Statistica Sinica **10**(2000), 885-903

ESTIMATION IN LONGITUDINAL RANDOM EFFECTS MODELS WITH MEASUREMENT ERROR

John Buonaccorsi, Eugene Demidenko* and Tor Tosteson*

*University of Massachusetts and *Dartmouth Medical School*

Abstract: Estimation of the regression parameters and variance components in a longitudinal mixed model with measurement error in a time-varying covariate is considered. The positive bias in variance estimators caused by covariate measurement error in a normal linear mixed model has recently been identified and studied (Tosteson, Buonaccorsi and Demidenko (1997)). The methods suggested there for correction of the bias involve convenient adaptations of existing software for a particular model. In this paper, we study alternative methods of estimation which achieve higher efficiencies and extend readily to a more general class of models. Full and pseudo-maximum likelihood estimators under normality are considered as is a pseudo-moment approach relying on initial estimation of nuisance parameters. The latter lead to a "regression calibration" method for estimating the regression parameters, in which a substitution is made for the unknown covariates, followed by a correction for estimation of the variance parameters. It is shown that for some cases this yields the pseudo-maximum likelihood estimates and, in these cases, the resulting estimators are highly efficient relative to the full maximum likelihood estimators. We first consider a model with no additional data, where identifiability follows from assumptions about the longitudinal model for the unobserved true covariates, and then describe some extensions to cases where either replicate or validation data is available. We illustrate with an example investigating the relationship between dietary and serum beta-carotene.

Key words and phrases: Longitudinal model, maximum likelihood, measurement error, random effects.

1. Introduction

The linear mixed model (Laird and Ware (1982)) is a popular choice for the treatment of longitudinal models with random effects. As in most regression models, difficulties can arise when some of the predictors/covariates are subject to measurement error. For example, in epidemiologic studies time-varying covariates such as dietary intakes or other exposure variables are often mismeasured. Fuller (1987) and Carroll, Stefanski and Ruppert (1995) provide a comprehensive treatment of measurement error in standard linear and nonlinear regression models, where the main objective is inferences for the regression coefficients. Recent work has begun investigating the role of measurement error in mixed models, both linear (Tosteson, Buonaccorsi and Demidenko (1998), hereafter referred to as TBD, Paterno and Amemiya (1996), Wang, Lin, Gutierrez and Carroll (1998)) and nonlinear (Wang and Davidian (1996), Carroll, Lin and Wang (1997), Wang, Lin, Gutierrez and Carroll (1998)) and Wang, Lin and Gutierrez (1999)). In this paper we examine estimation of both regression coefficients and variance parameters for a class of linear mixed models with measurement error in a time-varying covariate.

We begin with consideration of the model treated by Paterno and Amemiya (1996) and TBD in which

$$\mathbf{Y}_i \mid (\mathbf{X}_i, \mathbf{D}_i) = \mathbf{X}_i\boldsymbol{\beta} + \gamma\mathbf{D}_i + \mathbf{Z}\boldsymbol{\nu}_i + \boldsymbol{\epsilon}_i, i = 1, \ldots, n. \tag{1}$$

Here for "subject" i, $\mathbf{Y}_i$ is an $t \times 1$ random vector of outcomes, $\mathbf{X}_i$ and $\mathbf{Z}$ are known matrices of size $t \times p$ and $t \times q$ respectively, $\mathbf{D}_i$ is a $t \times 1$ vector of time-varying covariates subject to measurement error, $\boldsymbol{\nu}_i$ is a $q \times 1$ vector of random effects with mean $\mathbf{0}$ and covariance $\boldsymbol{\Omega}$, $\boldsymbol{\epsilon}_i$ is a $t \times 1$ random vector with mean $\mathbf{0}$ and covariance $\sigma_\epsilon^2\mathbf{I}$, and $\boldsymbol{\beta}$ $(t \times 1)$ and γ (scalar) are unknown parameters. The $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\nu}_i$ are assumed independent of each other and independent over subjects. Hence the $\mathbf{Y}_i$ are independent and $\mathrm{Cov}\,(\mathbf{Y}_i \mid \mathbf{X}_i, \mathbf{D}_i) = \mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z} + \sigma_\epsilon^2\mathbf{I}$. The fixed effects part of the model has been split into two pieces to isolate the portion involving the error-prone predictors, while the use of scalar γ implies that the effect of the predictor D is the same for each repeated measure within a subject. The assumption that $\mathbf{Z}$ is known prohibits the mismeasured D variable from entering into the random effects part of the model. Typically $\mathbf{Z}$ will capture random subject effects and random trends over time, as illustrated in the example in Section 4.

The measurement error is assumed additive with

$$\mathbf{W}_i = \mathbf{D}_i + \boldsymbol{\delta}_i, \tag{2}$$

where $\mathbf{W}_i$ is the error-prone measure of $\mathbf{D}_i$, $E(\boldsymbol{\delta}_i) = \mathbf{0}$, and $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n$ are assumed independent. In the absence of any additional data, some restriction must be made in order to account for measurement error. Paterno and Amemiya (1996) allow a general $\boldsymbol{\Sigma}_{\delta i}$ for $\mathrm{Cov}\,(\boldsymbol{\delta}_i)$, but without additional data they impose a restriction on the mean model, discussed in more detail in the next section. TBD were motivated to consider this model by an application where $\mathbf{Y}_i$ is a 6×1 vector of serum beta-carotene measurements, and $\mathbf{D}_i$ is a 6×1 vector of diet intakes of beta-carotene, measured by $\mathbf{W}_i$, based on a food frequency questionnaire. In the absence of replicate food frequency measurements or validation data, TBD used a restricted longitudinal model for the $\mathbf{D}_i$; see (7). Using this model, the focus of TBD was the bias in certain naive estimators and the development of a simple approach for correcting for measurement error to apply to the diet data. In addition to questions of efficiency, the methods used there do not readily extend to more general settings.

Section 2 provides more detail on the models, and Section 3 develops estimators using full and pseudo-maximum likelihood estimators under normality and a pseudo-moment approach without normality. An analysis of the beta-carotene example is presented in Section 4, followed by efficiency calculations in Section 5. Until Section 6, we limit ourselves to (1) in conjunction with (2) and (7) in order to examine the main issues in a relatively simple setting. Section 6 outlines extensions of the methodology to more complicated settings, including the use of replicate measurements, the use of validation data under linear measurement error and the treatment of more general longitudinal models.

2. Models

In addition to (1) and (2), a structural model is assumed in which the $\mathbf{D}_i$ are independent with $E(\mathbf{D}_i) = \boldsymbol{\mu}_{Di}$ and $\mathrm{Cov}\,(\mathbf{D}_i) = \boldsymbol{\Sigma}_D$. Both $\boldsymbol{\delta}_i$ in (2) and $\mathbf{D}_i$ are independent of random quantities in (1). Thus,

$$E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta} + \gamma\boldsymbol{\mu}_{Di},\ \ E(\mathbf{W}_i) = \boldsymbol{\mu}_{Di},\ \ \mathrm{Cov}\,(\mathbf{Y}_i) = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \boldsymbol{\sigma}^2_\epsilon\mathbf{I} + \gamma^2\boldsymbol{\Sigma}_D, \quad (3)$$

$$\mathrm{Cov}\,(\mathbf{W}_i) = \boldsymbol{\Sigma}_W = \boldsymbol{\Sigma}_D + \sigma^2_\delta\mathbf{I},\ \text{and}\ \mathrm{Cov}\,(\mathbf{Y}_i,\mathbf{W}_i) = \boldsymbol{\Sigma}_{YW} = \gamma\boldsymbol{\Sigma}_D. \quad (4)$$

Under multivariate normality this induces the model

$$\mathbf{Y}_i \mid (\mathbf{X}_i,\mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + \gamma\mathbf{Q}_i + \boldsymbol{\eta}^*_i \quad (5)$$

where $\mathbf{Q}_i = (\mathbf{I} - \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1})\boldsymbol{\mu}_{Di} + \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1}\mathbf{W}_i$, and $\boldsymbol{\eta}^*_i$ has covariance

$$\boldsymbol{\Psi} = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma^2_\epsilon\mathbf{I} + \gamma^2\boldsymbol{\Sigma}_D(\mathbf{I} - \boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_D). \quad (6)$$

Our methods of estimation are motivated by this conditional model but are generally robust to normality in that they provide consistent estimators under (3) and (4) only.

Naive estimation refers to fitting (1) using $\mathbf{W}_i$ in place of $\mathbf{D}_i$. It is clear from (5) that naive estimators of either γ or $\boldsymbol{\Sigma}$ will usually be biased, but it is not always possible to immediately identify the bias since the fixed effects part of (5) is not always of the form $\mathbf{X}_i\boldsymbol{\beta}^* + \gamma^*\mathbf{W}_i$, nor does Ψ usually always have the form $\mathbf{Z}'\boldsymbol{\Omega}^*\mathbf{Z} + \sigma^{*2}_\epsilon\mathbf{I}$. There are some special cases (see TBD and Wang, Lin, Gutierrez and Carroll (1998)) where the induced model is of the same form as (1) and the biases are easily established. The asymptotic biases can also be examined through the estimating equations defining the naive estimators, an approach taken by Wang, Lin, Gutierrez and Carroll (1998). Their Sections 3.1 and 4.1 provide a bias analysis for special cases of our linear model, with $\mathbf{X}_i = \mathbf{1}$ and $\mathbf{Z} = \mathbf{1}$. We will not pursue further details concerning bias here.

In the absence of additional data, some assumption must be made in order to allow a correction for measurement error. Until Section 6 we follow TBD, and enforce identifiability by assuming the error-prone covariates follow the mixed model

$$\mathbf{D}_i = \mathbf{A}_i\boldsymbol{\alpha} + \mathbf{R}\boldsymbol{\phi}_i, \tag{7}$$

where $\mathbf{A}_i$ is a known $t \times r$ matrix, $\boldsymbol{\alpha}$ is an $r \times 1$ vector of parameters, $\mathbf{R}$ is a known $t \times q$ matrix of rank $q < t$ and $\boldsymbol{\phi}_i$ is a $q \times 1$ random effect with mean $\mathbf{0}$ and a nonsingular covariance $\boldsymbol{\Omega}_D$. The $\mathbf{A}_i\boldsymbol{\alpha}$ captures the fixed effects part of the model for dietary intake, while $\mathbf{R}$ models the random effects. Similar to $\mathbf{Z}$ in (1), it allows for random subject effects and random time trends over subjects. Note that $\boldsymbol{\Sigma}_D = \mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}'$, which is singular.

