Generalized Score Test of Homogeneity for Mixed Effects Models: Supplement *

By Hongtu Zhu and Heping Zhang[1]
Columbia University and Yale University

Summary

In this document, we introduce five additional examples that motivate our research on testing homogeneity. In particular, we show how to check the assumptions for two linear mixed models to validate the asymptotic distributions of the score test statistics. Then, we give the detailed derivation for the score test statistic. We present the result about the local power of the score test statistics under a class of mixed effect models. Finally, we compare our method with some existing ones.

*[1]Supported in part by NIH Grant DA12468, DA017713 and DA016750.

AMS 2000 subject classifications. Primary 62F05; secondary 62F40.

Key words and phrases. Functional central limit theorem, latent variable, random quadratic form, score test, variance component.

1 Five Additional Examples

In the following, we will give five additional examples to illustrate the potential applications of the score test statistics proposed in Zhu and Zhang (2005). Unless stated otherwise, we will refer to, without re-introduction of, the notation and terminology in Zhu and Zhang (2005).

We give two examples having unidentified nuisance parameters under the null hypothesis, which include factor analysis model and random coefficient model. Other examples with unidentified nuisance parameters can be found in Hansen (1996), Andrews (2001) and references therein. We also apply the score test statistics in an epileptic study. Finally, we study linear mixed effect model and crossed design linear mixed effect model in detail and demonstrate how to apply the score test statistics and check the assumptions to ensure their asymptotic properties. In particular, the last two examples reveal that our theory covers the case with both large n and large m_i. To our knowledge, this is new for testing homogeneity of mixed effects models.

1.1 Factor Analysis Model

Consider the following factor analytic measurement model for an $m \times 1$ manifest random vector Y_i:

$$Y_i = \mu + \Lambda \mathbf{b}_i + \epsilon_i \quad i = 1, \cdots, n, \tag{1}$$

where $\mu(m \times 1)$ is the mean vector, $\Lambda(m \times q)$ is the factor loading matrix, $\mathbf{b}_i$ $(q \times 1)$ is a factor score vector with distribution $N[\mathbf{0}, \Sigma(\gamma)]$, and ϵ_i $(m \times 1)$ is a random vector of measurement errors with distribution $N[\mathbf{0}, \Phi_\epsilon]$, where $\Phi_\epsilon = \mathrm{diag}(\phi_1, \cdots, \phi_m)$. In addition, $\mathbf{b}_i$'s are independent of each other. For confirmatory factor analysis models, $\Sigma(\gamma)$ is usually an unstructured covariance matrix whereas Λ has a special pattern. We are interested in whether factor scores are warranted, that is, $H_0: \ \Sigma(\gamma) = 0$. It should be noted that the factor loading matrix Λ only appears under the alternative hypothesis, but not under the null hypothesis. The same question generally arises from latent variable models with mixed continuous and polytomous data; see Shi and Lee (2001) and Sammel, Ryan and Legler (1997).

1.2 Random Coefficient Model

Consider the following random coefficient model:

$$y_i = \mathbf{x}_i^T \beta + f(\mathbf{z}_i, \gamma_{(1)}) b_i + \epsilon_i \quad i = 1, \cdots, n, \tag{2}$$

where $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. We are interested in testing the null hypothesis that the variance of the random coefficient is zero. When $\sigma_b = 0$, the parameter $\gamma_{(1)}$ cannot be identified.

Specific examples of $f(\mathbf{z}_i, \gamma_{(1)})$ include Box-Cox transformation ($f(\mathbf{z}_i, \gamma_{(1)}) = (z_i^{\gamma_{(1)}} - 1)/\gamma_{(1)}$), structure change ($f(\mathbf{z}_i, \gamma_{(1)}) = \mathbf{1}(i/n \leq \gamma_{(1)})z_i$), and threshold models ($f(\mathbf{z}_i, \gamma_{(1)}) = \mathbf{1}(z_i \leq \gamma_{(1)})z_i$). Note that the score statistic in Lin (1997) and Hall and Praestgaard (2001) cannot handle this kind of random coefficient models.

1.3 An Epileptic Study

Table 2 of Thall and Vail (1990) presented a data set from a clinical trial of 59 epileptics. These epileptic patients were randomly assigned to treatment (T=1) and placebo (T=0) group as an adjutant to the standard chemotherapy. Every patient reported the number of seizures in each of the four 2-week observation periods. Thall and Vail (1990) proposed to use a Poisson random effect model. Specifically, conditional on $\mathbf{b}_i = (b_{i1}, b_{i2})^T$, the seizure count, y_{ij}, for the i-th patient in the j-th visit is assumed to follow Poisson distribution with mean μ_{ij} such that $\log \mu_{ij}(\mathbf{b}_i) = \mathbf{x}_{ij}^T\beta + b_{i1} + b_{i2}\text{Visit}_j/10$. The covariates $\mathbf{x}_{ij}$ include the intercept term, the logarithm of a pre-experiment baseline count of seizures (B), treatment (T), their interaction (B$\times$T), the logarithm of the patient's age and a variable, Visit_j, for each of four clinic visits (-3, -1, 1, 3). We are interested in testing whether the random effects $\mathbf{b}_i$ are warranted for these data. Without considering the random effects, we obtained the maximum likelihood estimate of β as $(-2.80, 0.95, -1.34, 0.56, 0.90, -0.29)$. S_O, S_P and S_S for this dataset are 32.30, 23.47, and 36.77, respectively. Setting $r_0 = 10,000$, we obtained the corresponding p-values as 0.038, 0.0005 and 0.0002, respectively. Thus, our tests offer significant support for including the random effects, and modest evidence for overdispersion.

1.4 Linear Mixed Effects Model

For simplicity, we follow the notation in Stram and Lee (1994) and write the linear mixed effects models as

$$Y_i = X_i\beta + Z_i\mathbf{b}_i + \epsilon_i, \tag{3}$$

where $Z_i = (\mathbf{z}_{i,1}, \cdots, \mathbf{z}_{i,m_i})^T$ is a known covariate matrix. The random coefficients $\mathbf{b}_i$'s $(q \times 1)$ and the residual vectors ϵ_i's $(m_i \times 1)$ are normally distributed such that $E[\mathbf{b}_i] = 0$, $E[\mathbf{b}_i\mathbf{b}_i^T] = \Sigma(\gamma)$, $E[\epsilon_i] = 0$ and $E[\epsilon_i\epsilon_i^T] = \phi I_{m_i}$, where I_{m_i} is the $m_i \times m_i$ identity matrix. Moreover, for $i \neq i'$, $\mathbf{b}_i$, $\mathbf{b}_{i'}$, ϵ_i and $\epsilon_{i'}$ are independent of each other.

