Functional Inference in Frailty Measurement Error Models for Clustered Survival Data Using the SIMEX Approach

Yi Li and Xihong Lin

We consider frailty models for clustered survival data in the presence of measurement errors in covariates. We first show that when the measurement error is accounted for in a full likelihood analysis but the distribution of the unobserved covariate is misspecified, the maximum likelihood estimators are asymptotically biased, especially for the variance component, whose bias can be substantial. We then discuss making inference using functional estimation via the SIMEX method where no distribution of the unobserved error-prone covariate is assumed. The SIMEX method is easy to implement by repeatedly fitting standard frailty models. We study the asymptotic properties of the SIMEX estimates and show that they are consistent and asymptotically normal. In simulation studies, we compare the SIMEX method and the likelihood method in terms of efficiency and robustness. We also propose a SIMEX score test for the variance component to test for the within-cluster correlation and evaluate its performance through simulation studies. The SIMEX variance component score test does not require specifying distributions for the random effect and the unobserved error-prone covariate, and is easy to implement by repeatedly fitting standard Cox models. The proposed methods are illustrated using the Kenya parasitemia data.

KEY WORDS: Asymptotic bias; Clustered survival data; Efficiency; Frailty model; Martingale residual; Measurement error; Robustness; Score test; SIMEX; Variance component.

1. INTRODUCTION

Many failure time regression applications involve covariates that are measured with error. For example, in nutritional studies, dietary intake (e.g., fat intake) is often measured based on a 24-hour food recall or a food frequency questionnaire and typically involves considerable noise (Carroll, Ruppert, and Stefanski 1995); in cardiovascular research, blood pressure is often subject to considerable hourly and daily variation and long-term blood pressure is difficult to ascertain (Carroll et al. 1995); and in AIDs studies, CD4 counts are often measured with a substantial amount of variability (Tsiatis, De Gruttola, and Wulfsohn 1995). For independent data under the Cox model, several authors (Prentice 1982; Hughes 1993; Zhou and Pepe 1995; Hu, Tsiatis, and Davidian 1998) have considered modeling measurement errors in covariates. In a similar context, Lin and Ying (1993) and Paik and Tsai (1997) have discussed the missing covariate problem.

Clustered failure time data arise in many contexts. For example, in familial studies, age at onset of a disease is recorded for multiple members of the same family, and in multicenter clinical trials, failure times are observed for multiple patients in each center. Frailty models provide a convenient framework for modeling intracluster correlation by assuming the dependence is induced by a shared frailty or a random effect (Clayton and Cuzick 1985; Hougaard 1986; Oakes 1989). An EM algorithm has been proposed to fit frailty models (Gill, Andersen, and Sorensen 1992), and the theoretical properties of frailty models have been studied in gamma frailty models (Murphy 1995; Parner 1998) and in general (e.g., gamma, normal, and inverse-Gaussian) frailty models (Kosorok, Lee, and Fine 2001). Approximate likelihood approaches (e.g., the penalized partial likelihood method) have also been proposed (see, e.g., McGilchrist 1993; Therneau and Grambsch 2000).

Limited work has been done on clustered failure time data with measurement errors in covariates. We present a data example in Section 7. The study was a cohort study on parasitemia among children in western Kenya (McElroy et al. 1997) in which 542 children from 309 households were followed over a 22-month period for the time to occurrence of parasitemia, which is an indicator for potential malaria, a disease accompanied by substantial mortality among young African children. The primary interest lies in studying the association between the risk of parasitemia and the exposure to infective mosquito bites after adjusting for other covariates. Because the observed infective mosquito bite exposure is subject to considerable error, the measurement error must be modeled while accounting for the correlation among children within the same household.

Two classes of approaches are often used for inference in the measurement error literature (Carroll et al. 1995): structural modeling and functional modeling. Structural modeling assumes a parametric distribution for the unobserved error-prone covariate X and proceeds through a likelihood analysis. Its validity often requires that the distribution of X be correctly specified. Functional modeling, on the other hand, makes no distributional assumption about the unobserved error-prone covariate X, and hence could be more robust but less efficient than structural modeling.

Li and Lin (2000) proposed frailty measurement error models, which jointly model the measurement error in covariate and the intracluster correlation for clustered failure time data. They showed that when the measurement error is ignored, the regression coefficients are attenuated and the variance component is overestimated asymptotically. They considered

Yi Li is Assistant Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *yili@jimmy.harvard.edu*). Xihong Lin is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109 (E-mail: *xlin@umich.edu*). This work was supported in part by U.S. National Cancer Institute grants R01 CA57253 (Li) and R29 CA76404 (Lin). The authors thank the editor, the associate editor, and two referees for their helpful comments and suggestions, and Peter McElroy for providing the Kenya parasitemia data.

Journal of the American Statistical Association
March 2003, Vol. 98, No. 461, Theory and Methods
DOI 10.1198/016214503388619210

structural modeling in frailty measurement error models by assuming a standard linear mixed model for the unobserved covariate X. Parameter estimation proceeded using a nonparametric maximum likelihood estimator (NPMLE) via the EM algorithm with the baseline hazard unspecified. The theoretical properties of the NPMLE were studied.

A question of substantial interest is the robustness of structural modeling in frailty measurement error models if the measurement error in a likelihood analysis is accounted for but the distribution of the error-prone covariate X is misspecified. In this article, we consider functional modeling in frailty measurement error models. To address the robustness issue of structural modeling, we first study the asymptotic bias in full likelihood estimation when the distribution of X is misspecified. Our asymptotic bias results show that the MLEs of the model parameters (especially the MLE of the variance component) could be subject to serious bias in structural modeling when the distribution of X is misspecified. We then discuss functional estimation in frailty measurement error models using the SIMEX method (Stefanski and Cook 1995; Wang, Lin, Gutierrez, and Carroll 1998), where no distribution of X is assumed. Compared to the NPMLE, the SIMEX method is easier to implement by repeatedly fitting standard no–measurement error frailty models (e.g., using S-PLUS.) We study the asymptotic properties of the SIMEX estimates and show that they are consistent and asymptotically normal. In simulation studies, we compare the SIMEX and the NPMLE methods in terms of efficiency and robustness.

A common problem in frailty models is to test for within-cluster correlation. Gray (1995) and Commenges and Andersen (1995) proposed such tests using score statistics for the variance component in frailty models when no measurement error is present. We extend these results to frailty measurement error models and propose a SIMEX variance component score test in the spirit of Lin and Carroll (1999) in generalized linear mixed measurement error models. A key feature of the SIMEX score test is its double robustness in the sense that no assumption need be made on the distributions of the frailty and the unobserved covariate X. It is easy to implement by repeatedly fitting standard Cox models using existing statistical software. We evaluate its performance through simulations.

The rest of the article is structured as follows. We state the frailty measurement error model (FMEM) in Section 2. In Section 3 we study the asymptotic biases in the MLEs using structural modeling when the distribution of the unobserved errorprone covariate X is misspecified. In Section 4 we propose functional estimation in FMEMs using the SIMEX method, study the asymptotic properties of the SIMEX, and propose the SIMEX score test for the variance component. We evaluate the performance of the SIMEX methods through simulations in Section 5 and apply the proposed methods to the Kenya parasitemia data in Section 6. We give discussions in Section 7.

2. THE FRAILTY MEASUREMENT ERROR MODEL

We present the FMEM within the framework of counting process. Denote by $T_{ij} = D_{ij} \wedge P_{ij} = min(D_{ij}, P_{ij})$ the observed survival time subject to right censoring, where D_{ij} is the failure time and P_{ij} is the censoring time, and by $\delta_{ij} = I(D_{ij} \le P_{ij})$ the censoring indicator for the jth subject ($j = 1, \ldots, n_i$) in the ith cluster ($i = 1, \ldots, m$), where $I(\cdot)$ is an indicator function. We assume that censoring is noninformative, that is, the P_{ij} are independent of the D_{ij} and the unobserved covariates X_{ij}, and the distribution of P_{ij} does not involve the model parameters $\mathbf{\Theta}$ (defined later). Let $Y_{ij}(t) = I(T_{ij} \ge t)$ be the at-risk process and $N_{ij}(t) = I(T_{ij} \le t, \delta_{ij} = 1)$ be the counting process. Conditional on the true unobserved error-prone covariate X_{ij}, other accurately measured covariates $\mathbf{Z}_{ij}(q \times 1)$, and a cluster-specific frailty b_i, the intensity for the counting process N_{ij} is

$$\lambda_{ij}(t; X_{ij}, \mathbf{Z}_{ij}, b_i) = Y_{ij}(t)\lambda_0(t)e^{X_{ij}\beta_x + \mathbf{Z}'_{ij}\boldsymbol{\beta}_z + b_i}, \tag{1}$$

where b_i follows some parametric distribution $F(\cdot; \theta)$, θ is a variance component, $(\beta_x, \boldsymbol{\beta}_z)$ are regression coefficients, and $\lambda_0(t)$ is an unspecified baseline hazard. Common choices of $F(\cdot; \theta)$ include log-gamma, normal and log inverse Gaussian (Clayton and Cuzick 1985; Hougaard 1986); for example, assume that if $F(\cdot)$ is log-gamma, then $\exp(b_i)$ has mean 1 and variance θ, and that if $F(\cdot)$ is normal, then $b_i \sim N(0, \theta)$.

Define $\mathbf{T}_i = (T_{i1}, \ldots, T_{in_i})'$, $\mathbf{\Delta}_i = (\delta_{i1}, \ldots, \delta_{in_i})'$, and $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{Y}_i$, and $\mathbf{N}_i$ similarly. Further, define $\mathbf{Y}' = (\mathbf{Y}'_1, \ldots, \mathbf{Y}'_m)$ and $\mathbf{N}$, $\mathbf{X}$, and $\mathbf{Z}$ similarly. Denote the unknown parameter vector by $\mathbf{\Theta} = \{\Lambda_0(\cdot), \beta_x, \boldsymbol{\beta}'_z, \theta\}$, where $\Lambda_0(t)$ is the integrated baseline hazard. Conditional on $(\mathbf{X}, \mathbf{Z})$, the log-likelihood function over a finite time interval $[0, \tau]$ ($\tau < \infty$) is

$$\begin{aligned}\ell_m(\mathbf{N}, \mathbf{Y}; \mathbf{X}, \mathbf{Z}, \mathbf{\Theta}) &= \sum_{i=1}^{m} \ell(\mathbf{N}_i, \mathbf{Y}_i; \mathbf{X}_i, \mathbf{Z}_i, \mathbf{\Theta}) \\ &= \sum_{i=1}^{m} \ln \int \left\{ \prod_{j=1}^{n_i} \prod_{t \le \tau} Y_{ij}^{\beta,\theta}(t)\lambda_0(t) \right\}^{\Delta N_{ij}(t)} \\ &\quad \times e^{-\sum_{j=1}^{n_i} \int_0^\tau Y_{ij}^{\beta,\theta}(t) d\Lambda_0(t)} \, dF(b_i; \theta), \end{aligned} \tag{2}$$

where $Y_{ij}^{\beta,\theta}(t) = Y_{ij}(t)\exp(X_{ij}\beta_x + \mathbf{Z}'_{ij}\boldsymbol{\beta}_z + b_i)$ and τ is a constant and usually is the study duration.

