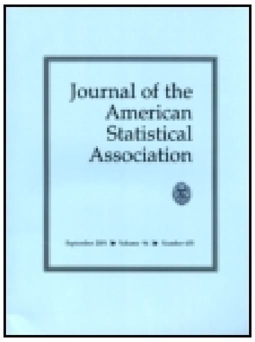

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Local Linear Quantile Regression

Keming Yu & M. C. Jones

To cite this article: Keming Yu & M. C. Jones (1998) Local Linear Quantile Regression, Journal of the American Statistical Association, 93:441, 228-237, DOI: 10.1080/01621459.1998.10474104

To link to this article: https://doi.org/10.1080/01621459.1998.10474104

Published online: 17 Feb 2012.

Local Linear Quantile Regression

Keming YU and M. C. JONES

In this article we study nonparametric regression quantile estimation by kernel weighted local linear fitting. Two such estimators are considered. One is based on localizing the characterization of a regression quantile as the minimizer of $E\{\rho_p(Y-a)|X=x\}$, where ρ_p is the appropriate "check" function. The other follows by inverting a local linear conditional distribution estimator and involves two smoothing parameters, rather than one. Our aim is to present fully operational versions of both approaches and to show that each works quite well; although either might be used in practice, we have a particular preference for the second. Our automatic smoothing parameter selection method is novel; the main regression quantile smoothing parameters are chosen by rule-of-thumb adaptations of state-of-the-art methods for smoothing parameter selection for regression mean estimation. The techniques are illustrated by application to two datasets and compared in simulations.

KEY WORDS: Bandwidth selection; Conditional quantile; Kernel estimator; Local linear regression; Reference chart; Rule of thumb.

1. INTRODUCTION

Although most regression investigations are concerned with the regression *mean* function $m(x)$, the conditional mean of a response Y given values x of a predictor X, other aspects of the conditional distribution of Y given X are also often of intererst. Quantile regression, the topic of this article, concerns the conditional *quantile* functions $q_p(x), 0 < p < 1$, of Y given $X = x$. Individual quantile functions, particularly the conditional median, are sometimes of interest, but more often one wishes to obtain a collection of conditional quantiles that when graphed give an impression of the entire conditional distribution of Y given X. Alternatively, pairs of extreme conditional quantiles map out a conditional prediction interval within which one expects the majority of individual points to lie. These "reference charts" are popular in medicine (see, e.g., Cole 1988) and have provided a stimulus for much of the recent statistical work in the area.

An example of this type is given in Figure 1. The dataset, a scatterplot of which is given in Figure 1a, comes from an anthropometric survey on triceps skinfold measurements of 892 girls and women up to age 50 in three Gambian villages, seen during the dry season of 1989. To understand "normal" variation in this measure with age, it is useful to look at a collection of estimated quantiles as a function of age as given in Figure 1b. (These quantiles, at .03, .1, .25, .5, .75, .9, and .97 levels, are those provided by the preferred version of our methodology; see Sec. 4.1.) The conditional quantiles map out the main features of development of triceps skinfold with age; both central and extreme quantiles show a decrease from 0 to about 10 years, but from there on, higher quantiles increase more rapidly than lower quantiles to a relatively steady state such that adulthood corresponds to a much greater variability in triceps skinfold than does childhood. These data were earlier used by Cole and Green (1992) for similar purposes.

The seminal (parametric) work of Koenker and Bassett (1978) was a major step forward in estimating conditional quantiles. In this article we are concerned with the nonparametric estimation of conditional quantile functions. Of course, many others since Koenker and Bassett have considered this problem, using a variety of ad hoc methods and adapting various smoothing techniques. For estimation of the regression mean, local polynomial fitting and particularly its special case local linear fitting have become increasingly popular. This is the stuff of lowess (Cleveland 1979; Cleveland and Devlin 1988; Stone 1977) and it has also been further recognized to have various advantages through recent work such as that of Cleveland and Loader (1996), Fan (1992), Fan and Gijbels (1992, 1995), Hastie and Loader (1993), and Ruppert and Wand (1995); see also Fan and Gijbels (1996) and Wand and Jones (1995, chap. 5). Unsurprisingly, local polynomial fitting, and in particular local linear fitting, can be adapted to quantile regression and its advantages will be maintained, as we describe later.

The purpose of this article is to develop local linear approaches to quantile regression in such a way that the results are immediately applicable in practice. The basic techniques are not novel to this article (Chaudhuri 1991; Fan, Hu, and Truong 1994; Fan, Yao and Tong 1996 have provided relevant theoretical background), but many details and insights are.

In fact, we present and develop two alternative local linear quantile regression methods that as we show in practical terms in Section 4, are broadly equivalent in terms of results. Although the user could certainly contemplate using either, we show that we have a preference for the second method that we describe. The first method is the more direct, however. The estimated quantile function $\hat{q}_p(x)$ is based on minimizing a local linear kernel weighted version of $E\{\rho_p(Y-a)|X=x\}$, where ρ_p is the "check" function

Keming Yu is Research Fellow and M. C. Jones is Reader, Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, U. K. The first author was supported by an Open University Research Studentship; this work forms part of his Ph.D. thesis. The authors thank Tim Cole and Patrick Royston for providing datasets and Matt Wand for providing S code implementing the Ruppert–Sheather–Wand bandwidth selection method. Comments of the referees were particularly helpful in improving the manuscript and the methodology concerning h_2.

Journal of the American Statistical Association
March 1998, Vol. 93, No. 441, Theory and Methods

