Shape testing in varying coefficient models

Peer-reviewed author version

Shape Testing in Varying Coefficient Models

M. Ahkim · I. Gijbels · A. Verhasselt

Abstract We consider varying coefficient models, which are an extension of the classical linear regression models in the sense that the regression coefficients are replaced by functions in certain variables (often time). Varying coefficient models have been popular in longitudinal data and panel data studies, and have been applied in fields such as finance and health sciences. We estimate the coefficient functions by splines. An important question in a varying coefficient model is whether a coefficient function is monotone or convex. We develop consistent testing procedures for monotonicity and convexity. Moreover, we provide procedures to test simultaneously the shapes of certain coefficient functions in a varying coefficient model. The tests use constrained and unconstrained regression splines. The performances of the proposed tests are illustrated on simulated data. We also give a real data application.

M. Ahkim
Department of Mathematics and Computer Science, Universiteit Antwerpen, Belgium

I. Gijbels
Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Belgium

A. Verhasselt Γ
Censtat, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Universiteit Hasselt, Belgium
E-mail: Anneleen.Verhasselt@uhasselt.be

1 Introduction

We consider varying coefficient models (VCMs) to study longitudinal data. VCMs were developed by Hastie and Tibshirani (1993). Such models have been widely applied to many scientific areas: environmental science, ecology, econometrics, epidemiology, etc. VCMs are an extension of the classic linear regression model where the coefficient corresponding to a covariate is assumed to be constant (independent of other variables). This assumption can lead to poor modeling when the data are for example time dependent. Therefore, the modeling strategy ought to be revised to increase flexibility and maintain interpretability (Fan and Wenyang (2008)). The extension consists of allowing the coefficients to depend on other variables. In this paper we consider the model

$$Y(t) = \mathbf{X}(t)^\top \boldsymbol{\beta}(t) + \varepsilon(t) = \beta_0(t) + \sum_{p=0}^{d} X^{(p)}(t)\beta_p(t) + \varepsilon(t), \qquad (1.1)$$

where $Y(t)$ is the response at time $t \in \mathcal{T} = [0,1]$; $\mathbf{X}(t) = (X^{(0)}(t), \ldots, X^{(d)}(t))^\top$ is the covariate vector at time t, with $X^{(0)}(t) \equiv 1$; $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_d(t))^\top$ is the vector of coefficient functions at time t. Note that t can be any variable and the domain $\mathcal{T}$ can be any interval. The stochastic error function $\varepsilon(t)$ has mean zero. We assume longitudinal data to expand the applicability. The j-th measurement of subject i $(t_{ij}, Y_{ij}, \mathbf{X}_{ij})$ for $1 \leq i \leq n$ and $1 \leq j \leq N_i$, is a sample from $(t, Y(t), \mathbf{X}(t))$, where t_{ij} is the observed time, Y_{ij} is the observed response of the ith subject at time t_{ij} and $\mathbf{X}_{ij} = (X_{ij}^{(0)}, \ldots, X_{ij}^{(d)})^\top$ is the corresponding observed covariate vector.

The observed covariates and responses are used for nonparametric estimation of the coefficient functions $\beta_0, \beta_1, \ldots, \beta_d$. This can be achieved by several nonparametric techniques. Local polynomial techniques are discussed in Hoover, Rice, Wu and Yang (1998) and Fan and Zhang (1999), among others. Huang, Wu and Zhou (2002) approximates the coefficient functions by spline functions using B-spline bases, and Antoniadis, Gijbels and Verhasselt (2012) use B-spline approximation in a variable selection context. Note that the optimal choice of the smoothing level for coefficient estimation need not be the optimal choice for the hypothesis testing, for more on this note see Zhang and Mei (2012) (p. 1945-1946).

It can be of interest to derive some conclusions about the shapes of the coefficient functions. In this paper, we develop tests for monotonicity and convexity. For example, a monotonically increasing coefficient of a time independent predictor indicates that the effect of this predicator on the response is

increasing. This can be important in, among other fields, medical sciences. See for example our study of schizophrenia patients (Sect. 8), where the 'Severity of Illness' is modeled by a VCM with covariate the binary variable whether the patient received a drug, with coefficients depending on time (with week as unit of time). The general finding was that the drug improved the health of the patients considerably. Since we are employing a VCM we also looked at the behaviour of the drug coefficient which revealed additional information on the evolution of the drug effect on the patients.

The inference question whether a coefficient function is constant already received a lot of attention in the literature. To the best of our knowledge there is yet no effective testing procedure for monotonicity and convexity in varying coefficient models. In the context of univariate regression, methods of estimation under a monotonicity constraint and testing for monotonicity have been widely discussed, see Bowman, Jones and Gijbels (1998), Ghosal, Sen and van der Vaart (2000),Wang and Meyer (2011) and references therein. In the context of varying coefficient models, not much has been written on this subject. However, Zhang, Mei and Wang (2013) extended the SiZer map approach to varying coefficient models where the local polynomial estimation technique is used, which reveals the statistically significant features of the coefficient functions. The SiZer approach leads to a good explanatory tool, for example for choosing the level of smoothness of each coefficient function.

We develop two testing procedures for monotonicity and convexity (concavity) using the nice properties of B-splines. This framework allows us to establish theoretical support for the testing procedures. The first testing procedure uses quadratic spline approximation and is a generalisation of the testing procedure developed by Wang and Meyer (2011). The second testing procedure uses cubic spline approximation with a very intuitive test statistic, namely the minimum of the estimated coefficient function over a certain grid. Then, the hypothesis is rejected if the test statistic is significantly smaller than zero. It will be noticed that the testing procedures are quite simple. In addition, they can be applied to various testing problem settings. Moreover, we also develop testing procedures for testing simultaneously different coefficient functions. A side result of this work is that we have shown that the first few derivatives of the B-spline estimator are uniformly consistent.

The remaining of the paper is presented as follows. In Sect. 2 we discuss the flexible B-spline estimator (Huang, Wu and Zhou (2004)), which is followed by a section on constrained spline estimation in VCMs, providing explanation on how to impose monotonicity on a spline function. The testing procedures for monotonicity are presented in Sect. 4, followed by the section on testing for convexity (concavity). When there are several covariates in the model, it can be of interest to test simultaneously the shape of different coefficient functions. This is discussed in Sect. 6. We use Matlab[i] to illustrate the performances of

---

[i] The code is available from
`https://www.dropbox.com/sh/dwdhuz7e6j459eu/AAAQ7lkBocy7ACM8bw8eHSkOa?dl=0`

the testing procedures on simulated data in Sect. 7, and to apply it to real data in Sect. 8. Sect. 9 contains a short conclusion.

2 Unconstrained spline estimation in VCMs

The coefficient functions are estimated by spline functions through a B-spline basis; for $p = 0, 1, \ldots, d$, $\beta_p(t) \approx \sum_{l=1}^{m_{p,n}} \alpha_{pl} B_{pl}(t; q_p)$, where $\{B_{pl}(\cdot; q_p) : l = 1, \ldots, K_{p,n} + q_p = m_{p,n}\}$ is the q_pth degree B-spline basis with $K_{p,n} + 1$ equidistant knots $\xi_{p0}, \xi_{p1}, \ldots, \xi_{pK_{p,n}}$ in $[0, 1]$. Throughout we let $K_n = \max_p K_{p,n}$ which is allowed to tend to infinity as n increases. Moreover we write m_p to denote $m_{p,n}$.

The B-spline estimator is obtained by minimizing the following expression with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^\top, \ldots, \boldsymbol{\alpha}_d^\top)^\top$, where $\boldsymbol{\alpha}_p = (\alpha_{p1}, \ldots, \alpha_{pm_p})^\top$ for $p = 0, \ldots, d$,

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(Y_{ij} - \sum_{p=0}^{d} \sum_{l=1}^{m_p} X_{ij}^{(p)} B_{pl}(t_{ij}; q_p) \alpha_{pl} \right)^2, \qquad (2.1)$$

or more compactly written

$$(\mathbf{Y} - \mathbf{U}\boldsymbol{\alpha})^\top \mathbf{W} (\mathbf{Y} - \mathbf{U}\boldsymbol{\alpha}). \qquad (2.2)$$

The definitions of the introduced notation are given in Appendix A. If $(\mathbf{U}^\top \mathbf{W} \mathbf{U})$ is invertible, the minimum of (2.2) is attained at

$$\hat{\boldsymbol{\alpha}} = (\mathbf{U}^\top \mathbf{W} \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{W} \mathbf{Y}. \qquad (2.3)$$

The B-spline estimate of $\boldsymbol{\beta}(t)$ is

$$\hat{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\hat{\boldsymbol{\alpha}} = (\hat{\beta}_0(t), \ldots, \hat{\beta}_d(t))^\top, \quad \text{with } \hat{\beta}_p(t) = \sum_{l=1}^{m_p} \hat{\alpha}_{pl} B_{pl}(t; q_p),$$

where $\mathbf{B}(t)$ is defined in Appendix A. Lemma 1 states that $(\mathbf{U}^\top \mathbf{W} \mathbf{U})$ is invertible with probability tending to one.

