

## On testing an unspecified function through a linear mixed effects model with multiple variance components

Yuanjia Wang\* and Huaihou Chen

*Department of Biostatistics, Mailman School of Public Health, Columbia University,  
New York, NY 10032, U.S.A.*

\*email: yuanjia.wang@columbia.edu

**SUMMARY:** We examine a generalized  $F$ -test of a nonparametric function through penalized splines and a linear mixed effects model representation. With a mixed effects model representation of penalized splines, we imbed the test of an unspecified function into a test of some fixed effects and a variance component in a linear mixed effects model with nuisance variance components under the null. The procedure can be used to test a nonparametric function or varying-coefficient with clustered data, compare two spline functions, test the significance of an unspecified function in an additive model with multiple components, and test a row or a column effect in a two-way analysis of variance model. Through a spectral decomposition of the residual sum of squares, we provide a fast algorithm for computing the null distribution of the test, which significantly improves the computational efficiency over bootstrap. The spectral representation reveals a connection between the likelihood ratio test (LRT) in a multiple variance components model and a single component model. We examine our methods through simulations, where we show that the power of the generalized  $F$ -test may be higher than the LRT, depending on the hypothesis of interest and the true model under the alternative. We apply these methods to compute the genome-wide critical value and  $p$ -value of a genetic association test in a genome-wide association study (GWAS), where the usual bootstrap is computationally intensive (up to  $10^8$  simulations) and asymptotic approximation may be unreliable and conservative.

**KEY WORDS:** Penalized splines; Likelihood ratio test; Generalized  $F$ -test; Test variance components; Genome-wide association study.

## 1. Introduction

With a mixed effects model representation of penalized splines (Speed 1991; Ruppert et al. 2003; Wand 2003), we imbed the test of an unspecified function into a test of some fixed effects and a variance component in a linear mixed effects model with multiple variance components. Tests involving variance components have non-standard null distributions because some parameters are on the boundary of the parameter space under the null. When data consists of independent subvectors both under the null and alternative, the asymptotic distribution of a likelihood ratio test (LRT) or a restricted LRT (RLRT) is a 50:50 mixture of chi-square distributions (Self and Liang 1987; Stram and Lee 1994). When the independence assumption is violated, Crainiceanu and Ruppert (2004) and Crainiceanu et al. (2005) discovered that the null distribution of the LRT is different from a 50:50 chi-square mixture in models with a single variance component, and that using a chi-square mixture distribution was conservative. For models with a single variance component, Crainiceanu and Ruppert (2004) took advantage of a spectral decomposition of the likelihood to propose a fast algorithm for computing the exact null distribution of the LRT and RLRT. This non-standard behavior was also observed for the degrees-of-freedom test proposed by Cantoni and Hastie (2002), and the null distribution needed to be computed by bootstrap.

To generalize these results to more complex models with nuisance variance components under the null, Greven et al. (2008) proposed to approximate the null distribution of the RLRT using a pseudo-likelihood ratio test theory (Liang and Self 1996). One constructs pseudo-outcomes by subtracting the best linear unbiased predictors (BLUPs) of nuisance random effects and applying methods developed for models with a single variance component to derive the null distribution of the RLRT. Although the procedure generally works well, in some models with highly correlated covariates and a nuisance variance component that is on the boundary of the parameter space, the regularity conditions of the pseudo-RLRT may not be satisfied and a conservative type I error rate has been observed (Greven et al. 2008; Scheipl et al. 2008). No simple spectral decomposition or exact distribution is available in the literature for testing a variance component in linear mixed models with multiple random effects.

We examine a generalized  $F$ -test of a variance component, where there are nuisance random effects under the null. The methods are applicable to testing an unspecified nonparametric function or varying-coefficient through penalized splines with clustered data, comparing two spline functions, testing the significance of an unspecified function in an additive model with multiple components, and testing a row or a column effect in a two-way analysis of variance model. We transform a test of a nonparametric function to a test of some fixed effects and a random effect in a linear mixed effects model with nuisance variance components under the null. We present a spectral decomposition to account for additional variance components in the model and develop a fast algorithm to compute the null distribution of the proposed test. The spectral representation is also used to compare the LRT with the pseudo-LRT, which reveals a connection with methods developed for the single variance component models, and sheds new insights on the geometry of the LRT in multiple variance components models. Compared to the LRT, the generalized  $F$ -test has a computational advantage – only a single linear mixed effects model will be fit under the alternative, which is an attractive feature when the test is carried out many times. For example, in a genome-wide association study (GWAS), the procedure is applied to compute the genome-wide critical value of a genetic association test with correlated family data, where the parametric bootstrap is computationally intensive due to the large number of simulations required (up to  $10^8$  repetitions) and the asymptotic approximation is unreliable and conservative at the extreme tails.

## 2. Models and examples

In this section, we first introduce several motivating examples and then describe the general modeling framework.

**Example 1:** *Test an unspecified function in a partially linear mixed effects model.* For clustered data, such as samples collected from a family study, let  $i$  index families (or clusters) and  $j$  index subjects in a family (or a cluster). Consider a partially linear mixed effects model

$$\begin{aligned} y_{ij} &= f(s_{ij}) + \mathbf{c}_{ij}^T \boldsymbol{\eta} + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i, \\ \alpha_i &\sim N(0, \sigma_\alpha^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \end{aligned} \tag{1}$$

where  $\alpha_i$ 's are independent family-specific random effects,  $f(s_{ij})$  is an unspecified baseline function relating the outcome to the covariates  $s_{ij}$ ,  $\mathbf{c}_{ij}$ 's are vectors of fixed effects with coefficients  $\boldsymbol{\eta}$ , and  $\varepsilon_{ij}$ 's are independent residual measurement errors. Our goal is to test the significance of the regression function  $f(s)$ , that is,

$$H_0 : f(s) = 0,$$

or the deviation of  $f(s)$  from an  $h$ th order polynomial function, that is,

$$H_0 : f(s) = \beta_0 + \beta_1 s + \cdots + \beta_h s^h.$$

Under the alternative, to incorporate a large class of functions, we specify  $f(s)$  to be a flexible spline function, such as

$$f(s) = \beta_0 + \beta_1 s + \cdots + \beta_h s^h + \sum_{k=1}^K b_k (s - \tau_k)_+^h,$$

where  $\tau_k, k = 1, \dots, K$ , are a sequence of knots, and  $(s - \tau)_+^h = (s - \tau)^h$  if  $s \geq \tau$ , and 0 otherwise. A sufficient number of knots will be used to guarantee flexibility. Let  $\mathbf{x}_{ij} = (1, s_{ij}, \dots, s_{ij}^h)^T$  and  $\mathbf{z}_{ij} = \{(s_{ij} - \tau_1)_+^h, \dots, (s_{ij} - \tau_K)_+^h\}^T$ . Under the alternative, we have the representation,  $f(s_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}$ , where  $\mathbf{b} = (b_1, \dots, b_K)^T$ .

To obtain a smooth fitted curve, one minimizes a penalized weighted least squares (Ruppert et al. 2003),

$$\sum_{i=1}^n (\mathbf{Y}_i - \mathbf{C}_i \boldsymbol{\eta} - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{C}_i \boldsymbol{\eta} - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}) + \frac{1}{\lambda} \boldsymbol{\Theta}^T \mathbf{D} \boldsymbol{\Theta},$$

where  $\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})^T$ ,  $\mathbf{C}_i = (\mathbf{c}_{i1}, \dots, \mathbf{c}_{in_i})^T$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^T$ ,  $\mathbf{Z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in_i})^T$ ,  $\boldsymbol{\Theta} = (\boldsymbol{\eta}^T, \boldsymbol{\beta}^T, \mathbf{b}^T)^T$ ,  $\mathbf{V}_i = \text{cov}(\mathbf{Y}_i)$ ,  $\lambda$  is a smoothing parameter,  $\mathbf{D} = \text{diag}(\mathbf{0}_{m+h+1}, \boldsymbol{\Sigma}^{-1})$ ,  $m$  is the dimension of  $\boldsymbol{\eta}$ , and  $\boldsymbol{\Sigma}$  is a known penalty matrix, depending on the spline basis used. For example,  $\boldsymbol{\Sigma} = \mathbf{I}_K$  with a truncated polynomial basis.

*Linear mixed effects model set up for the test:* With a mixed effects model representation of splines (Speed 1991; Ruppert et al. 2003; Wand 2003), solutions to the penalized weighted least squares are obtained from a linear mixed effects model, where  $b_k$  are treated as random effects and the smoothing parameter is specified as the ratio of two variance components.

