Statistical Methodology 8 (2011) 340–355

Contents lists available at ScienceDirect

Statistical Methodology

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

Efficient estimation in the Pareto distribution with the presence of outliers

U.J. Dixit, M. Jabbari Nooghabi *

Department of Statistics, University of Mumbai, Mumbai, India

ARTICLE INFO

Article history:
Received 16 June 2010
Received in revised form
31 January 2011
Accepted 31 January 2011

Keywords:
Pareto distribution
Maximum likelihood estimator
Uniformly minimum variance unbiased estimator
Probability density function
Cumulative distribution function
rth moment
Outliers
Insurance

ABSTRACT

The maximum likelihood (ML) and uniformly minimum variance unbiased estimators (UMVUE) of the probability density function (pdf), cumulative distribution function (cdf) and rth moment are derived for the Pareto distribution in the presence of outliers. It has been shown that MLE of pdf and cdf are better than their UMVUEs. At the end, these methods are illustrated with the help of real data from an insurance company.

1. Introduction

The Pareto distribution was originally used to describe the allocation of wealth among individuals since a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. It can be shown that using the graph of $f(x)$ (probability density function), the probability that owns a small amount of wealth per person is high. The probability then decreases steadily as wealth increases.

Another application of this distribution is for On-Line Analytical Processing (OLAP) view size estimation. Nadeau and Teorey [9] used the Pareto distribution for OLAP aims at gaining useful information quickly from large amounts of data residing in a data warehouse. To improve the quickness of response to queries, pre-aggregation is a useful strategy. However, it is usually impossible

* Corresponding author.
E-mail addresses: ulhasdixit@yahoo.co.in (U.J. Dixit), jabbarinm@yahoo.com (M. Jabbari Nooghabi).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2011.01.011

to pre-aggregate along all combinations of the dimensions. The multi-dimensional aspects of the data lead to a combinatorial explosion in the number and potential storage size of the aggregates.

Nadeau and Teore [9] have suggested to selectively pre-aggregate. Cost/benefit analysis involves estimating the storage requirements of the aggregates in question. They [9] presented an original algorithm for estimating the number of rows in an aggregate based on the Pareto distribution model. They also tested the Pareto model algorithm empirically against four published algorithms, and concluded that the Pareto model algorithm is consistently the best of these algorithms for estimating view size. Pareto distribution is also useful for finding the average of annuity.

In economics, where this distribution is used as an income distribution, the threshold parameter is some minimum income with a known value.

Asrabadi [1] derived the uniformly minimum variance unbiased estimator (UMVUE) of the probability density function (pdf), the cumulative distribution function (cdf) and the rth moment of Pareto distribution. Dixit and Jabbari Nooghabi [6] had obtained a maximum likelihood (ML) estimator of pdf and cdf and had shown that the ML estimators are more efficient than their UMVUEs. Further, it is shown that the MLE of the rth moment does not exist.

In insurance for modelling the claims where the minimum claim is the modal value, we can use the Pareto distribution (see http://www.brighton-webs.co.uk/distributions/pareto.asp). Also, according to Benktander [2], the Pareto distribution is useful for automobile insurance problems. For example, in a motor insurance, a claim of at least θ as a compensation can be made and claims below θ are not entertained. Here the parameter θ is known and we can fit the Pareto distribution with parameters α and θ to the data of claims, where θ is known and α is unknown. For more details about applications of the homogeneous case of the Pareto distribution see [9,1,2]. In the above example, we know that the vehicles involved are of different costs, of which some of them may have a very high cost and claim amounts vary according to the damage to the vehicles. So if a company assumes that claims of these vehicles (expensive/severe damaged vehicles) are β times higher than normal vehicles, the data of claims follow a Pareto distribution in the presence of outliers with parameters α, β and θ, where α is unknown, β, θ and the number of outliers are Dixit and Nasiri known. For the model of outliers refer to [3,5,7,8]. We do claim that this work is the first in estimation in the Pareto distribution with outliers.

Let a set of random variables $(X_1, X_2, \ldots, X_n)$ represent the claim amounts of a motor insurance company. It is assumed that claims of some of vehicles (expensive/severely damaged vehicle) are β times higher than normal vehicles.

Hence, we assume that the random variables $(X_1, X_2, \ldots, X_n)$ are such that any k (number of outliers) of them are distributed with pdf

$$f_2(x; \alpha, \beta, \theta) = \frac{\alpha(\beta\theta)^\alpha}{x^{\alpha+1}}, \quad 0 < \beta\theta \le x,\ \alpha > 0,\ \beta > 1,\ \theta > 0, \tag{1}$$

and the remaining $(n-k)$ random variables are distributed as

$$f_1(x; \alpha, \theta) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \quad 0 < \theta \le x,\ \alpha > 0. \tag{2}$$

In this paper, we have derived the ML and UMVU estimators of pdf and cdf of the above Pareto distribution in the presence of outliers. We assume that β, θ and k are known and α is unknown. At the end, we have given an example of claims in a motor insurance company.

2. Joint distribution of $(X_1, X_2, \ldots, X_n)$ with k outliers

The joint distribution of $(X_1, X_2, \ldots, X_n)$ in the presence of k outliers is given by

$$f(x_1, x_2, \ldots, x_n; \alpha, \beta, \theta) = \frac{\alpha^n\theta^{n\alpha}\beta^{k\alpha}}{C(n,k)}\left(\prod_{i=1}^{n} x_i\right)^{-(\alpha+1)} \sum_{A_1=1}^{n-k+1}\sum_{A_2=A_1+1}^{n-k+2}\cdots\sum_{A_k=A_{k-1}+1}^{n}\prod_{j=1}^{k}\mathbf{I}(x_{A_j}-\beta\theta), \tag{3}$$

where $C(n,k)=\frac{n!}{k!(n-k)!}$ and $\mathbf{I}$ is the indicator function defined as

$$\mathbf{I}(y)=\begin{cases}1 & y>0,\\ 0 & \text{otherwise.}\end{cases}$$

The marginal distribution of X_i $(i=1,2,\ldots,n)$ can be written as:

$$f(x_i;\alpha,\beta,\theta)=b\frac{\alpha(\beta\theta)^{\alpha}}{x_i^{\alpha+1}}\mathbf{I}(x_i-\beta\theta)+\bar{b}\frac{\alpha\theta^{\alpha}}{x_i^{\alpha+1}}\mathbf{I}(x_i-\theta),\quad \alpha>0,\ \beta>1,\ \theta>0, \tag{4}$$

where $b=\frac{k}{n}$, $\bar{b}=1-b$ and $(X_1,X_2,\ldots,X_n)$ are not independent (for more details see [3,5,7,8]).

3. UMVU estimator

Let $X_1,X_2,\ldots,X_n$ be a random sample of size n from the distribution (1) and (2), then $T=\prod_{i=1}^{n}X_i$ is a complete sufficient statistic for the unknown parameter α. According to the Lehmann Scheffe theorem if $h(x_1|t)=f^*(t)$ be the conditional pdf of X_1 given T, we have

$$\mathrm{E}[f^*(T)]=\int f^*(t)h^*(t)\mathrm{d}t=\int h(x_1|t)h^*(t)\mathrm{d}t=\int h(x_1,t)\mathrm{d}t=f(x_1),$$

where $h^*(t)$ is the pdf of T and $h(x_1,t)$ is the joint pdf of X_1 and T. Therefore $f^*(t)$ is the UMVUE of $f(x_1)$.

