© Heldermann Verlag
ISSN 0940-5151

Economic Quality Control
Vol 20 (2005), No. 1, 31 – 39

A Modified Quantile Estimator Using Extreme-Value Theory with Applications

M.B. Vermaat, R.J.M.M. Does and A.G.M. Steerneman

Abstract: Reliable predictions by means of quantiles constitute one of the most important tasks not only in statistics but in entire science. Quantiles may be estimated by using Extreme-Value Theory (EVT). However, the properties of many estimators based on this theory depend heavily on the actual location. In this paper modified estimators for the quantiles are derived, the properties of which are less sensitive with respect to location. Moreover, these modified quantile estimators are also symmetric with regard to the mean for symmetric distributions, which is not the case for some of the estimators based on the EVT. The modified quantile estimators are a limiting result of an infinity shift of location of the estimators proposed by Dekkers et al. (1989). The results may be used in establishing control limits for Shewhart control charts.

Key Words: Asymptotics, control limits, Shewhart control charts, extreme-value theory.

1 Introduction

The Extreme-Value Theory (EVT) deals with modelling the extremes. It can be applied to predict the occurrence of rare events. The prediction of flood levels, large jumps in the stock markets and extreme insurance claims are application areas of EVT. These application areas, i.e. insurance and finance, are treated in the textbook by Embrechts et al. (1999). A characteristic feature of these areas is the need for so-called one-sided predictions. Industry is another application area of the EVT. However, in industrial application often two-sided predictions are needed which require the determination of quantiles of a large and a small order.

Estimation of quantiles of large orders using the EVT is described in Dekkers et al. (1989), see also Weissman (1978) and Boos (1984). The quantile estimators of Dekkers et al. (1989) applied to quantiles of small order are for certain distributions, e.g. the uniform distribution, not quite satisfactory. It appears that for the uniform distribution on $[0, 1]$ the estimated quantiles of a large order $1 - q$ $(0 < q < 0.5)$ and of the corresponding small order q are not symmetric with regard to the average (i.e. 0.5). The reason for this is as follows. The quantile of small order is very close to zero. Dekkers et al. (1989) use a log-transformation to estimate the quantiles. The log-function is very steep in the neighborhood of zero, which causes asymmetry in the estimated small and large quantiles. To tackle this problem, a large shift of location is performed. We recalculate the quantiles for the transformed data. To get the quantiles for the original data, an inverse transformation of the calculated quantiles is executed.

The question is how large this shift of location has to be in order to get appropriate results. In other words, what happens if the shift of location goes to infinity? In this paper we derive the limiting values for quantile estimators for large and small order, respectively. As an illustration we calculate the quantiles of a uniform distribution on $[0, 1]$ based on the work of Dekkers et al. (1989) as well as the modified estimator.

This paper is organized as follows. In the next section we describe the estimation method for quantiles of large and small order based on Dekkers et al. (1989). In the subsequent section the modified estimators are introduced by a theorem. The theorem is proven in the appendix. In the next section an example is given using a uniform distribution on $[0, 1]$. We finish the paper with a conclusion.

2 Extreme-Value Theory

In this section we follow Dekkers et al. (1989) for the quantile estimation. Consider an independent sample $X_1, \ldots, X_n$ of a certain distribution F. We will assume throughout this paper that F is absolutely continuous. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the order statistics of this sample. Suppose that the sample $X_1, \ldots, X_n$ is such that the largest m and smallest m order statistics are all either on the positive reals or on the negative reals. Define

$$M_n^{(r)} = \frac{1}{m} \sum_{j=1}^{m} \left(\log \frac{X_{(n-j+1)}}{X_{(n-m)}} \right)^r \tag{1}$$

and

$$L_n^{(r)} = \frac{1}{m} \sum_{j=1}^{m} \left(\log \frac{X_{(j)}}{X_{(m+1)}} \right)^r \tag{2}$$

where the integer r takes the values $r = 1, 2$ and m is the number of upper respectively lower order statistics used in the estimation of quantiles of large and small order respectively.

