Statistics and Its Interface Volume 4 (2011) 65–72

A score test for variance components in a semiparametric mixed-effects model under non-normality

Yan Sun* and Jin-Ting Zhang†‡

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 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.

AMS 2000 subject classifications: Primary 62G10, 62G20; secondary 62G08.
Keywords and phrases: Extended quasi-likelihood, Laplace approximation, Local linear smoothing, Score test, Semiparametric mixed-effects model, Variance components.

1. INTRODUCTION

Longitudinal data frequently arise in many scientific areas, including biology, psychology, sociology and economics among others; examples may be found in Wu and Zhang (2006) and references therein. The measurements of longitudinal data were observed repeatedly over a time period on a number of subjects. The within-subject measurements are usually correlated with each other but the measurements from different subjects are usually assumed to be uncorrelated. Modeling of such longitudinal data has played an important role in scientific investigations.

Mixed-effects models provide an attractive tool for taking 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 Clayton 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:

$$y_{ij} = \eta(t_{ij}) + \mathbf{z}_{ij}^{\mathrm{T}}\mathbf{b}_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i;\ i = 1, \ldots, m, \tag{1.1}$$

where $\eta(t)$ is the nonparametric fixed-effects function, modeling the population mean function of the longitudinal data; $\mathbf{b}_i$ and $\mathbf{z}_{ij}$ are the q-dimensional parametric random-effects and the associated random-effects covariates; and ϵ_{ij} are the measurement errors. Throughout this paper, we assume, among others, that (1) the measurement errors ϵ_{ij} are i.i.d. with $\mathrm{E}\epsilon_{11} = 0$ and $\mathrm{Var}(\epsilon_{11}) = \sigma^2$; and (2) the random-effects $\mathbf{b}_i$ are i.i.d with $\mathrm{E}\mathbf{b}_1 = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{b}_1) = \mathbf{D}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a p-dimensional vector of unknown variance components varying in a parameter space Θ satisfying $\mathbf{D}(\mathbf{0}) = \mathbf{0}$. The magnitude of $\boldsymbol{\theta}$ can be used to measure the degree of overdispersion and correlation. Following Lin (1997), we postulate that each component of $\mathbf{D}(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$. Note that for imposing this assumption, we lose no generality since in most practical situations, except the symmetricity, we know nothing about $\mathbf{D}(\cdot)$ and we usually form $\boldsymbol{\theta}$ using the components of $\mathbf{D}(\cdot)$ on and above the main diagonal so that this assumption is automatically satisfied. We further assume that

$$E_\theta(\|\mathbf{b}_1\|^r) = o(\|\boldsymbol{\theta}\|^r), \text{ as } \|\boldsymbol{\theta}\| \to 0, \text{for all } r > 2. \tag{1.2}$$

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

*Sun's research was supported by the National Science Foundation of China (grant 10801093) and Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (3rd phase).
†Zhang's research was supported by the National University of Singapore Academic Research Grant R-155-000-085-112.
‡Corresponding author.

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 $\mathbf{b}_i$ are computed; and (2) for the given estimates of $\eta(t)$ and $\mathbf{b}_i$, 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 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:

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

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):

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

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 complicated 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, Lin (1997), Hall and Praestgaard (2001), Verbeke and Molenberghs (2003), Zhu and Zhang (2006) among others. According to our knowledge, less work has been done in the framework of semiparametric mixed-effects models. When the measurement errors ϵ_{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, in this paper, we propose and study a score test for the variance components in a semiparametric mixed-effects model which is robust against non-normality of the random-effects and the measurement errors.

The proposed score test admits several advantages. First of all, it is valid under a broad class of distributions for the random-effects and the measurement errors. This is because except for the first two moments, no particular distribution assumptions are made for the random-effects and the measurement errors. This is really in the spirit of generalized estimating equations which rely only on specification of the first two moments. The basic idea of using the score test under "working likelihood" or "working scores" has been used by other authors in different contexts; see, for example, Wang and He (2007) who developed a robust rank score test for linear quantile models with random-effects to detect differentially expressed genes in GeneChip studies. Secondly, the proposed score test can be simply constructed with the nuisance parameters estimated under the simple null model. This allows fast computation, especially when the sample size is large. Thirdly, the test statistic asymptotically has a simple chi-squared distribution with p degrees of freedom where p is the number of linearly-independent variance components in $\boldsymbol{\theta}$. Therefore it is easy to conduct the proposed score test using the usual χ^2-table. Last but not least, we expect that the proposed score test is useful in dealing with longitudinal data collected from scientific investigations in biology, psychology, sociology and economics among others. In Section 4, as an illustrative example, we shall apply the proposed score test to a real longitudinal data set collected in a clinical trial study.

