Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 138450, 18 pages
doi:10.1155/2012/138450

Research Article

New Bandwidth Selection for Kernel Quantile Estimators

Ali Al-Kenani and Keming Yu

Department of Mathematical Sciences, Brunel University, Uxbridge UBB 3PH, UK

Correspondence should be addressed to Ali Al-Kenani, mapgaja@brunel.ac.uk

Received 8 August 2011; Revised 26 September 2011; Accepted 10 October 2011

Academic Editor: Junbin B. Gao

We propose a cross-validation method suitable for smoothing of kernel quantile estimators. In particular, our proposed method selects the bandwidth parameter, which is known to play a crucial role in kernel smoothing, based on unbiased estimation of a mean integrated squared error curve of which the minimising value determines an optimal bandwidth. This method is shown to lead to asymptotically optimal bandwidth choice and we also provide some general theory on the performance of optimal, data-based methods of bandwidth choice. The numerical performances of the proposed methods are compared in simulations, and the new bandwidth selection is demonstrated to work very well.

1. Introduction

The estimation of population quantiles is of great interest when one is not prepared to assume a parametric form for the underlying distribution. In addition, due to their robust nature, quantiles often arise as natural quantities to estimate when the underlying distribution is skewed [1]. Similarly, quantiles often arise in statistical inference as the limits of confidence interval of an unknown quantity.

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random sample drawn from an absolutely continuous distribution function F with density f. Further, let $X_{(1)} \leq X_{(2)} \cdots \leq X_n$ denote the corresponding order statistics. For $(0 < p < 1)$ a quantile function $Q(p)$ is defined as follows:

$$Q(p) = \inf\{x : F(x) \geq p\}. \tag{1.1}$$

If $\hat{Q}(p)$ denotes pth sample quantile, then $\hat{Q}(p) = x_{([np]+1)}$ where $[np]$ denotes the integral part of np. Because of the variability of individual order statistics, the sample quantiles suffer from lack of efficiency. In order to reduce this variability, different approaches of estimating sample quantiles through weighted order statistics have been proposed. A popular class of these estimators is called kernel quantile estimators. Parzen [2] proposed a version of the kernel quantile estimator as below:

$$\tilde{Q}_K(p) = \sum_{i=1}^{n} \left[\int_{i-1/n}^{i/n} K_h(t-p)\,dt \right] X_{(i)}. \tag{1.2}$$

From (1.2) one can readily observe that $\tilde{Q}_K(p)$ puts most weight on the order statistics $X_{(i)}$, for which i/n is close to p. In practice, the following approximation to $\tilde{Q}_K(p)$ is often used:

$$\tilde{Q}_{AK}(p) = \sum_{i=1}^{n} \left[n^{-1} K_h(i/n - p) \right] X_{(i)}. \tag{1.3}$$

Yang [3] proved that $\tilde{Q}_K(p)$ and $\tilde{Q}_{AK}(p)$ are asymptotically equivalent in terms of mean square errors. Similarly, Falk [4] demonstrates that, from a relative deficiency perspective, the asymptotic performance of $\tilde{Q}_{AK}(p)$ is better than that of the empirical sample quantile.

In this paper, we propose a cross-validation method suitable for smoothing of kernel quantile estimators. In particular, our proposed method selects the bandwidth parameter, which is known to play a crucial role in kernel smoothing, based on unbiased estimation of a mean integrated squared error curve of which the minimising value determines an optimal bandwidth. This method is shown to lead to asymptotically optimal bandwidth choice and we also provide some general theory on the performance of optimal, data-based methods of bandwidth choice. The numerical performances of the proposed methods are compared in simulations, and the new bandwidth selection is demonstrated to work very well.

2. Data-Based Selection of the Bandwidth

Bandwidth plays a critical role in the implementation of practical estimation. Specifically, the choice of the smoothing parameter determines the tradeoff between the amount of smoothness obtained and closeness of the estimation to the true distribution [5]

Several data-based methods can be made to find the asymptotically optimal bandwidth h in kernel quantile estimators for $\tilde{Q}_{AK}(p)$ given by (1.3). One of these methods use derivatives of the quantile density for $\tilde{Q}_{AK}(p)$.

Building on Falk [4], Sheather and Marron [1] gave the MSE of $\tilde{Q}_{AK}(p)$ as follows. If f is not symmetric or f is symmetric but $p \neq 0.5$,

$$\text{AMSE}\left(\tilde{Q}_{AK}(p)\right) = \frac{1}{4}\mu_2(k)^2 \left[Q''(p)\right]^2 h^4 + p(1-p)\left[Q'(p)\right]^2 n^{-1} - R(K)\left[Q'(p)\right]^2 n^{-1} h, \tag{2.1}$$

where $R(K) = 2\int_{-\infty}^{\infty} uK(u)K^{-1}(u)du$, $\mu_2(k) = \int_{-\infty}^{\infty} u^2K(u)du$ and K^{-1} is the antiderivative of K.

If $Q' > 0$ then

$$h_{\text{opt}} = \alpha(K) \cdot \beta(Q) \cdot n^{-1/3}, \tag{2.2}$$

where $\alpha(K) = [R(K)/\mu_2(k)^2]^{1/3}$, $\beta(Q) = [Q'(p)/Q''(p)]^{2/3}$.

There is no single optimal bandwidth minimizing the AMSE$(\widetilde{Q}_{AK}(p))$ when F is symmetric and $p = 0.5$. Also, If $q = 0$, we need higher terms and the AMSE$(\widetilde{Q}_{AK}(p))$ can be shown to be

$$\text{AMSE}\left(\widetilde{Q}_{AK}(p)\right) = \left(\frac{1}{4} - \frac{1}{n}\right)h^4\left[Q''(p)\right]^2\mu_2(k)^2 + 2n^{-1}h^2\left[Q''(p)\right]^2 \int (q - ht)tK(t)j(t)dt, \tag{2.3}$$

where $j(t) = \int_{-\infty}^{t} xK(x)dx$, see Cheng and Sun [6].

