Estimation of the Extreme Value Type I Distribution by the Method of LQ-Moments

Ani Shabri and Abdul Aziz Jemain

Department of Mathematics, Faculty of Science, University Technology Malaysia, 81310 Skudai, Johor, Malaysia
Faculty of Science and Technology, School of Sciences Mathematic, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract: Problem statement: The study evaluated the effectiveness of the various quantile estimators of the LQ-moments method for estimating parameters of the Extreme Value Type 1 (EV1) distribution.

Approach: The performances of the LQ-moments were analyzed and compared against a widely used method of L-moments by using simulated samples of both EV1 and generalized Lambda distribution, focusing on small and moderate sample sizes.

Results: The analysis results showed that LQMOM method was more efficient in many cases especially for the upper tails of the distribution and for various sample sizes.

Conclusion: This study demonstrated that conventional LMOM was not optimal for the estimation of the EV1 distribution.

Key words: The Weighted kernel quantile, upper tail, LQ-moments, L-moments, quick estimator

INTRODUCTION

The Extreme Value Type I (EV1) distribution is widely used in various fields including hydrology for modeling extreme events\cite{5,10,14}. Despite its extensive use, however, there is generally no accepted method of estimating its parameters. Its successful application depends, doubtless, upon the accuracy with which its parameters can be estimated. Thus, the problem is one of selecting an appropriate method for estimating the EV1 distribution parameters.

The methods of PWM, ordinary product Moment (MOM) and Maximum Likelihood (ML) estimation are commonly used to estimate the parameter of the EV1 distribution. The method of ML is known to be asymptotically unbiased and optimal for the EV1 distribution\cite{13}. However, there is no guarantee that the ML-method is the best in small samples. Gumbel\cite{5} argued that the method of Maximum Likelihood Estimation (MLE) was very complicated and required numerical work and favored the Method Of Moments (MOM). Landwehr\cite{10} used the method of Probability Weighted Moments (PWM) and the related L-Moments (LMOM). They found that the method, in general, compared with the MLE and MOM methods. Raynal and Salas\cite{14} analyzed six different methods of parameter estimation and preferred PWM for large samples. Phien\cite{12} compared the MOM, ML, ME (maximum entropy) and PWM estimators for the EV1 distribution. PWM estimators were found to be best in terms of mean square error.

Mudholkar and Hutson\cite{11} extended L-moments to new moment like entities called LQ-moments to estimate the GEV distribution parameters. The LQ-moments are constructed by using functional defining the quick estimators, where the parameters of quick estimator take the values $p = 0, \alpha = 1$ for the median, $p = 1/4, \alpha = 1/4$ for the trimean and $p = 0.3, \alpha = 1/3$ for the Gastwirth, in places of expectations in L-moments. Ani and Jemain\cite{1-3} proposed the LQMOM based on the Weighted Kernel Quantile (WKQ) estimator in which the quick estimators parameters $\alpha$ and $p$ are not restricted, such as the median, trimean or Gastwirth but we explore an extended class of LQMOM with consideration combinations of $p$ and $\alpha$ values in the range 0 and 0.5.

The objective of this study is to analyze and compare statistically various quantile estimators of LQ-moments to estimate the parameters of the EV1 distributions. We considered five different quantile estimators namely the LIQ estimator, the L-quantile estimators, two of the weighted L-quantile estimators and the HDL estimator for estimating the sample LQ-moments. The performances of the LQMOM based on these quantile estimators were compared with the
method of the L MOM for various sample sizes using simulated samples of both EV1 and generalized Lambda distribution, focusing on small and moderate sample sizes.

