

2015-01-01

Bayesian Adaptive Penalized Splines In Nonparametric Regression And In Spectral Time Series Analysis

Luis Angel Mora

University of Texas at El Paso, lusmora4@gmail.com

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Mora, Luis Angel, "Bayesian Adaptive Penalized Splines In Nonparametric Regression And In Spectral Time Series Analysis" (2015).
Open Access Theses & Dissertations. 1104.
https://digitalcommons.utep.edu/open_etd/1104

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

BAYESIAN ADAPTIVE PENALIZED SPLINES IN NONPARAMETRIC
REGRESSION AND IN SPECTRAL TIME SERIES ANALYSIS

LUIS ANGEL MORA

Department of Mathematical Science

APPROVED:

Ori Rosen, Chair, Ph.D.

Joan Staniswalis, Ph.D.

Vanessa Lougheed, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

©Copyright

by

Luis Angel Mora

2015

to my

MOTHER and FATHER

with love

BAYESIAN ADAPTIVE PENALIZED SPLINES IN NONPARAMETRIC
REGRESSION AND IN SPECTRAL TIME SERIES ANALYSIS

by

LUIS ANGEL MORA

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Science

THE UNIVERSITY OF TEXAS AT EL PASO

August 2015

Acknowledgements

First and foremost, I would like to thank God for giving me the wisdom and courage in my life.

I want to thank deeply my advisor, Dr. Ori Rosen for taking me under his wing and giving me the opportunity to work along his side. It is a tremendous honor and privilege to be counted amongst his students. He was patient, supportive, but above all, he trusted my abilities even when I doubted in them. His belief in me and great sense of humor enriched my thesis experience, not only intellectually, but emotionally as well.

I want to also thank my committee members, Dr. Joan Staniswalis of the Department of Mathematical Sciences and Dr. Vanessa Loughheed of the Department Biological Sciences, both at The University of Texas at El Paso. Their presence on my committee was comforting and their support was invaluable.

Additionally, I want to thank the professors of statistics of the Department of Mathematical Sciences for all they do for the students: Dr. Panagais Moschopoulos, Dr. Amy Wagler, Dr. Naijun Sha, and Dr. Xiogang Su, all of whom I consider to be great role models. They are truly amazing individuals with exceptional abilities.

As a special note, I would like to extend my gratitude to the following individuals:

Dr. Joan Staniswalis

She always provided the nurturing and support that we all sometimes need. I did not have the opportunity to be a student in one of her classes, however, that did not stop me from discovering an even greater passion for statistics thanks to her. Moreover, I thank her for pushing me to my limits.

Dr. Ori Rosen

He was my mentor and didn't think twice being honest and blunt. I admired that of him, even though I think he never realized it. I learned from him the importance of being calm, but not hesitant in executing tasks. I thank him for sharing those values and his wealth of knowledge with me.

I want to thank my family for all their support and love. In the darkest times they gave me the strength and motivation to keep placing one foot in front of the other. They are my backbone and I could have not achieved this without them.

Abstract

A Bayesian approach to nonparametric regression using Penalized splines (P-splines) is presented. The approach uses the linear mixed model formulation of P-splines. The usual model assumes a single value for the smoothing parameter controlling the amount of smoothing of the fitted function. The main focus of the thesis is on spatially adaptive smoothing where the smoothing parameter is a function of the covariate so that different amounts of smoothing are applied in different regions of the covariate. An application to spectral time series analysis will be demonstrated. Markov chain Monte Carlo methods are used to make inference based on the posterior distribution.

Table of Contents

	Page
Acknowledgements	v
Abstract	vii
Table of Contents	viii
Chapter	
1 Introduction	1
1.1 Basis Functions	1
1.2 Time Series	3
2 Nonparametric Regression	6
2.1 Penalized Splines	6
2.2 Linear Mixed Models	7
2.3 The Linear Mixed Model	9
2.4 Bayesian Nonparametric Regression	10
2.5 Priors	12
2.6 Sampling Scheme	12
2.7 Application	14
3 Spatially Adaptive Smoothing	15
3.1 Literature Review	15
3.2 Bayesian Adaptive Penalized Splines	18
3.3 Model	18
3.4 Likelihood and Priors	19
3.5 Sampling Scheme	20
3.6 Simulation Study	22
4 Spatially Adaptive Smoothing for Spectral Analysis	28
4.1 Spectral Analysis	28

4.1.1	Stationarity	28
4.1.2	Periodogram	28
4.1.3	The Spectral Density	30
4.2	Model	30
4.3	Priors	31
4.4	Sampling Scheme	33
4.5	Simulation Study	34
5	Application	42
5.1	Southern Oscillation Index (SOI)	42
5.2	Analysis	43
5.3	Conclusion	44
Appendix		
A	Derivation of Sampling Schemes	50
A.1	BPS Sampling Scheme	50
A.2	BAPS Sampling Scheme	52
A.3	Whittle Sampling Scheme	55
B	Code	59
B.1	Bayesian Penalized Splines Code	59
B.2	BAPS Code	61
B.3	BAPS Code (Whittle estimate)	64
B.3.1	Non-Adaptive Whittle	64
B.3.2	Adaptive Whittle	67
	Curriculum Vitae	73

Chapter 1

Introduction

This chapter is devoted to a brief introduction to the methodology of nonparametric regression and time series. The idea is to estimate an unknown function via a finite number of basis functions. The estimation is done by using Penalized Splines (P-splines). The thesis first describes the methodology of nonparametric regression. Secondly, in Chapter 2 it will be shown that P-splines have a Linear Mixed Model (LMM) representation which can naturally be handled using a Bayesian approach. Thirdly, Chapter 3 explores spatially adaptive P-splines. These models are useful when the degree of smoothness of the underlying function varies with the covariate. Lastly, we employ spatially adaptive penalized splines in spectral estimation in Chapter 4, followed by an application to a real data set in Chapter 5. Simulations are done for each method. Markov chain Monte Carlo (MCMC) methods are used throughout the study.

1.1 Basis Functions

The use of P-splines was popularized by Eilers and Marx (1996). The methodology relies on a relatively small number of basis functions. Although, there is a wide variety of basis functions available, in what follows, we use truncated polynomials (Ruppert et al., 2003) given by

$$(x - \kappa)_+^p = \begin{cases} (x - \kappa)^p & \text{if } x \geq \kappa \\ 0 & \text{otherwise} . \end{cases}$$

The function $(x - \kappa)_+^p$ is constructed using a covariate value x and a knot κ that is located in the range of the covariate space. Any function of the above form is referred to as a

truncated polynomial of degree $p > 0$. A spline function is a linear combination of basis functions and it is written as

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{K_\kappa} b_j (x - \kappa_j)_+^p, \quad (1.1)$$

where κ_j is the j th knot in a set of knots $(\kappa_1, \dots, \kappa_{K_\kappa})$. The knots used in this thesis are quantiles of the observed covariate values.

As an example, Figure 1.1 displays the function $(x - \kappa)_+^p$ for $p = 1, 2$, and 3 . The data used is a sequence of 100 numbers in the interval $[0, 1]$ with 10 evenly spaced knots.

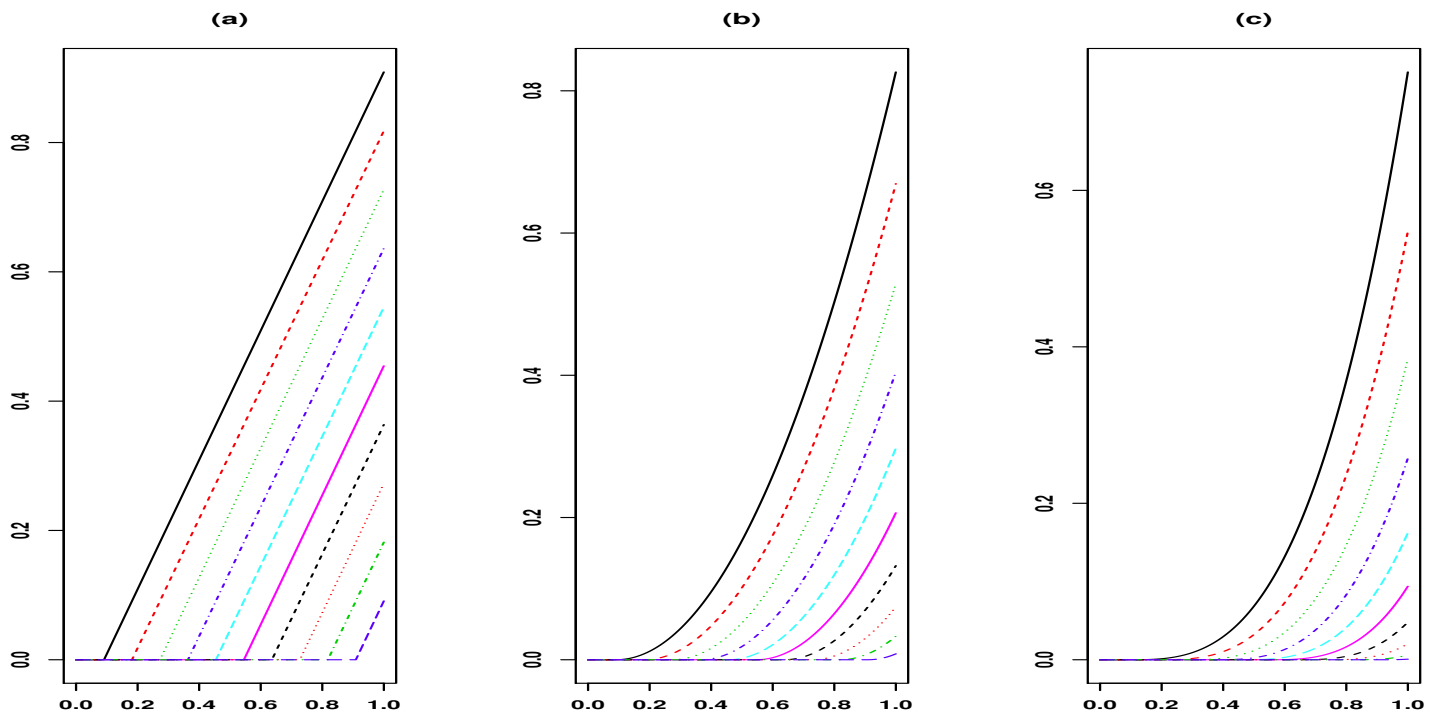


Figure 1.1: (a) Truncated polynomials of degree 1 (lines); (b) truncated polynomials of degree 2 (quadratic basis); (c) truncated polynomials of degree 3 (cubic basis).

1.2 Time Series

A time series is a sequence of random variables $\{Y_t\}$ at time points $t = 0, \pm 1, \pm 2, \pm 3, \dots$, which can also be viewed as a stochastic process. The expected value of the process at time t is given by

$$E(Y_t) = \mu_t, \quad t = 0, \pm 1, \pm 2, \pm 3, \dots,$$

and its autocovariance function, $\gamma_{t,s}$, is defined to be

$$E[(Y_t - \mu_t)(Y_s - \mu_s)] = \text{Cov}(Y_t, Y_s) = \gamma_{t,s},$$

for all time points $t, s = 0, \pm 1, \pm 2, \pm 3, \dots$. For a time series, analysis can be done in either the time domain or in the frequency domain. When we speak of *time domain* analysis, data are analyzed over a period of time. The time unit can be seconds, minutes, hours, etc.

The monthly values of the Southern Oscillation Index (SOI) are an example of a time series, see Figure 1.2. In Chapter 5, more details will be given for this particular time series.

Correlation properties of a time series are usually the basis for the analysis in the time domain. Parametric models that we wish to fit to a time series take this correlation into account. For example, consider the process $\{Y_t\}$. If we wish to fit the following autoregressive model of order 2

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \tag{1.2}$$

then the data y_1, \dots, y_n are used to estimate the parameters ϕ_1 and ϕ_2 . The error terms e_1, \dots, e_n are assumed independent, identically distributed random variables with mean zero and variance σ_e^2 .

In *frequency domain* analysis, a time series is analyzed based on its frequency properties. Consider the cosine wave

$$Y_t = R \cos(2\pi\omega t + \Phi) + e_t, \tag{1.3}$$

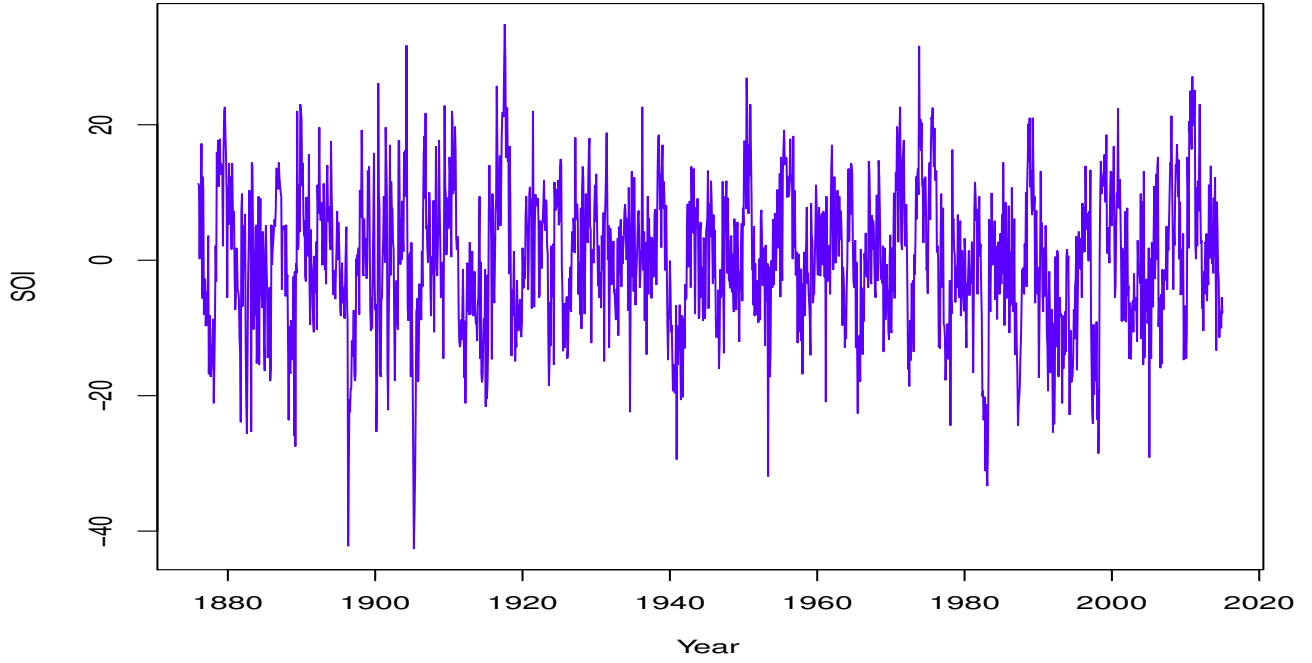


Figure 1.2: Monthly values of the Southern Oscillation Index (SOI).

where the amplitude of the cosine function (1.3) is $R > 0$, its frequency is ω , and Φ is its phase. The period of the wave, $1/\omega$, is the time it takes to complete a cycle. Model (1.3) is not very convenient in estimation since it is not linear in Φ . Using a Trigonometric identity, model (1.3) can be rewritten as

$$Y_t = A \cos(2\pi\omega t) + B \sin(2\pi\omega t) + e_t, \quad (1.4)$$

where $R = \sqrt{A^2 + B^2}$ and $\Phi = \arctan(-B/A)$, and conversely, $A = R \cos(\Phi)$ and $B = -R \sin(\Phi)$. For a fixed frequency ω , $\cos(2\pi\omega t)$ and $\sin(2\pi\omega t)$ are used as predictor variables and the A and B are estimated using ordinary least squares. In Chapter 4, we will see how this cosine wave helps motivate spectral analysis.

The thesis is structured as follows: Chapter 2 introduces penalized splines in a Bayesian approach and the core model of the thesis. Chapter 3 covers spatially adaptive non-parametric regression. In Chapter 4, we discuss spectral time series analysis and define

the model used for spectral analysis. Finally, in Chapter 5 we apply the adaptive model of Chapter 4 to a real data set.

Chapter 2

Nonparametric Regression

Unlike linear regression, nonparametric regression assumes that f is a smooth function, but not necessarily linear. Thus, the data analyst is not restricted by a pre-specified shape of the regression function. Consider the regression model

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2), \quad (2.1)$$

where $f(\cdot)$ is an unknown function. The goal of nonparametric regression is to estimate the function $f(\cdot)$.

2.1 Penalized Splines

Given a set of data $\{(x_1, y_1), \dots, (x_n, y_n)\}$, consider again model (2.1) where

$$f(x_i) = \beta_0 + \beta_1 x_i + \dots + \beta_p x_i^p + \sum_{j=1}^{K_\kappa} b_j (x_i - \kappa_j)_+^p. \quad (2.2)$$

In (2.2) $p > 0$ and $\{\kappa_j\}_{j=1}^{K_\kappa}$ are ordered fixed knots. Let $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$ and $\mathbf{b} = (b_1, \dots, b_{K_\kappa})'$ be the vectors of unknown parameters. Also, let

$$X = \begin{bmatrix} 1 & x_1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^p \end{bmatrix}, \quad Z = \begin{bmatrix} (x_1 - \kappa_1)_+^p & \dots & (x_1 - \kappa_{K_\kappa})_+^p \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+^p & \dots & (x_n - \kappa_{K_\kappa})_+^p \end{bmatrix}. \quad (2.3)$$

In addition, define $T = [X, Z]$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{b}')'$, and let $\mathbf{y} = (y_1, \dots, y_n)'$.

In ordinary linear regression, the estimators are given by

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \|\mathbf{y} - T\boldsymbol{\theta}\|^2 \right\} \quad (2.4)$$

and the regression fit is $\hat{\mathbf{y}} = T\hat{\boldsymbol{\theta}}$. However, the least-squares method will usually overfit the data, resulting in a rough estimate. To facilitate a smoother estimate, consider the minimization

$$\min_{\boldsymbol{\theta}} \left\{ \|\mathbf{y} - T\boldsymbol{\theta}\|^2 + \lambda \mathbf{b}'\mathbf{b} \right\}. \quad (2.5)$$

In (2.5), the roughness penalty term, $\lambda \mathbf{b}'\mathbf{b}$, leads to shrinking \mathbf{b} towards zero, thus resulting in a smoother fit compared to the one based on (2.4). The smoothing parameter λ controls the amount of smoothing. The larger λ , the smoother the resulting fit.

2.2 Linear Mixed Models

Mixed model methodology is used widely in applications such as longitudinal studies. In linear regression, one assumes that $f(x)$ depends linearly on x , i.e., $f(x) = \beta_0 + \beta_1 x$. The unknown parameters β_0 and β_1 are commonly estimated via the method of least squares. In Section 2.3 we will see that penalized splines can be formulated as linear mixed models. In this section we give some background on LMMs.

Observations are collected into groups or clusters in the mixed model setting. If we take for example longitudinal data, observations are collected repeatedly over time for individual subjects. These groups of data for individual subjects are independent, but usually correlated within-subjects. Two sources of variation are thus present, within groups and between groups. Accounting for within-subject correlation is one challenge in longitudinal data analysis which can be tackled by mixed models.

Consider the following study on pig weights over a period of nine weeks (Ruppert et al., 2003). Figure 2.1 shows the measurements pertaining to 48 pigs. Lines are drawn, connecting the measurements that belong to the same pig. Denote the weight of the i th pig in the j th week by $weight_{ij}$ and let $week_j = j$ be the week in which the measurements for a pig are recorded. If we consider the data as cross-sectional (i.e., there is only a single time measurement for a given pig rather than repeated time measurements), then we can

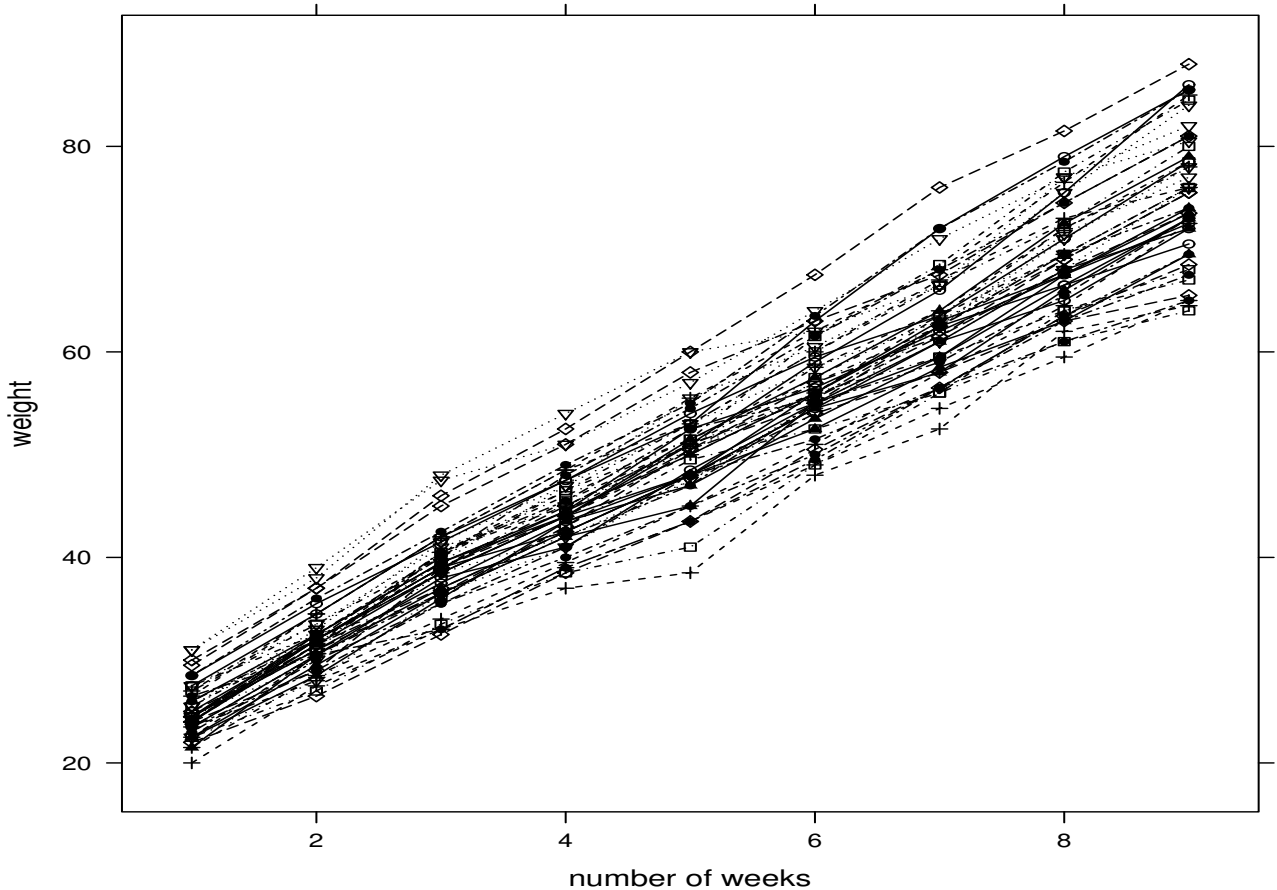


Figure 2.1: Lines connecting points belonging to the same pig.

fit an ordinary least squares model to the data,

$$weight_{ij} = \beta_0 + \beta_1 week_j + \epsilon_{ij}, \quad 1 \leq i \leq 48, \quad 1 \leq j \leq 9, \quad (2.6)$$

where the ϵ_{ij} are independent and identically distributed random variables (i.i.d) from $N(0, \sigma_\epsilon^2)$. However, model (2.6) does not account for the within-pig correlation of weight measurements. One solution to this drawback is to add an individual intercept α_i , for each pig i so that,

$$weight_{ij} = \alpha_i + \beta_1 week_j + \epsilon_{ij}. \quad (2.7)$$

Adding this extra parameter to each individual pig improves the estimate of the slope β_1 ,

but the practicality of model (2.7) is reduced, due to the large number of parameters. In addition, too much credence is placed on this random sample of pigs. We need to take into account the fact that this sample is only a subset of a broader population. A solution is to add to (2.6) random intercepts, b_1, \dots, b_{48} ,

$$weight_{ij} = \beta_0 + b_i + \beta_1 week_j + \epsilon_{ij}, \quad (2.8)$$

where the $b_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$ with variance component $\sigma_b^2 > 0$. The b_i 's are an example of a random effect that helps explain the randomness of other pig samples and helps account for the correlations of weight measurements within individual pigs.

