Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models

Naisyin WANG, Xihong LIN, Roberto G. GUTIERREZ, and Raymond J. CARROLL

We consider generalized linear mixed models (GLMMs) for clustered data when one of the predictors is measured with error. When the measurement error is additive and normally distributed and the error-prone predictor is itself normally distributed, we show that the observed data also follow a GLMM but with a different fixed effects structure from the original model, a different and more complex random effects structure, and restrictions on the parameters. This characterization enables us to compute the biases that result in common GLMMs when one ignores measurement error. For instance, in one common situation the biases in parameter estimates become larger as the number of observations within a cluster increases, both for regression coefficients and for variance components. Parameter estimation is described using the SIMEX method, a relatively new functional method that makes no assumptions about the structure of the unobservable predictors. Simulations and an example illustrate the results.

KEY WORDS: Asymptotic bias; Corrected penalized quasi-likelihood; Measurement error; Random effects; Variance components.

1. INTRODUCTION

Correlated data are frequently observed in various studies, such as longitudinal studies, clinical trials, and familial studies. Generalized linear mixed models (GLMMs) have become increasingly popular for analyzing such correlated and overdispersed data (see Breslow and Clayton 1993 for examples). A potential difficulty in making inference in GLMMs is that a full-likelihood analysis is burdened by often intractable numerical integration (although see McCulloch 1997 for Monte Carlo computation). Hence various approximate and Bayesian inference procedures have been proposed. The approximations include Laplace’s approximations (Breslow and Lin 1995; Liu and Pierce 1993), penalized quasi-likelihood (PQL) (Breslow and Clayton 1993; Schall 1991), and corrected penalized quasi-likelihood (CPQL) (Lin and Breslow 1996). The Bayesian procedures include EM-type algorithms (Gibson and Laird 1984) and the Gibbs sampler (Zeger and Karim 1991).

A common problem for analyzing correlated data is the presence of covariate measurement error. For example, it has been well documented in the literature that covariates such as blood pressure (Carroll, Ruppert, and Stefanski 1995), urinary sodium chloride (USC) level (Wang, Carroll, and Liang 1996), and exposure to pollutants (Tosteson, Stefanski, and Schafer 1989) are often subject to measurement error. In a longitudinal hypertension study, a patient’s hypertensive status may vary from one hospital visit to another due to different average USC levels prior to the hospital visit. A child’s respiratory status may change from time to time depending on different amounts of pollutants (e.g., NO₂ or ozone) to which the child is exposed at different times. In the Framingham Heart Study data that we examine in Section 8, a binary outcome for the presence or absence of left ventricular hypertrophy (LVH) was observed every 2 years in an 8-year period for 75 coronary heart disease patients. The primary interest was to study the association between the risk of LVH and the time-varying covariate systolic blood pressure (SBP), after adjusting for other covariates including baseline age, smoking status, body mass index, and exam number (1-4). Because it is appropriate to assume biologically that the risk of LVH depends on the average SBP prior to the exam rather than on the SBP measured at the exam, one needs to model the measurement error in SBP while accounting for correlation among multiple observations measured repeatedly over time for each patient.

The problem of measurement error with independent observations has a vast literature in linear models (Fuller 1987) and a growing literature in generalized linear models and other nonlinear models (Carroll et al. 1995). Generally, the literature distinguishes between functional modeling, in which nothing is assumed about the unobserved predictors, and structural modeling, in which specific assumptions are made about the distributional structure of these unobserved predictors. The effects of measurement error on the analysis of clustered data and ways to correct for these effects are not well understood, however.

In this article, we propose a new class of models, generalized linear mixed measurement error models (GLMMeMs), which model the correlation and the measurement error simultaneously (Sec. 2). We explore GLMMeMs from two directions: bias analysis and functional inference using the SIMEX method (Cook and Stefanski 1994). To illustrate the fundamental impact of measurement error and our pri-
primary findings, we concentrate on a simple but representative GLMMeM in our bias calculations (see Sec. 2.2); however, the proposed SIMEX method is applicable to the general GLMMeMs. In Sections 3 and 4 we study the bias in parameter estimation when the measurement error is not properly taken into account. The work here is facilitated by our showing that a GLMMeM can be viewed as a GLMM, with the same link function but with different fixed-effects and random-effects structures. This characterization enables us to compute the biases in parameter estimates resulting from ignoring measurement error. The bias analysis results are illustrated using several common GLMMs, including linear, logistic, and Poisson mixed models. The directions of the biases are complex and sometimes counterintuitive. For example, we show that in a particular but common setup, the biases in parameter estimates increase with the cluster size.

The description of the GLMMeM brings out the fact that likelihood estimation in this context requires the specification of a cluster-specific joint distribution for the unobserved covariates. Just as in the ordinary generalized linear model (GLM), concerns about robustness to distributional assumptions arise, but in GLMMeMs there are additional robustness concerns with respect to the covariance structure of the unobservables. In Section 5 in a special case we compute the biases resulting from a maximum likelihood analysis that accounts for measurement error but incorrectly specifies the within-cluster covariance structure of the unobservable.

In Section 6 we pursue functional estimation of regression coefficients and variance components and investigate the SIMEX procedure of Cook and Stefanski (1994). We also point out that the naïve regression calibration approach often yields inconsistent estimates of certain parameters in GLMMeMs. We provide numerical results of a simulation and an example in Sections 7 and 8, and concluding remarks in Section 9.

2. THE GENERALIZED LINEAR MIXED MEASUREMENT ERROR MODEL

2.1 The General Model

Suppose that the data are obtained from m independent clusters with outcome variable $Y_{ij}$, unobserved true covariate $X_{ij}(p_i \times 1)$, observed $X_{ij}$-related covariate $W_{ij}$, and other observed covariates $Z_{ij}(p_j \times 1)$ and $A_{ij}(q \times 1)$, where $i = 1, \ldots, m$ identifies the cluster; $j = 1, \ldots, n_i$ identifies subjects within clusters; and $(X_{ij}, Z_{ij})$ and $A_{ij}$ are associated with the fixed effects and random effects. That is, we consider the situations where the error-prone covariates $X_{ij}$ are associated with fixed coefficients. Given the covariates $X_{ij}, Z_{ij}, A_{ij}$ and an unobserved $q \times 1$ random-effects vector $b_i$, the observations $Y_{ij}$ in the $i$th cluster are assumed to be independent with means $\mu_{ij}^{b_i}$ and variances $\phi \mu_{ij}^{b_i^2}$, where $\phi$ is a scale parameter, $\kappa_{ij}$ is a prior weight (e.g., binomial denominator), and $u(\cdot)$ is a variance function. The GLMM of $Y$ given $X$ and $Z$ is constructed by assuming that the conditional mean $\mu_{ij}^{b_i}$ is related to $X_{ij}, Z_{ij}$, and $A_{ij}$ through a GLM,

$$g(\mu_{ij}^{b_i}) = \beta_0 + X_{ij}' \beta_{x} + Z_{ij}' \beta_{z} + A_{ij}' b_{i},$$

where $g(\cdot)$ is a monotonic differentiable link function, the random effects $b_{i}$ are independent of the covariates and are independent N(0, $D(\theta)$), and $\theta$ is an $l \times 1$ vector of variance components. Model (1) allows flexible correlation structures by assuming appropriate design matrix $A_{ij}$ and covariance matrix $D$ of the random effects $b_{i}$.

Define $Y_i = (Y_{i1}, \ldots, Y_{in_i})', X_i = (X_{i1}, \ldots, X_{in_i})'$, and $Z_i$ and $W_i$ similarly. The integrated quasi-likelihood of $Y_i$ given $(X_i, Z_i)$ in the $i$th cluster is

$$L_i(Y_i|X_i, Z_i; \beta, \theta) \propto |D|^{-1/2} \left( \sum_{j=1}^{n_i} l_{ij}(Y_{ij}|X_{ij}, Z_{ij}, b_i) - \frac{1}{2} b_i'^T D^{-1} b_i \right) db_i,$$

where $l_{ij}(Y_{ij}|X_{ij}, Z_{ij}, b_i) \propto \int_{Y_{ij} - u}^{Y_{ij}} \kappa_{ij}(Y_{ij} - u)/\{f(u)\} du$ denotes the conditional log quasi-likelihood of $Y_{ij}$ given $(X_{ij}, Z_{ij}, b_i)$ (see Breslow and Clayton 1993, eq. 2).

