Journal of Biometrics & Biostatistics
ISSN: 2155-6180
Li and Wang, J Biomet Biostat 2012, S7
DOI: 10.4172/2155-6180.S7-007
Research Article Open Access

Consistent Estimation in Generalized Linear Mixed Models with Measurement Error

He Li and Liqun Wang*
Department of Statistics, University of Manitoba, Canada

Abstract

We propose the instrumental variable method for consistent estimation of generalized linear mixed models with measurement error. This method does not require parametric assumptions for the distributions of the unobserved covariates or of the measurement errors, and it allows random effects to have any parametric distributions (not necessarily normal). We also propose simulation-based estimators for the situation where the marginal moments do not have closed forms. The proposed estimators are not only computationally attractive but also strongly root-n consistent. Moreover, the proposed estimators have a bounded influence function so they are robust against data outliers. The methodology is illustrated through simulation studies.

Keywords: Influence function; Instrumental variable; Longitudinal data; Measurement error; Mixed effects; Outliers; Robustness; Simulation-based estimator

Introduction

Generalized linear mixed models (GLMM) have been widely used in modeling longitudinal data where the response can be either discrete or continuous. Various estimation methods for GLMM have been developed in the literature [1,2,3]. However, estimation and inference in a GLMM remain very challenging when some of the covariates are not directly observed but are measured with error.

Measurement error is a common problem in medical and clinical research. For example, in epidemiologic studies time-varying covariates such as dietary intakes or other exposure variables are subject to measurement error; in HIV trials, the viral load and CD4+ cell counts are often mismeasured. It is well-known that simply substituting a proxy variable for the unobserved covariate in the model will generally lead to biased and inconsistent estimates of regression coefficients and variance components [4-6]. To account for the measurement error as well as the correlation in the longitudinal data, Wang et al. [5] proposed the simulation extrapolation (SIMEX) method to correct for the bias of the naive penalized quasi-likelihood estimator in a generalized linear mixed model with measurement error (GLMMeM), while Wang et al. [7], and Bartlett et al. [8] proposed a regression calibration (RC) approach. Although both RC and SIMEX approaches are practical and may produce satisfactory results when the measurement errors are small, it is known that in general they yield approximate but inconsistent estimators. Tosteson et al. [9] proposed a bias-corrected estimator but it was shown to be highly inefficient. Buonaccorsi et al. [10] proposed the likelihood based methods and Zhong et al. [11] studied the corrected score approach. However, the likelihood methods typically entail computation difficulties due to the multiple integrals and intractability of the likelihood function. Consequently, one usually relies on normality assumption for random effects, measurement error and residual error terms. Non- or semi-parametric approaches have also been considered for models with normal ME [12,13].

Instrumental variable (IV) method has been used by many researchers to overcome measurement error problems in cross-sectional data [6,14-17]. In practice, any variables correlated with the error-prone true covariates can serve as valid IV, e.g., a second independently measurement. Furthermore, the assumption of instrumental variable is weaker than that of replicate data because IV can be a biased observation for the true covariates [6,18].

In this paper, we propose an exact consistent estimation method for GLMMeM based on the methods of moments and instrumental variables. This method is easy to implement when the closed form of the moments exist. For the case where the marginal moments do not admit analytic forms, we propose simulation-based estimators (SBE). In particular, we use the simulation-by-parts technique of Wang [19] to construct the SBE to ensure its consistency even using finite number of simulated random points. 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, the proposed estimators have bounded influence functions so they are robust to data outliers.

The structure of the paper is as follows. In Section 2, we introduce the model and the proposed moments estimator. In Section 3, we construct the simulation-based estimator. In Section 4 we present simulation studies of finite sample performances of the proposed estimators. Finally, conclusions and discussion are contained in Section 5, whereas theoretical proofs of the theorems are given in Section 6.

The Model and Estimation

Consider the following generalized linear mixed model with measurement error (GLMMeM)

$$g^{-1}\left(E(y_{ij} \mid b_i, X_{ij})\right) = X_{ij}'\beta_x + Z_{ij}'\beta_z + B_{ij}'b_i,\ i = 1,\ldots,N,\ j = 1,\ldots,n_i \quad (1)$$

$$V(y_{ij} \mid b_i, X_{ij}) = \phi\nu(g(X_{ij}'\beta_x + Z_{ij}'\beta_z + B_{ij}'b_i)) \quad (2)$$

Where $y_{ij} \in IR$ is the j^{th} response for the i^{th} unit; $b_i \in IR^q$ is the random effect having mean zero and distribution f_b $(t;\ \theta)$ with unknown parameters $\theta \in IR^{p_b}$; $\beta_x \in IR^{p_x}$ and $\beta_z \in IR^{p_z}$ are vectors of

***Corresponding author:** Liqun Wang, Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2, Tel: (204) 474-6270; Fax: (204) 474-7621; E-mail: liqun_wang@umanitoba.ca

Received March 20, 2012; **Accepted** April 12, 2012; **Published** April 13, 2012

Citation: Li H, Wang L (2012) Consistent Estimation in Generalized Linear Mixed Models with Measurement Error. J Biomet Biostat S7:007. doi:10.4172/2155-6180.S7-007

fixed effects; $g^{-1}(\cdot)$ is a link function; $v(\cdot)$ is a known variance function and $\phi \in IR$ is a scalar parameter that may be known or unknown. It is assumed that y_{ij} given b_i are independent. Note that the estimation methods in this paper do not require the conditional distribution of y_{ij} given b_i to belong to an exponential family. Further, $Z_{ij} \in IR^{p_z}$ and $B_{ij} \in IR^q$ are known predictors observed without error; and $X_{ij} \in IR^{p_x}$ is unobservable. Instead one observes

$$W_{ij} = X_{ij} + \delta_{ij}, \tag{3}$$

where δ_{ij} is the vector of measurement errors.

Model (1) - (2) has been studied by various authors, e.g. Wang et al. [5]; Buonaccorsi et al. [10]; Zhong et al. [11]; Carroll et al. [6]. It is known that the parameters of the classical ME models generally require extra information in order to be identified [6,16]. Moreover, even if certain ME models are identifiable, additional information is useful to improve the efficiency of estimation [20]. The common source of additional data includes replicate measurements, validation data, instrumental variables, or knowledge of the measurement error distributions. Here we assume that one observes a set of instrumental variables $V_{ij} \in IR^{p_V}$ that is related to the error-prone predictor X_{ij} through

$$X_{ij} = m(V_{ij}; \gamma) + U_{ij}, \tag{4}$$

Where $m(\cdot)$ is a known function, $\gamma \in IR^{p_v}$ is a vector of unknown parameters, $U_{ij} \in IR^{p_x}$ is independent of V_{ij} and has mean zero and distribution $f_U(u;\alpha)$ with unknown parameters $\alpha \in IR^{p_u}$. Further, we assume that the ME δ_{ij} is independent of X_{ij}, V_{ij} and y_{ij}, $E(y_{ij} \mid X_{ij}, b_i) = E(y_{ij} \mid X_{ij}, V_{ij}, b_i)$ and $E(y_{ij}\, y_{ik} \mid X_{ij}, b_i) = E(y_{ij}\, y_{ik} \mid X_{ij}, V_{ij}, b_i)$ where $j \le k$. Following the convention of mixed modeling literature, throughout this paper all expectations are taken conditional on B_i and Z_i implicitly. There are no assumption on the functional forms of the distributions of X_{ij} and δ_{ij}. In this model, the observed variables are $(y_{ij}, W_{ij}', V_{ij}', Z_{ij}', B_{ij}')'$ and the parameter of interest is $\psi = (\beta_{x'}, \beta_{z'}, \theta', \alpha', \phi)'$.

