THE DATA-BASED CHOICE OF BANDWIDTH FOR KERNEL QUANTILE ESTIMATOR OF VAR

XIN YANG¹, SHAN CHAO YANG²
College of International Education¹
GuangXi University of Finance and Economics
Nanning, 530003, P. R. China

College of Mathematical Science²
Guangxi Normal University
Guilin, 541004, P. R. China

ABSTRACT. Value-at-Risk is an important risk measure, has been wildly applied in market practice and financial risk measurement. We use the smooth kernel estimator for the quantile proposed by Parzen (1979) as a VaR estimator and propose a data-based choice method of optimal bandwidth via normal reference distribution, which is easy to compute. The simulations show that the choice method of bandwidth is effective, and the smooth kernel estimator has better performance than the sample quantile estimator.

Keywords: VaR; Kernel Quantile Estimator; Choice of Bandwidth

1. Introduction. Suppose that \{X_i, i ≥ 1\} is a sequence of independent and identically distributed random variables. Denote the distribution function of \(X_i\) by \(F(x)\) and the p-quantile of the distribution function by \(Q(p) = F^{-1}(p) = \inf\{x : F(x) ≥ p\}\) for \(0 < p < 1\). Let the empirical distribution function of a sample \(X_1, X_2, \ldots, X_n\) be \(F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i ≤ x)\) and the empirical quantile function be \(Q_n(p) = F_n^{-1}(p) = \inf\{x : F_n(x) ≥ p\}\). Assume that \(X_{(1)}, X_{(2)}, \ldots, X_{(n)}\) are the order statistics of the sample. Then \(Q_n(p) = X_{([np]+1)}\), where \([x]\) is the integer part of \(x\). It is well-known that \(Q_n(p)\) is a sample estimator of the \(p\)-quantile \(Q(p)\).

The sample quantile is a very important statistics. Its applications are beyond the area of statistics. For example, the sample quantile is used to estimate Value-at-Risk (VaR), which is a risk measure and has been used widely in financial institutions for their risk management. Let \(X\) be a random variable describing the return of an asset or a portfolio. Given a positive value \(p ∈ (0,1)\), the \(1−p\) level VaR is \(VaR_p = −Q(p)\), which measures the maximum potential loss of a given asset or portfolio over a prescribed holding period at a given confidence level.

Early estimators of VaR are based on parametric models for the return distribution \(F\), for instance, Gaussian or t distributions, one can refer to Dowd (2000), Giot and Laurent (2003), Laurent and Peters (2002), and the references therein. The advantage of parametric methods is that it is easy to calculate, but actual distribution of return serials isn’t always
known. Non-parametric methods, without assuming statistical distribution, can deal with asymmetry and fat-tail problems of return serials effectively. This has caused wide attention among the world of financial management. Dowd (2001) used the sample quantile $-X_{(np)}$ to estimate VaR. Of course, $-X_{(np)}$ can be also used as a VaR estimator because it has the same asymptotical properties as $-X_{(np+1)}$. However, Berkowitz and O’Brien (2002) and Inui et al. (2005) reported that $-X_{(np)}$ tends to be conservative. In fact, the estimator $-X_{(np+1)}$ tends to be less than the real value of VaR while $-X_{(np)}$ tends to be more than the value of VaR. For other nonparametric methods, Gourieroux et al. (2000) first introduced VaR non-parametric kernel estimator. Chen and Tang (2005) proposed VaR non-parametric estimation based on kernel quantile estimator. Wei et al. (2010) used the kernel quantile estimator proposed by Parzen (1979) as a VaR estimator and studied its Bahadur representation, asymptotical mean square error under strongly mixing assumption. In this paper, we will consider the estimator. Let us to recall Parzen’s smooth kernel estimator for the quantile $Q(p)$ as follows

$$T_n(p) = \frac{1}{h_n} \int_0^1 Q_n(t) K \left( \frac{p-t}{h_n} \right) \, dt = \frac{1}{h_n} \sum_{i=1}^n X_i \int_{(i-1)/n}^{i/n} K \left( \frac{p-t}{h_n} \right) \, dt$$

where $h_n$ is a smoothing bandwidth and $K(\cdot)$ is a kernel function. The estimator is a linear combination of order statistics with kernel weights. Falk (1984) investigated the asymptotical mean square error of the estimator. Yang (1985) established a Bahadur representation in senses of probability and a mean squared convergence rate. Later, Sheather and Marron (1990) improved Falk’s asymptotical mean square error, gave an asymptotical optimal bandwidth, and studied the data-based choice method of the optimal bandwidth by using the estimators of $Q(p)$ and $Q'(p)$ as follows

