Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 50

Direct Simultaneous Inference in Additive Models and its Application to Model Undernutrition

Manuel Wiesenfarth, Tatyana Krivobokova, Stephan Klasen, Stefan Sperlich

First Version December 2010, Revised July 2011 (Title changed)

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Direct Simultaneous Inference in Additive Models and its Application to Model Undernutrition

Manuel Wiesenfarth[1] Tatyana Krivobokova[2] Stephan Klasen[3]

Stefan Sperlich[4]

20th July 2011

Abstract

This article proposes a simple and fast approach to build simultaneous confidence bands and perform specification tests for smooth curves in additive models. The method allows for handling of spatially heterogeneous functions and its derivatives as well as heteroscedasticity in the data. It is applied to study the determinants of chronic undernutrition of Kenyan children, with particular focus on the highly non-linear age pattern in undernutrition. Model estimation using the mixed model representation of penalized splines in combination with simultaneous probability calculations based on the volume-of-tube formula enable the simultaneous inference directly, i.e. without resampling methods. Finite sample properties of simultaneous confidence bands and specification tests are investigated in simulations. To facilitate and enhance its application, the method has been implemented in the R package *AdaptFitOS*.

Key words and phrases. Additive model, Confidence band, Undernutrition, Heteroscedasticity, Locally adaptive smoothing, Kenya, Penalized splines, Varying variance.

[1]CRC PEG, Georg-August-Universität Göttingen, Wilhelm-Weber-Str. 2, 37073 Göttingen, Germany

[2]CRC PEG and Institute for Mathematical Stochastics, Georg-August-Universität-Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany

[3]CRC PEG and Faculty of Economics, Georg-August-Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany

[4]Department of Economics, Université de Genève, 40 Bd du Pont d'Arve, 1211 Genève 4, Switzerland

1 Introduction

In empirical studies one is typically interested not only in estimation of parameters or curves, but also in statistical inference about these estimators. Constructing confidence intervals and performing corresponding specification tests are necessary tools for going beyond the first steps of data exploration. Compared to the finite-dimensional parametric case, inference about a smooth function f, say, in the univariate nonparametric regression context is much more involved. The pointwise confidence bands for $f(x)$ that are usually given do not assess the whole function. Another commonly used confidence band based on Bayesian smoothing splines proposed by Wahba (1983) (see also Nychka, 1988) is only valid in the average coverage sense. That is, the nominal coverage probability results by averaging the coverage probabilities for $f(x)$ at each sample point, so that the confidence band is valid neither at each point nor for the entire curve simultaneously. In general, both pointwise and Wahba (1983)'s confidence bands do not permit statements about the statistical significance of certain features in the underlying curve. Instead, one needs a simultaneous confidence band for f (from some suitable class of functions $\mathcal{F}$, say), which is typically based on its nonparametric estimator $\hat{f}$, and is given by $\left\{\hat{f}(x) - c\sqrt{\mathrm{Var}\{\hat{f}(x)\}},\ \hat{f}(x) + c\sqrt{\mathrm{Var}\{\hat{f}(x)\}},\ \forall x \in \mathcal{X}\right\}$, where c satisfies

$$\alpha = \inf_{f\in\mathcal{F}} P_f\left(\frac{|\hat{f}(x) - f(x)|}{\sqrt{\mathrm{Var}\{\hat{f}(x)\}}} > c, \forall x \in \mathcal{X}\right)$$

on some subspace of the predictor space $\mathcal{X}$ for a given $\alpha \in (0,1)$. Such a confidence band can be used, for example, in tests for functional form specification. Note that c depends crucially on f, which is unknown in practice.

There is an extensive theoretical literature on simultaneous confidence bands for models with a single curve. In a seminal paper, Bickel and Rosenblatt (1973) relate the asymp-

totic distribution of $\sup_{x\in\mathcal{X}} \left|\hat{f}(x) - \mathrm{E}\{\widehat{f}(x)\}\right|$ (that is, ignoring the bias $\mathrm{E}\{\widehat{f}(x)\} - f(x)$ that depends on the unknown f) to the distribution of the supremum of a Gaussian process. However, the convergence of these normal extremes is known to be exceedingly slow with $\log(n)^{-1}$ for sample size n, resulting in very poor performance in small samples. This has led to the development of confidence bands based on bootstrapping techniques in combination with slight undersmoothing, see for example Neumann and Polzehl (1998) and Claeskens and Van Keilegom (2003). In general, such resampling methods are extremely numerically demanding and the data-driven choice of an appropriate smoothing parameter is still an open (and difficult) issue. Hence, in applications with large number of observations and a complicated model structure bootstrapping techniques introduce an unacceptable computational burden.

For our study of undernutrition of children in Kenya we are confronted with a data set of nearly $5,000$ observations. The aim is to investigate the relationship between the so-called Z-score for height for age measuring chronic undernutrition (often called 'stunting') typically used by WHO (see e.g WHO, 1995) and various continuous covariates, modeled additively. Initial explorative analysis has indicated heteroscedasticity in the data and has shown that at least one component of the model needs to be estimated using locally adaptive methods. Such a task is hardly feasible for bootstrap based techniques.

Another approach to building simultaneous confidence bands is to consider the *tail probabilities* of suprema of Gaussian random processes, exploring its connection to the so-called volume-of-tube formula, see Sun (1993), Sun and Loader (1994) and Johansen and Johnstone (1990). As long as f can be estimated without a bias, this method yields very good results for $c \to \infty$ even in small samples, making resampling methods redundant. Recently, Krivobokova et al. (2010) have shown that using the mixed model representation of penalized splines (for a comprehensive overview see Ruppert et al., 2003) for the curve estimation in combination with the approach of Sun (1993) has several advantages

compared to other available techniques.

In this article we extend the flexible adaptive curve estimation of Krivobokova et al. (2008) to account for heteroscedasticity in the data and the approach of Krivobokova et al. (2010) to additive models in order to analyze the data on stunting by age in Kenya. Moreover, a new specification test is introduced. Note that employing penalized splines for estimation allows to avoid backfitting or marginal integration in additive models, and to obtain (adaptive) smoothing parameters from the corresponding (restricted) likelihood simultaneously with the main parameters of interest.

Simultaneous inference in additive models has to date not received much attention in the literature. Härdle et al. (2004) developed simultaneous confidence bands and specification tests for generalized additive models in the kernel regression context, and Wang and Yang (2009) for regression splines based on asymptotic considerations relying on a suboptimal choice of the number of knots. Härdle et al. (2001) proposed locally adaptive (via wavelets) and bandwidth adaptive specification tests for additive models.

The main advantage of the method we propose in this work is that one can obtain simultaneous confidence bands with very good small sample properties for sophisticated models – such as additive models with heterogeneous smooth components and heteroscedastic errors – instantly. Simple and fast calculations allow us also to perform model selection and specification tests without any additional effort. The approach is implemented in the R package *AdaptFitOS*, making it readily available for practitioners.

The paper is organized as follows. In Sections 2 and 3 additive models with penalized splines and the data are introduced. In Section 4 uniform confidence bands are considered, while a new model specification test is proposed in Section 5. The performance of our approach is investigated in Monte Carlo simulations in Section 6. The methods are used then to analyze the determinants of undernutrition of children in Kenya in Section 7 before we conclude in Section 8. Some of the technical details are deferred to the Appendix.

