Corrected-loss estimation for quantile regression with covariate measurement errors

By HUIXIA JUDY WANG, LEONARD A. STEFANSKI

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

judy_wang@ncsu.edu   len_stefanski@ncsu.edu

AND ZHONGYI ZHU

Department of Statistics, Fudan University, Shanghai 200433, China

zhuzy@fudan.edu.cn

SUMMARY

We study estimation in quantile regression when covariates are measured with errors. Existing methods require stringent assumptions, such as spherically symmetric joint distribution of the regression and measurement error variables, or linearity of all quantile functions, which restrict model flexibility and complicate computation. In this paper, we develop a new estimation approach based on corrected scores to account for a class of covariate measurement errors in quantile regression. The proposed method is simple to implement. Its validity requires only linearity of the particular quantile function of interest, and it requires no parametric assumptions on the regression error distributions. Finite-sample results demonstrate that the proposed estimators are more efficient than the existing methods in various models considered.

Some key words: Corrected loss function; Laplace distribution; Measurement error; Normal distribution; Quantile regression; Smoothing.

1. INTRODUCTION

In problems relating to econometrics, epidemiology and finance, the covariates of interest are often measured with errors. The measurement error, if ignored, often leads to bias in estimating the mean and quantile functions (Carroll et al., 2006; Wei & Carroll, 2009).

Less attention has been paid to quantile regression than to mean regression with a covariate measurement error. There are two main difficulties for correcting the bias in quantile regression caused by measurement error. First, a parametric regression-error likelihood is usually not specified in quantile regression. Second, unlike the mean, quantiles do not enjoy the additivity property, that is, the quantile of the sum of two random variables is not necessarily the sum of the two marginal quantiles. He & Liang (2000) proposed an estimation procedure that minimizes the quantile loss function of orthogonal residuals. This method assumes that the random errors in the response variable y and the measurement errors in the covariate x are independent and follow the same symmetric distribution. Assuming the existence of an instrumental variable, Hu & Schennach (2008) and Schennach (2008) proposed methods that require nonparametric modelling of densities such as that of y given x, and that of x given the instrumental variable. Wei & Carroll (2009) developed an iterative estimation procedure that requires estimating the

conditional density of y given x via modelling the entire quantile process, and this complicates the computation. In addition, Wei & Carroll's method relies on a strong global assumption, that is, estimation of the τth conditional quantile of y given x depends on the assumption that all the conditional quantiles below the τth are linear. In this paper, we propose a simple and consistent estimation procedure assuming a class of measurement error distributions. The proposed method avoids the symmetry assumption used in He & Liang (2000), and requires estimation only at the quantile of interest.

Whatever the approach taken, one must resolve the identifiability issue in measurement error models. In the proposed method, it is resolved by assuming a parametric form for the measurement error distribution whose parameters such as variance can be estimated. However, we leave the quantile regression error distribution completely unspecified.

We consider the linear quantile regression model

$$Q_\tau(y_j \mid x_j) = x_j^\mathrm{T} \beta_0(\tau) \quad (j = 1, \ldots, n), \tag{1}$$

where $Q_\tau(y_j \mid x_j)$ denotes the τth conditional quantile of the response variable y_j given by the covariate x_j, $\beta_0(\tau) \in \mathbb{R}^p$ is the coefficient vector and $\tau \in (0, 1)$ is the quantile level of interest. Our main interest is in estimating $\beta_0(\tau)$ when x_j is measured with an error. We assume an additive measurement error model, $w_j = x_j + u_j$, relating the surrogate w_j and x_j, where the $u_j \in \mathbb{R}^p$ are the independent and identically distributed measurement errors. Throughout, we assume that u_j is independent of x_j and y_j, and we drop τ in $\beta_0(\tau)$ for notational simplicity.

2. Proposed methods

2·1. *Corrected-loss estimator*

When x_j is measured without an error, β_0 can be estimated consistently by

$$\hat{\beta}_x = \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \sum_{j=1}^n \rho(y_j, x_j, \beta), \tag{2}$$

where $\rho(y, x, \beta) = \rho_\tau(y - x^\mathrm{T}\beta)$, $\rho_\tau(\epsilon) = \epsilon\{\tau - I(\epsilon < 0)\}$ is the quantile loss function and $I(\cdot)$ is the indicator function. The estimator $\hat{\beta}_x$ also satisfies

$$n^{-1} \sum_{j=1}^n \psi(y_j, x_j, \hat{\beta}_x) = o_p(1), \tag{3}$$

where $\psi(y, x, \beta) = x\{I(y - x^\mathrm{T}\beta < 0) - \tau\}$. Under model (1), $\mathrm{pr}(y < x^\mathrm{T}\beta_0 \mid x) = \tau$. Therefore, $E\{\psi(y, x, \beta_0)\} = 0$, and $\psi(y, x, \beta)$ is an unbiased estimating function for β_0.

When x_j is subject to error and we observe only a surrogate w_j, naively replacing x_j with w_j in (2) or (3) usually leads to inconsistent estimators, because $E\{\psi(y, w, \beta_0)\} = 0$ may not hold. To account for the measurement error, we construct corrected score functions of w that are unbiased for β_0 (Stefanski, 1989; Nakamura, 1990). However, in practice, it is challenging to determine the corrected scores, especially in quantile regression, as the quantile loss function $\rho_\tau(\epsilon)$ is not differentiable at $\epsilon = 0$. To overcome this difficulty, we approximate $\rho_\tau(\epsilon)$ by a smooth function $\rho(\epsilon, h)$ depending on a positive smoothing parameter h.

Let E^* denote the expectation with respect to w given y and x. Unless otherwise specified, we use E to denote the global expectation. We aim to find a corrected loss function $\rho^*(y, w, \beta, h)$ such that $E^*\{\rho^*(y, w, \beta, h)\} = \rho(y, x, \beta, h) \to \rho(y, x, \beta)$ pointwise in (y, x, β)

as $h \to 0$. Under some regularity conditions, β_0 is the unique minimizer of $E\{\rho(y, x, \beta)\}$. Therefore, minimizing the sample analog of $E\{\rho^*(y, w, \beta, h)\}$ leads to a consistent estimator of β_0, if h goes to zero at a suitable rate. Motivated by this idea, we define the corrected-loss quantile estimator as

$$\hat{\beta}_w = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{j=1}^{n} \rho^*(y_j, w_j, \beta, h).$$

In the next two subsections, we develop corrected-loss estimators for two measurement error models, normal and Laplace, because these two measurement error distributions provide reasonable error models in many applications. Our simulation study in §3 suggests that the proposed estimators are robust against misspecification of the measurement error distribution. The extension to a wider class of distribution families is discussed in §5.

2·2. *Normal measurement error*

Assume $\{y_j, w_j\}_{j=1}^{n}$ is a random sample with $w_j = x_j + u_j$, where $u_j \sim N(0, \Sigma)$ is a p-dimensional normal random vector that is independent of y_j and x_j; see Fuller (1987) and Carroll et al. (2006) for reviews on normal measurement errors in mean regression models.

We first review a useful result for normal random variables. Suppose that $\epsilon \sim N(\mu, \sigma^2)$ and that $g(\cdot)$ is a sufficiently smooth function. Let $u \sim N(0, 1)$ be independent of ϵ. Stefanski & Cook (1995) showed that $E[E\{g(\epsilon + i\sigma u) \mid \epsilon\}] = g(\mu)$, where $i = \sqrt{-1}$, the outer expectation is with respect to ϵ and the inner one is with respect to u given ϵ.

Motivated by the above result, we propose to approximate the quantile loss function $\rho_\tau(\epsilon)$ by an infinitely smooth function

$$\rho_N(\epsilon, h) = \epsilon \left\{ \tau - 1/2 + G_N(\epsilon/h) \right\},$$

where $G_N(x) = \pi^{-1}\mathrm{Si}(x) = \pi^{-1} \int_0^x \sin(t)/t \, dt$ is the sine integral function, which satisfies $\lim_{x \to \infty} \mathrm{Si}(x) = \pi/2$ and $\lim_{x \to -\infty} \mathrm{Si}(x) = -\pi/2$. By such an approximation, we have the following theorem.

THEOREM 1. *Suppose that* $\epsilon \sim N(\mu, \sigma^2)$. *Define* $A(\epsilon, \sigma^2, h) = E\{\rho_N(\epsilon + i\sigma u, h) \mid \epsilon\}$, *where* $u \sim N(0, 1)$ *is independent of* ϵ. *Then*

(i) $A(\epsilon, \sigma^2, h) = \epsilon(\tau - 1/2) + \pi^{-1} \int_0^{1/h} \{y^{-1}\epsilon \sin(y\epsilon) - \sigma^2 \cos(y\epsilon)\} \exp(y^2\sigma^2/2) \, dy$;
(ii) $E\{A(\epsilon, \sigma^2, h)\} = \rho_N(\mu, h)$.