$$\hat{Q}_n'(p) = \frac{1}{a_n^2} \sum_{i=1}^n X_i \int_{(i-1)/n}^{i/n} K'_* \left( \frac{p-t}{a_n} \right) \, dt$$

and

$$\hat{Q}_n''(p) = \frac{1}{b_n^2} \sum_{i=1}^n X_i \int_{(i-1)/n}^{i/n} K''_* \left( \frac{p-t}{b_n} \right) \, dt$$

where $K_*(\cdot)$ is a kernel function of order $m$ and symmetric about 0, and $a_n, b_n$ are also two bandwidths. At the same time, Sheather and Marron (1990) compared the performance of the kernel quantile estimator with the estimators of Harrel and Davis (1982) and Kaigh and Lachenbruch (1982) by using the samples of size 50 and 100 from the double-exponential, exponential, lognormal and normal distributions.

Under random censorship or randomly truncated data, Xiang (1995) and Zhou (2006) established a Bahadur representation of $T_n(p)$, the latter also gave the strong consistence, asymptotic normality of the estimator. For $\alpha$-mixing random variables, Wei et al. (2010) showed Bahadur representation in senses of almost surely convergence, strong consistence and mean square error. By simulations, they also reported that the mean absolute biases of
the estimator are much less than those of sample quantile for samples of size 100, 200, 500
and 1000 from normal distribution and student t distribution, whenever samples are
independent or dependent, and from light tail distribution or fat tail distribution. Recently,
Ajami et al. (2011) established the Bahadur-type representation of the kernel smooth
estimator under strong mixing and censored data, and showed that this estimator is strongly
consistent from the Bahadur representation. These studies mentioned above indicate that
the kernel estimator possesses many nice properties and its efficiency is better than that of
the sample quantile. Therefore, it is a reasonable estimator for the quantile.

It is well known that bandwidth selection for kernel estimator is a hard work. For the
estimator \( T_n(p) \), the data-based choice method of the asymptotical optimal bandwidth by
using the estimators (2) and (3) was studied in Sheather and Marron (1990). However,
because the estimators \( \hat{Q}_n(p) \) and \( \hat{Q}_n(p) \) are also kernel estimation, one further faces new
work to choose the bandwidths. To simplify the procedure, in this paper, we propose a
data-based choice method of bandwidth via normal reference distribution. Our simulation
results report that the choice method of bandwidth has good performance.

2. Mean Square Error and Optimal Bandwidth. Here let us recall the results about
asymptotical mean square error and asymptotical optimal bandwidth in Sheather and

**Theorem 1:** (Sheather and Marron, 1990, Theorem 1) Suppose that \( Q'' \) is continuous in a
neighborhood of \( p \) and that \( K \) is a density with compactly supported set \([-c, c]\),
symmetric about 0. Let \( \bar{K}(x) = \int_{-\infty}^{x} K(t) dt \). Then for all fixed \( p \in (0,1) \), apart from
\( p = 0.5 \) when \( F \) is symmetric,

\[
MSE[T_n(p)] = n^{-1} (Q'(p))^2 \left[ p(1-p) - 2h_n \int_{c}^{-c} xK(x)\bar{K}(x)dx \right] + \frac{1}{4} (Q''(p))^2 h_n^4 \left( \int_{c}^{-c} x^2 K(x)dx \right)^2 + o(n^{-1} h_n) + o(h_n^4)
\]

and for \( p = 0.5 \) when \( F \) is symmetric,

\[
MSE[T_n(0.5)] = n^{-1} (Q'(0.5))^2 \left[ 0.25 - h_n \int_{c}^{-c} xK(x)\bar{K}(x)dx + n^{-1} h_n^{-1} \int_{c}^{-c} K^2(x)dx \right] + o(n^{-1} h_n) + o(n^{-2} h_n^{-2})
\]

**Theorem 2:** (Sheather and Marron, 1990, Corollary 1) Supposed that the conditions given
in Theorem 1 hold. Then for all fixed \( p \in (0,1) \), apart from \( p = 0.5 \) when \( F \) is
symmetric, the asymptotically optimal bandwidth is given by

\[
h_{opt} = \left[ \frac{2(Q'(p))^2 \int_{c}^{-c} xK(x)\bar{K}(x)dx}{n(Q''(p))^2 \left( \int_{c}^{-c} x^2 K(x)dx \right)^2} \right]^{\frac{1}{3}}.
\]

