Functional and Structural Methods with Mixed Measurement Error and Misclassification in Covariates

Grace Y. Yi*[1], Yanyuan Ma[2], Donna Spiegelman[3] and Raymond J. Carroll[4]

Abstract

Covariate measurement imprecision or errors arise frequently in many areas. It is well known that ignoring such errors can substantially degrade the quality of inference or even yield erroneous results. Although in practice both covariates subject to measurement error and covariates subject to misclassification can occur, research attention in the literature has mainly focused on addressing either one of these problems separately. To fill this gap, we develop estimation and inference methods that accommodate both characteristics simultaneously. Specifically, we consider measurement error and misclassification in generalized linear models under the scenario that an external validation study is available, and systematically develop a number of effective functional and structural methods. Our methods can be applied to different situations to meet various objectives.

Some Key Words: External validation study; Functional measurement error modeling; Generalized linear models; Likelihood method; Measurement error; Misclassification; Regression calibration; Semiparametric regression; Simulation extrapolation algorithm; Structural measurement error modeling.

Short title: Misclassified and Mismeasured Data

[1] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, yyi@uwaterloo.ca; *the correspondence author

[2] Department of Statistics, Texas A&M University, TAMU 3143, College Station, TX 77843-3143, ma@stat.tamu.edu

[3] Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115, stdls@channing.harvard.edu

[4] Department of Statistics, Texas A&M University, TAMU 3143, College Station, TX 77843-3143, and School of Mathematical Sciences, University of Technology, Sydney, Broadway NSW 2007, carroll@stat.tamu.edu

1 Introduction

Mismeasurement in variables arises ubiquitously in practice, and it has long been a concern in various fields including clinical and epidemiological studies. In nutrition studies, for instance, food frequency questionnaires are commonly used to measure diet, and it is known that this instrument involves a large degree of variation and measurement error (e.g., Rosner et al. 1989). Measurement error may occur for different reasons. Sometimes variables may be difficult to observe precisely due to physical location or cost. Sometimes they are impossible to measure accurately due to their nature. For example, the level of exposure to radiation cannot be measured accurately (e.g., Pierce et al. 1992). In other situations, a variable may represent the average of a certain quantity over time, and any practical way of measuring such a variable necessarily incurs error.

Variables are often classified into two different categories leading to *measurement error* in continuous variables and *misclassification* of discrete variables. It is known that ignoring mismeasurement of variables often leads to biased results. For example, in the simple linear regression model where a covariate is subject to classical additive error, the estimate of the slope can be attenuated towards zero if the error in the covariate is ignored. The effect of mismeasurement in a covariate can be complex, generally depending on the form of the error model and the relationship between the response and the covariates as well as the distribution of the covariates. There is an enormous literature on this subject (e.g., Stefanski and Carroll, 1987; Nakamura, 1990; Carroll and Wand, 1990; Rosner et al., 1990; Rosner et al., 1992; Wang and Davidian, 1996; Wang et al., 1998; Lin and Carroll, 2000; Huang and Wang, 2001; Liang and Wang, 2005; Spiegelman et al., 2005; Zucker and Spiegelman, 2004; Zucker and Spiegelman, 2008; Sugar et al., 2007; Hall and Ma, 2007; Yi, 2008; Liang, 2009; Yi et al., 2011; Yi et al., 2012). Textbook treatments of measurement error in regression can be found in Fuller (1987), Gustafson (2004), Carroll et al. (2006) and Buonaccorsi (2010).

Although there has been extensive attention on either covariate measurement error (e.g., Carroll et al., 2006; Buonaccorsi, 2010) or covariate misclassification (e.g., Akazawa et al., 1998; Gustafson, 2004; Buonaccorsi, et al., 2005; Wang et al., 2008; Dalen et al., 2009; Buonaccorsi, 2010 and the references therein), relatively little work has been published addressing both characteristics simultaneously. The only work we are aware of is Spiegelman et al. (2000), where both continuous and discrete covariates are allowed to be subject to error or misclassification. However, this work was restricted to a binary outcome, using logistic regression and maximum likelihood to obtain estimates and inference. More discussion and a generalization of their work to allow for misspecification of the model describing the relationship between the true and observed covariates are given in Section 3.3.1.

Our goal is to develop a rich class of methods to handle data with both covariate measurement error and misclassification under general model frameworks. For convenience, we embed our approach within the framework of the generalized linear model, a class of models that are widely applied in practice. Regarding the measurement error and misclassification processes, we consider the scenario that an external validation study is available. We develop a number of estimation methods and inference tools which apply under a wide range of circumstances. Our investigation covers both functional and structural modeling strategies for the measurement error and misclassification processes. In particular, our likelihood method differs from that of Spiegelman et al. (2000) in several aspects which are discussed in Section 3.4.

Our paper contains the following sections. In Section 2, we describe the models for the data with mismeasured continuous covariates or misclassified discrete covariates, and in Section 3, we develop methods to correct bias induced from mismeasured or misclassified covariates. Asymptotic theory is described in Section 4. In Section 5, we further propose two methods that are approximate but easy to implement to partially correct for bias due to measurement error and misclassification. To assess the performance of our methods, we conduct simulation studies and present an empirical data analysis in Section 6. Concluding remarks are given in Section 7.

2 Notation and Model Setup

2.1 Response Model

Suppose there are n individuals in the main study. For $i = 1, ..., n$, let Y_i be the response variable for the ith subject. Let $\mathbf{W}_i$ be the vector of covariates that are precisely measured, $\mathbf{X}_i$ be the vector of error-prone covariates, and Z_i be a scalar binary covariate subject to misclassification. Extensions to multiple binary covariates subject to misclassification are straightforward but involve more complex notation.

We are interested in the relationship between the response variable Y_i and the true covariates $(\mathbf{X}_i, Z_i, \mathbf{W}_i)$. In particular, we link the response to the covariates using a parametric model $f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i; \boldsymbol{\beta})$ where f is a specified function and $\boldsymbol{\beta}$ is a vector of unknown parameters. The primary objective here is to conduct estimation and inference about the parameters in $\boldsymbol{\beta}$. To clearly demonstrate our proposed methods, here we consider a concrete model form: the generalized linear model. To be specific, assume that Y_i has the probability density or mass function from the exponential family

$$f(y_i) = \exp[\{y_i\theta - b(\theta)\}/d(\phi) + c(y_i, \phi)],$$

where $b(\cdot)$, $c(\cdot, \cdot)$ and $d(\cdot)$ are known functions, θ is a canonical parameter, and ϕ is a dispersion parameter. The mean and variance of Y_i are $b'(\theta)$ and $d(\phi)b''(\theta)$, respectively (McCullagh and Nelder 1989).

Let $\mu_i = E(Y_i|\mathbf{X}_i, Z_i, \mathbf{W}_i)$ and $v_i = \text{var}(Y_i|\mathbf{X}_i, Z_i, \mathbf{W}_i)$ denote the conditional mean and variance of Y_i, given covariates, respectively. It is customary to set $\mu_i = g^{-1}\{b'(\theta)\}$ and $v_i = h^{-1}[g^{-1}\{b'(\theta)\}]$ for some functions g and h, together with a regression model for $b'(\theta)$. More specifically, we consider the regression model

$$g(\mu_i) = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_x + Z_i\beta_z + \mathbf{W}_i^{\mathrm{T}}\boldsymbol{\beta}_w, \tag{1}$$

where $g(\cdot)$ is a known monotone function, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^{\mathrm{T}}, \beta_z, \boldsymbol{\beta}_w^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression parameters. An intercept may be included in $\boldsymbol{\beta}_w$ by including 1 in the covariate vector $\mathbf{W}_i$. Further, assume $v_i = h^{-1}(\mu_i, \phi)$, where $h(\cdot)$ is a known function and ϕ is the dispersion or scale parameter that is known or may be estimated. For instance, for binary data there is no ϕ and $v_i = \mu_i(1 - \mu_i)$.

2.2 Measurement Error and Misclassification Processes

Let $\mathbf{X}_i^*$ and Z_i^* be the observed measurements of $\mathbf{X}_i$ and Z_i, respectively, $i = 1, ..., n$. Suppose that the data come from a main study and a validation study. In the main study, there are no measurements of the error-prone covariates $\mathbf{X}_i$ and Z_i but measurements on other variables are available, while the external validation study has measurements on covariates only. That is, the available data from the main and the validation studies are $\{(y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i) : i \in \mathcal{M}\}$ and $\{(\mathbf{x}_i^*, z_i^*, \mathbf{x}_i, z_i, \mathbf{w}_i) : i \in \mathcal{V}\}$, respectively, where $\mathcal{M}$ and $\mathcal{V}$ contain n and m subjects, respectively. Here we assume that the subjects in those two studies are not the same, i.e, $\mathcal{M}$ and $\mathcal{V}$ do not overlap, and further assume that given $\mathbf{W}_i$, the conditional distribution of $(\mathbf{X}_i, Z_i, \mathbf{X}_i^*, Z_i^*)$ for $i \in \mathcal{V}$ is the same as that of $(\mathbf{X}_i, Z_i, \mathbf{X}_i^*, Z_i^*)$ for $i \in \mathcal{M}$ so that the information carried by the validation sample $\mathcal{V}$ can be transported to the main study $\mathcal{M}$ when carrying out inferences. This assumptio is similar to but different from the transportability assumption made by Spiegelman et al. (2000), who assumed that the conditional distribution of $(\mathbf{X}_i, Z_i)$ given $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is the same in both the main and the validation studies. The feasibility of a transportability assumption is basically justified by the nature of individual study designs. Our assumption is typically reasonable for scenarios where both main and external validation studies are carried out to the same population using the same data collection procedures.

Let $p_i = \text{pr}(Z_i^* = 0 | \mathbf{X}_i, Z_i = 1, \mathbf{W}_i)$ and $q_i = \text{pr}(Z_i^* = 1 | \mathbf{X}_i, Z_i = 0, \mathbf{W}_i)$ be the misclassification probabilities that may depend on the true covariates. Regression models for binary data can be employed to model the misclassification probabilities. Typically, we consider logistic regression models, bearing in mind that any parametric modeling can be employed for individual problems,

$$\text{logit}(p_i) = \alpha_{01} + \boldsymbol{\alpha}_{x1}^{\text{T}} \mathbf{X}_i + \boldsymbol{\alpha}_{w1}^{\text{T}} \mathbf{W}_i, \text{ and } \text{logit}(q_i) = \alpha_{00} + \boldsymbol{\alpha}_{x0}^{\text{T}} \mathbf{X}_i + \boldsymbol{\alpha}_{w0}^{\text{T}} \mathbf{W}_i \tag{2}$$

where $\boldsymbol{\alpha} = (\alpha_{01}, \boldsymbol{\alpha}_{x1}^{\text{T}}, \boldsymbol{\alpha}_{w1}^{\text{T}}, \alpha_{00}, \boldsymbol{\alpha}_{x0}^{\text{T}}, \boldsymbol{\alpha}_{w0}^{\text{T}})^{\text{T}}$ is the vector of regression parameters.

For the measurement error process, it is often reasonable to assume that $f(\mathbf{x}_i^*|z_i^*, \mathbf{x}_i, z_i, \mathbf{w}_i) = f(\mathbf{x}_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$, i.e. the $\mathbf{X}_i^*$ surrogate measurement is independent of surrogate Z_i^*, given the true covariates $(\mathbf{X}_i, Z_i, \mathbf{W}_i)$. A parametric model may then be employed to specify $f(\mathbf{x}_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$. Various options can be found in Carroll et al. (2006) for this purpose. As an example, we consider that $\mathbf{X}_i$ and $\mathbf{X}_i^*$ follow a regression model, i.e., given $(\mathbf{X}_i, Z_i, \mathbf{W}_i)$,

$$\mathbf{X}_i^* = \Gamma_x \mathbf{X}_i + \boldsymbol{\gamma}_z Z_i + \Gamma_w \mathbf{W}_i + \mathbf{e}_i, \tag{3}$$

where the error terms $\mathbf{e}_i$ are mean zero normal variables and are independent of other variables, Γ_x and Γ_w are conforming matrices, and $\boldsymbol{\gamma}_z$ is a vector of parameters. We write the vector formed by the elements of Γ_x, Γ_w as $\boldsymbol{\gamma}_x$ and $\boldsymbol{\gamma}_w$.

Different specification of the coefficient vectors or matrices features various measurement error models. For instance, setting Γ_w and Γ_x to be a zero and unit matrices respectively and $\boldsymbol{\gamma}_z$ to be a zero vector in (3) gives a classical additive model (Carroll et al. 2006); nonzero vector $\boldsymbol{\gamma}_z$ distinguishes different measurement error models corresponding to the two subpopulations categorized by $Z_i = 0$ or $Z_i = 1$. Again, measurement error models do not have to be restricted to the regression model (3); any parametric modeling of the measurement error process can be handled by our proposed methods.

3 Methodology

3.1 Likelihood Function

We assume conditional independence between the response variable Y_i and the surrogate measurements $(\mathbf{X}_i^*, Z_i^*)$, given the true covariates $(\mathbf{X}_i, Z_i, \mathbf{W}_i)$, i.e., measurement error and misclassification are nondifferential in the sense that $f(\mathbf{x}_i^*, z_i^*|y_i, \mathbf{x}_i, z_i, \mathbf{w}_i) = f(\mathbf{x}_i^*, z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$, or equivalently, $f(y_i|\mathbf{x}_i, z_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i) = f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i)$. Estimation of the model parameters relies on the factorization

$$f(y_i, \mathbf{x}_i, z_i, \mathbf{x}_i^*, z_i^*|\mathbf{w}_i) = f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i) f(\mathbf{x}_i^*, z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i) f(\mathbf{x}_i, z_i|\mathbf{w}_i), \tag{4}$$

where the last term $f(\mathbf{x}_i, z_i|\mathbf{w}_i)$ is a nuisance function. The factorization (4) allows us to model one type of variables at a time. Specifically, the first term $f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i)$ is determined by the response model (1), and the middle term $f(\mathbf{x}_i^*, z_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i)$ is determined by the measurement error and misclassification models (3) and (2), i.e., $f(\mathbf{x}_i^*, z_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i) = f(\mathbf{x}_i^* | z_i^*, \mathbf{x}_i, z_i, \mathbf{w}_i) f(z_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i) = f(\mathbf{x}_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i) f(z_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i)$. Estimation of the associated parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ can be carried out based on the validation data $\{(\mathbf{x}_i, z_i, \mathbf{w}_i, \mathbf{x}_i^*, z_i^*) : i \in \mathcal{V}\}$. The last term $f(\mathbf{x}_i, z_i | \mathbf{w}_i)$ features the probability distribution of the true covariate processes; and specification or nonspecification of this quantity leaves us room to take different estimation approaches.