Lemma. *The distribution of T is $h^*(t)$ as*

$$h^*(t)=\frac{\alpha^n\theta^{n\alpha}\beta^{k\alpha}}{\Gamma(n)}t^{-(\alpha+1)}[\ln(t)-k\ln(\beta)-n\ln(\theta)]^{n-1}\mathbf{I}(t-\beta^k\theta^n). \tag{5}$$

Proof. The joint pdf of (X_1,T) is obtained by making the following transformation.

$$\{y_1=x_1,y_2=x_2,\ldots,y_{n-1}=x_{n-1},t=x_1x_2\ldots x_n\}.$$

The Jacobian of transformation is $J=\frac{1}{y_1y_2\ldots y_{n-1}}$.

We have from (3)

$$h(y_1,y_2,\ldots,y_{n-1},t)=\frac{1}{y_1y_2\ldots y_{n-1}}f\left(y_1,y_2,\ldots,y_{n-1},\frac{t}{y_1y_2\ldots y_{n-1}}\right), \tag{6}$$

then integrating $y_2,y_3,\ldots,y_{n-1}$ over the respective range, the joint pdf of (Y_1,T) is

$$\begin{aligned}h(y_1,t)&=\frac{\alpha^n\theta^{n\alpha}\beta^{k\alpha}}{(n-2)!y_1}t^{-(\alpha+1)}\\&\quad\times\{b[\ln(t)-\ln(y_1)-(k-1)\ln(\beta)-(n-1)\ln(\theta)]^{n-2}\mathbf{I}(y_1-\beta\theta)\mathbf{I}(t-y_1\beta^{k-1}\theta^{n-1})\\&\quad+\bar{b}[\ln(t)-\ln(y_1)-k\ln(\beta)-(n-1)\ln(\theta)]^{n-2}\mathbf{I}(y_1-\theta)\mathbf{I}(t-y_1\beta^k\theta^{n-1})\}.\end{aligned} \tag{7}$$

Further, integrating $h(y_1,t)$ with respect to y_1 over the range of y_1, result is $h^*(t)$ given in (5). □

Now we obtain UMVUE of $f(x)$, $F(x)$ and rth moment.

Theorem 3.1. *For a given t*

(A) *$\hat{f}(x)$ is UMVUE of $f(x)$, where*

$$\begin{aligned}\hat{f}(x)&=\frac{(n-1)}{x[\ln(t)-k\ln(\beta)-n\ln(\theta)]^{n-1}}\\&\quad\times\{b[\ln(t)-\ln(x)-(k-1)\ln(\beta)-(n-1)\ln(\theta)]^{n-2}\mathbf{I}(t-x\beta^{k-1}\theta^{n-1})\mathbf{I}(x-\beta\theta)\\&\quad+\bar{b}[\ln(t)-\ln(x)-k\ln(\beta)-(n-1)\ln(\theta)]^{n-2}\mathbf{I}(t-x\beta^k\theta^{n-1})\mathbf{I}(x-\theta)\}.\end{aligned} \tag{8}$$

(B) *$\hat{F}(x)$ is UMVUE of $F(x)$, where*

$$\hat{F}(x) = 1 - \frac{1}{[\ln(t) - k\ln(\beta) - n\ln(\theta)]^{n-1}} \times \{b[\ln(t) - \ln(x) - (k-1)\ln(\beta) - (n-1)\ln(\theta)]^{n-1}\mathbf{I}(t - x\beta^{k-1}\theta^{n-1})\mathbf{I}(x - \beta\theta) + \bar{b}[\ln(t) - \ln(x) - k\ln(\beta) - (n-1)\ln(\theta)]^{n-1}\mathbf{I}(t - x\beta^{k}\theta^{n-1})\mathbf{I}(x - \theta)\}. \tag{9}$$

(C) *The rth moment of $X \sim \hat{f}(x)$ is the UMVUE of the rth moment of $X \sim f(x)$, where*

$$\widehat{\mathrm{E}(X^r)} = \frac{(n-1)!\theta^r(b\beta^r + \bar{b})}{[r(\ln(t) - k\ln(\beta) - n\ln(\theta))]^{n-1}} \left\{ t^r\beta^{-rk}\theta^{-rn} - \sum_{j=0}^{n-2} \frac{[r(\ln(t) - k\ln(\beta) - n\ln(\theta))]^j}{j!} \right\} \mathbf{I}(t - \beta^k\theta^n). \tag{10}$$

Proof. Case A: (8) can be proved by using the lemma.

Case B: We have

$$\begin{aligned}
\mathrm{E}(\hat{F}(x)) &= \int \hat{F}(x)h^*(t)\mathrm{d}t \\
&= 1 - \frac{\alpha^n\theta^{n\alpha}\beta^{k\alpha}}{(n-1)!} \\
&\quad \times \left\{ b\mathbf{I}(x - \beta\theta) \int_{x\beta^{k-1}\theta^{n-1}}^{\infty} t^{-(\alpha+1)}[\ln(t) - \ln(x) - (k-1)\ln(\beta) - (n-1)\ln(\theta)]^{n-1}\mathrm{d}t \right. \\
&\quad \left. + \bar{b}\mathbf{I}(x - \theta) \int_{x\beta^{k}\theta^{n-1}}^{\infty} t^{-(\alpha+1)}[\ln(t) - \ln(x) - k\ln(\beta) - (n-1)\ln(\theta)]^{n-1}\mathrm{d}t \right\}.
\end{aligned}$$

On simplification, we get $\mathrm{E}(\hat{F}(x)) = F(x)$.

Case C: We have

$$\begin{aligned}
\widehat{\mathrm{E}(X^r)} &= \int x^r\hat{f}(x)\mathrm{d}x = \frac{(n-1)}{[\ln(t) - k\ln(\beta) - n\ln(\theta)]^{n-1}} \\
&\quad \times \{b\mathbf{I}(t - \beta^k\theta^n) \int_{\beta\theta}^{t\beta^{1-k}\theta^{1-n}} x^{r-1}[\ln(t) - \ln(x) - (k-1)\ln(\beta) - (n-1)\ln(\theta)]^{n-2}\mathrm{d}x \\
&\quad + \bar{b}\mathbf{I}(t - \beta^k\theta^n) \int_{\theta}^{t\beta^{-k}\theta^{1-n}} x^{r-1}[\ln(t) - \ln(x) - k\ln(\beta) - (n-1)\ln(\theta)]^{n-2}\mathrm{d}x\}.
\end{aligned}$$

On simplification, we get $\widehat{\mathrm{E}(X^r)}$.

Further we have

$$\begin{aligned}
\mathrm{E}[\widehat{\mathrm{E}(X^r)}] &= \int_{\beta^k\theta^n}^{\infty} \widehat{\mathrm{E}(X^r)}h^*(t)\mathrm{d}t \\
&= \alpha^n\theta^{n\alpha+r}\beta^{k\alpha}r^{1-n}(b\beta^r + \bar{b}) \left[\beta^{-rk}\theta^{-rn} \int_{\beta^k\theta^n}^{\infty} t^{-\alpha+r-1}\mathrm{d}t \right. \\
&\quad \left. - \sum_{j=0}^{n-2} \frac{r^j}{j!} \int_{\beta^k\theta^n}^{\infty} t^{-\alpha-1}[\ln(t) - k\ln(\beta) - n\ln(\theta)]^j\mathrm{d}t \right].
\end{aligned}$$

