Likelihood Ratio Tests for Goodness-of-Fit of a Nonlinear Regression Model

Ciprian M. Crainiceanu* David Ruppert†

April 2, 2004

Abstract

We propose likelihood and restricted likelihood ratio tests for goodness-of-fit of nonlinear regression. The first order Taylor approximation around the MLE of the regression parameters is used to approximate the null hypothesis and the alternative is modeled nonparametrically using penalized splines. The exact finite sample distribution of the test statistics is obtained for the linear model approximation and can be easily simulated. We recommend using the restricted likelihood instead of the likelihood ratio test because restricted maximum likelihood estimates are not as severely biased as the maximum likelihood estimates in the penalized splines framework.

Short title: LRTs for nonlinear regression

Keywords: Fan-Huang goodness-of-fit test, mixed models, Nelson-Siegel model for yield curves, penalized splines, REML.

1 Introduction

Nonlinear regression models used in applications arise either from underlying theoretical principles or from the trained subjective choice of the statistician. From the perspective of goodness-of-fit testing these are null models and, once data become available, statistical testing may be used to validate or invalidate the original assumptions.

There are three main approaches to goodness-of-fit testing of a null regression model. The standard approach is to nest the null parametric model into a parametric supermodel (sometimes called the "full model") that is assumed to be true and use likelihood ratio tests (LRTs). This

*Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD, 21205 USA. Email: ccrainic@jhsph.edu

†School of Operational Research and Industrial Engineering, Cornell University, Rhodes Hall, NY 14853, USA. E-mail: dr24@cornell.edu

approach is very popular in linear regression and it makes sense when a *correct* supermodel is available. However, in some situations it is impossible to find a *correct* alternative parametric model.

A second approach is to obtain residuals from the regression analysis under the null model and base the test statistics on the estimated residuals. Brown, Durbin and Evans (1975) proposed the well known CUSUM statistic based on recursive residuals. Similar in spirit to the CUSUM statistic, the QS statistic is based on the least squares residual estimates and was proposed by Gardner (1969), extended by MacNeill (1978), and further extended by Perron (1991) for testing departures from a polynomial trend of a time series. Stute (1997) introduced a test using a marked empirical process based on the residuals, Diebolt and Zuber (1999) developed a test for nonlinear heteroscedastic regression models, and He and Zhou (2003) use quantile regression based on a CUSUM process of the gradient vector. Another class of tests based on residuals are the the Von Neumann type statistics (Hart, 1997). The appeal of such tests is that they do not require the specification of an alternative model and the hope is that they detect fairly general forms of lack of fit. However, tests based only on the null hypothesis ignore information about possible alternatives and thus may lose power with respect to parametric tests.

Our approach is to embed the parametric regression into a larger, semiparametric family (e.g., the parametric nonlinear regression model plus a nonparametric spline). This approach has been used for testing polynomial regressions by Cleveland and Devlin (1988), Azzalini and Bowman (1993), Hart (1997), Härdle, Mammen and Müller (1998), Crainiceanu and Ruppert (2004), Crainiceanu, Ruppert, Claeskens and Wand (2004), but the extension to nonlinear regression appears to be new. The success of such an approach relies on the fact that under the null hypothesis the fit is fully parametric and the distribution of the test statistic is usually easy to obtain. Also, by specifying a flexible semiparametric alternative the power of tests could be improved, though a power comparison would be needed to verify this conjecture.

In this paper we develop likelihood and restricted likelihood ratio tests, (R)LRTs, for testing the null hypothesis of nonlinear regression against a general alternative modeled nonparametrically. By nonlinear regression we mean that the conditional mean of the response variable given the covariates has a known functional form that is nonlinear in the parameters. In nonparametric regression the mean function has an unknown functional form and in this paper is modeled using basis functions (e.g., truncated power functions or B-splines for penalized splines). We propose to use a first order

Taylor approximation of the nonlinear function to obtain an approximate linear model and then use a LRT or RLRT for linearity as described, for example, by Crainiceanu and Ruppert (2004). We find that LRTs have poor power because the MLE strongly underestimates the smoothing parameter of penalized splines. Therefore, we recommend using the RLRT rather than the LRT.

Linearization of nonlinear regression models using a first order Taylor expansion of the conditional mean around the nonlinear MLE has been the standard tool for obtaining approximate confidence intervals for regression parameters (e.g. Bates and Watts, 1988; Seber and Wild, 1988). In this paper we use the linear Taylor approximation as the null model in a goodness-of-fit testing procedure. Deciding whether this approximation makes sense for a given data set has to be based on parametric tests of non-nested competing models. We recommend plotting the nonlinear MLE fit and its linear Taylor approximation on the same graph. Very often the two curves are visually indistinguishable, provided that a large enough sample is available and the nonlinear curve is smooth enough.

2 Testing for a nonlinear regression against a general alternative

Suppose that we want to test

$$H_0 : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{f}(\boldsymbol{\mathcal{X}}, \boldsymbol{\delta}) + \boldsymbol{\epsilon}\,, \tag{1}$$

where $\boldsymbol{L}$ is the matrix of covariates that enter the model linearly, $\boldsymbol{f}(\boldsymbol{\mathcal{X}}, \boldsymbol{\delta}) = \{f(\boldsymbol{x}_1, \boldsymbol{\delta}), \ldots, f(\boldsymbol{x}_n, \boldsymbol{\delta})\}^T$, $f(\boldsymbol{x}, \boldsymbol{\delta})$ is a nonlinear function in the parameters $\boldsymbol{\delta}$, n is the total number of observations, $\boldsymbol{\mathcal{X}}$ has ith row equal $\boldsymbol{x}_i^T$ which is the i-th value of the covariate $\boldsymbol{x}$, and $\boldsymbol{\epsilon}$ has a $N(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}_n)$ distribution. Denote by $\boldsymbol{\beta} = (\boldsymbol{\gamma}^T, \boldsymbol{\delta}^T)^T$ and by $\widehat{\boldsymbol{\beta}}_n = \left(\widehat{\boldsymbol{\gamma}}_n^T, \widehat{\boldsymbol{\delta}}_n^T\right)^T$ the MLE of $\boldsymbol{\beta}$ for model (1).

Our testing methodology consists of the following steps:

Step 1. Obtain the MLEs of parameters under the nonlinear regression model (1).

Step 2. Use a first order (linear) Taylor approximation of the nonlinear regression function around the MLEs of the parameters.

Step 3. Define the test for H_0 against a general alternative as the (R)LRT for the linear approximation at Step 2 against a general alternative. The variability of MLE's is ignored and critical values are obtained using finite sample results derived by Crainiceanu and Ruppert (2004) for the linear case.

