Chapter 3
Testing of Growth Curves with Cubic Smoothing Splines

Tapio Nummi and Nicholas Mesue

Abstract In this paper we present a novel method for testing growth curves when the analysis is based on spline functions. The new method is based on the use of a spline approximation. For the approximated spline model an exact F-test is developed. This method also applies under a certain type of correlation structures that are especially important in the analysis of repeated measures and growth data. We tested this method on the glucose data of Zerbe (J Am Stat Assoc 74:215–221, 1979) and also investigated it by simulation experiments. The new method proved to be a very powerful modeling and testing tool especially in situations, where the growth curve may not be easy to approximate using simple parametric models.

3.1 Introduction

Longitudinal research has an important role in various fields of science, for example in medicine, economics, social sciences, and engineering. The aim is to analyze the change caused, e.g., by growth, degradation, maturation, and ageing when individuals are followed over time or according to some other ordered sequence of measurements. In this paper the focus is on complete and balanced data. One of the most important statistical models for these data is the growth curve model of Potthoff and Roy (1964). The early development of this model was mainly based on the unstructured MANOVA assumption of the covariance matrix of independent random vectors (e.g., Khatri 1966 and Grizzle and Allen 1969). Later, however, more attention has been paid to modeling the covariance matrix by using parsimonious covariance structures (see, e.g., Azzalini 1987, Lee 1988 and Nummi 1997). For excellent reviews of the growth curve model we refer to the books by Kshirsagar and Smith (1995) and Pan and Fang (2002).

T. Nummi (✉) • N. Mesue
School of Health Sciences, FIN-33014 University of Tampere, Finland
e-mail: tan@uta.fi

Our approach is to use cubic smoothing splines to model the mean growth curve. As is very well known cubic smoothing splines are very flexible curves with interesting mathematical properties (see, e.g., Green and Silverman 1994). For an up-to-date summary of recent methods of smoothing splines and nonparametric regression we refer to Wu and Zhang (2006). Approximate inference with smoothing splines have been studied, e.g., in Eubank and Spiegelman (1990), Schimek (2000), and Cantoni and Hastie (2002). In their simulation study Liu and Wang (2004) compared six testing statistics. Nummi et al. (2011) provided a test of a regression model against spline alternative for correlated data. The main focus in these studies have been on testing the order of the polynomial model against a spline alternative. However, testing if two or more splines are equal would be very important in many applications. Nummi and Koskela (2008) introduced some results for the estimation and rough testing of growth curves when the analysis is based on spline functions. However, very little research about testing equality of smoothing splines, especially for correlated data, has been carried out so far. In this paper we focus on testing if the progression in time is equal over the set of correlated observations.

In Sect. 3.2 we introduce the basic growth model and its estimation using cubic smoothing splines. In Sect. 3.3 a spline approximation is introduced and a test for mean curves is developed. In Sect. 3.4 a computational example of Glucose data is presented and the method is also investigated by simulation experiments.

3.2   Basic Spline Growth Model and Some Properties

One of the most important statistical models for balanced complete multivariate repeated measures data is the GMANOVA (Generalized Multivariate Analysis of Variance Model) of Potthoff and Roy (1964). The model is often also refered to as the growth curve model. This model can be written as

$$\mathbf{Y} = \mathbf{TBA}' + \mathbf{E}, \tag{3.1}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ is the matrix of independent response vectors, $\mathbf{T}$ is a $q \times p$ within-individual design matrix, $\mathbf{A}$ is an $n \times m$ between-individual design matrix, $\mathbf{B}$ is an unknown $p \times m$ parameter matrix to be estimated, and $\mathbf{E}$ is a $q \times n$ matrix of random errors. It is assumed that the columns $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbf{E}$ are independently normally distributed as $\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$. In the original model formulation $\boldsymbol{\Sigma}$ was assumed to be an unstructured covariance matrix and the analyses were mainly based on the methods developed for linear models and multivariate analysis.

