The Canadian Journal of Statistics
Vol. 35, No. 2, 2007, Pages 233–248
La revue canadienne de statistique

A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models

Liqun WANG

Key words and phrases: Fixed effects; least squares; longitudinal data; measurement error; minimum distance estimator; random effect; repeated measurements; simulation-based estimator.

MSC 2000: Primary 62J02, 62H12; secondary 65C60, 65C05.

Abstract: Mixed effects models and Berkson measurement error models are widely used. They share features which the author uses to develop a unified estimation framework. He deals with models in which the random effects (or measurement errors) have a general parametric distribution, whereas the random regression coefficients (or unobserved predictor variables) and error terms have nonparametric distributions. He proposes a second-order least squares estimator and a simulation-based estimator based on the first two moments of the conditional response variable given the observed covariates. He shows that both estimators are consistent and asymptotically normally distributed under fairly general conditions. The author also reports Monte Carlo simulation studies showing that the proposed estimators perform satisfactorily for relatively small sample sizes. Compared to the likelihood approach, the proposed methods are computationally feasible and do not rely on the normality assumption for random effects or other variables in the model.

Une stratégie d'estimation commune pour les modèles non linéaires à effets mixtes et les modèles d'erreur de mesure de Berkson

Résumé : Les modèles à effets mixtes et les modèles d'erreur de mesure de Berkson sont très usités. Ils partagent certaines caractéristiques que l'auteur met à profit pour élaborer une stratégie d'estimation commune. Il considère des modèles dans lesquels la loi des effets aléatoires (ou des erreurs de mesure) est paramétrique tandis que celles des coefficients de régression aléatoires (ou de variables exogènes non observées) et des termes d'erreur ne le sont pas. Il propose une estimation des moindres carrés au second ordre et une approche par simulation fondées sur les deux premiers moments conditionnels de la variable endogène, sachant les variables exogènes observées. Les deux estimateurs s'avèrent convergents et asymptotiquement gaussiens sous des conditions assez générales. L'auteur fait aussi état d'études de Monte-Carlo attestant du bon comportement des deux estimations dans des échantillons relativement petits. Les méthodes proposées ne posent aucune difficulté particulière au plan numérique et au contraire de l'approche par vraisemblance, ne supposent ni la normalité des effets aléatoires, ni celle des autres variables du modèle.

1. INTRODUCTION

Mixed effects models and measurement error models are two classes of widely used statistical models in many scientific fields, e.g., in agriculture, biological and biomedical sciences, econometrics, environmental science, epidemiology and psychology. In econometrics they are usually called, respectively, panel data models and errors-in-variables models. Historically, these two classes of models have completely different origins and interpretations. Consequently, their statistical inference procedures have been studied separately in the literature.

Generally speaking, mixed effects models are used to analyze longitudinal or repeated measurement data arising from, e.g., clinical trials or pharmacokinetic and pharmacodynamic studies (Davidian & Giltinan 1995; Vonesh & Chinchilli 1997; Lindsey 1999), whereas panel data models are used to analyze longitudinal data arising in economics and business (Chamberlain 1984; Arellano & Honoré 2001; Hsiao 2003). In these models, the fixed and random effects are interpreted as common and individual specific parameters respectively and are treated as fixed or random unknown parameters. On the other hand, researchers are often interested in relations between a response variable and several predictor variables, some of which are unobservable or measured with substantial errors. For example, an epidemiologist studies lung cancer incidence of the residents in a city in relation to their exposures to certain pollutants in the air. While the amount of the pollutants are measured at certain monitoring stations in the city, the actual individual exposures vary around the observed values. This is a typical situation of Berkson-type measurement errors (Fuller 1987; Gustafson 2004).

In this paper, we show that the mixed effects models and Berkson measurement error models share a common statistical structure and can therefore be treated in a unified framework. Specifically, consider the general model

$$y_{it} = g(x_{it}, \xi_i, \theta) + \varepsilon_{it}, \quad t = 1, \ldots, T_i,\ i = 1, \ldots, N, \tag{1}$$

where $y_{it} \in \mathbb{R}$ is the response variable, $x_{it} \in \mathbb{R}^k$ is the predictor variable, $\xi_i \in \mathbb{R}^\ell$ and $\theta \in \mathbb{R}^p$ are unknown parameters, and ε_{it} is the random error. Further, suppose

$$\xi_i = Z_i \varphi + \delta_i, \tag{2}$$

where $Z_i \in \mathbb{R}^{\ell \times q}$ is a matrix of design variables, $\varphi \in \mathbb{R}^q$ is the vector of fixed effects and $\delta_i \in \mathbb{R}^\ell$ is the vector of random effects, which is independent of Z_i and $X_i = (x_{i1}, \ldots, x_{iT_i})$. In addition, δ_i, $i = 1, \ldots, N$ are assumed to be independent and identically distributed with density $f_\delta(u, \psi)$, where $\psi \in \mathbb{R}^r$ is an unknown parameter. Finally, the random errors ε_{it}, $t = 1, \ldots, T_i$ are conditionally independent given X_i, Z_i, independent across $i = 1, \ldots, N$, and satisfy $\mathrm{E}\,(\varepsilon_{it} \,|\, X_i, Z_i, \delta_i) = 0$ and $\mathrm{E}\,(\varepsilon_{it}^2 \,|\, X_i, Z_i, \delta_i) = \sigma_\varepsilon^2$. In general, g is nonlinear in either or both of the predictor variables and unknown parameters. Since there is no assumption about the functional form of the distributions of X_i, Z_i and ε_{it}, (1) and (2) represent a semiparametric model. In this model, only y_{it}, X_i, Z_i are observable. Model (1) and (2) become a measurement error model if ξ_i represents the unobserved predictor variable. In particular, if $q = 1$ and $\varphi = 1$, then (2) is the so-called Berkson measurement error model (Berkson 1950; Fuller 1987; Gustafson 2004). In its general form, (2) represents the regression calibration model (Carroll, Ruppert & Stefanski 1995). Therefore, (1) and (2) incorporate both mixed effects and Berkson measurement error models.

The nonlinear mixed effects models have been intensively studied in recent years. The mainstream of the research focuses on the (normal) likelihood approach (e.g., Lindstrom & Bates 1990; Davidian & Gallant 1993; Ke & Wang 2001; Vonesh, Wang, Nie & Majumdar 2002; Wu 2002; Daimon & Goto 2003; Lai & Shih 2003a), though a nonparametric method has also been considered recently (Lai & Shih 2003b). In econometrics, the research concentrates on the likelihood approach and the generalized method of moments for dynamic panel data models (Wooldridge 1999; Arellano & Honoré 2001; Hsiao, Peseran & Tahmiscioglu 2002). The main challenge with the likelihood approach is that the numerical computation of the maximum likelihood estimators is usually difficult or intractable, especially in the case of multivariate random effects. Consequently, most existing approximate likelihood methods rely on the normality assumption for random effects and other variables in the model, which is not always realistic in applications. See, e.g., Hartford & Davidian (2000).

Despite their practical importance, nonlinear models with Berkson measurement errors have been less intensively studied in the literature. Two commonly used methods for estimation are regression calibration and simulation extrapolation (SIMEX) (Carroll, Ruppert & Stefanski 1995). These methods, however, yield only approximately consistent estimators and are therefore applicable to small measurement error situations. Another stream of research is consistent estimation without the assumption of normality, for example, Wang (2004) proposed a minimum distance estimator and a simulation-based estimator based on the first two conditional moments of the response variable given the observed predictor variables. The goal of the present paper is to extend the methods of Wang (2004) to models (1) and (2), which incorporate the nonlinear mixed effects and general regression calibration models for Berkson measurement errors. Specifically, we propose a second-order least squares estimator for parameters $\gamma = (\theta^\top, \varphi^\top, \psi^\top, \sigma^2)^\top$ based on the first two conditional moments of y_{it} given the observed covariates (X_i, Z_i). This estimator can be easily computed if the closed forms of the two conditional moments are available. For the more general case where the closed forms are difficult or impossible to obtain, we propose a simulation-based estimator. We show that both estimators are consistent and asymptotically normally distributed under fairly general regularity conditions. Moreover, Monte Carlo simulation studies of the finite sample performance of the proposed estimators and a real data application are also presented.

