

Testing linearity and relevance of ordinal predictors

Jan Gertheiss^{*,†}

*Department of Statistics
Ludwig-Maximilians-Universität Munich
Akademiestraße 1
80799 Munich, Germany*

e-mail: jan.gertheiss@stat.uni-muenchen.de

*Department of Statistics
North Carolina State University
2311 Stinson Drive
Raleigh, NC 27695-8203, USA*

and

Franziska Oehrlein

*Department of Statistics
Ludwig-Maximilians-Universität Munich
Akademiestraße 1
80799 Munich, Germany*

e-mail: franzi.oehrlein@gmx.de

Abstract: In a linear model relevance of a categorical predictor with ordered levels is typically tested by use of the standard F -test (known from statistical textbooks). Such a test can also be applied for testing whether the regression function is linear in the ordinal predictor's class labels. In this paper we propose an alternative (restricted) likelihood ratio test for these hypotheses which is especially suited for ordinal predictors and is based on the mixed model formulation of penalized dummy coefficients. We show in simulation studies that the new test is more powerful than the standard F -test in many situations. The advantage of the new test is especially striking when the number of ordered levels is moderate or large. Using the relationship to mixed effect models and robust existent fitting software obtaining the test and its null distribution is very fast; a fast R implementation is provided.

Keywords and phrases: Classical linear model, ordinal covariates, likelihood ratio tests, linear mixed models, smoothing, zero variance component.

Received August 2011.

1. Introduction

We consider an ordinal covariate x with levels $k = 0, \dots, K$. Such ordinal predictors are, for example, often found in the social sciences. Frequently, however,

^{*}To whom correspondence should be addressed.

[†]This research was supported in part by DFG project GE2353/1-1.

they are treated as continuous and the class label is directly used as explanatory variable in a (linear) regression model for predicting a response variable y . [Thomas \(1997\)](#), for example, used linear regression to describe the relationship between a manager's values and his or her success, with values being coded using an ordinal scale ranging from 1 (lowest priority) to 18 (highest priority). [Labowitz \(1970\)](#) used ordinally scaled prestige ratings to predict suicide rates for different occupations. Also [Kim \(1975, 1978\)](#) argues that Pearson's correlation can be used to quantify the association between two variables if one (or even both) of these is ordinal. But not only in the social sciences, ordinal covariates play an important role. In particular when rating schemes are used, variables are often ordinally scaled; and schemes like that are found in many fields, such as medicine. For example, cell characteristics may be graded on ordinal scales with 1 being the closest to normal tissue and 10 the most anaplastic ([Wolberg and Mangasarian, 1990](#)), or see [WHO \(2001\)](#), [Cieza et al. \(2004\)](#) and [Section 7](#) for an example from rehabilitation medicine.

In standard linear models the conditional expectation $E(y|x)$ of y given an ordinal predictor x , is assumed to have a simple form

$$E(y|x) = \alpha + \delta x,$$

or equivalently

$$y = \alpha + \delta x + \epsilon, \tag{1}$$

where ϵ is a zero-mean random variable. Typically, such models make the additional assumption that ϵ is normal and has variance σ^2 . So for given data (y_i, x_i) , $i = 1, \dots, n$, we have

$$y_i = \alpha + \delta x_i + \epsilon_i,$$

with independent identically distributed (iid) $\epsilon_i \sim N(0, \sigma^2)$. Such iid normal errors ϵ_i are assumed throughout this paper.

Even though model (1) is often used in practice, the variable x is ordinal and may alternatively be included using dummy-coding as

$$y = \alpha + \sum_{k=0}^K \beta_k m_k + \epsilon, \tag{2}$$

where the dummy variables m_k are defined as

$$m_k = \begin{cases} 1 & \text{if } x = k, \\ 0 & \text{otherwise.} \end{cases}$$

That means, a one-dimensional variable x coding the factor level by a single number is transformed into a multivariate variable $(m_0, \dots, m_K)^T$. For example, if x has levels $0, \dots, 5$, $x = 3$ is transformed into $(0, 0, 0, 1, 0, 0)^T$. Each dummy variable m_k has its own (dummy) regression coefficient β_k . For identifiability, we specify reference category $k = 0$, such that $\beta_0 = 0$.

Whether methods for interval-level data (as model (1)) can also be used for ordinal variables has long been debated, in particular in the social sciences; see, for example, [Winship and Mare \(1984\)](#), [Labowitz \(1970\)](#) and [Mayer \(1970, 1971\)](#). In the framework considered here, the question whether the linear model (1) is sufficient may be answered using a formal statistical test. Indeed, model (1) is a special case of the more flexible model (2), as (1) can be obtained from model (2) by imposing the restrictions

$$\beta_k = k\beta_1, \quad k = 2, \dots, K, \quad (3)$$

or, equivalently,

$$\beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, \quad k = 1, \dots, K - 1. \quad (4)$$

As restrictions (3) (and (4)) are linear and the parameter space of $\theta = (\alpha, \beta_1, \dots, \beta_K)^T$ from model (2) is \mathbb{R}^{K+1} , it follows that model (1) can be tested against model (2) using a standard F -test. If one is interested in the global test whether the predictor x is necessary, the null-hypothesis $\beta_k = 0$ for all $k = 1, \dots, K$ can also be tested using a standard F -test.

These F -tests are shortly summarized in the next section. Section 3 shows how penalized estimates for dummy-coded ordinal predictors can be obtained, and in Section 4 these estimates are derived in the context of linear mixed models. Using the mixed model formulation, we propose an alternative (restricted) likelihood ratio test for checking linearity – i.e., distinguish between model (1) and (2) – and checking relevance of an ordinal predictor (Section 5). Simulation studies in Section 6 show that the proposed test distinctly outperforms the standard F -test in many situations. In Section 7, real data examples are given.

2. Standard F-tests

Assuming normal errors ϵ , the F -test can be used to test linear null-hypotheses of the form $H_0: C\theta = d$, where $\theta = (\alpha, \beta_1, \dots, \beta_K)^T$ denotes the coefficient vector from the dummy model (2) and C is a $r \times (K + 1)$ matrix of rank r . The matrix C and vector d define the linear hypotheses that are to be tested. For testing, the likelihood ratio statistic

$$\lambda = \frac{\max_{\theta \in H_0} L(\theta)}{\max_{\theta \in H_0 \cup H_A} L(\theta)}$$

can be used, with $L(\theta)$ denoting the data likelihood evaluated at the parameter vector θ . H_A denotes the alternative hypothesis $C\theta \neq d$. Assuming H_0 is true, the monotone transformation

$$F = \frac{n - K - 1}{r} \{\lambda^{-2/n} - 1\} = \frac{n - K - 1}{r} \frac{\text{SSE}_0 - \text{SSE}}{\text{SSE}}$$

follows an F -distribution with $(r, n - K - 1)$ degrees of freedom (see, e.g., [Rao et al. \(2008\)](#)). SSE denotes the sum of squared errors under the full model, and SSE_0 is the sum of squared errors if restrictions $C\theta = d$ hold.

If model (1) is tested against model (2), then $r = K - 1$, $d = (0, \dots, 0)^T$, and

$$C = \begin{pmatrix} 0 & 2 & -1 & 0 & \cdots & 0 \\ 0 & 3 & 0 & -1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & K & 0 & \cdots & 0 & -1 \end{pmatrix},$$

or, equivalently,

$$C = \begin{pmatrix} 0 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

In the case when testing the global null hypothesis $H_0: \beta_k = 0$ for all $k = 1, \dots, K$, the C matrix becomes $C = (0|I_K)$, where I_K is the $K \times K$ identity matrix, and $r = K$.

3. Penalized estimates for ordinal predictors

When estimating the regression parameters in model (2) estimates may become unstable if the number of levels K is moderate or large. To address this problem Gertheiss and Tutz (2009) proposed a penalized maximum likelihood (ML) estimate. Specifically, the penalized log-likelihood

$$l_p(\theta) = l(\theta) - \gamma J(\theta) \tag{5}$$

is maximized, with $l(\theta)$ denoting the usual, non-penalized, likelihood. By adding the penalty $J(\theta)$, estimates of the regression parameters can be stabilized. The choice of the penalty $J(\theta)$ is crucial. Because the categories of the predictor x are ordered, it may be assumed that the response y (or its conditional expectation) changes smoothly between two adjacent levels k and $k - 1$ of x . This led Gertheiss and Tutz (2009) to propose the quadratic difference penalty

$$J(\theta) = \sum_{k=1}^K (\beta_k - \beta_{k-1})^2 = \beta^T \Omega_1 \beta, \tag{6}$$

where $\beta = (\beta_1, \dots, \beta_K)^T$, penalty matrix $\Omega_1 = D_1^T D_1$, and

$$D_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \tag{7}$$

An implementation of this penalty for ordinal predictors is provided in the R package `ordPens` (Gertheiss, 2011), and also given in the Appendix. The strength of the penalization in (5) is controlled by the penalty parameter γ . With $\gamma = 0$, pure ML estimates are obtained. With $\gamma \rightarrow \infty$, all estimates $\hat{\beta}_k$ equal $\beta_0 = 0$, which means that in the model response y is not influenced by predictor x . If $\hat{\beta}_k = 0 \forall k$ is desired for some $\gamma < \infty$, a group lasso (Yuan and Lin, 2006) penalty can be used (see Gertheiss et al. (2011) for details).

