Electronic Journal of Statistics
Vol. 5 (2011) 1424–1449
ISSN: 1935-7524
DOI: 10.1214/11-EJS647

Estimation via corrected scores in general semiparametric regression models with error-prone covariates

Arnab Maity*

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.
e-mail: amaity@ncsu.edu

and

Tatiyana V. Apanasovich†

Department of Biostatistics, Thomas Jefferson University, Philadelphia, Pennsylvania 19118, U.S.A.
e-mail: tatiyana.apanasovich@jefferson.edu

Abstract: This paper considers the problem of estimation in a general semiparametric regression model when error-prone covariates are modeled parametrically while covariates measured without error are modeled non-parametrically. To account for the effects of measurement error, we apply a correction to a criterion function. The specific form of the correction proposed allows Monte Carlo simulations in problems for which the direct calculation of a corrected criterion is difficult. Therefore, in contrast to methods that require solving integral equations of possibly multiple dimensions, as in the case of multiple error-prone covariates, we propose methodology which offers a simple implementation. The resulting methods are functional, they make no assumptions about the distribution of the mismeasured covariates. We utilize profile kernel and backfitting estimation methods and derive the asymptotic distribution of the resulting estimators. Through numerical studies we demonstrate the applicability of proposed methods to Poisson, logistic and multivariate Gaussian partially linear models. We show that the performance of our methods is similar to a computationally demanding alternative. Finally, we demonstrate the practical value of our methods when applied to Nevada Test Site (NTS) Thyroid Disease Study data.

AMS 2000 subject classifications: Primary 62G08; secondary 62G20.
Keywords and phrases: Generalized estimating equations, generalized linear mixed models, kernel method, measurement error, Monte Carlo Corrected Score, semiparametric regression.

Received June 2011.

*Supported in part by NIEHS grant R00ES017744. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health.

†Supported in part by NSF grant #0707106.

Contents

1. Introduction

Regression models with measurement errors arise frequently in practice and have attracted much attention in the statistical literature. Semiparametric regression models with errors in covariates have been considered by several authors in the attempt to develop measurement error calibration techniques when the errors are in the linear part of linear regression ([9]) or generalized linear regression ([10]) models. [30] used a method of moments and deconvolution to construct the calibration for the case of partially linear models when the mismeasured covariate appears in parametric and nonparametric parts. However all of the above methodologies take advantage of the fact that unknown parameters in a parametric part enter the model through a linear combination with error prone covariates. We consider a general semiparametric regression problem where parameters can enter the model through any known function of covariates. Recently, for a general semiparametric regression problem we proposed to utilize a popular alternative to regression-calibration, a simulation-extrapolation method (SIMEX, Apanasovich et al. [2]). We considered the situations where the mismeasured variable is modeled purely parametrically, purely nonparametrically, or when the mismeasured variable has components that are modeled both parametrically and nonparametrically. Even though SIMEX is a general-purpose, widely applicable method for correcting parameter estimates for the biases induced by measurement error in covariates, it suffers from relying on a rather heuristic extrapolation step ([4]).

Ma and Carroll ([15]), building upon work of [29], developed a functional methodology for a general semiparametric measurement error regression framework. They require a specification of a parametric distribution for error-prone covariates given all other covariates. Nevertheless the method is general, the estimators are still consistent and asymptotically normally distributed even when that distribution is misspecified. However, the implementation of Ma and Carroll's method ([15]) requires solving integral equations, which can be quite computationally expensive. More importantly, this procedure is computationally infeasible when the error prone covariates are multivariate, e.g., in repeated measures or longitudinal data settings. Moreover, the methodology is efficient only when the posited parametric distribution for error-prone covariates given all other covariates is correctly specified, which is not a practical assumption for many applications.

In this paper, we develop an alternative functional methodology where almost no assumptions are made about the distribution of an error prone covariate. We consider a classical additive measurement error model, where covariate X is unobserved and W is observed instead, such that $W = X + U$. The measurement error U is independent of any other observed variables and has a normal distribution with zero mean and variance $\Sigma_{u,0}$. The normal distribution is often used in the literature and is not too restrictive. There are many other ways to model the relationship between X and W, which give additivity in some functional scale, e.g. logarithmic. See [6] for more details on transformations of X.

Our method is based upon the idea of Monte Carlo corrected scores (Novick and Stefanski [18]), where Monte Carlo simulations help to determine the corrected score for a large class of models. To our knowledge, this is the first attempt to introduce these ideas into semiparametric framework. While our method uses complex-value arithmetic, it is relatively easy to implement in many standard software packages. Specifically, its theory falls into the framework of standard estimation methods for criterion functions in semiparametric problems (Lin and Carroll [12]), thus we are taking advantage of some well established results. Hence, the core of our method's implementation is the standard semiparametric estimation technique adjusted for complex valued covariates and performed with a relatively small number of Monte Carlo runs.

Examples of widely used regression models for which our methods apply include: linear model ([9]), Poisson regression model ([3, 18]), Gamma regression model ([21, 1, 17]). In our paper we devote attention to each of these models except Gamma. We use univariate and multivariate partially linear models to illustrate the relationship between Monte Carlo corrected score method and other methods which exist in the literature. Specifically, in the univariate case, we show that as the number of Monte Carlo iterations goes to infinity, the estimators converge to that of [9]; and in the case of multivariate partially linear models, to that of [12]. Moreover, when X is Gaussian, our univariate estimator is efficient (Ma and Carroll [15]). Further, we demonstrate through simulations that our multivariate estimator performs similar to the efficient one with the reasonable number of Monte Carlo iterations even though its limit is not an efficient estimator ([12]). Poisson, logistic and multivariate Gaussian regression models are studied via simulations to demonstrate the ease of implementation and generality of proposed methods.

Note that logistic regression model is an exception to our method and we only present heuristic arguments for this model. The problem lies in the fact that the logistic distribution function is not an entire function, which is an essential theoretical condition for our method to be applicable. However, Novick and Stefanski ([18]) noted that for measurement error variance of the magnitudes commonly encountered in applications one can still apply corrected score based methods, with only a minor bias (see p. 479 of Novick and Stefanski [18]). Results from numerical studies show that our method also performs reasonably well for the logistic case when the measurement error is moderate.

An outline of this paper is as follows. In Section 2, we review the estimation in a general semiparametric regression where there is no measurement error as studied in Lin and Carroll ([12]); and corrected score method proposed by [17] and Novick and Stefanski ([18]). Then we introduce Monte Carlo corrected score method to the estimation in a general semiparametric regression with error prone covariates and study the asymptotic behavior of the proposed estimators. Among other results we offer asymptotic standard errors accounting for the uncertainty due to estimation of measurement error covariance matrix.

In Section 3, we focus on special cases of univariate and multivariate partially linear models. We demonstrate that as the size of Monte Carlo correction sample increases, our estimators converge to the ones mentioned in the literature ([9, 12]). Moreover, we show that when the error prone covariate is Gaussian, our estimator is efficient in the univariate case and performs similar to the efficient estimator in the multivariate case.

In Section 4, we present a simulation study using several semiparametric models to illustrate the performance of our method. We start with partially linear Poisson model and show that our method produces only a small bias and appropriate coverage probability. We also use the simulation scenario of Ma and Carroll ([15]) applied to the logistic partially linear model with a quadratic effect of X. In this case we note that our method, despite lacking a theoretical foundation for the logistic regression, produces even slightly smaller mean squared errors than theirs, while being computationally far less challenging. Moreover, when we triple the variance of the measurement error used in the scenario, we still get relatively small errors in the estimators. Last, we report results of simulations in the case of multivariate partially linear model with multivariate measurement errors. Such a model would be computationally challenging for the competing methodology (Ma and Carroll [15]), while our methods offer ease of implementation and satisfactory performance.

In Section 5.1, we apply our method to Nevada Test Site (NTS) Thyroid Disease Study data and report the results. We also present a simulation study where the measurement errors in the covariates are comparable to those in the real NTS data example. We show that for the assumed amount of uncertainty in radiation exposure, our method performs reasonably well.

Finally, Section 6 gives a few brief concluding remarks. All technical details are collected in an appendix.

2. Methodology

We describe a general semiparametric regression when there is no measurement error as studied in Lin and Carroll ([12]) first. Assume that data (Y_i, X_i, Z_i), $i = 1, \ldots, n$ are independent replications of a $(p_y + p_x + 1)$-dimensional random vector (Y, X, Z). Let $\mathcal{B}$ denote the parameter of interest and $\theta(\cdot)$ be an infinitely dimensional nuisance parameter with true values of $\mathcal{B}_0$ and $\theta_0(\cdot)$ respectively. Let $\mathcal{L}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}$ be a criterion function in the sense that $\mathrm{E}[\mathcal{L}_{\mathcal{B}}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}|Z] = 0$ and $\mathrm{E}[\mathcal{L}_{\theta}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}|Z] = 0$, where here and in what follows, we use subscripts $\mathcal{B}$ and θ to denote the partial derivatives with respect to $\mathcal{B}$ and θ respectively. Suppose $K(\cdot)$ is a symmetric density function with variance 1 and define $K_h(v) = h^{-1}K(v/h)$, where h is the bandwidth. Let $G_i(z) = \{1, (Z_i - z)/h\}^{\mathrm{T}}$. Given a fixed value of $\mathcal{B} = \mathcal{B}^*$, the modified kernel estimate of $\widehat{\theta}(z, \mathcal{B}^*)$ is a solution of the local-linear estimating equations

$$n^{-1}\sum_{i=1}^{n} K_h(Z_i - z)G_i(z)\mathcal{L}_{\theta}\left\{Y_i, X_i, \mathcal{B}^*, (\alpha_0, \alpha_1)G_i(z)\right\} = 0 \tag{2.1}$$

for α_0, calling it $\widehat{\alpha}_0$. To estimate $\mathcal{B}$, Lin and Carroll ([12]) proposed profile and backfitting methods. The profile kernel estimator for $\mathcal{B}$ maximizes $\sum_{i=1}^{n} \mathcal{L}\{Y_i, X_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\}$ in $\mathcal{B}$ which is equivalent to solving the score equation

$$n^{-1}\sum_{i=1}^{n}\left[\mathcal{L}_{\mathcal{B}}\left\{Y_i, X_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\right\} + \mathcal{L}_{\theta}\left\{Y_i, X_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\right\}\widehat{\theta}_{\mathcal{B}}(Z_i, \mathcal{B})\right] = 0. \tag{2.2}$$

Maximization of the profile likelihood requires calculating $\widehat{\theta}_{\mathcal{B}}(Z_i, \mathcal{B}) = \partial\widehat{\theta}(Z_i, \mathcal{B})/\partial\mathcal{B}$, which can be computed by numerical differentiation. In some cases where the profile kernel methods may be difficult to implement numerically, a backfitting algorithm can be used instead. Suppose that the current estimate is $\mathcal{B}^*$, the updated backfitting estimate then maximizes $\mathcal{B}$ in the function $\sum_{i=1}^{n} \mathcal{L}\{Y_i, X_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B}^*)\}$ or is the solution of

$$n^{-1}\sum_{i=1}^{n}\mathcal{L}_{\mathcal{B}}\left\{Y_i, X_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B}^*)\right\} = 0.$$

The second step in backfitting iterations is to find a solution of the local linear estimating equations (2.1) using updated estimate of $\mathcal{B}$. Lin and Carroll ([12]) showed that profiling and backfitting are asymptotically equivalent, however to obtain $\sqrt{n}$-consistent estimator of $\mathcal{B}$, unlike profiling, undersmoothing of the nonparametric function is required by the backfitting method.