Often when analyzing growth data the true growth function is more or less unknown and there may not be any theoretical justification for any specific parametric form of the curve. Parametric models are then used for descriptive purposes rather than interpretative to summarize the information of development profile. A natural first choice in such situations is a low order polynomial curve.

However, in many cases these models may fail to reveal important features of the growth process and more complicated models are therefore also needed.

Our approach is to use the cubic smoothing splines to model the mean growth curve. As is very well known cubic smoothing splines are very flexible curves with interesting mathematical properties (see, e.g., Green and Silverman 1994). We can write the model (3.1) in a slightly more general form as (see also Nummi and Koskela 2008)

$$\mathbf{Y} = \mathbf{G}\mathbf{A}' + \mathbf{E}, \tag{3.2}$$

where $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_m)$ is the matrix of smooth mean growth curves in time points $t_1, t_2, \ldots, t_q$. We assume that the covariance matrix $\mathbf{\Sigma}$ takes certain type of parsimonious structure $\mathbf{\Sigma} = \sigma^2 \mathbf{R}(\theta)$ with covariance parameters θ. In sequel we refer to this model as the spline growth model (SGM). The growth curve model of Potthoff and Roy (1964) is now the special case $\mathbf{G} = \mathbf{TB}$. The smooth solution for $\mathbf{G}$ can be obtained by minimizing the penalized least squares (PLS) criterion

$$Q = \text{tr}[(\mathbf{Y} - \dot{\mathbf{G}})'\mathbf{H}(\mathbf{Y} - \dot{\mathbf{G}}) + \alpha \dot{\mathbf{G}}'\mathbf{K}\dot{\mathbf{G}}], \tag{3.3}$$

where we denote $\dot{\mathbf{G}} = \mathbf{G}\mathbf{A}'$, $\mathbf{H} = \mathbf{R}^{-1}$, and $\mathbf{K}$ is the so-called roughness matrix arising from the common roughness penalty $RP = \int g''^2$ and α is a fixed smoothing parameter. For cubic smoothing splines the roughness matrix is

$$\mathbf{K} = \nabla \mathbf{\Delta}^{-1} \nabla', \tag{3.4}$$

where the nonzero elements of banded $q \times (q-2)$ and $(q-2) \times (q-2)$ matrices ∇ and $\mathbf{\Delta}$, respectively, are

$$\nabla_{k,k} = \frac{1}{h_k}, \ \nabla_{k+1,k} = -\left(\frac{1}{h_k} + \frac{1}{h_{k+1}}\right), \ \nabla_{k+2,k} = \frac{1}{h_{k+1}} \tag{3.5}$$

and

$$\mathbf{\Delta}_{k,k+1} = \mathbf{\Delta}_{k+1,k} = \frac{h_{k+1}}{6}, \ \mathbf{\Delta}_{k,k} = \frac{h_k + h_{k+1}}{3}, \tag{3.6}$$

where $h_j = x_{j+1} - x_j$, $j = 1, 2, \ldots, (q-1)$, and $k = 1, 2, \ldots, (q-2)$. It can be shown that Q can be rewritten in an alternative form

$$Q = \text{tr}[\{\dot{\mathbf{G}} - (\mathbf{H} + \alpha \mathbf{K})^{-1}\mathbf{H}\mathbf{Y}\}'(\mathbf{H} + \alpha \mathbf{K})\{\dot{\mathbf{G}} - (\mathbf{H} + \alpha \mathbf{K})^{-1}\mathbf{H}\mathbf{Y}\} + c], \tag{3.7}$$

where c is a constant and $(\mathbf{H} + \alpha \mathbf{K})$ is a positive definite matrix. The function Q is minimized for given α and $\mathbf{H}$ when $\dot{\mathbf{G}} = (\mathbf{H} + \alpha \mathbf{K})^{-1}\mathbf{H}\mathbf{Y}$. This gives the spline estimator

$$\tilde{\mathbf{G}} = (\mathbf{H} + \alpha \mathbf{K})^{-1}\mathbf{H}\mathbf{Y}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}. \tag{3.8}$$

