

# On estimation and goodness-of-fit testing for the Pareto distribution

**T Nombebe**

 [orcid.org/0000-0002-5940-9738](https://orcid.org/0000-0002-5940-9738)

Thesis accepted in fulfilment of the requirements for the degree  
*Doctor of Philosophy in Science with Statistics* at the North-West  
University

Promoter: Prof JS Allison

Co-promoter: Prof IJH Visagie

Co-promoter: Prof L Santana

Assistant promoter: Prof J Ngatchou-Wandji

Graduation May 2024

# Acknowledgements

I wish to express appreciation and gratitude to the following individuals:

- My supervisors, Prof James Allison, Prof Leonard Santana and Prof Jaco Visagie for their time, invaluable guidance, for sharing expertise, consistent support, for being patient and for their ability to provide clarity in complex subjects. This study would not have been accomplished without your assistance and dedicated involvement in every step throughout the process. I am eternally grateful for the privilege of having learned under such esteemed scholars.
- Prof Joseph Ngatchou-Wandji with whom I had the privilege of collaborating during this endeavor, your expertise and collaborative spirit played a pivotal role in elevating the quality of my dissertation.
- Dr Lethani Ndwandwe (the King), thank you for your insightful discussions, constant encouragement, and unwavering support throughout the study.
- All of my colleagues in the department for their encouragement and support.
- My family, for their love and support.
- My partner, who supported me through this venture.

Most importantly, I thank Almighty God for providing the good health and well-being that were necessary to complete this study.

# Abstract

The Pareto distributions are a class of distributions containing several members with varying levels of flexibility. Two of these members, the Pareto Type I and the Lomax distributions, are of particular interest in this thesis. The Pareto Type I distribution is a popular choice of model in many applications where heavy tailed data are modelled. This flexible distribution is often used to model the tails of, for example, observed insurance claims or salaries. The Lomax distribution is a generalisation of the Pareto Type I distribution (and a special case of the Pareto Type II distribution) which was initially proposed as a model for business failure, but has since found a wide range of other applications.

We propose several classes of goodness-of-fit tests for the Pareto Type I distribution based on two characterisations. The first characterisation relates to the distribution of order statistics while the second is based on conditional expectation. We derive the asymptotic properties of the tests based on the first of these characterisations. Extensive Monte Carlo studies are included in order to examine the finite sample power performance of the proposed tests and it is found that these tests are competitive, often outperforming the existing tests considered.

In the Monte Carlo power studies mentioned above, we employ different estimators to fit the Pareto distribution. The choice of estimation technique is shown to have a surprisingly profound effect on the resulting empirical powers. This prompted research into a comparison between various parameter estimation techniques. The Lomax distribution is a popular variant of the Pareto class and it is frequently fit to observed financial data. We compare the performance of various classical estimation techniques to techniques based on the minimisation of some distance measure. Interestingly, these minimum distance estimators outperform the classical techniques in various settings.

The use of the goodness-of-fit tests developed in the thesis are illustrated using practical data. The same holds for the minimum distance estimators considered.

*Keywords:* Lomax distribution; goodness-of-fit; characterisation; conditional expectation; Pareto distribution.

# Preface

Chapter 1 provides an overview and outlines the objectives of the research. This thesis is presented in article format and comprises of three journal papers: one has been accepted for publication, one has been submitted for publication and is at the second round of revisions, and one is a draft to be submitted in the near future. The three articles are presented in Chapters 2, 3 and 4. In Chapter 5 we report the main findings along with conclusions and some directions for future research.

**First article:** *On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics*, submitted to *Statistics* and is currently in the second round of revisions.

**Second article:** *New classes of tests for the Pareto distribution based on a conditional expectation characterisation*, will be submitted to *Computational Statistics*.


**Third article:** *On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques*, was accepted for publication in the journal *Computation*.

The promoters agreed on co-authorship and gave consent for the use of these articles as part of the final thesis. The candidate was solely responsible for the initial planning and proposal of the thesis, literature searches, writing all programs used in the Monte Carlo studies, interpretation of results, as well as planning and writing of the articles and the entire thesis.

## STATEMENT BY CO-AUTHORS

Herewith is a statement of the co-authors giving permission that the articles may form part of this thesis.

*I hereby approve the articles and give my consent that these articles may be published as part of the thesis for the degree Doctorate of Philosophy in Science with Statistics of Miss T Nombebe.*



---

Prof JS Allison

28/09/2023

---

Date



---

Prof IJH Visagie

28/09/2023

---

Date



---

Prof L Santana

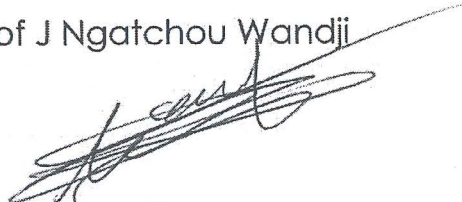
2023-09-28

---

Date

---

Prof J Ngatchou Wandji



---

Date

September 29, 2023

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview . . . . .	1
1.2	Objectives . . . . .	2
1.3	Thesis outline . . . . .	3
1.3.1	On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics . . . . .	3
1.3.2	New classes of tests for the Pareto distribution based on a conditional expectation characterisation . . . . .	3
1.3.3	On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques . . . . .	4
<b>2</b>	<b>Article 1: On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics</b>	<b>5</b>
<b>3</b>	<b>Article 2: New classes of tests for the Pareto distribution based on a conditional expectation characterisation</b>	<b>32</b>
<b>4</b>	<b>Article 3: On Fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques</b>	<b>61</b>
<b>5</b>	<b>Conclusion</b>	<b>90</b>
5.1	Overview of the goals of the papers presented . . . . .	90
5.2	Concluding remarks and future research . . . . .	91

# Chapter 1

## Introduction

### 1.1 Overview

The Pareto distribution, named after the Italian economist Vilfredo Pareto, is a probability distribution that finds application in a wide range of fields, including economics, finance, sociology, and natural sciences. It is well known for its ability to model phenomena where a small number of elements in a system accounts for a large proportion of the overall impact; this is often referred to as the “80-20” principle. Many real-world phenomena exhibit measurements with heavy-tailed behaviour and, as such, lend themselves to be modelled using the Pareto distribution. Due to its popularity, this distribution has enjoyed the attention of numerous researchers, resulting in a number of different versions of the Pareto distribution, including Types I, II, III, IV, and Generalised Pareto distributions.

However, in this PhD study we will only focus on the Type I distribution (both the one and two parameter versions) and a special case of the Type II Pareto distribution, known as the Lomax distribution. We will now briefly introduce these distributions as well as explain the relationship between them.

We say that a random variable  $X$  follows a Pareto Type I distribution with shape parameter  $\beta > 0$  if its cumulative distribution function (CDF) is

$$F_X(x) = 1 - x^{-\beta}, \quad x \geq 1,$$

with corresponding probability density function (PDF)

$$f_X(x) = \beta x^{-(\beta+1)}, \quad x \geq 1.$$

Consider the random variable  $Y = \sigma X$ , where  $\sigma > 0$  is a scale parameter. Then  $Y$  follows a general Pareto Type I distribution with CDF and PDF given by

$$F_Y(y) = 1 - \left(\frac{y}{\sigma}\right)^{-\beta}, \quad y \geq \sigma,$$

and

$$f_Y(y) = \beta\sigma^\beta y^{-(\beta+1)}, \quad y \geq \sigma,$$

respectively. Next, consider the random variable

$$Z = Y + \mu - \sigma,$$

where  $-\infty < \mu < \infty$  is a location parameter. The random variable  $Z$  follows a Pareto Type II distribution with CDF

$$F_Z(z) = 1 - \left[ 1 + \left( \frac{z - \mu}{\sigma} \right) \right]^{-\beta} \quad z \geq \mu, \quad (1.1)$$

and PDF

$$f_Z(z) = \frac{\beta}{\sigma} \left[ 1 + \left( \frac{z - \mu}{\sigma} \right) \right]^{-(\beta+1)} \quad z \geq \mu.$$

Setting  $\mu = 0$  in (1.1), we have a special case of the Pareto Type II distribution, called the Lomax distribution, which has found extended application in real world problems (see Chapter 4).

In Chapters 2 and 3 we propose and study new classes of goodness-of-fit tests for the Pareto Type I distribution. In Chapter 4 we compare various estimation techniques for the parameters of the Lomax distribution.

## 1.2 Objectives

The objectives of this thesis, which consists of three research papers, can be summarised as follows:

- To develop new classes of goodness-of-fit tests for the Pareto Type I distribution.
- To derive the asymptotic properties pertaining to some of these new classes of goodness-of-fit tests.
- To critically evaluate the performance of the newly proposed tests under different estimation methods using Monte Carlo simulation.
- To apply the newly proposed tests to real-world data settings and draw conclusions.
- To investigate different estimation methods used to estimate the shape and scale parameters of the Lomax distribution as well as determine whether the minimum distance estimators are suitable estimation alternatives for this distribution when compared to traditional methods of estimation.

### 1.3 Thesis outline

This section is a synopsis of the ‘golden thread’ running through the three research papers presented in this thesis. First note that the overall goal in this text is to investigate goodness-of-fit testing for the Pareto distribution and, tangentially, to compare various estimation methods and to determine the effect that the estimation methods have on the power of the tests.

Summaries of the three articles found in Chapters 2, 3 and 4, respectively, are given next. The relationship between these papers and the overarching theme is discussed and the connections between the papers are made clear.

#### 1.3.1 On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics

The first paper that will be presented is one in which a new class of goodness-of-fit tests based on a order statistic characterization is proposed. This paper can be considered a ‘straightforward’ goodness-of-fit article in which a novel goodness-of-fit test is proposed, its asymptotic properties derived, and then extensively studied via a finite sample simulation studies to assess its suitability and to compare its power performance relative to existing tests in the literature. The Pareto Type I distribution, using both the one and two parameter versions, were considered in this paper. An interesting consequence of this research was that it was found that the estimation method used to estimate parameters for these tests for the Pareto distribution had a notable effect on the powers of both the newly proposed tests and existing ones.

The paper’s abstract is given below.

*We propose new classes of goodness-of-fit tests for the Pareto Type I distribution. These tests are based on a characterization of the Pareto distribution involving order statistics. We derive the limiting null distribution of the tests and also show that the tests are consistent against fixed alternatives. The finite-sample performance of the newly proposed tests are evaluated and compared to some of the existing tests, where it is found that the new tests are competitive in terms of powers. The paper concludes with an application to a real world data set, namely the earnings of the 22 highest paid participants in the inaugural season of LIV golf.*

#### 1.3.2 New classes of tests for the Pareto distribution based on a conditional expectation characterisation

Following-up on the first paper, the second paper introduces classes of goodness-of-fit tests for the Pareto Type I distribution based on a conditional expectation characterisation. Once again the newly proposed tests are thoroughly investigated and the effect of a wider

variety of estimation techniques on the power of the tests is explored. Below is the abstract for this paper.

*We propose and investigate new classes of goodness-of-fit tests for the Pareto Type I distribution based on a characterisation involving a conditional expectation. A Monte Carlo power study is included in order to assess the finite sample performance of the newly developed tests for four different estimation methods. The results from the simulation study shows that the newly proposed tests are competitive in terms of power performance when compared to some existing tests. It also shows that the majority of tests produce their highest powers when the unknown shape parameter is estimated by the method of moments. A practical example, where we consider the annual salaries of English Premier League football players for two consecutive seasons, is also included to illustrate the use of the newly proposed tests.*

### **1.3.3 On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques**

Informed by the results of the previous papers, and with the aim towards future research related to the Pareto Type II distribution, this paper investigates the performance of different estimation methods when estimating the parameters of the Lomax distribution. A multitude of estimation techniques are investigated in an extensive Monte Carlo study to ascertain their suitability in the context of the Lomax distribution under a variety of settings. The abstract that accompanied the original article is given below:

*We investigate the performance of a variety of frequentist estimation techniques for the scale and shape parameters of the Lomax distribution. These methods include traditional methods such as the maximum likelihood estimator and the method of moments estimator. A version of the maximum likelihood estimator adjusted for bias is included as well. Furthermore, an alternative moment-based estimation technique, the  $L$ -moment estimator, is included, along with three different minimum distance estimators. The finite sample performances of each of these estimators are compared in an extensive Monte Carlo study. We find that no single estimator outperforms its competitors uniformly. We recommend one of the minimum distance estimators for use with smaller samples, while a bias-reduced version of maximum likelihood estimation is recommended for use with larger samples. In addition, the desirable asymptotic properties of traditional maximum likelihood estimators make them appealing for larger samples. We include a practical application demonstrating the use of the described techniques on observed data.*

In Chapter 5, we summarise the final outcomes and reach conclusions, and also highlight potential areas for further exploration.

## Chapter 2

# Article 1: On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics

The first article, *On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics*, is under review in *GSTA-2023-0063*. A summary of the guidelines to authors from the journal is now presented.

---

Manuscript	No page limit.
Title	The title should be concise and informative.
Abstract and Keywords	Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references and 4 to 6 keywords can be used for indexing purposes.
Tables	All tables are to be numbered using Arabic numerals. Tables should always be cited in text in consecutive numerical order. For each table, a table caption explaining the components of the table must be supplied.
References	The list of references should only include works that are cited in the text and that have been published or accepted for publication.
General formatting	A $\text{\LaTeX}$ template is provided for submission, Word files are also accepted.
Additional information	<a href="https://tinyurl.com/486kbupt">https://tinyurl.com/486kbupt</a>

---

# On classes of consistent tests for the Pareto distribution based on a characterization involving order statistics

Joseph Ngatchou–Wandji<sup>a</sup>, Thobeka Nombebe<sup>b</sup>, Leonard Santana<sup>b</sup>, James Allison<sup>b</sup>

<sup>a</sup>*Université de Rennes (EHESP) & Institut Élie Cartan de Lorraine, Nancy, France*

<sup>b</sup>*Pure and Applied Analytics, North–West University, Potchefstroom, South Africa*

## Abstract

We propose new classes of goodness-of-fit tests for the Pareto Type I distribution. These tests are based on a characterization of the Pareto distribution involving order statistics. We derive the limiting null distribution of the tests and also show that the tests are consistent against fixed alternatives. The finite-sample performance of the newly proposed tests are evaluated and compared to some of the existing tests, where it is found that the new tests are competitive in terms of powers. The paper concludes with an application to a real world data set, namely the earnings of the 22 highest paid participants in the inaugural season of LIV golf.

## 1 Introduction

Many real-world phenomena exhibit measurements with heavy-tailed behaviour and, as such, lend themselves to be modelled using the Pareto distribution. First developed by economist and socialist, Vilfredo Pareto (Pareto, 1897), the heavy-tailed Pareto distribution was initially used to model the unequal distribution of wealth in a population, but has found application in a number of other scenarios. Examples of situations modelled using this distribution include studies involving insurance claim premiums (Brazauskas, 2000; Beirlant et al., 2004; Mert & Saykan, 2005), studies involving medical insurance claims (Zisheng & Chi, 2006), and studies investigating gestation duration (Keiding et al., 2002, 2012), to name a few (see Arnold, 2015).

Due to its popularity, this distribution has enjoyed the attention of numerous researchers resulting in a number of different versions of the Pareto distribution, including the Types I, II, III, IV, and Generalised Pareto distributions. However, in this paper we will focus only on the use of the Type I Pareto distribution, which has cumulative distribution function (CDF) and probability density function (PDF) respectively given by

$$F_{\beta,\sigma}(x) = \begin{cases} 1 - \left(\frac{x}{\sigma}\right)^{-\beta}, & x \geq \sigma \\ 0, & x < \sigma \end{cases} \quad \text{and} \quad f_{\beta,\sigma}(x) = \begin{cases} \beta\sigma^\beta x^{-\beta-1}, & x \geq \sigma \\ 0, & x < \sigma \end{cases}$$

where  $\sigma > 0$  and  $\beta > 0$  denote respectively the scale and the shape parameters. The Type I version of the Pareto distribution with  $\sigma = 1$  has a number of practical applications as shown in a variety of old and new research works. In the earlier works, for example, Fisk (1961) and Steindl (1965) cited several examples of economic data which follow the Type I Pareto distribution, whereas Berger & Mandelbrot (1963) proposed using the Type I Pareto distribution in studies of error clusters in communication circuits. It has also been shown to be useful in applications where service times and queuing systems are modelled, as discussed in Harris (1968). More recent applications include using the Type I Pareto distribution to model the wealth distribution in the Forbes 400 list (Klass et al., 2006), and modelling the city size distribution in the United States (Ioannides & Skouras, 2013).

Hence, under these considerations, it is important to determine whether a realised data set,  $x_1, \dots, x_n$ , from a non-negative random variable  $X$  with distribution function  $F(x)$ , is well-described by a Type I Pareto distribution with parameters  $\beta$  and  $\sigma$ , denoted here by  $\mathcal{P}(\beta, \sigma)$ . In

this paper we will therefore consider using a goodness-of-fit (GOF) test to evaluate the following hypotheses regarding these data:

$H_0$  :  $X$  follows a  $\mathcal{P}(\beta, \sigma)$  distribution;  $\exists \beta, \sigma > 0$  such that  $F(x) = F_{\beta, \sigma}(x)$ ,  $x \in [\sigma, \infty)$ ,  
and

$H_1$  :  $X$  does not follow a  $\mathcal{P}(\beta, \sigma)$  distribution;  $\nexists \beta, \sigma > 0$  such that  $F(x) = F_{\beta, \sigma}(x)$ ,  $x \in [\sigma, \infty)$ .  
(1)

Several tests have been suggested to check the goodness-of-fit of Pareto distributions; the most commonly used formal goodness-of-fit tests for Pareto distributions are those based on the empirical distribution function (EDF), such as the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CvM) test, or Anderson-Darling (AD) test. These tests compare the empirical distribution of the data with the hypothesised theoretical Pareto distribution and assess the likelihood that the data were generated by a Pareto distribution. The results of these tests can be used to determine whether the Pareto distribution is a good fit for the data, or whether another distribution may be more appropriate.

Goodness-of-fit tests for the Pareto distribution have been discussed in Beirlant et al. (2006), Gulati & Shapiro (2008), Martynov (2009), and Rizzo (2009), among others. In Chu et al. (2019) a review of established tests for the Generalized Pareto, Pareto Type I and Pareto Type II distributions is provided, whereas goodness-of-fit tests based on a variety of different characterizations for the Pareto distribution can be found in Obradović et al. (2015), Obradović (2015), Volkova (2016), and Milošević & Obradović (2016b). Ndwandwe et al. (2022) provides an extensive review of the existing goodness-of-fit tests for the Pareto Type I distribution, focusing on the myriad characterisations of this distribution. In what follows we will refer to the Pareto Type I distribution as just the Pareto distribution.

In this paper we propose new classes of tests for the Pareto distribution. These tests are based on characterization of the Pareto distribution involving order statistics. In Section 2 we consider the case of the Type I Pareto with unit scale parameter. We present the characterization, introduce the new test statistics and derive the limiting null distribution of the tests and show that they are consistent against fixed alternatives. Section 3 is devoted to a discussion on the general Type I Pareto distribution. In Section 4 we compare the powers of our newly proposed tests with some existing tests (in the case of the general Type I Pareto distribution), while Section 5 illustrates the use of the tests in order to test the hypothesis that the 2022 season's earnings of LIV golfers (exceeding some known threshold), follows a Pareto distribution. The paper concludes in Section 6.

## 2 The Type I Pareto distribution with unit scale parameter

In this section, we study the case of a Type I Pareto distribution with unit scale parameter, that is,  $\mathcal{P}(\beta, 1)$ ,  $\beta > 0$ .

### 2.1 The test statistic

Consider the following characterisation of the Pareto distribution denoted here by  $\mathcal{P}(\beta, 1)$ , discussed in Allison et al. (2022).

**Characterisation.** *Let  $X_1, \dots, X_n$  be independent copies of a non-negative random variable  $X$  with common density function  $f$  and cumulative distribution function  $F$ . Let  $m$  be an integer such that  $2 \leq m \leq n$ . Then the random variables  $X_{\frac{1}{m}}$  and  $X_{(1)} = \min\{X_1, \dots, X_m\}$  have the same distribution if and only if  $F(x) = F_{\beta}(x) = F_{\beta, 1}(x)$ ,  $x \in \mathbb{R}$ ,  $\beta > 0$ .*

From this characterisation we have the following Theorem

**Theorem 1.** *Let  $X_1, \dots, X_n$  be copies of a non-negative random variable  $X$  with common density function  $f$  and cumulative distribution function  $F$ . Let  $m$  be an integer such that  $2 \leq m \leq n$ . Then the random variables  $X^{\frac{1}{m}}$  and  $\min\{X_1, \dots, X_m\}$  have the same distribution if and only if for any  $t \in \mathbb{R}$ ,*

$$\mathbb{E} \left\{ \frac{1}{m} \exp \left( -itX^{\frac{1}{m}} \right) - [1 - F(X)]^{m-1} \exp(-itX) \right\} = 0. \quad (2)$$

**Proof** - Let  $2 \leq m \leq n$ . It is well known that  $X_{(1)} = \min\{X_1, \dots, X_m\}$  has density function

$$\tilde{f}(x) = m [1 - F(x)]^{m-1} f(x), \quad x \in \mathbb{R}.$$

It is then clear that the random variables  $X^{\frac{1}{m}}$  and  $X_{(1)}$  have the same distribution if and only if they have the same characteristic functions, that is, if and only if for any  $t \in \mathbb{R}$ ,

$$\int_{\mathbb{R}} \exp \left( -itx^{\frac{1}{m}} \right) f(x) dx = m \int_{\mathbb{R}} \exp(-itx) [1 - F(x)]^{m-1} f(x) dx.$$

It is easy to see that the above equality is equivalent to

$$\int_{\mathbb{R}} \left\{ \frac{1}{m} \exp \left( -itx^{\frac{1}{m}} \right) - \exp(-itx) [1 - F(x)]^{m-1} \right\} f(x) dx = 0,$$

which can be written as

$$\mathbb{E} \left\{ \frac{1}{m} \exp \left( -itX^{\frac{1}{m}} \right) - [1 - F(X)]^{m-1} \exp(-itX) \right\} = 0.$$

□

From the characterisation and Theorem 1 we have that (2) characterises the  $\mathcal{P}(\beta, 1)$  distribution. Thus, suitable normalizations of empirical versions of the expectation in (2) can be used as basis for the construction of tests for that particular Pareto distribution. To this end, we propose the following test statistic

$$T_{m,n,w} = \int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}(t)|^2 w(t) dt, \quad (3)$$

where for all  $t \in \mathbb{R}$

$$S_{m,n,\beta}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \frac{1}{m} \exp \left( -itX_j^{\frac{1}{m}} \right) - X_j^{-\beta(m-1)} \exp(-itX_j) \right],$$

$\hat{\beta}_n = n / \sum_{j=1}^n \log(X_j)$  is the maximum likelihood estimator for  $\beta$ , and  $w(t)$  is a suitably chosen weight function satisfying some standard conditions (see Section 3).

**Proposition 1.** *Let  $2 \leq m \leq n$ . Then*

$$\int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}(t)|^2 w(t) dt = \int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}^\dagger(t)|^2 w(t) dt,$$

where for all  $t \in \mathbb{R}$ ,

$$S_{m,n,\beta}^\dagger(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \frac{1}{m} \nu \left( X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu(X_j; t) \right], \quad (4)$$

and  $\nu$  is the function defined on  $\mathbb{R} \times \mathbb{R}$  by

$$\nu(x; t) = \cos(tx) + \sin(tx). \quad (5)$$

**Proof** - Denote by  $\bar{z}$  the conjugate of any complex number  $z$ . One has the following equalities :

$$\begin{aligned} & \int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}(t)|^2 w(t) dt = \int_{\mathbb{R}} S_{m,n,\hat{\beta}_n}(t) \overline{S_{m,n,\hat{\beta}_n}(t)} w(t) dt \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \int_{\mathbb{R}} \left\{ \frac{1}{m^2} \exp \left[ -it \left( X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}} \right) \right] - \frac{1}{m} X_k^{-\hat{\beta}_n(m-1)} \exp \left[ -it \left( X_j^{\frac{1}{m}} - X_k \right) \right] \right. \\ & \quad \left. - \frac{1}{m} X_j^{-\hat{\beta}_n(m-1)} \exp \left[ -it \left( X_k^{\frac{1}{m}} - X_j \right) \right] + (X_j X_k)^{-\hat{\beta}_n(m-1)} \exp \left[ -it (X_j - X_k) \right] \right\} w(t) dt. \end{aligned}$$

By (6), since  $x \mapsto \sin(x)$  is an odd function, one has that

$$\begin{aligned} & \int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}(t)|^2 w(t) dt \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \int_{\mathbb{R}} \left\{ \frac{1}{m^2} \cos \left[ t \left( X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}} \right) \right] - \frac{1}{m} X_k^{-\hat{\beta}_n(m-1)} \cos \left[ t \left( X_j^{\frac{1}{m}} - X_k \right) \right] \right. \\ & \quad \left. - \frac{1}{m} X_j^{-\hat{\beta}_n(m-1)} \cos \left[ t \left( X_k^{\frac{1}{m}} - X_j \right) \right] + (X_j X_k)^{-\hat{\beta}_n(m-1)} \cos \left[ t (X_j - X_k) \right] \right\} w(t) dt. \end{aligned}$$

Using the identity  $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$  and the fact that the function  $x \mapsto \cos(x)\sin(x)$  is odd, one finally has :

$$\int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}(t)|^2 w(t) dt = \int_{\mathbb{R}} |S_{m,n,\hat{\beta}_n}^\dagger(t)|^2 w(t) dt.$$

□

For practical applications (and for the Monte-Carlo study in Section 4) we will choose  $w(t) = e^{-a|t|}$  and  $w(t) = e^{-at^2}$ , the choices of which lead to the following calculable forms of the test statistic:

$$\begin{aligned} T_{n,m,a}^{(1)} &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \left[ \frac{1}{m^2} \frac{2a}{a^2 + (X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}})^2} - \frac{1}{m} X_k^{-\hat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_j^{\frac{1}{m}} - X_k)^2} \right. \\ & \quad \left. - \frac{1}{m} X_j^{-\hat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_k^{\frac{1}{m}} - X_j)^2} + X_j^{-\hat{\beta}_n(m-1)} X_k^{-\hat{\beta}_n(m-1)} \frac{2a}{a^2 + (X_j - X_k)^2} \right]. \end{aligned}$$

and

$$\begin{aligned} T_{n,m,a}^{(2)} &= \frac{1}{n} \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n \left[ \frac{1}{m^2} \exp \left( -\frac{(X_j^{\frac{1}{m}} - X_k^{\frac{1}{m}})^2}{4a} \right) - \frac{1}{m} X_k^{-\hat{\beta}_n(m-1)} \exp \left( -\frac{(X_j^{\frac{1}{m}} - X_k)^2}{4a} \right) \right. \\ & \quad \left. - \frac{1}{m} X_j^{-\hat{\beta}_n(m-1)} \exp \left( -\frac{(X_k^{\frac{1}{m}} - X_j)^2}{4a} \right) + X_j^{-\hat{\beta}_n(m-1)} X_k^{-\hat{\beta}_n(m-1)} \exp \left( -\frac{(X_j - X_k)^2}{4a} \right) \right] \end{aligned}$$

respectively.

## 2.2 Large sample properties

In this section we study the asymptotic properties of the newly proposed tests under the null hypothesis as well as under fixed alternatives. It is well known that under  $H_0$ , the maximum

likelihood estimator of  $\beta$  is  $\hat{\beta}_n = n / \sum_{i=1}^n \log(X_i)$  and  $\mathbb{E}(\hat{\beta}_n) = n\beta / (n-1)$ . From this, under  $H_0$ , one can deduce the following equalities :

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta) &= \frac{n\beta}{\sum_{j=1}^n \log(X_j)} \left\{ \sqrt{n} \left[ \frac{1}{\beta} - \frac{\sum_{j=1}^n \log(X_j)}{n} \right] \right\} \\ \sqrt{n}(\hat{\beta}_n - \beta) &= \frac{\beta^2}{\sqrt{n}} \sum_{j=1}^n \left[ \frac{1}{\beta} - \log(X_j) \right] + o_P(1), \end{aligned}$$

where  $o_P(1)$  is a sequence of random variables converging in probability to 0.

Recalling that under  $H_0$ , for any  $j = 1, \dots, n$ ,  $\log(X_j)$  follows a gamma distribution with parameters 1 and  $\beta$ ,  $\sum_{j=1}^n \log(X_j)$  follows a gamma distribution with parameters  $n$  and  $\beta$  and  $1 / \sum_{j=1}^n \log(X_j)$  follows an inverse gamma distribution with parameters  $n$  and  $\beta$ , it is easy to see that the last equality results from the Central Limit Theorem (CLT) and the Law of Large Numbers (LLN) applied to the random variables  $\log(X_j)$ .

Let  $w(\cdot)$  be any continuous function satisfying

$$w(t) > 0, \lim_{t \rightarrow \pm\infty} w(t) = 0, t \in \mathbb{R}, \quad 0 < \int_{\mathbb{R}} w(t) dt < \infty \quad \text{and} \quad \int_{\mathbb{R}} \zeta(tx) w(x) dx = 0, t \in \mathbb{R}, \quad (6)$$

for any real-valued odd function  $\zeta$ . We recall that an odd real-valued function  $\zeta$  has the property  $\zeta(x) = -\zeta(-x)$ ,  $x \in \mathbb{R}$ .

Denote by  $\mathcal{C} = C(\mathbb{R}, \mathbb{R})$ , the set of  $\mathbb{R}$ -valued continuous functions defined on  $\mathbb{R}$ . Define on  $\mathcal{C}$  the metric

$$\rho(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{\rho_j(x, y)}{1 + \rho_j(x, y)}, \quad \forall j \geq 1, \quad \rho_j(x, y) = \sup_{\|w\| \leq j} |x(w) - y(w)|.$$

It is well known (see, for example, Bilodeau & Lafaye de Micheaux, 2005) that endowed with  $\rho$ ,  $\mathcal{C}$  is a separable Fréchet space, and that convergence in this metric corresponds to the uniform convergence on all compact sets. That is for all  $x, y \in \mathcal{C}$ ,  $\rho(x, y) = 0 \iff \forall j \geq 1, \rho_j(x, y) = 0$ .

For random elements  $x_n$  and  $y_n$  of  $\mathcal{C}$ ,  $\rho(x_n, y_n) \xrightarrow{P} 0 \iff \forall j \geq 1, \rho_j(x_n, y_n) \xrightarrow{P} 0$ .

**Proposition 2.** *Let  $2 \leq m \leq n$ . Under  $H_0$ , in  $\mathcal{C}$ , as  $n$  tends to infinity, in probability,*

$$S_{m, n, \hat{\beta}_n}^\dagger(\cdot) = \tilde{S}_{m, n, \beta}(\cdot) + o_P(1),$$

and for all  $t \in \mathbb{R}$ ,

$$\tilde{S}_{m, n, \beta}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \frac{1}{m} \nu \left( X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu(X_j; t) + \left[ \frac{1}{\beta} - \log(X_j) \right] \varphi(t) \right\}, \quad (7)$$

with  $\varphi$  standing for the function defined for any  $t \in \mathbb{R}$  by :

$$\varphi(t) = (m-1)\beta^4 \int_1^\infty x^{-\beta m - 2} \nu(x; t) dx. \quad (8)$$

**Proof** - Write for all  $t \in \mathbb{R}$ ,

$$S_{m,n,\widehat{\beta}_n}(t) = S_{m,n,\beta}(t) + \widehat{S}_{m,n}(t),$$

where

$$\widehat{S}_{m,n}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)} \right) \nu(X_j; t).$$

Now, from a first-order Taylor expansion, one has

$$X_j^{-\beta(m-1)} - X_j^{-\widehat{\beta}_n(m-1)} = (\widehat{\beta}_n - \beta)(m-1)\beta X_j^{-\beta(m-1)-1} + o_P(1).$$

Then, under  $H_0$ , for all  $t \in \mathbb{R}$ , one has the equalities :

$$\begin{aligned} \widehat{S}_{m,n}(t) &= (\widehat{\beta}_n - \beta)(m-1)\beta \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j^{-\beta(m-1)-1} \nu(X_j; t) + o_P(1), \\ &= \sqrt{n}(\widehat{\beta}_n - \beta) \left[ \frac{(m-1)\beta}{n} \sum_{j=1}^n X_j^{-\beta(m-1)-1} \nu(X_j; t) \right] + o_P(1). \end{aligned}$$

For all  $t \in \mathbb{R}$ , define the term in the brackets as  $\varphi_n(t)$ . By the law of large numbers, it is easy to see that  $\varphi_n(t)$  converges point-wise to  $\varphi(t)$  and that it is equicontinuous on every compact subset of  $\mathbb{R}$ . Therefore, it converges uniformly to  $\varphi(t)$  on any compact subset  $\Theta$  of  $\mathbb{R}$ . This result could also be obtained by applying Proposition 1 of Csörgö (1983).

As a consequence of the above convergence, uniformly in  $t \in \Theta$ ,

$$\widehat{S}_{m,n}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \frac{1}{\beta} - \log(X_j) \right] \varphi(t) + o_P(1)$$

and uniformly in  $t \in \Theta$ ,

$$\begin{aligned} S_{m,n,\widehat{\beta}_n}(t) &= S_{m,n,\beta}(t) + \widehat{S}_{m,n}(t) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \frac{1}{m} \nu \left( X_j^{\frac{1}{m}}; t \right) - X_j^{-\beta(m-1)} \nu(X_j; t) + \left[ \frac{1}{\beta} - \log(X_j) \right] \varphi(t) \right\} + o_P(1) \\ &= S_{m,n,\widehat{\beta}_n}^\dagger + o_P(1). \end{aligned}$$

As this holds for arbitrary compact  $\Theta$ , one can conclude that as  $n$  tends to infinity, in probability,

$$\rho(S_{m,n,\widehat{\beta}_n}, S_{m,n,\widehat{\beta}_n}^\dagger) \longrightarrow 0,$$

which establishes the proposition. □

Now, let  $k$  be the real-valued function defined for any  $(x, t) \in \mathbb{R} \times \mathbb{R}$  by

$$k(x, t) = \frac{1}{m} \nu \left( x^{\frac{1}{m}}; t \right) - x^{-\beta(m-1)} \nu(x; t) + \left[ \frac{1}{\beta} - \log(x) \right] \varphi(t).$$

Also consider the function  $K$  defined for any  $t \in \mathbb{R}$  by

$$K(t) = \int_{\mathbb{R}} k(x, t) dF(x),$$

and  $F_n$  the empirical cumulative distribution function of  $X_1, X_2, \dots, X_n$ . Then one sees that

$$\varpi_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n [k(X_j, t) - K(t)] = \int_{\mathbb{R}} k(x, t) d\{\sqrt{n}[F_n(x, t) - F(t)]\}.$$

Under suitable conditions (see Csörgö, 1983),  $\varpi_n$  converges weakly to a zero-mean Gaussian process with covariance kernel

$$\mathbb{E}[\varpi_n(t)\varpi_n(s)] = \int_{\mathbb{R}} k(x, t)k(x, s) dF(x) - K(t)K(s).$$

**Theorem 2.** *Let  $m \in \{2, \dots, n\}$ , fixed. If  $\beta > 1/m$ , then under  $H_0$ ,  $\tilde{S}_{m,n,\beta}(\cdot)$  converges weakly in  $\mathcal{C}$  to a zero-mean Gaussian process  $S_m(\cdot)$  with covariance kernel  $\Gamma_m$  defined for any  $s, t \in \mathbb{R}$  by*

$$\begin{aligned} \Gamma_m(s, t) &= \beta \int_1^\infty \left\{ \frac{1}{m^2} \nu\left(x^{\frac{1}{m}}; t\right) \nu\left(x^{\frac{1}{m}}; s\right) + x^{-2\beta(m-1)} \nu(x; t) \nu(x; s) \right. \\ &\quad + \left(\frac{1}{\beta} - \log(x)\right)^2 \varphi(t)\varphi(s) - \frac{1}{m} x^{-\beta(m-1)} \nu(t; s) \nu\left(x^{\frac{1}{m}}; s\right) \\ &\quad + \frac{1}{m} \left(\frac{1}{\beta} - \log(x)\right)^2 \nu\left(x^{\frac{1}{m}}; t\right) \varphi(s) - \frac{1}{m} x^{-\beta(m-1)} \nu(x; t) \nu\left(x^{\frac{1}{m}}; s\right) \\ &\quad - x^{-\beta(m-1)} \left(\frac{1}{\beta} - \log(x)\right) \nu(x; t) \varphi(s) + \frac{1}{m} \left(\frac{1}{\beta} - \log(x)\right) \varphi(t) \nu\left(x^{\frac{1}{m}}; s\right) \\ &\quad \left. - x^{-\beta(m-1)} \left(\frac{1}{\beta} - \log(x)\right) \varphi(t) \nu(x; s) \right\} x^{-\beta-1} dx. \end{aligned} \quad (9)$$

**Proof** - We prove this result by showing that  $\tilde{S}_{m,n,\beta}(\cdot)$  is tight and its finite-dimensional distributions converge to those of any zero-mean Gaussian process with covariance kernel  $\Gamma_m$ . For this, we check the conditions (i), (i)\* and (ii)\* of Csörgö (1983).

As  $k(x, t)$  is bounded with respect to  $t$  on any compact subset of  $\mathbb{R}$ , so is  $|k(x, t)|^{2+\delta}$ , for any  $\delta > 0$ . Thus, one can find  $t_0 \in \Theta$  such that for any  $x \geq 1$

$$\sup_{t \in \Theta} |k(x, t)|^{2+\delta} = |k(x, t_0)|^{2+\delta}.$$

Consequently,

$$\begin{aligned} \int_{\mathbb{R}} \sup_{t \in \Theta} |k(x, t)|^{2+\delta} dF(x) &= \int_{\mathbb{R}} |k(x, t_0)|^{2+\delta} dF(x) \\ &\leq Cst \int_{\mathbb{R}} \left[ 1 + x^{-(2+\delta)\beta(m-1)} + \left(\frac{1}{\beta} + \log(x)\right)^{2+\delta} \right] dF(x) \\ &< \infty. \end{aligned}$$

This establishes the convergence of the finite-dimensional distributions of  $\tilde{S}_{m,n,\beta}$  and the point (i)\* of Csörgö (1983).

It remains to show (ii)\*. One can write :

$$|k(x, t) - k(x, s)| \leq \frac{1}{m} \left| \nu \left( x^{\frac{1}{m}}; t \right) - \nu \left( x^{\frac{1}{m}}; s \right) \right| + x^{-\beta(m-1)} |\nu(x; t) - \nu(x; s)| \\ + \left| \frac{1}{\beta} - \log(x) \right| |\varphi(t) - \varphi(s)|.$$

By a first-order Taylor expansion of the function  $t \mapsto \nu(x; t) = \cos(xt) + \sin(xt)$ , one has :

$$|k(x, t) - k(x, s)| \leq \frac{x^{\frac{1}{m}}}{m} |t - s| + x^{-\beta(m-1)+1} |t - s| + Cst \left| \frac{1}{\beta} - \log(x) \right| |t - s|.$$

Then, one easily sees that for any  $s, t \in \Theta$ , one can find  $\alpha \in (0, 1]$  such that

$$|k(x, t) - k(x, s)| \leq |t - s|^\alpha Cst \left( \frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left| \frac{1}{\beta} - \log(x) \right| \right) \\ = |t - s|^\alpha M(x, v(t, s)),$$

where  $v$  is any  $\Theta$ -valued function defined on  $\Theta \times \Theta$ , and for any  $x \geq 1$  and  $t \in \Theta$ ,  $M$  stands for the function defined as

$$M(x, t) = Cst \left( \frac{x^{\frac{1}{m}}}{m} + x^{-\beta(m-1)+1} + \left| \frac{1}{\beta} - \log(x) \right| \right).$$

Since  $M(x, t)$  does not depend on  $t$ ,  $\sup_{t \in \Theta} M^2(x; t) = M^2(x; t)$  and by our assumptions

$$\int_{\mathbb{R}} \sup_{t \in \Theta} M^2(x; t) dF(x) = \int_{\mathbb{R}} M^2(x; t) dF(x) < \infty.$$

From all these, by Csörgö (1983), one can conclude that

$$\int_{\mathbb{R}} k(x, t) d\{\sqrt{n}[F_n(x, t) - F(t)]\}$$

converges weakly to any zero-mean Gaussian process with covariance kernel

$$\int_{\mathbb{R}} k(x, t)k(x, s) dF(x) - K(s)K(t).$$

From the following equality

$$\int_{\mathbb{R}} k(x, t) d\{\sqrt{n}[F_n(x, t) - F(t)]\} = \tilde{S}_{m,n,\beta}(t) - \sqrt{n}K(t), \quad (10)$$

since under  $H_0$ ,  $K(t) = 0$ ,  $\tilde{S}_{m,n,\beta}(\cdot)$  converges weakly to the Gaussian process invoked in the theorem. This establishes the theorem.  $\square$

**Theorem 3.** Let  $m \in \{2, \dots, n\}$ , fixed. Assume that under  $H_1$ ,  $\mathbb{E}[\log(X_1)] < \infty$ . Then under  $H_1$ , in probability, for any  $t \in \mathbb{R}$ ,  $S_{m,n,\hat{\beta}_n}(t)$  has the same asymptotic behavior as  $\sqrt{n}Q(t)$ , where

$$Q(t) = \mathbb{E} \left[ \frac{1}{m} \exp \left( -itX_1^{\frac{1}{m}} \right) - [1 - F(X_1)]^{(m-1)} \exp(-itX_1) \right] \\ + \mathbb{E} \left\{ \left[ X_1^{-\frac{(m-1)}{\mathbb{E}[\log(X_1)]}} - [1 - F(X_1)]^{(m-1)} \right] \exp(-itX_1) \right\}, t \in \mathbb{R}.$$

**Proof** - Define, for any  $t \in \mathbb{R}$ ,

$$\dot{S}_{m,n,F} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \frac{1}{m} \exp \left( -itX_j^{\frac{1}{m}} \right) - [1 - F(X_j)]^{(m-1)} \exp(-itX_j) \right].$$

Now, adding and subtracting one has

$$\frac{S_{m,n,\hat{\beta}_n}(t)}{\sqrt{n}} = \frac{\dot{S}_{m,n,F}(t)}{\sqrt{n}} + \frac{1}{n} \sum_{j=1}^n \left\{ X_j^{-\hat{\beta}(m-1)} - [1 - F(X_j)]^{(m-1)} \right\} \exp(-itX_j) + o_P(1).$$

Then by the LLN the first and second terms in the right-hand side of the above equation converge point-wise respectively to

$$Q_1(t) = \mathbb{E} \left[ \frac{1}{m} \exp \left( -itX_1^{\frac{1}{m}} \right) - [1 - F(X_1)]^{(m-1)} \exp(-itX_1) \right]$$

and

$$Q_2(t) = \mathbb{E} \left\{ \left[ X_1^{-\frac{(m-1)}{\mathbb{E}[\log(X_1)]}} - [1 - F(X_1)]^{(m-1)} \right] \exp(-itX_1) \right\}.$$

This establishes the theorem. □

**Theorem 4.** Let  $w$  be any function satisfying (6). Let  $m \in \{2, \dots, n\}$ , fixed.

(i)- If  $\beta > 1/m$ , then, under  $H_0$ , as  $n$  tends to infinity, in distribution

$$T_{m,n,w} \longrightarrow \int_{\mathbb{R}} S_m^2(t)w(t)dt,$$

where  $S_m$  is the Gaussian process invoked in Theorem 2.