The model is completed by adding the measurement error structure. The most convenient structure is additive error, so that

$$W_{ij} = X_{ij} + U_{ij},$$

where $U_{ij}$ are independent of the $X_{ij}$ and are independent N(0, $\Sigma_{xu}$). When $W$ and $X$ are scalar, we write the measurement variance $\Sigma_{xu}$ simply as $\sigma_u^2$. Model (3) is essential to our analytical closed-form bias calculations in Sections 3–5 but not for numerical bias calculations, estimation, and inference. Neither the independence of the $U_{ij}$ nor the additivity is required (see Sec. A.5 in the Appendix).

The joint integrated quasi-likelihood in the $i$th cluster is

$$L_i(Y_i|X_i, W_i|Z_i) = \int L_i(Y_i|X_i, Z_i) L_i(W_i|X_i, Z_i) L_i(Z_i|X_i) dX_i,$$

where $L_i(Y_i|Z_i)$ is the likelihood function of $X_i$ (so far unspecified) and $L_i(W_i|X_i, Z_i)$ is the error distribution, which is often assumed to be independent of $Z_i$ as in (3). The dependence of the quasi-likelihood on the within-cluster conditional distribution of the unobserved $X_i$’s leads to issues of model robustness.

A special case is instructive. Suppose that $X_{ij}$ is scalar and that $X_i = 1, \gamma_0 + Z_{i0} + e_{i0}$, where $1$ is an $n_i \times 1$ vector of $1$s and $e_{i0}$ given $Z_{i0}$ is normally distributed with mean $0$ and covariance matrix $\Sigma_{xu}$. Denote an $n_i \times n_i$ identity matrix by $I_i$ and the reliability matrix by $\Lambda_i = \Sigma_{xu}/(\Sigma_{xu} + \text{cov}(U_i))^{-1}$. Note that $\Sigma_{xu}$ and $\Lambda_i$ depend on $i$ through their dimensions $n_i$, but the set of unknown parameters in $\Sigma_{xu}$ and $\Lambda_i$ does not depend on $i$. We can write $X_i = (I_i - \Lambda_i)(1, \gamma_0 + Z_{i0} + e_{i0}) + \Lambda_i W_i + b_i^*$, where the sum of the first two terms corresponds to $E(X_i|W_i, Z_i)$ and $b_i^* = X_i - E(X_i|W_i, Z_i) = (I_i - \Lambda_i)e_{i0} - \Lambda_i U_i$ is N(0, $(I_i - \Lambda_i)^2 \text{cov}(U_i)$) and is independent of $b_i$ and $W_i$. The independence between $b_i^*$ and $W_i$ can be easily checked by
showing \( \text{cov}(b_i^*, W_i) = 0 \). The expression of \( X_i \) indicates that the \( j \)-th component of \( X_i \) can be written as, say,

\[
X_{ij} = \omega_{ij} + \eta_i^T Z_i^T \alpha_z + W_i^T \alpha_{w,i} + C_{ij}^T b_i^*.
\]

(5)

Because \( b_i^* = X_i - E(X_i | W_i, Z_i) \) and \( Y_i \) is independent of \( W_i \), given \( X_i, (Y_i | W_i, Z_i, b_i^*, b_i^*) \) has the same distribution as \( (Y_i, Z_i, b_i^*) \) and follows the same conditional generalized linear model (1), except that \( X_i \) is replaced by (5). In other words, the observed data \( (Y_i | W_i, Z_i) \) follow a GLMM as follows:

\[
g(b_{i,j,w}) = (\beta_0 + \alpha_{z,i} \beta_z) + W_i^T \alpha_{w,i} + (\eta_i^T Z_i Z_i^T \beta_x + Z_i^T \beta_z) + (A_i^T b_i + C_{ij}^T \beta_x b_i^*),
\]

where \( b_{i,w} = (b_i^T, b_i^*)^T \) denotes a vector of the new random effects.

Equation (6) shows that ignoring the measurement error may result in misspecifying both the fixed-effects and random-effects structures. The cluster size \( n_c \), which is implicit in the reliability matrix \( \Lambda_z \), also plays an important role in the asymptotic bias in maximum likelihood estimator (MLE) under a misspecified model as \( n_c \to \infty \). A general approach for bias analysis is to maximize the probability limit of the log quasi-likelihood of the misspecified model, which is equal to its expectation, when model (6) is true. Calculations of this expectation often involve numerical integration. Appendix Section A.5 gives a brief discussion on numerical integration techniques using Gaussian–Hermite quadrature. When the misspecified model and the observed data \( (Y_i, W_i, Z_i) \) models follow the same type of GLMMs, bias calculations can be greatly simplified by using the correspondence of their mean models (see Sec. 3). The foregoing general bias calculation strategy applies to an arbitrary GLMM. A more complicated GLMM structure requires no extra procedures than the ones used for a simple structure, except that the results will be more complicated. We hence consider simple GLMMs in our bias analysis to show the fundamental impact of measurement error and our primary findings, and to demonstrate the basic techniques used in bias calculations.

2.2 Specific Models Considered in Bias Analysis

In the bias analyses in Sections 3–5, for simplicity we assume that \( n_i = n_c \), the \( X_i \) are scalar, and simple random intercept GLMM \( g(\mu_{i,x}) = \beta_0 + \beta_2 X_i + b_i \), where the \( b_i \) are independent \( N(0, \theta) \). We distinguish between two cases depending on the likelihood structure \( L_i(X_i) \) of \( X_i \).

The homogeneous case occurs when \( X_i \) are marginally independent and have the same distribution irrespective of the cluster, so that

\[
X_{ij} = \mu_x + e_{ij},
\]

(7)

where the \( e_{ij} \) are independent \( N(0, \sigma_e^2) \). In the heterogeneous case, conditional on a cluster random effect \( \beta_i \), the distributions of the \( X_{ij} \) differ from cluster to cluster. In the version that we study here theoretically, we assume that the conditional cluster means differ, so

\[
X_{ij} = \mu_x + \alpha_i + e_{ij},
\]

(8)

where the \( \alpha_i \) independent of the model random effect \( b_i \), are independent \( N(0, \sigma_{\alpha}^2) \).

Although the two models that we consider here have simple structures, the bias calculation techniques used in Sections 3–5 are applicable to more complicated cases. Specifically, as indicated in Section 2.1, we can accommodate the covariates measured without error \( Z_i \) by treating \( Z_i \) as fixed and further allow for multivariate \( X_{ij} \) and a more complicated structure of the random effects \( b_i \). For example, to accommodate multivariate \( X_{ij} \), we can define \( T_i = (X_{i1}, \ldots, X_{im}) \), define \( W_i \) analogously, and modify (6).

To appreciate the practical differences between the homogeneous and heterogeneous models, we consider an ozone exposure example. When the subjects are from the same site, one distribution can be used to describe the behavior of the short-term average ozone exposures for all subjects, and the homogeneous model is appropriate. In this case the variations in ozone exposure may be mainly seasonal. On the other hand, if these subjects are from different neighborhoods, then the heterogeneous model, which accommodates cross-cluster variation, should be used. Clearly, a homogeneous model is a special case of the heterogeneous model \( (\sigma_{\alpha}^2 = 0) \), and this seems to indicate that one should consider only the heterogeneous model. However, assuming a heterogeneous model while the homogeneous model holds results in estimators with unnecessarily large variance.

