

# Estimation of sparse functional quantile regression with measurement error: a SIMEX approach

CARMEN D. TEKWE\*

*Department of Epidemiology and Biostatistics, Indiana University, Bloomington, IN 47405, USA*  
ctekwe@iu.edu

MENGLI ZHANG

*Department of Statistics, Oregon State University, Corvallis, OR 97331, USA*

RAYMOND J. CARROLL 

*Department of Statistics, Texas A&M University, College Station, TX 77843, USA*

YUANYUAN LUAN

*Department of Epidemiology and Biostatistics, Indiana University, Bloomington, IN 47405, USA*

LAN XUE

*Department of Statistics, Oregon State University, Corvallis, OR 97331, USA*

ROGER S. ZOH

*Department of Epidemiology and Biostatistics, Indiana University, Bloomington, IN 47405, USA*

STEPHEN J. CARTER

*Department of Kinesiology, Indiana University, Bloomington, IN 47405, USA*

DAVID B. ALLISON

*Department of Epidemiology and Biostatistics, Indiana University, Bloomington, IN 47405, USA*

MARCO GERACI 

*MEMOTEF Department, School of Economics, Sapienza - University of Rome, Rome, Italy and  
Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA*

## SUMMARY

Quantile regression is a semiparametric method for modeling associations between variables. It is most helpful when the covariates have complex relationships with the location, scale, and shape of the outcome distribution. Despite the method's robustness to distributional assumptions and outliers in the outcome, regression quantiles may be biased in the presence of measurement error in the covariates. The impact of function-valued covariates contaminated with heteroscedastic error has not yet been examined previously;

\*To whom correspondence should be addressed.

although, studies have investigated the case of scalar-valued covariates. We present a two-stage strategy to consistently fit linear quantile regression models with a function-valued covariate that may be measured with error. In the first stage, an instrumental variable is used to estimate the covariance matrix associated with the measurement error. In the second stage, simulation extrapolation (SIMEX) is used to correct for measurement error in the function-valued covariate. Point-wise standard errors are estimated by means of nonparametric bootstrap. We present simulation studies to assess the robustness of the measurement error corrected for functional quantile regression. Our methods are applied to National Health and Examination Survey data to assess the relationship between physical activity and body mass index among adults in the United States.

**Keywords:** Functional data analysis; Obesity; Physical activity; Spline basis splines; Wearable accelerometer.

## 1. INTRODUCTION

Quantile regression (Koenker and Bassett, 1978) assesses the effects of covariates on quantiles of an outcome of interest. In contrast, mean regression does not go beyond location-shift effects. An exception is when the mean regression model is based on a distribution whose location parameter is also related to the distribution's scale and/or shape. Such is the case with the Poisson distribution, which however imposes a rigid structure on moments of higher order (Winkelmann, 2006; Geraci, 2016). When parametric approaches are too restrictive or inappropriate, quantile regression is a flexible alternative that allows assessments of covariate effects across the entire distribution of an outcome, including the lower and upper tails of the distribution, leading to a complete examination of the location, scale, and shape of the conditional distribution (Geraci, 2016). Quantile regression models have been applied in a variety of settings including finance, biostatistics, economics, and ecology (Yu and others, 2003; Tsai, 2012; Baur and others, 2012; Machado and Mata, 2005; Martins and Pereira, 2004; Fitzenberger and others, 2022; Planque and Buffaz, 2008; Schröder and others, 2005; Bottai and others, 2014; Geraci and Bottai, 2007).

In obesity studies, body mass index (BMI) is often used as an indicator to identify subjects at risks for obesity. Therefore, assessment of the impacts of interventions or covariates on obesity requires analysis of data from individuals whose BMI falls in the upper quantile of the BMI distribution (Koenker and Hallock, 2001; Bottai and others, 2014; Geraci and Bottai, 2007). Despite that, most statistical analysis of BMI in obesity studies use mean regression-based approaches where the effects of covariates on BMI are estimated using ordinary least squares methods (Bottai and others, 2014). A drawback to the use of mean regression approaches is that they focus on modeling the mean of BMI, even though most research on obesity focuses on the upper quantiles of the distribution. Thus, standard analytic techniques based on mean regression models, provide incomplete answers to questions related to individuals at high risk for obesity, that is, individuals whose BMI values fall in the upper quantiles of the BMI distribution. To answer obesity-related questions, modeling of the full distribution of BMI is crucial. Wearable technology is increasingly used to monitor physical activity when researchers assess how lifestyle interventions contribute to reducing the risk of obesity. Wearable devices produce continuous and objective physical activity data that are best viewed as functions or curves, rather than as vectors. Current available methods in functional data analysis can be applied productively to such data; however, there are some unanswered statistical questions that must first be answered to permit more accurate assessments of physical activity in obesity and biomedical studies. Because true patterns of physical activity are not directly observable, wearable devices are used to monitor physical activity, thus introducing measurement error. Measurement error arises in device-based measures of physical activity due to several reasons including: (1) limitations in the assessment of certain forms of exercises such nonambulatory activities; (2) variability in the device's ability to predict physical activity intensity at various levels; and (3) errors related to the prediction equations used to predict physical activity (Corder and others, 2007; Bassett, 2012; Robertson and others, 2011;

Crouter and others, 2006; Jacobi and others, 2007; Warolin and others, 2012; Rothney and others, 2008; Jadhav and others, 2021). Also, we have previously shown that not accounting for potential measurement error in assessing the effects of wearable device-based physical activity measures on health outcomes can lead to biased and underestimated effects of physical activity (Tekwe and others, 2019). Additionally, data collected from wearable devices are usually not simple vectors or scalar values but rather curvilinear functions of time that require the use of functional data approaches in their analyses (Silverman and Ramsay, 2005). Finally, little is known of how measurement errors with complex heteroscedastic error structures influence conclusions drawn regarding the role of physical activity in obesity prevention. The methods proposed in this manuscript will address these analytic limitations.

There is a rich literature on statistical approaches to mean regression with scalar-valued covariates measured with error; however, comparatively little work has been done to address biases associated with measurement error in quantile regression (Wang and others, 2012; Wei and Carroll, 2009). He and Liang (2000) proposed an estimation procedure based on minimizing the quantile loss function of the *orthogonal* residuals from a model with an imprecisely observed scalar-valued covariate. Their proposed method assumed that the response errors and measurement errors were independent and followed the same spherically symmetric distribution (He and Liang, 2000). Wei and Carroll (2009) took an iterative approach that assumed linearity of the quantile functions and provided an approach based on estimating equations to fit quantile regression models with a scalar-valued covariate affected by measurement error. More recently, Wang and others (2012) developed a method under the assumption of a parametric additive measurement error model based on either the normal distribution or the heavier-tailed Laplace distribution for the measurement errors. Otherwise, they followed the traditional semiparametric approach of Koenker and Bassett (1978) and did not introduce distributional assumptions on the regression errors. The resulting estimators were shown to be efficient and robust against misspecification of the measurement error distribution. Finally, Chen and Müller (2012) considered the estimation of conditional quantile regression with function-valued covariates in a generalized functional regression framework. In their strategy, the observed functional data were regarded as irregularly observed and noisy measures of a true functional latent covariate, while the noise was handled using principal analysis by conditional expectation.

Our work provides novel contributions to quantile regression and functional data analysis in several ways. First, our approach can be used in studies evaluating relationships between function-valued covariates measured with complex error structures and conditional quantile functions of a continuous outcome. Second, we treat the functional covariate as a surrogate measure for a true function-valued covariate. An instrumental variable belonging in the same functional space is used to estimate the covariance matrix for the measurement error. Thus, we estimate the variance of the measurement error through the covariance between the instrumental variable and the unbiased surrogate measure for the unobserved true function-valued covariate. We applied our proposed methods to the National Health and Nutrition Examination Survey (NHANES) 2005–2006 database to assess impacts of daily activity counts on BMI.

## 2. FUNCTIONAL QUANTILE REGRESSION WITH MEASUREMENT ERROR

### 2.1. The model

Let  $\{Y, X, \mathbf{Z}\}$  be a triplet consisting of, respectively, a continuous scalar-valued random variable  $Y$ , a random function  $X = \{X(t), t \in [0, 1]\}$  on the unit interval assumed to be square integrable, and a  $p \times 1$  vector of error-free covariates  $\mathbf{Z} \in \mathbb{R}^p$ . For a given  $\tau \in (0, 1)$ , the  $\tau$ th conditional quantile function is defined as  $Q_{Y|X, \mathbf{Z}}(\tau) \equiv F_{Y|X, \mathbf{Z}}^{-1}(\tau)$  where  $F(Y|X, \mathbf{Z}) = P(Y \leq y|X, \mathbf{Z})$  is the cumulative distribution function of  $Y$  conditional on  $X, \mathbf{Z}$ . Our proposed functional quantile regression model with measurement error (CFQR-ME) is composed of two models: a response model for  $Y$  and a measurement error model.

Consider a sample of size  $n$ . The CFQR-ME for the  $i$ th subject,  $i = 1, \dots, n$ , is given by

$$Q_{Y_i|X_i, \mathbf{Z}_i}(\tau) = \int_0^1 \beta(\tau, t) X_i(t) dt + \mathbf{Z}_i \boldsymbol{\alpha}(\tau), \quad (2.1)$$

$$W_i(t) = X_i(t) + U_i(t), \quad (2.2)$$

where  $\beta(\tau, t)$  is an unknown functional coefficient for the  $\tau$ th conditional quantile of the  $i$ th response  $Y_i$ , while  $\boldsymbol{\alpha}(\tau)$  is a  $p \times 1$  vector of coefficients associated with the  $p$  error-free covariates,  $\mathbf{Z}_i$ . Further,  $X_i(t)$  is a function-valued covariate that is not directly observable but approximated by  $W_i(t)$ . The latter serves as a surrogate for  $X_i(t)$  and is subjected to a measurement error,  $U_i(t)$ ,  $i = 1, \dots, n$ .

We approximate  $\beta(\tau, t)$  in (2.1) using a polynomial spline, that is  $\beta(\tau, t) \approx \sum_{k=1}^{K_n} \gamma_k(\tau) b_k(t)$ , where  $\{\gamma_k(\tau)\}_{k=1}^{K_n}$  are unknown spline coefficients for the  $\tau$ th quantile function, while  $\{b_k(t)\}_{k=1}^{K_n}$  is a set of spline basis functions on the unit interval. The number of knots,  $K_n$ , is chosen to be large enough to capture the shape of  $\beta(\tau, t)$ , and  $K_n$  is allowed to depend on  $\tau$ . The models in (2.1)–(2.2) become

$$Q_{Y|X, \mathbf{Z}}(\tau) \approx \sum_{k=1}^{K_n} \gamma_k(\tau) X_{ik} + \mathbf{Z}_i \boldsymbol{\alpha}(\tau), \quad (2.3)$$

$$W_{ik} = X_{ik} + U_{ik} \quad k = 1, \dots, K_n, \quad (2.4)$$

after defining  $X_{ik} = \int_0^1 X_i(t) b_k(t) dt$ ,  $W_{ik} = \int_0^1 W_i(t) b_k(t) dt$ , and  $U_{ik} = \int_0^1 U_i(t) b_k(t) dt$ . Following the spline basis expansion, the model reduces to a linear quantile regression model with measurement errors. However,  $K_n$  increases with increasing sample sizes.

## 2.2. Instrumental variables

The presence of measurement error renders the model in (2.3)–(2.4) nonidentifiable without additional information (Carroll and others, 2006; Fuller, 2009; Tekwe and others, 2014, 2016). Here, we consider the case in which an instrumental variable (Bollen, 2012) for  $\{X(t)\}$  is available in the data. Alternatively, such information may come from repeated measurements of  $\{W(t)\}$  or a known variance–covariance matrix of  $\{W(t)\}$  given  $\{X(t)\}$ . Because  $\{X(t)\}$  is measured with error, simply replacing it with its surrogate measure,  $\{W(t)\}$ , would bias the association between  $\{X(t)\}$  and the quantiles of  $Y$  (Wang and others, 2012; Wei and Carroll, 2009). To be considered an instrumental variable, the covariate is assumed to be independent of the measurement error. This is often referred to as instrument exogeneity. This condition is difficult to test directly since the measurement error is unknown (Bollen, 2012). Instrumental variables have been used in a variety of measurement error problems including radiation epidemiology where the true radiation dose received at the time of exposure is unknown and not directly observable. In this setting, chromosome aberrations and acute symptoms of radiation exposure were used as instrumental variables for true radiation dose (Tekwe and others, 2014, 2016). They have also been used in public health survey data where family income is assumed to be measured with error, and family expenditures on culture and entertainment, education level, and health care were all used as instrumental variables for family income (Guan and others, 2019).

