

## Testing for Constancy in Varying Coefficient Models

Peer-reviewed author version

AHKIM, Mohamed & VERHASSELT, Anneleen (2018) Testing for Constancy in Varying Coefficient Models. In: Communications in statistics. Theory and methods, 47(4), p. 890-911..

DOI: 10.1080/03610926.2017.1300271

Handle: <http://hdl.handle.net/1942/23406>

# Testing for Constancy in Varying Coefficient Models

M. Ahkim<sup>1</sup> and A. Verhasselt<sup>2</sup>

1. Department of Mathematics - Computer Sciences, Universiteit Antwerpen, Belgium
2. Censtat, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Universiteit Hasselt, Belgium

November 30, 2016

## 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 (for example time), the covariates are also allowed to depend on other variables. Varying coefficient models are popular in longitudinal data and panel data studies, and have been applied in fields such as finance and health sciences. We consider longitudinal data and estimate the coefficient functions by the flexible B-spline technique. An important question in a varying coefficient model is whether an estimated coefficient function is statistically different from a constant (or zero). We develop testing procedures based on the estimated B-spline coefficients by making use of nice properties of a B-spline basis. Our method allows longitudinal data where repeated measurements for an individual can be correlated. We obtain the asymptotic null distribution of the test statistic. The power of the proposed testing procedures are illustrated on simulated data where we highlight the importance of including the correlation structure of the response variable and on real data.

**Key words:** varying coefficient models; hypothesis testing; quadratic forms.

## 1 Introduction

In this paper, we consider varying coefficient models (Hastie and Tibshirani (1993)) 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  $t$ ). More precisely, we study a varying coefficient model of the form

$$Y(t) = \mathbf{X}(t)' \boldsymbol{\beta}(t) + \varepsilon(t) = \beta_0(t) + \sum_{p=1}^d X^{(p)}(t) \beta_p(t) + \varepsilon(t), \quad (1)$$

where  $Y(t)$  is the response at time  $t \in \mathcal{T} = [0, T]$ ;  $\mathbf{X}(t) = (X^{(0)}(t), \dots, X^{(d)}(t))'$  is the covariate column-vector at time  $t$ , with  $X^{(0)}(t) \equiv 1$ ;  $\boldsymbol{\beta}(t) = (\beta_0(t), \dots, \beta_d(t))'$  is the column-vector of coefficients at time  $t$ . We assume longitudinal data, that is, each of the  $n$  independently chosen subjects is measured repeatedly over a time period in  $\mathcal{T}$ . The  $j$ -th measurement of subject  $i$  for  $1 \leq i \leq n$  and  $1 \leq j \leq N_i$ , is a sample from  $(t, Y(t), \mathbf{X}(t))$ , which is denoted by  $(t_{ij}, Y_{ij}, \mathbf{X}_{ij})$ , where  $t_{ij}$  is the observed time,  $Y_{ij}$  is the observed response of the  $i$ -th individual at time  $t_{ij}$  and  $\mathbf{X}_{ij} = (X_{ij}^{(0)}, \dots, X_{ij}^{(d)})'$  is the corresponding observed covariate vector, satisfying

$$Y_{ij} = \mathbf{X}_{ij}' \boldsymbol{\beta}(t_{ij}) + \varepsilon_{ij}. \quad (2)$$

The total number of observations is  $N = \sum_{i=1}^n N_i$ .

There are several nonparametric techniques to estimate the coefficient curves  $\beta_0(t), \dots, \beta_d(t)$  for longitudinal data. The local polynomial technique and the smoothing spline method are discussed in Hoover et al. (1998), Fan and Zhang (1999) developed a two step local polynomial estimation procedure. Huang et al. (2002) elaborated on an estimating method where coefficient functions are approximated by B-spline basis expansions and Antoniadis et al. (2012) consider P-spline estimation of the varying coefficients. We use the flexible B-spline estimation method as described in Huang et al. (2002).

In the context of model (1), it is important to know whether certain coefficient functions are constant. Since this means that a covariate does not have a varying but a constant effect on the response and a simpler model can be used instead. The purpose of this paper is to establish a method to test these kind of hypotheses for longitudinal data models. For an overview of this question in non-longitudinal data models we refer to for example Li et al. (2011) and the references therein. Huang et al. (2002) constructed a test statistic based on the difference of the residual sum of squares under the null (coefficient is constant) and the alternative hypothesis (coefficient is varying), but do not acquire asymptotical results of their approach. They obtain critical values via a bootstrap strategy, which imposes the need of a relatively large sample size at a high computational cost.

Our method extends the technique by Li et al. (2011) for non-longitudinal data to longitudinal data with correlated error structures, where the coefficient functions are estimated based on a B-spline basis expansion. Their approach makes fully use of the estimated coefficients and the nice properties of B-splines. A main advantage of this approach, besides its simplicity and high power, is that it can be extended to other interesting hypotheses.

The test statistic follows (asymptotically if the coefficient functions are not spline functions) a Fisher distribution. The main difficulty is incorporating the weight matrix (when longitudinal data are used) and the correlation structure of the errors. The novelty of our approach furnishes this issue. We prove that our test statistic follows asymptotically a generalized Fisher (notation:  $F$ ) distribution. This generalized  $F$ -distribution is the exact null distribution if the coefficient functions are splines.

In Section 2 we describe the B-spline estimator. In Section 3 the testing procedure and asymptotic results are presented. The proofs can be found in the Appendix. The performance of our method compared to Huang et al. (2002) are illustrated with numerical simulations in Section 4 and real data applications are discussed in Section 5. Finally, we end with a conclusion in Section 6. When there is no information on the correlation matrix we propose a bootstrap approach, see Section 6. Finally, we end with a conclusion in Section 7.

## 2 Spline estimation

### 2.1 B-spline estimator

In this section we briefly recall the B-spline estimator in varying coefficient models, see Huang et al. (2004). The assumption is that each component of  $\boldsymbol{\beta}(t) = (\beta_0(t), \dots, \beta_d(t))'$  can be approximated by a B-spline basis expansion, i.e., for each  $p = 0, \dots, d$ ,  $\beta_p(t) \approx \sum_{l=1}^{m_p} \alpha_{pl} B_{pl}(t; q_p)$ , where  $\{B_{pl}(\cdot; q_p) : l = 1, \dots, K_p + q_p = m_p\}$  is the normalized (i.e.  $\sum_{j=1}^{m_p} B_j(\cdot; q_p) = 1$ )  $q_p$ -th degree B-spline basis with  $K_p + 1$  equidistant knots  $\xi_{p0}, \xi_{p1}, \dots, \xi_{pK_p}$  in  $\mathcal{T}$ . Let  $\mathbb{G}_p$  denote the space spanned by this basis.

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

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

where  $w_i$  denote the weight for subject  $i$ , often  $w_i = \frac{1}{N_i}$  is used. More compactly written, we solve

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{U}_i \boldsymbol{\alpha})' \mathbf{W}_i (\mathbf{Y}_i - \mathbf{U}_i \boldsymbol{\alpha}), \quad (3)$$

where

$$\begin{aligned}
\mathbf{Y}_i &= (Y_{i1}, \dots, Y_{iN_i})'; & \mathbf{Y} &= (\mathbf{Y}_1, \dots, \mathbf{Y}_n)'; \\
\mathbf{U}'_{ij} &= \mathbf{X}'_{ij} \mathbf{B}(t_{ij}) \in \mathbb{R}^{1 \times \dim}; \\
\mathbf{U}_i &= (\mathbf{U}_{i1}, \dots, \mathbf{U}_{iN_i})' \in \mathbb{R}^{N_i \times \dim}, & \text{where } \dim &= \sum_{p=0}^d m_p; \\
\mathbf{U} &= (\mathbf{U}_1, \dots, \mathbf{U}_n)' \in \mathbb{R}^{N \times \dim}; \\
\mathbf{B}(t) &= \begin{pmatrix} B_{01}(t; q_0) & \dots & B_{0m_0}(t; q_0) & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \ddots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & B_{d1}(t; q_d) & \dots & B_{dm_d}(t, q_d) \end{pmatrix} \in \mathbb{R}^{(d+1) \times \dim}; \\
\mathbf{W}_i &= \text{diag}(w_i, \dots, w_i) \in \mathbb{R}^{N_i \times N_i} & \text{(a diagonal matrix with } N_i \text{ times} \\ & & w_i \text{ on the diagonal);} \\
\mathbf{W} &= \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n) \in \mathbb{R}^{N \times N} & \text{(a block diagonal matrix} \\ & & \text{with the matrices } \mathbf{W}_i \text{ on the diagonal).}
\end{aligned}$$

If  $(\mathbf{U}'\mathbf{W}\mathbf{U})$  is invertible, then (3) has a unique solution

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

Huang et al. (2004) proved that under Assumption 1 in Appendix B, the matrix  $(\mathbf{U}'\mathbf{W}\mathbf{U})$  is invertible with probability tending to 1. Throughout we assume Assumption 1. Then, the B-spline estimator of  $\boldsymbol{\beta}(t)$  is

$$\hat{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\hat{\boldsymbol{\alpha}} = (\hat{\beta}_0(t), \dots, \hat{\beta}_d(t))', \quad \text{with } \hat{\beta}_p(t) = \sum_{l=1}^{m_p} \hat{\alpha}_{pl} B_{pl}(t; q_p).$$

By Theorem 1 in Huang et al. (2004), the estimator  $\hat{\boldsymbol{\beta}}$  is consistent under Assumption 1 and if  $\lim_n K_n \log(K_n)/n = 0$ , where  $K_n = \max_{p=0, \dots, K_p} K_p$ .

## 2.2 Some properties of spline approximations

The motivation for our test statistics are based on the following nice properties of B-spline approximations.

