

Bootstrap variance estimation for Nadaraya quantile estimator

K.Y. Cheung · Stephen M.S. Lee

Received: 18 April 2007 / Accepted: 29 December 2008 / Published online: 14 February 2009
© Sociedad de Estadística e Investigación Operativa 2009

Abstract Under a suitable choice of bandwidth, Nadaraya's estimator of the p th quantile yields smaller mean squared error than the unsmoothed p th sample quantile. We investigate the problem of bootstrap estimation of the variance of the Nadaraya quantile estimator and show that the error of the variance estimator can be reduced by smoothing the bootstrap. A novel approach, which calibrates the order p of the bootstrapped Nadaraya quantile estimates, is shown to reduce the error further. A simulation study is reported on the empirical performance of the proposed modified bootstrap variance estimators.

Keywords Nadaraya quantile estimator · Order-calibration · Smoothed bootstrap

Mathematics Subject Classification (2000) 62G05 · 62G09 · 62G30

1 Introduction

Suppose that $\mathcal{X} = (X_1, \dots, X_n)$ constitute a random sample of size n drawn from a distribution F . Let $(X_{(1)}, \dots, X_{(n)})$ be the order statistic of \mathcal{X} such that $X_{(j)}$ denotes its j th smallest datum. For a fixed $p \in (0, 1)$, assume that F has a continuous and positive density f on $F^{-1}(\mathcal{O})$ for an open neighborhood \mathcal{O} containing p . Denote

Supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7131/00P).

K.Y. Cheung · S.M.S. Lee (✉)

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Pokfulam, Hong Kong

e-mail: smslee@hkusua.hku.hk

K.Y. Cheung

e-mail: celiacky@graduate.hku.hk

by $\xi_p = F^{-1}(p)$ the unique p th quantile of F . Estimation of ξ_p has found many applications in real life problems. For example, a risk manager may want to prescribe a lower bound, in the form of a lower quantile, for the value of a portfolio. Children's growth charts provide age-specific quantiles of body measurements such as height, weight and head circumference for parents' reference. An industrial production process is often regulated such that a particular feature of the product will have a desired probability to stay within an acceptable range, given in the form of a pair of quantiles.

The p th sample quantile $X_{([np]+1)}$, where $[\cdot]$ denotes the integer part function, is a natural and consistent estimator for ξ_p but suffers from a lack of efficiency due to the variability of individual order statistics. Nadaraya (1964) defines an alternative estimator $\hat{\xi}_{p,h}$ to be the p th quantile of the kernel-smoothed empirical distribution of \mathcal{X} based on a second-order kernel k and a bandwidth h . Azzalini (1981) establishes a second-order approximation to the mean squared error (MSE) of $\hat{\xi}_{p,h}$ under optimal h , which implies that $\hat{\xi}_{p,h}$ is more efficient than $X_{([np]+1)}$. We extend Azzalini's (1981) results to the more general case where $\hat{\xi}_{p,h}$, which we shall term the Nadaraya quantile estimator, is constructed using a kernel k of even order $r \geq 2$.

The problem of variance estimation for $X_{([np]+1)}$ has been well studied in the literature. The unsmoothed bootstrap (Hall and Martin 1988a) and the r th-order kernel-smoothed bootstrap (Hall et al. 1989) estimate $\text{Var}(X_{([np]+1)})$ subject to relative errors of orders $O(n^{-1/4})$ and $O(n^{-r/(2r+1)})$, respectively. Cheung and Lee (2005) show that the m out of n bootstrap method, which modifies the unsmoothed bootstrap by drawing bootstrap samples of size m , instead of n , from the empirical distribution of \mathcal{X} , yields a relative error of order smaller than $O(n^{-1/4})$. Variance estimation for $\hat{\xi}_{p,h}$ has, however, received relatively little attention, despite its clear practical relevance when concern about accuracy stipulates that $\hat{\xi}_{p,h}$ rather than $X_{([np]+1)}$ be used for estimating ξ_p . It can be shown that $\text{Var}(\hat{\xi}_{p,h})$ is asymptotically equivalent to $n^{-1}p(1-p)f(\xi_p)^{-2}$. Although the latter provides a closed-form expression useful for approximating $\text{Var}(\hat{\xi}_{p,h})$, its direct computation requires estimation of $f(\xi_p)$, which often results in estimates of unsatisfactory accuracy for finite samples. The approximation does not fare much better even if the true $f(\xi_p)$ is used: see Maritz and Jarrett (1978) and Figs. 1 and 2 below for related empirical results. Better alternatives are provided by the bootstrap, which we shall explore in this paper.

The ordinary, unsmoothed, bootstrap draws a large number of bootstrap samples, each of size n , from \mathcal{X} , and estimates $\text{Var}(\hat{\xi}_{p,h})$ by the sample variance of the Nadaraya quantile estimates, each of order p and based on bandwidth h , calculated from the bootstrap samples. The smoothed bootstrap calculates the variance estimator similarly except that the Nadaraya quantile estimates are calculated from (smoothed) bootstrap samples drawn from a kernel density estimate of f based on an s th-order kernel and bandwidth b . We propose further a novel approach to estimating $\text{Var}(\hat{\xi}_{p,h})$, which modifies the smoothed bootstrap variance estimator by calibrating the order p in the calculation of the bootstrapped Nadaraya quantile estimates. We shall denote by $\widehat{\text{VAR}}_{p,0,h}$, $\widehat{\text{VAR}}_{p,b,h}$ and $\widehat{\text{VAR}}_{q,b,h}$ the ordinary bootstrap, smoothed bootstrap and order-calibrated smoothed bootstrap estimators of $\text{Var}(\hat{\xi}_{p,h})$ respectively, where q is the calibrated order of the Nadaraya quantile estimate in the last method. Accuracy of the above three bootstrap variance estimators remains

