Statistical Modelling 2007; **7**(4): 377–389

Improved double kernel local linear quantile regression

M C Jones[1] and Keming Yu[2]
[1]Department of Statistics, The Open University, UK
[2]Department of Mathematical Sciences, Brunel University, UK

Abstract: As sample quantiles can be obtained as maximum likelihood estimates of location parameters in suitable asymmetric Laplace distributions, so kernel estimates of quantiles can be obtained as maximum likelihood estimates of location parameters in a general class of distributions with simple exponential tails. In this paper, this observation is applied to kernel quantile regression. In doing so, a new double kernel local linear quantile regression estimator is obtained which proves to be consistently superior in performance to the earlier double kernel local linear quantile regression estimator proposed by the authors. It is also straightforward to compute and more readily affords a first derivative estimate. An alternative method of selection for one of the two bandwidths involved also arises naturally but proves not to be so consistently successful.

Key words: asymmetric Laplace distribution; bandwidth selection; exponential tails; maximum likelihood

Received August 2006; **revised** April 2007; **accepted** April 2007

1 Introduction

The topic of quantile regression (Koenker, 2005; Koenker and Bassett, 1978) in which various quantiles of the distribution of a response variable—rather than its mean—are modelled as functions of values of covariates, is by now a well-established technique with a burgeoning theoretical and practical literature. This paper is a contribution to nonparametric quantile regression in which the dependence of quantiles of the response distribution on a single covariate is investigated though kernel-weighted local linear methods (for example, Fan and Gijbels, 1996; Loader, 1999). (The single covariate case is considered both for convenience—extension to multiple covariates could be done in a standard way—and because it tends to be of more widespread importance when quantiles are of interest as, for example, in reference and growth charts with dependence on age.) A currently successful approach is our own double kernel local linear quantile regression methodology (Yu and Jones, 1998). In this

Address for correspondence: Keming Yu, Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK. E-mail: keming.yu@brunel.ac.uk

 10.1177/1471082X0700700407

paper, we offer a new formulation of this approach which gives rise to its consistently superior modification which we, therefore, recommend for practical use.

It has long been recognised that maximum likelihood estimation in the asymmetric Laplace distribution (for example, Kotz *et al.*, 2001, Chapter 3) with skewness parameter fixed is equivalent to nonparametric estimation of a particular quantile of the underlying distribution. This has been exploited in a quantile regression context by Koenker and Machado (1999), and Yu and Moyeed (2001). Jones (2007a) introduced a family of distributions with simple exponentially decaying tails which includes the asymmetric Laplace distribution and, for example, the log F distribution as special cases. In that paper, it was noted that maximum likelihood estimation in any member of the new family of distributions is equivalent to kernel estimation of a particular quantile. In this paper, we extend that insight to kernel-weighted local linear quantile regression and find that the new formulation gives rise to the consistently advantageous modification to the double kernel local linear quantile regression method mentioned above. In addition, the new approach affords an alternative method for the selection of one of the two bandwidths involved, which works tolerably well.

The outline of the paper is as follows. The ideas of Jones (2007a) are reviewed in Section 2. See Section 3 for details of the new double kernel local linear quantile regression method which include: a description of the method and its predecessor (Section 3.1); a note on the simple computation of the new estimator (Section 3.2); reference to its asymptotic theory (Section 3.3); the new approach to bandwidth selection (Section 3.4); simulation evidence (Section 3.5); and remarks on quantile crossing (Section 3.6). Two examples are considered in Section 4. Our recommendation and some further brief comments close the paper in Section 5.