The MLE of $\mathbf{\Theta}$ does not exist for the likelihood function defined earlier when $\Lambda_0(t)$ is a continuous function. Hence we expand the parameter space of $\Lambda_0(t)$ to include discrete functions by replacing $\lambda_0(t)$ with $\Delta\Lambda_0(t)$ in (2) (Murphy 1995; Parner 1998; Kosorok et al. 2001). This gives

$$\begin{aligned}\ell_m(\mathbf{N}, \mathbf{Y}; \mathbf{X}, \mathbf{Z}, \mathbf{\Theta}) &= \sum_{i=1}^{m} \ln \int \left\{ \prod_{j=1}^{n_i} \prod_{t \le \tau} Y_{ij}^{\beta,\theta}(t)\Delta\Lambda_0(t) \right\}^{\Delta N_{ij}(t)} \\ &\quad \times e^{-\sum_{j=1}^{n_i} \int_0^\tau Y_{ij}^{\beta,\theta}(t) d\Lambda_0(t)} \, dF(b_i; \theta). \end{aligned} \tag{3}$$

Denote by W_{ij} the observed X_{ij}-related covariate. The FMEM is completed by assuming an additive measurement error model relating W_{ij} and X_{ij},

$$W_{ij} = X_{ij} + U_{ij}, \tag{4}$$

where the U_{ij} are independent of the X_{ij} and the P_{ij} and are independent following $N(0, \sigma_u^2)$, and σ_u^2 is the measurement error variance. Define $\mathbf{W}_i$ and $\mathbf{W}$ similarly to $\mathbf{X}_i$ and $\mathbf{X}$.

We assume that the measurement error is nondifferential, that is, $L(\mathbf{N}_i, \mathbf{Y}_i; \mathbf{X}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\Theta}) = L(\mathbf{N}_i, \mathbf{Y}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\Theta})$, where $L(\cdot) = \exp\{\ell(\cdot)\}$. This implies that conditional on the true covariates $(\mathbf{X}_i, \mathbf{Z}_i)$, the observed covariate $\mathbf{W}_i$ does not contain additional information about $(\mathbf{N}_i, \mathbf{Y}_i)$. The likelihood of the observed data $(\mathbf{N}_i, \mathbf{Y}_i, \mathbf{W}_i)$ is

$$L(\mathbf{N}_i, \mathbf{Y}_i, \mathbf{W}_i; \mathbf{Z}_i) = \int L(\mathbf{N}_i, \mathbf{Y}_i; \mathbf{X}_i, \mathbf{Z}_i) L(\mathbf{W}_i; \mathbf{X}_i) \times L(\mathbf{X}_i; \mathbf{Z}_i)\, d\mathbf{X}_i, \quad (5)$$

where the likelihood $L(\mathbf{W}_i; \mathbf{X}_i)$ is under (4) and $L(\mathbf{X}_i; \mathbf{Z}_i)$ is the likelihood function of $\mathbf{X}_i$ conditional on $\mathbf{Z}_i$, whose specification concerns model robustness and distinguishes structural modeling and functional modeling. Li and Lin (2000) assumed a linear mixed model for $\mathbf{X}_i$ in their structural modeling to allow possible correlation of the X_{ij} within the same cluster. They proposed using a nonparametric maximum likelihood estimator (NPMLE) to estimate the model parameters with the baseline hazard $\lambda_0(t)$ unspecified. In this article, to make inference more robust, we do not assume a distribution for $\mathbf{X}_i$, and we proceed with functional estimation. We motivate the need for this approach by conducting (in the next section) an asymptotic bias analysis in MLE when the distribution of X is misspecified.

Before proceeding to our asymptotic bias analysis, we remark that we have made no distributional assumption about the censoring process except that it is noninformative. Under this noninformative censoring assumption, the inference based on the likelihood (5) is valid for any distribution of the censoring times. In other words, as pointed out by a referee, we allow the censoring times to be correlated within each cluster. To see this, write the joint likelihood of the failure time process and the censoring process as $\widetilde{L}(\mathbf{N}, \mathbf{Y}, \mathbf{W}) = L(\mathbf{N}, \mathbf{Y}, \mathbf{W}; \mathbf{Z}, \boldsymbol{\Theta}) L_c(\mathbf{N}, \mathbf{Y})$, where $L_c(\mathbf{N}, \mathbf{Y})$ is the likelihood involving the censoring process. Specifically, $L_c(\mathbf{N}, \mathbf{Y})$ involves the joint density of P_{ij}, which might be correlated, but under the assumption of noninformative censoring, $L_c(\mathbf{N}, \mathbf{Y})$ does not depend on $\boldsymbol{\Theta}$, and maximizing $\widetilde{L}(\mathbf{N}, \mathbf{Y}, \mathbf{W})$ with respect to $\boldsymbol{\Theta}$ is equivalent to maximizing $L(\mathbf{N}, \mathbf{Y}, \mathbf{W}; \mathbf{Z}, \boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta}$.

3. ASYMPTOTIC BIAS IN MAXIMUM LIKELIHOOD ESTIMATORS WHEN THE DISTRIBUTION OF X IS MISSPECIFIED

Examination of the likelihood function (5) suggests that a full likelihood analysis requires specification of the distribution of X, which could be difficult and sometimes may not be desirable. A question of substantial interest is what will happen if one accounts for measurement error in a likelihood analysis but incorrectly specifies the distribution of X. In this section we study the asymptotic biases in the MLEs when the distribution of X is misspecified.

We assume in our bias analysis that the cluster size is $n_i = n < \infty$ and that the number of clusters is $m \to \infty$. We consider a simple random intercept frailty model (Clayton and Cuzick 1985),

$$\lambda_{ij}(t; X_{ij}, b_i) = \lambda_0(t) e^{\beta_x X_{ij} + b_i}, \quad (6)$$

where $b_i \sim N(0, \theta)$. We consider the asymptotic biases in the MLEs of β_x and θ when the covariance of $\mathbf{X}_i$ is misspecified. Suppose that the true X model is a random intercept model,

$$X_{ij} = \mu_x + a_i + \epsilon_{ij}, \quad (7)$$

where $a_i \sim N(0, \sigma^2_{x\mu})$ and $\epsilon_{ij} \sim N(0, \sigma^2_x)$, and one misspecifies the X model as an independent model,

$$X_{ij} = \mu_x + \epsilon_{ij}, \quad (8)$$

where $\epsilon_{ij} \sim N(0, \sigma^2_x)$. Model (7) reduces to (8) when $\sigma^2_{x\mu} = 0$.

Denote by $\boldsymbol{\Theta} = \{\lambda_0(t), \mu_x, \sigma^2_x, \beta_x, \theta\}$ the unknown parameter vector and by $\boldsymbol{\Theta}_{\text{ind}}$ the asymptotic limit of the MLE of $\boldsymbol{\Theta}$ under the independent X model when the random intercept X model is true. Then $\boldsymbol{\Theta}_{\text{ind}}$ maximizes

$$\begin{aligned} &E\{\ell_{\text{ind}}(\mathbf{T}_i, \boldsymbol{\Delta}_i, \mathbf{W}_i; \boldsymbol{\Theta}_{\text{ind}})\} \\ &= \int \cdots \int \ell_{\text{ind}}(\mathbf{T}_i, \mathbf{X}_i + \sigma_u \mathbf{U}_i; \boldsymbol{\Theta}_{\text{ind}}) \\ &\quad \times \left\{ \prod_{j=1}^{n} f(T_{ij}; X_{ij}, b_i; \boldsymbol{\Theta})\, dT_{ij} d\Phi(U_{ij}; 1) \right\} \\ &\quad \times d\Phi(b_i; \theta) d\Phi(\mathbf{X}_i - \mu_x \mathbf{1}; \sigma_{x\mu} \mathbf{1}\mathbf{1}^T + \sigma^2_x \mathbf{I}), \end{aligned}$$

where $f(\cdot)$ is a density function, $\Phi(\cdot)$ is a normal cumulative distribution function, $\mathbf{1}$ is an $n \times 1$ vector of 1s, $\mathbf{I}$ is an $n \times n$ identity matrix, the expectation is taken with respect to the true likelihood $L(\mathbf{T}_i, \boldsymbol{\Delta}_i, \mathbf{W}_i)$ assuming the random intercept X model (7), and $\ell_{\text{ind}}(\mathbf{T}_i, \boldsymbol{\Delta}_i, \mathbf{W}_i; \boldsymbol{\Theta}_{\text{ind}})$ is the log likelihood assuming the X model is misspecified as the independent model (8). Maximization is with respect to $\boldsymbol{\Theta}_{\text{ind}}$ for fixed $\boldsymbol{\Theta}$. The asymptotic limit $\boldsymbol{\Theta}_{\text{ind}}$ does not have a closed form. We hence study the asymptotic bias numerically. Because $E\{\ell_{\text{ind}}(\mathbf{T}_i, \boldsymbol{\Delta}_i, \mathbf{W}_i; \boldsymbol{\Theta}_{\text{ind}})\}$ involves multiple-dimensional integration, we numerically evaluate it using a combination of the methods of Gauss–Hermite quadrature and Monte Carlo simulations. The Newton–Raphson algorithm is then used for maximization.

For simplicity, we assumed in our numerical calculations that there was no censoring and a constant baseline hazard, that is, $\lambda_0(t) = \lambda_0$. The parameter configuration used in our numerical calculations was cluster size $n = 2$; $\ln(\lambda_0) = -2$, $\beta_x = 2$, $\theta = .5$; and $\mu = 1, \sigma^2_x = 1$. We varied the measurement error variance σ^2_u from 0 to 1. Figure 1 shows the asymptotic relative biases in the MLEs $\beta_{x,\text{ind}}$ and θ_{ind} assuming the independent X model when the random intercept X model is true. The biases are plotted as a function of the measurement error variance σ^2_u $(0 \le \sigma^2_u \le 1)$ for several values of $\sigma^2_{x\mu}$ ranging from 0 to 1. The MLE β_{ind} underestimates β_x, but its bias is small. The MLE θ_{ind}, on the other hand, overestimates θ, and its bias can be substantial, especially when σ^2_u and $\sigma^2_{x\mu}$ are moderate or large. Clearly, when $\sigma^2_{x\mu} = 0$, there is no bias in the MLEs $\beta_{x,\text{ind}}$ and θ_{ind}.

The asymptotic bias analysis suggests that even when one accounts for measurement error in a likelihood analysis, misspecification of the distribution of the unobserved covariate X could result in biased parameter estimation. Hence it is of substantial interest to develop a more robust inference procedure without specifying the distribution of X. We propose one such procedure using functional modeling in the next section.