Using the parametrization in the equation (6) of Zhu and Zhang (2005), we see that $W_{i,j}(\gamma)$ equals to $W(\gamma)$ for $i = j$ and zero otherwise. After some calculations, for model (3), we have

$U_{i,j} = e_{i,j}/\phi$, $V_{i,j} = 1/\phi$ and

$$b_{K,K'}(\gamma) = \begin{cases} \mathbf{z}_{(i,j)}^T W(\gamma)\mathbf{z}_{(i,j')}, & i = i' \\ 0, & i \neq i' \end{cases},$$

where $e_{i,j} = y_{i,j} - \mathbf{x}_{i,j}^T\beta$. We also have

$$T_O(\gamma) = \sum_K \mathbf{z}_K^T W(\gamma)\mathbf{z}_K[e_K^2 - \phi]/\phi^2 \text{ and } T_P(\gamma) = \sum_{i=1}^n \sum_{j\neq j'} \mathbf{z}_{i,j}^T W(\gamma)\mathbf{z}_{i,j'} e_{i,j} e_{i,j'}/\phi^2.$$

Furthermore, we have $\phi^{1/2}U_{i,j} \sim N(0,1)$, $E[U_{i,j}^2] = \phi^{-1}$ and $E[U_{i,j}^4] = 3\phi^{-2}$ for all (i,j)'s. Thus, $E[T_O(\gamma)T_O(\gamma')] = 2\sum_{(i,j)} \mathbf{z}_{i,j}^T W(\gamma)\mathbf{z}_{i,j}\mathbf{z}_{i,j}^T W(\gamma')\mathbf{z}_{i,j}\phi^{-2}$ and

$$E[T_P(\gamma)T_P(\gamma')] = 2\sum_{i=1}\sum_{j\neq j'} \mathbf{z}_{i,j}^T W(\gamma)\mathbf{z}_{i,j'}\mathbf{z}_{i,j}^T W(\gamma')\mathbf{z}_{i,j'}\phi^{-2}.$$

As discussed in Zhu and Zhang (2005), we need to replace ξ_* with its estimator under H_0. Let $\hat{\beta}$ and $\hat{\phi}$ be the maximum likelihood estimates of β and ϕ under H_0. In this example, we have $\hat{\beta} = (X^TX)^{-1}X^TY$ and $\hat{\phi} = N^{-1}Y^T(I_N - P_X)Y$, where $Y = (Y_1^T, \cdots, Y_n^T)^T$, $X = (X_1^T, \cdots, X_P^T)^T$ and $P_X = X(X^TX)^{-1}X^T$. Thus, we have

$$\hat{\beta} - \beta_* = (\sum_K \mathbf{x}_{i,j}^T\mathbf{x}_{i,j})^{-1}\sum_K e_{i,j}\mathbf{x}_{i,j} \text{ and } \hat{\phi} - \phi_* = N^{-1}\sum_K (e_{i,j}^2 - \phi_*) + O_P(N^{-1}),$$

which give the exact form of F_K for each $K = (i,j)$. Moreover, we can calculate that $J_N(\gamma) = (0, N^{-1}\phi_*^{-2}\sum_K \mathbf{z}_K^T W(\gamma)\mathbf{z}_K)^T$, and $\hat{T}_O(\gamma) = \sum_K \mathbf{z}_K^T W(\gamma)\mathbf{z}_K(\hat{e}_K^2 - \hat{\phi})/\hat{\phi}^2$ can be approximated by

$$\sum_K \{\mathbf{z}_K^T W(\gamma)\mathbf{z}_K - N^{-1}[\sum_K \mathbf{z}_K^T W(\gamma)\mathbf{z}_K]\}(e_{K,*}^2 - \phi_*)/\phi_*^2,$$

where $\hat{e}_K = y_K - \mathbf{x}_K^T\hat{\beta}$ and $e_{K,*} = y_K - \mathbf{x}_K^T\beta_*$. Finally, we get $I_{EO}(\gamma) = 2\sum_K\{\mathbf{z}_K^T W(\gamma)\mathbf{z}_K - N^{-1}[\sum_K \mathbf{z}_K^T W(\gamma)\mathbf{z}_K]\}^2/\hat{\phi}^2$. When all $\mathbf{z}_K$s are the same, $\hat{T}_O(\gamma)$ is exactly equal to zero and $\hat{T}_S(\gamma)$ reduces to $\hat{T}_P(\gamma)$. When $q = 1$ and $\mathbf{z}_K = 1$ for all K, our results are similar to those of Jacqmin-Gadda and Commenges (1995).

In terms of $T_P(\gamma)$, we see that $x_K = U_K/\sqrt{\phi_*} = (y_K - \mathbf{x}_K^T\beta_*)/\sqrt{\phi_*}$. For model (3), some assumptions like (C1) are trivial and others such as (C4) are necessary conditions. Without loss of generality, we will validate (C2) and (C3) in two simple cases.

The first model assumes that $m_1 = \cdots = m_n = 2$. Then, we have

$$X_P(\gamma) = \frac{\sum_{i=1}^n \mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2}e_{i,1}e_{i,2}}{\sqrt{\sum_{i=1}^n [\mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2}]^2}} = \sum_{K,K'} c_{K,K'}(\gamma)x_K x_{K'}.$$

Assumption (C2) relates to the maximum of the absolute eigenvalues of matrix $C(\gamma) = (c_{K,K'})$, denoted by $\mu_{\max}[C(\gamma)]$. After some calculations, we get

$$\mu_{\max}[C(\gamma)] = \frac{\max_i |\mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2})|}{\sqrt{\sum_{i=1}^n [\mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2}]^2}}.$$

Assumption(C2) requires that $\mu_{\max}[C(\gamma)]$ converges to zero uniformly. After some algebraic manipulations, we have

$$\sum_{i=1}^n [\partial_{\gamma_t} c_{K,K'}(\gamma)]^2 \leq C + C\frac{[\sum_{i=1}^n \mathbf{z}_{i,1}^T \partial_{\gamma_t} W(\gamma)\mathbf{z}_{i,2}\mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2}]^2}{[\sum_{i=1}^n (\mathbf{z}_{i,1}^T W(\gamma)\mathbf{z}_{i,2})^2]^2}.$$

If the supreme of the right hand side of above inequality is finite, then (C3) holds.

The second example is that $\mathbf{z}_{i,j} = \mathbf{z}$ for all (i,j)'s, e.g., $\mathbf{z} = \mathbf{1}_q$. In this simple case, we are interested in whether the simplest variance structure of ϵ_i is adequate. Thus, $T_P(\gamma) = \mathbf{z}^T W(\gamma)\mathbf{z} \sum_{i=1}^n \sum_{j\neq j'} e_{i,j}e_{i,j'}\phi^{-2}$ and $C_P(\gamma,\gamma') = 2\mathbf{z}^T W(\gamma)\mathbf{z}\mathbf{z}^T W(\gamma')\mathbf{z}\sum_{i=1}^n m_i(m_i-1)\phi^{-2}$. We have $X_P(\gamma) = \sum_{i=1}^n \sum_{j\neq j'} e_{i,j}e_{i,j'}\phi^{-1}/\sqrt{2\sum_{i=1}^n m_i(m_i-1)}$, which is independent of γ. Therefore, (C3) are unnecessary and others are trivial except (C2). Due to the special structure of $C = (c_{K,K'})$, we can show that

$$\mu_{\max}[C] = \max_i (m_i - 1)/\sqrt{2\sum_{i=1}^n m_i(m_i-1)}.$$

There are too many sequences $\{(n, m_1, \cdots, m_n)\}$ satisfying condition (C3); for example, $\{m_i = i : i = 1, \cdots, n\}$ is such a sequence. The forgoing discussions lead to that $X_P(\gamma)$ converges to the standard normal distribution.

Replacing ξ_* by $\hat{\xi}$ in $X_P(\gamma)$ gives $\hat{X}_P(\gamma)$. It follows that the asymptotic variance of $\hat{X}_P(\gamma)$ is given by $I_{EP}(\gamma) = 2\sum_{i=1}^n \sum_{j\neq j'} (\mathbf{z}_{i,j}^T W(\gamma)\mathbf{z}_{i,j})^2\phi_*^{-2}$. To consider the process $\hat{X}_P(\gamma)$, we need to introduce $U_K(s_K, \xi) = (y_K - \mathbf{x}_K^T\beta)/\phi$. In addition, $U_K(s_K,\xi)$ can be decomposed into two terms:

$$\frac{(y_K - \mathbf{x}_K^T\beta_*)}{\phi_*}\frac{\phi_*}{\phi} + \mathbf{x}_K^T(\beta_* - \beta)\phi_*^{-1}\frac{\phi_*}{\phi}.$$