Since $(y - w^{\mathsf{T}}\beta) \mid (y, x) \sim N(y - x^{\mathsf{T}}\beta, \beta^{\mathsf{T}}\Sigma\beta)$, we define the corrected quantile loss function as

$$\rho_N^*(y, w, \beta, h) = A(y - w^{\mathsf{T}}\beta, \beta^{\mathsf{T}}\Sigma\beta, h).$$

By Theorem 1, $E^*\{\rho_N^*(y, w, \beta, h)\} = \rho_N(y - x^{\mathsf{T}}\beta, h) \doteq \rho_N(y, x, \beta, h) \to \rho(y, x, \beta)$ pointwise in (y, x, β) as $h \to 0$. Let $\mathcal{B}$ denote a compact subset of $\mathbb{R}^p$ that contains β_0. The corrected quantile estimator is then defined as

$$\hat{\beta}_N = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \sum_{j=1}^{n} \rho_N^*(y_j, w_j, \beta, h).$$

In applications, often only one or two covariates are measured with an error. Our proposed method accommodates such scenarios as special cases. Throughout the paper, we let the first

component of x be 1, corresponding to the intercept, so there is no measurement error in the first component. For example, if we assume that only the pth component of x is subject to measurement error $u \sim N(0, \sigma^2)$, then we have

$$\Sigma = \begin{pmatrix} 0_{q \times q} & 0_q \\ 0_q^{\mathrm{T}} & \sigma^2 \end{pmatrix} \quad (q = p - 1),$$

where 0_q and $0_{q \times q}$ denote a q-dimensional vector and a $q \times q$ matrix with zero elements, respectively. Consequently, the corrected quantile loss function becomes $\rho_N^*(y, w, \beta, h) = A(y - w^{\mathrm{T}}\beta, \beta_p^2\sigma^2, h)$, where β_p is the pth element of β. The same parameterization applies to the correction for a Laplace measurement error described in § 2·3.

2·3. *Laplace measurement error*

We consider the situation where the measurement error follows a multivariate Laplace distribution. The Laplace distribution is often used for modelling data with tails heavier than normal. We refer to Stefanski & Carroll (1990), Hong & Tamer (2003), Richardson & Hollinger (2005), Purdom & Holmes (2005), Visscher (2006) and McKenzie et al. (2008) for discussions of Laplace measurement errors. We first introduce a multivariate Laplace distribution adopted from Kotz et al. (2001, Ch. 6), and give a lemma stating some related properties.

DEFINITION 1. *A random vector $X \in \mathbb{R}^p$ has a multivariate asymmetric Laplace distribution if its characteristic function is $\Psi(t) = (1 + t^{\mathrm{T}}\Sigma t/2 - i\mu^{\mathrm{T}}t)^{-1}$ for $t \in \mathbb{R}^p$, where $\mu \in \mathbb{R}^p$ and Σ is a $p \times p$ nonnegative definite symmetric matrix. In the following, we write $X \sim \mathrm{AL}_p(\mu, \Sigma)$. If $\mu = 0$, then $\mathrm{AL}_p(0, \Sigma)$ corresponds to a symmetric multivariate Laplace distribution. In addition, $\mathrm{AL}_1(0, \sigma^2)$ is the classical univariate Laplace (1774) distribution $L(\mu, \sigma^2)$ if and only if $\mu = 0$.*

LEMMA 1. *Let $X \sim \mathrm{AL}_p(\mu, \Sigma)$. Then*

(i) *the mean and covariance matrix of X are $E(X) = \mu$, and $\mathrm{cov}(X) = \Sigma + \mu\mu^{\mathrm{T}}$;*
(ii) *if $\mu = 0$, then for any constant a and vector $b \in \mathbb{R}^p$, the random variable $a + b^{\mathrm{T}}X \sim L(a, \sigma^2)$, where $\sigma^2 = b^{\mathrm{T}}\Sigma b$, and $L(a, \sigma^2)$ is the standard univariate Laplace distribution with mean a and variance σ^2.*

Suppose that the measured covariates are $w_j = x_j + u_j$, where $u_j \sim \mathrm{AL}_p(0, \Sigma)$, independent of x_j and y_j. Our corrected loss function is based on the following theorem.

THEOREM 2. *Suppose that the random variable ϵ follows the univariate Laplace distribution $L(\mu, \sigma^2)$. If $g(\epsilon)$ is a twice-differentiable function of ϵ, then*

$$E\{g(\epsilon) - (\sigma^2/2)g^{(2)}(\epsilon)\} = g(\mu),$$

where $g^{(2)}(\epsilon)$ is the second derivative of $g(\epsilon)$.

Let $K(\cdot)$ denote a kernel density function and define $G_L(x) = \int_{u < x} K(u)\,du$. In our numerical studies, we choose $K(\cdot)$ as the probability density function of $N(0, 1)$. We consider the smoothed quantile loss function $\rho_L(\epsilon, h) = \epsilon\{\tau - 1 + G_L(\epsilon/h)\}$. For Laplace measurement error, by Lemma 1, $(y - w^{\mathrm{T}}\beta) \mid (y, x) \sim L(y - x^{\mathrm{T}}\beta, \sigma^2)$, where $\sigma^2 = \beta^{\mathrm{T}}\Sigma\beta$. Let $\epsilon = y - w^{\mathrm{T}}\beta$.

Define the corrected quantile loss function as

$$\rho_L^*(y, w, \beta, h) = \rho_L(\epsilon, h) - \frac{\sigma^2}{2} \frac{\partial^2 \rho_L(\epsilon, h)}{\partial \epsilon^2}$$

$$= \epsilon(\tau - 1) + \epsilon G_L\left(\frac{\epsilon}{h}\right) - \frac{\sigma^2}{2}\left\{\frac{2}{h}K\left(\frac{\epsilon}{h}\right) + \frac{\epsilon}{h^2}K'\left(\frac{\epsilon}{h}\right)\right\}. \tag{4}$$

By Theorem 2, $E^*\{\rho_L^*(y, w, \beta, h)\} = \rho_L(y - x^{\mathrm{T}}\beta, h) \doteq \rho_L(y, x, \beta, h) \to \rho(y, x, \beta)$ pointwise in (y, x, β) as $h \to 0$.

The corrected quantile estimator is therefore defined as

$$\hat{\beta}_L = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \sum_{j=1}^{n} \rho_L^*(y_j, w_j, \beta, h).$$

2·4. *Large sample properties*

To establish the asymptotic results in this paper, we make the following assumptions.

Assumption 1. The samples $\{(y_j, x_j) : j = 1, \ldots, n\}$ are independent and identically distributed.

Assumption 2. The vector β_0 is an interior point of the parameter space $\mathcal{B}$, a compact subset of $\mathbb{R}^p$.

Assumption 3. The expectation $E(\|x_j\|^2)$ is bounded, and $E(x_j x_j^{\mathrm{T}})$ is a positive definite $p \times p$ matrix.

Assumption 4. Let $e_j = y_j - x_j^{\mathrm{T}}\beta(\tau)$. The conditional density of e_j, $f_j(e_j \mid x_j)$, is bounded from infinity, and it is bounded away from zero and has a bounded first derivative in the neighbourhood of zero.

Assumption 5. For each j, $E(e_j^2 \mid x_j)$ is bounded as a function of τ.

Assumption 6. The kernel function $K(u)$ is a bounded probability density function having finite fourth moment and is symmetric about the origin. In addition, $K(u)$ is twice-differentiable, and its second derivative $K^{(2)}(u)$ is bounded and Lipschitz continuous on $(-\infty, \infty)$.

Theorem 3 states the strong consistency of the proposed estimators for normal and Laplace measurement errors, respectively.

THEOREM 3. (i) *Suppose that the measurement error* $u_j \sim N(0, \Sigma)$, *that Assumptions* 1–5 *hold, and that* $h \to 0$ *and* $h = c(\log n)^{-\delta}$, *where* $\delta < 1/2$ *and* c *is some positive constant. Then* $\hat{\beta}_N \to \beta_0$ *almost surely as* $n \to \infty$. (ii) *If the measurement error* $u_j \sim \mathrm{AL}_p(0, \Sigma)$ *and Assumptions* 1–4 *and* 6 *hold,* $h \to 0$, *and* $(nh)^{-1/2} \log n \to 0$, *then* $\hat{\beta}_L \to \beta_0$ *almost surely as* $n \to \infty$.

