Quantile Regression With Measurement Error

Ying WEI and Raymond J. CARROLL

Regression quantiles can be substantially biased when the covariates are measured with error. In this paper we propose a new method that produces consistent linear quantile estimation in the presence of covariate measurement error. The method corrects the measurement error induced bias by constructing joint estimating equations that simultaneously hold for all the quantile levels. An iterative EM-type estimation algorithm to obtain the solutions to such joint estimation equations is provided. The finite sample performance of the proposed method is investigated in a simulation study, and compared to the standard regression calibration approach. Finally, we apply our methodology to part of the National Collaborative Perinatal Project growth data, a longitudinal study with an unusual measurement error structure.

KEY WORDS: Correction for attenuation; Growth curves; Longitudinal data; Measurement error; Quantile regression; Regression calibration; Regression quantiles.

1. INTRODUCTION AND MOTIVATION

Quantile regression, proposed by Koenker and Bassett (1978), has emerged as an important statistical methodology. By estimating various conditional quantile functions, quantile regression complements the focus of classical least squares regression on the conditional mean and offers a systematic strategy for examining how covariates influence the entire response distribution. It has been used in a wide range of applications including economics, biology, ecology, and finance.

Often, the covariates of interest, here denoted by $\mathbf{x}$, are not observable and instead are measured with error. It is well known that such errors can sometimes lead to substantial attenuation of estimated effects in mean regression (Carroll et al. 2006). As we illustrate in Section 5, the regression quantiles can also be seriously biased when the covariates contain measurement errors. This paper aims at developing statistical methods and theory that yield consistent quantile estimation in the presence of covariate measurement error.

There is some work on measurement error in quantile regression. He and Liang (2000) considered the case that errors in the response y and $\mathbf{x}$ are independent and follow the same symmetric distribution. Their approach yields consistent estimates. However, the equal distribution assumption is very strong and difficult to verify in practice. Chesher (2001) used a small error variance approximation approach, which does not require distributional assumptions on the response y. However, it does not yield consistent estimation, and the calculation is difficult when the error distribution depends on the covariates. Hu and Schennach (2008) and Schennach (2008) proved nonparametric identification of a nonparametric quantile function under various settings where there is an instrumental variable measured on all sampling units. There are many differences between our approach and theirs in terms of generality: our model is less general. There are also differences in implementation: they use sieve-based estimation, which requires choice of tuning constants such as the number of sieve terms as well as constraints on the sieve basis functions in order to estimate densities such as that of the response given the true covariates; our method relies on a simple, straightforward but novel EM-type implementation.

We consider a family of linear quantile regression models

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_{0,\tau} + \epsilon_i(\tau), \tag{1}$$

where y_i is the response for the ith individual, $\mathbf{x}_i$ is its corresponding covariate, and ϵ_i is the error term, whose τth quantile is zero conditional on $\mathbf{x}_i$. The distribution of ϵ_i may depend on $\mathbf{x}_i$. Moreover, we assume that Model (1) holds for all the τ's, that is, all the conditional quantiles are linear in $\mathbf{x}_i$ with quantile-specific coefficient $\boldsymbol{\beta}_{0,\tau}$. A special case of Model (1) is the well-known location scale model,

$$y_i = \mathbf{x}_i^\top \boldsymbol{\gamma}_1 + (\mathbf{x}_i^\top \boldsymbol{\gamma}_2)\epsilon_i, \tag{2}$$

where $\epsilon_i \sim F_\epsilon$ is independent of $\mathbf{x}_i$. This location-scale model implies that the τth conditional quantile coefficient $\boldsymbol{\beta}_{0,\tau} = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 F_\epsilon^{-1}(\tau)$.

An outline of this paper is as follows. Section 2 describes the basic methodology, while in Section 3 we describe the algorithm in detail. Section 4 gives asymptotic theory. Section 5 describes a simulation study. In Section 6 we apply our methodology to the National Collaborative Perinatal Project (NCPP) investigating the effect of body size in childhood on body size in adulthood. Section 7 gives concluding remarks. Technical details are given in an Appendix.

2. SEMIPARAMETRIC JOINT ESTIMATING EQUATIONS

2.1 Preliminaries and the Case That x Is Observed

Suppose $\{y_i, \mathbf{x}_i\}$ is a random sample from Model (1) with sample size n, where $\mathbf{x} = (x_1, \ldots, x_p)^\top$ is a p-dimensional covariate. Then an estimating equation for $\boldsymbol{\beta}_{0,\tau}$ can be written as

$$n^{-1} \sum_{i=1}^n \Psi_\tau (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_\tau) \mathbf{x}_i = 0, \tag{3}$$

Yin Wei is Assistant Professor, Department of Biostatistics, Columbia University, 722 West 168th St., New York, NY 10032 (E-mail: *ying.wei@columbia.edu*). Raymond J. Carroll is Distinguished Professor of Statistics, Nutrition and Toxicology, Department of Statistics, Texas A&M University, TAMU 3143, College Station, TX 77843-3143 (E-mail: *carroll@stat.tamu.edu*).

where $\Psi_\tau(u) = \tau - I\{u < 0\}$, $I\{\cdot\}$ denotes the indicator function, and $\boldsymbol{\beta}_\tau \in \mathbf{R}^p$ is a p-dimensional unknown coefficient vector. Actually, of course, the indicator function means that (3) may not have an exact zero, and what instead is done is to recast the issue as a minimization problem, and then use linear programming to solve this minimization problem. Thus, (3) is a slight abuse of notation, but since everything else involving observed data is an estimating equation that will have a zero, we will use the estimating equation nomenclature. The solution of equations (3) is proven to be a consistent estimate of $\boldsymbol{\beta}_{0,\tau}$. When $\mathbf{x}$ is measured with error and instead only a surrogate $\mathbf{w}_i$ is observed, naively replacing $\mathbf{x}_i$ by the observed $\mathbf{w}_i$ will result in substantial bias. We construct new estimating equations which take the measurement errors into account, and result in consistent estimation of $\boldsymbol{\beta}_{0,\tau}$. The new estimating equations take the form

$$\mathbf{S}_n^0(\boldsymbol{\beta}_\tau) = n^{-1} \sum_{i=1}^n \int_{\mathbf{x}} \Psi_\tau(y_i - \mathbf{x}^\top \boldsymbol{\beta}_\tau)\mathbf{x} \cdot f(\mathbf{x}|y_i, \mathbf{w}_i) \, d\mathbf{x} = 0, \quad (4)$$

where $f(\mathbf{x}|y_i, \mathbf{w}_i)$ is the conditional density of $\mathbf{x}$ given the observed $(y_i, \mathbf{w}_i)$. The integration in (4) makes the function continuous in its argument. The summand of (4) is $E_{\mathbf{x}}\{\Psi_\tau(y - \mathbf{x}^\top\boldsymbol{\beta}_\tau)\mathbf{x}|y, \mathbf{w}\}$, the conditional mean of the original score function given the observed y and $\mathbf{w}$. Letting $\Psi_{\text{new}}(y, \mathbf{w}, \boldsymbol{\beta}_\tau) = \int_{\mathbf{x}} \Psi_\tau(y - \mathbf{x}^\top \boldsymbol{\beta}_\tau)\mathbf{x} \cdot f(\mathbf{x}|y, \mathbf{w}) \, d\mathbf{x}$, it is easy to show that $E_y[\Psi_{\text{new}}(y, \mathbf{w}, \boldsymbol{\beta}_{0,\tau})|\mathbf{w}] \equiv 0$ for all $\mathbf{w}$. Therefore, $\Psi_{\text{new}}(y, \mathbf{w}, \boldsymbol{\beta}_\tau)$ is an unbiased estimating function, that is, has mean zero, and will be the basis for constructing estimating equations. We further impose the usual surrogacy condition that $f(y|\mathbf{x}, \mathbf{w}) = f(y|\mathbf{x})$, which means the contaminated $\mathbf{w}$ does not provide additional information about the response y if the true covariate $\mathbf{x}$ is known.

2.2 Two Technical Challenges

Although Equation (4) provides valid estimating equations for the coefficients of interests, $\boldsymbol{\beta}_{0,\tau}$, solving such equations is challenging, mainly due to the following two reasons. First, unlike the classical approaches in mean regression, *the conditional density $f(\mathbf{x}|y_i, \mathbf{w}_i)$ does not have any prespecified parametric form.* To get a better understanding of this, we can rewrite the conditional density $f(\mathbf{x}|y_i, \mathbf{w}_i)$ under the surrogacy condition by

$$f(\mathbf{x}|y_i, \mathbf{w}_i) = \frac{f(y_i|\mathbf{x})f(\mathbf{x}|\mathbf{w}_i)}{\int_{\mathbf{x}} f(y_i|\mathbf{x})f(\mathbf{x}|\mathbf{w}_i) \, d\mathbf{x}}. \quad (5)$$

In the spirit of quantile regression, we leave the error distribution of ϵ_i in Model (1) unspecified. Therefore, $f(y|\mathbf{x})$, and consequently, $f(\mathbf{x}|y_i, \mathbf{w}_i)$ does not have a parametric form. However, we get around this problem by noting that we can link the conditional density $f(y|\mathbf{x})$ to Model (1) by the following equation:

$$f(y|\mathbf{x}) = \lim_{\delta \to 0} \frac{\delta}{\mathbf{x}^\top[\boldsymbol{\beta}_0(\tau_y + \delta) - \boldsymbol{\beta}_0(\tau_y)]}, \quad (6)$$

where $\tau_y = \{\tau \in (0, 1): \mathbf{x}^\top \boldsymbol{\beta}_0(\tau) = y\}$, and $\boldsymbol{\beta}_0(\tau)$ is the true quantile coefficient viewed as a function of τ. To make the presentation clear, we note that $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), \dots, \beta_p(\tau))^\top \in \mathbf{R}^p \times (0, 1)$ is p-dimensional quantile coefficient process on the interval $(0, 1)$, and $\boldsymbol{\beta}_\tau = (\beta_{\tau,1}, \dots, \beta_{\tau,p})^\top \in \mathbf{R}^p$ is its evaluation specifically at quantile level τ. They are unknown parameters

in the estimating equations, while the previously defined $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\beta}_{0,\tau}$ are the corresponding true values.

Equation (6) is derived from the fact that the conditional quantile function $\mathbf{x}^\top \boldsymbol{\beta}_0(\tau)$ is the inverse function of the conditional distribution function $F(y|\mathbf{x})$. The density function is hence the reciprocal of the first derivative of the quantile function at the corresponding quantile level. This formulation reveals the second challenge in solving the estimating equations—*the conditional density $f(\mathbf{x}|y_i, \mathbf{w}_i)$ involves the entire quantile coefficient process $\boldsymbol{\beta}_0(\tau)$.* In other words, the estimating equations (4) need to be solved *jointly* for all the τ's, even if one is interested in a particular quantile level τ. Following the arguments above, we extend (4) to a semiparametric joint estimating equations

$$n^{-1} \sum_{i=1}^n \int_{\mathbf{x}} [\tau - I\{y_i - \mathbf{x}^\top \boldsymbol{\beta}(\tau) < 0\}]$$
$$\cdot \mathbf{x} \cdot f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}(\tau)\} \, d\mathbf{x} = 0, \quad (7)$$

where $f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}(\tau)\} = f\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\}f(\mathbf{x}|\mathbf{w}_i)/\int_{\mathbf{x}} f\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\}f(\mathbf{x}|\mathbf{w}_i) \, d\mathbf{x}$, and $f\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\}$ is the conditional density function of y_i given $\mathbf{x}$ that is induced by quantile function $\mathbf{x}^\top \boldsymbol{\beta}(\tau)$, that is, $f\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\} = F'\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\}$ and $F\{y_i|\mathbf{x}; \boldsymbol{\beta}(\tau)\} = \inf\{\tau \in (0, 1): \mathbf{x}^\top \boldsymbol{\beta}(\tau) > y\}$. We use $f\{\mathbf{x}|y, \mathbf{w}; \boldsymbol{\beta}(\tau)\}$ and $f\{y|\mathbf{x}; \boldsymbol{\beta}(\tau)\}$ to indicate their dependence on the entire unknown quantile process $\boldsymbol{\beta}(\tau)$. The aforementioned true densities $f(\mathbf{x}|y, \mathbf{w})$ and $f(y|\mathbf{x})$ can be written as $f(\mathbf{x}|y, \mathbf{w}) = f\{\mathbf{x}|y, \mathbf{w}; \boldsymbol{\beta}_0(\tau)\}$ and $f(y|\mathbf{x}) = f\{y|\mathbf{x}; \boldsymbol{\beta}_0(\tau)\}$. We note that the estimating function of (7), $\int_{\mathbf{x}}[\tau - I\{y - \mathbf{x}^\top\boldsymbol{\beta}(\tau) < 0\}] \cdot \mathbf{x} \cdot f\{\mathbf{x}|y, \mathbf{w}; \boldsymbol{\beta}(\tau)\} \, d\mathbf{x}$, is a function of τ, and its conditional mean at $\boldsymbol{\beta}_0(\tau)$,

$$E_y\left[\int_{\mathbf{x}} [\tau - I\{y - \mathbf{x}^\top \boldsymbol{\beta}_0(\tau) < 0\}]\right.$$
$$\left. \cdot \mathbf{x} \cdot f\{\mathbf{x}|y, \mathbf{w}; \boldsymbol{\beta}_0(\tau)\} \, d\mathbf{x}\middle|\mathbf{w}\right] \equiv 0 \quad (8)$$

for all the τ and $\mathbf{w}$, the estimating equations (7) are hence unbiased joint estimating equations. On the other hand, we say that the estimating equations (7) are semiparametric, since the conditional quantile function of y is a parametric function of $\mathbf{x}$ at given τ, but the coefficients $\boldsymbol{\beta}(\tau)$ are nonparametric functions of τ. The estimating equations (7) involve the infinite dimensional parametric space $\boldsymbol{\beta}(\tau) \in \mathbf{R}^p \times (0, 1)$.