We propose using a function-valued instrumental variable to obtain a reasonable estimate of  $\Sigma_{uu}$ , the covariance matrix of the measurement error,  $\{U(t)\}$ . We now present how the instrumental variable allows us to estimate the covariance matrix of the measurement error.

Let  $M_i(t)$ ,  $i = 1, \dots, n$ , be the function-valued instrumental variable observed for the  $i$ th subject. Assume the  $\{M_i(t)\}$ 's are reciprocally independent and that the instrumental variable is uncorrelated with

the measurement error,  $\text{Cov}\{M_i(s), U_i(t)\} = 0$ , for any  $s, t \in [0, 1]$ . In addition to (2.1)–(2.2), we introduce the model

$$M_i(t) = \delta(t)X_i(t) + \eta_i(t), \quad (2.5)$$

where  $\delta(t)$  is an unknown coefficient function that quantifies the relationship between  $M_i(t)$  and  $X_i(t)$ , and  $\eta_i(t)$  is a zero mean error-term independent of  $U_i(t)$ . The instrumental variable is not required to be an unbiased measure for  $X_i(t)$ ; however, it is correlated with  $X_i(t)$ . If  $\delta(t)$  in (2.5) is known, then the equation becomes

$$M_i^*(t) = X_i(t) + \eta_i^*(t), \quad (2.6)$$

where  $M_i^*(t) = M_i(t)/\delta(t)$ , and  $\eta_i^*(t) = \eta_i(t)/\delta(t)$  is the new error term. Let  $M_{ik}^* = \int_0^1 M_i^*(t)b_k(t)dt$  and  $\eta_{ik}^* = \int_0^1 \eta_i^*(t)b_k(t)dt$ , for  $k = 1, \dots, K_n$ . Following the spline basis expansion, (2.6) becomes

$$M_{ik}^* = X_{ik} + \eta_{ik}^*. \quad (2.7)$$

Define  $\mathbf{M}_i^* = (M_{i1}^*, \dots, M_{iK_n}^*)^T$ ,  $\mathbf{U}_i = (U_{i1}, \dots, U_{iK_n})^T$ , and  $\mathbf{W}_i = (W_{i1}, \dots, W_{iK_n})^T$ . Let  $\Sigma_{UU}$  and  $\Sigma_{WW}$  be the covariance matrices of  $\mathbf{U}_i$  and  $\mathbf{W}_i$ , respectively, and  $\Sigma_{M^*W}$  be the covariance matrix between  $\mathbf{M}_i^*$  and  $\mathbf{W}_i$ . Then, a reasonable estimate of  $\Sigma_{UU}$  can be obtained using

$$\widehat{\Sigma}_{UU} = \widehat{\Sigma}_{WW} - \widehat{\Sigma}_{M^*W}, \quad (2.8)$$

where  $\widehat{\Sigma}_{WW}$  and  $\widehat{\Sigma}_{M^*W}$  are both sample covariance matrices of dimension  $K_n \times K_n$ .

The estimator in (2.8) requires that the coefficient function  $\delta(t)$  be known. Because  $\delta(t)$  is often unknown and unavailable in practice, it can be estimated from the data. This is accomplished by noting that equations (2.2) and (2.5) imply that for any  $t \in [0, 1]$ ,

$$E\{M_i(t)\} = \delta(t)E\{X_i(t)\} = \delta(t)E\{W_i(t)\}.$$

Therefore, we propose a simple data-driven ratio estimator for  $\delta(t)$  by

$$\widehat{\delta}(t) = \frac{\frac{1}{n} \sum_{i=1}^n M_i(t)}{\frac{1}{n} \sum_{i=1}^n W_i(t)}.$$

We note that  $\widehat{\delta}$  can be unstable when the denominator is close to 0. However, an improved estimator can be obtained by smoothing the ratio of the estimates as functions of  $t$ .

### 2.3. Assumptions

Our goal is to estimate the unknown parameters in (2.3). For that, we introduce the following additional assumptions:

- [A1]  $\text{Cov}(\mathbf{X}_i, \mathbf{U}_i) = \mathbf{0}$ .
- [A2]  $E(\mathbf{W}_i | \mathbf{X}_i) = \mathbf{X}_i$  and  $\text{Cov}(\mathbf{W}_i | \mathbf{X}_i) = \Sigma_{UU}$ .
- [A3]  $\text{Cov}(\mathbf{W}_i, \mathbf{U}_i) \neq \mathbf{0}$ , and  $\text{Cov}(\mathbf{W}_i, \mathbf{X}_i) \neq \mathbf{0}$ .
- [A4]  $\text{Cov}(U_{ik}, U_{ij}) \neq 0$  for  $k \neq j$ .
- [A5]  $\mathbf{U}_i \sim \text{MVN}(\mathbf{0}, \Sigma_{UU})$ .

Assumptions A1 and A2 represent the usual classical measurement error assumptions. In Assumption A3, we provide the correlated measurement errors associated with the function-valued covariate for each subject  $i$ . Assumption A5 introduces normality on  $\mathbf{U}_i$ , but it can be relaxed.

### 3. SIMULATION EXTRAPOLATION

The SIMEX approach for measurement error correction is a simulation-based method for additive measurement error problems where the measurement error processes are simulated using Monte Carlo (Carroll and others, 2006). In that setting, the covariance of the measurement error,  $\Sigma_{UU}$ , is assumed to be known or is estimable from the observed data (Cook and Stefanski, 1994). SIMEX has been previously used in quantile regression with scalar-valued covariates measured with error. For example, Mao and Wei (2017) used SIMEX in censored quantile regression, and Shang (2012) applied SIMEX to model student growth percentiles.

The steps in SIMEX are as follows: (i) simulate additional measurement errors with increasing variances and add them to the observed data,  $\mathbf{W}_i$ ; (ii) obtain regression coefficient estimates from the model of interest using the simulated data prone to errors; (iii) repeat the simulation and estimation steps a large number of times; (iv) obtain the averaged values of the estimates to be used to determine the relationship between the measurement-error-induced bias and the variance of the errors; and (v) extrapolate the relationship back to the case of no measurement error, (see Carroll and others, 2006; Cook and Stefanski, 1994; Stefanski and Cook, 1995; Carroll and Stefanski, 1997; Mao and Wei, 2017, for details).

#### 3.1. Simulation step

We now provide the details of the application of the SIMEX to the case of quantile regression with a function-valued covariate measured with error. Our goal is to estimate  $\beta(\tau, t)$  and  $\alpha(\tau)$  at quantile level  $\tau$ , after measurement error correction. Following the basis expansion described in (2.3) and (2.4), the parameter to be estimated is  $\theta(\tau) = \{\gamma(\tau)^T, \alpha(\tau)^T\}^T \in K_{n+p}$ , where  $\gamma(\tau) = \{\gamma_1(\tau), \dots, \gamma_{K_n}(\tau)\}^T$ . By plugging  $\mathbf{W}_i$  into (2.3) for  $\mathbf{X}_i$ , we obtain  $\tilde{\theta}(\tau)$ , the naive estimator of the regression coefficients. Now, in addition to  $\mathbf{W}_i$ , we simulate a sequence of remeasurements of  $\mathbf{W}_i$  with increasing variability. Let  $U_i(\lambda) \sim \text{MVN}(0, \lambda \Sigma_{uu})$ . The additional data are simulated as

$$\mathbf{W}_i(\lambda_j) = \mathbf{W}_i + U_i(\lambda_j), j = 1, \dots, L$$

for a sequence  $0 < \lambda_1 < \dots < \lambda_L$ . Thus, data sets with increasingly larger measurement error variances are simulated. For the  $j$ th data set, the measurement error variance is  $(1 + \lambda_j) \Sigma_{uu}$ .

The regression coefficients are then estimated by

$$\tilde{\theta}_j(\tau) = \underset{\theta(\tau)}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau \{Y_i - \mathbf{W}_i(\lambda_j) \gamma(\tau) - \mathbf{Z}_i \alpha(\tau)\}, \quad (3.9)$$

where  $\rho_\tau(a) = a\{\tau - I(a < 0)\}$  is the quantile regression check function (Koenker and Hallock, 2001),  $\theta(\tau) = \{\gamma(\tau), \alpha(\tau)\}$ , and  $I(\cdot)$  denotes the indicator function. The optimization problem in (3.9) can be solved using linear programming algorithms (e.g., as implemented in the `quantreg` package in R or `PROC QUANTREG` in SAS). The simulation and estimation steps are repeated for a large number of times, and the average value of the parameter is calculated at each level of contamination. Let  $\hat{\theta}_j(\tau)$  denote such an estimate at  $\lambda_j$ , for  $j = 1, \dots, L$ .

### 3.2. Extrapolation step

In the extrapolation step, we use a regression model with data  $\{\lambda_j, \hat{\theta}_j(\tau)\}_{j=0}^L$  to understand the relationship between the averaged error-contaminated estimates and the magnitude of the measurement errors. Here,  $\lambda_0 = 0$  and  $\hat{\theta}_0(\tau)$  reduces to the naive estimator obtained using the original observed data,  $\mathbf{W}_i$ . We consider a quadratic extrapolation model, for  $j = 0, \dots, L$ ,

$$\hat{\theta}(\tau, \lambda_j) + \varepsilon_j = \phi_0(\tau) + \phi_1(\tau)\lambda_j + \phi_2(\tau)\lambda_j^2 + \varepsilon_j, \quad (3.10)$$

where  $\phi_0(\tau)$ ,  $\phi_1(\tau)$  and  $\phi_2(\tau)$  are unknown coefficients, each of length  $K_n + p$ , which can be estimated by ordinary least squares. As a result, let  $\hat{\theta}(\tau, \lambda)$  be the estimated extrapolation function. While (3.10) provides the regression equation for quadratic extrapolation, other methods such as linear or nonlinear extrapolations can be used. The linear extrapolation function is a linear function of  $\lambda_j$  in  $\hat{\theta}(\tau, \lambda_j)$ , while the nonlinear extrapolation function is the inverse function of  $\lambda_j$  in  $\hat{\theta}(\tau, \lambda_j)$ . Results from our numerical studies and application illustrated that the quadratic extrapolation had the best performance in terms of bias reduction and computational stability.