The next three sections are devoted to studying the asymptotic biases in regression coefficients and variance component estimators under three misspecified models. To facilitate the bias analysis, it is helpful to rewrite (6) in the special cases under consideration. The calculations outlined in Appendix Section A.1 show that the observed data under the heterogeneous GLMM satisfy

\[
g(b_{i,x,j}) = \beta_0^* + \beta_2^* W_i + \beta_3^* T_i + \hat{b}_i^* + \hat{b}_i^*,
\]

(9)

where

\[
\beta_0^* = \beta_0 + (1-\lambda) \hat{\lambda} \mu_x \beta_x,
\]

\[
\beta_2^* = \lambda \beta_x,
\]

\[
\beta_3^* = (1-\lambda)(1-\hat{\lambda}) \beta_x,
\]

\[
\hat{\lambda} = (\sigma_e^2 + \sigma_\alpha^2) / (\sigma_e^2 + \sigma_\alpha^2 + n_i \sigma_{\alpha}^2),
\]

\[
\lambda = \sigma_e^2 / (\sigma_e^2 + \sigma_\alpha^2),
\]

\[
W_i = n_i^{-1} \sum_{j=1}^n W_{ij},
\]

and

\[
\hat{b}_i = (b_{i1}, b_{i2}, \ldots, b_{im})^T.
\]

The random effects \( b_{i} \) and \( b_{i,j}^* \) are independent of \( W_i \), and are mutually independent and distributed as \( N(0, \theta) \) and \( N(0, \gamma) \), where \( \theta = \beta_2 + (1-\lambda)(1-\hat{\lambda}) \beta_2^2 n_i / n \) and \( \gamma = \lambda \sigma_e^2 \beta_2^2 n_i / n \). The exact expressions of \( b_{i}^* \) and \( b_{i,j}^* \) are given in Appendix Section A.1.
3. BIAS IN THE NAIVE ESTIMATOR UNDER THE HOMOGENEOUS MODEL

In this section we study the asymptotic biases in naive estimators of regression coefficients and variance component when the homogeneous model (7) holds. The naive estimator is defined as the estimator under the model that ignores the measurement error,

$$g(\mu_{yi, w}) = \beta_0 + \beta_2W_{ij} + b_i.$$  \hspace{1cm} (10)

From (9), the homogeneous model has \( \hat{\lambda} = 1, \beta_2 = 0 \), and \( \theta = \theta' \) and thus can be written as

$$g(\beta_{ij, w}^b) = \beta_0^b + \beta_2^bW_{ij} + b_i^b$$ \hspace{1cm} (11)

where the \( b_i \) are independent \( N(0, \theta) \) and the \( b_{ij}^b \) are independent \( N(0, \gamma) \).

Because conditional on the \( b_i \) and the \( W_{ij} \), the model for the observed data (11) corresponds to an overdispersed GLM, the bias analysis of the naive estimators can proceed by comparing the conditional mean \( E(Y_{ij}|W_{ij}, b_i) \) and the conditional variance \( \text{var}(Y_{ij}|W_{ij}, b_i) \) under the naive model (10) with those under the homogeneous model (11). Note that this conditional mean and variance under the homogeneous model can be easily obtained by integrating out \( b_{ij}^b \) from (11).

Although the naive model correctly assumes that the \( Y_{ij} \) are independent conditional on \( W_{ij} \) and \( b_i \), it may misspecify both the mean and variance conditional on \( W_{ij} \) and \( b_i \). When there is a correspondence between \( E(Y_{ij}|W_{ij}, b_i) \) and \( \text{var}(Y_{ij}|W_{ij}, b_i) \) under these two models, they follow the same type of GLMM, and hence the asymptotic biases in regression coefficients and the variance component can be easily derived. Otherwise, the calculations are often difficult, and closed-form expressions for the biases are not always available.

3.1 The Linear Mixed Model for Gaussian Data

It can be easily shown that in the linear mixed model, the observed data also follow a linear random intercept model, with the terms \( b_{ij}^b \) in (11) absorbed into the within-cluster variance in the responses. The naive estimators hence asymptotically converge to \( \beta_{0, \text{naive}} = \beta_0 + (1 - \lambda)\gamma^2\beta_{2, \text{naive}} = \lambda\beta_2 \), and \( \theta_{\text{naive}} = \theta \). Thus the naive estimators of the regression coefficients are asymptotically biased in a usual way, but the naive estimator of the variance component is asymptotically unbiased.

3.2 The Probit, Logistic, and Log-linear Mixed Models for Binary Data

When \( Y_{ij} \) given \( X_{ij} \) follows a probit random intercept model, so too do the observed data \( Y_i \) given \( W_i \). Let \( \tau = (1 + \gamma)^{1/2} = (1 + \lambda\gamma^2\beta_2)^{1/2} \); the derivations outlined in Appendix Section A.2 show that \( \beta_{0, \text{naive}} = \beta_0 + (1 - \lambda)\gamma\beta_2 / \tau, \beta_{2, \text{naive}} = \lambda\beta_2 \tau, \) and \( \theta_{\text{naive}} = \theta / \tau^2 \). These results indicate that the naive estimators of both \( \beta_2 \) and \( \theta \) are asymptotically biased toward 0. For the logistic model, exact closed-form results are not available. But by approximating the standard logistic distribution function by the distribution function of a mean 0 normal random variable that has standard deviation \( \sigma = 15\gamma / (16\gamma^2) \approx 1.7 \), similar calculations show that the same bias expressions obtain, but with \( \tau \) replaced by \( \tau^* = (1 + \lambda\gamma^2\beta_2^2 / \gamma^2)^{1/2} \).

The log-linear model assumes a log link function and is useful in ecological studies, where the disease rates may be low. Using the identity \( \int \exp(a + \tau \phi)d\phi (\phi / \sigma) = \exp(a + \sigma^2 / 2) \) for any constants a and \( \sigma \), where \( \Phi(\cdot) \) denotes the cumulative probability function of a standard normal random variable, similar calculations to those given in Appendix Section A.2 show that \( (Y_i, W_i) \) also follows a log-linear random intercept model and that the naive estimators converge to \( \beta_{0, \text{naive}} = \beta_0 + (1 - \lambda)\mu_0^2\beta_2 + \gamma / \tau, \beta_{2, \text{naive}} = \lambda\beta_2 \), and \( \theta_{\text{naive}} = \theta \). These results, except for the intercept, agree with those in the linear mixed model.

3.3 The Poisson Mixed Model for Count Data

Define \( \theta^* = \theta + \gamma, d_{ij} = \exp(\theta^*/2 + \beta_2 + \beta_{2, \text{naive}}X_{ij}), \) and \( c_{ij} = \exp(\theta^*/2 + \beta_0 + \beta_{2, \text{naive}}W_{ij}), \) where \( \beta_0 \) and \( \beta_2 \) are defined in Section 2.2 with \( \lambda = 1 \). Then the conditional mean and covariance of \( Y_i \) given \( W_i \) under the Poisson mixed model are \( E(Y_{ij}|W_{ij}) = c_{ij}, \text{var}(Y_{ij}|W_{ij}) = c_{ij} + c^2_{ij}\{\exp(\theta^* - 1) - 1\}, \) and \( \text{cov}(Y_{ij}, Y_{ik}|W_{ij}, W_{ik}) = c_{ij}c_{ik}\{\exp(\theta^* - \gamma - 1)\} \), and those of \( Y_i \) given \( X_i \) follow the same structure except that \( c_{ij} \) is replaced by \( d_{ij} \) and that \( \text{cov}(Y_{ij}, Y_{ik}|X_{ij}, X_{ik}) = d_{ij}d_{ik}\{\exp(\theta^* - 1)\} \). The lack of correspondence between the two covariances reveals that \( (Y|W) \) do not follow the same GLMM structure. Thus the approach based on the conditional mean correspondence is not applicable.

Using the techniques of reparameterization and applying the properties of sufficient statistics and maximum likelihood estimators, the calculations outlined in Appendix Section A.3 show that \( \beta_{0, \text{naive}} = \beta_0^b + (\theta + \gamma - \theta_{\text{naive}}) / 2, \beta_{2, \text{naive}} = \lambda\beta_2 \), and

$$\theta_{\text{naive}} = \theta + \log \left\{ \frac{(n - 1) + \exp(\beta_{2, \text{naive}}^2)}{(n - 1) + \exp(\beta_{2, \text{naive}}^2)} \right\}. \hspace{1cm} (12)$$

Equation (12) suggests that the naive estimator \( \theta_{\text{naive}} \) over-estimates \( \theta \) and that its bias depends on the cluster size \( n \) and decreases monotonically to 0 as the cluster size \( n \) goes to infinity.

4. BIAS IN THE NAIVE ESTIMATOR UNDER THE HETEROGENEOUS MODEL

In this section we study the asymptotic bias of the naive estimator when the heterogeneous model (8) holds. We see from (9) that the heterogeneous conditional quasi-likelihood \( (Y|W) \) also corresponds to a GLMM, with the same random-effects structure as in the homogeneous case but with an additional fixed effect—namely, the within-cluster mean of the \( W \)'s. When the heterogeneous model is true, a comparison of (9) and (11) suggests that the naive model (10) misspecifies both the fixed-effects and the random-effects structures. This double misspecification makes the bias analysis much more complicated, and closed-form solutions are often not available.