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. 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 $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$, $\boldsymbol{\eta}_i = (\eta(t_{i1}), \ldots, \eta(t_{in_i}))^{\mathrm{T}}$, and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})^{\mathrm{T}}$. The extended quasi-likelihood (Nelder and Pregibon 1987) of the variance components σ^2 and $\boldsymbol{\theta}$ can be expressed as

$$(2.1)\qquad l(\sigma^2, \boldsymbol{\theta}) = \prod_{i=1}^{m} E\Big\{ \exp\Big[l_i(\sigma^2; \mathbf{b}_i) \Big] \Big\},$$

where given the random-effects $\mathbf{b}_i$, $l_i(\sigma^2; \mathbf{b}_i)$ is the i-th conditional extended log-quasi-likelihood given by $l_i(\sigma^2; \mathbf{b}_i) = -\frac{1}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\eta}_i - \mathbf{Z}_i\mathbf{b}_i)^{\mathrm{T}}(\mathbf{y}_i - \boldsymbol{\eta}_i - \mathbf{Z}_i\mathbf{b}_i) - \frac{n_i}{2}\log\sigma^2$. 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

(2.2)

$$
\begin{aligned}
l_a(\sigma^2, \boldsymbol{\theta}) &= \sum_{i=1}^{m} l_i(\sigma^2; \mathbf{0}) \\
&\quad + \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{m}\left\{\frac{\partial l_i(\sigma^2;\mathbf{0})}{\partial \mathbf{b}_i}\frac{\partial l_i(\sigma^2;\mathbf{0})}{\partial \mathbf{b}_i^{\mathrm{T}}} + \frac{\partial^2 l_i(\sigma^2;\mathbf{0})}{\partial \mathbf{b}_i \partial \mathbf{b}_i^{\mathrm{T}}}\right\}\mathbf{D}(\boldsymbol{\theta})\right) \\
&= -\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\eta})^{\mathrm{T}}(\mathbf{y}-\boldsymbol{\eta}) - \frac{n}{2}\log\sigma^2 \\
&\quad + \frac{1}{2\sigma^4}(\mathbf{y}-\boldsymbol{\eta})^{\mathrm{T}}\boldsymbol{\Omega}(\mathbf{y}-\boldsymbol{\eta}) - \frac{\mathrm{tr}(\boldsymbol{\Omega})}{2\sigma^2},
\end{aligned}
$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\eta}_m^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_m^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{\Omega} = \mathrm{diag}(\mathbf{Z}_1\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}_1^{\mathrm{T}}, \ldots, \mathbf{Z}_m\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}_m^{\mathrm{T}})$.

Note that in the above extended quasi-likelihood expressions, the nuisance parameter $\boldsymbol{\eta}$ and the response vector $\mathbf{y}$ are suppressed for simplicity of the presentation. When the measurement errors ϵ_{ij} are normally distributed, the extended quasi-likelihood (2.1) reduces to the usual likelihood of σ^2 and $\boldsymbol{\theta}$. 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 later 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 $\boldsymbol{\theta}$ 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 σ^2 is a 1-dimensional nuisance parameter while $\eta(t)$ is an infinite-dimensional nuisance parameter. We will also use the nuisance parameter $\tau = \mathrm{Var}(\epsilon_{11}^2)$ 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 σ^2 and τ. 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 automatic 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., $\eta(t_{ij}) \approx \eta(t) + \eta'(t)(t_{ij}-t) = \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\pi}$ in the neighborhood of t, where $\mathbf{x}_{ij} = (1, t_{ij}-t)^{\mathrm{T}}$ and $\boldsymbol{\pi} = (\pi_0,\ \pi_1)^{\mathrm{T}}$ with $\pi_0 = \eta(t), \pi_1 = \eta'(t)$. The local linear estimator of $\eta(t)$ is then defined as $\hat{\eta}(t) = \hat{\pi}_0 = \mathbf{e}_{1,2}^{\mathrm{T}}\hat{\boldsymbol{\pi}}$, where $\mathbf{e}_{1,2} = (1,0)^{\mathrm{T}}$, and $\hat{\boldsymbol{\pi}}$ is the minimizer of the weighted least squares criterion $\sum_{i=1}^{m}\sum_{j=1}^{n_i}[y_{ij} - \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\pi}]^2 K_h(t_{ij}-t)$ where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being 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 $\hat{\eta}(t) = \mathbf{e}_{1,2}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1} \times \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}$ where the weight matrix $\mathbf{W} = \mathrm{diag}(K_h(t_{11}-t), \ldots, K_h(t_{1n_1}-t), \ldots, K_h(t_{m1}-t), \ldots, K_h(t_{mn_m}-t))$ and $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \ldots, \mathbf{X}_m^{\mathrm{T}})^{\mathrm{T}}$, with $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^{\mathrm{T}}$. The bandwidth h can be chosen by various methods, e.g., GCV; for more details, the reader is referred to Wu and Zhang (2006).

