TRIMMED L-MOMENTS: USE IN MODELING OF THE WAGE DISTRIBUTION

DIANA BÍLKOVÁ
University of Economics, Prague; Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mail: diana.bilkova@vse.cz

Abstract

The present paper deals with the use of order statistics as a method of TL-moments of parameter estimation. L-moments, having been introduced as a robust alternative to classical moments of probability distributions, provide some theoretical advantages over conventional moments. However, L-moments and their estimations lack some robust features associated with TL-moments that represent an alternative robust version of the former, called trimmed L-moments. This paper focuses on the use of TL-moments in the construction of models of the wage distribution. Three-parametric lognormal curves represent the basic theoretical distribution whose parameters were estimated simultaneously by three methods of point parameter estimation – those of TL-moments, L-moments and maximum likelihood in combination with Cohen's method, their accuracy being then evaluated. 328 wage distributions formed the subject of the research, whose purpose is to highlight the advantages of the trimmed L-moment method over the classical method of L-moments in its application to economic data, namely in modelling the wage distribution. Another objective of the research is to verify the higher accuracy of the classical L-moments method compared to the maximum likelihood one, despite the latter being considered one of the most accurate parameter estimation methods. Both the higher accuracy of the trimmed L-moments method in comparison to that of L-moments and advantages of the latter over the maximum likelihood method have been demonstrated.

Key words: *TL-moments of probability distribution, sample TL-moments, models of wage distribution.*

1. Introduction

In all developed countries, the economist interest in wages and incomes of the population comes from efforts for objective resolution of issues related to living standards of the population. Forecasts of wage and income distribution constructed on the basis of wage and income models allow to solve the questions of this type successfully. Wage and income models can be used in assessing the standard of living or in international comparisons of living standards, which is the sum of living conditions. Entrepreneurs would also take into account the distributions of wages and incomes of the population when considering their market opportunities. These distributions can be also used when considering the various tax burden, etc. This paper is focused only on the wage distribution models.

The substance of the conventional moment method of parameter estimation is based on the fact that the sample and relevant theoretical moments of the probability distribution are built into equality. General and central moments can be combined. This method of parameter estimation is although very easy to use, however it is very inaccurate. In particular, the estimation of the theoretical variance using its sample counterpart is substantially inaccurate.

The main aim of this paper is to compare the accuracy of the use of trimmed L-moments with accuracy of the use of classic L-moments when estimating the parameters of three-parametric lognormal curves as the models of wage distribution. Another aim is to compare the accuracy between the use of conventional method of L-moments and maximum likelihood method.

The advantages of L-moment method are obvious when applied to small data sets, predominantly in the fields of hydrology, meteorology and climatology, considering extreme precipitation in particular. L-moments have been introduced as a robust alternative to classical moments of probability distributions, see Hosking (1990). However, L-moments and their estimates lack some robust features specific to TL-moments, the latter representing an alternative robust version of the former, the so-called trimmed L-moments. The main aim of this paper is to utilize the two methods of parameter estimation in large data sets from the economic sphere and compare their accuracy to that of the maximum likelihood method. Three-parametric lognormal curves represent the basic theoretical probability distribution (see Johnson et al., 1994).

A number of authors (e.g. Bartošová and Longford, 2014; Pavelka and Löster, 2013; Pivoňka and Löster, 2014; Šimpach and Pechrová, 2013) deal with the issue of the labor market and living standards of the population of the Czech Republic. A number of authors (e.g. Malá, 2014; Marek, 2013; Marek and Vrabec, 2013; Pavelka et al., 2014) study directly the problems of wages and incomes. Such authors as Malá (2013) and Malec and Malec (2013) research some statistical methods, which can used for economic data.

The first part of this paper is purely theoretical, and it contains a detailed description of the methods used including a description of the data used. The second part of this article is then applicable. It contains the results obtained, discussion and conclusions received.

2. Theory and Methods

An alternative robust version of L-moments will be introduced now. This modification of L-moments is called the “trimmed L-moments” and is noted TL-moments. In this modification of L-moments, the expected values of order statistics of a random sample (in L-moments definition of probability distributions) are replaced by the expected values of order statistics of a larger random sample, the sample size growing in such a way that it corresponds to the total size of the adjustment, as shown below.

TL-moments have certain advantages over conventional L-moments and central moments. TL-moment of probability distribution may exist even if the corresponding L-moment or central moment of this probability distribution does not exist, as it is the case of the Cauchy distribution. Sample TL-moments are more resistant to outliers in the data. The method of TL-moments is not intended to replace the existing robust methods, but rather as their supplement, particularly in situations with outliers in the data.

2.1 TL-moments of Probability Distribution

Let X be a continuous random variable that has a distribution with distribution function $F(x)$ and with quantile function $x(F)$. Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics of random sample of sample size n, which comes from the distribution of random variable X.

