QUANTILE REGRESSION ESTIMATES FOR A CLASS OF LINEAR AND PARTIALLY LINEAR ERRORS-IN-VARIABLES MODELS *

Xuming He
Department of Statistics
University of Illinois

Hua Liang
Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin, Germany

Abstract

We consider the problem of estimating quantile regression coefficients in errors-in-variables models. When the error variables for both the response and the manifest variables have a joint distribution that is spherically symmetric but otherwise unknown, the regression quantile estimates based on orthogonal residuals are shown to be consistent and asymptotically normal. We also extend the work to partially linear models when the response is related to some additional covariate.

Key Words and Phrases: Kernel, linear regression, semiparametric model, errors-in-variables, regression quantile.

*The research of Xuming He was supported in part by National Science Foundation Grant SBR 9617278. The work of Hua Liang was partially supported by Alexander von Humboldt Foundation. Corresponding Author: Xuming He, Department of Statistics, University of Illinois, 725 S. Wright, Champaign, IL 61820. Email: x-he@uiuc.edu.

1 INTRODUCTION

Regression analysis is routinely carried out in all areas of statistical applications. It helps explain how a dependent variable Y relates to independent variables X. Most authors consider the estimation or inference problems based on data observed on both X and Y variables. However, the covariates are not always observable without error. If X is observed subject to random error, the regression model is usually called errors-in-variables (EV) model. A careful study of such models is often needed, as the standard results on regression models do not carry over. The best-known is the effect of attenuation for the likelihood based estimators without any correction for the measurement error in X. A detailed coverage on linear errors-in-variables models can be found in Fuller (1987). More recent work on nonlinear models with measurement errors can be found in Carroll, Ruppert, and Stefanski (1995). The literature on EV models are mainly confined to estimating the conditional mean function of Y given X, assuming Gaussian errors. In the present paper, we attempt to consider conditional median and other quantile functions as pioneered by Koenker and Bassett (1978) for a class of unspecified error distributions. For the usefulness of conditional quantiles, see examples and discussions found in Efron (1991) and He (1997), among many others.

Let's start with the EV model $Y_i = X_i^T\beta + \epsilon_i$ and $W_i = X_i + U_i$ $(i = 1, 2, \cdots, n)$, where $X_i \in R^p$ are unobservable explanatory variables, $W_i \in R^p$ are manifest variables, $Y_i \in R$ are the responses, and $(\epsilon_i, U_i) \in R^{p+1}$ are independent with a common error distribution that is spherically symmetric. Spherical symmetry implies that ϵ_i and each component of U_i have the same distribution, which ensures model identifiability. A special case of such EV models with Gaussian errors and known variance ratio is frequently considered in the literature. Multivariate t-distributions are additional examples for the error structure.

We restrict ourselves to structural models where X_i are independently and identically distributed random variables. If X_i are non-stochastic designs, the model is said to have a functional relationship, see Fuller (1987) for details.

The least squares estimator of β based on $\sum_i (Y_i - W_i^T\beta)^2$ is known to be biased towards zero. It is instructive to consider the quantile regression under the same spirit, but we work with the population version with $p = 1$ for clarity. We ask which $b \in R$ minimizes

$E\rho_\tau(Y - bW)$, where ρ_τ is the τ-th quantile objective function defined as