Lemma 1 (Huang et al. (2004)) *Suppose Assumptions $1-3$ and Assumption 5 (in Appendix B) hold. The probability that $(\mathbf{U}^\top \mathbf{W} \mathbf{U})$ is invertible tends to 1, when $n \to \infty$.*

Let $\mathcal{G}_q(\varXi)$ denote the space of spline functions of degree q with set of knots $\varXi$. Let $\mathcal{G} = \mathcal{G}_{q_0}(\varXi_0) \times \ldots \times \mathcal{G}_{q_d}(\varXi_d)$, where $\varXi_p$ are set of knots in $[0, 1]$ for $p = 0, \ldots, d$. The sup-norm of a function is denoted by $\| \cdot \|_\infty$. For a $\mathbf{g}^* = (g_0^*, \ldots, g_d^*)^\top \in \mathcal{G}$, we define $\|\boldsymbol{\beta} - \mathbf{g}^*\|_\infty := \max_p \|\beta_p - g_p^*\|_\infty$. Let $\rho_n = \inf_{\mathbf{g}^* \in \mathcal{G}} \|\boldsymbol{\beta} - \mathbf{g}^*\|_\infty$. Throughout we work with equidistant knots and assume that $\lim_{n \to \infty} \rho_n = 0$, i.e., the unknown function $\boldsymbol{\beta}$ can be uniformly approximated by spline functions of certain fixed degrees as the number of subjects n and the number of knots increases. For example $\rho_n = O(K_n^{-2})$ when the coefficients $\beta_0, \ldots, \beta_d$ have bounded second derivatives and $\mathcal{G}$ is the product

space of spline functions of degree 3 such that $\lim_{n\to\infty}\min_{p=0,\ldots,d} K_{p,n} = \infty$ (Schumaker (2007), Theorem 6.27).

We call the estimator $\hat{\boldsymbol{\beta}}(\cdot)$ uniform consistent if $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty = o_P(1)$. Under certain conditions we have the uniform consistency of $\hat{\boldsymbol{\beta}}$ and its derivatives. The uniform consistency for B-spline estimation of the functions themselves is available in the literature (see e.g.Huang et al. (2004)), but the uniform consistency result for the derivative functions was not available up to now. Denote by $B^l([0,1])$ the set of real functions with domain $[0,1]$ who have a bounded l-th derivative.

Theorem 1. *Suppose $\beta_p(\cdot) \in B^{q_p+1}([0,1])$. Set*

$$r_n^2 = \frac{K_n^2}{n^2} \sum_{i=1}^{n} \left(\frac{1}{N_i} \left(1 - \frac{1}{K_n}\right) + \frac{1}{K_n} \right).$$

Then, under Assumptions 1−5,

$$\|\hat{\beta}_p^{(v)} - \beta_p^{(v)}\|_\infty = O_P(K_n^v \rho_n + K_n^{v-q_p-1} + K_n^v r_n),$$

for $v = 0, \ldots, q_p$, where $\beta_p^{(v)}(\cdot)$ is the v-th derivative of $\beta_p(\cdot)$.

The consistency of the testing procedures is based on the approximation power given in Theorem 1, which is proved in Online Resource (see Section 2 therein). It follows immediately from Theorem 1 that $\hat{\boldsymbol{\beta}}(\cdot)$ and its derivatives are also uniform consistent.

Corollary 1. *Suppose $\beta_p(\cdot) \in B^{q_p+1}([0,1])$ for $p = 0, \ldots, d$. Then, under Assumptions 1−5,*

$$\|\hat{\boldsymbol{\beta}}^{(v)} - \boldsymbol{\beta}^{(v)}\|_\infty = O_P(K_n^v \rho_n + K_n^{v-\min_{p=0,\ldots,d} q_p-1} + K_n^v r_n),$$

for $v = 0, \ldots, \min_{p=0,\ldots,d} q_p$.

3 Constrained spline estimation in VCMs

The derivative of a spline function $g(t) = \sum_{j=1}^{m} \gamma_j B_j(t;q)$ having distance $\frac{1}{K}$ between the equidistant knots, is (de Boor (2001), page 116)

$$g'(t) = K \sum_{j=1}^{m-1} \Delta\gamma_{j+1} B_j(t;q-1), \tag{3.1}$$

where $\Delta\gamma_{j+1} = \gamma_{j+1} - \gamma_j$. In the lemma below it is established that when $q = 2$, monotonicity of $g(t)$ in the knots $\xi_0, \cdots, \xi_K$ is equivalent to monotonicity on the whole domain $[\xi_0, \xi_K]$.

Lemma 2 *If $q = 2$, then $g'(t) \geq 0$ for all $t \in [0,1]$ if and only if $g'(\xi_i) \geq 0$ for $i = 0, 1, \ldots, K$.*

Proof. The derivative of a quadratic spline is piecewise linear. Therefore, the derivative is positive if and only if it is positive at the knots. ∎

3.1 Constrained splines

In Sect. 4 we describe our testing procedures for the hypothesis that a certain coefficient function is increasing (decreasing). The estimation of the null distribution is based on a constrained estimation of the relevant function. Below we discuss the constraints on the B-spline coefficients which need to be added to obtain an increasing (decreasing) estimator.

3.1.1 Quadratic splines

The derivative of a quadratic spline function $g(t) = \sum_{j=1}^{m} \gamma_j B_j(t; 2)$ with B-spline basis $B_1(\cdot; 2), \ldots, B_m(\cdot; 2)$ which are based on the equidistant knots $\xi_0, \ldots, \xi_K$ is a linear spline function $g'(t) = K \sum_{j=1}^{m-1} \Delta \gamma_{j+1} B_j(t; 1)$ with B-spline basis $B_1(\cdot; 1), \ldots, B_{m-1}(\cdot; 1)$. Define the matrix $\mathbf{S} \in \mathbb{R}^{(K+1) \times (K+2)}$ which consists of B-spline derivatives at the knots; $\mathbf{S}_{ij} = B'_j(\xi_{i-1}; 2)$. By Lemma 2, the function g is increasing if and only if

$$\mathbf{S}\boldsymbol{\gamma} \geq \mathbf{0} \in \mathbb{R}^{K+1}, \tag{3.2}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^\top$. When the objective is to estimate an increasing function, we can minimize (2.1) under the constraint (3.2) to obtain the constrained estimator $\hat{\boldsymbol{\alpha}}^{cs}$.

3.1.2 Cubic splines

Unlike for quadratic spline estimation where linear constraints at the knots can impose monotonicity, for cubic spline estimation we are required to impose quadratic constraints at the knots. This is described in Section 1 in Online Resource. We obtain the constrained cubic spline estimator $\hat{\boldsymbol{\alpha}}^{cs}$ by minimizing (2.1) under the corresponding constraints given in Online Resource. We use the same notation for the constrained quadratic spline estimator when there is no ambiguity.

3.2 Selection of number of knots

Here, the B-spline estimator is attained by fixing the degree vector and allowing only the knot vector to vary. We resolve to a cross validation method to obtain a desired knot vector $(K_{0,n}, \ldots, K_{d,n})$. We delete subject i with all its repeated measurements from the original data to obtain the training data

with which we determine the B-spline estimator $\hat{\boldsymbol{\alpha}}^{-i}$. This is executed for all the subjects $i = 1, \ldots, n$ so that we can compute the cross validation score

$$CV(K_{0,n}, \ldots, K_{d,n}) = \sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{U}_i \hat{\boldsymbol{\alpha}}^{-i}\|_2^2. \qquad (3.3)$$

The desired $(K_{0,n}, \ldots, K_{d,n})$ is the minimizer of (3.3). This approach is called leave-one-subject-out cross validation (Huang et al. (2004) and references therein). One can also resolve to the v-fold cross validation method. Here we partition the data in equal parts (with respect to the subjects) $P_1, P_2, \ldots, P_v$ where all the information of one subject is contained in only one part. Then a training data is formed by deleting one part P_i with which we determine the B-spline estimator $\hat{\boldsymbol{\alpha}}^{-P_i}$ and compute the cross validation score for the deleted part. The total cross validation score which we seek to minimize is

$$CV_v(K_{0,n}, \ldots, K_{d,n}) = \sum_{i=1}^{v} \|\mathbf{Y}_{P_i} - \mathbf{U}_{P_i} \hat{\boldsymbol{\alpha}}^{-P_i}\|_2^2. \qquad (3.4)$$