2.3 The Linear Mixed Model

The linear mixed model can be generalized and rewritten in a compact form. This leads to

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}. \quad (2.9)$$

The expected value and variance-covariance matrix of the random vectors in expression (2.9) are given by

$$E \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \text{Cov} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} G & \mathbf{0} \\ \mathbf{0} & R \end{bmatrix}, \quad (2.10)$$

respectively. In the pig example, $G = \sigma_b^2 I_{K_\kappa}$, and $R = \sigma_\epsilon^2 I_n$ where I_{K_κ} is a $K_\kappa \times K_\kappa$ identity matrix, I_n is a $n \times n$ identity matrix, and σ_b^2 and σ_ϵ^2 are positive constants. Estimation of the fixed and random effects can be done by making the following distributional assumptions (Robinson, 1991), on $(\mathbf{y}|\mathbf{b})$ and \mathbf{b} ,

$$(\mathbf{y}|\mathbf{b}) \sim N(X\boldsymbol{\beta} + Z\mathbf{b}, R), \quad \mathbf{b} \sim N(\mathbf{0}, G).$$

The joint density for (\mathbf{y}, \mathbf{b}) is then

$$\begin{aligned} p(\mathbf{y}, \mathbf{b}) &= p(\mathbf{y}|\mathbf{b})p(\mathbf{b}) \\ &\propto \exp \left\{ -\frac{1}{2}(\mathbf{y} - T\boldsymbol{\theta})'R^{-1}(\mathbf{y} - T\boldsymbol{\theta}) - \frac{1}{2}\mathbf{b}'G^{-1}\mathbf{b} \right\}. \end{aligned} \quad (2.11)$$

One way of deriving the estimators uses Henderson's justification (Henderson, 1950). Maximizing the joint density (2.11) with respect to $\boldsymbol{\beta}$ and \mathbf{b} , is equivalent to minimizing the criterion

$$(\mathbf{y} - T\boldsymbol{\theta})'R^{-1}(\mathbf{y} - T\boldsymbol{\theta}) + \mathbf{b}'G^{-1}\mathbf{b}, \quad (2.12)$$

which leads to the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{b}')'$. Moreover, we can easily express (2.12) as follows:

$$(\mathbf{y} - T\boldsymbol{\theta})'R^{-1}(\mathbf{y} - T\boldsymbol{\theta}) + \boldsymbol{\theta}'H\boldsymbol{\theta}, \quad H = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G^{-1} \end{bmatrix}. \quad (2.13)$$

By differentiating the the above expression and equating the derivative to zero, we can solve for $\boldsymbol{\theta}$ and write its BLUP as

$$\hat{\boldsymbol{\theta}} = (T'R^{-1}T + H)^{-1}T'R^{-1}\mathbf{y}. \quad (2.14)$$

The fitted values are therefore $\text{BLUP}(\mathbf{y}) = X\hat{\boldsymbol{\beta}} + Z\hat{\mathbf{b}} = T\hat{\boldsymbol{\theta}}$.

Note that criterion (2.12) is similar to the penalized spline criterion (2.5). Dividing (2.5) by σ_ϵ^2 , leads to spline criterion as

$$\begin{aligned} &= \frac{1}{\sigma_\epsilon^2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 + \frac{\lambda}{\sigma_\epsilon^2} \mathbf{b}'\mathbf{b} \\ &= \frac{1}{\sigma_\epsilon^2} (\mathbf{y} - T\boldsymbol{\theta})'(\mathbf{y} - T\boldsymbol{\theta}) + \frac{\lambda}{\sigma_\epsilon^2} \mathbf{b}'\mathbf{b}. \end{aligned} \quad (2.15)$$

Comparing (2.12) to (2.15), it becomes clear that for the P-splines, $R = \sigma_\epsilon^2 I_n$ and $G = \frac{\sigma_\epsilon^2}{\lambda} I_n \equiv \sigma_b^2 I_{K_\kappa}$ with $\lambda = \sigma_\epsilon^2 / \sigma_b^2$. It is evident that the penalized spline criterion for a spline is exactly the BLUP criterion for a mixed model. Thus, P-splines can be written as mixed models with a smoothing parameter λ . In the next section we will see how to perform the estimation in a Bayesian setting.

2.4 Bayesian Nonparametric Regression

In this section, a Bayesian approach to penalized splines is examined. The Bayesian philosophy in statistics is based on the practice of treating parameters as random variables. By

adopting this philosophy in mixed models, all parameters are treated as random variables. These include the variance components and the fixed effects of the mixed model.

The main characteristics of Bayesian analysis are

1. Placing prior distributions on the unknown parameters before looking at the data;
2. All inference is based on the posterior distribution of all the parameters.
3. Bayesian inference is usually implemented by MCMC methods.

In particular, Bayesian inference is based on a set of unknown parameters, say,

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_N).$$

Any prior beliefs or characterizations of the parameters can be modeled by a probability density function $p(\boldsymbol{\theta})$. Consider now a vector of observed data

$$\mathcal{X} = (X_1, \dots, X_n)$$

with a probability distribution that depends on the parameter vector $\boldsymbol{\theta}$. The likelihood $\mathcal{L}(\mathcal{X}|\boldsymbol{\theta})$ represents the relationship between the parameter vector and the observed data and we think of $\mathcal{L}(\mathcal{X}|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$. From Bayes Theorem, the posterior distribution is given by

$$p(\boldsymbol{\theta}|\mathcal{X}) = \frac{\mathcal{L}(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (2.16)$$

Under a squared loss function, the Bayes estimator of $\boldsymbol{\theta}$ is the posterior mean $E(\boldsymbol{\theta}|\mathcal{X})$. In most cases, the posterior mean is mathematically intractable. MCMC methods facilitate multidimensional integration by simulating from the posterior distribution

$$p(\boldsymbol{\theta}|\mathcal{X}) \propto \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (2.17)$$

and then computing summary statistics of the distribution.

2.5 Priors

In this section, the priors for the the mixed model are described. As before, P-splines can be represented as the mixed model,

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}.$$

Let $(\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2)$ be the vector of the fixed effects, random effects, and variance components. For a fully Bayesian approach, prior distributions are placed on $(\boldsymbol{\beta}, \mathbf{b}, \sigma_b^2, \sigma_\epsilon^2)$. The priors are as follows:

1. The prior placed on $\boldsymbol{\beta}$ is $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2 I_{p+1})$.
2. The prior placed on \mathbf{b} is $\mathbf{b} \sim N(\mathbf{0}, \sigma_b^2 I_{K_\kappa})$.

The priors places on σ_ϵ^2 and σ_b^2 are inverse gamma priors with pdf $p(x) \propto x^{-(a+1)}e^{-b/x}$ for $x > 0$, $a > 0$, and $b > 0$. We then write

3.

$$\sigma_\epsilon^2 \sim IG(A_\epsilon, B_\epsilon), \quad A_\epsilon > 0, \quad B_\epsilon > 0. \quad (2.18)$$

4.

$$\sigma_b^2 \sim IG(A_b, B_b), \quad A_b > 0, \quad B_b > 0. \quad (2.19)$$

These prior specifications for the variance components are sensitive to the selection of the hyperparameters $A_\epsilon, B_\epsilon, A_b$, and B_b . Selecting values for these parameters must be addressed carefully since different values may lead to different results.

2.6 Sampling Scheme

In the context of the mixed model methodology, let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{b}')'$ be the parameter vector containing the fixed and random effects. The posterior distribution for the mixed model is given by

$$p(\boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2, \sigma_b^2 | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) p(\sigma_\epsilon^2) p(\mathbf{b} | \sigma_b^2) p(\sigma_b^2) p(\boldsymbol{\beta}).$$

Gibbs sampling is one of the main MCMC methods. A Gibbs sampling algorithm is implemented by considering the full conditional posterior distributions for the individual parameters. The full conditional posterior distribution for $\boldsymbol{\theta}$ is expressed as follows:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}, \sigma_\epsilon^2, \sigma_b^2) &\propto \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 - \frac{1}{2\sigma_b^2} \|\mathbf{b}\|^2 - \frac{1}{2\sigma_\beta^2} \|\boldsymbol{\beta}\|^2 \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \left(\|\mathbf{y} - T\boldsymbol{\theta}\|^2 + \frac{\sigma_\epsilon^2}{\sigma_b^2} \|\mathbf{b}\|^2 \right) - \frac{1}{2\sigma_\beta^2} \|\boldsymbol{\beta}\|^2 \right\}. \end{aligned} \quad (2.20)$$

Equation (2.20) can be further expanded by collecting terms and completing the square. The result is a multivariate normal distribution, $N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$, where

$$\boldsymbol{\mu}_\theta = (T'T + \sigma_\epsilon^2 D^{-1})^{-1} T'\mathbf{y}, \quad \boldsymbol{\Sigma}_\theta = \sigma_\epsilon^2 (T'T + \sigma_\epsilon^2 D^{-1})^{-1}, \quad (2.21)$$

where $D = \text{diag}(\sigma_\beta^2, \dots, \sigma_\beta^2, \sigma_b^2, \dots, \sigma_b^2)$.

The full conditional distributions for the variance components σ_ϵ^2 and σ_b^2 are inverse gamma distributions, i.e.,

$$(\sigma_\epsilon^2 | \mathbf{y}, \boldsymbol{\beta}, \mathbf{b}) \sim IG \left(\frac{n}{2} + A_\epsilon, \frac{1}{2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 + B_\epsilon \right) \quad (2.22)$$

and

$$(\sigma_b^2 | \mathbf{b}) \sim IG \left(\frac{K_\kappa}{2} + A_b, \frac{1}{2} \|\mathbf{b}\|^2 + B_b \right). \quad (2.23)$$

The Gibbs sampler is used to sample from $p(\boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2, \sigma_b^2 | \mathbf{y})$ by sampling from the full conditional distributions presented above. The sampling scheme for the Bayesian approach to P-splines iterates over the following steps:

1. Sample the parameter vectors $\boldsymbol{\beta}$ and \mathbf{b} jointly as $\boldsymbol{\theta}$ from the multivariate normal distribution

$$N \left\{ (T'T + \sigma_\epsilon^2 D^{-1})^{-1} T'\mathbf{y}, \sigma_\epsilon^2 (T'T + \sigma_\epsilon^2 D^{-1})^{-1} \right\}.$$

2. Sample σ_ϵ^2 from the inverse gamma distribution, $IG(A_\epsilon + \frac{1}{2}n, B_\epsilon + \frac{1}{2}\|\mathbf{y} - T\boldsymbol{\theta}\|^2)$.
3. Sample σ_b^2 from the inverse gamma distribution, $IG(A_b + \frac{1}{2}K_\kappa, B_b + \frac{1}{2}\|\mathbf{b}\|^2)$.

2.7 Application

A brief application (Ruppert et al., 2003) is presented in this section. The LIDAR data consist of $n = 221$ observations. The covariate, *range*, is the distance that light travels before coming back to its source, and the dependent variable is *log ratio*, which is the logarithm of the ratio of light received from two laser sources. Again, assume

$$y_i = f(x_i) + \epsilon_i.$$

The above sampling scheme was applied to the LIDAR data with 10,000 iterations, the first 2000 of which were used as a burn-in period. Estimates using basis functions of degree $p = 1$ and 2 are displayed. The number of knots for this application is $K_\kappa = 40$. Pointwise 95% credible intervals are also shown for each estimate (see Figure 2.2). The credible intervals are calculated by computing the 2.5th percentile and 97.5th percentile of the MCMC iterations of $T\hat{\theta}$ at each covariate value. From the plot, both estimates provide a good fit for the data.

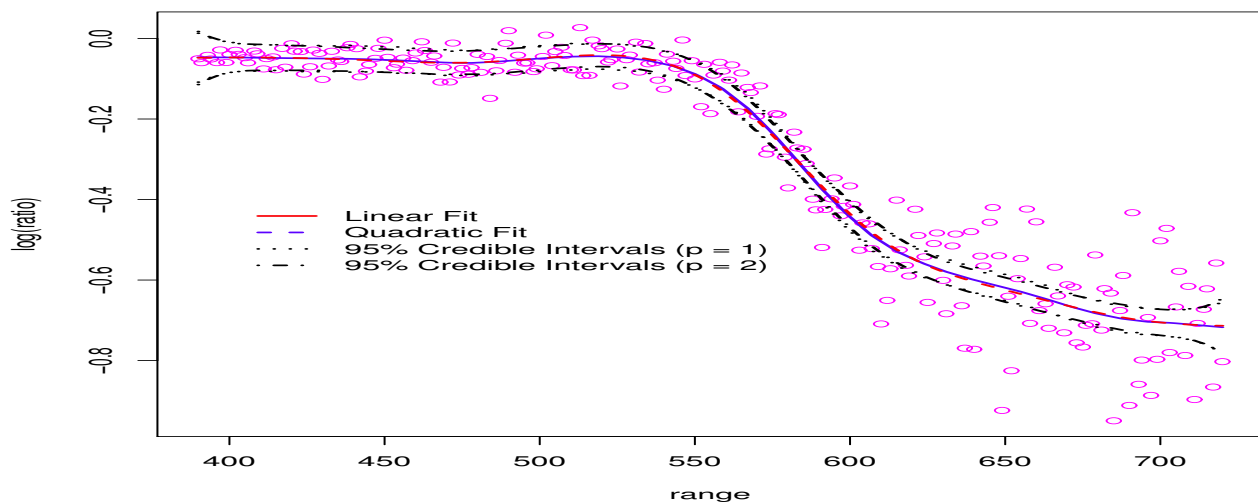


Figure 2.2: P-spline estimates of degrees $p = 1$ (solid line) and $p = 2$ (dashed line) to the LIDAR data along with pointwise 95% credible intervals.

Chapter 3

Spatially Adaptive Smoothing

3.1 Literature Review

There is a large literature on spatially adaptive smoothing. Ruppert and Carroll (2000) achieve spatial adaptability by using local smoothing parameters on the difference penalties of the regression coefficients. They then model the logarithm of the penalties as a linear spline. Denison et al. (1998) proposed a model using piecewise polynomials to estimate a function $f(x)$. They fit low order polynomials that are non zero between knots whose locations are selected adaptively using MCMC methods. Their method fared well compared to the wavelet methods of Donoho and Johnstone (1994).

More recently, DiMatteo et al. (2001) proposed adaptive models using free-knot splines. Their method is referred to as Bayesian adaptive regression splines (BARS). It incorporates knots that are located within the range of the data and assumes $f(x)$ is a cubic regression spline constructed using cubic B-spline basis functions. The prior placed on the number of knots is Poisson. DiMatteo et al. (2001) show that the mean square error (MSE) of their method is significantly smaller compared to the method of Denison et al. (1998).

Baladandayuthapani et al. (2005) model the smoothing parameter as a function of the covariate. The same idea is used by Crainiceanu et al. (2007). A set of knots $\kappa_1, \dots, \kappa_{K_\epsilon}$ is fixed over the range of x , with spline coefficients $a_i, \dots, a_{K_\epsilon}$. More specifically, the smoothing parameter is modeled as a P-spline of degree d constructed using truncated polynomials. Their method was compared to that of Ruppert and Carroll (2000) who computed the average mean square error (AMSE) of over 100 simulations. Even though Crainiceanu et al. (2007)'s method performed slightly better than that of Ruppert and

Carroll (2000), it was outperformed by the former when applied to the family of Doppler functions:

$$f(x) = \sqrt{x(1-x)} \sin \left\{ \frac{2\pi(1 + 2^{(9-4j)/5})}{x + 2^{(9-4j)/5}} \right\},$$

where j determines the severity of the oscillations in the function. The AMSE reported for the Doppler function with $j = 3$ was 0.0011 and 0.0012 for non-spatially adaptive and spatially adaptive smoothing parameters, respectively. Knot selection schemes are also used often for spatially adaptive estimations. For example, Zhou and Shen (2001) presented an algorithm that selects the optimal candidate knots for a regression spline. The algorithm is an improvement over traditional stepwise knot selection schemes by relocating knots closer to optimal knots, searching with more frequency intervals that need more knots, and using local spline regression to reduce computation. Several functions were used for testing the method of Zhou and Shen (2001), in addition to an application to a signal processing data set. Results demonstrated that the knot selection was a good estimate for the signal processing example.

In the frequentist realm, another method involving free-knot splines is that of Miyata and Shen (2003). These authors propose a knot selection scheme where estimation is performed using an evolutionary algorithm to find the optimal tuning parameter for the cubic B-spline representation of the model. This method is then compared to other methods in the literature such as that of Denison et al. (1998), DiMatteo et al. (2001) and Zhou and Shen (2001). The method of Miyata and Shen (2003) performs relatively well across all their simulations. However, when testing it method for the “hump” function, $h(x) = \sin(x) + 2 \exp(-30x^2)$, $x \in [-2, 2]$, the method manages to smooth the right tail of the function but undersmooths its left tail. In Section 3.6, we will see this function again for the simulation study presented in this chapter using the BAPS model.

Local bandwidth selection methods are also used successfully in spatially adaptive smoothing by Staniswalis (1989). This method estimates the optimal local bandwidth, $b_L(x)$, by firstly providing a consistent estimator for the MSE of the kernel estimates, and then minimizing this estimator using $\hat{b}_L(x)$. Additionally, asymptotic normality of the

estimate of $b_L(x)$ is proved. This method proved to be superior to the general global bandwidth method also presented in her study. The method involving local bandwidth selection provides adequate smoothing in areas where a global bandwidth undersmooths the simulation data. A hybrid nonparameteric model is additionally presented by Staniswalis and Yandell (1992). They combine properties of cubic smoothing splines and adaptive kernel estimators to achieve spatially adaptive estimates. The data are first smoothed using a local bandwidth kernel estimator, followed by applying a penalized likelihood to compute a global fit to the presmoothed data.

Following the hybrid methodology, a model combining characteristics of adaptive regression splines and smoothing splines is proposed by Lou and Wahba (1997). This procedure is called hybrid adaptive splines (HAS). It uses a forward stepwise regression procedure to select the basis functions from a set of basis, and then uses the selected basis functions in penalized regression. The method is compared to the wavelet methods of Donoho and Johnstone (1995). Among the seven examples presented in their study, the HAS procedure for spatial adaptivity and the SUREShrink method of Donoho and Johnstone (1995) provide good estimates to examples 1-5. For examples 6-7 a non-adaptive method outperformed both the HAS and the SUREShrink methods.

Another class of spatially adaptive methods involves mixing splines. Wood et al. (2002) presents a Bayesian model that is a mixture of smoothing splines, each having its own smoothing parameter over the domain of the covariate. The mixture model uses individual weights that depend on the covariate value x . What allows the model to be spatially adaptive is fact that the weights vary over the covariate space. Selecting the appropriate number of spline components is done via a modified Bayesian information criterion (BIC). The mixture-of-splines estimate was similar in performance to the methods of Denison et al. (1998), Lou and Wahba (1997), and Smith and Kohn (1996). In some cases the mixture-of-splines estimate performed better than the other methods across the four test functions that were considered.

3.2 Bayesian Adaptive Penalized Splines

The trade-off between bias and variance is controlled by the smoothing parameter λ as was mentioned in Chapter 2. However, for functions with varying oscillations or functions with discontinuities, a single smoothing parameter is inadequate (Scheipl and Kneib, 2009). In this chapter, we introduce spatially adaptive smoothing. Allowing λ to be locally adaptive improves the accuracy of inference and reduces the mean squared error (Ruppert and Carroll, 2000). The following method is referred to as Bayesian Adaptive Penalized Splines (BAPS).

3.3 Model

Consider the model

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

where

$$f(x_i) = \beta_0 + \beta_1 x_i + \cdots + \beta_p x_i^p + \sum_{j=1}^{K_\kappa} b_j (x_i - \kappa_j)_+^p. \quad (3.1)$$

In Chapter 2, $\mathbf{b} \sim N(\mathbf{0}, \sigma_b^2 I_{K_\kappa})$, i.e., σ_b^2 is common to all the b_j , $j = 1, \dots, K_\kappa$. To make the model spatially adaptive, Yue et al. (2012) proposed spatially adaptive precisions δ_j ,

$$(b_j | \delta_j) \stackrel{ind}{\sim} N(0, \delta_j^{-1}), \quad j = 1, \dots, K_\kappa. \quad (3.2)$$

Moreover, define δ_j as

$$\delta_j = \delta \exp(\gamma_j), \quad (3.3)$$

where δ is a scale parameter.

Let $\iota_1, \dots, \iota_{K_\iota}$ be a second layer of knots that covers the range of the knots $\kappa_1, \dots, \kappa_{K_\kappa}$. The model proposed by Yue et al. (2012) then specifies γ_j as

$$\gamma_j = \sum_{k=1}^{K_\iota} b_{\gamma k} I(\kappa_j \geq \iota_k), \quad j = 1, \dots, K_\kappa, \quad (3.4)$$

i.e., as a piecewise constant or

$$\gamma_j = \sum_{k=0}^{K_\iota} b_{\gamma k}(\kappa_j - \iota_k)_+, \quad j = 1, \dots, K_\kappa, \quad (3.5)$$

i.e., as truncated lines. A normal homoscedastic prior with precision η is placed on the new random effects $b_{\gamma k}$,

$$(b_{\gamma k}|\eta) \stackrel{iid}{\sim} N(\mathbf{0}, \eta^{-1} I_{K_\iota+q}), \quad (3.6)$$

for $k = 0, \dots, K_\iota$ when $q = 1$ (truncated lines) and for $k = 1, \dots, K_\iota$ when $q = 0$ (piecewise constant). Equations (3.4) and (3.5) are P-splines of degree 0 and 1, respectively.