Nonetheless, we can calculate the asymptotic bias in the naive estimator when the cluster size \( n \) goes to infinity.
ity. Specifically, our calculations show that the bias in the naive estimator becomes the same as in the homogeneous case when \( n \to \infty \), except that \( \theta \) is replaced by \( \theta + (1 - \lambda)^2 \sigma^2_{\beta_x} \beta_x^2 \). This can be easily shown by noting that for the heterogeneous GLMM (9), as \( n \to \infty \), we have \( \hat{\lambda} \to 0, \beta_x' \to (1 - \lambda)\beta_x \), \( \theta' \to \theta \), and \( \hat{W}_i = \mu_i + a_i + \hat{u}_i + \hat{e}_i \to \mu_i + a_i \). Consequently, the heterogeneous GLMM in (9) becomes

\[
g(\mu_{ij, w}) = \beta_0' + \beta_1' W_{ij} + b_{ij} + b_{ij}^a,
\]

(13)

where \( \beta_0' = \beta_0 + (1 - \lambda)\mu_0 \beta_x, b_{ij} = b_i + (1 - \lambda)\beta_x a_i \).

Now follows \( N(0, \theta + (1 - \lambda)^2 \sigma^2_{\beta_x} \beta_x^2) \), and \( b_{ij}^a \) stays the same. Because within a cluster (13) has the same structure as (11), and because the cluster size is infinite, the maximum likelihood estimator within-cluster intercept (and slope) must have the same form, except that \( b_i \) is replaced by \( b_{ij}^a \) and hence \( \theta \) is replaced by \( \theta + (1 - \lambda)^2 \sigma^2_{\beta_x} \beta_x^2 \).

In the next two sections we use analytic and numerical approaches to study the asymptotic bias in the naive estimator in the linear and logistic mixed models when the cluster size \( n \) is finite.

4.1 The Linear Mixed Model

Denote the residual variance by \( \sigma^2 \), which corresponds to the scale parameter \( \phi \) in Section 2. Let the probability limits of the naive estimators as \( m \to \infty \) be \( \beta_{naive} = \{\beta_{0,naive}, \beta_{x,naive}\}^{T} \) and \( \theta = \{\theta_{naive}, \sigma^2_{naive}\} \). Then \( \beta_{naive} \) and \( \theta \) are solutions of the following equations, which are the probability limits of the linear mixed model score equations as \( m \to \infty \) (Harville 1977):

\[
E[V^T V^{-1}(Y - \nu\beta_{naive})] = 0
\]

(14)

and

\[
\frac{1}{2} \left[ E(Y - \nu\beta_{naive})^T V^{-1} \frac{\partial V}{\partial \theta_i} \right] 
\times V^{-1}(Y - \nu\beta_{naive}) - \text{tr} \left( V^{-1} \frac{\partial V}{\partial \theta_i} \right) \right] = 0,
\]

(15)

where \( V = (1, W) \), \( V = \sigma^2_{naive} I + \sigma_{naive}^2 J \) is an \( n \times n \) matrix of 1s, and the expectations are taken with respect to both \( W \) and \( Y \). For simplicity, here we have removed the subscripts \( i \) of \( Y_i \) and \( W_i \), because they are identically distributed. Repeatedly using the equality

\[
E(X^T B X) = \text{tr}(BV) + \mu_{z}^2 BB \mu_{z},
\]

(16)

which holds for any positive definite matrix \( B \) and any random vector \( X \) following \( N(\mu_{z}, V_{x}) \), the calculations outlined in Appendix Section A.4 yield \( \beta_{0,naive} = \beta_0 + (1 - \lambda_0)\mu_0 \beta_x, \beta_{x,naive} = \lambda_0 \beta_x, \theta_{naive} = \theta + (1 - \lambda_0)^2 \sigma^2_{\beta_x} \beta_x^2 \), and

\[
\sigma^2_{naive} = \sigma^2 + \{(1 - \lambda_0)^2 \sigma^2_{\beta_x} + \lambda_0^2 \sigma^2_{\beta_x} \} \beta_x^2,
\]

(17)

where

\[
\lambda_0 = \frac{\sigma^2_{\beta_x} + \sigma^2_{\beta_x} \{1 + (n - 1)\theta_{naive}/\sigma^2_{naive}\}}{\sigma^2_{\beta_x} + \{\sigma^2_{\beta_x} + \sigma^2_{\beta_x} \} \{1 + (n - 1)\theta_{naive}/\sigma^2_{naive}\}}.
\]

A closed-form solution to (17) and (18) does not seem to be available. However, defining \( \rho = \theta_{naive}/\sigma^2_{naive} \), using (17) and writing (17) and (18) as joint equations of \( (\lambda_0, \rho) \), one can easily solve them numerically.

Because \( \lambda \leq \lambda_0 \leq 1 \), the naive estimator \( \beta_{x,naive} \) is still attenuated, but to a lesser extent compared to that in the homogeneous case. In contrast, \( \theta_{naive} \) and \( \sigma^2_{naive} \) generally overestimate their true counterparts. Our theoretical calculations show that the values of \( \beta_{0,naive}, \beta_{x,naive}, \theta_{naive}, \) and \( \sigma^2_{naive} \) all depend on the cluster size \( n \). Some tedious calculations show that \( \partial \lambda_0/\partial \theta < 0 \). Hence \( \lambda_0 \) is a decreasing function of \( n \), and the biases in all naive estimators except \( \sigma^2_{naive} \) become more serious when \( n \) increases. As \( n \uparrow \infty, \lambda_0 \downarrow 1 \), and we obtain the results in the last paragraph before Section 4.1.

In Figures 1a and 1b, we numerically evaluate the biases in \( \beta_{x,naive} \) and \( \theta_{naive} \) for \( \sigma^2_{\beta_x} \) varying between 0 and 1. For each plot, we obtained four curves corresponding to \( n = 2, 5, 10, \) and \( \infty \). The parameter configurations were \( \beta_0 = 0, \beta_x = 2, \theta = 1, \sigma^2 = 1, \sigma^2_{\beta_x} = 1.5, \) and \( \mu_0 = 0, \sigma^2_{\beta_x} = 1 \). The relative bias is defined as the bias of a parameter divided by its true value. Note the major features of the plot—the naive estimator of \( \beta_x \) is attenuated, the naive estimator \( \theta_{naive} \) overestimates \( \theta \), and the biases in \( \beta_{x,naive}, \) and \( \theta_{naive} \) increase as \( n \) increases.

4.2 The Logistic Mixed Model

Our interest here is in calculating the bias in \( \theta_{naive} \) in the logistic mixed model when \( n \) is finite and the heterogeneous model is true. Because there is no closed-form expression for these biases, even in the probit case, we calculated the asymptotic bias by numerically maximizing the probability limit of the log-likelihood of the naive model, which is calculated by its expectation, when the heterogeneous model is true. We brieﬂy describe our numerical methods in Appendix Section A.5. The parameter configurations used in our numerical calculations are identical to those used in Section 4.1. As shown in Figures 1c and 1d, the naive estimator \( \beta_{x,naive} \) underestimates \( \beta_x \) as usual, and its bias becomes larger as \( \sigma^2_{\beta_x} \) and \( n \) increase, whereas the bias in \( \theta_{naive} \) is no longer monotonic in \( \sigma^2_{\beta_x} \) and its direction depends on \( \sigma^2_{\beta_x} \) and \( \sigma^2_{\beta_x} \). For example, \( \theta_{naive} \) underestimates \( \theta \) when \( \sigma^2_{\beta_x} \) is close to 0 and overestimates \( \theta \) when \( \sigma^2_{\beta_x} \) increases. For \( n = 2 \), the measurement error effect on \( \theta_{naive} \) is less pronounced compared to its effect on \( \beta_{x,naive} \) however, there is substantial bias when \( n = \infty \). This simply points out once again that cluster size is important in the bias of estimates computed by ignoring measurement error.

