

Improved double kernel local linear quantile regression

M.C. Jones

The Open University, UK

and Keming Yu

Brunel University, UK

Summary. 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 so doing, 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 particularly straightforward to compute. An alternative method of selection for one of the two bandwidths involved also arises naturally but proves not to be so consistently successful.

Keywords: Asymmetric Laplace distribution; Bandwidth selection; Exponential tails; Maximum likelihood

Address for correspondence: M.C. Jones, Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.
E-mail: `m.c.jones@open.ac.uk`

1. Introduction

The topic of *quantile regression* (Koenker and Bassett, 1978, Koenker, 2005), 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 through kernel-weighted local linear methods (e.g. Fan and Gijbels, 1995, 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 paper, we offer a new formulation of this approach which gives rise to a consistently superior modification of it which we therefore recommend for practical use.

It has long been recognised that maximum likelihood estimation in the asymmetric Laplace distribution (e.g. 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 (2006a) 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 is 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 (2006a) 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); asymptotic theory (Section 3.3); the new approach to bandwidth selection (Section 3.4); simulation evidence (Section 3.5); an example (Section 3.6); and remarks on quantile crossing (Section 3.7). Our recommendation and some further brief comments close the paper in Section 4.

2. The family of distributions

The four-parameter version of the general family of distributions with simple exponential tails proposed by Jones (2006a) has density $\sigma^{-1}f_G\{\sigma^{-1}(x - \mu)\}$ where

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

Here, $\mu \in \mathcal{R}$, $\sigma, \alpha, \beta > 0$ and $\mathcal{K}_G(\cdot, \cdot)$ is the normalisation constant. $G^{[2]}$ is the first iterated (left-tail) distribution function defined by $G^{[2]}(x) = \int_{-\infty}^x G(t)dt = E\{(x - X_G)I(X_G < x)\}$ (e.g. 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 $0 < \gamma \leq 1$ as $|x| \rightarrow \infty$.) Then, regardless of the choice of admissible G ,

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

If X_G is the constant zero, the asymmetric Laplace distribution ensues. If G is the logistic distribution, the log F distribution (e.g. Jones, 2006b) 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 (e.g. Barndorff-Nielsen and Blaesild, 1983), see Jones (2006a).

The asymmetric Laplace distribution is, however, a three-parameter distribution because only two out of σ , α and β are identifiable. Jones (2006a,b) 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

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

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

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

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

and, with \mathcal{K}_U obtainable from Jones (2006a),

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

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, \dots, 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

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

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 (Nadaraya, 1964, Azzalini, 1981). 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 (2006a) 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) , $i = 1, \dots, n$, have conditional distribution $Y_i|X_i \sim \sigma^{-1}f_G[\sigma^{-1}\{Y_i - \mu(X_i)\}]$ 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 (e.g. 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

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

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:

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

where

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

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 $G\{\sigma^{-1}(\mu - Y_i)\}$, $i = 1, \dots, n$. This resulted in the estimating equation

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

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 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

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

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 $\mu_1 \rightarrow -\infty$, (*) $\rightarrow S_1^-$ and as $\mu_1 \rightarrow \infty$, (*) $\rightarrow S_1^+$. 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 and speedy computation.

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 $n \rightarrow \infty$. For theoretical purposes we suppose the support of the X -marginal density $f_X(x)$ to be $(0, 1)$ and assume x to be an ‘interior point’ unaffected by boundary considerations (we do not bother with the analogous result for boundary points here). The main theoretical result is Theorem 1 to follow; its proof is outlined in the Appendix. Its conditions are the same as those in Yu and Jones (1998):

- (1) The necessary partial derivatives of the joint density function $f(x, y)$, the joint distribution function $F(x, y)$ and $f_X(x)$ exist and are bounded and continuous;
- (2) $f_X(x) > 0$ and the conditional density $f(y|x) > 0$ and bounded;
- (3) The population conditional quantiles $q_\beta(x)$ are unique;
- (4) The two bandwidths h_1 and σ have the form $dn^{-\gamma}$, $0 < \gamma < 1$;
- (5) The kernels K with support $[-1, 1]$ and g are each second order symmetric. (The support restriction is a standard one for convenience.)

