Lack-of-fit tests in semiparametric mixed models

Gerda Claeskens, Huijuan Ding and Maarten Jansen

Lack-of-Fit Tests in Semiparametric Mixed Models

Gerda Claeskens, Huijuan Ding and Maarten Jansen

K.U.Leuven

ORSTAT and University Center for Statistics

Naamsestraat 69, 3000 Leuven

{Gerda.Claeskens, Huijuan.Ding}@econ.kuleuven.be, Maarten.Jansen@wis.kuleuven.be

January 2007

Abstract

In this paper we obtain the asymptotic distribution of restricted likelihood ratio tests in mixed linear models with a fixed and finite number of random effects. We explain why for such models the often quoted 50:50 mixture of a chi-squared random variable with one degree of freedom and a pointmass at zero does *not* hold. Our motivation is a study of the use of wavelets for lack-of-fit testing within a mixed model framework. Even though wavelets have received a lot of attention in the last say 15 years for the estimation of piecewise smooth functions, much less is known about their ability to check the adequacy of a parametric model when fitting the observed data. In particular we study the testing power of wavelets for testing a hypothesized parametric model within a mixed model framework. Experimental results show that in several situations the wavelet-based test significantly outperforms the competitor based on penalized regression splines. The obtained results are also applicable for testing in mixed models in general, and shed some new insight into previous results.

Keywords: Lack-of-fit test, likelihood ratio test, mixed models, one-sided test, penalization, restricted maximum likelihood, variance components, wavelets.

1 Introduction

The main aim of this paper is to construct test statistics based on wavelets for testing a parametric null model against a nonparametric alternative model. The proposed tests possess a similarity to the adaptive tests which use penalized spline regression models in a mixed model framework. Our simulations show that the wavelet based tests outperform the spline based tests in several situations, and are comparable in other settings where the spline tests are known to perform well.

A second result in this paper is the extension to testing in mixed models in general (not necessarily using wavelets). The asymptotic distribution that we obtain explains the often reported simulation results which seemingly deviate significantly from the assumed theoretical results. The asymptotic distribution for tests in mixed models is often described to be a mixture of a chi-squared random variable with one degree of freedom an a pointmass at zero. Pinheiro and Bates (2000), for example, observed in simulations that the mixture coefficients were approximating the values 0.65 and 0.35, rather than 0.5 and 0.5. We obtain an asymptotic distribution where the mixture proportions depend on the design of the fixed and random effects. In the case of a growing number of random effects, the proportions converge to 0.5, but not in general for testing with a fixed number of random effects.

The use of wavelets for lack-of-fit testing is advantageous for a multitude of reasons. Due to their multiscale nature, one possible application is scale dependent testing for random effects. Second, the test proposed in this paper detects piecewise smooth alternatives, i.e., functions with jumps, sharp peaks or high frequency alternations. This is in contrast to existing tests, which restrict themselves to smooth alternatives. Tests assuming smoothness of the alternative comprise the orthogonal series based tests with special emphasis on the order selection tests of Eubank and Hart (1992), Hart (1997), Aerts et al. (1999, 2000) and the Neyman smooth type tests, mainly used for goodness of fit testing for density functions of Ledwina (1994), see also Fan (1996). In this latter paper a goodness-of-fit test based on wavelet thresholding is proposed that builds further on the Neyman test. Spokoiny (1996) studies the optimal rates for adaptive tests based on wavelets, using L_2 distances.

Our proposed test is more in line with the restricted likelihood ratio tests used in combination with penalized splines, see Crainiceanu and Ruppert (2004), Claeskens (2004), and Crainiceanu et al. (2005). In order to test a parametric null hypothesis, a semiparametric alternative model is constructed using the hypothesized model as the parametric part and wavelet basis functions for the nonparametric part. This is embedded in a mixed model framework where the parametric part is taken as the set of fixed effects and where the wavelet coefficients are random effects. A thresholding procedure is used to select a relevant subset of wavelet basis functions to be used for the test. The particular advantage of constructing a mixed effects model is that

the hypothesis test reduces to testing whether the single variance component of the random effects, say σ_w^2, is equal to 0. The test statistic is the restricted likelihood ratio statistic comparing the semiparametric alternative model to the null model.

The use of wavelets and thresholding in the testing procedure is new. Because of the thresholding step, which acts as a pre-test or variable selection procedure, the distribution of the test statistic is different from that in case spline basis functions are used. Thresholding is a non-linear procedure and requires a different asymptotic theory. In addition to asymptotic distribution results, we use a bootstrap resampling procedure to obtain P-values. Interestingly, the obtained distribution has a much wider application area than testing lack of fit using wavelets.

2 A mixed effects wavelet model

In this first step we start with testing for polynomial models in one variable. Extensions to other settings are provided in Section 5. We wish to test using data $(y_i, x_i), i = 1, \ldots, n$ whether the mean of Y given the covariate x is a polynomial of degree q,

$$H_0 : E(Y) = \beta_0 + \beta_1 x + \ldots + \beta_q x^q, \tag{1}$$

where the coefficients $\beta_0, \ldots, \beta_q$ are left unspecified. A nonparametric lack-of-fit test contrasts this null model with a semiparametric alternative model of the form

$$Y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_q x_i^q + g(x_i) + \varepsilon_i,$$

where the function g is unspecified. The constructed test statistic shall use wavelets to estimate g. We first perform a wavelet transformation of $g(\cdot)$, employing the residuals of the null model fit.

The wavelet decomposition of $g(x)$ equals

$$g(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k} \psi_{j,k}(x)$$

where $\hat{c}_{j_0,k}$ and $\hat{d}_{j,k}$ are, respectively, empirical scaling function coefficients and wavelet coefficients obtained by a discrete wavelet transform (DWT) for a given wavelet basis. The scaling functions $\phi_{j_0,k}(x)$ with their coefficients constitute a coarse scale, smooth approximation of the observations. Under the null hypothesis this part is not significant (i.e., close to zero), as we are decomposing a residual $g(x)$, and the actual smooth approximation of the observations has been captured by the parametric part. Possible sharp transitions, jumps, peaks, are typically described by large wavelet coefficients. This observation explains the usage of threshold procedures in

wavelet based smoothing. Thresholding is used here to set up the design matrix $\mathbf{Z}$ (in a data-adaptive fashion). This matrix is a submatrix of the full reconstruction matrix in the perfect reconstruction:

$$[g(x_1), \ldots, g(x_n)]^T =$$

$$
\begin{bmatrix}
\phi_{j_0,0}(x_1) & \ldots & \phi_{j_0,2^{j_0}-1}(x_1) & \psi_{j_0,0}(x_1) & \ldots & \psi_{J-1,2^{J-1}}(x_1) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_{j_0,0}(x_n) & \ldots & \phi_{j_0,2^{j_0}-1}(x_n) & \psi_{j_0,0}(x_n) & \ldots & \psi_{J-1,2^{J-1}}(x_n)
\end{bmatrix}
\begin{bmatrix}
\hat{c}_{j_0,0} \\
\vdots \\
\hat{c}_{j_0,2^{j_0}-1} \\
\hat{d}_{j_0,0} \\
\vdots \\
\hat{d}_{J,2^{J}-1}
\end{bmatrix}.
$$

As already mentioned, the scaling basis functions $\phi_{j_0,k}(x_i)$ are not part of the design. Thresholding is applied to the empirical wavelet coefficients $\hat{d}_{j,k}$ and $\mathbf{Z}$ contains in its columns those wavelet basis functions corresponding to the empirical wavelet coefficients above the threshold. The threshold used here is the *universal threshold*, further detailed in Section 4. In the event that no wavelet coefficient survives the threshold, a fixed, single wavelet basis function is selected.

