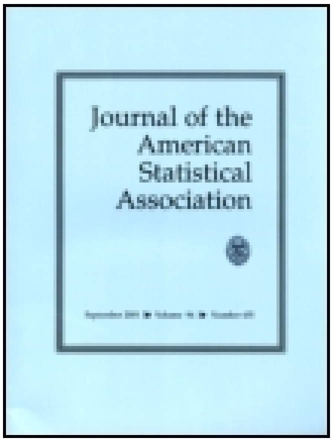


Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Kernel Quantile Estimators

Simon J. Sheather [a] & J. S. Marron [b]
[a] Australian Graduate School of Management, University of New South Wales , Kensington , New South Wales , 2033 , Australia
[b] Department of Statistics , University of North Carolina , Chapel Hill , NC , 27514
Published online: 28 Feb 2012.

Kernel Quantile Estimators

SIMON J. SHEATHER and J. S. MARRON*

For an estimator of quantiles, the efficiency of the sample quantile can be improved by considering linear combinations of order statistics, that is, L estimators. A variety of such methods have appeared in the literature; an important aspect of this article is that asymptotically several of these are shown to be kernel estimators with a Guassian kernel, and the bandwidths are identified. It is seen that some implicit choices of the smoothing parameter are asymptotically suboptimal. In addition, the theory of this article suggests a method for choosing the smoothing parameter. How much reliance should be placed on the theoretical results is investigated through a simulation study. Over a variety of distributions little consistent difference is found between various estimators. An important conclusion, made during the theoretical analysis, is that all of these estimators usually provide only modest improvement over the sample quantile. The results indicate that even if one knew the best estimator for each situation, one can expect an average improvement in efficiency of only 15%. Given the well-known distribution-free inference procedures (e.g., easily constructed confidence intervals) associated with the sample quantile, as well as the ease with which it can be calculated, it will often be a reasonable choice as a quantile estimator.

KEY WORDS: L estimators; Nonparametric; Quantiles; Smoothing parameter.

1. QUANTILE ESTIMATORS

The estimation of population quantiles is of great interest when one is not prepared to assume a parametric form for the underlying distribution. In addition, quantiles often arise as the natural thing to estimate when the underlying distribution is skewed. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with absolutely continuous distribution function F. Let $X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)}$ denote the corresponding order statistics. Define the quantile function Q to be the left-continuous inverse of F, given by $Q(p) = \inf\{x : F(x) \geq p\}$ $(0 < p < 1)$. For $0 < p < 1$, denote the pth quantile of F by ξ_p [i.e., $\xi_p = Q(p)$].

A traditional estimator of ξ_p is the pth sample quantile given by $SQ_p = X_{([np]+1)}$, where $[np]$ denotes the integral part of np. The main drawback to sample quantiles is that they experience a substantial lack of efficiency, caused by the variability of individual order statistics.

An obvious way of improving the efficiency of sample quantiles is to reduce this variability by forming a weighted average of all of the order statistics, using an appropriate weight function. These estimators are commonly called L estimators. The problem then becomes one of choosing the weight function.

A popular class of L estimators is called kernel quantile estimators. Suppose that K is a density function symmetric about 0 and that $h \to 0$ as $n \to \infty$. Let $K_h(\cdot) = h^{-1}K(\cdot/h)$. Then, one version of the kernel quantile estimator is given by $KQ_p = \sum_{i=1}^{n} [\int_{i-1/n}^{i/n} K_h(t - p)\, dt] X_{(i)}$. This form can be traced to Parzen (1979, p. 113). Clearly, KQ_p puts most weight on the order statistics $X_{(i)}$, for which i/n is close to p. Yang (1985) established the asymptotic normality and mean squared consistency of KQ_p. Falk (1984) investigated the asymptotic relative deficiency of the sample quantile with respect to KQ_p. Padgett (1986) generalized the definition of KQ_p to right-censored data. In this article, we obtain an expression for the value of the smoothing parameter h that minimizes the asymptotic mean squared error (MSE) of KQ_p and discuss the implementation of a sample-based version of it.

In practice, the following approximation to KQ_p is often used: $KQ_{p,1} = \sum_{i=1}^{n} [n^{-1}K_h(i/n - p)]X_{(i)}$. Yang (1985) showed that KQ_p and $KQ_{p,1}$ are asymptotically equivalent in mean square. If all of the observations X_i are multiplied by -1, then in general $KQ_{p,1}(-X_1, -X_2, \ldots, -X_n) \neq -KQ_{1-p,1}(X_1, X_2, \ldots, X_n)$. This is because the $X_{(n-i+1)}$ weight of $KQ_{p,1}$ differs from the $X_{(i)}$ weight of $KQ_{1-p,1}$. This problem can be overcome by replacing i/n in the definiton of $KQ_{p,1}$ by either $(i - \frac{1}{2})/n$ or $i/(n + 1)$, yielding the following estimators: $KQ_{p,2} = \sum_{i=1}^{n} [n^{-1}K_h((i - \frac{1}{2})/n - p)]X_{(i)}$ and $KQ_{p,3} = \sum_{i=1}^{n} [n^{-1}K_h(i/(n + 1) - p)]X_{(i)}$. The weights for each of these last three estimators do not generally sum to 1. Thus if a constant c is added to all of the observations X_i then in general $KQ_{p,i}(X_1 + c, X_2 + c, \ldots, X_n + c) \neq KQ_{p,i}(X_1, X_2, \ldots, X_n) + c$, for $i = 1, 2, 3$. This problem with these three estimators can be overcome by standardizing their weights by dividing them by their sum. If this is done, $KQ_{p,2}$ becomes

$$KQ_{p,4} = \sum_{i=1}^{n} K_h\left(\frac{i - \frac{1}{2}}{n} - p\right) X_{(i)} \Big/ \sum_{j=1}^{n} K_h\left(\frac{j - \frac{1}{2}}{n} - p\right).$$

In this article we establish asymptotic equivalences between KQ_p, $KQ_{p,1}$, $KQ_{p,2}$, $KQ_{p,3}$, and $KQ_{p,4}$. Each of these estimators can be motivated as adaptations of kernel regression smoothers. See Härdle (in press, chap. 3) for a discussion of these smoothers.

