L-ESTIMATION FOR LINEAR HETEROSCEDASTIC MODELS

by

ROGER KOENKER

and

QUANSHUI ZHAO

University of Illinois at Urbana-Champaign

ABSTRACT

L-estimators based on a weighted regression quantile process are considered for a class of linearly heteroscedastic regression models. It is shown that the resulting estimators are "efficient" in the sense introduced by Gutenbrunner(1992).

1. INTRODUCTION

L-statistics, or linear combinations of order statistics, offer a rich source of estimators for the (univariate) one-sample problem. Bickel and Lehmann (1975) in their survey of descriptive statistics for nonparametric models conclude:

> Of the three classes considered [M, L, and R estimators], it is found that trimmed expectations (and certain other weighted quantiles) are the only ones which are both robust and whose estimators have guaranteed high efficiency ...

There have been several suggestions for extending the L-estimator approach to the linear model. Bickel's (1973) one-step approach was pioneering, but suffered from a lack of equivariance. This was later remedied in the important work of Welsh (1987) who constructed one-step L-estimators for the linear model which satisfied natural equivariance requirements as well as exhibiting asymptotic behavior analogous to that of one-sample L-statistics.

Two other approaches to L-estimators for linear models have been suggested by work on quantile regression. Koenker and Bassett (1978) considered discrete linear combinations of "regression quantiles" which they defined as

$$\hat{\beta}(\tau) = \arg\min_{b \in \mathbf{R}^p} \sum \rho_\tau(y_i - x_i' b)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. Since these p-dimensional analogues of the sample quantiles have asymptotic behavior like that of their one-sample counterparts it was straightforward to establish that estimators of the form

$$\sum_{i=1}^{m} w_i \hat{\beta}_n(\tau_i)$$

had analogous asymptotic behavior to the "systematic statistics" investigated by Mosteller (1946), Bennett (1952) and many others. This approach was later extended to general weight functions

$$\tilde{\beta}_n^\nu = \int_0^1 \hat{\beta}_n(\tau) d\nu(\tau)$$

in Koenker and Portnoy (1987), Koenker and Portnoy (1989), and Gutenbrunner and Jurečková(1992). An excellent recent treatment of similar models from the standpoint of M-estimation is contained in Carroll and Ruppert(1988).

1.1 L-estimators for linear models with iid errors

Consider the classical linear model

$$y_i = x_i' \beta + u_i$$

with iid $\{u_i\}$ from distribution function F, and $x_{1i} = 1$ for $i = 1, \ldots, n$. Let ν be a finite signed measure concentrated on a compact subinterval of $(0, 1)$, then under mild further conditions on F and the design sequence we have, setting $e_1 = (1, 0, \ldots, 0)'$,

$$\sqrt{n}(\tilde{\beta}_n^\nu - \beta(\nu, F)) \longrightarrow^{\mathcal{D}} \mathcal{N}(0, \sigma^2(\nu, F) Q_0^{-1})$$

where

$$\beta(\nu, F) = \nu(0,1)\beta + \mu(\nu, F)e_1$$
$$\mu(\nu, F) = \int F^{-1}(t)d\nu(t)$$
$$\sigma^2(\nu, F) = \int \int \frac{t \wedge s - st}{f(F^{-1}(s))f(F^{-1}(t))} d\nu(t)d\nu(s),$$

and

$$Q_0 = \lim_{n \to \infty} n^{-1}(X'X).$$

This result directly parallels the theory for one-sample L-statistics. Note that for probability measures ν, or more generally for any ν such that $\nu(0,1) = 1$, $\tilde{\beta}_n^\nu$ is location equivariant; while if ν has total mass zero, $\tilde{\beta}_n^\nu$ is location invariant. In the latter case natural estimators of scale may be constructed by requiring that $\nu(0,u] \leq 0$ for all $u \in (0,1)$. To illustrate, the asymptotically optimal L-estimator for β when F is Cauchy would employ $\nu(A) = \int_A J(u)du$ where

$$J(u) = 2\cos(2\pi u)(\cos(2\pi u) - 1)$$

and the asymptotically optimal L-estimator for scale when F is Cauchy would use

$$J(u) = 8\cos^3(\pi(u - 1/2))\sin(\pi(u - 1/2)).$$

Note that the optimal Cauchy score function for location is negative in the tails so extreme order statistics receive negative weight. In applications it would be advisable to trim the Cauchy J functions to remove the efffect of the extreme order statistics on the estimator. A more conventional example might be the optimal B-robust L-estimators at the normal model. See, e.g., Hampel *et al* (1986, p. 124). For location in this case we have the familiar trimmed mean,

$$J(u) = (1 - 2\alpha)^{-1} I(\alpha < u < 1 - \alpha)$$

while for scale we have for $0 < \alpha_0 < \alpha_1 < \frac{1}{2}$,

$$J(u) = \Phi^{-1}(u) I(\alpha_0 < u < \alpha_1 \quad \text{or} \quad 1 - \alpha_1 < u < 1 - \alpha_0)/(2\alpha_1 - 2\alpha_0).$$

See Welsh (1990) for a detailed treatment of the latter estimator as well as an excellent general discussion of L-estimation of scale.