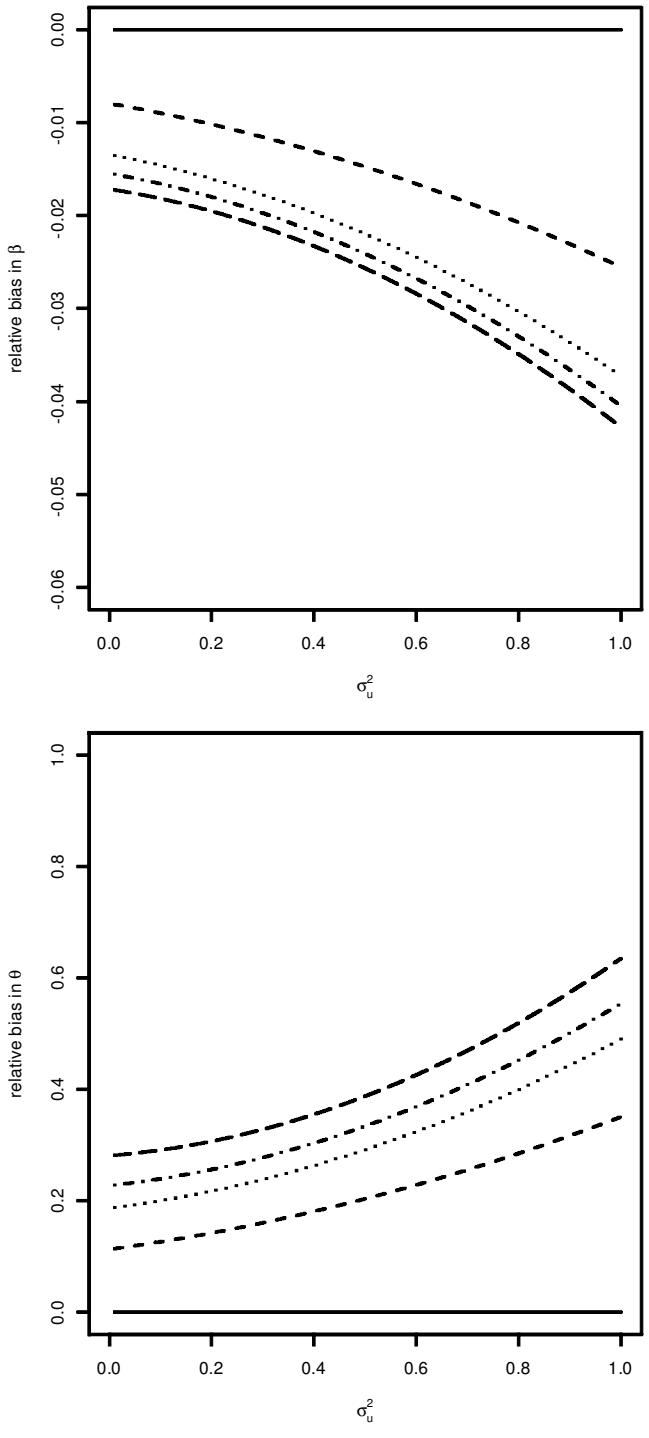

Figure 1. Asymptotic Relative Biases in the MLEs of β_x and θ Assuming the Independent X Model When the Random Intercept X Model is True as Functions of σ_u^2 and $\sigma_{x\mu}^2$. The cluster size $n = 2$. The true parameter values are $\lambda_0 = \exp(-2), \beta_x = 2, \theta = .5, \mu_x = 1$, and $\sigma_x^2 = 1$. The five curves in each plot correspond to —— (the horizontal axis) $\sigma_{x\mu}^2 = 0$; – – – $\sigma_{x\mu}^2 = .25$; - - - $\sigma_{x\mu}^2 = .50$; – - – $\sigma_{x\mu}^2 = .75$; — — — $\sigma_{x\mu}^2 = 1.00$.

4. THE SIMEX APPROACH IN THE FRAILTY MEASUREMENT ERROR MODEL

4.1 The SIMEX Estimation Procedure

The simulation extrapolation (SIMEX) method (Stefanski and Cook 1994) is a simulation-based functional estimation method for measurement error models wherein no distributional assumption is made on the unobserved covariate $\mathbf{X}_i$. We explain the SIMEX procedure using Figure 2, which shows the application of SIMEX to the Kenya parasitemia data. The two parameters of interest are the coefficient of the number of infective bites β_x and the variance component θ. A key feature of the SIMEX is that only standard frailty models in the form of (1) need to be fitted repeatedly, and hence it is easy to implement. For example, the most recent versions of S-PLUS and R (Therneau and Grambsch 2000) can fit frailty models for some common choices of frailty distributions (e.g., normal and log-gamma).

Suppose that the measurement error variance σ_u^2 is known or is estimated as $\hat{\sigma}_u^2$. Denote the parameter vector by $\boldsymbol{\Theta} = \{\Lambda_0(t), \beta_x, \boldsymbol{\beta}_z', \theta\}$. The SIMEX procedure consists of two steps, a simulation step and an extrapolation step. In the simulation step, for a given $\xi > 0$, one generates a large number of simulated datasets. Specifically, for each $c = 1, \ldots, C$, where C is large (e.g., $C = 100$), one generates simulated data, $W_{ij,c}$, by adding to the W_{ij} independent errors with mean 0 and variance $\sqrt{\xi}\sigma_u$ as $W_{ij,c} = W_{ij} + \sqrt{\xi}\sigma_u U_{ij,c}^*$, where the $U_{ij,c}^*$ are generated independently from $N(0, \sigma_u^2)$. Then one fits the naive frailty model by replacing X_{ij} in (1) with $W_{ij,c}$ as if $W_{ij,c}$ measured X_{ij} without error to compute the naive estimate $\widehat{\boldsymbol{\Theta}}_c(\xi)$. We used the EM algorithm of Nielsen et al. (1992) to fit these naive models, because in no–measurement error frailty models this method yields consistent estimates of the model parameters (Parner 1998; Kosorok et al. 2001), which are required for the consistency of the SIMEX estimate in the measurement error case. One next calculates the sample mean of these naive estimates as $\widehat{\boldsymbol{\Theta}}(\xi) = C^{-1}\sum_{c=1}^{C}\widehat{\boldsymbol{\Theta}}_c(\xi)$. One does this for a series of values of ξ, [e.g., $\xi = (0, .5, 1.0, 1.5, 2.0)$] and plots the resulting naive estimates $\widehat{\boldsymbol{\Theta}}(\xi)$ versus ξ. These are shown in small solid squares in Figure 2 for β_x and θ.

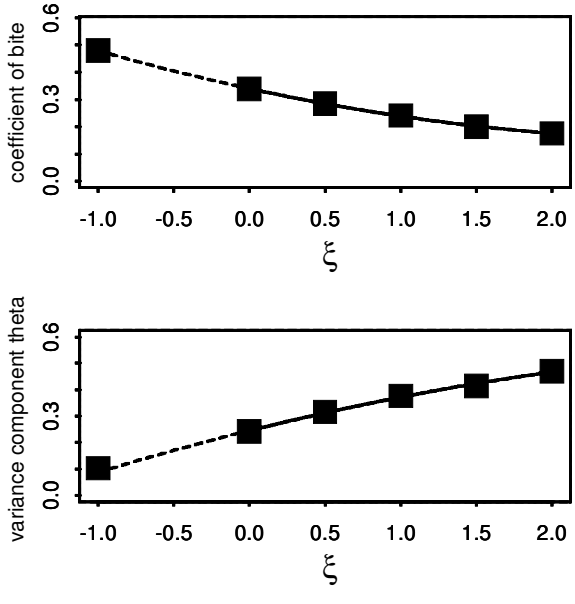

Figure 2. SIMEX Extrapolations for the Regression Coefficient of BITE β_B and the Variance Component θ When $\sigma_u^2 = .20$ in the Kenya Parasitemia Data.

In the extrapolation step, a regression model (e.g., a quadratic model) is fit to these averaged naive estimates $\widehat{\boldsymbol{\Theta}}(\xi)$ as a function of ξ, shown in solid curves in Figure 2 for β_x and θ. Extrapolating it back to $\xi = -1$ (no measurement error), represented by the dashed curves in Figure 2, yields the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$. Consistency and asymptotic normality of the SIMEX estimate are demonstrated in Theorems 1 and 2 in Section 4.2.

The covariance of the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ can be calculated similarly to the method described by Stefanski and Cook (1995). Specifically, let $\widehat{\boldsymbol{\Omega}}(\xi) = \frac{1}{C}\sum_{c=1}^{C} \widehat{\boldsymbol{\Omega}}\{\mathbf{N}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}_c(\xi)\}$, where $\widehat{\boldsymbol{\Omega}}\{\mathbf{N}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}_c(\xi)\}$ denotes the estimated model-based covariance of $\widehat{\boldsymbol{\Theta}}_c(\xi)$ by fitting the naive frailty model to $\mathbf{W}_c(\xi)$. Let $\widehat{\boldsymbol{\Omega}}^*(\xi)$ denote the sample covariance of the $\widehat{\boldsymbol{\Theta}}_c(\xi)$, that is, $\widehat{\boldsymbol{\Omega}}^*(\xi) = (C-1)^{-1}\sum_{c=1}^{C}\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widehat{\boldsymbol{\Theta}}(\xi)\}\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widehat{\boldsymbol{\Theta}}(\xi)\}'$. Next, let $\widehat{\boldsymbol{\Gamma}}(\xi) = \widehat{\boldsymbol{\Omega}}(\xi) - \widehat{\boldsymbol{\Omega}}^*(\xi)$. The covariance of the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ can be calculated by fitting a regression model (e.g., a quadratic model) of $\widehat{\boldsymbol{\Gamma}}(\xi)$ as a function of ξ and extrapolating it back to $\xi = -1$. Technical justification of this variance estimation procedure is given in Theorem 3 in Section 4.2.

4.2 Asymptotic Properties of the SIMEX Estimator

When the outcome variable follows a parametric model, consistency and asymptotic normality of the SIMEX estimate of a finite-dimensional parameter vector have been shown by Carroll, Küchenhoff, Lombard, and Stefanski (1996). The frailty model (1) is a semiparametric model, in which the unspecified cumulative baseline hazard $\Lambda_0(t)$ is an infinite parameter. Hence the unknown parameter vector $\boldsymbol{\Theta}$ is infinite-dimensional. The results of Carroll et al. (1996) cannot be directly applied to this case. The asymptotic properties of the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ are harder to study and rely on the semiparametric theory.

We assume in this section that the regularity conditions (A)–(G) of Kosorok et al. (2001) hold, where conditions (E)–(G) specify assumptions about a general frailty distribution $F(\cdot)$. Under these regularity conditions, we prove consistency and asymptotic normality of the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ assuming a general frailty distribution $F(\cdot)$ in Theorems 1 and 2. These asymptotic results apply to common frailty measurement error models, such as those assuming normal, log-gamma, and log-inverse Gaussian frailties. We provide theoretical justification of the SIMEX variance estimate described in Section 4.1 in Theorem 3.

For simplicity, in our proof we assume that $n_i = n$ and consider the model with only an error-prone covariate X. The proof with $\mathbf{Z}_i$ present is similar with more complicated notation. We also assume that there are no ties in failure times with probability 1. Define H_p to be the product space of bounded variation functions and the real value Euclidean space with norm $\|h\| = \|h_1\|_v + |h_2| + |h_3| < p$, where p is assumed to be finite, $h_1(t)$ is a function of bounded variation on $[0, \tau]$, $h_2, h_3 \in R^1$, and $\|h_1\|_v$ is defined as the sum of the absolute value of $h_1(0)$ and the total variation of $h_1(\cdot)$ on $[0, \tau]$.

When there is no measurement error, under these regularity conditions (A)–(G), Kosorok et al. (2001) showed consistency and asymptotic normality of the MLE of $\boldsymbol{\Theta}$ for frailty models assuming a general frailty distribution, where $\Lambda_0(t)$ was estimated by a step function with steps taken at each observed failure time. Kosorok et al. (2001) showed that common frailty distributions, including normal, log-gamma, and log-inverse Gaussian, all satisfied these regularity conditions and that the MLE of $\boldsymbol{\Theta}$ is consistent and asymptotic normal under these frailty models.