Suppose that the function  $\beta_p(t)$  is a constant  $c_p$ , then  $\beta_p(t) = c_p = \sum_{l=1}^{m_p} \alpha_{pl} B_{pl}(t; q_p) \in \mathbb{G}_p$ . Indeed, this holds since constant functions on  $\mathcal{T}$  are contained in  $\mathbb{G}_p$ . Moreover, we have that  $\boldsymbol{\alpha}_p = (c_p, \dots, c_p)' \in \mathbb{R}^{m_p \times 1}$ , since normalized B-splines are used and the functions  $B_{pl}(t; q_p)$  ( $l = 1, \dots, m_p$ ) form a basis of  $\mathbb{G}_p$ . Therefore there is a one-to-one relation between the constancy of the function  $\beta_p$  and the vector of B-spline coefficients  $\boldsymbol{\alpha}_p$  being constant.

### 3 Testing constancy of coefficient functions

In this section we consider the problem of testing whether the  $p$ -th coefficient  $\beta_p(t)$  of a varying coefficient model is really varying. We develop a testing procedure to test for constancy, i.e. test

$$H_0 : \beta_p \text{ is a constant function} \quad \text{versus} \quad \neg H_0 : \beta_p \text{ is not a constant function.} \quad (4)$$

Li et al. (2011)'s technique is based on the vector of first order differences  $\mathbf{D}_1 \hat{\boldsymbol{\alpha}}_p$  where

$$\mathbf{D}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \dots & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \dots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(m_p-1) \times m_p}.$$

They use linear splines. However, it should be noted that splines of any degree could be used (as noted in Section 2.2) which allows to generalize their asymptotic results. We extend their approach to our longitudinal data model with correlated errors where the coefficient functions are estimated by B-splines of any degree.

We first give a test for the more restrictive hypothesis that all coefficient functions are constant:

$$H_1 : \beta_p(\cdot) \text{ is a constant function for } p = 0, \dots, d \quad \text{versus} \quad \neg H_1. \quad (5)$$

Under hypothesis  $H_1$  all coefficient functions are modeled by spline functions. Let us consider the model where all coefficient functions are spline functions, i.e.  $\beta_p(t) = \sum_{l=1}^{m_p} \alpha_{pl} B_{pl}(t; q_p)$  and

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

with  $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$ ,  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_n)'$  and  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iN_i})'$  for  $i = 1, \dots, n$ . For this model, testing problem (5) is equivalent to

$$H_1^* : \mathbf{L}'_1 \boldsymbol{\alpha} = \mathbf{0} \quad \text{versus} \quad \neg H_1^* : \mathbf{L}'_1 \boldsymbol{\alpha} \neq \mathbf{0},$$

where

$$\mathbf{L}'_1 = \begin{pmatrix} \mathbf{D}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{D}_1 \end{pmatrix} \in \mathbb{R}^{(dim-d-1) \times dim}.$$

#### 3.1 Construction of the test statistic

Let us assume that  $\text{Cov}(\boldsymbol{\varepsilon}_i) = \sigma^2 \mathbf{V}_i$  and  $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{V}$ , where  $\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_n)$  and  $\mathbf{V}_i$  are correlation matrices. Recall the B-spline estimator of model (1)

$$\hat{\boldsymbol{\alpha}} = (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'\mathbf{Y},$$

where  $\underline{\mathbf{U}} = \mathbf{W}^{\frac{1}{2}}\mathbf{U}$  and  $\underline{\mathbf{Y}} = \mathbf{W}^{\frac{1}{2}}\mathbf{Y}$ . Let  $P_{\underline{\mathbf{U}}} = \underline{\mathbf{U}}(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'$  and  $\tilde{\mathbf{Y}} = \mathbb{E}(\mathbf{Y}|\mathcal{X})$ , where  $\mathcal{X} = \{(\mathbf{X}_{ij}, t_{ij}); i = 1, \dots, n, j = 1, \dots, N_i\}$ . Throughout the remainder of the paper we condition on  $\mathcal{X}$ . Let  $\underline{\mathbf{V}} = \mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}$ .

### Testing hypothesis (5)

*Remark 1.* If we would use Li et al. (2011)'s test statistic in our longitudinal case, then we would obtain the “test statistic”

$$\frac{\dim - d - 1}{N - \dim} \frac{\hat{\boldsymbol{\alpha}}' \mathbf{L}_1 (\mathbf{L}'_1 (\underline{\mathbf{U}}' \underline{\mathbf{U}})^{-1} \mathbf{L}_1)^{-1} \mathbf{L}_1 \hat{\boldsymbol{\alpha}}}{\underline{\mathbf{Y}}' (\mathbf{I}_N - P_{\underline{\mathbf{U}}}) \underline{\mathbf{Y}}}, \quad (6)$$

where  $\mathbf{I}_N$  is the identity matrix of dimension  $N \times N$ . Unlike the case in Li et al. (2011) where the statistic (6), forms a ratio of two *independent*  $\chi^2$  variables, we here have a ratio of two *dependent*  $\chi^2$  variables (see the proof of Theorem 1). The novelty of our approach allows to incorporate the weight matrix  $\mathbf{W}$  and the correlation matrix  $\mathbf{V}$  arising from longitudinal data models. This generalization is not straightforward.

Our test statistic is based on the fact that under hypothesis (5)

$$\mathbb{E}(\underline{\mathbf{L}}'_1 \hat{\boldsymbol{\alpha}} | \mathcal{X}) = \mathbb{E}(\underline{\mathbf{L}}'_1 \hat{\boldsymbol{\alpha}} | \mathcal{X}) = \mathbf{0},$$

where  $\underline{\mathbf{L}}'_1 = \mathbf{L}'_1 (\underline{\mathbf{U}}' \underline{\mathbf{V}}^{-1} \underline{\mathbf{U}})^{-1} (\underline{\mathbf{U}}' \underline{\mathbf{U}})$ ,  $\hat{\boldsymbol{\alpha}} := (\underline{\mathbf{U}}' \underline{\mathbf{U}})^{-1} \underline{\mathbf{U}}' \underline{\mathbf{V}}^{-1} \underline{\mathbf{Y}}$ . Note that  $\hat{\boldsymbol{\alpha}}$  and  $\underline{\mathbf{L}}'_1$  are precisely introduced to obtain a ratio of independent quadratic forms of which the distribution is known, see below.

Since  $\underline{\mathbf{Y}} \sim N(\mathbf{W}^{\frac{1}{2}}\tilde{\mathbf{Y}}, \sigma^2 \underline{\mathbf{V}})$ , we have that  $\hat{\boldsymbol{\alpha}} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma})$ , where

$$\boldsymbol{\mu} = (\underline{\mathbf{U}}' \underline{\mathbf{U}})^{-1} \underline{\mathbf{U}}' \underline{\mathbf{V}}^{-1} \mathbf{W}^{1/2} \tilde{\mathbf{Y}} \quad \text{and} \quad \boldsymbol{\Sigma} = (\underline{\mathbf{U}}' \underline{\mathbf{U}})^{-1} \underline{\mathbf{U}}' \underline{\mathbf{V}}^{-1} \underline{\mathbf{U}} (\underline{\mathbf{U}}' \underline{\mathbf{U}})^{-1}.$$

Next we define two quadratic forms in normal variables. The first is

$$Q_1 = \underline{\mathbf{Y}}' (\mathbf{I}_N - P_{\underline{\mathbf{U}}}) \underline{\mathbf{Y}},$$

the second

$$Q_2 = \hat{\boldsymbol{\alpha}}' \underline{\mathbf{L}}_1 (\underline{\mathbf{L}}'_1 \boldsymbol{\Sigma} \underline{\mathbf{L}}_1)^{-1} \underline{\mathbf{L}}'_1 \hat{\boldsymbol{\alpha}}.$$

Our test statistic for hypothesis (5) is a ratio of these (stochastic) quadratic forms, namely

$$T_1 = \frac{\dim - d - 1}{N - \dim} \frac{Q_1}{Q_2}.$$

Note that  $T_1$  does not depend on the marginal variance  $\sigma^2$ . Theorem 1 states the exact null distribution of  $T_1$ . When  $t_1 \in \mathbb{R}$  is a realization of  $T_1$ , the  $p$ -value  $p_1$  to test  $H_1$  is defined to be

$$p_1 = F_{T_1}(t_1), \quad (7)$$

where  $F_{T_1}$  is the null distribution function of  $T_1$ , since  $Q_2$  is relatively small under the null hypothesis. By Theorem 1 we know that under the null hypothesis in (5),  $T_1$  follows a generalized  $F$ -distribution of the type

$$\frac{(\sum_{i=1}^l c_i X_i) / (\sum_{i=1}^l m_i)}{Y/n},$$

where the components of  $(X_1, \dots, X_l, Y)$  are independent with  $X_i \sim \chi^2(m_i)$ ,  $Y \sim \chi^2(n)$  and all  $c_i > 0$ . Dunkl and Ramirez (2001) gave exact and numerically tractable expressions of the cumulative distribution function for this kind of generalized  $F$ -distributions. We have implemented the cumulative distribution function of this distribution in Matlab.

Note that in the case we have data without repeated measurements, i.e. the matrices  $\mathbf{W}$  and  $\mathbf{V}$  are the identity matrix, the test statistic

$$\frac{N - \dim}{\dim - d - 1} \frac{Q_2}{Q_1}$$

is exactly the test statistic used by Li et al. (2011) (see (6) in Remark 1) and follows the  $F$ -distribution with degrees of freedom  $(\dim - d - 1, N - \dim)$  under the null hypothesis in (5).

