A Score Test for Variance Components in a Semiparametric Mixed-Effects Model under Non-normality

Yan Sun and Jin-Ting Zhang

Abstract

In this paper, we propose a score test for variance components in a semiparametric mixed-effects model when the random-effects and measurement errors are not normally distributed. The asymptotic null distribution of the test statistic is shown to be a simple chi-squared distribution with the degrees of freedom being the number of linearly-independent variance components. The simulation results show that the proposed score test is robust against the non-normality of both the random-effects and the measurement errors and performs well in terms of both size and power. The score test is illustrated via an application to a real longitudinal data set collected in a clinical trial study.

Key words: Extended quasi-likelihood; Laplace approximation; local linear smoothing; score test; semiparametric mixed-effects model; variance components.

1 Introduction

Mixed-effects models provide an attractive tool to take the within-subject and between-subject variations of longitudinal data into account. Both parametric and nonparametric regression models have been extended by incorporating random-effects properly.
into the models (Wu and Zhang 2006). Although these parametric and nonparametric mixed-effects models have been enthusiastically accepted by both practitioners and researchers (Breslow and Calyton 1993), substantial theoretical and practical challenges remain. A natural question is whether or not the inclusion of random-effects and the accompanying, often-cumbersome mixed-effects modelling methodologies is necessary for a particular longitudinal data set. In this paper, we shall discuss this problem in the framework of a semiparametric mixed-effects model.

Suppose we have an experiment with \( m \) independent subjects with the \( i \)-th subject having \( n_i \) measurements over time. Let \( y_{ij} \) denote the responses for the \( i \)-th subject at design time points \( t_{ij} \). Consider the following semiparametric mixed-effects (SPME) model:

\[
y_{ij} = \eta(t_{ij}) + z_{ij}^T b_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m,
\]

where \( \eta(t) \) is the nonparametric fixed-effects function, modeling the population mean function of the longitudinal data; \( b_i \) and \( z_{ij} \) are the \( q \)-dimensional parametric random-effects and the associated random-effects covariates; and \( \epsilon_{ij} \) are the measurement errors. Throughout this paper, we assume, among others, that (1) the measurement errors \( \epsilon_{ij} \) are i.i.d. with \( \text{E}\epsilon_{11} = 0 \) and \( \text{Var}(\epsilon_{11}) = \sigma^2 \); and (2) the random-effects \( b_i \) are i.i.d with \( \text{E}b_1 = 0 \) and \( \text{Cov}(b_1) = D(\theta) \) where \( \theta \) is a \( p \)-dimensional vector of unknown variance components varying in a parameter space \( \Theta \) satisfying \( D(0) = 0 \). The magnitude of \( \theta \) can be used to measure the degree of overdispersion and correlation. Following Lin (1997), without loss of generality, we postulate that each component of \( D(\theta) \) is a linear function of \( \theta \). We further assume that

\[
\text{E}_\theta(\|b_1\|^r) = o(\|\theta\|^r), \quad \text{as} \quad \|\theta\| \to 0, \quad \text{for all} \quad r > 2.
\]

This moment condition is satisfied if the random-effects have an exponential-family distribution (McCullagh and Nelder 1989, P.350), or a mixture of exponential-family distributions (Johnson and Kotz 1970, P.88).

The SPME model (1.1) was first considered by Wang (1998) who estimated \( \eta(t) \) using a smoothing spline approach. A number of approaches for fitting the SPME model (1.1) and its various generalizations are given in Wu and Zhang (2006, Chap. 8). In all these approaches, a smoothing technique, e.g., smoothing spline (Wang 1998), is often needed to approximate \( \eta(t) \) with the approximation controlled by a smoothing parameter. For a grid of given values of the smoothing parameter, the SPME model (1.1) is often fitted via the following two steps: (1) given the variance
components, $\eta(t)$ and $b_i$ are computed; and (2) for the given estimates of $\eta(t)$ and $b_i$, the variance components are updated via some EM-algorithm. These two steps are repeated a number of times until convergence. The smoothing parameter is chosen according to some criterion (Wu and Zhang 2006, Chap. 8). The whole process is often-cumbersome and time-consuming mainly due to the inclusion of the parametric random-effects, especially when both random-effects and measurement errors are not normally distributed.

To check if the inclusion of the parametric random-effects in the SPME model (1.1) is necessary is equivalent to test the following hypothesis:

$$H_0 : \theta = 0 \text{ versus } H_1 : \theta \neq 0.$$  \hfill (1.3)

When $H_0$ is valid, the SPME model (1.1) reduces to its null model, i.e., the following population mean model (Wu and Zhang 2006):

$$y_{ij} = \eta(t_{ij}) + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m.$$  \hfill (1.4)

Compared with the full SPME model (1.1), the null model (1.4) is much simpler. It can be much more easily fitted by a number of techniques (Wu and Zhang 2006, Chap. 8).

To test $H_0$ (1.3), one should avoid fitting the complicate and time-consuming SPME model (1.1). For this end, a score test may be preferred. In many situations, the score test is an appealing competitor to both the likelihood ratio and the Wald-type tests because it only requires fitting the null model instead of the alternative model and estimating the nuisance parameters under the null model. In addition, all the three tests share the same local power (Cox and Hinkeley 1974, Chap.9). Many authors have studied the score tests in the framework of the parametric mixed-effects model; see, for example, Commenges and Jacqmin-Gadda (1997), Lin (1997), Hall and Praestgaard (2001), Verbeke and Molenberghs (2003), Zhu and Zhang (2006) among others. Up to our knowledge, less work has been done in the framework of semiparametric mixed-effects models. When the measurement errors $\epsilon_{ij}$ are normally distributed, the score test proposed by Zhu and Fung (2004) can be applied to test the null hypothesis in (1.3). However, simulation studies conducted in Section 3 show that Zhu and Fung’s (2004) testing procedure is not robust against non-normality of the measurement errors. Therefore, it is still worthwhile to propose and study a score test which is robust against non-normality of both the random-effects and the measurement errors.
The score test proposed in this paper can be simply constructed with the nuisance parameters estimated under the null model. In addition, we show that the test statistic asymptotically has a chi-squared distribution with \( p \) degrees of freedom where \( p \) is the number of linearly-independent variance components in \( \theta \). Therefore it is easy to conduct the proposed score test.