(ii)- Under  $H_1$ , if  $\mathbb{E}[\log(X_1)] < \infty$ , then as  $n$  tends to infinity,

$$T_{m,n,w} \longrightarrow \infty.$$

**Proof** - For Part (i), first observe that since the  $X_i$ 's are iid, under  $H_0$ , one has by simple computations

$$\mathbb{E} \left[ \tilde{S}_{m,n,\beta}^2(t) \right] = \mathbb{E} \left\{ \frac{1}{m} \nu \left( X_1^{\frac{1}{m}}; t \right) - X_1^{-\beta(m-1)} \nu(X_1; t) + \left[ \frac{1}{\beta} - \log(X_1) \right] \varphi(t) \right\}^2.$$

Denote by  $B_r \subset \mathbb{R}$  a ball of radius  $r$ , and  $\overline{B}_r$  its complementary in  $\mathbb{R}$ . Integrating both sides of the above equality with respect to  $w(t)dt$  on  $\overline{B}_r$ , one has :

$$\begin{aligned} & \int_{\overline{B}_r} \mathbb{E} \left[ \tilde{S}_{m,n,\beta}^2(t) \right] w(t)dt \\ &= \int_{\overline{B}_r} \mathbb{E} \left\{ \frac{1}{m} \nu \left( X_1^{\frac{1}{m}}; t \right) - X_1^{-\beta(m-1)} \nu(X_1; t) + \left[ \frac{1}{\beta} - \log(X_1) \right] \varphi(t) \right\}^2 w(t)dt. \end{aligned}$$

Since the functions  $t \mapsto \nu(x, t)$  and  $t \mapsto \varphi(t)$  are bounded, since  $w(t) \rightarrow 0$  as  $t$  tends to infinity, it is easy to see that as  $r$  tends to infinity, the right-hand side of the last equality converges to 0. From an adaption of Theorem 2.3 of Bilodeau & Lafaye de Micheaux (2005) with  $f(x) = x^2$  and  $\alpha = 1$ , one has that, under  $H_0$ , as  $n$  tends to infinity,

$$T_{m,n,w} \rightarrow \int_{\mathbb{R}} S_m^2(t)w(t)dt.$$

For the proof of the second part, it follows easily from Theorem 3 that for larger values of  $n$ ,

$$T_{m,n,w} = n \int_{\mathbb{R}} |Q(t)|^2 w(t)dt.$$

Under  $H_1$ , there exists a  $t_1 \in \mathbb{R}$ , such that  $Q_1(t_1) \neq 0$ . The quantity  $mQ_2(t)$  is the difference between the characteristic function of  $X_{(1)} = \min\{X_1, \dots, X_m\}$  under  $H_0$  ( $F = F_{1/E[\log(X_1)]}$ ) and its characteristic function under  $H_1$ . Since  $F \neq F_{1/E[\log(X_1)]}$  under  $H_1$ ,  $Q_2(t_2) \neq 0$  for some  $t_2 \in \mathbb{R}$ . This means that under  $H_1$ , there is some  $t_0 \in \mathbb{R}$  for which  $Q(t_0) \neq 0$ . By the continuity of  $|Q|$ ,  $\int_{\mathbb{R}} |Q(t)|^2 w(t)dt > 0$ . Whence, under  $H_1$ , as  $n$  tends to infinity, in probability,

$$T_{m,n,w} \rightarrow \infty.$$

□

Now, assume that  $w(\cdot)$  is the density function (with respect to the Lebesgue's measure) of some positive measure  $\mu$  with support  $\mathbb{R}$ . Let  $\mathcal{L}_2 = L_2(\mu)$  be the collection of functions  $g$  defined on  $\mathbb{R}$  such that  $\int_{\mathbb{R}} g^2(t)d\mu(t) < \infty$ . For  $h_1, h_2, h \in \mathcal{L}_2$ ,  $\langle h_1, h_2 \rangle = \int_{\mathbb{R}} h_1(t)h_2(t)d\mu(t)$  and  $\|h\|_{\mathcal{L}_2} = \langle h, h \rangle^{\frac{1}{2}}$  respectively stand for the usual inner product and norm on  $\mathcal{L}_2$ .

From our assumptions, it is easy to prove that the function  $\Gamma_m(s, t)$  defined by (9) is a positive semidefinite kernel. Consequently, the integral operator  $\nabla_{\Gamma_m}$  defined on  $\mathcal{L}_2$  by

$$\nabla_{\Gamma_m} h(t) = \int_{\mathbb{R}} \Gamma_m(s, t)h(s)d\mu(s), \quad t \in \mathbb{R}. \quad (11)$$

admits eigenvalues  $\xi_1, \xi_2, \dots$  sorted so that  $\xi_1 \geq \xi_2 \geq \dots \geq 0$ , and eigenfunctions  $g_1, g_2, \dots$  which form an orthonormal basis for  $\mathcal{L}_2$ .

**Corollary 1.** *Under the conditions of Theorem 4, under  $H_0$ ,  $T_{m,n,w}$  has asymptotically the same distribution as  $\sum_{j \geq 1} \xi_j \chi_j^2$ , where  $\xi_j$  and  $\chi_j^2$ ,  $j \geq 1$ , are respectively the eigenvalues of  $\nabla_{\Gamma_m}$  and i.i.d. random variables following a chi-squared distribution with one degree of freedom.*

**Proof** - The Gaussian process  $S_m(\cdot)$  defined in Theorem 2 can be viewed as a random element of  $\mathcal{L}_2$ . Its Karhunen-Loève representation is given by

$$S_m(t) = \sum_{j=1}^{\infty} G_j f_j(t), \quad t \in \mathbb{R},$$

where for all  $j \geq 1$ ,  $G_j = \langle S_m(\cdot), f_j \rangle$  are independent zero-mean Gaussian random variables with variances  $\xi_j$ . Thus,  $\|S_m(\cdot)\|_{\mathcal{L}_2}^2 = \sum_{j=1}^{\infty} G_j^2$ . Recalling that  $E(G_j^2) = \xi_j \geq 0$ ,  $j \geq 1$ , for nil  $\xi_j$ 's, the corresponding  $G_j$ 's are nil in probability. For positive  $\xi_j$ 's, one can observe that  $Z_j = G_j/\sqrt{\xi_j}$ ,  $j \geq 1$ , are iid standard Gaussian random variables. Thus,

$$\int_{\mathbb{R}} S_m^2(t) d\mu(t) = \|S_m(\cdot)\|_{\mathcal{L}_2}^2 = \sum_{j=1}^{\infty} \xi_j Z_j^2.$$

□

One can approximate the distribution of  $\sum_{j=1}^{\infty} \xi_j \chi_j^2$  by that of  $\sum_{j=1}^J \xi_j \chi_j^2$  for any integer  $J$  large enough. Since the  $\xi_j$ 's are unknown, they can be estimated by the eigenvalues of the operator  $\nabla_{\widehat{\Gamma}_{m,n}}(t)$ , where  $\widehat{\Gamma}_m(t)$  is any consistent estimator of  $\Gamma_m(t)$ . A possible way is to estimate them by the  $\widehat{\xi}_j$ 's from the integral equations

$$\nabla_{\widehat{\Gamma}_{m,n}} \widehat{f}_j = \widehat{\xi}_j \widehat{f}_j, \quad j \geq 1.$$

A natural estimator of  $\Gamma_m(s, t)$  can be obtained by taking the empirical counterpart in the expression given in (9), in which  $\beta$  is replaced by  $\widehat{\beta}_n$ . Some indications on the computation of the cumulative distribution function of  $\sum_{j=1}^J \widehat{\xi}_j \chi_j^2$  can be found in Ngatchou-Wandji (2009) or in Fan et al. (2017). We will not pursue this further in this paper.

### 3 The case of the general Type I Pareto distribution

In this section, we indicate how to treat the case where the observations follow a more general  $\mathcal{P}(\beta, \sigma)$  distribution. That is, we consider testing the following more general hypotheses :

$$H_0 : \exists \beta, \sigma > 0 \text{ such that } F(x) = F_{\beta, \sigma}(x), \quad x \in \mathbb{R}$$

and

$$H_1 : \nexists \beta, \sigma > 0 \text{ such that } F(x) = F_{\beta, \sigma}(x), \quad x \in \mathbb{R},$$

where we recall that

$$F_{\beta, \sigma}(x) = \begin{cases} 1 - \left(\frac{x}{\sigma}\right)^{-\beta}, & x \geq \sigma \\ 0, & x < \sigma \end{cases}$$

**Remark 1.** *The above testing problem can be related to that of the preceding sections by using the fact that a non-negative random variable  $X$  follows a  $\mathcal{P}(\beta, \sigma)$  distribution if and only if the scaled random variable  $X/\sigma$  follows a  $\mathcal{P}(\beta, 1)$  distribution. This can be seen easily by observing that :*

- if  $X \sim \mathcal{P}(\beta, \sigma)$  then for any  $x \geq 1$ ,

$$\mathbb{P}\left(\frac{X}{\sigma} \leq x\right) = \mathbb{P}(X \leq \sigma x) = 1 - x^{-\beta}$$

- if  $X/\sigma \sim \mathcal{P}(\beta, 1)$  then for any  $x \geq \sigma$ ,

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X}{\sigma} \leq \frac{x}{\sigma}\right) = 1 - \left(\frac{x}{\sigma}\right)^{-\beta}.$$

Now, let  $Y_1, \dots, Y_n$  be an independent and identically distributed sample following a  $\mathcal{P}(\beta, \sigma)$  distribution. By Remark 1, the current testing problem can be done as the one studied in Section 2, by considering the scaled observations  $X_j = Y_j/\sigma$ ,  $j = 1, \dots, n$ . In the case  $\sigma$  is unknown, one would rather consider  $X_j = Y_j/\hat{\sigma}_n$ ,  $j = 1, \dots, n$ , where  $\hat{\sigma}_n$  is a consistent estimator of  $\sigma$  (the maximum likelihood estimator of  $\sigma$  is  $\hat{\sigma}_n = \min\{Y_1, \dots, Y_n\}$ ). The last argument is motivated by the fact that for larger values of  $n$ , the  $Y_j/\hat{\sigma}_n$ 's are approximately the  $Y_j/\sigma$ 's which are iid with a  $\mathcal{P}(\beta, 1)$  distribution.

Then, the corresponding test statistic would be computed based on the  $X_j = Y_j/\hat{\sigma}_n$ 's while its asymptotic behavior would be studied with the  $X_j = Y_j/\sigma$ 's. With this, Theorem 2 becomes

**Theorem 5.** *Let  $m \in \{2, \dots, n\}$ , fixed. If  $\beta > 1/m$ , then under  $H_0$ ,  $\tilde{S}_{m,n,\beta}(\cdot)$  converges weakly in  $\mathcal{C}$  to a zero-mean Gaussian process  $S_m(\cdot)$  with covariance kernel  $\Gamma_m$  defined for any  $s, t \in \mathbb{R}$  by*

$$\begin{aligned} \Gamma_m(s, t) &= \beta \int_1^\infty \left\{ \frac{1}{m^2} \nu \left[ \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; t \right] \nu \left( \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; s \right) + \left( \frac{x}{\sigma} \right)^{-2\beta(m-1)} \nu \left( \frac{x}{\sigma} ; t \right) \nu \left( \frac{x}{\sigma} ; s \right) \right. \\ &+ \left[ \frac{1}{\beta} - \log \left( \frac{x}{\sigma} \right) \right]^2 \varphi(t) \varphi(s) - \frac{1}{m} \left( \frac{x}{\sigma} \right)^{-\beta(m-1)} \nu(t; s) \nu \left[ \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; s \right] \\ &+ \frac{1}{m} \left\{ \frac{1}{\beta} - \log \left[ \left( \frac{x}{\sigma} \right) \right] \right\}^2 \nu \left[ \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; t \right] \varphi(s) - \frac{1}{m} \left( \frac{x}{\sigma} \right)^{-\beta(m-1)} \nu \left[ \left( \frac{x}{\sigma} \right) ; t \right] \nu \left[ \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; s \right] \\ &- \left( \frac{x}{\sigma} \right)^{-\beta(m-1)} \left\{ \frac{1}{\beta} - \log \left[ \left( \frac{x}{\sigma} \right) \right] \right\} \nu \left[ \left( \frac{x}{\sigma} \right) ; t \right] \varphi(s) + \frac{1}{m} \left\{ \frac{1}{\beta} - \log \left[ \left( \frac{x}{\sigma} \right) \right] \right\} \varphi(t) \nu \left[ \left( \frac{x}{\sigma} \right)^{\frac{1}{m}} ; s \right] \\ &\left. - \left( \frac{x}{\sigma} \right)^{-\beta(m-1)} \left\{ \frac{1}{\beta} - \log \left[ \left( \frac{x}{\sigma} \right) \right] \right\} \varphi(t) \nu \left[ \left( \frac{x}{\sigma} \right) ; s \right] \right\} \left( \frac{x}{\sigma} \right)^{-\beta-1} dx. \end{aligned} \quad (12)$$

**Proof** - The proof can be handled in the same lines as the proof of Theorem 2 and is left to the reader.  $\square$

The null distribution of the test statistic can be computed along the same lines as at the end of Section 2. One can use a natural estimator of  $\Gamma_m(s, t)$  obtained by taking the empirical counterpart in the expression given in (12), where  $\beta$  and  $\sigma$  are replaced by consistent estimators  $\hat{\beta}_n$  and  $\hat{\sigma}_n$ .

Also, the consistency of the test can be handled by establishing a result similar to Theorem 3 and another similar to the second part of Theorem 4.

## 4 Monte Carlo simulation study and results

This section contains the results of a Monte Carlo study where the finite-sample performance of the newly proposed tests  $T_{n,m,a}^{(1)}$  and  $T_{n,m,a}^{(2)}$  are compared to the following existing tests for the hypothesis in (1); i.e., the general Type I Pareto distribution:

- The traditional Kolmogorov-Smirnov ( $KS_n$ ) and Cramer-von Mises ( $CV_n$ ) tests.
- Two tests proposed by Zhang (2002) based on the likelihood ratio, with test statistics given by

$$ZA_n = - \sum_{j=1}^n \left[ \frac{\log \left\{ 1 - X_{(j)}^{-\hat{\beta}_n} \right\}}{n - j + \frac{1}{2}} + \frac{\log \left\{ X_{(j)}^{-\hat{\beta}_n} \right\}}{j - \frac{1}{2}} \right]$$

and

$$ZB_n = \sum_{j=1}^n \left[ \log \left( \frac{\left( 1 - X_{(j)}^{-\hat{\beta}_n} \right)^{-1} - 1}{\frac{\binom{n-\frac{1}{2}}{j-\frac{3}{4}}}{(j-\frac{3}{4})-1}} \right) \right]^2,$$

where  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$  denote the order statistics of  $X_1, X_2, \dots, X_n$ .

- A test based on entropy utilizing the Kullback-Leibler divergence measure (see, e.g., Ahrari et al. (2002)). The test statistic is given by

$$KL_{n,m} = -H_{n,m} - \log(\hat{\beta}_n) + (\hat{\beta}_n + 1) \frac{1}{n} \sum_{j=1}^n \log(X_j),$$

where

$$H_{n,m} = \frac{1}{n} \sum_{j=1}^n \log \left\{ \binom{n}{2m} (X_{(j+m)} - X_{(j-m)}) \right\}$$

is an estimator for the entropy, with  $X_{(j)} = X_{(1)}$  for  $j < 1$ ,  $X_{(j)} = X_{(n)}$  for  $j > n$ , and  $m$  is a window width subject to  $m \leq \frac{n}{2}$ .

We implement the test for  $m = 1$  and  $m = 10$ .

- A test based on the empirical characteristic function proposed by Meintanis (2009b). The test is a weighted  $L2$  distance between the empirical characteristic function of transformed data and the characteristic function of the standard uniform distribution. Based on the transformation  $\hat{U}_j = F_{\hat{\beta}_n}(X_j)$ ,  $j = 1, \dots, n$ , the test statistic is

$$M_{n,a} = \frac{1}{n} \sum_{j,k=1}^n \frac{2a}{(\hat{U}_j - \hat{U}_k)^2 + a^2} + 2n \left[ 2 \tan^{-1} \left( \frac{1}{a} \right) - a \log \left( 1 + \frac{1}{a^2} \right) \right] - 4 \sum_{j=1}^n \left[ \tan^{-1} \left( \frac{\hat{U}_j}{a} \right) + \tan^{-1} \left( \frac{1 - \hat{U}_j}{a} \right) \right].$$

The value of the tuning parameter is set to  $a = 0.5$  and  $a = 1$  in order to obtain the Monte Carlo results presented.

- A test based on the Mellin transform proposed by Meintanis (2009a). The test statistic is given by

$$G_{n,a} = \frac{1}{n} \left[ (\hat{\beta}_n + 1)^2 \sum_{j,k=1}^n I_w^{(0)}(X_j X_k) + \sum_{j,k=1}^n I_w^{(2)}(X_j X_k) + 2(\hat{\beta}_n + 1) \sum_{j,k=1}^n I_w^{(1)}(X_j X_k) \right] \\ + \hat{\beta}_n \left[ n \hat{\beta}_n I_w^{(0)}(1) - 2(\hat{\beta}_n + 1) \sum_{j=1}^n I_w^{(0)}(X_j) - 2 \sum_{j=1}^n I_w^{(1)}(X_j) \right],$$

where

$$I_w^{(m)}(t) = \int_0^\infty (t-1)^m \frac{1}{x^t} w(t) dt, \quad m = 0, 1, 2.$$

Choosing  $w(x) = e^{-ax}$ , one has

$$I_a^{(0)}(x) = (a + \log x)^{-1},$$

$$I_a^{(1)}(x) = \frac{1 - a - \log x}{(a + \log x)^2},$$

and

$$I_a^{(2)}(x) = \frac{2 - 2a + a^2 + 2(a-1) \log x + \log^2 x}{(a + \log x)^3}.$$

We present results for  $a = 0.5$  and  $a = 2$ .

- A test proposed by Allison et al. (2022). The test statistic measures the difference between the empirical distribution of  $\min\{X_1, \dots, X_m\}$  and the  $V$ -empirical distribution of  $\sqrt[m]{X}$ , defined as

$$\Delta_{n,m}(x) = \frac{1}{n} \sum_{j=1}^n I\{X_j^{\frac{1}{m}} \leq x\} - \frac{1}{n^m} \sum_{j_1, \dots, j_m=1}^n I\{\min(X_{j_1}, \dots, X_{j_m}) \leq x\}.$$

Based on  $\Delta_{n,m}$ , the authors propose the following test statistic

$$I_{n,m} = \int_1^\infty \Delta_{n,m}(x) dF_n(x).$$

We show results for  $m = 2$ .

## 4.1 Simulation settings

A Monte Carlo study is carried out to examine the empirical power performance of the tests discussed in the previous sections against various fixed alternative distributions; this includes those listed in Table 1 along with two Pareto mixture distributions. Note that the alternatives listed in Table 1 natively have support  $(0, \infty)$  and so we were required to *shift* these distributions by  $\sigma = 1$  unit to ensure that the simulated data has the same support as the Pareto distribution.

The first of the two mixture distributions that we use in this study places mixture probability  $1 - p$  on the Pareto distribution with parameter  $\theta = 3$  ( $\mathcal{P}(3)$ ) and probability  $p$  on the lognormal distribution with parameters  $\mu = -2.69$  and  $\theta = 2$  ( $LN(-2.69, 2)$ ). The second family of mixture distributions similarly places probability  $1 - p$  on the  $\mathcal{P}(3)$  distribution and probability  $p$  on the

Table 1: Summary of various choices of the alternative distributions.

Alternative	Density function	Notation
Gamma ( $\theta$ )	$\frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}e^{-(x-1)}$	$\Gamma(\theta)$
Weibull( $\lambda, \theta$ )	$\frac{\theta}{\lambda} \left(\frac{x-1}{\lambda}\right)^{\theta-1} \exp\left\{-\left(\frac{x-1}{\lambda}\right)^\theta\right\}$	$W(\lambda, \theta)$
Log-normal( $\mu, \theta$ )	$\exp\left\{-\frac{1}{2}\left(\frac{\log(x-1)-\mu}{\theta}\right)^2\right\} / \{\theta(x-1)\sqrt{2\pi}\}$	$LN(\mu, \theta)$
Linear failure rate( $\theta$ )	$(1+\theta(x-1))\exp(-(x-1)-\theta(x-1)^2/2)$	$LF(\theta)$
Beta exponential( $\theta$ )	$\theta e^{-(x-1)}(1-e^{-(x-1)})^{\theta-1}$	$BE(\theta)$
Dhillon( $\theta$ )	$\frac{\theta+1}{x+1} \exp\left\{-\left(\log(x+1)\right)^{\theta+1}\right\} (\log(x+1))^\theta$	$DH(\theta)$
Log-Weibull( $\theta$ )	$\frac{(1+\theta)\log^\theta x}{x^{2+\theta}}$	$LW(\theta)$
Benini( $\theta$ )	$\frac{\exp^{-\theta\log^2 x}}{x^2}(1+2\theta\log x)$	$BN(\theta)$

Weibull distribution with parameter  $\lambda = 0.5$  and  $\theta = 0.25$  ( $W(0.5, 0.25)$ ). These parameter configurations are chosen to ensure that both distributions used in the mixtures share the same expected values. Random variates from many of these distributions can be obtained using, for example, the R package `Power` (de Micheaux & Tran, 2016).

The results are given in Tables 2–5 and display the estimated powers (the percentage of times the null hypothesis is rejected in  $MC = 20\,000$  independent Monte Carlo replications) calculated for sample sizes  $n = 20$  and  $n = 30$ . Note that the results from the test statistic  $T_{n,m,a}^{(1)}$  are omitted from the simulation results as they were found to be similar to  $T_{n,m,a}^{(2)}$ . The ‘warp-speed’ bootstrap method (Giacomini et al., 2013) is employed in order to simultaneously calculate the bootstrap critical values and Monte Carlo approximations of the power. In addition, all results are calculated by estimating the parameters using either maximum likelihood estimation (MLE),

$$\hat{\sigma}_{MLE} = \min(X_1, \dots, X_n) \quad \text{and} \quad \hat{\beta}_{MLE} = \frac{n}{\sum_{i=1}^n \log(X_i / \hat{\sigma}_{MLE})}$$

or method of moments estimation (MME),

$$\hat{\sigma}_{MME} = \frac{\bar{X}_n(\hat{\beta}_{MME} - 1)}{\hat{\beta}_{MME}} \quad \text{and} \quad \hat{\beta}_{MME} = \frac{n\bar{X} - \min(X_1, \dots, X_n)}{n(\bar{X}_n - \min(X_1, \dots, X_n))}.$$

The results obtained when MLE is employed are given in Tables 2 and 4, whereas the results associated with MME are given in Tables 3 and 5. A significance level of 5 % is used throughout the study and all calculations were executed using R v4.2.2 (R Core Team, 2021). Note that all of the code used in these simulations can be found here: <http://bit.ly/3EwubZf>.

The empirical power results for the two test statistics developed in this paper make use of specific configurations of the tuning parameters  $m$  and  $a$ , as these combinations were found to produce high powers in a separate exploratory preliminary simulation. Specifically, for  $T^{(2)}$ , the parameter settings used are the pairings of  $m = 2$  &  $a = 0.5$ ,  $m = 3$  &  $a = 0.5$  and  $m = 4$  &  $a = 0.5$ .

## 4.2 Simulation results

We begin by discussing the results obtained using MLE estimators for the parameters for  $n = 20$  and  $n = 30$ , respectively, and then move on to the results obtained when using MME estimation. The two highest values for each alternative distribution (rows) are highlighted to make

comparison easier. From Tables 2 to 5 it is clear that all the tests closely maintain the nominal significance level of 5%.

When considering the estimated powers under MLE estimation, the test  $KL_{n,10}$  does best for the majority of the alternatives given in Table 1, closely followed by  $T_{n,2,0.5}^{(2)}$ . However, we note that while the  $KL_{n,10}$  test has good performance among these non-mixture alternatives, it has remarkably poor performance when applied to the mixture distributions; displaying almost no power for this sequence of alternatives. When considering the two mixture alternatives, the statistic  $T_{n,4,0.5}^{(2)}$  produces some of the highest estimated powers across all mixing proportions used. In particular, it has the best performance for the Pareto-log-normal mixture alternative and is tied for the best performing test statistic (with  $ZA_n$  and  $ZB_n$ ) for the Pareto-Weibull mixture.

Turning our attention to the performance of the tests under the MME estimation scheme, we find that the  $G_{n,2}$  test statistic generally produces the best power performance among the non-mixture alternatives, followed, once again, by  $T_{n,2,0.5}^{(2)}$ . We note that, in stark contrast to the powers obtained under MLE estimation, the  $KL_{n,10}$  test is not competitive at all when using the MME estimation technique for the alternatives in Table 1. For the mixture alternatives, the test proposed in this paper produces some of the highest estimated powers for almost all of their parameter configurations. In particular, our test with setting  $m = 4$  and  $a = 0.5$  performs best for both of these mixture alternatives under almost all mixing proportions considered.

It is clear that the proposed test produces higher estimated powers for the alternatives in Table 1 when using MME estimation compared to the corresponding powers obtained under MLE estimation. Conversely, these tests calculated under MLE have much higher powers than their MME counterparts when considering the mixture alternatives. As is common with these kinds of analyses, we conclude that there is no single test with the best overall power performance, but we find that our proposed test is competitive against the majority of other tests encountered in the literature when either MLE or MME estimation is used; this is true for both mixed and non-mixed alternatives.

Finally, it is clear from Tables 2 to 5 that, for the test statistic  $T_{n,m,a}^{(2)}$ , the choice of the parameters  $m$  and  $a$  potentially have a pronounced impact on the estimated powers produced. This fluctuation in power performance is briefly explored in Table 6 for the ‘Benini’ and ‘log-Weibull’ alternatives using a variety of different configurations of the parameters  $m$  and  $a$ . These particular alternatives were selected because they can be considered local alternatives; setting the parameter  $\theta$  to zero the Pareto distribution is obtained, with deviations from Pareto occurring when  $\theta > 0$ . The powers are obtained using a sample size of  $n = 20$  and the parameters  $\sigma$  and  $\beta$  are estimated using MME. From Table 6 it is clear that, in general, the highest powers are associated with smaller values of both  $m$  and  $a$ ; the powers taper off as both  $m$  and  $a$  increase. This tendency is observed in all six alternatives considered here but was also found to be true for many of the other alternatives considered in the main simulation study. There were a few exceptions to this, but we only report these results as they are representative of the common trend observed.

Table 2: Estimated powers for sample size  $n = 20$  based on MLE.

	$KS_n$	$CV_n$	$ZA_n$	$ZB_n$	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_n$	$T_{n,2,0.5}^{(2)}$	$T_{n,3,0.5}^{(2)}$	$T_{n,4,0.5}^{(2)}$
$\mathcal{P}(5,1)$	5	5	5	6	5	4	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	23	25	<b>38</b>	<b>37</b>	26	1	25	23	26	12	28	15	26	32
$\Gamma(0.8)$	7	7	6	5	<b>10</b>	<b>18</b>	7	4	1	4	3	7	3	2
$\Gamma(1.0)$	15	16	10	4	12	<b>39</b>	16	13	2	11	8	<b>19</b>	9	3
$\Gamma(1.2)$	25	29	19	6	18	<b>56</b>	29	25	3	21	16	<b>32</b>	18	7
$W(0.5)$	34	37	<b>52</b>	<b>50</b>	35	1	38	35	49	29	39	30	39	46
$W(0.8)$	6	6	6	6	<b>10</b>	<b>15</b>	6	4	2	2	3	7	3	2
$W(1.2)$	29	<b>32</b>	21	6	18	<b>63</b>	31	28	6	24	19	<b>32</b>	19	7
$W(1.5)$	47	<b>56</b>	42	14	30	<b>83</b>	55	52	21	48	38	<b>53</b>	38	19
$LN(0,1,0)$	13	13	7	3	8	<b>26</b>	13	11	1	10	8	<b>14</b>	9	3
$LN(0,1,2)$	7	7	4	2	6	<b>16</b>	7	6	1	4	4	<b>9</b>	4	2
$LN(0,1,5)$	4	4	3	3	<b>6</b>	<b>7</b>	4	3	2	2	3	<b>6</b>	3	2
$LN(0,2,5)$	29	32	37	33	21	0	33	31	<b>43</b>	34	32	25	38	<b>41</b>
$LFR(0.2)$	20	23	15	6	16	<b>51</b>	22	19	4	17	11	<b>24</b>	12	4
$LFR(0.5)$	24	27	18	6	18	<b>60</b>	27	23	6	22	14	<b>28</b>	14	5
$LFR(0.8)$	27	<b>30</b>	21	7	19	<b>64</b>	<b>30</b>	27	8	24	17	29	16	6
$LFR(1.0)$	28	<b>32</b>	23	7	21	<b>66</b>	31	28	9	25	18	30	17	6
$BE(0.5)$	19	21	<b>35</b>	<b>34</b>	26	2	22	19	22	9	22	11	21	26
$BE(0.8)$	8	8	7	6	<b>11</b>	<b>21</b>	8	6	1	5	4	8	3	2
$BE(1.0)$	15	16	10	4	13	<b>39</b>	15	12	2	11	7	<b>19</b>	8	3
$BE(1.5)$	34	41	28	9	20	<b>66</b>	40	37	7	33	25	<b>44</b>	29	12
$DH(0.2)$	6	6	3	3	6	<b>11</b>	6	4	1	3	3	<b>7</b>	4	2
$DH(0.4)$	11	12	6	2	7	<b>23</b>	11	10	1	7	6	<b>13</b>	8	3
$DH(0.6)$	21	22	12	4	11	<b>36</b>	<b>23</b>	20	1	16	13	<b>23</b>	16	7
$DH(0.8)$	30	<b>34</b>	22	6	16	<b>50</b>	33	31	3	27	21	<b>34</b>	25	12
$LW(0.1)$	4	4	3	3	<b>5</b>	<b>7</b>	4	3	2	3	3	4	3	3
$LW(0.3)$	<b>8</b>	<b>8</b>	4	2	6	<b>17</b>	<b>8</b>	7	1	5	4	<b>8</b>	6	5
$LW(0.5)$	15	15	9	3	9	<b>29</b>	<b>16</b>	14	1	11	8	<b>16</b>	12	11
$LW(0.7)$	25	<b>29</b>	16	5	13	<b>44</b>	28	26	2	21	17	25	22	21
$BN(0.1)$	4	4	3	4	<b>6</b>	<b>7</b>	4	3	2	2	3	4	3	3
$BN(0.3)$	6	6	4	4	<b>7</b>	<b>15</b>	6	4	1	3	3	5	4	3
$BN(0.5)$	<b>8</b>	<b>8</b>	5	3	7	<b>20</b>	<b>8</b>	6	1	5	4	6	5	4
$BN(0.7)$	<b>11</b>	<b>11</b>	6	3	8	<b>26</b>	<b>11</b>	9	1	7	5	8	6	6
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.1$	7	8	<b>9</b>	<b>9</b>	6	3	8	8	<b>9</b>	<b>9</b>	8	<b>9</b>	<b>9</b>	<b>9</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.3$	18	21	24	21	11	1	20	23	<b>25</b>	21	21	22	<b>25</b>	<b>27</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.5$	34	39	43	37	20	0	38	41	<b>45</b>	37	39	40	<b>45</b>	<b>48</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.7$	56	62	63	55	33	0	61	64	<b>67</b>	58	57	62	<b>67</b>	<b>69</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.1$	6	6	<b>8</b>	<b>8</b>	6	4	7	7	7	7	7	7	7	<b>8</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.3$	11	13	<b>19</b>	<b>18</b>	11	2	13	14	16	14	15	14	16	17
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.5$	24	26	<b>37</b>	34	20	1	25	27	32	23	30	25	32	<b>35</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.7$	39	43	<b>54</b>	50	33	0	42	45	49	37	47	40	49	<b>54</b>

Table 3: Estimated powers for sample size  $n = 20$  based on MME.

	$KS_n$	$CV_n$	$ZA_n$	$ZB_n$	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_n$	$T_{n,2,0.5}^{(2)}$	$T_{n,3,0.5}^{(2)}$	$T_{n,4,0.5}^{(2)}$
$\mathcal{P}(5,1)$	5	5	5	5	5	4	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	4	5	5	5	5
$\Gamma(0.5)$	10	10	10	11	<b>26</b>	1	11	8	9	4	12	6	12	<b>15</b>
$\Gamma(0.8)$	16	18	15	17	10	18	17	19	15	<b>22</b>	10	<b>21</b>	15	12
$\Gamma(1.0)$	34	41	37	40	13	37	40	43	36	<b>48</b>	22	<b>45</b>	37	31
$\Gamma(1.2)$	54	61	57	61	20	53	60	63	57	<b>68</b>	39	<b>64</b>	58	51
$W(0.5)$	12	10	14	16	<b>35</b>	1	12	8	12	5	16	14	15	<b>20</b>
$W(0.8)$	16	18	15	17	10	15	18	19	12	<b>22</b>	8	<b>20</b>	14	11
$W(1.2)$	53	60	57	61	20	61	59	62	58	<b>67</b>	42	<b>63</b>	57	50
$W(1.5)$	71	79	78	81	32	<b>82</b>	78	80	80	<b>82</b>	66	81	77	71
$LN(0,1.0)$	32	38	37	38	9	24	37	<b>40</b>	37	<b>44</b>	24	38	<b>40</b>	36
$LN(0,1.2)$	24	29	27	27	7	15	27	<b>30</b>	24	<b>33</b>	12	27	27	25
$LN(0,1.5)$	14	16	16	16	7	6	15	<b>17</b>	10	<b>18</b>	5	14	12	11
$LN(0,2.5)$	12	11	<b>31</b>	<b>36</b>	28	0	12	9	3	8	22	11	11	11
$LFR(0.2)$	41	49	45	50	17	50	48	50	45	<b>55</b>	30	<b>52</b>	45	38
$LFR(0.5)$	45	53	51	57	19	<b>59</b>	52	55	51	<b>61</b>	36	<b>59</b>	49	42
$LFR(0.8)$	47	55	54	59	20	<b>62</b>	54	57	55	<b>61</b>	40	59	53	45
$LFR(1.0)$	50	58	57	62	21	<b>65</b>	57	60	58	<b>63</b>	42	60	54	47
$BE(0.5)$	8	8	8	10	<b>26</b>	3	9	6	7	4	10	5	8	<b>11</b>
$BE(0.8)$	19	21	18	21	11	20	21	22	18	<b>26</b>	10	<b>23</b>	18	15
$BE(1.0)$	35	40	36	40	14	37	39	42	36	<b>48</b>	22	<b>44</b>	37	33
$BE(1.5)$	64	71	69	73	23	64	70	73	69	<b>77</b>	51	<b>76</b>	69	63
$DH(0.2)$	22	25	24	23	7	10	24	<b>26</b>	18	<b>28</b>	10	21	21	20
$DH(0.4)$	35	43	41	41	9	21	41	<b>44</b>	35	<b>47</b>	19	40	39	36
$DH(0.6)$	50	57	54	55	14	34	56	<b>59</b>	51	<b>62</b>	33	55	55	51
$DH(0.8)$	61	69	66	67	19	47	68	<b>71</b>	64	<b>74</b>	46	<b>71</b>	67	63
$LW(0.1)$	13	16	17	16	6	6	15	17	10	<b>18</b>	6	<b>18</b>	12	12
$LW(0.3)$	28	34	33	32	8	16	33	<b>36</b>	27	<b>38</b>	14	34	29	31
$LW(0.5)$	42	50	47	47	11	27	48	<b>52</b>	43	<b>55</b>	26	<b>52</b>	45	47
$LW(0.7)$	54	64	61	62	16	41	62	65	58	<b>68</b>	39	<b>66</b>	60	61
$BN(0.1)$	13	<b>16</b>	14	15	6	7	15	<b>16</b>	10	<b>17</b>	6	15	11	12
$BN(0.3)$	20	24	20	21	8	14	22	24	18	<b>27</b>	10	<b>25</b>	19	20
$BN(0.5)$	22	27	23	25	8	19	25	<b>28</b>	23	<b>31</b>	14	<b>28</b>	24	24
$BN(0.7)$	26	31	27	29	9	25	30	32	28	<b>36</b>	19	<b>33</b>	28	29
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.1$	<b>6</b>	<b>6</b>	<b>6</b>	5	<b>7</b>	3	<b>6</b>	<b>6</b>	<b>6</b>	<b>6</b>	5	<b>6</b>	<b>6</b>	<b>6</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.3$	12	13	12	10	12	1	13	12	13	11	9	12	<b>14</b>	<b>16</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.5$	24	25	24	22	22	0	26	25	27	20	14	24	<b>29</b>	<b>32</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.7$	43	46	45	42	36	0	46	45	48	35	18	42	<b>50</b>	<b>54</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.1$	5	5	5	5	<b>6</b>	4	5	5	5	5	5	5	5	<b>6</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.3$	7	7	7	6	<b>11</b>	2	7	7	8	6	8	7	8	<b>9</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.5$	12	14	13	12	<b>20</b>	1	14	13	15	10	13	12	16	<b>20</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.7$	23	25	27	25	<b>34</b>	0	26	24	27	17	22	22	30	<b>35</b>

Table 4: Estimated powers for sample size  $n = 30$  based on MLE.