The discussion in this paper is limited to orthogonal wavelet bases. Orthogonality implies that wavelet coefficients of observational data with additive, normal, uncorrelated and homoscedastic noise are themselves homoscedastic, normal and uncorrelated (hence independent). We also impose that the wavelet basis is orthogonal to the polynomial basis $\{1, x, ..., x^q\}$ of the parametric part of our model. This ensures that the corresponding coefficients are independent. The orthogonality between wavelets and polynomials is well known in the literature as the *vanishing moments* condition. Vanishing moments are important in general applications for another reason: if the smooth intervals of a function can be well approximated by a polynomial, then the inner products of that function with wavelet basis functions are small, whenever the support of the wavelet function does not contain a singular point (jump, peak). As a consequence, most coefficients are close to zero, leading to a sparse representation that can be easily compressed or denoised using thresholds for wavelet coefficients. Daubechies (1988) proposed several different orthogonal wavelet families that have compact support with various degrees of smoothness and numbers of vanishing moments, which can be used for our problem. They are called Daubechies wavelets, Coiflets and Symlets. We choose a wavelet in the Daubechies families with at least $q + 1$ vanishing moments.

This brings us to the following semi-parametric model

$$Y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_q x_i^q + \sum_{k=1}^{K_n} u_k \psi_k(x_i) + \varepsilon_i, \tag{2}$$

or, in matrix notation, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$. The design matrices of fixed and random effects are given by

$$
\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^q \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \ldots & x_n^q \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \psi_1(x_1) & \ldots & \psi_{K_n}(x_1) \\ \vdots & \ddots & \vdots \\ \psi_1(x_n) & \ldots & \psi_{K_n}(x_n) \end{bmatrix}
$$

where $\psi_k, k = 1, ..., K_n$, are wavelet basis functions, and ε_i are independent identically distributed $N(0, \sigma_\varepsilon^2)$.

As explained above, the wavelet basis functions included in the matrix $\boldsymbol{Z}$ are those for which the corresponding wavelet coefficients are larger than the threshold. It is important to realize that both the choice of basis functions, as well as the number K_n is random and chosen in a data-driven way. At this point the wavelet model differs substantially from the spline model, where the spline basis functions as well as their total number are fixed beforehand, not data-driven.

In the mixed regression wavelet model, we explicitly assume that the coefficients β remain fixed, and that the wavelet coefficients u_k are independent and identically distributed random variables having a normal distribution $N(0, \sigma_w^2)$. The introduction of random effects on the u_ks results in equivalence of best linear unbiased predictors and estimators obtained via generalized least squares (GLS). Angelini and Leblanc (2003) show that the wavelet estimator in a mixed effect model coincides with the solution to a certain regularization problem over a reproducing kernel Hilbert space.

3   Description of the test statistic

With respect to the polynomial-wavelet full mixed model (2), the reduced model under the null hypothesis is the parametric polynomial

$$
H_0 : Y = \beta_0 + \beta_1 x + \ldots + \beta_q x^q + \varepsilon.
$$

Testing H_0 against the two-sided alternative that the conditional mean response has any different structure in the mixed model representation is equivalent with testing the now one-sided hypothesis

$$
H_0 : \sigma_w^2 = 0 \quad \text{versus} \quad H_a : \sigma_w^2 > 0.
$$

The mixed model formulation dramatically reduces the dimensionality of the testing problem. Otherwise a nonparametric test in the same setting requires testing whether all K_n wavelet coefficients are equal to 0, while now we can test whether the single variance component σ_w equals zero.

We shall employ the profile restricted log-likelihood ratio test. With $\lambda = \sigma_u^2 / \sigma_\varepsilon^2$, the restricted log-likelihood of the data under the alternative model H_a, with σ_ε^2 substituted by its restricted maximum likelihood (REML) estimator is, up to a constant not depending on the data,

$$\mathcal{L}(\lambda) = -\frac{1}{2} \log |\boldsymbol{V}_\lambda| - \frac{1}{2} \log |\boldsymbol{X}^T \boldsymbol{V}_\lambda^{-1} \boldsymbol{X}| - \frac{n - q - 1}{2} \log \left[\boldsymbol{Y}^T \boldsymbol{P}(\lambda)^T \boldsymbol{V}_\lambda^{-1} \boldsymbol{P}(\lambda) \boldsymbol{Y} \right].$$

The projection matrix is defined as $\boldsymbol{P}(\lambda) = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{V}_\lambda^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}_\lambda^{-1}$, with $\boldsymbol{V}_\lambda = \boldsymbol{I}_n + \lambda \boldsymbol{Z} \boldsymbol{I}_{K_n} \boldsymbol{Z}^T$ and $\boldsymbol{I}_a$ is an identity matrix of dimension $a \times a$. The restricted log-likelihood under H_0 is obtained by setting $\lambda = 0$. The restricted profile likelihood ratio statistic is $\mathcal{R}_n = 2\{\mathcal{L}(\hat{\lambda}) - \mathcal{L}(0)\}$.

Since the parameter of interest, σ_w^2, under the null hypothesis is on the boundary of its parameter space $[0, \infty)$, the classical result that $\mathcal{R}_n \to \chi_1^2$ in distribution under H_0 does not hold. Under certain independence assumptions that the response variable vector can be partitioned into J_n independent identically distributed subvectors with $J_n \to \infty$, the asymptotic theory of Self and Liang (1987) and Stram and Lee (1994) suggest that, under H_0, the asymptotic distribution of a likelihood ratio statistic is a 50:50 mixture between a χ_0^2 and χ_1^2, where χ_0^2 means a point mass at zero. In a setting of testing with penalised regression splines, Crainiceanu and Ruppert (2004) found the distribution of $\mathcal{R}_n$ for finite samples to be different from the asymptotic result. Claeskens (2004) obtained conditions for random spline models under which those results apply. One assumption was that the number of spline coefficients needs to increase to infinity at a rate $o(n)$. This assumption does not hold for this setting, where rather K_n goes to zero under the null hypothesis as n grows. Neither do the results of Self and Liang (1987), nor the more general results of Vu and Zhou (1997) directly apply. Both papers assume that the score value converges to a normal random variable. In this paper we obtain the asymptotic distribution of the restricted likelihood ratio test for mixed models with a fixed and finite number of columns in the random effects matrix. The obtained results also explain the different mixture proportions observed by simulations (see, for example, Pinheiro and Bates, 2000) who observed in simulations that the mixture coefficients were rather approximating the values 0.65 and 0.35. We show in Section 4 that the exact proportions depend on the design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ through the eigenvalues of the matrix $\boldsymbol{Z}^t \boldsymbol{P}(0) \boldsymbol{Z}$. In case there is only one random effect (Z has only a single column) then the mixing proportions are given by $P(|N| \leq 1)$ and $P(|N| > 1)$ for $N \sim N(0, 1)$, which are equal to about 0.68 and 0.32. When the number of columns grows ($K_n \to \infty$), the mixing proportions converge to the values 0.5, 0.5. The theoretical results in this paper are not only of interest to the specific case of hypothesis testing with wavelets, but are of general interest in testing in mixed models with a finite number of random effects where the results of Self and Liang (1987) and Vu

and Zhou (1997) are not applicable.