The linear Taylor approximation of $f(\boldsymbol{x}, \boldsymbol{\delta})$ is

$$m(\boldsymbol{x}, \boldsymbol{\delta}) = f(\boldsymbol{x}, \widehat{\boldsymbol{\delta}}_n) + \left\{ \frac{\partial}{\partial \delta} f\left(\boldsymbol{x}, \widehat{\boldsymbol{\delta}}_n\right) \right\}^T \left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}_n\right) . \tag{2}$$

The linear part of the regression described in model (1) is left unchanged because the first order Taylor approximation does not affect linear functions. In the following we will treat $\widehat{\boldsymbol{\delta}}_n$ as a constant, in which case $m(\boldsymbol{x}, \boldsymbol{\delta})$ is a linear function in $\boldsymbol{\delta}$. We replace the null hypothesis (1) with

$$H_0 : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{m}(\boldsymbol{\mathcal{X}}, \boldsymbol{\delta}) + \boldsymbol{\epsilon} . \tag{3}$$

where $\boldsymbol{m}(\boldsymbol{\mathcal{X}}, \boldsymbol{\delta}) = \{m(\boldsymbol{x}_1, \boldsymbol{\delta}), \ldots, m(\boldsymbol{x}_n, \boldsymbol{\delta})\}^T$. Let t be the number of δ parameters, $\boldsymbol{W}$ the $n \times t$ dimensional matrix with the i-th row equal to

$$\boldsymbol{W}_i^T = \left\{ \frac{\partial}{\partial \delta_1} f(\boldsymbol{x}_i, \widehat{\boldsymbol{\delta}}_n), \ldots, \frac{\partial}{\partial \delta_t} f(\boldsymbol{x}_i, \widehat{\boldsymbol{\delta}}_n) \right\} ,$$

and $\boldsymbol{y} = \boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\mathcal{X}}, \widehat{\boldsymbol{\delta}}_n) + \boldsymbol{W}\widehat{\boldsymbol{\delta}}_n$. The null model (3) is approximated by the following linear model

$$\boldsymbol{y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{\epsilon} . \tag{4}$$

The second step is to embed the function $m(\boldsymbol{x}, \boldsymbol{\delta})$ into a larger class of functions and use R(LRT)s for testing model (3) against a general alternative. The linearization (2) is needed to define RLRTs and could be omitted if only LRTs were used. However, as will be shown, for the hypotheses we will consider LRTs have serious problems, and the use of RLRTs is highly desirable.

For simplicity, we now assume that the covariate is one-dimensional and denote it by x instead of $\boldsymbol{x}$. The extension to the multivariate case will be discussed later. To model the alternative we consider the class of spline functions which is flexible enough to describe a large class of functions and suitable for testing

$$s\left(x, \boldsymbol{\theta}\right) = \alpha_0 + \alpha_1 x + \ldots + \alpha_p x^p + \sum_{k=1}^{K} b_k \left(x - \kappa_k\right)_+^p , \tag{5}$$

where $\theta = (\alpha_0, \ldots, \alpha_p, b_1, \ldots, b_K)^T$ is the vector of regression coefficients, and $\kappa_1 < \kappa_2 < \ldots < \kappa_K$ are fixed knots. Following Ruppert (2002), we consider a number of knots that is large enough (typically 5 to 20) to ensure the desired flexibility, and κ_k is the sample quantile of x's corresponding to probability $k/(K+1)$, but results hold for any other choice of knots.

The basic idea is to test the null model (3) against the alternative model

$$H_A : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{m}(\boldsymbol{\mathcal{X}}, \boldsymbol{\delta}) + \boldsymbol{s}(\boldsymbol{\mathcal{X}}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} , \tag{6}$$

where $\boldsymbol{s}(\boldsymbol{\mathcal{X}}, \boldsymbol{\theta}) = \{s(x_1, \boldsymbol{\theta}), \ldots, s(x_n, \boldsymbol{\theta})\}^T$. To avoid unidentifiable models under the alternative, the function $s(x, \boldsymbol{\theta})$ includes only those monomials that do not appear already in model (4). For example, if $\boldsymbol{L\gamma}$ contains an intercept then the spline function will not contain the monomial of degree zero.

To avoid overfitting the data, the criterion to be minimized is a penalized sum of squares

$$\sum_{i=1}^{n} \left\{ y_i - \boldsymbol{L}_i^T \boldsymbol{\gamma} - \boldsymbol{W}_i^T \boldsymbol{\delta} - s\left(x_i, \boldsymbol{\theta}\right) \right\}^2 + \frac{1}{\lambda} \boldsymbol{\theta}^T \boldsymbol{D} \boldsymbol{\theta} \,, \tag{7}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$, λ is the smoothing parameter, $\boldsymbol{L}_i^T$ is the i-th row of the matrix $\boldsymbol{L}$, and $\boldsymbol{D}$ is a known positive semi-definite matrix. The penalty $\int \left\{ m^{(2)}(x, \boldsymbol{\theta}) \right\}^2 dx$ used for smoothing splines can be achieved with $\boldsymbol{D}$ equal to the sample second moment matrix of the second derivatives of the spline basis functions. However, in this paper we focus on matrices $\boldsymbol{D}$ of the form

$$\boldsymbol{D} = \begin{bmatrix} \mathbf{0}_{p+1 \times p+1} & \mathbf{0}_{p+1 \times K} \\ \mathbf{0}_{K \times p+1} & \boldsymbol{\Sigma}^{-1} \end{bmatrix},$$

where $\boldsymbol{\Sigma}$ is a known positive definite matrix and $\mathbf{0}_{ml}$ is an $m \times l$ matrix of zeros. This type of matrix $\boldsymbol{D}$ penalizes the coefficients of the spline basis functions $(x - \kappa_k)_+^p$ only and will be used in the remainder of the paper. A standard choice is $\boldsymbol{\Sigma} = \boldsymbol{I}_K$ but other matrices can be used according to the specific application.

Let $\boldsymbol{X}$ be the $n \times (p+1)$ matrix with the i-th row $\boldsymbol{X}_i = (1, x_i, \ldots, x_i^p)$ with some of the columns possibly deleted if they already appear in $\boldsymbol{L}$, and $\boldsymbol{Z}$ be the $n \times K$ matrix with i-th row $\boldsymbol{Z}_i = \left\{ (x_i - \kappa_1)_+^p, \ldots, (x_i - \kappa_K)_+^p \right\}$ containing truncated power functions of x_i. Observe that the penalized spline fitting criterion (7) when divided by the variance of the error process, σ_ϵ^2, can be written as

$$\frac{1}{\sigma_\epsilon^2} \left\| \boldsymbol{y} - \boldsymbol{L\gamma} - \boldsymbol{W\delta} - \boldsymbol{X\alpha} - \boldsymbol{Zb} \right\|^2 + \frac{1}{\lambda \sigma_\epsilon^2} \boldsymbol{b}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{b} \,, \tag{8}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_p)^T$ and $\boldsymbol{b} = (b_1, \ldots, b_K)^T$. Define $\sigma_b^2 = \lambda \sigma_\epsilon^2$, consider the vectors $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ as unknown fixed parameters and the vector $\boldsymbol{b}$ as a set of random parameters with $E(\boldsymbol{b}) = 0$ and $\mathrm{cov}(\boldsymbol{b}) = \sigma_b^2 \boldsymbol{\Sigma}$. If $(\boldsymbol{b}^T, \boldsymbol{\epsilon}^T)^T$ is a normal random vector and $\boldsymbol{b}$ and $\boldsymbol{\epsilon}$ are independent then model (6) has a linear mixed model (LMM) representation (Brumback et al., 1999):

$$\boldsymbol{y} = \boldsymbol{L\gamma} + \boldsymbol{W\delta} + \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon}, \;\; \mathrm{cov} \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\epsilon} \end{pmatrix} = \begin{bmatrix} \sigma_b^2 \boldsymbol{\Sigma} & 0 \\ 0 & \sigma_\epsilon^2 \boldsymbol{I}_n \end{bmatrix}. \tag{9}$$

For this model $E(\boldsymbol{y}) = \boldsymbol{L\gamma} + \boldsymbol{W\gamma} + \boldsymbol{X\alpha}$ and $\mathrm{cov}(\boldsymbol{y}) = \sigma_\epsilon^2 \boldsymbol{V}_\lambda$, where

$$\boldsymbol{V}_\lambda = \boldsymbol{I}_n + \lambda \boldsymbol{Z\Sigma Z}^T \,.$$

In the LMM described in (9) $\boldsymbol{L}$, $\boldsymbol{W}$ and $\boldsymbol{X}$ correspond to fixed effects and $\boldsymbol{Z}$ corresponds to random effects or parameters. Also, (8) is, up to an additive constant, twice the negative log-likelihood.