3.2 Pseudo-Likelihood Method

Because measurements of the response variable are only available for the main study, one might attempt to estimate $\boldsymbol{\beta}$ using the observed likelihood contributed by the subjects in the main study, which is immediate from the factorization (4):

$$L_i = \int\int f(y_i | \mathbf{x}_i, z_i, \mathbf{w}_i) f(\mathbf{x}_i^*, z_i^* | \mathbf{x}_i, z_i, \mathbf{w}_i) f(\mathbf{x}_i, z_i | \mathbf{w}_i) d\eta(\mathbf{x}_i) d\eta(z_i), \tag{5}$$

where $d\eta(\cdot)$ represents the dominating measure which is either Lebesgue or counting measure for continuous or discrete random variable. This method requires modeling the covariate distribution $f(\mathbf{x}_i, z_i | \mathbf{w}_i)$, which can be restrictive sometimes and will be relaxed in Section 3.3. Let $\boldsymbol{\delta}$ be the vector of parameters governing the covariate process $f(\mathbf{x}_i, z_i | \mathbf{w}_i)$, and $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\vartheta}^{\mathrm{T}})^{\mathrm{T}}$. Then under regularity conditions including that $\boldsymbol{\theta}$ is identifiable, maximizing $\prod_{i \in \mathcal{M}} L_i$ with respect to $\boldsymbol{\theta}$ yields a consistent estimator of $\boldsymbol{\theta}$.

This approach is conceptually easy to implement. However, it overlooks the available measurements from the validation data set, and furthermore, using the main study data alone would usually lead to nonidentifiability issues for the model parameters (Küchenhoff 1990). To overcome these limitations, we propose a pseudo-likelihood method for estimation of $\boldsymbol{\theta}$, where the validation data serve as the basis for modeling and estimation pertaining to the true covariate process $f(\mathbf{x}_i, z_i | \mathbf{w}_i)$. In principle, $f(\mathbf{x}_i, z_i | \mathbf{w}_i)$ can be factorized as either the product of $f(\mathbf{x}_i | z_i, \mathbf{w}_i)$ and $f(z_i | \mathbf{w}_i)$ or the product of $f(z_i | \mathbf{x}_i, \mathbf{w}_i)$ and $f(\mathbf{x}_i | \mathbf{w}_i)$. To be consistent with our model setup in Section 2, here we break the covariate distribution $f(\mathbf{x}_i; z_i | \mathbf{w}_i)$ into two parts, $f(\mathbf{x}_i | z_i, \mathbf{w}_i)$ and $f(z_i | \mathbf{w}_i)$, and use the validation data $\{(\mathbf{x}_i, z_i, \mathbf{w}_i, \mathbf{x}_i^*, \mathbf{z}_i^*) : i \in \mathcal{V}\}$ to proceed with modeling and estimation procedures. Define

$$L_{i,\mathrm{cov}} = f(\mathbf{x}_i, z_i, \mathbf{x}_i^*, z_i^*|\mathbf{w}_i) = f(\mathbf{x}_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)f(z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)f(\mathbf{x}_i, z_i|\mathbf{w}_i),$$

and $\mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) = \partial \mathrm{log}(L_{i,\mathrm{cov}})/\partial\boldsymbol{\vartheta}$.

Estimation of $\boldsymbol{\theta}$ can proceed by maximizing $\prod_{i\in\mathcal{M}} L_i \cdot \prod_{i\in\mathcal{V}} L_{i,\mathrm{cov}}$ with respect to $\boldsymbol{\theta}$, or by jointly solving

$$\begin{pmatrix} \sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) + \sum_{i\in\mathcal{V}} \partial\mathrm{log}(L_i)/\partial\boldsymbol{\vartheta} \\ \sum_{i\in\mathcal{M}} \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta}) \end{pmatrix} = \mathbf{0}, \tag{6}$$

where $\mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \partial\mathrm{log}(L_i)/\partial\boldsymbol{\beta}$. Alternatively, one can use a pseudo-likelihood algorithm. Specifically, we first use the validation study to solve $\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) = \mathbf{0}$ for an estimator of $\boldsymbol{\vartheta}$, say, $\widehat{\boldsymbol{\vartheta}}$. Then, replacing $\boldsymbol{\vartheta}$ with the estimate $\widehat{\boldsymbol{\vartheta}}$ and then solving $\sum_{i\in\mathcal{M}} \mathbf{S}_i(\boldsymbol{\beta}, \widehat{\boldsymbol{\vartheta}}) = \mathbf{0}$ results in an estimator, denoted by $\widehat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$. Since $\mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})$ is free of the parameter $\boldsymbol{\beta}$ for $i \in \mathcal{V}$, under regularity conditions, this pseudo-likelihood procedure leads to the same estimator as that obtained by jointly solving

$$\{\textstyle\sum_{i\in\mathcal{V}}\mathbf{S}_{i,\mathrm{cov}}^{\mathrm{T}}(\boldsymbol{\vartheta}), \sum_{i\in\mathcal{M}}\mathbf{S}_i^{\mathrm{T}}(\boldsymbol{\beta}, \boldsymbol{\vartheta})\}^{\mathrm{T}} = \mathbf{0}. \tag{7}$$

The joint method based on (6) is statistically more efficient while the pseudo-likelihood procedure based on (7) is computationally easier to implement (Gong and Samaniego 1981). In the sequel, our discussion is focused on the pseudo-likelihood procedure; modifications for accommodating the joint method are straightforward.

3.3 Estimating Function Method

3.3.1 Basic Theory

We now explore a semiparametric approach to protect against possible model misspecification of $f(\mathbf{x}_i, \mathbf{z}_i|\mathbf{w}_i)$. Let $\mathbf{S}_\beta(y_i, \mathbf{x}_i, z_i, \mathbf{w}_i) = \partial\mathrm{log}\{f(y_i\ \mathbf{x}_i, z_i, \mathbf{w}_i; \boldsymbol{\beta})\}/\partial\boldsymbol{\beta}$ be the score function determined by the response model (1). If $\mathbf{X}_i$ and Z_i were observed precisely, $\boldsymbol{\beta}$ could be directly obtained by solving the sample version of $E\{\mathbf{S}_\beta(Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i; \beta)\} = 0$. Since $\mathbf{X}_i$ and Z_i are not observed and only the surrogates $\mathbf{X}_i^*$ and Z_i^* are available, we have to rely on the "observed" score function $\mathbf{U}_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) = E_{(X,Z)|(Y,X^*,Z^*,W)}\{\mathbf{S}_\beta(Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i)\}$, where the expectation $E_{(X,Z)|(Y,X^*,Z^*,W)}$ is evaluated with respect to the joint distribution of $\mathbf{X}_i$ and Z_i, given $(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$. The joint probability density function $f(\mathbf{x}_i, z_i|y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$ is

$$\frac{f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i)f(\mathbf{x}_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)f(z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)f(\mathbf{x}_i, z_i|\mathbf{w}_i)}{\int f(y_i|\mathbf{c}, t, \mathbf{w}_i)f(\mathbf{x}_i^*|\mathbf{c}, t, \mathbf{w}_i)f(z_i^*|\mathbf{c}, t, \mathbf{w}_i)f(\mathbf{c}, t|\mathbf{w}_i)d\eta(\mathbf{c})d\eta(t)}, \tag{8}$$

where $f(y_i|\mathbf{x}_i, z_i, \mathbf{w}_i)$ is determined by the response model (1), and $f(\mathbf{x}_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$ and $f(z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$ are determined by the measurement error model (3) and misclassification model (2), respectively.

We now consider the functional modeling strategy that leaves $f(\mathbf{x}_i, z_i|\mathbf{w}_i)$ unspecified. Our strategy consists of proposing a possibly misspecified model of the density function of $\mathbf{X}_i$ and Z_i, denoted $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i; \boldsymbol{\delta}^*)$, and use it as a working model. We let $f^*(\mathbf{x}_i, z_i|y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$ denote the corresponding working density function obtained from (8) except for replacing the true density $f(\mathbf{x}_i, z_i|\mathbf{w}_i)$ with the working density $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i; \boldsymbol{\delta}^*)$. Similarly, we use $E^*_{(X,Z)|(Y,X^*,Z^*,W)}$ to denote the expectation evaluated with respect to the joint working density $f^*(\mathbf{x}_i, z_i|y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$. Define

$$\mathbf{U}^*_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) = E^*_{(X,Z)|(Y,X^*,Z^*,W)}\{\mathbf{S}_\beta(Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i)\} \tag{9}$$

as the working version of $\mathbf{U}_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ under the working density $f^*(\mathbf{x}_i, z_i|y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$.

To find an unbiased estimating function for the $\boldsymbol{\beta}$ parameter, we use a projection method. The discussion in the following has similar spirit as that of Tsiatis and Ma (2004), while the development will be more complex due to the involvement of two additional processes. These two processes are required to feature a true discrete covariate Z_i and its misclassified value Z_i^*. To be specific, we assume that the working density $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i; \boldsymbol{\delta}^*)$ has the same support as that of the true density function $f(\mathbf{x}_i, z_i|\mathbf{w}_i)$. There exists a function $a(\mathbf{X}_i, Z_i, \mathbf{W}_i)$ that satisfies the identity

$$\begin{aligned} &E_{(Y,X^*,Z^*)|(X,Z,W)}[E^*_{(X,Z)|(Y,X^*,Z^*,W)}\{a(\mathbf{X}_i, Z_i, \mathbf{W}_i)\}] \\ &\qquad = E_{(Y,X^*,Z^*)|(X,Z,W)}\{\mathbf{U}^*_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)\}, \end{aligned} \tag{10}$$

where the expectation $E_{(Y,X^*,Z^*)|(X,Z,W)}$ is taken with respect to the joint density function $f(y_i, \mathbf{x}_i^*, z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$ that is determined by models (1), (2) and (3). Then an estimating function for $\boldsymbol{\beta}$ is given by

$$\mathbf{U}^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) = \mathbf{U}^*_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) - E^*_{(X,Z)|(Y,X^*,Z^*,W)}\{a(\mathbf{X}_i, Z_i, \mathbf{W}_i)\}. \tag{11}$$

Be definition of $a(\mathbf{X}_i, Z_i, \mathbf{W}_i)$, it is readily seen that this estimation function is unbiased, i.e., $E_{(Y,X^*,Z^*,W)}\{\mathbf{U}^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)\} = \mathbf{0}$. Therefore, under regularity conditions, a consistent estimate of $\boldsymbol{\beta}$ can be obtained from solving $\sum_{i\in\mathcal{M}} \mathbf{U}^*(y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i) = \mathbf{0}$.

3.3.2 Projections and the Robust Two-Step Method

The idea behind this method can be intuitively explained using the concept of "space" and "projection". If we think of an unbiased estimating function as a vector that is orthogonal to the tangent space spanned from the true model $f(y_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$ (called "the true tangent space"), then there are several ways to find an unbiased estimating function. One way is to directly project the true score function $\mathbf{S}_\beta(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ to the true tangent space and find the orthogonal residual vector. This approach usually requires the complete knowledge of the true distributions of the relevant variables. An alternative approach is to perform the projection by two steps using a working distribution as an intermediate stage. In the first step, we calculate the latent variable working score function $\mathbf{S}_\beta^*(Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i)$ based on the working model $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i)$, and subsequently construct the observed data working vector $\mathbf{U}_\beta^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ from the working density $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i)$ together with models (1),(2) and (3). In the second step, we further calculate the projection of the working score vector $\mathbf{U}_\beta^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ to the nuisance tangent space, which has the form $E^*_{(X,Z)|(Y,X^*,Z^*,W)}\{a(\mathbf{X}_i, Z_i, \mathbf{W}_i)\}$. The difference between the working score vector $\mathbf{U}_\beta^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ and its projection to the working tangent space turns out to be orthogonal to the true tangent space in this class of models. That is, the estimating function $\mathbf{U}^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is orthogonal to the true tangent space. When the working density $f^*(\mathbf{x}_i, z_i|\mathbf{w}_i; \boldsymbol{\delta}^*)$ coincides with the true density function $f(\mathbf{x}_i, z_i|\mathbf{w}_i)$, the difference vector $\mathbf{U}^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ coincides with the vector $\mathbf{U}(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ which is obtained from the direct projection approach, and hence this estimating function $\mathbf{U}^*(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ becomes semiparametric efficient (Tsiatis and Ma 2004). If Y_i is binary, semiparametric efficient estimating functions are also Fisher efficient; for epidemiological and clinical applications, this will often be the case. The appeal of the indirect projection approach lies in the relaxation of the knowledge of the true distribution $f(y_i, \mathbf{x}_i, z_i, \mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$. The only additional work is in solving (10), which, under many popular models, has a closed form solution, see Ma and Tsiatis (2006) and Ma and Ronchetti (2011).

We comment that our discussion here is suitable for the situation when models (1), (2) and (3) are correct, while the conditional distribution $f(\mathbf{x}_i, \mathbf{z}_i|\mathbf{w}_i)$ can be misspecified. Regardless of the correctness of a working model $f^*(\mathbf{x}_i, \mathbf{z}_i \mid \mathbf{w}_i)$ for the covariate process, consistency of the resulting estimators for the response parameters is always guaranteed; if the $f^*(\mathbf{x}_i, \mathbf{z}_i \mid \mathbf{w}_i)$ is correct, we can further ensure efficiency of the estimators.

Now we apply this projection method to handle estimation of the $\boldsymbol{\beta}$ parameter for our problem with a main study and a validation study. The notation is similar to that in Section 3.2 except for replacing the true density $f(\mathbf{x}_i, z_i)$ with a working density $f^*(\mathbf{x}_i, z_i; \boldsymbol{\delta}^*)$. To be specific, let $L^*_{i,cov}$ denote the counterpart of $L_{i,cov}$ in (6) with the true density $f(\mathbf{x}_i, z_i)$ replaced by the working density $f^*(\mathbf{x}_i, z_i; \boldsymbol{\delta}^*)$, $\boldsymbol{\vartheta}^* = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}, \boldsymbol{\delta}^{*\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\theta}^* = (\boldsymbol{\beta}, \boldsymbol{\vartheta}^{*\mathrm{T}})^{\mathrm{T}}$, and $\mathbf{S}^*_{i,cov}(\boldsymbol{\vartheta}^*) = \partial \log(L^*_{i,cov})/\partial\boldsymbol{\vartheta}^*$. Let $\mathbf{U}^*(\boldsymbol{\beta}, \boldsymbol{\delta}^*; Y_i, \mathbf{X}^*_i, Z^*_i, \mathbf{W}_i)$ be the estimating function determined in (11), where the function $a(\mathbf{X}_i, Z_i, \mathbf{W}_i)$ is the solution to the equation (10) which is calculated using $L^*_{i,cov}$ and (8), and $\mathbf{U}^*_\beta(Y_i, \mathbf{X}^*_i, Z^*_i, \mathbf{W}_i)$ is determined by (9). Then estimation of $\boldsymbol{\beta}$ can be carried out using a two-stage estimation algorithm. In Stage 1, solving $\sum_{i\in\mathcal{V}} \mathbf{S}^*_{i,cov}(\boldsymbol{\vartheta}^*) = \mathbf{0}$ leads to an estimator of $\boldsymbol{\vartheta}^*$, say, $\widehat{\boldsymbol{\vartheta}}^*$; in Stage 2, replace $\boldsymbol{\vartheta}^*$ with the estimate $\widehat{\boldsymbol{\vartheta}}^*$ and then solve $\sum_{i\in\mathcal{M}} \mathbf{U}^*(\boldsymbol{\beta}, \widehat{\boldsymbol{\vartheta}}^*; y_i, \mathbf{x}^*_i, z^*_i, \mathbf{w}_i) = \mathbf{0}$ for an estimator, $\widehat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$.

