Testing Generalized Linear Models Using Smoothing Spline Methods

Anna Liu, Wendy Meiring and Yuedong Wang

University of California, Santa Barbara

Abstract

This article considers testing the hypothesis of Generalized Linear Models (GLM) versus general smoothing spline models for data from exponential families. The tests developed are based on the connection between the smoothing spline models and the Bayesian models (Gu 1992). They are extensions of the locally most powerful (LMP) test by Cox, Koh, Wahba & Yandell (1988), the generalized maximum likelihood ratio (GML) test and the generalized cross validation (GCV) test by Wahba (1990) for Gaussian data. Null distribution approximations are considered and simulations are done to evaluate these approximations and the performance of the tests. Simulations show that the LMP and GML tests are more powerful for low frequency functions while the GCV test is more powerful for high frequency functions, which is also true for Gaussian data (Liu & Wang 2002). The tests are applied to data from the Wisconsin Epidemiology Study of Diabetic Retinopathy, the results of which confirm and provide more definite analysis than those of previous studies. The good performances of the tests make them useful tools for diagnosis of GLM.

Key words and phrases: Diagnosis, generalized cross validation, generalized maximum likelihood, locally most powerful test, hypothesis test, reproducing kernel hilbert space.

1 Introduction

Generalized Linear Models (GLM) are widely used to model responses with distributions from exponential families (McCullagh & Nelder 1989). These models assume that the mean of the response variable depends on covariates parametrically. Sometimes these parametric models are too restrictive and the parametric forms need to be verified to avoid misleading results. The usual scatter and residual plots are less informative for discrete data. Therefore general diagnostic and model building tools are especially important in practice. In this paper, we propose and compare three methods for testing GLM versus a general nonparametric alternative.

We use smoothing splines to model the alternative. For Gaussian data, Cox et al. (1988) showed that there is no uniformly most powerful (UMP) test, and they proposed a locally most powerful (LMP) test. Wahba (1990, Chapter 6) proposed two tests based on the generalized maximum likelihood (GML) and the generalized cross validation (GCV) scores. Simulations indicate that the LMP and GML tests are more powerful for detecting departure in the form of low frequency functions and the GCV test is more powerful for high frequency functions (Liu & Wang 2002). For data from exponential families, Zhang & Lin (2003) developed the score test under the semiparametric additive mixed model. When the data are Gaussian, the score test is exactly the same as the LMP test. Xiang & Wahba (1995) developed the symmetrized Kullback-Leibler (SKL) test based on the SKL distance between the functions estimated under the null and the alternative hypothesis. The SKL test has quite different characteristics from the GML, GCV and LMP tests (Liu & Wang 2002). It is not considered further in this paper.

In this paper, we extend the GML, GCV and LMP tests to data from exponential families. The extension of the LMP test turns out to be a special case of the score test by Zhang & Lin (2003), while the GML and GCV tests are new for data from exponential families. Simulations show that the comparative behaviors of the tests for Gaussian data (Liu & Wang 2002) remain true for data from exponential families. A motivating example is given in section 2. In section 3, a brief introduction to smoothing spline models for exponential families is given. In section 4, we present our tests. Simulations and the application of the tests to the motivating example are given in sections 5 and 6.

2 A motivating example

The Wisconsin Epidemiology Study of Diabetic Retinopathy (WESDR) data contains a sample of 2990 diabetic patients selected from an 11-county area in southern Wisconsin (Klein, Klein, Moss, Davis & DeMets 1988, Klein, Klein, Moss, Davis & DeMets 1989). We are interested in the younger onset group which consists of 256 insulin-dependent patients diagnosed as having diabetes before 30 years of age. None of the patients had diabetic retinopathy at the baseline. At the follow-up examination, all 256 patients were checked to see whether they had diabetic retinopathy. The response $y = 1$ if an individual had diabetic retinopathy at the follow-up and $y = 0$ otherwise. Several covariates were recorded. We only list the variables pertinent to our analysis: x_1, age in years at the time of baseline examination; x_2, duration of diabetes at the time of baseline examination; x_3, a measure of hyperglycaemia; x_4, systolic blood pressure in millimeters of mercury.

Wang (1994) considered five models including a GLM model (model II in Wang, 1994):

$$logit\{P(y=1|x_1,x_2,x_3,x_4)\} = \alpha_0 + \alpha_1 x_1 + \alpha_2 {x_1}^2 + \beta_1 x_2 + \beta_2 x_3 + \beta_3 x_4, \tag{1}$$

an additive model (model IV in Wang, 1994):

$$logit\{P(y=1|x_1,x_2,x_3,x_4)\} = f(x_1) + \beta_1 x_2 + \beta_2 x_3 + \beta_3 x_4, \tag{2}$$

where f is a smooth function. Note that the intercept in model (2) is absorbed into f. Based on leave-out-one cross validation estimate of Kullback-Leibler discrepancy and mean squared error, Wang (1994) concluded that model (2) gives the best prediction and model (1) gives almost as good prediction as model (2). It is then natural to ask if $f(x_1)$ in (2) is significantly better than the more parsimonious quadratic function in (1) from a hypothesis testing point of view. Figure 1 shows the estimate of f in (2) using a cubic smoothing spline and the corresponding 95% Bayesian confidence interval. For comparison, we also plot the constant, estimate of α_0 in model (1) with $\alpha_1 = \alpha_2 = 0$, and the quadratic age effect, $\hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_1^2$, from model (1). Bayesian confidence intervals are often used for inference and model building in the smoothing spline literature. However, it is well-known that the coverage of these Bayesian confidence intervals is neither simultaneous nor pointwise (Wahba 1990, Wang & Wahba 1995). Therefore, they usually do not provide a definite conclusion. In Figure 1, the 95% confidence interval contains the quadratic effect function which seems to suggest that f is not significantly different from a quadratic function. Furthermore, the 95% confidence interval contains the constant which seems to suggest that the age effect is not significant, a conclusion made in Klein et al. (1988) based on fitting GLM models with a linear age effect only. Assuming the other three covariates have linear effects, we will examine the form of the age effect with the tests developed in this paper.

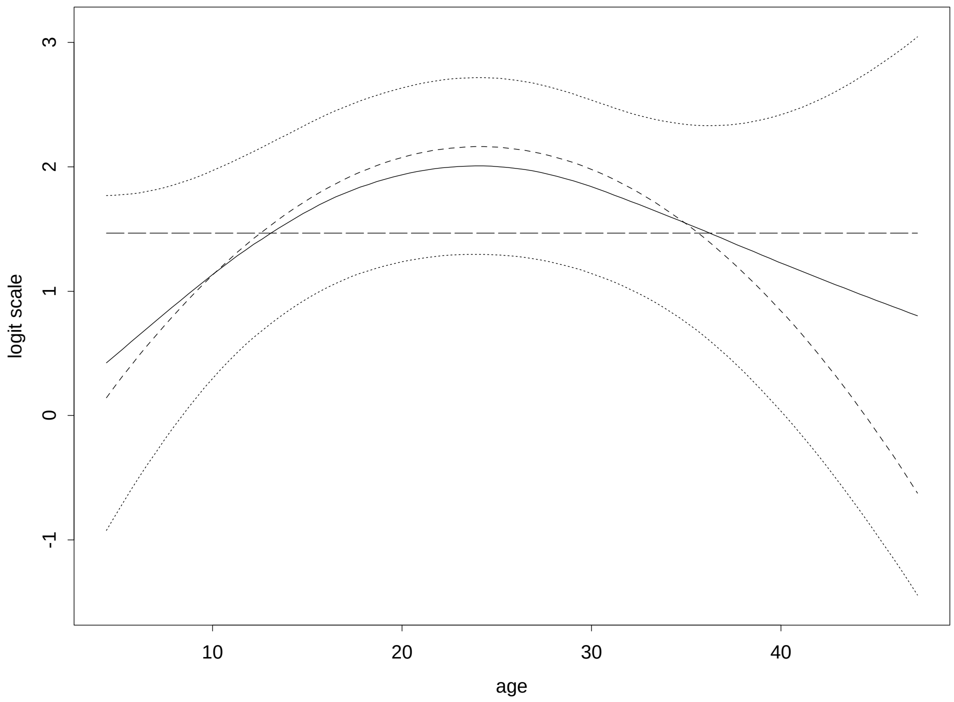

Figure 1: Estimates of the age main effect f in model (2) (solid line) using a cubic smoothing spline and 95% Bayesian confidence interval (dotted lines). The estimate of α_0 setting $\alpha_1 = \alpha_2 = 0$ is plotted as the long-dashed line. The quadratic age effect in model (1) is plotted as the short-dashed line.

3 Smoothing spline models for exponential families

In this section we briefly review smoothing spline models for exponential families, their corresponding Bayesian models and connections with generalized linear mixed effects models.