To estimate all unknown parameters in the model, we first note that substituting (4) into (3) results in a usual regression equation

$$E(W_{ij} \mid V_{ij}) = m(V_{ij}; \gamma) \tag{5}$$

which can be used to obtain consistent estimator for γ by least squares method. In practice, γ can be pre-estimated using an external sample or a subset of the main sample. We denote $X_i = (X_{i1}', X_{i2}', \ldots, X_{in_i}')'$, and denote W_i, V_i, Z_i, B_i and Y_i analogously. By model assumptions and the law of iterated expectation, we have the following moments

$$\begin{aligned} k_{1,ij}(\psi) &= E(y_{ij} \mid V_i) \\ &= E\,[\,E(y_{ij} \mid b_i, X_i, V_i) \mid V_i] \\ &= E\,[\,E(y_{ij} \mid b_i, X_i) \mid V_i] \\ &= E\left[g(X_{ij}'\beta_x + Z_{ij}'\beta_z + B_{ij}'b_i) \mid V_i \right] \\ &= \int g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right] f_b(t;\theta) f_U(u;\alpha)\,dtdu, \end{aligned} \tag{6}$$

and, similarly,

$$\begin{aligned} k_{2,ijk}(\psi) &= E(y_{ij} y_{ik} \mid V_i) \\ &= E\,[\,E(y_{ij} \mid b_i, X_i)\, E \mid E(y_{ik} \mid b_i, X_i) \mid V_i] \\ &= \int g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right] \times \\ &\quad g\left[(m(V_{ik};\gamma)+u)'\beta_x + Z_{ik}'\beta_z + B_{ik}'t \right] f_b(t;\theta) f_U(u;\alpha)\,dtdu + \\ &\quad \varphi_{jk}\phi \int v\{ g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right] \} f_b(t;\theta) f_U(u;\alpha)\,dtdu, \end{aligned} \tag{7}$$

and

$$\begin{aligned} k_{3,ijk}(\psi) &= E(y_{ij} W_{ik} \mid V_i) \\ &= E(y_{ij} X_{ik} \mid V_i) \\ &= E\,[X_{ik} E(y_{ij} \mid b_i, X_i) \mid V_i] \\ &= \int (m(V_{ik};\gamma)+u)\, g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right] f_b(t;\theta) f_U(u;\alpha)\,dtdu, \end{aligned} \tag{8}$$

where $\varphi_{jk} = 1$ if $j = k$ and zero otherwise. In the following we consider three popular GLMMeM examples.

Example 2.1: Consider a linear mixed model with continuous responses and an identity link function $g(\cdot)$. Assuming U_{ij} has mean zero and variance matrix αI, and b_i has mean zero and covariance matrix Σ_b, we have the explicit form of the moments

$$\kappa_{1,ij}(\psi) = m(V_{ij};\gamma)'\beta_x + Z_{ij}'\beta_z,$$

$$\kappa_{2,ijk}(\psi) = \kappa_{1,ij}(\psi)\kappa_{1,ik}(\psi) + B_{ij}\Sigma_b B_{ik}' + \varphi_{jk}\alpha\beta_x'\beta_x + \varphi_{jk}\phi,$$

$$k_{3,ijk}(\psi) = k_{1,ij}(\psi) m(V_{ik};\gamma) + \varphi_{jk}\alpha\beta_x$$

It is worth noting that no distributional assumptions are required for U_{ij} and b_i to obtain these moments.

Example 2.2: Consider a random intercept mixed Poisson model for counts, where $\log E(y_{ij} | b_i, Z_i, X_i, B_i) = \beta_0 + \beta_x x_{ij} + \beta_z z_i + \beta_{xz} x_{ij} z_i + b_i$ and $\phi = 1$; x_{ij}, z_i and b_i are scalars. Assuming $b_i : N(0, \theta)$ and $u_{ij} : N(0, \alpha I)$, we can derive the explicit forms of the moments as

$$\kappa_{1,ij}(\psi) = \exp\left(\beta_0 + (\beta_x + \beta_{xz} z_i) m(v_{ij};\gamma) + (\beta_x^2 + \beta_{xz}^2 z_i^2)\alpha/2 + \beta_z z_i + \theta/2 \right),$$

$$\kappa_{2,ijk}(\psi) = \kappa_{1,ij}(\psi)\kappa_{1,ik}(\psi) \exp\left((\beta_x^2 + \beta_{xz}^2 z_i^2)\alpha + \theta \right) + \varphi_{jk}\kappa_{1,ij}(\psi),$$

$$\kappa_{3,ijk}(\psi) = m(v_{ik};\gamma)\kappa_{1,ij}(\psi) + \varphi_{jk}(\beta_x + \beta_{xz} z_i)\alpha\kappa_{1,ij}(\psi).$$

Example 2.3: Consider a mixed logistic model for a binary response y_{ij}, where $\phi = 1$ and $g(\cdot)$ is the logistic distribution function. For this model we find

$$\kappa_{1,ij}(\psi) = \int g\left(m(V_{ij};\gamma)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right) f_b(t;\theta) f_U(u;\alpha)\,dtdu,$$

$$\kappa_{2,ijk}(\psi) = \int g\left(m(V_{ij};\gamma)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right) g\left(m(V_{ik};\gamma)'\beta_x + Z_{ik}'\beta_z + B_{ik}'t \right) f_b(t;\theta) f_U(u;\alpha)\,dtdu,$$

$$\kappa_{3,ijk}(\psi) = \int (m(V_{ik};\gamma)+u)\, g\left((m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t \right) f_b(t;\theta) f_U(u;\alpha)\,dtdu.$$

The above integrals are intractable but can be approximated using Monte Carlo simulators. This case will be treated in the next section.

Since γ is of secondary interest, it is treated as a nuisance parameter and is estimated by nonlinear least squares (NLS) method based on equation (5) as

$$\hat{\gamma}_N = \underset{\gamma \in \Omega_\gamma}{\operatorname{argmin}} \Psi_N(\gamma) = \underset{\gamma \in \Omega_\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} r_i'(\gamma) r_i(\gamma), \tag{9}$$

where $r_i'(\gamma) = (W_{ij} - m(V_{ij};\gamma), 1 \le j \le n_i)$. Under standard regularity conditions, $\hat{\gamma}_N - \gamma_0 = O_p(N^{-1/2})$. Then we replace γ in (6)-(8) by its least squares estimator $\hat{\gamma}_N$ and denote the moments as $\hat{\kappa}_{1,ij}, \hat{\kappa}_{2,ijk}$, and $\hat{\kappa}_{3,ijk}$ $\hat{\kappa}_{3,ijk}$ correspondingly. Throughout the paper, we denote the parameter space of a parameter vector, say ψ, by $\Omega\psi$. In particular, the parameter spaces of β_x and β_z are denoted as Ω_x and Ω_z respectively. Then the method of moments estimator (MME) for ψ is defined as

$$\hat{\psi}_N = \underset{\psi\in\Omega_\psi}{argmin} Q_N(\psi) = \underset{\psi\in\Omega_\psi}{argmin}\sum_{i=1}^{N}\hat{\rho}_{i'}(\psi)A_i\hat{\rho}_i(\psi), \quad (10)$$

where $\hat{\rho}_{i'}(\psi) = (y_{ij} - \hat{\kappa}_{1,ij}(\psi), y_{ij}y_{ik} - \hat{\kappa}_{2,ijk}(\psi), y_{ij}W_{ik} - \hat{\kappa}_{3,ijk}(\psi))$ and $A_i = A(V_i)$ is a nonnegative definite matrix that may depend on V_i.