Assumption 2 ensures the existence of $\hat{\beta}_N$ and $\hat{\beta}_L$, and the uniformity of the convergence of the minimand over $\mathcal{B}$, as required to prove the consistency. Assumptions 3 and 4 ensure that β_0 is the unique minimizer of $E\{\rho(y, x, \beta)\}$. With normal measurement error, because the corrected quantile loss function $\rho_N^*(\cdot)$ is complicated, Assumption 5 is used in the Appendix to bound the first-order expansion of $\rho_N^*(\cdot)$ uniformly over e_j. Assumption 5 is not needed for the Laplace measurement error. Assumption 6 specifies the conditions on the kernel function used in $\hat{\beta}_L$

for the Laplace measurement error. In Theorem 3, the rate of h differs for normal and Laplace measurement errors. This difference is related to the smoothness of the measurement error distribution. It is well known in the deconvolution literature that the rates of convergence are lower for smoother error distributions (Carroll & Hall, 1988; Fan, 1992).

We next establish asymptotic normality of the proposed estimators. For notational simplicity, let $\hat{\beta}$ denote the proposed corrected estimator, and $\rho^*(y, w, \beta, h)$ denote the corrected quantile loss function, for either normal or Laplace measurement errors. We make the following additional assumption.

Assumption 7. Let $\psi_1^*(y, w, \beta, h) = \partial \rho^*(y, w, \beta, h)/\partial \beta$ and $\psi_2^*(y, w, \beta, h) = \partial^2 \rho^*(y, w, \beta, h)/\partial \beta \partial \beta^{\mathrm{T}}$. As $n \to \infty$ and $h \to 0$, there exist positive definite matrices D and A such that $E\{\psi_1^*(y, w, \beta_0, h)^{\otimes 2}\} \to D$ and $E\{\psi_2^*(y, w, \beta_0, h)\} \to A$.

THEOREM 4. *Suppose that Assumptions* 1–7 *hold, and* $\hat{\beta}$ *is the consistent estimator of* β_0, *either* $\hat{\beta}_N$ *or* $\hat{\beta}_L$ *defined in* §§2·3 *and* 2·4. *Then* $n^{1/2}(\hat{\beta} - \beta_0) \to N(0, A^{-1}DA^{-1})$ *in distribution, as* $n \to \infty$.

2·5. *Estimated measurement error covariance matrix*

Thus far we have described our method under the assumption that the covariance matrix Σ is known. Applications where Σ is known exist, but are rare. The more common scenario is one in which an unbiased estimate, $\hat{\Sigma}$, is available. Analysis then proceeds using $\hat{\Sigma}$ as a plug-in estimator of Σ. A common design where this strategy is used is when each w_j is itself the average of m replicate measurements $w_{j,k}$ ($k = 1, \ldots, m$), each having variance $\Gamma = m\Sigma$. A consistent and unbiased estimator of Σ is $\hat{\Sigma} = \hat{\Gamma}/m$, where

$$\hat{\Gamma} = \{n(m-1)\}^{-1} \sum_{j=1}^{n} \sum_{k=1}^{m} (w_{j,k} - w_j)(w_{j,k} - w_j)^{\mathrm{T}}$$

is based on $n(m-1)$ degrees of freedom; see Liang et al. (2007). The application data in §4 have this structure with $m = 6$ in which case $\hat{\Sigma}$ is estimated on $5n$ degrees of freedom. In the Monte Carlo study in §3, we simulate this situation with $m = 2$.

Let σ be a q-dimensional vector consisting of the elements of the upper triangle of Σ including the diagonals, where $q = p(p+1)/2$. To reflect the dependence on σ, we let $\rho^*(y, w, \beta, h, \sigma)$ denote the corrected quantile loss function, for either normal or Laplace measurement errors. We next establish the asymptotic properties of the corrected estimator,

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\mathrm{argmin}} \sum_{j=1}^{n} \rho^*(y_j, w_j, \beta, h, \hat{\sigma}), \tag{5}$$

where $w_j = m^{-1} \sum_{k=1}^{m} w_{j,k}$. Let S_j be a q-vector consisting of the elements of the upper triangle of the matrix $\sum_{k=1}^{m}(w_{j,k} - w_j)(w_{j,k} - w_j)^{\mathrm{T}}$ including the diagonals. Define $\psi_{1\beta}^*(y, w, \beta, h, \sigma) = \partial \rho^*(y, w, \beta, h, \sigma)/\partial \beta$, $\psi_{2\beta}^*(y, w, \beta, h, \sigma) = \partial \psi_{1\beta}^*(y, w, \beta, h, \sigma)/\partial \beta$ and $\psi_{2\sigma}^*(y, w, \beta, h, \sigma) = \partial \psi_{1\beta}^*(y, w, \beta, h, \sigma)/\partial \sigma$. We replace Assumption 7 with the following Assumption 7′.

Assumption 7′. As $n \to \infty$ and $h \to 0$, $E\{\psi_{1\beta}^*(y, w, \beta_0, h, \sigma)^{\otimes 2}\} \to D$, $E\{\psi_{2\beta}^*(y, w, \beta_0, h, \sigma)\} \to A$, $E\{\psi_{2\sigma}^*(y, w, \beta_0, h, \sigma)\} \to B$, $E[\psi_{1\beta}^*(y_j, w_j, \beta_0, h, \sigma)\{S_j - m(m-1)\sigma\}^{\mathrm{T}}] \to C$, where A and D are $p \times p$-positive definite matrices, and B and C are $p \times q$ matrices.

THEOREM 5. *Under the conditions of Theorem 3 and Assumption 7′, the estimator $\hat{\beta}$ given in (5) is consistent and asymptotically normal with covariance matrix $A^{-1}D^*A^{-1}$, where $D^* = D + \{m(m-1)\}^{-2}BE[\{S_j - m(m-1)\sigma\}^{\otimes 2}]B^{\mathrm{T}} + \{m(m-1)\}^{-1}(CB^{\mathrm{T}} + BC^{\mathrm{T}})$.*

Remark 1. Compared with Theorem 4, the covariance of $\hat{\beta}$ has three additional terms due to the variation in the estimated measurement error variance. For normal measurement error, (y_j, w_j) are independent of $\hat{\Gamma}$, so the last two terms of D^* reduce to zero.

2·6. *Some computational issues*

Motivated by the method of Delaigle & Hall (2008), we propose a modified simulation-extrapolation-type strategy to choose the smoothing parameter h. The simulation and extrapolation method was introduced by Stefanski & Cook (1995) for estimation in a parametric setting; see also Stefanski (2000), Luo et al. (2006) among others. Delaigle & Hall (2008) showed how this strategy can be adapted to choose the smoothing parameter in nonparametric modelling.

Let $\hat{\beta}(h)$ be the corrected-loss estimator associated with smoothing parameter h. Define $M(h) = E[\{\hat{\beta}(h) - \beta_0\}^{\mathrm{T}}\Omega^{-1}\{\hat{\beta}(h) - \beta_0\}]$ as the mean squared error of $\hat{\beta}(h)$, where $\Omega = \mathrm{cov}\{\hat{\beta}(h)\}$. Ideally, we would like to find the optimal smoothing parameter $h_0 = \mathrm{argmin}\, M(h)$. However, since $M(h)$ depends on the unknown x_j, the minimization of $M(h)$ cannot be executed in practice. Instead, we develop two versions of $M(h)$ by simulating higher levels of measurement errors. Let $u_{b1}^*, \ldots, u_{bn}^*$ and $u_{b1}^{**}, \ldots, u_{bn}^{**}$ ($b = 1, \ldots, N_b$) denote independent and identically distributed random vectors from $N(0, \Sigma)$ for the normal measurement error or from $\mathrm{AL}_p(0, \Sigma)$ for Laplace measurement error depending on the error model assumed. Let $w_{bj}^* = w_j + u_{bj}^*$, $w_{bj}^{**} = w_{bj}^* + u_{bj}^{**}$, $\beta_b^*(h)$ and $\beta_b^{**}(h)$ as the corrected-loss estimators based on samples (y_j, w_{bj}^*) and (y_j, w_{bj}^{**}), respectively. Define

$$M_1(h) = N_b^{-1} \sum_{b=1}^{N_b} \{\beta_b^*(h) - \hat{\beta}(h)\}^{\mathrm{T}}(S^*)^{-1}\{\beta_b^*(h) - \hat{\beta}(h)\},$$

$$M_2(h) = N_b^{-1} \sum_{b=1}^{N_b} \{\beta_b^{**}(h) - \beta_b^*(h)\}^{\mathrm{T}}(S^{**})^{-1}\{\beta_b^{**}(h) - \beta_b^*(h)\},$$

where S^* and S^{**} are the sample covariance matrices of $\{\beta_b^*(h) - \hat{\beta}(h) : b = 1, \ldots, N_b\}$ and $\{\beta_b^{**}(h) - \beta_b^*(h) : b = 1, \ldots, N_b\}$, respectively. Let $\hat{h}_1 = \mathrm{argmin}_h\, M_1(h)$ and $\hat{h}_2 = \mathrm{argmin}_h\, M_2(h)$. Since w_{bj}^{**} measures w_{bj}^* in the same way that w_{bj}^* measures w_j, it is reasonable to expect that the relationship between $\hat{h}_2$ and $\hat{h}_1$ is similar to that between $\hat{h}_1$ and h_0. Therefore, back extrapolation can be used to approximate h_0. In our implementation, we use the linear extrapolation from the pair $(\log \hat{h}_1, \log \hat{h}_2)$ and define the second-order approximation to h_0 as $\hat{h} = \hat{h}_1^2/\hat{h}_2$.