**MATERIALS AND METHODS**

**Definition and properties of LQ-moments estimators:** Let \( X_1, X_2, \ldots, X_n \) be a random sample from a continuous distribution function \( F(.) \) with quantile function \( Q(u) = F^{-1}(u) \) and let \( X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n} \) denote the corresponding order statistics. Then the \( r \)th LQ-moments \( \xi_r \) is given by:[2]

\[
\xi_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \tau_{r, a}(X_{r-k:n}), \quad r = 1, 2, \ldots
\]  

where, \( 0 \leq a \leq 1/2, \quad 0 \leq p \leq 1/2: \)

\[
\tau_{r, a}(X_{r-k:n}) = pQ_{X_{r-k:n}}(\alpha) + (1 - 2p)Q_{X_{r-k:n}}(1 / 2) + pQ_{X_{r-k:n}}(1 - \alpha)
\]

\[
+ (1 - 2p)Q[B_{r-k:n}^{-1}(1 / 2)] + pQ[B_{r-k:n}^{-1}(1 - \alpha)]
\]  

Is the quick estimator of location and \( B_{r-k:n}^{-1}(\alpha) \) is the quantile of a beta random variable with parameter \( r-k \) and \( k+1 \) and \( Q(.) \) denotes the quantile estimator. The first four LQ-moments of the random variable \( X \) are defined as:

\[
\xi_1 = \tau_{r, a}(X_{1:n})
\]  

\[
\xi_2 = \frac{1}{4}[\tau_{r, a}(X_{2:n}) - \tau_{r, a}(X_{1:n})]
\]  

\[
\xi_3 = \frac{1}{4}[\tau_{r, a}(X_{3:n}) - 2\tau_{r, a}(X_{2:n}) + \tau_{r, a}(X_{1:n})]
\]  

\[
\xi_4 = \frac{1}{4}[\tau_{r, a}(X_{4:n}) - 3\tau_{r, a}(X_{3:n}) + 3\tau_{r, a}(X_{2:n}) - \tau_{r, a}(X_{1:n})]
\]

**Quantile estimators:** The sample quantiles estimators of the values of the population quantile \( Q(.) \), are used widely in a variety of applications such as a Q-Q plots and a box plot in the exploratory data analysis, non-parametric estimators involving statistics such as the quartiles and their ranges, to theoretical topics such as density function estimation.

Let \( X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n} \) be the corresponding order statistics. The population quantiles estimator of a distribution is defined as:

\[
Q(q) = F^{-1}(q) = \inf \{ x : F(x) \geq q \}, \quad 0 < q < 1
\]

where, \( F(x) \) is the distribution function.[3] The qth population quantile of \( F \) denoted by \( \xi_q \) and defined \( \xi_q = Q(q) \).

Sample quantiles have been studied in statistical literature.[4] The qth sample quantile is defined by:

\[
SQ_q = X_{[nq]+1}
\]

where, \([nq]\) denotes the integral part of \( nq \). The sample quantiles experience a substantial lack of efficiency, caused by the variability of individual order statistics. Mudholkar and Hutson[11] used the other quantile function estimator namely the Linear Interpolation Quantiles (LIQ) for constructed LQ-moments. The LIQ quantile is used commonly in statistical packages such as MINITAB, SAS, IMSL and S-PLUS[11].

A popular class of L-quantile estimator for improving the efficiency of sample quantiles uses equal weight to average over the order statistics and has been widely applied to reduce this variability.[15]. In recent years, researchers have studied weighted L-quantile estimator which use unequal weights for data points, to obtain better performance of estimators[7,8].

Huang and Brill[7] applied the level crossing empirical distribution function to propose a class of level crossing kernel quantile estimators. The theoretical and simulation results show that those estimators improve the efficiencies relative to the corresponding regular kernel quantile estimators. But selection of kernel or bandwidth of the kernel estimators has always been a sensitive problem[8].

Harrell and Davis[6] proposed an L-quantile estimator of \( \xi_q \), namely, HD quantile estimator which not only gives better efficiencies but also avoids the problems of selection of kernel or bandwidth. Sheather and Marron[15] showed the HD performs as well as other L-quantile estimators in large sample. Huang[8] use a level crossing empirical distribution function to propose a new estimator HDL which is a weighted version of HD. The theoretical and computational results show that the new estimator HDL is more efficient than the HD quantile in many cases, especially for the tails of the distributions and small sample sizes.

In this study, five different quantile estimators namely the LIQ estimator, the L-quantile estimators, two of the weighted L-quantile estimators and the HDL estimator were considered for estimating the sample LQ-moments. In the following, we discuss each of these estimators.
**Linear interpolation quantiles:** The Linear Interpolation Quantiles (LIQ)\(^9\) is given by:

\[
\text{LIQ}_q = (1 - \varepsilon)X_{\lfloor n'q \rfloor} + \varepsilon X_{\lfloor n'q \rfloor + 1}
\]

where, \(\varepsilon = n'q - \lfloor n'q \rfloor\), \(n' = n + 1\) and \(\lfloor nq \rfloor\) denotes the integral part of \(nq\).