If we put $z = \ln(t) - k\ln(\beta) - n\ln(\theta)$, then

$$\begin{aligned}
\mathrm{E}[\widehat{\mathrm{E}(X^r)}] &= \alpha^n \theta^{n\alpha+r}\beta^{k\alpha} r^{1-n}(b\beta^r + \bar{b})\left[\beta^{-rk}\theta^{-rn}\frac{(\beta^k\theta^n)^{r-\alpha}}{\alpha - r} - \sum_{j=0}^{n-2}\frac{r^j}{j!}\beta^{-k\alpha}\theta^{-n\alpha}\int_0^\infty z^j \mathrm{e}^{-\alpha z}\mathrm{d}z\right] \\
&= \alpha^n \theta^{n\alpha+r}\beta^{k\alpha} r^{1-n}(b\beta^r + \bar{b})\left[\frac{\beta^{-k\alpha}\theta^{-n\alpha}}{\alpha - r} - \beta^{-k\alpha}\theta^{-n\alpha}\frac{1}{\alpha}\sum_{j=0}^{n-2}\left(\frac{r}{\alpha}\right)^j\right] \\
&= \frac{\alpha\theta^r}{\alpha - r}(b\beta^r + \bar{b}) \\
&= \mathrm{E}(X^r),
\end{aligned}$$

by using

$$\sum_{j=0}^{n-2}\left(\frac{r}{\alpha}\right)^j = \frac{\alpha^{n-1} - r^{n-1}}{(\alpha - r)\alpha^{n-2}},$$

and the proof is complete. □

Note: UMVUE of α is

$$\hat{\alpha} = \frac{n-1}{\ln(t) - n\ln(\theta) - k\ln(\beta)}, \quad t > \beta^k\theta^n, \tag{11}$$

and

$$\mathrm{E}(\hat{\alpha}^r) = \frac{\Gamma(n-r)(n-1)^r}{\Gamma(n)}\alpha^r, \quad r \le n-1. \tag{12}$$

For $r = 1$, $\mathrm{E}(\hat{\alpha}) = \alpha$.

4. MSE of UMVU estimator

In this section, we obtain the MSE of $\hat{f}(x)$ and $\hat{F}(x)$.

Theorem 4.1. (A)

$$\begin{aligned}
MSE(\hat{f}(x)) = {} & \frac{(n-1)\alpha^{2n}\theta^\alpha}{(n-2)!x^{\alpha+2}}\{b^2\beta^\alpha \mathbf{A}_i^{2n-4}(\beta\theta)\mathbf{B}_i(\beta\theta) \\
& + \bar{b}^2\mathbf{A}_i^{2n-4}(\theta)\mathbf{B}_i(\theta) + 2b\bar{b}\mathbf{A}_i^{n-2}(\theta)\mathbf{A}_j^{n-2}(\beta\theta)\mathbf{B}_{i+j}(\theta)\} \\
& - \left[b\frac{\alpha(\beta\theta)^\alpha}{x^{\alpha+1}}\mathbf{I}(x-\beta\theta) + \bar{b}\frac{\alpha\theta^\alpha}{x^{\alpha+1}}\mathbf{I}(x-\theta)\right]^2,
\end{aligned} \tag{13}$$

(B)

$$\begin{aligned}
MSE(\hat{F}(x)) = {} & \frac{\alpha^{2n}\theta^\alpha}{(n-1)!x^\alpha}\{b^2\beta^\alpha \mathbf{A}_i^{2n-2}(\beta\theta)\mathbf{B}_i(\beta\theta) \\
& + \bar{b}^2\mathbf{A}_i^{2n-2}(\theta)\mathbf{B}_i(\theta) + 2b\bar{b}\mathbf{A}_i^{n-1}(\theta)\mathbf{A}_j^{n-1}(\beta\theta)\mathbf{B}_{i+j}(\theta)\} \\
& - \left(\frac{\theta}{x}\right)^{2\alpha}\left[b\beta^\alpha\mathbf{I}(x-\beta\theta) + \bar{b}\mathbf{I}(x-\theta)\right]^2,
\end{aligned} \tag{14}$$

where

$$\mathbf{A}_i^n(\theta) = \sum_{i=0}^{n} C(n,i)\left[\ln(\theta) - \ln(x)\right]^{n-i}\mathbf{I}(x-\theta),$$

and

$$\mathbf{B}_i(\theta) = \Gamma(i-n+2)\alpha^{-2-i}\sum_{l=0}^{i-n+1}\frac{[\alpha(\ln(x)-\ln(\theta))]^l}{l!}\mathbf{I}(x-\theta).$$

Proof. Case A: We can find $\mathrm{E}(\hat{f}(x))^2$ using the pdf of T in (5)

$$\mathrm{E}(\hat{f}(x))^2 = \frac{(n-1)\theta^\alpha\alpha^{2n}}{(n-2)!x^{\alpha+2}} \times [b^2\beta^\alpha\mathbf{A}_i^{2n-4}(\beta\theta)\mathbf{B}_i(\beta\theta) + \bar{b}^2\mathbf{A}_i^{2n-4}(\theta)\mathbf{B}_i(\theta) + 2b\bar{b}\mathbf{A}_i^{n-2}(\theta)\mathbf{A}_j^{n-2}(\beta\theta)\mathbf{B}_{i+j}(\theta)], \quad (15)$$

by using the following relation.

$$[t-\ln(x)+\ln(\theta)]^{2n-4} = \sum_{i=0}^{2n-4}C(2n-4,i)t^i[\ln(\theta)-\ln(x)]^{2n-4-i}.$$

Then, we obtain the MSE of $\hat{f}(x)$.

Case B: We can find $\mathrm{E}(\hat{F}(x))^2$ using (5).

Hence

$$\begin{aligned}\mathrm{E}(\hat{F}(x))^2 = {} & \frac{\alpha^{2n}\theta^\alpha}{(n-1)!x^\alpha}\{b^2\beta^\alpha\mathbf{A}_i^{2n-2}(\beta\theta)\mathbf{B}_i(\beta\theta) \\ & + \bar{b}^2\mathbf{A}_i^{2n-2}(\theta)\mathbf{B}_i(\theta) + 2b\bar{b}\mathbf{A}_i^{n-1}(\theta)\mathbf{A}_j^{n-1}(\beta\theta)\mathbf{B}_{i+j}(\theta)\} \\ & + 1 - 2\left(\frac{\theta}{x}\right)^\alpha[b\beta^\alpha\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)]. \end{aligned} \quad (16)$$

Then we get the MSE of $\hat{F}(x)$ and the proof is complete. □

5. Maximum likelihood estimator

Using (3), the likelihood equation for α is

$$\frac{n}{\alpha} + n\ln(\theta) + k\ln(\beta) - \sum_{i=1}^n \ln(x_i) = 0. \quad (17)$$

Hence solving the above equation, we get MLE of α as given below:

$$\tilde{\alpha} = \frac{n}{\sum_{i=1}^n \ln(x_i) - n\ln(\theta) - k\ln(\beta)}, \quad \sum_{i=1}^n \ln(x_i) > \ln(\theta^n\beta^k). \quad (18)$$

Using the property of MLE, we can obtain the estimator of pdf, cdf and the rth moment by using $\tilde{\alpha}$ instead of α in the pdf, cdf and rth moment, respectively. So

$$\tilde{f}(x) = \frac{\tilde{\alpha}\theta^{\tilde{\alpha}}}{x^{\tilde{\alpha}+1}}\left[b\beta^{\tilde{\alpha}}\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)\right], \quad \tilde{\alpha}>0,\ \theta>0,\ \beta>1, \quad (19)$$

$$\tilde{F}(x) = 1-\left(\frac{\theta}{x}\right)^{\tilde{\alpha}}\left[b\beta^{\tilde{\alpha}}\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)\right], \quad \tilde{\alpha}>0,\ \theta>0,\ \beta>1, \quad (20)$$

and

$$\widetilde{\mathrm{E}(X^r)} = \frac{\tilde{\alpha}\theta^r}{\tilde{\alpha}-r}(b\beta^r+\bar{b}), \quad \tilde{\alpha}>r,\ \theta>0,\ \beta>1. \quad (21)$$

Now we will find the distribution of $\tilde{\alpha}$.