### 3.3. The SIMEX estimator

Through extrapolation, we obtain  $\hat{\theta}(\tau)$  as the coefficient estimates in the case of no measurement error. In particular, the SIMEX estimator is obtained as

$$\hat{\theta}_{\text{SIMEX}}(\tau) = \lim_{\lambda \rightarrow -1} \hat{\theta}(\tau, \lambda).$$

We summarize our estimation process as follows:

1. Obtain the basis expansions of  $\{M(t)\}$  and  $\{W(t)\}$ .
2. Obtain  $\hat{\Sigma}_{UU}$  as in (2.8).
3. Choose a sequence of monotonically increasing positive small numbers for  $\lambda$ , such as  $\lambda = 0, 1/4, 1/2, \dots, 2$ .
4. Generate  $\mathbf{U} \sim \text{MVN}(\mathbf{0}, \hat{\Sigma}_{UU})$  and remeasurements for  $\mathbf{W}$  using  $\mathbf{W}(\lambda) = \mathbf{W} + \sqrt{\lambda}\mathbf{U}$  for each  $\lambda$ .
5. Perform the  $\tau$ th quantile regression of  $Y$  on  $\mathbf{W}(\lambda)$  to obtain  $\hat{\theta}(\tau, \lambda)$  for each  $\lambda$ .
6. Repeat the simulation and estimation steps for a total of  $S$  times. Let  $\hat{\theta}_s(\tau, \lambda)$  be the coefficient estimate in  $s$ th replication for  $s = 1, \dots, S$ . For each  $\lambda$ , obtain the averaged estimate  $\hat{\theta}(\tau, \lambda)$  across all  $S$  replications by

$$\hat{\theta}(\tau, \lambda) = \frac{1}{S} \sum_{s=1}^S \hat{\theta}_s(\tau, \lambda).$$

7. In the extrapolation step, an extrapolation function is formed for the relationship between the estimated  $\hat{\theta}(\tau, \lambda)$  and  $\lambda$ .
8. To obtain the extrapolation function,  $\hat{\theta}(\tau, \lambda)$  are regressed on  $\lambda$ .
9. Because  $\lambda = -1$  represents the condition of no measurement error, the predicted values for the SIMEX estimator are obtained by plugging  $\lambda = -1$  into the prediction equation obtained from the extrapolation function. The naive estimator is obtained from  $\lambda = 0$ . That is,  $\hat{\theta}_{\text{SIMEX}}(\tau) = \hat{\theta}(\tau, -1)$ , and  $\hat{\theta}_{\text{Naive}}(\tau) = \hat{\theta}(\tau, 0)$ .
10. The SIMEX and naive estimators of the functional coefficient are obtained as:

$$\hat{\beta}_{\text{SIMEX}}(\tau, t) \approx \sum_{k=1}^{K_n} \hat{\gamma}_k(\tau, \lambda = -1) b_k(t); \quad (3.11)$$



and

$$\hat{\beta}_{\text{Naive}}(\tau, t) \approx \sum_{k=1}^{K_n} \hat{\gamma}_k(\tau, \lambda = 0) b_k(t). \quad (3.12)$$

11. The measurement error corrected coefficient on the  $p$ th error-free covariate at the  $\tau$ th quantile level is estimated using  $\hat{\alpha}_{p,\text{SIMEX}}(\tau) = \hat{\theta}(\tau, \alpha_p, \lambda = -1)$ , while the uncorrected coefficient error-free coefficient is estimated as  $\hat{\alpha}_{p,\text{Naive}}(\tau) = \hat{\theta}(\tau, \alpha_p, \lambda = 0)$ .

We note that by applying the SIMEX approach, the impact of measurement error is automatically corrected for in all the model parameters, including the coefficients on the error-free covariates,  $\alpha(\tau)$ . For inference, nonparametric bootstrap confidence intervals (CIs) can be employed.

### 3.4. The number of basis functions

The proposed SIMEX approach depends on the choice of the number of basis functions for approximating the unknown functional coefficient. We propose to select the number of basis functions by Bayesian Information Criterion (BIC), which was also used is also used (Kato, 2012). For the naive method, the BIC is defined as

$$\text{BIC}(K_n) = \log \left\{ \frac{1}{n} \sum_{i=1}^n \rho_{\tau} \left[ Y_i - Z_i^T \hat{\alpha}(\tau) - \sum_{j=1}^{K_n} W_{ij} \hat{\gamma}_j(\tau) \right] \right\} + \frac{(K_n + p) \log(n)}{n}.$$

The optimal  $K_n$  is the one that minimizes the BIC value. We propose to first select  $K_n$  under the naive method with BIC, and to use the selected  $K_n$  with the SIMEX-based estimator.

## 4. APPLICATION TO BMI AND PHYSICAL ACTIVITY

Obesity is largely the result of chronic positive energy balance, which is primarily influenced by high dietary intake, low physical activity, or both (Spiegelman and Flier, 2001). Modification of lifestyle behaviors to improve energy balance may reduce obesity risk. We applied our methodology to investigate the relationship between wearable device-based measures of physical activity and BMI among adults in the United States. A total of 2106 adults 20–85 years of age who participated in the 2005–2006 cycle of the NHANES were included in our analysis. The average age of the participants was 43.3 (SD = 13.8), while their average BMI was 29.0 (SD = 6.9). There were 316 (15.01%) blacks, 1,413 (67.09%) whites, 282 (13.75%) Hispanics, and 95 (4.36%) other races; and 50.6% of the participants were female. The participants wore ActiGraph uniaxial accelerometers (model #AM7164) on their hips for at least 5 days during waking hours, except during water-related activities and sleep (Matthews and others, 2008; Troiano and others, 2008; Evenson and others, 2015). The devices measure vertical accelerations, recorded as “activity counts” that are attributed to body movements and were averaged in 60-s epochs. In general, the greater activity counts correspond to higher physical activity intensity. Each participant’s height and weight were obtained by NHANES examiners. We considered device-based measures of activity counts to be surrogate measures for the true function-valued covariate, that is, physical activity intensity. The instrumental variable,  $\{M(t)\}$ , used in our application was step counts obtained from the same device. In this application, we assumed that conditional on the true physical activity intensity, the activity counts, and step counts were independent. Thus, step counts were assumed to be correlated with physical activity intensity. However, it does not need to be an unbiased measure of physical activity intensity. On the other



hand, it is possible for Assumption A5 to be violated in our current application based on the NHANES data. Other distributions that may be considered for  $U_i$  instead of the normal one are the Poisson and asymmetric Laplace distributions since  $U_i$  represents the model error associated with activity counts, which are integer values and right skewed. Race/ethnicity (black, Hispanic, white, or Other), age, and sex were included in the analyses as error-free covariates. We aggregated the physical activity data into hourly averages to smooth out the noise that is typical of high-frequency measures. Prior to implementing the SIMEX algorithm, we investigated the optimal number of knots using BIC. We selected six knots for the 25th, 50th, and 95th percentiles and five knots for the 75th percentile. We applied sample weights to account for the oversampling of racial subgroups as per NHANES guidelines (Johnson and others, 2013).

Quadratic extrapolation with a grid of  $\lambda$  values ranging from 0 to 2 in steps of 0.05 (0.0, 0.05, 0.10, ..., 2.0) was used for the extrapolation step of the SIMEX algorithm as follows:

$$E\{\hat{\theta}(\tau, \lambda)\} = \phi_0(\tau) + \phi_1(\tau)\lambda + \phi_2(\tau)\lambda^2 \quad \lambda = 0.0, \dots, 2.0.$$

For the nonparametric bootstrap, we first resampled the original data without replacement. We then estimated the regression coefficients using the resampled data. Next, we repeated these two steps 100 times, resulting in 100 bootstrap samples with  $\hat{\beta}_b(t, \tau)$  and  $\hat{\alpha}_b(\tau)$ ,  $b = 1, \dots, 100$ . We computed the 95% pointwise bootstrap CIs as the 0.025 and 0.975 percentiles of  $\hat{\beta}_b(t, \tau)$  at each observed time point ( $t = 1, \dots, 24$ ). We also calculated the 95% bootstrap CI for each error-free covariate as the 0.025 and 0.975 percentiles of  $\hat{\alpha}_b(\tau)$ .

#### 4.1. Results

Table 1 shows the results of the application of our proposed model to the motivating example. One of the advantages of the SIMEX approach to correcting for measurement error is that the method automatically provides both measurement error adjusted and naive estimates of the parameters. From the table, we see that the presence of measurement error also impacts the error-free covariates. Our results show that the impact of measurement error on the error-free covariates depends on the covariate. Under both the naive and SIMEX-corrected methods, age, sex, and race were statistically significant after adjustment for the other covariates based on the 95% bootstrap CIs at the 25th percentile level of BMI. At the 50th percentile of BMI, we observed differences between the quantile levels of BMI for the black and other participants when compared to whites under the SIMEX approach. Under the SIMEX approach, there was no difference in the 50th percentile of BMI between Hispanics and whites. However, a statistically significant difference was observed under the naive method. There was statistically significant at the 75th percentile but not statistically significant at the 95th percentile level function of BMI. There were differences in the 75th and 95th quantile function of BMI between females and males, and also between black and white study participants. In general, the 95% CIs for the SIMEX corrected methods for measurement errors tended to be wider than the CIs based on the noncorrected methods.

In Figure 1, we provide the histogram of BMI. The histogram indicates that the distribution of BMI in our application was skewed. In Figure 2a and b, we provide spaghetti plots of both activity and step counts obtained from the wearable devices. Those plots indicated that both activity and step counts obtained from the devices were nonlinear functions of time. Quantile-specific plots of the SIMEX and naive estimators were provided in Figure 3. The results show that the relationship between activity counts and BMI depended on the time of activity. Based on the plots, we observed significant associations between activity counts and the 25th, 50th, and 75th quantile functions of BMI between the 20th and 24th hours, corresponding to after work hours based on the 95% pointwise CIs. This finding can be potentially due to anorexigenic effects of physical activity (Neumeier and others, 2016), where physical activity potentially suppresses appetite following mental work. Similarly, in the earlier hours, between the 10th and 15th

Table 1. *Estimated effects of the SIMEX- and naive-based estimators of  $\alpha(\tau)$ , the error-free coefficients, on quantile functions of BMI based on the application of our proposed methods to the NHANES database*

Covariates		SIMEX	95% CI	Naive	95% CI
25th quantile					
Age		0.08	0.06 to 0.10	0.08	0.06 to 0.10
Sex					
	Female	-1.10	-1.65 to -0.60	-1.10	-1.63 to -0.61
Race					
	Black	1.18	0.39 to 1.94	1.14	0.42 to 1.87
	Hispanic	1.26	(0.47 to 1.91	1.21	0.51 to 1.88
	Other	-1.39	-2.35 to -0.02	-1.40	-2.40 to -0.20
50th quantile					
Age		0.09	0.05 to 0.12	0.09	0.06 to 0.11
Sex					
	Female	-0.18	-0.94 to 0.42	-0.23	-0.99 to 0.41
Race					
	Black	1.56	0.46 to 2.46	1.66	0.61 to 2.48
	Hispanic	1.10	-0.03 to 2.14	0.97	0.04 to 1.98
	Other	-2.74	-3.48 to -0.80	-2.61	-3.45 to -0.89
75th quantile					
Age		0.09	0.05 to 0.12	0.08	0.05 to 0.11
Sex					
	Female	1.95	1.01 to 2.86	2.05	1.08 to 2.84
Race					
	Black	1.46	0.01 to 3.16	1.52	0.25 to 3.00
	Hispanic	-0.28	-1.34 to 0.85	-0.33	-1.26 to 0.84
	Other	-2.04	-4.70 to -0.48	-2.12	-4.57 to -0.62
95th quantile					
Age		0.03	-0.08 to 0.09	0.03	-0.08 to 0.09
Sex					
	Female	3.50	1.32 to 5.78	3.41	1.40 to 5.53
Race					
	Black	5.60	0.25 to 7.86	5.31	0.67 to 7.62
	Hispanic	0.45	-4.01 to 2.95	-0.16	-3.70 to 2.39
	Other	-1.64	-6.02 to 5.04	-0.95	-5.18 to 4.12

hours, we observed significant associations between activity counts and physical activity and the 25th, 50th, and 75 quantile functions of BMI since zero was completely included in the CIs. However, we did not observe any statistically significant associations between activity counts and the 95th quantile level of BMI after adjusting for the error-free covariates. This finding may be due to the accuracy of the device changing at higher BMI levels since higher values of BMI have been associated with irregular gait patterns which might not be detectable by the wearable device (Carter and others, 2022; Carter and others, 2019). Additionally, we also observed wider pointwise CIs and the 95th quantile when compared to the lower quantile levels. Finally, we did not observe differences between the error-corrected and naive-based estimators at all quantile levels of BMI as illustrated in Figure 3. We also obtained plots of SIMEX and

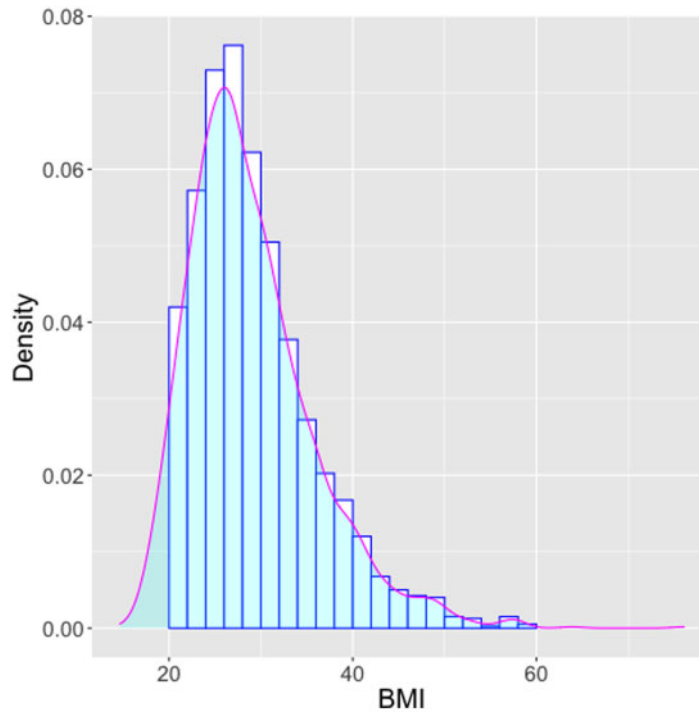


Fig. 1. Histogram of the BMI of the NHANES participants included in our study.