2 Additive models with penalized splines

Let us start with a simple additive model

$$Y_i = \beta_0 + \sum_{j=1}^{d} f_j(x_{ji}) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \;\; i = 1, \ldots, n, \tag{1}$$

where the constant β_0 is an intercept. Without loss of generality we assume non-random covariates to be scaled to the unit interval, i.e. $x_{j1}, \ldots, x_{jn} \in [0, 1]$, $j = 1, \ldots, d$. Each corresponding $f_j \in C^q[0, 1]$ is a q times continuously differentiable function and is centered at zero to ensure identifiability. To estimate f_j with penalized splines, we define for each f_j, $j = 1, \ldots, d$ a set of $k_j < n$ knots $\tau_j = \{0 < \tau_{j,1} < \ldots < \tau_{j,k_j} < 1\}$ and denote the corresponding spline space of degree p as $\mathcal{S}(p; \tau_j)$. This set consists of $p - 1$ times continuously differentiable functions, that are polynomials of degree p on each $[\tau_{j,i}, \tau_{j,i+1})$. Then, the penalized spline estimator is the solution to

$$\min_{s_j(x) \in \mathcal{S}(p;\tau_j),\, j=1,\ldots,d} \left[\sum_{i=1}^{n} \left\{ Y_i - \beta_0 - \sum_{j=1}^{d} s_j(x_i) \right\}^2 + \sum_{j=1}^{d} \lambda_j \int_0^1 \{s_j^{(q)}(x)\}^2 dx \right], \tag{2}$$

for some $q \leq p$. In principle, one can choose different spline degrees for each $\mathcal{S}(p; \tau_j)$ and different penalization orders q for each s_j, but we do not consider this generalization here. To solve (2), represent each $s_j(x)$ as a linear combination of $k_j + p + 1$ spline functions that form basis in $\mathcal{S}(p; \tau_j)$. We use B-splines in our implementation, although others are also certainly possible. Denote a row vector $B_j(x) = \{B_{j,1}(x, \tau_j), \ldots, B_{j,k_j+p+1}(x, \tau_j)\}$ to be some spline basis for $\mathcal{S}(p; \tau_j)$ and let $B_j = \{B_j(x_{j1})^t, \ldots, B_j(x_{jn})^t\}^t$ be the corresponding basis matrix. To obtain centered estimates for f_j, one uses the centered basis matrix $\tilde{B}_j = (\mathrm{I}_n - 1_n 1_n^t) B_j$, with 1_n as an n-dimensional column vector of ones. Now, representing each $s_j(x) = \tilde{B}_j(x)\beta_j$ allows to solve (2) as a minimization problem over β_j.

Smoothing parameters λ_j can be chosen using multivariate versions of cross-validation.

An alternative way to estimate smoothing parameters λ_j is to exploit the link between penalized splines and linear mixed models. Decompose each $\tilde{B}_j\beta_j = \tilde{B}_j(F_b^j b_j + F_u^j u_j) = X_j b_j + Z_j u_j$ in such a way that $(F_u^j)^t F_b^j = (F_b^j)^t D_j F_b^j = 0$ and $(F_u^j)^T D_j F_u^j = I_{\tilde{k}_j}$, where D_j is such that $\int_0^1 [\{\tilde{B}_j(x)\beta_j\}^{(q)}]^2 dx = \beta_j^t D_j \beta_j$ and $\tilde{k}_j = k_j + p + 1 - q$. This decomposition is not unique due to singularity of D_j. In our implementation we followed Durban and Currie (2003). Assuming

$$Y|u_1, \ldots, u_d = \beta_0 + \sum_{j=1}^{d}(X_j b_j + Z_j u_j) + \varepsilon, \quad u_j \sim \mathcal{N}(0, \sigma_{u_j}^2 I_{\tilde{k}_j}), \quad j = 1, \ldots, d, \tag{3}$$

for $Y = (Y_1, \ldots, Y_n)^t$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ leads to the standard linear mixed model with the best linear unbiased predictor being equal to the solution of (2) with $\lambda_j = \sigma^2/\sigma_{u_j}^2$. All mixed model parameters, including $\sigma^2/\sigma_{u_j}^2$, are estimated simultaneously by maximizing a single (restricted) likelihood function. In our further developments we will use the estimators for f_j that result from the mixed model representation of penalized splines (3), so that our estimator will have the form $\hat{f}_j(x) = \ell_j^t(x)Y$, with the smoothing matrix $\ell_j(x)$ given by

$$\ell_j(x) = (I - S_{-j})C_j\{C_j^t(I - S_{-j})C_j + \Lambda_j\}^{-1}C_j^t(x), \tag{4}$$

where model matrix $C_j = [X_j \;\; Z_j]$, penalty matrix $\Lambda_j = \sigma^2/\sigma_{u_j}^2 \mathrm{diag}(0_q, 1_{\tilde{k}_j})$ and $S_{-j} = C_{-j}(C_{-j}^t C_{-j} + \Lambda_{-j})^{-1}C_{-j}^t$ with $C_{-j} = [C_1, C_2, \ldots, C_{j-1}, C_{j+1}, \ldots, C_d]$ and $\Lambda_{-j} = \mathrm{blockdiag}(\Lambda_1, \Lambda_2, \ldots, \Lambda_{j-1}, \Lambda_{j+1}, \ldots, \Lambda_d)$. For practical implementation standard mixed models software can be used (e.g. function `lme` in R).

3 Data on children undernutrition in Kenya

Using the model introduced in the previous section we aim to investigate the data on undernutrition of Kenyan children. Acute and chronic undernutrition is among the most

serious health issues facing developing countries. It is not only an intrinsic indicator of well-being but also associated with morbidity, mortality, reduced labor productivity, etc. Moreover, some estimates claim that undernutrition is implicated in more than 50% of deaths in developing countries (Pelletier, 1994). Given the importance of nutrition for child development, a particular focus is on promoting adequate nutrition for children. Consequently, there is an abundant theoretical and empirical literature on the determinants of childhood undernutrition in developing countries (see Horton et al., 2009). However, most studies are limited to parametric approaches or simple descriptive methods, not accounting for the complex functional forms of the relationships and neglecting the high uncertainty due to the large variability in the data (e.g. Kabubo-Mariara et al., 2009 and Victora et al., 2010).

In this article we analyze the determinants of child undernutrition in Kenya, using the 2003 round of the Kenyan Demographic and Health Survey (KDHS2003, see Central Bureau of Statistics (CBS) Kenya et al., 2004). This includes information on $n = 4,561$ children. The data are cross-sectional, i.e. there are no repeated observations of the same individual. We focus on the Z-score for stuntingdefined as

$$Z_i = \frac{H_i - \text{med}(H)}{\sqrt{\text{Var}(H)}},$$

where H_i is the height of ith individual at a certain age and $\text{med}(H)$ and $\text{Var}(H)$ are the median and variance of the heights in a reference population of well-nourished and healthy children of the same age, respectively. By this normalization, a suitable Gaussian response is obtained and international comparability is aimed for. Note that our analysis is based on the new WHO child growth reference standard which was recently developed based on the assessment of child growth in health populations in six countries across the world. Roughly, as described in WHO (2006), to obtain $\text{med}(H)$ and $\sqrt{\text{Var}(H)}$ a generalized additive model for location, scale and shape (GAMLSS) was applied. Thereby, median

heights and standard deviation were estimated as smooth functions of age using cubic splines with degrees of freedom chosen by (G)AIC. Since children younger than 2 years were measured recumbent and children older than 2 years were measured standing, 0.7 cm were added to all observations of children older than 2 years prior to fitting the model. This estimated difference of 0.7 cm was obtained as the mean differences between measurements of recumbent length and standing height of children between 18 and 30 months from which both measurements are available. Further, some power transformation was applied to age prior to fitting in order to expand the age scale for low age values and compress it for larger age values. This was necessary in order to avoid oversmoothing for low age values where growth is much more rapid than for larger age values. After fitting, 0.7 cm were subtracted from the estimated median curve for all age values larger than 24 months.

Based on the literature on the determinants of chronic undernutrition (e.g. UNICEF, 1998), we start with the following simplified semiparametric model assuming i.i.d. Gaussian errors

$$Z_i = \beta_0 + f_1(age_i) + f_2(bmi_i) + f_3(mheight_i) + z_i'\gamma + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2),\ i = 1, \ldots, n, \quad (5)$$

where $f_1(age)$, $f_2(bmi)$ and $f_3(mheight)$ are smooth functions of the age of the child in months, the Body Mass Index (BMI, defined as weight in kg divided by the squared height in meters) of the mother and the mother's height, respectively. Constant smoothing parameters λ_j are assumed for all functions. Further, as control variables we add a set of covariates z including the numbers of years of education of the mother, the sex of the child as well as the location (rural/urban) and province of the household.

Some of the substantive questions for which a semi-parametric regression approach is particularly suitable concern the age effect. As shown in the literature on undernutrition (e.g. Belitz et al., 2010 and references therein), children in developing countries are usually born with an anthropometric status that is close to the median of the reference population.

