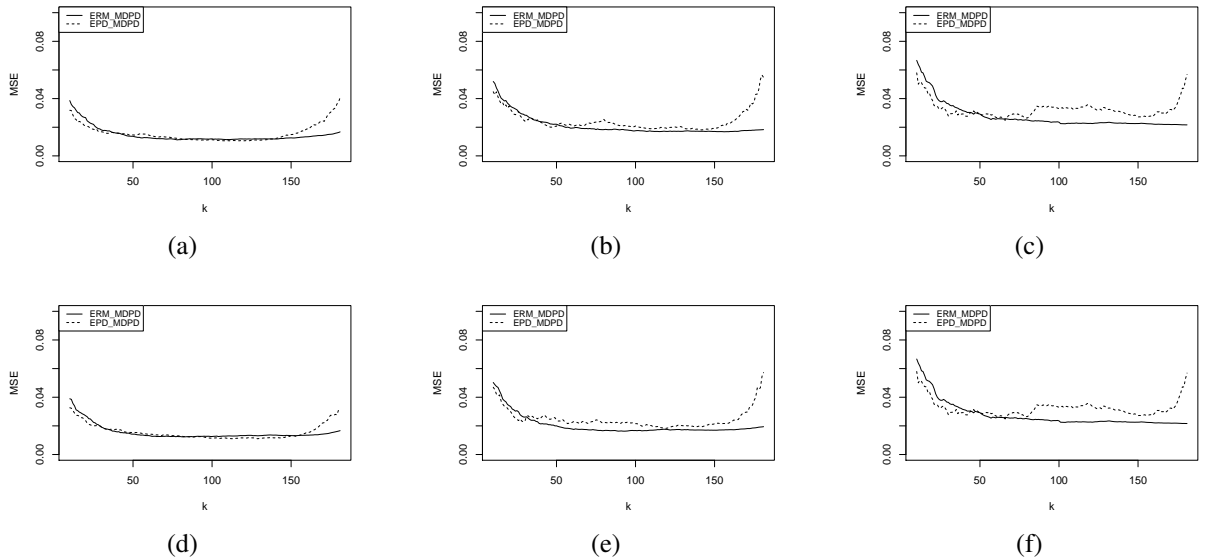


Figure 9: Fréchet distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .



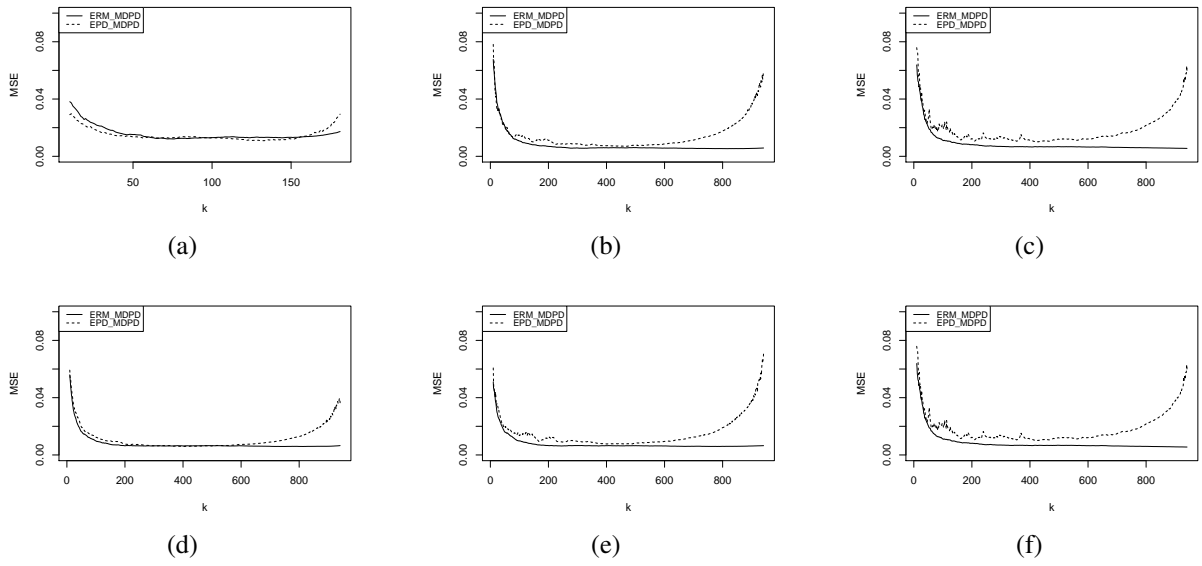


Figure 10: Fréchet distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

## A.2 Log gamma Distribution

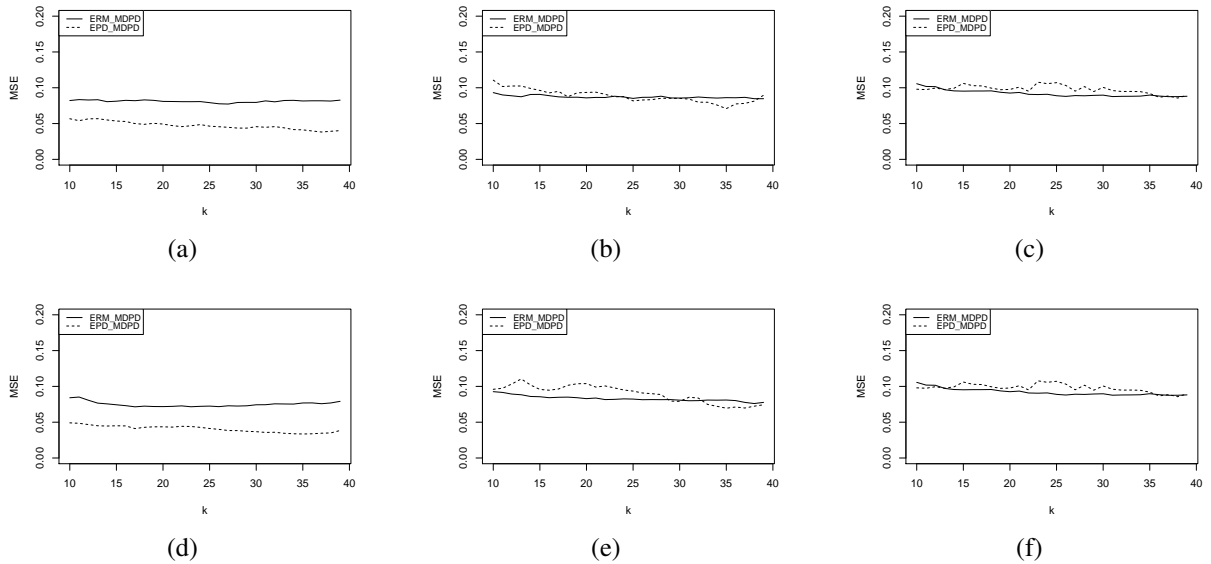


Figure 11: Log gamma distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

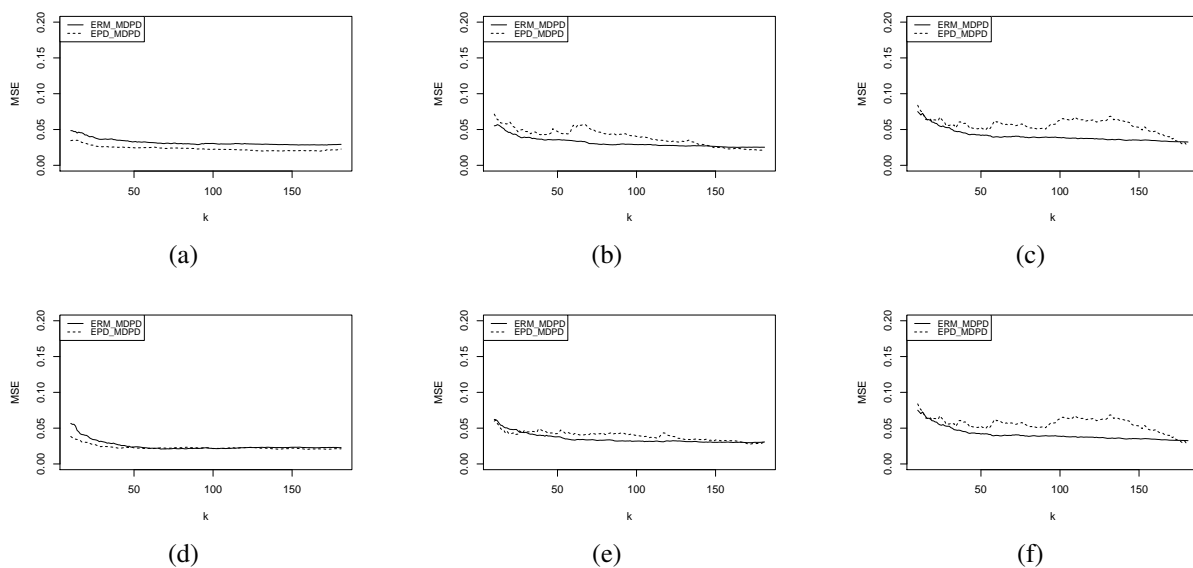


Figure 12: Log gamma distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