The rest of the paper is organized as follows. The main development of the score test and its asymptotic null distribution are given in Section 2. In Section 3, two simulation studies are conducted and they show that the proposed score test is robust against non-normality of the random-effects and measurement errors. In Section 4, we apply the proposed score test to a real longitudinal data set collected in a clinical trial study. Some concluding remarks are given in Section 5. The technical proof of a main result is given in the Appendix.

2 The Score Test

Under the SPME model (1.1), set \( y_i = (y_{i1}, \ldots, y_{in})^T \), \( \eta_i = (\eta(t_{i1}), \ldots, \eta(t_{in}))^T \), and \( Z_i = (z_{i1}, \ldots, z_{in})^T \). The extended quasi-likelihood (Nelder and Pregibon 1987) of the variance components \( \sigma^2 \) and \( \theta \) can be expressed as

\[
l(\sigma^2, \theta) = \prod_{i=1}^m E \left\{ \exp \left[ l_i(\sigma^2; \eta_i) \right] \right\},
\]

where given the random-effects \( b_i \), \( l_i(\sigma^2; \eta_i) \) is the \( i \)-th conditional extended log-quasi-likelihood given by

\[
l_i(\sigma^2; \eta_i) = -\frac{1}{2\sigma^2} (y_i - \eta_i - Z_ib_i)^T (y_i - \eta_i - Z_ib_i) - \frac{n}{2} \log \sigma^2.
\]

Following Solomon and Cox (1992), Brewslow and Lin (1995), by the moment assumption (1.2) imposed on the random-effects and by the Laplace expansion, the extended quasi-likelihood (2.1) can be approximated by

\[
l_a(\sigma^2, \theta) = \sum_{i=1}^m l_i(\sigma^2; 0) + \frac{1}{2} \text{tr} \left( \sum_{i=1}^m \left\{ \frac{\partial l_i(\sigma^2; \eta_i)}{\partial \eta_i} \frac{\partial l_i(\sigma^2; \eta_i)}{\partial b_i^T} + \frac{\partial^2 l_i(\sigma^2; \eta_i)}{\partial b_i \partial b_i^T} \right\} \Omega(\theta) \right)
\]

where \( \eta = (\eta_1^T, \ldots, \eta_m^T)^T \), \( y = (y_1^T, \ldots, y_m^T)^T \), and \( \Omega = \text{diag}(Z_1D(\theta)Z_1^T, \ldots, Z_mD(\theta)Z_m^T) \).

Note that in the above extended quasi-likelihood expressions, the nuisance parameter \( \eta \) and the response vector \( y \) are suppressed for simplicity of the presentation. When the measurement errors \( \epsilon_{ij} \) are normally distributed, the extended quasi-likelihood (2.1) reduces to the usual likelihood of \( \sigma^2 \) and \( \theta \). In this paper, however, this normality assumption is not imposed since it is not needed. In fact, the approximate extended
quasi-likelihood (2.2) is used only for deriving the score test statistic and the related estimators for the variance components and some nuisance parameters. It is not needed when we latter derive the asymptotic distribution of the proposed score test.

The score test can be constructed in four simple steps: (1) estimate the nuisance parameters under $H_0$; (2) calculate the score function under $H_0$ with the estimated nuisance parameters; (3) calculate the information matrix of $\theta$ under $H_0$ with the estimated nuisance parameters; and (4) form the score test using the score function and the information matrix.

First of all, we describe how to implement Step (1). Under $H_0$, the SPME model (1.1) reduces to the null model (1.4) where $\sigma^2$ is a 1-dimensional nuisance parameter while $\eta(t)$ is an infinite-dimensional nuisance parameter. We will also use the nuisance parameter $\tau = \text{Var}(\epsilon_{11}^2)$ in the development of the score test. We first construct the estimator of $\eta(t)$ under the null model (1.4) and then give the estimators of $\sigma^2$ and $\tau$.

Under the null model (1.4), various approaches for estimating $\eta(t)$ have been surveyed in Wu and Zhang (2006). In this paper, we shall adopt the well-known local linear method (Fan 1992) which has many good properties such as automatical boundary correction, design-adaptiveness among others.

The local linear method for estimating $\eta(t)$ can be briefly described as follows. Assume that for any fixed time point $t$, $\eta(t)$ has a second continuous derivative in a neighborhood of $t$. Then by Taylor’s expansion, $\eta(t_{ij})$ can be locally approximated by a linear function, i.e., $\eta(t_{ij}) \approx \eta(t) + \eta'(t)(t_{ij} - t) = x_{ij}^T \pi$ in the neighborhood of $t$, where $x_{ij} = (1, t_{ij} - t)^T$ and $\pi = (\pi_0, \pi_1)^T$ with $\pi_0 = \eta(t), \pi_1 = \eta'(t)$. The local linear estimator of $\eta(t)$ is then defined as $\hat{\eta}(t) = \hat{\pi}_0 = \mathbf{e}_{1,2}^T \hat{\pi}$, where $\mathbf{e}_{1,2} = (1, 0)^T$, and $\hat{\pi}$ is the minimizer of the weighted least squares criterion $\sum_{i=1}^m \sum_{j=1}^{n_i} [y_{ij} - x_{ij}^T \pi]^2 K_h(t_{ij} - t)$ where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ be the kernel function, usually a symmetric probability density function, and $h$ is the bandwidth, specifying the size of the local neighborhood of $t$ and controlling the smoothness of $\hat{\eta}(t)$.