3.4 Likelihood and Priors

This section is devoted to describing the priors used in the BAPS model. To begin, the prior placed on the fixed effects $\boldsymbol{\beta}$ is the same as in Chapter 2. The density $p(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \tau)$ is normal, i.e.,

$$(\mathbf{y}|\boldsymbol{\beta}, \mathbf{b}, \tau) \sim N(X\boldsymbol{\beta} + Z\mathbf{b}, \tau^{-1} I_n), \quad (3.7)$$

where τ is a precision parameter. The parameter τ has the improper Jeffrey's prior,

$$p(\tau) \propto \frac{1}{\tau}.$$

Let $\xi_1 = \delta/\tau$ and $\xi_2 = \eta/\delta$ be a parametrization of the smoothing parameters. It follows that $\delta = \tau\xi_1$. For a fully Bayesian approach, ξ_1 and ξ_2 need hyper prior specifications. The priors suggested by Yue et al. (2012) follow directly from the work of Liang et al. (2008) and Yue and Speckman (2010) where a Pareto prior is placed on ξ_1 ,

$$p(\xi_1|c) = \frac{c}{(c + \xi_1)^2}, \quad \xi_1 \geq 0, \quad c > 0, \quad (3.8)$$

and an inverse gamma prior with pdf $p(\xi_2|a, b) \propto \xi_2^{-(a+1)} e^{-b/\xi_2}$, $\xi_2 > 0$, $a > 0$, $b > 0$, is placed on ξ_2 . However, Yue et al. (2012) use a Pareto prior on ξ_2 as well. To ease the computation,

the Pareto distributions are expressed as the scale mixture $p(\xi_i|\rho_i) = \rho_i \exp(-\rho_i \xi_i)$ and $p(\rho_i|c_i) = c_i \exp(-c_i \rho_i)$ for $i = 1, 2$, i.e.,

$$p(\xi_i|c) = \int_0^\infty p(\xi_i|\rho_i)p(\rho_i|c)d\rho_i. \quad (3.9)$$

The joint distribution of the b_j 's is normal with zero mean vector and variance-covariance matrix $\delta_j I_{K_\kappa}$ (see Equation (3.2)). Using (3.3) we can write $(\mathbf{b}|\boldsymbol{\gamma}, \tau, \xi_1) \sim N(\mathbf{0}, (\tau \xi_1)^{-1} D_\gamma^{-1})$, where $D_\gamma = \text{diag}(e^{\gamma_1}, \dots, e^{\gamma_{K_\kappa}})$ and $\boldsymbol{\gamma} = Z_\gamma \mathbf{b}_\gamma$ where $\mathbf{b}_\gamma = (b_{\gamma_1}, \dots, b_{\gamma_{K_\kappa}})'$ and

$$Z_\gamma = \begin{bmatrix} (\kappa_1 - \iota_1)_+^q & \dots & (\kappa_1 - \iota_{K_\iota})_+^q \\ \vdots & \ddots & \vdots \\ (\kappa_{K_\kappa} - \iota_1)_+^q & \dots & (\kappa_{K_\kappa} - \iota_{K_\iota})_+^q \end{bmatrix},$$

that is, in the design matrix Z_γ the knots $\{\kappa_j\}_{j=1}^{K_\kappa}$ are treated as covariate values with knots $\{\iota_k\}_{k=1}^{K_\iota}$. The parameter vector $\mathbf{b}_\gamma \sim N(\mathbf{0}, \eta^{-1} I_{K_\iota+q})$. Note that $\eta = \delta \xi_2$ and $\delta = \tau \xi_1$, so $\eta = \tau \xi_1 \xi_2$. The prior on \mathbf{b}_γ can then be written as $(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2) \sim N(\mathbf{0}, (\tau \xi_1 \xi_2)^{-1} I_{K_\iota+q})$. This completes the prior specification for the BAPS model. Yue et al. (2012) show that the choice of priors leads to a proper posterior distribution.

3.5 Sampling Scheme

In this section, we present the sampling scheme for the BAPS model. Let $T = [X, Z]$. To sample from the posterior

$$\begin{aligned} p(\boldsymbol{\beta}, \mathbf{b}, \tau, \xi_1, \xi_2, \rho_1, \rho_2, \boldsymbol{\gamma}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{b}, \tau)p(\tau)p(\boldsymbol{\beta})p(\mathbf{b}|\boldsymbol{\gamma}, \tau, \xi_1)p(\xi_1|\rho_1)p(\xi_2|\rho_2) \\ &\times p(\rho_1|c_1)p(\rho_2|c_2)p(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2). \end{aligned}$$

The full conditional posterior distributions for the individual parameters $(\boldsymbol{\beta}, \mathbf{b}, \tau, \xi_1, \xi_2, \rho_1, \rho_2, \boldsymbol{\gamma})$ are as follows.

1. The parameter vectors $\boldsymbol{\beta}$ and \mathbf{b} are sampled jointly as $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{b}')$ from the multivariate normal distribution, $N(\boldsymbol{\mu}_\theta, Q_\theta)$, where

$$\boldsymbol{\mu}_\theta = \tau Q_\theta T' \mathbf{y},$$

$$Q_{\boldsymbol{\theta}} = (\tau T' T + \Lambda_y)^{-1},$$

and $\Lambda_y = \text{diag}(1/\sigma_{\beta}^2, \dots, 1/\sigma_{\beta}^2, \tau \xi_1 e^{\gamma_1}, \dots, \tau \xi_1 e^{\gamma_{K_{\kappa}}})$.

2. The parameter τ is sampled from the gamma distribution,

$$G\left(\frac{1}{2}(n + K_{\kappa} + K_{\iota} + q), \frac{1}{2}(\|\mathbf{y} - T\boldsymbol{\theta}\|^2 + \xi_1 \mathbf{b}' D_{\gamma} \mathbf{b} + \xi_1 \xi_2 \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma})\right).$$

3. The parameter ξ_1 is sampled from the gamma distribution,

$$G\left(\frac{1}{2}(K_{\kappa} + K_{\iota} + q) + 1, \frac{1}{2}\tau \mathbf{b}' D_{\gamma} \mathbf{b} + \frac{1}{2}\tau \xi_2 \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma} + \rho_1\right).$$

4. The parameter ξ_2 is sampled from the gamma distribution,

$$G\left(\frac{1}{2}(K_{\iota} + q) + 1, \frac{1}{2}\tau \xi_1 \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma} + \rho_2\right).$$

5. The parameter ρ_1 is sampled from the gamma distribution, $G(2, \xi_1 + c_1)$.

6. The parameter ρ_2 is sampled from the gamma distribution, $G(2, \xi_2 + c_2)$.

7. The $b_{\gamma k}$'s are sampled via a Metropolis-Hastings (M-H) step from

$$p(\mathbf{b}_{\gamma} | \boldsymbol{\gamma}, \mathbf{b}, \delta, \eta) \propto |D_{\gamma}|^{1/2} \exp\left\{-\frac{1}{2}\tau \xi_1 \mathbf{b}' D_{\gamma} \mathbf{b} - \frac{1}{2}\tau \xi_1 \xi_2 \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma}\right\}. \quad (3.10)$$

A new value \mathbf{b}_{γ}^* is proposed from $N(\hat{\mathbf{b}}_{\gamma}, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_{\gamma}})$, where $\hat{\mathbf{b}}_{\gamma} = \arg \max_{\mathbf{b}_{\gamma}} \log p(\mathbf{b}_{\gamma} | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)$ and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{b}_{\gamma}} = \left[-\frac{\partial^2 \log p(\mathbf{b}_{\gamma} | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)}{\partial \mathbf{b}_{\gamma} \partial \mathbf{b}'_{\gamma}} \Big|_{\mathbf{b}_{\gamma} = \hat{\mathbf{b}}_{\gamma}} \right]^{-1}.$$

The new value, $\mathbf{b}_{\gamma}^{(t+1)}$, satisfies

$$\mathbf{b}_{\gamma}^{(t+1)} = \begin{cases} \mathbf{b}_{\gamma}^* & \text{with probability } \alpha_{\mathbf{b}_{\gamma}} \\ \mathbf{b}_{\gamma}^{(t)} & \text{with probability } 1 - \alpha_{\mathbf{b}_{\gamma}}, \end{cases}$$

where $\mathbf{b}_{\gamma}^{(t)}$ is the current value. The acceptance probability is

$$\alpha_{\mathbf{b}_{\gamma}} = \min \left\{ \frac{p(\mathbf{b}_{\gamma}^* | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) g(\mathbf{b}_{\gamma}^{(t)} | \mathbf{b}_{\gamma}^*)}{p(\mathbf{b}_{\gamma}^{(t)} | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) g(\mathbf{b}_{\gamma}^* | \mathbf{b}_{\gamma}^{(t)})}, 1 \right\}, \quad (3.11)$$

where $g(\mathbf{b}_{\gamma}^* | \mathbf{b}_{\gamma}^{(t)})$ is the proposal density for \mathbf{b}_{γ}^* , i.e., $N(\hat{\mathbf{b}}_{\gamma}, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_{\gamma}})$.

In Yue et al. (2012)'s method, the $b_{\gamma k}$'s are sampled in blocks (see Yue and Speckman (2010)), rather than jointly. They split the vector \mathbf{b}_γ into \mathcal{V} blocks: $\mathbf{b}_\gamma = (\mathbf{b}_{\gamma 1}, \dots, \mathbf{b}_{\gamma \mathcal{V}})'$. They then use a random walk Metropolis-Hastings step with acceptance probability

$$\alpha_v = \min \left\{ \frac{\pi(\mathbf{b}_{\gamma v}^*)}{\pi(\mathbf{b}_{\gamma v}^{(t)})}, 1 \right\}$$

to update each block where the proposal distribution for the new value $\mathbf{b}_{\gamma v}^*$ is the prior distribution for \mathbf{b}_γ , i.e., $\pi(\mathbf{b}_{\gamma v}) = p(\mathbf{b}_{\gamma v} | \mathbf{b}, \mathbf{b}_{\gamma(-v)}, \boldsymbol{\gamma}, \tau, \xi_1, \xi_2, \boldsymbol{\gamma})$ and $\mathbf{b}_{\gamma(-v)}$ is the vector containing the remaining vectors.

Details on the sampling scheme can be found in Appendix A.2.

3.6 Simulation Study

In this section we conduct Monte Carlo simulations using the BAPS model. Based on the examples presented, we compare the performance of the BAPS model with its non-adaptive version, the Bayesian Penalized Splines (BPS) model. Two different functions taken from Yue et al. (2012) are used for the simulation. One is a smoothly varying function, and the other is a spatially inhomogeneous function. For each setting, we present 10 different simulations. The number of knots for the BAPS and the BPS models is $K_\kappa = 30$ with sub knots $K_\iota = 10$ for the BAPS model. Pointwise 95% credible intervals are also part of the displays. The fixed parameters were set to $c_1 = c_2 = 1$ in the posterior distributions for the parameters ρ_1 and ρ_2 in the BAPS model. For the BPS model $A_\epsilon = A_b = B_\epsilon = B_b = 0$

For the first setting, we consider a natural spline with knots located at the points (0.2, 0.6, 0.7) and coefficients (20, 4, 6, 11, 6). We evaluate the knots and coefficients at $n=101$ equally spaced points on $[0, 1]$. Gaussian noise with mean zero and standard deviation $\tau^{-1/2} = 0.9$ was added to the function values. The degree of the basis functions used for the estimation is $p = 2$. Figures 3.1 and 3.2 show the BAPS estimates and BPS estimates for the natural spline function. Figure 3.1 shows the results corresponding to five simulated samples while Figure 3.2 displays the results for the remaining five simulated

samples. In this case both estimates are good.

In the second setting, we generate data from the function

$$h(x) = \sin(x) + 2 \exp(-30x^2), \quad x \in [-2, 2]. \quad (3.12)$$

The function was evaluated at $n=101$ equally spaced points. Again, Gaussian noise with mean zero and standard deviation $\tau^{-1/2} = 0.3$ was added to the function values. The degree of basis functions used for the estimation for this setting is $p = 2$. From figures 3.3 and 3.4 we can appreciate the performance of the BAPS model. The BPS estimate is unable to flatten out the tails of the function. The BAPS estimate, however, is able to flatten out the tails and comes relatively closer than the BPS model at estimating the hump.

In the next chapter, we will apply the BAPS methodology to spectral time series analysis.

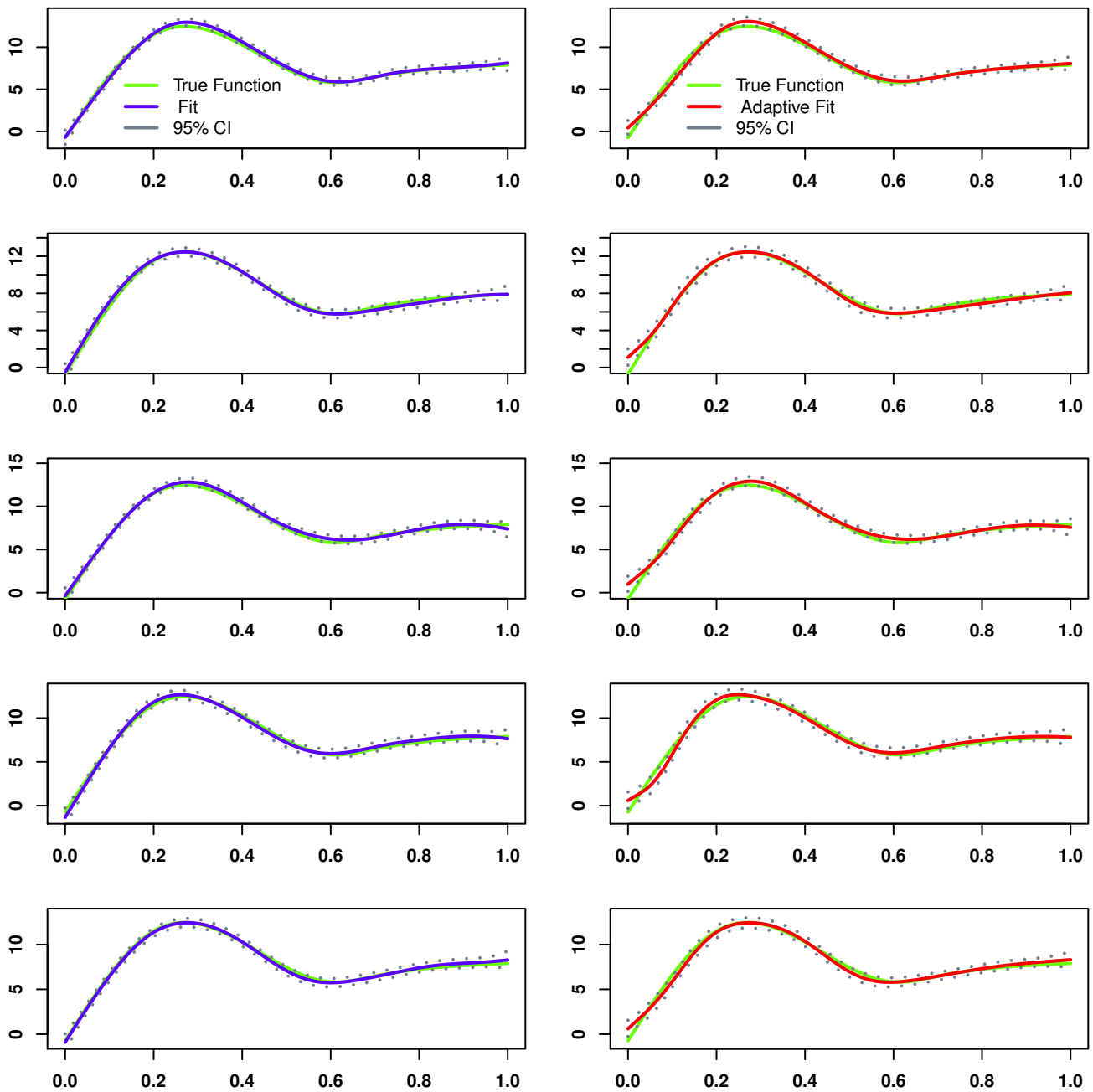


Figure 3.1: BAPS and BPS estimates for the natural spline. The first column displays the BPS estimates; the second column displays the BAPS estimates. In each column, the solid line represents the true function; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible intervals. (First 5 simulations)

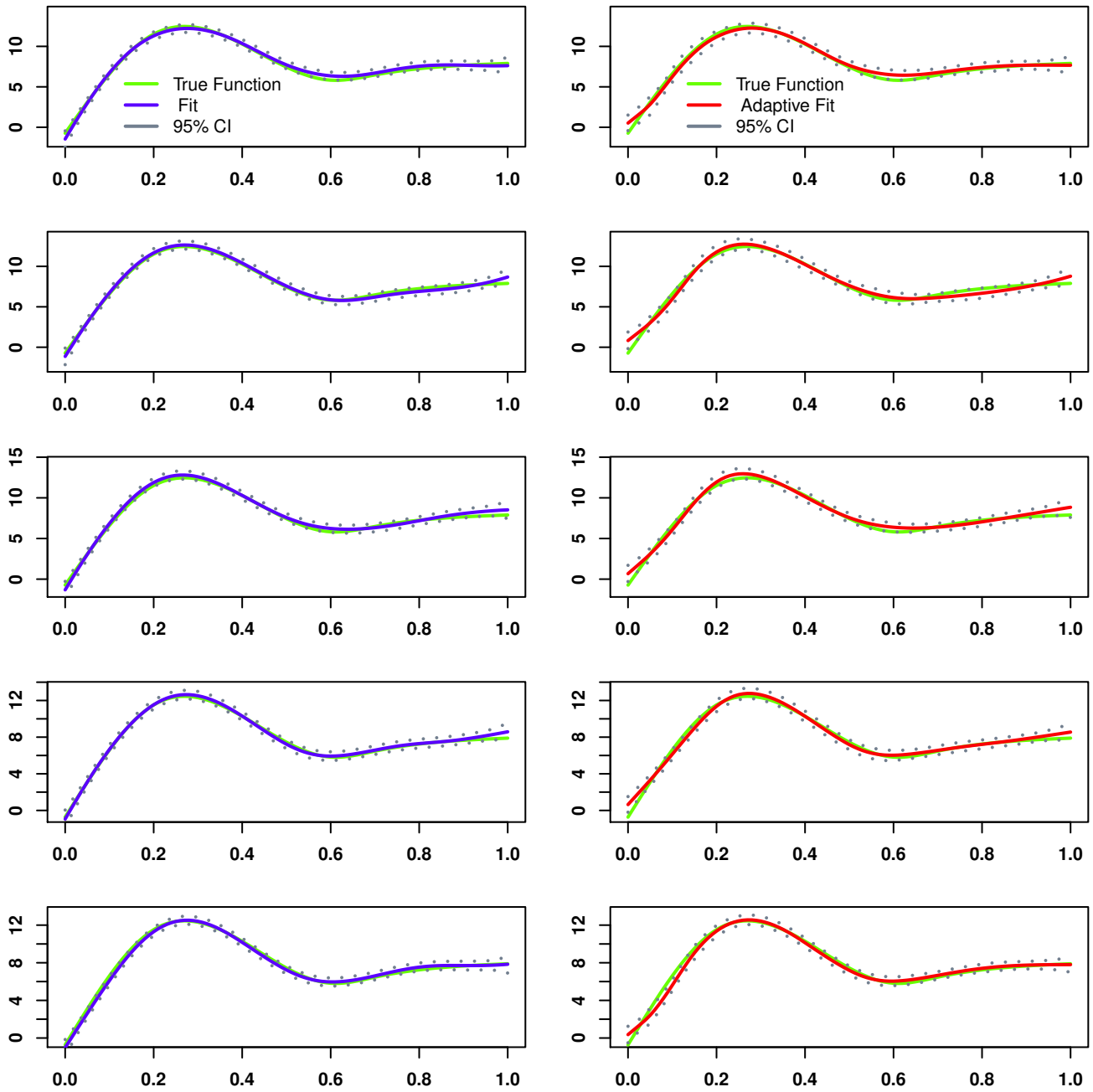


Figure 3.2: BAPS and BPS estimates for the natural spline. The first column displays the BPS estimates; the second column displays the BAPS estimates. In each column, the solid line represents the true function; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible intervals. (Second 5 simulations)

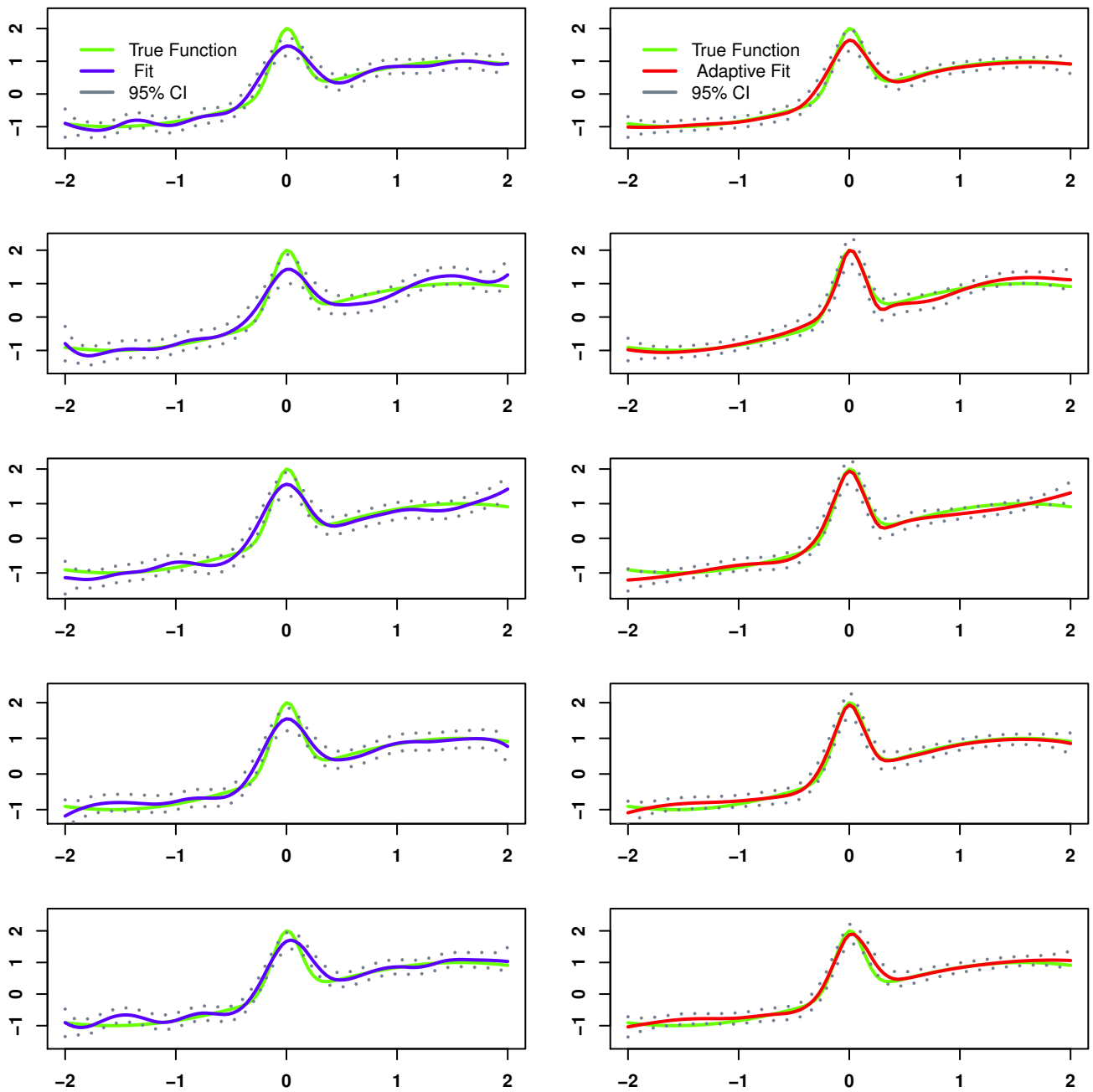


Figure 3.3: BAPS and BPS estimates for the function (3.12). The first column displays the BPS estimates; the second column displays the BAPS estimates. In each column, the solid line represents the true function; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible intervals.(First 5 simulations)

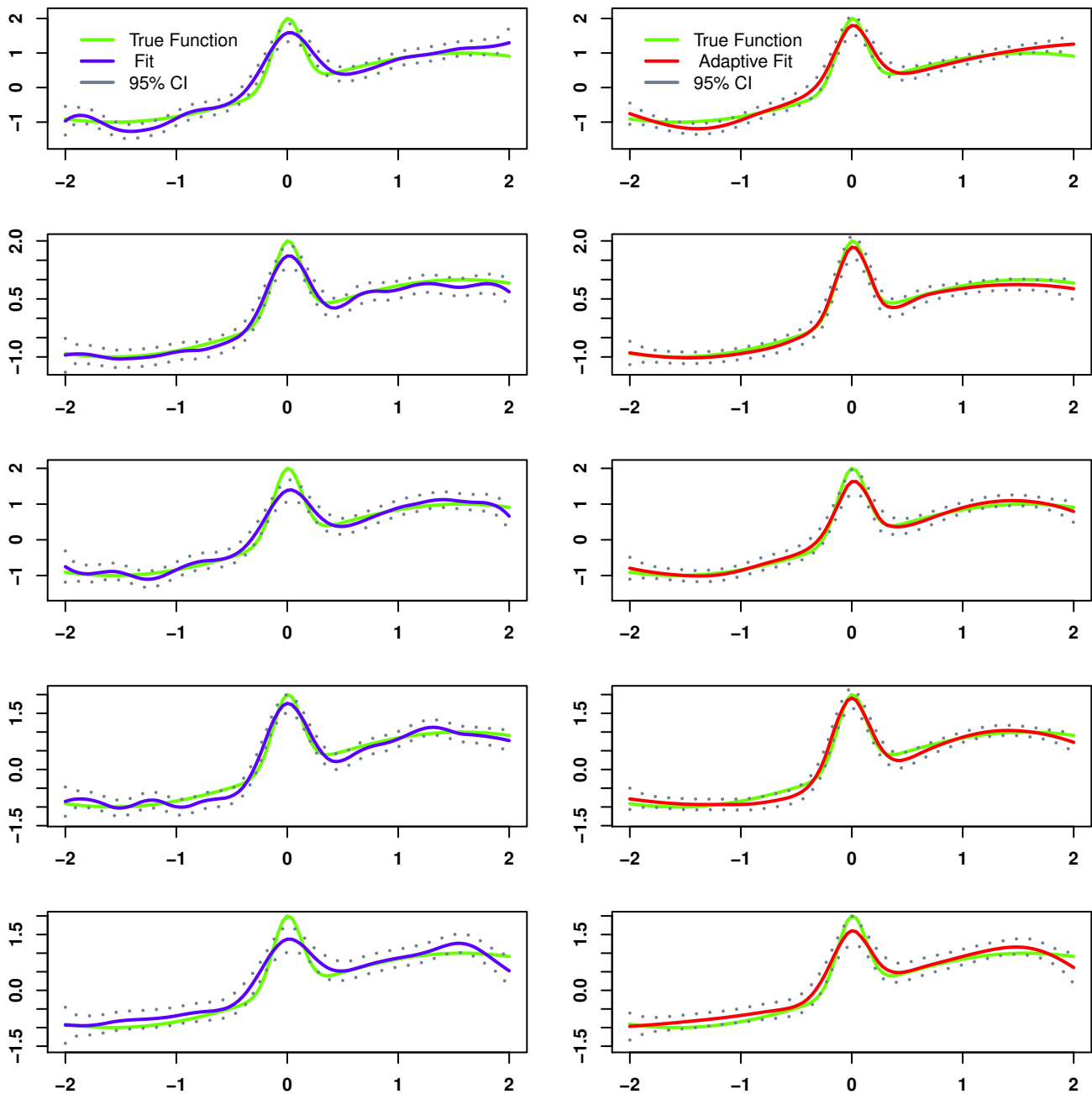


Figure 3.4: BAPS and BPS estimates for the function (3.12). The first column displays the BPS estimates; the second column displays the BAPS estimates. In each column, the solid line represents the true function; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible intervals. (Second 5 simulations)

Chapter 4

Spatially Adaptive Smoothing for Spectral Analysis

A brief introduction to time series was given in Chapter 1. It was mentioned that time series analysis can be done in either the time domain or the frequency domain. In this chapter we focus on the frequency domain.