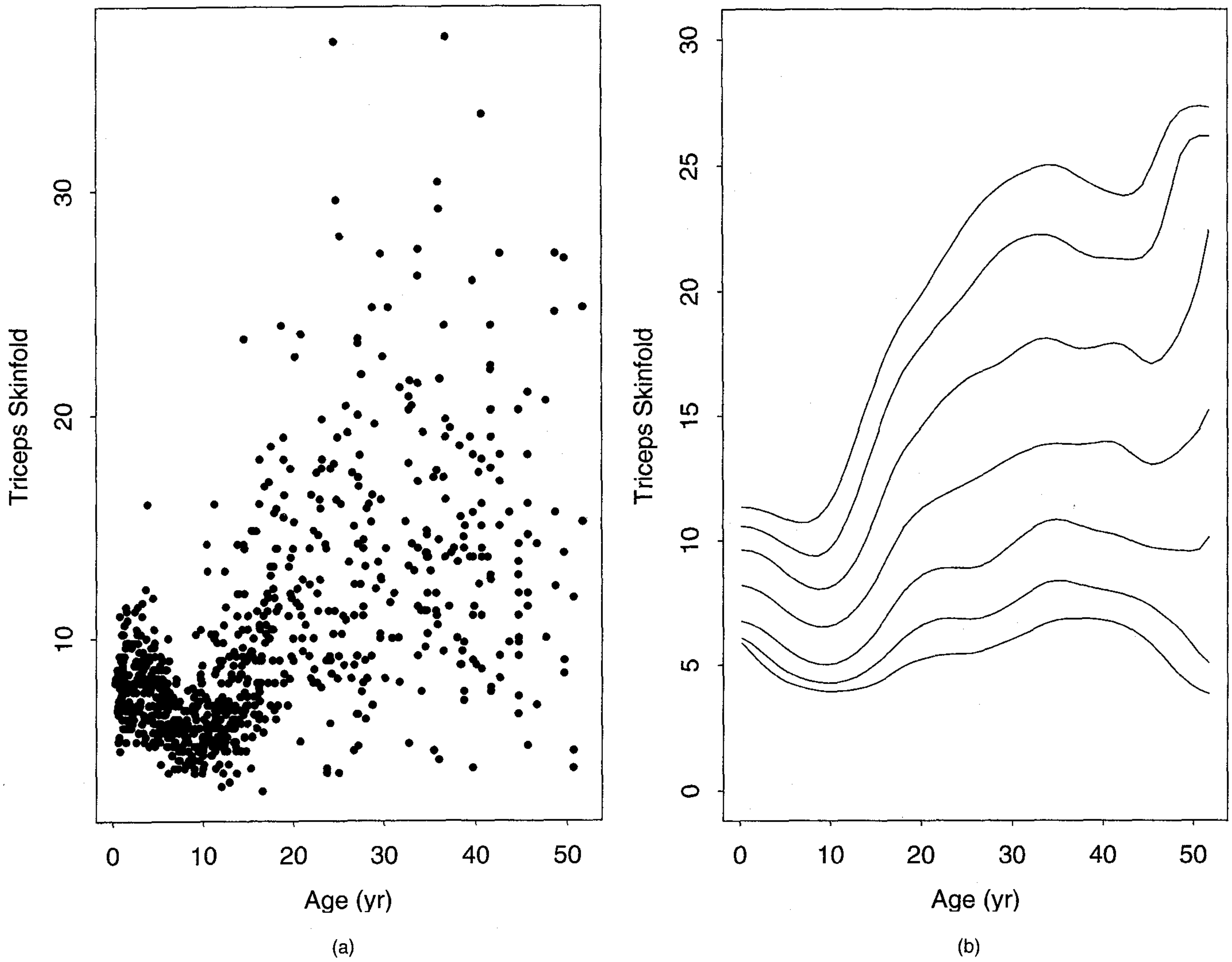

Figure 1. The Triceps Skinfold Data. (a) Scatterplot; (b) smoothed 3rd, 10th, 25th, 50th, 75th, 90th, and 97th quantile curves using double-kernel smoothing.

given by

$$\rho_p(z) = pzI_{[0,\infty)}(z) - (1-p)zI_{(-\infty,0)}(z), \tag{1}$$

where p indexes the conditional quantile of current interest. This method involves a kernel localization function K that we take to be a symmetric probability density function, and its scale parameter h is the bandwidth that controls the amount of smoothing applied to the data. Section 2.1 describes motivation and full description; Section 2.2, asymptotic mean squared error properties; and Section 2.3, a novel bandwidth selection rule covering all desired conditional quantiles at once. The latter is fully practical and successfully used.

In Section 3 we make a parallel development of an alternative "double-kernel" approach to conditional quantile estimation by first taking a kernel weighted local linear approach to estimating the conditional distribution function. In this case we allow two bandwidths, an additional one "in the y direction" to add to the existing one "in the x direction." Having to select a second bandwidth is an unappealing feature, but in Section 3.3 we show how to specify a rule-of-thumb for this that works in conjunction with the rule of Section 2.3 for the x direction bandwidth. The theoretical work of Section 3.2, plus practical experience, shows that the precise value of this second bandwidth is, unsurprisingly, not critical, but a practical rule (that depends on p) is still needed and is given here. To define the double-kernel quantile estimator $\tilde{q}_p$, solve

$$p = \frac{1}{\sum_j w_j(x;h_1)} \sum_j w_j(x;h_1)\Omega\left(\frac{\tilde{q}_p(x) - Y_j}{h_2}\right), \tag{2}$$

where Ω is the distribution function associated with a kernel density function W. Here $w_j(x;h_1)$ is the weight function associated with local linear fitting,

$$w_j(x;h_1) = K\left(\frac{x - X_j}{h_1}\right)[S_{n,2} - (x - X_j)S_{n,1}], \tag{3}$$

with

$$S_{n,l} = \sum_{i=1}^{n} K\left(\frac{x - X_j}{h_1}\right)(x - X_i)^l, \qquad l = 1, 2,$$

and the two bandwidths h_1 and h_2 relate to x and y direction smoothing. The right side of (2) is a (double-kernel) local linear estimate of the conditional distribution function, and the equation defines its inverse, our conditional quantile estimator. Further details on $\tilde{q}_p$ are given in Section 3.1.

Our preference for $\tilde{q}_p$ over $\hat{q}_p$ will be seen to reside in the smoother appearance of the former, its better squared error properties in simulations, and solution of a problem with $\hat{q}_p$ that allows estimated quantiles to cross.

2. LOCAL LINEAR CHECK FUNCTION MINIMIZATION

2.1 The Method

Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a set of independent observations from some underlying distribution $F(x, y)$ with density $f(x, y)$, and interest centers on the responses Y_i considered to be realizations from the conditional distribution $F(y|x)$ or density $f(y|x)$ of Y given $X = x$. A characterization of the pth conditional quantile $q_p(x)$ is as

$$q_p(x) = \operatorname{argmin}_a E\{\rho_p(Y - a)|X = x\},$$

with ρ_p given by (1). A first, "local constant" sample version of this might be to take

$$\bar{q}_p(x) = \operatorname{argmin}_a \sum_{i=1}^{n} \rho_p(Y_i - a) K\left(\frac{x - X_i}{h}\right).$$

Here h and K are the bandwidth and kernel function mentioned before.

In regression mean estimation, however, local linear fitting is nowadays considered superior to local constant fitting (see the references cited in Section 1). We thus consider the local linear version of the foregoing here. (Direct comparisons between local constant and local linear approaches in the conditional quantile estimation setting are made in Yu and Jones 1997.) The idea of the local linear fit is to approximate the unknown pth quantile $q_p(x)$ by a linear function $q_p(z) = q_p(x) + q_p'(x)(z - x) \equiv a + b(z - x)$ for z in a neighborhood of x. Locally, estimating $q_p(x)$ is equivalent to estimating a, whereas estimating $q_p'(x)$ is equivalent to estimating b. This motivates us to define an estimator by setting $\hat{q}_p(x) = \hat{a}$, where $\hat{a}$ and $\hat{b}$ minimize

$$\sum_{i=1}^{n} \rho_p(Y_i - a - b(X_i - x)) K\left(\frac{x - X_i}{h}\right). \tag{4}$$

As shown in Sections 2.2 and 2.3, this estimating method maintains the various advantages of local linear mean fitting, such as design adaptation and good boundary behavior, in a conditional quantile context. Formula (4) has also been addressed (essentially) by Chaudhuri (1991), Fan et al. (1994), and Koenker, Portnoy, and Ng (1992).

To calculate $\hat{q}_p$, we used an iteratively reweighted least squares algorithm, details of which have been provided by Yu (1997).

