

# **Empirical Bayesian Smoothing Splines for Signals with Correlated Errors: Methods and Applications**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

– Doctor rerum naturalium –

der Georg-August-Universität Göttingen

im Promotionsprogramm “Mathematik”

der Georg-August University School of Science (GAUSS)

vorgelegt von

MSc. Appl. Math. Francisco Rosales Marticorena

aus

Lima – Peru

Göttingen, im Juni 2016

## **Betreuungsausschuss**

Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik

Prof. Dr. Stephan von Cramon-Taubadel, Department für Agrarökonomie und Rurale Entwicklung

Prof. Dr. Thomas Kneib, Lehrstühle für Statistik und Ökonometrie

## **Mitglieder der Prüfungskommission**

Referent: Prof. Dr. Tatyana Krivobokova, Institut für Mathematische Stochastik

Korreferent: Prof. Dr. Stephan von Cramon-Taubadel, Department für Agrarökonomie und Rurale Entwicklung

## **Weitere Mitglieder der Prüfungskommission:**

Prof. Dr. Dominic Schuhmacher, Institut für Mathematische Stochastik

Prof. Dr. Gerlind Plonka-Hoch, Institut für Numerische und Angewandte Mathematik

Prof. Dr. Rusell Luke, Institut für Numerische und Angewandte Mathematik

Tag der mündlichen Prüfung: 22.06.2016

*To Auxilio Lacouture*

# Acknowledgements

I am deeply grateful to my principal advisor Prof. Dr. Krivobokova for her intellectual generosity and committed supervision. I extend my gratitude to my second advisor Prof. Dr. von Cramon-Taubadel for his feedback in the field of agricultural economics, and to my colleagues at the Georg-August-Universität Göttingen for the helpful discussions. In alphabetical order: Benjamin Eltzner, Friederike Greb, Carsten Holst, Nadja Klein, Johannes Martini, Marco Singer, Katja Schwartz, Paulo Serra, Max Sommerfeld, Inder Tecuapetla, Manuel Wiessenfarth, Carina Wollnik and Jisu Yoon. I would also like to acknowledge the research training groups 1644 (Scaling Problems in Statistics) and 2088 (Discovering Structure in Complex Data: Statistics meets Optimisation and Inverse Problems) for the financial support. Lastly, I thank the families Rosales Marticorena, and Marticorena Gamarra for their crucial assistance throughout my academic career.

# Abstract

Smoothing splines is a well established method in non-parametric statistics, although the selection of the smoothness degree of the regression function is rarely addressed and, instead, a two times differentiable function, i.e. cubic smoothing spline, is assumed. For a general regression function there is no known method that can identify the smoothness degree under the presence of correlated errors. This apparent disregard in the literature can be justified because the condition number of the solution increases with the smoothness degree of the function, turning the estimation unstable. In this thesis we introduce an exact expression for the Demmler-Reinsch basis constructed as the solution of an ordinary differential equation, so that the estimation can be carried out for an arbitrary smoothness degree, and under the presence of correlated errors, without affecting the condition number of the solution. We provide asymptotic properties of the proposed estimators and conduct simulation experiments to study their finite sample properties. We expect this new approach to have a direct impact on related methods that use smoothing splines as a building block. In this direction, we present extensions of the method to signal extraction and functional principal component analysis. The empirical relevance to our findings in these areas of statistics is shown in applications for agricultural economics and biophysics. R packages of the implementation of the developed methods are also provided.

# Table of Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Smoothing Splines . . . . .	2
1.2. Asymptotic Properties . . . . .	6
1.3. Open Problems and Objectives of the Thesis . . . . .	8
<b>2. Demmler-Reinsch Basis</b>	<b>11</b>
2.1. Ordinary Differential Equation . . . . .	11
2.2. A General Demmler-Reinsch Basis . . . . .	15
2.2.1. Eigenvalues . . . . .	15
2.2.2. Eigenvectors . . . . .	19
2.3. Comparison with Numerical Approximations . . . . .	20
<b>3. Smoothing Splines with Correlated Errors</b>	<b>25</b>
3.1. Statistical Model . . . . .	26
3.2. Estimators . . . . .	27
3.2.1. Empirical Bayes Function . . . . .	28
3.2.2. Smoothing Parameter . . . . .	30
3.2.3. Correlation Matrix . . . . .	31
3.2.4. Smoothness Class . . . . .	33
3.3. Asymptotic Properties <sup>1</sup> . . . . .	34
3.4. Implementation . . . . .	42
3.4.1. Statistical Algorithm . . . . .	42
3.4.2. Numerical Simulations . . . . .	44
3.5. Digression on Smoothness Classes . . . . .	51
<b>4. Extensions of Smoothing Splines with Correlated Errors</b>	<b>57</b>
4.1. Signal Decomposition Analysis . . . . .	57
4.1.1. Additive Models . . . . .	59

---

<sup>1</sup>This section corresponds to the work of Dr. Paulo Serra developed in a joint project at the Institute of Mathematical Stochastics at Göttingen University.

*Table of Contents*

---

4.1.2. Model for Signal Extraction . . . . .	62
4.1.3. Comparison with Other Methods . . . . .	64
4.1.4. Discussion . . . . .	73
4.2. Functional Principal Component Analysis . . . . .	74
4.2.1. Functional Data . . . . .	74
4.2.2. Principal Component Characterisation . . . . .	75
4.2.3. Statistical Algorithm . . . . .	77
4.2.4. Discussion . . . . .	79
<b>5. Applications</b>	<b>81</b>
5.1. Nonparametric Price Transmission . . . . .	81
5.1.1. Model . . . . .	83
5.1.2. Estimation . . . . .	85
5.1.3. Application . . . . .	85
5.1.4. Conclusions . . . . .	90
5.2. Stem Cell Dynamics . . . . .	92
5.2.1. Experiment . . . . .	93
5.2.2. Model . . . . .	93
5.2.3. Estimation . . . . .	95
5.2.4. Conclusions . . . . .	100
<b>6. Software</b>	<b>101</b>
6.1. R Package <code>ebsc</code> . . . . .	101
6.2. R Package <code>scsd</code> . . . . .	102
6.3. R Package <code>fpca</code> . . . . .	104
<b>7. Summary and Outlook</b>	<b>105</b>
<b>Appendices</b>	<b>107</b>
<b>A. Auxiliary Results</b>	<b>108</b>
<b>B. Additional Tables</b>	<b>111</b>

# 1. Introduction

Smoothing splines can be seen as the solution of a variational calculus problem where the aim is to obtain an approximation of the data that is accurate and yet smooth. The origins of the idea can be tracked to [Whittaker \[1923\]](#), [Schoenberg \[1964\]](#) and [Reinsch \[1967\]](#) who built the foundations of a general class of regularisation problems that are today a prominent research area in mathematical statistics. More specifically, the idea can be presented as the following (constrained) minimisation problem:

$$\min_{f \in C^q[x_1, x_n]} \int_{x_1}^{x_n} \{f^{(q)}(x)\}^2 dx : \sum_{i=1}^n \left\{ \frac{Y_i - f(x_i)}{\delta Y_i} \right\}^2 < S, \quad (1.1)$$

where  $(x_i, Y_i)$  are ordered data pairs such that  $x_1 < x_2 < \dots < x_n$ ,  $C^q[x_1, x_n]$  is the space of functions that have  $q$  continuous derivatives in  $[x_1, x_n]$ ,  $S$  is a redundant parameter and  $\delta Y_i$ 's are assumed to be known quantities. The original problem presented in (1.1) has an historical appeal, but it is mostly studied in its unconstrained form, where the trade-off between precision and smoothness of the solution is controlled by the Lagrangian parameter (see [Definition 6](#)). After [Reinsch \[1967\]](#), a vast part of the development of smoothing splines was due to [Demmler and Reinsch \[1975\]](#), [Wahba \[1978\]](#), [Speckman \[1985\]](#), among others. Moreover, spline methods in general have gained popularity not only due to the development of its theoretical properties portrayed in the publication of seminal monographs [cf. [de Boor, 1997](#), [Eubank, 1988](#), [Wahba, 1990](#), [Schumaker, 2007](#)], but also due to the availability of growingly amounts



of data and software packages that facilitated the application of new statistical methods, as illustrated in modern text books in non-parametric applied statistics [cf. [Hastie and Tibshirani, 1990](#), [Ruppert et al., 2003](#), [Wood, 2006](#)].

This thesis provides results of an smoothing spline estimator for a regression function of unknown smoothness degree, and under the presence of correlated errors. We present the subject by introducing some basic definitions; known asymptotic properties of the smoothing splines estimators; and some open problems that situate our work in the literature and motivate the objectives of the thesis.

## 1.1. Smoothing Splines

Splines are piecewise polynomials that join at points called *knots*. Throughout this dissertation we make the assumption that function  $f$  has support in  $[0, 1]$ , with knots located in an equidistant grid  $\underline{\tau} = \{\tau_j : \tau_j = j/(n - 1), j = 0, 1, \dots, n - 1\}$ . The smoothest polynomial spline space is given in the following definition

**Definition 1.** *The smoothest space of polynomial splines of degree  $p$  with knots placed at  $\underline{\tau}$  is defined as*

$$\mathcal{S}(p; \underline{\tau}) := \left\{ s : \exists s_1, \dots, s_{n-1} \in \mathcal{P}_p : s(x) = s_i(x) \text{ for } x \in [\tau_{i-1}, \tau_i], i = 1, \dots, n - 1 \right. \\ \left. \text{and } s_i^{(j)}(\tau_i) = s_{i+1}^{(j)}(\tau_i), j = 0, \dots, p - 1, i = 1, \dots, n - 2 \right\},$$

where  $\mathcal{P}_p$  denote an element of the polynomial space of degree  $p$ .

A particular class of polynomial spline space is the natural spline space, which arises as the solution of the variational problem posted by Reinsch (1.1).

**Definition 2.** The space of natural splines of degree  $2q - 1$  with knots  $\underline{\tau}$  is defined as

$$\mathcal{NS}(2q - 1; \underline{\tau}) := \{s \in \mathcal{S}(2q - 1; \underline{\tau}) : s|_{[0, \tau_1]}, s|_{[\tau_{n-3}, 1]} \in \mathcal{P}_{q-1}\},$$

where  $s|_{[0, \tau_1]}$  and  $s|_{[\tau_{n-3}, 1]}$  denote that the evaluation of function  $s(\cdot)$  is restricted to the domains  $[0, \tau_1]$  and  $[\tau_{n-3}, 1]$  respectively.

Hence a natural spline is  $2q - 2$  times continuously differentiable piecewise polynomial of degree  $2q - 1$  on  $[\tau_1, \tau_n - 2]$  and  $q - 1$  on the intervals  $[0, \tau_1]$  and  $[\tau_{n-3}, 1]$ . We introduce next the space that will use more intensively throughout this dissertation.

**Definition 3.** A Sobolev space of degree  $\beta$  is defined as

$$\mathcal{W}^\beta(M) := \left\{ f : f \in \mathcal{C}^{[\beta]-1}[0, 1], \|f^{(\beta)}\|^2 = \sum_{i=1}^{\infty} \theta_{\beta,i}^2 \nu_{\beta,i}^{\beta/[\beta]} < M^2 \right\} \quad (1.2)$$

An orthonormal basis of the space in Definition 3 is the Demmler-Reinsch basis.

**Definition 4.** A Demmler-Reinsch Basis of degree  $\beta$  is an orthonormal basis  $\{\psi_i(x)\}_{i=1}^{\infty}$  of  $\mathcal{W}^q(M)$  determined by

$$\int_0^1 \psi_{\beta,i}(x) \psi_{\beta,j}(x) dx = \delta_{i,j} = \nu_{\beta,i}^{-1} \int_0^1 \psi_{\beta,i}^{(\beta)}(x) \psi_{\beta,j}^{(\beta)}(x) dx. \quad (1.3)$$

Hence,  $f \in \mathcal{W}^q(M)$  can be represented as  $f = \sum_{i=1}^{\infty} \theta_{\beta,i} \psi_{\beta,i}$  for  $\theta_{\beta,i} = \int_0^1 f(x) \psi_{\beta,i}(x) dx$ , and  $\|f^{(\beta)}\|^2 = \sum_{i=1}^{\infty} \theta_{\beta,i}^2 \nu_{\beta,i} < M^2$ . Using Definitions 4 and 3 we can now re-state the definition of the Demmler-Reinsch basis for the natural spline case in Definition 2.

**Definition 5.** A Demmler-Reinsch basis of the natural spline space  $\mathcal{NS}(2q - 1, \underline{\tau})$  is defined by

$$\frac{1}{n} \sum_{i=1}^n \phi_{q,i}(x_k) \phi_{q,j}(x_k) = \delta_{i,j} = \eta_{q,i}^{-1} \int_0^1 \phi_{q,i}^{(q)}(x) \phi_{q,j}^{(q)}(x) dx. \quad (1.4)$$

As mentioned before the solution of the smoothing splines problem is an object in  $\mathcal{W}^q[0, 1]$ . Next we provide a formal definition of the problem.

**Definition 6.** *A smoothing spline estimator is the solution of the following minimization problem*

$$\min_{f \in \mathcal{W}^q[0,1]} \left[ \frac{1}{n} \sum_{i,j=1}^n \{Y_i - f(x_i)\} r_{i,j}^{-1} \{Y_j - f(x_j)\} + \lambda \int_0^1 \{f^{(q)}(x)\}^2 dx \right], \quad (1.5)$$

where  $\lambda \in \mathbb{R}_+$ ,  $q \in \mathbb{N}$ , data pairs  $(x_i, Y_i)$  follow  $Y_i = f(x_i) + \epsilon_i$ , and  $\{\epsilon_i\}_{i=1}^n$  is an homoscedastic stationary process with correlation matrix  $\mathbf{R}$  of elements  $\text{Cor}(\epsilon_i, \epsilon_j) = r_{i,j}$ . In matrix notation, the solution of (1.5) for an arbitrary basis  $\mathbf{C}_q$  of  $\mathcal{W}^q[0, 1]$  can be written as

$$\begin{aligned} \hat{\mathbf{f}}_{\lambda,q,\mathbf{R}} &= \mathbf{C}_q (\mathbf{C}_q^T \mathbf{R}^{-1} \mathbf{C}_q + \lambda n \mathbf{D}_q)^{-1} \mathbf{C}_q^T \mathbf{R}^{-1} \mathbf{Y} \\ &= \mathbf{S}_{\lambda,q,\mathbf{R}} \mathbf{Y}, \end{aligned} \quad (1.6)$$

where  $\mathbf{D}_q$  is a penalty matrix that depends on  $q$  and  $\mathbf{S}_{\lambda,q,\mathbf{R}}$  is referred as the smoother matrix.

Equation (1.6) has two clear extreme situations. Namely if  $\lambda = 0$ , it reduces to an interpolation problem, and if  $\lambda \rightarrow \infty$  it turns into the estimation of a polynomial of degree  $q - 1$ . To reduce the computational complexity of the solution in (1.6), the convenience of using an orthonormal basis obvious. The following two examples show its use for the independence case.

**Example 1.** *Given Definition 5 and its solution given in (1.6) for  $\text{Cor}(\epsilon_i, \epsilon_j) = \delta_{i,j}$ , the smoothing spline estimator can be written explicitly in terms of the Demmler-*

Reinsch basis as

$$\hat{f}(x) = \frac{1}{n} \sum_{l=1}^n \sum_{i=1}^n \frac{\phi_{q,i}(x_l)\phi_{q,i}(x)}{1 + \lambda n \eta_{q,i}} Y_l = \sum_{i=1}^n \frac{w_i \phi_{q,i}(x)}{1 + \lambda n \eta_{q,i}}, \quad (1.7)$$

where we have used  $\mathbf{D}_q = \text{diag}\{\eta_{q,i}\}$  and  $w_i = n^{-1} \sum_{l=1}^n \phi_{q,i}(x_l) Y_l$ .

In the next example we present a different approach to solve the penalised regression problem that consists in representing it as a linear mixed model (LMM) problem [cf. Ruppert et al., 2003, Wood, 2006].

**Example 2.** Given Definition 5 and its solution (1.6) for  $\text{Cor}(\epsilon_i, \epsilon_j) = \delta_{i,j}$  and matrices  $\mathbf{X} = \{\phi_{q,1}(x), \dots, \phi_{q,q}(x)\}$  and  $\mathbf{Z} = \{\eta_{q,q+1}^{-1/2} \phi_{q,q+1}(x), \dots, \eta_{q,n}^{-1/2} \phi_{q,n}(x)\}$ , it can be shown that the solution of the smoothing splines problem in (1.6) is equivalent to the solution of the following LMM problem

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad \mathbf{u} \sim \mathcal{N}\{\mathbf{0}, \sigma_u^2 \mathbf{I}_{n-q}\}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (1.8)$$

for  $\boldsymbol{\beta} \in \mathbb{R}^q$ ,  $\mathbf{u} \in \mathbb{R}^{n-q}$  and  $\mathbf{u}$  independent of  $\boldsymbol{\epsilon}$ . The smoothing parameter corresponding to this representation is given by  $\lambda = \sigma^2 / (n\sigma_u^2)$ . Solving (1.8) instead of (1.5) is advantageous because all parameters are estimated in a single step by the maximisation of a likelihood function.

From the previous examples it can be seen that the use of the Demmler-Reinsch basis reduces the computational complexity of the solution from  $O(n^3)$  to  $O(n^2)$  in both cases. The general case where  $\mathbf{R}$  is an arbitrary Toeplitz positive definite matrix is studied in detail in Chapter 3.

## 1.2. Asymptotic Properties

Here we summarise some asymptotic results of smoothing splines following [Schwartz \[2012\]](#). For a more detailed review on the integrated mean squared error, the average mean squared error, optimal rates of convergence or local properties of smoothing spline estimators, the reader is referred to [Rice and Rosenblatt \[1981\]](#), [Wahba \[1990\]](#), [Craven and Wahba \[1978\]](#), [Speckman \[1985\]](#) and [Nychka \[1995\]](#).

**Definition 7.** A positive sequence  $\{a_N\}^\infty$  is called an optimal rate of convergence of estimators on class of functions  $\mathcal{F}$  with respect to a norm  $\|\cdot\|$  if  $\exists c_1, c_2 > 0$  such that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\hat{f} \in \hat{\mathcal{F}}_{n,f}} \sup \mathbb{E} \left[ \|\hat{f}_n - f\|^2 \right] a_n^{-2} &\geq c_1 \\ \limsup_{n \rightarrow \infty} \inf_{\hat{f} \in \hat{\mathcal{F}}_{n,f}} \sup \mathbb{E} \left[ \|\hat{f}_n - f\|^2 \right] a_n^{-2} &\leq c_2 \end{aligned}$$

where  $\hat{\mathcal{F}}_{n,f}$  denotes a class of estimators of  $f$  based on a random sample of size  $n$ .

**Theorem 1.** [Speckman \[1985\]](#)

Let  $Y_{i,f} = f(x_i) + \epsilon_i$ ,  $i = 1, \dots, N$ ,  $f \in \mathcal{F}$ , where  $\{\epsilon_i\}_{i=1}^n$  is an i.i.d. sequence of random variables with variance  $\sigma^2 > 0$ . In [Definition 7](#) consider  $\mathcal{F} = \mathcal{W}^q[0, 1]$  and denote  $\hat{\mathcal{F}}_{f,n}$  as all linear estimators on  $\mathcal{F}$ . Then the optimal rate of convergence of an estimator on  $\mathcal{F}$  with respect to norm  $\|f\|_{n,2} = n^{-1} \{\sum_{i=1}^n f(x_i)^2\}^{1/2}$  is  $O(n^{-q/(2q+1)})$ .

Moreover, the global goodness of estimator  $\hat{f}$  can be measured by the integrated squared mean error

$$\begin{aligned} IMSE(\hat{f}) &:= \mathbb{E} \|\hat{f} - f\|_2^2 = \mathbb{E} \int_0^1 \left\{ \hat{f}(x) - f(x) \right\}^2 dx \\ &= \int_0^1 \left\{ \mathbb{E} \hat{f}(x) - f(x) \right\}^2 dx + \int_0^1 \mathbb{E} \left\{ \hat{f}(x) - \mathbb{E} \hat{f}(x) \right\}^2 dx, \end{aligned} \tag{1.9}$$

or its discrete version, i.e. the average mean squared error

$$\begin{aligned} AMSE(\hat{f}) &:= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{f}(x) - f(x) \right\}^2 dx \\ &= \sum_{i=1}^n \left\{ \hat{f}(x) - f(x) \right\}^2 + \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left\{ \hat{f}(x) - \mathbb{E} \hat{f}(x) \right\}^2. \end{aligned} \quad (1.10)$$

As  $n \rightarrow \infty$ , and given certain assumptions on the regularity of the design points, expressions (1.9) and (1.10) are asymptotically equivalent. Estimator  $\hat{f}$  is asymptotically optimal on  $\mathcal{F}$  with respect to the norm  $\|\cdot\|_2$  if its IMSE (AMSE) has the same order of magnitude as the corresponding squared optimal rate of convergence of estimators on  $\mathcal{F}$ . From Theorem 1 if  $\mathcal{W}^q[0, 1] = \mathcal{F}$ , asymptotically optimal linear estimators have IMSE (AMSE) of order  $O(n^{-2q/(2q+1)})$ .

Since functions  $\{\phi_{q,i}(x)\}_{i=1}^{\infty}$  build a complete orthonormal system in  $\mathcal{W}^q[0, 1]$  under the  $L_2$  inner product, one can write

$$f(x) = \sum_{l=1}^{\infty} f_l \phi_{q,l}(x), \text{ with } f_l = \int_0^1 f(x) \phi_{q,l}(x) dx \text{ and } \sum_{l=1}^{\infty} f_l^2 \eta_{q,l} < \infty,$$

and hence, the IMSE of smoothing splines can be written as

$$IMSE(\hat{f}) = \left[ \sum_{i=q+1}^{\infty} \frac{\lambda^2 \eta_{q,i}^2 f_i^2}{(1 + \lambda \eta_{q,i})^2} + \frac{1}{n} \sum \frac{\sigma^2}{(1 + \lambda \eta_{q,i})^2} + \frac{q\sigma^2}{n} \right] 1 + o(1). \quad (1.11)$$

The global asymptotic error of smoothing splines can then be obtained by looking at the terms in (1.11). The last two summands correspond to the variance and, given that  $\eta_{q,i} = \{(i-q)\pi\}^{2q}$  [cf. Speckman, 1985], are  $O(n^{-1}\lambda^{-1/2q})$ . Regarding the bias, its rate depends on the decay of  $f_i$ . Results presented in Utreras [1980] show that the bias of a function  $f \in \mathcal{W}^q[0, 1]$  decay at different rates depending on the natural boundary conditions. Namely it has been shown that the integrated squared bias has three different bound according to the following cases: i) if all boundary conditions hold  $O(\lambda^2)$ ; ii) if the  $j$ -th boundary condition do not hold  $O(\lambda^{(2j+1)/2q})$ ; and iii) if none

of the boundary conditions hold  $O(\lambda)$ . The global asymptotic error for smoothing splines then vary accordingly to each case. The bounds for the integrated squared bias can then be used to estimate the optimal smoothing parameter  $\lambda$  in each of the cases to obtain the asymptotically optimal IMSE. If all the boundary conditions hold, we obtain

$$IMSE(\hat{f}) = O(\lambda^2) + O(n^{-1}\lambda^{-1/q}),$$

with optimal  $\lambda \asymp n^{-q/(2q+1)}$  leading to  $IMSE(\hat{f}) = O(n^{-2q/(2q+1)})$ . Similarly, if the  $j$ -th boundary condition does not hold

$$IMSE(\hat{f}) = O(\lambda^{(2j+1)/q}) + O(n^{-1}\lambda^{-1/q}),$$

with optimal  $\lambda \asymp n^{-q/(2j+2)}$  providing  $IMSE(\hat{f}) = O(n^{-(2j+1)/(2j+2)})$ . Lastly, if no boundary conditions hold [cf. Craven and Wahba, 1978, Cox, 1983] one obtains

$$IMSE(\hat{f}) = O(\lambda) + O(n^{-1}\lambda^{-1/2q}),$$

with  $\lambda \asymp n^{-2q/(1+2q)}$  and hence  $IMSE(\hat{f}) = O(n^{-2q/(1+2q)})$ .

### 1.3. Open Problems and Objectives of the Thesis

Smoothing splines are a prominent area in mathematical statistics and there are plenty of theoretical results available in the literature, however some open problems remain and are of interest. In this dissertation we investigate the estimation of the degree  $q$  for certain  $f \in \mathcal{W}^q[0, 1]$  under the presence of correlated errors. The importance of the selection of the degree of the Sobolev space  $q$  was explicitly mentioned in Reinsch [1967], but to our knowledge no attempts have been done for its study except for the recent works by Krivobokova [2013] and Serra and Krivobokova [2016]. Moreover, even if the true value of  $q > 3$  is known in advance, the estimation procedure via numerical implementations of the Demmler-Reinsch basis [cf. Ruppert et al., 2003] is known

to be unstable. This situation has lead presumably to the nowadays standard cubic spline assumption, i.e.  $q = 2$ . In this thesis we present solutions to the aforementioned problems and study the following aspects of smoothing splines:

1. Computational stability. It is easy to see that the condition number of the solution (1.5) for an arbitrary  $\mathbf{C}_q$  increases exponentially with  $q$  via  $\mathbf{D}_q$ . This phenomenon has been documented in Wand and Ormerod [2008] for the O'Sullivan [1986] type penalisation. In this dissertation we consider an exact expression for the Demmler-Reinsch basis, where the eigenvectors and eigenvalues are assigned to the columns of  $\mathbf{C}_q$ , and the elements of the diagonal matrix  $\mathbf{D}_q$  respectively, so that computational stability is no longer an issue. In Chapter 2 we present our first contribution, which is an exact expression for the Demmler-Reinsch basis as the solution of certain ordinary differential equation.
2. Computational complexity. As shown in examples 1 and 2, the computational cost of the smoothing spline problem in the independence case can be considerably reduced if the Demmler-Reinsch basis is used. In Chapter 3 we show that this fact extends naturally to the correlated errors case. Moreover, in Chapter 4 we show that the computational efficiency is of crucial interest not only when large amounts of data are considered, but also when the basic model is used as a building block for more involved settings (see point 4 in this list).
3. Correlated Errors. Serra and Krivobokova [2016] introduced the idea of estimating  $q$  together with the smoothing parameter  $\lambda$  for the independence case. An extension of this work to the case where the correlation in the error term decays exponentially is presented in Chapter 3. In the same Chapter comparative simulation studies are presented to study the performance of our method in finite samples.
4. Extensions. The results presented in Chapter 3 are extended to models where empirical Bayes smoothing splines are used as a building block. Namely, in Chapter 4 we present extensions to additive models and functional data analysis.



These are new results and are presented in Chapters 3 and 4. Chapter 5 shows practical applications of these extensions.

This thesis is organised as follows. In Chapter 2 the Demmler-Reinsch basis is revisited as the solution of an ordinary differential equation and an exact expression is provided. Chapter 3 uses the results in chapter 2 and builds a general empirical Bayes smoothing splines model where the degree of the smoothness of the regression function, the structure of the error's correlation, and the smoothing parameter are unknown. Chapter 4 extends the results in Chapter 3 to additive models and functional data analysis. In Chapter 5 applications in agricultural economics and biophysics are presented to illustrate the use of the method. Chapter 6 provides a summarised description of the R software developed to implement the ideas presented in chapters 3 and 4. Chapter 7 closes the document with a summary and an outlook.

## 2. Demmler-Reinsch Basis

The smoothing splines problem can be solved under different bases, e.g. truncated polynomials, B-Splines, Thin-plate Splines, etc. The main advantage of the Demmler-Reinsch basis over the other alternatives is its orthonormality. Numerical approximations of this basis are available in the literature, see [cf. [Ruppert et al., 2003](#), [Wood, 2006](#)], however it is well known that such approximations become numerically unstable, or even computationally infeasible, as the smoothness degree  $q$  in  $\mathcal{W}^q[0, 1]$  increases. In this dissertation we study estimators of  $f \in \mathcal{W}^q[0, 1]$  without assuming any prior knowledge of the degree of the space, and hence it is of crucial importance to have a precise representation of the Demmler-Reinsch basis that allows for a computationally stable and efficient solution of the smoothing splines problem. In this chapter we present an exact expression of this basis as the solution of certain ordinary differential equation.

### 2.1. Ordinary Differential Equation

In this section the Demmler-Reinsch basis for  $\mathcal{W}^q[0, 1]$  is redefined as the solution of an ordinary differential equation. Consider the approximation error of a function  $f \in \mathcal{W}^q[0, 1]$  by the expansion of some arbitrary orthonormal basis  $\{\varphi_{q,i}\}_{i=1}^n$  as

$$D_N(\psi) := \sup_{f \in \mathcal{W}^q[0,1]} \frac{1}{n} \sum_{i=1}^n \left\{ f(x_i) - \sum_{j=1}^N \langle f, \varphi_{q,j} \rangle \varphi_{q,j}(x_i) \right\}^2,$$

where we have used the L-2 inner product. The quantity  $d_N := \inf_{\varphi_q} D_N(\varphi_q)$ , is often called Kolmogorov's diameter. Remarkably, by some simple algebra one can show that  $d_N = D_N(\phi_q) = \eta_{q,N+1}$ . Meaning that the Demmler-Reinsch basis is the ensemble of diagonals in the Sobolev's ellipsoid [cf. [Tikhomirov, 1986](#)] for details.

**Theorem 2.** *Cao [2008]. If an orthonormal basis in  $\mathcal{W}^q[0, 1]$  with eigenfunctions  $\{\varphi_{q,1}(x), \dots\}$  and corresponding eigenvalues  $\nu_{q,1} \dots$  solves the following ordinary differential equation*

$$(-1)^q \varphi_{q,i}^{(2q)}(x) - \nu_i \varphi_{q,i}(x) = 0 \quad \text{s.t.} \quad \varphi_{q,i}^{(l)}(0) = \varphi_{q,i}^{(l)}(1) = 0, \quad (2.1)$$

for  $l = q, q + 1, \dots, 2q - 1$  then it follows that

$$\langle \varphi_{q,j}^{(q)}, \varphi_{q,i}^{(q)} \rangle = \eta_{q,i} \delta_{j,i}.$$

*Proof.* We begin by noting that any  $f \in \mathcal{W}^q[0, 1]$  can be represented as the sum of two orthogonal functions, one of which is a polynomial. Namely

$$f(x) = P_f^{q-1}(x) + \{Ag(x) - \Pi^{q-1}Ag(x)\},$$

where the first term is the projection of  $f(x)$  onto a polynomial space of degree  $q - 1$ , i.e. given the projection operator  $\Pi^{m-1}$  one can write

$$P_f^{q-1}(x) = \Pi^{m-1}f(x) = \sum_{i=0}^{q-1} \langle f, \psi_i \rangle \psi_i(x),$$

for orthogonal polynomials  $\{\psi_i, \dots, \psi_{m-1}\}$ , whose elements are defined recursively, as e.g. in [Favard \[1935\]](#). For the second term define an operator  $A$  such that

$$Ag = \frac{1}{(q-1)!} \int_0^1 (x-t)_+^{q-1} g(x) dt, \quad \|g\| < 1,$$

and note that  $Ag$  itself contains polynomials, which justifies the form of the second

term in (2.1). Kolmogorov's diameter is hence given by

$$d_N = \inf_{\{\phi\} \perp \{\psi\}} \sup \|(Ag - \Pi^{q-1}Ag) - [Ag - \Pi^{q-1}Ag]_N^{\phi_q}\|^2,$$

where  $[f]_N^{\phi_q} = \sum_{k=1}^N \langle f, \phi_{q,k} \rangle \phi_{q,k}(x)$ . Moreover it is easy to see that for all  $k = 1, \dots, n$ , it holds that

$$d_N \phi_{q,k} = A^* A \phi_{q,k} - A^* \Pi^{q-1} A \phi_{q,k},$$

so that taking the  $2q$ -th derivative we obtain  $d_N \phi_{q,i}^{(2q)} = (-1)^q \phi_{q,i}$ , and hence the ordinary differential equation (with no boundary conditions) follows

$$(-1)^q \phi_{q,i}^{(2q)}(x) = \eta_{q,i} \phi_{q,i}(x).$$

The boundary conditions are obtained directly from the evaluation of  $\phi_{q,i}^{(l)}(x)$  at the borders for  $l = q, q+1, \dots, 2q-1$ . The previous ODE with respective border conditions fulfil

$$\begin{aligned} \langle \phi_{q,j}^{(q)}, \phi_{q,i}^{(q)} \rangle &= (-1)^m \langle \phi_{q,j}, \phi_{q,i}^{(2q)} \rangle \\ &= (-1)^m \int_0^1 \phi_{q,j}(x) \phi_{q,i}^{(2q)}(x) dx = \eta_{q,i} \langle \phi_{q,j}, \phi_{q,i} \rangle, \end{aligned}$$

where integration by parts was used  $q$  times. □

The importance of Theorem 2 is evident since it implies that obtaining a general solution for such ordinary differential equation, given the boundary conditions, leads to an explicit expression for the Demmler-Reinsch basis. Hereafter we consider the following definition

**Definition 8.** *A Demmler-Reinsch basis in  $\mathcal{W}^q[0, 1]$  is an orthonormal basis with eigenfunctions  $\{\phi_{q,i}(x)\}_{i=1}^n$  and eigenvalues  $\{\eta_{q,i}\}_{i=1}^n$ , where  $\{\phi_{q,i}(x)\}_{i=1}^q$  corresponds to an orthonormal polynomial basis of degree  $q-1$  such that  $\{\phi_{q,j}(x)\}_{j=1}^q \perp \{\phi_{q,i}(x)\}_{i=q+1}^n$ .*

## 2. Demmler-Reinsch Basis

---

The corresponding eigenvalues fulfil:  $0 = \eta_{q,1} = \dots = \eta_{q,q} < \eta_{q,q+1} \dots < \eta_{q,n}$ , and the eigenfunctions  $\{\phi_{q,i}(x)\}_{i=q+1}^n$  can be obtained as the solution of the ordinary differential equation

$$(-1)^q \phi_{q,i}^{(2q)}(x) - \eta_{q,i} \phi_{q,i}(x) = 0, \quad (2.2)$$

with boundary conditions

$$\phi_{q,i}(x)^{(l)}(0) = \phi_{q,i}^{(l)}(1) = 0, \quad l = q, q+1, \dots, 2q-1 \quad i = q+1, \dots, n. \quad (2.3)$$

The solution of the  $2q$ -order homogeneous linear equation (2.2) with conditions (2.3) has been studied in e.g. Polyanin and Zaitsev [2003]. In particular the characteristic polynomial  $P(\omega) = \omega^{2q} - (-1)^q \eta_{q,i} = 0$  has only complex conjugate roots if  $q \in 2\mathbb{N} - 1$  and a mixture of complex conjugate and real roots if  $q \in 2\mathbb{N}$ . Under this setting the eigenvectors can be written as

$$\begin{aligned} \phi_{q,i}(x) = & \sum_{\lambda_j \in \mathcal{S}(q)} \alpha_{q,j,i} \exp(-\lambda_j \mu_{q,i} x) + \sum_{\lambda_j \in \mathcal{S}(q)} \beta_{q,j,i} \exp(\lambda_j \mu_{q,i} x) \\ & + \gamma_{q,i} \cos(\mu_{q,i} x) + \delta_{q,i} \sin(\mu_{q,i} x), \end{aligned} \quad (2.4)$$

for  $\mu_{q,i} = \eta_{q,i}^{1/2q}$  and  $\mathcal{S}(q) := \cup_j \left\{ (-1)^{j/2q}, \overline{(-1)^{j/2q}} \right\}$  such that  $0 \leq j \leq q-2$  and  $j \in 2\mathbb{N} - 1$  if  $q$  is odd and  $j \in 2\mathbb{N}$  otherwise so that  $\#\{\mathcal{S}(q)\} = q-1$ . The solution for  $\mathbf{r} = \{\alpha_{q,1,i}, \dots, \alpha_{q,q-1,i}, \beta_{q,1,i}, \dots, \beta_{q,q-1,i}, \gamma_{q,i}, \delta_{q,i}\}^T$  can then be obtained by utilising the boundary conditions (2.3), which reduces the problem to solve the  $2q \times 2q$  linear system  $\mathbf{M}_{q,i} \cdot \mathbf{r}_i = \mathbf{0}$  together with the computation of the eigenvalues. The following example illustrates the problem.

**Example 3.** For  $q = 2$  the solution is given by

$$\phi_{2,i}(x) = \alpha_{2,1} \exp(-\mu_{2,i} x) + \beta_{2,1} \exp(\mu_{2,i} x) + \gamma_{2,1} \cos(\mu_{2,i} x) + \delta_{2,1} \sin(\mu_{2,i} x),$$

and we are interested in finding the null-space of

$$\begin{pmatrix} -1 & -1 & 1 & 0 \\ 1 & e^{2\mu_{2,i}} & -e^{\mu_{2,i}} \cos(\mu_{2,i}) & -e^{\mu_{2,i}} \sin(\mu_{2,i}) \\ 1 & -1 & 0 & 1 \\ -1 & e^{2\mu_{2,i}} & e^{\mu_{2,i}} \sin(\mu_{2,i}) & -e^{\mu_{2,i}} \cos(\mu_{2,i}) \end{pmatrix} \quad (2.5)$$

Since by definition  $\mathbf{M}_{2,i}$  is not full rank,  $\mu_{2,i}$  can be computed as the implicit solution  $\tilde{\mu}_{2,i} : |\mathbf{M}_{2,i}| = 0$ , so that  $\mathbf{r}_i$  solves  $\mathbf{M}_{2,i}|_{\mu_{2,i}=\tilde{\mu}_{2,i}} \cdot \mathbf{r}_i = \mathbf{0}$  and is such that  $\|\phi_{2,i}(x)\|_2 = 1$ .

In the next section we propose a method to circumvent the aforementioned procedure and obtain a direct formula for the Demmler-Reinsch basis.

## 2.2. A General Demmler-Reinsch Basis

In the first subsection we will show a general form to approximate the eigenvalues of a Demmler-Reinsch basis of an arbitrary degree. In the second subsection we use this results to present an expression for the eigenvectors.

### 2.2.1. Eigenvalues

In general for  $q \in \mathbb{N}$  and corresponding sequences  $\{\mu_{q,i}\}_{i=q+1}^n$  the procedure depicted in example 3 must be followed  $n - q$  times. However our problem has a structure that can be further exploited. Consider first the problem of computing the eigenvalues as the roots of the determinant for linear systems  $\mathbf{M}_{q,i}$ , with  $q$  small, i.e.  $q = 1, \dots, 4$ . Here we present the results

$$\begin{aligned}
|\mathbf{M}_{1,i}| &= -\sin(\mu_{1,i}) \\
|\mathbf{M}_{2,i}| &= -4e^{2\mu_{2,i}} (\cos(\mu_{2,i}) \cosh(\mu_{2,i}) - 1) \\
|\mathbf{M}_{3,i}| &= 6i\mu_{3,i}^{12} \sin\left(\frac{\mu_{3,i}}{2}\right) \left[ 2 \cos\left(\frac{\mu_{3,i}}{2}\right) \left\{ -\cos(\mu_{3,i}) + \cosh\left(\sqrt{3}\mu_{3,i}\right) + 4 \right\} - \right. \\
&\quad \left. 8 \cosh\left(\frac{1}{2}\sqrt{3}\mu_{3,i}\right) \right] \\
|\mathbf{M}_{4,i}| &= -64\mu_{4,i}^{22} \left[ \cos\left(\sqrt{2}\mu_{4,i}\right) + \cosh\left(\sqrt{2}\mu_{4,i}\right) + 2\sqrt{2} \sinh(\mu_{4,i}) \times \right. \\
&\quad \left\{ 2 \cos\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) \sinh\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) - \cos(\mu_{4,i}) \sinh\left(\sqrt{2}\mu_{4,i}\right) \right\} - \\
&\quad 4 \cosh\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) \left\{ \sqrt{2} \sin\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) \sin(\mu_{4,i}) + 2 \cos\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) \cos(\mu_{4,i}) \right\} + \\
&\quad \cosh(\mu_{4,i}) \left\{ 2\sqrt{2} \sin(\mu_{4,i}) \sin\left(\sqrt{2}\mu_{4,i}\right) + 3 \cos(\mu_{4,i}) \cos\left(\sqrt{2}\mu_{4,i}\right) - \right. \\
&\quad \left. 8 \cos\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) \cosh\left(\frac{\mu_{4,i}}{\sqrt{2}}\right) + \cos(\mu_{4,i}) \left( 3 \cosh\left(\sqrt{2}\mu_{4,i}\right) + 2 \right) \right\} + 6 \left. \right].
\end{aligned}$$

As it can be seen the determinant functions (and consequently their roots) get more involved as  $q$  increases. In fact the computational complexity of the determinant for a matrix  $\mathbf{M}_{q,i}$  of size  $2q \times 2q$  using the LU decomposition or the Turing factorisation takes  $O(q^6)$ , see. e.g [Golub and van Loan \[1996\]](#) and [Corless and Jeffrey \[1997\]](#) for details on the symbolic matrix case. We proceed by showing the estimation of  $\mu_{q,i}$  for these cases, provide a formula for  $i \rightarrow n$  and compare it with numerical approximations.