4   Asymptotic distribution theory

We first consider the case of the wavelet-based lack-of-fit test, which due to the thresholding, asymptotically can be considered as a special case of testing in mixed effects models with a single random effect. Next, we obtain the asymptotic distribution of a test on the variance component when there are a finite number (not depending on the sample size) of random effects.

4.1   Distribution of the wavelet-based lack-of-fit test

The thresholding scheme determines the columns to be selected in the matrix $\mathbf{Z}$. More precisely, for the universal threshold a column is selected if the corresponding scaled empirical wavelet detail coefficient

$$\hat{d}_{jk}/\hat{\sigma}_{\varepsilon} > \sqrt{2\log(n)}$$

where we use as scale estimator

$$\hat{\sigma}_{\varepsilon} = \text{med}\{|\hat{d}_{J-1,k} - \text{med}(\hat{d}_{J-1,k})|\}/0.6745$$

employing the coefficients at the finest level. Under the null hypothesis, by using properties of extremes of normal random variables, the probability that a scaled wavelet detail coefficients exceeds the threshold tends to zero if we apply the universal threshold (Donoho, 1995). This "statistical upper bound" is often replaced by less conservative methods, for instance, methods that control the false discovery rate (Benjamini and Hochberg, 1995) or even methods that do not concentrate on false positives but rather on the average squared error of the estimates reconstructed from an inverse transform. These methods include SURE (Stein's Unbiased Risk estimator) (Donoho and Johnstone, 1995) and cross validation (Nason, 1996) or generalised cross validation (Jansen et al., 1997).

We restrict the discussion in this paper to universal thresholds. As a consequence of the above mentioned property on the scaled wavelet detail coefficients, the statistic $\mathcal{R}_n$ would tend to zero in probability under the null hypothesis if we would not impose the minimum number of one column to be selected. By restricting this number of columns, under the null hypothesis, asymptotically the number of columns is equal to one and a non-trivial limit distribution results. In a different context, a similar construction happens with the data driven Neyman smooth tests (Ledwina, 1994) in testing goodness of fit, where the consistent model selection method BIC (Schwarz, 1978) is not allowed to pick zero as model order, but rather any strictly positive

integer. Thus the asymptotic distribution result is the same as that of when applying a restricted likelihood ratio test in a mixed effects model with a nonzero random effects matrix $\boldsymbol{Z}$ of dimension $n \times 1$.

Theorem 1 *In a normal linear mixed effects model, with a nonzero random effects matrix $\boldsymbol{Z}$ obtained by wavelet thresholding using the universal threshold, under H_0, the statistic $\mathcal{R}_n$ converges in distribution to a mixture distribution which consists of $0.5(N^2 - 1)^2$ where $N \sim N(0, 1)$, with probability $P(|N| \geq 1)$ and a point mass at zero with the complementary probability $P(|N| < 1)$.*

This theorem can be seen as a special case of testing for the zeroness of one variance component where the random effects design matrix $\boldsymbol{Z}$ has a fixed and finite number K different columns, see Theorem 2, where also the proof can be found.

Thus, in a mixed model with a single random effect, when testing whether the variance component is zero or positive, the asymptotic distribution is a mixture containing a pointmass at zero with probability $0.68 = P(|N| < 1)$, which is much larger than the value 0.5. The other component is not a chi-squared one random variable, but rather a squared 'normalized' chi-squared variable. This component of the mixture is associated with probability $0.32 = P(|N| \geq 1)$, which is much less than 0.5. The reason for not obtaining 0.5 turns out to be the fact that the score (first derivative of the log likelihood) does not converge to a normal distribution for growing sample size when the number of columns of $\boldsymbol{Z}$ is not growing to infinity with n. For more details we refer to the proof of Theorem 2.

4.2 Restricted likelihood ratio tests in mixed models

Consider a general linear mixed model, represented by the following matrix form

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon},$$

where the $n \times (q + 1)$ design matrix $\boldsymbol{X}$ contains all the fixed effects (including the intercept and not necessarily restricted to only polynomials in x), and the $n \times K$ matrix $\boldsymbol{Z}$ contains all random effects of the model (wavelet basis function are one example, splines are another, it can also be person-specific effects in a clinical trial, etc.). Assume that the random effects follow a normal distribution with mean zero and variance σ_u^2. Denote by $\mathcal{R}_n$ the restricted likelihood ratio statistic for testing

$$H_0 : \sigma_u^2 = 0 \text{ versus } H_a : \sigma_u^2 > 0,$$

thus $\mathcal{R}_n$ contrasts the fixed effects model with the full random effect model. Its asymptotic distribution is given by the next theorem.

Theorem 2 *Assume a normal linear mixed effects model, with a $n \times (q+1)$ design matrix $\boldsymbol{X}$ contains all the fixed effects (including the intercept) and a nonzero random effects matrix $\boldsymbol{Z}$ of dimension $n \times K$ and full rank $K < n - q - 3$. Denote by ξ_k ($k = 1, \ldots, K$) the eigenvalues of the matrix $\lim_{n \to \infty} \boldsymbol{Z}^t \boldsymbol{P}(0) \boldsymbol{Z}$ and consider for $k = 1, \ldots, K$, $N_k \sim N(0,1)$ all independent of each other.*

Then, under H_0, the statistic $\mathcal{R}_n$ converges in distribution to a mixture distribution which consists of the random variable

$$\frac{0.5\{\sum_{k=1}^{K} \xi_k (N_k^2 - 1)\}^2}{\sum_{k=1}^{K} \xi_k^2}$$

with probability $P(\sum_{k=1}^{K} \xi_k (N_k^2 - 1) \geq 0)$, and a point mass at zero with the complementary probability $P(\sum_{k=1}^{K} \xi_k (N_k^2 - 1) < 0)$.

Proof. We start with a spectral decomposition of the restricted likelihood ratio statistic (see Claeskens, 2004, Lemma 1) and take the first two derivatives with respect to λ. Under H_0,

$$\mathcal{L}'(0) = -\frac{1}{2}\sum_{k=1}^{K} \xi_k + \frac{1}{2}(\sum_{k=1}^{K} \xi_k N_k^2)(1 + o_P(1)),$$

$$\mathcal{L}''(0) = \frac{1}{2}\sum_{k=1}^{K} \xi_k^2 - \left\{ \sum_{k=1}^{K} \xi_k^2 N_k^2 - \frac{1}{2}\frac{1}{n-q-1}(\sum_{k=1}^{K} \xi_k N_k^2)^2 \right\}(1 + o_P(1)).$$

It follows that for n tending to infinity $E\{\mathcal{L}'(0)\} = 0$ and $E\{(\mathcal{L}'(0))^2\} = \frac{1}{2}\sum_{k=1}^{K} \xi_k^2$. Further,

$$E\{-\mathcal{L}''(0)\} = \frac{1}{2}\frac{1}{n-q-1}\{(n-q-3)\sum_{k=1}^{K} \xi_k^2 - (\sum_{k=1}^{K} \xi_k)^2\}.$$