Gutenbrunner and Jurečková(1992) introduce a second approach to L-statistics for the linear model based on the regression rankscore process,

$$\hat{a}_n(\tau) = \arg\max\{y'a | a \in [0,1]^n, \ X'a = (1-\tau)X'1\}$$

which is formally dual to the regression quantile problem in the sense of linear programming. For ν generated as $\nu(A) = \int_A J(t)dt$ with $\int_0^1 J(t)dt = 1$, they set $\hat{J} = diag(\hat{J}_{ni})$,

$$\hat{J}_{ni} = \int_0^1 \hat{a}_{ni}(t) dJ(t)$$

and for $J(t) > 0$ let

$$\check{\beta}_n^\nu = (X'\hat{J}X)^{-1} X'\hat{J}y$$

while for general $J = J^+ - J^-$ they define $\check{\beta}_n^\nu = \check{\beta}_n^{\nu+} - \check{\beta}_n^{\nu-}$. The simplest (and therefore perhaps most compelling) form of this is the so-called trimmed least squares estimator for which

$$J_\alpha(t) = (1-2\alpha)^{-1} I(\alpha < u < 1-\alpha)$$

which generates weights,

$$\hat{J}_{ni} = \hat{a}_{ni}(\alpha) - \hat{a}_{ni}(1-\alpha).$$

Noting that $\hat{a}_{ni}(\tau) = 1$ if $y_i > x_i'\hat{\beta}_n(\tau)$, equals zero if $y_i < x_i'\hat{\beta}_n(\tau)$ and takes some intermediate value otherwise, we see that for $J_\alpha(t)$ the weights $\hat{J}_{ni}$ are one if

$x_i'\hat{\beta}_n(\alpha) < y_i < x_i'\hat{\beta}_n(\alpha)$, $\hat{J}_{ni} = 0$ if y_i lies strictly outside this interval, and take an intermediate value otherwise, that is if $y_i = x_i'\hat{\beta}(\tau)$ for $\tau \in \{\alpha, 1-\alpha\}$. A simpler version of this, which used only $0-1$ weights was considered earlier by Ruppert and Carroll (1980). Gutenbrunner and Jurečková(1992) establish the asymptotic equivalence of $\tilde{\beta}_n^\nu$ and $\check{\beta}_n^\nu$ under iid error conditions as well as under contiguous alternatives. They also consider more general linear regression-scale models for which the asymptotic behavior of the two estimators diverge.

1.2 L-estimators for linearly heteroscedastic models

A more general, natural setting for quantile regression and L-estimators in particular is the linear heteroscedastic model

$$y_i = x_i'\beta + x_i'\gamma u_i$$

where again the $\{u_i\}$ are iid from F. In this model $\tilde{\beta}_n^\nu$ based on the unweighted regression quantiles can be shown (Gutenbrunner and Jurečková(1992)) to satisfy,

$$\sqrt{n}(\tilde{\beta}_n^\nu - \beta_n(\nu, F)) \longrightarrow^{\mathcal{D}} \mathcal{N}(0, \sigma^2(\nu, F) Q_1^{-1} Q_0 Q_1^{-1})$$

where e_1 is replaced by γ in the definition of $\beta(\nu, F)$, $Q_r = \lim n^{-1} X'\Gamma^{-r} X$, $r = 0, 1, 2$, and $\Gamma = diag(x_i'\gamma)$. To contrast the behavior of $\tilde{\beta}_n^\nu$ and $\check{\beta}_n^\nu$ under this linear heteroscedastic model it is convenient to contrast their linear representations. Thus,

$$\tilde{\beta}_n^\nu = K_1 \tilde{y}(\nu, F) + o_p(n^{-\frac{1}{2}})$$

where, as above, we will write, following Gutenbrunner and Jurečková(1992),

$$K_r = (X'\Gamma^{-r} X)^{-1} X'\Gamma^{-r}, \qquad r = 0, 1$$

$$\tilde{y}_i(\nu, F) = x_i'\beta + x_i'\gamma(\psi_{\nu,F}(u_i) - \mu(\nu, F))$$

and

$$\psi_{\nu,F}(u) = \int (t - I(F(u) \leq t))(f(F^{-1}(t)))^{-1} d\nu(t).$$

We may interpret this representation as establishing an asymptotic equivalence (to order $n^{-\frac{1}{2}}$) for $\tilde{\beta}_n^\nu$ and the pseudo-estimator defined by the weighted least squares regression of X on the pseudo-observations $\tilde{y}(\nu, F)$ using weights Γ^{-1}.

In contrast, from Gutenbrunner and Jurečková(1992, Theorem 3),

$$\breve{\beta}_n^\nu = K_0 \breve{y}(\nu, F) + K_1(\tilde{y}(\nu, F) - \breve{y}(\nu, F)) + o_p(n^{-\frac{1}{2}})$$

where

$$\breve{y}_i(\nu, F) = x_i'\beta + x_i'\gamma\breve{\psi}_{J,F}(u_i)$$
$$\breve{\psi}_{J,F}(u) = \left\{ \begin{array}{ll} J(F(u))[u - \mu(\nu, F)] & \text{if } J > 0 \\ \breve{\psi}_{J+,F}(u) - \breve{\psi}_{J-,F}(u) & \text{otherwise.} \end{array} \right\}$$