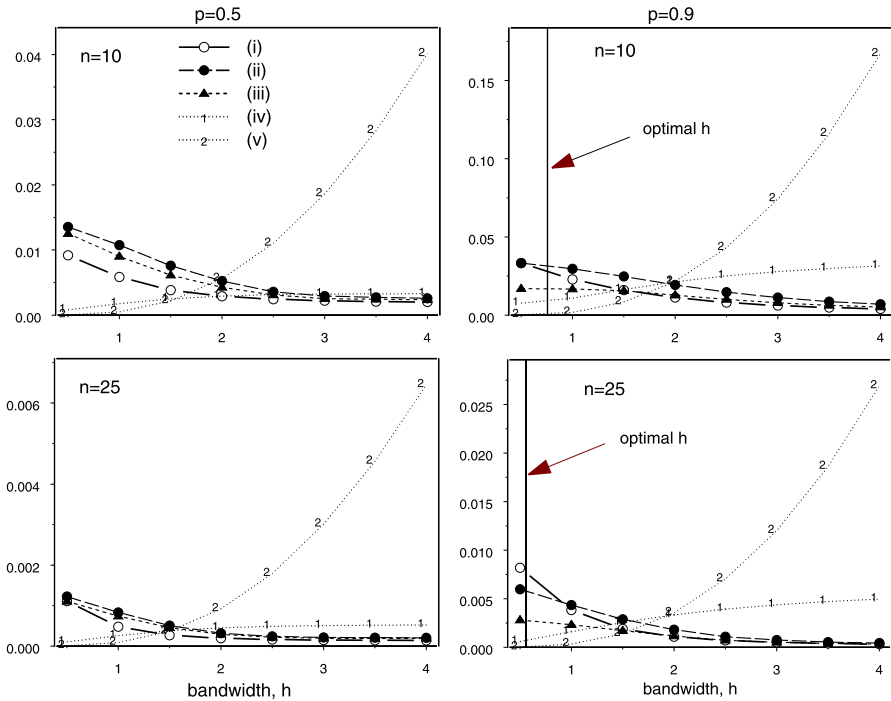


Fig. 1 Normal Example: mean squared errors of estimators of $\text{Var}(\hat{\xi}_{p,h})$ for $p = 0.5$ (left panel) and 0.9 (right panel), and for $n = 10$ (top panel) and 25 (bottom panel). The estimators considered are (i) ordinary bootstrap, (ii) smoothed bootstrap, (iii) order-calibrated smoothed bootstrap, (iv) first-order asymptotic approximation and (v) second-order asymptotic approximation

unexplored in the literature. We study in this paper their performance for the important case $h \propto n^{-1/(2r-1)}$ where $\hat{\xi}_{p,h}$ has the smallest asymptotic MSE given a fixed r . When $r = 2$, $\widehat{\text{VAR}}_{p,0,h}$ has relative error of order $O_p(n^{-1/3})$. Smoothing with $b \propto n^{-1/(2s+1)}$ reduces it to $O_p(n^{-s/(2s+1)})$, and order-calibration further to $O_p(n^{-(s+2)/(2s+5)})$ if $b \propto n^{-1/(2s+5)}$ and $q = p + b^s C_0$ for some constant C_0 . When $r > 2$, neither smoothing nor order-calibration improves upon $\widehat{\text{VAR}}_{p,0,h}$, which has relative error of order $O_p(n^{-2/(2r-1)})$.

The rest of the paper is organized as follows. Section 2 derives asymptotic expansions for the variance and MSE of $\hat{\xi}_{p,h}$ and shows that $\hat{\xi}_{p,h}$ has smaller error than $X_{(np+1)}$ if h is chosen appropriately. Section 3 considers estimation of $\text{Var}(\hat{\xi}_{p,h})$, for $h \propto n^{-1/(2r-1)}$, by the bootstrap estimators $\widehat{\text{VAR}}_{p,0,h}$, $\widehat{\text{VAR}}_{p,b,h}$ and $\widehat{\text{VAR}}_{q,b,h}$. Simulation results and concluding remarks are given in Sects. 4 and 5, respectively. Technical details are given in the Appendix.

2 Nadaraya quantile estimator

Suppose that F satisfies the following conditions for some $d \geq 4$.

(A.1) F is $(d + 1)$ -times continuously differentiable in a neighborhood of ξ_p .

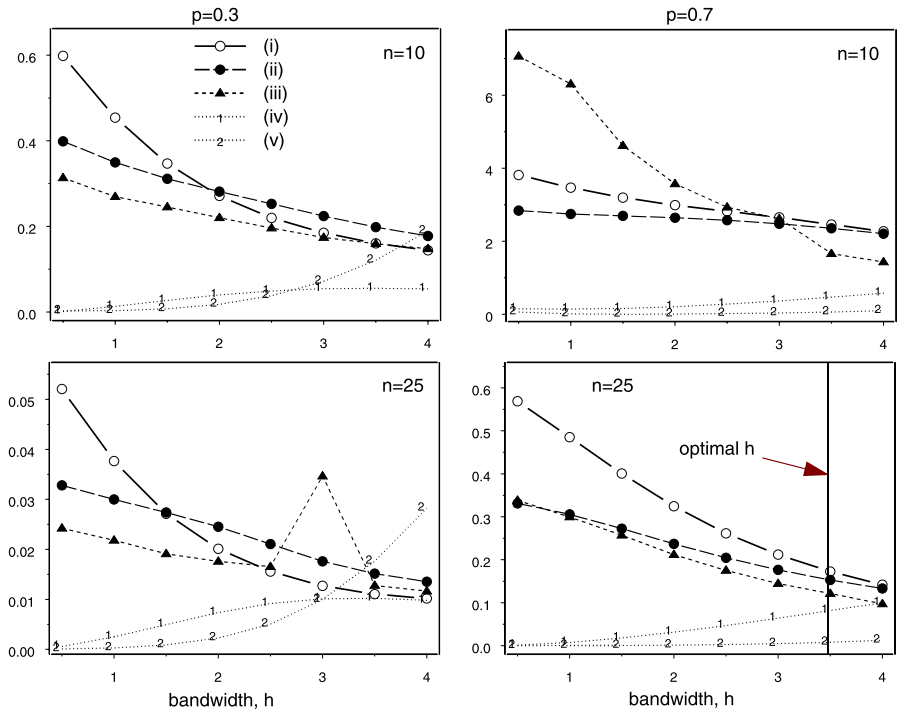


Fig. 2 Chi-squared Example: mean squared errors of estimators of $\text{Var}(\hat{\xi}_{p,h})$ for $p = 0.3$ (left panel) and 0.7 (right panel), and for $n = 10$ (top panel) and 25 (bottom panel). The estimators considered are (i) ordinary bootstrap, (ii) smoothed bootstrap, (iii) order-calibrated smoothed bootstrap, (iv) first-order asymptotic approximation and (v) second-order asymptotic approximation

- (A.2) $F^{(j)}$ is bounded for $1 \leq j \leq d + 1$, and the density $f \equiv F'$ satisfies $f(\xi_p) > 0$.
- (A.3) $\mathbb{E}|X|^\epsilon < \infty$ for some $\epsilon > 0$.

For definition of the Nadaraya quantile estimator $\hat{\xi}_{p,h}$, we consider the kernel density estimator

$$\hat{f}_h(x) = (nh)^{-1} \sum_{i=1}^n k((x - X_i)/h)$$

based on a bandwidth $h > 0$ and a kernel function $k = K'$ of even order $r \geq 2$ so that $\int_{-\infty}^{\infty} x^j k(x) dx = 1, 0$ and $\alpha_r \neq 0$ for $j = 0, 1 \leq j \leq r - 1$ and $j = r$, respectively. Assume the following conditions on h and k :

- (B.1) $h \rightarrow 0$ and $nh^{2+\epsilon} \rightarrow \infty$ as $n \rightarrow \infty$ for some $\epsilon > 0$.
- (B.2) k is bounded, absolutely integrable and symmetric about 0 .
- (B.3) $\int |x|^r k(x) dx < \infty$.