3.4 Robustness and Discussion for a Different Likelihood

A different but related modeling approach was taken by Spiegelman et al. (2000). We start from the joint likelihood of $(Y_i, \mathbf{X}_i, Z_i, \mathbf{X}^*_i, Z^*_i)$ given $\mathbf{W}_i$, while Spiegelman et al. (2000) start from the joint likelihood $(Y_i, \mathbf{X}_i, Z_i)$ given $(\mathbf{X}^*_i, Z^*_i, \mathbf{W}_i)$. Spiegelman et al. (2000) specify the density of $(\mathbf{X}_i, Z_i)$ given $(\mathbf{X}^*_i, Z^*_i, \mathbf{W}_i)$, which, referring to Section 3.3, we write here as $f(\mathbf{x}_i, z_i|\mathbf{X}^*_i, Z^*_i, \mathbf{W}_i, \boldsymbol{\vartheta})$. They then base estimation of $\boldsymbol{\beta}$ on the distribution of Y_i given $(\mathbf{X}^*_i, Z^*_i, \mathbf{W}_i)$. As suggested in the following theorem, our method results in a more efficient estimator for $\boldsymbol{\beta}$ than the method of Spiegelman et al. (2000) does. A proof is sketched in Appendix A.1.

Theorem 1 *Let $\widehat{\boldsymbol{\beta}}_{joint}$ be the estimator of $\boldsymbol{\beta}$ obtained from the joint likelihood of* $(Y_i, \mathbf{X}_i, Z_i, \mathbf{X}^*_i, Z^*_i)$

given $\mathbf{W}_i$*, and* $\widehat{\boldsymbol{\beta}}_{cond}$ *be the estimator of* $\boldsymbol{\beta}$ *obtained from using the conditional likelihood* $(Y_i, \mathbf{X}_i, Z_i)$ *given* $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$*. Then* $\widehat{\boldsymbol{\beta}}_{joint}$ *is asymptotically more efficient than* $\widehat{\boldsymbol{\beta}}_{cond}$*.*

We note that the approach of Spiegelman et al. (2000) is sensitive to misspecification of $f(\mathbf{x}_i, z_i|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta})$. However, it is straightforward to develop a two-stage approach similar to that in Section 3.3 that allows consistent estimation of $\boldsymbol{\beta}$ even if $f(\mathbf{x}_i, z_i|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta})$ is misspecified. Only minor changes are needed in the development in Section 3.3. Start with a working density $f^*(\mathbf{x}_i, z_i|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta})$. The working density function of $(\mathbf{X}_i, Z_i)$ given $(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is now not (8) but rather is

$$\frac{f(y_i|\mathbf{x}_i, z_i, \boldsymbol{\beta})f^*(\mathbf{x}_i, z_i|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta})}{\int f(y_i|\mathbf{c}, z_i, \boldsymbol{\beta})f^*(\mathbf{c}, t|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta})d\eta(\mathbf{c})d\eta(t)}. \tag{12}$$

Everything now is exactly the same as starting at (9), except expectations in the working model are based upon (12) rather than (8). The estimating function for $\boldsymbol{\beta}$ is still at (11) but using the working likelihood function (12), and then the two-step method discussed in Section 3.3.2 can be applied.

A major advantage of modeling $f(\mathbf{x}_i^*, z_i^*|\mathbf{x}_i, z_i, \mathbf{w}_i)$ in our approach is that this distribution is more likely to be transportable than is $f(\mathbf{x}_i, z_i|\mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$, and the distribution $f(\mathbf{x}_i^*, z_i^*, \mathbf{w}_i)$ can be estimated in the main study. As in Spiegelman et al. (2000), we can allow for a study in which the probability of selection into the validation component depends on $(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$. Our pseudo-likelihood method would follow exactly the same paradigm, while estimating function approach and the methods discussed in Section 5 would use weighting based on the probability of selection into the validation sample.

4 Asymptotic Results

In this section we establish the asymptotic results for the estimators resulted from the likelihood and estimating function methods. The proofs of the following results are sketched in the Appendix.

Theorem 2 *Assume that the ratio of the validation sample size* m *and main sample size* n *is bounded between two positive constants* c *and* C*. When the model* $f(\mathbf{x}_i, z_i \mid \mathbf{w}_i, \boldsymbol{\delta})$ *is* *correct, then the pseudo-likelihood estimator* $\widehat{\boldsymbol\beta}$ *obtained from (7) satisfies*

$$n^{1/2}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)\xrightarrow{d}\textit{Normal}(0,\boldsymbol\Sigma),\ n\to\infty,$$

where $\boldsymbol\Sigma=\mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}$*,*

$$\begin{aligned}\mathbf{A} &= E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol\beta,\boldsymbol\vartheta)}{\partial\boldsymbol\beta^{\mathrm{T}}}\right\},\\ \mathbf{B} &= \mathrm{var}\{\mathbf{S}_i(\boldsymbol\beta,\boldsymbol\vartheta)\}+(n/m)\mathbf{C}\ \mathrm{var}\{\mathbf{S}_{i,cov}(\boldsymbol\vartheta)\}\mathbf{C}^{\mathrm{T}},\ \textit{and}\\ \mathbf{C} &= E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol\beta,\boldsymbol\vartheta)}{\partial\boldsymbol\vartheta^{\mathrm{T}}}\right\}\left[E\left\{\frac{\partial \mathbf{S}_{i,cov}(\boldsymbol\vartheta)}{\partial\boldsymbol\vartheta^{\mathrm{T}}}\right\}\right]^{-1}.\end{aligned}$$

Theorem 3 *Assume the ratio of the validation sample size m and main sample size n is bounded between two positive constants c and C and the first equation of (7) is used to obtain* $\widehat{\boldsymbol\vartheta}$*. Let* $\widehat{\boldsymbol\beta}$ *be the estimator obtained from solving* $\sum_{i\in\mathcal{M}}\mathbf{U}_i^*(\boldsymbol\beta,\widehat{\boldsymbol\vartheta})=\mathbf{0}$*, where* $\mathbf{U}_i^*(\boldsymbol\beta,\widehat{\boldsymbol\vartheta})\equiv\mathbf{U}^*(y_i,\mathbf{x}_i^*,z_i^*,\mathbf{w}_i;\boldsymbol\beta,\widehat{\boldsymbol\vartheta})$*. Then regardless whether the model* $f(\mathbf{x}_i,z_i\mid\mathbf{w}_i,\boldsymbol\delta)$ *is correct or misspecified, the estimator* $\widehat{\boldsymbol\beta}$ *satisfies*

$$n^{1/2}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)\xrightarrow{d}\textit{Normal}(0,\boldsymbol\Sigma),\ n\to\infty,$$

where $\boldsymbol\Sigma=\mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}$*,*

$$\begin{aligned}\mathbf{A} &= E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol\beta,\boldsymbol\vartheta)}{\partial\boldsymbol\beta^{\mathrm{T}}}\right\},\\ \mathbf{B} &= \mathrm{var}\{\mathbf{U}_i^*(\boldsymbol\beta,\boldsymbol\vartheta)\}+(n/m)\mathbf{C}\ \mathrm{var}\{\mathbf{S}_{i,cov}(\boldsymbol\vartheta)\}\mathbf{C}^{\mathrm{T}},\ \textit{and}\\ \mathbf{C} &= E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol\beta,\boldsymbol\gamma,\alpha,\boldsymbol\delta)}{\partial(\boldsymbol\gamma^{\mathrm{T}},\alpha)}\right\}(I_p,\mathbf{0})\left[E\left\{\frac{\partial \mathbf{S}_{i,cov}(\boldsymbol\vartheta)}{\partial\boldsymbol\vartheta^{\mathrm{T}}}\right\}\right]^{-1}.\end{aligned}$$

Here p is the dimension of $(\boldsymbol\gamma^{\mathrm{T}},\alpha)^{\mathrm{T}}$*,* $\boldsymbol\vartheta=(\boldsymbol\gamma^{\mathrm{T}},\alpha,\boldsymbol\delta)^{\mathrm{T}}$ *is the true parameter value if the model* $f(\mathbf{x}_i,z_i\mid\mathbf{w}_i;\boldsymbol\delta)$ *is correct, and is the parameter that minimizes the Kullback-Leibler distance between the proposed model family and the true distribution that generated the data if the model is misspecified.*

Using the above two theorems, it is clear that while the likelihood method in Theorem 2 requires the model $f(\mathbf{x}_i,z_i\mid\mathbf{w}_i;\boldsymbol\delta)$ to be correct to yield a consistent estimator for $\boldsymbol\beta$, the estimating function method in Theorem 3 always yields consistent estimator for $\boldsymbol\beta$ whether $f(\mathbf{x}_i,z_i\mid\mathbf{w}_i;\boldsymbol\delta)$ is correct or not. We also note that the $\mathbf{C}$ matrix in Theorems 2-3 reflects the variability induced from estimation of nuisance parameters $\boldsymbol\vartheta$ using the validation data set.

5 Approximate Methods

5.1 Augmented Simulation-Extrapolation

If $\mathbf{X}_i$ and Z_i were precisely measured, inference for the parameters can be based on the likelihood function using the main study data $\mathcal{M}$, i.e., $L(\boldsymbol{\beta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\beta})$, where $L_i(\boldsymbol{\beta})$ is a probability density or mass function from the exponential family together with the regression model (1). Equivalently, under regularity conditions, the estimator is the root of the score functions $\mathbf{S}(\boldsymbol{\beta}) = \partial \log\{L(\boldsymbol{\beta})\}/\partial\boldsymbol{\beta}$.

In the presence of measurement error or misclassification, $L(\boldsymbol{\beta})$ is not computable because $(\mathbf{X}_i, Z_i)$ are unobserved. An intuitive method is to directly replace $(\mathbf{X}_i, Z_i)$ in $L(\boldsymbol{\beta})$ with the surrogate $(\mathbf{X}_i^*, Z_i^*)$. This method would, as shown in the context of measurement error alone, generally produce biased results. To correct induced biases, either completely or partially, one might be tempted to use existing methods that are developed to accommodate continuous or discrete mismeasured covariates. For example, it is appealing to develop a simulation based method by combining the simulation-extrapolation (SIMEX) method for continuous mismeasured covariates (Cook and Stefanski 1994) and the MC-SIMEX for misclassification (Küchenhoff et al. 2006).

To be specific, we consider the error model at (3), where $\mathbf{e}_i$ follows a normal distribution Normal$(\mathbf{0}, \boldsymbol{\Sigma}_e)$, and is independent of the true covariates and the response. Referring to the misclassification model (2), define $\boldsymbol{\alpha}_1 = (\alpha_{01}, \boldsymbol{\alpha}_{x1}^{\mathrm{T}}, \boldsymbol{\alpha}_{w1}^{\mathrm{T}})$ and $\boldsymbol{\alpha}_2 = (\alpha_{00}, \boldsymbol{\alpha}_{x0}^{\mathrm{T}}, \boldsymbol{\alpha}_{w0}^{\mathrm{T}})^{\mathrm{T}}$, and let $(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2)$ be their estimates computed using the validation data, so that the true and estimated misclassification probabilities are $\{p_i(\boldsymbol{\alpha}_1), q_i(\boldsymbol{\alpha}_2)\}$ and $\{p_i(\widehat{\boldsymbol{\alpha}}_1), q_i(\widehat{\boldsymbol{\alpha}}_2)\}$, respectively. Also let $\widehat{\boldsymbol{\Sigma}}_e$ be the estimate of $\boldsymbol{\Sigma}_e$ obtained from a linear regression analysis of model (3) in the validation data.

Following the simulation steps of Cook and Stefanski (1994) and Küchenhoff et al. (2006), one may create artificial surrogate measurements for $\mathbf{X}_i$ and Z_i, and then apply these measurements with other observed data to fit a model in order to portray the patterns of different error degrees on estimation; finally an estimator is obtained through extrapolating a regression model fitted to these patterns. This method can be quite time-consuming due to the intensive simulations required.

Alternatively, we propose an augmented simulation-extrapolation method. This procedure capitalizes on the unique feature associated with discrete variables, and is thus preferable. The idea works as follows. Using the discrete feature of Z_i, we first construct unbiased estimating functions to correct misclassification effects; in the second step we apply the SIMEX algorithm to these functions to further correct for measurement error effects induced in $\mathbf{X}_i^*$.

We write the score functions in (7) as $\mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i)$ by explicitly spelling out its dependence on the parameter as well as the data (Y_i, $\mathbf{X}_i$, Z_i, $\mathbf{W}_i$). Also explicitly accounting for the dependence of the misclassification probabilities on their parameters, define

$$\begin{aligned}
& \mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) \\
= \; & (1-p_i-q_i)^{-1}\Big[(1-Z_i^*)\{\mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i=0, \mathbf{W}_i)(1-p_i) - \mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i=1, \mathbf{W}_i)q_i\} \\
& -Z_i^*\{\mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i=0, \mathbf{W}_i)p_i - \mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i=1, \mathbf{W}_i)(1-q_i)\}\Big].
\end{aligned}$$

It can be shown that $E_{Z^*|Z}\{\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)\} = \mathbf{S}(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i, \mathbf{W}_i)$, where the conditional expectation is evaluated with respect to the conditional probability mass function $\mathrm{pr}(Z_i^*|Z_i, \mathbf{W}_i)$. That is, if $\mathbf{X}_i$ were not subject to measurement error, then estimating functions $\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$ can be used directly to produce a consistent estimator for the $\boldsymbol{\beta}$ parameter, because they are unbiased and computable.

Now we describe the augmented simulation-extrapolation method in detail. There are three basic steps:

1. Simulation Step: Given B (say, $B = 200$) and a sequence of M specified values $\{\lambda_1, \lambda_2, \cdots, \lambda_M\}$ with $\lambda_1 = 0$ (say, taken from $[0, 1]$), we artificially generate surrogates for $\mathbf{X}_i^*$ by adding additional noise from the measurement error and misclassification models. That is, we perform the following steps. Given $b = 1, 2, \cdots, B$, for each $\lambda = \lambda_1, \lambda_2, \cdots, \lambda_M$, generate $\mathbf{e}_{ib}$ from Normal$(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_e)$ and set $\mathbf{X}_i^*(b, \lambda)$ as $\mathbf{X}_{ib}^*(\lambda) = \mathbf{X}_i^* + \sqrt{\lambda}\mathbf{e}_{ib}$.

2. Estimation Step:

Replace $\mathbf{X}_i$ in the unbiased estimating functions $\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$ with $\mathbf{X}_i^*(b, \lambda)$, and solve $\mathbf{S}^*\{\boldsymbol{\beta}; Y_i, \mathbf{X}_{ib}^*(\lambda), Z_i^*, \mathbf{W}_i, \widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\} = \mathbf{0}$ to obtain an estimator $\widehat{\boldsymbol{\beta}}_b(\lambda)$. Define $\widehat{\boldsymbol{\beta}}(\lambda) = B^{-1} \sum_{b=1}^{B} \widehat{\boldsymbol{\beta}}_b(\lambda)$.

