Ann. Inst. Statist. Math.
Vol. 54, No. 3, 595–606 (2002)

ESTIMATION IN LINEAR MODELS WITH RANDOM EFFECTS AND ERRORS-IN-VARIABLES*

XU-PING ZHONG[1], WING-KAM FUNG[2] AND BO-CHENG WEI[1]

[1]*Department of Mathematics, Southeast University, Nanjing 210018, China*, e-mail: bcw@seu.edu.cn
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China*, e-mail: wingfung@hku.hk

(Received July 18, 2000; revised March 29, 2001)

Abstract. The independent variables of linear mixed models are subject to measurement errors in practice. In this paper, we present a unified method for the estimation in linear mixed models with errors-in-variables, based upon the corrected score function of Nakamura (1990, *Biometrika*, **77**, 127–137). Asymptotic normality properties of the estimators are obtained. The estimators are shown to be consistent and convergent at the order of $n^{-1/2}$. The performance of the proposed method is studied via simulation and the analysis of a data set on hedonic housing prices.

Key words and phrases: Corrected score function, errors-in-variable, fixed effects, random effects, measurement errors.

1. Introduction

There has been a great deal of interest in mixed effects models recently. These models are commonly used for analyzing longitudinal data and repeated measurements in biomedical, social and economical studies (see, for example, Diggle *et al.* (1994); Davidian and Giltinan (1995)). Statistical inference based on the common likelihood analysis is much involved because of the intractable numerical integration. Various methods have been proposed to tackle this problem; see for example, Breslow and Clayton (1993), Liu and Pierce (1993), Lin and Breslow (1996).

Independent variables, or covariates in the models are often measured with errors (Fuller (1987); Carroll *et al.* (1995)). Ordinary maximum likelihood estimators, without taking into account the measurement errors, are generally inconsistent (e.g. Armstrong (1985); Fuller (1987); Stefanski and Carroll (1987)). There are two useful approaches shown in the literature for dealing with measurement error models as pointed out by Hanfelt and Liang (1997). The corrected score approach of Nakamura (1990, 1992) successfully corrects for measurement errors in normal, Poisson, gamma and inverse Gaussian regression models, and in proportional hazards models (Hanfelt and Liang (1997)). Buzas and Stefanski (1996) expanded the potential for application of the methodology, and described an application to extreme-value binary regression. The other approach given by Stefanski and Carroll (1987) dealt with an unbiased score function based upon the estimating equation method. A full treatment of measurement error models can be found in Fuller (1987) and Carroll *et al.* (1995).

*The project is supported by a grant of The University of Hong Kong and the grant NSFC19631040 and NSFJ of China.

The combination of random effects and measurement errors for linear models is worth investigating. As pointed out by Davidian and Giltinan (1995) in their open problems (Chapter 12, p. 328), "it is often the case in practice that covariate values collected on individuals are measured with non-negligible errors" and the inference of these models "is not well-developed". Recently Wang *et al.* (1998) and Lin and Carroll (1999) investigated the bias of parameter estimates and variance component tests in generalized linear mixed measurement error models using the simulation extrapolation (SIMEX) approach. Related work on methods for longitudinal data with measurement errors was done by Palta and Lin (1996). In this paper, the corrected score function of Nakamura (1990) was employed to study linear mixed models with measurement errors.

In Section 2, we introduce the corrected score function of Nakamura (1990) for linear mixed models with measurement errors. Section 3 derives asymptotic results for the corrected score estimates of fixed and random effects. The estimates are shown to be consistent and convergent at the order of $n^{-1/2}$. The asymptotic normalities of the estimates are also described. Section 4 is concerned with the numerical aspect of the model. A simple algorithm is developed to obtain the estimates of the parameters and variance components. A simulation study is conducted. An analysis of a data set on hedonic housing-prices is given to illustrate the results in Section 5. Concluding remarks are given in Section 6.

2. The corrected score function

We study the following linear mixed model with errors-in-variables

$$Y = Z\beta + Ub + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I), \tag{2.1a}$$

$$X = Z + \delta, \qquad \delta \sim N(0, I \otimes \Lambda). \tag{2.1b}$$

In this model, Y is an $n \times 1$ vector of random variables whose observed values comprise the data points; Z and U are matrices of regressors with dimensions $n \times p$ and $n \times q$, respectively; β is a $p \times 1$ vector of parameters, which is the fixed effect; b is a $q \times 1$ vector of unobservable random effects with $b \sim N(0, \sigma^2\Sigma)$ for some Σ; ε is an $n \times 1$ vector of random errors as shown in (2.1a); and X is the observed value of Z with the measurement error δ as shown in (2.1b) for some Λ. We assume that σ^2, Σ and Λ are known and should be replaced by precise estimates in applications (see Section 4).

