Title	Derivative Estimation in Noisy Data; an Additive Penalty P-Spline Approach
Author(s)	Simpkin, Andrew
Publication Date	2010-12-20
Item record	http://hdl.handle.net/10379/2002

Derivative Estimation in Noisy Data; an Additive Penalty P-Spline Approach

Andrew Simpkin

PhD in Statistics

School of Mathematics, Statistics and Applied Mathematics

National University of Ireland, Galway

December 2010

Supervised by Dr. John Newell

Head of School Dr. Ray Ryan

Contents

List of Figures

List of Tables

Acknowledgements

This thesis would not have been possible without my supervisor and friend John Newell. Thank you for being a constant source of knowledge and encouragement. I would also like to thank John Hinde for all his advice and guidance over the past three years. I am indebted to IRCSET for funding this research and to Paul Eilers, Jutta Gampe and Giancarlo Camarda for their insightful contributions along the way.

I would like to dedicate this thesis to my parents Edward and Miriam, for everything.

Abstract

In many situations it is of primary interest to estimate the rate of change of the relationship between response and explanatory variables. In this thesis derivative estimation using spline smoothing is explored. A review of derivative estimates found as a by product of several popular spline smoothing techniques is provided. Concerns with these estimates are raised and an additive penalty method utilising the attractive properties of P-Splines is introduced. This approach is shown to improve on semiparametric and P-Spline derivative estimates in simulated smoothing scenarios. Variability bands for derivative estimates are developed for the additive penalty and P-Spline methods with these tested for coverage and precision in further simulations. Motivating examples in environmental, biomedical and astronomical applications are revisited throughout the thesis.

Introduction

In a range of disciplines it is often the case that, when analysing data, the derivative, or rate of change, of observed data is of primary interest. In the regression model $y = f(x) + \epsilon$ for example, one is often not interested in the underlying function f itself, but rather in the relative change $\frac{d}{dx}$ of f when increasing or decreasing x by a small value dx. In the situation where data are observed over time, the first derivative will correspond to velocity, the second to acceleration. Derivatives f' are also used in asymptotic approximations to obtain confidence intervals and optimal bandwidths for example. A further field of application for derivative estimation is change point problems. For instance, when analysing blood lactate data of elite athletes, one is interested in the workload at which the lactate level suddenly rises, which can be detected by finding the maximum of the second derivative (Newell et al. [39]).

A popular tool for derivative estimation is spline smoothing, with a large number of variants (e.g. smoothing splines, P-Splines) being available. Choice of smoothing parameter for these methods when the derivative is of primary concern is an area where there seems to be no consensus for the optimal choice. Generally the smoothing parameter is selected based on optimising $\hat{f}$ and not the derivative, which may lead to considerable undersmoothing.

The usual way of estimating derivatives is to take the derivatives as a by-product of the estimate $\hat{f}$. In other words, if $\hat{f}$ is an estimate of f, one considers $\frac{d^l}{dx^l}\hat{f}$ as an estimator of the lth derivative $f^{(l)}$, $l = 0, 1, 2, \ldots$. Several authors have pursued this idea (e.g. Heckman & Ramsay [21]), using splines with or without penalisation. The simplicity of this idea led to several papers which gave the impression that the entire issue of so called ‘nonparametric derivative estimation’ was solved. To a certain extent this has led to a lack of recent research on the topic, which is unfortunate as many open questions remain. Derivative estimates tend to be much less robust to horizontal and vertical outliers than estimators of the regression function. These estimates also suffer from boundary effects, Ramsay [42] noted that ‘typically one sees derivatives go wild at the extremes, and the higher the derivative, the wilder the behavior’. Further problems arise when it comes to smoothing parameter selection, where automatic routines, such as cross-validation, can be ‘poor guides’.

The main goals of this PhD thesis are

- To provide a comprehensive review of derivative estimation for noisy data.
- To outline the challenges faced in obtaining accurate derivative estimates when a nonlinear relationship between explanatory and response exists.
- To compare the performance of current methods for derivative estimation.
- To develop an approach to derivative estimation that achieves improved performance over current methods.
- To establish suitable variance estimates for these estimators in order to produce reliable variability bands.

The thesis is arranged as follows:

Chapter 1 introduces the datasets that have motivated this research. A nonlinear relationship between explanatory and response variables and considerable noise are present in each. The main question of interest for each dataset can be answered using a first and/or second derivative estimate. In the Winter Nutrient and Scottish Bird Count illustrations first derivative estimates are required to estimate whether significant changes in levels of nutrients and/or bird count are present over a period of time and to estimate when these changes transpire. In the Blood Lactate illustration a second derivative estimate is used to find an objective endurance marker for comparisons within and between athletes. The Astronomical data is somewhat complex in its background; it involves finding derivative estimates of a convolution of uncertain variables of different lengths as a means to estimate gravitational mass density distributions, which are otherwise impossible to measure.

In Chapter 2 basic methods for estimating the rate of change are discussed. Considering the situation of data (x_i, y_i) for $i = 1, \ldots, n$, the simplest approach to derivative estimation is to calculate the differences between consecutive response observations $y_i - y_{i-1}$ and divide by differences in corresponding predictor values $x_i - x_{i-1}$ for $i = 2, \ldots, n$. It becomes obvious that estimates from such first order differences are highly unstable for data in which a nonlinear relationship between the variables is apparent. A small simulation to display these shortcomings is presented. Some simple modelling based approaches to derivative estimation using linear models are then described.

Focus then moves in Chapter 3 to more complex modelling techniques capable of handling nonlinear relationships in observed data, i.e. smoothing. Derivative estimates are found as a by-product of a smooth estimate of the relationship between observed variables. Spline smoothing methods are motivated and several types of smoother are introduced including smoothing splines, mixed model smoothing and P-Splines. Application of P-Splines to the case of a derivative estimation problem involving a count response is detailed. A literature review of derivative estimation using spline smoothers is included. These methods are applied to the motivating applications and the resulting estimates are compared.

Chapter 4 directs attention solely to derivative estimation using P-Splines. P-Splines have become widely popular since their introduction (Eilers & Marx [12]). As far as derivative estimation is concerned, little work has been done to study performance of derivative estimates using P-Splines. Several simulation studies into the effects on P-Spline derivative estimates of sample size, variability, smoothing parameter selection and P-Spline components are provided.

Chapter 5 examines possible adjustments to the P-Spline method for derivative estimation which may help to improve goodness of fit. One such method is the use of an extra additive penalty term in the P-Spline framework. Motivation for this additive penalty model is provided. This method introduces yet more choices to the already choice-laden P-spline smoothing framework. These choices are explained, sensitivity to choice compared, and optimal decisions are made. R code for this method is included in the Appendix. Since one of the datasets, the Scottish Bird Count data, contains a response of counts the additive penalty method is extended in this thesis to deal with a Poisson response and derivative estimates are shown to be easily calculated. The performance of this modified appraoch in derivative estimation is compared to the methods of semiparametric regression and P-Splines.

Chapter 6 investigates further aspects of derivative estimates. Variability bands have not been fitted to derivative estimates from a P-Spline fit in the literature and so techniques for construction of these bands, as well as for the additive penalty method, are developed and tested rigorously for coverage. The variability bands developed for P-Spline and additive penalty derivative estimates are compared with bands from mixed model smoothing and those found through bootstrap resampling of residuals. Estimating certain features of a derivative is explored through some simulations using the additive penalty method. These findings are then compared with those using other spline smoothing techniques.

Finally, in Chapter 7, a summary of the main results provided is given, along with proposals for further research in the area of derivative estimation. A final conclusion for each of the motivating examples is also presented.

Chapter 1

Motivating Data

In this Chapter the four main motivating illustrations of analyses requiring derivative estimation are introduced. Each example poses a distinct problem involving derivative estimation in some sense. Later the functions which are used for simulations are introduced with reasoning provided for their choice.

1.1 Winter Nutrients Data

Researchers at the Marine Institute of Ireland collected water samples from the western Irish Sea between late November and early February in 1990/1991. EU member states measure winter nutrient concentration in marine waters as an indicator of trophic status. Nutrient inputs to the Irish Sea are numerous. By far the largest input, in terms of quantity, results from the flow of large volumes of water through St. George's Channel from the Celtic Sea.

Levels of two nutrients, Phosphate and NTRZ (which is a combination of Nitrate and Nitrite), were measured in samples taken from a depth of 3 metres (see Figures 1.1 and 1.2).

The main goals for this study were to model the rate of change of the levels of contamination over the course of the Winter and to estimate the times at which significant changes in contamination were found. Once a reliable first derivative estimate, along with confidence or variability bands, has been achieved then significant zero crossings (i.e. confidence/variability bands fully above or below zero of the derivative) indicate significant increases or decreases of contamination relative to the day of measurement. Once a significant increase or decrease has

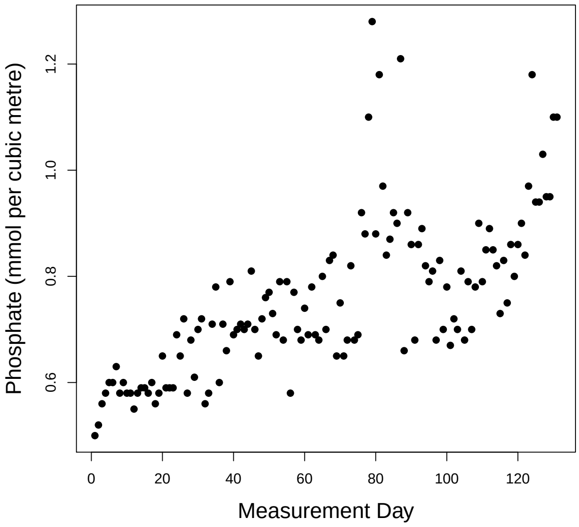


Figure 1.1: Phosphate contamination of the Irish Sea measured on 131 days during 1990 and 1991.

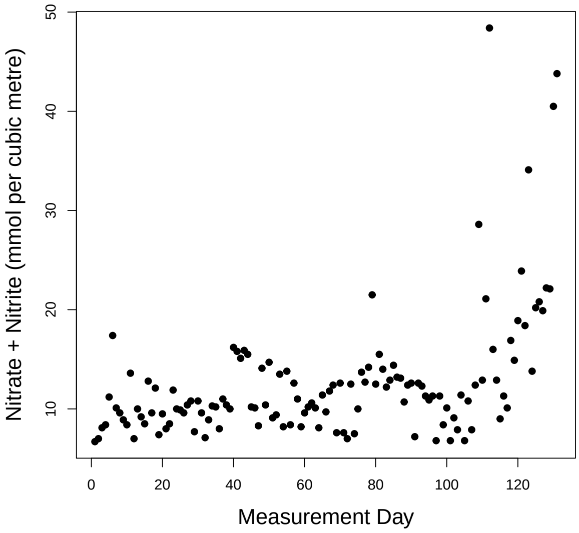


Figure 1.2: NTRZ contamination of the Irish Sea measured on 131 days during 1990 and 1991.

been found it is of interest to the researchers at what point in time it occurs.

1.2 Scottish Bird Count Data

Counts of 11 wetland bird species were collected annually at specified wetland locations across Scotland from 1974 to 2004 as part of an environmental study into the relationship between climate change and bird numbers. Water and wetland features have determined where people have settled, and how communities and economies have grown. The vision is to see healthy and biologically diverse rivers, lakes and wetlands in a landscape managed for the sustainable use of water. Healthy numbers of these birds, such as the Grey Plover (Figure 1.3), is a strong indicator of flourishing wetlands.

Figure 1.3: Winter plumaged Grey Plover.

Figure 1.4 displays 31 counts of Grey Plover taken annually between 1974 and 2004. The primary aim is to determine whether there has been a significant decrease in count over time. For conciseness, the results of the analysis of the count of Grey Plover only are contained in later chapters, it is assumed that the analysis carried out here is transferable to the other ten species where data are available. A Mann-Kendall test (Kendall [25]) could be performed in order to test for an overall trend present in the data. The Mann-Kendall test works by first identifying the sign of the change between consecutive observations, i.e. either -1, 0 or 1. The Mann-Kendall statistic S is the sum of all signs of slopes. A high value of S indicates an increasing trend, a low value a decreasing trend and a small value indicates no trend. However, a global trend indicator is not of interest here. It is necessary to find an accurate representation of the rate of change of the count relative to the year of measurement, i.e. the first derivative of the underlying function which describes the behaviour of the data. Finding an accurate first derivative estimate of the observed count along with variability bands would allow significant rate of change to be found in a similar fashion to the Winter Nutrients example.

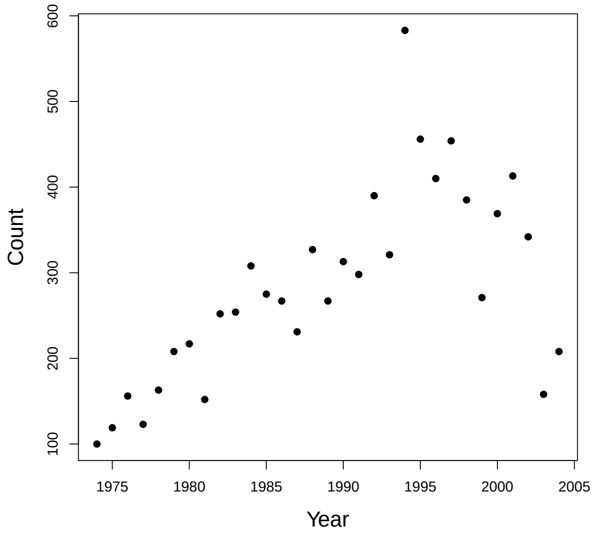


Figure 1.4: Count of Grey Plover 1974 to 2004.

The count of Grey Plover in Figure 1.4 clearly has a nonlinear relationship with Year. Therefore, simple linear regression methods are not suitable to model this relationship unless a transformation of count is taken. A polynomial regression could be used, however, the use of polynomial regression, as will be seen, has limitations for robust analysis of nonlinear relationships. Furthermore, since the data come in the form of counts, observations of the response variable should be treated as representations of a Poisson distributed random variable. Methods to estimate derivatives of a Poisson generated response are presented in Chapters 3 and Chapter 5.

1.3 Blood Lactate Data

Blood lactate testing is often used as a measure of endurance in elite athletes. Such a test involves collecting blood lactate at incremental workloads on a treadmill. A blood lactate test was performed on a group of 23 elite athletes in order to estimate the fitness levels of each athlete. Figure 1.5 shows a single athlete's blood lactate measured at 10 incremental workloads on a treadmill. The data are a discrete approximation to an underlying continuous system, i.e. the lactate response to incremental exercise curve, or 'lactate curve' for short.

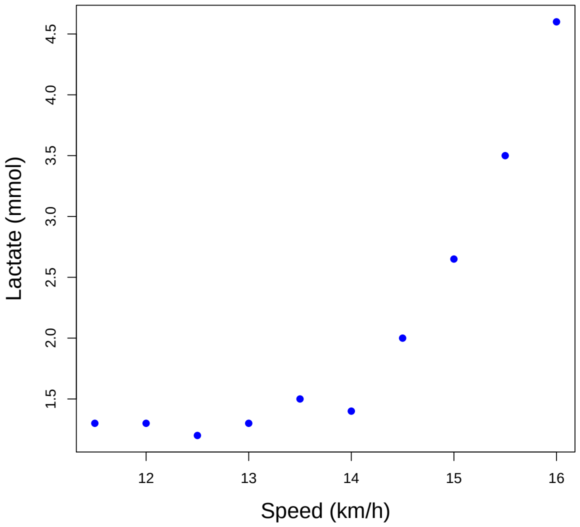


Figure 1.5: Blood lactate data for one individual measured at 10 speeds on a treadmill.

Several features of an individual's lactate data have been considered to be good predictors of endurance performance to track changes in fitness over time. Typically these features, or endurance markers, are used to monitor changes in aerobic fitness, set training regimes, and predict endurance performance. However, determination of these markers can be problematic.

Lundberg et al. [31] fit two linear splines which join at the location (workload or speed) of the lactate threshold (LT), a unique point at which lactate production shifts from an anaerobic to an aerobic state. The location where the two linear splines join is known as a knot. Least squares is used to estimate the knot location and the slopes and intercepts of the fitted lines. This technique is known as a broken stick model, with the 'break' occurring at the LT. Beaver et al. [3] propose log transformations of both the workload and observed blood lactate to obtain a better estimate of the LT.

Both of these methods have come in for criticism as the LT is not a known physiological entity and its existence is merely an assumption which suits these markers. The broken stick model assumption that lactate is linear post LT also leads to criticism, as does the fact that using linear regression is sensitive to outliers when the sample size is low. The log transformation of both explanatory and response assumes the increase in lactate post LT is exponential.