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

(2.3)

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

where $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\eta}}_1^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\eta}}_m^{\mathrm{T}})^{\mathrm{T}}$ with $\hat{\boldsymbol{\eta}}_i = (\hat{\eta}(t_{i1}), \ldots, \hat{\eta}(t_{in_i}))^{\mathrm{T}}$ 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 $\mathbf{u}_\theta = (u_{\theta_1}, \ldots, u_{\theta_p})^{\mathrm{T}}$ is easily calculated as

$$
\begin{aligned}
u_{\theta_r} &= \frac{\partial l_a(\sigma^2, \boldsymbol{\theta})}{\partial \theta_r}\Big|_{\boldsymbol{\theta}=\mathbf{0},\ \sigma^2=\hat{\sigma}^2,\ \boldsymbol{\eta}=\hat{\boldsymbol{\eta}}} \\
&= \frac{1}{2\hat{\sigma}^4}(\mathbf{y}-\hat{\boldsymbol{\eta}})^{\mathrm{T}}\left\{\dot{\boldsymbol{\Omega}}_r - \frac{\mathbf{I}_n}{n}\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r)\right\}(\mathbf{y}-\hat{\boldsymbol{\eta}}),
\end{aligned}
$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\dot{\boldsymbol{\Omega}}_r = \mathrm{diag}(\mathbf{Z}_1\dot{\mathbf{D}}_r\mathbf{Z}_1^{\mathrm{T}}, \ldots, \mathbf{Z}_m\dot{\mathbf{D}}_r\mathbf{Z}_m^{\mathrm{T}})$ with $\dot{\mathbf{D}}_r = \frac{\partial \mathbf{D}(\boldsymbol{\theta})}{\partial \theta_r}|_{\boldsymbol{\theta}=\mathbf{0}}$, $r = 1, 2, \ldots, p$.

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

$$
\begin{aligned}
\mathbf{V}_{\sigma^2\sigma^2} &= E\left(\frac{\partial l_a(\sigma^2, \boldsymbol{\theta})}{\partial \sigma^2}\frac{\partial l_a(\sigma^2, \boldsymbol{\theta})}{\partial \sigma^2}|\mathcal{D}\right)\Big|_{\boldsymbol{\theta}=\mathbf{0},\ \sigma^2=\hat{\sigma}^2,\ \tau=\hat{\tau}} \\
&= \frac{n\hat{\tau}}{4\hat{\sigma}^8}, \\
\mathbf{V}_{\sigma^2\theta_r} &= E\left(\frac{\partial l_a(\sigma^2, \boldsymbol{\theta})}{\partial \sigma^2}\frac{\partial l_a(\sigma^2, \boldsymbol{\theta})}{\partial \theta_r}|\mathcal{D}\right)\Big|_{\boldsymbol{\theta}=\mathbf{0},\ \sigma^2=\hat{\sigma}^2,\ \tau=\hat{\tau}} \\
&= \frac{\hat{\tau}}{4\hat{\sigma}^8}\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r),
\end{aligned}
$$

$$\mathbf{V}_{\theta_r\theta_s} = E\left(\frac{\partial l_a(\sigma^2,\boldsymbol{\theta})}{\partial\theta_r}\frac{\partial l_a(\sigma^2,\boldsymbol{\theta})}{\partial\theta_s}|\mathcal{D}\right)\Big|_{\boldsymbol{\theta}=\mathbf{0},\ \sigma^2=\hat\sigma^2,\ \tau=\hat\tau}$$
$$= \frac{1}{2\hat\sigma^4}\left\{\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r\dot{\boldsymbol{\Omega}}_s) - \mathrm{tr}(\dot{\boldsymbol{\Delta}}_r\dot{\boldsymbol{\Delta}}_s)\right\} + \frac{\hat\tau}{4\hat\sigma^8}\mathrm{tr}(\dot{\boldsymbol{\Delta}}_r\dot{\boldsymbol{\Delta}}_s),$$

where $\dot{\boldsymbol{\Delta}}_r = \mathrm{diag}(\dot{\boldsymbol{\Omega}}_r)$. It follows that

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

where $v_{rs}^{(1)} = \mathrm{tr}(\dot{\boldsymbol{\Omega}}_r\dot{\boldsymbol{\Omega}}_s) - \mathrm{tr}(\dot{\boldsymbol{\Delta}}_r\dot{\boldsymbol{\Delta}}_s)$, and $v_{rs}^{(2)} = \mathrm{tr}(\dot{\boldsymbol{\Delta}}_r\dot{\boldsymbol{\Delta}}_s) - \frac{1}{n}\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r)\mathrm{tr}(\dot{\boldsymbol{\Omega}}_s)$.

