#1783

# KERNEL QUANTILE ESTIMATORS

SIMON J. SHEATHER

and

J.S. MARRON

Simon J. Sheather is Lecturer, Australian Graduate School of Management, University of New South Wales, Kensington, NSW, 2033, Australia. J.S. Marron is Associate Professor, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, U.S.A. The research of the second author was partially supported by NSF Grant DMS–8701201 and by the Australian National University. The authors with to gratefully acknowledge some technical advice provided by Peter Hall.

# SUMMARY

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. The sample quantile is a popular nonparametric estimator of the corresponding population quantile. Being a function of at most two order statistics, sample quantiles experience a substantial loss of efficiency for distributions such as the normal. An obvious way to improve efficiency is to form a weighted average of several order statistics, using an appropriate weight function. Such estimators are called $L$-estimators. The problem then becomes one of choosing the weight function. One class of $L$-estimators, which uses a density function (called a kernel) as its weight function, are called kernel quantile estimators. The effective performance of such estimators depends critically on the selection of a smoothing parameter. An important part of this paper is a theoretical analysis of this selection. In particular, we obtain an expression for the value of the smoothing parameter which minimizes asymptotic mean square error. Another key feature of this paper is that this expression is then used to develop a practical data-based method for smoothing parameter selection.

Other $L$-estimators of quantiles have been proposed by Harrell and Davis (1982), Kaigh and Lachenbruch (1982) and Brewer (1986). The Harrell-Davis estimator is just a bootstrap estimator (Section 1). An important aspect of this paper is that we show that asymptotically all of these are kernel estimators with a Gaussian kernel and we identify the bandwidths. It is seen that the choices of smoothing parameter inherent in both the Harrell and Davis estimator and the Brewer estimator are asymptotically suboptimal. Our theory also suggests a method for choosing a previously not understood tuning parameter in the Kaigh-

Lachenbruch estimator.

The final point is an investigation of how much reliance should be placed on the theoretical results, through a simulation study. We compare one of the kernel estimators, using data-based bandwidths, with the Harrell-Davis and Kaigh-Lachenbruch estimators. Over a variety of distributions little consistent difference is found between these estimators. An important conclusion, also made during the theoretical analysis, is that all of these estimators usually provide only modest improvement over the sample quantile. Our 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; Smoothing parameter; Nonparametric; Quantiles.

## 1. QUANTILE ESTIMATORS

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with absolutely continuous distribution function $F$. Let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ denote the corresponding order statistics. Define the quantile function $Q$ to be the left continuous inverse of $F$ given by $Q(p) = \inf\{x : F(x) \ge p\}$, $0 < p < 1$. For $0 < p < 1$, denote the $p$th quantile of $F$ by $\xi_p$ [that is, $\xi_p = Q(p)$].

A traditional estimator of $\xi_p$ is the $p$th sample quantile which is 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 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 are called kernel quantile estimators. Suppose that $K$ is a density function symmetric about zero and that $h \to 0$ as $n \to \infty$. Let $K_h(.) = h^{-1} K(./h)$ then one version of the kernel quantile estimator is given by

$$KQ_p = \sum_{i=1}^{n} \left[ \int_{i-1/n}^{i/n} K_h(t-p)\,\mathrm{dt} \right] X_{(i)}.$$

This form can be traced back to Parzen (1979, p.113). Clearly, $KQ_p$ puts most weight on the order statistics $X_{(i)}$ for which $i/n$ is close to $p$. $KQ_p$ can also be motivated as an adaptation of the regression smoother of Gasser and Müller (1979). Yang (1985) established the asymptotic normality and mean square 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 paper, we obtain an expression for the value of the smoothing parameter $h$ which minimizes the asymptotic mean square error 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

$$KQ_{p,1} = \sum_{i=1}^{n} [n^{-1} K_h(i/n - p)] X_{(i)}.$$

This estimator is an adaptation of the regression smoother studied by Priestley and Chao (1972). Yang (1985) showed that $KQ_p$ and $KQ_{p,1}$ are asymptotically equivalent in mean square. If all 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 due to the fact that 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 definition of $KQ_{p,1}$ by either $(i - ½)/n$ or $i/(n+1)$, yielding the following estimators

$$KQ_{p,2} = \sum_{i=1}^{n} \left[ n^{-1} K_h \left[ \frac{i-½}{n} - p \right] \right] X_{(i)}$$

and

$$KQ_{p,3} = \sum_{i=1}^{n} \left[ n^{-1} K_h \left[ \frac{i}{n+1} - p \right] \right] X_{(i)}$$

The weights for each of these last three estimators do not in general sum to one. Thus if a constant $c$ is added to all 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

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

This estimator is an adaptation of the regression smoother proposed by Nadaraya (1964) and Watson (1964). In this paper we establish asymptotic equivalences between $KQ_p$, $KQ_{p,1}$, $KQ_{p,2}$, $KQ_{p,3}$ and $KQ_{p,4}$. See Härdle (1988) for further discussion and comparison of regression estimators.

Harrell and Davis (1982) proposed the following estimator of $\xi_p$

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

where $q = 1-p$ [see Maritz and Jarrett (1978) for related quantities]. While 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 paper, we also 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) also proposed an $L$-estimator of $\xi_p$. Their estimator is the average of $p$th sample quantiles from all $\binom{n}{k}$ subsamples of size $k$, chosen without replacement from $X_1, X_2, \ldots, X_n$. They show that their estimator may be written as

$$KL_p = \sum_{i=r}^{n+r-k} \frac{\binom{i-1}{r-1}\binom{n-i}{k-r}}{\binom{n}{k}} X_{(i)}$$

where $r = \lfloor p(k+1) \rfloor$. 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 Section 2 automatically provides a theory for choice of $k$ which minimizes the asymptotic mean square error 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 and Lachenbruch estimator. The latter is based on sampling without replacement while 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 have been in fact proposed and studied in a very recent paper by Kaigh and Cheng (1988). It is straightforward to use our methods to show this is also essentially a kernel estimator and use this to give a theory for choice of $k$.

Brewer (1986) proposed an estimator of $\xi_p$, based on likelihood arguments. His estimator is given by

$$B_p = \sum_{i=1}^{n} [n^{-1} \cdot \frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)} p^{i-1}(1-p)^{n-i}] X_{(i)}.$$

We also demonstrate an asymptotic equivalence between $KQ_p$ and $B_p$, for a particular value of the bandwidth which, as for $HD_p$, is asymptotically suboptimal.

