1137

Open Access

The Numerical Analysis of KL Quantile Estimates

Wei Jiang[1] and Yuhua Su[2],*

[1]*Department of Personnel, Hezhou University, Hezhou, Guangxi, 542899, China*

[2]*School of Science, Hezhou University, Hezhou, Guangxi, 542899, China*

Abstract: This article uses R software KL quantile estimate numerical simulation under different distributions and different sample values, and on this basis, two cases of KL SQ sample quantile quantile estimates and estimated mean square numerical errors, the simulation results show that: in most cases, KL quantile estimate of the mean square error is less than SQ sample quantile estimation mean square error; the truncated distribution (such as exponential distribution and uniform distribution) cut end of the estimated effect of KL quantile estimate is very good, and much better than the median estimate SQ sample points; heavy tail of the distribution (*e.g.* distribution) have an impact on the estimated effects of KL quantile estimates.

Keywords: KL quantile estimate, SQ sample quantile estimate, numerical analysis.

1. INTRODUCTION

Quantile estimation is one of the important research topics in statistics, it has been widely used in many areas of financial risk management, it is important quantile amount of risk, there are many non-scholars quantile parameter estimation for a more in-depth research, such as the sample quantile estimated weighted order statistics quantile estimates and nuclear quantile estimation [1-4]. quantile estimate has good theoretical properties through the study found and application value. As described in [5] Kaigh and Lachenbruch probability level under the conditions by numerical comparison KL quantile estimates and SQ sample estimate quantile analysis showed that KL quantile estimate generally superior to SQ sample quantile estimate; [6] discuss KL quantile by Jackknife variance estimation of probability level numerical simulation analysis, we found the KL estimate quantile estimation method is better than Bootstrap fitting effect; [7] in the overall distribution of uniform under the distribution and the sample value of 10 cases studied by numerical simulation the optimal value of KL quantile probability at different levels of the corresponding sub-sample size; [8] method through empirical analysis on a variety of VaR model and KL quantile estimate comparative analysis found that KL quantile measure is estimated to be well worth the risk of the portfolio.

Throughout these studies, we found that KL quantile estimate numerical simulation at high probability levels achieved some results, but for a small probability level ($p \leq 0.05$) numerical simulation under KL quantile estimation lacks. In order to further improve the KL quantile esti

mation results of numerical analysis to be able to replace KL quantile estimate value at risk VaR to measure the risks of financial products, we use the R software for KL quantile estimate numerical analysis, discuss its fitting effect and on this basis will KL SQ sample quantile quantile estimation and simulation analysis and comparison of estimates.

2. NUMERICAL SIMULATION OF KL QUANTILE ESTIMATES

2.1. KL Quantile Estimate Weights

KL quantile estimate is a weighted order statistics quantile Kaigh and Lachenbruch first proposed in 1982 estimated that it can serve as a broad sample quantile, which is defined as:

$$K_{[(k+1)p]k:n} = \frac{1}{C_n^k} \sum_{\alpha 1,\alpha 2 \cdots \alpha k} Y_{[(k+1)p]k:n}(X_{\alpha 1}, X_{\alpha 2}, \cdots X_{\alpha k}) \tag{1}$$

In order to better numerical analysis, (1) can also be written as the following expression:

$$K_{[(k+1)p]k:n} = \sum_{j=r}^{r+n-k} \frac{C_{j-1}^{r-1} C_{n-j}^{k-r}}{C_n^k} X_{j:n}, r = \left[(k+1)p\right] \tag{2}$$

From the above equation, right th order statistics corresponding weight, which does not depend on the overall weight distribution, only the parameters. To visually see KL next quantile estimated weight distribution, we take, made the situation weights (Fig. **1**), it can be seen from the figure, the weight distribution in the vicinity of quantile, weights greater range is the first 45 order statistics to the first 60 order statistics, the right to re-order statistics other almost negligible. Center coordinates of each layer of ancient pagoda in Table **1**:
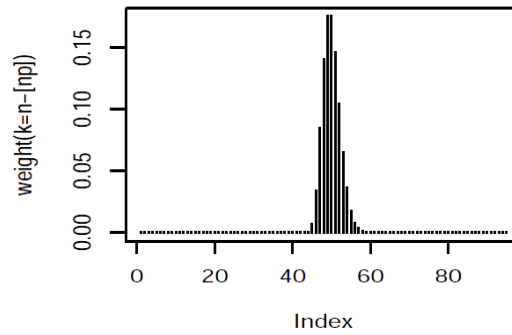
Fig. (1). KL quantile estimation of the weight distribution chart in n=500, k=0.1.

