Online Appendix for Wang and Chen, "On testing an unspecified function through a linear mixed effects model with multiple variance components"

A1. Proof of Theorem 1

We prove the theorem under the model (5) with $L = 2$. We present a spectral decomposition of RSS_0, RSS_1, T_1 and T_2 through two equivalent models. For simplicity, we suppress γ in the notation of $\boldsymbol{V}_0(\gamma)$ and suppress γ and λ in the notation of $\boldsymbol{V}_1(\gamma, \lambda)$. We first consider the case where γ is known and then proceed to estimated γ. Multiplying both sides of (5) by $\boldsymbol{V}_0^{-1/2}$ to obtain an equivalent model under the null hypothesis as

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}_0\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_1, \quad \tilde{\boldsymbol{\varepsilon}}_1 \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}}^2\boldsymbol{I}_n), \tag{A.1}$$

where $\tilde{\boldsymbol{Y}} = \boldsymbol{V}_0^{-1/2}\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}}_0 = \boldsymbol{V}_0^{-1/2}\boldsymbol{X}_0$. Under the alternative hypothesis, the equivalent model is

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}_1\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_2, \quad \tilde{\boldsymbol{\varepsilon}}_2 \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}}^2\tilde{\boldsymbol{V}}), \tag{A.2}$$

where $\tilde{\boldsymbol{X}}_1 = \boldsymbol{V}_0^{-1/2}\boldsymbol{X}_1$, $\tilde{\boldsymbol{V}} = \boldsymbol{V}_0^{-1/2}\boldsymbol{V}_1\boldsymbol{V}_0^{-1/2} = \boldsymbol{I}_n + \lambda\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T$, and $\tilde{\boldsymbol{Z}} = \boldsymbol{V}_0^{-1/2}\boldsymbol{Z}_2$.

Denote $p = \text{rank}(\tilde{\boldsymbol{X}}_1)$. Applying results from Patterson and Thompson (1971) and Kuo (1999) to the equivalent models (A.1) and (A.2), there exists an $n \times (n-p)$ matrix $\boldsymbol{W}$ such that

$$\begin{aligned}
&\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}_{n-p}, \quad \boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{I}_n - \tilde{\boldsymbol{X}}_1(\tilde{\boldsymbol{X}}_1^T\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1^T,\\
&\boldsymbol{W}^T\tilde{\boldsymbol{V}}\boldsymbol{W} = \boldsymbol{W}^T(\boldsymbol{I} + \lambda\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T)\boldsymbol{W} = \text{diag}\{1 + \xi_s(\gamma, \lambda)\},
\end{aligned}$$

where

$$\begin{aligned}
\xi_s(\gamma, \lambda) &= \text{eigen}_s(\lambda\boldsymbol{W}^T\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T\boldsymbol{W})\\
&= \lambda\ \text{eigen}_s(\boldsymbol{W}^T\boldsymbol{V}_0^{-1/2}\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{V}_0^{-1/2}\boldsymbol{W})\\
&= \lambda\rho_s(\gamma),
\end{aligned}$$

$\text{eigen}_s(\boldsymbol{M})$ denotes the sth eigenvalue of the matrix $\boldsymbol{M}$, and recall $\rho_s(\gamma) = \text{eigen}_s(\boldsymbol{W}^T\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T\boldsymbol{W})$. To obtain the matrix $\boldsymbol{W}$ explicitly, let $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ be the singular value decomposition (SVD) of $\tilde{\boldsymbol{P}}_1 = \boldsymbol{I}_n - \tilde{\boldsymbol{X}}_1(\tilde{\boldsymbol{X}}_1^T\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1$, and let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}^{1/2}$. Then by the idempotent of $\tilde{\boldsymbol{P}}_1$, we have $\boldsymbol{A}\boldsymbol{A}^T = \tilde{\boldsymbol{P}}_1$ and $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}_{n-p}$. Let $\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{U}_1^T$ be the SVD of $\boldsymbol{A}^T\tilde{\boldsymbol{V}}\boldsymbol{A}$, then $\boldsymbol{W} = \boldsymbol{A}\boldsymbol{U}_1$. To verify this, note that $\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{A}\boldsymbol{U}_1\boldsymbol{U}_1^T\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T = \tilde{\boldsymbol{P}}_1$, $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{U}_1^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{U}_1^T\boldsymbol{I}_{n-p}\boldsymbol{U}_1 = \boldsymbol{I}_{n-p}$, and

$$\boldsymbol{W}^T\tilde{\boldsymbol{V}}\boldsymbol{W} = \boldsymbol{U}_1^T\boldsymbol{A}^T\tilde{\boldsymbol{V}}\boldsymbol{A}\boldsymbol{U}_1 = \boldsymbol{U}_1^T\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{U}_1^T\boldsymbol{U}_1 = \boldsymbol{D}_1 = \text{diag}\{1 + \xi_s(\gamma, \lambda)\}. \tag{A.3}$$