Specifically, a mixed effects model representation is

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{C}_i \boldsymbol{\eta} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b} + \mathbf{U}_i \alpha_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \\ \mathbf{b} &\sim N(\mathbf{0}, \sigma_b^2 \boldsymbol{\Sigma}), \quad \alpha_i \sim N(0, \sigma_\alpha^2), \quad \boldsymbol{\varepsilon}_i \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_{n_i}), \end{aligned} \quad (2)$$

where the smoothing parameter  $\lambda = \sigma_b^2 / \sigma_\varepsilon^2$  and  $\mathbf{U}_i = \mathbf{1}_{n_i}$ . Through model (2), significance of  $f(s)$  can now be tested by

$$H_0 : \boldsymbol{\beta} = (\beta_0, \dots, \beta_h)^T = \mathbf{0}, \sigma_b^2 = 0.$$

Testing a nonparametric deviation from a polynomial function is simply through  $H_0 : \sigma_b^2 = 0$ .

Note that  $\alpha_i$  in model (2) are nuisance random effects under the null hypothesis.

**Example 2:** *Test an unspecified function in a partially linear mixed effects model with multiple variance components.* An extension to Example 1 is a partially linear mixed effects model with multiple variance components,

$$\begin{aligned} y_{ij} &= f(s_{ij}) + \mathbf{c}_{ij}^T \boldsymbol{\eta} + \alpha_{i0} + \alpha_{i1} u_{ij} + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i, \\ \alpha_{i0} &\sim N(0, \sigma_{\alpha_0}^2), \quad \alpha_{i1} \sim N(0, \sigma_{\alpha_1}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \end{aligned} \quad (3)$$

where  $\alpha_{i0}$  and  $\alpha_{i1}$  are independent random effects. It is again of interest to test  $H_0 : f(s) = 0$ . Here, there are two nuisance variance components under the null ( $\alpha_{i0}$  and  $\alpha_{i1}$ ). A model similar to (2) can be used to test this hypothesis.

**Example 3:** *Varying coefficient model.* In many applications, it is of interest to test an unspecified varying-coefficient or a group difference. Let  $g_i$  denote a group indicator. A flexible model with an unspecified baseline function and a varying-coefficient is

$$y_i = \mathbf{c}_i^T \boldsymbol{\eta} + f(s_i) + \beta(s_i) g_i + \varepsilon_i,$$

where  $\mathbf{c}_i$  is a vector of covariates,  $f(\cdot)$  is a spline function describing the relationship between the expected outcome and the covariate  $s_i$  in the baseline group, and  $\beta(\cdot)$  is the difference between the experiment group and the baseline group. The hypothesis of no group difference is  $H_0 : \beta(s) = 0$ . Using a linear mixed effects model representation of penalized splines on

$f(\cdot)$  and  $\beta(\cdot)$ , the model under the alternative can be expressed as

$$\begin{aligned} \mathbf{Y} &= \mathbf{C}\boldsymbol{\eta} + \mathbf{W}_1\boldsymbol{\beta}_1 + \mathbf{W}_2\boldsymbol{\beta}_2 + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \boldsymbol{\varepsilon}, \\ \mathbf{b}_1 &\sim N(\mathbf{0}, \sigma_{b_1}^2 \boldsymbol{\Sigma}_1), \quad \mathbf{b}_2 \sim N(\mathbf{0}, \sigma_{b_2}^2 \boldsymbol{\Sigma}_2), \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n), \end{aligned} \tag{4}$$

where  $\boldsymbol{\Sigma}_j$ 's are known penalty matrices and  $\mathbf{W}_j$  and  $\mathbf{Z}_j$  are related to the basis functions for  $f(\cdot)$  and  $\beta(\cdot)$ . Testing a group difference is through

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}, \sigma_{b_2}^2 = 0.$$

Here,  $\mathbf{b}_1$  are nuisance random effects under the null. In this example, the nuisance variance ratio,  $\gamma = \sigma_{b_1}^2 / \sigma_\varepsilon^2$ , can be regarded as the smoothing parameter for the baseline function.

**Example 4: Additive models.** Consider an additive model with two covariates,

$$y_i = f_1(s_{i1}) + f_2(s_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2),$$

where  $f_1(\cdot)$  and  $f_2(\cdot)$  are unspecified spline functions and the covariates  $s_{i1}$  and  $s_{i2}$  can be either correlated or independent. When one of the covariates, say  $s_{i2}$ , is of primary interest, one tests  $H_0 : f_2(s) = 0$ , which can also be assessed by testing fixed and random effects in a mixed effects model similar to (4).

Another example where this test is useful includes testing a fixed smoothing parameter with longitudinal data, for example, to test  $\lambda = \lambda_0$  in model (2). The methods can also be used to test a random slope in the presence of a random intercept, test a row or column effect in a two-way analysis of variance model, or test a random effect in a split-plot design through a two-level random effects model.

## 2.1 The general problem setting

All the examples above can be summarized as testing a hypothesis in a linear mixed effects model with multiple variance components. To be specific, our goal is to test

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}, \sigma_{b_L}^2 = 0, \text{ or } H_0 : \sigma_{b_L}^2 = 0$$

in the model

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_0\boldsymbol{\beta}_1 + \mathbf{W}\boldsymbol{\beta}_2 + \sum_{l=1}^L \mathbf{Z}_l\mathbf{b}_l + \boldsymbol{\varepsilon}, \\ \mathbf{b}_l &\sim N(\mathbf{0}, \sigma_{\mathbf{b}_l}^2 \boldsymbol{\Sigma}_l), \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}), \end{aligned} \quad (5)$$

where  $\mathbf{X}_0$  is the design matrix for the  $q$ -dimensional fixed effects under the null,  $\mathbf{X}_1 = (\mathbf{X}_0, \mathbf{W})$  is the design matrix for the  $p$ -dimensional fixed effects under the alternative (in some examples  $p - q = h + 1$ ),  $\mathbf{b}_l$ ,  $l = 1, \dots, L$ , are random effects independent of  $\boldsymbol{\varepsilon}$ , and  $\boldsymbol{\Sigma}_l$  are known matrices.

### 3. A generalized $F$ -test

We develop a generalized  $F$ -test by comparing the residual sum of squares ( $RSS$ ) of two models obtained from (5), similar to the classic ANOVA  $F$ -tests. Classic  $F$ -tests are based on computing  $RSS$  under a restricted model and a full model and dividing them by appropriate degrees of freedom. For tests involving variance components, it is unclear what degrees of freedom should be used (Hodge and Sargent 2001; Vaida and Blanchard 2005). Although the test statistic based on comparing  $RSS$  under the null and the alternative can still be constructed, its null distribution may be non-standard.

We present the test statistic under the general framework in Section 2.1. For the purpose of illustration, assume  $L = 2$  in model (5). It is easy to generalize to the case where  $L > 2$  (e.g., simulation scenario (c) in Tabel 1). Let  $\gamma$  denote the nuisance variance ratio,  $\sigma_{\mathbf{b}_1}^2/\sigma_{\boldsymbol{\varepsilon}}^2$ , and let  $\lambda = \sigma_{\mathbf{b}_2}^2/\sigma_{\boldsymbol{\varepsilon}}^2$ . Under the null hypothesis, the residual sum of squares is

$$RSS_0(\gamma) = \frac{1}{\sigma_{\boldsymbol{\varepsilon}}^2} \{\mathbf{Y} - \mathbf{X}_0 \widehat{\boldsymbol{\beta}}_0(\gamma)\}^T \mathbf{V}_0(\gamma)^{-1} \{\mathbf{Y} - \mathbf{X}_0 \widehat{\boldsymbol{\beta}}_0(\gamma)\},$$

where  $\widehat{\boldsymbol{\beta}}_0(\gamma) = \{\mathbf{X}_0^T \mathbf{V}_0(\gamma)^{-1} \mathbf{X}_0\}^{-1} \mathbf{X}_0^T \mathbf{V}_0(\gamma)^{-1} \mathbf{Y}$  and  $\mathbf{V}_0(\gamma) = \mathbf{I}_n + \gamma \mathbf{Z}_1 \boldsymbol{\Sigma}_1 \mathbf{Z}_1^T$ . Under the alternative, the residual sum of squares is

$$RSS_1(\gamma, \lambda) = \frac{1}{\sigma_{\boldsymbol{\varepsilon}}^2} \{\mathbf{Y} - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1(\gamma, \lambda)\}^T \mathbf{V}_1(\gamma, \lambda)^{-1} \{\mathbf{Y} - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1(\gamma, \lambda)\},$$

where  $\widehat{\boldsymbol{\beta}}_1(\gamma, \lambda) = \{\mathbf{X}_1^T \mathbf{V}_1(\gamma, \lambda)^{-1} \mathbf{X}_1\}^{-1} \mathbf{X}_1^T \mathbf{V}_1(\gamma, \lambda)^{-1} \mathbf{Y}$ , and  $\mathbf{V}_1(\gamma, \lambda) = \mathbf{V}_0(\gamma) + \lambda \mathbf{Z}_2 \boldsymbol{\Sigma}_2 \mathbf{Z}_2^T$ .