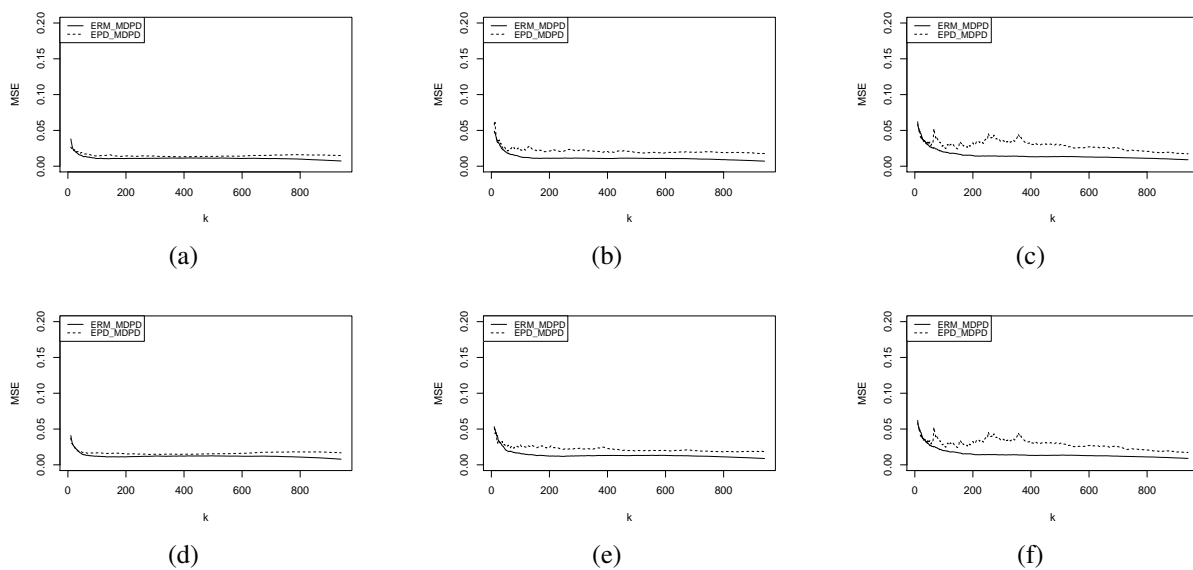


Figure 13: Log gamma distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

### A.3 Pareto Distribution

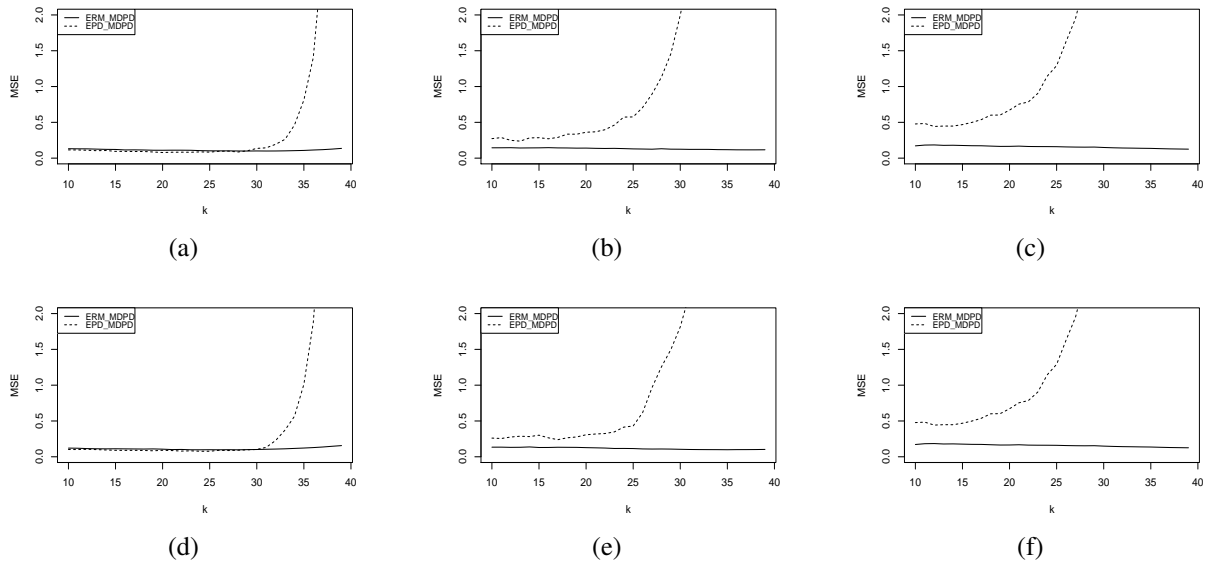


Figure 14: Log gamma distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

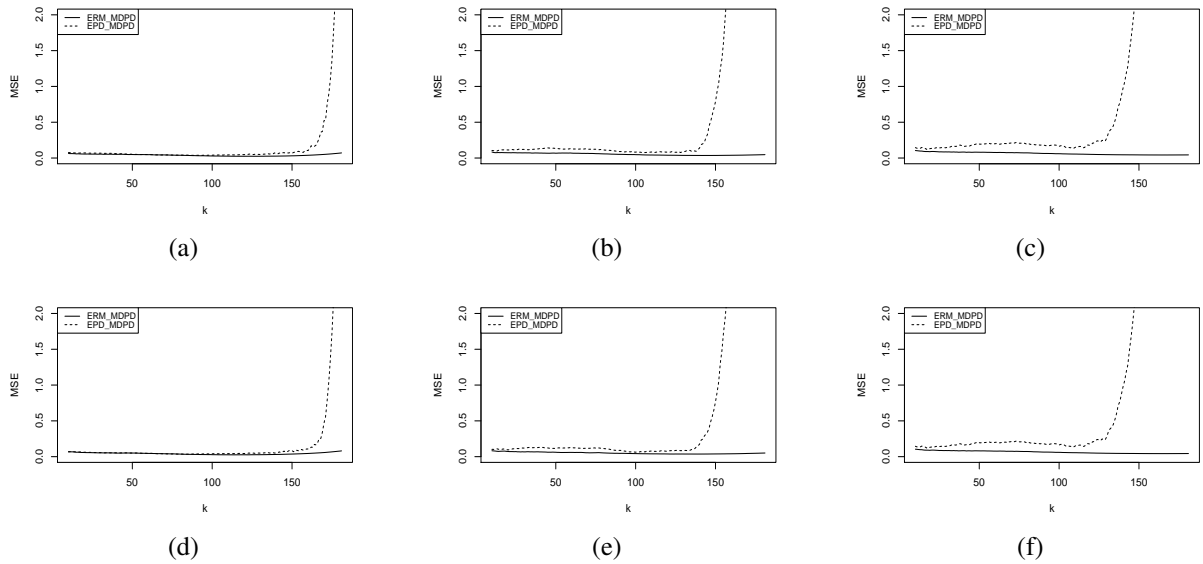


Figure 15: Log gamma distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

