PROPERTIES OF MOMENT METHOD ESTIMATES BASED ON L MOMENTS FROM RIGHT CENSORED DATA

Ivana Malá

Abstract

The most frequently used method for estimation of parameters is the maximum likelihood method. In the contribution the moment method based on L moments (instead of classical moments) is used and such robust estimates are compared to maximum likelihood estimates. The Monte Carlo study is presented in order to show the robustness of moment estimates especially for the skewed contaminated data. A sample from lognormal distribution is contamined by 0 %, 5 %, and 10 % of observations from other lognormal distribution. The situation is described by the finite mixture model with two components. The component membership is supposed to be known and one lognormal distribution is fitted into data as well as the mixture of two lognormal distributions. Results are compared with normal distributions fitted into logarithms of data. The dependence on the percentage of censored data is also of interest (censoring from 0 % (complete data) to 30 % of observations in the sample of 500 observations). All computations are done in R program.

Key words: censored data, L-moments, Monte Carlo simulation

JEL Code: C24, C63, C13

Introduction

The problem of estimation of parameters of probability distributions is frequently met in the statistical applications. The maximum likelihood method (ML) is accepted as general and asymptotically optimal method for the estimation of parameters. Method of moments (MM) is based on solving of equations of theoretical and sample moments with respect to the unknown parameters. The sample moments are sensitive to the outliers and the moment method (easy to be used) do not give generally satisfactory results. L-moments (and other robust analogies of classical moments) provide the characteristics of distribution more robust and they can replace classical moments in the moment method of estimation (Hosking, 1990; Bílková, 2013). Furthermore for heavy tailed distributions with only finite mean, this is a viable alternative to maximum likelihood (Delicadoa and Goriab, 2008, Mudholkar and Hutson,

1998) for a class of estimators based on QL-moments as well as L-moments that always exist. The L-moments, being linear functions of order statistic, are subject to less sampling variability, robust to outliers and the asymptotic results are reliable even for small samples. Hosking and Wallis (1987) applied MM with L-moments to extreme value distribution. They found that it performs better than method of moments and that both methods do well in small samples compared to maximum likelihood estimation. In the survival analysis highly skewed data with heavy tails (moreover with outliers) are usually used and the positive impact of robust moments is expected to be even stronger. In this text the sample L-moments from data including right censored observations are evaluated, the definition of sample L-moments was given by Wang et al., 2010. In the text the authors used L-moments for the estimation of parameters of extreme value and Weibull distributions with superior results to maximum likelihood estimates even in the presence of heavy censoring (up to 50 percent).

In the text the lognormal distribution is used. This skewed distribution is frequently used in the modelling of incomes or wages (Bílková, 2013). The moment method with L-moments was used for the modelling of the duration of unemployment in the Czech Republic in Malá, 2015 for interval censored data transformed to the right censored.

The aim of the text is to compare maximum likelihood estimates (MLE) with the estimates obtained by the moment method based on L-moments (MME) for contaminated lognormal distribution (possibly with right censored data). Two straightforward approaches are used, the lognormal distribution is fitted into data and normal distribution id fitted to the logarithm of data. There exist closed formulas for the theoretical values of L-moments of lognormal and normal distributions (Hosking, 1990). In the text the lognormal distribution is contaminated by 5 and 10 percent of different lognormal data and the censoring is up to 30 percent of observations. The presented results come from the large simulation study and it is not possible to illustrate the situation in general.

1 Description of the models

The nonparametric L-moments were proposed by Hosking in 1990 (Hosking, 1990). For a random sample of the size k from the distribution of a random variable X we denote the ordered sample $\left(X_{1:k}, X_{2:k}, ..., X_{k:k}\right),'$ where the first index refers to the rank and the second to the sample size. We define (for $k = 1, 2, ...$) the k-th L-moment λ_k as (Hosking, 1990)

$$\lambda_k = \frac{1}{k}\sum_{j=0}^{k-1}(-1)^j \binom{k-1}{j} E\left(X_{k-j:k}\right), \tag{1}$$

and the sample L-moment l_k (from a sample of the size n) by the sample version of (1)

$$l_k = \binom{n}{k}^{-1} \sum_{1\le i_1 < i_2 < \ldots < i_k \le n} \frac{1}{k}\sum_{j=0}^{k-1}(-1)^j \binom{k-1}{j} x_{i_{k-j}:n}. \tag{2}$$

