Latent-model Robustness in Joint Models for a Primary Endpoint and a Longitudinal Process

Xianzheng Huang[1,*], **Leonard A. Stefanski**[2], and **Marie Davidian**[2]

[1] Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, U.S.A

[2] Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A

Summary

Joint modeling of a primary response and a longitudinal process via shared random effects is widely used in many areas of application. Likelihood-based inference on joint models requires model specification of the random effects. Inappropriate model specification of random effects can compromise inference. We present methods to diagnose random effect model misspecification of the type that leads to biased inference on joint models. The methods are illustrated via application to simulated data, and by application to data from a study of bone mineral density in perimenopausal women and data from an HIV clinical trial.

Keywords

Censoring; Random effect; Remeasurement method; SIMEX

1. Introduction

It is often of interest to characterize the association between a primary endpoint and a longitudinal process and to also understand the inherent features of the longitudinal process. One popular approach is to link a regression model for the primary endpoint and a mixed effects model for the longitudinal process through joint dependence on latent random effects. It has been demonstrated (e.g., Hsieh, Tseng, and Wang, 2006) that appropriate parametric modeling of the random effects in joint models yields more effcient inference procedures and can also shed light on the underlying features of the longitudinal process. One concern in this approach is the sensitivity of inference to the model assumptions on random effects. In this article, we address the issue of robustness of estimators for the primary regression parameters to such assumptions. We call this aspect of robustness *latent-model robustness*.

The primary endpoint in the joint model setting can be a simple response such as a binary indicator of the presence of a disease, or more complex such as a possibly censored time-to-event. The Study of Women's Health Across the Nation (SWAN) (Sowers et al., 2003) provides an example of the former. Two objectives of SWAN are to characterize the association between an indicator of the evidence of osteopenia, a binary endpoint, and the underlying hormone patterns over the menstrual cycle in perimenopausal women, and to understand the underlying hormone patterns in this population. The hormone patterns cannot be observed directly but are observed through longitudinal progesterone levels derived from urine (PDG). AIDS Clinical Trials Group (ACTG) Protocol 175 (Hammer et al., 1996) is a setting where a joint model with time-to-event endpoint is a relevant framework. In this study, more than 2000 HIV-1-infected

*email: huang@stat.sc.edu.

subjects were followed for their CD4 counts from week 8 post baseline and every 12 weeks thereafter, and the "event" is defined as a composite of ≥50% decline in CD4, progression to AIDS, or death. It is of interest to study the prognostic value of CD4 counts and their inherent trajectories over time in this population. In both studies, the longitudinal measurements, PDG and CD4 counts, are subject to assay error and intra-subject variation.

Assuming multivariate normal random effects, Wulfsohn and Tsiatis (1997) obtained maximum likelihood estimators (MLEs) for the regression parameters in joint models with time-to-event endpoint. Wang, Wang, and Wang (2000) proposed three methods to estimate the primary regression parameters in joint models with simple endpoint. Their methods rely on the assumption that the random effects follow a multivariate normal distribution, and they noted the concern about the sensitivity of inference to the normality assumption. Song, Davidian, and Tsiatis (2002) modeled the random effects using a flexible seminonparametric (SNP) model to avoid the restrictive normal assumption. Li, Zhang, and Davidian (2004) proposed conditional score estimators (CSEs) for the primary regression parameters in joint models with simple endpoint. Tsiatis and Davidian (2001) also derived the CSEs for the regression parameters in joint models with time-to-event endpoint. The CSEs require no assumption on the random effects. However, the latent-model robustness of the CSEs is achieved at the expense of loss of effciency. The effects of model misspecification on random effects in joint models have been investigated by several authors. Through extensive simulation studies, Hsieh et al. (2006) demonstrated robustness of the MLEs against departure from the normal random effect assumption in joint models with time-to-event endpoint. Hsieh et al. (2006) concluded that the MLE is robust to random effect model misspecification when there is rich enough information from the longitudinal data. Also focusing on joint models with time-to-event endpoint, Rizopoulos, Verbeke, and Molenberghs (2008) investigated the effect of misspecifying the random effect model on the parameter estimators and their standard errors. They showed that the difference between the MLE obtained from the joint model with a misspecified random effect model and the MLE based on the correct model converges to zero as the number of repeated measurements per subject increases.

Assuming the two component models in a joint model correct, the MLE is consistent and effcient when the random effect model is correctly specified. Even with the robustness property of the MLE revealed by the aforementioned authors, a relevant question is whether or not the available longitudinal information in a particular data set is rich enough to yield an MLE insensitive to model misspecification. Diagnostic tools that can reveal adverse effects of model misspecification when they do exist are thus desired. Huang, Stefanski, and Davidian (2006) applied a remeasurement method to structural measurement error models to diagnose model misspecification on the unobservable true predictor. In this article, we use an improved remeasurement method to develop diagnostic tools for joint models. In Section 2, we formulate joint models generically. From a viewpoint different from that of Rizopoulos et al. (2008), we provide an explanation in Section 3 for the asymptotic latent-model robustness of the MLE when longitudinal data information is extensive enough. In Section 4, we describe the improved remeasurement method and apply it to joint models to diagnose random effect model misspecification; test statistics are also proposed to assess quantitatively the robustness of parameter estimators. The diagnostic methods are illustrated via simulation in Section 5. In Section 6, the proposed methods are applied to the SWAN and ACTG 175 data sets.

2. Joint Models

For subject i, $i = 1,\ldots, n$, denote by $\mathbf{Y}_i$ the primary endpoint, which is a scalar in joint models with simple endpoint, and is defined as a vector in joint models with time-to-event endpoint. Denote by $\mathbf{W}_i = (W_{i1},\ldots,W_{im_i})^T$ the set of longitudinal measurements recorded at times $\mathbf{t}_i =$

$(t_{i1},\ldots,t_{im_i})^T$ and by $\mathbf{H}_i$ the vector of observed covariates, for $i = 1,\ldots, n$. Finally, define $\mathbf{Q}_i^T = (\mathbf{Y}_i^T, \mathbf{W}_i^T, \mathbf{H}_i^T)^T$ as all the observed data from subject i, for $i = 1,\ldots, n$.

The two component models in a joint model are the model for the primary response $\mathbf{Y}_i$ and the model for the longitudinal process $\mathbf{W}_i$. Define $f_{\mathbf{Y}_i|\mathbf{X}_i\mathbf{H}_i}(y_i|x_i, h_i; \theta, \zeta)$ as the density function associated with the first component model, where θ is the vector of primary regression parameters that relate $\mathbf{Y}_i$ to $(\mathbf{X}_i, \mathbf{H}_i)$, ζ is a vector of nuisance parameters, and $\mathbf{X}_i$ is the $p \times 1$ vector of latent variables. Denote by $f_{\mathbf{X}_i|\mathbf{H}_i}^{(a)}(x_i|h_i; \tau^{(a)})$ the assumed density of $\mathbf{X}_i$ conditional on $\mathbf{H}_i$, where $\tau^{(a)}$ is a vector of model parameters. The second component model is derived from the linear mixed effects model,

$$\mathbf{W}_i = \mathbf{D}_i\mathbf{X}_i + \mathbf{U}_i, \tag{1}$$

where $\mathbf{D}_i$ is an $m_i \times p$ $(m_i > p)$ design matrix of rank p, $\mathbf{U}_i = (U_{i1},\ldots, U_{im_i})^T$ is the vector of intra-subject errors distributed according to $N_{m_i}(\mathbf{0}, \sigma^2\mathbf{I}_{m_i})$, and $\mathbf{I}_{m_i}$ is the $m_i \times m_i$ identity matrix. The density of $\mathbf{W}_i$ given $\mathbf{X}_i$, $f_{\mathbf{W}_i|\mathbf{X}_i}(w_i|x_i; \sigma^2)$, is thus $N_{m_i}(\mathbf{D}_i\mathbf{X}_i, \sigma^2\mathbf{I}_{m_i})$. It is assumed that $\mathbf{Y}_i$ and $\mathbf{W}_i$ are independent given $\mathbf{X}_i$ and $\mathbf{H}_i$ (Carroll et al., 2006, Section 2.5).