For corrected-loss approaches, one computational complication is that, in finite samples, the corrected objective function may not be globally convex in β; see also Stefanski (1989), Stefanski & Carroll (1985, 1987) and Nakamura (1990) for similar observations. If x_j is measured with a Laplace error, then $n^{-1}\sum_j \rho_L^*(y_j - w_j^{\mathrm{T}}\beta) \to -\infty$ or $+\infty$ when $\sigma^2 = \beta^{\mathrm{T}}\Gamma\beta \to \infty$, depending on the sign of the last term in brackets in (4). In such a case, the corrected objective function has no global minimum. However, it is locally convex around a local minimizer $\hat{\beta}$ that is the desired corrected-loss estimate. In our work, when solving the minimization problem for $\hat{\beta}$, we adopted the common strategy of starting from the naive estimator obtained by regressing

y_j on w_j, and then searched using the R (R Development Core Team, 2012) function optim with default options. This algorithm worked well in numerical studies.

3. Simulation study

We conduct a simulation study to investigate the performance of the proposed corrected-loss approaches. The data were generated from the model

$$y_j = 1 + x_j + (1 + \eta x_j)e_j \quad (j = 1, \dots, 200),$$

where $e_j \sim N(0, \sigma_e^2)$. Under the above model, the τth conditional quantile of y given x is $\beta_{01}(\tau) + \beta_{02}(\tau)x$ with $\beta_{01}(\tau) = 1 + \sigma_e\Phi^{-1}(\tau)$ and $\beta_{02}(\tau) = 1 + \eta\sigma_e\Phi^{-1}(\tau)$, and $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$. We further assume that the x_j are subject to measurement error following the model

$$w_j = x_j + u_j, \quad x_j \sim U(5, 5 + 12^{1/2}).$$

We consider four different cases. The measurement errors u_j are generated from $N(0, \sigma_u^2)$ in Cases 1–2, from $L(0, \sigma_u^2)$ in Case 3, and from the normalized χ_3^2 with mean zero and variance σ_u^2 in Case 4. We set $\eta = 0$ in Case 1, corresponding to a homoscedastic model, and $\eta = 0.2$ in Cases 2–4, corresponding to heteroscedastic models. We give e_j and u_j standard deviations $\sigma_e = \sigma_u = 0.5$, so the assumption required by He & Liang's method is satisfied in Case 1. In Cases 2–4 with heteroscedasticity, the variances of the regression errors depends on x_j and thus are on different scales with the measurement error.

For each case, 100 simulations are performed. Focusing on $\tau = 0.5$ and $\tau = 0.75$, we compare five estimators, including the naive estimator obtained from regressing y_j on w_j, He & Liang's estimator, the proposed corrected-loss estimator for normal measurement error, the proposed corrected-loss estimator for Laplace measurement error and Wei & Carroll's estimator obtained using the R program developed by Wei and Carroll with 20 iterations.

To make a fair comparison, in the implementation of He & Liang's method, we first transform y_i to $y_i^* = \lambda y$ with $\lambda = [E\{(1 + \eta x_j)^2\}]^{1/2}\sigma_e/\sigma_u$ to match the marginal variance of regression error with the measurement error variance. The resulting coefficient estimates are then transformed back to the original scale. For the proposed corrected-loss estimators and Wei & Carroll's approach, we generated an independent estimate $\hat{\sigma}_u^2$ of σ_u^2 based on n degrees of freedom as explained in §2·5 for each dataset. This simulates the situation in which each w_j is the average of two replicate measurements $w_{j,k} \sim N(x_j, \gamma_u^2)$ or $L(x_j, \gamma_u^2)$ $(k = 1, 2)$ with $\gamma_u^2 = 2\sigma_u^2$.

In the implementation of the two proposed methods, we choose the smoothing parameter h following the simulation and extrapolation procedure in §2·6 with $N_b = 20$. In Case 1, the mean $\hat{h}$ for the corrected-loss estimators for normal and Laplace errors are 2·82 and 2·29 at $\tau = 0.5$, and 1·04 and 1·09 at $\tau = 0.75$, respectively.

Figure 1 presents boxplots of $\hat{\beta}_k(\tau) - \beta_{0k}(\tau)$ $(k = 1, 2)$ at $\tau = 0.75$ from the five approaches. We omit the boxplots at $\tau = 0.5$ since the main observations are similar to those at $\tau = 0.75$. As expected, the naive estimator is seriously biased under all scenarios. He & Liang's estimator performs well in Case 1 when e_j and u_j have the same distribution, but has considerable bias in Cases 2–4 with heteroscedastic regression errors for estimation at $\tau = 0.75$. The two proposed estimators and Wei & Carroll's estimator successfully correct the bias for both homoscedastic and heteroscedastic models. Even though the proposed methods require parametric assumptions on the measurement error distribution, they are quite robust against model misspecification. The two methods perform very well not only in Cases 1–3 when the normal error assumption is used
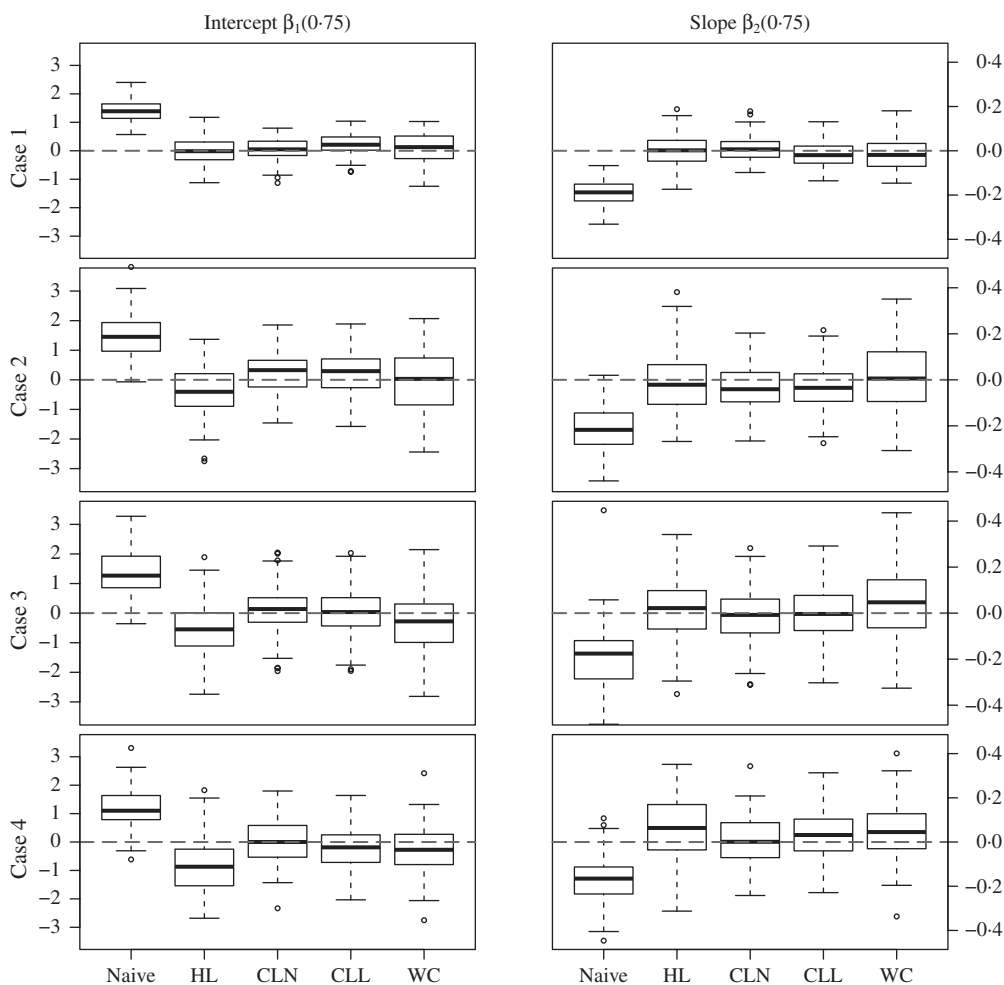
Fig. 1. Boxplots of $\hat{\beta}_k(\tau) - \beta_{0k}(\tau)$, $k = 1, 2$ for different methods in Cases 1–4 at $\tau = 0.75$. Naive, the naive method by regressing y_j on w_j; HL, He & Liang's method; CLN, corrected-loss estimator for normal measurement error; CLL, corrected-loss estimator for Laplace measurement error; WC, Wei & Carroll's method.

for the Laplace measurement error and vice versa, but also in Case 4 when the measurement error distribution is substantially right skewed.