We now show the spectral decomposition of RSS_0 and RSS_1. Let $\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{V}) \triangleq \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}$. Then we have

$$\begin{aligned} RSS_1(\gamma, \lambda) &= \frac{1}{\sigma_{\boldsymbol{\varepsilon}}^2}\tilde{\boldsymbol{Y}}^T\{\boldsymbol{I} - \boldsymbol{H}(\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{V}})\}\tilde{\boldsymbol{V}}^{-1}\{\boldsymbol{I} - \boldsymbol{H}(\tilde{\boldsymbol{X}}_1, \tilde{V})\}\tilde{\boldsymbol{Y}} \\ &= \frac{1}{\sigma_{\boldsymbol{\varepsilon}}^2}\tilde{\boldsymbol{Y}}^T\boldsymbol{W}\text{diag}[\{1 + \lambda\rho_s(\gamma)\}^{-1}]\boldsymbol{W}^T\tilde{\boldsymbol{Y}}. \end{aligned}$$

The last step follows from Patterson and Thompson (1971), Kuo (1999) and Crainiceanu and Ruppert (2004). Note that

$$\boldsymbol{W}^T\tilde{\boldsymbol{Y}} \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}}^2\text{diag}\{1 + \lambda_0\rho_s(\gamma_0)\}),$$

where γ_0 and λ_0 are the true values of γ and λ. Since under the null specified in (4) we have $\lambda_0 = \sigma_b^2/\sigma_{\boldsymbol{\varepsilon}}^2 = 0$, it follows that $\boldsymbol{W}^T\tilde{\boldsymbol{Y}} \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}}^2\boldsymbol{I}_{n-p})$ under the null. One desirable feature of this decomposition is that the distribution of $\boldsymbol{W}^T\tilde{\boldsymbol{Y}}$ under the null does not depend on the true values of the nuisance parameters θ_0 or γ_0. Under the null, we have

$$RSS_1 =^d \sigma_{\boldsymbol{\varepsilon}}^2\sum_{s=1}^{n-p}\frac{1}{1 + \lambda\rho_s(\gamma)}u_s^2, \tag{A.4}$$

where u_s are independent and identically distributed $N(0, 1)$ random variables. Under the alternative, we have

$$RSS_1 =^d \sigma_{\boldsymbol{\varepsilon}}^2\sum_{s=1}^{n-p}\frac{1 + \lambda_0\rho_s(\gamma_0)}{1 + \lambda\rho_s(\gamma)}u_s^2, \quad u_s \sim^{i.i.d.} N(0, 1). \tag{A.5}$$

Now we assess RSS_0. Again using the equivalent models (A.1) and (A.2), we have

$$
\begin{aligned}
RSS_0 &= \frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}^T\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{Y}} \\
&= \frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})+\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{Y}} \\
&= \frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{Y}}+\frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T\{\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{Y}} \\
&= \frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T\boldsymbol{W}\boldsymbol{W}^T\tilde{\boldsymbol{Y}} \qquad (A.6)\\
&+ \frac{1}{\sigma_{\varepsilon}^2}\tilde{\boldsymbol{Y}}^T[\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}-\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})\}]\tilde{\boldsymbol{Y}}. \qquad (A.7)
\end{aligned}
$$

It is easy to see that the second term (A.7) is the difference of residual sum of squares of least squares fit from two nested models, one with $\tilde{\boldsymbol{X}}_0$ as fixed effects and the other with $\tilde{\boldsymbol{X}}_1$ as fixed effects. Therefore it is standard to show that

$$\tilde{\boldsymbol{Y}}^T[\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_n)\}-\{\boldsymbol{I}-\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_n)\}]\tilde{\boldsymbol{Y}} =^d \sigma_{\varepsilon}^2\sum_{s=1}^{p-q}(\theta_s+v_s)^2,$$

where $v_s \sim N(0,1)$, θ_s is the sth component of $\{\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_n)-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_n)\}\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}$ (see for example, page 50 of Christensen 1996). In other words, (A.7) follows a χ^2 distribution with degrees of freedom $p-q$ and noncentrality parameter $\sum\theta_s^2$. Under the H_0, the expectation of $\tilde{\boldsymbol{Y}}$ is $\tilde{\boldsymbol{X}}_0\boldsymbol{\beta}_0$, and by $\{\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{X}}_0=\boldsymbol{0}$, we have $\boldsymbol{\beta}_0^T\tilde{\boldsymbol{X}}_0^T\{\boldsymbol{H}(\tilde{\boldsymbol{X}}_1,\boldsymbol{I}_{n-p})-\boldsymbol{H}(\tilde{\boldsymbol{X}}_0,\boldsymbol{I}_{n-p})\}\tilde{\boldsymbol{X}}_0\boldsymbol{\beta}_0=0$. Therefore (A.7) divided by σ_{ε}^2 follows a χ^2 with degrees of freedom $p-q$ and noncentrality parameter zero. Further observe that $\boldsymbol{W}^T\tilde{\boldsymbol{X}}_1=\boldsymbol{0}$ and $\boldsymbol{W}^T\tilde{\boldsymbol{X}}_0=\boldsymbol{0}$, we obtain that (A.7) is uncorrelated to (A.6). It follows that under the alternative,