Let $W_1 = \sum_{i=1}^{n-k} \ln\left(\frac{X_i}{\theta}\right)$ and $W_2 = \sum_{i=1}^{k} \ln\left(\frac{X_i}{\theta}\right)$. So

$$g(w_1) = \frac{\alpha^{n-k} w_1^{n-k-1}}{\Gamma(n-k)} \exp(-\alpha w_1), \quad w_1 > 0, \tag{22}$$

and

$$g(w_2) = \frac{\alpha^k \beta^{k\alpha} (w_2 - k\ln(\beta))^{k-1}}{\Gamma(k)} \exp(-\alpha w_2), \quad w_2 > k\ln(\beta). \tag{23}$$

Since W_1 and W_2 are independent then after some elementary algebra, we can find the distribution of $\tilde{\alpha}$. Hence

$$g(\tilde{\alpha}) = \frac{(\alpha n)^n}{\Gamma(n)(\tilde{\alpha})^{n+1}} \exp\left\{-\frac{\alpha n}{\tilde{\alpha}}\right\}, \quad \tilde{\alpha} > 0. \tag{24}$$

For more details see [4].

Theorem 5.1. (A) *$\tilde{f}(x)$ is a biased estimator of $f(x)$ and*

$$\mathrm{E}(\tilde{f}(x)) = \frac{1}{x\Gamma(n)} \sum_{j=0}^{n-2} \frac{(\alpha n)^{j+1}}{j!} \Gamma(n-j-1) \times \left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta) + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right]. \tag{25}$$

(B) *$\tilde{F}(x)$ is a biased estimator of $F(x)$ and*

$$\mathrm{E}(\tilde{F}(x)) = 1 - \frac{1}{\Gamma(n)} \sum_{j=0}^{n-1} \frac{(\alpha n)^j}{j!} \Gamma(n-j) \times \left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta) + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right]. \tag{26}$$

(C) *$\widetilde{\mathrm{E}(X^r)}$ is a biased estimator of $\mathrm{E}(X^r)$ and*

$$\mathrm{E}(\widetilde{\mathrm{E}(X^r)}) = \frac{\theta^r \mathrm{e}^{\frac{-n\alpha}{r}}}{\Gamma(n)} \left(b\beta^r + \bar{b}\right) \sum_{j=0}^{\infty} \Gamma(n+j) \left(\frac{r}{\alpha n}\right)^j \left[1 - \sum_{i=0}^{n+j-1} \frac{\left(\frac{n\alpha}{r}\right)^i}{i!}\right]. \tag{27}$$

Proof. Case A: We have

$$\begin{aligned}\mathrm{E}(\tilde{f}(x)) &= \int_0^\infty \tilde{f}(x) g(w)\mathrm{d}w \\ &= \frac{(\alpha n)^n}{x\Gamma(n)} \left[b\mathbf{I}(x-\beta\theta) \int_0^\infty \left(\frac{\beta\theta}{x}\right)^w \frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^n}\mathrm{d}w + \bar{b}\mathbf{I}(x-\theta) \int_0^\infty \left(\frac{\theta}{x}\right)^w \frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^n}\mathrm{d}w \right],\end{aligned}$$

where $g(w)$ is given in (24).

Note that

$$\left(\frac{\beta\theta}{x}\right)^w = \mathrm{e}^{w\ln\left(\frac{\beta\theta}{x}\right)} = \sum_{j=0}^{\infty} \frac{w^j \left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j}{j!}, \quad x \ge \beta\theta,$$

and

$$\left(\frac{\theta}{x}\right)^{w} = \mathrm{e}^{w\ln\left(\frac{\theta}{x}\right)} = \sum_{j=0}^{\infty} \frac{w^j\left(\ln\left(\frac{\theta}{x}\right)\right)^j}{j!}, \quad x \geq \theta.$$

Hence

$$\mathrm{E}(\tilde{f}(x)) = \frac{(\alpha n)^n}{x\Gamma(n)} \sum_{j=0}^{\infty} \frac{1}{j!}\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta) + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right]\int_0^{\infty} \frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n-j}}\mathrm{d}w,$$

which result in

$$= \frac{1}{x\Gamma(n)} \sum_{j=0}^{\infty} \frac{(\alpha n)^{j+1}}{j!}\Gamma(n-j-1)\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta) + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right],$$

where $j \leq (n-1)$.

Case B: We have

$$\mathrm{E}(\tilde{F}(x)) = \int_0^{\infty} \tilde{F}(x)g(w)\mathrm{d}w.$$

Then

$$\begin{aligned}\mathrm{E}(\tilde{F}(x)) &= 1 - \frac{(\alpha n)^n}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{1}{j!}\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta)\right. \\ &\qquad \left. + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right]\int_0^{\infty} \frac{\mathrm{e}^{\frac{-\alpha n}{w}}}{w^{n-j+1}}\mathrm{d}w \\ &= 1 - \frac{1}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{(\alpha n)^j}{j!}\Gamma(n-j)\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta) + \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\right].\end{aligned}$$

Case C: We can obtain

$$\mathrm{E}(\widetilde{\mathrm{E}(X^r)}) = \int_0^{\infty} \widetilde{\mathrm{E}(X^r)}g(w)\mathrm{d}w.$$

Hence

$$\begin{aligned}\mathrm{E}(\widetilde{\mathrm{E}(X^r)}) &= \frac{(\alpha n)^n}{\Gamma(n)}\theta^r\left(b\beta^r + \bar{b}\right)\sum_{j=0}^{\infty} r^j \int_r^{\infty} \frac{\mathrm{e}^{\frac{-\alpha n}{w}}}{w^{n+j+1}}\mathrm{d}w \\ &= \frac{\theta^r \mathrm{e}^{\frac{-n\alpha}{r}}}{\Gamma(n)}\left(b\beta^r + \bar{b}\right)\sum_{j=0}^{\infty}\Gamma(n+j)\left(\frac{r}{\alpha n}\right)^j\left[1 - \sum_{i=0}^{n+j-1}\frac{\left(\frac{n\alpha}{r}\right)^i}{i!}\right],\end{aligned}$$

by using

$$\frac{1}{w-r} = \frac{1}{w}\sum_{j=0}^{\infty}\left(\frac{r}{w}\right)^j, \quad w > r,$$

and the proof is complete. □

Note: $\mathrm{E}(\widetilde{\mathrm{E}(X^r)})$ tends to infinity as $n \to \infty$.

6. MSE of ML estimator

In the previous section, we found the MLE of $f(x)$, $F(x)$ and rth moment. Now we try to obtain the MSE of $\tilde{f}(x)$ and $\tilde{F}(x)$.

