Published in final edited form as:

Commun Stat Theory Methods. 2012 January 1; 41(23): 4240–4250. doi:10.1080/03610926.2011.568157.

Lack-of-fit Tests for Generalized Linear Models via Splines

CHIN-SHANG LI

Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, CA 95616

Abstract

Cubic B -splines are used to estimate the nonparametric component of a semiparametric generalized linear model. A penalized log-likelihood ratio test statistic is constructed for the null hypothesis of the linearity of the non-parametric function. When the number of knots is fixed, its limiting null distribution is the distribution of a linear combination of independent chi-squared random variables, each with one df. The smoothing parameter is determined by giving a specified value for its asymptotically expected value under the null hypothesis. A simulation study is conducted to evaluate its power performance; a real-life dataset is used to illustrate its practical use.

Keywords

B -spline; Generalized linear model; Generalized partially linear model; Penalized log-likelihood ratio test; Semiparametric generalized linear model

1. Introduction

Generalized linear models (GLM) offer a unifying framework of regression analysis for data from the exponential family of distributions, which include Gaussian, Poisson, and binomial distributions among others (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989). More specifically, in a GLM, the response variable Y is assumed to have a density (or probability mass) function of the following form

$$f(y; \theta, \varphi) = \exp \left(\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right), \quad (1)$$

where φ is a scale or nuisance parameter; θ is the natural parameter of interest; $a(\varphi)$, $b(\theta)$, and $c(y, \varphi)$ are functions of known form, which determine a specific member in the one-parameter exponential family of distributions. It can be shown that the mean response $\mu = E(Y) = b'(\theta)$ and the variance of the response $\text{var}(Y) = a(\varphi)b''(\theta) = a(\varphi)\nu(\mu)$, where $\nu(\mu) = b''(\theta)$ is a variance function depending only on μ . Let u be the covariate of interest and $\mathbf{x} = (x_0, x_1, \dots, x_{p-1})$ the other covariates in the model, where $x_0 = 1$. Under this formulation, we model the mean response $\mu = E(Y) = b'(\theta)$, related to the covariate vector $\mathbf{z} = (\mathbf{x}, u)$, through a monotonic differentiable link function g as follows:

$$g(\mu) = g[b'(\theta)] = \mathbf{x}\alpha + s(u). \quad (2)$$

Here $\alpha = (\alpha_0, \dots, \alpha_{p-1})^T$ is a vector of unknown p model parameters associated with the covariate vector $\mathbf{x}$. s is an unknown but smooth function of u , which is to be estimated from the data, while the effects of $\mathbf{x}$ remain linear. The model in (2) is a generalized partially

linear model and referred to as a semiparametric GLM, because it contains both parametric and nonparametric components.

There have been several methods of estimation proposed for estimating the parameters in (2), including, e.g., splines (Green and Yandell, 1985; Green, 1987; Hastie and Tibshirani, 1990; Green and Silverman, 1994; Ruppert, Wand, and Carroll, 2003), local likelihood (Kauermann and Tutz, 2001), and weighted quasi-likelihood (Severini and Staniswalis, 1994). Non-parametric regression techniques have been recently used for testing lack of fit of parametric regression models; see, e.g., Aerts, Claeskens, and Hart (1999), Cantoni and Hastie (2002), Crainiceanu and Ruppert (2004), Crainiceanu et al. (2005), and Liu and Wang (2004). Additionally, Hart (1997) gave a detailed survey on lack-of-fit tests using nonparametric regression techniques.

In contrast to these methods, we use a flexible alternative. That is, splines with a moderate number of knots are used. Specifically, the unknown functional form of the effect of the covariate u , $s(u)$, is modeled by a linear combination of fixed-knot cubic B -spline basis functions (Schoenberg, 1946; Curry and Schoenberg, 1966). The model fitting is achieved by maximizing the penalized log-likelihood. A second-order difference penalty on the adjacent B -spline coefficients is imposed in order to prevent overfitting while estimating model parameters. In addition to the added modeling flexibility, the proposed model can be used to test for the linear effect of u .

The paper is organized as follows. In Section 2, we introduce the structure of the semiparametric GLM and the method of estimation of the model parameters. A penalized log-likelihood ratio test is constructed for the linear effect of the covariate u . Section 3 presents simulation results to study the power performance of the proposed test. In Section 4, we illustrate the use of the proposed method with a real-life example. Some discussion is given in Section 5. Proofs and technical conditions are given in the Appendix.

2. Method

2.1 Estimation

The unknown smooth function s can be approximated by a variety of spline functions, but the s is approximated by using a fixed-knot cubic spline because it provides the best compromise between computational cost and smoothness. The plus-function terms are intuitive and, hence, the truncated power basis is attractive from a statistical point of view (Smith, 1979). However, the B -spline basis provides a numerically superior alternative basis to the truncated power basis. Therefore, we model s as a linear combination of cubic B -spline basis functions with q preselected knots. We choose the k th knot corresponding to the $\frac{k}{q+1}$ th sample quantile of the distinct values of u 's. Let $B_1(u), \dots, B_{q+4}(u)$ be the cubic B -spline basis functions. See de Boor (2001) for details of computation of B -splines and their mathematical properties. We are interested in testing the hypothesis that the effect of u is linear, i.e., testing $H_0: s(u) = \tau u$ versus $H_a: s(u) \neq \tau u$, where τ is an unknown parameter. The space of cubic B -splines includes the constant and linear functions and the constant is given in the parametric component of the model in (2). Therefore, to model s and easily specify the null hypothesis of the linear effect of the covariate u , the linear term is given separately in (3). Consequently, it is needed to drop any two of the $q + 4$ cubic B -spline basis functions to have the full-rank model

$$s(u) = \tau u + \sum_{k=1}^K \beta_k B_k(u). \quad (3)$$

Here $K = q + 2$. The β_k s are cubic B -spline coefficients.

Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)^T$ and $\mathbf{B}_u = (B_1(u), \dots, B_K(u))$. By writing $s(u) = \tau u + \mathbf{B}_u \boldsymbol{\beta}$, we express the mean response $\mu = b'(\theta)$ in (2) as $g(\mu) = g[b'(\theta)] = \mathbf{x}\boldsymbol{\alpha} + \tau u + \mathbf{B}_u \boldsymbol{\beta}$, which can be written as in vector notation

$$g(\mu) = g[b'(\theta)] = \mathbf{z}\boldsymbol{\xi} + \mathbf{B}_u \boldsymbol{\beta} = \mathbf{A}_z \boldsymbol{\psi}, \quad (4)$$

where $\mathbf{A}_z = (\mathbf{z}, \mathbf{B}_u) = (\mathbf{x}, u, \mathbf{B}_u)$ and $\boldsymbol{\psi} = (\boldsymbol{\xi}^T, \boldsymbol{\beta}^T)^T$ for $\boldsymbol{\xi} = (\boldsymbol{\alpha}^T, \tau)^T$. Let Y_i be the response from the i th subject with the covariate vector $\mathbf{z}_i = (\mathbf{x}_i, u_i)$, and Y_i follows the one-parameter exponential family of distributions as described by (1) with mean $\mu_i = b'(\theta_i)$, $i = 1, 2, \dots, n$. In most applications, the so-called canonical link function is used, i.e., $g(\cdot) = (b')^{-1}(\cdot)$. For example, the model in (2) is a semiparametric Poisson regression model if $g(\cdot)$ is a logarithmic function; a semiparametric logistic regression model if $g(\cdot)$ is a logit function. Therefore, we focus on the case of canonical link function. The model in (4) can be expressed as $\theta_i = g[b'(\theta_i)] = \mathbf{A}_{z_i} \boldsymbol{\psi}$.

By plugging $\theta_i = \mathbf{A}_{z_i} \boldsymbol{\psi}$ into $f(y_i, \theta_i, \varphi)$, $i = 1, \dots, n$, we can obtain the log-likelihood function $\ell(\boldsymbol{\psi}, \varphi) = \sum_{i=1}^n [y_i \mathbf{A}_{z_i} \boldsymbol{\psi} - b(\mathbf{A}_{z_i} \boldsymbol{\psi})] / a(\varphi) + \sum_{i=1}^n c(y_i, \varphi)$. The scale parameter φ does not affect the estimating equation that leads to the estimate of $\boldsymbol{\psi}$ and $a(\varphi)$ may be estimated consistently. Thus, for simplicity and without loss of generality, we take $a(\varphi) = 1$.

Additionally, $\sum_{i=1}^n c(y_i, \varphi)$ does not depend on $\boldsymbol{\psi}$. Therefore, the log-likelihood function for the model in (1) with $s(u) = \tau u + \mathbf{B}_u \boldsymbol{\beta}$ is $\ell(\boldsymbol{\psi}) = \sum_{i=1}^n [y_i \mathbf{A}_{z_i} \boldsymbol{\psi} - b(\mathbf{A}_{z_i} \boldsymbol{\psi})]$. For example, if $Y_i \sim \text{Poisson}(\mu_i)$, $\ell(\boldsymbol{\psi}) = \sum_{i=1}^n [y_i \mathbf{A}_{z_i} \boldsymbol{\psi} - \exp(\mathbf{A}_{z_i} \boldsymbol{\psi})]$. $\ell(\boldsymbol{\psi}) = \sum_{i=1}^n [y_i \mathbf{A}_{z_i} \boldsymbol{\psi} - \log(1 + \exp(\mathbf{A}_{z_i} \boldsymbol{\psi}))]$ if $Y_i \sim \text{Bernoulli}(\mu_i)$. Similarly, one can derive the log-likelihood functions for other members of the one-parameter exponential distribution family.

To estimate $\boldsymbol{\psi}$ and overcome the problem of overfitting the data, we employ the penalized likelihood technique introduced by Good and Gaskins (1971) in the context of nonparametric probability density estimation. Thus, by adding the roughness penalty $-(1/2)\boldsymbol{\beta}^T \boldsymbol{\mathcal{G}} \boldsymbol{\beta}$ to the $\ell(\boldsymbol{\psi})$ we obtain the penalized log-likelihood $\ell_p(\boldsymbol{\psi}, \lambda) = \ell(\boldsymbol{\psi}) - (1/2)\lambda \boldsymbol{\beta}^T \boldsymbol{\mathcal{G}} \boldsymbol{\beta}$. Here $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}^T \boldsymbol{\mathcal{D}}$ is a $K \times K$ matrix, where $\boldsymbol{\mathcal{D}}$ is a matrix representation of the difference operator Δ^2 . $\boldsymbol{\beta}^T \boldsymbol{\mathcal{G}} \boldsymbol{\beta} = \sum_{k=3}^K (\Delta^2 \beta_k)^2$, which is a discrete approximation to $\int s''(t)^2 dt$, where $\Delta^2 \beta_k = \beta_k - 2\beta_{k-1} + \beta_{k-2}$; see Eilers and Marx (1996). The smoothing parameter $\lambda > 0$ controls the tradeoff between fidelity to the data and smoothness of the fitting. Let $\boldsymbol{\mathcal{Q}}$ be a $(p+1+K) \times (p+1+K)$ matrix with zeros in the first $p+1$ rows and columns and $\boldsymbol{\mathcal{G}}$ in the remainder of the matrix. The $\ell_p(\boldsymbol{\psi}, \lambda)$ can then be rewritten as

$$\ell_p(\boldsymbol{\psi}; \lambda) = \ell(\boldsymbol{\psi}) - \frac{1}{2} \lambda \boldsymbol{\psi}^T \boldsymbol{\mathcal{Q}} \boldsymbol{\psi}. \quad (5)$$