By Chebychev's inequality, this quantity is strictly positive as long as $n > K+q+3$. For $n \to \infty$, $E\{-\mathcal{L}''(0)\} - E\{(\mathcal{L}'(0))^2\} \to 0$. For the remainder of the proof we follow Vu and Zhou (1997) with the exception that their conditions (B2) and (B5) do not hold. Condition (B2) requires that the Fisher information value converges to infinity, while in our case it converges to a strictly positive constant. This is, however, sufficient for the proof since only the positive definiteness is required. Their condition (B5) requires convergence to a standard normal random variable of the standardized score value

$$U_{n,K} = \mathcal{L}'(0)/[E\{(\mathcal{L}'(0))^2\}]^{1/2},$$

which in our case with fixed K has a different limit distribution. This means that we diverge from the proof of their Theorem 2.2 starting from their equation (4.18), to arrive at the representation

$$\mathcal{R}_n = |U_{n,K}|^2 - \inf_{\lambda \in [0,\infty)} |U_{n,K} - \lambda|^2 + o_P(1).$$

For the computation of the asymptotic distribution we need to consider two possibilities. Either $U_{n,K} \geq 0$, in which case $\mathcal{R}_n = U_{n,K}^2 + o_P(1)$, or $U_{n,K} < 0$ and $R_n = o_P(1)$. The event that $\mathcal{R}_n \to 0$ in probability happens with probability $P(U_{n,K} < 0)$ as claimed. $\square$

In case $K = K_n$ does not remain fixed but diverges to infinity, Theorem 1 of Claeskens (2004) formulates conditions under which sequences of the form U_{n,K_n} converge to a standard normal random variable. In such case the theory of Vu and Zhou (1997) is applicable and the resulting distribution is an equal mixture of a χ_1^2 distribution and a point mass at zero. The theorems above give the asymptotic distribution in case the approximation by a normal random variable is not appropriate. The mixing probabilities can either be simulated or computed via algorithms such as that of Davies (1980). This also clearly points out the connection with the exact likelihood ratio test of Crainiceanu et al. (2005) where the simulation starts directly from the representation of the test statistic in terms of the eigenvalues ξ_k and standard normal random variables.

As an example we apply Theorem 2 to the real dataset `Machines` of Pinheiro and Bates (2000) and look at the mixing proportions in the asymptotic distribution of the test statistic $\mathcal{R}_n$ under H_0. We expect for $K > 1$ the pointmass at zero to occur with a probability within the interval $(0.5, 0.68)$ for the random variables consisted in $\mathcal{R}_n$.

We model a fixed effect β_j for each type of machine or `Machine` factor and a random effect b_i for each worker or `Worker` factor. Because the workers represent a random sample from the population of interest, any interaction terms modeling difference between workers in changing from one machine to another will also be expressed as random effects b_{ij}. We further assume that all the random effects have the same variance matrix. The model can be written as

$$Y_{ijk} = \beta_j + b_i + b_{ij} + \varepsilon_{ijk}, \quad i = 1, ..., 6, \quad j = 1, ..., 3, \quad k = 1, ...3,$$
$$b_i, b_{ij} \sim \mathcal{N}(0, \sigma_b^2), \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2).$$

To express this model in its matrix/vector representation as described in the beginning of this subsection, we can define the formulae that generate design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ as model input matrices. To be more specific, We define $\boldsymbol{X}$ to be 54×3 fixed-effects design matrix and $\boldsymbol{Z}$ to be the 54×18 random-effects design matrix. So K equals 18 in this case. To apply Theorem 2, it is straight forward to obtain the eigenvalues of the matrix $\boldsymbol{Z}^t \boldsymbol{P}(0) \boldsymbol{Z}$ as well as the generated normal random variables $N_k \sim N(0, 1)$ for $k = 1, \ldots, K$. We can therefore calculate the probability $P(\sum_{k=1}^{K} \xi_k (N_k^2 - 1) < 0)$ via simulations. The result shows that the probability is 0.5742 after 10,000 simulations, which is within the interval $(0.5, 0.68)$ as expected, and differs from the value 0.5.

5 Extensions

5.1 Testing with additive alternatives

The wavelet mixed model can be easily extended to additive models. In such case an alternative model is built as

$$Y = X\beta + Z_1 u_1 + \ldots + Z_a u_a + \varepsilon.$$

The simplest case is under the assumption that all random vectors u_j are independent with the same variance component σ_u^2. Thresholding can be applied to each additive component. Theorem 1 still applies if we restrict the number of columns of all random effect matrices Z_j under the null hypothesis to be at least one. In a more general case, we could restrict all the random effects matrices Z_j to have the same dimension $n \times K$ with $K \geq 1$, which means that, for example, we allow the wavelet basis function to be different for the different additive components, but keep their column numbers the same. From lemma 1 of Claeskens (2004) follows that the relevant eigenvalues ξ_k are the non-zero eigenvalues of the matrix $\lim_{n \to \infty} \{Z_1^T P(0) Z_1 + \ldots + Z_a^T P(0) Z_a\}$. The mixture distribution of $0.5\{\sum_{k=1}^{K} \xi_k (N_k^2 - 1)\}^2 / \sum_{k=1}^{K} \xi_k^2$ where $N_k \sim N(0,1)$ with probability $P(\sum_{k=1}^{K} \xi_k (N_K^2 - 1) \geq 0)$ and a point mass at zero with the complementary probability $P(\sum_{k=1}^{K} \xi_k (N_K^2 - 1) < 0)$ holds asymptotically as in Theorem 2.

5.2 Testing for more than one variance component

When building the wavelet mixed effects model, we assumed an equal distribution of all wavelet coefficients. This had the advantage of circumventing the multiple or joint testing problem and reducing it to a single hypothesis test. In some circumstances this might not be optimal. Wavelet coefficients corresponding to coarse scales might have a different variance than those belonging to finer scales. One could even consider a different variance for each scale. We then have a set of variance components $\sigma_{u_1}^2, \ldots, \sigma_{u_a}^2$ which are all set to zero under a null hypothesis. The alternative hypothesis then consists of the mixed effects model with at least one non-zero variance component.

Another example is a mixed effect model with random effects both on the level of the hospital and the patient, there might also be a random patient-by-hospital interaction, resulting in three variance components. A relevant null hypothesis of interest might be whether all three random effects are zero, which could lead to simplifying the model.

For such examples an extension of the results above is required. As with boundary testing problems in general, the complexity dramatically increases when testing

for the zeroness of more than one variance component. Again, the main point of deviation with existing proofs is the lack of normality of the score statistic. Consider the linear mixed model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \ldots + \boldsymbol{Z}_a\boldsymbol{u}_a + \boldsymbol{\varepsilon},$$

with independent random effects $\boldsymbol{u_j} \sim N(\boldsymbol{0}, \sigma_{u_j}^2\boldsymbol{I})$. The fixed effects design matrix $\boldsymbol{X}$ has dimension $n \times (q+1)$, and includes a possible intercept. The design matrices $\boldsymbol{Z}_j$ of the random effects have dimension $n \times K_j$. The null hypothesis is

$$H_0 : \sigma_{u_1}^2 = \ldots = \sigma_{u_a}^2 = 0.$$

This is contrasted with the alternative hypothesis that at least one of the $\sigma_{u_j}^2 > 0$. Let $\lambda_j = \sigma_{u_j}^2/\sigma_\varepsilon^2$. The matrix representation of the first derivative of the restricted log likelihood with respect to the λ_j's, that is, the score vector, is obtained as