2.2 Mean Squared Error

An important way of assessing the performance of $\hat{q}_p(x)$ is by its (conditional or unconditional) mean squared error (MSE). For local linear conditional quantile fitting, the asymptotic form of $\text{MSE}(\hat{q}_p(x))$ (as $n \to \infty, h = h(n) \to 0$ and $nh \to \infty$) has already been given by Fan et al. (1994). We reproduce their result here because we make use of it in the following subsection. Under certain conditions given by Fan et al., including that x be not too near a boundary of the support of the design and a smoothness condition on $q_p(x)$,

$$\text{MSE}(\hat{q}_p(x)) \simeq \frac{1}{4} h^4 \mu_2(K)^2 q_p''(x)^2 + \frac{R(K)p(1-p)}{nhg(x)f(q_p(x)|x)^2}, \tag{5}$$

where $\mu_2(K) = \int u^2 K(u)\, du, R(K) = \int K^2(u)\, du$, and g is the "design density," the marginal density of X. Also from Fan et al. (1994), if $x = ch, 0 < c < 1$, is a boundary point of the design (and we take K to have support $[-1, 1]$ and g to have support [0, 1]), then

$$\text{MSE}(\hat{q}_p(ch)) \simeq \frac{1}{4} h^4 \alpha_c^2(K) q_p''(0+)^2 + \frac{\beta_c(K)p(1-p)}{nhg(0+)f(q_p(0+)|0+)^2},$$

where

$$\alpha_c(K) = \frac{a_2^2(c;K) - a_1(c;K)a_3(c;K)}{a_0(c;K)a_2(c;K) - a_1^2(c;K)},$$

$$\beta_c(K) = \frac{\int_{-1}^{c} \{a_2(c;K) - a_1(c;K)u\}^2 K(u)\, du}{\{a_0(c;K)a_2(c;K) - a_1^2(c;K)\}^2},$$

and

$$a_l(c;K) = \int_{-1}^{c} u^l K(u)\, du, \qquad l = 0, 1, 2.$$

Of course, $g(0+) = \lim_{z \downarrow 0} g(z)$.

These MSE expressions reflect two of the major advantages of local linear fitting and show that these advantages apply to the quantile regression problem as much as to regression mean estimation. These advantages are (a) no dependence of the asymptotic bias on the design density g, and indeed its dependence only on the simple quantile curvature function q_p''; and (b) automatic good behavior at boundaries, at least with regard to orders of magnitude, without the need for further boundary correction.

2.3 Bandwidth Selection

With the basic model in place, one must face the important bandwidth selection problem, as the quality of the curve estimates depends sensitively on the choice of h. For applications, a convenient and effective data-based rule will always be expected. However, almost nothing has been done so far about this problem in the context of estimating $q_p(x)$ [Fan and Gijbels (1996) indicated an alternative path to follow], and the selection problem is difficult even in the simpler case of kernel density estimation (Jones, Marron, and Sheather 1996).

Our starting point is the asymptotically optimal (interior) bandwidth that follows from Section 2.2. It is h_p, say, where

$$h_p^5 = \frac{R(K)p(1-p)}{n\mu_2(K)^2 q_p''(x)^2 g(x) f(q_p(x)|x)^2}. \tag{6}$$

This gives a relationship between optimal bandwidths for different values of p:

$$\left(\frac{h_{p_1}}{h_{p_2}}\right)^5 = \frac{p_1(1-p_1)}{p_2(1-p_2)} \frac{q''_{p_2}(x)^2 f(q_{p_2}(x)|x)^2}{q''_{p_1}(x)^2 f(q_{p_1}(x)|x)^2}. \quad (7)$$

We now simplify the relationship between h_{p_1} and h_{p_2} by making approximations to the unknown quantities involved. We fully recognize that in some cases these approximations will be far from the truth, but they afford a rule of thumb whose usefulness is to be seen in practice and simulations. First, even though $q_p(x)$ itself might vary considerably with x in terms of curvature, at any one point the second derivatives of any two quantiles will often be very similar. For example, if following the usual type of parametric regression model with identically distributed errors, the quantiles would be parallel and hence their second derivatives equal. More generally, it seems reasonable as a first-order approximation to take $q''_{p_1}(x) = q''_{p_2}(x)$. But equality is not an appropriate approximation for $f(q_p(x)|x)$, because this should be rather different for rather different p. However, we can turn to the usual type of rule-of-thumb calculations based on assuming a normal (conditional) distribution at this stage. Suppose for a moment that f is the density of a normal distribution with mean μ_x and variance σ_x^2. Then if ϕ and Φ are the standard normal density and distribution functions $f(q_p(x)|x) = \sigma_x^{-1}\phi(\Phi^{-1}(p))$, and so

$$f(q_{p_2}(x)|x)/f(q_{p_1}(x)|x) = \phi(\Phi^{-1}(p_2))/\phi(\Phi^{-1}(p_1)).$$

Using these approximations in (7) yields

$$\left(\frac{h_{p_1}}{h_{p_2}}\right)^5 = \frac{p_1(1-p_1)}{p_2(1-p_2)} \frac{\phi(\Phi^{-1}(p_2))^2}{\phi(\Phi^{-1}(p_1))^2},$$

which gives a neat, explicit, and practical way of modifying h with p.

In particular, when $p_2 = \frac{1}{2}$, we have

$$h_p^5 = \pi^{-1} 2p(1-p)\phi(\Phi^{-1}(p))^{-2} h_{1/2}^5.$$

It remains to find a method of bandwidth selection for the median. In fact, we can express the automatic bandwidth $h_{1/2}$ in terms of h_{mean}, the optimal choice of h for regression mean estimation, the automatic selection of which has already been considered elsewhere (Fan and Gijbels 1995; Ruppert, Sheather, and Wand 1995). From Fan (1993),

$$h_{\text{mean}}^5 = \frac{R(K)\sigma^2(x)}{n\mu_2(K)^2\{m''(x)\}^2 g(x)},$$

where $m(x)$ and $\sigma^2(x)$ are the conditional mean and variance. It follows that

$$\left(\frac{h_{\text{mean}}}{h_{1/2}}\right)^5 = \frac{4q''_{1/2}(x)^2\sigma^2(x)f(q_{1/2}(x)|x)^2}{m''(x)^2}.$$

By the same arguments as earlier, $q''_{1/2}(x)$ and $m''(x)$ should be similar and may be set equal, and use the normal distribution to argue that $\sigma^2(x)f(q_{1/2}(x)|x)^2$ should be replaced by $\phi(\Phi^{-1}(1/2))^2 = (2\pi)^{-1}$. Thus use

$$\left(\frac{h_{\text{mean}}}{h_{1/2}}\right)^5 = \frac{2}{\pi}.$$

In summary, we have the automatic bandwidth selection strategy for smoothing conditional quantiles as follows:

a. Use ready-made and sophisticated methods to select h_{mean}; we use the technique of Ruppert et al. (1995).
b. Use $h_p = h_{\text{mean}}\{p(1-p)/\phi(\Phi^{-1}(p))^2\}^{1/5}$ to obtain all other h_ps from h_{mean}.

To see how different bandwidths should be for different quantiles, the quantity $b(p) = \{p(1-p)/\phi(\Phi^{-1}(p))^2\}^{1/5}$ is plotted against p in Figure 2, and selected values are displayed in Table 1. These show the expected minimal smoothing for the median and a gradually increasing smoothing for the less central quantiles. The increase is of course symmetric in p and $1-p$. However, the change is very small for moderate p, and becomes considerable only as p gets close to 0 or 1. (In fact, $\lim_{p\to 0} b(p) = \lim_{p\to 1} b(p) = \infty$.)