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

$$T = \mathbf{u}_\theta^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{u}_\theta. \tag{2.5}$$

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

$$\tilde{T} = \mathbf{u}_\theta^{\mathrm{T}}\tilde{\mathbf{V}}^{-1}\mathbf{u}_\theta, \tag{2.6}$$

where $\tilde{\mathbf{V}}$ is the associated simplified information matrix of $\boldsymbol{\theta}$ with $\tilde{\mathbf{V}} = (\tilde v_{rs})_{1\le r,s\le p}, \tilde v_{rs} = \frac{1}{2\hat\sigma^4}\{\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r\dot{\boldsymbol{\Omega}}_s) - \frac{1}{n}\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r)\mathrm{tr}(\dot{\boldsymbol{\Omega}}_s)\}$. 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 comparison against the test $\tilde T$ defined in (2.6) which is constructed based on the assumption that the measurement errors ϵ_{ij} are normally distributed.

We generated the data from the following SPME model:

$$y_{ij} = 1+2\cos(2\pi t_{ij})+\mathbf{z}_{ij}^{\mathrm{T}}\mathbf{b}_i+\epsilon_{ij},\ j=1,\dots,n_i; i=1,\dots,m, \tag{3.1}$$

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 $t_j = j/(K+1), j=1,\dots,K$ with $K=10$ here. To obtain an imbalance design, we randomly remove some design time points for a subject at a rate of 10%, so that there are about 9 measurements per subject and $9m$ measurements for all the subjects. The resulting design time points are denoted as $t_{ij}, j=1,2,\dots,n_i; i=1,2,\dots,m$ 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 $\mathbf{b}_i$ and $\mathbf{z}_{ij}$ as b_i and z_i respectively. We generate z_i from the standard uniform distribution and the random-effects b_i $(i=1,\dots,m)$ 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 γ and v can be flexibly specified so that various cases of F can be considered. In Simulation 1, the following three cases of F are considered:

- 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, $1,000$ replications are 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 χ_1^2-distribution at the nominal significance level $\alpha=5\%$. The empirical powers of T and $\tilde T$ are defined as the proportions of the rejections in $1,000$ 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

Table 1. Empirical powers of T and $\tilde{T}$ in 1,000 replications at $\alpha = 5\%$

(Simulation 1)

m	random-effects	measurement errors	T	$\tilde{T}$
50	Null	$N(0,1)$	.042	.042
	(Case 1)	$\frac{1}{\sqrt{3}}t_3$	.041	.067
	Normal	$N(0,1)$	0.874	0.874
	(Case 2)	$\frac{1}{\sqrt{3}}t_3$	0.848	0.882
	Non-normal	$N(0,1)$	0.754	0.754
	(Case 3)	$\frac{1}{\sqrt{3}}t_3$	0.693	0.752
100	Null	$N(0,1)$	.044	.044
	(Case 1)	$\frac{1}{\sqrt{3}}t_3$	.044	.092
	Normal	$N(0,1)$	0.988	0.988
	(Case 2)	$\frac{1}{\sqrt{3}}t_3$	0.951	0.978
	Non-normal	$N(0,1)$	0.947	0.947
	(Case 3)	$\frac{1}{\sqrt{3}}t_3$	0.858	0.910

Table 2. Empirical powers of T and $\tilde{T}$ in 1,000 replications at $\alpha = 5\%$

(Simulation 2)

m	random-effects	measurement errors	T	$\tilde{T}$
50	Null	$N(0,1)$	.046	.046
	(Case 1)	$\frac{1}{\sqrt{3}}t_3$	.053	.222
	Normal	$N(0,1)$	0.938	0.938
	(Case 2)	$\frac{1}{\sqrt{3}}t_3$	0.917	0.959
	Non-normal	$N(0,1)$	0.706	0.706
	(Case 3)	$\frac{1}{\sqrt{3}}t_3$	0.743	0.818
100	Null	$N(0,1)$	.043	.043
	(Case 1)	$\frac{1}{\sqrt{3}}t_3$	.048	.261
	Normal	$N(0,1)$	0.998	0.998
	(Case 2)	$\frac{1}{\sqrt{3}}t_3$	0.987	0.997
	Non-normal	$N(0,1)$	0.928	0.928
	(Case 3)	$\frac{1}{\sqrt{3}}t_3$	0.918	0.960