It follows that $\tilde{\beta}_n^\nu$ and $\breve{\beta}_n^\nu$ have different asymptotic behavior when $\gamma \neq e_1$. Gutenbrunner (1992) has proposed a modification of $\breve{\beta}_n^\nu$ which "corrects" for the effect of the linear heteroscedasticity. For any estimator γ_n satisfying

$$\hat{\gamma}_n = \gamma + O_p(n^{-\frac{1}{2}})$$

he constructs

$$\hat{B}_n^\nu = \int (\hat{\beta}_{n1}(u) - \tilde{\beta}_{n1}^\nu)(\hat{a}_{ni}(u) - 1 - u) dJ(u)$$
$$\Gamma_n = diag(X\hat{\gamma}_n)$$

and shows that the estimator

$$\breve{\beta}_n^\nu = (X'\hat{J}\hat{\Gamma}^{-2}X)^{-1}X'(\hat{J}\hat{\Gamma}^{-2}y - \hat{\Gamma}^{-1}\hat{B}_n^\nu)$$

has the linear representation,

$$\breve{\beta}_n^\nu = \beta(\nu, F) + (X'\Gamma^{-2}X)^{-1}X'\Gamma^{-2}\Psi_{\nu,F} + o_p(n^{-\frac{1}{2}})$$

where $\Psi_{\nu,F} = (\psi_{\nu,F}(u_i))$ and $\psi_{\nu,F}$ is the influence function of the L-statistic $F \to \int F^{-1} d\nu$, defined above. It follows immediately that

$$\sqrt{n}(\breve{\beta}_n^\nu - \beta(\nu, F)) \longrightarrow^{\mathcal{D}} \mathcal{N}(0, \sigma^2(\nu, F)Q_2^{-1}).$$

The first term of this "efficient" L-estimator, $\breve{\beta}_n^\nu$ represents a natural reweighting of the original form of $\breve{\beta}_n^\nu$ to accomodate the heterogeneity in scale of the model. However, the second term is more surprising. Gutenbrunner (1992) refers to it as a " 'smooth Winsorizing' of residuals because we did not use the 'right' RQ's, namely the optimally weighted RQ's." He also notes that this term is closely related to the Winsorization employed in Welsh (1987) to construct one-step L-estimators for the linear model. In effect the first coordinate of the centered regression quantile process is used to estimate $F^{-1}(u) - \mu(\nu, F)$ and then to adjust the dual L-statistic.

The foregoing discussion raises a natural question: can one, by simply estimating an appropriately *weighted* regression quantile process, construct primal (and dual) L-estimators which achieve the same "efficient" asymptotic behavior as $\breve{\beta}_n^\nu$? An affirmative answer to this question is provided in the next section. Proofs are collected in the last section.

2. RESULTS

Consider the linearly heteroscedastic model

$$y_i = x_i'\beta + (x_i'\gamma)u_i \qquad i = 1, \ldots, n \tag{2.1}$$

where the random variables $\{u_i\}$ are iid with distribution function F. We will employ the following conditions

C1. $\sigma_i \equiv x_i'\gamma > 0, \quad i = 1, \ldots, n.$

C2. $n^{-1}\sum \sigma^{-r} x_i x_i' = Q_r + R_{rn}$ where $Q_r : r = 0, 1, 2$, are positive definite, and the maximum eigenvalues $\lambda_{max}(Q_{rn}) \longrightarrow 0$, for $r = 0, 1$ and $\lambda_{max}(R_{2n}) = O(n^{-1/4})$.

C3. $\sum_1^n \| x_i/\sigma_i \|^3 = O(n)$.

C4. $\max_i \| x_i/\sigma_i \| = O(n^{1/4})$.

C5. F has a density f and there exists $\epsilon > 0$, such that $s(u) \equiv f(F^{-1}(u)) > 0$ and $s'(u)$ is uniformly bounded for $u \in [\epsilon, 1-\epsilon]$.

Conditions C3 - C5 are adopted directly from Koenker and Portnoy (1987) with x_i/σ_i in place of x_i used there. C1 and C2 are somewhat modified versions of Gutenbrunner's (1992) conditions, and are necessary for the $\sqrt{n}$-consistency of the preliminary estimator of γ. Obviously, we can, if we so desire, restrict the form of (2.1) so that different vectors of covariates appear in the location and scale components of the regression function, at the cost of some increased notational complexity.

Under these conditions the results of Gutenbrunner and Jurečková(1992) and Gutenbrunner (1992) assure a $\sqrt{n}$ consistent estimator $\hat{\gamma}_n$ in the following sense: we will say that $\hat{\gamma}_n$ is $\sqrt{n}$-consistent up to scale if for some F dependent constant κ,

$$\hat{\gamma}_n = \kappa^{-1}\gamma + n^{-1/2}\kappa^{-1}\hat{\delta}_n$$

with $\hat{\delta}_n = O_p(1)$. Given such an estimator we may compute the weighted regression quantile process,

$$\hat{\beta}_n(\tau, \hat{\gamma}) = \arg\min_{b \in \mathbf{R}^p} \sum \hat{\sigma}_i^{-1} \rho_\tau (y_i - x_i b) \qquad (2.2)$$

where $\hat{\sigma}_i = x_i'\hat{\gamma}_n$. It is obvious that consistency up to scale is sufficient for the reweighting in (2.2) since only relative scale matters. Our main result is the following Bahadur representation of this weighted regression quantile process.

THEOREM 2.1 Let $\hat{\gamma}_n$ be $\sqrt{n}$-consistent up to scale, and $\beta(\tau) = \beta + \gamma F^{-1}(\tau)$. Then under C1-C5,

$$\sqrt{n}(\hat{\beta}(\tau, \hat{\gamma}) - \beta(\tau)) = \frac{Q_2^{-1}}{f(F^{-1}(\tau))} n^{-1/2} \sum_{i=1}^{n} \sigma_i^{-1} x_i \psi_\tau(u_i - F^{-1}(\tau)) + O_p(n^{-1/4} \log n)$$

uniformly for $\tau \in [\epsilon, 1-\epsilon]$, where $\psi_\tau(u) = \tau - I(u < 0)$.