$$RSS_0 =^d \sigma_{\varepsilon}^2\left\{\sum_{s=1}^{n-p}u_s^2+\sum_{s=1}^{p-q}(\theta_s+v_s)^2\right\},$$

where $v_s \sim N(0,1)$ are independent of u_s. Under the null, we have

$$RSS_0 =^d \sigma_{\varepsilon}^2\left(\sum_{s=1}^{n-p}u_s^2+\sum_{s=1}^{p-q}v_s^2\right), \quad u_s \sim^{i.i.d.} N(0,1), \quad v_s \sim^{i.i.d.} N(0,1). \qquad (A.8)$$

Putting together the spectral decompositions of RSS_0 and RSS_1 and by (6), we arrive at the decomposition for T_1 in (8) of Theorem 1. When γ is unknown but consistently estimated, $\widehat{\boldsymbol{V}}_0^{-1/2}\boldsymbol{V}_0\widehat{\boldsymbol{V}}_0^{-1/2} = \boldsymbol{I}_n + o_p(1)$. Therefore the variance of $\tilde{\boldsymbol{Y}}$ under the equivalent model (A.1) is approximately $\sigma_{\boldsymbol{\varepsilon}}^2\boldsymbol{I}$ and under the equivalent model (A.2) is approximately $\tilde{V}$. It follows that replacing γ_0 by $\hat{\gamma}$ in the expression (6), we obtain a spectral representation of T_2 in Theorem 1. Finally, we give the decomposition of the restricted likelihood, $f_n(\lambda, \gamma)$, in this Online Appendix A3 which completes the proof of Theorem 1.

A2. Simultaneous diagonalizing

Here we verify (10). Let $\boldsymbol{W}_*$ denote the matrix that simultaneously diagonalizes $\boldsymbol{Z}_1\boldsymbol{\Sigma}_1\boldsymbol{Z}_1^T$ and $\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T$. That is, $\boldsymbol{W}_*^T\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{W}_* = \mathrm{diag}(\mu_s)$, and $\gamma\boldsymbol{W}_*^T\boldsymbol{Z}_1\boldsymbol{\Sigma}_1\boldsymbol{Z}_1^T\boldsymbol{W}_* = \gamma\mathrm{diag}(\omega_s)$. Recall that $\rho_s(\gamma) = \mathrm{eigen}_s(\boldsymbol{W}^T\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T\boldsymbol{W})$, it suffices to show

$$
\begin{aligned}
\mathrm{eigen}_s(\boldsymbol{W}^T\tilde{\boldsymbol{Z}}\boldsymbol{\Sigma}_2\tilde{\boldsymbol{Z}}^T\boldsymbol{W}) &= \mathrm{eigen}_s(\boldsymbol{W}^T\boldsymbol{V}_0^{-1/2}\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{V}_0^{-1/2}\boldsymbol{W}) \\
&= \mathrm{eigen}_s(\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{V}_0^{-1/2}\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{V}_0^{-1/2}) \\
&= \mathrm{eigen}_s[\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{V}_0^{-1/2}\{\boldsymbol{I} - V_0^{-1/2}\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{V}_0^{-1}\boldsymbol{X}_1)\boldsymbol{X}_1\boldsymbol{V}_0^{-1/2}\}\boldsymbol{V}_0^{-1/2}] \\
&= \mathrm{eigen}_s[\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\{\boldsymbol{V}_0^{-1} - \boldsymbol{V}_0^{-1}\boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{V}_0^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{V}_0^{-1}\}] \\
&= \mathrm{eigen}_s[\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\{\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{X}_1, \boldsymbol{V}_0)\}^T\boldsymbol{V}_0^{-1}\{\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{X}_1, \boldsymbol{V}_0)\}] \\
&= \mathrm{eigen}_s\left[\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{W}_*\{\mathrm{diag}(1+\gamma\omega_s)\}^{-1}\boldsymbol{W}_*^T\right] \\
&= \mathrm{eigen}_s\left[\boldsymbol{W}_*^T\boldsymbol{Z}_2\boldsymbol{\Sigma}_2\boldsymbol{Z}_2{}^T\boldsymbol{W}_*\{\mathrm{diag}(1+\gamma\omega_s)\}^{-1}\right] \\
&= \mathrm{eigen}_s\left[\mathrm{diag}(\mu_s)\{\mathrm{diag}(1+\gamma\omega_s)\}^{-1}\right] = \mathrm{diag}\left(\frac{\mu_s}{1+\gamma\omega_s}\right).
\end{aligned}
$$

Here the fourth line is by straightforward matrix algebra and the fifth line is by various properties of $\boldsymbol{W}_*$ including that it diagonalizes $\boldsymbol{V}_0 = \boldsymbol{I} + \gamma\boldsymbol{Z}_1\boldsymbol{\Sigma}_1\boldsymbol{Z}_1{}^T$ and $\boldsymbol{W}_*\boldsymbol{W}_*^T = \boldsymbol{I} - \boldsymbol{H}(\boldsymbol{X}_1, \boldsymbol{I})$ (Kuo 1999).