L estimators with much different motivation include those proposed by Harrell and Davis (1982) and Kaigh and Lachenbruch (1982). The Harrell–Davis estimator of ξ_p is given by

$$HD_p = \sum_{i=1}^{n} \left[\int_{i-1/n}^{i/n} \frac{\Gamma(n + 1)}{\Gamma((n + 1)p)\Gamma((n + 1)q)} \times t^{(n+1)p-1}(1 - t)^{(n+1)q-1}\, dt \right] X_{(i)},$$

* Simon J. Sheather is Senior Lecturer, Australian Graduate School of Management, University of New South Wales, Kensington, New South Wales 2033, Australia. J. S. Marron is Associate Professor, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514. The research of Marron was partially supported by National Science Foundation Grant DMS-8701201 and the Australian National University. The authors gratefully acknowledge some technical advice from Peter Hall.

Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Theory and Methods

where $q = 1 - p$ [see Maritz and Jarrett (1978) for related quantities]. Although Harrell and Davis did not use such terminology, this is exactly the bootstrap estimator of $E(X_{((n+1)p)})$. In this case, an exact calculation replaces the more common evaluation by simulated resampling (see Efron 1979, p. 5). In this article we demonstrate an asymptotic equivalence between HD_p and KQ_p, for a particular value of the bandwidth h. It is interesting that the bandwidth is suboptimal, yet this estimator performs surprisingly well in our simulations. See Section 4 for further analysis and discussions.

Kaigh and Lachenbruch (1982) proposed an L estimator of ξ_p as well. Their estimator is the average of pth-sample quantiles from all $\binom{n}{k}$ subsamples of size k, chosen without replacement from $X_1, X_2, \ldots, X_n$. They showed that their estimator may be written as $KL_p = \sum_{i=r}^{n+r-k} [\binom{i-1}{r-1}\binom{n-i}{k-r}/\binom{n}{k}]X_{(i)}$, where $r = [p(k+1)]$. We establish an asymptotic equivalence between KQ_p and KL_p, where the bandwidth is a function of k. This relationship (together with the optimal bandwidth theory of Sec. 2) automatically provides a theory for the choice of k that minimizes the asymptotic MSE of KL_p. See Kaigh (1988) for interesting generalizations of the ideas behind KL_p.

Kaigh (1983) pointed out that HD_p is based on ideas related to the Kaigh–Lachenbruch estimator. The latter is based on sampling without replacement, whereas the former is based on sampling with replacement in the case $k = n$. A referee has pointed out one could thus generalize HD_p to allow arbitrary k, and this estimator as well as other generalizations were proposed and studied by Kaigh and Cheng (1988). It is straightforward to use our methods to show that this is also essentially a kernel estimator and use this to give a theory for choice of k.

Brewer (1986) proposed an estimator of ξ_p, based on likelihood arguments. His estimator is given by $B_p = \sum_{i=1}^{n} [n^{-1} \cdot [\Gamma(n+1)/\Gamma(i)\Gamma(n-i+1)]p^{i-1}(1-p)^{n-i}]X_{(i)}$. We demonstrate an asymptotic equivalence between KQ_p and B_p, for a particular value of the bandwidth that (as for HD_p) is asymptotically suboptimal.

2. ASYMPTOTIC PROPERTIES OF KQ_p AND RELATED ESTIMATORS

We note that the asymptotic results given in this section concerning kernel quantile estimators only describe the situation when p is in the interior of (0, 1) in the sense that h is small enough that the support of $K_h(\cdot - p)$ is contained in [0, 1]. Theorem 1 gives an expression for the asymptotic MSE of KQ_p. This extends the asymptotic variance result of Falk (1984). The proof of this result and all other results in this section are given in the Appendix.

Theorem 1. Suppose that Q'' is continuous in a neighborhood of p and that K is a compactly supported density, symmetric about 0. Let $K^{(-1)}$ denote the antiderivative of K. Then, for all fixed $p \in (0, 1)$, apart from $p = .5$ when F is symmetric,

$$\begin{aligned}\mathrm{MSE}(KQ_p) &= n^{-1}p(1-p)[Q'(p)]^2 \\ &\quad - 2n^{-1}h[Q'(p)]^2 \int_{-\infty}^{\infty} uK(u)K^{(-1)}(u)\,du \\ &\quad + \tfrac{1}{4}h^4[Q''(p)]^2 \left[\int_{-\infty}^{\infty} u^2K(u)\,du\right]^2 \\ &\quad + o(n^{-1}h) + o(h^4).\end{aligned}$$

When F is symmetric,

$$\begin{aligned}\mathrm{MSE}(KQ_{.5}) &= n^{-1}[Q'\tfrac{1}{2}]^2 \left\{.25 - h\int_{-\infty}^{\infty} uK(u)K^{(-1)}(u)\,du \right.\\ &\quad \left. + n^{-1}h^{-1}\int_{-\infty}^{\infty} K^2(u)\,du\right\} + o(n^{-1}h) + o(n^{-2}h^{-2}).\end{aligned}$$

Note that for reasonable choice of h (i.e., tending to 0 faster than $n^{-1/4}$) the dominant term of the MSE is the asymptotic variance of the sample quantile. The improvement [note $\int uK(u)K^{(-1)}(u)\,du > 0$] over the sample quantile of local averaging shows up only in lower-order terms (this phenomenon has been called *deficiency*), so it is relatively small for large samples. See Pfanzagl (1976) for deeper theoretical understanding and discussion of this phenomenon. The fact that there is a limit to the gains in efficiency that one can expect is verified in the simulation study in Section 4.