Newell et al. [39] [40] suggest the workload corresponding to the maximum second derivative of the lactate curve (D2LMax) as another such marker. There is an intuitive reason for measuring the point at which the maximum second derivative, or maximum change in slope, occurs. From Figure 1.5 one could imagine that a broken stick model, as described above, would be a good fit to the data. Attempting to estimate the point at which this 'break' occurs is equivalent to finding the point at which the maximum of the second derivative occurs. This marker has been shown (Newell et al. [39]) to be both reliable and replicable so that endurance can be compared within and between athletes.

The main question of interest here is to obtain a second derivative estimate which leads to an accurate estimate of the speed corresponding to its maximum.

1.4 Astronomical Data

Figures 1.6 and 1.7 display ρ_{gas} (gas density profile) and temperature emissions for a sample galaxy cluster (A1995) obtained by the X-ray measuring satellites Chandra and XMM Newton. The aim of this study is to find gravitational mass density distributions $\rho_{tot}(x_1, x_2, x_3)$ of clusters of galaxies. The equation of hydrostatic equilibrium (1.1) allows for the estimation of ρ_{tot} using differentiation together with the T and ρ_{gas} measurements

$$\frac{d}{dx}\frac{\frac{d}{dx}(\rho_{gas}kT\mu m_p)}{\rho_{gas}} = -4\pi G\rho_{tot} \tag{1.1}$$

where x (in arcmin) is a measure of distance, G is the universal gravitational constant, k is Boltzmann's constant, μ is the mean molecular weight in any cluster and m_p is the mass of the proton.

The primary analysis requires differentiation of $\rho_{gas}kT\mu m_p$ and then further differentiation of $\frac{\frac{d}{dx}(\rho_{gas}kT\mu m_p)}{\rho_{gas}}$. From this, estimates of ρ_{tot} can be ascertained. These estimates can be hugely important to researchers in astronomy.

The ρ_{gas} variable is not measured directly but is found through an astronomical model

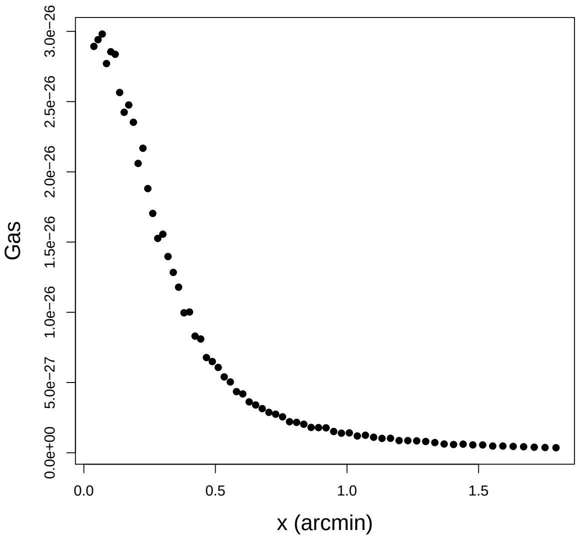


Figure 1.6: Gas readings at 64 distances in arcmin (an astronomical distance measure).

developed by collaborators on this project. However, there is measurement error in the temperature data. This uncertainty has contributions from instrumental errors as well as model sensitive factors. As such, lower and upper confidence estimates were provided along with an estimate of temperature at each of 8 arcmin. There are collaborators on this project who have doubts as to the accuracy of the set temperature measurements at the highest value of arcmin in Figure 1.7. From the extent of the bounds around this estimate the value is clearly highly unstable.

A further issue with these data is that the temperature variable contains far fewer observations (8) than ρ_{gas} (64). Hence, it is first necessary to predict values of temperature at the values of arcmin where ρ_{gas} has been observed. The ρ_{gas} measurements are collected at arcmin values far higher than the temperature data. Since extrapolation of the temperature data would be inaccurate due to the nonlinear nature of the data, it is predicted only over the range of arcmin where ρ_{gas} has been observed. This leaves 51 prediction points for the temperature variable.

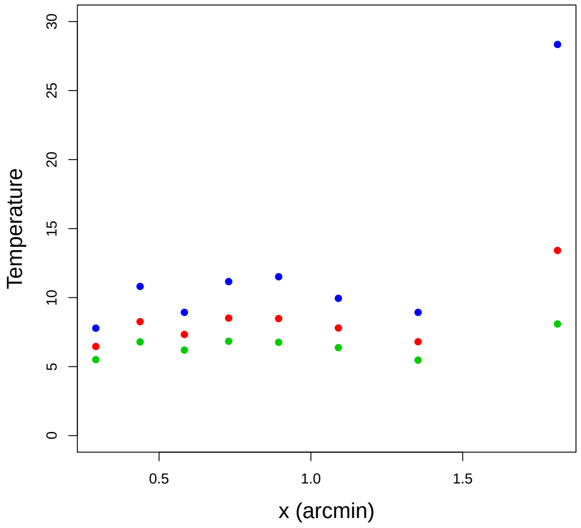


Figure 1.7: Temperature (low, green dots; medium, red dots; high, blue dots) readings at 8 distances in arcmin.

1.5 Simulated Functions

Known functions will be used to simulate data in order to test the performance for the derivative estimation methods which shall be introduced over the course of this thesis. These functions were chosen from the smoothing literature as they offer a range of trigonometric, polynomial and exponential components which lead to varied nonlinear characteristics and subsequent difficulties in derivative estimation. The functions chosen for simulation are displayed in Table 1.1 below, with each function and corresponding first and second derivatives plotted in Figures 1.8 to 1.13.

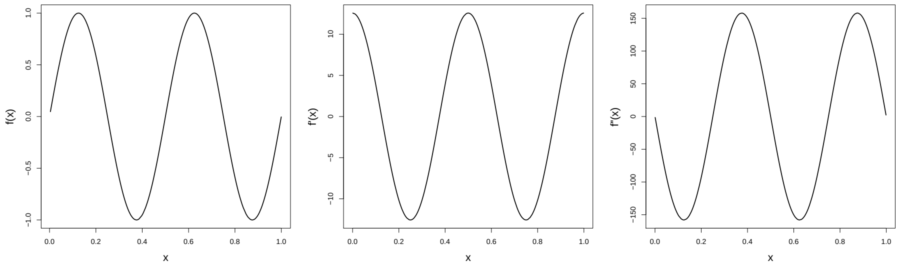


Figure 1.8: Plot of f_1, f_1' and f_1''

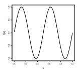
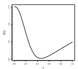
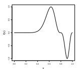
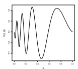
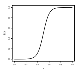
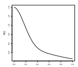
	$f_1 f_1$	$\sin(4\pi x)$
 f_2	f_2	$x + 2e^{-16x^2}$
 f_3	f_3	$(\sin(2\pi x^3))^3$
 f_4	f_4	$\sqrt{x(1-x)}\sin(2\pi(1.25)/(x+0.25))$
 f_5	f_5	$\frac{1}{1+e^{-20(x-0.5)}}$
 f_6	f_6	$-\cos(x-\pi/2) + 2e^{-16x^2}$

Table 1.1: Six functions under examination throughout this thesis.

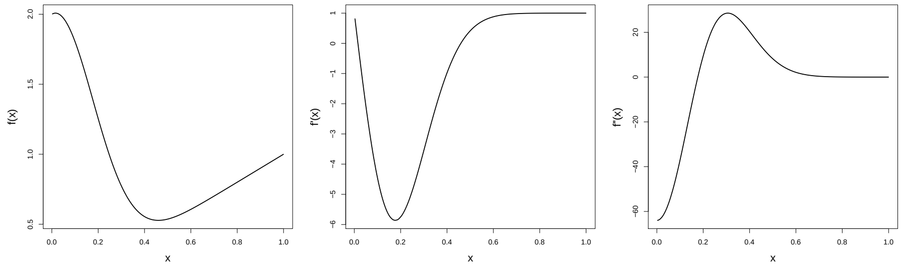


Figure 1.9: Plot of f_2, f_2' and f_2''

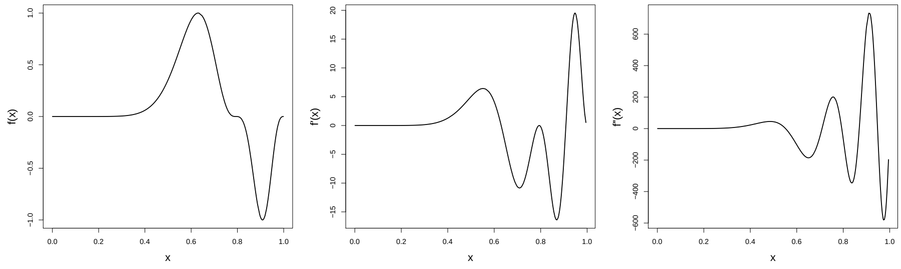


Figure 1.10: Plot of f_3, f_3' and f_3''

A simple sinusoidal function f_1 was taken from Härdle & Bowman [19]. Both the 'bump' and 'logit' functions f_2 and f_5 were taken from Ruppert [46] and f_6 is just a variant on the 'bump' function, with a cosine term added. The functions f_1, f_2, f_5 and f_6 were chosen to represent smoothing scenarios similar to that involved in the motivating illustrations. The

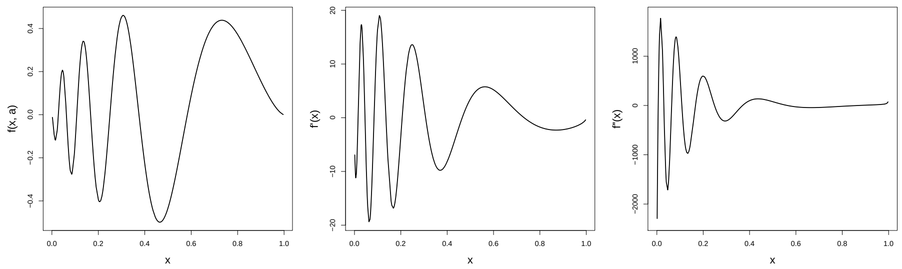


Figure 1.11: Plot of f_4, f_4' and f_4''

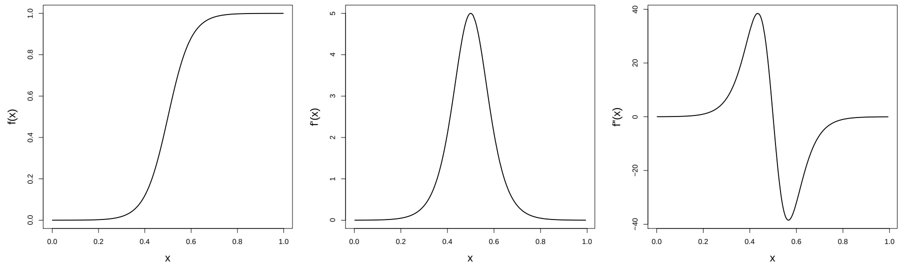


Figure 1.12: Plot of f_5, f_5' and f_5''

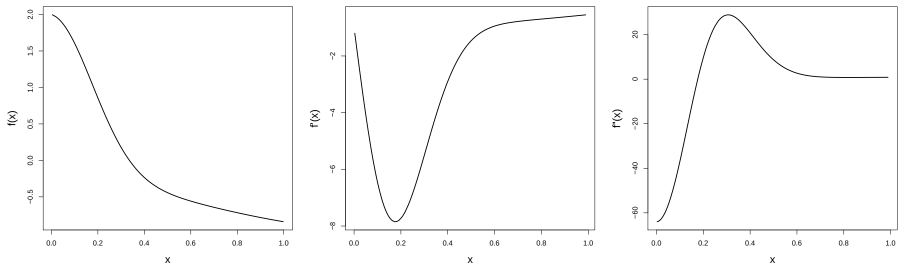


Figure 1.13: Plot of f_6, f_6' and f_6''

functions f_2 and f_6 have a single well defined maximum second derivative which is similar to a typical Blood Lactate curve. These will be used later to discover how the proposed methods are likely to perform in that example. The more complex sinusoidal f_3 is from Härdle [18] and the well known Doppler function f_4 was obtained from the paper by Donoho & Johnstone [10]. Both f_3 and f_4 were chosen since they present rather difficult smoothing (and therefore derivative estimation) problems in that a change in 'wiggliness' is apparent.

1.6 Chapter Summary

It is clear that these illustrations offer interesting and distinct problems that involve obtaining accurate derivative estimates. First derivative estimates are needed for a continuous (Winter Nutrients) and count (Grey Plover) response. Second derivative estimates are required for the Blood Lactate example where the sample size is small. In the Astronomical data two first derivative estimates are required once variables with different sample sizes have been combined. The six functions which shall be used in simulation studies have been introduced with reasoning provided behind each choice.

Discussion of what exactly one means by ‘rate of change’, along with some rather basic methods for estimating this quantity are introduced in the following Chapter. At several points throughout this thesis, the datasets profiled here will be revisited using the methods as they are introduced.

Chapter 2

Derivative Estimation Using Simple Techniques

In this Chapter some basic methods to estimate derivatives will be introduced, beginning with a simple data driven approach, then moving to more sophisticated modelling procedures. The nonlinear relationships between variables (e.g. Figure 1.1, 1.2, 1.4, 1.5, 1.6 and 1.7) in the motivating datasets from Chapter 1 suggest that modelling which can handle nonlinear relationships is required. Initially, polynomial regression is employed to achieve better performance in derivative estimation, although flaws with this modelling approach are found and discussed motivating the need for more complex procedures.

2.1 Derivative Estimation by Sequential Differences

When estimating the rate of change, or derivative, how the response variable y changes with respect to an explanatory variable x from observation to observation is of interest, i.e.

$$\frac{\Delta y_i}{\Delta x_i} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}. \tag{2.1}$$

When calculated for all $i = 1, \ldots, n-1$ a vector of sequential differences or 'slopes' y' is obtained. This is a first order difference estimate of the rate of change of the response. A second order difference estimate of the second derivative is found by calculating

$$\frac{\Delta^2 y_i}{\Delta^2 x_i} = \frac{y_{i+2} - 2y_{i+1} + y_i}{x_{i+2} - 2x_{i+1} + x_i} \tag{2.2}$$

for all $i = 1, \ldots, n-2$.

2.1.1 Example of a Sequential Difference Approach

The Grey Plover data, consisting of $n = 31$ counts taken annually from 1974 to 2004, are displayed in the left panel of Figure 2.1. The primary aim of the study is to investigate whether a significant decrease in count is evident. One approach to this question is to find an estimate for the rate of change and to compare this estimate with the line at $y' = 0$. A negative rate of change suggests a decrease in bird numbers (the notion of a 'significant' decrease will be discussed at length in later Chapters).

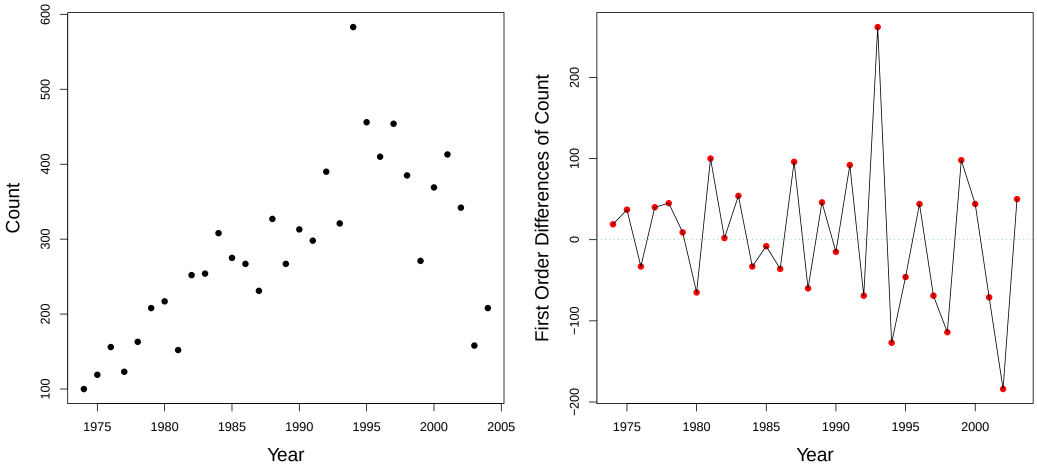


Figure 2.1: Grey Plover data 1974 to 2004 (left) with First Order Differences and interpolating fit (right).

Since the measurements are taken annually they are equally spaced such that, taking $x = Year$, $\Delta x_i = 1$ for all $i = 1, \ldots, n-1$. Using (2.1), the first order differences vector of length $n-1$ (for each order difference taken, an observation is lost) is displayed in the right panel of Figure 2.1. Note that 14 of the 30 first order differences were negative, i.e. a decrease in count was evident in 14 of the years. The first order differences have mean 3.6 and standard deviation 85.8. These summary statistics give an overall view of the likely change in levels of Grey Plover from 1974 to 2004. The primary goal, however, is not just to find whether a change occurs but, if a change is found, to identify the year that this transpires. This cannot be estimated using global measures or tests of trend. In order to investigate the performance of this estimate it is necessary to quantify the error of these fits.