By simple calculation, we have $\hat{\eta}(t) = \mathbf{e}_{1,2}^T (X^T W X)^{-1} X^T W Y$ where the weight matrix $W = \text{diag}(K_h(t_{11} - t), \ldots, K_h(t_{1n_1} - t), \ldots, K_h(t_{mn_1} - t), \ldots, K_h(t_{mn_m} - t))$ and $X = (X_1^T, \ldots, X_m^T)^T$, with $X_i = (x_{i1}, \ldots, x_{in_i})^T$. The bandwidth $h$ can be chosen by various methods, e.g., GCV; for more details, the readers are referred to Wu and Zhang (2006).

We are now ready to construct the estimators for $\sigma^2$ and $\tau$. At $\eta(t) = \hat{\eta}(t)$, the maximum extended quasi-likelihood estimator of $\sigma^2$ and the method of moment estimator
of \( \tau \) under \( H_0 \) are respectively

\[
\hat{\sigma}^2 = n^{-1}(y - \hat{\eta})^T(y - \hat{\eta}), \quad \hat{\tau} = n^{-1}\sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} - \hat{\eta}(t_{ij})\right]^4 - \hat{\sigma}^4, \tag{2.3}
\]

where \( \hat{\eta} = (\hat{\eta}_1^T, \ldots, \hat{\eta}_m^T)^T \) with \( \hat{\eta}_i = (\hat{\eta}(t_{i1}), \ldots, \hat{\eta}(t_{ini}))^T \) and \( n = \sum_{i=1}^m n_i \), the total number of measurements for the whole dataset.

In Step (2), the \( r \)-th entry of the score function \( u_\theta = (u_{\theta_1}, \ldots, u_{\theta_p})^T \) is easily calculated as

\[
u_{\theta_r} = \frac{\partial l_\eta(\sigma^2, \theta)}{\partial \theta_r} \bigg|_{\theta=0, \sigma^2=\hat{\sigma}^2, \eta=\hat{\eta}} = \frac{1}{2\hat{\sigma}^2}(y - \hat{\eta})^T \left\{ \hat{\Omega}_r - \frac{I_n}{n} \text{tr}(\hat{\Omega}_r) \right\} (y - \hat{\eta}),
\]

where \( I_n \) is the \( n \times n \) identity matrix, and \( \hat{\Omega}_r = \text{diag}(Z_1 \hat{D}_r Z_1^T, \ldots, Z_m \hat{D}_r Z_m^T) \) with \( \hat{D}_r = \frac{\partial \Omega_r}{\partial \theta_r} \bigg|_{\theta=0}, r = 1, 2, \ldots, p. \)

Throughout this paper, let \( \mathcal{D} = \{(t_{ij}, z_{ij}), j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, m\} \) denote the collection of the observed covariates and \( \text{diag}(A) \) denote the diagonal matrix formed by the diagonal entries of \( A \). In Step (3), the required information matrix of \( \theta \) is given by \( V = V_{\theta\theta} - V_{\theta\sigma} V_{\sigma^2 \sigma^2}^{-1} V_{\sigma \theta} \) where, by applying Lemma 1 in the Appendix, we have

\[
V_{\sigma^2 \sigma^2} = E \left( \frac{\partial^2 l_\eta(\sigma^2, \theta)}{\partial \sigma^2 \partial \sigma^2} \bigg|_\mathcal{D} \right) \bigg|_{\theta=0, \sigma^2=\hat{\sigma}^2, \tau=\hat{\tau}} = \frac{n \hat{\tau}}{4\hat{\sigma}^4},
\]

\[
V_{\sigma^2 \theta_r} = E \left( \frac{\partial^2 l_\eta(\sigma^2, \theta)}{\partial \sigma^2 \partial \theta_r} \bigg|_\mathcal{D} \right) \bigg|_{\theta=0, \sigma^2=\hat{\sigma}^2, \tau=\hat{\tau}} = \frac{\hat{\tau}}{4\hat{\sigma}^8} \text{tr}(\hat{\Omega}_r),
\]

\[
V_{\theta_r \theta_s} = E \left( \frac{\partial^2 l_\eta(\sigma^2, \theta)}{\partial \theta_r \partial \theta_s} \bigg|_\mathcal{D} \right) \bigg|_{\theta=0, \sigma^2=\hat{\sigma}^2, \tau=\hat{\tau}} = \frac{1}{2\hat{\sigma}^4} \left\{ \text{tr}(\hat{\Omega}_r \hat{\Omega}_s) - \text{tr}(\hat{\Delta}_r \hat{\Delta}_s) \right\} + \frac{\hat{\tau}}{4\hat{\sigma}^8} \text{tr}(\hat{\Delta}_r \hat{\Delta}_s),
\]

where \( \hat{\Delta}_r = \text{diag}(\hat{\Omega}_r) \). It follows that

\[
V = (v_{rs})_{1 \leq r, s \leq p}, \quad \text{with} \quad v_{rs} = \frac{1}{2\hat{\sigma}^4} v_{rs}^{(1)} + \frac{\hat{\tau}}{4\hat{\sigma}^8} v_{rs}^{(2)}, \tag{2.4}
\]

where \( v_{rs}^{(1)} = \text{tr}(\hat{\Omega}_r \hat{\Omega}_s) - \text{tr}(\hat{\Delta}_r \hat{\Delta}_s) \), and \( v_{rs}^{(2)} = \text{tr}(\hat{\Delta}_r \hat{\Delta}_s) - \frac{1}{n} \text{tr}(\hat{\Omega}_r) \text{tr}(\hat{\Omega}_s) \).