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = -\frac{1}{2}\text{tr}(\boldsymbol{Z}_j\boldsymbol{Z}_j^T\boldsymbol{V}_\lambda^{-1}\boldsymbol{P}(\lambda)) + \frac{n-q-1}{2} \cdot \frac{\boldsymbol{Y}^T\boldsymbol{P}(\lambda)\boldsymbol{V}_\lambda^{-1}\boldsymbol{Z}_j\boldsymbol{Z}_j^T\boldsymbol{V}_\lambda^{-1}\boldsymbol{P}(\lambda)\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{P}(\lambda)\boldsymbol{V}_\lambda^{-1}\boldsymbol{P}(\lambda)\boldsymbol{Y}},$$

for $j = 1, \ldots, a$ where $\boldsymbol{V}_\lambda = \boldsymbol{I}_n + \sum_{j=1}^a \lambda_j\boldsymbol{Z}_j\boldsymbol{Z}_j^T$ and $\boldsymbol{P}(\lambda)$ is as defined in Section 3. Note that $\boldsymbol{V}_0 = \boldsymbol{I}_n$. What is needed is the standardised score vector under the null hypothesis. After calculations similar to those needed for the proof of Theorem 2, this leads to a study of the following vector V_n with components

$$V_{n,j} = \frac{\frac{1}{2}\boldsymbol{U}^T\boldsymbol{P}(0)^T\boldsymbol{Z}_j\boldsymbol{Z}_j^T\boldsymbol{P}(0)\boldsymbol{U} - \frac{1}{2}\sum_{k=1}^{K_j}\xi_{j,k}}{\{\frac{1}{2}\sum_{k=1}^{K_j}\xi_{j,k}^2\}^{1/2}},$$

where for each $j = 1, \ldots, a$ the $\xi_{j,k}$'s are the non-zero eigenvalues of the matrix $\boldsymbol{Z}_j^T\boldsymbol{P}(0)\boldsymbol{Z}_j$ and U is a vector of standard normal random variables. Similar as in the one variance component models, this score vector needs to be projected on the cone defined by the boundary constraints. First, define the information matrix

$$\boldsymbol{G}_{n,\ell k} = E\{\frac{\partial \mathcal{L}}{\partial \lambda_\ell}(0,0)\frac{\partial \mathcal{L}}{\partial \lambda_k}(0,0)\} = \frac{1}{2}\text{tr}\{(\boldsymbol{Z}_\ell^T\boldsymbol{P}(0)\boldsymbol{Z}_k)(\boldsymbol{Z}_\ell^T\boldsymbol{P}(0)\boldsymbol{Z}_k)^T\}.$$

Since all variance components need to be non-negative, the parameter space for $(\lambda_1, \ldots, \lambda_a)$ is given by $\Omega = [0, +\infty)^a$. We now define the finite sample cone

$$C_{\Omega_n} = \left\{(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_a)^T = \boldsymbol{G}_n^{T/2} \cdot (\lambda_1, \ldots, \lambda_a)^T \; : \; (\lambda_1, \ldots, \lambda_a)^T \in \Omega\right\},$$

where $\boldsymbol{G}_n^{T/2}$ is the Cholesky square root matrix of $\boldsymbol{G}_n$.

In the case of one variance component, $C_{\Omega,n}$ is simply equal to $[0, +\infty)$. For the case of two variance components, a limit version of this cone can be explicitly

obtained, see Claeskens (2004), Section 5.2. In that case, with s the limit of $s_n = \boldsymbol{G}_{n,12}/\sqrt{\det(\boldsymbol{G}_n)}$, the limiting cone is given by

$$C_\Omega = \left\{ (\lambda_1, \lambda_2)^T : \ \lambda_1 - s\lambda_2 \geq 0, \ \lambda_2 \geq 0 \right\}.$$

In general, the Cholesky square root matrix is easy to compute in practice, but it is more difficult to state simple formulae. The asymptotic distribution of the restricted likelihood ratio test is now obtained by computing the limiting distance of the standardised score vector V_n to the limiting cone C_Ω. For finite sample results, we can compute the distance of V_n to $C_{\Omega,n}$. The main difference between this result and those obtained earlier in literature, is that V_n does not have a multivariate normal distribution. For an example of the special case where the score vector is bivariate normal in the limit, see Claeskens (2004), Theorem 2. Other examples are provided by Self and Liang (1987), see also Vu and Zhou (1997).

6   Simulation study

In this section, we first consider the empirical distribution of our proposed lack-of-fit test statistic $\mathcal{R}_n$ using the mixed wavelet model through an intensive Monte Carlo study. Next, we investigate the power properties of the test statistic and compare the results with those obtained by applying the bootstrap. We compare our test with the test based on penalized splines by Crainiceanu and Ruppert (2004) and Claeskens (2004), both in terms of size under the null and power.

As examples we considered testing the null hypothesis of a normal linear regression model $Y = \mu(x) + \varepsilon$ with the true $\mu(x) = 8x + 3$. We generated data according to this model with sample sizes equal to 64 and 128 and the covariates x_i $(i = 1, \ldots, n)$ are equidistant on the interval [0,1]. We use two types of wavelet basis functions, the Haar wavelet and Daubechies' wavelet with four vanishing moments. The primary resolution level for the wavelet transformation is set at $j_0 = 1$. To avoid the number of columns K_n of the random effects matrix $\boldsymbol{Z}$ to be zero after the thresholding procedure, we arbitrarily select the 10th column of the inverse discrete wavelet transform matrix and include it in the construction of $\boldsymbol{Z}$ matrix whenever K_n would be zero after thresholding.

For comparison, we also include the spline based test. In the penalized spline mixed model, we used a truncated linear spline basis where $\psi_j = \max\{(x - \kappa_j), 0\}$ for knots κ_j with 35 knots at sample quantiles. The test statistic for the penalized spline model is denoted by $\mathcal{R}_{ns}$.

6.1 Empirical distribution

We simulated 50,000 sets of Y's from the above polynomial null model, and obtained the corresponding restricted likelihood ratio statistics $\mathcal{R}_n$ and for the spline based test, $\mathcal{R}_{ns}$.

Under the null hypothesis, we found that the empirical distributions of the test statistics $\mathcal{R}_n$ by using different wavelet bases are almost the same, therefore we only present the histogram of $\mathcal{R}_n$ with the Daubechies 4 basis in Figure 1. Also included is the corresponding histogram of $\mathcal{R}_{ns}$, both for sample size 64.
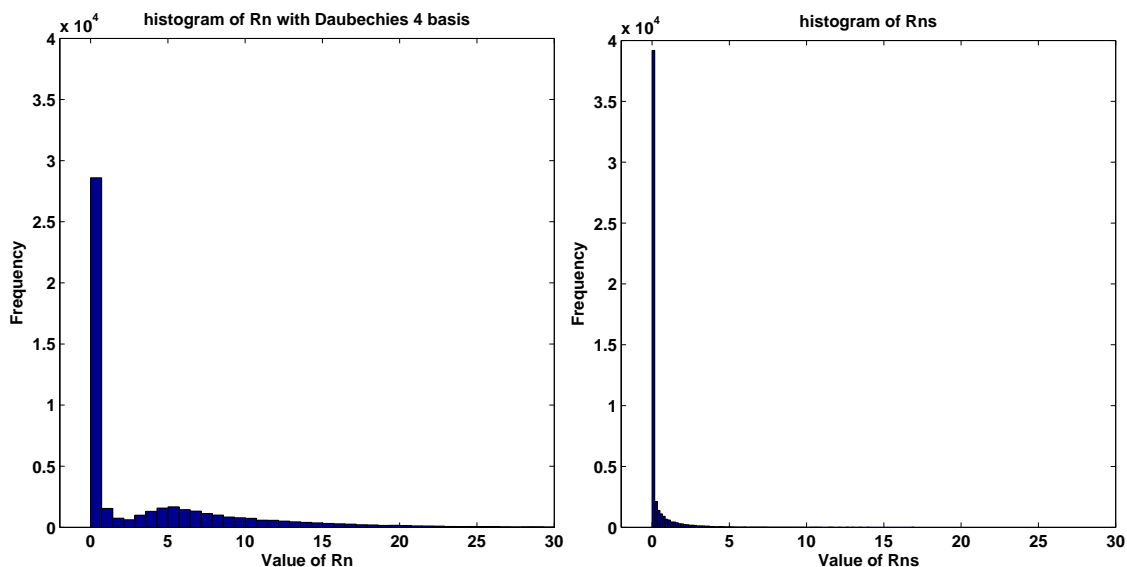