3. Extrapolation Step:

   For each component of $\widehat{\boldsymbol{\beta}}(\lambda)$ fit a regression model to each of the sequences $\{(\lambda, \widehat{\boldsymbol{\beta}}_r(\lambda)),$ $\lambda = \lambda_1, \lambda_2, ..., \lambda_M$ and extrapolate it to $\lambda = -1$, where $\widehat{\boldsymbol{\beta}}_r(\lambda)$ denotes the r^{th} component of $\widehat{\boldsymbol{\beta}}(\lambda)$. Let $\widehat{\boldsymbol{\beta}}_r$ denote the corresponding predicted values. Then $\widehat{\boldsymbol{\beta}}_{\text{asimex}} = (\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_{p_c})^{\mathrm{T}}$ is called the augmented-SIMEX estimator of $\boldsymbol{\beta}$, where $p_c = \dim(\boldsymbol{\beta})$.

The asymptotic theory for the Augmented SIMEX estimator $\widehat{\boldsymbol{\beta}}_{\text{asimex}}$ is given in Appendix A.4. Standard errors for $\widehat{\boldsymbol{\beta}}_{\text{asimex}}$ can be obtained using this theory, or, a computational cost, by bootstrapping.

5.2 Augmented Regression Calibration

Parallel to the augmented simulation-extrapolation above, we propose an augmented regression calibration (RC) method. By analogy with the augmented SIMEX method, we first correct for misclassification effects by using the unbiased estimating functions $\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i)$; then use standard regression calibration method to adjust for measurement error involved in $\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i)$. That is, replace $\mathbf{X}_i$ in the unbiased estimating functions $\mathbf{S}^*(\boldsymbol{\beta}; Y_i, \mathbf{X}_i, Z_i^*, \mathbf{W}_i)$ with its conditional mean $E(\mathbf{X}_i|\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$, and then solve

$$\mathbf{S}^*\{\boldsymbol{\beta}; Y_i, E(\mathbf{X}_i|\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i), Z_i^*, \mathbf{W}_i\} = \mathbf{0}$$

to obtain an augmented - RC estimator of $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}_{arc}$.

To implement this method, we need to estimate the conditional mean $E(\mathbf{X}_i|\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$, and this is done by applying standard regression procedures to the validation data in $\mathcal{V}$ using the conditional model for $\mathbf{X}_i$ on $\mathbf{W}_i$. Like the usual RC method, with linear regression or log-linear mean regression models, augmented-RC estimators are consistent; with logistic regression, the augmented-RC estimators would incur some degree of bias, although the magnitude is typically small (Spiegelman et al. 2000). Finally, the sandwich method can be employed to calculate the variance estimates for the augmented-RC estimator.

6 Empirical Studies

6.1 Simulation Studies

We performed extensive simulations to investigate the performance of the proposed methods, including the pseudo-likelihood method based on (7) and the estimating function method under both correct and misspecified latent variable distribution models. For comparison, we include the two approximate methods, augmented regression calibration and augmented SIMEX, where augmented regression calibration is also studied under both correct and misspecified latent variable distribution models. The validation sample size is set as $m = 500$ and the main study size is taken as $n = 1000$. One thousand simulations are run for each parameter configuration.

The true covariates X_i were independently generated from the uniform distribution $UNIF[-3.0, 4.0]$, and the discrete variables Z_i and W_i were independently simulated from a Bernoulli distribution with success probability 0.5. We generated X_i^* from the model $X_i^* = X_i + e_i$, where e_i is a centered normal random error with standard deviation half of that of X_i, and we generated Z_i^* from the Bernoulli distribution with the probability of misclassification 0.2 under both $Z_i = 0$ and $Z_i = 1$. These procedures were repeated m times to generate a validation sample $\{(X_i, Z_i, X_i^*, Z_i^*, W_i) : i = 1, \cdots, m\}$. To generate the data for the main study, we used the procedures above to generate n sets of covariates $\{(X_i, Z_i, X_i^*, Z_i^*, W_i) : i = 1, \cdots, n\}$, and then for each simulated true covariates (X_i, Z_i, W_i), we generated the response Y_i from the logistic regression model

$$\operatorname{logit}\{\operatorname{pr}(Y_i = 1 \mid X_i, Z_i, W_i)\} = \beta_0 + \beta_z Z_i + \beta_x X_i + \beta_w W_i, \tag{13}$$

with the true parameter values set as $\boldsymbol{\beta} = (\beta_0, \beta_z, \beta_x, \beta_w)^{\mathrm{T}} = (0.1, -1.0, 0.7, 0.5)^{\mathrm{T}}$, $i = 1, \cdots, n$. We then discarded $(X_i, Z_i), i = 1, \ldots, n$. Thus, the simulated data included a main study data $\{(Y_i, X_i^*, Z_i^*, W_i) : i = 1, \cdots, n\}$ and a separate validation sample $\{(X_i, Z_i, W_i, X_i^*, Z_i^*) : i = 1, \cdots, m\}$.

The validation data are used to fit the true model $f(x^*, z^* \mid x, z, w)$ that generated x^*, z^* to estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. In particular, the additive error model and the measurements of X_i's and X_i^*'s from the validation sample are used to estimate the measurement error variance.

The results of the seven different methods are reported in Figure 1 and Table 1, where the estimated standard errors ($\widehat{\text{sd}}$) were calculated using the results in Theorem 2 for the pseudo-likelihood estimators and those in Theorem 3 for the semiparametric estimators, with all the associated quantities evaluated at the estimated parameter values. The results for the regression calibration estimators were obtained from using Theorem 2 with X_i replaced by $E(X_i|X_i^*)$. From these results, it is clear that when the latent variable distribution model is correctly specified, the pseudo-likelihood method has the best performance in terms of both estimation bias and variability. However, as soon as this model was misspecified (here we misspecified the uniform distribution as normal), the pseudo-likelihood method showed severe bias. In contrast, the estimating function method retained a small bias regardless of model misspecification, and inference was quite precise judging from the close match between the sample and estimated standard deviations and the 95% confidence interval coverage rate and its nominal value. The two approximate methods, augmented regression calibration and augmented SIMEX, both reduced the estimation bias somewhat, but did not fully produce a consistent estimator as reflected from the nontrivial sample biases.

Our second simulation is similar to the first one, except that we now generated the latent variable from a normal distribution, and we increased the measurement error in X_i so that the standard deviation of e is about 90% of that of X_i. All other aspects of the data generation procedure remain unchanged. The corresponding results are given in Figure 2 and Table 2. As it can be clearly seen, similar conclusions can be drawn as in the first simulation.

To investigate how the performance of the proposed methods is affected by the validation sample size, we considered another scenario with the validation sample size m taken as one tenth of the main study size $n = 2000$. This ratio of m to n reflects the feature of the motivating data analyzed in Section 6.2. The simulation results are summarized in Figures 3-4 and Tables 3-4. The performance of the seven methods demonstrated the same patterns as observed previously. As expected, the results of inference are less precise mainly due to the smaller sample size of the validation study.

Based on the theoretical results as well as numerical performances, we hence recommend the estimating function method as the estimation and inference tool when both measurement error and misclassification exist in the covariates. If there is sufficient validation data to verify that a conjectured model for the latent variable distribution fits the data, then maximum pseudo-likelihood and regression calibration are also good alternatives.

In our simulation studies, we used the true parameter values as starting values for all the methods and reported the convergence values accepted by the default optimization procedure as point estimates. In addition, we experimented using the true parameter values plus a small random perturbation as starting values, and the final results were similar. This investigation pertains to the potential issue of local minimizers or multiple roots, as discussed in Section 7. Empirically, choosing sensible starting values may be helpful in real data analysis. For example, an estimate from a quick (and possibly approximate) method, such as the SIMEX or regression calibration approach, may serve as a good starting value.

6.2 Data Analysis

In this subsection, we illustrate our methods by analyzing data from the Women's Interview Study of Health (WISH) study (Brinton et al. 1995; Potischman et al. 1999). This was a case-control study in which the outcome variable Y_i is the indicator that a women (indexed by i) has breast cancer. Age and calories coming from protein and fat are potential risk factors for breast cancer. We let W_i be age. Our continuous variable X_i is the logarithm of the percentage of calories coming from protein, and the discrete variable is whether the percentage of calories coming from fat exceeds 30. The surrogates X_i^* and Z_i^* were measured by a food frequency questionnaire, and have both bias and substantial measurement error.

The main study consisted of 1,904 women for whom (Y_i, W_i, X_i^*, Z_i^*) were measured. There was also a validation study with measurements (X_i, Z_i, X_i^*, Z_i^*) for 180 subjects. These data consist of six 24-hour recalls completed one month apart, along with six days of dietary diaries from 2 sets of 3-day diaries. We treat the first dietary recall as an unbiased measure of a person's true intake. As Nusser et al. (1996) point out, "*it is well established that the characteristics of responses in a repeated survey are a function of the time in sample at which a responded is observed*". In response to this, they centered and scaled their data so that each day had the same mean and standard deviation as the first dietary recall, although unlike us, they did this in the transformed scale, and then back-transformed. The resulting 12 days of measurements were then averaged to get our definition of the true percentage of calories coming from protein and fat, and thus X_i and Z_i. For numerical stability, by subtraction and division we standardized each component of W_i, X_i and X_i^* so that they had mean zero and variance one in the validation study. Of course, the same subtraction and division was then used in the primary study. Such standardization has absolutely no impact on issues of statistical significance.

We considered the logistic regression model

$$\mathrm{logit}\{\mathrm{pr}(Y_i = 1 \mid X_i, Z_i, W_i)\} = \beta_0 + \beta_z Z_i + \beta_x X_i + \beta_w W_i.$$

In our illustration, for the misclassification and measurement error processes, we assumed that $\mathrm{pr}(Z_i^* = z_i^* | X_i, Z_i, W_i) = \mathrm{pr}(Z_i^* = z_i^* | Z_i)$ and $f(x_i^* | x_i, z_i, w_i) = f(x_i^* | x_i)$. The first assumption was reasonable based on a logistic regression of Z_i^* on Z_i, X_i, W_i, where the coefficients of X_i, W_i were both nonsignificant based on the validation data.

Similarly, the second assumption was also reasonable since a linear regression of X_i^* on X_i, Z_i, W_i yielded nonsignificant coefficients for Z_i and W_i. We denoted the misclassification probabilities $p_i = \mathrm{pr}(Z_i^* = 0 | Z_i = 1)$ and $q_i = \mathrm{pr}(Z_i^* = 1 | Z_i = 0)$, see (2). In the validation data, we estimated that $\mathrm{pr}(Z_i = 1) \approx 0.80$, $\mathrm{pr}(Z_i^* = 1) \approx 0.83$. In addition, we estimated that $\mathrm{pr}(Z_i = 1 | Z_i^* = 1) \approx 0.85$, $\mathrm{pr}(Z_i = 0 | Z_i^* = 0) \approx 0.48$, $\mathrm{pr}(Z_i^* = 1 | Z_i = 1) \approx 0.89$ and $\mathrm{pr}(Z_i^* = 0 | Z_i = 0) \approx 0.41$, all reflecting considerable misclassification. In terms of the measurement error process, we assumed a linear additive error model $X_i^* = \kappa_1 + \kappa_2 X_i + e_i$, where we estimated $\widehat{\kappa}_1 = 0.00$ and $\widehat{\kappa}_2 = 0.44$ based on the validation data. We assumed e_i to be normal with mean zero, variance σ_e^2, and independent of X_i. From the validation data, we estimated $(\kappa_1, \kappa_2, \sigma_e) = (0.00, 0.44, 0.90)$, reflecting considerable bias and measurement error in the FFQ for protein. The Kolmogorov-Smirnov test for normality based on this assumption yielded a p-value 0.976, which supports the normal error assumption. For the pseudo-likelihood method, we further assumed that X_i followed a standard normal distribution: this assumption was also supported by the Kolmogorov-Smirnov test with a p-value 0.968. To assess the impact of possible misspecification of this distribution, we also considered a case that the X_i's were assumed to follow a uniform distribution, even though this distributional assumption was not supported by the Kolmogorov-Smirnov test (the p-value is less than 0.0001).

We compared five methods. The two "naive" methods ignore the existence of measurement error and treat X_i^* as the same as X_i. One of the two naive methods takes into account the misclassification of Z_i^*, while the other even ignores the difference between Z_i^* and Z_i. Both naive estimators are carried out by performing pseudo-likelihood estimation. To correct for both measurement error and misclassification effects, we apply our methods - the pseudo-likelihood and estimating equation methods described in Section 3, and the augmented SIMEX and augmented RC methods discussed in Section 5.2. The analysis results are reported in Table 5 and are also summarized as follows.

- There were very strong corrections for measurement error. If we considered the analysis that ignored measurement error entirely, we found $(\widehat{\beta}_z, \widehat{\beta}_x) = (-0.20, -0.11)$. However, the estimating function results under either a normal or a uniform distribution for X were about $(-0.57, -0.58)$.

- The effect on the pseudo-likelihood estimator of differently specified distributions of (X_i, Z_i) given W_i was striking. It is seen that assuming normality yielded $\widehat{\beta}_x = -0.59$ with standard error 0.329, while assuming a uniform distribution yields $\widehat{\beta}_x = -0.24$ with standard error 0.127. Although it was not clear from what exact conditional distribution (X_i, Z_i) given W_i the data come, the Kolmogorov-Smirnov test provided support for the normal distribution (with p-value 0.9679) but not a uniform distribution (with p-value smaller than 0.0001). Figure 5) displays the corresponding QQ plots.

- The estimating function approach yielded almost identical estimates for β_x under either assumed normality or uniformity for X_i. However, incorrectly assuming uniformity increased the standard error estimate for β_x from 0.40 to 0.44.

Finally, we point out that the analyses we conducted here may appear not to accommodate the case-control study design. To be specific, the data we analyze were collected using a *retrospective* sampling strategy for case-control studies, but the model we use to fit the data was *prospective*. This discrepancy would, in general, make the analysis results invalid. However, under the logistic regression model, it is well-known that except for the intercept in the model, the case-control sampling design can be ignored. Indeed, fitting the data prospectively is equivalent to fitting the correct logistic model retrospectively, but with a different intercept. This equivalence was established by Prentice and Pyke (1979) for the case without covariate error, and discussed by Carroll, et al. (1993) for settings in the presence of error in covariates.

7 Discussion

In regression analysis, we often encounter covariates that are subject to both measurement error and misclassification. It is necessary to address biases induced by mismeasurement in order to carry out valid inferences. In this paper, we developed a number of functional and structural methods to handle data with a mix of measurement error and misclassification. Our methods can be applied to meet different objectives. The pseudo-likelihood method enjoys the efficiency property while the estimating function approach is attractive because of its robustness to model misspecification. The augmented SIMEX and augmented regression calibration methods are easy to implement, although they just partially correct for measurement error effects.