2.1.2 Quantifying Error

In the scenario where $y = f(x) + \epsilon$ it is required to estimate the underlying function f which best describes the relationship between x and y. The mean squared error (MSE) is the most popular method for measuring goodness of fit and is defined

$$MSE[\hat{f}] = \mathbb{E}[\hat{f} - f]^2, \tag{2.3}$$

where $\hat{f}$ is the estimator of f. With a little manipulation

$$\begin{aligned}
\mathbb{E}[\hat{f} - f]^2 &= \mathbb{E}[\hat{f}^2 - 2f\hat{f} + f^2] \\
&= \mathbb{E}[\hat{f}^2] - 2\mathbb{E}[f\hat{f}] + \mathbb{E}[f^2] \\
&= \mathbb{E}[\hat{f}^2] - 2f\mathbb{E}[\hat{f}] + f^2 \\
&= \mathbb{E}[\hat{f}^2] - \mathbb{E}[\hat{f}]^2 + \mathbb{E}[\hat{f}]^2 - 2f\mathbb{E}[\hat{f}] + f^2 \\
&= Var[\hat{f}] + (\mathbb{E}[\hat{f}] - f)^2
\end{aligned}$$

i.e.

$$MSE\hat{f}(x) = Var\hat{f} + Bias^2\hat{f}. \tag{2.4}$$

The MSE quantifies both the Variance and Bias of a fit; interpolation has high variance but zero bias while a linear fit will have low variance but high bias. This so called ‘bias-variance trade-off’ needs to be managed in order to find an optimum fit. A more detailed discussion of the bias-variance trade-off will be given later.

When quantifying the error in a real life application the function f is unknown and therefore MSE (2.3) is not practical. A natural surrogate is to consider the mean squared residuals $\hat{\epsilon}$

$$\hat{\epsilon} = \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{n}. \tag{2.5}$$

To measure performance in derivative estimation, it is necessary to simulate from a known function f such that derivatives $f^{(l)}$ $(l = 0, 1, 2, \dots)$ are known. When quantifying error the root mean squared error can be employed for derivative estimation. This measures both the variance and bias of the fitted derivative.

$$\text{RMSED}(l) = \frac{\sqrt{\sum_{i=1}^{n}(f^{(l)}(x_i) - \hat{f}^{(l)}(x_i))^2}}{n}. \tag{2.6}$$

For $l = 0$, (2.6) is known as RMSE and measures the error in the fit to the observed data $y_i = f^{(0)}(x_i)$.

2.1.3 Simulations for Sequential Difference Estimates

A small simulation study was performed to test the performance of first order difference estimates for the rate of change on the function f_1 from Section 1.5.

One thousand samples of $n = 50$ response values y were simulated from the function $f_1 = \sin 4\pi x$ (introduced in Section 1.5) with x uniform on $[-1, 1]$ and some error ϵ added such that $y = f_1(x) + \epsilon$, where $\epsilon \sim N(0, 1)$. A sample size of fifty is taken as it is close to the typical value of the sample sizes of the motivating examples (namely $n = 11, 31, 64$ and 131). Since it is known that $f_1' = 4\pi \cos 4\pi x$ the accuracy of the first order difference method can be measured using the root mean squared error of the l^{th} derivative (RMSED(l), (2.6)) of f_1.

Figure 2.2 displays the error observed using a first order difference estimate of the rate of change of the function $f_1 = \sin 4\pi x$. Without another method for comparison it is difficult to make judgements about the accuracy of this estimator. However, the range of f_1' is $[-10, 10]$, such that an average error of 2359 with standard deviation 20,261 is not very good performance!

Figure 2.3 displays one randomly selected replicate of these simulations with first order differences plotted along with the actual first derivative of f_1. There are outliers beyond -1500 which would have a large influence on the RMSED(1). Outliers such as these are most likely due to the sensitivity of differencing to unusual observations. This estimation procedure is not

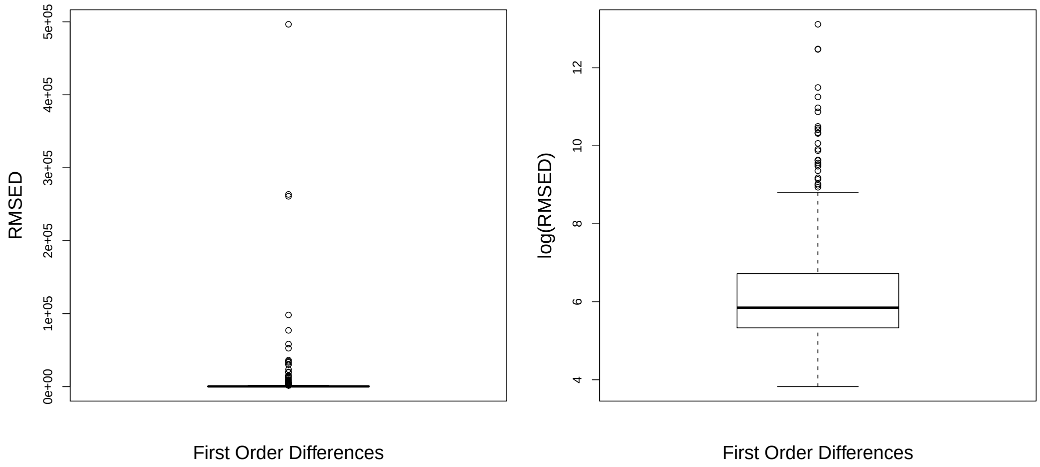


Figure 2.2: RMSED(1) found using First Order Differences (left panel) and log(RMSED(1)) (right panel).

a precise one, its simplicity hinders any accurate decision making and more complex methods for estimating rate of change are necessary. The estimate of the rate of change in the Grey Plover example using sequential differences thus comes into question.

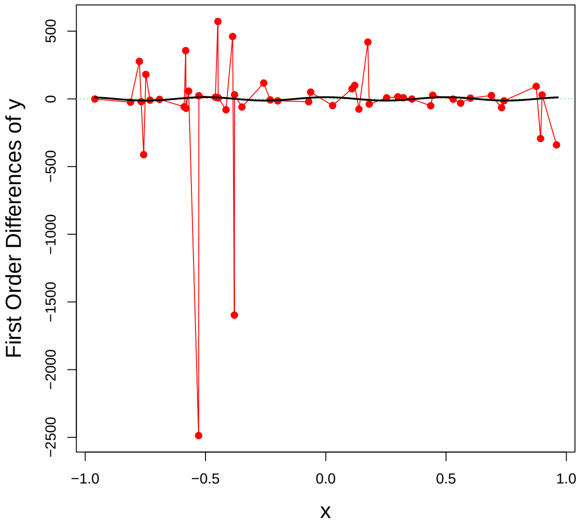


Figure 2.3: First Order Differences (red) compared to the actual first derivative of f_1 (bold curve).

2.1.4 Summary of Sequential Differences

Simply finding the first order difference between successive observations does not give an accurate guide to the likely rate of change of a response variable relative to an explanatory. The method is too volatile and the variance of the observed data is magnified when estimating the rate of change using a first order difference method. Whether there are methods involving differences of observations which can offer suitable derivative estimates has yet to be seen. Where little variance exists in a nonlinear relationship between variables it is possible that first order differences may offer decent estimates of the rate of change. There have been few attempts in the recent literature to use first order differences when estimating the rate of change. Tukey [59] discusses data driven smoothing approaches using sequential differences. Sangalli et al. [50] model the rate of change of blood flow in the coronary artery but abandon the use of simple differencing at an early stage in their paper in favour of a more practical approach. There now seems to be an argument to model the relationship between the variables and then attempt to discern from this an estimate for the rate of change.

2.2 Derivative Estimation Using Linear Models

Linear and polynomial regression methods are commonly used in statistical analyses. Here both modelling approaches are reviewed, applied to some of the motivating examples and tested for performance in derivative estimation through simulation.

2.2.1 Simple Linear Regression

The classical approach to modelling the relationship between a response variable y and a single explanatory variable x is written as

$$y = \beta_0 + \beta_1 x + \epsilon, \tag{2.7}$$

and is known as the simple linear regression model where errors ϵ are assumed independent and identically distributed (*iid*) Normal with zero mean and constant variance σ^2. The slope parameter β_1 is the average change in response per unit change in explanatory. Hypothesis tests allow conclusions to be drawn about the relationship between response and explanatory, but this model relies on the assumption that the response is linear in the parameters (β_0, β_1). If a nonlinear relationship exists one may transform y to force a linear relationship with x.

In terms of derivative estimation, the coefficient $\hat{\beta}_1$ is an estimate of the global slope parameter β_1. Once fitted (using the simple least squares methods), it provides a one number summary of the rate of change of the (linear) function underlying the observed data. In the datasets which have motivated this research, a single scalar estimate of the rate of change is unlikely to be adequate to answering the fundamental questions of interest.

2.2.2 Polynomial Regression

In order to model nonlinear relationships between the explanatory and response variable one can amend the simple regression model by transforming y or by introducing polynomial effects of x. A pth degree polynomial regression is of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \epsilon. \tag{2.8}$$

Thus first derivative estimates $\hat{y}'$ are found using

$$\hat{y}' = \hat{\beta}_1 + 2\hat{\beta}_2 x + \cdots + p\hat{\beta}_p x^{p-1} \tag{2.9}$$

where $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)$ is estimated using the least squares method.

2.2.3 Application of Derivative Estimation Using Polynomial Regression

Figure 2.4 exhibits polynomial regression fits of the Grey Plover data using quadratic, cubic and quintic polynomials in x (i.e. $Year$) along with a simple linear regression fit.

The linear regression fit is not suitable due to the nonlinear relationship which exists between $Year$ and the Grey Plover count. The polynomial regression method fits the data quite well as the degree of polynomial is increased. The quintic fit seems to give a good description of

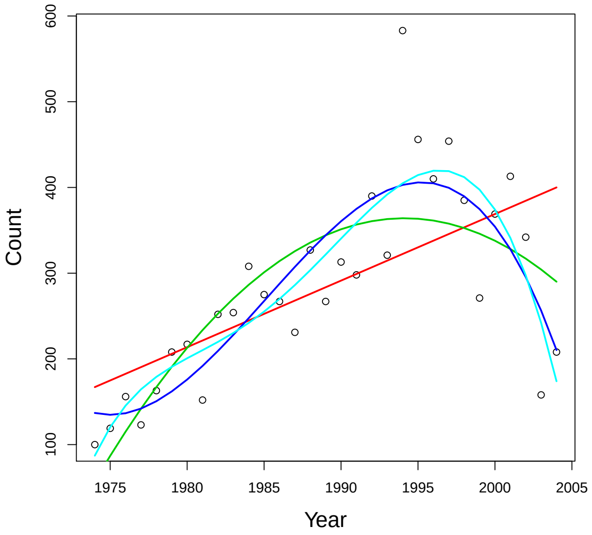


Figure 2.4: Regression Models for the Grey Plover data using linear (red), quadratic (green), cubic (blue) and quintic (cyan) polynomials.

the behaviour of the counts moving through the years. The mean squared residuals $\hat{\epsilon}$ (Section 2.1.2) of these fits is displayed in Table 2.1. It is evident that as the degree of the polynomial increases, the error decreases. However, at $p = 30$ (i.e. $n - 1$) no error would be present but an interpolating fit is not useful to the analysis because it has very high variance. This is the bias-variance trade-off argument once again.

p	ϵ_i
1	88.1
2	69.5
3	59.1
5	55.6

Table 2.1: Mean squared ϵ using degree p Polynomial Regression to fit the Grey Plover data.

Estimates of the rate of change of count with respect to year are shown in Figure 2.5. These estimates are a massive improvement on using first order differences.

In order to obtain a sensible estimate for rate of change it is necessary to fit a polynomial of degree $l + 2$, where l is the order of derivative needed. If a degree of $l + 1$ is selected then the lth derivative estimate will be linear which may not be a sensible estimate in nonlinear scenarios. Discounting the first year, both the cubic and quintic estimates show an increasing

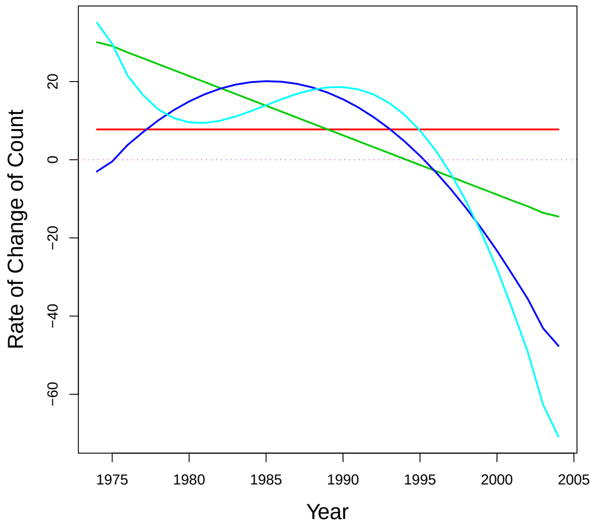


Figure 2.5: Estimates of the rate of change of count obtained by linear (red), quadratic (green), cubic (blue) and quintic (cyan) regression with reference line at 0 (dotted).

count until some time between 1995 and 2000.

2.2.4 Simulations Varying the Degree of a Polynomial Regression

In order to test the performance of polynomial regression derivative estimates, a simulation study was performed. To reiterate, since actual derivatives of data are unknown, derivative estimators can only be tested for performance using known functions.

Similarly to the simulations performed in Section 2.1.3, samples of size $n = 50$ were simulated 1000 times from the function $f_1 = \sin 4\pi x$ from Section 1.5. An error vector $\epsilon \sim N(0, \sigma^2)$ was added to f_1 with x uniform on $[-1, 1]$ and $\sigma = \frac{1}{3} range(f_1)$. The variance is set to a fraction of the range of the data to allow for realistic noise to be included in the simulated data. For example, if $\sigma = 1$ were chosen this would be negligible if a function's range was (-100,100). Polynomial regression with $p = 3, 5, 7, 9$ was used to estimate f_1 and f_1' with performance once again measured using the RMSE and RMSED(1) respectively.

Boxplots of the RMSE and RMSED(1) are provided in Figure 2.6. It is evident that as the degree of the polynomial is increased the RMSE decreases, i.e. the estimates of f_1 improve. Interestingly this monotonic improvement is not carried over to estimates of f_1'. One would logically believe that a better estimate of f_1 would garner a better estimate of f_1', but this is not the case. Indeed, results from Table 2.2 reveal that using a polynomial of degree $p = 7$ results in the best performance for estimating f_1'.

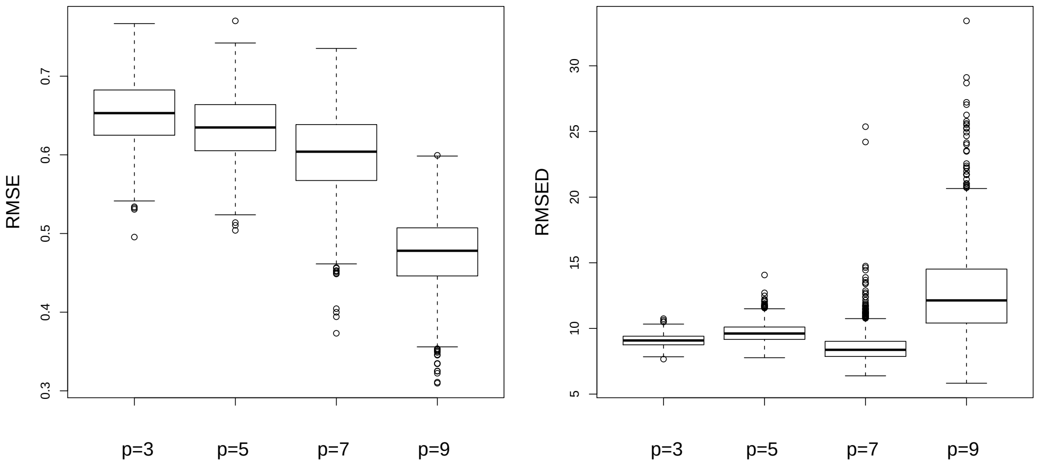


Figure 2.6: RMSE and RMSED(1) for estimates of f_1 (left) and f_1' (right) using polynomial regression.

p	RMSE	RMSED(1)
3	0.65(0.04)	9.1(0.47)
5	0.63(0.04)	9.7(0.77)
7	0.60(0.05)	8.6(1.34)
9	0.47(0.05)	12.8(3.51)

Table 2.2: Performance using degree p Polynomial Regression for the simulated data.