3 Likelihood Ratio Tests

In Section 2 we described how we can approximate testing for nonlinear regression by testing the null

$$H_0 : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\delta}) + \boldsymbol{\epsilon} \Leftrightarrow \boldsymbol{y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{10}$$

against the alternative

$$H_A : \boldsymbol{Y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\delta}) + s(\boldsymbol{x}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} \Leftrightarrow \boldsymbol{y} = \boldsymbol{L}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}\ , \tag{11}$$

where $\boldsymbol{b}$ and $\boldsymbol{\epsilon}$ are independent and have $N(0, \sigma_b^2 \boldsymbol{I}_K)$ and $N(0, \sigma_\epsilon^2 \boldsymbol{I}_n)$ distributions respectively. The null and alternative hypotheses can be written as

$$H_0 : \boldsymbol{\alpha} = 0 \text{ and } \sigma_b^2 = 0 \quad \text{vs.} \quad H_A : \boldsymbol{\alpha} \neq 0 \text{ or } \sigma_b^2 > 0\ , \tag{12}$$

which is testing for a null hypothesis that includes the assumption of zero random effects variance, $\sigma_b^2 = 0$ in a LMM with one random effects variance component. Crainiceanu and Ruppert (2004) defined likelihood ratio and restricted likelihood ratio tests for these type of hypotheses in the framework of LMMs with one variance component and derived their finite sample and asymptotic distributions.

Denote by $\boldsymbol{\mathcal{X}} = [\boldsymbol{L}|\boldsymbol{W}|\boldsymbol{X}]$ the design matrix of fixed effects under the alternative, by $\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^T\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^T$ and by $\mu_{s,n}$ and $\xi_{s,n}$ the K eigenvalues of the $K \times K$ matrices $\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}^T\boldsymbol{P}\boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}$ respectively. Then for testing hypotheses described in equation (12) the null finite sample distribution of the LRT is

$$\text{LRT}_n \stackrel{\mathcal{D}}{=} n\left(1 + \frac{\sum_{s=1}^{p+1} u_s^2}{\sum_{s=1}^{n-d} w_s^2}\right) + \sup_{\lambda \geq 0} f_n(\lambda), \tag{13}$$

where u_s for $s = 1, \ldots, K$, w_s for $s = 1, \ldots, n-d$, are independent $N(0,1)$, the notation $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution,

$$f_n(\lambda) = n \log\left\{1 + \frac{N_n(\lambda)}{D_n(\lambda)}\right\} - \sum_{s=1}^{K} \log(1 + \lambda\xi_{s,n}),$$

and

$$N_n(\lambda) = \sum_{s=1}^{K} \frac{\lambda\mu_{s,n}}{1+\lambda\mu_{s,n}} w_s^2 \,, \quad D_n(\lambda) = \sum_{s=1}^{K} \frac{w_s^2}{1+\lambda\mu_{s,n}} + \sum_{s=K+1}^{n-d} w_s^2 \,.$$

Here d is the number of columns of $\boldsymbol{\mathcal{X}}$ and $p+1$ is the number of columns of $\boldsymbol{X}$.

The RLRT_n is defined like the LRT_n using the restricted likelihood (Patterson and Thompson, 1971) instead of the likelihood function. Because this requires computing the likelihood of residuals after fitting the fixed effects, RLRT_n is appropriate only if the fixed effects are the same under the null and alternative, that is if we test for $\sigma_b^2 = 0$ only. This requirement is met when the linear part of the regression function already contains the monomials that appear in the spline function. For example, in many cases the regression function contains at least an intercept and the RLRT_n can be used with penalized piecewise constant splines. When the regression function does not contain an intercept, one can drop the intercept from the spline model as well; an example where this is appropriate is given in Section 7. Another option is to add an intercept and possibly other monomials to the regression function, at least for the purpose of testing goodness-of-fit; see the discussion of the three models in (15) in Section 4.

Crainiceanu and Ruppert (2004) showed that in this case

$$\text{RLRT}_n \stackrel{\mathcal{D}}{=} \sup_{\lambda \geq 0} \left[(n-p) \log \left\{ 1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right\} - \sum_{s=1}^{K} \log(1+\lambda\mu_{s,n}) \right] \,, \tag{14}$$

where notations are the same with the ones in equation (13). The equations (13) and (14) provide the spectral decompositions of the finite sample distributions of $(\text{R})\text{LRT}_n$ statistics in terms of the eigenvalues $\mu_{s,n}$ and $\xi_{s,n}$ and independent $N(0,1)$ random variables. While the form of these expressions may seem complex, both finite sample distributions are, in fact, very easy to simulate. Crainiceanu and Ruppert (2004) provided a fast simulation algorithm for these distributions and showed that the finite sample and asymptotic results differ from the results of Self and Liang (1987) and Stram and Lee (1994) derived for data that can be partitioned into a large number of i.i.d. subvectors.

4 Choosing the test: likelihood or restricted likelihood?

In the context of penalized splines, Crainiceanu, Ruppert and Vogelsang (2003) showed that the ML estimator of σ_b^2 has a strong downward bias, thus suggesting that LRT_n tends to favor the null model too often, which negatively affects its power. REML is also biased but the bias is less severe

than for ML. Crainiceanu, Ruppert, Claeskens and Wand (2004) showed in a simulation study that LRT_n has indeed less power than RLRT_n when testing for polynomial regression against a general alternative.

As an example of exactly how severe the effects of bias can be, consider the case of testing linear regression

$$H_0 : \; Y = \gamma_0 + \gamma_1 x + \epsilon$$

versus a general alternative modeled by a linear spline

$$H_A : \; Y = \gamma_0 + \gamma_1 x + \sum_{k=1}^{K} b_k (x - \kappa_k)_+ + \epsilon \; .$$

Following a procedure similar to the one described in Section 2 this may be reduced to testing

$$H_0 : \; \sigma_b^2 = 0 \quad \text{vs.} \quad H_A : \; \sigma_b^2 > 0 \; .$$

It can be shown that the null distribution of LRT_n in this case is practically point mass at zero. This is a consequence of the MLE property of correctly identifying the null model almost 100% of the time. Crainiceanu, Ruppert and Vogelsang (2003) calculated the probability mass at zero of LRT_n when σ_ϵ^2 remains constant and $\sigma_b^2 > 0$ increases (the true model is in the alternative) and showed that this probability decreases very slowly to zero. This property is not desirable because it means that, with high probability, the ML mistakenly identifies the null model as the true model. A related problem is that it is difficult to propose α-level tests when the null distribution of the test is very nearly point mass at zero. One solution would be to model the alternative as a quadratic instead of a linear spline

$$H_A : \; Y = \gamma_0 + \gamma_1 x + \alpha_2 x^2 + \sum_{k=1}^{K} b_k (x - \kappa_k)_+^2 + \epsilon$$

and transform the problem into testing

$$H_0 : \; \alpha_2 = 0, \sigma_b^2 = 0 \quad \text{vs.} \quad H_A : \; \alpha_2 \neq 0, \sigma_b^2 > 0 \; .$$

In this case it can be shown that the finite sample distribution of LRT_n is well approximated by a chi-squared distribution with one degree of freedom. While this solves the problem of designing an α-level test, the bias introduced by ML estimation will continue to affect the power of the test.

