Ann Inst Stat Math (2013) 65:167–189
DOI 10.1007/s10463-012-0369-6

Improved confidence intervals for quantiles

Yoshihiko Maesono · Spiridon Penev

Received: 30 May 2011 / Revised: 14 March 2012 / Published online: 18 July 2012

Abstract We derive the Edgeworth expansion for the studentized version of the kernel quantile estimator. Inverting the expansion allows us to get very accurate confidence intervals for the pth quantile under general conditions. The results are applicable in practice to improve inference for quantiles when sample sizes are moderate.

Keywords Edgeworth expansion · Quantile · Kernel quantile estimator · Confidence interval

1 Introduction

Although it is a central activity in financial and insurance applications, risk assessment is equally important in environmental regulation, in wildlife management, climate science and also in medicine. The demand for more precise methods of risk evaluation during the last decade has lead to renewed interest in some well established methods of inference. Undoubtedly, nonparametric inference for quantiles is among these methods. The risk measure VaR, widely used in the financial industry, is a tantamount quantile. The coherent risk measures (like conditional value at risk (e.g., Uryasev and Rockafellar 2001) also represent suitable transformations of quantiles. The Lorenz curve (Lorenz 1905), widely used is both economy and ecology, also represents transformation of quantiles. The papers Ogryczak and Ruszczyński (1999, 2002) give an

Y. Maesono (✉)
Faculty of Mathematics, Kyushu University, 744, Motooka, Nishi-ku, Fukuoka 819-0395, Japan
e-mail: maesono@math.kyushu-u.ac.jp

S. Penev
School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW 2052, Australia
e-mail: S.Penev@unsw.edu.au

applied focus on how transformations of quantiles can be used efficiently in the theory of dual stochastic dominance and related mean-risk models.

Although quantile estimation is a very old and well established area of inference and its asymptotic theory is well developed, a new insight is necessary to improve the inferential techniques for small to moderate samples. Very often, banks are using loss data in risk analysis. Even for a 20-year period (which is a large period given the dynamics in the financial industry), the accumulated data would amount to only 240 observations. The number of trading days on the stock market within a year is about 250. Applying first order asymptotic methods for estimation and confidence interval construction for sample sizes such as these may lead to non-precise coverage and any improvement is worth pursuing.

For a continuous random variable X with a cumulative distribution function $F(x)$, having density function $f(x)$ and satisfying $E|X| < \infty$, the p-th quantile is defined as $Q(p) = inf\{x : F(x) \geq p\}$. Given a sample $X_1, X_2, \ldots, X_n$ from F, the simplest estimator of $Q(p)$ is the sample quantile $\widehat{\xi}_{pn}$; that is, the p-th quantile of the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$. Under mild conditions, $n^{1/2}\{\widehat{\xi}_{pn} - Q(p)\}$ is asymptotically normal, with an asymptotic variance $\sigma^2 = \frac{p(1-p)}{f^2(Q(p))}$ (see e.g., Serfling 1980). This asymptotic variance happens to be large in the tails where the density $f(x)$ has small values, thus indicating difficulties in estimating extreme quantiles. Very often in practice, to overcome the difficulties one puts more structure in the model (for example, via extreme value distribution modeling) to improve inference in the tails. If one does not have enough evidence of such a structure and prefers to stay within the nonparametric modeling framework then a way to improve inference would be to consider alternative estimators and, for the purpose of confidence interval construction, to take into account the effect of higher order terms in the Edgeworth expansion. The inversion of such an Edgeworth expansion is supposed to give more precise confidence intervals for the quantile, thus improving the coverage accuracy for small to moderate samples.

One obvious choice of alternative estimator is the kernel quantile estimator

$$\hat{Q}_{p,h_n} = \frac{1}{h_n}\int_0^1 F_n^{-1}(x)K\left(\frac{x-p}{h_n}\right)\mathrm{d}x \tag{1}$$

where $F_n^{-1}(.)$ denotes the inverse of the empirical distribution function, $K(.)$ is a suitably chosen kernel, and h_n is a bandwidth. Using (1), we are weighting up different sampling quantiles by using weights that are determined by the kernel instead of just using only one empirical sample quantile. The choice of bandwidth is a judicious one and serves the purpose to ensure consistency, asymptotic normality, asymptotic bias elimination, and higher order accuracy of the estimator $\hat{Q}_{p,h_n}$. The estimator (1) has been studied by many researchers in the past (for a non-exclusive list, see Falk 1984, 1985; Sheather and Marron 1990; Xiang 1995a,b and the references therein).

Obviously (1) is an L-estimator since it can be written as a weighted sum of the order statistics $X_{(i)}, i = 1, 2, \ldots, n$:

$$\hat{Q}_{p,h_n} = \sum_{i=1}^{n} v_{i,n} X_{(i)}, \ v_{i,n} = \frac{1}{h_n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} K\left(\frac{x-p}{h_n}\right) \mathrm{d}x. \tag{2}$$

The Edgeworth expansion for the asymptotic distribution of the *standardized* version of (1) was recently obtained in Maesono and Penev (2011). In the same paper, the validity of the expansion is demonstrated. In addition, many numerical examples show the superior performance of the Edgeworth approximation in comparison to the simple normal approximation of the distribution of the kernel quantile estimator when sample sizes are small. As a partial case from the above expansion it is seen that, up to first order convergence, the kernel smoothed estimator does not improve upon the sample quantile. This finding coincides with earlier research presented originally in Falk (1985). The improvement with respect to the sample quantile can only show up in higher order terms of the MSE approximation (this phenomenon has been called *deficiency*). Much of the works of Falk (1984), Xiang (1995a,b) and Xiang and Vos (1997) and others has been directed towards showing advantages of the kernel quantile estimator with respect to the deficiency criterion.

In this paper, we consider the Edgeworth expansion of the *studentized* version of the kernel quantile estimator. Since the standard deviation of the estimator itself is typically unknown, it has to be estimated for confidence interval construction. Hence the Edgeworth expansion for the studentized version is necessary and is derived in this paper. Our result is used then to construct improved confidence intervals for quantiles.

2 The difficulties when using sample quantiles

Our approach is opposed to the use of sample quantile estimators and the inversion of their Edgeworth expansion for the same purpose. First, it is well known that due to the discreteness of the distribution of the sample quantile, obtaining a standard Edgeworth expansion for its distribution of an order $O(n^{-1/2})$ is the best one can get. Some modifications of the Edgeworth expansion have been proposed for the discrete case (Reiss 1990) which allow us to define the Edgeworth expansion of order $o(n^{-1/2})$ and beyond also for the discrete case. These expansions depend heavily on the left and right derivatives of F at the true quantile. Yet the resulting Edgeworth expansion for the empirical quantile is virtually impossible to use in practical settings. For example, the paper Janas (1993) gives theoretical derivation for the Edgeworth expansion of the sample quantile by treating both the standardized and the studentized version. It can be seen from the obtained expansions on page 320 that the $O(n^{-1/2})$ term contains a very complicated expression involving unknown characteristics of the density f.

Alternative expansions have been suggested in Hall and Sheather (1988) whereby the density $f(Q(p))$ in σ^2 has been estimated by a simple estimator based on the spacing of two order statistics with indices that are some distance apart. A finer version of the estimator has also been suggested in Hall (1991). The purpose in these cases has been to avoid the difficulty caused by the fact that there is no natural estimator of the variance of the sample quantile (e.g., Hall and Martin 1988).

Now we can spell out our agenda in this paper. The point we are making is that if we use the kernel quantile estimator (with a judicious choice of bandwidth and a suitable kernel) then we can get an estimator whose first order performance is equivalent to the sample quantile but with the additional advantage that it can be easily studentized; the Edgeworth expansion up to order $o(n^{-1/2})$ for the studentized version can be derived and the theoretical quantities involved in the expansion can easily be estimated in a coherent strategy via jackknife type estimators. In addition, a Cornish–Fisher inversion can be used to construct confidence intervals for the quantile. The resulting confidence intervals are more precise than the ones obtained via inversion of the normal approximation and can be utilized to improve the coverage accuracy for moderate sample sizes.

3 Notation and main results

A standard approach in the study of asymptotic properties of L-statistics is to first decompose them into an U-statistic plus a small-order remainder term and to apply asymptotic theory for the main term, that is, the U-statistic.

We stress on the new moments and theoretical difficulties in relation to the derivation of the Edgeworth expansion of order $o(n^{-1/2})$ for the kernel quantile estimator. They are common in both the standardized and studentized case and are related to the fact that the specific requirement on the bandwidth h_n for the sake of eliminating the asymptotic bias triggers the need to include further contributions from the terms of order $O(n^{-1})$ of the appropriately resulting U-statistic. These contributions have to be considered to achieve the desired Edgeworth expansion with remainder of order $o(n^{-1/2})$. The general result is involved, but its application in the most typical and practically relevant case of a symmetric compactly supported kernel $K(\cdot)$ is significantly simplified.