Figures 2.7 and 2.8 provide plots of one replicate simulation chosen at random along with estimates of f_1 and f_1' which call into question the accuracy of estimates for the Grey Plover data given in Figure 2.5.

The large variability in the higher degree polynomials causes overfitting estimates at different periods along the range of x. The results from Table 2.2 show that the $p = 7$ polynomial is the better estimator for f_1' and this seems to be resultant from the high variability of the nonic polynomial.

Figure 2.9 displays interpolation of the Grey Plover data. Interpolation is not the optimum method to explain a (nonlinear) relationship because it does not account for sampling error and an interpolated fit has very high variance.

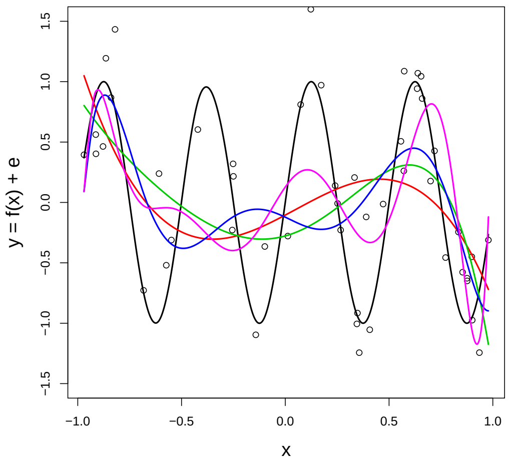


Figure 2.7: Estimates of f_1 (black curve) using cubic (red), quintic (green), septic (blue) and nonic (purple) polynomials.

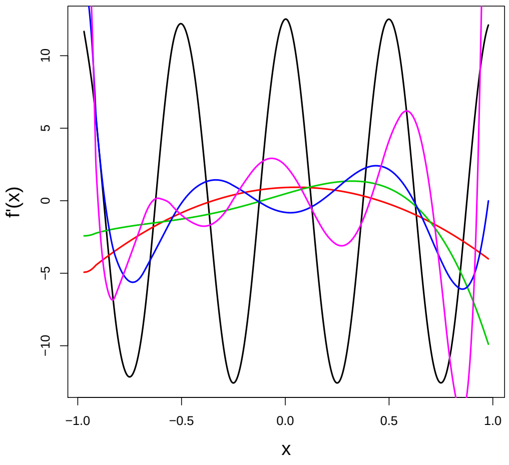


Figure 2.8: Estimates of f_1' (black curve) using cubic (red), quintic (green), septic (blue) and nonic (purple) polynomials.

Using higher degree polynomials can lower RMSE for estimation of f, but this does not always translate to improved RMSED(1), i.e. estimates of f'. The large variance contained in estimates which use a high degree global polynomial causes a high variability in estimates of f'.

The use of global polynomials offers no flexibility to local change. Since the bases on which these fits are made (i.e. powers of x) are global, a local change, say in a tail, will have a global

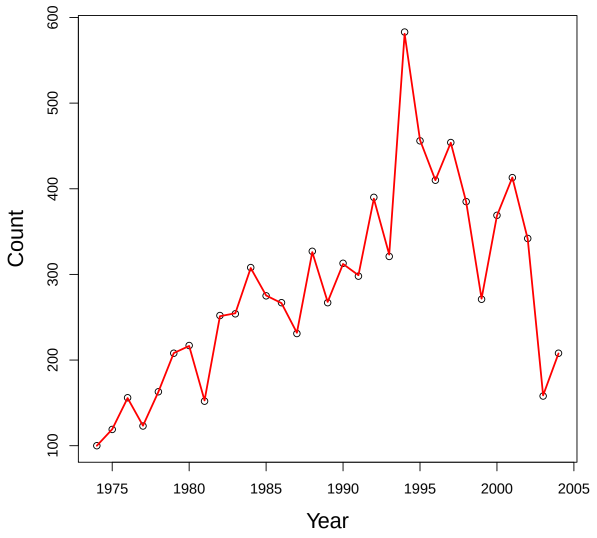


Figure 2.9: Interpolation of the Grey Plover.

effect.

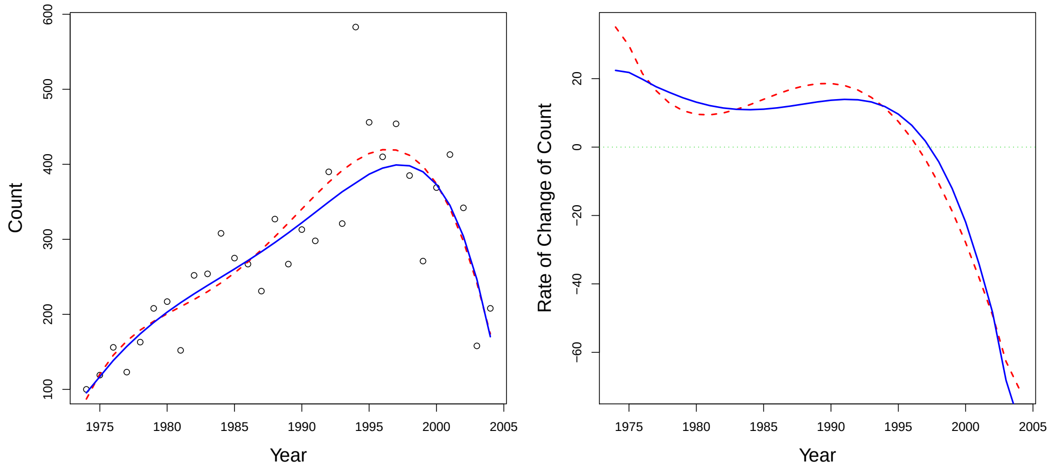


Figure 2.10: Estimates using $p = 5$ of the Grey Plover data (left) and the rate of change of Grey Plover (right) with (dashed red curve) and without (solid blue curve) the 1994 observation.

Figure 2.10 gives estimates of the Grey Plover data using a $p = 5$ polynomial where the 1994 observation of 583 birds has been removed. There is a global effect of removal of the observation caused by the rigidity of the polynomial basis. The removed observation at $Year = 1994$ also has an effect on $\hat{f}$ circa $Year = 1982$, which is magnified when using this fit to estimate f'. The higher the degree of the polynomial, the better the fit but the more sensitive to local change estimates become. Higher degree polynomials do not necessarily lead to improved estimates of rate of change. It seems that global polynomial fitting is not conducive for derivative estimation in noisy data.

2.3 Chapter Summary

Using both simulated and real data, the sequential differencing approach to estimating rate of change has been shown to possess too many problems to be a viable approach to the derivative estimation problem. Simple linear regression offers a global measure of trend which is not suitable for the datasets with which this thesis is concerned and for derivative estimation in all but a limited case. At first glance, polynomial regression offers a suitable fit and derivative estimate for the Grey Plover data. Strangely, monotonic increases in performance in derivative estimation with higher degree polynomials were not observed in the context studied here. The increased sensitivity of higher degree global polynomials leads to poor performance in simulated derivative estimation scenarios due to local change having global influence. It now seems that a more robust local polynomial structure needs to be in place to allow for better estimates of first and second derivatives.

Chapter 3

Derivative Estimation Using Spline Smoothing

Global polynomial fitting has failed to offer the flexibility required when estimating derivatives of noisy data. In this Chapter local methods for fitting curves to noisy data are motivated with several nonparametric regression or 'smoothing' techniques introduced and compared for performance in estimating first and second derivatives of simulated data The focus will be on spline smoothing methods with a short discussion on kernel versus spline methods. Smoothing splines, P-Splines, mixed model spline smoothing and some adaptive smoothing techniques will all be introduced with a guide to obtaining derivative estimates from each given. The datasets introduced in Chapter 1 are used to examine these methods graphically and simulations are used to measure performance empirically.

3.1 Derivative Estimation by Local Methods

Consider the situation

$$y = f(x) + \epsilon \tag{3.1}$$

where $\epsilon \sim N(0, \sigma^2)$ for some constant variance σ. The goal is to estimate f such that the most accurate estimates of f' and f'' can be found. Using global polynomials to model nonlinear relationships to obtain trustworthy estimates of rate of change is not plausible since a high degree polynomial basis is too sensitive when estimating derivatives (as seen in Section 2.2.4). Fortunately there exist a wide range of local modelling techniques which offer the flexibility required to model these relationships.

3.1.1 Knots, Splines and Bases

A simple fit of the Grey Plover data can be seen in the left panel of Figure 3.1. This piecewise linear fit, known as a broken stick model, serves as a starting point to the field of nonparametric regression and smoothing.

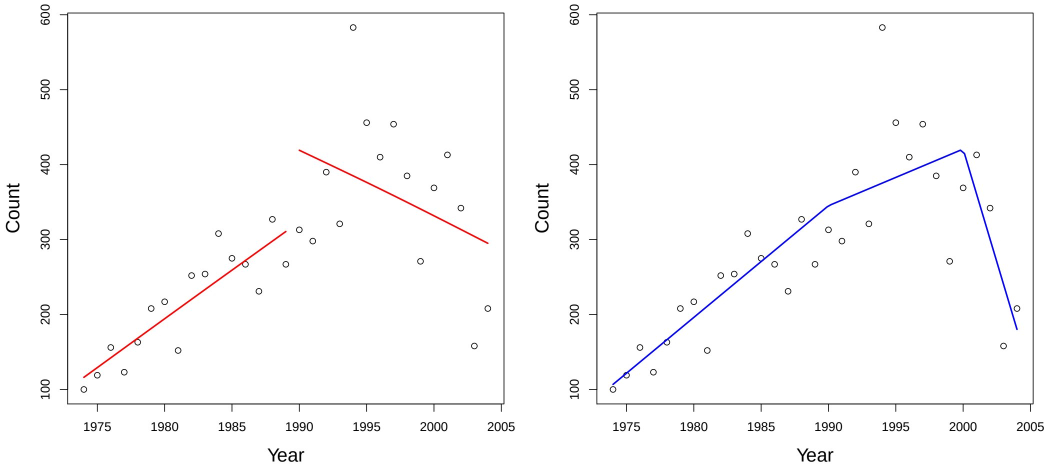


Figure 3.1: Piecewise Linear fit to the Grey Plover data using one (left) and two knots (right).

This fit has been achieved by choosing a breakpoint, or knot, in order to break up the range of the data. The number and position of these knots is a subjective choice, here one knot at $Year = 1990$ was selected. There are two paths to improve this fit. Firstly, joining the stick at the knot to make it continuous results in a more realistic description of the relationship and choosing further knots permits additional flexibility. The broken stick model, when applied to the Grey Plover data, can be written as

$$Count = \beta_0 + \beta_1 Year + \beta_2 (Year - 1990)_+ + \epsilon \tag{3.2}$$

where $a_+ = a$ if $a > 0$ and 0 otherwise. A continuous fit employing a further knot at $Year =$ 2000 is displayed in the right panel of Figure 3.1. The fitted values are obtained through

$$\hat{Count} = \hat{\beta}_0 + \hat{\beta}_1 Year + \hat{\beta}_2 (Year - 1990)_+ + \hat{\beta}_3 (Year - 2000)_+.$$

Taking a large number K of knots at locations $\kappa_1, \ldots, \kappa_K$ leads to what is known as the truncated linear spline basis

$$[1 \; x \; (x - \kappa_1)_+ \; \ldots \; (x - \kappa_K)_+]. \tag{3.3}$$

A spline of degree q is a qth degree polynomial between each pair of knots (κ_{k-1}, κ_k) which has q continuous derivatives in these intervals. In statistical analyses, spline functions offer the flexibility required to accurately describe nonlinear patterns which may exist in the relationship between variables.

Historically, splines were used as a tool by draftsmen in boat building to accurately draw curves by hand. Long thin strips of wood or metal, known as splines, were held in place by weights (or knots) and would, due to their elasticity, sag between these knots into the smoothest possible shape. A recreation of a classic spline tool is displayed in Figure 3.2.

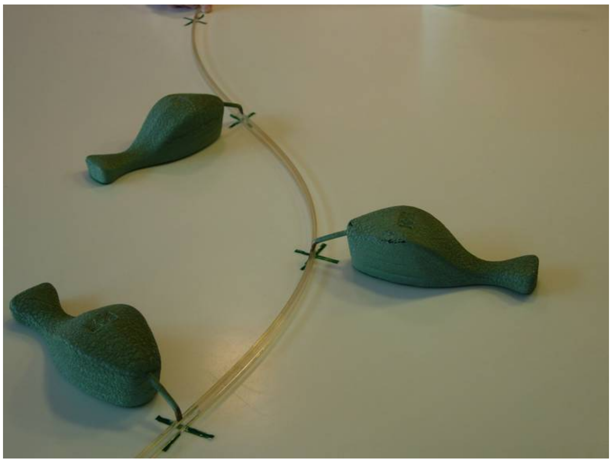


Figure 3.2: A recreation of a draftsman's spline

Mathematicians started studying the spline shape (Schoenberg [51]), and derived the piecewise polynomial formula known as the spline curve or spline function. A general form for the model in (3.2) is

$$y = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k (x - \kappa_k)_+ + \epsilon \tag{3.4}$$

for some weighting vector u_k. The number (K) and position (κ_k) of knots can be vital to the shape of the estimate obtained. To arrive at a satisfactory fit, one could cycle through different combinations of number and position of knots although this would be an arduous task if done subjectively. Automatic methods for estimating f by changing the number and position of knots are known as free-knot spline or regression spline methods. Alternatively, rather than changing a feature of the knots, one may attempt to control the influence of the knots through the coefficients u_k. This can be done using penalisation of a feature of the u_k (smoothing splines, P-Splines) or through mixed model smoothing. Moreover, the choice of basis is arbitrary, such that any general basis of dimension m

$$\phi_1(x), \ldots, \phi_m(x) \tag{3.5}$$

can replace (3.3). There are many types of bases which are used in practice, although here the emphasis is on the truncated power series

$$[1 \; x \; x^2 \; \ldots \; x^p \; (x - \kappa_1)_+^p \; \ldots \; (x - \kappa_K)_+^p], \tag{3.6}$$

cubic spline and B-Spline bases. For a comprehensive summary of basis types see Chapter 3 of Ramsay & Silverman [44].

The spline function f can be written as a linear combination of coefficients and bases

$$f = \sum_{j=1}^{m} \alpha_j \phi_j \tag{3.7}$$

where $\alpha = (\alpha_1, \ldots, \alpha_m)$ are the coefficients of the basis defining the behaviour of the estimate of f. Using X to denote the basis design matrix, f may be written

$$f = X\alpha.$$

Then from (3.1)

$$\begin{aligned}\epsilon^2 &= \|y - f\|^2 \\ &= \|y - X\alpha\|^2\end{aligned}$$

is to be minimised. Taking the derivative with respect to α and setting equal to zero gives the normal equations

$$2XX^T\alpha - 2Xy = 0.$$

The fitted coefficients α are found by solving

$$\hat{\alpha} = (X^TX)^{-1}X^Ty \tag{3.8}$$

and fitted values $\hat{y}$ are subsequently obtained using

$$\begin{aligned}\hat{y} &= X(X^TX)^{-1}X^Ty \\ &= Hy.\end{aligned} \tag{3.9}$$

The matrix which transforms the observed y into fitted values $\hat{y}$ is known as the hat matrix, denoted H, which has the property that the degrees of freedom (also known as the effective dimension) of the fit is the trace of H

$$df_{fit} = tr(H) \tag{3.10}$$

since $tr(H)$ gives the number of parameters of the fit in (3.9).

3.1.2 Penalties and Smoothing Parameters

There are several types of penalties which may be enforced in the fitting process in order to achieve an accurate estimate of f. Two of the most popular techniques are described later in this Chapter. A general penalised spline smoothing model may be written

$$\sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda P(f) \tag{3.11}$$

where $P(f)$ is a function penalising some feature of f and λ, known as the smoothing parameter, is a non negative constant which controls the smoothness of $\hat{f}$.

The smoothing parameter λ governs the balance between the smoothness and the accuracy of the fit, i.e. it controls the bias-variance trade-off. Setting λ too small, $\hat{f}$ tends towards the least squares fit which minimises $\sum_{i=1}^{n}[y_i - f(x_i)]^2$. This gives an low bias estimate with huge local variability, known as 'undersmoothing' (left panel of Figure 3.3).