2 The family of distributions

The four-parameter version of the general family of distributions with simple exponential tails proposed by Jones (2007a) has density $\sigma^{-1} f_G\{\sigma^{-1}(x-\mu)\}$ where

$$f_G(x) = \mathcal{K}_G^{-1}(\alpha, \beta) \exp\{\alpha x - (\alpha+\beta) G^{[2]}(x)\}. \tag{2.1}$$

Here, $\mu \in \mathcal{R}$, σ, α, $\beta > 0$ and $K_G(\cdot,\cdot)$ is the normalisation constant. $G^{[2]}$ is the first iterated (left-tail) distribution function defined by $G^{[2]}(x) = \int_{-\infty}^{x} G(t)dt = E\{(x - X_G)I(X_G < x)\}$ (for example, Bassan *et al.*, 1999) associated with what is conveniently taken to be a simple symmetric distribution G with density g, associated random variable X_G and no further unspecified parameters. (The only restriction on g is that its tails must not be so extremely heavy that $g(x) \sim |x|^{-(\gamma+1)}$ for $0 < \gamma \le 1$ as $|x| \to \infty$.) Then, regardless of the choice of admissible G,

$$f_G(x) \sim e^{\alpha x} \text{ as } x \to -\infty, \qquad f_G(x) \sim e^{-\beta x} \text{ as } x \to \infty.$$

If X_G is the constant zero, the asymmetric Laplace distribution ensues. If G is the logistic distribution, the log F distribution (for example, Jones, 2007b) ensues. If G is the uniform distribution, a distribution with a normal body and exponential tails is produced, which is the density associated with the classical Huber M-estimator of location (Huber, 1964) when $\alpha = \beta$. For many other special cases, including the hyperbolic distribution (for example, Barndorff-Nielsen and Blaesild, 1983), see Jones (2007a).

The asymmetric Laplace distribution is, however, a three-parameter distribution because only two out of σ, α and β are identifiable. Jones (2007a, 2007b) argues that, in practical terms, three parameters is the appropriate number for the log F distribution and other members of family (2.1) also. This is because σ, α and β are overall left-tail and right-tail scale parameters, respectively, and only two of these are practically necessary to describe the scale and skewness of the distribution. Of the various options available to us to remove one parameter from f_G, we choose to set $\alpha + \beta = 1$, and therefore work with $\sigma^{-1} f_G\{\sigma^{-1}(x-\mu)\}$ where

$$f_G(x) = \mathcal{K}_G^{-1}(1-\beta, \beta)\exp\{(1-\beta)x - G^{[2]}(x)\}, \tag{2.2}$$

$0 < \beta < 1$. The versions of the asymmetric Laplace and three-parameter log F and uniform-based distributions that are of interest, therefore, have densities

$$f_{AL}(x) = \beta(1-\beta)\begin{cases} \exp\{(1-\beta)x\} & \text{if } x \le 0, \\ \exp(-\beta x) & \text{if } x \ge 0, \end{cases}$$

$$f_{LF}(x) = \frac{\sin(\pi\beta)}{\pi}\frac{e^{(1-\beta)x}}{(1+e^x)}$$

and, with $\mathcal{K}_U$ obtainable from Jones (2007a),

$$f_U(x) = \mathcal{K}_U^{-1}(1-\beta,\beta)\begin{cases} \exp\{(1-\beta)x\} & \text{if } x < -1, \\ \exp\left\{-\beta(1-\beta) - \frac{1}{4}(x+2\beta-1)^2\right\} & \text{if } -1 \le x < 1, \\ \exp(-\beta x) & \text{if } x \ge 1, \end{cases}$$

respectively. Note that the version of the log F distribution written here is the natural exponential family generalized hyperbolic secant (NEF-GHS) distribution of Morris (1982).

Consider maximum likelihood estimation of the parameters μ and σ based on a random sample $X_1, ..., X_n$ from $\sigma^{-1} f_G\{\sigma^{-1}(x-\mu)\}$ with f_G given by (2.2) and β fixed. Then, the score equation corresponding to differentiation of the log-likelihood with respect to μ is

$$\frac{1}{\sigma}\left\{-n(1-\beta) + \sum_{i=1}^{n} G\left(\frac{X_i-\mu}{\sigma}\right)\right\} = 0$$

which yields

$$\frac{1}{n}\sum_{i=1}^{n} G\left(\frac{\mu - X_i}{\sigma}\right) = \beta. \tag{2.3}$$

But this shows that, for given σ, the value of μ satisfying (2.3) is nothing other than the standard inversion kernel quantile estimator at β (Azzalini, 1981; Nadaraya, 1964). This is because the left-hand side of (2.3) is the kernel estimator of the distribution function at the point μ with bandwidth σ and kernel distribution function G. (Maximum likelihood estimation of the location parameter of the NEF-GHS ($\log F$) distribution, for example, with β fixed is therefore kernel quantile estimation using the logistic density as kernel.) Also, the score equation corresponding to differentiation of the log-likelihood with respect to σ gives a method for selection of the bandwidth σ, although the quality of performance of this method as reported by Jones (2007a) is somewhat mixed.