For detailed comparison, Table 1 summarizes the mean squared errors of the different estimators. The two proposed corrected-loss estimators are more efficient than Wei & Carroll's estimator in all cases. In addition, since Wei & Carroll's estimator requires estimation of the whole quantile process simultaneously, it is computationally much more expensive than the proposed estimators when the focus is on one or a few quantile levels. The normal corrected-loss estimator is slower than the Laplace corrected-loss estimator, as the corrected loss function $\rho_N^*(\cdot)$ involves an integral that has no closed form and thus requires numerical integration. For one simulated dataset in Case 2 with $n = 200$, using R version 2·8·1 on a 3 GHz Dell computer, estimation at the median required 9·7 seconds for the Laplace corrected-loss estimator, 496 seconds for the normal corrected-loss estimator, and it took 1020 seconds for Wei & Carroll's estimator to obtain estimates at 39 quantile levels. The number of quantile levels required by Wei & Carroll's estimator grows with the sample size n, and thus the computation is even more challenging for larger datasets.

Table 1. *Mean squared errors of different estimators of the intercept $\beta_1(\tau)$ and slope $\beta_2(\tau)$ parameters. The values in the parentheses are the Monte Carlo standard deviations*

| | $100 \times \mathrm{MSE}\{\hat{\beta}_1(\tau)\}$ | | | | | $100 \times \mathrm{MSE}\{\hat{\beta}_2(\tau)\}$ | | | | |
	Naive	HL	CLN	CLL	WC	Naive	HL	CLN	CLL	WC
					Case 1					
$\tau = 0.5$	179	20	15	15	18	4·0	0·4	0·3	0·3	0·4
	(8)	(4)	(2)	(2)	(3)	(0·2)	(0·1)	(0·1)	(0·0)	(0·1)
$\tau = 0.75$	201	23	16	19	28	3·8	0·5	0·3	0·3	0·6
	(10)	(3)	(2)	(2)	(4)	(0·2)	(0·1)	(0·1)	(0·0)	(0·1)
					Case 2					
$\tau = 0.5$	192	64	47	46	67	4·4	1·5	1·1	1·1	1·6
	(16)	(9)	(6)	(6)	(9)	(0·4)	(0·2)	(0·1)	(0·1)	(0·2)
$\tau = 0.75$	275	78	58	56	105	5·8	1·5	1·1	1·2	2·3
	(24)	(12)	(7)	(7)	(12)	(0·5)	(0·2)	(0·1)	(0·1)	(0·3)
					Case 3					
$\tau = 0.5$	192	76	74	65	104	4·3	1·8	1·7	1·5	2·4
	(19)	(11)	(12)	(10)	(19)	(0·4)	(0·3)	(0·3)	(0·2)	(0·5)
$\tau = 0.75$	250	101	63	63	132	5·2	1·7	1·3	1·3	2·9
	(24)	(14)	(10)	(9)	(21)	(0·5)	(0·3)	(0·2)	(0·2)	(0·5)
					Case 4					
$\tau = 0.5$	205	69	56	53	87	4·6	1·6	1·3	1·2	1·9
	(19)	(10)	(7)	(8)	(14)	(0·4)	(0·2)	(0·2)	(0·2)	(0·3)
$\tau = 0.75$	194	157	59	58	75	4·0	2·5	1·2	1·2	1·8
	(19)	(17)	(8)	(8)	(13)	(0·4)	(0·3)	(0·2)	(0·2)	(0·3)

Naive, the naive method by regressing y on w; HL, He & Liang's method; CLN, corrected-loss estimator for normal measurement error; CLL, corrected-loss estimator for Laplace measurement error; WC, Wei & Carroll's method.

Table 2. *Coverage probabilities, %, of bootstrap confidence intervals with a nominal level of 95%*

| | $\beta_1(0.5)$ | | $\beta_2(0.5)$ | | $\beta_1(0.75)$ | | $\beta_2(0.75)$ | |
	CLN	CLL	CLN	CLL	CLN	CLL	CLN	CLL
Case 1	97	95	97	95	96	92	96	92
Case 2	95	96	95	96	91	94	91	94
Case 3	91	94	91	94	91	92	91	92
Case 4	98	95	98	95	92	92	92	92

CLN, the corrected-loss estimator for normal measurement error; CLL, the corrected-loss estimator for Laplace measurement error.

In quantile regression, it is challenging to estimate the asymptotic covariance of the quantile coefficients directly, as the covariance matrix involves unknown density functions that are difficult to estimate in finite samples. For practical implementation, we adopt a simple bootstrap approach through resampling (y_j, w_j) with replacement. To accommodate the variation in the estimation of σ_u, for each bootstrap, we obtain the proposed estimators by using the estimated σ_u calculated with the bootstrap sample of the internal replicates $w_{j,k}$. Bootstrap confidence intervals can be constructed by using the bootstrap standard error and the asymptotic normality of the proposed estimators. In each simulation run, 200 bootstrap samples are used to obtain the confidence intervals. Table 2 summarizes the coverage probabilities of 95% confidence intervals from the two proposed estimators. The bootstrap approach performs reasonably well. The confidence intervals of the proposed methods have empirical coverage probabilities close to the nominal level 95% even in cases where the parametric measurement error distribution is misspecified.

4. APPLICATION TO A DIETARY DATA

For illustration, we analyse a dietary dataset from the Women's Interview Survey of Health. These data are from 271 subjects, each completing a food frequency questionnaire and six 24-hour food recalls on randomly selected days. The food frequency questionnaire is a commonly used dietary assessment instrument in epidemiology studies; see Carroll et al. (1997) or Liang & Wang (2005), among others. We focus on studying the impacts of long-term usual intake, body mass index and age, on the food frequency questionnaire intake, measured as percent calories from fat. As the long-term intake cannot be observed due to measurement errors and other sources of variability, the 24-hour recalls were used to obtain error-prone measurements of intake.

We consider the following linear quantile regression and measurement error models:

$$Q_\tau(y_j|x_j, z_{j1}, z_{j2}) = \beta_1(\tau) + \beta_2(\tau)x_j + \beta_3(\tau)z_{j1} + \beta_4(\tau)z_{j2},$$
$$w_{j,k} = x_j + u_{j,k} \qquad\qquad (j = 1, \ldots, 271; \ k = 1, \ldots, 6),$$

where y_j, x_j, z_{j1} and z_{j2} are the food frequency questionnaire intake, the long-term usual intake, body mass index and age of the jth subject, $w_{j,k}$ is the kth food recall intake of the jth subject, $u_{j,k}$ is the measurement error with mean zero and variance $\gamma_u^2 = 6\sigma_u^2$, and the intake measurements are on the log scale. For illustration, we study quantile levels $\tau = 0.2, 0.5$ and 0.8. For this dataset, each subject j has six internal replicates of food recall intake, $w_{j,k}$ ($k = 1, \ldots, 6$). Therefore, we estimate γ_u^2 by $\hat\gamma_u^2 = (5n)^{-1}\sum_{j=1}^{n}\sum_{k=1}^{6}(w_{j,k} - w_j)^2 = 0.132$, where $w_j = 1/6\sum_{k=1}^{6} w_{j,k}$. Thus the estimated variance of w_j as a measurement of x_j is $\hat\sigma_u^2 = \hat\gamma_u^2/6 = 0.022$. The attenuation factor, $1 - \hat\sigma_u^2/\mathrm{var}(w_j)$, is estimated as 0.737. Using the simulation and extrapolation method, we chose h as $0.3, 0.55$ and 0.35 for the normal corrected-loss estimator, and $0.36, 0.45$ and 0.21 for the Laplace corrected-loss estimator, at $\tau = 0.2, 0.5$ and 0.8, respectively.