We will be using a compactly supported kernel $K(x)$ on $(-1, 1)$. The kernel is said to be of order m if

$$K \in L^2(-\infty, \infty),\ K^{(m)} \in Lip(\alpha) \text{ for some } \alpha > 0,$$
$$\int_{-1}^{1} K(x)\,\mathrm{d}x = 1,\ \int_{-1}^{1} x^i K(x)\,\mathrm{d}x = 0,\ i = 1, 2, \ldots, m-1,\ \int_{-1}^{1} x^m K(x)\,\mathrm{d}x \neq 0.$$

The basic assumption on the bandwidth is

$$h_n = o(n^{-1/4}) \quad \text{and} \quad \lim_{n\to\infty} (n^{1/4} h_n)^{-k} n^{-\beta} = 0 \tag{3}$$

for any $\beta > 0$ and any integer k. The default bandwidth h_n in our theoretical treatment will be $h_n = n^{-1/4}(\log n)^{-1}$.

Let $\{Y_i\}_{i=1,\ldots,n}$ be independent random variables uniformly distributed on $(0, 1)$ and define

$$\bar{Q}(p) = \frac{1}{h_n}\int_0^1 F^{-1}(x)K\left(\frac{x-p}{h_n}\right)\mathrm{d}x,$$
$$\hat{I}_x(Y_1) = I(Y_1 \le p + h_n x) - (p + h_n x),$$
$$g_{1n}(Y_1) = -\int_{-1}^1 Q'(p+h_n x)K(x)\hat{I}_x(Y_1)\,\mathrm{d}x,$$
$$\sigma_n^2 = Var(g_{1n}(Y_1)),$$
$$d_{1n} = \sigma_n^{-1}n^{-1/2}, \qquad d_{2n} = \sigma_n^{-1}n^{-3/2}h_n^{-1}, \qquad d_{3n} = \sigma_n^{-1}n^{-5/2}h_n^{-2},$$
$$g_{2n}(Y_1, Y_2) = -\int_{-1}^1 Q'(p+h_n x)K'(x)\hat{I}_x(Y_1)\hat{I}_x(Y_2)\,\mathrm{d}x,$$
$$g_{3n}(Y_1, Y_2, Y_3) = -\int_{-1}^1 Q'(p+h_n x)K^{(2)}(x)\hat{I}_x(Y_1)\hat{I}_x(Y_2)\hat{I}_x(Y_3)\,\mathrm{d}x,$$
$$\hat{g}_{1n}(Y_1) = -\frac{1}{2}\int_{-1}^1 Q'(p+h_n x)K^{(2)}(x)E[\hat{I}_x^2(Y_2)]\hat{I}_x(Y_1)\,\mathrm{d}x,$$
$$A_{1n} = \sum_{i=1}^n g_{1n}(Y_i), \qquad A_{2n} = \sum_{C_{n,2}} g_{2n}(Y_i, Y_j), \qquad A_{3n} = \sum_{C_{n,3}} g_{3n}(Y_i, Y_j, Y_k)$$

and $\hat{A}_{1n} = \sum_{i=1}^n \hat{g}_{1n}(Y_i)$.

We will use the notation $o_L(n^{-1/2})$ to indicate that

$$P(|o_L(n^{-1/2})| \ge n^{-1/2}\gamma_n) = o(n^{-1/2})$$

for some $\gamma_n \to 0$, as $n \to \infty$. In particular, we will single out the case where $\gamma_n = \frac{1}{\log n}$ via the notation $o_\ell(n^{-1/2})$, i.e., to indicate that

$$P(|o_\ell(n^{-1/2})| \ge n^{-1/2}(\log n)^{-1}) = o(n^{-1/2})$$

as $n \to \infty$.

Under additional assumption on the tails ($\int [F(x)(1-F(x))]^{1/5}dx < \infty$) and some further smoothness assumptions on the fourth order kernel $K(x)$ (i.e. $m = 4$) we obtained in Maesono and Penev (2011) the following stochastic expansion:

$$\begin{aligned}&\sigma_n^{-1}\sqrt{n}(\hat{Q}_{p,h_n} - \bar{Q}(p))\\ &\quad = d_{1n}A_{1n} + d_{2n}A_{2n} + d_{3n}A_{3n} + d_{3n}(n-1)\hat{A}_{1n} + \frac{\delta}{\sigma_n\sqrt{n}} + o_L(n^{-1/2}). \qquad (4)\end{aligned}$$

Hereby $\delta = Q'(p)(\frac{1}{2} - p) + \frac{1}{2}p(1-p)Q''(p)$. This expansion was further used to obtain the Edgeworth expansion for

$$P(\sqrt{n}(\hat{Q}_{p,h_n} - Q(p)) \le x\sigma_n)$$

in terms of

$$\begin{aligned} e_{1n} &= E[g_{1n}^3(Y_1)], \quad e_{2n} = E[g_{1n}(Y_1)g_{1n}(Y_2)g_{2n}(Y_1, Y_2)], \\ e_{3n} &= E[g_{1n}(Y_2)g_{1n}(Y_3)g_{2n}(Y_1, Y_2)g_{2n}(Y_1, Y_3)], \\ e_{4n} &= E[g_{1n}(Y_1)g_{1n}(Y_2)g_{1n}(Y_3)g_{3n}(Y_1, Y_2, Y_3)], \\ e_{5n} &= E[g_{1n}(Y_1)\hat{g}_{1n}(Y_1)] \quad \text{and} \quad e_{6n} = E[g_{2n}^2(Y_1, Y_2)] \end{aligned} \tag{5}$$

(see Maesono and Penev 2011, Theorem 2). Under the additional conditions $\int_{-1}^{1} K'(x)\,\mathrm{d}x = 0$ and $\int_{-1}^{1} K''(x)\,\mathrm{d}x = 0$ the expansion simplifies. For the purpose of later comparison with the results for the studentized case, we quote this final expansion (see *Remark 3*, Equation (17) in Maesono and Penev 2011):

$$\begin{aligned} &P(\sqrt{n}(\hat{Q}_{p,h_n} - Q(p)) \leq x\sigma_n) \\ &= \Phi(x) - \phi(x)\frac{x^2-1}{6n^{1/2}\sigma_n^3}\left(e_{1n} + \frac{3e_{2n}}{h_n}\right) - \frac{\delta}{\sigma_n\sqrt{n}}\phi(x) + o(n^{-1/2}). \end{aligned} \tag{6}$$

Now we move over to the new results concerning the *studentized* case. We will derive an asymptotic representation of the studentized quantile estimator with residual $o_L(n^{-1/2})$.

To be able to studentize the statistic $\hat{Q}_{p,h_n}$, we need a consistent estimator of σ_n^2. We propose the following jackknife type variance estimator

$$\hat{\sigma}_n^2 = (n-1)\sum_{i=1}^{n}\{\hat{Q}_{p,h_n}^{(i)} - \hat{Q}_{p,h_n}\}^2. \tag{7}$$

Here $\hat{Q}_{p,h_n}^{(i)}$ is based on the resulting $(n-1)$ observations from the original sample after the observation X_i is left out. For the sake of simplicity, we use same bandwidth h_n for $\hat{Q}_{p,h_n}^{(i)}$ and $\hat{Q}_{p,h_n}$.

Lemmas 1 and 2 given below for the studentized case relate to the stochastic expansions of orders $o_L(n^{-1/2})$ (or in particular $o_\ell(n^{-1/2})$).

Lemma 1 *Assume* $\int[F(x)(1-F(x))]^{1/5}\,\mathrm{d}x < \infty$. *Let* K *be a fourth order kernel (i.e.* $m = 4$*) and in addition* $K^{(5)} \in Lip(\alpha), \alpha > 0$ *holds. Let* $Q^{(5)}$ *be uniformly bounded in a neighbourhood of* p $(0 < p < 1)$ *and* $f(Q(p)) > 0$. *Further, choose* h_n *satisfying* (3). *Then we have*

$$\begin{aligned} \hat{\sigma}_n^2 = \sigma_n^2 + n^{-1}h_n^{-2}(2e_{5n} + e_{6n}) + n^{-1}B_{1n} + 2n^{-1}h_n^{-1}\hat{B}_{1n} + 2n^{-2}h_n^{-2}B_{2n} \\ + o_\ell(n^{-1/2}) \end{aligned}$$

where

$$\begin{aligned} &\zeta_{1n}(x) = g_{1n}^2(x) - \sigma_n^2, \quad \hat{\zeta}_{1n}(x) = E[g_{2n}(x, Y)g_{1n}(Y)], \\ &\zeta_{2n}(x, y) = E[g_{3n}(x, y, Y)g_{1n}(Y)] + E[g_{2n}(x, Y)g_{2n}(y, Y)], \\ &B_{1n} = \sum_{i=1}^{n}\zeta_{1n}(Y_i), \quad \hat{B}_{1n} = \sum_{i=1}^{n}\hat{\zeta}_{1n}(Y_i), \quad B_{2n} = \sum_{C_{n,2}}\zeta_{2n}(Y_i, Y_j) \end{aligned}$$

and Y *is an independent copy of* Y_i.