However, the covariance matrix $\mathbf{H}$ may not be known and therefore the estimator (3.8) maybe difficult to use in practical situations. Fortunately, it can be shown that in certain important special cases the general spline estimator (3.8) simplifies to simple linear functions of the original observations $\mathbf{Y}$. One obvious condition for such kind of simplification is

$$\mathbf{KR} = \mathbf{K} \tag{3.9}$$

and since now $\mathbf{K} = \mathbf{KH}$ the spline estimators $\tilde{\mathbf{G}}$ can be simplified as

$$\hat{\mathbf{G}} = (\mathbf{I} + \alpha\mathbf{K})^{-1}\mathbf{YA}(\mathbf{A'A})^{-1} = \mathbf{SYA}(\mathbf{A'A})^{-1}, \tag{3.10}$$

where the smoother matrix is $\mathbf{S} = (\mathbf{I} + \alpha\mathbf{K})^{-1}$. Covariance matrices satisfying the condition (3.9) have been studied in Nummi and Koskela (2008) and Nummi et al. (2011). Some important special cases of these structures useful for growth data are $\mathbf{R} = \mathbf{I}$, $\mathbf{R} = \mathbf{I} + \sigma_d^2\mathbf{11'}$, $\mathbf{R} = \mathbf{I} + \sigma_{d'}^2\mathbf{XX'}$ and $\mathbf{R} = \mathbf{I} + \mathbf{XDX'}$, where $\mathbf{X} = (\mathbf{1}, \mathbf{x})$ and $\mathbf{x}$ is a vector of q measuring times.

If we apply the result $\text{vec}(\mathbf{ABC}) = (\mathbf{C'} \otimes \mathbf{A})\text{vec}(\mathbf{B})$, where the vec operation rearranges the columns of a matrix underneath each other, we can write the basic model (3.2) in a vector form

$$\mathbf{y} = (\mathbf{A} \otimes \mathbf{I}_q)\mathbf{g},$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$ and $\mathbf{g} = \text{vec}(\mathbf{G})$. If the spline estimates are written in vector form we have

$$\hat{\mathbf{g}} = [(\mathbf{AA'})^{-1}\mathbf{A'} \otimes \mathbf{S}]\mathbf{y}$$

and the smoother of the whole data is

$$\hat{\mathbf{y}} = (\mathbf{P}_a \otimes \mathbf{S})\mathbf{y} = \mathbf{S}_*\mathbf{y}, \tag{3.11}$$

where we denote $\mathbf{P}_a = \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}$ and $\mathbf{S}_* = (\mathbf{P}_a \otimes \mathbf{S})$. The effective degrees of freedom of the smoother can now be given as

$$edf_* = \text{tr}(\mathbf{S}_*) = \text{tr}(\mathbf{P}_a \otimes \mathbf{S}) = \text{tr}(\mathbf{P}_a)\text{tr}(\mathbf{S}) = m \times edf, \tag{3.12}$$

where $edf = \text{tr}(\mathbf{S})$ is the effective degrees of freedom of the smoother $\mathbf{S}$. It is further easy to see that the generalized cross-validation criteria for choosing the smoothing parameter α take the form

$$GCV(\alpha) = \frac{\frac{1}{nq}\sum_{i=i}^{nq}[y_i - \hat{y}_i]^2}{(1 - \frac{m \times edf}{nq})^2}, \tag{3.13}$$

where y_i and $\hat{y}_i$ are individual elements of the observed and smoothed vectors $\mathbf{y}$ and $\hat{\mathbf{y}}$, respectively.

3.3 Testing of Mean Curves

It is very well known that exact tests may be difficult to develop when making statistical inference based on smoothing splines. Our interest in this study focuses on testing if the progression in time is the same in treatment groups considered. In this study an exact test based on spline approximations for testing growth curves is developed.