Finally in Step (4), the score test statistic is constructed as

\[
T = u_\theta^T V^{-1} u_\theta. \tag{2.5}
\]

Note that when \( \epsilon_{ij} \) are normally distributed, we have \( \tau = 2\sigma^4 \). In this case, the above score test statistic \( T \) can be simplified as

\[
\tilde{T} = u_\theta^T \tilde{V}^{-1} u_\theta, \tag{2.6}
\]
where $\tilde{V}$ is the associated simplified information matrix of $\theta$ with $\tilde{V} = (\tilde{v}_{rs})_{1 \leq r,s \leq p}$, $\tilde{v}_{rs} = \frac{1}{2\pi} \left\{ \text{tr}(\Omega_r\Omega_s) - \frac{1}{n} \text{tr}(\tilde{\Omega}_r)\text{tr}(\tilde{\Omega}_s) \right\}$. The simplified score test statistic (2.6) is similar to the one proposed by Zhu and Fung (2004) for a semiparametric mixed-effects model with the normality assumption imposed for the measurement errors.

The asymptotic distribution of the score test statistic (2.5) is very simple and is given in Theorem 1 below.

**Theorem 1** Under $H_0$ and the conditions in the Appendix, as $n \to \infty$, the score test statistic $T$ asymptotically follows a chi-squared distribution with $p$ degrees of freedom.

### 3 Simulation Studies

In this section, we shall present two simulation studies. We aim to examine the performance of the proposed score test $T$ defined in (2.5) via comparing it against the test $\tilde{T}$ defined in (2.6) which is constructed based on the assumption that the measurement errors $\epsilon_{ij}$ are normally distributed.

We generated the data from the following SPME model:

$$y_{ij} = 1 + 2 \cos(2\pi t_{ij}) + z_{ij}^T b_i + \epsilon_{ij}, \quad j = 1, \cdots, n_i; \quad i = 1, \cdots, m,$$

which is a special case of the SPME model (1.1) with $\eta(t) = 1 + 2 \cos(2\pi t)$. The design time points are first scheduled as $t_j = j/(K + 1), j = 1, \cdots, K$ with $K = 10$ here. To obtain an imbalance design, we randomly remove some design time points for a subject at a rate 10%, so that there are about 9 measurements per subject and 9$m$ measurements for all the subjects. The resulting design time points are denoted as $t_{ij}, j = 1, 2, \cdots, n_i; i = 1, 2, \cdots, m$ as in (3.1).

Simulation 1 aims to study the case when the random-effects are univariate and the random-effects covariates are time-independent. In this case, we write $b_i$ and $z_{ij}$ as $b_i$ and $z_i$ respectively. We generate $z_i$ from the standard uniform distribution and the random-effects $b_i$ ($i = 1, \cdots, m$) from the following normal mixture:

$$F = 0.25N(-0.75\gamma, \nu^2) + 0.75N(0.25\gamma, \nu^2),$$

which has mean 0 and variance $\theta = \frac{3}{16}\gamma^2 + \nu^2$. The tuning parameters $\gamma$ and $\nu$ can be flexibly specified so that various cases of $F$ can be considered. In Simulation 1, the following three cases of $F$ are considered:

$$F = \frac{3}{16}$$
Case 1: $\gamma = \nu = 0$ so that $\theta = 0$, specifying the null model.

Case 2: $\gamma = 0, \nu^2 = 2/5$ so that $\theta = 2/5$, specifying an alternative model with normal random-effects $b_i$.

Case 3: $\gamma = 1/2, \nu^2 = 9/64$ so that $\theta = 3/16$, specifying an alternative model with non-normal random-effects $b_i$.

These three cases of $F$ allow us to assess the empirical sizes and powers of $T$ and $\tilde{T}$ and compare their performance under normality and non-normality assumptions. To assess the effect of the number of subjects and the effect of the measurement error structure, we consider two choices of the number of subjects: $m = 50$ and $m = 100$, and two structures of the measurement errors: $N(0, 1)$ and $\frac{1}{\sqrt{3}}t_3$ (In both cases, the generated measurement errors have variance $\sigma^2 = 1$). For each choice of $m$, $F$, and the measurement error structures, 1000 replications were conducted. In each replication, the null model (1.4) is fitted using the local linear method described in Section 2 in which the well-known Epanechnikov kernel $K(t) = \frac{3}{4}(1 - t^2)_+$ is used with the bandwidth $h$ chosen by GCV. The score test statistics $T$ and $\tilde{T}$ are then computed using the method described in Section 2. The null hypothesis is rejected if the computed test statistic is larger than the critical value of the $\chi^2_p$-distribution with $p = 1$ here at the nominal significance level $\alpha = 5\%$. The empirical powers of $T$ and $\tilde{T}$ are defined as the proportions of the rejections in 1000 replications.

Table 1 presents the results of Simulation 1. As expected, when the measurement errors are normally distributed, $T$ and $\tilde{T}$ have the same empirical powers and sizes. However, when the measurement errors are $\frac{1}{\sqrt{3}}t_3$ distributed, the empirical sizes (Type-I errors) of $\tilde{T}$ are inflated, and much larger than those of $T$ which are not inflated. This explains why the empirical powers of $\tilde{T}$ are generally larger than those of $T$, possibly resulting in misleading results due to misspecifying the distribution of the measurement errors. In this sense, the proposed score test $T$ outperforms the existing score test $\tilde{T}$ proposed by Zhu and Fung (2004) constructed based on the normality assumption of the measurement errors. This situation becomes more serious when the random-effects are multivariate and the random-effects covariates are time-dependent as indicated by the results of Simulation 2 presented below.