**Testing hypothesis (4)** Let us return to hypothesis (4):

$$H_0 : \beta_p \text{ is a constant function} \quad \text{versus} \quad \neg H_0 : \beta_p \text{ is not a constant function.}$$

This time we are only interested in the coefficients belonging to the coefficient function  $\beta_p(\cdot)$ , therefore we apply the following transformation on  $\hat{\alpha}$

$$\mathbf{L}'_2 = \begin{pmatrix} \mathbf{0} & \dots & \mathbf{D}_1 & \dots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(m_p-1) \times \dim}.$$

The hypothesis in terms of the B-spline coefficients becomes

$$H_0^* : \mathbf{L}'_2 \boldsymbol{\alpha} = \mathbf{0} \quad \text{versus} \quad \neg H_0^* : \mathbf{L}'_2 \boldsymbol{\alpha} \neq \mathbf{0}. \quad (8)$$

The test statistic for hypothesis (4) is

$$T_2 = \frac{m_p - 1}{N - \dim} \frac{\mathbf{Y}'(\mathbf{I}_N - P_{\mathbf{U}})\mathbf{Y}}{\hat{\boldsymbol{\alpha}}' \mathbf{L}'_2 (\mathbf{L}'_2 \boldsymbol{\Sigma} \mathbf{L}'_2)^{-1} \mathbf{L}'_2 \hat{\boldsymbol{\alpha}}},$$

where  $\mathbf{L}'_2 = \mathbf{L}'_2(\mathbf{U}'\mathbf{V}^{-1}\mathbf{U})^{-1}(\mathbf{U}'\mathbf{U})$ . Theorem 2 states that the null distribution function of  $T_2$  denoted by  $F_{T_2}$  is asymptotically equal to the generalized  $F$ -distribution

$$\frac{m_p - 1}{N - \dim} \frac{\sum_{i=1}^k \lambda_i \chi^2(r_i)}{\chi^2(m_p - 1)},$$

with distribution function  $F_2$ , where  $\lambda_i, r_i$  and  $k$  are defined in Theorem 1. Suppose that  $t_2$  is an observed value for  $T_2$ . As in (7), the  $p$ -value is

$$p_2 = F_2(t_2).$$



**Theorem 1.** Assume that  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$  in model (2). If hypothesis  $H_0$  in (5) holds, then  $T_1$  follows the generalized  $F$ -distribution

$$\frac{\dim - d - 1}{N - \dim} \frac{\sum_{i=1}^k \lambda_i \chi^2(r_i)}{\chi^2(\dim - d - 1)},$$

with distribution function  $F_1$ , where  $\lambda_1, \dots, \lambda_k$  denote the nonzero distinct eigenvalues of

$$\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2} (\mathbf{I}_N - P_{\mathbf{U}})$$

with algebraic multiplicities  $r_1, \dots, r_k$  respectively, that satisfy  $\sum_{i=1}^k r_i = N - \dim$ , and where  $\chi^2(r_1), \dots, \chi^2(r_k), \chi^2(\dim - d - 1)$  are mutually independent.

The proof of Theorem 1 is given in Appendix D.

**Theorem 2.** Assume that  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$  in model (2). Define the random variable

$$\frac{m_p - 1}{N - \dim} \frac{\sum_{i=1}^k \lambda_i \chi^2(r_i)}{\chi^2(m_p - 1)},$$

with distribution function  $F_2$ , where  $\lambda_1, \dots, \lambda_k, r_1, \dots, r_k$  are defined in Theorem 1, and where  $\chi^2(r_1), \dots, \chi^2(r_k), \chi^2(m_p - 1)$  are mutually independent. Let  $\|\mathbf{A}\|$  denote the Frobenius norm of a matrix  $\mathbf{A}$ . Under  $H_0$  in (4) and if

$$\sqrt{M_{\xi_0} \left( N \rho_n^2 \|\mathbf{V}^{-1/2}\|^2 + \sqrt{N} \rho_n \|\mathbf{V}^{-1/2}\| \right) + M_{\eta_0} \left( N w_{\max} \rho_n^2 + N w_{\max} \rho_n \sqrt{w_{\max} N_i^{1/2}} \right)} \rightarrow 0, \quad (9)$$

where  $w_{\max} = \max_{i=1, \dots, n} w_i$ ,  $\rho_n$  is the approximation error (see Appendix A),  $M_{\xi_0}$  and  $M_{\eta_0}$  are the maxima of the density function of  $\chi^2(m_p - 1)$  and  $\sum_{i=1}^k \lambda_i \chi^2(r_i)$ , respectively, then

$$\lim_{n \rightarrow \infty} (F_{T_2}(t) - F_2(t)) = 0 \quad \text{uniformly in } t > 0.$$

The proof of Theorem 2 is given in Appendix E. A discussion on condition (9) is given in Appendix F.

**Testing a general hypothesis** The proof of Theorem 2 still holds when we test the general hypothesis

$$H_2^* : \mathbf{A}' \boldsymbol{\alpha} = \mathbf{a} \quad \text{versus} \quad \neg H_2^* : \mathbf{A}' \boldsymbol{\alpha} \neq \mathbf{a}, \quad (10)$$

where  $\mathbf{A}$  is a known fixed nonzero matrix, and  $\mathbf{a}$  is a known fixed vector, see (8). Then, as before we define  $\underline{\mathbf{A}}' = \mathbf{A}' (\underline{\mathbf{U}}' \underline{\mathbf{V}}^{-1} \underline{\mathbf{U}})^{-1} (\underline{\mathbf{U}}' \underline{\mathbf{U}})$ . Let  $r$  denote the number of rows of  $\underline{\mathbf{A}}'$ . The test statistic for hypothesis (10) is

$$T_3 = \frac{r}{N - \dim} \frac{\underline{\mathbf{Y}}' (\mathbf{I}_N - P_{\underline{\mathbf{U}}}) \underline{\mathbf{Y}}}{(\underline{\hat{\boldsymbol{\alpha}}}' \underline{\mathbf{A}} - \mathbf{a}') (\underline{\mathbf{A}}' \underline{\boldsymbol{\Sigma}} \underline{\mathbf{A}})^{-1} (\underline{\mathbf{A}}' \underline{\hat{\boldsymbol{\alpha}}} - \mathbf{a})}.$$

Denote its null distribution function by  $F_{T_3}$ . Define the random variable

$$\frac{r}{N - \dim} \frac{\sum_{i=1}^k \lambda_i \chi^2(r_i)}{\chi^2(r)},$$

with distribution function  $F_3$ , where  $\lambda_1, \dots, \lambda_k, r_1, \dots, r_k$  are defined in Theorem 2, and where  $\chi^2(r_1), \dots, \chi^2(r_k), \chi^2(r)$  are mutually independent. Under the conditions of Theorem 2 we have that

$$\lim_{n \rightarrow \infty} (F_{T_3}(t) - F_3(t)) = 0 \quad \text{uniformly in } t > 0.$$

Suppose  $t_3$  is a realization of  $T_3$ . As in (7), the  $p$ -value is

$$p_3 = F_3(t_3).$$

With (10) we can test several hypotheses on the coefficient functions due to nice properties of B-spline coefficients. For example; by the derivative property of B-splines (De Boor (2001)) we can test whether a coefficient function  $\beta_p$  is a polynomial of degree  $q$ . Then, the matrix  $\mathbf{A}$  is defined as taking the  $q$ -th order differences of the B-spline coefficients corresponding to  $\beta_p$  and  $\mathbf{a} = \mathbf{0}$ .

## 4 Simulation study

Here we discuss a simulation example where we illustrate the performance of our ratio of quadratic forms method (RQF) and compare it with Huang et al. (2002)'s method (Huang). We illustrate the importance of incorporating the correlation structure in the RQF method by also providing the RQF method where independence is assumed; referred to as RQFind.

We let the number of individuals be  $n = 30$ , the number of repeated measurements for individual  $i$  is  $N_i$  randomly generated from  $\{9, \dots, 12\}$  for  $i = 1, \dots, 30$ . For each individual  $i$ , the time points  $t_{ij}$ ,  $j = 1, \dots, N_i$  are equidistant in  $[0, 1]$ . We have a time dependent bivariate vector

$$\begin{pmatrix} X^{(1)}(t) \\ X^{(2)}(t) \end{pmatrix} \sim N(\mathbf{0}, \Sigma_X(t)), \quad \Sigma_X(t) = \begin{pmatrix} \frac{3}{2} & 1/(2+t) \\ 1/(2+t) & 2 \end{pmatrix}.$$

Intra-subject correlation is involved in this example. We consider two types of intra-subject correlated errors, the first is  $V_1$  where

$$\text{Corr}(\varepsilon(t_{ij}), \varepsilon(t_{ik})) = \frac{1}{2} \exp(-|t_{ij} - t_{ik}|), \quad 1 \leq j \neq k \leq N_i,$$

while the error terms of different subjects are mutually independent; the second is an exchangeable correlation structure  $V_2$

$$\text{Corr}(\varepsilon(t_{ij}), \varepsilon(t_{ik})) = 0.6, \quad 1 \leq j \neq k \leq N_i,$$

while the error terms of different subjects are mutually independent.

Furthermore, we use coefficient functions (with domain  $[0, 1]$ ):

$$\beta_0(t) = 0.5(e - e^{-1}) + b_0(e^{2t-1} - 0.5(e - e^{-1})),$$

$$\beta_1(t) = 4/3 + b_1(8t(1 - t) - 4/3) \text{ and } \beta_2(t) = 1 + b_2(2 \sin^2(2\pi t) - 1),$$

where changing the parameters  $b_0, b_1$  and  $b_2$ , change the level of constancy. We measure the performance of our test by varying the deviation of each coefficient function  $\beta_p$  ( $p = 0, 1, 2$ ) from a constant, that is  $b_p$  varies from 0 to 1 while  $b_k = 1$  for  $k \in \{0, 1, 2\} \setminus \{p\}$ .

Note that we have introduced modeling bias since the coefficient functions are not spline functions. The results are based on 200 simulated data sets. In the simulations below, we follow two approaches concerning the choice of the knots. The first approach fixes the knots  $(K_0, K_1, K_2) = (5, 5, 5)$ .

In the second approach we use a cross-validation (CV) method to obtain a desired knot vector  $K$ . Since  $n = 30$ , it is feasible to employ the leave-one-subject-out cross-validation method (Huang et al. (2004) and references therein). The advantage of deleting the whole subject is preserving any intra-subject correlation. We delete subject  $i$  from the original data to obtain the training data which we use to determine the B-spline estimator  $\hat{\boldsymbol{\alpha}}^{-i}$ . This is done for all the subjects  $i = 1, \dots, n$ , so that we can compute the cross-validation score

$$CV(K) = \sum_{i=1}^n \|\mathbf{Y}_i - \mathbf{U}_i \hat{\boldsymbol{\alpha}}^{-i}\|_2^2. \quad (11)$$

The desired  $K$  is the minimizer of (11) where we let  $K$  vary over  $\{5, 6, 7, 8, 9\}^3$ . The degree of the splines is fixed at 3.