Due to poorer nutrition and a poorer health and sanitary environment, many children begin to fall behind, first in weight, and then in growth so that a growth deficit begins to emerge. This is usually intensified in the so-called weaning crisis, which ranges from 4 to 8 months of age, when solid foods and liquids are introduced and the poor quality of these foods and liquids in many poor countries worsens the nutritional status of the child. As children's bodies then partly adapt to poorer nutritional and health environment (largely by becoming more resistant to pathogens, partly by the reduced energy needs for a smaller body, and partly through lower activity levels), stunting usually stabilizes at around age 2, i.e. no further deteriorations vis-a-vis a reference population of healthy children is observed. One of the important questions in the literature concerns the possibility of catch-up growth (see e.g. WHO, 1995), i.e. improvements of the stunting Z-score over time, particularly after age 2. Thus an important empirical question to ask is in which countries and under which contexts such catch-up growth (usually assumed to be possible particularly between age 2 and 3) is observed. This amounts to testing whether the slope of the age effect is significantly above 0 in some interval.

A second substantive question concerns the impact of the mother's nutritional status, typically proxied by her BMI, on child growth. Some studies (see e.g. Kandala et al., 2009) have found an inverse U-shape, where initially the BMI serves to improve the Z-score, but high levels of the BMI could signify poor quality nutrition which then leads to a worse nutritional status for the child. Again the shape of the curve is thus of interest here. Similar arguments can be made for the impact of mother's height on child height which is likely to be related to genetic transmission as well as inter-generational transmission of the economic status. Also here the shape of the curve is hard to guess in advance.

To answer these questions certain specification tests based on simultaneous confidence bands for additive models developed in the subsequent sections need to be employed.

4 Simultaneous confidence bands

4.1 The volume-of-tube formula

Sun and Loader (1994) suggested to build simultaneous confidence bands for a smooth function using the approximation to the *tail probability* of maxima of Gaussian random processes, which turned out to be connected to the volume-of-tube formula. In this case no bootstrap is necessary and the approach yields quite good results in small samples, once a function estimator is unbiased. For completeness we give here some details.

Consider model (1) with $d = 1$. Let $\tilde{f}(x) = \tilde{\ell}(x)^t Y$ be an unbiased estimator of f and assume λ to be known. This implies that $G(x) = \mathrm{Var}\{\tilde{f}(x)\}^{-1/2}\{\tilde{f}(x) - f(x)\} = \tilde{\ell}(x)^t\epsilon/\|\tilde{\ell}(x)\|$ is a zero mean Gaussian process with variance one and

$$\mathrm{Cov}\{G(x_1), G(x_2)\} = \left(\frac{\tilde{\ell}(x_1)}{\|\tilde{\ell}(x_1)\|}\right)^t \left(\frac{\tilde{\ell}(x_2)}{\|\tilde{\ell}(x_2)\|}\right) =: \eta^t(x_1)\eta(x_2),$$

with manifold $\{\eta(x) : x \in [0,1],\ \eta(x) = (\eta_1(x), \ldots, \eta_n(x))\}$. Then, according to Sun and Loader (1994), it holds for $c \to \infty$

$$\alpha = P\left(\sup_{x\in[0,1]} |G(x)| \geq c\right) = \frac{\kappa_0}{\pi}\exp\left(-c^2/2\right) + 2\{1 - \Phi(c)\} + o\left\{\exp(-c^2/2)\right\}, \qquad (6)$$

with $\kappa_0 = \int_0^1 \|\frac{d}{dx}\eta(x)\|dx$ the length of the manifold $\eta(x)$ and $\Phi(\cdot)$ the distribution function of a standard normal distribution. With this, the $100(1-\alpha)\%$ simultaneous confidence band for $f(x),\ x \in [0,1]$ has the form

$$f(x) \in \left[\tilde{f}(x) - c\sqrt{\mathrm{Var}\{\tilde{f}(x)\}},\ \tilde{f}(x) + c\sqrt{\mathrm{Var}\{\tilde{f}(x)\}}\right], \forall x \in [0,1],$$

where c is found by inverting (6). In practice, however, all nonparametric estimators

of f are biased and the smoothing parameter λ is estimated from the data, introducing extra variability. Both problems have been discussed in Krivobokova et al. (2010) who suggested (in the univariate case) to use instead of c a critical value c_m obtained from the mixed model representation of penalized splines (3). Some heuristic arguments and an extensive simulation study confirmed that this approach has very good small sample properties. Also, they showed that the variability due to the estimation of the smoothing parameter σ^2/σ_u^2 is negligible, once a small q is used. Thereby, one has to use enough knots (k proportional to $n^{\nu/(2q+1)}$, $\nu > 1$) to ensure that the approximation bias of the penalized spline estimator is negligible (approximation bias arises due to the fact that a smooth function f is replaced by a spline; it converges to zero with k^{-p-1}). For more details on the bias structure of penalized splines see Claeskens et al. (2009). In the next section we expound how c_m can be obtained for a general additive model.

4.2 Simultaneous confidence bands for additive models

We consider model (3) and assume that sufficiently many knots are taken, so that the approximation bias is small enough and one can replace $f_j(x)$ by $X_j(x)b_j + Z_j(x)u_j =: C_j(x)\theta_j$ directly. To obtain $c_{m,j}$ we consider the marginal distribution of Y, that is

$$Y \sim \mathcal{N}\left(\beta_0 + \sum_{j=1}^{d} X_j b_j,\ \sigma^2 I_n + \sum_{j=1}^{d} \sigma_{u_j}^2 Z_j Z_j^t\right).$$

With respect to this distribution we obtain a zero mean Gaussian process

$$G_{m,j}(x) = \frac{C_j(x)(\hat{\theta}_j - \theta_j)}{\sqrt{C_j(x)\mathrm{Cov}(\hat{\theta}_j - \theta_j)C_j(x)^t}} \sim \mathcal{N}(0,1),$$

where $\mathrm{Cov}(\hat{\theta}_j - \theta_j) = \{C_j^t(I_n - S_{-j})C_j + \Lambda_j\}^{-1}$ and

$$\mathrm{Cov}\{G_{m,j}(x_1), G_{m,j}(x_2)\} = \left(\frac{\ell_{m,j}(x_1)}{\|\ell_{m,j}(x_1)\|}\right)^t \left(\frac{\ell_{m,j}(x_2)}{\|\ell_{m,j}(x_2)\|}\right) =: \eta_{m,j}^t(x_1)\eta_{m,j}(x_2),$$

with $\ell_{m,j}(x) = \{C_j^t(I - S_{-j})C_j + \Lambda_j\}^{-1/2}C_j^t(x)$. Since $G_{m,j}(x)$ is a zero mean Gaussian process, we can apply the volume-of-tube formula to obtain $c_{m,j}$ from

$$P\left(\sup_{x\in[0,1]} |G_{m,j}(x)| \geq c_{m,j}\right) = \frac{\kappa_{m,j}}{\pi}\exp\left(-c_{m,j}^2/2\right) + 2\{1 - \Phi(c_{m,j})\} + o\left\{\exp(-c_{m,j}^2/2)\right\}, \quad (7)$$

with $\kappa_{m,j} = \int_0^1 \|\frac{d}{dx}\eta_{m,j}(x)\|dx$ as the length of the mixed model manifold. Now a confidence band around f_j based on a penalized spline estimator $\hat{f}_j$ is built as

$$\left[\hat{f}_j(x) - c_{m,j}\sqrt{\mathrm{Var}\{\hat{f}_j(x)\}},\ \hat{f}_j(x) + c_{m,j}\sqrt{\mathrm{Var}\{\hat{f}_j(x)\}}\right],$$

where $\mathrm{Var}\{\hat{f}_j(x)\} = \sigma^2\|\ell_j(x)\|^2$ with $\ell_j(x)$ defined in (4). A careful check shows that the proofs in Krivobokova et al. (2010) carry over to our complex case. Hence, this confidence band should have coverage probability close to the nominal level without further corrections. The critical value $c_{m,j}$ is obtained directly from (7) and no bootstrap is necessary. The small sample performance of this band is investigated in Section 6.1.