Simulation 2 aims to study the case when the random-effects are bivariate and the random-effects covariates are time-dependent. For simplicity, we specify the time-dependent covariates as $z_{ij} = [1, t_{ij}]^T$. The random-effects $b_i$ ($i = 1, \cdots, m$) are generated
Table 1: *Empirical powers of $T$ and $\hat{T}$ in 1000 replications at $\alpha = 5\%$.\(^\text{a}\)*

(Simulation 1)

<table>
<thead>
<tr>
<th>m random-effects measurement errors</th>
<th>$T$</th>
<th>$\hat{T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 Null (Case 1) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.042</td>
<td>0.042</td>
</tr>
<tr>
<td>Normal (Case 2) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.874</td>
<td>0.874</td>
</tr>
<tr>
<td>Non-normal (Case 3) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.754</td>
<td>0.754</td>
</tr>
<tr>
<td>100 Null (Case 1) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>Normal (Case 2) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>Non-normal (Case 3) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.947</td>
<td>0.947</td>
</tr>
</tbody>
</table>

Table 2: *Empirical powers of $T$ and $\hat{T}$ in 1000 replications at $\alpha = 5\%$.\(^\text{a}\)*

(Simulation 2)

<table>
<thead>
<tr>
<th>m random-effects measurement errors</th>
<th>$T$</th>
<th>$\hat{T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 Null (Case 1) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.046</td>
<td>0.046</td>
</tr>
<tr>
<td>Normal (Case 2) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.938</td>
<td>0.938</td>
</tr>
<tr>
<td>Non-normal (Case 3) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.706</td>
<td>0.706</td>
</tr>
<tr>
<td>100 Null (Case 1) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.043</td>
<td>0.043</td>
</tr>
<tr>
<td>Normal (Case 2) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>Non-normal (Case 3) $N(0, 1)$ $\sqrt{3}t_3$</td>
<td>0.928</td>
<td>0.928</td>
</tr>
</tbody>
</table>
from the following two-dimensional normal mixture:

\[
F = 0.25N_2(-0.75\gamma, \nu^2\Lambda) + 0.75N_2(0.25\gamma, \nu^2\Lambda),
\]

which has mean 0 and covariance matrix

\[
D(\theta) = \begin{pmatrix}
\theta_1 & \theta_2 \\
\theta_2 & \theta_3
\end{pmatrix} = \frac{3}{16} \gamma \gamma^T + \nu^2\Lambda.
\]

Set \( \Lambda = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 0.5 \end{pmatrix} \). As in Simulation 1, via specifying \( \gamma \) and \( \nu \), we specify the following three cases of \( F \) for study:

Case 1: \( \gamma = 0, \nu = 0 \) so that \( \theta = (\theta_1, \theta_2, \theta_3)^T = 0 \), specifying the null model.

Case 2: \( \gamma = 0, \nu^2 = 0.1 \) so that \( \theta = (0.1, 0.02, 0.05)^T \), specifying an alternative model with normal random-effects \( b_i \).

Case 3: \( \gamma = (0.4, -0.2)^T, \nu^2 = 0.05 \) so that \( \theta = (0.08, -0.005, 0.0325)^T \), specifying an alternative model with non-normal random-effects \( b_i \).

Other tuning parameters are the same as those in Simulation 1. Since the dimension of \( \theta \) is now \( p = 3 \), under \( H_0 \), the score tests \( T \) and \( \tilde{T} \) now asymptotically follow the \( \chi^2_3 \)-distribution. Therefore, the null hypothesis is now rejected if the computed test statistic is larger than the critical value of the \( \chi^2_3 \)-distribution at \( \alpha = 5\% \).

Table 2 presents the results of Simulation 2, which are similar to those of Simulation 1 in Table 1 except the empirical sizes of \( \tilde{T} \) in Simulation 2 are much more inflated from \( \alpha = 5\% \) than in Simulation 1. Thus, more serious misleading results (much larger Type-I errors) from using \( \tilde{T} \) in Simulation 2 than in Simulation 1 may be yielded when the measurement errors are not normally distributed. Therefore, from these two simulation studies, we shall recommend to use \( T \) instead of \( \tilde{T} \) in practice unless we have strong evidence showing that the measurement errors are indeed normally distributed.

### 4 An Application

We now apply the proposed score test \( T \) to an AIDS clinical study conducted by the AIDS Clinical Trials Group (ACTG). The study enrolled 517 HIV-1 infected patients in three antiviral treatments. The data considered here just consist of one of the treatment arms in which 166 patients were treated with a highly active antiretroviral therapy
(HAART) for 120 weeks during which CD4 cell counts were monitored at weeks 0, 4, 8, and every 8 weeks thereafter. However, each individual patient might not exactly follow the designed schedule, and missing clinical visits for CD4 cell measurements frequently occurred which makes the resulting longitudinal data set unbalanced. The number of CD4 cell count measurements per patient varies from 1 to 18.

The longitudinal data set was originally analyzed by Park and Wu (2006) using an nonparametric mixed-effects model. We are interest if the CD4 cell counts are different among the patients. We handled this problem via considering the following SPME model:

\[ y_{ij} = \eta(\text{week}_{ij}) + z_{ij}^T b_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, 2, \ldots, 166, \]  

(4.1)

where \( z_{ij} = (1, \text{week}_{ij}, \text{week}^2_{ij})^T \), \( b_i = (b_{i1}, b_{i2}, b_{i3})^T \) with mean vector \( 0 \) and covariance matrix

\[
D(\theta) = \begin{pmatrix}
\theta_1 & \theta_2 & \theta_3 \\
\theta_2 & \theta_4 & \theta_5 \\
\theta_3 & \theta_5 & \theta_6
\end{pmatrix},
\]

and considering the null hypothesis \( H_0 : \theta_1 = \theta_2 = \cdots = \theta_6. \) Since there is no information about the normality of the random-effects \( b_i \) and the measurement errors \( \epsilon_{ij} \), we applied the proposed score test \( T \) to the above problem. The computed test statistic \( T = 9445.3 \) and the associated P-value is 0, suggesting a strong rejection of the null hypothesis. That is, the inclusion of the random-effects in the SPME model (4.1) is strongly supported.

5 Concluding Remarks

In this paper, we propose and study a score test for variance components testing in the framework of the SPME model (1.1). The proposed score test can be easily constructed with the score function, the information matrix and the nuisance parameters computed under the null hypothesis. It has simple asymptotic null distribution and hence can be conducted easily without assuming normality of the random-effects and measurement errors. Although it may be tedious, the proposed score test can be extended to the framework of other semiparametric or time-varying coefficients mixed-effects models (Wu and Zhang 2006). When \( H_0 : \theta = 0 \) is rejected, it is of interest to test whether some of the variance components in \( \theta \) are zero. The studies in these directions are warranted.
Appendix: Proof of Theorem 1

In this appendix, we shall outline the proof of Theorem 1. First of all, we present the following lemma which was used to calculate the score function in Section 2 and will be used in the proof of Theorem 1. Its proof is straightforward and is hence omitted.

