STATISTICS IN TRANSITION new series, Winter 2014
Vol. 15, No. 1, pp. 133–144

APPLICATION OF QUANTILE METHODS TO ESTIMATION OF CAUCHY DISTRIBUTION PARAMETERS

Dorota Pekasiewicz[1]

ABSTRACT

Quantile methods are used for estimation of population parameters when other methods such as the maximum likelihood method and the method of moments cannot be applied. In the paper the percentile method, the quantile least squares method and its two modifications are considered. The proposed methods allow estimators to be obtained with smaller bias and smaller mean squared error than estimators of the quantile least squares method. The considered methods can be applied to estimation of the Cauchy distribution parameters. The results of the simulation analysis of the estimator properties have allowed conclusions to be drawn as concerning the application of the considered methods.

Key words: quantile , percentile method, quantile least square method, Cauchy distribution.

1. Introduction

Quantile estimation methods can be used for estimating parameters of different distributions, particularly in the cases when we cannot use the maximum likelihood method and the method of moments, for example, for heavy tailed distributions. We focus on the percentile method, the quantile least squares method and its modifications.

Since in the percentile method distribution quantiles are compared with sample quantiles, its application requires the formula for the quantile function. The number of required quantiles depends on the number of distribution parameters. The orders of selected quantile s have impact on properties of estimators. Quantiles that give the estimators with small mean squared errors can be different for different types of distribution estimates. In Aitchison and Brown (1975) the orders of quantiles which should be selected in estimating lognormal distribution parameters are given and in Pekasiewicz (2012) they are computed for the Pareto distribution.

[1] University of Łódź, Department of Statistical Methods. E-mail: pekasiewicz@uni.lodz.pl.

The quantile least squares method has the advantage over the percentile method that there is no need to determine the ranks of used order statistics. However, in the case of the Cauchy distribution, the application of minimum or maximum statistics leads to very large mean squared errors of the parameter estimators because extreme statistics have infinite variances. Rejecting extreme order statistics significantly improves the properties of the estimators. Hence, we suggest the truncated quantile least squares method. In the case of the Cauchy distribution, we reject a fixed number of the largest and the smallest order statistics. Rejecting the same number of quantiles on both sides of the distribution appears to be justified in view of the symmetry of the distribution. The use of this method requires determination of the number of rejected quantiles. The second of the proposed methods does not require assumptions about the number of truncated quantiles. In this method all possible estimators are calculated by the truncated least squares and the median of them is chosen.

The properties of the Cauchy distribution parameter estimators are analysed by the Monte Carlo method. The received results allow some conclusions to be drawn regarding the choice of ranks of the order statistics in the percentile method or the number of rejected order statistics.

2. The percentile method

The percentile method (PM) allows for estimation of unknown parameters $\theta_1, \theta_2, ..., \theta_s$ of the continuous random variable X distribution with cumulative distribution function $F(\cdot, \theta_1, \theta_2, ..., \theta_s)$ by comparing theoretical quantiles and empirical quantiles (Wywiał, 2004, Castillo et al., 2004).

Let $X_1, X_2, ..., X_n$ be an i.i.d. sample with a cdf F. Let us denote by $X_{p_i;n}$ the sample quantile of order p_i, $i = 1, ..., s$. Estimators of the parameters $\theta_1, \theta_2, ..., \theta_s$ are the statistics $\hat{\theta}_1^{pm}, \hat{\theta}_2^{pm}, ..., \hat{\theta}_s^{pm}$ that are solutions of the equations:

$$\begin{cases} X_{p_1;n} = F^{-1}(p_1, \theta_1, \theta_2, ..., \theta_s), \\ X_{p_2;n} = F^{-1}(p_2, \theta_1, \theta_2, ..., \theta_s), \\ \ldots\ldots \\ X_{p_s;n} = F^{-1}(p_s, \theta_1, \theta_2, ..., \theta_s), \end{cases} \tag{1}$$

where F^{-1} is the inverse of F.