A3. Spectral decompositions of (restricted) log-likelihood

Under the alternative, the profile log-likelihood is obtained by plugging $\hat{\boldsymbol{\beta}}_1(\gamma, \lambda)$ into the likelihood, which leads to

$$L(\gamma, \lambda) \quad = \quad -n\log\{RSS_1(\gamma, \lambda)\} - \log|\boldsymbol{V}_1|.$$

Using (A.4) and a result from Kuo (1999), we obtain that the log-likelihood assuming H_a has the spectral decomposition

$$L(\gamma, \lambda) = -n\log\left\{\sum_{s=1}^{K} \frac{u_s^2}{1+\lambda\rho_s(\gamma)} + \sum_{s=K+1}^{n-p} u_s^2\right\} - \sum_{s=1}^{K} \log\{1+\lambda\varphi_s(\gamma)\} - \sum_{s=1}^{K} \log(1+\gamma\omega_s).$$

Here we used the fact that only K eigenvalues $\rho_s(\gamma)$ are non-zero. Using (A.8) it is easy to obtain that the likelihood under the H_0 has the representation

$$L(\gamma, 0) = -n\log\left(\sum_{s=1}^{n-p} u_s^2 + \sum_{s=1}^{p-q} v_s^2\right) - \sum_{s=1}^{K} \log(1+\gamma\omega_s).$$

By the above spectral decompositions of $L(\gamma, \lambda)$ and $L(\gamma, 0)$ and some algebra, it follows that LRT_1 has the exact distribution shown in (11). Similar to T_2, when γ is unknown but can be consistently estimated, the null distribution of LRT_2 can be obtained by

$$n\log\left(1 + \frac{\sum_{s=1}^{p-q} v_s^2}{\sum_{s=1}^{n-p} u_s^2}\right) + \sup_{\lambda,\gamma} g_n(\gamma, \lambda).$$

Using a result from Kuo (1999), the restricted profile log-likelihood assuming H_a is

$$\begin{aligned} & RL(\gamma, \lambda) \\ = \quad & -(n-p)\log\left\{\sum_{s=1}^{K} \frac{u_s^2}{1+\lambda\rho_s(\gamma)} + \sum_{s=K+1}^{n-p} u_s^2\right\} - \sum_{s=1}^{K} \log\{1+\lambda\rho_s(\gamma)\} \\ & - \sum_{s=1}^{K} \log(1+\gamma\omega_s) - \log(|\boldsymbol{X}^T\boldsymbol{X}|), \end{aligned}$$

which is $f_n(\gamma, \lambda)$ in Theorem 1 aside from a constant. Assuming H_0, the corresponding restricted log-likelihood is

$$RL(\gamma, 0) = -(n-p)\log\left(\sum_{s=1}^{n-p} u_s^2\right) - \sum_{s=1}^{K} \log\left\{1 + \lambda\rho_s(\gamma)\right\} - \sum_{s=1}^{K} \log(1 + \gamma\omega_s) - \log(|\boldsymbol{X}^T\boldsymbol{X}|).$$

After some algebra, we obtain the null distributions of $RLRT_1$ and $RLRT_2$ presented in section 3.3.

A4. Additional information on the simulations

This section includes an expansion of Tables 1 and 2 in the main text. Here we report all empirical rejection rates and their confidence intervals (Tables A1 and A2 correspond to Table 1 in the main text, and Tables A3 and A4 correspond to Table 2 in the main text).

References

Christensen, R. (1996). Plane answers to complex questions: the theory of linear models. New York: Springer.

Kuo, B.S. (1999). Asymptotics of ML estimator for regression models with a stochastic trend component. *Econometrics Theory*, 15, 2449.

Patterson, H.D., and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-554.

Table A1: Type I error rates and confidence intervals of testing linearity (an unspecified function) or a random effect in five examples, based on 5,000 simulations*.