3.3 Conditional variance estimation

Let $\mathcal{X} = \{(t_{ij}, \mathbf{X}_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$. Conditioning on $\mathcal{X}$, we obtain by (2.3)

$$\text{Cov}(\hat{\boldsymbol{\alpha}} \mid \mathcal{X}) = (\mathbf{U}^\top \mathbf{W} \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{W} \mathbf{V} \mathbf{W} \mathbf{U} (\mathbf{U}^\top \mathbf{W} \mathbf{U})^{-1}, \qquad (3.5)$$

where the only unknown is $\text{Cov}(\boldsymbol{\varepsilon}) = \mathbf{V} = \text{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\text{Cov}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$ with $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iN_i})^\top$. More explicitly

$$(\mathbf{V}_i)_{jj'} = \text{Cov}(\varepsilon_i(t_{ij}), \varepsilon_i(t_{ij'})), \quad 1 \le j, j' \le N_i. \qquad (3.6)$$

Huang et al. (2004) estimate $\text{Cov}(\varepsilon_i(t_{ij}), \varepsilon_i(t_{ij'}))$ by a tensor product spline on $[0,1] \times [0,1]$ (Chapter 12, Schumaker (2007)), that is

$$\text{Cov}(\varepsilon_i(t), \varepsilon_i(s)) \approx \sum_{k,l=1}^{m_\varepsilon} u_{kl} B_k(t, q_\varepsilon) B_l(s, q_\varepsilon), \quad t, s \in [0,1], \ t \ne s, \qquad (3.7)$$

where we use a fixed set of B-splines $\{B_1(\cdot; q_\varepsilon), B_2(\cdot; q_\varepsilon), \ldots, B_{m_\varepsilon}(\cdot; q_\varepsilon)\}$ with degrees q_ε and equidistant knots in $[0,1]$, and let $u_{kl} = u_{lk}$. We impose the approximation in (3.7) only when $t \ne s$, since the covariance function $\text{Cov}(\varepsilon_i(t), \varepsilon_i(s))$ is not necessarily continuous at $t = s$, that is

$$\lim_{s \to t} \text{Cov}(\varepsilon_i(t), \varepsilon_i(s)) \ne \text{Cov}(\varepsilon_i(t), \varepsilon_i(t)),$$

see Diggle and Verbyla (1998) and Diggle (1988) for example. Moreover, $E(\varepsilon_i(t_{ij})\varepsilon_i(t_{ij'})) = \text{Cov}(\varepsilon(t_{ij}), \varepsilon(t_{ij'}))$, therefore we could estimate the coefficients u_{kl} by finding the minimizer of

$$\sum_{i=1}^{n} \sum_{j=1}^{N_i} \sum_{j'=j+1}^{N_i} \left(\varepsilon_i(t_{ij}), \varepsilon_i(t_{ij'}) - \sum_{k,l=1}^{m_\varepsilon} u_{kl} B_k(t_{ij}; q_\varepsilon) B_l(t_{ij'}; q_\varepsilon) \right)^2 \qquad (3.8)$$

if the error terms $\varepsilon_i(t_{ij})$ were observed. Since they are not observed, we replace them by the residuals $\hat{\varepsilon}_i(t_{ij}) = Y_{ij} - \mathbf{X}_{ij}^\top \hat{\boldsymbol{\beta}}(t_{ij})$ to obtain the minimizer $\hat{u}_{kl}$ $(k, l = 1, \ldots, m_\varepsilon)$.

For the estimation of $\sigma^2(t) = \mathrm{Cov}(\varepsilon(t), \varepsilon(t))$ we use the approximation $\sigma^2(t) \approx \sum_k v_k B_k(t; q_\varepsilon)$. As before, we minimize

$$\sum_{i=1}^n \sum_{j=1}^{N_i} \left(\varepsilon_i^2(t_{ij}) - \sum_{k=1}^{m_\varepsilon} v_k B_k(t_{ij}; q_\varepsilon) \right)^2 \tag{3.9}$$

conditioned by $v_k \geq 0$ to obtain $\hat{v}_k$, $k = 1, \ldots, m_\varepsilon$, and define the variance estimate $\hat{\sigma}^2(t) = \sum_{k=1}^{m_\varepsilon} \hat{v}_k B_k(t; q_\varepsilon)$. Under mild conditions, this yields a consistent estimator for the covariance function (Huang et al. (2004)).

4 Monotonicity tests in VCM

We test whether β_k ($k \in \{0, \ldots, d\}$) is increasing. Hence, the hypothesis

$$H_0 : \beta_k'(t) \geq 0 \text{ for all } t \text{ in } [0, 1], \qquad \text{versus} \qquad \neg H_0. \tag{4.1}$$

When $q_k = 2$ we use the idea of Wang and Meyer (2011) who worked with quadratic splines to test monotonicity in the univariate case.

To test whether $\beta_k(\cdot)$ is decreasing we use the varying coefficient model where we replace $X^{(k)}$ by $-X^{(k)}$. Then, we test whether the corresponding coefficient, which is equal to $-\beta_k(\cdot)$, is increasing.

4.1 Quadratic splines approximation

For quadratic spline functions, the monotonicity constraint is translated into a linear constraint on the B-spline coefficients (see Sect. 3.1.1). Define $\mathbf{C} = (\mathbf{0}_1, \mathbf{S}, \mathbf{0}_3)$ where $\mathbf{0}_1 \in I\!\!R^{(K_{k,n}+1) \times \sum_{j=0}^{k-1} m_j}$, $\mathbf{0}_3 \in I\!\!R^{(K_{k,n}+1) \times \sum_{j=k+1}^{d} m_j}$ are matrices with entries 0 and $\mathbf{S} \in I\!\!R^{(K_{k,n}+1) \times (K_{k,n}+2)}$ is the matrix of derivatives at the knots of B-splines corresponding to the coefficient $\beta_k(\cdot)$: $\mathbf{S}_{ij} = B_{kj}'(\xi_{k(i-1)}; 2)$, see (3.2). Hence, the estimate $\hat{\beta}_k$ is increasing if and only if

$$\mathbf{C}\hat{\boldsymbol{\alpha}} \geq 0.$$

In this paper we consider the simple test statistic $\min(\mathbf{C}\hat{\boldsymbol{\alpha}})$. For other possible ways to construct test statistics see e.g. Silvapulle and Sen (2005). Using the test statistic $\min(\mathbf{C}\hat{\boldsymbol{\alpha}})$ the pseudo algorithm to test hypothesis H_0 is as follows.

1. Determine the unconstrained estimator $\hat{\boldsymbol{\alpha}}$, and calculate $s_{\min}$, the minimum of the slopes at the knots; $s_{\min} = \min(\mathbf{C}\hat{\boldsymbol{\alpha}})$. This is the test statistic.
2. If $s_{\min}$ is non-negative, we do not reject H_0.
3. If $s_{\min} < 0$, determine the distribution of $s_{\min}$ under the null hypothesis and calculate the α percentile (see below) Q_α.
4. If $s_{\min}$ is smaller than the α percentile, then we reject H_0.

Below we discuss two approaches to determine the null distribution of $s_{\min}$.

4.1.1 Bootstrap method

We use a bootstrap method to determine the null distribution of $s_{\min}$. We start with setting residuals

$$\hat{\epsilon}_{ij} = Y_{ij} - \sum_{p=0}^{d} X_{ij}^{(p)} \hat{\beta}_p(t_{ij})$$

where $\hat{\boldsymbol{\beta}}(\cdot)$ is the unconstrained B-spline estimator and let

$$Y_{ij}^{ps} = \sum_{p=0}^{d} X_{ij}^{(p)} \hat{\beta}_p^{cs}(t_{ij}) + \hat{\epsilon}_{ij} \quad \text{for } i = 1, \ldots, n \quad \text{and } j = 1, \ldots, N_i$$

be a set of pseudo responses under the null hypothesis with $\hat{\boldsymbol{\beta}}^{cs}(\cdot) = (\hat{\beta}_0^{cs}(\cdot), \ldots, \hat{\beta}_d^{cs}(\cdot))^\top$ the constrained estimate putting the constraint on β_k. The bootstrap procedure to determine the null distribution of $s_{\min}$ goes as follows.

- Step 1: Resample n subjects (with all its repeated measurements) with replacement from

$$\{(Y_{ij}^{ps}, X_{ij}, t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$$

  to obtain the bootstrap sample $\{(Y_{ij}^{ps*}, X_{ij}^*, t_{ij}^*) : i = 1, \ldots, n, j = 1, \ldots, N_i^*\}$
- Step 2: Repeat the above sampling procedure B times.
- Step 3: Obtain the test statistic $s_{\min}^*$ from each bootstrap sample and derive the empirical distribution based on all $s_{\min}^*$.
- Step 4: Take the α percentile $\hat{Q}_\alpha$ of the empirical distribution in Step 3 and reject the null hypothesis if $s_{\min} < \hat{Q}_\alpha$, else do not reject the null hypothesis.