For the case  $q = 1$  we obtain the trivial solution  $\mu_{1,i} = (i - 1)\pi$ , while for the case  $q = 2$  one is already constrained to solutions as  $i \rightarrow n$ . Namely since  $\cosh(\mu_{2,i}) \neq 0$  for all  $\mu_{2,i} > 0$  and  $1/\cosh(\mu_{2,i}) \rightarrow 0$  as  $i \rightarrow n$ , it is clear that  $\mu_{2,i} = (i - 3/2)\pi$ . For the last case it is obvious that if  $\mu_{3,i}/\pi \in 2\mathbb{N}$  the roots read  $(i - 2)\pi$  and it can be easily shown that the same solution can be extended to  $\mu_{3,i}/\pi \in 2\mathbb{N} - 1$  as  $i \rightarrow n$ . Similar arguments follow for the case  $q = 4$ , where the following roots are found

$\mu_{4,i} = (i - 5/2)\pi$ . Given the previous results we conjecture

$$\lim_{i \rightarrow n} \eta_{q,i} = \left\{ \left( i - \frac{q+1}{2} \right) \pi \right\}^{2q}, \quad i = q+1, \dots, n. \quad (2.6)$$

The comparison of the implementation of (2.6) and numerical approximations via Newton search algorithms is presented in Table (2.1). As it can be seen the approximation is accurate up to the fourth decimal before the first 10 eigenvalues are computed. We do not report the results for  $q = 1$  since the result is exact and no numerical approximations are required. The comparisons for  $q = \{6, 7, 8, 9, 10\}$  were also performed reporting results in the same direction. The approximation formula (2.6) is compatible with Theorem 2.2 given in Speckman [1985], where it is shown that

$$\eta_{q,i} = \{(i - q)\pi\}^{2q} \{1 + o(1)\}, \quad \eta_{q,1} = \dots = \eta_{q,q} = 0, \quad i = q+1, \dots, n,$$

where the  $o(1)$  term is uniform over  $i = o(n^{2/(2q+1)})$  as  $n \rightarrow \infty$ . In this subsection we give a more precise expression of the eigenvalues.

**Theorem 3.** *The eigenvalues of a Demmler-Reinsch basis can be approximated by*

$$\lim_{n \rightarrow \infty} \eta_{q,i} = \left\{ \left( i - \frac{q+1}{2} \right) \pi \right\}^{2q}, \quad (2.7)$$

for  $q = 1, \dots, 6$ , and  $i = q+1, \dots, n$ .

*Proof.* Proceed by contradiction and take  $\mu_{q,i} = (i - (q+1)/2)\pi + \epsilon$ , for an arbitrarily small  $\epsilon \in \mathbb{R}$ . A straightforward Taylor expansion for the determinants around this point has the form

$$|\mathbf{M}_{q,i}| = m_q(\epsilon) + \sum_{i=1}^q \omega_i \left\{ \mu_{q,i} - \left( i - \frac{q+1}{2} \right) \pi - \epsilon \right\} + O(\epsilon^3),$$

for continuous functions  $m_q(\cdot)$ ,  $\omega_i \in \mathbb{R}$  and where the first two terms of the expansion



2. Demmler-Reinsch Basis

Table 2.1.: Roots in Determinant Functions\*

		$q = 2$		$q = 3$	
i	Roots (num. approx.)	Roots (Thm. 3)	Roots (num. approx.)	Roots (Thm. 3)	
3	<b>4.73004</b>	<b>4.71239</b>	–	–	
4	<b>7.8532</b>	<b>7.85398</b>	<b>9.42706</b>	<b>9.42478</b>	
5	10.9956	10.9956	12.5664	12.5664	
6	14.1372	14.1372	15.708	15.708	
7	17.2788	17.2788	18.8496	18.8496	
8	20.4204	20.4204	21.9911	21.9911	
9	23.5619	23.5619	25.1327	25.1327	
10	26.7035	26.7035	28.2743	28.2743	
11	29.8451	29.8451	31.4159	31.4159	
12	32.9867	32.9867	34.5575	34.5575	
13	36.1283	36.1283	37.6991	37.6991	
14	39.2699	39.2699	40.8407	40.8407	
15	42.4115	42.4115	43.9823	43.9823	
16	45.5531	45.5531	47.1239	47.1239	
17	48.6947	48.6947	50.2655	50.2655	

		$q = 4$		$q = 5$	
i	Roots (num. approx.)	Roots (Thm. 3)	Roots (num. approx.)	Roots (Thm. 3)	
5	<b>10.9958</b>	<b>10.9956</b>	–	–	
6	<b>14.1377</b>	<b>14.1372</b>	<b>12.5578</b>	<b>12.5664</b>	
7	17.2788	17.2788	<b>15.7077</b>	<b>15.708</b>	
8	20.4204	20.4204	<b>18.8497</b>	<b>18.8496</b>	
9	23.5619	23.5619	<b>21.9912</b>	<b>21.9911</b>	
10	26.7035	26.7035	25.1327	25.1327	
11	29.8451	29.8451	28.2743	28.2743	
12	32.9867	32.9867	31.4159	31.4159	
13	36.1283	36.1283	34.5575	34.5575	
14	39.2699	39.2699	37.6991	37.6991	
15	42.4115	42.4115	40.8407	40.8407	
16	45.5531	45.5531	43.9823	43.9823	
17	48.6947	48.6947	47.1239	47.1239	
18	51.8363	51.8363	50.2655	50.2655	
19	54.9779	54.9779	53.4071	53.4071	

\* The bolded rows highlight the cases  $i$  where the methods differ.

are used. Since  $m_q(\epsilon) = 0$  only when  $\epsilon = 0$ , the result follows.  $\square$

**Remark 1.** *In the context of periodic smoothing splines the resulting eigenvalues of the periodic Demmler-Reinsch basis should be normalised [cf. Schwartz, 2012] where the regression and penalised splines cases are also explored. This normalisation applies directly to the Fourier coefficients in the discrete approximation of the periodic Demmler-Reinsch basis following the theory of attenuation factors presented in Gautschi [1971], where it is shown that such factors arise whenever the approximation process that maps the ( $2\pi$  periodic) data into a periodic basis is linear and invariant under translation. Given certain conditions Gutknecht [1987] shows that it is possible to obtain attenuation factors for arbitrary data by utilising a combination of the translates of periodically extended box splines. In this chapter we considered a non-periodic basis and to our knowledge attenuation factors for these type of bases are not available in the literature.*

### 2.2.2. Eigenvectors

Given the convergence of the roots of the determinant functions in (2.6), the convergence of the null-space of systems  $\mathbf{M}_{q,i}(x)$  follows and the Demmler-Reinsch basis problem is solved up to a normalisation constant.

**Theorem 4.** *Given eigenvalues  $\eta_{q,i}$  as in Theorem 3, the corresponding eigenfunctions  $\{\phi_{q,i}(x)\}_{i=q+1}^n$  of the Demmler-Reinsch basis are given by*

$$\phi_{q,i}(x) = \sqrt{2} \left( \sum_{\lambda_j \in \mathcal{S}(q)} r_j \{ e^{-\lambda_j \mu_{q,i} x} + (-1)^{i+1} e^{-\lambda_j \mu_{q,i} (1-x)} \} \right) \quad (2.8)$$

$$+ \cos \left\{ \mu_{q,i} x + \left( \frac{q-1}{4} \right) \pi \right\}, \quad i = q+1, \dots, n, \quad (2.9)$$

where  $\mu_{q,i} = \eta_{q,i}^{1/2q}$  and  $\mathcal{S}(q) := \cup_j \{ (-1)^{j/2q}, \overline{(-1)^{j/2q}} \}$  such that  $0 \leq j \leq q-2$  and  $j \in 2\mathbb{N} - 1$  if  $q$  is odd and  $j \in 2\mathbb{N}$  otherwise.

*Proof.* We write the solution for  $i \rightarrow n$ , where the approximation of the eigenvalues as (2.6) holds, which implies that the coefficients in given in (2.4) for the solution of the ODE (2.2) can be written as quantities independent of  $i$ . From the boundary conditions (2.3) it follows immediately that  $\gamma_q = \delta_q$  and hence  $\beta_{q,j} = \alpha_{q,j}(-1)^{i+1} \exp\{-\lambda_j \mu_{q,i}\}$ , from which the condition  $\|\phi_{q,i}(t)\| = 1$  and basic trigonometric formulae lead to the result. □

### 2.3. Comparison with Numerical Approximations

In this section we compare the exact expressions for the eigenvalues and eigenvectors for the Demmler-Reinsch basis provided in Section 2.2, with a commonly used numerical approximation. Specifically we set  $\mathbf{C}_q$  in (1.6) to be a B-spline basis of degree  $2q - 1$  with the natural conditions at the borders and a penalisation matrix of the O’Sullivan [1986] type constructed from the Schumaker [2007] implementation. We then express the numerical approximation of the Demmler-Reinsch basis following the (standard) algorithm presented in Ruppert et al. [2003].

As mentioned in Section 1.3, the numerical instability of formulation (1.6) under the O’Sullivan [1986] type penalisation can be a severe problem. To circumvent this difficulty numerical approximations of the Demmler-Reinsch basis have been proposed and are available in the literature. The standard algorithm first computes the so-called Reinsch form of the smoother matrix by setting  $\mathbf{C}_q = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  in (1.6) to obtain  $\mathbf{S}_{\lambda,q,\mathbf{I}} = (\mathbf{I} + \lambda n \mathbf{K})^{-1}$  for

$$\mathbf{K} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^T \mathbf{D}_q \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T. \tag{2.10}$$

The spectral decomposition  $\mathbf{K} = \tilde{\mathbf{\Phi}}_q \text{diag}(\tilde{\eta}_{q,i}) \tilde{\mathbf{\Phi}}_q^T$  then retrieves the numerical approximations of the eigenvectors and eigenvalues of the Demmler-Reinsch basis, so that

the regression function can be estimated with (1.7) plugging in the corresponding numerical approximations.

Figure (2.1) shows a panel with a comparison between the eigenvectors when computed numerically and by expression (2.8) for a grid of  $n = 60$  equidistant points, revealing that the numerical approximation is unstable already for  $q \geq 3$  (see third and fourth rows in the panel). Specifically, the deterioration of the eigenvectors  $\phi_{q,i}(t)$  is noticeable at the borders of the unit interval and becomes more severe as  $q$  increases. Moreover such features become prominent as the sample size increases, to the point where the computation of the eigenfunctions is simply infeasible, which is clearly a problem induced by (2.10), i.e. in the spectral decomposition of  $\mathbf{C}_q$  and the computation of  $\mathbf{D}_q$ .

Other alternatives for the construction of the penalisation matrix are the [Wand and Ormerod \[2008\]](#) approach, where an exact matrix algebraic expression is derived for splines of arbitrary order, or the direct difference penalty suggested by [Eilers and Marx \[1996\]](#). In the former case the authors consider the general setting where the number of knots are less than or equal to the sample size, however we found instability issues in the penalisation matrix itself as  $q$  increase and as the number of knots approach the sample size, i.e. for smoothing splines. On the other hand, the [Eilers and Marx \[1996\]](#) so-called difference penalty have the nice property that is stable for an arbitrary  $q$ . Nonetheless, as pointed out by [Wand and Ormerod \[2008\]](#), its smoothers can have erratic extrapolation behaviour due to the differences between the exact integral penalty and its discrete approximation.

Figure (2.2) highlights the differences of the eigenvalues under the two proposed methods for the same grid of  $n = 60$  equidistant points. It is interesting to note that even though the attenuation factors for periodic smoothing splines presented in e.g. [Schwartz \[2012\]](#) are not directly applicable to our results, as explained in Remark 1, Figure (2.2) suggests that a correction of this form is remarkably accurate when making the comparison between the eigenvalues obtained directly from the ODE and

## 2. Demmler-Reinsch Basis

---

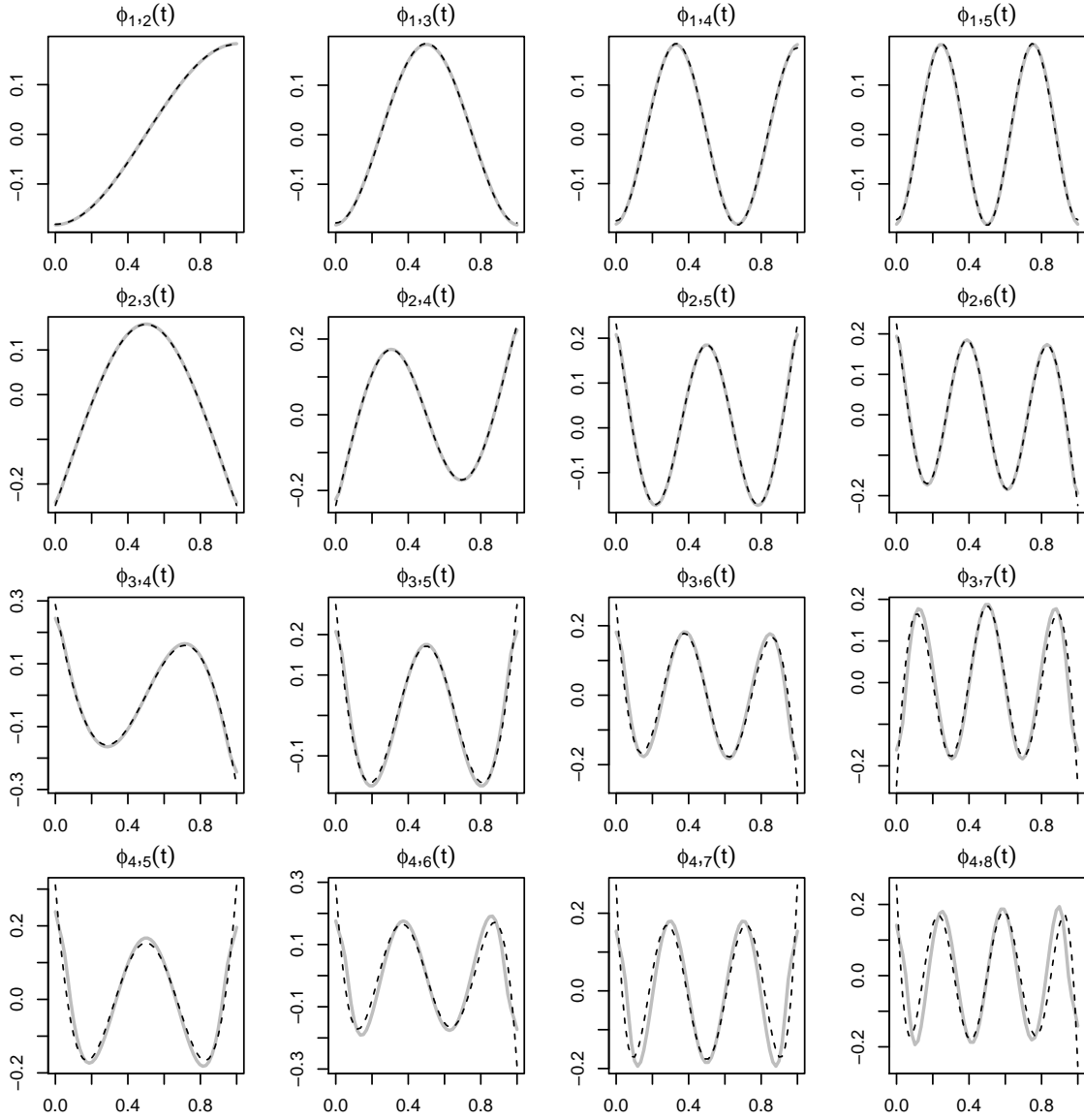


Figure 2.1.: Demmler-Reinsch Eigenfunctions. The continuous grey lines show the DR basis using the numerical approximation described in section 2.3. The black dashed lines correspond to the DR basis derived from the ODE presented in section 2.1. In both cases a sample size of  $n = 60$  was considered. The solutions for larger values of  $q$  are not reported because the numerical approximation becomes too unstable.

the numerical approximation presented in [Ruppert et al. \[2003\]](#). For completeness, here we present the attenuation factors for a periodic spline of degree  $2q - 1$  that were used to produce Figure (2.2). These are factors of the form

$$\tau_i = \frac{\text{sinc}(\pi i/N)^{2q}}{Q_{2q-2}(i/N)}, \quad \text{with} \quad Q_{2q-2}(z) := \sum_{l=-\infty}^{\infty} \text{sinc}\{\pi(z+l)\}^{2q},$$

where  $Q(\cdot)$  are the so-called Q-polynomials of [Schwartz \[2012\]](#). The interested reader can refer to [Gautschi \[1971\]](#) and references therein, where the general theory of attenuation factors is presented.

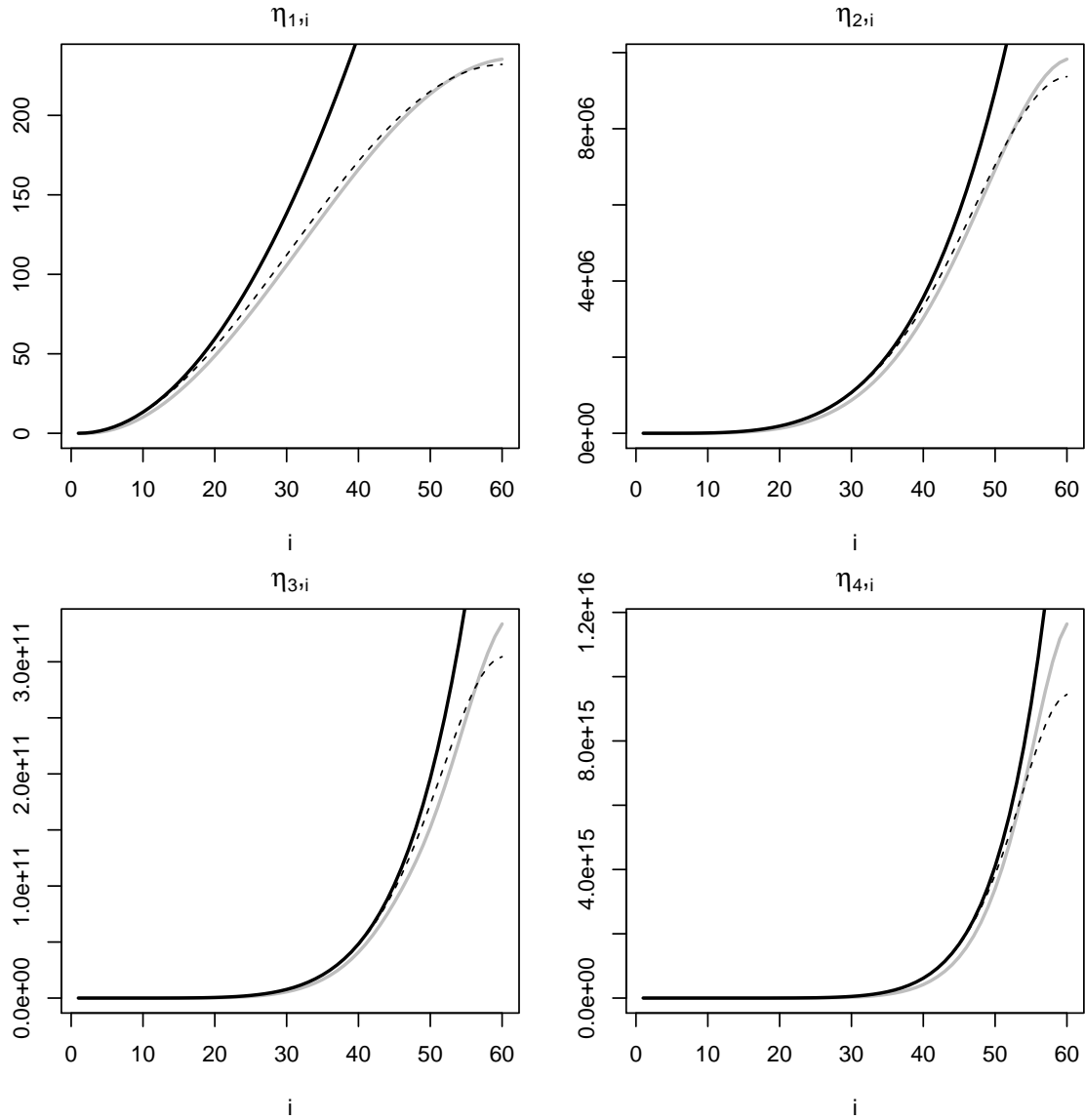


Figure 2.2.: Demmler-Reinsch Eigenvalues. The continuous grey lines show the DR basis using the numerical approximation described in section 2.3. The black dashed lines correspond to the DR basis derived from the ODE presented in section 2.1 when attenuation factors are considered. The solid black line shows the case when attenuation factors are omitted. In all cases a sample size of  $n = 60$  was considered. The solutions for larger values of  $q$  are not reported because the numerical approximation becomes too unstable.

# 3. Smoothing Splines with Correlated Errors

In this chapter we consider a likelihood based method for estimating the unknown function  $f \in \mathcal{W}^q[0, 1]$ , its smoothness class  $q$ , smoothing parameter  $\lambda$ , noise level  $\sigma^2$ , and correlation matrix of the noise  $\mathbf{R}$ . The approach consists of endowing the regression function (given  $\sigma^2$  and  $\mathbf{R}$ ) with a so-called partially informative Gaussian prior [cf. [Speckman and Sun, 2003](#)]. The specific prior depends on  $\lambda$ ,  $q$  and  $\sigma^2$  but is constant over  $\mathbf{R}$ . Under this prior, the data can be seen as a realisation of a linear mixed model (LMM) whose mean has as best linear unbiased predictor (BLUP) a smoothing spline with smoothing parameter  $\lambda$  and order  $q$ . The estimation of  $q$ ,  $\lambda$ ,  $\sigma^2$  and  $\mathbf{R}$  is then performed by the maximisation of the restricted profile log-likelihood function. Such maximisers are in fact empirical Bayes estimates for these parameters.

We propose a fully non-parametric method to estimate all model parameters by an iterative procedure consisting of two loops. In the inner loop  $\sigma^2$ ,  $\mathbf{R}$  and  $\lambda$  are estimated iteratively given  $q$  until convergence of  $\lambda$  is achieved. The outer loop repeats the previous iteration for different values of  $q = 1, 2, \dots$  and selects the value that maximises the restricted profiled log-likelihood (given known values for  $\lambda$ ,  $\sigma^2$  and  $\mathbf{R}$ ) for each  $q$ . We provide convergence properties of the estimators and shown that consistent estimates for  $\lambda$ ,  $\sigma^2$ , and  $\mathbf{R}$ , given certain  $q$ , can already be obtained after the first iteration. The finite sample properties of the method are studied in a Monte Carlo simulation study.



### 3.1. Statistical Model

We aim to solve the smoothing spline problem stated in Definition 6 under gaussianity. Namely, for given data pairs  $(x_i, Y_i)$  following

$$Y_i = f(x_i) + \epsilon_i,$$

where  $\{\epsilon_i\}_{i=1}^n$  is an homoscedastic stationary process with correlation matrix  $\mathbf{R}$  of elements  $\text{Cor}(\epsilon_i, \epsilon_j) = r_{i,j}$ . The estimation of  $f(\cdot)$  can be written as a the optimisation problem presented in (1.5), that is

$$\min_{f \in \mathcal{W}^q[0,1]} \left[ \frac{1}{n} \sum_{i,j=1}^n \{Y_i - f(x_i)\} r_{i,j}^{-1} \{Y_j - f(x_j)\} + \lambda \int_0^1 \{f^{(q)}(x)\}^2 dx \right],$$

where  $\lambda \in \mathbb{R}_+$ ,  $q \in \mathbb{N}$ . As stated in Chapter 1, the solution (1.6) can be written as  $\hat{f}_{\lambda,q,\mathbf{R}} = \mathbf{S}_{\lambda,q,\mathbf{R}} \mathbf{Y}$  where

$$\mathbf{S}_{\lambda,q,\mathbf{R}} = \mathbf{C}_q (\mathbf{C}_q^T \mathbf{R}^{-1} \mathbf{C}_q + \lambda n \mathbf{D}_q)^{-1} \mathbf{C}_q^T \mathbf{R}^{-1}, \quad (3.1)$$

is the *natural* smoother matrix. In a similar fashion we define a *naive* smoother matrix

$$\begin{aligned} \mathbf{S}_{\lambda,q,\mathbf{R}}^* &= \mathbf{R}^{1/2} \mathbf{C}_q (\mathbf{C}_q^T \mathbf{C}_q + \lambda n \mathbf{D}_q)^{-1} \mathbf{C}_q^T \mathbf{R}^{-1/2} \\ &= \mathbf{R}^{1/2} \mathbf{S}_{\lambda,q,\mathbf{I}} \mathbf{R}^{-1/2}, \end{aligned} \quad (3.2)$$

so that the interpretation of the latter is clear. The smoother first pre-whitens the data by pre-multiplying with the inverse of the square root of the correlation matrix, then applies the smoother for independent data, to finally undo the initial transformation. Moreover for  $\mathbf{C}_q^* = \mathbf{R}^{1/2} \mathbf{C}_q$ , it holds that

$$\mathbf{S}^* = \mathbf{C}_q^* (\mathbf{C}_q^{*T} \mathbf{R}^{-1} \mathbf{C}_q^* + \lambda n \mathbf{D}_q)^{-1} \mathbf{C}_q^{*T} \mathbf{R}^{-1},$$

meaning that the naive smoother is also a natural smoother for the same penalty matrix, but a different choice of design matrix  $\mathbf{C}_q$  (a choice which depends on the correlation matrix  $\mathbf{R}$ ). The two smoothers only coincide when  $\mathbf{R} = \mathbf{I}$ , which means that if the data are correlated, the same choice of  $\lambda$  and  $q$  for the two smoothers leads to different estimates. Reciprocally, given an estimate  $\hat{\mathbf{f}} = \mathbf{S}_{\lambda,q,\mathbf{R}}\mathbf{Y}$  one could ask if there exists  $\lambda^*$  such that  $\hat{\mathbf{f}}^* = \mathbf{S}_{\lambda^*,q,\mathbf{R}}^*\mathbf{Y} = \hat{\mathbf{f}}$ . This is indeed the case so that (3.1) and (3.2) are simply two different parameterisations (in terms of  $\lambda$ ) of the same estimator. The smoother matrix (3.1) is more natural when studying the asymptotic behaviour of our estimators for  $\lambda$  and  $q$ , while the smoother matrix (3.2) is more appropriate for the numeric implementation. Namely, using the Demmler-Reinsch basis (3.2) can be re-written as

$$\mathbf{S}_{\lambda,q,\mathbf{R}}^* = \mathbf{R}^{1/2}\Phi_q \text{diag} \{ (1 + \lambda n \eta_{q,i})^{-1} \} \Phi_q^T \mathbf{R}^{-1/2}, \quad (3.3)$$

where the computational complexity of the solution lies on the estimation of  $\mathbf{R}$ .

## 3.2. Estimators

We aim to estimate the regression function  $f \in \mathcal{W}^q[0, 1]$  via estimators for  $\lambda$ ,  $q$ ,  $\sigma^2$  and  $\mathbf{R}$ . However, there is a natural interdependence between  $\lambda$ ,  $q$ ,  $\sigma^2$  and  $\mathbf{R}$  so that these estimates cannot be attained directly. In particular, the estimation of  $f$  requires a reasonable estimate of  $\mathbf{R}$  and, conversely, the estimation of  $\mathbf{R}$  needs a good estimate of  $\mathbf{f}$  (and  $\sigma^2$ ), which creates a vicious circle. In this section we present estimators for  $\sigma^2$ ,  $\mathbf{R}$ ,  $\lambda$ , and  $q$  that can be interpreted as empirical Bayes estimators retrieved from an iterative maximisation procedure of the resulting marginal likelihood function.

### 3.2.1. Empirical Bayes Function

Consider the case where the design matrix  $\mathbf{C}_q$  is the Demmler-Reinsch basis. As presented in Example 2, in this case it is easy to see that  $\mathbf{X}(x) = \{\phi_{q,1}(x), \dots, \phi_{q,q}(x)\}$  and  $\mathbf{Z} = \{\eta_{q,q+1}^{-1/2}\phi_{q,q+1}(x), \dots, \eta_{q,n}^{-1/2}\phi_{q,n}(x)\}$  are the design matrices corresponding to the LMM representation of the smoothing splines problem. To estimate  $\sigma^2$ ,  $\mathbf{R}$ , and the spline parameters  $\lambda$  and  $q$  we use the empirical Bayes method by endowing  $\mathbf{f}$  with a prior and estimating the remaining model parameters from the respective marginal likelihood

$$\mathbf{f} \sim \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \quad \text{where } \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_{n-q}), \quad (3.4)$$

for  $\boldsymbol{\beta} \in \mathbb{R}^q$ ,  $\mathbf{u} \in \mathbb{R}^{n-q}$  and  $\mathbf{u}$  independent of  $\boldsymbol{\epsilon}$ . This is a partially informative Gaussian prior whose density is given by

$$\pi(\mathbf{f}|\sigma^2, \lambda, q) \propto \left| \frac{\mathbf{R}^{-1}(\mathbf{S}^{-1} - \mathbf{I})}{\sigma^2} \right|_+^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{f}^T \mathbf{R}^{-1}(\mathbf{S}^{-1} - \mathbf{I}) \mathbf{f} \right\}, \quad (3.5)$$

where  $|\cdot|_+$  denotes the product of the non-zero eigenvalues of the argument, and it should be noted that the prior does not depend on  $\mathbf{R}$ . This follows directly from the identity  $\mathbf{S}_R^{-1} - \mathbf{I} = \mathbf{R}(\mathbf{S}_I^{-1} - \mathbf{I})$ . Moreover under (3.4),  $\mathbf{Y}$  is a realisation from the following LMM

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_{n-q}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}) \quad (3.6)$$

where the best linear unbiased predictor  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\mathbf{u}}^T)^T$  of  $\boldsymbol{\theta}$  is known explicitly. Namely given  $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{Z}^T/(\lambda n)$ , it holds that

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \quad \text{and} \\ \hat{\mathbf{u}} &= (\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \lambda n \mathbf{I}_{n-q})^{-1} \mathbf{Z}^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}). \end{aligned} \quad (3.7)$$

In particular  $\hat{\mathbf{f}} = \mathbf{S}\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$ , that is, the solution coincides with the posterior mean corresponding to the prior (3.4). Now consider the estimation of  $\sigma^2$  from the

relation between the log-likelihood  $\ell_{\text{LMM}} = \ell_{\text{LMM}}(\sigma^2, \lambda, q, \mathbf{R})$  and the restricted log-likelihood  $\ell_{\text{RES}} = \ell_{\text{RES}}(\hat{\boldsymbol{\beta}}, \sigma^2, \lambda, q, \mathbf{R})$  of model (3.6), that is

$$\begin{aligned} \ell_{\text{RES}} &= \ell_{\text{LMM}} - \frac{1}{2} \log |\mathbf{X}^T (\sigma^2 \mathbf{V})^{-1} \mathbf{X}| \\ &= -\frac{n}{2} \log(\sigma^2) + \frac{1}{2} \log |\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})|_+ - \frac{1}{2\sigma^2} \mathbf{Y}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{Y}, \end{aligned} \quad (3.8)$$

where it is clear that the maximum with respect to  $\sigma^2$  (given  $\lambda$ ,  $q$  and  $\mathbf{R}$ ) can be obtained explicitly as

$$\hat{\sigma}^2 = \hat{\sigma}_{\lambda, q, \mathbf{R}}^2 = \mathbf{Y}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{Y}/n, \quad (3.9)$$

which can be plugged into (3.8) to obtain the restricted profile log-likelihood

$$\ell(\lambda, q, \mathbf{R}) = -\frac{n}{2} \log(\hat{\sigma}^2) + \frac{1}{2} \log |\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})|_+, \quad (3.10)$$

so that the estimates of  $\lambda$ ,  $q$  and  $\mathbf{R}$  are maximisers of this restricted profile log-likelihood.

As mentioned before, for computational purposes it is convenient to write the restricted log-likelihood in (3.8) in terms of the naive estimator. Denote by  $\mathbf{Y}^* = \mathbf{R}^{-1/2}\mathbf{Y}$  the pre-whitened data and let  $\ell_{\text{RES}}(\sigma^2, \lambda, q, \mathbf{I}; \mathbf{Y}^*)$  represent the respective restricted log-likelihood (with the dependence on the data made explicit) of the pre-whitened model  $\mathbf{Y}^* = \mathbf{f}^* + \boldsymbol{\epsilon}^*$ , with  $\boldsymbol{\epsilon}^* = \mathbf{R}^{-1/2}\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Straightforward matrix manipulations show that

$$\ell_{\text{RES}}(\sigma^2, \lambda, q, \mathbf{I}; \mathbf{Y}^*) = \ell_{\text{RES}}^*(\sigma^2, \lambda, q, \mathbf{R}; \mathbf{Y}) + \frac{1}{2} \log |\mathbf{R}|,$$

where  $\ell_{\text{RES}}^*(\sigma^2, \lambda, q, \mathbf{R}; \mathbf{Y})$  is exactly  $\ell_{\text{RES}}(\sigma^2, \lambda, q, \mathbf{R}; \mathbf{Y})$  from (3.8) with the natural smoother  $\mathbf{S}$  replaced with the naive smoother  $\mathbf{S}^*$ . Likewise,

$$-2\ell(\lambda, q, \mathbf{I}; \mathbf{Y}^*) = -2\ell^*(\lambda, q, \mathbf{R}; \mathbf{Y}) - \log |\mathbf{R}|.$$

We conclude that if for each  $q$  and  $\mathbf{R}$ ,  $\hat{\lambda}_{q,\mathbf{R}}$  and  $\hat{\lambda}_{q,\mathbf{R}}^*$  maximise  $\ell(\lambda, q, \mathbf{R}; \mathbf{Y})$  and  $\ell(\lambda, q, \mathbf{I}; \mathbf{Y}^*)$  respectively, then  $\hat{\mathbf{f}}_{\hat{\lambda}_{q,\mathbf{R}}}$  and  $\hat{\mathbf{f}}_{\hat{\lambda}_{q,\mathbf{R}}^*}$  coincide. The values of  $\hat{\lambda}_{q,\mathbf{R}}$  and  $\hat{\lambda}_{q,\mathbf{R}}^*$  will however be different, but can be related. Similarly, the corresponding estimator for  $\sigma^2$  is the estimator  $\hat{\sigma}^2$  with  $\mathbf{S}$  replaced with  $\mathbf{S}^*$ , that is

$$\hat{\sigma}^{*2} = \hat{\sigma}_{\lambda,q,\mathbf{R}}^{*2} = \mathbf{Y}^{*T}(\mathbf{I} - \mathbf{S}_I)\mathbf{Y}^*/n = \mathbf{Y}^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}^*)\mathbf{Y}/n.$$

For fixed  $q$  and  $\mathbf{R}$  the estimators  $\hat{\sigma}_\lambda^2$  and  $\hat{\sigma}_\lambda^{*2}$  are different, but they coincide when  $\lambda$  is set to the maximisers  $\hat{\lambda}_{q,\mathbf{R}}$  and  $\hat{\lambda}_{q,\mathbf{R}}^*$ , respectively. In practice, maximising  $\ell(\lambda, q, \mathbf{R}; \mathbf{Y})$  or  $\ell(\lambda, q, \mathbf{I}; \mathbf{Y}^*)$  directly to obtain estimates for  $\lambda$ ,  $q$ , and  $\mathbf{R}$  is not practical, so in the next subsections we define estimating equations that can be solved for this purpose.

### 3.2.2. Smoothing Parameter

Let  $\gamma$  represent  $\lambda$ ,  $q$ , or some parameter of  $\mathbf{R}$ . The restricted profile log-likelihood  $\ell(\lambda, q, \mathbf{R})$  satisfies

$$-2\hat{\sigma}^2 \frac{\partial \ell(\lambda, q, \mathbf{R})}{\partial \gamma} = \mathbf{Y}^T \frac{\partial}{\partial \gamma} \left\{ \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}) \right\} \mathbf{Y} - \hat{\sigma}^2 \operatorname{tr} \left[ (\mathbf{I} - \mathbf{S})^{-1} \mathbf{R} \frac{\partial}{\partial \gamma} \left\{ \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}) \right\} \right], \quad (3.11)$$

where it is straight forward to verify that

$$\frac{\partial}{\partial \gamma} \left\{ \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}) \right\} = -\mathbf{R}^{-1} \left\{ \frac{\partial \mathbf{R}}{\partial \gamma} \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}) + \frac{\partial \mathbf{S}}{\partial \gamma} \right\}.$$

For the case  $\gamma = \lambda$ , and using  $\partial \mathbf{S} / \partial \lambda = -(\mathbf{I} - \mathbf{S})\mathbf{S} / \lambda$ , the estimating equation for  $\lambda$  (up to an scaling factor) follows

$$T_\lambda(\lambda, q, \mathbf{R}) = \mathbf{Y}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{S}\mathbf{Y} - \hat{\sigma}^2 \operatorname{tr}(\mathbf{S}), \quad (3.12)$$

with  $\hat{\sigma}^2$  as defined in (3.9). Given  $q$  and  $\mathbf{R}$ , the solution  $\hat{\lambda}_{q,\mathbf{R}}$  of  $T_\lambda(\lambda, q, \mathbf{R}) = 0$  provides the desired result. Criterion (3.12) is convenient to derive asymptotics but it might be difficult to evaluate numerically. Instead  $\lambda$  can be obtain from  $\ell(\lambda, q, \mathbf{I}; \mathbf{Y}^*)$  to estimate it as the solution of  $T_\lambda(\lambda, q, \mathbf{I}; \mathbf{Y}^*) = 0$ . To reduce computational cost one can take advantage of the Demmler-Reinsch basis so that the estimating equation can be further simplified to

$$\begin{aligned} T_\lambda(\lambda, q, \mathbf{I}; \mathbf{Y}^*) &= \sum_{i=q+1}^n \frac{W_i^2 \lambda n \eta_{q,i}}{(1 + \lambda n \eta_{q,i})^2} - \hat{\sigma}^2 \sum_{i=q+1}^n \frac{1}{1 + \lambda n \eta_{q,i}}, \quad \text{for} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=q+1}^n \frac{W_i^2 \lambda \eta_{q,i}}{1 + \lambda n \eta_{q,i}}, \end{aligned} \quad (3.13)$$

and  $\mathbf{W} = \Phi^T \mathbf{Y}^*$ , which is the expression that we will use hereafter.