Model (7), in combination with (2), results in the mixed model

$$\mathbf{W}_i = \mathbf{A}_i\boldsymbol{\alpha} + \mathbf{R}\boldsymbol{\phi}_i + \boldsymbol{\delta}_i \tag{8}$$

for the observed $\mathbf{W}$'s. The lack of additional residual error in (7) therefore allows for estimation of $\boldsymbol{\alpha}$, σ_δ^2, and $\boldsymbol{\Omega}_D$ from the W data. The assumption in (7) may seem quite strong, but see TBD for further discussion of this model, its relationship to other work, and an assessment of the assumed model in relation to the beta-carotene data. Models which relax this assumption are discussed in Section 6.

Under (7), $\boldsymbol{\eta}_i^*$ in (5) can be written as $\mathbf{Z}\boldsymbol{\nu}_i+\mathbf{R}\boldsymbol{\tau}_i+\boldsymbol{\epsilon}_i$, where $\boldsymbol{\tau}_i$ has covariance $\boldsymbol{\Omega}_\tau = \gamma^2(\boldsymbol{\Omega}_D - \boldsymbol{\Omega}_D\mathbf{R}'\boldsymbol{\Sigma}_W^{-1}\mathbf{R}\boldsymbol{\Omega}_D)$ and so

$$\boldsymbol{\Psi} = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \mathbf{R}\boldsymbol{\Omega}_\tau\mathbf{R}' + \sigma_\epsilon^2\mathbf{I}. \tag{9}$$

A special case of interest is when $\mathbf{Z} = \mathbf{R}$, in which case $\mathbf{Z}\boldsymbol{\nu}_i + \mathbf{R}\boldsymbol{\tau}_i = \mathbf{Z}\boldsymbol{\tau}_i^*$ where $\boldsymbol{\tau}_i^*$ has covariance

$$\boldsymbol{\Omega}^* = \boldsymbol{\Omega} + \boldsymbol{\Omega}_\tau, \tag{10}$$

resulting in

$$\boldsymbol{\Psi} = \mathbf{Z}\boldsymbol{\Omega}^*\mathbf{Z}' + \sigma_\epsilon^2\mathbf{I}. \tag{11}$$

In this case the covariance structure is the same as in the original model (1) but with $\boldsymbol{\Omega}^*$ in place of $\boldsymbol{\Omega}$. As noted in TBD, a naive approach leads to overestimation of $\boldsymbol{\Omega}$.

In summary, until Section 6, the model under consideration is given by (1), (2) and (7), in which case (5) reduces to

$$\mathbf{Y}_i \mid (\mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + \gamma\mathbf{Q}_i + \mathbf{Z}\boldsymbol{\nu}_i + \mathbf{R}\boldsymbol{\tau}_i + \boldsymbol{\epsilon}_i \tag{12}$$

with $\boldsymbol{\Psi} = \mathrm{Cov}\,(\mathbf{Y}_i|\mathbf{W}_i)$ modeled by (9) or, when $\mathbf{Z} = \mathbf{R}$, by (11).

An alternative to (7) to ensure identifiability is to restrict the mean structure, but this appears to be of limited practical value. Consider the case with $\mathbf{X}_i = \mathbf{X}$ and $\mathbf{A}_i = \mathbf{A}$. Since $\boldsymbol{\mu}_D = \mathbf{A}\boldsymbol{\alpha}$ is identifiable from the $\mathbf{W}$ data, γ is identifiable from the mean structure if the matrix $[\mathbf{X}\ \boldsymbol{\mu}_D]$ is of full column rank. This will not hold if the columns of $\mathbf{A}$ are a subset of the columns of $\mathbf{X}$. In particular if one of the columns of $\mathbf{X}$ is $\mathbf{1}$ (a vector of 1's), as is often the case, then $\boldsymbol{\mu}_D$ cannot be of the form $c\mathbf{1}$ for some constant c. For the beta-carotene data, the mean diet is fairly constant over time, so this is not a useful strategy for that problem. Paterno and Amemiya (1996), treating a special case of our model with $\mathbf{X}_i = \mathbf{1}$, proceed under the assumption that $[\mathbf{1}, \boldsymbol{\mu}_D]$ is of full rank.

3. Estimation

3.1. Methods

Let $\boldsymbol{\omega} = vech(\boldsymbol{\Omega})$ and $\boldsymbol{\omega}_D = vech(\boldsymbol{\Omega}_D)$ contain the distinct components of $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_D$ respectively. See Fuller (1987, Appendix 4A) for a discussion of matrix-vector operations. Let $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$, where $\boldsymbol{\theta}_1' = (\boldsymbol{\beta}', \gamma, \boldsymbol{\omega}', \sigma_\epsilon^2)$ and $\boldsymbol{\theta}_2' = (\boldsymbol{\alpha}', \boldsymbol{\omega}_D', \sigma_\delta^2)$, contain the parameters in the model for $\mathbf{Y}|\mathbf{D}$ and the marginal model for $\mathbf{W}$ respectively. The elements of $\boldsymbol{\theta}_1$ are of primary interest.

Under normality $(\mathbf{Y}_i, \mathbf{W}_i)$ has density $f(\mathbf{y}_i, \mathbf{w}_i; \boldsymbol{\theta})$, the multivariate normal density with mean and covariance terms as given in (3). Equivalently $f(\mathbf{y}_i, \mathbf{w}_i; \boldsymbol{\theta}) = f(\mathbf{y}_i|\mathbf{w}_i; \boldsymbol{\theta}) f(\mathbf{w}_i; \boldsymbol{\theta}_2)$ where $f(\mathbf{y}_i|\mathbf{w}_i; \boldsymbol{\theta})$ is a normal density based on (12) and $f(\mathbf{w}_i|\boldsymbol{\theta}_2)$ is a normal density with mean $\boldsymbol{\mu}_{Wi}$ and covariance $\boldsymbol{\Sigma}_W$. The *full maximum likelihood estimator* (MLE) maximizes $L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \prod_i f(\mathbf{y}_i \mid \mathbf{w}_i; \boldsymbol{\theta}) f(\mathbf{w}_i; \boldsymbol{\theta}_2) = \prod_i f(\mathbf{y}_i, \mathbf{w}_i; \boldsymbol{\theta})$, as a function of $\boldsymbol{\theta}$. The full ML estimators are denoted by a $_{ML}$ subscript, e.g., $\hat{\gamma}_{ML}$.

We refer to *pseudo methods* as methods in which an estimate $\hat{\boldsymbol{\theta}}_2$ of $\boldsymbol{\theta}_2$ is first obtained from the $\mathbf{W}_i$ data, based on (8), and this is then used in estimating $\boldsymbol{\theta}_1$. From $\hat{\boldsymbol{\theta}}_2$, $\hat{\boldsymbol{\Sigma}}_D = \mathbf{R}\hat{\boldsymbol{\Omega}}_D\mathbf{R}'$ and $\hat{\boldsymbol{\Sigma}}_W = \mathbf{R}\hat{\boldsymbol{\Omega}}_D\mathbf{R}' + \hat{\sigma}_\delta^2\mathbf{I}$ are obtained.

The *pseudo-maximum likelihood estimator* (PMLE), as defined by Gong and Sammaniego (1981), maximizes $L(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)$ as a function of $\boldsymbol{\theta}_1$. Using (12), the pseudo-likelihood is based on the use of

$$\mathbf{Y}_i \mid \mathbf{W}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \hat{\mathbf{Q}}_i\gamma, \boldsymbol{\Psi}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)), \tag{13}$$

where $\hat{\mathbf{Q}}_i = (\mathbf{I} - \hat{\boldsymbol{\Sigma}}_D\hat{\boldsymbol{\Sigma}}_W^{-1})\mathbf{A}_i\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\Sigma}}_D\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{W}_i$ and $\boldsymbol{\Psi}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2) = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma_\epsilon^2\mathbf{I} + \gamma^2\hat{\boldsymbol{\Sigma}}_D(\mathbf{I} - \hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D)$. The pseudo-MLEs are denoted by a $_{PML}$ subscript, e.g., $\hat{\gamma}_{PML}$.

The form of $\boldsymbol{\Psi}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)$ means the pseudo-MLE's are not immediately fit using standard software. One approach to alleviate this problem is to fit $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \hat{\mathbf{Q}}_i\gamma + \boldsymbol{\eta}_i^*$ using standard linear mixed software, with the covariance structure for

$\boldsymbol{\eta}_i^*$ assumed to be one of the structures allowed by the fitting routine. This leads directly to estimators, say $\hat{\boldsymbol{\beta}}_C$ and $\hat{\gamma}_C$, for the regression coefficients. Estimation of the variance terms depends on the form of $\boldsymbol{\Psi}$. When $\mathbf{Z} = \mathbf{R}$, $\boldsymbol{\Psi} = \text{Cov}(\boldsymbol{\eta}_i^*) = \mathbf{Z}\boldsymbol{\Omega}^*\mathbf{Z}' + \sigma_\epsilon^2\mathbf{I}$, and the mixed model approach to fitting $\mathbf{Y}_i$ given $\hat{\mathbf{Q}}_i$, leads directly to $\hat{\sigma}^2_{\epsilon C}$ and $\hat{\boldsymbol{\Omega}}^*_C$, where $\hat{\boldsymbol{\Omega}}^*_C$ estimates $\boldsymbol{\Omega}^*$ in (10). From this, $\boldsymbol{\Omega}$ is estimated by $\hat{\boldsymbol{\Omega}}_C = \hat{\boldsymbol{\Omega}}^*_C - \hat{\boldsymbol{\Omega}}_{\tau C}$, where $\hat{\boldsymbol{\Omega}}_{\tau C} = \hat{\gamma}_C^2(\hat{\boldsymbol{\Omega}}_D - \hat{\boldsymbol{\Omega}}_D\mathbf{R}'\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{R}\hat{\boldsymbol{\Omega}}_D)$. Under normality, $\hat{\boldsymbol{\beta}}_C$ and $\hat{\gamma}_C$ are *regression calibration* (RC) estimators, as defined by Carroll, Stefanski and Ruppert (1995), since $\hat{\mathbf{Q}}_i$ is an estimate of $E(\mathbf{D}_i|\mathbf{W}_i) = (\mathbf{I} - \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1})\boldsymbol{\mu}_{Di} + \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1}\mathbf{W}_i$. For convenience we also refer to the variance components estimators as regression calibration estimators. This approach is readily extended to more general models as seen in Section 6.