	$KS_n$	$CV_n$	$ZA_n$	$ZB_n$	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_n$	$T_{n,2,0.5}^{(2)}$	$T_{n,3,0.5}^{(2)}$	$T_{n,4,0.5}^{(2)}$
$\mathcal{P}(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	33	36	<b>54</b>	<b>52</b>	35	2	37	33	36	16	41	21	36	44
$\Gamma(0.8)$	10	11	12	7	11	<b>26</b>	10	8	3	7	5	<b>14</b>	5	3
$\Gamma(1.0)$	24	29	23	8	16	<b>54</b>	28	28	9	25	16	<b>35</b>	21	10
$\Gamma(1.2)$	45	54	45	16	26	<b>76</b>	53	53	21	51	38	<b>59</b>	44	27
$W(0.5)$	49	54	<b>71</b>	<b>69</b>	49	1	54	50	66	41	58	43	55	63
$W(0.8)$	9	9	11	8	<b>12</b>	<b>22</b>	9	6	2	5	4	<b>12</b>	4	3
$W(1.2)$	50	59	49	16	27	<b>79</b>	58	58	31	55	42	<b>62</b>	47	31
$W(1.5)$	77	<b>86</b>	79	45	50	<b>95</b>	85	<b>86</b>	66	83	73	85	76	60
$LN(0,1,0)$	26	<b>30</b>	20	6	11	<b>36</b>	29	<b>30</b>	13	28	21	28	26	18
$LN(0,1,2)$	12	13	8	2	7	<b>21</b>	13	12	3	11	8	<b>15</b>	11	6
$LN(0,1,5)$	4	4	3	2	5	<b>9</b>	4	3	2	3	3	<b>9</b>	3	2
$LN(0,2,5)$	41	45	52	45	28	0	46	43	<b>57</b>	46	48	34	52	<b>57</b>
$LFR(0.2)$	35	42	34	10	21	<b>68</b>	41	39	16	38	25	<b>46</b>	29	15
$LFR(0.5)$	42	50	44	14	26	<b>79</b>	48	49	24	47	33	<b>54</b>	35	20
$LFR(0.8)$	47	57	51	18	30	<b>83</b>	55	55	30	54	38	<b>59</b>	40	24
$LFR(1.0)$	51	60	54	20	31	<b>85</b>	58	59	35	57	41	<b>61</b>	44	27
$BE(0.5)$	27	30	<b>50</b>	<b>49</b>	33	3	32	25	31	11	33	15	28	37
$BE(0.8)$	11	12	13	7	12	<b>30</b>	12	9	3	9	5	<b>14</b>	6	3
$BE(1.0)$	26	30	23	7	16	<b>53</b>	29	28	10	27	17	<b>35</b>	21	11
$BE(1.5)$	60	71	60	23	33	<b>85</b>	70	71	37	68	56	<b>74</b>	62	44
$DH(0.2)$	8	8	5	3	6	14	8	7	1	6	5	<b>9</b>	6	3
$DH(0.4)$	22	<b>24</b>	15	4	10	<b>30</b>	<b>24</b>	<b>24</b>	5	21	16	<b>24</b>	21	13
$DH(0.6)$	37	44	32	10	16	<b>50</b>	44	<b>45</b>	15	41	32	44	40	28
$DH(0.8)$	54	<b>62</b>	49	19	25	<b>68</b>	<b>62</b>	<b>62</b>	27	60	49	<b>62</b>	57	41
$LW(0.1)$	<b>5</b>	<b>5</b>	4	3	<b>5</b>	<b>8</b>	<b>5</b>	4	2	4	3	<b>5</b>	4	4
$LW(0.3)$	14	15	9	3	8	<b>21</b>	14	14	2	11	9	14	12	12
$LW(0.5)$	29	<b>34</b>	23	7	13	<b>41</b>	33	<b>34</b>	8	29	24	33	30	30
$LW(0.7)$	47	54	40	15	21	<b>60</b>	53	<b>55</b>	21	50	41	53	49	49
$BN(0.1)$	5	5	4	4	<b>6</b>	<b>8</b>	5	4	2	3	3	5	4	4
$BN(0.3)$	9	9	7	4	8	<b>19</b>	9	8	2	7	5	<b>11</b>	7	7
$BN(0.5)$	14	<b>16</b>	10	4	9	<b>28</b>	<b>16</b>	14	4	12	9	<b>16</b>	12	11
$BN(0.7)$	18	<b>20</b>	14	4	11	<b>35</b>	<b>20</b>	19	6	17	12	19	15	14
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.1$	8	8	<b>10</b>	<b>10</b>	7	3	8	9	<b>10</b>	9	<b>10</b>	9	<b>10</b>	<b>11</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.3$	23	26	<b>32</b>	27	13	1	26	28	<b>32</b>	26	30	27	<b>32</b>	<b>35</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.5$	48	54	58	50	26	0	53	56	<b>61</b>	50	57	54	<b>61</b>	<b>64</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.7$	73	79	79	71	45	0	78	80	82	73	76	78	<b>83</b>	<b>85</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.1$	6	7	<b>9</b>	<b>9</b>	6	4	7	7	8	7	8	7	8	<b>9</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.3$	15	17	<b>25</b>	<b>24</b>	14	2	17	18	21	16	21	17	21	23
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.5$	31	35	<b>50</b>	<b>48</b>	26	1	34	36	41	30	42	33	41	<b>48</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.7$	53	59	<b>72</b>	67	44	0	57	59	64	49	64	54	65	<b>69</b>

Table 5: Estimated powers for sample size  $n = 30$  based on MME.

	$KS_n$	$CV_n$	$ZA_n$	$ZB_n$	$KL_{n,1}$	$KL_{n,10}$	$M_{n,0.5}$	$M_{n,1}$	$G_{n,0.5}$	$G_{n,2}$	$A1_n$	$T_{n,2,0.5}^{(2)}$	$T_{n,3,0.5}^{(2)}$	$T_{n,4,0.5}^{(2)}$
$\mathcal{P}(5,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\mathcal{P}(10,1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.5)$	15	15	18	22	<b>35</b>	2	16	11	14	4	<b>27</b>	8	16	24
$\Gamma(0.8)$	21	23	19	24	12	<b>25</b>	22	23	18	<b>27</b>	12	<b>25</b>	18	14
$\Gamma(1.0)$	48	56	51	57	18	51	54	58	51	<b>64</b>	35	<b>61</b>	51	43
$\Gamma(1.2)$	74	81	78	82	30	73	80	83	78	<b>86</b>	61	<b>84</b>	78	70
$W(0.5)$	17	15	25	30	<b>49</b>	1	19	10	21	5	<b>42</b>	19	24	31
$W(0.8)$	21	24	19	23	13	20	23	24	16	<b>28</b>	9	<b>26</b>	17	13
$W(1.2)$	72	81	79	82	30	78	80	83	80	<b>86</b>	65	<b>84</b>	78	71
$W(1.5)$	91	95	95	<b>96</b>	53	95	95	<b>96</b>	95	<b>96</b>	89	<b>96</b>	94	90
$LN(0,1.0)$	49	57	56	57	14	34	55	<b>58</b>	57	<b>61</b>	40	56	<b>58</b>	56
$LN(0,1.2)$	33	39	39	<b>40</b>	9	20	37	<b>40</b>	36	<b>44</b>	20	36	37	36
$LN(0,1.5)$	17	<b>20</b>	<b>20</b>	<b>20</b>	7	8	19	<b>20</b>	13	<b>22</b>	6	16	15	13
$LN(0,2.5)$	13	11	38	<b>46</b>	35	0	13	8	5	6	<b>40</b>	13	14	16
$LFR(0.2)$	59	67	64	70	23	67	66	69	63	<b>73</b>	46	<b>71</b>	62	53
$LFR(0.5)$	64	73	72	77	27	77	72	75	70	<b>79</b>	56	<b>78</b>	68	59
$LFR(0.8)$	69	78	76	81	31	<b>82</b>	76	79	76	<b>82</b>	61	80	72	63
$LFR(1.0)$	71	79	79	82	32	<b>84</b>	78	81	78	<b>83</b>	64	<b>83</b>	74	66
$BE(0.5)$	11	10	14	18	<b>33</b>	3	12	7	9	3	<b>20</b>	5	10	16
$BE(0.8)$	25	29	24	28	13	29	28	29	23	<b>34</b>	14	<b>31</b>	22	17
$BE(1.0)$	49	58	52	58	18	51	56	59	52	<b>64</b>	35	<b>61</b>	54	45
$BE(1.5)$	84	90	88	90	38	83	89	91	88	<b>92</b>	77	<b>92</b>	88	83
$DH(0.2)$	27	33	32	32	8	13	32	<b>34</b>	26	<b>35</b>	13	28	28	27
$DH(0.4)$	51	59	56	56	14	28	57	<b>60</b>	52	<b>62</b>	33	54	55	53
$DH(0.6)$	69	77	74	75	22	48	76	<b>78</b>	73	<b>80</b>	54	75	76	72
$DH(0.8)$	81	87	85	86	32	65	86	88	85	<b>89</b>	71	<b>89</b>	88	82
$LW(0.1)$	17	<b>20</b>	<b>20</b>	<b>20</b>	7	7	19	<b>20</b>	14	<b>21</b>	7	<b>20</b>	14	15
$LW(0.3)$	40	<b>46</b>	<b>46</b>	45	11	20	45	<b>46</b>	39	<b>50</b>	22	<b>46</b>	41	43
$LW(0.5)$	61	69	67	67	18	38	68	71	63	<b>72</b>	43	<b>72</b>	65	66
$LW(0.7)$	76	83	81	82	27	56	82	<b>85</b>	80	<b>86</b>	63	84	81	82
$BN(0.1)$	17	<b>20</b>	18	19	7	8	19	<b>20</b>	13	<b>22</b>	7	19	14	15
$BN(0.3)$	26	30	27	28	9	18	29	<b>31</b>	23	<b>34</b>	14	<b>31</b>	25	26
$BN(0.5)$	33	39	34	37	11	26	37	40	33	<b>44</b>	21	<b>41</b>	34	35
$BN(0.7)$	38	44	40	43	12	33	43	46	41	<b>50</b>	26	<b>47</b>	42	41
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.1$	6	6	6	5	<b>7</b>	3	6	6	6	6	6	6	<b>7</b>	<b>7</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.3$	16	17	17	16	14	1	17	17	19	15	17	17	<b>20</b>	<b>22</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.5$	37	40	39	38	28	0	40	39	43	31	35	37	<b>46</b>	<b>50</b>
$\mathcal{P}(3,1) \& LN(-2.69,2), p=0.7$	63	66	65	64	49	0	66	65	69	52	48	60	<b>70</b>	<b>75</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.1$	5	5	5	5	<b>6</b>	4	5	5	5	5	5	5	5	<b>6</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.3$	9	9	9	9	<b>14</b>	2	9	9	10	8	<b>13</b>	9	11	<b>13</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.5$	18	20	22	22	<b>27</b>	1	20	19	23	14	26	18	24	<b>29</b>
$\mathcal{P}(3,1) \& W(0.5,0.25), p=0.7$	36	39	42	43	45	1	39	36	42	25	45	34	<b>46</b>	<b>53</b>

Table 6: Estimated powers of the test statistic  $T_{n,m,a}^{(2)}$  for varying choices of  $a$  and  $m$  (using MME estimation and sample size  $n = 20$ ) for 6 alternatives.

<b>LW(0.1)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$	<b>BN(0.1)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	18	12	12	12	$a = 0.5$	15	11	12	12
$a = 0.75$	13	13	11	11	$a = 0.75$	13	13	10	11
$a = 1$	11	12	12	9	$a = 1$	11	11	11	8
$a = 2$	9	9	9	9	$a = 2$	9	9	9	9

<b>LW(0.5)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$	<b>BN(0.5)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	52	45	47	48	$a = 0.5$	28	24	24	24
$a = 0.75$	48	49	42	44	$a = 0.75$	25	25	21	21
$a = 1$	44	44	44	30	$a = 1$	21	21	22	14
$a = 2$	30	30	30	30	$a = 2$	14	14	14	14

<b>LW(0.9)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$	<b>BN(0.9)</b>	$m = 2$	$m = 3$	$m = 4$	$m = 10$
$a = 0.5$	76	70	71	72	$a = 0.5$	36	31	31	32
$a = 0.75$	72	72	66	67	$a = 0.75$	32	32	28	28
$a = 1$	67	67	67	47	$a = 1$	28	28	28	19
$a = 2$	47	47	47	47	$a = 2$	19	19	19	19

## 5 Practical data application

To further investigate the behaviour of the test statistics studied in this paper, we now present the results obtained by applying those tests to a practical data set. The data set concerns the earnings for the 2022 inaugural season of LIV golf. LIV golf is a new golf series backed by the Saudi Arabian sovereign wealth fund which aims to be an alternative to the PGA Tour by attracting star players and providing larger paydays for winners. The data were obtained from [www.spotrac.com](http://www.spotrac.com) (2023) and collects the player earnings in the 2022 season. The data are shown in Table 7 and a box-plot of these values is provided in Figure 1.

It is clear from Figure 1 that the data values have a right-skewed tendency, with some extremely large outliers. The Pareto distribution is a distribution with heavy tails and is often used to model income above a specified threshold, so it is sensible to also consider this distribution as a possible model for these earnings values above a threshold. We therefore focus our attention on the LIV golf earnings above the threshold of 3 500 000, indicated in Figure 1 with a dashed grey line. The values above the threshold are extracted from the data set, scaled by dividing them by the *known* value of 3 500 000, and their empirical distribution is plotted in Figure 2. Overlaid on this empirical distribution plot are two parametric Pareto distributions: one where the parameter is estimated using MLE (producing the estimate  $\hat{\beta} = 1.781$ ) and the other where the parameter is estimated using MME (producing the estimate  $\hat{\beta}_n = 1.932$ ). From these figures it seems that the Pareto might be a good fit for the data, but to more formally test this assertion we now apply all of the tests discussed in this paper to the above-threshold, scaled LIV golf earnings data, with the results found in Table 8. The estimated  $p$ -values reported in Table 8 were obtained by making use of a parametric bootstrap employing  $B = 10\,000$  bootstrap replications from a Pareto distribution with parameter estimated using either the MLE or the

MME. From these  $p$ -values it is clear that *all* the tests do not reject the null hypothesis that the 2022 season's earnings of LIV golfers exceeding 3 500 000, follows a Pareto distribution.

Table 7: Data set: LIV golf earnings data set (accessed 2023-09-11).

33509017	16700083	13726499	13254785	11040714	9297500	8068000	8033500
7982785	7345000	7268167	5580314	5518500	4992618	4849000	4596000
4454500	4408000	4230964	4195367	3726583	3558666	3424333	3411314
3399100	3248000	3085000	2948500	2944950	2676500	2675833	2569917
2533833	2421714	2379700	2188285	2016350	1899000	1861667	1573000
1566999	1471200	1377000	1366000	1321950	1274285	1185000	1097000
1050000	1043000	1019000	958750	958750	840000	800000	800000
683000	651000	540000	530000	437000	360000	324000	268000
250000	232000	208750	190000	160000	156000	150000	140000
140000	133750	120000					

Table 8:  $p$ -values for the LIV golf earnings data set (accessed 2023-09-11).

Test	MLE	MME
$KS_n$	0.45	0.22
$CV_n$	0.55	0.24
$ZA_n$	0.80	0.50
$ZB_n$	0.68	0.50
$KL_{n,1}$	0.75	0.72
$KL_{n,10}$	0.39	0.46
$M_{n,0.5}$	0.58	0.27
$M_{n,1}$	0.40	0.20
$G_{n,0.5}$	0.38	0.22
$G_{n,2}$	0.30	0.16
$T_{n,2,0.5}^{(2)}$	0.47	0.22
$T_{n,3,0.5}^{(2)}$	0.36	0.22
$T_{n,4,0.5}^{(2)}$	0.37	0.24

## 6 Concluding remarks

In this paper, we propose two new classes of goodness-of-fit tests for the Pareto Type I distribution based on a characterization involving order statistics. In addition, we also derive the null distribution of these test statistics and show that the tests are indeed consistent. A Monte Carlo simulation study is presented to demonstrate the finite-sample performance of these tests under a variety of alternative distributions. Through the inclusion of other similar tests for the Pareto distribution, the simulation study also demonstrates that our tests are competitive when compared to these other tests. Finally, the choice of the two tuning parameters appearing in these tests was also roughly explored, with the finding that, when the tests are to be implemented in a practical setting, the choices  $a = 0.5$  and  $m = 2$  or  $m = 3$  can be recommended.

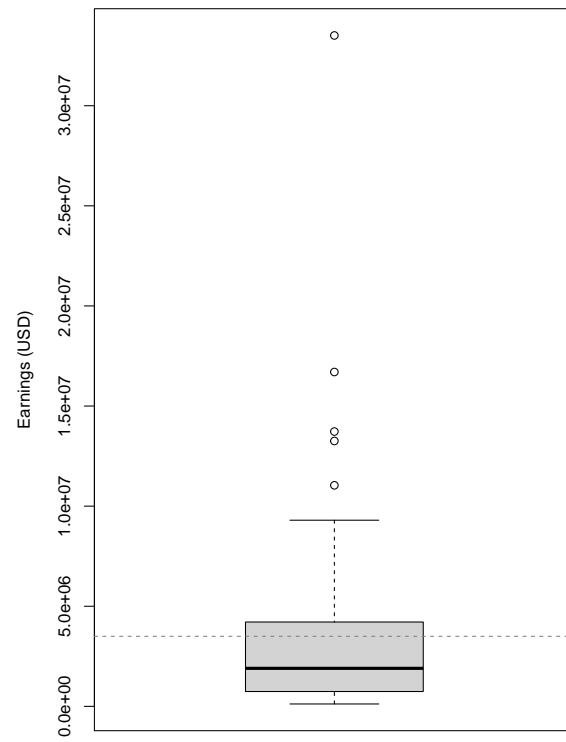


Figure 1: Box-plot of the LIV golf 2022 earnings. The dashed line represents the threshold of \$3 500 000.

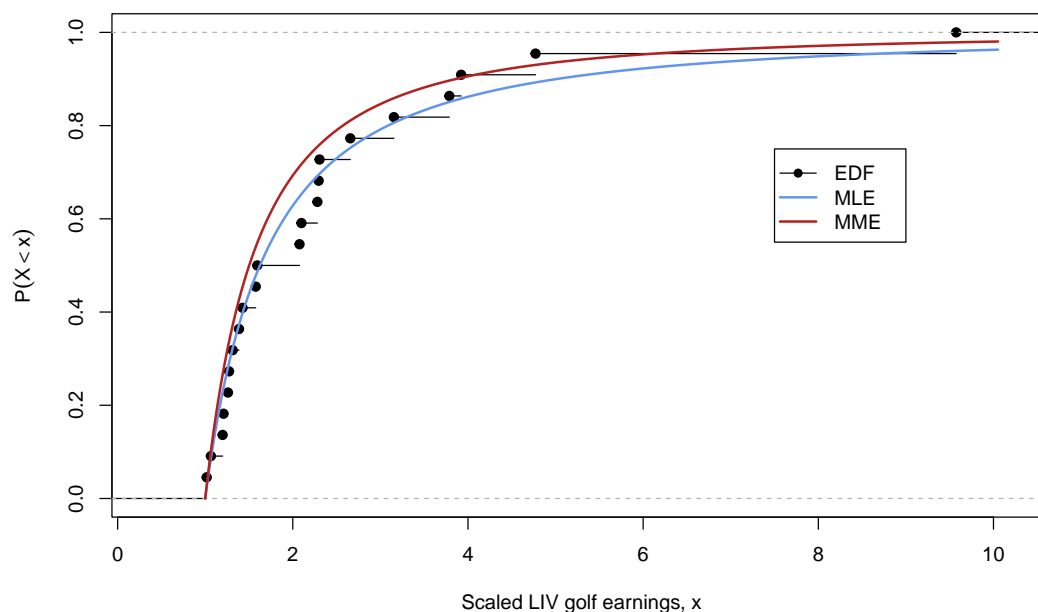


Figure 2: Empirical distribution function of the scaled above-threshold LIV golf 2022 earnings with two Pareto distribution overlays (one where the parameter is estimated via MLE and the other where it is estimated via MME).

## References

- Ahrari, V., Baratpour S. and Habibirad, A., & Fakoor, V. (2002). Goodness of fit tests for Rayleigh distribution based on quantiles. *Communications in Statistics*, 51(2), 341–357.
- Allison, J., Milošević, B., Obradović, M., & Smuts, M. (2022). Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. *Computational Statistics*, 37, 403–418.
- Arnold, B. (2015). *Pareto Distributions*. CRC Press, Taylor and Francis Group.
- Beirlant, J., de Wet, T., & Goegebeur, Y. (2004). Pricing risk when distributions are fat tailed. *Journal of Applied Probability*, 41A, 157–175.
- Beirlant, J., de Wet, T., & Goegebeur, Y. (2006). A goodness-of-fit statistic for the Pareto-type behavior. *Journal of Computational and Applied Mathematics*, 186, 99–116.
- Berger, J., & Mandelbrot, B. (1963). A new model for error clustering in telephone circuits. *IBM Journal of Research & Development*, 7, 224–236.
- Bilodeau, M., & Lafaye de Micheaux, P. (2005). A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, 9(2), 345–369.
- Brazauskas, V. (2000). Robust parametric modeling of the proportional reinsurance premium when claims are approximately Pareto-distributed. *Proceedings of the Business and Economic Statistics Section*, A, 144–149.

- Chu, J., Dickin, S., & Nadarajah, S. (2019). A review of goodness of fit tests for Pareto distributions. *Journal of Computational and Applied Mathematics*, *361*, 13–41.
- Csörgő, S. (1983). Kernel-transform empirical processes. *Journal of Multivariate Analysis*, *13*, 517–533.
- de Micheaux, P. L., & Tran, V. A. (2016). **Power**: A reproducible research tool to ease Monte Carlo power simulation studies for goodness-of-fit tests in R. *Journal of Statistical Software*, *69*(3), 1–42.
- Fan, Y., Lafaye de Micheaux, P., Penev, S., & Salopek, D. (2017). Multivariate nonparametric test of independence. *Journal of Multivariate Analysis*, *153*, 189–210.
- Fisk, P. (1961). The graduation of income distributions. *Econometrica*, *29*, 171–185.
- Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, *29*(3), 567–589.
- Gulati, S., & Shapiro, S. (2008). Goodness-of-fit tests for Pareto distribution. In *Statistical models and methods for biomedical and technical systems*, (pp. 259–274). Springer.
- Harris, C. M. (1968). The Pareto distribution as a queue service discipline. *Operations Research*, *16*, 307–313.
- Ioannides, Y., & Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, *73*, 18–29.
- Keiding, N., Hansen, O. K. H., Sorensen, D. N., & Slama, R. (2012). The current duration approach to estimating time to pregnancy. *Scandinavian Journal of Statistics*, *39*, 185–204.
- Keiding, N., Kvist, K., Hartvig, H., Tvede, M., & Juul, S. (2002). Estimating time to pregnancy from current durations in a cross-sectional sample. *Bio-statistics*, *3*, 565–578.
- Klass, O., Biham, O., Levy, M., Malcai, O., & Solomon, S. (2006). The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, *90*, 290–295.
- Martynov, G. (2009). Cramér-von Mises test for the Weibull and Pareto distributions. In *Proceedings of Dobrushin international conference Moscow*, (pp. 117–122).
- Meintanis, S. (2009a). Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity. *Journal of Statistical Planning and Inference*, *139*(2), 100–108.
- Meintanis, S. (2009b). A unified approach of testing for discrete and continuous Pareto laws. *Statistical Papers*, *50*(3), 569–580.
- Mert, M., & Saykan, Y. (2005). On a bonus-malus system where the claim frequency distribution is geometric and the claim severity distribution is Pareto. *Hacettepe Journal of Mathematics and Statistics*, *34*, 75–81.
- Milošević, B., & Obradović, M. (2016b). Two-dimensional Kolmogorov-type goodness-of-fit tests based on characterizations and their asymptotic efficiencies. *Journal of Nonparametric Statistics*, *28*, 413–427.
- Ndwandwe, L., Allison, J., Santana, L., & Visagie, I. (2022). Testing for the Pareto type I distribution: A comparative study. *arXiv preprint arXiv:2211.10088*.

- Ngatchou-Wandji, J. (2009). Testing for symmetry in multivariate distributions. *Statistical Methodology*, 6(3), 230–250.
- Obradović, M. (2015). On asymptotic efficiency of goodness of fit tests for Pareto distribution based on characterizations. *Filomat*, 29, 2311–2324.
- Obradović, M., Jovanović, M., & Milošević, B. (2015). Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. *Statistics*, 49, 1026–1041.
- Pareto, V. (1897). The new theories of economics. *Journal of Political Economy*, 4, 485–502.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.  
URL <https://www.R-project.org/>
- Rizzo, M. L. (2009). New goodness-of-fit tests for Pareto distributions. *ASTIN Bulletin*, 39, 69–715.
- Steindl, J. (1965). *Random Processes and the Growth of Firms*. Hafner Publishing.
- [www.spotracc.com](http://www.spotracc.com) (2023). LIV golf results by year. (Accessed 2023-09-11).  
URL <https://www.spotracc.com/liv/rankings/year/2022/>
- Volkova, K. (2016). Goodness-of-fit tests for Pareto distribution based on its characterization. *Statistical Methods and Applications*, 25, 351–373.
- Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 281–294.
- Zisheng, O., & Chi, X. (2006). Generalized Pareto distribution fit to medical insurance claims data. *Applied Mathematics - A Journal of Chinese Universities*, 21, 21–29.

## Chapter 3

# Article 2: New classes of tests for the Pareto distribution based on a conditional expectation characterisation

The second article, *New classes of tests for the Pareto distribution based on a conditional expectation characterisation* is currently a draft manuscript and has yet to be submitted to a journal. We plan to submit this paper to *Computational Statistics*.

---

Manuscript	No page limit.
Title	The title should be concise and informative.
Abstract and keywords	An abstract of 150 to 250 words and 4 to 6 keywords which can be used for indexing purposes.
Tables	All tables are to be numbered using Arabic numerals. Tables should always be cited in text in consecutive numerical order. For each table, please supply a table caption explaining the components of the table.
References	Cite references in the text by name and year in parentheses. The list of references should only include works that are cited in the text and that have been published or accepted for publication. Reference list entries should be alphabetized by the last names of the first author of each work.
General formatting	A L <sup>A</sup> T <sub>E</sub> X template is provided for submission.
Additional information	<a href="https://tinyurl.com/486kbupt">https://tinyurl.com/486kbupt</a>

---

# New classes of tests for the Pareto distribution based on a conditional expectation characterisation

T. Nombebe, J.S. Allison, L. Santana and I.J.H. Visagie

Pure and Applied Analytics, North-West Univeristy, Potchefstroom, South Africa

## Abstract

We propose and investigate new classes of goodness-of-fit tests for the Pareto Type I distribution based on a characterisation involving a conditional expectation. A Monte Carlo power study is included in order to assess the finite sample performance of the newly developed tests for four different estimation methods. The results from the simulation study shows that the newly proposed tests are competitive in terms of power performance when compared to some existing tests. It also shows that the majority of tests produce their highest powers when the unknown shape parameter is estimated by the method of moments. A practical example, where we consider the annual salaries of English Premier League football players for two consecutive seasons, is also included to illustrate the use of the newly proposed tests.

**Key words:** characterisation, conditional expectation, goodness-of-fit testing, Pareto distribution.

## 1 Introduction

Characterisations of probability distributions play an important role in understanding the properties of random variables. A characterisation of a distribution is a specific statistical or distributional attribute that *uniquely* identifies a given random variable (Koudou & Ley, 2014; Nagaraja, 2006). Characterisations are often exploited in order to construct goodness-of-fit tests for the distribution in question with the advantage being that such a test is often consistent. In recent years, the characterisation of probability distributions based on conditional expectations has grown in popularity, with a possible reason being that assuming knowledge of conditional expected values may be more

reasonable in many practical situations than assuming knowledge of the full probability distribution (Shanbhag, 1970).

When studying characterisations of probability distributions, Galambos & Kotz (1978) is an important reference, it collates and summarises the literature on this topic up to the date of publication. Specifically, this text provides an overview of the available characterisations of probability distributions using conditional expectations including, for example, the work of Shanbhag (1970) which characterised the exponential and geometric distributions in terms of their conditional expectations. The generalisations of this result by Hamdan (1972), Kotlarski (1972) and Shanbhag & Bhaskara Rao (1975), are also contained therein. However, the necessary and sufficient conditions for the implementation of the theory developed in the mentioned papers was only provided in Zoroa et al. (1990). Other important resources on the topic of conditional expectation include those by Gupta (1975) as well as Kotz & Shanbhag (1980). These references provide useful general representations of the entire distribution function of a random variable in terms of conditional expectations. Furthermore, Su & Huang (2000) extends the results of Ahmed (1991), Ruiz & Navarro (1994) and Ghitany et al. (1995) to characterise a distribution by considering the relationship between a conditional expectation and the hazard rate of the distribution.

While the literature discussed above generally pertains to large classes of distributions, this study specifically concentrates on the Pareto Type I distribution, simply referred to as the Pareto distribution below. The Pareto distribution is commonly used in various fields such as economics, finance, and insurance, where the occurrence of rare events or extreme values is of interest. Arnold (2015) gives an extensive historical survey of its use in the context of income distribution.

The density and distribution functions of the Pareto distribution are

$$f_{\beta,\sigma}(x) = \frac{\beta\sigma^{\beta+1}}{x^{\beta+1}}, \quad \text{and} \quad F_{\beta,\sigma}(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\beta}, \quad x \geq \sigma,$$

respectively, where  $\beta > 0$  is a shape parameter and  $\sigma > 0$  is a scale parameter. Note that the value of  $\sigma$  determines the support of the distribution. Since, in many practical applications, the support of the underlying distribution is known, the value of  $\sigma$  is frequently known *a priori*. This situation is demonstrated using the practical examples presented in Section 5. In these cases, a simple rescaling of the data can be employed in order to reduce the density and distribution functions above to

$$f_{\beta}(x) = \frac{\beta}{x^{\beta+1}}, \quad \text{and} \quad F_{\beta}(x) = 1 - x^{-\beta}, \quad x \geq 1. \quad (1.1)$$

In the remainder of this paper, we consider the case where  $\sigma$  is known, meaning that we define a Pareto distribution via the functions in (1.1).

We are interested in the construction of new goodness-of-fit tests for the Pareto distribution based on a characterisation involving a conditional expectation. Dallas (1976) investigated the characterisation of the Pareto and power distributions by using the relationship between these distributions and the exponential distribution. That is, if  $X$  is exponentially distributed, then  $e^X$  follows a Pareto distribution. The characterisation used in Dallas (1976) is based on the lack-of-memory property of the transformed variable and it is shown that this characterisation can be expressed in terms of conditional expectations rather than conditional distributions. Khan & Abu-Salih (1989) characterised many well-known continuous probability distributions, including the Pareto and power distributions, using conditional expectations of functions of order statistics. Wu & Lee (2001) employed conditional expectations of record values to characterise the generalised extreme value, power, and Pareto distributions. Lee (2003) discussed characterisations of the Pareto distribution by conditional expectations of record values.

The specific focus of this paper is a characterisation based on conditional expectation by Nofal & El Gebaly (2017), discussed in some detail in Section 2. Let  $X_1, \dots, X_n$  be a random sample from some unknown distribution  $F$ . We are interested in testing the composite hypothesis that these data are realised from the Pareto distribution;

$$H_0 : F(x) = F_\beta(x), \tag{1.2}$$

for all  $x > 1$  and some  $\beta > 0$ , tested against general alternatives.

The remainder of the paper is organised as follows. In Section 2, we introduce new classes of goodness-of-fit tests based on a conditional expectation characterisation proposed by Nofal & El Gebaly (2017) to test the hypothesis stated in (1.2). Section 3 is devoted to a discussion of the methods of parameter estimation used in the paper as well as the calculation of critical values. Results of a Monte Carlo simulation to compare the power performance of the newly proposed tests to some commonly used existing tests are presented in Section 4. Section 5 demonstrates the use of the tests applied to observed data. Finally, in Section 6 some concluding remarks are provided. Some details about the derivations of the test statistics are referred to the Appendix.

## 2 Newly proposed test statistics

In this section we develop classes of goodness-of-fit tests based on a conditional expectation characterisation of the Pareto distribution. This characterisation is proposed in Nofal & El Gebaly (2017)

and can be restated as follows:

**Theorem 2.1.** *Let  $X$  be a nonnegative continuous random variable with distribution, survival and density functions  $F$ ,  $S$  and  $f$ , respectively. Then  $X$  is said to follow the Pareto distribution with parameter  $\beta$  if, and only if, for all  $t \geq 1$*

$$E[X^s | X \geq t] = E(X^s)t^s, \quad s < \beta.$$

It easily follows from Theorem 2.1 that  $X$  has a Pareto distribution if, and only if, for all  $t \geq 1$

$$E[X^s I(X \geq t)] = S(t)E(X^s)t^s, \quad s < \beta, \quad (2.1)$$

where  $I(\cdot)$  is the indicator function.

From (2.1) we have that  $X$  follows a Pareto distribution if, and only if, for all  $t \geq 1$

$$\Psi_s(t) := E[X^s I(X \geq t)] - S(t)E(X^s)t^s = 0.$$

We can now estimate  $\Psi(t)$  from the observed data  $X_1, \dots, X_n$ . Two possible estimators for  $\Psi(t)$  are

$$\Psi_{n,s}^{[1]}(t) = \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - S_{\hat{\beta}}(t) \frac{t^s}{n} \sum_{j=1}^n X_j^s$$

and

$$\Psi_{n,s}^{[2]}(t) = \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - \frac{t^s}{n} \sum_{j=1}^n X_j^s \frac{1}{n} \sum_{j=1}^n I(X_j > t),$$

where  $S_{\hat{\beta}}(t) = t^{-\hat{\beta}}$ , for some estimator  $\hat{\beta}$  of the shape parameter  $\beta$ . Different estimation methods are discussed in Section 3. The only difference between these two estimators is the way in which  $S(t)$  is estimated. In  $\Psi_n^{[2]}$  we estimate the survival function non-parametrically, while in  $\Psi_n^{[1]}$  we use a parametric estimate,  $S_{\hat{\beta}}(t)$ .

If the null hypothesis is true, then both  $\Psi_{n,s}^{[1]}(t)$  and  $\Psi_{n,s}^{[2]}(t)$  should be close to zero. Therefore, we propose the following two weighted  $L2$ -type test statistics:

$$L_{n,s,a}^{[1]} = \int_1^\infty \left( \Psi_{n,s}^{[1]}(t) \right)^2 w_a(t) dt$$

and

$$L_{n,s,a}^{[2]} = \int_1^\infty \left( \Psi_{n,s}^{[2]}(t) \right)^2 w_a(t) dt,$$

where  $w_a(t)$  is a weight function ensuring the existence of the integrals above. Henceforth choose  $w_a(t) = t^{-a}$ , with  $a > 1$  being a user-defined tuning parameter.

After some simplification we obtain the following easily calculable forms of the two test statistics

$$L_{n,s,a}^{[1]} = \sum_{j=1}^n \sum_{k=1}^n \frac{X_j^s X_k^s}{n^2} \left\{ \frac{1 - \min(X_j, X_k)^{-a+1}}{a-1} - \frac{2(X_j^{s-\widehat{\beta}-a+1} - 1)}{s - \widehat{\beta} - a + 1} - \frac{1}{2s - 2\widehat{\beta} - a + 1} \right\}$$

and

$$L_{n,s,a}^{[2]} = \sum_{j=1}^n \sum_{k=1}^n \frac{X_j^s X_k^s}{n^2} \left\{ \frac{1 - \min(X_j, X_k)^{-a+1}}{a-1} - \frac{2(\min(X_j, X_k)^{s-a+1} - 1)}{n(s-a+1)} + \frac{\min(X_k, X_j)^{2s-a+1} - 1}{n^2(2s-a+1)} \right\}.$$

We can obtain two additional test statistics by noting that the right-hand side of (2.1) can be written as

$$S(t)E(X^s)t^s = P(X > t) \frac{\beta}{\beta-s} \frac{t^{s+1}}{\beta} \frac{\beta}{t} = \frac{f_\beta(t)t^{s+1}}{\beta-s}, \quad (2.2)$$

using the fact that  $E(X^s) = \beta/(\beta-s)$ , if  $X$  follows a Pareto distribution.

Substituting (2.2) into (2.1) we now have that  $X$  follows the Pareto law, if and only if, for all  $t \geq 1$

$$\Lambda_s(t) := E[X^s I(X \geq t)] - \frac{f_\beta(t)t^{s+1}}{\beta-s} = 0.$$

Again, we can estimate this quantity from the observed data  $X_1, \dots, X_n$ . Two possible estimators for  $\Lambda_s(t)$  are

$$\Lambda_{n,s}^{[1]}(t) = \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - \frac{f_{\widehat{\beta}}(t)t^{s+1}}{\widehat{\beta}-s}$$

and

$$\Lambda_{n,s}^{[2]}(t) = \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - \frac{\widehat{f}_h(t)t^{s+1}}{\widehat{\beta}-s},$$

where  $f_{\widehat{\beta}}(t) = \widehat{\beta}t^{-\widehat{\beta}-1}$  and  $\widehat{f}_h(t)$  is a kernel density estimator with bandwidth  $h$ . In  $\Lambda_{n,s}^{[1]}$  the density function is estimated parametrically, while in  $\Lambda_{n,s}^{[2]}$  we make use of a non-parametric kernel density estimator.

As before, we propose the following two weighted  $L_2$ -type test statistics

$$T_{n,s,a}^{[1]} = \int_1^\infty \left( \Lambda_{n,s}^{[1]}(t) \right)^2 w_a(t) dt$$

and

$$T_{n,s,a}^{[2]} = \int_1^\infty \left( \Lambda_{n,s}^{[2]}(t) \right)^2 w_a(t) dt.$$

Unfortunately, for the most common kernel functions used in  $\hat{f}_h$ , the test statistic  $T_{n,s,a}^{[2]}$  cannot be expressed in a simple closed-form expression. Even though it is possible to calculate the value of this statistic for any kernel using numerical integration, the results of some preliminary simulations show that this statistic exhibits erratic power, so we will not pursue it further in this paper.

The calculable form for  $T_{n,s,a}^{[1]}$  is given by

$$T_{n,s,a}^{[1]} = \frac{1}{n^2(a-1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s (1 - \min(X_j, X_k)^{-a+1}) - \frac{2\hat{\beta}}{n(\hat{\beta}-s)(s-\hat{\beta}-a+1)} \sum_{j=1}^n X_j^s (X_j^{s-\hat{\beta}-a+1} - 1) - \frac{\hat{\beta}^2}{(2s-2\hat{\beta}-a+1)(\hat{\beta}-s)^2}.$$

All of the newly proposed tests reject the null hypothesis for large values of the test statistics.

### 3 Parameter and critical value estimation

Below we consider four estimators for the shape parameter  $\beta$ ; the maximum likelihood estimator (MLE), the method of moments estimator (MME) and two minimum distance estimators. The MLE and MME are well known and we include only the required formulas. The first minimum distance estimator considered minimises squared differences while the second minimises a Cramér-von Mises type distance. After considering the estimation techniques, we turn our attention to the calculation of critical values.

#### 3.1 Parameter estimation

The MLE of  $\beta$  is given by

$$\hat{\beta} = \frac{n}{\sum_{j=1}^n \log X_j}.$$

The MME is obtained by equating the first moment of the fitted distribution to the first sample moment. Of course, this implicitly assumes that the first moment exists or, equivalently, that  $\beta > 1$ .

The MME is then

$$\tilde{\beta} = \frac{\bar{X}_n}{\bar{X}_n - 1},$$

where  $\bar{X}_n = \sum_{j=1}^n X_j/n$ .

We now consider the two minimum distance estimators. These estimators are obtained by minimising some discrepancy measure between  $F_{\hat{\beta}}$  and the empirical distribution function,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n I(X_j \leq x).$$

The Cramér–von Mises (CV) distance measure is based on the squared integral difference between  $F_n$  and  $F_\beta$ , and is given by

$$D_n^{CV}(\beta) = \int_1^\infty (F_n(x) - F_\beta(x))^2 f_\beta(x) dx, \quad (3.1)$$

with calculable form

$$D_n^{CV}(\beta) = \frac{1}{12n} + \sum_{j=1}^n \left( 1 - X_{(j)}^{-\beta} - \frac{2j-1}{2n} \right)^2,$$

with  $X_{(1)}, \dots, X_{(n)}$  denoting the order statistics. The resulting estimator is

$$\hat{\beta}_{CV} = \arg \min_{\beta} D_n^{CV}(\beta).$$

The final estimator considered is the so-called squared difference estimator. The discrepancy measure to be minimised is similar to (3.1), but the integrand is now weighted by  $F_n$ ;

$$\begin{aligned} D_n^{SD}(\beta) &= \int_1^\infty (F_n(x) - F_\beta(x))^2 dF_n(x) \\ &= \sum_{j=1}^n \left( 1 - X_{(j)}^{-\beta} - \frac{j}{n+1} \right)^2. \end{aligned}$$

The corresponding estimator is

$$\hat{\beta}_{SD} = \arg \min_{\beta} D_n^{SD}(\beta).$$

When computing MDEs, we use the `optim` function in R together with the “BFGS” optimisation method (see Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). These calculations require starting values for the optimisation procedure, the MMEs are used for this purpose.

We only consider two minimum distance estimators above. For a full treatment of the subject, the interested reader is referred to Wolfowitz (1953), Beran (1977, 1978), Boos (1981), as well as Parr & Schucany (1980). A recent reference on minimum distance estimators, specifically in the context of the Lomax distribution, is Nombebe et al. (2023).

### 3.2 Estimation of critical values

We now turn our attention to the calculation of numerical critical values. When using MLEs together with the transformed data

$$Y_j = X_j^{\hat{\beta}}, \quad j = 1, \dots, n,$$

the critical value are independent of the value of  $\hat{\beta}$ . In this case, we obtain the critical values via Monte Carlo simulation. However, when using the other estimation techniques, the critical value depends on the value of  $\hat{\beta}$ . In these cases, we employ the parametric bootstrap procedure outlined in Ndwandwe et al. (2023a).

## 4 Monte Carlo study

A Monte Carlo study is employed to examine the empirical power performance of the tests discussed in the previous sections against various fixed alternative distributions. The aim of this simulation study is to compare the finite-sample power performance of the newly proposed tests to the following existing goodness-of-fit tests for the Pareto distribution.