Let $\phi_*/\phi = 1 + N^{-1/2}h_\phi$, $\beta_* - \beta = N^{-1/2}\mathbf{h}_\beta$ and $\mathbf{h} = (h_\phi, \mathbf{h}_\beta^T)^T$. We have

$$\mu_K(N^{-1/2}\mathbf{h}) = EU_K(s_K, N^{-1/2}\mathbf{h}) = N^{-1/2}\mathbf{x}_{i,j}^T\mathbf{h}_\beta\phi_*^{-1}(1 + N^{-1/2}h_\phi).$$

Let $\tilde{U}_K(s_K, N^{-1/2}\mathbf{h}) = U_K(s_K, N^{-1/2}\mathbf{h}) - \mu_K(N^{-1/2}\mathbf{h}) = U_K(1+N^{-1/2}h_\phi)$. Now we can directly check (C6), (C7) and (C8), because

$$\tilde{U}_K(y_K, N^{-1/2}\mathbf{h})\tilde{U}_{K'}(s_{K'}, N^{-1/2}\mathbf{h}) - U_K U_{K'} = U_K U_{K'}[2N^{-1/2}h_\phi + N^{-1}h_\phi^2],$$
$$\tilde{U}_K(y_K, N^{-1/2}\mathbf{h}) - \tilde{U}_K(y_K, N^{-1/2}\mathbf{h}') = U_K N^{-1/2}(h_\phi - h_\phi').$$

We see that (C8) is given by

$$\sup_{\gamma\in\Gamma,||\mathbf{h}||_2\le M}\frac{|\sum_{K\neq K'}\mathbf{z}_K^TW(\gamma)\mathbf{z}_{K'}U_K\mathbf{x}_K^T\mathbf{h}_\beta|}{\sqrt{\sum_{K\neq K'}(\mathbf{z}_K^TW(\gamma)\mathbf{z}_{K'})^2}}\phi_*^{-1}N^{-1/2}(1+N^{-1/2}h_\phi)^2=o_p(1).$$

Some extra mild conditions are good enough to ensure (C8). So we omit the details. Until now, we have given all sufficient conditions for Theorems 3 and 5.

1.5 Crossed Design Linear Mixed Models

Following Lin (1997) and Hall and Praestgaard (2001), we consider the following cross-design linear mixed model

$$y_{i,j}=\mathbf{x}_{i,j}^T\beta+\mathbf{a}_i+\mathbf{c}_j+\epsilon_{i,j}, \tag{4}$$

for $i=1,\cdots,n$ and $j=1,\cdots,m$. For simplicity, we assume that $\mathbf{a}_i$, $\mathbf{c}_j$ and $\epsilon_{i,j}$ are independent of each other, $E\mathbf{a}_i=E\mathbf{c}_j=0$, $E\mathbf{a}_i^2=\sigma_a^2$, $E\mathbf{c}_j^2=\sigma_c^2$, and $\epsilon_{i,j}\sim N(0,\phi)$. Let $\mathbf{b}_1=(\mathbf{a}_1,\cdots,\mathbf{a}_n,\mathbf{c}_1,\cdots,\mathbf{c}_m)^T$, we have $E[\mathbf{b}_1\mathbf{b}_1^T]=\Sigma(\gamma)=\text{diag}(\sigma_a^2,\cdots,\sigma_a^2,\sigma_c^2,\cdots,\sigma_c^2)$. Define σ_T as $\sigma_a^2+\sigma_c^2$, we see that

$$\Sigma(\gamma)=\sigma_TW(\gamma)=\sigma_T\text{diag}(\cos^2(\gamma_1),\cdots,\cos^2(\gamma_1),\sin^2(\gamma_1),\cdots,\sin^2(\gamma_1)),$$

where $\gamma_1\in[-0.5\pi,0.5\pi]$ and $\cos^2(\gamma_1)=\sigma_a^2/\sigma_T$. Thus, (4) becomes

$$y_{i,j}=\mathbf{x}_{i,j}^T\beta+\mathbf{1}_{i,j}^T\mathbf{b}_1+\epsilon_{i,j}=\mathbf{x}_{i,j}^T\beta+\sigma_T^{1/2}\mathbf{1}_{i,j}^T\mathbf{u}_1+\epsilon_{i,j},$$

where $\mathbf{b}_1=\sigma_T^{1/2}\mathbf{u}_1$ and $\mathbf{1}_{i,j}$ is an R^{n+m} vector with both the ith and $(n+j)$th entries being 1 and all other entries being zero. Similar to the derivation in Example 2, we have $U_{i,j}=e_{i,j}/\phi$, $V_{i,j}=1/\phi$ and

$$b_{K,K'}(\gamma)=\mathbf{1}_K^TW(\gamma)\mathbf{1}_{K'}=\begin{cases}1, & K=K';\\ \cos^2(\gamma_1), & i=i',j\neq j';\\ \sin^2(\gamma_1), & i\neq i',j=j';\\ 0, & i\neq i',j\neq j'.\end{cases}$$

Therefore, $T_S(\gamma)=T_O(\gamma)+T_P(\gamma)$, where $T_O(\gamma)=\sum_{i=1}^n\sum_{j=1}^m(e_{i,j}^2-\phi)/\phi^2$ and

$$T_P(\gamma)=\sum_{i=1}^n\sum_{j\neq j'}\cos^2(\gamma_1)e_{i,j}e_{i,j'}/\phi^2+\sum_{j=1}^m\sum_{i\neq i'}\sin^2(\gamma_1)e_{i,j}e_{i',j}/\phi^2.$$

Thus, $E[T_O(\gamma)T_O(\gamma')]=nm\phi^{-2}$ and

$$E[T_P(\gamma)T_P(\gamma')]=2nm\phi^{-2}[(m-1)\cos^2(\gamma_1)\cos^2(\gamma_1')+(n-1)\sin^2(\gamma_1)\sin^2(\gamma_1')].$$

Since the crossed design linear mixed model reduces to a linear regression model under the null hypothesis, the maximum likelihood estimates of $\hat{\beta}$ and $\hat{\phi}$ in the previous example can be used here to replace ξ in all test statistics. By similar arguments, we obtain $\hat{T}_O(\gamma) = 0$ and $\hat{T}_P(\gamma) = \hat{T}_S(\gamma)$.

Let us discuss how Theorems 3-5 in Zhu and Zhang (2005) can be applied to derive the asymptotic distributions of $T_P(\gamma)$ and $\hat{T}_P(\gamma)$. We already know that $I_{TP}(\gamma) = 2nm\phi_*^{-2}[(m-1)\cos^4(\gamma_1) + (n-1)\sin^4(\gamma_1)]$. Therefore, nonzero elements $c_{(i,j),(i',j')}(\gamma)$ of $C(\gamma)$ include

$$\cos^2(\gamma_1) I_{TP}^{-1/2}(\gamma)/\phi_*, \quad \text{for all } i = i', j \neq j' \quad \text{and} \quad \sin^2(\gamma_1) I_{TP}^{-1/2}(\gamma)/\phi_* \quad \text{for all } i \neq i', j = j'.$$

Assumption (C.1) is trivial in this example and (C.5)-(C.8) can be validated by using the discussions in the previous example. So we will focus on checking (C.2), (C.3) and (C.4).