When estimating parameters θ_1, θ_2 for the random variable X with cdf $F(\cdot, \theta_1, \theta_2)$, frequently the quantiles of orders p_1, p_2 are chosen, such that $p_1 + p_2 = 1$.

For the Cauchy distribution with cdf $F(x)=\frac{1}{2}+\frac{1}{\pi}arctg\left(\frac{x-m}{\lambda}\right)$ and orders of quantile s p and $1-p$, equations (1) take the following form:

$$\begin{cases} X_{p;n} = m+\lambda \, \mathrm{tg}(\pi(p-0,5)), \\ X_{(1-p);n} = m+\lambda \, \mathrm{tg}(\pi(0,5-p)), \end{cases} \tag{2}$$

and the estimators are defined by the formulas:

$$\hat{m}^{pm} = \frac{X_{p;n} + X_{(1-p);n}}{2}, \tag{3}$$

$$\hat{\lambda}^{pm} = \frac{X_{(1-p);n} - X_{p;n}}{2\mathrm{ctg}(\pi p)}. \tag{4}$$

3. The Quantile least squares method and its modifications

The quantile least squares method (QLSM) estimates the unknown parameters $\theta_1, \theta_2, ..., \theta_s$ of random variable *X* with cdf *F* by minimizing the sum of squares of the differences between theoretical and empirical quantile s (Gilchrist, 2000; Castillo et al., 2004). Then, the function for which we calculate the global minimum has the following form:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1}^{n}\left(X_{i/n;n} - Q_{i/n}\right)^2, \tag{5}$$

where $X_{i/n;n}$ is the sample quantile of order $p_i = \frac{i}{n}$ from the i.i.d. sample $X_1, X_2, ..., X_n$ and $Q_{i/n} = F^{-1}\left(\frac{i}{n}, \theta_1, ..., \theta_n\right)$.

The estimators of parameters $\theta_1, \theta_2, ..., \theta_s$ obtained by QLSM are denoted by $\hat{\theta}_1^{qls}, \hat{\theta}_2^{qls}, ..., \hat{\theta}_s^{qls}$.

Using all available quantile orders can, however, result in unsatisfactory estimate properties or in some cases cannot be feasible. For the Cauchy distribution extreme statistics have infinite variance, which means that the mean squared errors of estimators based on them are very large. Therefore, the minimum and maximum statistics must be rejected for estimation of the Cauchy distribution parameters.

The first suggested modification of the quantile least squares method is rejecting a fixed number of quantile s, which we call the truncated quantile least squares method (TQLMS). In this case the estimators of distribution parameters $\theta_1, \theta_2, ..., \theta_s$ of the random variable *X* with distribution function $F(\cdot, \theta_1, \theta_2, ..., \theta_s)$

are statistics $\hat{\theta}_1^{tqls}, \hat{\theta}_2^{tqls}, ..., \hat{\theta}_s^{tqls}$, for which the following expression reaches a global minimum:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i \in I_n} \left(X_{p_i;n} - Q_{p_i} \right)^2 , \tag{6}$$

where $p_i = \frac{i}{n}$ and I_n is the subset of $\{1, 2, ..., n\}$.

For symmetric or close to symmetric distributions we suggest skipping k quantiles, where k is the even number, that is $\frac{k}{2}$ the smallest and $\frac{k}{2}$ the largest quantiles. Then, the function (6) takes the form:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1+k/2}^{n-k/2} \left(X_{p_i;n} - Q_{p_i} \right)^2 . \tag{7}$$

For asymmetric distributions with right or left heavy tail we suggest skipping k the largest or the smallest order statistics, respectively. Then, one of the functions expressed by the formula:

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=1}^{n-k} \left(X_{p_i;n} - Q_{p_i} \right)^2 \tag{8}$$

or

$$G(\theta_1, \theta_2, ..., \theta_s) = \sum_{i=k}^{n} \left(X_{p_i;n} - Q_{p_i} \right)^2 , \tag{9}$$

is minimized, respectively for right and left heavy tailed distribution.