In this paper, we consider the case where covariate X is unobserved and instead W is observed such that (possibly after transformation)

$$W = X + U,$$

where U is independent of any other observed variables and follows a Normal distribution with mean 0 and covariance matrix $\Sigma_{u,0}$. Measurement error induces bias in estimating equations (2.1) and (2.2), which results in biased estimators of the parameters. Thus, the purpose of the current study is to modify (2.1) and (2.2) and obtain unbiased estimating equations corrected for measurement error, as we describe next.

2.1. Monte-Carlo corrected score estimation

A function $\widetilde{\mathcal{L}}\{Y, W, \mathcal{B}_0, \theta_0(Z)\}$ is a criterion function if

$$\mathrm{E}[\widetilde{\mathcal{L}}\{Y, W, \mathcal{B}, \theta(Z)\}|Y, X, Z] = \mathcal{L}\{Y, X, \mathcal{B}, \theta(Z)\}$$

for any $\mathcal{B}$, $\theta(Z)$ from the parameter space ([17]). Novick and Stefanski ([18]), considering parametric models, proposed a general method to construct corrected score functions based on the results from complex analysis. Let $\widetilde{W}$ be a complex random vector

$$\widetilde{W} = W + \iota V,$$

where $\iota = \sqrt{-1}$ and V is a normal random vector with mean 0 and covariance matrix $\Sigma_{u,0}$. [26] showed that if $f(\cdot)$ is an entire function then, under integrability conditions

$$\mathrm{E}\{f(\widetilde{W})|X\} = \mathrm{E}[\mathrm{Re}\{f(\widetilde{W})\}|X] = f(X),$$

where $\mathrm{Re}(\cdot)$ denotes the real part of its argument.

Assume that $\mathcal{L}(\cdot)$ is an entire function of its second argument. We define the corrected criterion function as

$$\widetilde{\mathcal{L}}\{Y, W, \mathcal{B}_0, \theta_0(Z)\} = \mathrm{E}(\mathrm{Re}[\mathcal{L}\{Y, \widetilde{W}, \mathcal{B}_0, \theta_0(Z)\}]|Y, W, Z).$$

However, the required conditional expectation is not always easy to obtain analytically. Novick and Stefanski ([18]) proposed to use Monte Carlo integration to approximate the conditional expectation in the parametric models they considered. We introduce Monte Carlo correction into our semiparametric models so that Monte Carlo corrected criterion function becomes

$$\mathcal{R}(\cdot) = M^{-1} \sum_{m=1}^{M} \mathrm{Re}[\mathcal{L}\{Y, \widetilde{W}_m, \mathcal{B}_0, \theta_0(Z)\}],$$

where here and in what follows $(\cdot)$ denotes a real argument $\{Y, W, \underline{V}, \mathcal{B}_0, \theta_0(Z)\}$, where $\underline{V} = (V_1, \ldots, V_M)$. Here M is the Monte Carlo correction sample size. We will suppress the dependence on M in $\mathcal{R}(\cdot)$ for notational convenience. Note that $\mathcal{R}(\cdot)$ is a real valued function of real arguments and is a *criterion function* in the sense we discussed in the beginning of the section. Therefore, analogous to (2.1), we propose to estimate $\theta_0(z)$ by solving

$$n^{-1}\sum_{i=1}^{n} K_h(Z_i - z)G_i(z)\mathcal{R}_\theta\left\{Y_i, W_i, \underline{V}_i, \mathcal{B}^*, (\alpha_0, \alpha_1)G_i(z)\right\} = 0 \tag{2.3}$$

for α_0, setting $\widehat{\theta}(z, \mathcal{B}^*) = \widehat{\alpha}_0$ at some fixed $\mathcal{B}^*$.

There are two methods to estimate $\mathcal{B}$:

1. The profile kernel estimator $\widehat{\mathcal{B}}_{\mathrm{pf}}$ maximizes $n^{-1}\sum_{i=1}^{n}\mathcal{R}\{Y_i, W_i, \underline{V}_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\}$ in $\mathcal{B}$ solving

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}[\mathcal{R}_{\mathcal{B}}\{Y_i, W_i, \underline{V}_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\} + \mathcal{R}_\theta\{Y_i, W_i, \underline{V}_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\} \\ \times \widehat{\theta}_{\mathcal{B}}(Z_i, \mathcal{B})] = 0. \end{aligned} \tag{2.4}$$

2. The backfitting kernel estimator, $\widehat{\mathcal{B}}_{\mathrm{bf}}$ is obtained at convergence of the following iterations. Based on the current estimate, $\widehat{\mathcal{B}}_{\mathrm{cur}}$, solve for $\mathcal{B}$

$$n^{-1}\sum_{i=1}^{n}\mathcal{R}_{\mathcal{B}}\left\{Y_i, W_i, \underline{V}_i, \mathcal{B}, \widehat{\theta}(Z_i, \widehat{\mathcal{B}}_{\mathrm{cur}})\right\} = 0$$

call $\widehat{\mathcal{B}}_{\mathrm{new}}$, and then solve the local linear estimating equation (2.3) using $\widehat{\mathcal{B}}_{\mathrm{new}}$.

2.2. Asymptotic properties

In this section, we derive the asymptotic properties of our method in the case where the measurement error covariance matrix $\Sigma_{u,0}$ is known, see Section 2.4 for the case where it is estimated. The results given in Lin and Carroll ([12]) for the profiling and backfitting methods are true for any *criterion function* as long as various conditions are satisfied: these conditions translate into A1-A4, given in the Appendix. We assume that conditional expectations of $W, \underline{V}$ given Y, Z, X and the first and second order partial derivatives of $\mathcal{L}$ with respect to $\mathcal{B}$ and θ exist and are interchangeable. Define

$$\begin{aligned} \theta_{\mathcal{B}}(z, \mathcal{B}_0) &= -E\{\mathcal{R}_{\theta\mathcal{B}}(\cdot)|Z = z\}/E\{\mathcal{R}_{\theta\theta}(\cdot)|Z = z\} \\ &= -E[\mathcal{L}_{\theta\mathcal{B}}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}|Z = z]/E[\mathcal{L}_{\theta\theta}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}|Z = z]; \\ \Omega(z) &= f_Z(z)E\{\mathcal{R}_{\theta\theta}(\cdot)|Z = z\} \\ &= f_Z(z)E[\mathcal{L}_{\theta\theta}\{Y, X, \mathcal{B}_0, \theta_0(Z)\}|Z = z]; \\ \mathcal{V} &= E\{\mathcal{R}_{\mathcal{B}\mathcal{B}}(\cdot) + \mathcal{R}_{\mathcal{B}\theta}(\cdot)\theta_{\mathcal{B}}^{\mathrm{T}}(Z, \mathcal{B}_0)\} \\ &= E[\mathcal{L}_{\mathcal{B}\mathcal{B}}\{Y, X, \mathcal{B}_0, \theta_0(Z)\} + \mathcal{L}_{\mathcal{B}\theta}\{Y, X, \mathcal{B}_0, \theta(Z)\}\theta_{\mathcal{B}}^{\mathrm{T}}(Z, \mathcal{B}_0)]. \end{aligned}$$

Then the following result is a direct consequence of the main results of Lin and Carroll ([12]).

Theorem 2.1. *Assume that* (Y_i, Z_i, W_i), $i = 1, \dots, n$ *are independent and identically distributed; and* $\widehat{\mathcal{B}}_{\rm pf}$ *and* $\widehat{\theta}(\cdot)$ *are estimates obtained by using (2.3) and (2.4). Suppose further that the bandwidth* $h \propto n^{-c}$ *with* $1/5 \le c \le 1/3$. *Let* $\theta^{(2)}(z)$ *be the second derivative of* $\theta(z)$ *and* $\phi_2 = \int z^2 K(z)dz$. *Then, for the nonparametric part*

$$\widehat{\theta}(z, \widehat{\mathcal{B}}_{\rm pf}) - \theta_0(z) = (h^2/2)\phi_2\theta_0^{(2)}(z) - n^{-1}\sum_{i=1}^{n} K_h(Z_i - z)\mathcal{R}_{i\theta}(\cdot)/\Omega(z)$$
$$- \theta_{\mathcal{B}}(z, \mathcal{B}_0)^{\rm T}\mathcal{V}^{-1}n^{-1}\sum_{i=1}^{n}\{\mathcal{R}_{i\mathcal{B}}(\cdot) + \mathcal{R}_{i\theta}(\cdot)\theta_{\mathcal{B}}(Z_i, \mathcal{B}_0)\} + o_p(n^{-1/2});$$

and for the parametric part

$$\begin{aligned} n^{1/2}(\widehat{\mathcal{B}}_{\rm pf} - \mathcal{B}_0) &= -\mathcal{V}^{-1}n^{-1/2}\sum_{i=1}^{n}\{\mathcal{R}_{i\mathcal{B}}(\cdot) + \mathcal{R}_{i\theta}(\cdot)\theta_{\mathcal{B}}(Z_i, \mathcal{B}_0)\} + o_p(1) \\ &\Rightarrow \textit{Normal}(0, \mathcal{V}^{-1}\mathcal{F}\mathcal{V}^{-T}), \end{aligned}$$

where $\mathcal{F} = \textit{var}[\mathcal{R}_{\mathcal{B}}(\cdot) + \mathcal{R}_{\theta}(\cdot)\theta_{\mathcal{B}}(Z, \mathcal{B}_0)]$.

Theorem 2.2. *Make the same assumptions as in Theorem 2.1 except that* $nh^4 \to 0$. *Then the backfitting estimator* $\widehat{\mathcal{B}}_{\rm bf}$ *has the same limiting distribution as does the profile estimator* $\widehat{\mathcal{B}}_{\rm pf}$.