To derive the consistency and asymptotic normality of $\hat{\psi}_N$, we make the following assumptions.

Assumption 1: g (-) and v (-) are continuously differentiable; $m(v;\ \cdot)$ is a Lebegue measurable function of v and is continuously differentiable with respect to γ.

Assumption 2: $(y_i, W_i, V_i, Z_i, B_i, n_i)$, $i=1,\ldots,N$, 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, there exists a positive function $G(v,t,u)$ satisfying $E\left[\|A\|\left(\int G(V,t,u)(\|m(V,\gamma)+u\|+1)dtdu\right)^2\right] < \infty$, such that $g^2\left[(m(v,\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right] f_b(t;\theta) f_U(u;\alpha)$ and $v\left\{g\left[(m(v,\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right]\right\} f_b(t;\theta) f_U(u;\alpha)$ are bounded by $G(v,t,u)$.

Assumption 3: The parameter space is compact.

Assumption 4: $E[\rho_i(\psi) - \rho_i(\psi_0)]' A_i [\rho_i(\psi) - \rho_i(\psi_0)] = 0$ if and only if $\psi = \psi_0$.

Assumption 5: g (-)and v (-) are twice continuously differentiable; $f_b(t;\ \theta)$ and $f_U(u;\alpha)$ are twice continuously differentiable w.r.t to θ and α respectively in some open subsets $\theta_0 \in \Omega_{\theta 0} \subset \Omega_\theta$ and $\alpha \in \Omega_{\alpha 0} \in \Omega\alpha$. Furthermore, all first and second order partial derivatives of $g\left[(m(V_{ij},\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right] f_b(t;\theta) f_U(u;\alpha)$ and $v\left\{g\left[(m(V_{ij},\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right]\right\} f_b(t;\theta) f_U(u;\alpha)$ w.r.t (ψ,γ) are bounded absolutely by the positive function $G(v,t,u)$ given in Assumption 2.

Assumption 6: The matrices

$$D_\psi = E\left[\frac{\partial\rho_i'(\psi_0)}{\partial\psi} A_i \frac{\partial\rho_i(\psi_0)}{\partial\psi'}\right], \quad (11)$$

$$D_\gamma = E\left[\frac{\partial r_{i'}(\gamma_0)}{\partial\gamma}\frac{\partial r_i(\gamma_0)}{\partial\gamma'}\right] \quad (12)$$

are nonsingular.

Theorem 1: As $N\to\infty$,

1. under assumptions 1-4, $\hat{\psi}_N \xrightarrow{a.s.} \psi_0$;
2. under assumptions 1-6, $\sqrt{N}(\hat{\psi}_N - \psi_0) \xrightarrow{L} N(0, D_\psi^{-1} C D_\psi^{-1})$, where

$$C = E(C_1 C_1) \quad (13)$$

$$C_1 = \frac{\partial\rho_{i'}(\psi_0)}{\partial\psi} A_i\rho_i(\psi_0) + D_{\psi\gamma} D_\gamma^{-1}\frac{\partial r_{i'}(\gamma)}{\partial\gamma} r_i(\gamma) \quad (14)$$

$$D_{\psi\gamma} = E\left[\frac{\partial\rho_{i'}(\psi_0)}{\partial\psi} A_i \frac{\partial\rho_i(\psi_0)}{\partial\gamma'}\right] \quad (15)$$

The second term in equation (14) is the correction term due to the first-step estimation of γ. If γ_0 is known or estimated using an independent sample from the main sample, then this term vanishes and the most efficient weight is given by $A_i^{opt} = E[\rho_i(\psi_0)\rho_{i'}(\psi_0) \mid V_i]^{-1}$ [21]. In practice, direct calculation of A_i^{opt} is not feasible since it involves unknown parameters to be estimated. One possible solution is using a two-stage procedure. First, minimize $Q_N(\psi)$ using $A_i = I$ to obtain the first stage estimator $\hat{\psi}_{N1}$. Second, estimate A_i^{opt} by any nonparametric method or

$$A_i^{opt} = \left(\frac{1}{N}\sum_{i=1}^{N}\rho_i(\hat{\psi}_{N1})\rho'_i(\hat{\psi}_{N1})\right)^{-1}, \quad (16)$$

and minimizing $Q_N(\psi)$ again using A_i^{opt} to obtain the second stage estimator $\hat{\psi}_{N2}$. In practice, the calculation of A_i^{opt} may be difficult or inaccurate due to its high dimension, so one may consider using certain diagonal weight matrix. A detailed discussion on the choice of A_i^{opt} can be found in Li and Wang (2011).

In general, MME can be computed using Newton-Raphson algorithm as

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

where $\hat{\psi}^{(\tau)}$ denotes the estimate of ψ at the τ^{th} iteration,

$$\frac{\partial Q_N(\hat{\psi}^{(\tau)})}{\partial\psi} = 2\sum_{i=1}^{N}\frac{\partial\rho_{i'}(\hat{\psi}^{(\tau)})}{\partial\psi}A_i\rho_i(\hat{\psi}^{(\tau)}), \quad (17)$$

$$\frac{\partial^2 Q_N(\hat{\psi}^{(\tau)})}{\partial\psi\partial\psi'} = 2\sum_{i=1}^{N}\left[\frac{\partial\rho_{i'}(\hat{\psi}^{(\tau)})}{\partial\psi}A_i\frac{\partial\rho_i(\hat{\psi}^{(\tau)})}{\partial\psi'} + (\rho_{i'}(\hat{\psi}^{\tau})A_i\otimes I)\frac{\partial vec(\partial\rho_{i'}(\hat{\psi}^{(\tau)})/\partial\psi)}{\partial\psi'}\right]. \quad (18)$$

Since the second term in (18) has expectation zero, it can be ignored for computational simplicity.

When using the weight (16), the MME is able to safeguard against influential measurements. In particular, the influence function (IF) at a single contaminated data point v for subject l takes the form

$$IF(v;\hat{\psi}_N,F) = -D_\psi(\hat{\psi}_N(F))^{-1}\frac{\partial\hat{\rho}_{l'}(v;\hat{\psi}_N(F))}{\partial\psi}A_l(v;\hat{\psi}_N(F))\hat{\rho}_l(v;\hat{\psi}_N(F)), \quad (19)$$

where F is the underlying distribution and $D\psi$ is given in (11). If $\hat{\psi}_N$ is computed using the estimated weight (16), then analogous to Li and Wang [22] we can prove that $\|IF(v;\hat{\psi}_N,F)\|\to 0$ as $\|v\|\to\infty$. Therefore, the influence function of $\hat{\psi}_N$ is bounded and $\hat{\psi}_N$ has a redescending property, so it is robust to influential observations or outliers.