In this alternative robust modification of L-moments, the expected value $E(X_{r-j:r})$ is replaced by that of $E(X_{r+t_1-j:r+t_1+t_2})$. For each r, we increase the size of a random sample from the original r to $r + t_1 + t_2$, working only with the expected values of these r modified order statistics $X_{t_1+1:r+t_1+t_2}$, $X_{t_1+2:r+t_1+t_2}$, …, $X_{t_1+r:r+t_1+t_2}$ by trimming t_1 and t_2 (the lowest and highest value, respectively, from a conceptual sample). This modification is called the r-th trimmed L-moment (TL-moment) and marked $\lambda_r^{(t_1,t_2)}$ Thus, TL-moment of the r-th order of a random variable X is defined as

$$\lambda_r^{(t_1,t_2)} = \frac{1}{r}\cdot\sum_{j=0}^{r-1}(-1)^j\cdot\binom{r-1}{j}\cdot E(X_{r+t1-j:r+t1+t2}), \quad r = 1, 2, \ldots, \tag{1}$$

as can be found in Elamir and Seheult (2003). It is apparent that the TL-moments simplify to L-moments, when $t_1 = t_2 = 0$. Although we can also consider applications, where the values of trimming are not equal, i.e. $t_1 \neq t_2$, we focus here only on symmetric case $t_1 = t_2 = t$. Then equation (1) can be rewritten

$$\lambda_r^{(t)} = \frac{1}{r}\cdot\sum_{j=0}^{r-1}(-1)^j\cdot\binom{r-1}{j}\cdot E(X_{r+t-j:r+2t}), \quad r = 1, 2, \ldots. \tag{2}$$

Thus, for example, $\lambda_1^{(t)} = E(X_{1+t:1+2t})$ is the expected value of median from conceptual random sample of sample size $1 + 2t$. It is necessary here to note that $\lambda_1^{(t)}$ is equal to zero for distributions, which are symmetrical around zero.

First four TL-moments have the form for $t = 1$ as follows

$$\lambda_1^{(1)} = E(X_{2:3}), \tag{3}$$

$$\lambda_2^{(1)} = \frac{1}{2}E(X_{3:4} - X_{2:4}), \tag{4}$$

$$\lambda_3^{(1)} = \frac{1}{3}E(X_{4:5} - 2X_{3:5} + X_{2:5}), \tag{5}$$

$$\lambda_4^{(1)} = \frac{1}{4}E(X_{5:6} - 3X_{4:6} + 3X_{3:6} - X_{2:6}). \tag{6}$$

Note that the measures of location (level), variability, skewness and kurtosis of the probability distribution are based on $\lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_3^{(1)}$ a $\lambda_4^{(1)}$.

The expected value $E(X_{r:n})$ can be written using the formula (e.g. Elamir and Seheult, 2003)

$$E(X_{r:n}) = \frac{n!}{(r-1)!\cdot(n-r)!}\cdot\int_0^1 x(F)\cdot[F(x)]^{r-1}\cdot[1-F(x)]^{n-r}\mathrm{d}F(x). \tag{7}$$

Using equation (7) we can re-express the right side of equation (2) as

$$\lambda_r^{(t)} = \frac{1}{r}\cdot\sum_{j=0}^{r-1}(-1)^j\cdot\binom{r-1}{j}\cdot\frac{(r+2t)!}{(r+t-j-1)!\cdot(t+j)!}\cdot\int_0^1 x(F)\cdot[F(x)]^{r+t-j-1}\cdot[1-F(x)]^{t+j}\,\mathrm{d}F(x), \; r = 1, 2, \ldots. \tag{8}$$

It is necessary to be noted here that $\lambda_r^{(0)} = \lambda_r$ is a normal the r-th L-moment without any trimming.

Expressions (3) − (6) for the first four TL-moments, where $t = 1$, can be written in an alternative manner

$$\lambda_1^{(1)} = 6\cdot\int_0^1 x(F)\cdot[F(x)]\cdot[1-F(x)]\,\mathrm{d}F(x), \tag{9}$$

$$\lambda_2^{(1)} = 6 \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot [2\,F(x) - 1]\, \mathrm{d}\, F(x), \tag{10}$$

$$\lambda_3^{(1)} = \frac{20}{3} \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot \{5[F(x)]^2 - 5\,F(x) + 1\}\, \mathrm{d}\, F(x), \tag{11}$$

$$\lambda_4^{(1)} = \frac{15}{2} \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot \{14[F(x)]^3 - 21[F(x)]^2 + 9[F(x)] - 1]\, \mathrm{d}\, F(x). \tag{12}$$

Distribution may be identified by its TL-moments, although some of its L-moments or conventional central moments do not exit; for example $\lambda_1^{(1)}$ (expected value of median of conceptual random sample of sample size three) exists for Cauchy's distribution, although the first L-moment λ_1 does not exist.

TL-skewness $\tau_3^{(t)}$ and TL-kurtosis $\tau_4^{(t)}$ are defined analogously as L-skewness and L-kurtosis

$$\tau_3^{(t)} = \frac{\lambda_3^{(t)}}{\lambda_2^{(t)}}, \tag{13}$$

$$\tau_4^{(t)} = \frac{\lambda_4^{(t)}}{\lambda_2^{(t)}}. \tag{14}$$

2.2 Sample TL-moments

Let $x_1, x_2, \ldots, x_n$ is a sample and $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$ is an ordered sample. Expression

$$\hat{E}(X_{j+1:j+l+1}) = \frac{1}{\binom{n}{j+l+1}} \cdot \sum_{i=1}^{n} \binom{i-1}{j} \cdot \binom{n-i}{l} \cdot x_{i:n}, \tag{15}$$

is considered to be an unbiased estimation of expected value of the $(j + 1)$-th order statistic $X_{j+1:j+l+1}$ in conceptual random sample of sample size $(j + l + 1)$, as follows from Elamir and Seheult (2003). Now we will assume that we replace the expression $E(X_{r+t-j:r+2t})$ by its unbiased estimation in the definition of the r-th TL-moment $\lambda_r^{(t)}$ in (2)

$$\hat{E}(X_{r+t-j:r+2t}) = \frac{1}{\binom{n}{r+2t}} \cdot \sum_{i=1}^{n} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j} \cdot x_{i:n}, \tag{16}$$

which we gain by assigning $j \to r + t - j - 1$ a $l \to t + j$ in (15). Now we obtain the r-th sample TL-moment

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \hat{E}(X_{r+t-j:r+2t}), \quad r = 1, 2, \ldots, n - 2t, \tag{17}$$

i.e.