The previous theorem can be shown to hold for the normal and other reasonable infinite-support positive kernels, using a straightforward but tedious truncation argument. The results of Theorem 1 can easily be extended to higher-order kernels (i.e., those giving faster rates of convergence at the price of taking on negative values). Nevertheless, we do not state our results for higher-order kernels, since this would tend to obscure the important points concerning the asymptotic equivalences between estimators. Azzalini (1981) considered estimators of quantiles obtained by inverting kernel estimators of the distribution function and obtained a result related to our Theorem 1. Theorem 1 produces the following corollary.

Corollary 1. Suppose that the conditions given in Theorem 1 hold. Then, for all p, apart from $p = .5$ when F is symmetric, the asymptotically optimal bandwidth is given by $h_{\text{opt}} = \alpha(K) \cdot \beta(Q) \cdot n^{-1/3}$, where

$$\alpha(K) = \left[2\int_{-\infty}^{\infty} uK(u)K^{(-1)}(u)\,du \Big/ \left\{\int_{-\infty}^{\infty} u^2K(u)\,du\right\}^2\right]^{1/3} \tag{2.1}$$

and $\beta(Q) = [Q'(p)/Q''(p)]^{2/3}$. With $h = h_{\text{opt}}$,

$$\mathrm{MSE}(KQ_p) = n^{-1}p(1-p)[Q'(p)]^2 + O(n^{-4/3}). \tag{2.2}$$

When F is symmetric and $p = .5$, taking $h = O(n^{-1/2})$ makes the first two terms in h of the MSE of $KQ_{.5}$ the same order, and $\mathrm{MSE}(KQ_{.5}) = .25n^{-1}[Q'(\tfrac{1}{2})]^2 + O(n^{-3/2})$. Nevertheless, because the term in hn^{-1} is negative and the term in $n^{-2}h^{-1}$ is positive, there is no single bandwidth that minimizes the asymptotic MSE of $KQ_{.5}$ when F is symmetric. Instead, any h satisfying h = constant $\cdot\, n^{-m}$ $(0 < m \le 1/2)$ will, for large values of the constant, produce an estimator with smaller asymptotic MSE than $SQ_{.5}$.

We next present a theorem that establishes some asymptotic equivalences between the different forms of the kernel quantile estimator. In veiw of (2.2), we deem the two kernel quantile estimators $KQ_{p,i}$ and $KQ_{p,j}$ asymptotically equivalent when, for reasonable values of h, $E[(KQ_{p,i} - KQ_{p,j})^2] = o(n^{-4/3})$.

Theorem 2. Suppose that K is compactly supported and has a bounded second derivative. Then, (a) for $hn^{2/3} \to \infty$, KQ_p and $KQ_{p,2}$ are asymptotically equivalent; (b) for $hn^{2/3} \to \infty$, $KQ_{p,2}$ and $KQ_{p,1}$ are asymptotically equivalent; (c) for $hn^{5/6} \to \infty$, $KQ_{p,1}$ and $KQ_{p,3}$ are asymptotically equivalent; and (d) for $hn^{5/6} \to \infty$, KQ_p and $KQ_{p,4}$ are asymptotically equivalent.

The first assumption of this theorem rules out the normal kernel, but this and other reasonable infinite-support kernels can be handled by a straightforward but tedious truncation argument. The second assumption does not include the rectangular of Epanechnikov kernels. For a discussion of these and other kernels, see Härdle (in press). Similar results can be obtained for these, but slightly different methods of proof are required. These extensions of Theorem 2 are omitted because the space required for their proof does not seem to justify the small amount of added generality.

Finally, in this section we present a series of lemmas showing that in large samples HD_p, KL_p, and B_p are essentially the same as KQ_p for specific choices of K and h.

Lemma 1. Let $q = 1 - p$ (where $0 < p < 1$) and $\beta = \alpha + O(1)$. Then, as $\alpha \to \infty$,

$$\frac{\Gamma(p\alpha + q\beta)}{\Gamma(p\alpha)\Gamma(q\beta)}\, x^{p\alpha-1}(1 - x)^{q\beta-1} \to [2\pi pq/\alpha]^{-1/2}\exp(-\alpha(x - p)^2/2pq),$$

in the sense that

$$\frac{\Gamma(p\alpha + q\beta)}{\Gamma(p\alpha)\Gamma(q\beta)}\,[p + (pq/\alpha)^{1/2}y]^{p\alpha-1} \times [q - (pq/\alpha)^{1/2}y]^{q\alpha-1}(pq/\alpha)^{1/2} = [2\pi]^{-1/2}\exp(-\tfrac{1}{2}y^2) + O(\alpha^{-1/2}).$$

It follows from Lemma 1, with $\alpha = \beta = n + 1$, that in large samples HD_p is essentially the same as KQ_p, with K the standard normal density and

$$h = [pq/(n + 1)]^{1/2}. \tag{2.3}$$

We see from Theorem 1 that HD_p is asymptotically suboptimal, being based on $h = O(n^{-1/2})$ rather than $h = O(n^{-1/3})$, resulting in weights that are too concentrated in a neighborhood of p. See Yashizawa, Sen, and Davis (1985) for an interesting and closely related result in the case $p = \frac{1}{2}$. Understanding KL_p in large samples requires a further lemma.