The second modification of the quantile least squares method is the median-quantile least squares method (MQLSM). The estimators of $\theta_1, \theta_2, ..., \theta_s$ parameters of the random variable X from the random sample $X_1, X_2, ..., X_n$ are statistics $\hat{\theta}_1^{mq}, \hat{\theta}_2^{mq}, ..., \hat{\theta}_s^{mq}$ of the following form:

$$\hat{\theta}_i^{mq} = Me\left(\hat{\theta}_{i;k=2}^{tqls}, \hat{\theta}_{i;k=4}^{tqls}, ..., \hat{\theta}_{i;k=r}^{tqls} \right) \qquad \text{for } i = 1, ..., s , \tag{10}$$

where $\hat{\theta}_{i;k=2}^{tqls}, \hat{\theta}_{i;k=4}^{tqls}, ..., \hat{\theta}_{i;k=r}^{tqls}$ are estimators of the parameter θ_i obtained through the truncated quantile least squares method with k quantiles left out and $r = n - 2$, when n is even and $r = n - 3$, when n is odd.

The proposed modifications can be used to estimate the Cauchy distribution parameters.

The application of the truncated quantile least squares method for the Cauchy distribution is related to the minimization of the function:

$$G(m,\lambda) = \sum_{i=1+k/2}^{n-k/2} \left(X_{p_i;n} - m + \lambda \operatorname{ctg}(\pi p_i) \right)^2 . \qquad (11)$$

Therefore, it requires solving the system of equations:

$$\begin{cases} \sum_{i=1+k/2}^{n-k/2} \left(X_{p_i;n} - m + \lambda \operatorname{ctg}(\pi p_i) \right) = 0, \\ \sum_{i=1+k/2}^{n-k/2} \left(X_{p_i;n} - m + \lambda \operatorname{ctg}(\pi p_i) \right) \operatorname{ctg}(\pi p_i) = 0. \end{cases} \qquad (12)$$

The estimators for parameters m and λ, received by the truncated quantile least squares method (TQLSM) are defined by the formulas:

$$\hat{m}^{tqls} = \frac{\sum_{i=1+k/2}^{n-k/2} X_{p_i;n}}{n-k} - \frac{\sum_{i=1+k/2}^{n-k/2} X_{p_i;n} \sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}(\pi p_i) - (n-k) \sum_{i=1+k/2}^{n-k/2} X_{p_i;n} \operatorname{ctg}(\pi p_i)}{(n-k)\left(\sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}(\pi p_i) \right)^2 - (n-k)^2 \sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}^2(\pi p_i)} \sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}(\pi p_i), \qquad (13)$$

$$\hat{\lambda}^{tqls} = \frac{\sum_{i=1+k/2}^{n-k/2} X_{p_i;n} \sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}(\pi p_i) - (n-k) \sum_{i=1+k/2}^{n-k/2} X_{p_i;n} \operatorname{ctg}(\pi p_i)}{(n-k) \sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}^2(\pi p_i) - \left(\sum_{i=1+k/2}^{n-k/2} \operatorname{ctg}(\pi p_i) \right)^2}, \qquad (14)$$

where k is a fixed even number, $X_{p_i;n}$ is the quantile from the i.i.d. sample $X_1, X_2, ..., X_n$ and $p_i = \frac{i}{n}$ for $i = 1 + \frac{k}{2}, ..., n - \frac{k}{2}$.

The estimators of the Cauchy distribution parameters received by median-quantile least squares method (MQLSM) have the following form:

$$\hat{m}^{mq} = Me(\hat{m}_{k=2}^{tqls}, \hat{m}_{k=4}^{tqls}, ..., \hat{m}_{k=r}^{tqls}), \qquad (15)$$

$$\hat{\lambda}^{mq} = Me(\hat{\lambda}_{k=2}^{tqls}, \hat{\lambda}_{k=4}^{tqls}, ..., \hat{\lambda}_{k=r}^{tqls}), \qquad (16)$$

where $r = n - 2$, when n is even and $r = n - 3$, when n is odd.

4. Simulation analysis of Cauchy distribution parameter estimators

The properties of the Cauchy distribution parameter estimators were studied using the Monte Carlo methods. The following methods were considered:

- the percentile methods with different orders of selected quantiles,
- the truncated quantile least squares methods with different number of rejected quantile s from tails of the distribution,
- the median-quantile least squares method.