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \frac{1}{\binom{n}{r+2t}} \cdot \sum_{i=1}^{n} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j} \cdot x_{i:n}, \quad r = 1, 2, \ldots, n - 2t, \tag{18}$$

which is unbiased estimation of the r-th TL-moment $\lambda_r^{(t)}$. Note that for each $j = 0, 1, \ldots, r-1$, values $x_{i:n}$ in (18) are nonzero only for $r + t - j \leq i \leq n - t - j$ due to the combinatorial numbers. Simple adjustment of the equation (18) provides an alternative linear form

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{i=r+t}^{n-t} \left[\frac{\sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j}}{\binom{n}{r+2t}} \right] \cdot x_{i:n}, \tag{19}$$

as noted by Elamir and Seheult (2003). For example, we obtain for $r = 1$ for the first sample TL-moment

$$l_1^{(t)} = \sum_{i=t+1}^{n-t} w_{i:n}^{(t)} \cdot x_{i:n}, \tag{20}$$

where the weights are given by

$$w_{i:n}^{(t)} = \frac{\binom{i-1}{t} \cdot \binom{n-i}{t}}{\binom{n}{2t+1}}. \tag{21}$$

The above results can be used to estimate TL-skewness and TL-kurtosis by simple ratios

$$t_3^{(t)} = \frac{l_3^{(t)}}{l_2^{(t)}}, \tag{22}$$

$$t_4^{(t)} = \frac{l_4^{(t)}}{l_2^{(t)}}. \tag{23}$$

We can choose integer $t = n\alpha$ representing the amount of the adjustment from each end of the sample, where α is a certain proportion, where $0 \leq \alpha < 0{,}5$. More on the TL-moments is for example in Elamir and Seheult (2003). Table 1 contains the expressions and ratios for TL-moments and expressions for parameter estimations of chosen probability distributions obtained using the method of TL-moments ($t = 1$).

2.3 Estimation Accuracy Evaluation

It is also necessary to assess the suitability of constructed model or choose a model from several alternatives, which is made by some criterion, which can be a sum of absolute deviations of the observed and theoretical frequencies for all intervals

$$S = \sum_{i=1}^{k} \left| n_i - n\,\pi_i \right| \tag{24}$$

or known criterion χ^2

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\,\pi_i)^2}{n\,\pi_i}, \tag{25}$$

where n_i are the observed frequencies in individual intervals, π_i are the theoretical probabilities of membership of statistical unit into the i-th interval, n is the total sample size of corresponding statistical file, $n \cdot \pi_i$ are the theoretical frequencies in individual intervals, $i = 1, 2, \ldots, k$, and k is the number of intervals.

Table 1: Formulas for TL-moments and ratios of TL-moments and formulas for parameter estimates made by the method of TL-moments of chosen probability distributions (t = 1)

Distribution	TL-moments and ratios of TL-moments	Parameter estimation	Distribution	TL-moments and ratios of TL-moments	Parameter estimation
Normal	$\lambda_1^{(1)} = \mu$ $\lambda_2^{(1)} = 0{,}297\,\sigma$ $\tau_3^{(1)} = 0$ $\tau_4^{(1)} = 0{,}062$	$\hat{\mu} = l_1^{(1)}$ $\hat{\sigma} = \frac{l_2^{(1)}}{0{,}297}$	Cauchy	$\lambda_1^{(1)} = \mu$ $\lambda_2^{(1)} = 0{,}698\,\sigma$ $\tau_3^{(1)} = 0$ $\tau_4^{(1)} = 0{,}343$	$\hat{\mu} = l_1^{(1)}$ $\hat{\sigma} = \frac{l_2^{(1)}}{0{,}698}$
Logistic	$\lambda_1^{(1)} = \mu$ $\lambda_2^{(1)} = 0{,}500\,\sigma$ $\tau_3^{(1)} = 0$ $\tau_4^{(1)} = 0{,}083$	$\hat{\mu} = l_1^{(1)}$ $\hat{\sigma} = 2 l_2^{(1)}$	Exponential	$\lambda_1^{(1)} = \frac{5\alpha}{6}$ $\lambda_2^{(1)} = \frac{\alpha}{4}$ $\tau_3^{(1)} = \frac{2}{9}$ $\tau_4^{(1)} = \frac{1}{12}$	$\hat{\alpha} = \frac{6 l_1^{(1)}}{5}$

Source: Elamir and Seheult (2003), the author.

The question of the appropriateness of the given curve for model of the distribution of wage is not entirely conventional mathematical-statistical problem in which we test the null hypothesis

H_0: The sample comes from the supposed theoretical distribution

against the alternative hypothesis

H_1: non H_0,

because in goodness of fit tests in the case of wage distribution we meet frequently with the fact that we work with large sample sizes and therefore the tests would almost always lead to the rejection of the null hypothesis. This results not only from the fact that with such large sample sizes the power of the test is so high at the chosen significance level that the test uncovers all the slightest deviations of the actual wage distribution and a model, but it also results from the principle of construction of the test. But practically we are not interested in such small deviations, so only gross agreement of the model with reality is sufficient and we so called "borrow" the model (curve). Test criterion χ^2 can be used in that direction only tentatively. When evaluating the suitability of the model we proceed to a large extent subjective and we rely on experience and logical analysis. However, this criterion is commonly applied when selecting an appropriate model of wage distribution. Akaike Information Criterion and Bayesian Information Criterion require use of a likelihood function, but two other methods were used here to estimate the parameters.