**Lemma 1** Suppose \( \epsilon_1, \cdots, \epsilon_M \) are i.i.d. with \( E(\epsilon_1) = 0 \) and \( \text{Var}(\epsilon_1) = \sigma^2 \). Then for any two constant and symmetric matrices \( A : M \times M \) and \( B : M \times M \), we have

\[
E\left\{ \epsilon^T A \epsilon \epsilon^T B \epsilon \right\} = 2\sigma^4 \left[ \text{tr}(AB) - \text{tr} \left\{ \text{diag}(A) \text{diag}(B) \right\} \right] + \tau \text{tr}(\text{diag}(A) \text{diag}(B)) + \sigma^4 \text{tr}(A) \text{tr}(B),
\]

where \( \epsilon = (\epsilon_1, \cdots, \epsilon_M)^T \) and \( \tau = \text{Var}(\epsilon_1^2) \).

We now list some notations and the required conditions for Theorem 1. Let \( \alpha = (\alpha_1, \cdots, \alpha_p)^T \) be any given constant vector. Set

\[
R_n = \sum_{r=1}^p \alpha_r \left\{ \Omega_r - \frac{\text{tr}(\Omega_r)}{n} I_n \right\}, \quad \epsilon = (\epsilon_1^T, \cdots, \epsilon_m^T)^T, \quad \epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{im})^T.
\]

Further, write \( V^{(1)} = (v^{(1)}_{rs})_{1\leq r,s \leq p}, V^{(2)} = (v^{(2)}_{rs})_{1\leq r,s \leq p} \), where \( v^{(1)}_{rs} \) and \( v^{(2)}_{rs} \) are as defined in Section 2. Finally define the norms of a vector \( a \) and a matrix \( A \) as \( \|a\| = (a^T a)^{1/2} \) and \( \|A\| = \left\{ \text{tr}(AA^T) \right\}^{1/2} \) respectively. The following regularity conditions are imposed for Theorem 1:

1. The number of measurements per subject is bounded, i.e., \( n_i < C, i = 1, \cdots, m \) for some \( 0 < C < \infty \). Again \( n = \sum_{i=1}^m n_i \) denotes the total number of measurements for the whole dataset.

2. The measurement errors \( \epsilon_{ij} \) are i.i.d. with \( E\epsilon_{11}^4 < \infty \); the random-effects covariates satisfy \( \max_{1 \leq i \leq m; 1 \leq j \leq n_i} E\|z_{ij}\|^4 < \infty \).

3. The largest eigenvalues of matrices \( D_r, r = 1, \cdots, p \) are bounded, and for some \( \delta > 0 \), we have \( \max_{1 \leq i \leq m, 1 \leq j \leq n_i} E(\|z_{ij}\|\|z_{il}\|)^{2+\delta} < \infty \).

4. There exist nonnegative definite matrices \( V^{(1)}_0 \) and \( V^{(2)}_0 \) such that \( n^{-1}V^{(1)} \xrightarrow{P} V^{(1)}_0, n^{-1}V^{(2)} \xrightarrow{P} V^{(2)}_0 \), and \( V_0 = \frac{1}{2\pi^2} V^{(1)}_0 + \frac{\tau}{4\pi^2} V^{(2)}_0 \) is positive definite.

5. The marginal density \( f(t) \) has a compact support \( \mathcal{T} \), and is Lipschitz continuous on \( \mathcal{T} \). In addition, \( f(t) \neq 0, t \in \mathcal{T} \).

6. The second derivative function \( \eta''(t) \) is bounded on \( \mathcal{T} \) and \( \eta''(t) \neq 0, t \in \mathcal{T} \).
(7) The kernel function $K(u)$ is a symmetric probability density function having a compact support, e.g., $[-1, 1]$.

(8) The bandwidth $h$ satisfies $h \to 0$, $nh^2 \to \infty$, and $nh^8 \to 0$.

Condition (1) is satisfied by almost all longitudinal data; otherwise, the associated data are often referred to as functional data. Conditions (3) and (4) are needed for applying the Lindeberg-Feller central limit theorem when the random-effects and measurement errors are non-normal. We assume them for easy presentation although it is difficult to check in practice. From the proof of Theorem 1, we can see that $V^{(1)}$ and $V^{(2)}$ must be nonnegative matrices since $\text{Var}(J_n|D) = 2\sigma^4 \alpha^T V^{(1)} \alpha$ and $\text{Var}(J_n|D) = \tau \alpha^T V^{(2)} \alpha$ must be nonnegative for all $\alpha$. Therefore, Condition (4) is easily satisfied. Conditions (5)-(8) are regularity conditions for local linear smoothing for the null model (1.4). Under the null model, we can show that the optimal bandwidth for $\hat{\eta}(t)$ is $h = O(n^{-\frac{1}{5}})$ which satisfies Condition (8).

**Proof of Theorem 1** Using Lemma 1 in Fan and Zhang (1999) and the same arguments as those establishing Theorem 1 of their paper, we can show that under $H_0$ and for any given point $t \in T$, the asymptotic conditional bias and variance of $\hat{\eta}(t)$ are

$$\text{Bias} \{ \hat{\eta}(t)|D \} = O_P(h^2), \quad \text{Var} \{ \hat{\eta}(t)|D \} = O_P\{ (nh)^{-1} \},$$

uniformly in $t$.