In each case the bias and the mean squared error were estimated over 20000 repetitions. For the percentile method the dependences of the estimator bias and the mean squared error on the quantile order are presented in Figures 1 and 2. The size of the sample *n* was set to 100.

From the obtained results it can be concluded that, for the considered distribution, if *p* increases both the bias and the mean squared error of the estimators decrease, but only to a certain point. When this point is exceeded the precision can deteriorate. For $p \approx 0.45$ the estimator $\hat{m}^{pm}$ of parameter m has the best of properties, and for $p \approx 0.25$ the estimator $\hat{\lambda}^{pm}$ has the smallest mean squared error.

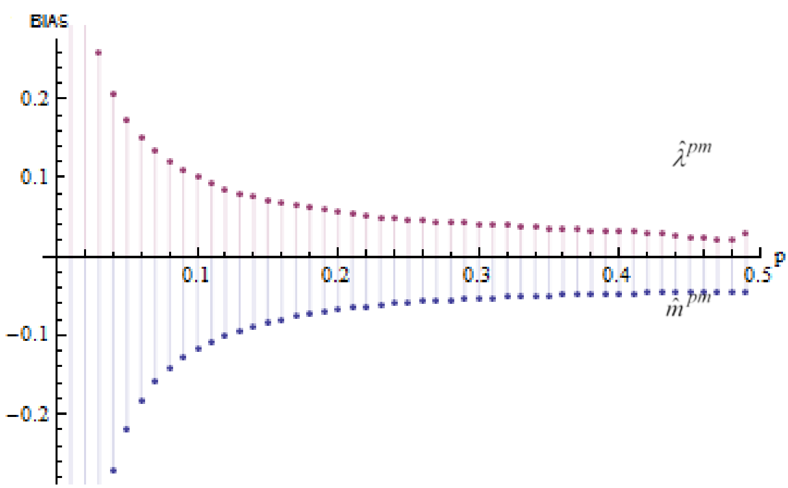

Figure 1. Dependence of the bias of *Ca*(0,3) parameter estimators obtained by the percentile methods on the quantile order

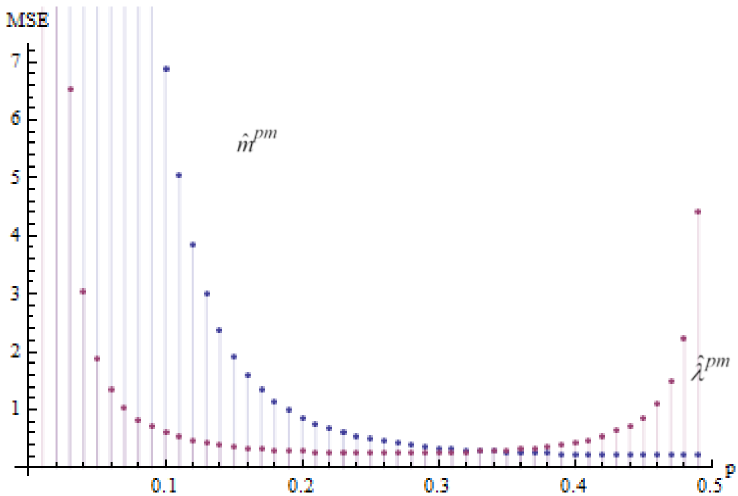

Figure 2. Dependence of the mean squared error of *Ca*(0,3) parameter estimators obtained by the percentile methods on the quantile order

The estimated values of the bias and the mean squared error for selected orders of quantile , for chosen parameters of the Cauchy distribution are shown in Table 1. The study included sample sizes n = 60 and n = 100. Pekasiewicz (2012) gives results for the Cauchy distribution with other parameters.