### 3.2.3. Correlation Matrix

Consider  $\gamma$  a parameter of  $\mathbf{R}$  only, and assume the dependence of  $\mathbf{R}$  on  $\gamma$ , is sufficiently smooth. Using the definition of the natural smoother  $\mathbf{S}$  and

$$\frac{\partial \mathbf{S}}{\partial \gamma} = -\mathbf{S} \frac{\partial \mathbf{R}}{\partial \gamma} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}), \quad \text{whence} \quad \frac{\partial}{\partial \gamma} \left\{ \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \right\} = -\mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \frac{\partial \mathbf{R}}{\partial \gamma} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}),$$

the estimating equation (3.11) for a parameter  $\gamma$  of  $\mathbf{R}$  follows

$$T_\gamma(\lambda, q, \mathbf{R}) = \mathbf{Y}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \frac{\partial \mathbf{R}}{\partial \gamma} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{Y} - \hat{\sigma}^2 \text{tr} \left\{ \frac{\partial \mathbf{R}}{\partial \gamma} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \right\}, \quad (3.14)$$

which can be further simplified. Note that since  $\mathbf{R}$  is symmetric Toeplitz and hence fully specified by its first row:  $(1, \mathbf{r}^T) = (1, r_1, \dots, r_{n-1})$ . If we define  $\mathbf{D}_k$  to be the  $n \times n$  upper-shift matrix, i.e., the matrix whose entries are  $D_{k,i,j} = \delta_{k,j-i}$ ,  $i, j = 1, \dots, n$ ,  $k = 1, \dots, n-1$ , then we can express

$$\mathbf{R} = \mathbf{I} + \sum_{i=k}^{n-1} r_k (\mathbf{D}_K + \mathbf{D}_k^T), \quad \text{so that} \quad \frac{\partial \mathbf{R}}{\partial r_k} = \mathbf{D}_K + \mathbf{D}_k^T, \quad k = 1, \dots, n-1.$$

Moreover, given  $\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}) = (\mathbf{I} - \mathbf{S})^T \mathbf{R}^{-1}$ ,  $\text{tr} \{ \mathbf{D}_k \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \} = \text{tr} (\mathbf{D}_k \mathbf{R}^{-1}) \{1 + o(1)\}$  and using  $\lambda n \rightarrow \infty$ , one can re-write the estimating equations for elements  $r_k, k = 1, \dots, n - 1$  of  $\mathbf{R}$  as

$$\begin{aligned} T_{\mathbf{r},k}(\lambda, q, \mathbf{r}) &= \mathbf{Y}^T (\mathbf{I} - \mathbf{S})^T \mathbf{R}^{-1} \mathbf{D}_k \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{Y} - \hat{\sigma}^2 \text{tr} (\mathbf{D}_k \mathbf{R}^{-1}) \\ &= \mathbf{v}^T \mathbf{R}^{-1} \mathbf{D}_k \mathbf{R}^{-1} \mathbf{v} - \hat{\sigma}^2 \text{tr} (\mathbf{D}_k \mathbf{R}^{-1}) \\ &= \text{tr} \{ \mathbf{R}^{-2} (\mathbf{D}_k \mathbf{v} \mathbf{v}^T - \hat{\sigma}^2 \mathbf{D}_k \mathbf{R}) \}, \end{aligned}$$

where we set  $\mathbf{v} = (\mathbf{I} - \mathbf{S}) \mathbf{Y}$  and we have taken advantage of the resulting quadratic form to write it as a trace. Moreover if we assume the noise to be short range dependent,  $\|\mathbf{R} - \rho \mathbf{I}\|_{op} \rightarrow 0$  as  $n \rightarrow \infty$  for some  $\rho \neq 0$ . Meaning that solving for  $r_k$  in  $T_{\mathbf{r},k}(\lambda, q, \mathbf{r}) = 0$  is asymptotically equivalent to solving  $\text{tr} (\mathbf{D}_k \mathbf{v} \mathbf{v}^T) = \hat{\sigma}^2 \text{tr} (\mathbf{D}_k \mathbf{R})$ . Hence

$$(n - k) \hat{\sigma}^2 r_k = \mathbf{v}^T \mathbf{D}_k \mathbf{v} \tag{3.15}$$

gives an explicit (approximate) solution for each  $r_k$ . Unfortunately the resulting estimate  $\hat{\mathbf{R}}$  is not necessarily a positive matrix, and it is not consistent for the true correlation matrix in operator norm. A common approach to solve this problems is to taper the estimate. Define the estimators  $\hat{r}_k = \hat{r}_{k,\lambda,q,\hat{\sigma}^2}$  for  $r_k$

$$\hat{r}_k = \frac{(\mathbf{Y} - \hat{\mathbf{f}})^T \mathbf{D}_k (\mathbf{Y} - \hat{\mathbf{f}})}{(n - k) \hat{\sigma}^2} = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n-k} \frac{(Y_i - \hat{f}_i)(Y_{i+k} - \hat{f}_{i+k})}{(n - k)}, \quad k = 1, \dots, n - 1, \tag{3.16}$$

and define the following tapered estimator of  $\mathbf{R}$

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_{\lambda,q,\hat{\sigma}^2,d_n} = \mathbf{I} + \sum_{k=1}^{d_n} \hat{r}_k w_k (\mathbf{D}_k + \mathbf{D}_k^T), \tag{3.17}$$

where  $d_n \leq n - 1$  is any non-decreasing sequence of positive integers, and  $w_k = w_{k,n}$  are appropriate weights chosen to ensure that the estimate is positive definite. For

the selection of  $d_n$  and  $w_k$  the interested reader can refer to [Xiao and Wu \[2012\]](#).

There are many alternatives in the literature to characterise the error's correlation, which allow for a direct estimation of the correlation matrix without assuming any prior estimation of the regression function [cf. [Hart, 1991](#), [Hall and Van Keilegom, 2003](#)] for an AR(p) parametric approach and [Herrmann et al. \[1992\]](#) for a non-parametric approach that handles a broader variety of correlation structures. In principle, any method that delivers a consistent estimator for  $\mathbf{R}$  could be used. However representation (3.17) is less restrictive since it only assumes exponential decay in the autocorrelation function of a short range dependent error process and, hence, is preferred.

### 3.2.4. Smoothness Class

The interdependence between the estimators for  $\lambda$  and  $\mathbf{R}$  does not affect the estimation of  $q$ , hence  $\lambda$  and  $\mathbf{R}$  can be estimated for each value of  $q \in \{1, \dots, \lfloor \log(n) \rfloor\}$  under consideration. In fact once consistent estimates for the correlation matrix of the noise are available, the problem of estimating  $q$  under correlation  $\mathbf{R}$  can be reduced to the problem of estimating  $q$  in a model with  $\mathbf{R} = \mathbf{I}$ , which was studied in [Serra and Krivobokova \[2016\]](#). Here, we apply this approach to the pre-whitened data  $\mathbf{Y}^* = \hat{\mathbf{R}}^{-1/2} \mathbf{Y}$ , where  $\hat{\mathbf{R}}$  is a consistent estimator of  $\mathbf{R}$ . Once again making use of the Demmler-Reinsch basis one can write  $\mathbf{S}_{\lambda,q,\mathbf{I}} = \Phi \text{diag} \{(1 + \lambda n \eta_{q,i})^{-1}\} \Phi^T$ , and since  $\partial(n\eta_{q,i})/\partial q = n\eta_{q,i} \log(n\eta_{q,i})/q$ , whence

$$\frac{\partial \mathbf{S}_{\lambda,q,\mathbf{I}}}{\partial q} = -\frac{1}{q} \Phi \mathbf{D}_{\lambda,q} \Phi^T, \text{ with } \mathbf{D}_{\lambda,q} = \text{diag} \left\{ \frac{\lambda n \eta_{q,1} \log(n\eta_{q,1})}{(1 + \lambda n \eta_{q,1})^2}, \dots, \frac{\lambda n \eta_{q,1} \log(n\eta_{q,n})}{(1 + \lambda n \eta_{q,n})^2} \right\},$$



it follows that up to a scaling factor

$$\begin{aligned}
 T_q(\lambda, q, \mathbf{I}; \mathbf{Y}^*) &= \mathbf{Y}^{*T} \boldsymbol{\Phi} \mathbf{D}_{\lambda, q} \boldsymbol{\Phi}^T \mathbf{Y}^* - \hat{\sigma}^2 \mathbf{I} \sum_{i=q+1}^n \frac{\log(n\eta_{q,i})}{1 + \lambda n \eta_{q,i}} \\
 &= \boldsymbol{\mathcal{Y}}^T (\hat{\mathbf{R}} \mathbf{R}^{-1})^{-1/2} \boldsymbol{\Phi} \mathbf{D}_{\lambda, q} \boldsymbol{\Phi}^T (\hat{\mathbf{R}} \mathbf{R}^{-1})^{-1/2} \boldsymbol{\mathcal{Y}} - \hat{\sigma}^2 \mathbf{I} \sum_{i=q+1}^n \frac{\log(n\eta_{q,i})}{1 + \lambda n \eta_{q,i}} \\
 &= T_q(\lambda, q, \mathbf{I}; \boldsymbol{\mathcal{Y}}) \{1 + o_P(1)\},
 \end{aligned}$$

where  $\boldsymbol{\mathcal{Y}} = \mathbf{R}^{-1/2} \mathbf{Y}$  and the last equality holding if  $\hat{\mathbf{R}}$  is consistent for  $\mathbf{R}$  in operator norm, and  $\mathbf{R}$  has eigenvalues bounded away from zero and infinity. If  $\mathbf{R}$  is the true correlation matrix of the noise, then the coordinates of  $\boldsymbol{\mathcal{Y}}$  are independent. The conclusion is that the naive criterium  $T_q(\lambda, q, \mathbf{I}; \mathbf{Y}^*)$  is asymptotically equivalent to  $T_q(\lambda, q, \mathbf{I}; \boldsymbol{\mathcal{Y}})$  which is of the form proposed in [Serra and Krivobokova \[2016\]](#), that is

$$\begin{aligned}
 T_q(\lambda, q, \mathbf{I}; \mathbf{Y}^*) &= \sum_{i=q+1}^n \frac{W_i^2 \lambda n \eta_{q,i} \log(n\eta_{q,i})}{(1 + \lambda n \eta_{q,i})^2} - \hat{\sigma}^2 \sum_{i=q+1}^n \frac{\log(n\eta_{q,i})}{1 + \lambda n \eta_{q,i}} \\
 \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=q+1}^n \frac{W_i^2 \lambda \eta_{q,i}}{1 + \lambda n \eta_{q,i}}
 \end{aligned} \tag{3.18}$$

where  $\mathbf{W} = \boldsymbol{\Phi}^T \mathbf{Y}^*$ . An estimator of  $q$  is obtained by solving  $T_q(\hat{\lambda}_q^*, q, \mathbf{I}; \mathbf{Y}^*) = 0$ ,  $q \in \{1, \dots, \lfloor \log(n) \rfloor\}$ , where  $\hat{\lambda}_q^*$  is the naive estimator that solves  $T_\lambda(\lambda, q, \mathbf{I}; \mathbf{Y}^*) = 0$ .

### 3.3. Asymptotic Properties<sup>1</sup>

We now describe the behaviour of the estimators from the [Section 3.2](#). Henceforth let  $\mathbb{P}$ , (resp.  $\mathbb{E}$ ,  $\mathbb{V}$ ) represent probability (resp. expectation, variance) with respect to  $\mathcal{N}(\mathbf{f}, \sigma^2 \mathbf{R})$ , where  $f \in L_2$ ,  $\sigma^2 > 0$ ,  $\mathbf{R}$  are the true values of the parameters of interest which determine the distribution of the data.

The first theorem describes the behaviour of the smoothing parameter when the cor-

---

<sup>1</sup>This section corresponds to the work of Dr. Paulo Serra developed in a joint project at the Institute of Mathematical Stochastics at Göttingen University.

### 3. Smoothing Splines with Correlated Errors

---

relation in the noise is ignored, i.e., when we set  $\mathbf{R} = \mathbf{I}$ .

**Theorem 5** (Consistency of preliminary estimate of  $\lambda$ ). *Let  $f \in \mathcal{W}^\beta(M)$  and assume that  $\|f^{(\beta)}\|^2 > 0$ . Assume that the first row of  $\mathbf{R}$  is absolutely summable. Assume in addition that the eigenvalues of  $\mathbf{R}$  are bounded away from zero and bounded away from infinity, and denote by  $\tau$  the spectral density at zero of the noise process. Denote by  $\hat{\lambda}_{q,\mathbf{I}}$  the solution to  $\mathbb{E}T_\lambda(\lambda, q, \mathbf{I}) = 0$ ,  $\lambda > 0$ . Then  $\hat{\lambda}_{q,\mathbf{I}}$  is consistent for the oracle  $\lambda_{q,\mathbf{I}}$ , in that  $\hat{\lambda}_{q,\mathbf{I}}/\lambda_{q,\mathbf{I}} \rightarrow 1$ , in  $\mathbb{P}$ -probability, as  $n \rightarrow \infty$ . Furthermore, if  $\max(1, \tau)/2 < q \leq \max\{\beta > \max(1, \tau)/2 : f \in \mathcal{W}^\beta(M)\}$ , then*

$$\lambda_{q,\mathbf{I}} = \left[ \frac{n\|f^{(q)}\|^2}{\sigma^2\{\tau\kappa_q(0, 2) - (\tau - 1)\kappa_q(0, 1)\}} \{1 + o(1)\} \right]^{-\frac{2q}{2q+1}}, \quad (3.19)$$

and, if  $f \in \mathcal{W}^\beta$ ,  $\beta > 1/2$ , and  $q > \max(\beta, \tau/2)$ , then

$$\lambda_{q,\mathbf{I}} \geq \left[ \frac{n\|f^{(\beta)}\|^2}{\sigma^2\{\tau\kappa_q(0, 2) - (\tau - 1)\kappa_q(0, 1)\}} \{1 + o(1)\} \right]^{-\frac{2q}{2\beta+1}}. \quad (3.20)$$

*Proof.* Note that  $\mathbf{y} = \mathbf{R}^{-1/2}\mathbf{Y} \sim \mathcal{N}(\mathbf{R}^{-1/2}\mathbf{f}, \sigma^2\mathbf{I})$  so that if  $\mathbf{L}$  is a symmetric matrix, then  $\mathbb{E}(\mathbf{y}^T\mathbf{L}\mathbf{y}) = \mathbf{f}^T\mathbf{R}^{-1/2}\mathbf{L}\mathbf{R}^{-1/2}\mathbf{f} + \sigma^2\text{tr}(\mathbf{L})$ . Denote by  $\tau$  the spectral density of  $\mathbf{R}$  at 0.

First note that for fixed  $\lambda, q, \mathbf{R}$ ,

$$\mathbb{E}(\hat{\sigma}^2) = \frac{1}{n} \left[ \mathbf{f}^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{f} + \sigma^2\text{tr}\{\mathbf{R}\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\} \right], \quad (3.21)$$

so that by (A.7) and (A.11), for fixed  $\lambda$  and  $q$ ,  $\mathbb{E}(\hat{\sigma}_{\mathbf{I}}^2) = \sigma^2\{1 + o(1)\}$ , uniformly over  $\mathbf{R}$ .

Similar computations show that for each  $\lambda, q, \mathbf{R}$ ,

$$\mathbb{E}T_\lambda(\lambda, q, \mathbf{R}) = \frac{1}{n} \left[ \mathbf{f}^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{S}\mathbf{f} + \sigma^2\text{tr}\{\mathbf{R}\mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{S}\} - \mathbb{E}\hat{\sigma}^2\{\text{tr}(\mathbf{S})\} \right],$$

so that in particular, for  $\mathbf{R} = \mathbf{I}$ ,

$$\mathbb{E}T_\lambda(\lambda, q, \mathbf{I}) = \frac{1}{n} \left[ \mathbf{f}^T (\mathbf{I} - \mathbf{S}_\mathbf{I}) \mathbf{S}_\mathbf{I} \mathbf{f} - \sigma^2 \operatorname{tr} \{ \mathcal{R} \mathbf{S}_\mathbf{I}^2 - (\mathcal{R} - \mathbf{I}) \mathbf{S}_\mathbf{I} \} \{1 + o(1)\} \right].$$

To ensure that a solution to the estimating equation  $\mathbb{E}T_\lambda(\lambda, q, \mathbf{I}) = 0$  exists, the trace above must be positive. By (A.7),

$$\operatorname{tr} \{ \mathcal{R} \mathbf{S}_\mathbf{I}^2 - (\mathcal{R} - \mathbf{I}) \mathbf{S}_\mathbf{I} \} = \{ \tau \kappa_q(0, 2) - (\tau - 1) \kappa_q(0, 1) \} \lambda^{-1/(2q)} \{1 + o(1)\},$$

which is positive if, and only if,  $\tau < 2q$ . We conclude that if  $\tau < 2q$ , then the solution to the estimating equation when  $f \in \mathcal{W}^q$ ,  $\mathbf{R} = \mathbf{I}$ , and  $\tau < 2q$ , is

$$\lambda_{q, \mathbf{I}} = \left[ \frac{n \|f^{(q)}\|^2}{\sigma^2 \{ \tau \kappa_q(0, 2) - (\tau - 1) \kappa_q(0, 1) \}} \{1 + o(1)\} \right]^{-\frac{2q}{2q+1}}, \quad (3.22)$$

and if  $f \in \mathcal{W}^\beta$ ,  $q > \max(\beta, \tau/2)$ , then

$$\lambda_{q, \mathbf{I}} \geq \left[ \frac{n \|f^{(\beta)}\|^2}{\sigma^2 \{ \tau \kappa_q(0, 2) - (\tau - 1) \kappa_q(0, 1) \}} \{1 + o(1)\} \right]^{-\frac{2q}{2\beta+1}}. \quad (3.23)$$

To show that  $\hat{\lambda}_{q, \mathbf{I}}$  is consistent for these oracles, it suffices to show that for each  $q$ ,  $T_\lambda(\lambda, q, \mathbf{I}) - \mathbb{E}T_\lambda(\lambda, q, \mathbf{I})$  converges uniformly over  $\lambda$  to 0, in probability. Since we will need this to hold in greater generality, we write

$$T_\lambda(\lambda, q, \mathbf{R}) = \frac{1}{n} \mathbf{y}^T \mathbf{L} \mathbf{y} - \frac{1}{n} \mathbf{y}^T \mathbf{L}' \mathbf{y} \frac{\operatorname{tr}(\mathbf{S})}{n},$$

where  $\mathbf{y} = \Phi_q^T \mathcal{R}^{-1/2} \mathbf{Y} \sim \mathcal{N}(\mathbf{y}, \sigma^2 \mathbf{I})$ , for  $\mathbf{y} = \Phi_q^T \mathcal{R}^{-1/2} \mathbf{f}$  (which is in  $\ell_2$  since  $\mathcal{R}$  has eigenvalues bounded away from 0) and the symmetric matrices  $\mathbf{L}$  and  $\mathbf{L}'$  are

$$\mathbf{L} = \Phi_q^T \mathcal{R}^{1/2} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{S} \mathcal{R}^{1/2} \Phi_q, \quad \mathbf{L}' = \Phi_q^T \mathcal{R}^{1/2} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathcal{R}^{1/2} \Phi_q.$$

Define the collection  $\Lambda_n$  of all  $\lambda$  such that  $\lambda = o(1)$ , and  $\lambda n \rightarrow \infty$ . We can write

### 3. Smoothing Splines with Correlated Errors

---

$\sup_{\lambda \in \Lambda_n} |T_\lambda(\lambda, q, \mathbf{R}) - \mathbb{E}T_\lambda(\lambda, q, \mathbf{R})|$  as the sum of two terms, one involving  $\mathbf{L}$ , and one involving  $\mathbf{L}'$ . We treat these separately.

Consider the term involving  $\mathbf{L}$  (the term involving  $\mathbf{L}'$  is controlled in the same way). If the smallest eigenvalue of  $\mathbf{R}$  is at least  $\delta$ , then the eigenvalues of  $\mathbf{L}$  (and  $\mathbf{L}'$ ) are at most  $\bar{\tau}/\delta$ , where  $\bar{\tau}$  is the largest eigenvalue of  $\mathbf{R}$ . By the Cauchy-Schwarz inequality, and the triangle inequality,

$$\begin{aligned} \frac{1}{n} |\mathbf{y}^T \mathbf{L} \mathbf{y} - \mathbb{E}(\mathbf{y}^T \mathbf{L} \mathbf{y})| &= \frac{1}{n} |(\mathbf{y} - \mathbf{y})^T \mathbf{L} (\mathbf{y} - \mathbf{y}) + 2(\mathbf{y} - \mathbf{y})^T \mathbf{L} \mathbf{z} - \sigma^2 \text{tr}(\mathbf{L})| \\ &\leq \frac{1}{n} (\mathbf{y} - \mathbf{y})^T \mathbf{L} (\mathbf{y} - \mathbf{y}) + \frac{2}{n} \sqrt{\mathbf{y}^T \mathbf{L} \mathbf{z} (\mathbf{y} - \mathbf{y})^T \mathbf{L} (\mathbf{y} - \mathbf{y})} + \frac{\sigma^2}{n} \text{tr}(\mathbf{L}) \\ &\leq \frac{\bar{\tau}}{n\delta} (\mathbf{y} - \mathbf{y})^T (\mathbf{y} - \mathbf{y}) + 2\bar{\tau} \sqrt{\frac{\mathbf{z}^T \mathbf{z}}{n} \frac{1}{n\delta^2} (\mathbf{y} - \mathbf{y})^T (\mathbf{y} - \mathbf{y})} + \frac{\sigma^2}{n} \text{tr}(\mathbf{L}), \end{aligned}$$

almost surely. If  $n\delta^2 \rightarrow \infty$  then it suffices to control  $(\mathbf{y} - \mathbf{y})^T (\mathbf{y} - \mathbf{y}) / (n\delta^2)$ , which does not depend on  $\lambda$ , and  $\text{tr}(\mathbf{L})/n$ , which is deterministic. By (A.7),

$$\frac{\sigma^2}{n} \text{tr}(\mathbf{L}) = \frac{\sigma^2}{n} \text{tr} \{ \mathbf{R} \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{S} \} = \frac{\sigma^2 \bar{\tau}}{n\delta} (\delta\lambda)^{-1/(2q)} \kappa_q(1, 1) \{1 + o(1)\},$$

which goes to zero if  $\delta \geq O\{n^{-(2q-1)/(2q+1)}\}$ . By Markov's inequality, since the coordinates of  $\mathbf{y}$  are independent, for any  $\epsilon > 0$ ,

$$\mathbb{P}\left(\left|\frac{1}{n\delta^2} (\mathbf{y} - \mathbf{y})^T (\mathbf{y} - \mathbf{y})\right| > \epsilon \mid q, \mathbf{R}\right) \leq \frac{1}{\epsilon^2 n^2 \delta^4} \mathbb{E}\left(\left|(\mathbf{y} - \mathbf{y})^T (\mathbf{y} - \mathbf{y})\right|^2 \mid q, \mathbf{R}\right) \leq \frac{2\sigma^2 + 4\mathbf{y}^t \mathbf{y}/n}{\epsilon^2 n \delta^4},$$

which goes to 0 as long as  $n\delta^4 \rightarrow \infty$ . (Note that if the eigenvalues of  $\mathbf{R}$  are at least  $\underline{\tau}$ , then  $\mathbf{y}^T \mathbf{y} \leq \mathbf{f}^t \mathbf{f} / \underline{\tau} = n \|f\|^2 / \underline{\tau} \{1 + o(1)\}$ .) Conclude that

$$\sup_{\lambda \in \Lambda_n} |\mathbf{y}^T \mathbf{L} \mathbf{y} - \mathbb{E}(\mathbf{y}^T \mathbf{L} \mathbf{y})| / n \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty,$$

as long as  $\delta \gg \max(n^{-1/4}, n^{-(2q-1)/(2q+1)})$ . In particular we conclude that  $\hat{\lambda}_{q, \mathbf{I}} / \lambda_{q, \mathbf{I}} \rightarrow 1$ , in probability as long as  $\delta \gg \max(n^{-1/4}, n^{-(2q-1)/(2q+1)})$ .

□

If  $\mathbf{R} = \mathbf{I}$  then  $\tau = 1$  in which case the constant in the denominator becomes  $\kappa_q(0, 2)$  and one just recuperates the oracles for the i.i.d. noise setting. For details see [Serra and Krivobokova \[2016\]](#).

Note also the lower bound on  $q$  based on the spectral density of the noise process. If the constraint does not hold then the criterium for  $\lambda$  will be asymptotically positive. This implies that the likelihood decreases with  $\lambda$ , such that approximate solutions to the estimating equation for  $\lambda$  will under-smooth the data.

For  $\mathbf{x} \in \mathbb{R}^n$  let  $\|\mathbf{x}\|^2$  denote  $\mathbf{x}^T \mathbf{x}$ . The risk of the smoothing spline estimator is, by (A.7), (A.9), and (A.10), for all large enough  $n$ ,

$$\begin{aligned} \mathbb{E}\|\mathbf{f} - \mathbf{S}\mathbf{Y}\|^2/n &= \mathbb{E}\|\mathbf{S}\boldsymbol{\epsilon}\|^2/n + \|(\mathbf{I} - \mathbf{S})\mathbf{f}\|^2/n = \sigma^2 \text{tr}(\mathbf{S}^2 \mathbf{R})/n + \mathbf{f}^T (\mathbf{I} - \mathbf{S})^2 \mathbf{f}/n \\ &\leq \tau(\rho\lambda)^{-1/(2q)} \kappa_q(0, 2)/n + (\rho\lambda)^{1 \wedge \frac{\beta}{q}} \|f^{(q \wedge \beta)}\|^2 \{1 + o(1)\}, \end{aligned}$$

where  $\rho$  is the spectral density associated with the entries of  $\mathbf{R}$  at zero; cf. (A.3). The risk is therefore affected by the presence of the correlation in the noise and by the specific candidate  $\mathbf{R}$  used in the smoother via the respective spectral densities at zero. Setting  $\mathbf{R} = \mathbf{I}$  and  $\lambda = \hat{\lambda}_{\mathbf{I}}$  results in an spline estimate whose risk is of the order of the minimax risk for  $f \in \mathcal{W}^{q \wedge \beta}$ . However, by estimating the correlation structure consistently, the (bound on the) risk can be reduced. The next theorem describes the behaviour of the estimator  $\hat{\mathbf{R}}_{\lambda, \mathbf{R}}$  when  $\mathbf{R} = \mathbf{I}$ , and  $\lambda = \hat{\lambda}_{\mathbf{I}}$ .

**Theorem 6** (Consistency of preliminary estimate of  $\mathbf{R}$ ). *Assume that the conditions of the previous theorem hold, and consider the estimator  $\hat{\lambda}_{q, \mathbf{I}}$ . Suppose that the entries of  $\mathbf{R}$  satisfy  $r_k = O(k^{-\alpha})$ ,  $\alpha > 1$ . Then, if  $\mathbf{R} = \mathbf{I}$  and  $\lambda = \hat{\lambda}_{q, \mathbf{I}}$ , the estimator  $\hat{\mathbf{R}} = \hat{\mathbf{R}}_{\lambda, q, \mathbf{R}, d_n}$  is consistent for  $\mathbf{R}$  in operator norm in the sense that,*

$$\mathbb{P} \left[ \nu(\hat{\mathbf{R}} - \mathbf{R}) \leq C \left\{ \sqrt{\frac{d_n \log(d_n)}{n}} + d_n^{-\min(\alpha-1, 2)} + d_n n^{-\frac{\min(\beta, q)}{2 \min(\beta, q) + 1}} \right\} \right] \rightarrow 1, \quad n \rightarrow \infty,$$

where  $\nu(\cdot)$  denotes the spectral radius, and  $C > 0$  is some universal constant. The

### 3. Smoothing Splines with Correlated Errors

---

terms in the rate are balanced for  $d_n = O\left(n^{\frac{\min(\beta,q)/\min(\alpha,3)}{2\min(\beta,q)+1}}\right)$ , resulting in the rate

$$n^{-\frac{\min(\beta,q)}{2\min(\beta,q)+1} - \frac{\min(\alpha-1,2)}{\min(\alpha,3)}}.$$

*Proof.* We prove the consistency of the estimator  $\hat{\mathbf{R}}_\lambda$  at  $\lambda = \hat{\lambda}_{q,\mathbf{I}}$ . For a square matrix  $\mathbf{L}$  let  $\nu(\mathbf{L})$  represent the spectral radius of  $\mathbf{L}$ ; this is the largest eigenvalue (in absolute value) of  $\mathbf{L}$ .

Since  $\mathbf{f}$  is unknown to us, define  $\tilde{r}_k = (\mathbf{Y} - \mathbf{f})^T \mathbf{D}_k (\mathbf{Y} - \mathbf{f}) / \{n\hat{\sigma}^2\}$  and note that

$$\nu(\hat{\mathbf{R}} - \mathcal{R}) \leq \nu(\hat{\mathbf{R}} - \tilde{\mathbf{R}}) + \nu(\tilde{\mathbf{R}} - \mathcal{R}),$$

where  $\tilde{\mathbf{R}}$  is  $\hat{\mathbf{R}}$  as defined in (3.17) with each  $\hat{r}_k$  replaced with  $\tilde{r}_k$ . The right-most term in the previous display converges to zero in probability (at a rate) by Theorem 4 of Xiao and Wu [2012]. If  $d_n \rightarrow \infty$ ,  $d_n < n - 1$ , and  $r_k = O(k^{-\alpha})$ ,  $\alpha > 1$ , then the rate is of order  $\sqrt{d_n \log(d_n)/n} + d_n^{-\min(\alpha-1,2)}$ . The optimal choice for  $d_n$  is to take it of order  $n^{1/\{2\min(\alpha-1,2)+1\}}$ , in which in which case one obtains the rate  $n^{-\min(\alpha-1,2)/\{2\min(\alpha-1,2)+1\}}$ ; cf. [Xiao and Wu, 2012, p. 475]. The remaining term is bounded using Gershgorin's circle Theorem [Horn and Johnson, 1990, Theorem 6.1.1]:

$$\nu(\hat{\mathbf{R}} - \tilde{\mathbf{R}}) \leq 2 \sum_{k=1}^{d_n} w_k |\hat{r}_k - \tilde{r}_k|,$$

where the differences in the summands above satisfy

$$n\hat{\sigma}^2(\hat{r}_k - \tilde{r}_k) = \boldsymbol{\epsilon}^T (\mathbf{D}_k + \mathbf{D}_k^T) (\mathbf{f} - \hat{\mathbf{f}}) + (\mathbf{f} - \hat{\mathbf{f}})^T \mathbf{D}_k (\mathbf{f} - \hat{\mathbf{f}}).$$

The two terms in the sum above are  $o_P(n)$  when we set  $\lambda = \hat{\lambda}_{q,\mathbf{I}}$ , implying consistency of  $\hat{\mathbf{R}}$  in operator norm. Since  $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = O_P(n)$ , and  $(\mathbf{f} - \hat{\mathbf{f}}_\lambda)^T (\mathbf{f} - \hat{\mathbf{f}}_\lambda) = O_P(n^{-\min(\beta,q)/\{2\min(\beta,q)+1\}})$  when  $\lambda = \hat{\lambda}_q^*$ , by the Cauchy-Schwarz inequality, we conclude that the right-hand-side above is  $O_P(n n^{-\min(\beta,q)/\{2\min(\beta,q)+1\}})$  and conclude that

the rate (in operator norm) of the estimate for the correlation matrix is

$$\sqrt{d_n \log(d_n)/n} + d_n^{-\min(\alpha-1,2)} + d_n n^{-\min(\beta,q)/\{2\min(\beta,q)+1\}}.$$

□

Even without knowledge of  $\alpha$  or  $\beta$ , picking  $d_n$  such that  $d_n \rightarrow \infty$  and  $d_n = o(n^{1/4})$  will still lead to a consistent estimator of  $\mathcal{R}$ . In the following theorem we show that using any consistent estimator of  $\mathcal{R}$  improves the estimation of the smoothing parameter  $\lambda$ , in the sense that the estimate is going to be consistent for the same oracle as if  $\mathcal{R}$  were known.

**Theorem 7** (Consistency of estimate of  $\lambda$ ). *Suppose the conditions of Theorem 5 hold. Denote by  $\hat{\lambda}_{q, \hat{\mathbf{R}}}$  the solution to  $\mathbb{E}T_\lambda(\lambda, q, \mathbf{I}; \hat{\mathbf{R}}^{-1/2} \mathbf{Y}) = 0$ ,  $\lambda > 0$ , where  $\hat{\mathbf{R}}$  is consistent in operator norm for  $\mathcal{R}$ . Then  $\hat{\lambda}_{q, \hat{\mathbf{R}}}$  is consistent for the oracle  $\lambda_{q, \mathcal{R}}$ , in that  $\hat{\lambda}_{q, \hat{\mathbf{R}}}/\lambda_{q, \mathcal{R}} \rightarrow 1$ , in  $\mathbb{P}$ -probability, as  $n \rightarrow \infty$ . Furthermore, if  $\max(1, \tau)/2 < q \leq \max\{\beta > \max(1, \tau)/2 : f \in \mathcal{W}^\beta(M)\}$ , then*

$$\lambda_{q, \mathcal{R}} = \frac{1}{\tau} \left[ \frac{n \|f^{(q)}\|^2}{\sigma^2 \tau \kappa_q(0, 2)} \{1 + o(1)\} \right]^{-\frac{2q}{2q+1}}, \quad (3.24)$$

and, if  $f \in \mathcal{W}^\beta$ ,  $\beta > 1/2$ , and  $q > \max(\beta, \tau/2)$ , then

$$\lambda_{q, \mathcal{R}} \geq \frac{1}{\tau} \left[ \frac{n \|f^{(\beta)}\|^2}{\sigma^2 \tau \kappa_q(0, 2)} \{1 + o(1)\} \right]^{-\frac{2q}{2\beta+1}}. \quad (3.25)$$

Furthermore, for each  $q$ , these oracles match the oracles for  $\lambda$  when  $\mathcal{R}$  is known.

*Proof.* We now look at the behaviour of  $\hat{\lambda}_{q, \hat{\mathbf{R}}}$ . From the previous section we know that the eigenvalues of  $\hat{\mathbf{R}} = \hat{\mathbf{R}}_{\hat{\lambda}, q, \mathbf{I}, d_n}$  are bounded away from zero in probability if  $d_n$  is chosen appropriately. We conclude that in probability  $\{T_\lambda(\lambda, q, \mathbf{R}) - \mathbb{E}T_\lambda(\lambda, q, \mathbf{R})\}|_{\mathbf{R}=\hat{\mathbf{R}}}$  converges to 0, uniformly over  $\lambda \in \Lambda_n$ . It remains to show that  $\mathbb{E}T_\lambda(\lambda, q, \mathbf{R})|_{\mathbf{R}=\hat{\mathbf{R}}} -$

$\mathbb{E}T_\lambda(\lambda, q, \mathbf{R})$  also converges to zero in probability, uniformly over  $\lambda$ .

First note that since in probability the eigenvalues of  $\hat{\mathbf{R}}$  are bounded away from zero and from infinity, for each  $f \in \mathcal{W}^\beta$ ,

$$\mathbf{f}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{f} \Big|_{\mathbf{R}=\hat{\mathbf{R}}} = O_P\{n\lambda^{1 \wedge (\beta/q)} \|f^{(\beta \wedge q)}\|^2\},$$

which is  $o_P(n)$  uniformly over  $\lambda \in \Lambda_n$ . On the other hand,

$$\text{tr}\{\mathbf{R}\hat{\mathbf{R}}^{-1}(\mathbf{I}-\mathbf{S})\} = n\{1+o(1)\} + \text{tr}\{(\mathbf{R}-\hat{\mathbf{R}})\hat{\mathbf{R}}^{-1}(\mathbf{I}-\mathbf{S})\} = n\left[1+o(1)+O_P\{d_n\lambda^{-1/(2q)}\}\right].$$

Conclude that uniformly over  $\lambda \in \Lambda_n$ ,  $\mathbb{E}[\sigma^2|\lambda, q, \mathbf{R}] \Big|_{\mathbf{R}=\hat{\mathbf{R}}} = \sigma^2\{1+o_P(1)\}$ , as long as  $d_n = O\{n^{(2q-1)/(2q)}\}$ . This means that if we abbreviate  $\hat{\mathbf{S}} = \mathbf{S}_{\lambda, \hat{\mathbf{R}}}$  and  $\mathbf{S} = \mathbf{S}_{\lambda, \mathbf{R}}$ , then  $n\{\mathbb{E}T_\lambda(\lambda, q, \mathbf{R}) \Big|_{\mathbf{R}=\hat{\mathbf{R}}} - \mathbb{E}T_\lambda(\lambda, q, \mathbf{R})\}$  is (up to smaller order terms)

$$\mathbf{f}^T \hat{\mathbf{R}}^{-1}(\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{S}}\mathbf{f} - \mathbf{f}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{S}\mathbf{f} + \sigma^2 \text{tr}\{\mathbf{R}\hat{\mathbf{R}}^{-1}(\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{S}} - \hat{\mathbf{S}}\} + \sigma^2 \text{tr}\{\mathbf{S}^2\}.$$

We bound the difference between the quadratic forms. Using the identity  $\mathbf{S}_{\hat{\mathbf{R}}}^{-1} - \mathbf{I} = \mathbf{R}(\mathbf{S}_{\hat{\mathbf{R}}}^{-1} - \mathbf{I})$  it follows that  $\mathbf{S}_{\hat{\mathbf{R}}} - \mathbf{S}_{\mathbf{R}} = \mathbf{S}_{\mathbf{R}}(\mathbf{R} - \hat{\mathbf{R}})(\mathbf{S}_{\hat{\mathbf{R}}}^{-1} - \mathbf{I})\mathbf{S}_{\hat{\mathbf{R}}}$ . Then,

$$\begin{aligned} & \mathbf{f}^T \hat{\mathbf{R}}^{-1}(\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{S}}\mathbf{f} - \mathbf{f}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S})\mathbf{S}\mathbf{f} = \\ & = \mathbf{f}^T(\hat{\mathbf{R}}^{-1} - \mathbf{R}^{-1})(\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{S}}\mathbf{f} - \mathbf{f}^T \mathbf{R}^{-1}(\mathbf{I} - \hat{\mathbf{S}} - \mathbf{S})(\mathbf{S} - \hat{\mathbf{S}})\mathbf{f}. \end{aligned}$$

The previous display is  $o_P(n)$ . The difference between the traces is

$$\sigma^2 \text{tr}\{(\mathbf{R} - \hat{\mathbf{R}})\hat{\mathbf{R}}^{-1}(\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{S}}\} + \sigma^2 \text{tr}\{\mathbf{S}^2 - \hat{\mathbf{S}}^2\}.$$

The first of these terms is controlled as before and is  $o_P(n)$ , uniformly over  $\lambda \in \Lambda_n$ , as long as  $d_n = O\{n^{(2q-1)/(2q)}\}$ ; the second term is easily shown to be of the same order as  $n\nu(\hat{\mathbf{R}} - \mathbf{R}) = o_P(n)$ , for appropriate  $d_n$ .



It remains to compute the oracle for  $\lambda$ . We have

$$\mathbb{E}T_\lambda(\lambda, q, \mathbf{R}) = \frac{1}{n} \left[ \mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}) \mathbf{S} \mathbf{f} - \sigma^2 \text{tr}(\mathbf{S}^2) \right].$$

Using (A.7), (A.9), and (A.10), conclude that the oracle  $\lambda_{q, \mathbf{R}}$  satisfies the following: when  $f \in \mathcal{W}_q$ , and  $q > 1/2$ , then

$$\lambda_{q, \mathbf{R}} = \frac{1}{\tau} \left[ \frac{n \|f^{(q)}\|^2}{\sigma^2 \tau \kappa_q(0, 2)} \{1 + o(1)\} \right]^{-\frac{2q}{2q+1}}, \quad (3.26)$$

and if  $f \in \mathcal{W}^\beta$ ,  $q > \beta$ , then

$$\lambda_{q, \mathbf{R}} \geq \frac{1}{\tau} \left[ \frac{n \|f^{(\beta)}\|^2}{\sigma^2 \tau \kappa_q(0, 2)} \{1 + o(1)\} \right]^{-\frac{2q}{2\beta+1}}. \quad (3.27)$$

□

## 3.4. Implementation

The specific dependence of the restricted profile log-likelihood on  $\lambda$ ,  $\mathbf{R}$  and  $q$  makes it infeasible to find the maximisers directly. However, in section 3.2 we defined estimates for each parameter given the others. This suggests an iterative procedure to approximate the maximisers of the restricted profile log-likelihood.