Figure 1: Histograms of test statistics $\mathcal{R}_n$n and $\mathcal{R}_{ns}$ under the null hypothesis when the sample size is 64. The wavelet based test uses the Daubechies 4 basis function, while the spline based test uses a truncated linear spline basis with 35 knots.

We first discuss the spline based test. Crainiceanu and Ruppert (2004) already pointed out the discrepancy between the expected 0.50: 0.50 mixture of a pointmass at zero and a chi-squared random variable with one degree of freedom. Indeed, we observe that the null distribution of the restricted likelihood ratio statistic $\mathcal{R}_{ns}$ obtained from the simulation has $P_0 = 0.6592$ probability mass at zero which is close to the 65:35 mixture of a point mass at zero and a χ_1^2 distribution as suggested by Pinheiro and Bates (2000). The 0.90, 0.95, and 0.99 quantiles of this distribution are approximately $q_{0.9} = 0.9225$, $q_{0.95} = 1.8666$ and $q_{0.99} = 4.2957$.

14

The histogram of $\mathcal{R}_n$ has two peaks, with one higher peak at the value zero and the other one around 5. Seemingly it does not follow the mixture distribution between $0.5(N^2 - 1)^2$ with probability 0.68 and a point mass at zero with the complementary probability 0.32 as we expect from Theorem 1. The explanation for this is that one of the conditions required by Theorem 1, namely that the random effects matrix Z is of dimension $n \times 1$, is not always satisfied in our finite sample simulation. It is only asymptotically that the probability that a scaled wavelet detail coefficients exceeds the universal threshold goes to zero. In practice, also under the null hypothesis, it happens that there is more than one column in the resulting Z matrix, which in turn gives large values of $\mathcal{R}_n$ and contributes to the second peak in the histogram.
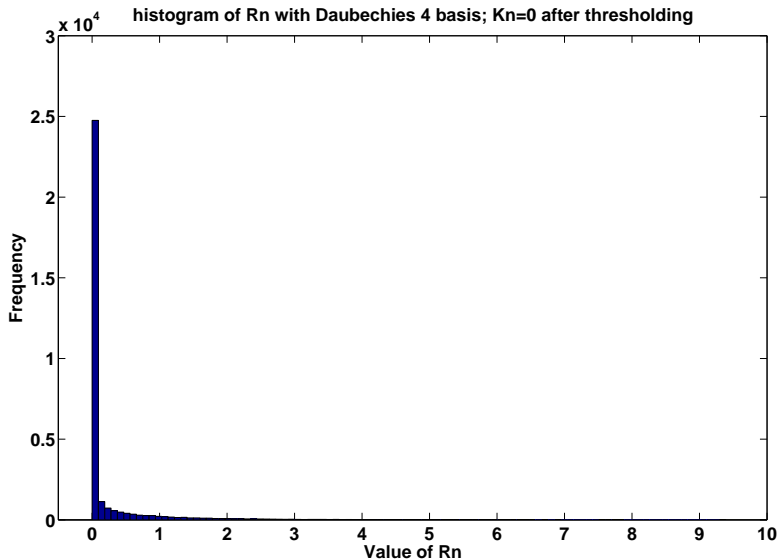

Figure 2: Histogram of the values of the test statistics $\mathcal{R}_n$ using the Daubechies 4 wavelet basis with sample size 64, only considering the cases where $K_n = 0$ after thresholding with the universal threshold.

Figure 2 presents the histogram of the wavelet-based test $\mathcal{R}_n$ that is constructed by only considering 31,742 out of 50,000 cases where the corresponding number of columns of the $\boldsymbol{Z}$ matrix is exactly zero after thresholding under the null hypothesis. In this case, Theorem 1 immediately applies. The observed probability mass at zero is 0.6823, coming very close to the theoretical value of $0.6827 = P(|N| \leq 1)$, even with this small sample size of 64. For increasing sample size we observed that the proportion of times that $K_n = 0$ under the null hypothesis, also increases.

In order to investigate the level of the test, in addition to the critical values obtained from the asymptotic distribution, we also use the empirical critical values, obtained from 50,000 simulated sets of data under the null hypothesis. We obtain

these values for the wavelet based test $\mathcal{R}_n$ (for both types of wavelet basis functions) as well as for the spline based test $\mathcal{R}_{ns}$. The nominal α levels of the lack-of-fit tests are set at 1%, 5% and 10%.

We now use 10,000 and 1000 independently simulated sets of data from the above models and calculate the test statistics $\mathcal{R}_n$ and $\mathcal{R}_{ns}$. Tables 1 and 2 report the frequencies of exceeding the $1 - \alpha$ quantile at different α levels by using our wavelet mixed model as well as the penalized spline model.

Table 1 shows that with this small sample size, the approximation of the asymptotic distribution in the tail is not working quite well. Simulations with bigger sample sizes (results not shown) indicate only slow convergence to the nominal values for the wavelet-based test, which might have its main reason in properties of the thresholding procedure. Using the empirical critical values obtained from the large simulation study of size 50,000 does give good results for all three nominal levels of 0.10. 0.05 and 0.01, and for both types of wavelet basis functions.

For the results of the spline based test, in Table 2, we include the rejection proportion under the assumption that $\mathcal{R}_{ns}$ follows the $50\chi_0^2 : 50\chi_1^2$ mixture distribution. We observe that for these models the simulated rejection proportions tend to their nominal values when based on the empirical distributions, but deviate from the assumed asymptotic 50:50 mixture distribution in the penalized spline model.

Table 1: Simulated rejection probabilities under the null hypothesis for the test in the mixed wavelet model.