Simulation-Based Estimator

The numerical computation of MME $\hat{\psi}_N$ is straightforward if the moments in (6)-(8) admit explicit forms. However, sometimes the integrals involved in these moments are intractable. In this case, we propose a simulation-based approach. The basic idea is to replace the integrals with their Monte Carlo simulators as follows. First, generate random points t_{is} and u_{is},i = 1,2,…,N; s = 1,2,…,2S; from known densities $l(t)$ and $h(u)$. Then use the first half of the points tis and u_{is},s = 1,2,…,S to compute

$$\kappa_{1,ij}^1(\psi) = \frac{1}{S}\sum_{s=1}^{S}g\left[(m(V_{ij};\gamma)+u_{is})'\beta_x + Z_{ij}'\beta_z + B_{ij}'t_{is}\right]\frac{f_b(t_{is};\theta)f_U(u_{is};\alpha)}{l(t_{is})h(u_{is})} \quad (20)$$

$$\kappa_{2,ijk}^1(\psi) = \frac{1}{S}\sum_{s=1}^{S}g\left[(m(V_{ij};\gamma)+u_{is})'\beta_x + Z_{ij}'\beta_z + B_{ij}'t_{is}\right] g\left[(m(V_{ik};\gamma)+u_{is})'\beta_x + Z_{ik}'\beta_z + B_{ik}'t_{is}\right]\frac{f_b(t_{is};\theta)f_U(u_{is};\alpha)}{l(t_{is})h(u_{is})} + \varphi_{jk}\phi\frac{1}{S}\sum_{s=1}^{S}v\left(g\left[(m(V_{ij};\gamma)+u_{is})'\beta_x + Z_{ij}'\beta_z + B_{ij}'t_{is}\right]\right)\frac{f_b(t_{is};\theta)f_U(u_{is};\alpha)}{l(t_{is})h(u_{is})} \quad (21)$$

$$\kappa_{3,ijk}^1(\psi) = \frac{1}{S}\sum_{s=1}^{S}(m(V_{ik};\gamma)+u_{is})g\left[(m(V_{ij};\gamma)+u_{is})'\beta_x + Z_{ij}'\beta_z + B_{ij}'t_{is}\right]\frac{f_b(t_{is};\theta)f_U(u_{is};\alpha)}{l(t_{is})h(u_{is})} \quad (22)$$

and similarly use the second half of the points t_{is} and u_{is},s = S+1,S+2,…

Citation: Li H, Wang L (2012) Consistent Estimation in Generalized Linear Mixed Models with Measurement Error. J Biomet Biostat S7:007. doi:10.4172/2155-6180.S7-007

,2S and to compute $\kappa^2_{1,ij}(\psi)$, $\kappa^2_{2,ijk}(\psi)$ and $\kappa^2_{3,ijk}(\psi)$. It is easy to see that $\kappa^t_{1,ij}(\psi)$, $\kappa^t_{2,ijk}(\psi)$ and $\kappa^t_{3,ijk}(\psi)$, t = 1,2 are unbiased estimators for $\kappa_{1,ij}(\psi)$, $\kappa_{2,ijk}(\psi)$ and $\kappa_{3,ijk}(\psi)$ respectively. Finally, the simulation-based estimator (SBE) for ψ is defined as

$$\hat{\psi}_{N,S} = \underset{\psi\in\Omega_\psi}{argmin} Q_{N,S}(\psi) = \underset{\psi\in\Omega_\psi}{argmin}\sum_{i=1}^{N}\hat{\rho}_{i,1}'(\psi)A_i\hat{\rho}_{i,2}(\psi), \qquad (23)$$

where $\hat{\rho}_{i,t}(\psi) = (y_{ij} - \hat{\kappa}^t_{1,ij}(\psi), y_{ij}y_{ik} - \hat{\kappa}^t_{2,ijk}(\psi), y_{ij}W_{ik} - \hat{\kappa}^t_{3,ijk}(\psi))'$. We refer this simulation technique as simulation-by-parts since $\hat{\rho}_{i,1}(\psi)$ and $\hat{\rho}_{i,2}(\psi)$ are constructed by using two independent sets of random points. The benefit of simulation by parts is that $\hat{\rho}_{i,1}(\psi)$ and $\hat{\rho}_{i,2}(\psi)$ are conditionally independent given $(Y_i, W_i, V_i, Z_i, B_i)$ so that $Q_{N,S}(\psi)$ is an unbiased simulator for $Q_N(\psi)$ for finite S. It is worth noting that the construction of simulated moments only requires b_i and U_{ij} to have certain known parametric forms (not necessary normal). For example, one can follow Davidian and Gallant [23] and Zhang and Davidian [24] to represent the density of b_i and U_{ij} by the standard seminonparametric density which includes normal, skewed, multi-modal, fat- or thin-tailed densities. One can also impose the Tukey (g,h) family distribution [25] for b_i as well which is generated by a single transformation of the standard normal and covers a variety of distributions.

Theorem 2: Suppose that $Supp(l) \supseteq Supp(fb(\cdot;\theta))$ for all $\theta \in \Omega_{\theta 0}$, and $Supp(h) \supseteq Supp(f_U(\cdot;\alpha))$ for all $\alpha \in \Omega_{\alpha 0}$. Then for any fixed $S > 0$, as $N\to\infty$,

1. under assumptions 1-4, $\hat{\psi}_{N,S} \xrightarrow{a.s.} \psi_0$;
2. under assumptions 1-6, $\sqrt{N}(\hat{\psi}_{N,S} - \psi_0) \xrightarrow{L} N(0, D_\psi^{-1}C_S D_\psi^{-1})$, where

$$C_S = E\left(C_{1S}C_{1S}'\right), \qquad (24)$$

$$2C_{1S} = \frac{\partial\rho_{i,1}'(\psi_0)}{\partial\psi}A_i\rho_{i,2}(\psi_0) + \frac{\partial\rho_{i,2}'(\psi_0)}{\partial\psi}A_i\rho_{i,1}(\psi_0) + 2D_{\psi\gamma}D_\gamma^{-1}\frac{\partial r_i'(\gamma)}{\partial\gamma}r_i(\gamma) \qquad (25)$$

Note that the above asymptotic results do not require the simulation size S tends to infinity because we use the simulation-by-parts technique to approximate moments. This is fundamentally different from other simulation-based methods in the literature which typically require S goes to infinity to obtain consistent estimators. However, due to the approximation of marginal moments, $\hat{\psi}_{N,S}$ is generally less efficient than $\hat{\psi}_N$. In general, analogous to the Corollary 4 in Wang [19] we can show that the efficiency loss caused by simulation decreases at the rate $O(1/S)$.

Monte Carlo Simulation Studies

In this section, we evaluate the finite sample behavior of the proposed estimators, and compare them with the naive maximum likelihood estimates that ignores measurement error. We carried out Monte Carlo replications in each simulation study and reported the biases and the root mean square errors (RMSE). All computations are done in R [26] and the naive ML estimates are obtained from glmmPQL package.

In the first simulation study, we considered the mixed Poisson model in Example 2.2. We simulated δ_{ij} from $N(0,1)$, set z_i = 1 for half the sample and 0 for the remainder, and set N = 100,300 and n = 4. In addition, we set x_{ij} = 1.5 + 0.5v_{ij} + u_{ij}, v_{ij} ~ $N(0,1)$, and u_{ij} ~ $N(0,0.25)$. Table 1 reports the simulation results. For the fixed effects associated with ME,β_0, β_x and β_{xz}, the MME is almost unbiased while the naive MLE is severely downward biased and attenuated towards zero. The MME is considerably more efficient than the naive MLE in terms of smaller RMSE. With the increase of sample size from N = 100 to 300, the RMSE and biases of MME are decreasing while the ones from the naive MLE stay almost the same. For exactly measured effect β_z, the MME still provides a better estimate in terms of biases and RMSE which may because z_i interacts with x_{ij}. The naive MLE for β_z is also biased towards zero. However, with the increase of sample sizes, the biases and RMSE reduces for the naive MLE as well as the MME. For the random effect θ, surprisingly both estimators provide quite satisfactory estimators with no apparent biases.