We note that like most estimation methods, iterative numerical algorithms, such as the Newton-Raphson or Fisher scoring schemes, are often needed to obtain estimators when implementing the proposed methods. In general, local maximizers or multiple roots may arise when implementing likelihood-based methods or estimating equations approaches. While evaluation of the likelihood function at local maximizers allows us to identify the global maximizer, choosing a suitable estimator from multiple roots of estimating equations may not be straightforward. When multiple roots occur with using the proposed estimating equations, one may follow the criteria by Heyde and Morton (1998) to discriminate the consistent estimator from multiple roots of estimating equations. More discussion on dealing with multiple roots of estimating equations can be found in Hanfelt and Liang (1995) and Heyde (1997, Section 13.2 and Section 13.3).

In this paper, we consider the main study/external validation study design. We can readily modify the proposed methods to accommodate other settings as well, such as a validation sample that is either internal or external, as discussed by Guo and Little (2011).

In contrast to our methods, Wang et al. (2008) explored the use of expected estimating equations to handle data with measurement error and misclassification. Instead of assuming the availability of a validation sample, Wang et al. (2008) investigated the situation with repeated surrogate measurements taken for associated response or covariate variables. The methods developed by Wang et al. (2008) emphasize the evaluation of conditional expectations of relevant quantities, and this requires specification of a distributional assumption for the true covariates, but our methods provide tools which apply to both settings where distributions of the true covariates are known as well as situations where distributions of the true covariates are left unspecified.

Measurement error or misclassification are ubiquitous in practice, and most available work deals with one of the two features separately but not both simultaneously. In this paper, we directed our attention to the common problem that measurement error and misclassification exist concurrently in the data analysis. We developed a rich class of methods to handle data with both covariate measurement error and misclassification under general model frameworks. Our investigation covers both functional and structural modeling strategies for measurement error and misclassification processes. Our methods can be applied to different situations to meet various objectives.

Acknowledgments

The authors thank the review team for their helpful comments. Yi's research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Ma's research was supported by grants from the National Science Foundation (DMS-1000354, DMS-1206693) and the National Institute of Neurological Disorder and Stroke (R01-073671). Spiegelman's research was supported by grants from NIH/NIEHS (R01 ES 09411) and NIH/NCI (R01-CA050597). Carroll's research was supported by grants from the National Cancer Institute (R37-CA057030), the National Institute of Neurological Disorder and Stroke (R01-073671) and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

Appendix

A.1 Efficiency comparison of the two likelihood methods

Note that the observed data from the validation and main studies are $\{(\mathbf{X}_i, Z_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) : i \in \mathcal{V}\}$ and $\{(Y_i, \mathbf{X}_i^*, Z_i^*, \mathbf{W}_i) : i \in \mathcal{M}\}$, respectively. We compute the likelihood of the observed data given the $\mathbf{W}_i$ as follows

$$\mathcal{L}_{\text{joint}}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \prod_{i \in \mathcal{V}} f(\mathbf{x}_i, z_i, \mathbf{x}_i^*, z_i^* | \ \mathbf{w}_i, \boldsymbol{\vartheta}) \times \prod_{j \in \mathcal{M}} f(y_j, \mathbf{x}_j^*, z_j^* | \ \mathbf{w}_j, \boldsymbol{\beta}, \boldsymbol{\vartheta}). \tag{A.1}$$

On the other hand, Spiegelman, et al. (2000) compute the likelihood function for the data of the validation and main studies by conditioning on $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$. Specifically, the likelihood of $(\mathbf{X}_i, Z_i)$ conditional on $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is

$$f(\mathbf{x}_i, z_i | \ \mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta}) = \frac{f(\mathbf{x}_i^*, z_i^*, \mathbf{x}_i, z_i | \ \mathbf{w}_i, \boldsymbol{\vartheta})}{\int\int f(\mathbf{x}_i^*, z_i^* | \ \mathbf{s}, t, \mathbf{w}_i, \boldsymbol{\alpha}, \boldsymbol{\gamma}) f(\mathbf{s}, t | \ \mathbf{w}_i, \boldsymbol{\delta}) d\eta(\mathbf{s}) d\eta(t)},$$

and the likelihood of Y_i given $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is

$$f(y_i | \ \mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\beta}, \boldsymbol{\vartheta}) = \frac{f(y_i, \mathbf{x}_i^*, z_i^* | \ \mathbf{w}_i, \boldsymbol{\beta}, \boldsymbol{\vartheta})}{\int\int f(\mathbf{x}_i^*, z_i^* | \ \mathbf{s}, t, \mathbf{w}_i, \boldsymbol{\alpha}, \boldsymbol{\gamma}) f(\mathbf{s}, t | \ \mathbf{w}_i, \boldsymbol{\delta}) d\eta(\mathbf{s}) d\eta(t)}.$$

Thus, the observed data likelihood given all the $(\mathbf{X}_i^*, Z_i^*, \mathbf{W}_i)$ is

$$\begin{aligned} \mathcal{L}_{\text{cond}}(\boldsymbol{\theta}) &= \prod_{i \in \mathcal{V}} f(\mathbf{x}_i, z_i | \ \mathbf{x}_i^*, z_i^*, \mathbf{w}_i, \boldsymbol{\vartheta}) \prod_{j \in \mathcal{M}} f(y_j | \ \mathbf{x}_j^*, z_j^*, \mathbf{w}_j, \boldsymbol{\beta}, \boldsymbol{\vartheta}) \\ &= \mathcal{L}_{\text{joint}}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) / \mathcal{M}(\boldsymbol{\vartheta}), \end{aligned} \tag{A.2}$$

where

$$\begin{aligned} \mathcal{M}(\boldsymbol{\vartheta}) &= \prod_{i \in \mathcal{V}} \int\int f(\mathbf{x}_i^*, z_i^* | \ \mathbf{s}, t, \mathbf{w}_i, \boldsymbol{\alpha}, \boldsymbol{\gamma}) f(\mathbf{s}, t | \ \mathbf{w}_i, \boldsymbol{\delta}) d\eta(\mathbf{s}) d\eta(t) \\ &\quad \times \prod_{j \in \mathcal{M}} \int\int f(\mathbf{x}_j^*, z_j^* | \ \mathbf{s}, t, \mathbf{w}_j, \boldsymbol{\alpha}, \boldsymbol{\gamma}) f(\mathbf{s}, t | \ \mathbf{w}_j, \boldsymbol{\delta}) d\eta(\mathbf{s}) d\eta(t), \end{aligned}$$

which is the likelihood of the $(\mathbf{X}_i^*, Z_i^*)$ given the $\mathbf{W}_i$ for all the data of the validation and main studies.

Let $\widehat{\boldsymbol{\theta}}_{joint} = (\widehat{\boldsymbol{\beta}}_{joint}^{\mathrm{T}}, \widehat{\boldsymbol{\vartheta}}_{joint}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}}_{cond} = (\widehat{\boldsymbol{\beta}}_{cond}^{\mathrm{T}}, \widehat{\boldsymbol{\vartheta}}_{cond}^{\mathrm{T}})^{\mathrm{T}}$ be the estimators of $\boldsymbol{\theta}$ that are obtained by maximizing (A.1) and (A.2), respectively. By likelihood theory, as $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{cond} - \boldsymbol{\theta})$ has a normal distribution with mean zero and covariance matrix J_{cond}^{-1}, where $J_{cond} = \lim_{n \to \infty} n^{-1} E\{-\partial^2 \text{log}(\mathcal{L}_{\text{cond}}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}\}$. After some algebra, the inverse of the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{cond} - \boldsymbol{\beta})$ is

$$[\text{asyvar}\{\sqrt{n}(\widehat{\boldsymbol{\beta}}_{cond} - \boldsymbol{\beta})\}]^{-1} = J_{\beta\beta}^{cond} - J_{\beta\beta}^{cond}(J_{\vartheta\vartheta}^{cond})^{-1}(J_{\beta\vartheta}^{cond})^{\mathrm{T}}, \tag{A.3}$$

where

$$J_{\beta\beta}^{cond} = \lim_{n \to \infty} n^{-1} E\{-\partial^2 \text{log}(\mathcal{L}_{\text{cond}}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\mathrm{T}}\},$$

$$J^{cond}_{\beta\vartheta} = \lim_{n\to\infty} n^{-1}E\{-\partial^2\mathrm{log}(\mathcal{L}_{\mathrm{cond}})/\partial\boldsymbol{\beta}\partial\boldsymbol{\vartheta}^{\mathrm{T}}\},$$

and

$$J^{cond}_{\vartheta\vartheta} = \lim_{n\to\infty} n^{-1}E\{-\partial^2\mathrm{log}(\mathcal{L}_{\mathrm{cond}})/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^{\mathrm{T}}\}.$$

By analogy and the relationship (A.2),the inverse of the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{joint} - \boldsymbol{\beta})$ is given by

$$[\mathrm{asyvar}(\sqrt{n}(\widehat{\boldsymbol{\beta}}_{joint} - \boldsymbol{\beta})\}]^{-1} = J^{cond}_{\beta\beta} - J^{cond}_{\beta\beta}(J^{cond}_{\vartheta\vartheta} + J_{\vartheta\vartheta})^{-1}(J^{cond}_{\beta\vartheta})^{\mathrm{T}}, \qquad (A.4)$$

where $J_{\vartheta\vartheta} = \lim_{n\to\infty} n^{-1}E\{-\partial^2\mathrm{log}\mathcal{M}(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^{\mathrm{T}}\}$. Comparing (A.3) and (A.4) implies that $\widehat{\beta}_{joint}$ is more efficient than $\widehat{\beta}_{cond}$.

A.2 Sketched proof of Theorem 2

Assume we have m observations in the validation data set and n observations in the main data set. Using (7), we obtain

$$\mathbf{0} = m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\widehat{\boldsymbol{\vartheta}}) = m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\} m^{1/2}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) + o_p(1).$$

or

$$m^{1/2}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) = -\left[E\left\{\frac{\partial \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\}\right]^{-1} m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) + o_p(1).$$

We also have

$$\begin{aligned}
\mathbf{0} &= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{S}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\vartheta}}) \\
&= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \\
&\qquad + E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1) \\
&= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&\quad - E\left\{\frac{\partial \mathbf{S}_i(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\} \sqrt{n/m}\left[E\left\{\frac{\partial \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\}\right]^{-1} m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta}) \\
&\qquad + o_p(1 + \sqrt{n/m}).
\end{aligned}$$

This yields the desired results.

A.3 Sketched proof of Theorem 3

Using the result in White (1982) and the proof of Theorem 2, we have

$$m^{1/2}(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}) = -\left[E\left\{\frac{\partial \mathbf{S}_{i,\text{cov}}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^{\text{T}}}\right\}\right]^{-1} m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\text{cov}}(\boldsymbol{\vartheta}) + o_p(1),$$

where $\boldsymbol{\vartheta}$ is the true parameter value if the model is correct and is the parameter that minimizes the Kullback-Leibler distance between the proposed model family and the true distribution that generated the data if the model is incorrect. Since the dimension of $(\boldsymbol{\gamma}^{\text{T}},\alpha)^{\text{T}}$ as p, we have

$$m^{1/2}\left\{\begin{pmatrix}\widehat{\boldsymbol{\gamma}}\\ \widehat{\alpha}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\gamma}\\ \alpha\end{pmatrix}\right\} = -\,(I_p,\mathbf{0})\left[E\left\{\frac{\partial \mathbf{S}_{i,\text{cov}}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^{\text{T}}}\right\}\right]^{-1} m^{-1/2}\sum_{i\in\mathcal{V}} \mathbf{S}_{i,\text{cov}}(\boldsymbol{\vartheta}) + o_p(1).$$

Using the estimating equation, we have

$$\begin{aligned}
\mathbf{0} &= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{U}_i^*(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\vartheta}})\\
&= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}^{\text{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\
&\qquad + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^{\text{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}) + o_p(1)\\
&= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}^{\text{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\
&\qquad + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^{\text{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})\\
&\qquad + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,\boldsymbol{\delta})}{\partial (\boldsymbol{\gamma}^{\text{T}},\alpha)}\right\} n^{1/2}\left\{\begin{pmatrix}\widehat{\boldsymbol{\gamma}}\\ \widehat{\alpha}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\gamma}\\ \alpha\end{pmatrix}\right\} + o_p(1).
\end{aligned}$$

The construction of $\mathbf{U}_i^*$ insures

$$E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^{\text{T}}}\right\} = \mathbf{0}$$

because $\mathbf{U}^*$ is in the space orthogonal to the nuisance tangent space spanned by the score functions with respect to the parameters involved in the model $f_{\mathbf{X},Z|\mathbf{W}}(\mathbf{x},z\mid\mathbf{w})$. Thus, we have

$$\begin{aligned}
&\mathbf{0}\\
&= n^{-1/2}\sum_{i\in\mathcal{M}} \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta}) + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}^{\text{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\
&\qquad + E\left\{\frac{\partial \mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,\boldsymbol{\delta})}{\partial (\boldsymbol{\gamma}^{\text{T}},\alpha)}\right\} n^{1/2}\left\{\begin{pmatrix}\widehat{\boldsymbol{\gamma}}\\ \widehat{\alpha}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\gamma}\\ \alpha\end{pmatrix}\right\} + o_p(1)
\end{aligned}$$

$$= n^{-1/2}\sum_{i\in\mathcal{M}}\mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta}) + E\left\{\frac{\partial\mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\vartheta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$
$$-(n/m)^{1/2}E\left\{\frac{\partial\mathbf{U}_i^*(\boldsymbol{\beta},\boldsymbol{\gamma},\alpha,\boldsymbol{\delta})}{\partial(\boldsymbol{\gamma}^{\mathrm{T}},\alpha)}\right\}(I_p,\mathbf{0})\left[E\left\{\frac{\partial\mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^{\mathrm{T}}}\right\}\right]^{-1} m^{-1/2}\sum_{i\in\mathcal{V}}\mathbf{S}_{i,\mathrm{cov}}(\boldsymbol{\vartheta})$$
$$+o_p(1+\sqrt{n/m}),$$

completing the argument.

A.4 Asymptotic Theory for Augmented SIMEX

Let vec$(\boldsymbol{\Sigma}_e)$ be the vector of unique elements of a matrix $\boldsymbol{\Sigma}_e$ and denote the symmetric square root of $\boldsymbol{\Sigma}_e$ as $g_{\mathrm{ssr}}\{\mathrm{vec}(\boldsymbol{\Sigma}_e)\}$.