4.1.2 Multivariate normal method

This approach is useful when we have normal errors and is also considered in Wang and Meyer (2011). Assume normal errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V}). \tag{4.2}$$

We need the function $Pr(s_{\min} \leq r)$, $r \in \mathbb{R}$. Since $E(\mathbf{Y}|\mathcal{X}) \approx \mathbf{U}\boldsymbol{\alpha}$ and $s_{\min} = \min(\mathbf{C}\hat{\boldsymbol{\alpha}})$, we have that $\mathbf{C}\hat{\boldsymbol{\alpha}}$ is, conditioned on $\mathcal{X}$, approximately normal with mean $\mathbf{C}\boldsymbol{\alpha}$ and covariance

$$\boldsymbol{\Sigma} = \mathbf{C}(\mathbf{U}^\top \mathbf{W} \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{W} \mathbf{V} \mathbf{W} \mathbf{U} (\mathbf{U}^\top \mathbf{W} \mathbf{U})^{-1} \mathbf{C}^\top. \tag{4.3}$$

We obtain the expression

$$Pr(s_{\min} \leq r) = 1 - \int \cdots \int_{\{\mathbf{z}|\mathbf{z} - r\mathbf{1} \geq 0\}} \phi(\mathbf{z}; \mathbf{C}\boldsymbol{\alpha}, \boldsymbol{\Sigma}) d\mathbf{z}, \tag{4.4}$$

where $\mathbf{z}, \mathbf{1} = (1, 1, \ldots, 1)^\top \in I\!\!R^{(K_{k,n}+1)\times 1}$, and $\phi(\cdot \; ; \mathbf{C}\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ denotes the multivariate normal density function with mean $\mathbf{C}\boldsymbol{\alpha}$ and covariance $\boldsymbol{\Sigma}$. We can compute (4.4) only if $\boldsymbol{\alpha}$ and $\mathbf{V}$ are known. It is clear that we can only approximate $\boldsymbol{\alpha}$ and $\mathbf{V}$. The suggestion is to use $\hat{\boldsymbol{\alpha}}^{cs}$, the constrained estimator, for approximating $\boldsymbol{\alpha}$, and the unconstrained $\hat{\boldsymbol{\alpha}}$ for estimating $\mathbf{V}$ as in Sect. 3.3; if we would use $\hat{\boldsymbol{\alpha}}^{cs}$ instead, it would lead to a biased estimate for $\mathbf{V}$ under the alternative hypothesis. Therefore, we use

$$P_{\mathbf{C}\hat{\boldsymbol{\alpha}}^{cs}, \hat{\boldsymbol{\Sigma}}}(r) = 1 - \int \cdots \int_{\{\mathbf{z}|\mathbf{z}-r\mathbf{1}\geq 0\}} \phi(\mathbf{z}; \mathbf{C}\hat{\boldsymbol{\alpha}}^{cs}, \hat{\boldsymbol{\Sigma}})d\mathbf{z}, \qquad (4.5)$$

where

$$\hat{\boldsymbol{\Sigma}} = \mathbf{C}(\mathbf{U}^\top\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{W}\hat{\mathbf{V}}\mathbf{W}\mathbf{U}(\mathbf{U}^\top\mathbf{W}\mathbf{U})^{-1}\mathbf{C}^\top. \qquad (4.6)$$

The estimated α percentile is determined as follows

$$\hat{Q}_\alpha = \inf\{r \mid P_{\mathbf{C}\hat{\boldsymbol{\alpha}}^{cs}, \hat{\boldsymbol{\Sigma}}}(r) \geq \alpha\}. \qquad (4.7)$$

4.1.3 Consistency of the test for quadratic splines

The following theorem states that the type II error tends to 0 when the coefficient function function is strictly increasing.

Theorem 2. *Assume that $K_n\rho_n + K_n^{q_k} + K_n r_n = o(1)$. Under Assumptions 1–5, if $\inf_{t\in[0,1]} \beta_k'(t) = \delta > 0$, then $P(s_{min} < \min(0, \hat{Q}_\alpha)) = o(1)$.*

The proof of this theorem can be found in Section 3 in Online Resource.

4.2 Cubic splines approximation

For cubic splines we look at the minimum of the derivative of

$$\hat{\beta}_k'(t) = K_{k,n} \sum_{j=1}^{m_k-1} \Delta\hat{\alpha}_{k(j+1)} B_j(t; q_k - 1)$$

for $t \in [0, 1]$, see equation (3.1), where $q_k = 3$. The degrees for other coefficient functions can be arbitrary. Hypothesis H_0 holds if and only if $\beta_k'(\cdot)$ is nonnegative on its domain. In practice we work with a grid G of $[0, 1]$, say $G = \{0, 0.001, 0.002, \cdots, 1\}$. Then we determine the minimum of $\hat{\beta}_k(\cdot)$ over the grid G which will be the test statistic, i.e. the test statistic is $\hat{\beta}_k(c)$ for gridpoint c. When the test statistic is nonnegative we do not reject H_0. In the other case we want to measure how plausible the negative test statistic is. Therefore we look at the α percentile of the null distribution of $\hat{\beta}_k(c)$. The pseudo algorithm for this approach is as follows.

1. Compute $s = \min_{g\in G} \hat{\beta}_k'(g)$
2. If $s \geq 0$, do not reject H_0.

3. If $s < 0$, choose a $c \in G$ such that $s = \hat{\beta}'_k(c) < 0$. Determine the null distribution of s (see below) and check whether s is smaller than the α percentile $\tilde{Q}_\alpha$.

4. If $s < \tilde{Q}_\alpha$, reject H_0, else do not reject H_0.

We are left with determining $\tilde{Q}_\alpha$ in Step 3, hence it suffices to find the null distribution of s.

Note that we could have used the same approach when working with quadratic splines. However, in that case we know that the minimum over the grid G of the linear spline function $\hat{\beta}'_k(\cdot)$ is attained at a knot. This fact makes that this approach is not appropriate when we use quadratic spline estimation.

4.2.1 Bootstrap method

The bootstrap method to determine the null distribution of s is similar as before (Sect. 4.1.1).

4.2.2 Asymptotic normality

Another approach to estimate the null distribution is motivated by the asymptotic normality of $\hat{\beta}'_k(t)$. Define

$$\mathbf{b}(t; q_k - 1) = (B_1(t; q_k - 1), B_2(t; q_k - 1), B_3(t; q_k - 1), \cdots B_{m_k - 1}(t; q_k - 1))^\top$$

$$(4.8)$$

and let $\mathbf{D} \in I\!R^{(m_k - 1) \times dim}$ denote the matrix such that

$$\mathbf{D}\hat{\boldsymbol{\alpha}} = (\Delta\hat{\alpha}_{k2}, \Delta\hat{\alpha}_{k3}, \cdots, \Delta\hat{\alpha}_{km_k})^\top.$$

Hence, we need to find the null distribution of $\hat{\beta}'_k(c) = K_{k,n}\mathbf{b}(c; q_k - 1)^\top \mathbf{D}\hat{\boldsymbol{\alpha}}$. This leads to the following result, for which the proof is given in Section 4 in Online Resource.

Theorem 3. *Under Assumptions 1−6 in Appendix B, where $\beta_p(\cdot)$ has a bounded fourth derivative with $q_p \geq 3$ for all p. Suppose $q_k = 3$ and $\lim_n \frac{K_n^9}{n \max_i N_i} = \infty$, then*

$$\frac{\hat{\beta}'_k(c) - \beta'_k(c)}{\sqrt{\mathrm{Var}\left(\hat{\beta}_k(c) - \beta'_k(c)\right)}} \xrightarrow{d} N(0, 1) \qquad as\ n \to \infty. \qquad (4.9)$$

We can use this asymptotic normality result if we can estimate the variance and $\beta'_k(c)$. For the variance, we have

$$\mathrm{Var}\left(\hat{\beta}_k(c) - \beta'_k(c)\right) = \mathrm{Var}(K_{k,n}\mathbf{b}(c; q_k - 1)^\top \mathbf{D}(\hat{\boldsymbol{\alpha}} - E(\hat{\boldsymbol{\alpha}}))) = K_{k,n}^2 \mathbf{b}(c; q_k - 1)^\top$$
$$\mathbf{D}(\mathbf{U}^\top\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{U}(\mathbf{U}^\top\mathbf{W}\mathbf{U})^{-1}\mathbf{D}^\top\mathbf{b}(c; q_k - 1) \in I\!R. \quad (4.10)$$

We can estimate $\mathbf{V}$ by $\hat{\mathbf{V}}$ as it is described in Sect. 3.3 to obtain the following estimate of the variance (4.10)

$$\hat{v}_k = K_{k,n}^2 \mathbf{b}(c; q_k - 1)\mathbf{D}(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}\mathbf{U}^\top \mathbf{W}\hat{\mathbf{V}}\mathbf{W}\mathbf{U}(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}\mathbf{D}^\top \mathbf{b}(c; q_k - 1)^\top. \tag{4.11}$$

We now have an estimate for the variance, but we still need to estimate $\beta_k'(c)$ to obtain the (estimated) null distribution of $\hat{\beta}'(c)$. Therefore we take $K_{k,n}\mathbf{b}(c; q_k - 1)\mathbf{D}\hat{\boldsymbol{\alpha}}^{cs}$ as a good approximation for $\beta_k'(c)$, where $\hat{\boldsymbol{\alpha}}^{cs}$ is the constrained cubic spline estimator.