Theorem 6.1. (A)

$$\begin{aligned}
\mathrm{MSE}(\tilde{f}(x)) = {} & \frac{1}{x^2\Gamma(n)}\sum_{j=0}^{n-3}\frac{(\alpha n)^{j+2}}{j!}\Gamma(n-j-2) \\
& \times\left[b^2\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)+\bar{b}^2\left(2\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)\right. \\
& \left.+\ 2b\bar{b}\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j\mathbf{I}(x^2-\beta\theta^2)\right] \\
& -2\frac{\alpha\theta^\alpha}{x^{\alpha+2}\Gamma(n)}\left[b\beta^\alpha\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)\right] \\
& \times\sum_{j=0}^{n-2}\frac{(\alpha n)^{j+1}}{j!}\Gamma(n-j-1)\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)\right. \\
& \left.+\ \bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)\right]+\frac{\alpha^2\theta^{2\alpha}}{x^{2\alpha+2}}\left[b\beta^\alpha\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)\right]^2. \qquad (28)
\end{aligned}$$

(B)

$$\begin{aligned}
\mathrm{MSE}(\tilde{F}(x)) = {} & \frac{1}{\Gamma(n)}\sum_{j=0}^{n-1}\frac{(\alpha n)^j}{j!}\Gamma(n-j) \\
& \times\left[-2b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)-2\bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)\right. \\
& +b^2\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)+2b\bar{b}\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j\mathbf{I}(x^2-\beta\theta^2) \\
& \left.+\bar{b}^2\left(2\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)\right] \\
& +\frac{2}{\Gamma(n)}\left[1-\left(\frac{\theta}{x}\right)^\alpha\left(b\beta^\alpha\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)\right)\right]\sum_{j=0}^{n-1}\frac{(\alpha n)^j}{j!}\Gamma(n-j) \\
& \times\left[b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)+\bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)\right] \\
& +\left(\frac{\theta}{x}\right)^{2\alpha}[b\beta^\alpha\mathbf{I}(x-\beta\theta)+\bar{b}\mathbf{I}(x-\theta)]^2. \qquad (29)
\end{aligned}$$

Proof. Case A:

$$\begin{aligned}
\mathrm{E}(\tilde{f}(x))^2 &= \int_0^\infty(\tilde{f}(x))^2g(w)\mathrm{d}w \\
&= \frac{(\alpha n)^n}{x^2\Gamma(n)}\left[b^2\mathbf{I}(x-\beta\theta)\int_0^\infty\left(\frac{\beta\theta}{x}\right)^{2w}\frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n-1}}\mathrm{d}w+\bar{b}^2\mathbf{I}(x-\theta)\int_0^\infty\left(\frac{\theta}{x}\right)^{2w}\frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n-1}}\mathrm{d}w\right.
\end{aligned}$$

$$+\ 2b\bar{b}\mathbf{I}(x^2-\beta\theta^2)\int_0^\infty\left(\frac{\beta\theta^2}{x^2}\right)^w\frac{\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n-1}}\mathrm{d}w\Bigg],$$

and to evaluate these integrals, we use the following equations.

$$\left(\frac{\beta\theta}{x}\right)^{2w}=\mathrm{e}^{2w\ln\left(\frac{\beta\theta}{x}\right)}=\sum_{j=0}^{\infty}\frac{w^j\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j}{j!},\quad x\geq\beta\theta,$$

$$\left(\frac{\theta}{x}\right)^{2w}=\mathrm{e}^{2w\ln\left(\frac{\theta}{x}\right)}=\sum_{j=0}^{\infty}\frac{w^j\left(2\ln\left(\frac{\theta}{x}\right)\right)^j}{j!},\quad x\geq\theta,$$

and

$$\left(\frac{\beta\theta^2}{x^2}\right)^{w}=\mathrm{e}^{w\ln\left(\frac{\beta\theta^2}{x^2}\right)}=\sum_{j=0}^{\infty}\frac{w^j\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j}{j!},\quad x^2\geq\beta\theta^2.$$

Hence

$$\mathrm{E}(\tilde{f}(x))^2=\frac{1}{x^2\Gamma(n)}\sum_{j=0}^{\infty}\frac{(\alpha n)^{j+2}}{j!}\Gamma(n-j-2)\left[b^2\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)\right.$$
$$\left.+\ \bar{b}^2\left(2\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)+2b\bar{b}\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j\mathbf{I}(x^2-\beta\theta^2)\right].$$

This implies that

$$\mathrm{E}(\tilde{f}(x))^2=\frac{1}{x^2\Gamma(n)}\sum_{j=0}^{n-3}\frac{(\alpha n)^{j+2}}{j!}\Gamma(n-j-2)\left[b^2\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)\right.$$
$$\left.+\ \bar{b}^2\left(2\ln\left(\frac{\theta}{x}\right)\right)^j\mathbf{I}(x-\theta)+2b\bar{b}\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j\mathbf{I}(x^2-\beta\theta^2)\right],\tag{30}$$

since $\Gamma(y)$ is defined for $y>0$.

So by using some elementary algebra, we can get the MSE of $\tilde{f}(x)$.

Case B: We can obtain $\mathrm{E}(\tilde{F}(x))^2$ as

$$\mathrm{E}(\tilde{F}(x))^2=\int_0^\infty(\tilde{F}(x))^2g(w)\mathrm{d}w$$
$$=1+\frac{(\alpha n)^n}{\Gamma(n)}\left[-2b\mathbf{I}(x-\beta\theta)\int_0^\infty\frac{\left(\frac{\beta\theta}{x}\right)^w\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n+1}}\mathrm{d}w-2\bar{b}\mathbf{I}(x-\theta)\int_0^\infty\frac{\left(\frac{\theta}{x}\right)^w\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n+1}}\mathrm{d}w\right.$$
$$+b^2\mathbf{I}(x-\beta\theta)\int_0^\infty\frac{\left(\frac{\beta\theta}{x}\right)^{2w}\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n+1}}\mathrm{d}w+\bar{b}^2\mathbf{I}(x-\theta)\int_0^\infty\frac{\left(\frac{\theta}{x}\right)^{2w}\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n+1}}\mathrm{d}w$$
$$\left.+\ 2b\bar{b}\mathbf{I}(x^2-\beta\theta^2)\int_0^\infty\frac{\left(\frac{\beta\theta^2}{x^2}\right)^w\mathrm{e}^{-\frac{\alpha n}{w}}}{w^{n+1}}\mathrm{d}w\right]$$
$$=1+\frac{1}{\Gamma(n)}\sum_{j=0}^{n-1}\frac{(\alpha n)^j}{j!}\Gamma(n-j)\left[-2b\left(\ln\left(\frac{\beta\theta}{x}\right)\right)^j\mathbf{I}(x-\beta\theta)\right.$$

$$- 2\bar{b}\left(\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta) + b^2\left(2\ln\left(\frac{\beta\theta}{x}\right)\right)^j \mathbf{I}(x-\beta\theta)$$
$$+ 2b\bar{b}\left(\ln\left(\frac{\beta\theta^2}{x^2}\right)\right)^j \mathbf{I}(x^2-\beta\theta^2) + \bar{b}^2\left(2\ln\left(\frac{\theta}{x}\right)\right)^j \mathbf{I}(x-\theta)\Bigg]. \quad (31)$$

Therefore we can obtain the MSE of $\tilde{F}(x)$ and the proof is complete. □

Note: If k is unknown, then k can be selected by evaluating the likelihood for different values of k choosing the one that maximizes the likelihood.

7. Comparison of MLE and UMVUE

In order to get an idea of the efficiency of the two types of estimation i.e MLE and UMVUE, we have generated a sample of size 10, 15, 20, ..., 40 from the Pareto distribution in the presence of outliers with $k = 1, 2, 3$, $\alpha = 0.5, 1, 1.5, 2$, $\beta = 1.5, 2$ and $\theta = 0.5, 1$. For example, for $k = 1$, $\alpha = 0.5$, $\beta = 1.5$ and $\theta = 0.5$ a sample of size 10 is generated such that a sample of size 9 is taken from (2) and a sample of size one is taken from (1). For these observations, we have calculated exact MSE of the estimators from (13), (14), (28) and (29). Then, we have taken the average of MSE of 10 observations. This process is repeated 1000 times. Further, these 1000 MSEs were divided by 1000 using R software. Then we have plotted all these graphs in Figs. 1–4.