If the measurement error of Z is negligible, i.e. $\delta = 0$, then $E(Y \mid b) = \mu' = \mu_f + \mu_r$ where $\mu_f = Z\beta$ is the fixed component and $\mu_r = Ub$ is the random component, which is the standard linear mixed model (Harville (1977)). For (2.1a), denote the joint probability density of Y and b by $f(y, b; \beta, Z)$ and $l(\beta, b; Z, Y) = \log f(Y, b; \beta, Z)$, which may be regarded as the joint log-likelihood (Robinson (1991)). From (2.1a) we have

$$l(\beta, b; Z, Y) = c(\sigma^2) - \frac{1}{2\sigma^2}(Y - Z\beta - Ub)^T(Y - Z\beta - Ub) - \frac{1}{2\sigma^2}b^T\Sigma^{-1}b, \tag{2.2}$$

where $c(\sigma^2) = -(n/2)\log(2\pi\sigma^2) - (q/2)\log(2\pi\sigma^2) - (1/2)\log|\Sigma|$. For given variance components σ^2 and Σ, Robinson (1991) proposed the best linear unbiased prediction method to estimate the fixed and random effects by solving the equations $\partial l/\partial\beta = 0$ and $\partial l/\partial b = 0$ (see also Harville (1977)). Following this approach and taking $\partial l/\partial b = 0$, we get $\tilde{b}(\beta, Z) = (U^TU + \Sigma^{-1})^{-1}U^T(Y - Z\beta)$. Substituting this formula into (2.2), rearranging the expression and using a well-known identity $(I+U\Sigma U^T)^{-1} = I-U(U^TU+\Sigma^{-1})^{-1}U^T$, we obtain

$$l_p(\beta; Z, Y) = l(\beta, \tilde{b}(\beta); Z, Y) = c(\sigma^2) - \frac{1}{2\sigma^2}(Y - Z\beta)^T V^{-1}(Y - Z\beta), \tag{2.3}$$

where

$$V = I + U\Sigma U^T, \quad \text{and} \quad V^{-1} = I - U(U^TU + \Sigma^{-1})^{-1}U^T.$$

When the covariate Z is measured with non-negligible errors as shown in (2.1b) and the correlated structure arises from the random effects, if we simply replace Z by X in (2.2) and (2.3), then the expectations of $\partial l(\beta, b; X, Y)/\partial b$ and $\partial l_p(\beta; X, Y)/\partial \beta$ with respect to Y and b, evaluated at the true parameter β_t, are generally not equal to zero. Therefore, the estimates obtained from the score functions are not consistent in general. Recently, Wang *et al.* (1998) discussed the bias of estimates due to measurement errors. Various ways are proposed in dealing with measurement error models, and one simple, yet useful approach is based upon the corrected score method by Nakamura (1990) (see also Nakamura (1992), Hanfelt and Liang (1997), Gimenz and Bolfarine (1997) and Zhong *et al.* 2000). The method proposes to find the corrected score function whose expectation with respect to the measurement error distribution coincides with the usual score function in Z (Nakamura (1990)). In the following, we will extend the idea to deal with measurement error models containing random effects as well.

Let E^* and Var* respectively denote the conditional mean and variance with respect to X given b and Y. The corrected log likelihood $l^*(\beta, b; X, Y)$ for our model should satisfy

$$E^*\{\partial l^*(\beta, b; X, Y)/\partial b\} = \partial l(\beta, b; Z, Y)/\partial b, \tag{2.4}$$

$$E^*\{\partial l_p^*(\beta; X, Y)/\partial \beta\} = \partial l_p(\beta; Z, Y)/\partial \beta. \tag{2.5}$$

The following equation is useful to find such a l^*,

$$E^*(X^TV^{-1}X) = Z^TV^{-1}Z + \mathrm{tr}(V^{-1})\Lambda. \tag{2.6}$$

Given Λ, l^* is obtained as

$$l^*(\beta, b; X, Y) = c(\sigma^2) - \frac{1}{2\sigma^2}\{(Y - X\beta - Ub)^T(Y - X\beta - Ub) - \mathrm{tr}(V^{-1})\beta^T\Lambda\beta\} - \frac{1}{2\sigma^2}b^T\Sigma^{-1}b. \tag{2.7}$$

When $V = I$ in (2.7), this reduces to the corrected log likelihood for normal regression models proposed by Nakamura ((1990), p. 131). A corrected likelihood equation $\partial l^*/\partial b = 0$ for b admits an explicit solution

$$\tilde{b}(\beta, X) = (U^TU + \Sigma^{-1})^{-1}U^T(Y - X\beta) = \Sigma U^TV^{-1}(Y - X\beta). \tag{2.8}$$

By an analogous derivation to (2.3), we have

$$l_p^*(\beta; X, Y) = l^*(\beta, \tilde{b}(\beta, X); X, Y) = c(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^TV^{-1}(Y - X\beta) + \frac{1}{2\sigma^2}\mathrm{tr}(V^{-1})\beta^T\Lambda\beta. \tag{2.9}$$

It is straightforward to verify (2.4) and (2.5) with the l^* (2.7) and l_p^* (2.9). A corrected score function for β is then obtained as

$$\frac{\partial l_p^*(\beta; X, Y)}{\partial \beta} = \frac{1}{\sigma^2} X^T V^{-1}(Y - X\beta) + \frac{1}{\sigma^2}\operatorname{tr}(V^{-1})\Lambda\beta. \tag{2.10}$$

Using formula (2.6), we get (2.5), as expected.