The proof of this theorem is rather complicated and is given in next section. It should be noted that when $\gamma = e_1$ the model (1.3) simplifies to a linear homoscedastic model, therefore Theorem 2.1 here implies Theorem 2.1 in Koenker and Portnoy (1987). However, we have removed condition X4 in Koenker and Portnoy (1987), which is very restrictive. Given this representation the following result is immediate.

THEOREM 2.2 Under the conditions of Theorem 2.1, let ν be a finite signed measure on $[0, 1]$, vanishing off the interval $[\epsilon, 1-\epsilon]$ for $\epsilon \in (0, 1/2)$. Then

$$\tilde{\beta}_n^\nu(\hat{\gamma}) = \int_0^1 \hat{\beta}_n(\tau, \hat{\gamma}) d\nu(\tau)$$

satisfies

$$\sqrt{n}(\tilde{\beta}_n^\nu(\hat{\gamma}_n) - \beta(\nu, F)) \longrightarrow^{\mathcal{D}} \mathcal{N}(0, \sigma^2(\nu, F) Q_2^{-1})$$

where $\beta(\nu, F) = \int_0^1 \beta(t) d\nu(t)$ and $\sigma^2(\nu, F)$ is as defined above.

Finally, by replacing the original unweighted regression rankscore process by the weighted dual process corresponding to (2.2)

$$\hat{a}_n(\tau, \hat{\gamma}_n) = \arg\max\{y'\hat{\Gamma}^{-1}a | a \in [0,1]^n,\ X'\hat{\Gamma}^{-1}a = (1-\tau)X'\hat{\Gamma}^{-1}1\}$$

and computing

$$\hat{J}_{ni} = \int_0^1 \hat{a}_{ni}(t, \hat{\gamma}_n) dJ(t)$$

and

$$\check{\beta}_n^{\nu}(\hat{\gamma}_n) = (X'\hat{J}\hat{\Gamma}^{-2}X)^{-1}X'\hat{J}\hat{\Gamma}^{-2}y,$$

Theorem 2.1 and the results of Gutenbrunner and Jurečková(1992) imply the following result.

THEOREM 2.3 Under the conditions of Theorem 2.1, let $\nu(A) = \int_A J(u)du$ for J of bounded variation, with $J(u) = 0$ on $[0,\epsilon] \cap [1-\epsilon, 1]$ for $\epsilon \in (0, 1/2)$ and $\nu(0,1) = 1$. Then

$$\sqrt{n}(\check{\beta}_n^{\nu}(\hat{\gamma}_n) - \beta(\nu, F)) \longrightarrow^{\mathcal{D}} \mathcal{N}(0, \sigma^2(\nu, F)Q_2^{-1})$$

REMARKS. Theorems 2.2 and 2.3 establish that the primal and dual L-estimators, $\tilde{\beta}_n^{\nu}(\hat{\gamma}_n)$ and $\check{\beta}_n^{\nu}(\hat{\gamma}_n)$ respectively, are *efficient L-statistics corresponding to* ν in the sense of Gutenbrunner (1992). For the primal L-estimators of Theorem 2.2, if ν is chosen to satisfy $\nu(0,1) = 1$ and $\int_0^1 F^{-1}(t)d\nu(t) = 0$, then either will be an estimator for β; while $\tilde{\beta}_n^{\nu}$ can be an scale L-estimator of γ if ν is chosen to satisfy $\nu(0,1) = 0$ and $\int_0^1 F^{-1}(t)d\nu(t) = 1$. Whether the dual L-estimators can be used to estimate γ is, to us, unclear. It may be noted that the restriction imposed on $\int_0^1 F^{-1}(t)d\nu(t)$, is trivially satisfied if we choose ν to be symmetric about 1/2, for location and antisymmetric about 1/2 for scale, when F is a symmetric distribution. Note, finally, that the dependence of estimators on $\hat{\gamma}$ will be suppressed in the subsequent development.

3. PROOFS

The proof of Theorem 2.1 is decomposed into a series of lemmas. The first lemma follows from Theorem 2.1 and Lemma A.2 of Koenker and Portnoy(1987).

LEMMA 3.1 Let

$$V(\Delta,\tau) = n^{-1/2}\sum_1^n \sigma_i^{-1} x_i \psi_\tau(u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1}x_i'\Delta) \qquad (3.1)$$

where $\Delta \in R^k$, $\tau \in (0,1)$ and for $K > 0$ define,

$$D_{n\epsilon} = \{(\Delta,\tau) : \tau \in [\epsilon, 1-\epsilon],\ \parallel \Delta \parallel \leq K\sqrt{\log n}\}. \qquad (3.2)$$

Then under C1-C5,

$$\sup_{D_{n\epsilon}} \parallel V(\Delta,\tau) - V(0,\tau) + f(F^{-1}(\tau))Q_2\Delta \parallel = O_p(n^{-1/4}\log n). \qquad (3.3)$$

LEMMA 3.2 Under C1, C2 and C4,

$$\sup_{\tau\in[\epsilon,1-\epsilon]} \parallel V(0,\tau) \parallel = O_p(\sqrt{\log n}).$$

