REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

Comparison of estimates using L and TL moments and other robust characteristics of distributional shape and tail heaviness *

Authors: Ivana Malá
– Department of probability and statistics, University of Economics in Prague, Czech Republic (malai@vse.cz)

Václav Sládek
– Department of probability and statistics, University of Economics in Prague, Czech Republic (vaclav.sladek@vse.cz)

Filip Habarta
– Department of probability and statistics, University of Economics in Prague, Czech Republic (filip.habarta@vse.cz)

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

Abstract:

- Correct identification of a probability distribution is crucial in many areas of parametric statistics, inappropriate choice of the model can result in misleading or even incorrect decisions. In the text, we study the performance of robust characteristics of skewness and kurtosis of probability distributions that are less sensitive to outliers than the characteristics based on classical product moments. We use Monte Carlo simulation to illustrate properties of various robust (mainly quantile type) characteristics of skewness and kurtosis and compare them to the L-skewness (TL-skewness) and L-kurtosis (TL-kurtosis). The bias, standard and mean squared error of estimators are compared using simulations for standard normal, Laplace, Student, gamma and beta distributions and sample sizes ranged from 10 to 500 observations. The selected distributions gain symmetric and asymmetric unimodal distributions with different tail heaviness.

Key-Words:

- *robust characteristics, L-moments, TL-moments, skewness, kurtosis.*

AMS Subject Classification:

- 49A05, 78B26, 65C05.

*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

1. INTRODUCTION

Correct identification of a probability distribution is essential in many areas of parametric statistics, from the modelling of probability distributions to the regression modelling (assuming a dependent variable distribution), multivariate statistics, extreme-value analysis or time series analysis. The assumption about distribution form is crucial for parametric statistics, the correct or at least suitable choice of distribution allows a wide range of parametric procedures to be applied; in case of inappropriate choice, the results might be misleading or even incorrect. To test such the assumption, a large spectrum of statistical goodness-of-fit tests is available. For the general information on the sample, empirical distribution (histogram of data or nonparametric kernel density estimate) can be plotted. Sample characteristics of the location, variability, shape and concentration also can be evaluated. The typical sample characteristics are (raw, centred or standardised) product moments: mean, sample variance, coefficient of skewness and coefficient of kurtosis. Theoretical and sample moments are used not only to describe the distribution but also in the choice of suitable distribution to model the data or in inferential statistics. For example, if the normal distribution of data is assumed, the absolute value of the sample coefficient of skewness is supposed to be small, and the coefficient of kurtosis close to three. A frequently used test of normality Jarque-Bera compares theoretical and sample coefficients of skewness and kurtosis. The moment matching method can be applied to estimate parameters, the equations of theoretical and sample moments are solved explicitly or numerically with respect to the unknown parameters.

In the definition of the coefficient of skewness, a finite third raw moment is needed and for a finite value of the coefficient of kurtosis, the finite fourth raw moment is required. Moreover, the sample coefficients of skewness and kurtosis are strongly dependent on the sample size and the presence of outliers in the data. With higher sample product moments, the impact of outliers becomes more substantial. If the sample is drawn from long- or heavy-tailed distributions, we expect multiple outliers in the data and the use of more robust methods is essential.

We analyse data of this type in many fields of applications. For this reason, more robust characteristics and its estimates can be preferred to describe the distribution. There are various robust characteristics whose estimates are based on sample quantiles, which are more robust than sample product moments. Robust quantile characteristics of the distribution shape were presented, e.g., by [5], [10], [17] and [9], those of tail heaviness being dealt with by [7], [19] and many others. Hosking in [11] defined L-moments as a linear combination of order statistics, a robust alternative to product moments (robust moment characteristics). According to [11], [13], [14], [2] and other authors, estimates of L-moments are more reliable than those of product moments. TL-moments, defined as a trimmed linear combination of order statistics, are even more robust than L-moments. They were applied by [8] to describe the probability distribution.

The present article is focused on the estimation of the distribution shape and tail heaviness using robust moments and quantile characteristics. We treat L- and TL-moments in comparison with product moments and robust quantile characteristics. In this paper, we consider random samples from both symmetric (Student, normal, and Laplace) and asymmetric (gamma and beta) probability distributions. The latter ones being flexible, a different combination of their parameters allows us to obtain different shapes of the distribution, including asymmetric distributions. The aim of the article is to compare estimated characteristics (bias and both standard and mean squared errors) of the shape and tail heaviness characteristics of distributions depending on the distribution and size of the random sample, employing Monte Carlo simulations. The calculation was performed in the program R ([20]), using predefined and author-written functions (cf. [22] and [2]). Along with formulas and a short description of the considered robust moment and quantile characteristics, the following methodology section also contains the algorithm of the simulation study. Results and inferences drawn from Monte Carlo simulation are summarised in the next part of the paper, the concluding section assessing the outcomes of the simulation.

2. METHODS

2.1. L moments

Let X be a continuous random variable with a cumulative distribution function $F(x)$, quantile function $Q(x)$ and let $X_{1:n},\ X_{2:n},\ ...,\ X_{n:n}$ be an ordered sample of the size n drawn from the distribution of the random variable X. L-moments are defined in [11] as a linear combination of order statistics, the rth L-moment λ_r being as follows:

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \ldots, \tag{2.1}$$

where $EX_{r-k:r}$ is an expected value of the $(r-k)$th order statistics from a sample of size r. L-moments, a robust alternative to product moments, are used to describe random variables similarly as the classical product moments. The most used L-moments are those of order $r = 1, 2, 3$, and 4. The λ_1 is equal to the expected value of the variable X, describing its level, λ_2 indicating variability, λ_3 shape, and λ_4 tail heaviness of the distribution. Hosking and Wallis in [14] mentioned a dimensionless version of L-moments which is independent of the distribution scale and more useful than the unbounded version. Dimensionless L-moments τ are called L-moment ratios. They are defined as

$$\tau_r = \lambda_r/\lambda_2, \quad r = 3, 4, \ldots, \tag{2.2}$$

where τ_3 (L-skewness) and τ_4 (L-kurtosis) are most widely used in selection of probability distributions. Common properties of L-moments and L-moment ratios are as follows:

- they are defined for all distributions with finite expected values (finite values of higher moments are not required)
- there are not two distributions with the same values of all L-moments
- $-1 < \tau_r < 1$ for $\quad r = 3, 4, \ldots$
- $\frac{1}{4}(3\tau_3^2 - 1) \leq \tau_4 < 1$, and
- $\lambda_3 = \tau_3 = 0$ for symmetric distributions.

Estimates of λ_r and τ_r are based on an ordered random sample $X_{1:n}, X_{2:n}, ..., X_{n:n}$ drawn from the probability distribution of X. Estimates can be calculated using the following formulas

$$\hat{\lambda}_r = \frac{1}{r}\binom{n}{r}^{-1} \sum_{i=1}^{n}\sum_{j=0}^{r-1}(-1)^j \binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-i}{j} X_{i:n}, \tag{2.3}$$

$$\hat{\tau}_r = \hat{\lambda}_r / \hat{\lambda}_2, \quad r = 3, 4, \ldots. \tag{2.4}$$

Statistical characteristics of estimates are available, e.g. in [2] or [11]).

The R package Lmoments [18] provides functions for evaluation of both symmetric and asymmetric sample moments. R packages packages lmomco [3] and Lmoments [16] allow a wide range of calculations based on L-moments.