Let $\mathbf{\Omega} = (\theta^T, \tau^{(a)T}, \sigma^2, \zeta^T)^T$ be the $d \times 1$ vector of all unknown parameters in the joint model. Inference on θ is of central interest. The MLE for $\mathbf{\Omega}$ maximizes the observed data likelihood, to which the contribution from subject i is given by, for $i = 1,\ldots, n$,

$$f_{\mathbf{Y}_i,\mathbf{W}_i|\mathbf{H}_i}(y_i, w_i|h_i; \mathbf{\Omega}) = \int f_{\mathbf{Y}_i|\mathbf{X}_i,\mathbf{H}_i}(y_i|x_i, h_i; \theta, \zeta) f_{\mathbf{W}_i|\mathbf{X}_i}(w_i|x_i; \sigma^2) f_{\mathbf{X}_i|\mathbf{H}_i}^{(a)}(x_i|h_i; \tau^{(a)}) dx_i. \tag{2}$$

For the SWAN data, the primary response is binary with $Y_i = 1$ indicating absence of osteopenia (bone mineral density above the 33rd percentile), and $Y_i = 0$ indicating presence, for $i = 1,\ldots, 632$. Li et al. (2004) analyzed these data and assumed a logistic model for Y_i,

$$\Pr(Y_i = 1|\mathbf{X}_i, \mathbf{H}_i) = \{1 + \exp(-\beta_0 - \beta_1^T\mathbf{X}_i - \beta_2^T\mathbf{H}_i)\}^{-1}, \tag{3}$$

where $\mathbf{H}_i$ includes covariates such as age and ethnicity indicator, and $\mathbf{X}_i = (X_{1i}, X_{2i})^T$ is a bivariate latent variable. The observed longitudinal process $\mathbf{W}_i$ is the recorded natural log of PDG over one menstrual cycle, the length of which is standardized to a reference of 28 days. Li et al. posited a piecewise linear mixed effects model for $\mathbf{W}_i$ given by $W_{ij} = X_{1i} + X_{2i}(t_{ij} - 1.4)_+ - 2X_{2j}(t_{ij} - 2.1)_+ + U_{ij}$, $i = 1,\ldots, 632, j = 1,\ldots, m_i$, where $u_+ = uI(u > 0)$, $I(\cdot)$ is the indicator function, t_{ij} is in units of 10 days, and $6 \le m_i \le 14$. Here, then, X_{1i} denotes the subject-specific natural log PDG up to day 14, and X_{2i} is the subject-specific "slope" of the symmetric rise (days 14–21) and fall (days 21–28) of natural log PDG over a standardized cycle. In this example, $\theta = (\beta_0, \beta_1^T, \beta_2^T)^T$, and there is no ζ in model (3).

For the ACTG 175 data, the response of interest is a time-to-event T_i, for $i = 1,\ldots, 2279$. Define $\mathbf{Y}_i = (V_i, \Delta_i)^T$, where $V_i = \min(T_i, C_i)$, C_i is the censoring time, and $\Delta_i = I(T_i \le C_i)$. Song et al. (2002), who analyzed these data, assumed that censoring, intra-subject errors, and timing of measurements are noninformative, and specified the first component model as the proportional hazards model (PHM)

$$\begin{aligned}\lambda_i(u|\mathbf{X}_i, H_i) \quad &= \lim_{du \to 0} du^{-1} \Pr(u \le T_i < u + du | T_i \ge u, \mathbf{X}_i, H_i) \\ &= \lambda_0(u) \exp\{\gamma(X_{1i} + X_{2i}u) + \eta H_i\},\end{aligned} \quad (4)$$

where $\lambda_0(u)$ is an unspecified baseline hazard function, H_i is a treatment indicator, and $\mathbf{X}_i = (X_{1i}, X_{2i})^T$ is a bivariate latent variable, with $(X_{1i} + X_{2i}u)$ representing the true post-12-week $\log_{10}$ CD4 count of subject i at time u. The observed post-12-week $\log_{10}$ CD4 count is given by, for $i = 1,\dots, 2279, j = 1,\dots, m_i$,

$$W_{ij} = X_{1i} + X_{2i}t_{ij} + U_{ij}. \quad (5)$$

The density of $\mathbf{Y}_i$ given $\mathbf{X}_i$ and H_i is

$$\begin{aligned}f_{Y_i|\mathbf{X}_i, H_i}(\boldsymbol{y}_i | \boldsymbol{x}_i, h_i; \theta, \lambda_0) &= [\lambda_0(V_i) \exp\{\gamma(x_{1i} + x_{2i}V_i) + \eta h_i\}]^{\Delta_i} \\ &\exp\left[-\int_0^{V_i} \lambda_0(u) \exp\{\gamma(x_{1i} + x_{2i}u) + \eta h_i\} du \right].\end{aligned}$$

In this example, $\boldsymbol{\theta} = (\gamma, \alpha)^T$, and $\lambda_0(u)$ can be viewed as the nuisance parameter $\boldsymbol{\zeta}$ in the first component model.

Throughout the article we assume both component models in the joint models are correctly specified, and we focus on the assumed latent variable model, $f_{\mathbf{X}_i|\mathbf{H}_i}^{(a)}(\boldsymbol{x}_i | \boldsymbol{h}_i; \tau^{(a)})$.

3. Expected Robustness

Consistency of the MLE is guaranteed when either $\sigma^2 = 0$ or the assumed random effect model is correct. Neither are likely to hold in practice, and thus the relevant issues are sensitivity of the MLE to the random effect model assumption and how to study the effects of model misspecification if they exist. Several authors (Song et al., 2002; Hsieh et al., 2006; and Rizopoulos et al., 2008) reported intriguing latent-model robustness under joint model setting. Hsieh et al. (2006) provided a heuristic explanation for this phenomenon. Rizopoulos et al. (2008) showed for survival models with finite dimensional parameter space that the score vector under the misspecified model is close to the correct score vector when m_i is large enough. In this section, we provide a new explanation for the robustness property of the MLE through the following result.

Theorem 1

Denote the ordinary least squares estimator for $\mathbf{X}_i$ by $\hat{\mathbf{X}}_{m_i}$, i.e., $\widehat{\mathbf{X}}_{m_i} = (\mathbf{D}_i^T \mathbf{D}_i)^{-1} \mathbf{D}_i^T \mathbf{W}_i$. The ratio of the density in (2) and the following expression,

$$f_{\mathbf{W}_i|\widehat{\mathbf{X}}_{m_i}}(\boldsymbol{w}_i | \widehat{\boldsymbol{x}}_{m_i}; \sigma^2) f_{Y_i|\widehat{\mathbf{X}}_{m_i}, \mathbf{H}_i}(\boldsymbol{y}_i | \widehat{\boldsymbol{x}}_{m_i}, \boldsymbol{h}_i; \theta, \zeta) f_{\mathbf{X}_i|\mathbf{H}_i}^{(a)}(\widehat{\boldsymbol{x}}_{m_i} | \boldsymbol{h}_i; \tau^{(a)}), \quad (6)$$

approaches one as the longitudinal information increases without bound.

The proof is given in Web Appendix A. The intuition of this result is that, when the longitudinal data information is rich enough, $\mathbf{X}_i$ can be well estimated by $\hat{\mathbf{X}}_{m_i}$ so that it is as if $\mathbf{X}_i$ were

observed like fixed effects instead of being latent quantities, and thus the dependence of likelihood inference on the assumed model for $\mathbf{X}_i$ weakens. Note in (6) that θ appears only in $f_{\mathbf{Y}_i|\hat{\mathbf{X}}_{m_i},\mathbf{H}_i}(\cdot)$. Consequently, the MLE derived from the likelihood based on (6) does not depend on $f_{\mathbf{X}_i|\mathbf{H}_i}^{(a)}(\cdot)$ and thus neither will the MLE based on (2) as the longitudinal information increases. The key issue in practice is knowing when the longitudinal information is great enough for the MLE to achieve a desired degree of robustness. We next describe an improved remeasurement method for assessing robustness of the MLE in a particular data set.