PROOF. Without loss of generality, assume $\sigma_i \equiv 1$, and provisionally that $x_i \in R^1$. Then

$$V(0,\tau) = n^{-1/2}\sum_1^n x_i(\tau - I(F(u_i) < \tau)).$$

As is well known, u_i iid F implies $U_i = F(u_i)$ is iid uniform on $[0,1]$. Thus

$$E[x_i(\tau - I(U_i < \tau))] = 0, Var(x_i(\tau - I(U_i < \tau))) = x_i^2\tau(1-\tau)$$

and $|x_i(\tau - I(U_i < \tau))| \leq \max_{i\leq n}|x_i| = O(n^{1/4})$. By Bernstein's Theorem (see Serfling 1980, p. 95)

$$P(|\sum_1^n x_i(\tau - I(U_i < \tau))| \geq nt) \leq 2\exp(-\frac{n^2t^2}{2\tau(1-\tau)\sum_1^n x_i^2 + O(n^{5/4})t}).$$

Choose $t = \sqrt{\lambda \log n/n}$, and $\lambda > 0$ so

$$P(|V(0,\tau)| \geq \sqrt{\lambda\log n}) \leq 2\exp(-\frac{\lambda\log n}{\lambda_{\max}(Q_2)/2 + o(1)}).$$

As in the proof of Lemma A.2 in Koenker and Portnoy (1987), by using the chaining argument we have

$$\sup_{\tau\in[\epsilon,1-\epsilon]} |V(0,\tau)| = O_p(\sqrt{\log n}).$$

When the dimension of x_i is greater than one, the result follows from the fact that

$$\sup_{\tau\in[\epsilon,1-\epsilon]} \| V(0,\tau) \| \leq (\sum_1^k (\sup_{\tau\in[\epsilon,1-\epsilon]} |V_j(0,\tau)|)^2)^{1/2} = O_p(\sqrt{\log n})$$

where $V_j(0,\tau)$ is the j-th component of $V(0,\tau)$. □

The next lemmas are used to approximate $V(\Delta,\tau)$ by

$$\hat{V}(\Delta,\tau) = n^{-1/2}\sum_1^n \frac{1}{\kappa\hat{\sigma}_i} x_i \psi_\tau(u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1}x_i'\Delta) \tag{3.4}$$

where $\kappa\hat{\sigma}_i = \sigma_i + n^{-1/2}x_i'\hat{\delta}_n$. If the approximation is sufficiently accurate, the result in Lemma 3.1 is true even if $V(\Delta,\tau)$ is replaced by $\hat{V}(\Delta,\tau)$.

LEMMA 3.3 Under C1-C5,

$$\sup_{D_{n\epsilon}} \| \hat{V}(\Delta,\tau) - V(\Delta,\tau) \| = O_p(n^{-1/4}\sqrt{\log n}).$$

Further,

$$\sup_{D_{n\epsilon}} \| \hat{V}(\Delta,\tau) - V(0,\tau) + f(F^{-1}(\tau))Q_2\Delta \| = O_p(n^{-1/4}\log n) \tag{3.5}$$

PROOF Since $\hat{\gamma}_n$ is $\sqrt{n}$-consistent up to scale, $\kappa\hat{\sigma}_i = \sigma_i(1 + n^{-1/2}\sigma_i^{-1}x_i'\hat{\delta}_n)$, so by C4 we have $\| n^{-1/2}\sigma_i^{-1}x_i'\hat{\delta}_n \| = O_p(n^{-1/4})$. By Taylor's expansion,

$$(1 + n^{-1/2}x_i'\sigma_i^{-1}\hat{\delta}_n)^{-1} = 1 - n^{-1/2}\sigma_i^{-1}x_i'\hat{\delta}_n + n^{-1} \| x_i/\sigma_i \|^2 O_p(1)$$

thus

$$\frac{1}{\kappa\hat{\sigma}_i} - \frac{1}{\sigma_i} = \frac{1}{\sigma_i}(-n^{-1/2}\sigma_i^{-1}x_i'\hat{\delta}_n + n^{-1} \| x_i/\sigma_i \|^2 O_p(1)). \tag{3.6}$$

It follows that

$$\begin{aligned}
&\sup_{D_{n\epsilon}} \| \hat{V}(\Delta,\tau) - V(\Delta,\tau) \| \\
\leq &\sup_{D_{n\epsilon}} \| n^{-1/2} \sum_1^n \frac{1}{\sigma_i^2} x_i x_i' \psi_\tau (u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1} x_i'\Delta) \| O_p(n^{-1/2}) \\
&+ n^{-1} \sum_1^n \| x_i/\sigma_i \|^3 O_p(n^{-1/2})
\end{aligned} \qquad (3.7)$$

By C3, the second term in the right hand side of the inequality is $O_p(n^{-1/2})$. As for the first term, let

$$T(\Delta,\tau) = n^{-1/2} \sum_1^n \frac{1}{\sigma_i^2} x_i x_i' \psi_\tau (u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1} x_i'\Delta).$$

Then by Lemma 3.4 below, we have

$$\sup_{D_{n\epsilon}} \| T(\Delta,\tau) - T(0,\tau) - ET(\Delta,\tau) \| = O_p(\log n).$$