Define

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(y) dy.$$

The Nadaraya estimator of ξ_p is given by $\hat{\xi}_{p,h} = \hat{F}_h^{-1}(p)$. The conditions (A.1)–(A.3) and (B.1)–(B.3) are typical regularity conditions sufficient for asymptotic normality of \hat{f}_h , \hat{F}_h and $\hat{\xi}_{p,h}$: see, for example, Reiss (1981) and Ralescu and Sun (1993). Note that imposition of a greater d strengthens (A.1) and (A.2) and enables us to establish asymptotic properties of $\hat{\xi}_{p,h}$ constructed using a higher kernel order r . Nadaraya (1964) and Azzalini (1981) establish first- and second-order approximations to $\hat{\xi}_{p,h}$ respectively for the special case $r = 2$. We prove in the Appendix asymptotic expansions for the variance and MSE of $\hat{\xi}_{p,h}$ as given in the following lemma.

Lemma 1 *Under conditions (A.1)–(A.3) and (B.1)–(B.3) and assuming that $r \leq d - 2$, we have*

$$\text{Var}(\hat{\xi}_{p,h}) = \frac{p(1-p)}{nf(\xi_p)^2} - \frac{2h\beta}{nf(\xi_p)} + O(n^{-3/2} + n^{-1}h^2 + n^{-1/2}h^{2r} + h^{3r}) \tag{1}$$

and

$$\begin{aligned} \text{MSE}(\hat{\xi}_{p,h}) &= \frac{p(1-p)}{nf(\xi_p)^2} - \frac{2h\beta}{nf(\xi_p)} + \frac{h^{2r} f^{(r-1)}(\xi_p)^2 \alpha_r^2}{r!^2 f(\xi_p)^2} \\ &\quad + O(n^{-3/2} + n^{-1}h^2 + n^{-1/2}h^{2r} + h^{2r+1}), \end{aligned} \tag{2}$$

where $\beta = \int uk(u)K(u) du$.

Minimizing the expansion (2) suggests setting

$$h = h_{\text{opt}} \equiv \left(\frac{\beta f(\xi_p) r!^2}{nr f^{(r-1)}(\xi_p)^2 \alpha_r^2} \right)^{1/(2r-1)}, \tag{3}$$

provided that $f^{(r-1)}(\xi_p) \neq 0$, so that the MSE is minimized to be

$$\begin{aligned} \text{MSE}(\hat{\xi}_{p,h_{\text{opt}}}) &= \frac{p(1-p)}{nf(\xi_p)^2} - \left(2 - \frac{1}{r}\right) \left(\frac{\beta^{2r} r!^2}{n^{2r} r \alpha_r^2 f^{(r-1)}(\xi_p)^2 f(\xi_p)^{2(r-1)}} \right)^{1/(2r-1)} \\ &\quad + O(n^{-(2r+1)/(2r-1)} + n^{-3/2}). \end{aligned} \tag{4}$$

Under conditions (A.1)–(A.3), we see from Duttweiler (1973) Theorem 2 that the bias of $X_{(\lfloor np \rfloor + 1)}$ has order $O(n^{-3/4})$, and obtain by Hall and Martin (1988a) Theorem 2.1 an expansion for its variance. It follows that $\text{MSE}(X_{(\lfloor np \rfloor + 1)}) = n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-3/2})$. Comparison with (4) gives that

$$\text{MSE}(X_{(\lfloor np \rfloor + 1)}) - \text{MSE}(\hat{\xi}_{p,h_{\text{opt}}}) \sim \Delta n^{-2r/(2r-1)}$$

for some positive constant $\Delta = O(1)$, so that $\hat{\xi}_{p,h_{\text{opt}}}$ outperforms $X_{(\lfloor np \rfloor + 1)}$ by having a smaller MSE asymptotically. It is clear that the reduction in MSE has order $n^{-2r/(2r-1)}$ and becomes more pronounced as r increases, suggesting advantages of using a higher kernel order r . However, the true effects of increasing r for finite samples depend also on a number of other factors including f , $f^{(r-1)}$ and the choice of k .

Table 1 Normal Example: mean squared errors of the Nadaraya p th quantile estimator based on the Epanechnikov kernel and optimal bandwidth h_{opt} , and the p th sample quantile for $n = 10$

p		0.10	0.25	0.45	0.60	0.75	0.90
MSE($X_{(\lfloor np \rfloor + 1)}$)		0.2971	0.1781	0.1478	0.1761	0.1776	0.4071
MSE($\hat{\xi}_{p, h_{\text{opt}}}$),	(i) $r = 2$	0.2052	0.1424	0.1393	0.1382	0.1441	0.2031
	(ii) $r = 4$	0.1780	0.1270	0.1039	0.1089	0.1278	0.1766
	(iii) $r = 6$	0.1671	0.1321	0.1082	0.1133	0.1326	0.1667
MSE($\hat{\xi}_{p, \hat{h}_{\text{opt}}}$),	(i) $r = 2$	1.0903	0.5766	0.1371	0.2177	0.5887	1.0636
	(ii) $r = 4$	0.3186	0.1867	0.1307	0.1352	0.1833	0.3505
	(iii) $r = 6$	0.2188	0.1577	0.1312	0.1333	0.1576	0.1889

The optimal formula for h_{opt} involves the density f and its derivatives, which are usually unknown. An estimate of h_{opt} can be obtained by substituting $X_{(\lfloor np \rfloor + 1)}$ for ξ_p , and kernel estimates for $f(\xi_p)$ and $f^{(r-1)}(\xi_p)$ into (3). Park and Marron (1990) argue that this ‘plug-in’ approach is the best available bandwidth selection method in density estimation. More data-dependent bandwidth selection methods for kernel density estimation can be found in, for example, Staniswalis (1989) and Chiu (1991). Härdle et al. (1990) propose a cross-validation approach to fixing the bandwidths in the kernel estimates of f and its derivatives.