2.3 The Continuous Case

To estimate $\boldsymbol{\beta}_0(\tau)$, we assume that they are smooth functions on $(0, 1)$, and approximate $\boldsymbol{\beta}(\tau)$ in (7) by natural linear splines with common internal knots $\Omega = \{\varepsilon = \tau_1 < \tau_2 < \cdots < \tau_{k_n} = 1 - \varepsilon\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_{\tau_1}^\top, \boldsymbol{\beta}_{\tau_2}^\top, \dots, \boldsymbol{\beta}_{\tau_{k_n}}^\top)^\top$ be the set of quantile coefficients at quantile levels Ω. We define a natural linear spline $\mathfrak{L}_\theta(\tau): [0, 1] \to \mathbf{R}^p$ as p continuous, piecewise linear functions on [0,1] which satisfies $\mathfrak{L}_\theta(\tau_k) = \boldsymbol{\beta}_{\tau_k}$, and is subject to the constraints that $\mathfrak{L}_\theta'(0) = \mathfrak{L}_\theta'(1) = \mathbf{0}$. With a sufficient numbers of knots, that is, $k_n \to \infty$ and $\varepsilon \to 0$, the difference between $\boldsymbol{\beta}(\tau)$, and its spline approximation $\mathfrak{L}_\theta(\tau)$ is negligible

(de Boor 2001). Consequently, we also approximate the conditional density function $f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}(\tau)\}$ by

$$f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}(\tau)\}$$

$$\approx f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta})$$

$$\hat{=} f(y_i|\mathbf{x}; \boldsymbol{\theta})f(\mathbf{x}|\mathbf{w}_i) \Big/ \int_{\mathbf{x}} f(y_i|\mathbf{x}; \boldsymbol{\theta})f(\mathbf{x}|\mathbf{w}_i)\, d\mathbf{x}, \quad (9)$$

where

$$f(y_i|\mathbf{x}; \boldsymbol{\theta}) = \sum_{k=1}^{k_n} \frac{\tau_{k+1} - \tau_k}{\mathbf{x}^\top \boldsymbol{\beta}_{\tau_{k+1}} - \mathbf{x}^\top \boldsymbol{\beta}_{\tau_k}} I\{\mathbf{x}^\top \boldsymbol{\beta}_{\tau_k} \le y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{\tau_{k+1}}\}.$$

In other words, we approximate the quantile function $\mathbf{x}^\top \boldsymbol{\beta}(\tau)$ by its spline approximation $\mathbf{x}^\top \mathfrak{L}_{\boldsymbol{\theta}}(\tau)$. In this way, we only need to solve the estimating equations (7) for the grid of internal knots, τ_k's. We hence reduce the infinite dimensional estimating equations (7) to a finite dimensional case

$$S_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}) \otimes \mathbf{x} \cdot f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta})\, d\mathbf{x} = 0,$$

$$(10)$$

where $\Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}) = \{\Psi_{\tau_1}(y_i - \mathbf{x}^\top \boldsymbol{\beta}_{\tau_1}), \dots, \Psi_{\tau_{k_n}}(y_i - \mathbf{x}^\top \times \boldsymbol{\beta}_{\tau_{k_n}})\}^\top$ is a k_n-dimensional vector, and $\otimes$ stands for Kronecker product. We note that $\Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}) \otimes \mathbf{x}$ is a $k_n \times p$ dimensional vector, which consists of k_n sets of original estimating functions $\{\Psi_{\tau_k}(y_i - \mathbf{x}^\top \boldsymbol{\beta}_{\tau_k})\mathbf{x}\}_{k=1,\dots,k_n}$ on quantile levels Ω; while $f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta})$ is the approximated conditional density of $\mathbf{x}$ given the observed $(y_i, \mathbf{w}_i)$. We call (10) the *working estimating equations*, which can be viewed as a spline approximation of the unbiased semiparametric joint estimating equations defined in (7). Solving such equations directly is not easy since they involve integration. In the next section, we outline an iterative EM-type algorithm to obtain the solution of (10).

3. ESTIMATION ALGORITHM

3.1 Preliminaries

The crux of all measurement error problems which have a likelihood flavor is the estimating of the distribution of $\mathbf{x}$ given $\mathbf{w}$. Our method, as well as many others, depends on being able to estimate this distribution reasonably well. In practice, this distribution needs to be estimated, and estimating the distribution of $\mathbf{x}$ given $\mathbf{w}$ depends on the context of the problem.

In nutritional epidemiology, it is fairly common to use parametric and semiparametric methods to transform the observed $\mathbf{w}$-data to normality, and to then assume that the measurement error model is additive with $\mathbf{x}$ normally distributed; see Nusser et al. (1990, 1997a, 1997b). In other cases, with replicates of the observed but error-prone predictor, the transformation is to normality and homoscedasticity of the measurement errors (Eckert, Carroll, and Wang 1997), with a flexible model for the distribution of $\mathbf{x}$. See also Carriquiry (2003) for other methods. In the simulation study presented later section, we used this additive transformation model as the basis for estimating $f(\mathbf{x}|\mathbf{w})$.

In some instances, when it can be assumed that $\mathbf{w} = \mathbf{x} + \mathbf{u}$ where there are replicated values of $\mathbf{w}$, the distribution of the measurement errors $\mathbf{u}$ as well as the latent variable $\mathbf{x}$ can be estimated nonparametrically (Li and Vuong 1998; Delaigle, Hall,

and Meister 2008), and hence so too can the distribution of $\mathbf{x}$ given $\mathbf{u}$ be estimated nonparametrically. In the next subsection, we first assume that the crucial distribution is known, and then show how the method is modified when it is estimated. We conjecture that interesting and possibly not parametric-rate properties arise when the distribution of $\mathbf{x}$ given $\mathbf{w}$ is estimated nonparametrically.

3.2 Basic Algorithm

In this section, we outline an iterative algorithm to obtain a solution of the working estimating equations (10). We will establish in the next section the consistency property of the resulting estimates. The algorithm can be viewed as an nonparametric analogue of the EM algorithm, since the basic components involve iteratively updating the conditional distribution $f(\mathbf{x}|y_i, \mathbf{w}_i, \boldsymbol{\theta})$ and quantile coefficients $\boldsymbol{\theta}$. However, we do not have specific likelihood functions to work with as in classical EM algorithms. Let ν be the indicator of iteration steps, the main steps of the algorithm are the following:

Step 1. Set initial values of $\boldsymbol{\theta}$ based on uncorrected quantile regression.

Step 2. Update the distribution $f^{(\nu)}(\mathbf{x}|y_i, \mathbf{w}_i)$ based on the $\widehat{\boldsymbol{\theta}}^{(\nu-1)} = (\widehat{\boldsymbol{\beta}}_{\tau_1}^{(\nu-1)}, \dots, \widehat{\boldsymbol{\beta}}_{\tau_{k_n}}^{(\nu-1)})$, that is,

$$f^{(\nu)}(\mathbf{x}|y_i, \mathbf{w}_i) = \frac{f(y_i|\mathbf{x}, \widehat{\boldsymbol{\theta}}^{(\nu-1)})f(\mathbf{x}|\mathbf{w}_i)}{\int_{\mathbf{x}} f(y_i|\mathbf{x}, \widehat{\boldsymbol{\theta}}^{(\nu-1)})f(\mathbf{x}|\mathbf{w}_i)\, d\mathbf{x}},$$

where

$$f(y_i|\mathbf{x}, \widehat{\boldsymbol{\theta}}^{(\nu-1)}) = \sum_{k=1}^{k_n} \frac{\tau_k - \tau_{k-1}}{\mathbf{x}^\top (\widehat{\boldsymbol{\beta}}_{\tau_k}^{(\nu-1)} - \widehat{\boldsymbol{\beta}}_{\tau_{k-1}}^{(\nu-1)})}$$

$$\times I(\mathbf{x}^\top \widehat{\boldsymbol{\beta}}_{\tau_{k-1}}^{(\nu-1)} \le y_i < \mathbf{x}^\top \widehat{\boldsymbol{\beta}}_{\tau_k}^{(\nu-1)}).$$

Step 3. Estimate $\widehat{\boldsymbol{\theta}}^{(\nu)} = (\widehat{\boldsymbol{\beta}}_{\tau_1}^{(\nu)}, \dots, \widehat{\boldsymbol{\beta}}_{\tau_{k_n}}^{(\nu)})$ based on the new estimating function $\Psi_{\text{new}}(y_i, \mathbf{w}_i; \boldsymbol{\beta}_\tau)$ evaluated at $f^{(\nu)}(\mathbf{x}|y_i, \mathbf{w}_i)$. In order to perform this step, we have to make a numerical approximation to an integral. We do this via translating the problem into a weighted quantile regression problem. Let $\widetilde{x}_i = (\widetilde{\mathbf{x}}_{i,1}, \widetilde{\mathbf{x}}_{i,2}, \dots, \widetilde{\mathbf{x}}_{i,m})$ is a fine grid of possible $\mathbf{x}_i$ values, akin to a set of abscissas in Gaussian quadrature. Then the new sample estimating equations are

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \Psi_{\tau_k}(y_i - \widetilde{\mathbf{x}}_{i,j}^\top \boldsymbol{\beta}_{\tau_k})\widetilde{\mathbf{x}}_{i,j} f^{(\nu)}(\widetilde{x}_{i,j}|y_i, \mathbf{w}_i) = 0,$$

$$k = 1, \dots, k_n,$$

which is a weighted quantile regression with response y_i over the covariates $\widetilde{x}_{i,j}$ with weights $f^{(\nu)}(\mathbf{x}_{i,j}|y_i, \mathbf{w}_i)$.

Note that the original quantile regression estimating function $\Psi_\tau(y - \mathbf{x}^\top \boldsymbol{\beta}_\tau)\mathbf{x}$ can be viewed as the first derivative of the logarithm of an asymmetric Laplace distribution (Koenker and Machado 1999)

$$f(y|\mathbf{x}; \boldsymbol{\beta}_\tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left[-\frac{y - \mathbf{x}^\top \boldsymbol{\beta}_\tau}{\sigma}\{\tau - I(y - \mathbf{x}^\top \boldsymbol{\beta}_\tau \le 0)\}\right]$$

with respect to $\boldsymbol{\beta}_\tau$. The convergence of the proposed algorithm follows from classical results on the EM algorithm (McLachlan and Krishnan 2008, page 19).

The estimation algorithm involves a turning parameter, the number of quantile levels. The necessary number of quantile levels depends on the underlying distribution of y. In our numerical investigations, we found 40 evenly spaced quantile levels worked well even for heavy-tail distributions such as the log-normal. However, when the dimension of $\mathbf{x}$ exceeds 2, evaluating $f^{(v)}(\mathbf{x}|y, \mathbf{w})$ on a fine grid of $\mathbf{x}$ in Step 3 could be computationally undesirable. Instead, we can simulate $\widetilde{\mathbf{x}}_{i,j}$ from the conditional density of $\mathbf{x}$ given $\mathbf{w}$ to ensure a sufficient number of $\widetilde{\mathbf{x}}_{i,j}$ with high densities, and then use importance weights to adjust the bias due to the difference between $f^{(v)}(\mathbf{x}|y, \mathbf{w})$ and $f(\mathbf{x}|\mathbf{w})$. This Monte Carlo integration approach can be implemented to reduce the computational burden. We delay until Section 6 the details of this modified algorithm implementing MC integration.

3.3 When the Distribution of x Given w Is Estimated

Denoting $\widetilde{f}(\mathbf{x}|\mathbf{w})$ as the estimated conditional distribution of $\mathbf{x}$ given $\mathbf{w}$, we further approximate the working estimating equations (10) by

$$\widetilde{S}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^{\top}\boldsymbol{\theta}) \otimes \mathbf{x} \cdot \widetilde{f}(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}) \, d\mathbf{x} = 0,$$

(11)

in which $\widetilde{f}(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta})$ is the conditional distribution of $\mathbf{x}$ given y_i and $\mathbf{w}_i$ with $\widetilde{f}(\mathbf{x}|\mathbf{w}_i)$ and $\boldsymbol{\theta}$. The working estimating equations (11) involve two approximations: (a) $\boldsymbol{\beta}(\tau)$ is approximated by a linear spline; and (b) the conditional density $f(\mathbf{x}|\mathbf{w})$ is approximated by its estimator. The estimation algorithm remains unchanged except that we need to replace the $f(\mathbf{x}|\mathbf{w}_i)$ in Step 2 by $\widetilde{f}(\mathbf{x}|\mathbf{w}_i)$. Once $f(\mathbf{x}|\mathbf{w})$ is estimated, it stays the same in all the iterations. The iteration converges to the solution of the approximated working estimation equations (11), which we denote as $\widehat{\boldsymbol{\theta}}_n$. Let $\widehat{\boldsymbol{\beta}}_n(\tau) = \mathfrak{L}_{\widehat{\boldsymbol{\theta}}_n}(\tau)$ be the natural linear spline extended from $\widehat{\boldsymbol{\theta}}_n$. We show in next section that $\widehat{\boldsymbol{\beta}}_n(\tau)$ is a consistent estimator of $\boldsymbol{\beta}_0(\tau)$ under certain conditions, especially that $\widetilde{f}(\mathbf{x}|\mathbf{w})$ is a consistent estimator of $f(\mathbf{x}|\mathbf{w})$, $k_n \to \infty$ and $k_n n^{-1} \to 0$.

4. ASYMPTOTIC PROPERTIES

For a vector $\mathbf{x}$, we use $\|\mathbf{x}\|$ to denote its Euclidean norm, and use $|\mathbf{x}|$ for its componentwise absolute values. By $|\mathbf{a}| < 1$, we mean that each component of $\mathbf{a}$ is bounded by ± 1.

In this section, we first list and discuss sufficient conditions for the consistency of $\widehat{\boldsymbol{\beta}}_n(\tau)$, with the main result summarized in Theorem 1. We first introduce the conditions on the covariates $(\mathbf{x}, \mathbf{w})$.