## 2. ASYMPTOTIC PROPERTIES OF $KQ_p$ AND RELATED ESTIMATORS

We begin this section by noting 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 $K_h(. - p)$ is contained in [0, 1]. Theorem 1 gives an expression for the asymptotic mean square error of $KQ_p$. This extends the asymptotic variance result of Falk (1984). The proof of this result and all other results in the section are given in the Appendix.

*Theorem 1.* Suppose that $Q''$ is continuous in a neighbourhood of $p$ and that $K$ is a compactly supported density, symmetric about zero. Let $K^{(-1)}$ denote the antiderivative of $K$. Then for all fixed $p \in (0, 1)$, apart from $p = 0.5$ when $F$ is symmetric

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

When $F$ is symmetric

$$MSE\,(KQ_{0.5}) = n^{-1}\,[Q'(\tfrac{1}{2})]^2 \left\{0.25 - h \int_{-\infty}^{\infty} u\,K(u)\,K^{(-1)}(u)\,\mathrm{du} \right.$$
$$\left. + n^{-1}\,h^{-1} \int_{-\infty}^{\infty} K^2(u)\,\mathrm{du} \right\} + o(n^{-1}h) + o(n^{-2}h^{-2}).$$

Note that for reasonable choice of $h$ (i.e. tending to zero 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 will be 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 above 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 be easily extended to higher order kernels (that is, those giving faster rates of convergence at the price of taking on negative values). However, 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 $p = 0.5$ when $F$ is symmetric, the asymptotically optimal bandwidth is given by $h_{opt} = \alpha(K) \,.\, \beta(Q) \,.\, n^{-1/3}$ where

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

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

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

When $F$ is symmetric and $p = 0.5$ taking $h = O\,(n^{-1/2})$ makes the first two terms in $h$ of the MSE of $KQ_{0.5}$ the same order and

$$MSE\,(KQ_{0.5}) = 0.25\, n^{-1}\, [Q'(½)]^2 + O\,(n^{-3/2}).$$

However, as the term in $hn^{-1}$ is negative and the term in $n^{-2}h^{-1}$ is positive there is no single bandwidth which minimizes the asymptotic mean square error of $KQ_{0.5}$ when $F$ is symmetric. Instead any $h$ satisfying $h = \text{constant}\,.\,n^{-m}$ $(0 < m \le ½)$ will, for large values of the constant, produce an estimator with smaller asymptotic mean square error than $SQ_{0.5}$.

We next present a theorem which establishes some asymptotic equivalences between the different forms of the kernel quantile estimator. In view of (2.2), we shall deem the two kernel quantile estimators $KQ_{p,i}$ and $KQ_{p,j}$ as "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

(i) for $h\, n^{2/3} \to \infty$, $KQ_p$ and $KQ_{p,2}$ are asymptotically equivalent;

(ii) for $h\, n^{2/3} \to \infty$, $KQ_{p,2}$ and $KQ_{p,1}$ are asymptotically equivalent;

(iii) for $h\, n^{5/6} \to \infty$, $KQ_{p,1}$ and $KQ_{p,3}$ are asymptotically equivalent;

(iv) for $h\, n^{5/6} \to \infty$, $KQ_p$ and $KQ_{p,4}$ are asymptotically equivalent.

The first assumption of the above theorem, rules out the normal kernel. However, 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 or Epanechnikov kernels. For a discussion of these and other kernels see Silverman (1986). However, similar results can be obtained for these, but slightly different methods of proof are required. These extensions of the above theorem 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 which show 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 $\alpha \to \infty$

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

in the sense that

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

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

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

We see from Theorem 1 that $HD_p$ is asymptotically suboptimal, being based on $h = O\,(n^{-\frac{1}{2}})$ rather than $h = O\,(n^{-\frac{1}{3}})$, resulting in weights which are too concentrated in a neighborhood of $p$. See Yashizawa et. al. (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 $0 < p < 1$), $i/n = p + O\,(k^{-\frac{1}{2}})$ and $r = pk + O(1)$ with $k = o\,(n)$ then as $n \to \infty$ and $k \to \infty$

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

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

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

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 $0 < p < 1$), $i/(n+1) = p + O\,(n^{-\frac{1}{2}})$ then as $n \to \infty$

$$\frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)} p^i q^{n-i}$$

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

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 $h = O(n^{-1/3})$.

For related asymptotic equivalence results, see Takeuchi (1971). Similar, but slightly weaker equivalences, have been obtained by Yang (1985, Theorem 3) between $KQ_p$ and $KQ_{p,1}$ and by Zelterman (1988) between $KQ_{p,1}$, $HD_p$ and $KL_p$. Pranab K. Sen has pointed out in private communication 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 $p = 0.5$ 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 zero (that is, $\int_{-\infty}^{\infty} K_*(u)\, du = 1$,

$\int_{-\infty}^{\infty} u^i K_*(u)\,du = 0 \quad i = 1, 2, \ldots, m-1$ and $\int_{-\infty}^{\infty} u^m K_*(u)\,du < \infty$).

The resulting estimate of the asymptotically optimal bandwidth is given by

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

where $\hat{\beta} = [\hat{Q}_m'(p) \Big/ \hat{Q}_m''(p)]^{2/3}$ 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 neighbourhood of $p$ and that $K_*$ is a compactly supported kernel of order $m$, symmetric about zero. The asymptotically optimal bandwidth for $\hat{Q}_m'(p)$ is given by

$$a_{opt} = \mu_m(K_*) \cdot \gamma_m(Q) \cdot n^{-1/(2m+1)}$$

where

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

and $\gamma_m(Q) = [Q'(p) \Big/ Q^{(m+1)}(p)]^{2/(2m+1)}$. The asymptotically optimal bandwidth for $\hat{Q}_m''(p)$ is given by

$$b_{opt} = \tau_m(K_*) \cdot \delta_m(Q) \cdot n^{-1/(2m+3)}$$

where

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

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

In view of the above theorem, we can choose the bandwidths for $\hat{Q}_m'(p)$ and $\hat{Q}_m''(p)$ to be $a = c_m' \cdot \mu_m(K_*) \cdot n^{-1/(2m+1)}$ and $b = c_m'' \cdot \tau_m(K_*) \cdot n^{-1/(2m+3)}$ 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 $Q'(0.5)$.