4.1 Spectral Analysis

4.1.1 Stationarity

The idea of stationarity is that the mean and variance of a time series are constant for all time points t . A process is said to be weakly stationary, if it has a constant mean and

$$\gamma_{t,s} = \text{Cov}(Y_{t-s}, Y_0) = \text{Cov}(Y_0, Y_{t-s}) = \text{Cov}(Y_0, Y_{|t-s|}) = \gamma_{|t-s|}, \quad (4.1)$$

for all time points t, s . Unlike weak stationarity, a process $\{Y_t\}$ is strictly stationary, when the joint distributions for the variables Y_{t_1}, \dots, Y_{t_n} and $Y_{t_1-\ell}, \dots, Y_{t_n-\ell}$ are the same for all time points $t = 1, \dots, n$ and lags ℓ (Cryer and Chan, 2008).

4.1.2 Periodogram

In Chapter 1, we briefly discussed the frequency domain and showed how the cosine wave (1.3) can be expanded using Trigonometric identities. In this section, the saturated cosine model will be introduced. For any time series sample y_1, \dots, y_n , where n is odd, we can

write exactly

$$Y_t = A_0 + \sum_{j=1}^{(n-1)/2} [A_j \cos(2\pi\omega_j t) + B_j \sin(2\pi\omega_j t)] \quad (4.2)$$

for $t = 1, \dots, n$ and suitably chosen coefficients (see Shumway and Soffer (2011)). The coefficients A_j and B_j can be found using regression results. The Fourier frequencies, j/n , $j = 1, \dots, (n-1)/2$ lead to simple expressions for A_j and B_j , since for these frequencies the sines and cosines are orthogonal. These expressions for A_0 , the A_j 's and B_j 's are $\hat{A}_0 = \bar{Y}$,

$$\hat{A}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(2\pi t j/n), \quad \text{and} \quad \hat{B}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(2\pi t j/n), \quad (4.3)$$

respectively. In the case where n is even, the estimates for A_0 , A_j 's and B_j 's are similar.

Given data y_1, \dots, y_n , the discrete Fourier transform (DFT), where $i = \sqrt{-1}$, is defined as

$$d(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t e^{-2\pi i \omega_j t} \quad (4.4)$$

for $j = 0, 1, \dots, n-1$ and $\omega_j = j/n$ (Fourier frequencies). The periodogram is defined as

$$I_n(\omega_j) = \left| d(\omega_j) \right|^2, \quad j = 0, 1, \dots, n-1. \quad (4.5)$$

Let the cosine and sine transforms be defined as

$$d_C(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t \cos(-2\pi\omega_j t) \quad \text{and} \quad d_S(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t \sin(-2\pi\omega_j t), \quad (4.6)$$

receptively, where $\omega_j = j/n$, $j = 0, 1, \dots, n-1$ (Shumway and Soffer, 2011). Then

$$I_n(\omega_j) = d_C^2(\omega_j) + d_S^2(\omega_j). \quad (4.7)$$

It follows that

$$\begin{aligned} I_n(\omega_j) &= \frac{n}{4} [\hat{A}_j^2 + \hat{B}_j^2] \\ &= \frac{1}{n} \left| \sum_{t=1}^n y_t e^{-2\pi i \omega_j t} \right|^2. \end{aligned} \quad (4.8)$$

The overall behavior of time series can then be summarized by identifying the cosine-sine pairs in the time series. The relative strength of the cosine-sine pairs is determined by the heights at the various frequencies $\omega_j = j/n$, $j = 0, 1, \dots, n-1$ (Cryer and Chan, 2008).

4.1.3 The Spectral Density

If the autocovariance function, γ_ℓ , of a stationary process satisfies

$$\sum_{\ell=-\infty}^{\infty} |\gamma_\ell| < \infty, \quad (4.9)$$

then it can be represented as

$$\gamma_\ell = \int_{-1/2}^{1/2} f(\omega) \exp(-2\pi i \ell \omega) d\omega, \quad \ell = 0, \pm 1, \pm 2, \dots \quad (4.10)$$

where the spectral density, $f(\omega)$, has the representation

$$f(\omega) = \sum_{\ell=-\infty}^{\infty} \gamma_\ell \exp(-2\pi i \ell \omega), \quad -1/2 \leq \omega \leq 1/2. \quad (4.11)$$

The spectral density is symmetric over the domain $-1/2 \leq \omega \leq 1/2$, i.e., $f(\omega) = f(-\omega)$, periodic and nonnegative. From (4.10), for $\ell = 0$, $\gamma_0 = \text{Var}(Y_t) = \int_{-1/2}^{1/2} f(\omega) d\omega$. This shows that the total variance of the time series is the integral of $f(\omega)$ over the entire range of frequencies. All mathematical properties of probability density functions can be applied to spectral densities with the exception that the area under the spectral density is $\text{Var}(Y_t)$.

4.2 Model

Let $\omega_m = m/n$, for $m = 1, \dots, M$, be the Fourier frequencies where $M = \lfloor (n-1)/2 \rfloor$. The notation $\lfloor w \rfloor$ represents the largest integer not greater than w . Again, let $C = [X, Z]$ where X denotes the $m \times (p+1)$ matrix whose i th row is $(1, \omega_i, \dots, \omega_i^p)$ and Z is a matrix whose i th row is equal to $((\omega_i - \kappa_1)_+^p, \dots, (\omega_i - \kappa_{K_\kappa})_+^p)$. Given a stationary process $\{Y_t\}$, Whittle (1957) shows that the values $I_n(\omega_m)$ are approximately independent and exponentially distributed with mean $f(\omega_m)$, i.e., the pdf of $I_n(\omega_m)$ is given by

$$p(I_n(\omega_m)) = f(\omega_m)^{-1} \exp(-I_n(\omega_m)/f(\omega_m)). \quad (4.12)$$

Moreover, he shows that for large n the likelihood for $\{Y_t\}$, conditional on f , can be approximated by

$$\begin{aligned}
p(\mathbf{y}|f) &= \prod_{m=1}^M f(\omega_m)^{-1} \exp \{-I_n(\omega_m)/f(\omega_m)\} \\
&= \prod_{m=1}^M \exp \{-\log f(\omega_m) - I_n(\omega_m)/f(\omega_m)\} \\
&= \prod_{m=1}^M \exp \{-\log f(\omega_m) - I_n(\omega_m) \exp[-\log f(\omega_m)]\}.
\end{aligned} \tag{4.13}$$

Let $-\log f(\omega_m) = \mathbf{c}'_m \boldsymbol{\theta}$, where \mathbf{c}'_m is the m th row of C . Then (4.13) becomes

$$p(\mathbf{y}|f) = \exp \left\{ \sum_{m=1}^M [\mathbf{c}'_m \boldsymbol{\theta} - I_n(\omega_m) \exp(\mathbf{c}'_m \boldsymbol{\theta})] \right\}. \tag{4.14}$$

Whittle likelihood has been used before to estimate the spectral density. Early references include Wahba (1980) who modeled the log-spectral density using cubic smoothing splines and Pawitan and O'Sullivan (1994) who used a penalized version of the Whittle likelihood. The next section will be devoted to prior specification on the parameters for this model.

4.3 Priors

In this section, the priors for the Whittle likelihood result in a spatially adaptive estimate of the spectral density, similar to the spatially adaptive estimates in Chapter 3. Firstly, the prior placed on the fixed coefficients $\boldsymbol{\beta}$ remains the same, as in Chapter 2. Recall that in order to make the estimate spatially adaptive, we placed a normal prior on the b_j 's with variance δ_j^{-1} . For spectral smoothing, we make a slight alteration to the precisions. The parameter τ no longer appears in the likelihood, so δ in this chapter is simply a scale parameter, no longer equal to $\tau\xi_1$, as was the case in Chapter 2. Then, let $\delta_j = \delta \exp(\gamma_j)$. The prior placed on \mathbf{b} is

$$(\mathbf{b}|\boldsymbol{\gamma}, \delta) \sim N(\mathbf{0}, \delta^{-1} D_{\boldsymbol{\gamma}}^{-1}),$$

where $D_\gamma = \text{diag}(e^{\gamma_1}, \dots, e^{\gamma_{K_\kappa}})$ and

$$\gamma_j = \sum_{k=1}^{K_\iota} b_{\gamma k} I(\kappa_j \geq \iota_k), \quad j = 1, \dots, K_\kappa,$$

when $q = 0$ (piecewise constant) or

$$\gamma_j = \sum_{k=0}^{K_\iota} b_{\gamma k} (\kappa_j - \iota_k)_+, \quad j = 1, \dots, K_\kappa,$$

when $q = 1$ (truncated lines). Next, as in the BAPS model of Chapter 3, the prior placed on the random effects $b_{\gamma k}$ is normal with mean zero and precision η , i.e.,

$$(\mathbf{b}_\gamma | \eta) \sim N(\mathbf{0}, \eta^{-1} I_{K_\iota + q}).$$

The prior placed on $\boldsymbol{\theta}$ is thus

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\boldsymbol{\beta}) \times p(\mathbf{b}) \\ &\propto \exp \left\{ -\frac{1}{2\sigma_\beta^2} \boldsymbol{\beta}' \boldsymbol{\beta} - \frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} \right\} \\ &= \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta} \right\}, \end{aligned}$$

where $\Lambda = \text{diag}(1/\sigma_\beta^2, \dots, 1/\sigma_\beta^2, \delta e^{\gamma_1}, \dots, \delta e^{\gamma_{K_\kappa}})$. To complete the prior specification, the work of Gelman (2006) is utilized to place priors on the standard deviations $\delta^{-1/2}$ and $\eta^{-1/2}$. The priors on these standard deviations are independent folded t distributions whose density is $p(x) \propto [1 + (x/G)^2/\nu]^{-(\nu+1)/2}$, $x > 0$, where ν and G are fixed parameters. For the simulations in this chapter, we let $\nu = 2$ and $G = 10^5$. The folded t distribution has a scale mixture representation which can be utilized to simplify the sampling scheme. A folded t prior on $\delta^{-1/2}$ is equivalent to the scale mixture

$$(\delta^{-1} | \nu_1, g_1) \sim IG(\nu_1/2, \nu_1/g_1), \quad g_1 \sim IG(1/2, \nu_1/G_1^2).$$

Similarly, a folded t prior on $\eta^{-1/2}$ can be represented by

$$(\eta^{-1} | \nu_2, g_2) \sim IG(\nu_2/2, \nu_2/g_2), \quad g_2 \sim IG(1/2, \nu_2/G_2^2),$$

see Wand et al. (2012).

4.4 Sampling Scheme

The posterior distribution for this model is given by

$$p(\boldsymbol{\beta}, \mathbf{b}, \mathbf{b}_\gamma, \delta, \eta, g_1, g_2 | \mathbf{I}) \propto p(\mathbf{I} | \boldsymbol{\beta}, \mathbf{b}) p(\mathbf{b} | \boldsymbol{\gamma}, \delta) p(\delta | g_1) p(g_1) p(\mathbf{b}_\gamma | \eta) p(\eta | g_2) p(g_2) p(\boldsymbol{\beta}),$$

where $\mathbf{I} = (I_n(\omega_1), \dots, I_n(\omega_M))'$. To draw $(\boldsymbol{\beta}, \mathbf{b}, \mathbf{b}_\gamma, \delta, \eta, g_1, g_2)$ from their joint posterior distribution, MCMC methods are employed as follows:

1. The parameter vectors $\boldsymbol{\beta}$ and \mathbf{b} are sampled jointly as $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{b}')$ via a Metropolis-Hastings step from its conditional posterior distribution whose logarithm is

$$\log p(\boldsymbol{\theta} | \mathbf{I}, C) = \sum_{m=1}^M [\mathbf{c}'_m \boldsymbol{\theta} - I_n(\omega_m) \exp(\mathbf{c}'_m \boldsymbol{\theta})] - \frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta}, \quad (4.15)$$

where \mathbf{c}'_m is the m th row of C .

2. The $b_{\gamma k}$'s are sampled via a Metropolis-Hastings step from

$$p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \delta, \eta) \propto |D_\gamma|^{1/2} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} - \frac{\eta}{2} \mathbf{b}'_\gamma \mathbf{b}_\gamma \right\}. \quad (4.16)$$

3. The parameter δ is sampled from the gamma distribution, $G(\frac{1}{2}(K_\kappa + \nu_1), \frac{1}{2} \mathbf{b}' D_\gamma \mathbf{b} + \frac{\nu_1}{g_1})$.
4. The parameter g_1 is sampled from the inverse gamma distribution,

$$IG \left(\frac{1}{2}(\nu_1 + 1), \nu_1 \delta + \frac{1}{G_1^2} \right).$$

5. The parameter η is sampled from the gamma distribution,

$$G \left(\frac{1}{2}(K_\iota + q + \nu_2), \frac{1}{2} \mathbf{b}'_\gamma \mathbf{b}_\gamma + \frac{\nu_2}{g_2} \right).$$

6. The parameter g_2 is sampled from the inverse gamma distribution,

$$IG \left(\frac{1}{2}(\nu_2 + 1), \nu_2 \eta + \frac{1}{G_2^2} \right).$$

For further details on the derivations of the full conditional posterior distributions for the various parameters and the Metropolis-Hastings steps, refer to Appendix A.3.

4.5 Simulation Study

In this section we explore the model presented via Monte Carlo simulations. Samples are generated from an AR(2) model,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t, \quad e_t \sim N(0, \sigma_e^2).$$

In the first simulation setting, $\phi_1 = 1.5$ and $\phi_2 = -0.75$ while in the second simulation setting, $\phi_1 = 0.1$ and $\phi_2 = 0.4$. The theoretical spectral density for an AR(2) process is

$$f(\omega) = \frac{\sigma_e^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega)}. \quad (4.17)$$

This spectral density may have different shapes depending on the values of ϕ_1 and ϕ_2 . The first setting has peaked spectrum, while the second one has trough spectrum. Ten samples of size 1000 are generated from each setting. Each sample was run for 2000 MCMC iterations with a burn-in period of 500. Two versions of the model were fit to the samples: spatially adaptive and non spatially adaptive. Quadratic basis functions and $K_\kappa = 30$ knots were used for the non-adaptive and adaptive fits. For the adaptive case, $K_\ell = 10$ knots with basis functions of degree $q = 0$ are used. Additionally, the plots also display the “smoothed” periodogram using a third (frequentist) method presented by Cryer and Chan (2008). For the Bayesian methods, pointwise 95% credible intervals are also displayed. For the frequentist methods, pointwise 95% confidence intervals are computed as follows:

$$\log[\bar{f}(\omega)] + \log \left[\frac{df}{\chi_{df, 1-\alpha/2}^2} \right] \leq \log[f(\omega)] \leq \log[\bar{f}(\omega)] \leq + \log \left[\frac{df}{\chi_{df, \alpha/2}^2} \right], \quad (4.18)$$

where $\bar{f}(\omega)$ is the smoothed periodogram given by

$$\bar{f}(\omega) = \sum_{\ell=-a}^a W_a(\ell) I_n \left(\omega + \frac{\ell}{n} \right). \quad (4.19)$$

In equation (4.19),

$$W_a(\ell) = \frac{1}{2a+1}, \quad -a \leq \ell \leq a.$$

The degrees of freedom in equation (4.18) are given by

$$df = \frac{2}{\sum_{\ell=-a}^a W_a^2(\ell)},$$

and $\chi_{df,1-\alpha/2}^2$ and $\chi_{df,\alpha/2}^2$ denote the $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$, percentiles, respectively, of a chi-square distribution on df degrees of freedom.

Figures 4.1 and 4.2 display the first and second five samples, respectively, generated from the AR(2) process with $\phi_1 = 1.5$ and $\phi_2 = -0.75$. The left column in each figure displays the non-adaptive estimates, the middle column displays the adaptive estimates, and the last column displays the estimates of the frequentist method. Each plot shows the true log spectral density, the estimate, and pointwise 95% credible (or confidence) intervals. The true spectral density has a prominent peak at frequency $\omega = 80/1000 = 0.08$, where $\log f(0.08) = 4.15855$. Overall, the adaptive estimates outperform the non-adaptive estimates in both settings. The non-adaptive estimate is unable to flatten out the tail, unlike the adaptive method, that manages to flatten out the tail and catch the peak with better accuracy. For the trough spectral density, (Figures 4.4 and 4.5) there is little difference between the adaptive and non-adaptive fits. To show the three methods more clearly, Figures 4.3 and 4.6 display the fit for a particular sample.

In the next chapter, we will use a real data set to apply the method presented in this thesis.

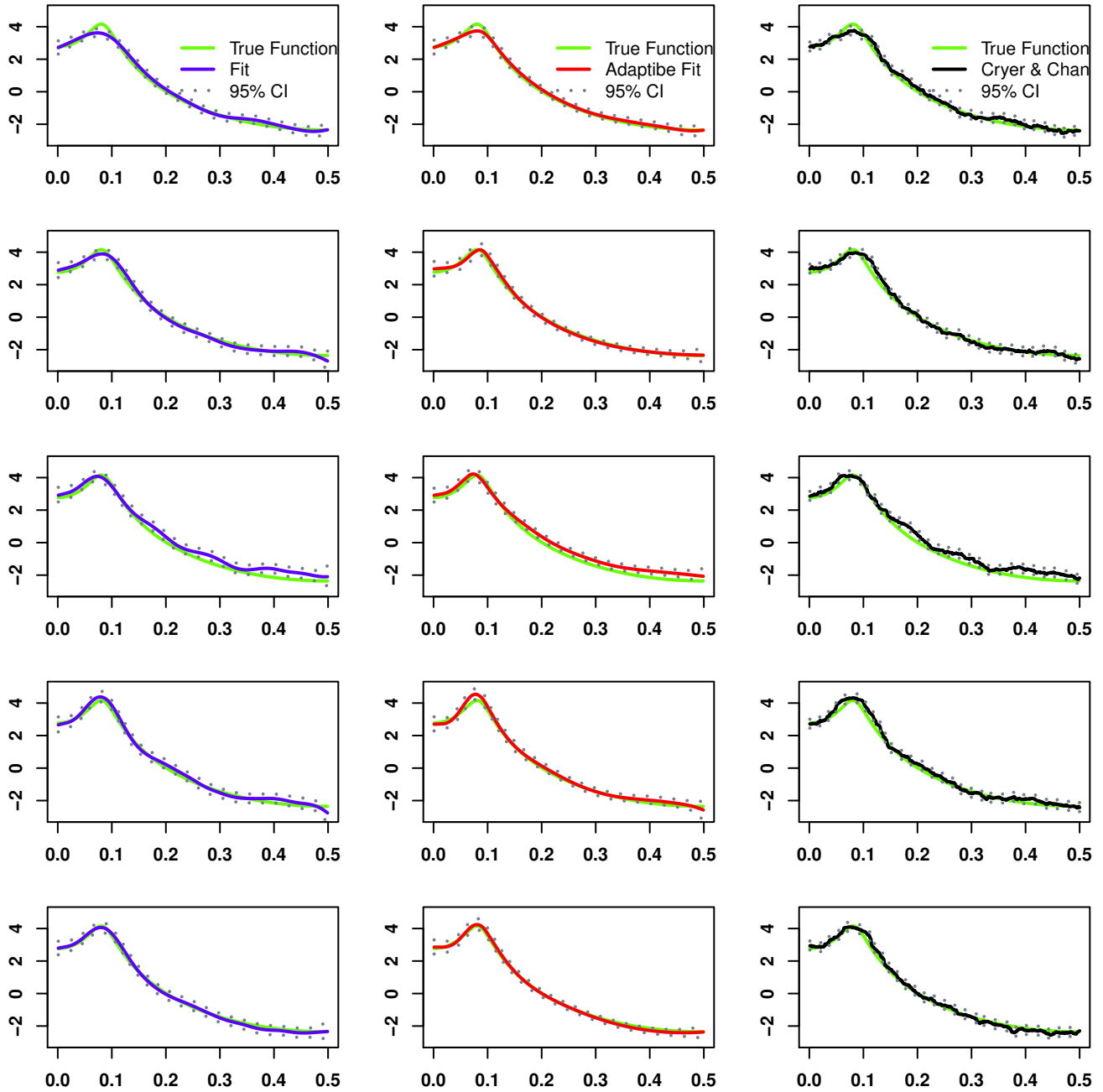


Figure 4.1: Estimates of the log spectral density based on the simulated samples from the AR(2) process with parameters $\phi_1 = 1.5$ and $\phi_2 = -0.75$. The first column displays the non-adaptive estimates; the second column displays the adaptive estimates; and the third column displays the estimates using the frequentist method. In each column, The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible (or confidence) intervals. (First 5 simulations).