From the graphs, it has been seen that ML estimators of pdf and cdf are more efficient than their UMVU estimators.

From (27), we can conclude that expected value of the MLE of rth moment does not exist, hence in this case UMVU estimator is better.

8. An illustrative example

For insurance company one of its services is motor insurance. A claim of at least 500,000 Rials (Iranian Rials) as compensation for the motor insurance can be made. The vehicles involved are of different costs, of which some of them may have a very high cost. Claim amounts vary according to the damage to the vehicles. The company had assumed that claims of expensive/severely damaged vehicles are 1.5 times higher than the normal vehicles. In this paper, we have drawn a random sample of size 20 of the claim amounts. It is observed that such claims follow the Pareto distribution in the presence of outliers with parameters α, β and θ, where α is unknown, $\beta = 1.5$ and $\theta = 500{,}000$ and the number of outliers (k) is unknown. One should note that for normal vehicles claims below 500,000 Rials are not entertained.

Here, even if company assumed a different value of β ($\beta > 1$), for example $\beta = 2$, the data of claims has the Pareto distribution in the presence of outliers with parameters α, β and θ, where α is unknown, $\beta = 2$ and $\theta = 500{,}000$ and the number of outliers (k) is unknown.

The data of claims from an insurance company in Iran records for the year 2008 is given below:

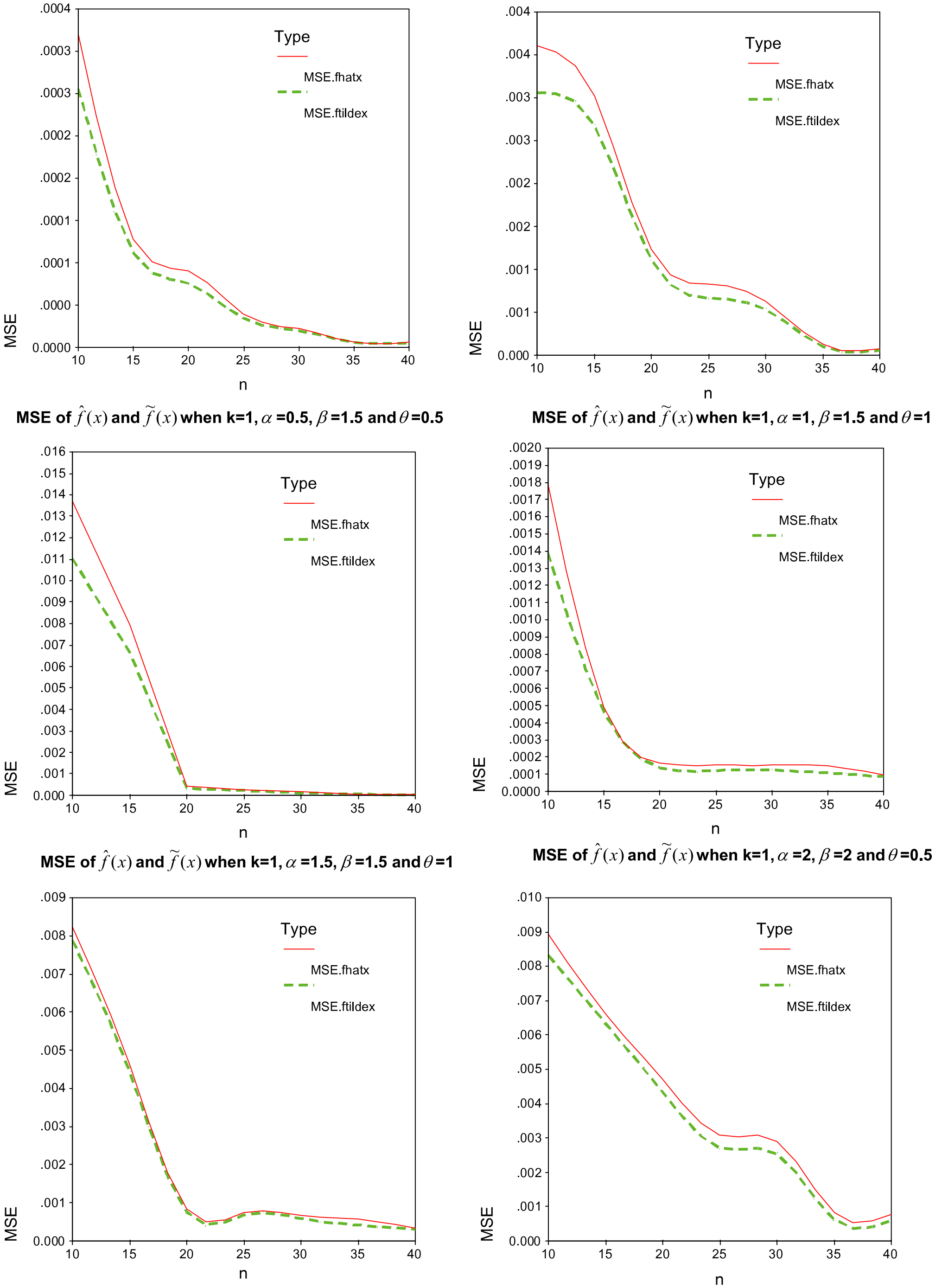
$$\begin{array}{c} 750{,}000,\ 780{,}000,\ 630{,}000,\ 1750{,}000,\ 1450{,}000 \\ 3000{,}000,\ 7650{,}000,\ 4210{,}000,\ 890{,}000,\ 950{,}000 \\ 1240{,}000,\ 1800{,}000,\ 1630{,}000,\ 9020{,}000,\ 4750{,}000 \\ 3250{,}000,\ 1135{,}000,\ 1326{,}000,\ 1280{,}000,\ 760{,}000. \end{array}$$

So $\hat{\alpha}$ (UMVUE of α) and $\tilde{\alpha}$ (MLE of α) for $k = 1, 2, 3$ are shown in Table 1.

Also from the likelihood function corresponding to k, $\mathbf{L}(\underline{x}; \hat{\alpha})$ and $\mathbf{L}(\underline{x}; \tilde{\alpha})$ for $k = 1, 2, 3$ are shown in Table 2.

The likelihood function is maximized for $k = 1$, $\hat{\alpha} = 0.7786114$ and $\tilde{\alpha} = 0.819591$.

Therefore for $n = 20$, $k = 1$, $\beta = 1.5$ and $\theta = 500{,}000$ the final result of UMVUE and MLE of $f(x)$ and $F(x)$ corresponding to $\hat{\alpha}$, $\tilde{\alpha}$ and the first observation (750,000) are given in Table 3.



MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=1, α =0.5, β =1.5 and θ =0.5

MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=1, α =1, β =1.5 and θ =1

MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=1, α =1.5, β =1.5 and θ =1

MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=1, α =2, β =2 and θ =0.5

MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=2, α =0.5, β =1.5 and θ =0.5

MSE of $\hat{f}(x)$ and $\tilde{f}(x)$ when k=2, α =1, β =1.5 and θ =1

Fig. 1. Comparison the MSE of the estimators of pdf respect to observation generated from the Pareto distribution in the presence of outliers.