Remark 2.1. One can show that

$$\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3,$$

where $\mathcal{F}$ is $\mathcal{F}_1$ in the absence of measurement error,

$$\mathcal{F}_1 = \mathrm{var}[\mathcal{L}_{\mathcal{B}}\{Y, Z, X, \mathcal{B}_0, \theta_0(Z)\} + \mathcal{L}_{\theta}\{Y, Z, X, \mathcal{B}_0, \theta_0(Z)\}\theta_{\mathcal{B}}(Z, \mathcal{B}_0)],$$

$\mathcal{F}_2$ is the additional variation due to the use of corrected scores

$$\mathcal{F}_2 = \mathrm{E}(\mathrm{var}[\widetilde{\mathcal{L}}_{\mathcal{B}}\{Y, Z, W, \mathcal{B}_0, \theta_0(Z)\} + \widetilde{\mathcal{L}}_{\theta}\{Y, Z, W, \mathcal{B}_0, \theta_0(Z)\}\theta_{\mathcal{B}}(Z, \mathcal{B}_0)|Y, Z, X])$$

and $\mathcal{F}_3$ is the additional variation due to the use of Monte Carlo method

$$\mathcal{F}_3 = \mathrm{E}(\mathrm{var}[\mathcal{R}_{\mathcal{B}}(\cdot) + \mathcal{R}_{\theta}(\cdot)\theta_{\mathcal{B}}(Z, \mathcal{B}_0)|Y, Z, W]) = O(M^{-1}).$$

Remark 2.2. Estimation of the asymptotic variance of $\widehat{\mathcal{B}}_{\rm pf}$ or $\widehat{\mathcal{B}}_{\rm bf}$ is a straightforward exercise. To construct such estimates, all the expectations in the definitions of $\mathcal{V}$ and $\mathcal{F}$ are replaced by sums and all the regression functions are replaced by kernel estimates.

2.3. Multivariate measurement error models

Consider longitudinal or repeated measures data, where for each subject we observe L responses $\mathbf{Y} = (Y_1, \ldots, Y_L)$; and predictors $\mathbf{X} = (X_1, \ldots, X_L)$ and $\mathbf{Z} = (Z_1, \ldots, Z_L)$, where each Z_j is scalar. The underlying loglikelihood function is taken to be of the form $\mathcal{L}\{\mathbf{Y}, \mathbf{X}, \mathcal{B}_0, \theta_0(Z_1), \ldots, \theta_0(Z_L)\}$. Here the key feature is that the nonparametric component $\theta_0(\cdot)$ is evaluated multiple times per individual. For notational convenience, we use boldface to denote the multivariate version of the corresponding observations, and let $\underline{\theta}_0(\mathbf{Z}) = \{\theta_0(Z_1), \ldots, \theta_0(Z_L)\}$. In the measurement error settings under consideration, instead of observing X_j, we observe $W_j = X_j + U_j$, $j = 1, \ldots, L$ and assume that vec$(\mathbf{U})$, where "vec" is a vector form of a matrix, has a Normal distribution with mean zero and covariance matrix $\Sigma_{u,0}$. Let $\widetilde{\mathbf{W}}_m = \mathbf{W} + \iota \mathbf{V}_m$ for $m = 1, \ldots, M$, where vec$(\mathbf{V}_m)$ has a Normal distribution with mean zero and covariance matrix $\Sigma_{u,0}$. Then Monte Carlo corrected criterion function is given by

$$M^{-1} \sum_{m=1}^{M} \mathrm{Re}\Big[\mathcal{L}\{\mathbf{Y}, \widetilde{\mathbf{W}}_m, \mathcal{B}, \underline{\theta}(\mathbf{Z})\}\Big],$$

and our asymptotic results apply.

2.4. Estimation of error covariance matrix

We now consider the case in which the measurement error covariance matrix $\Sigma_{u,0}$ is estimated from replicated measurements. Suppose we observe $R \geq 2$ replicates $\{W_{i(r)}\}_{r=1}^{R}$, where $W_{i(r)} = X_i + U_{i(r)}$ and $U_{i(r)}$ has Normal$(0, \Sigma_{u,0})$. Then a root-n consistent estimate $\widehat{\Sigma}_u$ of $\Sigma_{u,0}$ is the sample covariance matrix of the terms $W_{i(r)}$s is $n^{-1}\sum_{i=1}^{n} S_i$ where

$$S_i = \frac{\sum_{r=1}^{R}\{W_{i(r)} - \overline{W}_{i\cdot}\}\{W_{i(r)} - \overline{W}_{i\cdot}\}^T}{(R-1)}, \text{ and } \overline{W}_{i\cdot} = R^{-1}\sum_{r=1}^{R} W_{i(r)}.$$

Let $\gamma = \mathrm{vech}(\Sigma_u)$, where "vech" is the vector half, i.e., the vector of the unique elements of Σ_u. Then we have that $\widehat{\gamma} - \gamma_0 = n^{-1}\sum_i \mathrm{vech}(S_i - \Sigma_{u,0})$. Since V can be written as $\Sigma_u^{1/2} e$ where e comes from a standard Normal distribution, we can redefine the score equations (2.4) as

$$n^{-1}\sum_{i=1}^{n}[\mathcal{R}_{\mathcal{B}}\{Y_i, \overline{W}_{i(r)}, \Sigma_u^{1/2}\underline{e}_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B}, \Sigma_u)\} + \mathcal{R}_{\theta}\{Y_i, \overline{W}_{i(r)}, \Sigma_u^{1/2}\underline{e}_i, \mathcal{B}, \widehat{\theta}(Z_i, \mathcal{B})\} \times \widehat{\theta}_{\mathcal{B}}(Z_i, \mathcal{B}, \Sigma_u)] = 0.$$

Let subscript γ denote a partial derivative with respect to γ. Then following Section 4 of Lin and Carroll ([12]), we have the following asymptotic expansion for the profile estimator:

$$\begin{aligned} n^{1/2}(\widehat{\mathcal{B}}_{\mathrm{pf}} - \mathcal{B}_0) &= -\mathcal{V}^{-1}\bigg[n^{-1/2}\sum_{i=1}^{n}\{\mathcal{R}_{i\mathcal{B}}(\cdot) + \mathcal{R}_{i\theta}(\cdot)\theta_{\mathcal{B}}(Z_i, \mathcal{B}_0, \Sigma_{u,0})\} \\ &\qquad + \mathcal{V}_{\mathcal{B}\gamma} n^{1/2}(\widehat{\gamma} - \gamma)\bigg] + o_p(1) \\ &= -\mathcal{V}^{-1}\bigg(n^{-1/2}\sum_{i=1}^{n}[\mathcal{R}_{i\mathcal{B}}(\cdot) + \mathcal{R}_{i\theta}(\cdot)\theta_{\mathcal{B}}(Z_i, \mathcal{B}_0, \Sigma_{u,0}) \\ &\qquad + \mathcal{V}_{\mathcal{B}\gamma}\{\mathrm{vech}(S_i - \Sigma_{u,0})\}]\bigg) + o_p(1), \end{aligned}$$

where $\mathcal{V}_{\mathcal{B}\gamma} = \mathrm{E}\{\mathcal{R}_{i\mathcal{B}\gamma}(\cdot) + \theta_{\mathcal{B}}(Z_i, \mathcal{B}_0, \Sigma_{u,0})\mathcal{R}^{\mathrm{T}}_{i\theta\gamma}(\cdot)\}$. The covariance matrix of $n^{1/2}(\widehat{\mathcal{B}}_{\mathrm{pf}} - \mathcal{B}_0)$ follows from the above expressions and its estimator can be constructed as discussed in Remark 2.2.

3. Special case: Partially linear model

Two regression examples, the Univariate and Multivariate Partially Linear Models, are considered to illustrate the relationship between proposed Monte Carlo corrected score method and other methods that exist in the literature. Poisson and Logistic Partially Linear Models are also studied via simulations in the next section to demonstrate the general applicability of proposed methods.

3.1. Univariate partially linear model

Estimation in the partially linear model with error prone covariates are described in [9]. In this section we derive the asymptotic distribution of our estimates explicitly and compare our estimates to that of [9].

Consider the model

$$Y_i = X_i^{\mathrm{T}}\gamma_0 + \theta_0(Z_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ has a Normal distribution with mean zero and variance σ_0^2. Let $\beta = (\gamma^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$ and choose the loglikelihood to be our criterion function

$$\mathcal{L}\{Y, X, \theta(Z), \beta\} = -\log(\sigma^2)/2 - (2\sigma^2)^{-1}\{Y - X^{\mathrm{T}}\gamma - \theta(Z)\}^2.$$

Then, the corrected loglikelihood as discussed in the previous section is

$$\begin{aligned} &\mathcal{R}\{Y, W, \widetilde{V}, \theta(Z), \beta\} \\ &= -\log(\sigma^2)/2 - M^{-1}\sum_{m=1}^{M}\mathrm{Re}[(2\sigma^2)^{-1}\{Y - (W + \iota V_m)^{\mathrm{T}}\gamma - \theta(Z)\}^2] \\ &= -\log(\sigma^2)/2 - (2\sigma^2)^{-1}\bigg[\{Y - W^{\mathrm{T}}\gamma - \theta(Z)\}^2 - \gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\bigg]. \end{aligned}$$

Also, define

$$\begin{aligned}\Gamma &= E[\{X - E(X|Z)\}(\epsilon - U^{\mathrm{T}}\gamma_0)^2\{X - E(X|Z)\}^{\mathrm{T}}] + E(UU^{\mathrm{T}}\epsilon^2)\\ &\quad + E\{(UU^{\mathrm{T}} - \Sigma_{u,0})\gamma_0\gamma_0^{\mathrm{T}}(UU^{\mathrm{T}} - \Sigma_{u,0})^{\mathrm{T}}\};\\ \mathcal{S} &= \mathrm{cov}\{X - E(X|Z)\};\\ \tau^2 &= E\{(\epsilon - U^{\mathrm{T}}\gamma_0)^2 - (\sigma_0^2 + \gamma_0^{\mathrm{T}}\Sigma_{u,0}\gamma_0)\}^2;\\ \mathcal{C} &= \Sigma_{u,0}\gamma_0 + [E\{U(U^{\mathrm{T}}\gamma_0)^3\} - \Sigma_{u,0}\gamma_0\gamma_0^{\mathrm{T}}\Sigma_{u,0}\gamma_0]/(2\sigma_0^2).\end{aligned}$$

Then we have the following result:

Theorem 3.1. *Let $\widehat{\gamma}$ and $\widehat{\sigma}^2$ denote the estimates based on our method. Then jointly,*

$$n^{1/2}\begin{pmatrix}\widehat{\gamma} - \gamma_0\\ \widehat{\sigma}^2 - \sigma_0^2\end{pmatrix} \Rightarrow \mathit{Normal}\left\{\mathbf{0}, \begin{bmatrix}\mathcal{S}^{-1}\Gamma\mathcal{S}^{-T} + R_{11} & 2\sigma^2\mathcal{S}^{-1}\mathcal{C} + R_{12}\\ \bullet & \tau^2 + R_{22}\end{bmatrix}\right\},$$

where $R_{11}, R_{12}, R_{22} \to 0$ as $M \to \infty$.