Lemma 2. Let $q = 1 - p$ (where $0 < p < 1$), $i/n = p + O(k^{-1/2})$, and $r = pk + O(1)$, with $k = o(n)$. Then, as $n \to \infty$ and $k \to \infty$,

$$\frac{\binom{i-1}{r-1}\binom{n-i}{k-r}}{\binom{n}{k}} = n^{-1}\frac{\Gamma(k + 1)}{\Gamma(r)\Gamma(k - r + 1)} \times \left(\frac{i}{n}\right)^{r-1}\left(1 - \frac{i}{n}\right)^{(k-r+1)-1}\left(1 + O\left(\frac{k}{n}\right)\right).$$

Putting Lemmas 1 and 2 together, we find that in large samples KL_p is essentially the same as $KQ_{p,1}$, with K the standard normal density and

$$h = [pq/k]^{1/2}. \tag{2.4}$$

Corollary 1 can therefore be used to find an expression for the asymptotically optimal value of k. Finally, Brewer's estimator, B_p, requires a slightly different lemma.

Lemma 3. Let $q = 1 - p$ (where $0 < p < 1$) and $i/(n + 1) = p + O(n^{-1/2})$. Then, as $n \to \infty$,

$$\frac{\Gamma(n + 1)}{\Gamma(i)\Gamma(n - i + 1)}\, p^i q^{n-i} = [2\pi pq/(n + 1)]^{-1/2} \times \exp\left\{-\left(\frac{i}{n + 1} - p\right)^2 \Big/ \frac{2pq}{n + 1}\right\}[1 + O(n^{-1/2})].$$

It follows from Lemma 3 that in large samples B_p is essentially the same as $KQ_{p,3}$ with K the standard normal density and $h = [pq/n]^{1/2}$. We see from Theorem 1 that like HD_p, B_p is asymptotically suboptimal, since it is based on $h = O(n^{1/2})$ rather than $h = O(n^{-1/3})$.

Similar but slightly weaker equivalences were obtained by Yang (1985, theorem 3) between KQ_p and $KQ_{p,1}$ and by Zelterman (1988) among $KQ_{p,1}$, HD_p, and KL_p. P. K. Sen (private communication, 1988) pointed out that another way of deriving our results would be through standard U-statistic theory.

3. DATA-BASED CHOICE OF THE BANDWIDTH

In this section we propose a data-based choice of h, the smoothing parameter of KQ_p, for all p, apart from $p = .5$ when F is symmetric.

We see from Corollary 1 that for a given choice of K, the asymptotically optimal value of h depends on the first and second derivatives of the quantile function. Thus estimates of $Q'(p)$ and $Q''(p)$ are necessary for a data-based choice of h. If the first and second derivatives of K exist, then we can estimate these quantities by the first and second derivatives of KQ_p. Since interest is in the ratio $[Q'(p)/Q''(p)]^{2/3}$, it seems natural to consider higher-order kernels in an attempt to keep the problems associated with ratio estimation at bay. This results in the estimators

$$\hat{Q}'_m(p) = \sum_{i=1}^{n}\left[\int_{i-1/n}^{i/n} a^{-2}K'_*(a^{-1}(t - p))\,dt\right]X_{(i)}$$

and

$$\hat{Q}''_m(p) = \sum_{i=1}^{n} \left[\int_{i-1/n}^{i/n} b^{-3} K''_*(b^{-1}(t-p))\, dt \right] X_{(i)},$$

where K_* is a kernel of order m, symmetric about 0 [i.e., $\int_{-\infty}^{\infty} K_*(u)\, du = 1$, $\int_{-\infty}^{\infty} u^i K_*(u)\, du = 0$ $(i = 1, 2, \ldots, m-1)$, and $\int_{-\infty}^{\infty} u^m K_*(u)\, du < \infty$].

The resulting estimate of the asymptotically optimal bandwidth is given by

$$\hat{h}_{\text{opt}} = \alpha(K) \times \hat{\beta} \times n^{-1/3}, \tag{3.1}$$

where $\hat{\beta} = [\hat{Q}'_m(p)/\hat{Q}''_m(p)]^{2/3}$ and $\alpha(K)$ is given by (2.1). The problem is then to choose values for the bandwidths a and b that result in an asymptotically efficient $\hat{\beta}$.

Theorem 3. Suppose that $Q^{(m+2)}$ is continuous in a neighborhood of p and that K_* is a compactly supported kernel of order m, symmetric about 0. The asymptotically optimal bandwidth for $\hat{Q}'_m(p)$ is given by $\alpha_{\text{opt}} = \mu_m(K_*) \times \gamma_m(Q) \times n^{-1/(2m+1)}$, where

$$\mu_m(K_*) = \left[\frac{(m!)^2 \int_{-\infty}^{\infty} K_*^2(u)\, du}{2m \left\{ \int_{-\infty}^{\infty} u^m K_*(u)\, du \right\}^2} \right]^{1/(2m+1)}$$

and $\gamma_m(Q) = [Q'(p)/Q^{(m+1)}(p)]^{2/(2m+1)}$. The asymptotically optimal bandwidth for $\hat{Q}''_m(p)$ is given by $b_{\text{opt}} = \tau_m(K_*) \times \delta_m(Q) \times n^{-1/(2m+3)}$, where

$$\tau_m(K_*) = \left[\frac{3(m!)^2 \int_{-\infty}^{\infty} \{K'_*(u)\}^2\, du}{2m \left\{ \int_{-\infty}^{\infty} u^m K_*(u)\, du \right\}^2} \right]^{1/(2m+3)}$$

and $\delta_m(Q) = [Q'(p)/Q^{(m+2)}(p)]^{2/(2m+3)}$.