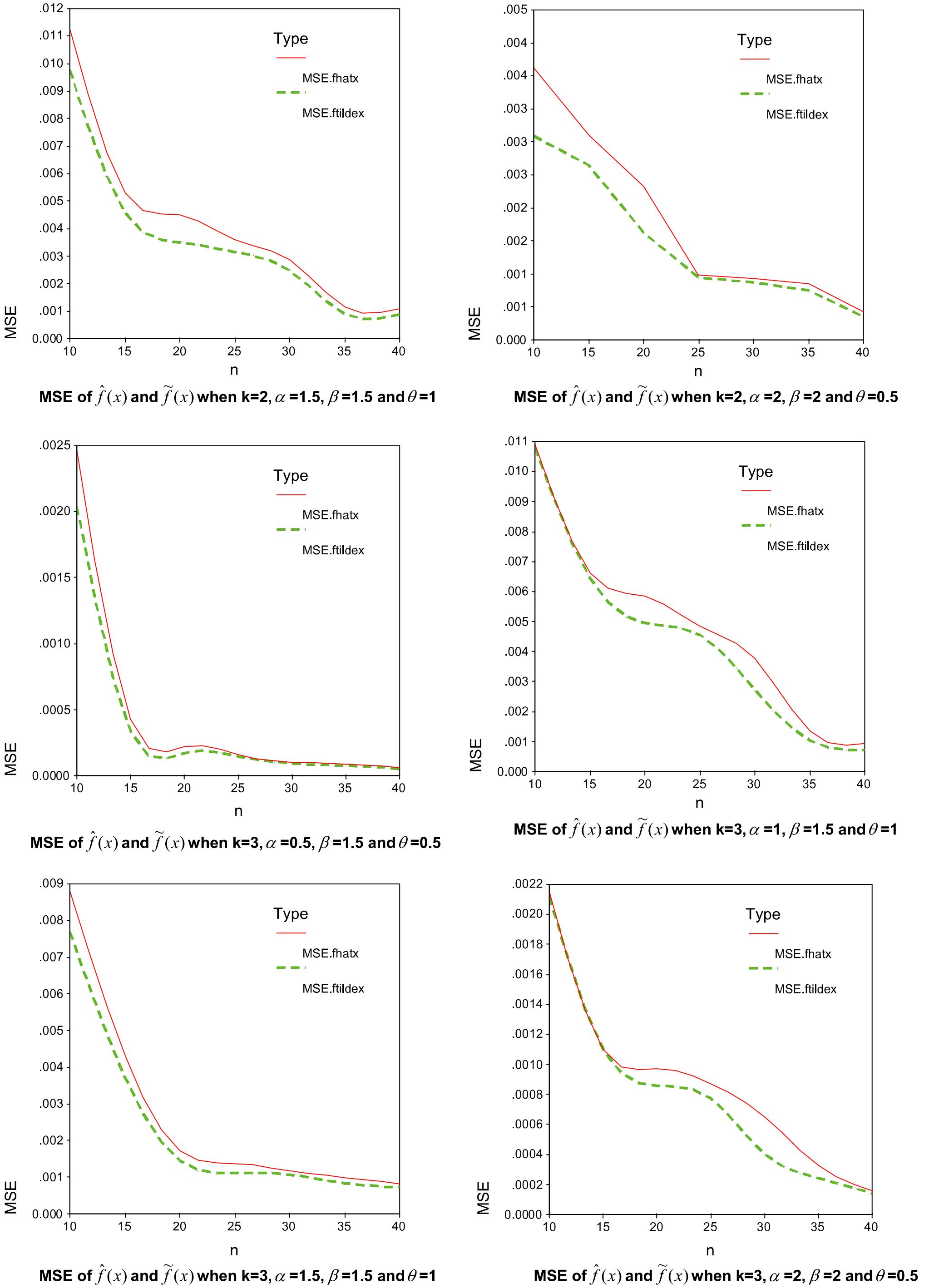


Fig. 2. Comparison the MSE of the estimators of pdf respect to observation generated from the Pareto distribution in the presence of outliers.

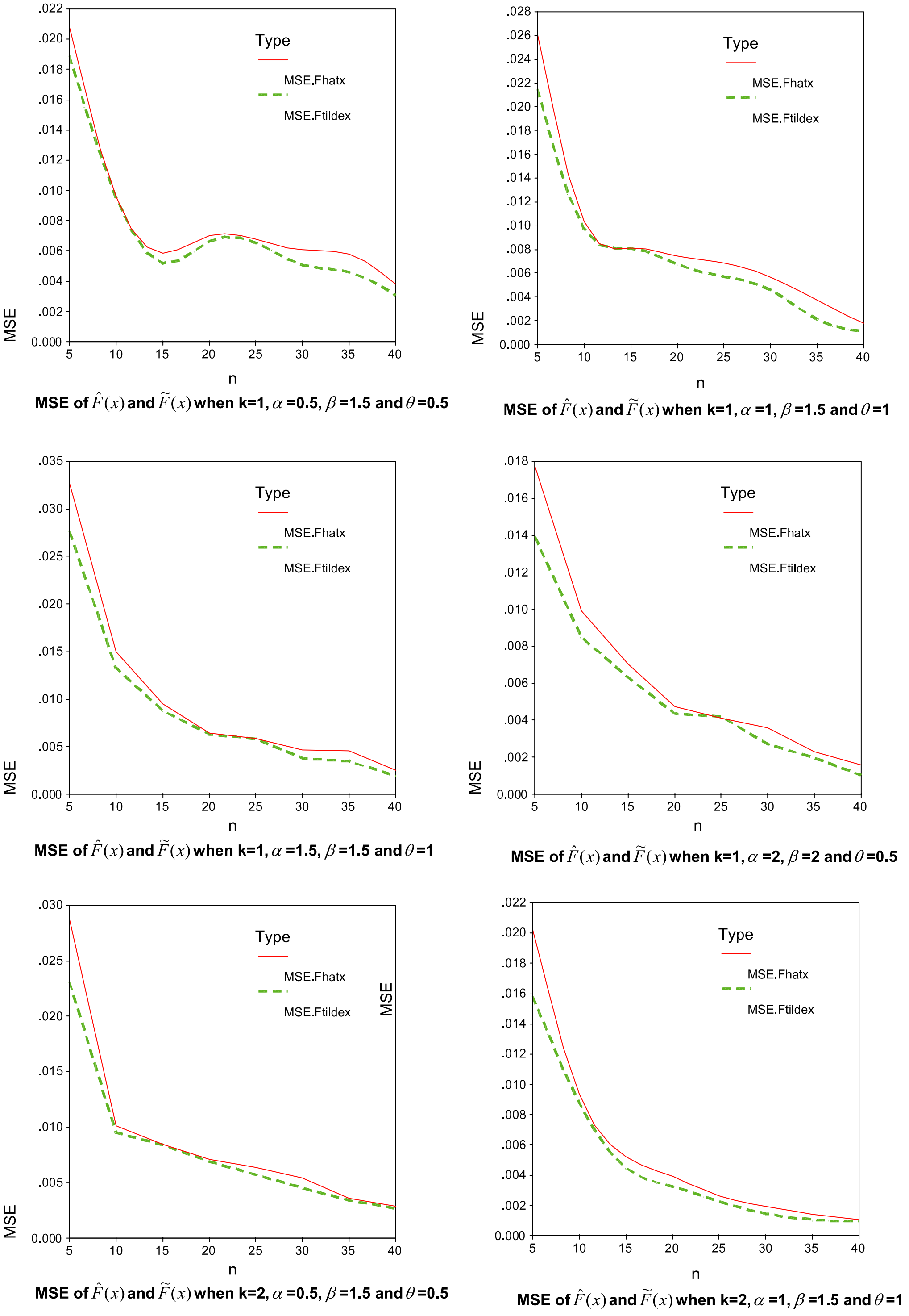
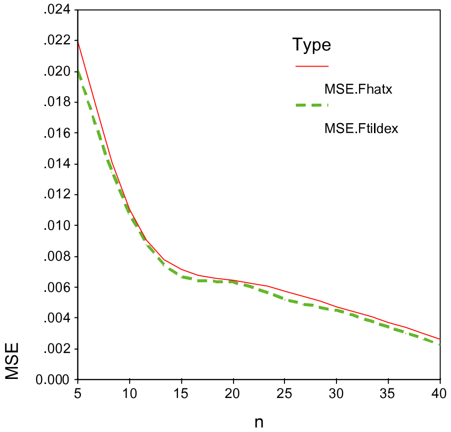
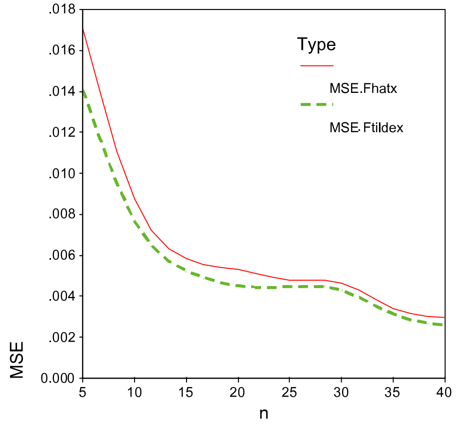


