The Canadian Journal of Statistics
Vol. 47, No. 2, 2019, Pages 262–280
La revue canadienne de statistique

Linear mode regression with covariate measurement error

Xiang LI and Xianzheng HUANG

Department of Statistics, University of South Carolina, Columbia, SC, U.S.A.

Key words and phrases: Bandwidth; corrected score; deconvoluting kernel; simulation-extrapolation.

MSC 2010: Primary 62J05; secondary 62G08

Abstract: We consider estimating the mode of a response given an error-prone covariate. It is shown that ignoring measurement error typically leads to inconsistent inference for the conditional mode of the response given the true covariate, as well as misleading inference for regression coefficients in the conditional mode model. To account for measurement error, we first employ the Monte Carlo corrected score method (Novick & Stefanski, 2002) to obtain an unbiased score function based on which the regression coefficients can be estimated consistently. To relax the normality assumption on measurement error this method requires, we propose another method where deconvoluting kernels are used to construct an objective function that is maximized to obtain consistent estimators of the regression coefficients. Besides rigorous investigation on asymptotic properties of the new estimators, we study their finite sample performance via extensive simulation experiments, and find that the proposed methods substantially outperform a naive inference method that ignores measurement error. *The Canadian Journal of Statistics* 47: 262–280; 2019 © 2019 Statistical Society of Canada

Résumé: Les auteures considèrent l'estimation du mode d'une variable réponse étant donné une covariable sujette à l'erreur. Le fait d'ignorer les erreurs de mesure mène souvent à un manque de convergence du mode conditionnel à la vraie covariable, ainsi qu'à une inférence trompeuse découlant des coefficients de régression du modèle conditionnel pour le mode. Pour tenir compte des erreurs de mesure, les auteures exploitent la méthode du score corrigé par Monte Carlo (Novick et Stefanski, 2002) afin d'obtenir une fonction score sans biais à partir de laquelle des estimateurs convergents des coefficients de régression peuvent être obtenus. Afin d'assouplir l'hypothèse de normalité des erreurs de mesure requise par cette méthode, les auteures proposent une autre approche dans laquelle des noyaux de déconvolution permettent de construire une fonction objective qui est maximisée pour obtenir des estimateurs convergents des coefficients de régression. En plus d'une étude rigoureuse des propriétés asymptotiques des nouveaux estimateurs, les auteures décrivent leur performance pour des échantillons finis à l'aide d'expériences de simulation substantielles. Elles constatent que la méthode proposée offre une performance considérablement meilleure que les méthodes d'inférence naïves qui ignorent les erreurs de mesure. *La revue canadienne de statistique* 47: 262–280; 2019 © 2019 Société statistique du Canada

1. INTRODUCTION

Regression analysis has been a standard platform to study the association between a response, Y, and covariates of interest, X. The majority of the literature on regression analysis is devoted to mean regression, where the mean of Y given X is the focal point of inference. There also exists a large body of work on quantile regression, where one infers quantiles of Y conditioning on X

Additional Supporting Information may be found in the online version of this article at the publisher's website.

* *Author to whom correspondence may be addressed.*
E-mail: huang@stat.sc.edu

(Koenker, 2005). In contrast, there have been much less study on mode regression (Lee, 1989; Yao & Li, 2014; Chen et al., 2016), which aims to characterize the mode of Y given X. The mode of a distribution is an informative summary feature that is more of interest than the mean or quantiles in many applications (Parzen, 1962), such as biology (Hedges & Shah, 2003), economy (Huang & Yao, 2012), meteorology (Hyndman, Bashtannyk & Grunwald, 1996), astronomy (Bamford et al., 2008) and traffic engineering (Einbeck & Tutz, 2006), where the underlying distributions of Y given X are often skewed. In these referenced works, the most likely value of Y given a covariate value, as opposed to some average value of the response, is of scientific interest; and a location measure that is resistant to outliers, such as the mode, is more appealing. In these applications, some covariates cannot be measured directly or precisely, and only data for their error-contaminated surrogates are collected.

To address complications caused by error-prone covariates, a good collection of methods for mean regression that account for covariate measurement error have been developed (Carroll et al., 2006; Fuller, 2009; Yi, 2017). There are also some approaches that take measurement error into consideration in quantile regression (He & Liang, 2000; Wei & Carroll, 2009; Wang, Stefanski, & Zhu, 2012). However, there is little research on mode regression in the presence of measurement error in covariates. The only work we are aware of is Zhou & Huang (2016), in which the authors proposed nonparametric methods to estimate the mode of Y given X based on kernel density estimators. Differing from the nonparametric route they took, here we consider a class of linear mode regression models, following the footsteps of existing works on mean regression (Fuller, 2009) and quantile regression (He & Liang, 2000; Wei & Carroll, 2009; Wang, Stefanski, & Zhu, 2012) with measurement error, where one starts by considering the conditional mean or quantiles as some linear function of covariates. This class of mode regression models has been mostly investigated by econometricians (Lee, 1989, 1993; Kemp & Silva 2012), and all existing works assume error-free covariates. To the best of our knowledge, we are the first to investigate linear mode regression with covariate measurement error.

The rest of the article is organized as follows. We first formulate the class of linear measurement error mode models in Section 2, and provide some preliminary analysis on the effect of measurement error on inference when one ignores measurement error. We propose two methods to estimate the regression coefficients in the model that account for measurement error in Section 3. Both methods depend on the choice of bandwidth, and we present a strategy of choosing a suitable bandwidth in Section 4. Section 5 reports simulation studies where we compare the two proposed methods with a naive method that ignores measurement error, using estimates from the method proposed by Yao & Li (2014) applied to error-free data as benchmarks. Section 6 presents an application of the three methods to dietary data collected from the Women's Interview Survey of Health. We point out extensions of the proposed methods under more general settings and discuss follow-up research agendas in Section 7.

2. PREAMBLE

2.1. Data and Models

Suppose that the observed data consist of n independent data points, $\{(Y_j, W_j)\}_{j=1}^n$, where $\{W_j\}_{j=1}^n$ are surrogates of the unobserved covariate values, $\{X_j\}_{j=1}^n$, and Y_j given X_j follows a distribution specified by the probability density function $f_{Y|X}(y \mid x)$, for $j = 1, \ldots, n$. As in Grund & Hall (1995), we assume that $f_{Y|X}(y \mid x)$ has a unique largest mode; in particular, we assume a linear model for this conditional mode,

$$y_M(x) = \text{Mode}(Y_j \mid X_j = x) = \beta_0 + \beta_1 x \quad \text{for } j = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$ is the regression coefficient vector containing parameters to be estimated.

A classical additive measurement error model is assumed in this article,

$$W_j = X_j + U_j, \tag{2}$$

where U_j is the nondifferential measurement error (Carroll et al., 2006, Section 2.5) for $j = 1, \ldots, n$, following a distribution specified by the density function $f_U(u)$, of which the mean is zero and variance is σ^2.

Measurement error in (2) being nondifferential essentially implies that, conditioning on X, Y and W are independent, where the index j is suppressed when we refer to a generic data point, X_j, Y_j, or W_j, for $j \in \{1, \ldots, n\}$. For model identifiability reasons, we assume $f_U(u)$ is entirely known, including parameters associated with the distribution. Considerations for cases where extra data are available to infer $f_U(u)$ are given in Section 7. Finally, we consider a univariate covariate for illustration purposes in the majority of the study, and discuss in Section 7 generalization to multivariate covariates that may include some error-free components.

2.2. Naive Inference

Denote by $y^*_M(w)$ the mode of the conditional density of Y given $W = w$, $f_{Y|W}(y \mid w)$. In the context of linear mode regression, a naive inference method infers $y^*_M(w)$ assuming, as in (1), $y^*_M(w) = \beta^*_0 + \beta^*_1 w$, where β^*_0 and β^*_1 are the regression coefficients in this posited mode model. In what follows, we use an example to demonstrate that naive inference for the mode function can be misleading.

Suppose Y given $X = x$ follows a distribution with mean $m(x) = \alpha_0 + \alpha_1 x$ and standard deviation $\sigma(x) = \gamma_0 + \gamma_1 x$, where α_0, $\alpha_1 (\neq 0)$, γ_0, and γ_1 are constants free of x. In addition, suppose $X \sim N(\mu_X, \sigma^2_X)$ and $U \sim N(0, \sigma^2)$, where μ_X and σ^2_X are the mean and variance of X, respectively. Then, conditioning on $W = w$, Y follows a distribution with mean and standard deviation given by (Fuller, 2009)

$$\begin{aligned} m^*(w) &= \alpha_0 + (1-\lambda)\alpha_1\mu_X + \lambda\alpha_1 w, \\ \sigma^*(w) &= \sqrt{\{\gamma_0 + (1-\lambda)\gamma_1\mu_X + \lambda\gamma_1 w\}^2 + (1-\lambda)\alpha_1^2\sigma_X^2}, \end{aligned} \tag{3}$$

respectively, where $\lambda = \sigma^2_X/(\sigma^2_X + \sigma^2)$ is the reliability ratio (Carroll et al., 2006, Section 3.2.1).

