Supplementary Material for "Goodness of Fit Tests for Linear Mixed Models"

Min Tang, Eric V. Slud and Ruth M. Pfeiffer

University of Maryland, National Cancer Institute

S1 Assumptions for Theorem 2

We first state and comment on the assumptions Miller (1977) made to ensure consistency and asymptotic normality of the MLE for parameters in equation (1) in Section 2 of the manuscript,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sum_{r=1}^{R} \mathbf{Z}_r \boldsymbol{\alpha}_r + \boldsymbol{\varepsilon}. \tag{S1.1}$$

Assumption **A.1** The partitioned matrix $[\mathbf{X} : \mathbf{Z}_r]$ has rank greater than p, $r = 1, \ldots, R$.
Assumption **A.2** The matrices $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_R$, defined as $\mathbf{G}_r = \mathbf{Z}_r\mathbf{Z}_r^T$, $r = 1, \ldots, R$, are linearly independent; that is, $\sum_{r=0}^{R} \tau_r \mathbf{G}_r = 0$ implies $\tau_r = 0$, $r = 0, 1, \ldots, R$.
Assumption **A.3** Each m_r, $r = 1, \ldots, R$, tend to infinity.
Assumption **A.4** Let $m_0 = N$. Then for each $s, t = 0, 1, \ldots, R$, either $\lim_{N\to\infty} m_s/m_t = \rho_{st}$ or $\lim_{N\to\infty} m_t/m_s = \rho_{ts}$ exists. If $\rho_{st} = 0$, then let $\rho_{ts} = \infty$ for notational convenience.

Without loss of generality, let $\mathbf{Z}_r$ be labeled so that for $s < t$, $\rho_{st} > 0$; i.e., the m_r are in decreasing order of magnitude. Generate a partition of the integers $0, 1, \ldots, R$, $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_c$, so that for indices r in the same set $\mathbf{S}_s$, the associated m_r's have the same order of magnitude as follows:
i) $r_0 = 0$; $\mathbf{S}_0 = \{0\}$; $r_1 = 1$.
ii) For $s = 1, 2, \ldots$, it is true that $r_s \in \mathbf{S}_s$. Then for $r = r_s + 1, r_s + 2, \ldots$, include r in $\mathbf{S}_s$ until $\rho_{r_s,r} = \infty$; call the first value of r where this occurs r_{s+1}; then $r_{s+1} \in \mathbf{S}_{s+1}$.
iii) Continue as in step ii) until R has been placed in a set. Call this set $\mathbf{S}_c$.
There are then $c + 1$ sets in partitions, $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_c$, and $\mathbf{S}_s = \{r_s, \ldots, r_{s+1} - 1\}$.

For each $r = 1, 2, \ldots, R$, $r \in \mathbf{S}_s$ for some $s = 1, 2, \ldots, c$. Define sequences K_r (depending on N) as follows:
$K_r = \text{rank}[\mathbf{Z}_{r_s} : \mathbf{Z}_{r_s+1} : \cdots : \mathbf{Z}_R] - \text{rank}[\mathbf{Z}_{r_s} : \cdots : \mathbf{Z}_{r-1} : \mathbf{Z}_{r+1} : \cdots : \mathbf{Z}_R]$, $r = 1, 2, \ldots, R$,
$K_0 = N - \text{rank}[\mathbf{Z}_1, \ldots, \mathbf{Z}_R]$.
(The K_r so defined are closely related to the degrees of freedom of sums of squares in the analysis of variance.)
Assumption **A.5** Each of the $\lim_{N\to} K_r/m_r$, $r = 1, \ldots, R$ exists and is positive.

Let $\mathbf{V}_0 = \sum_{r=1}^{R} \sigma_r^2 \mathbf{G}_r$ be the true covariance matrix.
Assumption **A.6** There exists a sequence K_{R+1} (depending on N) increasing to infinity such that the $p \times p$ matrix $\mathbf{C}_0$ defined by $\mathbf{C}_0 = \lim_{N\to\infty}[\mathbf{X}^T\mathbf{V}_0^{-1}\mathbf{X}]/K_{R+1}$ exists and is positive definite.

Define the $(R + 1) \times (R + 1)$ matrix $\mathbf{C}_1$ by

$$[\mathbf{C}_1]_{st} = \frac{1}{2} \lim_{N\to\infty} [\text{tr}\mathbf{V}_0^{-1}\mathbf{G}_s\mathbf{V}_0^{-1}\mathbf{G}_t]/K_s^{\frac{1}{2}}K_t^{\frac{1}{2}}, \quad s, t = 0, 1, \ldots, R.$$

Assumption **A.7** Each of the limits used in defining $[\mathbf{C}_1]_{st}$ exists, $s, t = 0, 1, \ldots, R$. The matrix $\mathbf{C}_1$ is positive definite.