See Appendix B for the proof and the expressions of R_{11}, R_{12} and R_{22}.

Remark 3.1. It is important to note that as R_{ij}s vanish, the limiting distribution of our estimators becomes the same as of estimators in [9]. Ma and Carroll ([15]) showed that the estimator in [9] is exactly the same as theirs when posited $p(X)$, $p(W|X)$ and $p(Y|X,Z)$ are all normal, proving that our estimator is also efficient when $M = \infty$.

3.2. Multivariate partially linear model

We illustrate our approach on the following multivariate partially linear measurement error model discussed in Lin and Carroll ([12])

$$Y_{ij} = X_{ij}^{\mathrm{T}}\beta_0 + \theta_0(Z_{ij}) + e_{ij}, \tag{3.1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, L$, where $\mathbf{e}_i = (e_{i1}, \ldots, e_{iL})^{\mathrm{T}}$ has a Normal distribution with mean zero and covariance matrix $\Sigma_{\epsilon,0}$. Let $\mathcal{B} = (\beta, \Sigma_\epsilon)$ be the parameter of interest. Then the criterion function ignoring the measurement errors is given by

$$\begin{aligned}&\mathcal{L}\{\mathbf{Y}, \mathbf{X}, \mathcal{B}, \theta(\widetilde{Z})\}\\ &\quad = (1/2)\log\{\det(\Sigma_\epsilon^{-1})\} - (1/2)\{\mathbf{Y} - \mathbf{X}\beta - \theta(\mathbf{Z})\}^{\mathrm{T}}\Sigma_\epsilon^{-1}\{\mathbf{Y} - \mathbf{X}\beta - \theta(\mathbf{Z})\}.\end{aligned}$$

The Monte-Carlo Corrected Scores criterion function is given by

$$\begin{aligned}\mathcal{R}(\cdot) &= M^{-1}\sum_{m=1}^{M}\mathrm{Re}[\mathcal{L}\{\mathbf{Y}, \widetilde{\mathbf{W}}_m, \mathcal{B}, \theta(\mathbf{Z})\}]\\ &= (1/2)\log\{\det(\Sigma_\epsilon^{-1})\} - (1/2)\{\mathbf{Y} - \mathbf{W}\beta - \theta(\mathbf{Z})\}^{\mathrm{T}}\Sigma_\epsilon^{-1}\{\mathbf{Y} - \mathbf{W}\beta - \theta(\mathbf{Z})\}\\ &\qquad + (1/2)\beta^{\mathrm{T}}\left(M^{-1}\sum_{m=1}^{M}\mathbf{V}_m^{\mathrm{T}}\Sigma_\epsilon^{-1}\mathbf{V}_m\right)\beta.\end{aligned}$$

The backfitting algorithm is easy to apply in this case. Given the current estimates, $\widehat{\mathcal{B}}_{\rm cur} = (\widehat{\beta}_{\rm cur}, \widehat{\Sigma}_{\epsilon,{\rm cur}})$, the new estimates are given by

$$\widehat{\beta}_{\rm new} = \left[n^{-1}\sum_{i=1}^{n}\left\{\mathbf{W}_i^{\rm T}\widehat{\Sigma}_{\epsilon,{\rm cur}}^{-1}\mathbf{W}_i - M^{-1}\sum_{m=1}^{M}(\mathbf{V}_{im}^{\rm T}\widehat{\Sigma}_{\epsilon,{\rm cur}}^{-1}\mathbf{V}_{im})\right\}\right]^{-1} \times n^{-1}\sum_{i=1}^{n}\mathbf{W}_i^{\rm T}\widehat{\Sigma}_{\epsilon,{\rm cur}}^{-1}\{\mathbf{Y}_i - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\rm cur})\};$$

and

$$\widehat{\Sigma}_{\epsilon,{\rm new}} = n^{-1}\sum_{i=1}^{n}\Bigg[\{\mathbf{Y}_i - \mathbf{W}_i\widehat{\beta}_{\rm cur} - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\rm cur})\} \times \{\mathbf{Y}_i - \mathbf{W}_i\widehat{\beta}_{\rm cur} - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\rm cur})\}^{\rm T} - M^{-1}\sum_{m=1}^{M}(\mathbf{V}_{im}\widehat{\beta}_{\rm cur}\widehat{\beta}_{\rm cur}^{\rm T}\mathbf{V}_{im}^{\rm T})\Bigg].$$

Profile pseudolikelihood estimates are also easily constructed. Let $\mathcal{S}$ be a smoother matrix as in [11] and define $\mathcal{Y} = (Y_{11}, \ldots, Y_{nm})^{\rm T}$ and $\mathcal{T} = (\mathbf{W}_1^{\rm T}, \ldots, \mathbf{W}_n^{\rm T})^{\rm T}$. Let $\mathcal{T}_* = (I - \mathcal{S})\mathcal{T}$, $\mathcal{Y}_* = (I - \mathcal{S})\mathcal{Y}$ and $\widetilde{\Sigma}_\epsilon = I_n \bigotimes \Sigma_\epsilon$. Then for given Σ_ϵ, the profile estimate of β is given by

$$\hat{\beta}_{pf} = \left\{\mathcal{T}_*^{\rm T}\widetilde{\Sigma}_\epsilon^{-1}\mathcal{T}_* - \sum_i\left(M^{-1}\sum_m \mathbf{V}_{im}^{\rm T}\Sigma_\epsilon^{-1}\mathbf{V}_{im}\right)\right\}^{-1}\mathcal{T}_*^{\rm T}\widetilde{\Sigma}_\epsilon^{-1}\mathcal{Y}_*.$$

A simple estimate of $\Sigma_{\epsilon,0}$ can be constructed by first forming the working independence estimate of β_0, then applying the above equation to obtain $\widehat{\Sigma}_{\epsilon,new}$ and iterating the steps until convergence.

Remark 3.2. Estimation of the error covariance matrix $\Sigma_{u,0}$ and its impact on limiting distribution theory for estimation of $\mathcal{B}_0$ is described in Section 2.4.

Remark 3.3. As $M \to \infty$, our estimators converges to those given in Lin and Carroll ([12]), see Appendix C for a sketch of proof.

3.3. Performance with respect to efficient methods

We show that under the assumption that X is generated from a Gaussian distribution, our method using different $M < \infty$, as well as Lin and Carroll's procedure ([12]), which is equivalent to our method with $M = \infty$, performs very similar to the semiparametrically efficient method.

Suppose X is from a Normal distribution with mean $\mu_{x,0}$ and covariance matrix $\Sigma_{x,0}$; and for simplicity of notation we let β_0 be a scalar. Then the criterion function becomes

$$\begin{aligned}\mathcal{L}_{\rm G}(\cdot) \quad = \quad & -(1/2)\big\{\log(|\mathcal{J}(\mu_x, \Sigma_x, \Sigma_\epsilon)|) + [\mathbf{Y} - \mathcal{Q}\{\mathbf{W}, \beta, \theta(\mathbf{Z}), \mu_x, \Sigma_x\}]^{\rm T} \\ & \times \{\mathcal{J}(\mu_x, \Sigma_x, \Sigma_\epsilon)\}^{-1}[\mathbf{Y} - \mathcal{Q}\{\mathbf{W}, \beta, \theta(\mathbf{Z}), \mu_x, \Sigma_x\}] \\ & + \log(|\Sigma_x + \Sigma_u|) - (1/2)(\mathbf{W} - \mu_x)^{\rm T}(\Sigma_x + \Sigma_u)^{-1}(\mathbf{W} - \mu_x)\big\},\end{aligned}$$

where

$$\begin{aligned}\mathcal{Q}\{\mathbf{W},\beta,\theta(\mathbf{Z}),\mu_x,\Sigma_x\} &= \beta\mu_x+\theta(\mathbf{Z})+\beta\Sigma_x(\Sigma_x+\Sigma_u)^{-1}(\mathbf{W}-\mu_x);\\ \mathcal{J}(\beta,\Sigma_x,\Sigma_\epsilon) &= \Sigma_\epsilon+\beta^2\Sigma_x(\Sigma_x+\Sigma_u)^{-1}\Sigma_u.\end{aligned}$$

By the results of Lin and Carroll ([12]), the estimators based on $\mathcal{L}_{\mathrm{G}}(\cdot)$ are semiparametrically efficient.

We compared our method with the optimal one (see discussion above) via a simulation study under the following scenario. We set $L=3$ and let $\beta_0=0.7$, $\theta_0(z)=0.5\cos(2z)-1$. We chose $\mu_{x,0}=(-1,-1,-1)^{\mathrm{T}}$, $\Sigma_{\epsilon,0}=I_3+0.3(J_3-I_3)$, $\Sigma_{x,0}=I_3$, and $\Sigma_{u,0}=0.3I_3+0.2J_3$, where J_k denotes the $k\times k$ matrix with all the elements equal to one. We generated Z from a Uniform on $[0,\pi]$ distribution.

Under this setup, we generated $1,000$ data sets following the model given by (3.1) with $n=200$. We used Epanechnikov kernel with the bandwidth estimated as $\widehat{\sigma}_z n^{-1/3}$, where $\widehat{\sigma}_z$ is the sample standard deviation of Z. Using each data set we constructed backfitting estimator of β_0 using our method with different values of M ranging from 100 to 500; Lin and Carroll method ([12]), which is ours when $M=\infty$; and semiparametrically efficient method (using $\mathcal{L}_G(\cdot)$). Root mean squared error (RMSE) of $\widehat{\beta}$ does not differ much between $M=100$ and $M=\infty$ (as in Lin and Carroll [12]) and is equal to 0.1432. Hence we show that our methods do not require many Monte Carlo runs. In fact in our numerical studies we find $M=150$ satisfactory. RMSE for the semiparametrically efficient method is 0.1421. This indicates that the efficiency of our method is very close to the optimal (0.77 percent loss).

