17

Estimation in Mixed-effects Models with Measurement Error

Liqun Wang

Department of Statistics, University of Manitoba

CONTENTS

“nobreak

Abstract

In this chapter we survey some recent methods for estimation in generalized linear mixed models with measurement error. The focus is on a semiparametric method that does not require parametric assumptions for the distributions of the unobserved covariates or of the measurement error, and it allows random effects to have any parametric distribution (not necessarily normal). The approach is based on the methods of moments and instrumental variables. We also introduce a simulation-based estimation method for the case where the marginal moments do not have closed forms. The proposed estimators are not only computationally attractive but also root-n consistent under fairly general conditions. Some numerical examples are presented.

Keywords:

Longitudinal data, mixed effects, conditional models, marginal models, measurement error, instrumental variable, semiparametric estimation, simulation-based method.

17.1 Introduction

Measurement error is common in longitudinal data analysis. For example, in epidemiologic studies covariates such as blood pressure, cholesterol level, exposure to air pollutants, dietary intakes, etc. are often measured with error. In HIV trials, the viral load and CD4+ T-cell counts are also subject to measurement error. It is well-known that simply substituting a proxy variable for the unobserved covariate in a model will generally lead to biased and inconsistent estimates of regression coefficients and variance components (e.g., Wang and Davidian 1996; Tosteson, Buonaccorsi and Demidenko 1998; Wang et al 1998; Carroll at al 2006). To account for the measurement error as well as the intra-unit correlation in the longitudinal data, Wang et al (1998) proposed the simulation extrapolation (SIMEX) method to correct for the bias of the naive penalized quasi-likelihood estimator in a generalized linear mixed model, while Wang, Lin and Guitierrez (1999), and Bartlett, de Stavola and Frost (2009) used the regression calibration (RC) approach under the availability of validation data. Although both RC and SIMEX approaches may produce satisfactory results when the measurement errors are small, in general they yield approximate but inconsistent estimators.

On the other hand, Buonaccorsi, Demidenko and Tosteson (2000) studied the likelihood based methods, and Zhong, Fung and Wei (2002) used the corrected score approach. However, the likelihood methods typically entail computational difficulties due to intractability of the likelihood function that involves multiple integrals. Consequently, one usually relies on normality assumption for random effects, measurement error and regression random error term. Non- or semi-parametric approaches have also been considered for models with normal measurement error (Tsiatis and Davidian 2001; Pan, Zeng and Lin 2009). Moreover, Liu and Wu (2007), Wang et al (2008), Yi, Liu and Wu (2011) and Yi, Ma and Carroll (2012) studied the joint problem of covariate measurement error and missing response in longitudinal data models. Most of the above mentioned methods assume that the measurement error variance is known or can be estimated by using validation or replicate data. However, note that the repeated longitudinal measurements taken at different time points cannot be used as replicate data in this context unless the mismeasured covariates are time-invariant.

The instrumental variable (IV) method has been used by many researchers to overcome measurement error problems in cross-sectional data analysis (e.g., Fuller 1987; Buzas and Stefanski 1996; Carroll et al 2006; Wang and Hsiao 2011; Abarin and Wang 2012; Xu, Ma and Wang 2015; and Guan et al. 2019). It is also used in the longitudinal data models with measurement error by, e.g., Li and Wang (2012a) and Abarin et al (2014). In practice, any variable that is correlated with the error-prone covariates and independent of the measurement error can serve as valid IV, e.g., a second independent measurement, or repeated measurements at different time points in longitudinal studies. Hence the IV assumption is weaker than that of replicate data because IV can be a biased observation for the true covariates (Carroll and Stefanski 1994; Carroll et al 2006).

In this chapter, we introduce a semiparametric estimation approach for the generalized linear mixed models with measurement error based on the methods of moments and instrumental variables. This method is easy to implement when the closed forms of the moments exist. We also propose a simulation-based estimator for the case where the marginal moments do not admit closed forms. The proposed estimators are root-n consistent and do not require the parametric assumptions for the distributions of the unobserved covariates or of the measurement errors. Further, these estimators have bounded influence functions so they are robust to data outliers (Li and Wang 2012b).

The rest of this chapter is organized as follows. In Section 17.2 we introduce the IV estimation method in the linear mixed measurement error model. This method is extended to the generalized linear mixed measurement error model in Section 17.3. In Section 17.4 we introduce the simulation-based estimation for the case where the closed forms of the marginal moments are not available. In Section 17.5 we deal with numerical computation issues and present Monte Carlo simulation examples. The asymptotic theories and mathematical proofs are given in Sections 17.6 and 17.7 respectively. Finally, conclusions and discussion are given in Section 17.8.

17.2 Linear Mixed Measurement Error Model

Let Y_{ij} be the response variable of individual i at time j, where $i = 1, 2, ..., n$, $j = 1, 2, ..., m$, $X_{ij} \in I\!\!R^{p_x}$ and $Z_{ij} \in I\!\!R^{p_z}$ be the covariates measured with and without measurement error respectively. Then the linear mixed model is defined as

$$Y_{ij} = X_{ij}^\top \beta_x + Z_{ij}^\top \beta_z + B_{ij}^\top b_i + \varepsilon_{ij}, \tag{17.1}$$

where B_{ij} is a vector of predictors, b_i is the vector of unobserved random effects and ε_{ij} is the random error. Throughout this chapter we assume that Z_{ij} and B_{ij} are exogenous and all expectations are taken conditional on them, however, they are suppressed to simplify notation. Further, we assume that the random effects b_i have mean $E(b_i|X_i) = 0$ and variance $E(b_i b_i^\top|X_i) = \Sigma_b$, where $X_i = (X_{i1}, X_{i2}, ..., X_{im})^\top$, and the random errors $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{im})^\top$ satisfy $E(\varepsilon_i|X_i, b_i) = 0$ and $E(\varepsilon_i \varepsilon_i^\top|X_i, b_i) = \phi I$, where I is the identity matrix.

Thus, under the model assumptions the first two marginal moments of the response variables are given by

$$E(Y_{ij}|X_i) = X_{ij}^\top \beta_x + Z_{ij}^\top \beta_z \tag{17.2}$$

and

$$E(Y_{ij}Y_{ik}|X_i) = (X_{ij}^\top \beta_x + Z_{ij}^\top \beta_z)(X_{ik}^\top \beta_x + Z_{ik}^\top \beta_z) + B_{ij}^\top \Sigma_b B_{ik} + d_{jk}\phi, \tag{17.3}$$

where $d_{jk} = 1$ if $j = k$ and zero otherwise.