Remark 1 *Assumption **A.1** requires that the fixed effects not be confounded with any of the random effects. **A.2** requires that the random effects not be confounded with each other. Assumptions **A.1**–**A.2** are sufficient to guarantee identifiability of the MLE $\hat{\boldsymbol{\theta}}$. Assumptions **A.3-A.7**, which correspond to Assumptions* 3.1 − 3.5 *in Miller (1977), are used to ensure the consistency of the MLE. Assumption **A.3** is natural and necessary for the consistency property of MLE estimators of both $\boldsymbol{\beta}$ and the variance components σ_ϵ^2 and $\sigma_r^2, r = 1, \ldots, R$, because the sample size used to estimate $\boldsymbol{\beta}$ and σ_ϵ^2 is N and the sample size used to estimate σ_r^2 is m_r. Assumptions **A.6**–**A.7** are used to establish the existence and positive definiteness of the limiting variance-covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$.* □

In addition to **A.1**–**A.7** taken from Miller (1977), we also require the following Assumptions to ensure the existence of components in the variance covariance matrix $\mathbf{\Sigma} = \mathbf{H} - \mathbf{\Lambda}\mathbf{J}_{\beta\beta}^{-1}\mathbf{\Lambda}^T$ for the test statistic.

Assumption **A.8** For any cell partition $E_1, \ldots, E_L$, $\Lambda_l = \lim_{N\to\infty} \sum_{k=1}^{N} I_{\{\mathbf{x}_k \in E_l\}} \mathbf{x}_k^T / N$ exists for each $l = 1, \ldots, L$.

Assumption **A.9** $\mathbf{H} = \lim_{N\to\infty} \mathbf{F}\mathbf{V}\mathbf{F}^T$ exists and is positive definite, with $\mathbf{F}$ given in equation (13) in the main paper.

Remark 2 *Assumption **A.8** ensures the existence of $\mathbf{\Lambda}$, that has elements defined in (14) in the main paper. Assumption **A.9** ensures the existence and positive definiteness of $\boldsymbol{H}$ in $\mathbf{\Sigma}$, where $\boldsymbol{H}$ denotes the limiting variance covariance matrix for $\{\boldsymbol{f} - \boldsymbol{e}(\boldsymbol{\beta}_0)\}/N$.* □

S2 Proof of Theorem 2

Let $\mathbf{J}$ be the limit of the sample information matrix per observation,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\beta\beta} & \mathbf{J}_{\beta\psi} \\ \mathbf{J}_{\beta\psi}^T & \mathbf{J}_{\psi\psi} \end{bmatrix}. \tag{S2.1}$$

Under model (S1.1), $\mathbf{J}_{\beta\psi} = 0$ (Wand 2007, equation (3)), and $\mathbf{J}$ and $\mathbf{J}^{-1}$ are block diagonal matrices. By Taylor series expansion of the score function $S(\hat{\boldsymbol{\theta}})$, we obtain

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \approx (-\frac{1}{N}\frac{\partial S(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}})^{-1}\frac{1}{\sqrt{N}}S(\boldsymbol{\theta}_0) \approx \mathbf{J}^{-1}\frac{1}{\sqrt{N}}S(\boldsymbol{\theta}_0), \tag{S2.2}$$

where $A \approx B$ means that $A - B \approx o_P(1)$ as $N \to \infty$.

As $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{V})$, the score function for $\boldsymbol{\beta}$, corresponding to the first p components of $S(\boldsymbol{\theta})$, is $S_\beta(\boldsymbol{\theta}) = \frac{\partial}{\partial \beta} l(\boldsymbol{\theta}) = \mathbf{X}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, where $l(\boldsymbol{\theta})$ is the log-likelihood function. By extracting the first p components of (S2.2), we have

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\approx \mathbf{J}_{\beta\beta}^{-1}\frac{1}{\sqrt{N}}S_\beta(\boldsymbol{\theta}_0) \\ &= \mathbf{J}_{\beta\beta}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0)/\sqrt{N}.\end{aligned}$$

Thus,

$$
\sqrt{N}\left(\begin{array}{c}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}/N\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{array}\right)\approx\left(\begin{array}{c}\frac{1}{\sqrt{N}}[I_{[\mathbf{x}_1\in E_1]}\cdots I_{[\mathbf{x}_N\in E_1]}]\\ \vdots\\ \frac{1}{\sqrt{N}}[I_{[\mathbf{x}_1\in E_L]}\cdots I_{[\mathbf{x}_N\in E_L]}]\\ \frac{1}{\sqrt{N}}\mathbf{J}_{\beta\beta}^{-1}\mathbf{X}^T\mathbf{V}^{-1}\end{array}\right)(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}_0)
$$

$$
=\mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}_0),
$$

which is a linear combination of Gaussian random variables, with $\mathbf{D}$ being a $(L+p)\times N$ matrix.