For frailty models with no measurement error, the score function of $\boldsymbol{\Theta}$ is obtained by differentiating (2) with respect to β, θ, and the jump sizes of $\Lambda_0(\cdot)$. An alternative way to form the score function is to consider one-dimensional submodels through the estimators and differentiate at the estimator (Murphy 1995); that is, set $\boldsymbol{\Theta}_s(\mathbf{h}) = \boldsymbol{\Theta}(\mathbf{1} + s \cdot \mathbf{h})$, where $\mathbf{1} = (1, 1, 1)$, s is a scalar, $\mathbf{h} = \{h_1(\cdot), h_2, h_3\} \in H_p$, and for any $\tilde{\mathbf{h}} \in H_p$,

$$\boldsymbol{\Theta}(\tilde{\mathbf{h}}) = \left\{\int_0^{\cdot} \tilde{h}_1(t)\, d\Lambda_0(t), \tilde{h}_2\beta, \tilde{h}_3\theta\right\}. \tag{9}$$

Then $\widehat{\boldsymbol{\Theta}}$ solves

$$\begin{aligned}\frac{\partial}{\partial s}\ell_m\{\mathbf{N}, \mathbf{Y}; \mathbf{X}, \boldsymbol{\Theta}_s(\mathbf{h})\}\big|_{s=0} &= S_m(\mathbf{N}, \mathbf{Y}; \mathbf{X}, \boldsymbol{\Theta})(\mathbf{h}) \\ &= \sum_{i=1}^{m} S(\mathbf{N}_i, \mathbf{Y}_i; \mathbf{X}_i, \boldsymbol{\Theta})(\mathbf{h}) \\ &= 0\end{aligned}$$

for any $\mathbf{h} = (h_1, h_2, h_3)$, where $S_m(\cdot)(\mathbf{h})$ is the Gateaux derivative of ℓ_m. Because $S_m(\cdot)(\mathbf{h})$ is uniformly continuous with respect to $\boldsymbol{\Theta}$ and $\mathbf{h}$, it is a linear operator on $\mathbf{h}$ and is also a Fréchet derivative of ℓ_m (Wouk 1979, thm. 12.1.4, p. 268). The form of the score operator $S(\cdot)$ is given in Appendix A. Equation (9) indicates that the parameter space of $\boldsymbol{\Theta}$ can be considered as a subset of $\ell^\infty(H_p)$, which is the space of bounded real functions on H_p under the supremum norm $\|A\| = \sup_{\mathbf{h}\in H_p}|A(\mathbf{h})|$. Therefore, the score function is a random map from $l^\infty(H_p)$ to itself for all finite p.

Now consider the frailty measurement error model, where X_{ij} is not observed in (1) and is related to the observed covariate W_{ij} by the measurement error model (4). To study the asymptotic properties of the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$, we first note that for each set of simulated data $\mathbf{W}_c(\xi)$ $(c = 1, \ldots, C)$ obtained in the simulation step, where $\mathbf{W}_c(\xi) = \{\mathbf{W}'_{1,c}(\xi), \ldots, \mathbf{W}'_{m,c}(\xi)\}'$ and $\mathbf{W}_{i,c}(\xi) = \{W_{i1,c}(\xi), \ldots, W_{in,c}(\xi)\}'$, the corresponding naive estimate $\widehat{\boldsymbol{\Theta}}_c(\xi)$ is obtained by fitting the naive frailty model, which is model (1) with X_{ij} replaced by $W_{ij,c}$, and hence solves

$$\frac{1}{m}\sum_{i=1}^{m} S\left\{\mathbf{N}_i, \mathbf{Y}_i; \mathbf{W}_{i,c}(\xi), \widehat{\boldsymbol{\Theta}}_c(\xi)\right\}(\mathbf{h}) = 0. \tag{10}$$

Computing the sample mean $\widehat{\boldsymbol{\Theta}}(\xi) = \frac{1}{C}\sum_{c=1}^{C}\widehat{\boldsymbol{\Theta}}_c(\xi)$, the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ is obtained by extrapolating $\widehat{\boldsymbol{\Theta}}(\xi)$ back to -1, that is, $\widehat{\boldsymbol{\Theta}}_{simex} = \widehat{\boldsymbol{\Theta}}(-1)$. We assume the same regularity conditions (A)–(G) of Kosorok et al. (2001). Consistency and asymptotic normality of the SIMEX estimator $\widehat{\boldsymbol{\Theta}}_{simex}$ are stated in Theorems 1 and 2. The proof of Theorem 1 follows work of Parner (1998) and Kosorok et al. (2001)

and is thus omitted; the proof of Theorem 2 is given in Appendix B. Following Kosorok et al. (2001), all common frailty distributions, including normal, log-gamma, and log-inverse Gaussian, can be shown to satisfy these regularity conditions. Hence our results hold for all of these common frailty measurement error models.

Theorem 1 (Consistency). Denote the true parameter by $\boldsymbol{\Theta}_0$. Under the regularity conditions (A)–(G) of Kosorok et al. (2001), assuming that the true extrapolant function is used, the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex} \to \boldsymbol{\Theta}_0$ in probability.

Theorem 2 (Normality). Under the regularity conditions (A)–(G) of Kosorok et al. (2001), assuming that the true extrapolant function is used and $\boldsymbol{\Theta}(\xi) = \lim_{m\to\infty} \widehat{\boldsymbol{\Theta}}(\xi)$, $\sqrt{m}\{\widehat{\boldsymbol{\Theta}}(\xi) - \boldsymbol{\Theta}(\xi)\} \to \mathcal{G}(\xi)$ in distribution, where $\mathcal{G}(\xi)$ is a tight mean 0 Gaussian process on $\ell^\infty(H_p)$ whose variance process $\operatorname{cov}[\mathcal{G}(\xi)(\mathbf{h}), \mathcal{G}(\xi)(\mathbf{g})]$ is given in Appendix C. The SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex}$ satisfies $\sqrt{m}\{\widehat{\boldsymbol{\Theta}}_{simex} - \boldsymbol{\Theta}_0\} \to \mathcal{G}^*$ in distribution, where $\mathcal{G}^*$ is a tight mean 0 Gaussian process on $\ell^\infty(H_p)$ whose variance process $\operatorname{cov}[\mathcal{G}^*(\xi)(\mathbf{h}), \mathcal{G}^*(\xi)(\mathbf{g})]$ is given in Appendix C.

We next provide theoretical justification of the SIMEX covariance estimator given in Section 4.1. One can easily see that as $C \to \infty$, $\widehat{\boldsymbol{\Theta}}(\xi) \to \widetilde{\boldsymbol{\Theta}}(\xi) = E\{\widehat{\boldsymbol{\Theta}}_c(\xi); \mathbf{N}, \mathbf{Y}, \mathbf{W}\}$ with probability 1. When C is large, the variance process of the SIMEX estimator can be estimated using the following theorem, the proof of which is given in Appendix C.

Theorem 3 (SIMEX Variance Estimator). As $C \to \infty$, the variance process of the SIMEX estimator $\widehat{\boldsymbol{\Theta}}_{simex}$ satisfies for any $\mathbf{h}, \mathbf{g} \in H_p$,

$$\begin{aligned}
&\operatorname{cov}\left\{\widehat{\boldsymbol{\Theta}}_{simex}(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_{simex}(\mathbf{g})\right\} \\
&\quad = \lim_{\xi\to-1} \operatorname{cov}\left[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{g})\right] \\
&\qquad - \lim_{\xi\to-1} \operatorname{cov}\left[\left\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\right\}(\mathbf{h}),\right. \\
&\qquad\qquad \left.\left\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\right\}(\mathbf{g})\right]. \qquad (11)
\end{aligned}$$

Theorem 3 shows that the variance of the SIMEX estimator in the semiparametric frailty measurement error model, which contains an infinite-dimensional parameter $\Lambda_0(t)$, has properties similar to those of a parametric model, which contains only finite-dimensional parameters. Specifically, this variance consists of two parts. The first term in (11) is due to sampling variability. In fact, it corresponds to the model-based variance of the estimator of $\boldsymbol{\Theta}$ using the "true" data $(\mathbf{N}, \mathbf{Y}, \mathbf{X})$ when there is no measurement error; the second term in (11) is due to measurement error variability. These results show that the conventional SIMEX variance method of Stefanski and Cook (1995) described in Section 4.1 can be applied to semiparametric frailty measurement error models.

A major difficulty in the standard error estimation under frailty models is that one must invert a $K \times K$ Fisher information matrix, where K is the number of finite-dimensional parameters plus the number of observed distinct failure times. For large datasets, this calculation is often difficult (Keiding, Andersen, and Klein 1997; Parner 1998; Therneau and Grambsch 2000). We hence adopt the profile likelihood approach (Hu et al. 1998; Murphy and van der Vaart 2000; Li and Lin 2000). Specifically, the profile log-likelihood of $\boldsymbol{\Theta}$ is defined as $\ell_p(\boldsymbol{\Theta}) = \sup_{\Lambda_0} \ell(\boldsymbol{\Theta}, \Lambda_0)$, where the baseline cumulative hazard Λ_0 is the nuisance parameter. Then the covariance of the MLE $\widehat{\boldsymbol{\Theta}}$ is estimated by $\widehat{\operatorname{cov}}(\widehat{\boldsymbol{\Theta}}) = \{-\partial^2 \ell_p(\widehat{\boldsymbol{\Theta}})/\partial\boldsymbol{\Omega}\,\partial\boldsymbol{\Theta}'\}^{-1}$. Numerical differentiation is often used to calculate this derivative.

4.3 The SIMEX Variance Component Test

As in frailty models, a problem of common interest in frailty measurement error models is to test for within-cluster correlation. This can be done by testing the variance component $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Difficulties with this test are that the null hypothesis lies on the boundary of the parameter space, and that the standard Wald and likelihood ratio statistics do not follow a chi-squared distribution asymptotically. We consider a SIMEX score test that follows a chi-squared distribution asymptotically. Key features of this proposed test are that it is double-robust in the sense that no distributional assumptions on the frailty and the unobserved covariate X are needed, and it can be easily calculated by repeatedly fitting standard Cox models.