The paper is organized as follows. Section 2 gives some examples to motivate our estimation methods. Section 3 introduces the second-order least squares estimator and gives its consistency and asymptotic normality. Section 4 presents the simulation-based estimator and its asymptotic properties. Monte Carlo simulation studies and a real data application are presented in Section 5. The regularity conditions are given in Section 6, and conclusions and discussion of possible extensions of the proposed methods are given in Section 7. Finally, proofs of the theorems are provided in the Appendix.

2. EXAMPLES AND MOTIVATION

In this section we motivate our approach using two simple examples. It is well known that the linear measurement error model under normal distributions is nonidentifiable. Interestingly, nonlinear models with Berkson measurement errors are generally identifiable (Rudemo, Ruppert & Streibig 1989). Moreover, the models can usually be identified using the first two conditional moments of the response variable given the observed predictor variables (Wang 2003, 2004). In this section, we demonstrate that this remains true for nonlinear mixed effects models. To simplify the notation, we consider the cases where the random effects δ_i have zero mean and the unknown parameters in their distributions consist of variances and covariances only, so that $\psi = (\psi_i, 1 \le i \le \ell, \psi_{ij}, 1 \le i < j \le \ell)^\top$. Moreover, we denote the conditional expectation given X_i, Z_i as $\mathrm{E}_i(\,\cdot\,) = \mathrm{E}\,(\,\cdot \mid X_i, Z_i)$.

Example 1. First consider an exponential model $y_{it} = \xi_{1i}\exp(-\xi_{2i}x_{it}) + \varepsilon_{it}$ and $\xi_i = \varphi + \delta_i$ (so that $Z_i = I_2$ is a two-dimensional identity matrix). For this model, we assume that δ_i is normally distributed with zero mean and variances and covariance $\psi = (\psi_1, \psi_2, \psi_{12})^\top$. Then the first two moments of y_{it} are

$$\begin{aligned}\mathrm{E}_i(y_{it}) &= \mathrm{E}_i(\varphi_1 + \delta_{1i})\exp\{-x_{it}(\varphi_2 + \delta_{2i})\} + \mathrm{E}_i(\varepsilon_{it}) \\ &= (\varphi_1 - \psi_{12}x_{it})\exp(-\varphi_2 x_{it} + \psi_2 x_{it}^2/2), \qquad (3)\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}_i(y_{it}y_{is}) &= \mathrm{E}_i(\varphi_1 + \delta_{1i})^2\exp\{-(\varphi_2 + \delta_{2i})(x_{it} + x_{is})\} + \mathrm{E}_i(\varepsilon_{it}\varepsilon_{is}) \\ &= \left[\psi_1 + \{\varphi_1 - \psi_{12}(x_{it} + x_{is})\}^2\right] \\ &\quad \times \exp\{-\varphi_2(x_{it} + x_{is}) + \psi_2(x_{it} + x_{is})^2/2\} + \sigma_{its}, \qquad (4)\end{aligned}$$

where $\sigma_{its} = \sigma_\varepsilon^2$ if $t = s$, and zero otherwise. Since (3) is a usual nonlinear regression equation, it is clear that φ_1, φ_2, ψ_2 and ψ_{12} can be consistently estimated by the least squares method. Similarly, ψ_1 can be consistently estimated by applying the least squares method to (4) with $t \ne s$, and σ_ε^2 can be estimated by (4) with $t = s$. Therefore the model is identifiable using the first two moments of y_{it} given X_i, Z_i.

Example 2. Now consider the growth model studied by Lindstrom & Bates (1990), and many other researchers,

$$y_{it} = \frac{\xi_i}{1 + \theta_1\exp(\theta_2 x_{it})} + \varepsilon_{it}, \qquad (5)$$

where $\xi_i = \varphi + \delta_i$. For this model the first two conditional moments of y_{it} are

$$\mathrm{E}_i(y_{it}) = \frac{\varphi}{1 + \theta_1 \exp(\theta_2 x_{it})}, \tag{6}$$

$$\mathrm{E}_i(y_{it} y_{is}) = \frac{\varphi^2 + \psi_1}{\{1 + \theta_1 \exp(\theta_2 x_{it})\}\{1 + \theta_1 \exp(\theta_2 x_{is})\}} + \sigma_{its}, \tag{7}$$

where σ_{its} is defined as in the previous example. Again, θ_1, θ_2 and φ can be consistently estimated by (6) and the nonlinear least squares method, while ψ_1 and σ_ε^2 can be consistently estimated by (7). Hence this model is identifiable using the first two moments of y_{it}. In Section 5, this model will be applied to the well-known orange tree data (Draper & Smith 1998).

From the above examples it is easy to see that in many situations parameters in nonlinear mixed effects models can be identified and consistently estimated using the first two conditional moments of y_{it} given X_i, Z_i. Although the closed forms of the conditional moments can be obtained in both examples, it is easy to see that the identifiability holds more generally. In fact, identifiability can always be achieved by imposing appropriate restrictions on unknown parameters, as has usually been done in practice. Unfortunately, given its theoretical and practical importance, general solutions to identifiability of nonlinear mixed effects models do not exist. In practice, it is usually done in an heuristic way.

3. SECOND-ORDER LEAST SQUARES ESTIMATOR

Motivated by the examples in the previous section, we consider a minimum distance estimator for models (1) and (2) based on the first two moments of the response variable. Let $\gamma = (\theta^\top, \varphi^\top, \psi^\top, \sigma^2)^\top$ denote the vector of model parameters and $\Gamma = \Theta \times \Phi \times \Psi \times \Sigma \subset \mathbb{R}^{p+q+r+1}$ the corresponding parameter space which is assumed to be compact. Then under the model assumptions given in Section 1, the first two conditional moments of y_{it} given the observed covariates X_i, Z_i are

$$\begin{aligned}\mu_{it}(\gamma) &= \mathrm{E}_\gamma(y_{it} \mid X_i, Z_i) \\ &= \int g(x_{it}, u, \theta) f_\delta(u - Z_i\varphi; \psi)\, du, \end{aligned} \tag{8}$$

$$\begin{aligned}\nu_{its}(\gamma) &= \mathrm{E}_\gamma(y_{it} y_{is} \mid X_i, Z_i) \\ &= \int g(x_{it}, u, \theta) g(x_{is}, u, \theta) f_\delta(u - Z_i\varphi; \psi)\, du + \sigma_{its}, \end{aligned} \tag{9}$$

where $\sigma_{its} = \sigma_\varepsilon^2$ if $t = s$, and zero if $t \neq s$. Throughout this paper all integrals are taken to be over the space $\mathbb{R}^\ell$. Then the second-order least squares estimator (SLS) for γ is defined as

$$\hat\gamma_N = \underset{\gamma\in\Gamma}{\operatorname{argmin}} Q_N(\gamma) = \underset{\gamma\in\Gamma}{\operatorname{argmin}} \sum_{i=1}^N \rho_i^\top(\gamma) W_i \rho_i(\gamma), \tag{10}$$

where $\rho_i(\gamma) = \big(y_{it} - \mu_{it}(\gamma), 1 \le t \le T_i, y_{it}y_{is} - \nu_{its}(\gamma), 1 \le t \le s \le T_i\big)^\top$ and $W_i = W(X_i, Z_i)$ is a nonnegative definite matrix which may depend on X_i, Z_i.

Now we investigate the asymptotic properties of $\hat\gamma_N$. To simplify the notation, we present our theoretical results for the case where $T_i = T$, $i = 1, \ldots, N$. The extension of the results to more general cases will be discussed in Section 7. The regularity conditions for the consistency and asymptotic normality of $\hat\gamma_N$ are given in Section 6. In particular, Assumption A3 and the dominated convergence theorem imply that the partial derivatives

$$\frac{\partial \rho_i^\top(\gamma)}{\partial\gamma} = -\left(\frac{\partial \mu_{it}(\gamma)}{\partial\gamma}, 1 \le t \le T, \frac{\partial \nu_{its}(\gamma)}{\partial\gamma}, 1 \le t \le s \le T\right)$$

exist and are as given after A4 of Section 6. Throughout the paper the true parameter value of models (1) and (2) is denoted by $\gamma_0 \in \Gamma$.