In (Hosking, 1990) the moment method is used to estimate parameters of probability distributions with L-moments instead of the classical moments. In case of the lognormal distribution $LN(\mu,\sigma^2)$ two equation are to be sold with respect to two unknown parameters μ and σ^2. Using Hosking, 1990 or Bílková, 2013 we obtain

$$\begin{aligned} l_1 &= \exp(\mu+\sigma^2) \\ l_2 &= \left(2\Phi\left(\frac{\sigma}{\sqrt{2}}\right)-1\right)\exp(\mu+\sigma^2/2), \end{aligned} \tag{3}$$

where Φ is the cumulative distribution function of the standard normal distribution. For the normal distribution we have

$$\begin{aligned} l_1 &= \mu, \\ l_2 &= \sigma/\pi. \end{aligned} \tag{4}$$

Equations (3) were sold numerically, equations (4) have a straightforward solution.

There are right censored observations and exact values present in the analysed datasets. In (Wang et al., 2010) the method of evaluating the sample L-moments is given as (instead of the formula (2))

$$l_k = \sum_{j=1}^{n} x_{(j:n)} u_{j(k)}, \tag{5}$$

where $x_{(0)} = 0,\ p = k-r,\ q = r+1,$ $B_{p,q}(.)$ is the incomplete beta function, $\hat{S}$ is a Turnbull estimate of the survival function (Klein and Moeschberger, 1998) and

$$u_{j(k)} = \sum_{r=1}^{k} \frac{1}{k}(-1)^r \binom{k-1}{r}\left[B_{p,q}\left(1-\hat{S}(t_{j:n})\right) - B_{p,q}\left(1-\hat{S}(t_{j-1:n})\right)\right]. \tag{6}$$

The regularity conditions for consistency of estimates and for asymptotic normality are given in Wang et al., 2010. In the text of Wang and all. bootstrap is used in order to estimate standard deviations of parameters, in this text we use Monte Carlo simulation that is described below.

The estimator given in (5) is implemented in the R package (R CORE TEAM, 2015) lmomco by Asquit, 2015.

2 Description of the simulation

In the simulation study two component mixtures of lognormal distributions were analysed with one major component ($100\alpha\%$ of observations) and one small component ($100(1-\alpha)\%$ of observations) describing the contamination of data. We suppose the component membership to be known (this assumption is very limited). We suppose *cens* % of right censored observations included in the data. For selected values of parameters

$$\left(0<\alpha\leq 1, 0\leq cens\leq 1, \mu_1\in R, \mu_2\in R, \sigma_1^2>0, \sigma_2^2>0\right)$$

B=500 of independent random samples were generated and six models were estimated:

1. Two components mixture

    ML estimates of $\alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$

    MM estimates with the use of *L*-moments

    based on observed data (lognormal distribution is fitted)

    the logarithm of data (normal distribution is fitted),

2. One lognormal distribution

    ML estimates of μ, σ^2

    MM estimates with the use of *L*-moments based on normal and lognormal distributions.

The models evaluated in 2. reflect the robustness with respect to the contamination, estimated values are compared (as in (7)) with the theoretical values of the major component (referred as component 1). In the Table 1 the estimates of the mean squared deviation in the form

$$\frac{1}{n}\sum_{j=1}^{B}\left(\text{estimated value-theoretical value}\right)^2, \tag{7}$$

where B is number of generated samples (replications). In the text B was selected to be 500 due to the time consuming numerical procedures.