We first briefly review the score test for $H_0 : \theta = 0$ in the frailty model (1) when there is no measurement error, that is, when X_{ij} is observed. Such tests have been proposed by Gray (1995) and Commenges and Andersen (1995). Under H_0, the data are independent and follow the standard Cox model,

$$\lambda(t) = \lambda_0(t) e^{\beta_x X_{ij} + \mathbf{Z}_{ij}'\boldsymbol{\beta}_z}. \qquad (12)$$

Let $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) \exp\{X_{ij}\beta_x + \mathbf{Z}_{ij}'\boldsymbol{\beta}_z\}\, d\Lambda_0(s)$. Under the null hypothesis $H_0 : \theta = 0$, M_{ij} is a martingale residual with respect to the filtration $\mathcal{F}_t = \sigma\{\mathbf{X}, \mathbf{Z}, \mathbf{N}(u), \mathbf{Y}(u+) : 0 \le u \le t\}$. Gray (1995) defined a class of general martingale based residuals as

$$\begin{aligned}
R_{ij}(t) = &\int_0^t f_{ij}(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, s)\, dM_{ij}(s) \\
&- \int_0^t \eta_{ij}^{(1)}(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, s)\, dM_{++}(s), \qquad (13)
\end{aligned}$$

where $\boldsymbol{\beta}' = (\beta_x, \boldsymbol{\beta}_z')$; $f_{ij}(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, s)$ is any function that is sufficiently smooth with respect to $\mathbf{X}$, continuous with respect to β, and left continuous with respect to s; and $\eta_{ij}^{(k)}(t) = f_{ij}^k(t) S_{ij}(t)/S_{++}(t)$. Here $S_{ij}(t) = \exp(X_{ij}\beta_x + \mathbf{Z}_{ij}'\boldsymbol{\beta}_z) Y_{ij}(t)$, and the subscript "++" denotes a summation over all subjects. The left continuity assumption of $f_{ij}(\cdot)$ is to guarantee the predictability of $f_{ij}(\cdot)$ and $\eta_{ij}^{(k)}(t)$ with respect to the filtration $\mathcal{F}_t$. When $f_{ij}(\cdot) = 1$, (13) gives the basic martingale residual (Gray 1995).

The score test is based on the square sum of the within-cluster residual sums $\sum_{i=1}^m R_{i+}^2(t)$, where the subscript "$i+$" means summing over the subjects within the ith cluster. Under the null hypothesis, $R_{i+}(t)$ is a martingale, and the expectation of $R_{i+}^2(t)$ is the expectation of the variance

process of $R_{i+}(t)$, that is, $\langle R_{i+}, R_{i+}\rangle(t)$ (Fleming and Harrington 1991). It can be shown that

$$V_i(t) = \langle R_{i+}, R_{i+}\rangle(t) = \int_0^t \left[\eta_{i+}^{(2)}(s) - \left\{\eta_{i+}^{(1)}(s)\right\}^2\right] S_{++}(s)\, d\Lambda_0(s),$$

which is estimated by $\widehat{V}_i(t) = \int_0^t [\hat{\eta}_{i+}^{(2)}(s) - \{\hat{\eta}_{i+}^{(1)}(s)\}^2]\, dN_{++}(s)$ at $\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is the standard maximum partial likelihood estimate of $\boldsymbol{\beta}$ obtained by fitting (12). The score statistic for θ under $H_0 : \theta = 0$ is $\widehat{Q} = m^{-1}\sum_{i=1}^m \{\widehat{R}_{i+}^2(\tau) - \widehat{V}_i(\tau)\}$ evaluated at $\widehat{\boldsymbol{\beta}}$; that is,

$$\widehat{Q}(\mathbf{N}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}; \widehat{\boldsymbol{\beta}}) = m^{-1}\sum_{i=1}^m \left[\left\{\sum_{j=1}^{n_i}\int_0^t f_{ij}\{\mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\beta}}, s\}\, dN_{ij}(s) - \int_0^t \eta_{ij}^{(1)}\{\mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\beta}}, s\}\, dN_{++}(s)\right\}^2 - \int_0^t \left[\hat{\eta}_{i+}^{(2)}(s) - \left\{\hat{\eta}_{i+}^{(1)}(s)\right\}^2\right] dN_{++}(s)\right].$$

Under H_0, $\sqrt{m}\widehat{Q}$ is asymptotically normal with mean 0 and variance V. The exact forms of V and its estimate $\widehat{V}$ have been given by Gray (1995, appendix). It follows that the score statistic for testing $H_0 : \theta = 0$ is

$$T = \sqrt{m}\widehat{Q}/\sqrt{\widehat{V}}, \tag{14}$$

which follows $N(0, 1)$ asymptotically. Note that this test is valid for any distribution $F(\cdot)$ of the random effect b_i in (1) and is thus robust to the specification of the frailty distribution.

Now consider the frailty measurement error model (1) and (4), where X_{ij} is not observed and is measured with error by W_{ij}. We propose a SIMEX score test for $H_0 : \theta = 0$ by extending the foregoing variance component score test to the measurement error case. When there is no measurement error, the test statistic (14) is exactly the setup studied by Stefanski and Cook (1995, sec. 5.2) when the error variance is known and that studied by Carroll et al. (1996) when the error variance is estimated. We propose using SIMEX to estimate the numerator of the test statistic T by treating $\widehat{Q}(\cdot)$ as if it were a parameter and applying the SIMEX variance method to calculate the variance of this "estimator." Denoting the results by $\widehat{Q}_{simex}(\cdot)$ and $\widehat{V}_{simex}(\cdot)$, the SIMEX score statistic for testing $H_0 : \theta = 0$ is simply

$$T_{simex} = \sqrt{m}\widehat{Q}_{simex}\Big/\sqrt{\widehat{V}_{simex}}.$$

Following Lin and Carroll (1999), who considered a similar SIMEX variance component test in generalized linear mixed measurement error models, we can show that this SIMEX score statistic follows a chi-squared distribution asymptotically under H_0. Note that in this test no distributional assumptions are needed for the random effect b_i and the unobserved covariate X, and one may calculate it by fitting a standard Cox model (using, e.g., SAS PROC PHREG) for each simulated dataset in the SIMEX simulation step.

5. SIMULATION STUDIES

5.1 Parameter Estimation

We conducted simulation studies to evaluate the finite-sample performance of the SIMEX approach and to compare the SIMEX approach with the full likelihood approach of Li and Lin (2000) in terms of robustness and efficiency. In each simulated dataset, we generated survival times D_{ij} within each cluster by the conditional hazard $\lambda_{ij}(t) = \lambda_0(t)\exp(\beta_x X_{ij} + \beta_z Z_{ij} + b_i)$, $j = 1, \ldots, n$, $i = 1, \ldots, m$, where the X_{ij} and Z_{ij} were generated from standard normal $N(0, 1)$, and $b_i \sim N(0, \theta = 0.5)$. We considered correlated censoring times within each cluster. For each cluster, the censoring times P_{ij} were generated from the following model from Clayton (1978) (see also Oakes 1989; Clayton and Cuzick 1985):

$$P(P_{i1} > t_1, \ldots, P_{in} > t_n) = \{(1 - t_1/r)^\alpha + \cdots + (1 - t_n/r)^\alpha - (n-1)\}^{-1/\alpha},$$

where $\alpha \geq 0$. Marginally each P_{ij} is uniform on $[0, r]$, and the magnitude of α measures the intracluster dependence among the P_{ij}, with $\alpha = 0$ corresponding to independence. The cluster-level frailties b_i were generated according to a normal distribution, $N(0, \theta)$.

We considered the following combinations of experiments: the baseline hazard $\lambda_0(t) = 2t$; $n = 3$, $m = 40$; $\beta_x = 2$, $\beta_z = 1$; $\alpha = 1$, and r was chosen to give four different censoring proportions (0%, 30%, 50%, 80%); $C = 50$ in the SIMEX simulation step; and the measurement error variance $\sigma_u^2 = .5$. We ran 500 replications and applied the SIMEX method to analyze each dataset. In the extrapolation step of the SIMEX, we used a quadratic function. We set the frailty variance as $\theta = .5$. We calculated the sample means of the parameter estimates and compared the profile likelihood–based standard errors (SEs) with the empirical SEs.

We further compared the SIMEX with the full likelihood approach (Li and Lin 2000) in terms of efficiency and robustness. Specifically, the full likelihood approach requires specification of the distribution of unobserved covariate X and hence might not be robust when the distribution of X is misspecified. Unlike the full likelihood method, the SIMEX does not require specification of the distribution of X and would be expected to be more robust but less efficient. Our simulation setup was similar to the foregoing one, except that the true X_{ij} were generated from model (7) with both a_i and ϵ_{ij} generated from the mixture normal distribution

$$\pi N\{-(1-\pi)\mu, \sigma^2\} + (1-\pi)N\{\pi\mu, \sigma^2\}, \tag{15}$$

which has mean 0 and variance $\pi(1-\pi)\mu^2 + \sigma^2$. For a_i, we set $\pi = .25$, $\mu = 1.5$, and $\sigma^2 = .5 - \pi(1-\pi)\mu^2 = .078$,

whereas for ϵ_{ij}, we set $\pi = .25$, $\mu = 2$, and $\sigma^2 = 1 - \pi(1 - \pi)\mu^2 = .25$. The choices of π, μ and σ^2 allowed the distributions of both a_i and ϵ_{ij} to be bimodal. The full likelihood–based NPMLEs (Li and Lin 2000) were calculated by incorrectly assuming that X followed model (7) with a_i and ϵ_{ij} normally distributed. We ran 500 replications. In the SIMEX calculations, we used a quadratic extrapolation function. We repeated the experiment for the number of clusters $m = 100$.

The results are reported in Table 1. The SIMEX method performed well and had minimal biases in the estimates of the regression coefficients and the variance component. However, the NPMLEs were biased. The biases in the NPMLEs of the regression coefficients $\boldsymbol{\beta}$ were relatively small, about 10%–15%; however, the bias in the NPMLE of the variance component θ was substantial and could be as high as 75%. As the censoring proportion increased, the biases in the NPMLEs of $\boldsymbol{\beta}$ became higher and the bias in the NPMLE of θ became lower. The biases persisted as the number of clusters increased to $m = 100$. The SIMEX method, on the other hand, performed consistently well in all scenarios and had minimal biases. Its performance improved as the number of clusters increased to $m = 100$. The profile likelihood–based SEs agreed well with the empirical SEs. These results are consistent with our theoretical asymptotic bias calculations in Section 3.

We next examined the efficiency loss of the SIMEX approach compared with the full likelihood approach when the distribution of the unobserved covariate X was correctly specified in the full likelihood approach. Because the SIMEX method is a semiparametric method and makes no distributional assumption about X, we would expect it to be less efficient compared to the full likelihood approach. For each parameter configuration, we ran 500 simulations and applied both the SIMEX method and the NPMLE method to each simulated dataset.

The results are displayed in Table 2. Both the SIMEX estimates and the NPMLEs have little bias; however, the SIMEX estimates had larger SEs than the NPMLEs. For the variance component, the SEs of the SIMEX estimates were slightly higher (about 10%), whereas for the regression coefficients, the SEs of the SIMEX estimate were considerably higher (about 30%–50%) than their NPMLE counterparts. Hence, in terms of mean squared errors (MSEs), the SIMEX method was subject to little efficiency loss for the variance component estimate but considerable efficiency loss for the regression coefficient estimates compared with the MLEs. As the proportion of censoring increased, the efficiency loss of the SIMEX method compared to the MLE increased.