5. Bias in the Homogeneous Maximum Likelihood Estimator Under the Heterogeneous Model

In this section we study the asymptotic bias in the MLE assuming the homogeneous model (7) when the heterogeneous model (8) in fact is true. The major issue is: What happens when one accounts for measurement error in a likelihood analysis, but incorrectly models the covariance of
the unobserved predictors? The calculations in this special case give some idea of the biases that can occur in structural modeling with an incorrectly specified structural model.

In our calculations we assume for simplicity that \( \sigma_u^2 \) is known. The unknown parameters under the homogeneous model are \((\beta_0, \beta_x, \theta)\) and \((\mu_x, \sigma_x^2)\). By writing (4) as \( L_0(Y, W) = L_0(Y_i | W_i) | L_i(W_i) \), a comparison of the homogeneous model (11) with the heterogeneous model (9) reveals that the bias in the homogeneous MLE comes from two sources: one from misspecification of the marginal likelihood of \( W_i \) and the other from misspecification of the likelihood structure of \( Y_i \) given \( W_i \). It is easily seen that the former ignores the cluster-level random effect \( \alpha_i \). In contrast to the naive model, the homogeneous GLMM (11) assumed by the latter misspecifies only the fixed-effects structure.

Denote the asymptotic limits of the homogeneous MLEs of \((\beta_0, \beta_x, \theta)\) and \((\mu_x, \sigma_x^2, \lambda)\) when the heterogeneous model is true by \((\beta_{0, \text{hom}}, \beta_{x, \text{hom}}, \theta_{\text{hom}})\) and \((\mu_{x, \text{hom}}, \sigma_{x, \text{hom}}^2, \lambda_{\text{hom}})\). Because the \( W_i \) are sufficient statistics for \((\mu_x, \sigma_x^2)\), some calculations give \( \mu_{x, \text{hom}} = \mu_x, \sigma_{x, \text{hom}}^2 = \sigma_x^2 + \sigma_{\alpha x}^2 \), and \( \lambda_{\text{hom}} = \sigma_{x, \text{hom}}^2 / (\sigma_{x, \text{hom}}^2 + \sigma_{\alpha x}^2) = (\sigma_x^2 + \sigma_{\alpha x}^2) / (\sigma_x^2 + \sigma_{\alpha x}^2) \).
the biases in $\beta_{z,\text{naive}}$ and $\theta_{\text{naive}}$ increase with $n$.  

In Figure 2 we numerically study the asymptotic biases in $\beta_{z,\hom}$ and $\theta_{\hom}$. The parameter configurations and setup in Figure 2 are identical to those used in Figure 1. These figures reflect our theoretical results. Specifically, the bias in $\beta_{z,\hom}$ is less than the bias in $\beta_{z,\text{naive}}$ and increases with $n$ (see Fig. 2a). As expected, Figures 1b and 2b are identical.

5.2 The Logistic Mixed Model

We now study the bias in the homogeneous MLE in the logistic mixed model when the heterogeneous model is true. Because no closed-form solution is available, we evaluated the bias by numerically maximizing the expectation of the log-likelihood of the homogeneous model when the heterogeneous model is true. We used numerical integration techniques similar to those in Section 4.2 to calculate this expectation; see Appendix Section A.5 for details.

The same parameter configurations as those in Figure 1 were used in our numerical bias calculations. The results are given in Figures 2c and 2d. Our calculations show that for $n = 2$, $\beta_{z,\hom}$ slightly overestimates $\beta_z$. Its bias slightly increases with $\sigma^2_x$ in the range that we consider. Contrary to the naive case, here the regular attenuation in $\beta_z$ estimator is not present for $n = 2$. The homogeneous MLE $\theta_{\hom}$ overestimates $\theta$ for both $n = 2$ and $n = \infty$, and the bias tends to be larger as $n$ increases.

It is interesting to compare the biases in naive estimates with the biases in homogeneous MLEs. The naive estimate of $\beta_z$ is much more biased than its homogeneous MLE counterpart. However, in figures not provided here, comparisons of $\theta_{\hom}$ and $\theta_{\text{naive}}$ for various values of $\sigma^2_x$ indicate that when $\sigma^2_x$ is small, the biases for $\theta_{\hom}$ and $\theta_{\text{naive}}$ have different directions.

6. SIMULATION EXTRAPOLATION AND FUNCTIONAL METHODS

Carroll et al. (1995) have drawn a distinction between functional modeling, in which nothing is assumed about the distribution of the X’s, and structural modeling, in which a parametric model (e.g., homogeneous or heterogeneous normal) is assumed and the MLE is computed. Functional methods have the advantage that when they apply, they are model robust. Two common functional methods are regression calibration and simulation extrapolation (SIMEX). We discuss their application in GLMMeMs in this section.

6.1 Inconsistency of the Regression Calibration Approach

The regression calibration method simply replaces $X$ by an estimate of $E(X|W, Z)$, and applies the naive method to these imputed values. Using (6) and (9) and noticing the sum of the first three terms is $E(X|W_i, Z_i)$, Wang, Lin, and Gutierrez (1997) showed that regular regression calibration in GLMMeMs often correctly specifies the fixed-effects structure but may misspecify the random-effects structures.
Because the regression coefficients and variance components are often not orthogonal in GLMMs, regression calibration can yield biased estimates of both $\langle \beta_2, \beta_2 \rangle$ and $\theta$, especially the latter.

Specifically, under the homogeneous X model, the regression calibration estimators of $\beta_0, \beta_2$, and $\theta$ are unbiased in the linear mixed model and biased in the logistic mixed model, with the asymptotic limits equal to $\beta_0/\tau^*, \beta_2/\tau^*$, and $(\tau^*)^{-2}\theta$, where $\tau^*$ is defined in Section 3.2. Under the heterogeneous X model, the regression calibration estimators of $\beta_0, \beta_2$, and $\theta$ converge to $\beta_0, \beta_2$, and $\theta'$ in the linear case, where $\theta'$ is defined in (9), and converge to $\beta_0/\tau^*, \beta_2/\tau^*$, and $(\tau^*)^{-2}\theta'$ in the logistic case. (For other bias analysis results and how to correct for the bias in naive regression calibration estimator, see Wang, Lin, and Gutierrez, 1997.)

6.2 Simulation Extrapolation Estimation

SIMEX is a simulation-based measurement error method, a full description of which has been given by Carroll et al. (1995) and Cook and Stefanski (1994). Rather than repeating the content of these references, we explain the
SIMEX procedure using Figure 3, which shows application of SIMEX to the LVH example in Section 8. The two parameters of interest are the regression coefficient of the log-transformed BP \( \beta_x \) and the variance component \( \theta \). For more details about this example, see Section 8.

Let the estimated value of \( \sigma_{\epsilon}^2 \) be \( \hat{\sigma}_{\epsilon}^2 \). The SIMEX method consists of two steps. The first step, the simulation step, is to establish the naive estimates if the measurement error were \((1 + \xi)\sigma_{\epsilon}^2 \). A simple empirical method is to add to the terms \( W_{ij} \) a normally distributed random variable with mean 0 and variance \( \xi \sigma_{\epsilon}^2 \), then recompute the naive estimates. Doing this only once may be misleading, because it introduces simulation variability, so instead one repeats the procedure a large number \( B \) times and computes the median of the resulting parameter estimates. For example, to estimate \( \beta_x \) in Figure 3, one does so for each \( \xi = (0.5, 1.0, 1.5, 2.0) \) and plots the resulting naive estimates of \( \beta_x \) versus \( \xi \). These are shown in small solid squares in Figure 3. Comparing the solid quadratic line connecting them to a plot such as Figure 1c, which is the bias curve of the naive estimate of \( \beta_x \) resulting from ignoring measurement error, the solid curve in Figure 3 corresponds to part of the curve in Figure 1c where \( \sigma_{\epsilon}^2 \geq \hat{\sigma}_{\epsilon}^2 \). The rest of the curve where \( \sigma_{\epsilon}^2 < \hat{\sigma}_{\epsilon}^2 \) is “hidden.” Therefore, the solid curve in Figure 3 is referred as partial bias plot.