In order to obtain h_{opt} we need to estimate $Q' = q$ and $Q'' = q'$. It follows from (1.3) that the estimator of $Q' = q$ can be constructed as follows:

$$\widetilde{q}_{AK}(p) = \widetilde{Q}'_{AK}(p) = \sum_{i=1}^{n} X_{(i)}\left[K_a\left(\frac{(i-1)}{n} - p\right) - K_a\left(\frac{i}{n} - p\right)\right]. \tag{2.4}$$

Jones [7] derived that the AMSE$(\widetilde{q}_{AK}(p))$ as

$$\text{AMSE}(\widetilde{q}_{AK}(p)) = \frac{a^4}{4}\mu_2(k)^2\left[q''(p)\right]^2 + \frac{1}{na}\left[q(p)\right]^2 \int K^2(y)dy. \tag{2.5}$$

By minimizing (2.5), we obtain the asymptotically optimal bandwidth for $\widetilde{Q}'_{AK}(p)$:

$$a^*_{\text{opt}} = \left[\frac{\left[Q'(p)\right]^2 \int K^2(y)dy}{n\left[Q'''(p)\right]^2\mu_2(k)^2}\right]^{1/5}. \tag{2.6}$$

To estimate $Q'' = q'$ in (2.2), we employ the known result

$$\widetilde{Q}''_{AK}(p) = \frac{d}{dp}\widetilde{Q}'_{AK}(p) = \frac{1}{a^2}\sum_{i=1}^{n} X_{(i)}\left[K'\left(\frac{(i-1)/n - p}{a}\right) - K'\left(\frac{i/n - p}{a}\right)\right], \tag{2.7}$$

and it readily follows that

$$a^{**}_{\text{opt}} = \left[\frac{3\left[Q'(p)\right]^2 \int K'^2(x)dx}{n\left[Q'''''(p)\right]^2 \mu_2(k)^2} \right]^{1/7} \tag{2.8}$$

which represents the asymptotically optimal bandwidth for $\widetilde{Q}''_{AK}(p)$. By substituting $a = a^*_{\text{opt}}$ in (2.4) and $a = a^{**}_{\text{opt}}$ in (2.7) we can compute h_{opt}.

3. Cross-Valdation Bandwidth Selection

When measuring the closeness of an estimated and true function the mean integrated squared (MISE) defined as

$$\text{MISE}(h) = E\int_0^1 \left\{\widetilde{Q}(p) - Q(p)\right\}^2 dp \tag{3.1}$$

is commonly used as a global measure of performance.

The value which minimises MISE(h) is the optimal smoothing parameter, and it is unknown in practice. The following ASE(h) is the discrete form of error criterion approximating MISE(h):

$$\text{ASE}(h) = \frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{Q}\left(\frac{i}{n}\right) - Q\left(\frac{i}{n}\right)\right\}^2. \tag{3.2}$$

The unknown $Q(p)$ is replaced by $\widehat{Q}(p)$ and a function of cross-validatory procedure is created as:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{Q}_{-i}\left(\frac{i}{n}\right) - \widehat{Q}\left(\frac{i}{n}\right)\right\}^2, \tag{3.3}$$

where $\widetilde{Q}_{-i}(i/n)$ denotes the kernel estimator evaluated at observation x_i, but constructed from the data with observation x_i omitted.

The general approach of crossvalidation is to compare each observation with a value predicted by the model based on the remainder of the data. A method for density estimation was proposed by Rudemo [8] and Bowman [9]. This method can be viewed as representing each observation by a Dirac delta function $\delta(x - x_i)$, whose expectation is $f(x)$, and contrasting this with a density estimate based on the remainder of the data. In the context of distribution functions, a natural characterisation of each observation is by the indicator function $I(x - x_i)$ whose expectation is $F(x)$. This implies that the kernel method for density estimation can be expressed as

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i), \tag{3.4}$$

when $h \to 0$ $K_h(x - x_i) \to \delta(x - x_i)$.

The kernel method for distribution function

$$\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - x_i}{h}\right), \tag{3.5}$$

where W is a distribution function, h is the bandwidth controls the degree of smoothing. When $h \to 0$

$$W\left(\frac{x - x_i}{h}\right) \longrightarrow I(x - x_i), \tag{3.6}$$

where $I(x - x_i)$ is the indicator function

$$I(x - x_i) = \begin{cases} 1, & \text{if } x - x_i \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

Now, from (1.3) when $h \to 0$

$$\tilde{Q}_{\text{AK}}(p) \longrightarrow \delta\left(\frac{i}{n} - p\right) \text{X}_{(i)}, \tag{3.8}$$

and thus a cross-validation function can be written as

$$\text{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \int_0^1 \left\{ \delta\left(\frac{i}{n} - p\right) \text{X}_{(i)} - \tilde{Q}_{-i}\left(\frac{i}{n}\right) \right\}^2 dp. \tag{3.9}$$

The smoothing parameter h is then chosen to minimise this function. By subtracting a term that characterise the performance of the true (p) we have

$$H(h) = \text{CV}(h) - \frac{1}{n}\sum_{i=1}^{n} \int_0^1 \left\{ \delta\left(\frac{i}{n} - p\right) \text{X}_{(i)} - Q\left(\frac{i}{n}\right) \right\}^2 dp \tag{3.10}$$

which does not involve h. By expanding the braces and taking expectation, we obtain

$$H(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1\left\{\tilde{Q}_{-i}^2\left(\frac{i}{n}\right) - 2\delta\left(\frac{i}{n}-p\right)X_{(i)}\tilde{Q}_{-i}\left(\frac{i}{n}\right) + 2\delta\left(\frac{i}{n}-p\right)X_{(i)}Q\left(\frac{i}{n}\right) - Q^2\left(\frac{i}{n}\right)\right\}dp. \tag{3.11}$$

When $n \to \infty$ the (np)th order statistic $x_{(np)}$ is asymptotically normally distributed

$$\begin{aligned}
x_{(np)} &\sim \mathrm{AN}\left(Q(p), \frac{p(1-p)}{n[f(Q(p))]^2}\right),\\
E\{H(h)\} &= E\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^1\left\{\tilde{Q}_{-i}^2\left(\frac{i}{n}\right) - 2\delta\left(\frac{i}{n}-p\right)X_{(i)}\,\tilde{Q}_{-i}\left(\frac{i}{n}\right) + 2\delta\left(\frac{i}{n}-p\right)X_{(i)}\,Q\left(\frac{i}{n}\right)\right.\right.\\
&\qquad\left.\left. - Q^2\left(\frac{i}{n}\right)\right\}dp\right],\\
E\{H(h)\} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1\left[E\left\{\tilde{Q}_{-i}^2\left(\frac{i}{n}\right)\right\} - 2\delta\left(\frac{i}{n}-p\right)Q\left(\frac{i}{n}\right)E\left\{\tilde{Q}_{-i}\left(\frac{i}{n}\right)\right\} + 2\delta\left(\frac{i}{n}-p\right)Q^2\left(\frac{i}{n}\right)\right.\\
&\qquad\left. - Q^2\left(\frac{i}{n}\right)\right]dp,\\
E\{H(h)\} &= E\int_0^1\left\{\tilde{Q}_{n-1}\left(\frac{i}{n}\right) - Q\left(\frac{i}{n}\right)\right\}^2 dp,
\end{aligned} \tag{3.12}$$

where the notation $\tilde{Q}_{n-1}(i/n)$ with positive subscript denotes a kernel estimator based on a sample size of $n-1$. The proceeding arguments demonstrate that $\mathrm{CV}(h)$ provides an asymptotic unbiased estimator of the true $\mathrm{MISE}(h)$ curve for a sample size $n-1$. The identity at (3.12) strongly suggests that crossvalidation should perform well.