$$\rho_\tau(r) = \tau \max\{r, 0\} + (1 - \tau) \max\{-r, 0\}. \tag{1.1}$$

Note that the solution to minimizing $E\rho_\tau(Y - c)$ over $c \in R$ is the τ-th quantile of Y. If the conditional quantile of Y given W is linear in W, then it is the solution to minimizing $E\rho_\tau(Y - a + bW)$ over $a, b \in R$. Consider the special case where X, U, ϵ are independent and normally distributed with mean zero and variances σ_x^2, σ^2 and σ^2 respectively. Then, (Y, W) is bivariate normal, and the conditional distribution of Y given W is normal with mean $\beta\sigma_x^2 W/(\sigma^2 + \sigma_x^2)$ and variance $v_0^2 = \{(\beta_x^2 + \sigma^2)(\beta^2\sigma_x^2 + \sigma^2) - \beta^2\sigma_x^2\}/(\sigma_x^2 + \sigma^2)$. Thus, for the τ-th quantile problem, we obtain $a = \Psi^{-1}(\tau)v_0$, and $b = \beta\sigma_x^2/(\sigma^2 + \sigma_x^2)$. This produces the well-known attenuation for the slope parameter. However, we further note that in general, the conditional quantile of Y given W is not linear in W, so the slope parameter from regressing Y directly on W would result in bias in a more complicated manner.

In the case of least squares estimation for the conditional mean, a number of authors have proposed methods for correction of the measurement error effects. Likelihood arguments of Lindley (1947) and Madansky (1959) lead to a minimization of

$$\sum_i \left(\frac{Y_i - W_i^T b}{\sqrt{1 + |b|^2}} \right)^2 \tag{1.2}$$

for Gaussian errors. A common interpretation of this weighted least squares method is that $(Y_i - W_i^T b)/\sqrt{1 + |b|^2}$ is the orthogonal residual rather than the vertical distance in the regression space. In Section 2, we consider regression quantile estimation for linear EV model by applying the loss function (1.1) to orthogonal residuals. Under some mild conditions, the resulting quantile estimate is consistent and asymptotically normal. We also note that without knowing a parametric form for the error distribution of (ϵ, U), the spherical symmetry is essential for the consistency. The median regression estimates are also compared with the L_2 estimates from (1.2) through a small scale simulation study. These ideas are extended to partly linear models in Section 3, where we adjust for the nonparametric part of the model using an idea of orthogonal projection. It is shown that the quantile estimate for the parametric component attains the same asymptotic efficiency as if the nonparametric component of the model were known. Proofs of the main results in the paper are provided in Section 4.

In the present paper, the identifiability of the EV model is resolved through a classical means by imposing some assumption on the joint error structure. Depending on the nature of the problem in practice, other means of identification might be more appropriate. In some cases, the distribution of the measurement error U may be estimated. In some others, instrumental variables may be available. Further research is clearly needed to identify and analyze appropriate methods of estimating the regression quantile estimates.

2 Linear EV Models

The median regression is better known in the statistical literature as the least absolute deviation regression. In this case, Brown (1982) discussed the approach of estimating covariates X_i to obtain

$$\beta = argmin_{b,x_1,\cdots,x_n} \sum_i \{|Y_i - x_i^T b| + |W_i - x_i|\}, \tag{2.1}$$

and concluded that the procedure will under- or overestimate the slope parameter. In this section, we assume that an intercept term α is in the model in addition to the p dimensional latent variable X: $Y_i = \alpha + X_i^T\beta + \epsilon_i$. We propose to compute the τ-th quantile estimate by minimizing $Q(a,b) = n^{-1}\sum_i \rho_\tau(Y_i - a - W_i^T b)/\sqrt{1+|b|^2}$ over $a \in R$, $b \in R^p$, where $|b|$ denotes the L_2 norm of the vector b.

Note that the loss function ρ_τ is differentiable everywhere except at the point of zero. The directional derivatives of $Q(a,b)$ at the solution $(\hat{\alpha},\hat{\beta})$ are all non-negative, which implies that

$$\sum_i \psi_\tau\Big(\frac{Y_i - a - W_i^T b}{\sqrt{1+|b|^2}}\Big) = O(\#\{h\}),$$

and

$$\sum_i \Big(W_i + \frac{Y_i - a - W_i^T b}{1+|b|^2} b\Big)\psi_\tau\Big(\frac{Y_i - a - W_i^T b}{\sqrt{1+|b|^2}}\Big) = O(\sum_{i\in h} W_i) \tag{2.2}$$

at $(a,b) = (\hat{\alpha},\hat{\beta})$, where $\{i \in h\}$ is the index set for zero residuals. Even though the solution $(\hat{\alpha},\hat{\beta})$ does not satisfy a usual estimating equation exactly, it does so approximately as the number of zero residuals for any linear fit is less than or equal to $p+1$ with probability one, provided that the distribution of (W,Y) is continuous.