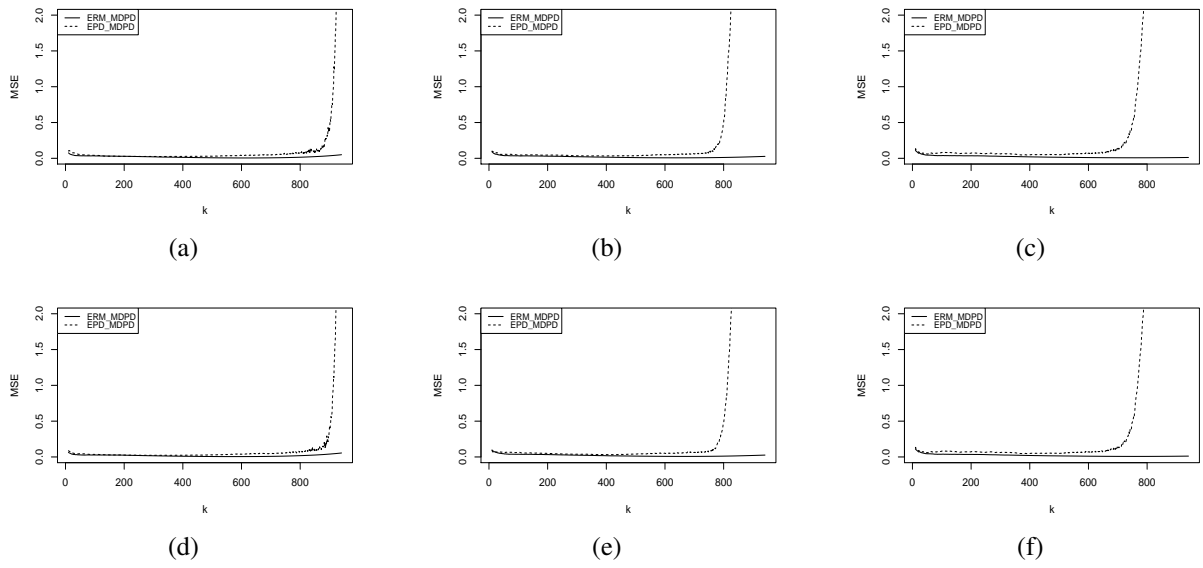


Figure 16: Log gamma distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

## B Contamination with different distribution

### B.1 Fréchet Distribution

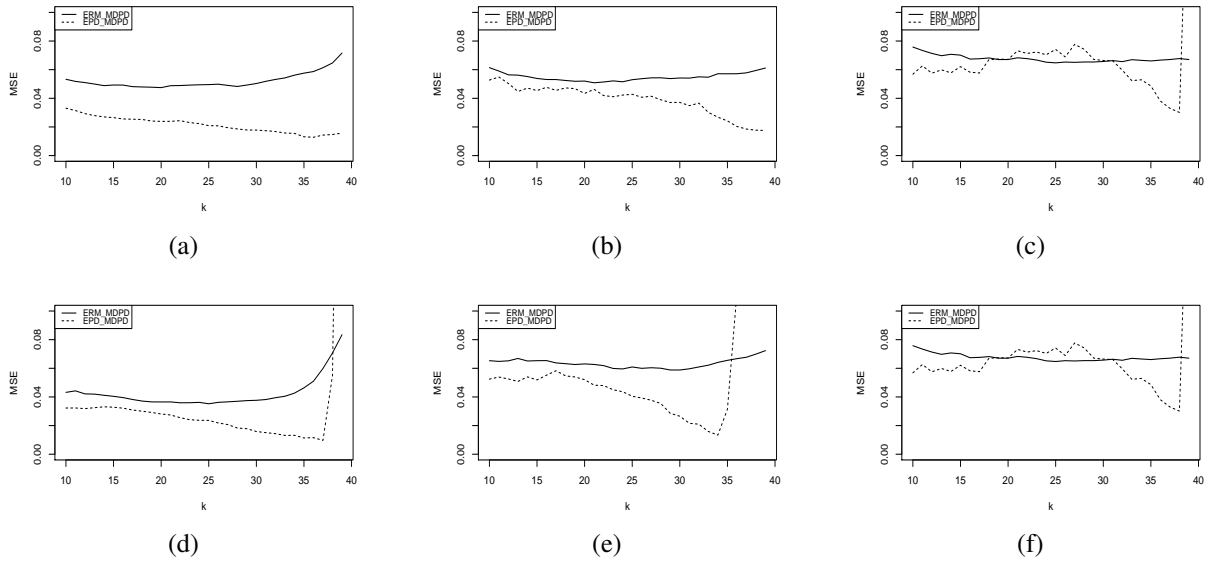


Figure 17: Fréchet distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

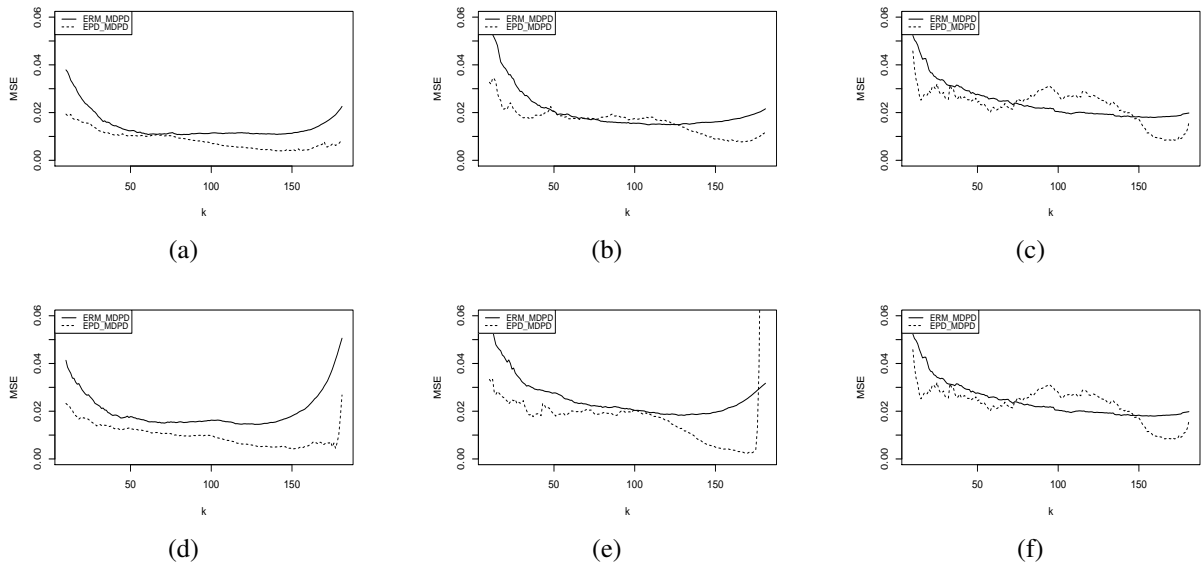