Under the assumption that the true coefficient vector β is non-linear but smooth, it has been shown in Gertheiss and Tutz (2009) that parameter estimation and prediction can be substantially improved using the proposed penalized likelihood compared to unpenalized likelihood, simple ridge estimation, or linear regression on the category labels.

Similarly, if deviations from a linear function – that is, from model (1) – are to be penalized, the penalty has the form

$$J(\theta) = \sum_{k=1}^{K-1} (\beta_{k+1} - 2\beta_k + \beta_{k-1})^2 = \beta^T \Omega_2 \beta, \quad (8)$$

with penalty matrix $\Omega_2 = D_2^T D_2$, and

$$D_2 = \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (9)$$

In principle, penalties of higher order are also possible, but we found these two penalties to be especially useful for ordinal predictors because: (1) they incorporate the ordinal scale level of x ; and (2) penalize derivations from a model without predictor x , and a linear model based on the class labels, respectively.

The penalty parameter γ can be chosen based on the data. Popular approaches are (leave-one-out/K-fold/generalized) cross-validation or information criteria, such as AIC or BIC. In the next section we propose to use the duality between penalized likelihood and linear mixed models, which allows the estimation of the penalty parameter, γ , using ML or REML; this also provides a natural framework for testing hypotheses of constancy $\beta_1 = \cdots = \beta_K = \beta_0 = 0$ and linearity.

4. Penalized estimates and linear mixed models

4.1. Linear mixed models

In a linear mixed model with one variance component it is assumed that

$$y = X\xi + Zu + \epsilon, \quad (10)$$

where $y = (y_1, \dots, y_n)$ is the vector of response values, ξ is the p -dimensional vector of fixed effects, u is the q -dimensional vector of random effects, $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ is the error vector, and X and Z are adequate design matrices (see, for example, [Ruppert et al. \(2003\)](#), or [Crainiceanu and Ruppert \(2004\)](#)). For the vectors of random variables u and ϵ , we assume

$$E \begin{pmatrix} u \\ \epsilon \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{cov} \begin{pmatrix} u \\ \epsilon \end{pmatrix} = \begin{pmatrix} \tau^2 G & 0 \\ 0 & \sigma^2 I_n \end{pmatrix},$$

where I_n denotes the $n \times n$ identity matrix (iid errors are assumed). Under the assumption of normal u and ϵ , we have

$$y|u \sim N(X\xi + Zu, \sigma^2 I_n), \quad u \sim N(0, \tau^2 G),$$

and maximizing the likelihood of (y, u) over the unknown ξ and u , leads to the criterion (cf. [Ruppert et al., 2003](#))

$$\sigma^{-2}(y - X\xi - Zu)^T(y - X\xi - Zu) + u^T(\tau^2 G)^{-1}u,$$

or, equivalently,

$$(y - X\xi - Zu)^T(y - X\xi - Zu) + (\sigma^2/\tau^2)u^T G^{-1}u.$$

Minimization of these loss functions leads to estimates

$$\begin{pmatrix} \tilde{\xi} \\ \tilde{u} \end{pmatrix} = (M^T M + (\sigma^2/\tau^2)B)^{-1} M^T y, \quad (11)$$

with $M = (X|Z)$ and

$$B = \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} \end{pmatrix}.$$

4.2. Penalized dummy variables in linear mixed models

Formula (11) shows that the vector $(\tilde{\xi}, \tilde{u})^T$ is a ridge type estimator with penalty parameter (σ^2/τ^2) where only the random effects are penalized. This equivalence between ridge and mixed model estimates offers the opportunity to obtain penalized estimates of dummy coefficients by using mixed models methodology. If penalty (6) is employed, we can write $\xi = \alpha$, $u = \beta$ (with β as defined in (6)), $G^{-1} = \Omega_1$ (Ω_1 is an invertible $K \times K$ matrix), $\gamma = (\sigma^2/\tau^2)$. Design matrices X and Z contain ones and dummy variables m_1, \dots, m_K , respectively.

However, the model can be simplified by using the reparametrization $u = (u_1, \dots, u_K)^T$ with

$$u_k = \beta_k - \beta_{k-1}, \quad k = 1, \dots, K. \quad (12)$$

Then Z contains modified dummy variables z_1, \dots, z_K with

$$z_k = \begin{cases} 1 & \text{if } x \geq k, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

In this case, G^{-1} is just the $K \times K$ identity matrix I_K , since with (12) we have $u = D_1\beta$ (see (7)).

Deriving the penalized estimator with second-order difference penalty (8) using a linear mixed model with identity matrix G^{-1} is only slightly more involved. The linear part of the dummy-coded design matrix has to be separated from the non-linear part. This can be done by writing $X = (1|\chi)$ with $\chi = (x_1, \dots, x_n)^T$ (remember: x_i are observations of $x \in \{0, \dots, K\}$) and $\xi = (\alpha, \delta)$. Now, the random effects u_1, \dots, u_{K-1} have to specify deviations of model (2) from linearity. To achieve this, we write

$$u_k = \beta_{k+1} - 2\beta_k + \beta_{k-1}. \tag{14}$$

Parameterizations $(\beta_1, \dots, \beta_K)^T$ and $(\delta, u_1, \dots, u_{K-1})^T$ are equivalent because

$$\begin{pmatrix} \delta \\ u_1 \\ \vdots \\ u_{K-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix},$$

and

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 2 & 1 & \ddots & & \vdots \\ 3 & 2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ K & \cdots & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} \delta \\ u_1 \\ \vdots \\ u_{K-1} \end{pmatrix}.$$

Thus, the design matrix Z in model (10) needs to contain variables z_1, \dots, z_{K-1} with

$$z_k = \begin{cases} x - k & \text{if } x > k, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

In Appendix A, it is explained more explicitly how design matrix Z and variables z_k are constructed.

4.3. Estimation of variance/penalty parameters

The advantage of using the mixed model formulation of the penalized estimates is that the penalty parameter γ equals σ^2/τ^2 , and likelihood methods for estimating σ^2 and τ^2 can be used for estimating γ . A direct consequence of this is that testing the two hypotheses we are interested in can be reduced to testing certain zero variance components. For estimating σ^2 and τ^2 , ML or restricted maximum likelihood (REML) estimation are typically used. ML estimation is based on the the marginal model

$$y \sim N(X\xi, \Sigma),$$

with covariance matrix $\Sigma = \tau^2 ZGZ^T + \sigma^2 I_n$. The log-likelihood of y under this model is

$$l(\xi, \Sigma) = -\frac{1}{2} \{n \log(2\pi) + \log |\Sigma| + (y - X\xi)^T \Sigma^{-1} (y - X\xi)\}.$$

Maximization over the unknown parameters in ξ and Σ leads to ML estimates. For any fixed Σ , $l(\xi, \Sigma)$ is maximized over ξ by $\tilde{\xi} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$, which equals $\tilde{\xi}$ from (11), cf. Ruppert et al. (2003). If G is just the identity matrix, as it is for our penalized estimates, matrix Σ becomes $\tau^2 Z Z^T + \sigma^2 I_n$, where only τ^2 and σ^2 are unknown. Crainiceanu and Ruppert (2004) use another parametrization: With $\vartheta = \tau^2 / \sigma^2$ and $V_\vartheta = I_n + \vartheta Z Z^T$, we have $\Sigma = \sigma^2 V_\vartheta$. So the log-likelihood becomes

$$l(\xi, \sigma^2, \vartheta) = -\frac{1}{2} \left\{ n \log(\sigma^2) + \log |V_\vartheta| + \frac{(y - X\xi)^T V_\vartheta^{-1} (y - X\xi)}{\sigma^2} + n \log(2\pi) \right\}.$$

A second parameter estimation criterion for our model is the restricted log-likelihood (cf. Harville, 1977; Crainiceanu et al., 2005)

$$l_R(\xi, \sigma^2, \vartheta) = l(\xi, \sigma^2, \vartheta) + \frac{p}{2} \log(\sigma^2) - \frac{1}{2} \log |X^T V_\vartheta^{-1} X|,$$

where p denotes the number of columns of the fixed effects design matrix X . In case of a first-difference penalty on adjacent dummy coefficients we have $p = 1$; if derivations from linearity are penalized, $p = 2$.