For the error-prone covariates X_{ij} we assume an additive classical measurement error model, so that its observed surrogate is

$$X_{ij}^* = X_{ij} + U_{ij}, \tag{17.4}$$

where U_{ij} is the vector of measurement errors satisfying $E(U_{ij}|X_{ij}) = 0$. Further, we assume that instrumental variables $W_{ij} \in I\!R^{p_w}$ are available that are related to X_{ij} through

$$X_{ij} = \Gamma_w W_{ij} + \delta_{ij}, \tag{17.5}$$

where $\Gamma_w \in I\!R^{p_x \times p_w}$ is a matrix of unknown parameters, δ_{ij} is the random error with $E(\delta_{ij}|W_i) = 0$ and $E(\delta_{ij}\delta_{ik}^\top|W_i) = d_{jk}\Sigma_\delta$, and $W_i = (W_{i1}, W_{i2}, ..., W_{im})^\top$. In order to ensure model identifiability, one needs to have at least as many IVs as the number of unobserved variables X_{ij}, so it is assumed that $p_w \geq p_x$ and Γ_w has full row rank (Wang 2021). Moreover, the IV W_{ij} may contain extra variables such as constant 1 and components of Z_{ij} that are highly correlated with X_{ij}. It is further assumed that the random effects and errors satisfy $E(U_{ij}|X_i, W_i) = 0$, $E(\epsilon_{ij}|X_i, X_i^*, W_i) = 0$ and $E(b_i|X_i, X_i^*, W_i) = 0$. These assumptions imply that

$$E(Y_{ij}|X_i, W_i) = E(Y_{ij}|X_i) \tag{17.6}$$

and

$$E(Y_{ij}Y_{ik}|X_i, W_i) = E(Y_{ij}Y_{ik}|X_i). \tag{17.7}$$

Note that in model (17.1), (17.4) and (17.5), there is no assumption on the functional forms of the distributions of unobserved variables, random effects and errors. The observed variables are $(Y_{ij}, X_{ij}^{*\top}, W_{ij}, Z_{ij}^\top, B_{ij}^\top)^\top$ and the parameters of main interest are $\psi = (\beta_x^\top, \beta_z^\top, \theta^\top, \alpha^\top, \phi)^\top$, where θ is the vector of unknown variance parameters in Σ_b and α is the vector of unknown variance parameters in Σ_δ. Throughout this chapter, we denote the parameter space of ψ by Ω_ψ.

The parameter estimation is based on the first two conditional moments of the responses and observed covariates given the instrumental variables. Specifically, by equation (17.6) and the law of iterated expectation, the conditional expectation of Y_{ij} given W_i is given by

$$\begin{aligned}
\mu_{ij}(\psi) &= E(Y_{ij}|W_i) \\
&= E\left[E(Y_{ij}|X_i, W_i)|W_i\right] \\
&= E\left[E(Y_{ij}|X_i)|W_i\right] \\
&= E(X_{ij}|W_i)^\top \beta_x + Z_{ij}^\top \beta_z \\
&= (\Gamma_w W_{ij})^\top \beta_x + Z_{ij}^\top \beta_z.
\end{aligned} \tag{17.8}$$

Similarly, the second conditional moments are given by

$$\begin{aligned}
\nu_{ijk}(\psi) &= E(Y_{ij}Y_{ik}|W_i) \\
&= E\left[E(Y_{ij}Y_{ik}|X_i)|W_i\right] \\
&= E\left[(X_{ij}^\top\beta_x + Z_{ij}^\top\beta_z)(X_{ik}^\top\beta_x + Z_{ik}^\top\beta_z)|W_i\right] + B_{ij}^\top\Sigma_b B_{ik} + d_{jk}\phi \\
&= (\Gamma_w W_{ij})^\top\beta_x + Z_{ij}^\top\beta_z)((\Gamma_w W_{ik})^\top\beta_x + Z_{ik}^\top\beta_z) + B_{ij}^\top\Sigma_b B_{ik} + d_{jk}(\beta_x^\top\Sigma_\delta\beta_x + \phi) \\
&= \mu_{ij}(\psi)\mu_{ik}(\psi) + B_{ij}^\top\Sigma_b B_{ik} + d_{jk}(\beta_x^\top\Sigma_\delta\beta_x + \phi)
\end{aligned} \tag{17.9}$$

and

$$\begin{aligned}
\kappa_{ijk}(\psi) &= E(Y_{ij}X_{ik}^*|W_i) \\
&= E(Y_{ij}X_{ik}|W_i) \\
&= E\left[E(Y_{ij}|X_i)X_{ik}|W_i\right] \\
&= E\left[(X_{ij}^\top\beta_x + Z_{ij}^\top\beta_z)X_{ik}|W_i\right] \\
&= \left[(\Gamma_w W_{ij})^\top\beta_x + Z_{ij}^\top\beta_z\right](\Gamma_w W_{ik}) + d_{jk}\Sigma_\delta\beta_x \\
&= \mu_{ij}(\psi)(\Gamma_w W_{ik}) + d_{jk}\Sigma_\delta\beta_x.
\end{aligned} \tag{17.10}$$

Next we construct semiparametric IV estimator for ψ based on these conditional moments. First, note that substituting (17.5) into (17.4) results in a usual regression equation

$$E(X_{ij}^*|W_{ij}) = \Gamma_w W_{ij} \tag{17.11}$$

and therefore Γ_w can be estimated by the least squares estimator

$$\hat{\Gamma}_w = \left(\sum_{i=1}^n X_i^{*\top} W_i\right)\left(\sum_{i=1}^n W_i^\top W_i\right)^{-1}, \tag{17.12}$$

where $X_i^* = (X_{i1}^*, X_{i2}^*, ..., X_{im}^*)^\top$ and $W_i = (W_{i1}, W_{i2}, ..., W_{im})^\top$. Then we substitute $\hat{\Gamma}_w$ for Γ_w in (17.8)-(17.10) and denote the conditional moments correspondingly as $\hat{\mu}_{ij}, \hat{\nu}_{ijk}$, and $\hat{\kappa}_{ijk}$. Finally, the IV estimator (IVE) for ψ is defined as

$$\hat{\psi}_n = \operatorname*{argmin}_{\psi\in\Omega_\psi} Q_n(\psi) := \sum_{i=1}^n \hat{\rho}_i^\top(\psi)A_i\hat{\rho}_i(\psi), \tag{17.13}$$

where $\hat{\rho}_i^\top(\psi) = (Y_{ij} - \hat{\mu}_{ij}(\psi), Y_{ij}Y_{ik} - \hat{\nu}_{ijk}(\psi), Y_{ij}X_{ik}^* - \hat{\kappa}_{ijk}(\psi), j = 1, ..., m; k = j, ..., m)$ and $A_i = A(W_i)$ is a nonnegative definite matrix that may depend on W_i. A numerical example of this estimator is given in Section 17.5.