The quantile regression can be viewed as a special class of M-estimators. In this direction, several authors have studied their properties mainly from robustness point of view. Zamar (1989) considered orthogonal regression M-estimators based on the idea of minimizing a robust scale. Cheng and Van Ness (1992) derived bounded influence M- and GM-estimators for Gaussian EV models. Such estimators provide some degree of protection against deviation from the Gaussian assumptions. The quantile estimation problem we consider in the present paper differs from the robust M-estimation literature in several ways. For instance, we do not have a central error model (such as Gaussian). The quantiles are of special interest for non-Gaussian models. They are not just alternative methods for the least squares estimation of the conditional mean, but are designed to estimate quantiles directly for their own sake. Besides, the M-estimators with more general loss functions such as those considered in Cheng and Van Ness (1992) are not scale equivariant unless a preliminary scale estimate is available.

To illustrate the method of quantiles, we consider a simple example as follows. We have measurements of the brain weight (in grams) and the body weight (in kilograms) of 28 animals. The data were given in Rousseeuw and Leroy (1987, p. 57). This sample was taken from larger data sets in Jerison (1973). We assume that the conditional quantiles of the log brain weight are linear in the log body weight. We also take the view that the body weights are measured with some error. By assuming that the regression error and the measurement error have a symmetric joint distribution, we computed the 25-th, 50-th and 75-th quantiles, see Figure 1(a). The slopes for the three quartile lines are 0.68, 0.74 and 0.71 respectively. By contrast, if we assume Gaussian homoscedastic errors, the quartiles can be obtained as in Figure 1(b) using parallel lines of slope 0.496. It is clear that a few outliers that do not seem to follow the Gaussian distribution in the regression equation have inflated the spread between quartiles. The regression quantile approach allows for heavier-tailed errors without having to specify it more exactly. Rousseeuw and Leroy (1987) computed a robust estimator of the regression with a slope parameter of 0.75 and an approximate 95% confidence interval of (0.6848, 0.8171). The robust estimate corrected the bias due to the outliers but an exact error distribution (say Gaussian) for the "good" data must be specified to compute quantiles. Besides, the quantiles obtained this way would not be consistent for the population with outliers included.

The rest of the section is devoted to the asymptotic properties of the quantile estimates. From the technical point of view, the quantile estimate involves a non-differentiable score function, and some of the Taylor-type expansions typically used for studying smoothing M-estimators are not *directly* applicable. However, the asymptotic expansions derived by He and Shao (1996) can be used. But first, we state the consistency result for the quantile estimates.

Theorem 2.1: Consider a random sample (W_i, Y_i) from the linear EV model

$$Y_i = \alpha + X_i^T\beta + \epsilon_i, \quad W_i = X_i + U_i \tag{2.3}$$

where the distribution of (ϵ_i, U_i) is spherically symmetric with finite first moment. Assume that q_τ is the unique solution to $E\rho_\tau(\epsilon_i - q) = 0$, and let $\alpha_\tau = \alpha + q_\tau\sqrt{1+|\beta|^2}$. Then the quantile estimate $(\hat{\alpha_\tau}, \hat{\beta}_\tau)$ that minimizes

$$Q(a,b) = n^{-1}\sum_i \rho_\tau\Big(\frac{Y_i - a - W_i^T b}{\sqrt{1+|b|^2}}\Big) \tag{2.4}$$

over (a, b) converges strongly to (α_τ, β).

Note that q_τ is the τ-th quantile of ϵ_i. The key requirement for consistency is that $D(a,b) = E\rho_\tau(\epsilon - U^T b - (a-\alpha) - X^T(b-\beta))/\sqrt{1+|b|^2}$ has a unique minimum at $(a,b) = (\alpha + q_\tau\sqrt{1+|\beta|^2}, \beta)$. This is true if $(\epsilon - U^T b)/\sqrt{1+|b|^2}$ has the same distribution for all b, which is implied by our assumption that (ϵ, U) is spherically symmetric.