Define $\mathcal{X}_i = (1,\mathbf{X}_i^{\mathrm{T}},\mathbf{W}_i^{\mathrm{T}})^{\mathrm{T}}$, and recall that $\boldsymbol{\alpha}_1 = (\alpha_{01},\boldsymbol{\alpha}_{x1}^{\mathrm{T}},\boldsymbol{\alpha}_{w1}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\alpha}_2 = (\alpha_{00},\boldsymbol{\alpha}_{x0}^{\mathrm{T}},\boldsymbol{\alpha}_{w0}^{\mathrm{T}})^{\mathrm{T}}$. Define $H^{(1)}(x) = H(x)\{1-H(x)\}$, the derivative of $H(x) = \exp(x)/\{1+\exp(x)\}$, and $\mathcal{C}_k = E\{\mathcal{X}_i\mathcal{X}_i^{\mathrm{T}}H^{(1)}(\mathcal{X}_i^{\mathrm{T}}\boldsymbol{\alpha}_k)\}$ for $k=1,2$. Then it is readily seen that for $k=1,2$,

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1) = (n/m)^{1/2}\mathcal{C}_1^{-1}m^{-1/2}\sum_{i=1,i\in\mathcal{V}}^{m}\left\{Z_i^*-1-H(\mathcal{X}_i^{\mathrm{T}}\boldsymbol{\alpha}_1)\right\}+o_p(1); \quad (\mathrm{A.5})$$
$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_2-\boldsymbol{\alpha}_2) = (n/m)^{1/2}\mathcal{C}_2^{-1}m^{-1/2}\sum_{i=1,i\in\mathcal{V}}^{m}\left\{Z_i^*-H(\mathcal{X}_i^{\mathrm{T}}\boldsymbol{\alpha}_2)\right\}+o_p(1). \quad (\mathrm{A.6})$$

For the error model (3), we use linear regression in the validation data to estimate $\{\Gamma_x,\gamma_z,\Gamma_w\}$, and let $\{\widehat{\Gamma}_x,\widehat{\gamma}_z,\widehat{\Gamma}_w\}$ denote the resulting estimate. Remembering that $\mathbf{e}_i = \mathbf{X}_i^*-\Gamma_x\mathbf{X}_i-\boldsymbol{\gamma}_zZ_i-\Gamma_w\mathbf{W}_i$, we write $\widehat{\mathbf{e}}_i = \mathbf{X}_i^*-\widehat{\Gamma}_x\mathbf{X}_i-\widehat{\boldsymbol{\gamma}}_zZ_i-\widehat{\Gamma}_w\mathbf{W}_i$. Then we estimate $\boldsymbol{\Sigma}_e$ by

$$\widehat{\boldsymbol{\Sigma}}_e = (m-p_c)^{-1}\sum_{i=1,i\in\mathcal{V}}^{m}\widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i^{\mathrm{T}},$$

where p_c is the dimension of $(\mathbf{X}_i,Z_i,\mathbf{W}_i)$. Because the residuals are uncorrelated with the predictors, it is obvious that

$$n^{1/2}\{\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_e)-\mathrm{vec}(\boldsymbol{\Sigma}_e)\} = (n/m)^{1/2}m^{-1/2}\sum_{i=1,i\in\mathcal{V}}^{m}\{\mathrm{vec}(\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}})-\mathrm{vec}(\boldsymbol{\Sigma}_e)\}+o_p(1). \quad (\mathrm{A.7})$$

The limit (A.7) does not involve the estimated regression parameters in model (3).

It is convenient to rewrite the Augmented SIMEX procedure as follows. Generate $\mathbf{e}_{ib*}\sim$ Normal$(\mathbf{0},\mathbf{I})$. Then $\widehat{\boldsymbol{\beta}}_b(\lambda)$ is the solution to

$$\mathbf{0} = n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathbf{S}^*[\widehat{\boldsymbol{\beta}}_b(\lambda),Y_i,\mathbf{X}_i^*+\sqrt{\lambda}g_{ssr}\{\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_e)\}\mathbf{e}_{ib*},Z_i^*,\mathbf{W}_i,\widehat{\boldsymbol{\alpha}}_1,\widehat{\boldsymbol{\alpha}}_2].$$

It is clear that $\widehat{\boldsymbol{\beta}}_b(\lambda) = \boldsymbol{\beta}_b(\lambda)+o_p(1)$, where $\boldsymbol{\beta}_b(\lambda)$ is the solution to

$$\mathbf{0} = E\left(\mathbf{S}^*[\boldsymbol{\beta}_b(\lambda),Y_i,\mathbf{X}_i^*+\sqrt{\lambda}g_{\mathrm{ssr}}\{\mathrm{vec}(\boldsymbol{\Sigma}_e)\}\mathbf{e}_{ib*},Z_i^*,\mathbf{W}_i,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2]\right).$$

Define

$$\mathbf{U}_{ib}(\lambda)=\mathbf{S}^*[\boldsymbol{\beta}_b(\lambda),Y_i,\mathbf{X}_i^*+\sqrt{\lambda}g_{\rm ssr}\{\mathrm{vec}(\boldsymbol{\Sigma}_e)\}\mathbf{e}_{ib*},Z_i^*,\mathbf{W}_i,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2],$$

and also define

$$\begin{aligned}\mathcal{F}_{\beta}(\lambda)&=-E\left(\partial\mathbf{U}_{ib}(\lambda)/\partial\boldsymbol{\beta}^{\rm T}\right);\\ \mathcal{F}_{\Sigma}(\lambda)&=E\left(\partial\mathbf{U}_{ib}(\lambda)/\partial\mathrm{vec}^{\rm T}(\boldsymbol{\Sigma}_e)\right);\\ \mathcal{F}_{\alpha,1}(\lambda)&=E\left(\partial\mathbf{U}_{ib}(\lambda)/\partial\boldsymbol{\alpha}_1^{\rm T}\right);\\ \mathcal{F}_{\alpha,2}(\lambda)&=E\left(\partial\mathbf{U}_{ib}(\lambda)/\partial\boldsymbol{\alpha}_2^{\rm T}\right).\end{aligned}$$

Then, by standard estimating equation calculations, we see that

$$\begin{aligned}n^{1/2}\mathcal{F}_{\beta}(\lambda)\{\widehat{\boldsymbol{\beta}}_b(\lambda)-\boldsymbol{\beta}_b(\lambda)\}&=n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathbf{U}_{ib}(\lambda)+\mathcal{F}_{\Sigma}(\lambda)n^{1/2}\{\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_e)-\mathrm{vec}(\boldsymbol{\Sigma}_e)\}\\ &\quad+\mathcal{F}_{\alpha,1}(\lambda)n^{1/2}(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1)+\mathcal{F}_{\alpha,2}(\lambda)n^{1/2}(\widehat{\boldsymbol{\alpha}}_2-\boldsymbol{\alpha}_2)+o_p(1).\end{aligned}$$

Using (A.5)-(A.7), define

$$\begin{aligned}\mathbf{V}_{ib}(\lambda)&=\mathcal{F}_{\Sigma}(\lambda)\{\mathrm{vec}(\mathbf{e}_i\mathbf{e}_i^{\rm T})-\mathrm{vec}(\boldsymbol{\Sigma}_e)\}+\mathcal{F}_{\alpha,1}(\lambda)\mathcal{C}_1^{-1}\left\{Z_i^*-1-H(\mathcal{X}_i^{\rm T}\boldsymbol{\alpha}_1)\right\}\\ &\quad+\mathcal{F}_{\alpha,2}(\lambda)\mathcal{C}_2^{-1}\left\{Z_i^*-H(\mathcal{X}_i^{\rm T}\boldsymbol{\alpha}_2)\right\}.\end{aligned}$$

Then we have that

$$\begin{aligned}n^{1/2}\{\widehat{\boldsymbol{\beta}}_b(\lambda)-\boldsymbol{\beta}_b(\lambda)\}&=\mathcal{F}_{\beta}^{-1}(\lambda)n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathbf{U}_{ib}(\lambda)\\ &\quad+(n/m)^{1/2}\mathcal{F}_{\beta}^{-1}(\lambda)n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathbf{V}_{ib}(\lambda)+o_p(1).\end{aligned}\tag{A.8}$$

If we define $\widetilde{\mathbf{U}}_i(\lambda)=\mathcal{F}_{\beta}^{-1}(\lambda)B^{-1}\sum_{b=1}^{B}\mathbf{U}_{ib}(\lambda)$, $\widetilde{\mathbf{V}}_i(\lambda)=\mathcal{F}_{\beta}^{-1}(\lambda)B^{-1}\sum_{b=1}^{B}\mathbf{V}_{ib}(\lambda)$, and $\boldsymbol{\beta}(\lambda)=B^{-1}\sum_{b=1}^{B}\boldsymbol{\beta}_b(\lambda)$, we have shown that

$$\begin{aligned}n^{1/2}\{\widehat{\boldsymbol{\beta}}(\lambda)-\boldsymbol{\beta}(\lambda)\}&=n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\widetilde{\mathbf{U}}_i(\lambda)\\ &\quad+(n/m)^{1/2}n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\widetilde{\mathbf{V}}_i(\lambda)+o_p(1).\end{aligned}\tag{A.9}$$

There is a known function $g_{\rm asimex}(\cdot)$, which is explicit in the case of polynomial extrapolation, such that $\widehat{\boldsymbol{\beta}}_{\rm asimex}=\mathbf{g}_{\rm asimex}\{\widehat{\boldsymbol{\beta}}(\lambda_1),...,\widehat{\boldsymbol{\beta}}(\lambda_M)\}$. Let

$$\mathbf{g}_{j,\rm asimex}=\partial\mathbf{g}_{\rm asimex}\{\boldsymbol{\beta}(\lambda_1),...,\boldsymbol{\beta}(\lambda_M)\}/\partial\boldsymbol{\beta}^{\rm T}(\lambda_j).$$

Then, by the delta-method,

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{\rm asimex}-\boldsymbol{\beta}_{\rm asimex})=\textstyle\sum_{j=1}^{M}\mathbf{g}_{j,\rm asimex}n^{1/2}\{\widehat{\boldsymbol{\beta}}(\lambda_j)-\boldsymbol{\beta}(\lambda_j)\}+o_p(1),$$

where $\boldsymbol{\beta}_{\rm asimex}=\{\boldsymbol{\beta}^{\rm T}(\lambda_1),\cdots,\boldsymbol{\beta}^{\rm T}(\lambda_M)\}^{\rm T}$.

Define

$$\mathcal{G}_i=\textstyle\sum_{j=1}^{M}\mathbf{g}_{j,\rm asimex}\widetilde{\mathbf{U}}_i(\lambda_j);\qquad\mathcal{H}_i=\textstyle\sum_{j=1}^{M}\mathbf{g}_{j,\rm asimex}\widetilde{\mathbf{V}}_i(\lambda_j).$$

Using (A.9), this means that

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\beta}}_{\rm asimex}-\boldsymbol{\beta}_{\rm asimex})&=n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathcal{G}_i+(n/m)^{1/2}n^{-1/2}\textstyle\sum_{i=1,i\in\mathcal{M}}^{n}\mathcal{H}_i\\ &\quad+o_p(1).\end{aligned}\tag{A.10}$$

By the central Limit Theorem, (A.10) converges in distribution to Normal$(\mathbf{0},\boldsymbol{\Sigma}_{\rm asimex})$ as $n\to\infty$, where $\boldsymbol{\Sigma}_{\rm asimex}=\mathrm{cov}(\mathcal{G}_i)+\rho\mathrm{cov}(\mathcal{H}_i)$, and $\rho=\lim_{n\to\infty}(n/m)$. The limiting covariance matrix $\boldsymbol{\Sigma}_{\rm asimex}$ can be estimated by replacing all population terms by their sample versions to form $\widehat{\mathcal{G}}_i$ and $\widehat{\mathcal{H}}_i$.

References

Akazawa, K., Kinukawa, K. and Nakamura, T. (1998). A note on the corrected score function adjusting for misclassification. *Journal of the Japanese Statistical Society*, 28, 115-123.

Brinton, L. A., Daling, J. R., Liff, J. M., Schoenberg, J. B., Malone, K. E., Stanford, J. L., Coates, R. J., Gammon, M. D., Hanson, L. and Hoover, R. N. (1995). Oral contraceptives and breast cancer risk among younger women. *Journal of the National Cancer Institute*, 87, 827-835.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications.* Chapman & Hall/CRC.

Buonaccorsi, J. P., Laake, P. and Veierod, M. (2005). A note on the effect of misclassification on bias of perfectly measured covariates in regression. *Biometrics*, 61, 831-836.

Carroll, R. J., Gail, M. H., and Lubin, J. H. (1993). Case-control studies with errors in covariates. *Journal of the American Statistical Association*, 88, 185-199.

Carroll, R. J., Lombard, F., Küchenhoff, H. and Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in structural measurement error models. *Journal of the American Statistical Association*, 91, 242-250.

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

Carroll, R. J. and Wand, M. P. (1990). Semi-parametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society, Series B*, 53, 573-587.

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.

Dalen, I., Buonaccorsi, J. P., Sexton, J., Laake, P., and Thoresen, M. (2009). Correcting for misclassification of a categorized exposure in binary regression using replication data. *Statistics in Medicine*, 28, 3368-3410.

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

Gong, G. and Samaniego, F. G. (1981). Pseudo maximum likelihood estimation: Theory and application. *The Annals of Statistics*, 9, 861-869.

Guo, Y. and Little, R. J. (2011). Regression analysis with covariates that have heteroscedastic measurement error. *Statistics in Medicine*, published online: May 17, 2011.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology.* Chapman & Hall/CRC.

Hall, P. and Ma, Y. (2007). Measurement error models with unknown error structure. *Journal of the Royal Statistical Society, Series B*, 69, 429-446.

Hanfelt, J. J. and Liang, K.-Y. (1995). Approximate likelihood ratios for general estimating functions. *Biometrika*, 82, 461-477.

Heyde, C. C. (1997). *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.* Springer-Verlag New York, Inc.

Heyde, C. C. and Morton, R. (1998). Multiple roots in general estimating equations. *Biometrika*, 85, 954-959.

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*, 95, 1209-1219.

Küchenhoff, H. (1990). *Logit- und Probit regression mit Fehlen in den Variabeln.* Frankfurt am Main: Anton Hain.

Küchenhoff, H., Mwalili, S. M. and Lesaffre, E. (2006). A general method for dealing with misclassification in regression: The misclassification SIMEX. *Biometrics*, 62, 85-96.

Liang, H. (2009). Generalized partially linear mixed-effects models incorporating mismeasured covariates. *Annals of the Institute of Statistical Mathematics*, 61, 27-46.

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

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95, 520-534.

Ma, Y. and Ronchetti, E. (2011). Saddlepoint test in measurement error models. *Journal of the American Statistical Association*, 106, 147-156.

Ma, Y. and Tsiatis, A. A. (2006). Closed form semiparametric estimators for measurement error models. *Statistica Sinica*, 16, 183-193.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.

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

Nusser, S. M., Carriquiry, A. L., Dodd, K. W. and Fuller, W. A. (1996). A semiparametric transformation approach to estimating usual intake distributions. *Journal of the American Statistical Association*, 91, 1440–1449.

Pierce, D. A, Stram, D. O., Vaeth , M. and Schafer, D. (1992). Some insights into the errors in variables problem provided by consideration of radiation dose-response analyses for the A-bomb survivors. *Journal of the American Statistical Association*, 87, 351-359.

Potischman, N., Carroll, R. J., Iturria, S., Mittl, B., Curtin, J., Thompson, F. and Brinton, L. (1999). Comparison of the 60- and 100-item NCI-block questionnaires with validation data. *Nutrition and Cancer*, 34, 70-85.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403-411.

Rosner, B., Spiegelman, D., and Willett, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*, 132, 734-745.

Rosner, B., Spiegelman, D., and Willett, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology*, 136, 1400-1413.

Rosner, B. A., Willett, W. C. and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8, 1051-1070.

Spiegelman, D., Rosner, B. and Logan, R. (2000). Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association*, 95, 51-61.