We carried out a simulation study to compute the MSE’s of $\hat{\xi}_{p, h_{\text{opt}}}$ and $X_{(\lfloor np \rfloor + 1)}$ from 8000 random samples of size 10 drawn from the standard normal distribution for $p = 0.10, 0.25, 0.45, 0.60, 0.75$ and 0.90 . The kernels k used for $\hat{\xi}_{p, h_{\text{opt}}}$ have support on $[-1, 1]$ and are defined, for $|x| \leq 1$, by (i) $k(x) = (3/4)(1 - x^2)$ (the Epanechnikov kernel), (ii) $k(x) = 15(7x^4 - 10x^2 + 3)/32$ and (iii) $k(x) = 105(5 - 35x^2 + 63x^4 - 33x^6)/256$, which have orders 2, 4 and 6, respectively. We also obtained the MSE results for $\hat{\xi}_{p, \hat{h}_{\text{opt}}}$, where \hat{h}_{opt} was the plug-in estimate of h_{opt} , in which ξ_p was estimated by $X_{(\lfloor np \rfloor + 1)}$, and $f, f^{(r-1)}$ by their kernel estimates based on bandwidths found by Härdle et al. (1990) cross-validation method. Table 1 summarizes our findings, which show that $\text{MSE}(\hat{\xi}_{p, h_{\text{opt}}})$ is significantly smaller than $\text{MSE}(X_{(\lfloor np \rfloor + 1)})$ and that use of a higher order r is advantageous. Empirical estimation of h_{opt} affects the MSE of the Nadaraya estimate adversely. Indeed, $\hat{\xi}_{p, \hat{h}_{\text{opt}}}$ based on $r = 2$ is no longer favorable compared to $X_{(\lfloor np \rfloor + 1)}$. However, increasing the order to $r = 6$ rectifies the situation remarkably, rendering $\hat{\xi}_{p, \hat{h}_{\text{opt}}}$ a smaller MSE than that of $X_{(\lfloor np \rfloor + 1)}$.

3 Variance estimation for Nadaraya quantile estimator

We consider in this section estimation of $\text{Var}(\hat{\xi}_{p, h})$ by the bootstrap and its variants. To fix ideas we focus on the important case where h has the order of h_{opt} , that is, $h \propto n^{-1/(2r-1)}$, but is not necessarily equal to h_{opt} precisely. Our results thus apply to a reasonably wide choice of h of practical interest, without dwelling unduly on the issue of optimal selection of h .

3.1 Ordinary bootstrap

Let $\mathcal{X}^* = (X_1^*, \dots, X_n^*)$ denote a generic bootstrap sample of size n drawn randomly with replacement from \mathcal{X} . Define

$$\hat{F}_h^*(x) = \sum_{i=1}^n K((x - X_i^*)/h)/n,$$

the bootstrap version of \hat{F}_h . Then the ordinary bootstrap estimates $\text{Var}(\hat{\xi}_{p,h})$ by $\widehat{\text{VAR}}_{p,0,h} = \text{Var}(\hat{F}_h^{*-1}(p)|\mathcal{X})$. We prove in the Appendix the following theorem.

Theorem 1 *Under the conditions of Lemma 1 and assuming that $h \propto n^{-1/(2r-1)}$, we have*

$$\widehat{\text{VAR}}_{p,0,h}/\text{Var}(\hat{\xi}_{p,h}) = 1 + \begin{cases} O_p(n^{-1/3}), & r = 2, \\ O_p(n^{-2/(2r-1)}), & r > 2. \end{cases}$$

We see that the convergence rate of $\widehat{\text{VAR}}_{p,0,h}$ becomes slower as the kernel order r increases, which is a price to pay for the reduction in $\text{MSE}(\hat{\xi}_{p,h})$ offered by use of a large r as has been discussed in Sect. 2.

It is interesting to note from Hall and Martin (1988a) that the ordinary bootstrap estimator of $\text{Var}(X_{([np]+1)})$ has relative error of order $O_p(n^{-1/4})$, suggesting that bootstrap variance estimation is less efficient for sample quantile than for Nadaraya quantile estimator if the latter is based on a second-order kernel, which gives an $O_p(n^{-1/3})$ relative error.

3.2 Smoothed bootstrap

The smoothed bootstrap has been extensively studied by, for example, Efron (1982), Silverman and Young (1987) and De Angelis and Young (1992). Falk and Janas (1992), Janas (1993), De Angelis et al. (1993) and Ho and Lee (2005) consider smoothed bootstrap confidence intervals for ξ_p .

We describe below a smoothed bootstrap procedure for estimating $\text{Var}(\hat{\xi}_{p,h})$. Let $\tilde{X}_1^*, \dots, \tilde{X}_n^*$ constitute a smoothed bootstrap sample drawn from a kernel density estimate $\hat{\ell}_b$ of f , where

$$\hat{\ell}_b(x) = (nb)^{-1} \sum_{i=1}^n \ell((x - X_i)/b),$$

$b > 0$ denotes the bandwidth and ℓ is an s th-order kernel function for even $s \geq 2$. It gives rise to a smoothed bootstrap version of \hat{F}_h , namely

$$\tilde{F}_{b,h}^*(x) = \sum_{i=1}^n K((x - \tilde{X}_i^*)/h)/n.$$

Then the smoothed bootstrap estimator of $\text{Var}(\hat{\xi}_{p,h})$ is $\widehat{\text{VAR}}_{p,b,h} = \text{Var}(\tilde{F}_{b,h}^{*-1}(p)|\mathcal{X})$. Its relative error is given by the following theorem, which we prove in the Appendix.

Theorem 2 Assume the conditions of Theorem 1, that $s \leq d - 2$, that $b \rightarrow 0$ and $nb^3 / \log n \rightarrow \infty$ as $n \rightarrow \infty$, and that (B.2) and (B.3) hold with k replaced by ℓ . Then

$$\widehat{\text{VAR}}_{p,b,h} / \text{Var}(\hat{\xi}_{p,h}) = 1 + O_p(h^2 + n^{1/2}h^{2r} + nh^{3r} + n^{-1/2}b^{-1/2} + b^s).$$

Furthermore, if we take $b \propto n^{-1/(2s+1)}$, then

$$\widehat{\text{VAR}}_{p,b,h} / \text{Var}(\hat{\xi}_{p,h}) = 1 + \begin{cases} O_p(n^{-s/(2s+1)}), & r = 2, \\ O_p(n^{-2/(2r-1)}), & r > 2. \end{cases}$$

Theorem 2 asserts that when $r = 2$, bootstrap estimation of $\text{Var}(\hat{\xi}_{p,h})$ can be made more accurate by smoothing with a bandwidth b of order $n^{-1/(2s+1)}$, reducing the relative error from $O_p(n^{-1/3})$ to $O_p(n^{-s/(2s+1)})$. In principle, the latter can be made arbitrarily close to $O_p(n^{-1/2})$ by choosing a sufficiently high kernel order s , in which case, however, $\hat{\ell}_b(x)$ necessarily takes on negative values for some x and poses practical difficulties if smoothed bootstrap samples need be simulated from $\hat{\ell}_b$. Hall and Murison (1993) and Lee and Young (1994) correct for negativity to make the smoothed bootstrap procedure computationally feasible even for $s > 2$. Whether their corrections apply to the present context without impinging on the reduction in relative error remains to be studied. When $s = 2$ so that $\hat{\ell}_b$ is a proper density function, the optimal relative error of $\widehat{\text{VAR}}_{p,b,h}$ has order $O_p(n^{-2/5})$, which already improves upon the $O_p(n^{-1/3})$ relative error given by $\widehat{\text{VAR}}_{p,0,h}$.