Also, write

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

Theorem 1. Under Conditions (1)–(5), if $h_1 \rightarrow 0$, $\sigma \rightarrow 0$ and $nh_1 \rightarrow \infty$, then

$$\begin{aligned} \text{MSE}\{\hat{\mu}(x)\} &\simeq \frac{1}{4f^2\{q_\beta(x)|x\}} [\mu_2(K)h_1^2 F^{20}\{q_\beta(x)|x\} + \mu_2(G)\sigma^2 F^{02}\{q_\beta(x)|x\}]^2 \\ &+ \frac{\int K^2(t)dt}{nh_1 f_X(x) f^2\{q_\beta(x)|x\}} \left[\beta(1-\beta) - \sigma f\{q_\beta(x)|x\} \int G(t)\{1-G(t)\}dt \right] \\ &+ o(h_1^4 + \sigma^4 + \sigma/nh_1) \end{aligned}$$

where $\mu_2(L) = \int x^2 L(t)dt$.

Inter alia, in Yu and Jones (1998) this 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 $q_\beta(x)$ 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 section, although we note, as before, that precise choice of σ is not as crucial as that of h_1 .

3.4. Bandwidth selection

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 section 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

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

Again, this proves to be nice and simple to solve computationally. For given μ and μ_1 , write $W_i = \mu + \mu_1(X_i - x) - Y_i$, $i = 1, \dots, n$. 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 $\sigma \rightarrow 0$ and to $-\infty$ as $\sigma \rightarrow \infty$. 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}_β ; the version of the new double kernel local linear estimator (3.2) using the ‘old’ bandwidth selection methodology will be denoted \hat{q}_β^0 and that using the different value for σ described in Section 3.4, \hat{q}_β^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, pp.234–235). 100 replications were made, integrated squared errors (ISEs) were computed over suitable ranges and median ISEs reported in Table 1.

* * * Table 1 about here * * *

Comparing \hat{q}_β , (3.2), with \tilde{q}_β , (3.3), shows a performance consistently as good and sometimes considerably better for the former. We are, therefore, happy to recommend \hat{q}_β 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 outperform the single kernel local linear check function approach to kernel quantile regression. The comparison of bandwidth selection methods between \hat{q}_β^0 and \hat{q}_β^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 \hat{q}_β continues to display a (sometimes considerable) improvement over \tilde{q}_β .

3.6. Example

We 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 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 Fig. 1(a), those associated with the new method together with the new bandwidth methodology (3.4) in Fig. 1(b). The general level of triceps

skinfold decreases a little up to age 10 and there is then a rapid increase upwards 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.

* * * Fig. 1 about here * * *

Comparable pictures have been produced for these data by the LMS method in Cole and Green (1992, Fig. 2) employing two subjectively chosen levels of smoothing and in Yu and Jones (1998, Fig. 1(b)) employing our earlier double 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 Fig. 2, although with differences at a detailed level. The new bandwidth selection method has yielded a picture (Fig. 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.

3.7 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 & 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. Closing remarks

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) above) 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 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 that, in contrast to our earlier approach, the new approach also affords immediate estimation of the first derivative with respect to x of the regression quantile through the value of $\hat{\mu}_1 = \hat{\mu}_1(x)$ obtained through (3.2) (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 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.

References

- Azzalini, A. (1981) A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.
- Barndorff-Nielsen, O. and Blaesild, P. (1983) Hyperbolic distributions. In *Encyclopedia of Statistical Sciences* (eds N.L. Johnson *et al.*), pp. 700–707. New York: Wiley.
- Bassan, B., Denuit, M. and Scarsini, M. (1999) Variability orders and mean differences. *Statist. Probab. Lett.*, **45**, 121–130.
- Chaudhuri, P., Doksum, K. and Samarov, A. (1997) On average derivative quantile regression. *Ann. Statist.*, **25**, 715–744.
- Cole, T.J. and Green, P.J. (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Statist. Med.*, **11**, 1305–1319.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
- Huber, P.J. (1964) Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73–101.
- Jones, M.C. (2006a) On a class of distributions with simple exponential tails. Under consideration. Open University Department of Statistics Technical Report 06/02; see <http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm>

- Jones, M.C. (2006b) The logistic and the log F distribution. In *Handbook of the Logistic Distribution*, Second Edition (ed. N. Balakrishnan). New York: Dekker, to appear. Open University Department of Statistics Technical Report 06/01; see <http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm>
- Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press. Boston: Birkhauser.
- Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R. and Machado, J.A.F. (1999) Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.*, **94**, 1296–1310.
- Kotz, S., Kozubowski, T.J. 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, C.N. (1982) Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65–80.
- Nadaraya, E.A. (1964) Some new estimates for distribution functions. *Theor. Probab. Applic.*, **15**, 497–500.
- Ruppert, D., Sheather, S.J. and Wand, M.P. (1995) An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257–1270.
- Yu, K. (1997) Smooth regression quantile estimation. Ph.D. thesis, Department of Statistics, The Open University.
- Yu, K. and Jones, M.C. (1998) Local linear quantile regression. *J. Amer. Statist. Assoc.*, **93**, 228–237.
- Yu, K. and Moyeed, R.A. (2001) Bayesian quantile regression. *Statist. Probab. Lett.*, **54**, 437–447.