5. (Restricted) likelihood ratio tests

When testing for the relevance of the ordinal predictor x , we have $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$. Using the random effects u_1, \dots, u_K defined in (12), this hypothesis is equivalent to

$$H_0 : \tau^2 = 0 \text{ (equivalently, } \vartheta = 0),$$

because the coefficients u_k have mean zero and variance τ^2 . The alternative hypothesis is

$$H_A : \tau^2 > 0 \text{ (equivalently, } \vartheta > 0),$$

which means that dummy coefficients are nonzero.

When testing whether dummy coefficients follow a linear trend over the categories of x , we define the random effects u_1, \dots, u_{K-1} as done in (14), and test again $H_0: \tau^2 = 0$ (equivalently, $\vartheta = 0$) against $H_A: \tau^2 > 0$ (equivalently, $\vartheta > 0$).

Since $\vartheta = \tau^2 / \sigma^2 = \gamma^{-1}$, the null-hypotheses given are equivalent to maximum penalization $\gamma \rightarrow \infty$ when employing penalties (6) and (8), respectively.

The test problems introduced above are nonstandard because under the null hypothesis the parameter of interest (τ^2 , resp. ϑ) is on the boundary of the parameter space. Nevertheless, the (log) likelihood ratio test statistic LRT_n

may be used. Following [Crainiceanu and Ruppert \(2004\)](#) and [Crainiceanu et al. \(2005\)](#), we define

$$\text{LRT}_n = 2 \left\{ \sup_{H_0 \cup H_A} l(\xi, \sigma^2, \vartheta) - \sup_{H_0} l(\xi, \sigma^2, \vartheta) \right\}. \quad (16)$$

Because in our case the fixed effects contained in ξ are the same under the null and the alternative hypotheses, we can also use the restricted likelihood ratio statistic (cf. [Crainiceanu et al., 2005](#))

$$\text{RLRT}_n = 2 \left\{ \sup_{H_0 \cup H_A} l_R(\xi, \sigma^2, \vartheta) - \sup_{H_0} l_R(\xi, \sigma^2, \vartheta) \right\}. \quad (17)$$

[Crainiceanu and Ruppert \(2004\)](#) derived the finite sample and asymptotic distributions of LRT_n and RLRT_n when testing for a zero random effects variance in linear mixed models with one variance component. Because of the mixed model representation of penalized dummy variables, these results can also be used for testing relevance and linearity of ordinal predictors.

With g denoting the number of random effects (that means, $g = K$ or $g = K - 1$, depending on the penalty employed), let $\mu_{s,n}$ and $\nu_{s,n}$ be the g eigenvalues of matrices $Z^T(I_n - X(X^T X)^{-1} X^T)Z$ and $Z^T Z$, respectively. Then, as a special case of Theorem 1 in [Crainiceanu and Ruppert \(2004\)](#),

$$\text{LRT}_n =^{\mathcal{D}} \sup_{\vartheta \geq 0} \left[n \log \left\{ 1 + \frac{N_n(\vartheta)}{D_n(\vartheta)} \right\} - \sum_{s=1}^g \log(1 + \vartheta \nu_{s,n}) \right], \quad (18)$$

where $=^{\mathcal{D}}$ denotes equality in distribution, and

$$N_n(\vartheta) = \sum_{s=1}^g \frac{\vartheta \mu_{s,n}}{1 + \vartheta \mu_{s,n}} w_s^2, \quad D_n(\vartheta) = \sum_{s=1}^g \frac{w_s^2}{1 + \vartheta \mu_{s,n}} + \sum_{s=q+1}^{n-p} w_s^2,$$

and w_s are independent $N(0, 1)$, $s = 1, \dots, n - p$; see also [Crainiceanu et al. \(2003\)](#). For the restricted likelihood ratio statistic (see [Crainiceanu and Ruppert, 2004](#)), we obtain

$$\text{RLRT}_n =^{\mathcal{D}} \sup_{\vartheta \geq 0} \left[(n - p) \log \left\{ 1 + \frac{N_n(\vartheta)}{D_n(\vartheta)} \right\} - \sum_{s=1}^g \log(1 + \vartheta \nu_{s,n}) \right]. \quad (19)$$

[Crainiceanu and Ruppert \(2004\)](#) also give an algorithm for simulating the distributions of LRT_n and RLRT_n , which is implemented in the R package `RLRsim` ([Scheipl, 2010](#)). Since REML estimates of the variance parameters are preferable to ML (cf. [Ruppert et al., 2003](#)) and restricted likelihood based tests have been seen to perform better than those using the unrestricted likelihood ([Morrell, 1998](#)), we will focus on RLRT_n .

6. Simulation studies

To investigate the performance of the proposed restricted likelihood ratio test, and to compare it to the standard F -test, we run several simulation studies. We will first consider the null hypothesis of constancy/relevance, i.e., $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$. Then we will check linearity of dummy coefficients, i.e., $H_0: \beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, k = 1, \dots, K - 1$. All computations are done using the statistical program R (R Development Core Team, 2010). For calculating the proposed restricted likelihood ratio statistic, we use the R package `amer` (Scheipl, 2011), since user defined basis functions (e.g., z_k as defined in (13), resp. (15)) can easily be added. To perform the test, we use `RLRsim` (Scheipl, 2010). A more detailed description of the corresponding R-code is given in Appendix B.

6.1. Testing for relevance

We consider three scenarios with three different regression functions that are defined on the categories $0, \dots, K$ of predictor x , see the left panels in Figures 1, 2 and 3. The response y is generated according to model (2) with standard normal error. We assume a sample size of $n = 100$. The effect strength is controlled by the parameter $a \in [0, 1]$, where $a = 0$ means that y is not influenced by x , i.e., the null hypothesis $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$ is true (green lines). For each regression function considered, the red line corresponds to $a = 1$. The other lines in the left panels of Figures 1, 2 and 3 correspond to $a = 0.1, 0.2, \dots, 0.9$. We consider two nonlinear functions (Figure 1 and 2) and a linear one (Figure 3), and investigate the power of the proposed restricted likelihood ratio test (RLRT) as a function of a for different values of K (middle/right panels of Figures 1, 2 and 3). Note, since $k = 0, \dots, K$, e.g., $K = 4$ means that predictor x has 5 levels. The power of the test is computed as the proportion of rejections of H_0 over 10,000 (independent) repetitions of data generation and testing. The significance level is $\alpha = 0.05$ (dotted lines in Figure 1, 2 and 3).

We compare the proposed RLRT (red lines) to some alternative tests. The standard F -test is carried out by the R function `anova()` from the base distribution (R Development Core Team, 2010). We compare the intercept model (fitted by `lm(y~1)`) and the model with dummy-coded predictor x (fitted by `lm(y~factor(x))`). In addition, we fit a linear model with the group labels of x as predictors (`lm(y~x)`), and compare it to the intercept model. The results of these F -tests are indicated as F_{anova} and F_{lm} , respectively.

Alternatively to linear modeling of x , analysts may also fit a smooth regression model using the `gam()` function from the `mgcv` package (Wood, 2006; Wood, 2011), and check significance of the smooth term (however, this would incorrectly treat x as continuous). For comparison, we also add results for such a test to Figures 1, 2 and 3, when a P-spline basis (Eilers and Marx, 1996) is used and derivations from a constant line over categories of x are penalized. The number of basis functions is chosen as the default value in `mgcv`, which is 9 in our case and seems sufficient since the true regression functions (see Figures 1, 2 and 3) are