In contrast, REML is less biased than ML and the power of RLRT_n is better than the power of LRT_n. Because REML has better properties than ML and can only be defined for linear models,

we prefer to first linearize the model (1) and then embed it in a LMM's. However, by inspecting the hypotheses (12) we see that the null contains restrictions on the fixed effects parameters. More precisely, the coefficients of the monomials in the spline function that are not already contained in the linear part of the regression are constrained to zero under the null. There are two ways to use RLRT_n in this situation. One way is to ensure that there are no restrictions on the parameters $\boldsymbol{\alpha}$. This can be achieved if all parameters $\boldsymbol{\alpha}$ are already contained in the regression. For example if the nonlinear regression contains an intercept and a piecewise constant spline ($p = 0$) is used to model the alternative then no restrictions are imposed on the $\boldsymbol{\alpha}$ parameters. If the null regression contains a linear trend the same property can be achieved using a linear spline. When this strategy does not work testing (10) versus (11) can be done, testing sequentially model M_2 versus model M_3 and model M_1 versus M_2 where

$$\begin{array}{lrcl} M_1: & \boldsymbol{y} & = & \boldsymbol{L\gamma} + \boldsymbol{W\delta} + \epsilon \\ M_2: & \boldsymbol{y} & = & \boldsymbol{L\gamma} + \boldsymbol{W\delta} + \boldsymbol{X\alpha} + \epsilon \\ M_3: & \boldsymbol{y} & = & \boldsymbol{L\gamma} + \boldsymbol{W\delta} + \boldsymbol{X\alpha} + \boldsymbol{Zb} + \epsilon\,. \end{array} \tag{15}$$

where $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\alpha}$ are fixed effects parameters and $\boldsymbol{b} \sim N(0, \sigma_b^2)$ are random effects or parameters.

Both M_1 and M_2 are fixed effects linear models with M_1 nested within M_2. Therefore, standard LRT_n (or the equivalent F-test) can be used to test M_1 versus M_2. M_3 is a linear mixed model and testing M_2 versus M_3 is equivalent to testing whether the random effects variance σ_b^2 is zero. Unlike the case of testing directly M_1 versus M_3, testing M_2 versus M_3 does not involve restrictions on the fixed effects parameters and RLRT_n can be used.

5 Testing for an exponential regression function

Suppose that we are interested in testing the null of exponential regression

$$H_0:\ Y_i = \gamma_1 + \delta_1 \exp\left(\delta_2 x_i\right) + \epsilon_i\,. \tag{16}$$

This regression is a particular case of the regression equation (1) where the $\boldsymbol{L}$ matrix is simply an $n \times 1$ column of ones, $\boldsymbol{\gamma} = \gamma_1$, $\boldsymbol{\delta} = (\delta_1, \delta_2)^T$ and

$$f(x, \boldsymbol{\delta}) = \delta_1 \exp\left(\delta_2 x\right)\,.$$

As discussed in Section 2 we first linearize the nonlinear part of the regression. Note that

$$\frac{\partial}{\partial \delta_1} f(x, \boldsymbol{\delta}) = \exp(\delta_2 x)\,, \quad \frac{\partial}{\partial \delta_2} f(x, \boldsymbol{\delta}) = \delta_1 x \exp(\delta_2 x)\,.$$

For a given set of data, let $(\widehat{\gamma}_1, \widehat{\delta}_1, \widehat{\delta}_2)^T$ be the MLE of $(\gamma_1, \delta_1, \delta_2)^T$ for model (16). We approximate the null hypothesis (16) with its linearized version

$$H_0 : \; Y_i = \gamma_1 + m(x_i, \boldsymbol{\delta}) + \epsilon_i \,, \tag{17}$$

where

$$m(x, \boldsymbol{\delta}) = \widehat{\delta}_1 \exp\left(\widehat{\delta}_2 x\right) + \exp\left(\widehat{\delta}_2 x\right)(\delta_1 - \widehat{\delta}_1) + \widehat{\delta}_1 x \exp\left(\widehat{\delta}_2 x\right)(\delta_2 - \widehat{\delta}_2) \,,$$

and $\widehat{\boldsymbol{\delta}}$ is treated as constant, just like it would be treated in a parametric bootstrap. We can model the alternative using a piecewise constant spline

$$H_A : \; Y_i = \gamma_1 + m(x_i, \boldsymbol{\delta}) + \sum_{k=1}^{K} b_i I(x_i > \kappa_k) + \epsilon_i \,, \tag{18}$$

where $I(\cdot)$ denotes the indicator function, b_i are i.i.d. $N(0, \sigma_b^2)$ independent of ϵ_i which are i.i.d. $N(0, \sigma_\epsilon^2)$. We do not need to add an α parameter for the zero degree monomial of the spline because the null model already contains the intercept γ_1.

Let $\boldsymbol{L}$ be an $n \times 1$ column of ones corresponding to the intercept, $\boldsymbol{W}$ be the $n \times 2$ matrix with the i-th row

$$\boldsymbol{W}_i^T = \left\{ \exp\left(\widehat{\delta}_2 x_i\right) \;\; \widehat{\delta}_1 x_i \exp\left(\widehat{\delta}_2 x_i\right) \right\} ,$$

$$\boldsymbol{y}_i = \boldsymbol{Y}_i - f(x_i, \widehat{\boldsymbol{\delta}}) + \boldsymbol{W}_i^T \widehat{\boldsymbol{\delta}} \,,$$

and $\boldsymbol{Z}$ be the $n \times K$ matrix with the i-th row

$$\boldsymbol{Z_i} = \{ I(x_i > \kappa_1), \ldots, I(x_i > \kappa_K) \} \; .$$

The alternative model can be written as

$$\boldsymbol{y} = \boldsymbol{L}\gamma_1 + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{Z}\boldsymbol{b} + \epsilon,$$

and testing the null (17) versus the alternative (18) is reduced to testing

$$H_0 : \; \sigma_b^2 = 0 \quad \text{vs.} \quad H_A : \; \sigma_b^2 > 0 \,,$$

and RLRT_n of these hypotheses can be used. Define $\boldsymbol{\mathcal{X}} = [\boldsymbol{L}|\boldsymbol{W}]$ and let $\mu_{s,n}$ be the K eigenvalues of the matrix $\boldsymbol{Z}^T \boldsymbol{P} \boldsymbol{Z}$. The finite sample distribution of RLRT_n can be obtained using the simulation of its spectral decomposition (14).

As examples we considered testing the null hypothesis of exponential regression (16) with $n = 100$, $\gamma_1 = 1$, $\delta_1 = 1$, $\delta_2 = -1$, $\sigma_\epsilon = 0.05$. We used four sets of x's in $[0, 1]$ one for each of the following distributions: equally spaced, uniform, BETA$(5, 1)$ and BETA$(1, 5)$. For each set of x's we simulated once a set of $\boldsymbol{Y}$'s from (16) and obtained the MLE of the parameters in the nonlinear regression (16). We then obtained $100,000$ simulations from the distributions of LRT_n and RLRT_n statistics using (13) and (14).

Not surprisingly, the null distribution of the LRT_n has more than 0.99 mass at zero for each of the four cases, which is consistent with our discussion in Section 4. The null distribution of RLRT_n changed very little over the four cases considered and had $p_0 = 0.6$ probability mass at zero and the 0.9, 0.95 and 0.99 quantiles of the distribution were approximately $q_{0.9} = 1.23$, $q_{0.95} = 2.22$ and $q_{0.99} = 4.82$ respectively. In comparison, the quantiles for the $50 : 50$ χ^2_0 and χ^2_1 are $q_{0.9} = 1.64$, $q_{0.95} = 2.71$ and $q_{0.99} = 5.41$.