Using this asymptotic representation of the variance estimator, we have the following representation of the studentized kernel quantile estimator.

Lemma 2 *Under the assumptions of Lemma 1 we have*

(i)

$$\sigma_n\hat{\sigma}_n^{-1} = 1 + n^{-1}h_n^{-2}\sigma_n^{-2}\left(\frac{3}{2}\sigma_n^{-2}e_{3n} - e_{5n} - \frac{1}{2}e_{6n}\right) - \frac{1}{2}n^{-1}\sigma_n^{-2}B_{1n}$$
$$-n^{-1}h_n^{-1}\sigma_n^{-2}\hat{B}_{1n} - n^{-2}h_n^{-2}\sigma_n^{-2}D_{2n} + o_\ell(n^{-1/2}) \tag{8}$$

where

$$\begin{aligned}\eta_{2n}(x, y) &= E[g_{3n}(x, y, Y)g_{1n}(Y) + g_{2n}(x, Y)g_{2n}(y, Y)]\\ &\quad -3\sigma_n^{-2}E[g_{2n}(x, Y)g_{1n}(Y)]E[g_{2n}(y, Y)g_{1n}(Y)],\\ D_{2n} &= \sum_{C_{n,2}}\eta_{2n}(Y_i, Y_j).\end{aligned}$$

(ii)

$$\begin{aligned}&\sqrt{n}\hat{\sigma}_n^{-1}\{\hat{Q}_{p,h_n} - \bar{Q}(p)\}\\ &= \nu_n + d_{1n}A_{1n} + d_{2n}\Lambda_{2n} + d_{3n}\Lambda_{3n} + d_{3n}n\hat{\Lambda}_{1n} + o_L(n^{-1/2})\end{aligned} \tag{9}$$

where

$$\begin{aligned}\nu_n &= -\frac{1}{2}\sigma_n^{-3}n^{-1/2}e_{1n} - \sigma_n^{-3}n^{-1/2}h_n^{-1}e_{2n} + \frac{\delta}{\sigma_n\sqrt{n}},\\ \lambda_{2n}(x, y) &= g_{2n}(x, y) - \sigma_n^{-2}\Big\{E[g_{2n}(x, Y)g_{1n}(Y)]g_{1n}(y)\\ &\qquad + E[g_{2n}(y, Y)g_{1n}(Y)]g_{1n}(x)\Big\}\\ &\quad -\frac{1}{2}h_n\sigma_n^{-2}\Big[\{g_{1n}^2(x) - \sigma_n^2\}g_{1n}(y) + \{g_{1n}^2(y) - \sigma_n^2\}g_{1n}(x)\Big],\\ \lambda_{3n}(x, y, z) &= g_{3n}(x, y, z)\\ &\quad -\sigma_n^{-2}\Big\{g_{1n}(x)\eta_{2n}(y, z) + g_{1n}(y)\eta_{2n}(x, z) + g_{1n}(z)\eta_{2n}(x, y)\\ &\quad +\hat{\zeta}_{1n}(x)g_{2n}(y, z) + \hat{\zeta}_{1n}(y)g_{2n}(x, z) + \hat{\zeta}_{1n}(z)g_{2n}(x, y)\Big\},\\ \hat{\lambda}_{1n}(x) &= \hat{g}_{1n}(x) + \sigma_n^{-2}\left\{\left(\frac{3}{2}\sigma_n^{-2}e_{3n} - e_{5n} - \frac{1}{2}e_{6n}\right)g_{1n}(x)\right.\\ &\quad \left. +E[\eta_{2n}(x, Y)g_{1n}(Y)] + E\left[g_{2n}(x, Y)\hat{\zeta}_{1n}(Y)\right]\right\},\\ \Lambda_{2n} &= \sum_{C_{n,2}}\lambda_{2n}(Y_i, Y_j),\ \Lambda_{3n} = \sum_{C_{n,3}}\lambda_{3n}(Y_i, Y_j, Y_k),\ \hat{\Lambda}_{1n} = \sum_{i=1}^{n}\hat{\lambda}_{1n}(Y_i)\end{aligned}$$

and Y is an independent copy of Y_i.

We are now ready to state the theorem.

Theorem 1 *Under same regularity conditions as in Lemma 1 we have*

$$P(\sqrt{n}\hat{\sigma}_n^{-1}(\hat{Q}_{p,h_n} - \bar{Q}(p)) \le x) = S_n(x - \nu_n) + o(n^{-1/2}). \tag{10}$$

Here

$$S_n(x) = \Phi(x) - \phi(x)\left\{\frac{x^2-1}{6n^{1/2}\sigma_n^3}\left(e_{1n} + \frac{3e_{2n}^*}{h_n}\right) + \frac{1}{nh_n^2}\right.$$
$$\left.\times\left(\frac{x}{4\sigma_n^2}\{4e_{5n}^* + e_{6n}^*\} + \frac{x^3-3x}{6\sigma_n^4}\{3e_{3n}^* + e_{4n}^*\} + \frac{x^5 - 10x^3 + 15x}{8\sigma_n^6}{e_{2n}^*}^2\right)\right\}$$

where

$$\begin{aligned}
e_{2n}^* &= E[g_{1n}(Y_1)g_{1n}(Y_2)\lambda_{2n}(Y_1, Y_2)],\\
e_{3n}^* &= E[g_{1n}(Y_2)g_{1n}(Y_3)\lambda_{2n}(Y_1, Y_2)\lambda_{2n}(Y_1, Y_3)],\\
e_{4n}^* &= E[g_{1n}(Y_1)g_{1n}(Y_2)g_{1n}(Y_3)\lambda_{3n}(Y_1, Y_2, Y_3)],\\
e_{5n}^* &= E[g_{1n}(Y_1)\hat{\lambda}_{1n}(Y_1)] \quad \text{and} \quad e_{6n}^* = E[\lambda_{2n}^2(Y_1, Y_2)].
\end{aligned}$$

We note that $e_{2n}^*, e_{3n}^*, \ldots, e_{6n}^*$ can be expressed by $e_{1n}, e_{2n}, \ldots, e_{6n}$ via

$$\begin{aligned}
e_{2n}^* &= -e_{2n} - h_n e_{1n},\\
e_{3n}^* &= \sigma_n^{-2}e_{2n}^2 + o(n^{-1/4}),\\
e_{4n}^* &= -2e_{4n} - 3e_{3n} + 6\sigma_n^{-2}e_{2n}^2,\\
e_{5n}^* &= \frac{7}{2}\sigma_n^{-2}e_{3n} + \sigma_n^{-2}e_{4n} + \frac{1}{2}e_{5n} - \frac{1}{2}e_{6n} - 3\sigma_n^{-4}e_{2n}^2,\\
e_{6n}^* &= e_{6n} - 2\sigma_n^{-2}e_{3n} + 2\sigma_n^{-4}e_{2n}^2 + o(n^{-1/4}).
\end{aligned} \tag{11}$$

The next theorem shows that the chosen bandwidth $h_n = n^{-1/4}(\log n)^{-1}$ allows us to eliminate the asymptotic bias when a kernel of order $m = 4$ is used.

Theorem 2 *Under the same regularity conditions as in Lemma 1 we have*

$$P(\sqrt{n}\hat{\sigma}_n^{-1}(\hat{Q}_{p,h_n} - Q(p)) \le x) = S_n(x - \nu_n) + o(n^{-1/2}). \tag{12}$$

Expanding the function $S_n(x - \nu_n)$ around x and keeping the $O(\frac{1}{\sqrt{n}})$ terms only we can also write (12) as follows:

$$P(\sqrt{n}\hat{\sigma}_n^{-1}(\hat{Q}_{p,h_n} - Q(p)) \le x) = \Phi(x) - \phi(x)\Bigg\{\frac{\delta}{\sigma_n\sqrt{n}} + \frac{-2x^2-1}{6n^{1/2}\sigma_n^3}e_{1n} + \frac{-x^2-1}{2n^{1/2}\sigma_n^3 h_n}e_{2n} + \frac{1}{nh_n^2}\Bigg[\frac{x}{4\sigma_n^2}(2e_{5n} - e_{6n}) + \frac{-x^3+3x}{2\sigma_n^4}e_{3n} - \frac{x^3}{3\sigma_n^4}e_{4n} + \frac{x^5+2x^3-41x}{8\sigma_n^6}e_{2n}^2\Bigg]\Bigg\} + o(n^{-1/2}). \quad (13)$$

The derivations that lead from (12) to (13) are omitted to save space.