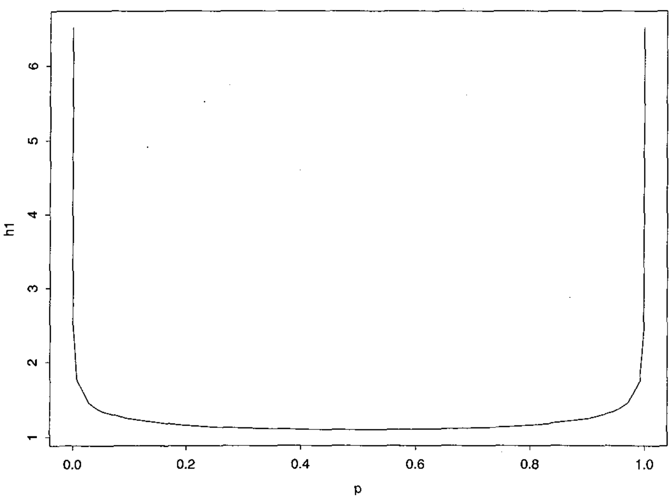

Figure 2. The Relationship Between h_p/h_{mean} and p Used by the Rule of Thumb.

3. LOCAL LINEAR DOUBLE-KERNEL SMOOTHING

3.1 The Method

Introduce another symmetric density kernel W and its distribution function Ω and note that

$$\int_{-\infty}^{y} W_{h_2}(Y_j - u)\,du = \Omega\left(\frac{y - Y_j}{h_2}\right).$$

Then, as the bandwidth $h_2 \to 0$,

$$E\left\{\Omega\left(\frac{y-Y}{h_2}\right)\middle| X = x\right\} \approx F(y|x).$$

Moreover, a local linear approach is suggested by the further approximation

$$E\left\{\Omega\left(\frac{y-Y}{h_2}\right)\middle| X = z\right\} \approx F(y|z) \approx F(y|x) + \dot{F}(y|x)(z-x) \equiv a + b(z-x),$$

Table 1. The Relationship Between Several h_ps and h_{mean} by Rule of Thumb

p	h
.025 or .975	$1.48h_{\text{mean}}$
.03 or .97	$1.44h_{\text{mean}}$
.05 or .95	$1.34h_{\text{mean}}$
.1 or .9	$1.24h_{\text{mean}}$
.25 or .75	$1.13h_{\text{mean}}$
.5	$1.095h_{\text{mean}}$

where $\dot{F}(y|x) = \partial F(y|x)/\partial x$. Then define $\tilde{F}_{h_1,h_2}(y|x) = \tilde{a}$, where

$$(\tilde{a}, \tilde{b}) = \operatorname{argmin} \sum_i \left(\Omega\left(\frac{y - Y_i}{h_2}\right) - a - b(X_i - x) \right)^2 \times K\left(\frac{x - X_i}{h_1}\right).$$

This *conditional distribution function estimator* is closely related to the conditional density function estimator of Fan et al. (1996). Explicitly,

$$\tilde{F}_{h_1,h_2}(y|x) = \frac{1}{\sum_j w_j(x;h_1)} \sum_j w_j(x;h_1)\Omega\left(\frac{\tilde{y} - Y_j}{h_2}\right), \quad (8)$$

where the weights $w_j(x;h_1)$ are given by (3). Note that this estimator can range outside [0, 1]. This does not give algorithmic difficulties in solving (9) for $0 < p < 1$. Also, $\tilde{F}_{h_1,h_2}$ is continuous with $\tilde{F}_{h_1,h_2}(-\infty|x) = 0, \tilde{F}_{h_1,h_2}(\infty|x) = 1$.

So to return to conditional quantile estimation, define $\tilde{q}_p(x)$ in principle to satisfy $\tilde{F}_{h_1,h_2}(\tilde{q}_p(x)|x) = p$ so that

$$\tilde{q}_p(x) = \tilde{F}^{-1}_{h_1,h_2}(p|x). \quad (9)$$

But $\tilde{F}_{h_1,h_2}(y|x)$ very occasionally may not be monotone everywhere in y. To overcome this difficulty, in our implementation we simply select $\tilde{q}_{1/2}(x)$ as any value that satisfies Equation (9). Then for $p > 1/2$, we take $\hat{q}_p(x)$ to be the largest value such that (9) holds; likewise, for $p < 1/2$, we take $\hat{q}_p(x)$ to be the smallest solution to (9).

Although it is perfectly possible for $\hat{q}_{p_1}$ and $\hat{q}_{p_2}$ to cross one another, the foregoing algorithm ensures that the same behavior does not occur with $\tilde{q}_{p_1}$ and $\tilde{q}_{p_2}$. This is an attraction of $\tilde{q}$ relative to $\hat{q}$. (Again, see Yu 1997 for further details of the computational algorithm that we used.)

This alternative approach, via the conditional distribution function, is attractive but suffers from the disadvantage of having to specify a second bandwidth h_2 as well as the bandwidth h_1 which plays much the same role as the bandwidth h in Section 2. Unsurprisingly, it turns out that the estimates are not very sensitive to the value of h_2, and we provide a rule of thumb for h_2 in Section 3.3. But the choice $h_2 = 0$ is not attractive to us, because it results in a discontinuous conditional quantile estimate (which is inelegant, at least for small samples).

3.2 Mean Squared Error

We need the following conditions here (patterned after Fan 1992, and Fan and Gijbels 1995, in the mean estimation case. These conditions are also a suitable specialization of the requirements of Fan et al. 1994):

1. The necessary partial derivatives of $F(x,y), f(x,y)$ and $g(x)$ exist and are bounded and continuous at both interior and boundary points.
2. $g(x) > 0$ and the conditional density $f(y|x) > 0$ and bounded.
3. The population conditional quantiles $q_p(x)$ are unique.
4. The two bandwidths have the form: $dn^{-\beta}, 0 < \beta < 1$.
5. The kernels W and K are each second-order symmetric.

Also, write

$$F^{ab}(q_p(x)|x) = \frac{\partial^{ab}}{\partial z^a \partial y^b} F(y|z)|_{x,q_p(x)}.$$

We give the following pointwise properties of $\tilde{q}_p(x)$, assuming that x is not a boundary point.

Theorem 1. Under Conditions 1–5, if $h_1 \to 0, h_2 \to 0$, and $nh_1 \to \infty$, then

$$\begin{aligned} \operatorname{MSE}(\tilde{q}_p(x)) &\simeq \frac{1}{4} \{\mu_2(K)h_1^2 F^{20}(q_p(x)|x)/f(q_p(x)|x) \\ &\quad + \mu_2(W)h_2^2 F^{02}(q_p(x)|x)/f(q_p(x)|x)\}^2 \\ &\quad + \frac{R(K)}{nh_1 g(x) f^2(q_p(x)|x)} (p(1-p) - h_2 f(q_p(x)|x)\alpha(W)) \\ &\quad + o(h_1^4 + h_2^4 + h_2/nh_1), \end{aligned}$$

where $\alpha(W) = \int \Omega(t)(1 - \Omega(t))\, dt$.

As for the boundary behavior of the estimator, we offer Theorem 2. Again, and without loss of generality, we consider left boundary points $x = ch_1, 0 < c < 1$.