4. Diagnostic methods

4.1 Remeasurement Method (SIMEX)

The remeasurement method in Huang et al. (2006) is derived from the SIMEX method developed by Cook and Stefanski (1994) and Stefanski and Cook (1995), also described in Carroll et al. (2006, Chapter 5). To motivate our improved remeasurement method, we first review the remeasurement method of Huang et al. (2006) in the joint model context.

The remeasurement method involves further contaminating $\mathbf{W}_i$ and reestimating $\mathbf{\Omega}$ based on the contaminated-enhanced data. Specifically, for each prespecified positive constant λ:

- *Step 1*. For $b = 1,\ldots, B$, generate the bth λ-remeasured data set, denoted by $\{\mathbf{Q}_{b,i}(\lambda)\}_{i=1}^n$, where $\mathbf{Q}_{b,i}(\lambda)=\{\boldsymbol{Y}_i^T, \mathbf{W}_{b,i}(\lambda)^T, \mathbf{H}_i^T\}^T$, by taking

$$\mathbf{W}_{b,i}(\lambda)=\mathbf{W}_i+ \sqrt{\lambda}\sigma\mathbf{Z}_{b,i}, \tag{7}$$

  where $\mathbf{Z}_{b,i}$ are independent m_i-dimensional standard normal random errors, for $i = 1, \ldots, n$ and $b = 1,\ldots, B$.

- *Step 2*. Estimate the parameters based on $\{\mathbf{Q}_{b,i}(\lambda)\}_{i=1}^n$. Denote by $\hat{\boldsymbol{\theta}}_b(\lambda)$ the estimate for $\boldsymbol{\theta}$, and by $\hat{\mathbf{\Omega}}_b(\lambda)$ the entire estimated parameter vector, for $b = 1,\ldots, B$.

- *Step 3*. Compute $\widehat{\theta}_B(\lambda)=B^{-1}\sum_{b=1}^B \widehat{\theta}_b(\lambda)$. Similarly define $\widehat{\mathbf{\Omega}}_B(\lambda)=B^{-1}\sum_{b=1}^B \widehat{\mathbf{\Omega}}_b(\lambda)$.

- *Step 4*. Plot $\hat{\boldsymbol{\theta}}_B(\lambda)$ versus $\lambda \geq 0$, where $\hat{\boldsymbol{\theta}}_B(0) = \hat{\boldsymbol{\theta}}(0)$ is the estimate based on $\{\mathbf{Q}_i\}_{i=1}^n$. This plot is referred to as SIMEX plot.

A SIMEX plot where $\hat{\boldsymbol{\theta}}_B(\lambda)$ remains relatively constant across λ indicates robustness.

The above procedure has two drawbacks. First, the remeasured $\mathbf{W}_{b,i}(\lambda)$ defined in (7) depends on the unknown σ. Second, $\mathbf{\Omega}$ is estimated B times in step 2 in order to obtain $\hat{\mathbf{\Omega}}_B(\lambda)$, which is computationally burdensome. The improved remeasurement method we now propose overcomes both drawbacks.

First, to generate remeasured data free of parameters, we define

$$\mathbf{W}_{b,i}(\lambda)=\mathbf{W}_{1,bi}(\lambda)+\mathbf{W}_{2,i}(\lambda), \tag{8}$$

where

$$\mathbf{W}_{1,bi}(\lambda) = \mathbf{P}_{\mathbf{D}_i}\mathbf{W}_i + \sqrt{\lambda}\mathbf{D}_i(\mathbf{D}_i^T\mathbf{D}_i)^{-1/2}\mathbf{T}_{b,i}^T\mathbf{W}_i, \tag{9}$$

$$\mathbf{W}_{2,i}(\lambda) = \sqrt{1+\lambda}(\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\mathbf{W}_i, \tag{10}$$

$\mathbf{P}_{\mathbf{D}_i} = \mathbf{D}_i(\mathbf{D}_i^T\mathbf{D}_i)^{-1}\mathbf{D}_i^T$, $\mathbf{T}_{b,i} = (\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\mathbf{Z}_{b,i}\{\mathbf{Z}_{b,i}^T(\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\mathbf{Z}_{b,i}\}^{-1/2}$, and the elements in the $m_i \times p$ matrix $\mathbf{Z}_{b,i}$ are independent standard normal random variables. It can be shown that $\mathbf{Z}_{b,i}^T(\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\mathbf{Z}_{b,i}$ is positive definite almost surely when $m_i \geq 2p$ so that $\{\mathbf{Z}_{b,i}^T(\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\mathbf{Z}_{b,i}\}^{-1/2}$ exists almost surely. The construction of the new $\mathbf{W}_{b,i}(\lambda)$ in (8) is in the spirit of the *empirical SIMEX* discussed in Section 5.3.1.3 in Carroll et al. (2006). As elaborated in Section 4.2, $\mathbf{W}_{1,bi}(\lambda)$ is a suboptimal, normally distributed, unbiased estimator for $\mathbf{D}_i\mathbf{X}_i$, and $\mathbf{W}_{2,i}(\lambda)$ is a normal unbiased estimator for zero, with the combined variance-covariance matrix of $\mathbf{W}_{1,bi}(\lambda)$ and $\mathbf{W}_{2,i}(\lambda)$ equal to $(1+\lambda)\sigma^2\mathbf{I}_{m_i}$, which coincides with the variance-covariance matrix of the old $\mathbf{W}_{b,i}(\lambda)$ defined in (7).

Second, to avoid repeated estimation of $\mathbf{\Omega}$ using the remeasured data, we construct a new system of estimating equations at $\lambda > 0$. Assume that $\hat{\mathbf{\Omega}}(0)$ solves the vector estimating equation evaluated at the observed data given by

$$\sum_{i=1}^{n}\psi(\mathbf{Q}_i;\mathbf{\Omega}) = \mathbf{0}, \tag{11}$$

for some $d \times 1$ vector-valued function $\psi(\cdot; \mathbf{\Omega})$. The functional form of $\psi(\cdot; \mathbf{\Omega})$ depends on estimation procedure. We defer specification of $\psi(\cdot; \mathbf{\Omega})$ until Section 5 where specific joint models and target estimators are considered in simulation. Based on the remeasured data, we solve the following vector estimating equation evaluated at all B sets of λ-remeasured data for an estimator of $\mathbf{\Omega}$,

$$\sum_{i=1}^{n}\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}\} = \mathbf{0}, \tag{12}$$

where $\mathbf{Q}_i^{(B)}(\lambda) = \{\mathbf{Q}_{b,i}(\lambda)\}_{b=1}^{B}$, and $\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}\} = B^{-1}\sum_{b=1}^{B}\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}\}$, for $i = 1,\ldots,n$. Denote by $\tilde{\mathbf{\Omega}}_B(\lambda)$ the solution to (12) and by $\tilde{\boldsymbol{\theta}}_B(\lambda)$ the corresponding estimator for $\boldsymbol{\theta}$. Using $\tilde{\mathbf{\Omega}}_B(\lambda)$ in place of $\hat{\mathbf{\Omega}}_B(\lambda)$ in the remeasurement method is appealing for two reasons. First, while $\tilde{\mathbf{\Omega}}_B(\lambda)$ is obtained by solving only one vector estimating (12), $\hat{\mathbf{\Omega}}_B(\lambda)$ requires solving B vector estimating equations,

$$\sum_{i=1}^{n}\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}\} = \mathbf{0}, \quad b = 1,\ldots,B. \tag{13}$$

Second, the summand in (12) is usually "smoother" than that in (13), thus solving (12) is often easier than solving (13). To be consistent in notation, we define $\tilde{\mathbf{\Omega}}(0)$ as the estimator based on $\{\mathbf{Q}_i\}_{i=1}^{n}$, which is the same as $\hat{\mathbf{\Omega}}(0)$.