In order to calculate $\mu_{\max}[C(\gamma)]$, we need to solve the characteristic equation $\det[xI_{mn} - C(\gamma)] = 0$. We obtain

$$\det[xI_{mn} - C(\gamma)] = [\det(A_1 - B_1)]^{n-1} \det[A_1 + (n-1)B_1],$$

where $B_1 = -I_{TP}^{-1/2}(\gamma)\phi_*^{-1}\sin^2(\gamma_1) I_m$ and

$$A_1 = [x + I_{TP}^{-1/2}(\gamma)\phi_*^{-1}\cos^2(\gamma_1)]I_m - I_{TP}^{-1/2}(\gamma)\phi_*^{-1}\cos^2(\gamma_1)\mathbf{1}_m\mathbf{1}_m^T.$$

We can solve $\det(A_1 - B_1) = 0$ and $\det[A_1 + (n-1)B_1] = 0$. Finally, we can obtain four characteristic roots as follows: $-I_{TP}^{-1/2}(\gamma)\phi_*^{-1}$, $I_{TP}^{-1/2}(\gamma)\phi_*^{-1}[(m-1)\cos^2(\gamma_1) - \sin^2(\gamma_1)]$, $I_{TP}^{-1/2}(\gamma)\phi_*^{-1}[(n-1)\sin^2(\gamma_1) - \cos^2(\gamma_1)]$ and $I_{TP}^{-1/2}(\gamma)\phi_*^{-1}[(n-1)\sin^2(\gamma_1) + (m-1)\cos^2(\gamma_1)]$. Obviously, we have

$$\mu_{\max}[C(\gamma)] = \frac{[(n-1)\sin^2(\gamma_1) + (m-1)\cos^2(\gamma_1)]}{\sqrt{2nm[(m-1)\cos^4(\gamma_1) + (n-1)\sin^4(\gamma_1)]}} \leq \frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{m-1}}.$$

A sufficient condition of (C.2) is that both m and n tend to infinity.

To check (C.3), we have

$$\sum_{K \neq K'} [\partial_{\gamma_1} c_{K,K'}(\gamma)]^2 = \frac{(n-1)(m-1)\sin^2(\gamma_1)\cos^2(\gamma_1)}{2[(m-1)\cos^4(\gamma_1) + (n-1)\sin^4(\gamma_1)]^2} \leq \frac{\sqrt{(n-1)(m-1)}}{[(m-1)\cos^4(\gamma_1) + (n-1)\sin^4(\gamma_1)]}.$$

The right hand side is bounded by $(m+n-2)/\sqrt{(n-1)(m-1)}$. So a sufficient condition for (C.3) is that the limit of m/n must be a positive constant c_0, as both m and n tend to infinity. That is, m and n increase at the same order. However, if m/n (or n/m) converges to zero, the left hand side of the above inequality may tend to infinity.

Some simple calculations lead to

$$\rho_N(\gamma,\gamma') = \frac{(m-1)\cos^2(\gamma_1)\cos^2(\gamma_1') + (n-1)\sin^2(\gamma_1)\sin^2(\gamma_1')}{\sqrt{[(m-1)\cos^4(\gamma_1) + (n-1)\sin^4(\gamma_1)][(m-1)\cos^4(\gamma_1') + (n-1)\sin^4(\gamma_1')]}}.$$

As $m/n \to 0$ and $N \to \infty$, we see that the limit of $\rho_N(\gamma,\gamma')$ is one; thus, $T_P(\gamma)$'s for all $\gamma \in \Gamma$ are identical. In this case, we can arbitrarily set $\gamma = 0$ and make inference by using $T_P(0)$. If $m/n \to c_0$, we see that

$$\rho(\gamma,\gamma') = \frac{c_0\cos^2(\gamma_1)\cos^2(\gamma_1') + \sin^2(\gamma_1)\sin^2(\gamma_1')}{\sqrt{[c_0\cos^4(\gamma_1) + \sin^4(\gamma_1)][c_0\cos^4(\gamma_1') + \sin^4(\gamma_1')]}}.$$

2 Derivation of Score Test Statistic

Following Zhu and Zhang (2005), we can write the log-likelihood function as

$$\mathcal{L}_n(\sigma_T|\beta,\gamma,\Phi) = \log\left\{\int \prod_{i=1}^n \prod_{j=1}^{m_i} p(y_{i,j}|\psi_{i,j}(\mathbf{x}_{i,j}^T\beta, f_{i,j}(\mathbf{z}_{i,j},\gamma_{(1)})^T\mathbf{u}_i\sigma_T^{1/2}),\Phi)dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\},$$

where $F(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)$ is the distribution function of $(\mathbf{u}_1,\cdots,\mathbf{u}_n)$. Let $\eta_{i,j} = f_{i,j}(\mathbf{z}_{i,j},\gamma_{(1)})^T\mathbf{u}_i$, $t_{i,j} = \sigma_T^{1/2}\eta_{i,j}$ and $H_{i,j}(\eta_{i,j}\sigma_T^{1/2}) = p(y_{i,j}|\psi_{i,j}(\mathbf{x}_{i,j}^T\beta, f_{i,j}(\mathbf{z}_{i,j},\gamma_{(1)})^T\mathbf{u}_i\sigma_T^{1/2}),\Phi)$. So, we can have

$$\mathcal{L}_n(\sigma_T|\beta,\gamma,\Phi) = \log\left\{\int \prod_{i=1}^n \prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\}.$$

Thus, we have

$$\lim_{\sigma_T\to 0^+}\frac{\partial}{\partial\sigma_T}\mathcal{L}_n(\sigma_T|\beta,\gamma,\Phi) = \lim_{\sigma_T\to 0^+}\frac{\partial}{\partial\sigma_T}\log\left\{\int \prod_{i=1}^n \prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\}$$

$$= \lim_{\sigma_T\to 0^+}\left(\left\{\frac{\partial\sqrt{\sigma_T}}{\partial\sigma_T}\frac{\partial}{\partial\sqrt{\sigma_T}}\left[\int \prod_{i=1}^n \prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right]\right\}\times\right.$$
$$\left.\left[\int \prod_{i=1}^n \prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right]^{-1}\right).$$

The second term on the right hand side of the above equation converges to a constant as $\sigma_T \to 0^+$, so we omit them temporary and focus the first term on the right side. In the following, we assume that we can change the order of integration and derivative. Now, because $\partial\sqrt{\sigma_T}/\partial\sigma_T = 0.5/\sqrt{\sigma_T}$, it follows from the L'Hopital's rule (Buck and Buck, 1965; p.92) that the first term is given by

$$
\begin{aligned}
&0.5 \lim_{\sigma_T \to 0^+} \frac{\frac{\partial}{\partial\sqrt{\sigma_T}}\left\{\int \prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\}}{\sqrt{\sigma_T}} \\
&= 0.5 \lim_{\sigma_T \to 0^+} \frac{\frac{\partial^2}{\partial^2\sqrt{\sigma_T}}\left\{\int \prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\}}{\frac{\partial\sqrt{\sigma_T}}{\partial\sqrt{\sigma_T}}} \text{L'Hopital's rule} \\
&= 0.5 \lim_{\sigma_T \to 0^+} \left\{\int \frac{\partial^2}{\partial^2\sqrt{\sigma_T}} \prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma)\right\}.
\end{aligned}
$$

Now, we only need to consider the integrand

$$
A = \frac{\partial^2}{\partial^2\sqrt{\sigma_T}} \prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2}).
$$

We can show that

$$
\begin{aligned}
A &= \frac{\partial}{\partial\sqrt{\sigma_T}}[\frac{\partial}{\partial\sqrt{\sigma_T}} \log\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\}\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\}] = \{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\} \times \\
&\left[\frac{\partial^2}{\partial^2\sqrt{\sigma_T}} \log\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\} + \left\{\frac{\partial}{\partial\sqrt{\sigma_T}} \log\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\}\right\}^2\right].
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\left\{\frac{\partial}{\partial\sqrt{\sigma_T}} \log\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\}\right\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\partial}{\partial\sqrt{\sigma_T}} \log\{H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\partial t_{i,j}}{\partial\sqrt{\sigma_T}} \frac{\partial}{\partial t_{i,j}} \log\{H_{i,j}(t_{i,j})\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \eta_{i,j} \frac{\partial}{\partial t_{i,j}} \log\{H_{i,j}(t_{i,j})\}
\end{aligned}
$$

and

$$
\begin{aligned}
&\left\{\frac{\partial^2}{\partial^2\sqrt{\sigma_T}} \log\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} H_{i,j}(\eta_{i,j}\sigma_T^{1/2})\}\right\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \eta_{i,j} \frac{\partial}{\partial\sqrt{\sigma_T}} \left\{\frac{\partial}{\partial t_{i,j}} \log\{H_{i,j}(t_{i,j})\}\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \eta_{i,j} \frac{\partial t_{i,j}}{\partial\sqrt{\sigma_T}} \frac{\partial^2}{\partial^2 t_{i,j}} \log\{H_{i,j}(t_{i,j})\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} {\eta_{i,j}}^2 \frac{\partial^2 \log\{H_{i,j}(t_{i,j})\}}{\partial^2 t_{i,j}}.
\end{aligned}
$$