4. Theoretical Properties

From (3.1), we can write $\mathrm{MISE}(h) = \int_0^1 \mathrm{bias}^2(\tilde{Q}_K(p))dp + \int_0^1 \mathrm{var}(\tilde{Q}_K(p))dp$.

Sheather and Marron [1] have shown that

$$\mathrm{bias}\left(\tilde{Q}_K(p)\right) = \frac{1}{2}\,h^2\mu_2(k)Q''(p) + 0\left(h^2\right). \tag{4.1}$$

while Falk [4, page 263] proved that

$$\mathrm{var}\left(\tilde{Q}_K(p)\right) = p(1-p)\left[Q'(p)\right]^2 n^{-1} - R(K)\left[Q'(p)\right]^2 n^{-1}h + 0\left(n^{-1}h\right). \tag{4.2}$$

On combining the expressions for bias and variance we can express the mean integrated square error as

$$\begin{aligned}\text{MISE}(h) = \frac{1}{4}h^4\mu_2(k)^2\int_0^1 \left[Q''(p)\right]^2 dp + p(1-p)\int_0^1 \left[Q'(p)\right]^2 dp\, n^{-1} \\ - R(K)\int_0^1 \left[Q'(p)\right]^2 dp\ n^{-1}h + 0\left(h^4 + n^{-1}h\right),\end{aligned} \tag{4.3}$$

and for $C_1 = p(1-p)\int_0^1 [Q'(p)]^2 dp$, $C_2 = R(K)\int_0^1 [Q'(p)]^2 dp$ and $C_3 = \mu_2(k)^2\int_0^1 [Q''(p)]^2 dp$ the MISE can be expressed as

$$\text{MISE}(h) = C_1 n^{-1} - C_2 n^{-1}h + \frac{1}{4}C_3 h^4 + 0\left(h^4 + n^{-1}h\right). \tag{4.4}$$

Therefore, the asymptotically optimal bandwidth is $h_0 = Cn^{-1/3}$, where $C = \{C_2/C_3\}^{1/3}$.

We can see from (3.12) that $H(h)$ may be a good approximation to MISE(h) or at least to that function evaluated for a sample of size $n-1$ rather than n. Additionally, this is true if we adjusted $H(h)$ by adding the quantity

$$J_n = \int_0^1 \left\{ \left(\widehat{Q}(p) - Q(p)\right)^2 - E\left(\widehat{Q}(p) - Q(p)\right)^2 \right\}. \tag{4.5}$$

This quantity is demean and does not depend on h which makes it attractive for obtaining a particularly good approximation to MISE(h).

Theorem 4.1. *Suppose that $Q(p)$ is bounded on $[0,1]$ and right continuous at the point 0, and that K is a compactly supported density and symmetric about 0. Then, for each $\delta, \varepsilon, C > 0$,*

$$H(h) + J = \text{MISE}(h) + 0_2\left\{ \left(n^{-3/2} + n^{-1}h^{3/2} + n^{-1/2}h^3\right) n^{\delta} \right\} \tag{4.6}$$

with probability 1, uniformly in $0 \le h \le Cn^{\delta}$, as $n \to \infty$.

(An outline proof of the above theorem is in the appendix).

From the above theorem, we can conclude that minimisation of $H(h)$ produces a bandwidth that is asymptotically equivalent to the bandwidth h_0 that minimises MISE(h).

Corollary 4.2. *Suppose that the conditions of previous theorem hold. If $\widehat{h}$ denotes the bandwidth that minimises $CV(h)$ in the range $0 \le h \le Cn^{\delta}$, for any $C > 0$ and any $0 \le \varepsilon \le 1/3$, then*

$$\frac{\widehat{h}}{h_0} \longrightarrow 1 \tag{4.7}$$

with probability 1 as $n \to \infty$.

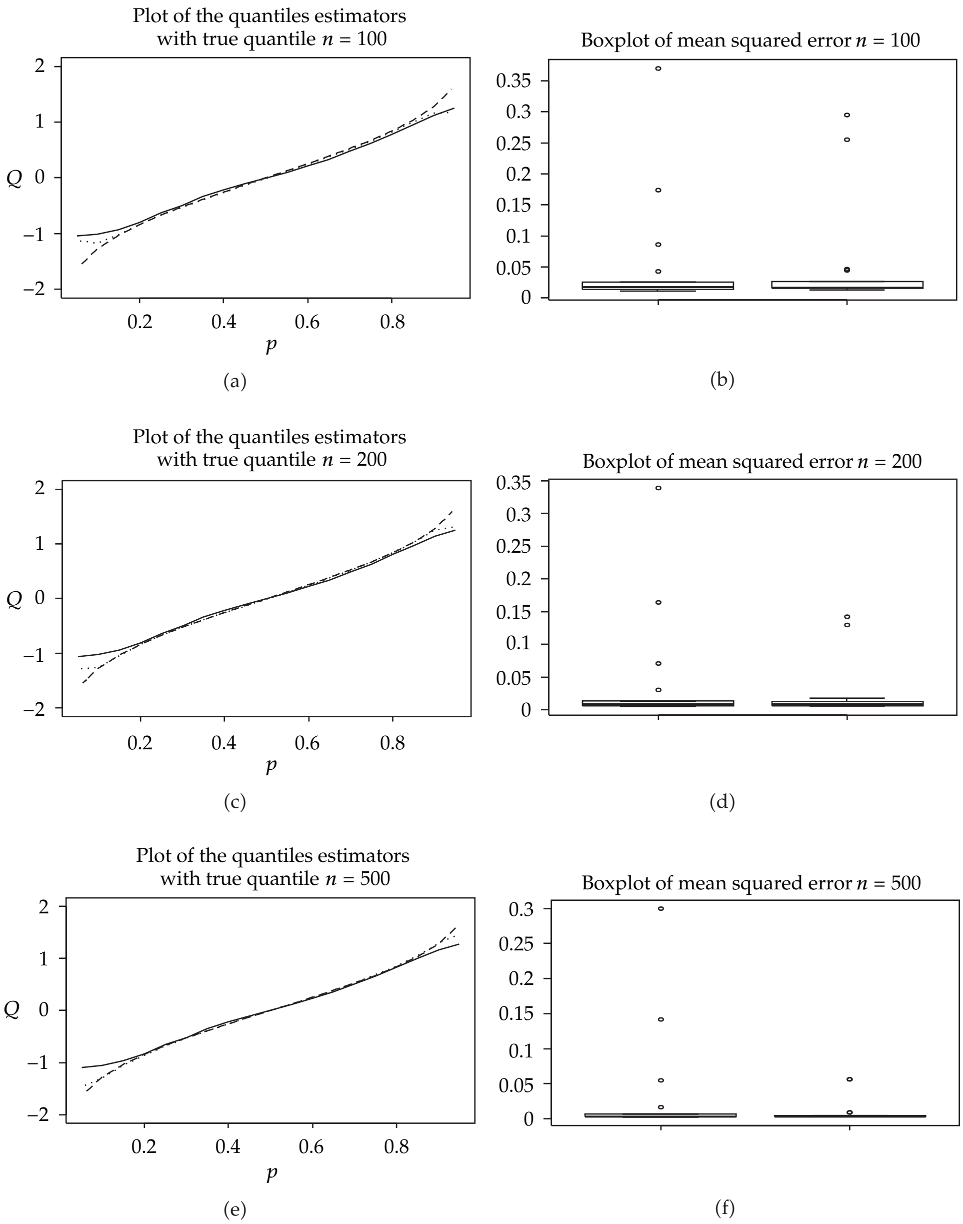

Figure 1: Left panel: plots of the quantile estimators for method 1 (solid line), method 2 (dotted line), and true quantile (dashed line) for different sample sizes and for data from a normal distribution. Right panel: box plots of mean squared errors for the quantile estimators for method 1 and method 2 for different sample sizes.