- The traditional Kolmogorov-Smirnov ( $KS_n$ ), Cramer-von Mises ( $CV_n$ ), Anderson-Darling ( $AD_n$ ) and modified Anderson-Darling ( $MA_n$ ) tests.
- A test proposed by Zhang (2002) based on the likelihood ratio. The test statistic has calculable form

$$ZA_n = - \sum_{j=1}^n \left[ \frac{\log \left\{ 1 - X_{(j)}^{-\hat{\beta}_n} \right\}}{n - j + \frac{1}{2}} + \frac{\log \left\{ X_{(j)}^{-\hat{\beta}_n} \right\}}{j - \frac{1}{2}} \right].$$

- A test based on the Mellin transform proposed by Meintanis (2009). The test statistic is

$$G_{n,a} = \frac{1}{n} \left[ (\hat{\beta}_n + 1)^2 \sum_{j,k=1}^n I_w^{(0)}(X_j X_k) + \sum_{j,k=1}^n I_w^{(2)}(X_j X_k) + 2(\hat{\beta}_n + 1) \sum_{j,k=1}^n I_w^{(1)}(X_j X_k) \right] \\ + \hat{\beta}_n \left[ n \hat{\beta}_n I_w^{(0)}(1) - 2(\hat{\beta}_n + 1) \sum_{j=1}^n I_w^{(0)}(X_j) - 2 \sum_{j=1}^n I_w^{(1)}(X_j) \right],$$

where

$$I_w^{(m)}(t) = \int_0^\infty (t-1)^m \frac{1}{x^t} w(t) dt, \quad m = 0, 1, 2.$$

For the choice  $w(t) = e^{-at}$ , we have that

$$\begin{aligned} I_a^{(0)}(t) &= (a + \log t)^{-1}, \\ I_a^{(1)}(t) &= \frac{1 - a - \log t}{(a + \log t)^2}, \\ I_a^{(2)}(t) &= \frac{2 - 2a + a^2 + 2(a - 1) \log t + \log^2 t}{(a + \log t)^3}. \end{aligned}$$

We show powers obtained using  $a = 2$  in the numerical results below as this has been shown to produce good powers.

- A test based on the difference between two empirical characteristic functions proposed by Ndwandwe et al. (2023b). The test statistic is

$$NA_{n,m,a} = n \int_{-\infty}^{\infty} |\phi_{n,m}(t) - \xi_{n,m}(t)|^2 e^{-at^2} dt,$$

where

$$\phi_{n,m}(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j^{1/m}}$$

and

$$\xi_{n,m}(t) = \frac{1}{n^m} \sum_{k_1=1}^n \dots \sum_{k_m=1}^n e^{it \min(X_{k_1}, \dots, X_{k_m})}.$$

The test statistic has the following simplified form

$$\begin{aligned} NA_{n,m,a} &= \frac{1}{n} \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n \left\{ \exp \left[ \frac{-\left(X_{(j)}^{1/m} - X_{(k)}^{1/m}\right)^2}{4a} \right] - 2nv_{j,m} \exp \left[ \frac{-\left(X_{(j)} - X_{(k)}^{1/m}\right)^2}{4a} \right] \right. \\ &\quad \left. + n^2 v_{j,m} v_{k,m} \exp \left[ \frac{-\left(X_{(j)} - X_{(k)}\right)^2}{4a} \right] \right\}, \end{aligned}$$

where

$$v_{j,m} := \frac{1}{n^m} [(n - j + 1)^m - (n - j)^m].$$

For implementing the test, Ndwandwe et al. (2023b) recommend choosing  $m = 3$  and  $a = 2$ ; results for these choices are therefore shown in the tables below.

All tests reject for large values of the test statistics.

## 4.1 Monte Carlo setup

Our aim is to compare the numerical sizes and powers achieved by the newly proposed tests to those of the existing tests for the Pareto distribution discussed above. We consider the powers achieved against a range of fixed alternative distributions with support  $x > 1$  in various settings. Two different sample sizes,  $n = 20$  and  $n = 30$ , are used, as well as different estimation methods for  $\beta$ . The alternative distributions considered, together with their density functions and the notation used for these distributions, are specified in Table 1.

Table 1: Probability density functions for choices of the alternative distributions.

Alternative	Density function	Notation
Pareto ( $\theta$ )	$\frac{\theta}{x^{\theta+1}}$	$P(\theta)$
Gamma ( $\theta$ )	$\frac{1}{\Gamma(\theta)}(x-1)^{\theta-1}e^{-(x-1)}$	$\Gamma(\theta)$
Log-normal( $\theta$ )	$\exp\left\{-\frac{1}{2}(\log(x-1)/\theta)^2\right\} / \{\theta(x-1)\sqrt{2\pi}\}$	$LN(\theta)$
Linear failure rate( $\theta$ )	$(1+\theta(x-1))\exp(-(x-1)-\theta(x-1)^2/2)$	$LF(\theta)$
Beta exponential( $\theta$ )	$\theta e^{-(x-1)}(1-e^{-(x-1)})^{\theta-1}$	$BE(\theta)$
Tilted Pareto( $\theta$ )	$\frac{1+\theta}{(x+\theta)^2}$	$TP(\theta)$
Log-Weibull( $\theta$ )	$\frac{(1+\theta)\log^\theta x}{x^{2+\theta}}$	$LW(\theta)$
Benini( $\theta$ )	$\frac{\exp(-\theta\log^2 x)}{x^2}(1+2\theta\log x)$	$BN(\theta)$

A nominal significance level of 5% is used throughout and all calculations are performed using R v4.2.2 (R Core Team, 2021). The estimated powers, reported in Tables 4 to 7, are the percentages of rejections of the null hypothesis in 10 000 independent Monte Carlo replications (rounded to the nearest integer). In an attempt to aid comparison, the highest and second highest powers against each alternative distribution are printed in bold and highlighted in black and grey, respectively. When  $\beta$  is estimated by its MLE, the empirical critical values of all the tests are obtained using  $MC = 50\,000$  independent Monte Carlo replications. For all other estimation techniques, the critical values are obtained using Giacomini et al.'s (2013) 'warp-speed' bootstrap procedure over each of the  $MC = 10\,000$  Monte Carlo replications.

For both sample sizes considered, we calculate the numerical powers obtained when estimating  $\beta$  using the four different estimators discussed in Section 3. However, the minimum distance estimators are omitted from the power tables due to their inferior power performance compared to those associated with the MLE and MME. In order to illustrate the differences between the numerical

powers achieved by the different estimation techniques, consider Figure 1. This figure illustrates the empirical powers against the  $TP$ ,  $LW$  and  $BN$  alternative distributions, calculated when  $n = 20$ , for each of the four estimation methods for the tests  $L_{n,0.1,3}^{[2]}$ ,  $T_{n,0.1,3}^{[1]}$  and  $AD_n$ . These three distributions have the Pareto distribution as a special case when  $\theta = 0$ , with deviations from the null hypothesis for  $\theta > 0$ . That is, they can be used to illustrate contiguous alternatives to the Pareto distribution as  $\theta$  increases above zero. When considering the numerical powers shown in Figure 1, it is clear that MME produces the highest power, while the second highest powers are obtained using MLE. The two minimum distance estimators are associated with the lowest powers in each of the distributions considered. This trend is also observed when considering other alternative distributions. As a result, we concern ourselves only with the powers achieved using the MLE and the MME in the power tables to follow.

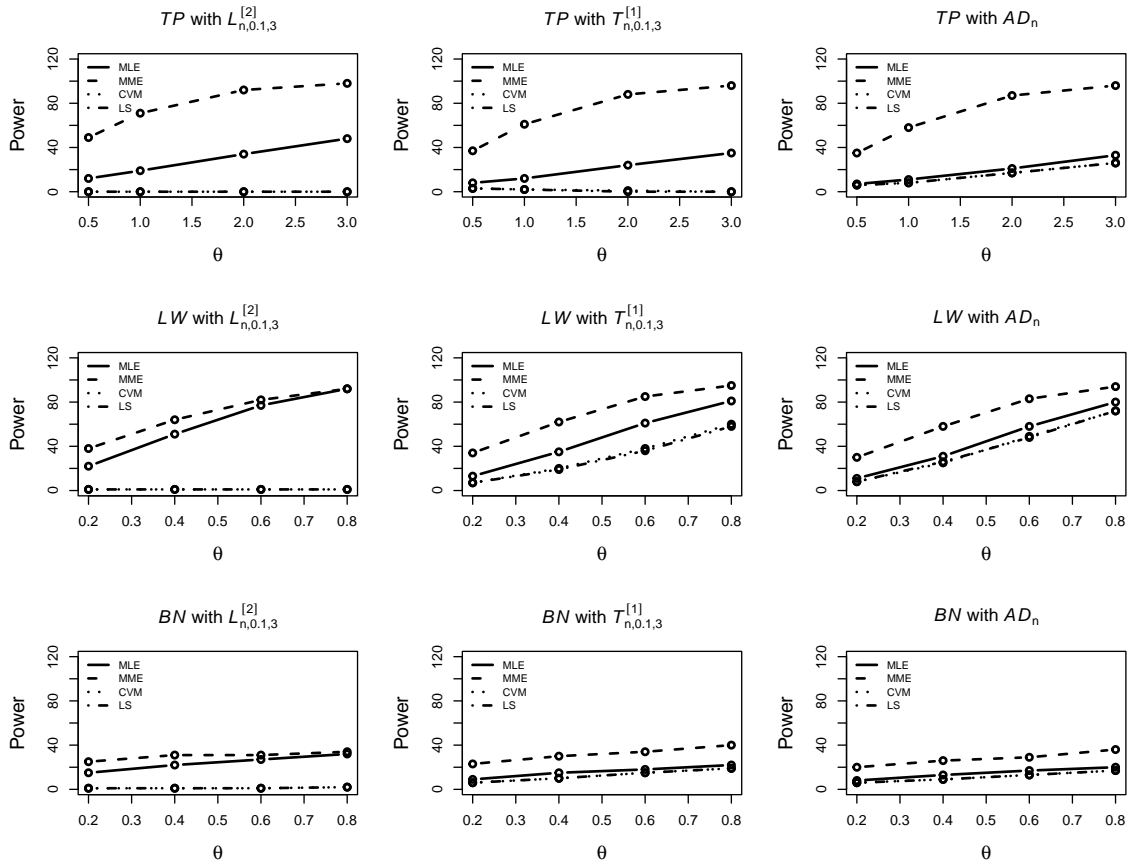


Figure 1: Numerical powers achieved using the different estimation techniques.

Each of the newly proposed tests contains two tuning parameters:  $a$  and  $s$ . The construction of the tests implicitly assumes that at least the  $s^{\text{th}}$  moment of the underlying distribution exists and so we recommend using fairly small values for  $s$ . A preliminary power study indicated that the combination of  $a = 3$  and  $s = 0.1$  generally provide high powers against a wide range of alternative distributions. As a result, we recommend this combination of tuning parameters when using the test. The empirical powers reported in Tables 4 to 7 were obtained using these parameters.

## 4.2 Discussion of numerical results

In the discussion presented below we first consider those results associated with the MLE before turning our attention to the empirical powers obtained using the MME. Thereafter we compare these two sets of results to one another.

### 4.2.1 Sizes and powers obtained using MLE

Consider the numerical sizes and powers obtained using MLE, see Tables 4 and 6.

All of the tests considered largely maintain the specified nominal significance level of 5% for both of the sample sizes used. As is to be expected, the powers of the tests generally increase with sample size.

The tests that generally achieve the highest powers against the specified alternative distributions are the  $L_{n,0.1,3}^{[2]}$  and  $NA_{n,3,2}$  tests. Specifically, the  $L_{n,0.1,3}^{[2]}$  test is shown to produce very high powers against  $LN$ ,  $TP$ ,  $LW$  and  $BN$  for the majority of the choices of the parameters of these alternative distributions. This test either outperforms all competitors outright, or produces powers close to the best performing competitor against these distributions for both of the sample sizes considered. However, we note that while the test has excellent performance in the majority of cases investigated, it produces no power for the alternatives  $\Gamma(0.4)$ ,  $LN(2.5)$ , and  $BE(0.5)$ .

### 4.2.2 Sizes and powers obtained using MME

Interestingly, the two tests with the best performance when using the MLE, namely  $L_{n,0.1,3}^{[2]}$  and  $NA_{n,3,2}$ , are undersized when using the MME for both sample sizes. This phenomenon is especially

pronounced in the case of the  $L_{n,0.1,3}^{[2]}$ . However, both of these tests achieve high powers against a wide range of alternatives. In fact, these two tests, together with  $MA_n$  and  $T_{n,0.1,3}^{[1]}$  generally provide the highest power against the widest range of alternative distributions. The most powerful of the newly proposed tests remains  $L_{n,0.1,3}^{[2]}$  test which is shown to be especially powerful against the  $TP$  and  $LW$  distributions.

### 4.2.3 A comparison between the results obtained using MLE and MME

For each of the tests considered, the average power against all of the alternatives considered is higher when using MME than is the case when using MLE. That is, both the existing and new tests produce higher powers when estimating the shape parameter of the Pareto distribution using MME rather than MLE. This finding is supported by the empirical results shown in Figure 1. As a result, we recommend using MME in order to do parameter estimation regardless of the test used.

Based on all of the numerical results presented, we would recommend using either  $L_{n,0.1,3}^{[2]}$  or  $NA_{n,3,2}$ , with  $\beta$  estimated using MME, when testing the hypothesis that data are realised from a Pareto distribution.

## 5 Practical data application

As was mentioned in the introduction, the Pareto distribution is used as a model for a wide variety of phenomena. In the field of extreme value theory, this distribution is often used to model the tails of a distribution, especially in cases where we expect *a priori* that the underlying distribution will exhibit a heavy tail.

Below we consider the observed annual salaries of English Premier League (EPL) football players for the seasons 2021–2022 as well as 2022–2023. Specifically, we are interested in determining whether or not the Pareto distribution is an appropriate model for the distribution of the salaries of players in excess of 10 million GBP. As a result, the hypothesis of interest is whether or not the reported salaries, rescaled by a factor of 10 million, are realised from the Pareto distribution. We consider the salaries from the two seasons separately. Both data sets used are obtained from [www.spotracc.com](http://www.spotracc.com) (2023). These salaries (reported in millions of GBP) can be found in Table 2. The sample sizes are 20 and 28, respectively.

Table 2: Practical data example: EPL earnings above 10 million GBP for the 2021-2022 and 2022-2023 seasons (values are reported in millions).

2021–2022				
10.31	10.40	10.40	10.40	10.40
10.40	11.44	13.00	13.00	14.14
15.08	15.08	15.60	15.60	16.90
17.68	18.20	19.50	20.80	26.80
2022–2023				
10.40	10.40	10.40	10.40	10.40
10.40	10.40	10.40	11.44	11.44
11.70	12.48	13.00	13.00	13.00
13.78	15.08	15.34	15.47	15.60
15.60	16.90	17.68	18.20	18.20
19.50	19.50	20.80		

We use each of the tests discussed in Section 4 to test the hypothesis of interest for both of the data sets considered. The code for this application can be found here:

<https://tinyurl.com/CondExp2023>.

However, before turning our attention to these formal hypothesis testing procedures, we present some graphical tests below. The empirical distribution functions (EDF) of each of the data sets are plotted in Figure 2. These distributions are overlaid with two fitted Pareto distributions in each case; these correspond to the fitted distributions obtained using MLE and MME.

The distributions showcased in Figure 2 indicate that the salaries in the 2021–2022 season correspond quite closely to what would be expected under the Pareto distribution. While no glaringly obvious departure from this distribution is visible in the comparison for the 2022–2023 salaries, the tail of the empirical distribution seems to be lighter than is the case for the fitted distributions. Furthermore, around the 15 million GBP mark (indicated by 1.5 on the  $x$ -axis), the empirical distribution seems to be systematically lower than the fitted distributions.

Another graphical test that we employ is side-by-side violin plots (with overlaid boxplots). These are shown, for both data sets considered, in Figure 3. One can see from these plots that both data sets have a pronounced positive third central sample moment, which is indicative of a right-skewed

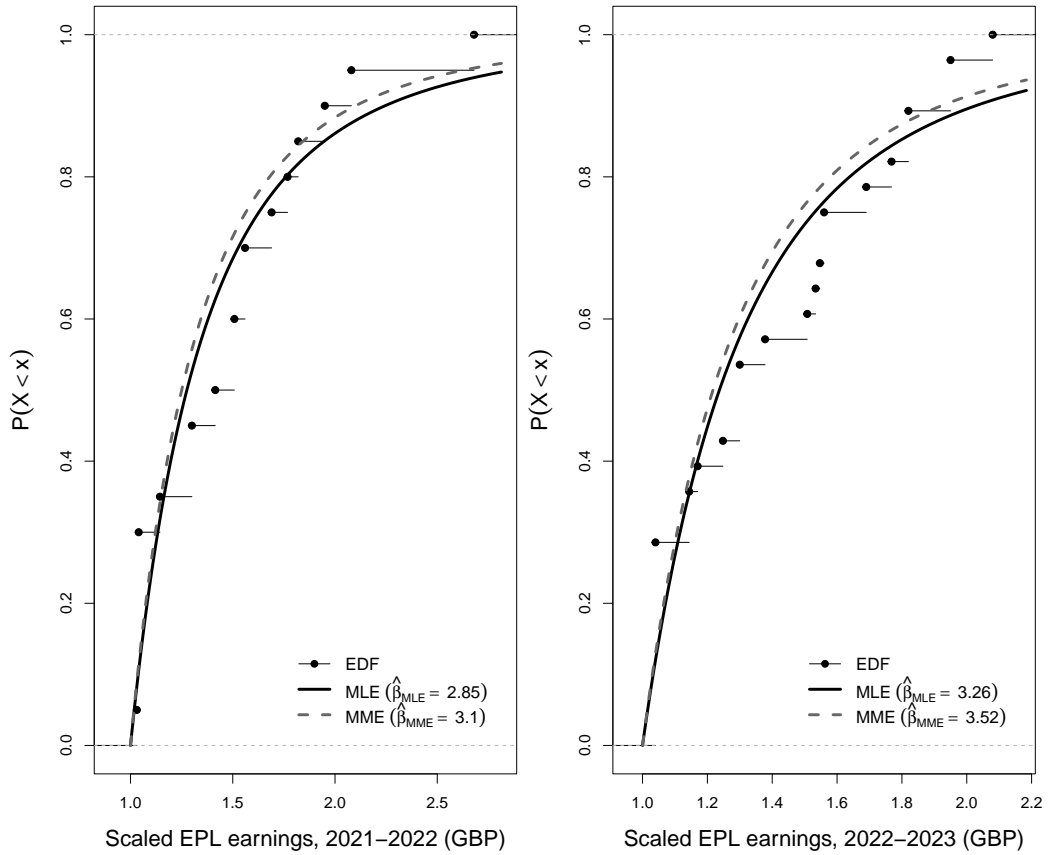


Figure 2: EDF plots of the EPL earnings above 10 million GBP for the two seasons; two theoretical Pareto CDFs are overlaid using MLE and MME for the parameter.

underlying distribution. Again, it should be noted that the salaries associated with 2021–2022 seem to be realised from a distribution with a heavier tail than is the case for the salaries in the 2022–2023 season.

We now formally test the hypothesis that each of the data sets considered are realised from the Pareto distribution. That is, we compute the  $p$ -values for each test considered. Table 3 shows the estimated  $p$ -values obtained by using a parametric bootstrap with  $B = 10\,000$  bootstrap replications, that is, the bootstrap samples are obtained from a Pareto distribution with the parameter estimated using either the MLE or MME.

The  $p$ -values from the first data set show that the majority of the tests do not reject the null hypothesis at a nominal significance level of 10%. With the second data set, the majority of tests,

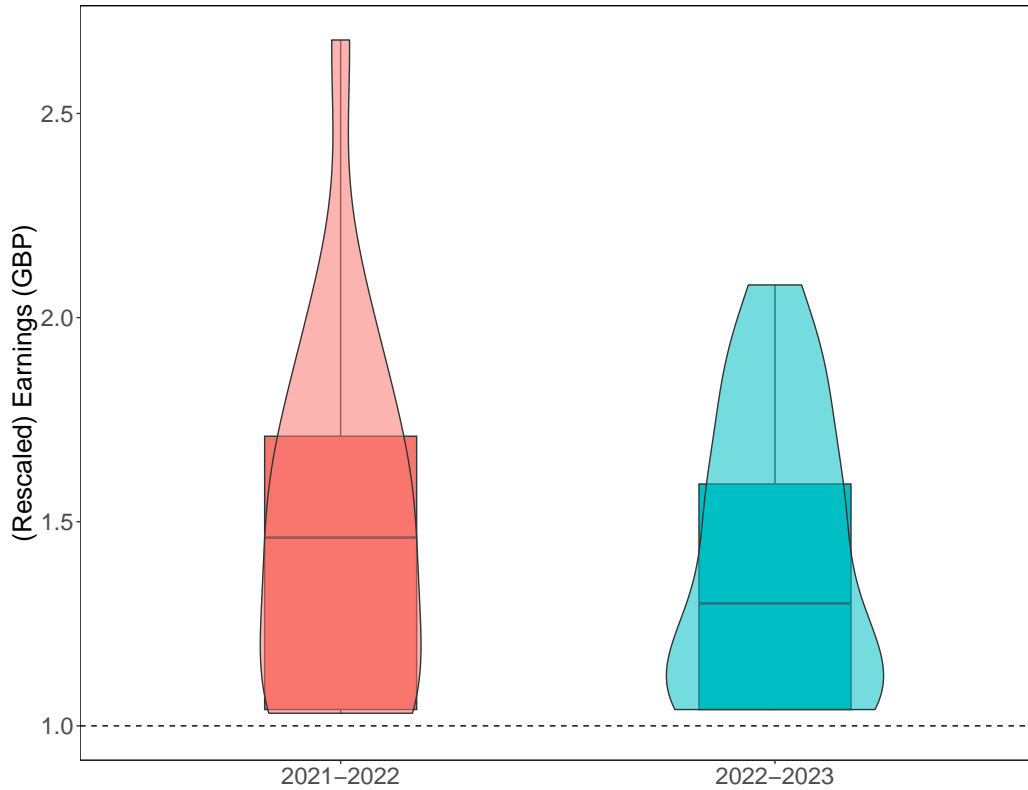


Figure 3: Side-by-side violin plots (with overlaid boxplots) of the EPL earnings above 10 million GBP for the two seasons.

with the exception of  $KS_n$ ,  $G_{n,2}$ ,  $NA_{n,3,2}$ , and  $L_{n,0.1,3}^{[2]}$ , reject the null hypothesis that the data are realised from a Pareto distribution. To summarise these results, we conclude that the Pareto distribution is an appropriate model for the salaries in the 2021–2022 season, but not necessarily for the 2022–2023 season.

## 6 Conclusions

In this paper we proposed and investigated new tests for the hypothesis that observed data are realisations from a Pareto distribution. These tests are based on a characterisation involving a conditional expectation. Using a simulation study, the finite sample power performance of the newly

Table 3:  $p$ -values (multiplied by 100) for testing the Pareto null hypothesis for the earnings above 10 million GBP for the EPL's 2021–2022 and 2022–2023 seasons (highlighted values are significant at a  $\alpha = 0.1$  level).

	2021–2022		2022–2023	
	MLE	MME	MLE	MME
$KS_n$	17	15	16	12
$CV_n$	10	11	9	8
$AD_n$	10	11	6	6
$MA_n$	11	9	5	4
$ZA_n$	20	21	2	2
$G_{n,2}$	41	32	30	20
$NA_{n,3,2}$	35	31	25	21
$L_{n,0.1,3}^{[1]}$	9	9	6	6
$L_{n,0.1,3}^{[2]}$	38	27	32	20
$T_{n,0.1,3}^{[1]}$	9	10	6	6

proposed tests are compared to that of the existing tests under a range of alternative distributions and overall the newly proposed test,  $L_{n,s,a}^{[2]}$ , performs very well when compared to the other tests. From extensive Monte Carlo simulations we recommend using  $a = 3$  and  $s = 0.1$  as this choice produced good powers for the majority of the alternatives considered.

Table 4: The approximated sizes and powers of the tests under MLE estimation,  $n = 20$ .

	$KS_n$	$CV_n$	$AD_n$	$MA_n$	$ZA_n$	$G_{n,2}$	$NA_{n,3,2}$	$L_{n,0.1,3}^{[1]}$	$L_{n,0.1,3}^{[2]}$	$T_{n,0.1,3}^{[1]}$
P(4)	5	5	5	5	5	5	5	5	5	5
P(10)	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.4)$	49	52	<b>77</b>	42	<b>75</b>	48	26	59	0	60
$\Gamma(0.8)$	9	10	10	<b>12</b>	11	9	<b>16</b>	9	11	9
$\Gamma(1.0)$	24	30	25	30	29	32	<b>41</b>	26	<b>36</b>	25
$\Gamma(1.3)$	57	68	64	67	68	73	<b>79</b>	65	<b>77</b>	64
LN(1.0)	56	66	64	54	<b>80</b>	69	72	68	<b>86</b>	67
LN(1.2)	24	30	27	23	<b>43</b>	33	37	31	<b>50</b>	30
LN(1.6)	5	5	4	5	6	5	<b>7</b>	5	<b>8</b>	5
LN(2.5)	26	29	<b>44</b>	24	27	26	13	<b>34</b>	0	<b>34</b>
LFR(0.3)	34	42	36	45	39	44	<b>55</b>	37	<b>46</b>	36
LFR(0.6)	41	50	44	53	47	53	<b>63</b>	44	<b>54</b>	44
LFR(0.9)	44	54	48	56	50	56	<b>67</b>	48	<b>57</b>	48
LFR(1.2)	47	57	52	<b>60</b>	54	<b>60</b>	<b>70</b>	51	<b>60</b>	51
BE(0.5)	18	20	<b>42</b>	16	<b>39</b>	15	10	24	0	25
BE(0.8)	11	12	11	<b>14</b>	12	11	<b>19</b>	10	<b>14</b>	10
BE(1.0)	24	30	25	30	29	32	<b>41</b>	26	<b>36</b>	26
BE(1.5)	69	80	77	79	81	85	<b>89</b>	78	<b>88</b>	77
TP(0.5)	7	8	7	7	9	8	<b>11</b>	8	<b>12</b>	8
TP(1.0)	12	13	11	12	14	13	<b>18</b>	12	<b>19</b>	12
TP(2.0)	22	26	21	23	25	26	<b>34</b>	24	<b>34</b>	24
TP(3.0)	32	38	33	35	36	39	<b>48</b>	36	<b>48</b>	35
LW(0.2)	12	13	11	11	16	13	<b>18</b>	13	<b>22</b>	13
LW(0.4)	30	36	31	30	40	38	<b>44</b>	35	<b>51</b>	35
LW(0.6)	51	62	58	54	66	65	<b>70</b>	61	<b>77</b>	61
LW(0.8)	72	82	80	77	85	86	<b>88</b>	81	<b>92</b>	81
BN(0.2)	9	10	8	9	11	10	<b>14</b>	9	<b>15</b>	9
BN(0.4)	14	16	13	15	17	16	<b>23</b>	15	<b>22</b>	15
BN(0.6)	17	20	17	19	20	21	<b>29</b>	18	<b>27</b>	18
BN(0.8)	20	24	20	23	23	25	<b>33</b>	22	<b>32</b>	22

Table 5: The approximated sizes and powers of the tests under MME estimation,  $n = 20$ .

	$KS_n$	$CV_n$	$AD_n$	$MA_n$	$ZA_n$	$G_{n,2}$	$NA_{n,3,2}$	$L_{n,0.1,3}^{[1]}$	$L_{n,0.1,3}^{[2]}$	$T_{n,0.1,3}^{[1]}$
P(4)	5	6	5	5	5	5	4	5	2	5
P(10)	5	5	5	5	5	5	4	5	2	5
$\Gamma(0.4)$	41	43	<b>72</b>	31	<b>73</b>	34	10	34	0	34
$\Gamma(0.8)$	14	16	15	<b>20</b>	11	16	<b>19</b>	17	15	17
$\Gamma(1.0)$	37	45	42	<b>47</b>	33	<b>47</b>	<b>50</b>	46	39	<b>47</b>
$\Gamma(1.3)$	76	84	83	84	76	<b>87</b>	<b>87</b>	85	79	85
LN(1.0)	71	81	83	74	<b>90</b>	84	<b>86</b>	83	82	85
LN(1.2)	41	49	50	45	<b>58</b>	55	53	51	<b>58</b>	53
LN(1.6)	13	16	17	<b>21</b>	14	17	16	16	<b>26</b>	18
LN(2.5)	11	9	<b>27</b>	26	22	6	<b>36</b>	9	13	10
LFR(0.3)	46	55	52	<b>58</b>	44	<b>58</b>	<b>62</b>	56	46	56
LFR(0.6)	51	60	56	<b>63</b>	47	62	<b>66</b>	62	47	62
LFR(0.9)	57	65	63	<b>69</b>	53	67	<b>70</b>	66	51	66
LFR(1.2)	58	66	63	<b>70</b>	54	69	<b>71</b>	68	50	68
BE(0.5)	13	12	<b>32</b>	10	<b>34</b>	7	3	11	2	10
BE(0.8)	18	20	19	<b>25</b>	14	21	<b>25</b>	22	18	23
BE(1.0)	37	44	41	<b>46</b>	31	45	<b>50</b>	45	38	<b>46</b>
BE(1.5)	84	91	90	90	86	<b>93</b>	<b>93</b>	91	86	92
TP(0.5)	27	32	35	<b>40</b>	30	33	28	33	<b>49</b>	37
TP(1.0)	46	54	58	<b>63</b>	52	55	44	55	<b>71</b>	61
TP(2.0)	76	82	87	<b>88</b>	82	84	67	83	<b>92</b>	<b>88</b>
TP(3.0)	90	94	<b>96</b>	<b>96</b>	93	94	80	94	<b>98</b>	<b>96</b>
LW(0.2)	25	30	30	32	30	<b>34</b>	31	32	<b>38</b>	<b>34</b>
LW(0.4)	50	57	58	57	55	<b>62</b>	59	60	<b>64</b>	<b>62</b>
LW(0.6)	73	82	83	80	80	<b>85</b>	84	84	82	<b>85</b>
LW(0.8)	88	94	94	92	93	<b>95</b>	<b>95</b>	<b>95</b>	92	<b>95</b>
BN(0.2)	18	20	20	<b>23</b>	16	22	22	21	<b>25</b>	<b>23</b>
BN(0.4)	24	28	26	29	22	30	<b>32</b>	29	<b>31</b>	30
BN(0.6)	26	32	29	32	25	<b>34</b>	<b>38</b>	33	31	<b>34</b>
BN(0.8)	32	38	36	38	31	<b>40</b>	<b>44</b>	39	34	<b>40</b>

Table 6: The approximated sizes and powers of the tests under MLE estimation,  $n = 30$ .

	$KS_n$	$CV_n$	$AD_n$	$MA_n$	$ZA_n$	$G_{n,2}$	$NA_{n,3,2}$	$L_{n,0.1,3}^{[1]}$	$L_{n,0.1,3}^{[2]}$	$T_{n,0.1,3}^{[1]}$
P(4)	5	5	5	5	5	5	5	5	5	5
P(10)	5	5	5	5	5	5	5	5	5	5
$\Gamma(0.4)$	66	70	<b>89</b>	60	<b>88</b>	64	49	76	0	76
$\Gamma(0.8)$	12	14	13	<b>17</b>	14	13	<b>20</b>	12	14	12
$\Gamma(1.0)$	35	45	39	47	43	48	<b>56</b>	39	<b>49</b>	39
$\Gamma(1.3)$	76	87	85	88	87	<b>91</b>	<b>93</b>	84	<b>91</b>	83
LN(1.0)	77	86	87	77	<b>96</b>	89	88	88	<b>97</b>	88
LN(1.2)	36	44	44	35	<b>63</b>	48	50	46	<b>68</b>	46
LN(1.6)	6	6	5	5	<b>8</b>	6	<b>8</b>	5	<b>9</b>	6
LN(2.5)	39	43	<b>58</b>	35	40	36	26	48	0	<b>49</b>
LFR(0.3)	49	61	56	<b>65</b>	59	<b>65</b>	<b>73</b>	55	63	54
LFR(0.6)	57	69	64	<b>73</b>	67	<b>73</b>	<b>80</b>	62	69	61
LFR(0.9)	61	74	69	<b>78</b>	71	<b>78</b>	<b>84</b>	67	73	66
LFR(1.2)	66	78	74	<b>81</b>	75	<b>81</b>	<b>86</b>	72	77	71
BE(0.5)	27	30	<b>55</b>	23	<b>53</b>	20	19	35	0	36
BE(0.8)	15	17	15	<b>21</b>	17	16	<b>25</b>	14	16	14
BE(1.0)	35	43	38	47	42	<b>48</b>	<b>55</b>	39	<b>48</b>	38
BE(1.5)	86	94	93	94	95	96	<b>97</b>	93	<b>97</b>	92
TP(0.5)	9	9	8	8	10	9	<b>12</b>	9	<b>14</b>	9
TP(1.0)	15	17	15	16	17	18	<b>23</b>	16	<b>25</b>	16
TP(2.0)	31	37	32	34	34	38	<b>44</b>	35	<b>47</b>	34
TP(3.0)	45	54	49	51	48	57	<b>63</b>	52	<b>64</b>	52
LW(0.2)	16	19	16	16	22	20	<b>24</b>	18	<b>30</b>	18
LW(0.4)	42	51	48	44	56	55	<b>59</b>	51	<b>68</b>	50
LW(0.6)	70	81	80	76	84	86	<b>87</b>	81	<b>91</b>	80
LW(0.8)	89	95	95	93	96	<b>97</b>	<b>97</b>	95	<b>99</b>	95
BN(0.2)	11	12	10	12	13	13	<b>17</b>	11	<b>17</b>	11
BN(0.4)	18	22	18	22	21	24	<b>29</b>	20	<b>29</b>	19
BN(0.6)	25	30	25	29	29	33	<b>38</b>	27	<b>38</b>	27
BN(0.8)	30	36	32	35	34	39	<b>46</b>	33	<b>45</b>	33

Table 7: The approximated sizes and powers of the tests under MME estimation,  $n = 30$ .

	$KS_n$	$CV_n$	$AD_n$	$MA_n$	$ZA_n$	$G_{n,2}$	$NA_{n,3,2}$	$L_{n,0.1,3}^{[1]}$	$L_{n,0.1,3}^{[2]}$	$T_{n,0.1,3}^{[1]}$
P(4)	5	5	5	5	5	5	4	5	1	5
P(10)	5	5	5	5	5	5	4	5	2	5
$\Gamma(0.4)$	56	58	<b>85</b>	45	<b>86</b>	47	27	49	0	49
$\Gamma(0.8)$	19	21	21	<b>27</b>	13	21	<b>26</b>	23	21	24
$\Gamma(1.0)$	53	61	59	<b>65</b>	48	63	<b>65</b>	62	54	63
$\Gamma(1.3)$	90	<b>96</b>	95	<b>96</b>	91	<b>96</b>	<b>96</b>	<b>96</b>	92	<b>96</b>
LN(1.0)	89	95	<b>96</b>	90	<b>98</b>	95	<b>96</b>	95	95	<b>96</b>
LN(1.2)	55	66	68	61	<b>78</b>	70	69	69	<b>74</b>	70
LN(1.6)	16	20	22	<b>26</b>	20	21	18	20	<b>32</b>	23
LN(2.5)	13	10	<b>31</b>	<b>31</b>	<b>31</b>	5	<b>53</b>	9	12	9
LFR(0.3)	65	74	73	<b>78</b>	61	76	<b>77</b>	75	61	74
LFR(0.6)	69	79	76	<b>81</b>	65	79	<b>82</b>	79	65	79
LFR(0.9)	75	82	79	<b>85</b>	70	83	<b>84</b>	82	66	82
LFR(1.2)	76	84	82	<b>86</b>	73	85	<b>86</b>	85	68	85
BE(0.5)	14	13	<b>40</b>	11	<b>44</b>	7	4	11	2	10
BE(0.8)	23	26	25	<b>33</b>	15	26	<b>30</b>	28	24	28
BE(1.0)	52	60	58	<b>65</b>	45	63	<b>66</b>	61	54	62
BE(1.5)	95	98	98	98	97	<b>99</b>	<b>99</b>	98	96	98
TP(0.5)	33	37	42	<b>47</b>	38	40	32	39	<b>60</b>	45
TP(1.0)	58	66	70	<b>73</b>	63	67	50	67	<b>83</b>	<b>73</b>
TP(2.0)	89	93	95	<b>96</b>	93	93	77	93	<b>98</b>	<b>96</b>
TP(3.0)	97	<b>99</b>	<b>99</b>	<b>99</b>	<b>99</b>	<b>99</b>	88	<b>99</b>	<b>100</b>	<b>99</b>
LW(0.2)	33	39	40	39	37	42	39	41	<b>49</b>	<b>43</b>
LW(0.4)	67	75	76	73	74	78	77	78	<b>82</b>	<b>79</b>
LW(0.6)	89	94	94	92	93	<b>95</b>	<b>95</b>	<b>95</b>	<b>95</b>	<b>95</b>
LW(0.8)	98	99	99	99	99	<b>100</b>	99	99	99	<b>100</b>
BN(0.2)	23	28	28	<b>31</b>	22	29	29	29	<b>33</b>	<b>31</b>
BN(0.4)	33	39	38	<b>41</b>	28	<b>41</b>	<b>43</b>	40	37	<b>41</b>
BN(0.6)	37	43	41	44	32	<b>45</b>	<b>49</b>	44	39	<b>45</b>
BN(0.8)	41	47	46	48	38	<b>50</b>	<b>55</b>	49	44	49

## References

- Ahmed, A.-H. (1991). Characterization of beta, binomial, and Poisson distributions. *IEEE Transactions on reliability*, 40(3), 290–295.
- Arnold, B. C. (2015). *Pareto Distributions*. CRC Press, Taylor and Francis Group.
- Beran, R. J. (1977). Minimum Pareto distance estimates for parameter models. *The Annals of Statistics*, 5, 445–463.
- Beran, R. J. (1978). An efficient and robust adaptive estimator of location. *The Annals of Statistics*, 6, 292–313.
- Boos, D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association*, 76, 663–670.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6, 76–90.
- Dallas, A. (1976). Characterizing the Pareto and power distributions. *Annals of the Institute of Statistical Mathematics*, 28, 491–497.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *Computer Journal*, 13(3), 317–322.
- Galambos, J., & Kotz, S. (1978). *Characterizations of Probability Distributions (Lecture Notes in Mathematics 675)*. New York: SpringerVerlag.
- Ghitany, M., El-Saidi, M., & Khalil, Z. (1995). Characterization of a general class of life-testing models. *Journal of Applied Probability*, 32(2), 548–553.
- Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29, 567–589.
- Goldfarb, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computation*, 24(109), 23–26.
- Gupta, R. C. (1975). On characterization of distribution by conditional expectation. *Communications in Statistics – Theory and Methods*, 4(1), 99–103.
- Hamdan, M. (1972). On a characterization by conditional expectations. *Technometrics*, 14(2), 497–499.

- Khan, A., & Abu-Salih, M. (1989). Characterizations of probability distributions by conditional expectation of order statistics. *Metron*, *47*, 171–181.
- Kotlarski, I. I. (1972). On a characterization of some probability distributions by conditional expectations. *Sankhyā: The Indian Journal of Statistics, Series A*, (pp. 461–466).
- Kotz, S., & Shanbhag, D. (1980). Some new approaches to probability distributions. *Advances in Applied Probability*, *12*(4), 903–921.
- Koudou, A. E., & Ley, C. (2014). Characterizations of GIG laws: A survey. *Probability Surveys*, *11*, 161–176.
- Lee, M.-Y. (2003). Characterizations of the Pareto distribution by conditional expectations of record values. *Communications – Korean Mathematical Society*, *18*(1), 127–132.
- Meintanis, S. G. (2009). A unified approach of testing for discrete and continuous Pareto laws. *Statistical Papers*, *50*(3), 569–580.
- Nagaraja, H. (2006). Characterizations of probability distributions. In *Springer Handbook of Engineering Statistics*, (pp. 79–95). Springer.
- Ndwandwe, L., Allison, J. S., Santana, L., & Visagie, I. J. H. (2023a). Testing for the Pareto type I distribution: a comparative study. *Metron*, *81*, 215–256.
- Ndwandwe, L., Allison, J. S., Smuts, M., & Visagie, I. J. H. (2023b). On a new class of tests for the Pareto distribution using Fourier methods. *Stat*, (p. e566).
- Nofal, Z. M., & El Gebaly, Y. M. (2017). New characterizations of the Pareto distribution. *Pakistan Journal of Statistics and Operation Research*, *13*, 63–74.
- Nombebe, T., Allison, J. S., Santana, L., & Visagie, I. J. H. (2023). On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques. *Computation*, *44*(11), <https://doi.org/10.3390/computation11030044>.
- Parr, W. C., & Schucany, W. R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association*, *75*, 616–624.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.  
URL <https://www.R-project.org/>

- Ruiz, J. M., & Navarro, J. (1994). Characterization of distributions by relationships between failure rate and mean residual life. *IEEE Transactions on Reliability*, *43*(4), 640–644.
- Shanbhag, D. (1970). The characterizations for exponential and geometric distributions. *Journal of the American Statistical Association*, *65*(331), 1256–1259.
- Shanbhag, D. N., & Bhaskara Rao, M. (1975). A note on characterisations of probability distributions based on conditional expected values. *Sankhyā: The Indian Journal of Statistics, Series A*, *37*(2), 297–300.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, *24*(111), 647–656.
- Su, J.-C., & Huang, W.-J. (2000). Characterizations based on conditional expectations. *Statistical Papers*, *41*, 423–435.
- www.spotrac.com (2023). EPL salary rankings. (Accessed 2023-09-18).  
URL <https://www.spotrac.com/epl/rankings/year/2021/>
- Wolfowitz, J. (1953). Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, *5*, 9–23.
- Wu, J.-W., & Lee, W.-C. (2001). On the characterization of generalized extreme value, power function, generalized Pareto and classical Pareto distributions by conditional expectation of record values. *Statistical Papers*, *42*, 225–242.
- Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 281–294.
- Zoroa, P., Ruiz, J., & Marin, J. (1990). A characterization based on conditional expectations. *Communications in Statistics – Theory and Methods*, *19*(8), 3127–3135.