Figure 18: Fréchet distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\epsilon = 0.05$ ; and bottom-most row:  $\epsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

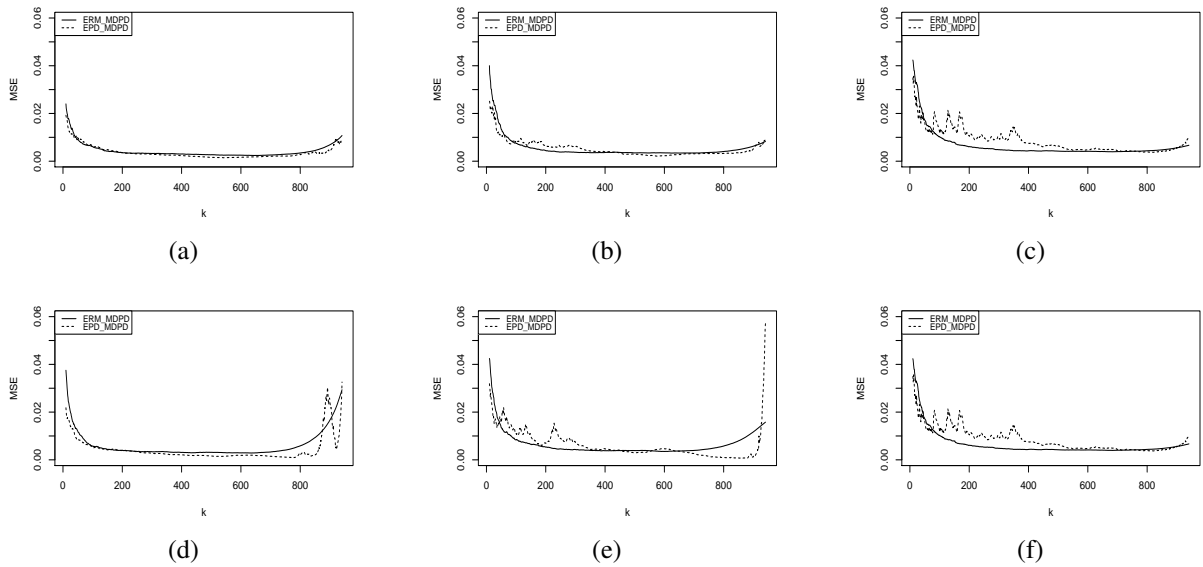


Figure 19: Fréchet distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\epsilon = 0.05$ ; and bottom-most row:  $\epsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