We use $LN(2;1)$ as a major population distribution (X_1, 95 and 90 percent of generated datasets, $\alpha = 0.9, 0.95$) and $LN(0;4)$ (X_2) as the distribution of contamination. We obtain

$$\text{components: } E(X_1) = 12.18, \sqrt{D(X_1)} = 15.97, E(X_2) = 2.72, \sqrt{D(X_2)} = 6.87$$

$$\text{mixture: } \alpha = 0.95 \; E(X) = 11.24, \sqrt{D(X)} = 15.57$$
$$0.9 \;\; E(X) = 11.71, \sqrt{D(X)} = 15.78.$$

The selected sample size of 500 observations gives (in the mean) for $\alpha = 0.9$ 450/50 observations and for $\alpha = 0.95$ the frequencies are 475/25. Moreover the datasets with no censored data (0 %) and 5, 10, 20 and 30 % of right censored data were used and values of (7) were evaluated ($cens = 0, 5, 10, 20,$ and 30).

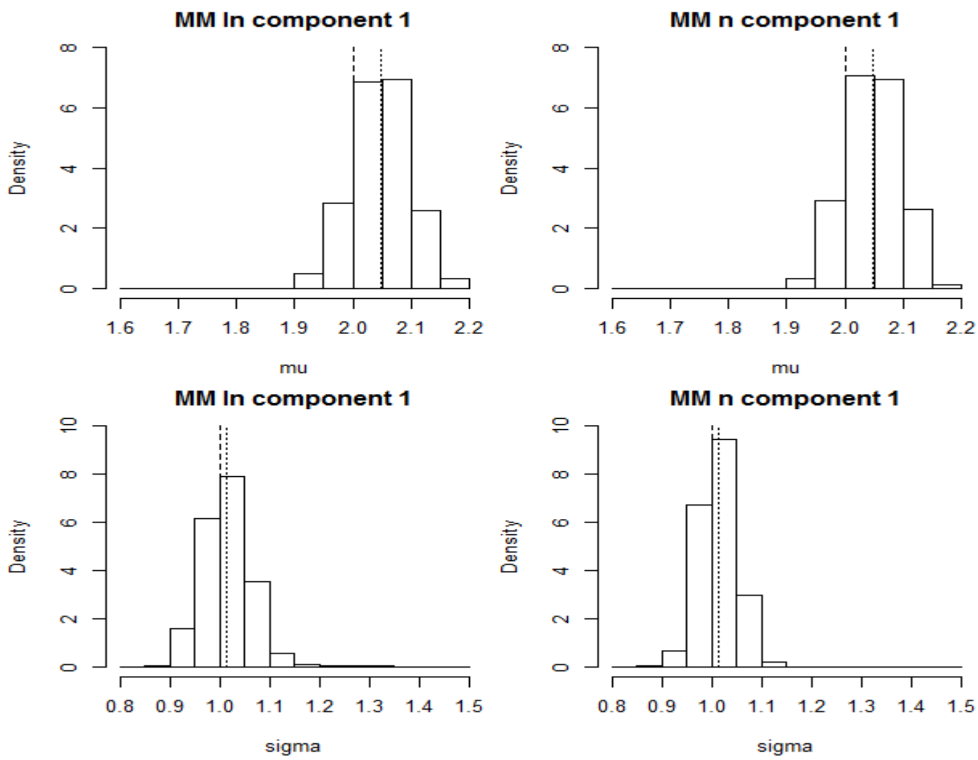
3 Results

The comparison of all above mentioned models is shown in the Table 1, where values of the mean squared deviation (7) are given. The results for the component 1 (columns 1-3) and for the mixture (columns 7-9) provides the comparison of approaches and the impact of contaminated data to the estimates (when one distribution for the whole sample is used instead of the mixture). In the Tables 1 (as well in the Table 2) the estimates in the second group (columns 4-6) are evaluated from a very small sample (50 or 25 observations as stated above) of censored observations; middle 3 columns in the table are included but no definite and valuable results can be derived. We refer to the MLE estimates in the lognormal distribution as (MLE ln) and to the moment method as MM, with *ln* for the lognormal and *n* for the normal distribution. All three types of estimates seems to be comparable, the use of normal fit to logarithmic data is preferred to the fit of the lognormal distribution to the data.

Histograms of estimated parameters (from 500 replications) are given only for models based on the moment method and for the choice $\alpha = 0.9, cens = 5\,\%$. In the Figure 1 histograms for the mixture model (we use information about component membership, the quality of fit is expected to be better) and in the Figure 2 the single lognormal distribution model. By the dashed lines the theoretical value and the mean of estimates is shown. The distributions of parameters are approximately normally distributed.