## A Appendix

This appendix contains the details about the simplification of the three newly proposed test statistics.

**The first test statistic is:**

$$\begin{aligned}
L_{n,s,a}^{[1]} &= \int_1^\infty \left( \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - \frac{t^s}{n} \sum_{j=1}^n X_j^s t^{-\hat{\beta}} \right)^2 t^{-a} dt. \\
&= \underbrace{\int_1^\infty \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) t^{-a} dt}_{T_1} \\
&\quad - \underbrace{\int_1^\infty \frac{2}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) t^{s-\hat{\beta}} t^{-a} dt}_{T_2} \\
&\quad + \underbrace{\int_1^\infty \frac{t^{2s}}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{-2\hat{\beta}} t^{-a} dt}_{T_3}
\end{aligned}$$

We consider each term separately:

$$\begin{aligned}
T_1 &= \int_1^{\min(X_j, X_k)} \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{-a} dt \\
&= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) \frac{t^{-a+1}}{-a+1} \Big|_1^{\min(X_j, X_k)} \\
&= \frac{1}{n^2(a-1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) [1 - \min(X_j, X_k)^{-a+1}]
\end{aligned}$$

$$\begin{aligned}
T_2 &= \frac{2}{n^2} \int_1^{X_j} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{s-\hat{\beta}-a} dt \\
&= \frac{2}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s \frac{t^{s-\hat{\beta}-a+1}}{s-\hat{\beta}-a+1} \Big|_1^{X_j} \\
&= \frac{2}{n^2(s-\hat{\beta}-a+1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s [(X_j)^{s-\hat{\beta}-a+1} - 1]
\end{aligned}$$

$$\begin{aligned}
T_3 &= \int_1^\infty \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{2s-2\hat{\beta}-a} dt \\
&= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s \left. \frac{t^{2s-2\hat{\beta}-a+1}}{2s-2\hat{\beta}-a+1} \right|_1^\infty \\
&= -\frac{1}{n^2(2s-2\hat{\beta}-a+1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s
\end{aligned}$$

Therefore,

$$\begin{aligned}
L_{n,s,a}^{[1]} &= \sum_{j=1}^n \sum_{k=1}^n \frac{X_j^s X_k^s}{n^2} \left[ \frac{1}{(a-1)} [1 - \min(X_j, X_k)^{-a+1}] \right. \\
&\quad \left. - \frac{2}{(s-\hat{\beta}-a+1)} [X_j^{s-\hat{\beta}-a+1} - 1] \right. \\
&\quad \left. - \frac{1}{(2s-2\hat{\beta}-a+1)} \right]
\end{aligned}$$

The second test statistic is:

$$\begin{aligned}
L_{n,s,a}^{[2]} &= \int_1^\infty \left( \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) - \frac{t^s}{n} \sum_{j=1}^n X_j^s \frac{1}{n} \sum_{j=1}^n I(X_j > t) \right)^2 t^{-a} dt. \\
&= \underbrace{\int_1^\infty \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) t^{-a} dt}_{T_1} \\
&\quad - \underbrace{\int_1^\infty \frac{2t^s}{n^3} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) t^{-a} dt}_{T_2} \\
&\quad + \underbrace{\int_1^\infty \frac{t^{2s}}{n^4} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) t^{-a} dt}_{T_3}
\end{aligned}$$

where

$$\begin{aligned}
T_1 &= \int_1^{\min(X_j, X_k)} \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{-a} dt \\
&= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) \left. \frac{t^{-a+1}}{-a+1} \right|_1^{\min(X_j, X_k)} \\
&= \frac{1}{n^2(a-1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s I(X_j > t) I(X_k > t) [1 - \min(X_j, X_k)^{-a+1}]
\end{aligned}$$

$$\begin{aligned}
T_2 &= \int_1^{\min(X_j, X_k)} \frac{2t^s}{n^3} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{-a} dt \\
&= \frac{2}{n^3} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s \frac{t^{s-a+1}}{s-a+1} \Big|_1^{\min(X_j, X_k)} \\
&= \frac{2}{n^3(s-a+1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s [\min(X_j, X_k)^{s-a+1} - 1]
\end{aligned}$$

$$\begin{aligned}
T_3 &= \int_1^{\min(X_j, X_k)} \frac{1}{n^4} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s t^{2s-a} dt \\
&= \frac{1}{n^4} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s \frac{t^{2s-a+1}}{2s-a+1} \Big|_1^{\min(X_j, X_k)} \\
&= \frac{1}{n^4(2s-a+1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s [\min(X_j, X_k)^{2s-a+1} - 1]
\end{aligned}$$

Combine the terms:

$$\begin{aligned}
L_{n,s,a}^{[2]} &= \sum_{j=1}^n \sum_{k=1}^n \frac{X_j^s X_k^s}{n^2} \left[ \frac{1}{(a-1)} [1 - \min(X_j, X_k)^{-a+1}] \right. \\
&\quad - \frac{2}{n(s-a+1)} [\min(X_j, X_k)^{s-a+1} - 1] \\
&\quad \left. + \frac{1}{n^2(2s-a+1)} [\min(X_k, X_j)^{2s-a+1} - 1] \right].
\end{aligned}$$

**The third test statistic:**

$$T_{n,s,a}^{[1]} = \int_1^\infty \left( \left( \frac{1}{n} \sum_{j=1}^n X_j^s I(X_j > t) \right) - \frac{\hat{\beta} t^{-\hat{\beta}-1}}{\hat{\beta}-s} t^{s+1} \right)^2 t^{-a} dt$$

The first term is the same as with the previous two tests. The middle term is:

$$\begin{aligned}
\int_1^\infty \frac{2}{n} \sum_{j=1}^n X_j^s I(X_j > t) \frac{\hat{\beta}}{\hat{\beta}-s} t^{-\hat{\beta}+s-a} dt &= \int_1^{X_j} \frac{2}{n} \sum_{j=1}^n X_j^s \frac{\hat{\beta}}{\hat{\beta}-s} t^{-\hat{\beta}+s-a} dt \\
&= \frac{2\hat{\beta}}{n} \sum_{j=1}^n X_j^s \frac{t^{s-\hat{\beta}-a+1}}{s-\hat{\beta}-a+1} \Big|_1^{X_j}
\end{aligned}$$

$$= \frac{2\hat{\beta}}{n(\hat{\beta} - s)(s - \hat{\beta} - a + 1)} \sum_{j=1}^n X_j^s \left[ (X_j)^{s - \hat{\beta} - a + 1} - 1 \right]$$

The third term is

$$\begin{aligned} \int_1^\infty \frac{\hat{\beta}^2}{(\hat{\beta} - s)^2} t^{2(-\hat{\beta}-1)} t^{2(s+1)} t^{-a} dt &= \int_1^\infty \frac{\hat{\beta}^2}{(\hat{\beta} - s)^2} t^{2s-2\hat{\beta}-a} dt \\ &= \frac{\hat{\beta}^2}{(\hat{\beta} - s)^2} \frac{t^{2s-2\hat{\beta}-a+1}}{2s-2\hat{\beta}-a+1} \Big|_1^\infty \\ &= -\frac{\hat{\beta}^2}{(\hat{\beta} - s)^2(2s-2\hat{\beta}-a+1)} \end{aligned}$$

Combine the terms:

$$\begin{aligned} T_{n,s,a}^{[1]} &= \frac{1}{n^2(a-1)} \sum_{j=1}^n \sum_{k=1}^n X_j^s X_k^s [1 - \min(X_j, X_k)^{-a+1}] \\ &\quad - \frac{2\hat{\beta}}{n(\hat{\beta} - s)(s - \hat{\beta} - a + 1)} \sum_{j=1}^n X_j^s [(X_j)^{s - \hat{\beta} - a + 1} - 1] \\ &\quad - \frac{\hat{\beta}^2}{(\hat{\beta} - s)^2(2s - 2\hat{\beta} - a + 1)}. \end{aligned}$$

## Chapter 4

# Article 3: On Fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques

The third article, *On fitting the Lomax distribution: A comparison between minimum distance estimators and other estimation techniques* has been published in *Computation*. A summary of the guidelines to authors from the journal is now presented.

---

Manuscript	The article should include the most recent and relevant references in the field. The structure should include an Abstract, Keywords, Introduction, Materials and Methods, Results, Discussion, and Conclusions (optional) sections.
Title	The title of your manuscript should be concise and relevant.
Abstract and keywords	The abstract should be a total of about 200 words maximum. The abstract should be a single paragraph. Three to ten pertinent keywords need to be added after the abstract.
Tables	All table columns should have an explanatory heading. To facilitate the copy-editing of larger tables, smaller fonts may be used, but no less than 8 pt. in size.
References	References must be numbered in order of appearance in the text (including table captions and figure legends) and listed individually at the end of the manuscript.
General formatting	A L <sup>A</sup> T <sub>E</sub> X template is provided for submission.
Additional information	<a href="https://tinyurl.com/y4lu7f9f">https://tinyurl.com/y4lu7f9f</a>

---

Article

# On Fitting the Lomax Distribution: A Comparison between Minimum Distance Estimators and Other Estimation Techniques

Thobeka Nombebe, James Allison, Leonard Santana  and Jaco Visagie \* 

School of Mathematical and Statistical Sciences, North-West University, Potchefstroom 2520, South Africa

\* Correspondence: jaco.visagie@nwu.ac.za

**Abstract:** In this paper, we investigate the performance of a variety of frequentist estimation techniques for the scale and shape parameters of the Lomax distribution. These methods include traditional methods such as the maximum likelihood estimator and the method of moments estimator. A version of the maximum likelihood estimator adjusted for bias is included as well. Furthermore, an alternative moment-based estimation technique, the *L*-moment estimator, is included, along with three different minimum distance estimators. The finite sample performances of each of these estimators are compared in an extensive Monte Carlo study. We find that no single estimator outperforms its competitors uniformly. We recommend one of the minimum distance estimators for use with smaller samples, while a bias-reduced version of maximum likelihood estimation is recommended for use with larger samples. In addition, the desirable asymptotic properties of traditional maximum likelihood estimators make them appealing for larger samples. We include a practical application demonstrating the use of the described techniques on observed data.

**Keywords:** Lomax distribution; estimation; minimum distance estimators



**Citation:** Nombebe, T.; Allison, J.; Santana, L.; Visagie, J. On Fitting the Lomax Distribution: A Comparison between Minimum Distance Estimators and Other Estimation Techniques. *Computation* **2023**, *11*, 44. <https://doi.org/10.3390/computation11030044>

Received: 27 November 2022

Revised: 15 February 2023

Accepted: 16 February 2023

Published: 23 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Pareto distribution is a heavy-tailed distribution that was originally developed in the 19th century to model the distribution of income among individuals [1]. However, in the years following its introduction it has been extensively modified and changed to produce several variants, referred to as the Type I, II, III, and IV Pareto distributions, as well as the so-called generalised Pareto distribution (GPD). The focus of this paper is on the Pareto Type II distribution, in which the location parameter is zero, which is known as the Lomax distribution. This rather popular distribution was originally introduced to model business failure data [2], and has subsequently been employed as a model in a wide range of settings. This distribution has proven useful in modelling survival times when censoring is present, e.g., in the lifetimes of electric power transformers [3]; as the times are left-truncated and right-censored, when they are modelled it is found that the Lomax distribution provides the best fit among the various distributions considered. As another example, the exponential Lomax and Weibull Lomax are a good fit for right-censored lung cancer survival data [4].

Other scenarios in which the Lomax distribution has been used as a model include wealth distribution [5], the distribution of queueing service times [6], life testing [7], and the sizes of files on a computer server [8], to name only a few.

We can say that a random variable  $X$  follows a Lomax distribution with scale parameter  $\sigma > 0$  and shape parameter  $\beta > 0$  if its cumulative distribution function (CDF) is

$$F_{\sigma,\beta}(x) = 1 - \left[1 + \left(\frac{x}{\sigma}\right)\right]^{-\beta}, \quad x > 0,$$

with a probability density function (PDF) of

$$f_{\sigma,\beta}(x) = \frac{\beta}{\sigma} \left(1 + \frac{x}{\sigma}\right)^{-(\beta+1)}, \quad x > 0.$$

In this paper, we investigate different approaches used for estimating the parameters of the Lomax distribution. We draw parallels with the methods used to estimate the parameters of other distributions in order to ascertain how one might go about obtaining reliable estimators. For example, when considering the estimation of the parameters of the GPD, a number of approaches have been studied, such as maximum likelihood estimation (MLE), method of moment estimation (MME),  $L$ -moment estimation (LME), and probability-weighted moment estimation (PWME). However, not all approaches are equally viable in all settings. It was noted in [9], for example, that MLE experiences difficulty when estimating the parameters of the GPD under a variety of configurations of the shape and scale parameter values. Similarly, ref. [10] found that the usefulness of MLE is restricted to very large sample sizes; instead, the use of PWME is advocated in case of moderate sample sizes. Several studies similar to the one presented in this paper have been conducted for other distributions; to mention only a few examples, ref. [11] considered improved estimators for the generalised gamma distribution using small samples, whereas [12,13] compared a variety of different estimators for the parameters of the unit-gamma and Weibull distributions, respectively. For the Lomax distribution in particular, studies include [14,15]. These papers are concerned with the estimation of Lomax parameters using diverse traditional methods of estimation, including MLE, MME, and PWME. The results of these methods have been compared to one another and recommendations have been made concerning the use of these more traditional estimation approaches. From previous papers, it is clear that MLE approaches are not particularly reliable when used with moderately large samples. The conclusion from these studies is that though traditional approaches have utility in certain settings, none are generally applicable.

The aim of this paper is twofold. First, we provide an overview of estimation techniques used for the Lomax distribution and compare the performance of various estimators. Second, we focus on the minimum distance estimators (MDEs), which represent an unexplored avenue for estimating the parameters of the Lomax distribution. While a number of these procedures have already been examined in [16] for the GPD, focusing on the Lomax distribution can provide specific insight into the usefulness of these methods for this particular distribution. In addition, the MDEs we consider here differ substantially from those proposed in [16,17] in that their paper only considered density power divergence measures, whereas we additionally explore distribution function-based methods.

Generally speaking, MDEs measure the difference between an empirical estimate of the probability density or distribution function and the theoretical function [18]. The use of minimum distance estimation measures has been advocated by a number of authors, including [19–21], when modelling data containing extreme values, as these estimators have good robustness qualities.

This paper studies the general case in which both the scale ( $\sigma$ ) and shape ( $\beta$ ) parameters of the Lomax distribution are unknown. We start by considering the myriad different traditional estimation methods proposed for related distributions, including the  $L$ -moment estimator, the MLE, and the MME. We then proceed to propose the use of MDEs in the hope that these may represent competitive alternatives to the more traditional estimators. Although we do not consider any Bayesian estimators in this study, we refer interested readers to [22] and the very recent paper by [23], where Bayesian estimators for the Lomax distribution are discussed and compared, as well as to [24], where Bayesian estimation of the two-parameter gamma distribution is discussed. The performance of the estimators are compared to one another through a comprehensive numerical study involving Monte Carlo simulations in which the finite sample variance, relative bias, and mean squared error (MSE) of each estimator are approximated. In addition, an omnibus measure allowing

the MSE of the estimation of both  $\sigma$  and  $\beta$  to be gauged simultaneously is employed in the simulation study.

The remainder of this article is organised as follows. In Section 2, the different parameter estimation methods are introduced and discussed. Next, the Monte Carlo study is presented in Section 3, including a detailed discussion of the simulation results. In Section 4, we apply the different estimators to a real world example. Finally, in Section 5, we conclude the paper by making a number of general observations regarding the estimators and provide recommendations concerning their use.

## 2. Estimation of Parameters

In this section, we discuss the parameter estimation methods used to estimate the shape and scale parameters of the Lomax distribution. Included here are traditional methods of parameter estimation as well as methods based on optimising a variety of distance measures. In addition to the performance of the MDEs, we provide an overview of the performance of a wide variety of competing estimators. This facilitates recommendations regarding the choice of estimator to be used in various settings. Theoretical details relevant to the estimation of the parameters are provided in each case.

For the remainder of the section, we assume that we have data  $X_1, X_2, \dots, X_n$  which are independently and identically distributed (i.i.d.) from the Lomax distribution with parameters  $\beta > 0$  and  $\sigma > 0$ . The order statistics based on this sample are denoted using  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ .

### 2.1. Method of Moment Estimators (MMES)

MME is one of the most widely used estimation techniques. It has been used to investigate key aspects of probability distributions, such as central tendency, spread, skewness, and kurtosis. However, for the Lomax distribution, due to the fact that with shape parameter  $\beta$  only moments lower than  $\beta$  are finite, this approach is often not ideal for the parameters of the Lomax distribution. It is ultimately presented here and in the Monte Carlo in Section 3.1 in order to illustrate its functionality for a range of values of  $\beta$ .

To derive the MMEs for the Lomax distribution, we can note that the  $r^{\text{th}}$  theoretical raw moment of the distribution is

$$E(X^r) = \mu_r = \frac{\sigma^r \Gamma(\beta - r) \Gamma(1 + r)}{\Gamma(\beta)}, \quad 0 < r < \beta \quad r = 1, 2, \dots$$

The first and second theoretical raw moments are then provided by

$$E(X) = \mu_1 = \frac{\sigma}{\beta - 1}, \quad \text{and} \quad E(X^2) = \mu_2 = \frac{2\sigma^2}{(\beta - 1)(\beta - 2)}. \tag{1}$$

From (1), we can express  $\beta$  and  $\sigma$  as

$$\beta = \frac{2\mu_2 - 2\mu_1^2}{\mu_2 - 2\mu_1^2} \quad \text{and} \quad \sigma = \mu_1(\beta - 1). \tag{2}$$

By simply substituting  $\mu_1$  and  $\mu_2$  with their sample estimators in (2), the MMEs of these parameters are obtained:

$$\hat{\beta}_{MM} = \frac{2\hat{\mu}_2 - 2\hat{\mu}_1^2}{\hat{\mu}_2 - 2\hat{\mu}_1^2} \quad \text{and} \quad \hat{\sigma}_{MM} = \hat{\mu}_1(\hat{\beta}_{MM} - 1),$$

where  $\hat{\mu}_r = \frac{1}{n} \sum_{i=1}^n X_i^r, \quad r = 1, 2.$

### 2.2. L-Moment Estimators (LMEs)

L-moments provide an alternative way to describe the shapes of probability distributions, and are defined as the expected values of linear combinations of order statistics. They were proposed in [25] as robust alternatives to classical moments. The method of L-moments has advantages over classical moments, including unbiasedness, robustness to the presence of outliers in the data, and lower sensitivity to sampling variability, especially when the data of interest are heavy-tailed, as is the case in our context. If the mean of the distribution exists, it follows that all of the L-moments exist and uniquely define the distribution [25]. The  $r^{\text{th}}$  population L-moment is provided by

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1, 2, \dots,$$

where  $r$  is the integer order of the L-moment and  $E(X_{r-k:r})$  is the expectation of the  $(r - k)^{\text{th}}$  order statistic of a sample of size  $r$ . The sample L-moment is provided by

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1 < \dots < i_r \leq n} \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{i_r-k:n} \quad r = 1, 2, \dots, n.$$

Therefore, the expressions of the first two sample L-moments are

$$l_1 = \frac{1}{n} \sum_{i=1}^n X_{i:n} = \bar{X} \quad \text{and} \quad l_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_{i>j} (X_{i:n} - X_{j:n}),$$

where  $\bar{X}$  is the sample mean.

Similar to the method of moments, the L-moments can be used to obtain parameter estimates by equating the theoretical L-moments of the distribution to the corresponding L-moments of a sample. Therefore, the L-moment estimators of the parameters of the Lomax distribution are provided by

$$\hat{\beta}_{LM} = \frac{l_2}{2l_2 - l_1} \quad \text{and} \quad \hat{\sigma}_{LM} = \frac{l_1^2 - l_1 l_2}{2l_2 - l_1};$$

see [15]. L-moments can be a good starting point for the iterative numerical procedure needed to obtain maximum likelihood estimates [26].

**Remark 1.** Note that the probability-weighted moment estimators (PWMEs) proposed in [27] were considered for this paper as well; however, they were omitted because they turn out to simply be linear combinations of the LMEs [25,28].

### 2.3. Maximum Likelihood Estimators (MLEs)

Maximum likelihood estimation is one of the most popular methods for estimating the unknown parameters of probability distributions. The popularity of MLE can be attributable to its desirable asymptotic properties, such as unbiasedness and consistency. However, for small sample sizes, these properties may not hold true, resulting in biased MLEs.

In this section, we first use the MLEs to estimate the unknown parameters; then, in the section that follows, we consider a bias-adjusted approach which reduces the bias of the MLEs to order  $O(n^{-1})$ ; see [14].

### 2.3.1. MLE of the Lomax Distribution

The MLEs for the Lomax distribution are obtained by maximising the log-likelihood function:

$$\ell := \ell(\beta, \sigma; X_1, \dots, X_n) = n \log(\beta) - n \log(\sigma) - (1 + \beta) \sum_{i=1}^n \log\left(1 + \frac{X_i}{\sigma}\right). \tag{3}$$

Differentiating the log-likelihood with respect to  $\beta$  and  $\sigma$ , respectively, yields the equations

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n \log\left(1 + \frac{X_i}{\sigma}\right), \tag{4}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1 + \beta}{\sigma} \sum_{i=1}^n \frac{X_i}{\sigma + X_i}. \tag{5}$$

Unfortunately, we cannot obtain closed-form solutions for both estimators using (4) and (5); thus, numerical procedures are required for optimisation purposes. To simplify the optimisation problem, the expressions can be reduced to the optimisation of a single parameter using the concentrated likelihood estimation approach (the approach followed in [29] using their `f1omax` function). This is accomplished by first noting that when setting (4) to zero and solving, we obtain the MLE  $\hat{\beta}_{ML}$  expressed in terms of  $\hat{\sigma}_{ML}$ :

$$\hat{\beta}_{ML} = \frac{n}{\sum_{i=1}^n \log(1 + X_i/\hat{\sigma}_{ML})}. \tag{6}$$

The maximised log-likelihood  $\hat{\ell} := \ell(\hat{\beta}_{ML}, \hat{\sigma}_{ML}; X_1, \dots, X_n)$  can therefore be expressed solely in terms of the MLE  $\hat{\sigma}_{ML}$  by substituting the form of  $\hat{\beta}_{ML}$  provided in (6) and  $\hat{\sigma}_{ML}$  into (3):

$$\hat{\ell} = -n \log\left(\frac{1}{n} \sum_{i=1}^n \log\left(1 + \frac{X_i}{\hat{\sigma}_{ML}}\right)\right) - n \log(\hat{\sigma}_{ML}) - n \left(\frac{1}{n} \sum_{i=1}^n \log\left(1 + \frac{X_i}{\hat{\sigma}_{ML}}\right) + 1\right).$$

Then, using R’s one-dimensional `optimize` [30] function, the MLEs  $\hat{\sigma}_{ML}$  and  $\hat{\beta}_{ML}$  (via (6)) are obtained.

### 2.3.2. MLEs of the Lomax Distribution Adjusted for Bias (MLE.b)

For small sample sizes, the MLEs sometimes perform poorly. In order to solve this problem, [14] proposed the use of second-order bias-adjusted versions of the MLEs. To obtain these estimators, the bias of the MLE is first approximated and then subtracted from the MLE to produce the bias-corrected estimate.

Let  $K$  be the Fisher information matrix for the Lomax distribution, as provided in [14]. Furthermore, it can be shown that

$$K^{-1} = \begin{bmatrix} \frac{\sigma^2}{n\beta}(\beta + 2)(\beta + 1)^2 & \frac{\sigma\beta}{n}(\beta + 1)(\beta + 2) \\ \frac{\sigma\beta}{n}(\beta + 1)(\beta + 2) & \frac{\beta^2}{n}(\beta + 1)^2 \end{bmatrix}.$$

The  $O(n^{-1})$  bias is obtained in [14] as

$$\text{Bias}\begin{pmatrix} \hat{\sigma}_{ML} \\ \hat{\beta}_{ML} \end{pmatrix} = K^{-1} \text{Avec}(K^{-1}) + O(n^{-2}),$$

where  $\text{vec}(\cdot)$  is an operator that, when applied to a  $(m \times n)$  matrix, produces a single column vector of dimension  $(mn \times 1)$  by piling up the column vectors below one another; here, the matrix  $A$  is provided by

$$A = n \begin{bmatrix} \frac{2\beta}{\sigma^3(\beta+2)(\beta+3)} & \frac{-1}{\sigma^2(\beta+1)(\beta+2)} & \frac{\beta}{\sigma^2(\beta+2)^2} & \frac{-1}{\sigma(\beta+1)^2} \\ \frac{-1}{\sigma^2(\beta+1)(\beta+2)} & 0 & \frac{-1}{\sigma(\beta+1)^2} & \frac{1}{\beta^3} \end{bmatrix}.$$

The bias-adjusted estimators proposed in [14] are then obtained by subtracting an estimator of this approximated bias from the original MLEs:

$$\begin{pmatrix} \hat{\sigma}_{MLB} \\ \hat{\beta}_{MLB} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{ML} \\ \hat{\beta}_{ML} \end{pmatrix} - \widehat{Bias} \begin{pmatrix} \hat{\sigma}_{ML} \\ \hat{\beta}_{ML} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{ML} \\ \hat{\beta}_{ML} \end{pmatrix} - \hat{K}^{-1} \widehat{A} \text{vec}(\hat{K}^{-1}), \tag{7}$$

where

$$\hat{K} = K|_{\sigma=\hat{\sigma}_{ML};\beta=\hat{\beta}_{ML}} \quad \text{and} \quad \hat{A} = A|_{\sigma=\hat{\sigma}_{ML};\beta=\hat{\beta}_{ML}}.$$

The bias-adjusted estimators in (7) have simple closed-form expressions, making them computationally appealing.

**Remark 2.** Note that a robust alternative to maximum likelihood for parameter estimation of univariate continuous distributions is the maximum product of spacings estimator (MPSE) proposed by [31,32]. As this approach essentially employs a discretised analogue of the log-likelihood, maximisation yields similar results to the MLE, and for this reason is omitted from our simulations.

#### 2.4. Minimum Distance Estimators (MDEs) Based on Distribution Functions

MDEs are estimators that minimise (for example) some distance measure between the theoretical CDF and the empirical distribution function (EDF) of the observed sample data. Note that we define the EDF as

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n I(X_{j:n} \leq x),$$

where  $I(\cdot)$  is the indicator function and  $n$  is the sample size. The main idea is to determine how close the theoretical CDF is to the EDF; the distance between two probability distributions captures the difference in information between them. Minimum distance estimation was first subjected to in-depth study in a series of papers discussed in [33], and has subsequently been considered a good method for deriving efficient and robust estimators in [19,34,35] as well as in [20], among others.

Several minimum distance estimation methods have been proposed, including those derived from empirical distribution functions, such as Cramér–von Mises, and those derived from entropy, such as Kullback–Liebler divergence. We discuss these estimators in two separate sections; the first is based on distance measures between distribution functions, while the second considers distances between density functions. Below, we consider the Cramér–von Mises (CvM) and squared difference (SD) statistics.

In Section 3, we calculate these MDEs using the `optim` function in R; specifically we use the “BFGS” optimisation method (see [36–39]). This is a quasi-Newton method reliant on gradients to perform optimisation. We use the LMEs as initial values, as they are available in closed-form and generally perform well compared to other estimators.

##### 2.4.1. A Cramér–Von Mises Distance Measure (MDE.CvM)

The Cramér–von Mises (CvM) statistic is based on the squared integral difference between the EDF and the theoretical distribution. It has the form

$$Q_n = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x).$$

We can define the CvM distance measure used to estimate the parameters of the Lomax distribution using the MDE approach as follows:

$$D_n^{CM}(\sigma, \beta) = \int_0^\infty [F_n(x) - F_{\sigma,\beta}(x)]^2 f_{\sigma,\beta}(x) dx,$$

where  $F_{\sigma,\beta}(x)$  and  $f_{\sigma,\beta}(x)$  are the CDF and PDF of the Lomax distribution with scale parameter  $\sigma$  and shape parameter  $\beta$ , respectively. This distance measure permits a simple calculable form provided by

$$\begin{aligned} D_n^{CM}(\sigma, \beta) &= \frac{1}{12n} + \sum_{j=1}^n \left( F_{\sigma,\beta}(X_{j:n}) - \frac{2j-1}{2n} \right)^2 \\ &= \frac{1}{12n} + \sum_{j=1}^n \left( 1 - \left( \frac{\sigma}{X_{j:n} + \sigma} \right)^\beta - \frac{2j-1}{2n} \right)^2. \end{aligned}$$

Finally, the resulting estimators for  $\sigma$  and  $\beta$  can be expressed simply as the values of those parameters that minimise this distance measure:

$$(\hat{\sigma}_{CM}, \hat{\beta}_{CM}) = \arg \min_{(\sigma,\beta)} D_n^{CM}(\sigma, \beta).$$

#### 2.4.2. A ‘Squared Difference’ Distance Estimator (MDE.SD)

An alternative to the Cramér–von Mises distance measure is the much simpler ‘squared difference’ measure. This distance measure focuses only on the squared difference between the quantities  $F_n$  and  $F$ , and has the following general expression:

$$S_n = \int_{-\infty}^\infty [F_n(x) - F(x)]^2 dF_n(x).$$

Now, using the Lomax distribution function  $F_{\sigma,\beta}(\cdot)$ , the distance measure can be defined as

$$D_n^{SD}(\sigma, \beta) = \int_0^\infty [F_n(x) - F_{\sigma,\beta}(x)]^2 dF_n(x),$$

where, upon simplification, we obtain the following tractable calculation form:

$$\begin{aligned} D_n^{SD}(\sigma, \beta) &= \sum_{j=1}^n \left( F_{\sigma,\beta}(X_{j:n}) - \frac{j}{n+1} \right)^2 \\ &= \sum_{j=1}^n \left( 1 - \left( \frac{\sigma}{X_{j:n} + \sigma} \right)^\beta - \frac{j}{n+1} \right)^2. \end{aligned}$$

Therefore, the estimators obtained by minimising this ‘squared difference’ distance measure can be expressed as

$$(\hat{\sigma}_{SD}, \hat{\beta}_{SD}) = \arg \min_{(\sigma,\beta)} D_n^{SD}(\sigma, \beta).$$

#### 2.5. Minimum Distance Estimators (MDEs) Based on Density Functions ( $\phi$ -Divergence Distance Measures)

The estimators discussed below are similar to those discussed in Section 2.4, the principal difference being that the distance to be minimised is between density functions instead of distribution functions. The resulting estimators are often referred to as  $\phi$ -divergence distance measures, and represent a broad class of distance measures that describe the distance between two densities  $f$  and  $g$ , defined as

$$\delta(f, g) = E \left[ \phi \left( \frac{f(X)}{g(X)} \right) \frac{g(X)}{f(X)} \right], \tag{8}$$

where  $X$  is a random variable from a distribution function with density  $f$  and  $\phi(\cdot)$  is a convex function such that  $\phi(1) = 0$  and  $\left. \frac{\partial^2 \phi(x)}{\partial x^2} \right|_{x=1} = \phi''(1) > 0$ . In (8), we set  $g(x)$  to be the Lomax density function,  $f_{\sigma,\beta}(x)$ , and set  $f(x)$  to be some empirical estimator for the density,  $\hat{f}(x)$ , to obtain the following distance measure to be minimised in order to estimate the parameters  $\sigma$  and  $\beta$ :

$$D_n^\phi(\sigma, \beta) = \frac{1}{n} \sum_{j=1}^n \phi \left( \frac{\hat{f}(X_j)}{f_{\sigma,\beta}(X_j)} \right) \frac{f_{\sigma,\beta}(X_j)}{\hat{f}(X_j)}.$$

The required parameters estimates are

$$(\hat{\sigma}, \hat{\beta}) = \arg \min_{(\sigma,\beta)} D_n^\phi(\sigma, \beta).$$

In this setting, we estimate the density using the kernel density estimator  $\hat{f}_h(x)$ , defined as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n k \left( \frac{x - X_i}{h} \right),$$

where  $k(\cdot)$  is the kernel function chosen to be the standard normal density function and  $h$  is the bandwidth, chosen using Silverman’s rule-of-thumb ( $h = 0.9n^{-1/5} \min(s, (q_3 - q_1)/1.34)$ , where  $q_1$  and  $q_3$  denote the sample quartiles and  $s$  denotes the sample standard deviation [40]). The practical implementation of the kernel density estimator is executed using the density function in R [30].

Different choices of the function  $\phi(\cdot)$  lead to different forms of the distance measure described above. Here, we consider three choices of  $\phi(\cdot)$  resulting in the Kullback–Liebler (KL) divergence, chi-square ( $\chi^2$ ) divergence, and total variation (TV) distance. These choices are described below.

### 2.5.1. Kullback–Liebler $\phi$ -Divergence Distance Measure (MDE.KL)

The Kullback–Liebler  $\phi$ -divergence distance measure can be obtained by setting

$$\phi(t) = t \log(t).$$

The resulting distance measure is

$$D_n^{KL}(\sigma, \beta) = \frac{1}{n} \sum_{j=1}^n \log \left( \frac{\hat{f}_h(X_j)}{f_{\sigma,\beta}(X_j)} \right),$$

and we denote the estimator by  $(\hat{\sigma}_{KL}, \hat{\beta}_{KL})$  [17]; note that this Kullback–Liebler form simply represents a robust extension of MLE. This approach is followed in [16], where a general class of density power divergent estimators are employed to estimate the parameters of the GPD.

### 2.5.2. Chi-Squared $\phi$ -Divergence Distance Measure (MDE. $\chi^2$ )

Setting  $\phi(t) = (t - 1)^2$  yields the chi-square  $\phi$ -divergence distance measure, denoted as

$$D_n^{CS}(\sigma, \beta) = \frac{1}{n} \sum_{j=1}^n \frac{(\hat{f}_h(X_j) - f_{\sigma,\beta}(X_j))^2}{f_{\sigma,\beta}(X_j) \hat{f}_h(X_j)},$$

and we can express the resulting estimator as  $(\hat{\sigma}_{CS}, \hat{\beta}_{CS})$

### 2.5.3. Total Variation $\phi$ -Divergence Distance Measure (MDE.TV)

When specifying  $\phi(t) = |t - 1|$ , we obtain the total variation  $\phi$ -divergence distance measure, provided by

$$D_n^{TV}(\sigma, \beta) = \frac{1}{n} \sum_{j=1}^n \left| 1 - \frac{f_{\sigma, \beta}(X_j)}{\hat{f}_h(X_j)} \right|,$$

and we can designate the resulting estimator by  $(\hat{\sigma}_{TV}, \hat{\beta}_{TV})$ .

## 3. Finite Sample Results

Below, we consider the Monte Carlo settings before turning our attention to the results.

### 3.1. Monte Carlo Simulation Settings

In order to compare the performance of the various estimators to one another, a comprehensive Monte Carlo simulation study was conducted. The simulation study sets out to approximate several distributional properties of the estimators, including the expected value, variance, mean squared error (MSE), and relative bias (RB), which we define to be the calculated bias divided by the true parameter value. Using these various approximations, it is possible to compare the estimators to one another in a sensible manner.

The Monte Carlo simulation was conducted by simulating  $MC = 150,000$  samples of size  $n \in \{30, 50, 100, 200, 500\}$  from the Lomax distribution using a variety of parameter settings. For the  $\sigma$  parameter, the values used included  $\sigma \in \{1, 2\}$ , whereas for the  $\beta$  parameter we used  $\beta \in \{1.1, 1.5, 2, 2.1\}$ . We considered only values of  $\beta$  exceeding 1, as the mean of the distribution is infinite if  $\beta \leq 1$ ; see (1). Furthermore, the value of  $\beta = 2.1$  is included in order to consider the case in which variance exists.

For each randomly generated sample, the parameters were estimated and collected to ultimately obtain the approximations of the distributional properties mentioned. All calculations were performed using R v4.2.2 [30].

In addition to calculating the MSE for each parameter separately, a combined MSE yielding a single value was calculated as follows: for every pair of estimates, define

$$\zeta_j = \left\| \begin{bmatrix} \hat{\beta}_j \\ \hat{\sigma}_j \end{bmatrix} - \begin{bmatrix} \beta \\ \sigma \end{bmatrix} \right\|^2 = (\hat{\beta}_j - \beta)^2 + (\hat{\sigma}_j - \sigma)^2,$$

then calculate  $TMSE = \frac{1}{MC} \sum_{j=1}^{MC} \zeta_j$  as the measure of the total mean squared error (TMSE) of the estimation technique.

Tables 1–4 present the Monte Carlo approximations of the expected value, variance, RB, and MSE for all estimators of  $\sigma$  and  $\beta$  as well as the TMSE described above for sample sizes  $n = 30$  and  $50$ . The remaining tables for samples sizes  $n = 100, 200$ , and  $500$  are presented in the Appendix A Tables A1–A6.