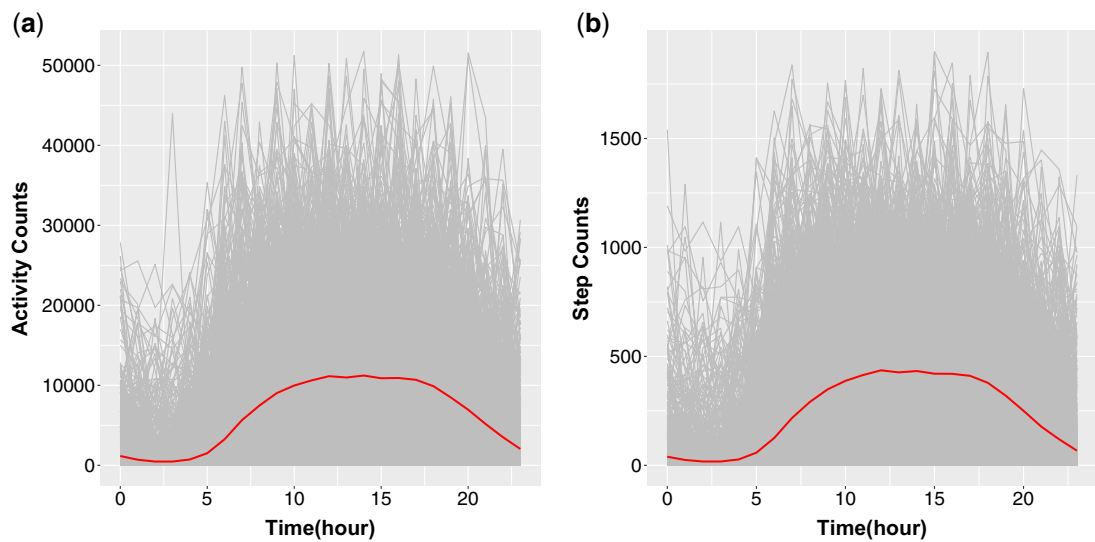


Fig. 2. Wearable device-based (a) activity counts and (b) step counts from the NHANES data base plotted against time.

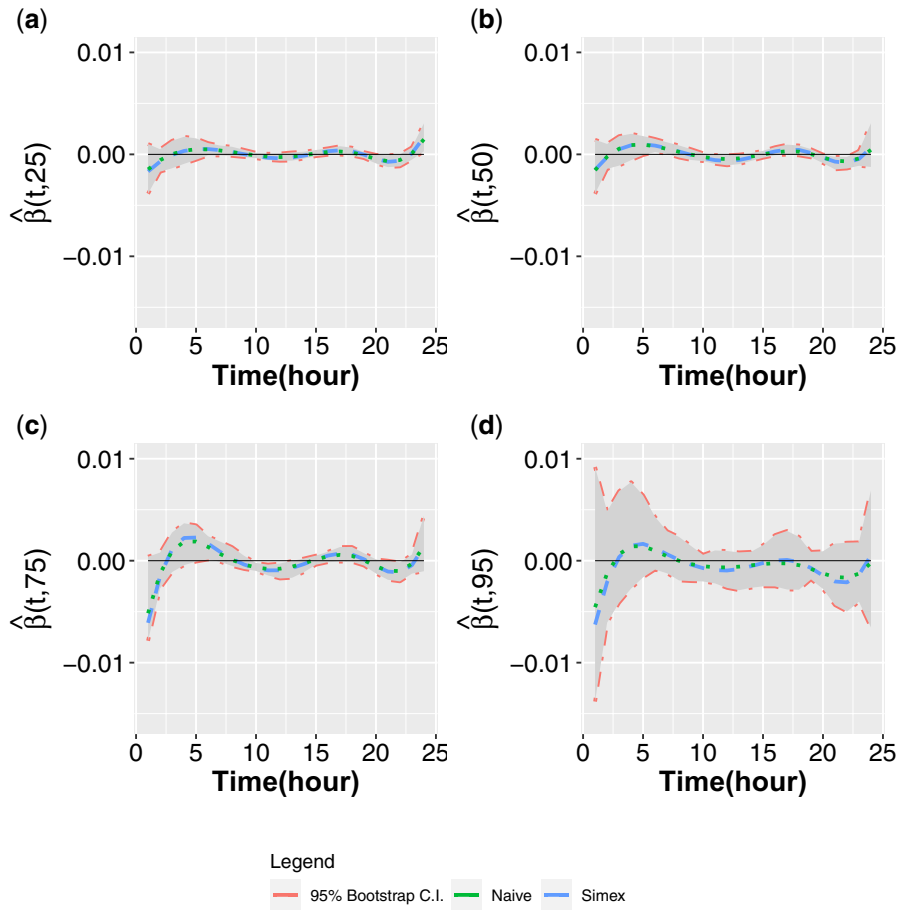


Fig. 3. Estimated quantile-specific plots of the SIMEX- and naive-based estimates of  $\beta(t, \tau)$  from our application to the NHANES database. We estimated the effects of the wearable device-based measures of activity counts on the 25th, 50th, 75th, and 95th quantile functions of BMI after adjusting for age, sex, and race. The shaded areas are the 95% point-wise bootstrap CIs. The dashed line represents the average SIMEX estimates of  $\beta(t, \tau)$  from the bootstrap. The dotted line represents the average naive estimates of  $\beta(t, \tau)$  of bootstrap.

naive-based estimators of the race by sex effects at all the quantile levels for subjects with average age and activity counts (Figure 4). We observed that under both the SIMEX- and naive-based methods, the BMI values for the Other race tended to be lowest, while the BMI values for blacks tended to be highest. The differences in the BMI values between the racial groups considered increased with increasing quantiles for both females and males.

## 5. SIMULATIONS

We conducted simulations to study the performance of our proposed method. Sensitivity analysis was also performed to assess the sensitivity of the proposed methods to the magnitude of measurement error and to the violation of the instrument exogeneity assumption.

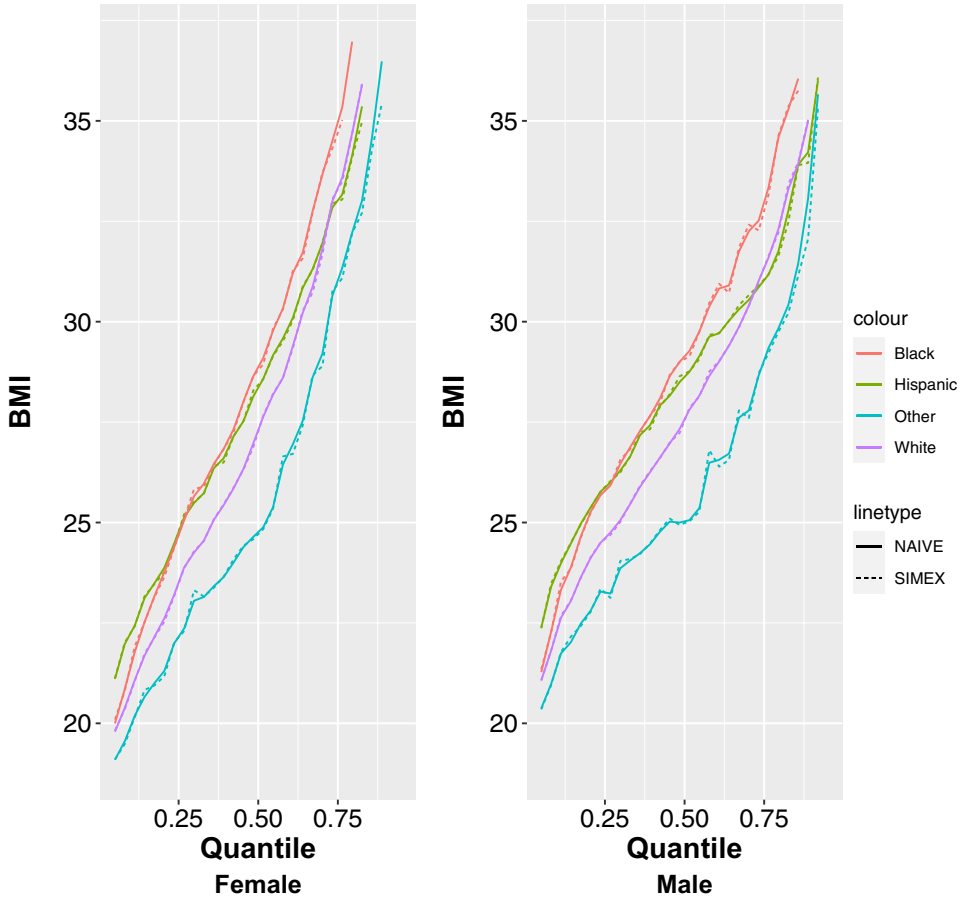


Fig. 4. Plots of predicted BMI at the varying quantile levels by demographics. Each plot compares the distribution of BMI estimated using the naive and SIMEX methods for different racial/ethnic groups by sex from the NHANES database. The left figure corresponds to the female study participants with average age and average levels of activity counts by race. The right figure corresponds to male participants with average age and average activity counts by race.

### 5.1. Simulation study 1

We first generated 500 data sets with sample sizes  $n = (200, 500, 1000, 2000)$  independently from model (2.1) with  $X(t) = 1/\{1 + \exp[8(t - 0.5)]\} + \varepsilon_x(t)$ ,  $W(t) = X(t) + U(t)$ , and  $M(t) = \delta(t)X(t) + \eta(t)$  for  $t \in [0, 1]$ . The error terms  $\varepsilon_x(t)$ ,  $U(t)$ , and  $\eta(t)$  were generated independently from Gaussian processes with zero mean and covariances with compound symmetry structures with correlations of  $\rho = 0.5$ , and constant variances on the diagonals with  $\sigma_x^2 = \text{Var}\{\varepsilon_x(t)\}$ ,  $\sigma_U^2 = \text{Var}\{U(t)\}$ , and  $\sigma_\eta^2 = \text{Var}\{\eta(t)\}$ . To obtain different magnitudes of measurement error, we considered  $\sigma_x = (0.75, 1.0, 1.5)$ ,  $\sigma_U = (0.75, 1.0)$ , and  $\sigma_\eta = 0.25$ . The outcome,  $Y$ , was simulated from a Gaussian distribution with mean  $\mu_Y = \int_0^1 \sin(2\pi t)X(t)dt$  and standard deviation  $\sigma_Y = 0.1$ . Error-free covariates were not considered and we assumed  $\delta(t) = 1$  in the first set of simulations.

Let  $\hat{\beta}^r(t, \tau)$  be the estimator of  $\beta(t, \tau)$  in the  $r$ th replication and  $\bar{\beta}(t, \tau) = \frac{1}{500} \sum_{l=1}^{500} \hat{\beta}^l(t, \tau)$ , the averaged estimate over a total of 500 replications. Let  $\{t_l\}_l^{n_{\text{grid}}}$  be a sequence of equally spaced grid points on  $[0, 1]$ .