Let $y_i, i = 1, \cdots, n$, be responses with density from the exponential family

$$p(y_i|\xi_i, \sigma^2) = \exp\left\{(y_i\xi_i - h(\xi_i))/a_i(\sigma^2) + c_i(y_i, \sigma^2)\right\}, \tag{3}$$

where ξ_i is the canonical parameter and σ^2 is the dispersion parameter. Let $t_i, i = 1, \cdots, n$, be measurements of covariates $t \in \mathcal{T}$, where $\mathcal{T}$ is an arbitrary set which can be either univariate or multivariate. Denote $\mu_i = E(y_i|t_i) = h'(\xi_i)$, $i = 1, \cdots, n$. Let $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$ and $\mathbf{y} = (y_1, \cdots, y_n)'$. Our goal is to investigate how $\mu = E(y|t)$ depends on t. Specifically, we assume that $g(\mu) = f(t)$ where $g(\cdot)$ is a known link function and the unknown function $f(\cdot)$ models the covariate effects. We assume that $f \in \mathcal{H}$, a Reproducing Kernel Hilbert Space (RKHS) on $\mathcal{T}$. Usually $\mathcal{H}$ can be decomposed as $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ where $\mathcal{H}_0 = span\{\phi_1(t), \cdots, \phi_M(t)\}$, a finite dimensional space, and $\mathcal{H}_1$ is a RKHS with a reproducing kernel, say, $R^1(s,t)$ for any generic s and t in $\mathcal{T}$. For example, for polynomial splines on $\mathcal{T} = [0, 1]$,

$$\mathcal{H} = \mathcal{W}_m = \left\{f|f, \cdots, f^{(m-1)} \text{ are absolutely continuous, } f^{(m)} \in \mathcal{L}_2[0,1]\right\}. \tag{4}$$

$\mathcal{W}_m$ can be decomposed to the direct sum of $\mathcal{H}_0 = span\{\phi_1(t), \cdots, \phi_m(t)\}$ with $\phi_\nu(t) = t^{\nu-1}/(\nu-1)!$, $\nu = 1, \cdots, m$, and a RKHS with reproducing kernel

$$R^1_m(s,t) = \int_0^{min(s,t)} (s-u)^{m-1}(t-u)^{m-1} du/((m-1)!)^2. \tag{5}$$

We will assume that $a_i(\sigma^2) = \sigma^2/\varpi_i$, where σ^2 is a dispersion parameter which may be unknown and ϖ_i's are known weights. Let $f_i = f(t_i)$ and $\mathbf{f} = (f_1, \cdots, f_n)'$. Denote the likelihood of $\mathbf{y}$ given $\mathbf{f}$ as

$$p(\mathbf{y}|\mathbf{f}) = \exp\left\{-\frac{1}{\sigma^2} l(\mathbf{y}|\mathbf{f})\right\},$$

where

$$l(\mathbf{y}|\mathbf{f}) = -\sum_{i=1}^n \varpi_i \{y_i\xi_i - h(\xi_i)\} - \sigma^2 \sum_{i=1}^n c_i(y_i, \sigma^2). \tag{6}$$

The smoothing spline estimate of f, $\hat{f}_\lambda$, is defined as the minimizer of the penalized log likelihood (Gu 1992, Wahba, Wang, Gu, Klein & Klein 1995):

$$l(\mathbf{y}|\mathbf{f}) + (n/2)\lambda||P_1 f||^2, \tag{7}$$

where P_1 is the orthogonal projection of f onto $\mathcal{H}_1$ and λ is a smoothing parameter controlling the trade-off between the smoothness of the estimate and the goodness of fit.

The solution to the minimization problem (7) has the form

$$\hat{f}_\lambda(t) = \sum_{i=1}^{M} d_i\phi_i(t) + \sum_{i=1}^{n} c_i R^1(t, t_i)\ , \tag{8}$$

where $\mathbf{c} = (c_1, \cdots, c_n)'$ and $\mathbf{d} = (d_1, \cdots, d_M)'$ can be calculated by the Newton-Raphson procedure (Gu 1992, Wahba et al. 1995).

Wahba (1978) established a connection between the smoothing spline models and a Bayesian model, which was used later in developing the GML, LMP and GCV tests for Gaussian data (Cox et al. 1988, Wahba 1990). Gu (1992) extended this connection to data from exponential families. Consider the following prior for f:

$$F(t) = \sum_{i=1}^{M} \theta_i\phi_i(t) + b^{\frac{1}{2}} Z(t), \tag{9}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M)' \sim N(0, aI)$, $Z(t)$ is a Gaussian process independent of $\boldsymbol{\theta}$ with $E(Z(t)) = 0$ and $E(Z(s)Z(t)) = R^1(s,t)$. Let $b = \sigma^2/n\lambda$ and $a \to \infty$. The posterior distribution of $F(t)$ can be approximated by a Gaussian distribution with mean $\hat{f}_\lambda$ given in (8) (Gu 1992).

The connection between smoothing spline models and linear mixed effect models for Gaussian data has been well established (Wang 1998). We now extend this connection to non-Gaussian data. Denote $T_{n\times M} = \{\phi_\nu(t_i)\}_{i=1}^{n}{}_{\nu=1}^{M}$ and $\Sigma_{n\times n} = \{R^1(t_i, t_j)\}_{i=1}^{n}{}_{j=1}^{n}$. Consider the following generalized linear mixed effect model (GLMM) (Breslow & Clayton 1993):

$$g(E(\mathbf{y}|\mathbf{c})) = T\mathbf{d} + \Sigma\mathbf{c}, \tag{10}$$

where $\mathbf{d}$ is a vector of fixed effects, $\mathbf{c}$ is a vector of random effects and $\mathbf{c} \sim N(0, b\Sigma^+)$ with Σ^+ being the Moore-Penrose generalized inverse of Σ. It is easily seen that $n\lambda\|P_1 f\|^2 = n\lambda\mathbf{c}'\Sigma\mathbf{c} = bn\lambda\mathbf{c}'(b\Sigma^+)^+\mathbf{c} = \sigma^2\mathbf{c}'(b\Sigma^+)^+\mathbf{c}$ (Wahba 1990, Chapter 1). Therefore, the penalized likelihood (7) for the spline model is the same as the penalized quasi-likelihood (PQL) (Breslow & Clayton 1993, equation (6)) for the GLMM model (10) up to a multiplying constant σ^2.

4 Hypothesis tests

We are interested in testing a parametric model for $f \in \mathcal{H}$. As in Xiang & Wahba (1995), we can decompose $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ under a suitably defined norm in such a way that $\mathcal{H}_0$ is the model space of the parametric model. Therefore, we consider the hypothesis

$$\mathrm{H}_0 : f \in \mathcal{H}_0, \qquad \mathrm{H}_1 : f \in \mathcal{H} \text{ and } f \notin \mathcal{H}_0, \tag{11}$$

where $\mathcal{H}_0 = span\,\{\phi_\nu, \nu = 1, \cdots, M\}$.

It is easy to see that $\lambda = \infty$ in (7), or equivalently $b = 0$ in the corresponding Bayesian (9) and mixed effects models (10), leads to $f \in \mathcal{H}_0$. Thus the hypothesis (11) can be represented as

$$\mathrm{H}_0 : \lambda = \infty, \qquad \mathrm{H}_1 : \lambda < \infty, \tag{12}$$

or

$$\mathrm{H}_0 : b = 0, \qquad \mathrm{H}_1 : b > 0. \tag{13}$$

In the following, we develop the GML and LMP tests for hypothesis (13) based on the marginal density of $\mathbf{y}$ under the Bayesian model, and the GCV test based on the GCV criteria (Wahba et al. 1995).

4.1 Marginal density of y and its approximation

Notice that under the Bayesian model (9), $\mathbf{f}|\boldsymbol{\theta} \sim N(T\boldsymbol{\theta}, b\Sigma)$. Let $q(\mathbf{f})$ be the unconditional density of $\mathbf{f}$ assuming a flat prior for $\boldsymbol{\theta}$. It is derived in Gu (1992) that

$$q(\mathbf{f}) \propto b^{-\frac{n-M}{2}} |\Sigma|^{-\frac{1}{2}} |T'\Sigma^{-1}T|^{-\frac{1}{2}} \exp\left(-\frac{1}{2b}\mathbf{f}' B \mathbf{f}\right),$$

where

$$B = \Sigma^{-1} - \Sigma^{-1}T(T'\Sigma^{-1}T)^{-1}T'\Sigma^{-1}.$$

We include part of the normalizing term ignored in Gu (1992) because it depends on b, the parameter of interest for our hypothesis. When $b = 0$, $\mathbf{f}$ has a flat prior and $q(\mathbf{f}) \propto 1$.

The marginal density of $\mathbf{y}$ is

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f})q(\mathbf{f})d\mathbf{f}. \tag{14}$$

The integral in (14) usually does not have a closed form since $l(\mathbf{y}|\mathbf{f})$ in (6) is not quadratic in $\mathbf{f}$. Similar to Gu (1992), we use the Laplace method to approximate the integral. Let $\hat{\mathbf{f}}$ be the mode of $p(\mathbf{y}|\mathbf{f})q(\mathbf{f})$. Gu (1992) showed that $\hat{\mathbf{f}} = (\hat{f}_\lambda(t_1), \cdots, \hat{f}_\lambda(t_n))'$ where $\hat{f}_\lambda(t)$ is given in (8).