When $r > 2$, smoothing has apparently no effects on the order of relative error of the bootstrap variance estimator.

We remark that the optimal convergence rate asserted by Theorem 2 is valid for any choice of bandwidth $b \propto n^{-1/(2s+1)}$. Selection of an ‘‘optimal’’ value for b is not crucial to rendering $\widehat{\text{VAR}}_{p,b,h}$ the best convergence rate. Empirical determination of b in practice can be done by, for example, a double bootstrap procedure, to which we shall return in Sect. 3.3 in more detail.

3.3 Order-calibrated smoothed bootstrap

The technique of calibration has been successfully applied to problems of bootstrap confidence interval and bias correction to reduce error: see, for example, Beran (1987) and Hall and Martin (1988b). It can be applied in the present context treating the quantile order p as the target for calibration. When applied to the smoothed bootstrap estimator, this gives rise to a new estimator $\widehat{\text{VAR}}_{q,b,h}$ of $\text{Var}(\hat{\xi}_{p,h})$ for which $q = p + o(1)$ denotes the calibrated order of the quantile. As can be deduced from (12) in the Appendix, the smoothed bootstrap estimator $\widehat{\text{VAR}}_{p,b,h}$ introduces a bias term $-n^{-1}b^s(1 - 2p)C_{0f}(\xi_p)^{-2}$, which can be eliminated by careful tuning of the calibrated order q . Subsequent selection of b to minimize the relative error yields an improvement upon the smoothed bootstrap for the case $r = 2$, as asserted in the following theorem. The proof is outlined in the Appendix.

Theorem 3 *Under the conditions of Theorem 2, we have*

$$\widehat{\text{VAR}}_{q,b,h} / \text{Var}(\hat{\xi}_{p,h}) = 1 + \begin{cases} O_p(n^{-(s+2)/(2s+5)}), & r = 2, \\ O_p(n^{-2/(2r-1)}), & r > 2, \end{cases}$$

provided that $b \propto n^{-1/(2s+5)}$ and $q = p + C_0 b^s$, where

$$C_0 = 2\gamma_s p(1-p) \{s!(1-2p)f(\xi_p)\}^{-1} \{f^{(s)}(\xi_p) - f^{(s-1)}(\xi_p)f'(\xi_p)/f(\xi_p)\}$$

and $\gamma_s = \int u^s \ell(u) du$.

As with smoothing, proper calibration of p is advantageous if $r = 2$, with the effect of reducing the relative error from $O_p(n^{-s/(2s+1)})$ to $O_p(n^{-(s+2)/(2s+5)})$.

Practical determination of q involves estimation of densities and their derivatives, which is notoriously difficult. Noting that the choice $q = p + C_0 b^s$ actually minimizes $\text{MSE}(\widehat{\text{VAR}}_{q,b,h}) = \mathbb{E}[(\widehat{\text{VAR}}_{q,b,h} - \text{Var}(\hat{\xi}_{p,h}))^2]$ asymptotically, an alternative approach to computing q is given by the following double bootstrap procedure. Draw a large number, C say, of first-level smoothed bootstrap samples from $\hat{\ell}_{b_0}$ with $b_0 \gg b$. From the c th such sample, draw B second-level smoothed bootstrap samples based on bandwidth b and kernel ℓ . Select T distinct values of the form $p + \lambda b^s \in (0, 1)$ by varying λ and denote them by $q_1 < \dots < q_T$. For each q_t and each c , calculate the sample variance, $\widehat{\text{VAR}}_{q_t,b,h}^{*(c)}$ say, of the B Nadaraya q_t th-order quantile estimates obtained respectively from the B second-level smoothed bootstrap samples that stem from the c th first-level sample. Estimate $\text{MSE}(\widehat{\text{VAR}}_{q_t,b,h})$ by $\text{MSE}_t^* \equiv C^{-1} \sum_{c=1}^C (\widehat{\text{VAR}}_{q_t,b,h}^{*(c)} - \widehat{\text{VAR}}_{p,b_0,h})^2$. Fit the quadratic function $\text{MSE}_t^* = b_0 + b_1 q_t + b_2 q_t^2$, $t = 1, \dots, T$, by the least squares method to obtain the coefficients b_0, b_1, b_2 . Calculate $q = -b_1/(2b_2)$, which minimizes $b_0 + b_1 q + b_2 q^2$. It can be shown that this value of q differs from its theoretical optimal value by $O_p(b_0^s b^s + n^{-1/2} b_0^{-s-1/2} b^s)$.

4 Simulation study

We conducted a simulation study of the effects of smoothing and order-calibration on bootstrap estimation of $\text{Var}(\hat{\xi}_{p,h})$ for $n = 10$ and 25 , and for $p = 0.10, 0.30, 0.50, 0.70$ and 0.90 . The underlying distribution was taken to be the standard normal and the chi-squared on 5 degrees of freedom, whose densities satisfy (A.1)–(A.3) for an arbitrarily large d . The Epanechnikov kernel, which satisfies (B.2) and (B.3), was used for both k and ℓ , so that $r = s = 2$. The bandwidth h was chosen to be $0.5j$ for $j = 1, 2, \dots, 8$. The bandwidth b for the smoothed bootstrap was found by minimizing the MSE estimated by the double bootstrap. The resulting empirical bandwidth b was also used for determining the calibrated order q based on the double bootstrap algorithm described in Sect. 3.3. In both cases, we set $b_0 = 2$, $T = 9$ and $B = C = 400$ in the double bootstrap procedure. The MSE of the variance estimator was each approximated by averaging over 250 random samples. For comparison, we calculated also the squared errors of the first- and second-order asymptotic approximations to $\text{Var}(\hat{\xi}_{p,h})$ given respectively by $n^{-1} p(1-p)f(\xi_p)^{-2}$

and $n^{-1}p(1-p)f(\xi_p)^{-2} - 2n^{-1}h\beta f(\xi_p)^{-1}$ based on the true $f(\xi_p)$: see (1) for details. The results provide insight into the best level of accuracy one might expect of methods that plug in density estimates directly into asymptotic expressions for estimating $\text{Var}(\hat{\xi}_{p,h})$.

Figures 1 and 2 plot the MSE's of the various bootstrap estimators and the squared errors of the two asymptotic approximations for the normal and chi-squared data, respectively. For conciseness, we report only the more representative cases, that is, $p = 0.5$ and 0.9 for the normal data, and $p = 0.3$ and 0.7 for the chi-squared data. We note that the optimal h_{opt} is invalid when $f^{(r-1)}(\xi_p) = 0$. This happens for the normal case where $p = 0.5$ and for the chi-squared case where $p \approx 0.3$. The value of h_{opt} is indicated by a vertical line in Figs. 1 and 2 whenever it lies within the range of h under study.