Proposition 1. *Under* (1), (2) *and* (7) *with* $\mathbf{Z} = \mathbf{R}$, *the regression calibration estimators are equivalent to the pseudo-maximum likelihood estimators under normality, so long as* $\hat{\boldsymbol{\Omega}}_C$ *is nonnegative and* $\mathbf{Z}\hat{\boldsymbol{\Omega}}_C\mathbf{Z}' + \hat{\sigma}_\epsilon^2\mathbf{I}$ *is positive definite.*

To show Proposition 1, define $\boldsymbol{\theta}_1^*$ to be the same as $\boldsymbol{\theta}_1$ but with $\boldsymbol{\omega}^* = vech(\boldsymbol{\Omega}^*)$ in place of $\boldsymbol{\omega}$. Now $\boldsymbol{\omega} = \boldsymbol{\omega}^* - \gamma^2\boldsymbol{\pi}$, where $\boldsymbol{\pi} = vech(\boldsymbol{\Omega}_D - \boldsymbol{\Omega}_D\mathbf{R}'\boldsymbol{\Sigma}_W^{-1}\mathbf{R}\boldsymbol{\Omega}_D)$. With $\boldsymbol{\theta}_2$, and hence σ_δ^2 and $\boldsymbol{\pi}$, known the transformation from $\boldsymbol{\theta}_1^*$ to $\boldsymbol{\theta}_1$ is one to one. Hence as long as the estimates fall within the parameter space, these are the pseudo-maximum likelihood estimators.

When $\mathbf{Z} \neq \mathbf{R}$, the fitting for the variance components with this approach is less straightforward since $\boldsymbol{\Psi}$ is no longer of the form $\mathbf{Z}\boldsymbol{\Omega}^*\mathbf{Z} + \sigma_\epsilon^2\mathbf{I}$. One approach is to fit the mixed model leaving $\boldsymbol{\Psi}$ unstructured and then use a linear least squares approach based on treating $\hat{\boldsymbol{\Psi}} - \mathbf{R}\hat{\boldsymbol{\Omega}}_{\tau C}\mathbf{R}' = \hat{\boldsymbol{\Psi}} - \hat{\gamma}^2\hat{\boldsymbol{\Sigma}}_D(\mathbf{I} - \hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D)$ as if it were unbiased for $\mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma_\epsilon^2\mathbf{I}$. That is, let $\mathbf{V} = vec(\hat{\boldsymbol{\Psi}} - \hat{\gamma}^2\hat{\boldsymbol{\Sigma}}_D(\mathbf{I} - \hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D))$ and write $vec(\mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma_\epsilon^2\mathbf{I}) = \mathbf{B}\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}' = (\boldsymbol{\omega}', \sigma_\epsilon^2)$ and $\mathbf{B} = [(\mathbf{Z} \otimes \mathbf{Z})\mathcal{D}, vech(\mathbf{I})]$ where $\mathcal{D}$ is a matrix of 0's and 1's for which $vec(\mathbf{A}) = \mathcal{D}vech(\mathbf{A})$. Using least squares the variance components are estimated via

$$\hat{\boldsymbol{\sigma}} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}\mathbf{B}'\mathbf{V}. \tag{14}$$

Some type of weighting could also be used.

For the case with $\mathbf{Z} = \mathbf{R}$ and constant $\mathbf{X}_i$, TBD used a pseudo-approach (referred to here as *pseudo-method A* or PMA), by fitting $\mathbf{Y}_i = \boldsymbol{\mu}^* + \boldsymbol{\Gamma}^*\mathbf{W}_i + \mathbf{Z}\boldsymbol{\nu}_i^* + \boldsymbol{\epsilon}_i$ where $\boldsymbol{\Gamma}^*$ is a $t \times t$ matrix and $\boldsymbol{\nu}_i^*$ has covariance $\boldsymbol{\Omega}^* = \boldsymbol{\Omega} + \boldsymbol{\Omega}_\tau$, leading to $\hat{\boldsymbol{\Gamma}}^*_A$, $\hat{\boldsymbol{\Omega}}^*_A$, and $\hat{\sigma}^2_{\epsilon A}$. From (5), $\boldsymbol{\Gamma}^* = \gamma\boldsymbol{\Sigma}_{\mathbf{D}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}$, suggesting $\hat{\gamma}_A = \mathbf{1}'\hat{\boldsymbol{\Gamma}}^*\mathbf{1}/\mathbf{1}'\hat{\boldsymbol{\Sigma}}_D\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{1}$ while (11) suggests $\hat{\boldsymbol{\Omega}}_A = \hat{\boldsymbol{\Omega}}^*_A - \hat{\boldsymbol{\Omega}}_{\tau A}$, where $\hat{\boldsymbol{\Omega}}_{\tau A} = \hat{\gamma}_A^2(\hat{\boldsymbol{\Omega}}_D - \hat{\boldsymbol{\Omega}}_D\mathbf{R}'\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{R}\hat{\boldsymbol{\Omega}}_D)$.

One advantage of this approach is that $\hat{\boldsymbol{\Gamma}}^*$ can be used to explore other structures if $\boldsymbol{\Gamma}\mathbf{D}_i$ replaces $\gamma\mathbf{D}_i$, where $\boldsymbol{\Gamma}$ is a $t \times t$ matrix. Another advantage is that under normality $(\hat{\boldsymbol{\mu}}^*, \hat{\boldsymbol{\Gamma}}^*, \hat{\boldsymbol{\Omega}}^*)$ is independent of $(\hat{\boldsymbol{\Sigma}}_D, \hat{\boldsymbol{\Sigma}}_W)$, which simplifies calculation of some asymptotic variances.

When $\mathbf{X}_i = \mathbf{X}$ and $\mathbf{A}_i = \mathbf{A}$, the method of moments approach in Section 4.2 of Fuller (1987), leading to nonlinear least squares, can be used. Let

$$\boldsymbol{\chi}_i = \begin{bmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_\chi = \begin{bmatrix} \mathbf{Z\Omega Z}' + \sigma_\epsilon^2 \mathbf{I} + \gamma^2 \mathbf{R\Omega}_D \mathbf{R}' & \gamma \mathbf{R\Omega}_D \mathbf{R}' \\ \gamma \mathbf{R\Omega}_D \mathbf{R}' & \mathbf{R\Omega}_D \mathbf{R}' + \sigma_\delta^2 \mathbf{I} \end{bmatrix}. \tag{15}$$

If $\mathbf{S}$ denotes the sample variance-covariance matrix of the $\boldsymbol{\chi}_i$, then $E(vech(\mathbf{S})) = vech(\boldsymbol{\Sigma}_\chi) = \mathbf{m}(\boldsymbol{\theta})$ where $\mathbf{m}(\boldsymbol{\theta})$ is a vector whose components are functions (some of them nonlinear) of $\boldsymbol{\theta}$. Fuller provides details of a generalized weighted nonlinear least squares approach with accompanying asymptotic theory, but this approach is not pursued further here due to the complexity of implementing the approach.

3.2. Asymptotic properties

The following lemma provides the information matrix under normality, which is used for both calculation of the full or pseudo-MLE and specification of their asymptotic covariance matrix.

Lemma 1. (Magnus and Neudecker (1988, p.325)). *Let $\boldsymbol{\chi}_i$, $i = 1, \ldots, n$, be independent $N(\boldsymbol{\mu}_{\chi_i}(\boldsymbol{\theta}), \Sigma_\chi(\boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is a $k \times 1$ vector of parameters. Then the information matrix for $\boldsymbol{\theta}$ is*

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_i \Big[\Big(\frac{\partial \boldsymbol{\mu}_{\chi_i}}{\partial \boldsymbol{\theta}}\Big)' \boldsymbol{\Sigma}_\chi{}^{-1} \Big(\frac{\partial \boldsymbol{\mu}_{\chi_i}}{\partial \boldsymbol{\theta}}\Big)\Big] + \frac{1}{2}\Big(\frac{\partial vec(\boldsymbol{\Sigma}_\chi)}{\partial \boldsymbol{\theta}}\Big)' \Big(\boldsymbol{\Sigma}_\chi{}^{-1} \otimes \boldsymbol{\Sigma}_\chi{}^{-1}\Big) \Big(\frac{\partial vec(\boldsymbol{\Sigma}_\chi)}{\partial \boldsymbol{\theta}}\Big).$$

For the current setting, $\boldsymbol{\chi}_i$ and $\boldsymbol{\Sigma}_\chi$ are as given in (15), $\boldsymbol{\mu}'_{\chi_i} = [(\mathbf{X}_i \boldsymbol{\beta} + \gamma \mathbf{A}_i \boldsymbol{\alpha})', (\mathbf{A}_i \boldsymbol{\alpha})']$, and the necessary derivatives are given in the Appendix. With $\boldsymbol{\theta}' = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2']$, the information matrix is partitioned as

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{11} \; \mathbf{I}_{12} \\ \mathbf{I}_{12}' \; \mathbf{I}_{22} \end{bmatrix}. \tag{16}$$

Computing the pseudo-MLE involves only the use of $\mathbf{I}_{11}$.

A careful statement of asymptotic properties requires some assumptions about the sequences of $\mathbf{X}_i$'s and $\mathbf{A}_i$'s. *It is assumed that these conditions are met in the following two propositions.* The needed conditions are met for instance, if the $(\mathbf{X_i}, \mathbf{A_i})$ are i.i.d. or, with fixed $\mathbf{X}$ and $\mathbf{A}$, if the asymptotics are viewed in the context of replicating the current design. Let $\boldsymbol{\Sigma}_{22}$ denote the asymptotic variance of $\hat{\boldsymbol{\theta}}_2$, the estimator of $\boldsymbol{\theta}_2$ from the $\mathbf{W}$ data only. It is also assumed that $\hat{\boldsymbol{\theta}}_2$ is obtained from one of the usual methods of estimation for linear mixed models and so is asymptotically normal. Let $\sim AN$ denote "is distributed asymptotically normal" and $\mathbf{A}()$ refer to asymptotic covariance matrix.

Proposition 2. *Under* (1), (2) *and* (7) *and multivariate normality,* $\hat{\boldsymbol{\theta}}_{1ML}$ *and* $\hat{\boldsymbol{\theta}}_{1PML}$ *are consistent with* $\hat{\boldsymbol{\theta}}_{1ML} \sim AN(\boldsymbol{\theta}_1, \mathbf{A}(\hat{\boldsymbol{\theta}}_{1ML}))$ *and* $\hat{\boldsymbol{\theta}}_{1PML} \sim AN(\boldsymbol{\theta}_1, \mathbf{A}(\hat{\boldsymbol{\theta}}_{1PML}))$. *Here*

$$\mathbf{A}(\hat{\boldsymbol{\theta}}_{1ML}) = \mathbf{I}_{11}^{-1} + \mathbf{I}_{11}^{-1}\mathbf{I}_{12}(\mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12})^{-1}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}, \tag{17}$$

and

$$\mathbf{A}(\hat{\boldsymbol{\theta}}_{1PML}) = \mathbf{I}_{11}^{-1} + \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\Sigma}_{22}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}. \tag{18}$$

Proposition 3. *Under* (1), (2) *and* (7) *and with* $\mathbf{Z} = \mathbf{R}$, $\hat{\boldsymbol{\theta}}_{1C}$ *(the RC estimator of* $\boldsymbol{\theta}_1$*) is consistent and* $\hat{\boldsymbol{\theta}}_{1C} \sim AN(\boldsymbol{\theta}_1, \mathbf{A}(\hat{\boldsymbol{\theta}}_{1C}))$ *with* $\mathbf{A}(\hat{\boldsymbol{\theta}}_{1C}) = \boldsymbol{\Sigma}_{1|2} + \mathbf{P}\boldsymbol{\Sigma}_{22}\mathbf{P}'$, *where* $\boldsymbol{\Sigma}_{1|2}$ *is the asymptotic covariance of* $\hat{\boldsymbol{\theta}}_{1C}$ *if* $\boldsymbol{\theta}_2$ *were known and* $\mathbf{P}$ *is defined at* (24).