Let E^+ and Var^+ denote the expectation and variance with respect to Y and b, and $E = E^+E^*$, then it follows from (2.5) that

$$E\left\{\frac{\partial l_p^*(\beta_t, X, Y)}{\partial \beta}\right\} = 0, \tag{2.11}$$

where β_t is the true value of β. This indicates that the corrected score function is unbiased.

Note that we can obtain the corrected log likelihood (2.9) from (2.3) and (2.6) directly without (2.7). However, to obtain the estimates of random effects, the corrected joint likelihood (2.7) must be employed. The estimates are obtained using the standard procedures as described by Robinson (1991) which performed well for random effects models without measurement errors; see also Harville (1977). These procedures will yield meaningful random effects estimates.

From (2.10), we obtain the corrected observed information as

$$I^*(\beta; X, Y) = -\frac{\partial^2 l_p^*(\beta; X, Y)}{\partial \beta \partial \beta^T} = \sigma^{-2}\{X^T V^{-1} X - \operatorname{tr}(V^{-1})\Lambda\}. \tag{2.12}$$

A corrected observed information is given as

$$I(\beta; Z, Y) = E^*\{I^*(\beta; X, Y)\} = \sigma^{-2}(Z^T V^{-1} Z), \tag{2.13}$$

and a corrected Fisher information is

$$I^+(\beta; Z) = E^+E^*\{I^*(\beta; X, Y)\} = \sigma^{-2}(Z^T V^{-1} Z). \tag{2.14}$$

They are identical to each other since our model (2.1) is linear.

3. Corrected estimates and asymptotic properties

From (2.10) and (2.8), we have corrected estimates

$$\hat{\beta} = \{X^T V^{-1} X - \operatorname{tr}(V^{-1})\Lambda\}^{-1} X^T V^{-1} Y, \tag{3.1}$$

and

$$\hat{b} = (U^T U + \Sigma^{-1})^{-1} U^T (Y - X\hat{\beta}) = \Sigma U^T V^{-1}(Y - X\hat{\beta}). \tag{3.2}$$

The predicted value $\hat{Y} = X\hat{\beta} + U\hat{b}$ of Y coincides with PY; and the residual vector $\hat{e} = Y - \hat{Y} = QY$, where $P = I - V^{-1} + V^{-1}X(X^T V^{-1} X - \operatorname{tr}(V^{-1})\Lambda)^{-1} X^T V^{-1}$ and $Q = I - P$. If $V = I$, i.e. the random effects do not exist, then $\hat{Y}$ and $\hat{e}$ reduce to the results of Nakamura (1990).

Thus, we need to derive asymptotic results for the estimates. It should be noted that Propositions 1 and 2 of Nakamura ((1990), p. 129) may not be applicable to model

(2.1), since the components of Y are not mutually independent. We assume that all the derivatives related to the likelihood exist and the parameter β is identifiable. It is also assumed that as $n \to \infty$, the following limits exist: $n^{-1}Z^TV^{-1}Z$, $n^{-1}Z^TU$, $n^{-1}(U^TU+\Sigma^{-1})$, $n^{-1}Z^TV^{-2}Z$, $n^{-1}\operatorname{tr}(V^{-1})$, and $n^{-1}\operatorname{tr}(V^{-2})$. The existence of the first three limits is also assured in Lee and Nelder ((1996), p. 651), and the last three limits always exist since V^{-1} is positive definite and its eigenvalues are all less than 1.

LEMMA 1. *Under the above assumptions, we have*

$$X^TV^{-1}X = Z^TV^{-1}Z + \operatorname{tr}(V^{-1})\Lambda + O_p(n^{1/2}). \tag{3.3}$$

PROOF. See the Appendix for details. □

THEOREM 1. *$\hat\beta$ is asymptotically normally distributed. The asymptotic mean and variance of $\hat\beta$ are respectively given as β_t and*

$$\operatorname{avar}(\hat\beta) = \sigma^2(Z^TV^{-1}Z)^{-1} + (Z^TV^{-1}Z)^{-1}B(Z^TV^{-1}Z)^{-1}, \tag{3.4}$$

where $B = \{\sigma^2\operatorname{tr}(V^{-1}) + \beta_t^T(Z^TV^{-2}Z)\beta_t\}\Lambda$.