2.2. TL moments

Elamir and Seheult in [8] introduced TL-moments (trimmed L-moments) as a robust version of L-moments defined by the formula

$$\lambda_r^{(t)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} EX_{r+t-k:r+2t}, \quad r = 1, 2, \ldots, \tag{2.5}$$

where t represents the number of trimmed expected values from both sides of the sample. Zero weight is assigned to expected (mean) values, which are thus considered as outliers. Trimming can be either symmetric or asymmetric, the choice of the respective approach depends on the nature of the data. In the asymmetric form, the ordered sample is trimmed by t_1 values from left and t_2 values from right. The $EX_{r+t-k:r+2t}$ in 2.5 is then changed to $EX_{r+t_1-k:r+t_1+t_2}$. The TL-moment is then denoted by $\lambda_r^{(t_1,t_2)}$.

In this article, we focus only on symmetric trimming with $t = 1$; this choice is usually applied in the literature (also in [8]) as well as in practical applications. Trimming of one value is sufficient to overcome the problem of finite values of the moments for example for Cauchy distribution and enables the existence of all TL-moments to be finite ([8]), as only the expected values of minimum and maximum are not defined and their trimming allows the calculation of all TL-moments. The existence of TL-moments depends on the existence of expected values of ordered statistics and sometimes, more trimmed values should be used. For the Pareto distribution, from the formula (6) in [1], the number of necessary trimmed values depends on the shape parameter. The smaller the parameter, the higher the number of trimmed values. TL-moments can be used for the description of the distribution of a random variable. $\lambda_1^{(t)}$ is equal to the expected value of the random variable (if it exists), describing its level; $\lambda_2^{(t)}$ quantifying variability, $\lambda_3^{(t)}$ shape and $\lambda_4^{(t)}$ tail heaviness of the distribution.

The TL-moment ratios for the shape of distribution $\tau_3^{(t)}$(TL-skewness) and tail heaviness $\tau_4^{(t)}$ (TL-kurtosis)

$$\tau_r^{(t)} = \lambda_r^{(t)}/\lambda_2^{(t)}, \quad r = 3, 4, \ldots. \tag{2.6}$$

Both characteristics are location and scale invariant ([8]). Main properties of TL-moments and TL-moment ratios are as follows:

- we obtain the L-moments for $t = 0$ (or $t_1 = t_2 = 0$)
- for $r \geq 3$ applies (see [15] for the more general equation for the trimming (t_1, t_2) instead of symmetric (t, t) denoted by (t))

$$|\tau_r^{(t)}| \leq \frac{2(t+1)!(r+2t)!}{r(t+r-1)!(2+2t)!} \tag{2.7}$$

- $\left|\tau_3^{(1)}\right| \leq \frac{10}{9}$ and $|\tau_4^{(1)}| \leq \frac{5}{4}$ (substituting $t = 1$ to (2.7))
- $\tau_3^{(t)} = 0$ for symmetric distributions.

Sample counterparts of (symmetric) TL-moments $\lambda_r^{(t)}$ and $\tau_r^{(t)}$ are based on an ordered random sample of size n

$$\hat{\lambda}_r^{(t)} = \frac{1}{r} \sum_{i=t+1}^{n-t} \frac{\sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{i-1}{r+t-k-1} \binom{n-i}{t+k}}{\binom{n}{t+2t}} x_{i:n}, \tag{2.8}$$

$$\hat{\tau}_r^{(t)} = \hat{\lambda}_r^{(t)}/\hat{\lambda}_2^{(t)}, \quad r = 3, 4, \ldots. \tag{2.9}$$

The R packages TLmoments [18] and lmomco [3] provides a wide range of useful functions for application of both symmetric and asymmetric sample TL-moments. In our analysis, we applied the former one.

2.3. Quantile characteristics of the distributional shape

In classical parametric statistics, the product moment coefficient of skewness is used as a third standardised raw moment

$$\alpha_3 = E(X - EX)^3/(VarX)^{3/2}, \tag{2.10}$$

in the sample version ($X_i, \quad i = 1, 2, \dots, n$) based on the sample moments

$$a_3 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^3 / \left[\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right]^{2/3}, \tag{2.11}$$

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

In the paper, we use more robust characteristics of shape based on robust moments as L-skewness (2.2) and TL-skewness (2.6) and quantiles. The first characteristic of the distribution shape mentioned above is the medcouple (referred to as MC_F); see [6]. The sample version is defined as

$$MC_F = median_{i,j;\, x_i < x_j}\, h(x_i, x_j), \tag{2.12}$$

where h is a kernel function measuring the difference in the distances of x_i and x_j to the sample median $\tilde{x}$. This function is given by

$$h(x_i, x_j) = \left[(x_j - \tilde{x}) - (\tilde{x} - x_i)\right] / (x_j - x_i) \tag{2.13}$$

for $x_i \neq x_j$. If $x_i = \tilde{x}$, the value of $h(\tilde{x}, x_j) = 1$ for $x_j > Q(0.5)$; $x_j = \tilde{x}$ gives $h(x_i, \tilde{x}) = -1$ for $x_i < \tilde{x}$. If x_j is infinitely larger than $\tilde{x}$, h is closed to 1. On the other hand, if x_i is infinitely smaller than $\tilde{x}$, h approaches -1. Thus, the medcouple is not influenced by the presence of extreme values in a random sample; there are no larger/smaller values than ± 1. The medcouple is defined for all continuous distribution functions, existence of the expected value of any distribution moment is not needed. The functional form of the characteristics is given in [6].

Bowley in [5] introduced the characteristic of shape based solely on distribution quartiles. It is called the Bowley coefficient of skewness (BC) and is defined as

$$BC = \{[Q(0.75) - Q(0.5)] - [Q(0.5) - Q(0.25)]\}/[Q(0.75) - Q(0.25)]. \tag{2.14}$$

Some authors use the term “quartile skewness“ instead of the Bowley coefficient.

Hinkley [10] introduced the generalisation of Bowley measure

$$\nu_1(p) = \{[Q(1-p) - Q(0.5)] - [Q(0.5) - Q(p)]\}/[Q(1-p) - Q(p)] \tag{2.15}$$

for $p \in (0, 1)$. It is obvious that the Bowley coefficient is a special case of 2.15 for $p = 0.25$.

If we use (in 2.15) the first and sevenths octiles $Q(0.125)$ and $Q(0.875)$, we obtain the octile skewness (OC)

$$OC = \{[Q(0.875)-Q(0.5)]-[Q(0.5)-Q(0.125)]\}/[Q(0.875)-Q(0.125)]. \tag{2.16}$$

Groeneveld and Meeden in [9] proposed the coefficient of skewness (GMC) given by

$$GMC = [EX - Q(0.5)]/E|X - Q(0.5)|. \tag{2.17}$$

The last robust quantile characteristic considered is the Pearson coefficient (PC) introduced by Kendall and Stuart in [17]. Its formula is based on 2.17, where instead of the expected value of the absolute deviation between x_i and the median, they employ the standard deviation $\sqrt{VarX}$ of a distribution:

$$PC = [EX - Q(0.5)]/\sqrt{VarX}. \tag{2.18}$$

GMC is defined only for distributions with a finite value $E|X|$, whereas PC is defined for that with a finite variance.