With $\mathbf{X}_i = \mathbf{X}$ and $\mathbf{A}_i = \mathbf{A}$, Proposition 2 follows from Parke (1986), while the discussion in Lehmann (1983, Section 6.6) applies for more general settings. Since $\hat{\boldsymbol{\theta}}_2$ is obtained by fitting a mixed model to the W data, under normality, $\boldsymbol{\Sigma}_{22}$ is available from standard mixed model theory; see for example Jennrich and Schluchter (1986). Notice that $\mathbf{I}_{11}^{-1}$ is the asymptotic covariance of the MLE of $\boldsymbol{\theta}_1$ under normality if $\boldsymbol{\theta}_2$ were known. The proof of Proposition 3 is given in the Appendix based on an estimating equation approach. The robust asymptotic covariance matrix is reduced there to a highly simplified form, useful for both analytical and computational purposes. We also sketch a proof of the extension of Proposition 3 to the case where $\mathbf{Z} \neq \mathbf{R}$ and the variance parameters are estimated via (14).

4. Example

We consider the beta-carotene example introduced in Section 1, where $\mathbf{Y}_i$ is a 6×1 vector of serum beta-carotene measurements, $\mathbf{D}_i$ is a 6×1 vector of true diet intakes of beta-carotene and $\mathbf{W}_i$ is a 6×1 vector of measured intakes based on a food frequency questionnaire. Our analysis assumes $\mathbf{X}_i = \mathbf{A}_i = \mathbf{Z} = \mathbf{R}$, where

$$\mathbf{R}' = \begin{bmatrix} 1\,1\,1\,1\,1\,1 \\ 0\,1\,2\,3\,4\,5 \end{bmatrix}.$$

Model (7) becomes $\mathbf{D}_i = \mathbf{R}\boldsymbol{\alpha} + \mathbf{R}\boldsymbol{\phi}_i$, implying a linear trend in true diet values over time, with the line being random over subjects. With $\boldsymbol{\nu}_i' = (\boldsymbol{\nu}_{i1}, \boldsymbol{\nu}_{i2})$ and $\mathbf{Z} = \mathbf{R}$, (1) becomes $Y_{it}|D_{it} = \boldsymbol{\beta}_0 + \boldsymbol{\nu}_{i1} + (\boldsymbol{\beta}_1 + \boldsymbol{\nu}_{i2})(t-1) + \gamma D_{it} + \epsilon_{it}$. This allows a random trend over time in serum level after conditioning on dietary intake, with $\mathbf{\Omega}$ containing the variances and covariance for the random intercept and slope over subjects. This could capture, for example, changes in the effect of diet on serum level as an individual ages. The validity of (7) for this problem is discussed in TBD.

Fitting the longitudinal model $\mathbf{W}_i = \mathbf{A}\boldsymbol{\alpha} + \mathbf{R}\boldsymbol{\phi}_i + \boldsymbol{\delta}_i$ using maximum likelihood yields $\hat{\boldsymbol{\alpha}}' = (1.2719, .0143)$, $\hat{\sigma}^2_\delta = .119$, and

$$\hat{\mathbf{\Omega}}_D = \begin{bmatrix} .217 & -.011 \\ -.011 & .0044 \end{bmatrix}.$$

Since $\mathbf{Z} = \mathbf{R}$, by Proposition 1 the pseudo-MLEs are equivalent to the regression calibration estimators. The approximate variance matrix was estimated under normality, using (18) with $\mathbf{I}(\hat{\boldsymbol{\theta}})$ in place of $\mathbf{I}(\boldsymbol{\theta})$, and with $\mathbf{\Sigma}_{22}$ replaced by $\hat{\mathbf{\Sigma}}_{22}$, which is the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}_2$ obtained from the longitudinal fit using the $\mathbf{W}_i$. The pseudo method A estimates and standard errors were obtained as described in Section 3 and TBD. The estimates here differ slightly from those in TBD since REML estimates, rather than ML estimates, were used there at each stage. The naive results are obtained by fitting (1) with $\mathbf{W}_i$ in place of $\mathbf{D}_i$. Table 1 displays estimates with estimated standard errors.

Table 1. Estimates and estimated standard errors (in parentheses) for the beta-carotene data.

Parameter	PML/RC	PMA	Naive
$\widehat{\beta}_0$	4.73 (.11)	4.66 (.145)	5.10 (.062)
$\widehat{\beta}_1$	.0049 (.007)	.0113 (.022)	.0091 (.007)
$\widehat{\gamma}$	.454 (.093)	.491 (.106)	.156 (.030)
$\widehat{\omega}_{11}$	.29 (.040)	.29 (.039)	.32 (.042)
$\widehat{\omega}_{12}$	−.0053 (.0046)	−.0059 (.0045)	−.0072 (.0046)
$\widehat{\omega}_{22}$	.0018 (.00096)	.0019 (.0009)	.0024 (.00091)
$\widehat{\sigma}^2_\varepsilon$	.097 (.0055)	.091 (.0052)	.095 (.0054)

The most dramatic effect of the correction for measurement error here is on the estimate of γ which changes from .156 in the naive analysis to .454 when we correct for measurement error. The corrected estimates of ω_{11} and ω_{22} are smaller than the naive estimates (reductions of 9.4% and 25% respectively), indicative of the fact that naive estimation of the variance components tend to produce overestimates.

5. Efficiency Calculations and Simulations

We first examine the efficiency of the PML and PMA estimators for γ, ω_{11}, ω_{22} and σ^2_ϵ under normality. Efficiency is measured by the ratio of the asymptotic variance of the estimator to the asymptotic variance of the MLE. Because the design matrices $\mathbf{X}$ and $\mathbf{A}$ are constant over i, this is the same as the ratio of approximate variances at any n. Based on the example the case with $\mathbf{X} = \mathbf{A} = \mathbf{Z} = \mathbf{R}$, as described in the preceding section, is used. Throughout $\sigma_\epsilon^2 = 0.094$, $\boldsymbol{\alpha}' = (1.25, .012)$, $\boldsymbol{\beta}' = (4.64, -0.007)$,

$$\boldsymbol{\Omega}_D = \begin{bmatrix} .247 & -.0158 \\ -.0158 & .0046 \end{bmatrix} \quad \text{and } \boldsymbol{\Omega} = \begin{bmatrix} .324 & -.01 \\ -.01 & .0021 \end{bmatrix}.$$

Either γ or σ_δ^2 is then varied, with the other held fixed. When γ varies σ_δ^2, is held fixed at .118, while when σ_δ^2 varies, γ is held fixed at .49. The results are presented in Figures 1 and 2. The PML estimators are highly efficient across all parameterizations and for each of the estimators. It can be argued that the estimator of σ_ϵ^2 is fully efficient. In this setting, little is sacrificed by the use of the PML/RC estimators rather than the ML estimators. The PMA estimators are in general rather inefficient compared to the PML/RC estimators. The loss of efficiency, especially with respect to estimation of γ, is not surprising since, as described in Section 3.1, these estimators fit a $\boldsymbol{\Gamma}^*\mathbf{W_i}$ term in trying to allow for a more general model than $\gamma\mathbf{D}$ for the effect of $\mathbf{D}$.

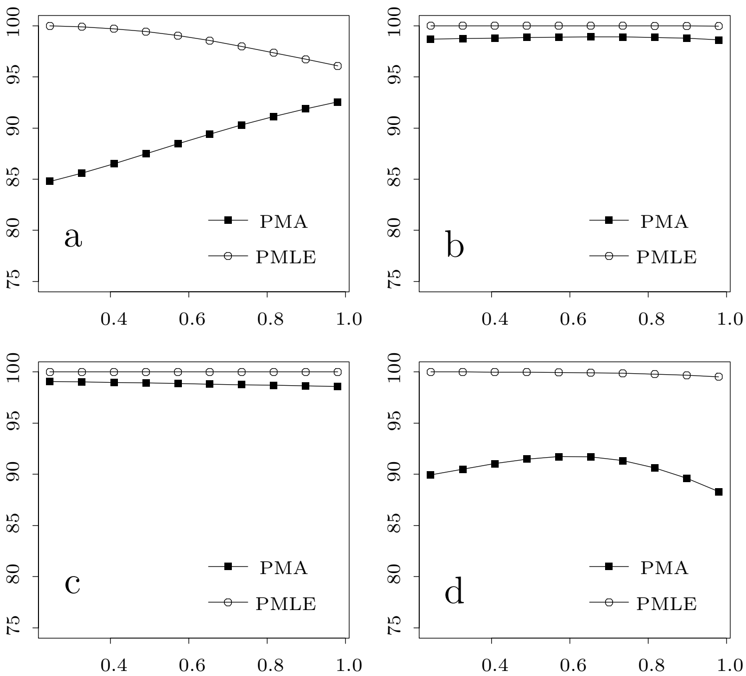

Figure 1. Asymptotic relative efficiency of PML and PMA estimators as a function of γ with $\sigma_\delta^2 = .118$. (a) $\hat{\gamma}$, (b) $\hat{\omega}_{11}$, (c) $\hat{\sigma}_\epsilon^2$, (d) $\hat{\omega}_{22}$.

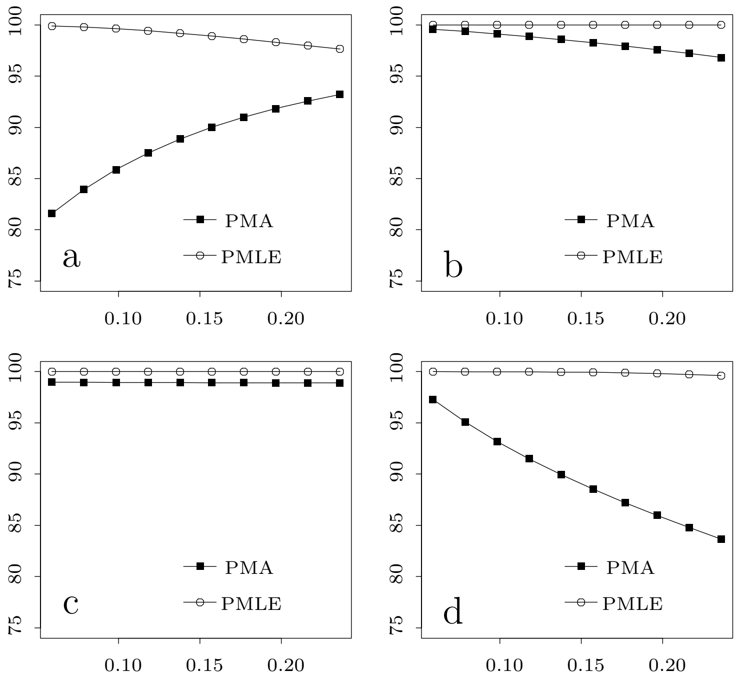

Figure 2. Asymptotic relative efficiency of PML and PMA estimators as a function of σ_δ^2, with $\gamma = .49$. (a) $\hat{\gamma}$, (b) $\hat{\omega}_{11}$, (c) $\hat{\sigma}_\epsilon^2$, (d) $\hat{\omega}_{22}$.