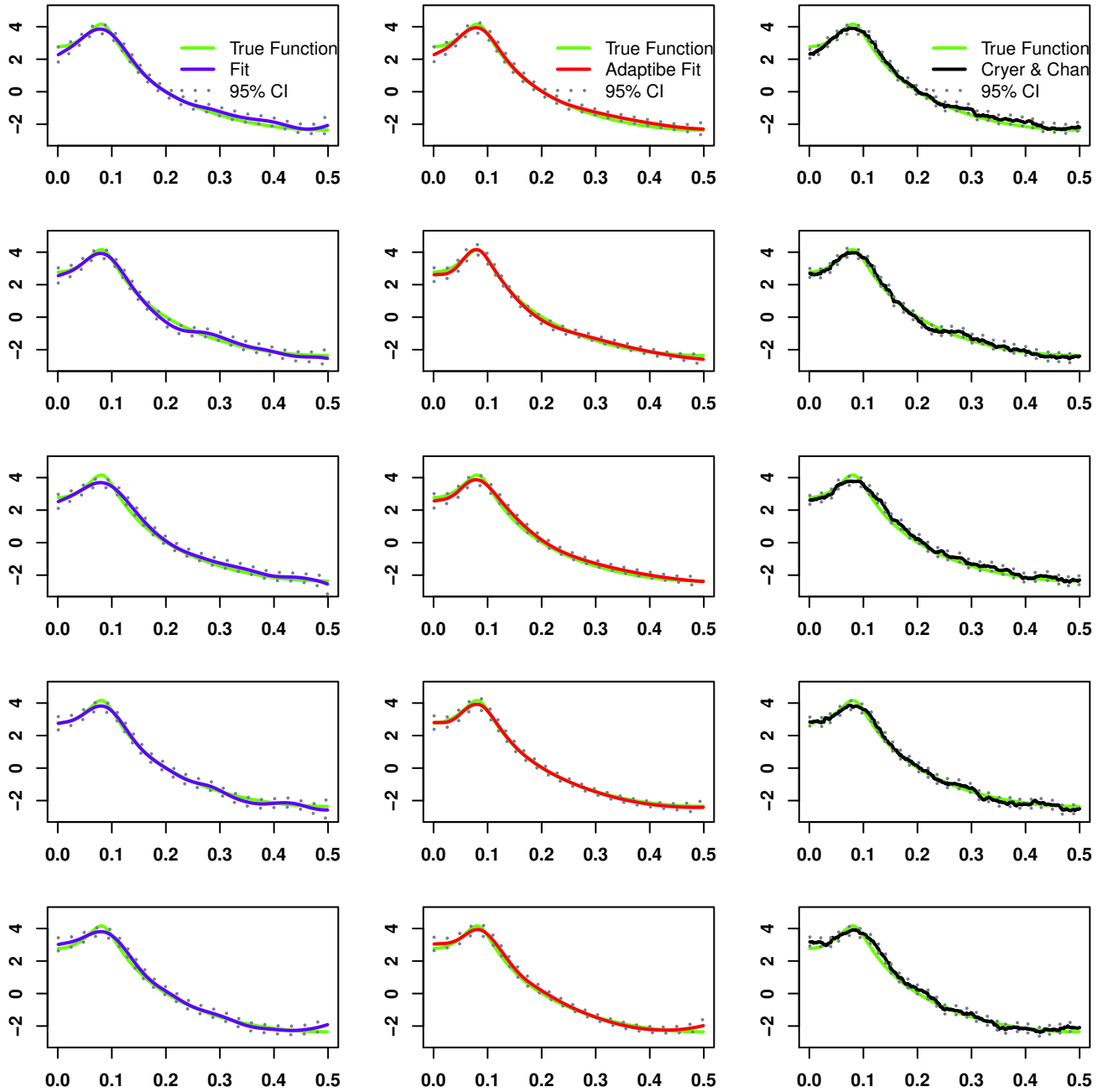


Figure 4.2: Estimates of the log spectral density based on the simulated samples from the AR(2) process with parameters $\phi_1 = 1.5$ and $\phi_2 = -0.75$. The first column displays the non-adaptive estimates; the second column displays the adaptive estimates; and the third column displays the estimates using the frequentist method. In each column, The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible (or confidence) intervals. (Second 5 simulations).

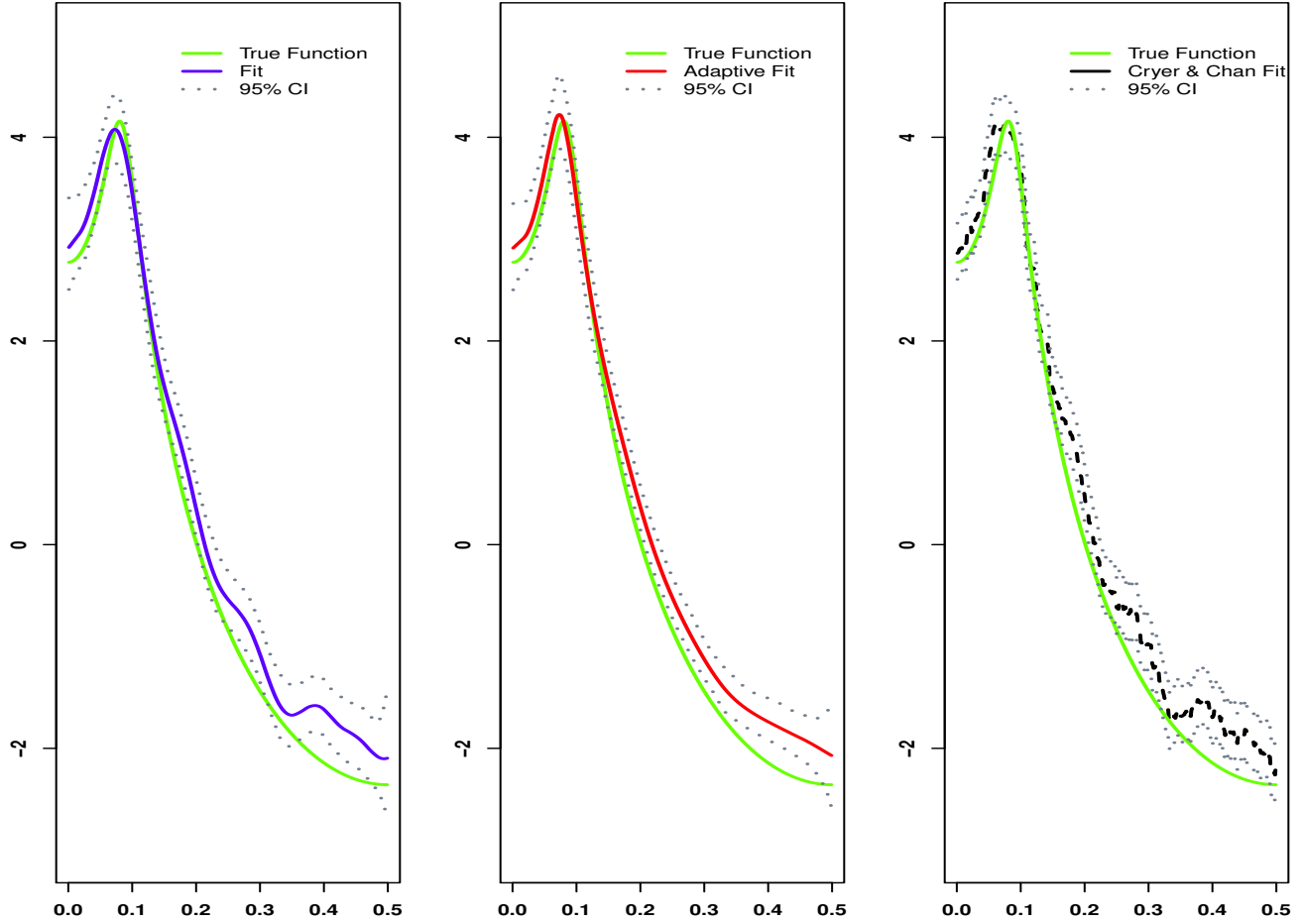


Figure 4.3: Log spectral estimates of the 3rd simulated sample from the AR(2) process with parameters $\phi_1 = 1.5$ and $\phi_2 = -0.75$. The first plot displays the non-adaptive estimate; the second plot displays the adaptive estimate; and the third plot displays the estimate using the frequentist method. The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible (or confidence) intervals.

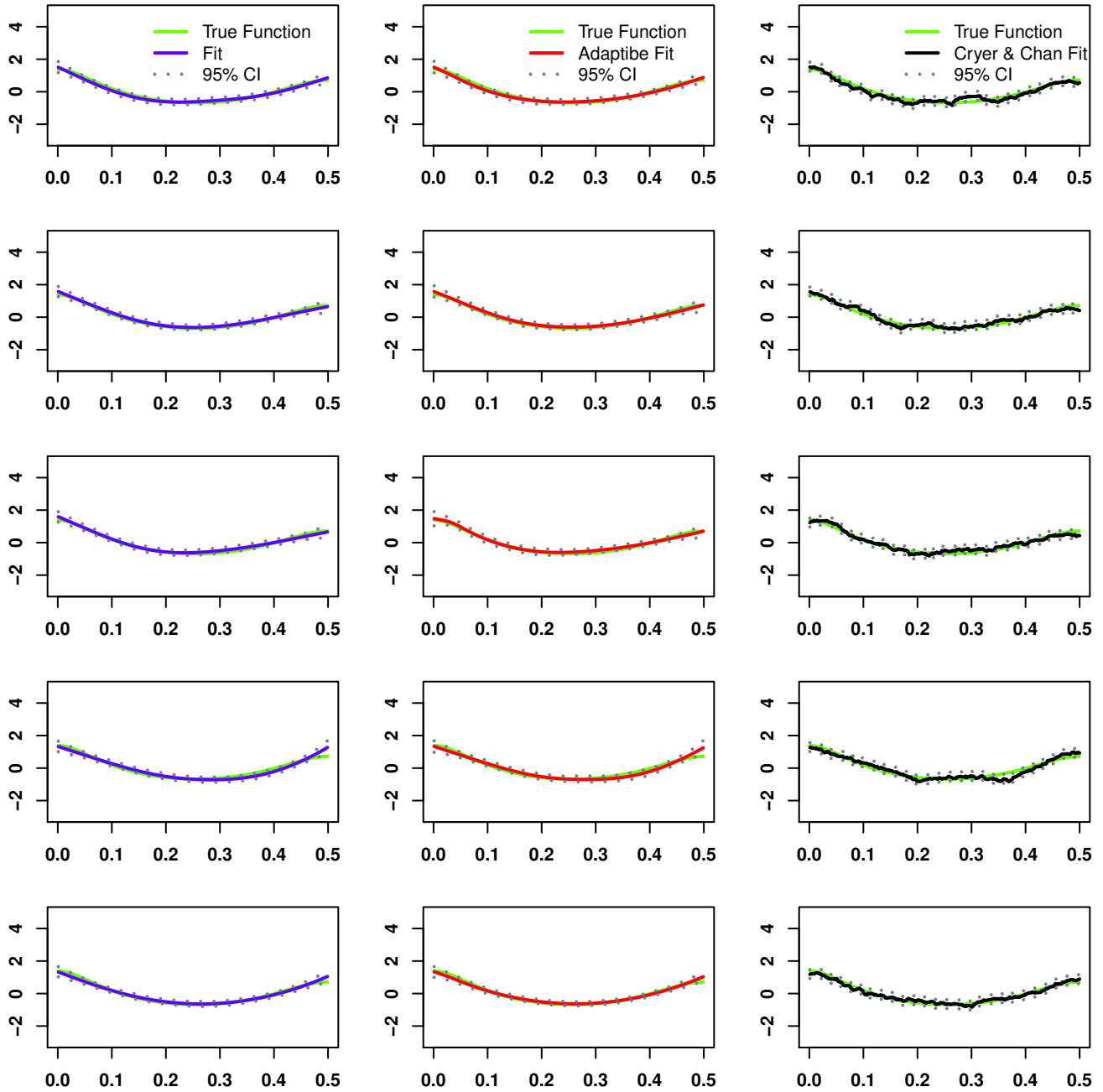


Figure 4.4: Estimates of the log spectral density based on the simulated samples from the AR(2) process with parameters $\phi_1 = 0.1$ and $\phi_2 = 0.4$. The first column displays the non-adaptive estimates; the second column displays the adaptive estimates; and the third column displays the estimates using the frequentist method. In each column, The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible (or confidence) intervals. (First 5 simulations).

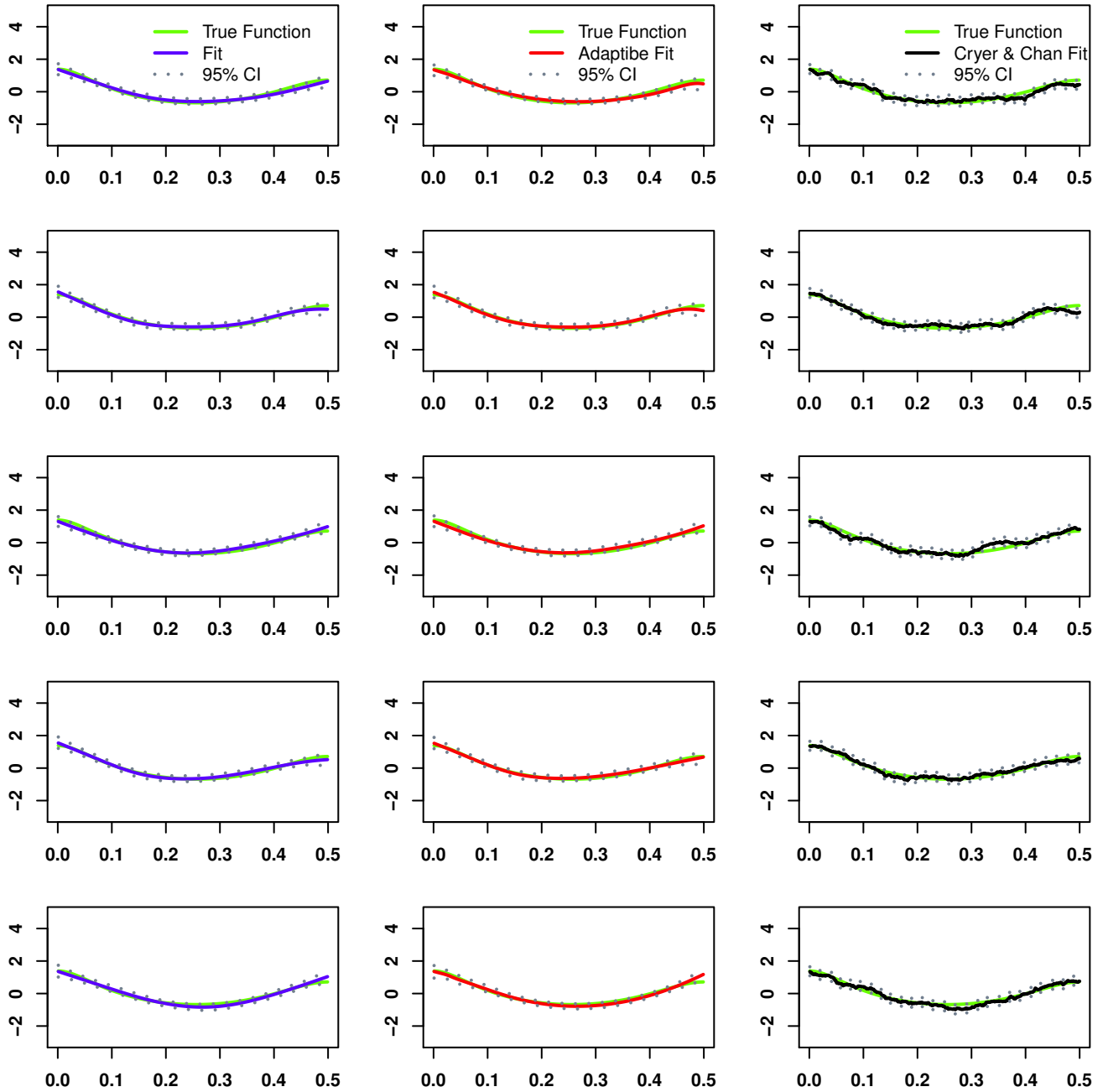


Figure 4.5: Estimates of the log spectral density based on the simulated samples from the AR(2) process with parameters $\phi_1 = 0.1$ and $\phi_2 = 0.4$. The first column displays the non-adaptive estimates; the second column displays the adaptive estimates; and the third column displays the estimates using the frequentist method. In each column, The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% credible (or confidence) intervals. (Second 5 simulations)

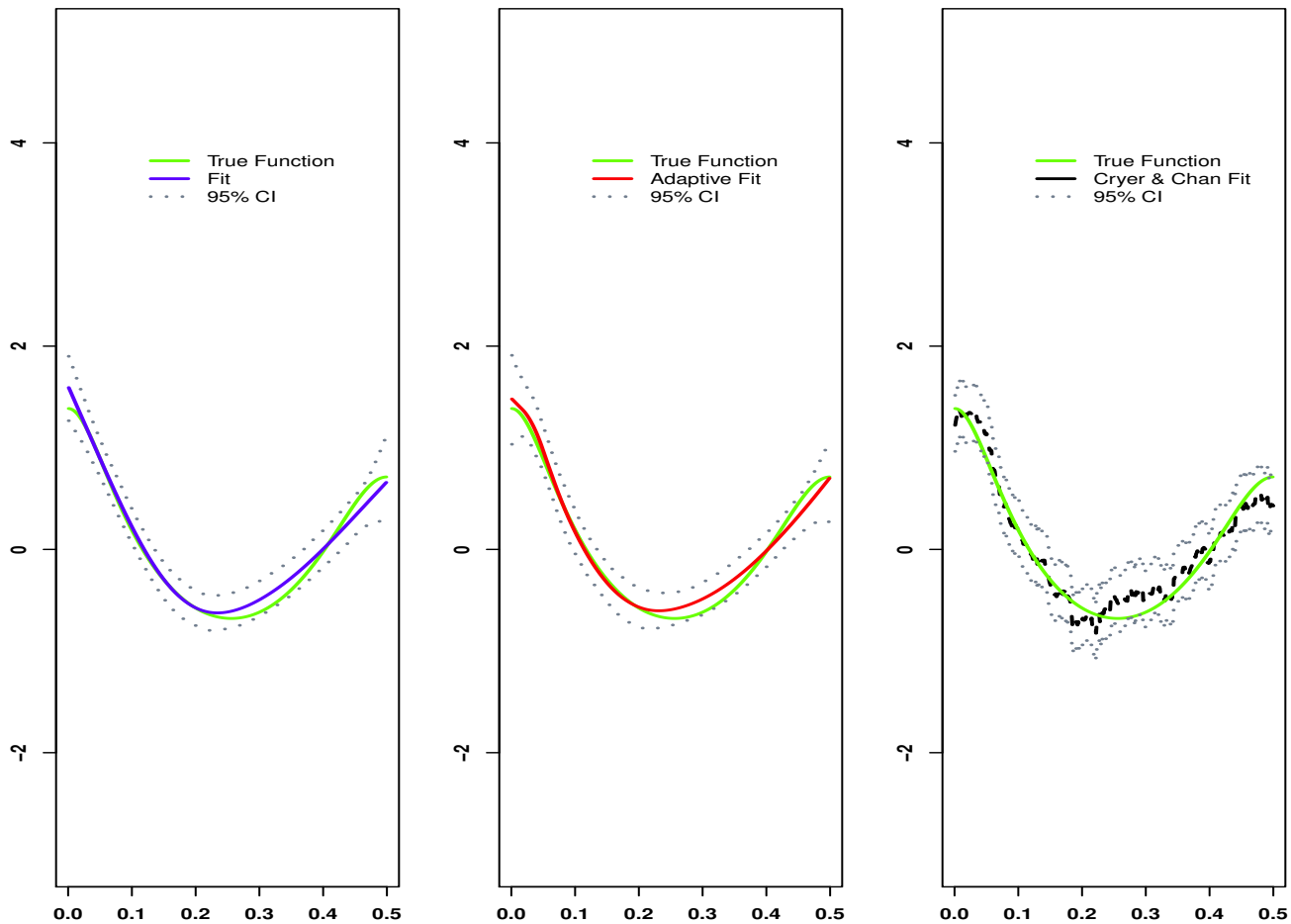


Figure 4.6: Log spectral estimates of the 3rd simulated sample from the AR(2) process with parameters $\phi_1 = 0.1$ and $\phi_2 = 0.4$. The first plot displays the non-adaptive estimate; the second plot displays the adaptive estimate; and the third plot displays the estimate using the frequentist method. The solid line represents the true spectral density; the dashed line represents the estimate; and the dotted lines represent pointwise 95% (or confidence) credible intervals.

Chapter 5

Application

In this chapter, we apply the method of smoothing the spectrum of a time series to a real data set. The example used is the El Niño Southern Oscillation (ENSO) phenomenon, where we model the spectrum of one indicator of ENSO, that is, the Southern Oscillation Index (SOI).

5.1 Southern Oscillation Index (SOI)

The El Niño Southern Oscillation (ENSO) phenomenon is the change of temperature in the Tropical Eastern Pacific Ocean that results in weather and climate episodes around the world. There has been much debate whether the intensity and frequency of ENSO has changed over the last century due to human-induced global warming. One particular indicator for ENSO that was used by Rosen et al. (2009) was the Southern Oscillation Index (SOI), which is the center of this application. The monthly standardized anomaly of the mean sea-level pressure differences between Tahiti and Darwin is what we refer to as the SOI. Negative values of the SOI that are below -8 are representative of El Niño episodes and are associated with the warming of the central and eastern Pacific Ocean. Positive values of the SOI above +8 are representative of La Niña episodes and are associated with warm sea temperatures to the north of Australia. The data are available from the Australian Bureau of Meteorology at <http://www.bom.gov.au/climate/current/soihtml1.shtml> and consist of measurements from the year 1876 to present.

Figure 5.1 shows the monthly values of the SOI. The sample size is $n = 1673$, with measurement values from January 1876 to June 2015.