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 $\mathbf{z}_{ij} = [1, t_{ij}]^{\mathrm{T}}$. The random-effects $\mathbf{b}_i$ $(i = 1, \ldots, m)$ are generated from the following two-dimensional normal mixture:

$$F = 0.25N_2\Big(-0.75\boldsymbol{\gamma},\ \nu^2\Lambda\Big) + 0.75N_2\Big(0.25\boldsymbol{\gamma},\ \nu^2\Lambda\Big),$$

which has mean $\mathbf{0}$ and covariance matrix

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

Set $\mathbf{\Lambda} = \left(\begin{smallmatrix} 1 & 0.2 \\ 0.2 & 0.5 \end{smallmatrix}\right)$. As in Simulation 1, via specifying $\boldsymbol{\gamma}$ and ν, we specify the following three cases of F for study:

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

Case 2: $\boldsymbol{\gamma} = \mathbf{0}$, $\nu^2 = 0.1$ so that $\boldsymbol{\theta} = (0.1,\ 0.02,\ 0.05)^{\mathrm{T}}$, specifying an alternative model with normal random-effects $\mathbf{b}_i$.

Case 3: $\boldsymbol{\gamma} = (0.4, -0.2)^{\mathrm{T}}$, $\nu^2 = 0.05$ so that $\boldsymbol{\theta} = (0.08, -0.005, 0.0325)^{\mathrm{T}}$, specifying an alternative model with non-normal random-effects $\mathbf{b}_i$.

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

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 a nonparametric mixed-effects model. We are interested in whether the CD4 cell counts are different among the patients. We handled this problem via the following SPME model:

(4.1)
$$y_{ij} = \eta(\text{week}_{ij}) + \mathbf{z}_{ij}^{\mathrm{T}}\mathbf{b}_i + \epsilon_{ij},\ j = 1, \ldots, n_i; i = 1, 2, \ldots, 166,$$

where $\mathbf{z}_{ij} = (1, \text{week}_{ij}, \text{week}_{ij}^2)^{\mathrm{T}}, \mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})^{\mathrm{T}}$ with mean vector $\mathbf{0}$ and covariance matrix

$$\mathbf{D}(\boldsymbol{\theta}) = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_2 & \theta_4 & \theta_5 \\ \theta_3 & \theta_5 & \theta_6 \end{pmatrix},$$

and considering the null hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_6 = 0$. Since there is no information about the normality of the random-effects $\mathbf{b}_i$ and the measurement errors ϵ_{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 model. 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 (Carter and Yang 1986; Wu and Zhang 2006). When heterogeneity across clusters is taken into account with the inclusion of random-effects in the model, the score test may be used to ascertain whether the fixed-effects model without the random-effects is adequate to fit the data. Other potential applications include a homogeneity test for variation over groups. For example, in genetic epidemiology, such tests can be used to study if familial aggregation of a disease, which may be determined by genetic factors, is homogeneous in different families; in Cobb-Douglas production function studies, it is of practical interest to test the homogeneity of the industrial variation in the effects of input on production. When the null hypothesis $H_0 : \boldsymbol{\theta} = 0$ is rejected, it is of practical interest to test if some of the variance components in $\boldsymbol{\theta}$ are zero. The studies in these directions are warranted.

ACKNOWLEDGMENTS

The authors thank the Co-editor and a referee for their invaluable comments which helped improve the presentation of the paper substantially. Thanks also go to the second author's research assistant Ms. Jing Han for her help in editing the tables presented in the paper.

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. The proof of Lemma 1 can be found in Sun and Zhang (2010), an online version of this paper.

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

$$E\Big\{\boldsymbol{\epsilon}^T\mathbf{A}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\mathbf{B}\boldsymbol{\epsilon}\Big\} = 2\sigma^4\Big[tr(\mathbf{AB}) - tr\{diag(\mathbf{A})diag(\mathbf{B})\}\Big] + \tau\, tr\{diag(\mathbf{A})diag(\mathbf{B})\} + \sigma^4 tr(\mathbf{A})tr(\mathbf{B}),$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_M)^T$ and $\tau = Var(\epsilon_1^2)$.