The Extreme-Value Theory deals with the tail behavior of distributions. These tails can be modelled by an extreme-value distribution, which is determined by an extreme-value index γ (cf. Dekkers et al. (1989)). If we do not make any assumptions on γ, we may use the moment estimator for it. For quantiles of large order it is defined by

$$\hat{\gamma}_n = M_n^{(1)} + 1 - \frac{1}{2} \left\{ 1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right\}^{-1} \tag{3}$$

and for quantiles of small order we can define its equivalent

$$\hat{\delta}_n = L_n^{(1)} + 1 - \frac{1}{2}\left\{1 - \frac{(L_n^{(1)})^2}{L_n^{(2)}}\right\}^{-1} \tag{4}$$

Furthermore, the $(1-q)$-quantile of the distribution function F for $0 < q < \frac{1}{2}$ is estimated in Dekkers, Einmahl and De Haan (1989) as

$$\hat{F}^{-1}_{n,\mathrm{DEH}}(1-q;\hat{\gamma}_n) = X_{(n-m)} + \frac{\left(\frac{m}{nq}\right)^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n}(1 - \min(\hat{\gamma}_n, 0))X_{(n-m)}M_n^{(1)} \tag{5}$$

The q-quantile of the distribution function F for $0 < q < \frac{1}{2}$ is estimated as

$$\hat{F}^{-1}_{n,\mathrm{DEH}}(q;\hat{\delta}_n) = X_{(m+1)} + \frac{\left(\frac{m}{nq}\right)^{\hat{\delta}_n} - 1}{\hat{\delta}_n}(1 - \min(\hat{\delta}_n, 0))X_{(m+1)}L_n^{(1)} \tag{6}$$

3 Convergence of the Quantiles

In this section we state our main result. First we shift the original observations with a constant K. For the transformed observations we estimate the quantiles with Dekkers' et al. (1989) method. Applying the inverse transformation we obtain estimators for the quantiles of the original data. In Theorem 1 we formulate the result for $K \to \infty$.

Theorem 1 Define $X^* = X + K$, where X is a random variable with an absolutely continuous probability distribution function F and K is a constant. Suppose that $\frac{m}{nq} \geq 1$ and suppose that the largest m and smallest m order statistics are all either on the positive reals or on the negative reals.

a) If $K \to \infty$ then the modified $(1-q)$-quantile estimator with $0 < q < \frac{1}{2}$ is

$$\hat{F}^{-1}_{n,\mathrm{MDEH}}(1-q) = X_{(n-m)} + D\left(\frac{m}{nq}\right)\frac{1}{m}\sum_{j=1}^{m}(X_{(n-j+1)} - X_{(n-m)}) \tag{7}$$

with

$$D\Big(\frac{m}{nq}\Big) = \frac{(\frac{m}{nq})^{G_n} - 1}{G_n}(1 - \min(G_n, 0)) \tag{8}$$

$$G_n = 1 - \frac{1}{2(1-Q_n)}$$

and

$$Q_n = \frac{\left(\frac{1}{m}\sum\limits_{j=1}^{m}(X_{(n-j+1)} - X_{(n-m)})\right)^2}{\frac{1}{m}\sum\limits_{j=1}^{m}(X_{(n-j+1)} - X_{(n-m)})^2}$$

b) If $K \rightarrow \infty$ then the modified q-quantile estimator with $0 < q < \frac{1}{2}$ is

$$\hat{F}^{-1}_{n,\mathrm{MDEH}}(q) = X_{(m+1)} + \bar{D}\left(\frac{m}{nq}\right)\frac{1}{m}\sum_{j=1}^{m}(X_{(j)} - X_{(m+1)}) \tag{9}$$

with

$$\bar{D}\Big(\frac{m}{nq}\Big) = \frac{(\frac{m}{nq})^{\bar{G}_n} - 1}{\bar{G}_n}(1 - \min(\bar{G}_n, 0)) \tag{10}$$

$$\bar{G}_n = 1 - \frac{1}{2(1-\bar{Q}_n)}$$

and

$$\bar{Q}_n = \frac{\left(\frac{1}{m}\sum_{j=1}^{m}(X_{(j)} - X_{(m+1)})\right)^2}{\frac{1}{m}\sum_{j=1}^{m}(X_{(j)} - X_{(m+1)})^2}$$

The proof of Theorem 1 is given in the appendix.

Remark: A necessary restriction in Theorem 1 is that that $\frac{m}{nq} \geq 1$, otherwise $D(\frac{m}{nq})$ and $\bar{D}(\frac{m}{nq})$ in (8) and (10) do not converge for all G_n and $\bar{G}_n$ in $(-\infty, \frac{1}{2})$, respectively. This restriction follows to a certain extent from Dekkers et al. (1989). They distinguish two situations:

1. finite case (Theorem 4.1 in Dekkers et al. (1989)): the number m of order statistics is fixed, $m > c$ with $c \in (0, \infty)$ and suppose that $q = q_n \rightarrow 0$, $nq_n \rightarrow c$, for $n \rightarrow \infty$;
2. infinite case (Theorem 5.1 in Dekkers et al. (1989)): the number m of order statistics is not fixed. Suppose $q_n \rightarrow 0$, $nq_n \rightarrow \infty$ $(n \rightarrow \infty)$ and $m(n) = nq_n$.

