Biostatistics (2003), **4**, 1, *pp.* 57–74
Printed in Great Britain

Hypothesis testing in semiparametric additive mixed models

DAOWEN ZHANG*
Department of Statistics, North Carolina State University, USA
dzhang2@stat.ncsu.edu

XIHONG LIN
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA
xlin@umich.edu

SUMMARY

We consider testing whether the nonparametric function in a semiparametric additive mixed model is a simple fixed degree polynomial, for example, a simple linear function. This test provides a goodness-of-fit test for checking parametric models against nonparametric models. It is based on the mixed-model representation of the smoothing spline estimator of the nonparametric function and the variance component score test by treating the inverse of the smoothing parameter as an extra variance component. We also consider testing the equivalence of two nonparametric functions in semiparametric additive mixed models for two groups, such as treatment and placebo groups. The proposed tests are applied to data from an epidemiological study and a clinical trial and their performance is evaluated through simulations.

Keywords: Equivalence test; Goodness of fit; Longitudinal data; Mixed models; Nonparametric regression; Polynomial test; Score test; Variance components.

1. INTRODUCTION

Linear mixed models (Laird and Ware, 1986) and their extension generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993) are widely used to analyse clustered data such as longitudinal data. A key assumption in these models is that the conditional mean of the outcome variable given the random effects depends on the covariates parametrically. Since this parametric assumption in GLMMs is strong and may not be appropriate for data with complex covariate effects, Lin and Zhang (1999) proposed generalized additive mixed models (GAMMs) that allow for flexible modeling of the covariate effects by replacing the linear predictor in GLMMs with an additive combination of nonparametric functions of covariates and random effects.

It is of substantial interest to test whether a simple parametric GLMM can fit the data well compared with a more complicated nonparametric GAMM. First, regression coefficients in an appropriate parametric GLMM might have an appealing practical interpretation. Second, despite its substantial flexibility, estimation in a GAMM is computationally much more intensive than estimation in a GLMM. One is often interested in having a tool to check for the goodness-of-fit of a parametric model where the covariate effects are assumed to be some fixed-degree polynomial.

*To whom correspondence should be addressed

One example is the longitudinal study of respiratory infection in 275 Indonesian children (Diggle *et al.*, 1994), which will be analysed in Section 5. Each child was examined every quarter up to six consecutive quarters for the presence of respiratory infection. One is interested in the age effect on the risk of respiratory infection. Diggle *et al.* (1994) fit a logistic mixed model assuming a quadratic age effect, while Lin and Zhang (1999) allowed the age effect to be nonparametric. A question of interest is whether the simple quadratic age model is appropriate compared with the nonparametric model.

For independent data, several authors considered testing whether a nonparametric function is a fixed-degree polynomial. Cox and Koh (1989) derived a test statistic for testing the adequacy of polynomial regression based on the smoothing spline formulation of the nonparametric function. Härdle *et al.* (1998) proposed a likelihood-ratio-based test using bootstrap to compare parametric generalized linear models with semiparametric generalized partial linear models. For additional references, see Eubank and Hart (1992); Azzalini and Bowman (1993) and Fan and Huang (2001), among others. These authors all assume the data are independent. Little work has been done to test the goodness of fit of parametric models against alternative nonparametric models for correlated clustered data.

A second problem of common interest in many applications of nonparametric regression is to compare nonparametric covariate effects between two groups. For example, Breslow and Clayton (1993) used a GLMM to analyse epileptic seizure count data from a clinical trial of an anti-epileptic drug and found different nonlinear baseline seizure count effects between the treatment group and the control group. It is desirable to develop an equivalence test for the baseline seizure count effects between the two groups. Several authors considered testing the equivalence of two nonparametric functions for independent Gaussian data. See Härdle and Marron (1990); Hall and Hart (1990) and Young and Bowman (1995), among others. Several tests were recently developed to test the equivalence of curves for longitudinal Gaussian data, see, for example, Fan and Lin (1998) and Zhang *et al.* (2000). Little work however has been done on testing the equivalence of two nonparametric functions for correlated non-Gaussian data.

We tackle these two problems in this paper. Specially, we consider a goodness-of-fit test for polynomial regression versus nonparametric regression for clustered Gaussian and non-Gaussian data, such as longitudinal data, using semiparametric additive mixed models (SAMMs), a special case of GAMMs. The test is based on the mixed-model formulation of the smoothing spline estimator of the nonparametric function in SAMMs. By treating the inverse of the smoothing parameter as an extra variance component, the variance component score test developed for parametric GLMMs (Lin, 1997) is adapted to construct a goodness-of-fit test of polynomial regression in SAMMs. Due to the special structure of the smoothing matrix, a scaled chi-square test is proposed for the goodness-of-fit test. We also consider an equivalence test for two nonparametric functions between two groups in SAMMs for clustered data. Simulation studies are conducted to evaluate the performance of these tests.

This paper is organized as follows. In Section 2, we present the model. In Section 3, we discuss the polynomial tests for clustered Gaussian and non-Gaussian responses in SAMMs. We present the equivalence test for two nonparametric functions in Section 4. We apply these two tests to Indonesian respiratory infection data and to epileptic seizure count data in Section 5. In Section 6, we report the results from simulation studies to evaluate the performance of these two tests. We conclude the paper with discussion in Section 7.

2. SEMIPARAMETRIC ADDITIVE MIXED MODELS

In this section, we present SAMMs for clustered data, and briefly discuss the estimation procedure. These models are special cases of GAMMs considered by Lin and Zhang (1999). Let the data consist of a response variable y_{ij} for the jth observation ($j = 1, \ldots, n_i$) of the ith cluster ($i = 1, \ldots, m$), a scalar covariate t_{ij}, a $p \times 1$ covariate vector s_{ij} associated with fixed effects, and a $q \times 1$ covariate vector z_{ij} associated with random effects. Conditional on a $q \times 1$ vector of random effects b_i, the y_{ij} are assumed to be independent with conditional means $\mathrm{E}(y_{ij}|b_i) = \mu_{ij}^b$ and conditional variances $\mathrm{var}(y_{ij}|b_i) = \phi\omega_{ij}^{-1}v(\mu_{ij}^b)$, where ϕ is a dispersion parameter, ω_{ij} is a known prior weight, and $v(\cdot)$ is a variance function. The SAMM assumes the conditional mean μ_{ij}^b takes the form

$$g(\mu_{ij}^b) = f(t_{ij}) + s_{ij}^T\alpha + z_{ij}^T b_i, \tag{1}$$

where $f(t)$ is an arbitrary smooth function, α is a $p \times 1$ vector of fixed effects and $g(\cdot)$ is a known link function. It is further assumed that the random effects b_i are independent and have a normal distribution $\mathrm{N}\{0, D_0(\theta)\}$, where θ is a vector of variance components.

Denote $y = (y_{11}, \dots, y_{1n_1}, \dots, y_{m1}, \dots, y_{mn_m})^T$, $D(\theta) = \mathrm{diag}(D_0, \dots, D_0)$, $b = (b_1^T, \dots, b_m^T)^T$. Suppose $f(t) \in W_2^{(h)}$, where h is a positive integer, $f^{(h)}(t)$ denotes the hth derivative of $f(t)$, and

$$W_2^{(h)} = \{g(t)|g(t), g'(t), \dots, g^{(h-1)}(t) \text{ absolutely continuous}, \int \{g^{(h)}(t)\}^2\,\mathrm{d}t < \infty\}.$$

Since the likelihood of model (1) involves integration over the random effects b and $f(\cdot)$ is a nonparametric function, following Lin and Zhang (1999), we estimate $\{f(t), \alpha\}$ given θ by maximizing the following double-penalized quasi-likelihood (DPQL) function with respect to $\{f(\cdot), \alpha, b\}$

$$\ell_{dp}\{f(.), b, \alpha, \theta; y\} = -\frac{1}{2\phi}\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}(y_{ij}, \mu_{ij}^b) - \frac{1}{2}b^T D^{-1} b - \frac{\lambda}{2}\int \{f^{(h)}(t)\}^2\,\mathrm{d}t, \tag{2}$$

where $d_{ij}(y_{ij}, \mu_{ij}^b) = -2\int_{y_{ij}}^{\mu_{ij}^b} \omega_{ij}(y_{ij} - u)/v(u)\,\mathrm{d}u$ is the conditional deviance function, and λ is a smoothing parameter that controls the goodness of fit of the model and the roughness of function $f(t)$. It can be easily shown that given (θ, λ), the estimate of $f(t)$ that maximizes the DPQL (2) is a natural smoothing spline of order h.