By some standard arguments and by the definitions of $\hat{\sigma}$ and $\hat{\tau}$ in (2.3), it is straightforward to show that under $H_0$ both $\hat{\sigma}^2$ and $\hat{\tau}$ are consistent for $\sigma^2$ and $\tau$ respectively. Therefore, by Condition (4), we have $n^{-1} V \frac{L}{n} \frac{1}{2\hat{\sigma}^2} V^{(1)}_0 + \frac{\tau}{2\hat{\tau}^2} V^{(2)}_0 = V_0$. Theorem 1 will follow if we can show that under $H_0$, for any given $\alpha = (\alpha_1, \ldots, \alpha_p)^T$, we have

$$n^{-1/2} \alpha^T u_\theta \xrightarrow{L} N(0, \alpha^T V_0 \alpha).$$

(A.2)

For this end, we write

$$\alpha^T u_\theta = \frac{1}{2\hat{\sigma}^2} \left\{ e^T (R_n - Q_n) e + e^T Q_n e + E[(\hat{\eta} - \eta)^T|D] R_n E[(\hat{\eta} - \eta)|D] + [\hat{\eta} - E(\hat{\eta}|D)]^T R_n [\hat{\eta} - E(\hat{\eta}|D)] - 2e^T R_n E[(\hat{\eta} - \eta)|D] \right.$$

$$- 2e^T R_n [\hat{\eta} - E(\hat{\eta}|D)] + 2[\hat{\eta} - E(\hat{\eta}|D)]^T R_n E[(\hat{\eta} - \eta)|D] \left. \right\}$$

$$\equiv \frac{1}{2\hat{\sigma}^2} \left\{ J_{n1} + J_{n2} + J_{n3} + J_{n4} - 2J_{n5} - 2J_{n6} + 2J_{n7} \right\},$$

where $Q_n = \text{diag}(R_n)$, a diagonal matrix having the same diagonal entries as $R_n$. 13
By straightforward calculation, we have

$$J_{n1} = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n} \sum_{l=1,l \neq j}^{n} z_{ij}^T \left( \sum_{r=1}^{p} \alpha_r \bar{D}_r \right) \bar{z}_{il} \epsilon_{il} \right\}.$$  

In addition, \(E(J_{n1} | D) = \sigma^2 \text{tr}(R_n - Q_n) = 0\). Applying Lemma 1, we have

$$\text{Var}(J_{n1} | D) = E(J_{n1} | D)^2 = 2\sigma^4 \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ z_{ij}^T \left( \sum_{r=1}^{p} \alpha_r \bar{D}_r \right) \bar{z}_{il} \right]^2 - \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ z_{ij}^T \left( \sum_{r=1}^{p} \alpha_r \bar{D}_r \right) \bar{z}_{ij} \right]^2 \right\} = 2\sigma^4 \alpha^T V^{(2)} \alpha.$$  

By Condition (4), as \(n \to \infty\), we have \(E\left\{ n^{-1/2} J_{n1} \right\} = 0\) and \(\text{Var}\left\{ n^{-1/2} J_{n1} \right\} \to 2\sigma^4 \alpha^T V^{(1)}_0 \alpha\). Thus, under Conditions (1)-(4) and by the Lindeberg-Feller central limit theorem, as \(n \to \infty\), we have that \(n^{-\frac{1}{2}} J_{n1} \xrightarrow{L} N\left(0, 2\sigma^4 \alpha^T V^{(1)}_0 \alpha\right)\).

Since \(Q_n\) is a diagonal matrix and \(\text{tr}(Q_n) = 0\), we have \(E(J_{n2} | D) = \sigma^2 \text{tr}(Q_n) = 0\). Applying Lemma 1, we have

$$\text{Var}(J_{n2} | D) = E(J_{n2}^2 | D) = \tau \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ z_{ij}^T \left( \sum_{r=1}^{p} \alpha_r \bar{D}_r \right) \bar{z}_{ij} \right]^2 - \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ z_{ij}^T \left( \sum_{r=1}^{p} \alpha_r \bar{D}_r \right) \bar{z}_{ij} \right]^2 \right\} = \tau \alpha^T V^{(2)} \alpha,$$

where \(\tau = \text{Var}(\epsilon_{11}^2)\) as defined before. By Condition (4), we have \(\text{Var}\left\{ n^{-1/2} J_{n2} \right\} \to \tau \alpha^T V^{(2)}_0 \alpha\), as \(n \to \infty\). Since \(Q_n\) is a diagonal matrix, \(J_{n2}\) is a sum of independent variables. By Conditions (1)-(4), and by the Lindeberg-Feller central limit theorem, as \(n \to \infty\), we have \(n^{-\frac{1}{2}} J_{n2} \xrightarrow{L} N\left(0, \tau \alpha^T V^{(2)}_0 \alpha\right)\). Since \(\text{Cov}(J_{n1}, J_{n2}) = E[E(J_{n1}J_{n2} | D)] = 2\sigma^4 E[\text{tr}\left( (R_n - Q_n)Q_n \right)] = 0\), \(J_{n1}\) and \(J_{n2}\) are uncorrelated. Therefore, as \(n \to \infty\), we have

$$n^{-\frac{1}{2}} \{ J_{n1} + J_{n2} \} \xrightarrow{L} N\left(0, \alpha^T [2\sigma^4 V^{(1)}_0 + \tau V^{(2)}_0] \alpha\right).$$

Then we will have

$$n^{-\frac{1}{2}} \alpha^T u_0 \xrightarrow{L} N\left(0, \frac{1}{4\sigma^8} \alpha^T \left\{ 2\sigma^4 V^{(1)}_0 + \tau V^{(2)}_0 \right\} \alpha\right),$$

if we can show that

$$n^{-1/2}(J_{n3} + J_{n4} - 2J_{n5} - 2J_{n6} + 2J_{n7}) = o_P(1). \quad (A.3)$$

14
The expression (A.3) then follows by noticing that \( \frac{1}{2\sigma^2} \left\{ 2\sigma^4 V_0^{(1)} + \tau V_0^{(2)} \right\} = \frac{1}{2\sigma^2} V_0^{(1)} + \frac{\tau}{2\sigma^2} V_0^{(2)} = V_0 \).