4.2 Equivalence Between Two Versions of the Remeasurement Method

The improved remeasurement method is more efficient computationally, and it still retains the key features necessary for diagnosing model misspecification. First, note that, for the old $\mathbf{W}_{b,i}(\lambda)$ defined in (7), one has $\mathbf{W}_{b,i}(\lambda)|\mathbf{X}_i \sim N_{m_i}\{\mathbf{D}_i\mathbf{X}_i, (1+\lambda)\sigma^2\mathbf{I}_{m_i}\}$, just like $\mathbf{W}_i|\mathbf{X}_i \sim N_{m_i}\{\mathbf{D}_i\mathbf{X}_i, \sigma^2\mathbf{I}_{m_i}\}$ except for the inflated variance, $(1+\lambda)\sigma^2$. This feature is important because it implies that the density of $\mathbf{Q}_{b,i}(\lambda)$ is identical to that of $\mathbf{Q}_i$ except for the measurement error variance. Therefore, if the observed data density given in (2) is correct, then replacing σ^2 with $(1+\lambda)\sigma^2$ in (2) gives the correct density of the λ-remeasured data. With the correct likelihood, consistent MLE for all sizes of λ is achieved, resulting in a constant SIMEX plot asymptotically. Conversely, a nonconstant SIMEX plot indicates model misspecification.

We show now that the new $\mathbf{W}_{b,i}(\lambda)$ defined in (8) has the same feature as that of the old $\mathbf{W}_{b,i}(\lambda)$. Because $\mathbf{W}_i|\mathbf{X}_i \sim N_{m_i}(\mathbf{D}_i\mathbf{X}_i, \sigma^2\mathbf{I}_{m_i})$, it is obvious by (10) that

$$\mathbf{W}_{2,i}(\lambda)|\mathbf{X}_i \sim N_{m_i}\{\mathbf{0}, \ (1+\lambda)\sigma^2(\mathbf{I}_{m_i} - \mathbf{P}_{\mathbf{D}_i})\}. \tag{14}$$

To derive the distribution of $\mathbf{W}_{1,bi}(\lambda)$ given $\mathbf{X}_i$, we first consider the distribution of $\mathbf{W}_{1,bi}(\lambda)$ given $\mathbf{T}_{b,i}$. By (9), $\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i} \sim N_{m_i}[E\{\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i}\}, \text{var}\{\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i}\}]$, where, by noting that $\mathbf{T}_{b,i}^T\mathbf{D}_i=\mathbf{0}$,

$$E\{\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i}\}=(\mathbf{P}_{\mathbf{D}_i} + \sqrt{\lambda}\mathbf{D}_i\mathbf{L}_i\mathbf{T}_{b,i}^T)\mathbf{D}_i\mathbf{X}_i=\mathbf{D}_i\mathbf{X}_i,$$

in which $\mathbf{L}_i=(\mathbf{D}_i^T\mathbf{D}_i)^{-1/2}$ such that $\mathbf{L}_i\mathbf{L}_i^T=(\mathbf{D}_i^T\mathbf{D}_i)^{-1}$. Then by realizing that $\mathbf{T}_{b,i}^T\mathbf{P}_{\mathbf{D}_i}=\mathbf{0}$, and $\mathbf{T}_{b,i}^T\mathbf{T}_{b,i}=\mathbf{I}_p$, we have

$$\text{var}\{\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i}\}=(\mathbf{P}_{\mathbf{D}_i} + \sqrt{\lambda}\mathbf{D}_i\mathbf{L}_i\mathbf{T}_{b,i}^T)\sigma^2\mathbf{I}_{m_i}(\mathbf{P}_{\mathbf{D}_i} + \sqrt{\lambda}\mathbf{T}_{b,i}\mathbf{L}_i^T\mathbf{D}_i^T)=(1+\lambda)\sigma^2\mathbf{P}_{D_i}.$$

That is, $\mathbf{W}_{1,bi}(\lambda)|\mathbf{T}_{b,i} \sim N_{m_i}\{\mathbf{D}_i\mathbf{X}_i(1+\lambda)\sigma^2\mathbf{P}_{D_i}\}$, and thus

$$\mathbf{W}_{1,bi}(\lambda)|\mathbf{X}_i \sim N_{m_i}\{\mathbf{D}_i\mathbf{X}_i, \ (1+\lambda)\sigma^2\mathbf{P}_{\mathbf{D}_i}\}. \tag{15}$$

Lastly, straightforward algebra reveals that, given $\mathbf{X}_i$, cov $\{\mathbf{W}_{1,bi}(\lambda), \mathbf{W}_{2,i}(\lambda)\} = \mathbf{0}$. Combining (14) and (15), we have $\mathbf{W}_{b,i}(\lambda)|\mathbf{X}_i \sim N_{m_i}\{\mathbf{D}_i\mathbf{X}_i, (1+\lambda)\sigma^2\mathbf{I}_{m_i}\}$, as desired. The new definition of $\mathbf{W}_{b,i}(\lambda)$ given in (8) is assumed in the sequel.

Second, we prove that $\hat{\mathbf{\Omega}}_B(\lambda)$ and $\tilde{\mathbf{\Omega}}_B(\lambda)$ defined in Section 4.1 are asymptotically equivalent. Assume that the vector equation

$$E[\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}(\lambda)\}]=\mathbf{0} \tag{16}$$

uniquely defines $\mathbf{\Omega}(\lambda)$, where the expectation is taken with respect to the true density of $\mathbf{\Omega}_{b,i}(\lambda)$. Recall that $\widehat{\mathbf{\Omega}}_b(\lambda)$ is the solution to (13), for $b = 1,\ldots, B$. A first-order Taylor expansion of (13) around $\mathbf{\Omega}(\lambda)$ and rearrangement of terms gives

$$n^{1/2}\left\{\widehat{\mathbf{\Omega}}_b(\lambda) - \mathbf{\Omega}(\lambda)\right\} = n^{-1/2}\mathbf{A}_1^{-1}\{\mathbf{\Omega}(\lambda)\}\sum_{i=1}^{n}\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}(\lambda)\}+o_p(1), \tag{17}$$

for $b = 1,\ldots, B$, where $\mathbf{A}_1\{\mathbf{\Omega}(\lambda)\}$ is equal to $E[-\partial\psi\{\mathbf{Q}_{b,i}(\lambda); \mathbf{\Omega}\}/\partial\mathbf{\Omega}^T]$ evaluated at $\mathbf{\Omega}(\lambda)$, and the expectation is taken with respect to the true density of $\mathbf{Q}_{b,i}(\lambda)$. Averaging (17) over $b = 1, \ldots, B$ for any finite B gives

$$\begin{aligned}n^{1/2}\left\{\widehat{\mathbf{\Omega}}_B(\lambda) - \mathbf{\Omega}(\lambda)\right\} &= n^{-1/2}\mathbf{A}_1^{-1}\{\mathbf{\Omega}(\lambda)\}\sum_{i=1}^{n}B^{-1}\sum_{b=1}^{B}\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}(\lambda)\}+o_p(1)\\ &= n^{-1/2}\mathbf{A}_1^{-1}\{\mathbf{\Omega}(\lambda)\}\sum_{i=1}^{n}\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}(\lambda)\}+o_p(1).\end{aligned} \tag{18}$$