Combining the results above, we show that the first-order right derivative of $\mathcal{L}_n(\sigma_T|\beta,\gamma,\Phi)$ at $\sigma_T = 0$ is given by

$$\begin{aligned} T_S(\gamma|\beta,\Phi) &= 2\lim_{\sigma_T\to 0^+}\left\{\frac{\partial \mathcal{L}_n}{\partial \sigma_T}(\sigma_T|\beta,\gamma,\Phi)\right\} = \int\left[\left\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\partial \log p(y_{i,j}|\psi_{i,j}(\mathbf{x}_{i,j}^T\beta;t_{i,j}))}{\partial t_{i,j}}(0)\eta_{i,j}\right\}^2\right. \\ &+ \left.\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\eta_{i,j}]^2\left\{\frac{\partial^2 \log p(y_{i,j}|\psi_{i,j}(\mathbf{x}_{i,j}^T\beta;t_{i,j}))}{\partial t_{i,j}^2}(0)\right\}\right] dF(\mathbf{u}_1,\cdots,\mathbf{u}_n|\gamma). \end{aligned}$$

3 Asymptotic Local Power

In this section, we mainly study asymptotic local power of the score test statistics under certain conditions. We assume that the true value $\xi_* = (\beta_*, \phi_*)$ of $\xi = (\beta, \phi)$ is an interior point of the parametric space Ξ. Let $\eta = (\sigma_T, \xi)$. We denote $L_N(\sigma_T, \xi|\gamma)$ as the corresponding log-likelihood function. Because $L_N(0, \xi|\gamma)$ is independent of γ, we define $\hat{\eta}$ to be the maximum likelihood estimate of η under H_0, which does not depend on γ under H_0. Furthermore, we now define the $\tilde{\eta}(\gamma)$ to be the maximum likelihood estimate of η for any given $\gamma \in \Pi$ under H_1. In addition, we define

$$\frac{\partial L_N(\eta|\gamma)}{\partial \eta} = S_{\eta,N}(\gamma|\eta) = \begin{pmatrix} S_{\sigma,N}(\gamma|\eta) \\ S_{\xi,N}(\gamma|\eta)\end{pmatrix}, \quad S_{\eta,N}(\gamma) = S_{\eta,N}(\gamma|\eta)|_{\eta=\eta_*} = \begin{pmatrix} S_{\sigma,N}(\gamma) \\ S_{\xi,N}(\gamma)\end{pmatrix},$$

and $J_{\eta,N}(\gamma) = E\{\partial^2 L_N(\eta|\gamma)/\partial\eta\partial\eta^T\}|_{\eta=\eta_*}$.

We need the following higher level conditions to establish the asymptotic properties.

Assumption S.1. $\sup_{\gamma\in\Gamma} ||\tilde{\eta}(\gamma) - \eta_*|| \to 0$ and $||\hat{\eta} - \eta_*|| \to 0$ in probability.

Assumption S.2. Assume that

$$O_p(1) = L_N(\eta|\gamma) = L_N(\eta_*|\gamma) + \sqrt{N}(\eta - \eta_*)^T S_{\eta,N}(\gamma) - \frac{N}{2}(\eta - \eta_*)^T J_{\eta,N}(\gamma)(\eta - \eta_*) + o_p(1), \quad (5)$$

holds uniformly for all $\sqrt{N}||\eta - \eta_*|| \le C_0$, where C_0 is any positive scalar. In addition, $S_{\eta,N}(\gamma)$ and $J_{\eta,N}(\gamma)$ are $(q_1 + 2) \times 1$ random vector and $(q_1 + 2) \times (q_1 + 2)$ symmetric random matrix, respectively. Moreover, $\sup_{\gamma\in\Gamma} ||S_{\eta,N}(\eta|\gamma)|| = O_p(1)$ and

$$C_2 \ge \sup_{\gamma\in\Gamma} \mu_{\min}[J_{\eta,N}(\gamma)] \ge \inf_{\gamma\in\Gamma} \mu_{\min}[J_{\eta,N}(\gamma)] \ge 4C_l^2 > 0$$

holds almost surely for some fixed C_l and C_2, where $\mu_{\max}$ and $\mu_{\min}$ represents the minimum and maximum eigenvalues of a matrix.

Assumption S.3. $(S_{\eta,N}(\cdot), J_{\eta,N}(\cdot)) \Rightarrow (S_\eta(\cdot), J_\eta(\cdot))$, where $\{(S_\eta(\gamma), J_\eta(\gamma)) : \gamma \in \Gamma\}$ has bounded continuous sample paths with probability one, where $\Rightarrow$ denotes weak convergence of a stochastic process under the uniform metric. Moreover, the $(q_1+2) \times (q_1+2)$ matrix $J_\eta(\cdot)$ is symmetric and $\infty > \sup_{\gamma\in\Gamma} \lambda_{\max}[J_\eta(\gamma)] \geq \inf_{\gamma\in\Gamma} \lambda_{\min}[J_\eta(\gamma)] > 0$ holds almost surely.

COMMENTS: Assumption S.1 ensures that both $\tilde{\eta}(\gamma)$ and $\hat{\eta}$ are consistent for $\eta_* = (0, \xi_*)$ uniformly over $\gamma \in \Gamma$. In general, this assumption can be proved by using Assumption $\mathbf{1}^*$ of Andrews (2001). Assumptions S.1-S.3 can be validated by using Assumptions $\mathbf{1}^*$ and $\mathbf{2}^{2*}$ in Andrews (2001). See Zhu and Zhang (2005) for how to validate Assumption S.3.

We need the following notation:

$$S_\eta(\gamma) = \begin{pmatrix} S_\sigma(\gamma) \\ S_\xi(\gamma) \end{pmatrix}, \quad J_\eta(\gamma) = \begin{bmatrix} J_{\sigma\sigma}(\gamma) & J_{\sigma\xi}(\gamma) \\ J_{\xi\sigma}(\gamma) & J_{\xi\xi}(\gamma) \end{bmatrix},$$

$$Z_{\eta,N}(\gamma) = J_{\eta,N}(\gamma)^{-1} S_{\eta,N}(\gamma) = \begin{pmatrix} Z_{\sigma,N}(\gamma) \\ Z_{\xi,N}(\gamma) \end{pmatrix}, \quad Z_\eta(\gamma) = J_\eta(\gamma)^{-1} S_\eta(\gamma) = \begin{pmatrix} Z_\sigma(\gamma) \\ Z_\xi(\gamma) \end{pmatrix}.$$

Here, $L_N(\eta_*|\gamma)$, $L_N(\hat{\eta}|\gamma)$, $S_\xi(\gamma)$ and $J_{\xi\xi}(\gamma)$ do not depend on γ, so we drop γ from them. Let $\mathbf{e}_1 = (1, 0, \cdots, 0)^T \in R^{q_1+2}$.

THEOREM S.1. *Suppose Assumptions S.1-S.3 hold. Then, under the null hypothesis, we have the following results:*

(a) $\sup_{\gamma\in\Gamma} ||\tilde{\eta}(\gamma) - \eta_*|| = O_p(N^{-1/2})$.