Define two standardized mean residuals, $e = \{Y - m(X)\}/\sigma(X)$ and $e^* = \{Y - m^*(W)\}/\sigma^*(W)$. Denote by $e_M(x)$ the mode of e given $X = x$, and by $e^*_M(w)$ the mode of e^* given $W = w$. One can show that

$$y_M(x) = m(x) + \sigma(x)e_M(x) = \alpha_0 + \alpha_1 x + (\gamma_0 + \gamma_1 x)e_M(x),$$

and similarly

$$\begin{aligned} y^*_M(w) &= m^*(w) + \sigma^*(w)e^*_M(w) \\ &= \alpha_0 + (1-\lambda)\alpha_1\mu_X + \lambda\alpha_1 w + \\ &\sqrt{\{\gamma_0 + (1-\lambda)\gamma_1\mu_X + \lambda\gamma_1 w\}^2 + (1-\lambda)\alpha_1^2\sigma_X^2}\, e^*_M(w). \end{aligned} \tag{4}$$

Comparing $y_M(x)$ and $y^*_M(w)$, one can see that, even if $e_M(x)$ and $e^*_M(w)$ are both constant functions, the naive mode $y^*_M(w)$ is not a linear function in w unless $\gamma_1 = 0$ or $\lambda = 1$, whereas the true mode $y_M(x)$ is linear in x if $e_M(x)$ does not depend on x.

This example illustrates that effects of measurement error on mode regression are in general far more complicated than those in the context of mean regression, and one typically misspecifies the functional form of the mode model in the naive mode regression. In contrast, by (3), when

$m(x)$ is linear in x, $m^*(w)$ is also linear in w when X and U are independent normal random variables, and thus the functional form of the mean model is not misspecified in the naive mean regression. In this example, if $\gamma_1 = 0$, then β_1^* in the naive mode model revealed in (4) reduces to $\lambda\alpha_1$, which is attenuated compared to $\beta_1 = \alpha_1$ in (1) when $e_M(x)$ is free of x.

3. PROPOSED METHODS

3.1. Inference in the Absence of Measurement Error

Given a fixed y in the support of Y, $Q_h(y) = n^{-1}\sum_{j=1}^{n} K_h(Y_j - y)$ is the local constant kernel density estimator (Silverman, 1986) of the density of Y evaluated at y, $f_Y(y)$, where $K(t)$ is a kernel, h is the bandwidth and $K_h(t) = K(t/h)/h$. Since the mode of Y maximizes its density function $f_Y(y)$, a sensible estimator for the mode of Y is the maximizer of $Q_h(y)$. Motivated by this viewpoint, in the absence of covariate measurement error, Yao & Li (2014) proposed to estimate $\boldsymbol{\beta}$ by maximizing

$$Q_h(\beta) = \frac{1}{n}\sum_{j=1}^{n} K_h(Y_j - \beta_0 - \beta_1 X_j). \tag{5}$$

Setting $K(t)$ as the standard normal density, Yao & Li (2014) developed an expectation-maximization algorithm to compute the estimate of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$. In addition, they derived the order of the bias and variance of $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$ as $n \to \infty$, and established its asymptotic normality.

Naive implementation of Yao and Li's method using error-contaminated data is to substitute X_j with W_j in (5), resulting in a naive objective function one maximizes with respect to $\boldsymbol{\beta}$. Denote by $\hat{\boldsymbol{\beta}}_{\mathrm{NV}}$ the resultant naive estimator of $\boldsymbol{\beta}$. To account for measurement error, we revise this naive method from two perspectives.

3.2. Monte Carlo Corrected Score Method

Maximizing $Q_h(\boldsymbol{\beta})$ in (5) with respect to $\boldsymbol{\beta}$ is equivalent to solving the score equations for $\boldsymbol{\beta}$, $\sum_{j=1}^{n}\Psi(Y_j, X_j; \boldsymbol{\beta}) = \mathbf{0}$, where $\Psi(Y_j, X_j; \boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})K_h(Y_j - \beta_0 - \beta_1 X_j)$. In the presence of measurement error, naively applying Yao and Li's method amounts to using the naive score, $\Psi(Y, W; \boldsymbol{\beta})$, in place of the true score, $\Psi(Y, X; \boldsymbol{\beta})$. One way to correct this naive score-based estimation for measurement error is to construct a score function that depends on (Y, W), whose expectation conditioning on (Y, X) is equal to $\Psi(Y, X; \boldsymbol{\beta})$. This leads to the corrected score method (Nakamura, 1990), which has found its successes in linear mean regression, several nonlinear mean regression models (Carroll et al., 2006, Chapter 7) and some survival models (Song & Huang, 2005; Wang, 2006; Zucker & Spiegelman, 2008) with covariate measurement error.

Although the idea of correcting the naive score by using an unbiased estimator of the true score leads to a general strategy to account for measurement error, such an unbiased estimator, referred to as a corrected score, often does not exist in closed form. Novick & Stefanski (2002) developed a Monte Carlo procedure to numerically obtain a corrected score under the assumption that $U \sim N(0, \sigma^2)$ and $\Psi(Y, X; \beta)$ is an entire function with respect to its second argument (Boas, 2011). By using the standard normal kernel in (5), we have the true score $\Psi(Y, X, \boldsymbol{\beta})$ as an entire function in X, which allows us to follow the Monte Carlo procedure to obtain an estimator of $\boldsymbol{\beta}$ via the following four-step algorithm.

- MC-1: For $b = 1, \ldots, B$, generate independent random errors, $\{U_{b,j}\}_{j=1}^{n}$, from $N(0, \sigma^2)$.
- MC-2: Form the complex-valued data, $\{\widetilde{W}_{b,j} = W_j + iU_{b,j}\}_{j=1}^{n}$, where i is the imaginary unit, for $b = 1, \ldots, B$.

- MC-3: Compute $\Psi_{\text{MC},B}(Y_j, W_j; \boldsymbol{\beta}) = B^{-1} \sum_{b=1}^{B} \text{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}$, where $\text{Re}(t)$ denotes the real part of a complex-valued t.
- MC-4: Solve the estimating equations for $\boldsymbol{\beta}$,

$$\sum_{j=1}^{n} \Psi_{\text{MC},B}(Y_j, W_j; \boldsymbol{\beta}) = \mathbf{0}. \tag{6}$$

Denote the resultant estimator as $\hat{\beta}_{\text{MC}}$.

By proving that $\text{E}[\text{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\} \mid (Y_j, X_j)] = \Psi(Y_j, X_j; \boldsymbol{\beta})$, Novick & Stefanski (2002) showed that $\text{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}$ is a corrected score that involves extra noise due to its dependence on $U_{b,j}$. A corrected score that is free of the extra noise is $\text{E}[\text{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\} \mid (Y_j, W_j)]$, which usually cannot be derived analytically. This motivates MC-3 above, where one computes the average of $\{\text{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}, b = 1, \dots, B\}$ as an approximation of the aforementioned expectation. Clearly, this empirical mean, $\Psi_{\text{MC},B}(Y_j, W_j; \boldsymbol{\beta})$, is also a corrected score, referred to as the Monte Carlo corrected score. Using the fact that $\hat{\boldsymbol{\beta}}_{\text{MC}}$ is an M-estimator that solves the estimating equations (6) constructed from an unbiased score function, Novick & Stefanski (2002, Section 5) established the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{\text{MC}}$. Finally, they demonstrated that, even when the assumption of U being normally distributed or the true score function being complete is violated, $\hat{\boldsymbol{\beta}}_{\text{MC}}$ is often less biased than the counterpart naive estimator.

3.3. Corrected Kernel Method

Even though the Monte Carlo corrected score method enjoys a certain degree of robustness to the normality assumption on U, an alternative method that is well justified for more general error distributions is desirable. This motivates us to correct the naive method from a different angle. Instead of correcting the naive score function, we propose to correct the naive objective function for measurement error. This is accomplished by constructing an unbiased estimator of the summand in (5), $K_h(Y - \beta_0 - \beta_1 X)$, based on (Y, W).

Since the objective function $Q_h(\beta)$ originates from a kernel density estimator, such unbiased estimators are readily available in Carroll & Hall (1988) and Stefanski & Carroll (1990), where the authors considered nonparametric density estimation in the presence of measurement error. Following their construction of a deconvoluting kernel, one can show that, conditioning on (Y, X), an unbiased estimator of $K_h(Y - \beta_0 - \beta_1 X)$ is $K_h^*(Y - \beta_0 - \beta_1 W)$, where $K_h^*(t) = K^*(t/h)/h$, and

$$K^*(t) = \frac{1}{2\pi} \int e^{-ist} \frac{\phi_K(s)}{\phi_U(-\beta_1 s/h)} ds, \tag{7}$$

in which $\phi_K(s)$ is the Fourier transform of $K(t)$, and where $\phi_U(s)$ is the characteristic function of U that does not vanish, both assumed to be even, and the integration is over the real line.