For given $\lambda > 0$, let $\hat{\boldsymbol{\psi}}_\lambda = (\hat{\boldsymbol{\xi}}_\lambda, \hat{\boldsymbol{\beta}}_\lambda)$ be the value of $\boldsymbol{\psi}$ maximizing the $\ell_p(\boldsymbol{\psi}, \lambda)$, which is the maximum penalized likelihood estimate (MPLE) of $\boldsymbol{\psi}$ and can be obtained by solving the following weighted least-squares equations $(\mathbf{A}^T \mathbf{W} \mathbf{A} + \lambda \boldsymbol{\mathcal{Q}}) \boldsymbol{\psi} = \mathbf{A}^T \mathbf{W} \mathbf{y}$. Here $\mathbf{A} = [\mathbf{Z} \quad \mathbf{B}]$ is an $n \times (p+1+K)$ matrix for $\mathbf{Z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)^T$ being an $n \times (p+1)$ matrix and $\mathbf{B} = (\mathbf{B}_{u_1}^T, \dots, \mathbf{B}_{u_n}^T)^T$ an $n \times K$ matrix. $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ is an $n \times n$ matrix for $w_i = b''$

$(A_{z_i}\boldsymbol{\psi})$. $\boldsymbol{\mathcal{Y}} = (\tilde{y}_1, \dots, \tilde{y}_n)^T$ is a working response vector for $\tilde{y}_i = A_{z_i}\boldsymbol{\psi} + [y_i - b'(A_{z_i}\boldsymbol{\psi})]/b''(A_{z_i}\boldsymbol{\psi})$.

2.2 A penalized log-likelihood ratio test

It turns out from the model in (3) that testing $H_0: s(u) = \tau u$ is equivalent to testing the hypothesis that $\boldsymbol{\beta} = \mathbf{0}$. It is noted that under $H_0: \boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{\varphi}\boldsymbol{\psi} = \mathbf{0}$ and, hence, $\ell_p(\boldsymbol{\psi}; \lambda) = \ell(\boldsymbol{\psi})$.

Let $\boldsymbol{\xi}_0 = (\alpha_0^T, \tau_0)^T$ be the true value of $\boldsymbol{\xi} = (\boldsymbol{\alpha}^T, \tau)^T$ under $H_0: \boldsymbol{\beta} = \mathbf{0}$. Let $\widehat{\boldsymbol{\psi}}_0 = (\widehat{\boldsymbol{\xi}}_0^T, \mathbf{0}^T)^T$ be the restricted MPLE of $\boldsymbol{\psi}_0 = (\alpha_0^T, \tau_0, \mathbf{0}^T)^T = (\boldsymbol{\xi}_0^T, \mathbf{0}^T)^T$ under $H_0: \boldsymbol{\beta} = \mathbf{0}$. That is, $\widehat{\boldsymbol{\psi}}_0$ is the value of $\boldsymbol{\psi}$ that maximizes the $\ell_p(\boldsymbol{\psi}; \lambda)$ subject to $\boldsymbol{\beta} = \boldsymbol{\mathcal{L}}^T \boldsymbol{\psi} = \mathbf{0}$, where $\boldsymbol{\mathcal{L}} = [\mathbf{0}_{K \times (p+1)} \quad \mathbf{I}_{K \times K}]^T$ is a $(p+1+K) \times K$ matrix for $\mathbf{I}_{K \times K}$ being an identity matrix of order K .

To test $H_0: \boldsymbol{\beta} = \mathbf{0}$, because the model parameters are estimated by maximizing the penalized log-likelihood, by analogy with the unpenalized parametric log-likelihood procedures we construct a penalized log-likelihood ratio test (PLRT) statistic as follows:

$$Q_\lambda = 2[\ell_p(\widehat{\boldsymbol{\psi}}_\lambda; \lambda) - \ell_p(\widehat{\boldsymbol{\psi}}_0; \lambda)], \quad (6)$$

where $\ell_p(\widehat{\boldsymbol{\psi}}_0; \lambda) = \ell(\widehat{\boldsymbol{\psi}}_0) - (1/2)\lambda \widehat{\boldsymbol{\psi}}_0^T \boldsymbol{\mathcal{L}} \widehat{\boldsymbol{\psi}}_0 = \ell(\widehat{\boldsymbol{\psi}}_0)$ because $\boldsymbol{\varphi}\widehat{\boldsymbol{\psi}}_0 = \mathbf{0}$. The main difference between the PLRT statistic Q_λ and the deviance statistic (Hastie and Tibshirani, 1990) is that the rough penalty is not included in the deviance.