Remark 1 If we use a kernel satisfying the conditions $\int_{-1}^{1} K'(x)\,\mathrm{d}x = 0$ and $\int_{-1}^{1} K''(x)\,\mathrm{d}x = 0$, we can show that $e_{3n},\ e_{4n},\ e_{5n},\ e_{6n}$ and e_{2n}^2 are all $O(h_n)$. For instance, it is easy to see that

$$\begin{aligned} &E[\hat{I}_{x_1}(Y_2)\hat{I}_{x_2}(Y_3)\hat{I}_{x_3}(Y_1)\hat{I}_{x_3}(Y_2)\hat{I}_{x_4}(Y_1)\hat{I}_{x_4}(Y_3)] \\ &\quad = p^3(1-p)^3 + h_n b_n(x_1, x_2, x_3, x_4) \end{aligned}$$

where b_n is a bounded function of $x_1,\ x_2,\ x_3$ and x_4. Using the Taylor expansion of $Q'(.)$, we can show that $e_{3n} = O(h_n)$. Then a simplification occurs in (13) and we get just

$$\begin{aligned} &P(\sqrt{n}\hat{\sigma}_n^{-1}(\hat{Q}_{p,h_n} - Q(p)) \le x) \\ &\quad = \Phi(x) - \phi(x)\left\{\frac{\delta}{\sigma_n\sqrt{n}} + \frac{-2x^2-1}{6n^{1/2}\sigma_n^3}e_{1n} + \frac{-x^2-1}{2n^{1/2}\sigma_n^3 h_n}e_{2n}\right\} + o(n^{-1/2}). \quad (14) \end{aligned}$$

The two expressions (14) and (6) can be compared now to appreciate the difference in the expansions in the standardized and in the studentized case.

4 Numerical comparisons

In the paper Maesono and Penev (2011), we have demonstrated on a range of examples the great accuracy of the Edgeworth expansion and the sensible improvement effect in the approximation of the distribution of the standardized kernel quantile estimator when we move over from the normal to the Edgeworth approximation of order $o(n^{-1/2})$. This effect could be seen even for sample sizes such as $n = 15, 30, 40, 50$. For larger n the two approximations become closer to each other with the Edgeworth expansion dominating the normal virtually uniformly in the whole range of values of the argument.

The most significant application of the observed phenomenon is in constructing accurate confidence interval for the quantile $Q(p)$ when the sample size is small to moderate. For a given level α, instead of constructing it in a symmetric way as

$\hat{Q}_{p,h_n} \pm z_{\alpha/2}\hat{\sigma}_n/\sqrt{n}$, one can improve the coverage accuracy by using $(\hat{Q}_{p,h_n} + c_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}, \hat{Q}_{p,h_n} + c_{\alpha/2}\hat{\sigma}_n/\sqrt{n})$ with the quantile values $c_{1-\alpha/2}$ and $c_{\alpha/2}$ obtained by inverting numerically the Edgeworth approximation. Similar approach can be used for constructing one-sided confidence intervals. We did not pursue these avenues in Maesono and Penev (2011) since, for the approach to be really practicable, we need the Edgeworth expansion for the *studentized* kernel quantile estimator. The results of this paper can help us to develop the relevant methodology.

We should note however that in the studentized case, we can not expect such precise results to hold for sample sizes as small as the ones that were suitable in the standardised case. Indeed, estimating the inversion of the Edgeworth approximation in the studentized case involves estimating quantities such as δ, σ_n and e_{1n}, e_{2n}. Looking at their definition we see that their estimators are bound to have large variability since their estimators involve high order statistics. The variability will be increased when the quantile is to be estimated is at the more extreme end. Our numerical experimentation shows that sample sizes less than 200 may give misleading results especially for the Edgeworth-based method.

We could choose different compactly supported fourth order kernels satisfying the requirements of our theorems but the effect of the kernel is not that crucial as long as it satisfies the requirement that the sign of the quantity ψ in (15) is positive:

$$\psi = \int_{-1}^{1} yK(y)M(y)\,\mathrm{d}y,\ M(y) = \int_{-1}^{y} K(x)\,\mathrm{d}x > 0. \tag{15}$$

The latter requirement implies that under regularity conditions the kernel quantile estimator is better than the sample quantile in terms of deficiency (see, e.g. Falk 1984).

Here, we only present results obtained with the following symmetric fourth order kernel first suggested by Müller (1984):

$$K(x) = \frac{315}{512}(11x^8 - 36x^6 + 42x^4 - 20x^2 + 3)I(|x| \le 1). \tag{16}$$

Using the kernel (16) brings about significant simplification due to the fact that the integral of its first and second derivatives is equal to zero. See Remark 1. As a result, the simple form of the Edgeworth expansion (14) can be used in our numerical work.

We need consistent estimators of $e_{1n}, e_{2n}/h_n, \delta$ and σ_n^2 in order to implement (14). Based on a popular idea of Hinkley and Wei (1984) the following jackknife estimators can be used:

Denoting $\hat{h}_1(i) = \hat{Q}_{p,h_n} - \hat{Q}^{(i)}_{p,h_n}$, a consistent estimator of e_{1n} is given by

$$\hat{e}_{1n} = \frac{(n-1)^3}{n}\sum_{i=1}^{n} \hat{h}_1^3(i).$$

Next we define

$$\hat{h}_2(i, j) = n\hat{Q}_{p,h_n} - (n-1)\{\hat{Q}^{(i)}_{p,h_n} + \hat{Q}^{(j)}_{p,h_n}\} + (n-2)\hat{Q}^{(i,j)}_{p,h_n}$$

whereby $\hat{Q}^{(i,j)}_{p,h_n}$ denotes the kernel quantile estimator calculated by using the original sample from which the ith and jth observation have been deleted. As noted earlier, we use the same bandwidth also in the definition of $\hat{Q}^{(i)}_{p,h_n}$ and of $\hat{Q}^{(i,j)}_{p,h_n}$. Then the consistent jackknife estimator of e_{2n}/h_n is given by

$$\frac{\hat{e}_{2n}}{h_n} = \frac{(n-1)^2}{n}\sum_{i=1}^{n}\sum_{i\neq j}^{n}\hat{h}_1(i)\hat{h}_1(j)\hat{h}_2(i, j).$$

A consistent estimator $\hat{\delta}$ of the second order bias δ is given by

$$\hat{\delta} = (n-1)\sum_{i=1}^{n}(\hat{Q}^{(i)}_{p,h_n} - \hat{Q}_{p,h_n}).$$

The consistent variance estimator $\hat{\sigma}_n^2$ is given in (7).

Regarding the choice of the bandwidth, we emphasize that our approach avoids (at least in asymptotic terms) the choice of the bandwidth and any $h_n = cn^{-1/4}(\log n)^{-1}$ with $c > 0$ should be fine. Of course finite sample performance can be influenced by the choice of c and some tuning might be needed for the range of sample sizes of primary interest. In our simulations, we have used the same value $h_n = n^{-1/4}(\log_{10} n)^{-1}$ for all scenarios and all sample sizes with which we have experimented.

Tables 1 and 2 illustrate the resulting coverage probabilities in comparison to the nominal coverage probabilities when calculating the lower confidence interval (Table 1) and a symmetric confidence interval (Table 2). The probabilities have been approximated in a simulation study where 50,000 replications have been performed and the empirical ratio of coverages of the true quantile has been adopted as the "true". We have found out that at the above number of replications a stabilization occurs and increasing the replications to, say, 100,000, only changes the last digit. Both tables represent the result of confidence interval construction for the 90th quantile point, i.e., $p = .90$, for three scenarios. We have simulated with other ranges of values of p, too but for the purpose of brevity we do not include other results in this paper.

We simulate data to three different scenarios. These are: a particular gamma distribution (Chi-square with 4 degrees of freedom), the standard exponential and the standard normal distribution. We assume that the shape of the distribution is unknown and apply the methods for estimation and confidence interval construction from the current paper. We point out that although the true parameter values for δ, for example, as well as of other relevant parameter values do involve information from higher order derivatives of the density of the data, none of these has been used in the estimation, all quantities of interest have been estimated via the relevant jackknife estimator.