17.3 Generalized Linear Mixed Model

Generalized linear mixed models (GLMM) have been widely used to analyze longitudinal data where the response variable can be either discrete or continuous. Various estimation methods for GLMM have been developed in the literature (e.g, Breslow and Clayton 1993; Durbin and Koopman 1997; Rabe-Hesketh, Skrondal and Pickles 2002). However, estimation and inference in a GLMM remain very challenging when some of the covariates are measured with error.

In this section we study the model

$$E(Y_{ij}|X_i,b_i) = g(\eta_{ij}), \tag{17.14}$$
$$V(Y_{ij}|X_i,b_i) = \phi h(\eta_{ij}), \tag{17.15}$$

where $\eta_{ij} = X_{ij}^\top\beta_x + Z_{ij}^\top\beta_z + B_{ij}^\top b_i$ is the linear predictor, $\beta_x \in I\!R^{p_x}$ and $\beta_z \in I\!R^{p_z}$ are vectors of fixed effects, b_i is the random effect having mean zero and distribution $f_b(b;\theta)$ with unknown variance parameters $\theta \in I\!R^{p_b}$. The link function $g(\cdot)$ and variance function $h(\cdot)$ are known functions and $\phi \in I\!R$ is a scale parameter that may be known or unknown. Note that the estimation methods in this chapter do not require the conditional distribution of Y_{ij} given b_i to belong to an exponential family. However, we assume that $Y_{ij}, j = 1,2,...,m$ are conditionally independent given b_i and, therefore, we have

$$E(Y_{ij}^2|X_i,b_i) = g^2(\eta_{ij}) + \phi h(\eta_{ij}) \tag{17.16}$$

and for $j \neq k$,

$$E(Y_{ij}Y_{ik}|X_i,b_i) = E(Y_{ij}|X_i,b_i)E(Y_{ik}|X_i,b_i) = g(\eta_{ij})g(\eta_{ik}). \tag{17.17}$$

Model (17.14)–(17.15) has been studied by various authors, e.g., Wang et al (1998), Buonaccorsi, Demidenko and Tosteson (2000), Zhong, Fung and Wei (2002) and Carroll et al (2006).

Again, suppose that X_{ij} is not observable and we assume the measurement error model (17.4) and instrumental model (17.5). We further assume that W_i, δ_i, b_i are mutually independent and, similar to (17.6) and (17.7), that

$$E(Y_{ij}|X_i,b_i,W_i) = E(Y_{ij}|X_i,b_i) \tag{17.18}$$

and

$$E(Y_{ij}Y_{ik}|X_i,b_i,W_i) = E(Y_{ij}Y_{ik}|X_i,b_i). \tag{17.19}$$

We extend the estimation method in the previous section to model (17.14)-(17.15). To simplify notation, let $g_{ij}(b,t,\beta,\gamma) = g((\Gamma_w W_{ij} + t)^\top\beta_x + Z_{ij}^\top\beta_z + B_{ij}^\top b)$, where $\gamma = \operatorname{vec}\Gamma_w$ is the vector consisting of the columns of Γ_w, and $h_{ij}(b,t,\beta,\gamma)$ is defined similarly. First, by model assumptions and the law of iterated expectation, the first conditional moments of Y_{ij} given W_i are given by

$$\begin{aligned}\mu_{ij}(\psi) &= E(Y_{ij}|W_i)\\ &= E\left[E(Y_{ij}|X_i,b_i)|W_i\right]\\ &= \int g_{ij}(b,t,\beta,\gamma)f_b(b;\theta)f_\delta(t;\alpha)\,db\,dt.\end{aligned} \tag{17.20}$$

Similarly, the second moments are given by

$$\begin{aligned}\nu_{ijj}(\psi) &= E(Y_{ij}^2|W_i)\\ &= \int [g_{ij}^2(b,t,\beta,\gamma) + \phi h_{ij}(b,t,\beta,\gamma)]f_b(b;\theta)f_\delta(t;\alpha)\,db\,dt,\end{aligned} \tag{17.21}$$

and for $j \neq k$,

$$\begin{aligned} \nu_{ijk}(\psi) &= E(Y_{ij}Y_{ik}|W_i) \\ &= \int g_{ij}(b,t,\beta,\gamma)g_{ik}(b,t',\beta,\gamma)f_b(b;\theta)f_\delta(t;\alpha)f_\delta(t';\alpha)\,db\,dt\,dt'. \quad (17.22) \end{aligned}$$

Further, we have

$$\begin{aligned} \kappa_{ijj}(\psi) &= E(Y_{ij}X^*_{ij}|W_i) \\ &= E\left[E(Y_{ij}|X_i,b_i)X_{ij}|W_i\right] \\ &= \int g_{ij}(b,t,\beta,\gamma)(\Gamma_w W_{ij}+t)f_b(b;\theta)f_\delta(t;\alpha)\,db\,dt \\ &= \mu_{ij}(\psi)(\Gamma_w W_{ij}) + \int t g_{ij}(b,t,\beta,\gamma)f_b(b;\theta)f_\delta(t;\alpha)\,db\,dt \quad (17.23) \end{aligned}$$

and for $j \neq k$,

$$\begin{aligned} \kappa_{ijk}(\psi) &= E(Y_{ij}X^*_{ik}|W_i) \\ &= E\left[E(Y_{ij}|X_i,b_i)X_{ik}|W_i\right] \\ &= E\left[E(Y_{ij}|X_i,b_i)(\Gamma_w W_{ik})|W_i\right] \\ &= \mu_{ij}(\psi)(\Gamma_w W_{ik}). \quad (17.24) \end{aligned}$$

Again, we substitute the least squares estimator $\hat{\Gamma}_w$ for Γ_w in (17.20)-(17.24) and denote the resulting moments as $\hat{\mu}_{ij}, \hat{\nu}_{ijk}$, and $\hat{\kappa}_{ijk}$ correspondingly. Then the IV estimator (IVE) for $\psi = (\beta_x^\top, \beta_z^\top, \theta^\top, \alpha^\top, \phi)^\top$ is given by

$$\hat{\psi}_n = \underset{\psi\in\Omega_\psi}{\operatorname{argmin}}\, Q_n(\psi) := \sum_{i=1}^n \hat{\rho}_i{}^\top(\psi) A_i \hat{\rho}_i(\psi), \quad (17.25)$$

where $\hat{\rho}_i^\top(\psi) = (Y_{ij} - \hat{\mu}_{ij}(\psi), Y_{ij}Y_{ik} - \hat{\nu}_{ijk}(\psi), Y_{ij}X^*_{ik} - \hat{\kappa}_{ijk}(\psi), j = 1, ..., m; k = j, ..., m)$ and $A_i = A(W_i)$ is the weight matrix. The asymptotic properties of the IVE $\hat{\psi}_n$ are given in Section 17.6 while a numerical example is given in Section 17.5.