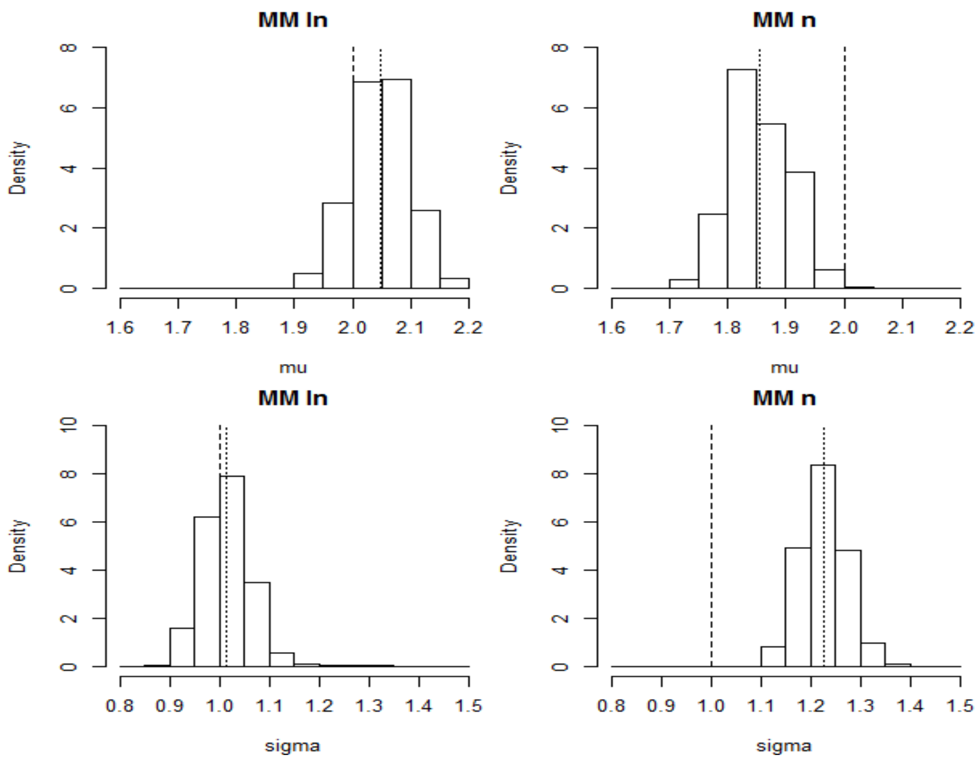
In the Table 2 estimated expected values (first row) and standard deviations (second row) are given for both components of the mixture and for the fit of one distribution to all data (in the same structure as in the Table 1). Moreover the distribution of estimated expected values is shown in the Figures 3 and 4 for $\alpha = 0.9, cens = 0.05$ and $0.15.$ Histograms for the component 1 are in the first row (lognormal fit (left) and normal fit (right)) and a single fitted distribution in the second row (lognormal fit (left) and normal fit (right)). The distributions of

Fig. 1: Histograms of estimated parameters with 5 % of censored data; mixture model