Yang (1985) proposed an alternative method of obtaining a data-base choice of the bandwidth, $h$. This method uses the bootstrap to estimate the mean square error of $KQ_p$ over a grid of values of $h$. The value of $h$ that minimizes this estimated mean square error is used as the bandwidth for $KQ_p$. Padgett and Thombs (1986) have extended this approach to right-censored data. There are two disadvantages associated with this approach. The first is the massive amount of computation required to compute the data-based bandwidth. Secondly, an estimate of $\xi_p$ is used as the value of $\xi_p$ in the calculation of the bootstrap estimates of mean square error: [Yang (1985) used the sample quantile for this purpose.] An appealing feature of the bootstrap approach is it does not employ asymptotic motivation.

Another bandwidth selector, based on cross-validation, has been proposed by Zelterman (1988). This approach is not directly comparable to ours, because our goal is to find the best bandwidth for a given $p$, while cross-validation yields a single bandwidth which attempts to optimize a type of average over $p$.

## 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 pseudo-random samples of size 50 and 100 were generated from the double exponential, exponential, lognormal and normal distributions. Over the 1,000 samples, we calculated the mean square error for the estimators given below at the 0.05, 0.1, 0.25, 0.5, 0.75, 0.9 and 0.95 quantiles.

To implement the data-based algorithm of the previous section, 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 $\hat{Q}_2''(p)$ while 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 $\delta_4(Q)$ (that is, 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 the values possible. This can be demonstrated by calculating these functionals for a family of distributions such as the generalized lambda distribution (Ramberg et al., 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 strategy of using the same $\beta(Q)$ and hence the same bandwidth for all underlying distributions as is essentially done by $HD_p$ and $B_p$.

— Table 1 here —

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 kernel, given in Müller (1984), was also used to estimate $Q''(p)$

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

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

$$\hat{Q}_m'(p) \simeq \sum_{i=1}^{n} \left[ n^{-1} a^{-2} K_*' \left[ a^{-1} \left[ \frac{i-½}{n} - p \right] \right] \right] X_{(i)}$$

$$\hat{Q}_m''(p) \simeq \sum_{i=1}^{n} \left[ n^{-1} b^{-3} K_*'' \left[ b^{-1} \left[ \frac{i-½}{n} - p \right] \right] \right] X_{(i)}.$$

Three different values of each of the constants $c_2'$, $c_2''$ and $c_4''$ were used for each value of $p$. Experience with $\hat{h}_{opt}$, 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 [0.05, 1.5] was set equal to the closest endpoint of this interval.

The values of the constants $c_2'$, $c_2''$ and $c_2'$, $c_4''$ which consistently produced the smallest mean square error for $KQ_{p,4}$ over the four distributions considered in this study are given in Table 2. We denote by $KQ_{p,4}^{(1)}$ 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 $KQ_{p,4}^{(2)}$ denote $KQ_{p,4}$ based on $h$ obtained from $\hat{Q}_2'(p)$ and $\hat{Q}_4''(p)$, using the values of $c_2'$ and $c_4''$ given in Table 2.

— Table 2 here —

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 $h_{opt} = [pq/k_{opt}]^{1/2}$. 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 (that is, ratio of mean square errors) 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 $KQ_{p,4}^{(2)}$ gave moderately superior performance to $KQ_{p,4}^{(1)}$ and $HD_p$ producing smaller mean square errors in 26 and 28 out of the 44 combinations, respectively. $KQ_{p,4}^{(2)}$ 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 $k$. However, $KL_p$ based on the fixed choices of $k$ generally performed worse than both $KQ_{p,4}^{(2)}$ 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, which is given by (2.3), estimate is fixed at a point which is often not too far from the optimum bandwidth in samples of size 50 and 100. Figure 1 contains plots of the asymptotic mean square error of $KQ_p$, obtained from the expression given in Theorem 1, for the 0.1 and 0.9 quantiles of the lognormal distribution when $n = 50$. The asymptotically 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 0.1 quantile these two bandwidths are close together, while for the 0.9 quantile they are well separated. This explains why the Harrell-Davis estimator performs better for the 0.1 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 0.9 quantile the center of the distribution of the data-based bandwidths is close to the optimum bandwidth while for the 0.1 quantile it is not. This explains the better performance of $KQ_{p,4}^{(2)}$ for the 0.9 quantile.

— Figure 1 here —

Because of the noise inherent in our data-based bandwidths, we considered using a fixed bandwidth for $KQ_p$ which was less arbitrary than the bandwidth for $HD_p$. The bandwidth we chose corresponds to the asymptotically optimal when the underlying distribution is normal. (This is undefined at $p = 0.5$ for which we set $h$ equal to the bandwidth corresponding to an exponential distribution.) We denote this estimator by $KQN_p$. $KQN_p$ had larger mean square error than $HD_p$ and $KQ_{p,4}^{(2)}$ in 23 and 27 out of the 44 combinations, respectively.

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 $SQ_{(p)}$.

— Figure 2 here —

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

Proof of Theorem 1. We first consider all $p$, apart from $p = 0.5$ when $F$ is symmetric. Since $K$ is compactly supported and $Q''$ is continuous in a neighborhood of $p$, we find using (4.6.3) of David (1981) that

$$\text{Bias}\,(KQ_p) = \sum_{i=1}^{n} [\int_{i-1/n}^{i/n} K_h\,(t-p)\,\text{dt}]\,\{Q\,(\frac{i}{n+1}) - Q\,(p)\} + O\,(n^{-1})$$

$$= \int_0^1 K_h\,(t-p)\,\{Q\,(t) - Q\,(p)\}\,\text{dt} + O\,(n^{-1})$$

$$= \tfrac{1}{2} h^2\,[\int_{-\infty}^{\infty} u^2 K\,(u)\,\text{du}]\,Q''(p) + o\,(h^2) + 0(n^{-1}).$$

Falk (1984, p.263) proved that

$$\text{Var}\,(KQ_p) = n^{-1}\,p\,(1-p)\,[Q'(p)]^2 - n^{-1}\,h\,[Q'(p)]^2 \int_{-\infty}^{\infty} u\,K\,(u)\,K^{(-1)}(u)\,\text{du} + o\,(n^{-1}h).$$

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, $KQ_{0.5}$ is symmetrically distributed about its mean $\xi_{0.5}$. The expression for $MSE$ $(KQ_{0.5})$ is found by extending Falk's expansion for Var $(KQ_p)$ to include the next term.