17.4 Simulation-based Estimation

The conditional moments in (17.20)–(17.24) involve multiple integrals. In the case where these integrals admit explicit forms, the numerical optimization of the objective function $Q_n(\psi)$ in (17.25) is straightforward. However, sometimes it is difficult or impossible to obtain closed forms of these integrals, e.g., in a logistic model. In this case, the integrals can be estimated using the following importance sampling technique. First, choose some known densities $f_{0b}(b)$ and $f_{0\delta}(t)$ such that their supports cover the supports of $f_b(b,\theta)$ and $f_\delta(t,\alpha)$ respectively. Then generate independent samples $b_s \sim f_{0b}(b)$ and $t_s, r_s \sim f_{0\delta}(t)$, $s = 1, 2, ..., S$, and let

$$\mu_{ij,1}(\psi) = \frac{1}{S}\sum_{s=1}^S g_{ij}(b_s,t_s,\beta,\gamma)\frac{f_b(b_s;\theta)f_\delta(t_s;\alpha)}{f_{0b}(b_s)f_{0\delta}(t_s)}, \quad (17.26)$$

$$\nu_{ijk,1}(\psi) = \frac{1}{S}\sum_{s=1}^{S} g_{ij}(b_s,t_s,\beta,\gamma)g_{ik}(b_s,r_s,\beta,\gamma)\frac{f_b(b_s;\theta)f_\delta(t_s;\alpha)f_\delta(r_s;\alpha)}{f_{0b}(b_s)f_{0\delta}(t_s)f_{0\delta}(r_s)}, j\neq k, \quad (17.27)$$

further, generate another set of samples $b_s \sim f_{0b}(b)$ and $t_s, r_s \sim f_{0\delta}(t)$, $s = S+1, S+2, ..., 2S$ to calculate $\mu_{ij,2}(\psi)$ and $\nu_{ijk,2}(\psi)$ respectively. Other moments are estimated similarly. It is easy to see that such simulated moments $\mu_{ij,l}(\psi)$, $\nu_{ijk,l}(\psi)$ and $\kappa_{ijk,l}(\psi)$, $l = 1, 2$ are unbiased estimators for $\mu_{ij}(\psi)$, $\nu_{ijk}(\psi)$ and $\kappa_{ijk}(\psi)$ respectively. Finally, the simulation-based IV estimator (SIVE) for ψ is defined as

$$\hat{\psi}_{n,S} = \underset{\psi\in\Omega_\psi}{\operatorname{argmin}}\, Q_{n,S}(\psi) := \sum_{i=1}^{n} \hat{\rho}_{i,1}^\top(\psi)A_i\hat{\rho}_{i,2}(\psi), \quad (17.28)$$

where $\hat{\rho}_{i,l}^\top(\psi) = (Y_{ij}-\hat{\mu}_{ij,l}(\psi), Y_{ij}Y_{ik}-\hat{\nu}_{ijk,l}(\psi), Y_{ij}X_{ik}^*-\hat{\kappa}_{ijk,l}(\psi), j = 1, ..., m; k = j, ..., m)$, $l = 1, 2$. Note that, since $\hat{\rho}_{i,1}(\psi)$ and $\hat{\rho}_{i,2}(\psi)$ are constructed by using two independent importance samples, they are conditionally independent given the observed data $(Y_i, X_i^*, W_i, Z_i, B_i)$. Therefore $Q_{n,S}(\psi)$ is an unbiased simulator for $Q_n(\psi)$ in (17.25) for finite S. A numerical example of this estimator is given in Section 17.5.

17.5 Numerical Computation and Simulations

In principle, the estimators IVE and SIVE can be computed using Newton-Raphson algorithm as

$$\hat{\psi}^{(\tau+1)} = \hat{\psi}^{(\tau)} - \left(\frac{\partial^2 Q_n(\hat{\psi}^{(\tau)})}{\partial\psi\partial\psi^\top}\right)^{-1}\frac{\partial Q_n(\hat{\psi}^{(\tau)})}{\partial\psi},$$

where $\hat{\psi}^{(\tau)}$ denotes the estimate of ψ at the τ^{th} iteration, and the gradient and Hessian matrix are given by

$$\frac{\partial Q_n(\psi)}{\partial\psi} = 2\sum_{i=1}^{n}\frac{\partial\rho_i(\psi)^\top}{\partial\psi}A_i\rho_i(\psi), \quad (17.29)$$

and

$$\frac{\partial^2 Q_n(\psi)}{\partial\psi\partial\psi^\top} = 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i(\psi)^\top}{\partial\psi}A_i\frac{\partial\rho_i(\psi)}{\partial\psi^\top} + (\rho_i(\hat{\psi})^\top A_i\otimes I)\frac{\partial\text{vec}(\partial\rho_i(\psi)^\top/\partial\psi)}{\partial\psi^\top}\right] \quad (17.30)$$
$$\approx 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i(\psi)^\top}{\partial\psi}A_i\frac{\partial\rho_i(\psi)}{\partial\psi^\top}\right].$$

respectively. The last approximation holds because the second term in (17.30) has expectation zero at the true value of the parameters and therefore can be omitted for computational simplicity.

Another issue is the choice of the weight matrix $A_i = A(W_i)$ in the objective functions $Q_n(\psi)$ and $Q_{n,S}(\psi)$. Although any nonnegative definite matrix satisfying the regularity conditions in Section 17.6 will result in consistent estimators, the optimal choice that yields most efficient estimator is $A_i^{opt} = E[\rho_i(\psi_0)\rho_i^\top(\psi_0)|W_i]^{-1}$ (Abarin and Wang 2006, Wang

2007). Since the optimal weight depends on the unknown parameters to be estimated, it needs to be pre-estimated by plugging in a first-stage estimator such as the naive estimator that ignores measurement error or by the IVE (SIVE) calculated using the identity weight $A_i = I$. All numerical examples in this section used the following estimate

$$\hat{A}^{opt} = \left(\frac{1}{n} \sum_{i=1}^{n} \rho_i(\hat{\psi}_n^{(1)}) \rho_i^{\top}(\hat{\psi}_n^{(1)}) \right)^{-1}, \tag{17.31}$$

where $\hat{\psi}_n^{(1)}$ is the first-stage estimator. A more detailed discussion on the choice of A_i^{opt} can be found in Li and Wang (2012b, 2013).