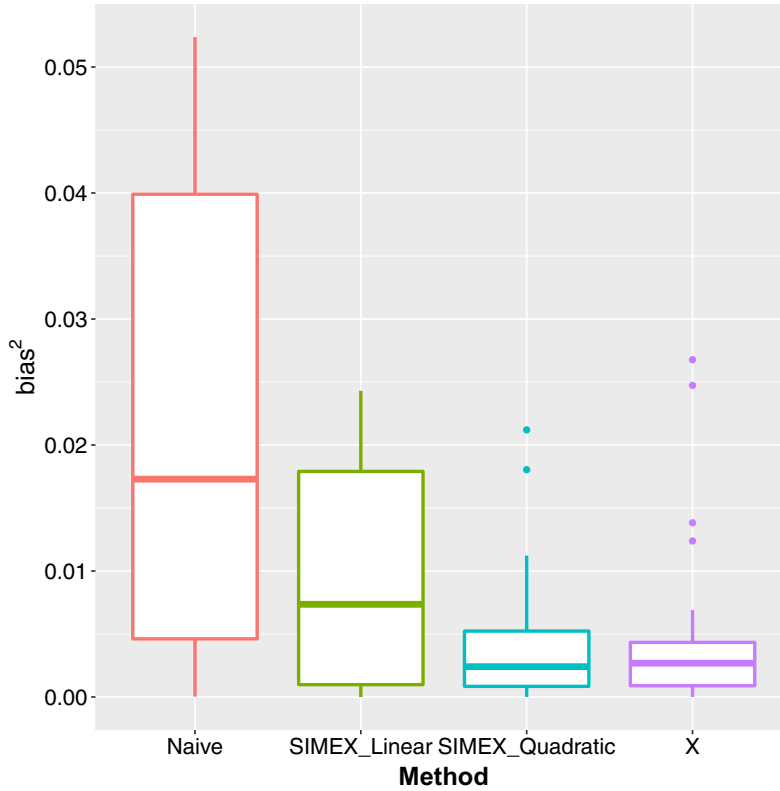


Fig. 5. Boxplots of squared biases from the naive, SIMEX with linear extrapolation, SIMEX with quadratic extrapolation, and oracle (X) methods from Simulation Study 1.

We define the average squared bias of  $\hat{\beta}(t, \tau)$  as  $ABias^2\{\hat{\beta}(t, \tau)\} = \frac{1}{n_{\text{grid}}} \sum_{l=1}^{n_{\text{grid}}} \{\bar{\beta}(t_l, \tau) - \beta_{\text{true}}(t_l, \tau)\}^2$ , the average sample variance as

$$Avar\{\hat{\beta}(t, \tau)\} = \frac{1}{n_r} \sum_{r=1}^{n_r} \frac{1}{n_{\text{grid}}} \sum_{l=1}^{n_{\text{grid}}} \{\hat{\beta}^r(t_l, \tau) - \bar{\beta}(t_l, \tau)\}^2,$$

and the average integrated mean square error as  $AIMSE\{\hat{\beta}(t, \tau)\} = ABias^2\{\hat{\beta}(t, \tau)\} + Avar\{\hat{\beta}(t, \tau)\}$ . The calculations of  $ABias^2$ ,  $Avar$ , and  $AIMSE$  were conducted at quantiles levels  $\tau \in (0.25, 0.50, 0.75, 0.95)$ .

In the first set of simulations, we examined the performance of the estimators under the oracle  $\{X(t)\}$ , SIMEX, and naive  $\{W(t)\}$  methods. The simulations were performed at the 50th quantile level with 500 replicates,  $n = 1000$ ,  $t = 200$ ,  $\sigma_x = 1.5$  and  $\sigma_w = 0.75$ . The boxplots of  $Abias^2$  are provided in Figure 5, which illustrates that the biases obtained under the naive method was the largest, followed by the biases obtained under the SIMEX estimators with the linear and quadratic extrapolation approaches. The oracle estimator had the lowest bias. We therefore recommend the quadratic extrapolation method when the SIMEX approach is employed for measurement error correction in this model setting.

In another simulation study, we assessed the impact of sample sizes on the estimation. Table 2 summarizes the estimation results of the three different methods at sample sizes  $n = (200, 500, 1000, 2000)$  when  $\sigma_x = 1.5$  and  $\sigma_U = 0.75$ . The results indicated that the bias of the proposed SIMEX method was

Table 2. The effects of sample sizes on the performance of  $\widehat{\beta}(t, \tau)$  obtained under the three estimators considered for Simulation Study 1 where error-free covariates were excluded. The averaged squared bias (ABias<sup>2</sup>), averaged sample variance (Avar), and averaged integrated mean squared error (AIMSE) of the oracle ( $X$ ), SIMEX, and naive estimators for sample sizes  $n = (200, 500, 1000, 2000)$ ;  $\tau = (0.25, 0.50, 0.75, 0.95)$ ; with  $\sigma_X = 1.5$ , and  $\sigma_U = 0.75$

$n$	Oracle			SIMEX			NAIVE		
	ABias <sup>2</sup>	AVar	AIMSE	ABias <sup>2</sup>	AVar	AIMSE	ABias <sup>2</sup>	AVar	AIMSE
25th quantile									
200	0.0033	0.0395	0.0428	0.0040	0.0872	0.0911	0.0238	0.0361	0.0599
500	0.0037	0.0149	0.0187	0.0037	0.0361	0.0398	0.0234	0.0131	0.0366
1000	0.0034	0.0082	0.0116	0.0037	0.0211	0.0248	0.0236	0.0072	0.0308
2000	0.0026	0.0050	0.0076	0.0030	0.0118	0.0148	0.0225	0.0043	0.0268
50th quantile									
200	0.0040	0.0305	0.0345	0.0045	0.0721	0.0766	0.0252	0.0269	0.0521
500	0.0039	0.0123	0.0162	0.0042	0.0316	0.0358	0.0236	0.0109	0.0345
1000	0.0034	0.0067	0.0102	0.0038	0.0182	0.0220	0.0228	0.0062	0.0289
2000	0.0029	0.0042	0.0070	0.0031	0.0113	0.0143	0.0223	0.0036	0.0260
75th quantile									
200	0.0041	0.0364	0.0405	0.0039	0.0854	0.0893	0.0232	0.0325	0.0556
500	0.0041	0.0147	0.0188	0.0043	0.0381	0.0424	0.0246	0.0136	0.0382
1000	0.0031	0.0084	0.0114	0.0029	0.0219	0.0248	0.0218	0.0076	0.0293
2000	0.0028	0.0048	0.0076	0.0030	0.0126	0.0156	0.0224	0.0045	0.0269
95th quantile									
200	0.0016	0.0995	0.1010	0.0011	0.2338	0.2349	0.0212	0.0938	0.1150
500	0.0018	0.0361	0.0379	0.0013	0.0867	0.0880	0.0214	0.0345	0.0559
1000	0.0018	0.0200	0.0219	0.0017	0.0451	0.0469	0.0211	0.0190	0.0401
2000	0.0012	0.0115	0.0127	0.0013	0.0248	0.0261	0.0212	0.0103	0.0315

close to that of the oracle method  $\{X(t)\}$ , and both biases slowly decreased with increasing sample sizes. Additionally, the bias of the SIMEX method was noticeably smaller than that of the naive method for all sample sizes and quantiles. Those results numerically confirmed the effectiveness of SIMEX in reducing estimation bias due to measurement error. The Avar decreased with increasing sample sizes under all three methods, while the Avar of the naive method was the smallest. However, because of the nonignorable bias, the SIMEX had better overall performance than the naive method with smaller AIMSE for relatively larger sample sizes ( $n = 1000, 2000$ ).

We also investigated the impacts of varying levels of  $\sigma_X$  and  $\sigma_U$  on the performance of the three estimators. For that, we performed simulations with  $n = (500, 1000)$ ,  $\sigma_X = (0.75, 1.0, 1.5)$ , and  $\sigma_U = (0.75, 1.0)$ , with  $t = 200$  grid points. Boxplots of the ABias<sup>2</sup> are provided in Figure 6. Our simulations indicated that as the value of  $\sigma_U$  increased or the value of  $\sigma_X$  decreased, the biases associated with the SIMEX and naive methods both increased due to higher magnitudes of measurement error relative to the error associated with  $\{X(t)\}$ . These trends were illustrated in Figure 7 for the 50th quantile function. Similar findings were observed across all the quantile functions considered. We also note that as  $\sigma_W$  increased, the differences among the three methods became more apparent. In summary, the SIMEX estimator was closer to the oracle  $\{X(t)\}$ , and the estimation biases associated with the naive estimator were much larger than those associated with both the SIMEX and oracle methods.



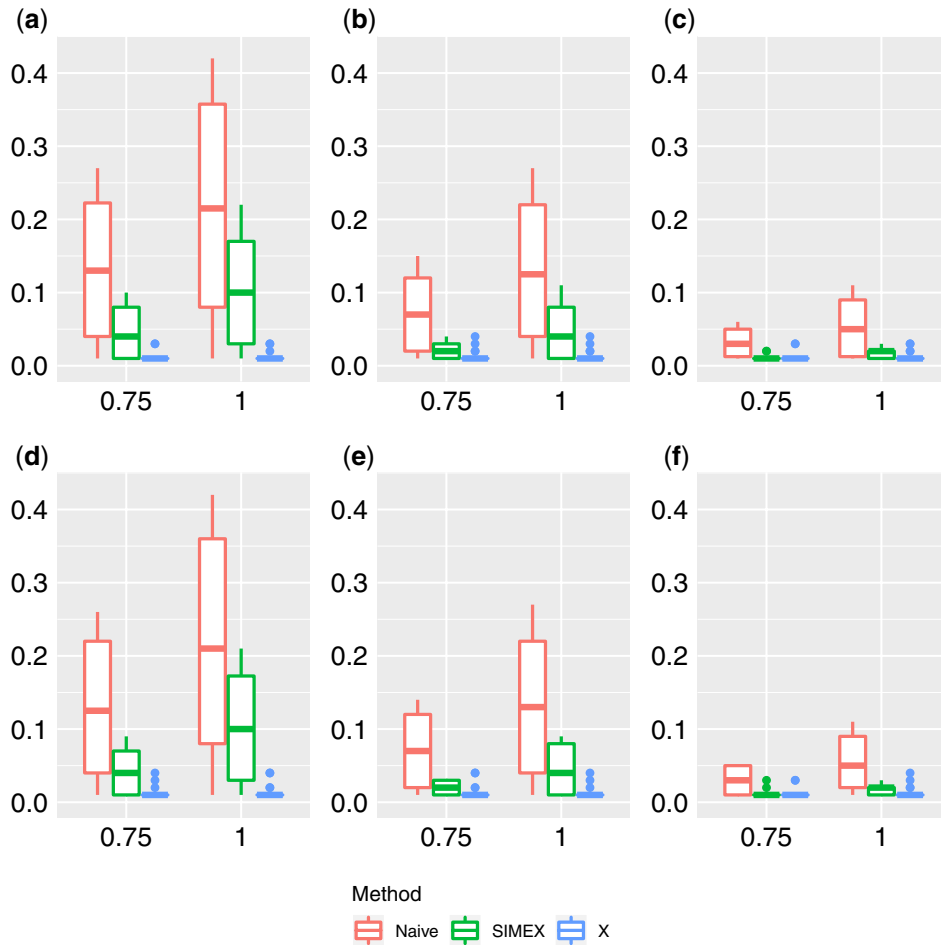


Fig. 6. Boxplots of  $\text{Abias}^2$  by the different estimators at the 50th percentile from Simulation Study 1. Each plot compares the boxplots of the averaged squared biases for the naive, SIMEX, and oracle(X) estimators of  $\beta(t, 50)$  at the 50th quantile level when  $\sigma_U = 0.75$  and 1. The top three figures correspond to  $n = 500$ , with (a)  $\sigma_X = 0.75$ , (b)  $\sigma_X = 1$ , and (c)  $\sigma_X = 1.5$ . The bottom three figures correspond to  $n = 1000$  with, (d)  $\sigma_X = 0.75$ , (e)  $\sigma_X = 1$ , and (f)  $\sigma_X = 1.5$ .