Theorem 2. Suppose that the conditions of Theorem 1 hold and that $q_p(x)$ is bounded on [0, 1] and right continuous at the point 0. Then the MSE of the estimator $\tilde{q}_p(x)$ at a boundary point is given by

$$\begin{aligned} \operatorname{MSE}(\tilde{q}_p(ch_1)) &\simeq \frac{1}{4} \{\alpha_c(K)h_1^2 F^{20}(q_p(0+)|0+)/f(q_p(0+)|0+) \\ &\quad + \mu_2(W)h_2^2 F^{02}(q_p(0+)|0+)/f(q_p(0+)|0+)\}^2 \\ &\quad + \frac{\beta_c(K)}{nh_1 g(0+) f^2(q_p(0+)|0+)} \\ &\quad \times (p(1-p) - h_2 f(q_p(0+)|0+)\alpha(W)) \\ &\quad + o(h_1^4 + h_2^4 + h_2/nh_1). \end{aligned}$$

Proofs of these theorems (which, unlike the results of Section 2.2 do not appear elsewhere) may be found in the Appendix.

If we choose $h_1 \gg h_2$, then the leading terms in the MSE of $\tilde{q}_p(x)$ are essentially the squared bias and variance terms in Theorem 1 for the check function approach, and they involve only h_1. In fact, these differ from the terms in (5) in only one respect: In the bias term of Theorem 1, the quantity $F^{20}(q_p(x)|x)/f(q_p(x)|x)$, involving the second derivative with respect to x of the conditional distribution function, replaces $q_p''(x)$, the second derivative with respect to x of the quantile function itself. The rest of the MSE expression in Theorem 1 then give us guidance on how to choose h_2. Bandwidth selection for $\tilde{q}_p$ is pursued in the next section.

3.3 Bandwidth Selection

The remarks at the end of Section 3.2 imply that for h_1's choice, we may still use the rules of thumb developed in Section 2.3; all that changes is that $F^{20}(q_p(x)|x)/f(q_p(x)|x)$ replaces $q_p''(x)$ in (6) and (7), and thus the rule-of-thumb arguments remain as valid (or otherwise) for this function as they did for the previous function.

That said, what we have glossed over so far is that one version of optimality, convergence rate–wise, is to choose h_2 as large as possible (subject to $h_2 \to 0$) in Theorem 1 because of the guaranteed nonpositivity of the second variance term, subject to not inflating the bias. This implies $h_2 \sim h_1$ and in particular the choice $h_2 = h_1$. But the result of this is to replace $F^{20}(q_p(x)|x)/f(q_p(x)|x)$ in the optimal h_1 formula with $\{F^{20}(q_p(x)|x) + F^{02}(q_p(x)|x)\}/f(q_p(x)|x)$. With the latter expression involved, we would no longer be so assured of our h_1 rule-of-thumb argument, and we regard this as much more crucial to our implementation.

Therefore, we pursue an alternative rule of thumb for h_2 assuming $h_2 < h_1$. This argument too results in MSE of order $n^{-4/5}$. Intuitively, in any case, the choice of h_2 should not be very critical, because it concerns a smoothing at the distribution function level. Moreover, we have experimented with various formulas for h_2 in practice and have observed little impact, at least for small changes in h_2. But in practice, it still remains to specify a value for h_2 (and to ensure that the value chosen is never so extreme as to make an unnecessary impact). Also, we found from practical computation that h_2 should not be specified too small relative to h_1.

We thus concentrate on the order $h_1^2h_2^2 + h_2/(nh_1)$ terms of $\text{MSE}(\tilde{q}_p(x))$. Although the latter is always negative, the former can be negative or positive at different x's. The optimal h_2 is either $h_2 = \infty$ or

$$h_2 = \left(\frac{R(K)\alpha(W)}{\mu_2(K)\mu_2(W)}\right) \times \frac{f(q_p(x)|x)}{g(x)F^{20}(q_p(x)|x)F^{02}(q_p(x)|x)} \frac{1}{nh_1^3}. \tag{10}$$

We concentrate on the latter (with moduli as necessary) as a proxy for the whole situation (because it remains a sensible value even when suboptimal). Regarding h_1 and h_2 as functions of p, this gives

$$\frac{h_{2,p_1}h_{1,p_1}^3}{h_{2,p_2}h_{1,p_2}^3} = \frac{f(q_{p_1}(x)|x)|F^{20}(q_{p_2}(x)|x)F^{02}(q_{p_2}(x)|x)|}{f(q_{p_2}(x)|x)|F^{20}(q_{p_1}(x)|x)F^{02}(q_{p_1}(x)|x)|}. \tag{11}$$

As when we considered automatic selection of h_1 in Section 2.3, we now suppose that $|F^{20}(q_{p_1}(x)|x)| = |F^{20}(q_{p_2}(x)|x)|$ and make a parametric approximation to $f(q_p(x)|x)$ and $|F^{02}(q_p(x)|x)|$. We take the double-exponential distribution as a guideline rather than the normal this time, because the latter has a zero derivative at its median; that is, take $f(q_p(x)|x) = \lambda\{(1-p)I(p \geq 1/2) + pI(p < 1/2)\}$, $|F^{02}(q_p(x)|x)| = \lambda^2\{(1-p)I(p \geq 1/2) + pI(p < 1/2)\}$. In particular, this yields the especially simple approximation $(h_{2,1/2}h_{1,1/2}^3)^{-1}h_{2,p}h_{1,p}^3 = 1$. Still, our automatic specification is not complete. We choose at this stage to take $h_{2,1/2} = h_{1,1/2}^2$. The main justification for this is to ensure the rate for $h_{2,p}$ is as suggested by (10), which, as $h_{1,p} \to 0$, ensures $h_{2,p} \ll h_{1,p}$. But sometimes in practice $h_{1,p} > 1$, so $h_{2,p}$ would be larger than $h_{1,p}$. We disallow this and, together with a bar on choosing $h_{2,p}$ too small relative to $h_{1,p}$ (these stop extremes of behavior; sensitivity to moderate values of $h_{2,p}$ is not strong), utilize the following rule of thumb for $h_{2,p}$:

$$h_{2,p} = \max\left(\frac{h_{1,1/2}^5}{h_{1,p}^3}, \frac{h_{1,p}}{10}\right) \quad \text{if } h_{1,1/2} < 1 \quad \text{and} \quad \frac{h_{1,1/2}^4}{h_{1,p}^3} \text{ otherwise.} \tag{12}$$

The nonoptimality of our thinking here is evident, but we repeat that all we need is a "reasonable" formula for h_2, which this is, obtained after considerable experimentation with other specifications.

4. PRACTICAL PERFORMANCE

First, we illustrate the methodology on two sets of data taken from the literature. Then we report some simulation results.

Throughout, the standard normal kernel is used as K and the uniform kernel as W (where necessary) in all computations. Two S programs were written for the two methods of iterative computation of quantiles, and two other S programs were used for computation of the local linear fitted regression mean and its derivative.