The performance of the testing procedure is illustrated by the power, namely the probability  $P(H_0 \text{ is rejected} | \neg H_0)$ , that should be as close as possible to 1. Since we do not know the exact distribution of the test statistic under  $\neg H_0$ , we report an empirical estimate of the power based on simulations, namely we report the percentage of rejecting  $H_0$ . In Figures 1-4 the power functions for each  $b_p \in \{0, 0.1, 0.2, \dots, 1\}$ ,  $p = 0, 1, 2$  are shown. The average computing time for a fixed knot vector of RQF is about 0.6 seconds, while the bootstrap method in Huang et al. (2002) took 22 seconds on average (bootstrap size  $B = 200$ ).

When  $b_p = 0$  the power functions attain approximately the theoretical level of 5%, and increase to 1 when  $b_p$  increases. The RQF method performs better than the method of Huang et al. (2002) in all our examples, except in Figure 1(a) where the method of Huang et al. (2002) performs slightly better. Especially when  $\mathbf{V} = V_2$  (exchangeable correlation) where the differences are bigger. We also note that including the correlation matrix in the

RQF method results in higher power than when using the independence correlation matrix. Moreover, even when the correlation is misspecified in the RQF method (RQFind), the power for  $\beta_1$  and  $\beta_2$  are higher than the method of Huang et al. (2002).

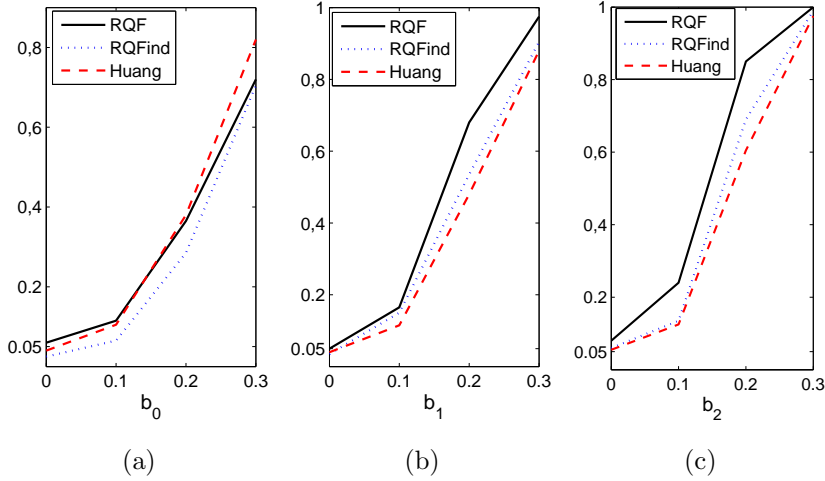


Figure 1: Simulation example ( $n = 30$ ) with  $V = V_1$  and knots  $K = (5, 5, 5)$ . The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dashed line) and Huang (red dotted line) respectively.

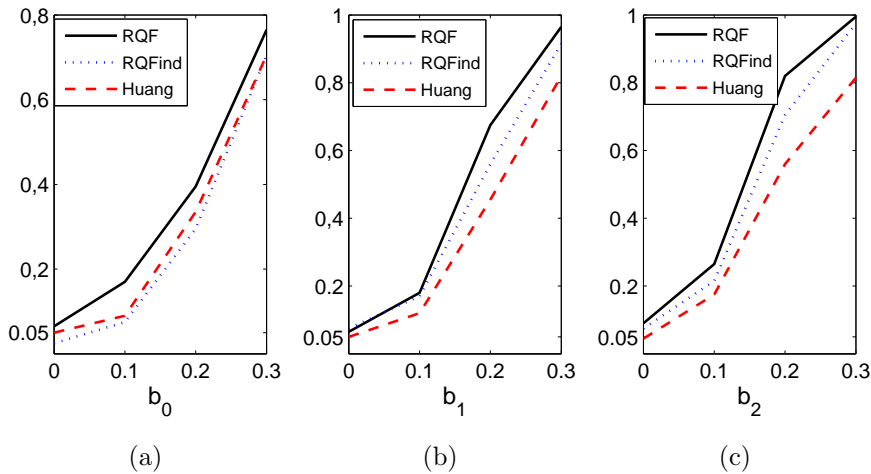


Figure 2: Simulation example ( $n = 30$ ) with  $V = V_1$  and where knots are determined by CV. The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dashed line) and Huang (red dotted line) respectively.

We now consider the same example, but with a bigger sample size. We let  $n = 60$  and the number of repeated measurements  $N_i$  are chosen randomly from  $[18, 24]$ . The RQF method takes on average 12.3 seconds for fixed knots, while the method of Huang et al. (2002) needs 260 seconds on average (bootstrap size  $B=200$ ). As for the knot selection we use leave-10-subjects-out cross-validation where we divide the data in 6 fixed parts (also

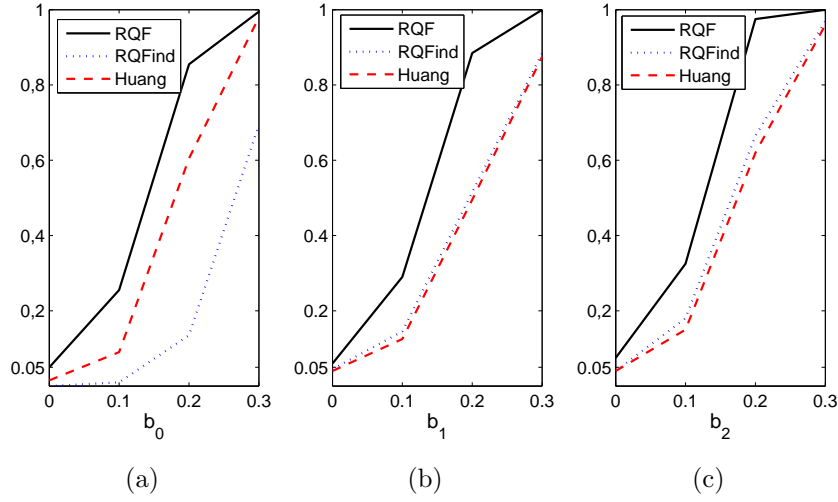


Figure 3: Simulation example ( $n = 30$ ) with  $V = V_2$  and knots  $K = (5, 5, 5)$ . The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dashed line) and Huang (red dotted line) respectively.

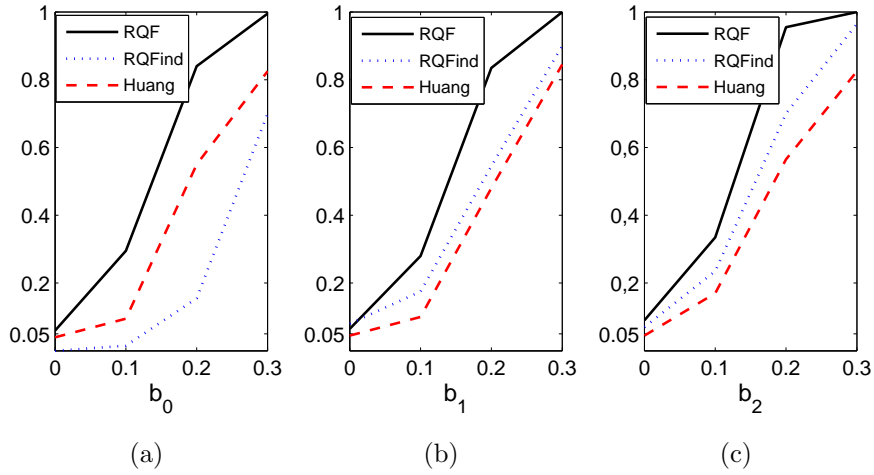


Figure 4: Simulation example ( $n = 30$ ) with  $V = V_2$  and where knots are determined by CV. The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dashed line) and Huang (red dotted line) respectively.

denoted by CV). The results can be found in Figures 5-6. Note that the power increases faster than in the case  $n = 30$ , as could be expected.

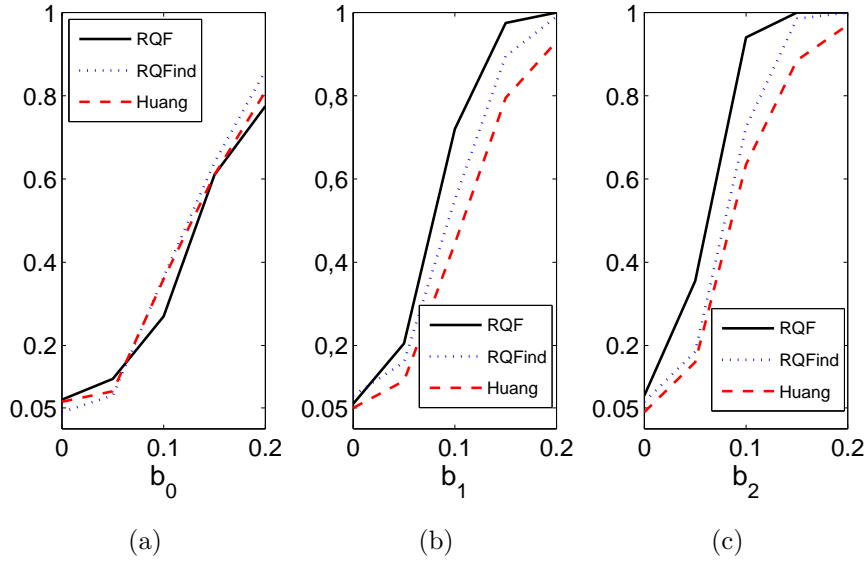


Figure 5: Simulation example ( $n = 60$ ) with  $V = V_1$  and where knots are determined by CV. The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dot line) and Huang (red dash line ) respectively.