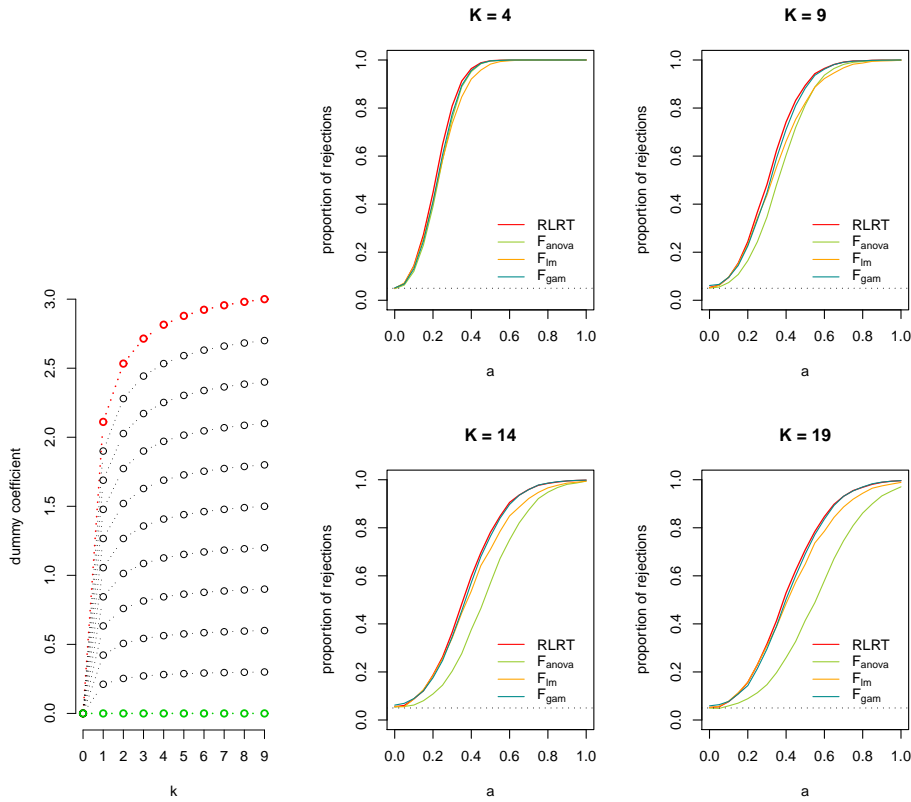


FIG 1. Left: The first set of considered regression functions (for $K = 9$) when testing null hypothesis $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , F_{lm} and F_{gam} are applied.

quite smooth and can hence be well approximated by a rather low-dimensional B-spline basis. The test used is the approximative F -test proposed by Wood (2006), which is part of the default gam-output. So the test is labeled as F_{gam} .

From Figures 1, 2 and 3 it follows that in these cases RLRT is among the most powerful tests. It clearly outperforms F_{anova} , particularly when the number of levels of x is high. F_{lm} is only superior to RLRT if the true underlying regression function is linear (3). F_{gam} does not seem to be as powerful as RLRT for ordinal predictors (particularly in Figures 2 and 3). Furthermore, F_{gam} may not be a reliable α -level test, as, for example, shown by Scheipl et al. (2008). Though in Figures 1, 2 and 3 it seems that F_{gam} has correct size, this is not the case. In our simulations for $a = 0$, the proportion of rejections of H_0 was mostly higher than 0.05 when F_{gam} was applied. More precisely, if $K \in \{9, 14, 19\}$, the (observed) size of F_{gam} was between 0.06 and 0.07 (which can be seen if the respective plots in Figure 1, 2 and 3 are zoomed in). F_{anova} , F_{lm} and RLRT, by contrast, have correct size.

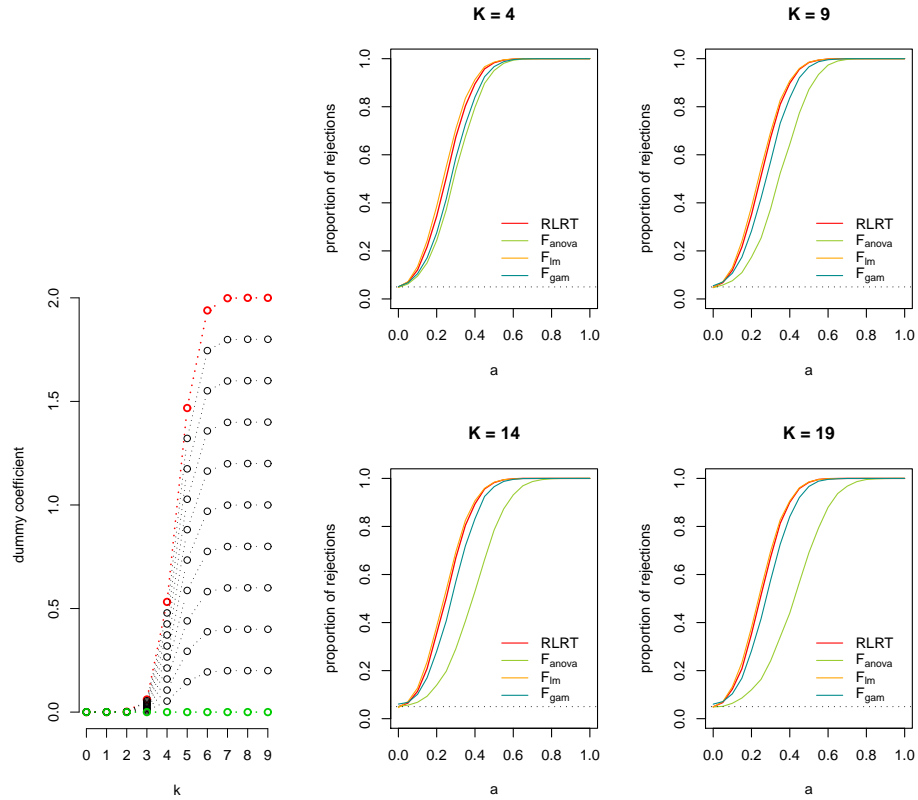


FIG 2. Left: The second set of considered regression functions (for $K = 9$) when testing null hypothesis $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , F_{lm} and F_{gam} are applied.

6.2. Testing for linearity

To investigate if the proposed restricted likelihood ratio test for ordinal predictors is also suitable for testing linearity, we consider three types of nonlinear functions, see the left panels in Figures 4, 5 and 6. Now the extent of non-linearity is controlled by parameter $a \in [0, 1]$, where $a = 0$ means that the conditional expectation of y given predictor x is a linear function of the x -levels $0, \dots, K$. That means, (assuming dummy model (2)) the null hypothesis $H_0: \beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, k = 1, \dots, K - 1$, is true – indicated by green lines in the left panels of Figures 4, 5 and 6. Red lines correspond to $a = 1$ again, and the other lines to $a = 0.1, 0.2, \dots, 0.9$.

We compare the proposed RLRT, the standard F -test F_{anova} (as described in Section 2) and the approximative F -test F_{gam} from *mgcv* (Wood, 2006; Wood, 2011) when x is treated as a continuous covariate. For F_{gam} , we use the procedure

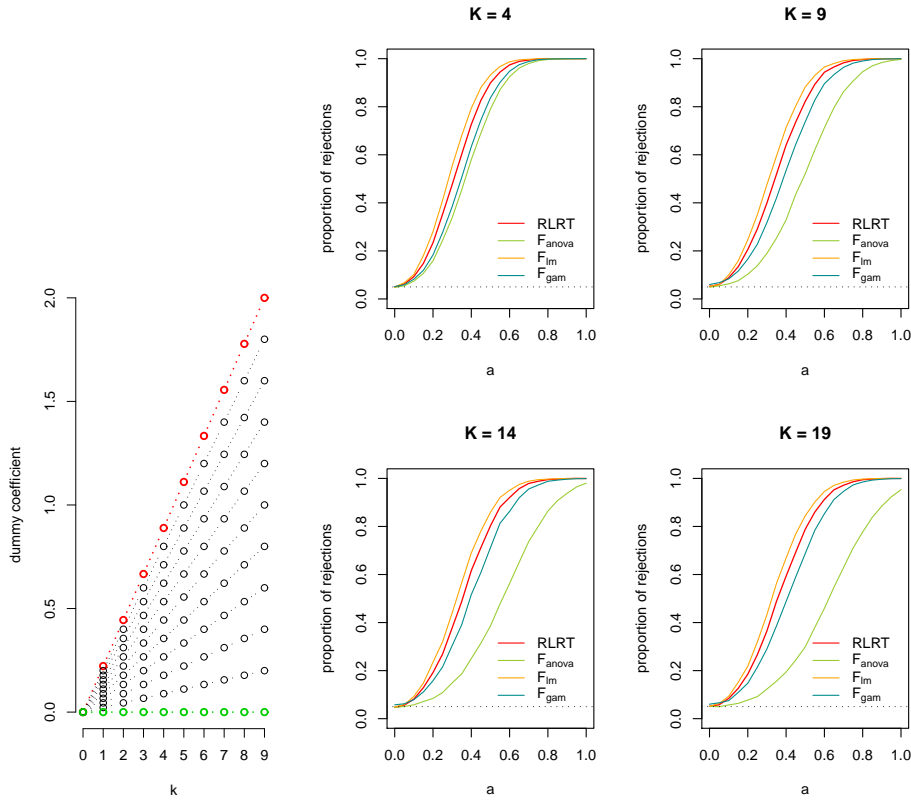


FIG 3. Left: The third set of considered regression functions (for $K = 9$) when testing null hypothesis $H_0: \beta_1 = \dots = \beta_K = \beta_0 = 0$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , F_{lm} and F_{gam} are applied.

suggested by Scheipl et al. (2008): We fit a model in which the smooth term represents only the deviations from linearity, which is done by fitting a model with an explicit linear term in x and an additional P-spline term in x with first-order difference penalization. We also tried the more general approximative F -test for comparing two nested GAMs as proposed by Wood (2006). But in our case this test was far away from being an α -level test (sometimes with rejection rates of around 20%, or even higher, for $a = 0$).