Wavelet basis	nRuns	Quantiles of test statistics $\mathcal{R}_n$			% Reject (Emp.Distr $K_n = 0$) [a]			% Reject (Asym.Distr) [b]		
		0.90	0.95	0.99	0.10	0.05	0.01	0.14	0.10	0.05
Haar	50000	11.5501	15.7995	24.8890	–	–	–	0.1389	0.1256	0.1211
Haar	10000	11.3974	15.7794	24.1571	0.0974	0.0499	0.0086	0.1388	0.1252	0.1208
Haar	1000	12.7089	15.5370	25.5231	0.1160	0.0480	0.0110	0.1250	0.1044	0.0997
Daubechies4	50000	11.5168	15.6165	24.7621	–	–	–	0.1417	0.1296	0.1244
Daubechies4	10000	11.8182	16.0044	24.6747	0.1053	0.0534	0.0098	0.1420	0.1304	0.1238
Daubechies4	1000	11.5248	16.2152	26.5490	0.1010	0.0570	0.0170	0.1352	0.1254	0.1189

[a] compare with empirical distribution (50,000 runs)
[b] compare with $0.68 * (0.5 * (\chi_1 - 1)^2) + 0.32 * 0$ distribution for $\mathcal{R}_n$ when its corresponding $K_n = 0$ after thresholding.

Table 2: Simulated rejection probabilities under the null hypothesis for the test in the penalized spline mixed model

nRuns	Quantiles of test statistics $\mathcal{R}_{ns}$			Reject (Emp.Distr) [a]			% Reject (Asym.Distr)[b]		
	0.90	0.95	0.99	0.10	0.05	0.01	0.10	0.05	0.01
50000	0.9225	1.8666	4.2957	–	–	–	0.0585	0.0285	0.0048
10000	0.9225	1.8946	4.3232	0.1000	0.0510	0.0103	0.0593	0.0286	0.0053
1000	1.1480	2.2792	4.0764	0.1160	0.0630	0.0090	0.0740	0.0390	0.0030
50000	0.9225	1.8666	4.2957	–	–	–	0.0585	0.0285	0.0048
10000	0.9424	1.9462	4.3375	0.1026	0.0528	0.0105	0.0618	0.0305	0.0053
1000	1.1480	2.2792	4.0764	0.1160	0.0630	0.0090	0.0740	0.0390	0.0030

[a] compare with empirical distribution (50,000 runs)
[b] compare with $0.5 * \chi_1^2 + 0.5 * 0$ distribution

6.2 Testing power

In this section, we present the results of comparative simulations whose purpose was to investigate the power properties of the tests using $\mathcal{R}_n$ and $\mathcal{R}_{ns}$. Our test functions are the familiar wavelet test functions, Donoho and Johnstone's (1994) 'Blocks','Bumps','Doppler', and 'Heavisine' function, as well as the cosine function $0.4\cos(\pi j x_i)$ for $j = 0, 0.1, \ldots, 9$. The last function is included since it is expected that the spline based test will perform well for this alternative function. Data are generated under an alternative model of the form $Y = \mu(x) + b \cdot g(x) + \varepsilon$ where $g(\cdot)$ is one of the test functions above, and $\mu(\cdot)$ is the null hypothesis linear model. We use the same null model described in the previous section, and obtain simulated rejection probabilities under a sequence of alternative models with increasing coefficients or frequencies of the test functions (constant b in the above model).

The following wavelets bases are used, Daubechies 4 for 'Bumps', 'Doppler', 'Cosine', Haar wavelets for 'Blocks' and Symmlet 8 for 'HeaviSine'.

Simulated power curves, based on 1000 simulated data sets, are depicted in Figures 3–5. Simulated power curves are shown using the critical values obtained from the empirical distribution under the null hypothesis based on 50,000 simulated data sets. This guarantees a fair comparison of the methods.

In Figure 3 we clearly observe that the new test in the wavelet mixed model has higher power than that in the penalized spline model for 'Blocks' and 'Bumps' functions which have high frequency alternations or sharp peaks. Results are very comparable for the doppler function, where the power curves are nearly identical. Wavelet based tests are, as expected, to perform less well for low frequency alternatives such as the 'HeaviSine' function and 'Cosine' function, as shown in Figures 4 and 5.
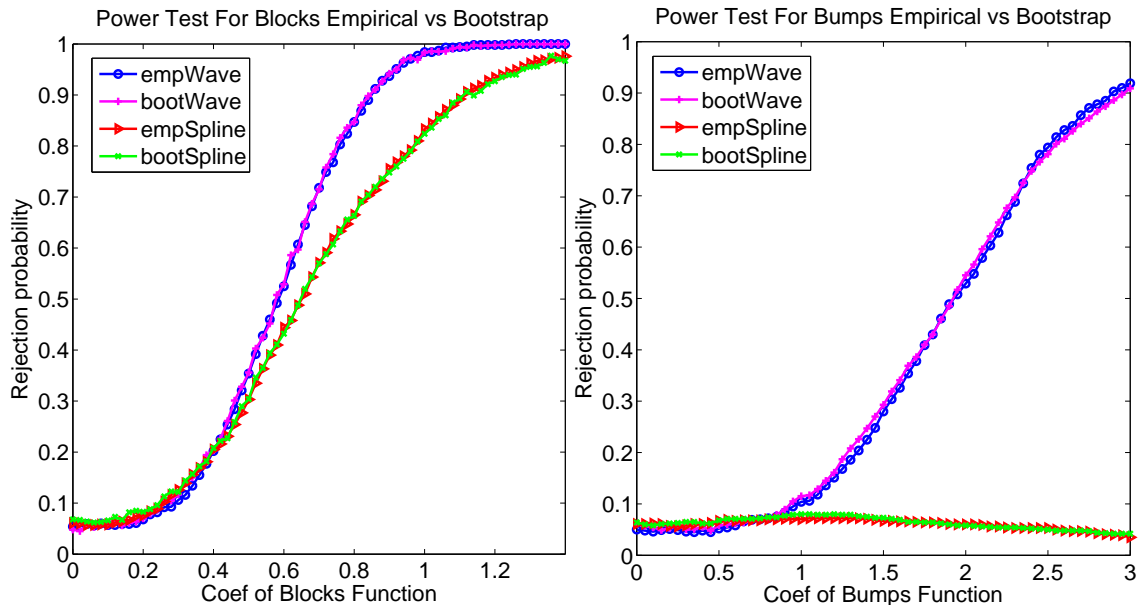
Figure 3: Simulated power curves of wavelet and spline based tests for the Blocks and Bumps alternative functions, using critical values from both the empirical and bootstrapped distribution.