All robust quantile and moment characteristics of the distribution shape 2.14 – 2.18 are defined on a range of values $[-1, 1]$ (except TL-moments, $|\tau_3^{(1)}| \leq 1.11$). For symmetric distributions, they are equal to zero. This allows us to compare properties of their estimates (bias and MSE) using an absolute value basis. The same applies for asymmetric distributions as they are functionally bounded (despite particular characteristics acquiring different values). The robust characteristics considered are applied only to the distributions for which they are defined (this is not the case, e.g., of Student distribution with 1 degree of freedom [the Cauchy distribution] and all characteristics based on classical moments or L-moments).

Sample counterparts of characteristics 2.14 – 2.18 are obtained by substituting the mean for the EX, sample standard error or more robust median absolute deviation for the standard deviation. There is not a generally accepted method for evaluation of sample quantiles. In the present paper, we use linear interpolation of the inverse of the empirical cumulative distribution function in the form

$$\hat{Q}(p) = x_{\lfloor h\rfloor:n} + (h - \lfloor h\rfloor)(x_{\lfloor h\rfloor+1:n} - x_{\lfloor h\rfloor:n}), \tag{2.19}$$

where $h = (n-1)p + 1$ and $\lfloor h\rfloor$ is the floor function.

2.4. Quantile characteristics of the tail heaviness

The moment coefficient of the kurtosis is defined as the fourth standardised raw moment

$$\alpha_4 = E(X - EX)^4/(VarX)^2, \tag{2.20}$$

in the sample version based on the sample moments

$$a_4 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^4 / \left[\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right]^2. \tag{2.21}$$

This sample formula is a biased estimator to α_4 (as well as for α_3 given in 2.11) even for a sample from the normal distribution ([12]).

The first robust characteristics of tail heaviness based on octiles is the Moors coefficient of kurtosis (MKC); see [19]. The coefficient is defined as

$$MKC = \{[Q(0.875) - Q(0.625)] + [Q(0.375) - Q(0.125)]\}/[Q(0.625) - Q(0.125)]. \tag{2.22}$$

MKC exists for any continuous distribution taking all positive values.

Crown and Siddiqui in [7] introduced their coefficient defined as

$$CKC = [Q(1-\alpha) - Q(\alpha)]/[Q(1-\beta) - Q(\beta)], \tag{2.23}$$

for $\alpha, \beta \in (0, 0.5)$. The authors recommend using $\alpha = 0.025$ and $\beta = 0.25$ for the normal distribution. In our study, we follow Crown and Siddiqui recommendation regarding the considered probability distributions because many of them are symmetric. The suitability of this arrangement for asymmetric distributions is also open to analysis.

The last characteristic of tail heaviness considered, defined by Schmid and Trede in [21] as a special case of 2.23, selects $\alpha = 0.125$ and $\beta = 0.25$

$$PKC = [Q(0.875) - Q(0.125)]/[Q(0.75) - Q(0.25)]. \tag{2.24}$$

All the above robust quantile characteristics of tail heaviness exist for any distribution, no finite moment values being necessary. The range of values of these characteristics is $[0, \infty]$. A comparison of the characteristics of tail heaviness estimates based on their absolute values (means and standard deviations) is not appropriate because of various possible ranges of values; modification of characteristics (expected value and variance) is thus applied (see 2.25).

2.5. The algorithm of simulation and methods of comparison

The simulation study employs the Monte Carlo methodology, which assumes knowledge of the theoretical distribution and its parameters, random samples being drawn from the distribution and values of all characteristics of interest being computed. Let us consider samples ranging from 10 to 500 observations, because the greatest differences in estimate characteristics are expected to concentrate in small-sized samples. The samples comprising 100 and more observations are used to analyse the convergence of estimation bias and variability.

In the simulation study, we have chosen three symmetric and two asymmetric distributions:

- Student distribution ($t(\nu)$, degrees of freedom $\nu = 1, 2, 3$),
- standard normal distribution ($N(0; 1)$),
- Laplace distribution ($La(\mu; b), \mu = 0$, the location parameter, $b = 10$ the scale of the distribution),
- gamma distribution ($Gamma(\theta; k), \theta = 2$ the shape parameter, $k = 2$ the scale parameter),
- beta distribution ($Beta(\theta_1; \theta_2), \theta_1 = 2$, $\theta_2 = 5$ the shape parameters).

In Table 1, all characteristics introduced in 2.1 – 2.4 for selected distributions are given.

The Student distribution is a symmetric continuous probability distribution, bell-shaped (similar to the normal distribution) and heavy-tailed (unlike the normal distribution). With an increasing number of degrees of freedom, the shape of the Student distribution converges to that of the normal distribution. We focus on how the above affects the properties of estimates. Gamma and beta distributions are also well-established distribution families. Their shape can be modified by setting different values of parameters. We choose combinations of parameter values to obtain a positively skewed shape.

The Laplace distribution is also called the double exponential distribution. It is symmetrically shaped like Student and normal distributions. The distinction between probability density functions of Laplace and normal distributions lies in that the latter is expressed as the squared difference, while the former as the absolute difference from their means, respectively. The Laplace distribution as a result has heavier tails than the normal distribution. Some of the chosen robust quantile and moment characteristics exist only if the distribution has one or more defined raw moments (Table 1). L-moments, for example, exist only for the distribution with a finite expected value, which does not apply to the Student distribution with one degree of freedom (Cauchy distribution). The existence of the moment characteristics of skewness (coefficient of skewness) assumes finite 3rd raw moment, for the coefficient of kurtosis, we need finite 4th raw moment. Estimate calculations are done only if the characteristics are defined for a particular distribution.

Some 50,000 times random samples were generated from the considered distributions for sample sizes $10 - 500$, point estimates and standard errors of analysed characteristics calculated as the mean and the sample standard deviation of 50,000 generated values. Bias in characteristics of shape is shown by the difference between their theoretical value and estimated expected value. We compare the variability of characteristics using estimates of their standard errors calculated as sample standard deviations and mean squared error MSE.

characteristic	$t(1)$	$t(3)$	$\mathbb{N}(0;1)$	$La(0;10)$	$gamma(2;2)$	$beta(2;5)$
EX	-	0	0	0	4	217
$\sqrt{VarX}$	-		1	14.142	2.828	0.126
α_3	-	0	0	0	1.414	0.596
α_4	-	-	3	3	0.142	2.88
τ_3	-	0	0	0	0.235	0.123
$\tau_3^{(1)}$	0	0	0	0	0.150	0.080
MC_F	0	0	0	0	0.225	0.128
BC	0	0	0	0	0.172	0.095
OC	0	0	0	0	0.287	0.160
GMC	0	0	0	0	0.306	0.165
PC_{MAD}	0	0	0	0	0.227	0.133
PC_{SD}	-	0	0	0	0.227	0.133
τ_4	-	0.035	0.123	0.035	0.071	0.090
τ_4^1	0.077	0.041	0.063	0.041	1.731	0.048
Pearson	8.630	0.547	1.706	0.547	1.262	1.659
MKC	8.663	0.590	1.233	0.590	3.078	1.181
CKC	24,628,907	5.325	2.906	5.325	3.078	2.619

Table 1: Basic characteristics of probability distributions selected in the simulation. We use (-) for non-existence.