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

$$\mathbf{R}_n = \sum_{r=1}^{p} \alpha_r \left\{ \dot{\boldsymbol{\Omega}}_r - \frac{\mathrm{tr}(\dot{\boldsymbol{\Omega}}_r)}{n}\mathbf{I}_n \right\},$$

$$\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^{\mathrm{T}}, \ldots, \boldsymbol{\epsilon}_m^{\mathrm{T}})^{\mathrm{T}}, \boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^{\mathrm{T}}.$$

Further, write $\mathbf{V}^{(1)} = (v_{rs}^{(1)})_{1\le r,s\le p}, \mathbf{V}^{(2)} = (v_{rs}^{(2)})_{1\le r,s\le p}$, where $v_{rs}^{(1)}$ and $v_{rs}^{(2)}$ are as defined in Section 2. Finally define the norms of a vector $\mathbf{a}$ and a matrix $\mathbf{A}$ as $\|\mathbf{a}\| = (\mathbf{a}^{\mathrm{T}}\mathbf{a})^{\frac{1}{2}}$ and $\|\mathbf{A}\| = \{\mathrm{tr}(\mathbf{A}\mathbf{A}^{\mathrm{T}})\}^{\frac{1}{2}}$ respectively. The following regularity conditions are imposed for Theorem 1:

1. (1) The number of measurements per subject is bounded, i.e., $n_i < C, i = 1, \ldots, m$ for some $0 < C < \infty$. Again $n = \sum_{i=1}^m n_i$ denotes the total number of measurements for the whole dataset.
2. (2) The measurement errors ϵ_{ij} are i.i.d. with $E\epsilon_{11}^4 < \infty$; the random-effects covariates satisfy $\max_{1\le i\le m; 1\le j\le n_i} E\|\mathbf{z}_{ij}\|^4 < \infty$.
3. (3) The largest eigenvalues of matrices $\dot{\mathbf{D}}_r, r = 1, \ldots, p$ are bounded, and for some $\delta > 0$, we have $\max_{1\le i\le m, 1\le j,l\le n_i} \mathrm{E}(\|\mathbf{z}_{ij}\|\|\mathbf{z}_{il}\|)^{2+\delta} < \infty$.
4. (4) There exist nonnegative definite matrices $\mathbf{V}_0^{(1)}$ and $\mathbf{V}_0^{(2)}$ such that $n^{-1}\mathbf{V}^{(1)} \xrightarrow{P} \mathbf{V}_0^{(1)}, n^{-1}\mathbf{V}^{(2)} \xrightarrow{P} \mathbf{V}_0^{(2)}$, and $V_0 = \frac{1}{2\sigma^4}\mathbf{V}_0^{(1)} + \frac{\tau}{4\sigma^8}\mathbf{V}_0^{(2)}$ is positive definite.
5. (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. (6) The second derivative function $\eta''(t)$ is bounded on $\mathcal{T}$ and $\eta''(t) \neq 0, t \in \mathcal{T}$.
7. (7) The kernel function $K(u)$ is a symmetric probability density function having a compact support, e.g., $[-1, 1]$.
8. (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 Conditions (3) and (4) for easy presentation although it is difficult to check in practice. From the proof of Theorem 1, we can see that $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ must be nonnegative matrices since $\mathrm{Var}(J_{n1}|\mathcal{D}) = 2\sigma^4\boldsymbol{\alpha}^T\mathbf{V}^{(1)}\boldsymbol{\alpha}$ and $\mathrm{Var}(J_{n2}|\mathcal{D}) = \tau\boldsymbol{\alpha}^T\mathbf{V}^{(2)}\boldsymbol{\alpha}$ must be nonnegative for all $\boldsymbol{\alpha}$. 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 $t \in \mathcal{T}$, the asymptotic conditional bias and variance of $\hat{\eta}(t)$ are

(A.1)
$$\mathrm{Bias}\{\hat{\eta}(t)|\mathcal{D}\} = O_P(h^2),\ \mathrm{Var}\{\hat{\eta}(t)|\mathcal{D}\} = O_P\{(nh)^{-1}\},$$

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 σ^2 and τ respectively. Therefore, by Condition (4), we have $n^{-1}\mathbf{V} \xrightarrow{P} \frac{1}{2\sigma^4}\mathbf{V}_0^{(1)} + \frac{\tau}{4\sigma^8}\mathbf{V}_0^{(2)} = \mathbf{V}_0$. Theorem 1 will follow if we can show that under H_0, $n^{-1/2}\mathbf{u}_\theta \xrightarrow{L} N(0,\mathbf{V}_0)$ as $n\to\infty$. It is equivalent to show that under H_0, for any given $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_p)^T$, we have

$$n^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}_\theta \xrightarrow{L} N(0,\boldsymbol{\alpha}^T\mathbf{V}_0\boldsymbol{\alpha}). \tag{A.2}$$