THEOREM 1. *As $N \to \infty$, the second-order least squares estimator SLS $\hat{\gamma}_N$ has the following properties:*

1. *Under* A1–A2, $\hat{\gamma}_N \xrightarrow{\text{a.s.}} \gamma_0$.
2. *Under* A1–A4, $\sqrt{N}\,(\hat{\gamma}_N - \gamma_0) \xrightarrow{L} \mathsf{N}(0, B^{-1}CB^{-1})$, *where*
$$C = \mathrm{E}\left\{\frac{\partial \rho_i^\top(\gamma_0)}{\partial \gamma} W_i \rho_i(\gamma_0)\rho_i^\top(\gamma_0) W_i \frac{\partial \rho_i(\gamma_0)}{\partial \gamma^\top}\right\} \tag{11}$$
and
$$B = \mathrm{E}\left\{\frac{\partial \rho_i^\top(\gamma_0)}{\partial \gamma} W_i \frac{\partial \rho_i(\gamma_0)}{\partial \gamma^\top}\right\}. \tag{12}$$
Furthermore, with probability one,
$$B = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \left\{\frac{\partial \rho_i^\top(\hat{\gamma}_N)}{\partial \gamma} W_i \frac{\partial \rho_i(\hat{\gamma}_N)}{\partial \gamma^\top}\right\} \tag{13}$$
and
$$4C = \lim_{N\to\infty} \frac{1}{N} \frac{\partial Q_N(\hat{\gamma}_N)}{\partial \gamma} \frac{\partial Q_N(\hat{\gamma}_N)}{\partial \gamma^\top}, \tag{14}$$
where
$$\frac{\partial Q_N(\gamma)}{\partial \gamma} = 2\sum_{i=1}^N \frac{\partial \rho_i^\top(\gamma)}{\partial \gamma} W_i \rho_i(\gamma).$$
3. *The above results hold if* A3 *is replaced by* A3′.

In the rest of this section, we briefly discuss the choice of the weighting matrix W_i in the computation of $\hat{\gamma}_N$. First, theoretically a natural question is how to choose W_i to obtain the most efficient estimator. To answer this question, we rewrite C as

$$C = \mathrm{E}\left\{\frac{\partial \rho_i^\top(\gamma_0)}{\partial \gamma} W_i V_i W_i \frac{\partial \rho_i(\gamma_0)}{\partial \gamma^\top}\right\},$$

where

$$V_i = \mathrm{E}\left\{\rho_i(\gamma_0)\rho_i^\top(\gamma_0) \mid X_i, Z_i\right\}.$$

Then by the matrix form of the Cauchy–Schwartz inequality we have

$$B^{-1}CB^{-1} \geq \mathrm{E}\left\{\frac{\partial \rho_i^\top(\gamma_0)}{\partial \gamma} V_i^{-1} \frac{\partial \rho_i(\gamma_0)}{\partial \gamma^\top}\right\}^{-1} \tag{15}$$

(in that the difference between the left- and right-hand sides is nonnegative definite), and the lower bound is attained with $W_i = V_i^{-1}$ in B and C (Hansen 1982; Abarin & Wang 2006). In practice, however, the use of V_i is infeasible because it depends on the unknown parameters to be estimated. A possible solution is similar to the two-stage procedure used in generalized least squares estimation (e.g., Amemiya 1974; Gallant 1987, ch. 5). First, minimize $Q_N(\gamma)$ using the identity weight $W_i = I$ to obtain the first-stage estimator $\tilde{\gamma}_N$. Second, estimate V_i by

$$\widehat{V}_i = \frac{1}{N} \sum_{i=1}^N \rho_i(\tilde{\gamma}_N)\rho_i^\top(\tilde{\gamma}_N),$$

and then minimize $Q_N(\gamma)$ again with $W_i = \widehat{V}_i^{-1}$ to obtain the second-stage estimator $\hat{\gamma}_N$. Since the asymptotic covariance matrix of $\hat{\gamma}_N$ will be the same as the right-hand side of (15), it is asymptotically more efficient than the first-stage estimator $\tilde{\gamma}_N$. However, because W_i is of dimension $T(T+3)/2$, it is only practical to use the optimal weight when T is not very large. For large T, either the identity matrix or certain block diagonal matrices can be used.

4. SIMULATION-BASED ESTIMATOR

The SLS $\hat{\gamma}_N$ of the previous section can be computed using the usual numerical optimization procedures if closed forms of $\mu_{it}(\gamma)$ and $\nu_{its}(\gamma)$ are available. Sometimes, however, explicit forms of the integrals in (8) and (9) may be difficult or impossible to obtain. In practice, the numerical optimization of an objective function involving multiple integrals can be troublesome, especially when the dimension of the integral is higher than two or three. To overcome this computational difficulty, in this section we consider a simulation-based approach in which the integrals are simulated by Monte Carlo methods such as importance sampling.

The simulation-based estimator can be constructed in the following way. First, choose a known density function $h(u)$ and generate an independent and identically distributed random sample $\{u_{ij},\ j=1,\ldots,2S,\ i=1,\ldots,N\}$ from it. Then approximate $\mu_{it}(\gamma)$ and $\nu_{its}(\gamma)$ respectively using the corresponding Monte Carlo simulators

$$\mu_{it,1}(\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{g(x_{it},u_{ij},\theta)f_\delta(u_{ij}-Z_i\varphi;\psi)}{h(u_{ij})} \tag{16}$$

and

$$\nu_{its,1}(\gamma) = \frac{1}{S}\sum_{j=1}^{S}\frac{g(x_{it},u_{ij},\theta)g(x_{is},u_{ij},\theta)f_\delta(u_{ij}-Z_i\varphi;\psi)}{h(u_{ij})} + \sigma_{its}. \tag{17}$$

Similarly, we construct another set of simulators $\mu_{it,2}(\gamma)$, $\nu_{its,2}(\gamma)$ using the second half of the simulated points $\{u_{ij},\ j=S+1,\ldots,2S,\ i=1,\ldots,N\}$. Finally, the simulation-based estimator (SBE) for γ is defined by

$$\hat{\gamma}_{N,S} = \underset{\gamma\in\Gamma}{\mathrm{argmin}} Q_{N,S}(\gamma) = \underset{\gamma\in\Gamma}{\mathrm{argmin}} \sum_{i=1}^{N}\rho_{i,1}^{\top}(\gamma)W_i\rho_{i,2}(\gamma), \tag{18}$$

where $\rho_{i,j}(\gamma) = \big(y_{it}-\mu_{it,j}(\gamma), 1\le t\le T_i, y_{it}y_{is}-\nu_{its,j}(\gamma), 1\le t\le s\le T_i\big)^{\top}$, $j=1,2$. It is easy to see that $\mu_{it,j}(\gamma)$ and $\nu_{its,j}(\gamma)$ approximate $\mu_{it}(\gamma)$ and $\nu_{its}(\gamma)$ respectively as S is sufficiently large. Moreover, because $\rho_{i,1}(\gamma)$ and $\rho_{i,2}(\gamma)$ are conditionally independent given Y_i, X_i, Z_i, we have $\mathrm{E}[\rho_{i,1}^{\top}(\gamma)W_i\rho_{i,2}(\gamma)] = \mathrm{E}[\rho_i^{\top}(\gamma)W_i\rho_i(\gamma)]$. Thus $Q_{N,S}(\gamma)$ is an unbiased simulator for $Q_N(\gamma)$. For the simulation-based estimator $\hat{\gamma}_{N,S}$, we have the following results.