In the rest of this section, we calculate some numerical examples to demonstrate the proposed estimation procedures and carry out Monte Carlo simulation studies to evaluate their finite sample performance. For comparison we also calculate the naive maximum likelihood estimate (nMLE) that ignores measurement error. In each example, we carry out 1000 Monte Carlo runs and report the biases (BIAS) and the mean absolute errors (MAE) of the estimators. All computations are done using statistical programming language `R` (R Development Core Team 2009). The numerical optimization was done using function `nlm()` and the nMLE are calculated using package `lme4`.

17.5.1 Linear Mixed Model

First we consider the linear mixed model

$$Y_{ij} = \beta_x X_{ij} + \beta_z Z_{ij} + b_i + \varepsilon_{ij}, \tag{17.32}$$

where X_{ij}, Z_{ij} are scalars, $B_{ij} \equiv 1$, $b_i \sim (0, \theta)$ (any distribution with mean 0 and variance θ), and $\varepsilon \sim (0, \phi)$. Further, suppose $X_{ij} = \gamma_1 + \gamma_w W_{ij} + \delta_{ij}$ with $\delta_{ij} \sim (0, \alpha)$, and $X_{ij}^* = X_{ij} + U_{ij}$ with $U_{ij} \sim (0, \sigma_u^2)$. For this model the conditional moments (17.8)–(17.10) become

$$\mu_{ij}(\psi) = \beta_x(\gamma_1 + \gamma_w W_{ij}) + \beta_z Z_{ij},$$

$$\nu_{ijk}(\psi) = \mu_{ij}(\psi)\mu_{ik}(\psi) + \theta + d_{jk}(\alpha\beta_x^2 + \phi)$$

and

$$\kappa_{ijk}(\psi) = \mu_{ij}(\psi)(\gamma_1 + \gamma_w W_{ij}) + d_{jk}\alpha\beta_x.$$

In the simulation study, the data are generated as follows. First generate independent variates $W_{ij} \sim N(0, 1)$ and $X_{ij} = 0.7W_{ij} + \delta_{ij}$ with $\delta_{ij} \sim N(0, \alpha = 0.1)$. Then, generate $X_{ij}^* = X_{ij} + U_{ij}$ with $U_{ij} \sim N(0, 0.1)$. The covariate $Z_{ij} = j, j = 1, 2, ..., m$. Finally, generate $b_i \sim N(0, \theta = 0.2)$ and the responses Y_{ij} according to (17.32) with $\varepsilon_{ij} \sim N(0, \phi = 0.5)$ and parameter values $(\beta_0, \beta_x, \beta_z) = (1.5, 1, -0.2)$. The sample sizes are $n = 100, 300$ respectively with repetition $m = 4$ in each case.

The simulation results are reported in Table 17.1. These results show clearly that the naive MLE for β_x is biased and the bias persists when the sample size increases. In contrast, the proposed IVE corrects the bias effectively. In addition, when the sample size increases, both the bias and mean absolute error of the IVE decrease rapidly. For the dispersion

Table 17.1
BIAS (MAE) of the naive MLE and IVE for the linear model in Example 17.5.1.

	$n = 100$		$n = 300$	
Parameter	nMLE	IVE	nMLE	IVE
$\beta_0 = 1.5$	-0.001 (0.084)	-0.020 (0.095)	0.003 (0.048)	-0.006 (0.055)
$\beta_x = 1$	-0.144 (0.144)	-0.072 (0.080)	-0.144 (0.144)	-0.028 (0.037)
$\beta_z = -0.2$	0.000 (0.028)	0.003 (0.030)	-0.001 (0.016)	0.000 (0.018)
$\theta = 0.2$	0.002 (0.040)	-0.013 (0.049)	-0.000 (0.022)	-0.005 (0.028)
$\phi = 0.5$	0.085 (0.086)	-0.015 (0.045)	0.085 (0.086)	-0.006 (0.027)

parameter ϕ, The IVE has significantly smaller bias and mean absolute error than the naive MLE. In some cases, the IVE has slightly larger MAE than the naive MLE, which is due to the numerical variability of the optimization algorithm used.

17.5.2 Mixed Poisson Model

Next we consider a mixed Poisson model where $g(\eta_{ij}) = h(\eta_{ij})$, $\phi = 1$ and

$$E(Y_{ij}|X_i, b_i) = \exp(\beta_0 + \beta_x X_{ij} + \beta_z Z_{ij} + b_i). \tag{17.33}$$

Further, suppose that $b_i \sim N(0, \theta)$ and $X_{ij} = \gamma_1 + \gamma_w W_{ij} + \delta_{ij}$ with $\delta_i \sim N(0, \alpha I)$. Then the moments (17.20)–(17.24) have the following closed forms. First, using the formula of the normal moment generating function, we obtain

$$\begin{aligned} \mu_{ij}(\psi) &= E(Y_{ij}|W_i) \\ &= \exp(\beta_0 + \beta_x(\gamma_1 + \gamma_w W_{ij}) + \beta_z Z_{ij}) E\exp(\beta_x \delta_{ij}) E\exp(b_i) \\ &= \exp(\beta_0 + \beta_x(\gamma_1 + \gamma_w W_{ij}) + \beta_z Z_{ij} + \beta_x^2 \alpha/2 + \theta/2). \end{aligned} \tag{17.34}$$

Similarly, we have

$$\begin{aligned} \nu_{ijj}(\psi) &= E(Y_{ij}^2|W_i) \\ &= \mu_{ij}(\psi) + \exp[2(\beta_0 + \beta_x(\gamma_1 + \gamma_w W_{ij}) + \beta_z Z_{ij})] E\exp(2\beta_x \delta_{ij}) E\exp(2b_i) \\ &= \mu_{ij}(\psi) + \mu_{ij}^2(\psi)\exp(\alpha\beta_x^2 + \theta) \end{aligned} \tag{17.35}$$

and, for $j \neq k$,

$$\begin{aligned} \nu_{ijk}(\psi) &= E(Y_{ij}Y_{ik}|W_i) \\ &= \exp(\beta_0 + \beta_x(\gamma_1 + \gamma_w W_{ij}) + \beta_z Z_{ij}) \exp(\beta_0 + \beta_x(\gamma_1 + \gamma_w W_{ik}) + \beta_z Z_{ik}) \\ &\quad \times E\exp(\beta_x \delta_{ij}) E\exp(\beta_x \delta_{ik}) E\exp(2b_i) \\ &= \mu_{ij}(\psi)\mu_{ik}(\psi)\exp(\theta). \end{aligned} \tag{17.36}$$

Table 17.2
BIAS (MAE) of the naive MLE and IVE for the Poisson model in Example 17.5.2.