	N = 100		*N* = 300	
Parameter	Naive MLE	MME	Naive MLE	MME
β_0 = 1.00	1.37 (1.39)	-0.22 (0.27)	1.38 (1.38)	-0.20 (0.23)
β_x = 1.00	-0.79 (0.79)	-0.05 (0.08)	-0.79 (0.79)	-0.07 (0.08)
β_z = -0.50	0.38 (0.47)	0.06 (0.17)	0.36 (0.40)	0.05 (0.16)
β_{xz} = 0.25	-0.20 (0.22)	-0.01 (0.10)	-0.19 (0.20)	-0.02 (0.08)
θ = 1.00	-0.03 (0.21)	-0.04 (0.19)	0.03 (0.28)	-0.05 (0.15)

Table 1: Biases(RMSE) for the parameter estimates in the random intercept Poisson models.

In the second simulation study, we considered a logistic model for binary responses. In particular, we adopted the following model used in the simulation studies by Wang et al. [5]:

$$logit(Pr(y_{ij} = 1|\, b_i, x_{ij}, z_{ij})) = \beta_0 + \beta_x x_{ij} + \beta_z z_{ij} + b_i \qquad (26)$$

where b_i ~ $N(0,0.5)$, z_{ij} ~ $N(0,1)$, and δ_{ij} ~ $N(0,1)$. In addition, we assumed an IV variable is observed that relates to x_{ij} though x_{ij} = 1.5 + 0.5v_{ij} +u_{ij}, v_{ij} : $N(0,1)$ and u_{ij} : N(0,0.5). In the present simulation, we selected N = 50,100 and n = 3. The closed form of the marginal moments are not available so we applied the SBE in this case. To compute the SBE, we chose the density of $N(0,2)$ to be $h(u)$ and $l(t)$, and generated independent points u_{is} and t_{is}, s = 1,…,2S, using S = 1000. The simulation results are presented in Table 2. For the fixed effects associated with ME, β_0 and β_x, SBE is almost unbiased while the naive ML is severely downward biased and attenuated towards zero. With the increase of sample size from N = 50 to 100, the RMSE and biases of MME are decreasing while the ones from the naive ML stay almost the same. This is the same findings as the ones in the first simulation study. For exactly measured effect β_z, both estimates seem to be unbiased; however, the naive ML provides a better estimates in terms of smaller biases and RMSE. With the increase of sample size, the RMSE and biases from both methods are decreasing. For the random effect, the naive ML overestimates θ with larger biases as well as RMSE. With the increase of sample size, both estimators lead to smaller biases and RMSE.

In the third simulation study, we considered the linear mixed model in Example 2.1 with N = 100,300 and n = 4. We simulated ϕ,v_{ij}, and u_{ij} independently from a standard normal distribution and set x_{ij}= 0.2+1.4 v_{ij}+ u_{ij}. The random effect was generated from either a normal distribution or a t(3) distribution. Table 3 reports the simulation results. For β_0 and β_x, the MME is almost unbiased while the naive MLE is severely biased. The MME has much smaller RMSE than MLE. With the increase of sample size from N = 100 to 300 to, the RMSE and biases of MME are decreasing while the ones from the naive MLE stay almost the same. The performance of MME and MLE for fixed effects under normal random effects and non-normal random effects are very similar. For the random effect θ, both estimators provide quite satisfactory estimators with no apparent biases. However, the RMSE from naive MLE increased significantly when the random effects were shifted from normal to t(3). For the residual error variance ϕ, naive MLE produces huge bias while MME stays unbias. These findings demonstrated that

Citation: Li H, Wang L (2012) Consistent Estimation in Generalized Linear Mixed Models with Measurement Error. J Biomet Biostat S7:007. doi:10.4172/2155-6180.S7-007

	N = 50		N = 100	
Parameter	Naive MLE	MME	Naive MLE	MME
β_0 = 0.00	1.65 (1.69)	0.02 (0.15)	1.61 (1.62)	0.01 (0.08)
β_x = 2.00	-1.31 (1.32)	0.07 (0.72)	-1.32 (1.32)	0.03 (0.12)
β_z = 1.00	-0.10 (0.22)	-0.05 (0.49)	-0.11 (0.16)	0.01 (0.22)
θ = 0.50	0.64 (1.06)	0.11 (1.09)	0.51 (0.65)	0.05 (0.13)

Table 2: Biases(RMSE) for the parameter estimates in the random intercept Logistic models.

		N =100		N =300	
Distribution	Parameter	Naive MLE	MME	Naive MLE	MME
Normal	β_0 = 0.00	0.11 (0.16)	0.01 (0.09)	0.10 (0.12)	0.01 (0.06)
	β_x = 2.00	-0.51 (0.51)	-0.05 (0.07)	-0.51 (0.51)	-0.02 (0.05)
	θ = 0.25	0.00 (0.18)	0.00 (0.25)	0.00 (0.11)	0.01 (0.17)
	ϕ =1	2.99 (3.00)	0.09 (0.46)	3.00 (3.01)	0.06 (0.31)
t(3)	β_0 = 0.00	0.10 (0.15)	0.00 (0.10)	0.10 (0.12)	0.01 (0.06)
	β_x = 2.00	-0.50 (0.51)	-0.04 (0.07)	-0.50 (0.51)	-0.02 (0.05)
	θ = 0.25	0.01 (0.38)	-0.02 (0.29)	-0.01 (0.20)	-0.02 (0.17)
	ϕ =1	2.98 (2.99)	0.05 (0.48)	2.99 (3.00)	0.05 (0.28)

Table 3: Biases(RMSE) for the parameter estimates in the random intercept Poisson models.

Concluding Remarks

We proposed consistent estimators for the generalized linear mixed models with errors in variables. These estimators are constructed by combining the method of instrumental variables and the method of moments. The strong consistency and asymptotic normality of the estimators are obtained under mild regularity conditions. The proposed approach does not require the parametric assumptions for the distributions of the unobserved covariates or the normality assumption for the random effects. In the simulation studies the proposed estimators perform well in finite sample situations.

In comparison with the regression calibration and simulation extrapolation [5,7], the proposed method is exactly consistent and computationally less intensive. It is therefore more reliable even if the measurement errors are large. Moreover, the instrumental data is usually less expensive to collect than validation data which is generally required by the regression calibration method. Unlike the likelihood-based method [10], the proposed method does not rely on normality distributional assumptions on the random effects or unobserved covariates and is easier to compute. As noted in Carroll and Stefanski [18], the assumption for an instrumental variable is weaker than a replicate for the error-prone true variable. In general, if the assumption of replicate cannot be verified, one may wish to investigate the applicability of using it as an instrumental variable. Therefore, the proposed method is less restrictive than the ones relying on replicates. A special case of measurement error is that the true unobserved covariate is known to be time-invariant (i.e., $X_{ij} = X_i$) but its surrogate W_{ij} is measured over time. In this case, W_{ij} can be used as the instruments for X_i and hence no extra data is required for the model identification and estimation. The proposed estimators can also be extended to the case of the generalized linear mixed models with Berkson measurement error. However, the use of instrumental variable is not necessary because the model is identifiable by using the first two moments (i.e., equation (6) and equation (7)). The proof is straightforward by following Wang [27].