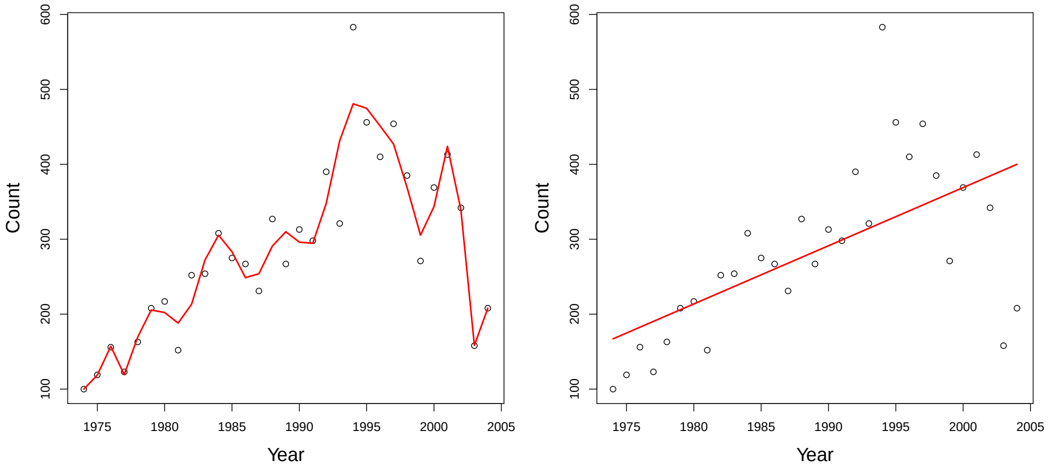


Figure 3.3: Undersmoothing (left) and oversmoothing (right) of the Grey Plover data.

If λ is very large the most smooth fit is achieved which would be unable to describe a typical noisy data situation. This choice causes large smoothing bias but small variability, referred to as 'oversmoothing' (right panel of Figure 3.3).

The choice of λ is therefore critical in spline smoothing and as such there exist several methods for determining the 'correct' choice. One could select the smoothing parameter subjectively by simply choosing λ based on observing curves and selecting the one which graphically 'looks right'. However, there is a need for an automatic selection of λ, perhaps for the inexperienced user, or just as a starting point for a subjective choice. To this end there exist several automatic methods for selecting λ.

3.1.2.1 Cross-Validation

Cross-validation (CV) [61] [62] is the most popular automatic smoothing parameter selection tool. It is generally built into packages for computational spline smoothing methods. The data are first split into two disjoint sets. The function f is estimated using information from one of the sets and is then used to predict the outcomes from the second. The predicted and observed values of the second set can then be compared. The process can be altered using different partitions of the data, however, leave-one-out cross-validation is generally used. In this case one data point is removed and predicted from the rest of the observations. This is repeated for each x_i, $i = 1, \ldots, n$ where the estimator with (x_i, y_i) left out is denoted by $\hat{f}_{-i}$. The CV choice of λ is given by minimising

$$CV_\lambda = \sum_{i=1}^{n} [y_i - \hat{f}_{-i}(x_i; \lambda)]^2. \tag{3.12}$$

This method can quickly become computationally hard to manage for large n since $\hat{f}_{-i}$ must be computed for $i = 1, \ldots, n$ which leads to a problem of order n^2. The algorithm can be simplified such that the cross-validation score can be computed more easily. Let H_λ be the hat matrix such that

$$H_\lambda y = \begin{bmatrix} \hat{f}(x_1; \lambda) \\ \hat{f}(x_2; \lambda) \\ \vdots \\ \hat{f}(x_n; \lambda) \end{bmatrix}.$$

It has been shown (Hutchinson & de Hoog [24]) that one may now write the cross-validation score (3.12) as

$$CV_\lambda = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_{ii,\lambda}} \right)^2 \tag{3.13}$$

where $h_{ii,\lambda}$ are the diagonal elements of H_λ.

3.1.2.2 Generalised Cross Validation

The computation discussed above can be reduced yet further by replacing the diagonal hat matrix entries by their average i.e.

$$\frac{1}{n}\sum_{i=1}^{n} h_{ii,\lambda} = \frac{1}{n}tr(H_\lambda). \tag{3.14}$$

This is known as generalised cross-validation (GCV), and was developed by Craven & Wahba [7]. The criterion to find the optimum smoothing parameter is to minimise λ over

$$GCV_\lambda = \sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - n^{-1}tr(H_\lambda)}\right)^2. \tag{3.15}$$

It is usually the case that the value of λ given by GCV does not differ by much from the CV value, but it is computationally much quicker to use GCV and therefore advisable for large n. Kohn et al. [26] compare CV and GCV when estimating both f and f' using a penalised spline method. They find that for unequally spaced data GCV performs better, but otherwise these criteria are comparable.

3.1.2.3 Akaike's Information Criterion and the Bayesian Information Criterion

Other selection criteria include Akaike's information criterion or AIC (Akaike [1]) and the Bayesian information criterion or BIC (Schwarz [53]). Each attempts to balance the least squares estimate with sensible smoothness. For AIC, λ is found by minimising

$$AIC_\lambda = \log\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \frac{2df_{fit,\lambda}}{n} \tag{3.16}$$

where $df_{fit,\lambda} = tr(H_\lambda)$. For the BIC, λ minimises

$$BIC_\lambda = \log\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \frac{\log(n)df_{fit,\lambda}}{n}. \tag{3.17}$$

3.2 Derivative Estimation Using Splines

Surprisingly, there are few papers in the literature which use derivative estimation based on splines as part of an analysis. Especially rare are those which notice, and/or deal with, problems when estimating some lth derivative $f^{(l)}$ ($l = 0, 1, 2, \dots$). Ramsay [42] gives his thoughts on boundary problems in derivative estimates:

"None of the current methods works [sic] well at the boundaries, in spite of claims to the contrary in the local polynomial smoothing literature. Methods that control bias pay a savage price in variance. Typically one sees derivatives go wild at the extremes, and the higher the derivative, the wilder the behavior."

Ramsay & Li [43] describe how to obtain the rate of change of an underlying function f which has been estimated by smoothing splines. Heckman & Ramsay [21] propose using L-Splines, a form of penalised spline, to obtain derivative estimates. Zhou & Wolfe [66] derive properties of regression spline derivative estimates but do not discuss problems which may be inherent with these. Newell & Einbeck [38] describe problems found using both spline and kernel derivative estimation techniques through a small simulation study. Sangalli et al. [50] use a free-knot spline model to estimate the curvature of three dimensional heart data and give a nice summary of derivative estimation from this type of model but do not notice any potential problems with their results.

There are many spline smoothing methods which can be constructed through the building blocks of knots, penalisation and bases discussed in Section 3.1. Here some of the more popular techniques which are commonly used to estimate f and its derivatives are profiled.

3.2.1 Smooothing Splines

Smoothing splines [45] are a popular penalisation based smoothing method. Knots are placed at each observation, i.e. $K = n$, and a natural cubic spline basis is used, i.e. outside the range of x, the basis is constrained to be linear and inside the range it consists of cubic splines. Using the definition of a spline of degree q from Section 3.1, a cubic spline is a cubic polynomial on each interval $[\kappa_k, \kappa_k + 1]$ for $k = 1, \ldots, K - 1$ which has two continuous derivatives (i.e. continuous first and second derivatives) at each κ_k.

Using least squares in (3.8) would lead to interpolation of the data (since $K = n$), which is not a sensible estimate of f. To avoid interpolation, a roughness penalty is introduced on the estimate $\hat{f}$. In order to hinder wild fluctuations of $\hat{f}$, large second derivatives of $\hat{f}$ are penalised, i.e. find f which minimises the penalised sum of squares equation

$$PENSS = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_X f''(x)^2 dx. \qquad (3.18)$$

It can be shown (de Boor [9]) that the function f which minimises (3.18) is a cubic smoothing spline. Green & Silverman [17] give an excellent description of the construction of a smoothing spline fit $\hat{f}$. Fortunately, the software package R contains the function `smooth.spline` which allows the user to fit cubic smoothing splines to data and `smooth.Pspline`, in the **psplines** library, to fit smoothing splines of arbitrary order. In practice, derivative estimates are found by numerical methods and in R through the `predict` function. One must be careful to choose a spline basis of order $l + 2$, where l is the highest order of derivative to be estimated, to avoid linear and constant estimates in higher order derivatives of interest.

3.2.1.1 Application of Smoothing Splines to the Winter Nutrients Example

The Winter Nutrients dataset contains 131 salinity adjusted values of NTRZ and Phosphate measured in 1990/1991. Figure 3.4 exhibits smoothing spline fits to the NTRZ data.

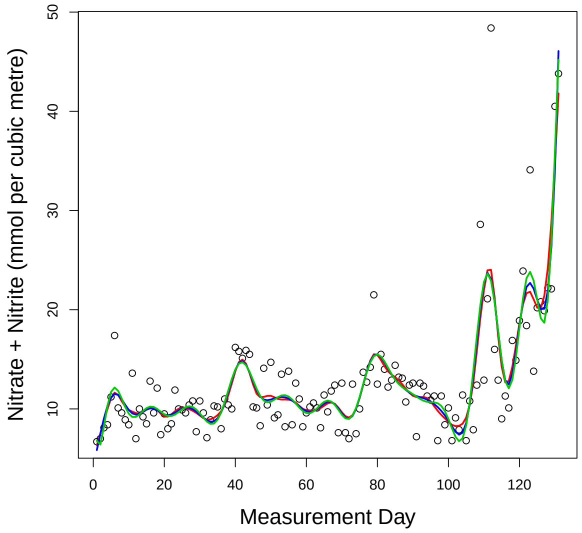


Figure 3.4: NTRZ fits using cubic (red), quintic (blue) and septic (green) smoothing splines

.

The similarity of these smooths shows that the degree of natural spline basis used is not an important choice for the smooth fit to the data. Each smooth does offer an accurate representation of the data, but this was also achieved through polynomial regression with a high enough degree in Chapter 2. Using local basis functions and coefficients gives a more robust model because local changes have only local effects and polynomials with a degree of just three are required to achieve the smooth fits in Figure 3.4. A global polynomial of degree greater than ten would be required to accurately describe the NTRZ data judging by the major oscillations of curves in Figure 3.4.

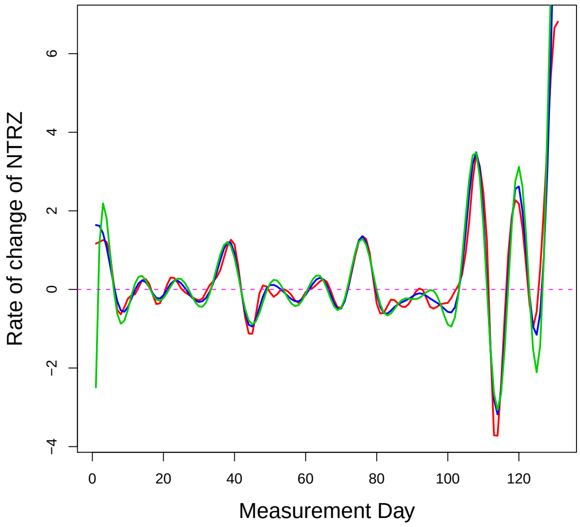


Figure 3.5: Estimates of the rate of change of NTRZ using cubic (red), quintic (blue) and septic (green) smoothing splines.

The main aim of the Winter Nutrients project is, however, not to model the relationship, but to identify periods of significant increase or decrease in contamination. It is therefore necessary to estimate the rate of change of contamination. First derivative estimates using smoothing splines for the NTRZ data from 1990 are provided in Figure 3.5. The difference of chosen basis degree is more evident in these derivative estimates. Small changes in estimates of the underlying function explaining the data are magnified when finding estimates of derivatives of this unknown function. The higher the degree, the greater the sensitivity. All three fits agree that there is a large amount of fluctuation in the rate of change of NTRZ with periods of increase, decrease and no change. Quite a large discrepancy between estimates is noticeable in the tails and in some of the major peaks and troughs towards the end of Winter.

3.2.1.2 Comparisons of Polynomial Regression and Smoothing Splines

To measure the performance in derivative estimation using smoothing splines a simulation study was carried out. Once again it is necessary to use simulated data from known functions in order to record performance when estimating f' and f''.

Datasets of size $n = 50$ were simulated 1000 times from the function $f_2 = x + 2e^{-16x^2}$ (from Section 1.5) with x uniform on $[-1, 1]$ and some Gaussian error $\epsilon \sim N(0, \sigma^2)$ added. The error was set to $\sigma = \frac{1}{3} range(f)$. This function is used as it resembles the Winter Nutrients data, specifically the Phosphate data from 1990. Recall the variance is set to a fraction of the range of the data to simulate realistic noise (Section 2.2.4). For each of the 1000 replicates the RMSE, RMSED(1) and RMSED(2) were recorded. The methods under investigation are quintic smoothing splines (since the objective is to obtain smooth second derivative estimates) as well as a septic polynomial fit for comparative purposes.

The boxplots in Figure 3.6 represent the summary statistics of the RMSE, RMSED(1) and RMSED(2). The septic polynomial is preferred in terms of RMSE although, as has been mentioned, the global basis on which it is formed is rigid and therefore unable to deal with local changes locally. The smoothing spline derivative estimates appear superior to the global polynomials due to the problems using higher degree global polynomials.

The outliers which can be seen in the RMSED(1) and RMSED(2) boxplots are of concern however. For estimates of f_2' there are 11 out of the 1000 replicates where the smoothing spline fit has a higher RMSED(1) than the median RMSED(1) from the septic polynomial. More worrying perhaps is that for estimates of f_2'' roughly 30 errors from the smoothing splines are considerably higher than the median error of the polynomial derivative estimate.

	Mean (sd)	Median
Quintic SS	1.27(0.19)	1.25
Septic Poly	0.43(0.07)	0.43

Table 3.1: Mean and median RMSE using smoothing splines for the simulated data.

Tables 3.1, 3.2 and 3.3 provide evidence of this outlier effect on the smoothing splines derivative estimates where the median and mean errors grow further apart.

Overall the smoothing spline method offers improved estimates for the rate of change and

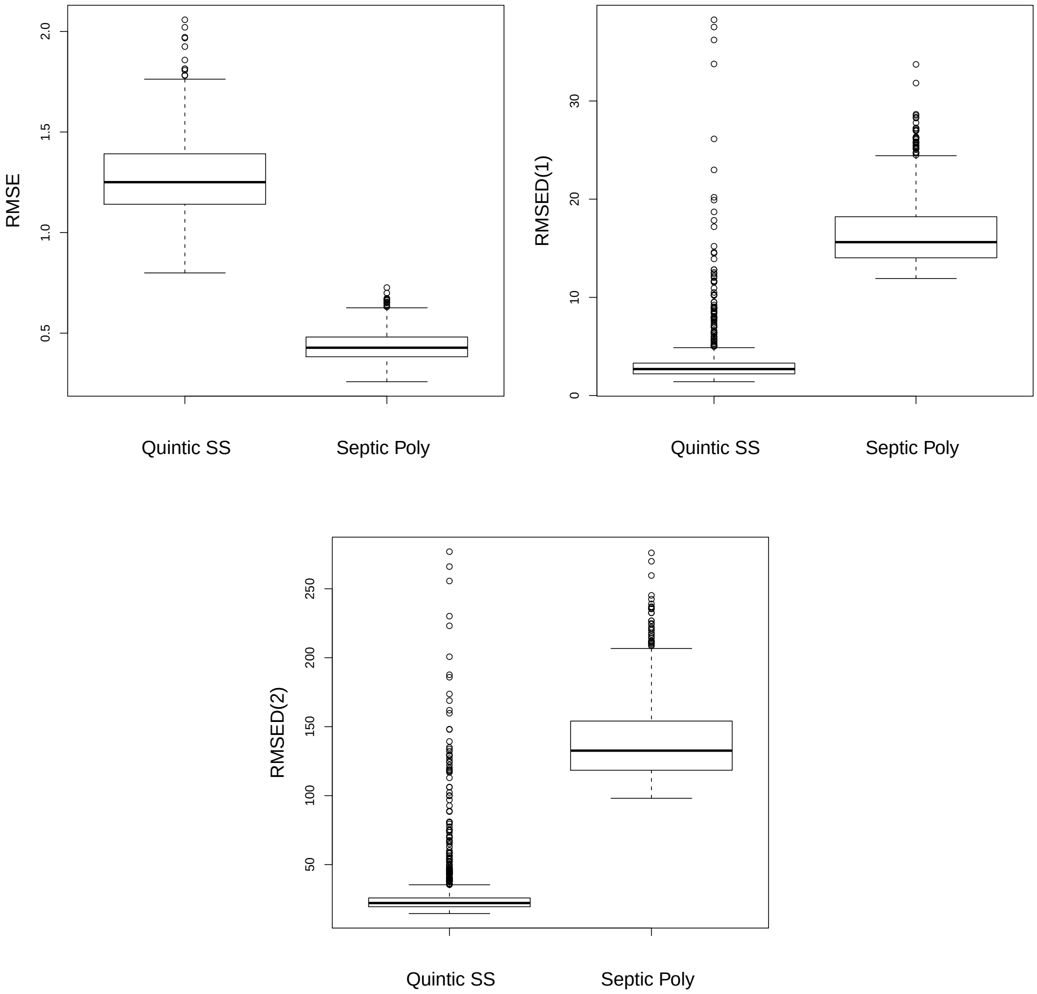


Figure 3.6: RMSE, RMSED(1) (top row) and RMSED(2) (bottom) for estimates of f_2, f_2' and f_2'' using smoothing splines and polynomial regression.