PROOF. It follows from (3.1) and (3.3) that

$$\begin{aligned}\hat\beta &= \{n^{-1}Z^TV^{-1}Z + O_p(n^{-1/2})\}^{-1}n^{-1}X^TV^{-1}Y\\ &= \{I_p + O_p(n^{-1/2})\}^{-1}(n^{-1}Z^TV^{-1}Z)^{-1}n^{-1}X^TV^{-1}Y\\ &= \{I_p + O_p(n^{-1/2})\}(n^{-1}Z^TV^{-1}Z)^{-1}n^{-1}X^TV^{-1}Y\end{aligned}$$

where $\{I_p + O_p(n^{-1/2})\}^{-1} = I_p + O_p(n^{-1/2})$ is obtained from Taylor series expansion. So we have

$$\sqrt{n}\hat\beta = \{I_p + O_p(n^{-1/2})\}(n^{-1}Z^TV^{-1}Z)^{-1}\frac{1}{\sqrt{n}}X^TV^{-1}Y. \tag{3.5}$$

We will obtain asymptotic properties of $\xi = X^TV^{-1}Y/\sqrt{n}$. Let $V^{-1/2} = \Gamma\Phi\Gamma^T$ denote the spectral decomposition of $V^{-1/2}$, where $\Gamma\Gamma^T = I_n$, $\Phi = \operatorname{diag}(\lambda_1^{-1/2},\ldots,\lambda_n^{-1/2})$ and λ_i's are the eigenvalues of V. Then we have

$$\xi = \frac{1}{\sqrt{n}}X^TV^{-1}Y = \frac{1}{\sqrt{n}}X^T\Gamma\Phi\Gamma^TV^{-1/2}Y = \frac{1}{\sqrt{n}}\tilde X^T\Phi\tilde Y,$$

where

$$\begin{aligned}\tilde X &= \Gamma^TX \sim N(\Gamma^TZ, I_n\otimes\Lambda),\\ \tilde Y &= \Gamma^TV^{-1/2}Y \sim N(\Gamma^TV^{-1/2}Z\beta_t, \sigma^2I_n).\end{aligned}$$

The a-th element of ξ is given by

$$\xi_a = \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde X_{ia}\lambda_i^{-1/2}\tilde Y_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n \alpha_i.$$

Since α_i's are independent and the limit of $\operatorname{Var}(\xi_a)$ exists as $n\to\infty$ (see below), by the central limit theorem, ξ_a is asymptotically normal.

Moreover, since the limit of $n^{-1}Z^TV^{-1}Z$ exists, let $M = n^{-1}Z^TV^{-1}Z$, then (3.5) can be written as

$$\sqrt{n}\hat{\beta} = (n^{-1}Z^TV^{-1}Z)^{-1}\frac{1}{\sqrt{n}}X^TV^{-1}Y + O_p(n^{-1/2}) = M^{-1}\xi + O_p(n^{-1/2}). \tag{3.6}$$

It follows from $E(X^TV^{-1}Y) = Z^TV^{-1}Z\beta_t$ or $E(\xi) = M\sqrt{n}\beta_t$ that $\sqrt{n}(\hat{\beta} - \beta_t)$ is asymptotically normal with mean 0.

To find the asymptotic variance of $\hat{\beta}$, (3.6) is rewritten as follows:

$$\sqrt{n}(\hat{\beta} - \beta_t) = M^{-1}\xi - M^{-1}M\sqrt{n}\beta_t + O_p(n^{-1/2}) = M^{-1}(\xi - E(\xi)) + O_p(n^{-1/2}).$$

So we have $\text{avar}(\sqrt{n}\hat{\beta}) = M^{-1}\,\text{Var}(\xi)M^{-1}$. The variance of ξ can be obtained by

$$\begin{aligned}\text{Var}(\xi) &= E^+\{\text{Var}^*(\xi)\} + \text{Var}^+\{E^*(\xi)\}\\ &= n^{-1}E^+(Y^TV^{-2}Y\Lambda) + n^{-1}\,\text{Var}^+(Z^TV^{-1}Y)\\ &= n^{-1}E^+(Y^TV^{-2}Y\Lambda) + n^{-1}\sigma^2(Z^TV^{-1}Z).\end{aligned}$$

Since $E^+(Y^TV^{-2}Y) = \sigma^2\,\text{tr}(V^{-1}) + \beta_t^T(Z^TV^{-2}Z)\beta_t$, $\text{Var}(\xi) = n^{-1}\{B + \sigma^2(Z^TV^{-1}Z)\}$ whose limit exists as $n \to \infty$ by the assumptions. This completes the proof. □

The variance decomposition (3.4) is useful in applications. If there is no measurement error, i.e. $\Lambda = 0$, then $\text{avar}(\hat{\beta}) = \sigma^2(Z^TV^{-1}Z)^{-1}$, which is the variance for mixed models. If there is no random effect, i.e. $V = I_n$, then $B = \{n\sigma^2 + \beta_t^T(Z^TZ)\beta_t\}\Lambda$.

The asymptotic variance (3.4) can be also expressed as

$$I^+(\beta; Z)^{-1} + I^+(\beta; Z)^{-1}(\sigma^{-4}B)I^+(\beta; Z)^{-1}; \tag{3.7}$$

see (2.14). Thus, (3.4) is estimated by

$$I^*(\hat{\beta}; X, Y)^{-1} + I^*(\hat{\beta}; X, Y)^{-1}(\hat{\sigma}^{-4}\hat{B})I^*(\hat{\beta}; X, Y)^{-1}. \tag{3.8}$$

Theorem 1 is an extension of Nakamura ((1990), equations (3) and (4)) to the linear mixed models, and implies that $\hat{\beta}$ is consistent (see (3.6)). Thus, we have

COROLLARY 1. *Let β_t be the true value of β, then $\hat{\beta}$ is consistent in probability and $\hat{\beta} - \beta_t = O_p(n^{-1/2})$.*

The $\hat{\beta}$ is strongly consistent to the order $n^{-1/2}$ under a fixed effect model (Nakamura (1990)). However, it is weakly consistent with the same order of convergence in our model, since it has an extra correlated structure arising from the random effects.