We also compared the performances of the bootstrap point-wise CIs for the SIMEX-based estimators through simulations with empirical standard errors. For that, we simulated a data set with  $\sigma_X = 1.5$  and  $\sigma_U = 0.75$  for sample sizes  $n = (200, 500, 1000)$ . We next simulated 500 bootstrap samples and obtained the SIMEX-based estimate of the coefficient function for each bootstrap sample. At each grid point, a bootstrap 95% CI was calculated using the standard error from the 500 bootstrap samples. We also calculated the empirical 95% point-wise CIs based on the 500 simulations. Both CIs were derived under the assumption of asymptotic normality of  $\hat{\beta}(t)$  where the bootstrap of  $\beta(t)$  was obtained as the average of the  $\hat{\beta}(t)$  across the 500 bootstrap samples. Our results indicated that the bootstrap CIs were generally wider than the empirically based CIs; however, the gap between the two CIs decreased with increasing sample sizes (see Figure 8).

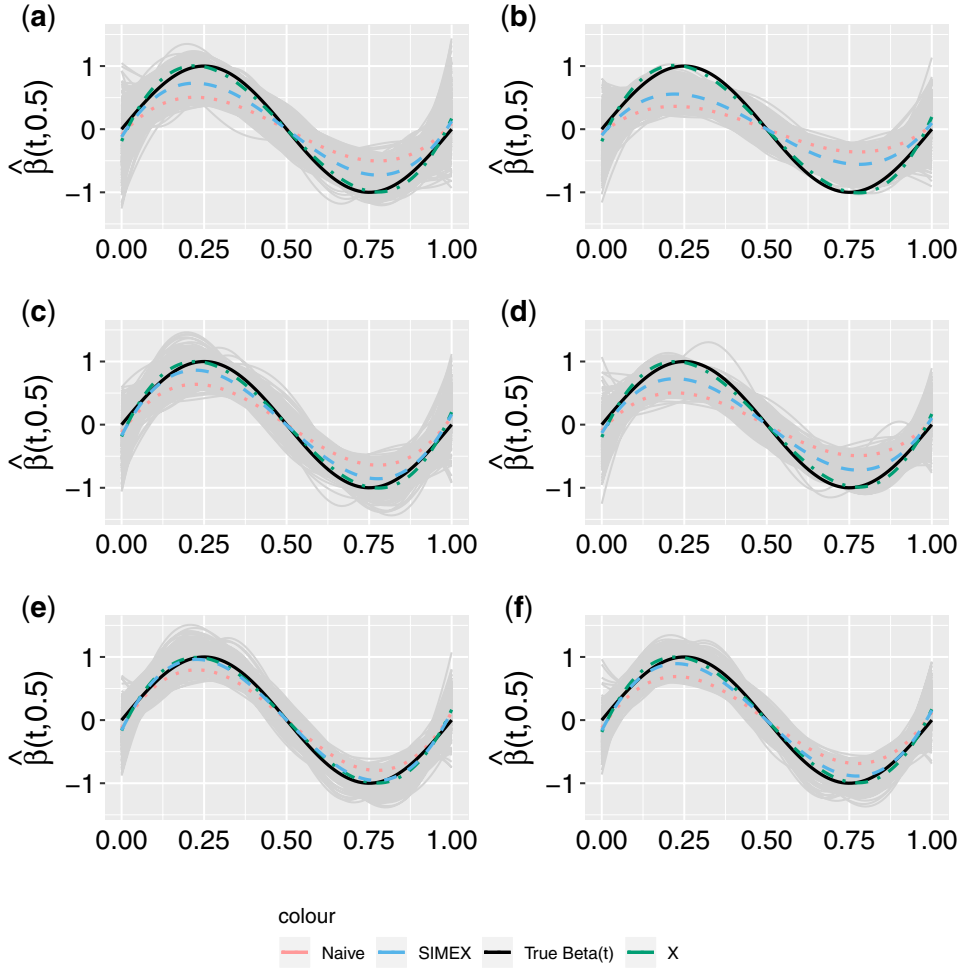


Fig. 7. Plots of the estimated regression coefficient functions for the 50th quantile function from Simulation Study 1 when  $n = 1000$ , with (a)  $\sigma_X = 0.75, \sigma_U = 0.75$ ; (b)  $\sigma_X = 0.75, \sigma_U = 1$ ; (c),  $\sigma_X = 1, \sigma_U = 0.75$ ; (d),  $\sigma_X = 1, \sigma_U = 1$ ; (e)  $\sigma_X = 1.5, \sigma_U = 0.75$ ; and (f)  $\sigma_X = 1.5, \sigma_U = 1$ . In each plot, the solid line represents the true coefficient function, while the dotted, dashed, and dot-dashed lines represent the averaged coefficient estimators over 500 replicates from the naive, the SIMEX, and the oracle methods, respectively. The estimated coefficient functions from the SIMEX-based approach were plotted in the background as gray lines based on 500 replicates.

## 5.2. Simulation Study 2

We also conducted another set of simulations to assess the impacts of measurement error on  $\beta(t, \tau)$  and  $\alpha(\tau)$ , the quantile-specific coefficients on the error-free and function-valued covariates, respectively. We generated 500 data sets with sample sizes  $n = (200, 500, 1000, 2000)$  independently from model (2.1) with  $X(t) = 1/[1 + \exp[8(t - 0.5)]] + 1 + \varepsilon_x(t)$ ,  $W(t) = X(t) + U(t)$ , and  $M(t) = \delta(t)X(t) + \eta(t)$  for  $t \in [0, 1]$ . The  $\delta(t)$  was generated as  $\delta(t) = c \sin(2\pi t) + 1$ , where  $c$  is a constant. The error terms  $\varepsilon_x(t)$ ,  $U(t)$ ,  $\eta(t)$ , and  $Y$  were generated as described in Simulation Study 1. We considered  $\sigma_x = 1.5$ ,  $\sigma_U = 0.75$ , and  $\sigma_\eta = 0.25$ . Two error-free covariates were included. The first error-free covariate,  $Z_1$ , was generated independently of  $X(t)$  from a Gaussian process with 0 mean and variance of  $0.5^2$ , i.e.,

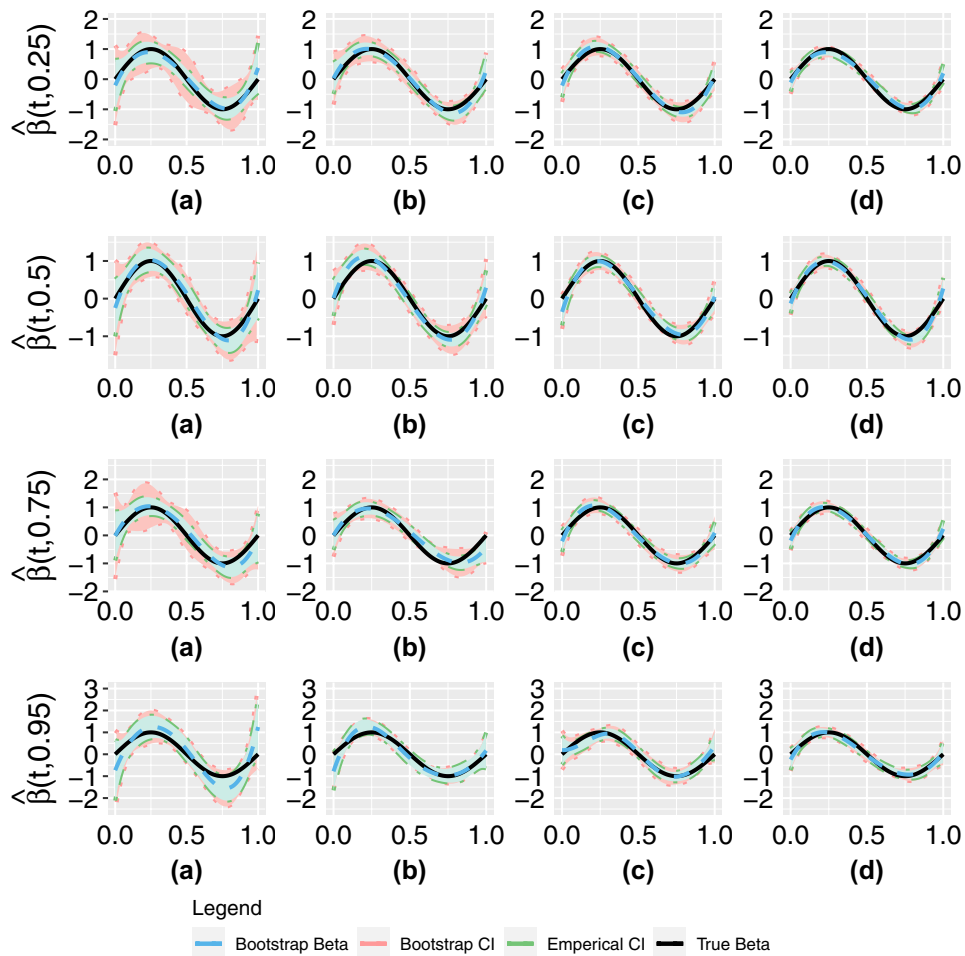


Fig. 8. Comparisons of bootstrap point-wise and empirical CIs by sample sizes for all the quantile functions considered for Simulation Study 1.  $\hat{\beta}(t, \tau)$  when  $\sigma_X = 1.5$ ,  $\sigma_W = 0.75$  at the 25th, 50th, 75th, and 95th quantiles for  $n =$  (a) 200, (b) 500, (c) 1000, and (d) 2000. The dotted lines represent the upper and lower bounds of 95% CI of  $\hat{\beta}(t, \tau)$  from bootstrap. The solid line represents the true  $\beta(t, \tau)$ . The dot-dashed lines represent the upper and lower bounds of the 95% CI for the empirical  $\hat{\beta}(t, \tau)$  from the simulations. While the dashed line represents the average  $\hat{\beta}(t, \tau)$  from bootstrap.

$Z_1 \sim N(0, 0.5^2)$ . While the second error-free covariate,  $Z_2$ , was generated from a Bernoulli distribution such that the probability of success was 0.6, i.e.,  $Z_2 \sim \text{Bernoulli}(0.6)$ . The simulations were performed at the 50th percentile with 500 replicates.

To determine how the presence of error-free covariates influence  $\hat{\beta}(t, \tau)$ , we compared the results in Table 3 when  $c = 0$  to the results for the 50th percentile in Table 2. In general, the impact of error-free covariates on the biases associated with  $\beta(t, \tau)$  in the presence of error-free covariates depended on the sample size. For the larger sample sizes ( $n \geq 500$ ), including the error-free covariates led to smaller biases associated with the function-valued covariate.

Table 3. Assessing the impacts of assuming  $\delta(t)$  is known or unknown and estimated on  $\widehat{\beta(t, 0.50)}$  from Simulation Study 2 in the presence of error-free covariates at the 50th quantile function  $\{t = 100, \rho = 0.5, \tau = 0.5, c = (0, 0.25, 0.5, 0.75), sd_x = 1.5, sd_w = 0.75, n = (200, 500, 1000, 2000)\}$ . SIMEX 1 and SIMEX 2 represent the SIMEX-based estimator when  $\delta(t)$  is estimated or assumed to be known, respectively.