Proof of Theorem 2. We only give the details for (i). The proofs of (ii), (iii) and (iv) follow in a similar manner. Let

$$W_{n,h}(i) = \int_{i-1/n}^{i/n} K_h\,(t-p)\,\text{dt} - n^{-1}\,K_h\,(\frac{i-\tfrac{1}{2}}{n} - p).$$

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 $O\,(n\,h)$, we find using (4.6.1) and (4.6.3) of David (1981) that

$$E\,[KQ_p - KQ_{p,2}]^2 = E\,[\sum_{i=1}^{n} W_{n,h}(i)\,X_{(i)}]^2$$

$$= E\,[\sum_{i\in S} W_{n,h}(i)\,\{X_{(i)} - E\,(X_{(i)})\}]^2 + [\sum_{i\in S} W_{n,h}(i)\,E\,(X_{(i)})]^2$$

$$= O\,(n^{-4}\,h^{-4})$$

$$= o\,(n^{-4/3})$$

if $h\,n^{2/3} \longrightarrow \infty$ as $n \longrightarrow \infty$.

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

## 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 Functions," unpublished manuscript.

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

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

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

Gasser, T. and Müller, H.G. (1979), "Kernel Estimation of Regression Functions," *Lecture Notes in Mathematics* 757, 144–154.

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

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

Härdle, W. (1988), *Applied Nonparametric Regression*, unpublished manuscript.

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.

Kaigh, W.D. (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.

Nadaraya, E.A. (1964), "On Estimating Regression", *Theory of Probability and Its Applications*, 9, 141–142.

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 Modelling," *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.

Priestley, M.B. and Chao, M.T. (1972), "Nonparametric Function Fitting," *Journal of the Royal Statistical Society*, Series B, 34, 385–392.

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.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Takeuchi, K. (1971), "A Uniformly Asymptotically Efficient Estimator of a Location Parameter", *Journal of the American Statistical Association*, 66, 292–301.

Watson, G.S. (1964), "Smooth Regression Analysis", *Sankhya Series A*, 26, 359–372.

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.

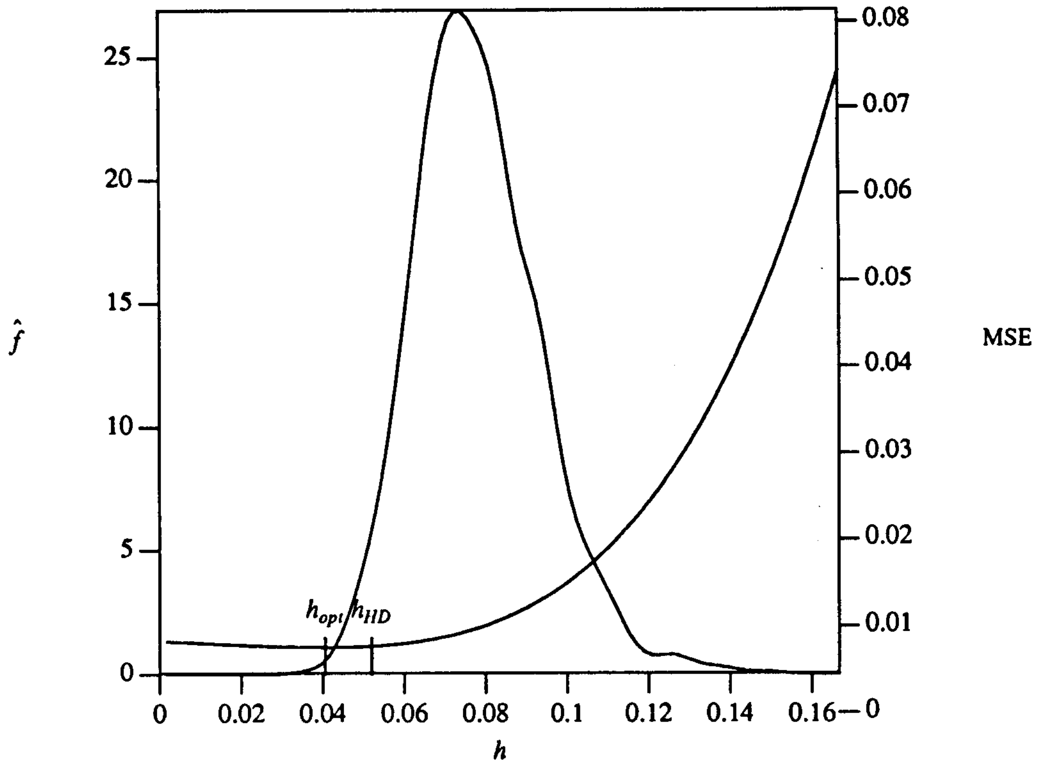
**Table 1** *Values of Functionals of $Q$ in Asymptotically Optimal Bandwidths for Kernel Estimates of $Q(p)$, $Q'(p)$ and $Q''(p)$.*

| $p$ | | Double Exponential | Normal | Lognormal | Exponential |
|---|---|---|---|---|---|
| 0.05, 0.95 | $\beta(Q)$ | 0.14 | 0.16 | 0.12, 0.29 | 0.97, 0.14 |
| | $\gamma_2(Q)$ | 0.07 | 0.08 | 0.05, 0.40 | 0.73, 0.07 |
| | $\delta_2(Q)$ | 0.05 | 0.05 | 0.03, 0.11 | 0.57, 0.05 |
| | $\delta_4(Q)$ | 0.03 | 0.03 | 0.01, 0.07 | 0.40, 0.03 |
| 0.1, 0.9 | $\beta(Q)$ | 0.22 | 0.27 | 0.18, 0.73 | 0.93, 0.22 |
| | $\gamma_2(Q)$ | 0.12 | 0.14 | 0.08, 0.40 | 0.70, 0.12 |
| | $\delta_2(Q)$ | 0.08 | 0.09 | 0.05, 0.16 | 0.55, 0.08 |
| | $\delta_4(Q)$ | 0.05 | 0.06 | 0.03, 0.08 | 0.38, 0.05 |
| 0.25, 0.75 | $\beta(Q)$ | 0.40 | 0.61 | 0.33, 0.98 | 0.83, 0.40 |
| | $\gamma_2(Q)$ | 0.25 | 0.31 | 0.15, 0.29 | 0.60, 0.25 |
| | $\delta_2(Q)$ | 0.18 | 0.22 | 0.09, 0.17 | 0.47, 0.18 |
| | $\delta_4(Q)$ | 0.12 | 0.13 | 0.05, 0.09 | 0.32, 0.10 |
| 0.5 | $\beta(Q)$ | – | – | 0.54 | 0.63 |
| | $\gamma_2(Q)$ | – | – | 0.23 | 0.44 |
| | $\delta_2(Q)$ | – | – | 0.14 | 0.33 |
| | $\delta_4(Q)$ | – | – | 0.08 | 0.22 |