DF	char.	10	20	30	50	70	90	100	150	200	300
1	$\tau_3^{(1)}$	**0.328**	**0.264**	**0.235**	0.201	0.178	0.162	0.155	0.132	0.117	0.097
1	MC_F	0.335	0.276	0.236	**0.190**	**0.163**	**0.145**	**0.1382**	**0.114**	**0.099**	**0.081**
1	BC	0.393	0.307	0.261	0.210	0.180	0.159	0.152	0.126	0.109	0.090
1	OC	*0.439*	*0.357*	*0.306*	*0.253*	*0.218*	*0.196*	*0.186*	*0.155*	*0.136*	*0.111*
2	τ_3	0.304	0.255	0.231	0.202	0.18	*0.177*	*0.165*	*0.146*	*0.133*	*0.115*
2	$\tau_3^{(1)}$	**0.236**	**0.150**	**0.124**	**0.093**	**0.078**	**0.069**	**0.066**	**0.054**	**0.046**	**0.038**
2	MC_F	0.285	0.235	0.201	0.162	0.139	0.123	0.118	0.096	0.084	0.069
2	BC	*0.358*	*0.278*	0.237	0.189	0.161	0.143	0.135	0.111	0.097	0.079
2	OC	0.350	0.263	0.226	0.179	0.151	0.136	0.128	0.105	0.092	0.075
2	GMC	0.346	0.282	*0.249*	*0.209*	*0.186*	0.170	0.164	0.142	0.127	0.106
3	τ_3	0.245	0.187	0.161	0.132	0.115	0.104	0.099	0.083	0.073	0.061
3	$\tau_3^{(1)}$	0.215	**0.128**	**0.099**	**0.075**	**0.062**	**0.055**	**0.052**	**0.042**	**0.036**	**0.029**
3	MC_F	0.274	0.226	0.194	0.156	0.134	0.119	0.113	0.093	0.081	0.065
3	BC	0.353	*0.274*	*0.233*	*0.186*	*0.159*	*0.141*	*0.134*	*0.112*	*0.095*	*0.078*
3	OC	0.330	0.243	0.203	0.163	0.138	0.123	0.116	0.095	0.083	0.068
3	GMC	0.301	0.232	0.197	0.158	0.136	0.121	0.115	0.097	0.083	0.068
3	PC_{MAD}	*0.443*	0.280	0.224	0.174	0.148	0.129	0.122	0.099	0.085	0.069
3	PC_{SD}	**0.203**	0.156	0.131	0.103 5	0.088	0.078	0.075	0.061	0.053	0.043

Table 2: Student distribution ($\nu = 1,2,3$; $n = 10,20,30,50,70,90,100,150,200,300$), characteristics of shape (2.4, 2.9, 2.12, 2.14–2.18). Estimated standard errors of estimates.

Because of the wide range of tail heaviness values, which acquire different ones for given distributions, a comparison using bias and standard error does not induce relevant inferences. Therefore, we use modified bias and modified MSE characteristics, which are the same as classical ones divided by the (squared) theoretical value of a robust characteristic. All the characteristics are put on the same level, the coefficient of variation being based on such an approach. Finally, we obtain the ratio of the value of bias and MSE to the theoretical value of the robust characteristic. The modified bias and modified MSE are calculated as (θ

DF	Sample	10	20	30	50	70	90	100	150	200	300
1	$\tau_4^{(1)}$	-0.250	-0.193	-0.156	-0.115	-0.074	-0.054	*-0.048*	*-0.042*	-0.038	*-0.030*
1	PKC	**0.095**	**0.110**	**0.065**	**0.045**	**0.019**	**0.016**	**0.011**	**0.010**	**0.009**	**0.008**
1	MKC	0.106	0.129	0.075	0.051	0.020	0.017	0.012	0.011	0.010	0.008
1	CKC	*3.028*	*3.746*	*3.419*	*2.768*	*0.111*	*0.119*	0.026	0.042	*0.054*	0.021
2	τ_4	*-0.224*	*-0.183*	*-0.157*	*-0.125*	*-0.093*	*-0.075*	*-0.069*	*-0.063*	*-0.059*	*-0.050*
2	$\tau_4^{(1)}$	**-0.010**	**-0.022**	-0.022	-0.017	-0.012	-0.009	-0.008	-0.007	-0.006	-0.005
2	PKC	0.024	0.026	0.018	0.013	0.004	0.004	0.002	0.002	0.002	0.002
2	MKC	0.025	0.030	0.020	0.013	0.002	0.003	0.001	0.001	0.002	0.002
2	CKC	-0.118	-0.037	0.014	0.021	-0.047	-0.026	-0.041	-0.029	-0.021	-0.025
3	τ_4	-0.127	*-0.096*	*-0.077*	*-0.058*	-0.040	-0.031	-0.028	-0.025	-0.023	-0.018
3	$\tau_4^{(1)}$	0.075	0.031	0.018	**0.009**	**0.003**	**0.000**	**0.000**	-0.001	**-0.001**	**0.000**
3	PKC	**0.018**	**0.018**	**0.014**	0.010	0.004	0.004	0.002	0.002	0.002	0.002
3	MKC	0.019	0.021	0.016	0.010	0.003	0.003	0.001	**0.001**	0.001	0.001
3	CKC	*-0.138*	-0.087	-0.051	-0.036	*-0.049*	*-0.034*	*-0.041*	*-0.032*	*-0.026*	*-0.025*

Table 3: Student distribution ($\nu = 1, 2, 3$; $n = 10, 20, 30, 50, 70, 90, 100, 150, 200, 300$), modified bias of estimates (defined in 2.25) of characteristics of kurtosis (2.4, 2.9, 2.22–2.24).

is a parameter, $\hat{\theta}$ is its estimate)

$$E(\hat{\theta} - \theta)/\theta, \quad Var(\hat{\theta})/\theta^2. \tag{2.25}$$

This allows for a statistical comparison between the properties of different estimates.

3. Results

3.1. Characteristics of skewness

The bias in estimated characteristics is low for symmetric distributions (Student, normal and Laplace). Figure 1 shows its development depending on sample sizes. For the Student distribution with one degree of freedom (Cauchy distribution), the characteristics τ_3, GMC, and PC are undefined (EX does not exist) and they are not included in the figure (see also Figure 2). Bias curves are similar for all estimates, differing only in their levels, converging to zero (represented by the dashed line) with an increasing number of observations. Estimation bias is volatile in small samples (up to 100 observations), which show no constant development. The figure indicates the highest bias in BC for each symmetric distribution, τ_3 and $\tau_3^{(1)}$ being among the estimates with the lowest bias value. In absolute terms, however, it is obvious that estimation bias is generally small. Values for Student distribution are shown in Table 2. The lowest values (the best performance for the particular distribution and degrees of freedom) are highlighted in bold letters, the highest in italics (the worst performance in the block in the table). The TL-skewness is superior in bias from two degrees of freedom and performs well. For the Cauchy distribution, the L-skewness is undefined (this characteristic is equivalent to TL-skewness for $t = 0$ (no trimmed values)), and the TL-skewness is a first possible defined value concerning the number of trimmed values.