Assumption 1. The covariate $\mathbf{x}$ has bounded support $\mathcal{X}$, and:

(i) the conditional density $f(\mathbf{x}|\mathbf{w})$ is bounded away from infinity for all $(\mathbf{x}, \mathbf{w})$;

(ii) there exists a consistent estimator $\widehat{f}(\mathbf{x}|\mathbf{w})$ of $f(\mathbf{x}|\mathbf{w})$, such that, $\forall \mathbf{x}$,

$$\sup_{\mathbf{w}} |\widehat{f}(\mathbf{x}|\mathbf{w}) - f(\mathbf{x}|\mathbf{w})| = o_p(1).$$

Remark 1. Assumption 1(i) is quite mild. The assumption that $\mathbf{x}$ has compact support is needed in the proof, but we think the reason is more to do with the method of proof than to the actual requirement. Our simulations does not obey this restriction, although due to the nature of the data, our empirical example does obey the restriction.

Recall that $\boldsymbol{\beta}_0(\tau)$ is the true quantile coefficient function, and $\boldsymbol{\beta}_{0,\tau}$ is the true value at quantile level τ, then for any $\mathbf{x} \in \mathcal{X}$, $\mathbf{x}^{\top}\boldsymbol{\beta}_0(\tau)$ defines a conditional quantile function. We further define a functional $h_{\mathbf{x}}(\tau) = 1/\mathbf{x}^{\top}\boldsymbol{\beta}_0'(\tau)$, which is the density of y given $\mathbf{x}$ at the τth quantile. We call this the conditional quantile density function. Its reciprocal is known as the sparsity function (Welsh 1988 and Koenker and Xiao 2004). With these definitions, we now introduce the smoothness conditions on $\boldsymbol{\beta}_0(\tau)$.

Assumption 2. The true coefficient $\boldsymbol{\beta}_0(\tau)$ are smooth functions on $(0, 1)$, and for any $\mathbf{x} \in \mathcal{X}$,

(i) $0 < h_{\mathbf{x}}(\tau) < \infty$ and $\lim_{\tau \to 0} h_{\mathbf{x}}(\tau) = \lim_{\tau \to 1} h_{\mathbf{x}}(\tau) = 0$;

(ii) there exist constants M and $v_1, v_2 > -1$ such that its first derivative is bounded by

$$\sup_{\mathbf{x}} h_{\mathbf{x}}'(\tau) < M\tau^{v_1}(1 - \tau)^{v_2}.$$

(12)

Remark 2. The first condition of Assumption 2 implicitly assumes that the conditional density $f(y|\mathbf{x})$ is continuous, bounded away from zero and infinity, and diminishes to zero as τ goes to 0 and 1. The assumption that $0 < h_{\mathbf{x}}(\tau) < \infty$ is fairly standard. The equivalent version of this assumption is that $0 < f(\epsilon_i) < \infty$, and this is commonly assumed in the quantile regression literature; see, for example, Portnoy (2003) and Koenker (2004). The second condition is on the tail behavior of $f(y|\mathbf{x})$, noting that $h_{\mathbf{x}}'(\tau)$ determines how smooth the density function diminishes as the quantile level goes to the two ends. The condition (12) is fairly general, and covers a wide range of distribution families, such as exponential, Gaussian, and Student t distributions.

We now make two further definitions:

- Recall that $\mathbf{S}_n^0(\boldsymbol{\beta}_\tau)$ is defined at (4), and let $\mathbf{S}^0(\boldsymbol{\beta}_\tau) = E\{\mathbf{S}_n^0(\boldsymbol{\beta}_\tau)\}$ be it expectation at the τth quantile.
- Recall that $\mathbf{S}_n(\boldsymbol{\theta})$ is defined at (10). Let $\mathbf{S}_n(\boldsymbol{\beta}_\tau)$ be the $p \times 1$ subset of $\mathbf{S}_n(\boldsymbol{\theta})$ that corresponds to the τth quantile, and let $\mathbf{S}(\boldsymbol{\beta}_\tau) = E\{\mathbf{S}_n(\boldsymbol{\beta}_\tau)\}$ be its expectation.

We make the following assumptions.

Assumption 3. The true coefficient $\boldsymbol{\beta}_{0,\tau}$ is the unique solution to the equation $\mathbf{S}^0(\boldsymbol{\beta}_\tau) = 0$, for all $\tau \in (0, 1)$, and there exist a $\boldsymbol{\beta}_\tau^*$ that uniquely solves the equation $\mathbf{S}(\boldsymbol{\beta}_\tau) = 0$, for all $\tau \in (0, 1)$.

Assumption 4. There exists a compact set $\boldsymbol{\Theta} \in \mathbf{R}^p$, such that

$$|\widetilde{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)| \leq \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \times \Omega} |\widetilde{\mathbf{S}}_n(\boldsymbol{\theta})| + o_p(1).$$

Remark 3. Assumption 3 is the identifiability condition that is commonly assumed in the quantile regression literature, while Assumption 4 is used to ensure that the solution to the ap-

proximated working estimating equations is confined to a compact set $\boldsymbol{\Theta}$, which is a standard condition for M and Z estimators.

Moreover, if we are willing to assume that the difference between the true coefficient process $\boldsymbol{\beta}_0(\tau)$ and its spline approximation is negligible, that is, $\boldsymbol{\beta}_0(\tau)$ are linear natural splines with a fixed number of knots K, we obtain the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$. The results are summarized in Theorem 2 below. Additional assumptions for asymptotic normality are listed as follows.

Assumption 2*. The coefficients $\boldsymbol{\beta}_0(\tau)$ are continuous linear splines on $[0, 1]$ with internal knots $\Omega = \{0 < \tau_1 < \tau_2 < \cdots < \tau_K < 1\}$.

Let $\Psi_{\text{new}}(y_i, \mathbf{w}_i, \boldsymbol{\theta}) = \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}) \otimes \mathbf{x} \cdot f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}) \, d\mathbf{x}$, then $S_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \Psi_{\text{new}}(y_i, \mathbf{w}_i, \boldsymbol{\theta})$. We further denote $\mathbf{V}_n = \text{var}\{S_n(\boldsymbol{\theta}_0)\}$ and $\mathbf{D}_n = \frac{\partial}{\partial \boldsymbol{\theta}_0} S_n(\boldsymbol{\theta}_0)$. We make the following additional assumptions.

Assumption 5. There exists a nonnegative definite matrix $\mathbf{V}$ such that $\mathbf{V}_n \to \mathbf{V}$ as $n \to \infty$.

Assumption 6. There exists a positive definite matrix $\mathbf{D}$, such that $\mathbf{D}_n \to \mathbf{D}$ in probability as $n \to \infty$.

Theorem 1. Under Assumptions 1–4, for $k_n \to \infty$, $k_n n^{-1} \to 0$, $\widehat{\boldsymbol{\beta}}_n(\tau)$ is a consistent estimator of $\boldsymbol{\beta}_0(\tau)$, that is,

$$\sup_{\tau \in [1/(k_n+1), k_n/(k_n+1)]} \|\widehat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)\| = o_p(1).$$

Theorem 2. Under Assumptions 1, 2*, and 3–6,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to \text{N}(0, \boldsymbol{\Sigma}),$$

in distribution as $n \to \infty$, where $\boldsymbol{\Sigma} = \mathbf{D}^{-1} \mathbf{V} \mathbf{D}^{-1}$.

The proofs of the two theorems are provided in the Appendix.

5. SIMULATION STUDY

5.1 Model Setup

To understand the effects of measurement errors and to demonstrate the performance of our method, we used a location-scale quantile regression model

$$y_i = \beta_1 + \beta_2 x_i + (\gamma_1 + \gamma_2 x_i)\epsilon_i, \tag{13}$$

where $\epsilon_i = \text{Normal}(0, 1)$. It follows that the actual quantile function of y given x is $\beta_{\tau,1} + \beta_{\tau,2} x$ with $\beta_{\tau,1} \equiv 0$ and $\beta_{\tau,2} = 2 + 0.5 \Phi^{-1}(\tau)$. We further assume that the x_i's are measured with error following one of two models:

Model I (Additive).

$$w_i = x_i + U_i, \quad \text{with } x_i \sim \text{N}(4, 1) \text{ and } U_i \sim \text{N}(0, \sigma_u^2).$$

Model II (Multiplicative).

$$\log(w_i) = \log(x_i) + U_i,$$

$$\text{with } \log(x_i) \sim \text{N}(2, 1) \text{ and } U_i \sim \text{N}(0, \sigma_u^2).$$

In Model I, w and x follow normal distributions, while in Model II, they follow log-normal distributions. We kept the variances of x in Model I or $\log(x)$ in Model II and ϵ to be 1, so that the two models have a constant signal-noise ratio. For each model, we assumed the measurement error U_i follows a normal distribution with mean 0 and variances $\sigma_u^2 = 0.25$ and 0.5. These choices correspond to moderate and larger attenuation which equal $R = 4/5$ and $R = 2/3$, respectively.

Regression model scale. It is important to note that the model (13) for the quantile regression function is in the *original* scale of x_i. In Model II, we are merely stating that the measurement errors are multiplicative.

5.2 Estimation of $f(x|w)$

To estimate the conditional density of x given w, we assume that the w_i's were observed from the following model:

$$\Lambda(w_i, \lambda) = \Lambda(x_i, \lambda) + U_i,$$

$$U_i \sim \text{N}(0, \sigma_u^2) \text{ and } \Lambda(x_i, \lambda) \sim \text{N}(0, \sigma_x^2), \tag{14}$$

where the function $\Lambda(\cdot)$ is the Box–Cox transformation function, that is, $\Lambda(Z, \lambda) = \log(Z)$ if $\lambda = 0$ and $= (Z^\lambda - 1)/\lambda$ otherwise.

The true power parameters λ in Models I and II are 1 and 0, respectively, but are assumed unknown. We estimate λ by maximizing the log-likelihood function of the w_i's, that is,

$$\widehat{\lambda} = \arg\max_\lambda L_n(\lambda)$$

$$= -\frac{n}{2} \sum_{i=1}^n \frac{[\Lambda(w_i, \lambda) - \overline{\Lambda}(\cdot, \lambda)]^2}{n} + (\lambda - 1) \sum_{i=1}^n \log(w_i),$$

where $\overline{\Lambda}$ is the sample mean. This is a transformation that tries to make $\Lambda(w, \lambda) = \text{Normal}(\mu, \sigma^2)$. For understanding the measurement error, we assume that there exists a subset of 100 replicates $(w_{i,1}, w_{i,2})$ for estimating σ_u^2. With replicates, the variance σ_u^2 can be estimated by half of the sample variance of the difference of the transformed $w_{i,1}$ and $w_{i,2}$, that is, $\Lambda(w_{i,1}, \widehat{\lambda}) - \Lambda(w_{i,2}, \widehat{\lambda})$, so that

$$\widehat{\sigma}_u^2 = \widehat{\text{var}}\{\Lambda(w_{i,1}, \widehat{\lambda}) - \Lambda(w_{i,2}, \widehat{\lambda})\}/2.$$

With the estimated $\widehat{\lambda}$ and $\widehat{\sigma}^2$, the estimated $f(x|w)$ is

$$\widehat{f}(x|w) = \widehat{s}^{-1} \Phi\left(\frac{\Lambda(x, \widehat{\lambda}) - \widehat{\mu}}{\widehat{s}}\right) x^{\widehat{\lambda} - 1}, \tag{15}$$

where $\Phi(\cdot)$ is the density function of the standard normal distribution, and $\widehat{\mu}$ and $\widehat{s}$ are the estimated conditional mean and standard deviation of the transformed $\Lambda(x, \widehat{\lambda})$ given the observed w. Let $\overline{\Lambda}_w$ and $\widehat{\sigma}_w^2$ be the sample mean and variance of $\Lambda(w_i, \widehat{\lambda})$. Then $\widehat{\sigma}_x^2 = \widehat{\sigma}_w^2 - \widehat{\sigma}_u^2$ is an unbiased estimator of the variance of x. It then follows that $\widehat{\mu}$ and $\widehat{s}$ in the density (15) are

$$\widehat{\mu} = \overline{\Lambda}_w + (\widehat{\sigma}_x^2/\widehat{\sigma}_w^2)\{\Lambda(w, \widehat{\lambda}) - \overline{\Lambda}_w\},$$

$$\widehat{s} = \sqrt{\widehat{\sigma}_x^2\{1 - (\widehat{\sigma}_x^2/\widehat{\sigma}_w^2)\}}.$$

5.3 Estimators Considered

We performed 100 simulations with $n = 500$ for each of these models, and computed four estimators:

(a) the naive estimator that replaces x by w;

(b) the regression calibration estimator, which replaces x by an estimate mean of x given w via the linear regression of x on w, or $\log(x)$ on $\log(w)$. Specifically, we replace x_i in Model I by $(1 - \widehat{R})\bar{w} + \widehat{R}w_i$, and replace x_i in Model II by $\exp\{(1 - \widehat{R})\overline{\log(w)} + \widehat{R}\log(w_i) + \widehat{\sigma}_u^2/2\}$, where $\widehat{R}$ and $\widehat{\sigma}_u^2$ and are estimated based on the set of replicates. When applied to ordinary linear regression, the regression calibration estimator is consistent for estimating (β_1, β_2), that is, for the mean regression curve;

(c) our method with 40 evenly spaced quantile levels (internal knots), but assuming $f(x|w)$ is known;

(d) our method with 40 evenly spaced quantile levels (internal knots), and the conditional density $f(x|w)$ is estimated based on Model (14).

When applying the proposed estimation algorithm, we chose Ω to be a set of 40 evenly spaced quantile levels, the convergence criterion is set to be the average of $|\widehat{\boldsymbol{\theta}}^{(\nu)} - \widehat{\boldsymbol{\theta}}^{(\nu-1)}| < 0.01$, and the maximum iteration steps is 50. In approach (d), the power λ is estimated using the w_i's, while the variance σ_u^2 is estimated using the 100 pairs of replicates $(w_{i,1}^*, w_{i,2}^*)$.