While the simulation results are enough to compute p-values for testing the exponential regression it may be helpful to have a theoretical distribution that provides tail probabilities. Consider the two-parameter family of distributions of the random variables

$$a \times U \times D$$

where $a > 0$ is an unknown parameter, U has a Bernoulli distribution with $P(U = 0) = p$ and D is a random variable with a χ^2_1 distribution. To find a distribution that approximates well the finite sample distributions of RLRT_n we match the quantiles corresponding to probabilities 0.6 and 0.9 and obtain $p = 0.6$ and $a = 0.93$.

6 Level and power of the tests

Our proposed tests involve two approximations that can influence their size under the null: replacing the nonlinear by a linear model and assuming that the MLE $\widehat{\beta}$ is fixed. We investigate the effect of these approximations in the case of testing for an exponential regression, as described in Section 5. We compare our test with the $T_n^{\text{AN},1}$ test of Fan and Huang (2001) both in terms of size under the null and power. Note that if the null hypothesis is linear in the parameters, then our tests are exact but $T_n^{\text{AN},1}$ is not. Here the subscript n denotes the sample size.

We now investigate the size of these tests when the null is the exponential regression described in equation (16). We use the RLRT described in Section 5 with equally spaced covariates in $[0, 1]$

and the alternative modeled by a piecewise constant spline with $K = 15$ knots. We used $10,000$ simulations from the model (16) with $\gamma_1 = 1$, $\delta_1 = 1$, $\delta_2 = -1$, $\sigma_\epsilon = 0.05$ and calculated RLRT_n and $T_n^{\text{AN},1}$. For RLRT_n and $T_n^{\text{AN},1}$ Table 1 reports the frequencies of exceeding the $1-\alpha$ quantile for different values of α. For $T_n^{\text{AN},1}$ we used the upper quantile Table 1 from Fan and Huang (2001). The worst Monte Carlo standard error is approximately $\sqrt{(0.1)(0.9)/10000} \approx 0.3\%$. For these examples, the approximate level of the RLRT_n tends to be much closer to the true level than for $T_n^{\text{AN},1}$, especially for smaller values of n.

Table 1: Level of tests in $10,000$ simulations

α	$T_{50}^{\text{AN},1}$	$T_{50}^{\text{AN},1}$	LRT_{50}	RLRT_{100}
0.01	0.007	0.020	0.010	0.011
0.05	0.029	0.044	0.049	0.052
0.10	0.054	0.073	0.100	0.110

In a simulation study Fan and Huang (2001) showed that the $T_n^{\text{AN},1}$ test performs well against a variety of alternatives and we compare the power of this test with the power of the RLRT_n where the alternative is

$$Y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i \tag{19}$$

and $d = 0$ corresponds to the null hypothesis. Figure 1–(a), (b) shows the power of $T_n^{\text{AN},1}$ adjusted and unadjusted for size compared with the power of the RLRT_n. The unadjusted test uses the critical values given by Fan and Huang (2001). The adjusted test uses critical values found by simulation and therefore is exact. We used level $\alpha = 0.05$ tests under the null. For RLRT_n the power curve adjusted for size is indistinguishable from the unadjusted curve. In both cases the RLRT_n test is more powerful than $T_n^{\text{AN},1}$ over the relevant range of values of d.

In addition to this simulation study, Crainiceanu et al., (2004) provide a comparative simulation power study for tests of linearity, including F, F-type, and Von-Neuman type tests in addition to LRT_n and RLRT_n, the main conclusion being that RLRT_n is easy to use and has good power properties, while LRT_n is less powerful due to large ML estimation bias in the variance component.

7 Example: Price of zero-coupon bonds

An important problem in finance is to estimate spot rates from bond prices. We will use zero-coupon bonds which pay no principal or interest until maturity and at maturity pay a fixed amount called the par value. Let $P(t)$ be the price of a zero-coupon bond maturing at time t expressed as a percentage of the par value. Let $r(t)$ be the spot rate, or yield, and let

$$D(t) = e^{-tr(t)}$$

be the discount function. The yield $r(t)$ is the average interest earned on a zero-coupon bond maturing at time t. A commonly used model, e.g., by James and Webber (2000), is

$$\frac{P(t_i)}{100} = D(t_i) + \epsilon_i \,. \tag{20}$$

The noise ϵ_i is due to a number of factors. For example, prices are from the last sale of a bond and will be somewhat stale, with the prices of less liquid bonds typically being the most stale. Also, some bonds sell at a premium because of special liquidity or other advantages but these effects are not considered to be part of the yield function $r(t)$ but rather are defined as noise (Tuckman, 2002).

A series of parametric models have been proposed to model the spot rate, and implicitly the discount function. To illustrate our methodology we will use the STRIPS prices at the end of June 1994. The sample size of the data is $n = 116$. "STRIPS" is an acronym meaning "Separate Trading of Registered Interest and Principal of Securities," and a STRIPS is a synthetic zero-coupon bond constructed from the coupon or principal payments of Treasury bonds (Tuckman, 2002). The data come from the fixed income data base (Warga, 1995). As an illustration, we will test whether the spot rate curve is given by the Nelson and Siegel (1985) model which is

$$r(t, \boldsymbol{\delta}) = \delta_0 + (\delta_1 + \delta_2 t)\, e^{-\delta_3 t} \,. \tag{21}$$

(As James and Webber (2000) mention, the Nelson and Siegel curve is suitable as a model for either the spot rate or the forward rate which is the $d/dt\{t\, r(t)\}$. Here we are using the Nelson/Siegel curve to model the spot rate.) To obtain starting values for the parameters $\boldsymbol{\delta}$ for model (20) we use the following fitting procedure proposed by Tuckman (2002)

$$-\frac{\log\{P(t_i)/100\}}{t_i} = \delta_0 + \delta_1 e^{-\delta_3 t} + \delta_2 t e^{-\delta_3 t} + e_i \,. \tag{22}$$

Division by t_i creates severe heteroscedasticity, so estimates from (22) will be inefficient, but they are satisfactory as starting values. For the parameter δ_3 we considered a grid of 100 points equally spaced on the log scale between $[-3.5, 1.5]$. For each fixed value of the δ_3 on this grid, model (22) is linear so the profile likelihood of δ_3 can be computed easily. The initial parameter estimates are obtained by maximizing the profile likelihood function over this grid of values for δ_3 and then obtaining the linear least-squares estimates of δ_1 and δ_2 corresponding to the maximizing value of δ_3. Figure 2 displays $-\log\{P(t)/100\}/t$ recorded in June 1994 versus time to maturity t for 116 bonds, as well as the estimated spot rates using model (21) using these estimates.

The estimated parameters using the grid search are used as initial values in the nonlinear maximization algorithm of the likelihood of model (20) using the parametric form (21) for the spot rate function. The residuals for this model are plotted in Figure 3. If $\widehat{\boldsymbol{\delta}}$ denotes the MLE of $\boldsymbol{\delta}$ for model (20) then the first order Taylor approximation is

$$\frac{P(t_i)}{100} = \exp\left\{-t_i r(t_i, \widehat{\boldsymbol{\delta}})\right\} - \boldsymbol{W}_i^T \widehat{\boldsymbol{\delta}} + \boldsymbol{W}_i^T \boldsymbol{\delta} + \epsilon_i \,, \tag{23}$$

where

$$\boldsymbol{W}_i = \left\{ \frac{\partial D}{\partial \delta_0}(t_i, \widehat{\boldsymbol{\delta}}), \frac{\partial D}{\partial \delta_1}(t_i, \widehat{\boldsymbol{\delta}}), \frac{\partial D}{\partial \delta_2}(t_i, \widehat{\boldsymbol{\delta}}), \frac{\partial D}{\partial \delta_3}(t_i, \widehat{\boldsymbol{\delta}}) \right\} .$$

As we discussed in Section 2 we treat $\widehat{\boldsymbol{\delta}}$ as fixed and fit model (23) as a linear model in $\boldsymbol{\delta}$. The residuals from this fit are also presented in Figure 3.