Let $\mathbf{I}(\boldsymbol{\psi})$ be the Fisher's information matrix as follows:

$$\mathbf{I}(\boldsymbol{\psi}) = \begin{bmatrix} \mathbf{I}_{\xi\xi}(\boldsymbol{\psi}) & \mathbf{I}_{\xi\beta}(\boldsymbol{\psi}) \\ \mathbf{I}_{\beta\xi}(\boldsymbol{\psi}) & \mathbf{I}_{\beta\beta}(\boldsymbol{\psi}) \end{bmatrix} = E \left[-\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^T} \right] = \sum_{i=1}^n b''(A_{z_i}\boldsymbol{\psi}) \mathbf{A}_{z_i}^T \mathbf{A}_{z_i}.$$

Let $\mathbf{I}_{\beta\beta|\xi}(\boldsymbol{\psi}) = \mathbf{I}_{\beta\beta}(\boldsymbol{\psi}) - \mathbf{I}_{\beta\xi}(\boldsymbol{\psi}) \mathbf{I}_{\xi\xi}^{-1}(\boldsymbol{\psi}) \mathbf{I}_{\xi\beta}(\boldsymbol{\psi})$. The asymptotic results of Q_λ are stated in the following theorem and the proof of the theorem is given in the Appendix.

Theorem 1—Assume that conditions (C_1) and (C_2) in the Appendix are satisfied. Let λ_n be the sequence of smoothing parameters with $\lim_{n \rightarrow \infty} \lambda_n/n$ finite, which may be 0. When the number of knots is fixed, under H_0

$$Q_\lambda \xrightarrow{D} Q = \sum_j \gamma_j \chi_j^2, \quad (7)$$

where the χ_j^2 s are independent chi-squared random variables each with one degree of freedom, and the γ_j s are the eigenvalues of $\lim_{n \rightarrow \infty} \mathbf{I}_{\beta\beta|\xi}(\boldsymbol{\psi}_0) [\mathbf{I}_{\beta\beta|\xi}(\boldsymbol{\psi}_0) + \lambda_n \mathcal{G}]^{-1}$

It can be seen from this theorem that for any $\alpha \in (0, 1)$ the test rejecting H_0 if the observed value of Q_λ exceeds the $100(1 - \alpha)$ th percentile of the distribution of Q has an asymptotic significance level α . There have been several methods developed for calculation of the distribution of Q by, e.g., Imhof (1961), Davies (1973, 1980), Sheil and O'Muircheartaigh (1977), and Castaño-Martínex and López-Blázquez (2005). The algorithm of Davies (1980) is implemented in this work.

Because $Q_\lambda \xrightarrow{D} Q$ under H_0 , one can have the asymptotically expected value of Q_λ under H_0 as follows:

$$\sum \gamma_j = \text{trace} \left\{ \lim_{n \rightarrow \infty} \mathbf{I}_{\beta\beta|\xi}(\boldsymbol{\psi}_0) [\mathbf{I}_{\beta\beta|\xi}(\boldsymbol{\psi}_0) + \lambda_n \mathcal{J}]^{-1} \right\}. \quad (8)$$

The asymptotically expected value would be the degrees of freedom of the test for the case of unpenalized log-likelihood (i.e., $\lambda = 0$). Although the Q_λ in (6) does not have an asymptotic chi-squared distribution when $\lambda = 0$, for simplicity and, hence, by following other researchers (e.g., Buja, Hastie, and Tibshirani, 1989), we adopt the terminology, degrees of freedom, for the test. We do not specify a value for the λ directly. Instead, in practice we give a specified value for the degrees of freedom, replace $\boldsymbol{\psi}_0$ with $\hat{\boldsymbol{\psi}}_0$ in (8), and find the value of λ , denoted by $\hat{\lambda}$, which gives the specified degrees of freedom. Therefore, $Q_{\hat{\lambda}} = 2[\ell_p(\hat{\boldsymbol{\psi}}_{\hat{\lambda}}; \hat{\lambda}) - \ell_p(\hat{\boldsymbol{\psi}}_0; \hat{\lambda})]$ is used as a test statistic.

3. Simulation study

We conducted a Monte Carlo simulation study to assess the performance of the proposed PLRT in Section 2. We assumed the covariate vector takes the form $\mathbf{z}_i = (\mathbf{x}_i, u_i)$, $i = 1, \dots, 100$. Here $\mathbf{x}_i = (x_{i0}, x_{i1})$, where $x_{i0} = 1$ and x_{i1} were generated from the uniform $[0.5, 1.5]$ distribution. u_i were generated from the uniform $[0, 1]$ distribution. One thousand replications were conducted for each experimental configuration and all tests used nominal level $\alpha = 0.05$.

To assess the empirical power of the test, we generated data y_i from the following model

$$\log[\mu(\mathbf{z}_i; \boldsymbol{\xi}, b)] = \alpha_0 + \alpha_1 x_{i1} + s(u_i; \boldsymbol{\tau}, b), \quad i = 1, 2, \dots, 100. \quad (9)$$

Here $\boldsymbol{\xi} = (\mathbf{a}^T, \boldsymbol{\tau}^T)^T$ for $\mathbf{a} = (\alpha_0, \alpha_1) = (1, 0.7)^T$. $s(u_i; \boldsymbol{\tau}, b) = \boldsymbol{\tau} u_i + b \cos(3\pi u_i)$, where $\boldsymbol{\tau} = 1.5$ and $b \cos(3\pi u_i)$ are used as alternatives for $b = 0, 0.02, \dots, 0.28$. It is noted that when $b = 0$, the model in (9) produces the null model

$$\log[\mu(\mathbf{z}_i)] = \alpha_0 + \alpha_1 x_{i1} + \boldsymbol{\tau} u_i = \log[\mu(\mathbf{z}_i; \boldsymbol{\xi}, 0)]. \quad (10)$$

The goal was to test $H_0: s(u) = \boldsymbol{\tau} u = s(u; \boldsymbol{\tau}, 0)$, which is equivalent to testing $b = 0$.