5.4 Simulation Results and Discussion

In Figures 1 and 2, we present the resulting estimated slope (Figure 1) and intercept (Figure 2) functions from the four approaches, both Models I and II and with attenuation $R = 4/5$. The top panel plots the mean of the estimated slope functions with the red solid line representing the true slope function. The middle panel illustrates the sample distributions of $\widehat{\beta}_{\tau,(k)} - \beta_\tau$ from approaches (a)–(d) under Model I at selected quantile levels 0.1, 0.5, and 0.9, where $\widehat{\beta}_{\tau,(k)}$ is the estimated coefficient from the kth Monte Carlo sample, and β_τ is the true value. The bottom panel is its counterpart for Model II. As expected, the naive estimator (dotted gray lines) is badly biased for both normally distributed $\mathbf{x}$ and skewed $\mathbf{x}$. The regression calibration estimator (dashed gray lines) worked fairly well for normally distributed $\mathbf{x}$, however, it is badly biased for skewed $\mathbf{x}$. Such bias is more evident under more severe contamination. The proposed method successfully corrected the bias, and brings the estimates fairly close to the true values with all the quantile levels. Moreover, the difference between the estimates using the true $f(x|w)$ and the estimated $f(x|w)$ are small. Similar results (not presented in this paper) were obtained for the more severe attenuation rate $R = 2/3$.

5.5 Simulation When the Estimated $f(\mathbf{x}|\mathbf{w})$ Deviates From the True Function

In the previous simulation study, the conditional density $f(x|w)$ is estimated in the correct model setting. The mean difference between the estimates from the true density and the estimated density is rather small. It is of interest to assess how sensitive the method is to the estimation of $f(x|w)$. To do this, we replace the density function of U in both Models I and II by a t distribution with 3 degrees of freedom, that is:

Model I*.

$$w_i = x_i + U_i, \quad \text{with } x_i \sim N(4, 1) \text{ and } U_i \sim (\sigma_u/\sqrt{3})t_3.$$

Model II*.

$$\log(w_i) = \log(x_i) + U_i,$$

$$\text{with } \log(x_i) \sim N(2, 1) \text{ and } U_i \sim (\sigma_u/\sqrt{3})t_3.$$

The scalar $\sigma_u/\sqrt{3}$ is used to maintain the same error variance as in the previous simulation. We repeated exactly the same estimation procedure (d) assuming U is Gaussian. Consequently, the estimated $f(x|w)$ deviates systematically from the underlying true density. In Figure 3, we plot the resulting estimated slope functions under misspecified Models I and II with attenuation $R = 4/5$, and compare them with the naive and regression calibration estimates. The solid dark-gray lines are estimated coefficients from the misspecified $f(x|w)$, and the long-dashed dark-gray lines are those from the correctly estimated ones. As expected, the estimated coefficients from the misspecified $f(x|w)$ do exhibit some modest bias, but still much less biased than the naive estimates (the dotted light-gray lines) and those from regression calibration (the dot-dashed light-gray lines).

6. APPLICATION

6.1 Data, Model, and Algorithm

We applied our method to part of the National Collaborative Perinatal Project (NCPP; Terry, Wei, and Essenman 2007). The data set included 232 women who were born at Columbia Presbyterian Medical Center from 1959–1963. Their growth measurements, weight, and height were carefully taken by clinical researchers at birth and at 4 months, 1 year, and 7 years. These ages are known to be critical times for growth. The public health researchers were interested in studying the long-term impact of early growth on adult body size. We consider therefore the quantile regression model

$$Q_\tau(Y) = \beta_{\tau,0} + \beta_{\tau,1}\mathbb{S}_{t_1} + \beta_{\tau,2}\mathbb{S}_{t_2} + \beta_{\tau,3}\mathbb{S}_{t_3} + \beta_{\tau,4}\mathbb{S}_{t_4}, \quad (16)$$

where Y is an individual subject's body mass index (BMI) at age 20, $\mathbb{S}_t$ stands for her weight at age t, and (t_1, t_2, t_3, t_4) are the four target ages at birth, 4 months, 1 year, and 7 years.

However, these subjects did not in fact all attend clinic at exactly these scheduled times. Since children grow relatively quickly especially at young ages, one or two week's deviation from the target time may result in substantial measurement error in S_t. If we pretend the actual observation times are the true ones, the coefficient estimates will therefore be biased, as we have demonstrated in our simulations. We apply our method to obtain consistent estimation of the β_τ's.

Model for $f(\mathbf{x}|\mathbf{w})$. Suppose $\mathbb{S}_{i,j}$ is the jth measurement of the ith subject taken at age $t_{i,j}$. We assume that they are observed from a underlying weight path of the ith subject, denoted as $\mathbb{S}_i(t)$. Let $\mathbf{T}_0 = (t_1, t_2, t_3, t_4)$ be the set of target ages; and $\mathbf{T}_i = (t_{i,1}, t_{i,2}, t_{i,3}, t_{i,4})$ be the actual measurement ages of the ith subject. Note that $\mathbb{S}_{t_1}$ is birth weight and is accurately measured at birth, that is, $t_{i,1} \equiv t_1$. We have three covariates of inter-

(a) Estimated slope coefficients from methods (a)–(d); $R = 0.8$

(b) Box plot of $\widehat{\beta}_{\tau,2} - \beta_{\tau,2}$'s from methods (a)–(d) at quantile level 0.1, 0.5, and 0.9; Model I, $R = 0.8$

(c) Box plot of $\widehat{\beta}_{\tau,2} - \beta_{\tau,2}$'s from methods (a)–(d) at quantile level 0.1, 0.5, and 0.9; Model II, $R = 0.8$

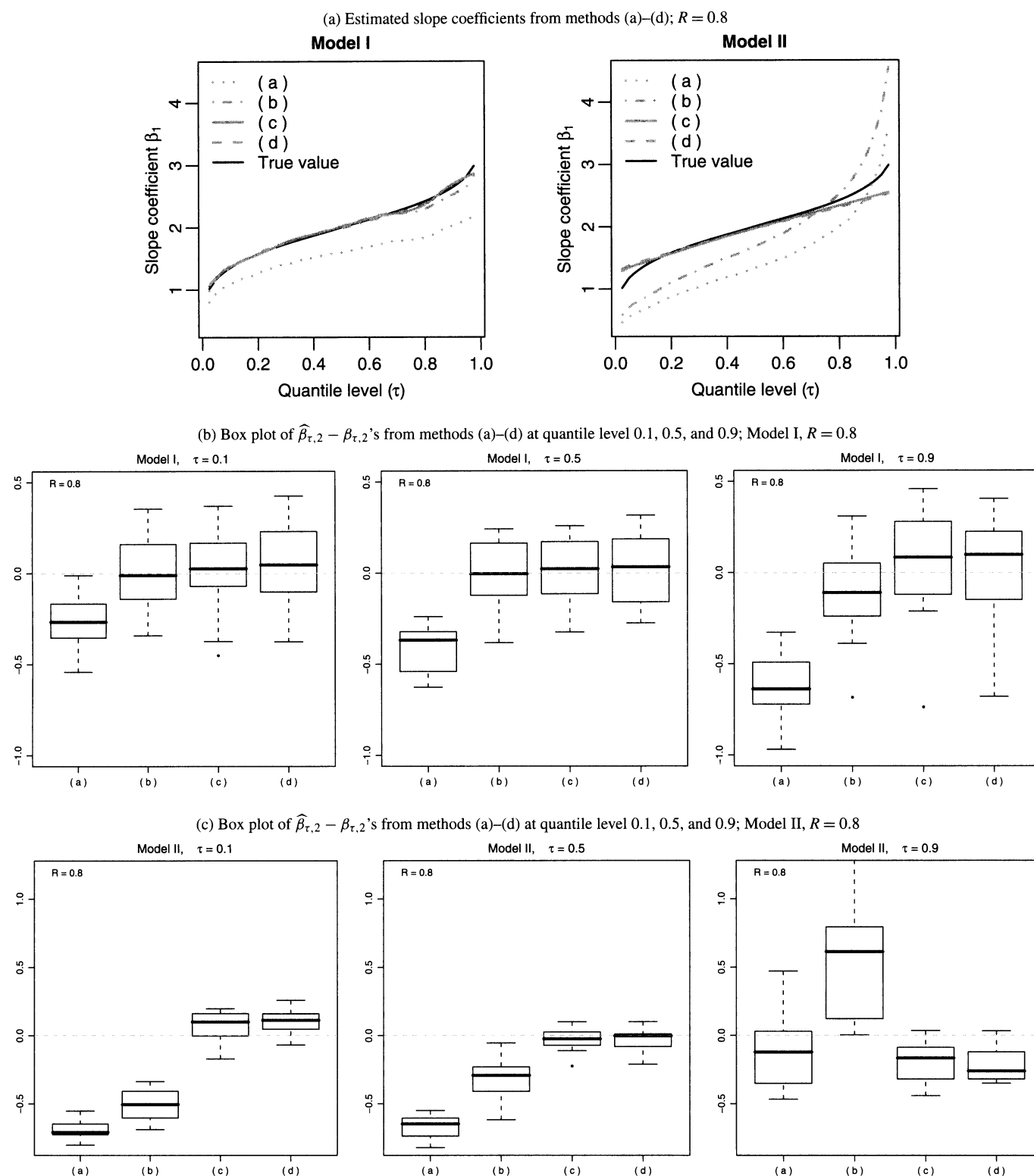

Figure 1. Comparison of the estimated slope coefficients from methods (a)–(d) under the contamination rate $R = 0.8$. In sub-figure (a), the black curve is the true coefficient function. The gray solid and dashed lines are the estimated coefficient functions from the proposed method using the true and estimated $f(x|w)$ respectively. The gray dotted line is the estimated coefficients from the naive method; the gray dash-dotted line is that of the regression calibration method. Note that the box plots are box plots of biases.

est that are unobserved, denoted as $\mathbf{x}_i = \{\mathbb{S}_i(t_2), \mathbb{S}_i(t_3), \mathbb{S}_i(t_4)\}^\top$. In addition, we denote $\mathbf{w}_i = \mathbb{S}_i(\mathbf{T}_i) = \{\mathbb{S}_i(t_1), \mathbb{S}_i(t_{i,2}), \mathbb{S}_i(t_{i,3}), \mathbb{S}_i(t_{i,4})\}^\top$ as the actual weights at birth and the individual measurement times $t_{i,2}, t_{i,3}$, and $t_{i,4}$. Finally, we estimate the density $f(\mathbf{x}_i|\mathbf{w}_i)$ by the following linear mixed model:

$$\log(\mathbb{S}_{i,j}) = \alpha_{i,0} + \alpha_{i,1}t_{i,j} + \alpha_{i,2}t_{i,j}^2 + \alpha_{i,3}\mathbb{S}_{i,1} + e_{i,j},$$

$$i = 1, \ldots, n \text{ and } j = 2, 3, 4. \quad (17)$$

(a) Estimated intercept coefficients from methods (a)–(d); $R = 0.8$

(b) Box plot of $\widehat{\beta}_{\tau,1} - \beta_{\tau,1}$'s from methods (a)–(d) at quantile level 0.1, 0.5, and 0.9; Model I, $R = 0.8$

(c) Box plot of $\widehat{\beta}_{\tau,1} - \beta_{\tau,1}$'s from methods (a)–(d) at quantile level 0.1, 0.5, and 0.9; Model II, $R = 0.8$
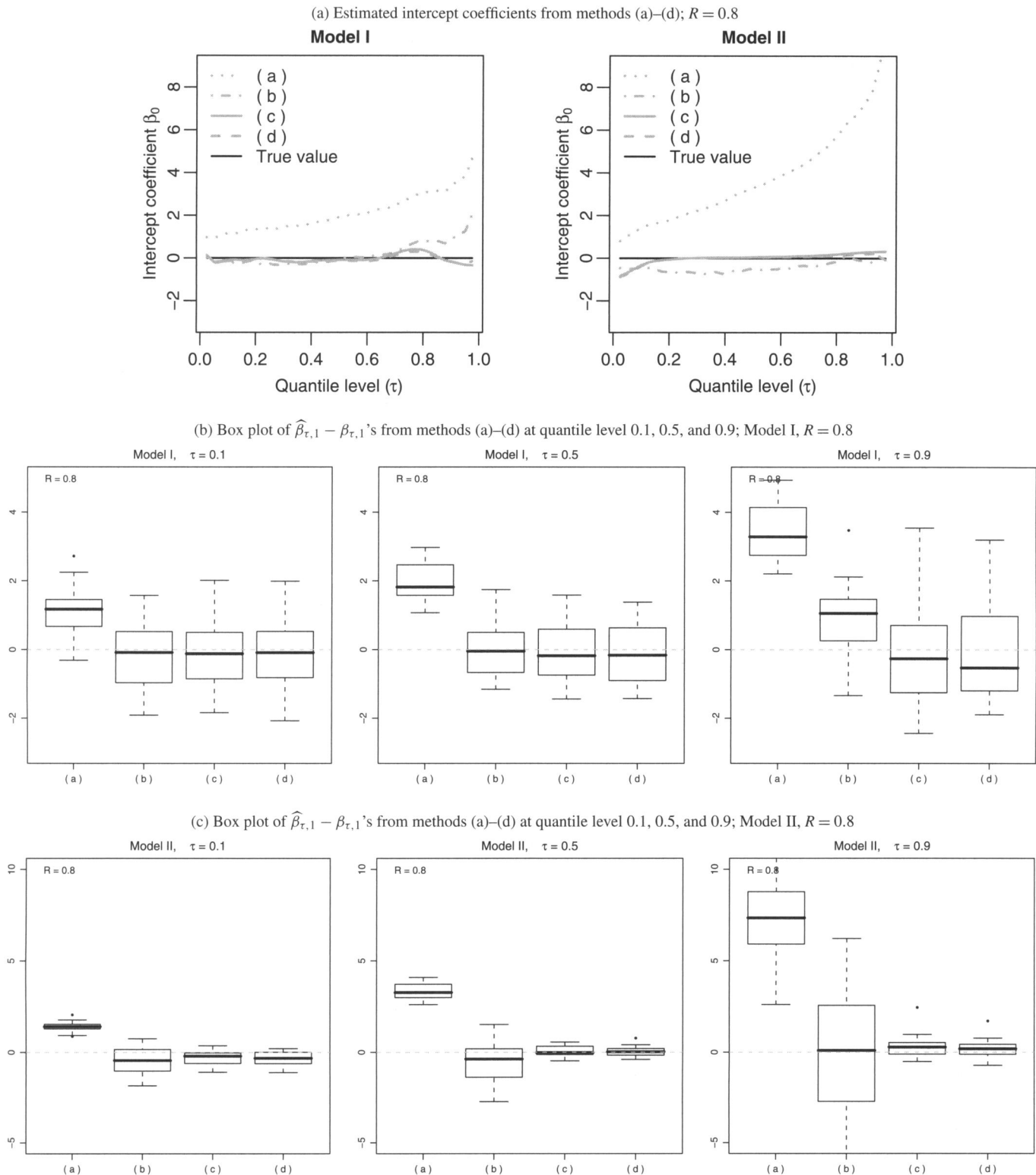

Figure 2. Comparison of the estimated intercept coefficients from methods (a)–(d) under the contamination rate $R = 0.8$. In sub-figure (a), the black curve is the true coefficient function. The gray solid and dashed lines are the estimated coefficient functions from the proposed method using the true and estimated $f(x|w)$, respectively. The gray dotted line is the estimated coefficients from the naive method; the gray dash-dotted line is that of the regression calibration method. Note that the box plots are box plots of biases.