3 The improved double kernel local linear quantile regression estimator

3.1 The method and its predecessor

When interested in the regression quantile associated with probability β, we assume the data (X_i, Y_i), $i = 1, \ldots, n$, have conditional distribution $Y_i|X_i \sim \sigma^{-1} f_G\left[\sigma^{-1}\{Y_i - \mu(X_i)\}\right]$ where f_G is given by (2.2), β is fixed and $\mu(x)$ is an unspecified smooth function of x which is the focus of estimation. Initially, fix σ as well. We now follow the usual kernel-weighted local (linear) likelihood paradigm (for example, Loader, 1999) and proceed by localising the log-likelihood ℓ associated with (2.2). This is done by introducing a further kernel (density) K and bandwidth h_1 and taking $\mu = \mu(x)$ to have (locally) linear form. It is

$$\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}\right) \ell(Y_i; \mu(X_i), \sigma) = \text{constant} + \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_1}\right)\Bigg[-\log\sigma + (1-\beta)\frac{\{Y_i - \mu - \mu_1(X_i - x)\}}{\sigma} - G^{[2]}\left\{\frac{Y_i - \mu - \mu_1(X_i - x)}{\sigma}\right\}\Bigg]. \tag{3.1}$$

We then maximise over μ and μ_1 for each x. Differentiating with respect to μ and μ_1 is the source of the following pair of estimating equations:

$$\beta = \frac{1}{\sum_{i=1}^{n} v_i^{(k)}(x)} \sum_{i=1}^{n} v_i^{(k)}(x) G\left\{\frac{\mu + \mu_1(X_i - x) - Y_i}{\sigma}\right\}, \qquad k = 0, 1 \tag{3.2}$$

where

$$v_i^{(k)}(x) = (X_i - x)^k K\left(\frac{x - X_i}{h_1}\right).$$

Note that the obtained value of $\mu_1 = \mu_1(x)$ is an estimate of the derivative of $\mu(x)$.

The above, new method involves two kernels (and two bandwidths, h_1 and σ) and is therefore a double kernel local linear quantile regression estimator in the same sense as the method of that name in Yu and Jones (1998). The method in Yu and Jones (1998)—in current notation—estimated the distribution function by kernel-weighted local least squares fitting of a linear function to $G\{\sigma^{-1}(\mu - Y_i)\}$, $i = 1, \ldots, n$. This resulted in the estimating equation

$$\beta = \frac{1}{\sum_{i=1}^n w_i(x)} \sum_{i=1}^n w_i(x) G\left(\frac{\mu - Y_i}{\sigma}\right) \tag{3.3}$$

where

$$w_i(x) = v_i^{(0)}(x) S_2(x) - v_i^{(1)}(x) S_1(x) \quad \text{and} \quad S_k(x) = \sum_{i=1}^n v_i^{(k)}(x).$$

The new approach appears to give a more principled formulation to double kernel local linear quantile regression than did the old one, which specified the kernel G and its interaction with the local linear smoothing in a more ad hoc manner. The two approaches would have yielded the same estimator had we considered locally constant rather than linear estimators, namely the solution for μ to

$$\beta = \frac{1}{\sum_{i=1}^n v_i^{(0)}(x)} \sum_{i=1}^n v_i^{(0)}(x) G\left(\frac{\mu - Y_i}{\sigma}\right).$$

3.2 Computation

Nice features of the pair of equations (3.2) are that (a) when $k = 0$, the right-hand side of the corresponding equation is a monotone increasing function of μ (for given μ_1 and σ) which takes all values between 0 and 1; (b) when $k = 1$, rewrite the corresponding equation as

$$\beta S_1(x) = \sum_{i=1}^n v_i^{(1)}(x) G\left\{\frac{\mu + \mu_1(X_i - x) - Y_i}{\sigma}\right\}. \tag{3.4}$$