	$n = 100$		$n = 300$	
Parameter	nMLE	IVE	nMLE	IVE
$\beta_0 = 1.2$	0.324 (0.324)	0.120 (0.142)	0.318 (0.318)	0.083 (0.098)
$\beta_x = 0.9$	-0.200 (0.200)	-0.112 (0.117)	-0.200 (0.200)	-0.074 (0.076)
$\beta_z = -0.2$	-0.001 (0.017)	0.002 (0.020)	0.001 (0.009)	0.003 (0.014)
$\theta = 0.2$	0.017 (0.031)	-0.054 (0.063)	0.018 (0.023)	-0.056 (0.058)

Finally,

$$\begin{aligned}\kappa_{ijk}(\psi) &= E(Y_{ij}X^*_{ik}|W_i)\\ &= E(Y_{ij}|W_i)(\gamma_1+\gamma_w W_{ik})\\ &\quad + \exp(\beta_0+\beta_x(\gamma_1+\gamma_w W_{ik})+\beta_z Z_{ik})E[\delta_{ik}\exp(\beta_x\delta_{ij})]E\exp(b_i)\\ &= \mu_{ij}(\psi)(\gamma_1+\gamma_w W_{ik}+d_{jk}\alpha\beta_x). \end{aligned} \tag{17.37}$$

In the simulations the covariates are generated similarly as in Example 17.5.1, except now $X_{ij} = 1.5 + 0.5W_{ij} + \delta_{ij}$ and the responses Y_{ij} are generated from Poisson distribution with conditional mean (17.33) and true parameter values $(\beta_0, \beta_x, \beta_z) = (1.2, 0.9, -0.2)$. The sample sizes are $n = 100, 300$ and $m = 4$.

The simulation results are given in Table 17.2. Again, these results show clearly that the naive MLE for β_0 and β_x are severely biased and the biases persist for the large sample. In contrast, the proposed IVE corrects the bias effectively, and its bias and mean absolute error decrease significantly when sample size increases. Again, in some cases the IVE has slightly larger MAE than the naive MLE due to the numerical instability in the optimization algorithm.

17.5.3 Mixed Logistic Model

In the third example we consider a mixed logistic model for a binary response Y_{ij} taking values 0 or 1 with probability

$$P(Y_{ij} = 1|X_{ij}, b_i) = g(\beta_0 + \beta_x X_{ij} + \beta_z Z_{ij} + b_i), \tag{17.38}$$

where $g(\eta_{ij}) = 1/(1 + \exp(-\eta_{ij}))$. In this model the variance function $h(\eta_{ij}) = g(\eta_{ij})(1 - g(\eta_{ij}))$ and $\phi = 1$. For this model the conditional moments in (17.20)–(17.24) do not have closed forms and therefore we apply the simulation-based procedure in Section 17.4. In addition, since Y_{ij} is binary, $E(Y_{ij}^2|X_i, b_i) = E(Y_{ij}|X_i, b_i)$ and, therefore, the moments ν_{ijj} in (17.21) are not included in the objective function $Q_{n,S}(\psi)$.

In the simulation study, the covariates are generated similarly as in Example 17.5.1, except now $X_{ij} = 0.5W_{ij} + \delta_{ij}$ and the response Y_{ij} are generated from Bernoulli distribution with conditional probability (17.38) and true parameter values $(\beta_0, \beta_x, \beta_z) = (0.4, 0.6, -0.2)$.

Table 17.3
BIAS (MAE) of the naive MLE and SIVE for the logistic model in Example 17.5.3.

	$n = 200$		$n = 400$	
Parameter	nMLE	SIVE	nMLE	SIVE
$\beta_0 = 0.4$	-0.002 (0.144)	0.017 (0.182)	-0.007 (0.100)	0.015 (0.133)
$\beta_x = 0.6$	-0.141 (0.153)	0.018 (0.139)	-0.139 (0.142)	0.012 (0.106)
$\beta_z = -0.2$	0.002 (0.051)	-0.007 (0.064)	0.002 (0.037)	-0.003 (0.046)
$\theta = 0.2$	-0.029 (0.100)	0.049 (0.224)	-0.042 (0.079)	0.000 (0.185)

The sample sizes are $n = 200, 400$ and $m = 4$. To compute the SIVE, we chose the normal density of $N(0, 2)$ for both $f_{0b}(b)$ and $f_{0\delta}(t)$, and generate independent samples b_s, t_s, r_s, $s = 1, ..., 2S$ with $S = 5000$.

The simulation results are shown in Table 17.3. Similar to the previous examples, these results show that the naive MLE for β_x is biased and the bias persists for the large sample. In contrast, the proposed IVE corrects the bias effectively, and its bias and mean absolute error decrease significantly when sample size increases. Again, in some cases the IVE has larger MAE than the naive MLE due to the numerical instability in the optimization algorithm.

17.6 Asymptotic Theory

In this section we outline the asymptotic theories for the proposed estimators and the associated regularity conditions. For the sake of generality, we present the conditions and results in terms of the generalized linear mixed models, however, some of these conditions can be simplified for the linear mixed models. As in Section 17.2 and 17.3, let ψ and Ω_ψ be the parameter vector and its parameter space respectively. Similarly, let $\gamma = \operatorname{vec}\Gamma_w$, Ω_γ be the corresponding parameter space, and $\Omega = \Omega_\psi \times \Omega_\gamma$ be the joint parameter space. Finally, let ψ_0 and γ_0 denote the true values with which the observed data are generated.

To obtain the consistency and asymptotic normality for the IVE and SIVE, we make the following assumptions.

Assumption 17.1. $g(\cdot)$ and $h(\cdot)$ are continuously differentiable functions.

Assumption 17.2. $(Y_i, X_i^*, W_i, Z_i, B_i)$, $i = 1, ..., n$ are independent and identically distributed and satisfy $E\left[\|A_i\|\left(Y_{ij}^4 + \|Y_{ij}W_{ij}\|^2 + 1\right)\right] < \infty$. Further,

$$E\left[\|A_i\| \int \sup_{\Omega} g_{ij}^2(b, t, \beta, \gamma) f_b(b; \theta) f_\delta(t; \alpha)\, db\, dt\right] < \infty \tag{17.39}$$

and

$$E\left[\|A_i\| \int \sup_{\Omega} h_{ij}(b, t, \beta, \gamma) f_b(b; \theta) f_\delta(t; \alpha)\, db\, dt\right] < \infty. \tag{17.40}$$

Assumption 17.3. The parameter space Ω_ψ is compact.