There are many equivalent expressions for the natural spline estimate of $f(t)$ (Cox and Koh, 1989; Green and Silverman, 1994; Wahba, 1990). For numerical reasons, we consider the smoothing spline representation given by Kimeldorf and Wahba (1971). Denote by $t^0 = (t_1^0, \dots, t_r^0)^T$ an $r \times 1$ vector of ordered distinct t_{ij} and by f the vector of $f(t)$ evaluated at t^0. Without loss of generality, assume $0 < t_1^0 < \dots < t_r^0 < 1$. The hth-order smoothing spline estimator $f(t)$ can be expressed as

$$f(t) = \sum_{k=1}^{h} \delta_k\phi_k(t) + \sum_{l=1}^{r} a_l R(t, t_l^0), \tag{3}$$

where $\{\phi_k(t)\}_{k=1}^h$ is a basis for the space of polynomials of order $h - 1$ (for example, $\phi_k(t) = t^{k-1}/(k-1)!$, $k = 1, \dots, h$) and $R(t, s)$ is defined by

$$R(t, s) = \frac{1}{[(h-1)!]^2}\int_0^1 (s-u)_+^{h-1}(t-u)_+^{h-1}\,\mathrm{d}u,$$

where $(s - u)_+ = s - u$ if $s \leqslant u$ and 0 otherwise.

Denote $\delta = (\delta_1, \dots, \delta_h)^T$ and $a = (a_1, \dots, a_r)^T$. Under this smoothing spline representation, f can be written as $f = T\delta + \Sigma a$ and the penalty in DPQL (2) has the expression

$$\frac{\lambda}{2}\int \{f^{(h)}(t)\}^2\,\mathrm{d}t = \frac{\lambda}{2}a^T\Sigma a, \tag{4}$$

where T is a $r \times h$ matrix with the (k, l)th element equal to $\phi_l(t_k^0)$, and Σ is a positive definite matrix with (k, l)th element equal to $R(t_k^0, t_l^0)$.

Let $n = \sum_{i=1}^m n_i$ and denote by N an $n \times r$ incidence matrix that maps t^0 into t_{ij}. Denote by S the matrix with the ith row block $s_i = (s_{i1}, \ldots, s_{in_i})^T$ and denote $\beta = (\delta^T, \alpha^T)^T$, $X = (NT, S)$, $B = N\Sigma$, $Z = \text{diag}(z_1, \ldots, z_m)$, where $z_i = (z_{i1}, \ldots, z_{in_i})^T$, and $\mu^b = (\mu_{11}^b, \ldots, \mu_{1n_1}^b, \ldots, \mu_{m1}^b, \ldots, \mu_{mn_m}^b)^T$. The penalty term (4) suggests that a can be treated as random effects following $\text{N}(0, \tau\Sigma^{-1})$ with $\tau = 1/\lambda$, and f as a linear combination of the fixed effects δ and the random effects a. Under this mixed-model representation of the smoothing spline estimator of f, SAMM (1) can be written as the following GLMM:

$$g(\mu^b) = X\beta + Ba + Zb, \tag{5}$$

where β is the fixed effect and $a \sim \text{N}(0, \tau\Sigma^{-1})$ and $b \sim \text{N}(0, D(\theta))$ are independent random effects.

This GLMM representation takes the same form as that Lin and Zhang (1999) used for natural cubic spline estimators. We hence adapt their approach by calculating the DPQL estimators of f and α via estimating (β, a, b) in (5) using the PQL method of Breslow and Clayton (1993), and estimating τ and θ simultaneously using the approximate maximum marginal likelihood approach by treating τ as an extra variance component. Specifically, define the working vector by $Y = X\beta + Ba + Zb + \Delta(y - \mu^b)$, where $\Delta = \text{diag}\{g'(\mu_{ij}^b)\}$, and the working weight matrix by $W = \text{diag}\{w_{ij}\}$ where $w_{ij} = \{\phi\omega_{ij}^{-1} v(\mu_{ij}^b)[g'(\mu_{ij}^b)]^2\}^{-1}$. One iteratively fits the working linear mixed model

$$Y = X\beta + Ba + Zb + \epsilon,$$

where $\epsilon \sim N(0, W^{-1})$. Then f and α are estimated by the BLUP estimators $T\hat{\delta} + \Sigma\hat{a}$ and $\hat{\alpha}$ respectively, and τ and θ are estimated by the restricted maximum likelihood (RML) estimators at convergence. For detailed justification of this estimation procedure, see Lin and Zhang (1999).

3. THE POLYNOMIAL TESTS

In this section, we propose a test for whether the nonparametric function $f(t)$ in SAMM (1) is equal to an $(h-1)$th-order polynomial for clustered Gaussian and non-Gaussian data. For example, if $h = 2$, we test the linearity of $f(t)$. This test provides a tool to check for the goodness of fit of a simple parametric GLMM against a nonparametric model. It is based on the mixed-model representation of the smoothing spline estimator of $f(t)$ given in Section 2 by testing the variance component $\tau = 0$ using the score test. A key feature of the proposed test is that it can be easily implemented by fitting a simple parametric GLMM by assuming $f(t)$ is an $(h-1)$th-order polynomial, and does not require fitting the more complicated nonparametric model (1).

3.1 *The polynomial test for gaussian responses*

We first consider the polynomial test when the response variable y is normally distributed with the identity link function. The SAMM (1) becomes

$$y_{ij} = f(t_{ij}) + s_{ij}^T\alpha + z_{ij}^T b_i + \epsilon_{ij}, \tag{6}$$

where $b_i \sim N\{0, D_0(\theta)\}$ and $\epsilon_{ij} \sim N(0, \phi)$. The following development remains valid for correlated residuals ϵ_{ij}, such as an auto-regression process. Denote by Φ the space of polynomials of order $h - 1$. We are interested in testing $H_0 : f(t) \in \Phi$ versus $H_1 : f(t) \in W_2^{(h)} - \Phi$. Using the results in Section 2, if one estimates $f(\cdot)$ by an hth-order smoothing spline, the semiparametric model (6) has a linear mixed model representation

$$y = X\beta + Ba + Zb + \epsilon, \tag{7}$$

where $a \sim \text{N}(0, \tau\Sigma^{-1})$, $b \sim \text{N}(0, D(\theta))$ and $\epsilon \sim \text{N}(0, \phi I)$.

The natural spline expression of $f(t)$ in (3) implies that $f(t)$ is an $(h-1)$th-order polynomial if and only if $a = 0$, which is equivalent to $H_0 : \tau = 0$ under linear mixed model representation (7). Since the natural spline estimate of $f(t)$ is the optimal estimate in $W_2^{(h)}$ using the penalized likelihood approach, we hence approximate our hypothesis testing problem by testing $H_0 : \tau = 0$ versus $H_1 : \tau > 0$ in linear mixed model (7). In other words, this test is equivalent to testing the null hypothesis that $f(t)$ is an $(h-1)$th-order polynomial against the alternative hypothesis that $f(t)$ is a natural spline function. An alternative motivation for testing $H_0 : \tau = 0$ versus $H_1 : \tau > 0$ is that the spline estimator of $f(t)$ can be derived from a Bayesian perspective by assuming $f(t)$ to have the following improper prior (Wahba, 1990)

$$f(t) = \sum_{k=1}^{h} \delta_k \phi_k(t) + \tau^{1/2} W(t), \tag{8}$$

where δ has a flat prior, and $W(t)$ is an $(h-1)$-fold integrated Wiener process. Obviously, $f(t)$ is an $(h-1)$th-order polynomial if and only if $\tau = 0$.

The null hypothesis $H_0 : \tau = 0$ places τ on the boundary of the parameter space. Self and Liang (1987) showed that the likelihood ratio and Wald statistics typically do not follow a chi-square distribution asymptotically. However, the score statistic often still follows a chi-square distribution asymptotically under some regularity conditions and can be easily calculated (Lin, 1997). We hence consider a score test for $H_0 : \tau = 0$ under model (7). However, unlike the variance component score test in GLMMs considered by Lin (1997), due to the special structure of the design matrix Σ, the score statistic for $H_0 : \tau = 0$ in (7) does not follow a chi-squared distribution, but follows a mixture of chi-square distributions, which is often called a chi-bar squared distribution (Robertson *et al.*, 1988).