The right-hand side of (3.4), (*) say, is a monotone increasing function of μ_1 (for given μ and σ). Write $S_1 = S_1^+ + S_1^-$ where S_1^+ is the sum of its positive contributions and S_1^- is the sum of its negative contributions. Then, as $\mu_1 \to -\infty$, $(*) \to S_1^-$, and as $\mu_1 \to \infty$, $(*) \to S_1^+$. When $S_1 = 0$, the limits are < 0 and > 0, respectively. When $S_1 > 0$, $S_1^- \leq 0$ and $S_1^+ \geq S_1$. When $S_1 < 0$, $S_1^- \leq S_1$ and $S_1^+ > 0$. In each case, all possible values of the left-hand side of (3.4) can be taken by the right-hand side. These observations make for simple implementation and computation that is speedy, though of the same order as that of Yu and Jones (1998). Note, however, that with the current method an estimate of the first derivative comes at no extra computational cost.

3.3 Theory

Double kernel local linear methods (3.2) and (3.3) are sufficiently similar to afford precisely the same asymptotic bias, variance and hence mean squared error when theory is performed in the standard kernel smoothing manner, assuming that h_1 and σ tend to zero at some appropriate speed as $n \to \infty$. See the technical report version of this paper (Jones and Yu, 2006) for details.

3.4 Bandwidth selection

Inter alia, in Yu and Jones (1998) the main theoretical result was used to provide rule-of-thumb bandwidth selectors for h_1 and σ. The former is based on an explicit approximate relation between the value of h_1 appropriate to estimating the βth quantile and the value appropriate to estimating the conditional mean function; any existing bandwidth selection methodology for ordinary local linear regression can then be used, for which we took that of Ruppert *et al.* (1995). Estimation of σ was based on numerous simplifying assumptions which, interestingly, involved the Laplace distribution. This rule-of-thumb for σ will be replaced in the next paragraph, although we note, as before, that precise choice of σ is not as crucial as that of h_1.

An alternative method for selecting the value of σ is available by maximising the localised log-likelihood (3.1) over σ as well as μ and μ_1. Note that this amounts to local constant fitting of the scale parameter and hence, in bandwidth selection terms, to a local bandwidth selection method. (Our work in this paragraph says nothing about choosing h_1, so we will combine the new method for selecting σ with the existing method for h_1.) The relevant estimating equation is

$$0 = \sum_{i=1}^{n} v_i^{(0)}(x) \left(\{Y_i - \mu - \mu_1(X_i - x)\} \left[\beta - G\left\{ \frac{\mu + \mu_1(X_i - x) - Y_i}{\sigma} \right\} \right] - \sigma \right). \tag{3.5}$$

Again, this proves to be nice and simple to solve computationally. For given μ and μ_1, write $W_i = \mu + \mu_1(X_i - x) - Y_i$, $i = 1, ..., n$. The derivative of the right-hand side of (3.5) with respect to σ is

$$-\sum_{i=1}^{n} v_i^{(0)}(x) \left\{ \frac{W_i^2}{\sigma^2} g\left(\frac{W_i}{\sigma}\right) + 1 \right\} < 0.$$

Moreover, the left-hand side of (3.5) tends to

$$(1-\beta)\sum_{i=1}^{n} v_i^{(0)}(x) W_i I(W_i \geq 0) - \beta \sum_{i=1}^{n} v_i^{(0)}(x) W_i I(W_i < 0) > 0$$

as $\sigma \to 0$, and to $-\infty$ as $\sigma \to \infty$. So, again, the solution can be found by simple methods.

3.5 Simulation evidence

We repeat the simulation study reported in Yu and Jones (1998) to compare the new estimators with our earlier version of the double kernel local linear estimator (3.3), which we shall now denote $\tilde{q}_\beta$; the version of the new double kernel local linear estimator (3.2) using the 'old' bandwidth selection methodology will be denoted $\hat{q}_\beta^0$ and that using the different value for σ described in Section 3.4, $\hat{q}_\beta^1$. We used the normal kernel as K and the uniform kernel as G throughout, to match with the choices made in Yu and Jones (1998).