To model the alternative for model (23) we used a penalized linear spline with $K = 10$ knots, where the kth knot was the sample quantile of observed times to maturity corresponding to probability $k/(K+1)$. Thus, the alternative model was

$$\frac{P(t_i)}{100} = \exp\left\{-t_i r(t_i, \widehat{\boldsymbol{\delta}})\right\} - \boldsymbol{W}_i^T \widehat{\boldsymbol{\delta}} + \boldsymbol{W}_i^T \boldsymbol{\delta} + s(t_i, \boldsymbol{\theta}) + \epsilon_i \,, \tag{24}$$

where $s(t_i, \boldsymbol{\theta})$ is a spline function. Because the discount function $D(\cdot)$ has the property that $D(0) = 1$ we used a spline function without intercept

$$s(t, \boldsymbol{\theta}) = \alpha_1 t + \sum_{k=1}^{K} b_k (t - \kappa_k)_+ \,.$$

Using the same notations as elsewhere in the paper, note that the matrix $\boldsymbol{L}$ is the empty matrix and we have the following three nested models

$$\begin{array}{lrcl} M_1: & \boldsymbol{y} & = & \boldsymbol{W}\boldsymbol{\delta} + \epsilon \\ M_2: & \boldsymbol{y} & = & \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{X}\boldsymbol{\alpha} + \epsilon \\ M_3: & \boldsymbol{y} & = & \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{b} + \epsilon \,, \end{array} \tag{25}$$

where $\boldsymbol{y}$ is defined like in Section 5, $\boldsymbol{X}$ is an $n \times 1$ vector with the i-th entry equal to t_i, and $\boldsymbol{Z}$ is an $n \times K$ matrix corresponding to random effects with the i-th row

$$\boldsymbol{Z}_i = \{(t_i - \kappa_1)_+, \ldots, (t_i - \kappa_K)_+\} \ .$$

Models M_1 and M_2 contain only fixed effects whereas M_3 contains random effects $\boldsymbol{b}$. Figure 3 shows the residuals for model M_3 which corresponds to the spline regression using REML estimation of the smoothing parameter. Also, Figure 4 zooms in on the residuals for model M_3, indicating that the spline model can safely be considered as the full model.

Our goal is to test model M_1 versus model M_3. Table 2 indicates strong evidence against model M_1 because for this data set $\text{LRT}_n = 356.60$ with a null finite sample distribution that can be well approximated by a χ^2_1 distribution. It may be surprising that the null distribution of LRT_n when testing M_1 versus M_3, that is testing $\alpha_1 = 0$ and $\sigma^2_b = 0$, is well approximated by χ^2_1 rather than a 50:50 mixture of χ^2_0 and χ^2_1. It may be even more surprising that the distribution of LRT_n when testing M_2 versus M_3, that is testing $\sigma^2_b = 0$, is practically the Dirac measure at zero. These results are due to the strong downward bias of the MLE of σ^2_b, and can influence the power of the test. In this particular example LRT_n rejects the null because there is so much evidence against the null. However, LRT_n should be used carefully in cases when the null is not rejected. RLRT_n can be used to test only model M_2 versus M_3 because both models have the same fixed effects and in this case RLRT_n rejects M_2 in favor of M_3. We used one million simulations to estimate relevant quantiles of the null finite sample distribution of RLRT_n as described in Section 3. Using the same quantile matching technique described in Section 5 we found that the finite sample distribution of RLRT_n can be well approximated by the distribution of $0.90\,UD$, where U has a Bernoulli distribution with $P(U = 0) = 0.63$ and D has a χ^2_1 distribution, independent of U. In Table 2 this distribution is denoted by $0.63\chi^2_0 : 0.37(0.90\chi^2_1)$.

8 Multidimensional Covariates

The case where $\boldsymbol{x}$ is one-dimensional (and then denoted here as x) is quite common in nonlinear regression, but there are many applications where $\boldsymbol{x}$ is multivariate in which case the linearization (2) is unchanged, but the spline model (5) needs to changed. One possibility would be to use a fully multivariate spline, either a tensor product spline or radial basis functions (Ruppert, Wand, and Carroll, 2003). However, this degree of complication is probably not needed for goodness-of-fit

Table 2: LRT_n and RLRT_n and null finite sample distributions for testing the null of model (20) with the parametric form for the spot rate described by (21). The alternative is modeled by penalized splines. Zero-Coupon Bond Prices: June 1994.

	LRT_n		RLRT_n	
Models Tested	Value	Appoximate Null Distribution	Value	Approximate Null Distribution
M1/M2	218.69	χ^2_1		
M2/M3	137.91	χ^2_0	159.53	$0.63\chi^2_0 : 0.37(0.90\chi^2_1)$
M1/M3	356.60	χ^2_1		

testing and we recommend instead an additive spline model consisting of a sum of models of form (5), one for each component of $\boldsymbol{x}$. For optimal fitting of such a model, one might use a separate variance component for each component of $\boldsymbol{x}$ which would take into account differing amounts of roughness among the components. However, for goodness-of-fit testing a common variance component should be adequate, though more research on this question would be useful. With a common variance component, (8) would still be used with some obvious changes: $\boldsymbol{X}$ and $\boldsymbol{Z}$ would be expanded to contain, respectively, monomials in all of the components of $\boldsymbol{x}$ and truncated power functions in all of the components of $\boldsymbol{x}$. There would be corresponding changes in $\boldsymbol{\alpha}$ and $\boldsymbol{b}$.

9 Summary

We described (R)LRT for testing the null hypothesis of nonlinear regression versus a general alternative modeled by penalized splines. The proposed strategy is to use a first order Taylor approximation of the nonlinear conditional mean around the MLE as the null model and penalized splines to describe a flexible alternative. We discuss why in this framework LRTs have poor power properties and we propose RLRTs instead. The linearization procedure allows the use of RLRT, because REML is only defined for linear regression models.

This testing strategy can be viewed as testing for a null hypothesis that includes zero random effects variance in a LMM with one random effects variance component. The spectral decomposition of the test statistics is used as the basis of a fast simulation algorithm that provides the exact finite sample distributions.

Acknowledgment

We thank Professor Yan Yu for supplying the STRIPS data.

REFERENCES

Brown, R.L., Durbin, J., Evans, J.M., (1975). Techniques for testing the constancy of regression relationship over time. *Journal of the Royal Statistical Society - Series B*, 37, 149–163.

Brumback, B., Rice, J.A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association*, 93, 961–994.

Brumback, B., Ruppert, D., Wand, M.P., (1999). Comment on "Variable selection and function estimation in additive nonparametric regression using data-based prior" by Shively, Kohn, and Wood, *Journal of the American Statistical Association*, 94, 794–797.

Crainiceanu, C.M., and Ruppert, D., (2004). Likelihood ratio tests in linear mixed models with one variance component, *Journal of Royal Statistical Society - Series B*, 66, 165–185.

Crainiceanu, C.M., Ruppert, D., Claeskens, G., and Wand, M.P., (2004). Exact Likelihood Ratio Tests for Penalized Splines. *under revision*, available at www.orie.cornell.edu/~davidr/papers.