Besides being used for density estimation in the works of Carroll et al., (2006), Fan & Truong (1993) also used a deconvoluting kernel similar to (7) to construct a local constant estimator of $E(Y \mid X = x)$ in the presence of measurement error in X. Replacing the naive quantity, $K_h(Y - \beta_0 - \beta_1 W)$, with its unbiased estimtor defined above, $K_h^*(Y - \beta_0 - \beta_1 W)$, gives the corrected objective function to be maximized with respect to $\boldsymbol{\beta}$,

$$Q_h^*(\boldsymbol{\beta}) = \frac{1}{n} \sum_{j=1}^{n} K_h^*(Y_j - \beta_0 - \beta_1 W_j).$$

We call this method the corrected kernel method and denote the resultant estimator as $\hat{\boldsymbol{\beta}}_{\text{CK}}$. One existing work that also corrects an objective function for measurement error is Wang, Stefanski, & Zhu (2012) in the context of linear quantile regression. In this work, the authors derived a smooth function depending on (Y, W), of which the conditional expectation given (Y, X) approaches to the true objective function as the smoothing parameter involved in the smooth function shrinks to zero.

Stefanski & Carroll (1990) studied the validity of the construction of (7) and its properties for two types of measurement error distributions, namely ordinary smooth error distributions and super smooth error distributions (Fan, 1991). Their definitions are as follows.

Definition 1. *The distribution of U is ordinary smooth of order b if, as* $|t| \to \infty$, $d_0|t|^{-b} \le |\phi_U(t)| \le d_1|t|^{-b}$ *for some positive constants* d_0, d_1 *and b.*

Definition 2. *The distribution of U is super smooth of order b if, as* $|t| \to \infty$, $d_0|t|^{b_0} \exp(-|t|^b/d_2) \le |\phi_U(t)| \le d_1|t|^{b_1} \exp(-|t|^b/d_2)$ *for some positive constants* d_0, d_1, d_2, b, b_0 *and* b_1.

For example, Laplace distributions are ordinary smooth of order $b = 2$, and normal distributions are super smooth of order $b = 2$. We derive the asymptotic bias and variance of $\hat{\boldsymbol{\beta}}_{\text{CK}}$ under each type of measurement error distributions, and also establish its asymptotic normality. These findings are summarized in the following two theorems. Detailed proofs are provided in Appendices A and B in the Supplementary Material. Lemmas referenced in the theorems along with their proofs are given in Appendix C in the Supplementary Material.

Denote by $g(\epsilon \mid x)$ the density of the mode residual, $\epsilon = Y - \beta_0 - \beta_1 x$. To prove the theorems, conditions on $g(\epsilon \mid x)$ and the covariate are listed under Conditions G in Appendix A in the Supplementary Material. These assumptions are also imposed in Yao & Li (2014) and are indeed mild assumptions satisfied in a wide range of applications. Additional conditions concerning $K(t)$ and $\phi_U(t)$ that are required for proving the following two theorems are also provided in Appendix A in the Supplementary Material. Conditions on $K(t)$ are imposed mainly to guarantee integrability of functions of the forms $t^{\ell_1}\phi_K^2(t)$ and $t^{\ell_1}K^{(\ell_2)}(t)$ for some positive integers ℓ_1 and ℓ_2. Essentially, these conditions suggest that $\phi_K(t)$ and $K^{(\ell_2)}(t)$ tail off fast enough as $|t| \to \infty$, which can be easily satisfied by choosing an adequate kernel such as the one we use for the corrected kernel method in the simulation study reported in Section 5. Conditions imposed on $\phi_U(t)$ are also mainly about how fast $\phi_U^{(\ell)}(t)$ tails off as $|t| \to \infty$ for some nonnegative integer ℓ.

Theorem 1. *Under Conditions G and conditions in Lemma C, there exists a maximizer of* $Q_h^*(\boldsymbol{\beta})$*, denoted by* $\hat{\boldsymbol{\beta}}_{CK}$*, such that, as* $n \to \infty$ *and* $h \to 0$,

(i) when U follows an ordinary smooth distribution of order b, if $nh^{7+2b} \to 0$*, then*

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left(\sqrt{\frac{1}{nh^{3+2b}}}\right); \tag{8}$$

(ii) when U follows a super smooth distribution of order b, if $\exp(2|\beta_1|^b h^{-b}/d_2))/(nh^{b_6}) \to 0$*, where* $b_6 = \max\{3 - 2\min(b_2, b_3), 5 - 2\min(b_2, b_3, b_4), 7 - 2\min(b_2, b_3, b_4, b_5)\}$*, in which* b_ℓ*, for* $\ell = 2, 3, 4, 5$*, are defined in Lemma C, then*

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left\{\exp\left(\frac{|\beta_1|^b}{d_2 h^b}\right)\sqrt{\frac{1}{nh^{3-2\min(b_2,b_3)}}}\right\}. \tag{9}$$

The error rates presented in Theorem 1 combine the rate of bias, appearing in the big-O part of (8) and (9), and the rate of standard deviation, as in the big-O_p part of (8) and (9), of $\hat{\boldsymbol{\beta}}_{\text{CK}}$. Three observations are worth pointing out regarding these rates. First, the bias rate is not affected by measurement error, and coincides with the bias rate of Yao and Li's estimator in the absence of measurement error (Yao & Li 2014, Theorem 2.2). Second, compared to the variance rate of Yao and Li's estimator in the absence of measurement error (Yao & Li 2014, Theorem 2.2), the variance rates here are inflated due to measurement error. By setting $b = 0$, the variance rates suggested by (8) and (9) reduce to $O_p\{1/(nh^3)\}$, which is the variance rate of Yao and Li's estimator. Setting $b = 0$ is equivalent to setting $\sigma^2 = 0$, which leads to an error-free covariate. Third, comparing (8) and (9) reveals that the convergence rate of $\hat{\boldsymbol{\beta}}_{\text{CK}}$ in the presence of super smooth measurement error is much slower than that when U is ordinary smooth. This is in line with the findings in density estimation (Carroll & Hall, 1988; Stefanski & Carroll, 1990), local polynomial estimation in mean regression (Fan & Truong, 1993; Delaigle, Fan & Carrollm, 2009; Huang & Zhou, 2017) and nonparametric mode regression (Zhou & Huang, 2016) in the presence of different types of measurement error.

Moments of certain functions that involve Fourier transforms are derived in Appendix C to show Theorem 1. Results regarding these moments, along with strategies for deriving them, are also useful for establishing the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\text{CK}}$, although additional assumptions listed under Conditions N in Appendix A in the Supplementary Material are needed as well.

Theorem 2. *Under Conditions N and the same assumptions imposed in Theorem 1, for the maximizer of $Q_h^*(\boldsymbol{\beta})$, $\hat{\boldsymbol{\beta}}_{CK}$, that satisfies the properties in Theorem 1,*

(i) if U follows an ordinary smooth distribution of order b,

$$\sqrt{nh^{3+2b}}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2\mu_2 J^{*-1}Q/4\right) \xrightarrow{d} N(0,\, J^{*-1}K_L J^{*-1}), \quad \text{as } n \to \infty,$$

where K_L is a constant matrix, $Q = \lim_{n\to\infty} n^{-1}\sum_{j=1}^n E\{g^{(3)}(0|X_j)\tilde{X}_j\}$, and $J^* = \lim_{n\to\infty} n^{-1}\sum_{j=1}^n E\{g^{(2)}(0|X_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}\}$, in which $\tilde{X}_j = (1, X_j)^{\mathrm{T}}$;

(ii) if U follows a super smooth distribution of order b,

$$\left\{Var(\hat{\boldsymbol{\beta}}_{CK})\right\}^{-1/2}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - \frac{1}{4}h^2\mu_2 J^{*-1}Q\right) \xrightarrow{d} N(0,\, 1), \quad \text{as } n \to \infty,$$

where $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{CK}) = O[\exp\{2|\beta_1|^b/(d_2h^b)\}/\{nh^{3-2\min(b_2,b_3)}\}]$, and, for a generic positive definite matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of the positive definite square root of it.

4. BANDWIDTH SELECTION

Kernel-based methods are typically sensitive to the choice of bandwidths. To address the complication in bandwidth selection due to measurement error, Delaigle & Hall (2008) developed a strategy for smoothing parameter selection that combines simulation-extrapolation (SIMEX) (Cook & Stefanski, 1994; Stefanski & Cook, 1995) and cross validation. We apply this strategy to choose a bandwidth h following the algorithm described next, where we aim to choose an h that optimizes inference for $\boldsymbol{\beta}$ in some sense. Generically denote by $\hat{\boldsymbol{\beta}}_h$ an estimator of $\boldsymbol{\beta}$ under consideration with the bandwidth fixed at h based on observed data $\{(Y_j, W_j)\}_{j=1}^n$.