On the other hand, if $D_1(b) = E\rho_\tau(\epsilon - U^T b)/\sqrt{1+|b|^2}$ is not constant, but achieves its minimum at some $b = b_1$ different from β, then the minimum of $D(a,b)$ cannot be attained at the true parameters for all distributions of X. Let $d = D_1(\beta) - D_1(b_1) > 0$. When the dispersion of X is sufficiently small relative to that of $\epsilon - U^T b_1$, we would have $D(\alpha, \beta) = D_1(\beta) = D_1(b_1) - d > D(\alpha, b_1)$.

Under the consistency framework, the quantile estimate has a Gaussian limiting distribution. Let f be the density of ϵ.

Theorem 2.2: Under the conditions of Theorem 2.1, we further assume that $E(X) = 0$, $\Sigma_x = E(XX^T)$ is positive definite, $f(e + q_\tau) - f(q_\tau) = O(e^{1/2})$ as $e \to 0$, and $E\epsilon^2 < \infty$. Then,

$$\sqrt{n}(\hat{\alpha}_\tau - \alpha_\tau) \to N(0, \tau(1-\tau)f^{-2}(q_\tau)),$$

$$\sqrt{n}(\hat{\beta}_\tau - \beta) \to N(0, \Sigma_\beta)$$

in distribution, where $\Sigma_\beta = f^{-2}(q_\tau)(1+|\beta|^2)\Sigma_x^{-1}Q\Sigma_x^{-1}$ with $\xi = (\epsilon - U^T\beta)/\sqrt{1+|\beta|^2} - q_\tau$, and

$$Q = \tau(1-\tau)\Sigma_x + Cov\{\psi_\tau(\xi)(U + \xi\beta/\sqrt{1+|\beta|^2})\}. \tag{2.5}$$

If there were no measurement errors, the second term of (2.5) would be absent. We can view the second part of Q as the additional uncertainty on the slope estimate due to measurement errors. If (ϵ, U) is multivariate normal, the expression of Q simplifies to

$$\mathcal{Q} = \tau(1-\tau)\Sigma_x + E\{UU^T\psi_\tau^2(\xi)\} - \frac{E\{\xi^2\psi_\tau^2(\xi)\}}{1+|\beta|^2}\beta\beta^T.$$

A consistent estimate of the intercept α can be obtained by $\frac{1}{2}(\hat{\alpha}_\tau + \hat{\alpha}_{1-\tau})$. The quantity q_τ can be estimated using $\hat{q}_\tau = \frac{\hat{\alpha}_\tau - \hat{\alpha}_{1-\tau}}{2\sqrt{1+|\hat{\beta}_\tau|^2}}$. At $\tau = .5$, we have $\alpha_\tau = \alpha$ and $q_\tau = 0$.

To gain some understanding of how well the proposed estimator works, we run a small Monte Carlo experiment. We draw samples of size n=100 from the following model $Y_i = \alpha + \beta X_i + \epsilon_i$ and $W_i = X_i + U_i$, where X_i is uniformly distributed in $(0, \sqrt{12})$, and (ϵ_i, U_i) is either standard bivariate normal or has a bivariate t-distribution with 3 degrees of freedom. The mean squares errors for the median estimates of both α and β are estimated from 500 runs. The estimates are compared with the maximum likelihood estimates under Gaussian errors. Some results are given in Table 1. Note that the conditional mean and the median for the model are the same, so the comparison can be made on the same scale.

Table 1: MSE for the L_1 and L_2 Estimates

β	L_2 Estimates (a)	L_1 Estimates (a)	L_2 Estimates (b)	L_1 Estimates (b)
1	0.0230, 0.0056	0.035, 0.0089	2.2177, 0.7104	0.4742,0.1396
2	0.0495, 0.0120	0.0846, 0.0207	2.2936, 0.7364	0.7132, 0.2071
10	1.0887, 0.2768	1.6943, 0.4188	13.745, 3.6205	8.8325, 2.2509

(a) refers to bivariate normal error, and (b) for bivariate t_3. The two numbers in each entry are the MSE's for the intercept and slope estimates respectively.