For data generation, we use dummy model (2), again with standard normal error and $n = 100$. As before, we set $\alpha = 0.05$ and give rejection rates (computed over 10,000 simulated data sets) for the different tests as a function of a (right panels of Figures 4, 5 and 6). For a true function as given in Figures 4 and 6, RLRT distinctly outperforms F_{gam} and F_{anova} , with differences between RLRT and F_{anova} becoming larger with increasing K . For a small number of levels K , F_{anova} and F_{gam} perform almost equally. If K is actually small and the true

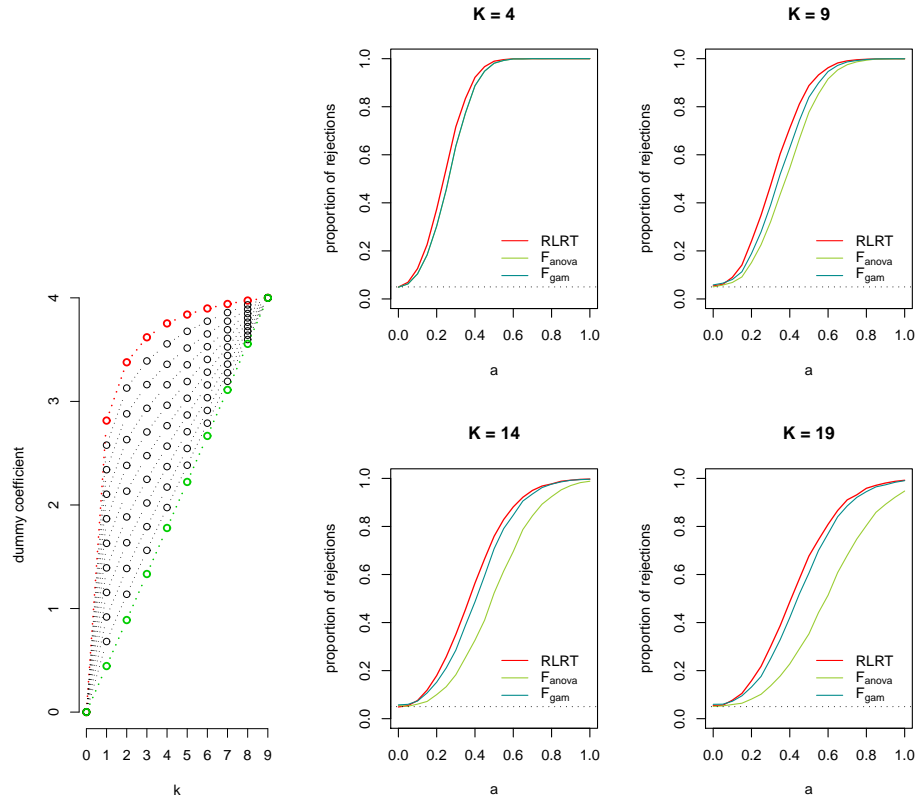


FIG 4. Left: The first set of considered regression functions (for $K = 9$) when testing the null hypothesis of linearity $H_0: \beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, k = 1, \dots, K - 1$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , and F_{gam} are applied.

regression function tends towards a smoothed step function (Figure 5, left), F_{anova} and F_{gam} are superior to RLRT, as seen in Figure 5 (top). But also in this case, such superiority is not found anymore when the number of levels increases.

7. Real data examples

7.1. Rent standard data

We consider data from the Munich rent standard 2003. All larger German cities publish so-called rent standards for having guidelines available to tenants, landlords, renting advisory boards and experts. Rent standards are used, in particular, to determine the local comparative rent, cf. Gertheiss and Tutz (2010).

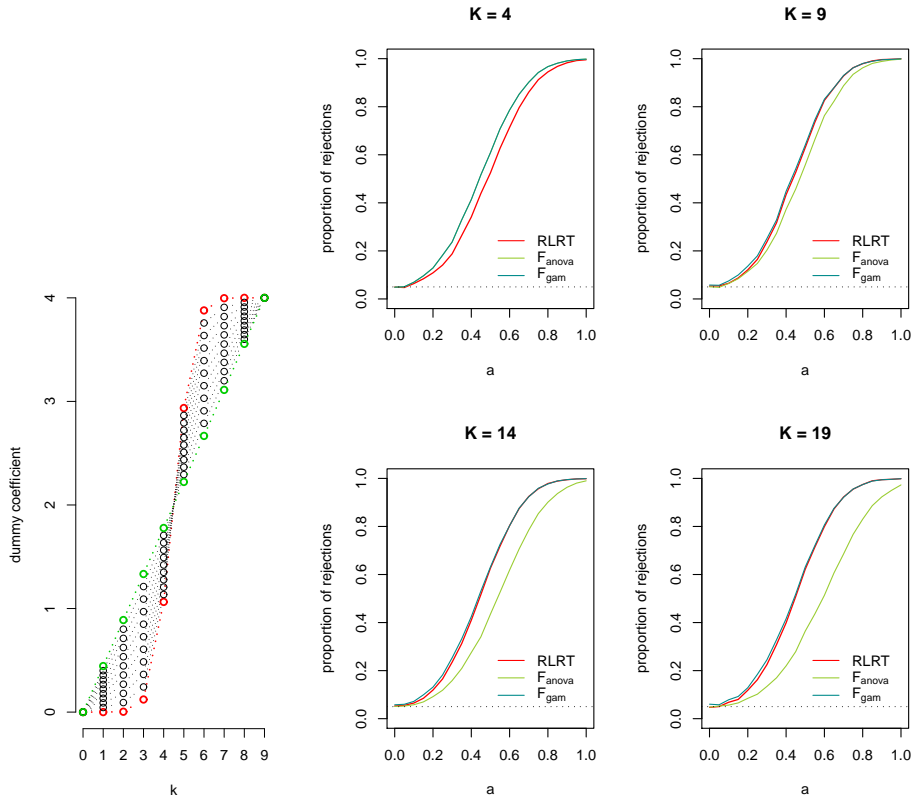


FIG 5. Left: The second set of considered regression functions (for $K = 9$) when testing the null hypothesis of linearity $H_0: \beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, k = 1, \dots, K - 1$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , and F_{gam} are applied.

Some of the data the Munich rent standard is based on can be found in the data archive of the Department of Statistics at the University of Munich.¹ We use the data to investigate the relationship between the rent per square meter (in Euro) and the number of rooms of the corresponding apartment, but we only consider new buildings, with ‘new’ meaning (according to the official rent standard definitions) that the building has been constructed after 1977. In Figure 7 (left) boxplots of the observed rents per square meter are shown for apartments with 1, 2, ..., 6 rooms. It seems that the rent (per square meter) decreases for apartments with more rooms. When looking at the mean rents in each group – that is, fitting a dummy model by pure ML – it seems that rents decrease at first, but then increase for apartments with 5 and 6 rooms (green \circ in Figure 7, right). If