### 3.4.1. Statistical Algorithm

A numerically robust procedure for the estimation of  $\lambda$ ,  $q$  and  $\mathbf{R}$  is not obvious. Here we present the algorithmic implementation for the naive versions of the estimators in two loops: an inner loop to compute  $\hat{\lambda} = \hat{\lambda}|_{q=\bar{q}}$  and  $\hat{\mathbf{R}} = \hat{\mathbf{R}}|_{q=\bar{q}}$  given an element of  $\mathcal{Q}_n = \{1, \dots, \lfloor \log(n) \rfloor\}$ , say  $\bar{q}$ ; and an outer loop where the inner loop is repeated for each element in  $\mathcal{Q}_n$ .

### Inner loop

Consider an element  $\bar{q} \in \mathcal{Q}$  and proceed as follows

1. Initialisation. Set  $\mathbf{R}^{(0)} = \mathbf{I}$  to compute  $\lambda^{(0)}$  as the solution of  $T_\lambda(\lambda, \bar{q}, \mathbf{I}) = 0$  and obtain  $\hat{\mathbf{f}}^{(0)}(\lambda_{\bar{q}, \mathbf{R}^{(0)}}, \bar{q}, \mathbf{R}^{(0)})$  and  $\mathbf{e}^{(0)} = \mathbf{e}^{(0)}(\lambda_{\bar{q}, \mathbf{R}^{(0)}}, \bar{q}, \mathbf{R}^{(0)})$ .
2. Iteration. Compute  $\hat{\mathbf{R}}^{(l)} = \hat{\mathbf{R}}(\mathbf{e}^{(l-1)})$  for  $l = 1, \dots, L$  and de-correlate the data via  $\mathbf{Y}^* = \{\hat{\mathbf{R}}^{(l)}\}^{-1/2} \mathbf{Y}$  and solve  $T_\lambda(\lambda, \bar{q}, \mathbf{I}; \mathbf{Y}^*) = 0$  with solution  $\lambda^{(l)}$ . Update  $\lambda^{(l)}$ ,  $\hat{\mathbf{f}}^{(l)}$  and  $\mathbf{e}^{(l)}$ .
3. Stopping. Repeat the iteration step until convergence in  $\lambda^{(l)}(\bar{q}, \mathbf{R}^{(l)})$  is achieved, say, at  $\tilde{l} < L$ . Collect  $\lambda^{(\tilde{l})}$  and  $\mathbf{R}^{(\tilde{l})}$ . If  $\tilde{l} = L$  conclude that no convergence is achieved for  $q = \bar{q}$ .

### Outer loop

1. Solve the estimating equation  $T_{\bar{q}} = T_q(\hat{\lambda}^{(\tilde{l})}, \bar{q}, \mathbf{I}; \{\hat{\mathbf{R}}^{(\tilde{l})}\}^{-1/2} \mathbf{Y})$  for each  $q \in \mathcal{Q}$  in ascending order.
2. Select the optimal  $q$  as  $\hat{q} = \lfloor \tilde{q} \rfloor$ . If  $\lambda^{(l)}|_{q=\hat{q}}$  does not converge,  $T_{\bar{q}}$  cannot be computed. If this occurs, stop and constraint the selection of  $q$  to  $\{1, \dots, \bar{q}\}$ .

As argued in subsection 3.2.4, for each  $q$  the final estimates for  $\mathbf{f}$ ,  $\sigma^2$ , and  $\mathbf{R}$  will be the same as if the natural estimators had been used, even though the corresponding value of the smoothing parameter  $\lambda$  will differ. This means that the asymptotic behaviour of the natural estimators for  $\mathbf{f}$ ,  $\sigma^2$ , and  $\mathbf{R}$  (which follows from the results in section 3.3) also applies to the naive estimators whose computation we outline in the current section. If, however, one is interested in the value of the natural smoothing parameter  $\hat{\lambda}_q$ , it can be obtain by solving  $\mathbf{S}_{\hat{\lambda}_q, \mathbf{R}}^* \mathbf{Y} = \mathbf{S}_{\hat{\lambda}_q, \mathbf{R}} \mathbf{Y}$ .

### 3.4.2. Numerical Simulations

In all settings the Monte Carlo sample is  $N = 100$ , the sample size is  $n = 500$ , and the design points are fixed and equidistant  $t = i/n, i = 1, \dots, n$ . We consider five mean functions properly scaled so their range lies in  $[0, 1]$ . Namely we use:

$$\begin{aligned} f_1(x) &= \sum_{i=\beta+1}^n \phi_{\beta,i}(x) \{\pi(i - (\beta + 1)/2)\}^{-\beta-0.1} \cos(2i), \quad \beta = 3 \\ f_2(x) &= \cos(5\pi x) \\ f_3(x) &= \sum_{i=\beta+1}^n \phi_{\beta,i}(x) \{\pi(i - (\beta + 1)/2)\}^{-\beta-0.1} \cos(2i), \quad \beta = 2 \\ f_4(x) &= 3\vartheta(x, 30, 17) + 2\vartheta(x, 3, 11), \quad \vartheta(x, p, q) = \{\Gamma(p + q)/(\Gamma(p)\Gamma(q))\}x^{p-1}(1 - x)^{q-1} \end{aligned}$$

Functions  $f_1(x)$  and  $f_2(x)$  were presented in [Serra and Krivobokova \[2016\]](#) for the case when errors are not correlated. Function  $f_3(x)$  is in  $\mathcal{W}^2[0, 1]$ , that is, it portrays the particular case where a cubic smoothing spline is, indeed, the correct choice to model the data. Function  $f_4(x)$  is introduced here and will be used in subsection [4.1.2](#) due to its flexibility under different choices of parameters  $p$  and  $q$ . [Figure \(3.1\)](#) presents a graphical depiction of the mean functions, accompanied with the corresponding simulated data for a noise level of  $\sigma = 2^{-4}$ . The same signal to noise ratio is used for all simulated dependence structures.

The simulation study aims to evaluate the performance of the algorithm described in subsection [3.4.1](#) given different correlation settings varying the dependence strength and the complexity of its structure. In particular we consider error structures generated from  $\epsilon_i = \phi_1\epsilon_{i-1} + \phi_2\epsilon_{i-2} + \epsilon_i$  with  $\epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ . [Figure \(3.2\)](#) shows the selected cases inside the parameter space of stationarity that allows us to study different correlation patterns commonly found in practice. The specific behaviour of the autocorrelation functions for these structures is presented in [Figure \(3.3\)](#).

To situate our work in the literature we compare our method with the *plug-in* estimators by [Hall and Van Keilegom \[2003\]](#) and [Herrmann et al. \[1992\]](#). In a nutshell, both

are kernel regression methods that use *difference-based* estimators for the autocorrelation function, which are later plugged into the kernel regression problem so that the bandwidth selection takes into account the dependence structure of the error. Additionally we consider the LMM representation of cubic smoothing splines with AR(p) errors as, in practice, it is the most common choice of the (unknown) smoothness class of the regression function.

It is important to note that none of the three alternative methods consider the estimation of parameter  $q$ , and hence an additional layer of complexity is added in our procedure. Another important difference is that the only assumption we make about the dependence structure of the error is its exponential decay, contrary to the more restrictive AR(p) assumption of [Hall and Van Keilegom \[2003\]](#) and the cubic smoothing splines cases. For clarity, the simulation study considers different scenarios for parameters  $q$  and  $\mathbf{R}$  as presented in Table (3.1).

Table 3.1.: Simulated scenarios

		$\mathbf{R}$	
		Unknown	known
$q$	Unknown	I	II
	known	III	IV

In scenarios I and II we consider all functions, while for scenarios III and IV only functions  $f_1(x)$  and  $f_3(x)$  are studied since the smoothness class for these non-analytic smooth functions is known by construction. All correlation structures depicted in Figure (3.2) are used for all scenarios. The results of the simulation study for each scenario are summarised in one table reporting the average of  $\text{MSE} = \sum_{i=1}^M (\hat{f}(x_i) - f(x_i))^2 / M$  and its standard deviation. Table (3.3) presents the assessment of the empirical Bayes smoothing spline method for scenario I, i.e. when the structure of the error and the smoothness class of the regression function are unknown, which is in practice the most common case. As it can be seen, the results of our method (EBS) have a better performance under all correlation structures and with respect to methods HVK and HER. In very few cases, however, we observe that the LMM

alternative is the best option. This result is not surprising since LMM uses the correct AR(p) characterisation of the noise, and EBS computes it non-parametrically. In fact, in scenario II, where we assume the noise structure is known, the EBS method performs better under all correlation structures and for all functions besides  $f_3(\cdot)$ , see Table (3.4). Tables III and IV reveal very similar information for the case when  $q$  is known, and thus the results are placed in appendix B. An interesting expected result is presented in both, Table (B.1) and Table (B.2), where the superiority of the LMM method is clear under all correlation settings for function  $f_3(x)$ . This is of course expected since  $q = 2$  is the true smoothness class of this function, and LMM assumes cubic smoothing splines.

Table (3.2) presents some statistics for the computation of the smoothness class for scenario I. We first note that functions  $f_1(x)$  and  $f_3(x)$  can be correctly identified, where the true values of  $q$  are 3 and 2 respectively. When the smoothness class of the functions is very large (or in fact infinite) we expect to underestimate  $q$ . In the case of function  $f_2(x)$ , i.e., the cosine function, we find values of  $q$  between 5 and 3; while for  $f_4(x)$  we consistently find a values of  $q$  between 2 and 3.

Table 3.2.: Simulation Results: Smoothness Class

	$f_1$		$f_2$		$f_3$		$f_4$	
$(\phi_1, \phi_2)$	Mean	Mode	Mean	Mode	Mean	Mode	Mean	Mode
(0, 0)	3.00	3	4.76	5	2.01	2	2.88	3
(0.2, 0)	2.85	3	4.22	5	2.09	2	2.59	3
(0.4, 0)	2.67	3	3.69	3	2.30	2	2.21	2
(0.6, 0)	2.94	3	3.24	3	3.46	2	2.58	2
(0.52, -0.3)	3.12	3	4.41	4	2.20	2	2.92	3
(0.3, -0.52)	3.59	3	4.40	4	3.09	2	3.34	3
(0, 0.3)	2.74	3	3.62	3	2.62	2	2.48	2
(0, -0.3)	3.05	3	4.31	5	2.36	2	3.07	3
(0, 0.6)	3.85	3	3.48	3	3.68	3	3.20	2
(0, -0.6)	4.10	3	4.85	5	3.27	2	4.02	3

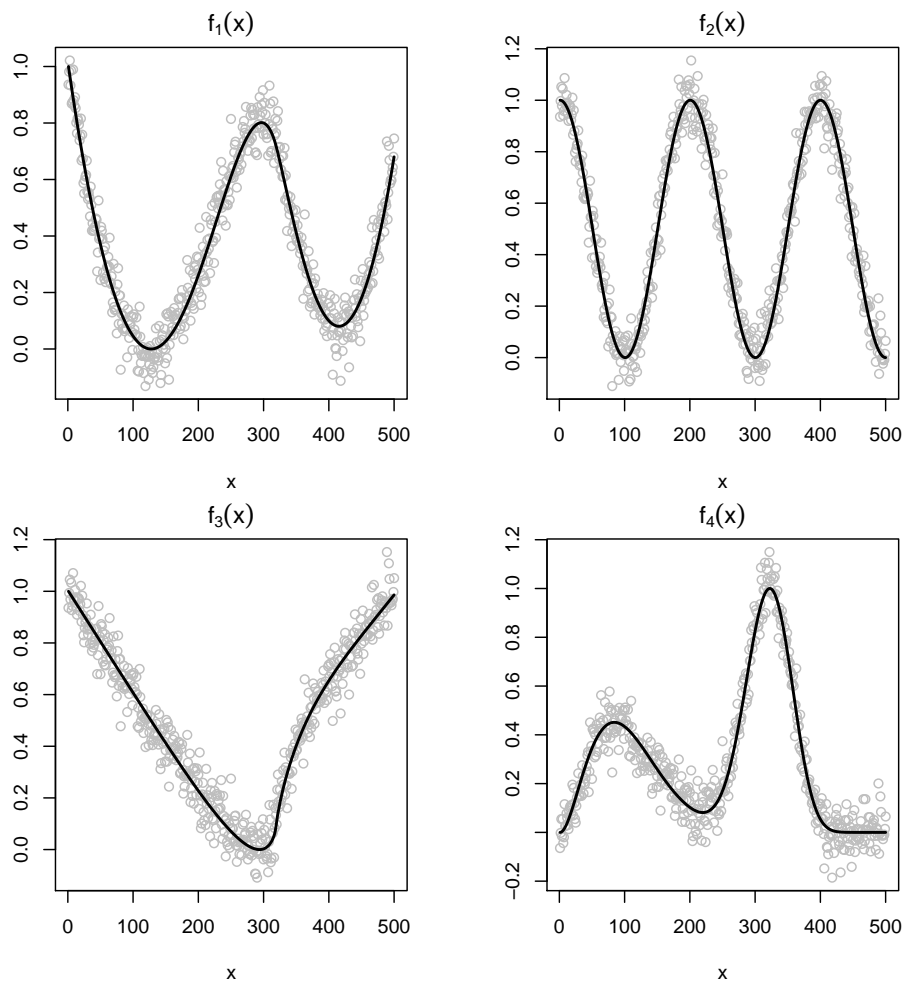


Figure 3.1.: Simulation setting. Simulated mean functions as black continuous lines and simulated data sample (for the independence case) as grey circles.

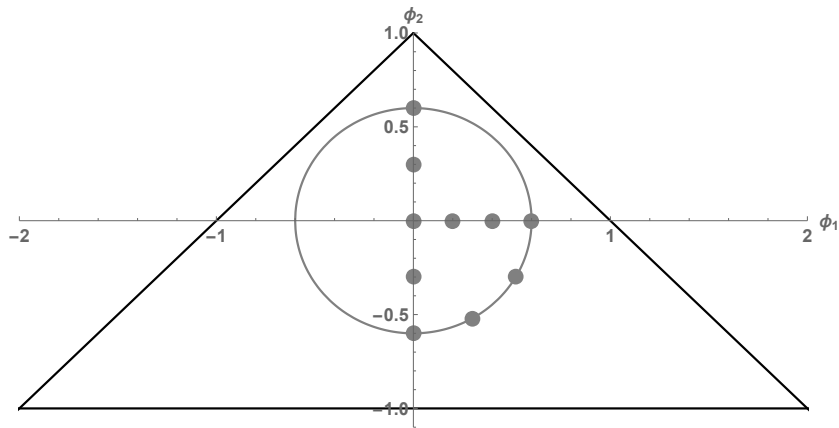


Figure 3.2.: Simulated error. The area inside the triangle represents the stationarity space for an AR2 processes. The grey points represent the cases selected for the simulation study.

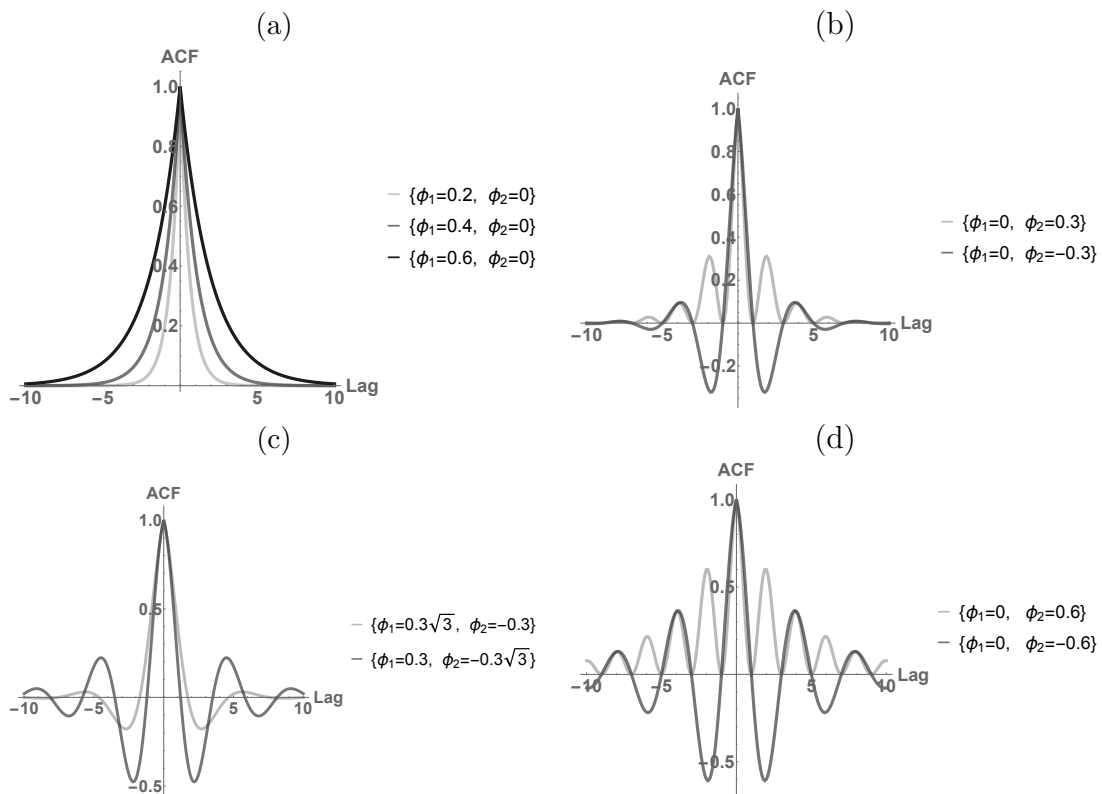


Figure 3.3.: Autocorrelation functions of the  $AR2(\phi_1, \phi_2)$  processes depicted by the grey points in Figure (3.2).

3. Smoothing Splines with Correlated Errors

Table 3.3.: Simulation Results: MSE

$(\phi_1, \phi_2)$	(0, 0)	(0.2, 0)	(0.4, 0)	(0.6, 0)	(0.52, -0.3)	(0.3, -0.52)	(0, 0.3)	(0, -0.3)	(0, 0.6)	(0, -0.6)
$f_1$										
EBS	<b>0.11</b> (0.04)	<b>0.18</b> (0.07)	<b>0.27</b> (0.08)	0.45 (0.14)	<b>0.14</b> (0.06)	<b>0.06</b> (0.02)	<b>0.21</b> (0.08)	<b>0.08</b> (0.04)	0.42 (0.16)	0.05 (0.02)
HER	0.29 (0.08)	0.33 (0.09)	0.37 (0.10)	0.49 (0.15)	0.31 (0.09)	0.25 (0.04)	0.34 (0.10)	0.25 (0.04)	0.47 (0.16)	0.25 (0.03)
LMM	0.14 (0.04)	0.20 (0.07)	0.28 (0.08)	<b>0.42</b> (0.13)	0.17 (0.06)	0.07 (0.02)	0.22 (0.07)	<b>0.08</b> (0.02)	<b>0.40</b> (0.13)	<b>0.04</b> (0.01)
HVK	0.19 (0.06)	0.66 (0.15)	0.78 (0.23)	1.34 (0.34)	0.65 (0.16)	0.59 (0.06)	1.19 (0.21)	0.60 (0.07)	2.08 (0.41)	0.58 (0.06)
$f_2$										
EBS	<b>0.11</b> (0.04)	<b>0.19</b> (0.08)	<b>0.29</b> (0.09)	<b>0.48</b> (0.18)	<b>0.16</b> (0.07)	<b>0.06</b> (0.02)	<b>0.24</b> (0.10)	<b>0.11</b> (0.06)	<b>0.48</b> (0.17)	<b>0.04</b> (0.03)
HER	0.28 (0.08)	0.32 (0.11)	0.39 (0.13)	0.51 (0.19)	0.29 (0.09)	0.23 (0.05)	0.35 (0.14)	0.24 (0.06)	0.50 (0.17)	0.20 (0.04)
LMM	0.19 (0.04)	0.26 (0.07)	0.37 (0.09)	0.57 (0.15)	0.23 (0.06)	0.10 (0.03)	0.31 (0.08)	<b>0.11</b> (0.02)	0.55 (0.13)	0.06 (0.01)
HVK	0.53 (0.11)	2.97 (0.30)	3.03 (0.37)	4.82 (1.29)	2.95 (0.41)	2.90 (0.14)	5.61 (0.48)	2.91 (0.17)	5.25 (1.15)	2.85 (0.12)
$f_3$										
EBS	<b>0.15</b> (0.04)	0.21 (0.06)	0.29 (0.08)	0.46 (0.21)	0.19 (0.06)	0.11 (0.06)	0.25 (0.07)	0.10 (0.03)	0.48 (0.20)	0.09 (0.05)
HER	0.49 (0.07)	0.53 (0.09)	0.57 (0.11)	0.66 (0.17)	0.51 (0.07)	0.47 (0.04)	0.54 (0.11)	0.47 (0.05)	0.66 (0.15)	0.45 (0.03)
LMM	<b>0.15</b> (0.04)	<b>0.20</b> (0.06)	<b>0.28</b> (0.08)	<b>0.42</b> (0.12)	<b>0.17</b> (0.05)	<b>0.08</b> (0.02)	<b>0.24</b> (0.07)	<b>0.09</b> (0.02)	<b>0.39</b> (0.12)	<b>0.05</b> (0.01)
HVK	0.16 (0.04)	0.33 (0.09)	0.45 (0.15)	0.62 (0.20)	0.32 (0.07)	0.18 (0.05)	0.52 (0.13)	0.28 (0.04)	0.85 (0.26)	0.14 (0.02)
$f_4$										
EBS	<b>0.14</b> (0.04)	<b>0.23</b> (0.08)	0.36 (0.09)	<b>0.51</b> (0.16)	<b>0.18</b> (0.07)	<b>0.07</b> (0.03)	0.29 (0.09)	<b>0.09</b> (0.04)	0.52 (0.19)	<b>0.05</b> (0.02)
HER	0.47 (0.11)	0.49 (0.13)	0.58 (0.17)	0.71 (0.23)	0.48 (0.13)	0.41 (0.08)	0.54 (0.16)	0.42 (0.07)	0.68 (0.21)	0.40 (0.04)
LMM	0.17 (0.04)	0.24 (0.07)	<b>0.34</b> (0.08)	0.52 (0.15)	0.20 (0.06)	0.09 (0.02)	<b>0.28</b> (0.08)	0.10 (0.02)	<b>0.49</b> (0.13)	<b>0.05</b> (0.01)
HVK	0.36 (0.10)	0.86 (0.15)	1.37 (0.45)	1.98 (0.35)	1.11 (0.44)	0.79 (0.10)	1.79 (0.28)	0.82 (0.21)	2.44 (0.75)	0.76 (0.05)

$\|\mathbf{f} - \hat{\mathbf{f}}\|_2$ : mean and standard deviations (in parenthesis) are reported. Quantities scaled by  $10^4$ .



3. Smoothing Splines with Correlated Errors

Table 3.4.: Simulation Results (Scenario II): MSE

$(\phi_1, \phi_2)$	(0, 0)	(0.2, 0)	(0.4, 0)	(0.6, 0)	(0.52, -0.3)	(0.3, -0.52)	(0, 0.3)	(0, -0.3)	(0, 0.6)	(0, -0.6)
$f_1$										
EBS	<b>0.11</b> (0.04)	<b>0.16</b> (0.06)	<b>0.23</b> (0.08)	<b>0.39</b> (0.13)	<b>0.13</b> (0.05)	<b>0.06</b> (0.02)	<b>0.18</b> (0.06)	<b>0.06</b> (0.02)	<b>0.38</b> (0.12)	<b>0.03</b> (0.01)
HER	0.29 (0.08)	0.33 (0.09)	0.37 (0.10)	0.49 (0.15)	0.31 (0.09)	0.26 (0.05)	0.35 (0.10)	0.25 (0.04)	0.47 (0.16)	0.24 (0.03)
LMM	0.14 (0.04)	0.20 (0.07)	0.28 (0.08)	0.42 (0.13)	0.17 (0.06)	0.07 (0.02)	0.22 (0.07)	0.08 (0.02)	0.41 (0.12)	0.04 (0.01)
HVK	0.19 (0.06)	0.66 (0.15)	0.78 (0.23)	1.34 (0.34)	0.66 (0.16)	0.60 (0.08)	1.19 (0.21)	0.60 (0.08)	2.04 (0.43)	0.58 (0.06)
$f_2$										
EBS	<b>0.13</b> (0.04)	<b>0.19</b> (0.07)	<b>0.27</b> (0.08)	<b>0.42</b> (0.13)	<b>1.58</b> (0.58)	<b>0.64</b> (0.20)	<b>2.17</b> (0.73)	<b>0.73</b> (0.21)	<b>3.86</b> (1.20)	<b>0.38</b> (0.11)
HER	0.28 (0.08)	0.32 (0.11)	0.39 (0.13)	0.51 (0.19)	2.91 (0.88)	2.22 (0.56)	3.46 (1.39)	2.46 (0.56)	4.69 (1.66)	0.49 (0.17)
LMM	0.19 (0.04)	0.26 (0.07)	0.37 (0.09)	0.57 (0.15)	2.25 (0.62)	1.00 (0.24)	3.07 (0.82)	1.10 (0.25)	5.37 (1.29)	0.61 (0.13)
HVK	0.53 (0.11)	2.97 (0.30)	3.03 (0.37)	4.83 (1.29)	29.50 (4.04)	28.56 (1.70)	56.02 (4.80)	29.24 (1.64)	49.93 (12.24)	28.89 (1.32)
$f_3$										
EBS	0.17 (0.04)	0.23 (0.06)	0.29 (0.08)	<b>0.41</b> (0.12)	0.20 (0.05)	0.09 (0.02)	0.26 (0.07)	0.10 (0.02)	<b>0.39</b> (0.12)	0.06 (0.01)
HER	0.49 (0.07)	0.53 (0.09)	0.57 (0.11)	0.66 (0.17)	0.50 (0.07)	0.46 (0.04)	0.55 (0.11)	0.47 (0.05)	0.66 (0.14)	0.46 (0.03)
LMM	<b>0.15</b> (0.04)	<b>0.20</b> (0.06)	<b>0.28</b> (0.08)	<b>0.41</b> (0.12)	<b>0.17</b> (0.05)	<b>0.08</b> (0.02)	<b>0.24</b> (0.07)	<b>0.09</b> (0.02)	<b>0.39</b> (0.12)	<b>0.05</b> (0.01)
HVK	0.16 (0.04)	0.33 (0.09)	0.45 (0.15)	0.62 (0.20)	0.32 (0.07)	0.18 (0.05)	0.53 (0.13)	0.28 (0.05)	0.85 (0.25)	0.15 (0.02)
$f_4$										
EBS	<b>0.13</b> (0.04)	<b>0.19</b> (0.07)	<b>0.29</b> (0.08)	<b>0.45</b> (0.14)	<b>0.16</b> (0.06)	<b>0.07</b> (0.02)	<b>0.23</b> (0.07)	<b>0.08</b> (0.02)	<b>0.43</b> (0.13)	<b>0.04</b> (0.01)
HER	0.47 (0.11)	0.49 (0.13)	0.58 (0.17)	0.72 (0.23)	0.48 (0.13)	0.42 (0.08)	0.54 (0.16)	0.42 (0.07)	0.70 (0.22)	0.21 (0.03)
LMM	0.17 (0.04)	0.24 (0.07)	0.34 (0.08)	0.52 (0.15)	0.20 (0.06)	0.09 (0.02)	0.28 (0.08)	0.10 (0.02)	0.51 (0.13)	0.06 (0.01)
HVK	0.36 (0.10)	0.86 (0.15)	1.37 (0.45)	1.98 (0.35)	1.11 (0.44)	0.79 (0.09)	1.79 (0.27)	0.81 (0.18)	2.43 (0.75)	0.76 (0.06)

$\|\mathbf{f} - \hat{\mathbf{f}}\|_2$ : mean and standard deviations (in parenthesis) are reported. Quantities scaled by  $10^4$ .

### 3.5. Digression on Smoothness Classes

In this section a digression on the estimation of the smoothness class of function  $f \in \mathcal{W}^q[0, 1]$  is presented following Krivobokova [2013]. In this document a different estimator for  $q$  is constructed, where the main idea is to select a value of  $q$  such that the estimating equations for the smoothing parameter under the frequentist framework  $\lambda_f$  and Bayesian framework  $\lambda_{r|f}$  are equivalent. Namely

$$\hat{\mathcal{R}}_q(\hat{\lambda}) = \frac{1}{n} \left[ \mathbf{Y}^T (\mathbf{I}_n - \mathbf{S}_{\hat{\lambda},q}) \mathbf{S}_{\hat{\lambda},q}^2 \mathbf{Y} - \sigma^2 \{ \text{tr}(\mathbf{S}_{\hat{\lambda},q}^2) - q \} \right], \quad (3.28)$$

is suggested for some  $\hat{\lambda} = \lambda_f = \lambda_{r|f}$ . The asymptotic behaviour of this estimator is described in the following theorem.

**Theorem 8.** *Estimator  $\hat{\mathcal{R}}_q(\hat{\lambda})$  converges to*

$$\mathcal{R}_q(\lambda) = \frac{1}{n} \left[ \mathbf{f}^T (\mathbf{I}_n - \mathbf{S}_{\lambda,q}) \mathbf{S}_{\lambda,q}^2 \mathbf{f} - \sigma^2 \{ \text{tr}(\mathbf{S}_{\lambda,q}^2) - q \} \right]$$

via

$$\lambda_{r|f}^{1/(4q)} (\hat{\mathcal{R}}_q(\hat{\lambda}_r) - \mathcal{R}_q(\lambda_{r|f})) \xrightarrow{D} \mathcal{N} \left( 0, \frac{\sigma^4}{384c(\rho)} \left[ C_1(q) + \frac{C_2(q)}{5} \right] \right), \quad (3.29)$$

where

$$\begin{aligned} C_1(q) &= \frac{(2q-1)(8q^2+4q-1)^2}{(2q+1)q^5 \text{sinc}\{\pi/(2q)\}} \\ C_2(q) &= \frac{(2q+1)(2q-1)(4q-1)(6q-1)}{q^5 \text{sinc}\{\pi/(2q)\}}. \end{aligned}$$

*Proof.* The proof follows from Lemma 3 in Krivobokova [2013] and uses (A.6). Here it is organised in three steps.

*Step 1: convergence of  $\mathcal{R}_q(\hat{\lambda}_r)$*

From the aforementioned results it holds that

$$(\hat{\lambda}_r - \lambda_{r|f}) \xrightarrow{D} \mathcal{N} \left( 0, \lambda_{r|f}^{2+1/(2q)} \frac{2c(\rho) \operatorname{sinc}\{\pi/(2q)\} q}{12q^2 - 3} \right) \text{ as } n \rightarrow \infty. \quad (3.30)$$

To apply the delta method for function  $n\mathcal{R}_q(\lambda)$  we need to show that  $\partial\{n\mathcal{R}_q(\lambda)\}/\partial\lambda$  exists and that it is different than zero when evaluated at  $\lambda_{r|f}$ . Namely

$$\frac{\partial\{n\mathcal{R}_q(\lambda)\}}{\partial\lambda} = \frac{1}{\lambda} \left[ 3\mathbf{f}^T (\mathbf{I}_n - \mathbf{S}_{\lambda,q}) \mathbf{S}_{\lambda,q}^3 \mathbf{f} - 2\mathbf{f}^T (\mathbf{I}_n - \mathbf{S}_{\lambda,q}) \mathbf{S}_{\lambda,q}^2 \mathbf{f} + 3\sigma^2 \operatorname{tr}\{(\mathbf{I}_n - \mathbf{S}_{\lambda,q}) \mathbf{S}_{\lambda,q}^3\} \right]$$

whose existence is obvious. Furthermore it can be shown that

$$\begin{aligned} \left. \frac{\partial\{n\mathcal{R}_q(\lambda)\}}{\partial\lambda} \right|_{\lambda=\lambda_{r|f}} &= \frac{\sigma^2}{\lambda_{r|f}} \left[ (\operatorname{tr}\{\mathbf{S}_{\lambda_{r|f},q}^2\} - q) + 3\operatorname{tr}\{(\mathbf{I}_n - \mathbf{S}_{\lambda_{r|f},q}) \mathbf{S}_{\lambda_{r|f},q}^3\} \right], \\ &= \frac{\sigma^2 \lambda_{r|f}^{-1-1/(2q)}}{16c(\rho)} \left[ \frac{(2q-1)(8q^2+4q-1)}{q^3 \operatorname{sinc}\{\pi/(2q)\}} \right] \{1 + o(1)\}, \end{aligned}$$

where in the last step we made use of (A.6). The delta method can be then applied using (3.30) and obtain

$$\lambda_{r|f}^{1/(4q)} (\mathcal{R}_q(\hat{\lambda}_r) - \mathcal{R}_q(\lambda_{r|f})) \xrightarrow{D} \mathcal{N} \left( 0, \frac{\sigma^4}{384c(\rho)} C_1(q) \right), \quad (3.31)$$

with

$$C_1(q) = \frac{(2q-1)(8q^2+4q-1)^2}{(2q+1)q^5 \operatorname{sinc}\{\pi/(2q)\}},$$

which decreases with  $q$ .

*Step 2: convergence of  $\hat{\mathcal{R}}_q(\lambda_{r|f})$*

### 3. Smoothing Splines with Correlated Errors

---

For  $\hat{\mathcal{R}}_q(\lambda)$  defined as in (3.28), under some fixed  $\lambda$  and the assumption of Gaussian errors, it is easy to verify that

$$\begin{aligned}\mathbb{E}\{n\hat{\mathcal{R}}_q(\lambda)\} &= \mathbf{f}^T(\mathbf{I}_n - \mathbf{S}_{\lambda,q})\mathbf{S}_{\lambda,q}^2\mathbf{f} - \sigma^2\{\text{tr}(\mathbf{S}_{\lambda,q}^3) - q\}, \\ \text{var}\{n\hat{\mathcal{R}}_q(\lambda)\} &= 2\sigma^2[2\mathbf{f}^T(\mathbf{I}_n - \mathbf{S}_{\lambda,q})^2\mathbf{S}_{\lambda,q}^4\mathbf{f} + \sigma^2\{\text{tr}(\mathbf{I}_n - \mathbf{S}_{\lambda,q})^2\mathbf{S}_{\lambda,q}^4\}].\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}\{n\hat{\mathcal{R}}_q(\lambda_{r|f})\} &= \frac{\sigma^2\lambda_{r|f}^{-1/(2q)}}{8c(\rho)} \left[ \frac{(2q-1)}{q^2\text{sinc}\{\pi/(2q)\}} \right] \{1 + o(1)\}, \\ \text{var}\{n\hat{\mathcal{R}}_q(\lambda_{r|f})\} &= \frac{\sigma^4\lambda_{r|f}^{-1/(2q)}}{1920c(\rho)} \left[ \frac{(2q+1)(2q-1)(4q-1)(6q-1)}{q^5\text{sinc}\{\pi/(2q)\}} \right] \{1 + o(1)\},\end{aligned}$$

where we have once again made use of (A.6). To apply the central limit theorem (CLT), consider the Demmler-Reinsch basis, i.e.  $\mathbf{S}_{\lambda,q} = \mathbf{\Phi}_q \text{diag}\{(1 + \lambda n \eta_{q,i})^{-1}\} \mathbf{\Phi}_q^T$ , so that  $\text{tr}(\mathbf{S}_{\lambda,q}^l) = \sum_{i=q+1}^n (1 + \lambda n \eta_{q,i})^{-l}$ . Hence we can define

$$n[\hat{\mathcal{R}}_q(\lambda_{r|f}) - \mathbb{E}\hat{\mathcal{R}}_q(\lambda_{r|f})] = \sum_{i=q+1}^n (d_i^2 - b_i^2 - \sigma^2) \frac{\lambda_{r|f} n \eta_{q,i}}{(1 + \lambda_{r|f} n \eta_{q,i})^3} + o_p(1) =: \sum_{i=q+1}^n \xi_i,$$

where  $\{\xi_i, i = q+1, \dots, n\}$  is a sequence of independent (but non-identically distributed) random variables with  $\mathbb{E}(\xi_i) = o(1)$  and  $s_n^2 = \sum_{i=q+1}^n \text{var}(\xi_i) = 2\sigma^4 \text{tr}\{(\mathbf{I}_n - \mathbf{S}_{\lambda_{r|f},q})^2 \mathbf{S}_{\lambda_{r|f},q}^4\} \{1 + o(1)\} = \text{const} \lambda_{r|f}^{-1/(2q)}$ . Moreover it can be shown that  $\text{var}(\xi_i) = o(1)$ , for which there must exist some constant  $B$  bounding  $\mathbb{E}(\xi_i^2)$ . Then for a finite  $\mathbb{E}(\xi_i^4)$ , we can write

$$\frac{1}{s_n^4} \sum_{i=q+1}^n \mathbb{E}(\xi_i^4) < \frac{1}{s_n^4} B \sum_{i=q+1}^n \mathbb{E}(\xi_i^2) = \frac{B}{s_n^2},$$

which converges to 0 as  $n \rightarrow \infty$ , fulfilling the Lyapunov condition. Hence we apply the CLT to obtain

$$\lambda_{r|f}^{1/(4q)} (\hat{\mathcal{R}}_q(\lambda_{r|f}) - \mathcal{R}_q(\lambda_{r|f})) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^4}{1920c(\rho)} C_2(q)\right), \quad (3.32)$$

with

$$C_2(q) = \frac{(2q+1)(2q-1)(4q-1)(6q-1)}{q^5 \text{sinc}\{\pi/(2q)\}},$$

which decreases with  $q$ .