4.3 Simultaneous bands for additive models with spatially heterogeneous components and heteroscedastic errors

So far we assumed a constant error variance σ^2. This assumption of homoscedasticity may often be violated when the variance changes with some covariate or depends on $E(Y)$. Further, we assumed constant smoothing parameters λ_j, which may be too restrictive for functions that exhibit strong spatial heterogeneity. For example, the mean function

can change rapidly for low covariate values and remains rather constant afterwards. This makes necessary to penalize little in one part of the covariate support and more severely in another, which is referred to as locally adaptive smoothing. To relax these assumptions, we define $u_{js} \sim \mathcal{N}\{0, \sigma^2_{u_j}(\tau_{j,s})\},\ s = 1, \ldots, k_j$ and $\varepsilon_i \sim \mathcal{N}\{0, \sigma^2(\tilde{x}_i)\},\ i = 1, \ldots, n$, where $\tilde{x}$ is one of the covariates or some linear combination of them. Assuming that the variance processes $\sigma^2_{u_j}(\tau_j)$ and $\sigma^2(\tilde{x})$ are smooth functions, we model them with penalized splines and estimate using the link to mixed models. More precisely, we define a hierarchical mixed model

$$\begin{aligned} Y &= \beta_0 + \sum_{j=1}^{d}(X_j b_j + Z_j u_j) + \varepsilon, \quad \varepsilon|v \sim N(0,\ \sigma^2\Sigma_\varepsilon), \quad u_j|w_j \sim N(0,\ \Sigma_{u_j}), \\ \Sigma_\varepsilon &= \mathrm{diag}\{\exp(X_v\gamma + Z_v v)\}, \quad v \sim N(0,\ \sigma_v^2 I_{k_v}), \\ \Sigma_{u_j} &= \mathrm{diag}\{\exp(X_{w_j}\delta_j + Z_{w_j} w_j)\}, \quad w_j \sim N(0,\ \sigma^2_{w_j} I_{k_{w_j}}), \end{aligned} \tag{8}$$

where X_v and Z_v are obtained by decomposing B_j in the same fashion as described in Section 2, but based on a smaller number of knots $k_v \ll k_j$. In contrast, X_{w_j} and Z_{w_j} are obtained by decomposing the basis matrix $B_j = \{B_j(\tau_{j,1}, \tau_{w_j})^t, \ldots, B_j(\tau_{j,k_j}, \tau_{w_j})^t\}^t$. This basis matrix is obtained by treating knots τ_j as observations and choosing as knots τ_{w_j} a smaller subset of τ_j. All parameters of this model can be estimated from the corresponding (restricted) likelihood. A similar idea was suggested in a fully Bayesian framework with $d = 1$ and MCMC techniques by Crainiceanu et al. (2007). To overcome the numerically intensive computations of the latter, Krivobokova et al. (2008) suggested to use the Laplace approximation of the likelihood. They have shown, that the resulting estimator is nearly identical to the Bayesian one, but can be obtained with considerably smaller numerical effort. In Appendix A.1 we extend the method of Krivobokova et al. (2008) to the model with heteroscedastic errors and provide some details on the estimation procedure.

The smoothing matrix for the given penalized spline estimators has now the form

$$\ell_j(x) = \Sigma_\varepsilon^{-1}(I - S_{-j})C_j\{C_j^t\Sigma_\varepsilon^{-1}(I - S_{-j})C_j + \Lambda_j\}^{-1}C_j^t(x) \tag{9}$$

with $\Lambda_j = \sigma^2\,\text{blockdiag}(0_q, \Sigma_{u_j}^{-1})$ and $S_{-j} = C_{-j}(C_{-j}^t\Sigma_\varepsilon^{-1}C_{-j} + \Lambda_{-j})^{-1}C_{-j}^t\Sigma_\varepsilon^{-1}$. Note that $\text{Var}\{\hat{f}_j(x)\} = \sigma^2\ell_j(x)^t\Sigma_\varepsilon\ell_j(x)$. Then, simultaneous confidence bands can be obtained as described in Section 4.2. Since k_{w_j} and k_ν are both typically very small (5 – 10 subknots are usually sufficient), following the arguments of Krivobokova et al. (2010) one can show that the variability due to estimation of Σ_ϵ and Σ_{u_j} is negligible for sufficiently large n and small q. Our simulation study in Section 6.2 confirms this. The approach can also be used for investigating the statistical significance of features like dips and bumps. In order to do so, choose $q \geq 2$ and build the simultaneous confidence band around the estimated first derivative of f_j (see Ruppert et al., 2003, Chapter 6.8).

Thus, using the mixed model representation of penalized splines one can estimate complex additive models with varying smoothing parameters and varying residual variance easily and obtain simultaneous confidence bands for the corresponding functions without additional effort.

5 A new specification test

The constructed simultaneous confidence bands can now be used for testing a parametric regression specification versus a quite general nonparametric alternative modeled by penalized splines. That is, we test the hypotheses

$$H_0:\ f_j(x_j) = f_j^0(x_j) \text{ vs } H_1:\ f_j(x_j) = f_j^0(x_j) + g_j(x_j),\ \forall x_j \in [0,1],$$

with $f_j^0(x_j)$ as a pre-specified polynomial function, whereas $g_j(x_j)$ is an unspecified deviation. The idea is to write $f_j(x_j) = f_j^0(x_j) + Z_j u_j$ and to exploit the orthogonality of $f_j^0(x_j)$ and $Z_j u_j$. Then, the above test is equivalent to testing $H_0 : \ Z_j u_j = 0$. This hypothesis can be checked by constructing a simultaneous confidence band around $g_j(x_j) = Z_j u_j$. Since any spline function of degree q can be decomposed into a $q-1$ degree polynomial and a remainder, we can always choose such ψ_l that $f_j^0(x_j) = \sum_{l=1}^{q-1} b_l x^l$.

Obviously, the test procedure corresponds to checking whether the confidence band for $Z_j u_j$ uniformly encloses the zero line coinciding with the test statistic

$$T_j = \sup_{x \in [0,1]} \left(|Z\hat{u}_j| / \sqrt{\mathrm{Var}_{Y|u}\{Z\hat{u}_j\}} \right).$$

Rejection of H_0 takes place if $T_j > c^*_{m,j}$. The critical value $c^*_{m,j}$ and $\mathrm{Var}_{Y|u}\{Z\hat{u}_j\}$ are obtained by setting $C_{-j} = [X_1, Z_1, \ldots, X_{j-1}, Z_{j-1}, X_j, X_{j+1}, Z_{j+1}, \ldots, X_d, Z_d]$ and $C_j = Z_j$, as well as appropriate adjustment of Λ_j and Λ_{-j} and then proceeding as in the previous sections. Note that p-values can be obtained directly by calculating the tail probabilities for given T_j using the volume-of-tube formula.

By exploiting the decomposition of a spline function, improved power is obtained compared to the test strategy proposed in Claeskens and Van Keilegom (2003), for example. They build their proposed test on the simultaneous confidence band around f_j itself with the hypotheses $H_0 : \ f_j(x_j) = f_j^0(x_j)$ vs $H_1 : \ f_j(x_j) \neq f_j^0(x_j), \ \forall x_j \in [0,1]$, and rely on local polynomials for estimation and bootstrapping to obtain the critical value. Thereby, the data-driven smoothing parameters choice is still an open problem.

Just like for the confidence bands, our test has the advantage of performing well in small samples and of being analytically available, i.e. no bootstrap or Monte Carlo simulation is necessary (as in Härdle et al., 2004, for example). In particular, this test is preferable to F-type tests as used in the R package `mgcv`, which tend to underestimate p-values

when smoothing parameters are estimated. As we will show in Monte Carlo simulations in Section 6.3, the proposed test not only performs competitively compared to restricted likelihood ratio tests (RLRT, see e.g. Crainiceanu et al., 2005), it also allows to incorporate spatially adaptive smoothed curves without any additional effort.

6 Monte Carlo studies

6.1 Simulation 1: Simultaneous confidence bands for additive models

First, we generate data from model (1) for $d = 3$ with homogeneous functions and i.i.d. Gaussian errors. The covariates are taken to be independent and uniformly distributed over $[0, 1]$. The true functions f_j, shown in Figure 1(a) – (c), are simulated according to

$$\begin{aligned} f_1(x) &= \sin^2\{2\pi(x - 0.5)\}, \\ f_2(x) &= \frac{6}{10}\beta_{30,17}(x) + \frac{4}{10}\beta_{3,11}(x), \\ f_{31}(x) &= x(1 - x), \end{aligned}$$

with $\beta_{l,m} = \Gamma(l+m)\{\Gamma(l)\Gamma(m)\}^{-1}x^{l-1}(1-x)^{m-1}$. Functions f_1 and f_2 were also considered in Krivobokova et al. (2010), while f_{31} was used by Claeskens and Van Keilegom (2003). We scaled all three functions such that their standard deviations are all equal to one providing comparable signal-to-noise ratios (SNR).