Let $\mathbf{u} = (u_1, \cdots, u_n)' = \partial l/\partial \mathbf{f}$ with $u_i = -\varpi_i(y_i - \mu_i)\left\{h''(\xi_i)g'(\mu_i)\right\}^{-1}$, and $W = \partial^2 l/\partial \mathbf{f}\partial \mathbf{f}' = diag(w_1, \cdots, w_n)$ with $w_i = \varpi_i\left\{h''(\xi_i)(g'(\mu_i))^2\right\}^{-1} - \varpi_i(y_i - \mu_i)\partial\left\{h''(\xi_i)g'(\mu_i)\right\}^{-1}/\partial f_i$. Here l denotes $l(\mathbf{y}|\mathbf{f})$. Since the second term in w_i has expectation 0, it is ignored in the remainder of the paper (Breslow & Clayton 1993). Let W_c and $\mathbf{u}_c$ be W and $\mathbf{u}$ evaluated at $f = \hat{f}_\lambda$. Let $\mathbf{y}_c = \hat{\mathbf{f}} - W_c^{-1}\mathbf{u}_c$ be the adjusted working variable. Expanding $l(\mathbf{y}|\mathbf{f})$ around $\hat{\mathbf{f}}$ leads to

$$l(\mathbf{y}|\mathbf{f}) \approx \frac{1}{2}(\mathbf{f} - \mathbf{y}_c)'W_c(\mathbf{f} - \mathbf{y}_c) + C,$$

where $C = l(\mathbf{y}|\hat{\mathbf{f}}) - \frac{1}{2}\mathbf{u}_c{}'W_c^{-1}\mathbf{u}_c$ which is independent of $\mathbf{f}$.

Some algebra (Appendix A) shows that $p(\mathbf{y})$ is approximately proportional to

$$p(\mathbf{y}) \propto C_0|V|^{-\frac{1}{2}}|T'V^{-1}T|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{y}_c'(V^{-1} - V^{-1}T(T'V^{-1}T)^{-1}T'V^{-1})\mathbf{y}_c\right\}, \tag{15}$$

where $V = b\Sigma + \sigma^2 W_c^{-1}$ and $C_0 = (2\pi)^{\frac{n}{2}}\sigma^n\exp(-\frac{1}{\sigma^2}C)|W_c|^{-\frac{1}{2}}$. It is easy to verify that aside from C_0, (15) is proportional to the restricted log likelihood for variance component b

based on the GLMM model (10) (Equation (13) after plugging in (10) in Breslow & Clayton, 1993). Let $\tilde{\mathbf{y}} = W_c^{1/2}\mathbf{y}_c$, $\tilde{\Sigma} = W_c^{1/2}\Sigma W_c^{1/2}$, $\tilde{T} = W_c^{1/2}T$ and the QR decomposition of $\tilde{T}$ be $(\tilde{Q}_1 \quad \tilde{Q}_2)(\tilde{R} \quad \mathbf{0})'$. Let UDU' be the spectral decomposition of $\tilde{Q}_2'\tilde{\Sigma}\tilde{Q}_2$ where $D = diag(\lambda_{\nu n}, \nu = 1, \cdots, n-M)$, and each $\lambda_{\nu n}$ is an eigenvalue with $\lambda_{1n} \geq \lambda_{2n} \geq \cdots \geq \lambda_{n-M\ n}$. Let $\mathbf{z} = (z_1, \cdots, z_{n-M})' = U'\tilde{Q}_2'\tilde{\mathbf{y}}$. Following arguments similar to the derivations right after Corollary 2.1 in Gu (1992) and Appendix B, (15) can be shown to be equivalent to

$$p(\mathbf{y}) \propto C_1 \prod_{\nu=1}^{n-M} (b\lambda_{\nu n} + \sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{\nu=1}^{n-M} \frac{z_\nu^2}{b\lambda_{\nu n} + \sigma^2}\right\} \tag{16}$$

where $C_1 = (2\pi)^{\frac{n}{2}}\sigma^n \exp(-\frac{1}{\sigma^2}C)|\tilde{R}|^{-1}$. Notice that $\lambda_{\nu n}$ and z_ν depend on b and C_1 depends on both σ^2 and b. C_1 can be further approximated using the same techniques as in Breslow & Clayton (1993). First it is easy to see that $\mathbf{u}_c'W_c^{-1}\mathbf{u}_c$ is the Pearson chi-square statistic. Let $l_s(\mathbf{y})$ denote (6) for the saturated model where $\boldsymbol{\mu}$ is estimated by the data $\mathbf{y}$. Thus in $l_s(\mathbf{y})$, ξ_i is the solution to the equation $y_i = \mu_i = h'(\xi_i)$. Replacing the Pearson chi-square statistic by the deviance, $2(l(\mathbf{y}|\hat{\mathbf{f}}) - l_s(\mathbf{y}))$, we have

$$C_1 \approx (2\pi)^{\frac{n}{2}}\sigma^n \exp\left(-\frac{1}{\sigma^2}l_s(\mathbf{y})\right)|\tilde{R}|^{-1}. \tag{17}$$

As in Breslow & Clayton (1993), we ignore the dependence of W_c on b. As a result, we treat both $\tilde{R}$ and C_1 as independent of b. However, C_1 in (17) still depends on σ^2. Therefore, we distinguish two different situations when σ^2 is known or unknown.

4.2 GML and LMP tests with known σ^2

For most applications, the dispersion parameter σ^2 is known. For example $\sigma^2 = 1$ for Binomial and Poisson data without over- and under-dispersion. In such cases, we denote $p(\mathbf{y})$ as $L(b|\mathbf{y})$ and define the GML statistic as

$$t_{GML} = \frac{L(0|\mathbf{y})}{\sup_b L(b|\mathbf{y})}. \tag{18}$$

To compute t_{GML}, we first apply approximations (16) and (17) to both the numerator and the denominator. The terms C_1 in the numerator and denominator are cancelled since they are independent of b. However, the dependence of $\lambda_{\nu n}$ and z_ν on b makes it difficult to use (16) directly. Therefore, we propose the following further approximation.

Let $\hat{b} = \sigma^2/n\hat{\lambda}$, where $\hat{\lambda}$ is any smoothing parameter estimate in (7) (Wahba et al. 1995, Lin, Wahba, Xiang, Gao, Klein & Klein 2000, Wood & Kohn 1998, Gu & Xiang 2001). Let $\tilde{\lambda}_{\nu n}$ and $\tilde{z}_\nu$ be $\lambda_{\nu n}$ and z_ν evaluated at $\hat{b}$. According to (16), we consider $\tilde{\mathbf{z}} = (\tilde{z}_1, \cdots, \tilde{z}_{n-M})'$ as from distribution

$$\tilde{\mathbf{z}} \sim N(0, b\tilde{D} + \sigma^2 I), \tag{19}$$

where $\tilde{D} = diag(\tilde{\lambda}_{\nu n})$. Let $\tilde{b}$ be the maximizer of the likelihood of $\tilde{\mathbf{z}}$ based on (19). We then approximate the denominator in (18) with the maximized likelihood of $\tilde{\mathbf{z}}$. Let $\check{\mathbf{z}} = (\check{z}_1, \cdots, \check{z}_{n-M})$ denote $\mathbf{z}$ under the null hypothesis. Finally, the GML test statistic is approximated by

$$t_{GML} \approx \frac{\prod_{\nu=1}^{n-M} (\tilde{b}\tilde{\lambda}_{\nu n} + \sigma^2)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\sum_{\nu=1}^{n-M} \tilde{z}_\nu^2/(\tilde{b}\tilde{\lambda}_{\nu n} + \sigma^2)\right\}}{\exp\left\{\frac{1}{2}\sum_{\nu=1}^{n-M} \check{z}_\nu^2/\sigma^2\right\}}, \tag{20}$$

where a constant is dropped. Note that this form of t_{GML} is different from the Gaussian case in Wahba (1990) because σ^2 is known here.

The LMP test statistic is defined as

$$t_{LMP} = \frac{\partial}{\partial b} \log L(b|\mathbf{y})|_{b=0}.$$

Again, the approximation (16) with C_1 in (17) is used for $L(b|\mathbf{y})$. Since the LMP test is defined assuming the null hypothesis is true, we further approximate (16) with z_ν and $\lambda_{\nu n}$ replaced by $\check{z}_\nu$ and $\check{\lambda}_{\nu n}$, which are calculated under the null hypothesis. Then the test statistic is approximated by

$$t_{LMP} \approx \sum_{\nu=1}^{n-M} \check{\lambda}_{\nu n} \check{z}_{\nu}^{2} - \sigma^2 \sum_{\nu=1}^{n-M} \check{\lambda}_{\nu n}, \tag{21}$$

where a multiplying constant $1/2\sigma^4$ is dropped. Each $\check{\lambda}_{\nu n}$ depends on data solely through W_c. Assuming W_c varies very slowly as a function of the mean (Breslow & Clayton 1993), we drop the last term in (21) and finally we have

$$t_{LMP} \approx \sum_{\nu=1}^{n-M} \check{\lambda}_{\nu n} \check{z}_{\nu}^{2}. \tag{22}$$

4.3 GML and LMP tests with unknown σ^2

We use $L(b, \sigma^2|\mathbf{y})$ to denote $p(\mathbf{y})$ subsequently. For the GML test, we extend the definition to

$$t_{GML} = \frac{\sup_{\sigma^2} L(0, \sigma^2|\mathbf{y})}{\sup_{b,\sigma^2} L(b, \sigma^2|\mathbf{y})}. \tag{23}$$

Again, (16) and (17) can be used to approximate $L(b, \sigma^2|\mathbf{y})$. However, the dependence of C_1 on σ^2 makes it difficult to obtain an explicit formula for (23). In general, one may use numerical methods to compute it based on the approximations. In the following we derive explicit formulas for two important special cases.