When the error distribution is normal, the median regression estimate has a relative efficiency of above 60% for all three β values considered. In the case of $t(3)$ as errors, the relative efficiency moves above 1. The deficiency for the least squares based estimate is higher for smaller β. When $\beta = 1$, the median estimate is about 5 times more efficient. This type of comparative results are typical. The median estimate has the desirable property of robustness. We also wish to add that the objective functions for both approaches are nonlinear and non-convex, so the computational complexity for finding the estimates are similar.

3 Partially Linear EV Models

Partially linear models drew a lot of attention in the 80's due to their flexibility in incorporating nonparametric relationship for some covariates while keeping the simplicity of linear regression on other variables. Engle, Granger, Rice and Weiss (1986) provided a good example for such semiparametric models in studying the relation between weather and electricity sales. Heckman (1986), Speckman (1988), Chen (1988), and He and Shi (1996) considered the asymptotics of partially linear models. They show that the parameters in the linear component can be estimated as efficiently as if the nonparametric component were known. Cuzick (1992a,b) considered adaptive estimation to achieve efficiency when the error distribution is unknown. Liang and Cheng (1993) provided some results on the second-order efficiency.

In this section, we consider the quantile estimate of the slope parameter β in the partially linear EV model

$$Y_i = X_i^T \beta + g(T_i) + \epsilon_i, \quad W_i = X_i + U_i \tag{3.1}$$

under the same structure as in Section 2 except that a nonparametric relation $g(T)$ enters the model additively. The intercept term is absorbed in g. We assume that T is an observable variable defined on $[0,1]$, and

Assumption 3.1. $E(X|T=t) = 0$ for all $t \in [0,1]$.

A sufficient condition for the assumption 3.1 is independence of T and X. The Gaussian likelihood based estimator for partially linear models has been recently studied by Liang, Härdle and Carroll (1997). The following projection operation is useful and defined first.

Let $\omega_{ni}(t) = \omega_{ni}(t; T_1, \ldots, T_n)$ be probability weight functions depending only on the design points $T_1, \ldots, T_n$. For any sequence of variables or functions $(S_1, \ldots, S_n)$, we define $\mathbf{S}^T = (S_1, \ldots, S_n)$, $\widetilde{S}_i = S_i - \sum_{j=1}^n \omega_{nj}(T_i) S_j$, and $\widetilde{\mathbf{S}}^T = (\widetilde{S}_1, \ldots, \widetilde{S}_n)$. The conversion from $\mathbf{S}$ to $\widetilde{\mathbf{S}}$ will be applied to W_i and Y_i.

Choices of the weight function $\omega_{ni}(t)$ will be made clear later. The estimator $\widetilde{\beta}_n$ we consider minimizes

$$\sum_i \rho_\tau \Big(\frac{\widetilde{Y}_i - \widetilde{W}_i^T b}{1 + |b|^2} \Big) \tag{3.1}$$

over $b \in R^p$. We suppress the dependence of $\widetilde{\beta}_n$ on τ in this section.

The weights are assumed to satisfy the following Assumption 3.2. They are essentially the same as Assumption 1.3 of Liang et al. (1997).

Assumption 3.2. Weight functions $\omega_{ni}(\cdot)$ satisfy:

$$(i) \qquad \sum_{j=1}^{n} \omega_{ni}(T_j) = 1, \quad for\ any\ i,$$

$$(ii) \qquad \max_{1 \le i,j \le n} \omega_{ni}(T_j) = O(b_n),$$

$$(iii) \qquad \max_{1 \le i \le n} \sum_{j=1}^{n} \omega_{nj}(T_i) I(|T_j - T_i| > c_n) = O(c_n),$$

for some $b_n = o(1)$ and $c_n = \log n/(nb_n)$.