Spiegelman, D., Zhao, B. and Kim, J. (2005). Correlated errors in biased surrogates: study designs and methods for measurement error correction. *Statistics in Medicine*, 24, 1657-1682.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika*, 74, 703-716.

Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics*, 59, 1140–1150.

Sugar, E. A., Wang, C.-Y. and Prentice, R. L. (2007). Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics*, 63, 143-151.

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

Wang, C. Y., Huang, Y., Chao, E. C. and Jeffcoat, M. K. (2008). Expected estimating equations for missing data, measurement error, and misclassification, with application to longitudinal nonignorable missing data. *Biometrics*, 64, 85-95.

Wang, N. and Davidian, M. (1996). A note on covariate measurement error in nonlinear mixed models. *Biometrics*, 83, 801-812.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-25.

Yi, G. Y. (2008). A simulation-based marginal method for longitudinal data with dropout and mismeasured covariates. *Biostatistics*, 9, 501-512.

Yi, G. Y., Liu, W. and Wu, L. (2011). Simultaneous inference and bias analysis for longitudinal data with covariate measurement error and missing responses. *Biometrics*, 67, 67-75.

Yi, G. Y., Ma, Y. and Carroll, R. J. (2012). A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. *Biometrika*, 99, 151-165.

Zucker, D. M. and Spiegelman, D. (2004). Inference for the proportional hazards model with misclassified discrete-valued covariates. *Biometrics*, 60, 324-334.

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

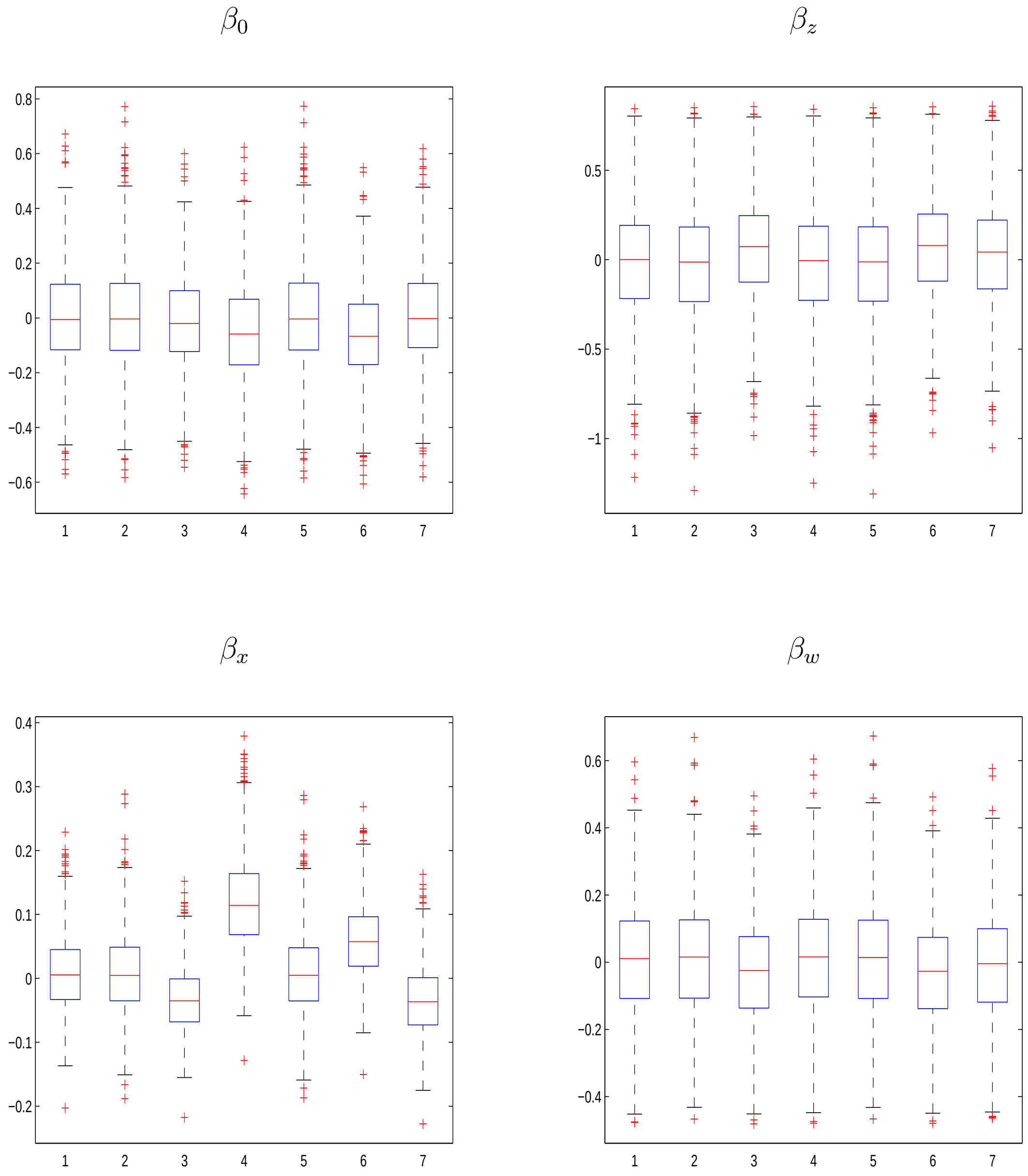

Figure 1: Boxplots of the biases of the seven estimators for β_0, β_z, β_x and β_w in Simulation 1. The seven estimators are respectively pseudo-likelihood (1) , estimating function (2) and regression calibration (3) estimators under uniform distribution model for X, pseudo-likelihood (4), estimating function (5) and regression calibration (6) estimators under normal distribution model for X, and SIMEX estimator (7).

Table 1: Results of Simulation 1 in Section 6.1 based on 1,000 data sets, $m = 500, n = 1000$ and X is normal. Here $(\beta_0, \beta_z, \beta_x, \beta_w)$ are defined in (13). Mean (est), standard deviation (sd), the average of the estimated standard deviation ($\widehat{\text{sd}}$) and 95% confidence interval coverage are reported for likelihood methods, estimating function methods, regression calibration methods, all based on both uniform and normal latent variable distribution models. SIMEX estimation results are reported in the left last block. The true latent variable distribution is uniform.

	β_0	β_z	β_x	β_w	β_0	β_z	β_x	β_w
true	0.1	-1.0	0.7	0.5	0.1	-1.0	0.7	0.5
	maximum likelihood							
	uniform				normal			
est	0.101	-1.012	0.707	0.511	0.049	-1.019	0.818	0.516
sd	0.184	0.303	0.060	0.170	0.184	0.305	0.073	0.172
$\widehat{\text{sd}}$	0.183	0.299	0.060	0.166	0.184	0.300	0.073	0.168
95%CI	95.2%	94.8%	95.7%	95.4%	94.1%	94.9%	67.3%	95.4%
	estimating function							
	uniform				normal			
est	0.107	-1.027	0.709	0.511	0.107	-1.027	0.709	0.511
sd	0.194	0.313	0.064	0.176	0.194	0.314	0.066	0.176
$\widehat{\text{sd}}$	0.185	0.300	0.063	0.174	0.185	0.300	0.063	0.174
95%CI	93.8%	94.5%	95.6%	95.8%	93.8%	94.4%	95.2%	95.9%
	regression calibration							
	uniform				normal			
est	0.088	-0.936	0.666	0.472	0.039	-0.930	0.759	0.472
sd	0.170	0.277	0.050	0.157	0.168	0.274	0.058	0.157
$\widehat{\text{sd}}$	0.169	0.273	0.050	0.153	0.167	0.270	0.057	0.153
95%CI	95.2%	93.4%	88.3%	94.9%	93.2%	93.3%	84.4%	94.8%
	simulation extrapolation							
est	0.103	-0.970	0.665	0.495	0.103	-0.970	0.665	0.495
sd	0.178	0.287	0.055	0.165	0.178	0.287	0.055	0.165

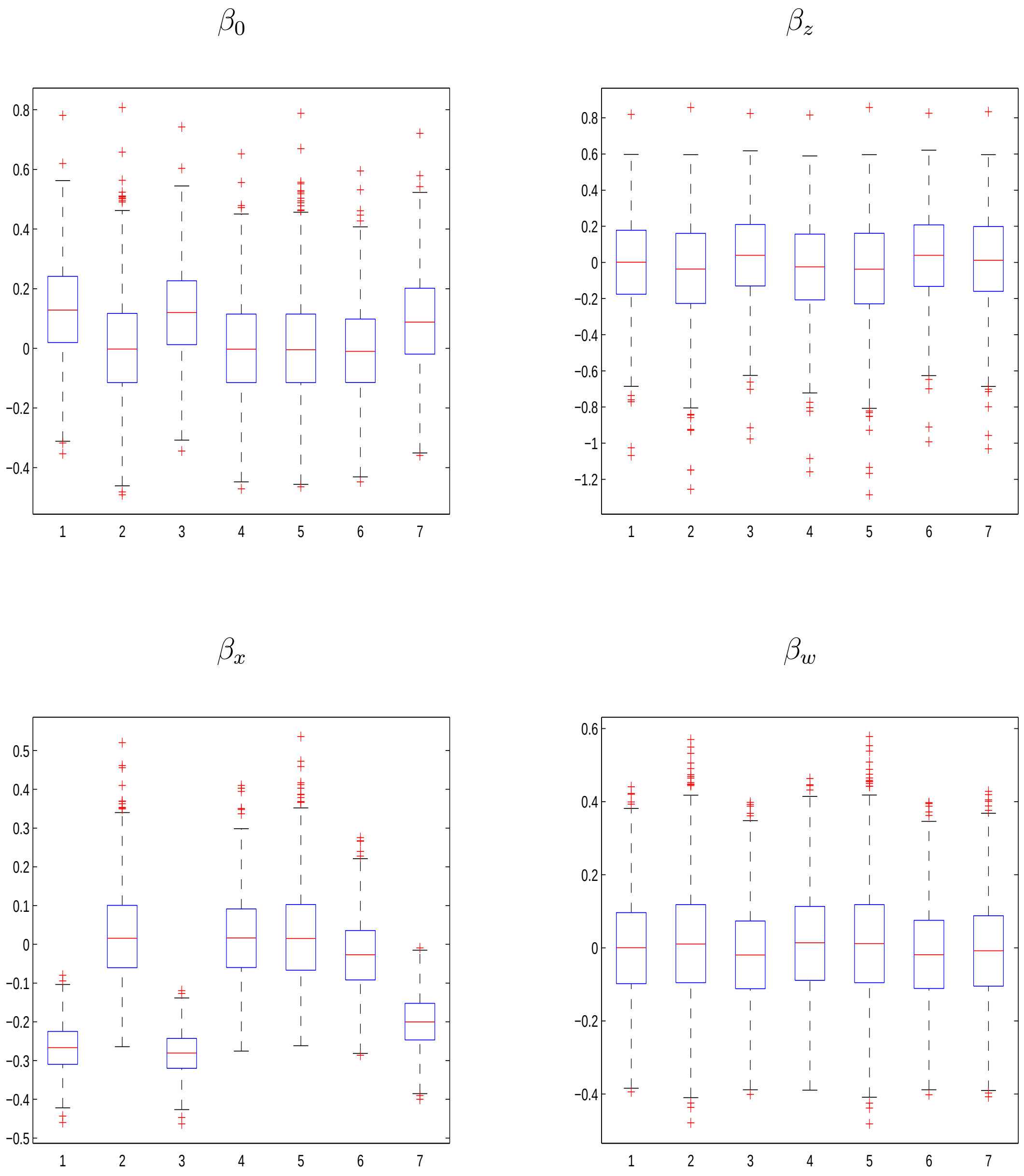

Figure 2: Boxplots of the biases of the seven estimators for β_0, β_z, β_x and β_w in Simulation 2. The seven estimators are respectively pseudo-likelihood (1) , estimating function (2) and regression calibration (3) estimators under uniform distribution model for X, pseudo-likelihood (4), estimating function (5) and regression calibration (6) estimators under normal distribution model for X, and SIMEX estimator (7).

Table 2: Results of Simulation 2 in Section 6.1 based on 1,000 data sets, $m = 500, n = 1000$ and X is uniform. Here $(\beta_0, \beta_z, \beta_x, \beta_w)$ are defined in (13). Mean (est), standard deviation (sd), the average of the estimated standard deviation ($\widehat{\text{sd}}$) and 95% confidence interval coverage are reported for likelihood methods, estimating function methods, regression calibration methods, all based on both uniform and normal latent variable distribution models. SIMEX estimation results are reported in the left last block. The true latent variable distribution is normal.

	β_0	β_z	β_x	β_w	β_0	β_z	β_x	β_w
true	0.1	-1.0	0.7	0.5	0.1	-1.0	0.7	0.5
	maximum likelihood							
	uniform				normal			
est	0.231	-1.008	0.433	0.500	0.100	-1.034	0.721	0.514
sd	0.158	0.256	0.061	0.146	0.166	0.265	0.113	0.151
$\widehat{\text{sd}}$	0.161	0.257	0.060	0.146	0.167	0.266	0.109	0.150
95%CI	89.6%	96.0%	1.8%	94.7%	95.8%	96.1%	95.3%	94.5%
	estimating function							
	uniform				normal			
est	0.104	-1.049	0.723	0.513	0.104	-1.049	0.724	0.514
sd	0.177	0.284	0.122	0.166	0.178	0.284	0.129	0.167
$\widehat{\text{sd}}$	0.170	0.271	0.118	0.161	0.170	0.272	0.118	0.161
95%CI	94.5%	95.9%	95.7%	94.8%	94.1%	95.7%	93.9%	94.6%
	regression calibration							
	uniform				normal			
est	0.221	-0.968	0.419	0.480	0.093	-0.967	0.675	0.481
sd	0.152	0.244	0.055	0.140	0.156	0.244	0.094	0.140
$\widehat{\text{sd}}$	0.155	0.245	0.054	0.140	0.156	0.245	0.088	0.140
95%CI	90.2%	95.5%	0%	94.1%	95.9%	95.5%	91.6%	94.2%
	simulation extrapolation							
est	0.191	-0.992	0.500	0.492	0.191	-0.992	0.500	0.492
sd	0.157	0.252	0.068	0.145	0.157	0.252	0.068	0.145