THEOREM 2. *Suppose that* $\mathrm{Supp}\,(h) \supseteq \mathrm{Supp}\,(f_\delta(\,\cdot\,;\psi))$ *for all* $\psi\in\Psi$. *Then, as* $N\to\infty$, $\hat{\gamma}_{N,S}$ *has the following properties:*

1. *Under* A1–A2, $\hat{\gamma}_{N,S} \xrightarrow{\text{a.s.}} \gamma_0$.
2. *Under* A1–A4, $\sqrt{N}\,(\hat{\gamma}_{N,S}-\gamma_0) \xrightarrow{L} \mathsf{N}(0, B^{-1}C_SB^{-1})$, *where*

$$\begin{aligned} 2C_S &= \mathrm{E}\left\{\frac{\partial\rho_{i,1}^{\top}(\gamma_0)}{\partial\gamma}W_i\rho_{i,2}(\gamma_0)\rho_{i,2}^{\top}(\gamma_0)W_i\frac{\partial\rho_{i,1}(\gamma_0)}{\partial\gamma^{\top}}\right\} \\ &\quad + \mathrm{E}\left\{\frac{\partial\rho_{i,1}^{\top}(\gamma_0)}{\partial\gamma}W_i\rho_{i,2}(\gamma_0)\rho_{i,1}^{\top}(\gamma_0)W_i\frac{\partial\rho_{i,2}(\gamma_0)}{\partial\gamma^{\top}}\right\}. \end{aligned} \tag{19}$$

Furthermore, with probability one,

$$4C_S = \lim_{N\to\infty} \frac{1}{N} \frac{\partial Q_{N,S}(\hat\gamma_{N,S})}{\partial \gamma} \frac{\partial Q_{N,S}(\hat\gamma_{N,S})}{\partial \gamma^\top}. \tag{20}$$

3. *The above results hold if* A3 *is replaced by* A3′.

The SBE $\hat\gamma_{N,S}$ is generally less efficient than the SLS $\hat\gamma_N$, due to the simulation approximation of $\rho_i(\gamma)$ through $\rho_{i,1}(\gamma)$ and $\rho_{i,2}(\gamma)$. A natural question is how much efficiency is lost due to simulation. The following Corollary shows that the efficiency loss caused by simulation decreases at the rate $O(1/S)$. The proof is completely analogous to that of Corollary 4 in Wang (2004) and is therefore omitted.

COROLLARY 1. *Under the conditions of Theorem 2,*

$$\begin{aligned} C_S &= C + \frac{1}{2S}\mathrm{E}\left\{\frac{\partial \rho_i^\top W_i(\rho_{i1}-\rho_i)}{\partial\gamma}\frac{\partial(\rho_{i1}-\rho_i)^\top W_i\rho_i}{\partial\gamma^\top}\right\} \\ &\quad + \frac{1}{4S^2}\mathrm{E}\left\{\frac{\partial(\rho_{i1}-\rho_i)^\top W_i(\rho_{i2}-\rho_i)}{\partial\gamma}\frac{\partial(\rho_{i2}-\rho_i)^\top W_i(\rho_{i1}-\rho_i)}{\partial\gamma^\top}\right\}, \end{aligned}$$

where $\rho_i = \rho_i(\gamma_0)$ *and* ρ_{ij} *is the summand in* $\rho_{i,1}(\gamma_0) = \sum_{j=1}^S \rho_{ij}/S$.

The above result also provides a practical guidance to the choice of the simulation size S. For example, one can control the efficiency loss by choosing a large enough value of S. Asymptotically, the importance density $h(u)$ has no effect on the efficiency of the estimator, as long as it satisfies the condition of Theorem 2. In practice, however, the choice of $h(u)$ will affect the finite sample variances of the Monte Carlo estimators such as $\mu_{it,1}(\gamma)$. Theoretically, the best choice of $h(u)$ is proportional to the absolute value of the integrand, which is $g(x_{it}, u_{ij}, \theta) f_\delta(u_{ij} - Z_i\varphi; \psi)$ for $\mu_{it,1}(\gamma)$. Practically, however, a density close to being proportional to the integrand is a good choice.

5. SIMULATION STUDIES AND APPLICATION

This section is an account of simulation studies we carried out to demonstrate the finite sample performances of the proposed estimators. Specifically, we simulate the exponential model of Example 1 and a linear-exponential model with Berkson measurement errors. In addition, we apply our methods to the well-known orange tree data set. In all simulations, we calculate the first-stage SLS (SLS1) using identity weight and the second-stage SLS (SLS2) using estimated optimal weight. For these and other estimators, we calculate the Monte Carlo means, the simulation standard errors (SSE) and the root mean squared errors (RMSE).

Example 3. First consider the exponential model given in Example 1, which has two correlated random effects. For this model, the closed forms of the first two moments of y_{it} are given in (3) and (4), so that the SLS can be computed by directly minimizing $Q_N(\gamma)$ in (10).

For comparison, we also calculate the quasilikelihood estimators of $\beta = (\varphi_1, \varphi_2, \psi_2, \psi_{12})$ base on the first moment condition (3). In particular, the estimators are calculated by solving the estimating equation

$$\sum_{i=1}^N D_i V_i^{-1}(Y_i - \mu_i) = 0,$$

where $Y_i = (y_{it}, 1 \le t \le T_i)^\top$, $V_i = V(Y_i)$, $\mu_i = (\mu_{it}, 1 \le t \le T_i)^\top$ and D_i is the matrix of partial derivatives of μ_i with respect to all parameters in β. In the quasilikelihood approach, the other "variance parameters" ψ_1, σ_ε^2 have to be estimated through additional estimating equations because they do not appear in μ_i and therefore are not identified by (3). To simplify the computation, we omit the quasilikelihood estimation of ψ_1 and σ^2 and use their true values in the above estimating equation.

The data were generated using $x_{it} = x_t \sim \mathcal{U}(0, 5)$ and $\varepsilon_{it} \sim \mathsf{N}(0, \sigma_\varepsilon^2)$. We have considered two sets of sample sizes (N, T), which are $(20, 5)$ and $(40, 7)$. In each case, 1000 Monte Carlo replications were carried out. The computation was done using the statistical package R on a workstation running Windows XP. The results are reported in Tables 1 and 2. These results show that both SLS estimators using the identity and optimal weights do not have apparent biases and they have similar SSE and RMSE. The reason that the SLS2 does not improve SLS1 significantly may be due to the fact that the weighting matrix is not easily estimated accurately with relatively small sample sizes. Further, it is clear that the quasilikelihood estimator (QLE) has finite sample biases for most parameters and has smaller SSE but larger RMSE than both SLS estimators.

We have also tried other distributions for random effects, such as χ^2 distributions with low degrees of freedoms. The results obtained follow patterns similar to those in Tables 1 and 2.

TABLE 1: Simulation results of Example 3 with sample sizes $N = 20, T = 5$.

True	$\varphi_1 = 10$	$\varphi_2 = 5$	$\psi_1 = 1$	$\psi_2 = 0.7$	$\psi_{12} = 0.5$	$\sigma^2 = 1$
SLS1	9.9024	4.9369	1.0032	0.6803	0.5003	0.9827
SSE	0.0499	0.0229	0.0092	0.0055	0.0055	0.0051
RMSE	1.5816	0.7264	0.2915	0.1749	0.1733	0.1612
SLS2	9.8597	4.9365	0.9940	0.6913	0.5012	0.9395
SSE	0.0442	0.0214	0.0092	0.0056	0.0055	0.0051
RMSE	1.4030	0.6785	0.2919	0.1768	0.1734	0.1722
QLE	11.2574	5.4979	-	0.6056	0.4935	-
SSE	0.0333	0.0186	-	0.0051	0.0055	-
RMSE	1.6392	0.7707	-	0.1868	0.1743	-

Example 4. Now consider a measurement error model $y_i = \theta_1 \xi_{1i} + \theta_2 \exp(\theta_3 \xi_{2i}) + \varepsilon_i$ and $\xi_i = Z_i + \delta_i$, where $\varepsilon_i \sim \mathsf{N}(0, \sigma_\varepsilon^2)$ and $\delta_i \sim \mathsf{N}[(0, 0)^\top, \text{diag}\,(\psi_1, \psi_2)]$. Here we have omitted the index t everywhere, since $T_i = 1$. For this model, the conditional moments (8) and (9) have closed forms

$$\mu_i(\gamma) = \theta_1 Z_{1i} + \theta_2 \exp(\theta_3 Z_{2i} + \theta_3^2 \psi_2 / 2)$$

and

$$\nu_i(\gamma) = \theta_1^2 (Z_{1i}^2 + \psi_1) + \theta_2^2 \exp(2\theta_3 Z_{2i} + 2\theta_3^2 \psi_2) + 2\theta_1 \theta_2 Z_{1i} \exp(\theta_3 Z_{2i} + \theta_3^2 \psi_2 / 2) + \sigma_\varepsilon^2,$$

so that the SLS can be computed by minimizing $Q_N(\gamma)$ in (10). For the purpose of demonstration, we also compute the simulation-based estimator. To compute the SBE, we choose the density of $N[(0, 0)^\top, \text{diag}\,(5, 5)]$ to be $h(u)$, and generate independent points $u_{ij} \sim h(u)$, $j = 1, \ldots, 2S$, $i = 1, \ldots, N$ using $S = 1000$. Further, the simulated moments $\mu_{i,j}(\gamma), \nu_{i,j}(\gamma)$, $j = 1, 2$ are calculated according to (16) and (17). Finally, the SBE $\hat{\gamma}_{N,S}$ is calculated by minimizing $Q_{N,S}(\gamma)$ in (18) using the identity weight. The SBE using the estimated optimal weight has also been calculated but the numerical results are very similar and are not reported here. The data were generated using $Z_i \sim \mathsf{N}[(1, 2)^\top, \text{diag}\,(1, 2)]$ and sample size $N = 50$. In this simulation, 500 Monte Carlo replications were carried out. For comparison, we also included the ordinary nonlinear least squares estimates ignoring the measurement errors. The computation was done using the package MATLAB on a workstation running Windows XP.

