MOMENT MATCHING METHOD FOR ESTIMATION OF PARAMETERS BASED ON ROBUST MOMENTS

Ivana Malá

Abstract

The moment method is frequently used to estimate parameters of probability distributions. It can also be applied to obtain initial values for a numerical procedure of maximum likelihood estimation. As sample moments, we can use classical product moments sensitive to outliers, or replace them by more robust variants (in this text L- and TL- moments). Based on these moments, we can construct equations to be solved to estimate unknown parameters. We set equal theoretical moments (as a function of unknown parameters) to the sample moments evaluated from data. We use a Monte Carlo simulation to illustrate the properties of the estimation method for the lognormal distribution (sample sizes 50, 100, 200 observations). The moment method is used for product moments, L-moments and TL- moments and compared with MLE estimators. Simultaneously, the dependence of a number of trimmed values in TL moments is of interest. The data are contaminated with 5% of outliers through a mixture of two distributions.

Key words: moment method, Monte Carlo simulation, L-moment, TL-moment

JEL Code: C14, C63

Introduction

The moment method of estimation of parameters of probability The moment matching method (MME) is a widely used method of estimation of parameters. The idea is to find values of the unknown parameters that result in a match between the theoretical (or population) and sample moments evaluated from data. The first use is attributed to Pearson, who at the end of the 19th century, applied the concept to a mixture of two normal distributions (Pearson, 1894).

Moment method estimates are consistent, but in general, the maximum likelihood method (MLE) is preferred and the moment method is used as quick information about the unknown parameter or an initial guess for a numerical search for MLE estimates.

In the text, the lognormal distribution is of interest as the frequently applied income distribution. The distribution is positively skewed, and in a sample, we often observe extreme values. These values can cause a shift in the product moments. For this reason, the more robust moment method based might be more convenient. The moment method based on classical moments is also sensitive to outliers, as in formulas the product moment, including powers of differences from the mean are used. The more robust approach needs the application of some robust moments in equations of theoretical and sample moments. In this text, we present moment method based on L-moments (Hosking, 1990) (LMME) and TL-moments (Elamir, Seheult, 2003; Mat Jan, 2016) (referred to as TLMME).

1 Method

For a random variable X with the lognormal distribution $LN(\mu;\sigma^2)$ we have $\mu = E(\ln X)$, $\sigma^2 = Var(\ln X)$ and $\ln X \sim N(\mu;\sigma^2)$. For this distribution, the closed formulas are available for almost all computations.

- For the MLE estimator we obtain

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} \ln X_i = \overline{\ln X},\ \ \hat{\sigma}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln X_i - \hat{\mu})^2}.$$

- For a classical moment method based on Pearson product moments we obtain

$$\hat{\mu}_{MME} = \ln\left(\frac{\bar{X}^2}{\sqrt{M_2'}}\right),\ \hat{\sigma}_{MME} = \sqrt{\ln\left(\frac{M_2'}{\bar{X}^2}\right)},$$

where M_2' is a second sample moment defined as

$$M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

- As a robust version of this method, we use L-moments instead of classical product moments. If we denote theoretical L-loments by $L_k, k = 1,2,...$ and by $l_k, k = 1,2,...$ their sample counterparts sample L-moments based on the sample of the size $n.$ For the moment method based on L-moments we solve (with respect to unknown distribution parameters) equations $l_k = L_k, k = 1,2,...$ instead of $M_k = \mu_k, k = 1,2,...$ where μ_k is a $k-th$ theoretical product moment and M_k its sample counterpart; for two unknown parameters, we need two equations. In (Hosking, 1990 or Bílková, 2006, Table 1) we obtain formulas for unknown parameters of three parametric lognormal distribution with a shift parameter ξ. For three parameters, three equations are needed. The third normalized moment called L-skewness $\tau_3 = L_3 / L_2$ is preferred to the third L-moment. If we use only two moments, we obtain two equations (Hosking, 1990, Bílková, 2006)

$$L_1 = \mathrm{e}^{\hat{\mu}_{LMME} + \hat{\sigma}^2_{LMME}/2} = l_1,$$

$$L_2 = \mathrm{e}^{\hat{\mu}_{LMME} + \hat{\sigma}^2_{LMME}/2} \left[2\Phi\left(\hat{\sigma}_{LMME} / \sqrt{2} \right) - 1) \right] = l_2,$$

where Φ is the standard normal cumulative distribution function. From these equations, we obtain moment estimates (based on L-moments)

$$\hat{\mu}_{LMME} = \ln(l_1) - \hat{\sigma}_{LMME} / 2, \ \hat{\sigma}_{LMME} = \sqrt{2} \ z_{0.5(1+l_2/l_1)},$$

where z is the quantile function of the standard normal distribution. The same result can be obtained in the package *lmomco* in R (Asquith, 2020). The function parln3 estimates three parametric lognormal distribution and the two parametric version is achieved by setting the shift parameter $\xi = 0$.