*Step 3: convergence of  $\hat{\mathcal{R}}_q(\hat{\lambda}_r)$*

One can evaluate (3.28) at random  $\hat{\lambda}_r$  to obtain

$$\hat{\mathcal{R}}_q(\hat{\lambda}_r) = \frac{1}{n}[\mathbf{Y}^T(\mathbf{I}_n - \mathbf{S}_{\hat{\lambda}_r})\mathbf{S}_{\hat{\lambda}_r}^2\mathbf{Y} - \sigma^2\{\text{tr}(\mathbf{S}_{\hat{\lambda}_r}^2) - q\}], \quad (3.33)$$

such that

$$\begin{aligned} \mathbb{E}\{n\hat{\mathcal{R}}_q(\hat{\lambda}_r)\} &= \mathbb{E}[\mathbb{E}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r)|\hat{\lambda}_r = \hat{\lambda}_r.]] = \mathbb{E}[n\mathcal{R}_q(\hat{\lambda}_r.)], \\ \text{var}\{n\hat{\mathcal{R}}_q(\hat{\lambda}_r)\} &= \mathbb{E}[\text{var}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r)|\hat{\lambda}_r = \hat{\lambda}_r.]] + \text{var}[\mathbb{E}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r)|\hat{\lambda}_r = \hat{\lambda}_r.]] \\ &= \mathbb{E}[\text{var}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r.)]] + \text{var}[n\mathcal{R}_q(\hat{\lambda}_r.)]. \end{aligned}$$

Using (3.31) it is clear that

$$\mathbb{E}[n\mathcal{R}_q(\hat{\lambda}_r.)] \rightarrow n\mathcal{R}_q(\lambda_{r|f}) \quad \text{and} \quad \text{var}[n\mathcal{R}_q(\hat{\lambda}_r.)] \rightarrow \frac{\sigma^4 \lambda_{r|f}^{-1/(2q)} C_1(q)}{384c(\rho)}$$

as  $n \rightarrow \infty$ . Furthermore applying the delta method on  $\text{var}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r.)]$  and using (3.32), it can be shown that

$$\mathbb{E}[\text{var}[n\hat{\mathcal{R}}_q(\hat{\lambda}_r.)]] \rightarrow \text{var}[n\hat{\mathcal{R}}_q(\lambda_{r|f})] = \frac{\sigma^4 \lambda_{r|f}^{-1/(2q)} C_2(q)}{1920c(\rho)}.$$

Hence, to make use of the CLT we build random variables

$$\begin{aligned}
 n[\hat{\mathcal{R}}_q(\hat{\lambda}_r) - \mathbb{E}\hat{\mathcal{R}}_q(\hat{\lambda}_r)] &= \sum_{i=q+1}^n \left\{ \left[ d_i^2 \frac{\hat{\lambda}_r n \eta_{q,i}}{(1 + \hat{\lambda}_r n \eta_{q,i})^3} \right] - \left[ b_i^2 \frac{\lambda_{r|f} n \eta_{q,i}}{(1 + \lambda_{r|f} n \eta_{q,i})^3} \right] \right. \\
 &\quad \left. - \left[ \sigma^2 \left( \frac{1}{(1 + \hat{\lambda}_r n \eta_{q,i})^2} - \frac{1}{(1 + \lambda_{r|f} n \eta_{q,i})^3} \right) \right] \right\} + o_p(1) \\
 &=: \sum_{i=q+1}^{k+n} \xi_i,
 \end{aligned}$$

where  $\{\xi_i, i = q + 1, \dots, n\}$  is a sequence of independent (but non-identically distributed) random variables with  $\mathbb{E}(\xi_i) = o(1)$  and  $s_n^2 = \text{const} \lambda_{r|f}^{-1/(2q)}$ . Using similar arguments as the ones given in the previous step is then easy to find a bound

$$\frac{1}{s_n^4} \sum_{i=q+1}^n \mathbb{E}(\xi_i^4) < \frac{1}{s_n^4} B \sum_{i=q+1}^n \mathbb{E}(\xi_i^2) = \frac{B}{s_n^2},$$

which goes to 0 as  $n \rightarrow \infty$  for some  $B$  bounding  $\mathbb{E}(\xi_i^2)$  and  $\mathbb{E}(\xi_i^4) < \infty$ . The application of the CLT then leads to the result

$$\lambda_{r|f}^{1/(4q)} (\hat{\mathcal{R}}_q(\hat{\lambda}_r) - \mathcal{R}_q(\lambda_{r|f})) \xrightarrow{D} \mathcal{N} \left( 0, \frac{\sigma^4}{384c(\rho)} \left[ C_1(q) + \frac{C_2(q)}{5} \right] \right), \quad (3.34)$$

with  $C_1(q)$  and  $C_2(q)$  as in (3.31) and (3.32) respectively.  $\square$

Since the estimator  $\hat{\mathcal{R}}_q(\hat{\lambda}_{r|f})$  can be seen as another criteria to estimate the smoothness class of the function  $f \in \mathcal{W}^q[0, 1]$ , it is only natural to compare it with the estimating equation  $T_q(\lambda, q, \mathbf{I})$  presented in Chapter 3. For clarity here we re-write (a properly scaled version of) the estimators

$$\begin{aligned}
 \hat{\mathcal{R}}_q(\lambda_{r|f}) &= \sum_{i=1}^n \left\{ \frac{W_i^2 \lambda_{r|f} n \eta_{q,i} - \hat{\sigma}^2 (1 + \lambda_{r|f} n \eta_{q,i})}{(1 + \lambda_{r|f} n \eta_{q,i})^3} \right\} + \hat{\sigma}^2 q \\
 T_q(\lambda) &= \sum_{i=1}^n \left\{ \frac{W_i^2 \lambda n \eta_{q,i} - \hat{\sigma}^2 (1 + \lambda n \eta_{q,i})}{(1 + \lambda n \eta_{q,i})^3} \right\} \log(n \eta_{q,i}) (1 + \lambda n \eta_{q,i}),
 \end{aligned}$$

where the difference in the selection of the smoothing parameters is made explicit.

### 3. Smoothing Splines with Correlated Errors

---

To compare the estimators denote  $\lambda_{r|f} = \alpha\lambda$  for  $\alpha \in \mathbb{R}_+$  and consider  $\sigma^2 \rightarrow 0$ . Straightforward manipulations show that both solutions are equivalent for  $\alpha > 1$ , i.e.  $\lambda_{r|f} < \lambda$ , which is expected since it is known that smoothing splines under the frequentist framework (i.e. for  $\lambda_f = \lambda_{r|f}$ ) undersmooths the estimation. A more detailed comparison between the two proposed estimators of the smoothness class is outside the scope of this thesis.

## 4. Extensions of Smoothing Splines with Correlated Errors

In this chapter we extend the model presented in Chapter 3 using additive models and functional data. In Section 4.1 we present the general formulation of additive models under our framework and further explore its use as a tool for signal decomposition, i.e. decoupling a signal in trend, seasonal and error components. Section 4.2 presents the use of the empirical Bayes smoothing splines when the collected data are curves (as opposite to points) and their co-movement is studied by functional principal components.

### 4.1. Signal Decomposition Analysis

Identification and understanding of different components of a time series is of great importance in the applied sciences, and there are many methods available to cope with the problem. In economics the decoupling of trend and seasonal components in macroeconomic signals like GDP is of great interest to identify the so-called economic cycle. Likewise, in electrical engineering signal processing demands the use of tools to discriminate between the frequency information of signals for the study of electronic circuits. A similar situation takes place in the life and atmospheric sciences, where the interest lies in investigating the smooth trajectory of certain variables that are considered to be functional in nature, e.g. growth or temperature data [cf. [Ramsay](#)



and Silverman, 1997, Ferraty and Vieu, 2006], so the distinction between signal and noise is a crucial matter.

Alexandrov et al. [2012] and Pollock [2006] provide a detailed discussion on the history of the time series decomposition problem, as well as on the advantages and disadvantages of various tools. In general, these methods form two main clusters: the so-called model-based approach and non-parametric techniques. Probably the most common tool in applied sciences is the model-based approach that requires the specification of time series either by an ARIMA or by a structured time series model. The former has been popularised thanks to the automatic software implementations given in TRAMO-SEATS [cf. Maravall and Caporello, 2004] and X-12-ARIMA [cf. Dagum, 1978, Findley et al., 1998], delivered by Banco de España and the Bureau of Census in the US respectively. The development of methods for structural time series models is more recent. In this case, the representation of the problem in state-space form [cf. Harvey, 1989, West and Harrison, 1997, Young et al., 1999, Godolphin, 2001, Zivot and Wang, 2006] has acquired great reputation since it allows to take advantage of the seminal Kalman [1960] filter as implemented, for example, in Koopman's STAMP package [cf. Mendelsohn, 2011]. In contrast to the model-based approach, non-parametric methods do not require any a priori specifications and are popular because of their simplicity. Among the most cited alternatives are the filters proposed by Henderson [1916], Butterworth [1930], Cleveland et al. [1990], Hodrick and Prescott [1997], Baxter and King [1999], and Christiano and Fitzgerald [2003]. As discussed by Kauermann et al. [2011], the main challenge of all non-parametric methods is the selection of a data driven smoothing parameter.

The method presented in subsection 4.1.2 contributes to the signal decomposition literature and can be considered a member of the non-parametric cluster, with the difference that not only the smoothing parameter is data driven, but also the selection of the smoothness class is considered and non-parametric errors are allowed. To present the topic we first introduce a general additive model framework for empirical

Bayes smoothing splines and take signal decomposition as an specific application under this setting. We propose an estimation algorithm and show simulation exercises to illustrate the performance of the method.

### 4.1.1. Additive Models

Consider the following extension of the smoothing splines model as presented in definition 6. Namely in an additive model we aim to solve

$$\min_{f_1 \in \mathcal{W}^{q_1}[0,1], \dots, f_d \in \mathcal{W}^{q_d}[0,1]} \left[ \frac{1}{n} \left( \mathbf{Y} - \sum_{k=0}^d \mathbf{f}_k \right)^T \mathbf{R}^{-1} \left( \mathbf{Y} - \sum_{k=0}^d \mathbf{f}_k \right) + \sum_{k=1}^d \lambda_k \int_0^1 \left\{ f_k^{(q_k)}(x) \right\}^2 dx \right] \quad (4.1)$$

for  $\lambda_1, \dots, \lambda_d \in \mathbb{R}_+$ ,  $q_1, \dots, q_d \in \mathbb{N}$ , and where data tuple  $(x_{i,1}, \dots, x_{i,k}, Y_i)$  follow  $Y_i = f(x_i) + \epsilon = \sum_{k=0}^d f_k(x_{i,k}) + \epsilon_i$  for  $f_0(x) := \beta_0$  and  $\{\epsilon_i\}_{i=1}^n$  is a homoscedastic stationary process with correlation  $\text{Cor}(\epsilon_i, \epsilon_j) = r_{i,j}$  and correlation matrix  $\mathbf{R}$ .

**Theorem 9.** *The smoother matrices corresponding to the minimisation problem in (4.1) are given by*

$$\begin{aligned} \mathbf{G}(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) &= \mathbf{S}_{[-j]} + (\mathbf{I} - \mathbf{S}_{[-j]})\mathbf{G}_j \\ \mathbf{G}_j(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) &= \mathbf{C}_j(\mathbf{C}_j^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}_{[-j]})\mathbf{C}_j + \lambda_j n \mathbf{D}_j)^{-1} \mathbf{C}_j^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{S}_{[-j]}), \end{aligned}$$

where  $\mathbf{C}_j = \mathbf{C}_{j,q_j}$  denotes an arbitrary basis of  $\mathcal{W}^{q_j}[0, 1]$ , and

$$\mathbf{S}_{[-j]}(\boldsymbol{\lambda}_{[-j]}, \mathbf{q}_{[-j]}, \mathbf{R}) = \mathbf{C}_{[-j]} (\mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} + \lambda_{[-j]} n \mathbf{D}_{[-j]})^{-1} \mathbf{C}_{[-j]}^T \mathbf{R}^{-1},$$

where the subindex  $[-j]$  denotes the exclusion of the  $j$ -th element in  $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_d\}$ ,  $\mathbf{q} = \{q_1, \dots, q_d\}$  and matrices  $\mathbf{C} = [\mathbf{C}_1, \dots, \mathbf{C}_d]$  and  $\mathbf{D} = \text{blockdiag}\{\mathbf{D}_1, \dots, \mathbf{D}_d\}$ .

*Proof.* The proof is an extension of Result 1 reported by Aerts et al. [2002] when correlated errors are considered. Without lost of generality, consider blocks  $j$  and

$[-j]$  so the design matrix reads  $\mathbf{C} = [\mathbf{C}_{[-j]}, \mathbf{C}_j]$ . To ease the notation we write  $\mathbf{G} = \mathbf{G}(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R})$  and  $\mathbf{S} = \mathbf{S}(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R})$ . By definition we have

$$\mathbf{G} = \mathbf{C} \begin{pmatrix} \mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} + \boldsymbol{\lambda}_{[-j]} n \mathbf{D}_{[-j]} & \mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_j \\ \mathbf{C}_j^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} & \mathbf{C}_j^T \mathbf{R}^{-1} \mathbf{C}_j + \boldsymbol{\lambda}_j n \mathbf{D}_j \end{pmatrix}^{-1} \mathbf{C}^T,$$

or simply  $\mathbf{G} = \mathbf{C} \mathbf{A} \mathbf{C}^T$ . Following block matrix inversion formulae we obtain

$$\mathbf{G} = \mathbf{S}_{[-j]} + \mathbf{C} \begin{pmatrix} \mathbf{B} \mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_j \mathbf{H} \mathbf{C}_j^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} \mathbf{B} & -\mathbf{B} \mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_j \mathbf{H} \\ -\mathbf{H} \mathbf{C}_j^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} \mathbf{B} & \mathbf{H} \end{pmatrix} \mathbf{C}^T$$

for  $\mathbf{B} = (\mathbf{C}_{[-j]}^T \mathbf{R}^{-1} \mathbf{C}_{[-j]} + \boldsymbol{\lambda}_{[-j]} n \mathbf{D}_{[-j]})^{-1}$ , and  $\mathbf{H} = (\mathbf{C}_j^T (\mathbf{I} - \mathbf{S}_{[-j]}) \mathbf{C}_j + \boldsymbol{\lambda}_j n \mathbf{D}_j)^{-1}$ . For the second summand straight forward matrix computations and proper factorisation allows us to further simplify the smoother matrix to

$$\begin{aligned} \mathbf{G} &= \mathbf{S}_{[-j]} + (\mathbf{I} - \mathbf{S}_{[-j]}) \mathbf{C}_j \mathbf{H} \mathbf{C}_j^T (\mathbf{I} - \mathbf{S}_{[-j]}) \\ &= \mathbf{S}_{[-j]} + (\mathbf{I} - \mathbf{S}_{[-j]}) \mathbf{G}_j, \end{aligned}$$

for  $\mathbf{G}_j = \mathbf{C}_j \mathbf{H} \mathbf{C}_j^T (\mathbf{I} - \mathbf{S}_{[-j]})$ , or more precisely

$$\mathbf{G}_j = \mathbf{C}_j \{ \mathbf{C}_j^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{[-j]}) \mathbf{C}_j + \boldsymbol{\lambda}_j n \mathbf{D}_j \}^{-1} \mathbf{C}_j \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{[-j]}).$$

□

The solution of (4.1) reads

$$\begin{aligned} \hat{\mathbf{f}}_j(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) &= \mathbf{G}_j(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) \mathbf{Y} \\ \hat{\mathbf{f}}(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) &= \sum_{k=1}^d \mathbf{G}_k(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) \mathbf{Y} = \mathbf{G}(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) \mathbf{Y}, \end{aligned} \quad (4.2)$$

which depends on parameters  $\boldsymbol{\lambda}$ ,  $\mathbf{q}$  and  $\mathbf{R}$ , whose values can be computed via estimating equations in an analogous fashion to the univariate case as presented in

Chapter 3. Namely, for design matrices  $\mathbf{X}_j(x_j) = \{\phi_{2,q_j}(x_j), \dots, \phi_{q_j,q_j}(x_j)\}$  and  $\mathbf{Z}_j(x_j) = \{\eta_{q_j,q_j+1}^{-1/2}\phi_{q_j,q_j+1}(x_j), \dots, \eta_{q_j,n}^{-1/2}\phi_{q_j,n}(x_j)\}$  one arrives to the standard linear mixed model

$$\mathbf{Y}|\mathbf{u}_1, \dots, \mathbf{u}_d = \beta_0 + \sum_{k=1}^d (\mathbf{X}_k \boldsymbol{\beta}_k + \mathbf{Z}_k \mathbf{u}_k) + \boldsymbol{\epsilon}, \quad \mathbf{u}_k \sim \mathcal{N}(0, \sigma_{u_k}^2, \mathbf{I}_{n-q_k}),$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2, \mathbf{I}_n), \quad k = 1, \dots, d,$$

where  $\beta_0$  is the intercept. The profiled restricted log-likelihood can then be derived to obtain

$$\ell(\boldsymbol{\lambda}, \mathbf{q}, \mathbf{R}) = -\frac{n}{2} \hat{\sigma}^2 + \frac{1}{2} \log |\mathbf{R}^{-1}(\mathbf{I} - \mathbf{G})|_+, \quad (4.3)$$

for  $\hat{\sigma}^2 = \mathbf{Y}^T \mathbf{R}^{-1}(\mathbf{I} - \mathbf{G})\mathbf{Y}/n$ . With the previous results the estimating equations for  $\boldsymbol{\lambda}$  and  $\mathbf{q}$  can be written as

$$T_{\lambda_j} = -2\lambda_j \hat{\sigma}^2 \frac{\partial \ell(\lambda_1, \dots, \lambda_d, q_1, \dots, q_d, \mathbf{I}; \mathbf{Y}^*)}{\partial \lambda_j}$$

$$= \mathbf{Y}^{*T}(\mathbf{I} - \mathbf{S}_{-j})(\mathbf{I} - \mathbf{G}_j)\mathbf{G}_j\mathbf{Y}^* - \hat{\sigma}^2 \text{tr}(\mathbf{G}_j) \quad (4.4)$$

$$T_{q_j} = -2q_j \hat{\sigma}^2 \frac{\partial \ell(\lambda_1, \dots, \lambda_d, q_1, \dots, q_d, \mathbf{I}; \mathbf{Y}^*)}{\partial q_j}$$

$$= -q_j \mathbf{Y}^{*T}(\mathbf{I} - \mathbf{S}_{-j}) \frac{\partial \mathbf{G}_j}{\partial q_j} \mathbf{Y}^* - \hat{\sigma}^2 \text{tr} \left\{ -q_j (\mathbf{I} - \mathbf{G}_j)^{-1} \frac{\partial \mathbf{G}_j}{\partial q_j} \right\}, \quad (4.5)$$

where  $\mathbf{Y}^* = \mathbf{R}^{-1/2}\mathbf{Y}$  and

$$-q_j \frac{\partial \mathbf{G}_j}{\partial q_j} = \mathbf{G}_j(\mathbf{I} - \mathbf{S}_{-j})^{-1} \mathbf{C}_j \text{diag} \{ \lambda_j n \eta_{j,i} \log(n \eta_{j,i}) \} \mathbf{C}_j^T \mathbf{G}_j.$$

The previous formulation is general and no assumptions regarding the basis  $\mathbf{C}_j$  were made. In the following subsections we will make intensive use of the Demmler-Reinsch basis to reduce the computational complexity of the solutions.

### 4.1.2. Model for Signal Extraction

For data pairs  $(x_i, Y_i)$ ,  $i = 1, \dots, n$ , we are interested in the following (sometimes called *classical*) smooth decomposition scheme:

$$Y_i = \tau(x_i) + \underbrace{\sum_{j=1}^l \alpha_j(x_i) \cos(x_i \omega_j) + \beta_j(x_i) \sin(x_i \omega_j)}_{\varsigma(x_i)} + \epsilon_i, \quad (4.6)$$

where  $\tau(\cdot)$  is an unknown smooth function that represents a deterministic trend and  $\varsigma(\cdot)$  describes the seasonal component given frequencies  $\omega_j$  and smooth functions  $\alpha_j(\cdot)$  and  $\beta_j(\cdot)$  that modulate seasonal patterns over  $x$  as varying coefficients [cf. [Hastie and Tibshirani, 1993](#)]. Given  $\{\epsilon_i\}_{i=1}^n$  a stationary homoscedastic error of unknown structure, we consider (4.6) in its simplest form, i.e.  $l = 1$  and  $\beta(x_i) := 0$  and aim to solve a variant of (4.1), namely

$$\min_{\tau \in \mathcal{W}^{q_1}, \alpha \in \mathcal{W}^{q_\alpha}} \left[ \frac{1}{n} (\mathbf{Y} - \boldsymbol{\tau} - \text{diag}\{\cos(x_i \omega)\} \boldsymbol{\alpha})^T \mathbf{R}^{-1} (\mathbf{Y} - \boldsymbol{\tau} - \text{diag}\{\cos(x_i \omega)\} \boldsymbol{\alpha}) + \lambda_\tau \int_0^1 \{\tau(x)^{(q_1)}\}^2 dx + \lambda_\alpha \int_0^1 \{\alpha(x)^{(q_\alpha)}\}^2 dx \right], \quad (4.7)$$

with corresponding solution given by (4.2) for  $d = 2$  and estimators  $\boldsymbol{\tau} = \mathbf{G}_1 \mathbf{Y}$  and  $\boldsymbol{\varsigma} = \mathbf{G}_2 \mathbf{Y}$  for the trend and seasonal components respectively. First note that if  $\mathbf{C}_1 = \boldsymbol{\Phi}_{q_1}$  and  $\mathbf{C}_2 = \tilde{\boldsymbol{\Phi}}_{q_2} = \text{diag}\{\cos(x_i \omega)\} \boldsymbol{\Phi}_{q_2}$ , then one can write the trend by spanning a Demmler-Reinsch basis obtaining the naive smoother matrices

$$\mathbf{S}_1(\lambda_1, q_1, \mathbf{R}) = \mathbf{R}^{1/2} \boldsymbol{\Phi}_{q_1} \text{diag}\{(1 + \lambda_1 n \eta_{q_1, i})^{-1}\} \boldsymbol{\Phi}_{q_1}^T \mathbf{R}^{-1/2} \quad (4.8)$$

$$\mathbf{G}_2(\lambda_1, \lambda_2, q_1, q_2, \mathbf{R}) = \tilde{\boldsymbol{\Phi}}_{q_2} \Lambda^{-1} \tilde{\boldsymbol{\Phi}}_{q_2}^T \mathbf{R}^{-1/2} \boldsymbol{\Phi}_{q_1} \text{diag}\left\{\frac{\lambda_1 n \eta_{q_1, i}}{1 + \lambda_1 n \eta_{q_1, i}}\right\} \boldsymbol{\Phi}_{q_1}^T \mathbf{R}^{-1/2}, \quad (4.9)$$

where

$$\Lambda = \tilde{\boldsymbol{\Phi}}_{q_2}^T \mathbf{R}^{-1/2} \boldsymbol{\Phi}_{q_1} \text{diag}\left\{\frac{\lambda_1 n \eta_{q_1, i}}{1 + \lambda_1 n \eta_{q_1, i}}\right\} \boldsymbol{\Phi}_{q_1}^T \mathbf{R}^{-1/2} \tilde{\boldsymbol{\Phi}}_{q_2} + \lambda_2 n \text{diag}\{\eta_{q_2, i}\}.$$

With the previous expressions the seasonal component can be obtained directly, and since the computation  $\mathbf{G}(\lambda_1, \lambda_2, q_1, q_2, \mathbf{R}) = \mathbf{S}_1 + (\mathbf{I} - \mathbf{S}_1)\mathbf{G}_2$  is immediate, the trend component can be recovered via  $\boldsymbol{\tau} = (\mathbf{G} - \mathbf{G}_2)\mathbf{Y}$ . The rest of the parameters can then be obtained by the corresponding estimating equations (4.4) and (4.5) following an iterative search algorithm analogue to procedure presented for the univariate case in Section 3.4.1 for grids  $(\lambda_1, \lambda_2)$  and  $(q_1, q_2)$  respectively. This direction carries a high computational cost and instead we consider a sequential approach.

### Statistical Algorithm

Assume  $l$  and  $\omega$  are known (if not, they can be estimated via e.g. Fourier analysis, etc.), and consider cubic smoothing splines to model both the trend and seasonal components, i.e.  $q_1^{(0)} = q_2^{(0)} = 2$ . Solve the (reduced) LMM problem under working independence and obtain  $\lambda_1^{(0)}, \lambda_2^{(0)}$  and  $\mathbf{e}^{(0)}$ . Compute a consistent estimator  $\hat{\mathbf{R}}^{(0)}$  and denote  $\mathbf{Y}_\varsigma = \{\mathbf{I} - \mathbf{G}_2(\boldsymbol{\lambda}^{(0)}, \mathbf{q}^{(0)}, \hat{\mathbf{R}}^{(0)})\}\mathbf{Y}$  as the de-seasonalised signal.

In principle, the smoothing parameter  $\lambda_1^{(0)}$  resulting from the LMM representation of the smoothing splines problem is known to be biased, [cf. Krivobokova and Kauer-[mann, 2007](#)] and hence it must be updated. The same is true for  $\lambda_2^{(0)}$ , however since the structure of the time varying coefficient is likely to be mild, we set  $q_2 = 2$  and use the estimator  $\hat{\lambda}_2 = \lambda_2^{(0)}$ , so this quantities are no further modified. The implementation of the second step retrieves an estimator for the smoothness class of the trend  $\hat{q}_1$  which must be consistent with the former assumption  $q_1^{(0)}$ . If this is not the case, we return to the first step and set  $q_1^{(0)} = q_1^{(0)} + 1$  and continue in this direction until  $\hat{q}_1 = q_1^{(0)}$ . Once this condition is fulfilled the estimator reads  $\hat{\mathbf{f}} = \mathbf{G}(\hat{\lambda}_1^{(\tilde{l})}, \lambda_2^{(0)}, \hat{q}_1, q_2^{(0)}, \hat{\mathbf{R}}^{(\tilde{l})})$  with trend and seasonal components  $\hat{\boldsymbol{\tau}} = \mathbf{G}_1(\hat{\lambda}_1^{(\tilde{l})}, \lambda_2^{(0)}, \hat{q}_1, q_2^{(0)}, \hat{\mathbf{R}}^{(\tilde{l})})$  and  $\hat{\boldsymbol{\varsigma}} = \mathbf{G}_2(\hat{\lambda}_1^{(\tilde{l})}, \lambda_2^{(0)}, \hat{q}_1, q_2^{(0)}, \hat{\mathbf{R}}^{(\tilde{l})})$  respectively.

### 4.1.3. Comparison with Other Methods

Five hundred data points are generated with the following setup for the decomposition scheme (4.6) with  $x_i = t_i / \sum_{i=1}^n t_i$ ,  $n = 500$ . The trend component is generated as  $\tau(x_i) = 6\beta_{a,b}(x_i)/10 + 4\beta_{c,d}(x_i)/10$ , with the beta function  $\beta_{a,b}(x_i) = \Gamma(a+b)\{\Gamma(a)\Gamma(b)\}^{b-1}$ ; the seasonal component follows  $\zeta(x_i) = \alpha_{e,f}(x_i) \cos(\omega x_i)$  where  $\alpha_{e,f}(x_i) = (1/2\pi)\{1 + e^2 + f^2 + 2e(f-1) \cos(\pi(2x_i - 1)) - 2f \cos(2\pi(2x_i - 1))\}^{-1}$ , with a period of 20 observations, i.e.  $\omega = 2\pi(n/20)$ ; and for the remainder component a first order autoregressive process is generated with the autocorrelation coefficient equal to 0.4. Altogether, we consider four scenarios for different values of the parameter vector  $\mathbf{p} = (a, b, c, d, e, f)$ : scenario 1 with  $\mathbf{p} = (3.3, 2, 5, 1, 0, 0)$ , scenario 2 with  $\mathbf{p} = (3.3, 2, 5, 1, 1.9, 0.9)$ , scenario 3 with  $\mathbf{p} = (30, 17, 3, 11, 0, 0)$  and scenario 4 with  $\mathbf{p} = (30, 17, 3, 11, 1.9, 0.9)$ . Furthermore, in all cases the trend, seasonal and remainder components are re-scaled so their variances are 1, 0.5 and 0.1 respectively to assure a reasonable signal-to-noise ratio. Figure (4.1) shows all four scenarios of the simulation study. To portray the capabilities of the method under the most general case, scenario 4 is selected, and model (4.7) is fitted.

**Remark 2.** *If the data were circular, the operation  $\hat{\boldsymbol{\tau}} = \mathbf{S}_1(\hat{\lambda}_1, \hat{q}_1, \hat{\mathbf{R}})\mathbf{Y}$  obtained after fitting the model could be seen as a stationary invariant linear filter with a unique impulse-response function  $\kappa(t)$  centred at the main diagonal of  $\mathbf{S}_1(\hat{\lambda}_1, \hat{q}_1, \hat{\mathbf{R}})$ . The effect of this matrix on input  $\mathbf{Y}$  is usually explored by taking the Fourier transform of the impulse-response function, i.e.  $K(\omega) = \sum_{j=-\infty}^{\infty} \exp\{-i\omega x_j\}\kappa(x_j)$ , also called transfer function. Figure (4.2) shows the impulse-response function, its Fourier transform and the resulting estimated trend for two choices of  $q_1$ . The grey line in panel (b) indicates the frequency of the seasonal component in the generated data, and shows that for  $q_1 = 2$  the impulse-response function of the trend component lie in a range of frequencies characterising the seasonal component, and hence captures both parts indistinguishably. In turn  $q_1 = 4$  reveals to be more adequate, as it can be seen in*

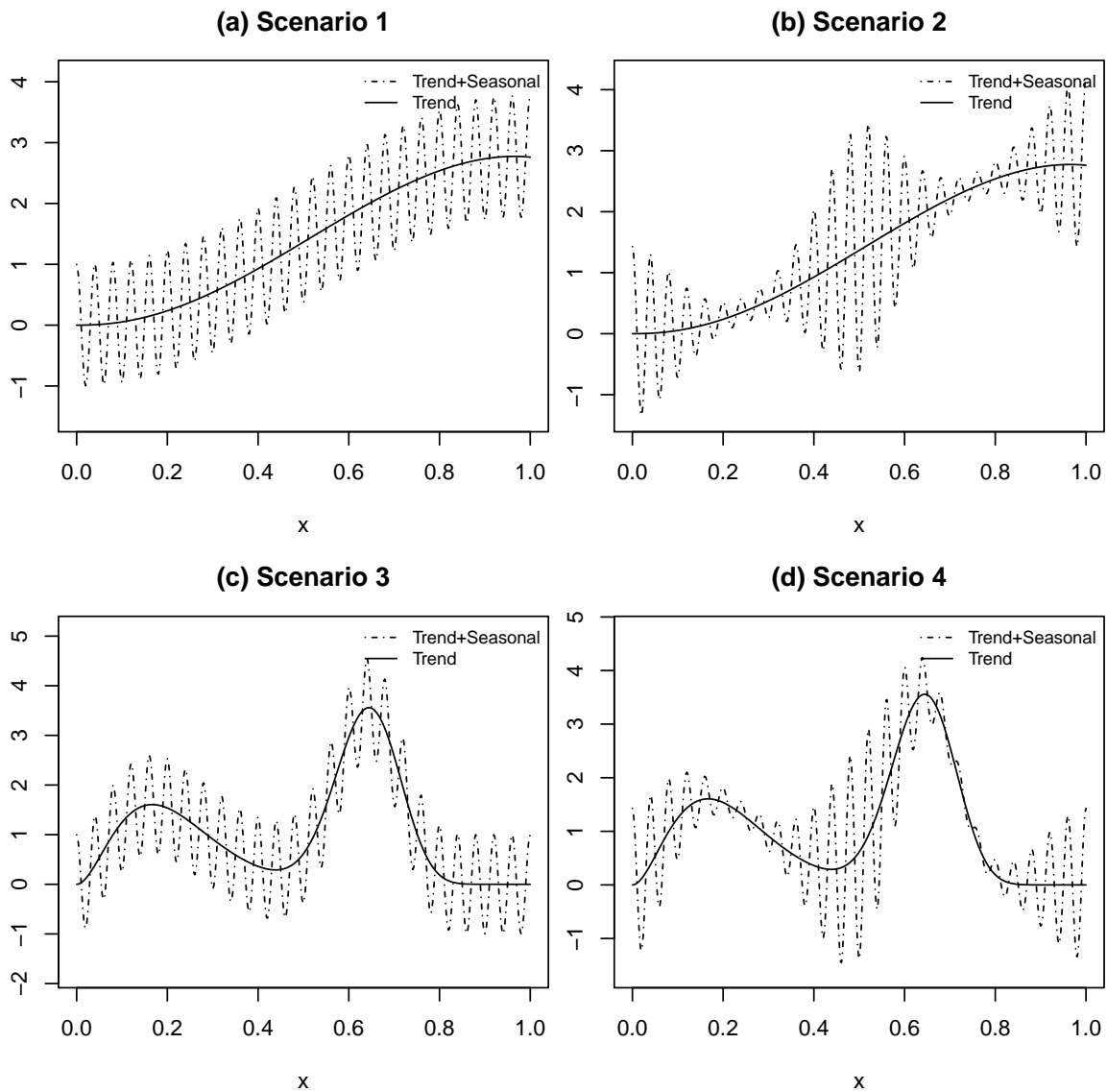


Figure 4.1.: Simulation scenarios in signal decomposition: (a) simple trend and seasonal components, (b) simple trend and varying seasonal component, (c) varying trend and simple seasonal component and (d) varying trend and seasonal components.



panel (c) of the same figure.

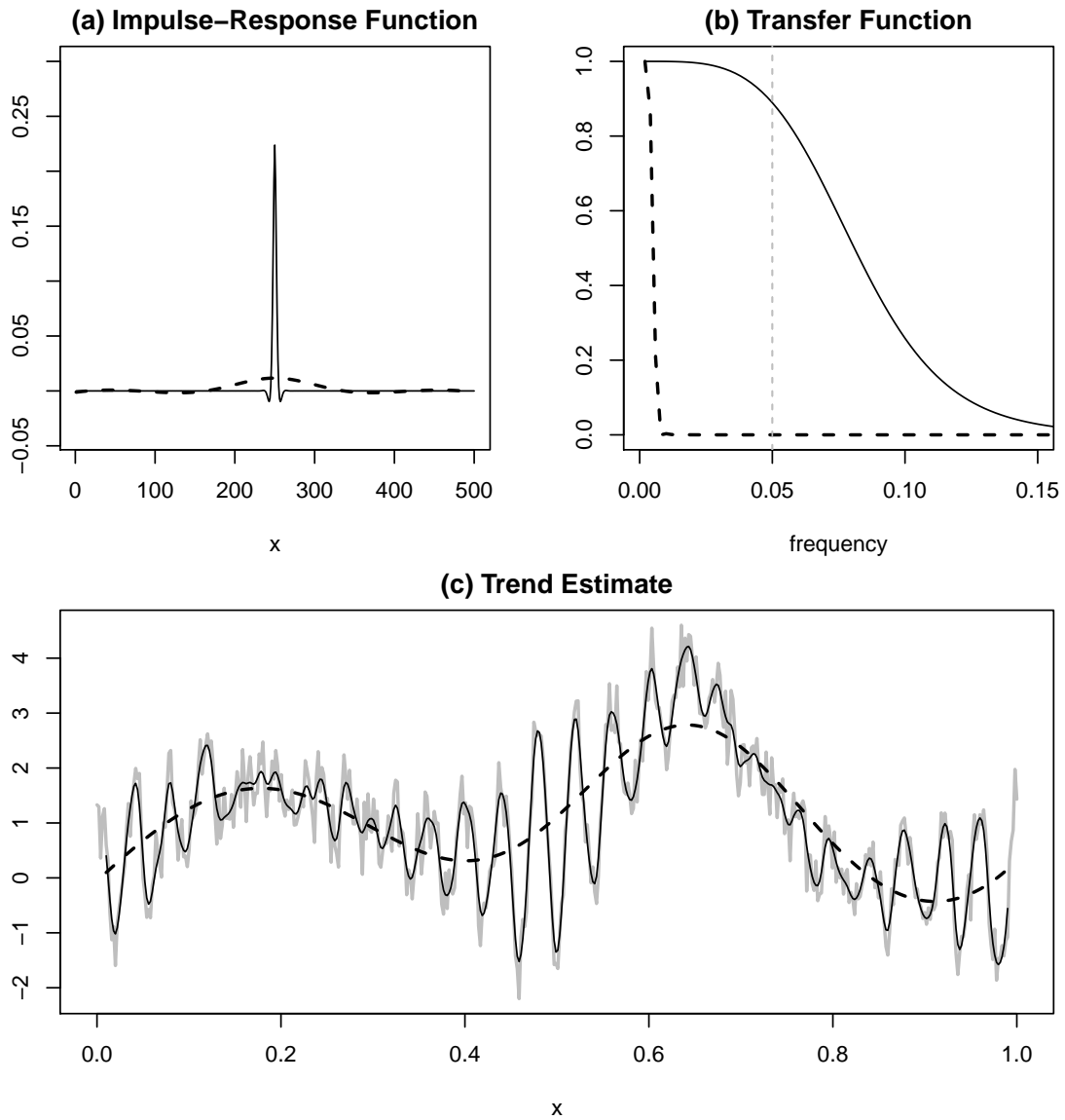


Figure 4.2.: (a) Impulse-Response Function; (b) Transfer Function; (c) Trend estimate. In all plots the continuous and dashed lines represent the cases when  $q_1 = 2$  and  $q_1 = 4$ , respectively. In (c) the simulated data are added as a grey line.

In what follows the performance of the splines based decomposition approach is compared with two alternatives commonly used by practitioners, namely the STL procedure and a state-space model (SSM) of a structured time series characterisation. The section is closed by giving a proper comparison of all three methods by a Monte Carlo

study.

### The STL procedure

The STL procedure for the decomposition of a time series into trend, seasonal and remainder components consists on systematic applications of the LOWESS smoother embedded in two loops: an inner loop that performs a seasonal smoothing updating the seasonal component, followed by a trend smoothing that updates the trend component; and an outer loop executed for robustness. LOWESS was developed in Cleveland [1979] as a robust version of the standard locally weighted polynomial regression estimator.

Consider, for example, the case of monthly data with yearly seasonality. In this case the updates at the  $(j + 1)$ -th pass of the inner loop would be computed in the following way: i) de-trend the series with the  $j$ th update of the trend component by  $y(x_i) - \tau^j(x_i)$ ; ii) build a set of cycle-subseries based on the de-trended series by grouping all Januaries, all Februaries, etc., smooth them with LOWESS, and build a (temporary) seasonal series  $c^{j+1}(x_i)$ ; iii) construct a low-pass filter of the (temporary) seasonal series  $L^{j+1}(x_i)$ ; iv) update the seasonal series de-trending  $c^{j+1}(x_i)$  by  $\zeta^{j+1}(x_i) = c^{j+1}(x_i) - L^{j+1}(x_i)$ ; v) de-seasonalise the original time series with  $y(x_i) - \zeta^{j+1}(x_i)$ ; and vi) update the trend component by smoothing the resulting de-seasonalised series by LOWESS and obtain  $\tau^{j+1}(x_i)$ . The outer loop updates are executed for robustness and operate by modifying the weighting functions in steps ii) and vi) of the inner loop so the effect of aberrant observations in the data (measured by the local magnitude of the remainder) is diminished. de-trended

### State space approach

The state space approach is commonly found in many applied disciplines to model structured time series. In particular, it can be used as a decomposition method by

making intensive use of the Kalman filter. Following Petris et al. [2009], model (4.6) can be represented in state-space form by

$$\begin{aligned} y_t &= \mathbf{Z}\boldsymbol{\varrho}_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \boldsymbol{\varrho}_t &= \mathbf{T}\boldsymbol{\varrho}_{t-1} + \boldsymbol{\zeta}_t \quad \boldsymbol{\zeta}_t \sim \mathcal{N}(0, \mathbf{Q}). \end{aligned}$$

The first expression is called measurement equation where  $\boldsymbol{\varrho} = (\tau_t, \delta_t, \varsigma_t, \varsigma_t^*)^t$  is a generally unobserved vector representing the state of the system at time  $t$ , and  $\mathbf{Z} = (1, 0, 1, 0)$ , so that  $y_t = \tau_t + \varsigma_t + \epsilon_t$ , i.e a decomposition similar to (4.6) holds. The second equation carries out the parametric structure imposed to the problem and it is called transition equation. Here  $\mathbf{T} = \text{blockdiag}\{\mathbf{T}_\tau, \mathbf{T}_\varsigma\}$  is defined, where

$$\mathbf{T}_\tau = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \mathbf{T}_\varsigma = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}$$

with only one frequency  $\omega$ , as the simulation setting requires. For the trend component this implies the use of a random walk model for its slope  $\delta_t$  and it is called the local linear trend model. As for the seasonal part, if one sets  $(\varsigma_0, \varsigma_0^*)^t = (\alpha, \gamma)^t$ , it can be seen that

$$\begin{pmatrix} \varsigma_i \\ \varsigma_i^* \end{pmatrix} \approx \begin{pmatrix} \cos(\omega i) & \sin(\omega i) \\ -\sin(\omega i) & \cos(\omega i) \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$$

holds for  $i = \{1, 2, \dots, n\}$  up to an error term, and hence  $\varsigma_t \approx \alpha \cos(\omega t) + \gamma \sin(\omega t)$  is a time varying seasonal component. For completeness we can define the error vector  $\boldsymbol{\zeta}_t = (\epsilon_t^\tau, \epsilon_t^\delta, \epsilon_t^\varsigma, \epsilon_t^{\varsigma^*})^t$  and  $\mathbf{Q} = \text{diag}\{\sigma_\tau^2, \sigma_\delta^2, \sigma_\varsigma^2, \sigma_{\varsigma^*}^2\}$  as its diagonal covariance matrix. All together the previous model is called Basic Structural Model and it is what is used in the simulation exercise. To setup the model, given the initial distribution  $\boldsymbol{\varrho}_0 = \mathcal{N}(m_0, C_0)$ , which can also be represented as a diffuse prior, parameters  $\sigma_\epsilon^2, \sigma_\tau^2, \sigma_\delta^2$  and  $\sigma_\varsigma^2$  can be estimated by means of maximum likelihood, so that the densities  $\pi(y_t|\boldsymbol{\varrho}_t)$  and  $\pi(\boldsymbol{\varrho}_t|\boldsymbol{\varrho}_{t-1})$  are specified. The smoothing problem of a model with this structure can be solved by computing the conditional distributions  $\pi(\boldsymbol{\varrho}_t|y_1, \dots, y_n)$

for any  $t < n$  starting from the filtering distribution  $\pi(\boldsymbol{\varrho}_n | y_1, \dots, y_n)$  and estimating backwards all state's history [cf. Harvey, 1989, Durbin and Koopman, 2001, Petris et al., 2009].