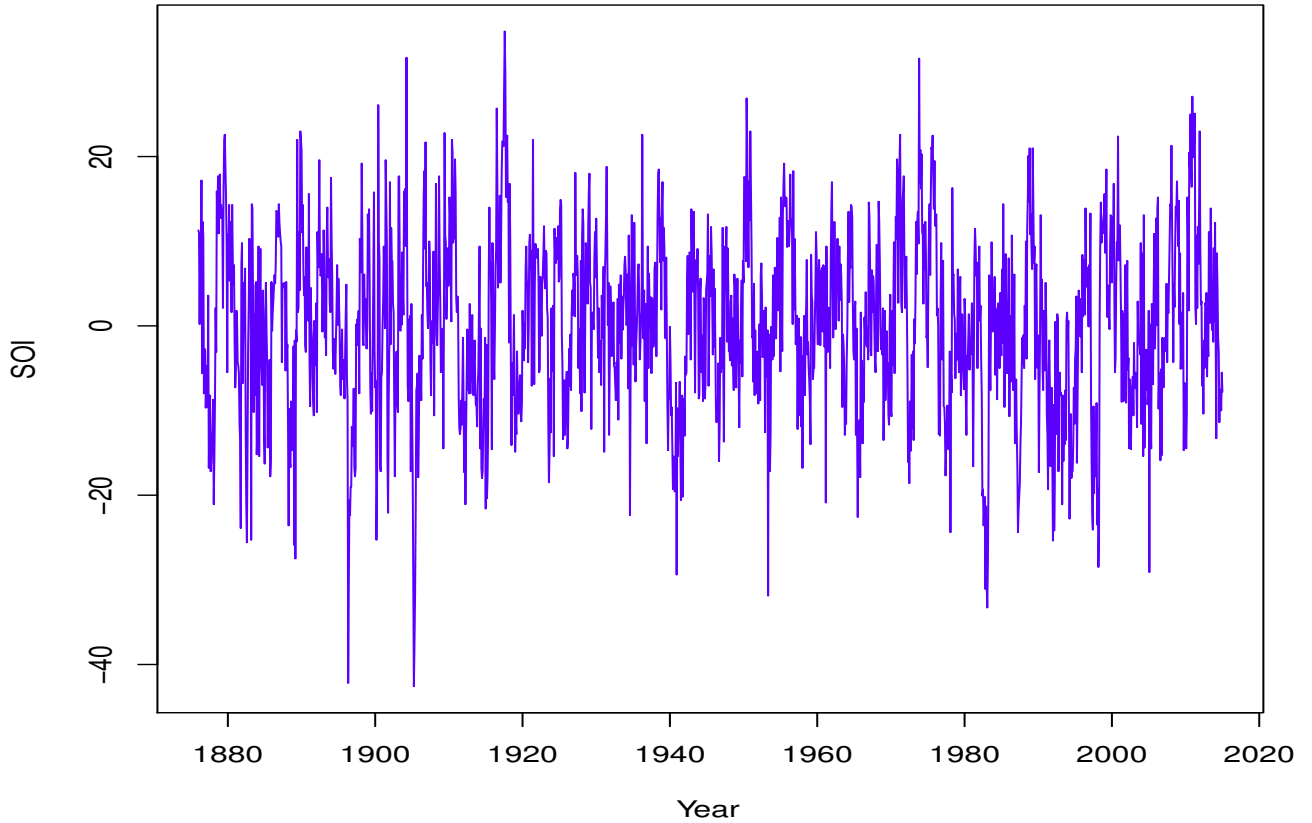


Figure 5.1: Monthly values of the Southern Oscillation Index (SOI).

5.2 Analysis

Two settings are considered for the application: The first setting utilizes quadratic basis functions to smooth the log periodogram of the SOI time series; the second setting utilizes cosine basis functions. The adaptive and non-adaptive methods were fit to the data running the algorithm for 2000 iterations with a burn-in period of 500. The number of knots for both fits is $K_\kappa = 50$. In the adaptive case, $K_\iota = 15$ knots are used for the piecewise constant basis functions. Figure 5.2 displays the log periodogram and the pointwise 95% credible intervals for the fits based on quadratic basis functions.

It is seen that in this case the non-adaptive fit does a poor job in the low frequency region where there seems to be a peak in the log periodogram. The adaptive fit is more satisfactory. Using the adaptive estimate there is a peak at $\omega = 35/1673 \approx 0.02$, that corresponds to a cycle of $1673/35 = 48$, corresponding to 4 years. Figure 5.3 is analogous to Figure 5.2 except that this time cosine basis functions are used for both the adaptive and non-adaptive methods. These basis functions (see Eubank (1999)) are given by

$$\frac{\sqrt{2}}{\pi(j-1)} \cos(\omega[j-1]\pi).$$

The corresponding fits are essentially the same for both the adaptive and the non-adaptive methods. The peak seems to be slightly better estimated compared to the fits corresponding to the polynomial basis functions, but the rest of the fit is wigglier. In this case the peak occurs at about 50 months, which is close to the result based on Figure 5.2.

5.3 Conclusion

This thesis focused on the study of spatially adaptive splines from a Bayesian point of view. As was shown through various examples, spatial adaptivity may provide better fits.

Periodogram for SOI

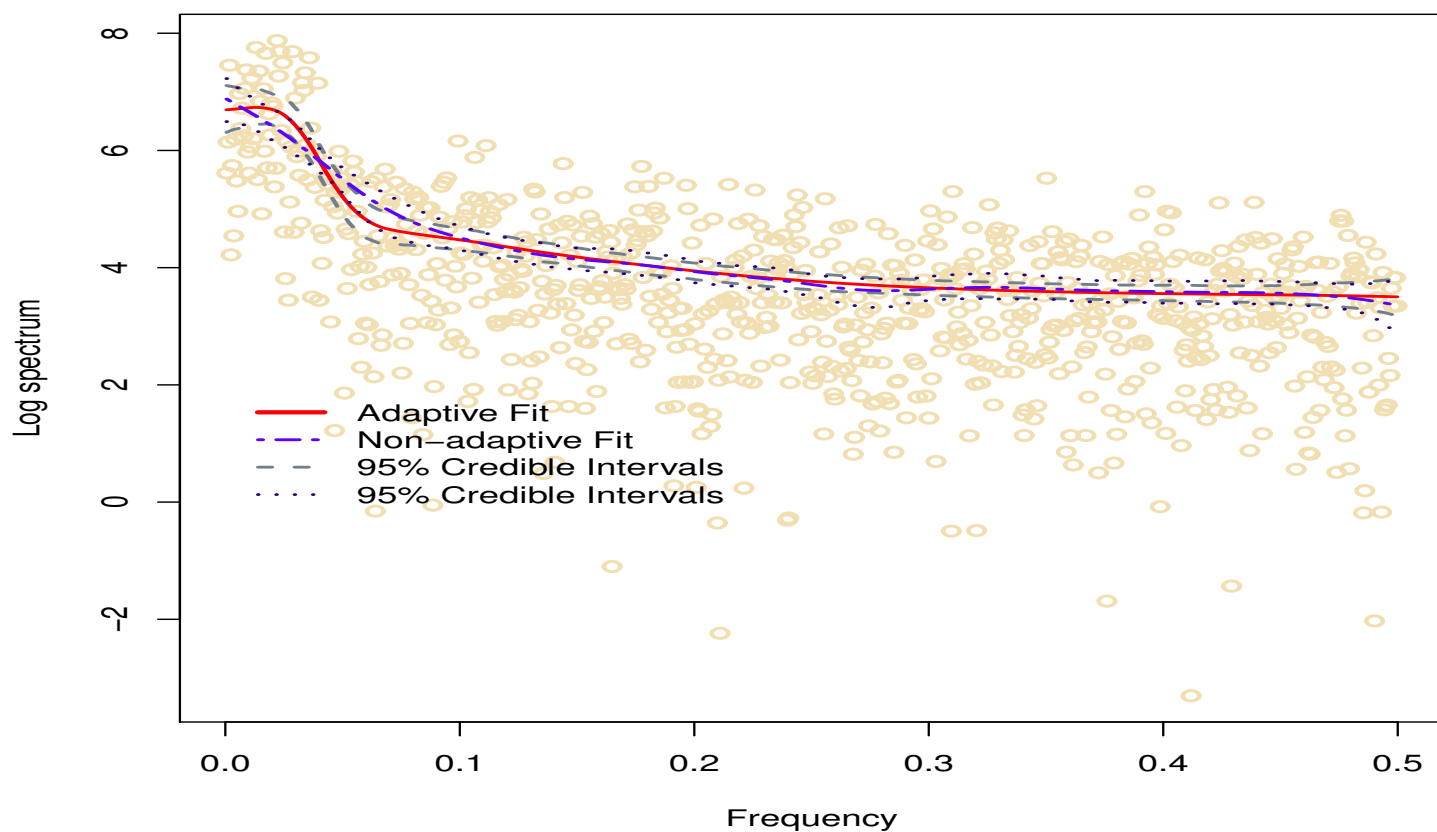


Figure 5.2: Adaptive estimate of the log spectrum (solid line); non-adaptive estimate of the log spectrum (dashed-dotted line); pointwise 95% credible intervals for the adaptive fit (dashed lines); and pointwise 95% credible intervals for the non-adaptive fit (dotted lines) against frequency for monthly values of the SOI. The fits are based on quadratic basis functions.

Periodogram for SOI

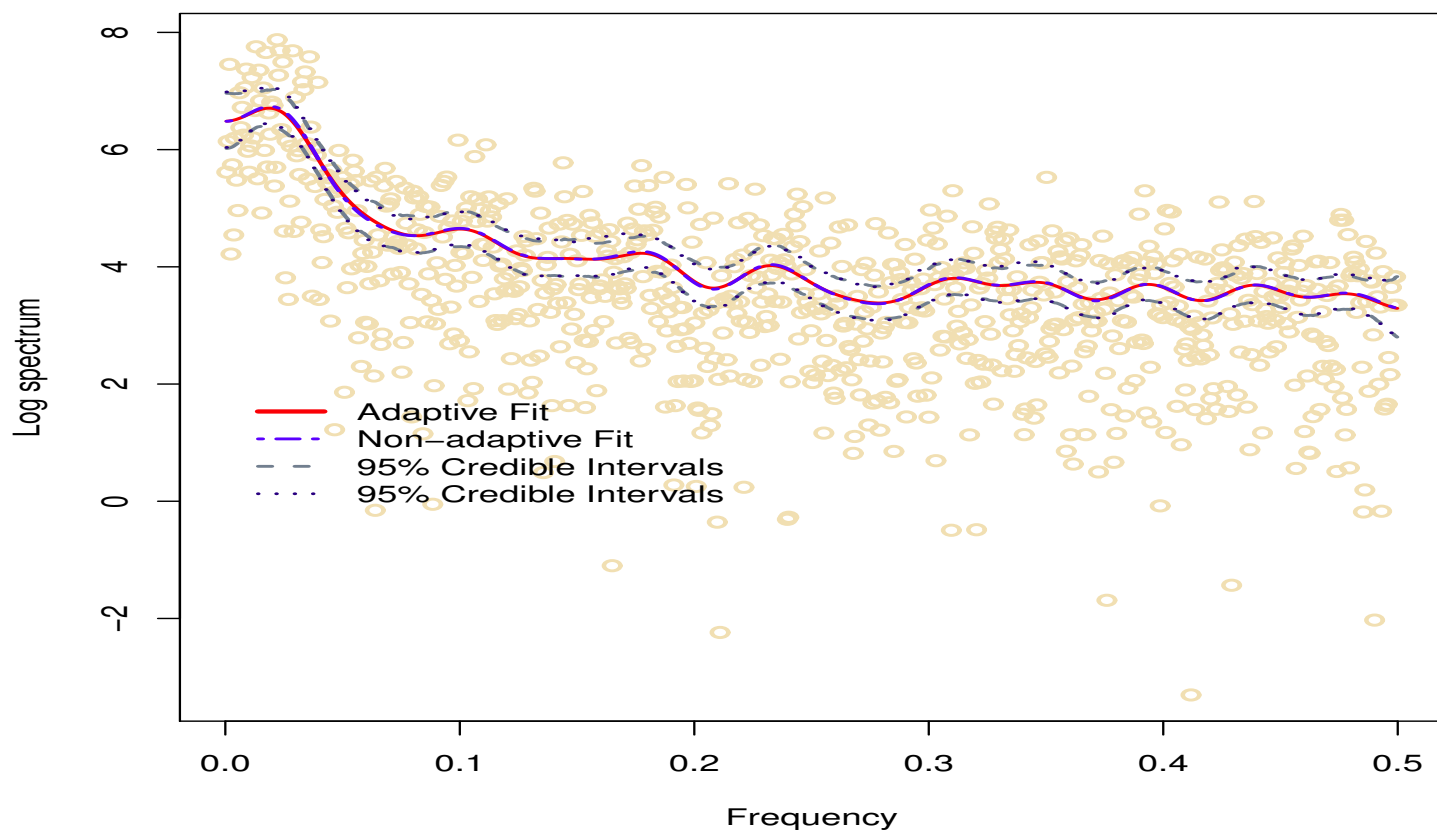


Figure 5.3: Adaptive estimate of the log spectrum (solid line); non-adaptive estimate of the log spectrum (dashed-dotted line); pointwise 95% credible intervals for the adaptive fit (dashed lines); and pointwise 95% credible intervals for the non-adaptive fit (dotted lines) against frequency for monthly values of the SOI. The fit are based on cosine basis functions.

References

- Baladandayuthapani, V., Mallick, B., and Carrol, R. (2005), “Spatially adaptive Bayesian regression rplines (P-splines),” *Journal of Computational and Graphical Statistics*, 14, 378–394.
- Crainiceanu, C., Ruppert, D., Carrol, R., Joshi, A., and Goodner, B. (2007), “Spatially adaptive Bayesian penalized splines with heteroscedastic errors,” *Journal of Computational and Graphical Statistics*, 16, 265–288.
- Cryer, J. D. and Chan, K.-S. (2008), *Time Series Analysis with Applications in R*, Springer, 2nd ed.
- Denison, D., Mallick, B., and Smith, A. (1998), “Automatic Bayesian curve fitting,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60, 333–350.
- DiMatteo, I., Genovese, C., and Kass, R. (2001), “Bayesian curve-fitting with free-knot splines,” *Biometrika*, 88, 1055–1071.
- Donoho, D. and Johnstone, I. (1994), “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, 3, 425–455.
- (1995), “Adapting to unknown smoothness via wavelent shrinkage,” *Journal of the American Statistical Association*, 90, 1200–1224.
- Eilers, P. and Marx, B. (1996), “Flexible smoothing with B-splines and penalties,” *Statistical Science*, 11, 89–102.
- Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, 2nd ed.

- Gelman, A. (2006), “Prior distribution for variace components in hierarchcal model (comment on article by Browne and Draper),” *Bayesian Analysis*, 1, 515–534.
- Henderson, C. R. (1950), “Estimation of genetic paramters (abstract),” *Annals of Mathematical Statistics*, 21, 309–310.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008), “Mixture g priors for Bayesian variable selection,” *Journal of the American Statistical Association*, 103, 410–423.
- Lou, Z. and Wahba, G. (1997), “Hybrid adaptive splines,” *Journal of the American Statistical Association*, 92, 107–116.
- Miyata, S. and Shen, X. (2003), “Adaptive free-knot splines,” *Journal of Computational and Graphical Statistics*, 12, 197–213.
- Pawitan, Y. and O’Sullivan, F. (1994), “Nonparametric spectral density estimation using penalized Whittle likelihood,” *Journal of the American Statistical Association*, 89, 600–610.
- Robinson, G. (1991), “That BLUP is a good thing: estimation of random effects,” *Statistical Science*, 6, 15–32.
- Rosen, O., Stoffer, D. S., and Wood, S. (2009), “Local spectral analysis via a Bayesian mixture of smoothing splines,” *Journal of the American Statistical Association*, 104, 249–262.
- Ruppert, D. and Carroll, R. (2000), “Spatially-adaptive penalties for spline fitting,” *Australian and New Zealand Journal of Statistics*, 42, 205–223.
- Ruppert, D., Wand, M., and Carrol, R. (2003), *Semiparametric Regression*, Cambridge University Press.

- Scheipl, F. and Kneib, T. (2009), “Locally adaptive bayesian P-splines with a normal-exponential-gamma Prior,” *Computational Statistics and Data Analysis*, 53, 3533–3552.
- Shumway, R. H. and Soffer, D. S. (2011), *Time Series Analysis and Its Applications*, Springer, 3rd ed.
- Smith, M. and Kohn, R. (1996), “Nonparametric regression using Bayesian variable selection,” *Journal of Econometrics*, 75, 317–343.
- Staniswalis, J. G. (1989), “Local bandwidth selection for kernel estimates,” *Journal of the American Statistical Association*, 84, 284–288.
- Staniswalis, J. G. and Yandell, B. S. (1992), “Locally adaptive smoothing splines,” *Journal of Statistical Computation and Simulation*, 43, 45–53.
- Wahba, G. (1980), “Automatic smoothing of the log periodogram,” *Journal of the American Statistical Association*, 75, 122–132.
- Wand, M., Ormerod, J. T., Padoan, S. A., and Frühwirth, R. (2012), “Mean field variational Bayes for elaborate distributions,” *Bayesian Analysis*, 7, 847–900.
- Whittle, P. (1957), “Curve and periodogram smoothing,” *Journal of the Royal Statistical Society. Series B (Methodology)*, 19, 38–63.
- Wood, S., Jiang, W., and Tanner, M. (2002), “Bayesian mixture of splines for spatially sdaptive nonparametric regression,” *Biometrika*, 89, 513–528.
- Yue, Y. and Speckman, P. (2010), “Nonstationary spatial Gaussian Markov random fields,” *Journal of Computational and Graphical Statistics*, 19, 96–116.
- Yue, Y., Speckman, P., and Sun, D. (2012), “Priors for Bayesian adaptive spline smoothing,” *Annals of the Institute of Statistical Mathematics*, 64, 577–613.
- Zhou, S. and Shen, X. (2001), “Spatially adaptive regression splines and accurate knot selection schemes,” *Journal of the American Statistical Association*, 96, 247–259.

Appendix A

Derivation of Sampling Schemes

A.1 BPS Sampling Scheme

The derivation of the full conditional distributions for the BPS model is provided in this section.

1. Sampling θ

$$\begin{aligned}
 p(\theta | \sigma_\epsilon^2, \sigma_b^2, \mathbf{y}) &\propto p(\mathbf{y} | \beta, \mathbf{b}, \sigma_\epsilon^2) p(\mathbf{b} | \sigma_b^2) p(\sigma_b^2) p(\beta) p(\sigma_\epsilon^2) \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{y} - T\theta)' (\mathbf{y} - T\theta) \right\} \times \exp \left\{ -\frac{1}{2\sigma_b^2} \mathbf{b}' \mathbf{b} \right\} \times \exp \left\{ -\frac{1}{2\sigma_\beta^2} \beta' \beta \right\} \\
 &= \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{y} - T\theta)' (\mathbf{y} - T\theta) - \frac{1}{2\sigma_b^2} \mathbf{b}' \mathbf{b} - \frac{1}{2\sigma_\beta^2} \beta' \beta \right\} \\
 &= \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{y}' - \theta' T') (\mathbf{y} - T\theta) - \frac{1}{2} \theta' D^{-1} \theta \right\} \\
 &= \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{y}' \mathbf{y} - 2\theta' T' \mathbf{y} + \theta' T' T \theta) - \frac{1}{2} \theta' D^{-1} \theta \right\} \\
 &\propto \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (-2\theta' T' \mathbf{y} + \theta' T' T \theta) - \frac{1}{2} \theta' D^{-1} \theta \right\} \\
 &= \exp \left\{ \frac{1}{\sigma_\epsilon^2} \theta' T' \mathbf{y} - \frac{1}{2\sigma_\epsilon^2} \theta' T' T \theta - \frac{1}{2} \theta' D^{-1} \theta \right\} \\
 &= \exp \left\{ \frac{1}{\sigma_\epsilon^2} \theta' T' \mathbf{y} - \frac{1}{2} \theta' \left[\frac{1}{\sigma_\epsilon^2} T' T + D^{-1} \right] \theta \right\}
 \end{aligned}$$

Note that $\Sigma_\theta = \left(\frac{1}{\sigma_\epsilon^2} T' T + D^{-1} \right)^{-1} = \sigma_\epsilon^2 (T' T + \sigma_\epsilon^2 D^{-1})^{-1}$. By solving for μ_θ , i.e., $\Sigma_\theta^{-1} \mu_\theta = \frac{1}{\sigma_\epsilon^2} T' \mathbf{y} \Rightarrow \mu_\theta = \frac{1}{\sigma_\epsilon^2} \Sigma_\theta T' \mathbf{y}$.

2. Sampling σ_b^2

$$\begin{aligned}
p(\sigma_b^2 | \mathbf{y}, \mathbf{b}, \boldsymbol{\beta}, \sigma_\epsilon^2) &\propto p(\mathbf{b} | \sigma_b^2) p(\sigma_b^2) \\
&\propto (\sigma_b^2)^{-\frac{K_\kappa}{2}} (\sigma_b^2)^{-(A_b+1)} \exp \left\{ -\frac{1}{2\sigma_b^2} \|\mathbf{b}\|^2 \right\} \times \exp \left\{ -\frac{B_b}{\sigma_b^2} \right\} \\
&= (\sigma_b^2)^{-\frac{K_\kappa}{2}} (\sigma_b^2)^{-(A_b+1)} \exp \left\{ -\frac{1}{2\sigma_b^2} \|\mathbf{b}\|^2 - \frac{B_b}{\sigma_b^2} \right\} \\
&= (\sigma_b^2)^{-(\frac{K_\kappa}{2} + A_b + 1)} \exp \left\{ -\frac{1}{\sigma_b^2} \left[\frac{1}{2} \|\mathbf{b}\|^2 + B_b \right] \right\}.
\end{aligned}$$

3. Sampling σ_ϵ^2

$$\begin{aligned}
p(\sigma_\epsilon^2 | \mathbf{y}, \mathbf{b}, \boldsymbol{\beta}, \sigma_b^2) &\propto p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{b}, \sigma_\epsilon^2) p(\sigma_\epsilon^2) \\
&\propto (\sigma_\epsilon^2)^{-\frac{n}{2}} (\sigma_\epsilon^2)^{-(A_\epsilon+1)} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 \right\} \times \exp \left\{ -\frac{B_\epsilon}{\sigma_\epsilon^2} \right\} \\
&= (\sigma_\epsilon^2)^{-(\frac{n}{2} + A_\epsilon + 1)} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 - \frac{B_\epsilon}{\sigma_\epsilon^2} \right\} \\
&= (\sigma_\epsilon^2)^{-(\frac{n}{2} + A_\epsilon + 1)} \exp \left\{ -\frac{1}{\sigma_\epsilon^2} \left[\frac{1}{2} \|\mathbf{y} - T\boldsymbol{\theta}\|^2 + B_\epsilon \right] \right\}
\end{aligned}$$

A.2 BAPS Sampling Scheme

The derivation of the full conditional distributions for the BAPS model is provided in this section.

1. Sampling θ

$$\begin{aligned}
p(\theta|\mathbf{b}, \beta, \gamma, \tau, \xi_1, \mathbf{y}) &\propto p(\mathbf{y}|\mathbf{b}, \theta, \tau)p(\mathbf{b}|\tau, \xi_1)p(\beta)p(\tau)p(\xi_1|\rho_1) \\
&\propto \exp\left\{-\frac{\tau}{2}(\mathbf{y} - T\theta)'(\mathbf{y} - T\theta)\right\} \times \exp\left\{-\frac{1}{2}\tau\xi_1\mathbf{b}'D_\gamma\mathbf{b}\right\} \times \exp\left\{-\frac{1}{2\sigma_\beta^2}\beta'\beta\right\} \\
&= \exp\left\{-\frac{\tau}{2}(\mathbf{y} - T\theta)'(\mathbf{y} - T\theta) - \frac{1}{2}\theta'\Lambda_y\theta\right\} \\
&= \exp\left\{-\frac{\tau}{2}(\mathbf{y}' - \theta'T')(\mathbf{y} - T\theta) - \frac{1}{2}\theta'\Lambda_y\theta\right\} \\
&= \exp\left\{-\frac{\tau}{2}(\mathbf{y}'\mathbf{y} - 2\theta'T'\mathbf{y} + \theta'T'T\theta) - \frac{1}{2}\theta'\Lambda_y\theta\right\} \\
&\propto \exp\left\{-\frac{\tau}{2}(-2\theta'T'\mathbf{y} + \theta'T'T\theta) - \frac{1}{2}\theta'\Lambda_y\theta\right\} \\
&= \exp\left\{\tau\theta'T'\mathbf{y} - \frac{\tau}{2}\theta'T'T\theta - \frac{1}{2}\theta'\Lambda_y\theta\right\} \\
&= \exp\left\{\tau\theta'T'\mathbf{y} - \frac{1}{2}\theta'[\tau T'T + \Lambda_y]\theta\right\}
\end{aligned}$$

Note that $Q_\theta = (\tau T'T + \Lambda_y)^{-1}$. By solving for μ_θ , i.e., $Q_\theta^{-1}\mu_\theta = \tau T'\mathbf{y} \Rightarrow \mu_\theta = \tau Q_\theta T'\mathbf{y}$.