The simulation setup of Yu and Jones (1998) involved four models, three quantiles ($\beta = 0.1, 0.5, 0.9$) and two sample sizes ($n = 100, 500$). We follow our earlier terminology by referring to the four models, Models 1–4, as '1. Almost linear quantiles, heteroscedastic', '2. Smooth "curvy" quantiles, homoscedastic', '3. Simple quantiles, skew distribution' and '4. Simple quantiles, heteroscedastic', respectively (formulae in Yu and Jones, 1998: 234–35). One hundred replications were made, integrated squared errors (ISEs) were computed over suitable ranges and median ISEs reported in Table 1.

Comparing $\hat{q}_\beta$, (3.2), with $\tilde{q}_\beta$, (3.3), shows a performance consistently as good and sometimes considerably better for the former. We are, therefore, happy to recommend $\hat{q}_\beta$ as an improved replacement for the double kernel local linear method of Yu and Jones (1998). Note that the predecessor double kernel local linear method in turn was claimed in our earlier paper to outperfom the single kernel local linear check function approach to kernel quantile regression. The comparison of

Table 1 Median ISEs, multiplied by 1000, for Yu and Jones (1998) double kernel local linear quantile estimator ($\tilde{q}_\beta$) and new double kernel local linear quantile estimator (3.2) with Yu and Jones bandwidth selection ($\hat{q}_\beta^0$) and bandwidth selection of Section 3.4 ($\hat{q}_\beta^1$)

		$n = 100$			$n = 500$		
		$\beta = 0.1$	$\beta = 0.5$	$\beta = 0.9$	$\beta = 0.1$	$\beta = 0.5$	$\beta = 0.9$
	$\tilde{q}_\beta$	1.86	1.01	4.90	0.51	0.48	0.93
1	$\hat{q}_\beta^0$	1.79	1.02	3.48	0.49	0.48	0.82
	$\hat{q}_\beta^1$	1.72	1.04	2.34	0.41	0.50	0.79
	$\tilde{q}_\beta$	259.4	256.0	262.1	217.5	172.3	207.2
2	$\hat{q}_\beta^0$	203.1	195.9	205.6	205.8	156.3	194.9
	$\hat{q}_\beta^1$	198.9	206.3	202.0	200.9	170.6	194.2
	$\tilde{q}_\beta$	304.1	276.3	311.9	182.9	127.2	218.9
3	$\hat{q}_\beta^0$	219.6	212.5	238.9	149.3	106.8	196.4
	$\hat{q}_\beta^1$	262.4	207.5	256.6	158.0	99.4	209.1
	$\tilde{q}_\beta$	72.6	51.6	79.7	60.0	30.6	67.6
4	$\hat{q}_\beta^0$	72.1	51.5	72.5	60.0	30.4	65.4
	$\hat{q}_\beta^1$	72.0	52.6	70.6	60.0	31.5	63.1

Note: The median ISEs are calculated for each of Models 1 to 4, three values of β and two values of n as shown in the table

bandwidth selection methods between $\hat{q}_\beta^0$ and $\hat{q}_\beta^1$ is far less clear cut. For Models 1 and, particularly, 4, results are much the same. For Model 2, the new bandwidth selector wins for the more extreme quantiles but loses at the median; for Model 3, the opposite effect is observed. For all these models, either version of $\hat{q}_\beta$ continues to display a (sometimes considerable) improvement over $\tilde{q}_\beta$.