5. A Simulation Study

A numerical study was conducted to compare the performances of the two bandwidth selection methods. Namely, the method presented by Sheather and Marron [1] and our proposed method.

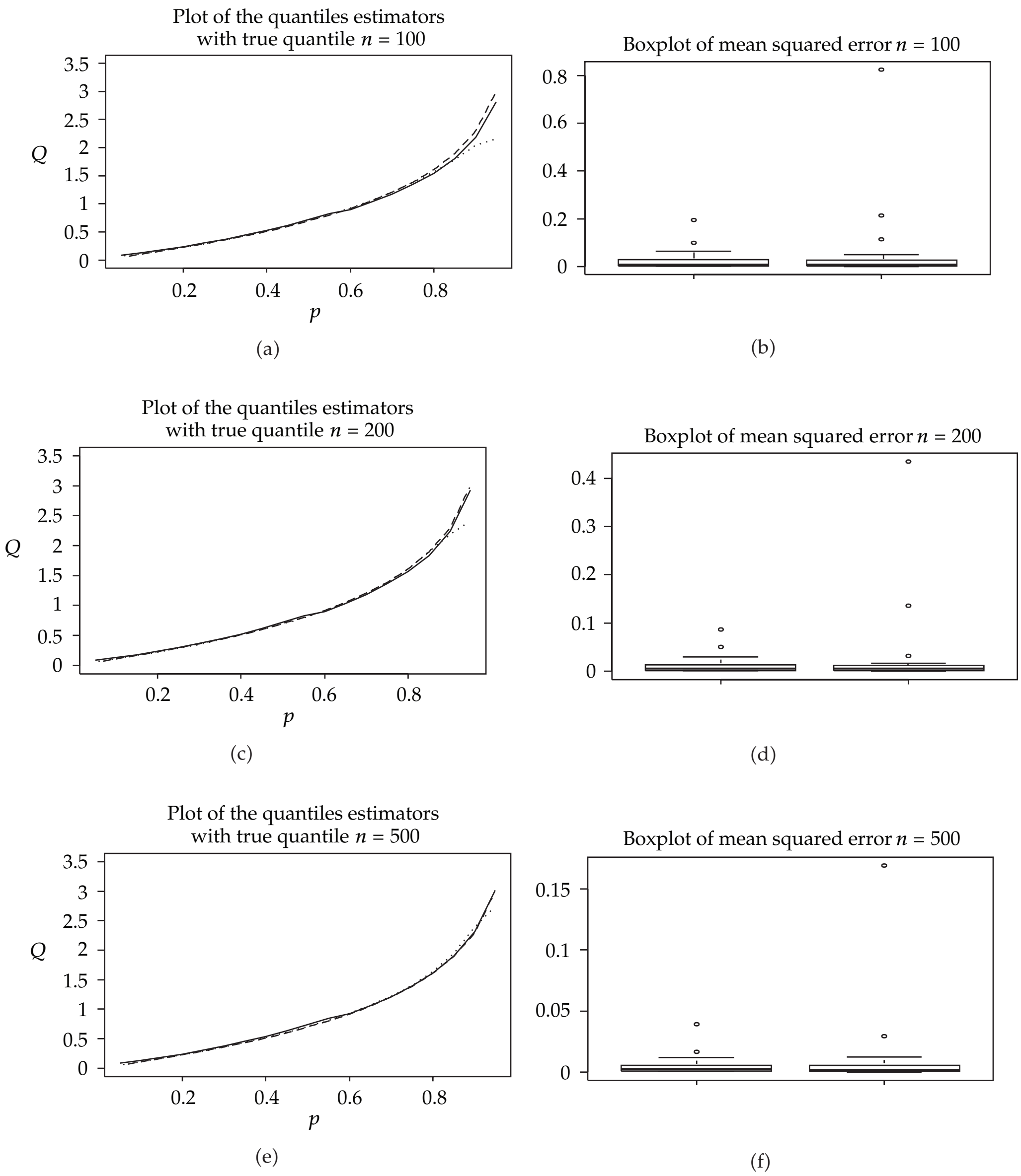

Figure 2: Left panel: plots of the quantile estimators for method 1 (solid line), method 2 (dotted line) and true quantile (dashed line) for different sample sizes and for data from an exponential distribution. Right panel: box plots of mean squared errors for the quantile estimators for method 1 and method 2 for different sample sizes.

In order to account for different shapes for our simulation study we consider a standard normal, Exp(1), Log-normal(0,1) and double exponential distributions and we calculate 18 quantiles ranging from $p = 0.05$ to $p = 0.95$. Through the numerical study the Gaussian kernel was used as the kernel function. Sample sizes of 100, 200 and 500 were used, with 100 simulations in each case. The performance of the methods was assessed through the