Crainiceanu, C.M., Ruppert, D., and Vogelsang, T.J. (2004). Some properties of Likelihood Ratio Tests in Linear Mixed Models. *submitted*, available at www.orie.cornell.edu/~davidr/papers.

Diebolt, J., Zuber, J., (1999). Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statistics and Pobability Letters*, 42, 53–60.

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

Gardner, L.A., (1969). On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics*, 10, 116–126.

Hart, J.D., (1997). Nonparametric smoothing and lack-of-fit tests. New York: Springer.

He, X., Zhou, L. X., (2003). A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, 98, 1013–1022.

James, J, Webber, N. (2000), *Interest Rate Modeling*, John Wiley & Sons, Chichester.

MacNeill, I.B., (1978), Properties of sequences of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times. *Annals of Statistics*, 6, 422–433.

Nelson, C.R., Siegel, A.F. (1985), Parsimonious modelling of yield curves, *Journal of Business*, 60, 473–489.

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

Perron, P., (1991), A test for changes in a polynomial trend function for a dynamic time series. *Mimeo*. Princeton Department of Economics.

Ruppert, D. (2002), Selecting the number of knots for penalized splines, *Journal of Computational and Graphical Statistics*, 11, 735–757.

Ruppert, D., Wand, M. P., Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge: Cambridge University Press.

Self, S.G., Liang, K.Y., (1987), Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions, *Journal of the American Statistical Association*, 82, 605–610.

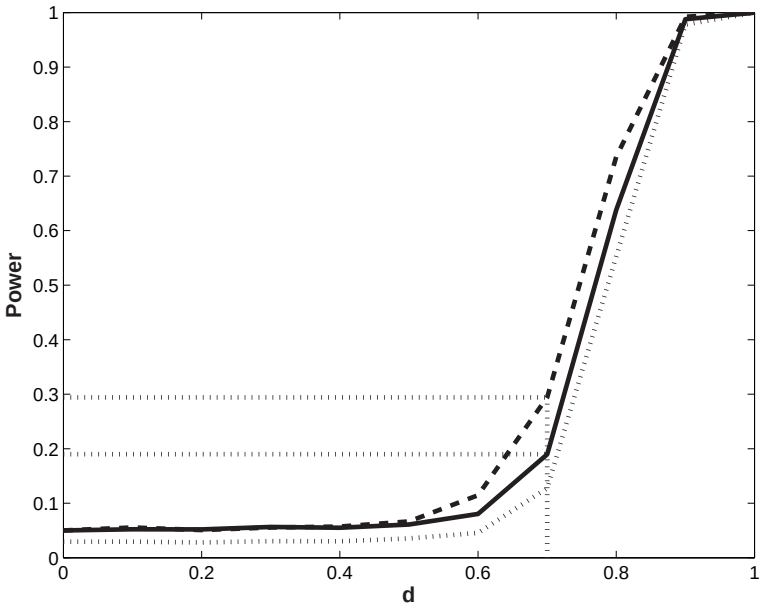
Stram, D.O., and Lee, J.W., (1994), Variance Components Testing in the Longitudinal Mixed Effects Model, *Biometrics*, 50, 1171–1177.

Stute, W. (1997), Nonparametric model checks for regression. *Annals of Statistics*, 25, 684–692.

Tuckman, B. (2002), *Fixed Income Securities, 2nd Ed.*, John Wiley & Sons, Hoboken, NJ.

Warga, A. (1995), Fixed Income Securities Database, Fixed Income Research Program, University of Houston. (Request information at www.uh.edu/˜awarga)

(a)



(b)

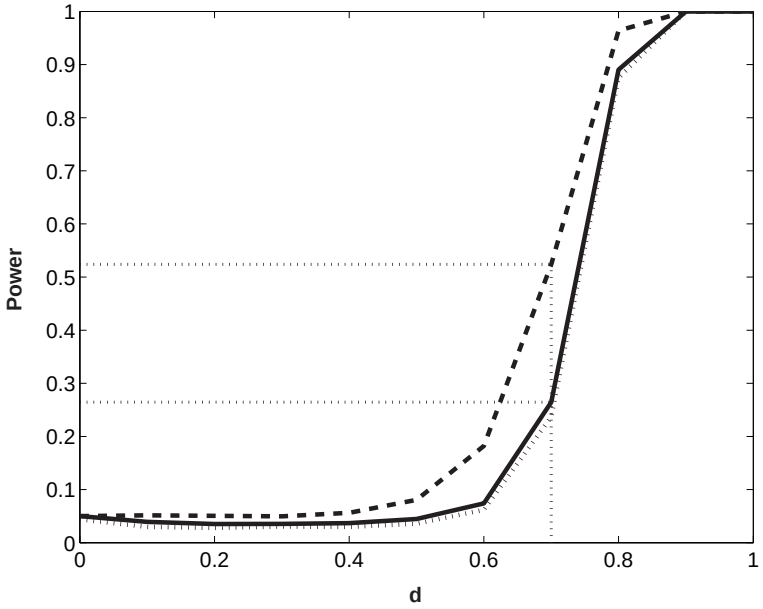


Figure 1: Power of RLRT_n "– –", size adjusted $T_n^{\mathrm{AN},1}$ "–", unadjusted $T_n^{\mathrm{AN},1}$ "..." for testing null hypothesis (16) against the alternative (19).

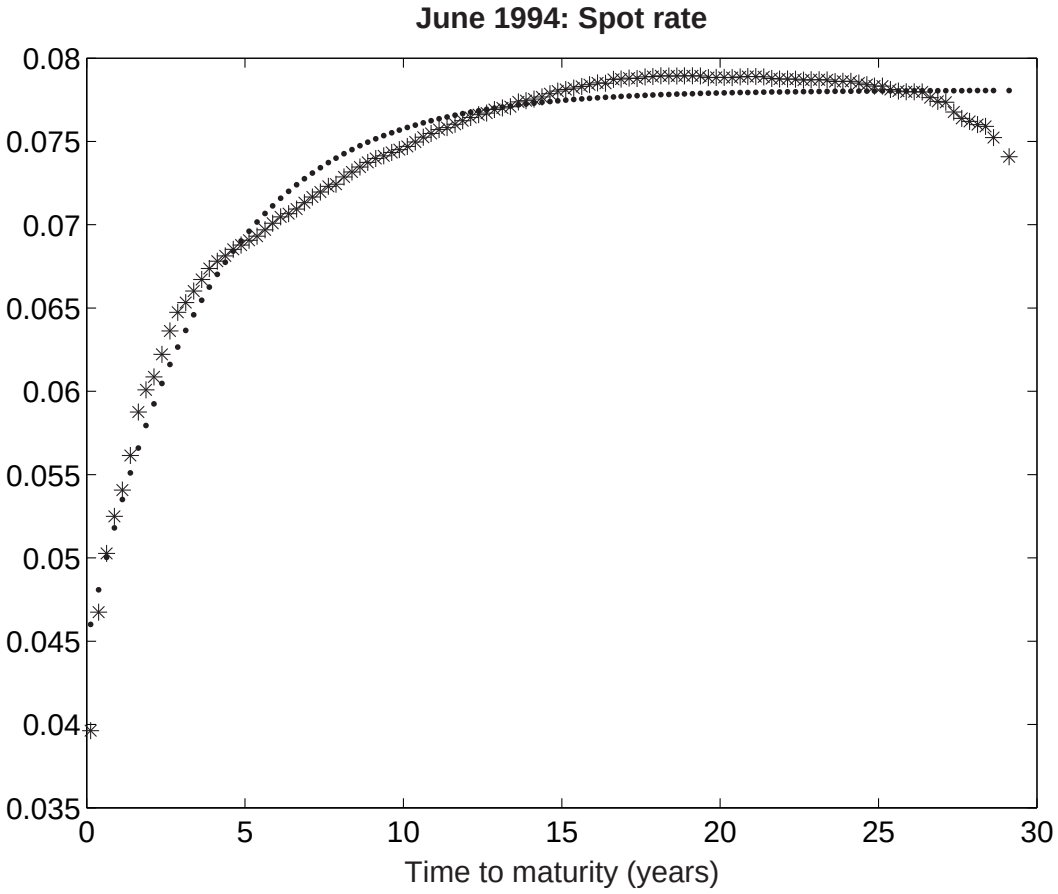


Figure 2: Plot of $-\log\{P(t_i)/100\}/t_i$ versus time to maturity – "*". Plot of estimated spot rates using model (21) – "•".