- The last estimator is obtained with the application of trimmed L-moments (Elamir, Seheult, 2003). The idea of this robust generalization of L-moments is similar to them, but the weight to the selected number of extreme (small or large) values is set to zero. By this process, large or small values are trimmed, and the number of such observations is a parameter to be selected. All formulas are given in (Elamir, Seheult, 2003, Mat Jan et al., 2016, Bílková, 2014). All L-moments are defined (finite) for the distributions with finite expected values, the TL moments are defined for all probability distributions without any assumption on the moments. The estimates are obtained by the numerical optimization of the function

$$(tl_1 - TL_1)^2 + (tl_2 - TL_2)^2,$$

where $TL_k, \ k = 1, 2$ are the first and the second TL-moments and $tl_k, \ k = 1, 2$ are their sample counterparts. For the numerical method, the classical moment estimate was used as the initial approximation. The estimated parameters depend on the proportion of trimmed values. In the literature, usually, only one observation from both sides (small and large values) are trimmed, this operation is sufficient for obtaining all trimmed moments finite. The dependence on the number of trimmed ordered statistics is presented in figures. We use the positively skewed lognormal distribution (and in the simulated data we add even more large values); for this reason, only data from extreme large values are trimmed (weights from 1 to 20 highest values are set to zero).

For all methods, the estimates are asymptotically normal. In Figures 1-3, the estimated ellipses as 95% confidence regions are shown. The MLE estimates are independent; the Fisher information matrix is diagonal with known diagonal elements depended on unknown parameters. It follows that the axes of the ellipses are parallel to the Cartesian system axes. The covariance matrices of all moment methods are estimated from the Monte-Carlo simulation. In the study, we compare estimates from a sample and samples contaminated with 5 % of outliers from normal distribution.

All calculations are performed in R (R CORE TEAM, 2015). The packages *lmomco* (Asquith, 2020), *lmoments* (Karvanen, 2006) and *tlmoments* (Lilienthal, 2019) were applied.

2 Simulation

To illustrate all these approaches, we perform a Monte Carlo simulation. For sample sizes of 50, 100, and 200 observations, we repeat 200 simulations for Figures 1-3 and 1,000 simulations for Figure 4. We select a lognormal distribution $LN(2.3;1.5^2)$, and we add a noise of 5% observations from the distribution $N(100;100)$ (mixture I) ($N(150;400)$, respectively, referred as mixture II). We obtain probabilities (for $X \sim LN(2.3;1.5^2)$)

$$P(X > 100) = 0.062, \quad P(X > 150) = 0.035.$$

The contamination causes a side maximum of the density in the mean values of the contaminating variable (the mode is equal to $\exp(2.3-1.5^2) = 1.05$), the distribution is highly skewed with a coefficient of skewness equal to 33.46. The mixture I has the higher value at the point 100 (selected standard deviation value 10) and mixture II at the point 150 (standard deviation 20).

We use three generated datasets: lognormal distribution, the mixture I and the mixture II. From these datasets we evaluate:

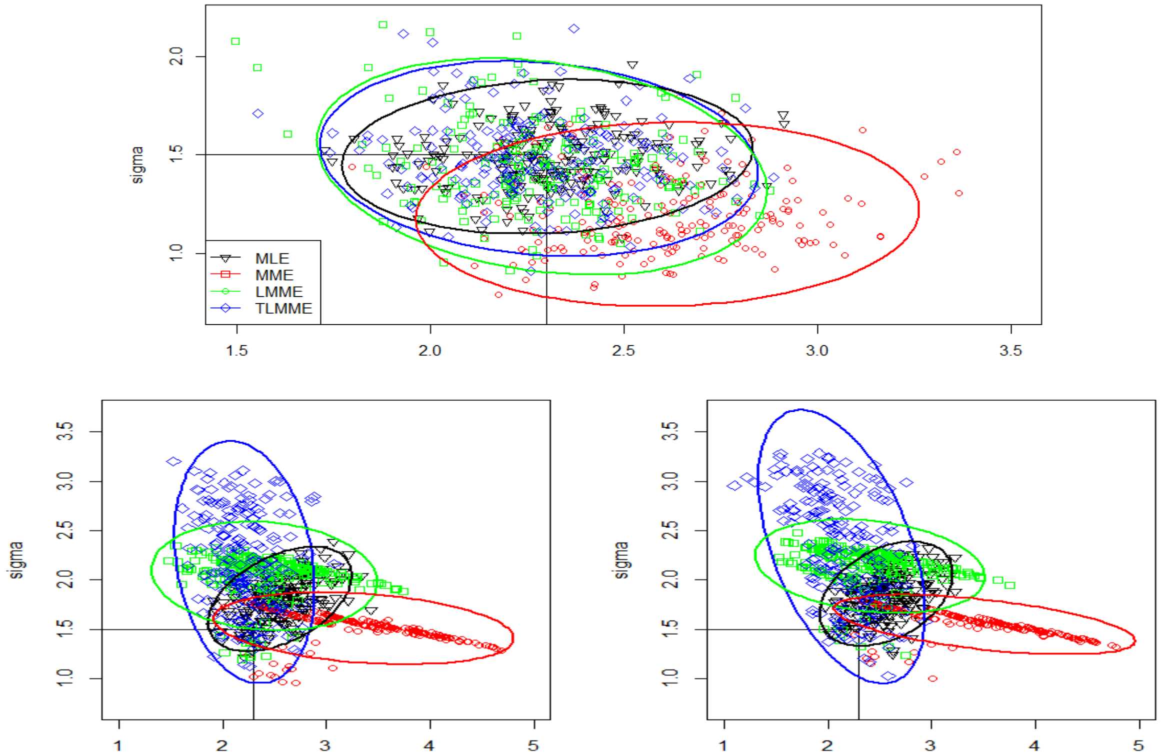
- maximum likelihood estimates (MLE, in black in Figures 1-3)
- moment method estimates based on the classical product moments (MME, in red in Figures 1-3)
- moment method estimates based on L-moments (LMME, in green in Figures 1-3)
- moment method estimates based on TL-moments (TLMME, in blue in Figures 1-3).

Moreover, the dependence on the number of trimmed values is shown for one observation from small values and 1-20 values from large values. It means 4-40% for the sample size of 50 observations, 2-20% for n = 100 and 1-10% for n = 200.

In Figures 1-3 we present 200 values of estimated parameters (μ on the horizontal axis, σ on the vertical axis). According to (Hosking, 1990; Elamir, Seheult, 2003; Bílková, 2014) the estimates are asymptotically normal. In Figures 1-3, the ellipses are estimated from the

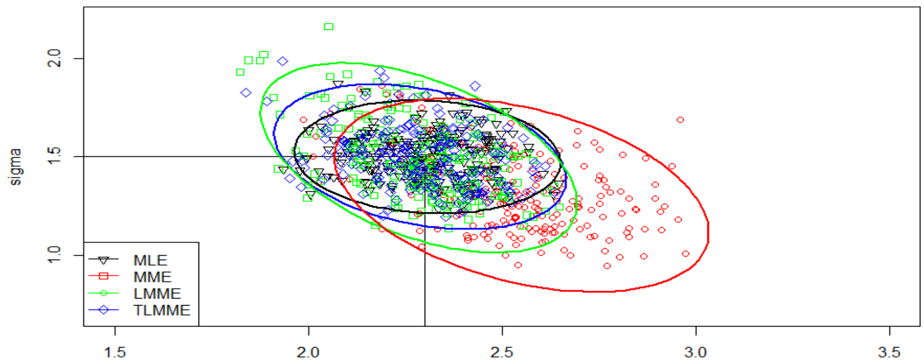
Monte Carlo sample (as a confidence region for a pair (μ,σ)). A normal approximation is hardly suitable for 50 observations. We include this size od data to show properties for small samples as the robust methods are supposed to be efficient for such samples.