	Mean (sd)	Median
Quintic SS	3.58(3.29)	2.83
Septic Poly	16.5(3.38)	15.6

Table 3.2: Mean and median RMSED(1) using smoothing splines for the simulated data.

	Mean (sd)	Median
Quintic SS	31.7(29.5)	23.3
Septic Poly	139.4(28.6)	132.6

Table 3.3: Mean and median RMSED(2) using smoothing splines for the simulated data.

second derivative over polynomial regression. The improved model behind these estimates gives greater flexibility when fitting curves to noisy data. However, a problem with some outlying poor fits to f_2' and f_2'' has been uncovered. Figure 3.7 exhibits estimates of f_2 and f_2' for the poorest of the 1000 fits to f_2' using smoothing splines which had RMSED(1) $= 37.8$.

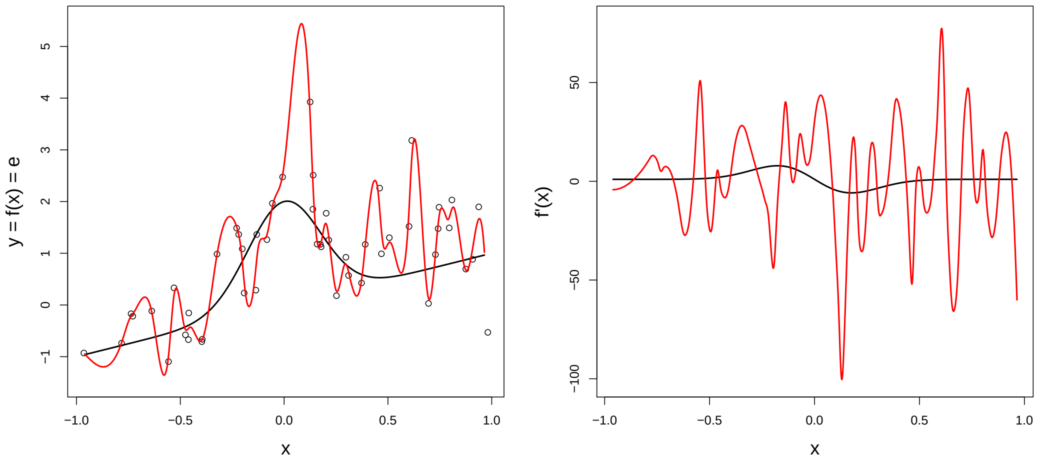


Figure 3.7: Estimates of f_2 and f_2' (bold curves) using smoothing splines (red curves).

Thankfully it seems the problem does not come from the derivative estimation procedure. The woefully undersmoothed estimate of f_2 causes the poor estimate of f_2'. In practice, the poor estimate of f_2 would be noticed and amended, leading to a better derivative estimate. The hugely erratic estimate of f_2 is caused by a value of λ which is too low and needs to be changed subjectively. Figure 3.8 provides several 'normal' estimates of f_2 along with the estimate from Figure 3.7 above so it can be seen how this estimate is very much atypical.

3.2.1.3 Altering the Value of λ

The rate of change of a variable is naturally more sensitive to outliers, and behaviour of derivatives is generally more erratic than for the functions underlying noisy data themselves. The smoothing parameter λ is chosen to minimise criteria pertaining to estimating f and not its derivatives. Thus it would seem intuitive to increase the smoothing parameter to achieve more settled derivative estimates. The R function `D1D2` in the **sfsmisc** package uses smoothing splines to find derivative estimates but adds a constant 'fudge' value to the cross-validation choice of λ. There is no record of this method in the literature and it exists only in this R package. Figure 3.9 compares results from using this with those found in the simulations from 3.2.1.2.

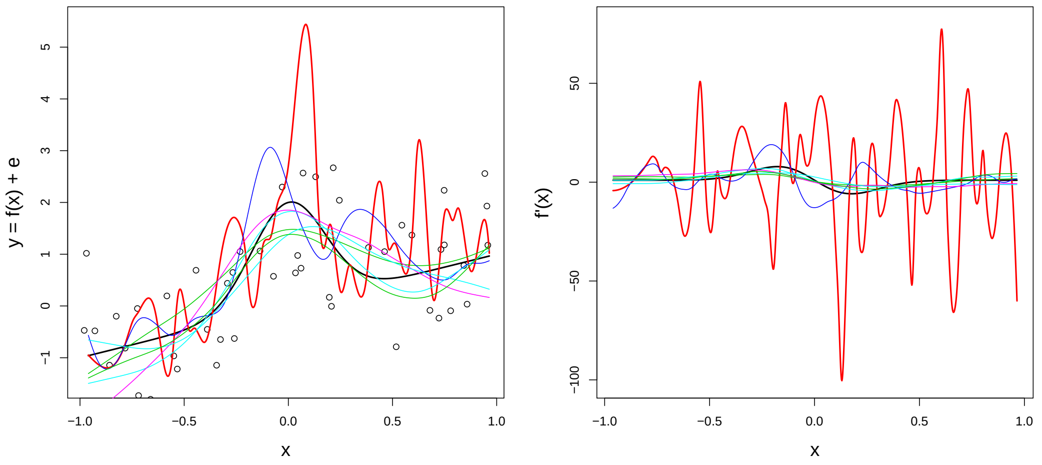


Figure 3.8: Using smoothing splines (multicoloured, with maximum error estimate in red) to estimate f_2 and f_2' (bold curves).

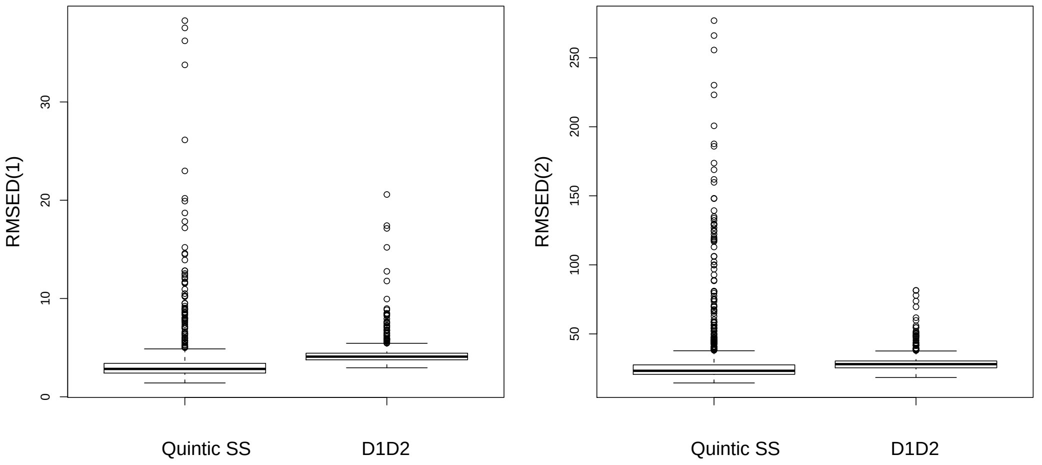


Figure 3.9: RMSED(1) (left) and RMSED(2) (right) for comparisons of quintic smoothing splines with `D1D2`.

The ‘fudged’ smoothing parameter does improve on the outlying errors but does not improve on the overall median or mean RMSED(·) for either first or second derivative estimates. The selection of the ‘fudge’ value is subjective with no automatic method provided and a mysterious default setting of 0.1384 is used. On closer inspection of the `D1D2` help file the following description of the ‘fudge’ is found:

"It is well known that for derivative estimation, the optimal smoothing parameter is larger (more smoothing) than for the function itself. `spar.offset` is really just a fudge offset added to the smoothing parameter. Note that in R's implementation of `smooth.spline`, `spar` is really on the $log\lambda$ scale.

When `deriv = 1:2` (as per default), both derivatives are estimated with the same smoothing parameter which is suboptimal for the single functions individually. Another possibility is to call `D1D2(*, deriv = k)` twice with $k = 1$ and $k = 2$ and use a larger smoothing parameter for the second derivative."

Increasing λ when estimates of f' and f'' are of interest eliminates some of the poorest fits to f' and f'' yet there is no noticeable overall improvement. From the results of these simulations there is no empirical evidence that this idea improves derivative estimation.

3.2.2 P-Splines

Another penalisation based spline smoothing method known as P-Splines (Eilers & Marx [12]) has become quite popular in the last decade. To begin, the K knots are (generally) spaced equally across the range of x. Little attention is paid to knot selection in the P-Spline framework, generally a modest number of knots is taken. Basis splines, or B-Splines (de Boor [9]), are used as the building blocks for the estimate of the underlying function explaining the data. A degree q B-Spline basis over K is made up of $m = K + q - 1$ B-Splines which are of degree q. Figure 3.10 displays a cubic B-spline basis over $K = 10$ knots, which results in $10 + 3 - 1 = 12$ cubic B-Splines.

B-Splines possess advantageous mathematical properties which make them the most common basis function in spline smoothing today. B-Splines offer flexibility as well as stability thanks to a banded structure of their design matrix. Except at the boundaries, each B-spline is positive over $q + 2$ knots (known as the compact support property) which leads to efficient computation since, at any given x, $q + 1$ B-Splines are non-zero. Another asset which makes B-Splines so attractive is that the bases are dispersed and lend themselves to large scale problems. From (3.7) the estimate $\hat{f}$ may be written as a combination of B-Splines and coefficients

$$f = B\alpha \tag{3.19}$$

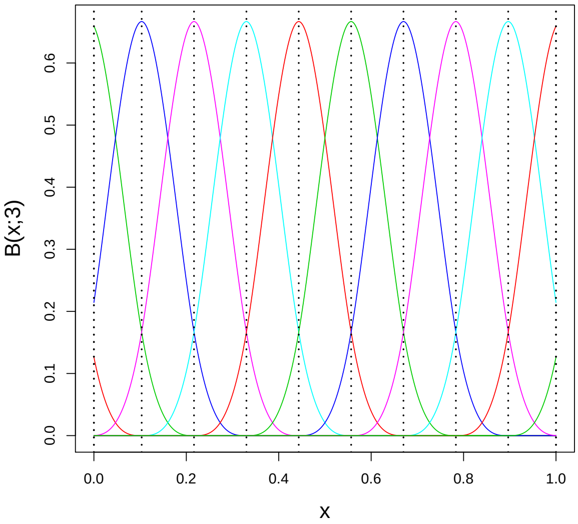


Figure 3.10: A cubic B-Spline basis using 10 knots which are marked by dashed lines.

where the B-Spline design matrix B is defined as

$$B = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \dots & B_m(x_1) \\ B_1(x_2) & B_2(x_2) & \dots & B_m(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & \dots & B_m(x_n) \end{pmatrix}.$$

The coefficients, α, of the B-Spline basis determine the scaling of the smooth estimate $\hat{f}$. Estimating α from (3.19) by minimising $(y - B\alpha)^2$ would lead to a close to interpolating fit (the fitted curve would oscillate K times). In order to avoid such an outcome, a roughness penalty is introduced on the difference between adjacent coefficients. This comes from the idea that if coefficients do not differ much from near neighbours, large jumps of the fitted curve are avoided. The difference penalty on the m coefficients is defined as

$$\Delta\alpha_j = \alpha_j - \alpha_{j-1} \tag{3.20}$$

for $j = 2, \ldots, m$. This may be written in matrix form as $\Delta\alpha = D\alpha$ where D is defined

$$D = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

for $m = 4$ and $n = 3$. The roughness of the fitted curve f is now measured as

$$\|\Delta\alpha\|^2 = \|D\alpha\|^2 . \tag{3.21}$$

There is a notational issue with P-Splines. In numerical analysis the term P-Splines refers to polynomial splines used in mathematical modelling whereas here it refers to penalised B-Splines. In a recent publication Eilers & Marx [13] propose changing the name of their method to PB-Splines to avoid confusion but for the course of this thesis P-Splines shall continue to refer to their method.

The function f is estimated by minimising

$$PENSS = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda (D_d \alpha)^2 \tag{3.22}$$

where the smoothing parameter λ controls the roughness of the fitted curve $\hat{f}$. Moreover, higher order difference penalties, d, may be selected, e.g.

$$\begin{aligned} D_2\alpha &= \Delta^2 \alpha_j \\ &= (\alpha_j - \alpha_{j-1}) - (\alpha_{j-1} - \alpha_{j-2}) \\ &= \alpha_j - 2\alpha_{j-1} + \alpha_{j-2} \end{aligned} \tag{3.23}$$

such that

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

and

$$\begin{aligned}D_3\alpha &= \Delta^3\alpha_j \\ &= (\alpha_j - 2\alpha_{j-1} + \alpha_{j-2}) - (\alpha_{j-1} - 2\alpha_{j-2} + \alpha_{j-3}) \\ &= \alpha_j - 3\alpha_{j-1} + 3\alpha_{j-2} - \alpha_{j-3}.\end{aligned} \tag{3.24}$$

Once again, λ may be selected subjectively by the user, although automatic methods, such as cross-validation, are generally used to optimise smoothing.

Derivative estimates are found as a by-product of the P-Spline fitting procedure. Once the fitted coefficients $\hat{\alpha} = (\hat{\alpha_1}, \ldots, \hat{\alpha_m})$ are obtained, derivatives are found using a formula due to de Boor [9]. For instance, the first derivative of a fitted B-Spline based curve with knots which are equally spaced is given by

$$\begin{aligned}f^{(1)}(x) &= \frac{\partial}{\partial x}\sum_{j=1}^{m} B_j(x;q)\alpha_j \\ &= (qh)^{-1}q\sum_{j=2}^{m}\Delta^1\alpha_j B_j(x;q-1)\end{aligned} \tag{3.25}$$

where h is the distance between adjacent knots. In general, the lth derivative of a fitted B-Spline curve with equally spaced knots is given by

$$f^{(l)}(x) = \prod_{l=1}^{L}((q+1-l)h)^{-1}(q+1-l)\sum_{j=l+1}^{m}\Delta^l\alpha_j B(x;q-l). \tag{3.26}$$

3.2.2.1 Application of P-Splines to the Blood Lactate Data

In the Blood Lactate data it is required to estimate the speed at which the maximum second derivative of the underlying lactate function occurs. The left panel of Figure 3.11 displays one individual's observed lactate at ten workloads on a treadmill along with a P-Spline smooth fit to the data.

The B-Spline basis used here is of degree 5 so that smooth second derivative estimates can be achieved. The right panel of Figure 3.11 exhibits the second derivative estimate of this lactate curve using P-Splines and the de Boor formula (3.26). The estimated maximum second

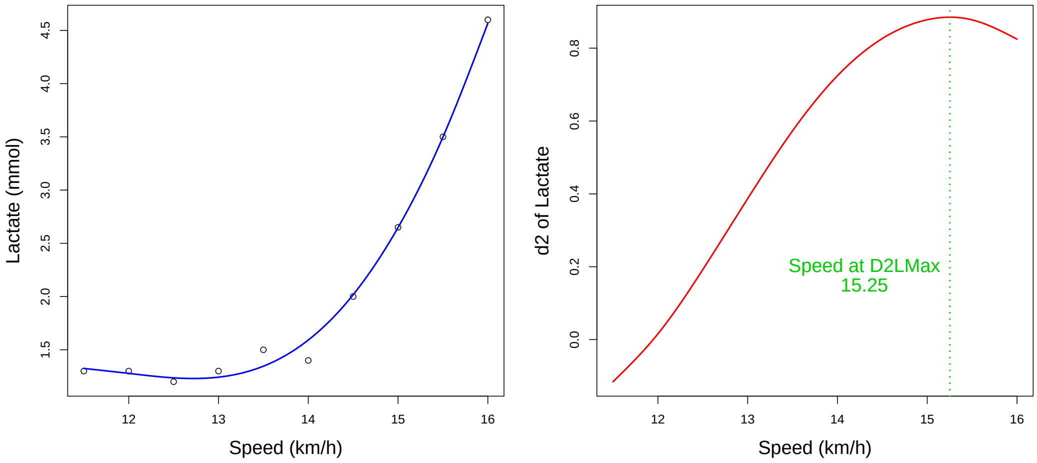


Figure 3.11: One individual's lactate data with estimate of lactate curve (left) and second derivative estimate (right) using P-Splines.

derivative of lactate is 0.885 which has a corresponding speed of 15.25 km/hr. This estimate could be used for comparison against other athletes in the group or longitudinally on the same individual if measured repeatedly over, say, a season. However, whether to trust this estimate is still an unknown. Simulations need to be performed to investigate how well P-Splines perform in terms of derivative estimation. These simulations will be carried out once more derivative estimation methods have been introduced.