4. Simulation study

4.1. Partially linear poisson regression model

We study the performance of our method via a simulation study. We considered the partially linear Poisson regression model where the response Y is Poisson distributed with mean $\lambda(X,Z)=X^{\mathrm{T}}\beta_0+\theta_0(Z)$. The true variable $X=(X_1,X_2)$ was generated from a bivariate standard Gaussian distribution, Z was generated from a Uniform on $[0,\pi]$ distribution. Error prone covariate was generated as $W=X+U$ where U followed a bivariate normal distribution with mean zero and a known covariance matrix $\Sigma_{u,0}$. We set $\Sigma_{u,0}=I_2$ and $\mathcal{B}_0=(\beta_{1,0},\beta_{2,0})=(0.2,0.2)$ and used two different functions: (1) $\theta_0(z)=5-0.5\cos(z)$ and (2) $\theta_0(z)=5-0.5\cos(2z)$. We generated $1,000$ samples of size $n=1,000$ and used $M=500$ as Monte Carlo correction sample size.

We employed Epanechnikov kernel to estimate the nonparametric function. We used the globally fixed bandwidth $h_n=\kappa\widehat{\sigma}_z n^{-1/3}$, where $\widehat{\sigma}_z$ is the estimated standard deviation of Z and κ is some selected positive number. We report the results for $\kappa=1$. Similar results were obtained for other values of κ ranging from 0.5 to 2. We used backfitting to estimate $\mathcal{B}_0$.

The results are displayed in Table 1. It is clear that our method produces only a small bias and favorable coverage probability.

Table 1

Results of the simulations using Poisson regression model. In nonparametric part: (1) is $\theta_0(z) = 5 - 0.5\cos(z)$ and (2) is $\theta_0(z) = 5 - 0.5\cos(2z)$. Reported are the mean, empirical standard errors (e.s.e.), root mean squared error (RMSE) and empirical coverage of 95% confidence intervals of $\beta_{1,0}$ and $\beta_{2,0}$ based on 1000 simulated data sets each with a sample size $n = 1000$

$\theta_0(z)$	estimation of $\beta_{1,0} = 0.2$				estimation of $\beta_{2,0} = 0.2$			
	mean	e.s.e.	RMSE	95%	mean	e.s.e.	RMSE	95%
(1)	0.203	0.039	0.039	0.951	0.199	0.038	0.038	0.952
(2)	0.204	0.040	0.041	0.954	0.202	0.041	0.041	0.955

4.2. Partially linear logistic model

We borrowed a simulation scenario applied to a logistic regression model from Ma and Carroll ([15]). As in their paper, we considered the model $\text{logit}\{\text{pr}(Y = 1|X, Z)\} = \beta_{1,0}X + \beta_{2,0}X^2 + \theta_0(Z)$, where $W = X + U$ and U is from Normal distribution with mean 0 and variance $\sigma^2_{u,0}$ with $\sigma^2_{u,0}$ known. We set $\sigma^2_{u,0} = 0.16$, $\mathcal{B}_0 = (\beta_{1,0}, \beta_{2,0}) = (0.7, 0.7)$ and used two different functions: (1) $\theta_0(z) = 0.5\cos(z) - 1$ and (2) $\theta_0(z) = 0.5\cos(2z) - 1$. The covariates X and Z were generated from Normal distribution with mean -1 and variance 1, and Uniform on $[0, \pi]$, respectively. We used the sample size of $n = 500$ and $M = 150$ Monte Carlo correction sample size. Backfitting with Epanechnikov kernel was used to estimate the model components. For the sake of comparison, we used the global bandwidth $h_n = \widehat{\sigma}_z n^{-1/3}$ as in Ma and Carroll ([15]).

Technically, the logistic regression setup as described above does not fall into our framework as the logistic distribution function is not entire in the complex plane. However, Novick and Stefanski ([18]) pointed out that for small measurement error variance one can still apply corrected score based methods, with only a minor bias. Specifically, Novick and Stefanski ([18]) followed the same paradigm to construct likelihood based corrected score, call it $\Psi_M(Y, W)$ and showed that $\Psi^*(Y, X) = E[\lim_{M\to\infty} \Psi_M(Y, W)|Y, X]$ is not the same as the true likelihood score function. However, they argued that the differences between the components of $\Psi^*(Y, X)$ and the true likelihood score function are small for measurement error variances of the magnitudes commonly encountered in applications (see p. 479 of Novick and Stefanski [18]).

The results of the simulation study are displayed in Table 2. It is evident that our method is comparable in both cases to that of Ma and Carroll ([15]) in terms of mean squared error and coverage probability, albeit with the small bias expected from the fact that the logistic function is not entire on the complex plane. It is clear from this numerical example that even though technically our method is not applicable in logistic regression, it performs quite well and is very close to that of Ma and Carroll ([15]).

The simulation was repeated for a much larger measurement error variance, $\sigma^2_{u,0} = 0.5$ versus $\sigma^2_{u,0} = 0.16$. The results are shown in Table 2. Again, our results indicate only a small bias and favorable coverage probability. Ma and Carroll ([15]) did not report results for this situation so it is not possible to compare our method with theirs.

Table 2

Results of the simulations using logit model. In nonparametric part: (1) is $\theta_0(z) = 0.5\cos(z) - 1$ and (2) is $\theta_0(z) = 0.5\cos(2z) - 1$. Reported are the mean, empirical standard errors (e.s.e.), root mean squared error (RMSE) and empirical coverage of 95% confidence intervals of $\beta_{1,0}$ and $\beta_{2,0}$ for two values of $\sigma^2_{u,0}$, based on 1000 simulated data sets each with a sample size $n = 500$. Our method is coded as MA in the column "Me" and the results from Ma and Carroll ([15]) are coded as MC. Ma and Carroll [15] did not consider the case of $\sigma^2_{u,0} = 0.5$

$\theta_0(z)$	$\sigma^2_{u,0}$	Me	estimation of $\beta_{1,0} = 0.7$				estimation of $\beta_{2,0} = 0.7$			
			mean	e.s.e.	RMSE	95%	mean	e.s.e.	RMSE	95%
(1)	0.16	MA	0.638	0.261	0.268	0.942	0.653	0.149	0.156	0.940
	0.16	MC	0.720	0.277	0.278	0.947	0.726	0.156	0.158	0.939
	0.50	MA	0.621	0.272	0.283	0.948	0.633	0.161	0.175	0.943
(2)	0.16	MA	0.615	0.238	0.253	0.943	0.639	0.135	0.148	0.942
	0.16	MC	0.727	0.276	0.277	0.951	0.728	0.155	0.158	0.940
	0.50	MA	0.600	0.250	0.269	0.942	0.618	0.155	0.175	0.945

4.3. Multivariate partially linear model

In this section we consider the partially linear Gaussian regression model with repeated measures. We generated data according to the model

$$Y_{ij} = X_{ij1}\beta_{1,0} + X_{ij2}\beta_{2,0} + \theta_0(Z_{ij}) + e_{ij},$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, L$, where $\mathbf{e}_i = (e_{i1}, \ldots, e_{iL})^{\mathrm{T}}$ has a Normal distribution with mean zero and covariance matrix $\Sigma_{\epsilon,0}$. We set $L = 3$ and let $\beta_{1,0} = \beta_{2,0} = 0.7$. The covariates $X_{ij\ell}$ were each generated from a Normal with mean -1 and variance 1 distribution and Z was generated from a Uniform on $[0, \pi]$ distribution. The error prone variable was generated as $W_{ij\ell} = X_{ij\ell} + U_{ij\ell}$, where $(U_{ij1}, U_{ij2}, U_{ij3})$ follows a normal distribution with mean zero and covariance matrix $\Sigma_{u,0} = 0.3I_3 + 0.2J_3$. Here J_k denotes the $k \times k$ matrix with all the elements equal to one.

We used two different functions: (1) $\theta_0(z) = 0.5\cos(z) - 1$ and (2) $\theta_0(z) = 0.5\cos(2z) - 1$, and considered two different covariance structures for $\Sigma_{\epsilon,0}$: (1) compound symmetry with common marginal variance 1 and correlation 0.3 and (2) AR(1) structure with autocorrelation parameter 0.4.

We generated $1,000$ samples of size $n = 200$ for each scenarion, and used $M = 150$ as Monte Carlo correction score sample size. The Gaussian kernel was used to estimate the nonparametric function and backfitting was used to estimate the parameters. We used the globally fixed bandwidth $h_n = \kappa\widehat{\sigma}_z n^{-1/3}$, where $\widehat{\sigma}_z$ is the estimated standard deviation of Z and κ is some selected positive number. We report the results for $\kappa = 1$. Similar results were obtained for other values of κ ranging from 0.5 to 2.

The results are displayed in Table 3. It is evident that our method produces only a small bias and favorable coverage probability for both cases: compound symmetry and AR(1) error structures.

Table 3

Results of the simulations using multivariate Gaussian measurement error. In nonparametric part: (1) is $\theta_0(z) = 0.5\cos(z) - 1$ and (2) is $\theta_0(z) = 0.5\cos(2z) - 1$. Reported are the mean, empirical standard errors (e.s.e.), root mean squared error (RMSE) and empirical coverage of 95% confidence intervals of $\beta_{1,0}$ and $\beta_{2,0}$ based on 1,000 simulated data sets each with a sample size $n = 200$ for different error correlation structures

Compound symmetry								
$\theta_0(z)$	estimation of $\beta_{1,0} = 0.7$				estimation of $\beta_{2,0} = 0.7$			
	mean	e.s.e.	RMSE	95%	mean	e.s.e.	RMSE	95%
(1)	0.687	0.137	0.138	0.945	0.691	0.144	0.145	0.947
(2)	0.685	0.143	0.144	0.949	0.689	0.152	0.152	0.947
AR(1)								
$\theta_0(z)$	estimation of $\beta_{1,0} = 0.7$				estimation of $\beta_{2,0} = 0.7$			
	mean	e.s.e.	RMSE	95%	mean	e.s.e.	RMSE	95%
(1)	0.687	0.137	0.138	0.948	0.690	0.144	0.145	0.949
(2)	0.686	0.143	0.143	0.946	0.689	0.152	0.153	0.946

5. Application

5.1. Nevada test site thyroiditis data example

In this section we apply our method to the Nevada test site (NTS) thyroid study data. The study was conducted in 1980's by the University of Utah. The original study is described in [27, 7] and [24]. The main idea of the study was to relate the incidence of thyroid related diseases to the exposure of radiation to the thyroid. In this study, 2,491 individuals, who were exposed to radiation as children, were tested for thyroid disease. The primary radiation exposure to the thyroid glands of these children came from the ingestion of milk and vegetables contaminated with radioactive isotopes of iodine. Recently, the dosimetry for the study was redone ([25]), and the study results were reported in [14].