In view of this theorem, we can choose the bandwidths for $\hat{Q}'_m(p)$ and $\hat{Q}''_m(p)$ to be $a = c'_m \times \mu_m(K_*) \times n^{-1/(2m+1)}$ and $b = c''_m \times \tau_m(K_*) \times n^{-1/(2m+3)}$, where c'_m and c''_m are constants calculated from $\gamma_m(Q)$ and $\delta_m(Q)$, respectively, assuming a distribution such as the normal. This approach has been used successfully by Hall and Sheather (1988) to choose the bandwidth of an estimator of $Q'(.5)$.

Alternative methods of obtaining a data-based choice of h were given by Yang (1985), Padgett and Thombs (1986), and Zelterman (1988).

4. MONTE CARLO STUDY

A Monte Carlo study was carried out to evaluate the performance of the data-based bandwidths for the kernel quantile estimator and to compare the performance of the kernel quantile estimator with the estimators of Harrell and Davis (1982) and Kaigh and Lachenbruch (1982).

Using subroutines from IMSL, 1,000 pseudorandom samples of size 50 and 100 were generated from the double-exponential, exponential, lognormal, and normal distributions. We calculated the MSE for the estimators given in the following at the .05, .1, .25, .5, .75, .9, and .95 quantiles.

To implement the data-based algorithm of Section 3, the order m of the kernel K_*, as well as the constants c'_m and c''_m, have to be chosen. A natural initial choice of K_* is a positive second-order kernel. Preliminary Monte Carlo results found that the performance of $\hat{\beta}$ based on $\hat{Q}'_2(p)$ and $\hat{Q}''_2(p)$ is dominated by the performance of $\hat{Q}''_2(p)$, whereas it is affected little by $\hat{Q}'_2(p)$. In fact, $\hat{Q}''_2(p)$ sometimes suffers from a large bias, which then translates into a large bias for $\hat{\beta}$. Thus a fourth-order kernel estimate of $Q''(p)$ was also included in the study.

Table 1 contains values of $\gamma_2(Q)$, $\delta_2(Q)$, and $\delta_4(Q)$ (i.e., the asymptotically optimal values of c'_2, c''_2, and c''_4) for the four distributions and the values of p considered in this study. These four distributions were chosen because the values of these functionals of Q include a wide cross-section of all possible values. This can be demonstrated by calculating these functionals for a family of distributions such as the generalized lambda distribution (Ramberg, Dudewicz, Tadikamalla, and Mykytka 1979). Also included in Table 1 are values of $\beta(Q)$. We can see from these values that there is a wide disparity between the optimal bandwidths of KQ_p for the four distributions. For example, $\beta(Q)$ for the exponential distribution is up to six times larger than that for the normal, lognormal, and double-exponential distributions. This seems to indicate that one should estimate $\beta(Q)$ rather than use the same $\beta(Q)$ and hence the same bandwidth for all underlying distributions, as is essentially done by HD_p and B_p.

In view of Lemmas 1, 2, and 3, we chose the Gaussian kernel $K(u) = [2\pi]^{-1/2} \exp(-\frac{1}{2}u^2)$ for this Monte Carlo study and used the form $KQ_{p,4}$ of the kernel quantile estimator. For the Gaussian kernel, $\int_{-\infty}^{\infty} uK(u)K^{(-1)}(u)\, du = 1/(2\sqrt{\pi})$. The Gaussian kernel was also used as K_* to estimate $Q'(p)$ and $Q''(p)$. The following fourth-order

Table 1. Values of Functionals of Q in Asymptotically Optimal Bandwidths for Kernel Estimates of Q(p), Q'(p), and Q''(p)

p	Double-exponential	Normal	Lognormal	Exponential
.05, .95				
$\beta(Q)$	.14	.16	.12, .29	.97, .14
$\gamma_2(Q)$	.07	.08	.05, .40	.73, .07
$\delta_2(Q)$	.05	.05	.03, .11	.57, .05
$\delta_4(Q)$	.03	.03	.01, .07	.40, .03
.1, .9				
$\beta(Q)$	.22	.27	.18, .73	.93, .22
$\gamma_2(Q)$	.12	.14	.08, .40	.70, .12
$\delta_2(Q)$	.08	.09	.05, .16	.55, .08
$\delta_4(Q)$	.05	.06	.03, .08	.38, .05
.25, .75				
$\beta(Q)$	.40	.61	.33, .98	.83, .40
$\gamma_2(Q)$	.25	.31	.15, .29	.60, .25
$\delta_2(Q)$	.18	.22	.09, .17	.47, .18
$\delta_4(Q)$	.12	.13	.05, .09	.32, .10
.5				
$\beta(Q)$	—	—	.54	.63
$\gamma_2(Q)$	—	—	.23	.44
$\delta_2(Q)$	—	—	.14	.33
$\delta_4(Q)$	—	—	.08	.22

kernel, given by Müller (1984), was used to estimate $Q''(p)$:

$$K_*(u) = (315/512)(3 - 20u^2 + 42u^4 - 36u^6 + 11u^8) \times I(-1 \le u \le 1).$$

To avoid integration, the following approximations to $\hat{Q}'_m(p)$ and $\hat{Q}''_m(p)$ were used:

$$\hat{Q}'_m(p) \simeq \sum_{i=1}^{n} \left[n^{-1}a^{-2}K'_* \left(a^{-1} \left(\frac{i-\frac{1}{2}}{n} - p \right) \right) \right] X_{(i)}$$

and

$$\hat{Q}''_m(p) \simeq \sum_{i=1}^{n} \left[n^{-1}b^{-3}K''_* \left(b^{-1} \left(\frac{i-\frac{1}{2}}{n} - p \right) \right) \right] X_{(i)}.$$

Three different values of each of the constants c'_2, c''_2, and c''_4 were used for each value of p. Experience with $\hat{h}_{\text{opt}}$, as given by (3.1), reveals that it can produce both small and large values when compared with h_{opt}. This is not surprising, since $\hat{\beta}$ is made up of a ratio of two estimates. To overcome this problem, any estimate $\hat{\beta}$ outside the interval [.05, 1.5] was set equal to the closest endpoint of this interval.