In the second step, the extrapolation step, a model is fit to the partial bias plot. A typical default is the quadratic, which we used. This is because quadratic curves often approximate the bias curves in Figure 1 well, and quadratic extrapolation works well in our simulation. We also experimented with fitting a quadratic to the log-transformed naive variance component estimates, to little positive effect for most of the cases considered. After a model is fit, the “hidden” parts of the figure are filled in by extrapolating the model to the values less than \( \hat{\sigma}_{\epsilon}^2 \), which are the dashed curves in Figure 3. The extrapolated value at \( \xi = -1 \) (zero error variance) is the SIMEX estimator.

Our calculations in Sections 3–5 show that the biases in the naive estimators are continuous functions of \( \sigma_{\epsilon}^2 \). Straightforward derivations using the M estimator arguments given by Wang, Lin, Gutierrez, and Carroll (1997, appendix) indicate that the asymptotic results given by Carroll et al. (1995) and Stefanski and Cook (1995) are directly applicable. Note that both of the latter works accommodate nonadditive or dependent measurement errors provided that the exact extrapolants are used and that the error distributions are normal. In our work we chose the correctly penalized quasi-likelihood method (CPQL) as our naive estimator, because compared to the naive MLE, it is more stable, easier to implement, and converges much faster and its performance is comparable to its MLE counterpart when the variance components are small or moderate. Because the CPQL estimator is an M estimator, the results of Stefanski and Cook (1995) can be applied. The CPQL procedure is briefly described in Appendix Section A.6. (For more details, see Breslow and Lin 1995 and Lin and Breslow 1996.)

### 7. Simulations

We conducted a simulation study to evaluate the finite-sample performance of various estimators. Binary observations \( Y_{ij} \) were generated within each cluster with conditional success probabilities satisfying logit \([P(Y_{ij} = 1 | X_{ij}, Z_{ij}, b_i)] = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + b_i k = 1, 2, \ldots, m, i = 1, 2, \ldots, n\). The following combinations of parameters were considered: (a) \( m = 50, 100, n = 3, 8 \), which are common sample sizes in longitudinal studies; and (b) homogeneous model (7) with \( \mu_0 = 0 \), within-cluster variance \( \sigma_x^2 = 1 \), and between-cluster variance \( \sigma_{\mu}^2 = 0 \) and heterogeneous model (8) with \( \mu_x = 0, \sigma_x^2 = 1, \) and \( \sigma_{\mu}^2 = 1.5 \). A moderate measurement error variance \( \sigma_{\epsilon}^2 = 0.25 \) was considered for both the homogeneous and heterogeneous models. The exactly measured covariate \( Z \) was generated independently from a standard normal distribution. Other parameters used to specify the \( X \) and \( W \) models were \( \theta = 0.5, \beta_0 = 0, \beta_2 = 1, \) and \( \beta_x = 2 \). There were 1,000 simulations for each parameter setting. A single run for one dataset using our C program with \( m = 100 \) and \( n = 3 \) took about 1.5 minutes on a SUN UltraSparc station, and about 4 minutes when \( n = 8 \). The estimators considered in the simulation study included the (artificial) estimates based on the true \( X \)'s, naive CPQL, and SIMEX/CPQL. For the SIMEX estimates, we set \( B = 100 \) and used quadratic extrapolations for all parameters (SIMEX-Q). The results are displayed in Tables 1 and 2.

We first comment on the homogeneous case (\( \sigma_{\mu}^2 = 0 \)). These results are largely consistent with our theory. The estimates of \( \beta_x \) and \( \beta_2 \) reflect attenuation and are reasonably well corrected by SIMEX/CPQL. The theoretical value of \( \beta_{\text{true}, \theta} - \beta_2 / \tau \) is close to 0.9 (\( \tau \) defined in Sec. 3.2), which is less biased than the estimate of \( \beta_2 \). As expected, there is a bias-variance trade-off, so that in estimating \( \beta_x \), SIMEX/CPQL is less biased but more variable than the naive estimate, which ignores measurement error. A similar phenomenon occurs for estimating the variance component \( \theta \). The only


Table 1. Simulation of Logistic Regression in the Homogeneous Case; that is, the Between-Cluster Variance is σ_{Zu}^2 = 0

<table>
<thead>
<tr>
<th>Cluster size</th>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>SE</th>
<th>MSE</th>
<th>Mean</th>
<th>SE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 3</td>
<td>β₂ = 2</td>
<td>TRUE</td>
<td>2.168</td>
<td>.310</td>
<td>2.093</td>
<td>.313</td>
<td>.106</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>1.516</td>
<td>.356</td>
<td>1.470</td>
<td>.217</td>
<td>.328</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>2.100</td>
<td>.411</td>
<td>2.010</td>
<td>.335</td>
<td>.125</td>
<td></td>
</tr>
<tr>
<td>n = 3</td>
<td>β₂ = 1</td>
<td>TRUE</td>
<td>1.088</td>
<td>.120</td>
<td>1.051</td>
<td>.219</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>.944</td>
<td>.033</td>
<td>.918</td>
<td>.168</td>
<td>.042</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>1.074</td>
<td>.136</td>
<td>1.036</td>
<td>.232</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td>n = 3</td>
<td>θ = 0.5</td>
<td>TRUE</td>
<td>.528</td>
<td>.314</td>
<td>.500</td>
<td>.290</td>
<td>.152</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>.438</td>
<td>.229</td>
<td>.407</td>
<td>.335</td>
<td>.121</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>.570</td>
<td>.337</td>
<td>.532</td>
<td>.428</td>
<td>.176</td>
<td></td>
</tr>
<tr>
<td>n = 8</td>
<td>β₂ = 2</td>
<td>TRUE</td>
<td>2.090</td>
<td>.074</td>
<td>2.066</td>
<td>.185</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>1.461</td>
<td>.322</td>
<td>1.447</td>
<td>.311</td>
<td>.322</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>1.567</td>
<td>.084</td>
<td>1.559</td>
<td>.210</td>
<td>.046</td>
<td></td>
</tr>
<tr>
<td>n = 8</td>
<td>β₂ = 1</td>
<td>TRUE</td>
<td>1.039</td>
<td>.036</td>
<td>1.033</td>
<td>.130</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>.907</td>
<td>.035</td>
<td>.902</td>
<td>.113</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>1.018</td>
<td>.039</td>
<td>1.010</td>
<td>.136</td>
<td>.019</td>
<td></td>
</tr>
<tr>
<td>n = 8</td>
<td>θ = 0.5</td>
<td>TRUE</td>
<td>.465</td>
<td>.083</td>
<td>.456</td>
<td>.198</td>
<td>.041</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAIVE</td>
<td>.369</td>
<td>.073</td>
<td>.359</td>
<td>.168</td>
<td>.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIMEX-Q</td>
<td>.466</td>
<td>.087</td>
<td>.451</td>
<td>.207</td>
<td>.045</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Here n refers to the number of observations per cluster, β₂ = 0, β₂ = 3, β₂ = 1, and θ = 5. The measurement error variance is σ_{Zu}^2 = . The Z's are generated as standard normal and random variables.

A logistic mixed model with random intercept was fit. Analysis of this dataset using our C program took about 4 minutes. The objective is to study the association between the risk of LVH and SBP after adjusting for the other covariates.

Initial analysis of the observed blood pressures themselves, or their residuals after regressing on Z, shows strong evidence in favor of the heterogeneous model, with approximately 3/4 of the observed variability due to cluster-to-cluster variation. Thus even in the absence of measurement error, we would conclude that σ_{Zu}^2 ≈ (1/2)σ_{Zu}^2.

Because blood pressures are obtained only every 2 years, and it is entirely possible that a person's SBP changes over time, there is no direct estimate of σ_{Zu}^2; this would require that SBP be obtained over a number of days within a relatively shorter period. To see this, consider the following argument. Suppose that within a cluster, X, given Z, follows a normal linear model with mean η_{W,1} + Z_νθ_ν and covariance matrix σ_{Zu}^2 I + σ_{Zu}^2 J. Then W_i given Z_i follows a normal linear model with the same mean but with covariance matrix (σ_{Zu}^2 + σ_{Zu}^2) I + σ_{Zu}^2 J. Thus the observed Framingham (W, Z) data alone can identify only the sum of σ_{Zu}^2 and σ_{Zu}^2, but not either component separately. One means of identification is to fix a value of σ_{Zu}^2, (e.g., σ_{Zu}^2 = 0), which assumes that latent blood pressure does not vary over the course of the study. Alternatively, identification is possible only from the assumed model for Y given (X, Z) and (outside the linear GLMM) from distributional assumptions concerning X.