Next consider the vector equation that uniquely defines $\mathbf{\Omega}^*(\lambda)$,

$$E[\psi\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}^*(\lambda)\}]=\mathbf{0}, \tag{19}$$

where the expectation is taken with respect to the true density of $\mathbf{Q}_i^{(B)}(\lambda)$. Because

$$E[\psi\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}\}]=B^{-1}\sum_{b=1}^{B}E[\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}\}]=E[\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}\}],$$

for any b and i, the solution to (16), $\mathbf{\Omega}(\lambda)$, also solves (19). By the uniqueness of the solution to (19), $\mathbf{\Omega}^*(\lambda) = \mathbf{\Omega}(\lambda)$. A first-order Taylor expansion of (12) around $\mathbf{\Omega}^*(\lambda)(= \mathbf{\Omega}(\lambda))$ gives

$$n^{1/2}\left\{\tilde{\mathbf{\Omega}}_B(\lambda) - \mathbf{\Omega}(\lambda)\right\} = n^{-1/2}\mathbf{A}_2^{-1}\{\mathbf{\Omega}(\lambda)\}\sum_{i=1}^{n}\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}(\lambda)\}+o_p(1), \tag{20}$$

where

$$\begin{aligned}\mathbf{A}_2\{\mathbf{\Omega}(\lambda)\} &= E[-\partial\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda);\mathbf{\Omega}\}/\partial\mathbf{\Omega}^T]|_{\mathbf{\Omega}=\mathbf{\Omega}(\lambda)}\\ &= B^{-1}\sum_{b=1}^{B}E[-\partial\psi\{\mathbf{Q}_{b,i}(\lambda);\mathbf{\Omega}\}/\partial\mathbf{\Omega}^T]|_{\mathbf{\Omega}=\mathbf{\Omega}_{(\lambda)}}\\ &= \mathbf{A}_1\{\mathbf{\Omega}(\lambda)\}.\end{aligned}$$

Finally, subtracting (20) from (18) reveals that $n^{1/2}\{\widehat{\Omega}_B(\lambda) - \tilde{\Omega}_B(\lambda)\} \xrightarrow{p} \mathbf{0}$ as $n \to \infty$.

4.3 Test of robustness

The SIMEX plot is a convenient graphical tool to visually assess latent-model robustness. However, due to the variation in the estimators, (non)robustness is not always evident from the SIMEX plot. We now define two test statistics to objectively assess robustness.

For a vector (or a square matrix) $\mathbf{\Pi}$, denote by $[\mathbf{\Pi}]_{(k)}$ the kth element (or diagonal element) of $\mathbf{\Pi}$. Analogous to the test statistic proposed in Huang et al. (2006), we define a test statistic to assess latent-model robustness based on the improved remeasurement method as

$$t_1^*(\lambda) = n^{1/2}\left\{\tilde{\Omega}_B(\lambda) - \tilde{\Omega}(0)\right\}_{(k)} / \sqrt{[\hat{v}_1]_{(k)}},$$

for $1 \le k \le d$, where $\hat{v}_1$ is an estimator for the variance-covariance matrix of $n^{1/2}\{\tilde{\Omega}_B(\lambda) - \tilde{\Omega}(0)\}$. A second test statistic we propose is defined by

$$t_2^*(\lambda) = n^{-1/2}\left[\sum_{i=1}^{n} \psi^{(B)}\left\{\mathbf{Q}_i^{(B)}(\lambda); \tilde{\Omega}_{-\sigma^2}(0), (1+\lambda)\tilde{\sigma}^2(0)\right\}\right]_{(k)} / \sqrt{[\hat{v}_2]_{(k)}},$$

where $\tilde{\Omega}_{-\sigma^2}(0)$ is $\tilde{\Omega}(0)$ excluding σ^2, and $\hat{v}_2$ is an estimator for the variance-covariance matrix of $n^{-1/2}\sum_{i=1}^{n}\psi^{(B)}\{\mathbf{Q}_i^{(B)}(\lambda); \tilde{\Omega}_{-\sigma^2}(0), (1+\lambda)\tilde{\sigma}^2(0)\}$. Note that, unlike $t_1^*(\lambda)$, computing $t_2^*(\lambda)$ does not require estimating Ω at $\lambda > 0$.

Define by $\Omega_{-\sigma^2}$ the parameter vector Ω excluding σ^2. Both test statistics are motivated by the fact that, if the estimators for $\Omega_{-\sigma^2}$ are robust, then $\Omega_{-\sigma^2}(\lambda) = \Omega_{-\sigma^2}(0)$ for $\lambda > 0$, and both test statistics should center at zero. The derivations for $\hat{v}_1$ and $\hat{v}_2$ are given in Web Appendix B. We also show in Web Appendix C that $t_1^*(\lambda)$ and $t_2^*(\lambda)$ are asymptotically equivalent for assessing robustness.

5. Simulation studies

5.1 Joint models with simple endpoint

We first demonstrate the proposed diagnostic methods applied to joint models with simple endpoint. A data set of size $n = 500$ is generated from a joint model with a binary response. The first component model is a logistic model, $\Pr(Y_i = 1 | \mathbf{X}_i) = \{1 + \exp(-\beta_0 - \boldsymbol{\beta}_1 \mathbf{X}_i)\}^{-1}$, where $\mathbf{X}_i = (X_{1i}, X_{2i})^T$, and $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T$. The true values of the primary regression parameters $\theta = (\beta_0, \beta_i^T)^T$ are $(-2, 1, 1)^T$. The latent variable $\mathbf{X}_i$ is generated from a location mixture bivariate normal (BVN), $(1-p)N_2(\boldsymbol{\delta}, \mathbf{I}_2) + pN_2(\mathbf{0}, \mathbf{I}_2)$, where $p = 0.4$ and $\boldsymbol{\delta} = (5, 0)^T$. The longitudinal measures $\mathbf{W}_i$ are generated according to (1), with $m_i = 5$, $t_{ij} = j$ for $j = 1, \ldots, 5$, $\mathbf{D}_i$ 5×2 with jth row equal to $(1, j)$, and $\mathbf{U}_i \sim N_5(\mathbf{0}, 0.6\mathbf{I}_5)$, for $i = 1, \ldots, 500$.

We consider four estimators for θ. One is the CSE derived in Li et al. (2004). The other three are the MLEs when the assumed models for $\mathbf{X}$ are a two-component location mixture BVN; a non-mixture BVN; and a model specified by the bivariate second-order SNP density (Zhang and Davidian, 2001) given by $f_{\mathbf{X}}^{(a)}(\boldsymbol{x}; \tau^{(a)}) = P_2^2\{\mathbf{R}^{-1}(\boldsymbol{x}_i - \mu)\}\varphi\{\mathbf{R}^{-1}(\boldsymbol{x}_i - \mu)\}|\mathbf{R}|^{-1}$, where

$P_2(z) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{20}z_1^2 + a_{11}z_1z_2 + a_{02}z_2^2$ for $z = (z_1, z_2)^T$, and the polynomial coefficients in $P_2(z)$ are constrained so that $f_{\mathbf{X}}^{(a)}(\boldsymbol{x}; \tau^{(a)})$ integrates to one. Among the four estimators, the CSE is robust by construction (Li et al., 2004), as is the MLE based on a mixture BVN, the correct model for $\mathbf{X}$. The other two MLEs are suspect, as the assumed random effect models are incorrect. We use the proposed diagnostic devices to evaluate the robustness of the estimators. The function $\boldsymbol{\psi}(\cdot; \boldsymbol{\Omega})$ in (11) associated with the CSE is the conditional score defined in Li et al. (2004); and $\boldsymbol{\psi}(\cdot; \boldsymbol{\Omega})$ associated with the MLE is given by

$$\psi(Y_i, \mathbf{W}_i; \boldsymbol{\Omega}) = \begin{pmatrix} \frac{1}{m_i-2}\mathbf{W}_i^T\{\mathbf{I}_{m_i} - \mathbf{D}_i(\mathbf{D}_i^T\mathbf{D}_i)^{-1}\mathbf{D}_i^T\}\mathbf{W}_i - \sigma^2 \\ \frac{\partial \log f_{\mathbf{W}_i}(\boldsymbol{w}_i; \tau^{(a)}, \sigma^2)}{\partial \tau^{(a)}} \\ \frac{\partial \log f_{Y_i, \mathbf{W}_i}(y_i, \boldsymbol{w}_i; \theta, \tau^{(a)}, \sigma^2)}{\partial \theta} \end{pmatrix},$$

where $f_{\mathbf{W}_i}(\boldsymbol{w}_i; \boldsymbol{\tau}^{(a)}, \sigma^2)$ is the marginal density of $\mathbf{W}_i$.