Due to the fact that the actual radiation doses in foods or in the thyroid gland of the individuals are not available, the estimated radiation doses are well known to be contaminated with measurement errors. Many authors have studied and described measurement error properties and analysis in this context ([20, 23, 16, 28, 13, 19, 22, 8]). A common approach is to build a large dosimetry model that attempts to convert the known data about above-ground nuclear testing to the radiation actually absorbed into the thyroid. Dosimetry calculations for individual subjects were based upon several variables, such as, age at exposure, gender, residence history, whether as a child the individual was breast-fed, and a diet questionnaire filled out by the parent focusing on milk consumption and vegetables. The data were then put into a complex model and for each individual, the point estimate of thyroid dose (the arithmetic mean of a lognormal distribution of dose estimates) and an associated error term (the geometric standard deviation) for the measurement error were reported.

It is typical to assume that radiation doses are estimated with a combination of Berkson and a classical measurement error ([20]). In the log-scale, true log-dose T is related to observed or calculated log-dose W by a latent intermediate

X via

$$\begin{aligned} T &= X + U_{\text{berk}}; \\ W &= X + U_{\text{class}}, \end{aligned}$$

where U_{berk} and U_{class} are the Berkson uncertainty and the classical uncertainty, respectively, with corresponding variances $\sigma^2_{u,\text{berk},0}$ and $\sigma^2_{u,\text{class},0}$ depending on the individual. It is typical to assume that the errors U_{berk} have Gaussian distributions. In the NTS study, the total uncertainty $\sigma^2_{u,\text{berk},0} + \sigma^2_{u,\text{class},0}$ is known but not the relative contributions. We will let 50% of the total uncertainty be classical in our study.

We take the incidence of thyroiditis (inflammation of the thyroid gland), Y, as the response variable. If the latent, X, could be observed then typically the total mean dose, $\mathrm{E}(T|X) = \exp(X + \sigma^2_{u,\text{berk},0}/2)$ would be the main predictor. In addition, we consider S, the sex of the patient and Z, age at exposure (standardized to have mean zero and variance 1), which are measured without measurement error. We include Z, age at exposure, nonparametrically into so called excess relative risk model

$$\mathrm{pr}(Y = 1|X, S, Z) = H[\beta_0 S + \log\{1 + \gamma_0 \exp(X + \sigma^2_{u,\text{berk},0}/2)\} + \theta_0(Z)], \quad (5.1)$$

where $H(\cdot)$ is the logistic distribution function, $\theta_0(\cdot)$ is an unknown function and γ_0 is called the excess relative risk.

We employed our method discussed in Section 2 to the model given by (5.1). Specifically, we used backfitting with Epanechnikov kernel and bandwidth equal to 1.5 (similar results were obtained for 1.0 and 2.0). We used $M = 100$ as the Monte Carlo correction sample size. We compared our method to the naive method, when one ignores the measurement error of both types. The estimated effect of gender, $\widehat{\beta}_1 \approx 1.75$ for both the naive and Monte Carlo corrected scores method. This can be explained by the fact that gender and radiation dose for an individual are essentially independent and hence the effect of gender is not affected by measurement error in radiation dose.

The estimated value of the relative risk parameter was 8.54 for the naive method and 17.19 for the proposed method. The effect of age, Z, is displayed in Figure 1 for both the naive and MCCS procedures. It is evident from the results that because of the difference in the estimate of the excess relative risk γ, there is a noticeable difference in the estimated age effect when measurement error is taken into account.

Remark 5.1. As noted in Section 4, the logistic regression setup does not fall into our framework as the logistic distribution function is not entire in the complex plane. To observe the performance of semiparametric-Monte Carlo corrected scores in this example, we compared our results to the well known SIMEX procedure ([5, 26]). Please refer to Apanasovich et. al. ([2]) for details. In short, we modeled the age effect parametrically by a quadratic polynomial and used a quartic extrapolant for SIMEX. The estimated value of the excess relative risk parameter was 15.92. We can see that in this case the SIMEX estimate is close to what proposed method produces.

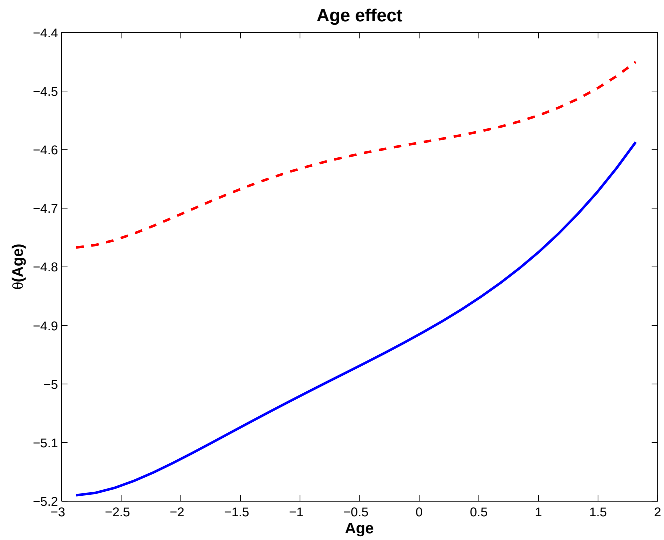

FIG 1. *Estimated age effect in the Nevada Test Site thyroiditis data. Solid line: the proposed estimate. Dashed line: the naive estimate ignoring the presence of measurement error.*

5.2. Simulation study mimicking the data example

To further assess the performance of our method when applied to the data example, we set up a simulation study where the measurement error present in the variables is of the same magnitude as the measurement error from the data example. Specifically, we generate responses using the model

$$\mathrm{pr}(Y = 1|X, S, Z) = H[\beta_0 S + \log\{1 + \gamma_0 \exp(X + \sigma^2_{u,\mathrm{berk},0}/2)\} + \theta_0(Z)],$$

where we generate S from a Binary distribution with rate 0.5, X from a Normal distribution with mean zero and variance 4; and Z from a Uniform distribution within $[-3, 2]$ interval. Following the data example, we set true $\beta_0 = 1.7$ and $\gamma_0 = 17.2$ and take $\theta_0(z)$ to be the assumed estimate of a true function. We generate 500 samples of size $n = 2500$ and, as in the example, use $M = 100$ as Monte Carlo corrected score runs.

As in the data example, the measurement error is Gaussian, a combination of Berkson and classical and we take the variance of classical uncertainty to be 50% of the total uncertainty present in the NTS data set.

The average of estimates for β_0 and γ_0 over the 500 generated data sets are 1.72 and 17.79, respectively, with empirical standard errors 0.18 and 4.69. It is evident that the proposed procedure performs quite well, giving low bias and moderate variability for the estimates when the magnitude of measurement errors is similar to that of the original NTS data.

6. Discussion

We consider the problem of estimation in a general semiparametric regression model when error-prone covariates are modeled parametrically while covariates measured without error are modeled nonparametrically. We propose to utilize ideas of corrected score methodology introduced by [18] for a purely parametric framework. Monte Carlo corrected scores method uses simulations and allows for simple implementation of the corrected score methodology in problems for which direct calculation of corrected scores is difficult. Its implementation is straightforward in any programming language that allows for complex-value arithmetic.

Following the common practice in a general semiparametric regression, our method is based upon profiling and backfitting for a criterion function. Our theoretical developments seem to be novel even though are supported by previously developed ideas. We demonstrate that our methods are general and include the existing in a literature methods for univariate and multivariate partially linear models as special cases. Moreover, we show that in a partially linear model and logistic model with quadratic effects of X, our method produces mean square errors similar to ones from the semiparametric efficient method. It should be noted that the theory of our method fails for logistic model, however it performs well even in presence of large measurement errors.

Here we have focused on the case where the covariate modeled nonparametrically is univariate. However, the idea of building a semiparametric criterion function using Monte-Carlo corrected scores can be applied to more general problems, e.g., additive models.

Acknowledgments

This work forms part of the first author's Ph.D. dissertation at Texas A&M University.

Appendix A: Conditions and assumptions

Regularity conditions. We require the following conditions.

A1. Distribution law of Z is absolutely continuous and has compact support $\mathcal{Z}$, its density $f_Z(\cdot)$ is differentiable on $\mathcal{Z}$, the derivative is continuous and $\inf_{z\in\mathcal{Z}} f_Z(z) > 0$. Moreover $\sup_{z\in\mathcal{Z}} |\theta_0(z)| \le M < \infty$. X also has a compact support, $\mathcal{X}$.

A2. Mixed partial derivatives $\frac{\partial^{r+t}}{\partial\mathcal{B}^r\partial\theta^t}\widetilde{\mathcal{L}}(Y, W, \mathcal{B}, \theta)$, $0 \le r, t \le 4$, $r + t \le 4$, exist for almost all (Y, W) and $\mathrm{E}\{|\frac{\partial^{r+t}}{\partial\mathcal{B}^r\partial\theta^t}\mathcal{L}(Y, X, \mathcal{B}, \theta)|^2\}$ are bounded.

A3. The smallest and the largest eigenvalues of matrix $\mathcal{V}$ are bounded away from zero and infinity. Moreover, $G(Z) =$

$$\mathrm{E}\left\{\frac{\partial}{\partial(\mathcal{B}^T,\theta^T)^T}\mathcal{L}(Y, X, \mathcal{B}_0, \theta)|_{\theta=\theta_0(Z)}\frac{\partial}{\partial(\mathcal{B}^T,\theta^T)}\mathcal{L}(Y, X, \mathcal{B}_0, \theta)|_{\theta=\theta_0(Z)} \mid Z\right\}$$

possesses a continuous derivative and $\inf_{z\in\mathcal{Z}} G(z) > 0$.

A4. $\frac{\partial^{k+l}}{\partial z^k \partial \mathcal{B}^l}\theta_0(z,\mathcal{B})$, $0 \le k+1 \le 3$, exist and continuous for almost all z and $\mathcal{B}$; and $||\frac{\partial^{k+l}}{\partial z^k \partial \mathcal{B}^l}\theta_0(z,\mathcal{B})||_\infty < \infty$.