¹<http://www.statistik.lmu.de/service/datenarchiv>

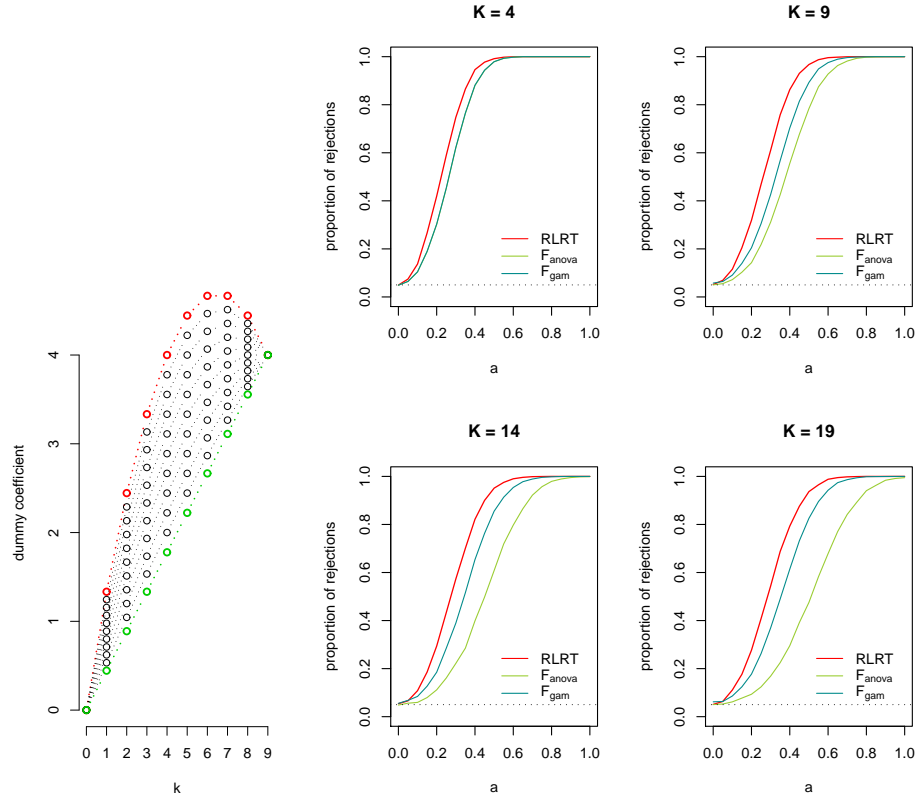


FIG 6. Left: The third set of considered regression functions (for $K = 9$) when testing the null hypothesis of linearity $H_0: \beta_{k+1} - 2\beta_k + \beta_{k-1} = 0, k = 1, \dots, K - 1$, which is true for $a = 0$ (green lines). The red lines correspond to $a = 1$, black lines to $a = 0.1, 0.2, \dots, 0.9$. Right: Proportions of rejections of H_0 as a function of a when RLRT, F_{anova} , and F_{gam} are applied.

a first-order difference penalty is imposed on adjacent dummy coefficients, rents for 4, 5 and 6 rooms seem to be (almost) equal (red * in Figure 7, right). When using a P-spline, fitted rents are very similar (blue \diamond in Figure 7, right).

To check whether there are significant differences between the rents per square meter in apartments with a different number of rooms, we may use classical ANOVA, that is, the standard F -test. Though rents seem to be different in Figure 7, the test indicates just borderline significance on the 0.1-level (p-value 0.0995). Results for the approximative F -test from the `gam`-output are similar (p-value 0.097). The proposed RLRT for ordinal predictors, however, indicates that rents are differing significantly on the 0.05-level (p-value 0.023). Nonlinearity that has been observed in Figure 7 is not indicated as significant by any of the considered tests. A possible explanation is that in the groups which are “responsible” for nonlinearity, namely 5 and 6, there are only a few observations

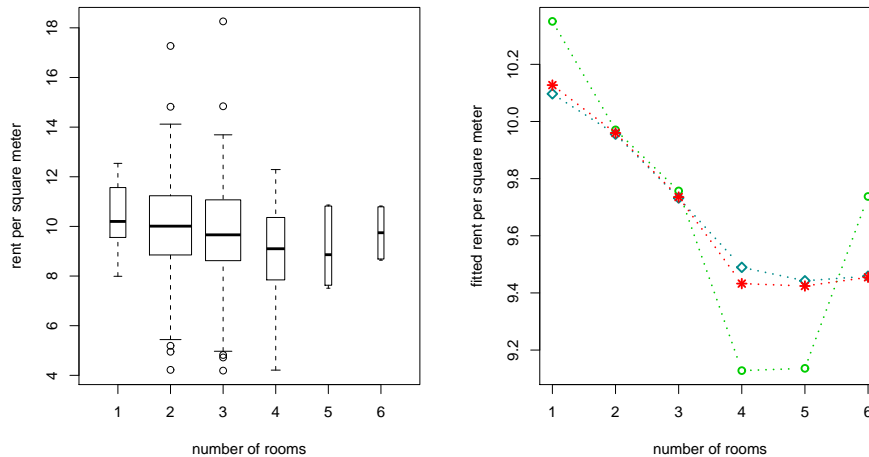


FIG 7. Boxplots of the observed rents per square meter (in Euro) against the number of rooms of the considered apartments (left), and the corresponding fitted mean rents (right). For fitting, we used unpenalized ML estimation of dummy coefficients (green \circ), penalized estimation with first-difference penalty on adjacent dummy coefficients with the penalty parameter estimated by REML (red $*$), and a penalized spline (using the *gam*-function from *mgcv*, blue \diamond).

available (5 and 4, respectively), as seen from Figure 7 (left), where widths of the boxes are proportional to the square-roots of the number of observations in each group.

7.2. ICF data

As another example, we consider patients with chronic widespread pain (CWP) and investigate the bivariate association between a score of physical health and the ICF category ‘individual attitudes of immediate family members’. The ICF – the *International Classification of Functioning, Disability and Health* – provides a unified and standard language and framework for the description of functioning and disability (see WHO, 2001). The ICF consists of about 1400 so-called ICF categories, each of which refers to a health or a health-related domain. Category e410 ‘individual attitudes of immediate family members’ is a so-called environmental factor. For environmental factors a differentiation is made between barriers and facilitators resulting in the coding scheme -4 (complete barrier), \dots , -1 (mild barrier), 0 (no barrier/facilitator), $+1$ (mild facilitator), \dots , $+4$ (complete facilitator). That means, e410 indicates whether, and to what extent, attitudes of family member are rather a barrier or a facilitator with respect to the patients’ daily life. We investigate the relationship between e410 and a physical health component summary measure which is based on the SF-36 questionnaire (Ware and Sherbourne, 1992; McHorney et al., 1993). The data analyzed have been collected in a multicenter, international study (sample

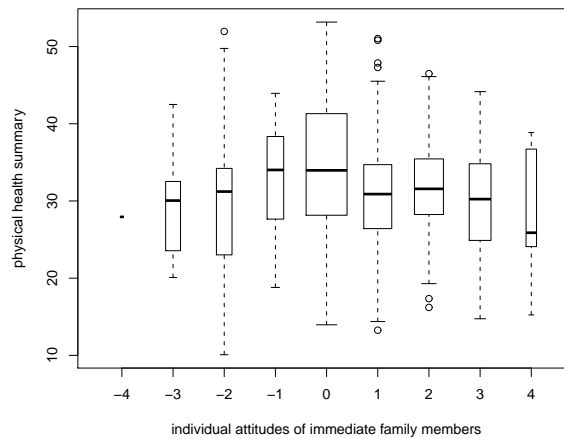


FIG 8. Boxplots of the observed physical health component scores for the different levels of ICF category ‘individual attitudes of immediate family members’ with ordinal coding scheme -4 (complete barrier), \dots , -1 (mild barrier), 0 (no barrier/facilitator), $+1$ (mild facilitator), \dots , $+4$ (complete facilitator).

size $n = 420$), and are part of the R package `ordPens` (Gertheiss, 2011), see Gertheiss et al. (2011) and Cieza et al. (2004) for details.

Figure 8 shows boxplots of the observed physical health component scores for the different levels of the ‘individual attitudes of immediate family members’. For testing whether there are differences between the levels, we use F_{anova} , F_{gam} and the proposed RLRT. All three tests indicate highly significant differences with p-values being given in Table 1. Figure 9 (left) shows the corresponding fitted mean health scores for each level of ICF category e410, where penalizing methods impose a penalty on deviations from constancy. The right panel of Figure 9 shows the analogous plot if deviations from linearity are penalized. The relationship seems highly nonlinear. And indeed, the corresponding tests indicate highly significant deviations from linearity, see Table 1 (right). It is also seen that the p-values of RLRT for both testing constancy and linearity are much smaller than those of F_{anova} and F_{gam} . Though all tests imply that the null-hypotheses of constancy and linearity can be rejected, the smaller p-values may become relevant if more than one ICF category is considered and p-values have to be adjusted.