3. Database

The database of the research consists in employees of the Czech Republic. There are a total set of all employees of the Czech Republic together and further the partial sets broken down by various demographic and socio-economic factors. The researched variable is the gross monthly wage in CZK (nominal wage). Data come from the official website of the Czech Statistical Office. The data was in the form of interval frequency distribution, since the individual data is not currently available. Researched period represents years 2003−2010. Employees of the Czech Republic were classified according to gender, job classification (CZ-ISCO), the classification of economic activities, age and educational attainment. Branch classification of economic activities (OKEC) has been replaced by classification of economic activities (CZ-NACE) during researched period. This fast therefore disrupts the continuity of the obtained time series during the analysis period.

The main classes of job classification CZ-ISCO form: managers (code 1000); professionals (code 2000); technicians and associate professionals (code 3000); clerical support workers (code 4000); service and sales workers (code 5000); skilled agricultural, forestry and fishery workers (code 6000); craft and related trades workers (code 7000); plant and machine operators, and assemblers (code 8000); elementary occupations (code 9000).

The main classes of branch classification of economic activities − OKEC (years 2003–2008) are: A-B − agriculture, fishing; C-E − industry; F − construction; G − trade, repairs; H − hotels and restaurants; I − transport, storage; J − financial intermediation; K − real estate, renting; L − public administration; M − education; N − health; O − other services.

The main classes of classification of economic activities − CZ-NACE (years 2009–2010) represent: A − agriculture, forestry and fishing; B-E − industry; F − construction; G − wholesale and retail trade, repair of motor vehicles and motorcycles; H − transportation and storage; I − accommodation and food service activities; J − information and communication; K − financial and insurance activities; L − real estate activities; M − professional, scientific and technical activities; N − administrative and support service activities; O − public administration and defense, compulsory social security; P − education; Q − human health and social work activities; R − arts, entertainment and recreation; S − other service activities.

Classification by age includes the following age intervals: to 19 years; from 20 to 24 years; from 25 to 29 years; from 30 to 34 years; from 35 to 39 years; from 40 to 44 years; from 45 to 49 years; from 50 to 54 years; from 55 to 59 years; from 60 to 64 years; from 65 years.
Classification according to educational attainment distinguishes the following five levels of educational attainment of the employee: primary education; apprenticeship; secondary with GCE; higher post-secondary schools; university.

4. Results and Discussion

Table 2 shows parameter estimations obtained using the three methods and the value of the criterion (24) for the total wage distribution in the Czech Republic, giving an approximate description of research outcomes for all 328 wage distributions. We found out that the method of TL-moments provided the most accurate results in almost all, with minor exceptions, wage distribution cases, the deviations having occurred mainly at both ends of the wage distribution due to extreme open intervals of an interval frequency distribution. Table 2 indicates that for the total wage distribution set for the whole Czech Republic in 2003–2010, the method of TL-moments always yields the most accurate output in terms of the S criterion. As for the research of all 328 wage distributions, the second most accurate results were produced by the method of L-moments, the deviations having occurred again at both ends of the distribution in particular. The latter method brought the second most accurate results in terms of all total wage distribution data sets over the period 2003–2010. In the majority of cases, the maximum likelihood method was the third most accurate approach. (For all cases, see Table 2.)

Table 2: Parameter estimations obtained using the three methods of point parameter estimation and the value of *S* criterion for total wage distribution in the Czech Republic

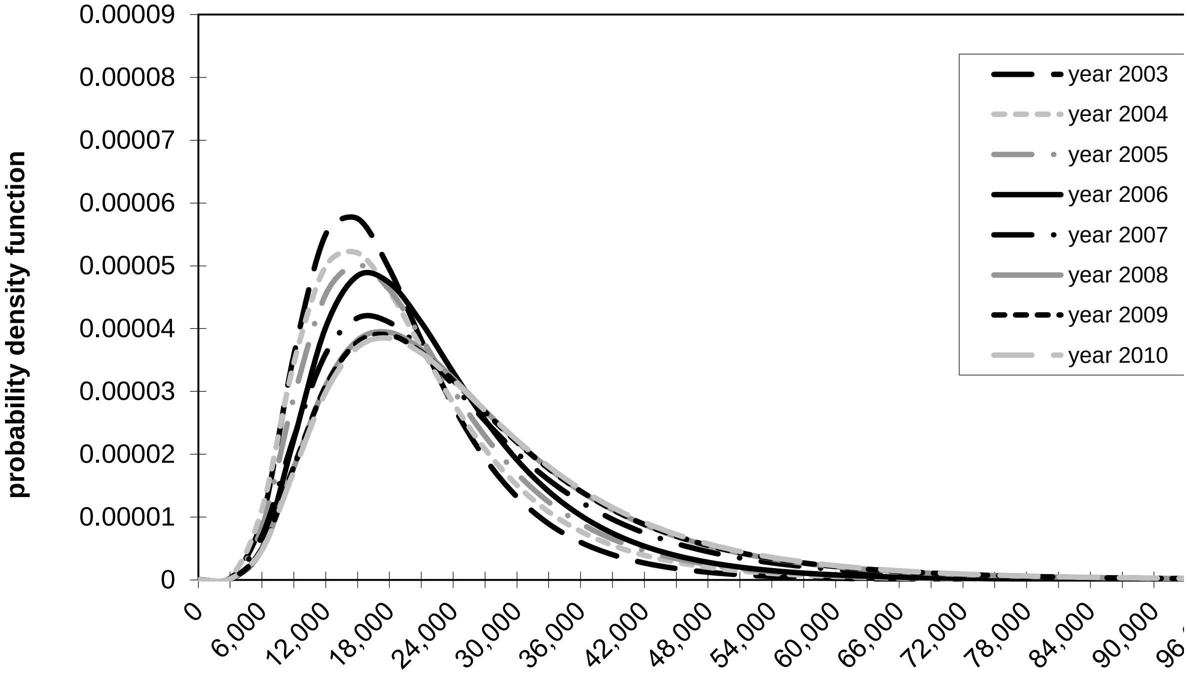
Year	Method of TL-moments			Method of L-moments			Maximum likelihood method		
	μ	σ^2	θ	μ	σ^2	θ	μ	σ^2	θ
2003	9.060	0.631	9,066	9.018	0.608	7,664	9.741	0.197	2.071
2004	9.215	0.581	8,552	9.241	0.508	6,541	9.780	0.232	0.222
2005	9.277	0.573	8,873	9.283	0.515	6,977	9.834	0.229	0.270
2006	9.314	0.578	9,383	9.284	0.543	7,868	9.891	0.211	0.591
2007	9.382	0.681	10,028	9.388	0.601	7,903	9.950	0.268	0.162
2008	9.439	0.689	10,898	9.423	0.624	8,755	10.017	0.264	0.190
2009	9.444	0.704	10,641	9.431	0.631	8,685	10.020	0.269	0.200
2010	9.482	0.681	10,617	9.453	0.621	8,746	10.034	0.270	0.201