Note that Assumption 3.2(iii) follows from (ii) if T_i are uniformly spaced on $[0,1]$. We now state the main result for partly linear models. Recall that spherical symmetry of the error distributions (ϵ, U) is always assumed as in Section 2.

Theorem 3.1: Suppose that g is Lipschitz, and for some $1 > \delta > 0$, $E\epsilon^{2+\delta} < \infty$ and $E|X|^{2+\delta} < \infty$. Under Assumption 3.1 and Assumption 3.2 with $b_n = n^{-(3-\delta_1)/4}$ and $\delta_1 = \delta/(2+\delta)$, $\tilde{\beta}_n$ is a consistent estimate of β, and

$$\sqrt{n}(\tilde{\beta}_n - \beta) \to N(0, \Sigma_\beta)$$

where the matrix Σ_β is the same as in Theorem 2.2.

Assumption 3.2 can be weakened slightly for the consistency part of the theorem, but we choose not to elaborate. Finite second moments of ϵ and X may be sufficient for asymptotic normality, but our proof requires existence of the $(2+\delta)$-th moments where $\delta > 0$ can be arbitrarily small.

To construct the weigh functions $\omega_{ni}(t)$, we may use a probability kernel K. Let h_n be a sequence of bandwidth parameters that tends to zero as $n \to \infty$. We propose to use

$$\omega_{nj}(t) = K\Big(\frac{t - T_j}{h_n}\Big) \Big/ \Big\{\sum_{i=1}^{n} K\Big(\frac{T_i - T_j}{h_n}\Big)\Big\} \quad 1 \le j \le n. \tag{3.3}$$

This choice can be justified by the following

Proposition 3.1: Suppose that $K(t)$ is a bounded and symmetric probability density function on [-1,1], $h_n = c/(nb_n)$ for some constant c, and the design points T_i are nearly uniform in the sense that $C_1/n \le \min\{|T_i - T_{i-1}|\} \le \max\{|T_i - T_{i-1}|\} \le C_2/n$ for some constants $C_1, C_2 > 0$. Then, the choice (3.3) satisfies assumption 3.2.

The proof of Proposition 3.1 is immediate. A particular example in the above construction is the Nadaraya-Watson kernel $K(t) = (15/16)(1-t^2)^2 I(|t| \le 1)$. Theorem 3.1 suggests using

$h_n = cn^{-(1+\delta_1)/4}$ for some small number $\delta_1 > 0$. Since the objective is to estimate β, our limited experience indicates that the choice of the bandwidth h_n here is not as critical as in direct nonparametric function estimation.

4 Proofs of Main Results

Proof of Theorem 2.1: Note that $Q(a,b)$ converges to $E\rho_\tau(\epsilon_1 - \frac{a-\alpha+X^T(b-\beta)}{\sqrt{1+|b|^2}})$. By the assumptions, this expectation has a unique minimum at $a = \alpha_\tau$ and $b = \beta$. Now consider any subsequence of $(\hat{\alpha}_\tau, \hat{\beta}_\tau)$. It is then easy to show by contradiction that (i) it is bounded, and (ii) any further subsequence that converges has the same limit (α_τ, β).

To see (i), note that if $a/\sqrt{1+|b|^2}$ is unbounded along the sequence, then, so is $Q(a,b)$. If a is bounded, but b is unbounded along the sequence, then $b/\sqrt{1+|b|^2}$ will converge for a further subsequence to, say, b_0 of unit length. This means that along the new subsequence $Q(a,b)$ will converge to $E\rho_\tau(\epsilon_1 - X^T b_0) > E\rho_\tau(\epsilon_1 - q_\tau)$, which leads to a contradiction. Similar arguments show (ii), and the proof is complete.