Denote $\gamma = (\beta^T, \theta^T, \phi)^T$ and $V = ZDZ^T + \phi I$. Let $\ell(\tau, \gamma; y)$ be the log-likelihood function of model parameters (τ, γ) under linear mixed model (7). Simple calculations show that the score statistic of τ for testing $H_0 : \tau = 0$ is

$$\mathcal{U}_\tau(\hat{\gamma}) = \left.\frac{\partial \ell(\tau, \gamma; y)}{\partial \tau}\right|_{\tau=0,\gamma=\hat{\gamma}} = \frac{1}{2}\left.\left\{(y - X\beta)^T V^{-1} B\Sigma^{-1} B^T V^{-1}(y - X\beta) - \mathrm{tr}(V^{-1} B\Sigma^{-1} B^T)\right\}\right|_{\gamma=\hat{\gamma}}, \tag{9}$$

where $\hat{\gamma} = (\hat{\beta}^T, \hat{\theta}^T, \hat{\phi})^T$ is the maximum likelihood estimate (MLE) of γ under the null parametric linear mixed model

$$y = X\beta + Zb + \epsilon. \tag{10}$$

Since $B\Sigma^{-1}B^T = N\Sigma N^T$ is not a block diagonal matrix, the score $\mathcal{U}_\tau(\gamma)$ cannot be written as a sum of m independent random variables, with the ith term being a quadratic function of the ith cluster data y_i. This suggests intuitively that the distribution of the standardized score statistic of $\mathcal{U}_\tau(\hat{\gamma})$ may not converge to a standard normal distribution, in contrast to the variance component score statistic of Lin (1997). A more rigorous proof of this result is given in Appendix A. Specifically, write $\mathcal{U}_\tau(\gamma)$ as $\mathcal{U}_\tau(\gamma) = U_\tau(y; \gamma) - e(\gamma)$, where $U_\tau(y; \gamma)$ and $e(\gamma)$ denote the first and the second terms in (9) respectively. Denoting by γ_0 the true value of γ, the results in Appendix A show that under $\tau = 0$, the distribution of $U_\tau(y; \gamma_0)$ is a mixture of chi-square distributions $\sum_{i=1}^{r} \psi_i \chi_{1i}^2$, where χ_{1i}^2 are independent random variables following a chi-square distribution with one degree of freedom. Here the weights ψ_i $(i = 1, \ldots, r)$ are the ordered non-zero eigenvalues of $V^{-1}N\Sigma N^T/2$ and decay rapidly to zero.

Since calculations of the ψ_i are often computationally intensive and the exact probability associated with a mixture of chi-square distributions is difficult to calculate, we use the Satterthwaite method to approximate the distribution of $U_\tau(y; \gamma_0)$ by a scaled chi-square distribution $\kappa\chi_\nu^2$, where the scale parameter κ and the degrees of freedom ν are calculated by equating the mean and variance of $U_\tau(y;\gamma_0)$ and those of $\kappa\chi^2_\nu$. Specifically, since $U_\tau(y;\gamma_0)$ is a quadratic function of y, its mean and variance are easy to calculate and are respectively $e = \mathrm{tr}(V^{-1}N\Sigma N^T)/2$ and $I_{\tau\tau} = \mathrm{tr}((V^{-1}N\Sigma N^T)^2)/2$. Simple calculations then give $\kappa = I_{\tau\tau}/2e$ and $\nu = 2e^2/I_{\tau\tau}$.

Since γ is unknown under H_0, we estimate it using its MLE $\hat{\gamma}$ by fitting model (10). The test statistic is $S(\hat{\gamma}) = U_\tau(y;\hat{\gamma})/\kappa$, and its distribution is approximated by χ^2_ν. To account for the fact that γ is estimated by its MLE $\hat{\gamma}$, we calculate κ and ν by replacing $I_{\tau\tau}$ with the efficient information $\tilde{I}_{\tau\tau} = I_{\tau\tau} - I_{\tau\vartheta}I^{-1}_{\vartheta\vartheta}I^T_{\tau\vartheta}$, where $\vartheta = (\theta,\phi)$ and

$$I_{\tau\vartheta} = \frac{1}{2}\mathrm{tr}\left(V^{-1}N\Sigma N^T V^{-1}\frac{\partial V}{\partial\vartheta}\right), \qquad I_{\vartheta\vartheta} = \frac{1}{2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial\vartheta}V^{-1}\frac{\partial V}{\partial\vartheta}\right). \tag{11}$$

To further adjust for the small-sample bias due to estimating β, we modify the test statistic $S(\hat{\gamma})$ by estimating ϑ using REML estimate under the null linear mixed model (10), and estimating κ and ν by replacing e with $\tilde{e} = \mathrm{tr}(PN\Sigma N^T)$, where $P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$ is the projection matrix under the null model (10), and replacing $\tilde{I}_{\tau\tau}$ by $\tilde{\mathcal{I}}_{\tau\tau} = \mathcal{I}_{\tau\tau} - \mathcal{I}_{\tau\vartheta}\mathcal{I}^{-1}_{\vartheta\vartheta}\mathcal{I}^T_{\tau\vartheta}$, where $\mathcal{I}_{\vartheta\vartheta}$ and $\mathcal{I}_{\tau\vartheta}$ are the same as $I_{\tau\vartheta}$ and $I_{\vartheta\vartheta}$ in (11) except that V^{-1} is replaced by the projection matrix P. This gives us a bias-corrected version S_R of the statistic S. Simulation studies show that S_R usually outperforms S. Therefore results are given only for the bias-corrected version S_R. Note that the test here is necessarily one-sided. We study the performance of this test statistic through simulations in Section 6.

The proposed test procedure has some attractive features. One only needs to fit the null linear mixed model (10). Unlike fitting SAMM (1) using DPQL, where one often needs to invert a high-dimensional matrix, this is not necessary when calculating the test statistics S and S_R. Hence computation of S and S_R is very easy. Especially, when $X_i\alpha$ and the random effects b_i are absent from model (6), we have the classical nonparametric regression model $y_i = f(t_i) + \epsilon_i$ for independent data, where $\epsilon_i \sim N(0,\sigma^2)$, and the test statistic S reduces to the simple form given by Cox and Koh (1989).

3.2 *The polynomial test for non-Gaussian responses*

In this section, we extend the REML version polynomial test proposed in Section 3.1 to SAMM (1) for non-Gaussian response y. If $f(t)$ is estimated by an $(h-1)$th-order smoothing spline, Lin and Zhang (1999) proposed to jointly estimate the smoothing parameter τ and the variance components θ by maximizing the marginal likelihood under the GLMM representation (5) of the SAMM (1)

$$\begin{aligned} L_{\mathrm{M}}(\tau,\vartheta;y) &\propto |D|^{-1/2}\tau^{-r/2}\int \exp\left\{\sum_{i=1}^m\sum_{j=1}^{n_i} -\frac{1}{2\phi}d_{ij}(y;\mu^b_{ij}) - \frac{1}{2}b^TD^{-1}b - \frac{1}{2\tau}a^T\Sigma a\right\}\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\beta. \\ &= |D|^{-1/2}\int \exp\left\{\sum_{i=1}^m\sum_{j=1}^{n_i} -\frac{1}{2\phi}d_{ij}(y;\mu^b_{ij}) - \frac{1}{2}b^TD^{-1}b - \frac{1}{2}u^Tu\right\}\mathrm{d}u\,\mathrm{d}b\,\mathrm{d}\beta, \end{aligned} \tag{12}$$

where $u = \tau^{-1/2}\Sigma^{1/2}a \sim \mathrm{N}(0,I)$ and in matrix notation μ^b satisfies $g(\mu^b) = X\beta + Zb + \sqrt{\tau}N\Sigma^{1/2}u$. Similar to Section 3.1, we test H_0 : $f(t)$ *is an* $(h-1)$*th-order polynomial* in SAMM (1) by testing H_0 : $\tau = 0$ in GLMM (5). Let $\ell_{\mathrm{M}}(\tau,\vartheta;y) = \log\{L_{\mathrm{M}}(\tau,\vartheta;y)\}$, and denote by $L(y|\beta,b,u)$ the conditional density of y given b and u, and by $L(b)$ and $L(u)$ the marginal densities of b and u respectively. It can be easily shown that

$$\frac{\partial\ell_{\mathrm{M}}(\tau,\vartheta;y)}{\partial\tau} = \frac{1}{2\tau^{1/2}L_{\mathrm{M}}(\tau,\vartheta;y)}\int L(y|\beta,b,u)L(b)L(u)u^T\Sigma^{1/2}N^TW\Delta(y-\mu^b)\,\mathrm{d}u\,\mathrm{d}b\,\mathrm{d}\beta, \tag{13}$$