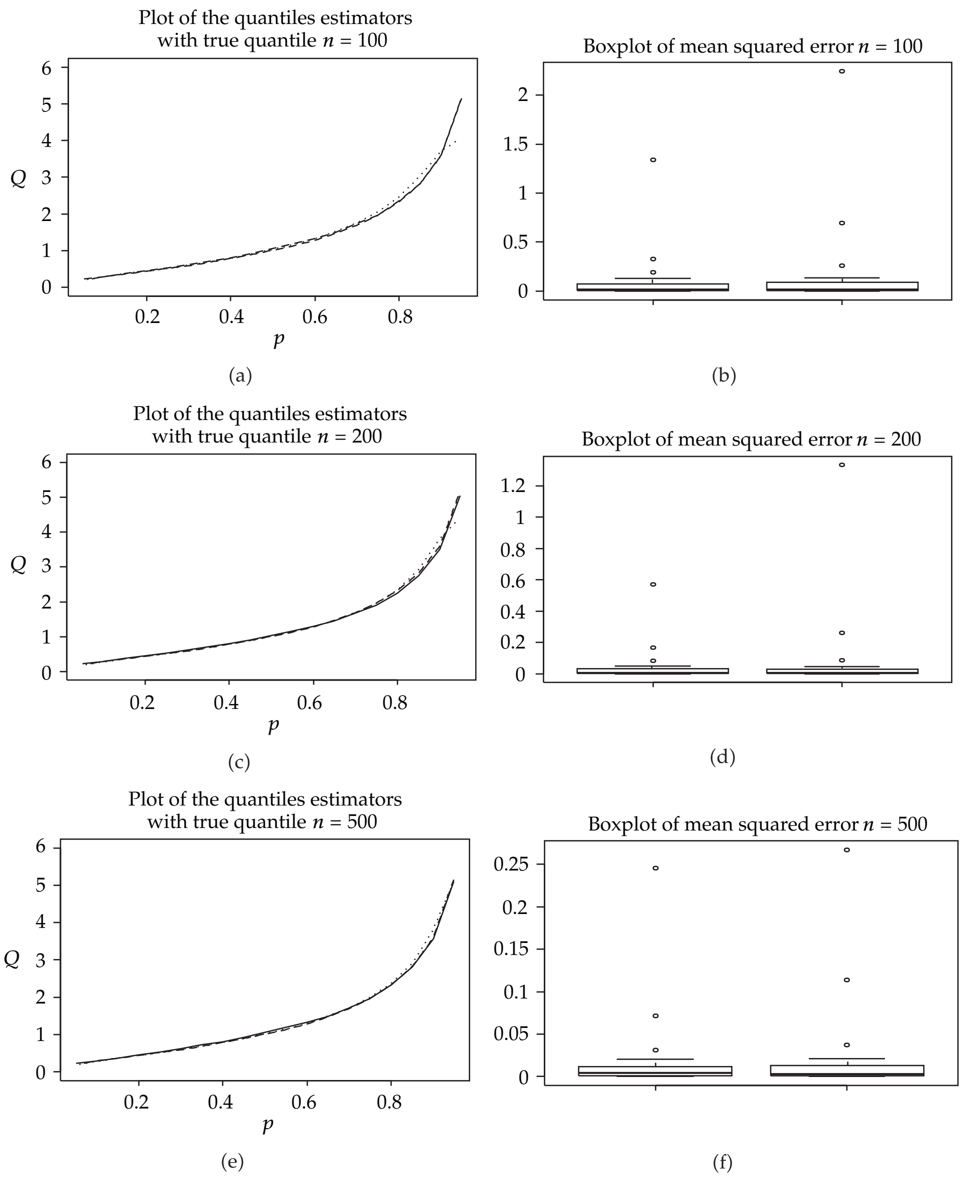

Figure 3: Left panel: plots of the quantile estimators for method 1 (solid line), method 2 (dotted line) and true quantile (dashed line) for different sample sizes and for data from a Log-normal distribution. Right panel: box plots of mean squared errors for the quantile estimators for method 1 and method 2 for different sample sizes.

mean squared errors criterion (MSE). $\mathrm{MSE}(h) = E\{\widetilde{Q}(p) - Q(p)\}^2$. And the relative efficiency (R.E)

$$\mathrm{R.E} = \left[\frac{\mathrm{MISE}_{\mathrm{Method}\ 2}(h_{\mathrm{Method}\ 2,\mathrm{opt}})}{\mathrm{MISE}_{\mathrm{Method}\ 1}(h_{\mathrm{Method}\ 1,\mathrm{opt}})}\right]. \tag{5.1}$$

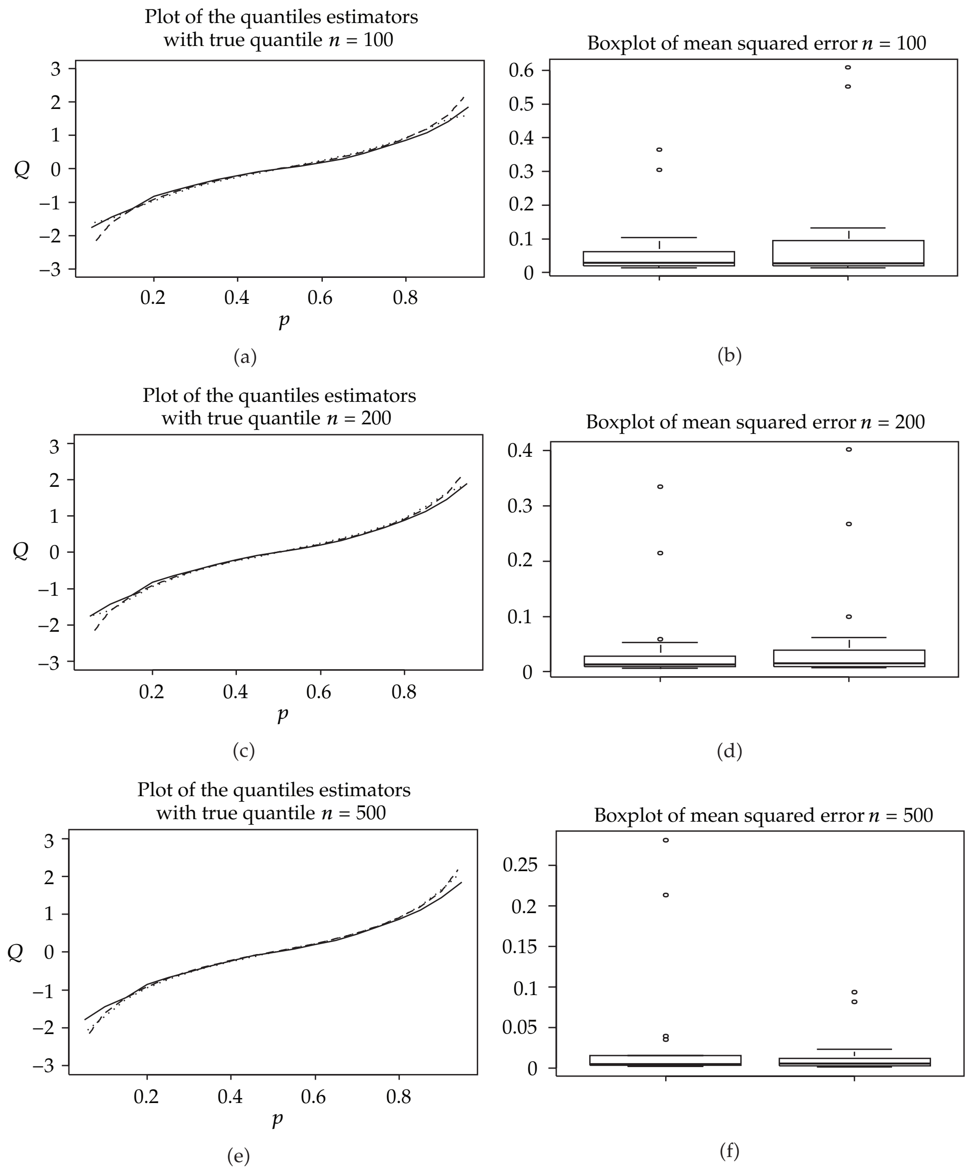

Figure 4: Left panel: plots of the quantile estimators for method 1 (solid line), method 2 (dotted line) and true quantile (dashed line) for different sample sizes and for data from a double exponential distribution. Right panel: box plots of mean squared error for the quantile estimators for method 1 and method 2 for different sample sizes.