(a) Partially linear model		$n = 10$	$n_i = 5$		
σ_b	0	0.01	0.1	1	10
Generalized F	0.052 [0.046, 0.059]	0.052 [0.046, 0.059]	0.054 [0.048, 0.060]	0.053 [0.047, 0.060]	0.054 [0.048, 0.061]
pseudo-RLRT	0.050 [0.044, 0.056]	0.049 [0.043, 0.056]	0.052 [0.046, 0.059]	0.052 [0.046, 0.059]	0.054 [0.048, 0.061]
$.5\chi^2_0 + .5\chi^2_1$	0.040 [0.034, 0.045]	0.043 [0.038, 0.049]	0.043 [0.037, 0.049]	0.042 [0.037, 0.048]	0.046 [0.040, 0.052]
(b) Partially linear model		$n = 10$	$n_i = 5$		
σ_α	0	0.01	0.1	1	10
Generalized F	0.052 [0.046, 0.058]	0.052 [0.046, 0.058]	0.048 [0.043, 0.055]	0.050 [0.044, 0.056]	0.055 [0.048, 0.061]
pseudo-LRT	0.050 [0.045, 0.057]	0.046 [0.041, 0.052]	0.049 [0.043, 0.055]	0.055 [0.048, 0.061]	0.054 [0.048, 0.061]
$.5\chi^2_2 + .5\chi^2_3$	0.038 [0.033, 0.044]	0.035 [0.030, 0.040]	0.036 [0.031, 0.041]	0.043 [0.037, 0.049]	0.040 [0.035, 0.046]
(c) Partially linear model		$n = 40$	$n_i = 5$		
σ_{α_1}	0	0.01	0.1	1	10
Generalized F	0.050 [0.044, 0.057]	0.047 [0.041, 0.053]	0.047 [0.041, 0.053]	0.052 [0.046, 0.059]	0.050 [0.044, 0.056]
pseudo-LRT	0.052 [0.046, 0.059]	0.047 [0.041, 0.053]	0.054 [0.048, 0.060]	0.053 [0.047, 0.060]	0.058 [0.051, 0.064]
$.5\chi^2_2 + .5\chi^2_3$	0.032 [0.027, 0.037]	0.030 [0.025, 0.034]	0.031 [0.026, 0.036]	0.034 [0.029, 0.039]	0.035 [0.030, 0.040]
(d)Varying coefficient model			$n = 50$		
σ_{b_1}	0	0.01	0.1	1	10
Generalized F	0.052 [0.046, 0.058]	0.048 [0.042, 0.054]	0.055 [0.049, 0.062]	0.053 [0.047, 0.060]	0.049 [0.044, 0.056]
pseudo-LRT	0.051 [0.045, 0.057]	0.050 [0.044, 0.056]	0.054 [0.048, 0.061]	0.054 [0.048, 0.061]	0.053 [0.047, 0.059]
$.5\chi^2_2 + .5\chi^2_3$	0.041 [0.035, 0.046]	0.040 [0.034, 0.045]	0.043 [0.037, 0.049]	0.042 [0.037, 0.048]	0.041 [0.036, 0.047]
(e)Additive model			$n = 50$		
σ_{b_1}	0	0.01	0.1	1	10
Generalized F	0.049 [0.043, 0.055]	0.052 [0.046, 0.058]	0.047 [0.041, 0.053]	0.055 [0.049, 0.062]	0.051 [0.045, 0.058]
pseudo-RLRT	0.046 [0.040, 0.052]	0.046 [0.040, 0.052]	0.045 [0.039, 0.051]	0.050 [0.044, 0.056]	0.054 [0.048, 0.060]
$.5\chi^2_0 + .5\chi^2_1$	0.023 [0.019, 0.028]	0.026 [0.021, 0.030]	0.026 [0.022, 0.031]	0.029 [0.024, 0.034]	0.032 [0.027, 0.037]

*: Entries are empirical rejects rates and their exact 95% confidence intervals.
(a): Testing for the random intercept with a nuisance unspecified function as in example 1.
(b): Testing for an unspecified function with a nuisance random intercept as in example 1.
(c): Testing for an unspecified function with nuisance random intercept and random slope as in example 2.
(d): Testing for a varying coefficient with a nuisance smooth term as in example 3.
(e): Testing for linearity of a smooth additive function with a nuisance smooth term as in example 4; $corr(t_{i1}, t_{i2})$=0.7.

Table A2: Type I errors and confidence intervals of testing linearity (an unspecified function) or a random effect in five examples, based on 5,000 simulations*.