Further De Haan and Rootzén (1993) reported that if $q < \frac{1}{n}$ nothing can be done unless one imposes extra conditions on F, hence the requirement $nq > 1$. This restriction in combination with the two situations mentioned implies the prerequisite $\frac{m}{nq} \geq 1$. De Haan and Rootzén (1993) made extra assumptions to analyse large quantile estimators if $q < \frac{1}{n}$.

4 Application

By monitoring industrial processes, we want to distinguish variation of 'common causes' from variation of 'special causes' (cf. Does et al., 1999). A well-known tool in monitoring is the control chart, which can be of various types. A control chart consists of a graph with time on the horizontal axis against a control characteristic (an individual measurement or a statistic such as mean or range) on the vertical axis. In this graph an upper control limit (UCL) and a lower control limit (LCL) are drawn.

An industrial production process is called statistically in-control if the values of the control characteristic are from one and the same desirable distribution. The interval between the control limits of a control chart constitutes a highly probable prediction in case the process operates in-control and any observation outside the interval indicates the occurrence of a 'special cause'. Control charts provide easy checks on the actual state of the process. The determination of the control limits is a research topic, which originates from Walter Shewhart (1931).

In Vermaat et al. (2003) EVT is used to determine control limits of control charts regarding the upper and lower control limit as quantiles of large and small order, respectively. Suppose we have $X_1, \ldots, X_k$ from a uniform distribution on $[0, 1]$, then for Shewhart-type control charts the control limits guarantee a false alarm risk of $q = 0.00135$ implying that quantiles of oder 0.99865 and of 0.00135 are needed.

The control limits based on the estimators by Dekkers et al. (1989) are defined by (see (6) respectively (5))

$$\mathrm{LCL}_{\mathrm{DEH}} = \hat{F}_{n,\mathrm{DEH}}(q, \hat{\gamma}_n)$$

and

$$\mathrm{UCL}_{\mathrm{DEH}} = \hat{F}_{n,\mathrm{DEH}}(1 - q, \hat{\gamma}_n)$$

The control limits based on the modified estimators are defined by (see (9) respectively (7))

$$\mathrm{LCL}_{\mathrm{MDEH}} = \hat{F}_{n,\mathrm{MDEH}}(q, \hat{\gamma}_n)$$

and

$$\mathrm{UCL}_{\mathrm{MDEH}} = \hat{F}_{n,\mathrm{MDEH}}(1 - q, \hat{\gamma}_n)$$

To illustrate the difference for both methods, we sample 10,000 observations from a uniform distribution on $[0, 1]$. Then we calculate the control limits based on both methods. We repeat this procedure 10,000 times and calculate the average of the control limits. For the control limits based on the estimators of Dekkers et al. (1989) we get $\mathrm{LCL}_{\mathrm{DEH}} = 0.000675$ and $\mathrm{UCL}_{\mathrm{DEH}} = 0.998617$. For the modified method we get $\mathrm{LCL}_{\mathrm{MDEH}} = 0.001388$ and $\mathrm{UCL}_{\mathrm{MDEH}} = 0.998617$. We see that the results for both UCLs are the same and are close to the theoretical value 0.99865. However, we see the $\mathrm{LCL}_{\mathrm{DEH}}$ differs substantially from the theoretical value of 0.00135. The $\mathrm{LCL}_{\mathrm{MDEH}}$ is close to this value. Moreover, the $\mathrm{LCL}_{\mathrm{MDEH}}$ is symmetric around 0.5 compared with the $\mathrm{UCL}_{\mathrm{MDEH}}$.

5 Conclusions

In this paper quantile estimators of Dekkers et al. (1989) are evaluated when a shift of location is applied. If the data are translated and the quantiles of large and small order are calculated, the reverse translated quantiles converge, and modified quantile estimators arise. The advantage of these new quantile estimators is that the corresponding estimates of large and small order are symmetric around the mean for a symmetric distribution, e.g. the uniform distribution. This is not the case for the estimator proposed by Dekkers et al. (1989). In our theory one can also calculate quantiles of negative values. Dekkers et al. (1989) considered only positive values and, if not, they advise to apply a simple shift. However, a shift does change the properties of the estimator, it is sensitive with respect to location changes. Our modified estimator is resistant in this respect. The modified estimator may be applied for establishing control limits.