Assumption 17.4. $E[(\rho_i(\psi) - \rho_i(\psi_0))^\top A_i(\rho_i(\psi) - \rho_i(\psi_0))] = 0$ if and only if $\psi = \psi_0$.

Assumption 17.5. $g(\cdot)$ and $h(\cdot)$ are twice continuously differentiable; $f_b(b;\theta)$ and $f_\delta(t;\alpha)$ are twice continuously differentiable with respect to θ and α respectively around their true values; all first and second order partial derivatives of $g_{ij}(b,t,\beta,\gamma)f_b(b;\theta)f_\delta(t;\alpha)$ and $h_{ij}(b,t,\beta,\gamma)f_b(b;\theta)f_\delta(t;\alpha)$ with respect to $(\psi^\top,\gamma^\top)^\top$ satisfy similar conditions as in (17.39) and (17.40).

Assumption 17.6. The matrix

$$D = E\left[\frac{\partial \rho_i^\top(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_i(\psi_0)}{\partial \psi'}\right] \tag{17.41}$$

is nonsingular.

Then we have the following results.

Theorem 17.1. Let $\hat{\psi}_n$ be the IVE defined in (17.25). Then, as $n \to \infty$,

1. under Assumption 17.1-17.4, $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$; and
2. under Assumption 17.1-17.6, $\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, D^{-1}CD^{-1})$, where

$$C = \operatorname*{plim}_{n\to\infty} \frac{1}{4n} \frac{\partial Q_n(\psi_0)}{\partial \psi} \frac{\partial Q_n(\psi_0)}{\partial \psi^\top}. \tag{17.42}$$

Further, for the simulation-based IVE of Section 17.4, we have the following results.

Theorem 17.2. Suppose that the supports of $f_b(b,\theta)$ and $f_\delta(t,\alpha)$ are contained in the supports of $f_{0b}(b)$ and $f_{0\delta}(t)$ respectively for all θ and α in the neighborhood of their true values. Let $\hat{\psi}_{n,S}$ be the SIVE defined in (17.28). Then for any fixed $S > 0$, as $n \to \infty$,

1. under Assumptions 17.1-17.4, $\hat{\psi}_{n,S} \xrightarrow{a.s.} \psi_0$;
2. under Assumptions 17.1-17.6, $\sqrt{n}(\hat{\psi}_{n,S} - \psi_0) \xrightarrow{d} N(0, D^{-1}C_S D)$, where

$$C_S = \operatorname*{plim}_{n\to\infty} \frac{1}{4n} \frac{\partial Q_{n,S}(\psi_0)}{\partial \psi} \frac{\partial Q_{n,S}(\psi_0)}{\partial \psi^\top}. \tag{17.43}$$

Note that the above asymptotic results do not require the simulation size S to tend to infinity because the objective function $Q_{n,S}(\psi)$ is constructed using the simulation-by-parts technique. This is fundamentally different from other simulation-based methods in the literature which typically require S tend to infinity to obtain consistent estimators. However, due to the Monte Carlo approximation of the marginal moments, $\hat{\psi}_{n,S}$ is generally less efficient than $\hat{\psi}_n$. In general, it can be shown that the efficiency loss caused by simulation decreases at the rate $O(1/S)$ (Wang 2004, 2007).

17.7 Mathematical Proofs

In this section we give a sketch of the proof of Theorem 17.1. Theorem 17.2 can be similarly proved. For more details, see Wang (2004, 2007) and Wang and Hsiao (2011).

17.7.1 Proof of Theorem 17.1.1

By Assumption 17.1 and the Dominated Convergence Theorem (DCT), we have the first-order Taylor expansion about γ_0

$$Q_n(\psi) = \sum_{i=1}^{n} \rho_i^\top(\psi) A_i \rho_i(\psi) + 2\sum_{i=1}^{n} \rho_i^\top(\psi,\tilde{\gamma}) A_i \frac{\partial \rho_i(\psi,\tilde{\gamma})}{\partial \gamma^\top}(\hat{\gamma}_n - \gamma_0), \tag{17.44}$$

where $\|\tilde{\gamma} - \gamma_0\| \le \|\hat{\gamma}_n - \gamma_0\|$ and $\hat{\gamma}_n = \operatorname{vec}\hat{\Gamma}_w$. Further, for any $1 \le i \le n$, by Assumption 17.1–17.3 and the uniform law of large numbers (ULLN), we have

$$\sup_{\psi\in\Omega_\psi} \left| \frac{1}{n}\sum_{i=1}^{n} \rho_i^\top(\psi) A_i \rho_i(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0, \tag{17.45}$$

where $Q(\psi) = E[\rho_i^\top(\psi) A_i \rho_i(\psi)]$. Similarly, by Assumption 17.1-17.3 we have

$$\begin{aligned} &\sup_{\Omega_\psi} \left\| \frac{1}{n}\sum_{i=1}^{n} \rho_i^\top(\psi,\tilde{\gamma}) A_i \frac{\partial \rho_i(\psi,\tilde{\gamma})}{\partial \gamma^\top}(\hat{\gamma}_n - \gamma_0) \right\| \\ &\le \sup_{\Omega} \left\| \frac{1}{n}\sum_{i=1}^{n} \rho_i^\top(\psi,\gamma) A_i \frac{\partial \rho_i(\psi,\gamma)}{\partial \gamma^\top} \right\| \|\hat{\gamma}_n - \gamma_0\| \xrightarrow{a.s.} 0. \end{aligned} \tag{17.46}$$

Therefore it follows from (17.44)–(17.46) that

$$\sup_{\Omega_\psi} \left| \frac{1}{n} Q_n(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0. \tag{17.47}$$

Furthermore, since $Q(\psi) = Q(\psi_0) + E[\rho_i(\psi) - \rho_i(\psi_0))^\top A_i(\rho_i(\psi) - \rho_i(\psi_0)]$, by Assumption 17.4, $Q(\psi) \ge Q(\psi_0)$ and the equality holds if and only if $\psi = \psi_0$. Thus, by Amemiya (1973, Lemma 3) we have $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$, as $n \to \infty$.

17.7.2 Proof of Theorem 17.1.2

By Assumption 17.5 and the DCT, the first derivative $\partial Q_n(\psi)/\partial\psi$ exists and has the first-order Taylor expansion in the neighborhood of ψ_0. Since $\hat{\psi}_n \xrightarrow{a.s.} \psi_0$, for sufficiently large n we have

$$\frac{\partial Q_n(\hat{\psi}_n)}{\partial \psi} = \frac{\partial Q_n(\psi_0)}{\partial \psi} + \frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial \psi \partial \psi^\top}(\hat{\psi}_n - \psi_0) = 0, \tag{17.48}$$

where $\left\|\tilde{\psi}_n - \psi_0\right\| \le \left\|\hat{\psi}_n - \psi_0\right\|$, and the first and second derivatives of $Q_n(\psi)$ are given in (17.29) and (17.30) respectively.