TABLE 2: Simulation results of Example 3 with sample sizes $N = 40, T = 7$.

True	$\varphi_1 = 10$	$\varphi_2 = 5$	$\psi_1 = 1$	$\psi_2 = 0.7$	$\psi_{12} = 0.5$	$\sigma^2 = 1$
SLS1	9.9178	4.8742	0.9959	0.6454	0.5104	0.9915
SSE	0.0475	0.0310	0.0089	0.0048	0.0055	0.0034
RMSE	1.5029	0.9888	0.2804	0.1614	0.1732	0.1073
SLS2	9.9049	4.8969	0.9971	0.6572	0.5055	0.9332
SSE	0.0391	0.0264	0.0091	0.0052	0.0054	0.0034
RMSE	1.2404	0.8406	0.2870	0.1691	0.1709	0.1269
QLE	11.4357	5.8306	-	0.6335	0.4920	-
SSE	0.0184	0.0129	-	0.0052	0.0055	-
RMSE	1.5491	0.9246	-	0.1759	0.1739	-

The results are reported in Table 3. These results show that with a moderate sample size $N = 50$, both SLS and SBE perform reasonably well, though the improvement of SLS2 over SLS1 seems not to be significant. Moreover, with a simulation size $S = 1000$, the SBE performs as well as the SLS, except for slightly higher standard deviations. As expected, the nonlinear least squares estimates (NLSE) are clearly biased for most parameters.

TABLE 3: Simulation results of Example 4 with sample size $N = 50, T = 1$. The simulation standard errors are in parentheses.

True	$\theta_1 = 3$	$\theta_2 = 2$	$\theta_3 = -1$	$\sigma_\varepsilon^2 = 1$	$\psi_1 = 1$	$\psi_2 = 1.5$
SLS1	2.9974	1.9291	-0.8699	1.0032	1.0133	1.3929
	(0.0065)	(0.0065)	(0.0045)	(0.0052)	(0.0057)	(0.0047)
SLS2	2.9990	1.9470	-0.8616	1.0057	0.9851	1.3554
	(0.0064)	(0.0068)	(0.0042)	(0.0053)	(0.0058)	(0.0034)
SBE	2.9845	1.9114	-0.8668	1.0000	1.0031	1.3853
	(0.0065)	(0.0062)	(0.0045)	(0.0053)	(0.0059)	(0.0046)
NLSE	3.0926	2.1114	-0.9466	37.7340	-	-
	(0.0062)	(0.0064)	(0.0061)	(4.8863)	-	-

Example 5. Finally we apply our methods to the orange tree data. The data are given in Draper & Smith (1998, p. 559) and contain the measurements on the trunk circumferences (y_{it}, in millimeters) of five orange trees taken at seven occasions (x_{it}, in days from December 31, 1968). The logistic growth model (5) in Example 2 has been used by many authors (e.g., Lindstrom & Bates 1990) to model this data set. Later, Pinheiro & Bates (1995) recalculated the estimates for a reparameterized form of the model

$$y_{it} = \frac{\varphi + \delta_i}{1 + \exp\{-(x_{it} - \theta_1)/\theta_2\}} + \varepsilon_{it},$$

where $\delta_i \sim (0, \psi)$ and $\varepsilon_{it} \sim (0, \sigma^2)$. For this model the first two conditional moments of y_{it} given X_i are similar to those in (6) and (7). Here we have calculated the SLS estimates using the identity weight, which are shown in Table 4. As a "gold standard", we have included the maximum likelihood estimates (MLE) and the linear mixed effects model approximation of the restricted MLE presented in Pinheiro & Bates (1995). We can see that the SLS are in line with the estimates obtained by the other two methods, and closer to the MLE than the LME for some instances. Among the three methods, the estimates for the random effect variance ψ look quite different, which is not surprising because its estimator is known to have a fairly large standard deviation.

TABLE 4: Estimation of the orange tree growth model in Example 5.

	θ_1	θ_2	φ	ψ	σ_ε^2
SLS	729.92	350.13	192.50	1002.41	61.00
MLE	727.91	348.07	192.05	1001.25	61.50
LME	722.56	344.16	191.05	990.29	61.56

6. REGULARITY CONDITIONS

This section contains regularity conditions that are required to derive the asymptotic properties of the SLS $\hat{\gamma}_N$ and the SBE $\hat{\gamma}_{N,S}$. In particular, for the consistency of the estimators, we assume the following conditions, where $\|\cdot\|$ denotes the Euclidean norm.

A1. *For each* $(\xi_i^\top, \theta^\top)^\top \in R^\ell \times \Theta$, $g(x_{it}, \xi_i, \theta)$ *is a measurable function of* x_{it} *and is continuous in* $(\xi_i^\top, \theta^\top)^\top \in R^\ell \times \Theta$ *for all* x_{it}. $f_\delta(u;\psi)$ *is continuous in* $\psi \in \Psi$ *for all* u. *Furthermore,* $\mathrm{E}\,\|W_i\|(y_{it}^4+1) < \infty$ *and* $\mathrm{E}\,\|W_i\|\left(\int \sup_\Gamma g^2(x_{it}, u, \theta) f_\delta(u - Z_i\varphi;\psi)\,du\right)^2 < \infty$.

A2. $\mathrm{E}\,\{\rho_i(\gamma) - \rho_i(\gamma_0)\}^\top W_i \{\rho_i(\gamma) - \rho_i(\gamma_0)\} = 0$ *if and only if* $\gamma = \gamma_0$.

The above conditions are common in the nonlinear regression literature. In particular, A1 is usually used to ensure the continuity and uniform convergence of $Q_N(\gamma)$. It is easy to see that the moment conditions in A1 are satisfied, e.g., if $g(x_{it}, \xi_i, \theta)$ is a polynomial and δ_i has a normal distribution. Moreover, A2 is the usual condition for identifiability of parameters, which implies that $Q_N(\gamma)$ has unique minimizer γ_0 for large N. For the asymptotic normality of our estimators, we assume further conditions as follows.

A3. *There exist open subsets* $\theta_0 \in \Theta_0 \subset \Theta$ *and* $\psi_0 \in \Psi_0 \subset \Psi$, *in which* $g(x_{it}, \xi_i, \theta)$ *is twice continuously differentiable with respect to* θ *and* $f_\delta(u;\psi)$ *is twice continuously differentiable with respect to both* u *and* ψ. *Furthermore, there exists positive function* $G(u, x, z)$ *satisfying*

$$\mathrm{E}\,\|W_i\| \left(\int G(u, X_i, Z_i)\,du \right)^2 < \infty,$$

such that all partial derivatives of order 0 to 2 of $g(x_{it}, u, \theta) f_\delta(u - Z_i\varphi;\psi)$ *and* $g(x_{it}, u, \theta) g(x_{is}, u, \theta) f_\delta(u - Z_i\varphi;\psi)$ *with respect to* (θ, φ, ψ) *are absolutely bounded by* $G(u, X_i, Z_i)$.

A4. *The matrix*

$$B = \mathrm{E}\left\{ \frac{\partial \rho_i^\top(\gamma_0)}{\partial \gamma} W_i \frac{\partial \rho_i(\gamma_0)}{\partial \gamma^\top} \right\}$$

is nonsingular.