When the variance components ratio  $\gamma$  is known to be  $\gamma_0$ , a generalized  $F$ -test can be defined

as

$$T_1 = \frac{RSS_0(\gamma_0) - RSS_1(\gamma_0, \hat{\lambda})}{RSS_1(\gamma_0, \hat{\lambda})/n},$$

where  $\hat{\lambda}$  is estimated by the restricted maximum likelihood (REML) under the alternative hypothesis. When  $\gamma$  is unknown, the test statistic is

$$T_2 = \frac{RSS_0(\hat{\gamma}) - RSS_1(\hat{\gamma}, \hat{\lambda})}{RSS_1(\hat{\gamma}, \hat{\lambda})/n},$$

where both  $\hat{\gamma}$  and  $\hat{\lambda}$  are obtained by REML under the alternative. Both test statistics are easy to compute with any standard statistical software. However, deriving their null distributions is not trivial because the null value of  $\sigma_{b_L}^2$  is on the boundary of the parameter space and the data cannot be partitioned as independent subvectors in certain models, such as (2).

Note that  $T_1$  or  $T_2$  is different from the  $F$  or  $R$  statistic examined in Cantoni and Hastie (2002). The latter statistics is based on  $RSS$  from a conditional model, that is,  $\sum_{i=1}^n (\mathbf{Y}_i - \widehat{\mathbf{Y}}_i)^2$ , where  $\widehat{\mathbf{Y}}_i = \mathbf{X}_{i1}\widehat{\boldsymbol{\beta}}_1 + \mathbf{W}_i\widehat{\boldsymbol{\beta}}_2 + \sum_{l=1}^L \mathbf{Z}_{il}\widehat{\mathbf{b}}_l$ , and  $\widehat{\mathbf{b}}_l$  are BLUPs of  $\mathbf{b}_l$ . The null distributions of these  $F$  and  $R$  tests are unknown in literature and need to be obtained through bootstrap or permutation to account for uncertainties in estimating smoothing parameters (or variance components). In contrast, the proposed  $T_1$  and  $T_2$  are based on marginal models of  $\mathbf{Y}$ . Although when there is a non-linear trend, the linear mixed effects model (5) is not the true model under the alternative, it holds exactly under the null. The null distribution of the proposed generalized  $F$ -test is computed from a valid model under the null and thereby the test is expected to maintain its size.

### 3.1 Spectral decomposition of the test statistic

Here we present a spectral decomposition to obtain the null distribution of  $T_1$  and  $T_2$  that accounts for the nuisance variance components still under the general framework (5) with  $L = 2$ .

**THEOREM 1:** *Let  $\rho_s(\gamma)$  denote the  $s$ th eigenvalue of  $\Sigma_2^{-1/2} \mathbf{Z}_2^T \mathbf{P}_{\mathbf{V}_0} \mathbf{Z}_2 \Sigma_2^{-1/2}$ , where  $\mathbf{P}_{\mathbf{V}_0} = \{\mathbf{V}_0(\gamma)^{-1} - \mathbf{V}_0(\gamma)^{-1} \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{V}_0(\gamma)^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{V}_0(\gamma)^{-1}\}$ . Let  $\omega_s$  denote the eigenvalues of  $\Sigma^{1/2} \mathbf{Z}_1^T \mathbf{P}_1 \mathbf{Z}_1 \Sigma^{1/2}$ , where  $\mathbf{P}_1 = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$ . Then under the null hypothesis (3),*



the generalized  $F$ -test  $T_1$  has the exact distribution

$$T_1 =^d \left\{ \sum_{s=1}^K \frac{\hat{\lambda} \rho_s(\gamma_0)}{1 + \hat{\lambda} \rho_s(\gamma_0)} u_s^2 + \sum_{s=1}^{p-q} v_s^2 \right\} / \left\{ \sum_{s=1}^K \frac{1}{1 + \hat{\lambda} \rho_s(\gamma_0)} u_s^2 + \sum_{s=K+1}^{n-p} u_s^2 \right\}, \quad (6)$$

where  $=^d$  denotes equality in distribution,  $u_s \sim^{i.i.d.} N(0, 1)$ ,  $s = 1, \dots, n - p$ ,  $v_s \sim^{i.i.d.} N(0, 1)$ ,  $s = 1, \dots, p - q$ ,  $\hat{\lambda} = \max_{\lambda} f_n(\gamma_0, \lambda)$ , and

$$\begin{aligned} f_n(\gamma, \lambda) = & -(n - p) \log \left\{ \sum_{s=1}^K \frac{u_s^2}{1 + \lambda \rho_s(\gamma)} + \sum_{s=K+1}^{n-p} u_s^2 \right\} \\ & - \sum_{s=1}^K \log\{1 + \lambda \rho_s(\gamma)\} - \sum_{s=1}^K \log(1 + \gamma \omega_s) \end{aligned} \quad (7)$$

is the spectral decomposition of the log-profile restricted likelihood under the alternative up to a constant. The null distribution of  $T_2$  is obtained by replacing  $\gamma_0$  by  $\hat{\gamma}$  in (6) and computing  $\hat{\gamma}$  and  $\hat{\lambda}$  by  $\max_{\gamma, \lambda} f_n(\gamma, \lambda)$ .

The proof of Theorem 1 is in the Online Appendix A.1. These decompositions allow for fast computation of the null distribution of  $T$  by avoiding permutation.

A similar decomposition can be used to obtain the distribution of the test statistics under the alternative, that is,

$$T_1 =^d \left\{ \sum_{s=1}^K \frac{(\hat{\lambda} - \lambda_0) \rho_s(\gamma_0)}{1 + \hat{\lambda} \rho_s(\gamma_0)} u_s^2 + \sum_{s=1}^{p-q} (\theta_s + v_s)^2 \right\} / \left\{ \sum_{s=1}^K \frac{1 + \lambda_0 \rho_s(\gamma_0)}{1 + \hat{\lambda} \rho_s(\gamma_0)} u_s^2 + \sum_{s=K+1}^{n-p} u_s^2 \right\},$$

where  $\gamma_0$  and  $\lambda_0$  are the true values of  $\gamma$  and  $\lambda$ , and  $\theta_s$  are related to the noncentrality parameter defined in the Online Appendix A1. This expression is useful for the fast computation of power under the alternative hypothesis without bootstrap. The distribution of  $T_2$  under the alternative has a similar representation, with  $\gamma_0$  replaced by  $\hat{\gamma}$ , and  $\hat{\gamma}$  and  $\hat{\lambda}$  computed by  $\max_{\gamma, \lambda} f_n(\gamma, \lambda)$ .

### 3.2 Fast algorithm to compute the null distribution

Taking advantage of the decompositions (6) and (7), one can obtain the exact null distribution of  $T_1$  rapidly with the following Algorithm A:

A0. Pre-simulation step: Compute eigenvalues  $\rho_s(\gamma_0)$ .

A1. Simulate  $n - p$  independent standard normal random variables,  $u_s$ , and  $p - q$  independent standard normal random variables,  $v_s$ .

- A2. Choose  $\lambda$  by maximizing  $f_n(\gamma_0, \lambda)$  in (7) over grid points  $\lambda_1, \dots, \lambda_m$ .
- A3. Compute the test statistic by (6) using the  $\lambda$  selected in step A2.
- A4. Repeat the above steps 1-3 for required number of repetitions.

Note that this algorithm is extremely fast, since both (6) and (7) only involve arithmetic operations and their computations are instantaneous. Using R (R Development Core Team, 2012), we were able to obtain about 80,000 simulations per minute on a Dell computer with 2.67GHz CPU and 4Gz memory.

To obtain the null distribution of  $T_2$ , the steps A0, A2, and A3 in the Algorithm A are replaced by the following Algorithm B:

- B0. Pre-simulation step: Compute eigenvalues  $\rho_s(\gamma_j)$  at grid points  $\gamma_1, \dots, \gamma_m$ .
- B2. With pre-computed  $\rho_s(\gamma_j)$ , choose  $\gamma$  and  $\lambda$  through maximizing  $f_n(\gamma, \lambda)$  in (7) over grid points  $\gamma_1, \dots, \gamma_m$  and  $\lambda_1, \dots, \lambda_m$ .
- B3. Compute the test statistic by (6) using  $\gamma$  and  $\lambda$  selected in step B2.

There are several desirable features regarding the numerical efficiency of Algorithm B. The eigen-decomposition to compute  $\rho_s(\gamma)$  in step B0 only needs to be done once before the simulation starts and the simulation replications are only applied to steps B1 through B3. The speed of eigen-decomposition depends on the column dimension of  $\mathbf{Z}_2$  (or the number of knots), which does not increase with the sample size or the number of nuisance variance components. The algorithm depends on the sample size through simulating  $n - p$  standard normal random variables, thus computation time increases minimally when  $n$  increases. In contrast, a permutation or bootstrap based approach for a mixed effects model may be slower when the number of random effects is large.