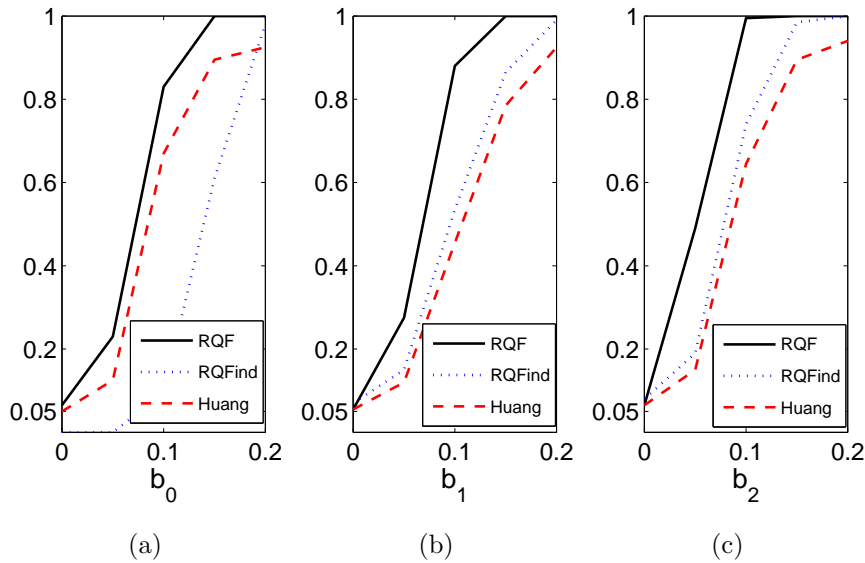


Figure 6: Simulation example ( $n = 60$ ) with  $V = V_2$  and where knots are determined by CV. The power functions for the hypothesis that (a)  $\beta_0$ , (b)  $\beta_1$  and (c)  $\beta_2$  respectively, are constant for RQF (black solid line), RQFind (blue dot line) and Huang (red dash line ) respectively.

## 5 Real data applications

### 5.1 Aids data

We apply our testing methodology to the AIDS data which is a subset of the Multicenter AIDS Cohort Study. This data set contains the repeated measurements of physical exami-

nations, laboratory results and CD4 cell percentages of 283 homosexual men who became HIV-positive between 1984 and 1991. CD4 cells play an important role in the body's immune system. The HIV virus destroys CD4 cells. The fewer functioning CD4 cells, the weaker the immune system and therefore the more vulnerable a person is to infections and illnesses. The patients would have measurements taken every 6 months, but due to certain individual's missing their appointments and the random infection moment, the number of repeated measurements varied per individual. The aim of the statistical analysis is to describe the trend of the mean CD4 percentage depletion over time (in years) explained by the effects of cigarette smoking, age at HIV infection and pre-HIV infection CD4 percentage. For more details about the design, methods and medical applications see Kaslow et al. (1987).

The model is

$$Y_{ij} = \beta_0(t_{ij}) + \beta_1(t_{ij})X_i^{(1)} + \beta_2(t_{ij})X_i^{(2)} + \beta_3(t_{ij})X_i^{(3)} + \varepsilon_{ij},$$

where  $Y_{ij}$  is the  $i$ -th individual CD4 percentage measured at time  $t_{ij}$ ,  $X_i^{(1)}$  is the smoking status of the  $i$ -th individual;  $X_i^{(1)}$  is 1 or 0 if the individual ever or never smoked,  $X_i^{(2)}$  is the  $i$ -th individual's centered age at HIV infection (obtained by subtracting the sample average age at HIV infection from the individual's age at HIV infection), and in a similar way we let  $X_i^{(3)}$  be the  $i$ -th individual's centered pre-HIV infection CD4 percentage. By centering, the baseline CD4 percentage  $\beta_0(t)$  represents the mean CD4 percentage  $t$  years after HIV infection of a homosexual individual with an average age at HIV infection, an average pre-HIV CD4 rate and who has never smoked cigarettes.

In our analysis we use the same B-spline bases as Huang et al. (2002) for the estimation of the coefficient functions, i.e.  $K = (1, 6, 2, 4)$  and  $q = (3, 3, 3, 3)$ .

Let us test whether a linear regression model makes more sense, hence we want to test hypothesis (5). The  $p$ -value for this test is 0.000 for RQFind. Therefore we strongly reject the hypothesis that all coefficient functions are constants. Hence, it makes sense to state a varying coefficient model. In Table 1 the  $p$ -values are presented for testing the constancy of each coefficient function of model (5.1), it includes the corresponding results of Huang et al. (2002). The inference procedure in Huang et al. (2002) is based on a resampling subject bootstrap to construct confidence regions and to perform hypothesis testing.

Table 1 shows that the results for testing the constancy of  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$  are the same with significance level 0.05. For  $\beta_3$ , the  $p$ -value for the method of Huang et al. (2002) is on the border of being significant, while our test "strongly" does not reject the hypothesis that  $\beta_3$  is a constant.

	RQFind	Huang et al. (2002)
<i>Null hypothesis</i>	<i>p-value</i>	<i>p-value</i>
$\beta_0(\cdot)$ is constant	0.000	0.000
$\beta_1(\cdot)$ is constant	0.495	0.176
$\beta_2(\cdot)$ is constant	0.153	0.301
$\beta_3(\cdot)$ is constant	0.575	0.059

Table 1: Aids data:  $p$ -values

## 5.2 Primary Biliary Cirrhosis Data

This data was collected by the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver, which was conducted from January 1974 to May 1984 (data available at <http://lib.stat.cmu.edu/datasets/pbcseq>). Primary biliary cirrhosis is an autoimmune disease of the liver marked by the slow progressive destruction of the small bile ducts of the liver. We look at multiple laboratory results of 312 patients, which were taken at specified visits at six months, one year, and annually thereafter. Many patients missed their scheduled appointments due to various reasons, hence the number of repeated measurements per patient differed.

The time (in days) dependent variables we consider are PT (prothrombin time in seconds) as the response variable, and Alb (albumin in gm/dl) as the predictor variable. Only the results of the first 1600 days are considered. The varying coefficient model is

$$PT(t) = \beta_0(t) + \beta_1(t)Alb(t) + \varepsilon(t).$$

The B-spline estimator of the coefficient functions, with  $m = (4, 4)$  and  $q = (3, 3)$ , for this data is given in Figure 7. These estimates suggest an increasing trend for the intercept and a decreasing trend for the slope function. As to be expected, we find that the  $p$ -value for testing hypothesis (4) is zero in both cases, i.e. the intercept function and the slope function are non-constant.

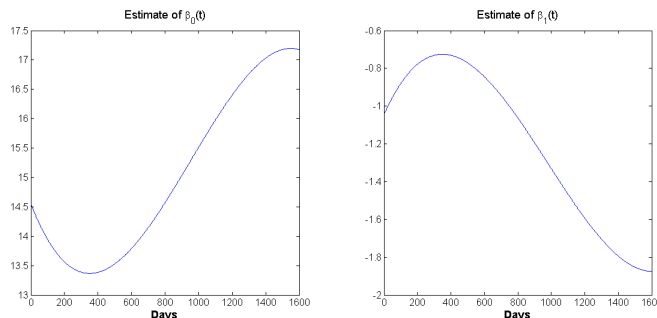


Figure 7: PBC data: estimates of the intercept function and the Albumin coefficient function.



## 6 Testing for constancy: unknown variance

When no prior information is available on the variance, it is natural to consider the RQF method where we plug in an estimate of  $\mathbf{V}$ . We have pursued this approach by estimating the variance as described in Huang et al. (2004), yielding bad results. There are several reasons for this unsuccessful attempt. Estimating  $\mathbf{V}$  is not sufficient, what is needed is a good estimation of  $(\mathbf{U}'\mathbf{V}\mathbf{U})^{-1}$ . Also, Theorems 1 and 2 only hold if  $\mathbf{V}$  is positive definite, therefore an estimator of  $\mathbf{V}$  should guarantee this. Instead, we propose a bootstrap approach when there is little information on  $\mathbf{V}$ .

Our bootstrap approach to test hypothesis (10) in VCMs is inspired by the bootstrap approach of Huang et al. (2002). The difference is our multidimensional test statistic and the use of Mahalanobis distances. An important step is to create pseudo data  $\{(Y_{ij}^{ps}, X_{ij}, t_{ij}) : i = 1, \dots, n, j = 1, \dots, N_i\}$  which satisfies the null hypothesis in (10).

Denote by  $\hat{\boldsymbol{\alpha}}^{cs}$  the estimator we obtain under the constraints imposed by the null hypothesis. Pseudo data  $\{(Y_{ij}^{ps}, X_{ij}, t_{ij}) : i = 1, \dots, n, j = 1, \dots, N_i\}$  are simulated by using  $\hat{\boldsymbol{\alpha}}^{cs}$ . The test statistic is  $\mathbf{A}^\top \hat{\boldsymbol{\alpha}} - \mathbf{a}$ , where  $\hat{\boldsymbol{\alpha}}$  is the estimator obtained without constraints. The null hypothesis is either rejected or not based on the bootstrap procedure given below.

- Step 1: Resample  $n$  subjects with replacement from

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

to obtain the bootstrap sample  $\{(Y_{ij}^{ps*}, X_{ij}^*, t_{ij}^*) : i = 1, \dots, n, j = 1, \dots, N_i^*\}$ .

- 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 (10) if  $(\mathbf{A}^\top \hat{\boldsymbol{\alpha}} - \mathbf{a} - \boldsymbol{\mu}_M)^\top \boldsymbol{\Sigma}_M (\mathbf{A}^\top \hat{\boldsymbol{\alpha}} - \mathbf{a} - \boldsymbol{\mu}_M) > M_{1-\alpha}$ , else do not reject the null hypothesis.