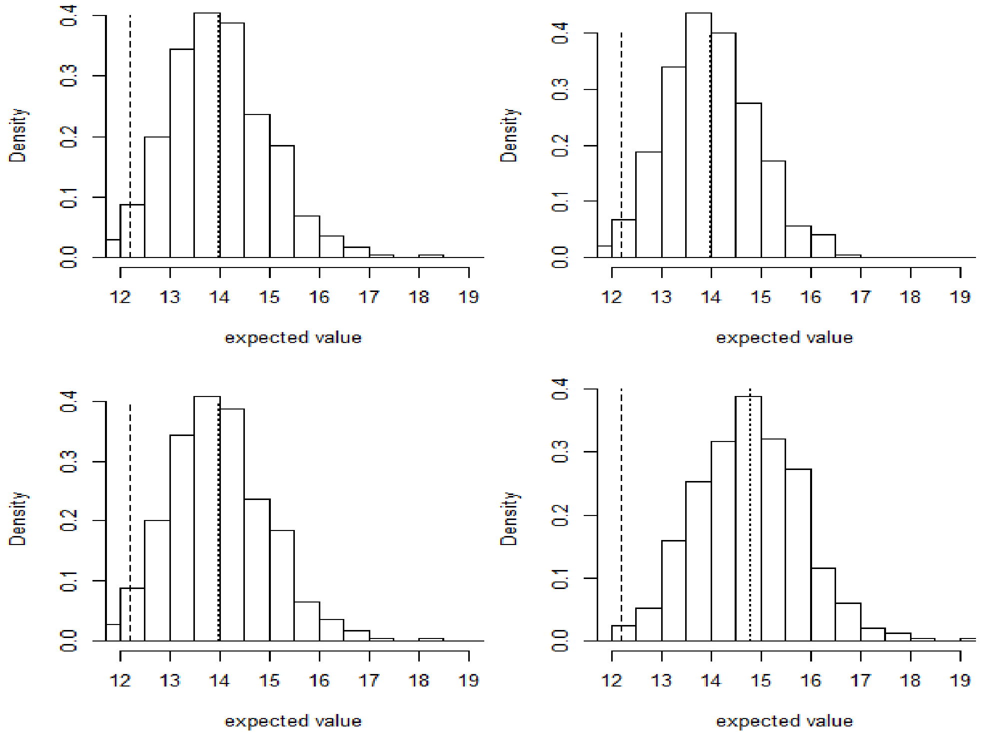
Source: own calculations

Fig. 2: Histograms of estimated parameters with 5 % of censored data; single distribution



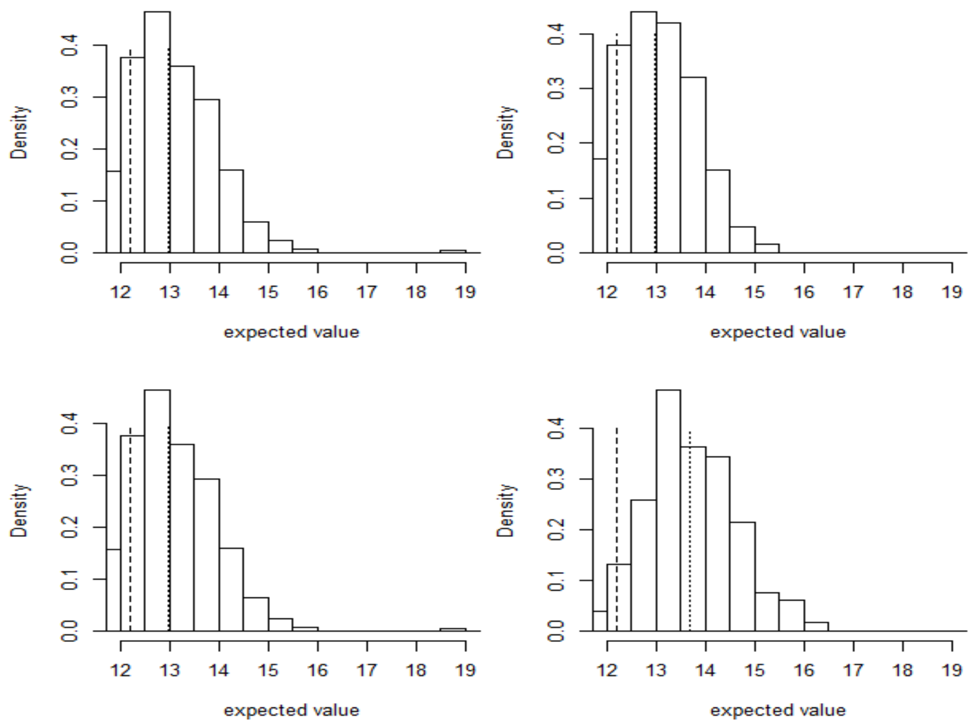
Source: own calculations

Fig. 3: Histograms of estimated expected values with 15 % of censored data, $\alpha = 0.9$. Dashed lines mark the theoretical value and the mean from estimates.