The first special case includes all distributions in the exponential family with C_1 in (17) independent of σ^2 or nearly so. It is easy to check that both Gaussian and Inverse Gaussian distributions have C_1 independent of σ^2 (see Appendix C for a proof for the Inverse Gaussian distribution). In this case we may again approximate $L(b, \sigma^2|\mathbf{y})$ by the normal density function in (19). Following the same steps as in the Gaussian case (Wahba 1990), one can show that the GML test in (23) is approximated by

$$t_{GML} = \frac{\sum_{\nu=1}^{n-M} \tilde{z}_{\nu}^{2}/(\tilde{\lambda}_{\nu n} + n\lambda_{GML})}{\Pi_{\nu=1}^{n-M}(\tilde{\lambda}_{\nu n} + n\lambda_{GML})^{-\frac{1}{n-M}}} \frac{1}{\sum_{\nu=1}^{n-M} \tilde{z}_{\nu}^{2}}, \tag{24}$$

where λ_{GML}, the Generalized Maximum Likelihood (GML) estimate of λ, is the minimizer of

$$M(\lambda) = \frac{\sum_{\nu=1}^{n-M} \tilde{z}_\nu^2/(\tilde{\lambda}_{\nu n} + n\lambda)}{\Pi_{\nu=1}^{n-M}(\tilde{\lambda}_{\nu n} + n\lambda)^{-\frac{1}{n-M}}}.$$

Another special case is the Gamma distribution where C_1 depends on σ^2. In this case, the GML test (23) can be simplified by profiling the likelihood $L(b, \sigma^2|\mathbf{y})$. It is shown in Appendix D that when σ^2 is small, the GML test statistic (23) can be approximated by

$$t_{GML} \approx \frac{\inf_\lambda \Pi_{\nu=1}^{n-M}(\lambda_{\nu n}/n\lambda + 1)^{\frac{1}{2}} \hat{\sigma}_\lambda^{n-M} \exp(n\hat{\sigma}_\lambda^2/6)}{\hat{\sigma}_\infty^{n-M} \exp(n\hat{\sigma}_\infty^2/6)}, \tag{25}$$

where $\hat{\sigma}_\lambda^2 = [9(1-M/n)^2 + 6\sum_{\nu=1}^{n-M} z_\nu^2/(\lambda_{\nu n}/\lambda + n)]^{1/2} - 3(1-M/n)$ and $\hat{\sigma}_\infty^2 = [9(1-M/n)^2 + 6\sum_{\nu=1}^{n-M} \check{z}_\nu^2/n]^{1/2} - 3(1-M/n)$.

To extend the LMP test statistic, let $\hat{\sigma}^2$ be any consistent estimate of σ^2. Define $I_{**} = -E(\frac{\partial^2 \log L(b,\sigma^2)}{\partial **})|_{b=0,\sigma^2=\hat{\sigma}^2}$ for each combination $**$ of b and σ^2, with the expectation taken under the null hypothesis. Let $I^{bb}(0,\hat{\sigma}^2) = (I_{bb} - I_{b\sigma^2} I_{\sigma^2\sigma^2}^{-1} I'_{b\sigma^2})^{-1}$. We extend the LMP test statistic as (Cox & Hinkley 1974, Chapter 9)

$$t_{appLMP} = \sqrt{I^{bb}(0,\hat{\sigma}^2)}\frac{\partial}{\partial b}\log L(b, \hat{\sigma}^2|\mathbf{y})|_{b=0}.$$

The extra square root term is used to account for the fact that σ^2 is estimated. Again, we approximate the likelihood $L(b, \sigma^2|\mathbf{y})$ by (16) with C_1 given in (17) which is independent of b. Since the test statistic is defined under the null hypothesis, we further approximate the likelihood by replacing z_ν and $\lambda_{\nu n}$ by $\check{z}_\nu$ and $\check{\lambda}_\nu$ respectively. Then the test statistic becomes

$$t_{appLMP} \approx \frac{\sqrt{I^{bb}(0,\hat{\sigma}^2)}}{2\hat{\sigma}^2}\left(\frac{1}{\hat{\sigma}^2}\sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}\check{z}_\nu^2 - \sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}\right). \tag{26}$$

In the following, we estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n-M}\sum_{\nu=1}^{n-M}\check{z}_\nu^2. \tag{27}$$

Simple derivations show that

$$I_{bb} = \frac{1}{2\hat{\sigma}^4}\sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}^2, \quad I_{b\sigma^2} = \frac{1}{2\hat{\sigma}^4}\sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}, \quad I_{\sigma^2\sigma^2} = \frac{n-M}{2\hat{\sigma}^4} - E\left(\frac{\partial^2 \log C_1}{\partial(\sigma^2)^2}\right).$$

For distributions with C_1 independent of σ^2, for example, the Gaussian and Inverse Gaussian distributions, $\partial^2 \log C_1/\partial(\sigma^2)^2 = 0$. Thus by definition

$$I^{bb}(0,\hat{\sigma}^2) = 2\hat{\sigma}^4\left(\sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}^2 - \frac{1}{n-M}\left(\sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}\right)^2\right)^{-1}.$$

Now plug the above formula into (26). By the same argument presented in section (4.2), we drop the terms in (26) which depend on $\check{\lambda}_{\nu n}$'s only and the test statistic becomes

$$t_{appLMP} \approx \sum_{\nu=1}^{n-M}\check{\lambda}_{\nu n}\check{z}_\nu^2 / \sum_{\nu=1}^{n-M}\check{z}_\nu^2. \tag{28}$$

This test statistic is a special case of the score test by Zhang & Lin (2003) for the same hypothesis (11) under the semiparametric additive mixed models but with a different derivation.

For distributions with C_1 dependent on σ^2, $E\left(\partial^2 \log C_1/\partial(\sigma^2)^2\right)$ needs to be calculated and plugged into (26) through $I^{bb}(0,\hat{\sigma}^2)$. For the Gamma distribution, from equation (36) in Appendix D, we have $\partial^2 \log C_1/\partial(\sigma^2)^2 \approx 0$ when σ^2 is small. Thus the approximate LMP test (28) can also be used for the Gamma distribution.

4.4 GCV test

For $\lambda < \infty$, the GCV function at convergence is defined as (Wahba et al. 1995)

$$V(\lambda) = \frac{\sum_{\nu=1}^{n-M}\tilde{z}_\nu^2/(1+\tilde{\lambda}_{\nu n}/n\lambda)^2}{[\sum_{\nu=1}^{n-M}1/(1+\tilde{\lambda}_{\nu n}/n\lambda)]^2}. \tag{29}$$

Let λ_{GCV} be the minimizer of $V(\lambda)$, which is the GCV estimate of the smoothing parameter. For $\lambda = \infty$ which corresponds to fitting the GLM under H_0, $\tilde{\mathbf{z}} = \check{\mathbf{z}}$ and (29) reduces to

$$V(\infty) = \sum_{\nu=1}^{n-M} \check{z}_\nu^2/(n-M).$$

As in the Gaussian case (Wahba 1990, Chapter 6), we define the GCV test statistic as $t_{GCV} = V(\lambda_{GCV})/V(\infty)$. Dropping a multiplying constant, we have

$$t_{GCV} = \frac{\sum_{\nu=1}^{n-M} \check{z}_\nu^2/(1+\tilde{\lambda}_{\nu n}/n\lambda_{GCV})^2}{[\sum_{\nu=1}^{n-M} 1/(1+\tilde{\lambda}_{\nu n}/n\lambda_{GCV})]^2} \frac{1}{\sum_{\nu=1}^{n-M} \check{z}_\nu^2}. \tag{30}$$

The GCV test does not require σ^2 to be known. Thus it can be applied both when σ^2 is known or unknown.

4.5 Null distributions

We reject H_0 when t_{GML} or t_{GCV} are too small, or t_{LMP} or t_{appLMP} are too large. In the Gaussian case the null distributions of the test statistics do not depend on nuisance parameters (Liu & Wang 2002). However, in the general class of exponential families, the null distributions of the test statistics depend on the true null function $f_0(t) = \sum_{i=1}^M d_i\phi_i(t)$ and thus depend on nuisance parameters $\mathbf{d}$. They also depend on σ^2 when it is unknown. Therefore, it is very difficult to obtain analytic results for the null distributions. Xiang & Wahba (1995) proposed the following bootstrap procedure to approximate the null distributions:

1. Fit the observed data with the model under H_0.
2. Generate bootstrap samples using the fitted model under H_0.
3. Calculate test statistics based on these bootstrap samples which form empirical null distributions.