Year	Criterion *S*	Criterion *S*	Criterion *S*
2003	108,437.01	133,320.79	248,331.74
2004	146,509.34	248,438.78	281,541.41
2005	137,422.05	231,978.79	311,008.23
2006	149,144.68	216,373.24	325,055.67
2007	198,670.74	366,202.87	370,373.62
2008	206,698.93	357,668.48	391,346.02
2009	193,559.55	335,999.20	359,736.37
2010	210,434.01	235,483.68	389,551.44

Source: the author.

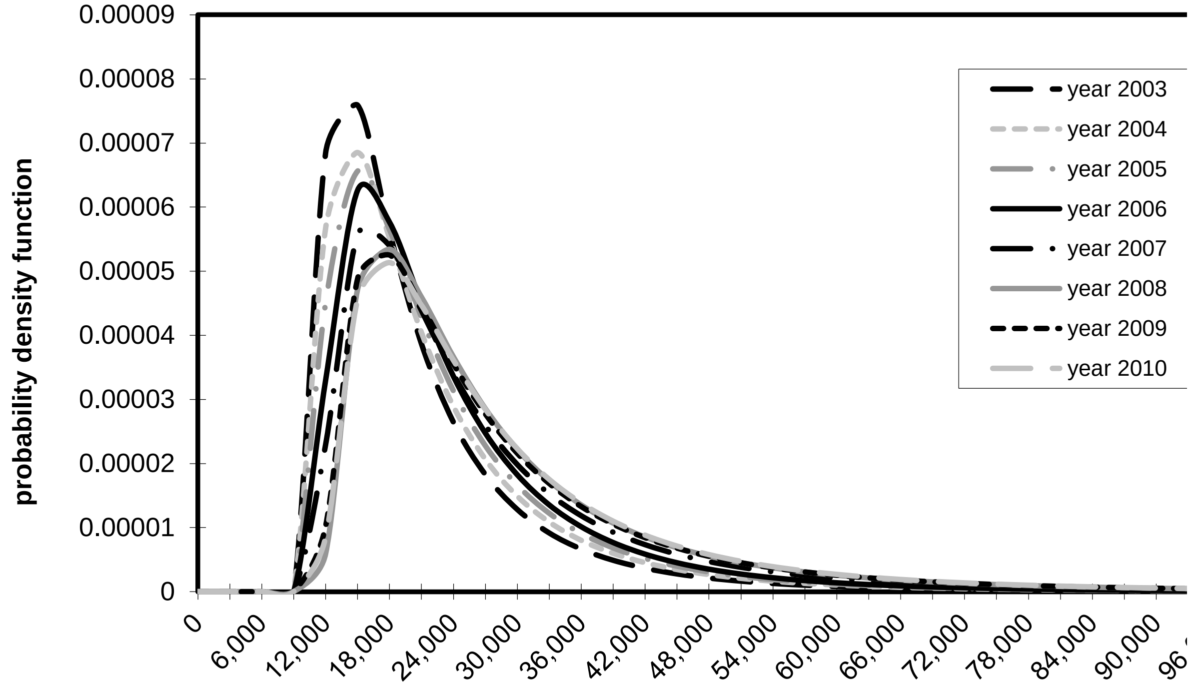
Figures 1–2 present the development of the probability density function of three-parameter lognormal curves with the parameters estimated employing the methods of TL-moments, L-moments and maximum likelihood, models of the total wage distribution for all employees of the Czech Republic being examined over the period 2003–2010 again. In comparison to the results obtained by the analysis of income distribution, we can see that the shapes of lognormal curves with the parameters estimated using L-moments and maximum likelihood methods (Figures 2 and 3) are similar to each other, differing greatly, however, from the shape of three-parameter lognormal curves with the parameters estimated by the method of TL-moments (Figure 1).