- SM-1: Generate M sets of further contaminated covariate data, $\{W_{m,j}^* = W_j + U_{m,j}^*\}_{j=1}^n$, for $m = 1, \ldots, M$, where $\{U_{m,j}^*, j = 1, \ldots, n\}_{m=1}^M$ are independent random errors generated from $f_U(u)$.

- SM-2: For $m = 1, \dots, M$, denote by $\hat{\boldsymbol{\beta}}^*_{h,m}$ the estimate of $\boldsymbol{\beta}$ based on data $\{(Y_j, W^*_{m,j})\}_{j=1}^n$ using the method under consideration. Find

$$h_1 = \underset{h>0}{\arg\min} \frac{1}{M} \sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}^*_{h,m} - \hat{\boldsymbol{\beta}}_h)^{\mathrm{T}} \boldsymbol{S}_{h,1}^{-1} (\hat{\boldsymbol{\beta}}^*_{h,m} - \hat{\boldsymbol{\beta}}_h),$$

  where $\boldsymbol{S}_{h,1}$ is the sample variance–covariance matrix of $\{\hat{\boldsymbol{\beta}}^*_{h,m} - \hat{\boldsymbol{\beta}}_h\}_{m=1}^M$.
- SM-3: Generate M sets of even further contaminated covariate data, $\{W^{**}_{m,j} = W^*_{m,j} + U^{**}_{m,j}\}_{j=1}^n$, for $m = 1, \dots, M$, where $\{U^{**}_{m,j}, j = 1, \dots, n\}_{m=1}^M$ are independent random errors generated from $f_U(u)$, which are also independent of $\{U^*_{m,j}, j = 1, \dots, n\}_{m=1}^M$.
- SM-4: For $m = 1, \dots, M$, denote by $\hat{\boldsymbol{\beta}}^{**}_{h,m}$ the estimate of $\boldsymbol{\beta}$ based on data $\{(Y_j, W^{**}_{m,j})\}_{j=1}^n$ using the method under consideration. Find

$$h_2 = \underset{h>0}{\arg\min} \frac{1}{M} \sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}^{**}_{h,m} - \hat{\boldsymbol{\beta}}^*_{h,m})^{\mathrm{T}} \boldsymbol{S}_{h,2}^{-1} (\hat{\boldsymbol{\beta}}^{**}_{h,m} - \hat{\boldsymbol{\beta}}^*_{h,m}),$$

  where $\boldsymbol{S}_{h,2}$ is the sample variance–covariance matrix of $\{\hat{\boldsymbol{\beta}}^{**}_{h,m} - \hat{\boldsymbol{\beta}}^*_{h,m}\}_{m=1}^M$.
- SM-5: Set the selected bandwidth as $h = h_1^2/h_2$.

The criterion we minimize in SM-2 and SM-4 is motivated by a theoretical optimal bandwidth given by $h_{\text{ideal}} = \arg\min_{h>0} E\{(\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{\Sigma}_h^{-1} (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})\}$, where $\boldsymbol{\Sigma}_h$ is the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_h$. The rationale behind this SIMEX procedure is that, as shown in Delaigle & Hall (2008), $\log(h_{\text{ideal}}) - \log(h_1) \approx \log(h_1) - \log(h_2)$ when σ^2 is small. And thus the value of h from SM-5 is a sensible approximation of h_{ideal}. Besides Delaigle & Hall (2008), Wang, Stefanski, & Zhu (2012) also used a similar strategy to select the smoothing parameter in their problem of linear quantile regression with covariate measurement error.

5. EMPIRICAL EVIDENCE

5.1. Simulation Design

To assess finite sample performance of the proposed estimators, we design comparative experiments where $\hat{\boldsymbol{\beta}}_{\text{NV}}$, $\hat{\boldsymbol{\beta}}_{\text{MC}}$ (with $B = 1000$) and $\hat{\boldsymbol{\beta}}_{\text{CK}}$ are obtained based on simulated error-prone data $\{(Y_j, W_j)\}_{j=1}^n$, as well as $\hat{\boldsymbol{\beta}}_{\text{YL}}$ based on the corresponding error-free data $\{(Y_j, X_j)\}_{j=1}^n$. The fourth estimator serves as a gold standard in the sense that estimators, naive or nonnaive, based on error-prone data are expected to be inferior in some regard than this estimator. Comparing the first three estimators with this reference estimator can shed light on how measurement errors compromise the naive estimator, and whether or not the two proposed nonnaive estimators improve over the naive estimator.

The kernel $K(t)$ used for obtaining $\hat{\boldsymbol{\beta}}_{\text{NV}}$, $\hat{\boldsymbol{\beta}}_{\text{MC}}$ and $\hat{\boldsymbol{\beta}}_{\text{YL}}$ is the standard normal density; and we use the kernel of which the Fourier transform is $\phi_K(t) = (1 - t^2)^3 I(-1 \le t \le 1)$ for $\hat{\boldsymbol{\beta}}_{\text{CK}}$. The choice of kernel for the corrected kernel method is in part dictated by the technical conditions on $\phi_K(t)$ that arise from deriving asymptotic properties of $\hat{\boldsymbol{\beta}}_{\text{CK}}$. To mitigate the effects of data-driven bandwidth selection on the proposed estimators, in the first part of simulation, we use an approximation of h_{ideal} given by $\hat{h}_{\text{ideal}} = \arg\min_{h>0} (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_h^{-1} (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\Sigma}}_h$ is a bootstrap estimate of $\boldsymbol{\Sigma}_h$ based on 100 bootstrap samples. Clearly, $\hat{h}_{\text{ideal}}$ cannot be computed in practice since $\boldsymbol{\beta}$ is unknown. In the second part of the simulation, we implement the SIMEX method described in Section 4, with $M = 10$, to select h for the proposed estimators. To preserve the integrity of $\hat{\boldsymbol{\beta}}_{\text{YL}}$, we run the Matlab code kindly provided by Professor Yao to compute $\hat{\boldsymbol{\beta}}_{\text{YL}}$ and $\hat{\boldsymbol{\beta}}_{\text{NV}}$, including their choice of bandwidth based on minimizing an estimate of the asymptotic mean squared error of Yao and Li's estimator of $\boldsymbol{\beta}$.

For ease of comparison, we follow the model setting in the simulation study presented in Yao & Li (2014) to generate error-free data. More specifically, for each of the two sample sizes, $n = 200$ and 400, the true covariate values $\{X_j\}_{j=1}^n$ are independent realizations from uniform(0, 1). Given X_j, the response is generated according to $Y_j = 1 + 3X_j + (1 + 2X_j)e_j$, for $j = 1, \ldots, n$, where $\{e_j\}_{j=1}^n$ are independent errors from $0.5N(-1, 2.5^2) + 0.5N(1, 0.5^2)$. For this error distribution, $e_M(x) \approx 1$ for all $x \in [0, 1]$, and thus $y_M(x) \approx 2 + 5x$. Ignoring rounding error, we have the true mode regression coefficients $\boldsymbol{\beta} = (2, 5)^{\text{T}}$. The error-contaminated covariate measurements $\{W_j\}_{j=1}^n$ are generated according to (2), with U following a Laplace distribution and a normal distribution, respectively, whose mean is zero and variance σ^2 is set at four levels to achieve reliability ratios $\lambda = 0.9, 0.85, 0.8, 0.75$.

5.2. Simulation Results

Under each of 16 model settings resulting from the combinations of n-$f_U(u)$-λ, 300 Monte Carlo replicate data sets of the form $\{(Y_j, X_j, W_j)\}_{j=1}^n$ are generated, producing 300 sets of estimates, $\{\hat{\boldsymbol{\beta}}_{\text{NV}}, \hat{\boldsymbol{\beta}}_{\text{MC}}, \hat{\boldsymbol{\beta}}_{\text{CK}}, \hat{\boldsymbol{\beta}}_{\text{YL}}\}$, among which $\hat{\boldsymbol{\beta}}_{\text{YL}}$ is not affected by the change in $f_U(u)$ or λ. Figure 1 presents the boxplots of these estimates when $n = 200$ for the case with Laplace measurement error when the approximated ideal bandwidth is used for $\hat{\boldsymbol{\beta}}_{\text{MC}}$ and $\hat{\boldsymbol{\beta}}_{\text{CK}}$. Figure 2 depicts the boxplots of the estimates when $n = 200$, U is normal, and the approximated ideal bandwidth is used for $\hat{\boldsymbol{\beta}}_{\text{MC}}$ and $\hat{\boldsymbol{\beta}}_{\text{CK}}$. We provide in Appendix D in the Supplementary Material figures of the counterpart boxplots when h is chosen by the SIMEX method for $\hat{\boldsymbol{\beta}}_{\text{MC}}$ and $\hat{\boldsymbol{\beta}}_{\text{CK}}$.