3.6 Quantile crossing

There seems to be no guarantee that two quantile curves produced by method (3.2) cannot cross; yet such an eventuality seems to be rare in practice and we have never yet observed its occurrence. To address the issue theoretically, define

$$S_k^g(x) = \sum_{i=1}^{n} v_i^{(k)}(x) g\left\{\frac{a + b(X_i - x) - Y_i}{\sigma}\right\}.$$

Then, differentiation with respect to β in (3.2) leads to the following neat representation for $a'_\beta(x) \equiv \partial a(x)/\partial\beta$:

$$a'_\beta(x) = \sigma \left\{ \frac{S_0(x)S_2^g(x) - S_1(x)S_1^g(x)}{S_0^g(x)S_2^g(x) - (S_1^g)^2(x)} \right\}.$$

The denominator of this expression is positive by the Cauchy-Schwartz inequality. Given that $S_0(x)S_2(x) - (S_1)^2(x)$ is also positive, the similar form of the numerator to both these other expressions gives hope that it is *usually* positive too. But that seems to be as much as can be said in general. We do have that $a'_\beta(x) > 0$ (i) when $S_1(x)S_1^g(x) < 0$; (ii) for h_1 sufficiently small, since then $S_1(x)S_1^g(x) = o(S_0(x)S_2^g(x))$; and (iii) for σ sufficiently large, since then $S_k^g(x) \sim S_k(x)$.

There is, however, a fallback in the unlikely event of quantile crossing rearing its head in a substantial way. The same modification can be applied to our algorithm as was employed at (9) of Yu and Jones (1998) to avoid quantile crossing. Given $\hat{\mu}_1(x)$, consider inversion of the $k = 0$ version of (3.2) to be of the form $\hat{q}_\beta(x) = \hat{F}_{h_1,\sigma}(\beta|x)$, then (i) select the median (at x) to be any value satisfying this equation for $\beta = 1/2$; (ii) for $\beta > 1/2$, take $\hat{q}_\beta(x)$ to be the largest value satisfying this equation; and (iii) for $\beta < 1/2$, take $\hat{q}_\beta(x)$ to be the smallest value satisfying the equation.

4 Examples

4.1 Example I

We first apply the methods of this paper to the triceps skinfold data of Cole and Green (1992). These data comprise 892 measurements of females up to age 50 years in Gambia in 1989. Following Cole and Green, we estimate seven quantiles, those corresponding to $\beta = 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97$, as functions of the explanatory variable age. The results of applying the new method (3.2) with the Yu and Jones (1998) bandwidth selection methodology are in Figure 1(a), and those associated with the new method together with the new bandwidth methodology (3.4) in Figure 1(b). The general level of triceps skinfold decreases a little up to age 10 years and there is then a rapid increase to a higher level in adult life. The major impression, however, is of heteroscedasticity: a small variability in girls is replaced by a much higher variability in women with indications of yet greater variability in the oldest participants in this study.

Comparable pictures have been produced for these data by the LMS method in Cole and Green (1992: Figure 2) employing two subjectively chosen levels of smoothing, and in Yu and Jones (1998: Figure 1(b)) employing our earlier double

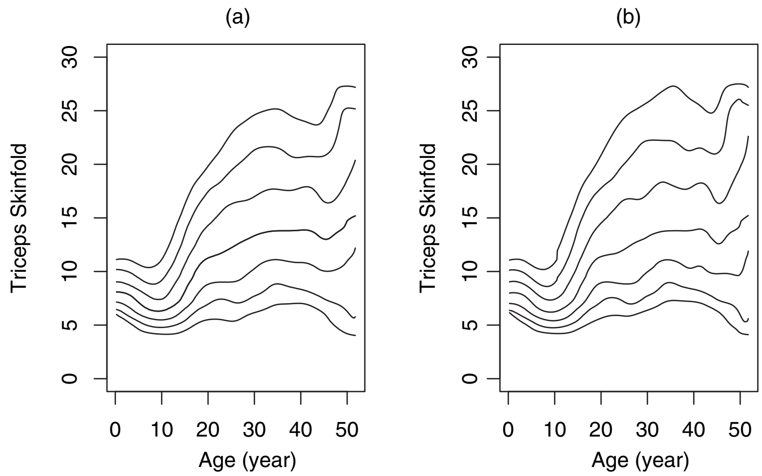

Figure 1 Smoothed $0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97$ quantiles for the triceps skinfold data using the new double kernel local linear method (3.2): (a) with bandwidths selected as in Yu and Jones (1998); (b) with bandwidth σ selected using (3.4)

kernel local linear method with automatic bandwidth choice. The results produced by old and new double kernel local linear methods with the same bandwidth selection method are very similar in this instance. These in turn are broadly comparable to the less smooth of the two sets of results in Cole and Green's (1992) Figure 2, although with differences at a detailed level. The new bandwidth selection method has yielded a picture (Figure 1(b)) with rather less smoothness. In this particular instance, we are inclined to prefer the results of the old bandwidth selection methodology to that of the new.