3.2.2.2 Generalised Smoothing with P-Splines

Until now, the Grey Plover data have been treated as having a continuous response. Here the idea of generalised smoothing is introduced to deal with situations where a response variable is thought to come from a Poisson process such as when the response is a vector of counts. Derivative estimates are obtained from this model and are compared to estimates when the response is modelled as a continuous variable. Generalised smoothing and derivative estimation approaches to handling a Poisson response are now developed. The assumption of Poisson distributed counts, i.e. $y \sim \text{Pois}(\mu)$, can be rash as this relies on the mean and variance being equal. For the purposes of simplicity, this assumption has been made here.

Consider the familiar smoothing scenario

$$y = f(x) + \epsilon.$$

It is still required to estimate f which best describes the relationship between x and y, although now the assumption that $y \sim \text{Pois}(\mu)$ is made such that

$$ln(\mu) = B\alpha \tag{3.27}$$

i.e.

$$\mu = e^{B\alpha}, \tag{3.28}$$

where B is the B-Spline design matrix and α is the vector of coefficients. The likelihood function is

$$L = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}.$$

Taking logs

$$ln(L) = l = \sum_{i=1}^{n} (y_i ln\mu_i - \mu_i - ln(y_i!))$$

and substituting (3.27) and (3.28) gives

$$l = \sum_{i=1}^{n} \left(y_i \sum_{j} B_j \alpha_j - e^{\sum_j B_j \alpha_j} - ln(y_i!) \right)$$

where $j = 1, \ldots, m$ is the number of coefficients to be estimated.

Applying the difference penalty on the α_j, the likelihood is written as

$$l^* = l - \frac{1}{2}\lambda \left\| D_d \alpha \right\|^2.$$

Differentiating with respect to α and setting to zero leads to the iteratively weighted least squares solution for the coefficients α_j, i.e.

$$\hat{\alpha}_{t+1} = (B^T \hat{W}_t B + \lambda D_d^T D_d)^{-1} B^T \hat{W}_t \hat{z}_t \tag{3.29}$$

where

$$z = B\alpha + W^{-1}(y - \mu)$$

and

$$\hat{W} = diag(\hat{\mu})$$

$$H = B(B^T\hat{W}B + \lambda D_d^T D_d)^{-1}B^T$$

with

$$\hat{\eta} = B\hat{\alpha}$$

which can be rewritten

$$B\hat{\alpha} = H\hat{z}.$$

Once the coefficients α_j have been estimated, derivative may be estimated using the de Boor formula (3.26). Here

$$\hat{f} = e^{B\hat{\alpha}}$$

such that

$$\hat{f}' = e^{B\hat{\alpha}}\frac{1}{h}B'\Delta\alpha \tag{3.30}$$

3.2.2.3 Application of *P*-Splines to the Grey Plover Example

Recall that the Grey Plover data consist of 31 counts taken annually between 1974 and 2004. In all examples using these data to date, the response has been treated as a continuous one. Since the response is a variable of counts this should be modelled accordingly. Using the standard *P*-Spline approach will be less effective since it does not take into account the constraints on the response, i.e. that they must be nonnegative integers. The left plot in Figure 3.12 exhibits smooth fits of the Grey Plover data using *P*-Splines where the count variable has been considered (blue) and not taken into account (red).

The standard *P*-Spline fit is slightly more smooth than the generalised approach. The same number of knots ($K = 6$), basis degree ($q = 3$), penalty order ($d = 2$) and method for smoothing parameter selection (CV) were taken such that any difference between the curves is

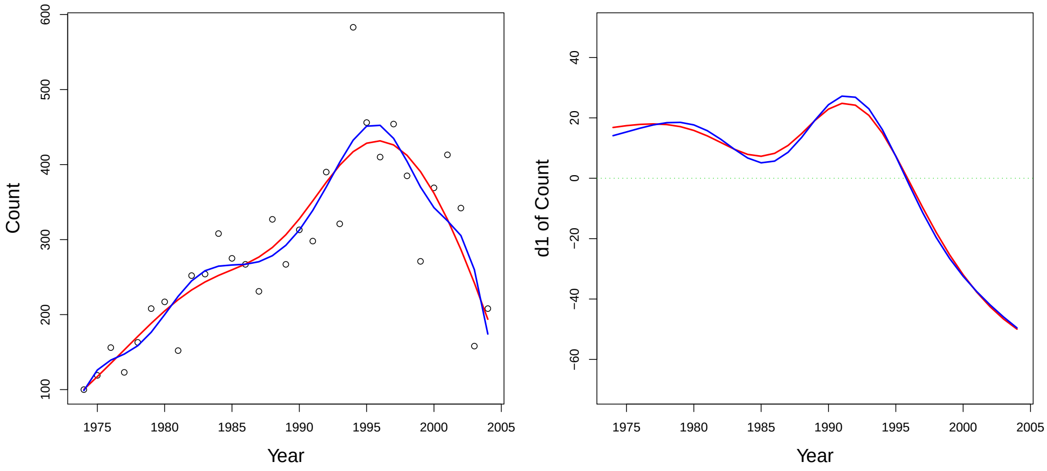


Figure 3.12: P-Spline fit (left) and derivative estimate (right) for the Grey Plover data modelled as counts (blue) and as continuous (red).

purely down to the standard versus generalised smoothing approach. The right panel of Figure 3.12 exhibits first derivative estimates obtained from each approach. Again, the generalised P-Spline approach is slightly more ‘wiggly’ and, given the better modelling approach taken, is likely to be a better estimate of the true underlying first derivative.

3.2.3 Mixed Model Smoothing

Among the many desirable attributes of mixed or random effects models is their applicability to the smoothing of noisy data. The use of mixed models in smoothing is comprehensively reviewed in the seminal paper by Wand [63] which also offers an excellent summary of the background to mixed model smoothing, also known as semiparametric regression. A fixed or parametric component such as that discussed in Section 2.2 is combined with a random component which allows for the flexibility needed to describe nonlinear relationships. The general linear mixed effects model is

$$y = X\beta + Zu + \epsilon \tag{3.31}$$

where X and Z are the fixed and random design matrices respectively, β are the fixed parameters, u are the random effects with zero mean and covariance matrix G and ϵ is the error vector with zero mean and covariance matrix R. Assuming that ϵ and u are uncorrelated leads to

$$Var \begin{pmatrix} u \\ \epsilon \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}. \tag{3.32}$$

Typically the ϵ_i are assumed to be *iid* Normal with variance σ_ϵ^2 such that $R = \sigma_\epsilon^2 I$ and similarly the u_k are thought to depend solely on the single variance parameter σ_u^2, i.e. $G = \sigma_u^2 I$. Recalling the linear spline estimator (3.4), the vector of random effects u of length K may be substituted to give

$$y = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k (x - \kappa_k)_+. \tag{3.33}$$

From (3.31), the fixed design matrix X becomes

$$X = [1 \ x],$$

the random design matrix Z corresponds to

$$Z = [(x - \kappa_1)_+ \ \ldots \ (x - \kappa_K)_+]$$

and

$$\begin{pmatrix} u \\ \epsilon \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 I & 0 \\ 0 & \sigma_\epsilon^2 I \end{bmatrix} \right).$$

Solving for β, σ_ϵ and σ_u using maximum likelihood and u by best prediction (see Chapter 9, McCulloch & Searle [36]) is equivalent to solving

$$\|y - X\beta - Zu\|^2 + \lambda \|u\|^2 \tag{3.34}$$

where $\lambda \equiv \sigma_\epsilon^2 / \sigma_u^2$ is the familiar smoothing parameter controlling the balance between the least squares fit to the data $\|y - X\beta - Zu\|^2$ by penalising large values of u. The solution of this minimisation problem is

$$\begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = (C^T C + \lambda D)^{-1} C^T y \tag{3.35}$$

where $C = [X\ Z]$ and $D = diag(0, 0, 1, \ldots, 1)$, i.e. the number of zeros beginning the diagonal of D corresponds to the number of fixed parameters (e.g. β_0, β_1) used in the modelling process.

Derivative estimation from a mixed model smooth can be demonstrated more clearly using a cubic spline estimator i.e.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} u_k (x - \kappa_k)_+^3 \tag{3.36}$$

so that $X = [1\ x\ x^2\ x^3]$ and $Z = [(x - \kappa_1)_+^3\ \ldots\ (x - \kappa_K)_+^3]$ in (3.31). The parameters β, u are estimated as in (3.35) and an estimate for the rate of change of y relative to x is given by

$$\hat{y}' = \hat{\beta}_1 + 2\hat{\beta}_2 x + 3\hat{\beta}_3 x^2 + \sum_{k=1}^{K} 3\hat{u}_k (x - \kappa_k)_+^2 \tag{3.37}$$

such that $X = [0\ 1\ x\ x^2]$ and $Z = [3(x - \kappa_1)_+^2\ \ldots\ 3(x - \kappa_K)_+^2]$ in (3.31). An estimate of the second derivative of f is

$$\hat{y}'' = 2\hat{\beta}_2 + 6\hat{\beta}_3 x + \sum_{k=1}^{K} 6\hat{u}_k (x - \kappa_k)_+ \tag{3.38}$$

such that $X = [0\ 0\ 2\ 6x]$ and $Z = [6(x - \kappa_1)_+\ \ldots\ 6(x - \kappa_K)_+]$ in (3.31). Once more, it is advisable to choose the degree of the basis to be $l + 2$ where l is the highest order derivative estimate required for analysis. Other bases may be used for mixed model smoothing and therefore derivative estimation, including the B-Spline basis (Currie & Durban [8]) and radial basis functions (Wand [63]).

3.2.3.1 Application of Mixed Model Smoothing to the Astronomical Illustration

In the Astronomical example it is first required to smooth both the temperature and gas profile data and to estimate the rate of change of the vector resulting from the multiplication of these smooths. The top row of Figure 3.13 displays semiparametric regression fits using a truncated polynomial basis of degree 5.

The three temperature smooths along with their average appear jagged, which is due to the small number ($n = 8$) of observations here. The mean temperature smooth was predicted at the x's for the observed gas values and the bottom panel of Figure 3.13 shows the resulting

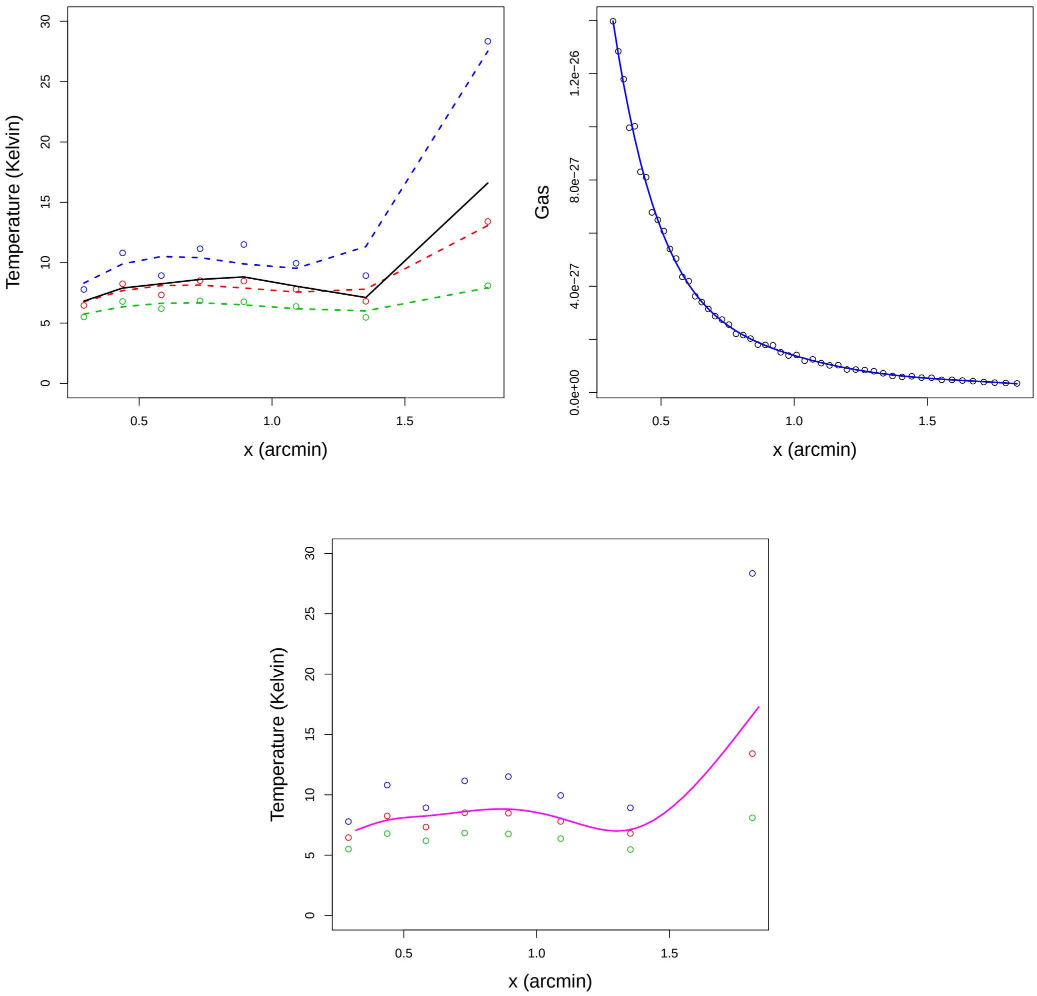


Figure 3.13: Top row: Semiparametric smooth fits to the temperature (black smooth is the average of the coloured smooths) and gas profiles from the astronomical data. Bottom: Temperature predicted at x for which ρ_{gas} was observed (purple).

curve.

The fitted values for the combination of the average temperature smooth with the smooth fit to the gas data are displayed in the left panel of Figure 3.14. Figure 3.14 also gives an estimate for the rate of change of $T\rho_{gas}$ using semiparametric method described in Section 3.2.3. As one moves away from the centre of the galaxy cluster, the decreasing of $T\rho_{gas}$ appears to level off. Whether this estimate is accurate cannot be assured based on the fact that the rate of change is unknown. In order to study the performance of this estimator, simulations must be carried out and these will follow toward the end of this Chapter.

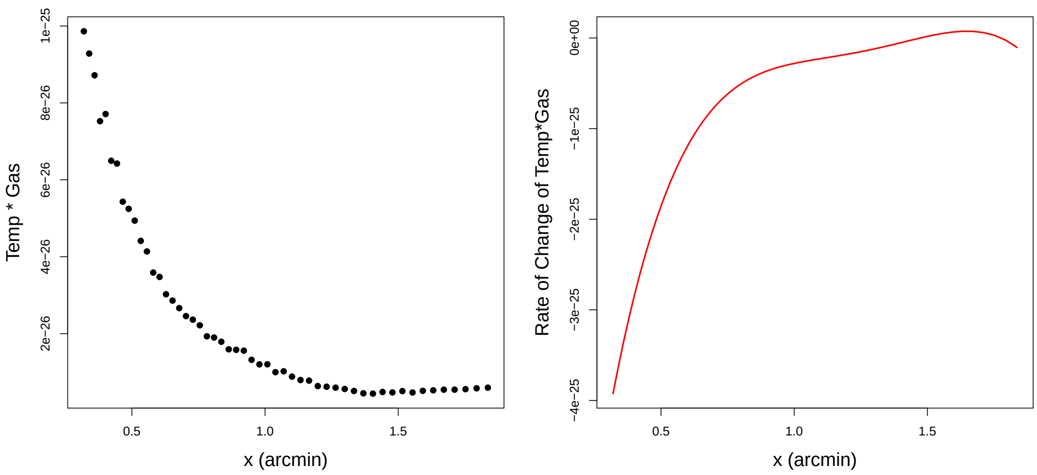


Figure 3.14: $T\rho_{gas}$ (left) with estimate for the rate of change using semiparametric smoothing (right)

3.2.4 Spatially Adaptive Smoothing

In smoothing splines, P-Splines and semiparametric regression, the smoothing parameter (λ) which acts to balance the bias-variance trade-off (Section 2.1.2), has been treated as constant. However, in situations where data display heteroscedacity, i.e. non constant variance over the range of explanatory variable, there are obvious concerns with the single constant λ approach. These situations can benefit from yet more flexibility, where the value of λ is allowed to vary to adjust to the needs of the data. Early spatially adaptive methods due to Friedman [15] and Luo & Wahba [32] offer a glimpse at the potential of such models. In the following sections two more recent methods are summarised and used for analysis in a derivative estimation context.