We consider three different sample sizes (300, 600 and 1000), $k_j = 40,\ j = 1, 2, 3$ knots and $\sigma \in \{0.33,\ 0.5,\ 1.0\}$, corresponding to medium, low and very low SNR, that is $\sqrt{\text{Var}\{f_j(x_j)\}}/\sigma \in \{3,\ 2,\ 1\}$. We used B-spline bases of degree three with penalties on the integrated squared second derivatives ($q = 2$) of the spline functions. Results for $k_j = 80$

knots were very similar and are therefore discarded. Table 1 shows the coverage rates based on a Monte Carlo sample size of 1000 and nominal coverage $100(1-\alpha)\% = 95\%$. All coverage rates are very close to the nominal level of 0.95, except for f_2 in the case of $\sigma = 1.0$ and $n = 300$. In the latter case, the SNR is too low for the given small sample size such that the second peak of function f_2 could not be recovered frequently. This led to coverage rates lower than the nominal level, since the confidence bands were not correctly centered. Note, however, that this setting is very extreme compared to common settings used in simulations to test the performance of other approaches to simultaneous confidence bands (e.g. Claeskens and Van Keilegom, 2003), where usually considerably larger signal-to-noise ratios are used. Compared to these studies, we thus find that our approach works rather well also in quite unfavorable data situations.

6.2 Simulation 2: Additive model with locally adaptive smoothed components and heteroscedasticity

In the second simulation study, function f_{31} of simulation 1 is replaced by function f_{32} shown Figure 1(d) which is defined as

$$f_{32}(x) = \exp\{-400(x-0.6)^2\} + \frac{5}{3}\exp\{-500(x-0.75)^2\} + 2\exp\{-500(x-0.9)^2\}.$$

This function was also considered e.g. in Krivobokova et al. (2008) and exhibits strong heterogeneity. Further, we introduce heteroscedasticity by specifying $\sigma(x_2) = \sigma - 0.2(x_2 - \overline{x}_2)$. We consider either (i) constant smoothing parameters and error variance or (ii) varying error variance $\sigma^2(x_2)$ and adaptive smoothing parameter $\lambda_3(\tau_3)$ for f_{32} ($k_{w_3} = k_v = 5$ knots). All other settings remain the same as in Section 6.1.

Table 1 shows the coverage rates for $100(1-\alpha)\% = 95\%$. Coverage probabilities for function f_1 are very close to the nominal level regardless whether heterogeneities are taken

into account or not except for $\sigma = 1,\ n = 300$ where the apparently worse overall model fit in (i) led to undercoverage. For function f_2 coverage probabilities improve considerably by taking heteroscedasticity into account such that rates of 0.94 or 0.95 are achieved except for the $\sigma = 1,\ n = 300$ case. Note the virtually identical average areas in (i) and (ii), i.e. the improvement is not ascribed to overall wider confidence bands. Locally adaptive estimation of f_{32} leads to a similar improvement and nearly perfect coverage rates were obtained, except for $n = 300$ and the very low SNR. Further, the average sizes of the bands are decreased notably, due to improved estimation of the horizontal part of f_{32}. However, estimation of the wiggly part of function f_{32} regularly failed for the smallest sample size or high noise settings, such that a slight undercoverage occurs in these cases. That is, although the volume-of-tube formula does not require $n \to \infty$, we observe improved coverage probabilities for increasing sample sizes, due to more precise function estimation.

Summarizing, the sample size must be large enough in low signal-to-noise settings such that the functions can properly be recovered, which is, however, a common feature to all approaches to confidence bands. Overall, we found the approach to perform very well even in these relatively complex models and extreme settings.

6.3 Simulation 3: Nonparametric specification test

We now compare the performance of the proposed test with the restricted likelihood ratio test of Crainiceanu et al. (2005). We consider additive models with i. i. d. Gaussian errors

$$\begin{aligned}
Y &= \mu_j(x_1, x_2, x_3) + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2 I),\ j = 1, 2, 3 \text{ with} \\
\mu_1(x_1, x_2, x_3) &= \varphi f_1(x_1) + x_2(1 - x_2) + f_2(x_2) + x_3 + f_{32}(x_3) \\
\mu_2(x_1, x_2, x_3) &= f_1(x_1) + x_2(1 - x_2) + \varphi f_2(x_2) + x_3 + f_{32}(x_3) \\
\mu_3(x_1, x_2, x_3) &= f_1(x_1) + x_2(1 - x_2) + f_2(x_2) + x_3 + \varphi f_{32}(x_3)
\end{aligned}$$

where $\varphi \in [0; 0.6]$ corresponds to the separation distance between the null and the alternative. We test for no effect, second degree polynomial and for linearity of the components $f_1^*(x_1) = \varphi f_1(x_1)$, $f_2^*(x_2) = x_2(1-x_2) + \varphi f_2(x_2)$ and $f_3^*(x_3) = x_3 + \varphi f_{32}(x_3)$, respectively. To do so, B-spline bases with $(p = 1,\ q = 1)$, $(p = 5,\ q = 3)$ and $(p = 3,\ q = 2)$, respectively, are used.

Further, we choose $\sigma = 0.33$, $n = 300$, $k_j = 40$, $j = 1, 2, 3$ and $k_{w_3} = 5$. (Results for $n = 600$ led to the same conclusions and are therefore not reported here.) Three Monte Carlo simulations with 1000 replications each were carried out. Critical values for the RLRT test were computed using the simulation based approximation to the RLRT distribution implemented in the R package `RLRsim` (see Scheipl et al., 2008 which also includes a comprehensive comparisons of RLRT with F-type tests). The power curves of the proposed test and the RLRT test are virtually identical. The rejection rates are given in Figure 2.

7 Studying undernutrition in Kenya

We start by estimating the model (5). For completeness, the parametric estimates are given in Table 2. Figure 3 shows the estimated functions based on B-splines with $p = 5$, $q = 3$ and $k_1 = 40$, $k_2 = k_3 = 30$. In Figure 3(b), the partial residuals seem to exhibit a larger variability for small BMI values than for large BMI values. This could indicate a dependency between the Body Mass Index and the variance of the error term, which we want to explore by modeling the error variance as a smooth function of *bmi*. Further, the bump between the ages of 30 and 50 months in the enlarged plot of $\hat{f}_1(age)$ shown as grey line in Figure 5(a) could be an artefact due to the constant smoothing parameter used. Since the Z-score decreases rapidly in the first 20 months and remains nearly constant afterwards, it seems reasonable to estimate the effect of age with a locally

adaptive smoothing parameter, as discussed in Section 4.3. Note that WHO (2006) also faced this problem of spatial heterogeneity in their derivation of reference standards used to construct the Z-scores. However, instead of locally adaptive smoothing, a rather crude approach was chosen to address the issue (see Section 3).

Naturally, neglecting these heterogeneities in smoothness and error variance could lead to wrong conclusions. Figure 5 shows the results for model (5) supplemented by these two features. The density plot in Figure 4(a) shows that now the residual distribution is reasonably close to the Gaussian distribution. That is, we can consider the distributional assumption for the validity of the given confidence bands to be fulfilled. The estimated smoothing parameter function $\hat{\lambda}_1(\tau_1)$ shown in Figure 4(b) penalizes the roughness of $f_1(age)$ more strongly for larger age values. This ensures that the 'wiggliness' of the age effect between 30 and 50 month disappears. The estimated function of the residual standard deviation in Figure 4(c) indicates a slightly decreasing trend with bmi, however, barely affecting the width of the confidence bands in Figure 5.

The resulting estimated fit of the mother's BMI is positive and statistically significant based on a 5% significance level, since the zero line lies not entirely inside the simultaneous confidence band. However, the effect of bmi is more or less linear and, according to the test proposed in Section 5, does not significantly deviate from the parametric fit (with a p-value of 0.652). That is, the inverted U shape of the effect of the mother's BMI mentioned before is not confirmed for our Kenyan data. Similarly, the estimated function of the mother's height ($mheight$) is virtually linear and does not significantly deviate from the parametric fit (p-value 1). Regarding the age effect, we find a clearly nonlinear relationship and a significant deviation from the parametric linear fit (the p-value of our test is < 0.0001). Note also that the hypothesis of a quadratic age effect would be rejected indicating that the commonly used parametric models quadratic in age are vulnerable to misspecification bias and inference for other variables of interest could be misleading.