Theoretical proofs

Proof of Theorem 1.1

By assumption 1 and the Dominated Convergence Theorem (DCT), we have the first-order Taylor expansion about γ_0.

$$Q_N(\psi) = \sum_{i=1}^{N} \rho_{i'}(\psi) A_i \rho_i(\psi) + 2\sum_{i=1}^{N} \rho_{i'}(\psi,\tilde{\gamma}) A_i \frac{\partial \rho_i(\psi,\tilde{\gamma})}{\partial \gamma'}(\hat{\gamma}_N - \gamma_0), \tag{27}$$

where $\|\tilde{\gamma}-\gamma_0\| \le \|\hat{\gamma}_N - \gamma_0\|$. Further, for any $1 \le i \le N$, by assumptions 1-3 and Cauchy-Schwartz inequality, we have

$$\|\rho_i(\psi)\|^2 \le 2\sum_j y_{ij}^2 + 2\sum_{j\le k} y_{ij}^2 y_{ik}^2 + 2\sum_{j\le k}\|y_{ij}W_{ik}\|^2$$
$$+2\sum_j\left(\int g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right] f_b(t;\theta) f_U(u;\alpha)dtdu\right)^2$$
$$+4\sum_{j\le k}\left(\int g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right]\right.$$
$$\left. g\left[(m(V_{ik};\gamma)+u)'\beta_x + Z_{ik}'\beta_z + B_{ik}'t\right] f_b(t;\theta) f_U(u;\alpha)dtdu\right)^2$$
$$+4\phi^2\sum_j\left(\int v\{g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right]\} f_b(t;\theta) f_U(u;\alpha)dtdu\right)^2$$
$$+2\sum_{j\le k}\left(\int \|m(V_{ik};\gamma)+u\| g\left[(m(V_{ij};\gamma)+u)'\beta_x + Z_{ij}'\beta_z + B_{ij}'t\right] f_b(t;\theta) f_U(u;\alpha)dtdu\right)^2$$

and therefore

$$E\sup_{\Omega_\psi}|\rho'_i(\psi)A_i\rho_i(\psi)| \le E\|A_i\|\sup_{\Omega_\psi}\|\rho_i(\psi)\|^2$$
$$\le 2n_i E\|A_i\| y_{ij}^2 + n_i(n_i+1)\left(E\|A_i\| y_{ij}^2 y_{ik}^2 + E\|A_i\|\|y_{ij}W_{ik}\|^2\right)$$
$$+2n_i E\|A_i\|\left(\int G(V_i,Z_i,t,u)dtdu\right)$$
$$+2n_i\left(n_i+1+2\sup_{\Omega_\phi}\phi^2\right)E\|A_i\|\left(\int G(V_i,Z_i,t,u)dtdu\right)^2$$
$$+n_i(n_i+1)E\|A_i\|\left(\int G(V_i,Z_i,t,u)\|m(V;\gamma_0)+u\|dtdu\right)^2$$
$$<\infty.$$

Hence by the uniform law of large numbers (ULLN),

$$\sup_{\psi\in\Omega_\psi}\left|\frac{1}{N}\sum_{i=1}^{N}\rho_{i'}(\psi)A_i\rho_i(\psi) - Q(\psi)\right| \xrightarrow{a.s.} 0, \tag{28}$$

where $Q(\psi) = E[\rho_{i'}(\psi)W_i\rho_i(\psi)]$. Similarly, by assumption 1-3 and 5 we can show

$$\left(E\sup_{\Omega_\psi,\Omega_\gamma}\left\|\rho_{i'}(\psi,\gamma)A_i\frac{\partial\rho_i(\psi,\gamma)}{\partial\gamma'}\right\|\right)^2 \le E\|A_i\|\sup_{\Omega_\psi,\Omega_\gamma}\|\rho_{i'}(\psi,\gamma)\|^2 E\|A_i\|\left\|\frac{\partial\rho_i(\psi,\gamma)}{\partial\gamma'}\right\|^2 < \infty,$$

then again by the ULLN,

$$\sup_{\Omega_\psi,\Omega_\gamma}\left\|\frac{1}{N}\sum_{i=1}^{N}\rho_{i'}(\psi,\gamma)A_i\frac{\partial\rho_i(\psi,\gamma)}{\partial\gamma'}\right\| = O(1)\ (a.s.)$$

Therefore,

$$\sup_{\Omega_\psi}\left\|\frac{1}{N}\sum_{i=1}^{N}\rho_{i'}(\psi,\tilde{\gamma})A_i\frac{\partial\rho_i(\psi,\tilde{\gamma})}{\partial\gamma'}(\hat{\gamma}_N-\gamma_0)\right\|$$
$$\le \sup_{\Omega_\psi,\Omega_\gamma}\left\|\frac{1}{N}\sum_{i=1}^{N}\rho_{i'}(\psi,\gamma)A_i\frac{\partial\rho_i(\psi,\gamma)}{\partial\gamma'}\right\|\|\hat{\gamma}_N-\gamma_0\| \xrightarrow{a.s.} 0. \tag{29}$$

It follows (27) - (??) that

$$\sup_{\Omega_\gamma}\left|\frac{1}{N}Q_N(\psi) - Q(\psi)\right| \xrightarrow{a.s.} 0. \tag{30}$$

Furthermore, $Q(\psi) = Q(\psi_0) + E[\rho_i(\psi)-\rho_i(\psi_0))'A_i(\rho_i(\psi)-\rho_i(\psi_0)]$, then by assumption 4, $Q(\psi) \ge Q(\psi_0)$ and the equality holds if and only if $\psi = \psi_0$. Thus, all conditions of Amemiya [28] Lemma 3 are satisfied and therefore $\hat{\psi}_N \xrightarrow{a.s.} \psi_0$, as N→∞.

Proof of Theorem 1.2

By assumption 5 and the DCT, the first derivative $\partial Q_N(\psi)/\partial\psi$

exists and has the first-order Taylor expansion in the open neighborhood $\Omega_{\psi 0} \in \Omega_\psi$ of ψ_0. Since $\hat{\psi}_N \xrightarrow{a.s.} \psi_0$, for sufficiently large N we have

$$\frac{\partial Q_N(\hat{\psi}_N)}{\partial \psi} = \frac{\partial Q_N(\psi_0)}{\partial \psi} + \frac{\partial^2 Q_N(\tilde{\psi}_N)}{\partial \psi \partial \psi'}(\hat{\psi}_N - \psi_0) = 0, \tag{31}$$

where $\|\tilde{\psi}_N - \psi_0\| \le \|\hat{\psi}_N - \psi_0\|$. The first and second derivative of $Q_N(\psi)$ in (31) are given in (17) and (18).

Analogous to the proof of Theorem 1.1, by assumption 1 - 5 and Cauchy-Schwartz inequality, we can verify that

$$E \sup_{\Omega_\psi} \left\| \frac{\partial \rho_{i'}(\psi)}{\partial \psi} A_i \frac{\partial \rho_i(\psi)}{\partial \psi'} \right\| \le E\|A_i\| \sup_{\Omega_\psi} \left\| \frac{\partial \rho_{i'}(\psi)}{\partial \psi} \right\|^2 < \infty$$

and

$$E \sup_{\Omega_\psi} \left\| (\rho_{i'}(\psi) A_i \otimes I) \frac{\partial vec(\partial \rho_{i'}(\psi)/\partial \psi)}{\partial \psi'} \right\| < \infty.$$

Therefore by the ULLN and Lemma 4 of Amemiya [28], we have

$$\frac{1}{2N} \frac{\partial^2 Q_N(\tilde{\psi})}{\partial \psi \partial \psi'} \xrightarrow{a.s.} E\left[\frac{\partial \rho_{i'}(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_i(\psi_0)}{\partial \psi'} + (\rho_{i'}(\psi_0) A_i \otimes I) \frac{\partial vec(\partial \rho_{i'}(\psi_0)/\partial \psi)}{\partial \psi'} \right] = D_\psi \tag{32}$$

where D_ψ is given in (11) and the second equality holds because

$$E\left[(\rho_{i'}(\psi_0) A_i \otimes I) \frac{\partial vec(\partial \rho_{i'}(\psi_0)/\partial \psi)}{\partial \psi'} \right] = 0.$$

Then by assumption 6 and (32), we rearrange (31) as

$$\sqrt{N}(\hat{\psi}_N - \psi_0) = (2D_\psi)^{-1} \left(-\frac{1}{\sqrt{N}} \frac{\partial Q_N(\psi_0)}{\partial \psi} \right) \tag{33}$$