(a) Partially linear model		$n = 100$	$n_i = 5$		
σ_b	0	0.01	0.1	1	10
Generalized F	0.054 [0.048, 0.061]	0.052 [0.046, 0.058]	0.047 [0.041, 0.053]	0.053 [0.046, 0.059]	0.049 [0.044, 0.056]
pseudo-RLRT	0.051 [0.045, 0.058]	0.050 [0.044, 0.057]	0.048 [0.042, 0.054]	0.052 [0.044, 0.056]	0.050 [0.044, 0.056]
$.5\chi^2_0 + .5\chi^2_1$	0.048 [0.042, 0.055]	0.048 [0.042, 0.054]	0.045 [0.040, 0.052]	0.050 [0.044, 0.056]	0.047 [0.041, 0.053]
(b) Partially linear model		$n = 100$	$n_i = 5$		
σ_α	0	0.01	0.1	1	10
Generalized F	0.046 [0.040, 0.052]	0.050 [0.045, 0.057]	0.047 [0.041, 0.053]	0.054 [0.048, 0.060]	0.048 [0.042, 0.055]
pseudo-LRT	0.040 [0.035, 0.046]	0.045 [0.039, 0.051]	0.041 [0.036, 0.047]	0.048 [0.043, 0.055]	0.052 [0.046, 0.058]
$.5\chi^2_2 + .5\chi^2_3$	0.023 [0.019, 0.028]	0.025 [0.021, 0.030]	0.023 [0.019, 0.028]	0.033 [0.028, 0.038]	0.033 [0.028, 0.038]
(c)Partially linear model		$n = 100$	$n_i = 5$		
σ_{α_1}	0	0.01	0.1	1	10
Generalized F	0.052 [0.046, 0.059]	0.056 [0.050, 0.063]	0.050 [0.044, 0.056]	0.050 [0.045, 0.057]	0.056 [0.049, 0.062]
pseudo-LRT	0.050 [0.044, 0.056]	0.050 [0.044, 0.056]	0.046 [0.040, 0.052]	0.048 [0.042, 0.055]	0.055 [0.049, 0.062]
$.5\chi^2_2 + .5\chi^2_3$	0.030 [0.025, 0.035]	0.033 [0.028, 0.038]	0.029 [0.024, 0.034]	0.031 [0.026, 0.036]	0.037 [0.032, 0.042]
(d)Varying coefficient model			$n = 500$		
σ_{b_1}	0	0.01	0.1	1	10
Generalized F	0.052 [0.046, 0.058]	0.046 [0.041, 0.052]	0.048 [0.043, 0.055]	0.048 [0.042, 0.054]	0.052 [0.046, 0.059]
pseudo-LRT	0.051 [0.045, 0.057]	0.052 [0.046, 0.059]	0.051 [0.045, 0.057]	0.054 [0.048, 0.060]	0.049 [0.043, 0.055]
$.5\chi^2_2 + .5\chi^2_3$	0.031 [0.026, 0.036]	0.032 [0.028, 0.038]	0.031 [0.027, 0.037]	0.032 [0.027, 0.037]	0.030 [0.025, 0.035]
(e)Additive model			$n = 500$		
σ_{b_1}	0	0.01	0.1	1	10
Generalized F	0.046 [0.040, 0.052]	0.050 [0.044, 0.057]	0.051 [0.045, 0.057]	0.052 [0.046, 0.058]	0.051 [0.045, 0.057]
pseudo-RLRT	0.039 [0.034, 0.045]	0.040 [0.035, 0.046]	0.042 [0.036, 0.048]	0.050 [0.044, 0.056]	0.051 [0.045, 0.057]
$.5\chi^2_0 + .5\chi^2_1$	0.018 [0.015, 0.022]	0.019 [0.016, 0.024]	0.017 [0.014, 0.021]	0.027 [0.023, 0.032]	0.027 [0.023, 0.032]

*: Entries are empirical rejects rates and their exact 95% confidence intervals.
(a): Testing for the random intercept with a nuisance unspecified function as in example 1.
(b): Testing for an unspecified function with a nuisance random intercept as in example 1.
(c): Testing for an unspecified function with nuisance random intercept and random slope as in example 2.
(d): Testing for a varying coefficient with a nuisance smooth term as in example 3.
(e): Testing for linearity of a smooth additive function with a nuisance smooth term as in example 4; $corr(t_{i1}, t_{i2})$=0.7.

Table A3: Power of and confidence intervals testing linearity (an unspecified function) or a random effect in six examples, based on 1,000 simulations*.