Simulations were run under normality, for sample sizes 100, 200 and 1000, with $\gamma = 0.49$ and $\sigma_\delta^2 = 0.118$. The remaining values were as used in the efficiency calculations. Table 2 provides standard errors for the PML and PMA estimators of γ, ω_{11}, ω_{12}, ω_{22}, and σ_ϵ^2. These results serve as a double check on the asymptotic variance expressions and also as a check as to what is lost by use of the asymptotic expressions at "small" sample sizes. At $n = 1000$ there is very close agreement between the theoretically calculated and simulated standard errors, and even at $n = 100$ the simulated standard errors of the PML estimators are still quite close to the asymptotic values.

For $n = 100$, some limited simulations were also run without normality to evaluate the performance of the RC estimators which, based on theory, are consistent. First a set of standard variables with mean 0 and variance 1 are created using a normal, double exponential or squared normal; the latter using $(Z^2 - 1)/2^{1/2}$ where Z has a standard normal distribution. These are then linearly transformed to create univariate or bivariate random quantities with the desired mean and covariance structure. Table 3 displays the simulated biases and standard errors. At this sample size, modest for epidemiologic studies but maybe large in other contexts, the biases are negligible. The standard errors for estimates of variance parameters can change substantially under non-normality. This points out the need to utilize the asymptotic covariance matrix in Proposition 3 rather than the normal-based one in Proposition 2 for some situations. We leave a fuller evaluation of the small sample performance of related confidence intervals and tests for future work.

Table 2. Asymptotic ($n = \infty$) and simulated standard errors under $\gamma = 0.49$ and $\sigma_\delta^2 = 0.118$, under normality. Based on 500 simulations.

Method	n	$\hat{\gamma}$	$\hat{\omega}_{11}$	$\hat{\omega}_{12}$	$\hat{\omega}_{22}$	$\hat{\sigma}_\epsilon^2$
MLE	∞	1.1529	0.5426	0.0622	0.0124	0.0665
PMA	∞	1.3178	0.5490	0.0637	0.0136	0.0672
	100	1.3893	0.5572	0.0655	0.0132	0.0683
	200	1.3413	0.5432	0.06245	0.0126	0.0651
	1000	1.3143	0.5426	0.0637	0.0136	0.0670
PMLE/RC	∞	1.1600	0.5427	0.0622	0.0124	0.0665
	100	1.1821	0.5372	0.0626	0.0125	0.0663
	200	1.1161	0.5360	0.0609	0.0121	0.0637
	1000	1.1634	0.5411	0.06216	0.0124	0.0669

Table 3. Simulated biases and standard errors with $\gamma = 0.49$, $\omega_{11} = .324$, $\omega_{12} = -.01$, $\omega_{22} = .0021$, $\sigma_\epsilon^2 = .094$, $\sigma_\delta^2 = 0.118$ and $n = 100$. Based on 500 simulations.

Model		$\hat{\gamma}$	$\hat{\omega}_{11}$	$\hat{\omega}_{12}$	$\hat{\omega}_{22}$	$\hat{\sigma}_\epsilon^2$
Normal	Bias	0.0307	0.0112	$-.0009$	0.0003	$-.0015$
Squared Normal	Bias	0.0095	$-.0277$	0.0014	0.0002	0.0026
Double Exp.	Bias	$-.0096$	0.0074	$-.0001$	0.0004	0.0033
Normal	SE	1.2024	0.4561	0.0510	0.0120	0.0825
Squared Normal	SE	1.1835	1.0031	0.0655	0.0169	0.1590
Double Exp.	SE	1.1163	0.8039	0.0708	0.0137	0.0845

6. Extensions

6.1. Separate estimation of the measurement error variance

To this point (7) has been used to ensure that $\mathbf{\Sigma}_D$ is identifiable without additional data. Proceeding under (7), when it does not hold, actually fits a model for $\mathbf{Y}_i|\boldsymbol{\mu}_{Di}$ rather than for $\mathbf{Y}_i|\mathbf{D}_i$. If there is replicate data to provide an estimate $\hat{\sigma}_\delta^2$ of the measurement error variance, then (7) can be generalized to allow Cov $(\boldsymbol{\xi}_i) = \mathbf{\Sigma}_D$ to be of full rank. This could be unstructured or take some particular form, such as $\mathbf{\Sigma}_D = \mathbf{R}\mathbf{\Omega}_D\mathbf{R}' + \sigma_e^2\mathbf{I}$. Now $\mathbf{\Sigma}_D$ can be estimated from the W data. As before $\hat{\mathbf{\Sigma}}_W = \hat{\mathbf{\Sigma}}_D + \hat{\sigma}_\delta^2\mathbf{I}$, while $\boldsymbol{\theta}_2' = (\boldsymbol{\alpha}', \boldsymbol{\sigma}_D', \sigma_\delta^2)$, where $\boldsymbol{\sigma}_D$ contains the unique parameters in $\mathbf{\Sigma}_D$.

For likelihood approaches the conditional distribution of $\hat{\sigma}^2_\delta$ given the collection of $(\mathbf{Y}_i, \mathbf{W}_i)$ values must be specified. It would usually be assumed to depend on parameters only through σ^2_δ. A typical assumption, which results under normality and the use of replicate W values at some time points on some individuals, is that $d\hat{\sigma}^2_\delta/\sigma^2_\delta$ is distributed chi-square with d degrees of freedom independent of the $(\mathbf{Y}_i, \mathbf{W}_i)$. Under normality, the information matrix is now $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}_1(\boldsymbol{\theta}) + \mathbf{I}_2(\sigma^2_\delta)$, where $\mathbf{I}_1(\boldsymbol{\theta})$ is calculated using Lemma 1 with $\boldsymbol{\Sigma}_\chi$ as in (15) but now with $\boldsymbol{\Sigma}_D$ replacing $\mathbf{R\Omega_D R}'$, while $\mathbf{I}_2(\sigma^2_\delta)$ has an entry only in the diagonal position corresponding to σ^2_δ, that being the information for σ^2_δ contained in $\hat{\sigma}^2_\delta$. The PMLE is still based on (13), with $\hat{\boldsymbol{\Sigma}}_D$ and $\hat{\boldsymbol{\Sigma}}_W$ appropriately modified but, as was the case earlier, it usually cannot be computed using standard mixed model software.

Regression calibration can still be used as in Section 3 for estimation of $\boldsymbol{\beta}$ and γ, but estimation of the variance parameters is more complicated. As before $\boldsymbol{\Psi} = \mathbf{Z\Omega Z}' + \sigma^2_\epsilon \mathbf{I} + \gamma^2 \boldsymbol{\Sigma}_D(\mathbf{I} - \boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_D)$. When $\boldsymbol{\Sigma}_D$ is *not* of the form $\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}'$, $\boldsymbol{\Psi}$ is *not* of the form $\mathbf{Z}\boldsymbol{\Omega}^*\mathbf{Z}' + \sigma^2_\epsilon\mathbf{I}$, even for the case with $Z = R$, nor does it take on any of the typical structures usually allowed by mixed model programs. One suggestion is to regress $\mathbf{Y}_i$ on $\mathbf{X}_i$ and $\hat{\mathbf{Q}}_i$ with an unstructured covariance matrix $\boldsymbol{\Psi}$, estimated by $\hat{\boldsymbol{\Psi}}$, and then use least squares as in (14). This approach usually yields consistent and asymptotically normal estimators without normality assumptions, as is sketched in the appendix.

In related work, Takeuchi and Ware (1990) and Takeuchi (1992) consider a model which in our notation has $E(\mathbf{Y}_i|\mathbf{D}_i) = \gamma\mathbf{D}_i$, $\mathrm{Cov}\,(\mathbf{Y}_i|\mathbf{D}_i) = \boldsymbol{\Sigma}$ (either unstructured or compound symmetric), and $\mathrm{Cov}\,(\boldsymbol{\delta}_i) = \boldsymbol{\Sigma}_\delta$ is known. The compound symmetry case corresponds to our model with $\mathbf{Z} = \mathbf{1}$. They addressed estimation of γ but not of the variance components.

6.2. More general structure for the D effects

Equation (1) can be generalized to $\mathbf{Y}_i \mid \mathbf{D}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{D}_i + \mathbf{Z}\boldsymbol{\nu}_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\Gamma}$ is now a $t \times t$ matrix of parameters, perhaps with some simplifying structure. To this point $\boldsymbol{\Gamma} = \gamma\mathbf{I}$ has been assumed. Retaining the additive measurement error assumption in (2), $\mathbf{Y}_i \mid \mathbf{W}_i = \boldsymbol{\mu}^*_i + \boldsymbol{\Gamma}^*\mathbf{W}_i + \boldsymbol{\eta}^*_i$, where $\boldsymbol{\mu}^*_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Gamma}(\boldsymbol{\mu}_{Di} - \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1}\boldsymbol{\mu}_{Di})$, $\boldsymbol{\Gamma}^* = \boldsymbol{\Gamma}\boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1}$, and $\boldsymbol{\eta}^*_i$ has covariance $\boldsymbol{\Psi} = \mathbf{Z\Omega Z}' + \sigma^2_\epsilon\mathbf{I} + \boldsymbol{\Gamma}\boldsymbol{\Sigma}_D\boldsymbol{\Gamma}' - \boldsymbol{\Gamma}\boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_D\boldsymbol{\Gamma}'$.

If there are no additional data and we incorporate (7), $\boldsymbol{\Gamma}^*\mathbf{W}_i$ can be rewritten as $\boldsymbol{\Gamma}_2\mathbf{B}_i$ where $\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}\mathbf{R}$ and $\mathbf{B}_i = \boldsymbol{\Omega}_D\mathbf{R}'(\boldsymbol{\Sigma}_D + \sigma^2_\delta\mathbf{I})^{-1}\mathbf{W}_i$. Since $\boldsymbol{\Gamma}$ enters the conditional mean and covariance of $\mathbf{Y} \mid \mathbf{W}$ only via $\boldsymbol{\Gamma}_2$, it is the unique components of $\boldsymbol{\Gamma}_2$, and functions of them, which are identifiable. If $\boldsymbol{\Gamma}$ is an arbitrary t by t matrix, it is not identifiable since $\boldsymbol{\Sigma}_D = \mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}'$ is not of full rank. However, with some structure to $\mathbf{\Gamma}$, it's distinct components may be identifiable, as happens when $\mathbf{\Gamma}$ is diagonal.