N	Method	c = 0			c = 0.25			c = 0.5			c = 0.75		
		ABias <sup>2</sup>	AVar	AMSE	ABias <sup>2</sup>	AVar	AMSE	ABias <sup>2</sup>	AVar	AMSE	ABias <sup>2</sup>	AVar	AMSE
200	Oracle	0.0048	0.0307	0.0356	0.0048	0.0307	0.0356	0.0048	0.0307	0.0356	0.0048	0.0307	0.0356
200	Naive	0.0270	0.0267	0.0540	0.0270	0.0267	0.0540	0.0270	0.0267	0.0540	0.0270	0.0267	0.0540
200	SIMEX 1	0.0055	0.3673	0.3950	0.0051	0.3844	0.4102	0.0049	0.4179	0.4425	0.0073	0.4845	0.5214
200	SIMEX 2	0.0060	0.3398	0.3704	0.0056	0.3557	0.3841	0.0050	0.3890	0.4143	0.0058	0.4606	0.4897
500	Oracle	0.0001	0.0138	0.0138	0.0001	0.0138	0.0138	0.0001	0.0138	0.0138	0.0001	0.0138	0.0138
500	Naive	0.0190	0.0121	0.0313	0.0190	0.0121	0.0313	0.0190	0.0121	0.0313	0.0190	0.0121	0.0313
500	SIMEX 1	0.0007	0.1739	0.1773	0.0006	0.1856	0.1885	0.0005	0.2080	0.2106	0.0011	0.2587	0.2642
500	SIMEX 2	0.0006	0.1647	0.1679	0.0056	0.3557	0.3841	0.0005	0.1898	0.1925	0.0007	0.2340	0.2378
1000	Oracle	0.0001	0.0076	0.0077	0.0001	0.0076	0.0077	0.0001	0.0076	0.0077	0.0001	0.0076	0.0077
1000	Naive	0.0202	0.0067	0.0271	0.0202	0.0067	0.0271	0.0202	0.0067	0.0271	0.0202	0.0067	0.0271
1000	SIMEX 1	0.0010	0.1120	0.1169	0.0009	0.1175	0.1219	0.0008	0.1341	0.1381	0.0011	0.1766	0.1821
1000	SIMEX 2	0.0009	0.0950	0.0993	0.0056	0.3557	0.3841	0.0008	0.1105	0.1146	0.0009	0.1436	0.1480
2000	Oracle	0.0001	0.0044	0.0045	0.0001	0.0044	0.0045	0.0001	0.0044	0.0045	0.0001	0.0044	0.0045
2000	Naive	0.0199	0.0039	0.0240	0.0199	0.0039	0.0240	0.0199	0.0039	0.0240	0.0199	0.0039	0.0240
2000	SIMEX 1	0.0008	0.0614	0.0653	0.0008	0.0648	0.0690	0.0008	0.0752	0.0793	0.0006	0.1035	0.1064
2000	SIMEX 2	0.0008	0.0527	0.0567	0.0056	0.3557	0.3841	0.0009	0.0627	0.0674	0.0008	0.0869	0.0909

Simulations were also conducted to assess the influence of  $\delta(t)$  on the SIMEX approach to estimating  $\beta(t, \tau)$ . To do this, we estimated  $\beta(t, \tau)$ , the SIMEX-based estimator at the different quantile levels under two assumptions related to  $\delta(t)$ . We first assumed an unknown but estimable  $\delta(t)$  (SIMEX1) as discussed in Section 2.2 and a known  $\delta(t)$  SIMEX2. We provide the results for the 50th percentile in Table 3. The finite sample properties associated with estimating  $\beta(t, 0.50)$  under the two assumptions were equivalent (Table 3).

To investigate the influence of misspecifying  $\delta(t)$  on estimating  $\beta(t, \tau)$ , additional simulations were performed with a possibly misspecified  $\delta(t)$  (SIMEXM) (results shown in Table S2 of Supplementary material available at *Biostatistics* online). The SIMEXM estimator was obtained under the assumption that  $\delta(t) = 1$ . Under the SIMEXM approach,  $\delta(t)$  is correctly specified only when  $c = 0$ . We computed  $\beta(t, \tau)$  under SIMEX1, SIMEX2, and SIMEXM. We found that misspecification of  $\delta(t)$  led to larger ABias<sup>2</sup>. We also observed that when  $c = 0$ , the biases associated with SIMEXM were comparable to those associated with SIMEX1 and SIMEX2. However, when  $c \neq 0$  and  $\delta(t) \neq 1$ , the biases associated with the SIMEXM estimators were much larger than the SIMEX1 and SIMEX2 estimators. Additionally, the simulation results suggested that the bias of the SIMEX-based estimator obtained under the raw (unsmoothed) estimation of  $\delta(t)$  was close to that of the oracle method and noticeably smaller than that of the naive method for all sample sizes (200, 500, 1000, 2000), when  $\frac{1}{n} \sum_{i=1}^n W_i(t)$  was not close to 0.

Simulation studies were also conducted to assess the influence of measurement errors on the error-free covariates at the 25th, 50th, 75th, and 95th percentiles with  $n = 500$  (Tables 4 and 5). Estimated absolute biases were used to compare the oracle, naive, and SIMEX-based estimators. The small values of the absolute biases all indicated that the presence of measurement error in  $W(t)$  did not influence the

Table 4. Assessing the impacts of measurement error on the continuous error-free covariate from Simulation Study 2. Absolute bias ( $|Bias|$ ), averaged sample variance ( $Avar$ ) and averaged integrated mean squared error ( $AIMSE$ ) for the continuous error-free covariate at the 25th, 50th, 75th, and 95th quantile function  $\{t = 100, \rho = 0.5, c = 0, 0.25, 0.5, 0.75, sd_x = 1.5, sd_w = 0.75, n = 500$ . SIMEX 1 and SIMEX 2 represent the SIMEX-based estimator when  $\delta(t)$  is estimated or assumed to be known, respectively.

$\tau$	Method	$c = 0$			$c = 0.25$			$c = 0.5$			$c = 0.75$		
		$ Bias $	Var	MSE	$ Bias $	Var	MSE	$ Bias $	Var	MSE	$ Bias $	Var	MSE
0.25	Oracle	0.00046	0.00016	0.00016	0.00046	0.00016	0.00016	0.00046	0.00016	0.00016	0.00046	0.00016	0.00016
0.25	Naive	0.00120	0.00018	0.00018	0.00120	0.00018	0.00018	0.00120	0.00018	0.00018	0.00120	0.00018	0.00018
0.25	SIMEX 1	0.00079	0.00025	0.00025	0.00064	0.00026	0.00026	0.00064	0.00027	0.00027	0.00064	0.00028	0.00028
0.25	SIMEX 2	0.00026	0.00026	0.00052	0.00026	0.00026	0.00051	0.00026	0.00026	0.00052	0.00026	0.00026	0.00053
0.5	Oracle	0.00021	0.00013	0.00013	0.00021	0.00013	0.00013	0.00021	0.00013	0.00013	0.00021	0.00013	0.00013
0.5	Naive	0.00032	0.00015	0.00015	0.00032	0.00015	0.00015	0.00032	0.00015	0.00015	0.00032	0.00015	0.00015
0.5	SIMEX 1	0.00045	0.00021	0.00021	0.00049	0.00020	0.00020	0.00029	0.00020	0.00020	0.00019	0.00020	0.00020
0.5	SIMEX 2	0.00021	0.00021	0.00042	0.00057	0.00057	0.00114	0.00020	0.00020	0.00041	0.00020	0.00020	0.00040
0.75	Oracle	0.00061	0.00014	0.00014	0.00061	0.00014	0.00014	0.00061	0.00014	0.00014	0.00061	0.00014	0.00014
0.75	Naive	0.00064	0.00018	0.00019	0.00064	0.00018	0.00019	0.00064	0.00018	0.00019	0.00064	0.00018	0.00019
0.75	SIMEX 1	0.00204	0.00029	0.00030	0.00192	0.00029	0.00030	0.00174	0.00029	0.00029	0.00135	0.00029	0.00029
0.75	SIMEX 2	0.00028	0.00028	0.00057	0.00029	0.00029	0.00058	0.00029	0.00029	0.00058	0.00029	0.00029	0.00058
0.95	Oracle	0.00081	0.00035	0.00035	0.00081	0.00035	0.00035	0.00081	0.00035	0.00035	0.00081	0.00035	0.00035
0.95	Naive	0.00154	0.00038	0.00039	0.00154	0.00038	0.00039	0.00154	0.00038	0.00039	0.00154	0.00038	0.00039
0.95	SIMEX 1	0.00096	0.00059	0.00059	0.00109	0.00059	0.00059	0.00136	0.00059	0.00059	0.00156	0.00060	0.00061
0.95	SIMEX 2	0.00060	0.00061	0.00121	0.00061	0.00061	0.00122	0.00058	0.00059	0.00117	0.00058	0.00058	0.00117

Table 5. Assessing the impacts of measurement error on the binary error-free covariate from Simulation Study 2. Absolute bias ( $|Bias|$ ), averaged sample variance ( $Avar$ ) and averaged integrated mean squared error ( $AIMSE$ ) for the continuous error-free covariate at the 25th, 50th, 75th, and 95th quantile function  $\{t = 100, \rho = 0.5, c = 0, 0.25, 0.5, 0.75, sd_x = 1.5, sd_w = 0.75, n = 500$ . SIMEX 1 and SIMEX 2 represent the SIMEX-based estimator when  $\delta(t)$  is estimated or assumed to be known, respectively.

$\tau$	Method	$c = 0$			$c = 0.25$			$c = 0.5$			$c = 0.75$		
		$ Bias $	Var	MSE	$ Bias $	Var	MSE	$ Bias $	Var	MSE	$ Bias $	Var	MSE
0.25	Oracle	0.00099	0.00017	0.00017	0.00099	0.00017	0.00017	0.00099	0.00017	0.00017	0.00099	0.00017	0.00017
0.25	Naive	0.00146	0.00018	0.00019	0.00146	0.00018	0.00019	0.00146	0.00018	0.00019	0.00146	0.00018	0.00019
0.25	SIMEX 1	0.00018	0.00025	0.00025	0.00025	0.00024	0.00024	0.00058	0.00025	0.00025	0.00084	0.00026	0.00026
0.25	SIMEX 2	0.00025	0.00025	0.00104	0.00026	0.00026	0.00090	0.00027	0.00027	0.00091	0.00027	0.00027	0.00091
0.5	Oracle	0.00007	0.00014	0.00014	0.00007	0.00014	0.00014	0.00007	0.00014	0.00014	0.00007	0.00014	0.00014
0.5	Naive	0.00012	0.00016	0.00016	0.00012	0.00016	0.00016	0.00012	0.00016	0.00016	0.00012	0.00016	0.00016
0.5	SIMEX 1	0.00017	0.00022	0.00022	0.00016	0.00022	0.00022	0.00026	0.00022	0.00022	0.00033	0.00023	0.00023
0.5	SIMEX 2	0.00022	0.00022	0.00067	0.00055	0.00055	0.00169	0.00022	0.00022	0.00052	0.00022	0.00022	0.00041
0.75	Oracle	0.00042	0.00018	0.00018	0.00042	0.00018	0.00018	0.00042	0.00018	0.00018	0.00042	0.00018	0.00018
0.75	Naive	0.00078	0.00019	0.00019	0.00078	0.00019	0.00019	0.00078	0.00019	0.00019	0.00078	0.00019	0.00019
0.75	SIMEX 1	0.00101	0.00028	0.00028	0.00080	0.00028	0.00028	0.00070	0.00027	0.00027	0.00121	0.00027	0.00027
0.75	SIMEX 2	0.00028	0.00028	0.00232	0.00028	0.00028	0.00220	0.00028	0.00028	0.00202	0.00028	0.00028	0.00163
0.95	Oracle	0.00173	0.00042	0.00042	0.00173	0.00042	0.00042	0.00173	0.00042	0.00042	0.00173	0.00042	0.00042
0.95	Naive	0.00052	0.00040	0.00040	0.00052	0.00040	0.00040	0.00052	0.00040	0.00040	0.00052	0.00040	0.00040
0.95	SIMEX 1	0.00100	0.00057	0.00057	0.00153	0.00058	0.00059	0.00141	0.00056	0.00056	0.00164	0.00054	0.00054
0.95	SIMEX 2	0.00058	0.00058	0.00154	0.00058	0.00059	0.00168	0.00059	0.00059	0.00196	0.00055	0.00056	0.00211

estimated coefficients associated with the two error-free covariates under all the quantile levels considered. The minimal impact of the measurement error on the estimated coefficients for the error-free covariates was consistent under both assumptions of known and unknown  $\delta(t)$ .