(a) Partially linear model			$n = 10$	$n_i = 5$	
σ_α	0.3	0.5	0.7	0.8	0.9
Generalized F	0.177 [0.154, 0.202]	0.469 [0.438, 0.500]	0.763 [0.735, 0.789]	0.835 [0.811, 0.857]	0.905 [0.885, 0.922]
pseudo-RLRT	0.180 [0.157, 0.205]	0.461 [0.430, 0.492]	0.763 [0.735, 0.789]	0.829 [0.804, 0.852]	0.904 [0.884, 0.922]
$.5\chi_0^2 + .5\chi_1^2$	0.162 [0.140, 0.186]	0.423 [0.392, 0.454]	0.739 [0.711, 0.766]	0.816 [0.791, 0.840]	0.891 [0.870, 0.910]
(b)Partially linear model			$n = 10$	$n_i = 5$	
d	0.3	0.5	0.7	1	1.2
Generalized F	0.143 [0.122, 0.166]	0.343 [0.314, 0.373]	0.596 [0.565, 0.627]	0.870 [0.848, 0.890]	0.971 [0.959, 0.980]
pseudo-LRT	0.143 [0.122, 0.166]	0.317 [0.288, 0.347]	0.559 [0.528, 0.590]	0.793 [0.767, 0.818]	0.945 [0.929, 0.958]
$.5\chi_2^2 + .5\chi_3^2$	0.106 [0.088, 0.127]	0.264 [0.237, 0.292]	0.505 [0.474, 0.536]	0.745 [0.717, 0.772]	0.929 [0.911, 0.944]
(c) Partially linear model			$n = 20$	$n_i = 5$	
σ_b	0.1	0.3	0.5	0.7	1.5
Generalized F	0.408 [0.377, 0.439]	0.576 [0.545, 0.607]	0.725 [0.696, 0.752]	0.830 [0.805, 0.853]	0.959 [0.945, 0.970]
pseudo-LRT	0.465 [0.434, 0.496]	0.599 [0.568, 0.630]	0.748 [0.720, 0.775]	0.826 [0.801, 0.849]	0.940 [0.923, 0.954]
$.5\chi_2^2 + .5\chi_3^2$	0.400 [0.369, 0.431]	0.554 [0.523, 0.585]	0.706 [0.677, 0.734]	0.796 [0.770, 0.820]	0.935 [0.918, 0.949]
(d) Varying coefficient model			$n = 50$		
d	0.2	0.4	0.6	0.8	1
Generalized F	0.088 [0.071, 0.107]	0.244 [0.218, 0.272]	0.490 [0.459, 0.521]	0.789 [0.762, 0.814]	0.922 [0.904, 0.938]
pseudo-LRT	0.089 [0.072, 0.108]	0.244 [0.218, 0.272]	0.438 [0.407, 0.469]	0.721 [0.692, 0.749]	0.855 [0.832, 0.876]
$.5\chi_2^2 + .5\chi_3^2$	0.071 [0.056, 0.089]	0.221 [0.196, 0.248]	0.394 [0.364, 0.425]	0.677 [0.647, 0.706]	0.830 [0.805, 0.853]
(e)Varying coefficient model			$n = 50$		
d	0.2	0.5	0.7	0.8	1
Generalized F	0.103 [0.085, 0.124]	0.404 [0.373, 0.435]	0.652 [0.622, 0.682]	0.839 [0.815, 0.861]	0.931 [0.913, 0.946]
pseudo-LRT	0.121 [0.101, 0.143]	0.429 [0.398, 0.460]	0.688 [0.658, 0.717]	0.835 [0.811, 0.857]	0.938 [0.921, 0.952]
$.5\chi_2^2 + .5\chi_3^2$	0.101 [0.083, 0.121]	0.389 [0.359, 0.420]	0.650 [0.620, 0.680]	0.813 [0.787, 0.837]	0.930 [0.912, 0.945]
(f) Additive model			$n = 50$		
σ_{b_2}	2	3	5	8	12
Generalized F	0.270 [0.243, 0.299]	0.459 [0.428, 0.490]	0.664 [0.634, 0.693]	0.812 [0.786, 0.836]	0.932 [0.915, 0.947]
pseudo-RLRT	0.266 [0.239, 0.295]	0.454 [0.423, 0.485]	0.651 [0.621, 0.681]	0.808 [0.782, 0.832]	0.925 [0.907, 0.941]
$.5\chi_0^2 + .5\chi_1^2$	0.206 [0.181, 0.232]	0.404 [0.373, 0.435]	0.590 [0.559, 0.621]	0.777 [0.750, 0.802]	0.909 [0.889, 0.926]

*: Entries are empirical rejects rates and their exact 95% confidence intervals.
(a): Testing for the random intercept with a nuisance unspecified function as in example 1.
(b): Testing for an unspecified function with a nuisance random intercept as in example 1.
(c): Testing for an unspecified function with nuisance random intercept and random slope as in example 2.
(d): Testing for a varying coefficient with a nuisance smooth term as in example 3, sine function.
(e): Testing for a varying coefficient with a nuisance smooth term as in example 3, polynomial function.
(f): Testing for linearity of a smooth additive function with a nuisance smooth term as in example 4; $corr(t_{i1}, t_{i2})$=0.7.

Table A4: Power and confidence intervals of testing linearity (an unspecified function) or a random effect in six examples, based on 1,000 simulations*.