Such a bootstrap procedure usually provides good approximations to the exact null distributions (Xiang & Wahba 1995), and it can be used for our tests here. However, it is computationally intensive. We propose the following alternative null distribution approximation method which applies to all tests except for the test (25). First note that for any observed data, once the hypothesis and $\mathcal{H}_0$ and $\mathcal{H}_1$ are decided, T and Σ are known and remain unchanged. We calculate W_c from the observed data, followed by $\tilde{\lambda}_{\nu n}$ and $\check{\lambda}_{\nu n}$ for $\nu \in \{1, \cdots, n-M\}$. According to (19), we then generate $\tilde{\mathbf{z}}$ and $\check{\mathbf{z}}$ from $N(0, I)$ under the null hypothesis $b = 0$. Note that for convenience, σ^2 is set to be 1 since when σ^2 is unknown, all the tests except (25) are transformation invariant to σ^2. For each set of generated $\tilde{\mathbf{z}}$, the estimated smoothing parameters $\tilde{b}$, λ_{GCV} and λ_{GML} are calculated. We then calculate the test statistics based on each realization of $\tilde{\mathbf{z}}$ and $\check{\mathbf{z}}$. By repeating this process many times, we obtain empirically approximated null distributions. We investigate these approximations through simulations in section 5.

5 Simulations

Two sets of simulations are presented in this section. The first set considers Binomial data and the second set considers Gamma data. In the first simulation, cubic smoothing splines (m=2 in (4)) are used. Let $n = 100$ and $t_i = (i-1)/(n-1), i = 1, \cdots, n$. We generate Binomial data $y_i \sim b(k, p(t_i)), i = 1, \cdots, n$, with two choices of k: $k = 1$ (Binary) and $k = 4$, and the following choices of $f(t) = logit\,(p(t))$:

$$f(t) = t + a(t-0.5)^2, \tag{31}$$

$$f(t) = a(2t-1)^3 - 2t, \tag{32}$$

$$f(t) = 1 + t + acos(6\pi t). \tag{33}$$

These three models are chosen to represent functions with increasing frequencies as illustrated in the top row of Figure 2. Four values of a are chosen for each form of model: $a = 0$, 3, 5 and 7 for (31), $a = 0$, 1.5, 3 and 5 for (32) and $a = 0$, 0.5, 1 and 1.5 for (33). We are testing whether each of these models is significantly different from a linear model: H_0: $f \in span\{1, t\}$. Thus H_0 is true when $a = 0$. Models with larger values of a are further away from H_0. It is therefore desirable that the null hypothesis be rejected more often for larger a.

In section 4.5 we noted the dependence of the null distributions on $\mathbf{d}$ and σ^2. If they were known, we could generate samples $\mathbf{y}$ from the true null model and compute test statistics. Repeating this process would give us empirical null distributions. The p-values calculated from these true empirical null distributions could then serve as a benchmark for the p-values calculated from the approximated null distribution as described in section 4.5. The true models (31), (32) and (33) are known completely in the simulations. However, they do not belong to the linear space $\mathcal{H}_0$ when $a > 0$ and the corresponding true null model is not readily available. In the following, we use the projection of the true function f onto $\mathcal{H}_0$ as its null model (Xiang & Wahba 1995), which we call the proxy null model. For any $f \notin \mathcal{H}_0$, define the proxy null model as

$$f^* = argmin_{g \in \mathcal{H}_0} RKL(g|f),$$

where

$$RKL(g|f) = \int_{\mathcal{T}} \left(h(g(t)) - E_f(y|t)g(t)\right) d(t)dt$$

is the relative Kullback-Leibler (RKL) distance with $h(\cdot)$ as defined in (3) and $d(t)$ as the sampling density of t. Xiang & Wahba (1995) showed that the GLM estimate of f under $\mathcal{H}_0$ converges to f^* as $n \to \infty$. Thus f^* can be used as the proxy null model for f. Note that f^* usually does not equal the function obtained by simply setting $a = 0$. In our simulations, we calculate f^* as

$$f^* = argmin_{g \in \mathcal{H}_0} \sum_{i=1}^{n} \{h(g(t_i)) - E_f(y|t_i)g(t_i)\}.$$

True null distributions are formed by 40000 test statistics based on samples generated from the proxy null models. For any chosen significance level, a decision can be made as to whether to accept or reject the null hypothesis. We use a significance level of 0.05 in the simulations. The proportions of rejections based on 2000 replications from each model are obtained. They are shown in Table 1, 2 and 3 with row entries labelled as "true". Similarly, the proportions of rejections based on approximated null distributions are labelled as "approx" in these tables. Figure 2 shows the proportion of rejections for each of the tests for each of the models obtained with approximated null distributions.

		k=1				k=4			
		a=0	a=3	a=5	a=7	a=0	a=3	a=5	a=7
LMP	true	0.050	0.171	0.371	0.565	0.046	0.536	0.909	0.990
	approx	0.044	0.183	0.408	0.595	0.055	0.537	0.908	0.990
GCV	true	0.052	0.127	0.249	0.328	0.046	0.376	0.810	0.967
	approx	0.036	0.101	0.211	0.397	0.042	0.361	0.795	0.965
GML	true	0.053	0.169	0.339	0.560	0.051	0.516	0.891	0.989
	approx	0.041	0.150	0.314	0.534	0.050	0.501	0.889	0.989

Table 1: Proportion of rejections in 2,000 replications under model (31) with Binary ($k = 1$) and Binomial ($k = 4$) data.

It is observed that all the tests hold their levels reasonably correctly (Tables 1, 2 and 3). Power increases as a increases for each form of models although very slowly for the LMP test for model (33). For the low frequency function (31), GML and the LMP tests are more

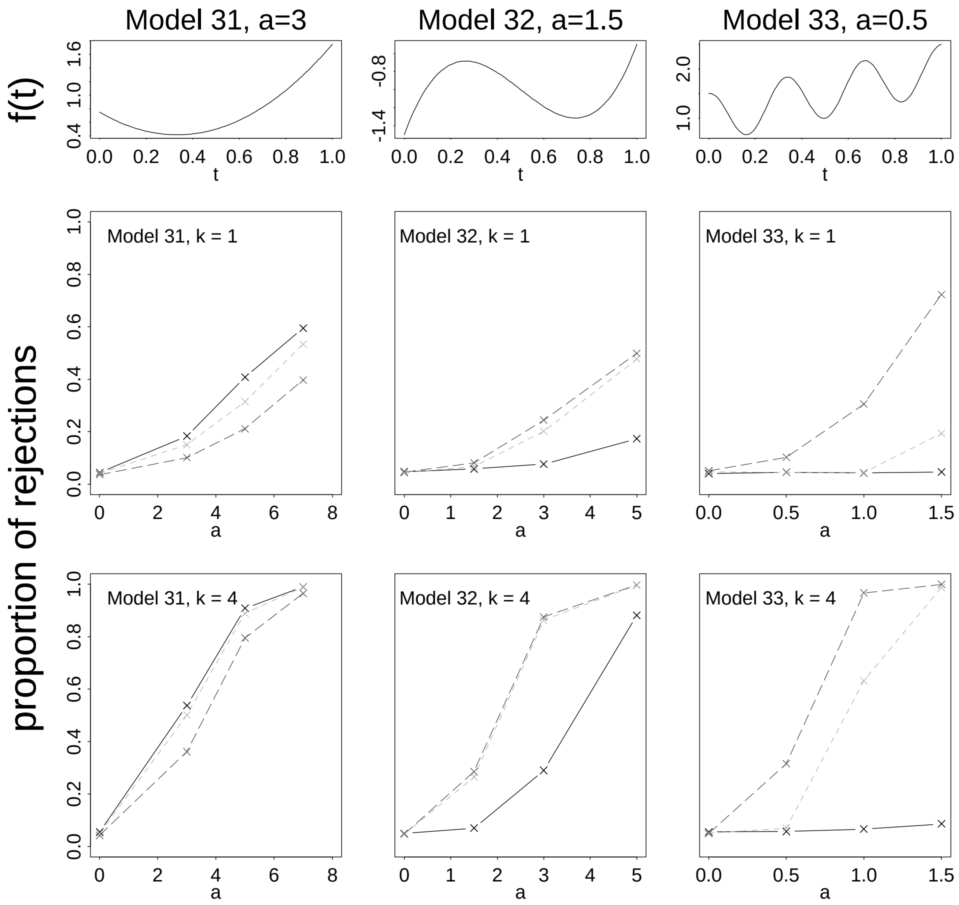

Figure 2: The left column of panels corresponds to model (31), the center column to model (32), and the right column to model (33). Top row: $f(t)$ for each model for one of the simulation a values. Middle row: Proportion of rejections in each of 2000 replications for each of the models, versus a for the Binary simulation ($k = 1$). Lowest row: Proportion of rejections in each of 2000 replications for each of the models, versus a for the Binomial simulation ($k = 4$). In each of the lower two rows, a line is drawn for each of the three tests. Solid lines correspond to LMP test results, long dashed lines correspond to GCV test results, and short dashed lines correspond to GML test results. Crosses indicate results using the approximated null distributions.