4.1 Data Examples

The first data example, on triceps skinfold, was described in Section 1. The conditional quantiles corresponding to the check function method are given in Figure 3. The bandwidths chosen by our selection method were $h_{\text{mean}} = 2.5, h_{.5} = 2.73, h_{.75} = h_{.25} = 2.82, h_{.9} = h_{.1} = 3.1$, and $h_{.97} = h_{.03} = 3.6$. These should be compared with the conditional quantiles obtained by the double-kernel method given in Figure 1b. Those quantiles used the foregoing values for h_1, and values of h_2 following from these by using (12).

The second dataset comprises the serum concentration, in grams per liter, of immunoglobulin-G in children age 6 months to 6 years. Royston and Altman (1994) used it with $n = 298$, but the data kindly sent by Dr. Royston contained $n = 300$ points, all of which we use. A scatterplot is given in Figure 4. The data came originally from Isaacs, Alt-

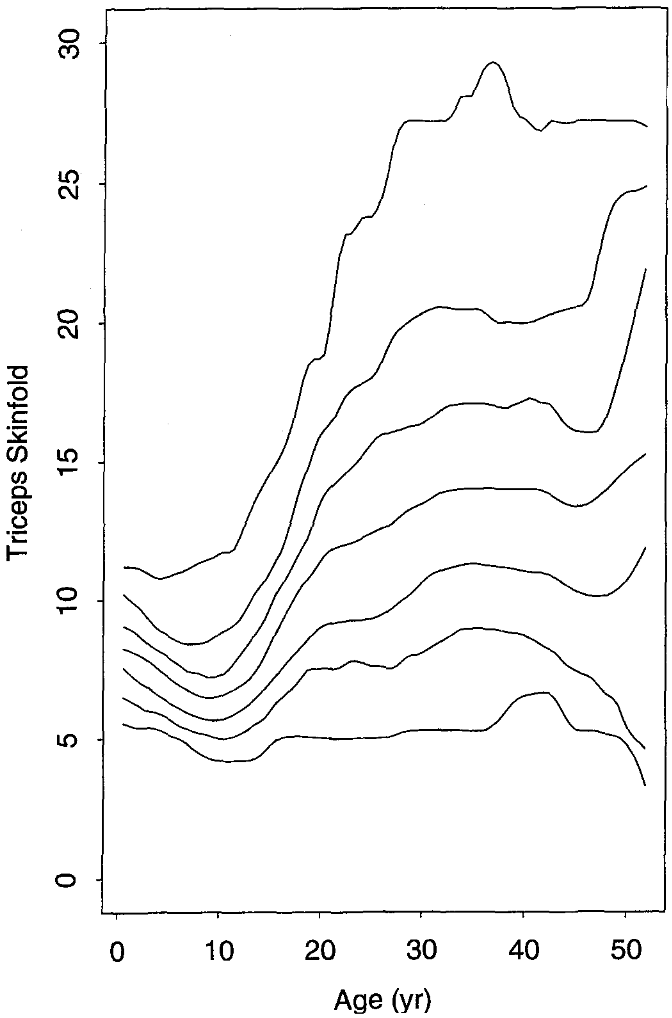

Figure 3. Smoothed 3rd, 10th, 25th, 50th, 75th, 90th, and 97th Quantile Curves for the Triceps Skinfold Data Using Single-Kernel Smoothing.

man, Tidmarsh, Valman, and Webster (1983) who "aimed to establish reference centiles for the serum concentration of certain immunoglobulins in children." Our estimated quantiles, corresponding to the check function and double-kernel approaches with rule-of-thumb bandwidth selection, are given in Figures 5a and 5b. The h or h_1 bandwidths in this case were $h_{\mathrm{mean}} = .5$, $h_{.5} = .54$, $h_{.75} = h_{.25} = .56$, $h_{.9} = h_{.1} = .62$, and $h_{.95} = h_{.05} = .67$.

The first impression to be gained from these figures, and other examples not shown, is that the messages yielded by the check function and double-kernel approaches about the conditional quantiles are broadly similar. That said, there is a distinct tendency for the double-kernel smooths to be smoother than the check function results. It seems that the kernel in the vertical direction is beneficial at least in producing more pleasing pictures.

Comparison of our Figures 1a and 3 with figure 2 of Cole and Green (1992) shows a very considerable degree of similarity in the results of our methods and theirs (which is an interesting semiparametric approach involving penalty functions). Something that we have observed quite generally is a certain "conservative" widening of the extreme quantiles in the double-kernel fit relative to the check function fit where the conditional distribution is "narrow."

The interesting comparison of our Figure 5 with Royston and Altman's (1994) figure 5 (where the regression mean is estimated by parametric methods) is that all our quantiles including the median display a marked peak at the larger ages which is not apparent in Royston and Altman's models. Comparing with the results of Isaacs et al. (1983) also, the main difference lies near the right-hand boundary where their results seem to us a little oversmooth and hence too flat.

In other data examples, not shown, some estimated quantiles have crossed each other under the check function approach but cannot under our double-kernel approach.

4.2 *Some Simulations*

Let Z denote a N(0, 1) random variable and E denote an exponential, mean 1, random variable, each Z or E being independent of the design variable X. Datasets were simulated from each of four models as follows:

1. Almost linear quantiles, heteroscedastic:

$$Y = \sin(.75X) + 1 + .3\sqrt{(\sin(.75X) + 1)}Z$$

$$X \sim \mathrm{N}(0, .0625)$$

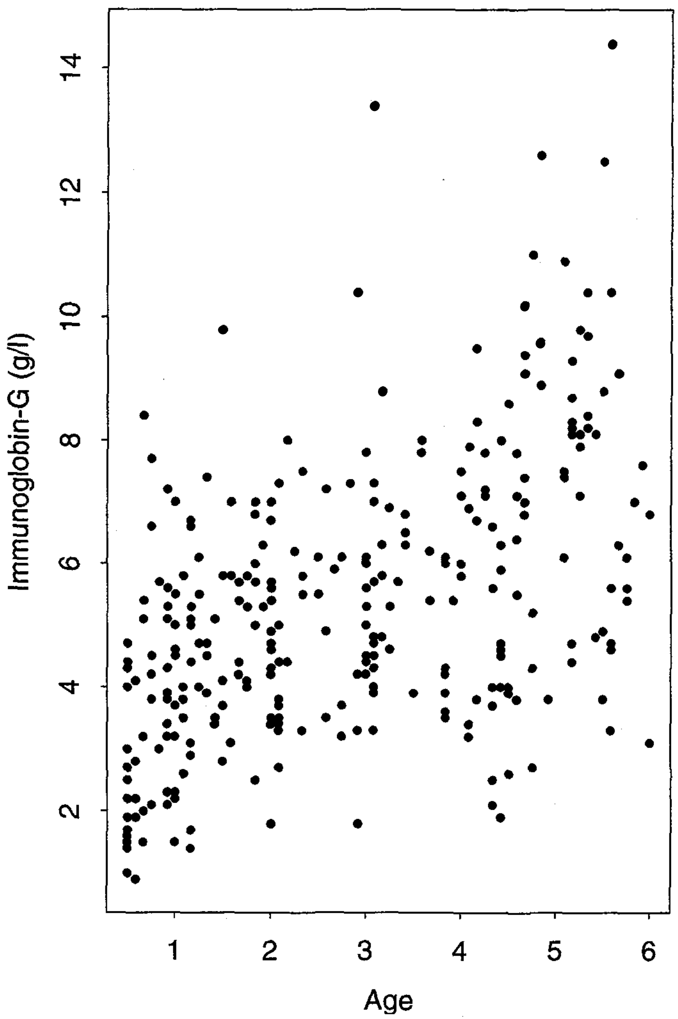

Figure 4. Scatterplot of the Immunoglobulin-G Serum Concentration Data.