### Comparison

There are several conceptual similarities between all three discussed methods. For example, both STL and splines based approach could be seen as stationary symmetric linear filters for each of its components modulated by different weighting (or impulse-response) functions. The link between the splines based method and the SSM can be seen if one writes the state space representation of cubic smoothing splines [cf. Zivot and Wang, 2006]. In general, all three methods require the setting of various parameters in order to be implemented. The differences among the procedures are various, but we argue in favor of the proposed splines method mainly in two fronts. Firstly, the splines based approach uses a non-parametric characterisation of the remainder. Secondly, the smoothness class of the trend component is chosen in a data driven manner.

An illustrative comparison of the methods to filter the simulated time series is shown in Figure 4.3. For the splines method the model is set up with  $q_2 =$  and  $\omega = 20$ , as indicated at the beginning of this section; and for the STL procedure  $d_\tau = 1$ ,  $h_\tau = 39$  are considered for the trend component and  $d_\zeta = 1$ ,  $h_\zeta = 7$  for the seasonal component. As it can be seen in panel (a) of Figure 4.3, these methods produce very similar results for the trend component, up to certain wiggleness in the STL case. The performance of STL is not explored for polynomials of degree greater than 1, and the differences between both methods are considered to be small in magnitude. Regarding the SSM, the required parameters to set up the model were obtained via maximum likelihood as in Petris et al. [2009]. The trend estimates obtained using SSM and splines based method are almost undistinguishable.

Even though all procedures allow for the variation of the seasonal part across time,

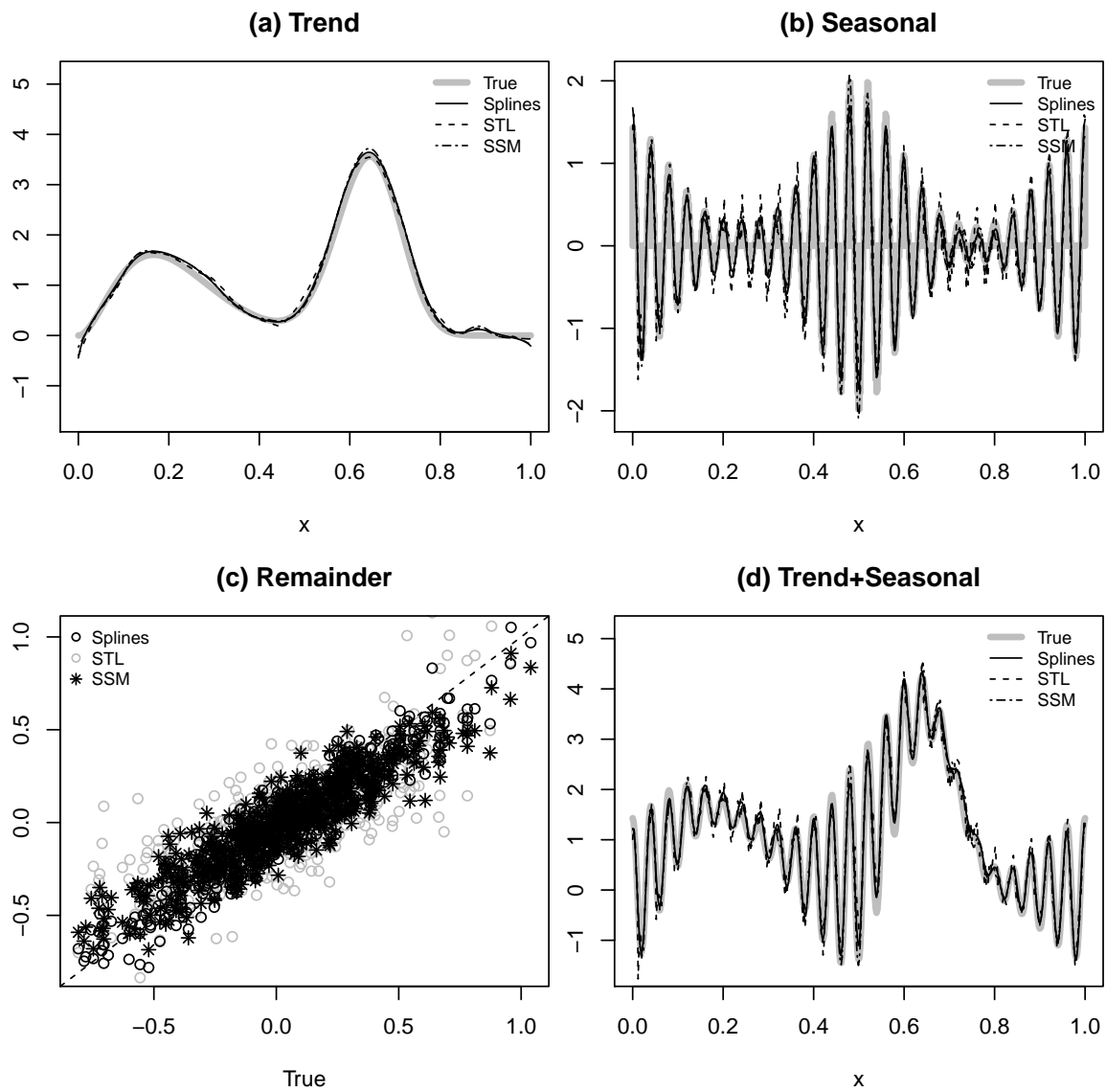


Figure 4.3.: Comparison between STL, SSM and splines method: (a) trend components, (b) seasonal components, (c) remainder components and (d) added trend and seasonal estimation comparison.

the SSM and the splines method seem to behave very similarly, while the STL fit is much more variable. To understand the difference in the shape of the seasonal components it can be noted that while the spline and SSM methods perform smoothing for the data series along the index  $x$ , the STL alternative does it for each cycle-subseries according to the span chosen for the LOWESS window. Clearly, once the smoothed sub-series are re-arranged according to the original time sequence, a rough path can be observed. Furthermore, in this example a span window of size  $h_\zeta = 7$  is chosen (a very small value compared the sample size  $n = 500$ ), which produces a high variance estimation for this component (with small bias). This choice was based on the inspection of the seasonal-diagnostic plot under different  $h_\zeta$  values, as suggested in Cleveland et al. [1990]. In fact, to our knowledge, there is no data driven method for the selection of any of the STL parameters, including the crucial smoothing parameter  $h_\zeta$ . Consequently, the seasonal component could be under-smoothed, as happens in this example. Regarding the comparison with the SSM method for the seasonal part, it appears that splines method performs particularly better than the SSM counterpart when a more structured seasonal component is considered (scenarios 2 and 4). This could be attributed to the fact that in SSM the smoothing parameters for the trend and seasonal component are implicitly set to be equal, and this can generate a bias in one of the components, in this case the seasonal part of the decomposition.

Lastly, regarding the scatterplots of the remainders in panel (c) of Figure (4.3), it is clear that the remainder part from the spline approach (black circles) and the SSM (black stars) follow better the true remainder than the STL (grey circles), which is an obvious result of the better performance of the spline method to fit the seasonal component, and, consequently, allows for a more accurate characterisation of an AR model for the remainder. The last panel of Figure (4.3) compares the estimations of the smooth part of the decomposition  $\tau(x)$  under the three methods showing overall similar results.

This section is closed by reporting the performance of the splines method when com-

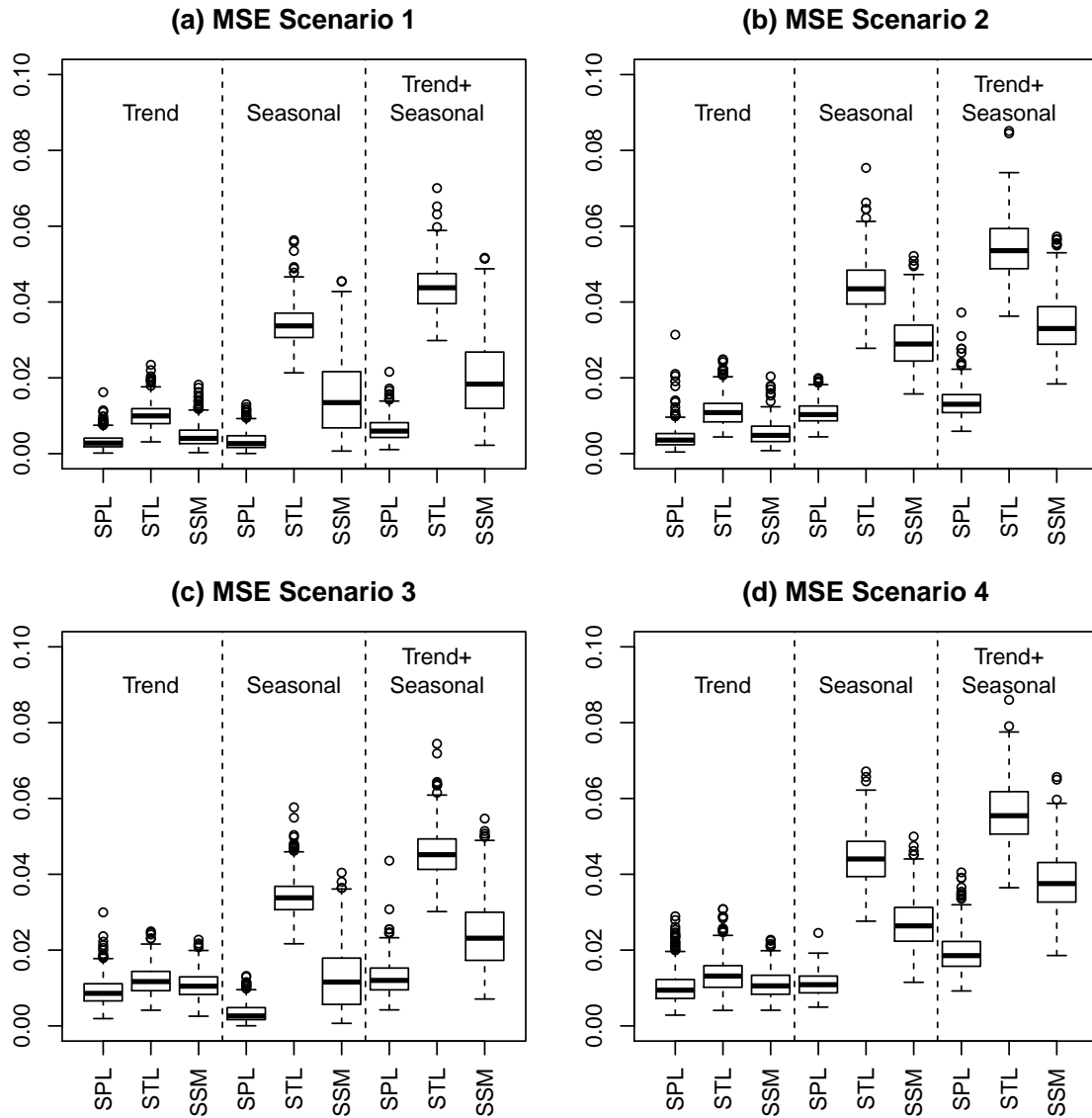


Figure 4.4.: AMSE component-wise comparison between the spline method (SPL in the figure), the STL and the SSM procedures in the Monte Carlo experiment considering four different scenarios for the components.

pared to the STL and SSM alternatives in a Monte Carlo simulation study considering  $M = 1000$  realisations of the remainder component for the four scenarios mentioned at the beginning of this section. The results are presented in terms of the component-wise average mean squared error considering the usual  $\text{AMSE}_j = \sum_{i=1}^M (\hat{g}_j(t_i) - g(t_i))^2 / M$  of each method  $j = \{\text{SPL}, \text{STL}, \text{SSM}\}$  for functions  $g \in \{\tau, \varsigma, f\}$  as defined in (4.6). The results are illustrated in Figure (4.4) and, as expected from the example presented in Figure (4.3), under all scenarios the methods are quite similar for the estimation of the trend with a slightly higher bias for the STL. However, when it comes to comparing the seasonal parts, the splines based procedure reveals its superiority, showing to be less variable and more accurate than its counterparts.

#### 4.1.4. Discussion

We presented a non-parametric new method that contributes to the signal decomposition literature in that it does not only allows for the estimation of the smoothing parameter, but also for the selection of the smoothness class of the trend component and handles correlated errors. Monte Carlo simulations shown the finite sample properties of the components' estimators, revealing that our method is more accurate than its benchmark competitors in the literature.

The relation between cubic smoothing splines and dynamic models is well known in the econometric literature [cf. Harvey, 1989, Zivot and Wang, 2006]. Moreover, this equivalence can be extended to more general types of smoothing splines where the second derivative in the penalty term is replaced by linear differential operators as studied by Kohn and Ansley [1987]. In principle, this suggests that it is possible to represent the empirical Bayes smoothing splines with correlated errors presented in Chapter 3 as a state space model where the smoothness class of the trend component is data driven. This selection approach is not known in the literature and could be of interest for practitioners in the applied sciences.



## 4.2. Functional Principal Component Analysis

Functional data is a term coined by Ramsay and Silverman [1997] to refer to *data curves* (as opposite to *data points*). Many statistical problems arise when dealing with such units of analysis, e.g. functional regression, functional principal component analysis (FPCA), commonly used for prediction, classification, etc. For a mathematical discussion on the subject the reader is referred to Ferraty and Vieu [2006] and Hsing and Eubank [2015]. There are many approaches to perform FPCA [cf. James et al., 2000, Guo et al., 2013], however none of them consider the possibility of seasonality in the observations or a data driven selection of the smoothness class of the underlying functional observations, which is the main contribution of this subsection.

### 4.2.1. Functional Data

Following Ferraty and Vieu [2006], a random variable  $\mathcal{F}$  is called a functional variable if it takes values in an infinite dimensional space (or functional space). Moreover, since  $\mathcal{F}$  (resp.  $f$ ) denotes a random curve (resp. its observation), the following identification is implicitly made  $\mathcal{F} = \{\mathcal{F}(x); x \in [0, 1]\}$  (resp.  $f = \{f(x); x \in [0, 1]\}$ ). The analysis takes place in a functional data set  $f_1, \dots, f_m$  of  $m$  functional variables  $\mathcal{F}_1, \dots, \mathcal{F}_m$  identically distributed as  $\mathcal{F}$ , where an observation  $f$  of  $\mathcal{F}$  is called functional data. Even though the data is functional in nature, we consider the possibility of correlated measurement errors with respect to  $x$  for each realisation  $f_1, \dots, f_m$ , so the characterisation of these curves. e.g. by means of Kernel regression, splines methods, wavelets, etc., should consider it. Moreover, since we would like to take into account the selection of the smoothness class of the data functions, the use of the empirical Bayes smoothing spline method for correlated errors presented in Chapter 3 follows naturally.

### 4.2.2. Principal Component Characterisation

Consider the function  $f_j(x)$  and its de-seasonalised version  $\tilde{f}_j(x) := f_j(x) - \varsigma_j(x)$ , for  $j = 1, 2, \dots, m$ , defined in the compact interval  $[0, 1]$  as functional data. More specifically, assume each  $\tilde{f}_j(x)$  is an independent realisation of the stochastic process  $\{\tilde{f}(x), x \in [0, 1]\}$  with mean  $\mathbb{E}[\tilde{f}(x)] = \tau(x)$  and covariance kernel  $\mathcal{K}(x, z) = \text{cov}\{\tilde{f}(x), \tilde{f}(z)\}$ ,  $x, z \in [0, 1]$ . Mercer's lemma states that if  $\int_0^1 \mathcal{K}(x, x) dx < \infty$  then there exists an orthonormal sequence of eigenfunctions  $\{\zeta_j\}_{j=1}^\infty$  and non-increasing, non-negative sequence of eigenvalues  $\{\kappa_j\}_{j=1}^\infty$  such that for  $(\mathcal{K}\zeta_j)(x) := \int_0^1 \mathcal{K}(x, z)\zeta_j(z) dz = \kappa_j\zeta_j(x)$ , it holds that

$$\mathcal{K}(x, z) = \sum_{j=1}^{\infty} \kappa_j \zeta_j(x) \zeta_j(z), \quad \text{and} \quad \sum_{j=1}^{\infty} \kappa_j = \int_0^1 \mathcal{K}(x, x) dx.$$

Hence we can write the Karhunen-Loève expansion as

$$\tilde{f}(x) = \tau(x) + \sum_{j=1}^{\infty} \sqrt{\kappa_j} \xi_j \zeta_j(x), \quad (4.10)$$

where  $\xi_j := \frac{1}{\sqrt{\kappa_j}} \int \tilde{f}(x) \zeta_j(x) dx$ ,  $\mathbb{E}[\xi_j] = 0$ ,  $\mathbb{E}[\xi_j, \xi_k] = \delta_{jk}$ ,  $j, k \in \mathbb{N}$  and  $\delta_{j,k}$  is the Kronecker delta. From (4.10) the implementation of a reduced rank model for a sample  $j = 1, 2, \dots, m$  as presented in e.g. James et al. [2000] can be written as

$$f_j(x) = \tau(x) + \varsigma_j(x) + \sum_{k=1}^h \psi_k(x) v_{jk} = \tau(x) + \varsigma_j(x) + \mathbf{\Psi}(x)^T \mathbf{v}_j \quad (4.11)$$

where  $h$  is a finite integer usually  $h \ll m$ ,  $\mathbf{\Psi}(x) = \{\psi_1(x), \psi_2(x), \dots, \psi_h(x)\}^T$  contain the  $k$ -th principal component (PC) function  $\psi_k(x)$  and  $\mathbf{v}_j = \{v_{j1}, v_{j2}, \dots, v_{jh}\}^T$  is the vector of PC scores for the  $j$ -th curve. Moreover, we can interpret each  $f_j(x)$  in (4.11) as a curve composed by an overall trend component  $\tau(x)$ , a subject specific seasonality  $\varsigma_j(x)$  and a subject specific deviation  $\eta_j(x) := \sum_{k=1}^h \psi_k(x) v_{jk}$ .

Given observation pairs  $(x_{i,j}, Y_{i,j})$  for  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , one can write a decomposition scheme similar to (4.6) as

$$Y_{i,j} = \tau(x_i) + \varsigma_j(x_i) + \eta_j(x_i) + \epsilon_{i,j}, \quad (4.12)$$

where  $\tau(x)$  is an unknown smooth function and  $\{\epsilon_{i,j}\}_{i=1}^n$  is the homoscedastic error measurement of curve  $j$  with unknown correlation structure, i.e.  $\text{Cor}(\epsilon_{i,j}, \epsilon_{k,j}) = r_{j,i,k}$  for each individual. The seasonal component is modelled as  $\varsigma_j(x) = \alpha_j(x) \cos(x_i\omega) + \beta_j(x) \sin(x_i\omega)$ , where  $\alpha_j(x)$  and  $\beta_j(x)$  are unknown smooth functions that modulate the seasonal pattern over time as presented in Subsection 4.1.2. For simplicity hereafter we consider  $\varsigma_j(x_i) = \alpha_j(x_i) \cos(x_i\omega)$  only. Furthermore, given  $\mathcal{W}^{q_h}[0, 1]$ ,  $h \in \{\tau, \alpha_1, \dots, \alpha_m, \psi_1, \dots, \psi_h\}$  and setting  $q_{h\alpha_1} = \dots = q_{h\alpha_m}$ , and  $q_{h\psi_1} = \dots = q_{h\psi_m}$  we can represent the unknown functions as  $\tau(x) = \mathbf{C}_\tau(x)\boldsymbol{\theta}_\tau$ ,  $\alpha_j(x) = \mathbf{C}_\alpha(x)\boldsymbol{\theta}_{\alpha_j}$  and  $\psi_j(x) = \mathbf{C}_\psi(x)\boldsymbol{\theta}_{\psi_j}$ , where the explicit dependence of the design matrices on  $q_h$  is omitted to ease the notation. For the last case we can also write  $\boldsymbol{\Psi}(x)^T = \mathbf{C}_\psi(x)\boldsymbol{\Theta}_\psi$  with  $\boldsymbol{\Theta}_\psi = \{\boldsymbol{\theta}_{\psi_1}, \dots, \boldsymbol{\theta}_{\psi_h}\}$  an  $n \times h$  matrix such that

$$\int \boldsymbol{\Psi}(x)\boldsymbol{\Psi}(x)^T dx = \boldsymbol{\Theta}_\psi^T \left( \int \mathbf{C}(x)^T \mathbf{C}(x) dx \right) \boldsymbol{\Theta}_\psi = \mathbf{I}_h,$$

holds, and the usual orthogonality requirements for the principal component curves are satisfied. It follows that  $\boldsymbol{\Theta}_\psi^T \boldsymbol{\Theta}_\psi = \mathbf{I}_h$  and thus the estimation problem in (4.11) is reduced to the computation of the spline coefficients  $\boldsymbol{\theta}_\tau$ ,  $\boldsymbol{\theta}_{\alpha_j}$ ,  $\boldsymbol{\theta}_{\psi_k}$  for  $j = 1, \dots, m$ , and  $k = 1, \dots, h$  in

$$\begin{aligned}
 \min_{\boldsymbol{\theta}_\tau, \boldsymbol{\theta}_{\alpha_j}, \boldsymbol{\theta}_{\psi_k}} & \left[ \frac{1}{n} \sum_{j=1}^m (\mathbf{Y}_j - \mathbf{C}_\tau \boldsymbol{\theta}_\tau - \tilde{\mathbf{C}}_\alpha \boldsymbol{\theta}_{\alpha_j} - \mathbf{C}_\psi \boldsymbol{\Theta} \mathbf{v}_j)^T \mathbf{R}^{-1} (\mathbf{Y}_j - \mathbf{C}_\tau \boldsymbol{\theta}_\tau - \tilde{\mathbf{C}}_\alpha \boldsymbol{\theta}_{\alpha_j} - \mathbf{C}_\psi \boldsymbol{\Theta} \mathbf{v}_j) \right. \\
 & + \lambda_\tau \boldsymbol{\theta}_\tau^T \left( \int_0^1 \{ \mathbf{C}_\tau(x)^{(q_\tau)} \}^T \{ \mathbf{C}_\tau(x)^{(q_\tau)} \} dx \right) \boldsymbol{\theta}_\tau \\
 & + \lambda_\alpha \boldsymbol{\theta}_{\alpha_j}^T \left( \int_0^1 \{ \mathbf{C}_\alpha(x)^{(q_\alpha)} \}^T \{ \mathbf{C}_\alpha(x)^{(q_\alpha)} \} dx \right) \boldsymbol{\theta}_{\alpha_j} \\
 & \left. + \lambda_\psi \sum_{k=1}^h \boldsymbol{\theta}_{\psi_k}^T \left( \int_0^1 \{ \mathbf{C}_\psi(x)^{(q_\psi)} \}^T \{ \mathbf{C}_\psi(x)^{(q_\psi)} \} dx \right) \boldsymbol{\theta}_{\psi_k} \right], \tag{4.13}
 \end{aligned}$$

where  $\tilde{\mathbf{C}}_\alpha = \text{diag}\{\cos(x_i \omega)\} \mathbf{C}_\alpha$ ,  $\boldsymbol{\Theta}^T \boldsymbol{\Theta} = \mathbf{I}_h$ ,  $\mathbf{v}_j$  is a random vector such that  $\mathbf{v}_j \sim \mathcal{N}(\mathbf{0}, \sigma_{v_j} \mathbf{I}_m)$  and we denote  $\boldsymbol{\Psi} = \mathbf{C}_\psi \boldsymbol{\Theta}$ . Given  $h$  and  $\omega$  it remains to estimate  $\mathbf{v}_j$ ,  $\lambda_\tau$ ,  $\lambda_\alpha$ ,  $\lambda_\psi$ ,  $\mathbf{R}_1, \dots, \mathbf{R}_d$ ,  $q_\tau$ ,  $q_\alpha$ ,  $q_\psi$  for which we turn to empirical Bayesian smoothing splines as presented in Chapter 3.

### 4.2.3. Statistical Algorithm

Since the solution of (4.13) is constraint to the knowledge of  $\mathbf{v}$ , in the next section we present an algorithm that computes  $\mathbf{v}$  iteratively until convergence is achieved on the  $h$  principal component functions  $\psi_k(x)$ . Our procedure is based on the work by Guo et al. [2013] with the difference that we allow for correlation in the errors and make use of the LMM representation of smoothing splines.

#### Initial values

1. Solve (4.13) for  $\mathbf{R}_1 = \dots = \mathbf{R}_m = \mathbf{I}_n$ ,  $\boldsymbol{\theta}_{\alpha_1} = \dots = \boldsymbol{\theta}_{\alpha_d} = \mathbf{0}$ ,  $\boldsymbol{\theta}_{\psi_1} = \dots = \boldsymbol{\theta}_{\psi_h} = \mathbf{0}$  to obtain  $\lambda_\tau$  and compute

$$\hat{\boldsymbol{\theta}}_\tau^{(0)} = (\mathbf{C}_\tau^T \mathbf{C}_\tau + \lambda_\tau n \mathbf{D}_\tau)^{-1} \mathbf{C}_\tau^T \mathbf{Y}$$

with  $\mathcal{C}_\tau = \mathbf{C}_\tau \otimes \mathbf{I}_m$  and  $\mathcal{D}_\tau = \mathbf{D}_\tau \otimes \mathbf{I}_m$ .

- Write the residuals for each curve as  $\mathbf{r}_j = \mathbf{Y}_j - \mathbf{C}_\tau \hat{\boldsymbol{\theta}}_\tau^{(0)}$  and fit the linear least squares model

$$\mathbf{r}_j = \mathbf{C}_\psi \boldsymbol{\Gamma}_j + \boldsymbol{\epsilon}_j,$$

to obtain  $\hat{\boldsymbol{\Gamma}}^{(0)} = \{\hat{\boldsymbol{\Gamma}}_1^{(0)}, \dots, \hat{\boldsymbol{\Gamma}}_N^{(0)}\}^T$ .

- Calculate the singular value decomposition

$$\hat{\boldsymbol{\Gamma}}^{(0)} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T,$$

and set  $\hat{\boldsymbol{\theta}}_{\psi_k}^{(0)} = \mathbf{V}_k \boldsymbol{\Sigma}_k$ , and hence  $\hat{\boldsymbol{\Psi}}^{(0)} = \mathbf{C}_\psi \hat{\boldsymbol{\Theta}}_\psi^{(0)}$ .

### Iterative procedure

Given  $h$ , initial value  $\hat{\boldsymbol{\Psi}}^{(0)}$  and parametrisation  $\boldsymbol{\Phi}$ , the main idea consists in updating the computation of  $\hat{\mathbf{v}}^{(l)}$  in (4.13) for the  $l$ -th iteration and the corresponding  $\hat{\boldsymbol{\Psi}}^{(l)}$  matrix until convergence is achieved. More precisely

- Solve (4.13) given  $\hat{\boldsymbol{\Theta}}_\psi^{(l-1)}$  and update  $\hat{\mathbf{v}}^{(l)}$ ,  $\hat{\lambda}_\tau^{(l)}$ ,  $\hat{\lambda}_\alpha^{(l)}$  and  $\hat{\mathbf{R}}^{(l)} = \text{blockdiag}\{\hat{\mathbf{R}}_1^{(l)}, \dots, \hat{\mathbf{R}}_m^{(l)}\}$  to compute

$$\begin{aligned} \hat{\boldsymbol{\theta}}_\tau^{(l)} &= \left\{ \mathbf{C}_\tau^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \mathbf{C}_\tau + \hat{\lambda}_\tau^{(l)} n \mathbf{D}_\tau \right\}^{-1} \mathbf{C}_\tau^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \mathbf{Y}, \\ \hat{\boldsymbol{\theta}}_\alpha^{(l)} &= \left\{ \mathbf{C}_\alpha^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \mathbf{C}_\alpha + \hat{\lambda}_\alpha^{(l)} n \mathbf{D}_\alpha \right\}^{-1} \mathbf{C}_\alpha^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \mathbf{Y}, \end{aligned}$$

with  $\mathcal{C}_\alpha = \mathbf{C}_\alpha \otimes \mathbf{I}_m$  and  $\mathcal{D}_\alpha = \mathbf{D}_\alpha \otimes \mathbf{I}_m$ ; and  $\mathcal{C}_\tau$  and  $\mathcal{D}_\tau$  as defined in the initialisation step.

- Write the residuals for as  $\mathbf{r}^{(l)} = \mathbf{Y} - \mathbf{C}_\tau \hat{\boldsymbol{\theta}}_\tau^{(l)} - \mathbf{C}_\alpha \hat{\boldsymbol{\theta}}_\alpha^{(l)}$  and fit the penalised least squares model

$$\mathbf{r}^{(l)} = \mathbf{C}_\psi \hat{\boldsymbol{\Theta}}_\psi^{(l-1)} \hat{\mathbf{v}}^{(l)} + \boldsymbol{\epsilon},$$

to update  $\hat{\Theta}_\psi^{(l)}$  by

$$\hat{\theta}_{\psi_k}^{(l)} = \left\{ \sum_{j=1}^N \left( \hat{v}_{jk}^{(l)} \right)^2 \mathbf{C}_\psi^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \mathbf{C}_\psi + \hat{\lambda}_\psi^{(l)} \mathbf{D}_\psi \right\}^{-1} \left\{ \sum_{j=1}^N \left( \hat{v}_{jk}^{(l)} \right) \mathbf{C}_\psi^T \left( \hat{\mathbf{R}}^{(l)} \right)^{-1} \left( \mathbf{Y}_j - \mathbf{C}_\tau \hat{\theta}_\tau^{(l)} - \mathbf{C}_\alpha \hat{\theta}_\alpha^{(l)} - \mathbf{C}_\psi \hat{\mathbf{Q}}_{jk}^{(l)} \right) \right\},$$

for

$$\hat{\mathbf{Q}}_{jk}^{(l)} = \sum_{l \neq k} \hat{\theta}_{\psi_l}^{(l)} \hat{v}_{jl}^{(l)}, \quad j = 1, \dots, N.$$

3. Construct  $\hat{\Theta}_\psi^{(l)}$  and use the QR decomposition to orthonormalise its columns.

With the new estimation of  $\hat{\Theta}_\psi^{(l)}$  go back to step one until convergence is achieved.

The previous algorithm takes place making the implicit assumption that  $f_j(x) \in \mathcal{W}^{q_j} [0, 1]$ , where the smoothness classes of the overall trend, seasonal component and subject specific variations take place and need to be estimated. Here we proceed as in Subsection 4.1.2 and take cubic smoothing splines for both seasonal and subject specific deviation components, and hence the problem reduces to find the smoothness class of the overall trend  $q_\tau$ .

The estimation of  $q_\tau$  is performed in an outer loop, after executing the previous algorithm (inner loop) for  $q_\tau \in \mathcal{Q}$ , and  $\mathcal{Q} = \{2, \dots, \lfloor \log(n) \rfloor\}$ . So that given certain  $q_\tau$ , we compute  $\mathbf{Y}_j^* = \hat{\mathbf{R}}^{-1/2} (\mathbf{Y}_j - \tilde{\mathbf{C}}_\alpha \hat{\theta}_{\alpha_j} - \mathbf{C}_\psi \hat{\Theta} v_j)$  and collect the  $\hat{q}_{\tau,j}$ 's that solve  $T_q(\lambda_j, q_\tau, \mathbf{I}; \mathbf{Y}_j^*) = 0$  for each  $j$  as presented in (3.18). We then assign a unique  $\hat{q}_\tau$  computed as  $\hat{q}_\tau = \text{Mode}\{\hat{q}_{\tau,j}\}$  and take the estimations that resulted from  $\hat{q}_\tau$  in the inner loop.

#### 4.2.4. Discussion

We presented a semi-parametric new method that contributes to the functional data literature in that we consider the possibility of correlated measurement errors with

respect to  $x$  for each realisation  $f_1, \dots, f_m$ , so the characterisation of these curves by means of smoothing splines accounts for it. Moreover, we also took into consideration the selection of the smoothness class of the data functions, and hence the link between FPCA and the empirical Bayes smoothing spline method presented in Chapter 3 is established naturally. We also provided an statistical algorithm for the estimation procedure.

The semi-parametric FPCA method presented is rather natural for functional data, where the dependence structure of each curve's measurement error is, in principle, mild and an ARMA(p,q) can be used to capture its features. It should also be noted that this modelling is different from the so-called *dependent* functional data analysis, where the dependence between functional observations  $f_1, \dots, f_m$  is taken into account. In our case we assumed that such observations are independent realisations of the same stochastic process. Extensions in this direction could be interesting for future research.

## 5. Applications

In this chapter we apply the methods described in sections 4.1 and 4.2 to agricultural economics and biophysics. In the first application we make use of signal decomposition to study price transmission for agricultural commodities, where the use of smoothing methods is motivated from the pork cycle conjecture [cf. [Ezekiel, 1938](#)]. In the second application we explore the multivariate co-movement of adult human mesenchymal cell data via functional principal component analysis. Specifically, stem cell's image area are modelled as independent realisations of certain functional stochastic process to study their time dynamics in a climatic chamber.

### 5.1. Nonparametric Price Transmission

Piecewise linear or regime-dependent cointegration methods are currently popular in the price transmission literature. See [Ihle \[2009\]](#) for a detailed revision on the topic. However, the methods have been criticised for making unrealistic assumptions about the functional form and the nature of the transition process between regimes. The often used threshold vector error correction model [cf. [Goodwin and Piggott, 2001](#), [Greb et al., 2013](#)] assumes that the price transition process between two prices changes abruptly the moment the difference between these prices crosses a certain threshold value. While this is clearly more flexible than assuming that a single price transmission process holds for all values of the difference between prices, it does not allow the price transmission process to change gradually. Other regime dependent



cointegration methods maintain a number of assumptions regarding functional form and the parametrisation of the transition between regimes. In response, some authors have proposed using non-parametric methods to estimate price transmission. Namely [Serra et al. \[2006a\]](#), and [Serra et al. \[2006b\]](#) made an important contribution to the price transmission literature by introducing the use of non-parametric methods in cointegration analysis. Our contributions are in this direction.

First, the non-parametric approach employed in [Serra et al. \[2006a\]](#), and [Serra et al. \[2006b\]](#) maintain the assumption that the long run equilibrium relationship between the prices being studied is constant and linear. There are many settings in which this assumption might be too restrictive. For example, policy changes such as the introduction of a tariff will affect the spatial equilibrium condition for prices in two locations and, thus, alter the long run relationship between these prices. In the vertical price transmission setting, changes in processing technology and shifts in market power can lead to changes in the long run equilibrium between prices at different levels of a food chain as in [Lloyd et al. \[2006\]](#). Finally, in both vertical and spatial settings the long run equilibrium might display a seasonal component, for example if product qualities or the costs of transportation between markets display seasonal fluctuations [cf. [Holst and von Cramon-Taubadel, 2011](#)]. Ideally one would account for such factors explicitly by including appropriate variables in the estimation of the long run relationship. However, in many cases the variables of interest are difficult or impossible to observe, especially at high frequency (weekly or daily) at which much commodity price transmission analysis is carried out. Hence, in the following study we propose a non-parametric vector error correction model (NPVECM) that allows for time-varying long run equilibrium relationships, that can capture drifting and seasonal components.

Second, while [Serra et al. \[2006a\]](#) and [Serra et al. \[2006b\]](#) employ local polynomial techniques, in recent years non-parametric estimation has been enriched by the refinement of spline methods [cf. [Kauermann et al., 2011](#)]. In particular, recent advances

by [Wiesenfarth et al. \[2012\]](#) allow for estimation and simultaneous direct inference without the need for re-sampling methods, which can reduce computational costs significantly. The NPVECM that we propose employs these techniques.

### 5.1.1. Model

Here we consider a non-parametric vector error correction model with time-varying cointegration. The method we propose can be seen as an extension of the classical method given by [Engle and Granger \[1987\]](#) that can be applied when the unconstrained expectation of the error correction term exhibits i) a non-linear deterministic variation over time and/or ii) a deterministic seasonal pattern. For estimation we use a two-step method. First we estimate the error correction term allowing for a time-varying cointegrating relationship following the results given in Chapter 4 regarding signal decomposition; and second we estimate the error correction process semi-parametrically using the penalised spline methods developed in [Wiesenfarth et al. \[2012\]](#).

Consider the following extension of the system given in [Engle and Granger \[1987\]](#):

$$\begin{aligned} x_{1,t} + \bar{\gamma}x_{2,t} &= e_t + g(t) \\ x_{1,t} + \bar{\beta}x_{2,t} &= W_t + h(t), \end{aligned} \tag{5.1}$$

measured at times  $t = 1, \dots, n$  for functions  $g(t)$  and  $h(t)$ . Assume  $e_t = \rho(e_{t-1})e_{t-1} + \mu_t$  and  $W_t = W_{t-1} + \epsilon_t$ , with  $\mu_t \sim \mathcal{N}(0, \sigma_\mu^2)$ ,  $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$  and  $\text{Cov}(\mu_t, \epsilon_t) = 0$ . Hence  $e_t$  is characterised as a smooth transmission auto regressive (STAR) model [cf. [Teräsvirta and Eliasson, 2001](#), [van Dijk et al., 2002](#)], where  $|\rho(e_{t-1})| \leq 1$  (in fact for the non-reaction regime one would expect  $\rho(e_{t-1}) = 1$  so that  $e_t$  behaves like a random walk). More specifically  $\vartheta(\cdot)$  is a transition function, e.g. logistic, exponential. Both  $\bar{\gamma}$  and  $\bar{\beta}$  are fixed in time. Under this setting  $g(t)$  can be seen as a time varying intercept that modulates the long term relationship between  $x_{1,t}$  and  $x_{2,t}$ . As defined here,  $e_t$  and  $W_t$  contain no seasonality. If the processes defined in (5.1) contains deterministic

## 5. Applications

---

seasonal patterns, e.g.  $\varsigma_i(t)$  in  $X_{i,t} = x_{i,t} + \varsigma_i(t)$  for each  $i = \{1, 2\}$ , these can be removed independently. In fact, equation (5.1) can be re-written as:

$$\begin{aligned} X_{1,t} + \bar{\gamma}X_{2,t} &= e_t + \{\varsigma_1(t) + \bar{\gamma}\varsigma_2(t)\} + g(t), \\ X_{1,t} + \bar{\beta}X_{2,t} &= W_t + \{\varsigma_1(t) + \bar{\beta}\vartheta_2(t)\} + h(t), \end{aligned} \quad (5.2)$$

where the first equation is of special interest since it accounts for common features mentioned in the previous section, i.e. a remainder with a drifting (possibly non-linear) component and a seasonal pattern in the error correction term [cf. [Holst and von Cramon-Taubadel, 2011](#)]. Solving (5.2) for  $x_{1,t}$  and  $x_{2,t}$  leads to:

$$\begin{aligned} x_{1,t} &= -\left(\frac{\bar{\beta}}{\bar{\gamma} - \bar{\beta}}\right) e_t + \left(\frac{\bar{\gamma}}{\bar{\gamma} - \bar{\beta}}\right) W_t - \left(\frac{\bar{\beta}g(t) - \bar{\gamma}h(t)}{\bar{\gamma} - \bar{\beta}}\right), \\ x_{2,t} &= +\left(\frac{1}{\bar{\gamma} - \bar{\beta}}\right) e_t - \left(\frac{1}{\bar{\gamma} - \bar{\beta}}\right) W_t + \left(\frac{g(t) - h(t)}{\bar{\gamma} - \bar{\beta}}\right). \end{aligned} \quad (5.3)$$

Since the seasonalities cancel out in the derivation of (5.3), subsequent analysis can focus on the de-seasonalised series. This is not a surprising feature, as a number of authors have suggested that whenever deterministic seasonality is present, it should be removed [cf. [Box and Jenkins, 1971](#), [Engle and Granger, 1987](#), [Hylleberg et al., 1990](#)]. From (5.3) we obtain the error correction representation by taking differences:

$$\begin{aligned} \Delta x_{1,t} &= \rho_1(e_{t-1}) + \tau_1(t) + \varepsilon_{1,t}, \\ \Delta x_{2,t} &= \rho_2(e_{t-1}) + \tau_2(t) + \varepsilon_{2,t}, \end{aligned} \quad (5.4)$$

where  $\rho_i(e_{t-1}) = \rho_i(e_{t-1})e_{t-1}$ ,  $\tau_1(t) = -(\bar{\beta}\Delta g(t) - \bar{\gamma}\Delta h(t))/(\bar{\gamma} - \bar{\beta})$ ,  $\tau_2(t) = (\Delta g(t) - \Delta h(t))/(\bar{\gamma} - \bar{\beta})$ ,  $\varepsilon_{1,t} = (\bar{\gamma}v_t - \bar{\beta}\mu_t)/(\bar{\gamma} - \bar{\beta})$  and  $\varepsilon_{2,t} = (\mu_t - v_t)/(\bar{\gamma} - \bar{\beta})$  so that in general  $\text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) \neq 0$  unless  $-\bar{\gamma}/\bar{\beta} = \sigma_\mu^2/\sigma_v^2$  holds. Moreover, the adjustment speeds can be computed as the first derivatives of  $\rho_i(e_{t-1})$ ,  $i = \{1, 2\}$ .