The values of the constants c'_2, c''_2 and c'_2, c''_4 that consistently produced the smallest MSE for $KQ_{p,4}$ over the four distributions considered in this study are given in Table 2. We denote by $KQ^{(1)}_{p,4}$ the kernel quantile estimator $KQ_{p,4}$ based on h obtained from $\hat{Q}'_2(p)$ and $\hat{Q}''_2(p)$, using the values of c'_2 and c''_2 given in Table 2. Similarly, we let $KQ^{(2)}_{p,4}$ denote $KQ_{p,4}$ based on h obtained from $Q'_2(p)$ and $Q''_4(p)$, using the values of c'_2 and c''_4 given in Table 2.

To implement the Kaigh and Lachenbruch estimate KL_p, one is faced with the problem of choosing its smoothing parameter k. Following Kaigh (1983), we chose k = 19, 39 when n = 50 and k = 39, 79 when n = 100 for this Monte Carlo study. In view of (2.4), the asymptotically optimal value of k can be found via the formula $h_{\text{opt}} = [pq/k_{\text{opt}}]^{1/2}$. Using this formula, the data-based choices of h were used to produce data-based choices of k.

The table of Monte Carlo results is too large to report here, so we simply give some highlights. As expected from the theory in Section 2, no quantile estimator dominated over the others, nor was any better than the sample quantile in every case. To get a feeling for how much improvement over the sample quantile was possible, we considered the increase in efficiency (i.e., ratio of MSE's) of the best of all estimators (for each of the 44 combinations of distribution, sample-size and quantile). This estimator, which is clearly unavailable in practice, was not much better than the sample quantile, with increases in efficiency ranging from 3% to 42% with an average of 15%. The kernel estimator $KQ^{(2)}_{p,4}$ gave moderately superior performance to $KQ^{(1)}_{p,4}$, and HD_p produced smaller MSE's in 26 and 28 of the 44 combinations, respectively. $KQ^{(2)}_{p,4}$ had even better performance when compared with the other estimators (although never dominating any of them). The two data-based choices of k for KL_p generally gave inferior performance to the Kaigh and Lachenbruch estimator based on the fixed but arbitrary choices of k. Nevertheless, KL_p based on the fixed choices of k generally performed worse than both $KQ^{(2)}_{p,4}$ and HD_p.

The reason for the somewhat surprisingly similar performance of the Harrell–Davis estimator and the kernel estimators can be explained as follows. There is quite a lot of variability in the data-based bandwidths for the kernel estimators, whereas the bandwidth inherent in the Harrell–Davis estimate, given by (2.3), is fixed at a point that is often not too far from the optimum bandwidth in samples of size 50 and 100. Figure 1 contains plots of the asymptotic MSE of KQ_p, obtained from the expression given in Theorem 1, for the .1 and .9 quantiles of the lognormal distribution when n = 50. The asymptotically

Table 2. Values of the Constants c'_2, c''_2 and c'_2, c''_4 That Consistently Produce the Smallest MSE for $KQ_{p,4}$

p	c'_2, c''_2	c'_2, c''_4
.05, .95	.75, .6	.75, .4
.1, .9	.2, .6	.2, .4
.25, .75	.6, .5	.6, .3
.5	.8, .3	.4, .2

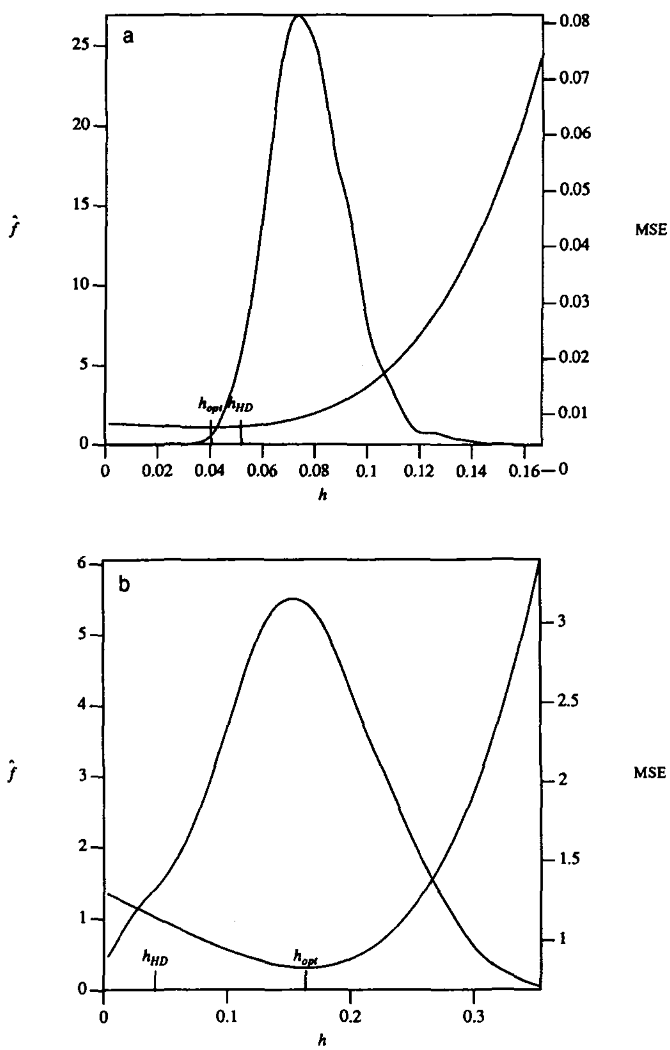
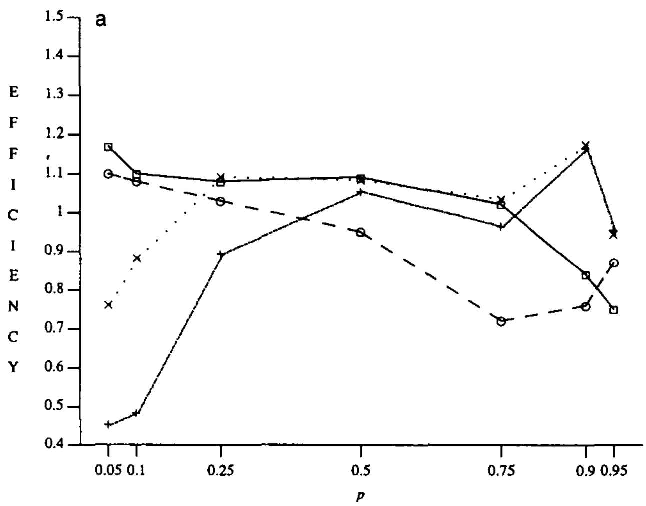