For illustrative purposes, we vary the measurement error variance between the two extremes σ_{Zu}^2 = 0 and σ_{Zu}^2 = 0, using the method-of-moments estimator of their sum (.016) to identify σ_{Zu}^2 exactly. In this illustration σ_{Zu}^2 is treated as fixed and known, and we thus used the standard error estimation methods of Stefanski and Cook (1995).
We expect from our theory that as $\sigma^2_\epsilon$ increases, the measurement error-corrected estimate of $\beta_2$ will increase and the estimate of $\hat{\theta}$ will decrease. The results confirm this. The estimate of $\beta_2$ increased from 2.79 with a standard error of 1.44 ($p$ value = .052) when $\sigma^2_\epsilon = 0$ to 3.92 with a standard error of 1.73 ($p$ value = .023) when $\sigma^2_\epsilon = .016$. This suggests statistically significant effect of SBP on LVH. Higher SBP is associated with a higher risk of LVH. The evidence becomes stronger when the measurement error is taken into account.

The estimate of $\hat{\theta}$ decreased from 2.05 with a standard error of 1.57 when $\sigma^2_\epsilon = 0$ to 1.85 with a standard error of 1.52 when $\sigma^2_\epsilon = .016$. We note that the nominal standard error of $\hat{\theta}$ cannot be used directly for testing $\theta = 0$, because the null hypothesis is on the boundary of the parameter space and the Wald statistic is not asymptotically distributed in a chi-square (Lin 1997). A SIMEX/score test for the variance component developed in an earlier version of this article (Wang, Lin, Gutierrez and Carroll 1997) indicates strong evidence for a nonzero variance component. Specifically, the $p$ value of the score test increased from .0001 when $\sigma^2_\epsilon = 0$ to .006 when $\sigma^2_\epsilon = .016$. Figure 3, discussed in Section 6.2, illustrates the SIMEX/CPQL extrapolations of $\beta_2$ and $\hat{\theta}$ when $\sigma^2_\epsilon = 0$.

The conclusions for the rest of the regression coefficients stay the same with or without taking the measurement errors into account; namely, except for the intercept, none is significant at .05 level. The SIMEX/CPQL and the naive CPQL estimates (standard errors) of the regression coefficients for intercept, age, smoking status, body mass index, and the exam number are $-23.92(8.45)$, $03(0.07)$, $-.75(0.73)$, $0.02(0.11)$, and $0.33(0.25)$ and $-17.76(6.88)$, $0.03(0.06)$, $-0.60(0.67)$, $0.05(0.11)$, and $0.33(0.25)$.

### 9. CONCLUDING REMARKS

In this article, we have shown that the effect of ignoring measurement error in GLMMEs can be to bias regression coefficient and variance component estimators. The reason is that even under normality assumptions, the observed data follow a GLMM but with the structure of the fixed effects and random effects being misspecified. For example, typically the observed data are overdispersed relative to the assumed model ignoring measurement error (Sec. 2).

We have been able to compute the biases in a number of cases (Secs. 3–5). Typically, the broad direction of the biases in regression coefficient estimators is similar to that in ordinary generalized linear models (GLIM), whereas the direction of bias when estimating the variance component varies from case to case. Our results show that there is an important effect of the within-cluster sample size on the biases, and indeed some of the worst biases may occur when such sample sizes are the largest. We note that the techniques that we provided in bias calculations apply to more complicated cases.

The measurement error literature makes a distinction between functional and structural modeling. The former makes no assumptions about the distribution of the unobservables, and the latter typically makes distributional assumptions. The appeal of functional modeling is one of model robustness. We showed in Section 5 that in GLMMEs there is an additional component of concern in structural modeling with respect to model robustness—namely, the assumed covariance structure of the unobservable predictors within a cluster.

The functional estimator that we considered (Sec. 6.2) is the SIMEX/CPQL method, largely because it has the potential to estimate parameters with minimal bias. Because CPQL is often fast even when the random-effects structures become more complicated, with respect to computational
concern we expect that using SIMEX/CPQL to fit the general GLMMeM in (1) would still be practical. Using the general theory of Carroll et al. (1995) and Stefanski and Cook (1995), one can show that the SIMEX approach can be applied to the cases with dependent measurement errors and multivariate X_{ij}.

In our example (Sec. 8), we noted a difficulty with the functional approach in GLMMs—namely, that without careful attention to experiment design, the structure of the measurement error (additivity, covariance) may not be identifiable from the observed covariates themselves. Thus in our example of logistic regression, estimating the measurement error covariance with any precision would require a structural approach; in the linear GLMM, only the covariance structure of the unobserved covariate must be specified. Clearly, if one is to consider the issue of measurement error in GLMMs, one must also pay attention to the design, and allow for estimation of the error covariance structure.

The appeal of functional methods in general, and the SIMEX procedure in this article in particular, is robustness against modeling the structure of the latent unobservables covariates X. But the formidable appeal of functionality and the ease of implementation of the SIMEX procedure must be balanced by the potential loss of information and efficiency incurred relative to a correctly modeled structural analysis. Based on some preliminary calculations, we conjecture that the loss of efficiency in functional estimation may not be very severe in many problems for estimating the regression parameters (β0, β1, β2), but can in some cases be considerable for estimating variance components. Even in the ordinary GLMM, the tradeoff between model robustness and efficiency is a subject under vigorous development, with many attempts to model the latent covariates flexibly. In the GLMMeM context, with its more complex structure involving both distributions and covariances of the latent variables, such flexible modeling is a challenging problem that clearly deserves attention.

**APPENDIX: DETAILED BIAS CALCULATIONS**

A.1 Derivation of Equation (9)

Under the heterogeneous X model, we have A_t = (σ_u^2 I + σ_{w_2}^2 J)((σ_u^2 + σ_{w_2}^2 I + σ_{w_2}^2 J)^{-1}) and e_{A,t} = o_t I + e_t. Using the equality (σ + b)^-1 = 1/(1 - b(a + bm)^-1) for any constants a and b, we have A_t = σ_u^2 (1 + μ_t)^-1(1 - μ_t)^-1 and X_t = λW_t + (1 - λ)(1 - μ_t)^-1(1 - μ_t)^-1 1. The 1st component of b_t is b_{1,t} = X_t - E(X_t|W_t) = b_{2,t}^-1 + b_{3,t}^-1 + b_{4,t}^-1 = λ(1 - λ)α_0 - (1 - λ)μ_t(1 - (1 - λ)μ_t) and the 1st component of b_{2,t} is b_{2,t}^-1 = (1 - λ)μ_t. Note that b_{3,t}^-1 and b_{4,t}^-1 are independent and are independent of W_t and b_t. Define b_t = b_t + β_2 t b_t^-1 b_t = b_t b_t^-1 b_t = b_t b_t^-1 b_t = b_t b_t^-1 b_t = b_t b_t^-1. We then have Equation (9).

A.2 Bias Derivations in the Homogeneous Poisson GLMMeM

The naive probit model is Φ⁻¹(μ_{niw}, w) = β_0 + β_1 W_{ij} + b_i and the Y_i are binary with conditional means μ_{niw} and conditional variances σ_{w_2}^2 μ_{niw}(1 - μ_{niw}). The naive probit model is Φ⁻¹(μ_{niw}, w) = β_0 + β_1 W_{ij} + b_i and b_i = b_i. To integrate out b_i, one can use the identity f(Φ(a + b) dΦ(b/γ) = Φ(a(1 + γ) - b/γ)) for any constants a and γ. It follows that E(Y_i|W_{ij}, b_i) satisfies Φ⁻¹(μ_{niw}, w) = (1 + γ)⁻¹/2(β_0 + β_1 W_{ij} + b_i). Hence we have the bias results in Section 3.2.

A.3 Bias Derivations in the Homogeneous Poisson GLMMeM

Let ζ = θ/2 + β_0 and ζ* = θ/2 + β_1. The marginal means of (Y_i|X_{ij}) and (Y_i|W_{ij}) are d_0 = exp(ζ + d_0 X_{ij}) and c_0 = exp(ζ + d_0 W_{ij}). This slight reparameterization allows θ to be expressed in the conditional variances and covariances of (Y_i|X_{ij}) and (Y_i|W_{ij}) and not in the conditional means. Simple calculations using the correspondence between the conditional means of the (X|Y) and the (Y|W) models show that the naive estimators of (ζ, θ) converge to (ζ*, θ*).