2. Sampling τ

$$\begin{aligned}
p(\tau|\gamma, \mathbf{b}, \xi_1, \xi_2, \mathbf{y}) &\propto p(\mathbf{y}|\beta, \mathbf{b}, \tau)p(\tau)p(\mathbf{b}|\tau, \xi_1)p(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2) \\
&\propto \tau^{\frac{n}{2}}(\tau\xi_1)^{\frac{K_\kappa}{2}}(\tau\xi_1\xi_2)^{\frac{K_\iota+q}{2}}\tau^{-1}\exp\left\{-\frac{\tau}{2}\|\mathbf{y} - T\theta\|^2 - \frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b} - \frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right\} \\
&\propto \tau^{\frac{n}{2}+\frac{K_\kappa}{2}+\frac{K_\iota+q}{2}-1}\exp\left\{-\frac{\tau}{2}\|\mathbf{y} - T\theta\|^2 - \frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b} - \frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right\} \\
&= \tau^{\frac{1}{2}(n+K_\kappa+K_\iota+q)-1}\exp\left\{-\tau\left[\frac{1}{2}\|\mathbf{y} - T\theta\|^2 + \frac{\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b} + \frac{\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right]\right\}, \quad \tau > 0.
\end{aligned}$$

3. Sampling ξ_1

$$\begin{aligned}
p(\xi_1|\boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_2) &\propto p(\mathbf{b}|\tau, \xi_1)p(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2)p(\xi_1|\rho_1) \\
&\propto (\tau\xi_1)^{\frac{K_\kappa}{2}}(\tau\xi_1\xi_2)^{\frac{K_\iota+q}{2}} \exp\left\{-\frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b}-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma-\rho_1\xi_1\right\} \\
&\propto \xi_1^{\frac{K_\kappa}{2}+\frac{K_\iota+q}{2}} \exp\left\{-\frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b}-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma-\rho_1\xi_1\right\} \\
&= \xi_1^{\frac{1}{2}(K_\kappa+K_\iota+q)+1-1} \exp\left\{-\xi_1\left[\frac{\tau}{2}\mathbf{b}'D_\gamma\mathbf{b}+\frac{\tau\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma+\rho_1\right]\right\}, \quad \xi_1 > 0.
\end{aligned}$$

4. Sampling ξ_2

$$\begin{aligned}
p(\xi_2|\mathbf{b}_\gamma, \tau, \xi_1, \rho_2) &\propto p(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2)p(\xi_2|\rho_2) \\
&\propto (\tau\xi_1\xi_2)^{\frac{K_\iota+q}{2}} \exp\left\{-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma-\rho_2\xi_2\right\} \\
&\propto \xi_2^{\frac{K_\iota+q}{2}} \exp\left\{-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma-\rho_2\xi_2\right\} \\
&= \xi_2^{\frac{K_\iota+q}{2}+1-1} \exp\left\{-\xi_2\left[\frac{\tau\xi_1}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma+\rho_2\right]\right\}, \quad \xi_2 > 0.
\end{aligned}$$

5. Sampling ρ_i 's for $i = 1, 2$

$$\begin{aligned}
p(\rho_i|\xi_i, c_i) &\propto p(\xi_i|\rho_i)p(\rho_i|c_i) \\
&\propto \rho_i \exp\{\xi_i\rho_i\} \times c_i \exp\{c_i\rho_i\} \\
&\propto \rho_i \exp\{-\xi_i\rho_i - c_i\rho_i\} \\
&= \rho_i^{2-1} \exp\{-\rho_i[\xi_i + c_i]\}.
\end{aligned}$$

6. Sampling the $b_{\gamma k}$'s

$$\begin{aligned}
p(\mathbf{b}_\gamma|\boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) &\propto p(\mathbf{b}_\gamma|\tau, \xi_1, \xi_2)p(\mathbf{b}|\tau, \xi_1) \\
&\propto \tau\xi_1^{\frac{K_\kappa}{2}}|D_\gamma|^{\frac{1}{2}} \exp\left\{-\frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b}\right\} \times \exp\left\{-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right\} \\
&= |D_\gamma|^{\frac{1}{2}} \exp\left\{-\frac{\tau\xi_1}{2}\mathbf{b}'D_\gamma\mathbf{b}-\frac{\tau\xi_1\xi_2}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right\}.
\end{aligned}$$

Partial derivatives:

$$\begin{aligned}
\log p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) &= \frac{1}{2} \sum_{j=1}^{K_\kappa} \mathbf{z}'_{\gamma_j} \mathbf{b}_\gamma - \frac{1}{2} \tau \xi_1 \sum_{j=1}^{K_\kappa} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_\gamma\} - \frac{1}{2} \tau \xi_1 \xi_2 \mathbf{b}'_\gamma \mathbf{b}_\gamma \\
\frac{\partial \log p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)}{\partial \mathbf{b}_\gamma} &= \frac{1}{2} Z'_\gamma \mathbf{1} - \frac{1}{2} \tau \xi_1 \sum_{j=1}^{K_\kappa} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_\gamma\} \mathbf{z}_{\gamma_j} - \tau \xi_1 \xi_2 \mathbf{b}_\gamma \\
\frac{\partial^2 \log p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)}{\partial \mathbf{b}_\gamma \partial \mathbf{b}'_\gamma} &= -\frac{1}{2} \tau \xi_1 \sum_{j=1}^{K_\kappa} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_\gamma\} \mathbf{z}_{\gamma_j} \mathbf{z}'_{\gamma_j} - \tau \xi_1 \xi_2 I_{K_\iota+q},
\end{aligned}$$

where \mathbf{z}_{γ_j} is the j th row of Z_γ . Propose a new value \mathbf{b}_γ^* from $N(\hat{\mathbf{b}}_\gamma, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_\gamma})$, where $\hat{\mathbf{b}}_\gamma = \arg \max_{\mathbf{b}_\gamma} \log p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)$ and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{b}_\gamma} = \left[-\frac{\partial^2 \log p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)}{\partial \mathbf{b}_\gamma \partial \mathbf{b}'_\gamma} \Big|_{\mathbf{b}_\gamma = \hat{\mathbf{b}}_\gamma} \right]^{-1}.$$

Accept the new value via

$$\mathbf{b}_\gamma^{(t+1)} = \begin{cases} \mathbf{b}_\gamma^* & \text{with probability } \alpha_{\mathbf{b}_\gamma} \\ \mathbf{b}_\gamma^{(t)} & \text{with probability } 1 - \alpha_{\mathbf{b}_\gamma}, \end{cases}$$

with acceptance probability

$$\alpha_{\mathbf{b}_\gamma} = \min \left\{ \frac{\pi(\mathbf{b}_\gamma^* | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) g(\mathbf{b}_\gamma^{(t)} | \mathbf{b}_\gamma^*)}{\pi(\mathbf{b}_\gamma^{(t)} | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) g(\mathbf{b}_\gamma^* | \mathbf{b}_\gamma^{(t)})}, 1 \right\},$$

such that $\pi(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2) = p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \tau, \xi_1, \xi_2)$ and $g(\mathbf{b}_\gamma^* | \mathbf{b}_\gamma^{(t)})$ is the proposal density for \mathbf{b}_γ^* , i.e., $N(\hat{\mathbf{b}}_\gamma, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_\gamma})$.

A.3 Whittle Sampling Scheme

The derivation of the full conditional distributions for the BAPS model for spectral analysis is provided in this section.

1. Sampling $\boldsymbol{\theta}$

$$\begin{aligned}
 p(\boldsymbol{\theta}|\mathbf{I}, C) &\propto p(\mathbf{y}|f) \times p(\boldsymbol{\theta}) \\
 &\propto \exp \left\{ \sum_{m=1}^M [\mathbf{c}'_m \boldsymbol{\theta} - I_m(\omega) \exp(\mathbf{c}'_m \boldsymbol{\theta})] \right\} \times \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta} \right\} \\
 &= \exp \left\{ \sum_{m=1}^M [\mathbf{c}'_m \boldsymbol{\theta} - I_m(\omega) \exp(\mathbf{c}'_m \boldsymbol{\theta})] - \frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta} \right\}.
 \end{aligned}$$

where \mathbf{c}'_m is the m th row of C .

Partial derivatives:

$$\begin{aligned}
 \log p(\boldsymbol{\theta}|\mathbf{I}, C) &= \sum_{m=1}^M [\mathbf{c}'_m \boldsymbol{\theta} - I_m(\omega) \exp(\mathbf{c}'_m \boldsymbol{\theta})] - \frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta} \\
 \frac{\partial \log p(\boldsymbol{\theta}|\mathbf{I}, C)}{\partial \boldsymbol{\theta}} &= \sum_{m=1}^M [\mathbf{c}_m - I_m(\omega) \exp(\mathbf{c}'_m \boldsymbol{\theta}) \mathbf{c}_m] - \Lambda \boldsymbol{\theta} \\
 \frac{\partial^2 \log p(\boldsymbol{\theta}|\mathbf{I}, C)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= - \sum_{m=1}^M [I_m(\omega) \exp(\mathbf{c}'_m \boldsymbol{\theta}) \mathbf{c}_m \mathbf{c}'_m] - \Lambda
 \end{aligned}$$

Vectorized Partial derivatives:

$$\begin{aligned}
 \log p(\boldsymbol{\theta}|\mathbf{I}, C) &= \sum_{m=1}^M [C \boldsymbol{\theta} - \exp(C \boldsymbol{\theta}) I_m(\omega)] - \frac{1}{2} \boldsymbol{\theta}' \Lambda \boldsymbol{\theta} \\
 \frac{\partial \log p(\boldsymbol{\theta}|\mathbf{I}, C)}{\partial \boldsymbol{\theta}} &= C' (\mathbf{1} - \mathbf{I} * \exp(C \boldsymbol{\theta})) - \Lambda \boldsymbol{\theta} \\
 \frac{\partial^2 \log p(\boldsymbol{\theta}|\mathbf{I}, C)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= -C' \text{diag}\{I_m(\omega) \exp(C \boldsymbol{\theta})\} C - \Lambda,
 \end{aligned}$$

where $*$ is the entrywise product of two matrices of the same dimension. Propose a new value $\boldsymbol{\theta}^*$ from $N(\hat{\boldsymbol{\theta}}, \hat{\Sigma}_{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{I}, C)$ and

$$\hat{\Sigma}_{\boldsymbol{\theta}} = \left[-\frac{\partial^2 \log p(\boldsymbol{\theta}|\mathbf{I}, C)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1}.$$

Accept the new value via

$$\boldsymbol{\theta}^{(t+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \alpha_{\boldsymbol{\theta}} \\ \boldsymbol{\theta}^{(t)} & \text{with probability } 1 - \alpha_{\boldsymbol{\theta}}, \end{cases}$$

with acceptance probability

$$\alpha_{\boldsymbol{\theta}} = \min \left\{ \frac{\pi(\boldsymbol{\theta}^*|\mathbf{I}, C)g(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(t)}|\mathbf{I}, C)g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}, 1 \right\}$$

such that $\pi(\boldsymbol{\theta}|\mathbf{I}, C) = p(\boldsymbol{\theta}|\mathbf{I}, C)$ and $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})$ is the proposal density for $\boldsymbol{\theta}^*$, i.e., $N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}})$.

2. Sampling the $b_{\gamma k}$'s

$$\begin{aligned} p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta) &\propto p(\mathbf{b}_{\gamma}|\eta)p(\mathbf{b}|\delta)p(\delta|g_1) \\ &\propto \delta^{\frac{K_{\kappa}}{2}} |D_{\gamma}|^{\frac{1}{2}} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_{\gamma} \mathbf{b} \right\} \times \exp \left\{ -\frac{\eta}{2} \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma} \right\} \\ &= |D_{\gamma}|^{\frac{1}{2}} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_{\gamma} \mathbf{b} - \frac{\eta}{2} \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma} \right\}. \end{aligned}$$

Partial derivatives:

$$\begin{aligned} \log p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta) &= \frac{1}{2} \sum_{j=1}^{K_{\kappa}} \mathbf{z}'_{\gamma_j} \mathbf{b}_{\gamma} - \frac{1}{2} \delta \sum_{j=1}^{K_{\kappa}} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_{\gamma}\} - \frac{1}{2} \eta \mathbf{b}'_{\gamma} \mathbf{b}_{\gamma} \\ \frac{\partial \log p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta)}{\partial \mathbf{b}_{\gamma}} &= \frac{1}{2} Z'_{\gamma} \mathbf{1} - \frac{1}{2} \delta \sum_{j=1}^{K_{\kappa}} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_{\gamma}\} \mathbf{z}_{\gamma_j} - \eta \mathbf{b}_{\gamma} \\ \frac{\partial^2 \log p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta)}{\partial \mathbf{b}_{\gamma} \partial \mathbf{b}'_{\gamma}} &= -\frac{1}{2} \delta \sum_{j=1}^{K_{\kappa}} b_j^2 \exp\{\mathbf{z}'_{\gamma_j} \mathbf{b}_{\gamma}\} \mathbf{z}_{\gamma_j} \mathbf{z}'_{\gamma_j} - \eta I_{K_{\kappa}+q} \end{aligned}$$

where \mathbf{z}_{γ_j} is the j th row of Z_{γ} . Propose a new value \mathbf{b}_{γ}^* from $N(\hat{\mathbf{b}}_{\gamma}, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_{\gamma}})$, where $\hat{\mathbf{b}}_{\gamma} = \arg \max_{\mathbf{b}_{\gamma}} \log p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta)$ and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{b}_{\gamma}} = \left[-\frac{\partial^2 \log p(\mathbf{b}_{\gamma}|\boldsymbol{\gamma}, \mathbf{b}, \eta, \delta)}{\partial \mathbf{b}_{\gamma} \partial \mathbf{b}'_{\gamma}} \Big|_{\mathbf{b}_{\gamma}=\hat{\mathbf{b}}_{\gamma}} \right]^{-1}.$$

Accept the new value via

$$\mathbf{b}_{\gamma}^{(t+1)} = \begin{cases} \mathbf{b}_{\gamma}^* & \text{with probability } \alpha_{\mathbf{b}_{\gamma}} \\ \mathbf{b}_{\gamma}^{(t)} & \text{with probability } 1 - \alpha_{\mathbf{b}_{\gamma}}, \end{cases}$$

with acceptance probability

$$\alpha_{\mathbf{b}_\gamma} = \min \left\{ \frac{\pi(\mathbf{b}_\gamma^* | \boldsymbol{\gamma}, \mathbf{b}, \eta, \delta) g(\mathbf{b}_\gamma^{(t)} | \mathbf{b}_\gamma^*)}{\pi(\mathbf{b}_\gamma^{(t)} | \boldsymbol{\gamma}, \mathbf{b}, \eta, \delta) g(\mathbf{b}_\gamma^* | \mathbf{b}_\gamma^{(t)})}, 1 \right\},$$

such that $\pi(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \eta, \delta) = p(\mathbf{b}_\gamma | \boldsymbol{\gamma}, \mathbf{b}, \eta, \delta)$ and $g(\mathbf{b}_\gamma^* | \mathbf{b}_\gamma^{(t)})$ is the proposal density for \mathbf{b}_γ^* , i.e., $N(\hat{\mathbf{b}}_\gamma, \hat{\boldsymbol{\Sigma}}_{\mathbf{b}_\gamma})$.

3. Sampling δ

$$\begin{aligned} p(\delta | \mathbf{b}, \boldsymbol{\gamma}, g_1) &\propto p(\mathbf{b} | \boldsymbol{\gamma}, \delta) p(\delta | g_1) p(g_1) \\ &\propto \delta^{\frac{K_\kappa}{2}} |D_\gamma|^{\frac{1}{2}} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} \right\} \times (\delta^{-1})^{-(\frac{\nu_1}{2}+1)} \exp \left\{ -\frac{\nu_1 \delta}{g_1} \right\} \\ &\propto \delta^{\frac{K_\kappa}{2}} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} \right\} \times (\delta^{-1})^{-(\frac{\nu_1}{2}+1)} \exp \left\{ -\frac{\nu_1 \delta}{g_1} \right\} \\ &= \delta^{\frac{K_\kappa}{2}} (\delta^{-1})^{-(\frac{\nu_1}{2}+1)} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} \right\} \times \exp \left\{ -\frac{\nu_1 \delta}{g_1} \right\} \\ &= \delta^{\frac{K_\kappa}{2} + \frac{\nu_1}{2} - 1} \exp \left\{ -\frac{\delta}{2} \mathbf{b}' D_\gamma \mathbf{b} - \frac{\nu_1 \delta}{g_1} \right\} \\ &= \delta^{\frac{\kappa x + \nu_1}{2} - 1} \exp \left\{ -\delta \left[\frac{1}{2} \mathbf{b}' D_\gamma \mathbf{b} + \frac{\nu_1}{g_1} \right] \right\}, \quad \delta > 0. \end{aligned}$$

4. Sampling g_1

$$\begin{aligned} p(g_1 | \delta) &\propto p(\delta | g_1) p(g_1) \\ &\propto \frac{(\nu_1/g_1)^{\frac{\nu_1}{2}}}{\Gamma(\nu_1/2)} \exp \left\{ -\frac{\nu_1 \delta}{g_1} \right\} \times g_1^{-\frac{3}{2}} \exp \left\{ -\frac{1}{G_1^2 g_1} \right\} \\ &\propto g_1^{-\frac{\nu_1}{2}} g_1^{-\frac{3}{2}} \exp \left\{ -\frac{\nu_1 \delta}{g_1} - \frac{1}{G_1^2 g_1} \right\} \\ &= g_1^{-\frac{\nu_1}{2} - \frac{3}{2}} \exp \left\{ -\frac{\nu_1 \delta}{g_1} - \frac{1}{G_1^2 g_1} \right\} \\ &= g_1^{-(\frac{\nu_1+1}{2}+1)} \exp \left\{ -\frac{1}{g_1} \left[\nu_1 \delta + \frac{1}{G_1^2} \right] \right\}, \quad g_1 > 0. \end{aligned}$$

5. Sampling η

$$\begin{aligned}
p(\eta|\mathbf{b}_\gamma, g_2) &\propto p(\mathbf{b}_\gamma|\delta)p(\eta|g_2)p(g_2) \\
&\propto \eta^{\frac{K_l+q}{2}} \exp\left\{-\frac{\eta}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma\right\} \times (\eta^{-1})^{-(\frac{\nu_2}{2}+1)} \exp\left\{-\frac{\nu_2\eta}{g_2}\right\} \\
&= \eta^{\frac{K_l+q}{2}} (\eta^{-1})^{-(\frac{\nu_2}{2}+1)} \exp\left\{-\frac{\eta}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma - \frac{\nu_2\eta}{g_2}\right\} \\
&= \eta^{\frac{K_l+q}{2} + \frac{\nu_2}{2} - 1} \exp\left\{-\frac{\eta}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma - \frac{\nu_2\eta}{g_2}\right\} \\
&= \eta^{\frac{K_l+q+\nu_2}{2} - 1} \exp\left\{-\eta\left[\frac{1}{2}\mathbf{b}'_\gamma\mathbf{b}_\gamma + \frac{\nu_2}{g_2}\right]\right\}, \quad \eta > 0.
\end{aligned}$$

6. Sampling g_2

$$\begin{aligned}
p(g_2|\eta) &\propto p(\eta|g_2)p(g_2) \\
&\propto \frac{(\nu_2/g_2)^{\frac{\nu_2}{2}}}{\Gamma(\nu_2/2)} \exp\left\{-\frac{\nu_2\eta}{g_2}\right\} \times g_2^{-\frac{3}{2}} \exp\left\{-\frac{1}{G_2^2 g_2}\right\} \\
&\propto g_2^{-\frac{\nu_2}{2}} g_2^{-\frac{3}{2}} \exp\left\{-\frac{\nu_2\eta}{g_2} - \frac{1}{G_2^2 g_2}\right\} \\
&= g_2^{-\frac{\nu_2}{2} - \frac{3}{2}} \exp\left\{-\frac{\nu_2\eta}{g_2} - \frac{1}{G_2^2 g_2}\right\} \\
&= g_2^{-(\frac{\nu_2+1}{2}+1)} \exp\left\{-\frac{1}{g_2}\left[\nu_2\eta + \frac{1}{G_2^2}\right]\right\}, \quad g_2 > 0.
\end{aligned}$$

Appendix B

Code

The notation mt and ms in the following code refers to the K_κ and K_ℓ knots, respectively.