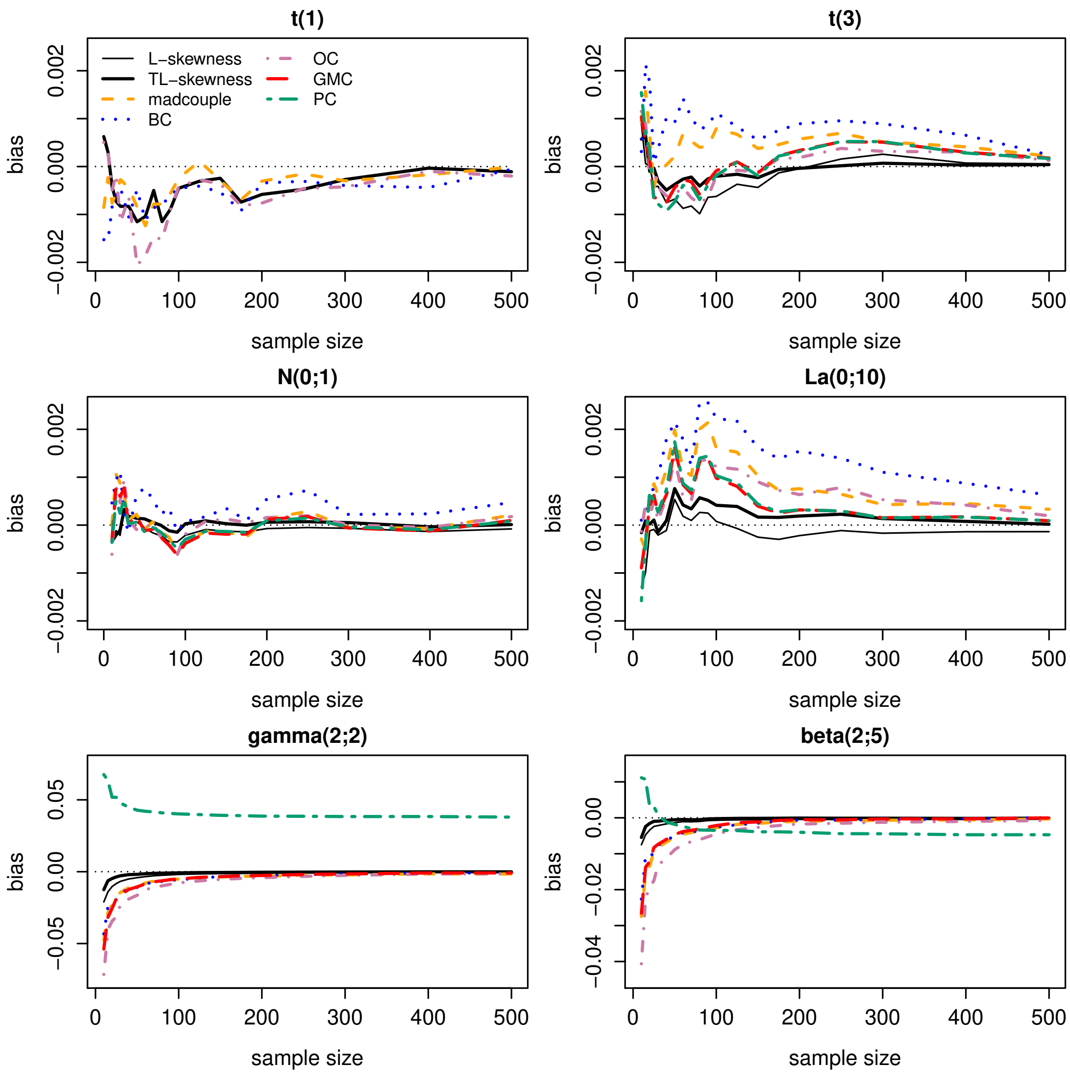

Figure 1: Estimated bias in distribution shape characteristics (2.4, 2.9, 2.12, 2.14, 2.16–2.18) for $n = 10$–500.

For the asymmetric distributions (last row in Figure 1) the τ_3 and $\tau_3^{(1)}$ exhibit the lowest absolute and relative values of bias for small samples (relative bias is equal to the estimation bias value divided by the actual value of the characteristic). An extremely high bias even for the samples with several hundred observations occurs in PC_{MAD}, where MAD is used as a standard deviation estimate in [19], its convergence being very slow. Both for small and large samples, bias is low in absolute terms and estimate convergences are relatively fast (except PC estimates).

Distribution shape estimates vary mostly in variability. Using the standard error of estimation, Figure 2 illustrates standard errors of estimates. For the Student distribution with one degree of freedom the characteristics τ_3, GMC,

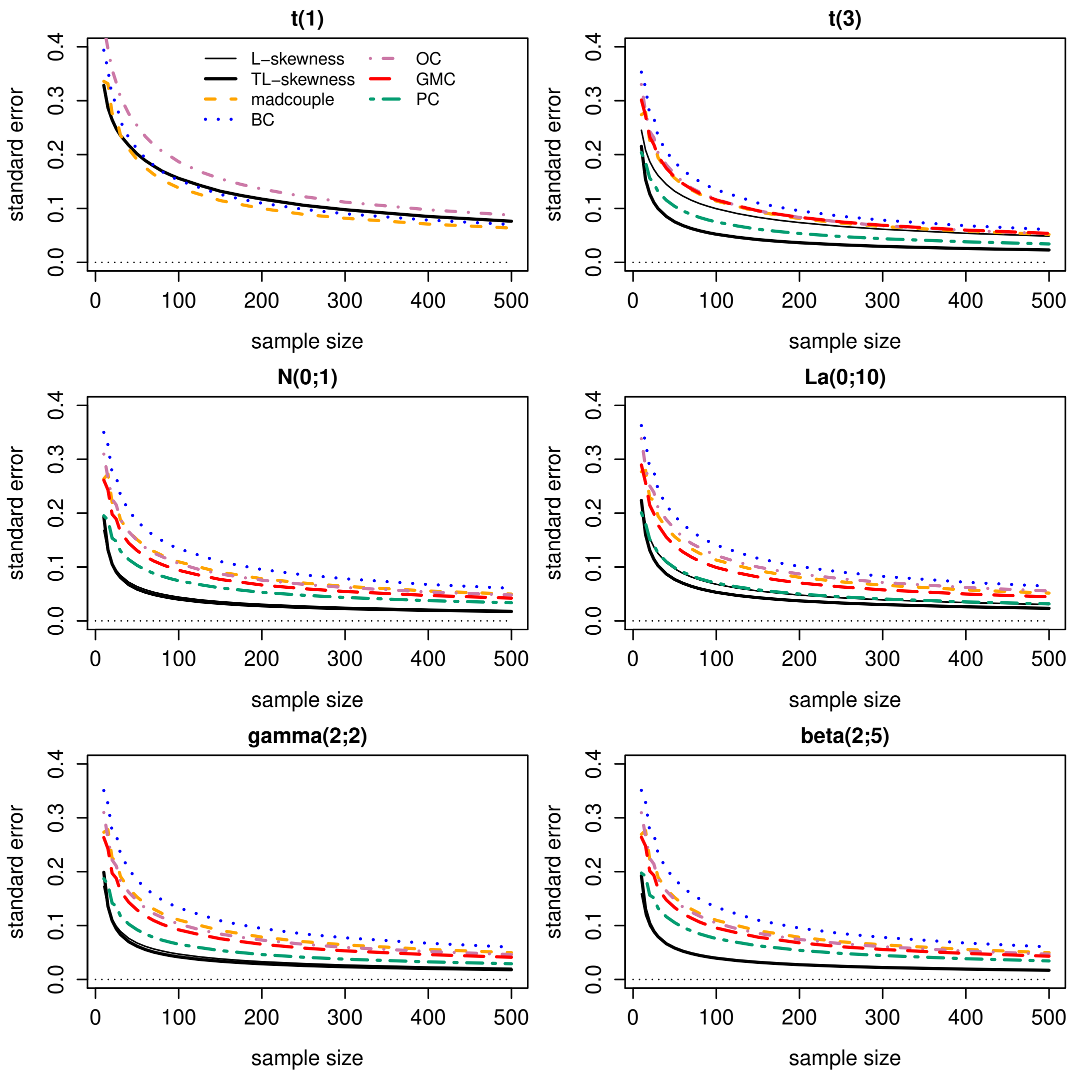

Figure 2: Standard errors of estimates of distribution shape characteristics (2.4, 2.9, 2.12, 2.14, 2.16–2.18) for $n = 10$–500.