where W and Δ were defined in Section 2. Using L'Hôpital's rule, some calculations show that the score of τ evaluated at $\tau = 0$ is

$$\begin{aligned}
\mathcal{U}_\tau &= \left.\frac{\partial \ell_{\mathrm{M}}(\tau, \vartheta; y)}{\partial \tau}\right|_{\tau=0} \\
&= \frac{1}{2L_M(0, \vartheta; y)} \int L(y|\beta, b)L(b)\{(y-\mu^b)^T \Delta W N \Sigma N^T W \Delta (y-\mu^b) - \mathrm{tr}(W N \Sigma N^T)\}\, \mathrm{d}b\, \mathrm{d}\beta \\
&= \tfrac{1}{2}\mathrm{E}\{(y-\mu^b)^T \Delta W N \Sigma N^T W \Delta (y-\mu^b) - \mathrm{tr}(W N \Sigma N^T)|y\},
\end{aligned} \tag{14}$$

where μ^b satisfies the null parametric GLMM

$$g(\mu^b) = X\beta + Zb, \tag{15}$$

with $b \sim N\{0, D(\theta)\}$, W and Δ are evaluated at μ^b, and the conditional expectation is taken under the null hypothesis $\tau = 0$ and by assuming $b \sim N\{0, D(\theta)\}$ and β a flat prior. The score statistic $\mathcal{U}_\tau$ for testing $H_0 : \tau = 0$ is evaluated at $\hat{\vartheta}$, the REML estimate of ϑ under the null GLMM (15).

Since the integral in (14) generally does not have a closed form except for normal responses y and the identity link function, and sometimes is high-dimensional, we approximate (14) using the Laplace method. Specifically, we show in Appendix B that (14) can be approximated using the Laplace method by

$$\mathcal{U}_\tau \approx \mathcal{U}_\tau(\hat{\beta}, \hat{\vartheta}) = \tfrac{1}{2}\{(Y - X\beta)^T V^{-1} N \Sigma N^T V^{-1} (Y - X\beta) - \mathrm{tr}(P N \Sigma N^T)\}|_{\hat{\beta}, \hat{\vartheta}}, \tag{16}$$

where $\hat{\beta}$ is the BLUP-type estimate of β and $\hat{\vartheta}$ is the REML estimate of ϑ, and Y is the working vector $Y = X\beta + Zb + \Delta(y - \mu^b)$ under the null GLMM (15), $P = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$ and $V = W^{-1} + ZDZ^T$. One can use the existing software such as the SAS macro %GLIMMIX to obtain the estimates $\hat{\beta}$ and $\hat{\vartheta}$ by fitting (15).

Examination of equation (16) suggests that it corresponds to the score equation of τ evaluated at $\tau = 0$ under the working linear mixed model at convergence

$$Y = X\beta + Zb + Ba + \epsilon, \tag{17}$$

by assuming $b \sim N\{0, D(\theta)\}$, $a \sim N(0, \tau\Sigma^{-1})$ and $\epsilon \sim N(0, W^{-1})$. We hence can apply the results in Section 3.1 to approximate the distribution of the score statistic $\mathcal{U}_\tau$ in (16) by replacing y in Section 3.1 by the working vector Y. Specifically, write $\mathcal{U}_\tau = U_\tau - \tilde{e}$, where U_τ and $\tilde{e}$ are the first term and the second term in (16). We approximate the distribution of U_τ by a scaled chi-square distribution by matching their means and variances. The mean of U_τ can be approximated by $\tilde{e} = \mathrm{tr}(P N \Sigma N^T)/2$, and the variance by $\tilde{\mathcal{I}}_{\tau\tau}$, where P was defined in the previous paragraph and $\tilde{\mathcal{I}}_{\tau\tau}$ was defined in Section 3.1 but under the working linear mixed model (17). The test proceeds by using the test statistic $S_R = U_\tau/\kappa$, which follows χ^2_ν approximately, where κ and ν are calculated similarly to those in Section 3.1 but under the working linear mixed model (17). Note that S_R here corresponds to the bias-corrected statistic S_R in Section 3.1.

Due to the Laplace approximation of the score function (14), we here have assumed normality of the working vector Y and implicitly used a Gaussian fourth-moment assumption for the working vector Y in the approximation of the variance of U_τ. This approximation may be less than satisfactory for sparse data such as binary data. We illustrate this test procedure through an application to the infectious disease data in Section 5.1 and evaluate its performance for different kinds of responses through simulations in Section 6

4. The equivalence test of two nonparametric functions

In many applications, such as clinical trials and epidemiological studies, we are often interested in comparing the overall response profiles between two groups, for example, a treatment group and a placebo group. In this section we consider testing the equivalence of two nonparametric functions between two groups.

Suppose the response y_{kij} of group $k (k = 1, 2)$ has conditional mean $\mu^b_{kij} = \mathrm{E}(y_{kij}|b_{ki})$ that satisfies the following SAMM:

$$g(\mu^b_{kij}) = f_k(t_{kij}) + s^T_{kij}\alpha_k + z^T_{kij}b_{ki}, \tag{18}$$

where all model components have the same specification as that given in Section 2 except that they are now group specific. For example, the subject-specific random effects $b_{ki} (k = 1, 2)$ may have different distributions $\mathrm{N}\{0, D_k(\theta_k)\}$. We are interested in testing the hypothesis $H_0 : f_1(t) = f_2(t)$, where both $f_1(t)$ and $f_2(t)$ are nonparametric functions, e.g. time profiles.

Denote by $[T_1, T_2]$ the interval that specifies the range of t_{kij} for both groups. Following Zhang *et al.* (2000), define an overall measure of the difference between $f_1(t)$ and $f_2(t)$ as $\int_{T_1}^{T_2}\{f_1(t) - f_2(t)\}^2\,\mathrm{d}t$. To test $H_0 : f_1(t) = f_2(t)$, we construct the test statistic

$$G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\} = \int_{T_1}^{T_2}\{\hat{f}_1(t) - \hat{f}_2(t)\}^2\,\mathrm{d}t, \tag{19}$$

where $\hat{f}_k(t)$ is the DPQL estimate of $f_k(t)$ defined in Section 2. A large value of $G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\}$ will suggest evidence against H_0.

We now study the distribution of $G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\}$ under H_0. When y is normally distributed, Zhang *et al.* (2000) studied the distribution of G under H_0 and showed that its exact distribution is difficult to derive. Noticing G can be written as a quadratic function of y, Zhang *et al.* (2000) approximated the distribution of $G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\}$ by a scaled chi-square distribution using the moment matching technique. We now extend their results to SAMM (18) for non-Gaussian data.

Denote by Y_k the working vector of SAMM (18) defined in Section 2 for group k. The results in Section 2 show that the DPQL estimator $\hat{f}_k$ can be obtained by iteratively fitting the working linear mixed model using group k data

$$Y_k = X_k\beta_k + Z_k b_k + B_k a_k + \epsilon_k,$$

where $b_k \sim N(0, D_k)$, $a_k \sim N(0, \tau_k I)$, $\epsilon_k \sim N(0, W_k^{-1})$, X_k, Z_k, B_k, W_k are defined similarly to those in Section 2 using group k data. One can show that at convergence, $\hat{f}_k(t)$ can be written as $\hat{f}_k(t) = c_k^T(t)Y_k$ for some vector function $c_k(t)$. Let $c(t) = \{c_1^T(t), -c_2^T(t)\}^T$ and $Y_0 = \{Y_1^T, Y_2^T\}^T$, then the test statistic $G\{\hat{f}_1(.), \hat{f}_2(.)\}$ can be written as a quadratic function of the joint working vector Y_0

$$G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\} = \int_{T_1}^{T_2} Y_0^T c(t)c(t)^T Y_0\,\mathrm{d}t = Y_0^T C Y_0, \tag{20}$$

where $C = \int_{T_1}^{T_2} c(t)c(t)^T\,\mathrm{d}t$ and the integration is evaluated elementwise.