If there are additional data for estimation of σ_δ^2, then a nonsingular $\mathbf{\Sigma}_D$ can be allowed and a general $\mathbf{\Gamma}$, as well as the variance components, can be estimated. This is closely related to the multivariate "errors in variables' problem (see Fuller (1987, Chapter 4) and Gleser (1992)) which allows general $\mathbf{\Gamma}$, no structure on $\mathbf{\Sigma} = \text{Cov}\,(\mathbf{Y}|\mathbf{D})$ and $\text{Cov}\,(\boldsymbol{\delta}_i) = \mathbf{\Sigma}_\delta$, assumed known or estimated.

6.3. Alternate measurement error models with validation data

In many cases, the measurement error model (2) needs to be modified to allow the mean of W to depend on D. Such models can be handled as long as there is some validation data (called internal or external, depending on whether the units are part of the main study or not) with both W and D values. This allows estimation of the measurement error parameters and the restricted model in (7) can be dropped. The likelihood based methods under normality are readily extended in principle, with the details depending on the nature of the additional data. As before though, computational implementation may be difficult and the distributional assumptions questionable. For some cases, the regression calibration approach is easily extended to handle these more complex models.

To illustrate, consider a common linear measurement error model at each time point (e.g., Carroll, Stefanski and Ruppert (1995, p.8) and Buonaccorsi (1989, 1990)) for which $\mathbf{W}_i = \lambda_0\mathbf{1} + \lambda_1\mathbf{D}_i + \boldsymbol{\delta}_i$, where λ_0 and λ_1 are scalars. This leads to $E(\mathbf{W}_i) = \lambda_0\mathbf{1} + \lambda_1\boldsymbol{\mu}_{Di}$, $\mathbf{\Sigma}_W = \lambda_1^2\mathbf{\Sigma}_D + \sigma_\delta^2\mathbf{I}$, and $\mathbf{\Sigma}_{YW} = \gamma\lambda_1\mathbf{\Sigma}_{YD}$. Under normality, $\mathbf{Y}_i|\mathbf{W}_i = \mathbf{X}_i\boldsymbol{\beta} + \gamma\mathbf{Q}_i + \boldsymbol{\eta}_i^*$ but now $\mathbf{Q}_i = \boldsymbol{\mu}_{Di} + \lambda_1\mathbf{\Sigma}_D\mathbf{\Sigma}_W^{-1}(\mathbf{W}_i - \lambda_0\mathbf{1} - \lambda_1\boldsymbol{\mu}_{Di})$ and $\boldsymbol{\eta}_i^*$ has variance $\mathbf{\Psi} = \mathbf{Z\Omega Z}' + \sigma_\epsilon^2\mathbf{I} + \gamma^2\mathbf{\Sigma}_D(\mathbf{I} - \lambda_1^2\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_D)$. In the regression calibration approach $\hat{\mathbf{Q}}_i$ is obtained based on estimated parameters from the validation data and we proceed as before. When the validation data is internal and a particular D_{it} is available, it makes sense to replace the tth element of $\hat{\mathbf{Q}}_i$ with it. As with our earlier results, this will usually give consistent estimators without normality. The development of the asymptotic covariance matrix depends on the type of validation data. For external data, asymptotic properties would be established in a manner similar to that used in the Appendix. With internal data, some of the main study data is also used in estimating the measurement error parameters and this needs to be accounted for; similar to what is done in Buonaccorsi (1990) in a related problem. This setting needs further study.

6.4. Extensions involving Z

Two important and related assumptions that have been made are that i) $\mathbf{Z}_i$ is constant over i, and ii) that no components of $\mathbf{Z}$ are measured with error.

Allowing a changing $\mathbf{Z}_i$ is readily absorbed into the likelihood approaches and into the regression calibration approach for estimating $\boldsymbol{\beta}$ and γ. The approach to estimating the variance parameters after fitting the model using $\hat{\mathbf{Q}}_i$ assumed that $\boldsymbol{\Psi}_i = \text{Cov}\,(\mathbf{Y}_i|\mathbf{W}_i, \mathbf{X}_i)$ was constant in i, so these methods need modification for varying $\mathbf{Z}_i$. The presence of measurement error in some parts of $\mathbf{Z}_i$ raises new methodological issues due to the product of the random effects and the measurement errors. These two problems arise together with random coefficients, for which $\mathbf{Y}_i|\mathbf{D}_i = \boldsymbol{\beta}\mathbf{1} + \gamma\mathbf{D}_i + \mathbf{Z}_i\boldsymbol{\nu}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{Z}_i = [\mathbf{1}, \mathbf{D}_i]$, so the mismeasured covariates enter into $\mathbf{Z}_i$. These problems are currently under investigation.

Acknowledgements

This research has been supported by NIH grants CA50597. We are grateful to Chii-Dean Lin for his comments and to Naisyin Wang for helpful discussion and for sharing preprints of her work.

Appendix

Derivatives for the Information Matrix

Let $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \gamma, \boldsymbol{\omega}', \sigma^2_\epsilon, \boldsymbol{\alpha}', \boldsymbol{\omega}'_D, \sigma^2_\delta)$. Then,

$$\frac{\partial \boldsymbol{\mu}_{\chi i}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{X}_i & \mathbf{A}_i\boldsymbol{\alpha} & 0 & 0 & \gamma\mathbf{A}_i & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{A}_i & 0 & 0 \end{bmatrix}.$$

Let $\boldsymbol{\Delta} = \mathbf{K} \otimes I$, where $\mathbf{K}$ is the commutation matrix, defined such that $vec(\mathbf{A}') = \mathbf{K}vec(\mathbf{A})$. Also, let $\mathcal{D}$ and $\mathcal{D}_D$ be the matrices of 0's and 1's such that $\tilde{\boldsymbol{\omega}} = vec(\boldsymbol{\Omega}) = \mathcal{D}vech(\boldsymbol{\Omega}) = \boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}_D = vec(\boldsymbol{\Omega}_D) = \mathcal{D}_D vech(\boldsymbol{\Omega}_D) = \boldsymbol{\omega}_D$, and define $\mathbf{S}_R = (\mathbf{R} \otimes \mathbf{R})\mathcal{D}_D$, $\mathbf{S}_Z = (\mathbf{Z} \otimes \mathbf{Z})\mathcal{D}$, and $\mathbf{i} = vec(\mathbf{I})$. Since $vec(\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}') = (\mathbf{R} \otimes \mathbf{R})\tilde{\boldsymbol{\omega}}_D$ and $vec(\mathbf{Z}\boldsymbol{\Omega}_D\mathbf{Z}') = (\mathbf{Z} \otimes \mathbf{Z})\tilde{\boldsymbol{\omega}}$,

$$\begin{aligned} vec(\boldsymbol{\Sigma}_\chi) &= \begin{bmatrix} \boldsymbol{\Delta} & 0 \\ 0 & \boldsymbol{\Delta} \end{bmatrix} \begin{bmatrix} vec(\mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma^2_\epsilon\mathbf{I} + \gamma^2\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}') \\ vec(\gamma\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}') \\ vec(\gamma\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}') \\ vec(\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}' + \sigma^2_\delta\mathbf{I}) \end{bmatrix} \\ &= \boldsymbol{\Delta}^+ \begin{bmatrix} \mathbf{S}_Z\boldsymbol{\omega} + \mathbf{i}\sigma^2_\epsilon + \gamma^2\mathbf{S}_R\boldsymbol{\omega}_D \\ \gamma\mathbf{S}_R\boldsymbol{\omega}_D \\ \gamma\mathbf{S}_R\boldsymbol{\omega}_D \\ \mathbf{S}_R\boldsymbol{\omega}_D + \mathbf{i}\sigma^2_\delta \end{bmatrix} \end{aligned}$$

where $\boldsymbol{\Delta}^+ = BlockDiagonal(\boldsymbol{\Delta}, \boldsymbol{\Delta})$. Hence,

$$\frac{\partial vec(\boldsymbol{\Sigma}_{\chi i})}{\partial \boldsymbol{\theta}} = \boldsymbol{\Delta}^+ \begin{bmatrix} 0 & 2\gamma\mathbf{S}_R\boldsymbol{\omega}_D & \mathbf{S}_Z & \mathbf{i} & 0 & \gamma^2\mathbf{S}_R & 0 \\ 0 & \mathbf{S}_r\boldsymbol{\omega}_D & 0 & 0 & 0 & \gamma\mathbf{S}_R & 0 \\ 0 & \mathbf{S}_r\boldsymbol{\omega}_D & 0 & 0 & 0 & \gamma\mathbf{S}_R & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{S}_R & \mathbf{i} \end{bmatrix}.$$

Proof of Proposition 3

Recall that $\hat{\boldsymbol{\theta}}_2$ is fit with the W data only. Let $\hat{\boldsymbol{\theta}}^{*'}_{1C} = (\hat{\boldsymbol{\beta}}'_C, \hat{\gamma}_C, \hat{\boldsymbol{\omega}}^{*'}_C, \hat{\sigma}^2_{\epsilon C})$ denote the estimates obtained by fitting $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \hat{\mathbf{Q}}_i\gamma + \mathbf{Z}\boldsymbol{\tau}^*_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\tau}^*_i$ has mean $\mathbf{0}$ and covariance $\boldsymbol{\Omega}^*$. The vector $\hat{\boldsymbol{\omega}}^*_C$ estimates $vech(\boldsymbol{\Omega}^*)$. Simplifying the estimating equations in normal mixed models (e.g., Section 5.3 of Crowder and Hand (1990) or Jennrich and Schluchter (1986)), the estimate of $\boldsymbol{\theta}^*_1$ solves

$$\begin{bmatrix} \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{X}_i & \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\hat{\mathbf{Q}}_i \\ (\sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\hat{\mathbf{Q}}_i)' & \sum_i \hat{\mathbf{Q}}'_i\boldsymbol{\Psi}^{-1}\hat{\mathbf{Q}}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \gamma \end{bmatrix} - \begin{bmatrix} \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{Y}_i \\ \sum_i \hat{\mathbf{Q}}'_i\boldsymbol{\Psi}^{-1}\mathbf{Y}_i \end{bmatrix} = \mathbf{0}, \tag{19}$$

$$\mathbf{Z}'\boldsymbol{\Psi}^{-1}\sum_i(\mathbf{r}_i\mathbf{r}'_i - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\mathbf{Z} = 0 \text{ and } \sum_i \mathbf{r}'_i\boldsymbol{\Psi}^{-2}\mathbf{r}_i - ntr(\boldsymbol{\Psi}^{-1}) = \mathbf{0}, \tag{20}$$