**Table 2** *Values of the constants $c_2{}'$, $c_2{}''$ and $c_2{}'$, $c_4{}''$ which consistently produce the smallest mean square error for $KQ_{p,4}$.*

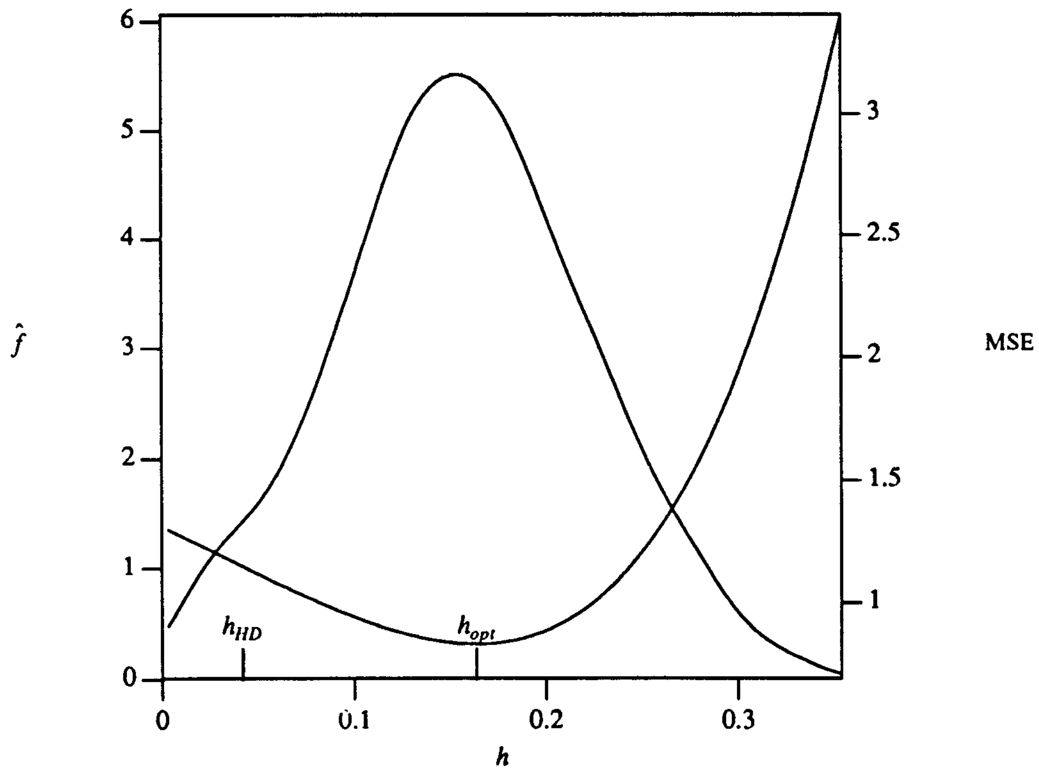
| $p$ | $c_2{}'$, $c_2{}''$ | $c_2{}'$, $c_4{}''$ |
|---|---|---|
| 0.05, 0.95 | 0.75, 0.6 | 0.75, 0.4 |
| 0.1, 0.9 | 0.2, 0.6 | 0.2, 0.4 |
| 0.25, 0.75 | 0.6, 0.5 | 0.6, 0.3 |
| 0.5 | 0.8, 0.3 | 0.4, 0.2 |

(a) $p = 0.1$



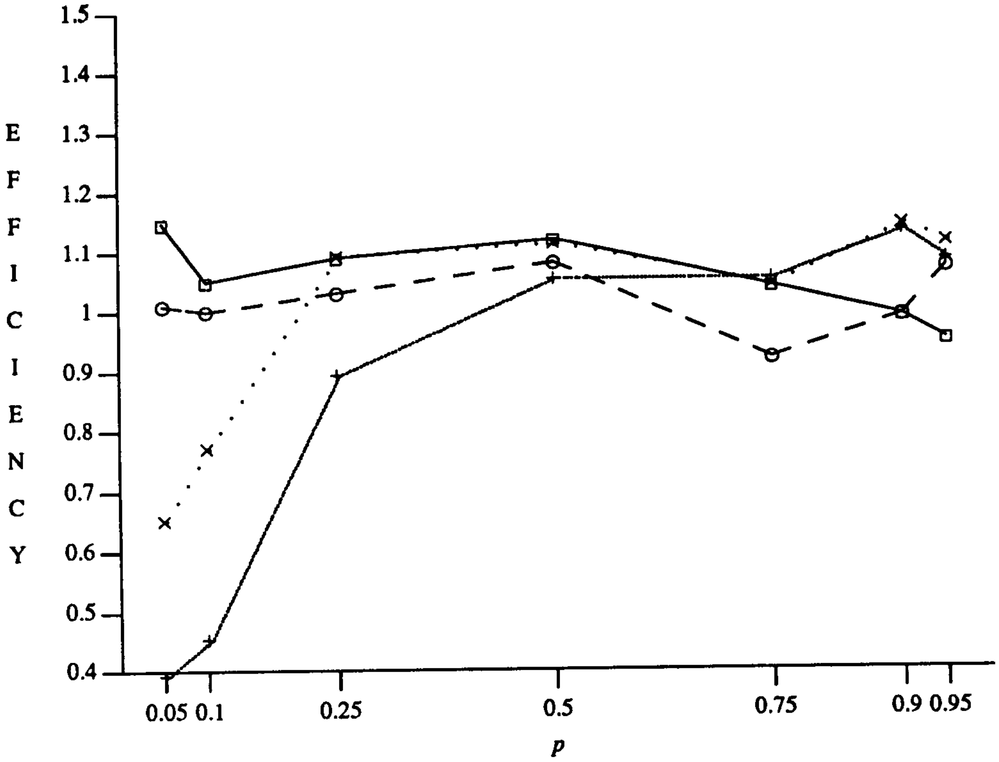
25
20
15
10
5
0
$\hat{f}$
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
MSE
$h_{opt}$
$h_{HD}$
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
$h$