Proof of Theorem 2.2: Since the quantile estimate satisfies (2.2), we evoke Corollary 2.2 of He and Shao (1996) for M-estimators. One can verify the assumptions needed for the Corollary by setting $r = 1$, $A_n = \lambda_{max}(Q)n$, and $D_n = nf(q_\tau)diag(1, (1+|\beta|^2)^{-1/2}\Sigma_x)$, where $\lambda_{max}(Q)$ denotes the maximum eigenvalue of Q. Furthermore, let $\xi_i = (\epsilon_i - U_i^T\beta)/\sqrt{1+|\beta|^2} - q_\tau$. It then follows that

$$\hat{\alpha}_\tau - \alpha_\tau = \{nf(q_\tau)\}^{-1}\sum_i \psi_\tau(\xi_i) + o(n^{-1/2}),$$

and

$$\hat{\beta}_\tau - \beta = \{nf(q_\tau)(1+|\beta|^2)^{-1/2}\Sigma_x\}^{-1}\sum_i (X_i + U_i + \xi_i\beta/\sqrt{1+|\beta|^2})\psi_\tau(\xi_i) + o(n^{-1/2}).$$

The routine application of the central limit theorem completes the proof of Theorem 2.2 with

$$\begin{aligned} Q &= E\{(X_i + U_i + \xi_i\beta/\sqrt{1+|\beta|^2})(X_i + U_i + \xi_i\beta/\sqrt{1+|\beta|^2})^T\psi_\tau^2(\xi_i)\} \\ &= E\psi_\tau^2(\xi_i)\Sigma_x + Cov\{(U + \xi\beta/\sqrt{1+|\beta|^2})\psi_\tau^2(\xi_i)\}. \end{aligned}$$

The following lemmas are useful for the proof of Theorems 3.1. Lemma 4.1 can be proven using Beinstein's inequality, see Liang et al. (1997, Lemma A4) for a similar proof. Lemma 4.2 can be verified using induction. We skip the details.

Lemma 4.1: For any sequence of independent variables $\{V_k, k = 1, \cdots, n\}$ with mean zero and finite $(2+\delta)$-th moment, and for a set of positive numbers $\{a_{ki},\ k, i = 1, \cdots, n\}$ such that $\sup_{1\le i,k\le n} |a_{ki}| \le n^{-p_1}$ for some $0 < p_1 < 1$ and $\sum_{j=1}^n a_{ji} = O(n^{p_2})$ for some $p_2 \ge \max\{0, 2/(2+\delta) - p_1\}$, it holds that

$$\max_{1\le i\le n} |\sum_{k=1}^n a_{ki} V_k| = O(n^{-(p_1-p_2)/2} \log n).$$

Lemma 4.2: Let $S_k = \sum_{i=1}^k A_i$ and $B_i \ge 0$. Then

$$\min_{1\le k\le n} S_k \max_{1\le k\le n} B_k \le \sum_{i=1}^n A_i B_i \le \max_{1\le k\le n} S_k \max_{1\le k\le n} B_k.$$

Proof of Theorem 3.1: Assume without loss of generality that δ is small so that $\delta_1 < 1/4$. Assumption 3.2(ii) and Lemma 4.1 (using $p_1 = (3-\delta_1)/4$ and $p_2 = 1 - \delta_1 - p_1$) imply that $\max_{1\le i\le n} |\omega_{nk}(T_i) V_k| = O(n^{-(1+\delta_1)/4} \log n)$ when $V_k = \epsilon_k$ or X_k or U_k. Assumption 3.2(iii) implies that $\max_{1\le i\le n} |\tilde{g}(T_i)| = O(c_n) = O(n^{-(1+\delta_1)/4} \log n)$.

For simplicity in notation, let $\tilde{a}_i = \widetilde{W}_i + \frac{\widetilde{Y}_i - \widetilde{W}_i^T \beta}{1+|\beta|^2} \cdot \beta$, $\tilde{b}_i = \psi_\tau\Big(\frac{\widetilde{Y}_i - \widetilde{W}_i^T \beta}{\sqrt{1+|\beta|^2}}\Big)$, $a_i = W_i + \frac{\epsilon_i - U_i^T \beta}{1+|\beta|^2} \cdot \beta$, and $b_i = \psi_\tau\Big(\frac{\epsilon_i - U_i^T \beta}{\sqrt{1+|\beta|^2}}\Big)$. Then it is straightforward to verify that

$$\max_{1\le i\le n} |\tilde{a}_i - a_i| = O(n^{-(1+\delta_1)/4} \log n),$$

$$\max_{1\le i\le n} |\tilde{b}_i - b_i| = O_p(n^{-(1+\delta_1)/4} \log n). \tag{4.1}$$