Examination of Tables 1 and 2 shows that the "true" coverage probabilities approach the nominal probabilities when the sample size increases. Since the confidence intervals are asymptotic in spirit, this fact demonstrates the consistency of the procedure. It is also seen in both Tables 1 and 2 that for the range of sample sizes considered

Table 1 Lower confidence intervals

Distribution	n	Nominal coverage					
		90 %		95 %		99 %	
		Normal	Edge	Normal	Edge	Normal	Edge
Gamma	200	0.85426	0.87672	0.90574	0.93016	0.96218	0.97692
Gamma	250	0.85404	0.87820	0.90876	0.92986	0.96350	0.97824
Gamma	350	0.86680	0.88640	0.91866	0.93838	0.96982	0.98110
Gamma	500	0.87478	0.89156	0.92754	0.94362	0.97588	0.98538
Exp(1)	200	0.84748	0.87310	0.90184	0.92850	0.95990	0.97674
Exp(1)	250	0.84974	0.87462	0.90576	0.93066	0.96274	0.97810
Exp(1)	350	0.86520	0.88640	0.91772	0.93894	0.96984	0.98308
Exp(1)	500	0.87294	0.89144	0.92418	0.94270	0.97392	0.98484
Normal	200	0.86684	0.88604	0.91726	0.93580	0.96976	0.98100
Normal	250	0.86436	0.88372	0.91780	0.93584	0.97076	0.98072
Normal	350	0.87338	0.88992	0.92334	0.94000	0.97388	0.98228
Normal	500	0.88168	0.89626	0.93078	0.94322	0.97646	0.98434

Table 2 Symmetric confidence intervals

Distribution	n	Nominal coverage					
		90 %		95 %		99 %	
		Normal	Edge	Normal	Edge	Normal	Edge
Gamma	200	0.86710	0.88094	0.91796	0.93136	0.96758	0.97612
Gamma	250	0.87210	0.88294	0.92150	0.93460	0.96972	0.97738
Gamma	350	0.87702	0.88812	0.92698	0.93838	0.97422	0.98114
Gamma	500	0.88524	0.89436	0.93460	0.94316	0.97952	0.98406
Exp(1)	200	0.86638	0.88114	0.91662	0.93080	0.96634	0.97542
Exp(1)	250	0.86884	0.88364	0.91986	0.93352	0.96902	0.97786
Exp(1)	350	0.87860	0.88962	0.92900	0.94044	0.97524	0.98158
Exp(1)	500	0.88262	0.89324	0.93246	0.94136	0.97830	0.98336
Normal	200	0.87446	0.88686	0.92392	0.93510	0.97316	0.97894
Normal	250	0.87516	0.88678	0.92698	0.93704	0.97456	0.97892
Normal	350	0.87918	0.89040	0.92918	0.93950	0.97690	0.98224
Normal	500	0.88406	0.89166	0.93322	0.94148	0.97862	0.983100

($n = 200$–500), the confidence intervals based on inverting the Edgeworth approximation *always* outperform the ones based on inverting the normal approximation. Comparison within each of the table also demonstrates that for the normal distribution scenario, both constructions give slightly better coverages. We do not have a rational explanation for this phenomenon, however.

When doing a cross-comparison between Tables 1 and 2, we observe that the symmetric confidence intervals have slightly better coverage (closer to the nominal) in comparison to the lower confidence intervals. This effect is also observed uniformly across all corresponding cells in the tables.

Remark 2 There is an interesting modification of the estimator for the case where a wavelet-based kernel of two arguments could be used instead. If $\varphi(x)$ is certain father wavelet and $\varphi_{js}(x) = 2^{s/2}\varphi(2^s x - j)$ denotes its dilation and translation, then the kernel of two arguments $K_s^*(x, y) = \sum_{j=-\infty}^{\infty} \varphi_{sj}(x)\varphi_{sj}(y)$ can be defined and could be used for alternative kernel quantile estimator construction. The estimator itself can be defined as

$$\hat{Q}_{p,s}^* = \frac{1}{n}\sum_{i=1}^{n} X_{(i)} K_s^* \left(\frac{i-1}{n}, p\right)$$

whereby now s replaces the role of the bandwidth. The coiflet-type wavelets satisfy the moment conditions required from a higher order kernel and could be used as alternatives to the kernel (16). However, although being almost symmetric they are not strictly symmetric. Besides, we expect the numerical effects by using this approach to be comparable to the effects of using the kernel (16) and we have not pursued the approach numerically in the current paper.

5 Proofs

We believe that the following two remarks will allow the reader to follow easier the main steps in the derivations in the paper.

Remark 3 If for a random variable R it holds that $E|R|^c = O(n^{-1/2-c/2-\delta})$ for some $c > 0$ and $\delta > 0$, then we have

$$P\{|R| > n^{-1/2}\log n^{-1}\} = o(n^{-1/2}),$$

i.e., $R = o_\ell(n^{-1/2})$. Thus we can ignore R when discussing asymptotic expansion up to order $n^{-1/2}$. See, e.g., Maesono (1997), p. 64.

Remark 4 Our proofs are derived in two steps. In the first step, using moment evaluation technique (20) for the Hoeffding decomposition (H-decomposition) (18) described below, a stochastic expansion of statistic of interest is shown to be decomposed into main term and a remainder of order $o_L(n^{-1/2})$. This fact allows us to ignore the effect of the remainder when the asymptotic distribution up to order $o(n^{-1/2})$ of the statistic is derived in the second step. The justification when to ignore the remainder R is discussed in Remark 3. Regarding the asymptotic distributions of the products involved in the definition of the studentized statistic, we use a standard technique of evaluating the terms in such a product. A comprehensive illustration of this technique with detailed proofs can be found, e.g., in Maesono (1997).

We use some general moment evaluations related to the H-decomposition. For independent identically distributed random variables $X_1, \ldots, X_n$ and a function $\nu(x_1, \ldots, x_r)$, centered ($E[\nu(X_1, \ldots, X_r)] = 0$), and symmetric in its arguments, we define

$$\rho_1(x_1) = E[\nu(x_1, X_2, \ldots, X_r)],$$
$$\rho_2(x_1, x_2) = E[\nu(x_1, x_2, \ldots, X_r)] - \rho_1(x_1) - \rho_1(x_2), \ldots,$$

and

$$\rho_r(x_1, x_2, \ldots, x_r) = \nu(x_1, x_2, \ldots, x_r) - \sum_{k=1}^{r-1} \sum_{C_{r,k}} \rho_k(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$$

where $\sum_{C_{r,k}}$ indicates that the summation is taken over all integers $i_1, \ldots, i_k$ satisfying $1 \le i_1 < \cdots < i_k \le r$. Then it holds

$$E[\rho_k(X_1, \ldots, X_k)|X_1, \ldots, X_{k-1}] = 0 \ \ a.s. \tag{17}$$

and

$$\sum_{C_{n,r}} \nu(X_{i_1}, \ldots, X_{i_r}) = \sum_{k=1}^{r} \binom{n-k}{r-k} A_k \tag{18}$$

where

$$A_k = \sum_{C_{n,k}} \rho_k(X_{i_1}, \ldots, X_{i_k}). \tag{19}$$

Equation (18) represents the H-decomposition of U-statistic with a sum of forward martingales $\{A_k\}_{n \ge k}$, for $k = 1, 2, \ldots, r$. Using Eq. (17) and the moment evaluations of martingales (Dharmadhikari et al. 1968), we have upper bounds of the absolute moments of A_k. For $q \ge 2$, if $E|\nu(X_1, \ldots, X_r)|^q < \infty$, there exists a positive constant C, which may depend on ν and F but not on n, such that

$$E|A_k|^q \le Cn^{qk/2} E|\rho_k(X_{i_1}, \ldots, X_{i_k})|^q. \tag{20}$$

Since a product of U-statistics with different kernels is a linear function of $U-$statistics, we can use (20) to continue the evaluation.

Proof of Lemma 1 From the definition, we have

$$\hat{Q}^{(i)}_{p,h_n} - \hat{Q}_{p,h_n} = -\int_{-\infty}^{\infty} \left\{ \int_{[F_n(x)-p]/h_n}^{[F_n^{(i)}(x)-p]/h_n} K(s)ds \right\} \mathrm{d}x$$

where $F_n^{(i)}(x) = \sum_{j\neq i}^n I(X_j \leq x)/(n-1)$. We now take Taylor expansion of the inner integral around the point $(F_n(x)-p)/h_n$. Further, we define for $j = 1, \dots, 6$

$$L_{jn}^{(i)} = -\frac{1}{j!}\int_{-\infty}^{\infty} K^{(j-1)}\Big(\frac{F_n(x)-p}{h_n}\Big)\Big(\frac{F_n^{(i)}(x)-F_n(x)}{h_n}\Big)^j \,\mathrm{d}x$$

and denote the remainder term by

$$R_n^{(i)} = -\int_{-\infty}^{\infty}\bigg(\int_{[F_n(x)-p]/h_n}^{[F_n^{(i)}(x)-p]/h_n} K(u)du\bigg)\,\mathrm{d}x - \sum_{j=1}^{6} L_{jn}^{(i)}.$$

Then we have

$$(n-1)\sum_{i=1}^{n}\{\hat{Q}_{p,h_n}^{(i)} - \hat{Q}_{p,h_n}\}^2 = (n-1)\sum_{i=1}^{n}\left\{\sum_{j=1}^{6} L_{jn}^{(i)} + R_n^{(i)}\right\}^2. \tag{21}$$

We see directly that

$$F_n^{(i)}(x) - F_n(x) = -\frac{1}{n-1}I_x^*(X_i) + \frac{1}{n(n-1)}\sum_{j=1}^{n} I_x^*(X_j)$$

holds where $I_x^*(X_i) = I(X_i \leq x) - F(x)$.