3.3.1 Spline Approximation

It has been demonstrated by Nummi et al. (2011) that the approximation discussed in this paper is quite good for relatively smooth data. More detailed consideration of spline approximations can be found, e.g., in Hastie (1996). In a general case the smoother matrix $\mathbf{S}$ is not a projection matrix and therefore certain results, e.g. in testing, developed for general linear models are not directly applicable. Our approach is to utilize an approximation for the smoother matrix $\mathbf{S}$ with the properties of a projection matrix. As discussed by Hastie (1996) the smoother matrix can be written as

$$\mathbf{S} = \mathbf{M}(\mathbf{I} + \alpha\Lambda)^{-1}\mathbf{M}', \tag{3.14}$$

where $\mathbf{M}$ is the matrix of q orthogonal eigenvectors of $\mathbf{K}$ and Λ is a diagonal matrix of corresponding q eigenvalues. It is easily seen that $\mathbf{K}$ and $\mathbf{S}$ share the same set of eigenvectors $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_q$ and the eigenvalues are connected such that the eigenvalues of $\mathbf{S}$ are $\gamma = 1/(1 + \alpha\lambda)$. In sequel we assume that eigenvectors $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_q$ are ordered according to the eigenvalues of $\mathbf{S}$. It is well known that the sequence of eigenvectors appears to increase in complexity like a sequence of orthogonal polynomials. The first two eigenvalues of $\mathbf{S}$ are always 1. We can set $\mathbf{m}_1 = 1/\sqrt{n}$ and $\mathbf{m}_2 = \mathbf{t}_*$, where $\mathbf{t}_* = (\mathbf{t} - \bar{t}\mathbf{1})/S_t$, $\bar{t}$ is the mean and $S_t = \sqrt{\sum_{i=1}^{q}(t_i - \bar{t})^2}$ is the square root of the sum of squares of the time points $t_1, \ldots, t_q$. Therefore the first two eigenvectors $\mathbf{m}_1$ and $\mathbf{m}_2$ span the subspace corresponding to the straight line model. In the mixed model formulation of the spline solution (e.g. Verbyla et al. 1999) this corresponds to the fixed part of the model. It is also easily observed that if the value of the smoothing parameter α increases the fit approaches the straight line model and the fitted line (fixed part) is not influenced by any specific choice of α.

Clearly, one obvious approximation of the spline fit (3.10) is the spline model

$$\bar{\mathbf{G}} = \mathbf{P}_m\mathbf{Y}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}, \tag{3.15}$$

where $\mathbf{P}_m = \mathbf{M}_*\mathbf{M}_*'$ and $\mathbf{M}_*$ contains the $c(\leq q)$ first eigenvectors of $\mathbf{M}$. This corresponds to minimizing the least squares (LS) criteria

$$Q' = \text{tr}(\mathbf{Y} - \dot{\mathbf{G}})'(\mathbf{Y} - \dot{\mathbf{G}}), \tag{3.16}$$

where $\dot{\mathbf{G}} = \mathbf{GA}'$. Note that the smoother matrix $\mathbf{S}$ and the smoothing parameter need not be computed here. However, the number of eigenvectors c from $\mathbf{K}$ used in the approximation needs to be estimated. This is easily done by, for example, using a modified generalized cross-validation criteria

$$GCV^*(c) = \frac{\frac{1}{nq}\sum_{i=i}^{nq}[y_i - \bar{y}_i]^2}{(1 - \frac{m \times c}{nq})^2},$$

(3.17)

where $\bar{y}_i$ is now computed using the formula (3.11) with $\mathbf{S}$ replaced by $\mathbf{P}_m$.