4.2 Example II

We now consider estimating the same set of conditional quantiles for the second dataset investigated by Cole and Green (1992) which consists of 'body weight in 4011 US girls aged between 1 and 21 years, obtained as part of the American HANES1 Health and Nutrition Survey'. In Figure 2(a), we display the result of applying the Yu and Jones (1998) methodology, while in Figure 2(b), we display the estimated quantiles obtained by the methodology of this paper. In this case, the old estimates are undersmoothed relative to the new, the latter being very similar to results (not shown) obtained by (3.2) with the old bandwidth methodology and pretty similar to Figure 5 of Cole and Green (1992). Here, the new methodology appears to give rather better results than the old.

In this scenario, it may also be of interest to consider estimates of the derivatives (with respect to x) of the conditional quantiles. As mentioned in Section 3.1, such estimates 'come for free' with the new methodology, provided we are happy to utilise the same bandwidth values, as we are in this case. The results are shown in Figure 3. Clearly, the growth velocities for all quantiles gradually increase up to age around 12, then decrease. Furthermore, Figure 3 makes it particularly clear that after age 12, the heavier girls' growth velocity decreases much faster than that of the lighter girls.

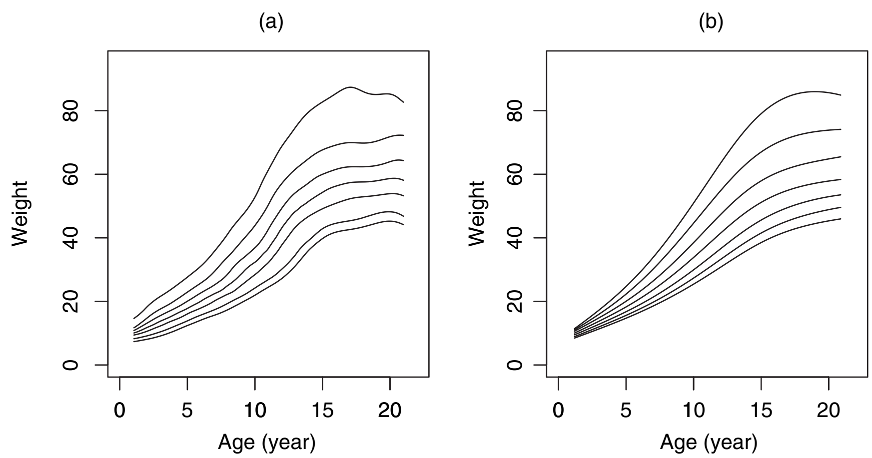

Figure 2 Smoothed 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97 quantiles for the US girls' weight data: (a) using the methodology of Yu and Jones (1998); (b) using the new double kernel local linear method (3.2) with bandwidth σ selected using (3.4)

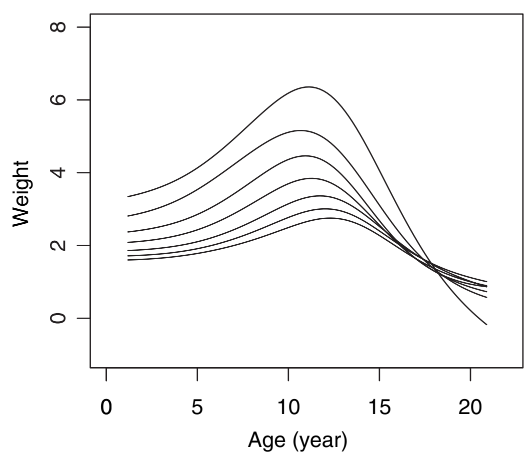

Figure 3 Estimates of the first derivatives of, reading from the bottom curve towards the left of the frame, the 0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97 quantiles for the US girls' weight data using the new double kernel local linear methodology