Further, for comparison purposes we refer to our proposed method and that of Sheather and Marron [1] as method 1 and method 2 respectively.

(a) Standard normal distribution (see Table 1 and Figure 1).

(b) Exponential distribution (see Table 2 and Figure 2).

(c) Log-normal distribution (see Table 3 and Figure 3).

(d) Double exponential distribution (see Table 4 and Figure 4).

Table 1: Mean squared errors results for bandwidth selection methods for different sample sizes and for data from a normal distribution.

p		$n = 100$	$n = 200$	$n = 500$
0.05	method 1	0.34841956	0.321073870	0.298936771
	method 2	0.29636758	0.164364738	0.090598082
0.10	method 1	0.07645956	0.065440575	0.054697205
	method 2	0.04947745	0.022846355	0.015566907
0.15	method 1	0.02291501	0.013920668	0.007384189
	method 2	0.02939708	0.013386234	0.005005849
0.20	method 1	0.01891919	0.009273746	0.003152866
	method 2	0.02228828	0.010094172	0.003812209
0.25	method 1	0.01596948	0.008581398	0.003000777
	method 2	0.01835912	0.008880639	0.003568772
0.30	method 1	0.01614981	0.008035667	0.003208531
	method 2	0.01639148	0.008299838	0.003375445
0.35	method 1	0.01461880	0.007677567	0.003534028
	method 2	0.01544790	0.007763629	0.003012045
0.40	method 1	0.01279474	0.007375428	0.002899081
	method 2	0.01494506	0.007248497	0.002661230
0.45	method 1	0.01224268	0.006128817	0.002183302
	method 2	0.01444153	0.006790490	0.002295830
0.55	method 1	0.01414050	0.006348893	0.001922013
	method 2	0.01373258	0.006702430	0.002099446
0.60	method 1	0.01375373	0.006392721	0.002007274
	method 2	0.01341763	0.006762798	0.002254869
0.65	method 1	0.01344773	0.006063502	0.002589679
	method 2	0.01290569	0.006801901	0.002507202
0.70	method 1	0.01320832	0.006394102	0.002456085
	method 2	0.01233948	0.007001064	0.002691678
0.75	method 1	0.01503264	0.007011867	0.002789939
	method 2	0.01219829	0.007216326	0.002679609
0.80	method 1	0.01604847	0.007246605	0.002715445
	method 2	0.01327836	0.007602346	0.002791240
0.85	method 1	0.01757171	0.009239589	0.004770755
	method 2	0.01740931	0.009522181	0.003848474
0.90	method 1	0.03192379	0.023292975	0.019942754
	method 2	0.03702774	0.018053976	0.012250413
0.95	method 1	0.15323893	0.147773963	0.150811561
	method 2	0.24825188	0.146840177	0.092517440

We can compute and summarize the relative efficiency of $h_{\text{Method 1,opt}}$ for the all previous distributions in Table 5 .

From Tables 1, 2, 3, and 4, for the all distributions, it can be observed that in 52.3% of cases our method produces lower mean squared errors, slightly wins Sheather-Marron method.

Also, from Table 5 which describes the relative efficiency for $h_{\text{Method 1,opt}}$ we can see $h_{\text{Method 1,opt}}$ more efficient from $h_{\text{Method 2,opt}}$ for all the cases except the standard normal distribution cases with $n = 200,\ 500$ and double exponential distribution cases with $n = 500$.

Table 2: Mean squared errors results for bandwidth selection methods for different sample sizes and for data from an exponential distribution.

p		$n = 100$	$n = 200$	$n = 500$
0.05	method 1	0.001687025	0.0014699990	0.0014107454
	method 2	0.0006023236	0.0002476745	$8.122873e-05$
0.10	method 1	0.001306211	0.0009229338	0.0007744410
	method 2	0.0008225254	0.0004075822	$1.749150e-04$
0.15	method 1	0.001589646	0.0008940486	0.0006237375
	method 2	0.0012963576	0.0006938287	$3.186597e-04$
0.20	method 1	0.002187990	0.0011477063	0.0006801504
	method 2	0.0019188172	0.0010358272	$4.746909e-04$
0.25	method 1	0.002916417	0.0015805678	0.0008156225
	method 2	0.0026838659	0.0014096523	$6.303538e-04$
0.30	method 1	0.003827511	0.0019724207	0.0010289166
	method 2	0.0036542688	0.0018358956	$7.948940e-04$
0.35	method 1	0.004919618	0.0025540323	0.0012720751
	method 2	0.0048301657	0.0023318358	$9.724792e-04$
0.40	method 1	0.005868113	0.0031932355	0.0016253398
	method 2	0.0060092243	0.0028998751	$1.170038e-03$
0.45	method 1	0.007267783	0.0039962426	0.0021094081
	method 2	0.0072785641	0.0035363816	$1.417269e-03$
0.55	method 1	0.011776976	0.0065148222	0.0039208447
	method 2	0.0110599156	0.0055548552	$2.154130e-03$
0.60	method 1	0.012864521	0.0070366699	0.0026965785
	method 2	0.0138585365	0.0070359561	$2.626137e-03$
0.65	method 1	0.018173097	0.0086476349	0.0031472559
	method 2	0.0169709413	0.0088832263	$3.255114e-03$
0.70	method 1	0.021125532	0.0111607501	0.0041235720
	method 2	0.0201049720	0.0114703180	$4.201740e-03$
0.75	method 1	0.024025836	0.0150785289	0.0057215181
	method 2	0.0229763952	0.0149490250	$5.812526e-03$
0.80	method 1	0.037367344	0.0204676368	0.0081595071
	method 2	0.0407106885	0.0181647976	$8.020787e-03$
0.85	method 1	0.057785539	0.0317404871	0.0098128398
	method 2	0.0838657681	0.0300656149	$1.134861e-02$
0.90	method 1	0.078797379	0.0426418410	0.0152139697
	method 2	0.1878456852	0.1117820016	$2.156987e-02$
0.95	method 1	0.121239102	0.0810135450	0.0284524316
	method 2	0.6668323836	0.4923732684	$1.478679e-01$

So, we may conclude that in terms of MISE our bandwidth selection method is more efficient than Sheather-Marron for skewed distributions but not for symmetric distributions.