For this end, we write

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}_\theta = \frac{1}{2\hat{\sigma}^4}\Big\{&\boldsymbol{\epsilon}^{\mathrm{T}}(\mathbf{R}_n - \mathbf{Q}_n)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{Q}_n\boldsymbol{\epsilon}\\
&+ E[(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^{\mathrm{T}}|\mathcal{D}]\mathbf{R}_n E[(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})|\mathcal{D}]\\
&+ [\hat{\boldsymbol{\eta}} - E(\hat{\boldsymbol{\eta}}|\mathcal{D})]^{\mathrm{T}}\mathbf{R}_n[\hat{\boldsymbol{\eta}} - E(\hat{\boldsymbol{\eta}}|\mathcal{D})]\\
&- 2\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{R}_n E[(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})|\mathcal{D}]\\
&- 2\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{R}_n[\hat{\boldsymbol{\eta}} - E(\hat{\boldsymbol{\eta}}|\mathcal{D})]\\
&+ 2[\hat{\boldsymbol{\eta}} - E(\hat{\boldsymbol{\eta}}|\mathcal{D})]^{\mathrm{T}}\mathbf{R}_n E[(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})|\mathcal{D}]\Big\}\\
\equiv \frac{1}{2\hat{\sigma}^4}\Big\{&J_{n1} + J_{n2} + J_{n3} + J_{n4}\\
&- 2J_{n5} - 2J_{n6} + 2J_{n7}\Big\},
\end{aligned}$$

where $\mathbf{Q}_n = \mathrm{diag}(\mathbf{R}_n)$, a diagonal matrix having the same diagonal entries as $\mathbf{R}_n$.

By straightforward calculation, we have

$$J_{n1} = \sum_{i=1}^m\left\{\sum_{j=1}^{n_i}\sum_{l=1,l\neq j}^{n_i}\mathbf{z}_{ij}^{\mathrm{T}}\Big(\sum_{r=1}^p\alpha_r\dot{\mathbf{D}}_r\Big)\mathbf{z}_{il}\epsilon_{ij}\epsilon_{il}\right\},$$

with $\mathrm{E}(J_{n1}|\mathcal{D}) = \sigma^2\mathrm{tr}(\mathbf{R}_n - \mathbf{Q}_n) = 0$ and $\mathrm{Var}(J_{n1}|\mathcal{D}) = \mathrm{E}(J_{n1}^2|\mathcal{D}) = 2\sigma^4\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}^{(1)}\boldsymbol{\alpha}$ obtained by applying Lemma 1. By Condition (4), as $n\to\infty$, we have $\mathrm{E}\{n^{-1/2}J_{n1}\} = 0$ and $\mathrm{Var}\{n^{-1/2}J_{n1}\} \to 2\sigma^4\boldsymbol{\alpha}^T\mathbf{V}_0^{(1)}\boldsymbol{\alpha}$. Thus, under Conditions (1)-(4) and by the Lindeberg-Feller central limit theorem, as $n\to\infty$, $n^{-\frac{1}{2}}J_{n1} \xrightarrow{L} N(0, 2\sigma^4\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_0^{(1)}\boldsymbol{\alpha})$.

Since $\mathbf{Q}_n$ is a diagonal matrix and $\mathrm{tr}(\mathbf{Q}_n) = 0$, we have $\mathrm{E}(J_{n2}|\mathcal{D}) = \sigma^2\mathrm{tr}(\mathbf{Q}_n) = 0$ and $\mathrm{Var}(J_{n2}|\mathcal{D}) = \mathrm{E}(J_{n2}^2|\mathcal{D}) = \tau\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}^{(2)}\boldsymbol{\alpha}$ obtained by applying Lemma 1 again, where $\tau = \mathrm{Var}(\epsilon_{11}^2)$ as defined before. By Condition (4), we have $\mathrm{Var}\{n^{-1/2}J_{n2}\} \to \tau\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_0^{(2)}\boldsymbol{\alpha}$, as $n\to\infty$. Since $\mathbf{Q}_n$ 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 $n^{-\frac{1}{2}}J_{n2} \xrightarrow{L} N(0,\tau\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_0^{(2)}\boldsymbol{\alpha})$. Since $\mathrm{Cov}(J_{n1},J_{n2}) = \mathrm{E}\{\mathrm{E}(J_{n1}J_{n2}|\mathcal{D})\} = 2\sigma^4\mathrm{E}[\mathrm{tr}((\mathbf{R}_n - \mathbf{Q}_n)\mathbf{Q}_n)] = 0$, J_{n1} and J_{n2} are uncorrelated. Therefore, as $n\to\infty$, we have $n^{-\frac{1}{2}}\{J_{n1}+J_{n2}\} \xrightarrow{L} N(0,\boldsymbol{\alpha}^{\mathrm{T}}[2\sigma^4\mathbf{V}_0^{(1)} + \tau\mathbf{V}_0^{(2)}]\boldsymbol{\alpha})$. Then we will have $n^{-\frac{1}{2}}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}_\theta \xrightarrow{L} N(0,\frac{1}{4\sigma^8}\boldsymbol{\alpha}^{\mathrm{T}}\{2\sigma^4\mathbf{V}_0^{(1)} + \tau\mathbf{V}_0^{(2)}\}\boldsymbol{\alpha})$, if we can show that