Again, A3 and A4 are regularity conditions commonly seen, which are sufficient for the asymptotic normality of the second-order least squares estimators. In particular, while A3 ensures that the first derivative of $Q_N(\gamma)$ admits a first-order Taylor expansion and the second derivative of $Q_N(\gamma)$ converges uniformly, A4 implies that the second derivative of $Q_N(\gamma)$ has a nonsingular limiting matrix. Again, it is easy to see that A3 and A4 are satisfied for the polynomial model $g(x_{it}, \xi_i, \theta)$ and the normal random effects δ_i.

Moreover, assumption A3 and the dominated convergence theorem imply that the first derivatives

$$\frac{\partial \rho_i^\top(\gamma)}{\partial \gamma} = -\left(\frac{\partial \mu_{it}(\gamma)}{\partial \gamma}, 1 \le t \le T_i, \frac{\partial \nu_{it}(\gamma)}{\partial \gamma}, 1 \le t \le s \le T_i\right)$$

exist and they are given by

$$\begin{aligned}
\frac{\partial \mu_{it}(\gamma)}{\partial \theta} &= \int \frac{\partial g(x_{it}, u, \theta)}{\partial \theta} f_\delta(u - Z_i\varphi; \psi)\, du, \\
\frac{\partial \mu_{it}(\gamma)}{\partial \varphi} &= -Z_i^\top \int g(x_{it}, u, \theta) \frac{\partial f_\delta(u - Z_i\varphi; \psi)}{\partial u}\, du, \\
\frac{\partial \mu_{it}(\gamma)}{\partial \psi} &= \int g(x_{it}, u, \theta) \frac{\partial f_\delta(u - Z_i\varphi; \psi)}{\partial \psi}\, du, \\
\frac{\partial \mu_{it}(\gamma)}{\partial \sigma^2} &= 0
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \nu_{its}(\gamma)}{\partial \theta} &= \int \frac{\partial g(x_{it}, u, \theta) g(x_{is}, u, \theta)}{\partial \theta} f_\delta(u - Z_i\varphi; \psi)\, du, \\
\frac{\partial \nu_{its}(\gamma)}{\partial \varphi} &= -Z_i^\top \int g(x_{it}, u, \theta) g(x_{is}, u, \theta) \frac{\partial f_\delta(u - Z_i\varphi; \psi)}{\partial u}\, du, \\
\frac{\partial \nu_{its}(\gamma)}{\partial \psi} &= \int g(x_{it}, u, \theta) g(x_{is}, u, \theta) \frac{\partial f_\delta(u - Z_i\varphi; \psi)}{\partial \psi}\, du, \\
\frac{\partial \nu_{itt}(\gamma)}{\partial \sigma^2} &= 1, \quad \frac{\partial \nu_{its}(\gamma)}{\partial \sigma^2} = 0,\ t \ne s.
\end{aligned}$$

Note that A3 entails the differentiability of $f_\delta(u; \psi)$ with respect to u. This condition can be replaced by the differentiability of $g(x_{it}, \xi_i, \theta)$ with respect to ξ_i, because through variable substitution, integrals (8) and (9) can be written as

$$\mu_{it}(\gamma) = \int g(x_{it}, Z_i\varphi + u, \theta) f_\delta(u; \psi)\, du, \tag{21}$$

$$\nu_{its}(\gamma) = \int g(x_{it}, Z_i\varphi + u, \theta) g(x_{is}, Z_i\varphi + u, \theta) f_\delta(u; \psi)\, du + \sigma_{its}. \tag{22}$$

Hence A3 can be modified as follows.

A3′. *There exist open subsets $\theta_0 \in \Theta_0 \subset \Theta$ and $\psi_0 \in \Psi_0 \subset \Psi$, in which $g(x_{it}, \xi_i, \theta)$ is twice continuously differentiable with respect to $(\xi_i^\top, \theta^\top)^\top$ and $f_\delta(u; \psi)$ is twice continuously differentiable with respect to ψ. Furthermore, there exists function $G(u, x, z)$ satisfying*

$$\mathrm{E}\, \|W_i\| \left(\int G(u, X_i, Z_i)\, du\right)^2 < \infty,$$

such that all partial derivatives of order 0 to 2 of $g(x_{it}, Z_i\varphi + u, \theta) f_\delta(u; \psi)$ and $g(x_{it}, Z_i\varphi + u, \theta) g(x_{is}, Z_i\varphi + u, \theta) f_\delta(u; \psi)$ with respect to (θ, φ, ψ) are absolutely bounded by $G(u, X_i, Z_i)$.

Under this assumption, the first derivatives of $\mu_{it}(\gamma)$ and $\nu_{it}(\gamma)$ with respect to φ become

$$\frac{\partial \mu_{it}(\gamma)}{\partial \varphi} = Z_i^\top \int \frac{\partial g(x_{it}, Z_i\varphi + u, \theta)}{\partial \xi} f_\delta(u; \psi)\, du,$$

and

$$\frac{\partial \nu_{its}(\gamma)}{\partial \varphi} = Z_i^\top \int \frac{\partial g(x_{it}, Z_i\varphi + u, \theta) g(x_{is}, Z_i\varphi + u, \theta)}{\partial \xi} f_\delta(u;\psi)\, du,$$

respectively, and all other derivatives remain the same as under A3.

Finally note that, since the simulated objective function $Q_{N,S}(\gamma)$ does not involve integrals any more, it is continuous in, and differentiable with respect to, γ, as long as functions $g(x_{it}, \xi_i, \theta)$ and $f_\delta(u;\psi)$ have these properties. In this sense, the simulation-based estimator requires weaker assumptions than the second-order least squares estimator.

7. CONCLUSIONS AND DISCUSSION

We have used a unified framework for estimation of the mixed effects and the Berkson measurement error models, which are presented in different contexts and have different interpretations in the literature. For the mixed effects models, this approach does not require the distribution of the random effects to be normal, nor does it need any parametric assumption for the distribution of the random errors in the regression equation. In the context of measurement error models, this approach produces *exactly* (rather than *approximately*) consistent estimators. The possible computational issue of minimizing a function that involves multiple integrals is addressed using the method of simulated moments, so that the proposed estimators are numerically always feasible. Limited Monte Carlo simulation studies show that the proposed estimators perform fairly satisfactorily for relatively small sample sizes and slightly better than the quasi-likelihood estimators, even though the latter uses more information than the former.

It is possible to extend the approach of this paper to more general situations. One possible extension is that ε_{it} and ε_{is} are correlated, so that $\mathrm{E}\,(\varepsilon_{it}\varepsilon_{is} \mid X_i, Z_i, \delta_i) = \sigma_{ts} \neq 0$. It is easy to see that the asymptotic results of the SLS and the SBE remain valid with minor modification of the asymptotic covariance matrix. The proofs can be modified easily by repeatedly using the Cauchy–Schwartz inequality. Another possible extension of the approach is to cover the situation where the individuals have unbalanced observations. In this case, T_i depends on i and may be different for $i = 1, \ldots, N$. Because now $\rho_i(\gamma)$, $i = 1, \ldots, N$ have different dimensions, the proofs of asymptotic normality of the estimators should be based on the central limit theorem of Lindeberg–Feller, instead of Lindeberg–Lévy. Future research should include investigation of the finite sample behavior of the proposed estimators through more extensive Monte Carlo simulation studies and comparisons with other existing methods in the literature.