The simulation results generally agree with our asymptotic findings. Especially notable is the improvement made by order-calibration upon the smoothed bootstrap, which is observed in all cases except for a few awkward figures found in the chi-squared case. Smoothing and order-calibration appear to be more effective in correcting the ordinary bootstrap under the chi-squared distribution, especially when h is small. The squared errors of both asymptotic approximations increase as h increases, with that of the second-order approximation increasing at a much faster rate except for $p = 0.7$ in the chi-squared case. They are in general found to be more accurate than the bootstraps for small h but less for large h .

5 Conclusion

We have shown, both theoretically and empirically, that the smoothed and order-calibrated smoothed bootstraps improve upon the ordinary bootstrap estimator of $\text{Var}(\hat{\xi}_{p,h})$ for the most important case where calculation of $\hat{\xi}_{p,h}$ is based on a bandwidth h of optimal order and a second-order kernel k . In particular, the $O_p(n^{-1/3})$ relative error of the ordinary bootstrap can be reduced to $O_p(n^{-2/5})$ and $O_p(n^{-4/9})$ respectively by $\widehat{\text{VAR}}_{p,b,h}$ and $\widehat{\text{VAR}}_{q,b,h}$ if the smoothed bootstrap is done with a second-order kernel ℓ , provided that $b \propto n^{-1/5}$ in the former case and $b \propto n^{-1/9}$ in the latter with $q = p + C_0b^2$. Use of a higher-order kernel for smoothing the bootstrap can reduce the error rate of the variance estimator further but will complicate the Monte Carlo procedure due to negativity of the kernel estimate $\hat{\ell}_b$. Hall and Murison (1993) suggest some methods of removing negativity of higher-order kernel density estimators, but their effects on the smoothed bootstrap variance estimators for $\hat{\xi}_{p,h}$ remain to be studied. Order-calibration provides a practically more appealing method for enhancing the accuracy of the smoothed bootstrap without relying on high-order kernels nor any negativity correction techniques. Numerical results also suggest that bootstrap methods have advantages over direct asymptotic approximations for estimating $\text{Var}(\hat{\xi}_{p,h})$.

We have also studied the effects of increasing the order r of the kernel k used in calculating $\hat{\xi}_{p,h}$. In general, $\hat{\xi}_{p,h}$ becomes more accurate for estimating ξ_p when r increases, as a consequence of the reduction in biases of the kernel estimators \hat{F}_h and \hat{f}_h . We have seen, however, then that bootstrap estimation of the corresponding

$\text{Var}(\hat{\xi}_{p,h})$ suffers from a slower convergence rate, upon which neither smoothing the bootstrap nor calibrating the quantile order can improve. It appears that use of a high kernel order r for calculating $\hat{\xi}_{p,h}$ is only recommendable when one is not concerned with estimating its variance in the practical problem being investigated.

In many practical applications, inference based on the quantile estimator requires also estimation of its sampling distribution. In the case of the sample quantile $X_{([np]+1)}$, Theorem 4.2.4 in Reiss (1989) establishes an Edgeworth expansion of the form

$$\Phi(t) + n^{-1/2}E_1(t) + O(n^{-1})$$

for the distribution function of the standardized $X_{([np]+1)}$, where Φ denotes the standard normal distribution function, and E_1 is a smooth function independent of n . The corresponding expansion for the standardized $\hat{\xi}_{p,h}$ based on a second-order kernel, derived in (A.20) of Ho and Lee (2005), has the form

$$\Phi(t) + n^{-1/2}\tilde{E}_1(t) + h\tilde{E}_2(t) + n^{1/2}h^2\tilde{E}_3(t) + O(n^{-1} + n^{3/2}h^4)$$

for some smooth functions \tilde{E}_j independent of n and h . Comparison of the two expansions suggests that the problem of distribution estimation for $\hat{\xi}_{p,h}$ remains essentially the same as that for $X_{([np]+1)}$ insofar as accuracy of the smoothed bootstrap is concerned, provided that h converges to zero sufficiently fast.

Some recent studies have been done to extend the Nadaraya quantile estimator to the case of dependent data. For example, Cai and Roussas (1997) establish asymptotic properties of $\hat{\xi}_{p,h}$ for positively or negatively associated observations; Sun and Cordero (2005) consider stationary strong mixing random variables; Youndjé and Vieu (2006) investigate consistency properties for long-memory stationary stochastic processes. Methods for estimating the distributional properties of $\hat{\xi}_{p,h}$ in these more general contexts are certainly worth investigating.

Appendix

Proof of Lemma 1

Write $Z_{1,h} = n^{1/2}\{\hat{F}_h(\xi_p) - \mathbb{E}\hat{F}_h(\xi_p)\}$ and $Z_{2,h} = (nh)^{1/2}\{\hat{f}_h(\xi_p) - \mathbb{E}\hat{f}_h(\xi_p)\}$. Under conditions somewhat weaker than those assumed in Lemma 1, Reiss (1981) shows that $Z_{1,h}$ is asymptotically normally distributed with zero mean. Also, conditions (B.1)–(B.3) imply the Lindeberg condition on the double array of random variables $\{k((\xi_p - X_i)/h)/h\}$, under which $Z_{2,h}$ is asymptotically normal with mean zero. The smoothness conditions on F enable us to expand the biases of $\hat{F}_h(\xi_p)$ and $\hat{f}_h(\xi_p)$ such that

$$\mathbb{E}\hat{F}_h(\xi_p) - F(\xi_p) = h^r \alpha_r f^{(r-1)}(\xi_p)/r! + O(h^{r+1}) = \bar{\mathcal{B}} \text{ say,}$$

$$\mathbb{E}\hat{f}_h(\xi_p) - f(\xi_p) = h^r \alpha_r f^{(r)}(\xi_p)/r! + O(h^{r+1}) = \mathcal{B} \text{ say.}$$

It then follows by inverting the Taylor expansion $p = \hat{F}_h(\hat{\xi}_{p,h}) = \hat{F}_h(\xi_p) + (\hat{\xi}_{p,h} - \xi_p)\hat{f}_h(\xi_p) + O_p(n^{-1} + h^{2r})$ that

$$\begin{aligned} \hat{\xi}_{p,h} - \xi_p &= \frac{p - \hat{F}_h(\xi_p)}{\hat{f}_h(\xi_p)} + O_p(n^{-1} + h^{2r}) \\ &= \frac{-n^{-1/2}Z_{1,h} - \bar{B}}{(nh)^{-1/2}Z_{2,h} + f(\xi_p) + \mathcal{B}} + O_p(n^{-1} + h^{2r}) \\ &= -\frac{n^{-1/2}Z_{1,h} + \bar{B}}{f(\xi_p)} \left\{ 1 + \frac{(nh)^{-1/2}Z_{2,h} + \mathcal{B}}{f(\xi_p)} \right\}^{-1} + O_p(n^{-1} + h^{2r}) \\ &= -\frac{n^{-1/2}Z_{1,h} + \bar{B}}{f(\xi_p)} \left\{ 1 - \frac{(nh)^{-1/2}Z_{2,h} + \mathcal{B}}{f(\xi_p)} \right\} \\ &\quad + O_p\{(nh)^{-1}(n^{-1/2} + h^r)\} + O_p(n^{-1} + h^{2r}) \\ &= -\frac{Z_{1,h}}{n^{1/2}f(\xi_p)} - \frac{\bar{B}}{f(\xi_p)} + \frac{Z_{1,h}Z_{2,h}}{nh^{1/2}f(\xi_p)^2} + \frac{Z_{2,h}\bar{B}}{n^{1/2}h^{1/2}f(\xi_p)^2} \\ &\quad + O_p(n^{-1} + h^{2r}). \end{aligned} \tag{5}$$