5.2 The SIMEX Variance Component Test

We also conducted a simulation study to evaluate the performance of the SIMEX variance component score test proposed in Section 4.2 in terms of its size and power. We set $f_{ij} \equiv 1$ in (13), corresponding to the basic martingale residuals. The design that we used in the simulation study was similar to that described in Section 5.1, except that the frailty was generated from the mixture normal distribution in (15) and censoring times P_{ij} were generated independently from uniform $[0, r]$ for simplicity. We considered two situations: $\pi = 0$, which assumes that the frailty follows a normal distribution,

Table 1. Robustness Comparison of SIMEX and NPMLE Based on 500 Replications When the Unobserved Frailty Follows a Normal Distribution $N(0, \theta)$ and the Unobserved X Follows a Mixture-Normal Distribution Under the Random-Effects Model (7)

			Full likelihood			SIMEX			
No. of clusters	Censoring level	Parameter	Estimate	SE_e	MSE	Estimate	SE_e	SE_p	MSE
$m = 40$	0%	θ	.81	.31	.19	.47	.37	.35	.14
		β_x	1.82	.33	.11	1.95	.43	.47	.18
		β_z	.91	.17	.04	.96	.23	.25	.06
	30%	θ	.73	.25	.13	.48	.38	.34	.15
		β_x	1.79	.41	.21	1.97	.54	.48	.29
		β_z	.91	.21	.05	.97	.33	.36	.11
	50%	θ	.75	.43	.25	.48	.48	.46	.23
		β_x	1.77	.48	.28	1.98	.56	.54	.31
		β_z	.79	.32	.13	.97	.42	.44	.17
	80%	θ	.62	1.03	1.12	.53	1.17	1.21	1.28
		β_x	1.63	.44	.30	1.97	.64	.58	.41
		β_z	.83	.42	.21	.96	.52	.47	.27
$m = 100$	0%	θ	.85	.22	.17	.47	.25	.21	.06
		β_x	1.83	.17	.05	2.02	.21	.24	.04
		β_z	.90	.12	.02	.97	.22	.19	.05
	30%	θ	.79	.23	.14	.47	.31	.34	.10
		β_x	1.73	.35	.17	2.05	.39	.43	.15
		β_z	.87	.18	.05	.94	.27	.24	.07
	50%	θ	.79	.33	.17	.53	.42	.39	.18
		β_x	1.73	.42	.25	1.97	.47	.51	.22
		β_z	.82	.28	.11	.98	.35	.32	.12
	80%	θ	.68	.87	.79	.52	.95	.92	.90
		β_x	1.70	.39	.24	1.97	.46	.42	.21
		β_z	.83	.34	.17	.94	.40	.39	.16

NOTE: In the NPMLE calculations, the distribution of X is misspecified as normal. The true values are $\beta_x = 2$, $\beta_z = 1$, and $\theta = .5$. The cluster size is $n = 3$, and the measurement error variance is $\sigma_U^2 = .5$. The MSEs are calculated using SE_e.

Table 2. Efficiency Comparison of SIMEX and NPMLE Based on 500 Replicates When the Unobserved Frailty Follows a Normal Distribution $N(0,\theta)$ and the Unobserved X Follows a Normal Distribution Under the Random-Effects Model (7)

		Full likelihood			SIMEX			
Censoring level	Parameter	Estimate	SE_e	MSE	Estimate	SE_e	SE_p	MSE
0%	θ	.48	.38	.14	.47	.45	.42	.20
	β_x	1.93	.32	.10	1.94	.42	.44	.16
	β_z	.95	.18	.04	.96	.26	.32	.07
30%	θ	.49	.43	.18	.46	.48	.51	.23
	β_x	1.98	.35	.13	2.04	.52	.55	.27
	β_z	1.02	.22	.05	.98	.32	.35	.11
50%	θ	.53	.44	.19	.47	.51	.48	.25
	β_x	1.94	.37	.14	1.97	.58	.54	.34
	β_z	.97	.21	.04	.94	.38	.34	.14
80%	θ	.52	.97	.95	.54	1.09	1.12	1.18
	β_x	2.06	.44	.19	1.92	.68	.71	.46
	β_z	1.02	.30	.09	.94	.44	.40	.20

NOTE: In the NPMLE calculation, the distribution of X is correctly specified as normal. The true values are $\beta_x = 2$, $\beta_z = 1$, and $\theta = .5$. The cluster number is $m = 40$, and the cluster size is $n = 3$. The measurement error variance is taken as $\sigma_u^2 = .5$.

and $\mu = 1$ and $\pi = .25$, which assumes that the frailty follows a bimodal mixture normal distribution. In both cases we varied the variance component $\theta = \pi(1-\pi)\mu^2 + \sigma^2$ as 0, .25, .50, and 1.00 to study the size and the power of the SIMEX variance component score test. For the purpose of comparison, we also calculated the size of the naive variance component score test obtained by ignoring the measurement error with X_{ij} replaced by W_{ij}. Note that we calculated both sizes and powers using a one-sided test. The nominal size was set to be .05, and two censoring proportions (0% and 50%) were considered. We ran 1,500 simulations.

Table 3 gives the sizes and the powers of the naive and the SIMEX variance component score tests. The level of the naive score test was too high, and its performance worsened as the frailty distribution departed from normality and followed a bimodal normal mixture. However, the SIMEX score test performed well in all cases, and its level was very close to the nominal value. As the variance component θ increased, the power of the SIMEX score test increased and quickly approached 1.

6. APPLICATION TO THE KENYA PARASITEMIA DATA

We applied the proposed methods to the analysis of Kenya parasitemia data (McElroy et al. 1997) introduced in Section 1. A total of 542 children from 309 households age 6 months–6 years were enrolled into the study between February 1986 and July 1987. At the entry into the study, the children were treated to eliminate blood-stage infection of parasitemia, and their blood films were examined 2 weeks after enrollment and found to be negative. They were then followed for the first recurrence of parasitemia for up to 22 months.

Table 3. Empirical Sizes and Powers of the Naive and SIMEX Score Tests for the Variance Component Observed in 1,500 Simulations

Frailty distribution	Censoring level	Method	$\theta = 0$	$\theta = .25$	$\theta = .50$	$\theta = 1.00$
Normal	0%	Naive	.086			
		SIMEX	.053	.234	.398	.896
	50%	Naive	.070			
		SIMEX	.043	.204	.353	.867
Bimodal	0%	Naive	.123			
		SIMEX	.055	.266	.442	.923
	50%	Naive	.093			
		SIMEX	.053	.247	.409	.903

In the first 2 weeks after enrollment, two field workers visited each household one night each week and took turns collecting mosquitos on another person's legs every 30 minutes throughout the night. The next morning these collected mosquitos were delivered to a laboratory, where the number of infective mosquitos was determined. The investigators were interested in studying the effects of the daily mean dose of infective bites in the first 2 weeks on the risk of recurrent parasitemia. The daily mean dose of infective bites in the first 2 weeks was calculated using the average of the two night measures and hence was measured with substantial error. The other covariates included sex (1 = F, 0 = M), age, and baseline parasitemia density. The average follow-up time was 9 months, and about 90% children experienced recurrent parasitemia during the follow-up. The average number of the daily mean dose of infective mosquito bites was .89. The baseline parasitemia density was log-transformed (LNBPD), and the daily mean dose of infective bites in the first 2 weeks was quartic root–transformed (BITE), to be consistent with the previous analysis (Li and Lin 2000).

We fitted a random intercept frailty measurement error model,

$$\lambda_{ij}(t; b_i) = \lambda_0(t)\exp\{\beta_1 \times \mathrm{BITE}_{ij} + \beta_2 \times \mathrm{AGE}_{ij} + \beta_3 \times \mathrm{SEX}_{ij} + \beta_4 \times \mathrm{LNBPD}_i + b_i\}, \quad (16)$$

where the frailty b_i follows $N(0, \theta)$. Here X = true BITE and Z = (AGE, GENDER, LNBPD). Because different children within the same household might enter into the study at different times, their numbers of infective mosquito bites might be different. Because the measurement error variance σ_u^2 was not available from the data, following Li and Lin (2000), we

conducted a sensitivity analysis by varying σ_u^2 from 0 (naive analysis), .08 (moderate error), or .20 (severe error). When fitting the models, we treated σ_u^2 as fixed and known.

We applied the SIMEX method to fit (16) and compared the results with those obtained using the NPMLE of Li and Lin (2000), who assumed X following a normal linear mixed model. Figure 3 shows the histogram of the observed BITE variable W and suggests that the distribution of X does not seem to be normally distributed. This indicates that the SIMEX approach could be more appropriate than the NPMLE. A quadratic extrapolant function was used in the SIMEX analysis.

The results are presented in Table 4. A greater risk of parasitemia was significantly associated with a higher daily mean dose of infective mosquito bites, older age, and a higher baseline parasitemia density. Ignoring the measurement error attenuated the regression coefficient estimates and inflated the variance component. The SIMEX estimates corrected the biases in the naive estimates. As σ_u^2 increased, the SIMEX estimates of the regression coefficients increased and the SIMEX estimates of the variance component θ decreased. For example, the SIMEX coefficient of BITE increased from .33 (SE = .11) when $\sigma_u^2 = 0$ to 1.05 (SE = .29) when $\sigma_u^2 = .20$. This indicates that accounting for the measurement error increased the magnitude of the estimated effects of BITE. The SIMEX variance component estimate decreased from .16 (SE = .11) to .12 (SE = .22) when σ_u^2 changed from 0 to .20. For this particular dataset, the results using the SIMEX and the NPMLE were similar.

We next applied the proposed SIMEX variance component score test to model (16) to test $H_0 : \theta = 0$, that is, no within-household correlation. The naive score statistic for $H_0 : \theta = 0$ calculated by ignoring the measurement error assuming that $\sigma_u^2 = 0$ was 1.37 (p value = .08). It implied that there was some evidence of the within-household correlation. The SIMEX score statistics that accounted for the measurement error were .87 (p value = .20) and .60 (p value = .27) when assuming that $\sigma_u^2 = .08$ and .20. Note that a one-sided test was used because $H_a : \theta > 0$ is one-sided. Unlike the naive score tests, the SIMEX score tests did not show a significant correlation among observations within the same household. Note that the standard error of the estimate of θ in Table 4 cannot be directly used to test for $H_0 : \theta = 0$, because the null hypothesis is on the boundary of the parameter space and the Wald statistic is not asymptotically distributed as a chi-square (Lin 1997).

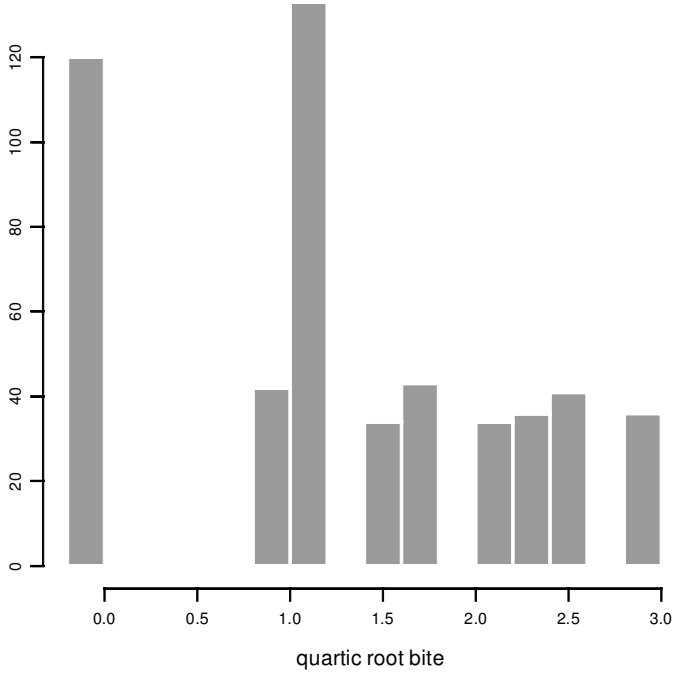

Figure 3. Histogram of the Quartic Root of BITE in the Kenya Parasitemia Data.