		k=1				k=4			
		a=0	a=1.5	a=3	a=5	a=0	a=1.5	a=3	a=5
LMP	true	0.053	0.052	0.068	0.153	0.054	0.075	0.330	0.875
	approx	0.047	0.058	0.076	0.173	0.049	0.070	0.290	0.881
GCV	true	0.048	0.093	0.261	0.477	0.053	0.289	0.879	0.997
	approx	0.046	0.081	0.244	0.499	0.047	0.285	0.875	0.997
GML	true	0.048	0.079	0.215	0.476	0.052	0.268	0.859	0.996
	approx	0.043	0.067	0.202	0.477	0.047	0.264	0.863	0.996

Table 2: Proportion of rejections in 2,000 replications under model (32) with Binary ($k = 1$) and Binomial ($k = 4$) data.

		k=1				k=4			
		a=0	a=0.5	a=1	a=1.5	a=0	a=0.5	a=1	a=1.5
LMP	true	0.049	0.051	0.053	0.051	0.050	0.052	0.065	0.084
	approx	0.040	0.045	0.043	0.046	0.055	0.057	0.066	0.086
GCV	true	0.045	0.096	0.293	0.730	0.051	0.319	0.967	0.999
	approx	0.051	0.102	0.305	0.723	0.049	0.315	0.967	0.999
GML	true	0.044	0.044	0.046	0.198	0.051	0.069	0.634	0.988
	approx	0.047	0.044	0.044	0.193	0.051	0.070	0.632	0.987

Table 3: Proportion of rejections in 2,000 replications under model (33) with Binary ($k = 1$) and Binomial ($k = 4$) data.

powerful. For the high frequency function (33), GCV test outperforms other tests. GML and GCV tests are more powerful than the LMP test for model (32). We also ran simulations with Poisson data, other functions and sample sizes. Similar results were obtained. We conclude that for detecting relatively low frequency functions, the GML test is preferred, and for detecting high frequency functions, the GCV test is preferred. This is consistent with the observations for corresponding tests in the Gaussian case (Liu & Wang 2002). The comparative behavior of the tests is intrinsic and is explained in that paper by different weighting schemes of the tests. The LMP test concentrates its weights on low frequency functions relative to the null hypothesis. Compared with the LMP test, the GML puts relatively larger weights on higher frequency functions, and the GCV test further increases the weights on higher frequency functions. Powers based on Binomial data (k=4) are larger than those based on Binary data (k=1). This is not surprising since the Binomial data is equivalent to having four replicates of Binary data at each point, thus four times the sample size. The null distribution approximation provides a practical way to apply our methods to real data. From the comparisons shown in the tables, it can be seen that they give satisfying approximations.

For the second set of simulations, we generate Gamma data from the following models of $f(t) = \log \mu$:

$$f(t) = 1 + t^2 + a\cos(3\pi t), \tag{34}$$

$$f(t) = 1 + t^2 + a\cos(6\pi t), \tag{35}$$

where $a = 0$, 0.05, 0.1, 0.2 in model (34) and $a = 0$, 0.1, 0.2, 1 in (35). All other settings remain the same as for the first set of simulations. We fit data using quintic smoothing splines (m=3) instead of cubic smoothing splines (m=2). The GML test (25), LMP test (28) and GCV test (30) are used to test a quadratic null hypothesis. The null distributions are approximated by bootstrapping as described at the beginning of Section 4.5. The results are shown in Table 4 and 5 for different values of σ^2.

It is seen that the comparative behavior of the tests remains the same. However, the LMP test does not perform well for the high frequency function model (35). Table 5 shows that

	$\sigma^2 = 0.01$				$\sigma^2 = 0.1$			
	a=0	a=0.05	a=0.1	a=0.2	a=0	a=0.05	a=0.1	a=0.2
LMP	0.05	0.887	1	1	0.051	0.162	0.529	0.982
GCV	0.051	0.835	1	1	0.052	0.133	0.44	0.967
GML	0.048	0.887	1	1	0.049	0.166	0.513	0.978

Table 4: Proportion of rejections in 2,000 replications under model (34) with Gamma data.

	$\sigma^2 = 0.01$				$\sigma^2 = 0.1$			
	a=0	a=0.1	a=0.2	a=1	a=0	a=0.1	a=0.2	a=1
LMP	0.05	0.024	0.002	0	0.051	0.054	0.041	0
GCV	0.051	1	1	1	0.052	0.209	0.821	1
GML	0.048	0.96	1	1	0.049	0.071	0.321	1

Table 5: Proportion of rejections in 2,000 replications under model (35) with Gamma data.

the power of the LMP test may even decrease as a increases. Further investigation shows that under the high frequency model, the estimate of σ^2 based on (27) becomes increasingly inflated as a increases, leading to the decrease in power. Liu & Wang (2002) shows that the LMP test does not perform as well as the GML and GCV tests for high frequency functions even for Gaussian data. Simulation results suggest that these results remain true for distributions in the exponential family. Therefore, we do not recommend the LMP test for high frequency functions.

6 Application to WESDR data

For the data presented in section 2, we first rescale the age variable such that $x_1 \in [0,1]$. Based on model (2), we are interested in testing the following three hypotheses with $\mathcal{W}_m$ as defined in (4):

1. $\mathrm{H}_0^1 : f \in span\{1\}$ vs $\mathrm{H}_1^1 : f \in \mathcal{W}_1$ and $f \notin span\{1\}$,
2. $\mathrm{H}_0^2 : f \in span\{1, x_1\}$ vs $\mathrm{H}_1^2 : f \in \mathcal{W}_2$ and $f \notin span\{1, x_1\}$,
3. $\mathrm{H}_0^3 : f \in span\{1, x_1, x_1^2\}$ vs $\mathrm{H}_1^3 : f \in \mathcal{W}_3$ and $f \notin span\{1, x_1, x_1^2\}$,

which correspond to no age effect, linear and quadratic age effects respectively. Note that three different polynomial spline spaces on [0, 1] are used for different hypotheses such that there is no penalty for functions under the null hypotheses. Let $t = (x_1, x_2, x_3, x_4)'$. The model spaces of model (2) for three hypotheses are $\mathcal{H} = \mathcal{W}_1 \otimes \mathcal{L}$, $\mathcal{H} = \mathcal{W}_2 \otimes \mathcal{L}$ and $\mathcal{H} = \mathcal{W}_3 \otimes \mathcal{L}$ respectively where $\mathcal{L} = span\{x_2, x_3, x_4\}$ represents the linear model space for duration, glycosylated haemoglobin and pressure. $\mathcal{H}$ is decomposed into $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ according to each of the hypotheses. The decompositions and their reproducing kernels are listed in Table 6.

	$\mathcal{H}$	$\mathcal{H}_0$	$\mathcal{H}_1$	R^1 for $\mathcal{H}_1$
1	$\mathcal{W}_1 \otimes \mathcal{L}$	$\mathcal{L}$	$(\mathcal{W}_1 \ominus span\{1\}) \otimes \mathcal{L}$	R_1^1
2	$\mathcal{W}_2 \otimes \mathcal{L}$	$span\{1, x_1\} \otimes \mathcal{L}$	$(\mathcal{W}_2 \ominus span\{1, x_1\}) \otimes \mathcal{L}$	R_2^1
3	$\mathcal{W}_3 \otimes \mathcal{L}$	$span\{1, x_1, x_1^2\} \otimes \mathcal{L}$	$(\mathcal{W}_3 \ominus span\{1, x_1, x_1^2\}) \otimes \mathcal{L}$	R_3^1

Table 6: The model spaces $\mathcal{H}$, their decompositions $\mathcal{H}_0$ and $\mathcal{H}_1$, and reproducing kernels of $\mathcal{H}_1$ for hypotheses 1, 2 and 3, where R_m^1 is given in (5).

The null distributions for the tests are approximated as described in section 4.5. The p-values of the approximated null distributions are listed in Table 7. One may also test these hypotheses using polynomial GLM. Specifically, one may test H_0^1, H_0^2 and H_0^3 by comparing a linear model vs a constant, a quadratic model vs a linear model and a cubic model vs a quadratic model respectively. P-values based on these nested polynomial GLM tests are added for comparison.

Hypothesis	LMP	GML	GCV	polynomial GLM
1	0.134	0.072	0.078	0.276
2	0.004	0.015	0.039	0.007
3	0.305	0.325	0.412	0.263

Table 7: P-values for the three hypotheses on age.