We further assume that

$$(\alpha_{i,0}, \alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3})^\top = N\{(\alpha_0, \alpha_1, \alpha_2, \alpha_3)^\top, \Sigma_\alpha\},$$

$$e_{i,j} = N(0, \sigma_\gamma^2).$$

Model (17) assumes that the logarithm of $\mathbb{S}_i(t)$ is a Gaussian process with mean $\mu_i(t) = \alpha_{i,0} + \alpha_{i,1}t + \alpha_{i,2}t^2 + \alpha_{i,3}\mathbb{S}_{i,1}$ and variance $(1, t, t^2, \mathbb{S}_{i,1})\Sigma_\alpha(1, t, t^2, \mathbb{S}_{i,1})^\top + \sigma_\gamma^2$. The covariance
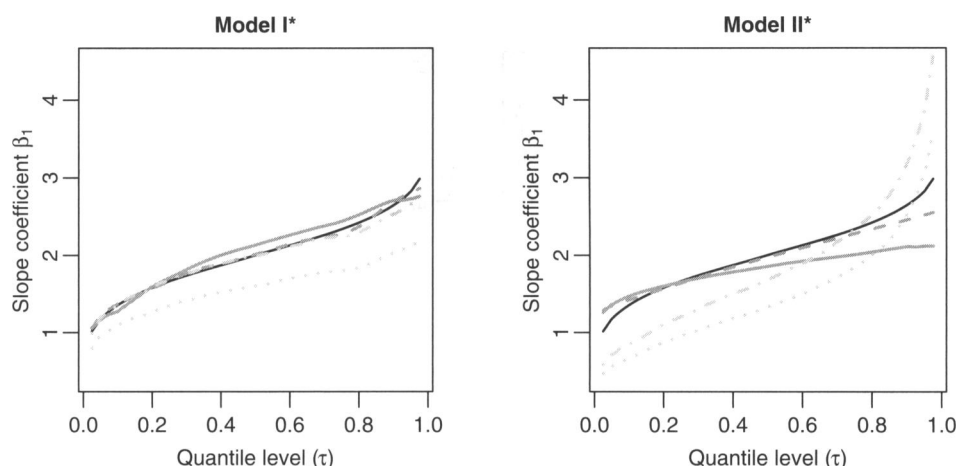
Figure 3. Comparison of the estimated slope coefficients when the measurement error model is misspecified with moderately contaminated rate $R = 4/5$. The subfigure on the left compares the estimated slope functions of Model I, while the one on the right compares those Model II. In all the subfigures, the sold black lines are the true slope coefficients. The solid dark-gray lines are estimators using the proposed method but with the misspecified $f(x|w)$, and the dashed dark-gray ones are estimators using the proposed method for the correct model. As a comparison, the dotted light-gray lines are the naive estimates and the dot-dashed light-gray ones are from regression calibration.

between $\mathbb{S}_i(t)$ and $\mathbb{S}_i(s)$ is $(1, t, t^2, \mathbb{S}_{i,1}) \Sigma_\alpha (1, s, s^2, \mathbb{S}_{i,1})^\top$. The log-transform on $\mathbb{S}$ is used due to the skewness of weight.

It then follows that $\mathbf{x}_i$ has a log-normal distribution, that is,

$$\log(\mathbf{x}_i) \sim \mathrm{N}(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i), \tag{18}$$

where

$$\widehat{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_i(\mathbf{T}_0) + \boldsymbol{\Sigma}_i(\mathbf{T}_0, \mathbf{T}_i) \boldsymbol{\Sigma}_i(\mathbf{T}_i)^{-1} \{\mathbf{w}_i - \boldsymbol{\mu}(\mathbf{T}_i)\},$$

$$\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}(\mathbf{T}_0) - \boldsymbol{\Sigma}_i(\mathbf{T}_0, \mathbf{T}_i) \boldsymbol{\Sigma}_i(\mathbf{T}_i)^{-1} \boldsymbol{\Sigma}_i(\mathbf{T}_0, \mathbf{T}_i)^\top.$$

Algorithmic Details. We choose $\Omega = \{\frac{1}{41}, \frac{2}{41}, \ldots, \frac{40}{41}\}$ to be a set of 40 evenly spaced quantile levels. The initial estimate $\widehat{\theta}_n^{(0)}$ is obtained by regressing Y over the observed weights assuming the all the $t_{i,j} = t_j$, the target ages. Note that $\mathbf{x}$ is three dimensional in this example, so evaluating $f(\mathbf{x}_i|\mathbf{w}_i)$ on a three-dimensional grid can be computationally undesirable. Consequently, we modify Step 3 in our iterative algorithm by using an importance sampling strategy as follows:

Step 3′(a). For each subject i, generate $m = 500$ $\widetilde{\mathbf{x}}_{i,j}$'s from the log-normal distribution in (18), the conditional distribution of $\mathbf{x}_i$ given $\mathbf{w}_i$.

Step 3′(b). Calculate the importance sampling weights for $\widetilde{\mathbf{x}}_{i,j}$ as

$$\delta_{i,j} = \frac{f^{(\nu)}(\widetilde{\mathbf{x}}_{i,j}|\mathbf{w}_i, y_i)}{f(\widetilde{\mathbf{x}}_{i,j}|\mathbf{w}_i)}$$

$$= \frac{f^{(\nu)}(y_i|\widetilde{\mathbf{x}}_{i,j}, w_{i,1})}{\int f^{(\nu)}(y_i|\widetilde{\mathbf{x}}_{i,j}, w_{i,1}) f(\widetilde{\mathbf{x}}_{i,j}|\mathbf{w}_i) \, d\mathbf{x}_i}.$$

Step 3′(c). Estimate $\widehat{\boldsymbol{\beta}}_{\tau_k}^{(\nu)}$'s by solving the following weighted estimating equations:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \Psi_{\tau_k}(y_i - (1, w_{i,1}, \widetilde{\mathbf{x}}_{i,j}^\top)\boldsymbol{\beta}_{\tau_k})(1, w_{i,1}, \widetilde{\mathbf{x}}_{i,j}^\top)^\top \delta_{i,j} = 0,$$

$$k = 1, \ldots, k_n.$$

We used the same convergence rule as in the simulation study.

6.2 Results

The resulting coefficient estimates are presented in Figure 4 (the solid lines). In Figure 4, for comparison we also plot the estimated coefficients from two alternative approaches. The long-dashed lines are the estimated coefficients from the original uncorrected quantile regression. Another possible approach to correct for measurement error is to fit an interpolation spline for each individual growth path, and use its fitted values at the target ages (t_2, t_3, t_4) as $\mathbf{x}$, the true weights of interests. This approach smooths individual paths to reduce the measurement errors, and has been used in Terry, Wei, and Essenman (2007). The resulting coefficient estimates using this interpolation approach are displayed in Figure 4 as short-dashed lines. We note that the estimated coefficients from our method differ considerably from the naive estimator, especially the coefficients for weights at 1 year and 7 years. In contrast, the fits from the interpolation approach agree with the naive estimators fairly well. Comparing with our estimates, both the naive estimates and the interpolation approach appear to underestimate the impact of weight at age 4 months at upper quantiles, but overestimate the impact of weights at 1 and 7 years at upper quantiles. We performed a bootstrap analysis and found that the differences between two estimates for weight at 1 year are significantly different from each other at quantile levels from 0.71 to 0.85 (as presented in Table 1). The observed differences for weight at 7 years are comparably smaller, and not of statistical significance.

6.3 Further Investigation via Simulated NCPP-Like Data

To understand the observed differences seen in the NCPP example, we generated synthetic data sets based on the estimated models (16), (17), and (18). To mimic the NCPP data, we chose the same sample size, $n = 232$, and use the original birth weight $\mathbb{S}_{i,1}$ and actual measurement ages $t_{i,2}, t_{i,3}$, and $t_{i,4}$ for all the subjects. We then generated the weights $\mathbb{S}_{i,2}, \mathbb{S}_{i,3}$, and $\mathbb{S}_{i,4}$ from the estimated model (17). In addition, we generated $\mathbf{x}_i$, the un-
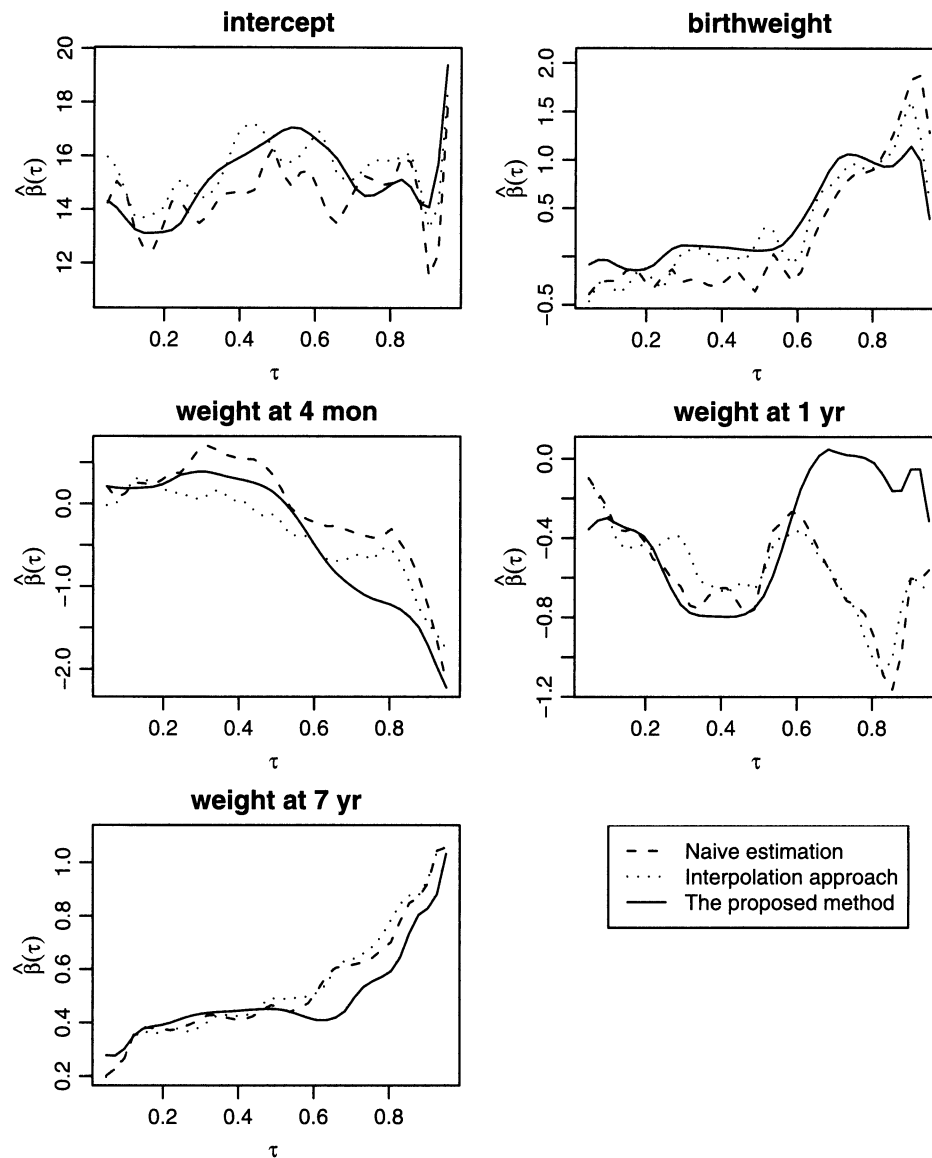
Figure 4. Estimated coefficients in the NCPP study and its comparison to original QR estimates and the interpolation approach.

derlying "true" weights at ages 4 months, 1 year, and 7 years from the estimated model (17). Finally, we generated BMI at 20 years from the estimated quantile model (16) by

$$Y_i^* = \widehat{\beta}_0(u_i) + \widehat{\beta}_1(u_i)\mathbb{S}_{i,1} + \widehat{\beta}_2(u_i)\mathbb{S}_{i,2} + \widehat{\beta}_3(u_i)\mathbb{S}_{i,3} + \widehat{\beta}_4(u_i)\mathbb{S}_{i,4},$$

where $\widehat{\beta}_k(\cdot)$ are smoothed the quantile coefficient process from the estimated quantile models (16), and u_i are iid random draws from Uniform$(0, 1)$. The generated datasets follow the distribution characterized by Models (16), (17), and (18).