We start evaluating the right hand side of (21). Our first step is to approximate the expression $(n-1)\sum_{i=1}^n \{L_{1n}^{(i)}\}^2$.

To this end, we first expand in Taylor series each of the expressions $L_{1n}^{(i)}$ around the point $(F(x)-p)/h_n$. Let us define for $k = 1, \dots, 6$

$$M_{kn}^{(i)} = -\frac{1}{(k-1)!}\int_{-\infty}^{\infty} K^{(k-1)}\Big(\frac{F(x)-p}{h_n}\Big) \times\Big(\frac{F_n(x)-F(x)}{h_n}\Big)^{k-1}\Big(\frac{F_n^{(i)}(x)-F_n(x)}{h_n}\Big)\,\mathrm{d}x$$

and

$$r_n^{(i)} = L_{1n}^{(i)} - \sum_{k=1}^{6} M_{kn}^{(i)}.$$

From the definition, we have

$$
\begin{aligned}
(M_{1n}^{(i)})^2 &= \left\{ \int_{-\infty}^{\infty} K\Big(\frac{F(x)-p}{h_n}\Big)\Big(\frac{F_n^{(i)}(x)-F_n(x)}{h_n}\Big)\,\mathrm{d}x \right\}^2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_n^{-2} K\Big(\frac{F(x)-p}{h_n}\Big) K\Big(\frac{F(y)-p}{h_n}\Big)\Big\{-\frac{1}{n-1} I_x^*(X_i) \\
&\quad + \frac{1}{n(n-1)}\sum_{j=1}^{n} I_x^*(X_j)\Big\}\Big\{-\frac{1}{n-1} I_y^*(X_i) + \frac{1}{n(n-1)}\sum_{j=1}^{n} I_y^*(X_j)\Big\}\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}
$$

Using Shorack and Wellner (1986), we have

$$
\begin{aligned}
&(n-1)\sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_n^{-2} K\Big(\frac{F(x)-p}{h_n}\Big) K\Big(\frac{F(y)-p}{h_n}\Big)\frac{1}{(n-1)^2} \\
&\quad \times I_x^*(X_i) I_y^*(X_i)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\int_{-1}^{1}\int_{-1}^{1} Q'(p+h_n x)Q'(p+h_n y)K(x)K(y)\hat{I}_x(Y_i)\hat{I}_y(Y_i)\,\mathrm{d}x\,\mathrm{d}y \\
&= \sigma_n^2 + \frac{1}{n}\sum_{i=1}^{n}\{g_{1n}^2(Y_i)-\sigma_n^2\} + o_\ell(n^{-1/2}).
\end{aligned}
$$

We can show that all other terms in the expansion of $(n-1)\sum_{i=1}^{n}(M_{1n}^{(i)})^2$ are of order $o_\ell(n^{-1/2})$, as in Remark 3, by noting that

$$
\begin{aligned}
&\left| E\left[\prod_{\ell=1}^{4}\sum_{i=1}^{n}\hat{I}_{x_\ell}(Y_i)\right]\right| \\
&\quad \le \left\{\prod_{\ell=1}^{4} E\left|\sum_{i=1}^{n}\hat{I}_{x_\ell}(Y_i)\right|^4\right\}^{1/4} = O(n^2)\prod_{\ell=1}^{4}\Big\{F(x_\ell)(1-F(x_\ell))\Big\}.
\end{aligned}
$$

Thus we have

$$
(n-1)\sum_{i=1}^{n}(M_{1n}^{(i)})^2 = \sigma_n^2 + \frac{1}{n}\sum_{i=1}^{n}\{g_{1n}^2(Y_i)-\sigma_n^2\} + o_\ell(n^{-1/2}). \tag{22}
$$

Similarly as for $(M_{1n}^{(i)})^2$, we can show that

$$
\begin{aligned}
&2(n-1)\sum_{i=1}^{n} M_{1n}^{(i)} M_{2n}^{(i)} \\
&\quad = 2n^{-2}h_n^{-1}\sum_{C_{n,2}} g_{2n}(Y_i,Y_j)\{g_{1n}(Y_i)+g_{1n}(Y_j)\} + o_\ell(n^{-1/2}).
\end{aligned} \tag{23}
$$

From direct computation, we have

$$
\begin{aligned}
&2(n-1)\sum_{i=1}^{n} M_{1n}^{(i)} M_{3n}^{(i)} \\
&\quad = 2n^{-3}h_n^{-2}\sum_{C_{n,3}} g_{3n}(Y_i, Y_j, Y_k)\{g_{1n}(Y_i)+g_{1n}(Y_j)+g_{1n}(Y_k)\} \\
&\qquad +2n^{-1}h_n^{-2}e_{5n} + o_\ell(n^{-1/2}).
\end{aligned} \tag{24}
$$

Using the same method, we can show that

$$
\begin{aligned}
(n-1)\sum_{i=1}^{n}\{M_{2n}^{(i)}\}^2 &= 2n^{-3}h_n^{-2}\sum_{C_{n,3}}\{g_{2n}(Y_i, Y_j)g_{2n}(Y_i, Y_k) + g_{2n}(Y_i, Y_j)g_{2n}(Y_j, Y_k) \\
&\quad + g_{2n}(Y_i, Y_k)g_{2n}(Y_j, Y_k)\} + n^{-1}h_n^{-2}e_{6n} + o_\ell(n^{-1/2}).
\end{aligned} \tag{25}
$$

We can also show that the remaining terms

$$
2(n-1)\sum_{i=1}^{n} M_{jn}^{(i)} M_{kn}^{(i)} = o_\ell(n^{-1/2}) \quad \text{for } j+k \geq 5,
$$

and

$$
2(n-1)\sum_{i=1}^{n} r_n^{(i)} M_{jn}^{(i)} = o_\ell(n^{-1/2})(j=1,\ldots,6).
$$

Details are omitted to save space. However, we note that the Lipschitz continuity condition $K^{(5)} \in Lip(\alpha)$ is required when evaluating the term $E|r_n^{(i)}|$, as a part of the evaluation of $2(n-1)\sum_{i=1}^{n} r_n^{(i)} M_{jn}^{(i)}$.

Our next step in the evaluation of the right hand side in (21) is to show that the terms $(n-1)\sum_{i=1}^{n} L_{jn}^{(i)} L_{kn}^{(i)}$ $(j+k \geq 3)$ and $(n-1)\sum_{i=1}^{n} L_{jn}^{(i)} R_n^{(i)}$ $(j \geq 1)$ are all $o_\ell(n^{-1/2})$. The details of these evaluations are omitted to save space.

Combining the above all evaluations from the contributions in the right hand side of (21), we can show that

$$
\begin{aligned}
&(n-1)\sum_{i=1}^{n}\{\hat{Q}_{p,h_n}^{(i)} - \hat{Q}_{p,h_n}\}^2 \\
&\quad = \sigma_n^2 + n^{-1}h_n^{-2}(2e_{5n}+e_{6n}) + \frac{1}{n}\sum_{i=1}^{n}\{g_{1n}^2(Y_i) - \sigma_n^2\} \\
&\qquad +2n^{-2}h_n^{-1}\sum_{C_{n,2}} g_{2n}(Y_i, Y_j)\{g_{1n}(Y_i)+g_{1n}(Y_j)\}
\end{aligned}
$$

$$
\begin{aligned}
&+2n^{-3}h_n^{-2}\sum_{C_{n,3}} g_{3n}(Y_i,Y_j,Y_k)\{g_{1n}(Y_i)\\
&+g_{1n}(Y_j)+g_{1n}(Y_k)\}+2n^{-3}h_n^{-2}\\
&\times\sum_{C_{n,3}}\{g_{2n}(Y_i,Y_j)g_{2n}(Y_i,Y_k)+g_{2n}(Y_i,Y_j)g_{2n}(Y_j,Y_k)\\
&+g_{2n}(Y_i,Y_k)g_{2n}(Y_j,Y_k)\}+o_\ell(n^{-1/2}).
\end{aligned}
$$

Using the H-decomposition, we will obtain the form of the asymptotic U-statistic. Let us introduce the statistics H_1 and H_2 via the decomposition

$$
\begin{aligned}
&2n^{-2}h_n^{-1}\sum_{C_{n,2}} g_{2n}(Y_i,Y_j)\{g_{1n}(Y_i)+g_{1n}(Y_j)\}\\
&\quad=2n^{-2}(n-1)h_n^{-1}H_1+2n^{-2}h_n^{-1}H_2.
\end{aligned}
$$

Using moment evaluation again, we can show that

$$
E\left|2n^{-2}h_n^{-1}H_2\right|^4=O(n^{-4}h_n^{-4})=O(n^{-1/2-2-1/2}(n^{1/4}h_n)^{-4}).
$$