Analogous to the proof of Theorem 17.1.1, by Assumption 17.1–17.5, the ULLN and Amemiya (1973, Lemma 4) we can show that

$$\frac{1}{2n}\frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial\psi\partial\psi^\top} \xrightarrow{a.s.} E\left[\frac{\partial\rho_i^\top(\psi_0)}{\partial\psi}A_i\frac{\partial\rho_i(\psi_0)}{\partial\psi^\top}\right] = D, \tag{17.49}$$

where D is given in (17.41). Since D is nonsingular by Assumption 17.6, for sufficiently large n, we can rewrite (17.48) as

$$\sqrt{n}(\hat{\psi}_n - \psi_0) = -\left(\frac{1}{2n}\frac{\partial^2 Q_n(\tilde{\psi}_n)}{\partial\psi\partial\psi^\top}\right)^{-1}\left(\frac{1}{2\sqrt{n}}\frac{\partial Q_n(\psi_0)}{\partial\psi}\right). \tag{17.50}$$

Further, similar to Wang (2007) and Wang and Hsiao (2011), it can be shown that

$$\frac{1}{2\sqrt{n}}\frac{\partial Q_n(\psi_0)}{\partial\psi} \xrightarrow{d} N(0, C), \tag{17.51}$$

where C is given in (17.42). Finally, the theorem follows from (17.48)–(17.51) and Slutsky's Theorem.

17.8 Conclusions and Discussion

In this chapter we proposed an instrumental variable approach to estimation of the generalized linear mixed models with measurement error. This approach is based on the first two conditional moments of the responses given the instrumental variables, and it does not require distributional assumption for the unobserved covariates and measurement error. The random effects can have any parametric distribution that is not necessarily normal. A simulation-based estimator is developed to overcome the computational difficulty when the marginal moments do not have closed forms. The proposed estimators are consistent and asymptotically normally distributed under general conditions. Numerical examples demonstrate that the proposed estimators effectively correct the attenuation bias in the naive maximum likelihood estimator caused by the measurement error. The computation of the estimators is done through numerical optimization. Therefore developing more efficient numerical optimization procedure will be very helpful.

Acknowledgements

The author is grateful to co-editor Paul Gustafson for his kind invitation to write this chapter and for his helpful comments and suggestions on the previous drafts. I am also thankful to Zhiyong Jin and Lin Xue for their assistance in R programming for the computation of the numerical examples. This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Bibliography

[1] Abarin, T. and Wang, L. (2006). Comparison of GMM with second-order least squares estimator in nonlinear models. *Far East Journal of Theoretical Statistics*, 20, 179-196.

[2] Abarin, T. and Wang, L. (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics*, 64, 475-493.

[3] Abarin, T., Li, H., Wang, L. and Briollais, L. (2014). On method of moments estimation in linear mixed effects models with measurement error on covariates and response with application to a longitudinal study of gene-environment interaction. *Statistics in Biosciences*, 6, 1-18.

[4] Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, 41, 997-1016.

[5] Bartlett, J.W., de Stavola, B.L. and Frost, C. (2009). Linear mixed models for replication data to efficiently allow for covariate measurement error. *Statistics in Medicine*, 28, 3158-3178.

[6] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistician Association*, 88, 9-25.

[7] Buonaccorsi, J., Demidenko, E. and Tosteson T. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica*, 10, 885-836.

[8] Buzas, J.S. and Stefanski L.A. (1996). Instrumental variable estimation in generalized measurement error models. *Journal of American Statistical Association*, 91, 999-1006.

[9] Carroll, R.J. and Stefanski, L.A. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine*, 13, 1265-1282.

[10] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd Edition, Chapman & Hall, London.

[11] Durbin, J. and Koopman, S.J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84, 669-684.

[12] Fuller, W. (1987). *Measurement Error Models*, New York: John Wiley & Sons.

[13] Guan, J., Cheng, H., Bollen, K.A., Thomas, D.R. and Wang, L. (2019). Instrumental variable estimation in ordinal Probit models with mismeasured predictors. *Canadian Journal of Statistics*, 47, 653–66.

[14] Li, D. and Wang, L. (2013). A semiparametric estimation approach for linear mixed models. *Communications in Statistics - Theory and Methods*, 42, 1982-1997.

[15] Li, H. and Wang, L (2012a). Consistent estimation in generalized linear mixed models with measurement error. *Journal of Biometrics and Biostatistics*, S7:007, doi:10.4172/2155-6180.S7-007.

[16] Li, H. and Wang, L. (2012b). A consistent simulation-based estimator in generalized linear mixed models. *Journal of Statistical Computation and Simulation*, 82, 1085-1103.

[17] Liu, W. and Wu, L. (2007). Simultaneous inference for semiparametric nonlinear mixed?effects models with covariate measurement errors and missing responses. *Biometrics*, 63, 342-350.

[18] Pan, W., Zeng, D. and Lin, X. (2009). Semiparametric transition measurement error models for longitudinal data. *Biometrics*, 65, 728-736.

[19] R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

[20] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1-21.

[21] Tsiatis, T. and Davidian, M. (2001) A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88, 447 - 458.

[22] Tosteson, T., Buonaccorsi, J., and Demidenko, E. (1998). Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine*, 17, 1959-1971.

[23] 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.

[24] Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *Annals of Statistics*, 32, 2559-2579.

[25] Wang, L. (2007). A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models. *The Canadian Journal of Statistics*, 35, 233-248.

[26] Wang, L. (2021). Identifiability in measurement error models. In G.Y. Yi, A. Delaigle and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 3. Chapman & Hall/CRC.

[27] Wang, L. and Hsiao, C. (2011). Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics*, 165, 30-44.

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

[29] Wang, N., Lin, X. and Guttierrez, R. G. (1999). A bias correction regression calibration approach in generalized linear mixed measurement error models. *Communication in Statistics - Theory and Methods*, 28, 217-233.

[30] Wang, N., Lin, X., Gutierrez, R. G. and Carroll, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association*, 93, 249-261.

[31] Xu, K., Ma, Y. and Wang, L. (2015). Instrument assisted regression for errors in variables models with binary response. *Scandinavian Journal of Statistics* 42, 104-117.

[32] 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.

[33] 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.

[34] Zhong, X.P., Fung, W.K. and Wei, B.C. (2002). Estimation in linear models with random effects and errors-invariables. *Annals of the Institute of Statistical Mathematics*, 54, 595-606.