Theorem 1 considers the asymptotic properties of the fixed effect estimators $\hat{\beta}$. For the estimation of random effects (Robinson (1991); Lee and Nelder (1996)) of which no analogous results are obtained in Nakamura (1990), we have

THEOREM 2. *Let $b_t = \tilde{b}(\beta_t)$ given in (2.8), then $\hat{b} - b_t = O_p(n^{-1/2})$ and it is asymptotically normally distributed with the asymptotic variance*

$$\text{avar}(\hat{b} - b_t) = M_1^{-1}M_2\,\text{avar}(\hat{\beta})M_2^TM_1^{-1},$$

where $M_1 = n^{-1}(U^TU + \Sigma^{-1})$ *and* $M_2 = n^{-1}U^TZ$, *whose limits exist by assumption.*

PROOF. From (2.8) and (3.2), we have

$$\begin{aligned}\hat{b} - b_t &= -(U^TU + \Sigma^{-1})^{-1}U^TX(\hat{\beta} - \beta_t)\\ &= -\{n^{-1}(U^TU + \Sigma^{-1})\}^{-1}\{n^{-1}U^TZ + O_p(n^{-1/2})\}(\hat{\beta} - \beta_t)\\ &= -M_1^{-1}M_2(\hat{\beta} - \beta_t) + O_p(n^{-1});\end{aligned}$$

here we use the result $U^TX = U^TZ + O_p(n^{1/2})$ whose proof is similar to (3.3). Then from Theorem 1 we get the desired results. □

The above results are similar to those given by Lee and Nelder ((1996), p. 651), for models without measurement errors.

4. Estimation procedure and its performance in simulation

We have obtained the corrected score function estimates and their asymptotic properties in the last section. This is done by computing the variance components σ^2, $\sigma^2\Sigma$ and Λ, by applying the method of Harville (1977), which has been found to be successful in estimation in various random effects models (see for example, Fellner (1986) and Schall (1991)). The method-of-moment estimate is employed for Λ (Fuller (1987); Carroll *et al.* (1995)).

Following Harville (1977) and Schall (1991), the linear models with random effects and errors-in-variables can be formulated as

$$Y = Z\beta + U_1b_1 + U_2b_2 + \cdots + U_cb_c + \varepsilon, \qquad X = Z + \delta, \tag{4.1}$$

where $\varepsilon \sim N(0, \sigma^2 I_n)$, U_i is $n \times q_i$, b_i is $q_i \times 1$ and $q = q_1 + q_2 + \cdots + q_c$. Let $U = (U_1, U_2, \ldots, U_c)$, $b^T = (b_1^T, b_2^T, \ldots, b_c^T)$. We take $b \sim N(0, \sigma^2\Sigma)$, and $b_1, \ldots, b_c, \varepsilon, X$ are all independent, where $\Sigma = \text{diag}(\sigma_1^2\sigma^{-2}I_{q_1}, \sigma_2^2\sigma^{-2}I_{q_2}, \ldots, \sigma_c^2\sigma^{-2}I_{q_c})$, as discussed by Harville (1977), Fellner (1986) and Schall (1991). Under this formulation, let $D_i = \sigma_i^2/\sigma^2U_i^TV^{-1}$, then from (3.2) we have $\hat{b}_i = D_i(Y - X\hat{\beta})$.

Since $\hat{\beta}$ and $\hat{b}$ are solutions to the simultaneous equations, $\partial l^*/\partial\beta = 0$ and $\partial l^*/\partial b = 0$, we have

$$\begin{pmatrix} X^TX - \text{tr}(V^{-1})\Lambda & X^TU \\ U^TX & U^TU + \Sigma^{-1} \end{pmatrix}\begin{pmatrix}\hat{\beta}\\ \hat{b}\end{pmatrix} \triangleq W\begin{pmatrix}\hat{\beta}\\ \hat{b}\end{pmatrix} = \begin{pmatrix}X^TY\\ U^TY\end{pmatrix}. \tag{4.2}$$

These equations form the basis for the following algorithm which is regarded as an extension of Harville (1977) and Schall's (1991) to deal with the extra measurement errors:

Step 1: Given estimates $\hat{\sigma}^2$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$, compute estimates $\hat{\beta}$ and $\hat{b}_1, \ldots, \hat{b}_c$ as the solutions to the linear equations (4.2), where Σ is evaluated at the current estimates of the variance components.

Step 2: Let T^* be the matrix formed by the last q rows and columns of W^{-1}, partitioned conformably with Σ as

$$\begin{pmatrix} T_{11}^* & \cdots & T_{1c}^* \\ \vdots & & \vdots \\ T_{c1}^* & \cdots & T_{cc}^* \end{pmatrix}.$$

Table 1. Simulation of linear mixed model with errors-in-variables based on 1000 repetitions for case(i): with measurement errors and case(ii): without measurement errors. The corrected score function estimator (CSFE), naive estimator based on X and that based on true Z are employed. The mean, standard error (SE) and mean squared error (MSE) of each estimator are obtained based on repetitions. ASE denotes the asymptotic standard error.