Acknowledgements: The authors thank the referees and the editor for their useful comments and suggestions.

Appendix

1 Proof of Theorem 1

a) Quantile of large order:

By replacing X with X^* the definitions of $\hat{\gamma}_n$ in (3) and $M_n^{(r)}$ in (1) are modified in $\hat{\gamma}_n^*$ and $M_n^{(r)^*}$. The $(1-q)$-quantile of the distribution function of X^* can be easily obtained from (5) by replacing $X_{(n-m)}, \hat{\gamma}_n$ and $M_n^{(1)}$ by $X^*_{(n-m)}, \hat{\gamma}_n^*$ and $M_n^{(r)^*}$, respectively. A modified estimator of the $(1-q)$-quantile of the distribution function F of the original random variable X is derived as the limiting value for $K \to \infty$ of

$$\hat{F}^{-1}_{n,\mathrm{DEH}}(1-q;\hat{\gamma}_n^*;K) = X^*_{(n-m)} + \frac{\left(\frac{m}{nq}\right)^{\hat{\gamma}_n^*}-1}{\hat{\gamma}_n^*}(1-\min(\hat{\gamma}_n^*,0))X^*_{(n-m)}M_n^{(1)^*} - K$$

First we calculate the limit of $\hat{\gamma}_n^*$:

$$\lim_{K\to\infty}\hat{\gamma}_n^* = \lim_{K\to\infty}\left\{M_n^{(1)^*}+1-\frac{1}{2}\left[1-\frac{(M_n^{(1)^*})^2}{M_n^{(2)^*}}\right]^{-1}\right\} \tag{11}$$

So, we have to consider

$$\frac{(M_n^{(1)^*})^2}{M_n^{(2)^*}} = \frac{\left(\frac{1}{m}\sum_{j=1}^{m}\log\frac{X^*_{(n-j+1)}}{X^*_{(n-m)}}\right)^2}{\frac{1}{m}\sum_{j=1}^{m}\left(\log\frac{X^*_{(n-j+1)}}{X^*_{(n-m)}}\right)^2}$$

Now we write

$$\begin{aligned}\log\frac{X^*_{(n-j+1)}}{X^*_{(n-m)}} &= \log\frac{X_{(n-j+1)}+K}{X_{(n-m)}+K}\\ &= \log\left(1+\frac{X_{(n-j+1)}-X_{(n-m)}}{X_{(n-m)}+K}\right)\\ &= \log(1+C_jH)\end{aligned}$$

where $C_j = X_{(n-j+1)} - X_{(n-m)}$ and $H = \frac{1}{X_{(n-m)}+K}$.

Since $K \to \infty$ we have $H \to 0$, and hence

$$\lim_{H\to 0}\frac{\log(1+C_jH)}{H} = C_j$$

This implies that

$$\lim_{H\to 0}\frac{\frac{1}{m}\sum_{j=1}^{m}\log(1+C_jH)}{H} = \frac{1}{m}\sum_{j=1}^{m}C_j$$

and

$$\lim_{H\to 0}\frac{\frac{1}{m}\sum_{j=1}^{m}\Big(\log(1+C_jH)\Big)^2}{H^2} = \frac{1}{m}\sum_{j=1}^{m}C_j^2$$

Hence, we find that

$$\lim_{K\to\infty}\frac{(M_n^{(1)^*})^2}{M_n^{(2)^*}} = \frac{(\frac{1}{m}\sum_{j=1}^{m}C_j)^2}{\frac{1}{m}\sum_{j=1}^{m}C_j^2}$$

By substitution of the C_j we get

$$\lim_{K\to\infty}\frac{(M_n^{(1)^*})^2}{M_n^{(2)^*}} = \frac{\Big(\frac{1}{m}\sum_{j=1}^{m}(X_{(n-j+1)}-X_{(n-m)})\Big)^2}{\frac{1}{m}\sum_{j=1}^{m}(X_{(n-j+1)}-X_{(n-m)})^2} = Q_n$$

Note that $0 < Q_n < 1$ with probability one. Further, we know that

$$\begin{aligned}\lim_{K\to\infty}M_n^{(1)^*} &= \lim_{K\to\infty}\frac{1}{m}\sum_{j=1}^{m}\log\frac{X^*_{(n-j+1)}}{X^*_{(n-m)}}\\ &= \lim_{K\to\infty}\frac{1}{m}\sum_{j=1}^{m}\log\left(\frac{X_{(n-j+1)}+K}{X_{(n-m)}+K}\right) = 0\end{aligned}$$