APPENDIX

Proof of Theorem 1.1. For any $1 \le i \le N$, by definition and the Cauchy–Schwartz inequality,

$$\begin{aligned}\|\rho_i(\gamma)\|^2 \quad &\le \quad 2\sum_t y_{it}^2 + 2\sum_{t\le s} y_{it}^2 y_{is}^2 + 2\sum_t \int g^2(x_{it}, u, \theta) f_\delta(u - Z_i\varphi;\psi)\, du \\ &\quad + 4\sum_{t\le s} \int g^2(x_{it}, u, \theta) f_\delta(u - Z_i\varphi;\psi)\, du \int g^2(x_{is}, u, \theta) f_\delta(u - Z_i\varphi;\psi)\, du \\ &\quad + 4T\sigma^4.\end{aligned}$$

It follows from assumption A1 that

$$\begin{aligned}\mathrm{E}\sup_\Gamma \rho_i^\top(\gamma) W_i \rho_i(\gamma) \quad &\le \quad \mathrm{E}\,\|W_i\| \sup_\Gamma \|\rho_i(\gamma)\|^2 \\ &\le \quad 2\sum_t \mathrm{E}\,\|W_i\| y_{it}^2 + 2\sum_{t\le s} \mathrm{E}\|W_i\| y_{it}^2 y_{is}^2 + 4T\sup_\Sigma \sigma^4 \mathrm{E}\|W_i\| \\ &\quad + 2\sum_t \mathrm{E}\,\|W_i\| \int \sup_\Gamma g^2(x_{it}, u, \theta) f_\delta(u - Z_i\varphi;\psi)\, du\end{aligned}$$

$$+ 4\sum_{t\le s}\left\{\mathrm{E}\,\|W_i\|\left(\int \sup_{\Gamma} g^2(x_{it},u,\theta) f_\delta(u - Z_i\varphi;\psi)\,du\right)^2\right\}^{1/2}$$
$$\times \left\{\mathrm{E}\,\|W_i\|\left(\int \sup_{\Gamma} g^2(x_{is},u,\theta) f_\delta(u - Z_i\varphi;\psi)\,du\right)^2\right\}^{1/2}$$
$$< \quad \infty.$$

Again by the Cauchy–Schwartz inequality,

$$\int \sup_{\Gamma} |g(x_{it},u,\theta)|^j f_\delta(u - Z_i\varphi;\psi)\,du < \infty, \quad j = 1, 2.$$

It follows from A1 and the dominated convergence theorem that $\rho_i(\gamma)$, and therefore $Q_N(\gamma)$, is continuous in $\gamma \in \Gamma$. Furthermore, by the uniform law of large numbers (Jennrich 1969, Th. 2), uniformly in $\gamma \in \Gamma$, $Q_N(\gamma)/N$ converges almost surely to

$$Q(\gamma) = \mathrm{E}\,\rho_i^\top(\gamma) W_i \rho_i(\gamma) = Q(\gamma_0) + \mathrm{E}\,\{\rho_i(\gamma) - \rho_i(\gamma_0)\}^\top W_i \{\rho_i(\gamma) - \rho_i(\gamma_0)\}. \tag{23}$$

It follows that $Q(\gamma) \ge Q(\gamma_0)$ and, by A2, equality holds if and only if $\gamma = \gamma_0$. Therefore by Lemma 3 of Amemiya (1973), we have $\hat\gamma_N \xrightarrow{\text{a.s.}} \gamma_0$.

Proof of Theorem 1.2. By Assumption A3 the first derivative $\partial Q_N(\gamma)/\partial\gamma$ exists and has a first-order Taylor expansion in an open neighbourhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_N(\hat\gamma_N)/\partial\gamma = 0$ and $\hat\gamma_N \xrightarrow{\text{a.s.}} \gamma_0$, for sufficiently large N we have

$$\frac{\partial Q_N(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_N(\tilde\gamma_N)}{\partial\gamma\partial\gamma^\top}(\hat\gamma_N - \gamma_0) = 0, \tag{24}$$

where $\|\tilde\gamma_N - \gamma_0\| \le \|\hat\gamma_N - \gamma_0\|$. The first derivative of $Q_N(\gamma)$ in (24) is given by

$$\frac{\partial Q_N(\gamma)}{\partial\gamma} = 2\sum_{i=1}^N \frac{\partial\rho_i^\top(\gamma)}{\partial\gamma} W_i \rho_i(\gamma),$$

where $\partial\rho_i^\top(\gamma)/\partial\gamma$ is given in Section 6 after A4. Since $\partial\rho_i^\top(\gamma)W_i\rho_i(\gamma)/\partial\gamma$ are independent and identically distributed by the central limit theorem we have, as $N \to \infty$,

$$\frac{1}{\sqrt{N}}\frac{\partial Q_N(\gamma_0)}{\partial\gamma} \xrightarrow{L} \mathsf{N}(0, 4C), \tag{25}$$

where C is given in (11). The second derivative of $Q_N(\gamma)$ in (24) is given by

$$\frac{\partial^2 Q_N(\gamma)}{\partial\gamma\partial\gamma^\top} = 2\sum_{i=1}^N\left\{\frac{\partial\rho_i^\top(\gamma)}{\partial\gamma} W_i \frac{\partial\rho_i(\gamma)}{\partial\gamma^\top} + (\rho_i^\top(\gamma) W_i \otimes I)\frac{\partial\,\mathrm{vec}\,(\partial\rho_i^\top(\gamma)/\partial\gamma)}{\partial\gamma^\top}\right\},$$

where $\otimes$ is the Kronecker product, I is the $2N(p+q+r+1)$ dimensional identity matrix, and

$$\frac{\partial\,\mathrm{vec}\,(\partial\rho_i^\top(\gamma)/\partial\gamma)}{\partial\gamma^\top} = -\left(\frac{\partial^2\mu_{it}(\gamma)}{\partial\gamma\partial\gamma^\top}, 1 \le t \le T, \frac{\partial^2\nu_{its}(\gamma)}{\partial\gamma\partial\gamma^\top}, 1 \le t \le s \le T\right)^\top.$$

Analogous to the proof of Theorem 1.1, by repeatedly using the Cauchy–Schwartz inequality and A3, we can verify that

$$\mathrm{E}\,\sup_{\Gamma}\left\|\frac{\partial\rho_i^\top(\gamma)}{\partial\gamma} W_i \frac{\partial\rho_i(\gamma)}{\partial\gamma^\top}\right\| \le \sum_{t,s}\mathrm{E}\,\|W_1\|\sup_{\Gamma}\left(\left\|\frac{\partial\mu_{it}(\gamma)}{\partial\gamma}\right\|^2 + \left\|\frac{\partial\nu_{its}(\gamma)}{\partial\gamma}\right\|^2\right) < \infty$$

and

$$\begin{aligned}
&\mathrm{E}\sup_{\Gamma}\left\|(\rho_i^\top(\gamma)W_i\otimes I)\frac{\partial\,\mathrm{vec}\,(\partial\rho_i^\top(\gamma)/\partial\gamma)}{\partial\gamma^\top}\right\| \\
&\le \sqrt{2\ell(p+q+r+1)}\left(\mathrm{E}\,\|W_i\|\sup_{\Gamma}\|\rho_i(\gamma)\|^2\mathrm{E}\|W_i\|\sup_{\Gamma}\left\|\frac{\partial\,\mathrm{vec}\,(\partial\rho_i^\top(\gamma)/\partial\gamma)}{\partial\gamma^\top}\right\|^2\right)^{1/2} \\
&< \infty.
\end{aligned}$$

That $(1/N)\partial^2 Q_N(\gamma)/\partial\gamma\partial\gamma^\top \xrightarrow{\text{a.s.}} \partial^2 Q(\gamma)/\partial\gamma\partial\gamma^\top$ uniformly in $\gamma\in\Gamma_0$ follows from the uniform law of large numbers. Therefore, by Lemma 4 of Amemiya (1973), we have

$$\frac{1}{N}\frac{\partial^2 Q_N(\tilde\gamma_N)}{\partial\gamma\partial\gamma^\top}\xrightarrow{\text{a.s.}}2B, \tag{26}$$

where B is given in (12). It follows then from (24)–(26), A4 and Slutsky's theorem (Amemiya 1985) that $\sqrt{n}\,(\hat\gamma_N-\gamma_0)\xrightarrow{L}\mathsf{N}(0,B^{-1}CB^{-1})$. Moreover, (13) and (14) can be similarly verified by Lemma 4 of Amemiya (1973).

Proof of Theorem 1.3. Under the alternative assumption A3′, the above derivation remains valid with minor modification. In fact, through variable substitution integrals in (8) and (9) can be written as in (21) and (22), respectively. Therefore the only change is that the derivatives of $\mu_{it}(\gamma)$ and $\nu_{its}(\gamma)$ with respect to φ now are calculated through the derivatives of $g(x_{it},\xi_i,\theta)$ with respect to ξ_i, instead of the derivatives of $f_\delta(u;\psi)$ with respect to u.