### 5.1.2. Estimation

The estimation of the NPVECM can be conducted in two steps. In the first step the error correction term is computed non-parametrically from (5.1), and in the second step the adjustment speeds are computed semi-parametrically from (5.4).

The first step is equivalent to the estimation of  $\bar{\gamma}$  and  $g(t)$ . A direct semi-parametric estimation to compute both simultaneously would be ideal. However, methods that use non-intrusive / data-driven smoothing parameter selection suffer from an identification problem in this setting: the magnitude of  $\bar{\gamma}$  tends to be underestimated since the inherent flexibility of  $g(t)$  allows it to explain most of the variance in  $x_{1,t}$ . Estimating  $\bar{\gamma}$  first by least squares and then estimating  $g(t)$  non-parametrically is also not possible as omitting  $g(t)$  in the initial least squares estimation of  $\bar{\gamma}$  would induce bias. Instead we follow Gardner [1975], and set  $\bar{\gamma} = 5/2$  to estimate function  $g(t)$  using the signal decomposition techniques presented in Chapter 4.

The second step can be performed by a direct semi-parametric extension of (5.4) following Wiesenfarth et al. [2012]. Specifically, we estimate:  $\Delta x_{i,t} = \rho_i(e_{t-1}) + \tau_i(t) + \varepsilon_{i,t}$  with  $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^2)$  and assuming  $\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t}) = \delta_{i,j}$  for  $\delta_{i,j}$  the kronecker delta and  $i = \{1, 2\}$ . From this model the adjustment speeds can be directly obtained as  $\hat{\rho}_i(e_{t-1}) = \partial \hat{\rho}_i(e_{t-1}) / \partial e_{t-1}$ , so that  $\beta(e_{t-1})$  can be recovered as  $-\hat{\rho}_1(e_{t-1}) / \hat{\rho}_2(e_{t-1})$ .

### 5.1.3. Application

Here we study the dynamics of vertical transmission between piglet and pork prices in Germany. The application uses 18 years of weekly prices from 1996 to 2013. The piglet price is an average of regional prices in Euro for 25 kg piglets in Lower Saxony, and the slaughter pig price is an average price for the main classes of slaughter pig in Euro/kg also in Lower Saxony, which is the largest pork producing region in Germany as reported in the *Land und Forstwirtschaftliche Zeitung*. The raw data and the corresponding de-seasonalised series are depicted in Figure (5.1).

Slaughter pig producers in Germany generally purchase piglets from specialised farrowing operations located in Germany but increasingly also in Denmark and the Netherlands. After feed, which accounts for roughly one-half of the cost of slaughter pig production, piglets are the second largest cost component with a share of roughly 40%. While slaughter pig producers are largely price takers on the markets for inputs such as feed, energy and labour, piglets are a specialised input for which there is essentially no alternative demand. Hence, piglet prices depend heavily on the expected profitability of slaughter pig production and, thus, slaughter pig prices. This, in turn, depends on conditions on pork markets, which are subject to well-known cyclical fluctuations, see [Berg and Huffaker \[2014\]](#) and the original conjecture provided by [Ezekiel \[1938\]](#).

[Holst and von Cramon-Taubadel \[2011\]](#) hypothesise that when pork prices are in the declining phase of a cycle, slaughter pig producers will be less interested in purchasing piglets, leading, *ceteris paribus*, to oversupply. On the resulting buyers' market for piglets, slaughter pig producers will be able to pass negative price shocks on to piglet producers more rapidly than positive shocks; i.e. price transmission between slaughter pig and piglet prices will be characterised by negative asymmetry as in [Meyer and von Cramon-Taubadel \[2004\]](#). The opposite will hold when pork prices are in the increasing phase of a cycle: slaughter pig producers will be eager to expand production, and the resulting sellers' market for piglets will be characterised by positive asymmetry – sometimes referred to as the “rockets and feathers” phenomenon, see [Bacon \[1991\]](#) – whereby piglet prices react quickly when slaughter pig prices increase, but slowly when they fall. [Holst and von Cramon-Taubadel \[2011\]](#) test this hypothesis by means of a piecewise linear error correction model. They first use the [Hodrick and Prescott \[1997\]](#) filter to extract a smooth cyclical component from the series of slaughter pig prices and divide their dataset into increasing and decreasing phases of pork prices. They then estimate a separate asymmetric error correction model for each of these phases and find support for the hypothesis of alternating regimes of positive and negative asymmetric price transmission.

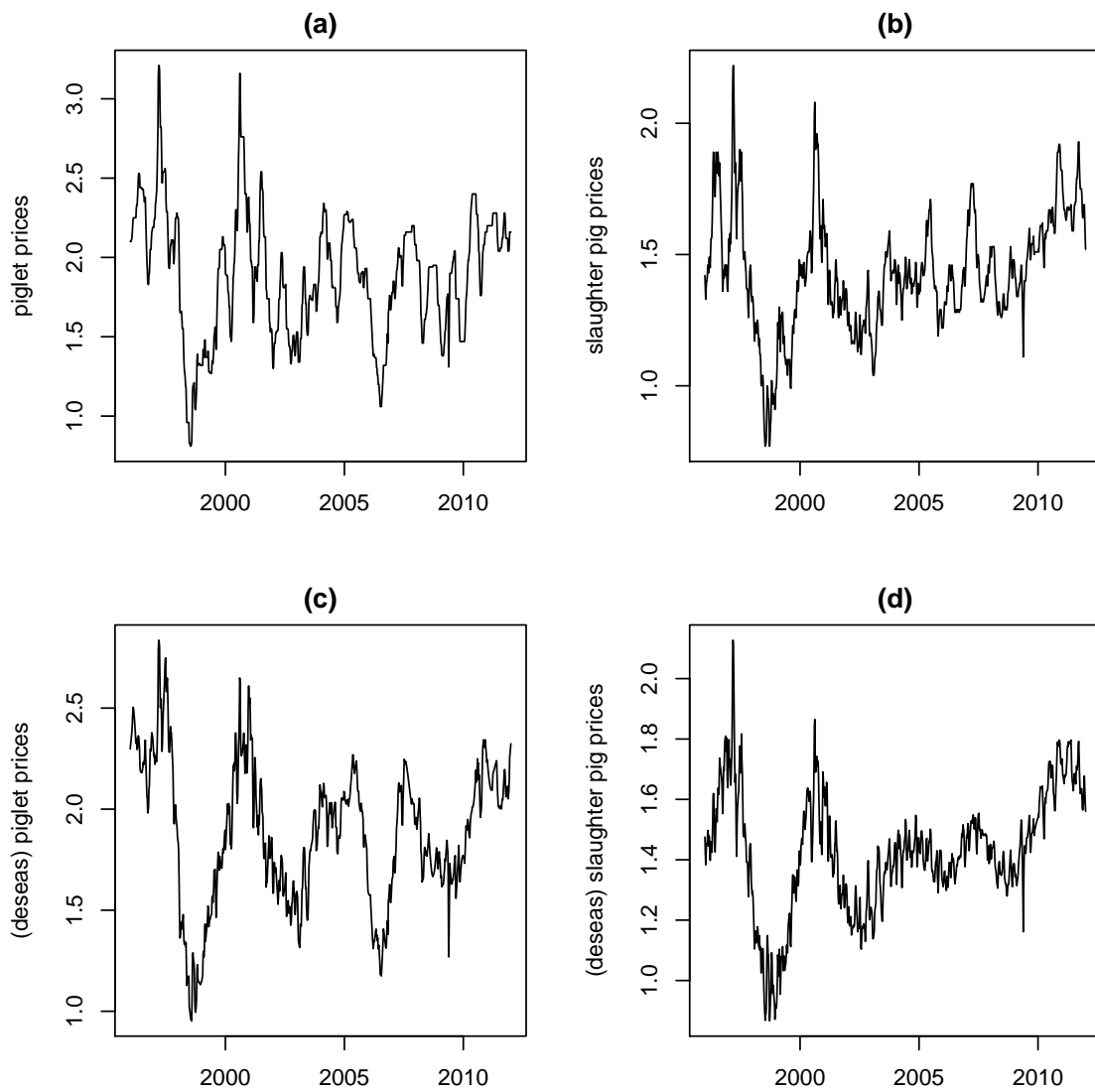


Figure 5.1.: Piglet and slaughter pig prices in Germany (Euro/piglet and Euro/kg, 1996-2013). Panels (a) and (b) show the raw data and panels (c) and (d) the corresponding de-seasonalised series.

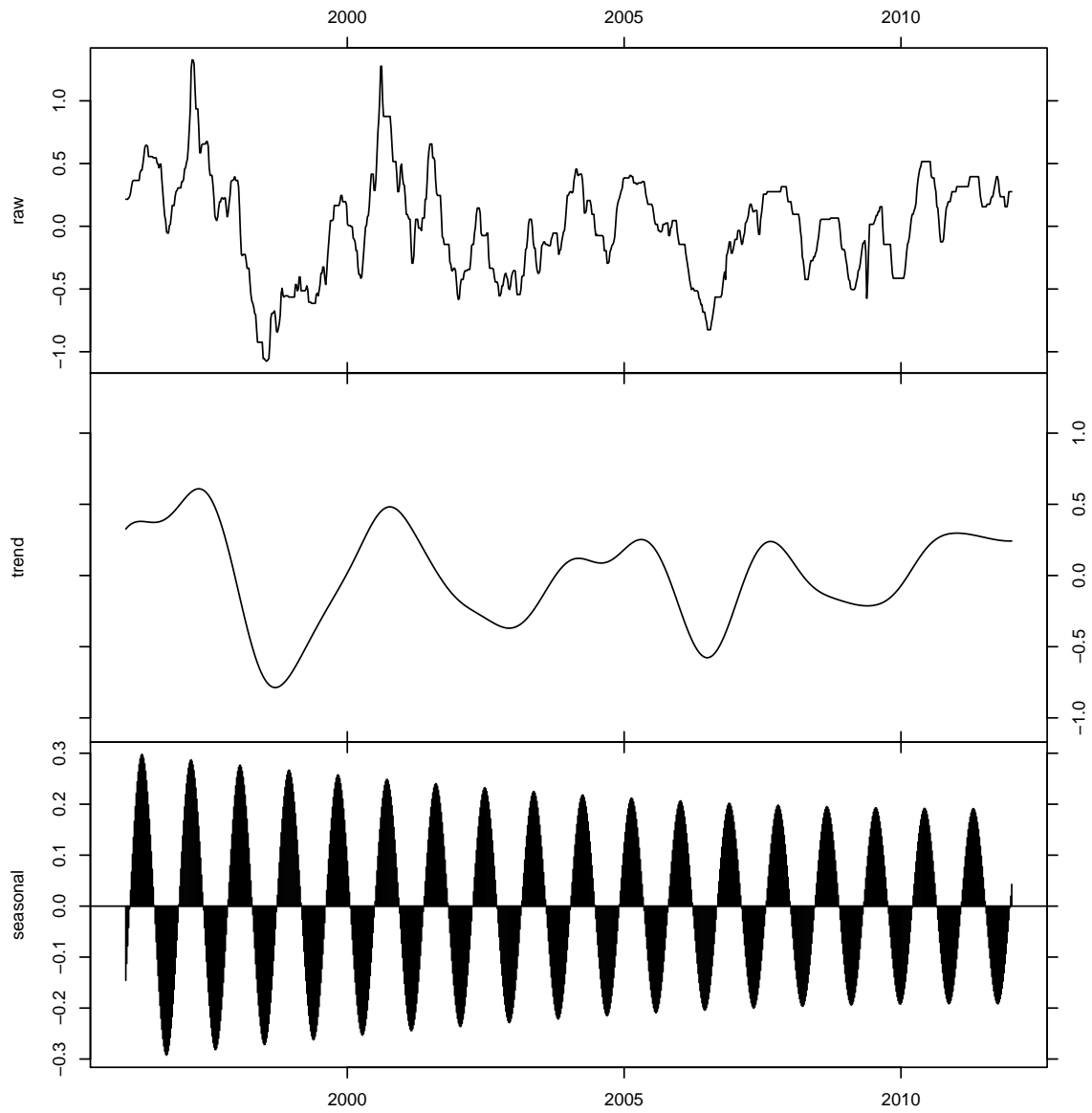


Figure 5.2.: Detailed decomposition of piglet prices.

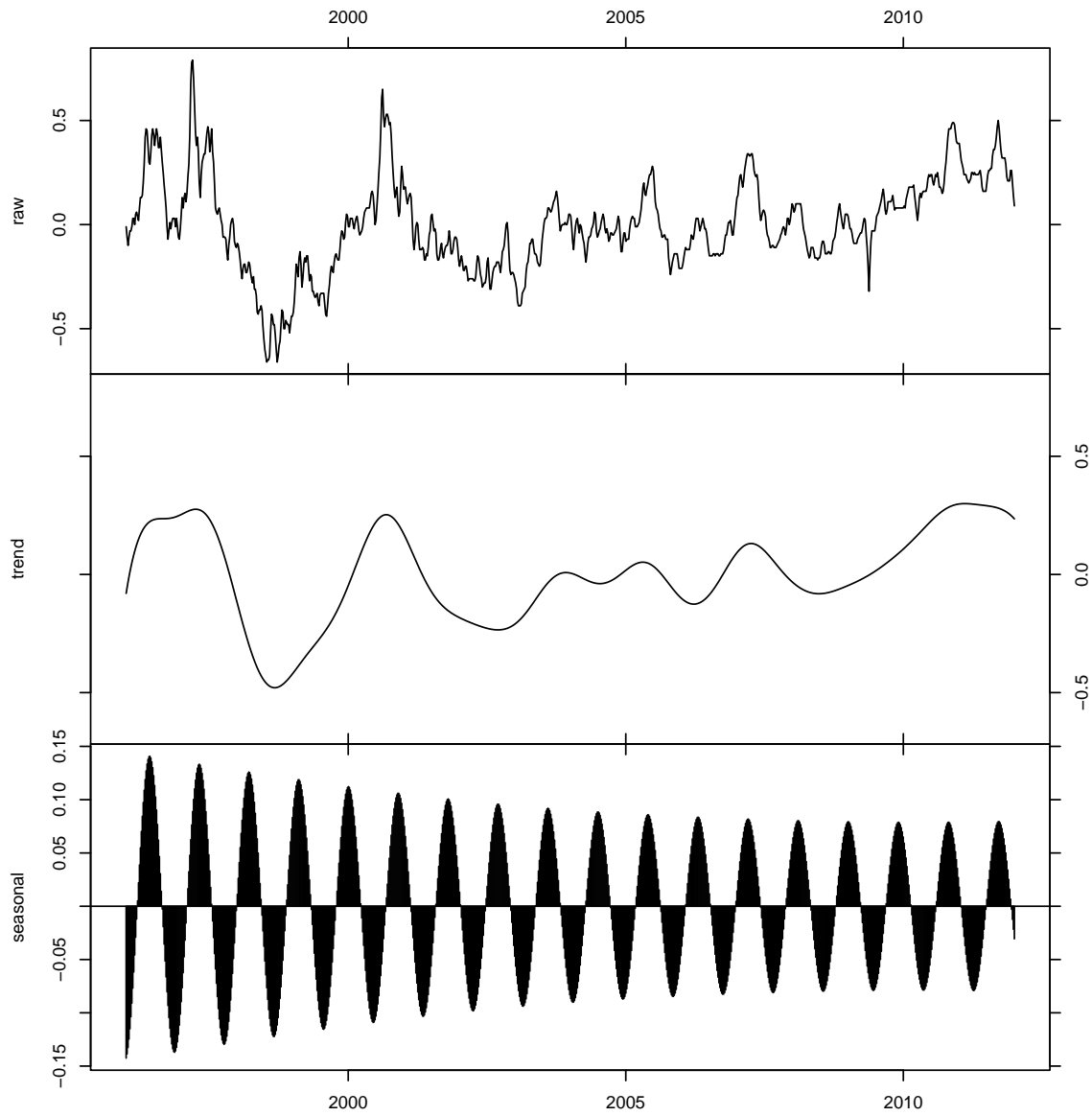


Figure 5.3.: Detailed decomposition of slaughter pig prices.



The upper panels of figure (5.1) display the series for piglets and slaughter pigs respectively. It shows that the piglet and the slaughter pig prices both display seasonal fluctuations, and hence the error correction term could inherit this feature. We extract the deterministic fluctuations in each series following Chapter 4 and illustrate the results for the piglet and slaughter pig prices in Figures (5.2) and (5.3) respectively. The adjustment responses estimated semi-parametrically using these series are presented in Figure (5.4), together with the corresponding 95% confidence intervals for two scenarios on the error correction term (ect).

Figure (5.4) shows that slaughter pig prices do not react significantly to any deviations from the long run equilibrium, i.e the estimated adjustment response is not significant over the entire range of deviations (panels b and d). Hence, as expected the burden of adjustment to disequilibrium is carried out by the piglet prices, which react to positive and negative deviations. These results suggest that piglet prices react more strongly when they are too high vis-à-vis slaughter pig prices, i.e. if the slaughter pig producers' margins are squeezed. When piglet prices are too low, which increases slaughter pig producers' margins, the reaction of piglet prices is of smaller magnitude.

### 5.1.4. Conclusions

The current price transmission literature largely draws on piecewise-linear or regime-dependent VECM specifications such as the threshold VECM and the Markov-switching VECM. While these specifications are more flexible than a linear VECM, they are all based on the assumption that the linear VECM is a valid parametrisation of the price transmission process within each individual regime is valid, and each is based on some parametric assumption about the process that governs switches between regimes. These assumptions have been criticised as too restrictive, leading some authors, e.g. [Serra et al. \[2006a\]](#), and [Serra et al. \[2006b\]](#), to propose the use of non-parametric estimation techniques. We used a fully data-driven non-parametric VECM estimation

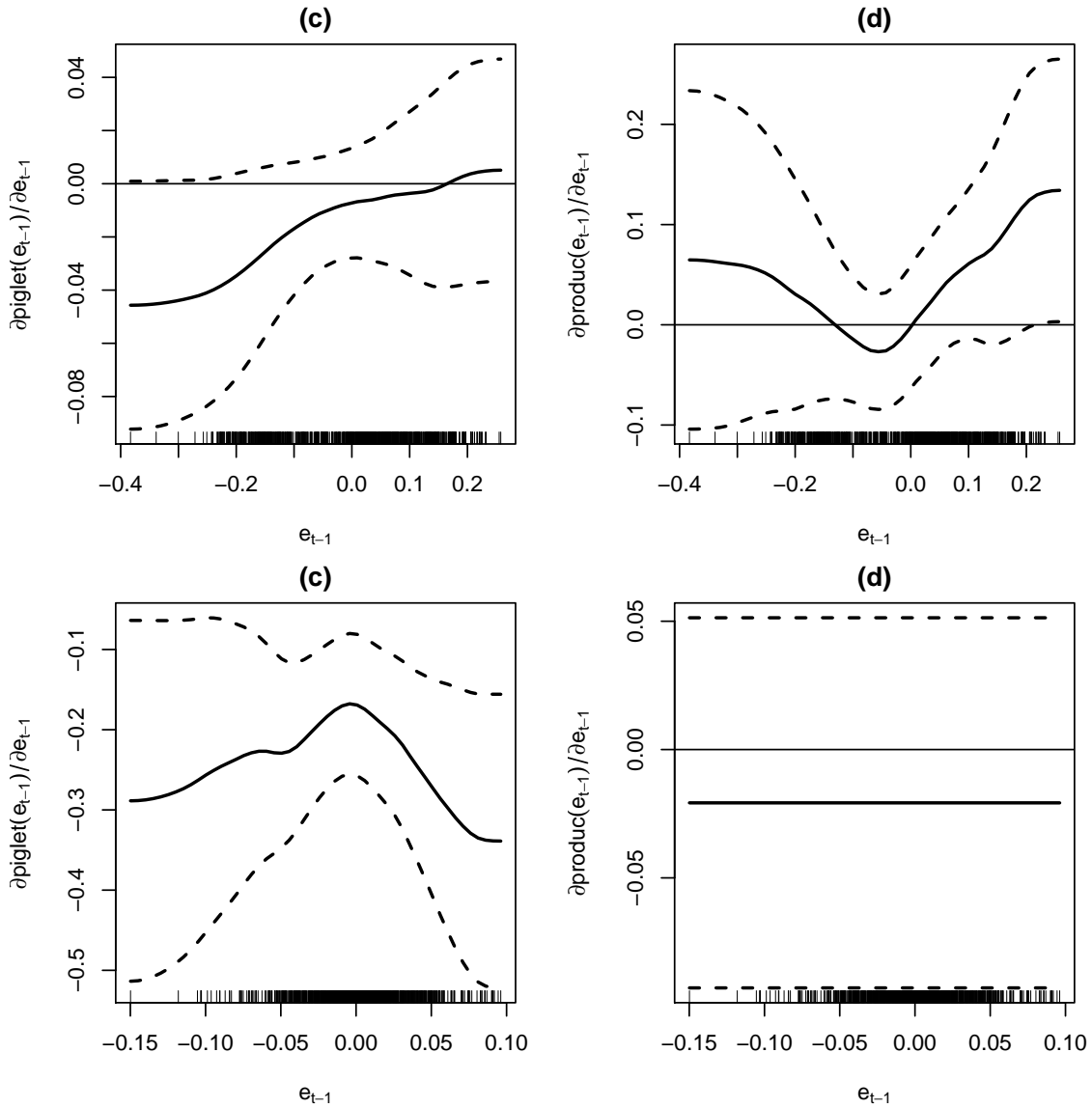


Figure 5.4.: Result of semi-parametric error correction model. (a) piglet speed of adjustment wrt error correction term; (b) slaughter pig speed of adjustment wrt error correction term; (c) piglet speed of adjustment wrt error correction term where the pork cycle was removed for each series; (d) slaughter pig speed of adjustment wrt error correction term where the pork cycle was removed for each series.

method that allows for non-parametric error correction but also a time varying long run equilibrium relationship.

### 5.2. Stem Cell Dynamics

Stem cells feature the ability to become different types of cells, which is useful for applications in medicine like tissue replacement from the patient's own cells [cf. [Inverardi and Ricordi, 2001](#)]. Adult human mesenchymal stem cells from bone marrow (hMSCs) are able to differentiate into various cell lines like fat, cartilage, bone or muscle cells. Most strikingly is the observation that providing a distinct mechanical stimulus in form of substrate stiffness is sufficient to guide hMSC differentiation [cf. [Engler et al., 2006](#)]. Furthermore, within the first 24 hours on such elastic hydrogels, the cells show a distinct, substrate elasticity dependent, actin-cytoskeleton structure that can be used as an early morphological marker for early stem cell differentiation as presented in [Zemel et al. \[2010\]](#).

To elucidate the complex interplay of integration of mechanical cues from the outside to bio-chemical signals it is imperative to understand the kinetics of the formation and structure of acto-myosin stress fibres. This could provide insight into the mechanisms of hMSC differentiation strategies, which are relevant for medical applications. In this context, massive parallel live-cell imaging of hMSCs were performed on polyacrylamide substrates with Young's modulus of 1, 10 and 30 kilo Pascals (kPa) to determine early stem cell differentiation to neuronal, muscle and bone tissue precursor cells. From a developmental biology point of view, it is expected that hMSCs on 10 kPa and 30 kPa polyacrylamide will show a much more similar behaviour when compared to the cells on 1 kPa.

### 5.2.1. Experiment

A detailed explanation of the experiment can be found in [Wollnik \[2016\]](#). Here we summarise its main aspects to provide some insight into the data generating process. Namely, adult human mesenchymal stem cells (hMSCS) from bone marrow were obtained from Lonza (Switzerland) and transfected with pCMV Lifeact-TagRFP (ibidi, Germany) using electroporation (Nucleofector, Lonza, Switzerland). Cells were cultured in DMEM (life technologies, Germany) + 10 % FBS (Sigma-Aldrich, USA) + 1 % Pen/Strep (life technologies, Germany) at 37° and 5 % CO<sub>2</sub>. After transfection, cells were allowed to rest for 48 hours in T75 cell culture flasks (Corning, USA) and were then seeded onto polyacrylamide (PA) gels of distinct elasticity, as described before in [Zemel et al. \[2010\]](#). Solutions for PA gels were mixed from 40% acrylamide solution (Bio-Rad, USA) and cross-linker bis-acrylamide 2% solution (Bio-Rad, USA), and subsequently polymerised using TEMED (Bio-Rad, USA) and APS (Sigma-Aldrich, USA) for 60 minutes. Hydrogels were then coated with collagen I (rat tail collagen I, Corning, USA) using the cross-linker Sulfo-SANPAH (Thermo Scientific, USA). Following this step, gels on cover glasses were glued into bottomless petri-dishes (ibidi, Germany) by using UV-curable glue (NOA 68, Norland products inc., USA). During live-cell imaging on an inverted microscope (Axio Observer.Z1, Zeiss, Germany), cells were kept in a climatic chamber (ibidi, Germany) at 37° and 5 % CO<sub>2</sub> throughout the imaging time of 24 hours. Images were recorded with a Zyla camera (Zyla sCMOS 4.2, Andor, United Kingdom) at time intervals of ten minutes between two subsequent images.

### 5.2.2. Model

To model the stem cell dynamics we use the area measurement of the cell's image across time and conduct the FPCA approach presented in Section 4.2 to study the area dynamics, see e.g. Figure (5.5), where images of 10kPa stem cells are provided

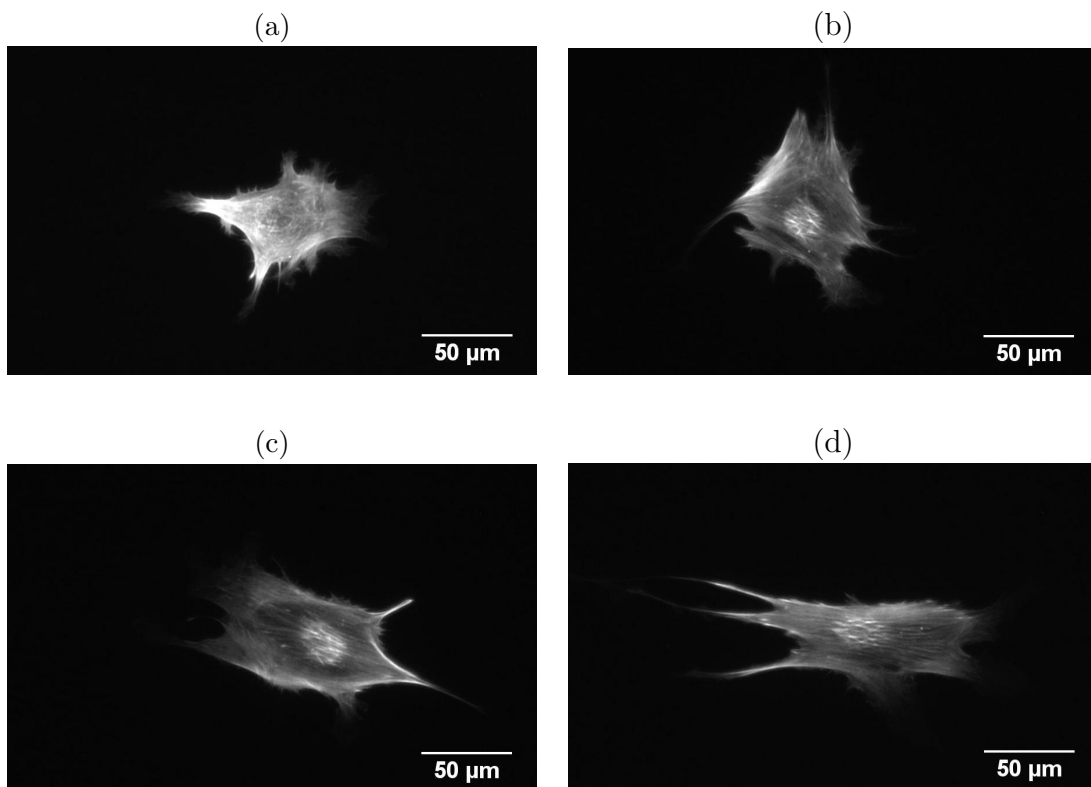


Figure 5.5.: Stem cell data example for 10 kilo Pascal. The panels show an arbitrary cell at: (a) hour 1; (b) hour 6; (c) hour 12; and (d) hour 24.

as an illustration. In particular, we assume that each group of cells is generated by a different data generating process as  $\boldsymbol{\tau}_{l,i}(x) = \{\boldsymbol{\tau}_l(x), x \in [0, 1]\}$ , where  $l$  represents the Young's modulus of the group and  $i$  is a subject specific realisation. The underlying assumption is that the time dynamic of the cell's area can be characterised as a smooth function in group  $l$ , and that each  $\boldsymbol{\tau}_{l,j}$  (each area process across time) is an independent realisation of that group. Namely, consider representation in (4.11) omitting seasonal components, that is

$$\boldsymbol{\tau}_{l,j}(x) = \boldsymbol{\tau}_l(x) + \sum_{k=1}^{h_l} \boldsymbol{\psi}_k(x)v_{j,k} + \boldsymbol{\epsilon}_{l,j}, \quad \boldsymbol{\epsilon}_{l,j} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_l),$$

where  $\boldsymbol{\tau}_k(x)$  represents the population mean in group  $l$ , and  $\boldsymbol{\eta}_{l,j} = \sum_{k=1}^{h_l} \boldsymbol{\psi}_k(x)v_{j,k}$  captures the subject specific deviation for each individual in their group, characterised as a linear combination of the first  $h_l$  principal components in the group's data.

### 5.2.3. Estimation

We estimate (5.5) for  $l = \{1\text{kP}, 10\text{kP}, 30\text{kP}\}$  separately but considering the same input variables in each case for comparative purposes. Namely we select  $h = 7$  based on the likelihood-ratio test to assess the significance of adding an additional principal component; To estimate the smoothness class of the population we applied criteria (3.18) for each individual for  $\mathbf{Y}_{l,j}^* = \hat{\mathbf{R}}_l^{-1/2}(\mathbf{Y}_{l,j} - \boldsymbol{\eta}_{l,j})$  and take the mode of the results as an estimator of  $q_\tau$  (in this case we obtained  $q_\tau = 3$ ). Additionally, inspection of the individual ACF's suggest the use of an AR(1) process for the error for all groups. Lastly regarding the subject specific deviations, we set  $q_\eta = 2$ , i.e a cubic spline without further analysis.

The results are reported in Figure (5.6) and reveal similar qualitative results to those found by Wollnik [2016]. Namely, the second row of plots in the same figure shows that the cell's area for 1 kilo Pascal portrays a significantly different behaviour when compare to its 10 and 30 kilo Pascal counterparts, however the differentiation among

## 5. Applications

---

the last two groups is not clear since the 95% confidence bands indeed cross each other. Moreover, the shaded grey area highlights hours 9, 10 and 11 in the overall trend for each group, where a change in the slopes of the overall trends is apparent. To study this feature further, we turn into the estimated functional principal components.

Figures (5.7) and (5.8) report the estimated functional principal components. Visual inspection suggests that the first component captures the curves intercept, while the second component mostly captures the groups' trends. Similarly, the third component reveals a landmark region for all groups that coincides with the highlighted grey area in Figure (5.6). For 1 and 10 kilo Pascal modulus this happens around the tenth hour, while for 30 kilo Pascal modulus the change point seems to be delayed for approximately two hours (see the shift to the right in the lower panel of Figure (5.7)). The interpretation of the second three most important functional principal components is more involved. The fourth principal component however shows a clear departure between the 1 kilo Pascal modulus and its counterparts, which happens notoriously around hours 14, 15 and 16, highlighted in the first panel of Figure (5.8).

For comparative purposes here we reproduce Table 4.6 from Wollnik [2016], where statistics of the point-wise mean per group are reported. Although the comparison is not direct, there are some qualitative similarities between both studies. Specifically there appears to be an initial state of growth until hour 10, which eventually lead to a plateau and a shrinkage in the 1 kilo Pascal cells. For the 10 kilo Pascal the shrinkage seems to be much more subtle, while for the 30 kilo Pascal the cells' area keep expanding until the end of the experiment.

Table 5.1.: Cell Area Development in Time

<b>Modulus</b>	Time Intervals				
	<b>0-5 h</b>	<b>5-10 h</b>	<b>10-15 h</b>	<b>15-20 h</b>	<b>20-24 h</b>
1 kPa	9.44%	0.55%	2.12%	−6.50%	−7.85%
10 kPa	14.82%	9.70%	0.41%	4.41%	−1.73%
30 kPa	13.63%	9.13%	4.31%	2.44%	2.57%

Source: Wollnik [2016]

## 5. Applications

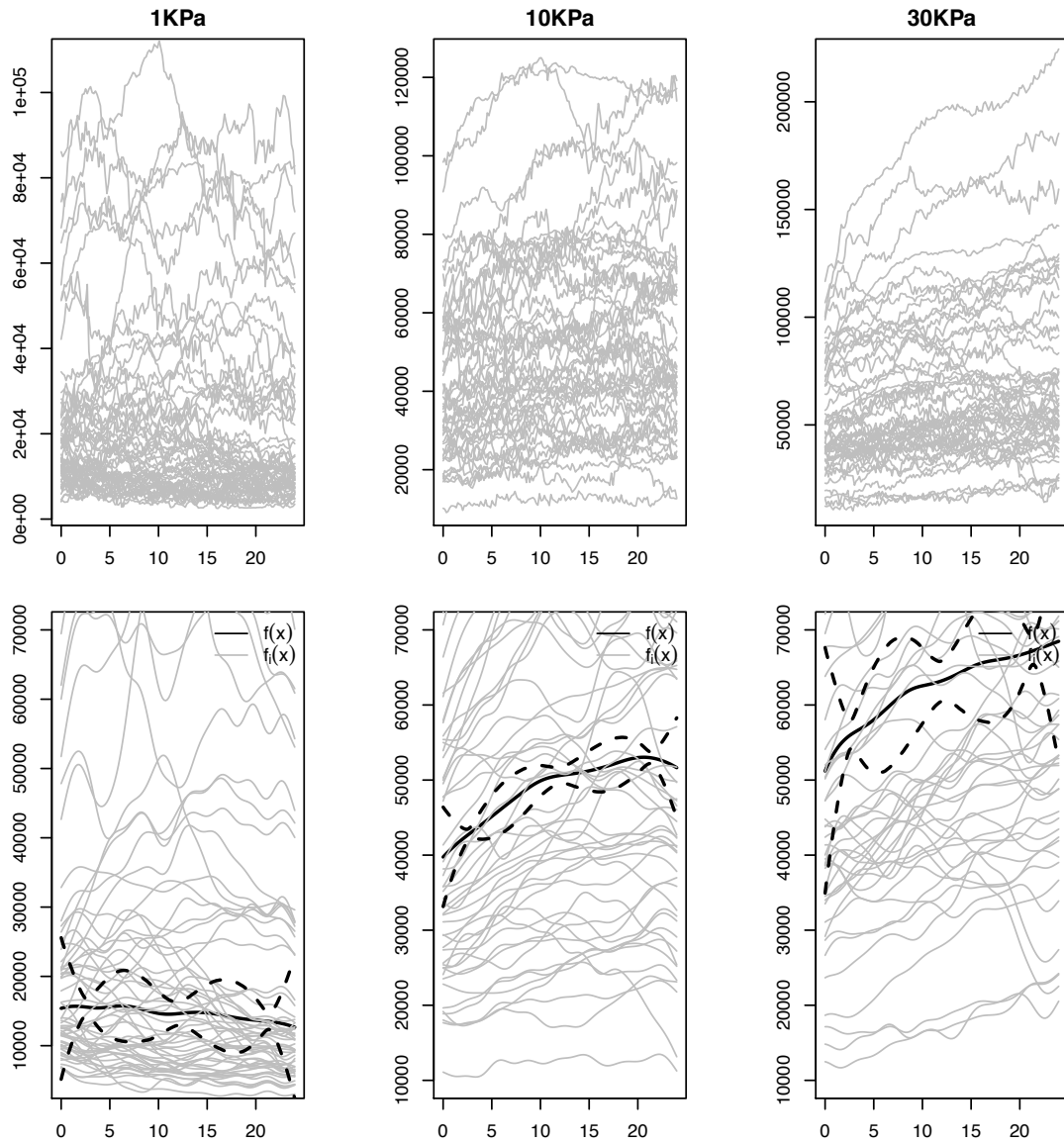


Figure 5.6.: FPCA for stem cell data. First row: raw data. Second row: population  $f(x)$  and subject specific estimations  $f_i(x)$ . The grey area correspond to hours 9, 10 and 11.



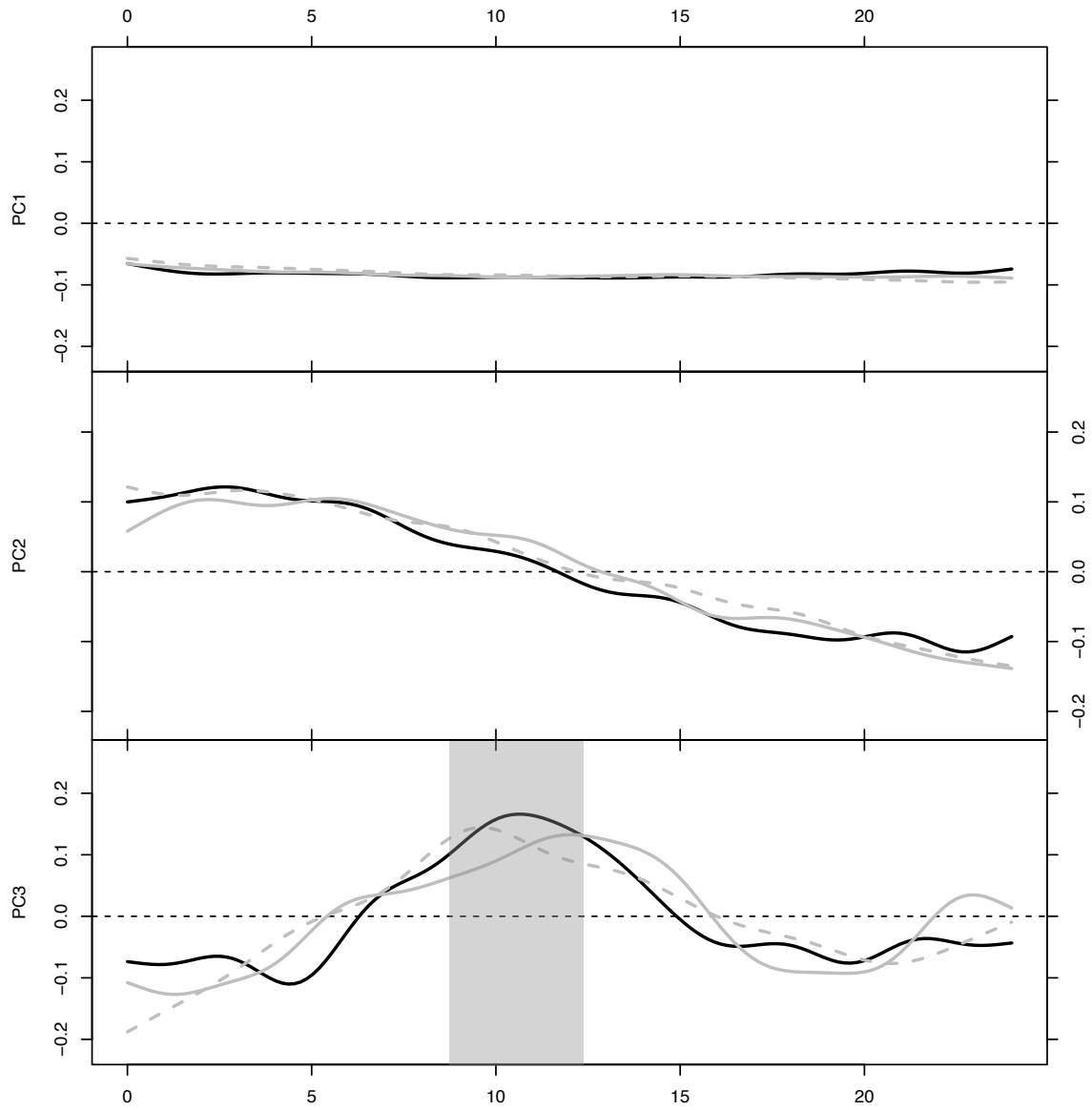


Figure 5.7.: First three functional principal components. Thick black line: 1 kilo Pascal; thick grey line: 10 kilo Pascal; dashed grey line: 30 kilo Pascal. The grey area for PC3 correspond to hours 9, 10 and 11.