Concerning parameter estimation, results of different ordinal penalties seem quite unlike (compare red $*$ in the left and right panel of Figure 9), whereas spline fits (blue \diamond) with different penalties are very similar. However, differences between the two types of ordinal penalties are mainly caused by differences with respect to category 4 and, more importantly, -4 . But the number of observations in these extreme categories is very low. In category 4 we have just 10 observations, and only one in -4 (see Figure 8), which makes estimates of respective dummy coefficients very unreliable. This uncertainty at the boundary is also reflected by pointwise confidence intervals as provided by `amer` or `mgcv` for splines

TABLE 1

Results of different tests in terms of p-values when testing relevance and linearity of the ordinal covariate ‘individual attitudes of immediate family members’. The physical health component score shown in Figure 8 is considered as response

	Testing Relevance p-values	Testing Linearity p-values
F_{anova}	$3.0 \cdot 10^{-4}$	$3.4 \cdot 10^{-4}$
F_{gam}	$3.4 \cdot 10^{-4}$	$5.1 \cdot 10^{-4}$
RLRT	$1.8 \cdot 10^{-5}$	$7.7 \cdot 10^{-6}$

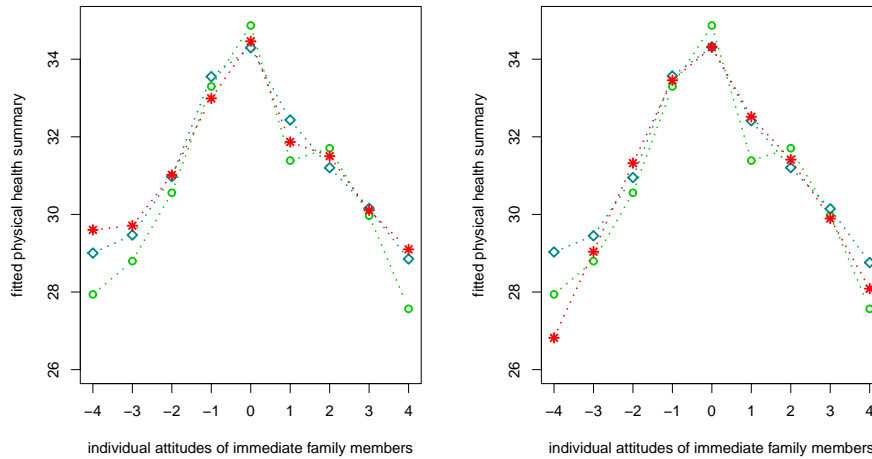


FIG 9. The fitted mean physical health scores for the different levels of e_{410} ‘individual attitudes of immediate family members’. For fitting, we used unpenalized ML estimation of dummy coefficients (green \circ), penalized estimation with first- (left) and second-order (right) difference penalty on adjacent dummy coefficients with the penalty parameter estimated by REML (red $*$), and a penalized spline (using the *gam*-function from *mgcv*, blue \diamond) where deviations from a constant (left) or linearity (right) are penalized.

(not shown). The test proposed here, by contrast, is not for hypotheses with respect to single coefficients, but for global hypotheses of constancy and linearity.

8. Summary and discussion

We proposed a new test for ordinal predictors in the classical linear model. When testing relevance or linearity of an ordinal covariate this can be interpreted as testing for a zero variance component in a linear mixed model. So the proposed test is based on the restricted likelihood ratio test for zero variance components in linear mixed models and uses the exact test distribution derived by Crainiceanu and Ruppert (2004).

We showed that in many situations the new test outperforms the standard F -test which is usually applied for categorical covariates, in particular – but

not only – when the number of levels of the independent variable is high. The proposed test was mostly also superior to (approximative) testing in additive models that may be used if the ordinal predictor is treated as continuous and modeled smoothly using splines. In addition, the latter test may not be a reliable α -level test. So the presented RLRT for ordinal predictors can be seen as a very powerful alternative to existing F -tests when the levels of a categorical covariate are ordered. The advantages of the new test over existing F -tests, which may lead to an increase in power, can be summarized as follows. While the standard F -test for categorical data only uses the predictor’s nominal scale level, the proposed test also takes the covariate’s ordinal nature into account – thus using more information. Differences between the new test and the approximate test based on splines (and treating the ordinal predictor as continuous) are, on the one hand, not such big. Both tests use basis expansions and a difference penalty on basis coefficients. On the other hand, however, there are two important differences: The proposed test is an exact test, whereas the approximate test is not. Furthermore, with dummy-coding instead of, e.g., B-splines, the new test uses basis functions that are somewhat natural for discrete data, as values between predictor levels cannot be observed anyway. When only relevance of the ordinal predictor is to be tested, predictor levels may also be directly included in a linear model, assuming a linear relationship between the predictor level and the response (if there is some dependence). If the relationship is in fact (almost) linear, the corresponding F -test (or the equivalent t -test) is a very powerful test. If the true dependence is, however, clearly nonlinear (as non-monotone, for example), this F -test is often outperformed by the proposed RLR test.

In some situations it is less relevant of course to test whether there is a linear relationship between class labels and the response; for example, when a movie rating scheme is considered with labels “very bad”, “bad”, “okay”, “good”, “very good”. In cases like that, however, it makes no sense either to assume such linear relationship when testing for relevance, whereas dummy-coding of the predictor seems to be adequate. Therefore the proposed RLRT is highly attractive when in such situations relevance of the ordinal covariate is to be tested.

Though in our simulations we only considered one ordinal predictor, the test is directly applicable to multiple regression problems if all additional covariates can be treated as fixed effects that are not be tested simultaneously (see [Crainiceanu et al., 2005](#)). In the case of more than one ordinal predictor that may be treated as a random effect, approximations of the RLRT distribution proposed by [Greven et al. \(2008\)](#) may be used.

For performing the test, existing software can be used and only a little additional implementation is necessary, as given in the Appendix. Instead of R package `amer`, `mgcv` could alternatively be used to fit the penalized regression model. Only the ordinal predictor has to be coded in the right way and the adequate penalty matrices need to be given. So ordinal predictors can be combined with several other types of predictors, including even functional covariates, for example. In addition, extensions to longitudinal or clustered data are possible by including additional random effects.

Acknowledgements

We thank Ciprian Crainiceanu, Sonja Greven, Fabian Scheipl, Gerhard Tutz and two anonymous referees for a number of very helpful comments and suggestions.

Appendix A: Construction of design matrix Z

We consider the linear mixed model (10) and the representation of penalized dummy variables in the linear mixed models framework (see Subsection 4.2).

Penalizing deviations from a constant line

If penalty (6) is employed, we define $\xi = \alpha$ and $u_k = \beta_k - \beta_{k-1}$, $k = 1, \dots, K$, and obtain the equivalent models

$$E[y|x = k] = \alpha + \beta_k \quad \text{and} \quad E[y|x = k] = \xi + u_1 + \dots + u_k,$$

if $k = 1, \dots, K$, and $E[y|x = 0] = \alpha$, resp. $E[y|x = 0] = \xi$, if $k = 0$. So if for observation i the ordinal predictor x has value x_i , the i th row of (random effects) design matrix Z has to be

$$\underbrace{(1, \dots, 1)}_{x_i}, \underbrace{(0, \dots, 0)}_{K-x_i},$$

which leads to the definition of variables z_1, \dots, z_K as given in (13).

Penalizing deviations from linearity

We define random effects u_k as second order differences $u_k = (\beta_{k+1} - \beta_k) - (\beta_k - \beta_{k-1}) = \beta_{k+1} - 2\beta_k + \beta_{k-1}$, $k = 1, \dots, K - 1$. The linear trend over categories (that is not to be penalized) is denoted by δ . Original dummy coefficients β_1, \dots, β_K are obtained from $\delta, u_1, \dots, u_{K-1}$ by the back-transformation (see Subsection 4.2)

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 2 & 1 & \ddots & & \vdots \\ 3 & 2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ K & \dots & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} \delta \\ u_1 \\ \vdots \\ u_{K-1} \end{pmatrix}.$$

So if $x = 0$, the model is still $E[y|x = 0] = \alpha$; if $x \geq 1$, we have equivalent models

$$E[y|x = k] = \alpha + \beta_k, \quad k = 1, \dots, K$$

and

$$\begin{aligned}
 E[y|x = 1] &= \alpha + 1\delta \\
 E[y|x = 2] &= \alpha + 2\delta + 1u_1 \\
 E[y|x = 3] &= \alpha + 3\delta + 2u_1 + 1u_2 \\
 E[y|x = 4] &= \alpha + 4\delta + 3u_1 + 2u_2 + 1u_3 \\
 &\vdots
 \end{aligned}$$

Therefore the i th row of (fixed effects) design matrix X is just $(1, x_i)$, with x_i denoting the value of ordinal predictor x observed at subject i . If $x_i \in \{0, 1\}$, the i th row of (random effects) design matrix Z is just $(0, \dots, 0)$. But if $x_i = 2$, we have $(1, 0, \dots, 0)$; if $x_i = 3$, $(2, 1, 0, \dots, 0)$; if $x_i = 4$, $(3, 2, 1, 0, \dots, 0)$; etc. In general, for $x_i \geq 4$, we have $(x_i - 1, \dots, 2, 1, 0, \dots, 0)$. So if $x = 5$, for example, the new variables z_1, z_2, \dots that are defined to build Z , have values

$$z_1 = 4, z_2 = 3, z_3 = 2, z_4 = 1, z_5 = \dots = z_{K-1} = 0,$$

which leads to definition (15).