The major computational step in Algorithm B is to maximize  $f_n(\gamma, \lambda)$  in (7). After obtaining  $\rho_s(\gamma_j)$ , computing  $f_n(\gamma_j, \lambda_l)$  only involves arithmetic operations, which can be done extremely rapidly. In addition, we studied a search algorithm with respect to  $\lambda$  at a fixed value of  $\hat{\gamma}$  in step B2 and found satisfactory performance in many simulation settings that offers further improvement on the numerical efficiency. In our data analysis example with a large sample size ( $n = 6309$ ), reduction in computing time was more than 3000 folds using the proposed algorithm when compared to the bootstrap. We obtained about 60,000 simulations per minute on a Dell computer with 2.67GHz CPU and 4Gz memory using R.

Under a special case, one can further speed up the pre-simulation step of Algorithm B. We

show in the Online Appendix A2 that when  $\mathbf{Z}_1 \boldsymbol{\Sigma} \mathbf{Z}_1^T$  and  $\mathbf{Z}_2 \boldsymbol{\Sigma}_2 \mathbf{Z}_2^T$  can be simultaneously diagonalized, one obtains

$$\rho_s(\gamma) = \frac{\mu_s}{1 + \gamma \omega_s}, \quad (8)$$

where  $\mu_s$  is the  $s$ th eigenvalue of  $\boldsymbol{\Sigma}_2^{1/2} \mathbf{Z}_2^T \mathbf{P}_1 \mathbf{Z}_2 \boldsymbol{\Sigma}_2^{1/2}$ . Therefore  $\rho_s(\gamma)$  has an explicit expression as a function of  $\gamma$  in this case. With this explicit formula, the eigen-decomposition to obtain  $\mu_s$  and  $\omega_s$  only needs to be done once in step B0 for all grid points of  $\gamma$ . Two matrices can be simultaneously diagonalized if they commute. In example 4 (additive model), for a balanced design where the two covariates are equally spaced and the knots are also equally spaced, the matrices  $\boldsymbol{\Sigma}_1^{-1/2} \mathbf{Z}_1^T \mathbf{P}_1 \mathbf{Z}_1 \boldsymbol{\Sigma}_1^{-1/2}$  and  $\boldsymbol{\Sigma}_2^{-1/2} \mathbf{Z}_2^T \mathbf{P}_1 \mathbf{Z}_2 \boldsymbol{\Sigma}_2^{-1/2}$  commute; therefore, they can be simultaneously diagonalized and the relation (8) holds.

Another speed up of the algorithm is useful when there is more than one nuisance variance component. In this case, step B2 of the algorithm can be replaced by a one-dimensional search where we fix nuisance parameters at values estimated under the alternative hypothesis. We present an example of an additive model with multiple nuisance variance components in Section 4.

### 3.3 Distribution of the (R)LRT with multiple variance components

The decomposition (6) allows a direct comparison between LRT and pseudo-LRT approximation. Let  $L(\gamma, \lambda) = -n \log\{RSS(\gamma, \lambda)\} - \log|\mathbf{V}_1(\gamma, \lambda)|$  denote the profile log-likelihood under the model (2) obtained by substituting the weighted least square estimate  $\hat{\beta}_1(\gamma, \lambda)$  into the likelihood. When the variance ratio  $\gamma_0$  is known, the likelihood ratio test is  $LRT_1 = \sup_{\lambda \geq 0} L(\gamma_0, \lambda) - \sup_{\lambda=0} L(\gamma_0, \lambda)$ . We show in Online Appendix A4 that under the null hypothesis,  $LRT_1$  has the exact distribution

$$LRT_1 = {}^d n \log \left( 1 + \frac{\sum_{s=1}^{p-q} v_s^2}{\sum_{s=1}^{n-p} u_s^2} \right) + \sup_{\lambda} g_n(\gamma_0, \lambda), \quad (9)$$

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

and  $\varphi_s(\gamma)$  are the eigenvalues of  $\Sigma_2^{-1/2} \mathbf{Z}_2^T \mathbf{V}_0^{-1} \mathbf{Z}_2 \Sigma_2^{-1/2}$ . When  $\gamma$  is unknown, we show in the Online Appendix A3 that the null distribution of  $LRT_2$  can be obtained as

$$LRT_2 =^d n \log \left( 1 + \frac{\sum_{s=1}^{p-q} v_s^2}{\sum_{s=1}^{n-p} u_s^2} \right) + \sup_{\lambda, \gamma} g_n(\gamma, \lambda).$$

When the hypothesis of interest only involves a variance component without fixed effects, RLRT is used instead of LRT. We can show that the exact null distribution of RLRT with  $\gamma$  known is

$$RLRT_1 =^d \sup_{\lambda \geq 0} h_n(\gamma_0, \lambda), \quad (10)$$

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

and the null distribution of RLRT with  $\gamma$  unknown is  $RLRT_2 =^d \sup_{\gamma \geq 0, \lambda \geq 0} h_n(\gamma, \lambda)$ .

Greven et al. (2008) computed the null distribution of RLRT by applying methods for the single variance component model (Crainiceanu and Ruppert 2004) through a pseudo-LRT using pseudo-outcomes  $\tilde{\mathbf{Y}}_i = \mathbf{Y}_i - \sum_{l=1}^{L-1} \mathbf{Z}_{il} \hat{\mathbf{b}}_l$ , where  $\hat{\mathbf{b}}_l$  are BLUPs of  $\mathbf{b}_l$ . From (9), it is easy to see that by replacing  $\mu_{s,n}$  in Crainiceanu and Ruppert (2004) with  $\rho_s(\gamma)$ , we arrive at their equation (9). Note that  $\mu_{s,n}$  are eigenvalues of  $\Sigma_2^{1/2} \mathbf{Z}_2^T \mathbf{P}_1 \mathbf{Z}_2 \Sigma_2^{1/2}$  denoted by  $\mu_s$  in this work. Since  $\rho_s(\gamma)$  are eigenvalues of  $\Sigma_2^{1/2} \mathbf{Z}_2^T \mathbf{P}_{\mathbf{V}_0} \mathbf{Z}_2 \Sigma_2^{1/2}$ , it follows that

$$\rho_s(0) = \mu_{s,n}.$$

Therefore, when  $\gamma_0 = 0$  or  $\hat{\gamma} = 0$ ,  $RLRT_1$  and  $RLRT_2$  reduce to the equation (9) in Crainiceanu and Ruppert (2004). In general, when  $\gamma > 0$  there is no explicit expression relating  $\rho_s(\gamma)$  to  $\mu_{s,n}$ . However, when relation (8) in Section 3.2 holds, we substitute (8) into (11) to arrive at the null distribution representation

$$RLRT_1 =^d \sup_{\lambda \geq 0} \left[ (n - p) \log \left\{ 1 + \frac{\sum_{s=1}^K \frac{\lambda \mu_{s,n}/c_{0,s}}{1 + \lambda \mu_{s,n}/c_{0,s}} u_s^2}{\sum_{s=1}^K \frac{u_s^2}{1 + \lambda \mu_{s,n}/c_{0,s}} + \sum_{s=K+1}^{n-p} u_s^2} \right\} - \sum_{s=1}^K \log \{1 + \lambda \mu_{s,n}/c_{0,s}\} \right],$$

where the constants  $c_{0,s} = (1 + \gamma_0 \omega_s)$ . Therefore, scaling  $\mu_{s,n}$  by  $1 + \gamma_0 \omega_s$  would equate the exact null distribution of the  $RLRT_1$  and the pseudo-RLRT in this case. The close relationship between (9) and equation (9) in Crainiceanu and Ruppert (2004) sheds lights

on the validity of approximating LRT in the multiple variance components model by a single variance component based approach in some cases observed in Greven et al. (2008).

## 4. Simulations

### 4.1 Overview of the simulation experiments

We performed simulation studies to examine the type I error rate and power of the generalized  $F$ -test and compare them with the 50:50 chi-square approximation and the pseudo-LRT (or pseudo-RLRT when applicable). In all simulations, we assume the nuisance variance ratio  $\gamma$  to be unknown and examine the performance of  $T_2$ . Performance of  $T_1$  is similar and results are omitted. In all experiments, we simulated covariates  $s_i$  from a uniform distribution with support  $[0,1]$ . We used a linear truncated polynomial basis with  $K = \min(n/10, 35)$  knots, and examined two sample sizes for each scenario. We obtained 5,000 replications to compute the null distribution of  $T_2$ , 5000 replications to compute the empirical rejection rate to assess type I error rate, and 1000 replications to assess power.