(b) $p = 0.9$

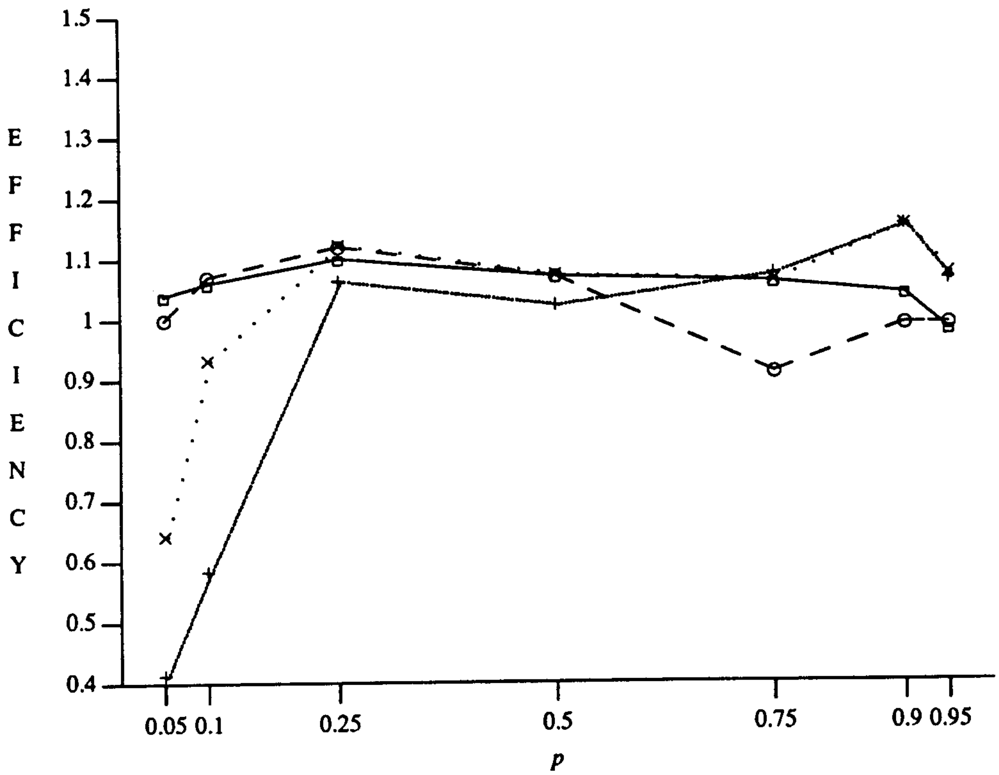


6
5
4
3
2
1
0
$\hat{f}$
3
2.5
2
1.5
1
MSE
$h_{HD}$
$h_{opt}$
0
0.1
0.2
0.3
$h$

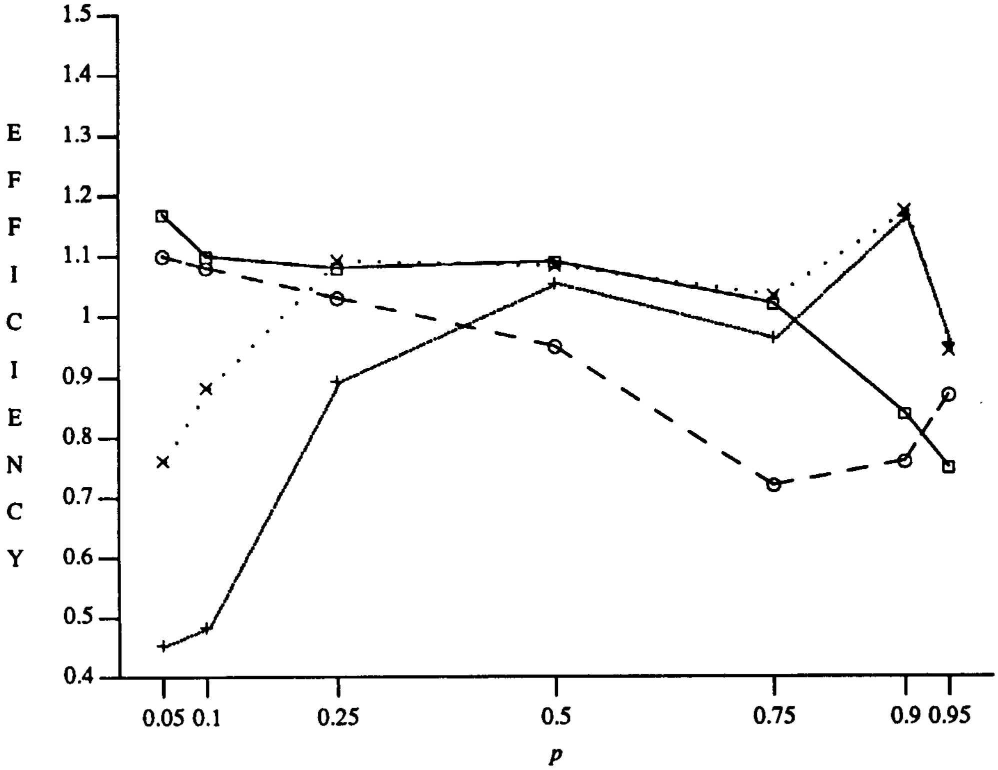
(a) Exponential $n = 50$



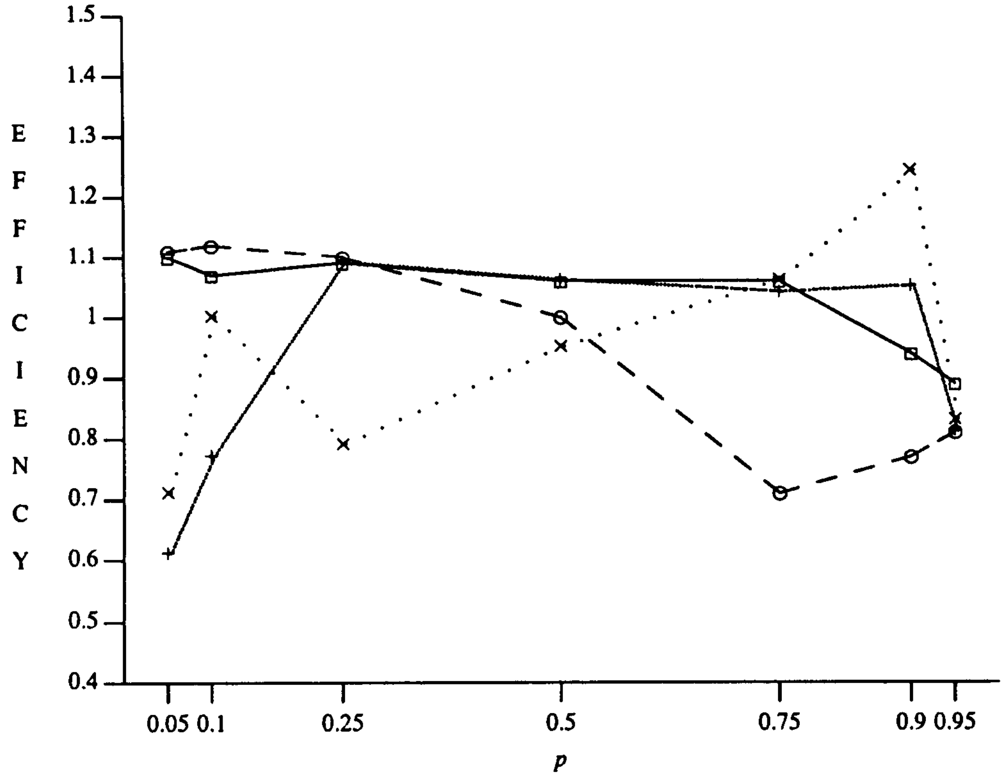

(b) Exponential $n = 100$



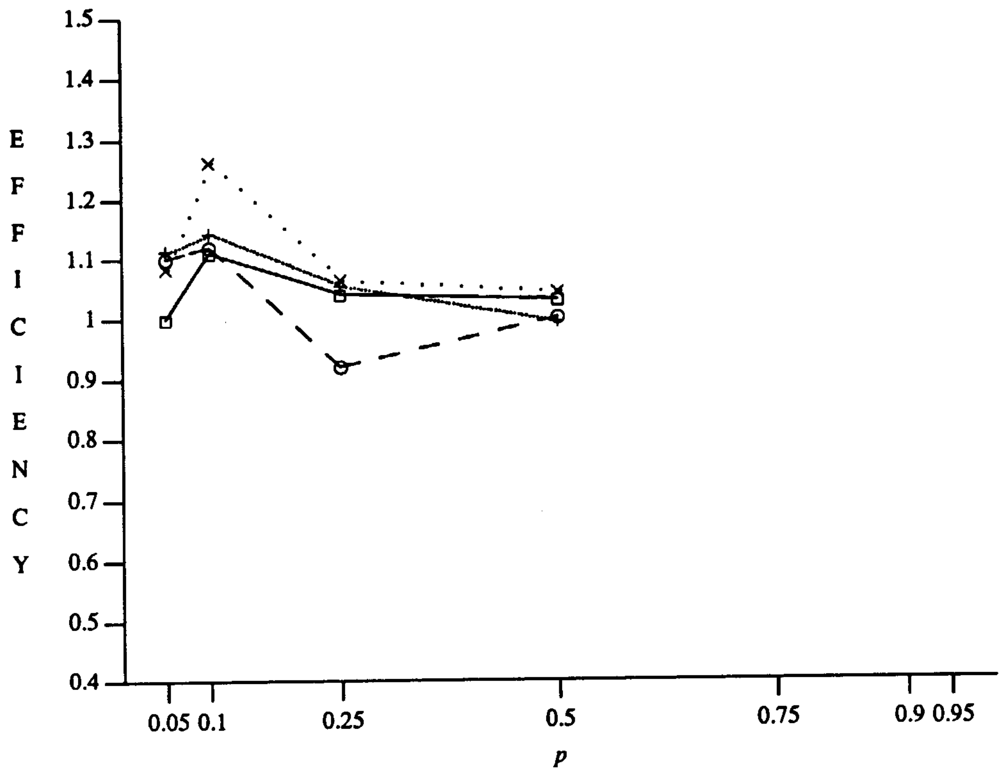
(c) Lognormal $n = 50$