Overall, results for the two proposed methods that account for measurement error with bandwidths selected via the SIMEX method are very similar to those when the approximated ideal bandwidths are used. Except for higher variability, the two proposed estimates are comparable with the estimates obtained in the absence of measurement error, $\hat{\boldsymbol{\beta}}_{\text{YL}}$; and the naive estimate, $\hat{\boldsymbol{\beta}}_{\text{NV}} = (\hat{\beta}_{\text{NV},0}, \hat{\beta}_{\text{NV},1})^{\text{T}}$, is compromised by measurement error in contrast. Under the current model setting, $\hat{\beta}_{\text{NV},1}$ attenuates more towards null as error contamination in the covariate is more severe, that is, as λ decreases; furthermore $\hat{\beta}_{\text{NV},0}$ deviates more from the truth from above.

Between the two proposed estimators, $\hat{\boldsymbol{\beta}}_{\text{MC}}$ appears to be more variable than $\hat{\boldsymbol{\beta}}_{\text{CK}}$, especially in the presence of Laplace measurement error. This is expected because the Monte Carlo corrected score involves simulated pseudo measurement error. This source of variability can be more prominent when a small B is used to construct the Monte Carlo corrected score, $\Psi_{\text{MC},B}$. But increasing B after certain point, say, going beyond the current level (1,000) in the presented simulation experiments, becomes less profitable in terms of efficiency gain, especially considering the added computational burden with a much larger B. Another reason for the observed higher variability when U follows a Laplace distribution can be due to applying the Monte Carlo corrected score method when the normality assumption on U is violated.

Although the corrected kernel method has neither aforementioned concern, computing the deconvoluting kernel requires some care as the integral that defines $K^*(t)$ in (7) can be computationally challenging, especially in the presence of normal measurement error (Delaigle & Gijbels, 2007). We use the fast Fourier transforms (Bailey & Swarztrauber, 1994) to compute these integrals, which can still be problematic at times when U is normal. To alleviate numerical inaccuracy in the numerical integration, we follow the suggestion in Meister (2004) and replace the normal characteristic function with the Laplace characteristic function in (7) even when U actually follows a normal distribution. The presented numerical results associated with $\hat{\boldsymbol{\beta}}_{\text{CK}}$ in this section are obtained using this treatment. We observe in our extensive numerical study

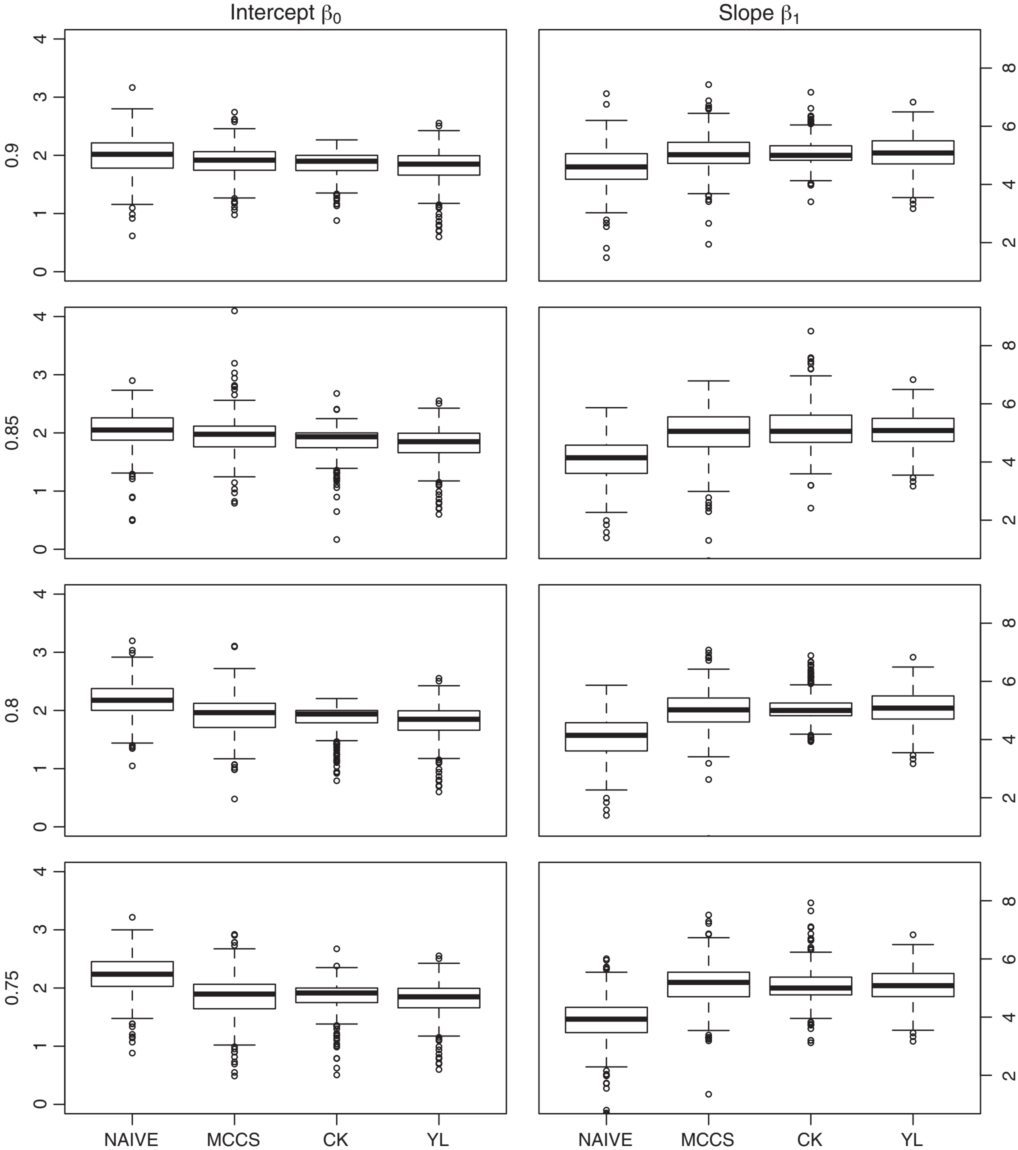

FIGURE 1: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is Laplace measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK) and Yao and Li's method (YL) in the absence of measurement error, respectively. The approximated theoretical optimal bandwidths are used for the Monte Carlo corrected score method and the corrected kernel method.

that, when the numerical integration using fast Fourier transforms goes through smoothly with $\phi_U(s)$ as the normal characteristic function, using a Laplace characteristic function instead does not cause noticeable changes in $\hat{\boldsymbol{\beta}}_{\text{CK}}$; and using the latter often leads to smoother numerical implementation. The robustness to and the benefit of Laplace measurement error assumption