We considered five simulation scenarios: (a) Testing  $\sigma_\alpha^2 = 0$  in model (2) of Example 1; (b) Testing  $f(s) = 0$  in model (1) of Example 1; (c) Testing  $f(s) = 0$  in model (3) of Example 2; (d) Testing  $\beta(t) = 0$  in Example 3; and (e) Testing linearity of  $f_2(t)$  in Example 4. The parameters common to the next two subsections are as follows. For case (a) through (d), we let  $\eta = 1$  and simulated  $c_{ij}$  from a uniform distribution with support  $[0,1]$ . We fixed  $\beta = (-0.2, 0.2)^T$  and  $\sigma_b^2 = 1$ . For case (c), the nuisance random intercepts in model (3) have variance  $\sigma_{\alpha_0}^2 = 1$ . For case (d), coefficients  $\beta_1 = (1, -0.5)$  in the varying coefficient model (4). We generated the binary group indicator from a Bernoulli distribution with a success probability of 0.5, and used the effect coding (-1/1 coding). For case (e), we generated covariates  $(s_{i1}, s_{i2})$  from uniform distributions with a correlation of 0.7. For (d) and (e), we centered the covariates  $s_i$  or  $(s_{i1}, s_{i2})$  to improve numerical stability. Lastly, we fixed  $\sigma_\varepsilon^2 = 1$  in all experiments.

## 4.2 Type I error rates

Here we examine the sensitivity of type I error rates of various tests to the presence of the nuisance random effects under the null. The empirical type I error rate was evaluated at a nominal level of 5% and at various values of the variance of the nuisance random effects that ranged from 0 to 100. We show the empirical type I error rates in Table 1 and show the confidence intervals of the error rates computed based on the exact binomial distribution in the Online Supplementary Material (Tables A1 and A2). From these tables, it is seen that the 50:50 chi-square approximation of Stram and Lee (1994) is conservative, regardless of the sample size and the value of the nuisance variance component, except in case (a) with a relatively large sample size. Similar to Greven et al. (2008) and Scheipl et al. (2008), the pseudo-LRT and pseudo-RLRT behave satisfactorily across various settings, except in case (e): a conservative type I error was observed for pseudo-RLRT when the nuisance variance is small and the covariates  $s_{i1}$  and  $s_{i2}$  are highly correlated in the additive model. In this case, when the nuisance variance component  $\sigma_{b_1}^2 = 0, 10^{-4}$  and 0.01, and  $n=500$ , the type I error rate of the pseudo-RLRT is smaller than the nominal level (i.e., the upper bound of the exact 95% confidence interval is smaller than 5%, Table A2 in the Online Supplementary Material). For other values of  $\sigma_{b_1}^2$ , the type I error rate of the pseudo-RLRT adheres to the nominal level. In this scenario, the proposed procedure is robust to the presence of nuisance variance component in the sense that it maintains the correct size for all values of  $\sigma_{b_1}^2$ .

In other simulation settings where the covariates have low correlation or the nuisance variance component is not on the boundary of the parameter space, the generalized  $F$ -test also has type I error close to the nominal level. For example, with the partially linear model, case (a) and (b) in Table 1 show that both the generalized  $F$ -test and the pseudo-RLRT maintain the nominal level of the type I error rate at all the values of nuisance variance. However, the 50:50 chi-square approximation is still conservative. In these settings, we investigated a speed-up of Algorithm  $B$  by replacing Step  $B2$  with a one-dimensional search of  $\lambda$  over  $\lambda_1, \dots, \lambda_m$ , while fixing  $\hat{\gamma}$  as estimated under the alternative. The procedure is instantaneous and provides satisfactory type I error rates similar to those reported in Table 1. In case (c), we investigate the performance of the proposed procedure when there

are multiple nuisance variance components under the null (random intercepts and random slopes), and show that the proposed test maintains the correct type I error rate.

### 4.3 Power of the tests

The next six sets of simulations compare the power of various tests. The true models under the alternative hypothesis are specified as follows. For case (a), we simulated data from model (2) with  $\sigma_b^2 = 1$  and let  $\sigma_\alpha^2$  take different values. For case (b), we fixed  $\sigma_\alpha^2 = 1$  and let  $f(t) = d * \sin(2\pi t)$  in the partially linear mixed effects model (1). The parameter  $d$  serves as a measure of the effect size. For case (c), we fixed both variances  $\sigma_{\alpha_0}^2$  and  $\sigma_{\alpha_1}^2$  at one in the partially linear model (3). We let  $\beta = (0.3, 0.3)^T$  for the case  $n = 40$ ,  $n_i = 5$ ;  $\beta = (0.1, 0.1)^T$  for the case  $n = 100$ ,  $n_i = 5$ ; and let the variance  $\sigma_b^2$  take different values. For the varying coefficient model, we fixed the nuisance variance component at  $\sigma_{b_1}^2 = 1$  and let  $\beta(t) = d * \sin(2\pi t)$  in scenario (d), let  $\beta(t) = d * (t + t^2)$  in scenario (e), and let  $d$  range from 0.1 to 1. For case (f), we fixed the nuisance variance component at  $\sigma_{b_1}^2 = 1$  in the additive model and let the variance component of interest,  $\sigma_{b_2}^2$ , take different values.

We present the numerical results in Tables A3 and A4 in the Online Appendix A.4 and depict the power comparisons in Figure 1 and Figure 2. We see that in all six scenarios, the 50:50 chi-square approximation is less powerful because it is a conservative procedure. The loss of power can be up to 16% compared to the proposed test. Depending on the hypothesis being tested and the true underlying model, there are some differences in power between the generalized  $F$ -test and the pseudo-LRT or RLRT. For example, in scenario (a), when testing for the random intercepts in model (2), there is no difference between the proposed and RLRT. In scenario (b), when testing the presence of a sine function in a partially linear model, the LRT is less powerful than the generalized  $F$ -test for several values of  $d$ . For scenario (c) and with a sample size of 200, the LRT appears to be slightly more powerful than the generalized- $F$  when  $\sigma_b^2 = 0.01$ .

For the varying coefficient model in case (d), when the function being tested is a sine function, the proposed test has a greater power than the pseudo-LRT. The power gain is up to 10%. When the function being tested is simpler (less number of modes), such as a quadratic function in scenario (e), the pseudo-LRT is slightly more powerful than the proposed  $F$ -test.

For the additive model in case (f), the proposed test behaves similarly to the pseudo-RLRT. A computational advantage of the generalized  $F$ -test is that it requires fitting only one linear mixed effects model. The reduction in computational burden is important when the test is applied many times, such as in a GWAS, where  $T_1$  or  $T_2$  is computed for up to a million single nucleotide polymorphisms (SNPs) along the genome.

## 5. Data analysis

In 2007, dense SNP genotyping (550,000 SNPs) was conducted in the Framingham Heart Study (FHS) to map genes associated with risk factors of cardiovascular disease (CVD). In this genome-wide association study (GWAS), the research goal is to test the association between a SNP and a risk factor for CVD. Here the outcome of interest is the systolic blood pressure (SBP), a complex trait influenced by both environmental and genetic factors. In the literature, the heritability of SBP is estimated to be high (between 30% to 60%, Levy et al. 2000), which suggests a substantial genetic contribution. Several GWAS were conducted to identify SNPs that can explain the high heritability (Levy et al. 2009). However, most work in the current literature do not account for the age-specific genetic effect.

We test for the genetic association by the proposed generalized  $F$ -test using the FHS baseline SBP data incorporating the age effect by a varying coefficient mixed effects model. There were 6309 subjects from 951 families with available SBP included in the analysis. In the literature on genetic analysis of SBP, a log-transformation is often applied to better satisfy the model assumptions (Byng et al. 2003; Cui and Sheffield 2003). Therefore, a log-transformation of SBP was also applied in our analysis. We show a scatter plot of SBP and  $\log(\text{SBP})$  against age in Figure 3.

Consider the varying-coefficient mixed effects model,

$$y_{ij} = \mathbf{c}_{ij}^T \boldsymbol{\eta} + \beta(s_{ij})g_{ij} + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_{ij}, \quad (11)$$

where  $y_{ij}$  is the log-transformed SBP of the  $j$ th subject in the  $i$ th family,  $\alpha_i$  is a family-specific random effect,  $g_{ij}$  is the genotype at a SNP for this subject coded as number of the minor alleles,  $s_{ij}$  is the subject's age,  $\mathbf{c}_{ij}$  are fixed effects (such as gender, age, age-squared,



body mass index (BMI) and BMI-squared), and  $\beta(s_{ij})$  is an additive genetic effect. It is of interest to test no genetic effect, that is,  $H_0 : \beta(s) = 0$  in (11). Under the alternative, usual practice specifies  $\beta(s)$  as a linear function. Wang et al. (2012) observed power loss of the linear model analysis when the true genetic effect is nonlinear. Since the true genetic effect is unknown in practice, a flexible model under the alternative is desirable. We specify  $\beta(s)$  using a quadratic truncated polynomial base under the alternative and test the association through penalized spline. With the mixed model representation, the test falls under the general framework in model (5) and we test  $H_0 : \beta_2 = 0, \sigma_{b_2}^2 = 0$ .