Assuming the working vector Y_k follows $N(0, V_k)$ approximately, where $V_k = Z_k D_k Z_K^T + \tau_k B_k B_k^T + W_k^{-1}$, the quadratic form in equation (20) suggests that $G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\}$ follows a mixture of chi-square distributions approximately. We hence approximate its distribution by a scaled chi-square $\kappa_*\chi^2_{\nu_*}$, where the scale parameter κ_* and the degrees of freedom ν_* are calculated by matching the approximate mean and variance of $G\{\hat{f}_1(.), \hat{f}_2(.)\}$ under H_0 and those of $\kappa_*\chi^2_{\nu_*}$. Denote by E_0 and V_0 the approximate mean and variance of Y_0 under H_0, then the approximate mean e_* and variance ψ_* of $G\{\hat{f}_1(.), \hat{f}_2(.)\}$ under H_0 can be calculated as

$$e_* = E_0^T C E_0 + \mathrm{tr}(CV_0), \qquad \psi_* = 2\mathrm{tr}(CV_0)^2 + 4E_0^T C V_0 C E_0,$$

where the unknown parameters are replaced by their DPQL estimates obtained under H_0. Since CE_0 is negligible under H_0, e_* and ψ_* can be further approximated by $e_* \approx \mathrm{tr}(CV_0)$ and $\psi_* \approx 2\mathrm{tr}(CV_0)^2$. Matching these mean and variance with those of $\kappa_* \chi^2_{\nu_*}$ gives the estimates of κ_* and ν_* as $\kappa_* = \psi_*/2e_*$ and $\nu_* = 2e_*^2/\psi_*$. Define $\chi^2_{\mathrm{obs}} = G_{\mathrm{obs}}/\kappa_*$, where G_{obs} denotes the observed value of G. Then the approximate p-value for the test statistic $G\{\hat{f}_1(\cdot), \hat{f}_2(\cdot)\}$ is given by $P[\chi^2_{\nu_*} > \chi^2_{\mathrm{obs}}]$. We evaluate the performance of this test through simulations in Section 6.

5. APPLICATIONS

5.1 *Application of the polynomial test to the infectious disease data*

Lin and Zhang (1999) applied SAMM (1) to analyse longitudinal data on respiratory infection in 275 Indonesian children. The children were examined every 3 months up to six consecutive visits for respiratory infection (0=no, 1=yes). The covariates of interest included age in years, xerophthalmia status (0=no, 1=yes), an eye condition of chronic vitamin A deficiency, sex (0=male, 1=female), height for age, seasonality and the presence of stunting (0=no, 1=yes). For a more detailed description of this data set, see Lin and Zhang (1999) and the reference cited therein.

Denote by y_{ij} the jth binary indicator of respiratory infection for the ith child, age_{ij} his/her age and s_{ij} the other covariates. Examination of the data suggested a strong non-linear age effect. Lin and Zhang (1999) hence fit the following semiparametric logistic mixed model to model the age effect nonparametrically

$$\mathrm{logit}\{\mathrm{pr}(y_{ij}|b_i)\} = s_{ij}^T\alpha + f(\mathrm{age}_{ij}) + b_i, \tag{21}$$

where the random intercept $b_i \sim \mathrm{N}(0, \theta)$, s_{ij} is a vector of the other covariates.

Figure 1 shows the estimated function $\hat{f}(\mathrm{age})$ and its 95% confidence interval. It suggests that the risk of respiratory infection increased during the first two years of life and then steadily decreased afterwards. An immediate question of interest was whether or not a quadratic function of age fits the data adequately. In other words, we were interested in testing $H_0 : f(\mathrm{age})$ *is a quadratic function* vs $H_1 : f(\mathrm{age})$ *is a smooth non-quadratic function*. We applied the polynomial test proposed in Section 3.2 to the data. The test statistic was $S_R = 5.73$ with 1.30 degrees of freedom, providing strong evidence (p-value=0.026) against the null hypothesis that $f(\mathrm{age})$ is quadratic in age.

Table 1 contrasts the estimates of the covariate effects and the variance components when the age effect is modeled quadratically and nonparametrically. The parameter estimates were similar, except that the regression coefficient estimate of the covariate *Stunted* was attenuated and the estimate of the variance component was inflated when a quadratic age model was assumed. This suggests that the covariate effects and the variance component were somewhat sensitive to the misspecification of the functional form of the age effect, and nonparametric modeling of $f(\mathrm{age})$ would be more preferable.

5.2 *Application of the equivalence test to the epileptic data*

Thall and Vail (1990) presented data from a clinical trial of 59 epileptics randomized to receive either the anti-epileptic drug progabide or a placebo, as an adjutant to standard chemotherapy. The response variable was the number of epileptic seizures in four two-week intervals. The covariates of interest included the epileptic seizure counts during a baseline period of eight weeks, logarithm of age (in years) at baseline, visit number (coded as $visit = -3, -1, 1, 3$) and a treatment indicator (0=placebo, 1=progabide). For a more detailed description of the study, see Thall and Vail (1990). Breslow and Clayton (1993) analysed this data set using GLMMs assuming random intercepts and random slopes. Their analysis indicated

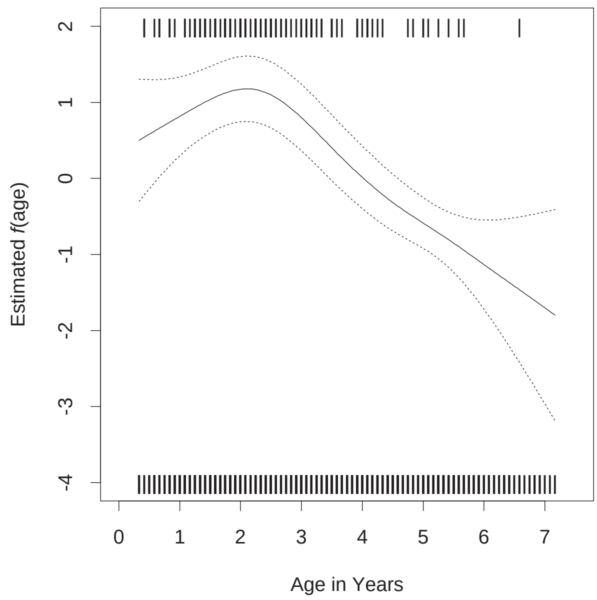

Fig. 1. Estimated $\hat{f}(age)$ (——) and its 95% pointwise confidence intervals (- - - -) under model (21) for the infectious disease data: the vertical strokes at 2 and -4 indicate the occurrence of 1s and 0s in the response.

Table 1. *Comparison of parameter estimates for the infectious disease data assuming the age effect to be quadratic and non-parametric*

	Quadratic age model		Nonparametric age model	
Covariate	Estimate	SE	Estimate	SE
Intercept	-2.12	0.22	-2.92	0.24
Vitamin A	0.53	0.46	0.52	0.46
Seasonal cosine	-0.58	0.17	-0.58	0.17
Seasonal sine	-0.16	0.17	-0.16	0.17
Sex	-0.48	0.24	-0.50	0.24
Height for age	-0.03	0.03	-0.03	0.02
Stunted	0.27	0.42	0.39	0.43
Age	-0.30	0.08		
Age square	-0.11	0.04		
θ (DPQL)	0.40	0.26	0.38	0.26
θ (CDPQL)	0.52	0.34	0.48	0.33

CDPQL: Corrected DPQL estimate (See Lin and Zhang, 1999)

some nonlinearity of the baseline seizure count effect and some degree of interaction between treatment and the baseline seizure count. We applied the equivalence test developed in Section 4 to investigate this interaction.

Denote by $k = 1$ the treatment group and $k = 0$ the placebo group. Let y_{ijk} be the seizure count for the ith subject on the jth visit in group k, t_{ik} be his/her baseline seizure count and s_{ijk} be the covariate vector consisting of centered logarithm of age at baseline and $visit/10$. For group k, conditional on the subject specific random effects $b_{ik} = (b_{0ik}, b_{1ik})^T \sim N(0, D_k)$, y_{ijk} was assumed to have a Poisson distribution

Table 2. *Estimates of the parameters in model (22) fitted to the epilepsy data*

	Placebo group		Treatment group	
Variable	Estimate	SE	Estimate	SE
Fixed effects				
Age	0.37	0.43	0.71	0.55
Visit/10	−0.20	0.26	−0.40	0.17
Random effects				
Intercept	0.17	0.02	0.38	0.12
Visit/10	1.04	0.50	0.10	0.17
Covariance	−0.06	0.13	0.15	0.13

with the conditional mean $\mu^b_{ijk} = \mathrm{E}(y_{ijk}|b_k)$ satisfying the SAMM

$$\log(\mu^b_{ijk}) = f_k(t_{ik}) + s^T_{ijk}\alpha_k + b_{0ik} + b_{1ik} \cdot visit_k/10, \tag{22}$$

where $f_k(\cdot)$ is a smooth function and α_k is the covariate effect. This model is similar to model IV in Table 4 of Breslow and Clayton (1993) except they modeled the baseline seizure effect linearly and the random effects were assumed to have the same distribution.