For by assumption 5 and DCT, we have the first-order Taylor expansion of $\frac{\partial Q_N(\psi_0)}{\partial \psi}$ about γ_0:

$$\frac{\partial Q_N(\psi_0)}{\partial \psi} = 2\sum_{i=1}^N \frac{\partial \rho_{i'}(\psi_0)}{\partial \psi} A_i \rho_i(\psi_0) + \frac{\partial^2 \tilde{Q}_N(\psi_0)}{\partial \psi \partial \gamma'}(\hat{\gamma} - \gamma_0), \tag{34}$$

where $\|\tilde{\gamma} - \gamma_0\| \le \|\hat{\gamma} - \gamma_0\|$ and

$$\frac{\partial^2 \tilde{Q}_N(\psi_0)}{\partial \psi \partial \gamma'} = 2\sum_{i=1}^N \left[\begin{array}{l} \frac{\partial \rho_{i'}(\psi_0, \tilde{\gamma})}{\partial \psi} A_i \frac{\partial \rho_i(\psi_0, \tilde{\gamma})}{\partial \gamma'} \\ + (\rho_{i'}(\psi_0, \tilde{\gamma}) A_i \otimes I) \frac{\partial vec(\partial \rho_{i'}(\psi_0, \tilde{\gamma})/\partial \psi)}{\partial \gamma'} \end{array} \right].$$

Similarly to the derivation of (32), we can show

$$\frac{1}{2N} \frac{\partial^2 \tilde{Q}_N(\psi_0)}{\partial \psi \partial \gamma'} \xrightarrow{a.s.} E\left[\frac{\partial \rho_{i'}(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_i(\psi_0)}{\partial \gamma'} \right] = D_{\psi\gamma}. \tag{35}$$

Then by (33)-(35), we have

$$\sqrt{N}(\hat{\psi}_N - \psi_0) = D_\psi^{-1} \left(-N^{-1/2} \sum_{i=1}^N \frac{\partial \rho_{i'}(\psi_0)}{\partial \psi} A_i \rho_i(\psi_0) \right) + D_\psi^{-1} D_{\psi\gamma} \sqrt{N}(\hat{\gamma} - \gamma_0) \tag{36}$$

Therefore, if $D_{\psi\gamma}$=0 we can ignore the effect of $\hat{\gamma}$ and simply treated it as a known constant. If $D_\psi \neq 0$, we need to make some adjustments to the asymptotic variance of . Since , we have the first-order Taylor expansion in the open neighborhood of γ_0

$$\frac{\partial \Psi(\hat{\gamma})}{\partial \gamma} = \frac{\partial \Psi(\gamma_0)}{\partial \gamma} + \frac{\partial^2 \Psi(\tilde{\gamma})}{\partial \gamma \partial \gamma'}(\hat{\gamma}_N - \gamma_0) = 0, \tag{37}$$

By assumption 6, we can have the following representation of $\hat{\gamma}_N$

$$\sqrt{N}(\hat{\gamma} - \gamma_0) = D_\gamma^{-1} \left(-N^{1/2} \sum_{i=1}^N \frac{\partial r_{i'}(\gamma)}{\partial \gamma} r_i(\gamma) \right) = N^{-1/2} \sum_{i=1}^N \left(-D_\gamma^{-1} \frac{\partial r_{i'}(\gamma)}{\partial \gamma} r_i(\gamma) \right), \tag{38}$$

where D_γ is given in (12). Then plug it back into (36), we have

$$\sqrt{N}(\hat{\psi}_N - \psi_0) = -D_\psi^{-1} N^{-1/2} \sum_{i=1}^N \left(\frac{\partial \rho_{i'}(\psi_0)}{\partial \psi} A_i \rho_i(\psi_0) + D_{\psi\gamma} D_\gamma^{-1} \frac{\partial r_{i'}(\gamma)}{\partial \gamma} r_i(\gamma) \right)$$

Finally, the theorem follows from (31) - (??), CLT and Slutsky's Theorem.

Proof of Theorem 2.1

By assumption 1 and the DCT, $Q_{N,S}(\psi)$ has the first-order Taylor expansion about γ_0,

$$Q_{N,S}(\psi) = \sum_{i=1}^N \rho_{i,1}'(\psi) A_i \rho_{i,2}(\psi)$$

$$+ \sum_{i=1}^N \left[\rho_{i,1}'(\psi, \tilde{\gamma}) A_i \frac{\partial \rho_{i,2}(\psi, \tilde{\gamma})}{\partial \gamma'} + \rho_{i,2}'(\psi, \tilde{\gamma}) A_i \frac{\partial \rho_{i,1}(\psi, \tilde{\gamma})}{\partial \gamma'} \right] (\hat{\gamma}_N - \gamma_0), \tag{39}$$

where $\|\tilde{\gamma}_N - \gamma_0\| \le \|\hat{\psi}_N - \psi_0\|$. Since $\rho_{i,1}$ and $\rho_{i,2}$ are conditionally independent given $(Y_i, W_i, V_i, Z_i, B_i)$, analogous to the proof of Theorem 1.1, by assumptions 1-3 and Cauchy-Schwartz inequality, we have

$$E\left(\sup_\Gamma |\rho_{i,1}'(\psi) A_i \rho_{i,2}(\psi)| \right) < \infty.$$

Hence by the ULLN, $\frac{1}{N}\sum_{i=1}^N \rho_{i,1}'(\psi) A_i \rho_{i,2}(\psi) \xrightarrow{a.s.} E\rho_{i,1}'(\psi) W_i \rho_{i,2}(\psi)$ uniformly in $\psi \in \Gamma$, where

$$E\rho_{i,1}'(\psi) W_i \rho_{i,2}(\psi) = E[E(\rho_{i,1}'(\psi) | Y_i, W_i, V_i, Z_i, B_i)$$
$$W_i E(\rho_{i,2}'(\psi) | Y_i, W_i, V_i, Z_i, B_i)] = Q(\psi).$$

Similar to proof of Theorem 1.1, we can show that

$$\sup_\Gamma \left\| \frac{1}{N} \sum_{i=1}^N \rho_{i,2}'(\psi, \tilde{\gamma}) A_i \frac{\partial \rho_{i,1}(\psi, \tilde{\gamma})}{\partial \gamma'} (\hat{\gamma}_N - \gamma_0) \right\|$$

$$\le \sup_{(\Gamma,\Upsilon)} \left\| \frac{1}{N} \sum_{i=1}^N \rho_{i,2}'(\psi, \tilde{\gamma}) A_i \frac{\partial \rho_{i,1}(\psi, \tilde{\gamma})}{\partial \gamma'} \right\| \|\hat{\gamma}_N - \gamma_0\| \xrightarrow{a.s.} 0. \tag{40}$$

It then follows that

$$\sup_\Gamma \left| \frac{1}{N} Q_{N,S}(\psi) - Q(\psi) \right| \xrightarrow{a.s.} 0. \tag{41}$$

It has been proved previously that $Q(\psi)$ attains a unique minimum at $\psi_0 \in \Gamma$. Therefore, by Lemma 3 of Amemiya (1973), $\hat{\psi}_{N,S} \xrightarrow{a.s.} \psi_0$, as $N \to \infty$.