Figure 1. Plots of the Asymptotic MSE of KQ_p Versus h for the (a) .1 and (b) .9 Quantiles of the Lognormal Distribution When n = 50. Estimates of the density of the data-based bandwidths are also included.

optimum bandwidth (h_{opt}) and the bandwidth inherent in the Harrell–Davis estimator (i.e., $h_{HD} = [pq/(n + 1)]^{1/2}$) are marked on the plots. In the case of the .1 quantile these two bandwidths are close together, whereas for the .9 quantile they are well separated. This explains why the Harrell–Davis estimator performs better for the .1 quantile. Also included in the plots are Gaussian kernel estimates of the density of the data-based bandwidths for $KQ_{p,4}^{(2)}$. Each density estimate is based on the 1,000 bandwidths obtained in the Monte Carlo study. The bandwidth for each density estimate was found using the plug-in method of Hall, Sheather, Jones, and Marron (1989). In the case of the .9 quantile the center of the distribution of the data-based bandwidths is close to the optimum bandwidth, whereas for the .1 quantile it is not. This explains the better performance of $KQ_{p,4}^{(2)}$ for the .9 quantile.

Because of the noise inherent in our data-based bandwidths, we considered using a fixed bandwidth for KQ_p that was less arbitrary than the bandwidth for HD_p. The bandwidth we chose corresponds to the one that is asymptotically optimal when the underlying distribution is normal. (This is undefined at $p = .5$, for which we set h equal to the bandwidth corresponding to an exponential distribution.) We denote this estimator by KQN_p. KQN_p had a larger MSE than HD_p and $KQ_{p,4}^{(2)}$ in 23 and 27 of the 44 combinations, respectively.



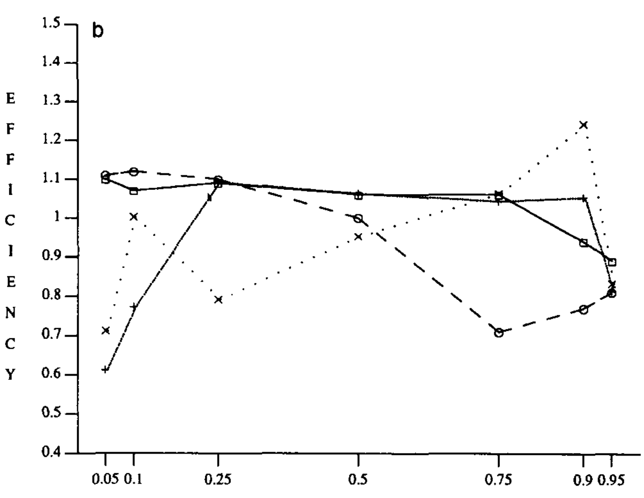


Figure 2. Plots of the Ratio of the MSE of Each of the Estimators HD_p (—□—), KQN (-○-), $KQ_{p,4}^{(1)}$ (⋯+⋯), and $KQ_{p,4}^{(2)}$ (⋯×⋯) to the Sample Quantile SQ_p: (a) Lognormal n = 50; (b) Lognormal n = 100.

Figure 2 is a plot of the efficiency of each of the estimators HD_p, $KQ_{p,4}^{(1)}$, $KQ_{p,4}^{(2)}$, and KQN with respect to the sample quantile $SQ_{(p)}$ for samples of size 50 and 100 from the lognormal distribution. A full set of such plots for each distribution and sample size can be found in Sheather and Marron (1989).

Figure 2 shows once again that apart from the extreme quantiles, there is little difference between various quantile estimators (including the sample quantile). Given the well-known distribution-free inference procedures (e.g., easily constructed confidence intervals) associated with the sample quantile, as well as the ease with which it can be calculated, it will often be a reasonable choice as a quantile estimator.

APPENDIX: PROOFS

Theorem 1. We first consider all p, apart from $p = .5$ when F is symmetric. Since K is compactly supported and Q'' is continuous in a neighborhood of p, we find [using eq. (4.6.3) of David (1981)] that

$$\begin{aligned}\text{bias}(KQ_p) &= \sum_{i=1}^{n}\left[\int_{i-1/n}^{i/n} K_h(t-p)\,dt\right] \times \left\{Q\left(\frac{i}{n+1}\right) - Q(p)\right\} + O(n^{-1})\\ &= \int_0^1 K_h(t-p)\{Q(t) - Q(p)\}\,dt + O(n^{-1})\\ &= \frac{1}{2}h^2\left[\int_{-\infty}^{\infty} u^2K(u)\,du\right]Q''(p) + o(h^2) + O(n^{-1}).\end{aligned}$$

Falk (1984, p. 263) proved that

$$\begin{aligned}\text{var}(KQ_p) &= n^{-1}p(1-p)[Q'(p)]^2\\ &\quad - n^{-1}h[Q'(p)]^2\int_{-\infty}^{\infty} uK(u)K^{(-1)}(u)\,du + o(n^{-1}h).\end{aligned}$$

Squaring the expression for the bias and combining it with the variance gives the result.