Based on these findings we conclude from (11) that

$$\lim_{K\to\infty} \hat{\gamma}_n^* = 1 - \frac{1}{2(1-Q_n)} = G_n \tag{12}$$

with $G_n \in (-\infty, \frac{1}{2})$ with probability one. Using (12) we obtain that

$$\frac{\left(\frac{m}{nq}\right)^{\hat{\gamma}_n^*} - 1}{\hat{\gamma}_n^*}(1 - \min(\hat{\gamma}_n^*, 0)) \to D\Big(\frac{m}{nq}\Big)$$

where

$$D\Big(\frac{m}{nq}\Big) = \frac{(\frac{m}{nq})^{G_n} - 1}{G_n}(1 - \min(G_n, 0))$$

It is supposed that $\frac{m}{nq} \geq 1$, so $D(\frac{m}{nq}) \geq 0$ and bounded, with probability one. Now we have

$$\begin{aligned}
&\lim_{K\to\infty} \hat{F}^{-1}_{n,\mathrm{DEH}}(1-q; \hat{\gamma}_n^*; K) \\
&\quad = \lim_{K\to\infty} \left\{ X^*_{(n-m)} + \frac{\left(\frac{m}{nq}\right)^{\hat{\gamma}_n^*} - 1}{\hat{\gamma}_n^*}(1 - \min(\hat{\gamma}_n^*, 0)) X^*_{(n-m)} M_n^{(1)^*} - K \right\} \\
&\quad = X_{(n-m)} + \lim_{K\to\infty} \frac{\left(\frac{m}{nq}\right)^{\hat{\gamma}_n^*} - 1}{\hat{\gamma}_n^*}(1 - \min(\hat{\gamma}_n^*, 0)) X^*_{(n-m)} M_n^{(1)^*} \qquad (13)
\end{aligned}$$

In order to evaluate (13), we focus on $X^*_{(n-m)} M_n^{(1)^*}$. Using the same notation for C_j and H as before, we derive that

$$\begin{aligned}
\lim_{K\to\infty} X^*_{(n-m)} M_n^{(1)^*} &= \lim_{K\to\infty} \frac{1}{m} \sum_{j=1}^{m} (X_{(n-m)} + K) \log \frac{X_{(n-j+1)} + K}{X_{(n-m)} + K} \\
&= \lim_{H\to 0} \frac{1}{m} \sum_{j=1}^{m} \frac{\log(1 + C_j H)}{H} \\
&= \frac{1}{m} \sum_{j=1}^{m} C_j
\end{aligned}$$

Hence, if $K \to \infty$ then $\hat{F}^{-1}_{n,\mathrm{DEH}}(1-q; \hat{\gamma}_n^*; K)$ converges to

$$\hat{F}^{-1}_{n,\mathrm{MDEH}}(1-q) = X_{(n-m)} + D\left(\frac{m}{nq}\right) \frac{1}{m} \sum_{j=1}^{m} \Big(X_{(n-j+1)} - X_{(n-m)}\Big)$$

with $D(\frac{m}{nq})$ as in (8).

b) Quantile of small order:

The proof for quantiles of small order goes analogously. □

References

[1] Boos, D.D. (1984): Using extreme value theory to estimate large percentiles. *Technometrics* 26, 33–39.

[2] Dekkers, A. L. M., Einmahl, J. H. J. and Haan, de L. (1989): A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* 17, 1833–1855.

[3] Does, R. J. M. M., Roes, K. C. B. and Trip, A. (1999): *Statistical process control in industry.* Kluwer Academic, Dordrecht, The Netherlands.

[4] Haan, de L. and Rootzéen, H. (1993): On the estimation of high quantiles. *Journal of Statistical Planning and Inference* 35, 1–13.

[5] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1999): *Modelling extremal events for insurance and finance.* Springer-Verlag, New York.

[6] Shewhart, W. A. (1931): *Economic control of quality of manufactured product.* Van Nostrand, Princeton, New York.

[7] Vermaat, M. B., Ion, R. A., Does, R. J. M. M. and Klaassen, C. A. J. (2003): A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory. *Quality and Reliability Engineering International* 19, 337–353.

[8] Weissman, I. (1978): Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* 73(364), 812–815.

M.B. Vermaat and R.J.M.M. Does
Institute for Business and Industrial Statistics, IBIS UvA
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands

A.G.M. Steerneman
Department of Econometrics
University of Groningen
P.O. Box 800
9700 AV Groningen
The Netherlands