Table 1. Differences between naive and the proposed estimates, and their bootstrap standard errors. $\widehat{\beta}$ is the estimated coefficient using the proposed method, while $\widehat{\beta}_{(a)}$ is that from the naive estimation. The standard error of $(\widehat{\beta} - \widehat{\beta}_{(a)})$ is calculated based on 50 bootstrap sample

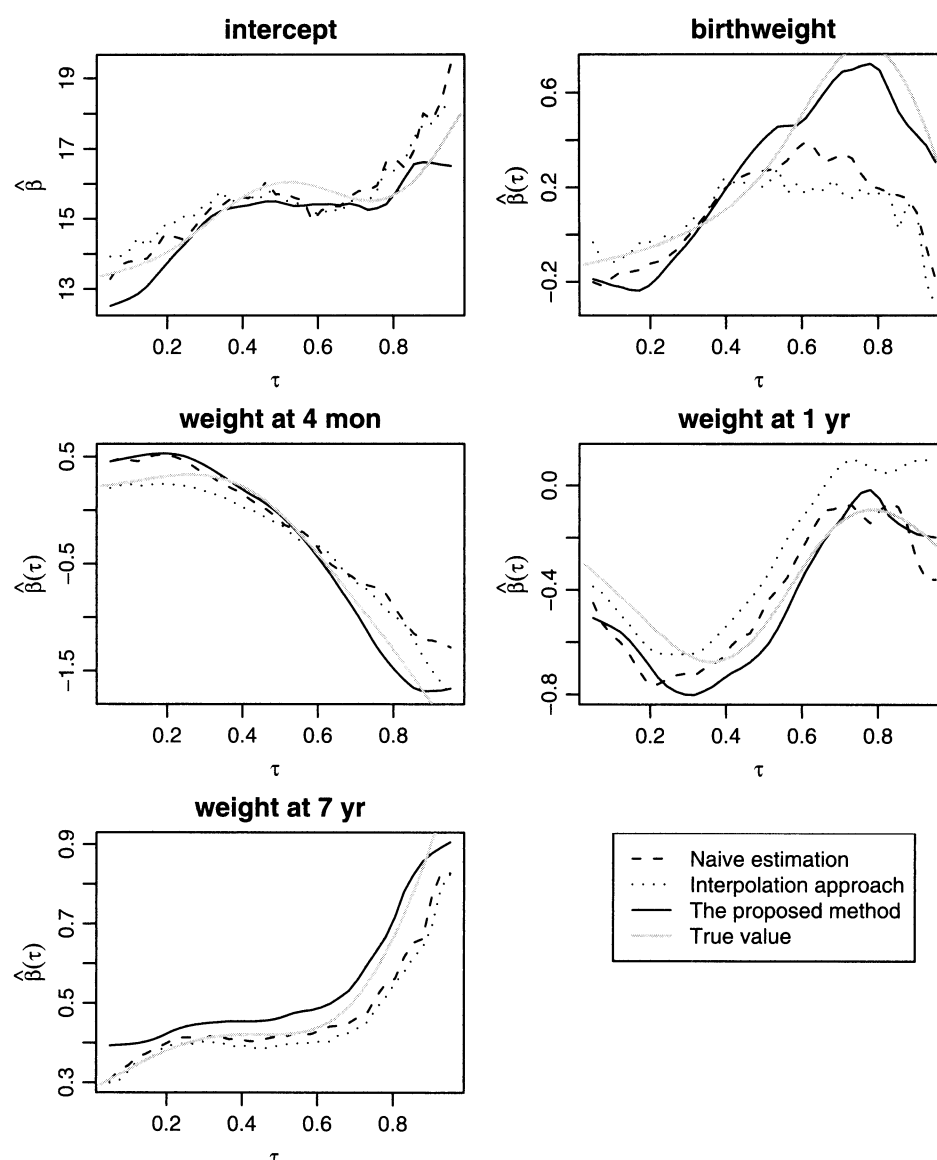| Weight at		Quantile level (τ)								
		0.66	0.68	0.71	0.73	0.76	0.78	0.80	0.83	0.85
1 yr	$\widehat{\beta} - \widehat{\beta}_{(a)}$	0.48	0.61	0.66	0.76	0.74	0.79	0.84	1.04	0.98
	Bootstrap SE	0.40	0.41	0.37	0.36	0.39	0.42	0.40	0.40	0.50
	p-value	0.114	0.069	0.038	0.019	0.029	0.030	0.018	0.005	0.024
7 yrs	$\widehat{\beta} - \widehat{\beta}_{(a)}$	−0.19	−0.17	−0.12	−0.09	−0.08	−0.12	−0.09	−0.16	−0.12
	Bootstrap SE	0.10	0.10	0.10	0.10	0.11	0.10	0.09	0.09	0.15
	p-value	0.969	0.960	0.883	0.827	0.767	0.881	0.828	0.958	0.973

Figure 5. Estimated coefficients using synthetic data sets and its comparison to original QR estimates and the interpolation approach.

We generated 20 synthetic data sets, and repeatedly estimated the coefficients using the naive approach, the interpolation approach and our proposed method. The results are presented in Figure 5. In Figure 5, the solid gray line represents the true coefficient process that is used to generate the synthetic data sets. The solid black lines are the estimated coefficients using our approach, the long-dashed lines are those from naive estimation, and the short-dashed lines used the interpolation approach. The naive estimates are close to the true values except for the birth weight coefficient, for which the naive estimates seriously underestimated the efficient at the upper quantiles. In contrast, our method appeared to largely correct the bias, as we had hoped and as the simulation led us to expect.

7. DISCUSSION

In this paper, we have proposed a new method to estimate linear quantile regression models when the covariates are measured with error. The method is based on constructing estimating equations jointly for all quantile levels $\tau \in (0, 1)$, and

avoids specifying a distribution for the response given the true covariates. The heart of the algorithm is an EM-type computation. The resulting estimated coefficients are consistent and, if the underlying function is characterized by a spline, asymptotically normally distributed. Numerical results show that the new estimator is promising in terms of correcting the bias arising from the errors-in-covariates, and compares favorably with alternative approaches.

In both simulation studies and an empirical data example, we estimated the conditional density $f(\mathbf{x}|\mathbf{w})$ based on a Gaussian model with appropriate transformation to accommodate possible heteroscedasticity and skewness. There are many other ways to estimate this distribution when the scale is known such that $\mathbf{w}$ has mean $\mathbf{x}$, for example, deconvolution with or without heteroscedasticity; see Staudenmayer, Ruppert, and Buonaccorsi (2008) for multiple references to the rapidly growing deconvolution literature as well as a semiparametric Bayesian approach. Our results show that use of these methods will still lead to consistent estimation of the quantile function, although

asymptotic distributions would have to be addressed separately if such deconvolution approaches were to be used.

The methodology can be extended to longitudinal or clustered data within the working independence context; see for example He, Zhu, and Fung (2002). In such an approach, one ignores the correlations but fixes up the estimated covariance matrix to account for the correlated responses, using so-called sandwich method. That is, each individual contributes an estimating function to the overall estimating equation, and it is those estimating functions to which the sandwich method is applied.

APPENDIX

Recall that $\boldsymbol{\beta}(\tau)$ is a $(p+1)$ dimensional unknown quantile coefficient function on $(0, 1)$, and $\theta = \{\boldsymbol{\beta}_{\tau_k} : \tau_k \in \Omega\}$ is the set of its quantile coefficients on the quantile level set Ω. Without loss of generosity, we assume in our proof that $\tau_k = k/(k_n + 1)$ such that $\Omega = \{1/(k_n + 1), 2/(k_n + 1), \ldots, k_n/(k_n + 1)\}$. Recall that $S_n(\theta)$ are the working estimating equations defined in (10), and $S(\theta)$ is its expectation. Similarly, we denote $S_n^0(\theta)$ as the estimating equations defined in (4) at quantile levels Ω, and $S^0(\theta)$ its expectation. With this notation, we introduce Lemmas A.1 and A.2 which will be used for the proof of consistency.

Lemma A.1. Under Assumptions 1–2, we have

$$k_n^{-1} \| S(\boldsymbol{\theta}_0) - S^0(\boldsymbol{\theta}_0) \| = o(1). \tag{A.1}$$

Proof. We first decompose $\| S(\boldsymbol{\theta}_0) - S^0(\boldsymbol{\theta}_0) \|$ as

$$\| S(\boldsymbol{\theta}_0) - S^0(\boldsymbol{\theta}_0) \|$$

$$= \left\| n^{-1} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}_0) \right. \right.$$

$$\left. \left. \times \mathbf{x} \big[f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big] dx \Big| \mathbf{w}_i \right\} \right\|$$

$$= \left\| n^{-1} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}_0) \right. \right.$$

$$\times \mathbf{x} \big[f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big]$$

$$\left. \left. \cdot I\{\mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)} < y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} dx \Big| \mathbf{w}_i \right\} \right\|$$

$$+ \left\| n^{-1} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}_0) \right. \right.$$

$$\times \mathbf{x} \big[f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big]$$

$$\left. \left. \cdot I\{y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} dx \Big| \mathbf{w}_i \right\} \right\|$$

$$+ \left\| n^{-1} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}_0) \right. \right.$$

$$\mathbf{x} \big[f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big]$$

$$\left. \left. \cdot I\{y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)}\} dx \Big| \mathbf{w}_i \right\} \right\|$$

$$\hat{=} I + II + III. \tag{A.2}$$

By construction, $f(\mathbf{x}|y_i, \mathbf{w}_i, \boldsymbol{\theta}_0) = 0$ for $y_i \ge \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}$, therefore,

$$II = \left\| n^{-1} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} \Psi(y_i - \mathbf{x}^\top \boldsymbol{\theta}_0) \mathbf{x} f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \right. \right.$$

$$\left. \left. \cdot I\{y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} dx \Big| \mathbf{w}_i \right\} \right\|$$

$$\le n^{-1} \sum_{k=1}^{k_n} \sum_{i=1}^{n} E_{y_i} \big\{ E_{\mathbf{x}} \big\{ \| \Psi_{k/(k_n+1)}(y_i - \mathbf{x}^\top \boldsymbol{\beta}_{0,k/(k_n+1)}) \mathbf{x} \| $$

$$\cdot I\{y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} | y_i, \mathbf{w}_i \} \big\}$$

$$= n^{-1} \sum_{k=1}^{k_n} \sum_{i=1}^{n} E_{\mathbf{x}} \big\{ E_{y_i} \big\{ \| \Psi_{k/(k_n+1)}(y_i - \mathbf{x}^\top \boldsymbol{\beta}_{0,k/(k_n+1)}) \mathbf{x} \| $$

$$\cdot I\{y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} | \mathbf{x}, \mathbf{w}_i \} | \mathbf{w}_i \}$$

$$\le n^{-1} k_n \sum_{i=1}^{n} E_{\mathbf{x}} \big\{ 2p \| \mathbf{x} \| E_{y_i} \{ \cdot I\{y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} | \mathbf{x}, \mathbf{w}_i \} | \mathbf{w}_i \}$$

$$= n^{-1} k_n \sum_{i=1}^{n} E_{\mathbf{x}} \big\{ 2p \| \mathbf{x} \| \operatorname{Prob}(y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)} | \mathbf{x}) | \mathbf{w}_i \}$$

$$= n^{-1} k_n/(k_n + 1) \sum_{i=1}^{n} E_{\mathbf{x}}(2p \| \mathbf{x} \| | \mathbf{w}_i)$$

$$= O(1). \tag{A.3}$$

The second last equation above is followed by the surrogacy condition, and the last equation is due to the fact that $\operatorname{Prob}(y_i > \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)} | \mathbf{x}) = 1/(k_n + 1)$ for all i's. Since $E_{\mathbf{x}}(\| \mathbf{x} \| | \mathbf{w}_i)$ is bounded for all the i's according to Assumption 1, it follows that $k_n^{-1} II = o(1)$. Using similar arguments, we can also show that $k_n^{-1} III = o(1)$. In what follows, we show that $k_n^{-1} I = o(1)$. Note that I can be bounded by

$$I \le n^{-1} \sum_{k=1}^{k_n} \sum_{i=1}^{n} E_{y_i} \left\{ \int_{\mathbf{x}} 2p \| \mathbf{x} \| \big| f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big| \right.$$

$$\left. \times I\{\mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)} \le y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} dx \right\}.$$

Since $\mathbf{x}$ has bounded support as indicated in Assumption 1, a sufficient condition for $k_n^{-1} I = o(1)$, according to Scheffe's theorem (Scheffe 1947), is that, for any $\mathbf{x} \in \mathcal{X}$, the following holds:

$$\max_i \big| f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) - f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} \big|$$

$$\times I\{\mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)} \le y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} = o_p(1).$$

Since $f(\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\theta}_0) = f(y_i|\mathbf{x}; \boldsymbol{\theta}_0) f(\mathbf{x}|\mathbf{w}) / \int_{\mathbf{x}} f(y_i|\mathbf{x}; \boldsymbol{\theta}_0) f(\mathbf{x}|\mathbf{w}) dx$ and $f\{\mathbf{x}|y_i, \mathbf{w}_i; \boldsymbol{\beta}_0(\tau)\} = f\{y_i|\mathbf{x}; \boldsymbol{\beta}_0(\tau)\} f(\mathbf{x}|\mathbf{w}) / \int_{\mathbf{x}} f\{y_i|\mathbf{x}; \boldsymbol{\beta}_0(\tau)\} f(\mathbf{x}|\mathbf{w}) dx$, it again suffices to show that $\forall \mathbf{x} \in \mathcal{X}$,

$$\max_i \big| f(y_i|\mathbf{x}; \boldsymbol{\theta}_0) - f\{y_i|\mathbf{x}; \boldsymbol{\beta}_0(\tau)\} \big|$$

$$\times I\{\mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)} \le y_i \le \mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}\} = o_p(1). \tag{A.4}$$