The child's nutritional status seems to be more or less constant for the first three or four months of age, which is, however, associated with high uncertainty. Then, as already been suggested by the nutritional literature, there is a virtually linear deterioration until some inflection point at about 20 months of age after which there seems to be some improvement. In order to investigate whether this catching-up is real, i.e. statistically significant, we compute the first derivative of the function of age, which is given in Figure 6. The slope observed after the inception point until approximately 28 months is marginally significant on a 5% level, since the 95% confidence band around the first derivative does not include the zero line for this range. Afterward, the zero line is included, meaning that we cannot reject the null hypothesis of no catching-up for ages larger than 28 months.

Despite the efforts of the WHO to improve the comparability of Z-scores by age by introducing a new reference standard derived from samples of comparable populations of children younger and older than 24 months, there still could be some problems. For example, the estimated derivative could also be picking up the fact that children younger than 2 years were measured recumbent and children older than 2 years were measured standing. To account for this, the reference standard was adjusted assuming a difference of 0.7 cm as described in Section 3. Note that this difference is associated with high uncertainty. Also, using the mean differences between recumbent length and height instead of median could make the estimate sensitive to a very likely skewed distribution. If this estimate is not appropriate or this difference is smaller for the Kenyan children (who were smaller in average than the sample of healthy children from well-to-do families), this could have led to the observed effect which therefore has to be treated with caution.

To see this, we show in Figure 6(b)-(c) what would happen if the difference between children measured lying down and standing was assumed to be only 0.3 cm. Given that the children in Kenya are generally much worse nourished than children in the reference standard, it might well be the case. As shown in the figure, if the difference were only

0.3 cm, the significant effect of catch-up growth would disappear. Similarly, if there is substantial age misreporting around that age group, the reliability of the finding of catch-up growth could be open to question.

8 Discussion

In this paper we constructed simultaneous confidence bands for additive models with varying residual variance and spatially heterogenous smooth components. In doing so, the use of the mixed model representation of penalized splines not only allows for the fast and efficient estimation of such complex models, it also helps to build simultaneous confidence bands with very good small sample properties instantly, that is without using bootstrap or other numerically demanding techniques. Moreover, this technique can be used to construct specification tests for the additive components. Our simulation study confirmed that the resulted coverage probabilities are very close to the nominal level even for small sample sizes and the specification test is competitive to simulation based alternatives. When studying data on undernutrition of children in Kenya the suggested model, the simultaneous confidence bands and corresponding specification tests provided useful insights into drivers of undernutrition of Kenyan children, particularly the highly non-linear age affect. Our analysis indicates a statistically significant improvement of the stunting score between ages of 23 and 28 months. This, however, could also be due to differences in height measurements of children younger/older than 24 months and therefore requires further investigation. For children older than 28 months, no evidence for catch-up growth with respect to the reference population is found. From a model selection point of view, our analysis emphasizes the importance of flexible estimation of the age effect in order to avoid misspecification bias in the fully parametric models that are frequently employed in this context. Note that the data exhibit both heterogeneity in the functional form of some additive components as well as heteroscedasticity.

Possible further extensions are to include random effects and multidimensional components into the additive model, as well as to account for possible serial correlations in the data. It is important to note that the confidence bands considered rely explicitly on the assumption of normality of the data. Even though for symmetric distributions and sufficiently large sample sizes this assumption is less crucial and good results are typically obtained (see Loader and Sun, 1997), some corrections would be needed for highly skewed data. The proposed approach is quite fast and can readily be applied to large data sets despite its nonparametric nature. It is implemented in the R package *AdaptFitOS*.

Acknowledgements.

The authors acknowledge the support of the German Research Foundation (Deutsche Forschungsgemeinschaft) as part of the Institutional Strategy of the University of Göttingen. We would also like to thank Jan Priebe for his help on processing the Kenyan data.

A Appendix

A.1 Estimation of varying residual variance

To keep the exposition as clear as possible we will cover the single covariate case with varying residual variance and constant smoothing parameter. Details on the estimation of a model with varying smoothing parameter and constant error variance are given in Krivobokova et al. (2008). Combination of varying smoothing parameter with varying residual variance, as well as the extension to additive models, is straightforward. Thus, in the following we provide details only on the estimation of the model

$$\begin{aligned} Y|u,v &\sim N(Xb+Zu, \sigma^2\Sigma_\varepsilon),\ u \sim N(0, \sigma_u^2 I_k), \\ \Sigma_\varepsilon &= \operatorname{diag}\{\exp(X_v\gamma + Z_v v)\},\ v \sim N(0, \sigma_v^2 I_{k_v}) \end{aligned} \tag{10}$$

The marginal likelihood of model (10) is given by

$$L(b, \gamma, \sigma_u^2, \sigma^2, \sigma_v^2) = (2\pi)^{-\frac{(n+k_v)}{2}} \sigma_v^{-k_v} \int_{\mathbb{R}^{k_v}} \exp\{-g(v)\} dv \tag{11}$$

where $2g(v) = \log|V| + v^t v/\sigma_v^2 + (Y - Xb)^t V^{-1}(Y - Xb)$, with $V = \sigma^2 \Sigma_\varepsilon + \sigma_u^2 ZZ^t$. Since the integral in (11) is not available analytically, we opt to use the Laplace approximation. This is justified because the approximation error is of order k_v/n (see Severini, 2000) and we assumed $k_v \ll n$. After applying the Laplace approximation, the log-likelihood corresponding to (10) results in

$$-2l(b, \gamma\, \sigma^2, \sigma_v^2, \sigma_u^2) \approx k_v \log \sigma_v^2 + \log|V(\hat{v})| + \log|I_{vv}(\hat{v})| + \frac{\hat{v}^t \hat{v}}{\sigma_v^2} + (Y - Xb)^t V^{-1}(\hat{v})(Y - Xb),$$

with $\hat{v}$ as a solution to $0 = \partial g(v)/\partial v = -Z_v^t \left[\{(ZZ^t)^{-1} Z\hat{b}\}^2 \sigma^2 \Sigma_\varepsilon - \mathrm{diag}(A_1)\right]/2 + v\sigma_v^{-2}$, where A_1 denotes the vector of diagonal elements of matrix $\mathcal{A} = Z(Z^t \sigma^{-2} \Sigma_\varepsilon^{-1} Z + \sigma_u^{-2} I_k)^{-1} Z^t (ZZ^t)^{-1} \sigma_u^{-2}$. The corresponding Fisher information matrix is given by $I_{vv}(v) = \mathrm{E}\,(\partial^2 g(v)/\partial v \partial v^t | v) = Z_v^t \mathrm{diag}(A_2) Z_v/2 + \sigma_v^{-2} I_{k_v}$. Here A_2 is the vector of diagonal elements of matrix $\mathcal{A}^2$. Introducing notations $\omega = (\gamma, v)$, $C_v = [X_v\ Z_v]$ and $D_v = \mathrm{diag}(0, I_{k_v})$, one can obtain estimates $\hat{\gamma}$ and $\hat{v}$ simultaneously from the iterated weighted least squares

$$\hat{\omega} = \frac{1}{2}\left(\frac{1}{2} C_v^t \mathrm{diag}(A_2) C_v + \sigma_v^{-2} D_v\right)^{-1} C_v \mathrm{diag}(A_2)\alpha, \tag{12}$$

with the working vector $\alpha = C_v \omega + \mathrm{diag}(A_2^{-1})\{(ZZ^t)^{-1} Z\hat{b}\}^2 \sigma^2 \Sigma_\varepsilon - \mathrm{diag}(A_1)$. The corresponding variance is estimated as

$$\hat{\sigma}_v^2 = \hat{v}^t \hat{v}/\mathrm{tr}\{Z_v^t \mathrm{diag}(A_2) Z_v I_{vv}^{-1}\}. \tag{13}$$

Thus, the parameters of model (10) can be estimated by iterating between estimation of

$\hat{b}$, $\hat{u}$, $\hat{\sigma}^2$, $\hat{\sigma}_u^2$ for a fixed ω and σ_v^2 using standard linear mixed model software and updating $\hat{\omega}$ and $\hat{\sigma}_v^2$ from (12) and (13). To use the restricted likelihood one has to replace $g(v)$ by $g_r(v) = g(v) + \log|X^t V^{-1} X|/2$.