**L-quantile estimators:** A popular class of L-quantile estimators is called Kernel Quantile (KQ) estimators of Sheather and Marron\(^{15}\) is given by:

\[
(\frac{i}{n} - q) \sum_j \int_{(t-q)h} K_h(t)dt X_{\lfloor jn \rfloor}
\]

where, \(K_h(t) = \frac{1}{\sqrt{2\pi}h}e^{-\frac{t^2}{2h^2}}\) is the Gaussian Kernel, \(h = \sqrt{\frac{q(1-q)}{n}}\) is an optimal bandwidth given in Corollary 1 of Sheather and Marron\(^ {15}\).

**Weighted L estimators:** Huang and Brill\(^7\) introduced the weighted L estimator called Weighted Kernel Quantile (WKQ). The weighted L estimators is given by:

\[
\text{WKQ}_q = \sum_j \left(\frac{1}{n} - q\right) \sum_j \int_{(t-q)h} K_h(t)dt X_{\lfloor jn \rfloor}
\]

where, \(K_h(t) = \frac{1}{\sqrt{2\pi}h}e^{-\frac{t^2}{2h^2}}\) is the Gaussian Kernel, \(h = \sqrt{\frac{q(1-q)}{n(n-1)}}\) is an optimal bandwidth given in (13).

**EV1 distribution:** The extreme value distribution type I (EV1) was introduced by Gumbel\(^5\) and is commonly known as Gumbel’s distribution. It one of the most widely used probability distribution functions for extreme value in hydrologic and meteorological studies for prediction of flood peaks, maximum rainfalls, maximum wind speed. The Cumulative Distribution Function (CDF) of EV1 distribution is:

\[
F(x) = \exp\left[\exp\left(-\frac{x - \mu}{\sigma}\right)\right], \quad -\infty < x < \infty
\]

The approximation form of WKQ\(_q\) are considered in this study are the following:

\[
\text{WKQ}_1 = \sum_j \left(\frac{1}{n} - q\right) \sum_j \int_{(t-q)h} K_h(t)dt X_{\lfloor jn \rfloor}
\]

\[
\text{WKQ}_2 = \sum_j \left(\frac{1}{n} - q\right) \sum_j \int_{(t-q)h} K_h(t)dt X_{\lfloor jn \rfloor}
\]

**New level crossing HDL quantile estimators:** The new level crossing qth HDL quantile estimators called HDL quantiles is given by:

\[
\text{HDL}_q = \sum_j \left(\frac{1}{n} - q\right) \sum_j \int_{(t-q)h} K_h(t)dt X_{\lfloor jn \rfloor}
\]

where:

\[
\rho_{n,i} = \sum_j w_{j,n}
\]

and \(w_{j,n}\) is given in (13).

\[
\xi_1 = \mu + \sigma \tau_{\rho_{n,i}}(X_{\lfloor i \rfloor})
\]
\[ \xi_2 = \frac{1}{2} \sigma [\tau_{\alpha, p}(X_{2,2}) - \tau_{\alpha, p}(X_{1,2})] \]  

(20)

The LQMOM estimators \( \mu \) and \( \sigma \) of the parameters are the solution of (19) and (20). \( \hat{\sigma} \) and \( \hat{\mu} \) can be estimated successively from Eq. 20 and 19 as:

\[ \hat{\sigma} = \frac{2 \xi_2}{\tau_{\alpha, p}(X_{2,2}) - \tau_{\alpha, p}(X_{1,2})} \]  

(21)

\[ \hat{\mu} = \hat{\xi}_1 - \hat{\sigma} \gamma \]  

(22)

Method of L-moments (LMOM): The LMOM estimators for the EV1 distributions are given by:

\[ \hat{\sigma} = \frac{L_2}{\log 2} \]  

(23)

\[ \hat{\mu} = b_1 \gamma - \hat{\sigma} \gamma \]  