Table 2 presents the parameter estimates of age and time under model (22). The results showed that both the treatment and the placebo reduced the change rate of the number of seizures over time. However, the seizure reduction rate over time for the placebo group was statistically not significant, while it was significant in the treatment group. This result was different from that in Breslow and Clayton (1993) who assumed a common linear time effect between the two groups. Older patients had a higher number of seizures and the effect of age was stronger in the treatment group, although the age effect was not significant in both groups.

Figure 2(a) presents the DPQL estimates $\hat{f}_0(t)$ and $\hat{f}_1(t)$ of the baseline seizure effects in the two treatment groups, while Figure 2(b) gives the difference of the two curves and its 95% pointwise confidence intervals. Breslow and Clayton (1993) assumed a linear baseline seizure effect. Figure 2(a) seems to suggest a linear baseline seizure effect in the placebo arm, but a nonlinear baseline seizure effect in the treatment arm. In both groups, patients with a higher number of baseline seizures were likely to have a higher number of seizures in subsequent weeks.

Figures 2(a) and (b) show that, compared with the placebo, the drug progabide seemed to reduce the number of seizures for those patients whose baseline seizure count was less than 80, but to increase the number of seizures for those patients whose baseline seizure count was greater than 80. We applied the equivalence test to test the hypothesis $H_0 : f_0(t) = f_1(t)$, i.e. whether the effect of the baseline seizure count was the same in the two treatment arms. The test statistic was 1.55 with 1.52 degrees of freedom, which provided no strong evidence for different effects of the baseline seizure count between the placebo and the treatment arms.

6. Simulation studies

6.1 *Simulation study for the polynomial test*

We conducted a simulation study to evaluate the performance of the polynomial test proposed in Section 3 for clustered data with different types of responses and different magnitudes of correlation. Each dataset

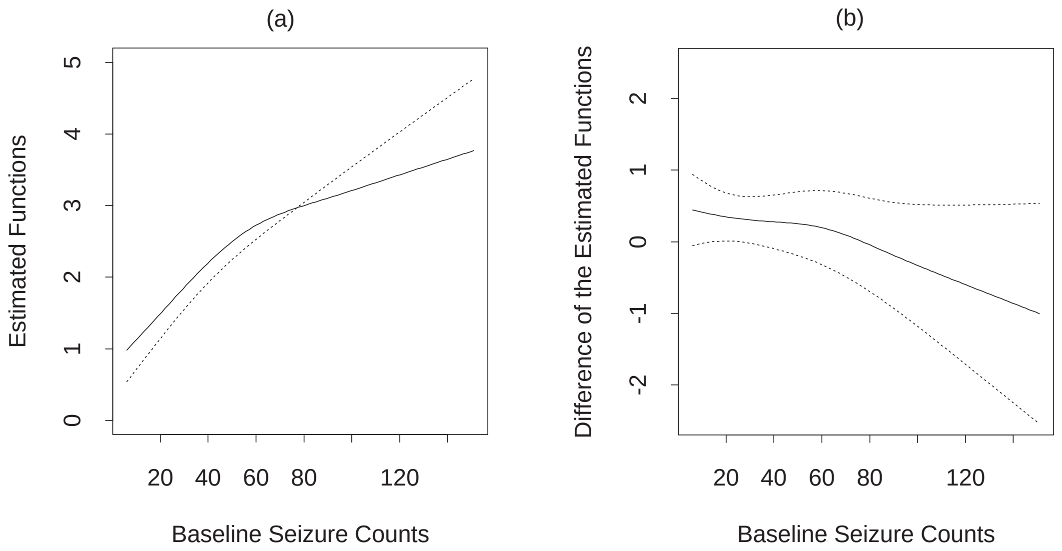

Fig. 2. (a) Estimated $\hat{f}_0(t)$ (——) for the placebo group and $\hat{f}_1(t)$ (- - - -) for the treatment group under model (22) for the epileptic data. (b) Difference (——) of $\hat{f}_0(t)$ and $\hat{f}_1(t)$ and its 95% pointwise confidence intervals (- - - -).

was composed of 100 clusters of size $n_i = 5$. Conditional on the cluster-specific random intercept $b_i \sim N(0, \theta)$ with $\theta = 0.5$ and 1, independent Gaussian, binary and binomial responses (with denominator $N = 8$) y_{ij} were generated respectively under the model

$$g\{\mathrm{E}(y_{ij}|b_i)\} = f_d(t_{ij}) + b_i,$$

where $g(\mu) = \mu$ for Gaussian response, and $g(\mu) = \mathrm{logit}(\mu)$ for binary and binomial responses. The scale parameter ϕ was estimated for Gaussian responses and was set to be one for binary and binomial responses. One hundred equally spaced points were generated for t in $[0, 2]$ as follows: $t_{ij} = [\mathrm{trun}\{(i + 4)/5\}/50] + 0.40(j - 1)$ for $i = 1, \ldots, 100$ and $j = 1, \ldots, 5$, where trun(.) denotes a truncation operator. We assumed the functions $f_d(t) = (0.25d)t \cdot \exp(2-2t) - t + 0.5$ ($d = 0, 1, 2, 3, 4$). We considered the linearity test in our simulations, i.e. H_0 : $f_d(t)$ is a linear function in t. Therefore, $d = 0$ corresponds to linearity, and as d increases, $f_d(t)$ becomes further away from being linear. The functions $f_d(t)$ are plotted in Figure 3.

Table 3 presents the empirical size and power of the linearity test (bias-corrected version) based on 2000 simulation runs for $d = 0$ and 1000 runs for $d = 1, 2, 3, 4$. We carried out more runs for the size calculations to ensure more accurate estimation of the empirical size of the test, since the nominal size was often set to be small. We set the nominal size to be 0.05. The results showed that the linearity test performed very well for Gaussian responses for different magnitudes of the variance component. The empirical sizes were very close to the nominal size and the powers of the test were high, and was not significantly affected by the magnitude of the variance component. The test was a little conservative and not very powerful for binary responses. However, when the binomial denominator increased to 8, the size of the test quickly approached the nominal value and the test became very powerful to detect nonlinearity for both values of the variance component. The test became slightly less powerful when the variance component became larger.

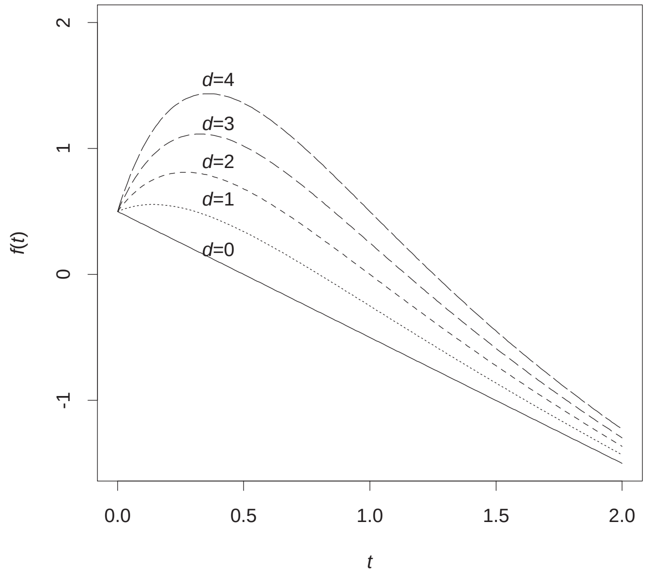

Fig. 3. Functions $f_d(t)$ ($d = 0, 1, 2, 3, 4$) used in the simulation studies for the polynomial test in Section 6.1.

Table 3. *Empirical sizes and powers of the linearity test for three types of data based on 1000 simulation runs*

Variance of random effects	Data type	Size* $d = 0$	Power $d = 1$	$d = 2$	$d = 3$	$d = 4$
$\theta = 0.5$	Gaussian	0.051	0.184	0.618	0.945	0.996
	Binary ($N = 1$)	0.043	0.072	0.152	0.276	0.424
	Binomial ($N = 8$)	0.054	0.288	0.816	0.996	1.000
$\theta = 1$	Gaussian	0.049	0.205	0.599	0.922	1.000
	Binary ($N = 1$)	0.042	0.067	0.127	0.263	0.421
	Binomial ($N = 8$)	0.047	0.289	0.784	0.984	1.000

*Simulation on size was based on 2000 runs.