As an alternative to working with the asymptotic distribution of the test statistic, we turn to bootstrap methods introduced by Efron (1979). For a review, see Davison and Hinkley (1997), and Efron and Tibshirani (1993). Large values of the restricted likelihood ratio (RLRT) statistic $\mathcal{R}_n$ supply evidence against H_0 and the level of evidence is measured by the P-value $p = P(\text{RLRT} \geq t_{obs}|H_0)$, where t_{obs} is the observed value of RLRT. A bootstrap value for p is obtained by comparing t_{obs} to the bootstrap distribution of RLRT under H_0. Hall and Wilson (1991) advocate resampling in a way that reflects the null hypothesis. For our testing situation this translates to first obtaining the residuals of the null model fit. Bootstrapping is then carried out by resampling these residuals, for which different procedures are possible. The parametric bootstrap employs the conventional estimate of the error variance $\hat{\sigma_{\varepsilon}}^2$ and resamples bootstrap errors ε^* with replacement from the normal distribution $N(0, \hat{\sigma_{\varepsilon}}^2)$, see Efron (1979) and Freedman (1981). For residual resampling (Efron and Tibshirani, 1986), in order to reflect the situation of the null hypothesis, bootstrap observations are constructed as $Y_i^* = X_i\hat{\beta} + \varepsilon_i^*, \ i = 1, \ldots, n$, where ε_i^* are the resampled null model residuals. Fitting the hypothesized model to the bootstrap data $(y_1^*, \ldots, y_n^*)$ yields the bootstrap test statistics. The entire process is then
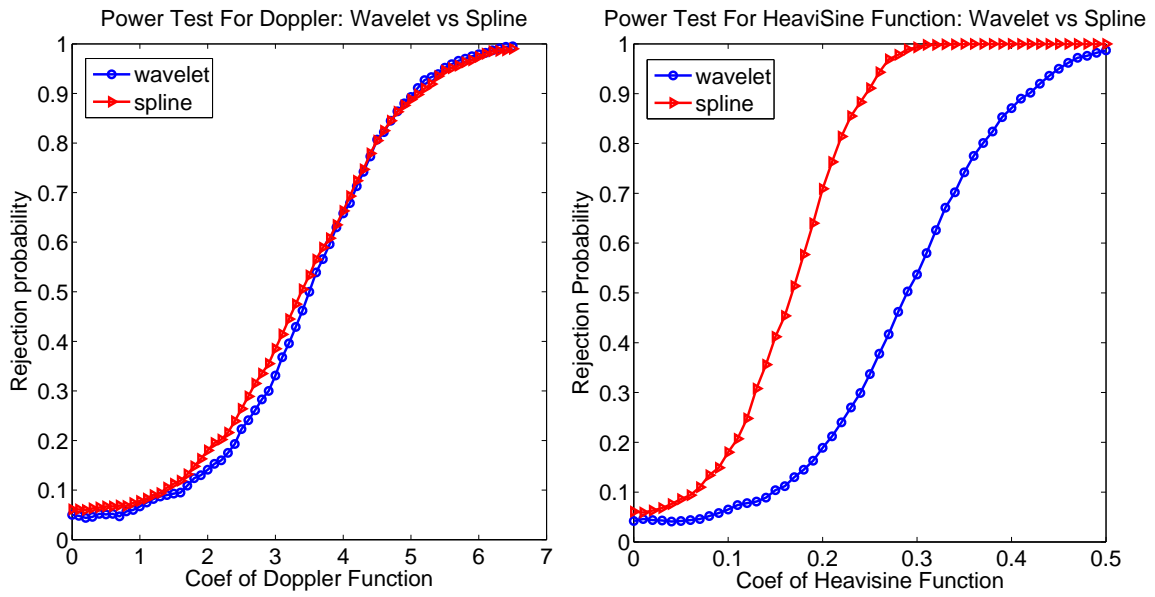
Figure 4: Simulated power curves of the wavelet and spline based tests for the Doppler and HeaviSine alternative functions, using the empirical critical values.
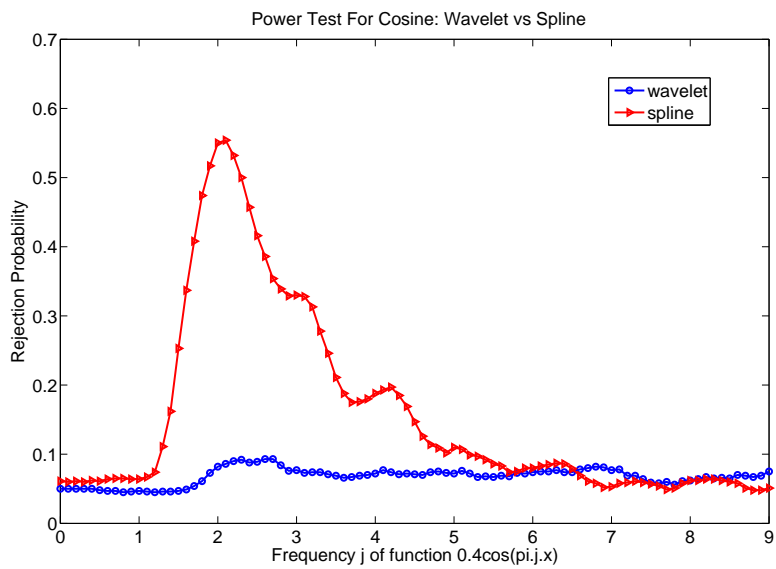

Figure 5: simulated power curves of the wavelet and spline based tests for the Cosine alternative function, using the empirical critical values.

repeated B times to obtain B test statistics. Finally, for a test with significance level α, we compare the original sample statistic to the $(1-\alpha)B$th percentile of the ordered bootstrap statistics.

A similar but smaller simulation study is performed to investigate the power properties of the test statistics under the bootstrapped distribution. We now run 1000 simulations each with 100 bootstrap runs to check the rejection probabilities. We implemented the residual bootstrap for the 'Blocks' function and the parametric bootstrap for the 'Bumps' function. The simulated power curves are shown in Figure 3. From the graph, we observe that the simulated power curves obtained from the bootstrap distribution and the empirical distribution are not distinguishable.

7   Discussion

The results on the asymptotic distribution of the likelihood ratio test in mixed linear models helps providing the 'missing link' in explaining the difference between simulation results and assumed theoretical results. While this paper is restricted to testing lack-of-fit in linear mixed models, it is expected that the results can be extended to nonlinear mixed models, or generalised linear mixed models.

The method of wavelets has shown to be quite fruitful in gaining testing power for situations which are not perfectly smooth (such as the blocks or bumps alternative). We believe that it might serve as an attractive companion to the penalised spline-based tests. Also for estimation purposes, this method looks promising. Studying estimation properties, however, is beyond the scope of this paper.

References

Aerts, M., Claeskens, G., and Hart, J. D. (1999). Testing the fit of a parametric function. *J. Am. Statist. Ass.*, 94:869–879.

Aerts, M., Claeskens, G., and Hart, J. D. (2000). Testing lack of fit in multiple regression. *Biometrika*, 87:405–424.

Angelini, C., D. C. D. and Leblanc, F. (2003). Wavelet regression estimation in nonparametric mixed effect models. *J. Multiv. Anal.*, 85:267–291.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statist. Soc., Ser. B*, 57(1):289–300.

Claeskens, G. (2004). Restricted likelihood ratio lack of fit tests using mixed spline models. *J. Royal Statist. Soc., Ser. B*, 66:909–926.

Crainiceanu, C., Ruppert, D., Claeskens, G., and Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika*, 92(1):91–103.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. Royal Statist. Soc., Ser. B*, 66:165–85.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996.

Davies, R. B. (1980). The distribution of a linear combination of χ^2 random variables (algorithm as 155). *Appl. Statist.*, 29:323–333.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge University Press.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, 90:1200–1224.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7:1–26.

Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Science*, 1:54–75.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Cambridge University Press, Cambridge.

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.*, 20:1412–1425.

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *J. Am. Statist. Ass.*, 91:674–688.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.*, 9:1218–1228.

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47:757–762.

21

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests.* Springer-Verlag, New York.

Jansen, M., Malfait, M., and Bultheel, A. (1997). Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44.

Ledwina, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Am. Statist. Ass.*, 89:1000–1005.

Nason, G. P. (1996). Wavelet shrinkage using cross validation. *J. Royal Statist. Soc., Ser. B*, 58:463–479.

Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer-Verlag, New York.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, 6:461–464.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Ass.*, 82:605–610.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498.

Stram, D. O. and Lee, J. W. (1994). Variance component testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177.

Vu, H. T. V. and Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist.*, 25:897–916.