**Table 1.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 1$  and  $n = 30$ ).

$n = 30$																		
Method	$\beta = 1.1$				$\sigma = 1$				$\sigma = 1$				TMSE					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	1.631	(1.388)	242.871	(0.455)	48.245	(26.164)	243.151	(0.538)	1.859	(1.432)	1301.721	(1.091)	85.949	(43.278)	1302.451	(1.278)	1545.602	(1.816)
MLEb	1.356	(1.111)	243.050	(0.501)	23.297	(0.964)	243.114	(0.501)	1.409	(0.978)	1302.131	(1.116)	40.923	(-2.203)	1302.290	(1.116)	1545.404	(1.617)
LME	1.752	(1.698)	1.443	(0.249)	59.261	(54.331)	1.867	(0.606)	1.965	(1.883)	3.538	(0.736)	98.479	(88.254)	4.489	(1.515)	6.356	(2.122)
MDE.CvM	1.289	(1.221)	1.638	(0.283)	17.165	(11.008)	1.673	(0.298)	1.282	(1.180)	3.860	(0.605)	28.231	(18.032)	3.940	(0.638)	5.613	(0.936)
MDE.SD	1.067	(1.021)	1.277	(0.140)	-3.044	(-7.145)	1.278	(0.146)	0.989	(0.919)	3.068	(0.304)	-1.133	(-8.109)	3.068	(0.310)	4.346	(0.457)
MDE. $\alpha^2$	1.400	(0.658)	142.589	(64.907)	27.266	(-40.158)	142.678	(65.102)	5.828	(3.689)	709.410	(10.567)	482.831	(268.909)	732.718	(17.798)	875.395	(82.901)
MDE.TV	4.310	(3.088)	308.406	(162.076)	291.837	(180.755)	318.709	(166.028)	13.234	(8.817)	3050.295	(1054.581)	1223.385	(781.723)	3199.942	(1115.683)	3518.651	(1281.711)
MDE.KL	1.538	(1.388)	7.991	(0.455)	39.804	(26.165)	7.783	(0.538)	1.663	(1.433)	21.395	(1.091)	66.250	(43.278)	21.833	(1.278)	29.616	(1.816)
MME	3.566	(3.183)	402.398	(1.534)	224.178	(189.368)	408.476	(5.873)	7.944	(6.573)	73,798.684	(8.391)	694.391	(57.345)	73,806.410	(39.454)	74,214.887	(45.327)
$\beta = 1.5$																		
Method	$\beta = 1.5$				$\sigma = 1$				$\sigma = 1$				TMSE					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	2.576	(1.904)	1328.191	(1.713)	71.749	(26.911)	1329.341	(1.876)	2.124	(1.446)	1248.991	(1.958)	112.409	(44.569)	1250.247	(2.157)	2579.587	(4.033)
MLEb	2.292	(1.617)	1328.862	(2.004)	52.818	(7.789)	1329.481	(2.017)	1.822	(1.140)	1249.680	(2.238)	82.176	(14.030)	1250.347	(2.238)	2579.828	(4.275)
LME	2.119	(2.020)	4.407	(0.555)	41.261	(34.635)	4.790	(0.825)	1.606	(1.503)	4.733	(0.686)	60.632	(50.340)	5.101	(0.939)	9.891	(1.764)
MDE.CvM	1.645	(1.527)	4.824	(0.618)	9.677	(1.796)	4.845	(0.619)	1.180	(1.056)	5.088	(0.647)	18.004	(5.774)	5.120	(0.650)	9.965	(1.269)
MDE.SD	1.319	(1.233)	4.058	(0.252)	-12.076	(-17.833)	4.091	(0.323)	0.884	(0.793)	4.285	(0.266)	-11.649	(-20.725)	4.298	(0.309)	8.389	(0.632)
MDE. $\alpha^2$	3.691	(2.407)	249.441	(86.961)	146.080	(60.489)	254.241	(87.784)	7.321	(4.588)	984.263	(221.195)	632.149	(38.751)	1024.217	(234.064)	1278.458	(321.848)
MDE.TV	8.985	(7.700)	779.077	(621.789)	498.994	(413.342)	835.096	(660.226)	17.796	(14.366)	3307.599	(2137.419)	1679.607	(1336.634)	3589.685	(2316.063)	4424.781	(2976.200)
MDE.KL	2.235	(1.904)	24.311	(1.713)	49.005	(26.910)	24.851	(1.876)	1.793	(1.446)	27.346	(1.958)	79.332	(44.568)	27.976	(2.157)	52.827	(4.032)
MME	5.033	(3.949)	2382.947	(5.832)	235.503	(163.261)	2395.410	(11.830)	5.627	(4.518)	2253.515	(7.249)	462.743	(351.785)	2274.913	(19.624)	4670.323	(31.454)
$\beta = 2$																		
Method	$\beta = 2$				$\sigma = 1$				$\sigma = 1$				TMSE					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	10.631	(2.483)	6,374,279.987	(4.146)	431.530	(24.163)	6,374,311.979	(4.380)	8.050	(1.425)	4,411,470.525	(2.571)	704.975	(42.546)	4,411,490.815	(2.752)	10,785,802.793	(7.131)
MLEb	10.379	(2.230)	6,374,284.603	(4.675)	418.975	(11.481)	6,374,312.324	(4.728)	7.866	(1.234)	4,411,473.226	(2.841)	686.612	(23.997)	4,411,490.960	(2.899)	10,785,803.284	(7.627)
LME	2.560	(2.340)	16.215	(0.813)	29.521	(17.002)	16.564	(0.929)	1.430	(1.352)	8.543	(0.533)	42.901	(23.178)	8.528	(0.596)	25.092	(1.525)
MDE.CvM	2.094	(1.817)	17.499	(0.951)	4.700	(-9.160)	17.508	(0.984)	1.125	(0.923)	8.779	(0.523)	12.484	(-7.229)	8.794	(0.528)	26.302	(1.512)
MDE.SD	1.667	(1.425)	16.001	(0.344)	-16.649	(-28.771)	16.111	(0.675)	0.850	(0.678)	8.127	(0.191)	-15.507	(-32.233)	8.150	(0.295)	24.261	(0.970)
MDE. $\alpha^2$	5.751	(4.396)	440.292	(260.872)	187.574	(119.806)	454.363	(266.611)	7.995	(5.744)	931.582	(424.717)	699.467	(474.422)	980.502	(447.221)	1434.864	(713.832)
MDE.TV	14.132	(12.812)	1313.540	(1152.541)	606.596	(540.613)	1460.715	(1269.438)	19.502	(16.988)	2718.308	(2104.381)	1850.233	(1598.769)	3060.626	(2359.973)	4821.341	(3629.411)
MDE.KL	3.080	(2.483)	58.598	(4.142)	53.982	(24.154)	59.763	(4.376)	1.880	(1.425)	34.722	(2.568)	88.012	(42.531)	35.497	(2.749)	95.260	(7.124)
MME	10.502	(4.847)	1,519,554.747	(13.880)	425.099	(142.374)	1,519,616.900	(21.988)	8.101	(3.617)	1,049,846.183	(8.816)	710.103	(261.746)	1,049,889.609	(15.667)	2,569,506.509	(37.655)
$\beta = 2.1$																		
Method	$\beta = 2.1$				$\sigma = 1$				$\sigma = 1$				TMSE					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	3.968	(2.583)	4052.394	(4.795)	88.931	(23.023)	4055.854	(5.028)	2.408	(1.417)	2121.562	(2.695)	140.780	(41.714)	2123.530	(2.869)	6179.385	(7.897)
MLEb	3.725	(2.339)	4053.619	(5.354)	77.387	(11.362)	4056.233	(5.410)	2.241	(1.249)	2122.147	(2.951)	124.091	(24.856)	2123.673	(3.013)	6179.905	(8.423)
LME	2.683	(2.392)	20.284	(0.857)	27.739	(13.913)	20.623	(0.942)	1.406	(1.215)	9.003	(0.512)	40.628	(21.536)	9.168	(0.588)	29.791	(1.500)
MDE.CvM	2.189	(1.867)	21.999	(1.033)	4.258	(-11.083)	22.007	(1.087)	1.120	(0.909)	9.521	(0.513)	12.045	(-9.116)	9.535	(0.522)	21.129	(1.609)
MDE.SD	1.741	(1.456)	20.114	(0.365)	-17.108	(-30.653)	20.243	(0.780)	0.847	(0.660)	8.863	(0.182)	-15.288	(-33.966)	8.886	(0.298)	39.242	(1.077)
MDE. $\alpha^2$	4.338	(4.970)	506.733	(224.538)	201.832	(136.686)	524.694	(332.775)	8.337	(6.143)	965.206	(485.821)	733.660	(514.313)	1019.025	(512.270)	1543.719	(845.044)
MDE.TV	13.028	(13.701)	1407.830	(1245.632)	615.639	(552.412)	1574.965	(1380.198)	19.563	(17.179)	2992.689	(2043.054)	1856.338	(1617.909)	2937.271	(2384.883)	4512.326	(3688.001)
MDE.KL	3.230	(2.583)	66.156	(4.790)	53.819	(23.009)	67.432	(5.023)	1.874	(1.417)	33.640	(2.692)	87.407	(41.694)	34.404	(2.866)	101.836	(7.889)
MME	7.206	(4.999)	10764.703	(15.688)	243.140	(138.039)	10790.702	(24.091)	5.098	(3.303)	5326.312	(8.978)	409.809	(250.286)	5343.071	(15.242)	16133.772	(39.333)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table 2.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 2$  and  $n = 30$ ).

n = 30																		
Method	$\beta = 1.1$				$\beta = 1.5$				$\beta = 2$				$\beta = 2.1$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	2.654	(1.336)	80,051.293	(0.316)	141.251	(21.473)	80,053.273	(0.372)	5.885	(2.335)	45,6819.670	(2.234)	194.234	(16.734)	456831.715	(2.346)	536,884.988	(2.718)
MLEb	2.366	(1.045)	80,052.166	(0.330)	115.071	(-4.971)	80,053.235	(0.333)	5.070	(1.512)	456,825.356	(2.129)	153.505	(-24.488)	456,831.736	(2.367)	536,884.971	(2.701)
LME	1.737	(1.646)	2.859	(0.195)	57.919	(9.670)	3.265	(0.494)	3.352	(3.070)	26.767	(1.619)	67.586	(53.514)	28.594	(2.764)	31.859	(3.258)
MDEcVM	1.333	(1.215)	3.468	(0.241)	21.183	(10.461)	3.522	(0.254)	2.406	(2.035)	33.547	(1.679)	20.284	(1.767)	33.712	(1.680)	37.234	(1.934)
MDESD	1.108	(1.021)	2.787	(0.118)	0.734	(-7.190)	2.787	(0.125)	1.867	(1.393)	26.395	(0.838)	-4.674	(-20.574)	26.412	(1.004)	29.209	(1.129)
MDE $_{\alpha}^2$	-1.802	(-2.472)	519.986	(458.735)	-265.855	(-324.714)	528.407	(471.400)	8.987	(5.647)	1797.035	(18.494)	349.530	(182.333)	1835.836	(31.792)	2364.243	(503.283)
MDETV	3.114	(2.000)	179.513	(57.251)	183.069	(81.857)	183.567	(58.062)	16.174	(9.895)	5035.891	(988.450)	708.695	(394.731)	5236.757	(1050.769)	5420.324	(1108.830)
MDEKL	1.549	(1.336)	12.625	(0.316)	40.824	(21.474)	12.826	(0.372)	3.011	(2.335)	139.137	(2.234)	50.568	(16.735)	140.159	(2.346)	152.985	(2.718)
MME	4.587	(3.041)	104,262.039	(0.985)	317.029	(176.484)	104,273.505	(4.752)	15.807	(11.049)	620,871.008	(20.274)	690.332	(452.457)	621037.492	(102.160)	725,330.997	(106.913)
n = 30																		
Method	$\beta = 1.1$				$\beta = 1.5$				$\beta = 2$				$\beta = 2.1$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	3.840	(1.768)	303,171.263	(1.239)	156.002	(17.885)	303,174.717	(1.311)	6.526	(2.470)	1,070,845.334	(4.535)	226.281	(23.307)	1,070,858.676	(4.756)	1,374,033.393	(6.067)
MLEb	3.540	(1.465)	303,172.728	(1.462)	135.984	(-2.336)	303,174.868	(1.463)	5.907	(1.846)	1,070,851.049	(5.240)	195.372	(-7.714)	1,070,859.178	(5.264)	1,374,034.046	(6.272)
LME	2.097	(1.953)	6.287	(0.469)	39.804	(30.189)	6.643	(0.674)	3.039	(2.735)	27.573	(1.824)	51.931	(36.773)	26.652	(2.365)	35.295	(3.039)
MDEcVM	1.633	(1.460)	7.163	(0.627)	8.881	(-2.673)	7.181	(0.629)	2.260	(1.892)	31.451	(2.132)	12.989	(-5.409)	31.519	(2.143)	38.699	(2.272)
MDESD	1.328	(1.157)	6.142	(0.251)	-11.492	(-20.868)	6.172	(0.349)	1.731	(1.432)	27.114	(0.878)	-13.461	(-28.579)	27.187	(1.200)	33.358	(1.549)
MDE $_{\alpha}^2$	2.601	(1.438)	211.051	(77.716)	73.417	(-4.124)	212.262	(77.719)	10.979	(6.649)	2166.770	(223.677)	448.972	(232.431)	2247.386	(245.285)	2459.648	(323.004)
MDETV	6.649	(5.472)	486.172	(352.656)	343.236	(264.823)	512.677	(368.434)	24.295	(18.812)	6861.384	(3870.786)	1114.764	(840.592)	7358.418	(4153.398)	7871.094	(4521.831)
MDEKL	2.112	(1.768)	24.414	(1.239)	40.796	(17.885)	24.788	(1.311)	3.193	(2.470)	113.933	(4.535)	59.672	(23.307)	115.357	(4.756)	140.145	(6.067)
MME	5.558	(3.715)	134,131.052	(3.674)	270.516	(147.511)	134,146.624	(8.570)	11.891	(8.226)	474,718.074	(16.034)	494.542	(311.289)	474,812.737	(54.794)	608,959.361	(63.364)
n = 30																		
Method	$\beta = 1.1$				$\beta = 1.5$				$\beta = 2$				$\beta = 2.1$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	4.651	(2.542)	40,321.568	(5.325)	132.353	(27.104)	40,328.326	(5.619)	5.796	(2.743)	90,043.271	(9.927)	189.790	(37.143)	90,057.079	(10.478)	130,585.405	(16.097)
MLEb	4.393	(2.282)	40,323.226	(5.900)	119.658	(14.078)	40,328.653	(5.979)	5.433	(2.376)	90,046.502	(10.951)	171.652	(18.822)	90,057.087	(11.055)	130,586.570	(17.072)
LME	2.626	(2.378)	14.325	(1.017)	31.302	(18.909)	14.717	(1.160)	2.773	(2.418)	29.559	(1.987)	38.651	(20.895)	30.156	(2.162)	44.873	(3.322)
MDEcVM	2.121	(1.849)	15.686	(1.569)	6.048	(-7.561)	15.700	(1.392)	2.206	(1.814)	32.171	(2.826)	10.306	(-9.319)	32.213	(2.861)	47.913	(4.453)
MDESD	1.675	(1.429)	14.142	(0.594)	-16.232	(-25.568)	14.247	(0.838)	1.656	(1.301)	29.152	(0.946)	-17.193	(-34.566)	29.270	(1.435)	43.517	(2.245)
MDE $_{\alpha}^2$	5.823	(4.584)	413.157	(263.767)	191.143	(129.198)	427.769	(270.442)	14.546	(10.837)	2668.173	(1287.792)	627.256	(441.825)	2825.555	(1365.867)	3253.324	(1636.309)
MDETV	13.483	(12.285)	1121.905	(989.562)	574.169	(514.248)	1253.765	(1095.335)	33.639	(29.381)	7502.526	(5733.947)	1581.944	(1369.057)	8503.495	(6483.635)	9757.260	(7578.970)
MDEKL	3.148	(2.542)	57.284	(5.324)	57.420	(27.099)	58.602	(5.618)	3.589	(2.743)	115.121	(9.924)	79.429	(37.137)	117.643	(10.475)	176.246	(16.093)
MME	8.056	(4.920)	91,678.128	(14.940)	302.823	(146.019)	91,714.198	(23.468)	11.400	(6.907)	202,885.798	(28.586)	470.023	(245.539)	202,272.819	(52.582)	293,987.017	(76.050)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table 3.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 1$  and  $n = 50$ ).

n = 50																		
Method	$\beta = 1.1$					$\sigma = 1$					TMSE							
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.337	(1.279)	17.087	(0.179)	21.575	(16.301)	17.143	(0.211)	1.362	(1.278)	26.143	(0.451)	36.187	(27.806)	26.273	(0.528)	43.417	(0.739)
MLEb	1.121	(1.063)	17.023	(0.088)	1.919	(-3.354)	17.024	(0.090)	1.008	(0.921)	25.944	(0.194)	0.758	(-7.915)	25.944	(0.201)	43.968	(0.291)
LME	1.619	(1.589)	0.566	(0.141)	47.154	(44.486)	0.835	(0.380)	1.805	(1.761)	1.175	(0.437)	80.507	(76.078)	1.823	(1.016)	2.658	(1.397)
MDEcVM	1.255	(1.226)	0.645	(0.204)	14.071	(11.420)	0.669	(0.219)	1.235	(1.192)	1.203	(0.430)	23.469	(19.140)	1.258	(0.467)	1.927	(0.686)
MDESD	1.123	(1.100)	0.533	(0.135)	2.048	(-0.008)	0.534	(0.135)	1.059	(1.025)	0.969	(0.286)	5.921	(2.521)	0.972	(0.287)	1.506	(0.422)
MDE <sub>v</sub> <sup>2</sup>	0.576	(0.454)	52.779	(44.175)	-47.631	(-58.730)	53.053	(44.592)	2.643	(2.302)	64.223	(1.739)	164.331	(130.241)	66.923	(3.435)	119.976	(48.028)
MDETV	1.469	(1.023)	44.461	(0.558)	33.536	(-7.039)	44.597	(0.564)	4.124	(2.883)	356.464	(5.102)	312.391	(188.318)	366.220	(8.648)	410.817	(9.212)
MDEKL	1.324	(1.279)	1.142	(0.179)	20.398	(16.301)	1.192	(0.211)	1.346	(1.278)	2.994	(0.451)	34.392	(27.806)	2.714	(0.528)	3.906	(0.739)
MME	2.819	(2.731)	18.958	(0.429)	156.255	(148.298)	21.912	(0.091)	6.047	(3.830)	35.892	(4.301)	504.663	(484.988)	61.360	(27.822)	83.272	(30.913)
n = 50																		
Method	$\beta = 1.5$					$\sigma = 1$					TMSE							
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.910	(1.742)	29.072	(0.600)	27.336	(16.113)	29.240	(0.658)	1.435	(1.261)	29.125	(0.655)	43.471	(26.132)	29.314	(0.723)	58.554	(1.382)
MLEb	1.594	(1.422)	29.160	(0.580)	6.259	(-5.177)	29.168	(0.586)	1.099	(0.922)	29.195	(0.606)	9.870	(-7.809)	29.204	(0.613)	58.373	(1.199)
LME	2.006	(1.939)	2.268	(0.375)	33.746	(29.268)	2.525	(0.568)	1.497	(1.428)	2.143	(0.431)	49.655	(42.835)	2.390	(0.614)	4.914	(1.182)
MDEcVM	1.643	(1.570)	2.388	(0.373)	9.524	(4.850)	2.409	(0.378)	1.159	(1.084)	2.216	(0.360)	15.887	(8.438)	2.241	(0.367)	4.650	(0.745)
MDESD	1.435	(1.378)	2.054	(0.228)	-4.353	(-8.123)	2.058	(0.243)	0.972	(0.915)	1.885	(0.224)	-2.778	(-8.537)	1.886	(0.232)	3.944	(0.474)
MDE <sub>v</sub> <sup>2</sup>	1.777	(1.190)	64.224	(1.210)	18.480	(-20.645)	64.301	(1.306)	3.088	(2.071)	197.156	(2.661)	208.762	(107.097)	201.513	(3.808)	265.814	(5.114)
MDETV	3.474	(2.185)	213.434	(49.292)	131.589	(45.654)	217.328	(49.760)	6.070	(3.667)	704.911	(118.023)	506.972	(266.731)	730.609	(125.137)	947.937	(174.897)
MDEKL	1.894	(1.742)	6.843	(0.600)	25.604	(16.114)	6.991	(0.658)	1.409	(1.261)	7.616	(0.655)	40.948	(26.132)	7.794	(0.723)	14.775	(1.382)
MME	3.525	(3.268)	68.985	(1.662)	135.021	(117.847)	73.086	(4.787)	3.913	(3.633)	56.723	(2.081)	291.549	(265.328)	65.223	(9.121)	138.309	(13.908)
n = 50																		
Method	$\beta = 2$					$\sigma = 1$					TMSE							
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	4.064	(2.322)	88,391.676	(1.469)	103.208	(16.076)	88,395.348	(1.572)	2.550	(1.262)	53,460.738	(0.839)	154.964	(26.202)	53,462.783	(0.907)	141,858.131	(2.479)
MLEb	3.687	(1.941)	88,393.256	(1.737)	84.370	(-2.952)	88,395.514	(1.740)	2.275	(0.985)	53,461.571	(0.967)	127.540	(-1.500)	53,462.841	(0.966)	141,858.335	(2.706)
LME	2.541	(2.362)	10.175	(0.656)	27.056	(18.080)	10.468	(0.787)	1.379	(1.251)	5.002	(0.373)	37.875	(25.133)	5.145	(0.436)	15.613	(1.223)
MDEcVM	2.147	(1.950)	11.017	(0.672)	7.369	(-2.515)	11.039	(0.674)	1.130	(0.911)	5.261	(0.311)	13.028	(-0.918)	5.278	(0.311)	16.317	(0.985)
MDESD	1.944	(1.571)	9.964	(0.366)	-7.324	(-16.474)	9.989	(0.474)	0.939	(0.917)	4.853	(0.174)	-6.884	(-18.329)	4.856	(0.207)	14.845	(0.682)
MDE <sub>v</sub> <sup>2</sup>	3.234	(1.907)	184.224	(8.655)	61.702	(-4.626)	185.745	(8.664)	3.858	(2.219)	289.266	(13.104)	285.820	(121.906)	297.434	(14.590)	483.179	(23.254)
MDETV	6.526	(5.154)	508.122	(325.038)	226.284	(157.708)	528.600	(334.985)	7.836	(5.878)	805.664	(426.266)	683.559	(487.831)	852.384	(450.061)	1380.985	(785.046)
MDEKL	2.681	(2.321)	30.242	(1.468)	34.065	(16.074)	30.706	(1.571)	1.523	(1.262)	15.553	(0.838)	52.308	(26.199)	15.826	(0.907)	46.533	(2.479)
MME	6.154	(4.024)	117,688.362	(4.352)	207.707	(101.200)	117,704.835	(8.448)	4.407	(2.830)	71,115.685	(2.552)	340.671	(182.997)	71,126.816	(5.901)	188,831.651	(14.349)
n = 50																		
Method	$\beta = 2.1$					$\sigma = 1$					TMSE							
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	3.683	(2.425)	15,372.855	(1.606)	75.395	(15.489)	15,375.260	(1.712)	2.144	(1.258)	9856.006	(0.824)	114.436	(25.797)	9857.250	(0.891)	25,232.509	(2.602)
MLEb	3.299	(2.037)	15,374.150	(1.936)	37.099	(-2.992)	15,375.485	(1.940)	1.881	(0.992)	9856.614	(0.965)	88.066	(-0.839)	9857.324	(0.965)	25,232.809	(2.905)
LME	2.628	(2.435)	11.262	(0.687)	25.125	(15.946)	11.541	(0.799)	1.354	(1.225)	5.092	(0.350)	35.444	(22.513)	5.218	(0.401)	16.759	(1.200)
MDEcVM	2.237	(2.022)	12.392	(0.728)	6.532	(-3.711)	12.411	(0.734)	1.122	(0.979)	5.413	(0.299)	12.212	(-2.110)	5.428	(0.299)	17.838	(1.033)
MDESD	1.912	(1.725)	11.068	(0.389)	-8.935	(-17.851)	11.103	(0.530)	0.929	(0.804)	4.978	(0.165)	-7.889	(-19.626)	4.983	(0.203)	16.086	(0.733)
MDE <sub>v</sub> <sup>2</sup>	3.456	(2.091)	202.280	(17.788)	64.571	(-0.425)	204.118	(17.788)	3.896	(2.281)	286.675	(21.194)	289.624	(128.129)	295.061	(22.836)	499.179	(40.024)
MDETV	7.182	(5.800)	576.829	(391.624)	242.007	(176.209)	602.653	(405.314)	8.175	(6.272)	829.944	(472.227)	717.500	(527.232)	881.419	(500.021)	1484.072	(905.335)
MDEKL	2.816	(2.425)	34.535	(1.606)	34.086	(15.490)	35.047	(1.712)	1.526	(1.258)	16.273	(0.824)	52.628	(25.797)	16.550	(0.891)	51.597	(2.602)
MME	5.797	(4.155)	30,610.691	(4.790)	176.088	(97.777)	30,624.158	(9.006)	3.925	(2.730)	22,527.694	(2.531)	292.536	(173.042)	22,536.102	(5.525)	53,160.260	(14.531)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table 4.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 2$  and  $n = 50$ ).

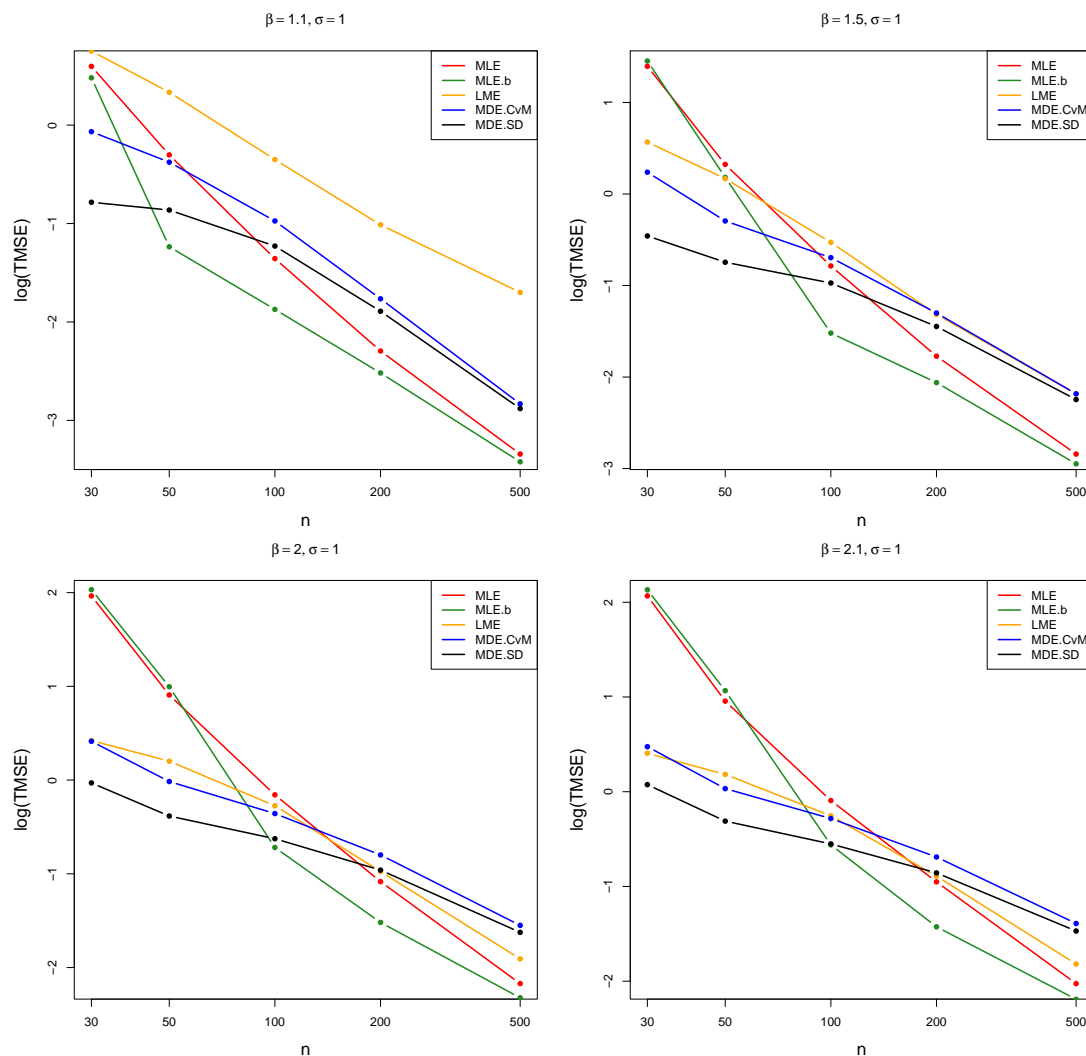
$n = 50$																		
Method	$\beta = 1.1$				$\beta = 1.5$				$\beta = 2$				$\beta = 2.1$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	1.321	(1.271)	7.676	(0.130)	20.075	(15.576)	7.724	(0.159)	2.446	(2.300)	58.608	(1.118)	22.202	(14.976)	58.806	(1.207)	66.531	(1.367)
MLEb	1.103	(1.053)	7.613	(0.848)	0.270	(-4.310)	7.613	(0.450)	1.795	(1.651)	58.000	(0.891)	-10.229	(-17.446)	58.041	(0.513)	65.654	(0.563)
LME	1.598	(1.571)	0.459	(0.112)	45.313	(42.838)	0.707	(0.334)	3.194	(3.113)	3.822	(1.118)	59.701	(55.631)	5.247	(2.356)	5.955	(2.689)
MDE CVM	1.262	(1.232)	0.599	(0.157)	14.738	(11.971)	0.625	(0.175)	2.283	(2.188)	4.956	(1.228)	14.135	(9.423)	5.036	(1.263)	5.661	(1.438)
MDE SD	1.131	(1.108)	0.444	(0.106)	2.851	(0.722)	0.445	(0.106)	1.862	(1.889)	3.498	(0.832)	-1.897	(-5.562)	3.499	(0.845)	3.944	(0.951)
MDE $\chi^2$	-2.443	(-2.559)	432.103	(430.062)	-322.096	(-332.608)	444.653	(443.445)	4.567	(4.071)	143.225	(4.875)	128.353	(103.562)	149.814	(9.165)	594.468	(452.410)
MDE TV	1.195	(0.966)	18.898	(0.343)	8.655	(-12.149)	18.907	(0.361)	5.975	(4.822)	493.736	(9.713)	198.729	(141.079)	509.530	(17.674)	528.437	(18.035)
MDE KL	1.314	(1.271)	1.379	(0.130)	19.457	(15.577)	1.425	(0.159)	2.427	(2.300)	13.544	(1.118)	21.357	(14.976)	13.726	(1.207)	15.151	(1.367)
MME	2.758	(2.688)	9.027	(0.306)	150.697	(144.400)	11.774	(2.829)	10.963	(10.449)	113.818	(11.941)	448.140	(422.473)	194.149	(83.335)	205.924	(86.164)
$n = 50$																		
Method	$\beta = 1.5$				$\beta = 2$				$\beta = 2$				$\beta = 2$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	1.401	(1.635)	1222.016	(0.360)	26.749	(8.975)	1222.169	(0.378)	2.790	(2.267)	3893.460	(1.445)	39.491	(13.340)	3894.057	(1.516)	514.226	(1.895)
MLEb	1.586	(1.316)	1222.114	(0.289)	5.734	(-12.252)	1222.113	(0.323)	2.130	(1.600)	3893.758	(1.045)	6.406	(-19.980)	3893.749	(1.205)	515.862	(1.528)
LME	1.945	(1.872)	2.313	(0.260)	29.683	(24.769)	2.511	(0.398)	2.829	(2.683)	9.341	(1.145)	41.442	(34.155)	10.028	(1.612)	12.539	(2.010)
MDE CVM	1.586	(1.471)	2.566	(0.264)	3.723	(-1.942)	2.569	(0.265)	2.122	(1.952)	10.429	(0.988)	6.081	(-2.408)	10.444	(0.991)	13.013	(1.256)
MDE SD	1.372	(1.303)	2.203	(0.167)	-8.515	(-13.118)	2.219	(0.206)	1.801	(1.662)	8.880	(0.637)	-9.964	(-16.576)	8.920	(0.751)	11.139	(0.957)
MDE $\chi^2$	1.272	(0.903)	59.656	(25.287)	-15.204	(-39.807)	59.708	(25.644)	4.937	(3.750)	372.382	(5.471)	146.837	(87.524)	381.004	(8.536)	440.712	(34.179)
MDE TV	2.517	(1.565)	104.788	(2.606)	67.820	(4.316)	105.822	(2.610)	8.412	(5.254)	120.454	(32.218)	320.602	(162.680)	1242.561	(42.804)	1348.383	(45.414)
MDE KL	1.773	(1.635)	7.148	(0.360)	18.170	(8.975)	7.222	(0.378)	2.549	(2.67)	32.956	(1.445)	27.432	(13.349)	33.257	(1.516)	40.479	(1.895)
MME	3.487	(3.133)	2302.496	(0.974)	132.488	(108.849)	2306.430	(3.640)	7.675	(6.982)	7317.771	(5.416)	283.772	(249.093)	7349.933	(30.235)	9656.363	(53.875)
$n = 50$																		
Method	$\beta = 2$				$\beta = 2$				$\beta = 2$				$\beta = 2$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	3.964	(2.294)	87,007.295	(1.892)	98.197	(14.702)	87,010.572	(1.979)	4.811	(2.410)	209,557.330	(3.584)	140.568	(20.522)	209,563.836	(3.753)	296,574.408	(5.731)
MLEb	3.502	(1.919)	87,008.800	(2.157)	79.611	(-4.072)	87,010.756	(2.164)	4.280	(1.873)	209,560.318	(4.021)	113.978	(-6.837)	209,564.118	(4.037)	296,574.873	(6.201)
LME	2.540	(2.349)	9.617	(0.813)	27.016	(17.454)	9.909	(0.935)	2.690	(2.420)	19.220	(1.570)	34.494	(21.013)	19.696	(1.747)	29.605	(2.682)
MDE CVM	2.086	(1.876)	10.282	(0.925)	4.282	(-6.206)	10.289	(0.940)	2.143	(1.848)	20.472	(1.684)	7.187	(-7.577)	20.493	(1.707)	30.782	(2.648)
MDE SD	1.800	(1.610)	9.404	(0.462)	-9.987	(-19.497)	9.444	(0.815)	1.792	(1.523)	18.879	(0.854)	-10.283	(-23.865)	18.922	(1.081)	28.366	(1.696)
MDE $\chi^2$	3.158	(1.929)	162.550	(13.023)	57.896	(-3.538)	163.890	(13.028)	7.065	(4.326)	826.859	(59.570)	253.272	(116.200)	852.512	(64.979)	1016.402	(78.007)
MDE TV	6.245	(5.002)	432.995	(282.844)	212.245	(150.104)	451.011	(291.854)	14.062	(10.739)	2329.501	(1232.378)	603.122	(436.940)	2474.988	(1308.737)	2925.999	(1600.591)
MDE KL	2.700	(2.294)	32.588	(1.891)	35.001	(14.698)	33.078	(1.977)	2.980	(2.410)	65.105	(3.583)	49.011	(20.517)	66.065	(3.751)	99.143	(5.729)
MME	6.068	(4.028)	116,001.156	(4.733)	203.376	(101.383)	116,016.928	(8.845)	8.449	(5.521)	278,863.728	(9.167)	322.454	(176.057)	278,903.459	(21.565)	394,920.587	(50.410)
$n = 50$																		
Method	$\beta = 2.1$				$\beta = 2.1$				$\beta = 2.1$				$\beta = 2.1$					
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE	TMSE	
MLE	3.792	(2.431)	12,233.774	(2.372)	80.558	(15.770)	12,236.554	(2.482)	4.344	(2.436)	31,226.068	(3.994)	117.180	(21.824)	31,231.352	(4.185)	43,467.907	(6.667)
MLEb	3.418	(2.053)	12,235.142	(2.725)	62.746	(-2.223)	12,236.797	(2.727)	3.842	(1.930)	31,228.517	(4.532)	92.091	(-3.518)	31,231.701	(4.537)	43,468.407	(7.264)
LME	2.639	(2.440)	10.710	(0.934)	25.676	(16.193)	11.001	(1.050)	2.643	(2.382)	18.616	(1.598)	32.159	(19.099)	19.029	(1.744)	30.030	(2.794)
MDE CVM	2.164	(1.947)	11.349	(1.078)	3.062	(-7.287)	11.353	(1.101)	2.110	(1.825)	19.753	(1.762)	5.496	(-8.728)	19.765	(1.793)	31.118	(2.894)
MDE SD	1.859	(1.662)	10.413	(0.524)	-11.467	(-20.869)	10.470	(0.716)	1.755	(1.495)	18.274	(0.868)	-12.230	(-25.263)	18.333	(1.124)	28.804	(1.839)
MDE $\chi^2$	3.521	(2.274)	190.586	(36.774)	67.675	(8.271)	192.604	(36.804)	7.400	(4.700)	864.422	(127.551)	269.986	(135.003)	893.573	(154.841)	1088.177	(171.645)
MDE TV	7.138	(5.888)	516.606	(365.525)	239.903	(180.380)	541.984	(379.871)	15.073	(11.903)	2439.587	(1445.992)	653.643	(495.143)	2610.470	(1544.049)	3152.454	(1923.921)
MDE KL	2.853	(2.431)	34.842	(2.371)	35.839	(15.766)	35.408	(2.481)	2.990	(2.436)	61.063	(3.992)	49.499	(20.517)	62.043	(4.183)	97.451	(6.663)
MME	5.909	(4.218)	16,449.516	(5.815)	181.398	(100.875)	16,463.918	(10.302)	7.785	(5.383)	47,121.621	(9.898)	289.267	(169.145)	47154.777	(21.342)	63,618.695	(31.645)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

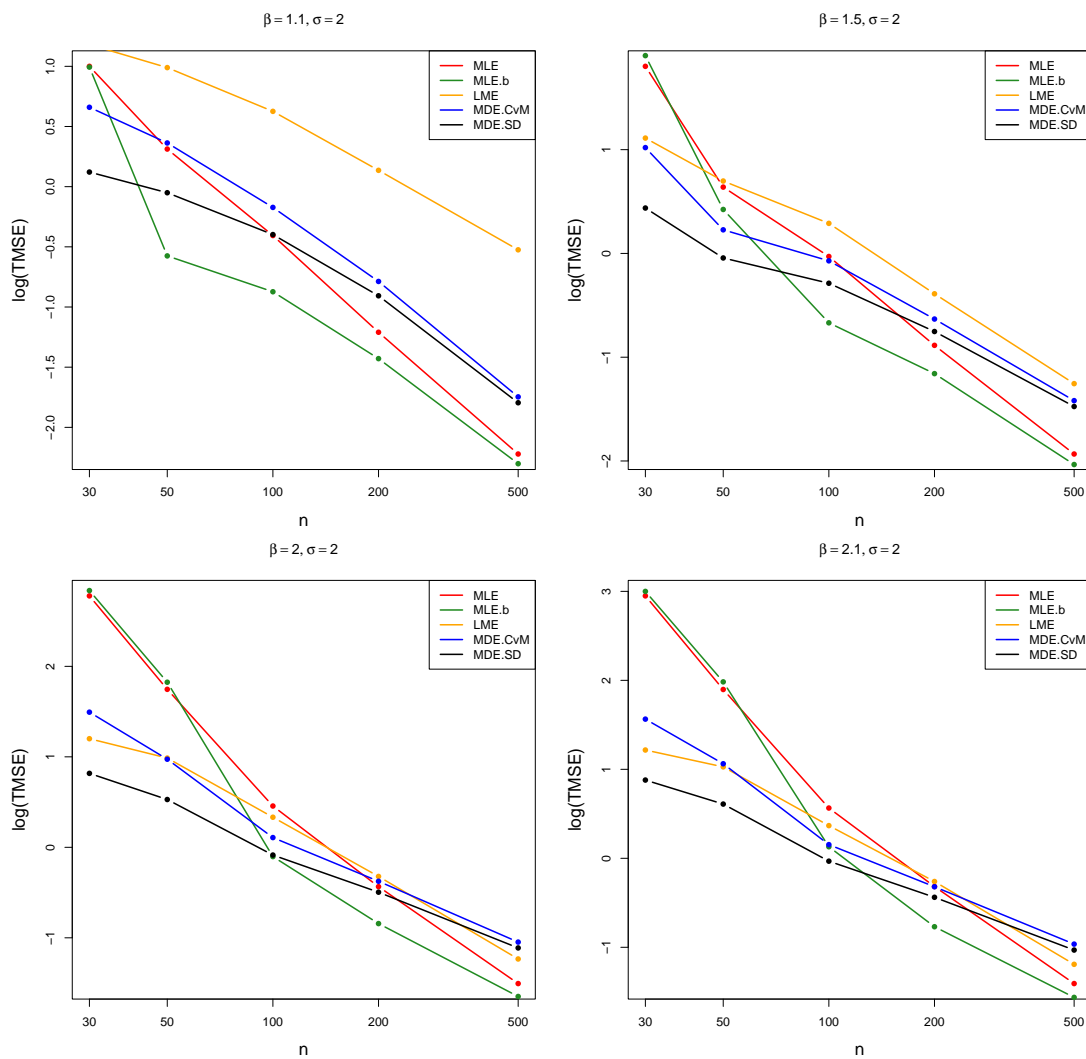
The main entries in Tables 1–4 and A1–A6 represent the unadjusted calculation of these quantities; however, due to issues with a small number of aberrant samples in the simulation, the optimisation produced extreme values for the estimators in these cases. Therefore, in order to better understand the behaviour of the estimators, a ‘trimmed’ version of each of these quantities is provided in parentheses. These were obtained by first removing the largest 1% of the estimator values produced in the simulation and then calculating the various metrics. The effect of this trimming is most pronounced for smaller samples.

Furthermore, each of the columns in the tables have a single value printed in bold that highlights which of the estimators perform ‘best’. Mean values in bold font indicate that the estimated values are closest to the true parameter value on average, whereas for variance, RB, MSE and TMSE bold indicates that these values are the smallest among the estimators considered.

Figures 1 and 2 show the logarithm of the TMSE as a function of sample size. We choose to use a logarithmic scale, as this increases the separation between the lines on the graph, which makes the figures easier to interpret.



**Figure 1.** Plots of the log of the trimmed TMSE measure against sample size for five different estimation techniques when  $\sigma = 1$ .



**Figure 2.** Plots of the log of the trimmed TMSE measure against sample size for five different estimation techniques when  $\sigma = 2$ .

**Remark 3.** The reason for setting the number of Monte Carlo simulations to the relatively high number of 150,000 was due to the extreme sampling variation observed with the MLE and MLE.b estimators. This variability was exacerbated whenever using a particularly small sample size combined with a  $\beta$  value close to 2; therefore, we thought it best to make the number of simulations as large as possible in order to mitigate this issue. The variability is likely due to numerical issues related to the small sample sizes; note, however, that as the sample increases to  $n = 100$  and above this artefact disappears almost entirely. In addition, when considering the trimmed version of the measures calculated for each estimator in the simulation it is apparent that the sampling variability exhibited by these estimators is damped, which we feel is a more accurate reflection of the true value of these measures.

### 3.2. Results

To obtain a better understanding of the performance of the estimators considered in this study, we present a brief discussion of the simulation results presented in Tables 1–4 and A1–A6. We compare the trimmed versions of the variance, RB, and MSE performance of the estimators to one another in various parameter settings in the Monte Carlo study.

We start by first noting that the RB, variance, and MSE of each of the estimation methods improve when the sample size increases. The most notable improvements in overall MSE as the sample size increases from 30 to 500 are in the MLE, MDE.TV, and

MME; however, as we mention later, the MME method should likely not be considered for estimation even for large samples. Interestingly, we find that the methods that perform well for smaller sample sizes do not always perform well for larger samples when compared to the competing methods. In particular, when the sample size is small ( $n = 30$  or  $n = 50$ ), it can be seen that in the majority of cases MDE.SD, MDE.CvM, and LME have the smallest variance, RB, and MSE values. Note that this small sample behaviour of the variance and MSE measures holds for all but the  $\beta = 1.1$  parameter. However, when considering the case in which the sample size is small and the parameter  $\beta$  is specified to be small, there are instances where the MDE.CvM exhibits comparatively larger bias than MDE.SD, LME, and even MLE.b. Thus, it is clear that the underlying parameter values play a role in the relative performance of the various estimation techniques considered. Finally, when considering the RB for the various estimators, we note that MDE.SD most frequently gives rise to negative RB values for small samples. Based on the generally positive RB values associated with the remaining estimators, it seems that these other estimators tend to overestimate the true parameter value in these cases. This behaviour of MDE.SD (and indeed all other MDE-type estimators that we considered) is difficult to explain, as their distributional properties depend on the parameters and do not permit simple closed-form expressions.