Thus we have $2n^{-2}h_n^{-1}H_2=o_\ell(n^{-1/2})$. Therefore we have

$$
\begin{aligned}
&2n^{-2}h_n^{-1}\sum_{C_{n,2}} g_{2n}(Y_i,Y_j)\{g_{1n}(Y_i)+g_{1n}(Y_j)\}\\
&\quad=2n^{-1}h_n^{-1}\sum_{i=1}^{n}E[g_{2n}(Y_i,Y)g_{1n}(Y)|Y_i]+o_\ell(n^{-1/2})
\end{aligned}
$$

whereby Y denotes an independent copy of Y_i. Similarly, using the H-decomposition, we can show that

$$
\begin{aligned}
&2n^{-3}h_n^{-2}\sum_{C_{n,3}} g_{3n}(Y_i,Y_j,Y_k)\{g_{1n}(Y_i)+g_{1n}(Y_j)+g_{1n}(Y_k)\}\\
&\quad=2n^{-2}h_n^{-2}\sum_{C_{n,2}}E[g_{3n}(Y_i,Y_j,Y)g_{1n}(Y)|Y_i,Y_j]+o_\ell(n^{-1/2})
\end{aligned}
$$

and

$$
\begin{aligned}
&2n^{-3}h_n^{-2}\sum_{C_{n,3}}\{g_{2n}(Y_i,Y_j)g_{2n}(Y_i,Y_k)+g_{2n}(Y_i,Y_j)g_{2n}(Y_j,Y_k)\\
&\qquad+g_{2n}(Y_i,Y_k)g_{2n}(Y_j,Y_k)\}\\
&\quad=2n^{-2}h_n^{-2}\sum_{C_{n,2}}E[g_{2n}(Y_i,Y)g_{2n}(Y_j,Y)|Y_i,Y_j]+o_\ell(n^{-1/2}).
\end{aligned}
$$

Thus we have the desired result. □

Proof of Lemma 2 *Part (i)* Using the Taylor expansion of

$$\sigma_n\hat{\sigma}_n^{-1} = \Big(1 + \frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2}\Big)^{-1/2}$$

we have

$$\Big(1 + \frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2}\Big)^{-1/2} = 1 - \frac{\hat{\sigma}_n^2 - \sigma_n^2}{2\sigma_n^2} + \frac{3}{8}\Big(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2}\Big)^2 + R_n^* \tag{26}$$

where

$$R_n^* = -\frac{15}{48}\Big(1+\Theta\Big)^{-7/2}\Big(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2}\Big)^3$$

and $0 \le |\Theta| \le \left|\frac{\hat{\sigma}_n^2-\sigma_n^2}{\sigma_n^2}\right|$ $a.s.$ First, we will obtain an approximation of $\frac{3}{8}\Big(\frac{\hat{\sigma}_n^2-\sigma_n^2}{\sigma_n^2}\Big)^2$. From the definition, we have

$$4n^{-2}h_n^{-2}\hat{B}_{1n}^2 = 4n^{-2}h_n^{-2}\sum_{i=1}^{n}\hat{\zeta}_{1n}^2(Y_i) + 8n^{-2}h_n^{-2}\sum_{C_{n,2}}\hat{\zeta}_{1n}(Y_i)\hat{\zeta}_{1n}(Y_j).$$

Using the moment evaluation, it is easy to see that

$$4n^{-2}h_n^{-2}\sum_{i=1}^{n}\hat{\zeta}_{1n}^2(Y_i) = 4n^{-1}h_n^{-2}E[\hat{\zeta}_{1n}^2(Y_i)] + o_\ell(n^{-1/2}).$$

Thus we have

$$\begin{aligned} 4n^{-2}h_n^{-2}\hat{B}_{1n}^2 &= 4n^{-1}h_n^{-2}E[\hat{\zeta}_{1n}^2(Y_i)] + 8n^{-2}h_n^{-2}\sum_{C_{n,2}}\hat{\zeta}_{1n}(Y_i)\hat{\zeta}_{1n}(Y_j) + o_\ell(n^{-1/2}) \\ &= 4n^{-1}h_n^{-2}e_{3n} + 8n^{-2}h_n^{-2}\sum_{C_{n,2}}\hat{\zeta}_{1n}(Y_i)\hat{\zeta}_{1n}(Y_j) + o_\ell(n^{-1/2}). \end{aligned} \tag{27}$$

Along the same lines we can show that

$$n^{-2}B_{1n}^2 = o_\ell(n^{-1/2}). \tag{28}$$

Furthermore, we have that

$$\begin{aligned} 4n^{-2}h_n^{-1}B_{1n}\hat{B}_{1n} &= 4n^{-2}h_n^{-1}\sum_{i=1}^{n}\{g_{1n}^2(Y_i) - \sigma_n^2\}E[g_{2n}(Y_i, Y)g_{1n}(Y)|Y_i] \\ &\quad +4n^{-2}h_n^{-1}\sum_{C_{n,2}}\Big[\{g_{1n}^2(Y_i) - \sigma_n^2\}E[g_{2n}(Y_j, Y)g_{1n}(Y)|Y_j] \end{aligned}$$

$$+\{g_{1n}^2(Y_j) - \sigma_n^2\}E[g_{2n}(Y_i, Y)g_{1n}(Y)|Y_i]\Big]$$
$$= o_\ell(n^{-1/2}).$$

Proceeding along the same lines, we can evaluate the remaining terms in $\frac{3}{8}(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2})^2$. The final result is

$$\frac{3}{8}\Big(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2}\Big)^2 = \frac{3}{2}n^{-1}h_n^{-2}\sigma_n^{-4}e_{3n} + 3n^{-2}h_n^{-2}\sigma_n^{-4}\sum_{C_{n,2}}\hat{\zeta}_{1n}(Y_i)\hat{\zeta}_{1n}(Y_j) + o_\ell(n^{-1/2}).$$

Finally, we will show that $R_n^* = o_\ell(n^{-1/2})$ holds. Similarly to the proof of Lemma 3 of Maesono (1997, p.78), it is sufficient to prove that $(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2})^3 = o_\ell(n^{-1/2})$. Using the H-decomposition, we can show that

$$\begin{aligned}
&n^{-3}h_n^{-3}\hat{B}_{1n}^3 \\
&\quad = n^{-1}h_n^{-1}\hat{B}_{1n}\{n^{-1}h_n^{-2}e_{3n} + 2n^{-2}h_n^{-2}\sum_{C_{n,2}}\hat{\zeta}_{1n}(Y_i)\hat{\zeta}_{1n}(Y_j) + o_\ell(n^{-1/2})\} \\
&\quad = o_\ell(n^{-1/2}).
\end{aligned}$$

Also, we can show that the remaining terms in $(\frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2})^3$ are all $o_\ell(n^{-1/2})$. Putting everything together in the right hand side of (26) we get (8).

Part (ii) Next we will obtain the asymptotic representation of the studentized quantile estimator. Observing that

$$\sqrt{n}\hat{\sigma}_n^{-1}\{\hat{Q}_{p,h_n} - \bar{Q}(p)\} = \sqrt{n}\sigma_n^{-1}\{\hat{Q}_{p,h_n} - \bar{Q}(p)\} \times \frac{\sigma_n}{\hat{\sigma}_n},$$

we will examine the product of the two terms on the right hand side. This means to multiply the stochastic expansions (4) and (8).

We note that $\sqrt{n}\sigma_n^{-1}\{\hat{Q}_{p,h_n} - \bar{Q}(p)\}$ is approximated by a standardized U-statistic plus a reminder of order $o_L(n^{-1/2})$. We start by examing the products of $d_{1n}A_{1n}$ with the main terms in (8):

$$\begin{aligned}
\frac{1}{2}d_{1n}n^{-1}\sigma_n^{-2}A_{1n}B_{1n} &= \frac{1}{2}d_{1n}\sigma_n^{-2}n^{-1}\sum_{i=1}^{n}\zeta_{1n}(Y_i)g_{1n}(Y_i) \\
&\quad + \frac{1}{2}d_{1n}n^{-1}\sigma_n^{-2}\sum_{C_{n,2}}\{\zeta_{1n}(Y_i)g_{1n}(Y_j) + \zeta_{1n}(Y_j)g_{1n}(Y_i)\} \\
&= \frac{1}{2}d_{1n}\sigma_n^{-2}e_{1n} + \frac{1}{2}d_{1n}n^{-1}\sigma_n^{-2}\sum_{C_{n,2}}\{\zeta_{1n}(Y_i)g_{1n}(Y_j) \\
&\quad + \zeta_{1n}(Y_j)g_{1n}(Y_i)\} + o_\ell(n^{-1/2}).
\end{aligned}$$

In a similar way

$$d_{1n}n^{-1}h_n^{-1}\sigma_n^{-2}A_{1n}\hat{B}_{1n} = d_{1n}h_n^{-1}\sigma_n^{-2}e_{2n} + d_{2n}\sigma_n^{-2}\sum_{C_{n,2}}\{\hat{\zeta}_{1n}(Y_i)g_{1n}(Y_j) + \hat{\zeta}_{1n}(Y_j)g_{1n}(Y_i)\} + o_\ell(n^{-1/2}).$$