Table 1. Distribution of four kinds of different values KL quantile estimation mean square error.

	k	100	300	500	700	900
N(0,1)	P=0.01	0.049100157	0.01676314	0.0147543	0.01436412	0.01372016
	P=0.05	0.005912452	0.00459786	0.00457461	0.00462179	0.00476439
T(n,4)	P=0.01	1.4467913	0.2164581	0.1657256	0.0146178	0.1416431
	P=0.05	0.02647125	0.01512642	0.01486125	0.01489964	0.01463412
E(1)	P=0.01	5.537835e-06	7.801613e-06	8.716285e-06	9.446734e-06	9.924467e-06
	P=0.05	4.001135e-05	4.613694e-06	4.916344e-05	4.984687e-05	5.201364e-05
R(0,1)	P=0.01	5.134619e-06	7.201364e-06	8.106348e-06	8.634612e-06	9.213461e-06
	P=0.05	3.700012e-05	4.301568e-05	4.491672e-05	4.713647e-05	4.831679e-05

Fig. (**2**) overall distribution (top) and distribution.

2.2. KL Quantile Estimation Fitting Results

Numerical simulation for intuitive quantile estimates reflect KL fitting results, respectively, under the standard normal distribution and the distribution of conditions, and the situation KL quantile estimation, in each case repeated 5000 times to give KL quintile estimate estimate curve shown in Fig. (**2**), Fig. (**3**). with triangle dotted line represents the estimated value of KL quantile, the solid line represents the true quantile overall simulation.

The overall distribution of Fig. (**3**) (top) and distribution (below)

As can be seen from the figure, when the probability $p \in (0.2, 0.3, \cdots 0.8)$, KL quantile estimate simulation results were quite good, a small amount of deviation in the head and tail section, when the probability $p \in (0.01, \cdots 0.05)$, the whole KL quantile estimate is smaller than the true value.

2.3. Numerical Analysis of Neutron Sample Size Selection

In the numerical analysis, the problem first consideration is the selection of the sub-sample size, reference Kaigh and Lachenbruch suggests, we use the mean square error of $E(K_{(\lceil k+1 \rceil p:k:n)} - Q(p))^2$ to get the sub-sub-sample of the sample size was defined as the minimum value of the numerical analysis, the numerical simulation of repetitions mean square error is calculated as follows:

$$MSE_{kl}^{i}(p) = \frac{1}{m} \sum_{i=1}^{m} (KL_i(p) - Q(p))^2 \tag{3}$$

In which, $KL_i(p)$ represents the i simulations of KL quantile estimates, $Q(p)$ indicates that the analog overall real quantile.

Selecting a different sub-sample size in the sample size of 1000 samples, $k = 100, 300, 500, 700, 900$, respectively, in normal N(1,0), uniform distribution $R(0,1)$, $(df = 4)$ the case of the exponential distribution $E(1)$, simulation was repeated 2000 times, and the calculated KL quantile estimates are square error are listed in Table **1**:

Data in the table shows:

(1) A uniform distribution $R(0,1)$ and exponential distribution $E(1)$ mean square error than normal distribution
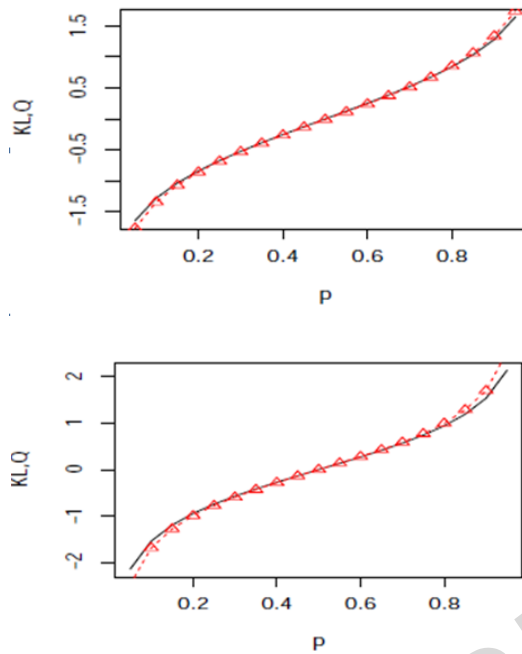
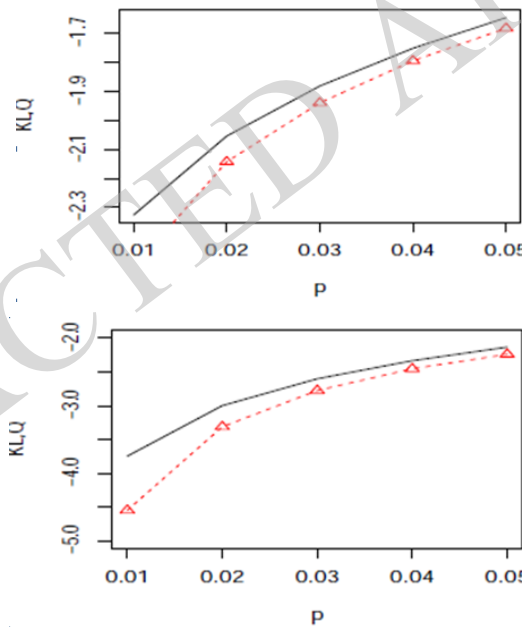**Fig. (2).** Overall distribution N(0,1) and t distribution (df=4).

Fig. (3). Overall distribution N(0,1) and t distribution (df=4).

$N(0,1)$ of the mean squared error is much smaller. Because uniform and left truncated exponential distribution, and the distribution was normal and the left tail, so the probability level is below 0.01 and 0.05, the above results will be reflected in the left end of the distribution. In other words, for a truncated distribution, the cut end of the KL-quantile is estimated to be better than the non-truncated tail end of KL quantile estimates. Mean square error.

(2) The mean square error is less than normal distribution, indicating heavy tails of the distribution of KL quantile estimate influential.

(3) The distribution t and the normal distribution, the majority of cases are sub-sample size increases, the corresponding mean square error reduction, and $k = 100$ obviously when $k = 100$ the mean square error of the mean square error is greater than when, while $k > 100$ reducing the amount

Table 2. The Probability level $p \leq 0.05$ MSE ratio (m=5000).

p	N(0,1)	R(0,1)	E(1)	Logis(0.1)	t(n,4)
n=100,k=99					
p=0.01	1.010806	0.9720669	1.0272021	1.0817680	0.7884654
p=0.02	1.016363	0.9941542	0.9812193	0.9813198	1.0284933
p=0.03	1.028179	1.0117961	1.0529235	1.0538652	1.0346324
p=0.04	1.038810	0.9800104	1.0136737	1.0145623	1.0195426
p=0.05	1.018850	0.9618123	0.9915477	0.9915977	1.0124652
n=300,k=150					
p=0.01	0.7568693	1.298419	1.284327	1.284672	0.6545892
p=0.02	0.7785453	1.204376	1.235618	1.234562	0.6999545
p=0.03	0.7628794	1.118624	1.162148	1.162453	0.6882634
p=0.04	1.0831150	1.133459	1.185218	1.185426	1.0254931
p=0.05	1.0319959	1.114265	1.209344	1.218924	0.9487163
n=500,k=375					
p=0.01	0.6112065	1.168156	1.134689	1.135261	0.4658554
p=0.02	0.9054816	1.141956	1.141536	1.146421	0.7954126
p=0.03	0.9918321	1.123640	1.086314	1.087316	0.9384561
p=0.04	0.5796267	1.097865	1.045948	1.045964	1.0175642
p=0.05	0.9438526	1.108178	1.048905	1.043512	0.8658321
n=1000,k=750					
p=0.01	0.9508651	1.124566	1.096991	1.098721	0.7854225
p=0.02	1.0889271	1.066131	1.048419	1.049124	1.0789954
p=0.03	1.0171082	1.071089	1.048987	1.034562	0.9810065
p=0.04	1.5946518	1.045789	1.048993	1.492546	1.0829875
p=0.05	1.0246314	1.065248	1.076351	1.074257	1.0356981

Table 3. The Probability level $p \in (0.1, 0.9)$ mean square error ratio (m=5000).