(24)

\[ L_2 = 2b_1 - b_6 \]

Where:

\[ b_1 = n \sum_{i=1}^{n} (i - 1) x_{in} \]

\[ b_6 = n \sum_{i=1}^{n} x_{in} \]

\[ \gamma = 0.577 \] is Euler’s constant

Monte Carlo simulations: Monte Carlo simulations have been carried out to investigate the effect of LQ moments methods based on five different quantile estimators compared with L-moments method. In each simulation a total of 1000 samples of size 10, 25 and 50 are used to generate random samples to obtain the quantile estimators of \( Q(q) \), \( q = 0.01, 0.05, 0.1, 0.025, 0.5, 0.75, 0.90, 0.95 \) and 0.99. Results are presented in Table 1 in terms of estimation efficiency (EFF) of the LQMOM method relative to LMOM method defined as:

\[ \text{EFF} = \frac{\text{RMSE(LQMOM)}}{\text{RMSE(LMOM)}} \]  

(25)

Values EFF >1 indicated that the LMOM method is superior to the LQMOM methods.

The simulation results of the Table 1 show that when the data are generated form the EV1 distribution, only the LQMOM based on WKQ1 is more efficient relative to the LMOM method in 16 out of 27 (equal to 59.26%). The LQMOM method based on the KQ, WKQ2 and HDL are in many cases significantly more efficient than the LMOM.

In the upper tails (\( q = 0.90, 0.95, 0.99 \)), the LQMOM based on the WKQ1 estimator has higher efficiency than LMOM equal to 100% cases. The LQMOM method based on the KQ, WKQ2 and HDL are in many cases significantly more efficient than the LMOM in 6 out of 9 (equal to 66.67%), in 7 out of 9 (equal to 77.78%) and in 6 out of 9 (equal to 66.67%) cases, respectively. The LIQ quantile estimator is significantly less efficient than the LMOM method.

RESULTS AND DISCUSSION

Simulation study for parent distribution function known: Although the true underlying distribution function is never known in practice, it is still useful to look at how estimation is affected by various methods when the distribution function is known. For this purpose, 1000 random samples of different \( n \) are generated from the EV1 distribution with the location and scale parameters \( (\mu, \sigma) \) were set 0 and 1 respectively.

Initially, parameters of EV1 were estimated by the LQMOM method using combinations of the quick estimators parameters (\( \alpha \) and \( p \)) values in the ranges 0-0.5. In the computer simulations the values of \( \alpha \) are 0.01(0.02)0.41 and \( p \) are 0.05(0.05)0.45 were chosen and all possible combination of \( \alpha \) and \( p \) were examined in order to find the best combination in term of RMSE. The presentation of our results will focus on the properties of quantile estimators because they are more direct practical interest.

The smallest RMSE of the LQMOM based on five quantile estimators obtained by simulation were compared with RMSE obtained using the LMOM method for sample sizes of \( n = 10, 25 \) and 50 with \( p = 0.01, 0.05, 0.1, 0.025, 0.5, 0.75, 0.90, 0.95 \) and 0.99. Results are presented in Table 1 in terms of estimation efficiency (EFF) of the LQMOM method relative to LMOM method defined as:

\[ \text{EFF} = \frac{\text{RMSE(LQMOM)}}{\text{RMSE(LMOM)}} \]  

(25)

Parent distribution function unknown: In practice, the true distribution function is never known. Thus it will be even more useful to look how estimation is affected by various methods when the assumed distribution function differs from the parent distribution function. In this study generalized Lambda distribution was considered to generate the random samples data.
Table 1: Estimation efficiency of the LQMOM method relative to LMOM method. Data are generated form the EV1 distribution

| Estimator | Sample size n = 10, 25, 50 |
|-----------|-----------------------------|
| LIQ       | 0.9613 0.9616 0.9706 0.9776 1.0333 1.0580 0.9943 0.9990 0.9989 |
| KQ        | 0.9826 0.9524 0.9280 0.9762 2.0749 1.0570 1.0591 1.0615 1.0596 |
| WKQ1      | 1.0786 1.0929 1.1025 1.1251 1.1176 1.0438 1.0278 1.0231 1.0259 |
| WKQ2      | 0.9992 0.9764 0.9573 1.0114 1.9813 1.0642 1.0627 1.0635 1.0600 |
| HDL       | 1.0024 0.9821 0.9660 0.9405 0.9893 1.0373 1.0514 1.0533 1.0592 |