6.2 *Simulation studies for the equivalence test*

We next conducted a simulation study to evaluate the performance of the equivalence test proposed in Section 4. The design of the simulation study was the same as that in Section 6.1. Since the performance of the test for Gaussian responses was reported in a simulation study by Zhang *et al.* (2000), we here considered binary and binomial responses (with the denominator $N = 8$) y_{ij}, which were generated for group $k = 1, 2$ under the model

$$\text{logit}\{\mathrm{E}(y_{ij}|b_{ik})\} = f_{kd}(t_{ij}) + b_{ik},$$

where $b_{i1} \sim N(0, \theta_1)$, $b_{i2} \sim N(0, \theta_2)$, two configurations of the variance components were considered $(\theta_1, \theta_2) = (0.5, 0.4)$ and $(1.0, 0.8)$, and $f_{kd}(t) = \frac{d}{4} f_k(t) + (1 - \frac{d}{4}) f(t)$ ($d = 0, 1, 2, 3, 4$; $k = 1, 2$), and $f_1(t) = t \cdot \exp(2 - 2t) - t + 0.5$, $f_2(t) = t \cdot \exp(2 - 2t) - 0.5$ and $f(t) = (f_1(t) + f_2(t))/2$, $t \in [0, 2]$. Therefore, $d = 0$ corresponds to the situation where the functions in the two groups are equal, and as d increases, they differ further more. These functions are plotted in Figures 4(a) and (b).

Table 4. *Empirical sizes and powers for the equivalence test for binary and binomial data based on 1000 simulation runs*

Variance of random effects	Data type	Size* $d = 0$	Power $d = 1$	$d = 2$	$d = 3$	$d = 4$
$\theta_1 = 0.5, \theta_2 = 0.4$	Binary ($N = 1$)	0.061	0.121	0.362	0.710	0.938
	Binomial ($N = 8$)	0.047	0.249	0.958	1.000	1.000
$\theta_1 = 1.0, \theta_2 = 0.8$	Binary ($N = 1$)	0.041	0.092	0.260	0.556	0.851
	Binomial ($N = 8$)	0.050	0.119	0.706	1.000	1.000

*Simulation on size was based on 2000 runs.

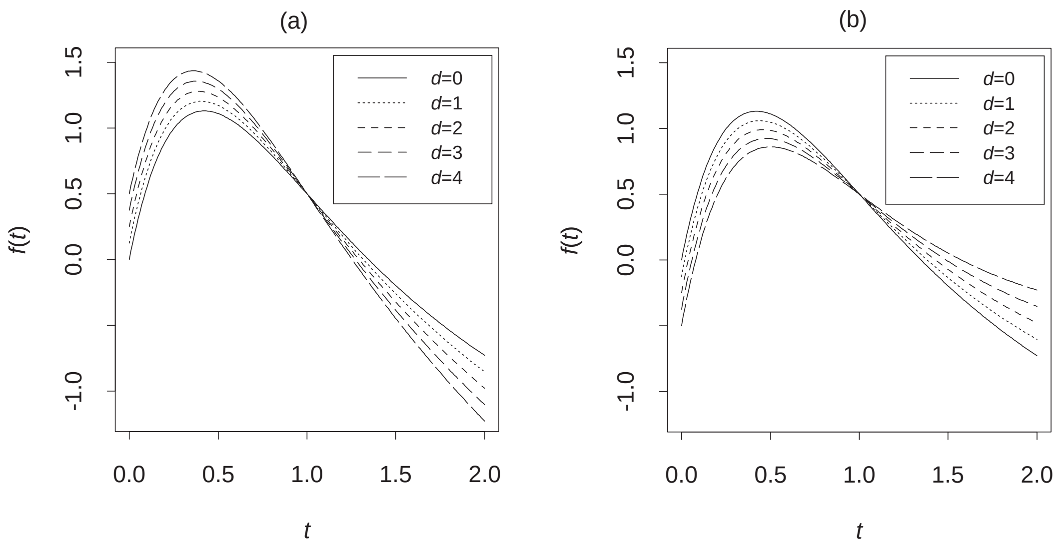

Fig. 4. (a) Functions $f_{1d}(t)$ ($d = 0, 1, 2, 3, 4$) used in the simulation studies for the equivalence test in Section 6.2. (b) Functions $f_{2d}(t)$ ($d = 0, 1, 2, 3, 4$) used in the simulation studies for the equivalence test in Section 6.2.

Table 4 presents the empirical size based on 2000 simulation runs for $d = 0$ and the empirical power based on 1000 runs for $d = 1, 2, 3, 4$. The nominal size of the test was 0.05. For binary data, the size of the test differed slightly from the nominal size for both sets of the variance components. The size was slightly anti-conservative when the variance component was smaller, while slightly conservative when the variance component was larger. This might be due to the fact that the DPQL estimates of the two nonparametric functions for binary data are biased (Lin and Zhang, 1999) and their biases may not cancel out under $H_0 : f_1(t) = f_2(t)$. The test was quite powerful to detect the difference in two nonparametric functions. When the binomial denominator N increased to 8, the size quickly approached the nominal value and the power was very high for both sets of variance components. For both binary and binomial data, the power was slightly affected by the variance component, and became a little higher when the variance component was smaller.

7. DISCUSSION

We have developed in this paper a test procedure for testing a parametric mixed model against a SAMM by testing whether the nonparametric function is some fixed-degree polynomial. The key idea is based on the mixed-effect representation of the natural spline estimator of the nonparametric function and that the inverse of the smoothing parameter can be treated as a variance component. We hence represent the SAMM as a working GLMM and proceed the test by a variance component score test. Unlike conventional variance component score tests, due to the special structure of the smoothing matrix, this polynomial test does not have a chi-square distribution asymptotically. We hence approximate its distribution by a scaled chi-square distribution. For non-Gaussian outcomes, the Laplace approximation is used when developing the test statistic to avoid possibly high-dimensional integration. Simulation studies show that for Gaussian outcomes the test performs very well in terms of size and power. For sparse data such as binary data, the performance of the test is less satisfactory. This is mainly due to the less satisfactory performance of the Laplace approximation for the score statistic and the implicit Gaussian fourth-moment assumption when estimating the variance of the score statistic. As the binomial denominator increases, its performance in terms of size and power quickly improves. The performance is slightly affected by the magnitude of the variance component.

We have used the PQL approach of Breslow and Clayton (1993) to obtain approximate REML estimates of the parameters under the null GLMM and used them to calculate the score test statistic (16) for the polynomial test. These estimates can be obtained using the existing software such as SAS macro %GLIMMIX. When the dimension of the random effects is manageable, one can also obtain the exact MLE of those parameters to remove one source of the bias in the score statistic using the existing software such as the SAS procedure NLMIXED. It is also of substantial interest in future research to calculate the exact score statistic (14) by numerically evaluating the required integral, for example, using adaptive Gaussian quadrature or Monte Carlo simulation methods.

We have also proposed in this paper an equivalence test for testing whether nonparametric functions are the same in two groups for correlated non-Gaussian data. This test extends the previous work of Zhang *et al.* (2000) for correlated Gaussian data to correlated non-Gaussian data. Our simulation results show that the proposed test perform reasonably well even for correlated binary data. As the binomial denominator increases, its performance rapidly improves. The performance of the test is slightly affected by the magnitude of the variance component.

The proposed test can be easily extended to test for the equivalence of nonparametric functions for more than two groups, as discussed in Zhang *et al.* (2000). It is of future research interest to extend other test procedures of the equivalence of nonparametric functions, such as the bootstrap test (Härdle and Marron, 1990) and the adaptive Neyman test (Fan and Lin, 1998), to SAMMs.

We have used the simple scaled chi-square distribution to approximate the distribution of the proposed test statistics. This approximation procedure can be improved and made a little more flexible by approximating their distribution by that of $\xi + \kappa\chi^2_\nu$, where constants ξ, κ and ν are obtained by moment matching technique as described in the paper.

Both the polynomial test and the equivalence test have been implemented in SAS macros. They are available from the authors upon request.

ACKNOWLEDGEMENT

This work was supported in part by grants from the U.S. National Cancer Institute.

APPENDIX A

Distribution of the Score Statistic for the Linearity Test for Gaussian Data

To study the asymptotic distribution of the score statistic $\mathcal{U}_\tau(\hat{\gamma})$ under $\tau = 0$, it is sufficient to study the asymptotic distribution $\mathcal{U}_\tau(\gamma_0)$ under $\tau = 0$, where $\gamma_0 = (\beta_0^T, \theta_0^T, \phi_0)^T$ denotes the true value of γ.