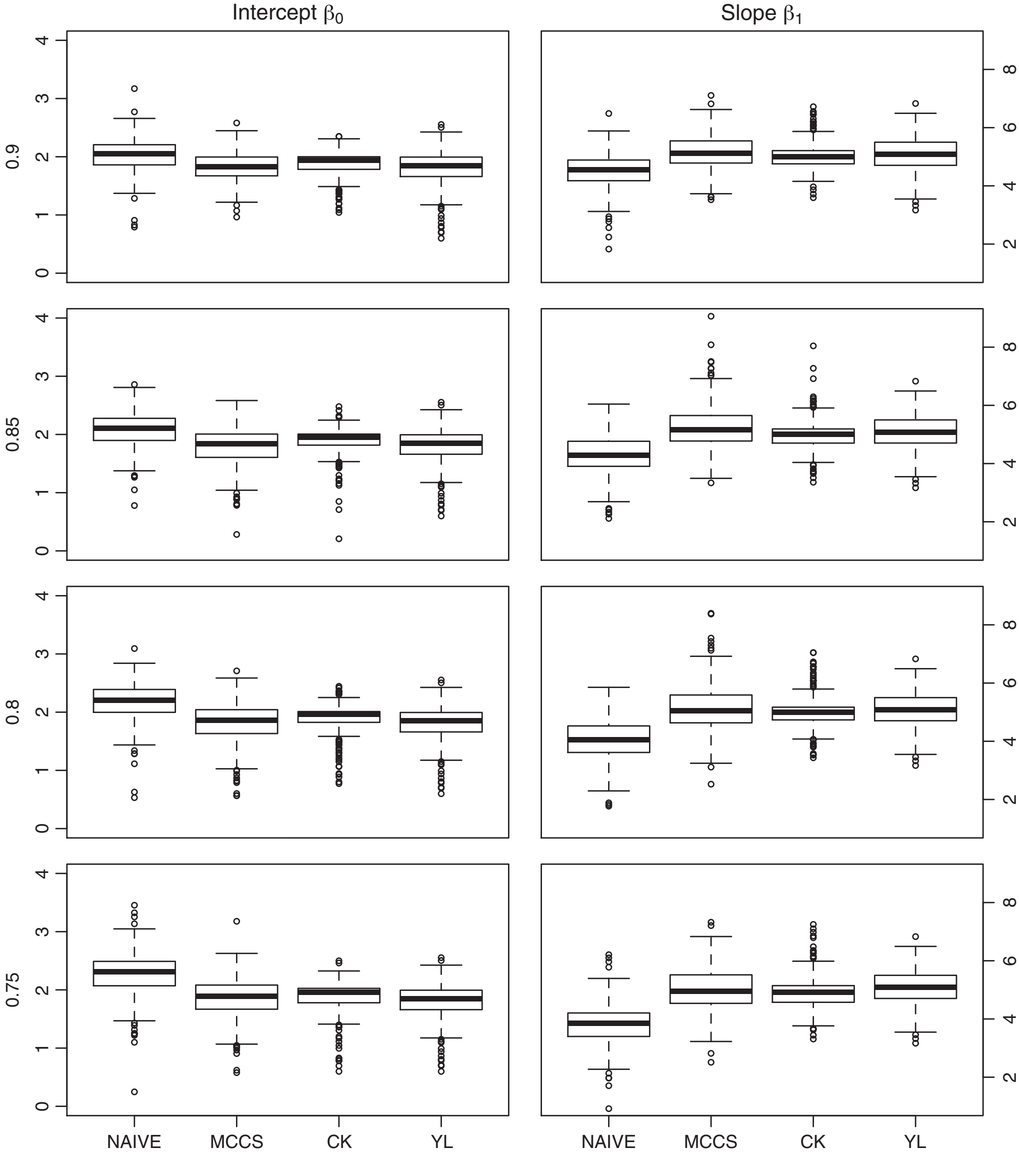

FIGURE 2: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is normal measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK) and Yao and Li's method (YL) in the absence of measurement error, respectively. The approximated theoretical optimal bandwidths are used for the Monte Carlo corrected score method and the corrected kernel method.

was noted and investigated by Meister (2004) and Delaigle (2008). For instance, Delaigle (2008) showed that, if the assumed error distribution and the true error distribution match in regard to the first two moments, the bias due to misspecifying the error distribution is of order $O(h^2) + o(\sigma^2)$ when a second-order kernel is used in a kernel density estimator.

TABLE 1: Monte Carlo averages of four sets of estimates over 300 Monte Carlo replicates when $\lambda = 0.75, 0.80$. Numbers in parentheses underneath the averages are empirical standard errors associated with the averages. The truth is $(\beta_0, \beta_1) = (2, 5)$.

	$\lambda = 0.75$				$\lambda = 0.80$			
	$n = 200$		$n = 400$		$n = 200$		$n = 400$	
Method	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
$U \sim$ Laplace$(0, \sigma^2)$								
Naive	2.26	3.82	2.29	3.85	2.18	4.04	2.22	4.05
	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
MCCS	1.96	4.77	1.84	5.08	1.86	5.06	1.80	5.18
	(0.04)	(0.08)	(0.03)	(0.07)	(0.03)	(0.07)	(0.02)	(0.05)
CK	1.72	5.14	1.72	5.19	1.74	5.14	1.73	5.18
	(0.03)	(0.05)	(0.02)	(0.04)	(0.02)	(0.05)	(0.02)	(0.03)
$U \sim N(0, \sigma^2)$								
Naive	2.21	3.88	2.27	3.92	2.17	4.07	2.18	4.16
	(0.02)	(0.05)	(0.01)	(0.03)	(0.02)	(0.05)	(0.01)	(0.03)
MCCS	2.02	4.62	1.92	4.88	1.75	5.18	1.78	5.17
	(0.04)	(0.09)	(0.02)	(0.04)	(0.02)	(0.05)	(0.03)	(0.06)
CK	1.77	5.02	1.79	4.97	1.80	4.99	1.84	4.99
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.04)
YL	1.83	5.08	1.87	5.05	1.83	5.08	1.87	5.05
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)

CK: corrected kernel method; MCCS: Monte Carlo corrected score method; YL: Yao and Li's method in the absence of measurement error.

Table 1 presents Monte Carlo averages of the four considered estimates across 300 replicates along with their empirical standard errors when $\lambda \in \{0.75, 0.8\}$. The same summary statistics for results obtained when $\lambda \in \{0.85, 0.9\}$ are tabulated in Appendix D in the Supplementary Material. Besides reinforcing the findings from Figures 1 and 2 that, compared to the naive estimator, the two proposed estimators are less compromised by measurement error and are closer to the benchmark estimator, these results also show that the performance of the proposed estimators improve in both accuracy and precision as the sample size increases. This is observed even for the Monte Carlo corrected score method in the presence of Laplace measurement error, a case this method is not designed for.

To this end, our focus has been on estimating $\boldsymbol{\beta}$. Because modes can be used to predict the outcome Y, we also compare predictions using estimated modes from the above three linear mode regression methods and the local linear mode estimation using the nonparametric method developed by Zhou & Huang (2016). Table 2 provides such comparison in terms of the empirical coverage probability of a prediction interval (band) of width $c\sigma_e$ centred around an estimated mode line (or curve) from a considered method across 300 Monte Carlo replicates, for $c = 0.1, 0.2, 0.5$. Here, σ_e is the standard deviation of e_j, which is around 2 in the simulation. According to Table 2, all four considered methods applying to error-prone data yield prediction

TABLE 2: Monte Carlo averages of proportions of observed responses captured by a prediction interval (band) of width $c\sigma_e$, for $c = 0.1, 0.2, 0.5$, associated with each method across 300 Monte Carlo replicates. Numbers in parentheses underneath the averages are 100×(empirical standard error) associated with the averages.

	$\lambda = 0.85$						$\lambda = 0.9$					
	$n = 200$			$n = 400$			$n = 200$			$n = 400$		
Method	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$
$U \sim$ Laplace$(0, \sigma^2)$												
Naive	0.09	0.18	0.39	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.07)	(0.11)	(0.15)	(0.07)	(0.10)	(0.15)	(0.07)	(0.10)	(0.14)	(0.06)	(0.08)	(0.11)
MCCS	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.06)	(0.08)	(0.11)	(0.06)	(0.08)	(0.09)	(0.06)	(0.08)	(0.10)	(0.05)	(0.07)	(0.10)
CK	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.19	0.40
	(0.06)	(0.09)	(0.11)	(0.06)	(0.08)	(0.10)	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)
NMR	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38
	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)	(0.06)	(0.09)	(0.013)	(0.05)	(0.08)	(0.11)
$U \sim N(0, \sigma^2)$												
Naive	0.09	0.18	0.39	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.07)	(0.11)	(0.17)	(0.06)	(0.10)	(0.15)	(0.06)	(0.09)	(0.12)	(0.06)	(0.09)	(0.12)
MCCS	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.06)	(0.08)	(0.11)	(0.05)	(0.08)	(0.10)	(0.06)	(0.08)	(0.09)	(0.05)	(0.07)	(0.09)
CK	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.10	0.19	0.40
	(0.06)	(0.09)	(0.12)	(0.05)	(0.07)	(0.10)	(0.06)	(0.09)	(0.11)	(0.05)	(0.07)	(0.09)
NMR	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38
	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)
YL	0.09	0.18	0.40	0.09	0.19	0.41	0.09	0.18	0.40	0.09	0.19	0.41
	(0.06)	(0.10)	(0.13)	(0.05)	(0.07)	(0.10)	(0.06)	(0.10)	(0.13)	(0.05)	(0.07)	(0.10)

CK: corrected kernel method; MCCS: Monte Carlo corrected score method; NMR: Zhou and Huang's nonparametric mode regression; YL: Yao and Li's method in the absence of measurement error.

intervals (bands) with similar empirical coverage probabilities as those from Yao and Li's linear mode regression method applying to error-free data. The observed similarity may not be surprising because prediction based on mean regression is also less affected by measurement error in covariates when compared to the amount that covariate effects estimation is affected (Buonaccorsi, 1995).

Instead of comparing prediction intervals centring around estimated modes, Table 3 presents a more close-up comparison of estimated modes themselves. In particular, Table 3 shows the Monte Carlo averages of the point-wise error associated with each method, |the estimated $y_M(x) - y_M(x)$|, at $x = 0.5, 0.9$. From this more close-up comparison, one can see that using error-prone data for mode estimation tends to produce more bias than when one uses error-free data; but our two proposed methods substantially alleviate the bias seen in the naive mode estimates. The nonparametric method shows no advantage when point-wise error of mode estimation is concerned, especially when the covariate value is near the boundary, for example, $x = 0.9$. We acknowledge that the current simulation setting is designed for linear mode regression, with data simulated from models with a linear mode function. Nonparametric mode regression makes no assumption on the functional form of the conditional mode function, and thus it is expected

TABLE 3: Monte Carlo averages of point-wise errors, |the estimated $y_M(x) - y_M(x)$|, associated with each method when $x = 0.5, 0.9$ across 300 Monte Carlo replicates. Numbers in parentheses are empirical standard error associated with the averages.