## 7 Conclusion

The RQF method was introduced in the varying coefficient model setting by Li et al. (2011) as well as its theoretical motivation. They illustrated on simulations the RQF method and showed that it is competitive with other methods in the literature. In this

paper, we have extended it to varying coefficient models for longitudinal data with intra-subject correlation. This method stands on its own due to its simplicity.

The simulations have shown that the RQF method is more powerful with considerably less computing time than the bootstrap method of Huang et al. (2002). Moreover, the RQF method allows to test a series of hypothesis, by adjusting the transformation matrix on the coefficients, see (10). For example one could test simultaneously whether certain coefficient functions are constant, constant with a prespecified constant, polynomial, etc.

Moreover, we paid attention to the situation where the correlation matrix  $\mathbf{V}$  is unknown by providing a bootstrap approach which is of the same spirit as the RQF method.

## Acknowledgement

We would like to thank the Editor and the referees for their detailed reading and very valuable comments on the manuscript.

M. Ahkim's research was supported by the Special Research Fund (BOF) of Universiteit Antwerpen [grant number 42FA070300FFB5994]. A. Verhasselt gratefully acknowledge support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) and the FWO research grant 1.5.137.13N. The infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI, was used for the simulations.

## 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.
- Bausch, J. (2013), 'On the Efficient Calculation of a Linear Combination of Chi-Square Random Variables with an Application in Counting String Vacua', *Journal of Physics A: Mathematical and Theoretical* **46**(1), 505202.
- De Boor, C. (2001), *A Practical Guide to Splines*, number v. 27 in 'Applied Mathematical Sciences', Springer.
- Dunkl, C. F. and Ramirez, D. E. (2001), 'Computation of the generalized f distribution', *Australian & New Zealand Journal of Statistics* **43**(1), 21–31.
- Fan, J. and Zhang, W. (1999), 'Statistical estimation in varying coefficient models', *The Annals of Statistics* **27**(5), 1491–1518.

- Hastie, T. and Tibshirani, R. (1993), ‘Varying-coefficient models’, *Journal of the Royal Statistical Society. Series B* **55**(4), 757–796.
- 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.
- Kaslow, R., Ostrow, D., Detel, R., Phair, J., Polk, B. and Rinaldo, C. (1987), ‘The Multicenter AIDS Cohort Study : rationale, organization, and selected characteristics of the participants’, *American Journal of Epidemiology* **126**, 310–318.
- Li, N., Xu, X. and Liu, X. (2011), ‘Testing the constancy in varying-coefficient regression models’, *Metrika* **74**(3), 409–438.
- Tan, W. Y. (1977), ‘On the distribution of quadratic forms in normal random variables’, *Canadian Journal of Statistics* **5**(2), 241–250.
- Wolkowicz, H. and Styan, G. P. (1980), ‘More bounds for eigenvalues using traces’, *Linear Algebra and its Applications* **31**(1), 1 – 17.

## Appendix A Notation

1. For a real valued function  $h$  on  $\mathcal{T}$ ,  $\|h\|_\infty = \sup_{t \in \mathcal{T}} |h(t)|$  denotes its supremum norm, while for a real vector valued function  $\mathbf{h} = (h_1, \dots, h_m)'$ , we let its supremum norm be  $\|\mathbf{h}\|_\infty = \max_{1 \leq i \leq m} \|h_i\|_\infty$ .
2. Let  $\mathbb{G} = \mathbb{G}_0 \times \dots \times \mathbb{G}_d$ . Define the function  $\mathbf{g}^*(t) = (g_0^*(t), \dots, g_d^*(t))'$  such that  $\|\boldsymbol{\beta} - \mathbf{g}^*\|_\infty = \rho_n = \inf_{\mathbf{g} \in \mathbb{G}} \|\boldsymbol{\beta} - \mathbf{g}\|_\infty$ . Let  $\boldsymbol{\alpha}^*$  denote the corresponding coefficient vector, i.e.  $\mathbf{g}^*(t) = \mathbf{B}(t)\boldsymbol{\alpha}^*$ . Throughout we assume that  $\lim_{n \rightarrow \infty} \rho_n = 0$ , i.e., the unknown function  $\boldsymbol{\beta}$  can be uniformly approximated by spline functions of a certain degree.

## Appendix B Assumptions

- Assumption 1.** 1. The observation times  $t_{ij}$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, n$ , 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  $\{(Y_i(t), X_i(t))\}$ ,  $i = 1, \dots, n$ . 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  $M_1 \leq f_T(t) \leq M_2$  for  $t \in \mathcal{T}$ .
2. The eigenvalues  $\eta_0(t), \dots, \eta_d(t)$  of  $\Sigma(t) = \mathbf{E}(\mathbf{X}(t)\mathbf{X}(t)')$  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  $M_3 \leq \eta_0(t) \leq \dots \leq \eta_d(t) \leq M_4$  for  $t \in \mathcal{T}$ .
3. There exists a positive constant  $M_5$  such that  $|X_p(t)| \leq M_5$  for  $t \in \mathcal{T}$  and  $p = 0, \dots, d$ .
4. There exists a positive constant  $M_6$  such that  $\mathbf{E}(\varepsilon(t)^2) \leq M_6 < \infty$  for  $t \in \mathcal{T}$ .
5.  $\limsup_n \frac{\max_p K_p}{\min_p K_p} < \infty$ .

These conditions are commonly used (e.g. Huang et al. (2004)) and are satisfied in many practical examples. As for Assumption 1.1, when dealing with deterministic time points we can replace this assumption by

$$\sup_{t \in \mathcal{T}} |F_n(t) - F_T(t)| = o(1/K_p),$$

for some distribution function  $F_T$  having a Lebesgue density function  $f_T$  which is bounded away from zero and infinity, uniformly over  $t \in \mathcal{T}$ , where  $F_n(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} 1_{t_{ij} < t}$  and  $1_{t_{ij} < t}$  is the indicator function (Huang et al. (2004)). Note that we do not assume zero modeling bias, since we allow the knots to increase to infinity.

## Appendix C Theorem of Tan (1977)

In the proof of Theorem 3 and 4 we need the following Lemma, based on Theorem 3.1 of Tan (1977).

**Lemma 1.** Let  $\mathbf{Z} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$  with  $\mathbf{V}$  invertible and  $Q = \mathbf{Z}'\mathbf{A}\mathbf{Z}$ , where  $\mathbf{A}$  is a real symmetric matrix. Then  $Q = \sum_{i=1}^k \lambda_i \chi^2(r_i, \theta_i^2)$  where  $\chi^2(r_i, \theta_i^2)$  are independent noncentral chi-square variables,  $\lambda_1, \dots, \lambda_k$  are the nonzero distinct eigenvalues of  $\mathbf{V}\mathbf{A}$  with algebraic multiplicities  $r_1, \dots, r_k$  respectively, and

$$\theta_j^2 = \boldsymbol{\mu}'\mathbf{V}^{-1}\mathbf{E}_j\boldsymbol{\mu},$$



Recall the definition of  $\boldsymbol{\alpha}^*$  (see Appendix A). Set  $\boldsymbol{\delta} = \mathbb{E}[\mathbf{Y}|\mathcal{X}] - \mathbf{U}\boldsymbol{\alpha}^*$ , then  $\|\boldsymbol{\delta}\|_\infty = O(\rho_n)$ . We can also write  $\underline{\mathbf{Y}} = \underline{\mathbf{U}}\boldsymbol{\alpha}^* + \mathbf{W}^{1/2}\boldsymbol{\delta} + \mathbf{W}^{1/2}\boldsymbol{\varepsilon}$ , so that under hypothesis  $H_0$  we obtain

$$\underline{\mathbf{Y}} \sim N(\underline{\mathbf{U}}\boldsymbol{\alpha}^* + \mathbf{W}^{1/2}\boldsymbol{\delta}, \sigma^2\underline{\mathbf{V}}).$$

Note that  $\mathbf{L}'_2\boldsymbol{\alpha}^* = 0$  under  $H_0$ , hence

$$\underline{\mathbf{L}}'_2\hat{\boldsymbol{\alpha}} = \underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\mathbf{W}^{1/2}\boldsymbol{\varepsilon} + \underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\mathbf{W}^{1/2}\boldsymbol{\delta},$$

so

$$\underline{\mathbf{L}}'_2\hat{\boldsymbol{\alpha}} \sim N_{m_p-1}(\underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\mathbf{W}^{1/2}\boldsymbol{\delta}, \sigma^2\underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{U}}(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{L}}_2).$$

Denote  $\boldsymbol{\Sigma}_2 := \text{Cov}(\underline{\mathbf{L}}'_2\hat{\boldsymbol{\alpha}})$ . We define

$$\xi_0 := (\mathbf{W}^{1/2}\boldsymbol{\varepsilon})'\underline{\mathbf{V}}^{-1}\underline{\mathbf{U}}(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{L}}_2\boldsymbol{\Sigma}_2^{-1}\underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\mathbf{W}^{1/2}\boldsymbol{\varepsilon} \sim \chi^2(m_p - 1).$$

Using Lemma 1, we obtain that

$$\begin{aligned} \eta_0 &:= (\mathbf{W}^{1/2}\boldsymbol{\varepsilon})'(\mathbf{I}_N - P_{\underline{\mathbf{U}}})(\mathbf{W}^{1/2}\boldsymbol{\varepsilon}) \sim \sum_{i=1}^l \lambda_i \chi^2(r_i), \\ \xi_1 &:= \underline{\mathbf{Y}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{U}}(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{L}}_2\boldsymbol{\Sigma}_2^{-1}\underline{\mathbf{L}}'_2(\underline{\mathbf{U}}'\underline{\mathbf{U}})^{-1}\underline{\mathbf{U}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{Y}} \sim \chi^2(m_p - 1, \gamma^2), \\ \eta_1 &:= \underline{\mathbf{Y}}'(\mathbf{I}_N - P_{\underline{\mathbf{U}}})\underline{\mathbf{Y}} \sim \sum_{i=1}^k \lambda_i \chi^2(r_i, \theta_i^2), \end{aligned}$$

where  $\gamma^2$  and  $\theta_i^2$  are specified in Lemma 1. Denote  $\tau_0 = \frac{\eta_0}{\xi_0}$  and  $\tau_1 = \frac{\eta_1}{\xi_1}$ . To prove Theorem 2, we need to show that

$$\lim_{n \rightarrow \infty} (F_{\tau_1}(t) - F_{\tau_0}(t)) = 0 \quad \text{uniformly in } t > 0. \quad (13)$$