To correct for the large number of tests performed in a GWAS, the genome-wide significance level is usually set as  $10^{-8}$  (Levy et al. 2009). We created  $10^8$  repetitions to examine the null distribution of the test statistic. The computing time for  $10^8$  null distribution simulations was around 28 hours on a Dell computer with 2.67GHz CPU and 4GHz memory, compared to an estimated  $10^5$  hours by bootstrap (estimated from performing 100 bootstraps). The long computing time for bootstrap is partially due to a large number of random effects (families). The genome-wide critical value was computed to be 7.84.

Levy et al. (2009) conducted a meta analysis of six GWAS on blood pressure traits and reported several promising regions that may harbor genes for SBP. We selected two promising candidate regions on chromosome 12 to analyze. There were 42 SNPs available in these regions (19 from region 88485Kb to 88605Kb and 23 from region 110305Kb to 110390Kb). The  $p$ -values were computed as the proportions of the simulated null test statistics greater than or equal to the observed. The top ranking SNPs are rs7136259, rs17249754, and rs11065898 with the respective  $p$ -values 0.00012, 0.00015, and 0.00068. The minor allele frequencies of these SNPs are 18%, 21%, and 34%, respectively. The first two top-ranking SNPs locate in the gene ATP2B1 (88585Kb and 110347Kb on chromosome 12) and the third one locates in the gene SH2B3 (88605Kb on chromosome 12). Both genes were reported to be associated with hypertension and blood pressure in several studies (Levy et al. 2009; Newton-Cheh et al. 2009). We present the fitted genetic effect  $\hat{\beta}(s)$  on the original scale for three SNPs in Figure 4, which suggests a nonlinear trend in all three cases. For SNPs rs7136259 and rs17249754, the genetic effect appears to be larger at the early and late ages

(before age 20 and after age 50), while for SNP rs11065898, the genetic effect fluctuates across time.

These analyses provide some evidence of the presence of gene-age interaction, which in some cases may not have been identified if the time-trend had been ignored (Wang et al. 2012). There are reports of empirical evidence and theoretical justification for genetic factors controlling time-varying developmental features of a phenotype in plant, animal, and human genetic literature (Province and Rao 1985, Rice 2002, He et al. 2010). Since aging is a complex biological process during which many physiological changes may take place, age may represent a surrogate of various unmeasured biological factors. Taking into account the gene-age interaction in a genetic association study may help with resolving some of the inconsistencies in replicating a genetic finding, and may increase power of association tests (e.g., Lasky-Su et al. 2008; Wang et al. 2012).

This data analysis example shows that the proposed procedure solves a computational problem encountered in obtaining reliable genome-wide significance level for large-scale studies such as GWAS, where permutation is practically unrealistic and asymptotic approximation may be conservative and unstable.

## 6. Discussion

In this work, we propose a test of an unspecified function through linear mixed effects models with more than one variance component. We present a spectral decomposition of the test statistic that offers fast computation of the null and alternative distributions and reveals the connection between (R)LRT for single and multiple variance component models. The computational gain over permutation and bootstrap is especially useful for large scale experiments, such as the GWAS or microarray gene expression studies. With a large scale study, improving power is important, therefore an asymptotic chi-square approximation that leads to a conservative decision is highly undesirable. We observe some power differences between the generalized  $F$ -test and the (R)LRT depending on the true underlying model. In practice, when the true underlying function is unknown, these tests may supplement each

other in terms of power. For the LRT or RLRT, the R package RLRsim (Scheipl et al. 2008) can be used to compute the null distribution efficiently.

Here, we use a reduced-rank penalized spline to represent an unknown function under the alternative. An extension to applying the test and its null distribution computation with smoothing splines or kernel machines is possible through mixed effects model representations. Lastly, the methods can be extended to O’Sullivan penalized splines using the B-splines basis functions through a penalty matrix presented in Wand and Ormerod (2008).

## 7. Supplementary Material

Web Appendices and Tables referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library.

## Acknowledgements

Wang’s research is supported by NIH grants AG031113-01A2 and NS073671-01. The Framingham data was obtained from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine (Contract No. N01-HC-25195). The authors wish to thank the associate editor and two anonymous reviewers for extremely helpful comments and suggestions that have greatly improved this work, and thank Ms. Christine Mauro for editorial assistance.

## References

- Byng, M.C., Fisher, S.A., Lewis, C.M., and Whittaker, J.C. (2003). Variance components linkage analysis for adjusted systolic blood pressure in the Framingham Heart Study. *BMC Genetics* 4 Suppl 1:S4.
- Cantoni, E. and Hastie, T. (2002). Degrees-of-freedom tests for smoothing splines. *Biometrika* 89 (2), 251-263.
- Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* 65, 165-185.

- Crainiceanu, C., Ruppert, D., Claeskens, G., and Wand, P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* 92, 91-103.
- Cui, J.S. and Sheffield, L.J. (2003). Bivariate variance-component analysis, with application to systolic blood pressure and total cholesterol levels in the Framingham Heart Study. *BMC Genetics* 4, Suppl 1:S81.
- Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines. *Statistical Science* 11, 89-121.
- Greven, S., Crainiceanu, C., Kühnhoft, H., and Peters, A. (2008). Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models. *Journal of Computational and Graphical Statistics*, 17 (4): 870-891.
- He, Q., Berg, A., Li, Y., Vallejos, C.E., and Wu, R. (2010). Mapping Genes for Plant Structure, Development and Evolution: Functional Mapping Meets Ontology. *Trends Genet* 26, 39-46.
- Hodge, J. and Sargent, D.J. (2001). Counting degrees of freedom in hierarchical and other richly - parametrised models. *Biometrika* 88, 367-379.
- Lasky-Su, J., Lyon, H.N., Emilsson, V., Heid, I.M., Molony, C., Raby, B.A., Lazarus, R., Klanderma, B., Soto-Quiros, M.E., Avila, L., Silverman, E.K. et. al. (2008). On the Replication of Genetic Associations: Timing Can Be Everything! *American Journal of Human Genetics* 82, 849-858.
- Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J. et al. (2009), Genome-wide Association Study of Blood Pressure and Hypertension, *Nature Genetics* 41, 677 - 687.
- Liang, K.-Y. and Self, S. G. (1996). On the Asymptotic Behaviour of the Pseudolikelihood Ratio Test Statistic. *Journal of the Royal Statistical Society, Series B* 58, 785-796.
- Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., et al. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics* 41(6), 666-676.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-103.
- R Development Core Team. (2012). R: A Language and Environment for Statistical Comput-

- ing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>.
- Rice, S.H. (2002). A general population genetic theory for the evolution of developmental interactions. *Proceedings of the National Academy of Sciences* 99, 15518-15523
- Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric regression*. New York: Cambridge University Press.
- Self, S. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of American Statistics Association* 82 (398), 605-610.
- Speed, T. (1991). Comment on that BLUP is a good thing: the estimation of random effects. *Statistical Science* 6, 42-44.
- Stram, D. and Lee, J.-W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* 50(3), 1171-1177.
- Scheipl, F., Greven, S., and Kühnhoﬀ, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics and Data Analysis* 52 (7): 3283-3299.
- Vaida, F. and Blanchard, S. (2005) Conditional Akaike information for mixed-effects models. *Biometrika* 92: 351-370
- Wand, M.P. (2003), Smoothing and mixed models, *Computational Statistics* 18, 223-249.
- Wand, M.P., and Ormerod, J.T. (2008). On semiparametric regression with O’Sullivan penalised splines. *Australia and New Zealand Journal of Statistics* 50, 179-198.
- Wang, Y., Huang, C., Fang, Y., Yang, Q., and Li, R. (2012). Flexible semiparametric analysis of longitudinal genetic studies by reduced rank smoothing. *Applied Statistics: Journal of the Royal Statistical Society, Series C*. 61, 1-24.

[Table 1 about here.]

[Figure 1 about here.]