and PC are undefined (see also Figure 1), for this reason, no lines are included. For the normal and Laplace distributions with relatively low kurtosis and absence of outliers, the standard errors are very close for both characteristics based on L-moments. Let us first summarise the results for the Student distribution (also see Table 3). The $\tau_3^{(1)}$ and MC_F show lower variability (especially in the case of small samples) than other estimates with one degree of freedom. The shape of the Student distribution can be estimated using L-moments and Pearson and GM coefficients only if the number of degrees of freedom is higher than one. The τ_3 has lower variability than the quantile-based estimates but higher than $\tau_3^{(1)}$ and MC_F. For small samples, the study outcomes confirm that the variability of τ_3 and $\tau_3^{(1)}$ decreases more sharply with an increasing number of degrees of freedom

DF	char.	10	20	30	50	70	90	100	150	200	300
1	$\tau_4^{(1)}$	1.025	0.579	0.465	0.365	0.314	0.278	0.265	0.221	0.194	0.160
1	PKC	**1.108**	**0.481**	**0.350**	**0.255**	**0.207**	**0.182**	**0.169**	**0.136**	**0.118**	**0.096**
1	MKC	1.327	0.581	0.422	0.308	0.250	0.220	0.205	0.165	0.142	0.116
1	CKC	*96.964*	*62.156*	*47.601*	*1.431*	*0.844*	*0.594*	*0.543*	*0.405*	*0.346*	*0.266*
2	τ_4	0.724	0.500	0.422	0.348	*0.308*	*0.281*	*0.271*	*0.235*	*0.213*	*0.184*
2	$\tau_4^{(1)}$	*1.384*	0.685	0.512	*0.372*	0.307	0.266	0.251	0.202	0.174	0.141
2	PKC	**0.462**	**0.275**	**0.224**	**0.175**	**0.146**	**0.130**	**0.122**	**0.100**	**0.086**	**0.071**
2	MKC	0.611	0.368	0.298	0.232	0.194	0.172	0.162	0.132	0.114	0.093
2	CKC	1.251	*0.931*	*0.685*	0.363	0.295	0.261	0.244	0.199	0.175	0.140
3	τ_4	0.805	0.518	0.422	0.332	0.286	0.256	0.245	0.204	0.179	0.149
3	$\tau_4^{(1)}$	*1.601*	*0.775*	*0.568*	*0.410*	*0.336*	*0.292*	*0.276*	*0.221*	*0.189*	*0.154*
3	PKC	**0.390**	**0.247**	**0.203**	**0.160**	**0.134**	**0.119**	**0.112**	**0.092**	**0.079**	**0.065**
3	MKC	0.537	0.343	0.279	0.220	0.184	0.163	0.154	0.125	0.108	0.089
3	CKC	0.623	0.447	0.351	0.269	0.222	0.200	0.189	0.156	0.137	0.111

Table 4: Student distribution ($\nu = 1, 2, 3$; $n = 10, 20, 30, 50, 70, 90, 100, 150, 200, 300$). Coefficient of variation of characteristics of kurtosis (2.4, 2.9, 2.12, 2.14– 2.18).

than the variability of other estimates. The PC_SD and $\tau_3^{(1)}$ are the estimates with the lowest variability for the Student distribution with three degrees of freedom, other estimates showing much higher variability.

The order of estimates showing the lowest variability for the sample containing ten observations drawn from the standard normal distribution is τ_3, $\tau_3^{(1)}$ and PC_{SD}. Other estimates show much higher variability. The difference in variability between τ_3 and $\tau_3^{(1)}$ decreases with an increasing number of observations. The variability of PC_SD declines slowly compared to τ_3 and $\tau_3^{(1)}$. Convergence in the variability of other estimates is not fast enough to reach the value of τ_3 and $\tau_3^{(1)}$. The last considered symmetric distribution is the Laplace distribution. As is the case with the normal distribution, τ_3 and $\tau_3^{(1)}$ and PC_SD are estimates with the lowest variability. Also, the speed of their variability convergence seems similar. The variability of other estimates is significantly higher, their convergence not being fast enough to achieve τ_3 and $\tau_3^{(1)}$ and PC_SD variability. The order of estimates arranged according to their variability is the same for gamma and beta distributions. The conclusions drawn are analogous to those concerning the normal distribution. The τ_3 and $\tau_3^{(1)}$ and PC_SD show the lowest variability, that of PC_SD decreases more slowly compared to τ_3 and $\tau_3^{(1)}$. Convergence in the variability of other estimates is not as fast τ_3 and $\tau_3^{(1)}$.

3.2. Characteristics of kurtosis

Given the inconsistent values of tail heaviness characteristics of considered distributions is shown in Figure 3 and for Student distribution in Table 3. For the Cauchy distribution $t(1)$, again the characteristics based on L-moments is undefined. The most biased estimate for the Student distribution with one degree of freedom is CKC, converging faster than the other considered estimates. It belongs to high-biased symmetric distribution estimates comparable with those