Table 3 summarizes the coefficient estimates $\hat\beta(\tau)$ from the naive method, He & Liang's method, Wei & Carroll's method, the normal corrected-loss and the Laplace corrected-loss methods at three quantile levels. The values in parentheses are the corresponding bootstrap standard errors, based on 200 bootstrap samples. In the implementation of He & Liang's method, we first transform y_j to $y_j^* = \hat\sigma_u y_j/s$ to put the variances of measurement and regression errors on the same scale, where s is the standard deviation of the estimated residuals obtained from the naive method at the median. The resulting coefficient estimates are then transformed back to the original scale. According to He & Liang (2000, Theorem 2.1), their estimator $\hat\beta_1(\tau)$ of the intercept converges to some quantity depending on $\beta_k(\tau)$ ($k = 2, 3, 4$) and the unknown τth quantile of the regression error. Therefore, we omit $\hat\beta_1(\tau)$ in Table 3.

By accounting for the measurement error, both normal and Laplace corrected-loss methods identify a stronger association between food frequency questionnaire intake and the long-term intake at all three quantiles than the naive method. For instance, the normal corrected-loss estimates of $\beta_2(\tau)$ increase by 26, 44 and 74% at $\tau = 0.2, 0.5$, and 0.8, respectively, compared with the naive estimates. In contrast, He & Liang's method gives a $\beta_2(\tau)$ estimate smaller than the naive estimates at $\tau = 0.8$. Both normal and Laplace corrected-loss methods suggest that body mass index has a significantly positive effect at $\tau = 0.8$, but He & Liang's method gives a larger $\beta_2(\tau)$ estimate associated with a large standard error, which leads to insignificance. All methods show that age has no significant effect on any of the three quantiles. The effect of body mass index increases with the quantile level, and the effect of the long-term intake decreased with the quantile level, which indicate some form of heteroscedasticity. Our simulation demonstrated that He & Liang's method gives biased estimates for such heteroscedastic data. Wei &

Table 3. *Estimates, standard errors, of the quantile coefficients from different methods for the Womens Interview Survey of Health data*: $\beta_2(\tau)$, $\beta_3(\tau)$ *and* $\beta_4(\tau)$ *correspond to the effects of long-term usual intake*, *body mass index and age on the* τ *th quantile of the food frequency questionnaire intake*, *respectively*

τ	Method	$\beta_2(\tau)$	$\beta_3(\tau)$	$10 \times \beta_4(\tau)$
0·2	Naive	0·65 (0·12)	−0·11 (0·18)	0·28 (0·42)
	HL	0·81 (0·18)	−0·11 (0·28)	0·35 (0·38)
	WC	0·95 (0·18)	−0·19 (0·25)	0·29 (0·39)
	CLN	0·82 (0·20)	−0·01 (0·17)	0·10 (0·28)
	CLL	0·81 (0·16)	−0·05 (0·13)	0·16 (0·26)
0·5	Naive	0·51 (0·10)	0·22 (0·16)	−0·01 (0·29)
	HL	0·71 (0·13)	0·49 (0·27)	0·00 (0·30)
	WC	0·70 (0·14)	0·24 (0·16)	−0·13 (0·33)
	CLN	0·73 (0·14)	0·31 (0·13)	0·04 (0·27)
	CLL	0·71 (0·13)	0·29 (0·15)	−0·00 (0·27)
0·8	Naive	0·4 (0·17)	0·5 (0·18)	−0·06 (0·37)
	HL	0·38 (0·26)	0·75 (0·44)	0·15 (0·41)
	WC	0·62 (0·15)	0·51 (0·21)	−0·18 (0·36)
	CLN	0·70 (0·16)	0·47 (0·15)	−0·05 (0·28)
	CLL	0·78 (0·24)	0·71 (0·16)	−0·09 (0·33)

Naive, the naive method; HL, He & Liang's method; CLN, corrected-loss estimator for normal measurement error; CLL, corrected-loss estimator for Laplace measurement error; WC, Wei & Carroll's method.

Carroll's method yields the same significance results as the Laplace corrected-loss method, but it is computationally much more expensive. Using the same computer, it took 218 hours to obtain the bootstrap standard error of Wei & Carroll's estimates with 20 iterations for each of the 200 bootstrap samples, while it required only 35 minutes for the Laplace corrected-loss method.

5. Discussion

Our proposed estimation procedure has the following general structure. Since the quantile loss function cannot be corrected in the manner of Stefanski (1989) and Nakamura (1990), we projected the function into a class of suitably smooth functions via kernel smoothing. The corrected-loss method was then applied to the smoothed quantile objective function. We balanced the bias and variance by choosing the smoothing parameter using the simulation and extrapolation method of Delaigle & Hall (2008). This strategy is general and can be used in other problems where correction is possible after some smoothing of the objective functions.

We assumed a class of measurement errors, including normal and Laplace, for identification purpose. The two proposed estimators both showed robustness against misspecification of the measurement error distribution in the simulation study. Considering the finite sample performance and the computational efficiency, we recommend the Laplace corrected-loss estimator for practical usage. The corrected-loss methods developed herein can be extended to a wider class of distribution families, as long as their characteristic functions are proportional to the inverse of a polynomial; see Hong & Tamer (2003) for related discussions. The degree of the polynomial puts constraints on the smoothness of the objective function. Such an extension is beyond the scope of this paper.

APPENDIX

Proof of Theorem 1. We first prove (i). By the definitions of $A(\epsilon, \sigma^2, h)$ and $\rho_N(\epsilon, h)$, we get $A(\epsilon, \sigma^2, h) = \epsilon(\tau - 1/2 + I_1/\pi) + I_2/\pi$, where

$$I_1 = E_z \left\{ \int_0^{(\epsilon+i\sigma z)/h} \sin(t)/t \, dt \right\}, \quad I_2 = E_z \left\{ i\sigma z \int_0^{(\epsilon+i\sigma z)/h} \sin(t)/t \, dt \right\}.$$

Recall that $\sin(x) = (e^{ix} - e^{-ix})/(2i)$ and $\cos(x) = (e^{ix} + e^{-ix})/2$. Then

$$I_1 = \int_{-\infty}^{\infty} (2\pi)^{-1} e^{-z^2/2} \int_0^{(\epsilon+i\sigma z)/h} \sin(t)/t \, dt \, dz = \int_0^{1/h} \frac{1}{y} e^{\sigma^2 y^2/2} \sin(y\epsilon) \, dy.$$

Applying similar arguments, we have

$$I_2 = \int_0^{1/h} \frac{1}{y} \int_{-\infty}^{\infty} i\sigma z (2\pi)^{-1} e^{-z^2/2} \sin\{y(\epsilon + i\sigma z)\} \, dz \, dy$$

$$= -\int_0^{1/h} \sigma^2 e^{y^2\sigma^2/2} (e^{iy\epsilon} + e^{-iy\epsilon})/2 \, dy = -\int_0^{1/h} \sigma^2 e^{y^2\sigma^2/2} \cos(y\epsilon) \, dy.$$

We next show (ii). For any $U \sim N(0, 1)$, it is easy to show that

$$E\{\sin(a + bU)\} = \int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-u^2/2} (2i)^{-1} \{e^{i(a+bu)} - e^{-i(a+bu)}\} \, du = e^{-b^2/2} \sin(a),$$

$$E\{U \sin(a + bU)\} = b e^{-b^2/2} \cos(a), \quad E\{\cos(a + bU)\} = e^{-b^2/2} \cos(a). \tag{A1}$$

By (A1), $E\{\epsilon \sin(y\epsilon)\} = E\{(\mu + \sigma U) \sin(y\mu + y\sigma U)\} = e^{-y^2\sigma^2/2} \{\mu \sin(y\mu) + y\sigma^2 \cos(y\mu)\}$. Therefore, we have

$$E\{A(\epsilon, \sigma, h)\} = \mu(\tau - 1/2) + \frac{1}{\pi} \int_0^{1/h} \frac{1}{y} e^{y^2\sigma^2/2} E\{\epsilon \sin(y\epsilon)\} - \frac{\sigma^2}{\pi} \int_0^{1/h} e^{y^2\sigma^2/2} E\{\cos(y\epsilon)\} \, dy$$

$$= \mu(\tau - 1/2) + \frac{1}{\pi} \int_0^{1/h} \frac{1}{y} \mu \sin(y\mu) \, dy = \rho_N(\mu, h). \qquad \square$$