p	N(0,1)	R(0,1)	E(1)	Logis(0.1)	t(n,4)
n=100,k=99					
p=0.1	1.0128988	0.9807856	0.9698752	1.0128752	0.9833462
p=0.3	1.0519751	0.9556412	0.9679602	0.9854321	1.0120571

Table **3**. contd…

p	N(0,1)	R(0,1)	E(1)	Logis(0.1)	t(n,4)
p=0.5	1.0452630	1.0120050	0.9789952	1.0335468	1.0208475
p=0.7	1.0359633	1.0352932	0.9478653	1.0245863	1.0151602
p=0.9	0.9852341	1.1208957	0.0887650	0.9884678	0.9115747
n=300,k=150					
p=0.1	1.017789	1.060127	1.082950	0.9913478	0.9263425
p=0.3	1.023146	0.988712	1.036429	1.0265459	0.9814652
p=0.5	1.024536	1.004112	1.016425	1.0435657	1.0643150
p=0.7	1.019123	1.008975	1.044890	1.0274456	1.0682158
p=0.9	1.062468	1.125321	1.010368	1.0057485	0.9924124
n=500,k=375					
p=0.1	1.0089521	1.0154682	1.063222	1.0153642	0.9834589
p=0.3	1.0345261	0.9821456	1.056485	1.0526522	0.9760453
p=0.5	1.0542136	1.0041057	1.007981	1.0946353	1.0215468
p=0.7	1.0625432	1.0732187	1.050623	1.0457638	1.0574523
p=0.9	1.0842164	1.0951249	1.002898	1.0090175	1.0592213
n=1000,k=750					
p=0.1	1.0487556	1.021031	1.0315264	1.0124501	1.0009012
p=0.3	1.0452677	1.002984	0.9755863	1.0447552	1.0256887
p=0.5	1.0123545	1.054726	1.0165217	1.0689157	1.0531254
p=0.7	1.0448791	1.042459	1.0017861	1.0364528	1.0674522
p=0.9	1.0623548	1.018120	0.9886742	1.0138762	1.0397856

of the mean square error small. It is believed that when $k \geq [n/3]$ the sub-sample size after selecting a certain value, the corresponding changes in the mean square error is not much different. Therefore, we generally can select a sub-sample size. Numerical simulation results obtained with the above [9] $p = 0.5$ was proposed, to take $k = [n/3]$, when $k = [3n/4]$ to take the point of view of other circumstances consistent.

In view of the above analysis, the sub-sample capacity behind the choice of numerical analysis $k = [3n/4]$, it is worth noting that when making the choice of k, $[(k+1)p]$ becomes zero, you should choose a bigger k, for example, when $n = 100$, when $p = 0.1$, k only take 99.

3. KL QUANTILE ESTIMATE ANALOG COM-PA-RATORS WITH SQL QUANTILE ESTI-MATES

Here we are two cases to the median estimate of K and SQ sample quantile estimate the mean square error of the numerical simulation comparison, steps taken in the simulation are as follows:

Step 1: Let $X_1, X_2 \cdots X_n$ are n samples were drawn at random from the whole of X;

Step 2: For a given level of probability p, $SQ(p) = X_{[(n+1)p]}$ $KL(p) = K_{[(k+1)p]:k:n}$ were calculated;

Step 3: 1,2 procedure was repeated $m = 5000$ to give an estimated value $SQ_i(p) = X_{[(n+1)p]}$

$KL_i(p) = K_{[(k+1)p]k:n}$, wherein $i = 1, 2, \cdots m$;

Step 4: SQ sample were calculated and KL quantile quantile estimate the mean square error:

$$MSE_{SQ}(p) = \frac{1}{m}\sum_{i=1}^{m}(SQ_i(p) - Q(p))^2 \qquad (4)$$

$$MSE_{KL}(p) = \frac{1}{m}\sum_{i=1}^{m}(KL_i(p) - Q(p))^2 \qquad (5)$$

Where $Q(p)$ is the true value of the overall simulation X quantile p;

Step 5: calculation error ratio $\rho = \dfrac{MSE_{SQ}}{MSE_{KL}}$ of both parties,

if $\rho > 1$, that is $MSE_{SQ} > MSE_{KL}$,

Indicating that KL quantile estimate is better than SQ sample quantile estimates. If $\rho < 1$, that is $MSE_{SQ} < MSE_{KL}$, indicating that KL is estimated that more than SQ sample quantile quantile estimated to be poor.