Figure 1: Development of the probability density function of three-parameter lognormal curves with parameters estimated using the method of TL-moments



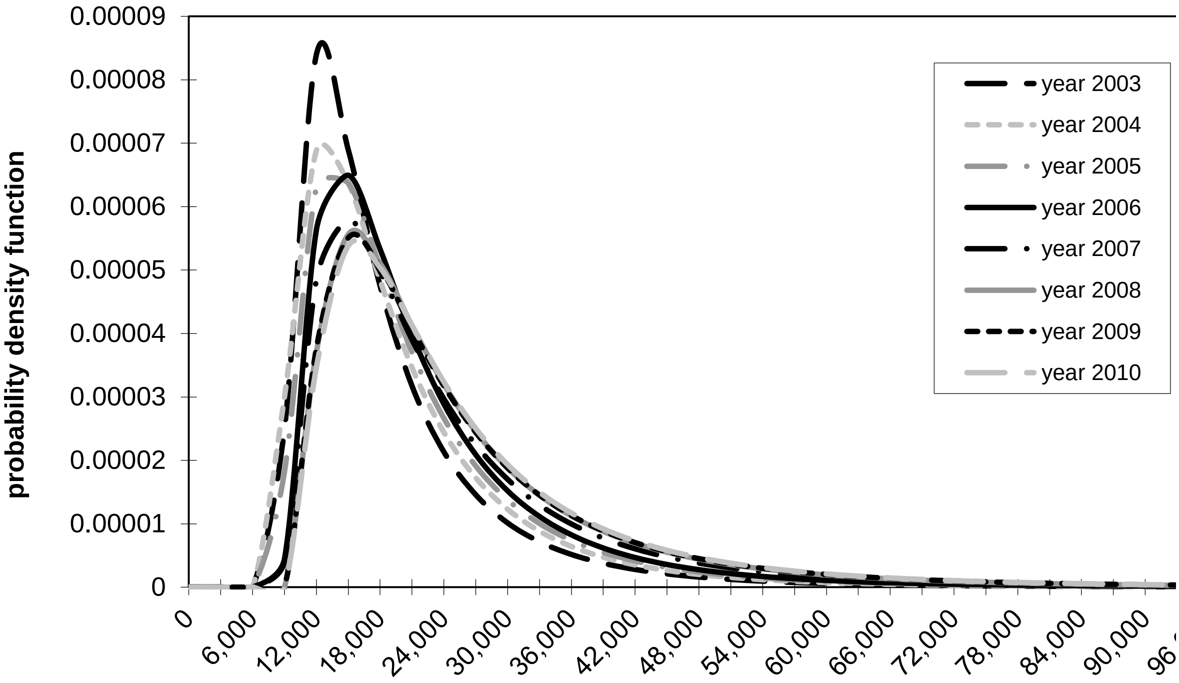
Source: the author.

Figure 2: Development of the probability density function of three-parameter lognormal curves with parameters estimated using the method of L-moments



Source: the author.

Figure 3: Development of the probability density function of three-parameter lognormal curves with parameters estimated using the maximum likelihood method

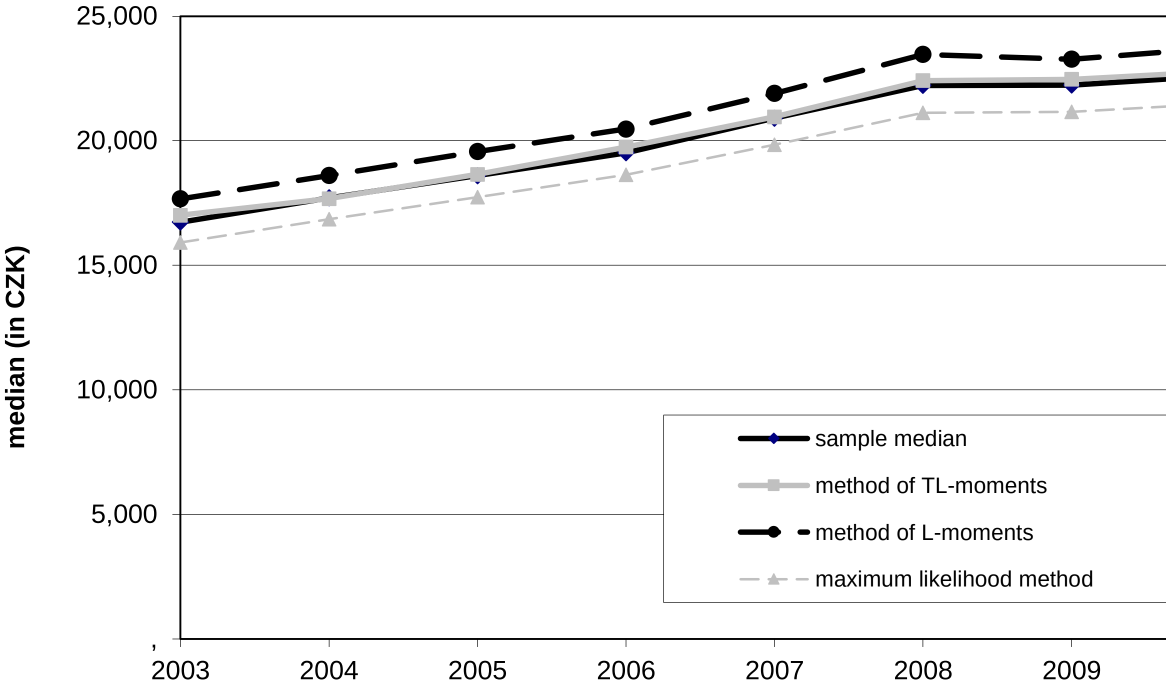


Source: the author.