And when \( h_n = h_{opt} \),
Supposed that the distribution function $F(x)$ has a density function $f(x)$. Then

$$Q'(p) = \frac{1}{f(Q(p))}, \quad Q''(p) = -\frac{f'(Q(p))}{f^2(Q(p))}. \quad (8)$$

Hence

$$\frac{Q'(p)}{Q''(p)} = -\frac{f^2(Q(p))}{f'(Q(p))}. \quad (9)$$

Because $F$ is symmetric, so $f(x)$ has a maximal value at $x = 0$, which implies that $f'(Q(0.5)) = 0$. From (9), $Q'(p)/Q''(p)$ will be infinity at $p = 0.5$. Therefore, these theorems require that $p$ is apart from 0.5.


From Theorem 2, we see that the asymptotical optimal bandwidth $h_{opt}$ depends on the first and second derivatives of the quantile function, and also on the kernel function. Here, we will use normal distribution as the population distribution to estimate the the quantile function $Q(p)$, while quadratic function as kernel function.

#### 3.1. Quadratic Kernel Function and Weight Function.

Let the quadratic kernel function

$$K(u) = \frac{3}{4}(1 - u^2)I_{[-1,1]}(u). \quad (10)$$

Then

$$\overline{K}(x) = \int_{x}^{1} K(u)du = \int_{-1}^{x} (1 - u^2)du = \frac{3}{4}(x - \frac{1}{3}x^3 + \frac{1}{3}) \quad \text{and}$$

$$\int_{-1}^{1} x^2 K(x)dx = 1/5, \quad \int_{-1}^{1} xK(x)\overline{K}(x)dx = \frac{9}{70}. \quad (11)$$

Denote the weight

$$\omega_{n,i}(p) = \int_{(i-1)/n}^{i/n} K((p-t)/h_n)dt. \quad \text{Let } a_{i,j} = \max\{i/n, p-h_n\} \text{ and } a_{2,i} = \min\{i/n, p+h_n\}.$$ 

Then

$$w_{n,i}(p) = \frac{3}{4} \int_{(i-1)/n}^{i/n} \left(1 - \frac{(p-t)^2}{h_n^2}\right) I(-1 < (p-t)/h_n < 1)dt$$

$$= \frac{3}{4h_n^2} \int_{(i-1)/n}^{i/n} \left(h_n^2 - (p-t)^2\right) I(p-h_n < t < p+h_n)dt$$

$$= \frac{3}{4h_n^2} \int_{a_{1,i}}^{a_{2,i}} \left(h_n^2 - (p-t)^2\right) I(p-h_n < t < p+h_n)dt. \quad (12)$$

It is clear that $\omega_{n,i}(p) = 0$ when $a_{i,j} \geq a_{2,i}$. And when $a_{i,j} < a_{2,i}$ we have
where \( g(x) = h_n^2 x + \frac{1}{4} (p - x)^3 \).

3.2. Optimal Bandwidth Based on Normal Distribution. We use normal distribution as a reference distribution of the population distribution. That is to assume that \( F(x) \sim N(\mu, \sigma^2) \). In the case, \( f'(x) = \frac{\mu x^2}{\sigma^2} f(x) \). From (9),

\[
\frac{Q'(p)}{Q''(p)} = -\frac{\sigma^2 f^2(Q(p))}{(\mu - Q(p)) f(Q(p))} = -\frac{\sigma^2 f(Q(p))}{\mu - Q(p)}.
\]

Thus, the optimal bandwidth

\[
h_{opt,N} = \sqrt[3]{\frac{\sigma^2 f^2(Q(p))}{(\mu - Q(p)) f(Q(p))}} = \sqrt[3]{\frac{45\sigma^4 f^2(Q(p))}{7n(\mu - Q(p))^2}}.
\]

Therefore, we have the estimator of the optimal bandwidth

\[
\hat{h}_{opt,N} = \sqrt[3]{\frac{45\sigma^4 f^2(\hat{Q}(p))}{7n(\overline{X} - \hat{Q}(p))^2}}.
\]

where \( X \) and \( S_n \) are sample mean and sample standard deviation, respectively, and \( \hat{Q}(p) \) is the quantile of \( N(\overline{X}, S_n^2) \).