The order term in the last equality follows by noting that $(nh)^{-2} = o(n^{-1})$, as implied by (B.1). Thus, we have

$$\begin{aligned} \text{Var}(\hat{\xi}_{p,h}) &= \frac{\mathbb{E}[Z_{1,h}^2]}{nf(\xi_p)^2} - \frac{2\mathbb{E}[Z_{1,h}^2Z_{2,h}]}{n^{3/2}h^{1/2}f(\xi_p)^3} \\ &\quad + O(n^{-3/2} + n^{-1/2}h^{2r} + n^{-1}h^r + h^{3r}), \end{aligned} \tag{6}$$

$$\begin{aligned} \text{MSE}(\hat{\xi}_{p,h}) &= \frac{\mathbb{E}[Z_{1,h}^2]}{nf(\xi_p)^2} - \frac{2\mathbb{E}[Z_{1,h}^2Z_{2,h}]}{n^{3/2}h^{1/2}f(\xi_p)^3} - \frac{4\mathbb{E}[Z_{1,h}Z_{2,h}]\bar{B}}{nh^{1/2}f(\xi_p)^3} + \frac{\bar{B}^2}{f(\xi_p)^2} \\ &\quad + O(n^{-3/2} + n^{-1/2}h^{2r} + n^{-1}h^r + h^{3r}). \end{aligned} \tag{7}$$

Invoking the smoothness conditions on F again to expand the mixed moments of $Z_{1,h}$ and $Z_{2,h}$, we obtain

$$\begin{aligned} \mathbb{E}[Z_{1,h}^2] &= p(1 - p) - 2h\beta f(\xi_p) + O(h^2), \\ \mathbb{E}[Z_{1,h}Z_{2,h}] &= h^{1/2} \left\{ \frac{1}{2} - p \right\} f(\xi_p) + O(h^{3/2}), \\ \mathbb{E}[Z_{1,h}^2Z_{2,h}] &= n^{-1/2}h^{1/2} \left\{ \frac{1}{3} - 2p(1 - p) \right\} f(\xi_p) + O(n^{-1/2}h^{3/2}). \end{aligned}$$

Substitution of the above into (6) and (7) yields (1) and (2), respectively.

Proof of Theorem 1

Write $Z_1^* = n^{1/2}\{\hat{F}_h^*(\hat{\xi}_{p,h}) - p\}$ and $Z_2^* = (nh)^{1/2}\{\hat{F}_h^{*\prime}(\hat{\xi}_{p,h}) - \hat{f}_h(\hat{\xi}_{p,h})\}$. Inversion of the Taylor expansion of $p \equiv \hat{F}_h^*(\hat{F}_h^{*-1}(p))$ about $\hat{F}_h^*(\hat{\xi}_{p,h})$ gives that

$$\hat{F}_h^{*-1}(p) - \hat{\xi}_{p,h} = -\frac{Z_1^*}{n^{1/2} \hat{f}_h(\hat{\xi}_{p,h})} + \frac{Z_1^* Z_2^*}{nh^{1/2} \hat{f}_h(\hat{\xi}_{p,h})^2} + O_p(n^{-1}). \tag{8}$$

Noting that $\mathbb{E}[Z_1^*|\mathcal{X}] = 0$, $\mathbb{E}[Z_1^* Z_2^*|\mathcal{X}] = O_p(h^{1/2})$, and that

$$\mathbb{E}[Z_1^{*2} | \mathcal{X}] = \mathbb{E}[Z_{1,h}^2] + O_p(n^{-1/2} + h^r), \mathbb{E}[Z_1^{*2} Z_2^* | \mathcal{X}] = O_p(n^{-1/2} h^{1/2}),$$

it follows that (8) has conditional mean of order $O_p(n^{-1})$ and conditional variance

$$\begin{aligned} \widehat{\text{VAR}}_{p,0,h} &= \frac{\mathbb{E}[Z_1^{*2}|\mathcal{X}]}{n \hat{f}_h(\hat{\xi}_{p,h})^2} - \frac{2 \mathbb{E}[Z_1^{*2} Z_2^*|\mathcal{X}]}{n^{3/2} h^{1/2} \hat{f}_h(\hat{\xi}_{p,h})^3} + O_p(n^{-3/2}) \\ &= \frac{p(1-p) - 2h\beta f(\xi_p)}{n \hat{f}_h(\hat{\xi}_{p,h})^2} + O_p(n^{-3/2} + n^{-1}h^2). \end{aligned} \tag{9}$$

It follows from (5) and the expansion \mathcal{B} that $\hat{\xi}_{p,h} - \xi_p = O_p(n^{-1/2} + h^r)$ and $\mathcal{B} = O(h^r)$, respectively. We then obtain

$$\begin{aligned} \hat{f}_h(\hat{\xi}_{p,h}) &= \hat{f}_h(\hat{\xi}_{p,h}) - \hat{f}_h(\xi_p) + (nh)^{-1/2} Z_{2,h} + \mathcal{B} + f(\xi_p) \\ &= f(\xi_p) + (nh)^{-1/2} Z_{2,h} + O_p(|\hat{\xi}_{p,h} - \xi_p|) + O_p(h^r) \\ &= f(\xi_p) \left\{ 1 + \frac{Z_{2,h}}{n^{1/2} h^{1/2} f(\xi_p)} \right\} + O_p(n^{-1/2} + h^r). \end{aligned} \tag{10}$$

It follows by substituting (10) into (9) that

$$\widehat{\text{VAR}}_{p,0,h} = \frac{p(1-p)}{nf(\xi_p)^2} - \frac{2h\beta}{nf(\xi_p)} - \frac{2Z_{2,h}p(1-p)}{n^{3/2}h^{1/2}f(\xi_p)^3} + O_p(n^{-3/2} + n^{-1}h^2). \tag{11}$$

Theorem 1 follows immediately by comparing (11) with (1).