Appendix: outline proof of Theorem 1

Let $\theta = (\mu, \mu_1)^T$ and $\hat{\theta} = (\hat{\mu}, \hat{\mu}_1)^T$ where $\hat{\theta}$ satisfies the pair of equations (3.2). Let $K_i = h_1^{-1}K\{h_1^{-1}(x - X_i)\}$, $q(x) = q_\beta(x)$, $\bar{q}(x, u) = q(x) + q'(x)(u - x)$, $\theta^* =$

$(nh_1)^{1/2}[\mu - q(x), h_1\{\mu_1 - q'(x)\}]^T$ and $Z_i = \{1, (X_i - x)/h_1\}^T$. Then

$$\mu + \mu_1(X_i - x) = \bar{q}(x, X_i) + a_n \theta^{*T} Z_i$$

with $a_n = (nh_1)^{-1/2}$. If $\hat{\theta}$ satisfies (3.2), then $\hat{\theta}^* = (nh_1)^{1/2}[\hat{\mu} - q(x), h_1\{\hat{\mu}_1 - q'(x)\}]^T$ satisfies

$$\beta \sum_{i=1}^n K_i Z_i = \sum_{i=1}^n G \left\{ \frac{\bar{q}(x, X_i) + a_n \theta^{*T} Z_i - Y_i}{\sigma} \right\} K_i Z_i.$$

Using Taylor expansion of the right-hand side of the above equation we have $nV_n = -na_n A_n \theta^* + o(h_1^2)$ where

$$A_n = \frac{1}{n\sigma} \sum_{i=1}^n g \left\{ \frac{\bar{q}(x, X_i) - Y_i}{\sigma} \right\} K_i (Z_i^T Z_i)$$

and

$$V_n = \frac{1}{n} \sum_{i=1}^n \left[G \left\{ \frac{\bar{q}(x, X_i) - Y_i}{\sigma} \right\} - \beta \right] K_i Z_i.$$

Note that $A_n = EA_n + O_P\{\sqrt{\text{Var}(A_n)}\}$ and that $EA_n = E_X(E_{Y|X} A_n)$.

Now, $\int g(u)du = 1$ and $\int ug(u)du = 0$ and because the support of K is $[-1, 1]$, we need consider only those X_i 's for which $|X_i - x| \leq h_1$. We have

$$\begin{aligned} E_{Y|X} g \left\{ \frac{\bar{q}(x, X_i) - Y_i}{\sigma} \right\} &= \int g \left\{ \frac{\bar{q}(x, X_i) - u}{\sigma} \right\} f(u|X_i) du \\ &= \sigma \int g(v) f\{q(x, X_i) - \sigma v|X_i\} dv \\ &= \sigma f\{q(x)|x\} + o(\sigma) \end{aligned}$$

so that, for $k, \ell = 1, 2$,

$$E\{(A_n)_{k\ell}|\mathcal{X}\} = [f\{q(x)|x\} + o(1)] \left\{ f(x)\mu_{k+\ell-2}(K) + h_1 f'(x)a_{i+j-1}(K) + o(h_1) \right\}$$

where $\mu_r(K) = \int u^r K(u)du$ which is zero if r is odd. Now let P and Q be 2×2 matrices having (k, ℓ) 'th entry equal to $\mu_{k+\ell-2}(K)$ and $\mu_{k+\ell-1}(K)$, respectively. Then, with $E_x = f(x)P$ and $F_x = f'(x)Q$, $E(A_n|\mathcal{X}) = [f\{q(x)|x\} + o(1)]\{E_x + h_1 F_x + o(h_1)\}$.

Noting that $\text{Var}\{(A_n)_{k\ell}|\mathcal{X}\} = O\{(nh_1)^{-1}\}$, $k, \ell = 1, 2$, we can write

$$\hat{\theta}^* \simeq -\frac{1}{a_n f\{q(x)|x\}} (E_x + h_1 F_x)^{-1} V_n,$$

with $V_n = n^{-1} \sum_{i=1}^n W_i$ and $W_i = (G[\sigma^{-1}\{\bar{q}(x, X_i) - Y_i\}] - \beta) K_i Z_i$. Then, along the same lines as Chapter 2 of Yu (1997) or the proofs of Lemmas 1 and 2 of Yu and Jones (1998), we calculate the first components of $E(W_i) = E_X\{E_{Y|X}(W_i)\}$ and $\text{Var}(W_i)$ and complete the proof immediately.

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

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

Fig. 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).