We also computed two log-likelihood ratio test (LRT) statistics for $H_0: b = 0$ to compare their power performances with the PLRT. The first LRT statistic for H_0 is

$Q_1 = 2[\ell(\hat{\boldsymbol{\xi}}, \hat{b}) - \ell(\hat{\boldsymbol{\xi}})] \xrightarrow{D} \chi^2(1)$ that has a chi-squared distribution with one degree-of-freedom. Here $(\hat{\boldsymbol{\xi}}, \hat{b})$ is the value of $(\boldsymbol{\xi}, b)$ maximizing the

$\ell(\boldsymbol{\xi}, b) = \sum_{i=1}^n [y_i(z_i \boldsymbol{\xi} + b \cos(3\pi u_i)) - \exp(z_i \boldsymbol{\xi} + b \cos(3\pi u_i))]$ for the data generating model in (9).

$\hat{\boldsymbol{\xi}}$ is the value of $\boldsymbol{\xi}$ maximizing the $\ell(\boldsymbol{\xi}) = \sum_{i=1}^n [y_i(z_i \boldsymbol{\xi}) - \exp(z_i \boldsymbol{\xi})]$ for the null model in (10).

The test gives an asymptotically optimal test that provides a useful comparison benchmark but is of limited practical value because its use requires very specific knowledge of the form of the alternative. We rejected H_0 at $\alpha = 0.05$ when the observed value of Q_1 exceeded $\chi_{0.95}^2(1)$ that denotes the 0.95-quantile of the chi-squared distribution with one degree-of-freedom.

To examine the power loss when the alternative parametric model is mis-specified, we fit data generated from the model in (9) by the misspecified alternative parametric model

$$\log[\mu(z_i; \xi, c)] = \alpha_0 + \alpha_1 x_{1i} + \tau u_i + c u_i^2. \quad (11)$$

The second LRT statistic for H_0 is $Q_2 = 2[\ell(\tilde{\xi}, \tilde{c}) - \ell(\tilde{\xi})] \xrightarrow{D} \chi^2(1)$. Here $(\tilde{\xi}, \tilde{c})$ is the value of (ξ, c) maximizing the $\ell(\xi, c) = \sum_{i=1}^n [y_i(z_i \xi + c u_i^2) - \exp(z_i \xi + c u_i^2)]$ for the misspecified model in (11). H_0 was rejected at $\alpha = 0.05$ when the observed value of Q_2 exceeded $\chi_{0.95}^2(1)$.

To study the type I error and power performance of the PLRT Q_λ under the different numbers of knots and different degrees of freedom, we considered 4, 4.5, and 5 degrees of freedom, each with $q = 10, 15, 20$, and 25 knots, to determine the value of λ while fitting the B -spline model. Under each of the combinations of different degrees of freedom and the different numbers of knots, the empirical significance level of the PLRT appeared to agree with the 0.05 nominal level. The empirical powers are shown in Figure 1. The power performances of the PLRTs, among the aforementioned combinations, are essentially the same. The first LRT, Q_1 , had the best power performance; however, the PLRT was competitive with the test and still retained the advantage of having the ability to detect more general alternatives. The second LR test, Q_2 , suffered a severe power loss and had the worst power performance among these tests; thus, it indicated the parametric test Q_2 was not optimal under the case where a misspecified parametric regression model was used to fit data.

4. Example

We illustrate the practical use of the methodology with the data set from a study of 173 nesting horseshoe crabs (Brockmann, 1996; Agresti, 2002). In the study, each female horseshoe crab had a male crab attached to her in her nest. The study investigated factors that affect whether the female crab had any other males, called satellites, residing nearby her. Explanatory variables included the female crab's color, which is light medium, medium, dark medium, or dark, spine condition, which is both good, one worn or broken, or both worn or broken, carapace width (W) in centimeters, and weight (Wt) in kgs. The response outcome for each female crab is her number of satellites (Sa).

We fit the Poisson log-linear regression to the data to produce the fitted log-linear regression model $\log \hat{\mu} = -0.801 + 0.531 \times I_{\{light\ medium\}} + 0.266 \times I_{\{medium\}} + 0.018 \times I_{\{dark\ medium\}} - 0.087 \times I_{\{both\ good\}} - 0.238 \times I_{\{one\ worn\ or\ broken\}} + 0.017 \times W + 0.497 \times Wt$, where

$\hat{\mu} = E(Sa)$. We employ the PLRT to evaluate whether there is a linear relation between the logarithm of the expected number of satellites, $\log \mu = \log E(Sa)$, and female crab's weight. Because the p-values of the PLRT associated with 4, 4.5, and 5 degrees of freedom, each calculated by using 10, 15, 20, and 25 knots, are less than 0.0035, the female crab's weight has a statistically significantly nonlinear effect on the logarithm of the expected number of satellites. Figures 2 and 3 show a cubic spline-fitted curve for the effect of female crab's weight obtained with each of the above combinations of different degrees of freedom and the different numbers of knots.

5. Discussion

A flexible method has been proposed for modeling the nonparametric component of the semiparametric generalized linear model in (2) via a linear combination of fixed-knot cubic B -splines with the second-order difference penalty on the adjacent cubic B -spline

coefficients to prevent overfitting to the data. Although automatic smoothing parameter selection methods such as cross-validation and generalized cross-validation, can be considered in the estimation procedure, we give a specified value for the degrees of freedom, and find the value of λ that gives the specified degrees of freedom. For comments on using fixed smoothing parameters, see, e.g., Buja et al. (1989, p. 545–546).