We first implement the improved remeasurement method on one simulated data set with $B = 50$ and $\lambda \in [0, 1]$ to construct SIMEX plots. Denote the four estimators as $\tilde{\theta}_B^{(c)}(\lambda), \tilde{\theta}_B^{(m)}(\lambda), \tilde{\theta}_B^{(n)}(\lambda)$, and $\tilde{\theta}_B^{(s)}(\lambda)$, where the superscript identifies the estimator: c, CSE; m, mixture BVN; n, BVN; s, SNP. Figure 1a and b contain the SIMEX plots of the first two elements in $\boldsymbol{\theta}$ for each of the four estimates. As expected, $\tilde{\theta}_B^{(c)}(\lambda)$ and $\tilde{\theta}_B^{(m)}(\lambda)$ appear to be robust as reflected by the nearly constant SIMEX plots. The estimate resulting from the flexible SNP modeling $\tilde{\theta}_B^{(s)}(\lambda)$ also has a relatively flat SIMEX plot. However, the SIMEX plot of $\tilde{\theta}_B^{(n)}(\lambda)$, which is based on the least flexible assumed model for $\mathbf{X}$ among all the considered models, is clearly distinguished from the other three. In order to observe the typical trend in SIMEX plots, we repeat this experiment 30 times and construct the average SIMEX plots. These appear in Figure 1c and d. Note the similarity with Figure 1a and b.

To assess the robustness objectively, we present $t_1^*(1)$ and $t_2^*(1)$ in Table 1 for the four types of estimators depicted in Figure 1a and b. In Table 1, the pattern of p-values is consistent with the visual impressions of Figure 1a and b. The operating characteristics of $t_1^*(\lambda)$ based on the improved remeasurement method are similar to those based on the original remeasurement method of Huang et al. (2006). To examine the operating characteristics of $t_2^*(\lambda)$, we compute $t_2^*(1)$ associated with $\tilde{\theta}_B^{(c)}(\lambda), \tilde{\theta}_B^{(m)}(\lambda)$, and $\tilde{\theta}_B^{(n)}(\lambda)$, respectively, for 500 replicate data sets generated from the same joint model as above. The percentages of $|t_2^*(1)|$ values exceeding $t_{0.975}(n - d)$ are presented in Table 2. The results of $t_2^*(1)$ for $\tilde{\theta}_B^{(c)}(\lambda)$ and $\tilde{\theta}_B^{(m)}(\lambda)$ indicate reasonable size of $t_2^*(1)$. The results of $t_2^*(1)$ associated with $\tilde{\beta}_{11,B}^{(n)}$ and $\tilde{\beta}_{12,B}^{(n)}$ suggest promising power. In combination, these results suggest that $t_2^*(\lambda)$ provides power for detecting the effects of latent model misspecification, while maintaining reasonable size.

5.2 Joint models with time-to-event endpoint

We now study the diagnostic methods on a joint model with possibly censored time-to-event endpoint. Each simulated data set has $n = 500$ subjects. The time-to-event is generated according to a PHM given by $\lambda_i(u|\mathbf{X}_i) = \lambda_0(u) \exp\{\gamma(X_{1i} + X_{2i}u)\}$, with $\gamma = -1$ and $\lambda_0(u) = I(u \geq 16)$. The bivariate latent variable $\mathbf{X}_i = (X_{1i}, X_{2i})^T$ is generated from a truncated BVN obtained by first generating $\mathbf{X}_i$ from a BVN with $E(\mathbf{X}_i) = (4.173, -0.0103)^T$, and $\{\text{var}(X_{1i}), \text{cov}(X_{1i},$

X_{2i}), var(X_{1i})} = (4.96, −0.0456, 0.012), then discarding the realizations with negative γX_{2i}. This causes around 46% truncation of the original BVN. The censoring distribution is exponential with mean 110, resulting in a censoring rate of around 25%. The longitudinal measures W_{ij} are generated according to (5) at times $t_{ij} = (0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80)$, with a 10% missingness rate at times $u \geq 16$. On average there are around six repeated measures for each subject under this configuration. The intra-subject error variance is $\sigma^2 = 0.15$.

Using the superscript convention introduced in Section 5.1, we consider three MLEs in this simulation, $\tilde{\gamma}_B^{(m)}$, $\tilde{\gamma}_B^{(n)}$, and $\tilde{\gamma}_B^{(s)}$, where the assumed SNP model is of first order; and the MLEs are obtained via the EM algorithm as described in Wulfsohn and Tsiatis (1997) and Song et al. (2002). The function $\psi(\cdot; \boldsymbol{\Omega})$ in (11) is the likelihood score in this case. Because it is often very time-consuming to estimate the parameters in the setting of joint models with time-to-event endpoint, we only compute $t_2^*(\lambda)$ to assess robustness. For a data set generated from the current joint model, the values of $t_2^*(1)$, with the associated p-values in the following parentheses, are found to be: $\tilde{\gamma}_B^{(m)}$, −1.56 (0.12); $\tilde{\gamma}_B^{(n)}$, −1.81 (0.07); and $\tilde{\gamma}_B^{(s)}$, −1.15 (0.25). As in the previous simulation, $\tilde{\gamma}_B^{(n)}$ exhibits the greatest evidence of non-robustness, although falling short of 0.05 level of significance. These results agree with the observations in Song et al. (2002) and Hsieh et al. (2006) under similar simulation settings. This is an example where the longitudinal information is great enough to yield the MLEs relatively insensitive to random effect model assumptions. Among 100 Monte Carlo (MC) replicates, the proportions of data sets that yield significant $t_2^*(1)$ are, 0.08, 0.12, and 0.06 for $\tilde{\gamma}_B^{(m)}$, $\tilde{\gamma}_B^{(n)}$, and $\tilde{\gamma}_B^{(s)}$, respectively, which suggests some gain in robustness from flexible modeling on $\mathbf{X}$ when the true model deviates from normal.

One complication arises when computing the proposed test statistics for joint models with time-to-event endpoint due to the dimensionality of $\boldsymbol{\Omega}$. Strictly speaking, the nuisance parameter ζ in the first component model is the baseline hazard function, $\lambda_0(u)$, which is infinite dimensional. Because the observed data likelihood is maximized when $\lambda_0(u) = 0$ at non-event time u (Song et al., 2002), we define $\zeta = \{\lambda_0(u_1), \ldots, \lambda_0(u_L)\}^T$, where $(u_1, \ldots, u_L)$ is the set of observed times-to-event, and L is the number of distinct times-to-event in the data set. This treatment of ζ yields a finite yet large dimension of $\boldsymbol{\Omega}$, since L is usually large. As shown in Web Appendix B, computing $t_1^*(\lambda)$ and $t_2^*(\lambda)$ involves the $d \times 1$ score vector and $d \times d$ Hessian matrix. It is formidable to implement the computation when d is large. Our current solution to this computational obstacle is to drop ζ from the parameter space when computing the score or Hessian. The tradeoff is that the variance estimators, $\hat{v}_1$ and $\hat{v}_2$, may be biased downward. The extent of underestimation depends on model configuration. For instance, in the foregoing simulation with 100 MC replicates, the ratio of the average of $\sqrt{\hat{v}_2}$ over the empirical standard deviation of the numerator of $t_2^*(1)$ are 0.94, 0.97, and 1.00 associated with $\tilde{\gamma}_B^{(m)}$, $\tilde{\gamma}_B^{(n)}$, and $\tilde{\gamma}_B^{(s)}$, respectively, with standard error (estimated via the jackknife method) around 0.07 for each ratio. To compare with the variance estimators when there is no such complication, we summarize in Table 3 the ratio of the mean of $\sqrt{\hat{v}_l}$ averaging across 100 MC replicates over the empirical standard deviation of the numerator of $t_l^*(1)$ from the simulation in Section 5.1, for $l = 1, 2$. The results in Table 3 indicate that $\hat{v}_1$ and $\hat{v}_2$ are reasonably reliable variance estimator in the setting of joint models with simple endpoint.