Given (ζ*, θ*), if the X's are observable, then the sufficient statistics for θ are (Y_i, X_{ij}, X_{ij}). Denoting η = exp(θ/2) - 1, the asymptotic limit of the MLE of ζ satisfies ζ = E(Y_i - d_0 Y_i - c_0 Y_i - c_0 W_{ij})/E(d_0 Y_i). The asymptotic limit of the naive estimator θ naive satisfies θ naive = E(Y_i - c_0 Y_i - c_0 W_{ij})/E(c_0 Y_i), where θ naive = exp(θ naive) - 1, and the expectation is taken with respect to (Y_i, W_{ij}) under the homogeneous Poisson model. Some calculations show that θ naive = exp(θ naive) - 1 = exp(θ) - 1 + (n - 1)((n - 1) + exp(ξ) - 1). Equation (12) follows immediately.

A.4 Bias Derivations in the Heterogeneous Linear GLMMeM

Let β = (β_0, β_1, β_2)^T and X = (1, X). Equation (14) can be written as E(1 X^T V^{-1} Υ) = θ_n = E(1 X^T V^{-1} Υ) β. Applying (16) to both sides of this equation, some algebra yields θ naive = θ naive and λ_2 given in (18). To solve (15) for θ naive and σ naive, we first calculate the mean and covariance of T = Y - 1W naive as μ_T = E(T) = 0 and V_T = cov(T)

V_T = (σ_u^2 + 2σ_{w_2}^2 σ_{w_2}^2 + σ_{w_2}^2 I) + (θ + 1 - λ)σ_u^2 I = (σ_u^2 + 2σ_{w_2}^2 σ_{w_2}^2 + σ_{w_2}^2 I) + (θ + 1 - λ)σ_u^2 I.

Using the equalities θ naive = (θ naive, θ naive) = I and (16), one can show that (15) is equivalent to tr(V^{-1} J V^{-1} T V^{-1}) = tr(V^{-1} J) and tr(V^{-1} V^{-1} T V^{-1}) = tr(V^{-1}). Because V and V_T have the same matrix structure, we have V = V_T. Equivalently, θ naive and σ naive satisfy (17).

A.5 Numerical Bias Calculations of the Naive Estimators

Denote the log-likelihood by l = log L. The naive estimators θ naive = (θ naive, θ naive) and θ naive maximize m_i=1 log L(Y_i | W_i, Z_i, θ naive) where Z naive = Z naive Y_i | W_i, Z_i, θ naive) takes the same form as (2) but X is replaced by W_i. Thus, the probability limit of the naive estimators is (θ naive, θ naive) maximizes E_θ naive(Y_i | W_i, Z_i, θ naive, θ naive) = m_i=1 log L(Y_i | W_i, Z_i, θ naive, θ naive) as m_i→∞, where the expectation is taken with respect to (Y_i, W_i, Z_i) conditional on Z_i. For simplicity, we remove the subscript i in the ensuing discussion. Using the identity E(Z) = E(E(Z|X)) and the independence of Y and W given X, Z, we first calculate

B_{i=1}^M \delta naive(Y_i, W_i, Z_i, θ naive, θ naive) = \int \delta naive(Y_i, X_i, θ naive, θ naive) L(Y_i, X_i, Z_i, θ naive, θ naive) dµ(X_i).
where $\nu(Y)$ denotes an appropriate probability measure of $Y$
and
\[ p_{\text{naive}}(Y, X, Z; \beta, \theta) = \int p_{\text{naive}}(Y, X + U, Z; \beta, \theta) L(U) dU. \] (A.2)

In the heterogeneous logistic GLMM fixed in Section 4.2, we have
\[ p_{\text{naive}}(Y, X, Z; \beta, \theta) = \int \prod_{j=1}^{n} \left( \sum_{i=1}^{m} \mu_{y_{ij}, u_i}(b) \phi(b) \right) d\Phi(u_1) \ldots d\Phi(u_m), \]
where $\mu_{y_{ij}, u_i}(b) = \{h_{u_i}(b)^{y_{ij}} \{1 - h_{u_i}(b)\}^{1-y_{ij}}, h_{u_i}(b) = H(\beta_0 + \beta_x X_j + \beta_u u_i + \theta_0^2 b), \text{ and } H(u) = 1 + \exp(-u)^{-1}.$

We next need to further take expectation of (A.1) with respect to $X$ conditional on $Z$. For the heterogeneous $X$ model considered in Section 4.2, $X$ follows $N(\mu_{X1}, V_x)$, with $V_x = \sigma_x^2 I + \sigma_{x1} J$. Let $T = A^{-1}(X - \mu_{x1})$, where $A$ is a lower triangular matrix satisfying $ATA = V_x$ obtained using the Cholesky decomposition of $V_x$. Denoting (A.1) by $E(X, Z; \beta, \theta)$, we have $E(X, Z; \beta, \theta) = E(f(X, Z; \beta, \theta)) = E \{f(X, Z; \beta, \theta)\} = E \{f(X, Z; \beta, \theta)\}$. Evaluation of required integration can be carried out by repeatedly using 20-point Gauss–Hermite quadrature for small $n$. Monte Carlo simulations can be used for large $n$. Using the change-of-variable technique for numerical integration has been discussed in detail by Davidian and Giltinan (1995, chap. 7). An optimization routine was used for maximization. This numerical technique can be applied to accommodate the cases with nonadditive and/or dependent measurement errors provided that the conditional distribution of $(W|X)$ is normal.

The foregoing procedures can be used to calculate the biases of the heterogeneous estimators discussed in Section 5. Specifically, we can replace $L_{\text{naive}}(Y, X, Z; \beta, \theta)$ in (A.1) and (A.2) by $p_{\text{naive}}(Y, X, Z; \beta, \theta)$, which is defined as $L(Y, X|Z)$ in (4) with $L(X|Z)$ being normal with mean $\mu_{x1}$ and covariance $\sigma_x^2 I$. We then maximize the resulting expectation with respect to $\beta$, $\theta$, $\mu_{x1}$, and $\sigma_x^2 I$. The maximizers are $\hat{\beta}_{\text{hom}}$, $\hat{\theta}_{\text{hom}}$, $\hat{\mu}_{x1}$, and $\hat{\sigma}_x^2$. A.6 The CPQL Method

A popular approximate inference procedure in the GLMM without measurement error (I) is the penalized quasi-likelihood (PQL) method of Schall (1991) and Breslow and Clayton (1993). Denote the right side of equation (1) by $p_{\text{naive}}(Y, X, Z; \beta, \theta)\text{ in (A.1) and (A.2)}$ by $L_{\text{naive}}(Y, X, Z; \beta, \theta)$, which is defined as $L(Y, X|Z)$ in (4) with $L(X|Z)$ being normal with mean $\mu_{x1}$ and covariance $\sigma_x^2 I$. We then maximize the resulting expectation with respect to $\beta$, $\theta$, $\mu_{x1}$, and $\sigma_x^2 I$. The maximizers are $\hat{\beta}_{\text{hom}}$, $\hat{\theta}_{\text{hom}}$, $\hat{\mu}_{x1}$, and $\hat{\sigma}_x^2$.

\[ y_{ij} = \beta_0 + X_{ij}^T \beta_x + Z_{ij}^T \beta + A_{ij}^T b_i + \epsilon_{ij}, \]
where the random effects $b_i$ follow $N(0, \Sigma(b))$. The key feature of PQL is that it can be easily implemented by iteratively fitting a linear mixed model to a modified dependent variable $y_{ij} = y_{ij} + g(\mu_{x1}(Y_{ij} - \mu_{x1}))$ as

\[ y_{ij} = \beta_0 + X_{ij}^T \beta_x + Z_{ij}^T \beta + A_{ij}^T b_i + \epsilon_{ij}, \]
where the random effects $b_i$ follow $N(0, \Sigma(b))$. The key feature of PQL is that it can be easily implemented by iteratively fitting a linear mixed model to a modified dependent variable $y_{ij} = y_{ij} + g(\mu_{x1}(Y_{ij} - \mu_{x1}))$ as