The focus is only on testing $H_0: s(u) = \tau u$, but the idea also can be used to test the null hypothesis of no effect of u , $H_0: s(u) = 0$, i.e., $(\tau, \beta^T)^T = \mathbf{0}$. It is exactly the same as constructing the PLRT statistic for $H_0: s(u) = \tau u$ to construct one for $H_0: s(u) = 0$. One can extend the methodology to examine the effects of several covariates of interest on the response variable.

Acknowledgments

The author is very grateful to the Editor and a referee whose helpful comments improved the presentation. This publication was made possible by Grant Number UL1 RR024146 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

References

- Agresti, A. Categorical Data Analysis. 2. New York: John Wiley and Sons; 2002.
- Aerts M, Claeskens G, Hart JD. Testing the fit of a parametric function. *Journal of the American Statistical Association*. 1999; 94:869–879.
- Brockmann HJ. Satellite male groups in horseshoe crabs, *Limulus polyphemus*. *Ethology*. 1996; 102:1–21.
- Buja A, Hastie T, Tibshirani R. Linear smoother and additive models (with Discussion). *Ann Statist*. 1989; 17:453–555.
- Cantoni E, Hastie T. Degrees-of-freedom tests for smoothing splines. *Biometrika*. 2002; 289:251–263.
- Castaño-Martínez A, López-Blázquez F. Distribution of a sum of weighted central chi-square variables. *Communications in Statistics: Theory and Methods*. 2005; 34:515–524.
- Crainiceanu CM, Ruppert D. Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*. 2004; 91:3552.
- Crainiceanu CM, Ruppert D, Claeskens G, Wand MP. Exact likelihood ratio tests for penalised splines. *Biometrika*. 2005; 92:91–103.
- Curry HB, Schoenberg IJ. On Pólya frequency functions IV: The fundamental splines and their limits. *J Analyse Math*. 1966; 17:71–107.
- Davies RB. Numerical inversion of a characteristic function. *Biometrika*. 1973; 60:415–417.
- Davies RB. The distribution of a linear combination of χ^2 random variables. *Applied Statistics*. 1980; 29:323–333.
- de Boor, C. *Applied Mathematical Sciences*. Vol. 27. New York: Springer-Verlag; 2001. A Practical Guide to Splines. Revised edition
- Eilers PHC, Marx BD. Flexible smoothing with B -splines and penalties. *Statist Sci*. 1996; 11:89–121.
- Good IJ, Gaskins RA. Nonparametric roughness penalties for probability densities. *Biometrika*. 1971; 58:255–277.
- Green, PJ.; Yandell, BS. *Lecture Notes in Statistics*. Vol. 32. Berlin: Springer; 1985. Semi-parametric generalized linear models; p. 44–55.
- Green PJ. Penalized likelihood for general semi-parametric regression models. *International Statistical Review*. 1987; 55:245–259.
- Green, PJ.; Silverman, BW. *Nonparametric Regression and Generalized Linear Models*. New York: Chapman & Hall; 1994.
- Hart, JD. *Nonparametric Smoothing and Lack-of-fit Tests*. New York: Springer Verlag; 1997.
- Hastie, TJ.; Tibshirani, RJ. *Generalized Additive Models*. New York: Chapman & Hall; 1990.

- Imhof JP. Computing the distribution of quadratic forms in normal variables. *Biometrika*. 1961; 48:419–426.
- Kauermann G, Tutz G. Testing generalized linear and semi-parametric models against smooth alternatives. *Journal of Royal Statistical Society, Series B*. 2001; 63:147–166.
- Liu A, Wang Y. Hypothesis testing in smoothing spline models. *Journal of Statistical computation and simulation*. 2004; 74:581–597.
- McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2. London: Chapman & Hall; 1989.
- Nelder JA, Wedderburn RWM. Generalized linear models. *Journal of Royal Statistical Society, Series A*. 1972; 135:370–384.
- Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. Cambridge University Press; 2003.
- Scheffé, H. *The Analysis of Variance*. New York: John Wiley and Sons; 1959.
- Schoenberg IJ. Contributions to the problem of approximation of equidistant data by analytic functions. *Quart Appli Math*. 1946; 4:45–99. 112–141.
- Severini TA, Staniswalis JG. Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*. 1994; 89:501–511.
- Sheil J, O’Muircheartaigh I. The distribution of non-negative quadratic forms in normal variables. *Applied Statistics*. 1977; 26:92–98.
- Smith PL. Splines as a useful and convenient statistical tool. *Am Stat*. 1979; 33:57–62.