If we note that for $(\Delta,\tau) \in D_{n\epsilon}$,

$$\begin{aligned}
\| ET(\Delta,\tau) \| &\leq n^{-1} \sum_1^n \| x_i/\sigma_i \|^3 O(\sqrt{\log n}) \\
&+ n^{-3/2} \sum_1^n \| x_i/\sigma_i \|^4 O(\log n) = O(\sqrt{\log n})
\end{aligned}$$

and by Lemma 3.5 below,

$$\sup_{\tau\in[\epsilon,1-\epsilon]} \| T(0,\tau) \| = O_p(n^{1/4}\sqrt{\log n}).$$

Then

$$\sup_{D_{n\epsilon}} \| T(\Delta,\tau) \| \leq O_p(n^{1/4}\sqrt{\log n}).$$

The first part of the lemma follows from (3.7) and (3.5) is obtained by using Lemma 3.1. □

LEMMA 3.4 Under C1-C5, $\sup_{D_{n\epsilon}} \| T(\Delta,\tau) - T(0,\tau) - ET(\Delta,\tau) \| = O_p(\log n)$.

PROOF We can use the procedures in the proofs of the proposition and Lemma A.2 in Koenker and Portnoy (1987). The proof consists of two parts. The first part is to get a probability inequality and the second part is to apply the chaining argument. Since the second part is similar to KP's argument without major change, we will omit the details. The only thing that needs to be proved is that for $\Delta \in D_{n\epsilon}$,

$$P(|T_j(\Delta,\tau) - T_j(0,\tau) - ET_j(\Delta,\tau) \| \geq \lambda \log n) \leq 2\exp(-\lambda \log n\ (1+o(1))) \quad (3.8)$$

where T_j is the jth component of T. Denote v_{ij} as the jth component of $x_i x_i'/\sigma_i^2$, and

$$\tilde{T}_j = n^{1/2}(T_j(\Delta,\tau) - T_j(0,\tau) - ET_j(\Delta,\tau))$$

so

$$T_j(\Delta,\tau) = n^{-1/2}\sum_1^n v_{ij}\psi_\tau(u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1}x_i'\Delta).$$

Adopting KP's notation, Markov's inequality yields

$$P(|\tilde{T}_j| \geq \lambda_n) \leq e^{-t\lambda_n}(M_j(t) + M_j(-t))$$

for $t > 0,\ \lambda_n > 0$, where $\log M_j(t) = \sum_1^n \log M_{ij}(t)$ and

$$\log M_{ij}(t) \leq c|n^{-1/2}\sigma_i^{-1}x_i'\Delta|(v_{ij}t)^2\exp(|v_{ij}t|)$$

for some constant c. Since by condition C4, we have

$$|v_{ij}| \leq \| x_i x_i'/\sigma_i^2 \| \leq \| x_i/\sigma_i \|^2 = O(n^{1/2})$$

thus, using C3,

$$\log M_j(t) \leq c \| \Delta \| \sum_1^n \| x_i/\sigma_i \|^3\ t^2\exp(Bn^{1/2}t) \leq c'n\ \sqrt{\log n}t^2\exp(Bn^{1/2}t)$$

where c' and B are constants. Finally, if we take $t = n^{-1/2}$ and $\lambda_n = \lambda n^{1/2}\log n$, so

$$P(|\tilde{T}_j| \geq \lambda n^{1/2}\log n) \leq 2\exp(-\lambda \log n + c'\sqrt{\log n}\; e^B)$$
$$= 2\exp(-\lambda \log n(1+o(1)))$$

and (3.8) follows. □

LEMMA 3.5 Under C1-C5, $\sup_{\tau\in[\epsilon,1-\epsilon]} \parallel T(0,\tau) \parallel = O_p(n^{1/4}\sqrt{\log n})$.

PROOF The proof is similar to that of Lemma 3.2. Again using Bernstein's Theorem

$$P(n^{1/2}|T_j(0,\tau)| > nt) \leq 2\exp\left(-\frac{n^2t^2}{2\tau(1-\tau)\sum_1^n \parallel x_i/\sigma_i \parallel^4 + t\; O(n^{3/2})}\right).$$

Set $t = \lambda n^{-1/4}\sqrt{\log n}$, so

$$2\tau(1-\tau)\sum_1^n \parallel x_i/\sigma_i \parallel^4 + t\; O(n^{3/2}) \leq O(n^{5/4}) + O(n^{5/4}\sqrt{\log n}) = O(n^{5/4}\sqrt{\log n}.)$$

For n large, there exists $B > 0$, such that

$$P(|T_j(0,\tau)| > \lambda n^{1/4}\sqrt{\log n}) \leq 2\exp(-B\lambda^2 n^{1/4}\sqrt{\log n}.)$$

The proof is completed using this inequality and the chaining argument. □

The next two lemmas use the procedure from Jurečková(1977, Lemma 5.2).

LEMMA 3.6 Let $D^0_{n\epsilon} = \{(\Delta,\tau) : \tau \in [\epsilon, 1-\epsilon],\ \parallel \Delta \parallel = K\sqrt{\log n}\}$. Then under C1-C5, there exists $K > 0$ such that $P(\inf_{D^0_{n\epsilon}}[-\Delta'\hat{V}(\Delta,\tau)] < n^{-1/4}(\log n)^2) \longrightarrow 0$.