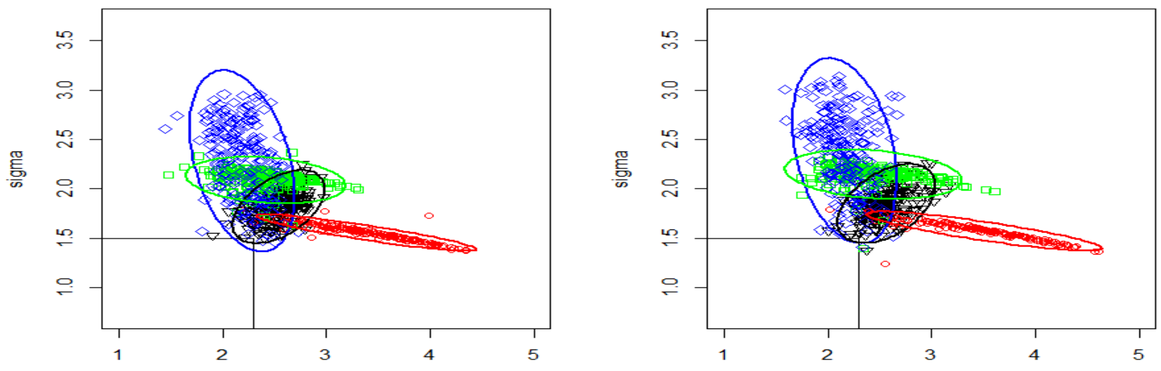
Fig. 1: Estimated parameters in the Monte Carlo study, *n*=50



Source: own calculations

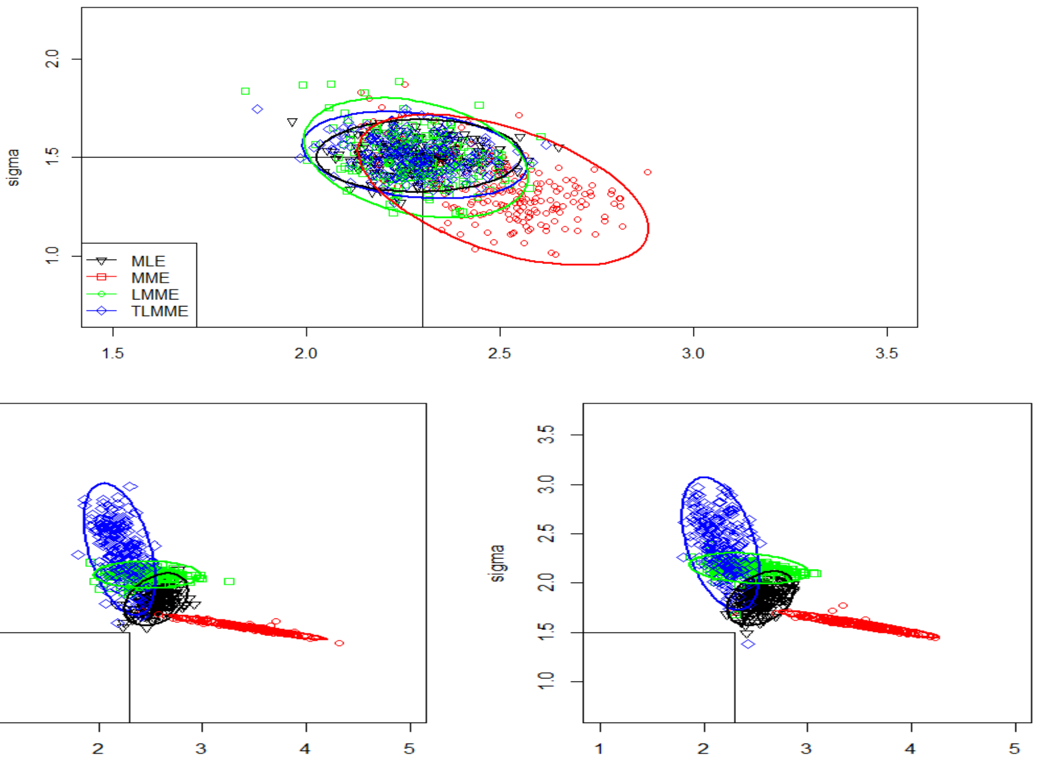
Fig. 2: Estimated parameters in the Monte Carlo study, *n*=100