Table 2: Estimation efficiency of the LQMOM method relative to LMOM method. Data are generated form the standard normal distribution

| Estimator | Sample size n = 10, 25, 50 |
|-----------|-----------------------------|
| LIQ       | 1.1606 1.0568 0.9894 1.0878 1.4081 1.3666 0.9669 1.1378 1.7151 |
| KQ        | 1.0593 0.9533 0.9254 1.0388 2.3060 1.0103 1.2468 2.1768 |
| WKQ1      | 1.1141 1.0453 0.9776 1.0519 1.8459 1.0131 1.0642 1.1648 1.1411 |
| WKQ2      | 0.9038 0.9071 0.9562 1.1849 1.4518 1.1376 0.9834 1.1905 1.8116 |
| HDL       | 1.0009 1.0105 1.0076 1.0925 1.4112 1.1986 1.5864 2.7071 |

Huang and Brill\[7\] used the Generalized Lambda Distribution (GLD) to compare the performance of the new quantile estimation method (WKQ) with the usual Kernel Quantile (KQ) estimation method. The GLD has four parameters and a wide variety of curve shapes. Hence it is useful for the representation of data when the underlying distribution is unknown. The quantile function of the GLD is given by:

$$Q(q) = a + [q^b - (1 - q)^c] / d, \quad 0 < q < 1, \quad b \neq 0 \quad (26)$$

Where:
- $a$ = A location parameter
- $b$ = A scale parameter
- $c$ and $d$ = Shape parameters.

By varying parameter values, Huang and Brill\[7\] constructed three generalized Lambda distributions, identified as the standard normal-like distribution ($a = 0, b = 0.1974, c = 0.1349, d = 0.1349$), the peaked symmetric distribution ($a = 0, b = -0.3203, c = -0.1359, d = -0.1359$) and the medium positive skewed distribution ($a = 0.6390, b = 0.0979, c = 0.0251, d = 0.0953$). The three GLDs then used as parent distributions in simulations to assess the performance of various quantile estimation of the LQMOM method compared to the LMOM method for estimating the parameters of the EV1 distribution.

The simulation result for the Q(q), q = 0.01, 0.05, 0.1, 0.025, 0.5, 0.75, 0.90, 0.95 and 0.99 for sample sizes of n = 10, 25 and 50 of the EV1 distribution are listed in terms of Estimation Efficiency (EFF) are shown in Table 3 and 4. The simulation results show that the LQMOM methods have better efficiencies in 56 out of 81 (equal to 69.14%) cases for WKQ1, in 62 out of 81 (equal to 76.54%) cases for KQ, in 50 out of 81 (equal to 61.73% cases for the HDL and for the LIQ estimator and in 60 out of 81 (equal to 74.07%) cases for the WKQ1 estimator.
In the upper tails (q = 0.90, 0.95, 0.99), the LQMOM based on all quantile estimators also always perform better than the LMOM. The KQ and WKQ2 have higher efficiency relative to the LMOM in 24 out of 27 (88.89%) respectively, followed by the HDL has 22 out 27 (81.48%), the WKQ1 has 20 out of 27 (74.07%) and the LIQ has 16 out of 27 (59.26%) cases overall.

Overall results presented in Table 2-4 show that the LQMOM based on the WKQ1 has 72 out of 108 (equal to 66.67%) cases, the KQ has 67 out of 108 (62.04%), the HDL has 54 out of 108 (50%), the WKQ2 has 65 out of 108 (60.19%) and the LIQ has 50 out of 108 (46.30%) cases with better efficiencies.

CONCLUSION

An accurate estimation of parameters of the Extreme Value Type 1 (EV1) distribution in statistical analysis of extremes is of considerable importance. In this study, we develop improved the LQMOM that does not impose restrictions on the value of the quick estimators parameters p and a but we explore an extended class of LQ-moments with consideration combinations of p and a values in the range 0 and 0.5. The method of the LQMOM based on five different quantile estimators were examined and compared their performances against a widely acceptable method of L-moments using simulated samples of both EV1 and generalized Lambda distribution.