Appendix B: R-code for the proposed RLRT for ordinal predictors

This appendix provides some example R code that can be used to carry out the proposed restricted likelihood ratio test for ordinal predictors. At first, we need some packages:

```
> library(amer)
> library(RLRsim)
```

Then we define the basis functions that can be used within `amer` to impose the suggested penalties via a mixed model formulation. Function `cd()` specifies dummy variables such that constancy over categories of the ordinal predictor can be tested, and `ld()` effects second-order difference penalization such that linearity can be tested. Note, functions given below take predictor vector \mathbf{x} with numeric values $1, 2, 3, \dots$

```
> cd <- function(x, by = NULL, allPen = FALSE, diag = FALSE, varying = NULL) {
+   n <- length(x)
+   call <- as.list(expand.call())
+   X <- matrix(0, n, 0)
+   Z <- matrix(rep(x, max(x) - 1), n, max(x) - 1)
+   Z <- Z - matrix(rep(1:(max(x) - 1), n), dim(Z)[1], dim(Z)[2], byrow = T)
+   Z[Z < 0] <- 0
+   Z[Z > 1] <- 1
+   res <- list(X = X, Z = Z)
+   attr(res, "call") <- as.call(call)
+   return(res)
+ }
> ld <- function(x, by = NULL, allPen = FALSE, diag = FALSE, varying = NULL) {
+   n <- length(x)
+   call <- as.list(expand.call())
+   X <- cbind(x - 1)
```

```

+   Z <- matrix(rep(x, max(x) - 2), n, max(x) - 2)
+   Z <- Z - matrix(rep(2:(max(x) - 1), n), dim(Z)[1], dim(Z)[2], byrow = T)
+   Z[Z < 0] <- 0
+   res <- list(X = X, Z = Z)
+   attr(res, "call") <- as.call(call)
+   return(res)
+ }

```

For reproducibility, we set

```
> set.seed(1701)
```

Now we generate a data set of size 100 with ordinal predictor x having levels $1, \dots, 10$:

```
> n <- 100
> x <- c(1:10, sample(1:10, n - 10, replace = T))
```

For given x , response y is normal, and the conditional expectation of y – the true regression function – is nonlinear in x .

```
> y <- 4/9 * (x - 1) - 1/30 * (x - 1) * (x - 10) + rnorm(100)
> dyx <- data.frame(y, x)
```

Now we assume that we would like to test if y is actually influenced by x . Therefore we fit a linear (mixed) model with first-difference penalization using `amer`:

```
> fm1 <- amer(y ~ cd(x), data = dyx, basisGenerators = c("cd"))
```

The proposed test can then be done using `exactRLRT()` from the package `RLRsim`, where for models with only one variance component only the model under the alternative need to be specified (see the help files of `exactRLRT` for details):

```
> RLR1 <- exactRLRT(fm1)
```

The test result is obtained by looking at the corresponding p-value:

```
> RLR1$p.value
```

```
[1] 0
```

To test if the influence of x on y is linear in the class labels of x , we fit a linear (mixed) model with second-order difference penalization, and carry out the test for a zero variance component:

```
> fm2 <- amer(y ~ ld(x), data = dyx, basisGenerators = c("ld"))
```

```
> RLR2 <- exactRLRT(fm2)
```

```
> RLR2$p.value
```

```
[1] 0.0021
```

For comparison, we also carry out the standard F -test:

```
> fm0 <- lm(y ~ x, data = dyx)
```

```
> fm2a <- lm(y ~ factor(x), data = dyx)
```

```
> anova(fm0, fm2a)
```

Analysis of Variance Table

```
Model 1: y ~ x
```

```
Model 2: y ~ factor(x)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	98	101.131				
2	90	83.467	8	17.664	2.3808	0.02247 *

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

- CIEZA, A., G. STUCKI, M. WEIGL, L. KULLMANN, T. STOLL, L. KAMEN, N. KOSTANJSEK, AND N. WALSH (2004). ICF Core Sets for chronic widespread pain. *Journal of Rehabilitation Medicine Suppl.* 44, 63–68.
- CRAINICEANU, C. M. AND D. RUPPERT (2004). Likelihood ratio test in linear mixed models with one variance component. *Journal of the Royal Statistical Society B* 66, 165–185. [MR2035765](#)
- CRAINICEANU, C. M., D. RUPPERT, G. CLAESKENS, AND M. P. WAND (2005). Exact likelihood ratio test for penalised splines. *Biometrika* 92, 91–103. [MR2158612](#)
- CRAINICEANU, C. M., D. RUPPERT, AND T. J. VOGELSANG (2003). Some properties of likelihood ratio tests in linear mixed models. *Technical Report*. (Available from <http://www.orie.cornell.edu/~davidr/papers/>).
- EILERS, P. H. C. AND B. D. MARX (1996). Flexible smoothing with B-splines and Penalties. *Statistical Science* 11, 89–121. [MR1435485](#)
- GERTHEISS, J. (2011). *ordPens: Selection and/or Smoothing of Ordinal Predictors*. R package version 0.1-7.
- GERTHEISS, J., S. HOGGER, C. OBERHAUSER, AND G. TUTZ (2011). Selection of ordinally scaled independent variables with applications to international classification of functioning core sets. *Applied Statistics* 60, 377–395. [MR2767852](#)
- GERTHEISS, J. AND G. TUTZ (2009). Penalized regression with ordinal predictors. *International Statistical Review* 77, 345–365.
- GERTHEISS, J. AND G. TUTZ (2010). Sparse modeling of categorical explanatory variables. *The Annals of Applied Statistics* 4, 2150–2180. [MR2829951](#)
- GREVEN, S., C. CRAINICEANU, H. KÜCHENHOFF, AND A. PETERS (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics* 17, 870–891. [MR2649072](#)
- HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320–338. [MR0451550](#)
- KIM, J.-O. (1975). Multivariate analysis of ordinal variables. *American Journal of Sociology* 81, 261–298.
- KIM, J.-O. (1978). Multivariate analysis of ordinal variables revisited. *American Journal of Sociology* 84, 448–456.
- LABOWITZ, S. (1970). The assignment of numbers to rank order categories. *American Sociological Review* 35, 515–524.
- MAYER, L. S. (1970). Comment on “the assignment of numbers to rank order categories”. *American Sociological Review* 35, 916–917.
- MAYER, L. S. (1971). A note on treating ordinal data as interval data. *American Sociological Review* 36, 519–520.
- MCHORNEY, C. A., J. E. WARE, AND A. E. RACZEK (1993). The MOS 36-item short-form health survey (SF-36): II. psychometric and clinical tests of

- validity in measuring physical and mental health constructs. *Medical Care* 31, 247–263.
- MORRELL, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* 54, 1560–1568.
- R DEVELOPMENT CORE TEAM (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- RAO, C. R., H. TOUTENBURG, SHALABH, AND C. HEUMANN (2008). *Linear Models and Generalizations*. Berlin: Springer. [MR2370506](#)
- RUPPERT, D., M. P. WAND, AND R. J. CARROLL (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press. [MR1998720](#)
- SCHEIPL, F. (2010). *RLRsim: Exact (Restricted) Likelihood Ratio tests for mixed and additive models*. R package version 2.0-5.
- SCHEIPL, F. (2011). *amer: Additive mixed models with lme4*. R package version 0.6.10.
- SCHEIPL, F., S. GREVEN, AND H. KÜCHENHOFF (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis* 52, 3283–3299. [MR2427346](#)
- THOMAS, C. B. (1997). The relationship between values and success for managers in large corporations. *Journal of Social Behavior and Personality* 12, 671–688.
- WARE, J. E. AND C. SHERBOURNE (1992). The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Medical Care* 30, 473–483.
- WHO (2001). *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization.
- WINSHIP, C. AND R. D. MARE (1984). Regression models with ordinal variables. *American Sociological Review* 49, 512–525.
- WOLBERG, W. H. AND O. L. MANGASARIAN (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceeding of the National Academy of Science* 87, 9193–9196.
- WOOD, S. N. (2006). *Generalized Additive Models: An Introduction with R*. London: Chapman & Hall. [MR2206355](#)
- WOOD, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B* 73, 3–36. [MR2797734](#)
- YUAN, M. AND Y. LIN (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society B* 68, 49–67. [MR2212574](#)