Table 1. The estimated bias and the mean square errors for the Cauchy distribution parameter estimators obtained by the percentile method

Distribution random variable	n	p	$\hat{BIAS}(\hat{m}^{pm})$	$\hat{BIAS}(\hat{\lambda}^{pm})$	$\hat{MSE}(\hat{m}^{pm})$	$\hat{MSE}(\hat{\lambda}^{pm})$
$Ca(0,3)$	60	0.05	- 4.4690	0.7057	241.5300	5.7885
		0.10	- 0.9033	0.2788	14.1888	1.2504
		0.15	- 0.4011	0.1824	3.7038	0.7143
		0.20	- 0.2385	0.1400	1.5905	0.5356
		0.25	- 0.1614	0.1155	0.8972	0.4656
	100	0,05	- 2.4252	0.3735	81.5948	1.8908
		0,10	- 0.5821	0.1630	6.6861	0.6084
		0,20	- 0.1521	0.0817	0.8736	0.2842
		0,30	- 0.0818	0.0567	0.3424	0.2618
		0,40	- 0.0575	0.0500	0.2312	0.4381
		0,45	- 0.0525	0.0397	0.2233	0.8682

Table 1. The estimated bias and the mean square errors for the Cauchy distribution parameter estimators obtained by the percentile method (cont.)

Distribution random variable	n	p	$\hat{BIAS}(\hat{m}^{pm})$	$\hat{BIAS}(\hat{\lambda}^{pm})$	$\hat{MSE}(\hat{m}^{pm})$	$\hat{MSE}(\hat{\lambda}^{pm})$
$Ca(3,2)$	60	0,05	- 3.1883	0.4949	143.4070	3.4363
		0,10	- 0.6597	0.2021	6.8335	0.6129
		0,15	- 0.2971	0.1285	1.7430	0.3321
		0,20	- 0.1729	0.0958	0.7408	0.2427
		0,25	- 0.1153	0.0759	0.4051	0.2056
	100	0,05	- 1.6606	0.2460	36.0447	0.8338
		0,10	- 0.3889	0.1115	3.1781	0.2740
		0,20	- 0.1032	0.0511	0.3826	0.1239
		0,30	- 0.0552	0.0372	0.1515	0.1136
		0,40	- 0.0402	0.0317	0.1022	0.1907
		0,45	- 0.0369	0.0361	0.0973	0.3835

The use of the truncated quantile least squares method for large k allowed the bias and the mean squared error to be significantly reduced, although rejecting too many quantiles gives poorer results. Dependences of the bias and the mean squared error on k are shown in Figures 3 and 4, respectively. The results for the selected parameters are given in Table 2.

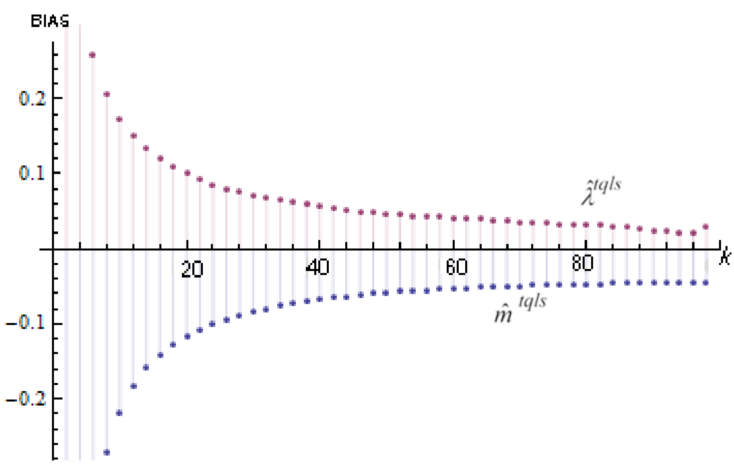

Figure 3. Dependence of the bias of $Ca(0,3)$ parameter estimators obtained by TQLSM on the number of rejected quantiles