	$\lambda = 0.85$				$\lambda = 0.9$			
	$n = 200$		$n = 400$		$n = 200$		$n = 400$	
Method	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$
$U \sim$ Laplace$(0, \sigma^2)$								
Naive	0.28 (0.01)	0.53 (0.02)	0.25 (0.01)	0.51 (0.02)	0.27 (0.01)	0.47 (0.02)	0.20 (0.01)	0.38 (0.01)
MCCS	0.22 (0.01)	0.45 (0.03)	0.16 (0.02)	0.33 (0.03)	0.18 (0.01)	0.35 (0.03)	0.15 (0.01)	0.30 (0.02)
CK	0.20 (0.01)	0.35 (0.02)	0.17 (0.01)	0.26 (0.01)	0.21 (0.01)	0.32 (0.01)	0.18 (0.01)	0.28 (0.01)
NMR	0.66 (0.01)	1.10 (1.12)	0.65 (0.01)	1.24 (0.12)	0.68 (0.02)	1.31 (0.12)	0.66 (0.01)	0.58 (0.15)
$U \sim N(0, \sigma^2)$								
Naive	0.28 (0.01)	0.59 (0.02)	0.25 (0.01)	0.54 (0.02)	0.25 (0.01)	0.46 (0.02)	0.21 (0.01)	0.39 (0.02)
MCCS	0.18 (0.01)	0.37 (0.02)	0.14 (0.01)	0.24 (0.01)	0.18 (0.01)	0.33 (0.02)	0.12 (0.01)	0.23 (0.01)
CK	0.21 (0.01)	0.33 (0.02)	0.21 (0.01)	0.26 (0.01)	0.20 (0.01)	0.40 (0.02)	0.18 (0.01)	0.25 (0.01)
NMR	0.64 (0.01)	0.95 (0.09)	0.34 (0.02)	0.52 (0.08)	0.63 (0.02)	1.28 (0.14)	0.68 (0.01)	1.05 (0.11)
YL	0.14 (0.01)	0.21 (0.01)	0.14 (0.01)	0.22 (0.01)	0.14 (0.01)	0.21 (0.01)	0.14 (0.01)	0.22 (0.01)

CK: corrected kernel method; MCCS: Monte Carlo corrected score method; NMR: Zhou and Huang's nonparametric mode regression; YL: Yao and Li's method in the absence of measurement error.

to exhibit higher variability and less accuracy in estimating the mode than methods that take into account a simple (and true) functional form. Scenarios where the data generating process involves a nonlinear mode function are where one can benefit from employing the nonparametric method, which are scenarios beyond the scope of the current article.

6. APPLICATION TO DIETARY DATA

In this section, we apply the proposed methods to a dietary data set from the Women's Interview Survey of Health. The data are from $n = 271$ subjects, each completing a food frequency questionnaire (FFQ) and six 24-h food recalls on randomly selected days. We focus on studying the impact of the long-term usual intake (X) on the FFQ intake measured as the percent calories from fat (Y) (Carroll, Freedman & Pee, 1997). Since the long-term intake cannot be measured directly, and the 24-h recalls can be viewed as error-contaminated surrogates of it, we used the average of these recalls from each subject as a surrogate (W) of this subject's long-term intake. Figure 3 provides the histogram of FFQ intake and the scatter plot of it versus the 24-h food recalls. The histogram indicates an underlying skewed distribution, and the scatter plot suggests existence of outliers in the observed data. These two features suggest that mode regression can provide valuable information regarding the association between a response and a covariate that mean regression may not capture.

For illustration purposes, we consider a linear mode regression model for the mode of Y_j given X_j, where X_j is not observed but its error-contaminated surrogate W_j is, where $W_j = \sum_{k=1}^{6} W_{j,k}/6$, in which $W_{j,k}$ is subject j's kth food recall, for $k = 1, \ldots, 6$ and $j = 1, \ldots, 271$. Using the six replicate measures for each underlying X_j, we estimate the variance of measurement errors associated with W_j by way of one-sixth of $\sum_{j=1}^{n} \sum_{k=1}^{6} (W_{j,k} - W_j)^2/(5n)$, following equation (4.3) in Carroll et al., (2006). This gives an estimate of the measurement error variance as $\hat{\sigma}^2 = 0.12$, and the corresponding estimated reliability ratio being 0.73.

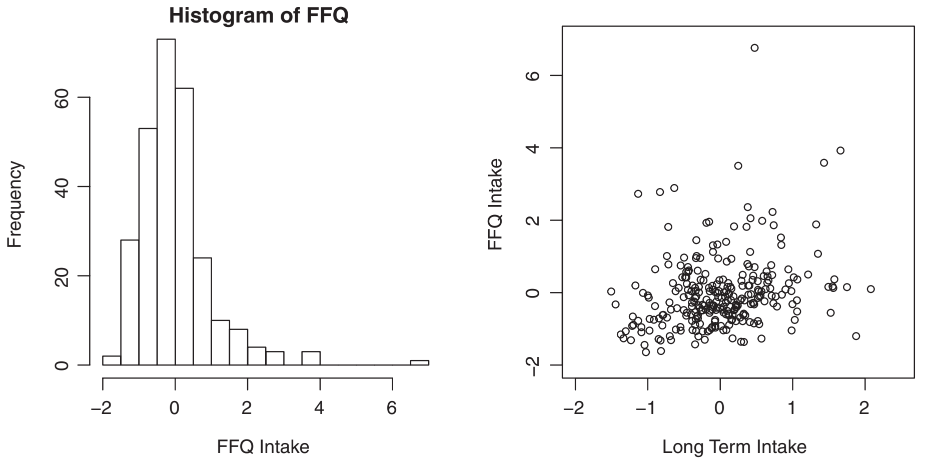

FIGURE 3: The histogram (on the left panel) of food frequency questionnaire intake and the scatter plot (on the right panel) of this quantity versus a surrogate of long-term intake for the dietary data.

We carry out the linear mode regression analysis using the naive method, the Monte Carlo corrected score method, the corrected kernel method assuming Laplace and normal measurement error, respectively, and we also implement the local linear mode estimation as the only fully nonparametric method. Table 4 presents the estimated regression coefficients from three linear mode regression methods. These results suggest that both proposed methods produce estimates of the covariate effect, β_1, that imply a stronger association between the FFQ intake and the long-term intake than the estimate from the naive method does. In particular, compared to the naive estimate, the estimated covariate effect from the Monte Carlo corrected score method increases by 29%, and the estimates from the corrected kernel method increase by 38% and 34% when assuming Laplace measurement error and normal measurement error, respectively.

This also gives an example where using the Laplace characteristic function and the normal characteristic function in the corrected kernel method yields very similar estimates. Figure 4 depicts three of these estimated mode regression lines, omitting the one from the corrected kernel method under the normality assumption and the estimated mode curve obtained by applying the local linear estimation in Zhou & Huang (2016). Computer codes for implementing the two proposed method for this data set are provided in Appendices E and F in the Supplementary Material. This code and the dietary data are also available as online Supplementary Material for interested readers to download.

7. DISCUSSION

7.1. Recap and Extensions

In this article, we propose two methods to infer regression coefficients in a linear mode model for a response given an error-prone covariate. The resultant inference for the covariate effect significantly improve over the naive inference from applying Yao and Li's method without accounting for measurement error. As demonstrated in the real data analysis in Section 6, estimating the measurement error variance is trivial when replicate measures of each underlying true covariate value are available. This treatment of unknown σ^2 has been a routine practice in the measurement error literature, where researchers typically observe little impact of estimating σ^2 on the final inference for covariate effects. The measurement error variance is the only piece of

TABLE 4: Regression coefficient estimates in the linear mode regression model from the naive method, the Monte Carlo corrected score method, and the corrected kernel method (assuming Laplace and normal U, respectively) using the dietary data. Numbers in parentheses are estimated standard deviations of the regression coefficient estimates resulting from 200 bootstrap samples.

Method	β_0	β_1
Naive	−0.27 (0.10)	0.36 (0.11)
MCCS	−0.10 (0.05)	0.48 (0.13)
CK-Laplace	−0.07 (0.05)	0.50 (0.12)
CK-Normal	−0.09 (0.06)	0.49 (0.13)

CK-Laplace: corrected kernel method assuming Laplace U; CK-Normal: corrected kernel method assuming normal U; MCCS: Monte Carlo corrected score method.