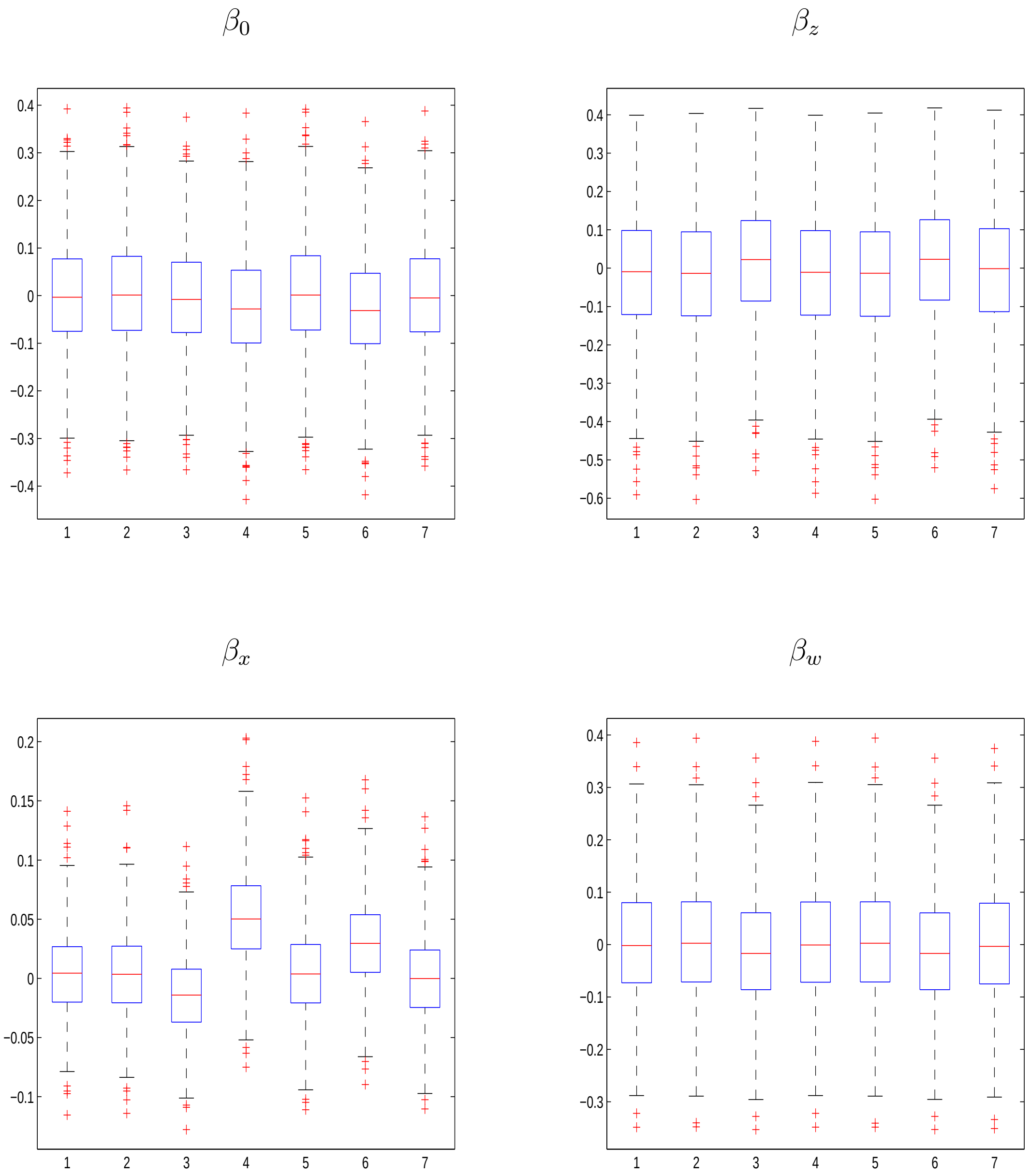

Figure 3: Boxplots of the biases of the seven estimators for β_0, β_z, β_x and β_w in Simulation 3. The seven estimators are respectively pseudo-likelihood (1) , estimating function (2) and regression calibration (3) estimators under uniform distribution model for X, pseudo-likelihood (4), estimating function (5) and regression calibration (6) estimators under normal distribution model for X, and SIMEX estimator (7).

Table 3: Results of Simulation 3 in Section 6.1 based on 1,000 data sets, $m = 200, n = 2000$ and X is normal. Here $(\beta_0, \beta_z, \beta_x, \beta_w)$ are defined in (13). Mean (est), standard deviation (sd), the average of the estimated standard deviation ($\widehat{\text{sd}}$) and 95% confidence interval coverage are reported for likelihood methods, estimating function methods, regression calibration methods, all based on both uniform and normal latent variable distribution models. SIMEX estimation results are reported in the left last block. The true latent variable distribution is uniform.

	β_0	β_z	β_x	β_w	β_0	β_z	β_x	β_w
true	0.1	-1.0	0.7	0.5	0.1	-1.0	0.7	0.5
	maximum likelihood							
	uniform				normal			
est	0.101	-1.016	0.705	0.504	0.078	-1.017	0.753	0.505
sd	0.116	0.162	0.036	0.114	0.117	0.162	0.040	0.114
$\widehat{\text{sd}}$	0.112	0.160	0.037	0.114	0.112	0.160	0.041	0.114
95%CI	94.6%	95.6%	94.7%	94.2%	93.7%	95.6%	77.9%	94.2%
	estimating function							
	uniform				normal			
est	0.105	-1.022	0.704	0.504	0.105	-1.022	0.704	0.504
sd	0.118	0.163	0.037	0.115	0.119	0.163	0.038	0.115
$\widehat{\text{sd}}$	0.107	0.149	0.036	0.114	0.107	0.149	0.036	0.114
95%CI	92.9%	93.3%	94.6%	94.5%	92.9%	93.2%	93.8%	94.5%
	regression calibration							
	uniform				normal			
est	0.097	-0.985	0.686	0.488	0.074	-0.983	0.730	0.488
sd	0.113	0.157	0.033	0.110	0.113	0.156	0.036	0.110
$\widehat{\text{sd}}$	0.108	0.154	0.034	0.110	0.108	0.153	0.036	0.110
95%CI	94.6%	94.2%	93.3%	94.7%	93.5%	94.2%	87.2%	94.9%
	simulation extrapolation							
est	0.101	-1.011	0.701	0.502	0.101	-1.011	0.701	0.502
sd	0.116	0.160	0.036	0.114	0.116	0.160	0.036	0.114

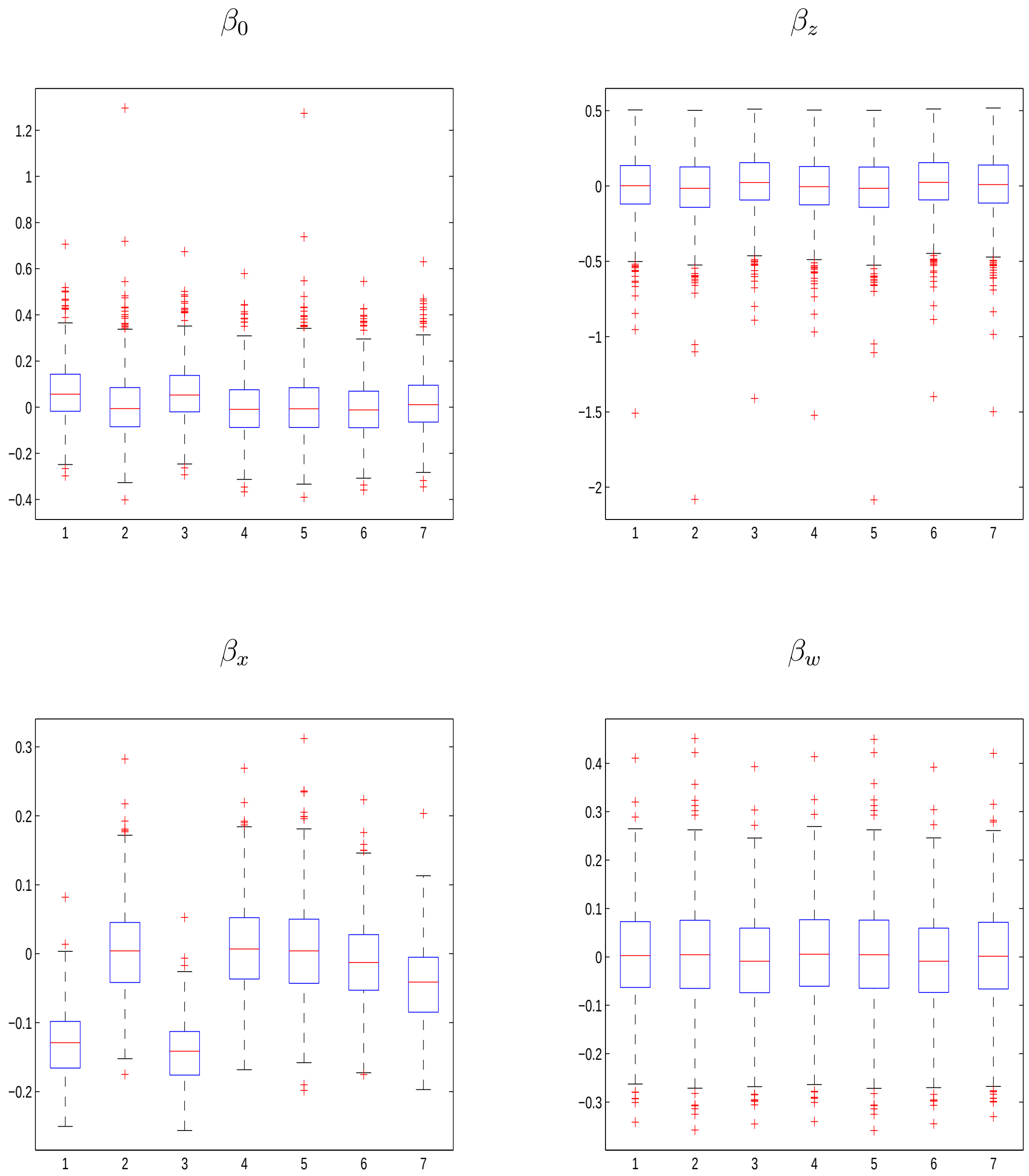

Figure 4: Boxplots of the biases of the seven estimators for β_0, β_z, β_x and β_w in Simulation 4. The seven estimators are respectively pseudo-likelihood (1), estimating function (2) and regression calibration (3) estimators under uniform distribution model for X, pseudo-likelihood (4), estimating function (5) and regression calibration (6) estimators under normal distribution model for X, and SIMEX estimator (7).

Table 4: Results of Simulation 4 in Section 6.1 based on 1,000 data sets, $m = 200, n = 2000$ and X is uniform. Here $(\beta_0, \beta_z, \beta_x, \beta_w)$ are defined in (13). Mean (est), standard deviation (sd), the average of the estimated standard deviation ($\widehat{\text{sd}}$) and 95% confidence interval coverage are reported for likelihood methods, estimating function methods, regression calibration methods, all based on both uniform and normal latent variable distribution models. SIMEX estimation results are reported in the left last block. The true latent variable distribution is normal.

	β_0	β_z	β_x	β_w	β_0	β_z	β_x	β_w
true	0.1	-1.0	0.7	0.5	0.1	-1.0	0.7	0.5
	maximum likelihood							
	uniform				normal			
est	0.163	-1.003	0.570	0.504	0.095	-1.008	0.708	0.507
sd	0.124	0.207	0.049	0.102	0.125	0.208	0.065	0.103
$\widehat{\text{sd}}$	0.124	0.204	0.047	0.103	0.125	0.206	0.062	0.103
95%CI	0.944%	0.950%	0.228%	0.957%	0.946%	0.951%	0.947%	0.955%
	estimating function							
	uniform				normal			
est	0.105	-1.021	0.704	0.507	0.104	-1.021	0.704	0.507
sd	0.140	0.218	0.065	0.108	0.140	0.219	0.069	0.108
$\widehat{\text{sd}}$	0.115	0.180	0.061	0.108	0.115	0.180	0.061	0.108
95%CI	0.905%	0.916%	0.937%	0.955%	0.905%	0.916%	0.924%	0.955%
	regression calibration							
	uniform				normal			
est	0.158	-0.979	0.557	0.492	0.092	-0.979	0.688	0.492
sd	0.121	0.201	0.046	0.100	0.122	0.200	0.060	0.100
$\widehat{\text{sd}}$	0.122	0.199	0.044	0.100	0.121	0.198	0.055	0.100
95%CI	0.946%	0.936%	0.122%	0.958%	0.944%	0.935%	0.908%	0.958%
	simulation extrapolation							
est	0.117	-0.998	0.656	0.502	0.117	-0.998	0.656	0.502
sd	0.123	0.205	0.057	0.102	0.123	0.205	0.057	0.102

Table 5: Analysis of the WISH Study. Estimates (est) and estimated standard deviations ($\widehat{\text{sd}}$) are reported for likelihood methods, estimating function methods, regression calibration methods, based on uniform (uniform) and normal (normal) latent variable distribution models. Naive and SIMEX results are also included. Here $(\beta_z, \beta_x, \beta_w)$ are the coefficients for Z, X and W, respectively.

	β_z	β_x	β_w	β_z	β_x	β_w
			maximum likelihood			
		normal			uniform	
est	-0.586	-0.589	0.251	-0.575	-0.236	0.244
$\widehat{\text{sd}}$	0.476	0.329	0.077	0.458	0.127	0.074
			estimating function			
		normal			uniform	
est	-0.571	-0.576	0.245	-0.572	-0.568	0.240
$\widehat{\text{sd}}$	0.218	0.399	0.076	0.254	0.440	0.077
			regression calibration			
		normal			uniform	
est	-0.556	-0.564	0.239	-0.560	-0.231	0.239
$\widehat{\text{sd}}$	0.442	0.279	0.072	0.443	0.117	0.072
		naive (treat x^* as x)			naive (treat x^*, z^* as x, z)	
est	-0.555	-0.105	0.239	-0.195	-0.105	0.237
$\widehat{\text{sd}}$	0.442	0.052	0.072	0.148	0.051	0.072
			simulation extrapolation			
est	-0.633	-0.184	0.240	-0.633	-0.184	0.240

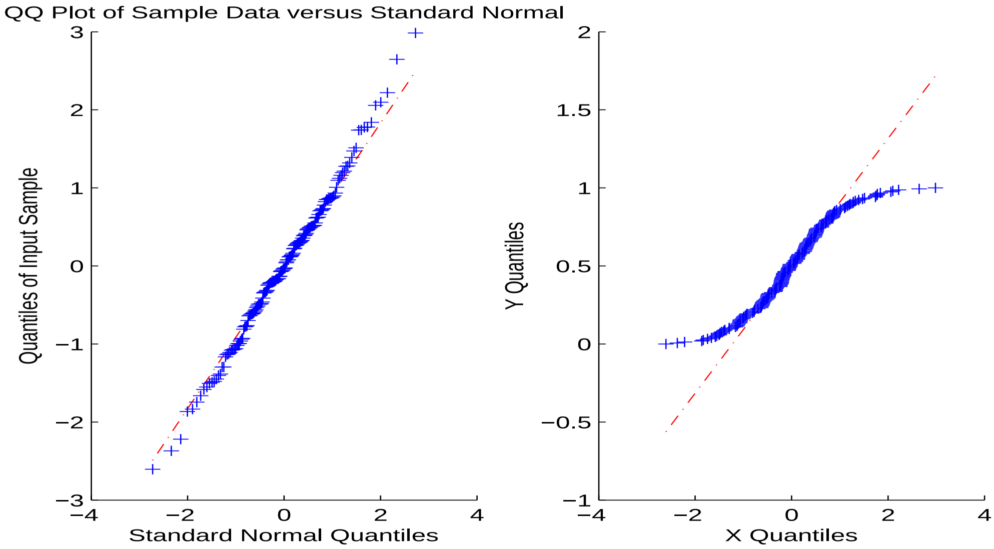

Figure 5: QQ plot: The left one is against normal quantiles, and the right one is against uniform quantiles