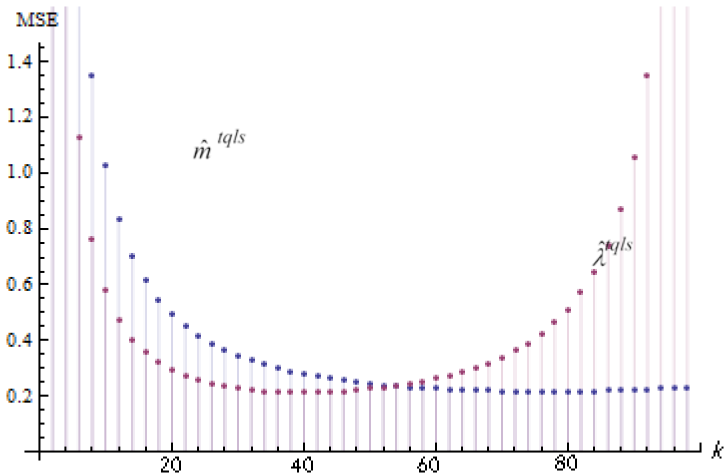

Figure 4. Dependence of the mean squared error of $Ca(0,3)$ parameter estimators obtained by TQLSM on the number of rejected quantiles

Table 2. The estimated bias and the mean squared errors for the Cauchy distribution parameter estimators obtained by the truncated quantile least squares method

Distribution random variable	n	k	$B\hat{IAS}(\hat{m}^{tqls})$	$B\hat{IAS}(\hat{\lambda}^{tqls})$	$M\hat{S}E(\hat{m}^{tqls})$	$M\hat{S}E(\hat{\lambda}^{tqls})$
$Ca(0,3)$	60	2	- 1.1615	0.4884	14.5741	9.2934
		10	- 0.3844	0.2749	2.0822	1.2637
		20	- 0.2024	0.1687	0.9150	0.5702
		30	- 0.1460	0.1296	0.6253	0.4364
		40	- 0.1182	0.1068	0.4968	0.3990
		50	- 0.1016	0.0943	0.4303	0.4137
	100	2	- 1.1474	0.4970	15.3895	12.9429
		10	- 0,2189	0.1736	1.0284	0.5808
		30	- 0.0840	0.0716	0.3446	0.2286
		50	- 0.0585	0.0471	0.2446	0.2264
		70	- 0.0492	0.0362	0.2184	0.3392
		90	- 0.0451	0.0240	0.2222	1.0580

Table 2. The estimated bias and the mean squared errors for the Cauchy distribution parameter estimators obtained by the truncated quantile least squares method (cont.)

Distribution random variable	n	k	$B\hat{IAS}(\hat{m}^{tqls})$	$B\hat{IAS}(\hat{\lambda}^{tqls})$	$M\hat{S}E(\hat{m}^{tqls})$	$M\hat{S}E(\hat{\lambda}^{tqls})$
$Ca(3, 2)$	60	2	- 0.8459	0.3919	3.3330	1.3605
		10	- 0.3745	0.1965	0.7766	0.4571
		20	- 0.2330	0.0424	0.4033	0.1809
		30	- 0.2575	0.0192	0.3496	0.1500
		40	- 0.2631	- 0.0129	0.3209	0.1615
		50	- 0.2553	-0.0232	0.2996	0.1710
	100	2	- 0.7546	0.3381	7.3597	4.4003
		10	- 0.1461	0.1177	0.4638	0.2525
		30	- 0.0562	0.0522	0.1570	0.1014
		50	- 0.0389	0.0405	0.1111	0.1014
		70	- 0.0317	0.0363	0.0979	0.1532
		90	- 0.0292	0.0332	0.0996	0.4758

The median-quantile least squares method was also used for estimation of the Cauchy distribution parameters. The properties of the estimators are presented in Table 3 for selected parameters. This method offers a more convenient estimation algorithm in comparison to the percentile method and the truncated quantile least squares method because it does not require additional assumptions about the quantile orders or the number of rejected quantiles.