Appendix B: Proof of Theorem 3.1

We start by noting that the loglikelihood is given by

$$\mathcal{L}(\bullet) \quad = \quad -\log(\sigma^2)/2 - \{Y - X^{\mathrm{T}}\gamma - \theta(Z)\}^2/(2\sigma^2).$$

Define $\epsilon^* = Y - W^{\mathrm{T}}\gamma - \theta(Z)$. Then by definition, we have

$$\mathcal{R}(\bullet) = -\log(\sigma^2)/2 - (\epsilon^{*2} - \gamma^{\mathrm{T}} M^{-1} \sum_{m=1}^{M} V_m V_m^{\mathrm{T}} \gamma)/(2\sigma^2).$$

Note that the parameter of interest is $\beta = (\gamma^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$.

Direct calculations yield

$$\begin{aligned}
\mathcal{R}_\beta(\bullet) &= \begin{pmatrix} (W\epsilon^* + M^{-1}\sum_m V_m V_m^{\mathrm{T}}\gamma)/\sigma^2 \\ -1/(2\sigma^2) + [\epsilon^{*2} - \gamma^{\mathrm{T}} M^{-1}\sum_m V_m V_m^{\mathrm{T}}\gamma]/(2\sigma^4) \end{pmatrix}; \\
\mathcal{R}_\theta(\bullet) &= \epsilon^*/\sigma^2; \\
\mathcal{R}_{\beta\beta}(\bullet) &= \\
&\begin{bmatrix} (-WW^{\mathrm{T}} + \frac{1}{M}\sum_m V_m V_m^{\mathrm{T}})/\sigma^2 & -[W\epsilon^* + \frac{1}{M}\sum_m V_m V_m^{\mathrm{T}}\gamma]/\sigma^4 \\ -[W\epsilon^* + \frac{1}{M}\sum_m V_m V_m^{\mathrm{T}}\gamma]/\sigma^4 & \frac{1}{2\sigma^4} - [\epsilon^{*2} - \gamma^{\mathrm{T}}\frac{1}{M}\sum_m V_m V_m^{\mathrm{T}}\gamma]/\sigma^6 \end{bmatrix}; \\
\mathcal{R}_{\beta\theta}(\bullet) &= \begin{pmatrix} -W/\sigma^2 \\ -\epsilon^*/\sigma^4 \end{pmatrix}; \\
\mathcal{R}_{\theta\theta}(\bullet) &= -1/\sigma^2.
\end{aligned}$$

Using these, we obtain that

$$\begin{aligned}
\theta_\beta(\bullet) &= -E\{\mathcal{R}_{\beta\theta}(\bullet)|Z\}/E\{\mathcal{R}_{\theta\theta}(\bullet)|Z\} = -\begin{Bmatrix} E(W|Z) \\ 0 \end{Bmatrix}; \\
E\{\mathcal{R}_{\beta\beta}(\bullet)\} &= \begin{bmatrix} -E(XX^{\mathrm{T}})/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{bmatrix} \\
E\{\mathcal{R}_{\beta\theta}(\bullet)\theta_\beta(\bullet)^{\mathrm{T}}\} &= \begin{bmatrix} (\sigma^2)^{-1}E\{WE(W|Z)^{\mathrm{T}}\} & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} E\{XE(X|Z)^{\mathrm{T}}\}/\sigma^2 & 0 \\ 0 & 0 \end{bmatrix}.
\end{aligned}$$

Hence, using the definition that $\mathcal{S} = \mathrm{var}\{X - E(X|Z)\}$, we derive that

$$\mathcal{V} \quad = \quad E\{\mathcal{R}_{\beta\beta}(\bullet) + \mathcal{R}_{\beta\theta}(\bullet)\theta_\beta(\bullet)^{\mathrm{T}}\} = \begin{bmatrix} -\mathcal{S}/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) & . \end{bmatrix}$$

Also, let $\mathcal{K} = \mathcal{R}_\beta + \mathcal{R}_\theta \theta_\beta$. Then we have from Theorem 2.1 that

$$n^{1/2}(\widehat{\beta} - \beta_0) \quad \Rightarrow \quad \text{Normal}(0, \mathcal{V}^{-1}\mathcal{F}\mathcal{V}^{-T}),$$

where $\mathcal{F} = \text{var}(\mathcal{K})$.

To complete the proof, we need to derive the asymptotic covariance matrix $\mathcal{V}^{-1}\mathcal{F}\mathcal{V}^{-T}$. First we note that

$$\begin{aligned} \mathcal{K} &= (1/\sigma^2)\begin{pmatrix} \{W - E(W|Z)\}\epsilon^* + M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma \\ -1/2 + [\epsilon^{*2} - \gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma]/(2\sigma^2) \end{pmatrix} \\ &= (1/\sigma^2)\begin{pmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{pmatrix}. \end{aligned}$$

Hence we have

$$\text{var}(\mathcal{K}) \quad = \quad (1/\sigma^4)\begin{pmatrix} \text{var}(\mathcal{K}_1) & \mathcal{F}_{12} \\ \mathcal{F}_{12}^{\mathrm{T}} & \text{var}(\mathcal{K}_2) \end{pmatrix},$$

where $\mathcal{F}_{12} = E(\mathcal{K}_1\mathcal{K}_2) - E(\mathcal{K}_1)E(\mathcal{K}_2)$. Now we derive,

$$\begin{aligned} \text{var}(\mathcal{K}_1) &= \text{var}\Bigg[\{X - E(X|Z) + U\}(\epsilon - U^{\mathrm{T}}\gamma) + M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\Bigg] \\ &= \text{var}\Bigg[\{X - E(X|Z)\}(\epsilon - U^{\mathrm{T}}\gamma) + U\epsilon - UU^{\mathrm{T}}\gamma + M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\Bigg] \\ &= \text{var}[\{X - E(X|Z)\}(\epsilon - U^{\mathrm{T}}\gamma)] + \text{var}(U\epsilon) + \text{var}\{(UU^{\mathrm{T}} - \Sigma_u)\gamma\} \\ &\quad + \text{var}\Bigg\{M^{-1}\sum_{m=1}^{M}(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\Bigg\} \\ &= \text{var}[\{X - E(X|Z)\}(\epsilon - U^{\mathrm{T}}\gamma)] + \text{var}(U\epsilon) + \text{var}\{(UU^{\mathrm{T}} - \Sigma_u)\gamma\} \\ &\quad + M^{-1}\text{var}\{(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\} \\ &= \Gamma + M^{-1}\text{var}\{(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\}, \end{aligned}$$

and

$$\begin{aligned} &(4\sigma^4)\text{var}(\mathcal{K}_2) = \text{var}(\epsilon^{*2}) + \text{var}\Bigg(\gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\Bigg) \\ &= E\{(\epsilon - U^{\mathrm{T}}\gamma)^4\} - (\sigma^2 + \gamma^{\mathrm{T}}\Sigma_u\gamma)^2 + M^{-2}\sum_{m=1}^{M}\text{var}\{\gamma^{\mathrm{T}}(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\} \\ &= E\{(\epsilon - U^{\mathrm{T}}\gamma)^2 - (\sigma^2 + \gamma^{\mathrm{T}}\Sigma_u\gamma)\}^2 + M^{-1}\text{var}\{\gamma^{\mathrm{T}}(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\} \\ &= \tau^2 + M^{-1}\text{var}\{\gamma^{\mathrm{T}}(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\}. \end{aligned}$$

Finally, we derive

$$\begin{aligned} E(\mathcal{K}_1) &= E\left[\{W - E(W|Z)\}\epsilon^* + M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\right] = -\Sigma_u\gamma + \Sigma_u\gamma = 0; \\ E(\mathcal{K}_2) &= E\left(-1/2 + \left[\epsilon^{*2} - \gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\right]/(2\sigma^2)\right) = 0; \\ E(\mathcal{K}_1\mathcal{K}_2) &= E\left(\left[\{W - E(W|Z)\}\epsilon^* + M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\right]\right. \\ &\qquad \left.\times\left[-1/2 + [\epsilon^{*2} - \gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma]/(2\sigma^2)\right]\right). \end{aligned}$$

Note that $\epsilon^* = \epsilon - U^{\mathrm{T}}\gamma$ is independent of X and Z. Hence using $W = X + U$, we observe

$$\begin{aligned} &E(\{W - E(W|Z)\}\epsilon^{*3}) = -3\sigma^2\Sigma_u\gamma - E\{U(U^{\mathrm{T}}\gamma)^3\}; \\ &E\left[\{W - E(W|Z)\}\epsilon^*\gamma^{\mathrm{T}} M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\right] = -\Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma; \\ &E\left[M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\epsilon^{*2}\right] = \Sigma_u\gamma\sigma^2 + \Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma; \\ &E\left[\left\{M^{-1}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\right\}\left\{\gamma^{\mathrm{T}} M^{-1}\sum_{l=1}^{M} V_l V_l^{\mathrm{T}}\gamma\right\}\right] \\ &\quad = E\left[M^{-2}\sum_{m=1}^{M}\sum_{l=1}^{M} V_m V_m^{\mathrm{T}}\gamma\gamma^{\mathrm{T}} V_l V_l^{\mathrm{T}}\gamma\right] \\ &\quad = E\left[M^{-2}\sum_{m=1}^{M} V_m V_m^{\mathrm{T}}\gamma\gamma^{\mathrm{T}} V_m V_m^{\mathrm{T}}\gamma\right] + E\left[M^{-2}\sum_{m=1}^{M}\sum_{l\neq m=1}^{M} V_l V_l^{\mathrm{T}}\gamma\gamma^{\mathrm{T}} V_l V_l^{\mathrm{T}}\gamma\right] \\ &\quad = M^{-1}E\{V_m(V_m^{\mathrm{T}}\gamma)^3\} + M^{-1}(M-1)\Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma. \end{aligned}$$

Hence we see that

$$\begin{aligned} E(\mathcal{K}_1\mathcal{K}_2) &= \Sigma_u\gamma/2 - 3\Sigma_u\gamma/2 - \{E\{U(U^{\mathrm{T}}\gamma)^3 - \Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma\}/(2\sigma^2) \\ &\quad - (1/2)\Sigma_u\gamma + (1/2)\Sigma_u\gamma + \Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma/(2\sigma^2) \\ &\quad - M^{-1}E\{V_m(V_m^{\mathrm{T}}\gamma)^3\}/(2\sigma^2) - M^{-1}(M-1)\Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma/(2\sigma^2) \\ &= -\Sigma_u\gamma - [E\{U(U^{\mathrm{T}}\gamma)^3\} - \Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma]/(2\sigma^2) \\ &\quad - M^{-1}E\{V_m(V_m^{\mathrm{T}}\gamma)^3\}/(2\sigma^2) + \{1 - M^{-1}(M-1)\}\Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma/(2\sigma^2). \end{aligned}$$

Finally, we obtain that the asymptotic covariance matrix is given by

$$\mathcal{V}^{-1}\mathcal{F}\mathcal{V}^{-T} = \begin{bmatrix} \mathcal{S}^{-1}\Gamma\mathcal{S}^{-T} + R_{11} & 2\sigma^2\mathcal{S}^{-1}\mathcal{C} + R_{12} \\ \cdot & \tau^2 + R_{22} \end{bmatrix},$$

where

$$\begin{aligned} R_{11} &= M^{-1}\mathcal{S}^{-1}\mathrm{cov}\{(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\}\mathcal{S}^{-T}; \\ R_{12} &= 2\sigma^2\mathcal{S}^{-1}[M^{-1}E\{V_m(V_m^{\mathrm{T}}\gamma)^3\} - \{1 - M^{-1}(M-1)\}\Sigma_u\gamma\gamma^{\mathrm{T}}\Sigma_u\gamma]; \\ R_{22} &= M^{-1}\mathrm{var}\{\gamma^{\mathrm{T}}(V_m V_m^{\mathrm{T}} - \Sigma_u)\gamma\}. \end{aligned}$$

Now the result follows from Theorem 2.1.