The estimated α percentile $\tilde{Q}_\alpha$ is the α percentile of a $N(K_{k,n}\mathbf{b}(c; q_k - 1)\mathbf{D}\hat{\boldsymbol{\alpha}}^{cs}, \hat{v}_k)$.

4.2.3 Consistency

Suppose we estimate $\mathbf{V}$ by $\hat{\sigma}^2\mathbf{I}_N$ where

$$\hat{\sigma}^2 = \frac{1}{N - dim}(\mathbf{Y} - \mathbf{U}\hat{\boldsymbol{\alpha}})^\top(\mathbf{Y} - \mathbf{U}\hat{\boldsymbol{\alpha}}),$$

where $dim = \sum_{p=0}^d m_p$.

When there is correlation, even with this misspecified estimator of $\mathbf{V}$, Theorem 4 says that as n increases to infinity we correctly reject the null hypothesis with probability tending to 1 when the coefficient β_k is strictly decreasing in a point in the domain. Also, when $\beta_k(\cdot)$ is strictly increasing we do not reject the null hypothesis with probability tending to 1.

Theorem 4. *We construct a grid G_n which depends on n, such that $\min G_n = 0$, $\max G_n = 1$ and the supremum distance between two consecutive grid points tends to zero as $n \to \infty$. Under Assumptions 1-5 and the condition $K_n\rho_n + K_n^{-q_k} + K_n r_n = o(1)$ such that $\lim_n \frac{K_n^3}{n} = 0$ we have the following:*

1. *Suppose $\inf_{t\in[0,1]} \beta_k'(t) = \delta < 0$, then $\lim_{n\to\infty} P(s \geq \min(0, \tilde{Q}_\alpha)) = 0$.*
2. *Suppose $\inf_{t\in[0,1]} \beta_k'(t) = \delta > 0$, then $\lim_{n\to\infty} P(s < \min(0, \tilde{Q}_\alpha)) = 0$.*

The proof of this theorem is deferred to Section 5 in Online Resource. Theorem 4 states the effectiveness of this method under certain conditions when we consider the space $\mathcal{S} = \{\beta_k(\cdot) \in B^4([0,1]) \mid \inf_{t\in[0,1]} \beta_k'(t) \neq 0\}$. This function space can be seen as the space of smooth functions excluding increasing functions with a flat part.

5 Testing convexity

It is also of interest to test for convexity or impose convexity when estimating a coefficient function. We want to test the hypothesis that $\beta_k(\cdot)$ is convex, hence

$$H_1 : \beta_k''(t) \geq 0 \text{ for all } t \text{ in } [0,1], \quad \text{versus} \quad \neg H_1. \tag{5.1}$$

The testing procedures are analogous to these in Sect. 4, except for the obvious adjustments since we now work with the second derivative. Moreover, the consistency results in Sect. 4 are carried over under the appropriate adjustments of the conditions.

5.1 Cubic spline

When $q_k = 3$, the second derivative of the B-spline estimate $\hat{\beta}_k(\cdot)$ is a linear spline function. The second derivative of $\hat{\beta}_k(\cdot)$ is nonnegative if and only if (as in (3.2))

$$\mathbf{T}\boldsymbol{\alpha}_k \geq 0, \tag{5.2}$$

where $\mathbf{T} \in I\!\!R^{(K_{k,n}+1) \times (K_{k,n}+2)}$ is the matrix of second derivatives at the knots of B-splines corresponding to the coefficient $\beta_k(\cdot)$: $\mathbf{T}_{ij} = B''_{kj}(\xi_{k(i-1)}; 3)$. Define $\mathbf{H} = (\mathbf{0}_1, \mathbf{T}, \mathbf{0}_3)$ where $\mathbf{0}_1 \in I\!\!R^{(K_{k,n}+1) \times \sum_{j=0}^{k-1} m_j}, \mathbf{0}_3 \in I\!\!R^{(K_{k,n}+1) \times \sum_{j=k+1}^{d} m_j}$ are matrices with entries 0. Then we proceed as in Sect. 4.1.1 using the test statistic $\min(\mathbf{H}\hat{\boldsymbol{\alpha}})$. Moreover, the estimator $\hat{\boldsymbol{\alpha}}^{cs}$ is obtained under constraint (5.2).

5.2 Quartic spline

When $q_k = 4$ we use the bootstrap method as in Sect. 4.2.1. For the asymptotic normality approach (Sect. 4.2.2) we need the α percentile of $N(K_{k,n}^2 \mathbf{b}(c_2; q_k - 2)\mathbf{D}_2\hat{\boldsymbol{\alpha}}^{cs}, \hat{v}_{2k})$ denoted by $\hat{Q}_{2,\alpha}$, where $c_2 = \operatorname{argmin}_{t \in G} \hat{\beta}_k(t)$ and $\mathbf{D}_2$ is the matrix which takes the second order differences of $\hat{\boldsymbol{\alpha}}^{cs}$ and

$$\hat{v}_{2k} = K_{k,n}^4 \mathbf{b}(c_2; q_k - 2)\mathbf{D}_2(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}\mathbf{U}^\top \mathbf{W}\hat{\mathbf{V}}\mathbf{W}\mathbf{U}(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}$$
$$\mathbf{D}_2^\top \mathbf{b}(c_2; q_k - 2)^\top. \tag{5.3}$$

The constrained estimator $\hat{\boldsymbol{\alpha}}_k^{cs}$ is obtained as in Sect. 3.1.2 since we need to constrain a quadratic spline function. The only difference is that we work with second order differences because we work with the second order derivative of a quartic spline function.

6 Simultaneous testing

We address how to conduct a simultaneous shape test, i.e. test simultaneously whether certain coefficient functions are monotone and/or convex. Suppose we want to test for b shapes (monotonicity and/or convexity). For example, we test whether $\beta_{i_1}(\cdot)$ is monotonic, $\beta_{i_2}(\cdot)$ is convex,..., $\beta_{i_b}(\cdot)$ is convex where $i_1, \ldots, i_b$ is contained in $\{0, 1, \ldots, d\}$. For simplicity we require that $i_1, i_2, \ldots, i_b$ is a

mutually different sequence. Let $\mathbf{s}$ denote the column vector with length b of the corresponding test statistics. Thus for the example above

$$\mathbf{s} = (\min_{t \in G} \hat{\beta}'_{i_1}(t), \min_{t \in G} \hat{\beta}''_{i_2}(t), \ldots, \min_{t \in G} \hat{\beta}''_{i_b}(t))^\top,$$

for a suitable grid $G \subset [0, 1]$.

The first way is to test for the relevant coefficient functions based on the Bonferroni correction. Thus, we do the test for each relevant coefficient function with level α/b. We do not reject the null hypothesis if all individual tests give a positive answer, i.e. all individual null hypotheses are not rejected. When we use cubic splines for testing monotonicity (see Sect. 4.2.3) and quartic splines for testing convexity (see Sect. 5.2), the Bonferroni correction method yields a consistent multiple test when we restrict to the appropriate space.

Otherwise, it is quite straightforward to use the bootstrap approach as before (Sect. 4.1.1 and Sect. 4.2.1). If all the components of $\mathbf{s}$ are nonnegative, then we do not reject the null hypothesis. In the other case, we use a bootstrap method to determine how plausible such a test statistic is observed under the null hypothesis. As before we start with setting residuals

$$\hat{\epsilon}_{ij} = Y_{ij} - \sum_{p=0}^{d} X_{ij}^{(p)} \hat{\beta}_p(t_{ij})$$

where $\hat{\boldsymbol{\beta}}$ is the unconstrained B-spline estimator and let

$$Y_{ij}^{ps} = \sum_{p=0}^{d} X_{ij}^{(p)} \hat{\beta}_p^{cs}(t_{ij}) + \hat{\epsilon}_{ij} \quad \text{for } i = 1, \ldots, n \quad \text{and } j = 1, \ldots, N_i$$

be a set of pseudo responses under the null hypothesis with $\hat{\boldsymbol{\beta}}^{cs}(\cdot) = (\hat{\beta}_0^{cs}(\cdot), \ldots, \hat{\beta}_d^{cs}(\cdot))^\top$ the constrained estimate which is obtained by adding the appropriate constraints. Thus, for the example above we put the monotonicity constraint on $\beta_{i_1}(\cdot)$, the convexity constraint on $\beta_{i_2}(\cdot)$, ..., the convexity constraint on $\beta_{i_b}(\cdot)$. Then the bootstrap procedure to determine whether we do not reject the null hypothesis is as follows.