For the hypothesis of a constant age effect, a 0.1 significance level would lead to rejection of the constant age effect by GML and GCV tests, but not by the LMP test and the linear GLM test. The large p-value of the linear GLM test led Klein et al. (1988) to conclude that the age variable is not significant. The lower power of the LMP test can be explained by the fact that for the constant null hypothesis, the actual close-to-quadratic age effect is a relatively high frequency function. All four tests reject the hypothesis that the age effect is linear at significance level 0.05. The LMP test is more powerful than the GML and GCV tests for testing the second hypothesis since the true age effect is close to quadratic and is a low frequency function relative to the linear null hypothesis. All four tests fail to reject the hypothesis that the age effect is quadratic.

7 Discussion

We extended the GML, GCV and LMP tests to data from exponential families. Approximations to the null distributions of the tests were developed so that the tests are easily implemented. They showed good approximations in our simulations. More accurate but computationally intensive methods may also be used (Xiang & Wahba 1995). Simulations and a real data application indicated that the GML test and the GCV test are preferred for relatively low and high frequency functions respectively.

The good performances of the tests motivate us to extend them further. Interesting circumstances include hypothesis testing for GLMM (Breslow & Clayton 1993) and generalized additive models (Hastie & Tibshirani 1990) for multivariate data, where SS ANOVA models can be used (Wahba et al. 1995).

Bayesian models for smoothing splines have been used to construct confidence intervals (Wahba 1983). They have good frequentist properties when smoothing parameters are estimated from data (Wahba 1983, Nychka 1988, Nychka 1990) even though some theoretical issues remain (Cox 1993, Shen & Wasserman 2001). The GML and LMP tests in this paper are derived from Bayesian models. Nevertheless, these Bayesian models and the many approximations involved serve as heuristic motivations rather than rigorous theoretical justifications to the final stand-alone test statistics. Empirical properties of these tests are evaluated by simulations. Future research on theoretical properties will be invaluable.

Acknowledgments

This work was supported by NIH Grants R01 GM58533. The authors thank Ronald Klein, MD and Barbara Klein, MD for providing the WESDR data. The authors also thank the associate editor and two referees for numerous useful comments that have greatly improved this article.

Appendix A

The marginal density defined in (14) is approximated by

$$p(\mathbf{y}) =$$
$$b^{-\frac{n-M}{2}} |\Sigma|^{-\frac{1}{2}} |T'\Sigma^{-1}T|^{-\frac{1}{2}} \exp\left(-\tfrac{C}{\sigma^2}\right) \int \exp\left\{-\tfrac{1}{2\sigma^2}(\mathbf{y}_c-\mathbf{f})' W_c (\mathbf{y}_c-\mathbf{f}) - \tfrac{1}{2b}\mathbf{f}'B\mathbf{f}\right\} d\mathbf{f}.$$

The integral can be simplified as follows:

$$\begin{aligned}
& \int \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}_c-\mathbf{f})' W_c (\mathbf{y}_c-\mathbf{f}) - \frac{1}{2b}\mathbf{f}'B\mathbf{f}\right\} d\mathbf{f} \\
= \; & \int \exp\left\{-\frac{1}{2}\left(\mathbf{f} - \left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)^{-1}\frac{W_c}{\sigma^2}\mathbf{y}_c\right)' \left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)\left(\mathbf{f} - \left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)^{-1}\frac{W_c}{\sigma^2}\mathbf{y}_c\right)\right\} \\
& \exp\left\{-\frac{1}{2}\mathbf{y}_c'\frac{W_c}{\sigma^2}\left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)^{-1}\frac{B}{b}\mathbf{y}_c\right\} d\mathbf{f} \\
= \; & (2\pi)^{\frac{n}{2}} \left|\frac{W_c}{\sigma^2}+\frac{B}{b}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{y}_c'\frac{W_c}{\sigma^2}\left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)^{-1}\frac{B}{b}\mathbf{y}_c\right\}.
\end{aligned}$$

According to equation (1.5.12) in Wahba (1990),

$$B = \lim_{a\to\infty}\left(aTT'+\Sigma\right)^{-1}.$$

Thus,

$$\begin{aligned}
& \frac{W_c}{\sigma^2}\left(\frac{W_c}{\sigma^2}+\frac{B}{b}\right)^{-1}\frac{B}{b} \\
= & \lim_{a\to\infty}\left(\sigma^2 W_c^{-1} + abTT' + b\Sigma\right)^{-1} \\
= & \lim_{a\to\infty}\left(abTT' + V\right)^{-1} \\
= & V^{-1} - V^{-1}T\left(T'V^{-1}T\right)^{-1}T'V^{-1}.
\end{aligned}$$

On the other hand, we use Harville (1977)'s Equation (5.1) by matching R with our $b\Sigma$, Z with the unit matrix and D with $\sigma^2 W_c^{-1}$. It can be shown that S is B/b in terms of our notion and the V matrices coincide. Thus we have

$$\begin{aligned}
& b^{-\frac{n-M}{2}}\left|\Sigma\right|^{-\frac{1}{2}}\left|T'\Sigma^{-1}T\right|^{-\frac{1}{2}}\left|\frac{W_c}{\sigma^2}+\frac{B}{b}\right|^{-\frac{1}{2}} \\
= & \left|\frac{W_c}{\sigma^2}\right|^{-\frac{1}{2}}\left|b\Sigma\right|^{-\frac{1}{2}}\left|T'\left(b\Sigma\right)^{-1}T\right|^{-\frac{1}{2}}\left|I+\frac{B}{b}\left(\sigma^2 W_c^{-1}\right)\right|^{-\frac{1}{2}} \\
= & \left|\frac{W_c}{\sigma^2}\right|^{-\frac{1}{2}}\left|V\right|^{-\frac{1}{2}}\left|T'V^{-1}T\right|^{-\frac{1}{2}}.
\end{aligned}$$

Thus (15) is derived.

Appendix B

Let $\tilde{V} = b\tilde{\Sigma} + \sigma^2 I$ and $\tilde{Q} = (\tilde{Q}_1, \tilde{Q}_2)$. Notice that $\tilde{Q}$ is orthogonal. We have

$$\begin{aligned}
& \left|V\right|^{-\frac{1}{2}}\left|T'V^{-1}T\right|^{-\frac{1}{2}} \\
= & \left|W_c\right|^{\frac{1}{2}}\left|\tilde{V}\right|^{-\frac{1}{2}}\left|\tilde{T}'\tilde{V}^{-1}\tilde{T}\right|^{-\frac{1}{2}} \\
= & \left|W_c\right|^{\frac{1}{2}}\left|\tilde{Q}'\tilde{V}^{-1}\tilde{Q}\right|^{\frac{1}{2}}\left|\tilde{Q}_1'\tilde{V}^{-1}\tilde{Q}_1\right|^{-\frac{1}{2}}\left|\tilde{R}\right|^{-1}.
\end{aligned}$$

Let $G = \tilde{Q}'\tilde{V}^{-1}\tilde{Q} = (\tilde{Q}'\tilde{V}\tilde{Q})^{-1}$ and write G as a block matrix $\begin{pmatrix} G_{11} & G_{12} \\ G_{12}' & G_{22} \end{pmatrix}$. Then the bottom right block of $G^{-1} = \tilde{Q}'\tilde{V}\tilde{Q}$, $\tilde{Q}_2'\tilde{V}\tilde{Q}_2$, can be represented by $\left(G_{22} - G_{12}'G_{11}^{-1}G_{12}\right)^{-1}$

(Rao 1973, Page 33). Therefore, $\left|G_{22} - G_{12}' G_{11}^{-1} G_{12}\right| = \left|\tilde{Q}_2' \tilde{V} \tilde{Q}_2\right|^{-1}$. Now

$$\begin{aligned}
\left|\tilde{Q}' \tilde{V}^{-1} \tilde{Q}\right| &= |G| \\
&= |G_{11}| \left|G_{22} - G_{12}' G_{11}^{-1} G_{12}\right| \\
&= \left|\tilde{Q}_1' \tilde{V}^{-1} \tilde{Q}_1\right| \left|\tilde{Q}_2' \tilde{V} \tilde{Q}_2\right|^{-1} \\
&= \left|\tilde{Q}_1' \tilde{V}^{-1} \tilde{Q}_1\right| \left|b \tilde{Q}_2' \tilde{\Sigma} \tilde{Q}_2 + \sigma^2 I\right|^{-1}.
\end{aligned}$$

Thus

$$\begin{aligned}
& |V|^{-\frac{1}{2}} \left|T' V^{-1} T\right|^{-\frac{1}{2}} \\
= \; & |W_c|^{\frac{1}{2}} \left|\tilde{R}\right|^{-1} \left|b \tilde{Q}_2' \tilde{\Sigma} \tilde{Q}_2 + \sigma^2 I\right|^{-\frac{1}{2}} \\
= \; & |W_c|^{\frac{1}{2}} \left|\tilde{R}\right|^{-1} \left|b U D U' + \sigma^2 I\right|^{-\frac{1}{2}} \\
= \; & |W_c|^{\frac{1}{2}} \left|\tilde{R}\right|^{-1} \left|b D + \sigma^2 I\right|^{-\frac{1}{2}}.
\end{aligned}$$

Appendix C

For data from the Inverse Gaussian distribution, $h(\xi_i) = -(-2\xi_i)^{1/2}$, $a_i(\sigma^2) = \sigma^2$, $\varpi_i = 1$, and $c_i(y_i, \sigma^2) = -\left\{\log(2\pi\sigma^2 y_i^3) + 1/(\sigma^2 y_i)\right\}/2$. It is easy to check that $l_s(\mathbf{y}) = \frac{\sigma^2}{2} \sum_{i=1}^n \log(2\pi\sigma^2 y_i^3)$. Thus C_1 in (17) is $(2\pi)^{\frac{n}{2}} \sigma^n \exp\left(-\sum_{i=1}^n (\log 2\pi\sigma^2 y_i^3)/2\right) |W_c|^{-\frac{1}{2}} = \prod_{i=1}^n y_i^{3/2} |W_c|^{-\frac{1}{2}}$ which is independent of σ^2.