3.3.2 Constructing a Test for Mean Spline Curves

First, consider the set of fitted spline curves

$$\hat{\mathbf{Y}} = \hat{\mathbf{G}}\mathbf{A}'.$$

(3.18)

As discussed in the previous section we may use the approximation

$$\hat{\mathbf{Y}} = \bar{\mathbf{G}}\mathbf{A}' = \mathbf{M}_*\hat{\Omega}\mathbf{A}',$$

(3.19)

where we denoted $\hat{\Omega} = \mathbf{M}'_*\mathbf{YA}(\mathbf{A}'\mathbf{A})^{-1}$. All the relevant information for testing mean profiles is now in the matrix $\hat{\Omega}$, which can now be considered to be an unbiased estimate of the unknown parameter matrix of the statistical model $E(\mathbf{Y}) = \mathbf{M}_*\Omega\mathbf{A}'$. Therefore in sequel we confine in testing linear hypothesis of the form

$$H_0 : \mathbf{C}\Omega\mathbf{D} = \mathbf{0},$$

where $\mathbf{C}$ and $\mathbf{D}$ are known $v \times c$ and $m \times g$ matrices with ranks v and g, respectively. Since $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\mathrm{vec}(\Omega)$, the vector form of H_0 is given by

$$H_0 : (\mathbf{D}' \otimes \mathbf{D})\omega = \mathbf{0},$$

where $\omega = \mathrm{vec}(\Omega)$. If we take the vector form of $\hat{\Omega}$, we get

$$\hat{\omega} = \mathrm{vec}(\hat{\Omega}) = [(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \otimes \mathbf{M}'_*]\mathbf{y}.$$

(3.20)

It is now easily seen that the covariance matrix of $\hat{\omega}$ is

$$Var(\hat{\omega}) = \sigma^2[(\mathbf{A}'\mathbf{A})^{-1} \otimes \mathbf{M}'_*\mathbf{RM}_*].$$

(3.21)

If we denote $Var(\mathbf{D}' \otimes \mathbf{C})\hat{\omega} = \mathbf{W}$, it is then obvious that under the null hypothesis

$$\mathbf{W}^{-1/2}(\mathbf{D}' \otimes \mathbf{C})\hat{\omega} \sim N_{vg}(\mathbf{0}, \mathbf{I})$$

and

$$\sigma^{-2} Q_* = \hat{\omega}'(\mathbf{D} \otimes \mathbf{C}')\mathbf{W}^{-1}(\mathbf{D}' \otimes \mathbf{C})\hat{\omega} \sim \chi^2_{vg}.$$

By using the results $\mathrm{tr}(\mathbf{AZ}'\mathbf{BZC}) = (\mathrm{vec}\, \mathbf{Z})'(\mathbf{CA} \otimes \mathbf{B})\mathrm{vec}\, \mathbf{Z}$, it is further easy to see that Q_* can be rewritten as

$$Q_* = \mathrm{tr}\{[\mathbf{D}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{D}]^{-1}[\mathbf{C}\hat{\Omega}\mathbf{D}]'[\mathbf{C}\mathbf{M}'_*\mathbf{R}\mathbf{M}_*\mathbf{C}']^{-1}[\mathbf{C}\hat{\Omega}\mathbf{D}]. \qquad (3.22)$$

If σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{1}{n(q-c)}\mathrm{tr}\mathbf{Y}'(\mathbf{I} - \mathbf{P}_m)\mathbf{Y}, \qquad (3.23)$$

it can be shown that $n(q-c) \times \hat{\sigma}^2 \sim \chi^2_{n(q-c)}$ and since Q_* and $\hat{\sigma}^2$ are independent testing can be based on the F-ratio. Then under the null hypothesis

$$F = \frac{Q_*/vg}{\hat{\sigma}^2} \sim F[vg, n(q-c)]. \qquad (3.24)$$

Testing can then be based on the quantiles of the F-distribution. However, in practical situations the matrix $\mathbf{R}$ contains unknown parameters that need to be estimated and therefore the distribution of F in general case is only approximate. However, if we are only interested in progression in time we can drop the first eigenvector $\mathbf{m}_1$ corresponding to the constant term in the approximation model (see Sect. 3.3.1). Therefore we can take $\mathbf{C} = [\mathbf{0}, \mathbf{I}]$, and if we assume the uniform covariance model $\mathbf{R} = d^2\mathbf{1}\mathbf{1}' + \mathbf{I}$, it can be shown that