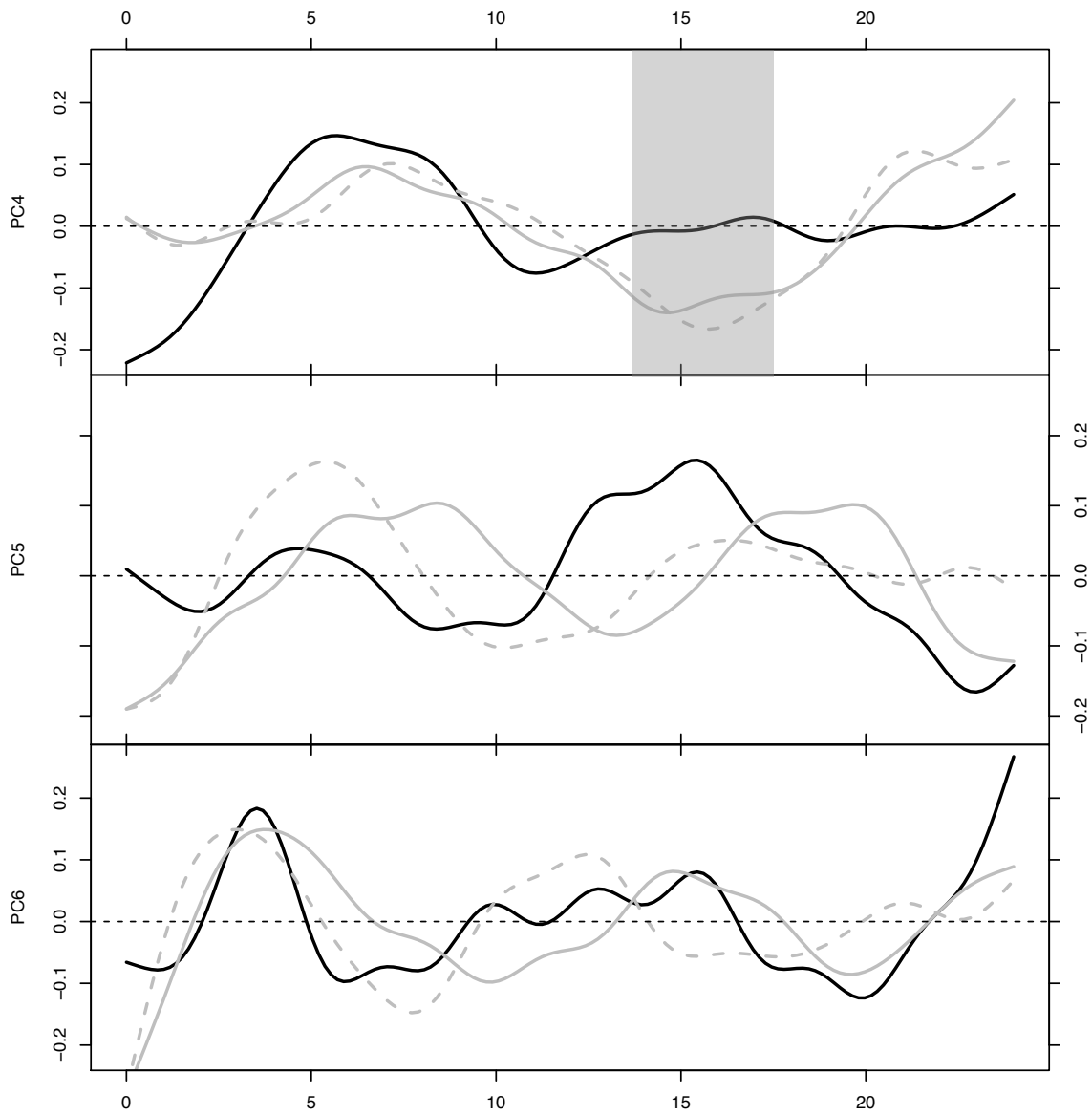


Figure 5.8.: Second three functional principal components. Thick black line: 1 kilo Pascal; thick grey line: 10 kilo Pascal; dashed grey line: 30 kilo Pascal. . The grey area for PC4 correspond to hours 14, 15 and 16.

### 5.2.4. Conclusions

Our method reports comparable results to those given in [Wollnik \[2016\]](#), but we consider they can be further improved. In principle, tracking only the cell's area is possibly insufficient to capture the whole dynamics of the cells, which are constructed as bidimensional objects. In this application we did not consider aspect ratio data (which is available from the experiment) because the roughness of its paths is a feature that does not adjust to the assumptions of the stochastic processes considered here. However other variables such as anisotropy or order parameter could also be explored. Additionally, it might be useful to consider a time varying AR(1) process for the error's measurement. This could be adequate since it is apparent that the dependence of the error weakens as the experiment progresses.

## 6. Software

The implementation of the ideas presented in Chapter 3 and Chapter 4 was developed in the R programming language. The computations and plots for the applications presented in Chapter 5 were conducted using the packages `scsd` and `fpca` for sections 5.1 and 5.2 respectively. Both packages depend on an additional package called `ebsc`, where the statistical model of the empirical smoothing splines under correlated errors was developed.

### 6.1. R Package `ebsc`

The approach proposed in Chapter 3 is implemented in the comprehensive R package `ebsc`. Its contents are based on the results presented in Chapter 3 and in [Serra and Krivobokova \[2016\]](#). To be able to manipulate numbers of large magnitude (resulting from the exact computation of the Demmler-Reinsch basis described in Chapter 2) we use the `Brobdingnag` library. To exploit the parallel structure of our implementation we require the `parallel` library. In principle these are the only dependencies of `ebsc`.

The main function of the package is called by

```
O<-ebsc(y,q.fixed,R,iterations,method,arma.order,parallel)
```

for arguments

---

<code>y</code>	One dimensional array of equidistant real values. No missing values are allowed.
<code>q.fixed</code>	Smoothness class of the function. If <code>q.fixed=NULL</code> , its optimal value is selected using the $T_q(\cdot)$ criteria described in Chapter 3.
<code>R</code>	Correlation matrix used to start the iterative process. If <code>R=NULL</code> , the identity matrix is used.
<code>iterations</code>	Maximum number of iterations. If <code>iterations=NULL</code> , 25 is used.
<code>method</code>	Method to estimate the model. If <code>method=NULL</code> the non-parametric estimation presented in Chapter 3 is done. If <code>method="MM"</code> the linear mixed model's representation of smoothing splines is fitted using the <code>nlme</code> package.
<code>arma.order</code>	Only necessary if <code>method="MM"</code> . It is a vector of the form $c(p, q)$ where <code>p</code> and <code>q</code> represent the AR and MA orders of the error process.
<code>parallel</code>	Option to select if parallel computation is to be used. If <code>parallel=NULL</code> a sequential estimation is performed.

The function returns the estimations of the smoothing parameter `lambda.hat`, the correlation `sigma2.hat` and `R.hat`, the smoothness class `q.hat` and the optimal fit `f.hat`.

## 6.2. R Package `scsd`

Signal decomposition under empirical Bayesian smoothing splines with correlated errors is implemented in the package `sdsc`. Details of the method are presented in subsection 4.1.2 of Chapter 4, hence the only dependency of the package is the library `ebsc` presented in the previous section.

The main function of the package is called by

```
O<-scsd(y,R,iterations,method,arma.order,parallel,method,h,q1.0)
```

for arguments

<code>y</code>	One dimensional array of equidistant real values. No missing values are allowed.
<code>R</code>	Correlation matrix used to start the iterative process. If <code>R=NULL</code> , the identity matrix is used.
<code>iterations</code>	Maximum number of iterations. If <code>iterations=NULL</code> , 25 is assumed.
<code>method</code>	Method to estimate the model. If <code>method=NULL</code> the non-parametric estimation presented in Chapter 3 is done. If <code>method="MM"</code> the linear mixed model's representation of smoothing splines is fitted using the <code>nlme</code> package.
<code>arma.order</code>	Only necessary if <code>method="MM"</code> . It is a vector of the form <code>c(p,q)</code> where <code>p</code> and <code>q</code> represent the AR and MA orders of the error process.
<code>parallel</code>	Option to select if parallel computation is to be used. If <code>parallel=NULL</code> a sequential estimation is performed.
<code>seas.struct</code>	Method to de-seasonalised the data. If <code>seas.struct=NULL</code> , the smoothing splines method is applied. Other options are <code>stl</code> and <code>d1m</code> .
<code>h</code>	Frequency of the seasonal component.
<code>q1.0</code>	Smoothness class of the trend component for the initial iteration.

The function returns the estimations of the smoothing parameter for the trend component `lambda.hat`, the correlation `sigma2.hat` and `R.hat`, the smoothness class of the trend component `q.hat` and the optimal fit for the seasonal component `seas.hat`, trend component `trend.hat` and overall fit `f.hat`. For more details the user can type `?scsd` in the terminal window.

### 6.3. R Package `fpca`

Functional principal component analysis using the mixed models representation of smoothing splines are developed in the package `fpca`. Details of the method are presented in subsection 4.2.2 of Chapter 4. The `fpca` package depends on libraries `nlme` and `ebsc` presented in the first section.

The main function of the package is called by

```
O<-fpca(DATA,k,q,correlation)
```

for arguments

<code>DATA</code>	Matrix where each column represents a sample measured in an equidistant grid of <code>nrow(DATA)</code> points. No missing values are allowed.
<code>k</code>	Number of eigenfunctions to construct subject specific deviations between 1 and <code>ncol(DATA)</code>
<code>q</code>	Smoothness class of the model specified as a vector of the form <code>c(q.pop,q.subj)</code> indicating the smoothness class of the overall trend and subject specific deviations respectively.
<code>correlation</code>	Correlation structure of the errors specified as <code>corARMA(p,q)</code> . If left empty a white noise error is assumed.

The function returns the population fit `f`, the subject specific deviations `di` and subject specific fits `fi`. Details of the fitted LMM can be inspect on the `est` object. For more details the user can type `?fpca` in the terminal window.

## 7. Summary and Outlook

Smoothing splines is a well established method in non-parametric statistics, although the selection of the smoothness class  $q$  for  $f \in \mathcal{W}^q[0, 1]$  is rarely addressed and, instead, a cubic smoothing spline, i.e.  $q = 2$ , is the standard assumption taken by practitioners. For a general regression function there is no known method to identify  $q$  under the presence of errors with correlation matrix  $\mathbf{R}$ . This apparent disregard in the literature can be justified because the condition number of the solution increases with  $q$ , turning the estimation unstable. In Chapter 2 we introduced an exact expression for the Demmler-Reinsch basis constructed as the solution of an ordinary differential equation. In Chapter 3 we presented an estimation procedure that can be carried out for an arbitrary  $q$  and for a general positive definite Toeplitz matrix  $\mathbf{R}$ , without affecting the condition number of the solution. Asymptotic properties of the proposed estimators were provided in Section 3.3, and Monte Carlo experiments were conducted to study their finite sample properties and reported in section 3.4.2. In Chapter 4 we presented extensions where the empirical Bayes smoothing splines method serves as a building block. Namely, we considered extensions to signal extraction and functional principal component analysis. In Chapter 5 the empirical relevance to our findings was shown by applications in agricultural economics and biophysics. Finally, Chapter 6 provided a summarised description of the R software developed to implement the ideas presented in chapters 3 and 4 and that were used to obtain the results reported in Chapter 5.

The studying of empirical Bayes smoothing splines with correlated errors is of great



importance. In theoretical terms, the results presented in Section 3.3 showed that the estimators for the smoothing parameter and correlation matrix are consistent under short range correlation for the provided statistical algorithm. Moreover, we showed that our method outperforms its benchmark competitors in the literature in finite samples via Monte Carlo experiments. We also have presented the relevance of the empirical Bayes smoothing splines method for real life applications. Namely, values of  $q = 5$  and  $q = 3$  were reported for the non-parametric price transmission and stem cell applications presented in Chapter 5. These results are not surprising and, in fact, values of  $q > 2$  have been previously observed in the literature. Consider for example the so-called SiZer feature significance problem, as presented in Hanning and Marron [2006], or the study of phase plots in functional data analysis as described by Ramsay and Silverman [1997], where the second or third derivatives (and not the regression function itself), are the focus of the study, and hence large values of  $q$  are presumed.

The present work can be extended in various directions. A possible field for future research is the study of the relation between state space models and smoothing splines of general degree  $q$ . As presented in Section 4.1, the work by Kohn and Ansley [1987] suggests that it is possible to represent the empirical Bayes smoothing splines with correlated errors presented in Chapter 3, as a state space model where the smoothness class of the trend component is data driven. This is a non-trivial extension of the, already well known, relation between cubic smoothing splines and state space models [cf. Harvey, 1989, Zivot and Wang, 2006]. Another interesting area for future developments is the extension of the FPCA method developed in Section 4.2 to the dependent functional data case, and to explore a method to further estimate the smoothness class of the subject specific deviations.

# Appendices

# A. Auxiliary Results

Here we present some auxiliary results that are required for Section 3.3 of Chapter 3.

Let  $\{\psi_k\}_{k \in \mathbb{N}}$  represent the Fourier basis functions such that for  $k \in \mathbb{N}$ ,

$$\psi_1(t) = 1, \quad \psi_{2k}(t) = \sqrt{2} \cos(2\pi kt), \quad \text{and} \quad \psi_{2k+1}(t) = \sqrt{2} \sin(2\pi kt). \quad (\text{A.1})$$

Define the matrix

$$\Psi = [n^{-1/2} \psi_j(t_i)]_{i,j=1}^n, \quad (\text{A.2})$$

so that  $\Psi^T \Psi = \Psi \Psi^T = \mathbf{I}$ .

We define the (rescaled) spectral density of the correlation function  $r$  as

$$\tilde{r}(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-2\pi i k \omega}. \quad (\text{A.3})$$

For two real matrices  $\mathbf{A}, \mathbf{B}$  of the same dimension, let  $\mathbf{A} \approx \mathbf{B}$  mean that the entries of  $\mathbf{A} - \mathbf{B}$  converge uniformly to zero, as  $n$  goes to infinity. By Proposition 4.5.2 of Brockwell and Davis [2009],

$$\Psi^T \mathbf{R} \Psi \approx \text{diag} \{ \tilde{r}(t_1), \tilde{r}(t_2), \tilde{r}(t_2), \tilde{r}(t_3), \tilde{r}(t_3), \dots \} = \text{diag}(\tilde{\mathbf{r}}). \quad (\text{A.4})$$

Assume that  $\tilde{r}$  is absolutely continuous. Note that by choice of the design points,  $\tilde{r}_1 = \tilde{r}(t_1) = \tilde{r}(0)$  which we define as  $\rho$  so that that for any sequence  $m_n = o(n)$ , the

## A. Auxiliary Results

---

eigenvalues satisfy  $|\tilde{r}_i - \rho| \rightarrow 0$ , uniformly over  $i = 1, \dots, m_n$ . This follows since

$$\sup_{i=1, \dots, m_n} |\rho - \tilde{r}(t_i)| = \sup_{i=1, \dots, m_n} |\tilde{r}(0) - \tilde{r}(t_i)| \leq \sup_{i=1, \dots, m_n} \int_0^{t_i} |\tilde{r}'(x)| dx \leq \int_0^{t_{m_n}} |\tilde{r}'(x)| dx,$$

which goes to zero as  $n \rightarrow \infty$ .

We now compute some traces and quadratic forms involving the smoother matrix  $\mathbf{S}_{\lambda, \mathbf{R}}$ . It is not difficult to see that if we denote  $\mathbf{\Delta} = \rho \mathbf{R}^{-1} - \mathbf{I}$ , then

$$\mathbf{S}_{\lambda, \mathbf{R}} = \mathbf{S}_{\rho\lambda, \mathbf{I}} \{ \mathbf{I} + \mathbf{\Delta}(\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \}. \quad (\text{A.5})$$

Since  $\mathbf{S}_{\lambda, \mathbf{R}} = \mathbf{R}^{1/2} \mathbf{C}_{\mathbf{R}} (\mathbf{C}_{\mathbf{R}}^T \mathbf{C}_{\mathbf{R}} + n\lambda\rho\mathbf{D})^{-1} \mathbf{C}_{\mathbf{R}}^T \mathbf{R}^{-1/2}$  where  $\mathbf{C}_{\mathbf{R}} = \mathbf{R}^{-1/2} \mathbf{C}$ , the eigenvalues of  $\mathbf{S}_{\lambda, \mathbf{R}}$  are between 0 and 1. For any sequence  $m_n$  as above we also have that the first  $m_n$  eigenvalues of  $\mathbf{\Delta}$  are uniformly close to zero, as  $n \rightarrow \infty$ , and the remaining eigenvalues are bounded. Hence conclude that the first  $m_n$  eigenvalues of  $\mathbf{I} + \mathbf{\Delta}(\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}})$  are uniformly  $1 + o(1)$ , as  $n \rightarrow \infty$ . Note also that by Lemma 1 of Krivobokova [2013] that for  $l \in \mathbb{N}$ , and  $m \in \mathbb{N} \cup \{0\}$ ,

$$\text{tr} \{ (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{I}})^m \mathbf{S}_{\lambda, \mathbf{I}}^l \} = \lambda^{-1/(2q)} \kappa_q(m, l) \{1 + o(1)\}, \quad (\text{A.6})$$

where  $\kappa_q(m, l) = \Gamma\{m+1/(2q)\} \Gamma\{l-1/(2q)\} / \{2\pi q \Gamma(m+l)\}$ ; cf. also Lemma 1 of Serra and Krivobokova [2016].

Let  $\mathbf{R}$  be a Toeplitz correlation matrix and denote by  $\tau$  the associated rescaled spectral density at 0, as defined in (A.3). Using identity (A.5), relation (A.6), and the dominated convergence theorem, conclude that for  $l \in \mathbb{N}$ , and  $m, s, t \in \mathbb{N} \cup \{0\}$ ,

$$\text{tr} \{ \mathbf{R}^t \mathbf{R}^s (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}})^m \mathbf{S}_{\lambda, \mathbf{R}}^l \} = \tau^t \rho^s (\rho\lambda)^{-1/(2q)} \kappa_q(m, l) \{1 + o(1)\}. \quad (\text{A.7})$$

Consider now the quadratic form  $\mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f}$ . Using identity (A.5), for

$m, l \in \mathbb{N} \cup \{0\}$ ,

$$(\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}})^m \mathbf{S}_{\lambda, \mathbf{R}}^l = (\mathbf{I} + \mathbf{S}_{\rho\lambda, \mathbf{I}} \Delta)^{-m} (\mathbf{I} - \mathbf{S}_{\rho\lambda, \mathbf{I}})^m \mathbf{S}_{\rho\lambda, \mathbf{I}}^l \{\mathbf{I} + \Delta(\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}})\}^l.$$

By use of the identity from the previous display and the same argument as with the trace, if  $\mathbf{f}^T (\mathbf{I} - \mathbf{S}_{\rho\lambda, \mathbf{I}}) \mathbf{S}_{\rho\lambda, \mathbf{I}} \mathbf{f}$  converges, then

$$\mathbf{f}^T (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f} = \mathbf{f}^T (\mathbf{I} - \mathbf{S}_{\rho\lambda, \mathbf{I}}) \mathbf{S}_{\rho\lambda, \mathbf{I}} \mathbf{f} \{1 + o(1)\},$$

so that we conclude that if  $\mathbf{f}^T (\mathbf{I} - \mathbf{S}_{\rho\lambda, \mathbf{I}}) \mathbf{S}_{\rho\lambda, \mathbf{I}} \mathbf{f}$  converges, then

$$\mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f} = \frac{1}{\rho} \mathbf{f}^T (\mathbf{I} - \mathbf{S}_{\rho\lambda, \mathbf{I}}) \mathbf{S}_{\rho\lambda, \mathbf{I}} \mathbf{f} \{1 + o(1)\}. \quad (\text{A.8})$$

Indeed by Lemma 2 in Krivobokova [2013], and Lemma 3 in Serra and Krivobokova [2016], the quadratic form above converges and

$$\mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f} = n\lambda \|f^{(q)}\|^2 \{1 + o(1)\}, \quad f \in \mathcal{W}_q, \quad (\text{A.9})$$

$$\mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f} \leq n\rho^{-1} (\rho\lambda)^{\beta/q} \|f^{(\beta)}\|^2 \{1 + o(1)\}, \quad f \in \mathcal{W}_\beta, \quad q > \beta. \quad (\text{A.10})$$

It also holds that that for  $l \in \mathbb{N} \cup \{0\}$ ,

$$\mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}}^l \mathbf{f} = \mathbf{f}^T \mathbf{R}^{-1} (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \mathbf{S}_{\lambda, \mathbf{R}} \mathbf{f} \{1 + o(1)\}. \quad (\text{A.11})$$

Using the identity in (A.5) one can (approximately) diagonalise  $\mathbf{S}_{\lambda, \mathbf{R}}$ . Let  $\Phi$  represent the Demmler-Reinsch basis matrix such that  $\Phi^T \Phi = \Phi \Phi^T = \mathbf{I}$ . Then

$$\Phi^T \mathbf{S}_{\lambda, \mathbf{R}} \Phi = \{\mathbf{I} + n\lambda\rho \text{diag}(\boldsymbol{\eta})\}^{-1} \{\mathbf{I} + \Phi^T \Delta (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{R}}) \Phi\}, \quad (\text{A.12})$$

where  $\boldsymbol{\eta}$  are the eigenvalues of the Demmler-Reinsch basis matrix.

## **B. Additional Tables**

B. Additional Tables

Table B.1.: Simulation Results (Scenario III): MSE

$(\phi_1, \phi_2)$	(0, 0)	(0.2, 0)	(0.4, 0)	(0.6, 0)	(0.52, -0.3)	(0.3, -0.52)	(0, 0.3)	(0, -0.3)	(0, 0.6)	(0, -0.6)
$f_1$										
EBS	<b>0.11</b> (0.04)	<b>0.16</b> (0.06)	<b>0.23</b> (0.07)	<b>0.39</b> (0.12)	<b>0.14</b> (0.06)	<b>0.06</b> (0.02)	<b>0.19</b> (0.07)	<b>0.07</b> (0.03)	<b>0.37</b> (0.14)	<b>0.04</b> (0.01)
HER	0.29 (0.08)	0.33 (0.09)	0.37 (0.10)	0.49 (0.15)	0.31 (0.09)	0.26 (0.05)	0.34 (0.10)	0.25 (0.04)	0.46 (0.15)	0.24 (0.03)
LMM	0.14 (0.04)	0.20 (0.07)	0.28 (0.08)	0.42 (0.13)	0.17 (0.06)	0.07 (0.02)	0.22 (0.07)	0.08 (0.02)	0.39 (0.12)	<b>0.04</b> (0.01)
HVK	0.19 (0.06)	0.66 (0.15)	0.78 (0.23)	1.34 (0.34)	0.66 (0.16)	0.60 (0.08)	1.19 (0.21)	0.60 (0.08)	2.04 (0.41)	0.58 (0.06)
$f_3$										
EBS	<b>0.15</b> (0.04)	0.21 (0.06)	0.29 (0.08)	0.48 (0.26)	0.19 (0.06)	0.11 (0.06)	0.25 (0.07)	0.10 (0.02)	0.50 (0.22)	0.09 (0.05)
HER	0.49 (0.07)	0.53 (0.09)	0.57 (0.11)	0.66 (0.17)	0.50 (0.07)	0.46 (0.04)	0.54 (0.11)	0.47 (0.05)	0.65 (0.14)	0.46 (0.03)
LMM	<b>0.15</b> (0.04)	<b>0.20</b> (0.06)	<b>0.28</b> (0.08)	<b>0.42</b> (0.12)	<b>0.17</b> (0.05)	<b>0.08</b> (0.02)	<b>0.24</b> (0.07)	<b>0.09</b> (0.02)	<b>0.39</b> (0.12)	<b>0.05</b> (0.01)
HVK	0.16 (0.04)	0.33 (0.09)	0.45 (0.15)	0.62 (0.20)	0.32 (0.07)	0.18 (0.05)	0.52 (0.13)	0.28 (0.05)	0.83 (0.26)	0.15 (0.02)

$\|\mathbf{f} - \hat{\mathbf{f}}\|_2$ : mean and standard deviations (in parenthesis) are reported. Quantities scaled by  $10^4$ .

B. Additional Tables

Table B.2.: Simulation Results (Scenario IV): MSE

$(\phi_1, \phi_2)$	(0, 0)	(0.2, 0)	(0.4, 0)	(0.6, 0)	(0.52, -0.3)	(0.3, -0.52)	(0, 0.3)	(0, -0.3)	(0, 0.6)	(0, -0.6)
$f_1$										
EBS	<b>0.11</b> (0.04)	<b>0.16</b> (0.06)	<b>0.22</b> (0.08)	<b>0.38</b> (0.12)	<b>0.13</b> (0.05)	<b>0.06</b> (0.02)	<b>0.18</b> (0.06)	<b>0.06</b> (0.02)	<b>0.34</b> (0.10)	<b>0.03</b> (0.01)
HER	0.29 (0.08)	0.33 (0.09)	0.37 (0.10)	0.49 (0.15)	0.31 (0.09)	0.26 (0.05)	0.34 (0.10)	0.25 (0.04)	0.46 (0.16)	0.24 (0.03)
LMM	0.14 (0.04)	0.20 (0.07)	0.28 (0.08)	0.42 (0.13)	0.17 (0.06)	0.07 (0.02)	0.22 (0.07)	0.08 (0.02)	0.39 (0.12)	0.04 (0.01)
HVK	0.19 (0.06)	0.66 (0.15)	0.78 (0.23)	1.34 (0.34)	0.66 (0.16)	0.60 (0.08)	1.19 (0.21)	0.60 (0.08)	2.05 (0.42)	0.58 (0.06)
$f_3$										
EBS	<b>0.15</b> (0.04)	<b>0.20</b> (0.06)	<b>0.28</b> (0.08)	<b>0.42</b> (0.12)	<b>0.17</b> (0.05)	<b>0.08</b> (0.02)	<b>0.24</b> (0.07)	<b>0.09</b> (0.02)	<b>0.39</b> (0.12)	<b>0.05</b> (0.01)
HER	0.49 (0.07)	0.53 (0.09)	0.57 (0.11)	0.66 (0.17)	0.50 (0.07)	0.46 (0.04)	0.55 (0.11)	0.47 (0.05)	0.65 (0.14)	0.46 (0.03)
LMM	<b>0.15</b> (0.04)	<b>0.20</b> (0.06)	<b>0.28</b> (0.08)	<b>0.42</b> (0.12)	<b>0.17</b> (0.05)	<b>0.08</b> (0.02)	<b>0.24</b> (0.07)	<b>0.09</b> (0.02)	<b>0.39</b> (0.12)	<b>0.05</b> (0.01)
HVK	0.16 (0.04)	0.33 (0.09)	0.45 (0.15)	0.62 (0.20)	0.32 (0.07)	0.18 (0.05)	0.53 (0.13)	0.28 (0.05)	0.83 (0.26)	0.15 (0.02)

$\|\mathbf{f} - \hat{\mathbf{f}}\|_2$ : mean and standard deviations (in parenthesis) are reported. Quantities scaled by  $10^4$ .



# Bibliography

- M. Aerts, G. Claeskens, and M. Wand. Some theory for penalized spline generalized additive models. *Journal of Statistical Planning and Inference*, 103:455–470, 2002.
- T. Alexandrov, S. Bianconcini, E. Dagum, P. Maass, and T. McElroy. A review of some modern approaches to the problem of trend extraction. *Econometric Reviews*, 31(6):593–624, 2012.
- J. Amikuzuno and S. von Cramon-Taubadel. Seasonal variation in price transmission between tomato markets in ghana. *Journal of African Economics*, 21:669–686, 2012.
- R. Bacon. Rockets and feathers: The asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Economics*, 13:211–218, 1991.
- M. Baxter and R. King. Measuring business cycles: approximate band-pass filters for economic time series. *Review of Economics and Statistics*, 81(4):575–593, 1999.
- E. Berg and R. Huffaker. What drives the german hog price cycle? diagnostic modeling of a nonlinear dynamic system. *Proceedings of the European Association of Agricultural Economists Congress*, 2014.
- G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. HOLDEN-DAY, Oackland–California, 1971.

- P. Brockwell and R. Davis. *Time series: Theory and Methods*. Springer Science & Business Media, 2009.
- S. Butterworth. On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer*, 7:536–541, 1930.
- Y. Cao. *Inégalités d’Oracle pour l’ Estimation de la Régression*. PhD thesis, Université I de Provence U.F.R. M.I.M, 2008.
- L. Christiano and T. Fitzgerald. The band pass filter. *International Economic Review*, 44(2):435–465, 2003.
- R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning. STL: A seasonal–trend decomposition procedure based on LOESS. *Journal of Official Statistics*, 6(1):3–73, 1990.
- W. Cleveland. Robust locally regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.
- R. Corless and D. Jeffrey. The Turing factorization of a rectangular matrix. *SIGSAM Bulletin*, 31(3):20–28, 1997.
- D. Cox. Asymptotics for m-type smoothing splines. *The Annals of Statistics*, 11: 530–551, 1983.
- P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1978.
- E. Dagum. Modeling, forecasting and seasonally adjusting economic time series with the X-11 ARIMA method. *Journal of the Royal Statistical Society. Series D*, 27(3): 203–216, 1978.
- C. de Boor. *A Practical Guide to Splines*. Springer, Berlin–Heidelberg, 1997.
- A. Demmler and C. Reinsch. Oscillation matrices with spline smoothing. *Numerische Mathematik*, 24:375–382, 1975.

- J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
- P. Eilers and B. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–102, 1996.
- P. Eilers, B. Marx, and M. Durban. 20 years of P-splines. *SORT*, 39(2):1–38, 2015.
- R. Engle and C. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55:251–276, 1987.
- A. Engler, S. Sen, H. Sweeney, and D. Discher. Matrix elasticity directs stem cell lineage specification. *Cell*, 126:677–689, 2006.
- R. Eubank. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker Inc., 1988.
- M. Ezekiel. The cobweb theorem. *The Quarterly Journal of Economics*, 52:255–280, 1938.
- J. Favard. Sur les polynôme de Tchebychev. *C.R. Acad. Sci. Paris*, 200:2052–2053, 1935.
- F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis*. Springer, 2006.
- D. Findley, B. Monsell, W. Bell, M. Otto, and B. Chen. New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, 16(2):127–177, 1998.
- B. Gardner. The farm-retail price spread in a competitive food industry. *American Journal of Agricultural Economics*, 57(3):399–409, 1975.
- W. Gautschi. Attenuation factors in practical Fourier analysis. *Numerische Mathematik*, 18(5):373–400, 1971.
- B. Geiger, J. Spatz, and A. Bershadsky. Environmental sensing through focal adhesions. *Nat. Rev. Mol. Cell Biol.*, 10:21–33, 2009.

- E. Godolphin. Decomposition of time series models in state-space form. *Computational Statistics and Data Analysis*, 50(9):2232–2246, 2001.
- G. Golub and F. van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore and London, 1996.
- B. Goodwin and N. Piggott. Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics*, 83:302–317, 2001.
- F. Greb, S. von Cramon-Taubadel, T. Krivobokova, and A. Munk. The estimation of threshold models in price transmission analysis. *American Journal of Agricultural Economics*, 95:900–916, 2013.
- M. Guo, L. Zhou, J. Huang, and W. Härdle. Functional data analysis of generalized regression quantiles. *Statistics and Computing*, 25(2):189–202, 2013.
- M. Gutknecht. Attenuation factors in multivariate Fourier analysis. *Numerische Mathematik*, 51(6):615–629, 1987.
- P. Hall and I. Van Keilegom. Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society B*, 65:443–456, 2003.
- J. Hanning and J. Marron. Advance distribution theory for SiZer. *Journal of the American Statistical Association*, 101:484–499, 2006.
- J. Hart. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53:173–187, 1991.
- A. Harvey. *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989.
- A. Harvey. *Forecasting with Unobserved Components Time Series Models*. Handbook of Economic Forecasting. Elsevier, 2006.

- T. Hastie and C. Sugar. Principal component models for sparse functional data. *Biometrika*, 87:587–602, 2000.
- T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, USA, 1990.
- T. Hastie and R. Tibshirani. Varying coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55:756–796, 1993.
- R. Henderson. Note on graduation by adjusted average. *Transactions of the American Society of Actuaries*, 17:43–48, 1916.
- E. Herrmann, T. Gasser, and A. Kneip. Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79:783–795, 1992.
- R. Hodrick and E. Prescott. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997.
- C. Holst and S. von Cramon-Taubadel. Einfluss des schweinezyklus auf die preistransmission zwischen ferkel und schlachtschweinepreisen in niedersachsen. *Schriften der Gesellschaft für Wirtschafts und Sozialwissenschaften des Landbaues*, 46:137–150, 2011.
- R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.
- T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley and Sons, Wes Sussex, United Kingdom, 2015.
- S. Hylleberg, R. Engle, C. Granger, and B. Yoo. Seasonal integration and cointegration. *Journal of Econometrics*, 44:215–238, 1990.
- R. Ihle. *Models for Analyzing Nonlinearities in Price Transmission*. PhD thesis, Georg-August Universitaet Goettingen, 2009.

- L. Inverardi and C. Ricordi. Tolerance and pancreatic islet transplantation. *Philosophical Transactions of the Royal Society B*, 356:759–765, 2001.
- G. James, T. Hastie, and C. Sugar. Principal component models for sparse functional data. *Biometrika*, 87:587–602, 2000.
- R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- G. Kauermann, T. Krivobokova, and W. Semmler. Filtering time series with penalized splines. *Studies in Nonlinear Dynamics & Econometrics*, 15(2), 2011.
- R. Kohn and C. Ansley. A new algorithm for spline smoothing based on smoothing a stochastic process. *Journal of Scientific and Statistical Computing*, 8, 1987.
- T. Krivobokova. Smoothing parameter selection in two frameworks for penalized splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):725–741, 2013.
- T. Krivobokova and G. Kauermann. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102(480):1328–1337, 2007.
- T. Lloyd, S. McCorriston, W. Morgan, and A. Rayner. Food scares, market power and price transmission: the UK BSE case. *European Review of Agricultural Economics*, 33:119–147, 2006.
- A. Maravall and G. Caporello. Program TSW: Revised reference manual. Technical Report, Research Dep., Bank of Spain. <http://www.bde.es>, 2004.
- T. McMurry and D. Politis. Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31:471–482, 2010.
- R. Mendelssohn. The STAMP software for state space models. *Journal of Statistical Software*, 41(2):1–18, 2011.

- J. Meyer and S. von Cramon-Taubadel. Asymmetric price transmission: a survey. *Journal of Agricultural Economics*, 55:581–611, 2004.
- D. Nychka. Splines as local smoothers. *The Annals of Statistics*, 23:1175–1197, 1995.
- F. O’Sullivan. A statistical perspective on ill-posed problems. *Statistical science*, 1(4):502–518, 1986.
- G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. Springer-Verlag, 2009.
- D. Pollock. Econometric methods of signal extraction. *Computational Statistics and Data Analysis*, 50(9):2268–2292, 2006.
- A. Polyanin and V. Zaitsev. *Handbook of Exact Solutions for Ordinary Differential Equations*. Chapman & Hall/CRC, 2003.
- J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer, 1997.
- C. Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10:177–183, 1967.
- J. Rice and M. Rosenblatt. Integrated mean squared error of a smoothing spline. *Journal of approximation theory*, 33:353–369, 1981.
- D. Ruppert, M. Wand, and R. Carroll. *Semiparametric Regression*. Cambridge University Press, New York, 2003.
- I. Schoenberg. Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences of the United States of America*, 52:947, 1964.
- L. Schumaker. *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge, United Kingdom., 2007.
- K. Schwartz. *A Unified Framework for Spline Estimators*. PhD thesis, Georg-August Universität Göttingen, 2012.

- P. Serra and T. Krivobokova. Adaptive empirical bayesian smoothing splines. *Bayesian Analysis*, DOI:10.1214/16-BA997, 2016.
- T. Serra, J. Gil, and B. Goodwin. Local polynomial fitting and spatial price relationships: price transmission in eu pork markets. *European Review of Agricultural Economics*, 33(3):415–436, 2006a.
- T. Serra, B. Goodwin, J. Gil, and M. Mancuso. Non-parametric modeling of spatial price relationships. *Journal of Agricultural Economics*, 57:501–522, 2006b.
- P. Speckman. Spline smoothing and optimal rates of convergence in nonparametric regression models. *The Annals of Statistics*, 13:970–983, 1985.
- P. Speckman and D. Sun. Fully Bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika*, 90(2):289–302, 2003.
- T. Teräsvirta and A. Eliasson. Non-linear error correction and the U.K. demand for broad money. *Journal of Applied Econometrics*, 16:1878–1993, 2001.
- V. Tikhomirov. *Fundamental Principles of the Theory of Extremal Problems*. John Wiley & Sons, 1986.
- H. Tong and K. Kim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B. (Methodological)*, 42(3):245–292, 1980.
- F. Utreras. Sur le choix du paramètre d’ajustement dans le lissage par fonctions spline. *Numerische Mathematik*, 34:15–28, 1980.
- D. van Dijk, T. Teräsvirta, and P. Franses. Smooth transition autoregressive models: a survey of recent developments. *Econometric Reviews*, 21(1):1–47, 2002.
- G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):364–372, 1978.
- G. Wahba. *Spline Models for Observational Data*, volume 59. SIAM, 1990.



- M. Wand and J. Ormerod. On semiparametric regression with O'Sullivan penalized splines. *Australian and New Zealand Journal of Statistics*, 50(2):179–198, 2008.
- M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1997.
- E. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–74, 1923.
- M. Wiesenfarth, T. Krivobokova, S. Klasen, and S. Sperlich. Direct simultaneous inference in additive models and its application to model undernutrition. *Journal of the American Statistical Association*, 107(500):1286–1296, 2012.
- C. Wollnik. *Structure and Dynamics of Stress Fibres in Adult Stem Cells*. PhD thesis, Georg-August Universität Göttingen, 2016.
- S. Wood. *Generalized Additive Models: an Introduction with R*. Chapman & Hall/CRC, 2006.
- H. Xiao and W. Wu. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1):466–493, 2012.
- P. Young, D. Pedregal, and W. Tych. Dynamic harmonic regression. *Journal of Forecasting*, 18(6):369–394, 1999.
- A. Zemel, F. Rehfeldt, A. Brown, D. Discher, and S. Safran. Optimal matrix rigidity for stress-fibre polarization in stem cells. *Nature Physics*, 6:468–473, 2010.
- E. Zivot and J. Wang. *Modeling Financial Time Series with S-Plus*. Springer Science & Business Media, Inc., 2006.