4. Simulations. In this section, some numerical simulations are carried out to evaluate the performance of the kernel quantile estimator (1) by using the data-based bandwidth via normal reference distribution (16), and to compare it with the sample quantile estimator.

We know that normal distribution is light tail distribution, while student t distribution is fat tail distribution. So we consider the samples that are from normal distribution with standard deviations 1 and 3, student t distribution with degrees of freedom 3 and 5. The sample sizes are 200, 500, 800 and 1000, and repeated 1000 times for each one. The probability level \( p \) is from 0.01 to 0.4 by step 0.01, i.e. \( p_i = 0.01 \times i \) for \( i = 1, 2, \cdots, m \), where \( m = 40 \).
FIGURE 1. The estimation of Quantile for standard normal population.

For standard normal population, Figure 1 reports the kernel quantile estimator $T_n(p)$ (dashed line), the sample quantile estimator $Q_n(p)$ (longdash line) and the quantile $Q(p)$ (solid line) for four cases of the sample sizes 200, 500, 800 and 1000. The figures illustrate that the estimator values are near to the real values, and they are getting better as sample size $n$ grows. Moreover, the kernel quantile estimator is more smooth and closer to the real value than the sample quantile estimator. For other populations, there are similar figures. Due to the space, we don’t show the figures.

Now we observe the relatively absolute biases of the estimators. Let
\[
RT_i = \frac{|T_n(p_i) - Q(p_i)|}{Q(p_i)} \times 100\%
\]
for $i = 1, 2, \cdots, m$, and $\overline{RT} = m^{-1} \sum_{i=1}^{m} RT_i$, which is the mean of the relatively absolute biases over $p$ for the kernel quantile estimator. Analogously let
\[
RQ_i = \frac{|Q_n(p_i) - Q(p_i)|}{Q(p_i)} \times 100\%
\]
for $i = 1, 2, \cdots, m$, and $\overline{RQ} = m^{-1} \sum_{i=1}^{m} RQ_i$. It is the mean of the relatively absolute biases over $p$ for the sample quantile estimator.

The the means of the relatively absolute biases for each population and each sample size are listed in Table 1. These show that the means of the relatively absolute biases are less than 5% for most cases, whether light or fat tail distribution. And the means of the
relatively absolute biases of the kernel quantile estimator are less than those of the sample quantile estimator. These results show that the data-based bandwidth via normal reference distribution (16) is effective.

**TABLE 1: Mean of the relatively absolute biases (%)**

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample size</th>
<th>$\sigma = 1$</th>
<th>$\sigma = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_n(p)$</td>
<td>$X_{[np]}+1$</td>
<td>$T_n(p)$</td>
</tr>
<tr>
<td>Normal distribution</td>
<td>200</td>
<td>4.135933</td>
<td>7.644296</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>3.882401</td>
<td>5.311367</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>2.923828</td>
<td>3.953348</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.946372</td>
<td>2.433437</td>
</tr>
<tr>
<td>$\nu$ distribution</td>
<td>$df = 3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T_n(p)$</td>
<td>$X_{[np]}+1$</td>
<td>$T_n(p)$</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>6.792439</td>
<td>7.050439</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>4.024788</td>
<td>4.714794</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>2.302889</td>
<td>4.293160</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.167786</td>
<td>3.893698</td>
</tr>
<tr>
<td></td>
<td>$df = 5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T_n(p)$</td>
<td>$X_{[np]}+1$</td>
<td>$T_n(p)$</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>6.792439</td>
<td>7.050439</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>4.024788</td>
<td>4.714794</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>2.302889</td>
<td>4.293160</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>2.167786</td>
<td>3.893698</td>
</tr>
</tbody>
</table>

5. **Conclusion.** Value-at-Risk, a very important risk measure, has been wildly applied in market practice and financial risk measurement. And estimating VaR has received a considerable amount of attention in literature. In this paper, we consider the smooth kernel estimator for the quantile proposed by Parzen (1979) as a VaR estimator. In view of the complexity to choose bandwidth for the kernel estimator, we propose a data-based choice method of optimal bandwidth via normal reference distribution. The choice method is easy to compute and the simulations show that it is effective.

**Acknowledgment.** This project is supported by the National Science Foundation of China(11061007), the Social Science Foundation of China(10CTJ004) and the National Science Foundation of Guanxi(2011GXNSFA018133).

**REFERENCES**