Let $F_{\mathbf{x}}(y_i) = \inf\{\tau : \mathbf{x}^\top \boldsymbol{\beta}_{0,\tau} \ge y_i\}$ be the quantile rank of y_i with respect to the probability measure induced by the quantile function $\mathbf{x}^\top \boldsymbol{\beta}_0(\tau)$, and let $h_{\mathbf{x}}(\tau) = 1/\mathbf{x}^\top \boldsymbol{\beta}_0'(\tau)$ be the density of y at the τth quantile. For any y_i that is bounded between $\mathbf{x}^\top \boldsymbol{\beta}_{0,1/(k_n+1)}$ and $\mathbf{x}^\top \boldsymbol{\beta}_{0,k_n/(k_n+1)}$, there exist a k_i such that $\mathbf{x}^\top \boldsymbol{\beta}_{0,k_i/(k_n+1)} \le y_i \le$

$\mathbf{x}^\top\boldsymbol{\beta}_{0,(k_i+1)/(k_n+1)}$. Consequently, the left side of (A.4) is equivalent to

$$\max_i \left| \frac{1}{k_n\mathbf{x}^\top(\boldsymbol{\beta}_{0,(k_i+1)/(k_n+1)} - \boldsymbol{\beta}_{0,k_i/(k_n+1)})} - \frac{1}{\mathbf{x}^\top\boldsymbol{\beta}_0'\{F_\mathbf{x}(y_i)\}} \right|$$

$$= \max_i \left| h_\mathbf{x}(\tau_i*) - h_\mathbf{x}\{F_\mathbf{x}(y_i)\} \right|$$

for some $k_i/(k_n+1) < \tau_i* < (k_i+1)/(k_n+1)$

$$= \max_i \left| h_\mathbf{x}'\{k_i/(k_n+1)\}O(k_n^{-1}) \right|. \tag{A.5}$$

According to Assumption 2,

$$h_\mathbf{x}'\{k_i/(k_n+1)\} < M\left(\frac{k_i}{k_n+1}\right)^{\nu_1}\left(1 - \frac{k_i}{k_n+1}\right)^{\nu_2}$$

$$< M\left(\frac{1}{k_n+1}\right)^{\nu_1}\left(1 - \frac{1}{k_n+1}\right)^{\nu_2}$$

$$+ M\left(\frac{k_n}{k_n+1}\right)^{\nu_1}\left(1 - \frac{k_n}{k_n+1}\right)^{\nu_2}$$

$$= O(k_n^{-\nu_1}) + O(k_n^{-\nu_2}). \tag{A.6}$$

Since $\nu_1, \nu_2 > -1$, (A.4) is implied by (A.6) and (A.5). The proof of Lemma A.1 is hence complete.

Lemma A.2. Recall that $\widetilde{S}_n(\boldsymbol{\theta})$ and $S_n(\boldsymbol{\theta})$ are two sets of score functions defined in (10) and (11), respectively. Let $S(\boldsymbol{\theta})$ be the expectation of $S_n(\boldsymbol{\theta})$. Then under Assumptions 1–4, for $k_n \to \infty$, $k_n n^{-1} \to 0$, we have the following uniform convergence:

$$\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widetilde{S}_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})\| = o_p(1) \quad \text{as } n \to \infty. \tag{A.7}$$

Proof. We first bound the left side of (A.7) by

$$\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widetilde{S}_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})\|$$

$$\leq \sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widetilde{S}_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|S_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})\|.$$

According to Assumption 1, $f(\mathbf{x}|\mathbf{w})$ is bounded away from infinity, and $\sup_\mathbf{w} |\widetilde{f}(\mathbf{x}|\mathbf{w}) - f(\mathbf{x}|\mathbf{w})| \to 0$ for any $\mathbf{x}$ as n goes to infinity. Therefore, for any i,

$$\left| f(y_i|\mathbf{x};\boldsymbol{\theta})\widetilde{f}(\mathbf{x}|\mathbf{w}_i) - f(y_i|\mathbf{x};\boldsymbol{\theta})f(\mathbf{x}|\mathbf{w}_i) \right| = o_p(1),$$

which further implies that $|\widetilde{f}(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\theta}) - f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\theta})| = o_p(1)$ for all i. Due to the boundness of $\mathbf{x}$, it follows from Sheffe's theorem that

$$\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widetilde{S}_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta})\| = o_p(1). \tag{A.8}$$

Therefore, to show (A.7), we only need to show that, for any ε,

$$\text{pr}\left(\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|S_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})\| > \varepsilon \right) \to 0, \tag{A.9}$$

as $n \to \infty$. In what follows, we will show (A.9) using Huber's chaining augment. Without loss of generality, we assume $\Theta\times\Omega = \bigcup_k\{\boldsymbol{\beta}: |\boldsymbol{\beta} - \boldsymbol{\beta}_{0,\tau_k}| < 1\}$. We partition the parameter space $\Theta\times\Omega$ into L_n disjoint small cubes Γ_l with diameters less than $q_n = C_1 k_n/n$, for some constant C_1. Let $\boldsymbol{\xi}_l$ be the center of the lth cube Γ_l. The probability of the left side of (A.9) is bounded by the sum of the following two probabilities, $P_1 + P_2$, where

$$P_1 = \text{pr}\left(\max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} k_n^{-1}\|S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\xi}_l) - S(\boldsymbol{\theta}) + S(\boldsymbol{\xi}_l)\| > \varepsilon/2 \right);$$

$$P_2 = \text{pr}\left(\max_{1\leq l\leq L_n} k_n^{-1}\|S_n(\boldsymbol{\xi}_l) - S(\boldsymbol{\xi}_l)\| > \varepsilon/2 \right).$$

We first note that

$$\|S_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\xi}_l)\|$$

$$\leq \left\| n^{-1}\sum_{i=1}^n \int_\mathbf{x} \{\Psi_\tau(y_i - \mathbf{x}^\top\boldsymbol{\theta}) - \Psi_\tau(y_i - \mathbf{x}^\top\boldsymbol{\xi}_l)\}\mathbf{x} \right.$$

$$\left. \cdot f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)\,d\mathbf{x} \right\|$$

$$+ \left\| n^{-1}\sum_{i=1}^n \int_\mathbf{x} \Psi_\tau(y_i - \mathbf{x}^\top\boldsymbol{\theta})\mathbf{x} \right.$$

$$\left. \cdot \{f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\theta}) - f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)\}\,d\mathbf{x} \right\|$$

$$= SS_1 + SS_2.$$

Moreover, under Assumption 1, there exist a constant C such that

$$\max_l \sup_{\boldsymbol{\theta}\in\Gamma_l} SS_1$$

$$\leq \max_l \sup_{\boldsymbol{\theta}\in\Gamma_l} \left\| n^{-1}\sum_{i=1}^n \int_\mathbf{x} I\{|\mathbf{x}^\top\boldsymbol{\xi}_l - y_i| \leq |\mathbf{x}^\top(\boldsymbol{\xi}_l - \boldsymbol{\theta})|\}\mathbf{1}_{k_n}\cdot\mathbf{x} \right.$$

$$\left. \cdot f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)\,d\mathbf{x} \right\|$$

$$\leq k_n \max_l n^{-1}\sum_{i=1}^n \int_\mathbf{x} I\{|\mathbf{x}^\top\boldsymbol{\xi}_l - y_i| \leq \|\mathbf{x}\|q_n\}\|\mathbf{x}\|\cdot f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)\,d\mathbf{x}$$

$$\leq C\cdot k_n\cdot\max_l n^{-1}\sum_{i=1}^n \int_\mathbf{x} I\{|\mathbf{x}^\top\boldsymbol{\xi}_l - y_i| \leq C\cdot q_n\}$$

$$\cdot f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)\,d\mathbf{x}$$

$$= C\cdot k_n\cdot\max_l n^{-1}\sum_{i=1}^n \text{pr}\{|\mathbf{x}^\top\boldsymbol{\xi}_l - y_i| \leq C\cdot q_n|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l\}.$$

Let $g_i(z)$ be the density of $(\mathbf{x}^\top\boldsymbol{\xi}_l - y_i)$ given $(y_i,\boldsymbol{\xi}_l,\mathbf{w}_i)$. Then $g_i(z)$ is also continuous and bounded away from zero and infinity. Following the mean value theorem, for any i there exist z_i^* such that $\text{pr}(|\mathbf{x}^\top\boldsymbol{\xi}_l - y_i| \leq Cq_n|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l) = 2Cq_ng_i(z_i^*)$. It follows that $\max_l\sup_{\boldsymbol{\beta}\in\Gamma_l} k_n^{-1}SS_1 = O_p(q_n)$.

On the other hand, a sufficient condition for $k_n^{-1}SS_2 = o_p(1)$ is that $\max_i |f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\theta}) - f(\mathbf{x}|y_i,\mathbf{w}_i;\boldsymbol{\xi}_l)| = o_p(1)$. Due to the boundedness of $f(\mathbf{x}|\mathbf{w})$, it again suffices to show that

$$\max_i |f(y_i|\mathbf{x};\boldsymbol{\theta}) - f(y_i|\mathbf{x};\boldsymbol{\xi}_l)| = o_p(1),$$

where $f(y|\mathbf{x};\boldsymbol{\theta})$ and $f(y|\mathbf{x};\boldsymbol{\xi}_l)$ are two density functions that are induced by the quantile functions $\mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\theta}}(\tau)$ and $\mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\xi}_l}(\tau)$. Moreover, denote $F_{\boldsymbol{\theta}}(y) = \inf\{\tau: \mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\theta}}(\tau) > y\}$ and $F_{\boldsymbol{\xi}_l}(y) = \inf\{\tau: \mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\xi}_l}(\tau) > y\}$ as the inverse functions of $\mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\theta}}(\tau)$ and $\mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\xi}_l}(\tau)$. Since $|\boldsymbol{\theta} - \boldsymbol{\xi}_l| = O(k_n/n)$, $\sup_\tau |\mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\theta}}(\tau) - \mathbf{x}^\top\mathfrak{L}_{\boldsymbol{\xi}_l}(\tau)| = o(1)$. Let $\boldsymbol{\theta}[k]$ and $\boldsymbol{\xi}_l[k]$ stand for the subset of coefficients $\boldsymbol{\theta}$ and $\boldsymbol{\xi}_l$ at the quantile level $k/(k_n+1)$, then by construction,

$$\mathfrak{L}_{\boldsymbol{\theta}}(\tau) = \begin{cases} \boldsymbol{\theta}[1], & \tau < 1/(k_n+1) \\ \boldsymbol{\theta}[k_n], & \tau > k_n/(k_n+1) \\ \boldsymbol{\theta}[\lfloor\tau k_n\rfloor] + \dfrac{\boldsymbol{\theta}[\lfloor\tau k_n\rfloor + 1] - \boldsymbol{\theta}[\lfloor\tau k_n\rfloor]}{1/k_n} \\ \quad \times\left(\tau - \dfrac{\lfloor\tau k_n\rfloor}{k_n+1}\right), & \text{else.} \end{cases}$$

And $\mathfrak{L}_{\xi_l}(\tau)$ has the same format. The difference between $\mathbf{x}^\top \mathfrak{L}_\theta(\tau)$ and $\mathbf{x}^\top \mathfrak{L}_{\xi_l}(\tau)$ is then bounded by

$$
\left| \mathbf{x}^\top \mathfrak{L}_\theta(\tau) - \mathbf{x}^\top \mathfrak{L}_{\xi_l}(\tau) \right|
$$

$$
\leq \begin{cases}
|\xi_l[1] - \theta[1]|, & \tau < 1/(k_n + 1) \\
|\xi_l[k_n] - \theta[k_n]|, & \tau > k_n/(k_n + 1) \\
2|\xi_l[\lfloor \tau k_n \rfloor] - \theta[\lfloor \tau k_n \rfloor]| \\
\quad + |\xi_l[\lfloor \tau k_n \rfloor + 1] - \theta[\lfloor \tau k_n \rfloor + 1]|, & \text{else.}
\end{cases}
$$

Since $|\theta - \xi_l| = O(k_n/n)$ implies $\max_k |\theta[k] - \xi_l[k]| = O(k_n/n)$, then the boundness above implies $\sup_\tau |\mathbf{x}^\top \mathfrak{L}_\theta(\tau) - \mathbf{x}^\top \mathfrak{L}_{\xi_l}(\tau)| = o_p(1)$. Consequently, if we denote $L_n = \min(\mathbf{x}^\top \theta[1], \mathbf{x}^\top \xi_l[1])$ and $U_n = \max(\mathbf{x}^\top \theta[k_n], \mathbf{x}^\top \xi_l[k_n])$, then we have $\sup_{y \in [L_n, U_n]} |F_\theta(y) - F_{\xi_l}(y)| = o(1)$. Since we can write $f(y|\mathbf{x}; \theta) = 1/\mathbf{x}^\top \mathfrak{L}'_\theta \{F_\theta(y)\}$ and $f(y|\mathbf{x}; \xi_l) = 1/\mathbf{x}^\top \mathfrak{L}'_{\xi_l} \{F_{\xi_l}(y)\}$ for $y \in [L_n, U_n]$. $\sup_{y \in [L_n, U_n]} |f(y|\mathbf{x}; \theta) - f(y|\mathbf{x}; \xi_l)| = o(1)$. Moreover, by construction, $f(y|\mathbf{x}; \theta) = f(y|\mathbf{x}; \xi_l) = 0$ for any $y > U_n$ or $y < L_n$. Combining these facts, we have $\max_i |f(y_i|\mathbf{x}; \theta) - f(y_i|\mathbf{x}; \xi_l)| = o_p(1)$, which in turn implies $k_n SS_2 = o_p(1)$. Following a similar argument, we can also show that, $\sup_{\theta \in \Gamma_l} \|S(\theta) - S(\xi_l)\| = o(1)$. It then follows that $P_1 = o(1)$.