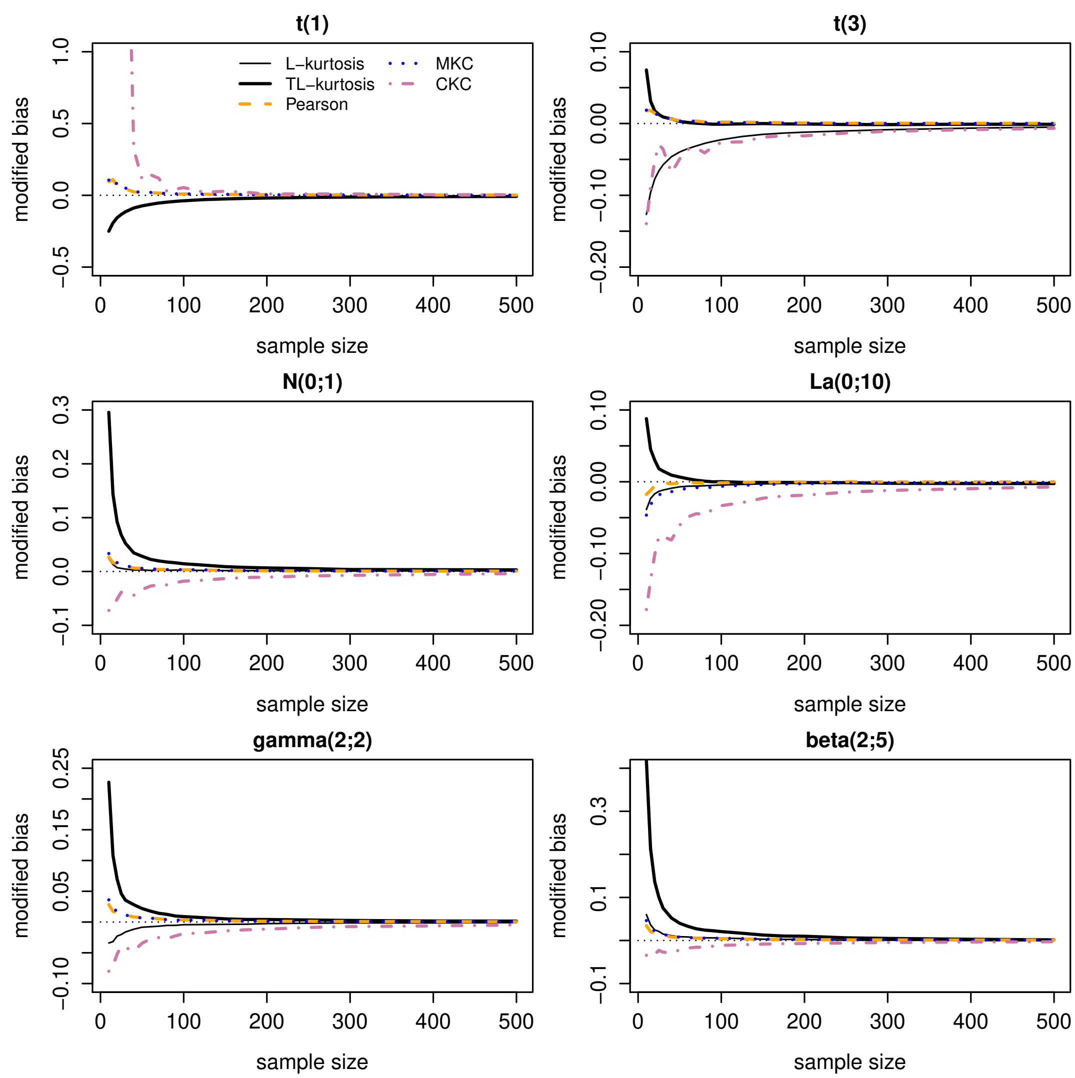

Figure 3: Modified estimation bias 2.25 in distribution tail characteristics (2.4, 2.9, 2.22–2.24) for $n = 10$–500.

for samples with 100 and more observations. The $\tau_4^{(1)}$ has a high value of modified bias for small samples drawn from normal and Laplace distributions. The τ_4, MKC, and PKC, on the other hand, are the least-biased estimates for all symmetric distributions (both small and large samples) considered. Their bias modification values are similar. Table 3 contains the values of the modified bias of estimates for the Student distribution. MKC and PKC are estimates with the lowest modified bias values for all degrees of freedom considered, including small samples. The modified bias of $\tau_4^{(1)}$ is close to MKC and PKC for two and more degrees of freedom. The τ_4 and CKC exhibit the highest modified bias for this distribution. The $\tau_4^{(1)}$ is the most biased estimate for both distributions considered. The $\tau_4^{(1)}$ overestimates its theoretical value (for a sample with

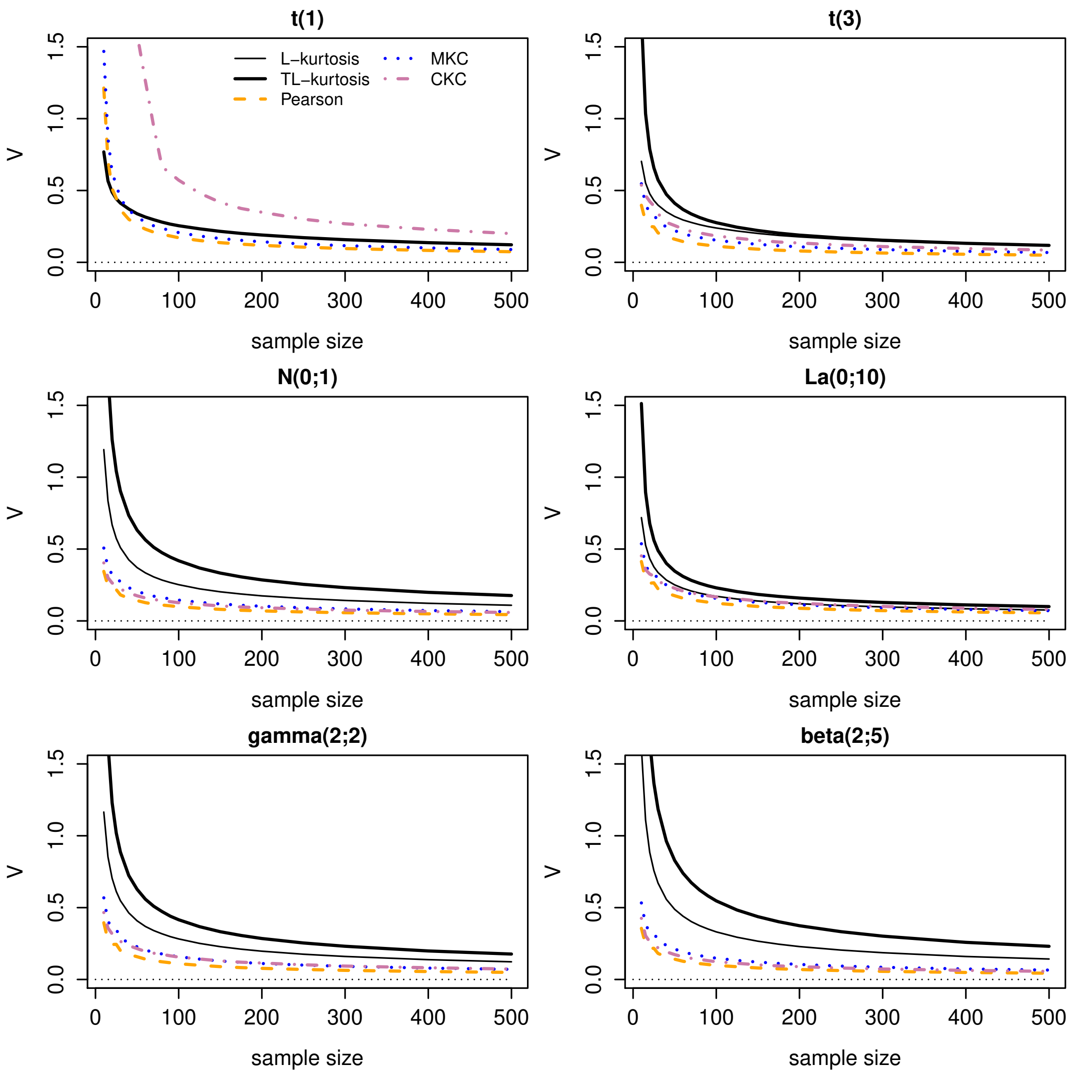

Figure 4: Variation coefficient of estimates of tail heaviness characteristics (2.4, 2.9, 2.22–2.24) for $n = 10$–500.