Table 4. Analysis Results of the Kenya Parasitemia Data

	Naive	*Moderate error*		*Severe error*	
Parameter	*NPMLE*	*NPMLE*	*SIMEX*	*NPMLE*	*SIMEX*
θ	.16(.11)	.14(.11)	.15(.22)	.11(.12)	.12(.22)
Bite	.33(.11)	.49(.15)	.63(.26)	.99(.23)	1.05(.29)
Age	.04(.02)	.04(.02)	.04(.02)	.04(.02)	.04(.02)
Gender	.06(.03)	.06(.04)	.06(.04)	.05(.04)	.05(.04)
Baseline bpd	.09(.02)	.09(.02)	.09(.03)	.09(.02)	.09(.03)

NOTE: Moderate error means measurement error variance $\sigma_U^2 = .08$, while severe error means $\sigma_U^2 = .20$. Estimates were calculated by the NPMLE and the SIMEX approaches. The SEs (values in the parentheses) are calculated by the profile likelihood approach.

7. DISCUSSION

In this article, we have proposed functional estimation using the SIMEX approach in frailty measurement error models. Our asymptotic bias analysis shows that if one accounts for measurement error in a full likelihood analysis but misspecifies the distribution of the unobserved covariate X, then the estimates will be biased—especially the estimate of the variance component, whose bias can be substantial. The SIMEX approach has an advantage of being more robust than the full likelihood approach in the sense that no distributional assumption on the unobserved covariate X is needed. It is also computationally more convenient than the full likelihood method, because one need only repeatedly fit standard (no measurement error) frailty models (using, e.g., S-PLUS) in the simulation step. It should be noted that the robustness of the SIMEX applies only to the distribution of the unobserved covariate X, not to the frailty distribution. In other words, a valid SIMEX estimation requires that the frailty distribution be correctly specified.

We have studied the asymptotic properties of the SIMEX estimates for frailty measurement error models and have shown that they are consistent and asymptotically normal under some regularity conditions. The results apply to common frailty distributions, such as normal and log-gamma. Our results extend those of Carroll et al. (1996) for parametric $Y|X$ models to semiparametric $Y|X$ models, such as frailty models.

Our simulation studies show that the SIMEX performs well in finite samples, and that there is a trade-off between robustness and efficiency of the SIMEX method and the full likelihood method. Specifically, when the distribution of the unobserved covariate X is misspecified, the NPMLEs are biased, especially the variance component estimate, whose bias can be substantial, whereas the SIMEX estimates are still close to

the true values. However, when the distribution of the unobserved covariate X is correctly specified in a likelihood analysis, the SIMEX estimates are subject to small efficiency loss in estimating the variance component but considerable efficiency loss in estimating the regression coefficients compared with the NPMLEs. Regarding the choice between the SIMEX method and the NPMLE method, we would suggest that one examines the distribution of the error-prone covariates W. If W deviates much from a normal distribution, then the SIMEX estimates may be more preferable, because they are robust to the distribution of the unknown covariates. But if the normality assumption of W were plausible, which might imply that the unobserved X is normal, then the NPMLE estimates would be more efficient and would be preferred. We also recommend applying both methods in practice as a sensitivity analysis and checking whether the results agree.

We have also proposed a simple SIMEX score test for the variance component in frailty measurement error models. This test, based on the martingale residuals, is an extension of Gray's (1995) test to the measurement error context. One key feature of the SIMEX score test is its double robustness; that is, no distributions need be assumed for the random effects and the unobserved covariate X. Further, it is easy to calculate by repeatedly fitting standard Cox models.

Our simulation studies indicate that the level of the naive variance component score test when ignoring the measurement error is higher (sometimes much higher) than the nominal value. In contrast, the SIMEX score test performs well, yielding a correct level and good statistical power. However, the SIMEX score test could be less powerful than a fully parametric score test constructed by assuming a parametric distribution for X. Unlike the SIMEX score test, which has a closed-form expression and can be easily implemented using existing statistical software, such a fully parametric score test can be difficult to construct because a closed-form expression is often not available, it often involves multiple-dimensional integration, and new statistical software usually needs to be developed. Future research is needed to compare the robustness and the power of the SIMEX variance component score test and a full likelihood–based score test. It would also be of interest to investigate whether the SIMEX variance component score test is locally most powerful.

APPENDIX A: DERIVATION OF THE SCORE OPERATOR

We assume in Appendixes A and B that the frailty b_i follows a normal distribution with mean 0 and variance θ. We rewrite the log likelihood function (3) as

$$\ell_m(\mathbf{N},\mathbf{Y};\mathbf{X},\boldsymbol{\Theta}) = \sum_{i=1}^{m} \ln\left[\prod_{j=1}^{n}\left\{\prod_{0\le t\le\tau} Y_{ij}(t)e^{X_{ij}\beta}\Delta\Lambda_0(t)\right\}^{\Delta N_{ij}(t)} \times G\{N_{i\cdot}(\tau), Y_{i\cdot}^{\beta}(\tau),\theta\}\right],$$

where $N_{i\cdot}(\tau)=\sum_{j=1}^{n} N_{ij}(\tau)$, $Y_{i\cdot}^{\beta}(\tau)=\sum_{j=1}^{n}\int_0^\tau Y_{ij}(t)\exp(X_{ij}\beta)d\Lambda_0(t)$, and $G\{n,x,\theta\} = E\{e^{nb_1 - x\exp(b_1)}\} = (-1)^n L^{(n)}(x;\theta)$ for $n \in \{0,1,2,\dots\}$. Here $L^{(n)}(\cdot)$ denotes the nth derivative of the Laplace transformation with respect to the random variable $\exp(b_1)$. Let $G_x\{n,x,\theta\} = \frac{\partial}{\partial x}G\{n,x,\theta\}$ and $G_\theta\{n,x,\theta\} = \frac{\partial}{\partial\theta}G\{n,x,\theta\}$. Some algebra shows that $G_x\{n,x,\theta\} = -G\{n+1,x,\theta\}$.

To calculate the score operator, choose the one-dimensional submodels $s:\to \boldsymbol{\Theta}_s(\mathbf{h}) = \boldsymbol{\Theta} + s(\int_0^\cdot h_1\, d\Lambda_0, h_2, h_3)$ for $\mathbf{h} = \{h_1(\cdot), h_2, h_3\}$. Then define the score operator by

$$\begin{aligned} &S_m(\mathbf{N},\mathbf{Y};\mathbf{X},\boldsymbol{\Theta})(\mathbf{h}) \\ &\quad= \frac{\partial}{\partial s}\ell_m\{\mathbf{N},\mathbf{Y};\mathbf{X},\boldsymbol{\Theta}_s(\mathbf{h})\}\big|_{s=0} \\ &\quad= \sum_{i=1}^{m}\{S_{\Lambda_0}(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_1) + S_\beta(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_2) \\ &\qquad + S_\theta(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_3)\}, \end{aligned}$$

where

$$\begin{aligned} &S_{\Lambda_0}(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_1) \\ &\qquad= \sum_{j=1}^{n}\int_0^\tau h_1(t)\,dN_{ij}(t) - \frac{G_{i,2}}{G_i}\sum_{j=1}^{n}\int_0^\tau Y_{ij}(t)e^{X_{ij}\beta}h_1\,d\Lambda_0(t), \end{aligned}$$

$$\begin{aligned} &S_\beta(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_2) \\ &\qquad= h_2\sum_{j=1}^{n}\int_0^\tau X_{ij}\,dN_{ij}(t) - h_2\frac{G_{i,2}}{G_i}\sum_{j=1}^{n}\int_0^\tau Y_{ij}(t)e^{X_{ij}\beta}X_{ij}\,d\Lambda_0(t), \end{aligned}$$

and

$$S_\theta(\mathbf{N}_i,\mathbf{Y}_i;\mathbf{X}_i,\boldsymbol{\Theta})(h_3) = h_3\frac{G_{i,3}}{G_i},$$

where $G_i = G\{N_{i\cdot}(\tau), Y_{i\cdot}^{\beta}(\tau),\theta\}$, $G_{i,2} = G_x\{N_{i\cdot}(\tau), Y_{i\cdot}^{\beta}(\tau),\theta\}$, and $G_{i,3} = G_\theta\{N_{i\cdot}(\tau), Y_{i\cdot}^{\beta}(\tau),\theta\}$.

APPENDIX B: PROOF OF NORMALITY (THEOREM 2)

With the same regularity conditions postulated by Kosorok et al. (2001), we first show that for a given $\xi > 0$, $\sqrt{m}(\widehat{\boldsymbol{\Theta}}(\xi) - \boldsymbol{\Theta}(\xi)) \to \mathcal{G}(\xi)$ weakly, where $\mathcal{G}(\xi)$ is a tight mean 0 Gaussian process. Specifically, the naive estimate $\widehat{\boldsymbol{\Theta}}_c(\xi)$ for data $(\mathbf{N},\mathbf{Y},\mathbf{W}_c)$ solves, for any $h \in H_p$,

$$\begin{aligned} \overline{S}_m\{\mathbf{W}_c(\xi), \widehat{\boldsymbol{\Theta}}_c(\xi)\}(\mathbf{h}) &= \frac{1}{m}\sum_{i=1}^{m} S\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_{i,c}(\xi), \widehat{\boldsymbol{\Theta}}_c(\xi)\}(\mathbf{h}) \\ &= 0, \end{aligned}$$

where $\mathbf{W}_{i,c}(\xi) = \mathbf{W}_i + \xi^{1/2}\mathbf{U}_{i,c}^*$. Hence by lemma 2 of Kosorok et al. (2001), that the information operator is a continuously invertible linear operator, we have that

$$\sqrt{m}\{\widehat{\boldsymbol{\Theta}}(\xi) - \boldsymbol{\Theta}(\xi)\} = m^{-1/2}\sum_{i=1}^{m}\mathcal{F}_i(\xi) + o_p(1),$$

where $\mathcal{F}_i(\xi) = -\dot{S}^{-1}\{\xi,\boldsymbol{\Theta}(\xi)\}[\chi_{C,i}\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i,\mathbf{U}_i^*,\xi,\boldsymbol{\Theta}(\xi)\}]$. Here the information operator $\dot{S}\{\xi,\boldsymbol{\Theta}(\xi)\} = E[S^{(1)}\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i + \xi^{1/2}\mathbf{U}_{i,c}^*,\boldsymbol{\Theta}(\xi)\}]$, $S^{(1)}$ is the Fréchet derivative of S with respect to $\boldsymbol{\Theta}$ and $\chi_{C,i}\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i,\mathbf{U}_i,\xi,\boldsymbol{\Theta}(\xi)\} = \frac{1}{C}\sum_{c=1}^{C} S\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i + \xi^{1/2}\mathbf{U}_{i,c}^*,\boldsymbol{\Theta}(\xi)\}$. Notice that

$$\begin{aligned} E\mathcal{F}_i &= E[\dot{S}^{-1}\{\xi,\boldsymbol{\Theta}(\xi)\}\chi_{C,i}\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i,\mathbf{U}_i^*,\xi,\boldsymbol{\Theta}(\xi)\}] \\ &= \dot{S}^{-1}\{\xi,\boldsymbol{\Theta}(\xi)\}E[\chi_{C,i}\{\mathbf{N}_i,\mathbf{Y}_i;\mathbf{W}_i,\mathbf{U}_i^*,\xi,\boldsymbol{\Theta}(\xi)\}] \\ &= 0. \end{aligned}$$

Because $\mathcal{F}_i$ are independent and identically distributed replicates, applying the functional central limit theorem (van der Vaart and Wellner 1996), we have $m^{-1/2}\sum_{i=1}^{m}\mathcal{F}_i \to \mathcal{G}(\xi)$, where $\mathcal{G}(\xi)$ is a tight Gaussian process on $\ell^\infty(H_p)$ with mean 0 and covariance process

$$\begin{aligned}\operatorname{cov}\{\mathcal{G}(\xi)(\mathbf{h}), \mathcal{G}(\xi)(\mathbf{g})\} &= \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\operatorname{cov}\{\mathcal{F}_i(\xi)(\mathbf{h}), \mathcal{F}_i(\xi)(\mathbf{g})\}\\ &= \int_0^{\cdot} h_1\sigma_\Lambda^{-1}(\xi)(\mathbf{g})\,d\Lambda_0 + h_2\sigma_\beta^{-1}(\xi)(\mathbf{g})\\ &\quad + h_3\sigma_\theta^{-1}(\xi)(\mathbf{g}),\end{aligned}$$

and $\{\sigma_\Lambda(\xi), \sigma_\beta(\xi), \sigma_\theta(\xi)\}$ are some continuously invertible linear operators from H_∞ to H_∞ with inverse $\{\sigma_\Lambda^{-1}(\xi), \sigma_\beta^{-1}(\xi), \sigma_\theta^{-1}(\xi)\}$. Hence $\sqrt{m}\{\widehat{\boldsymbol{\Theta}}(\xi) - \boldsymbol{\Theta}(\xi)\} \to \mathcal{G}(\xi)$ weakly.