6. Conclusion

In this paper we have a proposed a cross-validation-based-rule for the selection of bandwidth for quantile functions estimated by kernel procedure. The bandwidth selected by our

Table 3: Mean squared errors results for bandwidth selection methods for different sample sizes and for data from a Log-normal distribution.

p		$n = 100$	$n = 200$	$n = 500$
0.05	method 1	0.001663032	0.0010098573	0.0006568989
	method 2	0.002384136	0.0007270441	0.0003613541
0.10	method 1	0.001863141	0.0008438333	0.0002915013
	method 2	0.002601994	0.0008361475	0.0002981938
0.15	method 1	0.002633153	0.0013492870	0.0004451506
	method 2	0.002623552	0.0011943144	0.0003738508
0.20	method 1	0.003753458	0.0019922356	0.0006866399
	method 2	0.003107351	0.0014724525	0.0005685022
0.25	method 1	0.004956635	0.0027140878	0.0009886053
	method 2	0.004564382	0.0022952079	0.0008557756
0.30	method 1	0.006480195	0.0035603171	0.0015897314
	method 2	0.006436967	0.0031574264	0.0011938924
0.35	method 1	0.008858850	0.0047972372	0.0023446072
	method 2	0.008443129	0.0038626105	0.0015443970
0.40	method 1	0.010053969	0.0055989143	0.0022496198
	method 2	0.010893398	0.0051735721	0.0017579579
0.45	method 1	0.012998940	0.0069058362	0.0030102466
	method 2	0.013607931	0.0063606758	0.0019799551
0.55	method 1	0.019687850	0.0115431473	0.0051386226
	method 2	0.020581110	0.0100828810	0.0029554466
0.60	method 1	0.023881883	0.0129227902	0.0046644050
	method 2	0.025845419	0.0129081138	0.0040301844
0.65	method 1	0.032155537	0.0160476126	0.0056732073
	method 2	0.035737008	0.0167147469	0.0056528658
0.70	method 1	0.045027965	0.0249576836	0.0077709058
	method 2	0.042681315	0.0223936302	0.0077616346
0.75	method 1	0.060715676	0.0318891176	0.0121926243
	method 2	0.059276198	0.0323738749	0.0104119217
0.80	method 1	0.087694754	0.0450814911	0.0165993582
	method 2	0.090704630	0.0530374710	0.0168162426
0.85	method 1	0.140537374	0.0840290373	0.0311728395
	method 2	0.193857196	0.1131949907	0.0350218855
0.90	method 1	0.289944417	0.1642236062	0.0679038026
	method 2	0.552092689	0.2763301818	0.1112433633
0.95	method 1	1.119717137	0.4764026616	0.1984216218
	method 2	2.306672668	1.3159008668	0.2217620895

proposed method is shown to be asymptotically unbiased and in order to assess the numerical performance, we conduct a simulation study and compare it with the bandwidth proposed by Sheather and Marron [1]. Based on the four distributions considered the proposed bandwidth selection appears to provide accurate estimates of quantiles and thus we believe that the new bandwidth selection method is a practically useful method to get bandwidth for the quantile estimator in the form (1.3).

Table 4: Mean squared errors results for bandwidth selection methods for different sample sizes and for data from a double exponential distribution.

p		$n = 100$	$n = 200$	$n = 500$
0.05	method 1	0.35372420	0.288207742	0.251339747
	method 2	0.45458819	0.315704320	0.051385372
0.10	method 1	0.07123072	0.043684160	0.029307307
	method 2	0.14868952	0.097871072	0.023601368
0.15	method 1	0.05081769	0.025358946	0.009241326
	method 2	0.09377244	0.035207151	0.010910214
0.20	method 1	0.02489079	0.015360242	0.007647199
	method 2	0.04997348	0.024864359	0.008013159
0.25	method 1	0.01863802	0.012204904	0.004401402
	method 2	0.03117942	0.019101033	0.006247279
0.30	method 1	0.01869611	0.012031162	0.004145965
	method 2	0.02516932	0.014680335	0.004847191
0.35	method 1	0.01562279	0.009560873	0.003235724
	method 2	0.02017404	0.011355808	0.003513386
0.40	method 1	0.01430068	0.007860775	0.002493813
	method 2	0.01669505	0.009165203	0.002621345
0.45	method 1	0.01386331	0.007587705	0.002485022
	method 2	0.01529664	0.008221501	0.002104265
0.55	method 1	0.01501458	0.007801051	0.002013993
	method 2	0.01280613	0.007796411	0.002227569
0.60	method 1	0.01712203	0.009076922	0.002233672
	method 2	0.01394454	0.009475605	0.002791236
0.65	method 1	0.01946241	0.011129870	0.003521070
	method 2	0.01840894	0.012558998	0.003628169
0.70	method 1	0.02098394	0.011997405	0.003255335
	method 2	0.02333092	0.015792466	0.004534060
0.75	method 1	0.02791943	0.016885471	0.004419826
	method 2	0.02937457	0.019852122	0.005469359
0.80	method 1	0.03532806	0.021319714	0.005471649
	method 2	0.04294634	0.024757804	0.007270187
0.85	method 1	0.05463890	0.030489951	0.011338629
	method 2	0.08441144	0.035306415	0.012054182
0.90	method 1	0.09188621	0.058587164	0.030485192
	method 2	0.14755444	0.083844232	0.024399440
0.95	method 1	0.28184945	0.224432372	0.180645893
	method 2	0.51462209	0.319147435	0.076406491

Table 5: The relative efficiency (R.E) of $h_{\text{Method 1,opt}}$.

n	Standard normal dist.	Exponential dist.	Log normal dist.	Double exponential dist.
100	1.037276	2.636250	1.806082	1.520903
200	0.6986324	2.952808	2.096307	1.308667
500	0.4455828	2.324423	1.173547	0.4519134

Appendix

Step 1. Let $nH = S_1 - 2S_2$, where

$$S_1 = \sum_i \int_0^1 \left(\widetilde{Q}_{-i}(p) - Q(p)\right)^2, \qquad S_2 = \sum_i \int_0^1 \left(\delta\left(\frac{i}{n} - p\right)X_{(i)} - Q(p)\right)\left(\widetilde{Q}_{-i}(p) - Q(p)\right). \tag{A.1}$$

Step 2. With $D_i(p) = K_h(i/n - p)X_{(i)} - Q(p)$ and $D_i^0(p) = \delta(i/n - p)X_{(i)} - Q(p)$

$$\begin{aligned} S_1 &= (n-1)^{-2}n^2(n-2)\int_0^1 \left(\widetilde{Q}(p) - Q(p)\right)^2 + (n-1)^{-2}\sum_{i=1}^n \int_0^1 D_i^2(p), \\ S_2 &= (n-1)^{-1}n^2\int_0^1 \left(\widehat{Q}(p) - Q(p)\right)\left(\widetilde{Q}(p) - Q(p)\right) + (n-1)^{-1}\sum_{i=1}^n \int_0^1 D_i D_i^0(p). \end{aligned} \tag{A.2}$$