		$q = 50$	$m = 3$			$q = 50$	$m = 8$		
Parameter	Method	Mean	SE	MSE	ASE	Mean	SE	MSE	ASE
case(i)									
$\beta_1 = 1$	CSFE	1.005	0.132	0.0175	0.144	0.998	0.077	0.0059	0.070
	Naive	0.798	0.090	0.0488	0.090	0.801	0.053	0.0428	0.055
	True Z	0.998	0.057	0.0033	0.056	1.000	0.032	0.0010	0.032
$\beta_2 = 2$	CSFE	2.013	0.134	0.0182	0.134	2.011	0.079	0.0064	0.070
	Naive	1.603	0.092	0.1661	0.091	1.601	0.057	0.1626	0.054
	True Z	1.996	0.056	0.0032	0.056	2.000	0.032	0.0010	0.032
$\sigma_1^2 = 0.25$	CSFE	0.250	0.179	0.0323	—	0.249	0.080	0.0065	—
	Naive	0.267	0.151	0.0232	—	0.355	0.073	0.0163	—
	True Z	0.256	0.071	0.0051	—	0.250	0.060	0.0036	—
$\sigma^2 = 0.36$	CSFE	0.380	0.200	0.0403	—	0.384	0.080	0.0070	—
	Naive	1.323	0.151	0.9505	—	1.336	0.073	0.9584	—
	True Z	0.356	0.071	0.0051	—	0.358	0.060	0.0036	—
case(ii)									
$\beta_1 = 1$	CSFE	0.997	0.055	0.0030	0.062	1.000	0.032	0.0010	0.033
$\beta_2 = 2$	CSFE	2.000	0.056	0.0031	0.062	2.000	0.032	0.0010	0.033
$\sigma_1^2 = 0.25$	CSFE	0.249	0.050	0.0025	—	0.250	0.051	0.0026	—
$\sigma^2 = 0.36$	CSFE	0.377	0.050	0.0028	—	0.361	0.051	0.0026	—

Compute the estimates of σ^2 and $\sigma_1^2, \ldots, \sigma_c^2$ as

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^T \hat{V}^{-1}(Y - X\hat{\beta}) - \mathrm{tr}(\hat{V}^{-1})\hat{\beta}^T \hat{\Lambda}\hat{\beta}}{n - q^*}; \qquad \hat{\sigma}_i^2 = \frac{\hat{b}_i^T \hat{b}_i - \mathrm{tr}(\hat{D}_i^T \hat{D}_i \hat{V}^{-1})}{q_i - v_i^*},$$

where $v_i^* = \hat{\sigma}^2 \,\mathrm{tr}(T_{ii}^*)/\hat{\sigma}_i^2$ and $q^* = \sum_{i=1}^c (q_i - v_i^*)$ evaluated at their current estimates are penalties under the general linear mixed measurement error models, and the term $\mathrm{tr}(\hat{V}^-)\hat{\beta}^T\hat{\Lambda}\hat{\beta}$ and $\mathrm{tr}(\hat{D}_i^T \hat{D}_i \hat{V}^{-1})$ are the corrections for extra measurement errors.

If a necessary convergence criterion is satisfied, repeat Step 1 and quit; otherwise return to Step 1 to continue. Following the argument of Schall (1991), this algorithm is analogous to the algorithm yielding the maximum corrected likelihood estimates of the parameters in the normal random effects model; see also Fellner (1986).

The performance of estimators described above is evaluated using simulations. The response y_{ij} is simulated from the model $y_{ij} = z_{ij}^{(1)}\beta_1 + z_{ij}^{(2)}\beta_2 + b_{1j} + \varepsilon_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, q$, where q can be interpreted as the number of independent clusters and m is the cluster size in a longitudinal study, and the total size is $n = mq$ (Wang *et al.* (1998)). When this model is formulated in a matrix form as in (4.1), we write $c = 1$, $b_1 = (b_{11}, \ldots, b_{1q})^T$, $Y = (y_{11}, \ldots, y_{1q}, y_{21}, \ldots, y_{2q}, \ldots, y_{m1}, \ldots, y_{mq})^T$, and Z and ε are rewritten in accordance with Y. The following combinations were taken for simulation:

$q = 50$, $m = 3$ or 8, which are common sample sizes in longitudinal studies, $\beta_1 = 1$, $\beta_2 = 2$, $z_{ij} \sim N(0,1)$, $b_{1j} \sim N(0, \sigma_1^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, where $\sigma_1^2 = 0.5^2$, $\sigma^2 = 0.6^2$. We consider two cases: (i) with measurement errors ($\Lambda = \text{diag}(0.5^2, 0.5^2)$) and (ii) without measurement errors ($\Lambda = 0$). These parameter values are similar to those treated by Wang *et al.* (1998) and Nakamura (1990) in their simulation studies. The simulation study was conducted using the MATLAB software. For each combination of parameters, 1000 repetitions were performed.