## B.2 Log gamma Distribution

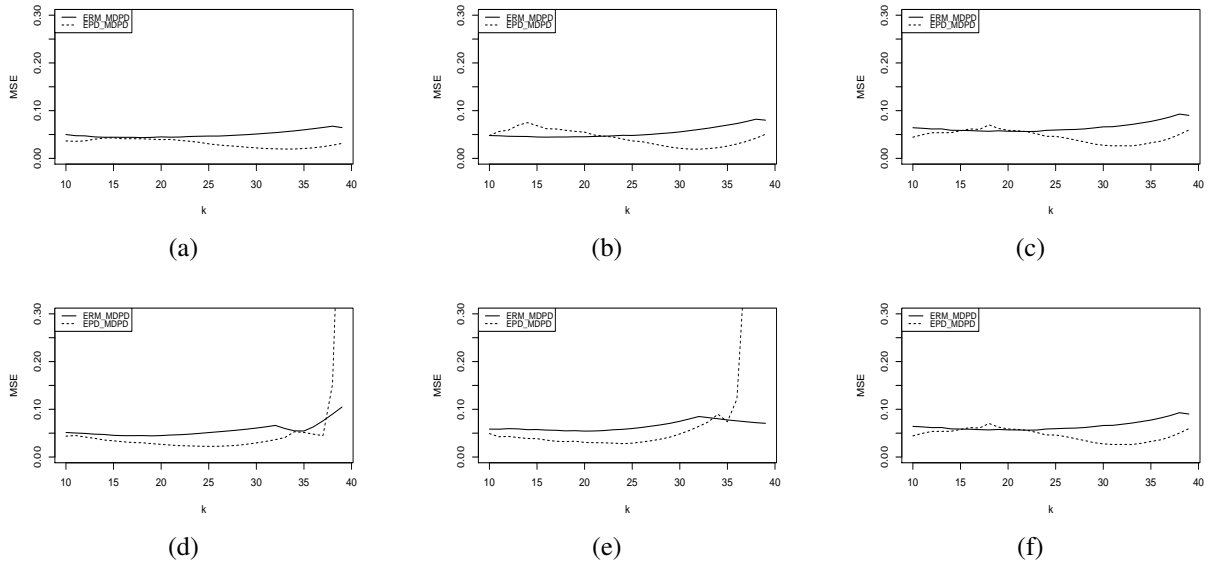


Figure 20: Log gamma distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

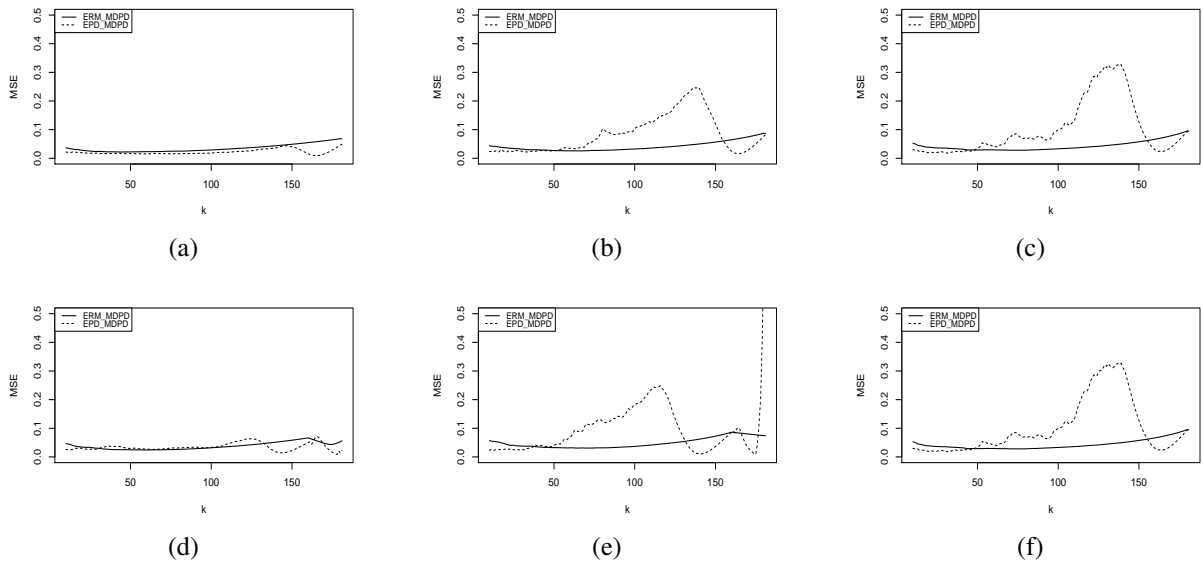


Figure 21: Log gamma distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

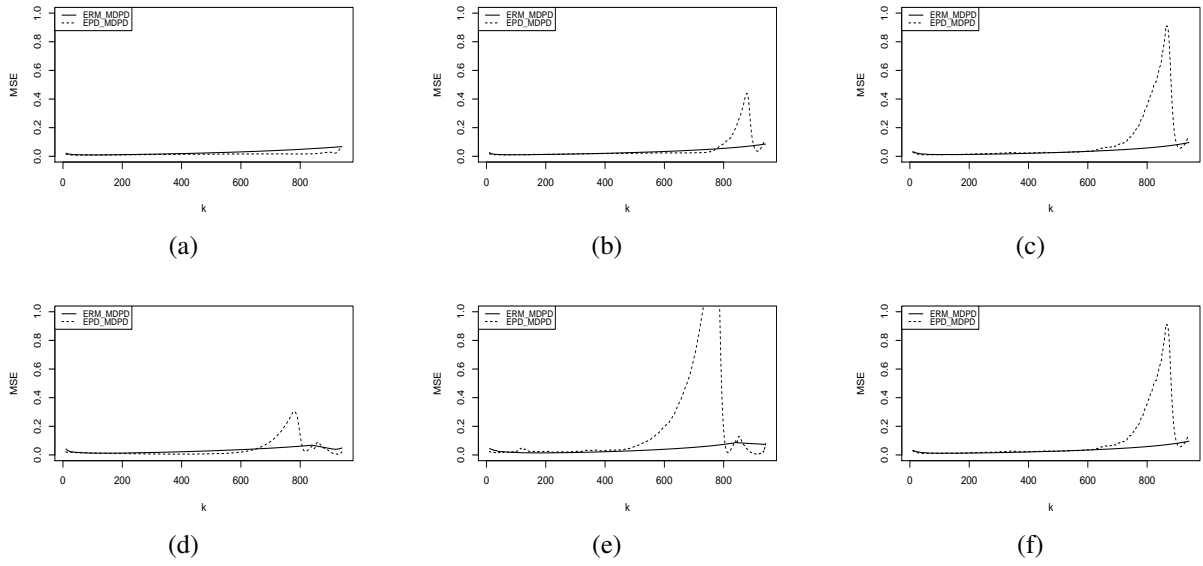