Appendix

In this appendix, we set out the essential steps for proving Theorem 1. The following conditions are required.

$$(C_1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} I(\psi) \text{ is finite and positive definite.}$$

$$(C_2) \quad \sup_{\substack{\psi^* \in B(\delta) \\ R^{p+1+K}: \|\psi^* - \psi\| < \delta}} \frac{1}{n} \sum_{i=1}^n \| [b''(A_{z_i} \psi^*) - b''(A_{z_i} \psi)] A_{z_i}^T A_{z_i} \| \rightarrow 0 \quad \text{as } \delta \downarrow 0, \text{ where } B(\delta) = \{ \psi^* \in R^{p+1+K}: \|\psi^* - \psi\| < \delta \}, \delta \downarrow 0 \text{ is an infinitesimal neighborhood of } \psi.$$

Let $\widehat{\psi}_{\lambda_n} = (\widehat{\xi}_{\lambda_n}^T, \widehat{\beta}_{\lambda_n}^T)^T$ be the value of ψ that maximizes $\ell_p(\psi; \lambda) = \ell(\psi) - (1/2)\lambda \psi^T \mathfrak{Q} \psi$ for a sample of size n when $\lambda = \lambda_n$. Since $\ell_p(\psi; \lambda_n)/\psi^T \widehat{\psi}_{\lambda_n} = 0$, by using (C_1) , we consider a Taylor expansion of $\ell_p(\psi; \lambda_n)/\psi^T \widehat{\psi}_{\lambda_n}$ around ψ_0 to have

$$0 = \partial \ell_p(\psi; \lambda_n)/\partial \psi|_{\psi=\psi_0} + \partial^2 \ell_p(\psi; \lambda_n)/(\partial \psi \partial \psi^T)|_{\psi=\psi_0} (\widehat{\psi}_{\lambda_n} - \psi_0) + o_p(\sqrt{n}),$$

where $o_p(a_n)$ indicates a vector whose components are uniformly $o_p(a_n)$. Because $\mathfrak{Q} \psi = 0$ under $H_0: \beta = 0$, $\ell_p(\psi; \lambda_n)/\psi^T \widehat{\psi}_{\lambda_n} = \ell(\psi)/\psi^T \widehat{\psi}_{\lambda_n} = U(\psi_0)$. Thus, it implies that under H_0

$$\sqrt{n}(\widehat{\psi}_{\lambda_n} - \psi_0) = n[I(\psi_0) + \lambda_n \mathfrak{Q}]^{-1} n^{-1/2} U(\psi_0) + o_p(1). \quad (12)$$

By using the assumption that $\lim_{n \rightarrow \infty} \lambda_n/n$ is finite, Liapounov’s Central Limit Theorem, and Cramér-Wold Theorem, one can show $n^{-1/2} U(\psi_0) \xrightarrow{D} N_{p+1+K}(\mathbf{0}, \lim_{n \rightarrow \infty} n^{-1} I(\psi_0))$ under H_0 . Therefore, using Slutsky’s Theorem and (C_2) can show

$\sqrt{n}(\widehat{\psi}_{\lambda_n} - \psi_0) \xrightarrow{D} N_{p+1+K}(\mathbf{0}, \lim_{n \rightarrow \infty} n V_{\lambda_n}(\psi_0))$ under H_0 , where $V_{\lambda_n}(\psi_0) = [I(\psi_0) + \lambda_n \mathfrak{Q}]^{-1} I(\psi_0) [I(\psi_0) + \lambda_n \mathfrak{Q}]^{-1}$. Then,

$$\sqrt{n} \widehat{\beta}_{\lambda_n} = \mathcal{L}^T \sqrt{n}(\widehat{\psi}_{\lambda_n} - \psi_0) \xrightarrow{D} N_K \left(\mathbf{0}, \lim_{n \rightarrow \infty} n V_{\beta \beta; \lambda_n}(\psi_0) \right), \quad (13)$$

where $V_{\beta \beta; \lambda_n}(\psi_0) = [I_{\beta \beta} \xi(\psi_0) + \lambda_n \mathfrak{Q}]^{-1} I_{\beta \beta} \xi(\psi_0) [I_{\beta \beta} \xi(\psi_0) + \lambda_n \mathfrak{Q}]^{-1}$.

Under H_0 : $\boldsymbol{\zeta}^T \boldsymbol{\psi} = \boldsymbol{\beta} = \mathbf{0}$, $\hat{\boldsymbol{\psi}}_0$ is the restricted MPLE of $\boldsymbol{\psi}_0$, so $\hat{\boldsymbol{\psi}}_0$ must satisfy the restricted penalized log-likelihood equations

$$\left. \frac{\partial}{\partial \boldsymbol{\psi}} \{ \ell_p(\boldsymbol{\psi}; \lambda_n) + (\mathcal{L}^T \boldsymbol{\psi} - \mathbf{0}) \mathbf{d} \} \right|_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_0} = \left. \frac{\partial}{\partial \boldsymbol{\psi}} \ell_p(\boldsymbol{\psi}; \lambda_n) \right|_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_0} + \mathcal{L} \mathbf{d} = \mathbf{0}, \quad (14)$$

$$\mathcal{L}^T \hat{\boldsymbol{\psi}}_0 = \mathbf{0},$$

where $\mathbf{d}$ is a $K \times 1$ vector of Lagrange multipliers. We use (C_1) and $\mathfrak{Q} \boldsymbol{\psi}_0 = \mathbf{0}$ to consider the Taylor expansion $n^{1/2} \ell_p(\boldsymbol{\psi}; \lambda_n) / \partial \boldsymbol{\psi} |_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_0} = n^{1/2} \mathbf{U}(\boldsymbol{\psi}_0) - n^{-1} [\mathbf{I}(\boldsymbol{\psi}_0) + \lambda_n \mathfrak{Q}] \sqrt{n}(\hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0) + o_p(1)$. In addition, $\sqrt{n} \mathcal{L}^T \hat{\boldsymbol{\psi}}_0 = \sqrt{n} \mathcal{L}^T (\hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0)$. One can then express $\sqrt{n}(\hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0)$ as