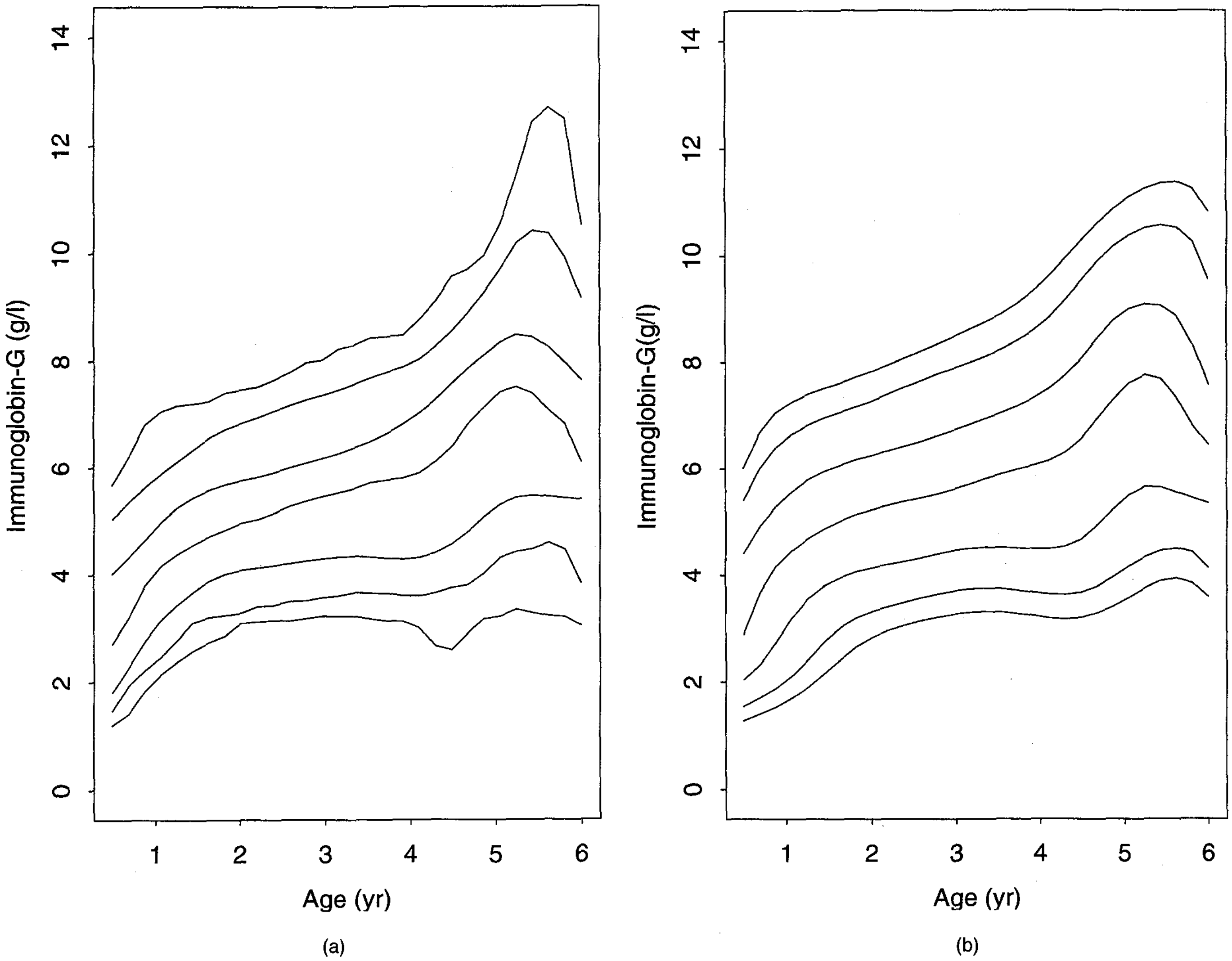

Figure 5. Smoothed 5th, 10th, 25th, 50th, 75th, 90th, and 95th Quantile Curves for the Immunoglobulin-G Data Using (a) Single-Kernel Smoothing and (b) Double-Kernel Smoothing.

2. Smooth "curvy" quantiles, homoscedastic:

$$Y = 2.5 + \sin(2X) + 2\exp(-16X^2) + .5Z \quad X \sim \mathrm{N}(0, 1)$$

3. Simple quantiles, skew distribution:

$$Y = 2 + 2\cos(X) + \exp(-4X^2) + E \qquad X \sim N(0, 1)$$

4. Simple quantiles, heteroscedastic:

$$Y = 2 + X + \exp(-X)(E - \log 2.6) \qquad X \sim U[0, 5].$$

In each case, 100 replications of sample sizes $n = 100$ and $n = 500$ were made. Three quantile values, $p = .1, .5$, and .9 were considered. Integrated squared errors (ISEs) were computed over ranges covering almost all of the data (and averaged over replications); those ranges were (1) $[-.5, .5]$, (2) $[-2, 2]$, (3) $[-2, 2]$, and (4) $[0, 5]$. The check, $\hat{q}_p$, and double, $\tilde{q}_p$, kernel estimators, implemented as discussed earlier, were compared. Results are shown in Table 2.

In most cases, the double kernel approach exhibits a lower ISE than the check function method, sometimes considerably so. On occasions where $\hat{q}_p$ wins, differences are not great. Comparisons with more methods have been made by Yu (1997).

4.3 Conclusion

Here we briefly gather our conclusions with a view to

Table 2. Median ISEs of Check ($\hat{q}_p$) and Double ($\tilde{q}_p$) Kernel Quantile Estimators, Multiplied by 1,000, for Each of Models 1–4 Defined in the Text, and for Three Values of p and Two Values of n as Shown in the Table

		n = 100			n = 500		
		p = .1	p = .5	p = .9	p = .1	p = .5	p = .9
1	$\hat{q}_p$	2.89	1.23	5.35	.92	.79	1.25
	$\tilde{q}_p$	1.87	1.00	4.93	.51	.48	.93
2	$\hat{q}_p$	279.1	267.2	287.8	247.7	191.3	262.6
	$\tilde{q}_p$	259.2	256.0	262.1	217.3	171.3	207.2
3	$\hat{q}_p$	296.5	284.0	306.7	197.7	131.0	216.6
	$\tilde{q}_p$	304.2	276.8	312.6	182.9	127.4	218.7
4	$\hat{q}_p$	72.5	56.0	98.2	60.2	32.8	80.7
	$\tilde{q}_p$	74.9	51.4	81.0	67.1	30.4	67.4

practice. We have shown that local linear conditional quantile estimation is feasible and practical. Results are at the least comparable with those produced by other approaches. As local linear (and polynomial) methods have gained popularity in regression mean estimation, so we feel that local linear methods should gain popularity in conditional quantile estimation, and for much the same reasons: The way they work is interpretable and natural, their asymptotic properties are tractable and informative, computational algorithms are feasible and they work at least as well as other methods, and better near boundaries.