Note that while MDE.KL, MDE.SD, and LME are relatively good estimators (in an MSE sense) when the sample size is large ( $n = 200$  or  $n = 500$ ), the MLE method now has the *best* performance among all methods considered in this setting.

Next, we consider the behaviour of the estimators as the parameter settings of the Monte Carlo change from one value to the next. As  $\beta$  increases from 1.1 to 2, the values of the approximated variance, RB, and MSE of the estimators for both  $\beta$  and  $\sigma$  have the tendency to increase as well. However, these values then immediately start to decrease again when  $\beta$  is strictly larger than 2.

When the value of  $\sigma$  increases from 1 to 2, the simulation approximation of the variance of the estimators of  $\sigma$  generally increases for all of the different estimation methods. The same is not necessarily true for the estimation of  $\beta$ , where we find that there are instances in which the approximated variance of the estimator of  $\beta$  (using a given estimation method) decreases when the true  $\sigma$  increases from 1 to 2, and increases in others.

When comparing the class of estimators based on minimum distance measures to the remaining methods, it can be readily seen that the MDE.SD and MDE.CvM estimators are the best performers in terms of the MSE for small to moderate samples. While LME is an excellent competitor, the MLE and MLE.b estimators perform relatively poorly in terms of the MSE in these cases. However, for larger sample sizes, the MLE and MLE.b clearly outperform the entire MDE class of estimators. Interestingly, the LME remains competitive here too, though it is almost never found to have the best MSE performance.

The performances of the estimators within the class of MDE methods differ wildly from one another. We have already noted that the MDE.SD and MDE.CvM methods generally have good MSE performance for small to moderate samples sizes; however, within this class we find that the MDE.TV and MDE. $\chi^2$  are among the worst performances. The MDE.KL method is something of a mixed bag; certain parameter settings yield good MSE performance, while others yield extremely poor MSE performance. Generally, however, we can state that the MDE.KL method is a better performer than MDE.TV and MDE. $\chi^2$ , although this does not place this estimator near the top performers among the estimation methods considered.

The overall worst performing estimator is MME, regardless of which value of  $\beta$  or  $\sigma$  is considered. While this estimator improves with an increase in sample size, it nonetheless has the worst performance when compared to all other estimation methods.

In Figures 1 and 2, the logarithm of the trimmed TMSE measure is plotted against the samples sizes in order to produce a more intuitive representation of the behaviour of a selection of estimators over varying sample sizes. The estimators MLE, MLE.b, LME, MDE.CvM, and MDE.SD are selected for inclusion in these figures, as according to the tables these estimators generally perform well (the TMSE and trimmed TMSE values of

the remaining estimators can be found in the tables). It can be seen that, as expected, the trimmed TMSE measure’s value generally decreases as the sample size increases. Note, however, that the trimmed TMSE indicates that the MLE and MLE.b estimators are erratic for small samples (see for example the graphs associated with  $\beta = 2$  in Figures 1 and 2). Furthermore, it is apparent that while the MLE and MLE.b estimators have among the worst trimmed TMSE performance for small sample sizes, this trend is reversed when the sample size increases. Conversely, it is clear that the MDE.CvM and MDE.SD estimators perform well in terms of TMSE for smaller sample sizes, in many cases being outperformed only when the sample size increases to  $n = 100$  and higher. It should be noted that while the performance of LME is comparable to the performance of MDE.CvM for choices of  $\beta$  larger than 2, it performs poorly for smaller  $\beta$  values. Furthermore, these numerical results indicate that the relative performance of the estimators depends to a certain degree on the specific parameter values considered.

#### 4. Practical Applications

We now apply the various Lomax parameter estimation techniques to a practical data set, namely, the airborne communication transceiver data shown in Table 5, originally discussed in [41], and previously analysed in [22].

**Table 5.** Repair times for the airborne communication transceiver dataset.

0.2	0.3	0.5	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8
0.8	1.0	1.0	1.0	1.0	1.1	1.3	1.5	1.5	1.5	1.5	2.0
2.0	2.2	2.5	2.7	3.0	3.0	3.3	3.3	4.0	4.0	4.5	4.7
5.0	5.4	5.4	7.0	7.5	8.8	9.0	10.3	22.0	24.5		

The dataset contains 46 observed repair times (measured in hours) of airborne communication transceiver equipment recorded during active operation. Assuming a Lomax distribution for these data, we use each of the estimation techniques to obtain the parameter estimates of the distribution. The resulting parameter estimates, along with 95% parametric bootstrap confidence intervals using  $B = 1000$  bootstrap replications, are shown in Table 6. Note that for each estimation method the parametric bootstrap samples used in these intervals are constructed by first estimating the parameters and then generating samples from a Lomax distribution with the estimators as the parameters [42]. The output in Table 6 indicates that the estimated parameter values are all reasonably close to one another, with the exception of MDE.CvM, MDE. $\chi^2$ , MDE.TV, and MME. The discrepancies observed for these four estimators relative to the other estimators considered is perhaps unsurprising in light of the irregular performance of these estimators observed in the Monte Carlo study presented earlier.

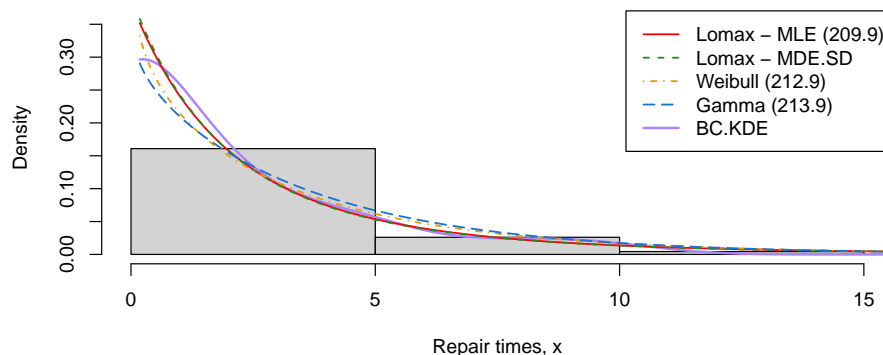
**Table 6.** Estimated parameters for the airborne communication transceiver dataset (with a 95% percentile bootstrap confidence interval).

Estimation Technique	$\hat{\beta}$	$CI_{0.95}$	$\hat{\sigma}$	$CI_{0.95}$
MLE	3.549	[1.540; 56.744]	9.236	[2.672; 166.270]
MLE.b	3.549	[1.222; 43.509]	9.236	[1.792; 138.151]
LME	3.387	[1.692; 16.621]	8.608	[3.171; 52.699]
MDE.CvM	4.105	[0.965; 34.123]	10.685	[1.223; 112.835]
MDE.SD	3.399	[0.933; 26.084]	8.670	[1.312; 78.789]
MDE. $\chi^2$	1.820	[0.583; 19.904]	7.706	[5.077; 178.639]
MDE.TV	2.402	[0.819; 105.179]	9.732	[5.293; 920.561]
MDE.KL	3.549	[1.474; 52.620]	9.236	[2.653; 191.968]
MME	4.385	[2.603; 81.117]	12.209	[6.540; 290.507]

The code for this example can be found here: <https://tinyurl.com/LomaxPracticalCode> (accessed on 26 November 2023).

Figure 3 shows the boundary-corrected kernel density estimator (BC.KDE) of the data (obtained using the `bckden` suite of functions in R’s `evmix` package [43]) overlaid on a simple histogram. Two separate Lomax density functions are added to the plot using two different methods of parameter estimation, namely, MLE and MDE.SD, the latter being chosen because this method performed well for smaller sample sizes. In addition, the fitted

two-parameter Weibull and gamma densities are added to the plot in order to facilitate comparison between the Lomax density and other theoretical densities containing two parameters. The parameters from both of these theoretical densities are estimated via maximum likelihood. Visual inspection of the graph indicates that the Lomax densities can be considered good candidate models, as they most closely match the empirical estimate of the density of these data, whereas the gamma distribution exhibits the greatest deviation from the others.



**Figure 3.** The kernel density estimate and fitted density functions based on the airborne communication transceiver repair time data (AIC values for the MLE estimated distributions provided in parentheses in the legend).

The Akaike information criterion (AIC) is calculated for each of the three distinct distributions by employing the maximum likelihood estimators, with the results presented in Table 7. From Table 7, it can be seen that the Lomax distribution has the lowest AIC value, closely followed by the AIC of the Weibull and gamma distributions, which supports the conclusions made above.

**Table 7.** The AIC values for the fitted distributions of the airborne communications transceiver data.

Weibull	Gamma	Lomax*
212.9	213.9	209.9

\* The AIC value reported for the Lomax distribution employs the MLE.

### 5. Conclusions

In this study, different parameter estimation methods for the two parameters of the Lomax distribution have been explored and discussed. We have reviewed *L*-moment estimators, maximum likelihood estimators (with and without a bias adjustment), method of moments estimators, and three different minimum distance estimators. The specific goals of the study were to provide an overview of the different estimation methods as well as to determine whether the minimum distance estimators are feasible estimation alternatives for this distribution when compared to traditional methods of estimation such as maximum likelihood. In an attempt to ascertain the properties of these estimators, all of the methods were compared to one another in a comprehensive Monte Carlo simulation while assuming different parametric values for small to large samples. Unsurprisingly, this study showed that while the MLEs perform well in large samples, they yield severely biased answers in small samples. The bias correction outlined in [14], however, noticeably reduces the bias of the MLE estimates for moderate to large sample sizes.

The traditional MME is found to be a universally poor performer in terms of MSE for all settings of the  $\beta$  and  $\sigma$  parameters and the given sample sizes. Our simulation showed that LME is stable under different parameter settings, including varying values of  $\beta$ ,  $\sigma$ , and sample size. Thus, while this estimator is almost never found to have the best performance in either small or large sample settings, it can be considered a ‘safe’ option for estimating of the parameters of the Lomax distribution.

Finally, several estimation methods based on distance measures were investigated, and were found to produce varied results. These methods all attempt to estimate the parameters of the Lomax distribution by considering different measures of distance measures that need to be minimised. The studied measures considered include the Cramér–von Mises (CvM), squared difference, and several phi-divergence measures (including the Kullback–Leibler (KL) divergence, total variation (TV) distance, and chi-square divergence). The results of the simulation study established that, for this class of MDE methods, the MDE.SD and MDE.CvM methods have the best overall MSE performance for small to moderate samples sizes; we would recommend either of these methods for the estimation of the Lomax parameters  $\sigma$  and  $\beta$  when sample sizes are smaller than 100. However, within this MDE class we find two of the worst performing methods, namely, the MDE.TV and MDE. $\chi^2$  methods. The MDE.KL method produces reasonably good MSE performance for the moderate to large sample sizes, however, as with MLE, the quality of the estimator fluctuates tremendously for smaller samples.

These results show that while no single estimator has the best overall performance, we can identify the distance-based measure, that is, MDE.SD, as generally having the best MSE performance for small sample sizes ( $n = 30$  and  $n = 50$ ), whereas MLE.b is found to be the better choice for moderate and large sample sizes ( $n = 100$ ,  $n = 200$  and  $n = 500$ ) for almost all of the  $\sigma$  and  $\beta$  values considered here.

The related literature contains finds competing views about the usefulness of MDEs and MLEs. In the case of multivariate data (specifically estimators for the parameters of a copula), [44] advocates for the use of MDEs because MLEs have the disadvantage that of requiring the density of the copula, whereas MDEs only require its empirical distribution estimate. However, in an extensive overview paper comparing MDEs and MLEs for parameter estimation of copulas, [45] found that MLEs have a lower computational cost while producing lower estimation biases. Naturally, the other main advantages of the traditional MLEs are the ease with which asymptotic standard errors can be obtained, the fact that they have asymptotically normal distributions, and, consequently, that confidence intervals for the parameter being estimated are easy to obtain. In contrast, the standard errors of the other estimators considered in this paper require either complex analytical derivations (possibly with no tractable forms) or require the use of resampling methods.

Finally, we note that another important application of the Lomax distribution, not mentioned yet in this paper, is found in autoregressive conditional duration (ACD) models, which model high frequency financial data; in this context, the duration between market events are of great interest. Without discussing the details of the model, we briefly describe the role of the Lomax distribution in ACD models here; for an in-depth exposition of ACD models, see [46,47]. We only note that the error terms of the ACD model are historically assumed to be exponential or Weibull distributed; see, for example, [48]. However, in [49,50], it is found that these choices do not adequately describe the characteristics of the observed data. In [51], the authors advocate the use of infinite mixtures of exponential distributions to model the distribution of the error term, finding that this choice improves the fit of the model. If the scale parameter of this mixing exponential distribution follows an inverse Gaussian distribution, then the resulting distribution of the error term is Lomax. It is therefore important for the efficient implementation of these ACD models that we are able to accurately estimate the parameters of this distribution. Another important avenue for future research can be found in goodness-of-fit tests for the Lomax and related distributions, as these tests check the assumptions underlying the ACD models.

**Author Contributions:** Conceptualization, T.N., J.A., L.S. and J.V.; Methodology, T.N., J.A., L.S. and J.V.; Writing—original draft, T.N., J.A., L.S. and J.V.; Writing—review and editing, T.N., J.A., L.S. and J.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The work of J.S. Allison, L. Santana and IJH Visagie are based on research supported by the National Research Foundation (NRF). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Additional Tables

Tables [A1–A6](#) report the results of the Monte Carlo experiments for sample sizes  $n = 100$ ,  $n = 200$ , and  $n = 500$ . Each table shows the empirical expected value, variance, RB, MSE, and TMSE for the various estimators considered, as in the body of the paper.

**Table A1.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 1$  and  $n = 100$ ).

$n = 100$																		
Method	$\beta = 1.1$					$\sigma = 1$					RB	MSE	TMSE					
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE								
MLE	1.204	(1.190)	0.089	(0.067)	9.443	(8.157)	0.099	(0.075)	1.161	(1.139)	0.215	(0.163)	16.122	(13.906)	0.241	(0.182)	0.340	(0.258)
MLEb	1.109	(1.099)	0.057	(0.046)	<b>0.805</b>	(-0.113)	0.057	(0.046)	1.007	(0.992)	<b>0.133</b>	(0.107)	<b>0.714</b>	(-0.850)	<b>0.133</b>	(0.107)	<b>0.190</b>	(0.154)
LME	1.471	(1.458)	0.079	(0.061)	33.698	(32.518)	0.216	(0.189)	1.593	(1.572)	0.232	(0.188)	59.307	(57.215)	0.584	(0.515)	0.880	(0.705)
MDE.CvM	1.201	(1.186)	0.138	(0.116)	9.186	(7.814)	0.148	(0.123)	1.153	(1.131)	0.286	(0.237)	15.314	(13.097)	0.309	(0.255)	0.457	(0.378)
MDE.SD	1.139	(1.126)	0.111	(0.095)	3.553	(2.360)	0.113	(0.095)	1.071	(1.052)	0.232	(0.195)	7.149	(5.203)	0.237	(0.197)	0.350	(0.293)
MDE. $\chi^2$	-0.012	(-0.034)	77.255	(77.879)	-101.064	(-103.103)	78.491	(79.164)	1.718	(1.544)	5.834	(0.394)	71.761	(54.417)	6.348	(0.690)	84.839	(79.854)
MDE.TV	0.766	(0.742)	0.318	(0.104)	-30.374	(-32.568)	0.430	(0.233)	1.852	(1.789)	2.022	(0.685)	85.230	(78.875)	2.749	(1.307)	3.178	(1.540)
MDE.KL	1.204	(1.190)	0.089	(0.067)	9.443	(8.157)	0.099	(0.076)	1.161	(1.139)	0.215	(0.163)	16.122	(13.906)	0.241	(0.182)	0.340	(0.258)
MME	2.421	(2.400)	0.150	(0.103)	120.116	(118.218)	1.896	(1.794)	5.579	(5.426)	5.865	(2.744)	457.943	(442.637)	26.836	(22.336)	28.732	(24.130)
$\beta = 1.5$																		
Method	$\beta = 1.5$					$\sigma = 1$					RB	MSE	TMSE					
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE								
MLE	1.654	(1.621)	0.429	(0.204)	10.260	(8.088)	0.452	(0.219)	1.165	(1.131)	0.482	(0.220)	16.474	(13.072)	0.509	(0.237)	0.962	(0.456)
MLEb	1.462	(1.436)	0.312	(0.103)	-2.534	(-4.297)	0.313	(0.108)	0.962	(0.934)	0.350	(0.107)	-3.802	(-6.589)	0.351	(0.111)	0.665	(0.219)
LME	1.834	(1.809)	0.263	(0.189)	22.265	(20.605)	0.375	(0.284)	1.331	(1.305)	0.294	(0.212)	33.066	(30.476)	0.403	(0.385)	0.778	(0.590)
MDE.CvM	1.605	(1.581)	0.317	(0.245)	7.013	(5.426)	0.328	(0.252)	1.114	(1.089)	0.317	(0.239)	11.370	(8.921)	0.330	(0.247)	0.658	(0.499)
MDE.SD	1.502	(1.482)	0.241	(0.191)	1.016	(-1.218)	0.241	(0.191)	1.021	(1.000)	0.243	(0.187)	2.148	(0.086)	0.244	(0.187)	0.485	(0.378)
MDE. $\chi^2$	1.007	(0.928)	5.467	(0.842)	-32.854	(-38.143)	5.710	(1.170)	1.512	(1.386)	12.397	(0.492)	51.212	(38.599)	12.459	(0.641)	18.369	(1.811)
MDE.TV	1.301	(1.164)	10.168	(0.383)	-13.251	(-22.423)	10.207	(0.496)	1.890	(1.668)	26.650	(0.938)	89.003	(66.827)	27.442	(1.385)	37.649	(1.881)
MDE.KL	1.654	(1.621)	0.429	(0.204)	10.261	(8.088)	0.453	(0.219)	1.165	(1.131)	0.483	(0.220)	16.474	(13.073)	0.510	(0.237)	0.963	(0.456)
MME	2.813	(2.762)	0.939	(0.399)	87.515	(84.162)	2.662	(1.993)	3.082	(3.031)	1.231	(0.582)	208.215	(203.059)	5.567	(4.705)	8.229	(6.698)
$\beta = 2$																		
Method	$\beta = 2$					$\sigma = 1$					RB	MSE	TMSE					
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE								
MLE	2.358	(2.185)	1253.813	(0.527)	17.902	(9.228)	1253.933	(0.561)	1.257	(1.139)	516.870	(0.274)	25.711	(13.943)	516.933	(0.293)	1770.866	(0.854)
MLEb	2.012	(1.835)	1253.708	(0.298)	<b>0.586</b>	(-8.264)	1253.700	(0.326)	1.007	(0.887)	516.806	(0.149)	<b>0.681</b>	(-11.339)	516.802	(0.162)	1770.503	(0.487)
LME	2.345	(2.298)	1.351	(0.416)	17.241	(14.878)	1.470	(0.504)	1.234	(1.199)	0.655	(0.216)	23.360	(9.926)	0.709	(0.256)	2.179	(0.760)
MDE.CvM	2.071	(2.020)	1.472	(0.475)	3.533	(1.004)	1.477	(0.476)	1.060	(1.026)	0.684	(0.223)	5.994	(2.559)	0.687	(0.224)	2.164	(0.700)
MDE.SD	1.909	(1.866)	1.257	(0.347)	-4.575	(-6.707)	1.266	(0.365)	0.959	(0.930)	0.585	(0.164)	-4.116	(-7.043)	0.587	(0.169)	1.852	(0.535)
MDE. $\chi^2$	1.784	(1.430)	36.496	(0.708)	-10.784	(-28.485)	36.542	(1.033)	1.795	(1.411)	43.330	(0.804)	79.511	(41.099)	43.962	(0.973)	80.504	(2.005)
MDE.TV	2.326	(1.816)	55.816	(1.368)	16.315	(-9.222)	55.922	(1.402)	2.282	(1.727)	68.700	(1.559)	128.234	(72.668)	68.344	(2.087)	124.266	(3.489)
MDE.KL	2.262	(2.185)	3.326	(0.527)	13.082	(9.228)	3.394	(0.561)	1.195	(1.139)	1.559	(0.274)	19.514	(13.943)	1.597	(0.293)	4.991	(0.854)
MME	3.583	(3.361)	1651.397	(1.147)	79.138	(68.033)	1653.891	(2.999)	2.380	(2.228)	680.774	(0.611)	137.966	(122.766)	682.673	(2.118)	2336.564	(5.117)
$\beta = 2.1$																		
Method	$\beta = 2.1$					$\sigma = 1$					RB	MSE	TMSE					
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE								
MLE	2.400	(2.290)	21.204	(0.585)	14.303	(9.028)	21.294	(0.621)	1.211	(1.135)	9.059	(0.272)	21.059	(13.518)	9.104	(0.290)	30.397	(0.911)
MLEb	2.024	(1.910)	21.064	(0.358)	-3.596	(-9.052)	21.069	(0.394)	<b>0.956</b>	(0.878)	8.989	(0.161)	-4.438	(-12.237)	8.991	(0.176)	30.060	(0.571)
LME	2.446	(2.387)	2.061	(0.451)	16.478	(13.664)	2.180	(0.534)	1.220	(1.180)	0.902	(0.210)	22.026	(8.025)	0.951	(0.242)	3.131	(0.776)
MDE.CvM	2.170	(2.105)	2.293	(0.532)	<b>3.314</b>	(0.242)	2.297	(0.532)	1.056	(1.015)	0.959	(0.222)	5.580	(1.477)	0.962	(0.222)	3.260	(0.753)
MDE.SD	1.995	(1.939)	1.989	(0.383)	-5.018	(-7.673)	<b>2.000</b>	(0.409)	0.953	(0.917)	<b>0.843</b>	(0.161)	-4.690	(-8.279)	<b>0.845</b>	(0.168)	<b>2.846</b>	(0.576)
MDE. $\chi^2$	1.957	(1.527)	47.611	(0.833)	-6.787	(-27.277)	47.631	(1.161)	1.853	(1.411)	50.949	(0.838)	85.315	(41.077)	51.677	(1.007)	99.307	(2.168)
MDE.TV	2.352	(1.944)	69.816	(1.653)	21.518	(-7.430)	70.020	(1.678)	2.357	(1.730)	78.235	(1.673)	135.743	(73.041)	77.077	(2.206)	147.097	(3.884)
MDE.KL	2.380	(2.290)	4.145	(0.585)	13.316	(9.028)	4.223	(0.621)	1.197	(1.135)	1.947	(0.272)	19.704	(13.518)	1.986	(0.290)	6.208	(0.911)
MME	3.625	(3.475)	29.416	(1.294)	72.610	(65.458)	31.741	(3.184)	2.243	(2.140)	12.877	(0.618)	124.317	(113.988)	14.422	(1.917)	46.163	(5.101)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table A2.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 2$  and  $n = 100$ ).

n = 100																		
Method	$\beta = 1.1$				$\sigma = 2$													
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	TMSE						
MLE	1.211	(1.198)	0.079	(0.060)	10.063	(8.923)	0.091	(0.070)	2.236	(2.198)	0.720	(0.557)	11.809	(9.916)	0.775	(0.596)	0.866	(0.666)
MLEb	<b>1.116</b>	<b>(1.106)</b>	<b>0.052</b>	<b>(0.042)</b>	<b>1.424</b>	<b>(0.890)</b>	<b>0.052</b>	<b>(0.042)</b>	1.941	(1.914)	<b>0.460</b>	<b>(0.368)</b>	-2.949	(-4.322)	<b>0.463</b>	<b>(0.376)</b>	<b>0.516</b>	<b>(0.418)</b>
LME	1.477	(1.465)	0.074	(0.060)	34.254	(33.208)	0.216	(0.194)	3.043	(3.001)	0.800	(0.663)	52.134	(50.309)	1.887	(1.676)	2.104	(1.870)
MDECvM	1.197	(1.184)	0.108	(0.089)	8.857	(7.626)	0.117	(0.096)	2.196	(2.156)	0.885	(0.721)	9.798	(7.797)	0.923	(0.746)	1.040	(0.842)
MDESd	1.137	(1.125)	0.088	(0.074)	3.277	(2.292)	0.090	(0.074)	<b>2.044</b>	<b>(2.008)</b>	0.727	(0.597)	<b>2.190</b>	<b>(0.418)</b>	0.729	(0.597)	0.819	(0.672)
MDE <sub>v</sub> <sup>2</sup>	-1.923	(-1.961)	308.580	(311.545)	-274.775	(-278.280)	317.714	(320.913)	3.408	(3.105)	15.342	(3.025)	70.386	(55.254)	17.324	(4.246)	335.038	(325.159)
MDETV	0.772	(0.751)	0.249	(0.101)	-29.833	(-31.699)	0.357	(0.223)	3.532	(3.427)	6.016	(2.313)	76.586	(71.367)	8.362	(4.351)	8.719	(4.574)
MDEKL	1.211	(1.198)	0.079	(0.060)	10.063	(8.923)	0.091	(0.070)	2.236	(2.198)	0.720	(0.557)	11.809	(9.916)	0.775	(0.596)	0.866	(0.666)
MME	2.429	(2.411)	0.136	(0.100)	120.913	(119.167)	1.902	(1.818)	10.919	(10.420)	52.861	(12.498)	445.964	(421.005)	132.414	(83.396)	134.316	(85.214)
n = 100																		
Method	$\beta = 1.5$				$\sigma = 2$													
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	TMSE						
MLE	1.622	(1.596)	0.365	(0.176)	8.129	(6.419)	0.380	(0.186)	2.252	(2.200)	1.531	(0.745)	12.610	(10.023)	1.595	(0.785)	1.975	(0.971)
MLEb	1.437	(1.419)	0.260	(0.094)	-4.209	(-5.387)	0.264	(0.101)	1.864	(1.828)	1.073	(0.382)	-6.787	(-8.598)	1.091	(0.411)	1.355	(0.512)
LME	1.821	(1.800)	0.253	(0.166)	21.433	(20.026)	0.356	(0.256)	2.625	(2.582)	1.089	(0.741)	31.259	(29.118)	1.480	(1.080)	1.836	(1.336)
MDECvM	1.546	(1.525)	0.281	(0.196)	3.070	(1.693)	0.284	(0.197)	2.101	(2.059)	1.077	(0.731)	5.048	(2.986)	1.087	(0.734)	1.370	(0.931)
MDESd	1.451	(1.435)	0.231	(0.155)	-3.255	(-4.447)	0.233	(0.159)	1.935	(1.996)	0.888	(0.580)	-3.268	(-5.115)	0.892	(0.591)	1.125	(0.750)
MDE <sub>v</sub> <sup>2</sup>	0.846	(0.801)	16.560	(15.088)	-43.601	(-46.627)	16.987	(15.577)	2.898	(2.744)	16.697	(1.679)	44.882	(37.210)	17.503	(2.233)	34.490	(17.810)
MDETV	1.230	(1.155)	3.956	(0.326)	-18.010	(-23.157)	4.029	(0.446)	3.539	(3.296)	39.080	(3.168)	76.964	(64.807)	41.449	(4.848)	45.478	(5.294)
MDEKL	1.622	(1.596)	0.365	(0.176)	8.130	(6.420)	0.380	(0.186)	2.252	(2.200)	1.531	(0.745)	12.611	(10.024)	1.595	(0.785)	1.975	(0.971)
MME	2.794	(2.756)	0.669	(0.344)	86.259	(83.724)	2.343	(1.921)	6.122	(6.023)	4.008	(2.145)	206.117	(201.172)	21.001	(18.353)	23.344	(20.254)
n = 100																		
Method	$\beta = 2$				$\sigma = 2$													
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	TMSE						
MLE	2.380	(2.154)	1376.246	(1.501)	18.987	(7.707)	1376.381	(0.524)	2.525	(2.223)	2299.944	(1.004)	26.244	(11.132)	2300.204	(1.054)	3676.385	(1.278)
MLEb	2.041	(1.812)	1376.175	(0.275)	2.074	(-9.377)	1376.168	(0.310)	2.039	(1.732)	2299.762	(0.821)	1.658	(-13.348)	2299.748	(0.365)	3675.916	(0.903)
LME	2.333	(2.278)	1.505	(0.394)	16.671	(13.898)	1.616	(0.472)	2.433	(2.356)	3.045	(0.796)	21.630	(17.812)	3.233	(0.923)	4.849	(1.394)
MDECvM	2.015	(1.959)	1.566	(0.394)	<b>0.744</b>	(-2.068)	1.566	(0.395)	2.404	(1.957)	3.074	(0.716)	1.724	(-2.166)	3.075	(0.718)	4.642	(1.114)
MDESd	1.863	(1.814)	1.379	(0.294)	-6.852	(-9.295)	1.398	(0.328)	1.847	(1.779)	2.752	(0.541)	-7.642	(-11.055)	2.776	(0.590)	4.174	(0.918)
MDE <sub>v</sub> <sup>2</sup>	1.725	(1.413)	28.719	(1.087)	-13.764	(-29.353)	28.795	(1.431)	3.409	(2.779)	115.965	(2.995)	70.431	(38.966)	117.948	(3.603)	146.743	(5.034)
MDETV	2.258	(1.802)	43.929	(1.343)	12.906	(-9.911)	43.995	(1.382)	4.354	(3.409)	191.143	(5.892)	117.721	(70.640)	196.684	(7.878)	240.680	(9.260)
MDEKL	2.241	(2.154)	3.589	(0.501)	12.050	(7.708)	3.647	(0.524)	2.342	(2.223)	6.888	(1.004)	17.088	(11.133)	7.005	(1.054)	10.652	(1.578)
MME	3.619	(3.350)	1768.602	(1.110)	80.940	(67.495)	1771.211	(2.953)	4.779	(4.419)	2951.022	(2.270)	138.930	(320.933)	2958.723	(8.120)	4729.933	(11.052)
n = 100																		
Method	$\beta = 2.1$				$\sigma = 2$													
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	TMSE						
MLE	2.420	(2.270)	72.245	(0.606)	15.245	(8.116)	72.347	(0.635)	2.427	(2.231)	120.427	(1.071)	21.347	(11.558)	120.609	(1.124)	192.956	(1.759)
MLEb	2.052	(1.899)	72.134	(0.382)	-2.286	(-9.593)	72.136	(0.422)	1.931	(1.730)	120.201	(0.643)	-3.474	(-13.514)	120.205	(0.716)	192.341	(1.139)
LME	2.446	(2.377)	2.152	(0.453)	16.473	(13.178)	2.272	(0.529)	2.421	(2.332)	3.693	(0.804)	21.034	(16.594)	3.870	(0.915)	6.142	(1.444)
MDECvM	2.112	(2.039)	2.263	(0.442)	<b>0.591</b>	(-2.887)	2.263	(0.446)	2.263	(1.937)	3.808	(0.717)	1.546	(-3.168)	3.809	(0.721)	6.073	(1.166)
MDESd	1.947	(1.884)	2.012	(0.326)	-7.264	(-10.295)	2.036	(0.375)	1.839	(1.756)	3.420	(0.536)	-8.056	(-12.192)	3.446	(0.595)	5.481	(0.968)
MDE <sub>v</sub> <sup>2</sup>	1.958	(1.531)	42.717	(1.002)	-6.750	(-27.114)	42.737	(1.326)	3.624	(2.815)	153.838	(3.423)	81.189	(40.747)	156.474	(4.087)	199.211	(5.413)
MDETV	2.588	(1.958)	66.640	(1.847)	13.240	(-6.783)	66.878	(1.868)	4.681	(3.469)	250.358	(7.044)	134.064	(73.445)	257.545	(9.202)	324.423	(11.070)
MDEKL	2.381	(2.270)	4.894	(0.606)	13.358	(8.117)	4.973	(0.635)	2.374	(2.231)	8.788	(1.071)	18.711	(11.559)	8.928	(1.124)	13.901	(1.759)
MME	3.680	(3.486)	101.650	(1.352)	75.219	(66.021)	104.144	(2.274)	4.528	(4.275)	172.607	(2.426)	126.379	(313.741)	178.995	(7.603)	283.139	(10.877)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table A3.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 1$  and  $n = 200$ ).

$n = 200$																		
Method	$\beta = 1.1$						$\beta = 1$						MSE	TMSE				
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	1.147	(1.139)	0.033	(0.028)	4.232	(3.583)	0.036	(0.030)	1.073	(1.061)	0.080	(0.067)	7.256	(6.134)	0.085	(0.071)	0.120	(0.101)
MLE <sub>b</sub>	1.106	(1.099)	0.028	(0.024)	0.512	(-0.072)	0.028	(0.024)	1.007	(0.997)	0.066	(0.056)	0.464	(-0.336)	0.066	(0.056)	0.095	(0.081)
LME	1.365	(1.357)	0.036	(0.030)	24.048	(23.385)	0.106	(0.097)	1.430	(1.418)	0.106	(0.092)	43.011	(41.801)	0.291	(0.266)	0.396	(0.363)
MDE <sub>CV</sub> M	1.157	(1.147)	0.064	(0.054)	5.163	(4.251)	0.067	(0.056)	1.085	(1.071)	0.130	(0.110)	8.507	(7.075)	0.137	(0.115)	0.204	(0.171)
MDE <sub>SD</sub>	1.128	(1.119)	0.058	(0.049)	2.561	(1.705)	0.058	(0.049)	1.048	(1.034)	0.118	(0.100)	4.774	(3.426)	0.120	(0.101)	0.179	(0.151)
MDE <sub>v<sup>2</sup></sub>	0.409	(0.400)	15.554	(15.703)	-62.774	(-63.594)	16.031	(16.192)	1.299	(1.233)	2.178	(0.103)	29.851	(23.340)	2.267	(0.158)	18.298	(16.350)
MDE <sub>TV</sub>	0.648	(0.638)	0.050	(0.041)	-41.109	(-41.962)	0.254	(0.254)	1.408	(1.386)	0.213	(0.167)	40.752	(38.647)	0.379	(0.317)	0.633	(0.571)
MDE <sub>KL</sub>	1.147	(1.139)	0.033	(0.028)	4.231	(3.582)	0.036	(0.030)	1.073	(1.061)	0.079	(0.067)	7.252	(6.133)	0.085	(0.071)	0.120	(0.101)
MME	2.234	(2.225)	0.041	(0.032)	103.100	(102.234)	1.328	(1.297)	6.674	(5.539)	148721.693	(3.235)	567.367	(453.901)	148752.893	(28.838)	148754.220	(25.135)
$n = 200$																		
Method	$\beta = 1.5$						$\beta = 1$						MSE	TMSE				
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	1.569	(1.555)	0.099	(0.080)	4.603	(3.684)	0.104	(0.083)	1.073	(1.059)	0.105	(0.084)	7.343	(5.920)	0.110	(0.087)	0.214	(0.170)
MLE <sub>b</sub>	1.487	(1.476)	0.075	(0.062)	-0.846	(-1.598)	0.075	(0.063)	0.988	(0.976)	0.078	(0.064)	-1.222	(-2.379)	0.078	(0.065)	0.153	(0.127)
LME	1.710	(1.697)	0.109	(0.092)	13.986	(13.100)	0.153	(0.130)	1.208	(1.194)	0.119	(0.100)	20.774	(19.392)	0.162	(0.138)	0.316	(0.268)
MDE <sub>CV</sub> M	1.576	(1.561)	0.158	(0.137)	5.041	(4.052)	0.163	(0.140)	1.078	(1.063)	0.150	(0.128)	7.800	(6.287)	0.157	(0.132)	0.320	(0.273)
MDE <sub>SD</sub>	1.526	(1.512)	0.139	(0.121)	1.720	(0.810)	0.139	(0.121)	1.034	(1.020)	0.133	(0.114)	3.412	(2.013)	0.134	(0.114)	0.273	(0.235)
MDE <sub>v<sup>2</sup></sub>	0.857	(0.839)	0.243	(0.091)	-42.851	(-44.069)	0.656	(0.528)	1.132	(1.106)	0.557	(0.160)	13.195	(10.592)	0.574	(0.171)	1.230	(0.699)
MDE <sub>TV</sub>	1.014	(0.995)	0.282	(0.131)	-32.367	(-33.678)	0.518	(0.387)	1.266	(1.238)	0.610	(0.220)	26.552	(23.753)	0.680	(0.277)	1.198	(0.663)
MDE <sub>KL</sub>	1.569	(1.555)	0.099	(0.080)	4.603	(3.684)	0.104	(0.083)	1.073	(1.059)	0.105	(0.084)	7.343	(5.920)	0.110	(0.087)	0.214	(0.170)
MME	2.506	(2.487)	0.170	(0.132)	67.047	(65.775)	1.182	(1.105)	2.745	(2.704)	0.453	(0.278)	174.452	(170.418)	3.496	(3.182)	4.678	(4.287)
$n = 200$																		
Method	$\beta = 2$						$\beta = 1$						MSE	TMSE				
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	2.121	(2.096)	0.282	(0.216)	6.037	(4.797)	0.296	(0.225)	1.087	(1.069)	0.143	(0.109)	8.682	(6.896)	0.150	(0.113)	0.446	(0.339)
MLE <sub>b</sub>	1.954	(1.937)	0.176	(0.143)	-2.303	(-3.160)	0.178	(0.147)	0.967	(0.955)	0.088	(0.070)	-3.276	(-4.510)	0.088	(0.072)	0.267	(0.219)
LME	2.220	(2.198)	0.261	(0.214)	10.993	(9.910)	0.310	(0.253)	1.146	(1.130)	0.132	(0.107)	14.572	(13.015)	0.153	(0.124)	0.463	(0.377)
MDE <sub>CV</sub> M	2.069	(2.046)	0.364	(0.308)	3.456	(2.279)	0.368	(0.310)	1.050	(1.034)	0.164	(0.138)	5.004	(3.298)	0.167	(0.140)	0.535	(0.450)
MDE <sub>SD</sub>	1.987	(1.966)	0.307	(0.263)	-0.666	(-1.718)	0.307	(0.254)	0.999	(0.985)	0.139	(0.119)	-0.096	(-1.541)	0.139	(0.119)	0.446	(0.383)
MDE <sub>v<sup>2</sup></sub>	1.360	(1.310)	1.481	(0.294)	-32.004	(-34.481)	1.890	(0.770)	1.191	(1.143)	1.433	(0.255)	19.146	(14.270)	1.470	(0.275)	3.360	(1.045)
MDE <sub>TV</sub>	1.570	(1.527)	0.962	(0.351)	-21.498	(-23.634)	1.147	(0.575)	1.288	(1.247)	0.920	(0.290)	28.837	(24.678)	1.003	(0.351)	2.150	(0.926)
MDE <sub>KL</sub>	2.121	(2.096)	0.282	(0.216)	6.037	(4.798)	0.296	(0.225)	1.087	(1.069)	0.143	(0.109)	8.682	(6.897)	0.150	(0.113)	0.446	(0.339)
MME	3.004	(2.969)	0.540	(0.405)	50.209	(48.437)	1.548	(1.343)	1.913	(1.887)	0.285	(0.210)	91.340	(88.721)	1.119	(0.998)	2.667	(2.341)
$n = 200$																		
Method	$\beta = 2.1$						$\beta = 1$						MSE	TMSE				
	Mean	Var	RB	MSE	Mean	Var	RB	MSE	Mean	Var	RB	MSE						
MLE	2.234	(2.207)	0.337	(0.256)	6.366	(5.076)	0.355	(0.267)	1.091	(1.072)	0.151	(0.114)	9.059	(7.221)	0.160	(0.119)	0.515	(0.386)
MLE <sub>b</sub>	2.045	(2.026)	0.208	(0.162)	-2.621	(-3.511)	0.211	(0.167)	0.963	(0.951)	0.092	(0.071)	-3.654	(-4.926)	0.093	(0.073)	0.304	(0.240)
LME	2.323	(2.300)	0.303	(0.246)	10.634	(9.525)	0.353	(0.286)	1.139	(1.123)	0.135	(0.109)	13.907	(12.330)	0.154	(0.124)	0.507	(0.410)
MDE <sub>CV</sub> M	2.171	(2.146)	0.424	(0.357)	3.401	(2.172)	0.430	(0.359)	1.049	(1.032)	0.170	(0.142)	4.893	(3.239)	0.172	(0.143)	0.602	(0.502)
MDE <sub>SD</sub>	2.081	(2.058)	0.355	(0.302)	-0.903	(-1.994)	0.355	(0.303)	0.996	(0.981)	0.143	(0.121)	-0.384	(-1.864)	0.143	(0.121)	0.499	(0.424)
MDE <sub>v<sup>2</sup></sub>	1.474	(1.414)	2.263	(0.366)	-29.827	(-32.678)	2.655	(0.837)	1.211	(1.156)	2.027	(0.280)	21.123	(15.604)	2.072	(0.304)	4.727	(1.141)
MDE <sub>TV</sub>	1.691	(1.638)	1.696	(0.417)	-19.454	(-22.012)	1.863	(0.631)	1.302	(1.253)	1.422	(0.305)	30.154	(25.252)	1.513	(0.369)	3.375	(0.999)
MDE <sub>KL</sub>	2.234	(2.207)	0.337	(0.256)	6.367	(5.076)	0.355	(0.267)	1.091	(1.072)	0.151	(0.114)	9.060	(7.222)	0.160	(0.119)	0.515	(0.386)
MME	3.115	(3.076)	0.647	(0.484)	48.330	(46.480)	1.677	(1.436)	1.843	(1.816)	0.297	(0.220)	84.276	(81.618)	1.007	(0.886)	2.685	(2.323)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table A4.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 2$  and  $n = 200$ ).