$$\mathbf{C}\mathbf{M}'_*\mathbf{R}\mathbf{M}_*\mathbf{C}' = \mathbf{C}\{d^2\mathbf{e}_1\mathbf{e}'_1 + \mathbf{I}\}\mathbf{C}' = \mathbf{I}, \qquad (3.25)$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)'$. Therefore the term Q_* simplifies to

$$Q_* = \mathrm{tr}[\mathbf{C}\hat{\Omega}\mathbf{D}]\{[\mathbf{D}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{D}]^{-1}[\mathbf{C}\hat{\Omega}\mathbf{D}]', \qquad (3.26)$$

which does not contain unknown parameters of the covariance matrix and therefore for this special case the distribution of the F-statistic is exact. This is an important result since the uniform covariance model is quite common and a good approximation in many situations. The F-test proposed here provides means to test if the progression in time is the same over treatment groups when the models are based on spline curves. Following the same kind of considerations it would be easy to develop an exact F-statistic to test if the progression around the fitted straight line (the so-called random part in mixed model formulation) is the same over treatment groups with the more general assumption of linear correlation structure $\mathbf{R} = \mathbf{XDX}' + \mathbf{I}$.

3.4   Computational Examples

3.4.1   Standard Glucose Tolerance Test

As the first computational example we consider the glucose data of Zerbe (1979).
In these data glucose tolerance tests were administered to 13 control and 20 obese
patients. Plasma inorganic phosphate measurements determined from blood samples
drawn 0, 0.5, 1, 1.5, 2, 3, 4, and 5 h after standard oral glucose dose were taken.
The curves plotted for the control and obese patients are plotted in Fig. 3.1. In
Fig. 3.1, two features of the plotted curves are quite obvious. First, there is a
considerable variation in patient's individual levels. Secondly, the functional form
of the dependency of plasma inorganic phosphate and time is quite complicated and
possibly different for control and obese patients. In Zerbe (1979) a polynomial of
degree of 4 was used to model this relationship.

To set up the spline growth model the between-individual design matrix $\mathbf{A}$ was
first defined. For 13 control patients the rows of $\mathbf{A}$ are $(1, 0)$, $i = 1, \ldots, 13$ and for
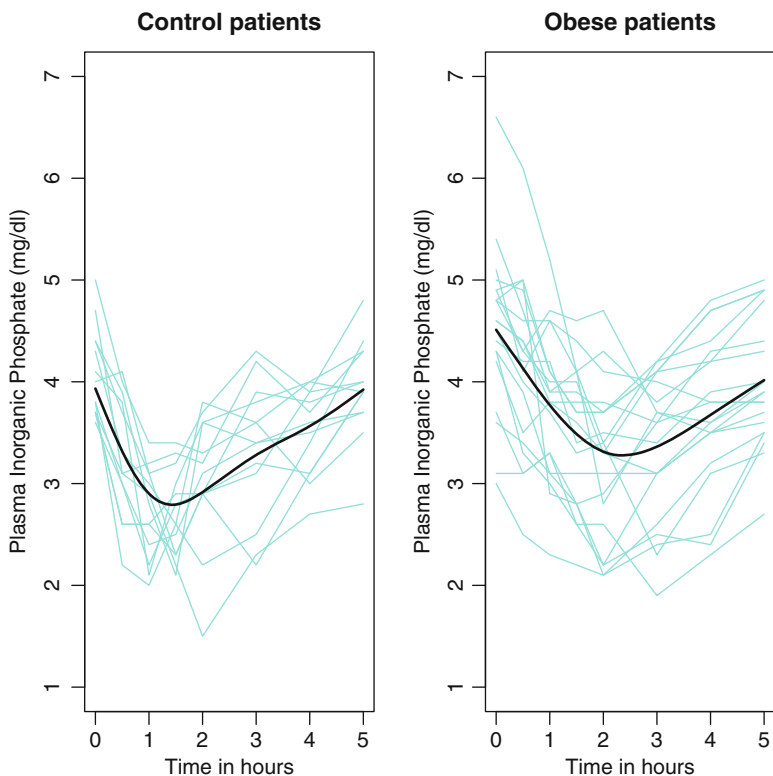