Note that both a_i and b_i are sequences of i.i.d. variables with mean zero and finite variances. By Lemma 4.2, we have for any β,

$$\sum_i \tilde{a}_i \tilde{b}_i - \sum_i a_i b_i = \sum_i (\tilde{b}_i - b_i)(\tilde{a}_i - a_i) + \sum_i (\tilde{b}_i - b_i) a_i - \sum (\tilde{a}_i - a_i) b_i = o_p(n^{1/2-\delta_1/4}). \tag{4.2}$$

This means that the parameter estimate $\widetilde{\beta}_n$ satisfies

$$\sum_i \Big(W_i + \frac{\epsilon_i - U_i^T \widetilde{\beta}_n}{1+|\widetilde{\beta}_n|^2} \cdot \widetilde{\beta}_n\Big) \psi_\tau\Big(\frac{\epsilon_i - U_i^T \widetilde{\beta}_n}{1+|\widetilde{\beta}_n|^2}\Big) = o_p(n^{1/2-\delta_1/4}).$$

Then the consistency of $\tilde{\beta}_n$ follows from similar arguments used for Theorem 2.1. Moreover, the same arguments for Theorem 2.2 lead to

$$\tilde{\beta}_n - \beta = -(1 + |\beta|^2)^{1/2}(nf(q_\tau)\Sigma_x)^{-1}\sum_i a_i b_i + o_p(n^{-1/2})$$

and the desired result follows.

REFERENCES

Brown, M.L.(1982), Estimating line estimation with error in both variables. *Journal of the American Statistical Association*, **77**, 71-79.

Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995), *Nonlinear Measurement Error Models.* Chapman and Hall, New York.

Chen, H. (1988).Convergence rates for parametric components in a partly linear model. *Annals of Statistics*, **16**, 136-146.

Cheng,C.L. and Ness J.W.V. (1992). Generalized $M-$estimators for errors-in-variables regression. *Annals of Statistics*, **20**, 385-397.

Cuzick, J. (1992a). Semiparametric additive regression. *Journal of the Royal Statistical Society, Series B*, **54**, 831-843.

Cuzick, J. (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Annals of Statistics*, **20**, 1129-1136.

Engle, R. F., Granger, C.W.J., Rice,J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.

Efron, (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, **1**, 93-125.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

He, X. (1997). Quantile curves without crossing. *American Statistician*, **51**, 186-192.

He, X. and Shi, P.D. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Analysis*, **58**, 162 -181.

He, X. and Shao, Q. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Annals of Statistics*, **24**, 2608-2630.

Heckman, N.E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.

Jerison, H.J. (1973). *Evolution of the Brain and Intelligence*, Academic Press, New York.

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

Liang, H. and Cheng, P. (1993). Second order asymptotic efficiency in a partially linear model. *Statistics & Probability Letters*, **18**, 73-84.

Liang, H., Härdle, W. and Carroll, R.J. (1997). Large sample theory in a semiparametric partially linear errors-in-variables model, submitted for publication.

Lindley,D.B.(1947), Regression lines and the functional relationship. *Journal of the Royal Statistical Society, Series B*, **9**, 219-244.

Madansky,A.(1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, **54**, 173-205.

Rousseeuw, P.J., and Leroy, A.M. (1987) *Robust Regression and Outlier Detection*, Wiley, New York.

Severini, T.A. and Wang, W. H. (1992). Generalized profile likelihood and conditionally parametric models, *Annals of Statistics*, **20**, 1768-1802 .

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.

Zamar, R.H.(1989). Robust estimation in the errors-in-variables model. *Biometrika*, **76**, 14-160.