(b)

$$L_N(\tilde{\eta}(\cdot)|\cdot) - L_N(\eta_*|\cdot) \Rightarrow 0.5\{S_\eta(\cdot)^T J_\eta(\cdot)^{-1} S_\eta(\cdot) - \inf_\lambda [\lambda - J_\eta(\cdot)^{-1} S_\eta(\cdot)]^T J_\eta(\cdot)[\lambda - J_\eta(\cdot)^{-1} S_\eta(\cdot)]\},$$

and

$$\sup_{\gamma\in\Gamma}[L_N(\tilde{\eta}(\gamma)|\gamma) - L_N(\eta_*|\gamma)] \Rightarrow 0.5 \sup_{\gamma\in\Gamma}\{S_\eta(\gamma)^T J_\eta(\gamma) S_\eta(\gamma) - \inf_\lambda [\lambda - J_\eta(\gamma)^{-1} S_\eta(\gamma)]^T J_\eta(\gamma)[\lambda - J_\eta(\gamma)^{-1} S_\eta(\gamma)]\}.$$

(c) $\sqrt{N}\tilde{\sigma}(\cdot) \Rightarrow \max\{Z_\sigma(\cdot), 0\}$.

(d) $\sup_{\gamma\in\Gamma}[L_N(\tilde{\eta}(\gamma)|\gamma)] - L_N(\eta_*) \Rightarrow 0.5 \sup_{\gamma\in\Gamma} \max^2\{Z_\sigma(\gamma), 0\}/(\mathbf{e}_1^T J_\eta(\gamma)^{-1}\mathbf{e}_1) + 0.5 S_\xi^T J_{\xi\xi}^{-1} S_\xi$.

(e) $L_N(\hat{\eta}) - L_N(\eta_*) \Rightarrow 0.5 S_\xi^T J_{\xi\xi}^{-1} S_\xi$.

(f) $\sup_{\gamma\in\Gamma}[L_N(\tilde{\eta}(\gamma)|\gamma)] - L_N(\hat{\eta}) \Rightarrow 0.5 \sup_{\gamma\in\Gamma} \max^2\{Z_\sigma(\gamma), 0\}/(\mathbf{e}_1^T J_\eta(\gamma)^{-1}\mathbf{e}_1)$.

Proof of Theorem S.1. Part (a) can be easily seen from the quadratic expansion in Assumption S.2 and consistency assumption in Assumption S.1. For given γ, it follows from Assumption S.2 that $L_N(\eta|\gamma) - L_N(\eta_*|\gamma)$ can be approximated by

$$0.5\{S_{\eta,N}(\gamma)^T J_{\eta,N}^{-1}(\gamma) S_{\eta,N}(\gamma) - [\sqrt{N}(\eta - \eta_*) - Z_{\eta,N}(\gamma)]^T J_{\eta,N}(\gamma)[\sqrt{N}(\eta - \eta_*) - Z_{\eta,N}(\gamma)]\} + o_p(1).$$

Furthermore, $L_N(\tilde{\eta}(\gamma)|\gamma) - L_N(\eta_*|\gamma)$ is close to

$$0.5\{Z_{\eta,N}(\gamma)^T J_{\eta,N}(\gamma) Z_{\eta,N}(\gamma) - \inf_{\lambda}[\lambda - Z_{\eta,N}(\gamma)]^T J_{\eta,N}(\gamma)[\lambda - Z_{\eta,N}(\gamma)]\}.$$

By using Assumption S.3 and continuous mapping theorem, we can prove part (b).

By noting that the parametric space of σ_T is $[0, \infty)$ and η_* is an interior point, we use some matrix calculation to obtain

$$S_{\eta,N}^T(\gamma) J_{\eta,N}(\gamma) S_{\eta,N}(\gamma) = Z_{\eta,N}^T(\gamma) J_{\eta,N}(\gamma)^{-1} Z_{\eta,N}(\gamma) = Z_{\sigma,N}(\gamma)^2/(\mathbf{e}_1^T J_{\eta,N}(\gamma)^{-1}\mathbf{e}_1) + S_\xi^T J_{\xi\xi}^{-1} S_\xi + o_p(1)$$

and $\inf_\lambda[\lambda - Z_{\eta,N}(\gamma)]^T J_{\eta,N}(\gamma)[\lambda - Z_{\eta,N}(\gamma)] = \inf_\sigma[\sigma - Z_{\sigma,N}(\gamma)]^2/[\mathbf{e}_1^T J_{\eta,N}(\gamma)^{-1}\mathbf{e}_1]^{-1}$. With these preparations, we can prove parts (c) and (d). Similarly, we can easily prove parts (e) and (f).

THEOREM S.2. *Suppose Assumptions S.1-S.3 hold. If $T_S(\gamma)$ is evaluated at the $\hat{\eta}$, the likelihood ratio statistic and the score test statistic S_S for testing H_0 against H_1 have the same asymptotic null distribution.*

Proof of Theorem S.2. Recall that

$$S_S = \sup_{\gamma\in\Gamma} \max{}^2 \left\{ \frac{\hat{T}_S(\gamma)}{\sqrt{I_{ES}(\gamma)}}, 0 \right\}.$$

At the $\hat{\eta}$, we see that

$$\mathbf{e}_1^T J_{\eta,N}^{-1}(\gamma) \begin{pmatrix} 2\hat{T}_S(\gamma) \\ 0 \end{pmatrix} = \mathbf{e}_1^T J_{\eta,N}^{-1}(\gamma) S_{\eta,N}(\gamma|\hat{\eta}) = \mathbf{e}_1^T J_{\eta,N}^{-1}(\gamma) S_{\eta,N}(\gamma) + o_p(1).$$

Thus, the left term of the equation above is equal to $2\hat{T}_S(\gamma)\mathbf{e}_1^T J_{\eta,N}^{-1}\mathbf{e}_1$, so we have

$$2\hat{T}_S(\gamma) = \frac{\mathbf{e}_1^T Z_\eta(\gamma)}{\mathbf{e}_1^T J_{\eta,N}(\gamma)^{-1}\mathbf{e}_1} + o_p(1) \quad \text{and} \quad I_{ES}(\gamma) = \frac{1}{4\mathbf{e}_1^T J_{\eta,N}(\gamma)^{-1}\mathbf{e}_1} + o(1).$$

Therefore, we can show that $S_S = \sup_{\gamma\in\Gamma} \max^2 \left\{ \mathbf{e}_1^T Z_\eta(\gamma)/\sqrt{\mathbf{e}_1^T J_\eta(\gamma)^{-1}\mathbf{e}_1}, 0 \right\} + o_p(1)$. Now, we finish the proof of Theorem S.2.

Theorem S.2 holds for most mixed models considered in the literature, because Assumptions S.1-S.3 are almost necessary for establishing asymptotic behavior of estimates and test statistics in general mixed model (Jiang, 1996). For instance, for the clustered mixed models, if the number of observations in each unit is finite, then Assumptions S.1-S.3 are obviously true. Even further, for the hierarchical and crossed model, we can show that Assumptions S.1-S.3 hold under some mild conditions. In general, we have the following Corollary.

COROLLARY 1. *Under the assumption of Lin's (1997) Proposition 1, Assumptions S.1-S.3 hold.*

Furthermore, we consider the asymptotic local power of the score testing statistics, because power considerations are important in study design and significance testing (Cox and Hinkley, 1975, p.103). It becomes apparent from the quadratic approximations that the distribution of $Z_\eta(\gamma)$ plays a critical role in determining the asymptotic local power of S_S. Therefore, it is worthwhile to explore its property under a sequence of local alternatives $(\eta_N(\mathbf{h}), \gamma^0)$ such that $\eta_N(\mathbf{h}) = \eta_* + N^{-1/2}\mathbf{h}$ and $\gamma = \gamma^0$, a true value of γ when $\mathbf{h} = (h_1, \cdots, h_{q_1+2})^T \neq \mathbf{0}$ and $h_1 > 0$.