Simulations were also performed to assess the impact of violation of the independence assumption between the instrumental variable and the measurement errors on  $\hat{\beta}(t, \tau)$  under all quantile functions considered (Tables S3–S6 of [Supplementary material](#) available at *Biostatistics* online). Under the assumptions of both known and unknown  $\delta(t)$ , the biases associated with the naive-based estimators were lower than the SIMEX-based estimators when  $\delta(t)$  is estimated. However, when  $c = 0$  and  $\delta(t) = 1$ , the biases associated with the SIMEX-based estimators were smaller than the naive-based approach. In summary, we find that when the independence assumption between  $M(t)$  and  $U(t)$  is violated, the SIMEX-based method performs better than the naive-based method when the  $\delta(t) = 1$ .

## 6. DISCUSSION

We studied the quantile regression models with a function-valued covariate with measurement error. To date, most approaches designed to address measurement error in quantile regression have focused on cases where the error-prone covariate is scalar valued. Although some authors have studied quantile regression models with function-valued covariates, they considered the error terms associated with the models as random errors, and not classical measurement errors with complex error structures. In our proposed methods, we considered the function-valued covariate,  $W(t)$ , to be a surrogate or observed measure of a latent function-valued covariate,  $X(t)$ , prone to heteroscedastic classical measurement error. We used a function-valued instrumental variable to estimate the covariance matrix of the measurement error and the SIMEX approach was used to estimate the error-corrected function-valued coefficient. Through simulations, we compared the finite sample performance of our proposed estimator to estimators obtained under the oracle method,  $X(t)$ , and the naive uncorrected approach,  $W(t)$ . Our results illustrated the importance of correcting for measurement error when there is a considerable amount of measurement error associated with the function-valued covariate. The biases of our proposed SIMEX-based estimator were equivalent to those estimated for  $X(t)$  as the sample size increased. The estimated variances for the naive-based methods tended to be smaller than those for both the SIMEX-based estimator and  $X(t)$ .

We successfully applied our methods to the NHANES database. In our application, we estimated the relationship between wearable device-based measures of physical activity and quantile functions of BMI among adults 20–85 years of age living in the United States. We found that the relationship between physical activity intensity and quantile functions of BMI depended on the timing of the physical activity after adjusting for error-free covariates including age, sex, and race.

## 7. SOFTWARE

We provide our simulation codes in the Github repository located at <https://github.com/ctekwe/Quantile-SIMEX->. An Rmd file is also included to provide instructions on how to use the provided codes.

## SUPPLEMENTARY MATERIAL

[Supplementary material](http://biostatistics.oxfordjournals.org) is available at <http://biostatistics.oxfordjournals.org>.

## ACKNOWLEDGMENTS

*Conflict of Interest:* None declared.

## FUNDING

The National Cancer Institute Supplemental Award (U01-CA057030-29S2) and the National Institute of Diabetes and Digestive and Kidney Diseases Award (1R01DK132385-01) to C.D.T.; the National Cancer Institute Award (U01-CA057030-29 to R.J.C.); the National Science Foundation Award DMS (1812258) and the National Institute of Diabetes and Digestive and Kidney Diseases Award (1R01DK132385-01) to L.X.; the National Cancer Institute Supplemental Award (U01-CA057030-29S1) and the National Institute of Diabetes and Digestive and Kidney Diseases Award (1R01DK132385-01) to R.S.Z.; the National Institute of Child Health and Human Development Award (1R03HD084807-01A1 to M.G.).

## REFERENCES

- BASSETT, D. R. (2012). Device-based monitoring in physical activity and public health research. *Physiological Measurement* **33**, 1769.
- BAUR, D. G., DIMPFL, T. AND JUNG, R. C. (2012). Stock return autocorrelations revisited: a quantile regression approach. *Journal of Empirical Finance* **19**, 254–265.
- BOLLEN, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology* **38**, 37–72.
- BOTTAI, M., FRONGILLO, E. A., SUI, X., O'NEILL, J. R., MCKEOWN, R. E., BURNS, T. L., LIESE, A. D., BLAIR, S. N. AND PATE, R. R. (2014). Use of quantile regression to investigate the longitudinal association between physical activity and body mass index. *Obesity* **22**, E149–E156.
- CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. New York, NY: Chapman and Hall.
- CARROLL, R. J. AND STEFANSKI, L. A. (1997). Asymptotic theory for the SIMEX estimator in measurement error models. In: *Advances in Statistical Decision Theory and Applications*. Birkhäuser Boston: Springer, pp. 151–164.
- CARTER, S. J., BARANAUSKAS, M. N., SINGH, H., MARTINS, C. AND HUNTER, G. R. (2022). Arte index revisited: linking biomarkers of cardio-metabolic health with free-living physical activity in postmenopausal women. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **322**, R292–R298.
- CARTER, S. J., ROGERS, L. Q., BOWLES, H. R., NORIAN, L. A. AND HUNTER, G. R. (2019). Inverse association between changes in energetic cost of walking and vertical accelerations in non-metastatic breast cancer survivors. *European Journal of Applied Physiology* **119**, 2457–2464.
- CHEN, K. AND MÜLLER, H. G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 67–89.
- COOK, J. R. AND STEFANSKI, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89**, 1314–1328.
- CORDER, K., BRAGE, S. AND EKELEND, U. (2007). Accelerometers and pedometers: methodology and clinical application. *Current Opinion in Clinical Nutrition & Metabolic Care* **10**, 597–603.
- CROUTER, S. E., CHURILLA, J. R. AND BASSETT, D. R. (2006). Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology* **98**, 601–612.
- EVENSON, K. R., WEN, F., METZGER, J. S. AND HERRING, A. H. (2015). Physical activity and sedentary behavior patterns using accelerometry from a national sample of United States adults. *International Journal of Behavioral Nutrition and Physical Activity* **12**, 20.
- FITZENBERGER, B., KOENKER, R., MACHADO, J. AND MELLY, B. (2022). Economic applications of quantile regression 2.0. *Empirical Economics*, **62**, 1–6.
- FULLER, W. A. (2009). *Measurement Error Models*. Hoboken, NJ: Wiley.



- GERACI, M. (2016). Qtools: a collection of models and other tools for quantile inference. *R Journal* **8**, 117–138.
- GERACI, M. AND BOTTAI, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* **8**, 140–154.
- GUAN, J., CHENG, H., BOLLEN, K. A., THOMAS, D. R. AND WANG, L. (2019). Instrumental variable estimation in ordinal probit models with mismeasured predictors. *Canadian Journal of Statistics* **47**, 653–667.
- HE, X. AND LIANG, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, 129–140.
- JACOBI, D., PERRIN, A. E., GROSMAN, N., DORÉ, M. F., NORMAND, S., OPPERT, J. M. AND SIMON, C. (2007). Physical activity-related energy expenditure with the RT3 and TriTrac accelerometers in overweight adults. *Obesity* **15**, 950–956.
- JADHAV, S., TEKWE, C. D. AND LUAN, Y. (2021). A function-based approach to model the measurement error in wearable devices. *Statistics in Medicine*. In press.
- JOHNSON, C. L., PAULOSE-RAM, R., OGDEN, C. L., CARROLL, M. D., KRUSZAN-MORAN, D., DOHRMANN, S. M. AND CURTIN, L. R. (2013). National health and nutrition examination survey. Analytic guidelines, 1999–2010.
- KATO, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics* **40**, 3108–3136.
- KOENKER, R. AND HALLOCK, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, **15**, 143–156.
- KOENKER, R. AND BASSETT, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.
- MACHADO, J. A. F. AND MATA, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* **20**, 445–465.
- MAO, G. AND WEI, Y. AND LIU, Y. (2017). Simex method for censored quantile regression with measurement error. *Communications in Statistics-Simulation and Computation* **46**, 7552–7560.
- MARTINS, P. S. AND PEREIRA, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics* **11**, 355–371.
- MATTHEWS, C. E., CHEN, K. Y., FREEDSON, P. S., BUCHOWSKI, M. S., BEECH, B. M., PATE, R. R. AND TROIANO, R. P. (2008). Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American Journal of Epidemiology* **167**, 875–881.
- NEUMEIER, W. H., GOODNER, E., BIASINI, F., DHURANDHAR, E. J., MENEAR, K. S., TURAN, B. AND HUNTER, G. R. (2016). Exercise following mental work prevented overeating. *Medicine and Science in Sports and Exercise* **48**, 1803.
- PLANQUE, B. AND BUFFAZ, L. (2008). Quantile regression models for fish recruitment–environment relationships: four case studies. *Marine Ecology Progress Series* **357**, 213–223.
- ROBERTSON, W., STEWART-BROWN, S., WILCOCK, E., OLDFIELD, MICHELLE AND THOROGOOD, M. (2011). Utility of accelerometers to measure physical activity in children attending an obesity treatment intervention. *Journal of Obesity* **2011**.
- ROTHNEY, M. P., SCHAEFER, E. V., NEUMANN, M. M., CHOI, L. AND CHEN, K. Y. (2008). Validity of physical activity intensity predictions by Actigraph, Actical, and RT3 accelerometers. *Obesity* **16**, 1946–1952.
- SCHRÖDER, H. K., ANDERSEN, H. E., KIEHL, K. AND KENKEL, N. (2005). Rejecting the mean: Estimating the response of fen plant species to environmental factors by non-linear quantile regression. *Journal of Vegetation Science* **16**, 373–382.
- SHANG, Y. (2012). Measurement error adjustment using the SIMEX method: an application to student growth percentiles. *Journal of Educational Measurement* **49**, 446–465.
- SILVERMAN, B. W. AND RAMSAY, J. O. (2005). *Functional Data Analysis*. New York, NY: Springer.

- SPIEGELMAN, B. M. AND FLIER, J. S. (2001). Obesity and the regulation of energy balance. *Cell* **104**, 531–543.
- STEFANSKI, L. A. AND COOK, J. R. (1995). Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association* **90**, 1247–1256.
- TEKWE, C. D., CARTER, R. L. AND CULLINGS, H. M. (2016). Generalized multiple indicators, multiple causes measurement error models. *Statistical Modelling* **16**, 140–159.
- TEKWE, C. D., CARTER, R. L., CULLINGS, H. M. AND CARROLL, R. J. (2014). Multiple indicators, multiple causes measurement error models. *Statistics in Medicine* **33**, 4469–4481.
- TEKWE, C. D., ZOH, R. S., YANG, M., CARROLL, R. J., HONVOH, G., ALLISON, D. B., BENDEN, M. AND XUE, L. (2019). Instrumental variable approach to estimating the scalar-on-function regression model with measurement error with application to energy expenditure assessment in childhood obesity. *Statistics in Medicine* **38**, 3764–3781.
- TROIANO, R. P., BERRIGAN, D., DODD, K. W., MASSE, L. C., TILERT, T., MCDOWELL, M. and others. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise* **40**, 181.
- TSAI, I. C. (2012). The relationship between stock price index and exchange rate in asian markets: a quantile regression approach. *Journal of International Financial Markets, Institutions and Money* **22**, 609–621.
- WANG, H. J., STEFANSKI, L. A. AND ZHU, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika* **99**, 405.
- WAROLIN, J., CARRICO, A. R., WHITAKER, L. E., WANG, L., CHEN, K. Y., ACRA, S. AND BUCHOWSKI, M. S. (2012). Effect of BMI on prediction of accelerometry-based energy expenditure in youth. *Medicine and Science in Sports and Exercise* **44**, 2428.
- WEI, Y. AND CARROLL, R. J. (2009). Quantile regression with measurement error. *Journal of the American Statistical Association* **104**, 1129–1143.
- WINKELMANN, R. (2006). Reforming health care: evidence from quantile regressions for counts. *Journal of Health Economics* **25**, 131–145.
- YU, K., LU, Z. AND STANDER, J. (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**, 331–350.

[Received March 23, 2021; revised February 17, 2022; accepted for publication March 9, 2022]