Fig. 3.1 Plasma inorganic phosphate measurements for control and obese patients

20 obese patients the rows of $\mathbf{A}$ are $(0, 1)$, $i = 14, \ldots, 33$. The minimum value of $GCV = 0.4484152$ is obtained at $\alpha = 0.09410597$. This gives the total effective degrees of freedom $edf_* = 9.310273$. The fitted curves are plotted in Fig. 3.1. It can be observed that the fitted spline curves very nicely depict the mean performance of measurements in both groups.

To test if the progression in time is the same in both groups we first determined the dimension c needed in the spline approximation. Minimizing the modified generalized cross-validation criterion gives $c = 5$. To test the null hypothesis we took

$$\mathbf{C} = [\mathbf{0}, \mathbf{I}_4] \quad \text{and} \quad \mathbf{D} = [1, -1]'.$$

Next we calculated the estimate $\mathbf{C}\hat{\Omega}\mathbf{D}$. This yields

$$\mathbf{C}\hat{\Omega}\mathbf{D} = (0.74897422, \ -0.03613939, \ 0.46201190, \ 0.54998030)'$$

and the residual variance estimate for this setup is $\hat{\sigma}^2 = 0.09408348$. For the covariance matrix $\mathbf{R}$ we assumed the uniform correlation model and therefore the exact version of the test statistics can be used. Then the value of Q_* is given as

$$Q_* = \text{tr}[\mathbf{C}\hat{\Omega}\mathbf{D}]\{[\mathbf{D}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{D}]^{-1}[\mathbf{C}\hat{\Omega}\mathbf{D}]' = 8.494923$$

and the value of the test statistics is then

$$F = \frac{Q_*/\nu g}{\hat{\sigma}^2} = \frac{8.494923/4}{0.09408348} = 22.57283.$$

If this is compared to the critical value $F_{0.95}(4, 99) = 2.447$, the null hypothesis of equal progression in mean plasma inorganic phosphate for control and obese patients is clearly rejected.

3.4.2   A Simulation Study

In order to demonstrate the advances of the methodology presented we conducted a simulation study. In this study two models were tested

$$y = 1 + 0.5 \times t + \epsilon, \tag{3.27}$$

$$y = 1 + 0.5 \times t + a \times \cos(0.4\pi t) + \epsilon, \tag{3.28}$$

with $t = 1, \ldots, 10$ and independent random errors $\epsilon_i \sim N(0, 1)$. The coefficient a takes the values 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7. The first group of growth curves consists of 100 random vectors generated from model 3.27 and the second consists of 100 random vectors generated from model 3.28. So, for each value
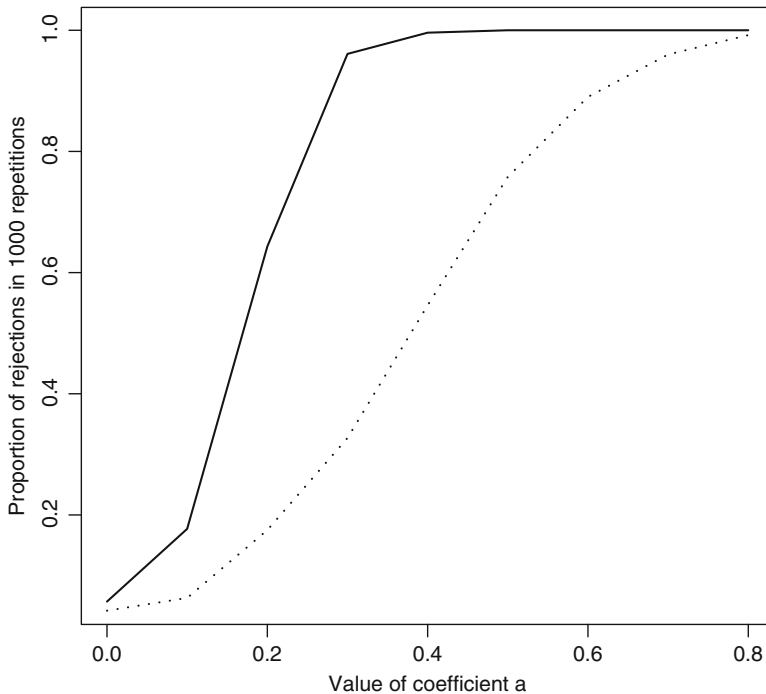
Fig. 3.2 Proportion of the rejections in 1,000 repetitions under models (3.27) and (3.28)