Next, suppose that F is symmetric. Since KQ_p is both location- and scale-equivariant, $KQ_{.5}$ is symmetrically distributed about its mean, $\xi_{.5}$. The expression for MSE($KQ_{.5}$) is found by extending Falk's expansion for var(KQ_p) to include the next term.

Theorem 2. We only give the details for (a). The proofs of (b)–(d) follow in a similar manner. Let $W_{n,h}(i) = \int_{i-1/n}^{i/n} K_h(t - p)\,dt - n^{-1}K_h[(i - \frac{1}{2})/n - p]$. Since $|W_{n,h}(i)| = O(n^{-3}h^{-3})$ and $W_{n,h}(i) = 0$ except for i in a set S of cardinality $O(nh)$, we find [using eqs. (4.6.1) and (4.6.3) of David (1981)] that

$$\begin{aligned}&E[KQ_p - KQ_{p,2}]^2\\ &= E\left[\sum_{i=1}^{n} W_{n,h}(i)X_{(i)}\right]^2\\ &= E\left[\sum_{i\in S} W_{n,h}(i)\{X_{(i)} - E(X_{(i)})\}\right]^2 + \left[\sum_{i\in S} W_{n,h}(i)E(X_{(i)})\right]^2\\ &= O(n^{-4}h^{-4})\\ &= o(n^{-4/3}),\end{aligned}$$

if $hn^{2/3} \to \infty$ as $n \to \infty$.

The proofs of Lemmas, 1, 2, and 3 follow from an application of Sterling's formula. The proof of Theorem 3 follows in the same manner as that of Theorem 1.

[*Received June 1988. Revised August 1989.*]

REFERENCES

Azzalini, A. (1981), "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68, 326–328.

Brewer, K. R. W. (1986), "Likelihood Based Estimation of Quantiles and Density Estimation," unpublished manuscript.

David, H. A. (1981), *Order Statistics* (2nd ed.), New York: John Wiley.

Efron, B. (1979), "Boostrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Falk, M. (1984), "Relative Deficiency of Kernel Type Estimators of Quantiles," *The Annals of Statistics*, 12, 261–268.

Hall, P., and Sheather, S. J. (1988), "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society*, Ser. B, 50, 381–391.

Hall, P., Sheather, S. J., Jones, M. C., and Marron, J. S. (1989), "An Optimal Data-Based Bandwidth Selection in Kernal Density Estimation," unpublished manuscript.

Härdle, W. (in press), *Applied Nonparametric Regression* (Econometric Society Monograph Series), Cambridge, U.K.: Cambridge University Press.

Harrell, F. E., and Davis, C. E. (1982), "A New Distribution-Free Quantile Estimator," *Biometrika*, 69, 635–640.

Kaigh, W. D. (1983), "Quantile Interval Estimation," *Communications in Statistics, Part A—Theory and Methods*, 12, 2427–2443.

——— (1988), "*O*-Statistics and Their Applications," *Communications in Statistics, Part A—Theory and Methods*, 17, 2191–2210.

Kaigh, W. D., and Cheng, C. (1988), "Subsampling Quantile Estimators and Uniformity Criteria," unpublished manuscript.

Kaigh, W. D., and Lachenbruch, P. A. (1982), "A Generalized Quantile Estimator," *Communications in Statistics, Part A—Theory and Methods*, 11, 2217–2238.

Maritz, J. S., and Jarrett, R. G. (1978), "A Note on Estimating the Variance of the Sample Median," *Journal of the American Statistical Association*, 73, 194–196.

Müller, H. (1984), "Smooth Optimum Kernel Estimators of Densities, Regression Curves and Modes," *The Annals of Statistics*, 12, 766–774.

Padgett, W. J. (1986), "A Kernel-Type Estimator of a Quantile Function From Right-Censored Data," *Journal of the American Statistical Association*, 81, 215–222.

Padgett, W. J., and Thombs, L. A. (1986), "Smooth Nonparametric Quantile Estimation Under Censoring: Simulations and Bootstrap Methods," *Communications in Statistics, Part B—Simulation and Computation*, 15, 1003–1025.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–131.

Pfanzagl, J. (1976), "Investigating the Quantile of an Unknown Distribution," *Contributions to Applied Statistics, Experentia Supplementum*, 22, 111–126.

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., and Mykytka, E. F. (1979), "A Probability Distribution and Its Uses in Fitting Data," *Technometrics*, 21, 201–214.

Sheather, S. J., and Marron, J. S. (1989), "Kernel Quantile Estimation," technical report, University of North Carolina, Dept. of Statistics.

Yang, S.-S. (1985), "A Smooth Nonparametric Estimator of a Quantile Function," *Journal of the American Statistical Association*, 80, 1004–1011.

Yashizawa, C. N., Sen, P. K., and Davis, C. E. (1985), "Asymptotic Equivalence of the Harrell-Davis Estimator and the Sample Median," *Communications in Statistics, Part A—Theory and Methods*, 14, 2129–2136.

Zelterman, D. (1988), "Smooth Nonparametric Estimation of the Quantile Function," unpublished manuscript.