Table 3. The estimated bias and the mean squared errors for the Cauchy distribution parameter estimators obtained by the median-quantile least squares method

Distribution random variable	n	$B\hat{IAS}(\hat{m}^{mq})$	$B\hat{IAS}(\hat{\lambda}^{mq})$	$M\hat{S}E(\hat{m}^{mq})$	$M\hat{S}E(\hat{\lambda}^{mq})$
$Ca(0, 3)$	60	- 0.1136	0.0846	0.4383	0.3861
	100	- 0.0659	0.0455	0.2468	0.2181
$Ca(3, 2)$	60	- 0.0780	0.0535	0.2005	0.1685
	100	- 0.0462	0.0340	0.1119	0.0983

In Table 4 the results of simulation analysis about the considered methods: the percentile method (PM) for selected p, the truncated quantile least squares method (TQLSM) and the median-quantile least squares method (MQLSM) for selected k are presented. The estimators obtained from the percentile methods have the smallest mean squared errors for the value of p given in the table. In the case of the truncated quantile least squares estimators for k=78 the estimator $\hat{m}$ has the smallest mean squared errors and the estimator $\hat{\lambda}$ for k =40. The number of k =52 was chosen as an intermediate value for comparison of the properties of the estimators.

Table 4. The estimated mean squared errors for the Cauchy distribution parameter estimators obtained by the considered quantile methods for n = 100

Distribution random variable	Estimators	Estimation method					
		PM	*TQLSM*				MQLSM
			k=2	k=40	k=52	k=78	
Ca(0,3)	$\hat{m}$	0.2227 (p= 0.44)	15.3895	0.2787	0.2399	0.2167	0.2468
	$\hat{\lambda}$	0.2536 (p= 0.27)	12.9429	0.2135	0.2318	0.4228	0.2181
Ca(3,2)	$\hat{m}$	0.0973 (p= 0,45)	7.3597	0.1273	0.1089	0.0972	0.1118
	$\hat{\lambda}$	0.1098 (p= 0,27)	4.4003	0.0957	0.1014	0.1910	0.0983

The results of the analysis indicate that rejecting a number of the smallest and the largest quantiles significantly improved properties of the Cauchy distribution parameter estimators as compared to the quantile least squares method, which rejects only extreme statistics. In both methods estimation of each parameter requires choosing different orders and different number of rejected quantiles, which ensures the smallest mean squared errors.

The application of the median-quantile least squares method gives results which are similar to the truncated quantile least squares method on condition that the appropriate quantile order is chosen.

5. Conclusions

Quantile methods are used for estimation of the Cauchy distribution parameters because the method of moments and the maximum likelihood method cannot be used. Practical conclusions as to their application result from the simulation analysis of the estimator properties. The appropriate value of the quantile orders in the percentile method and the number of rejected quantiles in the truncated quantile least squares method lead to estimators with small bias and mean squared error.

In the case of the Cauchy distribution, which is a heavy tailed distribution, rejecting a fixed number of the smallest and the largest quantiles significantly improves the properties of the parameter estimators. In order to minimize the mean squared errors of estimators, it is possible to use different number of rejected quantiles for each estimator.

The second suggested modification of the quantile least squares method is more convenient in applications, as it does not require any additional assumptions and allows estimators with good properties to be obtained.

Both methods can be applied to estimation of the Cauchy distribution parameters. The application of these methods to estimation of other distribution parameters requires simulation analysis of the quantile orders in the percentile methods and of the number of truncated quantiles in the truncated quantile least squares method.

REFERENCES

AITCHISON, J., BROWN, J. A. C., (1975). The lognormal distribution. Cambridge University Press, Cambridge.

CASTILLO, E., HADI, A. S., BALAKRISHNAN, N., SARABIA, J. M., (2004). Extreme value and related models with application in engineering and science. Wiley Interscience, A. John Wiley & Sons, Inc. New Jersey.

GILCHRIST, W. G., (2000). Statistical modelling with quantile functions. Chapman & Hall/CRT, Boca Raton.

PEKASIEWICZ, D., (2012). The use of simulation methods to study the properties of estimators obtained by quantile method, [in:] collective work edited by Z. E. Zielinski: „The role of information technology in economic and social sciences. Innovations and interdisciplinary implications”. Publishing House of Higher School of Commerce, 236–244.

WYWIAŁ, J., (2004). Introduction to statistical inference. Publishing House of University of Economics in Katowice, Katowice.