Although the approach is not novel to us, our implementation is. In particular, we have provided rules for the selection of bandwidths that appear to give sensible results. We have been a little surprised by the extent of differences between the two approaches we have developed. In particular, the "vertical" smoothing in the double-kernel approach appears to be advantageous in providing a reasonable extra smoothing over that inherent in the check function approach. It is also advantageous vis-a-vis noncrossing of quantiles. And its performance is the better of the two in our limited simulations. We believe that the double-kernel method in particular is a practically useful method, and one worthy of further fine-tuning for even better performance in the future.

APPENDIX: PROOF OF THEOREMS

First, the following lemma is easy to see from theorem 1 of Fan (1993) and theorem 5 of Fan and Gijbels (1995).

Lemma 1. Let

$$m(x,y) = E\{\Omega(h_2^{-1}(y-Y))|X=x\}$$

and define $\hat{m}_{h_1,h_2}(y|x)$ as the local linear kernel estimator of $m(x,y)$. Then, under the conditions of Theorem 1, as $n \to \infty$,

$$\sqrt{nh_1}\{\hat{m}_{h_1,h_2}(y|x) - m(x,y) - \frac{1}{2}\, h_1^2\mu_2(K)F^{20}(y|x)\} \to \mathrm{N}(0, g^{-1}(x)R(K)\sigma^2(x,y)),$$

where $\sigma^2(x,y) = \mathrm{var}\{\Omega(h_2^{-1}(y-Y))|X=x\}$.

Lemma 2. Under the conditions of Theorem 1,

$$m(x,y) = F(y|x) + \frac{1}{2}\, h_2^2\mu_2(W)F^{0,2}(y|x) + O(h_2^2)$$

and

$$\sigma^2(x,y) = F(y|x)(1-F(y|x)) - h_2F^{01}(y|x)\alpha(W) + O(h_2^2).$$

Under the conditions of Theorem 2,

$$m(0+,y_0) = F(y_0|0+) + \frac{1}{2}\, h_2^2\mu_2(W)F^{0,2}(y_0|0+) + O(h_2^2)$$

and

$$\sigma^2(0+,y_0) = F(y_0|0+)(1-F(y_0|0+)) - h_2F^{01}(y_0|0+)\alpha(W) + O(h_2^2).$$

Proof. These formulas follow by standard Taylor series approximation, as $h_2 \to 0$, allied to integration by parts.

Lemma 3. Let $\hat{f}_{h_1,h_2}(y|x) = \hat{m}^{01}_{h_1,h_2}(y|x)$; in fact, $\hat{f}_{h_1,h_2}(y|x)$ is the local linear kernel estimator of the conditional density discussed by Fan et al. (1996). Also, let $q_p^*(x)$ be some random point between $\tilde{q}_p(x)$ and $q_p(x)$. Then, under the conditions of Theorems 1 and 2, we have

$$\hat{f}_{h_1,h_2}(q_p^*(x)|x) = f(q_p(x)|x) + o_p(1)$$

and

$$\hat{f}_{h_1,h_2}(q_p^*(0+)|0+) = f(q_p(0+)|0+) + o_p(1).$$

Proof. This follows lemma 6 of Samanta (1989) and equations (6.4) and (6.5) of Fan (1993).

Proof of Theorem 1. (Similarly for Theorem 2) From Lemmas 1 and 2, we have

$$\begin{aligned} E\{F_{h_1,h_2}(y|x) - F(y|x)\}^2 &= \left\{\frac{1}{2}\,\{F^{20}(y|x)\}\mu_2(K)h_1^2 + \frac{1}{2}\,\{F^{02}(y|x)\}\mu_2(W)h_2^2\right\}^2 \\ &+ \frac{R(K)}{nh_1f(x)}\,(F(y|x)(1-F(y|x)) - f(y|x)\alpha(W)h_2) \\ &+ o(h_2^2/nh_1 + h_1^4 + h_2^4). \end{aligned}$$

Then Theorem 1 follows from Lemma 3 and the following equation:

$$\tilde{q}_p(x) - q_p(x) \simeq -\frac{\hat{m}_{h_1,h_2}(q_p(x)|x) - F(q_p(x)|x)}{f(q_p^*(x)|x)}.$$

[Received August 1996. Revised August 1997.]

REFERENCES

Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *The Annals of Statistics*, 2, 760–777.

Cleveland, W. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cleveland, W., and Devlin, S. J. (1988), "Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 83, 597–610.

Cleveland, W., and Loader, C. (1996), "Smoothing by Local Regression: Principles and Methods," in *Statistical Theory and Computational Aspects of Smoothing*, eds. W. Härdle and M. G. Schimek, Heidelberg: Physica-Verlag.

Cole, T. J. (1988), "Fitting Smoothed Centile Curves to Reference Data," *Journal of the Royal Statistical Society*, Ser. A, 151, 385–418.

Cole, T. J., and Green, P. J. (1992), "Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood," *Statistics in Medicine*, 11, 1305–1319.

Fan, J. (1992), "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

——— (1993), "Local Linear Regression Smoothing and Their Minimax Efficiencies," *The Annals of Statistics*, 21, 196–216.

Fan, J., and Gijbels, I. (1992), "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20, 2008–2036.

——— (1995), "Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation," *Journal of the Royal Statistical Society*, Ser. B, 57, 371–394.

——— (1996), *Local Polynomial Modelling and its Applications*, London: Chapman and Hall.

Fan, J., Hu, T.-C., and Truong, Y. K. (1994), "Robust Nonparametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433–446.

Fan, J., Yao, Q., and Tong, H. (1996), "Estimation of Conditional Densities and Sensitivity Measures," *Biometrika*, 83, 189–206.

Hastie, T., and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry," *Statistical Science*, 8, 120–143.

Isaacs, D., Altman, D. G., Tidmarsh, C. E., Valman, H. B., and Webster, A. D. B. (1983), "Serum Immunoglobulin Concentrations in Preschool

Children Measured by Laser Nephelometry: Reference Ranges for IgG, IgA, IgM," *Journal of Clinical Pathology*, 36, 1193–1196.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996), "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, 91, 401–407.

Koenker, R., and Bassett, G. S. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., Portnoy, S., and Ng, P. (1992), "Nonparametric Estimation of Conditional Quantile Functions," in L_1*-Statistical Analysis and Related Methods*, ed. Y. Dodge, Amsterdam: Elsevier, 217–229.

Royston, P., and Altman, D. G. (1994), "Regression using Fractional Polynomials of Continuous Covariates: Parimonious Parametric Modelling" (with discussion), *Applied Statistics*, 43, 429–467.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Ruppert, D., and Wand, M. P. (1995), "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 23, 1346–1370.

Samanta, M. (1989), "Non-Parametric Estimation of Conditional Quantiles," *Statistics and Probability Letters*, 7, 407–412.

Stone, C. J. (1977), "Consistent Nonparametric Regression," *The Annals of Statistics*, 4, 595–645.

Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.

Yu, K. (1997), "Smooth Regression Quantile Estimation," unpublished Ph.D. thesis, The Open University.

Yu, K., and Jones, M. C. (1997), "A Comparison of Local Constant and Local Linear Regression Quantile Estimation," *Computational Statistics and Data Analysis*, 25, 159–166.