Appendix D

For the Gamma distribution with density (3), $h(\xi_i) = -log(-\xi_i)$, $a_i(\sigma^2) = \sigma^2$, $\varpi_i = 1$, and $c_i(y_i, \sigma^2) = (\log y_i)/\sigma^2 - (log\sigma^2)/\sigma^2 - \log y_i - \log\Gamma(1/\sigma^2)$ (McCullagh & Nelder 1989). $1/\sigma^2$

is the shape parameter and $-\sigma^2/\xi_i$ is the scale parameter. Thus the mean is $-1/\xi_i$ and the variance is σ^2/ξ_i^2. It is easy to check that $l_s(\mathbf{y}) = n + n\log\sigma^2 + \sigma^2\sum_{i=1}^n \log y_i + n\sigma^2 \log\Gamma(1/\sigma^2)$. Up to an additive constant, we have

$$\begin{aligned}\log C_1 &= \frac{n}{2}\log\sigma^2 - \frac{1}{\sigma^2} l_s(\mathbf{y}) \\ &= \frac{n}{2}\log\sigma^2 - \frac{n}{\sigma^2} - n\left[\frac{\log\sigma^2}{\sigma^2} + \log\Gamma(\frac{1}{\sigma^2})\right] - \sum_{i=1}^n \log y_i.\end{aligned}$$

Using the formula (6.1.41) for $\Gamma(1/\sigma^2)$ in Abramowitz & Stegun (1970) and neglecting terms of order σ^6 and higher, we have $\log\sigma^2/\sigma^2 + \log\Gamma(1/\sigma^2) \approx -1/\sigma^2 + \log\sigma^2/2 + \sigma^2/12$. Thus

$$\log C_1 \approx -\frac{n\sigma^2}{12} - \sum_{i=1}^n \log y_i. \tag{36}$$

Note that $b = \sigma^2/n\lambda$. We rewrite $L(b, \sigma^2|\mathbf{y}) = L(\sigma^2/n\lambda, \sigma^2|\mathbf{y}) = LH(\lambda, \sigma^2|\mathbf{y})$, the likelihood of λ and σ^2. It can be easily verified that up to an additive constant

$$\begin{aligned}&\log LH(\lambda, \sigma^2|\mathbf{y}) \\ =\ &\log C_1 - \tfrac{n-M}{2}\log\sigma^2 - \tfrac{S_2}{2\sigma^2} - S_1 \\ \approx\ &-\tfrac{n\sigma^2}{12} - \tfrac{n-M}{2}\log\sigma^2 - \tfrac{S_2}{2\sigma^2} - S_1 - \textstyle\sum_{i=1}^n \log y_i\end{aligned} \tag{37}$$

where $S_1 = \frac{1}{2}\sum_{\nu=1}^{n-M} \log(\lambda_{\nu n}/n\lambda + 1)$ and $S_2 = \sum_{\nu=1}^{n-M} z_\nu^2/(\lambda_{\nu n}/n\lambda + 1)$.

Taking the derivative of $LH(\lambda, \sigma^2|\mathbf{y})$ with respect to σ^2, we have

$$\frac{\partial \log LH(\lambda, \sigma^2|\mathbf{y})}{\partial\sigma^2} \approx -\frac{n}{12} - \frac{n-M}{2\sigma^2} + \frac{1}{2\sigma^4}S_2.$$

It can be shown that $\hat\sigma_\lambda^2 = \left[9(1-M/n)^2 + 6\sum_{\nu=1}^{n-M} z_\nu^2/(\lambda_{\nu n}/\lambda + n)\right]^{1/2} - 3(1-M/n)$ is the unique positive root of the equation $\partial\log LH(\lambda, \sigma^2|\mathbf{y})/\partial\sigma^2 = 0$. Using $\hat\sigma_\lambda^2$ as an estimate of σ^2 and plugging $\hat\sigma_\lambda^2$ into (37), we have

$$\log LH(\lambda, \hat\sigma_\lambda^2|\mathbf{y}) \approx -\frac{n-M}{2}\log\hat\sigma_\lambda^2 - \frac{n\hat\sigma_\lambda^2}{6} - S_1. \tag{38}$$

Note that a new GML method of estimating λ for Gamma data can be developed based on maximizing (38). Now the GML test statistic reduces to

$$t_{GML} \approx \frac{LH(\infty, \hat{\sigma}^2_\infty|\mathbf{y})}{sup_\lambda LH(\lambda, \hat{\sigma}^2_\lambda|\mathbf{y})} = \frac{\inf_\lambda \Pi_{\nu=1}^{n-M}(\lambda_{\nu n}/n\lambda + 1)^{\frac{1}{2}} \hat{\sigma}_\lambda^{n-M} \exp(n\hat{\sigma}^2_\lambda/6)}{\hat{\sigma}_\infty^{n-M} \exp(n\hat{\sigma}^2_\infty/6)},$$

which is (25).

References

Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statisticial Association* **88**: 9–25.

Cox, D. (1993). An analysis of bayesian inference for nonparametric regression, *Annals of Statistics* **21**: 903–923.

Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, *Annals of Statistics* **16**: 113–119.

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

Gu, C. (1992). Penalized likelihood regression: A Bayesian analysis, *Statistica Sinica* **2**: 255–264.

Gu, C. and Xiang, D. (2001). Cross-validating non-Gaussian data: Generalized approximate cross-validation revisited, *Journal of Computational and Graphical Statistics* **10**: 581–591.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association* **72**: 320–338.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall.

Klein, R., Klein, B. E. K., Moss, S. E., Davis, M. D. and DeMets, D. L. (1988). Glycosylated hemoglobin predicts the incidence and progression of diabetic retinopathy, *Journal of the American Medical Association* **260**: 2864–2871.

Klein, R., Klein, B. E. K., Moss, S. E., Davis, M. D. and DeMets, D. L. (1989). Is blood pressure a predictor of the incidence or progression of diabetic retinopathy, *Archives of Internal Medicine* **149**: 2427–2432.

Lin, X., Wahba, G., Xiang, D., Gao, F., Klein, R. and Klein, B. (2000). Smoothing spline ANOVA models for large data sets with Bernoulli observations and the randomized GACV, *Annals of Statistics* **28**: 1570–1600.

Liu, A. and Wang, Y. (2002). Hypothesis testing in smoothing spline models, *Technical Report 387*, Department of Statistics and Applied Probability, University of California, Santa Barbara. Available at http://www.pstat.ucsb.edu/faculty/yuedong.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall, London.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines, *Journal of the American Statistical Association* **83**: 1134–1143.

Nychka, D. (1990). The average posterior variance of a smoothing spline and a consistent estimate of the average squared error, *Annals of Statistics* **18**: 415–428.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions, *Annals of Statistics* **29**: 687–714.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression, *Journal of the Royal Statistical Society B* **40**: 364–372.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline, *Journal of the Royal Statistical Society B* **45**: 133–150.

Wahba, G. (1990). *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy, *Annals of Statistics* **23**: 1865–1895.

Wang, Y. (1994). *Smoothing Spline Analysis of Variance of Data from Exponential Families*, PhD thesis, University of Wisconsin, Madison.

Wang, Y. (1998). Smoothing spline models with correlated random errors, *Journal of the American Statistical Association* **93**: 341–348.

Wang, Y. and Wahba, G. (1995). Bootstrap confidence intervals for smoothing spline estimates and their comparison to Bayesian confidence intervals, *Journal of Statistical Computation and Simulation* **51**: 263–279.

Wood, S. and Kohn, R. (1998). A Bayesian approach to robust binary nonparametric regression, *Journal of the American Statistical Association* **93**: 203–213.

Xiang, D. and Wahba, G. (1995). Testing the generalized linear model null hypothesis versus 'smooth' alternatives, *Technical Report 953*, Department of Statistics, University of Wisconsin, Madison. Available at http://www.stat.wisc.edu/∼wahba.

Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models, *Biostatistics* **4**: 57–74.

Running title: Nonparametric Tests for Generalized Linear Models

Anna Liu

Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106

E-mail: annaliu@pstat.ucsb.edu

Wendy Meiring

Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106

E-mail: meiring@pstat.ucsb.edu

Yuedong Wang (corresponding author)

Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106

E-mail: yuedong@pstat.ucsb.edu

Telephone: (805) 893-4870

Fax: (805) 893-2334