THEOREM S.3. *Under Assumptions S.1-S.3 and the alternatives $(\eta_N(\mathbf{h}), \gamma^0)$, $Z_{\eta,N}(\gamma)$ converges to $N(J_\eta(\gamma)^{-1}J_\eta(\gamma,\gamma^0)\mathbf{h}, J_\eta(\gamma)^{-1})$ in distribution for any given $\gamma \in \Gamma$, where*

$$J_\eta(\gamma, \gamma^0) = \lim_{N\to\infty} Cov[S_{\eta,N}(\gamma), S_{\eta,N}(\gamma^0)],$$

and $Z_{\eta,N}(\cdot)$ converges weakly to a Gaussian process with mean $J_\eta(\cdot)^{-1}J_\eta(\cdot,\gamma^0)\mathbf{h}$ and covariance components $J_\eta(\gamma_1)^{-1}J_\eta(\gamma_1,\gamma_2)J_\eta(\gamma_2)^{-1}$.

Proof of Theorem S.3. First, we have

$$L_N(\eta_* + N^{-1/2}\mathbf{h}|\gamma^0) - L_N(\eta_*) = \mathbf{h}^T S_{\eta,N}(\gamma^0) - \frac{1}{2}\mathbf{h}^T J_{\eta,N}(\gamma^0)\mathbf{h} + o_p(1).$$

Therefore, under H_0, we can show that $(Z_{\eta,N}(\gamma), L_N(\eta_* + N^{-1/2}\mathbf{h}|\gamma^0) - L_N(\eta_*))$ converges to normal distribution with mean $(\mathbf{0}^T, -0.5\sigma_{22})^T$ and covariance matrix

$$\begin{pmatrix} J_\eta^{-1}(\gamma) & J_\eta(\gamma)^{-1}J_\eta(\gamma,\gamma^0)\mathbf{h} \\ \mathbf{h}^T J_\eta(\gamma,\gamma^0)J_\eta(\gamma)^{-1} & \sigma_{22} \end{pmatrix}$$

where $\sigma_{22} = \mathbf{h}^T J_\eta(\gamma^0)\mathbf{h}$. By using LeCam third lemma (van der Vaart, 1998, p.90), we can show that $Z_{\eta,N}(\gamma)$ converges to $N(J_\eta(\gamma)^{-1}J_\eta(\gamma,\gamma^0)\mathbf{h}, J_\eta(\gamma)^{-1})$ in distribution. We also infer that the local alternatives $(\eta_N(\mathbf{h}), \gamma^0)$ are contiguous to the null η_*.

We need to check finite convergence and asymptotic equicontinuity of $Z_{\eta,N}(\gamma)$ under the local alternatives $(\eta_N(\mathbf{h}), \gamma^0)$. The finite convergence can be directly verified by generalizing the asymptotic distribution of $Z_{\eta,N}(\gamma)$ at one point to any finite points. With contiguous local alternatives, we can use the same methods to establish the asymptotic equicontinuity of $Z_{\eta,N}(\gamma)$ as we did in Theorem 3 of Zhu and Zhang (2005). Combing these two, we can establish that $Z_{\eta,N}$ converges weakly to the desirable Gaussian process under the local alternatives $(\eta_N(\mathbf{h}), \gamma^0)$.

Based on Theorems S.1-S.3, we know that under the local alternatives alternatives $(\eta_N(\mathbf{h}), \gamma^0)$, $Z_{\sigma,N}(\cdot)$ converges weakly to a Gaussian process $GP(\cdot)$ with marginal mean $\mu_\sigma(\gamma) = \mathbf{e}_1^T J_\eta(\gamma)^{-1}J_\eta(\gamma,\gamma^0)\mathbf{h}$ and covariance component $C_\sigma(\gamma_1,\gamma_2) = \mathbf{e}_1^T J_\eta(\gamma_1)^{-1}J_\eta(\gamma_1,\gamma_2)J_\eta(\gamma_2)^{-1}\mathbf{e}_1$. We also define

$$X(\gamma) = \frac{GP(\gamma) - \mu_\sigma(\gamma)}{\sqrt{\mathbf{e}_1^T J_\eta(\gamma)^{-1}\mathbf{e}_1}} \quad \text{and} \quad \mu(\gamma) = \frac{\mu_\sigma(\gamma)}{\sqrt{\mathbf{e}_1^T J_\eta(\gamma)^{-1}\mathbf{e}_1}}.$$

Note that for each given γ, $X(\gamma)$ follows standard normal, but $X(\cdot)$ is a centered Gaussian process with correlation component

$$\frac{C_\sigma(\gamma_1, \gamma_2)}{\sqrt{\mathbf{e}_1^T J_\eta(\gamma_1)^{-1}\mathbf{e}_1}\sqrt{\mathbf{e}_1^T J_\eta(\gamma_2)^{-1}\mathbf{e}_1}}$$

Let $\chi_{X,\alpha}$ be the critical value of $\sup_{\gamma\in\Gamma}\max\{X(\gamma), 0\}$ at the level of α. Therefore, the power of S_S under under the local alternatives alternatives $(\eta_N(\mathbf{h}), \gamma^0)$ is close to

$$P\left(\sup_{\gamma\in\Gamma}\max{}^2\{X(\gamma) + \mu(\gamma), 0\} > \chi^2_{X,\alpha}\right).$$

Inspired by Theorem 2 of Fan (1996), we use a grid of Γ, $\{\gamma_i : i = 1, \cdots, m\}$ to approximate Γ. Thus, we know that

$$\begin{aligned} & P\left(\sup_{\gamma\in\Gamma}\max{}^2\{X(\gamma) + \mu(\gamma), 0\} > \chi^2_{X,\alpha}\right) \geq P\left(\max_{1\leq i\leq m}\{X(\gamma_i) + \mu(\gamma_i), 0\} > \chi_{X,\alpha}\right) \\ = \; & 1 - P\left(\max_{1\leq i\leq m}\{X(\gamma_i) + \mu(\gamma_i), 0\} \leq \chi_{X,\alpha}\right) = 1 - P\left(\max_{1\leq i\leq m}\{X(\gamma_i) + \mu(\gamma_i)\} \leq \chi_{X,\alpha}\right). \end{aligned}$$

Based on the above result, we have the following result.

THEOREM S.4. *The score test statistic S_S has the power at the alternatives $(\eta_N(\mathbf{h}), \gamma^0)$ at least $1 - P\left(\cap_{i=1}^m A_i\right)$, where A_i denotes event $\{X(\gamma_i) \leq \chi_{X,\alpha} - \mu(\gamma_i)\}$.*

For given grid, we can numerically calculate the probability of event $\cap_{i=1}^m A_i$. If the grid includes γ^0, we have $\mu(\gamma^0) = h_1/\sqrt{\mathbf{e}_1^T J_\eta(\gamma^0)^{-1}\mathbf{e}_1}$. As $h_1 \to \infty$, we know that the probability of $\{X(\gamma^0) \leq \chi_{X,\alpha} - \mu(\gamma^0)\}$ converges to zero. So, the power of the score test statistic is close to one. We have shown how to establish the asymptotic local power for S_S. In fact, we can similarly study the asymptotic local power for S_O and S_P.

4 Comparisons between the Methods

In the literature, many authors have proposed score test statistics to test homogeneity in the framework of the GLMMs; see Liang (1987), Commenges and Jacqmin-Gadda (1997), Lin (1997), Hall and Praestgaard (2001), and Verbeke and Molenberghs (2003), for example. Specifically, we will highlight the advantages of our method over Lin's (1997) score test statistic, referred to as LS, and Hall and Praestgaard's (2001) projected score test, referred to as PLS.