PROOF Let $\eta_n = n^{-1/4}(\log n)^2$, then

$$\begin{aligned} &P(\inf_{D^0_{n\epsilon}}[-\Delta'\hat{V}(\Delta,\tau)] < \eta_n) \\ \leq &P(\inf_{D^0_{n\epsilon}}[-\Delta'\hat{V}(\Delta,\tau)] < \eta_n, \inf_{D^0_{n\epsilon}}[-\Delta'(V(0,\tau) - f(F^{-1}(\tau))Q_2\Delta)] \geq 2\eta_n) \\ &+ P(\inf_{D^0_{n\epsilon}}[-\Delta'(V(0,\tau) - f(F^{-1}(\tau))Q_2\Delta)] < 2\eta_n) \equiv I + II \end{aligned}$$

For the first part I, if we note the fact

$$\begin{aligned}
&\{\inf_{D^0_{n\epsilon}}[-\Delta'\hat{V}(\Delta,\tau)] < \eta_n, \inf_{D^0_{n\epsilon}}[-\Delta'(V(0,\tau) - f(F^{-1}(\tau))Q_2\Delta)] \geq 2\eta_n\} \\
\subseteq&\{\sup_{D^0_{n\epsilon}}(\Delta'(\hat{V}(\Delta,\tau) - V(0,\tau) + f(F^{-1}(\tau))Q_2\Delta)) \geq \eta_n\} \\
\subseteq&\{\sup_{D_{n\epsilon}} \| \hat{V}(\Delta,\tau) - V(0,\tau) + f(F^{-1}(\tau))Q_2\Delta \| \geq \eta_n/K\sqrt{\log n}\}
\end{aligned}$$

and $\eta_n/K\sqrt{\log n} = K^{-1}n^{-1/4}(\log n)^{3/2}$, which has higher order than $n^{-1/4}\log n$, it follows from (3.5) that $I \to 0$ as $n \to 0$.

For II, since for $(\Delta,\tau) \in D^0_{n\epsilon}$,

$$\begin{aligned}
&-\Delta'(V(0,\tau) - f(F^{-1}(\tau))Q_2\Delta) = -\Delta'V(0,\tau) + f(F^{-1}(\tau))\Delta'Q_2\Delta \\
\geq& - \| \Delta \| \| V(0,\tau) \| + \min_{\tau\in[\epsilon,1-\epsilon]} f(F^{-1}(\tau))\lambda_{\min}(Q_2) \| \Delta \|^2 \\
=& - K\sqrt{\log n} \| V(0,\tau) \| + \phi_0\lambda_0 K^2 \log n
\end{aligned}.$$

By assumption, $\phi_0 = \min_{\tau\in[\epsilon,1-\epsilon]} f(F^{-1}(\tau)) > 0$, $\lambda_0 = \lambda_{\min}(Q_2) > 0$, hence it follows from Lemma 3.2 that for large K,

$$\begin{aligned}
II &\leq P(-K\sqrt{\log n}\sup_{D^0_{n\epsilon}} \| V(0,\tau) \| + \phi_0\lambda_0 K^2 \log n < 2\eta_n) \\
&= P(\sup_{\tau\in[\epsilon,1-\epsilon]} \| V(0,\tau) \| \geq K\phi_0\lambda_0\sqrt{\log n} - o(1)) \longrightarrow 0.
\end{aligned}$$

□

LEMMA 3.7 Under C1-C5, for K as chosen in Lemma 3.6, let

$$D^+_{n\epsilon} = \{(\Delta,\tau) : \tau \in [\epsilon, 1-\epsilon], \ \| \Delta \| \geq K\sqrt{\log n}\}$$

Then

$$P(\inf_{D^+_{n\epsilon}} \| \hat{V}(\Delta,\tau) \| < K^{-1}n^{-1/4}(\log n)^{3/2}) \longrightarrow 0. \tag{3.9}$$

PROOF Since

$$\sum_1^n \frac{\sigma_i}{\kappa\hat{\sigma}_i}\rho_\tau(u_i - F^{-1}(\tau) - n^{-1/2}\sigma_i^{-1}x_i'\Delta\lambda), \ \lambda \geq 1$$

is a convex function in λ, the gradient of the function in λ, $-\Delta'\hat{V}(\lambda\Delta,\tau)$, is non-decreasing in λ. Hence for all Δ and $\lambda \geq 1$,

$$-\Delta'\hat{V}(\lambda\Delta,\tau) \geq -\Delta'\hat{V}(\Delta,\tau), \tag{3.10}$$

For $\| \Delta \| \geq K\sqrt{\log n}$, let $\Delta_1 = (\Delta/ \| \Delta \|) \cdot K\sqrt{\log n}$, thus $\| \Delta_1 \| = K\sqrt{\log n}$ and Δ can be represented by Δ_1 as $\Delta = \lambda\Delta_1$ for $\lambda \geq 1$. Hence by the Schwarz inequality

$$-\Delta_1'\hat{V}(\Delta,\tau) \leq \| \Delta_1 \| \cdot \| \hat{V}(\Delta,\tau) \|$$

and then by (3.10)

$$\| \hat{V}(\Delta,\tau) \| \geq -\Delta_1'\hat{V}(\Delta,\tau)/ \| \Delta_1 \| \geq -\Delta_1'\hat{V}(\Delta_1,\tau)/ \| \Delta_1 \|$$

Thus,

$$\begin{aligned} \inf_{\|\Delta\| \geq K\sqrt{\log n}} \| \hat{V}(\Delta,\tau) \| &\geq \inf_{D^0_{n\epsilon}} (-\Delta_1'\hat{V}(\Delta_1,\tau)/ \| \Delta_1 \|) \\ &= \inf_{D^0_{n\epsilon}} (-\Delta'\hat{V}(\Delta,\tau))/K\sqrt{\log n}. \end{aligned}$$