Due to the complication in variance estimators in joint models with time-to-event end-point, even though an insignificant value of $t_1^*(\lambda)$ or $t_2^*(\lambda)$ still indicates lack of evidence for

nonrobustness, one should be cautious when interpreting significant values of the test statistics. In that case, one needs to explore further whether or not the significant results are caused by overoptimistic variance estimators. For instance, one can use the bootstrap procedure to obtain a more reliable variance estimator, as outlined in Hsieh et al. (2006).

6. Application to SWAN and ACTG 175

6.1 SWAN

We now apply the diagnostic methods to the SWAN data. For simplicity, we exclude the observable covariates $\mathbf{H}_i$ from the first component model for the simple endpoint in (3) and posit the logistic model given by $\Pr(Y_i = 1|\mathbf{X}_i) = \{1 + \exp(-\beta_0 - \boldsymbol{\beta}_1 \mathbf{X}_i)\}^{-1}$. Three estimators for the primary regression parameter $\theta = (\beta_0, \beta_1^T)^T$ are considered, including the CSE, the MLE when assuming $\mathbf{X}_i$ follows a two-component location mixture BVN, and the MLE resulting from a BVN assumed model for $\mathbf{X}_i$. We compute $t_1^*(1)$ and $t_2^*(1)$ with B = 100 to assess the robustness of these three estimators. The resulting test statistics are presented in Table 4. The SIMEX plots for these three sets of estimates are given in Web Appendix D.

The statistics $t_1^*(1)$ indicate little evidence of nonrobustness for any of the three estimators for $\boldsymbol{\theta}$, which is also reflected by the SIMEX plots in Web Appendix D. The statistics $t_2^*(1)$ do not suggest strong evidence of nonrobustness either, but the values of $t_2^*(1)$ associated with $\tilde{\beta}_{0,B}^{(n)}$ and $\tilde{\beta}_{12,B}^{(n)}$ are much closer to being significant than those for the counterpart estimates in $\tilde{\theta}_B^{(c)}$ and $\tilde{\theta}_B^{(m)}$. Li et al. (2004) found that the estimated density for $\mathbf{X}_i$ "does not deviate considerably from multivariate normality." Their finding may explain why our diagnostic tools do not find strong evidence that $\tilde{\theta}_B^{(n)}$ is not robust.

6.2 ACTG 175

We now consider the ACTG 175 data with 2279 subjects and 350 events. This clinical trial found zidovudine alone to be an inferior treatment compared to the other three therapies, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine alone. We assume the PHM in (4) where $H_i = I$(treatment $\neq$ zidovudine for subject i). There is an average of 8.28 CD4 measurements per subject in this data set.

We compute $t_2^*(1)$ with $B = 30$ associated with three MLEs for $\boldsymbol{\theta} = (\gamma, \alpha)^T$, with assumed models for $\mathbf{X}_i$ as a two-component location mixture BVN, BVN, and the first-order SNP, respectively. The resulting statistics $t_2^*(1)$ are, for γ: $\tilde{\gamma}^{(m)}$, 1.37 (0.17); $\tilde{\gamma}^{(n)}$, 1.73 (0.08); and $\tilde{\gamma}^{(s)}$, 1.77 (0.08); for α: $\tilde{\alpha}^{(m)}$, 0.98 (0.32); $\tilde{\alpha}^{(n)}$, 1.32 (0.19); and $\tilde{\alpha}^{(s)}$ 0.42 (0.67). Therefore, there is not sufficient evidence to imply nonrobustness of the MLEs for $\boldsymbol{\theta}$ under any of the three assumed random effect models. This reconciles with the findings in Song et al. (2002).

7. Discussion

We have presented a graphical method and two test statistics for diagnosing latent-model robustness in joint models for a primary endpoint and a longitudinal process. The methods are designed to reveal sensitivity of the target estimator to model assumptions on the random effects in joint models. With these diagnostic tools, it is hopeful to find an appropriate and parsimonious random effect model to implement parametric inference as opposed to semipara-metric inference as in Li et al. (2004) and Song et al. (2002), which can be less eficient. Our diagnostic methods are closely related to the SIMEX method. Many authors (e.g., Li and Lin,

2003;Greene and Cai, 2007; He, Yi, and Xiong, 2007) used SIMEX for estimating regression parameters when covariates in survival models are measured with error, which is in line with the initial motivation of SIMEX developed in the framework of structural measurement error models. Our use of SIMEX is a new application of it as we do not use it for parameter estimation per se but mainly for assessing latent-model robustness.

As noted in Section 5.2, the variance estimators in the test statistics for joint models with time-to-event endpoint can be overly optimistic. More refined variance estimators for constructing the test statistics to assess latent-model robustness in these complicated joint models call for further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement Error in Nonlinear Models: A Modern Perspective. London: Chapman & Hall; 2006.

Cook J, Stefanski LA. Simulation extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association 1994;89:1314–1328.

Greene WF, Cai J. Measurement error in covariates in the marginal hazards model for multivariate failure time data. Biometrics 2004;60:987–996. [PubMed: 15606419]

Hammer SM, Katezstein DA, Hughes MD, Gundaker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. for the AIDS Clinical Group Study 175 Study Team. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine 1996;335:1081–1089. [PubMed: 8813038]

He W, Yi GY, Xiong J. Accelerated failure time models with covariates subject to measurement error. Statistics in Medicine 2007;26:4817–4832. [PubMed: 17436310]

Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. Biometrics 2006;62:1037–1043. [PubMed: 17156277]

Huang X, Stefanski LA, Davidian M. Latent-model robustness in structural measurement error models. Biometrika 2006;93:53–64.

Li Y, Lin X. Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. Journal of the American Statistical Association 2003;98:191–203.

Li E, Zhang D, Davidian M. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal parameters. Biometrics 2004;60:1–7. [PubMed: 15032767]

Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. Biometrika 2008;95:1–12.

Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. Biometrics 2002;58:742–753. [PubMed: 12495128]

Sowers MR, Finkelstein J, Ettinger B, Bondarenko I, Neer R, Cauley J, Sherman S, Greendale G. The association of endogenous hormone concentrations and bone mineral density measures in pre- and peri-menopausal women of four ethnic groups: SWAN. Osteoporosis International 2003;14:44–52. [PubMed: 12577184]

Stefanski LA, Cook J. Simulation extrapolation: The measurement error jackknife. Journal of the American Statistical Association 1995;90:1247–56.

Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika 2001;88:447–458.