The proposed corrected estimators, usual naive estimators ignoring measurement errors in X and score function estimators based on the true value Z are investigated. The summary results are presented in Table 1.

For the smaller sample size of $n = mq = 150$ with measurement errors, the corrected estimators perform very well. The mean values of $\hat{\beta}_1$ and $\hat{\beta}_2$ are very close to their true values. By contrast, the naive estimator are biased towards zero. The biases cause, in turn, the error variance estimate $\hat{\sigma}^2$ to be overestimated to a rather large extent. On the other hand, the corrected score method for estimating σ^2 is nearly unbiased. The method also performs well for estimating the variance component σ_1^2, but the naive estimator for σ_1^2 is nearly unbiased just by coincidence. The estimators based on Z are obviously consistent, as shown in Table 1. The entry ASE gives the asymptotic standard error of $\hat{\beta}$ using equation (3.8), which is in agreement with the observed standard error given in entry SE based on the 1000 repetitions in the simulation.

For the larger sample size $q = 50$, $m = 8$, we also observe a similar phenomenon; see the right columns of Table 1. The corrected score method gives nearly unbiased estimates for the regression coefficients and the variance components. However, the estimates of the naive method are biased. The corrected estimates outperform the naive estimates when measurement errors exist.

5. An example

Harrison and Rubinfeld (1978) constructed a hedonic housing price model to measure the willingness to pay for clean air. Their study utilized data for census tracts in the Boston Standard Statistical Area in 1970. The (logarithm) median value of the owner-occupied homes in the census tract was taken as the dependent variable in a fixed effects regression model. The independent variables chosen include attribute variables, neighborhood variables, accessibility variables, and one air pollution variable on the concentration of nitrogen oxides (NOX). A description of these data can be found in Harrison and Rubinfeld (1978).

The data of $n = 132$ census tracts within the 15 districts of the Boston city is selected for our study. We follow the regression model of Harrison and Rubinfeld (1978). However, the census tracts within districts are taken as repeated measurements and so a mixed effects linear model is employed. All independent variables can be measured precisely except the pollution variable NOX which is taken to have measurement errors. Table 2 gives the parameter estimates and t-ratios for our mixed effects model with errors-in-variable. The corresponding results under the fixed effects model of Harrison and Rubinfeld (1978) are listed for comparison. It is found that several parameter estimates of the two models are not too close to each other. If we look at the t-ratios, the magnitudes are generally smaller for our model. This is often the case since the standard errors for our model are higher due to random effects and measurement errors. In particular, the t-ratio for the pollution variable NOX^2 shows that the variable may not

Table 2. Corrected score function estimates (CSFE) of the linear mixed model with errors-in-variables, and maximum likelihood estimates (MLE) of Harrison-Rubinfeld fixed effects linear model, for the hedonic housing price data of Boston city. The t-ratios are in brackets.

Variable	CSFE	MLE
Intercept	9.14(15.5)	8.87(21.4)
RM^2	$-2.3 \times 10^{-3}(-0.49)$	$-5.3 \times 10^{-3}(-1.38)$
AGE	$1.1 \times 10^{-3}(0.31)$	$3.3 \times 10^{-3}(0.98)$
Log (DIS)	$1.4 \times 10^{-3}(0.005)$	$2.87 \times 10^{-2}(0.19)$
$(B - 0.63)^2$	$3.46 \times 10^{-1}(1.18)$	$1.76 \times 10^{-1}(0.85)$
log (STAT)	$-5.75 \times 10^{-1}(-5.08)$	$-6.71 \times 10^{-1}(-7.94)$
CRIM	$-7.6 \times 10^{-3}(-3.12)$	$-9.8 \times 10^{-3}(-3.88)$
CHAS	$2.2 \times 10^{-3}(0.02)$	$1.86 \times 10^{-1}(1.29)$
NOX^2	$-1.18 \times 10^{-2}(-1.40)$	$-1.15 \times 10^{-2}(-2.79)$
σ	0.255	0.330
σ_1	0.198	—

be statistically significant in our model, while in the fixed effects model it is significant.

6. Concluding remarks

In this paper, we propose the corrected score method for correcting for the measurement errors of independent variables in linear mixed models. We show that the proposed estimators are consistent to the order of $n^{-1/2}$. This order of convergence of the estimators is the same as that in the estimation for standard measurement errors models without random effects (Nakamura (1990)), and in the estimation for mixed-effects models without measurement errors. The combination of measurement errors and random effects does not decrease the order of convergence. The proposed estimators are also found to perform well in finite sample cases of the simulation study.

In the analysis of the hedonic housing price, our method does not indicate any significant effect of the NOX level after correcting for the effect of the measurement errors. Lin and Carroll (1999) also found a lower significance level after adjusting for measurement errors. In general, the signficance level after correcting for measurement errors may be higher or lower than naive ones depending on particular data, though both levels are asymptotically equal with each other (Stefanski and Carroll (1990)). Thus, more comprehensive investigation into the data would be needed in the future for further understanding of the data.