$n = 200$																		
Method	$\beta = 1.1$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.155	(1.148)	0.032	(0.027)	4.976	(4.352)	0.035	(0.029)	2.129	(2.101)	0.302	(0.257)	6.442	(5.370)	0.319	(0.269)	0.354	(0.298)
MLEb	1.113	(1.107)	<b>0.027</b>	<b>(0.023)</b>	<b>1.209</b>	<b>(0.648)</b>	<b>0.027</b>	<b>(0.023)</b>	<b>1.998</b>	<b>(1.978)</b>	<b>0.251</b>	<b>(0.216)</b>	<b>-0.122</b>	<b>(-1.079)</b>	<b>0.251</b>	<b>(0.216)</b>	<b>0.278</b>	<b>(0.240)</b>
LME	1.377	(1.370)	0.036	(0.031)	25.210	(24.580)	0.113	(0.104)	2.841	(2.819)	0.421	(0.371)	42.071	(40.930)	1.129	(1.041)	1.241	(1.145)
MDEcVM	1.158	(1.149)	0.055	(0.048)	5.231	(4.435)	0.059	(0.050)	2.134	(2.108)	0.456	(0.393)	6.683	(5.398)	0.474	(0.405)	0.533	(0.455)
MDESD	1.129	(1.121)	0.050	(0.044)	2.655	(1.906)	0.051	(0.044)	2.261	(2.037)	0.415	(0.358)	3.050	(1.836)	0.419	(0.360)	0.470	(0.404)
MDE <sub>v</sub> <sup>2</sup>	-0.708	(-0.728)	138.432	(139.790)	-164.363	(-166.183)	141.700	(143.131)	2.874	(2.516)	15.770	(0.960)	43.687	(25.825)	16.533	(1.227)	158.233	(144.357)
MDETV	0.663	(0.654)	0.050	(0.042)	-39.749	(-40.561)	0.241	(0.241)	2.798	(2.759)	0.813	(0.655)	39.921	(37.942)	1.450	(1.231)	1.691	(1.472)
MDEKL	1.155	(1.148)	0.032	(0.027)	4.976	(4.352)	0.035	(0.029)	2.129	(2.107)	0.302	(0.257)	6.441	(5.370)	0.319	(0.269)	0.354	(0.298)
MME	2.248	(2.238)	0.042	(0.033)	104.324	(103.488)	1.359	(1.329)	11.479	(10.883)	78.896	(18.161)	473.937	(444.125)	168.742	(97.060)	170.100	(98.389)
$\beta = 1.5$																		
Method	$\beta = 1.5$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.561	(1.548)	0.094	(0.077)	4.037	(3.172)	0.098	(0.079)	2.125	(2.098)	0.395	(0.323)	6.238	(4.906)	0.411	(0.333)	0.508	(0.412)
MLEb	1.480	(1.469)	<b>0.072</b>	<b>(0.060)</b>	-1.335	(-2.048)	<b>0.072</b>	<b>(0.061)</b>	1.956	(1.935)	<b>0.285</b>	<b>(0.248)</b>	-2.179	(-3.270)	<b>0.297</b>	<b>(0.253)</b>	<b>0.369</b>	<b>(0.314)</b>
LME	1.716	(1.705)	0.104	(0.089)	14.387	(13.552)	0.151	(0.130)	2.423	(2.397)	0.456	(0.390)	21.169	(19.869)	0.636	(0.548)	0.786	(0.678)
MDEcVM	1.539	(1.526)	0.130	(0.113)	2.611	(1.727)	0.131	(0.113)	2.081	(2.054)	0.483	(0.415)	4.034	(2.714)	0.490	(0.418)	0.621	(0.531)
MDESD	1.492	(1.486)	0.115	(0.100)	-0.538	(-1.356)	0.115	(0.100)	1.998	(1.974)	0.429	(0.370)	-0.995	(-1.323)	0.429	(0.371)	0.543	(0.471)
MDE <sub>v</sub> <sup>2</sup>	0.829	(0.813)	3.953	(3.960)	-44.710	(-45.806)	4.403	(4.433)	2.284	(2.229)	1.367	(0.636)	14.205	(11.433)	1.447	(0.688)	5.850	(5.121)
MDETV	1.023	(1.005)	0.163	(0.128)	-31.817	(-33.012)	0.390	(0.373)	2.546	(2.495)	1.152	(0.869)	27.308	(24.760)	1.451	(1.114)	1.841	(1.487)
MDEKL	1.561	(1.548)	0.094	(0.077)	4.037	(3.172)	0.098	(0.079)	2.125	(2.098)	0.395	(0.323)	6.238	(4.906)	0.411	(0.333)	0.508	(0.412)
MME	2.517	(2.498)	0.163	(0.130)	67.767	(66.563)	1.197	(1.127)	5.469	(5.388)	1.920	(0.996)	173.472	(169.404)	13.957	(12.475)	15.154	(13.602)
$\beta = 2$																		
Method	$\beta = 2$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	2.103	(2.078)	0.282	(0.209)	5.166	(3.883)	0.292	(0.215)	2.148	(2.112)	0.561	(0.420)	7.405	(5.601)	0.583	(0.432)	0.876	(0.648)
MLEb	1.940	(1.922)	<b>0.177</b>	<b>(0.139)</b>	-3.019	(-3.907)	<b>0.181</b>	<b>(0.145)</b>	1.914	(1.889)	<b>0.348</b>	<b>(0.273)</b>	-4.311	(-5.556)	<b>0.355</b>	<b>(0.286)</b>	<b>0.536</b>	<b>(0.430)</b>
LME	2.207	(2.185)	0.263	(0.211)	10.354	(9.236)	0.306	(0.245)	2.272	(2.240)	0.524	(0.422)	13.592	(12.025)	0.598	(0.480)	0.903	(0.725)
MDEcVM	2.015	(1.995)	0.292	(0.244)	<b>0.740</b>	<b>(-0.330)</b>	0.292	(0.245)	2.028	(1.998)	0.535	(0.443)	<b>1.398</b>	<b>(-0.089)</b>	0.536	(0.443)	0.829	(0.687)
MDESD	1.937	(1.918)	0.249	(0.210)	-3.137	(-4.103)	0.253	(0.217)	1.932	(1.905)	0.458	(0.383)	-3.382	(-4.731)	0.463	(0.392)	0.715	(0.609)
MDE <sub>v</sub> <sup>2</sup>	1.355	(1.303)	1.670	(0.298)	-32.250	(-34.857)	2.086	(0.784)	2.374	(2.276)	5.678	(1.028)	18.699	(13.776)	5.818	(1.104)	7.904	(1.888)
MDETV	1.563	(1.518)	1.031	(0.354)	-21.848	(-24.103)	1.222	(0.586)	2.568	(2.482)	3.462	(1.166)	28.383	(24.109)	3.785	(1.399)	5.006	(1.985)
MDEKL	2.103	(2.078)	0.282	(0.209)	5.167	(3.884)	0.292	(0.215)	2.148	(2.112)	0.561	(0.420)	7.406	(5.602)	0.583	(0.432)	0.876	(0.648)
MME	3.001	(2.965)	0.352	(0.410)	50.063	(48.256)	1.554	(1.341)	3.829	(3.777)	1.124	(0.842)	91.429	(88.865)	4.467	(4.001)	6.022	(5.342)
$\beta = 2.1$																		
Method	$\beta = 2.1$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	2.217	(2.188)	0.386	(0.254)	5.592	(4.211)	0.400	(0.262)	2.159	(2.120)	0.685	(0.452)	7.935	(6.011)	0.711	(0.466)	1.110	(0.728)
MLEb	2.031	(2.011)	<b>0.252</b>	<b>(0.161)</b>	-3.279	(-4.222)	<b>0.257</b>	<b>(0.169)</b>	1.908	(1.882)	<b>0.443</b>	<b>(0.281)</b>	-4.596	(-5.909)	<b>0.452</b>	<b>(0.295)</b>	<b>0.709</b>	<b>(0.464)</b>
LME	2.312	(2.288)	0.313	(0.247)	10.091	(8.929)	0.358	(0.282)	2.261	(2.229)	0.552	(0.436)	13.066	(11.454)	0.620	(0.489)	0.978	(0.771)
MDEcVM	2.105	(2.081)	0.341	(0.277)	<b>0.253</b>	<b>(-0.884)</b>	0.341	(0.278)	2.016	(1.985)	0.557	(0.448)	<b>0.801</b>	<b>(-0.754)</b>	0.557	(0.449)	0.898	(0.726)
MDESD	2.021	(2.000)	0.290	(0.237)	-3.748	(-4.774)	0.296	(0.247)	1.918	(1.890)	0.477	(0.386)	-4.088	(-5.498)	0.484	(0.398)	0.780	(0.645)
MDE <sub>v</sub> <sup>2</sup>	1.471	(1.410)	2.315	(0.377)	-29.934	(-32.876)	2.710	(0.853)	2.420	(2.310)	7.146	(1.154)	21.024	(15.484)	7.323	(1.250)	10.033	(2.103)
MDETV	1.685	(1.630)	1.656	(0.422)	-19.747	(-22.369)	1.828	(0.643)	2.597	(2.498)	5.049	(1.234)	29.859	(24.920)	5.406	(1.482)	7.234	(2.126)
MDEKL	2.217	(2.188)	0.386	(0.254)	5.592	(4.211)	0.400	(0.262)	2.159	(2.120)	0.686	(0.452)	7.936	(6.012)	0.711	(0.466)	1.111	(0.728)
MME	3.118	(3.077)	0.734	(0.499)	48.461	(46.528)	1.770	(1.454)	3.695	(3.641)	1.322	(0.901)	84.746	(82.026)	4.194	(3.593)	5.964	(5.047)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table A5.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 1$  and  $n = 500$ ).

$n = 500$																		
Method	$\beta = 1.1$					$\beta = 1$					$\beta = 0.9$							
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE			
MLE	1.117	(1.113)	0.012	(0.010)	1.546	(1.217)	0.012	(0.011)	1.026	(1.021)	0.027	(0.024)	2.640	(2.077)	0.028	(0.025)	0.040	(0.035)
MLE <sub>b</sub>	1.102	(1.098)	0.011	(0.010)	0.165	(-0.154)	0.011	(0.010)	1.002	(0.997)	0.026	(0.023)	0.205	(-0.338)	0.026	(0.023)	0.037	(0.033)
LME	1.288	(1.284)	0.016	(0.015)	17.082	(16.714)	0.052	(0.048)	1.309	(1.302)	0.047	(0.043)	30.900	(30.219)	0.143	(0.134)	0.194	(0.183)
MDE <sub>CV</sub> M	1.122	(1.117)	0.022	(0.019)	2.003	(1.516)	0.022	(0.019)	1.033	(1.025)	0.045	(0.039)	3.275	(2.500)	0.046	(0.039)	0.068	(0.059)
MDE <sub>SD</sub>	1.111	(1.106)	0.021	(0.018)	1.032	(0.555)	0.021	(0.019)	1.019	(1.011)	0.043	(0.037)	1.894	(1.135)	0.043	(0.038)	0.065	(0.056)
MDE <sub>1/2</sub>	0.579	(0.575)	0.009	(0.008)	-47.335	(-47.707)	0.280	(0.283)	1.273	(1.266)	0.037	(0.033)	27.297	(26.642)	0.112	(0.104)	0.392	(0.387)
MDE <sub>TV</sub>	0.652	(0.647)	0.016	(0.014)	-40.725	(-41.147)	0.216	(0.218)	1.351	(1.343)	0.055	(0.049)	35.093	(34.308)	0.178	(0.167)	0.394	(0.385)
MDE <sub>KL</sub>	1.117	(1.113)	0.012	(0.010)	1.545	(1.216)	0.012	(0.011)	1.026	(1.021)	0.027	(0.024)	2.637	(2.076)	0.028	(0.025)	0.040	(0.035)
MME	2.119	(2.115)	0.011	(0.009)	92.666	(92.273)	1.050	(1.039)	6.334	(5.811)	23513.706	(4.323)	533.382	(481.139)	23541.999	(27.472)	23543.048	(28.511)
$n = 500$																		
Method	$\beta = 1.5$					$\beta = 1$					$\beta = 0.5$							
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE			
MLE	1.529	(1.523)	0.032	(0.028)	1.926	(1.503)	0.033	(0.029)	1.031	(1.024)	0.033	(0.029)	3.052	(2.401)	0.034	(0.030)	0.067	(0.058)
MLE <sub>b</sub>	1.499	(1.493)	0.029	(0.026)	-0.053	(-0.453)	0.029	(0.026)	1.000	(0.993)	0.030	(0.027)	-0.044	(-0.659)	0.030	(0.027)	0.059	(0.052)
LME	1.622	(1.615)	0.046	(0.042)	8.133	(7.679)	0.061	(0.055)	1.121	(1.114)	0.049	(0.044)	12.119	(11.408)	0.064	(0.057)	0.125	(0.112)
MDE <sub>CV</sub> M	1.543	(1.533)	0.065	(0.057)	2.841	(2.228)	0.067	(0.059)	1.043	(1.034)	0.061	(0.053)	4.273	(3.271)	0.063	(0.054)	0.130	(0.113)
MDE <sub>SD</sub>	1.524	(1.515)	0.062	(0.055)	1.579	(0.985)	0.063	(0.055)	1.026	(1.017)	0.058	(0.051)	2.618	(1.742)	0.058	(0.051)	0.121	(0.106)
MDE <sub>1/2</sub>	0.860	(0.851)	0.053	(0.045)	-42.690	(-43.278)	0.463	(0.467)	1.024	(1.014)	0.060	(0.049)	2.434	(1.576)	0.061	(0.049)	0.524	(0.516)
MDE <sub>TV</sub>	0.971	(0.963)	0.060	(0.053)	-35.252	(-35.821)	0.340	(0.342)	1.063	(1.053)	0.060	(0.050)	6.278	(5.305)	0.064	(0.053)	0.404	(0.395)
MDE <sub>KL</sub>	1.529	(1.523)	0.032	(0.028)	1.926	(1.503)	0.033	(0.029)	1.031	(1.024)	0.033	(0.029)	3.052	(2.401)	0.034	(0.030)	0.067	(0.058)
MME	2.311	(2.302)	0.051	(0.044)	54.069	(53.487)	0.709	(0.688)	2.492	(2.451)	0.317	(0.312)	149.154	(145.139)	2.542	(2.227)	3.251	(2.915)
$n = 500$																		
Method	$\beta = 2$					$\beta = 1$					$\beta = 0.5$							
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE			
MLE	2.050	(2.039)	0.087	(0.074)	2.491	(1.936)	0.089	(0.076)	1.035	(1.028)	0.043	(0.037)	3.543	(2.756)	0.045	(0.038)	0.134	(0.114)
MLE <sub>b</sub>	1.992	(1.982)	0.075	(0.065)	-0.396	(-0.903)	0.075	(0.065)	0.994	(0.987)	0.037	(0.032)	-0.581	(-1.296)	0.037	(0.033)	0.113	(0.098)
LME	2.112	(2.101)	0.101	(0.090)	5.587	(5.033)	0.114	(0.100)	1.074	(1.066)	0.050	(0.044)	7.410	(6.625)	0.056	(0.049)	0.170	(0.148)
MDE <sub>CV</sub> M	2.056	(2.040)	0.168	(0.145)	2.780	(2.007)	0.171	(0.147)	1.038	(1.028)	0.075	(0.065)	3.828	(2.806)	0.076	(0.066)	0.248	(0.212)
MDE <sub>SD</sub>	2.023	(2.009)	0.157	(0.136)	1.164	(0.425)	0.157	(0.136)	1.018	(1.009)	0.070	(0.061)	1.857	(0.858)	0.070	(0.061)	0.228	(0.197)
MDE <sub>1/2</sub>	1.325	(1.308)	0.178	(0.149)	-33.737	(-34.576)	0.633	(0.627)	1.029	(1.015)	0.115	(0.094)	2.906	(1.468)	0.116	(0.094)	0.749	(0.722)
MDE <sub>TV</sub>	1.451	(1.437)	0.159	(0.138)	-27.426	(-28.163)	0.460	(0.455)	1.021	(1.009)	0.090	(0.075)	2.124	(0.911)	0.090	(0.075)	0.550	(0.530)
MDE <sub>KL</sub>	2.050	(2.039)	0.087	(0.074)	2.491	(1.936)	0.089	(0.076)	1.035	(1.028)	0.043	(0.037)	3.543	(2.756)	0.045	(0.038)	0.134	(0.114)
MME	2.688	(2.672)	0.166	(0.143)	34.384	(33.617)	0.639	(0.595)	1.646	(1.634)	0.088	(0.074)	64.589	(63.430)	0.506	(0.476)	1.145	(1.072)
$n = 500$																		
Method	$\beta = 2.1$					$\beta = 1$					$\beta = 0.5$							
	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE	Mean	Var	RB	MSE	TMSE			
MLE	2.157	(2.145)	0.105	(0.089)	2.738	(2.147)	0.108	(0.091)	1.038	(1.030)	0.046	(0.040)	3.835	(3.011)	0.048	(0.040)	0.155	(0.132)
MLE <sub>b</sub>	2.092	(2.081)	0.089	(0.077)	-0.365	(-0.988)	0.090	(0.078)	0.995	(0.987)	0.039	(0.034)	-0.536	(-1.279)	0.039	(0.034)	0.129	(0.112)
LME	2.215	(2.203)	0.117	(0.103)	5.494	(4.913)	0.130	(0.113)	1.072	(1.064)	0.051	(0.045)	7.188	(6.378)	0.056	(0.049)	0.187	(0.162)
MDE <sub>CV</sub> M	2.163	(2.145)	0.205	(0.176)	2.995	(2.166)	0.209	(0.178)	1.041	(1.030)	0.081	(0.070)	4.077	(2.987)	0.083	(0.071)	0.292	(0.249)
MDE <sub>SD</sub>	2.127	(2.113)	0.191	(0.164)	1.293	(0.596)	0.191	(0.164)	1.020	(1.010)	0.076	(0.065)	1.997	(0.960)	0.076	(0.065)	0.267	(0.230)
MDE <sub>1/2</sub>	1.428	(1.409)	0.215	(0.179)	-31.983	(-32.886)	0.666	(0.656)	1.040	(1.025)	0.126	(0.102)	4.031	(2.500)	0.128	(0.103)	0.794	(0.758)
MDE <sub>TV</sub>	1.553	(1.537)	0.185	(0.160)	-26.025	(-26.795)	0.484	(0.477)	1.024	(1.011)	0.095	(0.080)	2.998	(1.146)	0.096	(0.080)	0.580	(0.556)
MDE <sub>KL</sub>	2.157	(2.145)	0.105	(0.089)	2.738	(2.147)	0.108	(0.091)	1.038	(1.030)	0.046	(0.040)	3.835	(3.011)	0.048	(0.040)	0.156	(0.132)
MME	2.777	(2.760)	0.198	(0.171)	32.230	(31.427)	0.657	(0.646)	1.578	(1.566)	0.094	(0.079)	57.773	(56.903)	0.427	(0.399)	1.084	(1.006)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

**Table A6.** Comparison of different estimation methods for different values of  $\beta$  ( $\sigma = 2$  and  $n = 500$ ).

$n = 500$																		
Method	$\beta = 1.1$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.119	(1.115)	<b>0.011</b>	(0.010)	1.710	(1.386)	0.012	( <b>0.010</b> )	2.052	(2.041)	0.108	(0.096)	2.591	(2.039)	0.111	(0.098)	0.122	(0.108)
MLEb	<b>1.104</b>	(1.100)	0.011	(0.010)	<b>0.325</b>	(0.011)	<b>0.011</b>	(0.010)	<b>2.003</b>	(1.992)	<b>0.103</b>	( <b>0.090</b> )	<b>0.155</b>	(-0.377)	<b>0.101</b>	( <b>0.090</b> )	<b>0.112</b>	( <b>0.100</b> )
LME	1.290	(1.286)	0.017	(0.015)	17.252	(16.889)	0.053	(0.050)	2.615	(2.602)	0.196	(0.180)	30.754	(30.090)	0.574	(0.542)	0.627	(0.592)
MDECvM	1.124	(1.119)	0.021	(0.019)	2.199	(1.738)	0.022	(0.019)	2.065	(2.051)	0.174	(0.153)	3.265	(2.527)	0.178	(0.155)	0.199	(0.174)
MDESd	1.114	(1.109)	0.020	(0.018)	1.232	(0.782)	0.021	(0.018)	2.038	(2.023)	0.167	(0.147)	1.894	(1.172)	0.169	(0.148)	0.189	(0.166)
MDE <sub>v</sub> <sup>2</sup>	0.432	(0.426)	14.995	(15.144)	-60.746	(-61.249)	15.442	(15.597)	2.587	(2.511)	3.121	(0.124)	29.333	(25.569)	3.466	(0.385)	18.907	(15.983)
MDETV	0.652	(0.647)	0.017	(0.015)	-40.733	(-41.151)	0.218	(0.220)	2.676	(2.660)	0.211	(0.188)	33.779	(33.003)	0.667	(0.624)	0.885	(0.844)
MDEKL	1.119	(1.115)	0.011	(0.010)	1.710	(1.386)	0.012	(0.010)	2.052	(2.041)	0.108	(0.096)	2.591	(2.038)	0.111	(0.098)	0.122	(0.108)
MME	2.121	(2.117)	0.011	( <b>0.009</b> )	92.851	(92.461)	1.054	(1.043)	12.397	(11.746)	94.988	(24.968)	519.827	(487.399)	203.075	(119.991)	204.129	(121.034)
$\beta = 1.5$																		
Method	$\beta = 1.5$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	1.525	(1.519)	0.031	(0.028)	1.700	(1.279)	0.032	(0.028)	2.053	(2.040)	0.131	(0.115)	2.666	(2.018)	0.134	(0.117)	0.166	(0.145)
MLEb	<b>1.496</b>	(1.490)	<b>0.029</b>	( <b>0.025</b> )	<b>-0.269</b>	(-0.668)	<b>0.029</b>	( <b>0.025</b> )	<b>1.992</b>	(1.978)	<b>0.119</b>	( <b>0.105</b> )	<b>-0.412</b>	(-1.025)	<b>0.119</b>	( <b>0.105</b> )	<b>0.148</b>	( <b>0.131</b> )
LME	1.626	(1.619)	0.045	(0.041)	8.367	(7.914)	0.061	(0.055)	2.249	(2.235)	0.194	(0.175)	12.437	(11.729)	0.256	(0.230)	0.317	(0.285)
MDECvM	1.530	(1.521)	0.058	(0.051)	1.980	(1.416)	0.059	(0.052)	2.060	(2.043)	0.214	(0.188)	2.986	(2.166)	0.217	(0.190)	0.276	(0.242)
MDESd	1.511	(1.505)	0.055	(0.049)	0.737	(0.191)	0.055	(0.049)	2.027	( <b>2.011</b> )	0.204	(0.180)	1.360	(0.563)	0.204	(0.180)	0.260	(0.229)
MDE <sub>v</sub> <sup>2</sup>	0.863	(0.854)	0.053	(0.045)	-42.465	(-43.053)	0.459	(0.462)	2.053	(2.032)	0.242	(0.197)	2.670	(1.612)	0.244	(0.198)	0.703	(0.660)
MDETV	0.975	(0.967)	0.060	(0.053)	-34.974	(-35.541)	0.335	(0.337)	2.131	(2.111)	0.239	(0.202)	6.525	(5.533)	0.256	(0.214)	0.591	(0.552)
MDEKL	1.525	(1.519)	0.031	(0.028)	1.700	(1.279)	0.032	(0.028)	2.053	(2.040)	0.131	(0.115)	2.666	(2.018)	0.134	(0.117)	0.166	(0.145)
MME	2.315	(2.306)	0.051	(0.044)	54.343	(53.762)	0.716	(0.694)	4.369	(4.892)	1.266	(0.441)	148.433	(144.616)	10.079	(8.807)	10.795	(9.501)
$\beta = 2$																		
Method	$\beta = 2$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	2.043	(2.032)	0.085	(0.073)	2.173	(1.620)	0.087	(0.074)	2.062	(2.046)	0.170	(0.146)	3.095	(2.310)	0.174	(0.148)	0.261	(0.222)
MLEb	1.866	(1.976)	<b>0.074</b>	( <b>0.064</b> )	-0.692	(-1.197)	<b>0.074</b>	( <b>0.064</b> )	1.980	(1.966)	<b>0.147</b>	( <b>0.127</b> )	-0.966	(-1.210)	<b>0.147</b>	( <b>0.128</b> )	<b>0.221</b>	( <b>0.192</b> )
LME	2.107	(2.096)	0.101	(0.090)	5.355	(4.801)	0.113	(0.099)	2.141	(2.126)	0.200	(0.177)	7.071	(6.287)	0.220	(0.192)	0.333	(0.291)
MDECvM	2.031	(2.018)	0.140	(0.125)	1.535	(0.907)	0.141	(0.126)	2.044	(2.027)	0.253	(0.225)	2.196	(1.336)	0.255	(0.225)	0.395	(0.351)
MDESd	1.999	(1.987)	0.131	(0.118)	-0.043	(-0.444)	0.131	(0.118)	2.005	(1.989)	0.237	(0.211)	0.254	(-0.572)	0.237	(0.211)	0.368	(0.329)
MDE <sub>v</sub> <sup>2</sup>	1.323	(1.306)	0.179	(0.150)	-33.866	(-34.708)	0.638	(0.632)	2.056	(2.027)	0.465	(0.379)	2.821	(1.375)	0.468	(0.380)	1.106	(1.012)
MDETV	1.448	(1.433)	0.159	(0.138)	-27.609	(-28.342)	0.464	(0.460)	2.040	(2.015)	0.360	(0.301)	1.979	(0.772)	0.361	(0.302)	0.825	(0.761)
MDEKL	2.043	(2.032)	0.085	(0.073)	2.173	(1.620)	0.087	(0.074)	2.062	(2.046)	0.170	(0.146)	3.095	(2.310)	0.174	(0.148)	0.261	(0.222)
MME	2.687	(2.671)	0.168	(0.145)	34.337	(33.567)	0.640	(0.596)	3.294	(3.270)	0.355	(0.299)	64.680	(63.520)	2.028	(1.913)	2.668	(2.509)
$\beta = 2.1$																		
Method	$\beta = 2.1$					$\sigma = 2$					MSE	TMSE						
	Mean	Var	RB	MSE	RB	Mean	Var	RB	MSE									
MLE	2.150	(2.137)	0.101	(0.087)	2.359	(1.778)	0.104	(0.088)	2.066	(2.050)	0.180	(0.154)	3.317	(2.504)	0.184	(0.156)	0.288	(0.245)
MLEb	2.085	(2.074)	<b>0.087</b>	( <b>0.075</b> )	-0.714	(-1.240)	<b>0.087</b>	( <b>0.076</b> )	1.980	(1.965)	<b>0.153</b>	( <b>0.133</b> )	-1.012	(-1.245)	<b>0.154</b>	( <b>0.134</b> )	<b>0.241</b>	( <b>0.210</b> )
LME	2.210	(2.198)	0.116	(0.102)	5.218	(4.645)	0.128	(0.112)	2.136	(2.120)	0.202	(0.178)	6.794	(5.994)	0.221	(0.192)	0.349	(0.304)
MDECvM	2.131	(2.118)	0.164	(0.147)	1.488	(0.845)	0.165	(0.147)	2.043	(2.025)	0.264	(0.234)	2.127	(1.254)	0.264	(0.234)	0.429	(0.381)
MDESd	2.097	(2.084)	0.153	(0.137)	-0.159	(-0.773)	0.153	(0.137)	2.002	(1.986)	0.245	(0.219)	0.117	(-0.718)	0.245	(0.219)	0.398	(0.357)
MDE <sub>v</sub> <sup>2</sup>	1.425	(1.406)	0.216	(0.180)	-32.125	(-33.026)	0.671	(0.661)	2.078	(2.048)	0.506	(0.410)	3.915	(2.388)	0.512	(0.412)	1.183	(1.073)
MDETV	1.549	(1.533)	0.186	(0.160)	-26.222	(-26.988)	0.489	(0.482)	2.044	(2.019)	0.381	(0.319)	2.203	(0.956)	0.383	(0.320)	0.872	(0.801)
MDEKL	2.150	(2.137)	0.101	(0.087)	2.360	(1.778)	0.104	(0.088)	2.066	(2.050)	0.180	(0.154)	3.317	(2.504)	0.184	(0.156)	0.288	(0.245)
MME	2.776	(2.759)	0.200	(0.172)	32.179	(31.378)	0.656	(0.606)	3.157	(3.133)	0.372	(0.318)	57.842	(56.675)	1.711	(1.602)	2.367	(2.209)

Note: Main entries correspond to the untrimmed values; the 1% trimmed values are provided in parentheses.

## References

1. Pareto, V. The new theories of economics. *J. Political Econ.* **1897**, *4*, 485–502. [\[CrossRef\]](#)
2. Lomax, K. Business failures: Another example of the analysis of failure data. *J. Am. Stat. Assoc.* **1954**, *49*, 847–852. [\[CrossRef\]](#)
3. Mitra, D.; Kundu, D.; Balakrishnan, N. Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1019–1043. [\[CrossRef\]](#)
4. Abujarad, M.H.A.; Khan, A.A. JMASM 57: Bayesian Survival Analysis of Lomax Family Models with Stan (R). *J. Mod. Appl. Stat. Methods* **2021**, *19*, 12. [\[CrossRef\]](#)
5. Atkinson, A.B.; Harrison, A.J. *Distribution of Personal Wealth in Britain*; Cambridge University Press: Cambridge, UK, 1978.
6. Harris, C.M. The Pareto distribution as a queue service discipline. *Oper. Res.* **1968**, *2*, 307–308. [\[CrossRef\]](#)
7. Hassan, A.; Al-Ghamdi, A. Optimum step-stress accelerated life testing for Lomax distribution. *J. Appl. Sci. Res.* **2009**, *12*, 2153–2164.
8. Holland, O.; Golaup, A.; Aghvami, A.H. Traffic characteristics of aggregated module downloads for mobile terminal reconfiguration. *IEE Proc.* **2006**, *135*, 683–690. [\[CrossRef\]](#)
9. Hosking, J.R.M.; Wallis, J.R. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **1987**, *29*, 339–349. [\[CrossRef\]](#)
10. de Zea Bermudez, P.; Kotz, S. Parameter estimation of the generalized Pareto distribution—Part I. *J. Stat. Plan. Inference* **2010**, *140*, 1353–1373. [\[CrossRef\]](#)
11. Ramos, P.L.; Nascimento, D.C.; Ferreira, P.H.; Weber, K.T.; Santos, T.E.G.; Louzada, F. Modelling traumatic brain injury lifetime data: Improved estimators for the generalised gamma distribution under small samples. *PLoS ONE* **2019**, *14*, e0221332. [\[CrossRef\]](#)
12. Teimouri, M.; Hoseini, S.M.; Nadarajah, S. Comparison of estimation methods for the Weibull distribution. *Statistics* **2013**, *47*, 93–109. [\[CrossRef\]](#)
13. Dey, S.; Menezes, F.B.; Mazucheli, J. Comparison of estimation methods for unit-gamma distribution. *J. Data Sci.* **2019**, *17*, 768–801. [\[CrossRef\]](#)
14. Giles, D.E.; Feng, H.; Godwin, R.T. On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution. *Commun. Stat. Methods* **2013**, *42*, 1934–1950. [\[CrossRef\]](#)
15. Shakeel, M.; Rehmat, N.; Haq, M.A. Comparison of the robust parameters estimation methods for the two-parameters Lomax distribution. *Cogent Math.* **2017**, *4*, 3–5. [\[CrossRef\]](#)
16. Juarez, S.F.; Schucany, W.R. Robust and efficient estimation for the generalized Pareto distribution. *Extremes* **2004**, *7*, 237–251. [\[CrossRef\]](#)
17. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559. [\[CrossRef\]](#)
18. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; Chapman and Hall: Boca Raton, FL, USA, 2011.
19. Boos, D. Minimum distance estimators for location and goodness of fit. *J. Am. Stat. Assoc.* **1981**, *76*, 663–670. [\[CrossRef\]](#)
20. Parr, W.C.; Schucany, W.R. Minimum distance and robust estimation. *J. Am. Stat. Assoc.* **1980**, *75*, 616–624. [\[CrossRef\]](#)
21. Parr, W.C.; De Wet, T. On minimum Cramer-von Mises-norm parameter estimation. *Commun. Stat. Methods* **1981**, *10*, 1149–1166. [\[CrossRef\]](#)
22. Ferreira, P.H.; Ramos, R.; Ramos, P.L.; Gonzales, J.F.B.; Tomazella, V.L.D.; Ehlers, R.S.; Silva, E.B.; Louzada, F. Objective Bayesian analysis for the Lomax distribution. *Stat. Probab. Lett.* **2020**, *159*, 108677. [\[CrossRef\]](#)
23. He, D.; Sun, D.; Zhu, Q. Bayesian analysis for the Lomax model using noninformative priors. *Stat. Theory Relat. Fields* **2022**, 1–8. [\[CrossRef\]](#)
24. Son, Y.S.; Oh, M. Bayesian estimation of the two-parameter gamma distribution. *Commun. Stat. Comput.* **2006**, *35*, 285–293. [\[CrossRef\]](#)
25. Hosking, J.R.M. *L*-moments: Analysis and estimation of distributions using linear combinations of order statistic. *J. R. Stat. Soc.* **1990**, *52*, 105–124. [\[CrossRef\]](#)
26. Hosking, J.R.M. Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *J. R. Stat. Society. Ser. C Appl. Stat.* **1985**, *34*, 301–310.
27. Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C.; Wallis, J.R. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* **1979**, *52*, 105–124.
28. Asquith, W. *Univariate Distributional Analysis with L-moment Statistics Using R*; Create Space Independent Platform: Scotts Valley, CA, USA, 2011.
29. Deville, Y. *Renext: Renewal Method for Extreme Values Extrapolation*; R package version 3.1-0; Institut de Radioprotection et de Sûreté Nucléaire: Fontenay-aux-Roses, France, 2016.
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
31. Cheng, R.; Amin, N. Maximum product of spacings estimation with application to the lognormal distribution. *Math. Rep.* **1979**, *1*, 79.
32. Ranney, B. The maximum spacing method. An estimation method related to the maximum likelihood method. *Scand. J. Stat.* **1984**, *11*, 92–112.
33. Wolfowitz, J. Estimation by the minimum distance method. *Ann. Inst. Stat. Math.* **1953**, *5*, 9–23.

34. Beran, R.J. Minimum Pareto distance estimates for parameter models. *Ann. Stat.* **1977**, *5*, 445–463. [[CrossRef](#)]
35. Beran, R.J. An efficient and robust adaptive estimator of location. *Ann. Stat.* **1978**, *6*, 292–313.
36. Broyden, C.G. The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Its Appl.* **1970**, *6*, 76–90. [[CrossRef](#)]
37. Fletcher, R. A New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13*, 317–322. [[CrossRef](#)]
38. Goldfarb, D. A Family of Variable Metric Updates Derived by Variational Means. *Math. Comput.* **1970**, *24*, 23–26. [[CrossRef](#)]
39. Shanno, D.F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **1970**, *24*, 647–656. [[CrossRef](#)]
40. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman & Hall, Inc.: London, UK, 1986.
41. Von Alven, W.H. *Reliability Engineering*; Prentice Hall: Hoboken, NJ, USA, 1964.
42. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall, Inc.: New York, NY, USA, 1993.
43. Hu, Y.; Scarrott, C. evmix: An R package for Extreme Value Mixture Modeling, Threshold Estimation and Boundary Corrected Kernel Density Estimation. *J. Stat. Softw.* **2018**, *84*, 1–27. [[CrossRef](#)]
44. Liebscher, E.; Taubert, F.; Hetzer, J. Modelling multivariate data using product copulas and minimum distance estimators: An exemplary application to ecological traits. *Environ. Ecol. Stat.* **2022**, *29*, 1–24. [[CrossRef](#)]
45. Weiss, G. Copula parameter estimation by maximum-likelihood and minimum-distance estimator: A simulation study. *Comput. Stat.* **2011**, *26*, 31–54. [[CrossRef](#)]
46. Engle, R.F.; Russell, J.R. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **1998**, *66*, 1127–1162. [[CrossRef](#)]
47. Thas, O. *Comparing Distributions*; Springer: New York, NY, USA, 2010.
48. Meintanis, S.G.; Milosevic, B.; Obradovic, M. Goodness-of-fit tests in conditional duration models. *Stat. Pap.* **2020**, *61*, 123–140. [[CrossRef](#)]
49. Bauwens, L.; Giot, P. *Econometric Modelling of Stock Market Intraday Activity*; Kluwer Academic Publisher: Boston, MA, USA, 2001.
50. Bauwens, L.; Giot, P.; Grammig, J.; Veredas, D. A comparison of financial duration models via density forecasts. *Int. J. Forecast.* **2005**, *20*, 589–609. [[CrossRef](#)]
51. De Luga, G.; Gallo, G.M. Mixture processes for financial intradaily durations. *Stud. Nonlinear Dyn. Econom.* **2004**, *8*. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

# Chapter 5

## Conclusion

This chapter contains some conclusions as well as directions for future research. We begin by considering the goals of the papers contained in the thesis.

### 5.1 Overview of the goals of the papers presented

- **To develop new classes of goodness-of-fit tests for the Pareto Type I distribution.** New classes of goodness-of-fit tests are developed in both Chapters 2 and 3. In the former, we exploit a characterisation involving order statistics, while the tests proposed in the latter are based on a characterisation involving conditional expectations.
- **To derive the asymptotic properties pertaining to some of these new classes of goodness-of-fit tests.** The asymptotic properties of the tests based on the order statistics characterisation of Chapter 2 are extensively studied. The asymptotic null distributions of the tests are derived and consistency against a wide class of fixed alternatives is shown.
- **To critically evaluate the performance of the newly proposed tests under different estimation methods using Monte Carlo simulation.** Extensive Monte Carlo studies are carried out in both Chapters 2 and 3 to examine the numerical sizes and powers, against a wide range of fixed alternative distributions, achieved by the newly proposed tests. The calculated powers are compared to those of existing tests for the Pareto distribution and we find that the newly proposed tests are competitive, often outperforming the other tests considered.
- **To apply the newly proposed tests to real-world data settings and draw conclusions.** The newly proposed goodness-of-fit tests based on a characterisation involving a conditional expectation are applied to the observed annual salaries of English Premier League (EPL) football players for the seasons 2021–2022 as well as 2022–2023. The aim is to determine whether or not the Pareto distribution is

an appropriate model for the distribution of the salaries of players in excess of 10 million GBP in Chapter 3. We find that the Pareto distribution is an appropriate model for salaries in one of the seasons, while this is not the case for the other season. Furthermore, the newly proposed tests based on an order statistic characterisation are applied to test whether or not observed LIV golf earnings, in excess of 3.5 million USD, are realised from the Pareto distribution in Chapter 2. The results indicate that the Pareto distribution is an appropriate model for the data.

- **To investigate different estimation methods used to estimate the shape and scale parameters of the Lomax distribution as well as determine whether the minimum distance estimators are suitable estimation alternatives for this distribution when compared to traditional methods of estimation.** The different estimation methods are compared to one another in a comprehensive Monte Carlo study. This study, included in Chapter 4, considers various parameter combinations as well as a range of sample sizes. The results obtained indicate that, while a bias reduced version of maximum likelihood estimation is a good choice for estimation using large samples, a minimum distance estimator is recommended for use with smaller samples.

## 5.2 Concluding remarks and future research

The aim of this thesis is to contribute to the literature regarding the Pareto distributions, specifically that of the Pareto Type I and Lomax distributions. We propose several classes of goodness-of-fit tests for the Pareto Type I distribution and we demonstrate that, in addition to desirable asymptotic properties, the tests perform well in finite sample settings. Additionally, parameter estimation for the Lomax distribution is considered. Special attention is paid to estimators based on distance minimisation and it is found that these estimators often outperform classical techniques such as the maximum likelihood estimation and the method of moments, especially for small sample sizes. The use of all of the techniques developed in the thesis, including goodness-of-fit tests and estimation techniques, are illustrated using observed data.

We now provide some directions for future research. An important shortcoming in the literature can be filled when developing goodness-of-fit tests specifically for the Lomax distribution. Such a test will have application in parametric frailty models in survival analysis where the assumption is that the frailty parameter has a gamma distribution. If this is the case, then the distribution of the so-called residuals should follow a Lomax distribution under the null hypothesis. Furthermore, there is little available in the literature regarding the Pareto Types II, III and IV distributions. We plan to address this by developing goodness-of-fit tests for these distributions and to investigate and compare estimation techniques for these distributions. The larger number of parameters which require estimation when using these distributions results in an increase in the complexity associated with the development of the mentioned techniques.