We now briefly outline the proof of (A.4). First of all, by the law of large numbers and the conditional bias \( \text{Bias}\{\hat{\eta}(t)\} \) given in (A.1), we have \( n^{-\frac{1}{2}} J_{n3} = O_P \left\{ (nh^8)^{1/2} \right\} \).

Second, we can write

\[
J_{n4} = \sum_{i=1}^{m} [\hat{\eta}_i - E(\hat{\eta}_i|D)]^T \left[ \sum_{r=1}^{p} \alpha_r (Z_i, \bar{D}_i, \bar{Z}_i^T - n^{-1} I_{n_i} \sum_{k=1}^{m} \sum_{v=1}^{n_k} \bar{z}_{kv}^T \bar{D}_r, \bar{z}_{kv}) \right] [\hat{\eta}_i - E(\hat{\eta}_i|D)].
\]

Let \( \gamma_{ijl} \) be the \( (j,l) \)th component of matrix \( \sum_{r=1}^{p} \alpha_r (Z_i, \bar{D}_i, \bar{Z}_i^T - n^{-1} I_{n_i} \sum_{k=1}^{m} \sum_{v=1}^{n_k} \bar{z}_{kv}^T \bar{D}_r, \bar{z}_{kv}) \).

By the inequality \( |a^T Aa| \leq \|A\| \cdot \|a\|^2 \) and (A.1), we have

\[
E\{||J_{n4}|D\} \leq C(nh)^{-\frac{1}{2}} \sum_{i=1}^{m} \left( \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \gamma_{ijl}^2 \right)^{\frac{1}{2}} = O_P(h^{-\frac{1}{2}}),
\]

where \( C \) is some positive constant which can take different values at each appearance. Therefore, by \( |J_{n4}| \leq |J_{n4} - EJ_{n4}| + |EJ_{n4}| \) and Markov inequality, we have \( n^{-\frac{1}{2}} J_{n4} = O_P \left\{ (nh^2)^{-\frac{1}{2}} \right\} \).

Using (A.1) again, we have

\[
EJ_{n5}^2 = E \left[ E \left\{ J_{n5}^2|D \right\} \right] \leq C h^4 E \left[ (e^T R_n 1_n)^2 \right] \leq C \sigma^2 h^4 \text{tr}(R_n^2),
\]

where \( 1_n \) is an \( n \)-dimensional vector with all its elements being 1. By the Markov inequality, we have \( n^{-\frac{1}{2}} J_{n5} = O_P(h^2) \).

By straightforward calculation, we have

\[
(\hat{\eta}(t_{ij}) - E[\hat{\eta}(t_{ij})|D]) \epsilon_{il} = \frac{1}{nf(t_{ij})} \sum_{k=1}^{m} \sum_{v=1}^{n_k} K_h(t_{kv} - t_{ij}) \epsilon_{k\nu} \epsilon_{il} (1 + o_P(1)).
\]

It follows that

\[
J_{n6} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \gamma_{ijl} (\hat{\eta}(t_{ij}) - E[\hat{\eta}(t_{ij})|D]) \epsilon_{il} = \left\{ \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{\gamma_{ijl}}{f(t_{ij})} K_h(t_{il} - t_{ij}) \epsilon_{il}^2 + \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{v=1}^{n_k} \frac{\gamma_{ijl}}{f(t_{ij})} K_h(t_{iv} - t_{ij}) \epsilon_{iv} \epsilon_{il} \right. \]

\[
+ \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k \neq i}^{m} \sum_{l=1}^{n_i} \sum_{v=1}^{n_k} \frac{\gamma_{ijl}}{f(t_{ij})} K_h(t_{kv} - t_{ij}) \epsilon_{kv} \epsilon_{il} \left\} (1 + o_P(1)) \equiv \left\{ J_{n6,1} + J_{n6,2} + J_{n6,3} \right\} (1 + o_P(1)).
\]

By the boundness of the kernel function and the independence of measurement errors, we have

\[
\text{Var}\{J_{n6,1}|D\} \leq C \frac{1}{n^2 \sigma^2} \text{tr}(R_n^2) = O_P \left\{ (nh^2)^{-1} \right\},
\]

\[15\]
and

$$|E_{J_{n6,1}}| \leq C\sigma^2(nh)^{-1}\sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} E|\gamma_{i,j,l}| = O(h^{-1}).$$

Then by $$|J_{n6,1}| \leq |J_{n6,1} - EJ_{n6,1}| + |EJ_{n6,1}|$$ and the Markov inequality, it can be shown that $$n^{-\frac{1}{2}}J_{n6,1} = O_P((nh^2)^{-\frac{1}{2}})$$. By straightforward calculation and the boundedness of the kernel function, we have $$E(J_{n6,2}) = 0$$, $$\text{Var}(J_{n6,2}) = O(n^{-1}h^{-2})$$ and $$E(J_{n6,3}) = 0$$, $$\text{Var}(J_{n6,3}) = O(h^{-1})$$. Hence $$n^{-\frac{1}{2}}J_{n6,2} = O_P((nh)^{-1})$$ and $$n^{-\frac{1}{2}}J_{n6,3} = O_P((nh^2)^{-\frac{1}{2}})$$. Finally, by (A.1), we have

$$n^{-1}E\left\{J_{n7}^2|D\right\} \leq C\frac{h^4}{n}\text{tr}(R_n^2)\sum_{k=1}^{m} \sum_{l=1}^{n_i} E\left\{[\hat{\eta}(t_{kl}) - E(\hat{\eta}(t_{kl})|D)]^2|D\right\}(1 + o_P(1)) = O_P(h^3).$$

Therefore, we have $$n^{-\frac{1}{2}}J_{n7} = O_P(h^{\frac{1}{2}})$$. By Condition (8) and all the above, the expression (A.4) follows immediately. The proof of Theorem 1 is completed.

References