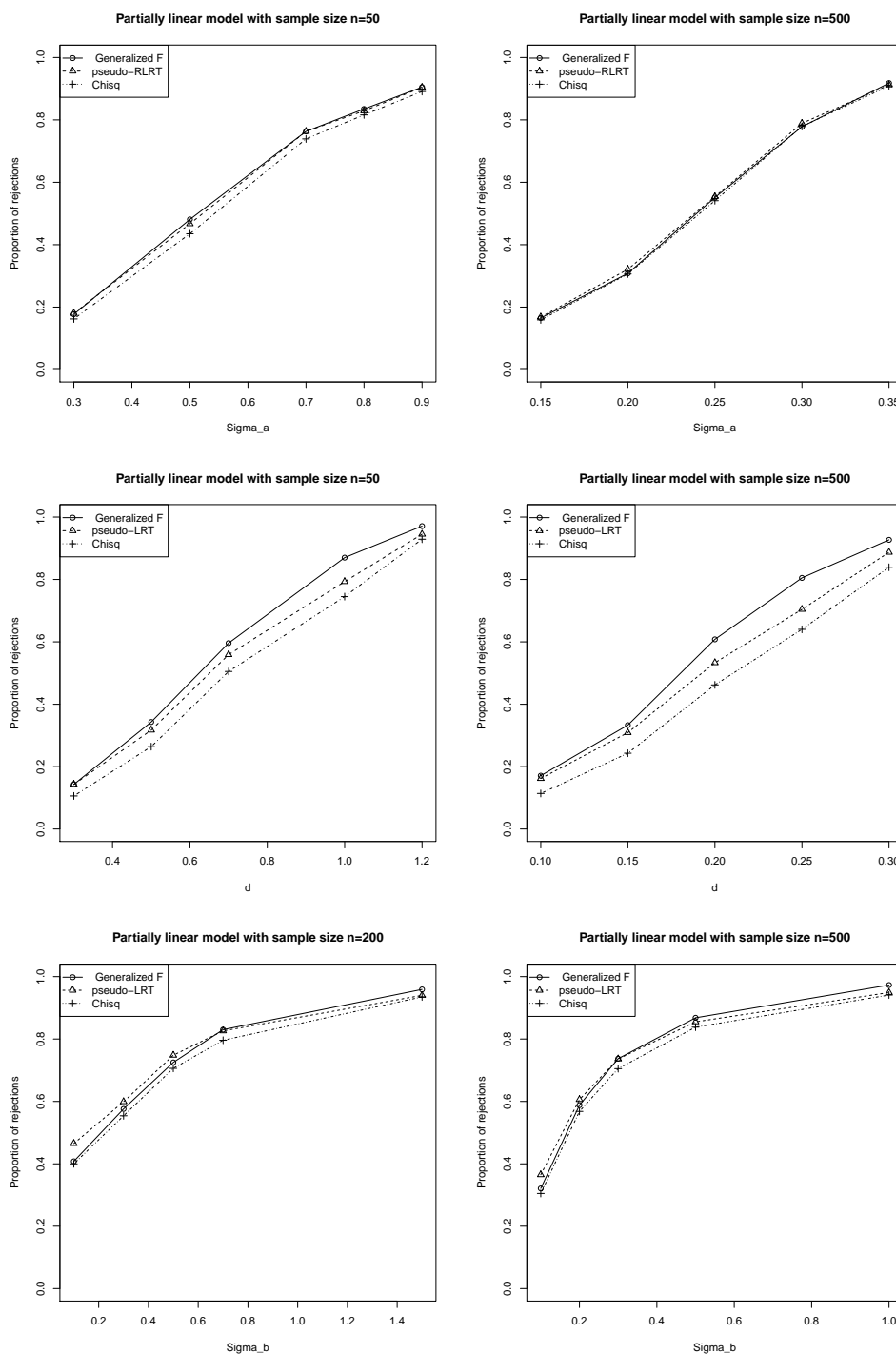
[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

*Received April 2011. Revised April 2011. Accepted April 2011.*

**Figure 1.** Power of various tests in scenarios (a) (top two panels), (b) (middle two panels), and (c) (bottom two panels) based on 1,000 simulations.

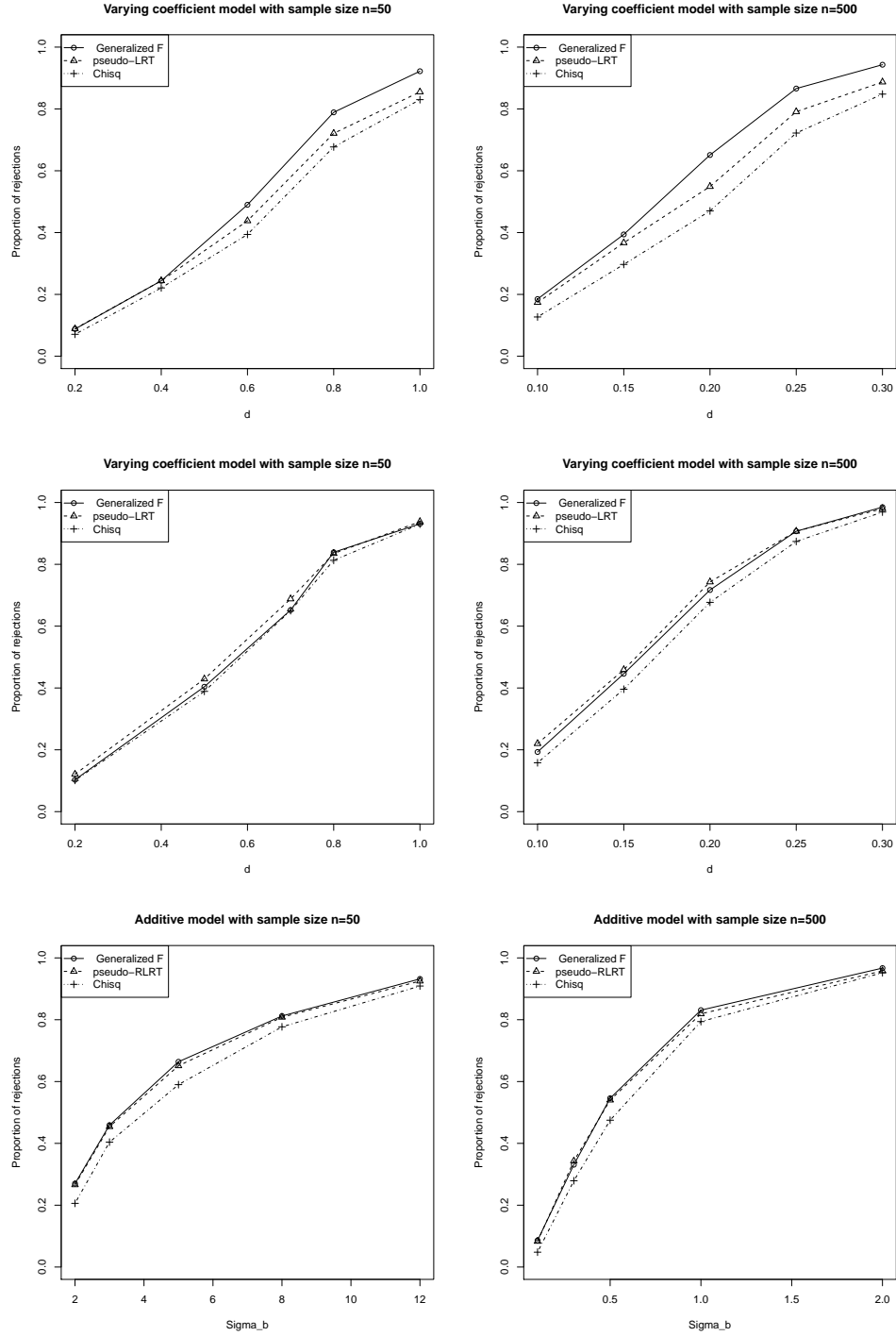


(a): Testing for the random intercept with a nuisance unspecified function in example 1.

(b): Testing for an unspecified function with a nuisance random intercept in example 1.

(c): Testing for an unspecified function with nuisance random intercepts and random slopes in example 2.

**Figure 2.** Power of various tests in scenarios (d) (top two panels), (e) (middle two panels), and (f) (bottom two panels) based on 1,000 simulations.



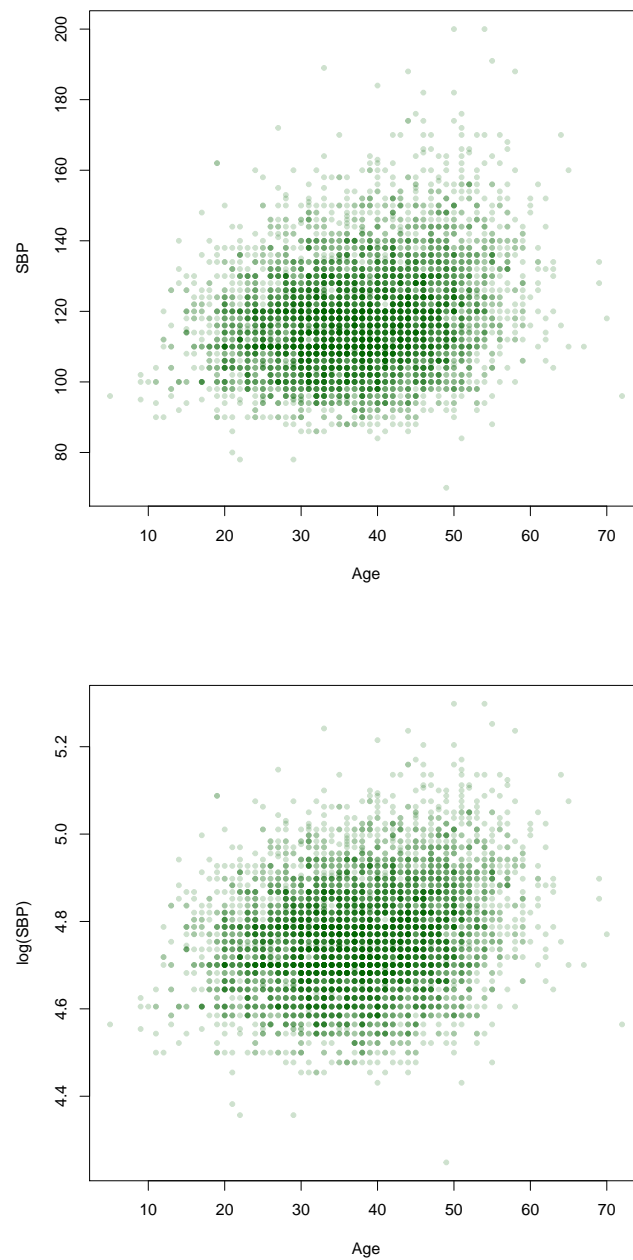
(d): Testing for a varying coefficient with a nuisance smooth term in example 3, sine function.