Further, by Lemma 3.6,

$$\begin{aligned} &P(\inf_{D^+_{n\epsilon}} \| \hat{V}(\Delta,\tau) \| < K^{-1}n^{-1/4}(\log n)^{3/2}) \\ &\qquad \leq P(\inf_{D^0_{n\epsilon}} (-\Delta'\hat{V}(\Delta,\tau)) < n^{-1/4}(\log n)^2) \longrightarrow 0. \end{aligned}$$

□

LEMMA 3.8 Under C1-C5, $\sup_{\tau\in[\epsilon,1-\epsilon]} \| \hat{\beta}_n(\tau,\hat{\gamma}) - \beta(\tau) \| = O_p(\sqrt{\log n/n})$.

PROOF Let $\hat{\Delta}(\tau) = \sqrt{n}(\hat{\beta}_n(\tau,\hat{\gamma}) - \beta(\tau))$, using Lemma A.2 in Carroll and Ruppert (1980), we have

$$\| \hat{V}(\hat{\Delta}(\tau),\tau) \| \leq (k+1)n^{-1/2} \max_{i\leq n} \| x_i/\kappa\hat{\sigma}_i \|$$

uniformly in τ. Using (3.6), we have

$$\begin{aligned} \| x_i/\kappa\hat{\sigma}_i \| &\leq \| x_i/\sigma_i \| + \| x_i/\sigma_i \|^2 O_p(n^{-1/2}) + \| x_i/\sigma_i \|^3 O_p(n^{-1}) \\ &= O(n^{1/4}) + O_p(1) + O_p(n^{-1/4}) = O_p(n^{1/4}) \end{aligned}$$

hence

$$\sup_{0\le\tau\le1} \| \hat{V}(\hat{\Delta}(\tau),\tau) \| = O_p(n^{-1/4}). \tag{3.11}$$

Now, combining (3.9) and (3.11) we have

$$\begin{aligned} &P(\sup_{\tau\in[\epsilon,1-\epsilon]} \| \hat{\beta}_n(\tau,\hat{\gamma}) - \beta(\tau) \| \ge K\sqrt{\log n/n}) \\ \le &P(\sup_{\tau\in[\epsilon,1-\epsilon]} \| \hat{\Delta}(\tau) \| \ge K\sqrt{\log n}, \sup_{\tau\in[\epsilon,1-\epsilon]} \| \hat{V}(\hat{\Delta}(\tau),\tau) \| < n^{-1/4}\log n) \\ &+ P(\sup_{\tau\in[\epsilon,1-\epsilon]} \| \hat{V}(\hat{\Delta}(\tau),\tau) \| \ge n^{-1/4}\log n) \\ \le &P(\inf_{D^+_{n\epsilon}} \| \hat{V}(\Delta,\tau) \| < n^{-1/4}\log n) + o(1) \\ \le &P(\inf_{D^+_{n\epsilon}} \| \hat{V}(\Delta,\tau) \| < K^{-1}n^{-1/4}(\log n)^{3/2}) + o(1) = o(1). \end{aligned}$$

□

PROOF OF THEOREM 2.1 The result follows from the combination of Lemma 3.3, Lemma 3.8 and (3.11). □

REFERENCES

Bennett, C.A. (1952) Asymptotic properties of ideal linear estimators, Ph.d. Dissertation, University of Michigan

Bickel, P. J. and E. L. Lehmann. (1975) Descriptive Statistics for Nonparametric Models II: Location, *Ann. Statist.*, **3**, 1045-1069

Bickel, P. (1973) On some analogues to linear combinations of order statistics in the linear model, *Ann. Statist.*, **1**, 597-616.

Carroll, R. J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, London: Chapman and Hall.

Gutenbrunner, C. (1992), Efficient L-estimators in Semiparametric Linear Heteroscedastic Models, *L1-Statistical Analysis and Related Methods*, (ed. Y. Dodge) 187-200, New York: Elsevier.

Gutenbrunner, C. and Jurečková, J. (1992) Regression quantile and regression rank score processes in the linear model and derived statistics, *Ann. Statist.*, **20**, 305-330.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley: New York.

Jurečková, J. (1977), Asymptotic Relations of M-estimates and R-estimates in Linear Regression Models, *Ann. Statist.*, **5**, 464-472.

Koenker, R. W. (1987), Discussion of "The Trimmed Mean in the Linear Model" by A. H. Welsh, *Ann. Statist.*, **15**, 39-44.

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

Koenker, R. and Portnoy, S. (1987) L-Estimation for Linear Models, *J. Am Stat. Assoc*, **82**, 851-857.

Portnoy, S. and Koenker, R (1989) Adaptive L-estimation of linear models, *Ann. Statist.*,**17**, 362-381.

Mosteller, F. (1946) On some useful "inefficient" statistics, *Ann. Math. Statist.*, **17**, 377-408.

Ruppert, D. and Carroll, R. J. (1980), Trimmed Least Square Estimation in the Linear Model, *J. Am Stat. Assoc*, **75**, 828-838.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons.

Welsh, A.H. (1987), The Trimmed Mean in the Linear Model, *Ann. Statist.*, **15**, 20-36.

Welsh, A.H. and H. L. Morrison (1990) Robust L-estimation of scale with an application in astronomy *J. of the Amer. Stat. Assn.* **85**, 729-743.