- Step 1: Resample n subjects with replacement from

$$\{(Y_{ij}^{ps}, X_{ij}, t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$$

  to obtain the bootstrap sample $\{(Y_{ij}^{ps*}, X_{ij}^*, t_{ij}^*) : i = 1, \ldots, n, j = 1, \ldots, N_i^*\}$
- Step 2: Repeat the above resampling procedure B times.
- Step 3: Obtain the test statistic vector from each bootstrap sample and derive the center of mass $\boldsymbol{\mu}_M$ and the sample covariance $\boldsymbol{\Sigma}_M$ of all test statistic vectors obtained from all the bootstrap samples. Then determine the sample distribution of all Mahalanobis distances.
- Step 4: Take the $(1 - \alpha)$ percentile $M_{1-\alpha}$ of the Mahalanobis distances obtained in Step 3 and reject the null hypothesis if $(\mathbf{s} - \boldsymbol{\mu}_M)^\top \boldsymbol{\Sigma}_M (\mathbf{s} - \boldsymbol{\mu}_M) > M_{1-\alpha}$, else do not reject the null hypothesis.

7 Simulation examples

In this section we evaluate the testing procedures on simulated data with significance level $\alpha = 0.05$. We assume a longitudinal model with mean zero errors. Let the number of individuals be n, the number of repeated measurements for individual i is N_i which is randomly generated from $\{\lfloor 3n/10 \rfloor, \ldots, \lfloor 4n/10 \rfloor\}$ for $i = 1, \ldots, n$. For each individual i, the time points t_{ij}, $j = 1, \ldots, N_i$ are equidistant in $[0, 1]$ and blurred by adding a random variable with distribution $N(0, 5 \cdot 10^{-5})$. For the error structure we consider two settings. In the first setting $\varepsilon(t_{ij}) \sim N(0, 0.6^2)$ and in the second $\varepsilon(t_{ij}) \sim Un[-1.73, 1.73]$. In both settings, the error terms from different subjects are independent and the intrasubject correlation is

$$\mathrm{Corr}(\varepsilon(t_{ij}), \varepsilon(t_{ik})) = 0.2, \qquad 1 \le i \le n, \ 1 \le j, k \le N_i, \ j \ne k. \qquad (7.1)$$

Throughout the simulations, the signal-to-noise ratio (SNR) is around 7. The SNR is defined by

$$\frac{\mathrm{Var}\left(\beta_0(t) + \sum_{p=0}^{d} X^{(p)}(t)\beta_p(t)\right)}{\mathrm{Var}\left(\varepsilon(t)\right)}.$$

In practice the SNR is estimated by its sample version. Moreover, the simulation results are based on 200 samples. First we study the performances for the test whether a certain coefficient function is increasing, followed by a study for testing the convexity of one coefficient function and for simultaneous testing.

7.1 Monotonicity tests

We consider $n = 50$ and $n = 100$. We use coefficient functions $\beta_0(t) = 0.25 + 2t$, $\beta_1(t) = f(t)$, $\beta_2(t) = -0.5 + 10(t - 0.5)^2$. The hypothesis (4.1) is tested for β_1 using several functions f. We consider

- $f_{1,a}(t) = -2 + 2(1 + t - a \exp(-50(t - 0.5)^2))$;
- $f_2(t) = 1.1$;
- $f_3(t) = 5(t - 0.25)^2$;
- $f_4(t) = 5(t - 0.25)_+^2$.

The function $f_{1,a}(\cdot)$ is taken from Bowman et al. (1998). This function is strictly monotone for $a = 0.15$, whereas a dip appears when $a = 0.30$ and more profoundly when $a = 0.45$. The function $f_3(\cdot)$ is a parabola which is strictly decreasing on $[0, 0.25]$ and strictly increasing on $[0.25, 1]$. Wang and Meyer (2011) used $f_4(\cdot)$, which is the same as the function $f_3(\cdot)$ but is zero for $t \le 0.25$, to show that certain testing procedures reject too often the null hypothesis in (4.1) when the increasing function has flat parts.

We employ the time dependent bivariate vector

$$\begin{pmatrix} X^{(1)}(t) \\ X^{(2)}(t) \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}(t)), \quad \boldsymbol{\Sigma}(t) = \begin{pmatrix} 1 & 1/(4 + t) \\ 1/(4 + t) & 1 \end{pmatrix}. \qquad (7.2)$$

$f(\cdot), n$	Characteristics	Normal errors			Uniform errors		
	of $f(\cdot)$	quadratic	cubic splines	cubic	quadratic	cubic splines	cubic
		2	3 (B)	3 (AN)	2	3 (B)	3 (AN)
$f_{1,0.15}(\cdot), 50$	increasing	0.045	0.010	0.045	0.025	0.015	0.025
$f_{1,0.15}(\cdot), 100$	function	0.030	0.005	0.020	0.015	0.005	0.010
$f_{1,0.3}(\cdot), 50$	non-increasing	0.445	0.340	0.900	0.450	0.340	0.955
$f_{1,0.3}(\cdot), 100$	with small dip	0.945	0.990	1.000	0.965	0.990	1.000
$f_{1,0.45}(\cdot), 50$	non-increasing	0.910	0.920	1.000	0.910	0.940	1.000
$f_{1,0.45}(\cdot), 100$	with large dip	1.000	1.000	1.000	1.000	1.000	1.000
$f_2(\cdot), 50$	constant	0.075	0.045	**0.335**	0.055	0.040	**0.385**
$f_2(\cdot), 100$	function	0.050	0.045	**0.705**	0.055	0.065	**0.625**
$f_3(\cdot), 50$	non-monotonic,	0.510	0.545	0.600	0.480	0.545	0.595
$f_3(\cdot), 100$	parabolic	0.875	0.935	1.000	0.835	0.950	1.000
$f_4(\cdot), 50$	increasing	0.050	0.045	**0.300**	0.045	0.050	**0.265**
$f_4(\cdot), 100$	but with flat part	0.020	0.045	**0.530**	0.035	0.040	**0.515**

Table 1: The rejection rates for the hypothesis that $\beta_1(\cdot)$ is increasing are stated based on 200 simulations. The functions in the most left column are consecutively substituted in $\beta_1(\cdot)$. In the case of cubic splines we differentiate between the bootstrap method (denoted by B) and the method based on asymptotic normality (AN).

We set for the cubic spline testing procedure $K_{0,50} = K_{1,50} = K_{2,50} = 4$ with $q_0 = q_1 = q_2 = 3$, whereas for the quadratic spline testing procedure $K_{0,50} = K_{2,50} = 4, K_{1,50} = 5$ with $q_0 = q_2 = 3, q_1 = 2$. The number of knots are chosen in such a way that the number of B-splines for each coefficient is the same irrespective of the degree we use.

The rejection rates are collected in Table 1. The quadratic spline testing procedure is based on the bootstrap method (Sect. 4.1.1) with bootstrap size equal to 200. For the cubic spline approach we look at both the bootstrap method and the method based on asymptotic normality (Sect. 4.2.1-4.2.2) where we restrict the grid G to $[0.05, 0.95]$ using 100 equidistant grid points. In the asymptotic normality approach (AN) there is the obstacle of estimating the covariance (see (4.10)). Note that Theorem 4 holds when we replace $\hat{\mathbf{V}}$ in (4.10) by $\hat{\sigma}^2 \mathbf{I}_N$. The rejection rates in the asymptotic normality approach where we use the true covariance matrix are very close (differences of less than 1%) to the rejection rates using the (misspecified) covariance $\hat{\sigma}^2 \mathbf{I}_N$, therefore we only report the results based on $\hat{\sigma}^2 \mathbf{I}_N$.