Source: own calculations

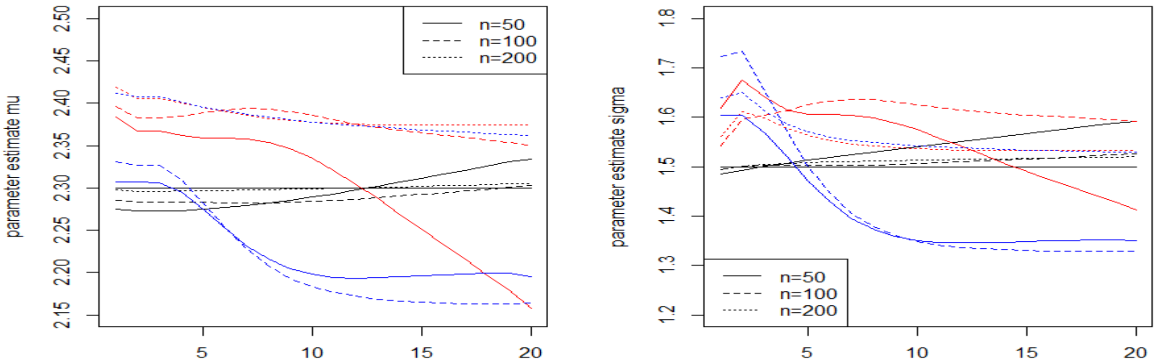
Fig. 3: Estimated parameters in the Monte Carlo study, *n*=200



Source: own calculations

In Figure 4, we present the dependence of estimates on the number of trimmed ordered values (from high values). We select from 1 to 20 values (horizontal line) and evaluated 1,000 values of estimates. In Figure 4, means of all estimates are presented, “pure” data in black, contaminated data in red (mixture I) and blue (mixture II). For 50 observations (solid line), 21 trimmed observations give us a sample of 29 observations. For this reason, the stability of estimates is relatively poor even in case of non-contaminated data. The estimates from 200 observations are precise, for the small number of trimmed values, both estimates are higher than the true value. The approaching to the real value is visible; it is faster for the parameter sigma.

Fig. 4: Estimated parameters μ and σ sigma, TL moments with trimmed values 1-20. Lognormal distribution in black, mixture I in red and mixture II in blue.



Source: own calculations

Conclusion

In the text, different possible methods of estimates of parameters are treated and compared with the use of Monte Carlo simulation. Small and moderate sample sizes from the lognormal distribution and the distribution contaminated with 5% of data from a normal distribution with the high expected value and two values of variances (but usually not the highest in the sample).

The estimates based on L-moments and TL-moments were proposed (Hasking (1990) and Elamir, Seheult, 2003) as a robust variant of a classical moment matching method. In our simulation, a maximum likelihood method performs better than other methods; the classical moment matching method being the worst. The selected distribution is highly skewed with outliers even in the pure data; in the simulation, we only add more high observations.

In the literature, usually only one observation is trimmed on both sides of the distribution. This choice is straightforward, as it allows the application for all distributions even without the finite first moment. For this reason, we can apply methods based on TL-moments for Cauchy distribution or Pareto distribution with a shape parameter less than (or equal to) 1. Moreover, the approach allows for non-symmetric trimming. In our text, we apply for our data multiple trimming from large values () and only one observation from small values.

Acknowledgment

The support of the grant no. F4/45/2020 provided by Internal grant agency of the University of Economics, Prague is gladly acknowledged.

References

1. Asquith, W. H. (2020). lmomco: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.3.6. 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. Bílková, D. (2014) Alternative Means of Statistical Data Analysis: L-Moments and TL-Moments of Probability Distributions. *Statistika*, 77-94.
4. 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.
5. Delignette-Muller M.L., Dutang, C. (2015). Fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64, 1-34.
6. Elamir, E.A.H., Seheult, A.H. (2003). Trimmed L-moments. *Comput. Statist. Data Analysis*, 43, 299-314.
7. 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.
8. Karvanen, J. (2006). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics & Data Analysis*, 51, 947-959.
9. Lilienthal, J. 2019. tlmoments: Calculate TL-Moments and Convert Them to Distribution Parameters. R package version 0.7.5. Retrieved from: https://cran.r-project.org/web/packages/TLMoments.
10. Mat Jan, N.A., Shabri, A., Ismail, S., Badyalina, B., Abadan, S.S., Yusof, N. (2016). Three-parameter lognormal distribution: Parametric estimation using l-moment and tl-moment approach. *Jurnal Teknologi (Sciences & Engineering)*, 78, 85–96.
11. Pearson, K. (1894). Contribution to the mathematical theory of evolution. *Philos Tr R Soc S-A*, 185, 71–110.
12. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/.

Contact

Ivana Malá
University of Economics, Prague
W. Churchill pl. 4, Prague 3, Czech Republic
malai@vse.cz