Therefore, as $N\to\infty$,

$$
\sqrt{N}\left(\begin{array}{c}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}/N\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{array}\right)\xrightarrow{D}N(\mathbf{0},\mathbf{A}_0),
$$

where

$$
\lim_{N\to\infty}\mathbf{D}\mathbf{V}\mathbf{D}^T=\left(\begin{array}{cc}\mathbf{H} & \boldsymbol{\Lambda}\mathbf{J}_{\beta\beta}^{-1}\\ \mathbf{J}_{\beta\beta}^{-1}\boldsymbol{\Lambda}^T & \mathbf{J}_{\beta\beta}^{-1}\end{array}\right)\equiv\mathbf{A}_0,
$$

and $\mathbf{H}=\lim_{N\to\infty}\mathbf{F}\mathbf{V}\mathbf{F}^T$ is a symmetric $L\times L$ matrix, with

$$
\mathbf{F}=\frac{1}{\sqrt{N}}\left(\begin{array}{c}I_{[\mathbf{x}_1\in E_1]}\cdots I_{[\mathbf{x}_N\in E_1]}\\ \vdots\\ I_{[\mathbf{x}_1\in E_L]}\cdots I_{[\mathbf{x}_N\in E_L]}\end{array}\right).
$$

and

$$
\boldsymbol{\Lambda}=\left(\begin{array}{c}\Lambda_1^T\\ \vdots\\ \Lambda_L^T\end{array}\right)_{L\times p}=\lim_{N\to\infty}\left(\begin{array}{c}\frac{1}{N}\sum_{k=1}^N I_{[\mathbf{x}_k\in E_1]}\mathbf{x}_k^T\\ \vdots\\ \frac{1}{N}\sum_{k=1}^N I_{[\mathbf{x}_k\in E_L]}\mathbf{x}_k^T\end{array}\right).
$$

The existence of $\mathbf{H}$ and $\boldsymbol{\Lambda}$ are ensured by Assumptions **A.8** and **A.9**. Also, by consistency of $\hat{\beta}$ and Taylor expansion,

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\hat{\boldsymbol{\beta}})\}=&\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}+\frac{1}{\sqrt{N}}\{\mathbf{e}(\boldsymbol{\beta}_0)-\mathbf{e}(\hat{\boldsymbol{\beta}})\}\\
\approx&\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}-\frac{1}{\sqrt{N}}\nabla\mathbf{e}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
\xrightarrow{P}&\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}-\boldsymbol{\Lambda}\sqrt{N}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0).
\end{aligned}
$$

Since $\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\hat{\boldsymbol{\beta}})\}$ is a linear combination of components of $\sqrt{N}\left(\begin{array}{c}\{\mathbf{f}-\mathbf{e}(\boldsymbol{\beta}_0)\}/N\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{array}\right)$, we get

$$
\frac{1}{\sqrt{N}}\{\mathbf{f}-\mathbf{e}(\hat{\boldsymbol{\beta}})\}\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{\Sigma}),
$$

with $\boldsymbol{\Sigma}=\mathbf{H}-\boldsymbol{\Lambda}\mathbf{J}_{\beta\beta}^{-1}\boldsymbol{\Lambda}^T$. Thus, $T=\{\mathbf{f}-\mathbf{e}(\hat{\boldsymbol{\beta}})\}^T\boldsymbol{\Sigma}^{-1}\{\mathbf{f}-\mathbf{e}(\hat{\boldsymbol{\beta}})\}/\sqrt{N}\xrightarrow{\mathcal{D}}\chi_k^2$, where $\boldsymbol{\Sigma}^{-1}$ denotes the generalized inverse of $\boldsymbol{\Sigma}$ and $k=\text{rank}(\boldsymbol{\Sigma})$.

Web Table 1

Table 1: Impact of cell partition on empirical power (Scenario II). $m = 500, E(N) = 1750, \beta_3 = .2, \sigma_a = 1, \sigma_\epsilon = .5, K = 1000$.

Partition	$\rho_{12} = 0$		$\rho_{12} = 0.3$	
	$L = 12$	$L = 42$	$L = 12$	$L = 42$
x_1	0.049	0.049	0.256	0.182
x_2	0.038	0.041	0.273	0.173
x_3	0.893	0.771	0.859	0.749
x_1, x_2	0.991	0.966	0.989	0.975
x_1, x_3	0.843	0.938	0.885	0.939
x_2, x_3	0.936	0.912	0.956	0.928