(e): Testing for a varying coefficient with a nuisance smooth term in example 3, polynomial function.

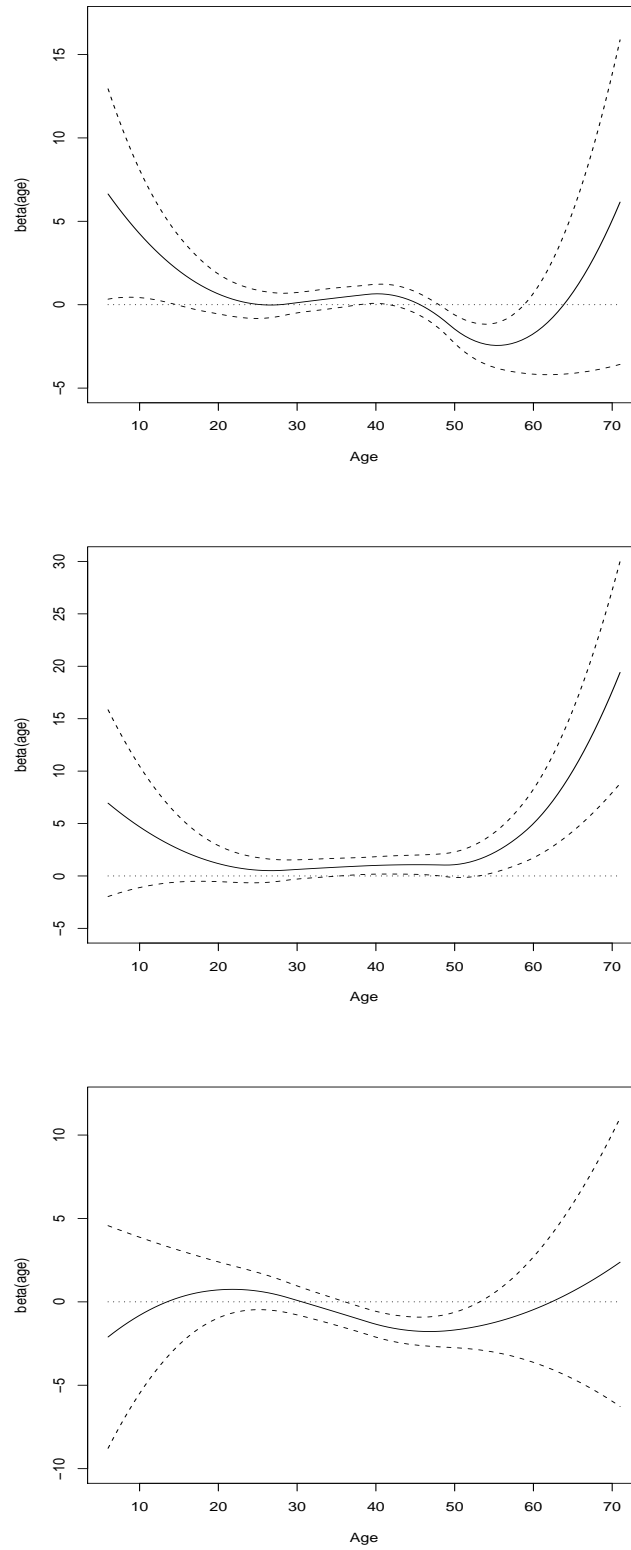
(f): Testing for linearity of a smooth additive function with a nuisance smooth term in example 4;  $\text{corr}(t_{i1}, t_{i2})=0.7$ .



**Figure 3.** Scatterplot of the SBP versus age (top panel) and  $\log(\text{SBP})$  versus age (bottom panel) in 6309 Framingham samples



**Figure 4.** Fitted age-specific genetic effect  $\hat{\beta}(s)$  (solid lines) and its pointwise confidence interval (dashed lines) for SNPs rs7136259 (top panel), rs17249754 (middle panel) and rs11065898 (bottom panel)



**Table 1**

*Type I error of testing linearity (an unspecified function) or a random effect in five examples, based on 5,000 simulations.*

(a) Partially linear model												
			$n = 10$			$n_i = 5$			$n = 100$			$n_i = 5$
$\sigma_b$	0	0.01	0.1	1	10	0	0.01	0.1	1	10	0	0.01
Generalized $F$	0.052	0.052	0.054	0.053	0.054	0.054	0.052	0.047	0.053	0.049	0.053	0.049
pseudo-LRT	0.050	0.049	0.052	0.052	0.054	0.051	0.050	0.048	0.052	0.050	0.052	0.050
$.5\chi_0^2 + .5\chi_1^2$	0.040	0.043	0.043	0.042	0.046	0.048	0.048	0.045	0.045	0.045	0.050	0.047
(b) Partially linear model												
			$n = 10$			$n_i = 5$			$n = 100$			$n_i = 5$
$\sigma_a$	0	0.01	0.1	1	10	0	0.01	0.1	1	10	0	0.01
Generalized $F$	0.052	0.052	0.048	0.050	0.055	0.046	0.050	0.047	0.054	0.048	0.054	0.048
pseudo-LRT	0.050	0.046	0.049	0.055	0.054	0.040	0.045	0.041	0.048	0.048	0.052	0.052
$.5\chi_2^2 + .5\chi_3^2$	0.038	0.035	0.036	0.043	0.040	0.023	0.025	0.023	0.023	0.033	0.033	0.033
(c) Partially linear model												
			$n = 40$			$n_i = 5$			$n = 100$			$n_i = 5$
$\sigma_{\alpha_1}$	0	0.01	0.1	1	10	0	0.01	0.1	1	10	0	0.01
Generalized $F$	0.050	0.047	0.047	0.052	0.050	0.052	0.056	0.050	0.050	0.050	0.050	0.056
pseudo-LRT	0.052	0.047	0.054	0.053	0.058	0.050	0.050	0.046	0.048	0.048	0.055	0.055
$.5\chi_2^2 + .5\chi_3^2$	0.032	0.030	0.031	0.034	0.035	0.030	0.033	0.029	0.031	0.031	0.037	0.037
(d) Varying coefficient model												
			$n = 50$			$n = 500$			$n = 500$			
$\sigma_{b_1}$	0	0.01	0.1	1	10	0	0.01	0.1	1	10	0	0.01
Generalized $F$	0.052	0.048	0.055	0.053	0.049	0.052	0.046	0.048	0.048	0.052	0.048	0.052
pseudo-LRT	0.051	0.050	0.054	0.054	0.053	0.051	0.052	0.051	0.054	0.049	0.054	0.049
$.5\chi_2^2 + .5\chi_3^2$	0.041	0.040	0.043	0.042	0.041	0.031	0.032	0.031	0.032	0.032	0.032	0.030
(e) Additive model												
			$n = 50$			$n = 500$			$n = 500$			
$\sigma_{b_1}$	0	0.01	0.1	1	10	0	0.01	0.1	1	10	0	0.01
Generalized $F$	0.049	0.052	0.047	0.055	0.051	0.046	0.050	0.051	0.052	0.051	0.052	0.051
pseudo-LRT	0.046	0.046	0.045	0.050	0.054	0.039	0.040	0.042	0.050	0.051	0.050	0.051
$.5\chi_0^2 + .5\chi_1^2$	0.023	0.026	0.026	0.029	0.032	0.018	0.019	0.017	0.027	0.027	0.027	0.027

- (a): Testing for the random intercept with a nuisance unspecified function in example 1.
- (b): Testing for an unspecified function with a nuisance random intercept in example 1.
- (c): Testing for an unspecified function with a nuisance random intercept and random slope in example 2.
- (d): Testing for a varying coefficient with a nuisance smooth term in example 3.
- (e): Testing for linearity of a smooth additive function with a nuisance smooth term in example 4;  $corr(t_{i1}, t_{i2})=0.7$ .