where $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \hat{\mathbf{Q}}_i\gamma$. Vectorizing, this can be written as $\mathbf{S}_1(\boldsymbol{\theta}^*_1, \hat{\boldsymbol{\theta}}_2) = \mathbf{0}$, where

$$\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2) = \begin{bmatrix} \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{X}_i\boldsymbol{\beta} + \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{Q}_i\gamma - \sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{Y}_i \\ (\sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\mathbf{Q}_i)'\boldsymbol{\beta} + \sum_i \mathbf{Q}'_i\boldsymbol{\Psi}^{-1}\mathbf{Q}_i\gamma - \sum_i \mathbf{Q}'_i\boldsymbol{\Psi}^{-1}\mathbf{Y}_i \\ \mathcal{D}(\mathbf{M}\otimes\mathbf{M})vec(\sum_i(\boldsymbol{\rho}_i\boldsymbol{\rho}'_i)) - \mathcal{D}(\mathbf{Z}'\otimes\mathbf{Z}')vec(\boldsymbol{\Psi}^{-1}) \\ \sum_i \boldsymbol{\rho}'_i\boldsymbol{\Psi}^{-2}\boldsymbol{\rho}_i - ntr(\boldsymbol{\Psi}^{-1}) \end{bmatrix} \tag{21}$$

with $\mathbf{M} = \mathbf{Z}'\boldsymbol{\Psi}^{-1}$ and $\rho_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Q}_i\gamma$. Similarly $\hat{\boldsymbol{\theta}}_2$ is the solution to a set of equations $\mathbf{S}_2(\boldsymbol{\theta}_2) = \mathbf{0}$ so together $\hat{\boldsymbol{\theta}}_2$ and $\hat{\boldsymbol{\theta}}^*_{1C}$ solve $\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2) = \mathbf{0}$ and $\mathbf{S}_2(\boldsymbol{\theta}_2) = \mathbf{0}$. Note that $\mathbf{S}_1$ depends on the Y and W values while $\mathbf{S}_2$ depends just on the W values, but this has been suppressed in the notation.

Under (1), (2) and (7), $E(\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2)) = \mathbf{0}$. From the theory of estimating equations (see Appendix A.3 of Carroll, Stefanski and Ruppert (1995)) under suitable conditions, $\hat{\boldsymbol{\theta}}^*_{\mathbf{C}}$ is consistent for $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}^*_{\mathbf{C}}$ is $AN(\boldsymbol{\theta}^*, \mathbf{H}^{-1}\mathbf{C}\mathbf{H}^{-1'})$, where AN stands for asymptotically normal and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}'_{12} & \mathbf{C}_{22} \end{bmatrix}$$

with $\mathbf{H}_{11} = E(\partial\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}^*_1)$, $\mathbf{H}_{12} = E(\partial\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2)$, $\mathbf{H}_{21} = E(\partial\mathbf{S}_2(\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}^*_1) = \mathbf{0}$, $\mathbf{H}_{22} = E(\partial\mathbf{S}_2(\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2)$, $\mathbf{C}_{11} = E(\mathbf{S}_1\mathbf{S}'_1)$, $\mathbf{C}_{12} = E(\mathbf{S}_1\mathbf{S}'_2)$, and $\mathbf{C}_{22} = E(\mathbf{S}_2\mathbf{S}'_2)$. Since $\mathbf{S}_2$ does not involve $\boldsymbol{\theta}^*_1$, $\mathbf{H}_{21} = \mathbf{0}$. It is easy to show that $E(\mathbf{S}_1|\mathbf{W_1}, \ldots, \mathbf{W_n}) = \mathbf{0}$ and hence $\mathbf{C}_{12} = E(\mathbf{S}_1\mathbf{S}'_2) = E(E(\mathbf{S}_1\mathbf{S}'_2|\mathbf{W_1}, \ldots, \mathbf{W_n})) = E(E(\mathbf{S}_1|\mathbf{W_1}, \ldots, \mathbf{W_n})\mathbf{S}'_2) = \mathbf{0}$. Using partitioned matrix results,

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}^{-1}_{11} & -\mathbf{H}^{-1}_{11}\mathbf{H}_{12}\mathbf{H}^{-1}_{22} \\ \mathbf{0} & \mathbf{H}^{-1}_{22} \end{bmatrix}$$

and so

$$\mathbf{A}(\hat{\boldsymbol{\theta}}^*) = \mathbf{H}^{-1}\mathbf{C}\mathbf{H}^{-1'} = \begin{bmatrix} \boldsymbol{\Sigma}^*_{11} & \boldsymbol{\Sigma}^*_{12} \\ \boldsymbol{\Sigma}^{*'}_{12} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \tag{22}$$

where $\mathbf{A}(\hat{\boldsymbol{\theta}}^*_{\mathbf{1C}}) = \boldsymbol{\Sigma}^*_{11} = \mathbf{H}_{11}^{-1}\mathbf{C}_{11}\mathbf{H}_{11}^{-1'} + \mathbf{H}_{11}^{-1}\mathbf{H}_{12}\boldsymbol{\Sigma}_{22}\mathbf{H}'_{12}\mathbf{H}_{11}^{-1'}$, $\mathbf{A}(\hat{\boldsymbol{\theta}}_2) = \boldsymbol{\Sigma}_{22} = \mathbf{H}_{22}^{-1}\mathbf{C}_{22}\mathbf{H}_{22}^{-1'}$, and $\mathbf{A}(\hat{\boldsymbol{\theta}}^*_{\mathbf{1C}}, \hat{\boldsymbol{\theta}}_2) = \boldsymbol{\Sigma}^*_{12} = -\mathbf{H}_{11}^{-1}\mathbf{H}_{12}\boldsymbol{\Sigma}_{22}$, where $\mathbf{A}()$ denotes asymptotic covariance matrix. A similar estimating equation argument implies that $\boldsymbol{\Sigma}^*_{1|2} = \mathbf{H}_{11}^{-1}\mathbf{C}_{11}\mathbf{H}_{11}^{-1'}$ is the asymptotic covariance of $\hat{\boldsymbol{\theta}}^*_{\mathbf{1C}}$ if $\boldsymbol{\theta}_2$ were known, and so

$$\mathbf{A}(\hat{\boldsymbol{\theta}}^*_{\mathbf{1C}}) = \boldsymbol{\Sigma}^*_{1|2} + \mathbf{H}_{11}^{-1}\mathbf{H}_{12}\boldsymbol{\Sigma}_{22}\mathbf{H}'_{12}\mathbf{H}_{11}^{-1'}. \tag{23}$$

Let $\mathbf{E} = (\sigma^2_\delta \boldsymbol{\Omega}_D^{-1} + \mathbf{R}'\mathbf{R})^{-1}$ and let $\dot{\mathbf{E}}_j$ denote the derivative of $\mathbf{E}$ with respect to the jth element of $\boldsymbol{\omega}_D$, $\dot{\mathbf{E}}_\delta$ the derivative of $\mathbf{E}$ with respect to σ^2_δ. Applying 4.A.9 in Fuller (1987, p.390) twice, $\dot{\mathbf{E}}_j = -\sigma^2_\delta \mathbf{E}^{-1}\boldsymbol{\Omega}_D^{-1}\dot{\boldsymbol{\Omega}}_{Dj}\boldsymbol{\Omega}_D^{-1}\mathbf{E}^{-1}$ and $\dot{\mathbf{E}}_\delta = -\mathbf{E}^{-1}\boldsymbol{\Omega}_D^{-1}\mathbf{E}^{-1}$, where $\dot{\boldsymbol{\Omega}}_{Dj}$ has a 1 in positions containing $\boldsymbol{\omega}_{Dj}$ (the jth component of $\boldsymbol{\omega}_D$) and 0's elsewhere. We transform to $\hat{\boldsymbol{\theta}}'_{1C} = (\hat{\boldsymbol{\beta}}'_C, \hat{\gamma}_C, \hat{\boldsymbol{\omega}}'_C, \hat{\sigma}^2_{\epsilon C})$, where $\hat{\boldsymbol{\omega}}_C = vech(\hat{\boldsymbol{\Omega}}_C) = \hat{\boldsymbol{\omega}}^*_C - \hat{\gamma}^2_C\hat{\sigma}^2_\delta \mathbf{q}$, with $\mathbf{q} = vech((\hat{\sigma}^2_\delta\hat{\boldsymbol{\Omega}}_D^{-1} + \mathbf{R}'\mathbf{R})^{-1})$ and we have used the equality $\hat{\boldsymbol{\Omega}}_D - \hat{\boldsymbol{\Omega}}_D\mathbf{R}'\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{R}\hat{\boldsymbol{\Omega}}_D = (\hat{\boldsymbol{\Omega}}_D^{-1} + \hat{\sigma}_\delta^{-2}\mathbf{R}'\mathbf{R})^{-1}$. Applying the delta method, $\mathbf{A}(\hat{\boldsymbol{\theta}}) = \mathbf{G}\mathbf{A}(\hat{\boldsymbol{\theta}}^*)\mathbf{G}'$, where $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2]$, with

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -2\gamma\sigma^2_\delta\mathbf{q} & \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \text{ and } \mathbf{G}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\boldsymbol{\omega}} & \mathbf{G}_\delta \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}.$$

Here $\mathbf{I}_a$ denotes an $a \times a$ identity matrix, $k = p(p+1)/2$, $\mathbf{G}_{\boldsymbol{\omega}} = -\gamma^2\sigma^2_\delta[vech(\dot{\mathbf{E}}_1), \ldots, vech(\dot{\mathbf{E}}_{n_{\boldsymbol{\omega}}})]$ and $\mathbf{G}_\delta = -\gamma^2(\sigma^2_\delta\dot{\mathbf{E}}_\delta + \mathbf{q})$. Now using (22), (23) and the structure of $\mathbf{G}$ yields $\mathbf{A}(\hat{\boldsymbol{\theta}}_{\mathbf{1C}}) = \mathbf{G}_1\boldsymbol{\Sigma}^*_{1|2}\mathbf{G}'_1 + \mathbf{P}\boldsymbol{\Sigma}_{22}\mathbf{P}'$, where

$$\mathbf{P} = \mathbf{G}_1\mathbf{H}_{11}^{-1}\mathbf{H}_{12} - \mathbf{G}_2. \tag{24}$$

Since the $\mathbf{G}_1$ matrix is what would be used if transforming with known $\boldsymbol{\theta}_2$, $\mathbf{G}_1\boldsymbol{\Sigma}^*_{1|2}\mathbf{G}'_1$ must equal $\boldsymbol{\Sigma}_{1|2}$, the asymptotic covariance of $\hat{\boldsymbol{\theta}}_{\mathbf{1C}}$ if $\boldsymbol{\theta}_2$ is known. This yields Proposition 3.