of *a* these two sets of growth curves were generated. The mean growth curves are then tested against the null hypothesis that the progression in time is the same in both groups. Two methods were utilized. The spline testing method presented in this paper and the second method utilized here was the basic parametric least squares fit of the third degree polynomial model. The power was estimated with the significance level 0.05 by counting the percentage of rejections in the 1,000 repetitions.

The results are shown in Fig. 3.2. Clearly, the spline test presented in this paper performed better than the test based on the least squares fit of the third degree polynomial. This is obviously due to the fact that the fit provided by the splines better depicts the peculiarities of the unknown growth function.

3.5 Concluding Remarks

Traditional analyses of growth curves are often based on simple parametric curves, which may not satisfactorily depict all the features of the growth process during the testing period. The method presented in this paper is based on cubic smoothing

splines, which provides a very flexible modeling tool for the analysis. However, very little research on the statistical inference (especially testing) of cubic smoothing splines for correlated data has been carried out. The novel test presented in this paper seems to provide a good alternative, especially when more accurate modeling of growth process is required.

References

Azzalini, A., (1987). Growth Curves analysis for patterned covariance matrices. In: New perspectives in theoretical and applied statistics (eds puri, M., Vilaplana, J.P. & Wertz, W) New York: Wiley, 63–70

Cantoni, E., & Hastie, T. (2002). Degrees-of-freedom tests for smoothing splines. *Biometrika*, *89*(2), 251–263.

Eubank, R.L., & Spiegelman, C.H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association, 85*(410), 387–392.

Green, P.J., & Silverman, B.W. (1994). *Nonparametric regression and generalized linear models*. London: Chapman and Hall.

Grizzle, J.E., & Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrics, 25*, 357–381.

Hastie, T. (1996). Pseudosplines. *Jounal of the Royal Statistical Society, Series B*, *58*, 379–396.

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics, 18*, 75–86.

Kshirsagar, A.M., & Smith, W.B. (1995). *Growth curves*. New York: Marcel Dekker.

Lee, J.C. (1988). Tests and model solution for general growth curve model. Biometrics, *47*, 147–159.

Liu, A., & Wang, Y. (2004). Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation*, *74*, 581–597.

Nummi, T. (1997). Estimation in random effects growth curve model. *Journal of Applied Statistics, 24*(2), 157–168.

Nummi, T., & Koskela, L. (2008). Analysis of growth curve data using cubic smoothing splines. *Journal of Applied Statistics*, *35*, 1–11.

Nummi, T., Jianxin, P., Siren, T., & Liu, K. (2011). Testing for cubic smoothing splines under dependent data. *Biometrics*, *67*(3), 871–875. DOI: 10.1111/j.1541-0420.2010.01537.x.

Pan, J., & Fang, K. (2002). *Growth curve models and statistical diagnostics*, Springer series in Statistics. New York: Springer.

Potthoff, R.F., & Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 5*, 313–326.

Schimek, M.G. (2000). Estimation and Inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference*, *91*, 525–540.

Verbyla, A.P., Cullis, B.R., Kenward, M.G., & Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Journal of the Royal Statistical Society, Series C*, *48*, 269–311.

Wu, L., & Zhang, J.T. (2006). *Nonparametric regression methods for longitudinal data analysis*. New Jersey: Wiley.

Zerbe, G.O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association*, *74*, 215–221.