Acknowledgements

The authors thank the referees for their very helpful comments that greatly improved the presentation of the paper.

Appendix

PROOF OF LEMMA 1. Using (2.1b), we have

$$n^{-1}\{X^TV^{-1}X - Z^TV^{-1}Z - \text{tr}(V^{-1})\Lambda\} = n^{-1}(Z^TV^{-1}\delta + \delta^TV^{-1}Z + C),$$

where $C = \delta^T V^{-1}\delta - \text{tr}(V^{-1})\Lambda$. Since $\delta \sim N(0, I_n \otimes \Lambda)$, we have

$$n^{-1/2} Z^T V^{-1}\delta \sim N(0, n^{-1} Z^T V^{-2} Z \otimes \Lambda).$$

By assumption, $n^{-1} Z^T V^{-2} Z$ exists as $n \to \infty$. So we have $n^{-1} Z^T V^{-1}\delta = O_p(n^{-1/2})$. Similarly, $n^{-1}\delta^T V^{-1} Z = O_p(n^{-1/2})$ holds. Now let the elements of C at the (a,b) position be C_{ab}. Then

$$C_{ab} = \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_{ia} V^{ij} \delta_{jb} - \sum_{i=1}^{n} V^{ii} \Lambda_{ab},$$

where $\delta = (\delta_{ia})$, $V^{-1} = (V^{ij})$, $\Lambda = (\Lambda_{ab})$, $i, j = 1, 2, \ldots, n$, and $a, b = 1, 2, \ldots, p$. It is easily seen that $E(\delta_{ia}\delta_{ib}) = \Lambda_{ab}$ and $E(C_{ab}) = 0$. Further, we have

$$\begin{aligned} E(C_{ab})^2 &= \sum_{i,j}\sum_{k,l} E(\delta_{ia}\delta_{jb}\delta_{ka}\delta_{lb}) V^{ij} V^{kl} - \Lambda_{ab}^2 \{\text{tr}(V^{-1})\}^2 \\ &= (\Lambda_{aa}\Lambda_{bb} + \Lambda_{ab}^2)\, \text{tr}(V^{-2}). \end{aligned}$$

The above result is obtained by summing the non-zero expected value terms with equal indices and with pairwise equal indices, while the other terms are all zeros. By assumption, $n^{-1}\,\text{tr}(V^{-2})$ exists, so we have $E(n^{-1/2} C_{ab})^2 = O_p(1)$ as $n \to \infty$ and then $n^{-1} C = O_p(n^{-1/2})$. Combining all the above results, we get (3.3). □

References

Armstrong, B. (1985). Measurement error in the generalized linear model, *Comm. Statist. Simulation Comput.*, **14**, 529–544.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *J. Amer. Statist. Assoc.*, **88**, 9–25.

Buzas, J. S. and Stefanski, L. A. (1996). A note on corrected-score estimation, *Statist. Probab. Lett.*, **28**, 1–8.

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

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

Fellner, W. H. (1986). Robust estimation of variance components, *Technometrics*, **28**, 51–60.

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

Gimenz, P. and Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models, *Austral. J. Statist.*, **39**, 325–344.

Hanfelt, J. J. and Liang K. Y. (1997). Approximate likelihood for generalized linear errors-in-variables models, *J. Roy. Statist. Soc. Ser. B*, **59**, 627–637.

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management*, **5**, 81–102.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and related problems (with discussion), *J. Amer. Statist. Assoc.*, **72**, 320–340.

Lee, Y. and Nelder, N. A. (1996). Hierarchical generalized linear models (with discussion), *J. Roy. Statist. Soc. Ser. B*, **58**, 619–678.

Lin, X. and Breslow, N. E. (1996). Bias correction in generalized linear models with multiple components of dispersion, *J. Amer. Statist. Assoc.*, **91**, 1007–1016.

Lin, X. and Carroll, R. J. (1999). SIMEX variance component tests in generalized linear mixed measurement error models, *Biometrics*, **55**, 613–619.

Liu, Q. and Pierce, D. A. (1993). Heterogeneity in Mantel-Haenzel-Type models, *Biometrika*, **80**, 543–556.

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

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error, *Biometrics*, **48**, 829–838.

Palta, M. and Lin, C.-Y. (1996). Latent variables, measurement error and method for analyzing longitudinal ordinal data, *Proceedings of American Statistical Association in Biometrics*, 340–345.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion), *Statist. Sci.*, **6**, 15–51.

Schall, R. (1991). Estimation in generalized linear models with random effects, *Biometrika*, **78**, 719–727.

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

Stefanski, L. A. and Carroll, R. J. (1990). Score tests in generalized linear measurement error models, *J. Roy. Statist. Soc. Ser. B*, **52**, 345–359.

Wang, N., Lin, X., Gutierrez, R. G. and Carroll, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models, *J. Amer. Statist. Assoc.*, **93**, 249–261.

Zhong, X. P., Wei, B. C. and Fung, W. K. (2000). Influence analysis for linear measurement error models, *Ann. Inst. Statist. Math.*, **52**, 367–379.