Figure 4 also informs about the accuracy of the examined methods of point parameter estimation. The figure shows the development of the sample median of gross monthly wage for the total set of all employees of the Czech Republic in the period 2003–2010 as well as the development of the respective theoretical median of three-parameter lognormal model curves with the parameters estimated by the three methods. It is observable from this figure that the curve following the course of the theoretical median of a three-parameter lognormal distribution with the parameters estimated using the method of TL-moments adheres the most to the curve showing the development of the sample median. The other two curves articulating the development of the theoretical median of three-parameter lognormal curves with the parameters estimated by L-moments and by maximum likelihood methods are relatively distant from the course of the sample median of the wage distribution.

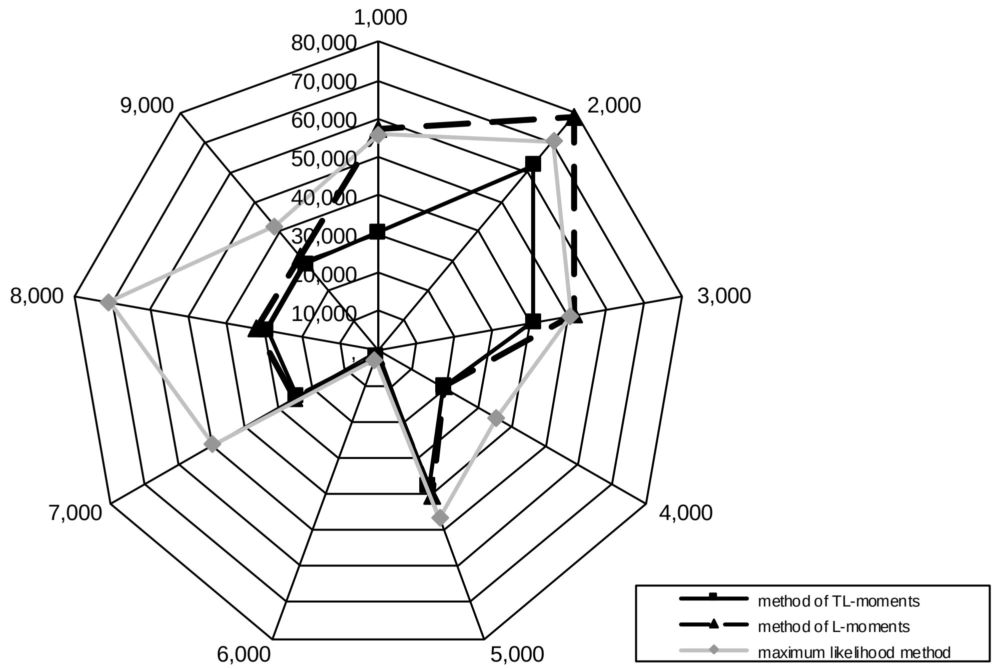
Figures 5 and 6 indicate the values of *S* criterion of 2010 wage distributions in terms of job category and five-year age intervals, respectively. High accuracy of the method of TL-moments in comparison to the other two methods of point parameter estimation is evident from the two figures, too. I is also clear from these figures that the method of conventional L-moments terminated as the second before the maximum likelihood method.

Figure 4: Development of the sample and theoretical median of three-parameter lognormal curves with parameters estimated using the three methods of parameter estimation



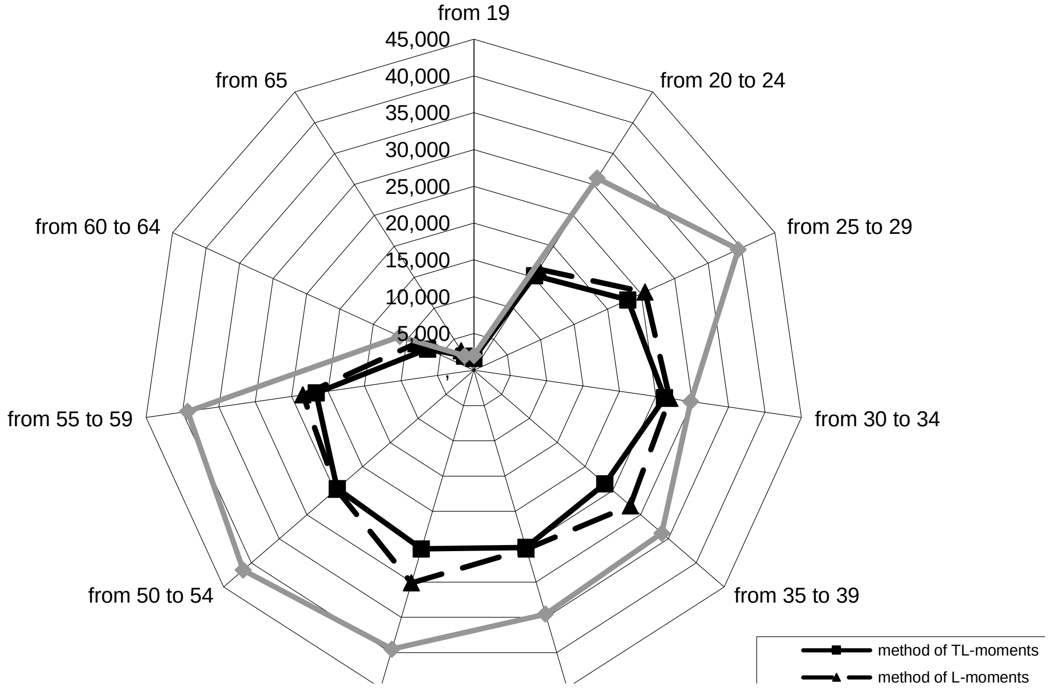
Source: the author.