Proof of Lemma 1. Assertion (i) can be obtained from representation (6.3.4) in Kotz et al. (2001), and (ii) is a direct conclusion of Proposition 6.8.1 in Kotz et al. (2001). $\square$

Proof of Theorem 2. Suppose that there exists a function $\bar{g}(\epsilon)$ such that $E\{\bar{g}(\epsilon)\} = g(\mu)$. We shall show that $\bar{g}(\epsilon) = g(\epsilon) - 0 \cdot 5\sigma^2 g^{(2)}(\epsilon)$. First recall that if $\epsilon \sim L(\mu, \sigma^2)$, then $f(\epsilon) = (\sqrt{2}\sigma)^{-1} e^{-\sqrt{2}|\epsilon - \mu|/\sigma}$. Denote $\sigma = \sqrt{2}b$. Therefore

$$E\{\bar{g}(\epsilon)\} = e^{-\mu/b} \int_{-\infty}^{\mu} \bar{g}(\epsilon) \frac{1}{2b} e^{\epsilon/b} \, d\epsilon + e^{\mu/b} \int_{\mu}^{\infty} \bar{g}(\epsilon) \frac{1}{2b} e^{-\epsilon/b} \, d\epsilon$$

$$= e^{-\mu/b} I_1(\mu) + e^{\mu/b} I_2(\mu) = g(\mu). \tag{A2}$$

Differentiating both sides of the equation (A2) with respect to μ gives

$$
\begin{aligned}
g'(\mu) &= -\frac{1}{b}e^{-\mu/b}I_1(\mu) + \frac{1}{b}e^{\mu/b}I_2(\mu) \\
&= -\frac{1}{b}e^{-\mu/b}I_1(\mu) + e^{-\mu/b}\bar{g}(\mu)\frac{1}{2b}e^{\mu/b} + \frac{1}{b}e^{\mu/b}I_2(\mu) - e^{\mu/b}\bar{g}(\mu)\frac{1}{2b}e^{-\mu/b}.
\end{aligned}
\tag{A3}
$$

Differentiating (A3) again with respect to μ, we get $b^{-2}\{e^{-\mu/b}I_1(\mu) + e^{\mu/b}I_2(\mu)\} - b^{-2}\bar{g}(\mu) = g^{(2)}(\mu)$. Thus, we have $\bar{g}(\mu) = g(\mu) - b^2 g^{(2)}(\mu) = g(\mu) - 0{\cdot}5\sigma^2 g^{(2)}(\mu)$. □

Proof of Theorem 3.   For easy demonstration, we first show (ii). Define

$$
M_L^*(w, \beta, h) = n^{-1}\sum_{j=1}^{n}\{\rho_L^*(y_j, w_j, \beta, h) - \rho_L^*(y_j, w_j, \beta_0, h)\},
$$

$$
M_L(x, \beta, h) = n^{-1}\sum_{j=1}^{n}\{\rho_L(y_j, x_j, \beta, h) - \rho_L(y_j, x_j, \beta_0, h)\},
$$

$$
M(x, \beta) = n^{-1}\sum_{j=1}^{n}\{\rho(y_j, x_j, \beta) - \rho(y_j, x_j, \beta_0)\}.
$$

By Theorem 2, $E\{M_L^*(w, \beta, h)\} = E\{M_L(x, \beta, h)\}$. Therefore,

$$
\begin{aligned}
|M_L^*(w, \beta, h) - E\{M(x, \beta)\}| &\leqslant |M_L^*(w, \beta, h) - E\{M_L^*(w, \beta, h)\}| \\
&\quad + |E\{M_L(x, \beta, h)\} - M_L(x, \beta, h)| + |M_L(x, \beta, h) - M(x, \beta)| \\
&\quad + |M(x, \beta) - E\{M(x, \beta)\}|.
\end{aligned}
\tag{A4}
$$

Following the arguments used for proving Horowitz (1998, Lemma 1), we can show that the following relations hold almost surely as $n \to \infty$:

$$
\sup_{\beta \in \mathcal{B}} |M(x, \beta) - E\{M(x, \beta)\}| = o(n^{-1/2}\log n),
$$

$$
\sup_{\beta \in \mathcal{B}} |M_L(x, \beta, h) - E\{M_L(x, \beta, h)\}| = o(n^{-1/2}\log n) + O(h).
\tag{A5}
$$

Since the corrected loss function $\rho_L^*(\cdot)$ involves the second derivative of $\rho_L(\cdot)$, similar to Horowitz (1998, Lemma 3(b)), we obtain

$$
\sup_{\beta \in \mathcal{B}} |M_L^*(w, \beta, h) - E\{M_L^*(w, \beta, h)\}| = o\{\log n/(nh)^{1/2}\} + O(h)
\tag{A6}
$$

almost surely. Furthermore, under Assumption 6, $\sup_\epsilon |\rho_L(\epsilon, h) - \rho_\tau(\epsilon)| = \sup_\epsilon |\epsilon\{G_L(\epsilon/h) - I(\epsilon > 0)\}| = \sup_t |ht G_L(-t)| \leqslant hE|Z| = O(h)$, where $t = |\epsilon/h|$, $Z \sim G_L(\cdot)$. Therefore,

$$
\sup_{\beta \in \mathcal{B}} |M_L(x, \beta, h) - M(x, \beta)| = O(h)
\tag{A7}
$$

almost surely. Combining (A4)–(A7), we have that as $h \to 0$ and $(nh)^{-1/2}\log n \to 0$,

$$
\sup_{\beta \in \mathcal{B}} |M_L^*(w, \beta, h) - E\{M(x, \beta)\}| = o(1)
$$

almost surely. By Assumptions 3 and 4, β_0 uniquely minimizes $E\{M_N(x, \beta)\}$ over $\mathcal{B}$. By White (1980, Lemma 2.2), $\hat{\beta}_L \rightarrow \beta_0$ almost surely. To prove (i), we define

$$M_N^*(w, \beta, h) = n^{-1} \sum_{j=1}^{n} \{\rho_N^*(y_j, w_j, \beta, h) - \rho_N^*(y_j, w_j, \beta_0, h)\},$$

$$M_N(x, \beta, h) = n^{-1} \sum_{j=1}^{n} \{\rho_N(y_j, x_j, \beta, h) - \rho_N(y_j, x_j, \beta_0, h)\}.$$

By Theorem 1, $E\{M_N^*(w, \beta, h)\} = E\{M_N(x, \beta, h)\}$. Therefore,

$$
\begin{aligned}
|M_N^*(w, \beta, h) - E\{M(x, \beta)\}| &\leqslant |M_N^*(w, \beta, h) - E\{M_N^*(w, \beta, h)\}| \\
&\quad + |M_N(x, \beta, h) - E\{M_N(x, \beta, h)\}| + |M_N(x, \beta, h) - M(x, \beta)| \\
&\quad + |M(x, \beta) - E\{M(x, \beta)\}|.
\end{aligned}
\tag{A8}
$$

Denote $G_N(x) = \int_{-\infty}^{x} K_N(u)\, du$, where $K_N(u) = \sin(u)/(u\pi)$. For any $t > 0$, there exists an integer number $k \geqslant 0$ such that $t \in (k\pi, (k+1)\pi]$, and

$$
\begin{aligned}
|G_N(-t)| &= \frac{1}{\pi} \left| \int_{-\infty}^{-t} \sin(x)/x\, dx \right| = \frac{1}{\pi} \left| \int_{t}^{(k+1)\pi} \sin(x)/x\, dx + \sum_{l=k+1}^{\infty} \int_{l\pi}^{(l+1)\pi} \sin(x)/x\, dx \right| \\
&\leqslant \frac{1}{\pi} \left| \int_{t}^{(k+1)\pi} \sin(x)/x\, dx \right| + \frac{1}{\pi} \left| \int_{(k+1)\pi}^{(k+2)\pi} \sin(x)/x\, dx \right| \leqslant 2/|t|.
\end{aligned}
$$

Therefore, we have almost surely

$$
\begin{aligned}
\sup_{\epsilon} |\rho_N(\epsilon, h) - \rho_\tau(\epsilon)| &= \sup_{\epsilon} |\epsilon\{\tau - 1 + G_N(\epsilon/h)\} - \epsilon\{\tau - I(\epsilon < 0)\}| \\
&= \sup_{\epsilon} |\epsilon\{G_N(\epsilon/h) - I(\epsilon > 0)\}| \\
&= \sup_{t} \{ht\, G_N(-t)\} = O(h), \quad t = |\epsilon/h|,
\end{aligned}
\tag{A9}
$$

and

$$\sup_{\beta \in \mathcal{B}} |M_N(x, \beta, h) - M(x, \beta)| = O(h). \tag{A10}$$