This is because under $\tau = 0$ and standard regularity conditions, the MLE $\hat{\gamma}$ is a $\sqrt{m}$ consistent estimator of γ_0.

In what follows, we assume that the null hypothesis $H_0 : \tau = 0$ holds, which implies that the null linear mixed model (10) is true. Let $M = (1/2)V^{-1}N\Sigma N^T V^{-1}$. Then we can write $\mathcal{U}_\tau(\gamma_0)$ as

$$\begin{aligned}\mathcal{U}_\tau(\gamma_0) &= (y - X\beta_0)^T M(y - X\beta_0) - \mathrm{tr}(V^{\frac{1}{2}} M V^{\frac{1}{2}})\\ &= \tilde{y}^T V^{\frac{1}{2}} M V^{\frac{1}{2}} \tilde{y} - \mathrm{tr}(V^{\frac{1}{2}} M V^{\frac{1}{2}})\end{aligned}$$

where $\tilde{y} = V^{-\frac{1}{2}}(y - X\beta_0)$ is distributed as N(0, I).

Let $\psi_1 \geqslant \cdots \geqslant \psi_r > 0$ be the ordered non-zero eigenvalues of $V^{\frac{1}{2}} M V^{\frac{1}{2}}$ and $\Psi = \mathrm{diag}(\psi_i)$. Let H be a $r \times n$ matrix consisting of the corresponding eigenvectors of ψ_i such that $HH^T = I$. We then have

$$\mathcal{U}_\tau(\gamma_0) = \tilde{y}^T H^T \Psi H \tilde{y} - \mathrm{tr}(\Psi) = \sum_{i=1}^{r} \psi_i (z_i^2 - 1)$$

check
lling.
or

where $z = (z_1, \ldots, z_c)^T = H\tilde{y}$ and z_i are independent random variables following a standard normal distribution.

Demmler and Reisch (1975) and Hastie and Tibshirani (1990)showed for cubic smoothing splines ($h = 2$) the eigenvalues of the smoothing matrix are dominated by the first few large ones and decrease rapidly to zero. Similar results hold for the eigenvalues of Σ. Since $V^{\frac{1}{2}} M V^{\frac{1}{2}} = V^{\frac{1}{2}} N \Sigma N^T V^{\frac{1}{2}}$, it follows that the eigenvalues ψ_i would have a similar behavior. Therefore the ratio

$$\frac{\max_{1 \leqslant i \leqslant r}\{\mathrm{var}[\psi_i(z_i^2 - 1)]\}}{\sum_{i=1}^{r} \mathrm{var}[\psi_i(z_i^2 - 1)]} = \frac{\psi_1^2}{\sum_{i=1}^{r} \psi_i^2}$$

would not go to zero as r, the number of distinct values of t_{ij} goes to infinity. It follows that the Lindeburg condition, which is a necessary and sufficient condition for asymptotic normality, fails. As a consequence, the standardized score statistic of $\mathcal{U}_\tau(\gamma_0)$ is not asymptotically normal when $m, r \to \infty$ (Serfling, 1980, Section 1.9).

APPENDIX B

Derivation of equation (16)

In what follows we assume that the null hypothesis $\tau = 0$ holds. Denote

$$\mathcal{U}_\tau^b = (y - \mu^b)^T \Delta W N \Sigma N^T W \Delta (y - \mu^b) - \mathrm{tr}(W N \Sigma N^T).$$

We can write $\mathcal{U}_\tau$ in (14) as

$$\mathcal{U}_\tau = \frac{1}{2} \frac{\int\int \mathcal{U}_\tau^b \exp\{\ell(y|\beta, b) + \ell(b)\}\, \mathrm{d}b\, \mathrm{d}\beta}{\int\int \exp\{\ell(y|\beta, b) + \ell(b)\}\, \mathrm{d}b\, \mathrm{d}\beta}. \tag{B.1}$$

Given θ, let $(\hat{\beta}, \hat{b})$ be the maximizer of $-\kappa(\beta, b) = \ell(y|\beta, b) + \ell(b)$.

We apply the Laplace method to both the numerator and the denominator of (B.1) by approximating $-\kappa(\beta, b)$ by a quadratic expansion about $(\hat{\beta}, \hat{b})$

$$-\kappa(\beta, b) \approx -\kappa(\hat{\beta}, \hat{b}) - \frac{1}{2}\{(\beta - \hat{\beta})^T, (b - \hat{b})^T\} Q \left\{ \begin{array}{c} \beta - \hat{\beta} \\ b - \hat{b} \end{array} \right\}, \tag{B.2}$$

where

$$\kappa''(\hat\beta, \hat b) \approx Q = \begin{pmatrix} X^T W X & X^T W Z \\ Z^T W X & Z^T W Z + D^{-1} \end{pmatrix}.$$

Denote by $Y = X\beta + Zb + \Delta(y - \mu^b)$ the working vector under the null GLMM (15). It follows that (B.1) can be approximated by

$$\mathcal{U}_\tau \approx \tfrac{1}{2}\mathrm{E}\{(Y - X\beta - Zb)^T W N \Sigma N^T W (Y - X\beta - Zb) - \mathrm{tr}(W N \Sigma N^T)\}, \qquad \text{(B.3)}$$

where the expectation is taken with respect to $(\beta, b) \sim N\{(\hat\beta, \hat b), Q^{-1}\}$. Further approximating Y and W by their values at $(\hat\beta, \hat b)$, some calculations show that (B.3) becomes

$$\begin{aligned}\mathcal{U}_\tau &\approx \tfrac{1}{2}\{(Y - X\hat\beta - Z\hat b)^T W N \Sigma N^T W (Y - X\hat\beta - Z\hat b) - \mathrm{tr}[\{W - W(X, Z) Q^{-1}(X, Z)^T W\} N \Sigma N^T]\} \\ &= \tfrac{1}{2}\{(Y - X\hat\beta)^T V^{-1} N \Sigma N^T V^{-1}(Y - X\hat\beta) - \mathrm{tr}(P N \Sigma N^T)\},\end{aligned}$$

where $V = ZDZ^T + W^{-1}$ and $P = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$.

References

AZZALINI, A. AND BOWMAN, A. (1993). On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society,* B **55**, 549–557.

BRESLOW, N. E. AND CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

COX, D. D. AND KOH, E. (1989). A smoothing spline based test of model adequacy in polynomial regression. *Annals of Ins. of Stat. Math.* **41**, 383–400.

DEMMLER, A. AND REISCH, C. (1975). Oscillation matrices with spline smoothing. *Numerische Mathematik* **24**, 375–382.

DIGGLE, P. J., LIANG, K-Y AND ZEGER, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

FAN, J. AND LIN, S. (1998). Test of significance when data are curves. *Journal of the American Statistical Association* **93**, 1007–1021.

FAN, J. AND HUANG, L. S. (2001). Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association* **96**, 640–652.

EUBANK, R. L. AND HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics* **20**, 1412–1425.

GREEN, P. J. AND SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

HALL, P. AND HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association* **85**, 1039–1049.

HÄRDLE, W. AND MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics* **18**, 63–89.

HÄRDLE, W., MAMMEN, E. AND MÜLLER, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association* **93**, 1461–1474.

HASTIE, T. J. AND TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

KIMELDORF, G. AND WAHBA, G. (1971). Some results on Tachebycheffian spline functions. *Journal of Mathematical Analysis and Applications* **33**, 82–95.

LAIRD, N. M. AND WARE, J. H. (1986). Random-effects models for longitudinal data. *Biometrics* **38**, 963–964.

LIN, X. (1997). Variance components testing in generalized linear models with random effects. *Biometrika* **84**, 309–326.

LIN, X. AND ZHANG, D. (1999). Inference in generalized additive mixed models using smoothing splines. *Journal of the Royal Statistical Society*, B **61**, 381–400.

ROBERTSON, T., WRIGHT, F. T. AND DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

SELF, S. AND LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New york: Wiley.

THALL, P. F. AND VAIL, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.

WAHBA, G. (1990). *Spline Methods for Observational Data*, Society for Industrial and Applied Mathematics. Philadelphia, PA.

YOUNG, S. G. AND BOWMAN, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* **51**, 920–931.

ZHANG, D., LIN, X. AND SOWERS, M. F. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics* **56**, 31–39.

[*Received July 27, 2001*; *revised January 7, 2002*; *accepted for publication January 14, 2002*]