Figure 22: Log gamma distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

### B.3 Pareto Distribution

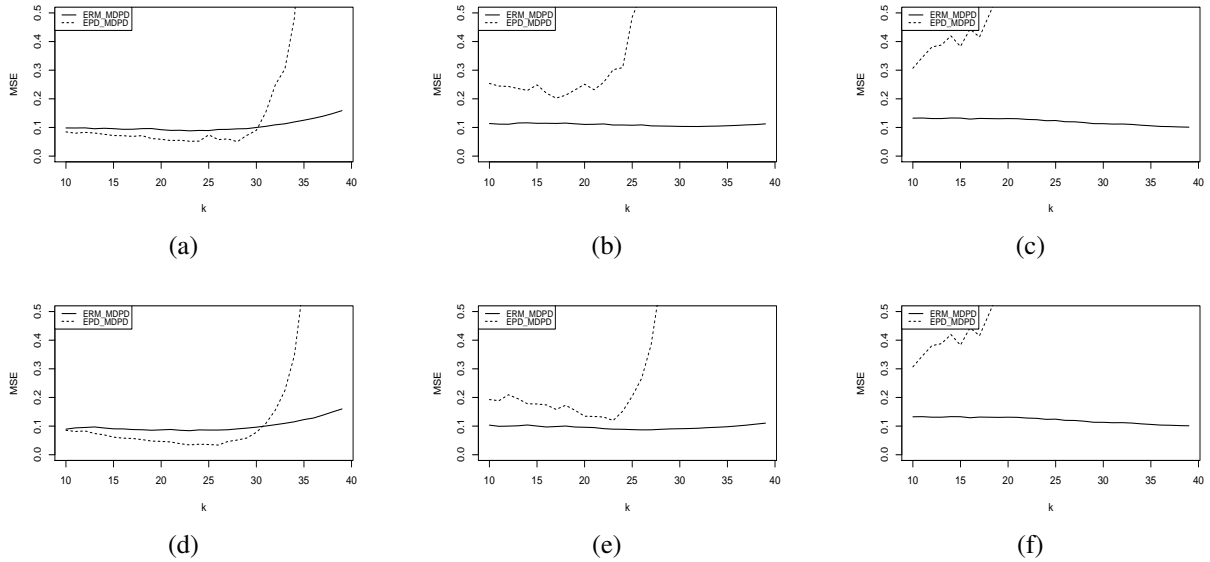


Figure 23: Pareto distribution with  $n = 50$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottommost row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