Under normality, $\boldsymbol{\Sigma}_{22} = \mathbf{H}_{22}^{-1} = \mathbf{I}_{22}^{-1}$, $\boldsymbol{\Sigma}^*_{1|2} = \mathbf{H}_{11}^{-1}$ and $\boldsymbol{\Sigma}_{1|2} = \mathbf{I}_{11}^{-1}$ are available from normal linear mixed models theory; see for example equations (11)- (13) in Jennrich and Schluchter (1986).

A more explicit expression for $\mathbf{H}_{12}$ is still needed. Recall that $\mathbf{Q}_i$ and $\boldsymbol{\rho}_i$ depend on $\boldsymbol{\theta}_2$. Since $\Upsilon = \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1} = \mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}'(\mathbf{R}\boldsymbol{\Omega}_D\mathbf{R}' + \sigma^2_\delta\mathbf{I})^{-1} = \mathbf{R}\mathbf{E}\mathbf{R}'$, $\dot{\Upsilon}_j = \mathbf{R}\dot{\mathbf{E}}_j\mathbf{R}'$ is the derivative of Υ with respect to the jth element of $\boldsymbol{\omega}_D$, $\dot{\Upsilon}_\delta = \mathbf{R}\dot{\mathbf{E}}_\delta\mathbf{R}'$ is derivative of Υ with respect to σ^2_δ and $\dot{\mathbf{Q}}_i = \partial\mathbf{Q}_i/\partial\boldsymbol{\theta}_2 = [(\mathbf{I} - \Upsilon)\mathbf{A}_i, \dot{\Upsilon}_1(\mathbf{W}_i - \mathbf{A}_i\boldsymbol{\alpha}), \ldots, \dot{\Upsilon}_{n_\omega}(\mathbf{W}_i - \mathbf{A}_i\boldsymbol{\alpha}), \dot{\Upsilon}_\delta(\mathbf{W}_i - \mathbf{A}_i\boldsymbol{\alpha})]$. Differentiating and simplifying,

$$\mathbf{H}_{12} = E(\partial\mathbf{S}_1(\boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2) = E\begin{bmatrix} \gamma\sum_i \mathbf{X}'_i\boldsymbol{\Psi}^{-1}\dot{\mathbf{Q}}_i \\ \sum_i((\boldsymbol{\Psi}^{-1}\mathbf{X}_i\beta)' + 2\gamma\mathbf{Q}'_i\boldsymbol{\Psi}^{-1} - \mathbf{Y}'_i\boldsymbol{\Psi}^{-1})\dot{\mathbf{Q}}_i \\ \mathcal{D}(\mathbf{M}\otimes\mathbf{M})\sum_i\mathbf{U}_i) \\ -2\gamma\sum_i\boldsymbol{\rho}'_i\boldsymbol{\Psi}^{-2}\dot{\mathbf{Q}}_i \end{bmatrix}, \tag{25}$$

where $\mathbf{U}_i = [\mathbf{u}_{i(1)}, \ldots, \mathbf{u}_{i(n_2)}]$, $\mathbf{u}_{i(m)} = vec(-\gamma\dot{\mathbf{Q}}_{i(m)}\boldsymbol{\Delta}_i' - \boldsymbol{\Delta}_i\dot{\mathbf{Q}}_{i(m)}' + \gamma^2[\mathbf{Q}_i\dot{\mathbf{Q}}_{i(m)}' + \dot{\mathbf{Q}}_{i(m)}\mathbf{Q}_i'])$, $\dot{\mathbf{Q}}_{i(m)}$ = the mth column of $\dot{\mathbf{Q}}_i$, and $\boldsymbol{\Delta}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}$. This leads to

$$\mathbf{H}_{12} = \begin{bmatrix} \gamma\sum_i \mathbf{X}_i'\boldsymbol{\Psi}^{-1}(\mathbf{I}-\Upsilon)\mathbf{A}_i & 0 & \ldots & 0 & 0 \\ \gamma\boldsymbol{\alpha}'\sum_i \mathbf{A}_i(\mathbf{I}-\Upsilon)\mathbf{A}_i & n\gamma\mathbf{F}_1 & \ldots & n\gamma\mathbf{F}_{n_{\boldsymbol{\omega}}} & n\gamma\mathbf{F}_\delta \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix},$$

where $\mathbf{F}_j = tr(\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}_D\dot{\Upsilon}_j')$ and $\mathbf{F}_\delta = tr(\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}_D\dot{\Upsilon}_\delta')$,

Extending Proposition 3

Under (7) with $\mathbf{Z} \neq \mathbf{R}$, or under the model of Section 6.1, one suggestion was to estimate the variance components using (14). Let $\boldsymbol{\theta}_1^* = (\boldsymbol{\beta}', \gamma, \boldsymbol{\psi}')$ where $\boldsymbol{\psi} = vec(\boldsymbol{\Psi})$. The estimating equations for $\boldsymbol{\beta}$, γ and $\boldsymbol{\psi}$ consist of (19) along with $\sum_i(\mathbf{r_i}\mathbf{r_i}' - \boldsymbol{\Psi}) = \mathbf{0}$, since $\hat{\boldsymbol{\Psi}}_{\mathbf{C}} = \sum_i \mathbf{r_i}\mathbf{r_i}'/n$. Now $\mathbf{S}_1(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2)$ will have the first two lines of (21) while the last two lines are replaced by $vec(\sum_i(\boldsymbol{\rho}_i\boldsymbol{\rho}_i')) - vec(\boldsymbol{\Psi}^{-1})$. The form of $\mathbf{S}_2$ will depend on which model is being used. The asymptotic covariance of $\hat{\boldsymbol{\theta}}^*$ will be of the same form as given in (22) with the new definitions of $\mathbf{S}_1$ and $\mathbf{S}_2$. This immediately yields the asymptotic covariance for $\hat{\boldsymbol{\beta}}_C$ and $\hat{\gamma}_C$. The asymptotic covariance matrix of $\hat{\boldsymbol{\sigma}}_C$ is $(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}(\mathbf{V})\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}$ where $\mathbf{A}(\mathbf{V})$ is the asymptotic covariance of $\mathbf{V} = \hat{\boldsymbol{\psi}} - vec(\hat{\gamma}^2\hat{\boldsymbol{\Sigma}}_D(\mathbf{I} - \hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D))$. Since $\hat{\boldsymbol{\Sigma}}_W = \hat{\boldsymbol{\Sigma}}_D + \hat{\sigma}_\delta^2\mathbf{I}$, it is easy to show that $\hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D = \mathbf{I} - \hat{\sigma}_\delta^2\hat{\boldsymbol{\Sigma}}_W^{-1}$, so $\hat{\boldsymbol{\Sigma}}_D(\mathbf{I} - \hat{\boldsymbol{\Sigma}}_W^{-1}\hat{\boldsymbol{\Sigma}}_D) = \hat{\boldsymbol{\Sigma}}_D\hat{\sigma}_\delta^2\hat{\boldsymbol{\Sigma}}_W^{-1} = \hat{\sigma}_\delta^2(\mathbf{I} + \hat{\sigma}_\delta^2\hat{\boldsymbol{\Sigma}}_D^{-1})^{-1}$. Hence $\mathbf{V} = \hat{\boldsymbol{\psi}} - \hat{\gamma}^2\hat{\sigma}_\delta^2 vec((\mathbf{I} + \hat{\sigma}_\delta^2\hat{\boldsymbol{\Sigma}}_D^{-1})^{-1})$ and $\mathbf{A}(\mathbf{V})$ can be derived in a relatively straightforward fashion from $\mathbf{A}(\hat{\boldsymbol{\theta}}^*)$. The matrix $\mathbf{H}_{12}$ in this case is

$$\mathbf{H}_{12} = \begin{bmatrix} \gamma\sum_i \mathbf{X}_i'\boldsymbol{\Psi}^{-1}E(\dot{\mathbf{Q}}_i) \\ (\sum_i \boldsymbol{\Psi}^{-1}\mathbf{X}_i\boldsymbol{\beta})'E(\dot{\mathbf{Q}}_i) + 2\gamma\sum_i E(\mathbf{Q}_i'\boldsymbol{\Psi}^{-1}\dot{\mathbf{Q}}_i) - \sum_i E(\mathbf{Y}_i'\boldsymbol{\Psi}^{-1}\dot{\mathbf{Q}}_i) \\ \mathbf{0} \end{bmatrix},$$

where $\dot{\mathbf{Q}}_i$ is as earlier but with $\boldsymbol{\Upsilon} = \boldsymbol{\Sigma}_D\boldsymbol{\Sigma}_W^{-1} = \mathbf{I} - \sigma_\delta^2(\sigma_\delta^2\mathbf{I} + \boldsymbol{\Sigma}_D)^{-1}$.

References

Buonaccorsi, J. P. (1989). Errors-in-variables with systematic biases. *Comm. Statist. Part A - Theory and Methods* **18**, 1001-1021.

Buonaccorsi, J. P. (1990). Double Sampling for exact values in some multivariate measurement error problems. *J. Amer. Statist. Assoc.* **85**, 1075-1082.

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

Carroll, R. J., Lin, X. and Wang, N. (1997). Generalized linear mixed measurement error models. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions* (Edited by T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren and R. D. Wolfinger). Springer-Verlag, New York.

Crowder, M. J. and Hand, D. J. (1990). *Analysis of Repeated Measures.* Chapman and Hall, London.

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

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

Gleser, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *J. Amer. Statist. Assoc.* **87**, 696-707.

Gong, G. and Sammaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* **9**, 861-869.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariances. *Biometrics* **42**, 805-820.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lehmann, E. (1983). *Theory of Point Estimation.* Wiley Interscience, New York.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

Parke, W. R. (1986). Pseudo maximum likelihood estimation: the asymptotic theory. *Ann. Statist.* **14**, 355-357.

Paterno, E. M. and Amemiya, Y. (1996). Random effect and random coefficient analysis with errors-in-variables. *ASA Proceedings of Business and Economic Statistics*, 76-79.

Takeuchi, M. and Ware, J. H. (1990). Measurement error in the balanced linear growth curve model. *International Biometric Conference Proceedings*, 183-199.

Takeuchi, M. (1992). Parameter Estimation in Linear Models for Longitudinal Data with Mismeasured Covariates. Unpublished D. Sc. thesis, Harvard University.

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

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

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.

Wang, N., Lin, X. and Guttierrez, R. B. (1999). A bias correction regression calibration approach in generalized linear mixed measurement error models. *Comm. Statist.- Theory and Methods* **28**, 217-233.

Department of Mathematics and Statistics, University of Massachusetts, Amherst MA 01003, U.S.A.

E-mail: johnpb@math.umass.edu

Section of Biostatistics and Epidemiology, Department of Community and Family Medicine, Dartmouth Medical School, Lebanon NH 03756, U.S.A.

E-mail:

(Received January 1999; accepted September 1999)