Let $\mathcal{Y}_i(l, k, m) = \int_\mathbf{x} \Psi_\tau(y_i - \mathbf{x}^\top \beta_{\tau_k}) x_m \cdot f(\mathbf{x}|y_i, \mathbf{w}_i; \theta) \, d\mathbf{x} \, I \theta \in \Gamma_l$. A sufficient condition for $P_2 = o(1)$ is that, for any β_{τ_k} and x_m,

$$
\mathrm{pr}\left\{ \max_{1 \leq l \leq L_n; 1 \leq k \leq k_n; 1 \leq m \leq p} \left| \sum_{i=1}^n \mathcal{Y}_i(l, k, m) - E\{\mathcal{Y}_i(l, k, m)\} \right| > \varepsilon \right\}
$$

$$
= o(1).
$$

Under Assumption 5, $|\mathcal{Y}_i(l, k, m)| < C$ for all the i's. Applying Bernstein's inequality to the probability term above, we have

$$
\mathrm{pr}\left(\max_{1 \leq l \leq L_n; 1 \leq k \leq k_n; 1 \leq m \leq p} n^{-1} \left| \sum_{i=1}^n \mathcal{Y}_i(l, k, m) - E\mathcal{Y}_i(l, k, m) \right| > \varepsilon \right)
$$

$$
\leq k_n \cdot p \cdot L_n \cdot \mathrm{pr}\left(n^{-1} \left| \sum_{i=1}^n \mathcal{Y}_i(l, k, m) - E\mathcal{Y}_i(l, k, m) \right| > \varepsilon \right)
$$

$$
\leq k_n \cdot p \cdot L_n \cdot \exp\left\{ -\frac{n^2 \varepsilon^2}{2nC^2 + (2/3)Cn\varepsilon} \right\} = o(1).
$$

We now have shown that both P_1 and $P_2 = o(1)$, which in turn implies that the uniform convergence (A.7) holds. Lemma A.2 is hence proved.

Proof of Consistency

Recall that $\widehat{\beta}_{\tau_k}$ is the $(p+1)$-dimensional estimated coefficient vector at quantile level τ_k based on the working estimating equations (10), and $\widehat{\beta}_{j,\tau_k}$ is its jth component. We further define $\widehat{\beta}_j(\tau)$ as the piece-wise linear function with $\widehat{\beta}_j(\tau_k) = \widehat{\beta}_{j,\tau_k}, \forall \tau_k \in \Omega$. To simplify the notation, we also denote $\widehat{\theta}_n = (\widehat{\beta}_{\tau_1}^\top, \widehat{\beta}_{\tau_2}^\top, \ldots, \widehat{\beta}_{\tau_{k_n}}^\top)^\top$ as the $p \times k_n$ dimensional estimated coefficient matrix.

For any $\delta > 0$, we define a compact set $B_\tau = \{\beta \in \mathbb{R}^{p+1} : \|\beta - \beta_{0,\tau}\| < \delta\}$, where $\beta_{0,\tau}$ is the true coefficients at the quantile level τ. We denote B_τ^c as its complementary set. Note that $S_n(\theta)$ are the working estimating equations defined in (10), and let $S(\theta)$ be its expectation. We define the distance

$$
d_n(\delta) = k_n^{-1} \left\{ \min_{\theta \in \{\Theta \cap B_\tau^c\} \otimes \Omega} \|S(\theta)\| - \|S(\theta_0)\| \right\} \quad \text{(A.10)}
$$

between the norm of the working estimating equations evaluated at the true coefficients θ_0 and the minimized norm when θ stays outside of $B_\tau \otimes \Omega$. In what follows, we show that $d_n(\delta) > 0$ under Assumptions 1 to 4.

Recall in Assumption 3 that θ_0 is the unique solution of $S^0(\theta) = 0$, we have $S^0(\theta_0) = 0$. Therefore, the convergence of (A.1) stated in

Lemma A.1 is equivalent to $k_n^{-1} \|S(\theta_0)\| = o(1)$. Moreover, since $\theta^* = (\beta_{\tau_1}^*, \ldots, \beta_{\tau_{k_n}}^*)$ is the unique solution of $S(\theta) = 0$ according to Assumption 3, it follows that $k_n^{-1} \|S(\theta_0) - S(\theta^*)\| \to 0$. Due to the continuity of $S(\cdot)$ and the uniqueness of θ^*, we have $k_n^{-1} \|\theta^* - \theta_0\| \to 0$, as n goes to infinity. Consequently, there exist K_δ, such that when $k_n > K_\delta$, we have that $k_n^{-1} \|\theta^* - \theta_0\| < \delta/2$, in other words, $\theta^* \in B_\tau \times \Omega$ for $k_n > K_\delta$. Due to the uniqueness of θ^*, for any $k_n > K_\delta$, we have

$$
d_n^*(\delta) = k_n^{-1} \left\{ \min_{\theta \in \{\Theta \cap B_\tau^c\} \otimes \Omega} \|S(\theta)\| - \|S(\theta^*)\| \right\} > 0. \quad \text{(A.11)}
$$

On the other hand, due to the continuity of $S(\cdot)$, for sufficiently larger k_n, we also have

$$
k_n^{-1} \|S(\theta^*) - S(\theta_0)\| < d_n^*(\delta)/2, \quad \text{(A.12)}
$$

Combining these (A.12) and (A.11), we have

$$
d_n(\delta) = k_n^{-1} \left[\min_{\theta \in \Theta \cap B_\tau^c} \|S(\theta)\| - \|S(\theta_0)\| \right] > d_n^*(\delta)/2 > 0, \quad \text{(A.13)}
$$

for sufficiently large k_n.

We now define the random event that

$$
E_n = \left\{ k_n^{-1} \max_{\theta \in \Theta \times \Omega} \left[\|\widetilde{S}_n(\theta) - S(\theta)\| \right] < d_n(\delta)/3 \right\},
$$

which, together with Assumption 4, implies that

$$
k_n^{-1} \|S(\widehat{\theta}_n)\| \leq k_n^{-1} \|\widetilde{S}_n(\widehat{\theta}_n)\| + d_n(\delta)/3 \quad \text{(A.14)}
$$

and

$$
k_n^{-1} \|\widetilde{S}_n(\theta_0)\| \leq k_n^{-1} \|S(\theta_0)\| + d_n(\delta)/3. \quad \text{(A.15)}
$$

Since $\widehat{\theta}_n$ is the minimizer of $\|\widetilde{S}_n(\theta)\|$, we have $\|\widetilde{S}_n(\widehat{\theta}_n)\| < \|\widetilde{S}_n(\theta_0)\|$, which, together with (A.14), shows that $k_n^{-1} \|S(\widehat{\theta}_n)\| \leq k_n^{-1} \|\widetilde{S}_n(\theta_0)\| + d_n(\delta)/3 \leq k_n^{-1} \|S(\theta_0)\| + 2d_n(\delta)/3$. Following Lemma A.2, $\lim_{n \to \infty} \mathrm{pr}(E_n) = 1$, which implies

$$
\lim_{n \to \infty} \mathrm{pr}\{k_n^{-1} \|S(\widehat{\theta}_n)\| \leq k_n^{-1} \|S(\theta_0)\| + 2d_n(\delta)/3\} \geq \lim_{n \to \infty} \mathrm{pr}(E_n) = 1.
$$

By the definition of B_τ and the fact that $d_n(\delta) > d_n^*(\delta)/2 > 0$, this in turn implies that $\lim_{n \to \infty} \mathrm{pr}(\widehat{\theta}_n \in B_\tau) = 1$, that is,

$$
\sup_{\tau \in [1/(k_n+1), k_n/(k_n+1)]} \|\widehat{\beta}_n(\tau) - \beta_0(\tau)\| = o_p(1).
$$

The consistency of $\widehat{\beta}_n(\tau)$ is hence proved.

Proof of Asymptotic Normality

Recall that $S_n(\theta) = n^{-1} \sum_{i=1}^n \Psi_{\text{new}}(y_i, \mathbf{w}_i, \theta)$, where $\Psi_{\text{new}}(y_i, \mathbf{w}_i, \theta) = \int_\mathbf{x} \Psi(y_i - \mathbf{x}^\top \theta) \otimes \mathbf{x} \cdot f(\mathbf{x}|y_i, \mathbf{w}_i; \theta) \, d\mathbf{x}$, and $\widetilde{S}_n(\theta)$ is its approximation replacing $f(\mathbf{x}|\mathbf{w}_i)$ in $f(\mathbf{x}|y_i, \mathbf{w}_i; \theta)$ by its estimate $\widetilde{f}(\mathbf{x}|\mathbf{w}_i)$. Following a similar argument as in Lemma A.2, we can show that for any decreasing sequence $d_n \to 0$, we have

$$
\sup_{\|\theta - \theta_0\| < d_n} n^{1/2} \|S_n(\theta) - S_n(\theta_0) - S(\theta) + S(\theta_0)\| = o_p(1). \quad \text{(A.16)}
$$

Theorem 1 implies, for fixed K, that $\widehat{\theta}_n$ is a consistent estimator of θ_0. The uniform convergence (A.16) hence implies

$$
n^{1/2} \|S_n(\widehat{\theta}_n) - S_n(\theta_0) - S(\widehat{\theta}_n) + S(\theta_0)\| = o_p(1). \quad \text{(A.17)}
$$

Note that $n^{1/2}\widetilde{S}_n(\widehat{\theta}_n) = o_p(1)$, and $\|\widetilde{S}_n(\widehat{\theta}_n) - S_n(\widehat{\theta}_n)\| = o_p(1)$, which is implied by (A.8) and Assumption 4, it follows that $S_n(\widehat{\theta}_n) \approx 0$ for large enough n. On the other hand, $S(\theta_0) = 0$ under Assumption 3. Therefore, the convergence (A.17) is equivalent to

$$
n^{1/2} \|S_n(\theta_0) - S(\widehat{\theta}_n)\| = o_p(1).
$$

Taylor expanding $S(\widehat{\theta}_n)$ around $S(\theta_0)$, we have

$$
\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\mathbf{D}_n^{-1} S_n(\theta_0) + o_p(1).
$$

Theorem 2 follows immediately from the Central Limit Theorem.

[Received August 2008. Revised February 2009.]

REFERENCES

Carriquiry, A. L. (2003), "Estimation of Usual Intake Distributions of Nutrients and Foods," *Journal of Nutrition*, 133, 601–608.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), Boca Raton, FL: Chapman & Hall/CRC Press.

Chesher, A. (2001), "Parameter Approximations for Quantile Regressions With Measurement Error," Working Paper CWP02/01, University College London, Dept. of Economics.

de Boor, C. (2001), *A Practical Guide to Splines, Applied Mathematical Sciences*, New York: Springer-Verlag.

Delaigle, A., Hall, P., and Meister, A. (2008), "On Deconvolution With Repeated Measurements," *The Annals of Statistics*, 36, 665–685.

Eckert, R. S., Carroll, R. J., and Wang, N. (1997), "Transformations to Additivity in Measurement Error Models," *Biometrics*, 53, 262–272.

He, X., and Liang, H. (2000), "Quantile Regression Estimates for a Class of Linear and Partially Linear Errors-in-Variables Models," *Statistica Sinica*, 10, 129–140.

He, X., Zhu, Z. Y., and Fung, W. K. (2002), "Estimation in a Semiparametric Model for Longitudinal Data With Unspecified Dependence Structure," *Biometrika*, 89, 579–590.

Hu, Y., and Schennach, S. M. (2008), "Identification and Estimation of Nonclassical Nonlinear Errors-in-Variables Models With Continuous Distributions Using Instruments," *Econometrica*, 76, 195–216.

Koenker, R. (2004), "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

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

Koenker, R., and Machado, J. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1309.

Koenker, R., and Xiao, Z. J. (2004), "Unit Root Quantile Autoregression Inference," *Journal of the American Statistical Association*, 99, 775–787.

Li, T., and Vuong, Q. (1998), "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

McLachlan, G. J., and Krishnan, T. (2008), *The EM Algorithm and Extensions*, New York: Wiley.

Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. (1997a), "A Semiparametric Transformation Approach to Estimating Usual Intake Distributions," *Journal of the American Statistical Association*, 91, 1440–1449.

Nusser, S. M., Carriquiry, A. L., Jensen, H. H., and Fuller, W. A. (1990), "A Transformation Approach to Estimating Usual Intake Distributions," in *Applied Statistics in Agriculture: Proceedings of the 1990 Kansas State University Conference on Applied Statistics in Agriculture*, eds. G. A. Milliken and J. R. Schwenke, Manhattan, KS, Kansas State University, pp. 120–132.

Nusser, S. M., Fuller, W. A., and Guenther, P. M. (1997b), "Estimating Usual Dietary Intake Distributions: Adjusting for Measurement Error and Nonnormality in 24-Hour Food Intake Data," in *Survey Measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwartz, and D. Trewin, New York: Wiley, pp. 689–709.

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of American Statistical Association*, 98, 1001–1012.

Scheffe, H. (1947), "A Useful Convergence Theorem in Probability Distributions," *Annals of Mathematical Statistics*, 18, 434–458.

Schennach, S. M. (2008), "Quantile Regression With Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. P. (2008), "Density Estimation in the Presence of Heteroskedastic Measurement Error," *Journal of the American Statistical Association*, 103, 726–736.

Terry, M. B., Wei, Y., and Essenman, D. (2007), "Maternal, Birth, and Early Life Influences on Adult Body Size in Women" (with discussions), *The American Journal of Epidemiology*, 166, 5–13. Author reply, 17–18.

Welsh, A. H. (1988), "Asymptotically Efficient Estimation of the Sparsity Function at a Point," *Statistics and Probability Letters*, 6, 427–432.