In the first case, the overall distribution of choice were normal $N(0,1)$, uniform $R(0,1)$, t distribution $(df = 4)$, exponential distribution $E(1)$, probability level p=0.01,0.02, 0.03,0.04,0.05, the corresponding value calculated according to the above steps, the calculation results shown in Table **2**.

The above table gives the small probability level ($p \le 0.05$) under different simulated data distribution ratio of the mean square error, these data show that:

(1) On the whole, in a small probability level KL quantile estimates in most cases the mean square error is less than SQ sample quantile estimate the mean square error. Therefore, we can think in general, KL quantile comparison SQ sample quantile estimate is superior.

(2) From the local point of view, in uniform and exponential distribution, KL SQ quantile estimate was significantly better than the median estimate of sample points, and the standard normal distribution, Points in the distribution and logistic, KL quantile estimates obvious advantages, and SQ sample quantile estimate considerably.

The second case, select the overall distribution were normal, uniform distribution, exponential distribution, the probability level = 0.1, 0.3, 0.5, 0.7, 0.9, calculate the corresponding values follow the above steps, the results shown in Table **3**.

The above table gives the small probability level ($p \le 0.05$) under different simulated data distribution ratio of the mean square error, these data show that:

(1) The vast majority of cases, K quantile estimate of the mean square error is significantly less than SQ sample quantile estimate the mean square error. Therefore, KL quantile

estimate significantly better than the SQ meter sample quantile estimates.

(2) In the case of the t distribution, there will be circumstances mean square error is less than 1, indicating the heavy-tailed distributions have an impact on the estimated quantile KL fitting effect [10].

CONCLUSION

By the previous simulation data analysis, we estimate the effect of K quantile estimates have a more comprehensive understanding of the above, the following conclusions:

(1) In most cases, KL quantile estimate of the mean square error is less than SQ sample quantile estimation mean square error;

(2) In the truncated distribution (such as exponential distribution and uniform distribution) of the cut end of the estimated effect of KL quantile estimate is very good, and much better than the median estimate of SQ sample points;

(3) Heavy tail of the distribution (eg distribution) have an impact on the estimated effects of KL quantile estimates.

The authors confirm that this article content has no conflicts of interest.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

REFERENCES

[1]   Z. Wang, and G. Yang, "Kernel-Copula function based on VaR", *The Research on Investment*, vol. 12, no. 2, pp. 132-140, 2013.
[2]   E. L. Lehmann, Elements of Large-Sample Theory, Springer, 2010, pp. 85-96.
[3]   P. Liu, and Z. Tianjun, "VaR and ES adjusted Sharpe ratio and empirical research based on fund ratings" *Number of Technical Economics*, vol. 31, no. 4, pp. 735-750, 2012.
[4]   W. Hardle, and S. Sperlich, "*Marlen Nonparametric and Semiparametric Models*", Springer Verlag Berlin and Heidelberg Gmb H & Co. K, 2010, pp. 132-140.
[5]   W. D. Kaigh, and P. A. Lachenbruch, "A generalized quantile estimator", *Communicationsin Statistics-Theory and Methods*, vol. 11, no. 19, pp. 2217-2238, 1982.
[6]   S. M. Steinberg, and C. E. Davis, "Distribution-free confidence intervals for quantiles insmall samples," *Communications in Statistics-Theory and Methods,* vol. 14, no. 4, pp. 979-990, 1985.
[7]   R. Zielinski, "Small-sample quantile estimators in a large nonparametric model", *Communications in Statistics-Theory and Methods,* vol. 3, no. 7, pp. 1223-1241, 2006.

[8]   T. V. Ramanathan, "Parametric and non-parametric estimation of value-at-risk", *The Journal of Risk Model Validation*, vol. 3, no.1, pp. 51-71, 2009.

[9]   S. M. Steinberg, "*Confidence Intervlas for Function of Quantiles Using Linear Combination of Order Statistics*", The University Of North Carolina, Carolina, 1983.

[10]   Y. Guo, X. Liao, and Z. Peng, "Shanghai and Shenzhen stock thick tails and VaR estimate," *Southwest China Normal University Natural Science*, vol. 37, no. 5, pp. 102-106, 2012.