References

Belitz, C., Hübner, J., Klasen, S., and Lang, S. (2010). Determinants of the socioeconomic and spatial pattern of undernutrition by sex in india: A geoadditive semi-parametric regression approach. In Kneib, T. and Tutz, G., editors, *Statistical Modelling and Regression Structures*, pages 155–179. Physica-Verlag HD.

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, pages 1071–1095.

Central Bureau of Statistics (CBS) Kenya, Ministry of Health (MOH) Kenya, and ORC Macro (2004). *Kenya Demographic and Health Survey 2003*. CBS, MOH, and ORC Macro, Calverton, Maryland.

Claeskens, G., Krivobokova, T., and Opsomer, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics*, 31(6):1852–1884.

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

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16(2):265–288.

Durban, M. and Currie, I. (2003). A note on P-spline additive models with correlated errors. *Computational Statistics*, 18(2):251–262.

Härdle, W., Huet, S., Mammen, E., and Sperlich, S. (2004). Bootstrap inference in

semiparametric generalized additive models. *Econometric Theory*, 20:265–300.

Härdle, W., Sperlich, S., and Spokoiny, V. (2001). Structural tests in additive regression. *Journal of the American Statistical Association*, 96:1333–1347.

Horton, S., Alderman, H., and Rivera, J. (2009). Hunger and malnutrition. In Lomborg, B., editor, *Global Crises, Global Solutions*. Cambridge University Press, Cambridge, 2nd ed.

Johansen, S. and Johnstone, I. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *The Annals of Statistics*, 18(2):652–684.

Kabubo-Mariara, J., Ndenge, G., and Mwabu, D. (2009). Determinants of children's nutritional status in Kenya: Evidence from demographic and health surveys. *Journal of African Economies*, 18(3):363.

Kandala, N., Fahrmeir, L., Klasen, S., and Priebe, J. (2009). Geo-additive models of childhood undernutrition in three sub-Saharan African countries. *Population, Space and Place*, 15(5):461–473.

Krivobokova, T., Crainiceanu, C., and Kauermann, G. (2008). Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics*, 17(1):1–20.

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, 105(490):852–863.

Loader, C. and Sun, J. (1997). Robustness of tube formula based confidence bands. *Journal of Computational and Graphical Statistics*, 6(2):242–250.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9(4):307–333.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83(404):1134–1143.

Pelletier, D. (1994). The relationship between child anthropometry and mortality in

developing countries: implications for policy, programs and future research. *Journal of nutrition*, 124:2047–2081.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge, U.K.

Scheipl, F., Greven, S., and Küchenhoff, H. (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(7):3283–3299.

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *The Annals of Probability*, 21(1):34–71.

Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22(3):1328–1345.

UNICEF (1998). *The state of the world's children 1998*. Oxford University Press, for UNICEF.

Victora, C., de Onis, M., Hallal, P., Blossner, M., and Shrimpton, R. (2010). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics*, 125(3).

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *The Annals of Statistics*, 45(1):133–150.

Wang, J. and Yang, L. (2009). Efficient and fast spline-backfitted kernel smoothing of additive models. *Annals of the Institute of Statistical Mathematics*, 61(3):663–690.

WHO (1995). Physical status: The use and interpretation of anthropometry. Technical Report 854, WHO Technical Report Series, Geneva: WHO.

WHO (2006). WHO child growth standards based on length/height, weight and age. *Acta Paediatrica*, 95(Supplement 450):76–85.

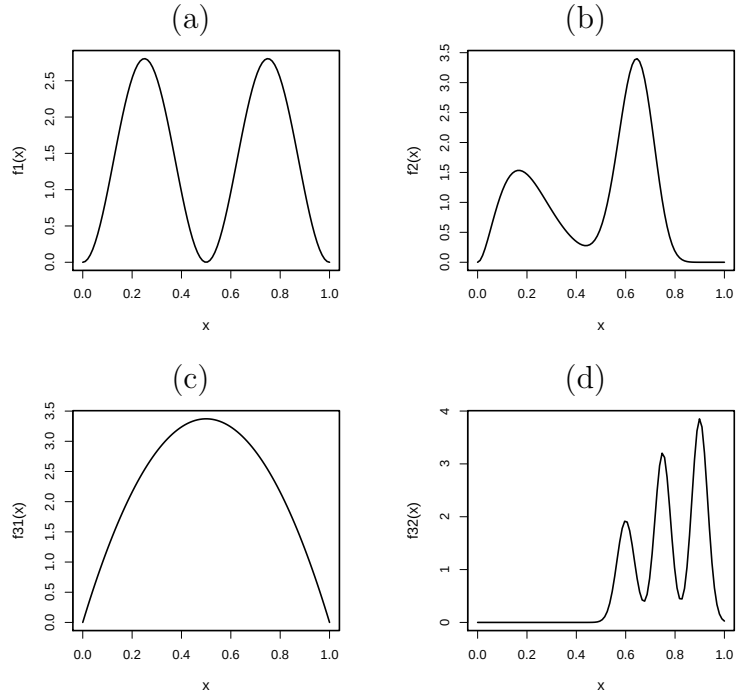

Figure 1: True functions in simulations 1 (top and bottom left) and 2 (top and bottom right) scaled to have variance 1.

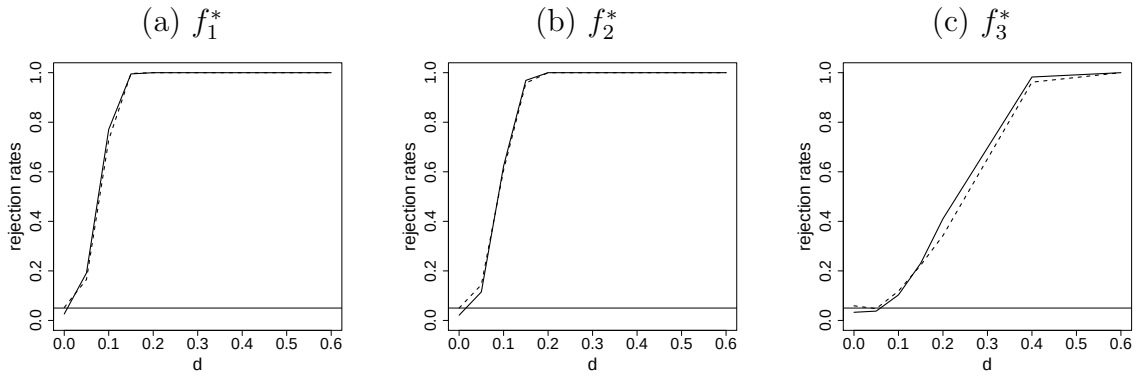

Figure 2: Empirical power curves of the proposed test (solid lines) and RLRT test (dashed lines) in simulation 3.

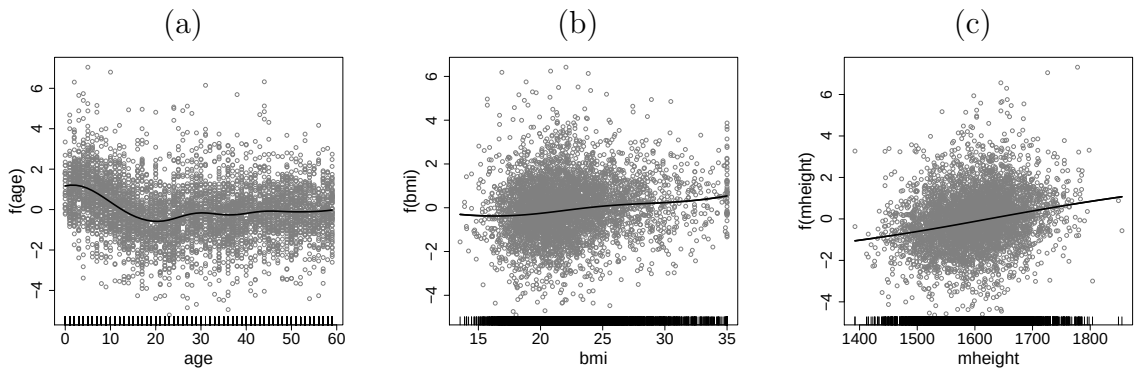

Figure 3: Estimated effects with corresponding partial residuals in model (5).