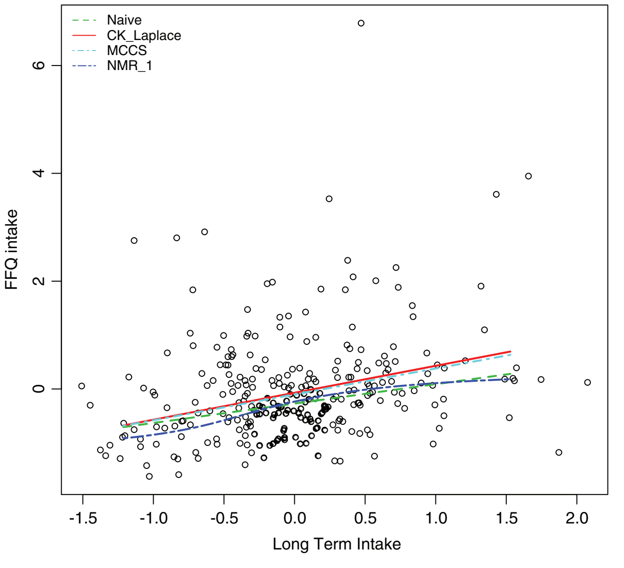

FIGURE 4: Dietary data overlaid with the estimated mode regression line from naively applying Yao and Li's method (green dashed line), the Monte Carlo corrected score method (cyan dot-dashed line), the corrected kernel method assuming Laplace measurement error (red solid line) and a local linear estimate of the mode curve (blue two-dashed line).

information regarding $f_U(u)$ required for implementing the Monte Carlo corrected score method since normal U is assumed for this method. To implement the corrected kernel method, the characteristic function of U, $\phi_U(t)$, is needed, which can also be easily estimated using replicate measures Delaigle, Hall & Meister (2008). Moreover, as noted in our simulation study and by several other authors (Meister, 2004; Delaigle, 2008; Delaigle, Fan & Carrollm 2009; Zhou & Huang, 2016), simply setting $\phi_U(t)$ as the Laplace characteristic function works well in most scenarios, which frees one from estimating the characteristic function altogether.

Both proposed methods can easily incorporate multiple covariates in the linear mode model. Indeed, Yao and Li's method is developed more generally with multivariate covariates, and the Monte Carlo corrected score method entails evaluating the score function used in Yao and

Li's method at simulated contaminated covariate data, hence one only needs to revise MC-1 in the algorithm in Section 3.2 accordingly to implement the Monte Carlo corrected score method with multivariate covariates. To implement the corrected kernel method when there are $p > 1$ covariates, some or all of which are prone to nondifferntial measurement error, one uses a multivariate characteristic function of $\boldsymbol{U} = (U_1, \dots, U_p)^{\mathrm{T}}$ in (7) evaluated at $-\boldsymbol{\beta}_1^{\mathrm{T}} s/h$, bearing in mind that $\phi_{U_\ell}(t) = 1$ if the ℓth covariate is error-free, for $\ell \in \{1, 2, \dots, p\}$.

7.2. Limitations and Future Research

When the normality assumption on U is violated in the Monte Carlo corrected score method, the bias of $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ can be substantial when the true measurement error distribution is skewed. The bias reduction achieved by $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$ when the Laplace characteristic function is used in the deconvoluting kernel can also be less impressive in this case. In addition, the proposed estimates may exhibit erratic behaviour when λ is small, regardless of the sample size. Besides the measurement error distribution, the bandwidth chosen by the computationally intensive SIMEX method may also contribute to such erratic behaviour. Further research is needed to develop a more effective bandwidth selector that is computationally efficient.

The mode residual distribution, $g(\epsilon \mid x)$, is left unspecified in our article, and thus the proposed methods are broadly applicable even when one lacks a parametric model for $f_{Y|X}(y \mid x)$. However, we impose a linear functional form for the mode of Y given $X = x$, making these methods semiparametric in nature. One follow-up research direction is to incorporate semiparametric components into the specification of $y_M(x)$. Yao & Xiang (2016) proposed a local polynomial mode estimation method and also considered a nonparametric varying coefficient mode regression model. Zhao et al., (2014) proposed a variable selection method based on a partially linear varying coefficient mode regression model. These works, and other existing works on semiparametric mode regression, all assume error-free covariates. Upon completion of the project reported in this article, we have started to explore partially linear mode regression in the presence of covariate measurement error.

SUPPLEMENTARY MATERIAL

Supplementary material available online includes proofs of Theorems 1 and 2, lemmas referenced in the theorems along with their proof, technical conditions stated in the theorems, additional simulation results mentioned in Section 5.2, and computer code for analyzing the dietary data using the two proposed methods, and the dietary data.

BIBLIOGRAPHY

Bailey, D. H. & Swarztrauber, P. N. (1994). A fast method for the numerical evaluation of continuous Fourier and Laplace transforms. *SIAM Journal on Scientific Computing*, 15, 1105–1110.

Bamford, S. P., Rojas, A. L., Nichol, R. C., Miller, C. J., Wasserman, L., Genovese, C. R., & Freeman, P. E. (2008). Revealing components of the galaxy population through non-parametric techniques. *Monthly Notices of the Royal Astronomical Society*, 391, 607–616.

Boas, R. P. (2011). *Entire Functions*. Academic Press, New York.

Buonaccorsi, J. P. (1995). Prediction in the presence of measurement error: General discussion and an example predicting defoliation. *Biometrics*, 51, 1562–1569.

Carroll, R., Ruppert, D., Stefanski, L., & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall/CRC, Boca Raton.

Carroll, R. J., Freedman, L., & Pee, D. (1997). Design aspects of calibration studies in nutrition, with analysis of missing data in linear measurement error models. *Biometrics*, 53, 1440–1457.

Carroll, R. J. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83, 1184–1186.

Chen, Y. -C., Genovese, C. R., Tibshirani, R. J., & Wasserman, L. (2016). Nonparametric modal regression. *The Annals of Statistics*, 44, 489–514.

Cook, J. R. & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association*, 89, 1314–1328.

Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statistica Sinica*, 18, 1025–1045.

Delaigle, A., Fan, J., & Carroll, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104, 348–359.

Delaigle, A. & Gijbels, I. (2007). Frequent problems in calculating integrals and optimizing objective functions: A case study in density deconvolution. *Statistics and Computing*, 17, 349–355.

Delaigle, A. & Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *Journal of the American Statistical Association*, 103, 280–287.

Delaigle, A., Hall, P., & Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, 36, 665–685.

Einbeck, J. & Tutz, G. (2006). Modelling beyond regression functions: An application of multimodal regression to speed-flow data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55, 461–475.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19, 1257–1272.

Fan, J. & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, 21, 1900–1925.

Fuller, W. A. (2009). *Measurement Error Models*. John Wiley & Sons, New York.

Grund, B. & Hall, P. (1995). On the minimisation of L^p error in mode estimation. *The Annals of Statistics*, 23, 2264–2284.

He, X. & Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, 10, 129–140.

Hedges, S. B. & Shah, P. (2003). Comparison of mode estimation methods and application in molecular clock analysis. *BMC Bioinformatics*, 4, 31.

Huang, M. & Yao, W. (2012). Mixture of regression models with varying mixing proportions: A semiparametric approach. *Journal of the American Statistical Association*, 107, 711–724.

Huang, X. & Zhou, H. (2017). An alternative local polynomial estimator for the error-in-variables problem. *Journal of Nonparametric Statistics*, 29, 301–325.

Hyndman, R. J., Bashtannyk, D. M., & Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5, 315–336.

Kemp, G. C. & Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, 170, 92–101.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Lee, M. -J. (1989). Mode regression. *Journal of Econometrics*, 42, 337–349.

Lee, M. -J. (1993). Quadratic mode regression. *Journal of Econometrics*, 57, 1–19.

Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32, 439–449.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77, 127–137.

Novick, S. J. & Stefanski, L. A. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97, 472–481.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33, 1065–1076.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, Boca Raton.

Song, X. & Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, 61, 702–714.

Stefanski, L. A. & Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics: A Journal of Theoretical and Applied Statistics*, 21, 169–184.

Stefanski, L. A. & Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, 90, 1247–1256.

Wang, C. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, 16, 235–253.

Wang, H. J., Stefanski, L. A., & Zhu, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*, 99, 405.

Wei, Y. & Carroll, R. J. (2009). Quantile regression with measurement error. *Journal of the American Statistical Association*, 104, 1129–1143.

Yao, W. & Li, L. (2014). A new regression model: Modal linear regression. *Scandinavian Journal of Statistics*, 41, 656–671.

Yao, W. & Xiang, S. (2016). Nonparametric and varying coefficient modal regression. arXiv: 1602.06609.

Yi, G. Y. (2017). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. Springer, New York.

Zhao, W., Zhang, R., Liu, J., & Lv, Y. (2014). Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Annals of the Institute of Statistical Mathematics*, 66, 165–191.

Zhou, H. & Huang, X. (2016). Nonparametric modal regression in the presence of measurement error. *Electronic Journal of Statistics*, 10, 3579–3620.

Zucker, D. M. & Spiegelman, D. (2008). Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in Medicine*, 27, 1911–1933.

Received 01 August 2017
Accepted 27 October 2018