Suppose that $\boldsymbol{\Xi} = (\xi_1, \ldots, \xi_K)'$, where K is the number of grid points of ξ used in the SIMEX simulation step. Using the foregoing results, we can easily show that

$$\sqrt{m}\{\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Xi}) - \boldsymbol{\Theta}(\boldsymbol{\Xi})\} = m^{-1/2}\sum\mathcal{F}_i(\boldsymbol{\Xi}) + o_p(1) \to \mathcal{G}_1(\boldsymbol{\Xi})$$

in distribution, where $\boldsymbol{\Theta}(\boldsymbol{\Xi}) = \{\boldsymbol{\Theta}(\xi_1), \ldots, \boldsymbol{\Theta}(\xi_K)\}'$, $\mathcal{F}_i(\boldsymbol{\Xi}) = \{\mathcal{F}_i(\xi_1), \ldots, \mathcal{F}_i(\xi_K)\}'$, and $\mathcal{G}_1(\boldsymbol{\Xi})$ is a tight mean 0 Gaussian process on $\ell^\infty(H_p)$ with covariance process

$$\operatorname{cov}\{\mathcal{G}_1(\boldsymbol{\Xi})(\mathbf{h}), \mathcal{G}_1(\boldsymbol{\Xi})(\mathbf{g})\} = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\operatorname{cov}\{\mathcal{F}_i(\boldsymbol{\Xi})(\mathbf{h}), \mathcal{F}_i(\boldsymbol{\Xi})(\mathbf{g})\}.$$

Suppose that $\boldsymbol{\Theta}(\xi)$ can be specified using a parametric model $g(\boldsymbol{\tau}, \xi)$ depending on a parameter $\boldsymbol{\tau}$. Assuming the true extrapolation function, we have $\boldsymbol{\Theta}_0 = g(\tau, -1)$ and $\widehat{\boldsymbol{\Theta}}_{simex} = g(\widehat{\boldsymbol{\tau}}, -1)$, where $\widehat{\boldsymbol{\tau}}$ solves

$$\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\{\mathbf{g}(\boldsymbol{\tau}, \boldsymbol{\Xi}) - \widehat{\boldsymbol{\Theta}}(\boldsymbol{\Xi})\} = 0$$

and $\dot{\mathbf{g}} = \partial\mathbf{g}/\partial\boldsymbol{\tau}'$. We then have

$$\begin{aligned}\sqrt{m}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}) = \{\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})\}^{-1}\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\sqrt{m}\{\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Xi}) - \boldsymbol{\Theta}(\boldsymbol{\Xi})\}\\ + o_p(1) \to \mathcal{G}_2(\boldsymbol{\Xi})\end{aligned}$$

in distribution, where $\mathcal{G}_2(\boldsymbol{\Xi}) = \{\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})\}^{-1}\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\mathcal{G}_1(\boldsymbol{\Xi})$. Because the SIMEX estimate $\widehat{\boldsymbol{\Theta}}_{simex} = g(\widehat{\boldsymbol{\tau}}, -1)$ and $\boldsymbol{\Theta}(-1) = g(\boldsymbol{\tau}, -1) = \boldsymbol{\Theta}_0$, using the delta method, we have that $\sqrt{m}\{\widehat{\boldsymbol{\Theta}}_{simex} - \boldsymbol{\Theta}_0\} \to \mathcal{G}_3$, where $\mathcal{G}_3 = \dot{\mathbf{g}}(\boldsymbol{\tau}, -1)'\{\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\}^{-1}\dot{\mathbf{g}}(\boldsymbol{\tau}, \boldsymbol{\Xi})'\mathcal{G}_1(\boldsymbol{\Xi})$ is a tight mean 0 Gaussian process on $\ell^\infty(H_p)$.

APPENDIX C: PROOF OF THE VARIANCE PROCESS FORM (THEOREM 3)

Because $\widetilde{\boldsymbol{\Theta}}(\xi) = E\{\widehat{\boldsymbol{\Theta}}_c(\xi); \mathbf{N}, \mathbf{Y}, \mathbf{W}\}$, $\widehat{\boldsymbol{\Theta}}(\xi) \to \widetilde{\boldsymbol{\Theta}}(\xi)$ almost surely as $C \to \infty$. Then $\operatorname{cov}\{\widehat{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}(\xi)(\mathbf{g})\} \to \operatorname{cov}\{\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})\}$. Now calculate

$$\begin{aligned}&\operatorname{cov}[\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\}(\mathbf{h}), \{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\}(\mathbf{g})]\\ &\quad= \operatorname{cov}[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{g})] + \operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\\ &\qquad- \operatorname{cov}[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})] - \operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{g})],\end{aligned}$$

where

$$\begin{aligned}\operatorname{cov}[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})] &= E[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h})\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\\ &\quad - E[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h})]E[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\\ &= E\{E[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h})\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g}); \mathbf{N}, \mathbf{Y}, \mathbf{W}]\\ &\quad - E[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}); \mathbf{N}, \mathbf{Y}, \mathbf{W}]E[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\}\\ &= E[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h})\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\\ &\quad - E[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h})]E[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]\\ &= \operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})].\end{aligned}$$

Similar calculations show that $\operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{g})] = \operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})]$. It follows that

$$\begin{aligned}&\operatorname{cov}[\{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\}(\mathbf{h}), \{\widehat{\boldsymbol{\Theta}}_c(\xi) - \widetilde{\boldsymbol{\Theta}}(\xi)\}(\mathbf{g})]\\ &\quad= \operatorname{cov}[\widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{h}), \widehat{\boldsymbol{\Theta}}_c(\xi)(\mathbf{g})] - \operatorname{cov}[\widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{h}), \widetilde{\boldsymbol{\Theta}}(\xi)(\mathbf{g})].\end{aligned}$$

This completes the proof.

[Received August 2000. Revised May 2002.]

REFERENCES

Carroll, R., Küchenhoff, H., Lombard, F., and Stefanski, L. A. (1996), "Asymptotics for the SIMEX Estimator in Structural Measurement Error Models," *Journal of the American Statistical Association*, 91, 242–250.

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

Clayton, D. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familiar Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Clayton, D. G., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model," *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Commenges, D., and Andersen, P. K. (1995), "Score Test of Homogeneity for Survival Data," *Lifetime Data Analysis*, 1, 145–159.

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.

Gray, R. J. (1995), "Tests for Variation Over Groups in Survival Data," *Journal of the American Statistical Association*, 90, 198–203.

Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671–678.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998), "Estimating the Parameters in the Cox Model When Covariate Variables Are Measured With Error," *Biometrics*, 54, 1407–1419.

Hughes, M. D. (1993), "Regression Dilution in the Proportional Hazards Model," *Biometrics*, 49, 1056–1066.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2001). "Semiparametric Inference for Proportional Hazards Frailty Regression Models," Technical Report 156, University of Wisconsin-Madison, Dept. of Biostatistics and Bioinformatics.

Li, Y., and Lin, X. (2000), "Covariate Measurement Errors in Frailty Models for Clustered Survival Data," *Biometrika*, 87, 4, 849–866.

Lin, D. Y., and Ying, Z. (1993), "Cox Regression With Incomplete Covariate Measurements," *Journal of the American Statistical Association*, 88, 1341–1349.

Lin, X. (1997), "Variance Component Testing in Generalized Linear Models With Random Effects," *Biometrika*, 84, 309–326.

Lin, X., and Carroll, R. J. (1999), "SIMEX Variance Component Tests in Generalized Linear Mixed Measurement Error Models," *Biometrics*, 55, 613–619.

McElroy, P. D., Beier, J. C., Oster, C. N., Onyango, F. K., Oloo, A. J., Lin, X., Beedle, C., and Hoffman, S. L. (1997), "Dose- and Time-Dependent Relations Between Infective Anopheles Inoculation and Outcomes of Plasmodium Falciparum Parasitemia Among Children in Western Kenya," *American Journal of Epidemiology*, 145, 945–956.

McGilchrist, C. A. (1993), "REML Estimation for Survival Models With Frailty," *Biometrics*, 49, 221–225.
Murphy, S. A. (1995), "Asymptotic Theory for the Frailty Model," *The Annals of Statistics*, 23, 182–198.
Murphy, S. A., and van der Vaart, A. W. (2000), "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–471.
Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992), "A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models," *Scandinavian Journal of Statististics*, 19, 25–43.
Oakes, D. (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.
Paik, M., and Tsai, W. (1997), "On Using the Cox Proportional Hazards Model With Missing Covariates," *Biometrika*, 84, 579–593.
Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-Frailty Model," *The Annals of Statistics*, 26, 183–214.
Prentice, R. L. (1982), "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model," *Biometrika*, 69, 331–342.
Stefanski, L. A., and Cook, J. R. (1995), "Simulation-Extrapolation: The Measurement Error Jackknife," *Journal of the American Statistical Association*, 90, 1247–1256.
Therneau, T. M., and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, New York: Springer.
Tsiatis, A. A., De Gruttola, V., and Wulfsohn, M. S. (1995), "Modeling the Relationship of Survival to Longitudinal Data Measured With Error. Applications to Survival and CD4 Counts in Patients With AIDS," *Journal of the American Statistical Association*, 90, 27–37.
van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Process*, New York: Springer Verlag.
Wang, N., Lin, X., Gutierrez, R., and Carroll R. J. (1998), "Bias Analysis and the SIMEX Approach in Generalized Linear Mixed Error Models," *Journal of the American Statistical Association*, 93, 249–261.
Wouk, A. (1979), *A Course of Applied Functional Analysis*, New York: Wiley.
Zhou, H., and Pepe, M. S. (1995), "Auxiliary Covariate Data in Failure Time Regression," *Biometrika*, 82, 139–149.