Wang CY, Wang N, Wang S. Regression analysis when covariates are regression parameters of a random effects models for observed longitudinal measurements. Biometrics 2000;56:487–495. [PubMed: 10877308]

Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics 1997;53:330–339. [PubMed: 9147598]

Zhang D, Davidian M. Linear mixed model with flexible distribution of random effects for longitudinal data. Biometrics 2001;57:795–802. [PubMed: 11550930]
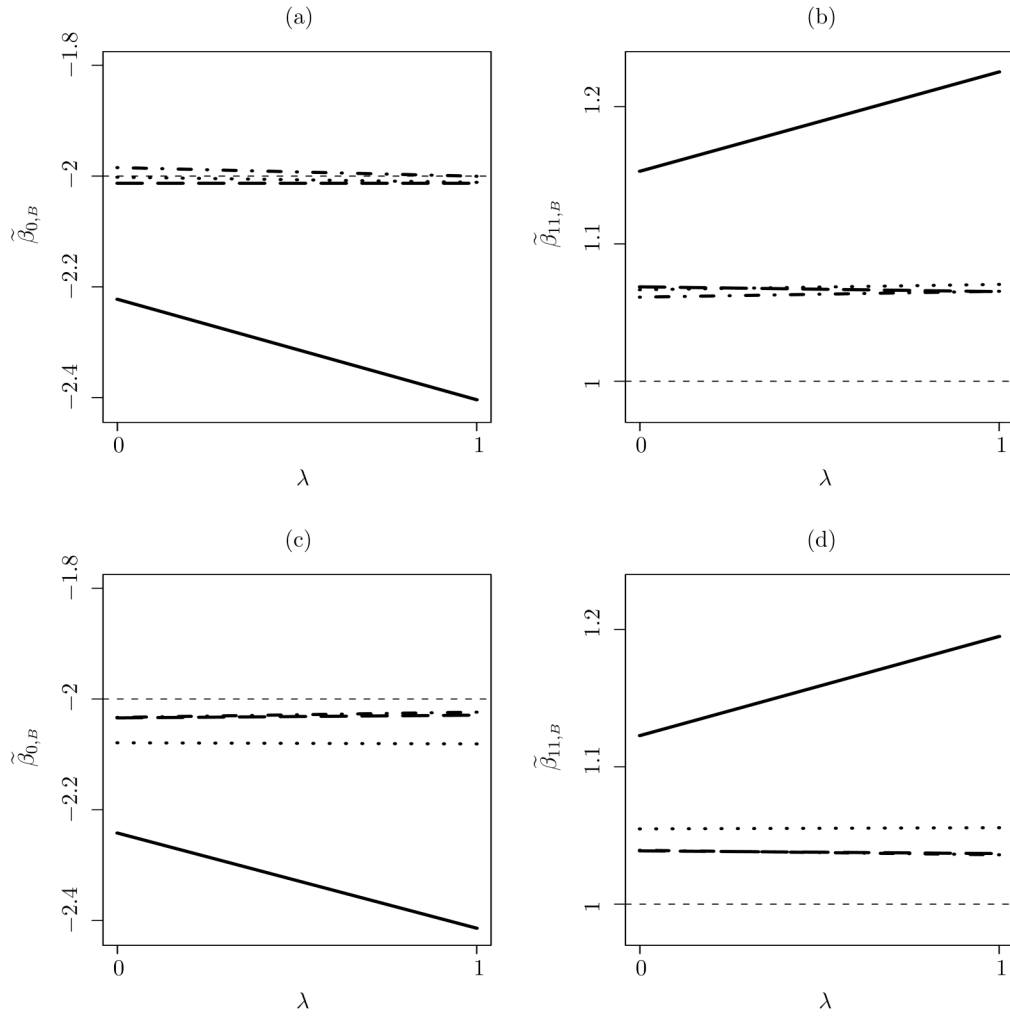
Figure 1.
Plots (a) and (b) are SIMEX plots for the MLEs of the first two elements in each of $\tilde{\theta}_B^{(c)}(\lambda), \tilde{\theta}_B^{(m)}(\lambda), \tilde{\theta}_B^{(n)}(\lambda)$, and $\tilde{\theta}_B^{(s)}(\lambda)$, computed from one simulated data set. Plots (c) and (d) are the average SIMEX plots from 30 Monte Carlo replicates. The line types are, $\tilde{\theta}_B^{(c)}(\lambda)$: long dashed; $\tilde{\theta}_B^{(m)}(\lambda)$: dash-dotted; $\tilde{\theta}_B^{(n)}(\lambda)$: solid; and $\tilde{\theta}_B^{(s)}(\lambda)$: dotted. The short dashed lines are the reference lines at the true values, $\beta_0 = -2$ and $\beta_{11} = 1$. The ranges of the vertical axes in (a) and (b) are set to be one estimated standard deviation of $\tilde{\theta}^{(n)}(0)$ below and above the average of the four types of estimates at $\lambda = 0$.

Table 1

Statistics $t_1^*(1)$ and $t_2^*(1)$ associated with $\tilde\theta_B^{(c)}$, $\tilde\theta_B^{(m)}$, $\tilde\theta_B^{(n)}$, and $\tilde\theta_B^{(s)}$ as depicted in Figure 1a and b from the simulation in Section 5.1. The numbers in parentheses are the *p*-values associated with the statistics.

Statistic	Parameter	CSE	Mixture BVN-MLE	BVN-MLE	SNP-MLE
$t_1^\square(1)$	$\beta_{0,B}$	0.001 (0.999)	0.80 (0.42)	3.53 (<0.001)	0.43 (0.67)
	$\beta_{11,B}$	0.46 (0.64)	−0.51 (0.61)	−3.37 (<0.001)	−0.42 (0.67)
	$\beta_{12,B}$	0.73 (0.47)	1.00 (0.32)	−0.26 (0.80)	1.83 (0.07)
$t_2^\square(1)$	$\beta_{0,B}$	−0.95 (0.34)	−0.53 (0.59)	−0.70 (0.48)	0.01 (0.99)
	$\beta_{11,B}$	−0.59 (0.56)	0.54 (0.59)	3.24 (0.001)	0.11 (0.91)
	$\beta_{12,B}$	−0.33 (0.74)	−1.47 (0.14)	−2.74 (0.006)	−0.21 (0.83)

Percentage of $t_2^*(1)$ that exceed $t_{0.975}(n-d)$ in absolute value among the 500 replicate data sets from the simulation in Section 5.1.

Parameter	CSE	Mixture BVN-MLE	BVN-MLE
β_0	0.06	0.04	0.06
β_{11}	0.05	0.04	0.95
β_{12}	0.03	0.05	0.56

Table 3

Ratio of the average of $\sqrt{\hat{\nu}_l}$ over the empirical standard deviation of the numerator of $t_l^*(1)$ from 100 MC replicates, for $l = 1, 2$, associated with $\tilde{\theta}_B^{(c)}$, $\tilde{\theta}_B^{(m)}$, and $\tilde{\theta}_B^{(n)}$ from the simulation in Section 5.1. The numbers in parentheses are the jackknife estimates for the standard errors of the ratios.

Statistic	Parameter	CSE	Mixture BVN-MLE	BVN-MLE
$\sqrt{\hat{\nu}_1}$	β_0	1.00 (0.07)	1.02 (0.08)	1.00 (0.09)
	β_{11}	0.99 (0.07)	1.04 (0.07)	1.00 (0.09)
	β_{12}	0.98 (0.06)	1.00 (0.07)	1.00 (0.07)
$\sqrt{\hat{\nu}_2}$	β_0	1.07 (0.07)	1.03 (0.06)	0.99 (0.06)
	β_{11}	1.02 (0.07)	1.05 (0.07)	1.01 (0.06)
	β_{12}	1.07 (0.08)	1.04 (0.07)	1.01 (0.07)

Table 4

Statistics $t_1^*(1)$ and $t_2^*(1)$ associated with $\overset{\sim(c)}{\theta_B}$, $\overset{\sim(m)}{\theta_B}$, and $\overset{\sim(n)}{\theta_B}$ for the SWAN data. The numbers in parentheses are the *p*-values associated with the statistics.

Statistic	Parameter	CSE	Mixture BVN-MLE	BVN-MLE
$t_1^{\square}(1)$	β_0	−0.18 (0.86)	−0.70 (0.48)	0.64 (0.52)
	β_{11}	−0.12 (0.90)	−0.67 (0.50)	0.57 (0.57)
	β_{12}	0.17 (0.87)	0.68 (0.50)	−0.82 (0.41)
$t_2^{\square}(1)$	β_0	0.42 (0.67)	0.87 (0.38)	1.64 (0.10)
	β_{11}	0.03 (0.98)	0.49 (0.62)	−0.83 (0.41)
	β_{12}	−0.02 (0.98)	−0.35 (0.73)	1.63 (0.10)