Proof of Theorem 2. The proof of Theorem 2.1 is analogous to that of Theorem 1.1. First, A3 implies that $Q_{N,S}(\gamma)$ is continuous in $\gamma\in\Gamma$. Then, by the uniform law of large numbers, we have, as $N\to\infty$, uniformly in $\gamma\in\Gamma$ that

$$\frac{1}{N}Q_{N,S}(\gamma)\xrightarrow{\text{a.s.}}\mathrm{E}\left\{\rho_{i,1}^\top(\gamma)W_i\rho_{i,2}(\gamma)\right\}=Q(\gamma).$$

Finally, $\hat\gamma_{N,S}\xrightarrow{\text{a.s.}}\gamma_0$ follows from (23), A2 and Lemma 3 of Amemiya (1973).

The proof of Theorem 2.2 is analog to that of Theorem 1.2. First, by A3 we have the first-order Taylor expansion of $\partial Q_{N,S}(\gamma)/\partial\gamma$ in a neighbourhood $\Gamma_0\subset\Gamma$ of γ_0

$$0=\frac{\partial Q_{N,S}(\gamma_0)}{\partial\gamma}+\frac{\partial^2 Q_{N,S}(\tilde\gamma_{N,S})}{\partial\gamma\partial\gamma^\top}(\hat\gamma_{N,S}-\gamma_0), \tag{27}$$

where $\|\tilde\gamma_{N,S}-\gamma_0\|\le\|\hat\gamma_{N,S}-\gamma_0\|$ and the first derivative of $Q_{N,S}(\gamma)$ is given by

$$\frac{\partial Q_{N,S}(\gamma)}{\partial\gamma}=\sum_{i=1}^N\left\{\frac{\partial\rho_{i,1}^\top(\gamma)}{\partial\gamma}W_i\rho_{i,2}(\gamma)+\frac{\partial\rho_{i,2}^\top(\gamma)}{\partial\gamma}W_i\rho_{i,1}(\gamma)\right\}.$$

Since $\rho_{i,1}(\gamma)$ has the same distribution as $\rho_{i,2}(\gamma)$, all terms in the above summation are independent and identically distributed and have the common covariance matrix $4C_S$ which is given in (19). It follows by the central limit theorem that, as $N\to\infty$,

$$\frac{1}{\sqrt{N}}\frac{\partial Q_{N,S}(\gamma_0)}{\partial\gamma}\xrightarrow{L}\mathsf{N}(0,4C_S). \tag{28}$$

Now, the second derivative in (27) is given by

$$\frac{\partial^2 Q_{N,S}(\gamma)}{\partial\gamma\partial\gamma^\top} = \sum_{i=1}^N\left\{\frac{\partial\rho_{i,1}^\top(\gamma)}{\partial\gamma}W_i\frac{\partial\rho_{i,2}(\gamma)}{\partial\gamma^\top}\right.$$

$$+ (\rho_{i,2}^{\top}(\gamma) W_i \otimes I_{p+q+r+1}) \frac{\partial \operatorname{vec}(\partial \rho_{i,1}^{\top}(\gamma)/\partial \gamma)}{\partial \gamma^{\top}} \Bigg\}$$

$$+ \sum_{i=1}^{N} \Bigg\{ \frac{\partial \rho_{i,2}^{\top}(\gamma)}{\partial \gamma} W_i \frac{\partial \rho_{i,1}(\gamma)}{\partial \gamma^{\top}}$$

$$+ (\rho_{i,1}^{\top}(\gamma) W_i \otimes I_{p+q+r+1}) \frac{\partial \operatorname{vec}(\partial \rho_{i,2}^{\top}(\gamma)/\partial \gamma)}{\partial \gamma^{\top}} \Bigg\}.$$

Again, by A3 and Lemma 4 of Amemiya (1973), uniformly in $\gamma \in \Gamma$,

$$\frac{1}{N} \frac{\partial^2 Q_{N,S}(\tilde{\gamma}_N)}{\partial \gamma \partial \gamma^{\top}} \xrightarrow{\text{a.s.}} \mathrm{E} \left\{ \frac{\partial \rho_{i,1}^{\top}(\gamma_0)}{\partial \gamma} W_i \frac{\partial \rho_{i,2}(\gamma_0)}{\partial \gamma^{\top}} + \frac{\partial \rho_{i,2}^{\top}(\gamma_0)}{\partial \gamma} W_i \frac{\partial \rho_{i,1}(\gamma_0)}{\partial \gamma^{\top}} \right\} = 2B. \qquad (29)$$

Therefore by (27)–(29) and Slutsky's theorem, we have $\sqrt{N}\,(\hat{\gamma}_{N,S} - \gamma_0) \xrightarrow{L} \mathsf{N}(0, B^{-1} C_S B^{-1})$. Moreover, (20) can be similarly shown by Lemma 4 of Amemiya (1973). Finally, the same argument as in the proof of Theorem 1.3 applies here, too.

ACKNOWLEDGEMENTS

The author is grateful to two anonymous referees, an Associate Editor and the Editor for their constructive comments and suggestions. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

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

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

T. Amemiya (1974). The nonlinear two-stage least-squares estimator. *Journal of Econometrics*, 2, 105–110.

T. Amemiya (1985). *Advanced Econometrics*. Basil Blackwell, Oxford.

M. Arellano & B. Honoré (2001). Panel data models: some recent developments. In *Handbook of Econometrics, Volume 5* (J. J. Heckman & E. Leamer, eds.), North-Holland, Amsterdam, pp. 3229–3296.

J. Berkson (1950). Are there two regressions? *Journal of American Statistical Association*, 45, 164–180.

R. J. Carroll, D. Ruppert & L. A. Stefanski (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, London, USA.

G. Chamberlain (1984). Panel data. In *Handbook of Econometrics, Volume 2* (Z. Griliches & M. D. Intriligator, eds.), Elsevier Science, Amsterdam, pp. 1247–1318.

T. Daimon & M. Goto (2003). Power-transformation mixed effects model with applications to pharmacokinetics. *Journal of the Japanese Society of Computational Statistics*, 15, 135–150.

M. Davidian & A. R. Gallant (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80, 475–488.

M. Davidian & D. M. Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall/CRC, Boca Raton, Florida.

N. R. Draper & H. Smith (1998). *Applied Regression Analysis*, Third edition, Wiley, New York.

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

A. R. Gallant (1987). *Nonlinear Statistical Models*, Wiley, New York.

P. Gustafson (2004). *Measurement Error and Misclassification in Statistics and Epidemiology - Impacts and Bayesian Adjustments*. Chapman & Hall, Boca Raton, Florida.

L. P. Hansen (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

A. Hartford & M. Davidian (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 34, 139–164.

C. Hsiao (2003). *Analysis of Panel Data*, Second edition, Econometric Society Monographs 34. Cambridge University Press, Cambridge.

C. Hsiao, M. H. Pesaran & A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, 109, 107–150.

R. I. Jennrich (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40, 633–643.

C. Ke & Y. Wang (2001). Semiparametric nonlinear mixed effects models and their applications. *Journal of the American Statistical Association*, 96, 1272–1281.

T. L. Lai & M.-C. Shih (2003a). A hybrid estimator in nonlinear and generalized linear mixed effects models. *Biometrika*, 90, 859–879.

T. L. Lai & M.-C. Shih (2003b). Nonparametric estimation of nonlinear mixed effects models. *Biometrika*, 90, 1–13.

J. K. Lindsey (1999). *Models for Repeated Measurements*, Second edition, Oxford University Press.

M. J. Lindstrom & D. M. Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673–687.

J. C. Pinheiro & D. M. Bates (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12–35.

M. Rudemo, D. Ruppert & J. C. Streibig (1989). Random-effect models in nonlinear regression with applications to bioassay. *Biometrics*, 45, 349–362.

E. F. Vonesh & V. M. Chinchilli (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.

E. F. Vonesh, H. Wang, L. Nie & D. Majumdar (2002). Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *Journal of the American Statistical Association*, 97, 271–283.

L. Wang (2003). Estimation of nonlinear Berkson-type measurement error models. *Statistica Sinica*, 13, 1201–1210.

L. Wang (2004). Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics*, 32, 2559–2579.

J. M. Wooldridge (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90, 77–97.

L. Wu (2002). A joint model for nonlinear mixed effects models with censoring and covariates measured with error, with applications to AIDS studies. *Journal of the American Statistical Association*, 97, 955–964.

Received 20 March 2006
Accepted 7 March 2007

Liqun WANG: liqun_wang@umanitoba.ca
Department of Statistics
The University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2