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

The expression (A.2) then follows by noticing that $\frac{1}{4\sigma^8}\{2\sigma^4\mathbf{V}_0^{(1)} + \tau\mathbf{V}_0^{(2)}\} = \frac{1}{2\sigma^4}\mathbf{V}_0^{(1)} + \frac{\tau}{2\sigma^8}\mathbf{V}_0^{(2)} = \mathbf{V}_0$. The proof of (A.3) is much involved and is hence omitted here for space saving. Interested readers are referred to Sun and Zhang (2010), an online version of this paper, for the complete proof of Theorem 1. □

Received 7 May 2010

REFERENCES

[1] Breslow, N. E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **88** 9–25.

[2] Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, **82** 81–91. MR1332840

[3] Carter, R.L. and Yang, M. C. K. (1986). Large sample inference in random coefficient regression models. *Communications in Statistics – Theory and Methods*, **15** 2507–2535. MR0853026

[4] Cox, D.R. and Hinkeley, D.V. (1974). *Theoretical statistics*. Chapman and Hall, London. MR0370837

[5] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87** 998–1004. MR1209561

[6] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.*, **27** 1491–1518. MR1742497

[7] Hall, D.B. and Praestgaard, J.T. (2001). Order-restricted score tests for homogeneity in generalized linear and nonlinear mixed models. *Biometrika*, **88** 739–751. MR1859406

[8] Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics, continuous univariate distributions-1*. Houghton-Mifflin, Boston. MR0270475

[9] Lin, X. (1997). Variance component testing in generalised linear models with random-effects. *Biometrika*, **84** 309–326. MR1467049

[10] McCullagh, P. G. and Nelder, J.A. (1989). *Generalized linear models*, 2nd ed. Chapman and Hall, London. MR0727836

[11] Nelder, J.A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, **74** 221–232. MR0903123

[12] Park, J. and Wu, H. (2006). Backfitting and local likelihood methods for nonparametric mixed-effects models with longitudinal data. *J. Statist. Plann. & Infer.*, **136** 3760–3782. MR2299164

[13] Solomon, P. J. and Cox, D. R. (1992). Nonlinear component of variance models. *Biometrika*, **79** 1–11. MR1158513

[14] Sun, Y. and Zhang, J. T. (2010). A score test for variance components in a semiparametric mixed-effects model under nonnormality. *Manuscript*, available at http://www.stat.nus.edu.sg/~zhangjt/SubPages/research.htm.

[15] Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59** 254–262. MR1987392

[16] Wang, Y. (1998). Mixed-effects smoothing spline analysis of variance. *J. R. Statist. Soc. B*, **60** 159–174. MR1625640

[17] Wang, H.X. and He, X. (2007). Detecting Differential Expressions in GeneChip Microarray Studies: A Quantile Approach. *J. Amer. Statist. Assoc.*, **102** 104–112. MR2293303

[18] Wu, H. and Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data: Mixed-Effects Modeling Approaches*. Wiley, New York. MR2216899

[19] Zhu, H. and Zhang, H. (2006). Generalized score test of homogeneity for mixed effects models. *Ann. Statist.*, **34** 1545–1569. MR2278367

[20] Zhu, Z. and Fung, W. K. (2004). Variance component testing in semiparametric mixed models. *J. Multi. Anal.*, **91** 107–118. MR2083907

Yan Sun
School of Economics
Shanghai Univ. of Finance and Economics
P.R. China
E-mail address: yansun2002cn@yahoo.com.cn

Jin-Ting Zhang
Dept of Stat. and Appl. Prob.
National Univ. of Singapore
Singapore
E-mail address: stazjt@nus.edu.sg