$$\sqrt{n}(\hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0) = n \mathcal{L} [\mathcal{L}^T (\mathbf{I}(\boldsymbol{\psi}_0) + \lambda_n \mathfrak{Q}) \mathcal{L}]^{-1} \mathcal{L}^T n^{-1/2} \mathbf{U}(\boldsymbol{\psi}_0) + o_p(1),$$

where $\boldsymbol{\zeta} = [\mathbf{I}_{(p+1) \times (p+1)} \mathbf{0}_{K \times (p+1)}]^T$ is a $(p+1+K) \times (p+1)$ matrix. Because by using a second-order Taylor expansion of $\ell_p(\hat{\boldsymbol{\psi}}_{\lambda_n}; \lambda_n)$ and $\ell_p(\hat{\boldsymbol{\psi}}_0; \lambda_n)$, one can have $2[\ell_p(\hat{\boldsymbol{\psi}}_{\lambda_n}; \lambda_n) - \ell_p(\boldsymbol{\psi}_0; \lambda_n)] = \mathbf{U}^T(\boldsymbol{\psi}_0) [\mathbf{I}(\boldsymbol{\psi}_0) + \lambda_n \mathfrak{Q}]^{-1} \mathbf{U}(\boldsymbol{\psi}_0) + o_p(1)$ and $2[\ell_p(\hat{\boldsymbol{\psi}}_0; \lambda_n) - \ell_p(\boldsymbol{\psi}_0; \lambda_n)] = \mathbf{U}^T(\boldsymbol{\psi}_0) \boldsymbol{\zeta} \boldsymbol{\zeta}^T [\mathbf{I}(\boldsymbol{\psi}_0) + \lambda_n \mathfrak{Q}]^{-1} \boldsymbol{\zeta}^T \mathbf{U}(\boldsymbol{\psi}_0) + o_p(1)$ under H_0 , using (12) can express

$Q_{\lambda_n} = 2[\ell_p(\hat{\boldsymbol{\psi}}_{\lambda_n}; \lambda_n) - \ell_p(\hat{\boldsymbol{\psi}}_0; \lambda_n)] = \widehat{\boldsymbol{\beta}}_{\lambda_n}^T [\mathbf{I}_{\beta\beta\xi}(\boldsymbol{\psi}_0) + \lambda_n \mathcal{G}] \widehat{\boldsymbol{\beta}}_{\lambda_n} + o_p(1)$. By means of a non-singular linear transformation (Scheffé, 1959, p. 418) and Slutsky's Theorem, one can show

$Q_{\lambda_n} \xrightarrow{D} Q$ under H_0 . Let $\hat{\lambda}$ be an estimator satisfying $\hat{\lambda}/\lambda_n \xrightarrow{P} 1$. From (13), we can show via

Slutsky's Theorem that $\sqrt{n} \widehat{\boldsymbol{\beta}}_{\hat{\lambda}} \xrightarrow{D} N_K(\mathbf{0}, \lim_{n \rightarrow \infty} n \mathbf{V}_{\beta\beta\xi; \lambda_n}(\boldsymbol{\psi}_0))$ under H_0 . Then, $\hat{\boldsymbol{\beta}}_{\hat{\lambda}}$ and $\hat{\boldsymbol{\beta}}_{\lambda_n}$

have the same limiting null distribution. Let $Q_{1, \lambda_n} = \widehat{\boldsymbol{\beta}}_{\lambda_n}^T [\mathbf{I}_{\beta\beta\xi}(\boldsymbol{\psi}_0) + \lambda_n \mathcal{G}] \widehat{\boldsymbol{\beta}}_{\lambda_n}$ and

$Q_{1, \hat{\lambda}} = \widehat{\boldsymbol{\beta}}_{\hat{\lambda}}^T [\mathbf{I}_{\beta\beta\xi}(\boldsymbol{\psi}_0) + \hat{\lambda} \mathcal{G}] \widehat{\boldsymbol{\beta}}_{\hat{\lambda}}$. It follows via the Slutsky's Theorem that $Q_{\lambda_n} = Q_{1, \lambda_n} + o_p(1)$, Q_{1, λ_n} and $Q_{1, \hat{\lambda}}$ have the same limiting null distribution. Thus, one can show that $Q_{\hat{\lambda}} = 2[\ell_p(\hat{\boldsymbol{\psi}}_{\hat{\lambda}}; \hat{\lambda}) - \ell_p(\hat{\boldsymbol{\psi}}_0; \hat{\lambda})] = Q_{1, \hat{\lambda}} + o_p(1)$, $Q_{1, \hat{\lambda}}$, Q_{1, λ_n} and Q_{λ_n} have the same limiting null distribution.

Figure 1.

Empirical power curve comparison of the PLRT ($d; q$), LRT₁ (i.e., Q_1), and LRT₂ (i.e., Q_2) for the null model $s(u) = \tau u$ versus the alternative model $s(u) = \tau u + b \cos(3\pi u)$.

Abbreviations: PLRT ($d; q$), penalized log-likelihood ratio test associated d degrees of freedom and q knots; LRT₁, log-likelihood ratio test with data generating model $s(u) = \tau u + b \cos(3\pi u)$ as alternative model; LRT₂, log-likelihood ratio test with misspecified model $s(u) = \tau u + cu^2$ as alternative model.

Figure 2. Cubic spline-fitted curves for the effect of female crab's weight by using 4, 4.5, and 5 degrees of freedom with 10, 15, 20, and 25 knots.

Figure 3.

Cubic spline-fitted curves for the effect of female crab's weight by using 4, 4.5, and 5 degrees of freedom with 10, 15, 20, and 25 knots.