From the results in Table 1 we see that the bootstrap outcomes (B) are comparable to each other and consistent as n increases, i.e. the rejection rates when the function is not increasing ($f_{1,0.3}(\cdot), f_{1,0.45}(\cdot)$ and $f_3(\cdot)$) tend to 1 while the rejection rate is about 0.05 (the testing level, as the sample size increases) for increasing functions ($f_{1,0.15}(\cdot), f_2(\cdot)$ and $f_4(\cdot)$). The asymptotic normality approach (AN) rejects too often for increasing functions which are not contained in $\mathcal{S}$ (see Sect. 4.2.3), i.e. the constant function and the increasing function with a flat part (bold numbers in Table 1). This seems to be inherent to the

approach, since the simulations for a much bigger $n = 500$ reveal extremely high rejection rates (0.995 for the constant function and 1 for the increasing function with a flat part). This is fully in line with the theoretical result in Sect. 4.2.3, in which the consistency of the AN testing procedure is established but for this excluding functions that have flat parts was needed. However, for functions in $\mathcal{S}$ ($f_{1,0.15}(\cdot), f_{1,0.30}(\cdot), f_{1,0.45}(\cdot)$ and $f_3(\cdot)$) the asymptotic normality approach performs better than the bootstrap methods. We could expect this bad performance for functions with flat parts since the AN approach is based on the pointwise asymptotic normality result given in Theorem 3, and for increasing functions with flat parts the test statistic s is attained at different grid points while the AN approach assumes the grid point where the minimum is attained to be unique. Furthermore, the results for normal errors and uniform errors are comparable. The first two rows of Table 1 suggest that the test level is not reached as n increases. For the AN approach this seems to be in contradiction with result (4.9) which uses the true coefficient function and the true variance with increasing number of knots. However, we estimate both the coefficient function and the variance with fixed number of knots (which do not increase as n increase). By our remark at the end of the previous paragraph, the 'problem' seems to lie with the estimation of the coefficient function (under the null hypothesis). The results of the bootstrap approach which uses the same coefficient function estimation agrees with this conclusion.

Therefore, the recommendation is to use the AN approach when it is known that the true coefficient function has no flat parts. Otherwise it is better to use the bootstrap approaches.

7.2 Convexity and simultaneous tests

In addition, we test the convexity of $\beta_0(t) = 0.25 + 2t$ in first instance and $\beta_1(\cdot) = f_{1,0.45}(\cdot)$ in second instance, using quartic splines and the bootstrap method. To illustrate the effectiveness of our simultaneous approach, we also test whether $\beta_0(\cdot)$ and $\beta_2(\cdot)$ are both increasing using cubic splines and the bootstrap method. Table 2 contains the results where we have fixed the knots $K_{0,50} = K_{1,50} = K_{2,50} = 4$. As before, we see that the powers tend to one and we do not reject too often when the null hypothesis holds.

8 Real data application

We consider data from the National Institute of Mental Health Schizophrenia Collaborative Study. Specifically, we study Item 79, 'Severity of Illness', of the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr and Klett (1966)). Item 79 was originally measured on a numerical scale ranging from 1 (normal, not at all ill) to 7 (among the most extremely ill). In this study, most patients were measured at weeks 0, 1, 3 and 6; however, a few patients were additionally

Hypothesis	$n = 50$	$n = 100$	$n = 150$
$\beta_0(\cdot)$ is convex (case one)	0.04	0.025	0
$\beta_1(\cdot)$ is convex (case one)	0	0.47	1
$\beta_0(\cdot)$ and $\beta_2(\cdot)$ are increasing (case two)	1	1	1
$\beta_0(\cdot)$ and $\beta_2(\cdot)$ are increasing (case three)	0.265	0.915	0.955

Table 2: The rejection rates stated are based on 200 simulations with normal errors. These results are based on the bootstrap method using cubic splines for monotonicity testing and quartic splines for convexity testing. Three cases are considered. In each case $\beta_0(t) = 0.25 + 2t$, and further in case one $\beta_1(t) = f_{1,0.45}(t)$, $\beta_2(t) = -0.5 + 10(t - 0.5)^2$; for case two we take $\beta_1(t) = f_{1,0.30}(t)$, $\beta_2(t) = -0.5 + 10(t - 0.5)^2$; for case three we use $\beta_1(t) = -0.5 + 10(t - 0.5)^2$, $\beta_2(t) = f_{1,0.30}(t)$.
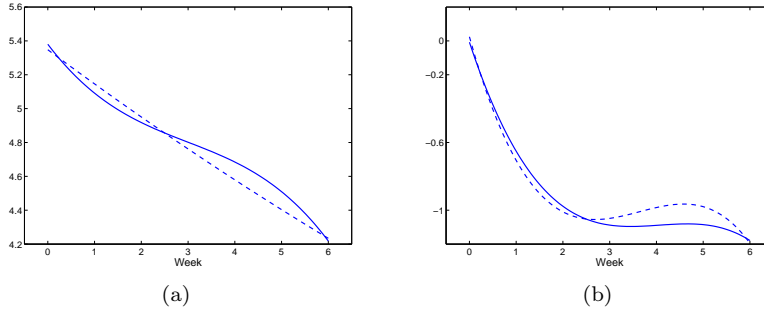

(a)   (b)

Fig. 8.1: Schizophrenia data. The full lines are the cubic spline estimations (i.e. the degree vector is $(3,3)$) of the coefficients $\beta_0(\cdot)$ and $\beta_1(\cdot)$, the dashed lines are the estimations when the degree vector is $(3,2)$. (a) contains the estimations of $\beta_0(\cdot)$, and (b) the estimations of $\beta_1(\cdot)$.

measured at weeks 2, 4 and 5. The $n = 437$ patients were randomly assigned to either receive a drug or a placebo. The data are available in R (Package 'vcrpart'). Previously, these data were studied by Hedeker and Gibbons (1997) who used a mixed-effects probit regression model to model the changes of the 'Severity of Illness' measurements. Here, we study the changing of Item 79 $(= Y)$ with the VCM

$$Y(week) = \beta_0(week) + \beta_1(week)Drug + \varepsilon(week), \qquad (8.1)$$

$Drug$ is a binary variable where $Drug = 1$ denotes a patient who received the drug, otherwise the patient received a placebo. The number of knots are determined by a 4-fold cross validation where the possible number of knots varies from 1 to 8. This yields the knot vector $(1,1)$ for both the degree vectors $(3,3)$ and $(3,2)$. As for the degree vector $(2,3)$, the knot vector obtained is $(2,1)$. In Figure 8.1 we depict the cubic spline estimators for $\beta_0(\cdot)$ and $\beta_1(\cdot)$ (the solid curves in respectively Figures 8.1(a) and (b)). We also present, as
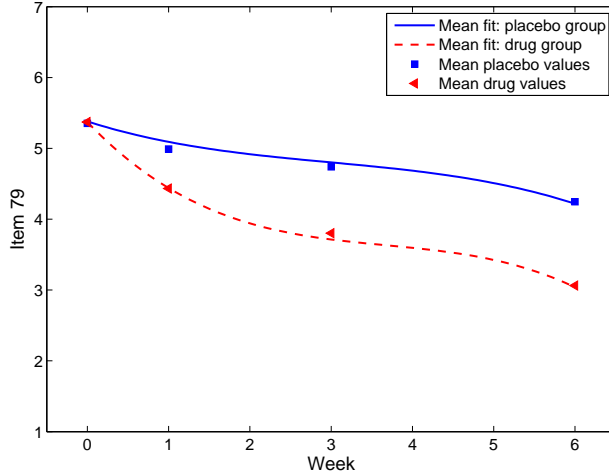
Fig. 8.2: Schizophrenia data. The mean fits from model (8.1) are shown of the placebo group and the drug group. The squares and triangles are the mean Item 79 measurements at weeks 0, 1, 3 and 6, of the placebo group and drug group, respectively.

dashed curves, the spline estimators when using the degree vector $(2, 3)$. The question of interest is how the drug affects the illness of the patients, that is the coefficient $\beta_1(\cdot)$. A negative $\beta_1(\cdot)$ which is decreasing proves that the drug is effective. The full line in Figure 8.1(b) suggests that such is the case. Moreover, from Figure 8.1(b) the drug effect drops quickly to reach a steady effect of -1 from week 3 onwards. Figure 8.1(a) shows an overall mildly decreasing trend for the baseline coefficient function $\beta_0(\cdot)$, revealing a little improvement of the illness over time (in weeks).

Figure 8.2 contains the mean fits, i.e. $\hat{\beta}_0(week) + \hat{\beta}_1(week)Drug$, for the placebo group and the drug group. We see that the varying of mean Item 79 measurements are well described by model (8.1) for both groups.

Table 3 contains the results of the monotonicity tests. The asymptotic normality results are included for completeness. However, as we have concluded in the end of Sect. 7, we base our judgments on the bootstrap approach. For test level 0.10 we reject the hypothesis that $\beta_0(\cdot)$ is increasing, moreover, the high p-values for the decreasing hypothesis indicates that $\beta_0(\cdot)$ is decreasing. For $\beta_1(\cdot)$ we totally reject the hypothesis that it is increasing. For cubic splines we do not reject the hypothesis that $\beta_1(\cdot)$ is decreasing. On contrary is the result for quadratic splines, because the cubic fit is monotonically decreasing while the quadratic fit is not from week 4 onwards (see Figure 8.1(b)). This discrepancy can be explained by noting that very few measurements were taken on weeks 4 and 5.