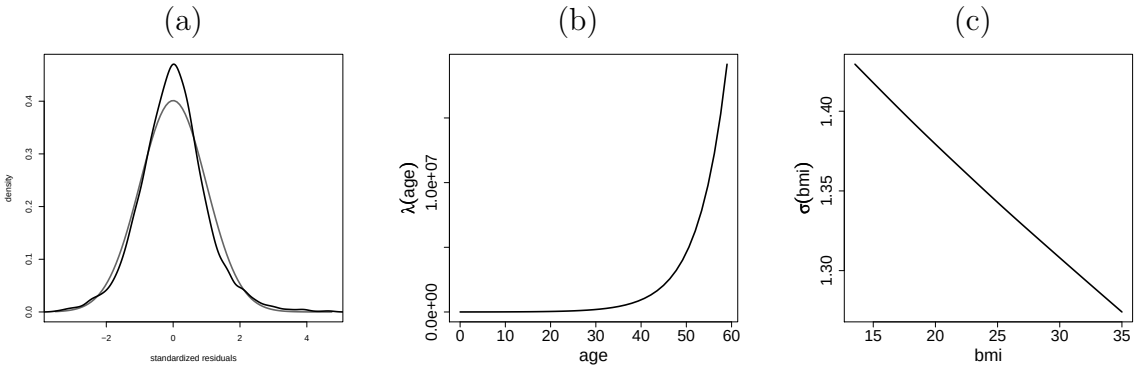

Figure 4: In (a), the gray line corresponds to the standard normal pdf. In (b) and (c), the estimated smoothing parameter function $\hat{\lambda}_1(\tau_1)$ and the estimated residual standard deviation $\hat{\sigma}(bmi)$ based on $k_v = k_{w_1} = 5$ knots are given.

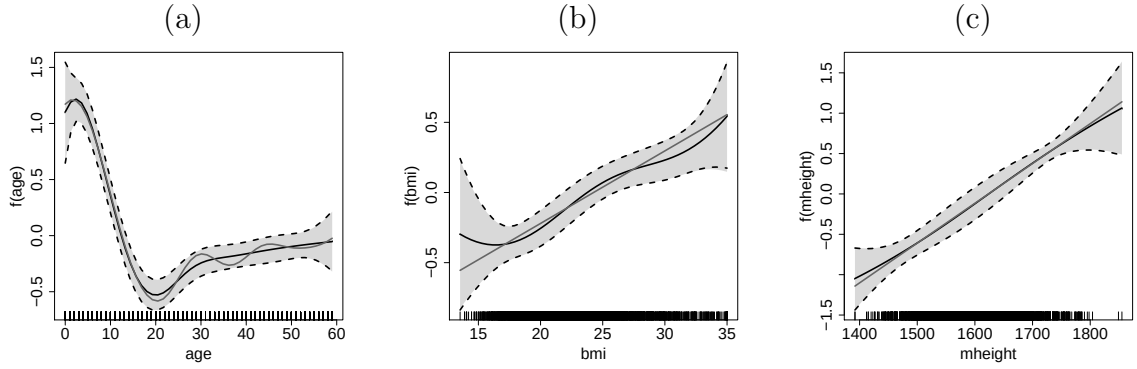

Figure 5: Estimated effects in the final model. As gray lines, in (a) the fit assuming constant smoothing parameter and in (b) and (c) the linear fits are superimposed. 95% simultaneous confidence bands assuming homoscedasticity (light gray area) and heteroscedasticity (dashed lines) are practically identical and can therefore hardly be distinguished.

Table 1: Coverage rates in simulations together with average areas in parenthesis. For simulation 2 results for either constant smoothing parameters and error variance (columns (i)) or varying error variance $\sigma^2(x_2)$ and adaptive smoothing parameter $\lambda_3(\tau_3)$ for f_{32} (columns (ii)) are given.

		Simulation 1			Simulation 2					
σ	n	f_1	f_2	f_{31}	f_1		f_2		f_{32}	
					(i)	(ii)	(i)	(ii)	(i)	(ii)
0.33	300	0.94	0.94	0.95	0.93	0.94	0.89	0.94	0.92	0.93
		(0.45)	(0.49)	(0.31)	(0.46)	(0.44)	(0.51)	(0.51)	(0.69)	(0.51)
	600	0.95	0.94	0.95	0.95	0.95	0.89	0.94	0.93	0.95
		(0.35)	(0.38)	(0.23)	(0.36)	(0.34)	(0.39)	(0.38)	(0.52)	(0.38)
	1000	0.96	0.95	0.96	0.95	0.94	0.88	0.95	0.94	0.95
		(0.28)	(0.31)	(0.19)	(0.29)	(0.27)	(0.32)	(0.31)	(0.42)	(0.29)
0.50	300	0.94	0.93	0.94	0.93	0.95	0.90	0.94	0.90	0.92
		(0.61)	(0.67)	(0.42)	(0.63)	(0.62)	(0.70)	(0.69)	(0.95)	(0.73)
	600	0.94	0.95	0.95	0.95	0.94	0.91	0.95	0.92	0.95
		(0.48)	(0.52)	(0.32)	(0.48)	(0.47)	(0.53)	(0.53)	(0.72)	(0.55)
	1000	0.95	0.95	0.96	0.95	0.95	0.91	0.94	0.93	0.95
		(0.39)	(0.42)	(0.26)	(0.39)	(0.39)	(0.43)	(0.43)	(0.59)	(0.42)
1.00	300	0.93	0.88	0.95	0.91	0.94	0.87	0.90	0.71	0.81
		(1.03)	(1.12)	(0.71)	(1.05)	(1.06)	(1.16)	(1.18)	(1.54)	(1.26)
	600	0.95	0.93	0.96	0.95	0.94	0.92	0.93	0.86	0.92
		(0.8)	(0.87)	(0.54)	(0.81)	(0.81)	(0.89)	(0.89)	(1.22)	(0.97)
	1000	0.94	0.94	0.97	0.95	0.94	0.93	0.94	0.89	0.92
		(0.66)	(0.72)	(0.44)	(0.67)	(0.66)	(0.73)	(0.73)	(1.01)	(0.76)

Table 2: Estimated effects of variables in z. Reference category for regional effects is Nairobi province.

	Estimate	Std. Error	t value	p-value
(Intercept)	-1.380	0.091	-15.090	0.000
yearsofedu	0.041	0.006	6.778	0.000
rural	-0.100	0.060	-1.658	0.097
female	0.197	0.041	4.868	0.000
central	-0.132	0.101	-1.300	0.194
coast	-0.071	0.102	-0.700	0.484
eastern	-0.163	0.106	-1.539	0.124
nyanza	-0.129	0.102	-1.275	0.202
rift valley	-0.151	0.098	-1.553	0.121
western	-0.248	0.099	-2.495	0.013
north eastern	0.547	0.127	4.316	0.000

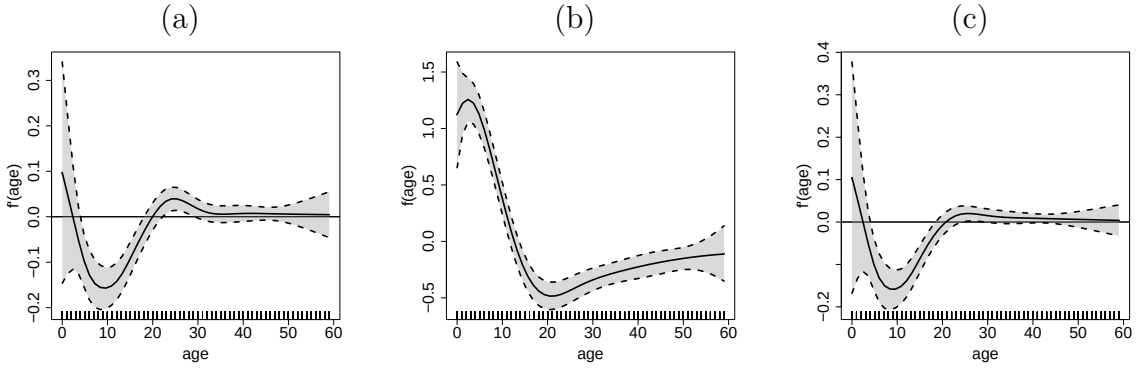

Figure 6: (a): Estimated first derivative of the age effect with 95% simultaneous confidence band. (b) and (c) the estimated age effect and its first derivative assuming that the recumbent length and standing height only differ by 0.3 cm.