Source: own calculations

Fig. 4: Histograms of estimated expected values with 5 % of censored data, $\alpha = 0.9$



Source: own calculations

estimates are positively skewed and the theoretical values are overestimated. The dependence on the censoring is well visible.

Tab. 1: Comparison of models; values of (7) for estimates of parameters.

	Component 1			Component 2			Single distribution		
Model	MM *ln*	MM *n*	MLE	MM *ln*	MM *n*	MLE	MM *ln*	MM *n*	MLE
$\alpha = 0.9, cens = 0$	0.0024	0.0023	0.0023	0.1392	0.0874	0.0874	0.0024	0.0424	0.0424
	0.0020	0.0011	0.0011	0.1198	0.0391	0.0382	0.0020	0.0510	0.0870
$\alpha = 0.9, cens = 5$	0.0046	0.0044	0.0044	0.1896	0.0962	0.0959	0.0046	0.0244	0.0227
	0.0025	0.0014	0.0014	0.1218	0.0436	0.0432	0.0025	0.0547	0.0960
$\alpha = 0.9, cens = 10$	0.0124	0.0124	0.0124	0.2808	0.1232	0.1229	0.0124	0.0108	0.0088
	0.0039	0.0023	0.0023	0.1255	0.0466	0.0469	0.0039	0.0604	0.1075
$\alpha = 0.9, cens = 20$	0.0466	0.0467	0.0464	0.6940	0.1491	0.2783	0.0076	0.0076	0.0051
	0.0092	0.0061	0.0059	0.0599	0.0466	0.0580	0.0092	0.0705	0.1299
$\alpha = 0.9, cens = 30$	0.1224	0.1214	0.1201	1.3728	0.5350	0.5474	0.1224	0.0408	0.0574
	0.0179	0.0143	0.0133	0.1817	0.0820	0.0934	0.0179	0.0895	0.1665
$\alpha = 0.95, cens = 0$	0.0022	0.0021	0.0021	0.2466	0.1561	0.1561	0.0022	0.0137	0.0137
	0.0022	0.0012	0.0012	0.2093	0.0825	0.0787	0.0022	0.0142	0.0266
$\alpha = 0.95, cens = 5$	0.0042	0.0041	0.0041	0.3719	0.1828	0.1833	0.0042	0.0047	0.0043
	0.0025	0.0013	0.0012	0.2218	0.0908	0.0894	0.0025	0.0165	0.0306
$\alpha = 0.95, cens = 10$	0.0118	0.0119	0.0118	0.4801	0.2193	0.2212	0.0119	0.0027	0.0028
	0.0044	0.0024	0.0024	0.2193	0.0936	0.0930	0.0044	0.0216	0.0394
$\alpha = 0.95, cens = 20$	0.0464	0.0468	0.0464	1.1616	0.3925	0.4076	0.0463	0.0192	0.0229
	0.0090	0.0056	0.0054	0.2545	0.1261	0.1356	0.0090	0.0300	0.0541
$\alpha = 0.95, cens = 30$	0.1200	0.1185	0.1173	2.0088	0.7820	0.7242	0.1199	0.0746	0.0864
	0.0159	0.0125	0.0133	0.3456	0.1908	0.1513	0.0159	0.0426	0.0732

Source: own computations

Conclusion

In the text a robust alternative to the classical method of maximum likelihood for the estimation of parameters is analysed. In the text the results of the simple simulation are presented and the properties of the method are briefly commented. The use of lognormal distributions enables two models to be fitted – lognormal distribution to the original data and normal distribution to the logarithms of data.

The results were found to be comparable to the maximum likelihood estimates and the method seems to be more robust with respect to the contamination of data. For a single distribution the robust methods are superior to MLE as it was expected. Both MM methods are comparable in the mixture model, but the former method seems to be better in case of the single distribution.

The MM method is (according to presented and not presented results) worth studying and potentially efficient especially in case of a heavy censoring.