Proof of Theorem 2.2

For sufficiently large N, by assumption 5 we have the first-order Taylor expansion of $\partial Q_{N,S}(\psi)/\partial \psi$ about ψ_0:

$$\frac{\partial Q_{N,S}(\psi_0)}{\partial \psi} + \frac{\partial^2 Q_{N,S}(\tilde{\psi})}{\partial \psi \partial \psi'}(\hat{\psi}_{N,S} - \psi_0) = 0, \tag{42}$$

where $\|\tilde{\psi} - \psi_0\| \le \|\hat{\psi}_{N,S} - \psi_0\|$ and the first and second derives are given by

$$\frac{\partial Q_{N,S}(\psi)}{\partial \psi} = \sum_{i=1}^N \left(\frac{\partial \hat{\rho}_{i,1}'(\psi)}{\partial \psi} A_i \hat{\rho}_{i,2}(\psi) + \frac{\partial \hat{\rho}_{i,2}'(\psi)}{\partial \psi} A_i \hat{\rho}_{i,1}(\psi) \right)$$

and

$$\frac{\partial^2 Q_{N,S}(\psi)}{\partial \psi \partial \psi'} = \sum_{i=1}^N \left[\frac{\partial \hat{\rho}_{i,1}'(\psi)}{\partial \psi} A_i \frac{\partial \hat{\rho}_{i,2}(\psi)}{\partial \psi'} + (\hat{\rho}_{i,2}'(\psi) A_i \otimes I) \frac{\partial vec(\partial \hat{\rho}_{i,1}'(\psi)/\partial \psi)}{\partial \psi'} \right]$$

$$+ \sum_{i=1}^N \left[\frac{\partial \hat{\rho}_{i,2}'(\psi)}{\partial \psi} A_i \frac{\partial \hat{\rho}_{i,1}(\psi)}{\partial \psi'} + (\hat{\rho}_{i,1}'(\psi) A_i \otimes I) \frac{\partial vec(\partial \hat{\rho}_{i,2}'(\psi)/\partial \psi)}{\partial \psi'} \right].$$

Similar to the derivation of (32), we can show $\frac{1}{N} \frac{\partial^2 Q_{N,S}(\psi)}{\partial \psi \partial \psi'}$ converges to

$$E\left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_{i,2}(\psi_0)}{\partial \psi'} + (\rho_{i,2}'(\psi_0) A_i \otimes I) \frac{\partial vec(\partial \rho_{i,1}'(\psi_0)/\partial \psi)}{\partial \psi'} \right]$$

$$+E\left[\frac{\partial \rho_{i,2}'(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_{i,1}(\psi_0)}{\partial \psi'} + (\rho_{i,1}'(\psi_0) A_i \otimes I) \frac{\partial vec(\partial \rho_{i,2}'(\psi_0)/\partial \psi)}{\partial \psi'}\right],$$

uniformly for all $\psi \in \Gamma$. Since

$$E\left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_{i,2}(\psi_0)}{\partial \psi'}\right] = E\left[\frac{\partial \rho_i'(\psi_0)}{\partial \psi} A_i \frac{\partial \rho_i(\psi_0)}{\partial \psi'}\right] = D_\psi$$

and

$$E\left[(\rho_{i,2}'(\psi_0) A_i \otimes I) \frac{\partial vec(\partial \rho_{i,1}'(\psi_0)/\partial \psi)}{\partial \psi'}\right] = 0,$$

we have

$$\frac{1}{N} \frac{\partial^2 Q_{N,S}(\psi)}{\partial \psi \partial \psi'} \xrightarrow{a.s.} 2D. \quad (43)$$

Again, $\partial Q_{N,S}(\psi_0) \partial \psi$ has the first-order Taylor expansion about γ_0:

$$\frac{\partial Q_{N,S}(\psi_0)}{\partial \psi} = \sum_{i=1}^{N} \left[\frac{\partial \rho_{i,1}'(\psi_0)}{\partial \psi} A_i \rho_{i,2}'(\psi_0) + \frac{\partial \rho_{i,2}'(\psi_0)}{\partial \psi} A_i \rho_{i,1}'(\psi_0)\right] + \frac{\partial^2 \tilde{Q}_{N,S}(\psi_0)}{\partial \psi \partial \gamma'} (\hat{\gamma}_N - \gamma_0). \quad (44)$$

Finally, Analogous to the proof of Theorem 1.2, the results follows from (38), (42)-(44), CLT and Slutsky's Theorem.

References

1. Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88: 9-25.
2. Durbin J, Koopman SJ (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84: 669-684.
3. Rabe-Hesketh S, Skrondal A, Pickles A (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. Stata J 2: 1-21.
4. Wang N, Davidian M (1996) A note on covariate measurement error in nonlinear mixed effects models. Biometrika 83: 801-812.
5. Wang N, Lin X, Gutierrez RG, Carroll RJ (1998) Bias analysis and SIMEX approach in generalized linear mixed measurement error models. J Am Stat Assoc 93: 249-261.
6. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C (2006) Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition, Chapman & Hall, London.
7. Wang N, Lin X, Guttierrez RG (1999) A bias correction regression calibration approach in generalized linear mixed measurement error models. Commun Stat Theory Methods 28: 217-232.
8. Bartlett JW, de Stavola BL, Frost C (2009) Linear mixed models for replication data to efficiently allow for covariate measurement error. Stat Med 28: 3158-3178.
9. Tosteson T, Buonaccorsi JP, Demidenko E (1998) Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. Stat Med 17: 1959-1971.
10. Buonaccorsi J, Demidenko E, Tosteson T (2000) Estimation in longitudinal random effects models with measurement error. Stat Sin 10: 885-903.
11. Zhong XP, Fung WK, Wei BC (2002) Estimation in linear models with random effects and errors-invariables. Ann Inst Statist Math 54: 595-606.
12. Tsiatis AA, Davidian M (2001) A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika 88: 447-458.
13. Pan W, Zeng D, Lin X (2009) Estimation in semiparametric transition measurement error models for longitudinal data. Biometrics 65: 728-736.
14. Fuller WA (1987) Measurement Error Models. New York: John Wiley & Sons.
15. Buzas JS, Stefanski LA (1996) Instrumental variable estimation in generalized measurement error models. J Am Stat Assoc 91: 999-1006.
16. Schennach MS (2007) Instrumental Variable Estimation of Nonlinear Errors-in-variables Models. Econometrica 75: 201-239.
17. Wang L, Hsiao C (2011) Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. J Econom 165: 30-44.
18. Carroll RJ, Stefanski LA (1994) Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Stat Med 13: 1265-1282.
19. Wang L (2004) Estimation of Nonlinear Models with Berkson Measurement Errors. Ann Statist 32: 2559-2579.
20. Schneeweiss H, Augustin T (2005) Some recent advances in measurement error models and methods. All Stat Arch 90: 183-197.
21. Abarin T, Wang L (2006) Comparison of GMM with Second-order Least Squares Estimator in Nonlinear Models. Far East Journal of Theoretical Statistics 20: 179-196.
22. Li H, Wang L (2011) A consistent simulation-based estimator in generalized linear mixed models. J Stat Comput Simul 1-19.
23. Davidian M, Gallant A R (1993) The nonl inear mixed effects model with a smooth random effects density. Biometrika 80: 475-488.
24. Zhang D, Davidian M (2001) Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics 57: 795-802.
25. Field CA, Genton MG (2006) The multivariate g-and-h distribution. Technometrics 48: 104-111.
26. 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
27. Wang L (2007) A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models. Can J Stat 35: 233-248.
28. Amemiya T (1973) Regression analysis when the dependent variable is truncated normal. Econometrica 41: 997-1016.
29. McFadden D (1989) A method of simulated moments for estimation of discrete response models with out numerical integration. Econometrica 57: 995-1026.

This article was originally published in a special issue, **Medical statistics: Clinical and experimental research** handled by Editor(s). Dr. Herbert Pang, Duke University, USA.