Appendix C: Proof of Remark 3.3

Recall that, for a fixed M, Monte Carlo corrected criterion function is given by

$$\begin{aligned} \mathcal{R}_M(\cdot) &= M^{-1}\sum_{m=1}^{M}\mathrm{Re}[\mathcal{L}\{\mathbf{Y}, \widetilde{\mathbf{W}}_m, \mathcal{B}, \theta(\mathbf{Z})\}] \\ &= (1/2)\log\{\det(\Sigma_\epsilon^{-1})\} - (1/2)\{\mathbf{Y} - \mathbf{W}\beta - \theta(\mathbf{Z})\}^{\mathrm{T}}\Sigma_\epsilon^{-1}\{\mathbf{Y} - \mathbf{W}\beta - \theta(\mathbf{Z})\} \\ &\quad + (1/2)\beta^{\mathrm{T}}\left(M^{-1}\sum_{m=1}^{M}\mathbf{V}_m^{\mathrm{T}}\Sigma_\epsilon^{-1}\mathbf{V}_m\right)\beta. \end{aligned}$$

We observe that as $M \to \infty$, $\mathcal{R}_M(\cdot) = \mathcal{R}_\infty(\cdot) + O_p(M^{-1/2})$, where

$$\begin{aligned} \mathcal{R}_\infty(\cdot) = {} & [\log\{\det(\Sigma_\epsilon^{-1})\} - \{\mathbf{Y} - \mathbf{T}\beta - \theta(\mathbf{Z})\}^{\mathrm{T}}\Sigma_\epsilon^{-1}\{\mathbf{Y} - \mathbf{T}\beta - \theta(\mathbf{Z})\} \\ & + \beta^{\mathrm{T}}E(\mathbf{V}_m^{\mathrm{T}}\Sigma_\epsilon^{-1}\mathbf{V}_m)\beta]/2. \end{aligned}$$

Note that $\mathcal{R}_\infty(\cdot)$ is exactly the same criterion function as in Lin and Carroll[12] (see their equation (22), p. 81). Hence as $M \to \infty$, the estimates based on $\mathcal{R}_\infty(\cdot)$ are given as follows: given the current estimates, $\widehat{\mathcal{B}}_{\mathrm{cur}} = (\widehat{\beta}_{\mathrm{cur}}, \widehat{\Sigma}_{\epsilon,\mathrm{cur}})$, the new estimates are given by

$$\begin{aligned} \widehat{\beta}_{\mathrm{new}} &= \left[n^{-1}\sum_{i=1}^{n}\{\mathbf{W}_i^{\mathrm{T}}\widehat{\Sigma}_{\epsilon,\mathrm{cur}}^{-1}\mathbf{W}_i - E(\mathbf{V}_m^{\mathrm{T}}\widehat{\Sigma}_{\epsilon,\mathrm{cur}}^{-1}\mathbf{V}_m)\}\right]^{-1} \\ &\quad \times n^{-1}\sum_{i=1}^{n}\mathbf{W}_i^{\mathrm{T}}\widehat{\Sigma}_{\epsilon,\mathrm{cur}}^{-1}\{\mathbf{Y}_i - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\mathrm{cur}})\}; \\ \widehat{\Sigma}_{\epsilon,\mathrm{new}} &= n^{-1}\sum_{i=1}^{n}[\{\mathbf{Y}_i - \mathbf{W}_i\widehat{\beta}_{\mathrm{cur}} - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\mathrm{cur}})\}\{\mathbf{Y}_i - \mathbf{W}_i\widehat{\beta}_{\mathrm{cur}} \\ &\quad - \widehat{\theta}(\mathbf{Z}_i, \widehat{\mathcal{B}}_{\mathrm{cur}})\}^{\mathrm{T}} - E(\mathbf{V}_m\widehat{\beta}_{\mathrm{cur}}\widehat{\beta}_{\mathrm{cur}}^{\mathrm{T}}\mathbf{V}_m^{\mathrm{T}})]. \end{aligned}$$

Again, as $M \to \infty$, we note that the estimating equation for β and Σ_ϵ are same as Lin and Carroll [12] (see their equation (23), p. 81). Hence the result.

References

[1] Al-Abood, A. M. and Young, D. H. (1986). The power of approximate tests for the regression coefficients in a gamma regression model. *IEEE Transactions On Reliability*, R-35, 216-220.

[2] Apanasovich, T.V., Carroll, R. J., Maity, A. (2009). SIMEX and standard error estimation in semiparametric measurement error models. *Electronic Journal of Statistics*, 3, 318-348. MR2497157

[3] Cameron, A.C. and Trivedi, P.K. (1998). Regression analysis of count data, Cambridge:Cambridge University Press. MR1648274

[4] Carroll, R. J., Ruppert, D., Crainiceanu, C. and Stefanski, L. A. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. London: CRC Press. MR2243417

[5] 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.

[6] Eckert, R.S., Carroll, R.J. and Wang, N. (1997). Transformations to additivity in measurement error models. *Biometrics*. 53, 262-272. MR1450184

[7] Kerber, R. L., Till, J. E., Simon, S. L., Lyon, J. L. Thomas, D. C., Preston-Martin, S., Rollison, M. L., Lloyd, R. D. and Stevens, W. (1993). A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *Journal of the American Medical Association*, 270, 2076-2083.

[8] Li, Y., Guolo, A., Owen Hoffman, F., and Carroll, R. J. (2007). Shared Uncertainty in Measurement Error Problems, with Application to Nevada Test Site Fallout Data. *Biometrics*, 63, 1226-36. MR2414601

[9] Liang, H., Haerdle, W. and Carroll, R.J. (1999). Estimation in a partially linear error-in-variables model. *Annals of Statistics*, 27, 1519-1535. MR1742498

[10] Liang, H. and Ren, H.B. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics*, 14, 237-250. MR2137900

[11] Lin, X., Wang, N., Welsh, A. and Carroll, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered data. *Biometrika*, 91, 177–193. MR2050468

[12] Lin, X. and Carroll, R.J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society, Series B*, 68, 69-88. MR2212575

[13] Lubin, J. H., Schafer, D. W. Ron, E., Stovall, M. and Carroll, R. J. (2004). A reanalysis of thyroid neoplasms in the Israeli tinea capitis study accounting for dose uncertainties. *Radiation Research*, 161, 359-368.

[14] Lyon, J. L., Alder, S. C., Stone, M. B., Scholl, A., Reading, J. C. Holubkov, R., Sheng, X. White, G. L., Hegmann, K. T., Anspaugh, L., Hoffman, F. O., Simon, S. L., Thomas, B., Carroll, R. J. & Meikle, A. W. (2006). Thyroid disease associated with

exposure to the Nevada Test Site radiation: a reevaluation based on corrected dosimetry and examination data. *Epidemiology*, 17, 604-614.

[15] Ma, Y. and Carroll, R.J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association*, 101, 1465-1474. MR2279472

[16] Mallick, B., Hoffman, F. O. and Carroll, R. J. (2002). Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada Test Site. *Biometrics*, 58, 13-20. MR1891038

[17] Nakamura, T. (1990). Corrected score functions for error-in-variable models: methodology and application to generalized linear models, *Biometrika*, 77, 127-137. MR1049414

[18] Novick, J.S. and Stefanski, L.A. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97, 472-481. MR1941464

[19] Pierce, D. A. and Kellerer, A. (2004). Adjusting for covariate errors with nonparametric assessment of the true covariate distribution. *Biometrika*, 91, 863-876. MR2126038

[20] Reeves, G. K., Cox, D. R., Darby, S. C. and Whitley, E. (1998). Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine*, 17, 2157-2177.

[21] Singpurwalla, N. D. (1971). A problem in accelerated Life testing. *Journal of the American Statistical Association*, 66, 841-845.

[22] Schafer, D. W. and Gilbert, E. S. (2006). Some statistical implications of does uncertainty in radiation dose-response analyses. *Radiation Research*, 166, 303-312.

[23] Schafer, D. W., Lubin, J. H., Ron, E., Stovall, M. and Carroll, R. J. (2001). Thyroid cancer following scalp irradiation: a reanalysis accounting for uncertainty in dosimetry. *Biometrics*, 57, 689-697. MR1859805

[24] Simon, S. L., Till, J. E., Lloyd, R. D., Kerber, R. L., Thomas, D. C., Preston–Martin, S., Lyon, J. L. and Stevens, W. (1995). The Utah Leukemia case–control study: dosimetry methodology and results. *Health Physics*, 68, 460–471.

[25] Simon, S. L., Anspaugh, L. R., Hoffman, F. O., et al. (2006). 2004 update of dosimetry for the Utah Thyroid Cohort Study. *Radiation Research*, 165, 208-222.

[26] Stefanski, L.A. and Cook, J.R. (1995). Simulation-Extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, 90, 1247-1256. MR1379467

[27] Stevens, W., Till, J. E., Thomas, D. C., et al. (1992). Assessment of leukemia and thyroid disease in relation to fallout in Utah: report of a cohort study of thyroid disease and radioactive fallout from the Nevada test site. University of Utah.

[28] Stram, D. O. and Kopecky, K. J. (2003). Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose

estimates are based on a complex dosimetry system: some observations. *Radiation Research*, 160, 408-417.

[29] Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, 91, 835-848. MR2126036

[30] Zhu, L. and Cui, H. (2003). A semiparametric regression model with errors in variables. *Scan. J. Statist.* 30, 429-442. MR1983135