Considering all factors of comparison, the LQMOM based on WKQ1 in many cases significantly more efficient than LMOM, when the data are generated from the EV1 distribution and other distributions. The LQMOM based on WKQ2, KQ and HDL have higher efficiency relative to the LMOM when only the data are simulated from the generalized Lambda distribution.
When the upper quantiles, $q \geq 0.90$ are considered, the LQMOM based on all quantile estimators except LIQ always perform better than LMOM for moderate and small sample sizes.

This study has demonstrated that the conventional LMOM is not optimal for the estimation of the EV1 distribution. The new method of estimation, denoted the LQMOM in many cases represents higher efficiency in the quantile estimation compared to the other quantile estimators. The simplicity and generally good performance of this method make it an attractive option for estimating quantiles in the EV1 distribution.

ACKNOWLEDGEMENT

The researchers would like to thanks Ministry of Science, Technology and Innovation, Malaysia for funding this research.

REFERENCES

1. Ani, S. and A.A. Jemain, 2006. LQ-moments: Application to the extreme value type I distribution. J. Applied Sci., 6: 993-997. http://adsabs.harvard.edu/abs/2006JApSc...6..993S
2. Ani, S. and A.A. Jemain, 2006. LQ-moments for statistical analysis of extreme events. J. Modern Applied Stat. Method., 6: 228-238. http://eprints.utm.my/7643/1/Anishabri2007_LQmomentsForStatisticalAnalysis.pdf
3. Ani, S. and A.A. Jemain, 2007. LQ-moments: Application to the generalized extreme value. J. Applied Sci., 7: 115-120. http://www.doaj.org/doaj?func=abstract&id=318059
4. David, H.A. and H.N. Nagaraja, 2003. Order Statistics. 3rd Edn., Wiley, New Jersey, ISBN: 0-471-38926-9.
5. Gumbel, E.J., 1958. Statistics of Extremes. Columbia University Press, New York, ISBN: 0486436047, pp: 377.
6. Harrell, F.E. and C.E. Davis, 1982. A new distribution-free quantile estimator. Biometrika, 69: 635-640. http://biomet.oxfordjournals.org/cgi/content/abstract/69/3/635
7. Huang, M.L. and P. Brill, 1999. A level crossing quantile estimation method. Stat. Probability Lett., 45: 111-119. http://cat.inist.fr/?aModele=afficheN&cpsidt=1967635
8. Huang, M.L., 2001. On a distribution-free quantile estimator. Comput. Stat. Data Anal., 37: 477-486. http://portal.acm.org/citation.cfm?id=568747.568752
9. Hyndman, R.J. and Y. Fan, 1996. Sample quantiles in statistical packages. Am. Stat., 50: 361-365. http://www.jstor.org/stable/2684934
10. Landwehr, J.M., N.C. Matalas and J.R. Wallis, 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour. Res., 15: 1055-1064. http://adsabs.harvard.edu/abs/1979WRR....15.1055L
11. Mudholkar, G.S. and A.D. Hutson, 1998. LQ-Moments: Analogs of L-moments. J. Stat. Plann. Inference, 71: 191-208. http://cat.inist.fr/?aModele=afficheN&cpsidt=1771266
12. Phien, H.N., 1987. A review of methods of parameter estimation for the extreme value type I distribution. J. Hydrol., 86: 391-398. http://adsabs.harvard.edu/abs/1987JHyd..86..391P
13. Rasmussen, P.F. and N. Gautam, 2003. Alternative PWM-estimators of the Gumbel distribution. J. Hydrol., 280: 265-271. http://adsabs.harvard.edu/abs/2003JHyd..280..265R
14. Raynal, J.A. and J.D. Salas, 1986. Estimation procedures for the type-I extreme value distribution. J. Hydrol., 87: 315-336. http://cat.inist.fr/?aModele=afficheN&cpsidt=7872247
15. Sheather, S.J. and J.S. Marron, 1990. Kernel quantile estimators. J. Am. Stat. Assoc., 85: 410-416. http://www.jstor.org/stable/2289777