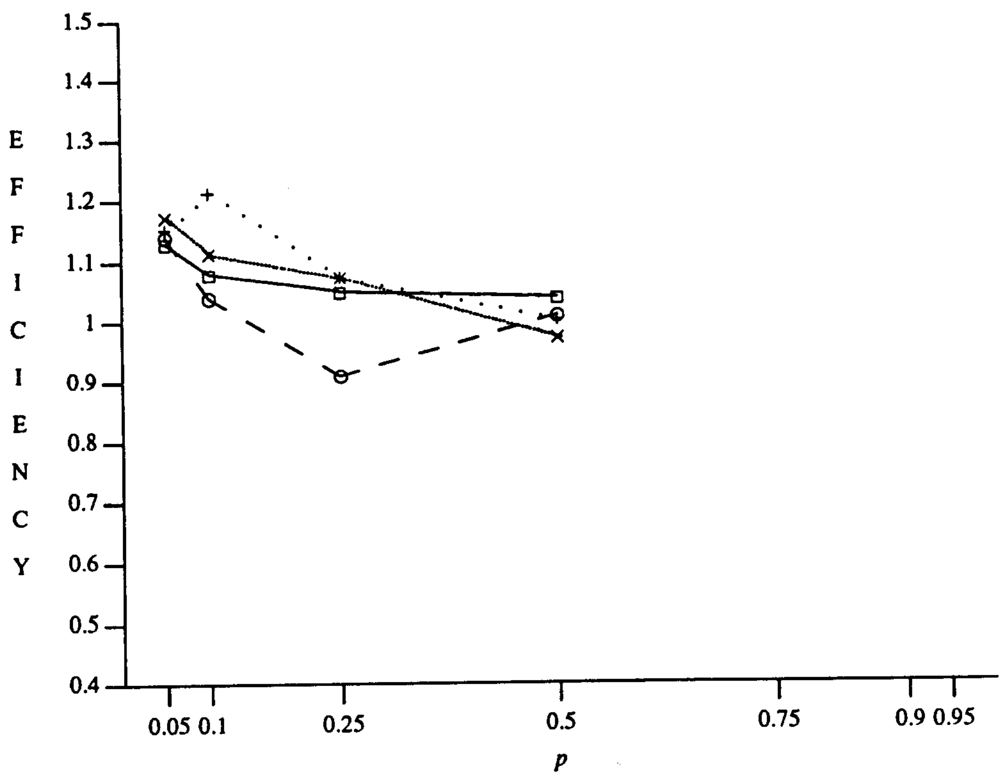

(d) Lognormal $n = 100$



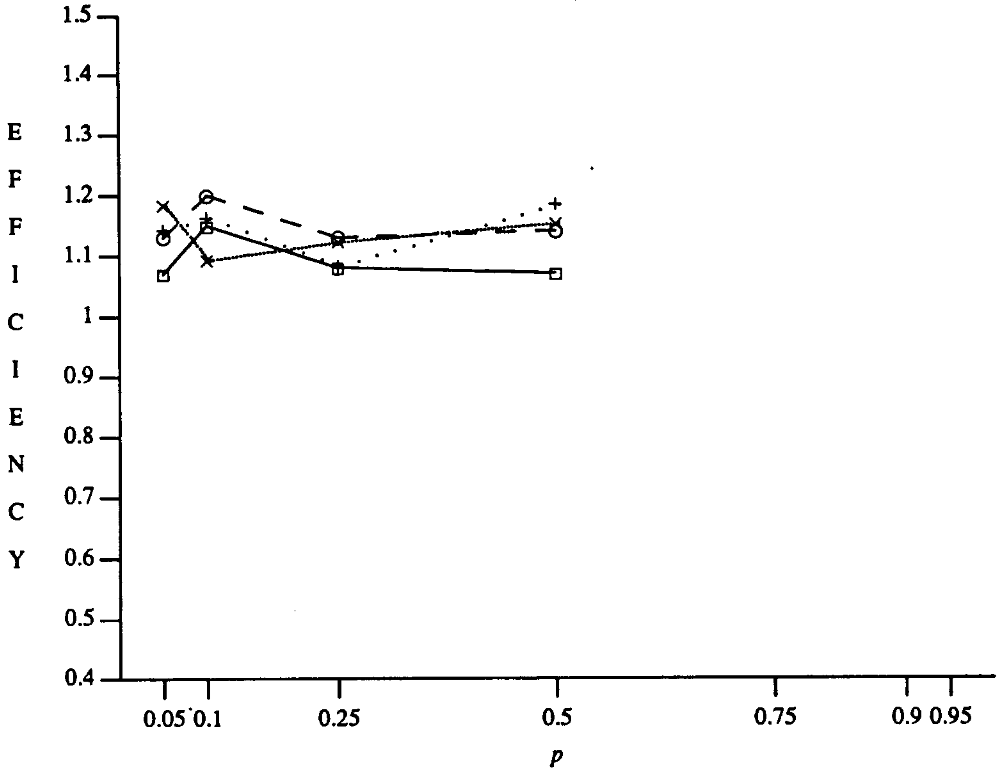
(e) Double Exponential $n = 50$



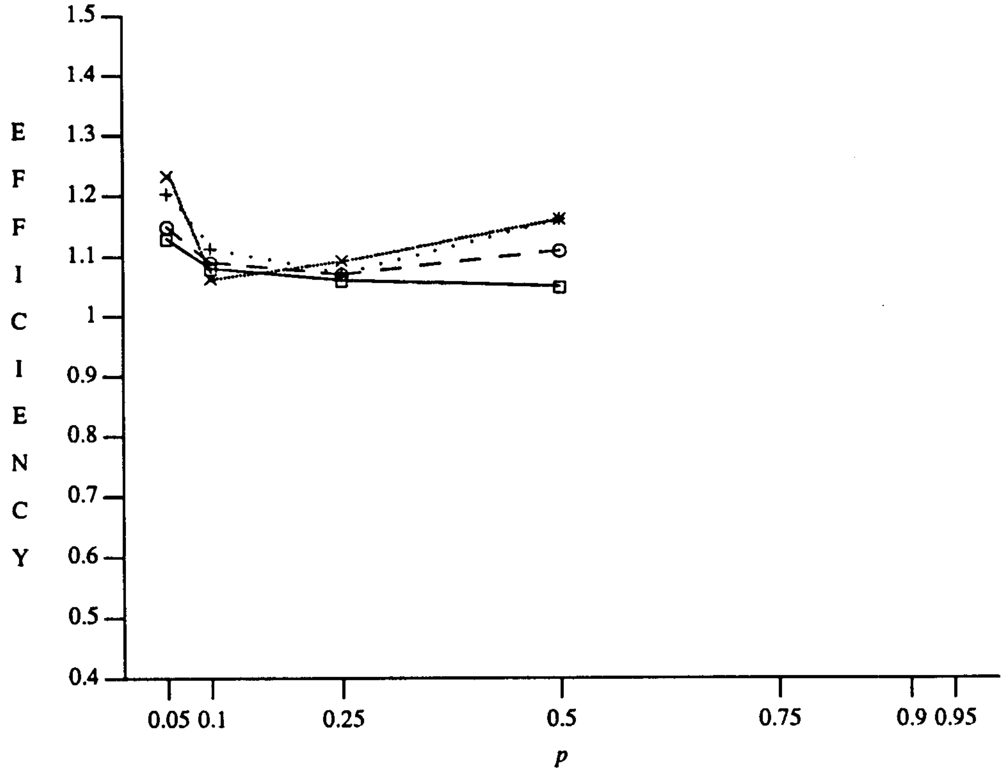

(f) Double Exponential $n = 100$



(g) Normal $n = 50$




(h) Normal $n = 100$



## FIGURE LEGENDS

Figure 1. Plots of the asymptotic mean square error of $KQ_p$ versus $h$ for the 0.1 and 0.9 quantiles of the lognormal distribution when $n = 50$. Estimates of the density of the data-based bandwidths are also included in the plots.

Figure 2. Plots of the ratio of the mean square error of each of the estimators $HD_p$ (——□——), $KQN$ (— — ○ — —), $KQ_{p,4}^{(1)}$ (..........+..........) and $KQ_{p,4}^{(2)}$ (. . . . × . . . .) to the sample quantile $SQ_p$.