From the definition of the terms A_{1n} and D_{2n} we have

$$\begin{aligned} &d_{1n}n^{-2}h_n^{-2}\sigma_n^{-2}A_{1n}D_{2n} \\ &= d_{3n}\sigma_n^{-2}\sum_{C_{n,3}}\{g_{1n}(Y_i)\eta_{2n}(Y_j, Y_k) + g_{1n}(Y_j)\eta_{2n}(Y_i, Y_k) + g_{1n}(Y_k)\eta_{2n}(Y_i, Y_j)\} \\ &\quad + d_{3n}\sigma_n^{-2}\sum_{C_{n,2}}\eta_{2n}(Y_i, Y_j)\{g_{1n}(Y_i) + g_{1n}(Y_j)\}. \end{aligned}$$

Using the moment evaluations for the H-decomposition, we can show that

$$\begin{aligned} &d_{3n}\sigma_n^{-2}\sum_{C_{n,2}}\eta_{2n}(Y_i, Y_j)\{g_{1n}(Y_i) + g_{1n}(Y_j)\} \\ &\quad = d_{3n}n\sigma_n^{-2}\sum_{i=1}^{n}E[\eta_{2n}(Y_i, Y)g_{1n}(Y)|Y_i] + o_\ell(n^{-1/2}). \end{aligned}$$

We next examine the products of $d_{2n}A_{2n}$ with two terms in (8). It is easy to see that

$$\frac{1}{2}d_{2n}n^{-1}\sigma_n^{-2}A_{2n}\hat{B}_{1n} = o_\ell(n^{-1/2})$$

holds. Again using the moment evaluations for the H-decomposition, we have

$$\begin{aligned} &d_{2n}n^{-1}h_n^{-1}\sigma_n^{-2}A_{2n}\hat{B}_{1n} \\ &= d_{3n}\sigma_n^{-2}\sum_{C_{n,3}}\{\hat{\zeta}_{1n}(Y_i)g_{2n}(Y_j, Y_k) + \hat{\zeta}_{1n}(Y_j)g_{2n}(Y_i, Y_k) + \hat{\zeta}_{1n}(Y_k)g_{2n}(Y_i, Y_j)\} \\ &= d_{3n}\sigma_n^{-2}\sum_{C_{n,3}}\{\hat{\zeta}_{1n}(Y_i)g_{2n}(Y_j, Y_k) + \hat{\zeta}_{1n}(Y_j)g_{2n}(Y_i, Y_k) + \hat{\zeta}_{1n}(Y_k)g_{2n}(Y_i, Y_j)\} \\ &\quad + d_{3n}n\sigma_n^{-2}\sum_{i=1}^{n}E[g_{2n}(Y_i, Y)\hat{\zeta}_{1n}(Y)|Y_i] + o_\ell(n^{-1/2}). \end{aligned}$$

Finally, we use a large deviation theorem for U-statistic (cf. Malevich and Abdalimov 1979) to show that the order of its product with the remainder is $o_L(n^{-1/2})$ (for the argument we refer to Maesono 1997, p. 81). All other terms in the cross product of (4) and (8) can be shown to be of smaller order by using the same moment evaluation technique. Hence we have the desired result. □

Proof of Theorem 1 Before beginning the proof of Theorem 1, we will formulate a Lemma that establishes a link between the terms $e^*_{in}(i = 2, \ldots, 6)$ as shown in the statement of the Theorem and the terms $e_{in}(i = 2, \ldots, 6)$ in (5). The latter have already been used by us when proving the Edgeworth expansion of the standardized kernel quantile estimator (Theorem 1 in Maesono and Penev 2011).

Lemma 3 *The relationships* (11) *hold.*

The proof of this Lemma is omitted since it follows from a simple moment computation.

Having established the relationships between the terms $e^*_{in}(i = 2, \ldots, 6)$ and the terms $e_{in}(i = 2, \ldots, 6)$ we can now finish the proof of Theorem 1, following the same steps in the derivation of the Edgeworth expansion for the studentized statistic as in the proof of Theorem 1 in Maesono and Penev (2011). The detailed argumentation of all steps is presented in Maesono and Penev (2011). Therefore we get (10). □

Proof of Theorem 2 Note that

$$\sqrt{n}(\hat{Q}_{p,h_n} - Q(p)) = \sqrt{n}(\hat{Q}_{p,h_n} - \bar{Q}(p)) + d_n$$

with $d_n = \sqrt{n}(\bar{Q}(p) - Q(p)) = \sqrt{n}(\frac{1}{h_n}\int_0^1 F^{-1}(x)K(\frac{x-p}{h_n})\,\mathrm{d}x - Q(p))$. Substituting $x - p = yh_n$ and applying Taylor expansion we get

$$d_n = O(\sqrt{n}h_n^m)\int_{-1}^{1} K(y)y^m\,\mathrm{d}y.$$

Note that $\sqrt{n}h_n^m = o(n^{-1/2})$ is ensured by the condition $m = 4$ and by the choice of h_n. Noticing that

$$\frac{\sqrt{n}(\hat{Q}_{p,h_n} - Q(p))}{\hat{\sigma}_n} = \frac{\sqrt{n}(\hat{Q}_{p,h_n} - \bar{Q}(p))}{\sigma_n}\sigma_n\hat{\sigma}_n^{-1} + \frac{\sqrt{n}d_n}{\sigma_n}\sigma_n\hat{\sigma}_n^{-1}$$

and using Part (i) of Lemma 2 we see that for the chosen bandwidth we have $\frac{\sqrt{n}d_n}{\sigma_n}\sigma_n\hat{\sigma}_n^{-1} = o_\ell(n^{-1/2})$. Hence the statement of the Theorem follows directly from Theorem 1. □

Acknowledgments The authors are very grateful to the editor and to the two external referees for their careful reading of an earlier submission. Their useful comments and suggestions helped us to improve both the style and the presentation of the paper.

References

Dharmadhikari, S. W., Fabian, V., Jogdeo, K. (1968). Bounds on the moments of martingales. *Annals of Mathematical Statistics, 39*, 1719–1723.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *The Annals of Statistics, 12*(1), 261–268.

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *The Annals of Statistics, 13*(1), 428–433.

Hall, P. (1991). Edgeworth expansions for nonparametric denisty estimators with applications. *Statistics, 22*(2), 215–232.
Hall, P., Martin, M. (1988). Exact convergence rate of bootstrap quantile variance estimator. *Probability Theory and Related Fields, 80*, 261–268.
Hall, P., Sheather, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society, Ser. B, 50*(3), 381–391.
Hinkley, D., Wei, B. (1984). Improvements of jackknife confidence limit methods. *Biometrika, 71*, 331–339.
Janas, D. (1993). A smoothed bootstrap estimator for a studentized sample quantile. *Annals of the Institute of Statistical Mathematics, 45*(2), 317–329.
Lorenz, M. (1905). Methods of measuring concentration of wealth. *Journal of the American Statistical Association, 9*, 209–219.
Maesono, Y. (1997). Edgeworth expansions of a studentized U-Statistic and a jackknife estimator of variance. *Journal of Statistical Planning and and inference, 61*, 61–84.
Maesono, Y., Penev, S. (2011). Edgeworth expansion for the kernel quantile estimator. *Annals of the Institute of Statistical Mathematics, 63*, 617–644.
Malevich, T.L., Abdalimov, B. (1979). Large deviation probabilities for U-statistics. *Theory of Probability and Its Applications, 24*, 215–219.
Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics, 12*(2), 766–774.
Ogryczak, W., Ruszczyński, A. (1999). From stochastic dominance to mean-risk models: Semideviations and risk measures. *European Journal of Operational Research, 116*, 33–50.
Ogryczak, W., Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization, 13*(1), 60–78.
Reiss, R. D. (1990). *Approximate Distributions of Order Statistics*, Springer, New York.
Serfling, R. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.
Sheather, S., Marron, J. (1990). Kernel quantile estimators. *Journal of the American Statistical Association, 85*(410), 410–416.
Shorack, G., Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
Uryasev S., Rockafellar, R. T. (2001). Conditional value-at-risk: optimization approach. In: Uryasev, S., Pardalos, P. (Eds.), *Stochastic optimization: Algorithms and applications* (pp. 411–435). Dodrecht: Kluwer.
Xiang, X. (1995a). Defficiency of the sample quantile estimator with respect to kernel quantile estimators for censored data. *Annals of Statistics, 23*(3), 836–854.
Xiang, X. (1995b). A Berry–Esseen theorem for the kernel quantile estimator with application to studying the defficiency of qunatile estimators. *Annals of the Institute of Statistical Mathematics, 47*(2), 237–251
Xiang, X., Vos, P. (1997). Quantile estimators and covering probabilities. *Journal of Nonparametric Statistics, 7*, 349–363