By arguments similar to Horowitz (1998, Lemma 3(a)), it is easy to see that

$$\sup_{\beta \in \mathcal{B}} |M_N(x, \beta, h) - E\{M_N(x, \beta, h)\}| = o((\log n)/n^{1/2}) \tag{A11}$$

almost surely. Let $\epsilon = y - w^{\mathrm{T}}\beta$ and $\sigma^2 = \beta^{\mathrm{T}}\Gamma\beta$. By Assumption 5, the compactness of $\mathcal{B}$ and the fact that $|\sin(t)/t| \leqslant 1$, we have

$$E\{\rho_N^*(y, w, \beta, h)^2\} \leqslant C_1 + C_2 E \left\{ \int_0^{1/h} (\epsilon^2 + \sigma^2)\, e^{t^2\sigma^2/2}\, dt \right\}^2 \leqslant C_1 + C_2 h^{-2}\, e^{\sigma^2/h^2} = \delta_n,$$

where C_1 and C_2 are some positive constants. By Nolan & Pollard (1987, Lemma 22) and Pollard (1984, Theorem 2.37),

$$\sup_{\beta \in \mathcal{B}} |M_N^*(w, \beta, h) - E\{M_N^*(w, \beta, h)\}| = o(\delta_n^{1/2} n^{-1/2} \log n) \tag{A12}$$

almost surely, which is $o(1)$ if $h = C(\log n)^{-\delta}$, where $\delta < 1/2$ and C is some positive constant. The above equation together with (A5), (A8) and (A9)–(A11) gives $\sup_{\beta \in \mathcal{B}} |M_N^*(w, \beta, h) - E\{M(x, \beta)\}| = o(1)$ almost surely The rest of the proof follows the same lines as that for Theorem 3(ii). $\quad\square$

Proof of Theorem 4. Let $\rho^*(y, w, \beta, h)$ denote the corrected quantile loss function for either normal or Laplace measurement errors. Define $\psi_1^*(y, w, \beta, h) = \partial \rho^*(y, w, \beta, h)/\partial \beta$, $\psi_2^*(y, w, \beta, h) = \partial^2 \rho^*(y, w, \beta, h)/\partial \beta \partial \beta^{\mathrm{T}}$. Furthermore, let $\rho_n^*(w, \beta) = n^{-1} \sum_{j=1}^{n} \rho^*(y_j, w_j, \beta, h)$, $\psi_{1n}^*(w, \beta) = \partial \rho_n^*(w, \beta)/\partial \beta$ and $\psi_{2n}^*(w, \beta) = \partial^2 \rho_n^*(w, \beta)/\partial \beta \partial \beta^{\mathrm{T}}$. Under the conditions of Theorem 2 and 3, similar to (A6) and (A12), we have $\sup_{\beta \in \mathcal{B}} |\psi_{2n}^*(w, \beta) - E\{\psi_{2n}^*(w, \beta)\}| = o_p(1)$. Taylor expansion gives $n^{1/2}(\hat{\beta} - \beta_0) = -E\{\psi_{2n}^*(w, \beta_0)\} n^{1/2} \psi_{1n}^*(w, \beta_0) + o_p(1)$. By Assumption 7, we have $\lim_{n \to \infty} E\{\psi_{2n}^*(w, \beta_0)\} = A$. On the other hand, $n^{1/2} \psi_{1n}^*(w, \beta_0) = n^{-1/2} \sum_{j=1}^{n} \psi_1^*(y_j, w_j, \beta_0, h)$. By using the results of Theorems 1–2 and methods like those used to obtain the asymptotic means and variances of kernel density estimators, we have $E\{\psi_1^*(y_j, w_j, \beta_0, h)\} = E\{\psi_1(y_j, x_j, \beta_0, h)\} = o(n^{-1/2})$ as $n \to \infty$ and $h \to 0$, where $\psi_1(y, x, \beta, h) = \partial \rho(y, x, \beta, h)/\partial \beta$ and $\rho(y, x, \beta, h)$ is the smoothed quantile loss function. Therefore, $n^{1/2} E\{\psi_{1n}^*(w, \beta_0)\} = n^{1/2} E\{\psi_1^*(y_j, w_j, \beta_0, h)\} = o(1)$, which together with the central limit theorem gives $n^{1/2} \psi_{1n}^*(w, \beta_0) \to N(0, D)$ in distribution. □

Proof of Theorem 5. Similar to the proof in Theorem 3, the consistency of $\hat{\beta}$ can be proven by using the fact that $\hat{\Sigma} - \Sigma = O_p(n^{-1/2})$. In addition, note that minimizing the objective function in (5) with the estimated covariance matrix $\hat{\Sigma}$ is equivalent to solving the following stacked estimating equation $[n^{-1/2} \sum_{j=1}^{n} \psi_{1\beta}^*(y_j, w_j, \beta, h, \sigma)^{\mathrm{T}}, n^{-1/2} \sum_{j=1}^{n} \{S_j - m(m-1)\sigma\}^{\mathrm{T}}] = (0_p^{\mathrm{T}}, 0_{p(p+1)/2}^{\mathrm{T}})^{\mathrm{T}}$. The asymptotic normality can be proven by following the same arguments as in the proof of Theorem 4 and by expanding the stacked estimating function. □

References

Carroll, R. J. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Am. Statist. Assoc.* **83**, 1184–6.

Carroll, R. J., Freedman, L. & Pee, D. (1997). Design aspects of calibration studies in nutrition, with analysis of missing data in linear measurement error models. *Biometrics* **53**, 1440–57.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. New York: Chapman and Hall.

Delaigle, A. & Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Am. Statist. Assoc.* **103**, 280–7.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Can. J. Statist.* **20**, 155–69.

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

He, X. & Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica* **10**, 129–40.

Hong, H. & Tamer, E. (2003). A simple estimator for nonlinear error in variable models. *J. Economet.* **117**, 1–19.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66**, 1327–52.

Hu, Y. & Schennach, S. M. (2008). Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions using instruments. *Econometrica* **76**, 195–216.

Kotz, S., Kozubowski, T. J. & Podgórski, K. (2001). *The Laplace Distribution and Generalizations*. Boston: Birkhäuser.

Laplace, P. S. (1774). Memoir on the probability of causes of events. *Mém. Acad. R. Sci. (Translated in Statist. Sci.* 1, 359–78.) **6**, 621–56.

Liang, H. & Wang, N. (2005). Partially linear single-index measurement error models. *Statist. Sinica* **15**, 99–116.

Liang, H., Wang, S. & Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika* **94**, 185–98.

Luo, X. H., Stefanski, L. A. & Boos, D. D. (2006). Tuning variable selection procedures by adding noise. *Technometrics* **48**, 165–75.

McKenzie, H., Jerde, C., Visscher, D., Merrill, E. & Lewis, M. (2008). Inferring linear feature use in the presence of GPS measurement error. *J. Envir. Ecol. Statist.* **16**, 531–46.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika* **77**, 127–37.

Nolan, D. & Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780–99.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Purdom, E. & Holmes, S. P. (2005). Error distribution for gene expression data. *Statist. Appl. Genet. Mol. Biol.* **4**, Article 16.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

RICHARDSON, A. D. & HOLLINGER, D. Y. (2005). Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. *Agric. Forest Meteorol.* **131**, 191–208.

SCHENNACH, S. M. (2008). Quantile regression with mismeasured covariates. *Economet. Theory* **24**, 1010–43.

STEFANSKI, L. A. (1989). Unbiased estimation of a nonlinear function of a normal-mean with application to measurement error models. *Commun. Statist.* A **18**, 4335–58.

STEFANSKI, L. A. (2000). Measurement error models. *J. Am. Statist. Assoc.* **95**, 1353–8.

STEFANSKI, L. A. & CARROLL, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13**, 1335–51.

STEFANSKI, L. A. & CARROLL, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703–16.

STEFANSKI, L. A. & CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 165–84.

STEFANSKI, L. A. & COOK, J. R. (1995). Simulation-extrapolation: the measurement error jackknife. *J. Am. Statist. Assoc.* **90**, 1247–56.

VISSCHER, D. R. (2006). GPS measurement error and resource selection functions in a fragmented landscape. *Ecography* **29**, 458–64.

WEI, Y. & CARROLL, R. J. (2009). Quantile regression with measurement error. *J. Am. Statist. Assoc.* **104**, 1129–43.

WHITE, H. (1980). Nonlinear regression on cross-sectional data. *Econometrica* **48**, 721–46.