Tab. 2: Comparison of models; estimated values of the expected value and the variance

	Component 1			Component 2			Single distribution		
Model	MM *ln*	MM *n*	MLE	MM *ln*	MM *n*	MLE	MM *ln*	MM *n*	MLE
$\alpha = 0.9, cens = 0$	12.2	12.2	12.2	7.5	8.0	8.2	12.2	12.8	14.0
	16.1	16.1	16.1	271.4	107.3	115.0	16.1	24.1	29.5
$\alpha = 0.9, cens = 5$	13.0	13.0	13.0	8.8	9.5	9.7	13.0	13.6	15.1
	17.6	17.5	17.5	879.8	172.1	176.3	17.6	26.0	32.8
$\alpha = 0.9, cens = 10$	14.0	14.0	14.0	9.8	11.6	12.0	14.0	14.8	16.6
	19.5	19.5	19.5	259.8	273.2	307.6	19.5	28.8	37.4
$\alpha = 0.9, cens = 20$	16.2	16.2	16.1	12.7	16.1	16.9	16.2	17.2	19.9
	24.4	23.8	23.7	286.4			24.4	34.5	47.3
$\alpha = 0.9, cens = 30$	19.4	19.4	19.3	16.5	25.3	27.7	19.4	20.8	25.1
	31.6	30.8	30.4	374.6			31.6	44.5	65.0
$\alpha = 0.95, cens = 0$	12.1	12.1	12.1	7.5	8.1	8.5	12.1	12.4	13.0
	16.1	16.0	16.0	444.5	249.5	268.6	16.1	19.5	22.1
$\alpha = 0.95, cens = 5$	13.0	13.0	13.0	8.6	10.3	11.0	13.0	13.2	14.0
	17.7	17.5	17.5	191.3	507.3	719.6	17.7	21.3	24.3
$\alpha = 0.95, cens = 10$	14.0	13.9	13.9	8.6	11.6	12.3	14.0	14.3	15.3
	19.7	19.4	19.4	147.5	823.3	779.3	19.7	23.6	27.6
$\alpha = 0.95, cens = 20$	16.2	16.1	16.1	15.4	26.6	32.6	16.2	16.7	18.1
	24.3	23.7	23.6	469.4			24.3	28.8	34.7
$\alpha = 0.95, cens = 30$	19.1	19.2	19.1	16.5	34.9	49.9	19.1	19.9	22.0
	31.0	30.3	30.0	357.8			31.0	36.5	45.3

Source: own computations

References

1. Asquith, W. H. (2015). lmomco: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.1.3. Retrieved from: http://www.cran.r-project.org/package=lmomco

2. Bílková, D. (2013). Modeling of Wage Distribution in Recent Years in the Czech Republic Using L-moments and the Prediction of Wage Distribution by Industry. *E & M Ekonomie a Management*, 16, 42-54.
3. Delicadoa, P., Goriab, M.N. (2008). A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. *Computational Statistics & Data Analysis*, 52, 1661-1673.
4. Delignette-Muller M.L., Dutang, C. (2015). Fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64, 1-34.
   Retrieved from: http://www.jstatsoft.org/v64/i04/.
5. Hosking, J. R. M., Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29, 339–349.
6. Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*. *Series B (Methodological)*, 52, 105-124.
7. Klein, J.P., Moeschberger, M.L. (1998). *Survival Analysis*, Techniques for Censored and Truncated Data. Springer.
8. Malá, I. (2015). Modelling of unemployment duration in the Czech Republic with the use of L-moments. In: *Applications of Mathematics and Statistics in Economics – AMSE*[CD ROM]. Jindřichův Hradec, 02.09.2015 – 06.09.2015. Praha: University of Economics, Prague, Oeconomica Publishing House, 2015.
9. Mudholkar, G. S., Hutson, A. D. (1998). LQ-moments: Analogs of L-moments. Journal of Statistical Planning and Inference, 71, 191-208.
10. R CORE TEAM 2015: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012. Retrieved from: http://www.r-project.org/.
11. Wang, D., Hutson, A. D., Miecznikowski, J. C. (2010). L-moment estimation for parametric survival models given censored data. *Statistical Methodology*, 7, 655-667.

Contact

Ivana Malá
University of Economics, Prague
W. Churchill pl. 4
130 67 Praha 3
malai@vse.cz