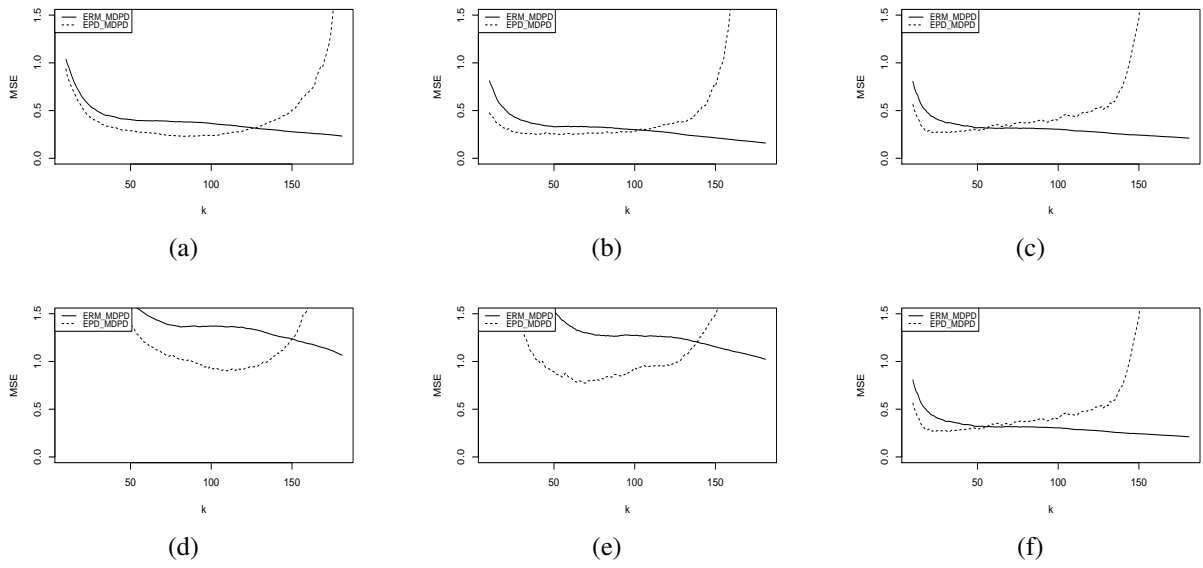


Figure 24: Pareto distribution with  $n = 200$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottom-most row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .

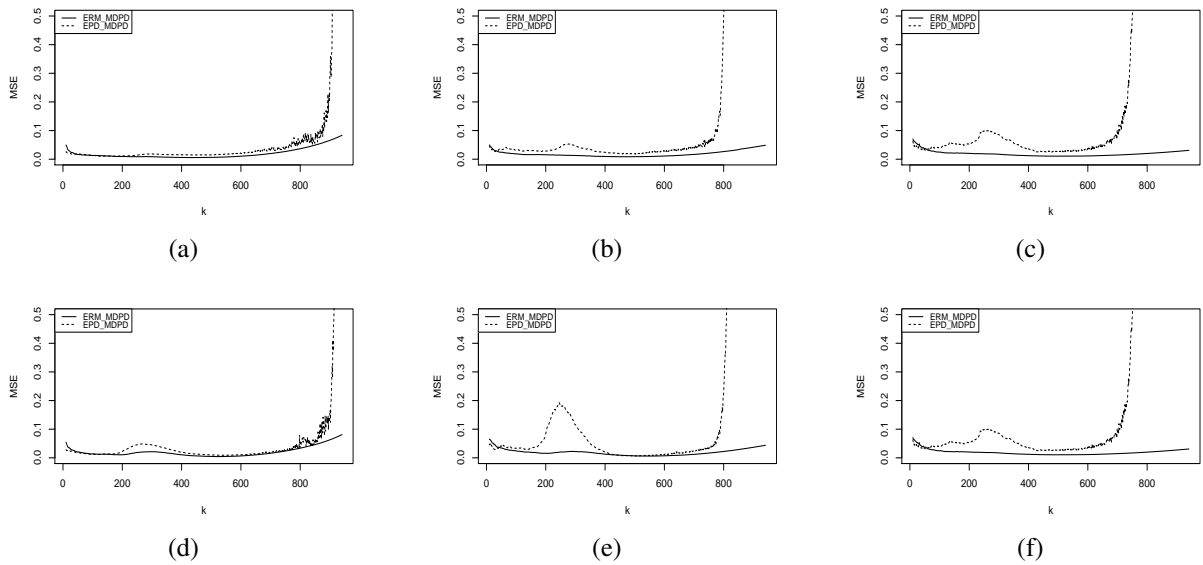


Figure 25: Pareto distribution with  $n = 1000$  and  $\gamma = 0.5$ . Topmost row:  $\varepsilon = 0.05$ ; and bottom-most row:  $\varepsilon = 0.15$ . Leftmost column:  $\alpha = 0.1$ ; middlemost column:  $\alpha = 0.5$ ; and rightmost column:  $\alpha = 1$ .