Our score test statistics cover more models than do LS and PLS. Lin (1997) and Hall and Praestgaard (2001) are based on a conditional generalized linear mixed model (GLMM) with

$$E(y_{i,j}|\mathbf{b}) = \mu_{i,j}(\mathbf{b}) \quad \text{and} \quad g(\mu_{i,j}(b)) = \mathbf{x}_{i,j}^T\beta + \mathbf{z}_{i,j}^T\mathbf{b},$$

where $\mathbf{b}$ denotes all possible random effects. In contrast, in Section 1 of Zhu and Zhang (2005), we assume that the probability density of $y_{i,j}$ depends on $\psi_{i,j}(\mathbf{b}_i; \beta, \gamma_{(1)})$ that $\psi_{i,j}(\mathbf{b}_i; \beta, \gamma_{(1)})$ may not be a function of the mean of $y_{i,j}$ conditional on the random effects. Moreover, $f(\mathbf{z}_{i,j}, \gamma_{(1)})$ may depend on an unknown parameter in Sections 1.1 and 1.2. In general, the existing methods are not applicable in some useful models including random coefficient models, genetic models and factor analysis models, when a hypothesis testing involves a nuisance parameter which is defined only under the alternative hypothesis (Davis, 1977; Andrews, 2001). In some application where there is no nuisance $\gamma_{(1)}$, we can show that our score test S_S reduces to the PLS. In this sense, S_S is a generalization of the existing test.

Both our score test statistics and PLS impose the positive definitive constraint on the variance components of Σ. We should note, however, that LS is not necessarily locally asymptotically most stringent as defined in equation (1.3) of Bhat & Nagnur (1963) (even under the GLMM), because it does not account for the fact that Σ, the covariance matrix of the random effects, has to be a positive semidefinite matrix. Lin (1997) suggested the asymptotical stringency of LS, but in remarks following their Theorem 2, Hall and Praestgaard (2001) argued against Lin's (1997) suggestion. Through simulations, we have shown that ignoring the constraint of Σ can lead to the loss of power of detecting heterogeneity (see Tables 1 and 2 in Zhu and Zhang, 2005). Similar findings have also been reported in Table 3 of Hall and Praestgaard (2001). Previous result revealed that the likelihood ratio (or score) statistic under the constrained alternative is uniformly powerful than that for the unconstrained case (Tsai, 1992).

Whether or not we need to explicitly impose the constraint on Σ (or specify the alternative model) for the score test is an important issue, and it is equivalent to choosing a one-sided or two-sided test for testing the variance component. Recently, Verbeke and Molenberghs (2003) suggest that choice between a two-sided/one-sided score test should be tightly linked to a unconstrained/constrained alternative space, and developing the one-sided score test is statistically sound and practically meaningful. Interestingly, Lin (1997) also suggested to use a one-sided score test to test for individual variance component; see the second paragraph on page 316 and the first paragraph on page 317 in Lin (1997). In contrast, for a general Σ, Lin (1997) considered an unconstrained parametrization, which may lead to a matrix that does not satisfy the covariance constraint. In general, the choice between one-sided and two-sided tests should also depend on the scientific question, the models and the interpretation of the parameters in those models.

Some of previous results about the PLS also appear in doubt. In particular, the PLS may not follow a mixture of χ^2 distributions as claimed in Hall and Praestgarrd (2001). In fact, the tangent cone given in Sections 3.2 and 3.3 of Hall and Praestgarrd (2001) are not correct. Even for a random effect model with unstructured Σ as discussed in Example 1.4, we can show that the tangent cone is in fact the convex cone given by $\{\Sigma : \Sigma \geq 0\}$, where $\Sigma \geq 0$ represents a positive semidefinite matrix. For simplicity, we consider a 2×2 Σ with three unknown components $(\sigma_{11}, \sigma_{22}, \sigma_{12})$ and the parametric space is given by

$$\Theta = \{(\sigma_{11}, \sigma_{22}, \sigma_{12}) : \sigma_{11} \geq 0, \sigma_{22} \geq 0, \sigma_{11}\sigma_{22} - \sigma_{12}^2 \geq 0\}.$$

The parameter under the null hypothesis is $\theta_0 = (0, 0, 0)$. Thus, $t(\Theta - \theta_0)$ converges to Θ as $t \to \infty$. It follows from a result due to Rockafeller and Wets (1998, page 198) that the tangent cone is Θ itself instead of $\{\theta : \sigma_{11} \geq 0, \sigma_{22} \geq 0\}$ as given in Section 3.2 of Hall and Praestgarrd (2001). Similar mistake was made in their Section 3.3. It is not even clear what happens if the Σ has a specific structure in most of applications. In contrast, our theoretical results fill up such gap, and our resampling method provides an attractive approach to conduct the homogeneity test.

Following Commenges and Jacqmin-Gadda (1997), we make an additional improvement by decomposing the score test into two components: S_O and S_P. We have shown that S_O mainly tests the overdispersion and S_P detects the homogeneity. Computationally, it is much easier to calculate S_P than both S_S and S_O, because the formula for computing S_P remains valid if the unknown estimate has the property of asymptotic normality (see equation (12) in Zhu and Zhang (2005)). However, if we choose different estimation methods, computations of S_S, LS and PLS need to be modified as a result of changing variances.

We have established the asymptotic null distributions of the score test statistics under weaker conditions than the existing methods. For instance, our results cover the situations with both large n and large m_i and a general design matrix for mixed models, because we only need to validate Assumption (C.2); see Section 1 in Zhu and Zhang (2005) for details. However, to obtain the asymptotic distribution, Lin (1997) requires either large n and bounded m_is, or a M-dependent structure and special design pattern in the mixed effect model.

If the dimension of γ is large, computing the score test can be computational intensive. Zhu and Zhang (2005) suggest to compute the score test on a grid and to search for the maximum. In many cases that we have examined, a rough grid work very well. See the examples in Andrews (2001). Similar finding has been reported in the linkage analysis by Zheng and Chen (2005). However, we agree that in some cases, choosing a rough grid could lead to loss of power. Further research along this direction is warranted.

REFERENCES

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69** 685-734.

Buck, R. C. and Buck, E. F. (1965). *Advanced Calculus.* McGraw-Hill Inc, New York.

Commenges, D. and Jacqmin-Gadda, H. (1997). Generalized score test of homogeneity based on correlated random effects models. *J.R.Stat.Soc, Ser.B Stat.Methodol.* **59** 157-171.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247-254.

Fan, J.Q. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J.Amer.Statist.Assoc.* **91** 674-688.

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

Hansen, B. E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413–430.

Liang, K. Y. (1987). A locally most powerfully test for homogeneity with many strata. *Biometrika* **74** 259-264.

Lin, X. H. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84** 309-326.

Jacqmin-Gadda, H. and Commenges, D. (1995). Tests of homogeneity for generalized linear models. *J.Amer.Statist.Assoc.* **90** 1237-1246.

Jiang, J. M. (1996). REML estimation: asymptotic behavor and related topics. *Ann.Statist.* **24** 255-286.

Rockafellar, R. T. and Wets, R. (1998). *Variational Analysis.* Springer, New York.

Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *J.R.Stat.Soc, Ser.B Stat.Methodol.* 59 667-678.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J.Amer.Statist.Assoc.* **82** 605-610.

Sen, P. K. and Silvapulle, M. J. (2002). An appraisal of some aspects of statistical inference under inequality constraints. *J. Statist. Plann. Inference.* **107**, 3-43.

Shi, J. Q. and Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *J.R.Stat.Soc, Ser.B Stat.Methodol.* **62** 77-87.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effect model. *Biometrics* **50** 1171-1177.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46** 657-671.

Tsai, M. T. (1992). On the power superiority of likelihood ratio tests for restricted alternatives. *J.Multivariate.Anal.* **42** 102-109.

van der Vaart (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* 254-262.

Zheng, G. and Chen, Z.H. (2005). Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrics*, **61**, 254-258.

Zhu, H. T. and Zhang, H. P. (2005). Generalized score test of homogeneity for mixed effects models, Technical Report, Yale University School of Medicine.