Some mathematical preparation is needed to prove (13). The Takagi factorization of  $(\mathbf{I}_N - P_{\underline{\mathbf{U}}})$  leads to a matrix  $\mathbf{G} \in \mathbb{R}^{(N-dim) \times N}$  such that

$$\mathbf{G}'\mathbf{G} = (\mathbf{I}_N - P_{\underline{\mathbf{U}}}), \quad \mathbf{G}\mathbf{G}' = \mathbf{I}_{N-dim}.$$

Throughout  $\|\mathbf{A}\|$  ( $\|\mathbf{c}\|$ ) denotes the Frobenius (Euclidean) norm of a matrix  $\mathbf{A}$  (vector  $\mathbf{c}$ ), and  $\langle \mathbf{a}, \mathbf{b} \rangle$  denotes the standard inproduct of vectors  $\mathbf{a}, \mathbf{b}$ . Let  $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \dots, \zeta_{N-dim})' = \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}$ , then  $\eta_0 = \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}\|_2^2 = \sum_{i=1}^{N-dim} \zeta_i^2$  where

$$\boldsymbol{\zeta} \sim N(\mathbf{0}, \mathbf{G}\sigma^2\underline{\mathbf{V}}\mathbf{G}').$$

Let  $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_{N-dim})' = \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}$ . Note that if  $\boldsymbol{\delta} = \mathbf{0}$ , then there is nothing to prove since in that case  $\xi_0 = \xi_1$  and  $\eta_0 = \eta_1$ , so we proceed with the case  $\boldsymbol{\delta} \neq \mathbf{0}$ . We also have that

$$N - dim = \text{Rank}(\mathbf{G}'\mathbf{G}) \leq \min(\text{Rank}(\mathbf{G}'), \text{Rank}(\mathbf{G})) = \text{Rank}(\mathbf{G}) \leq N - dim,$$

from which it follows that  $\boldsymbol{\nu} \neq \mathbf{0}$ . Define an orthogonal transformation  $\mathbf{T} \in \mathbb{R}^{(N-dim) \times (N-dim)}$  with first row equal to  $\boldsymbol{\nu}'/\|\boldsymbol{\nu}\|$  and let

$$\boldsymbol{\zeta}^* = (\zeta_1^*, \zeta_2^*, \dots, \zeta_{N-dim}^*)' := \mathbf{T}\boldsymbol{\zeta}.$$

We obtain the expressions

$$\begin{aligned} \eta_0 &= \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}\|_2^2 = \sum_{i=1}^{N-dim} (\zeta_i^*)^2 \\ \eta_1 &= \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon} + \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|_2^2 = \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}\|_2^2 + \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|_2^2 + \\ &\quad 2\langle \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}, \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta} \rangle \\ &= \|\boldsymbol{\zeta}\|^2 + \|\boldsymbol{\nu}\|^2 + 2\langle \boldsymbol{\zeta}, \boldsymbol{\nu} \rangle \\ &= \|\boldsymbol{\zeta}\|^2 + \|\boldsymbol{\nu}\|^2 + 2\zeta_1^* \|\boldsymbol{\nu}\| \\ &= (\|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\| + \zeta_1^*)^2 + \sum_{i=2}^{N-dim} (\zeta_i^*)^2. \end{aligned}$$

Therefore

$$\begin{aligned} |\eta_1 - \eta_0| &\leq \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 + 2\|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|\zeta_1^* \\ \mathbb{E}(|\eta_1 - \eta_0|) &\leq \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 + 2\|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\| \mathbb{E}(|\zeta_1^*|) \\ &= \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 + 2\|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\| \sqrt{\frac{2}{\pi}} \sqrt{\text{Var}(\zeta_1^*)}, \end{aligned} \quad (14)$$

since for a mean zero normal variable  $Z$  we have the property  $\mathbb{E}(|Z|) = \sqrt{\frac{2}{\pi}} \sqrt{\text{Var}(Z)}$ . Now  $\text{Var}(\boldsymbol{\zeta}^*) = \text{Var}(\mathbf{T}\boldsymbol{\zeta}) = \mathbf{T}\mathbf{G}\sigma^2\mathbf{V}\mathbf{G}'\mathbf{T}'$  and  $\mathbf{T}\mathbf{G}\mathbf{G}'\mathbf{T}' = \mathbf{I}_{N-dim}$ . We want to bound  $\text{Var}(\zeta_1^*)$ . Let  $\mathbf{b} = (b_1, b_2, \dots, b_N)$  denote the first row of the orthogonal matrix  $\mathbf{T}\mathbf{G}$ , then we know  $\|\mathbf{b}\| = 1$ , also denote by  $\mathbf{c}_1, \dots, \mathbf{c}_N$  the columns of  $\sigma^2\mathbf{V}$ . Using the fact  $\langle \mathbf{b}, \mathbf{c}_i \rangle \leq \sigma^2 \max_{i=1}^n w_i \sqrt{N_i}$  which is obtained by the Cauchy-Schwarz inequality, and the symmetric property of  $\sigma^2\mathbf{V}$ , we have that

$$\text{Var}(\zeta_1^*) = \sum_{i=1}^N b_i \langle \mathbf{b}, \mathbf{c}_i \rangle \leq \sum_{i=1}^N |b_i \langle \mathbf{b}, \mathbf{c}_i \rangle| \leq \sum_{i=1}^N |\langle \mathbf{b}, \mathbf{c}_i \rangle| \leq \sigma^2 N \max_{i=1}^n w_i \sqrt{N_i}.$$

Using the previous inequality, we can continue from equation (14) to obtain

$$\mathbb{E}(|\eta_1 - \eta_0|) \leq \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 + 2\|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\| \sqrt{\frac{2}{\pi}} \sigma \sqrt{N \max_{i=1}^n w_i \sqrt{N_i}}. \quad (15)$$

Let  $\mathbf{H} = \frac{1}{\sigma} \boldsymbol{\Sigma}_2^{-1/2} \mathbf{L}'_2 (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'\mathbf{W}^{-1/2}\mathbf{V}^{-1/2}$ , then  $\mathbf{H}\mathbf{H}' = \mathbf{I}_{m_p-1}$ ,  $\xi_0 = \|\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma\|^2$  and  $\xi_1 = \|\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma + \mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\delta}/\sigma\|^2$ . Analogously as in (15) we obtain

$$\mathbb{E}(|\xi_1 - \xi_0|) \leq \|\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\delta}/\sigma\|^2 + 2\|\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\delta}/\sigma\| \sqrt{\frac{2}{\pi}}, \quad (16)$$

since for any orthogonal transformation  $\mathbf{T}_2 \in \mathbb{R}^{(m_p-1) \times (m_p-1)}$ , the variance of the first component of  $\boldsymbol{\kappa}^* := \mathbf{T}_2 \boldsymbol{\kappa}$ , where  $\boldsymbol{\kappa} = \mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma$  is obtained by the entry with index (1, 1) of the matrix

$$\text{Cov}(\boldsymbol{\kappa}^*) = \frac{1}{\sigma^2} \mathbf{T}_2 \mathbf{H}\mathbf{V}^{-1/2} \sigma^2 \mathbf{V}\mathbf{V}^{-1/2} \mathbf{H}' \mathbf{T}_2' = \mathbf{I}_{m_p-1}.$$

Note that  $\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}$  and  $\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma$  are independent multivariate normal random vectors, because on the one hand

$$\text{Cov}(\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}, \mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma) = \mathbf{G}\mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{H}',$$

on the other hand, by the same argument as in (12)

$$\mathbf{G}'\mathbf{G}\mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{H}'\mathbf{H} = \mathbf{0},$$

from which we find that

$$\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{H}'\mathbf{H}\mathbf{H}' = \mathbf{G}\mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{H}' = \mathbf{0}.$$

Hence

$$\text{Cov}(\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\varepsilon}, \mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\varepsilon}/\sigma) = \mathbf{0}.$$

Fix a  $t > 0$ , then

$$\begin{aligned} F_{\tau_1}(t) - F_{\tau_0}(t) &= P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_0}{\xi_0} < t\right) \\ &= P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_1}{\xi_0} < t\right) + P\left(\frac{\eta_1}{\xi_0} < t\right) - P\left(\frac{\eta_0}{\xi_0} < t\right) \\ &\leq P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_1}{\xi_0} < t\right). \end{aligned} \quad (17)$$