5 Conclusions

The main practical recommendation of this paper is that the successful double kernel local linear quantile regression method of Yu and Jones (1998) (given at (3.3) earlier) be replaced by the consistently, if sometimes slightly, superior modified method that we have introduced in this paper (specifically given at (3.2)); the new version is also particularly simple to implement computationally. Note that we are claiming finite sample performance superiority of the new version of the method which is not reflected in the (first order) asymptotic equivalence of the two versions.

Our new approach to selection of one of the two bandwidths involved in the new double kernel local linear method is less consistently successful. This is probably because, in common with approaches involving the asymmetric Laplace distribution, the parametric distribution is used not as a 'serious' model for the distribution of the response variable—after all, a different distribution is fitted for each quantile estimated—but just as a convenient vehicle for (essentially) nonparametric estimation of quantile regression functions. There may well remain scope for improved bandwidth selection relative both to the method of Yu and Jones (1998), and to the method of this paper. For example, it can be argued that as (3.4) involves $v_i^{(1)}(x)$ rather than $v_i^{(0)}(x)$, a larger value of h_1 might be used in (3.4) to obtain $\hat{\sigma}$; initial exploration of this idea seems quite encouraging.

Finally, we note again that, in contrast to our earlier approach, the new approach also affords immediate estimation of the first derivative of the regression quantile (although one may also wish to use a larger value of h_1 if this is the principal estimation problem). Derivatives of regression quantiles are also important in, for example, measuring the return of education in labour economics or of investment in finance. For earlier work on local polynomial quantile derivative estimation, see Chaudhuri *et al*. (1997).

Acknowledgement

We wish to acknowledge the helpful and encouraging discussion of this work by participants at the ICMS Workshop on Quantile Regression, LMS Method and Robust Statistics in the 21st Century, held in Edinburgh in June 2006, and the comments of the referees.

References

Azzalini A (1981) A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–28.

Barndorff-Nielsen O and Blaesild P (1983) Hyperbolic distributions. In Johnson NL, Kotz S and Read CB eds, *Encyclopedia* *of Statistical Sciences*. New York: Wiley, 700–07.

Bassan B, Denuit M and Scarsini M (1999) Variability orders and mean differences. *Statistics and Probability Letters*, **45**, 121–30.

Chaudhuri P, Doksum K and Samarov A (1997) On average derivative quantile regression. *Annals of Statistics*, **25**, 715–44.

Cole TJ and Green PJ (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–19.

Fan J and Gijbels I (1996) *Local polynomial modelling and its applications*. London: Chapman and Hall.

Huber PJ (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

Jones MC and Yu K (2006) *Improved double kernel local linear quantile regression*. Open University Department of Statistics Technical Report 06/06; see http://statistics.open.ac.uk/Technical Reports/Technical ReportsIntro.htm

Jones MC (2007a) On a class of distributions with simple exponential tails. *Statistica Sinica*, forthcoming.

Jones MC (2007b) The logistic and the log *F* distribution. In Balakrishnan N ed., *Handbook of the Logistic Distribution*, Second Edition. Dekker, forthcoming.

Koenker R (2005) *Quantile regression*. Cambridge: Cambridge University Press.

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

Koenker R and Machado JAF (1999) Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 1296–310.

Kotz S, Kozubowski TJ and Podgórski K (2001) *The Laplace distribution and generalizations; a revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhauser.

Loader C (1999) *Local regression and likelihood*. New York: Springer.

Morris CN (1982) Natural exponential families with quadratic variance functions. *Annals of Statistics*, **10**, 65–80.

Nadaraya EA (1964) Some new estimates for distribution functions. *Theory of Probability and its Applications*, **15**, 497–500.

Ruppert D, Sheather SJ and Wand MP (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257–70.

Yu K and Jones MC (1998) Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–37.

Yu K and Moyeed RA (2001) Bayesian quantile regression. *Statistics and Probability Letters*, **54**, 437–47.