B.1 Bayesian Penalized Splines Code

```
library(MASS)
library(SemiPar)
library(MCMCpack)
k = 40
data(lidar)
x = lidar$range
y = lidar$logratio
n=length(y)
N=10000
degp=1
knots = quantile(unique(x), seq(0,1, length=(k+2))[-c(1,(k+2))])
dimnames(X) = NULL
Z = outer(x, knots, "-")
Z = Z*(Z>0)
Z=Z^degp
dimnames(Z) <- NULL
X=matrix(1,nrow=n, ncol=degp+1)
for(p in 1:degp){
  X[,p+1]=x^p
}
C=cbind(X,Z)
CT=t(C)
B=CT%*%C
betav=10000
gamma=matrix(0, nrow=N+1, ncol=k+degp+1)
vare=matrix(1, nrow=N+1, ncol=1)
varu=matrix(1, nrow=N+1, ncol=1)
Au=0
```

```

Bu=0
Aupost=(k/2)+Au
Ae=0
Be=0
Aepost=(n/2) + Ae
for(i in 1:N){
  print(i)
  D_inv=diag(c(rep(1/betav,degp+1),rep(1/varu[i],k)))
  gam=(B/vare[i])+D_inv
  sigma=solve(gam,diag(k+degp+1) ,tol=10^-50)
  mu=(sigma%*%CT%*%y)/vare[i]
  gamma[i+1,]=mvrnorm(1, mu, sigma)
  #Sample from inverse gamma of random effects
  U=gamma[i+1,(degp+2):(k+degp+1)]
  Bupost=Bu +0.5*(t(U)%*%U)
  varu[i+1]=rinvgamma(1,Aupost,Bupost)
  #Sample from inverse gamma of residuals
  G=gamma[i+1,]
  W=y-C%*%G
  Bepost=Be + 0.5*(t(W)%*%W)
  vare[i+1]=rinvgamma(1,Aepost,Bepost)
}
burn=2000
gammapost=apply(gamma[burn:N,],2,mean)
fit1=C%*%gammapost
#Credible Intervals
gammaa=gamma[burn:N,]
fitss=matrix(0, nrow=(N-burn+1),ncol=n)
for(b in 1:(N-burn+1)){
  fitss[b,]=C%*%gammaa[b,]
}
ci=apply(fitss,2, quantile,c(0.025,0.975))
lines(x, ci[1,], lty=4,lwd=1)
lines(x, ci[2,], lty=4,lwd=1)
#Plot
plot(x,y,col=6)
lines(x,fit1,lwd=2,col=4)
lines(x, ci[1,], lty=3,lwd=2,)
lines(x, ci[2,], lty=3,lwd=2)

```

B.2 BAPS Code

```
#Functions#####

#Natural Spline =====
#t=seq(0,1, length=n)
# B=c(20,4,6,11,6)
# nudos=c(0,0.2,0.6,0.7)
# Z=ns(t, knots=nudos, Boundary.knots=c(0,1))
# func=Z%*%B
# y=func+rnorm(n,0,.9)

#Hump function =====
t=seq(-2,2,length=n)
y=sin(t)+2*exp(-30*t^2)+ rnorm(n,0,0.3)
#####

mt=40
ms=10
degp=2
degq=1
mss=ms+degq
#t knots

tknots = quantile(unique(t), seq(0,1, length=(mt+2))[-c(1,(mt+2))])
Z = outer(t, tknots, "-")
Z = Z*(Z>0)
Z=Z^degp
dimnames(Z)=NULL

#s knots

sknots = quantile(unique(tknots), seq(0,1, length=(ms+2))[-c(1,(ms+2))])
sknots=c(0,sknots)
#sknots=sort(sknots, decreasing=FALSE)
Z_gamma = outer(tknots, sknots, "-")
Z_gamma = Z_gamma*(Z_gamma>0)
Z_gamma=Z_gamma^degq
dimnames(Z_gamma)=NULL

##For degq==0 Piecewise constants#####
#tknots = quantile(unique(t), seq(0,1, length=(mt+2))[-c(1,(mt+2))])
#Z=matrix(0,nrow=n,ncol=mt)
#for(a in 1:mt){
  Z[,a]=t-tknots[a]
#}
#Z=Z*(Z>0)
#Z=Z^degp
#sknots = quantile(unique(tknots), seq(0,1, length=(ms+2))[-c(1,(ms+2))])
```

```

#Z_gamma=matrix(0,nrow=mt,ncol=mss)
#for(k in 1:mss){
# Z_gamma[,k]=tknots>sknots[k]
#}
#####
#Data preparation
betasqrt=100
c=c(1,1)
X=matrix(1,nrow=n, ncol=degp+1)
for(d in 1:degp){
  X[,d+1]=t^d
}
T=cbind(X,Z)
Ttran=t(T)
R=Ttran%*%T
theta=matrix(1, nrow=N+1, ncol=mt+degp+1)
tau=matrix(1, nrow=N+1, ncol=1)
xi=matrix(1, nrow=2, ncol=N+1)
rho=matrix(1, nrow=2, ncol=N+1)
b_gamma=matrix(0,nrow=N+1, ncol=mss)
alpha=c()
trustobj=function(param,b,tau,xi1,xi2,Z_gamma){
  gamma=Z_gamma%*%param
  e=exp(gamma)
  D_gamma=diag(c(e))
  one=rep(1,mt)
  sum1=matrix(0, nrow=mss, ncol=1)
  sum2=matrix(0,mss,mss)
  for(j in 1:mt){
    coef=as.vector(exp(t(Z_gamma[j,])%*%param))
    sum1=sum1 + (b[j]^2)*coef*Z_gamma[j,]
    sum2=sum2 + (b[j]^2)*coef*Z_gamma[j,]%*%t(Z_gamma[j,])
  }
  f=0.5*(sum(gamma)-tau*xi1*t(b)%*%D_gamma%*%b-tau*xi1*xi2*t(param)%*%param)
  g=0.5*t(Z_gamma)%*%one -0.5*tau*xi1*sum1-tau*xi1*xi2*param
  H=-0.5*tau*xi1*sum2-tau*xi1*xi2*diag(mss)
  list(value = f, gradient = g, hessian = H)
}
for(i in 1:N){
  print(i)
  gamma=Z_gamma%*%b_gamma[i,]
  e=exp(gamma)

```

```

D_gamma=diag(c(e))
para=tau[i]*xi[1,i]
Lamda_y=diag(c(rep(1/betasqrt, degp+1), para*e))
Q_theta=tau[i]*R + Lamda_y
Q_theta_inv=chol2inv(chol(Q_theta))
#solve(Q_theta,tol=1.6912e-45)
mu_theta=tau[i]*(Q_theta_inv**Ttran**y)
#Drawing theta
theta[i+1,]=mvrnorm(1,mu_theta, Q_theta_inv)
#Drawing tau
b=theta[i+1,(degp+2):(mt+degp+1)]
U=theta[i+1,]
V=y-T**U
BDB=t(b)**D_gamma**b
Bsqr=t(b_gamma[i,])**b_gamma[i,]
tau_shape=0.5*(n+mt+mss)
tau1=t(V)**V
tau2=xi[1,i]*(BDB)
tau3=xi[1,i]*xi[2,i]*(Bsqr)
tau_rate=0.5*(tau1+tau2+tau3)
tau[i+1]=rgamma(1, tau_shape ,scale=1/tau_rate)
#Drawing xi1
xi1_shape=0.5*(mt+mss)+1
xi11=0.5*tau[i+1]*(BDB)
xi12=0.5*tau[i+1]*xi[2,i]*(Bsqr)
xi1_rate=xi11+xi12+rho[1,i]
xi[1,i+1]=rgamma(1,xi1_shape, scale=1/xi1_rate)
#Drawing xi2
xi2_shape=0.5*(mss) + 1
xi2_rate=0.5*(tau[i+1]*xi[1,i+1]*Bsqr) +rho[2,i]
xi[2,i+1]=rgamma(1,xi2_shape, scale=1/xi2_rate)
#Drawing rho_1 and rho_2
rho1_rate=xi[1,i+1] + c[1]
rho[1,i+1]=rgamma(1, 2, scale=1/rho1_rate)
rho2_rate=xi[2,i+1] + c[2]
rho[2,i+1]=rgamma(1, 2,scale=1/rho2_rate)
m=trust(trustobj,parinit=rep(0,mss),b=theta[i+1,(degp+2):(mt+degp+1)],tau=tau[i+1],xi1=xi[1,i+1],xi2=xi[2,i+1],
Z_gamma=Z_gamma,blather=TRUE,rinit=100,rmax=100,minimize=FALSE)
b_gamma_max=m$argument
H=m$hessian
b_gamma[i+1,]=mvrnorm(1,b_gamma_max,-solve(H))
#Metropolis-Hastings - Drawing b_gamma

```

```

gamma_prop=Z_gamma%*%b_gamma[i+1,]
D_gamma_prop=diag(c(exp(gamma_prop)))
log_proposal_curr=-0.5*(t(b_gamma[i,]-b_gamma_max)%*%(-H)%*%(b_gamma[i,]-b_gamma_max))
log_proposal_prop=-0.5*(t(b_gamma[i+1,]-b_gamma_max)%*%(-H)%*%(b_gamma[i+1,]-b_gamma_max))
log_lik_gamma_curr=0.5*(sum(gamma)-
tau[i+1]*xi[1,i+1]*t(b)%*%D_gamma%*%b-tau[i+1]*xi[1,i+1]*xi[2,i+1]*(t(b_gamma[i,])%*%b_gamma[i,]))
log_lik_gamma_prop=0.5*(sum(gamma_prop)-
tau[i+1]*xi[1,i+1]*t(b)%*%D_gamma_prop%*%b-tau[i+1]*xi[1,i+1]*xi[2,i+1]*(t(b_gamma[i+1,])%*%b_gamma[i+1,]))
M_H_ratio=log_lik_gamma_prop+log_proposal_curr-log_lik_gamma_curr-log_proposal_prop
alpha[i]=min(c(1,exp(M_H_ratio)))
u=runif(1)
if(u > alpha[i]){
  b_gamma[i+1,] = b_gamma[i,]
}
}
#mean(alpha)
burnin=2000
theta_post=apply(theta[burnin:N,], 2, mean)
BAPS_fit=T%*%theta_post
#Plots#####
plot(t,y,type="n")
# lines(t,sin(t)+2*exp(-30*t^2), lty=1)
#lines(t,func)
lines(x,BAPS_fit,lty=2,col='red')
#Credible Intervals#####
burn=2000
theta1=theta[burn:N,]
fits=matrix(0, nrow=(N-burn+1),ncol=n)
for(b in 1:(N-burn+1)){
  fits[b,]=T%*%theta1[b,]
}
ci=apply(fits,2, quantile,c(0.025,0.975))
lines(t, ci[1,], lty=4,lwd=1)
lines(t, ci[2,], lty=4,lwd=1)

```

B.3 BAPS Code (Whittle estimate)

B.3.1 Non-Adaptive Whittle

```
library(TSA)
```

```

library(tseries)
library(trust)
library(MASS)
library(MCMCpack)
library(mvtnorm)
library(mnormt)
dat=scan("/Users/luismora/Desktop/thesisdata.txt")
dat=dat-mean(dat)
#Trough Spectral Density: phi=c(0.1,0.4)#####
#Peak Spectral Density: phi=c(1.5, -0.75)#####
phi=c(0.1,0.4)
ts=arima.sim(model=list(ar=phi),n)
y=ts-mean(ts)
#True Spectral Density for AR(2) process#####
phi1=phi[1]^2
phi2=phi[2]^2
truesd=(1^2)/(1+phi1 + phi2 - 2*phi[1]*(1-phi[2]) *cos(2*pi*f)-2*phi[2]*cos(4*pi*f))
#####
n=length(dat)# or length(y)
plot(dat,type="l")
M=floor((n-1)/2);M
tran=fft(dat)/sqrt(n);tran
I=as.vector(abs(tran)^2)
A=I[2:(M+1)]
length(A)
plot(A,type='l')
f=(1:M)/n
#####
N=2000
burn=500
mt=30
degp=2
#####
When using cosine basis functions, let degp == 0
#####
#Data=====
# X=matrix(1,nrow=M, ncol=degp+1)
# for(k in 1:degp){
# X[,k+1]=x^k
# }
#t knots
tknots = quantile(unique(f), seq(0,1, length=(mt+2))[-c(1,(mt+2))])

```

```

Z=matrix(0,nrow=M,ncol=mt)
for(s in 1:mt){
  Z[,s]=f-tknots[s]
}
Z=Z*(Z>0)
Z=Z^degp
#
X=cbind(rep(1,M),f,f^2)
C=cbind(X,Z)
#
# C=matrix(1,nrow=length(f), ncol=(mt+1))
# for(j in 2:(mt+1)){
#   C[,j]=sqrt(2)*(cos((j-1)*pi*f)/(pi*(j-1)))
# }
# dim(C)
#=====
trustthe=function(the,Lambday,D,P){
  f=sum(D%*%the- exp(D%*%the)*P) - 0.5*t(the)%*%Lambday%*%the
  g=t(D)%*%(1- P*exp(D%*%the)) - Lambday%*%the
  h=-t(D)%*%diag(as.vector(P*exp(D%*%the)))*%D - Lambday
  list(value = f, gradient = g, hessian = h)
}
#Fixed Parameters=====
G=10^5
nu=2
#Parameters=====
theta=matrix(1,nrow=N+1,ncol=mt+degp+1)
delta=matrix(1,nrow=N+1,ncol=1)
g1=matrix(1,nrow=N+1, ncol=1)
alpha1=c()
sigmasqrt=100
gshape=0.5*(nu+1)
delta.a=0.5*(mt+nu)
for(i in 1:N){
  print(i)
  Lamda_y=diag(c(rep(1/sigmasqrt,degp+1), rep(delta[i],mt)))
  z=trust(trustthe,parinit=rep(0,mt+degp+1), Lambday=Lamda_y,D=C,P=A, blather=TRUE,rinit=100,rmax=100,minimize=FALSE)
  thetamax=z$argument
  h=-z$hessian
  #Drawing theta
  theta[i+1,]=mvrnorm(1,thetamax,chol2inv(chol(h)))
  #Metropolis-Hastings Step

```

```

whittle_lik_curr=sum(C%*%theta[i,]- exp(C%*%theta[i,])*A) - 0.5*t(theta[i,])%*%Lamda_y%*%theta[i,]
whittle_lik_pro=sum(C%*%theta[i+1,]- exp(C%*%theta[i+1,])*A) - 0.5*t(theta[i+1,])%*%Lamda_y%*%theta[i+1,]
proposal_theta_curr= -0.5*(t(theta[i,] - thetamax)%*(h)%*(theta[i,] - thetamax))
proposal_theta_prop= -0.5*(t(theta[i+1,] - thetamax)%*(h)%*(theta[i+1,] - thetamax))
MH_ratio= whittle_lik_pro + proposal_theta_curr - whittle_lik_curr - proposal_theta_prop
alpha1[i]=min(c(1,exp(MH_ratio)))
u1=runif(1)
if(u1 > alpha1[i]){
  theta[i+1,] = theta[i,]
}
b=theta[i+1,(2+degp):(mt+degp+1)]
#When using Cosine Basis : theta[i+1,2:(mt+1)]
#Drawing delta and g1
delta.b=0.5*t(b)%*%b + nu/g1[i]
delta[i+1]=rgamma(1,shape=delta.a,scale=1/delta.b)
g1b=nu*delta[i+1] + (1/G)^2
g1[i+1]=1/rgamma(1,shape=gshape,scale=1/g1b)
}
k=kernel('daniell',m=25)
s=spec(dat, kernel=k, ci.plot=TRUE, ci.col=NULL,log='yes',demean=FALSE)
lines(s$freq,log(s$spec),lty=3,lwd=2)
lines(s$freq,log(s$spec)+log(102/qchisq(.025,102)),col="navy",lty=4,lwd=1)
lines(s$freq,log(s$spec)+log(102/qchisq(.975,102)),col="navy",lty=4,lwd=1)
#Plots and Fit#####
mean(alpha1)
theta_post=apply(theta[burn:N,], 2, mean)
fit=-C%*%theta_post
plot(f,log(A),lwd=0.5)
lines(f,fit,lty=6,lwd=1)
#Credible Intervals#####
theta1=theta[burn:N,]
fits=matrix(0, nrow=(N-burn+1),ncol=M)
for(h in 1:(N-burn+1)){
  fits[h,]=-C%*%theta1[h,]
}
cii=apply(fits,2, quantile,c(0.025,0.975))
lines(f, cii[1,],lty=3,lwd=1)
lines(f, cii[2,],lty=3,lwd=1)

```

B.3.2 Adaptive Whittle

```
#=====
```

```

N=2000
burn=500
mt=50
ms=15
degp=2
degq=0
df=4
mss=ms+degq
#####
  When using cosine basis functions, let degp == 0
#####
#Data=====
# X=rep(1,M)
# t knots
# tknots = quantile(unique(f), seq(0,1, length=(mt+2))[-c(1,(mt+2))])
# Z = outer(f, tknots, "-")
# Z = Z*(Z>0)
# Z=Z^degp
# dimnames(Z)=NULL
# Z=matrix(0,nrow=M,ncol=mt)
# for(s in 1:mt){
#   Z[,s]=f-tknots[s]
# }
# Z=Z*(Z>0)
# Z=Z^degp
tknots = quantile(unique(f), seq(0,1, length=(mt+2))[-c(1,(mt+2))])
Z=matrix(0,nrow=M,ncol=mt)
for(a in 1:mt){
  Z[,a]=f-tknots[a]
}
Z=Z*(Z>0)
Z=Z^degp
#s knots
sknots = quantile(unique(tknots), seq(0,1, length=(ms+2))[-c(1,(ms+2))])
# # sknots=c(0,sknots)
# Z_gamma = outer(tknots, sknots, "-")
# Z_gamma = Z_gamma*(Z_gamma>0)
# # Z_gamma=Z_gamma^degq
# dimnames(Z_gamma)=NULL
# Z_gamma
Z_gamma=matrix(0,nrow=mt,ncol=mss)
for(k in 1:mss){

```

```

Z_gamma[,k]=tknots>sknots[k]
}
#Truncated Polynomials
# X=matrix(1,nrow=M, ncol=degp+1)
# for(k in 1:degp){
#   X[,k+1]=x^k
# }
X=cbind(rep(1,M),f,f^2)
C=cbind(X,Z)
#Cosine Basis Functions
# C=matrix(1,nrow=length(f), ncol=(mt+1))
# for(j in 2:(mt+1)){
#   C[,j]=sqrt(2)*(cos((j-1)*pi*f)/(pi*(j-1)))
# }
#=====
trustthe=function(the,Lambday,D,P){
  f=sum(D%*%the- exp(D%*%the)*P) - 0.5*t(the)%*%Lambday%*%the
  g=t(D)%*%(1- P*exp(D%*%the)) - Lambday%*%the
  h=-t(D)%*%diag(as.vector(P*exp(D%*%the)))*%D - Lambday
  list(value = f, gradient = g, hessian = h)
}
trustgam=function(param,b,del,et,Z_gamma){
  gamma=Z_gamma%*%param
  e=exp(gamma)
  D_gamma=diag(c(e))
  one=rep(1,mt)
  sum1=matrix(0, nrow=mss, ncol=1)
  sum2=matrix(0,mss,mss)
  for(j in 1:mt){
    coef=as.vector(exp(t(Z_gamma[j,])%*%param))
    sum1=sum1 + (b[j]^2)*coef*Z_gamma[j,]
    sum2=sum2 + (b[j]^2)*coef*Z_gamma[j,]%*%t(Z_gamma[j,])
  }
  f=0.5*(sum(gamma)- del*t(b)%*%D_gamma%*%b- et*t(param)%*%param)
  g=0.5*t(Z_gamma)%*%one - 0.5*del*sum1-et*param
  h=-0.5*del*sum2 - et*diag(mss)
  list(value = f, gradient = g, hessian = h)
}
#Fixed Parameters=====
G=10^5
nu=2
#Parameters=====

```

```

theta=matrix(1,nrow=N+1,ncol=mt+degp+1)
b_gamma=matrix(1,nrow=N+1,ncol=mss)
delta=matrix(1,nrow=N+1,ncol=1)
g1=matrix(1,nrow=N+1, ncol=1)
g2=matrix(1,nrow=N+1, ncol=1)
eta=matrix(1,nrow=N+1,ncol=1)
alpha1=c()
alpha2=c()
sigamsqr=100
gshape=0.5*(nu+1)
eta.a=0.5*(mss+nu)
delta.a=0.5*(mt+nu)
for(i in 1:N){
  print(i)
  gamma=Z_gamma%*%b_gamma[i,]
  D_gamma=diag(c(exp(gamma)))
  Lamda_y=diag(c(rep(1/sigamsqr,degp+1), delta[i]*exp(gamma)))
  z=trust(trustthe,parinit=rep(0,mt+degp+1), Lambday=Lamda_y,D=C,P=A, blather=TRUE,rinit=100,rmax=100,minimize=FALSE)
  thetamax=z$argument
  h=-z$hessian
  #Drawing theta
  #Sigma=(df/(df-2))*chol2inv(chol(h))
  theta[i+1,]=mvrnorm(1,thetamax,chol2inv(chol(h)))
  #rmvt(1,delta=thetamax,sigma=Sigma,df=4)
  #Metropolis-Hastings Step
  whittle_lik_curr=sum(C%*%theta[i,]- exp(C%*%theta[i,])*A) - 0.5*t(theta[i,])%*%Lamda_y%*%theta[i,]
  whittle_lik_pro=sum(C%*%theta[i+1,]- exp(C%*%theta[i+1,])*A) - 0.5*t(theta[i+1,])%*%Lamda_y%*%theta[i+1,]
  #proposal_theta_curr=dmvtn(theta[i,],delta=thetamax,sigma=Sigma,df=4,log=TRUE)
  #proposal_theta_prop=dmvtn(theta[i+1,],delta=thetamax,sigma=Sigma,df=4,log=TRUE)
  proposal_theta_curr= -0.5*(t(theta[i,] - thetamax)%*%(h)%*(theta[i,] - thetamax))
  proposal_theta_prop= -0.5*(t(theta[i+1,] - thetamax)%*%(h)%*(theta[i+1,] - thetamax))
  MH_ratio= whittle_lik_pro + proposal_theta_curr - whittle_lik_curr - proposal_theta_prop
  alpha1[i]=min(c(1,exp(MH_ratio)))
  u1=runif(1)
  if(u1 > alpha1[i]){
    theta[i+1,] = theta[i,]
  }
  b=theta[i+1,(degp+2):(mt+degp+1)]
  #When using Cosine basis: theta[i+1,2:(mt+1)]
  #Drawing b_gamma
  o=trust(trustgam,parinit=rep(0,mss),b=b,del=delta[i],et=eta[i],Z_gamma=Z_gamma,blather=TRUE,rinit=100,rmax=100,minimize=FALSE)
  b_gamma_max=o$argument

```

```

H=-o$hessian
#Drawing bgamma
b_gamma[i+1,]=mvrnorm(1,b_gamma_max,chol2inv(chol(H)))
#Metropolis-Hastings - Drawing b_gamma
gamma_prop=Z_gamma%*b_gamma[i+1,]
D_gamma_prop=diag(c(exp(gamma_prop)))
log_proposal_curr=-0.5*(t(b_gamma[i,]-b_gamma_max)%*(H)%*(b_gamma[i,]-b_gamma_max))
log_proposal_prop=-0.5*(t(b_gamma[i+1,]-b_gamma_max)%*(H)%*(b_gamma[i+1,]-b_gamma_max))
log_lik_gamma_curr=0.5*(sum(gamma)-delta[i]*t(b)%*D_gamma%*b - eta[i]*t(b_gamma[i,])%*b_gamma[i,])
log_lik_gamma_prop=0.5*(sum(gamma_prop)-delta[i]*t(b)%*D_gamma_prop%*b-eta[i]*t(b_gamma[i+1,])%*b_gamma[i+1,])
M_H_ratio=log_lik_gamma_prop+log_proposal_curr-log_lik_gamma_curr-log_proposal_prop
alpha2[i]=min(c(1,exp(M_H_ratio)))
u2=runif(1)
if(u2 > alpha2[i]){
  b_gamma[i+1,] = b_gamma[i,]
}
#Drawing delta and g1
delta.b=0.5*t(b)%*D_gamma%*b + nu/g1[i]
delta[i+1]=rgamma(1,shape=delta.a,scale=1/delta.b)
g1b=nu*delta[i+1] + (1/G)^2
g1[i+1]=1/rgamma(1,shape=gshape,scale=1/g1b)
#Drawing eta and g2
eta.b=0.5*t(b_gamma[i+1,])%*b_gamma[i+1,]+nu/g2[i]
eta[i+1]=rgamma(1,shape=eta.a,scale=1/eta.b)
g2b=nu*eta[i+1] + (1/G)^2
g2[i+1]=1/rgamma(1,shape=gshape,scale=1/g2b)
}
#####
k=kernel('daniell',m=25)
s=spec(dat, kernel=k, ci.plot=TRUE, ci.col=NULL,log='yes',demean=FALSE)
plot(f,log(A),col='wheat')
lines(s$freq,log(s$spec),lty=3,lwd=2)
lines(s$freq,log(s$spec)+log(102/qchisq(.025,102)),col="deepskyblue",lty=4,lwd=1)
lines(s$freq,log(s$spec)+log(102/qchisq(.975,102)),col="deepskyblue",lty=4,lwd=1)
#Plot and Fits#####
cbind(mean(alpha1),mean(alpha2))
theta_post=apply(theta[burn:N,], 2, mean)
BAPS_fit=-C%*theta_post
plot(f,log(A),type='n',lwd=1)
lines(f,BAPS_fit, lty=1,lwd=1)
#Credibel Intervals#####
theta=theta[burn:N,]

```

```

fits=matrix(0, nrow=(N-burn+1),ncol=M)
for(h in 1:(N-burn+1)){
  fits[h,]=-C*%theta[h,]
}
ci=apply(fits,2, quantile,c(0.025,0.975))
lines(f, ci[1,],lty=2,lwd=1)
lines(f, ci[2,],lty=2,lwd=1)

```

Curriculum Vitae

Luis Angel Mora was born on December 03, 1988. The first son of Maria de la Luz Mora and Roumaldo Mora, he graduated from Gadsden High School, Anthony, New Mexico, in the spring of 2007. Luis entered New Mexico State University in the fall of 2007. While pursuing his bachelor's degree in mathematics he worked as a mathematics tutor and library assistant at New Mexico State University. Luis received his bachelor's degree in mathematics with a minor emphasis in management in the summer of 2012.

In the fall of 2013, he entered the Graduate School of The University of Texas at El Paso. While pursuing a master's degree in Statistics he worked as a Teaching Assistant and supervised tutors at the mathematics tutoring center.

Permanent address: 653 Mesilla View Drive
Chaparral, New Mexico 88081