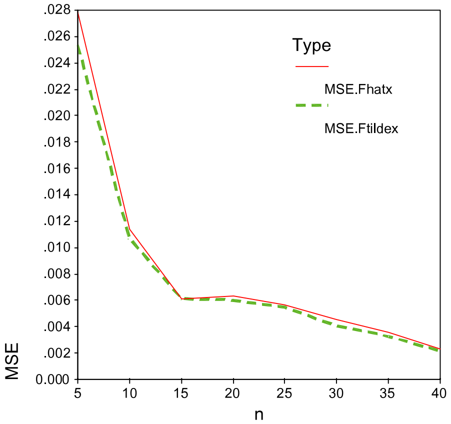
Fig. 3. Comparison the MSE of the estimators of cdf respect to observation generated from the Pareto distribution in the presence of outliers.



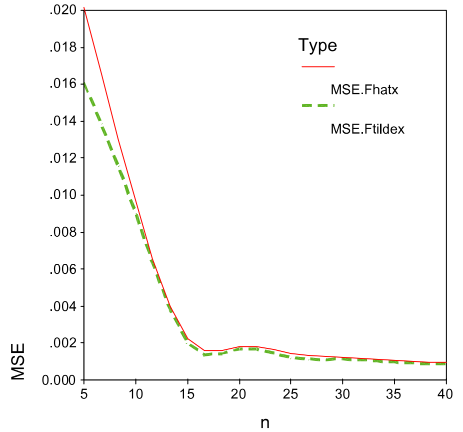
MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=2, α =1.5, β =1.5 and θ =1



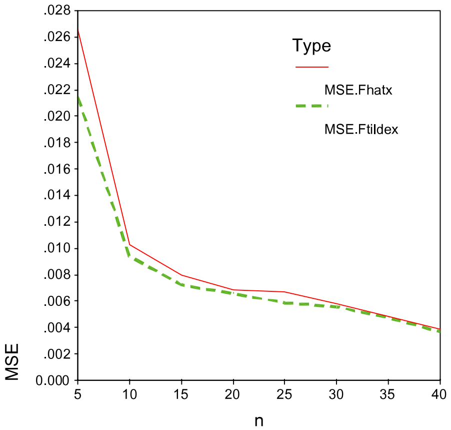
MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=2, α =2, β =2 and θ =0.5



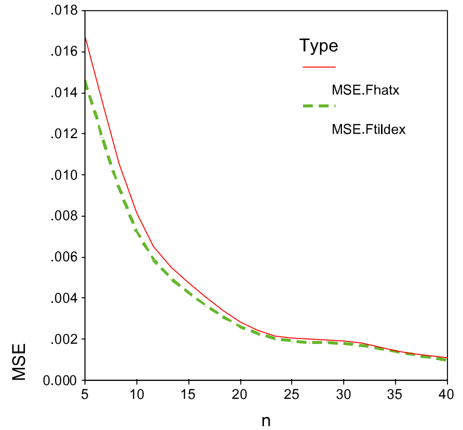
MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=3, α =0.5, β =1.5 and θ =0.5



MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=3, α =1, β =1.5 and θ =1



MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=3, α =1.5, β =1.5 and θ =1



MSE of $\hat{F}(x)$ and $\widetilde{F}(x)$ when k=3, α =2, β =2 and θ =0.5

Fig. 4. Comparison the MSE of the estimators of cdf respect to observation generated from the Pareto distribution in the presence of outliers.

Table 1
UMVUE and MLE of α for $\beta = 1.5$ and $\theta = 500{,}000$.

k	$\hat{\alpha}$	$\tilde{\alpha}$
1	0.7786114	0.819591
2	0.7917672	0.8334392
3	0.8053753	0.8477634

Table 2
The likelihood function corresponding to k.

k	$\mathbf{L}(\underline{x}; \hat{\alpha})$	$\mathbf{L}(\underline{x}; \tilde{\alpha})$
1	3.314979e−137	3.401843e−137
2	4.878569e−138	5.006404e−138
3	1.143291e−138	1.173248e−138

Table 3
UMVUE and MLE of pdf and cdf for $n = 20, k = 1, \beta = 1.5, \theta = 500{,}000$ and $x = 750{,}000$.

$\hat{f}(x)$	7.81366 e−07
$\tilde{f}(x)$	7.992615 e−07
$\hat{F}(x)$	0.2590218
$\tilde{F}(x)$	0.2686033

Acknowledgements

The authors are thankful to the editors and the referees for their valuable comments.

References

[1] B.R. Asrabadi, Estimation in the Pareto distribution, Metrika 37 (1990) 199–205.
[2] G. Benktander, A note on the most "dangerous" and skewest class of distribution, Astin Bulletin 2 (1963) 387–390.
[3] U.J. Dixit, Estimation of parameters of the gamma distribution in the presence of outliers, Communications in Statistics. Theory and Methods 18 (8) (1989) 3071–3085.
[4] U.J. Dixit, On the estimation of power of scale parameter of the gamma distribution in the presence of outliers, Communications in Statistics. Theory and Methods 20 (4) (1991) 1315–1328.
[5] U.J. Dixit, M.M. Ali, Jungsoo Woo, Efficient estimation of parameters of a uniform distribution in the presence of outliers, Soochow Journal of Mathematics 29 (4) (2003) 363–369.
[6] U.J. Dixit, M. Jabbari Nooghabi, Efficient estimation in the Pareto distribution, Statistical Methodology 7 (6) (2010) 687–691.
[7] U.J. Dixit, K.L. Moore, V. Barnett, On the estimation of the scale parameter of the exponential distribution in the presence of outliers generated from uniform distribution, Metron LIV (3–4) (1996) 201–211.
[8] U.J. Dixit, F.P. Nasiri, Estimation of parameters of the exponential distribution in the presence of outliers generated from uniform distribution, Metron LIX (3–4) (2001) 187–198.
[9] T.P. Nadeau, T.J. Teorey, A Pareto model for OLAP view size estimation, Information Systems Frontiers 5 (2) (2003) 137–147.