(a) Partially linear model	$n = 100$			$n_i = 5$	
σ_α	0.15	0.2	0.25	0.3	0.35
Generalized F	0.165 [0.143, 0.189]	0.308 [0.279, 0.338]	0.551 [0.520, 0.582]	0.778 [0.751, 0.803]	0.916 [0.897, 0.932]
pseudo-RLRT	0.168 [0.145, 0.193]	0.321 [0.292, 0.351]	0.554 [0.523, 0.585]	0.789 [0.762, 0.814]	0.913 [0.894, 0.930]
$.5\chi^2_0 + .5\chi^2_1$	0.159 [0.137, 0.183]	0.306 [0.278, 0.336]	0.541 [0.510, 0.572]	0.779 [0.752, 0.804]	0.910 [0.891, 0.930]
(b) Partially linear model	$n = 100$			$n_i = 5$	
d	0.1	0.15	0.2	0.25	0.3
Generalized F	0.171 [0.148, 0.196]	0.333 [0.304, 0.363]	0.608 [0.577, 0.638]	0.805 [0.779, 0.829]	0.927 [0.909, 0.942]
pseudo-LRT	0.162 [0.140, 0.186]	0.309 [0.280, 0.339]	0.533 [0.502, 0.564]	0.704 [0.675, 0.732]	0.887 [0.866, 0.906]
$.5\chi^2_2 + .5\chi^2_3$	0.114 [0.095, 0.135]	0.243 [0.217, 0.271]	0.462 [0.431, 0.493]	0.640 [0.609, 0.670]	0.839 [0.815, 0.861]
(c)Partially linear model	$n = 100$			$n_i = 5$	
σ_b	0.1	0.2	0.3	0.5	1
Generalized F	0.321 [0.292, 0.351]	0.588 [0.557, 0.619]	0.737 [0.709, 0.764]	0.868 [0.845, 0.888]	0.973 [0.961, 0.982]
pseudo-LRT	0.365 [0.335, 0.396]	0.606 [0.575, 0.636]	0.736 [0.708, 0.763]	0.855 [0.832, 0.876]	0.949 [0.933, 0.962]
$.5\chi^2_2 + .5\chi^2_3$	0.305 [0.277, 0.335]	0.568 [0.537, 0.599]	0.705 [0.676, 0.733]	0.838 [0.814, 0.860]	0.941 [0.925, 0.955]
(d) Varying coefficient model	$n = 500$				
d	0.1	0.15	0.2	0.25	0.3
Generalized F	0.185 [0.161, 0.210]	0.394 [0.364, 0.425]	0.651 [0.621, 0.681]	0.866 [0.843, 0.887]	0.943 [0.927, 0.957]
pseudo-LRT	0.174 [0.151, 0.199]	0.367 [0.337, 0.398]	0.549 [0.518, 0.580]	0.791 [0.764, 0.816]	0.887 [0.866, 0.906]
$.5\chi^2_2 + .5\chi^2_3$	0.127 [0.107, 0.149]	0.297 [0.269, 0.326]	0.470 [0.439, 0.501]	0.722 [0.693, 0.750]	0.848 [0.824, 0.870]
(e) Varying coefficient model	$n = 500$				
d	0.1	0.15	0.2	0.25	0.3
Generalized F	0.193 [0.169, 0.219]	0.446 [0.415, 0.477]	0.717 [0.688, 0.745]	0.907 [0.887, 0.924]	0.985 [0.975, 0.992]
pseudo-LRT	0.220 [0.195, 0.247]	0.459 [0.428, 0.490]	0.743 [0.715, 0.770]	0.907 [0.887, 0.924]	0.980 [0.969, 0.988]
$.5\chi^2_2 + .5\chi^2_3$	0.158 [0.136, 0.182]	0.396 [0.366, 0.427]	0.677 [0.647, 0.706]	0.874 [0.852, 0.894]	0.969 [0.956, 0.979]
(f) Additive model	$n = 500$				
σ_{b_2}	0.1	0.3	0.5	1	2
Generalized F	0.086 [0.069, 0.105]	0.331 [0.302, 0.361]	0.546 [0.515, 0.577]	0.831 [0.806, 0.854]	0.967 [0.954, 0.977]
pseudo-RLRT	0.083 [0.067, 0.102]	0.343 [0.314, 0.373]	0.540 [0.509, 0.571]	0.819 [0.794, 0.842]	0.958 [0.944, 0.970]
$.5\chi^2_0 + .5\chi^2_1$	0.048 [0.036, 0.063]	0.279 [0.251, 0.308]	0.475 [0.444, 0.506]	0.794 [0.768, 0.819]	0.952 [0.937, 0.964]

*: Entries are empirical rejects rates and their exact 95% confidence intervals.
(a): Testing for the random intercept with a nuisance unspecified function as in example 1.
(b): Testing for an unspecified function with a nuisance random intercept as in example 1.
(c): Testing for an unspecified function with nuisance random intercept and random slope as in example 2.
(d): Testing for a varying coefficient with a nuisance smooth term as in example 3, sine function.
(e): Testing for a varying coefficient with a nuisance smooth term as in example 3, polynomial function.
(f): Testing for linearity of a smooth additive function with a nuisance smooth term as in example 4; $corr(t_{i1}, t_{i2})$=0.7.