For the last inequality, since  $\eta_1$  and  $\xi_1$  are independent, and  $\eta_1$  and  $\xi_0$  are independent, we have that

$$\begin{aligned} P\left(\frac{\eta_1}{\xi_0} < t\right) &= \mathbb{E}_{\xi_0} \{P(\eta_1 \leq t\xi_0) | \xi_0\} \\ &= \mathbb{E}_{\xi_0} \left\{ \int_{\|\mathbf{x} + \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 \leq t\xi_0} f(\mathbf{x}) d\mathbf{x} \mid \xi_0 \right\} \\ &= \mathbb{E}_{\xi_0} \left\{ \int_{\|\mathbf{x}\|^2 \leq t\xi_0} f(\mathbf{x} - \mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}) d\mathbf{x} \mid \xi_0 \right\} \\ &\leq \mathbb{E}_{\xi_0} \left\{ \int_{\|\mathbf{x}\|^2 \leq t\xi_0} f(\mathbf{x}) d\mathbf{x} \mid \xi_0 \right\} \\ &= P\left(\frac{\eta_0}{\xi_0} < t\right), \end{aligned}$$

where  $f$  is the density function of the multivariate normal distribution

$$N_{N-\dim}(0, \mathbf{G}\sigma^2\mathbf{V}\mathbf{G}').$$



Continuing from equation (17) with  $k$  a positive real number

$$\begin{aligned}
P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_1}{\xi_0} < t\right) &= P(\xi_0 \leq \eta_1/t) - P(\xi_1 \leq \eta_1/t) \\
&= P(\xi_1 \leq \eta_1/t, \xi_0 \leq \eta_1/t) + \\
&\quad P(\xi_1 > \eta_1/t, \xi_0 \leq \eta_1/t) - P(\xi_1 \leq \eta_1/t) \\
&\leq P(\xi_1 > \eta_1/t, \xi_0 \leq \eta_1/t) \\
&= P(\xi_1 > \eta_1/t, \eta_1/t - k \leq \xi_0 \leq \eta_1/t) + \\
&\quad P(\xi_1 > \eta_1/t, \xi_0 < \eta_1/t - k) \\
&\leq P(\eta_1/t - k \leq \xi_0 \leq \eta_1/t) + P(\xi_0 - \xi_1 < -k) \\
&\leq M_{\xi_0}k + \frac{1}{k} \mathbb{E}(|\xi_0 - \xi_1|), \tag{18}
\end{aligned}$$

where  $M_{\xi_0}$  is the maximum of the density function of  $\xi_0$  (the Markov inequality is applied in (18)). Substitute

$$k = \sqrt{\frac{\mathbb{E}(|\xi_0 - \xi_1|)}{M_{\xi_0}}}$$

in (18) to find that

$$P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_1}{\xi_0} < t\right) \leq 2\sqrt{M_{\xi_0} \mathbb{E}(|\xi_0 - \xi_1|)},$$

and by (17) we obtain that for all  $t \geq 0$

$$F_{\tau_1}(t) - F_{\tau_0}(t) \leq 2\sqrt{M_{\xi_0} \mathbb{E}(|\xi_0 - \xi_1|)}.$$

On the other hand, we obtain in a similar fashion

$$\begin{aligned}
F_{\tau_1}(t) - F_{\tau_0}(t) &= P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_0}{\xi_0} < t\right) \\
&= P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_0}{\xi_1} < t\right) + P\left(\frac{\eta_0}{\xi_1} < t\right) - P\left(\frac{\eta_0}{\xi_0} < t\right) \\
&\geq P\left(\frac{\eta_1}{\xi_1} < t\right) - P\left(\frac{\eta_0}{\xi_1} < t\right) \\
&= -P(\eta_0 \leq t\xi_1) - P(\eta_1 \leq t\xi_1) \\
&= -P(\eta_1 \leq t\xi_1, \eta_0 \leq t\xi_1) + P(\eta_1 > t\xi_1, \eta_0 \leq t\xi_1) - \\
&\quad P(\eta_1 \leq t\xi_1) \\
&\geq -P(\eta_1 > t\xi_1, \eta_0 \leq t\xi_1) \\
&= -P(\eta_1 > t\xi_1, t\xi_1 - k \leq \eta_0 \leq t\xi_1) + \\
&\quad P(\eta_1 > t\xi_1, \eta_0 \leq t\xi_1 - k) \\
&\geq -P(t\xi_1 - k \leq \eta_0 \leq t\xi_1) + P(\eta_0 - \eta_1 < -k) \\
&\geq -M_{\eta_0}k + \frac{1}{k} \mathbb{E}(|\eta_0 - \eta_1|), \tag{19}
\end{aligned}$$

where  $M_{\eta_0}$  is the maximum of the density function of the random variable  $\eta_0$ . Substitute in (19)

$$k = \sqrt{\frac{\mathbb{E}(|\eta_0 - \eta_1|)}{M_{\eta_0}}}$$

to finally establish

$$\forall t > 0 : -2\sqrt{M_{\eta_0} \mathbb{E}(|\eta_0 - \eta_1|)} \leq F_{\tau_1}(t) - F_{\tau_0}(t) \leq 2\sqrt{M_{\xi_0} \mathbb{E}(|\xi_0 - \xi_1|)}. \quad (20)$$

Note that

$$\begin{aligned} \|\mathbf{H}\mathbf{V}^{-1/2}\boldsymbol{\delta}/\sigma\|^2 &\leq \|\mathbf{V}^{-1/2}\boldsymbol{\delta}/\sigma\|^2 = O(N\rho_n^2\|\mathbf{V}^{-1/2}\|^2) \\ \|\mathbf{G}\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 &\leq \|\mathbf{W}^{1/2}\boldsymbol{\delta}\|^2 = O(N\rho_n^2w_{\max}) \end{aligned}$$

since  $\mathbf{H}'\mathbf{H}$  and  $\mathbf{G}'\mathbf{G}$  are idempotent matrices, thus 0 and 1 are the only eigenvalues. Then by (14),(16) and (20), it follows that

$$\begin{aligned} |F_{\tau_1}(t) - F_{\tau_0}(t)| &= O\left(\sqrt{M_{\xi_0} \left(N\rho_n^2\|\mathbf{V}^{-1/2}\|^2 + \sqrt{N}\rho_n\|\mathbf{V}^{-1/2}\|\right)}\right) + \\ &O\left(\sqrt{M_{\eta_0} \left(Nw_{\max}\rho_n^2 + Nw_{\max}\rho_n\sqrt{\max_i w_i N_i^{1/2}}\right)}\right). \end{aligned}$$

■

## Appendix F Rate of convergence

In Theorem 2 we assume (9). We shed more light on this rate by assuming that  $\frac{N_{\max}}{N_{\min}}$  is bounded ( $N_{\max} = \max_{i=1,\dots,n} N_i$  and  $N_{\min} = \min_{i=1,\dots,n} N_i$ ),  $\frac{N_{\max}^{3/2}}{n} = o(1)$  and  $\frac{dim}{n} = o(1)$ . Suppose that subjects with equal number of repeated measurements have the same time points, we do not need this assumption if the correlation structure does not depend on time, as is the case with any time independent correlation structure.

For the first part we use the fact that  $M_{\xi_0} = O\left(\frac{1}{\sqrt{m_p-1}}\right)$  (Li et al. (2011)), thus the first part is bounded by

$$\sqrt{\left(\frac{N\rho_n^2\|\mathbf{V}^{-1/2}\|^2}{\sqrt{m_p}} + \frac{\sqrt{N}\rho_n\|\mathbf{V}^{-1/2}\|}{\sqrt{m_p}}\right)}.$$

**Bounding  $M_{\eta_0}$**  For the second part, we note that there is no closed form expression of the density function of a linear combination of chi-square variables (see Bausch (2013))

among others). However, we obtain a reasonable bound on  $M_{\eta_0}$  which is the maximum of the density of  $\sum_{i=1}^k \lambda_i \chi^2(r_i)$ .

First, it does not hold that  $r_i = 1$  for all  $i$ . To prove this, suppose otherwise, i.e.  $r_i = 1$  for all  $i$ . Then, by Theorem 1, we have  $k = \sum_{i=1}^k r_i = N - \dim$ . Next, we obtain a bound on  $k$ . We argue, as in the proof of Theorem 1, that to find a bound on  $k$  we restrict to the eigenspace  $\mathcal{E}_1 \subset \mathbb{R}^N$  of eigenvalue 1 of  $(\mathbf{I}_N - P_{\mathbf{U}})$ . Thus, restricting to  $\mathcal{E}_1$ , we only look at the number of positive eigenvalues of  $\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2}$  which is a block diagonal matrix. By the restriction on the time points (see above),  $\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2}$  contains at most  $N_{\max} - N_{\min} + 1$  different block matrices with dimensions not exceeding  $N_{\max}$ . Hence, the number of different positive eigenvalues does not exceed  $N_{\max}(N_{\max} - N_{\min} + 1)$ , i.e.  $k \leq N_{\max}(N_{\max} - N_{\min} + 1)$ . By assumption all  $r_i = 1$ , thus it should hold

$$N - \dim = \sum_i r_i = k \leq N_{\max}(N_{\max} - N_{\min} + 1). \quad (21)$$

Divide (21) by  $N$ , since  $N_{\max}/N_{\min}$  is bounded by  $C > 0$  and  $N_{\max}/n \rightarrow 0$ , we obtain from the previous inequality using also the fact  $N \geq nN_{\min}$ , that the left hand side is  $1 + o(1)$  while the right hand side is  $o(1)$ . This is a contradiction. Hence, there is a  $1 \leq j \leq k$  such that  $r_j > 1$ .

Also, we can write  $\sum_{i=1}^k \lambda_i \chi^2(r_i)$  as a sum of a scaled chi-square distribution  $\lambda_{\max} \chi^2(r_{\lambda_{\max}})$  and the remaining part, where  $\lambda_{\max} := \max_i \lambda_i$  is assumed to be an eigenvalue of a vector in  $\mathcal{E}_1$ . Moreover, we assume that  $r_{\lambda_{\max}} > 1$ . The density of this sum is a convolution which is bounded by  $O(\frac{1}{\lambda_{\max}})$  (after a small calculation). Moreover, by Theorem 2.1 of Wolkowicz and Styan (1980) we know that

$$\lambda_{\max} \geq \frac{\text{Tr}(\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2})}{N} = \sum_{i=1}^n N_i w_i / N \geq w_{\min}$$

since  $\mathbf{V}$  contains only ones on its diagonal. Hence we derived  $M_{\eta_0} = O(1/w_{\min})$ .

**Bound on (9)** By the discussion above, we have the following bound on (9)

$$\sqrt{\left( \frac{N \rho_n^2 \|\mathbf{V}^{-1/2}\|^2}{\sqrt{m_p}} + \frac{\sqrt{N} \rho_n \|\mathbf{V}^{-1/2}\|}{\sqrt{m_p}} \right)} + \sqrt{\left( \frac{w_{\max}}{w_{\min}} N \rho_n^2 + N \frac{w_{\max}}{w_{\min}} \rho_n \sqrt{\max_i w_i N_i^{1/2}} \right)}.$$