10 observations) in the cases of gamma and beta distributions by about 22 and 41%, respectively. Its decline is relatively sharp with an increasing number of observations, and bias is similar to that of other estimates for samples with 60 and more observations. The MKC and PKC show the lowest value of modified bias analogous to that for a symmetric distribution. τ_4 is close to MKC and PKC, and CKC estimate is less biased for asymmetric distributions than for symmetric ones.

The variability of estimates is quantified using the variation coefficient. Its development for symmetric distributions is shown in Figure 4. The $\tau_4^{(1)}$ has the lowest coefficient of variation for small samples (up to 25 observations) drawn from the Student distribution with one degree of freedom, its convergence being slower than that of MKC and PKC. If a sample consists of more than 25 observations, MKC and PKC are less variable than $\tau_4^{(1)}$. Table 4 shows values of the variation coefficient of CKC, $\tau_4^{(1)}$ and τ_4 for the Student distribution dependent on degrees of freedom. Interestingly, the variability of CKC declines markedly when degrees of freedom change from one to two (from 390.565 to 1.103 for a 10-observation sample), the variability of $\tau_4^{(1)}$ and τ_4 growing with an increase in degrees of freedom. The latter two estimates are more variable than other ones in the case of small samples generated from normal and Laplace distributions. For large samples, the variability of τ_4 is comparable with other estimates. PKC has the lowest variability for each symmetric distribution considered. For the asymmetric distributions both τ_4 and $\tau_4^{(1)}$ show fast convergence of variability. However, even for the sample with 500 observations, their variability is several times greater than that of MKC, PKC and CKC, the variability of $\tau_4^{(1)}$ being the highest. Therefore, in terms of variability, neither τ_4 nor $\tau_4^{(1)}$ are appropriate estimates of the tail heaviness of asymmetric distributions.

Because of estimation bias, comparison of estimates is made on the basis of the modified mean square error 2.25, the method taking into account both bias and variability of estimates. The development of modified MSE is similar to that of estimates of a coefficient of variation (not given in the text). MSE-based methodology provides results similar to those yielded by variation analysis.

4. Conclusion

Simulation results show that the bias of distribution shape estimates is low for both symmetric and asymmetric probability distributions. The main difference between estimates is in their variability (quantified by standard error). τ_3 and $\tau_3^{(1)}$ are estimates with small variability, the best robust quantile ones in terms of variability being MC_F and PC_SD. The variability of τ_3 and $\tau_3^{(1)}$ decreases more sharply than that of other estimates in the case of small samples with an increasing number of degrees of freedom of the Student distribution. The τ_3 and $\tau_3^{(1)}$ and PC_SD are the most appropriate estimates for symmetric (Student, normal and Laplace) and asymmetric (gamma and beta) distributions dealt with in this paper.

Some conclusions concerning tail heaviness estimates follow: τ_4 and $\tau_4^{(1)}$ has a high value of modified bias for small samples drawn from normal and Laplace distributions, $\tau_4^{(1)}$ is the most biased estimate for asymmetric distributions, and MKC and PKC are those with the lowest value of modified bias as far as Student distributions are concerned. We conclude that τ_4, MKC and PKC show the lowest bias for all the considered symmetric and asymmetric distributions (small and large samples alike), values of their modified bias being mutually comparable. The variability of $\tau_4^{(1)}$ and τ_4 increases with increasing degrees of freedom of the

Student distribution. As for small samples generated from Normal and Laplace distributions, $\tau_4^{(1)}$ and τ_4 are more variable than other estimates. The τ_4 variability is comparable to other estimates for large samples. While PKC indicates the lowest variability for each given symmetric distribution, the variability of τ_4 and $\tau_4^{(1)}$ is much higher than that of other tail heaviness estimates for asymmetric distributions.

Estimates of distributional shape based on L- and TL-moments possess the best characteristics (bias and variability), outperforming those yielded by a robust quantile approach in the situations considered. Our study, however, also confirms that robust quantile-based estimators produce more reliable tail heaviness estimation outcomes than those based on L- and TL-moments.

ACKNOWLEDGMENTS

Institutional support from the funds for the long-term conceptual advancement of science and research, number IP 400 040, at the Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic is gratefully acknowledged. We also acknowledge valuable suggestions from the referees.

REFERENCES

[1] ARNOLD, B. C. (2015). Pareto Distribution. In: Wiley Online Library. https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat01100.pub2.

[2] ASQUITH, W. H. (2011). *Distributional analysis with L-moment Statistics using the R environment for statistical computing,* Lubbock, Texas, Create Space Independent Publishing Platform.

[3] ASQUITH, W. H. (2020). lmomco: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.3.6. http://www.cran.r-project.org/package=lmomco.

[4] BONATO, M. (2011). Robust estimation of skewness and kurtosis in distributions with infinite higher moments, *Financial Research Letters*, **8**, 77–87.

[5] BOWLEY, A. L. (1920). *Elements of statistics*, New York, P. S. King & son, Limited.

[6] BRYS, G; HUBERT, M. and STRUYF, A. (2004). A Robust measure of skewness, *Journal of Computational and Graphical Statistics*, **13**, 996–1017.

[7] CROWN, E. and SIDDIQUI, M. (1967). Robust estimation of location, *The Statistician*, **62**, 353–389.

[8] ELSAYED, A. H. E. and SEHEULT, A. H. (2003). Trimmed L-moments, *Computational Statistics & Data Analysis*, **43**, 299–314.

[9] GROENEVELD, R. and MEEDEN, G. (1984). Measuring skewness and kurtosis, *The Statistician*, **33**, 391–399.

[10] HINKLEY, D. V. (1975). On power transformations to symmetry, *Biometrika*, **62**, 101–111.

[11] HOSKING, J. R. M. (1990). L-moments: Analysis and estimation of distribution using linear combinations of order statistics, *J. R. Statist. Soc. B*, **52**, 105–124.

[12] JOANES, D. N. and GILL, C. A. (1998). Comparing Measures of Sample Skewness and Kurtosis, *The American Statistician*, **47**, 183–189.

[13] HOSKING, J. R. M. (1992). Moments of L-moments? An example comparing two measures of distributional shape, *The American Statistician*, **46**, 186–189.

[14] HOSKING, J. R. M. and WALLIS, J. R. (1997). *Regional frequency analysis*, Cambridge, New York, Cambridge University Press.

[15] HOSKING, J. R. M. (2007). Some theory and practical uses of trimmed L-moments, *Journal of Statistical Planning and Inference*, **137**, 3024–3039.

[16] KARVANEN, J. (2019). Lmoments: L-moments and quantile mixtures. R package version 1.3-1 `http://www.cran.r-project.org/package=Lmoments`.

[17] KENDALL, M. and STUART, A. (1977). *The Advance theory of statistics*, London, Griffin.

[18] LILIENTHAL, J. (2019). tlmoments: Calculate TL-Moments and Convert Them to Distribution Parameters. R package version 0.7.5. `https://cran.r-project.org/web/packages/TLMoments`.

[19] MOORS, J. (1988). A quantile alternative for kurtosis, *The Statistician*, **37**, 25–32.

[20] R CORE TEAM (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

[21] SCHMID, F. and TREDE, M. (2003). Simple tests for peakedness, fat tails and leptokurtosis based on quantiles, *Computational Statistics & Data Analysis*, **43**, 1–12.

[22] YEE, T. W. (2016). VGAM: Vector Generalized Linear and Additive Models. R package version 1.0-2. `http://CRAN.R-project.org/package=VGAM`.