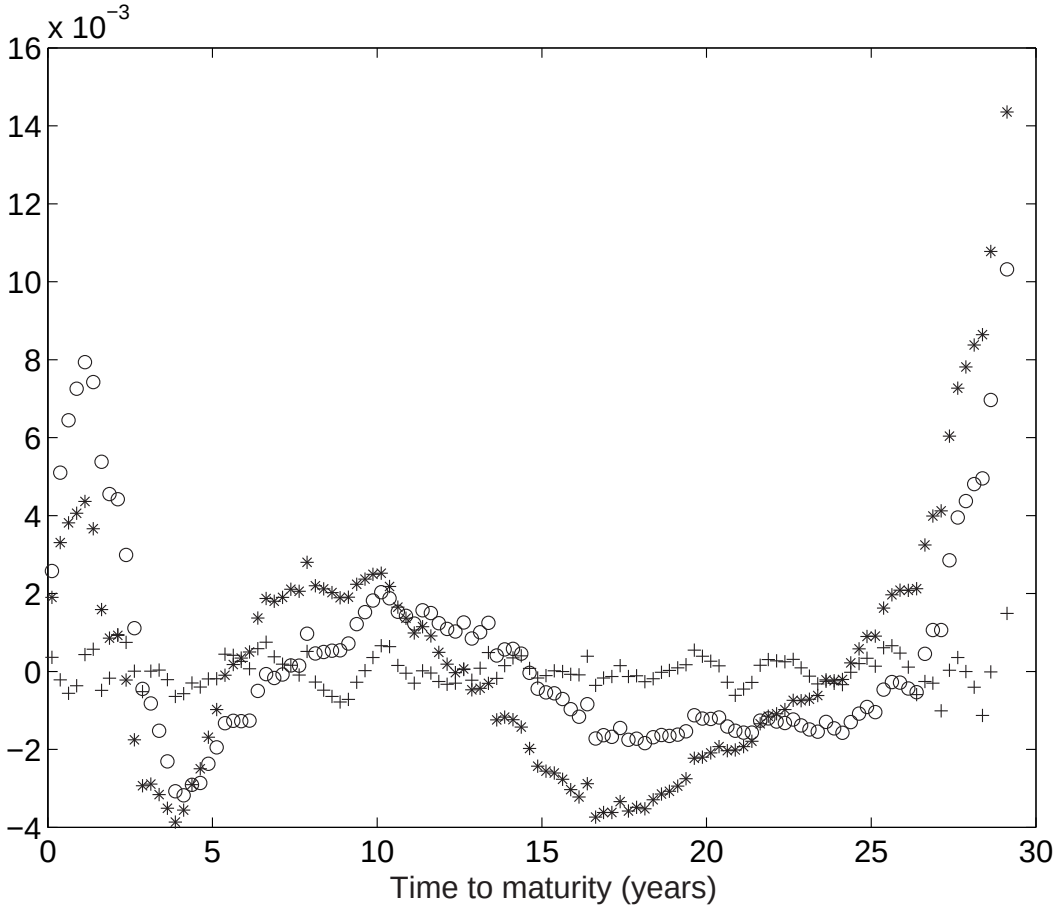


Figure 3: Residuals for model (20) using the parametric form (21) for spot rates – "*". Residuals for first order Taylor approximation of the model (20) – "o". Residuals for the alternative modeled by a linear penalized spline with $K = 10$ knots using REML estimation of the smoothing parameter – "+"

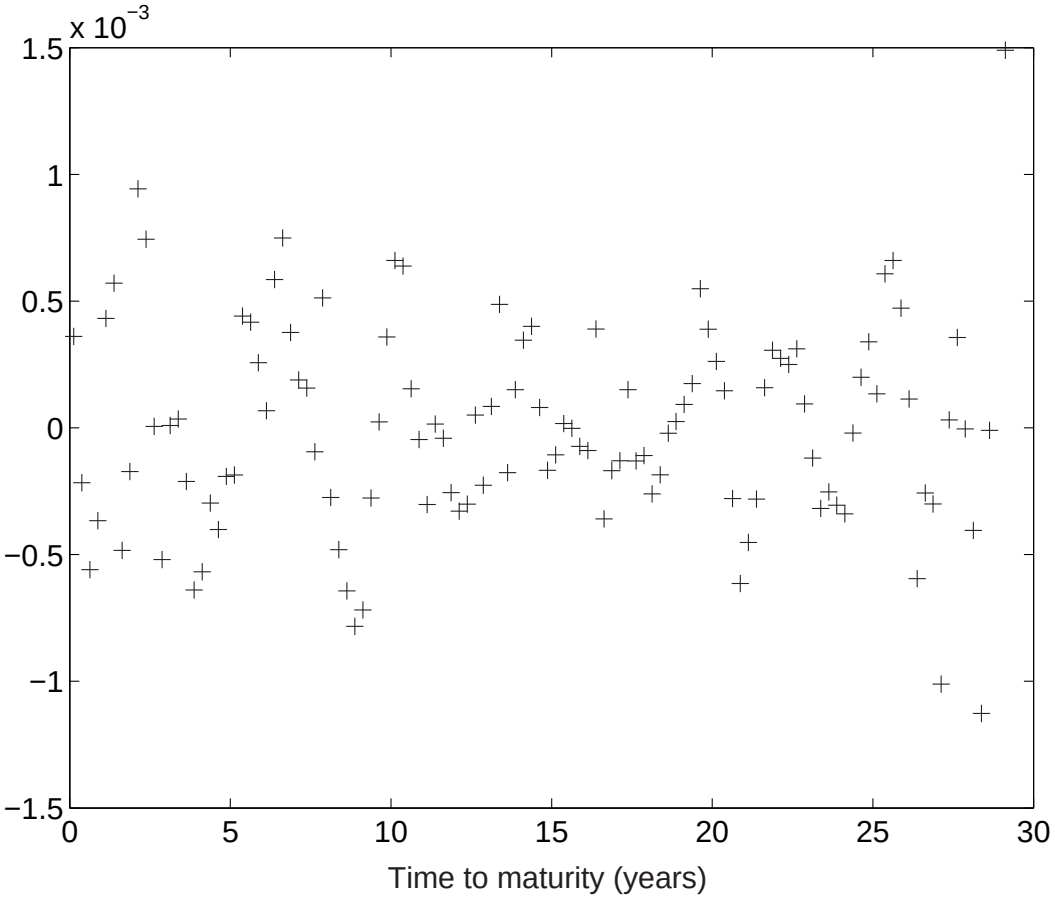


Figure 4: Zoom-in on the residuals for Model M_3 using REML estimation of the smoothing parameter.