Step 3. This step combines Steps 1 and 2 to prove that

$$\begin{aligned} H = &\left\{1 - (n-1)^{-2}\right\}\int_0^1 \left(\widetilde{Q}(p) - Q(p)\right)^2 + \frac{1}{n(n-1)^2}\sum_{i=1}^n \int_0^1 D_i^2(p) \\ &- 2\left\{1 + (n-1)^{-1}\right\}\int_0^1 \left(\widehat{Q}(p) - Q(p)\right)\left(\widetilde{Q}(p) - Q(p)\right) \\ &+ \frac{2}{n(n-1)}\sum_{i=1}^n \int_0^1 D_i(p)D_i^0(p). \end{aligned} \tag{A.3}$$

Step 4. This step establishes that

$$\begin{aligned} &E\left\{\int_0^1 \left(\widetilde{Q}(p) - Q(p)\right)^2\right\}^2 + E\left\{\int_0^1 \left(\widehat{Q}(p) - Q(p)\right)\left(\widetilde{Q}(p) - Q(p)\right)\right\}^2 = 0\left(n^{-2} + h^8\right), \\ &E\left\{n^{-3}\sum_{i=1}^n \int_0^1 D_i^2(p)\right\}^2 + \operatorname{var}\left(n^{-2}\sum_{i=1}^n \int_0^1 D_i D_i^0(p)\right)^2 = 0\left(n^{-3}\right). \end{aligned} \tag{A.4}$$

Step 5. This step combines Steps 3 and 4, concluding that

$$H + \int_0^1 \left(\widehat{Q}(p) - Q(p)\right)^2 = \int_0^1 \left(\widetilde{Q}(p) - \widehat{Q}(p)\right)^2 + 2(n-1)^{-1}\ \mu(h) + 0_2\left(n^{-3/2} + n^{-1}h^4\right), \tag{A.5}$$

where $\mu(h) = \int_0^1 E(D_i(p)D_i^0(p))$.

Let $U = 0_2(\xi)$, for a random variable $U = U(n)$ and a positive sequence $\xi = \xi(n)$

$$E\left(U^2\right) = 0\left(\xi^2\right). \tag{A.6}$$

Step 6. This step notes that $\int_0^1 (\widetilde{Q}(p) - \widehat{Q}(p))^2 = S + T$, where

$$\begin{gathered} S = n^{-2}\sum\sum_{i \neq j} g(X_i, X_j), \qquad T = n^{-2}\sum_{i=1}^{n} g(X_i, X_i), \\ g(X_i, X_j) = \int_0^1 \left\{K_h\left(\frac{i}{n} - p\right)X_{(i)} - \delta\left(\frac{i}{n} - p\right)X_{(i)}\right\}\left\{K_h\left(\frac{j}{n} - p\right)X_{(j)} - \delta\left(\frac{j}{n} - p\right)X_{(j)}\right\}dp, \end{gathered} \tag{A.7}$$

and that $S = S^{(1)} + S^{(2)} + (1 - n^{-1})g_0$, where

$$\begin{gathered} S^{(1)} = n^{-2}\sum\sum_{i \neq j}\{g(X_i, X_j) - g_1(X_i) - g_1(X_j) + g_0\}, \\ S^{(2)} = 2n^{-1}\left(1 - n^{-1}\right)\sum_{i=1}^{n}\{g_1(X_i) - g_0\}, \quad g_1(x) = E\{g(x, X_1)\},\ g_0 = E\{g_1(X_1)\}. \end{gathered} \tag{A.8}$$

Step 7. Shows that $E\{g(X_1, X_1)^2\} = 0(1)$, $E\{g(X_1, X_2)^2\} = 0(h^3)$, $E\{g_1(X_1)^2\} = 0(h^6)\mathrm{var}\{T\} = 0(n^{-3})$, $E(S^{(1)})^2 = 0(n^{-2}h^3)$ and $E(S^{(2)})^2 = 0(n^{-1}h^6)$.

Step 8. This step combines the results of Steps 5, 6, 7, obtaining

$$\begin{aligned} H + \int_0^1 \left(\widehat{Q}(p) - Q(p)\right)^2 &= E(T) + \left(1 - n^{-1}\right)g_0 + 2(n-1)^{-1}\mu(h) \\ &\quad + 0_2\left(n^{-3/2} + n^{-1}h^{3/2} + n^{-1/2}h^3\right) \\ &= \int_0^1 E\left(\widetilde{Q}(p) - \widehat{Q}(p)\right)^2 + 2(n-1)^{-1}\ \mu(h) + 0_2\left(n^{-3/2} + n^{-1}h^{3/2} + n^{-1/2}h^3\right). \end{aligned} \tag{A.9}$$

Step 9. This step notes that $\mu(h) = 0(h)$ and

$$\int_0^1 E\left(\widetilde{Q}(p) - \widehat{Q}(p)\right)^2 = \int_0^1 E\left(\widetilde{Q}(p) - Q(p)\right)^2 + \int_0^1 E\left(\widehat{Q}(p) - Q(p)\right)^2 - 2n^{-1}\mu(h). \tag{A.10}$$

Step 10. This step combines Steps 8 and 9, establishing that

$$\begin{aligned} H + \int_0^1 \left(\widehat{Q}(p) - Q(p)\right)^2 - \int_0^1 E\left(\widehat{Q}(p) - Q(p)\right)^2 \\ = \int_0^1 E\left(\widetilde{Q}(p) - Q(p)\right)^2 + 0_2\left(n^{-3/2} + n^{-1}h^{3/2} + n^{-1/2}h^3\right). \end{aligned} \tag{A.11}$$

This means that

$$E\{H + J - \mathrm{MISE}(h)\}^2 = 0_2\left(n^{-3/2} + n^{-1}h^{3/2} + n^{-1/2}h^3\right). \tag{A.12}$$

References

[1] S. J. Sheather and J. S. Marron, "Kernel quantile estimators," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 410–416, 1990.
[2] E. Parzen, "Nonparametric statistical data modeling," *Journal of the American Statistical Association*, vol. 74, pp. 105–131, 1979.
[3] S.-S. Yang, "A smooth nonparametric estimator of a quantile function," *Journal of the American Statistical Association*, vol. 80, no. 392, pp. 1004–1011, 1985.
[4] M. Falk, "Relative deficiency of kernel type estimators of quantiles," *The Annals of Statistics*, vol. 12, no. 1, pp. 261–268, 1984.
[5] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, UK, 1995.
[6] M. Y. Cheng and S. Sun, "Bndwidth selection for kernel quantile estimation," *Journal of Chines Statistical Association*, vol. 44, no. 3, pp. 271–295, 2006.
[7] M. C. Jones, "Estimating densities, quantiles, quantile densities and density quantiles," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 4, pp. 721–727, 1992.
[8] M. Rudemo, "Empirical choice of histograms and kernel density estimators," *Scandinavian Journal of Statistics*, vol. 9, no. 2, pp. 65–78, 1982.
[9] A. W. Bowman, "An alternative method of cross-validation for the smoothing of density estimates," *Biometrika*, vol. 71, no. 2, pp. 353–360, 1984.