Proof of Theorem 2

Write $Z_{3,b} = (nb)^{1/2}\{\hat{\ell}_b(\xi_p) - \mathbb{E}\hat{\ell}_b(\xi_p)\}$, which is asymptotically normal with zero mean according to the central limit theorem under the Lindeberg condition. The expansion (10) can be adapted to yield an analogous expansion for $\hat{\ell}_b(\hat{\xi}_{p,b})$, substitution of which for $f(\xi_p)$ in (1) yields

$$\begin{aligned} \widehat{\text{VAR}}_{p,b,h} &= \frac{p(1-p)}{nf(\xi_p)^2} - \frac{2Z_{3,b}p(1-p)}{n^{3/2}b^{1/2}f(\xi_p)^3} - \frac{b^s(1-2p)C_0}{nf(\xi_p)^2} - \frac{2h\beta}{nf(\xi_p)} \\ &+ O_p(n^{-3/2} + n^{-1}h^2 + n^{-1/2}h^{2r} + h^{3r} + n^{-3/2}b^{-1/2}h \\ &+ n^{-1}b^s h + n^{-1}b^{s+2}), \end{aligned} \tag{12}$$

which, on comparison with (1), implies that $\widehat{\text{VAR}}_{p,b,h}$ has relative error of order $O_p(h^2 + n^{1/2}h^{2r} + nh^{3r} + n^{-1/2}b^{-1/2} + b^s)$, proving the first assertion of Theorem 2. The second assertion follows immediately by minimizing the above order with respect to b .

Proof of Theorem 3

Replacing p by $q + \delta$ with $\delta = o(1)$ and expanding (12) in powers of δ , we obtain

$$\begin{aligned} \widehat{\text{VAR}}_{q,b,h} = & \frac{p(1-p)}{nf(\xi_p)^2} + \frac{\delta(1-2p)}{nf(\xi_p)^2} - \frac{2Z_{3,b}p(1-p)}{n^{3/2}b^{1/2}f(\xi_p)^3} - \frac{b^s(1-2p)C_0}{nf(\xi_p)^2} - \frac{2h\beta}{nf(\xi_p)} \\ & + O_p(n^{-3/2} + n^{-1}h^2 + n^{-1/2}h^{2r} + h^{3r} + n^{-3/2}b^{-1/2}h + n^{-1}b^s h \\ & + n^{-1}b^{s+2} + \delta^2n^{-1} + \delta n^{-3/2}b^{-1/2} + \delta n^{-1}b^s + \delta n^{-1}h). \end{aligned} \quad (13)$$

Setting $\delta = b^s C_0$ in (13) eliminates the bias term $-n^{-1}b^s(1-2p)C_0f(\xi_p)^{-2}$ and minimizes $\text{MSE}(\widehat{\text{VAR}}_{q,b,h})$, whose expansion can be derived from (1) and (13). Substitution of this δ into (13) gives $\widehat{\text{VAR}}_{q,b,h}/\text{Var}(\hat{\xi}_{p,h}) = 1 + O_p(h^2 + n^{1/2}h^{2r} + nh^{3r} + n^{-1/2}b^{-1/2} + b^{s+2})$. The optimal order of b follows by minimizing $\max\{n^{-1/2}b^{-1/2}, b^{s+2}\}$; hence the assertion of Theorem 3 follows.

References

- Azzalini A (1981) A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* 68:326–328
- Beran RJ (1987) Prepivoting to reduce level error of confidence sets. *Biometrika* 74:457–468
- Cai Z, Roussas GG (1997) Smooth estimate of quantiles under association. *Stat Probab Lett* 36:275–287
- Cheung KY, Lee SMS (2005) Variance estimation for sample quantiles using the m out of n bootstrap. *Ann Inst Stat Math* 57:279–290
- Chiu ST (1991) Bandwidth selection for kernel density estimation. *Ann Stat* 19:1883–1905
- De Angelis D, Young GA (1992) Smoothing the bootstrap. *Int Stat Rev* 60:45–56
- De Angelis D, Hall P, Young GA (1993) A note on coverage error of bootstrap confidence intervals for quantiles. *Math Proc Camb Philos Soc* 114:517–531
- Duttweiler DL (1973) The mean-square error of Bahadur's order-statistic approximation. *Ann Stat* 1:446–453
- Efron B (1982) The jackknife, the bootstrap and other resampling plans. SIAM, Philadelphia
- Falk M, Janas D (1992) Edgeworth expansions for studentized and prepivoted sample quantiles. *Stat Probab Lett* 14:13–24
- Hall P, Martin MA (1988a) Exact convergence rate of bootstrap quantile variance estimator. *Probab Theory Relat Fields* 80:261–268
- Hall P, Martin MA (1988b) On bootstrap resampling and iteration. *Biometrika* 75:661–671
- Hall P, Murison RD (1993) Correcting the negativity of high-order kernel density estimators. *J Multivar Anal* 47:103–122
- Hall P, DiCiccio TJ, Romano R (1989) On smoothing and the bootstrap. *Ann Stat* 17:692–704
- Härdle W, Marron JS, Wand MP (1990) Bandwidth choice for density derivatives. *J R Stat Soc, Ser B* 52:223–232
- Ho YHS, Lee SMS (2005) Iterated smoothed bootstrap confidence intervals for population quantiles. *Ann Stat* 33:437–462
- Janas D (1993) A smoothed bootstrap estimator for a studentized sample quantile. *Ann Inst Stat Math* 45:317–329
- Lee SMS, Young GA (1994) Practical higher-order smoothing of the bootstrap. *Stat Sin* 4:445–459

- Maritz JS, Jarrett RG (1978) A note on estimating the variance of the sample median. *J Am Stat Assoc* 73:194–196
- Nadaraya EA (1964) Some new estimates for distribution functions. *Theory Probab Appl* 9:497–500
- Park BU, Marron JS (1990) Comparison of data-driven bandwidth selectors. *J Am Stat Assoc* 85:66–72
- Ralescu SS, Sun S (1993) Necessary and sufficient conditions for the asymptotic normality of perturbed sample quantiles. *J Stat Plann Inference* 35:55–64
- Reiss R-D (1981) Nonparametric estimation of smooth distribution functions. *Scand J Stat* 8:116–119
- Reiss R-D (1989) *Approximate distributions of order statistics with applications to nonparametric statistics*. Springer, New York
- Silverman BW, Young GA (1987) The bootstrap: to smooth or not to smooth? *Biometrika* 74:469–479
- Staniswalis JG (1989) Local bandwidth selection for kernel estimates. *J Am Stat Assoc* 84:284–288
- Sun S, Cordero M (2005) Smooth quantile estimators under strong mixing: necessary and sufficient conditions on bandwidth for weak convergence. *J Stat Plann Inference* 128:397–409
- Youndjé É, Vieu P (2006) A note on quantile estimation for long-range dependent stochastic processes. *Stat Probab Lett* 76:109–116