Hypothesis	p-values		
	2	3 (B)	3 (AN)
$\beta_0(\cdot)$ is increasing	0.08	0.05	0.21
$\beta_0(\cdot)$ is decreasing	0.69	0.65	0.50
$\beta_1(\cdot)$ is increasing	0.00	0.00	0.02
$\beta_1(\cdot)$ is decreasing	0.03	0.70	0.35

Table 3: The p-values concerning the Schizophrenia data with bootstrap size $B = 500$. In the case of cubic splines we differentiate between the bootstrap method (denoted by B) and the method based on asymptotic normality (AN).

9 Conclusion and discussion

In this paper we presented two approaches for (simultaneous) shape testing in varying coefficient models. We showed in simulation examples that in general the bootstrap approach is consistent and yields the best results. As an application we modeled the 'Severity of Ilness' measurements from the Schizophrenia data by a varying coefficient model with covariate a binary variable which indicates whether the patient received a drug, and where the coefficients depend on the week number. Our interest has been the drug coefficient and whether it is decreasing, which would reveal how affective the drug is. It turned out that the drug coefficient is negative and mainly monotonically decreasing as the age increases.

Varying coefficient models are also applied in other contexts, in particular in the generalized linear models context and in survival analysis (see Fan and Wenyang (2008) for an overview). It should be noted that the bootstrap approach is quite universal because one mainly needs an estimation of the B-spline coefficients which satisfies the null hypothesis, i.e. one only needs to impose the relevant constraints on the B-spline coefficients.

Appendix A   Notation

In this paper two submultiplicative matrix norms are considered. For a real matrix $\mathbf{A} \in I\!\!R^{n_1 \times n_2}$, $\|\mathbf{A}\|_2$ denotes the Frobenius norm: $\|\mathbf{A}\|_2 = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{A}_{ij}^2}$. The norm $\|\cdot\|_\infty$ is defined by $\|\mathbf{A}\|_\infty = \max_{i=1,\ldots,n_1} \sum_{j=1}^{n_2} |\mathbf{A}_{ij}|$. Further, we use the notation

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^\top \qquad \mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^\top$$

$$\mathbf{B}(t) = \begin{pmatrix} B_{01}(t; q_0) \ldots B_{0m_0}(t; q_0) & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \ddots & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 0 \ldots 0 & B_{d1}(t; q_d) & \ldots & B_{dm_d}(t, q_d) \end{pmatrix},$$

$$\mathbf{B}(t) \in I\!\!R^{d \times dim} \text{ where } dim = \sum_{p=0}^{d} m_p$$

$$\mathbf{U}'_{ij} = \mathbf{X}'_{ij} \mathbf{B}(t_{ij}) \in I\!\!R^{1 \times dim} \qquad \mathbf{U}_i = (\mathbf{U}_{i1}, \ldots, \mathbf{U}_{iN_i})' \in I\!\!R^{N_i \times dim}$$

$$\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_n)' \in I\!\!R^{N \times dim}, \text{ where } N = \sum_{i=1}^{n} N_i$$

$$\mathbf{W}_i = \mathrm{diag}\left(N_i^{-1}, \ldots, N_i^{-1}\right) \in I\!\!R^{N_i \times N_i} \quad \text{(a diagonal matrix with } N_i \text{ times}$$
$$N_i^{-1} \text{ on the diagonal)}$$

$$\mathbf{W} = \mathrm{diag}\left(\mathbf{W}_1, \ldots, \mathbf{W}_n\right) \in I\!\!R^{N \times N} \quad \text{(a block diagonal matrix}$$
$$\text{with the matrices } \mathbf{W}_i \text{ on the diagonal)}$$

$$\boldsymbol{\varepsilon}_i = (\varepsilon(t_{i1}), \ldots, \varepsilon(t_{iN_i}))^\top \qquad \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)^\top$$
$$\mathbf{V}_i = E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top) \qquad \mathbf{V} = E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top)$$

Appendix B Assumptions

Assumptions:

1. The observation times t_{ij}, $j = 1, \ldots, N_i$, $i = 1, \ldots, n$, are chosen independently according to a distribution function $F_T(t)$ on $\mathcal{T}$. Moreover, they are independent of the response and the covariate process $\{(Y_i(t), X_{i1}(t), \ldots, X_{id}(t))\}$, $i = 1, \ldots, n$. The distribution function $F_T(t)$ has a Lebesgue density $f_T(t)$ that is bounded away from zero and infinity, uniformly over all $t \in \mathcal{T}$, that is, there exist positive constants M_1 and M_2 such that $M_1 \leqslant f_T(t) \leqslant M_2$ for all $t \in \mathcal{T}$.

2. The eigenvalues $\eta_0(t), \ldots, \eta_d(t)$ of $\boldsymbol{\Sigma}(t) = E(\mathbf{X}(t)\mathbf{X}(t)')$ are bounded away from zero and infinity, uniformly over all $t \in \mathcal{T}$, that is, there exist positive constants M_3 and M_4 such that $M_3 \leqslant \eta_0(t) \leqslant \ldots \leqslant \eta_d(t) \leqslant M_4$ for all $t \in \mathcal{T}$.

3. There exists a positive constant M_5 such that $|X_p(t)| \leqslant M_5$ for all $t \in \mathcal{T}$ and $p = 0, \ldots, d$.

4. There exists a positive constant M_6 such that $E(\varepsilon^2(t)) \leqslant M_6 < \infty$ for all $t \in \mathcal{T}$.

5. $\limsup_{n \to \infty} \left(\frac{\max_p m_p}{\min_p m_p}\right) < \infty$.

6. The process $\varepsilon(t)$ can be decomposed as the sum of two independent stochastic processes, $\varepsilon^{(1)}(t)$ and $\varepsilon^{(2)}(t)$, where $\varepsilon^{(1)}(t)$ is an arbitrary mean zero process and $\varepsilon^{(2)}(t)$ is a process of measurement errors that are independent at different time points and have mean zero and constant variance σ^2.

References

Antoniadis, A., Gijbels, I. and Verhasselt, A. (2012), 'Variable selection in varying coefficient models using p-splines', *Journal of Computational and Graphical Statistics* **21**(3), 638–661.

Bowman, A. W., Jones, M. C. and Gijbels, I. (1998), 'Testing monotonicity of regression', *Journal of Computational and Graphical Statistics* **7**(4), 489–500.

de Boor, C. (2001), *A Practical Guide to Splines*, number v. 27 *in* 'Applied Mathematical Sciences', Springer.

Diggle, P. J. (1988), 'An approach to the analysis of repeated measurements', *Biometrics* **44**(4), 959–971.

Diggle, P. and Verbyla, A. (1998), 'Non-parametric estimation of covariance structure in longitudinal data', *Biometrics* **54**, 401–415.

Fan, J. and Wenyang, Z. (2008), 'Statistical methods with varying coefficient models', *Statistics and its interface* **1**(1), 179–195.

Fan, J. and Zhang, W. (1999), 'Statistical estimation in varying coefficient models', *The Annals of Statistics* **27**(5), 1491–1518.

Ghosal, S., Sen, A. and van der Vaart, A. W. (2000), 'Testing monotonicity of regression', *The Annals of Statistics* **28**(4), 1054–1082.

Hastie, T. and Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4), 757–796.

Hedeker, D. and Gibbons, R. D. (1997), 'Application of random-effects pattern-mixture models for missing data in longitudinal studies', *Psychological Methods* **2**(6), 64–78.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998), 'Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data', *Biometrika* **85**(4), 809–822.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002), 'Varying-coefficient models and basis function approximations for the analysis of repeated measurements', *Biometrika* **89**(1), 111–128.

Huang, J. Z., Wu, C. O. and Zhou, L. (2004), 'Polynomial spline estimation and inference for varying coefficient models with longitudinal data', *Statistica Sinica* **14**(3), 763–788.

Lorr, M. and Klett, C. J. (1966), *Inpatient Multidimensional Psychiatric Scale: Manual.*, Palo Alto, CA: Consulting Psychologists Press.

Schumaker, L. (2007), *Spline Functions: Basic Theory*, Cambridge University Press.

Silvapulle, M. J. and Sen, P. K. (2005), *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*, John Wiley & Sons.

Wang, J. C. and Meyer, M. C. (2011), 'Testing the monotonicity or convexity of a function using regression splines', *Canadian Journal of Statistics* **39**, 89–107.

Zhang, H.-G. and Mei, C.-L. (2012), 'Sizer inference for varying coefficient models', *Communications in Statistics - Simulation and Computation* **41**(10), 1944–1959.

Zhang, H.-G., Mei, C.-L. and Wang, H.-L. (2013), 'Robust sizer approach for varying coefficient models', *Mathematical Problems in Engineering* **2013**(Article ID 547874), 13 pages.