Figure 5: Values of *S* criterion for three-parameter lognormal model curves with parameters estimated by methods of point parameter estimation (broken down by job category codes) – year 2010



Source: the author.

Figure 6: Values of *S* criterion for three-parameter lognormal model curves with parameters estimated by methods of point parameter estimation (broken down by age-year intervals) – year 2010



Source: the author.

5. Conclusion

A relatively new class of moment characteristics of the probability distribution has been introduced in this paper. The probability distribution characteristics of the location (level), variability, skewness and kurtosis have been constructed using L-moments and their robust extension – TL-moments method, the former (as an alternative to classical moments of probability distributions) lacking some robust features that are typical for the latter.

Sample TL-moments are linear combinations of sample order statistics assigning zero weight to a predetermined number of sample outliers. They are unbiased estimates of the corresponding TL-moments of probability distributions. The efficiency of TL-statistics depends on the choice of α – $l_1^{(0)}, l_1^{(1)}, l_1^{(2)}$, for instance, having the smallest variance (the highest efficiency) among other estimations for random samples of normal, logistic and double exponential distributions. Some theoretical and practical aspects of TL-moments need to be further researched anyway.

The accuracy of TL-moments method was compared to that of L-moments and the maximum likelihood method. Higher accuracy of the former approach in comparison to that of the latter two methods has been proved by 328 wage distribution data sets. Advantages of L-moments over the maximum likelihood method have been demonstrated by the present study as well. Two criterions for tackling wage distributions, respectively – namely the χ^2 criterion and the sum of all absolute deviations of the observed and theoretical frequencies for all intervals – have been employed. The χ^2 criterion values have always resulted in rejection of the null hypothesis about the supposed shape of the distribution due to large sample sizes typical for income and wage distribution at any significance level.

The advantages of the method of trimmed L-moments due to the conventional method of L-moments and to the maximum likelihood method and the advantages of the classical method of L-moments due to the maximum likelihood method were demonstrated when applied to the economic data (models of wage distribution). Presented methods of parameter estimation of continuous probability distributions have their practical use in all economic areas, where we work with continuous probability models whose parameters must be estimated based on sample data, for example when modeling the amount of damage in non-life insurance, etc.

Acknowledgements

The paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague (IP 400040).

References

[1] BARTOŠOVÁ, J., LONGFORD, N. 2014. A study of income stability in the Czech Republic by finite mixtures. In Prague Economic Papers, 2014, vol. 23, iss. 3, pp. 330–348.

[2] ELAMIR, E. A. H., SEHEULT, A. H. 2003. Trimmed L-moments. In Computational Statistics & Data Analysis, 2003, vol. 43, iss. 3, pp. 299–314.

[3] HOSKING, J. R. M. 1990. L-moments : Analysis and estimation of distributions using linear combinations of order statistics. In Journal of the Royal Statistical Society (Series B), 1990, vol. 52, iss. 1, pp. 105–124.

[4] JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N. 1994. Continuous univariate distributions. New York : Wiley, 1994. 756 pp. ISBN 0-471-58495-9.

[5] MALÁ, I. 2014. A multivariate mixture model for incomes of the Czech households in 2006–2010. In LÖSTER, T., PAVELKA, T. (eds.) The 8th International Days of Statistics and Economics. Slaný : Melandrium, 2014. ISBN 978-80-87990-02-5.

[6] MALÁ, I. 2013. Finite mixtures of lognormal and gamma distributions. In LÖSTER, T., PAVELKA, T. (eds.) The 7th International Days of Statistics and Economics. Slaný : Melandrium, 2013. ISBN 978-80-86175-87-4.

[7] MALEC, L., MALEC, M. 2013. Application of two-set multivariate statistical methods to the Czech Republic arrival tourism data. In LÖSTER, T., PAVELKA, T. (eds.) The 7th International Days of Statistics and Economics. Slaný : Melandrium, 2013. ISBN 978-80-86175-87-4.

[8] MAREK, L. 2013. Some aspects of average wage evolution in the Czech Republic. In LÖSTER, T., PAVELKA, T. (eds.) The 7th International Days of Statistics and Economics. Slaný : Melandrium, 2013. ISBN 978-80-86175-87-4.

[9] MAREK, L., VRABEC, M. 2013. Probability models for wage distributions. In VOJÁČKOVÁ, H. (ed.) The 31st International Conference on Mathematical Methods in Economics. Jihlava : College of Polytechnics Jihlava, 2013. ISBN 978-80-87035-76-4.

[10] PAVELKA, T., LÖSTER, T. 2013. Flexibility of the Czech labour market from a perspective of the employment protection index. In LÖSTER, T., PAVELKA, T. (eds.) The 7th International Days of Statistics and Economics. Slaný : Melandrium, 2013. ISBN 978-80-86175-87-4.

[11] PAVELKA, T., SKÁLA, M., ČADIL, J. 2014. Selected issues of the minimum wage in the Czech Republic. In E&M Economics and Management, 2014, vol. 17, iss. 4, pp. 30-45.

[12] PIVOŇKA, T., LÖSTER, T. 2014. Clustering of regions of the European Union by the labour market structure. In LÖSTER, T., PAVELKA, T. (eds.) The 8th International Days of Statistics and Economics. Slaný : Melandrium, 2014. ISBN 978-80-87990-02-5.

[13] ŠIMPACH, O., PECHROVÁ, M. 2013. Assessing the impact of standard of living on the life expectancy at birth using vector autoregressive model. In VOJÁČKOVÁ, H. (ed.) The 31st International Conference on Mathematical Methods in Economics. Jihlava : College of Polytechnics Jihlava, 2013. ISBN 978-80-87035-76-4.