3.2.4.1 Spatially Adaptive Penalised Splines

The method of Ruppert & Carroll [47] uses a large number K of knots with the kth knot κ_k placed at the $k/(K+1)$th sample quantile of x_i, for $i = 1, \dots, n$ observations. They use a truncated polynomial basis such that the model is written

$$y = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{k=1}^{K} \beta_{p+k}(x - \kappa_k)_+^p \tag{3.39}$$

where $\beta = (\beta_0, \ldots, \beta_p, \beta_{p+1}, \ldots, \beta_{p+K})$ is the vector of coefficients, $p \geq 1$ is an integer defining the degree of the basis and $\kappa_1 < \cdots < \kappa_K$ are fixed knots. They use a penalty on β_{p+k} for $k = 1, \ldots, K$ and $\hat{\beta}$ are estimated by minimising

$$\sum_{i=1}^{n} [y_i - f(x)]^2 + \sum_{k=1}^{K} \lambda(\kappa_k)\beta_{p+k}^2 \tag{3.40}$$

where $\lambda(\kappa_k)$ is a penalty function. To achieve a spatially varying smoothing parameter function $\lambda(\kappa_k)$, a set of subknots $\kappa_1^*, \ldots, \kappa_M^*$ is chosen where $M < K$ such that $\kappa_1^* = \kappa_1 < \cdots < \kappa_M^* = \kappa_K$. The penalty at each subknot κ_k^* is controlled with a smoothing parameter λ_k^*. The penalties at κ_k for $k = 1, \ldots, K$ are determined using interpolation of the subknot penalties. This leads to each knot κ_k having its own penalty $\lambda(\kappa_k)$ relying solely upon $\lambda^* = (\lambda_1^*, \ldots, \lambda_M^*)^T$, i.e. $\lambda(\kappa_k)$ for $k = 1, \ldots, K$ is a function of λ^*.

Fitted coefficients $\hat{\beta}_{\lambda^*}$ are found using

$$\hat{\beta}_{\lambda^*} = (X^T X + D_{\lambda^*})^{-1} X^T y \tag{3.41}$$

where X is a design matrix of the form $X = [1 \; x \; x^2 \; \ldots \; x^p \; (x - \kappa_1)_+^p \; \ldots \; (x - \kappa_K)_+^p]$ and D_{λ^*} is the penalty matrix with $p + 1$ zeros beginning the diagonal followed by $\lambda(\kappa_1), \ldots, \lambda(\kappa_K)$. Ruppert & Carroll [47] recommend selecting λ^* using GCV, i.e.

$$GCV_{\lambda^*} = \left(\frac{y - \hat{y}}{1 - df_{\lambda^*}/n} \right)^2 \tag{3.42}$$

where the degrees of freedom of the fit is the trace of the hat matrix

$$df_{\lambda^*} = tr(X(X^T X + D(\lambda^*))^{-1} X^T). \tag{3.43}$$

Selecting M values $\lambda_1^*, \ldots, \lambda_M^*$ simultaneously would require a large computational effort to search over an M dimensional space. To avoid this it is recommended to first set each $\lambda_1^*, \ldots, \lambda_M^*$ equal to the best global λ using GCV. Then each λ_k^* is varied sequentially (with $\lambda_1^*, \ldots \lambda_{k-1}^*, \lambda_{k+1}^*, \ldots, \lambda_M^*$ fixed) on a one dimensional grid centred at λ_k^*. The value λ_k^* is replaced with the global λ chosen by GCV on this grid. This process reduces the computational cost using M one dimensional searches in place of one M dimensional search. The iterative search can be repeated N_{iter} times such that the user-defined inputs to the method consist of K, M and N_{iter}. The authors provide simulations over combinations of these three and recommend that $M > 6$ and $N_{iter} > 2$ be avoided due to the computational cost required.

Once $\hat{\beta}_{\lambda^*}$ have been obtained it is simple to find derivative estimates as a by product of (3.40) using

$$\hat{f}'(x;\beta) = \hat{\beta}_1 + 2\hat{\beta}_2 x + \cdots + p\hat{\beta}_p x^{p-1} + \sum_{k=1}^{K} p\hat{\beta}_{p+k}(x-\kappa_k)_+^{p-1}. \tag{3.44}$$

No R library exists for the implementation of this approach, however MATLAB code was kindly sent by Professor Ruppert so that testing of this method could be undertaken (after the code was updated). The spatially adaptive smoothing parameter should offer improved flexibility which can handle the rather sensitive estimation of rate of change of noisy data and this will be tested against the constant smoothing parameter methods presented to date.

3.2.4.2 Adaptive Mixed Model Smoothing

The approach taken by Krivobokova et al. [27] is similar to the Ruppert & Carroll [47] model in that a set of subknots is taken on a truncated polynomial basis. However, they use both fixed and random coefficients, i.e. a pth degree model is of the form

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{k=1}^{K} u_k (x-\kappa_k)_+^p \tag{3.45}$$

where the number of knots is $K \geq min(n/4, 40)$ as per the suggestion of Ruppert [46]. For the truncated polynomial basis, the random coefficients are generally taken to be $u_k \sim N(0, \sigma_u^2 I)$ and $f(x)$ is found by minimising (3.34), i.e. β and u_k are obtained from (3.35). Semiparametric regression, as discussed in Section 3.2.3, uses just the single parameter σ_u^2 to shrink all of the random coefficients. The method allows for $u_1, \ldots, u_K$ to have different prior variances

$$u_k \sim N(0, \sigma_u^2(\kappa_k)) \tag{3.46}$$

for $k = 1, \dots, K$. The $\sigma_u^2(\kappa_k)$ are treated as a coming from a smooth function modelled as a log penalised spline

$$\sigma_u^2(\kappa) = \exp\left(\gamma_0 + \gamma_1\kappa + \cdots + \gamma_q\kappa^q + \sum_{c=1}^{C} v_k(\kappa - \kappa_c^*)_+^q\right) \tag{3.47}$$

where κ_c^*, $c = 1, \dots, C$ is a set of subknots covering the range of $\kappa_1, \dots, \kappa_K$, $\gamma_0, \dots, \gamma_q$ are the fixed and $v_k \sim N(0, \sigma_v^2)$ the random coefficients defining the $\sigma_u^2(\kappa_k)$ process for some sub-basis of degree $q < p$.

The parameters β, u, γ, v, σ_ϵ, σ_u and σ_v are estimated using ML or REML (see Krivobokova et al. [27] for full description). The R library **AdaptFit** accompanies their paper and the function `asp` allows for (somewhat) automatic implementation of the above. First derivative estimates can be obtained as a by-product of the model (3.45)

$$\hat{f}'(x) = \hat{\beta}_1 + 2\hat{\beta}_2 x + \cdots + p\hat{\beta}_p x^{p-1} + \sum_{k=1}^{K} \hat{u}_k p(x - \kappa_k)_+^{p-1}. \tag{3.48}$$

3.2.4.3 Application of Adaptive Methods to the Winter Nutrients Data

The spatially adaptive methods are suited to situations where data are smooth in one region and quite variable in another. The Winter Nutrients data, in particular the Phosphate measurements, exhibit this type of behaviour. Both adaptive smoothing methods are applied here to estimate the rate of change of the Phosphate observations.

Figure 3.15 displays smooth fits to the data along with estimates for the rate of change of Phosphate using both adaptive methods introduced in Section 3.2.4. The Phosphate data exhibit steady increases except between days 90 and 100 where a sudden jump in Phosphate level occurs. The adaptive/varying penalty methods should achieve superior results in fitting a curve to the data since this situation is what they were intended for.

The derivative estimates are quite different between the methods, with the adaptive mixed model approach having a much larger estimate for the size of the aforementioned jump in Phosphate than the spatially varying penalty method. It is clear from the right panel of Figure 3.15 that the adaptive mixed model fit takes full advantage of its adaptability. It is very smooth in the tails but is hugely descriptive at the jump between measurement 60 and 100. A smooth

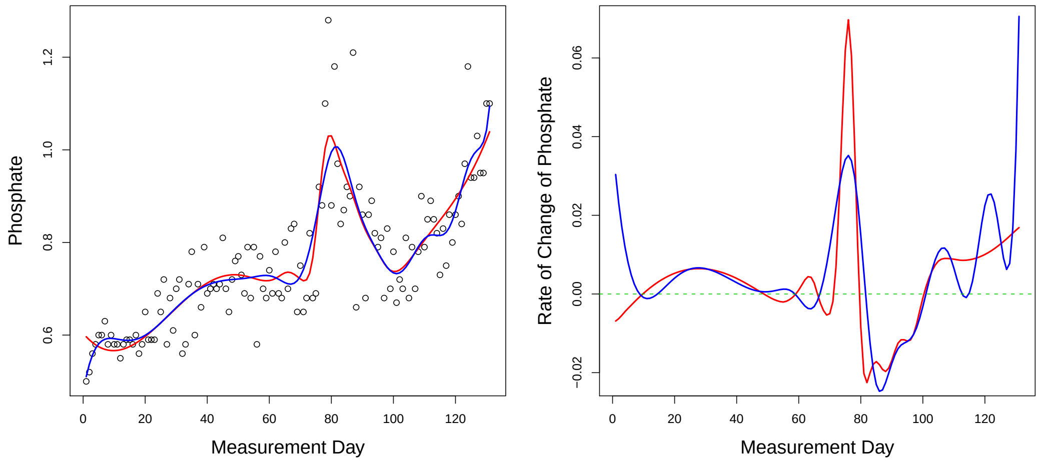


Figure 3.15: Adaptive estimate of the function underlying the Phosphate data (left) and of the rate of change of the Phosphate data (right) using spatially adaptive splines (blue) and the adaptive mixed model methods (red).

with a single constant smoothing parameter would be unable to achieve this fit.

Given the added flexibility allowed in these modelling approaches one would tend to believe these estimates of the data over methods employing a single fixed smoothing parameter λ. Whether this translates to better derivative estimates is a matter to be decided based on empirical investigations which now follow.

3.2.5 Simulation Study to Compare Derivative Estimation for Spline Smoothing Methods

A simulation study was performed to compare the performance of the derivative estimation methods which have been discussed in this Chapter. Under investigation were the quality of first and second derivative estimates found using smoothing splines, P-Splines, semiparametric regression, spatially adaptive splines and adaptive mixed model smoothing. Each approach was measured for performance using the RMSED(1) and RMSED(2) criteria.

The function $f_2 = x + 2e^{-16x^2}$ was used to mirror a typical noisy data situation. A thousand samples of $n = 50$ response values $y = f_2(x) + \epsilon$ were simulated with $\epsilon \sim N(0, \sigma^2)$, $\sigma = \frac{1}{3}range(f_2)$ and x uniform on $[-1, 1]$.

The boxplots in Figure 3.16 summarise the results of these simulations. The median RMSED(1) seems similar across the five approaches. What is striking is the difference in precision of performance. The mixed model approaches have more stable error than the penalised spline techniques, with smoothing splines and spatially adaptive splines displaying many RMSED(1) outliers and all three non-mixed model methods having many RMSED(2) outliers. This has been discussed and explored previously in Section 3.2.1.2 and wildly fluctuating smooth fits to the data were again discovered to be the cause.

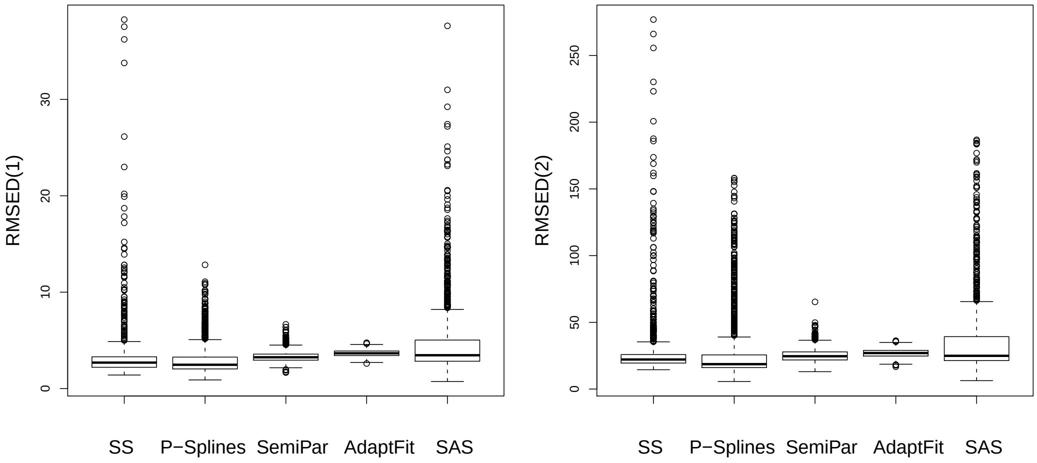


Figure 3.16: RMSED(1) and RMSED(2) for estimates of f_2' and f_2''.

In Table 3.4 the results appear to suggest that P-Splines (first derivative) and semiparametric smoothing (second derivative) achieve the best overall goodness of fit. In these simulations the mixed model techniques offer the more stable derivative estimation procedures.

Method	RMSED(1)	RMSED(2)
Smoothing Splines	3.58(3.29)	31.7(29.5)
P-Splines	3.09(1.77)	29.9(27.8)
Semiparametric Smoothing	3.34(0.61)	25.3(5.28)
Adaptive MM Smoothing	3.66(0.35)	26.9(3.13)
Spatially Adaptive Splines	4.90(4.01)	39.3(31.4)

Table 3.4: Mean (standard deviation) RMSED(1) and RMSED(2) comparing the derivative estimation performance of spline smoothing methods.

The results of simulations seem to show the P-Spline and semiparametric methods perform best out of the five methods discussed in terms of derivative estimation. Using one run of simulations from the function f_2 these results are investigated further. Figure 3.17 displays smooth estimates of f_2 and f_2' using each method.

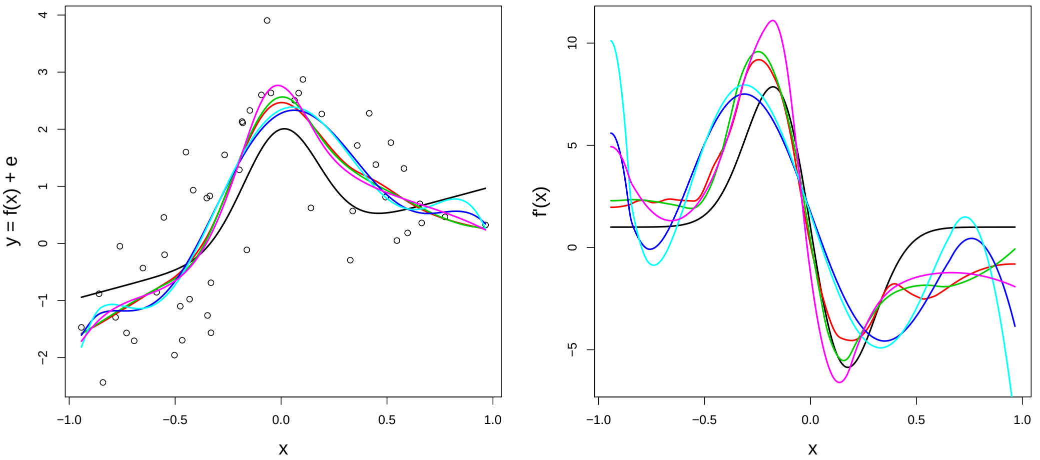


Figure 3.17: Estimates of f_2 (black curve, left) and f_2' (black curve, right) using smoothing splines (red), P-Splines (green), semiparametric smoothing (blue), adaptive mixed models (cyan) and a spatially varying penalty (purple).

Very similar results are found when estimating f_2. The specific error in this one run has caused all methods to overestimate the true function, where each sits above the actual function f_2 when the jump at $x \approx 0$ occurs.

The results for estimating f_2' show less similarities than those for f_2. None of the methods here achieves a very good estimate. The non-mixed model methods all follow the same pattern as the actual rate of change but display a scaling error nearly throughout the function. The methods exhibit superfluous oscillations (undersmoothing) in the tails, i.e. boundary effects. Subjectively it is difficult to choose a 'winner' from the right panel of Figure 3.17 but the P-Spline fit seems to be the most 'accurate' here.

3.3 Derivative Estimation Using Kernels

The other main branch of nonparametric smoothing methods is known collectively as Kernel Smoothing, with several subsidiary methods also available. Kernel smoothing uses local polynomials fitted in the neighbourhood of a point x_i $(i = 1, \ldots, n)$ using weighted least squares.

The size of neighbourhood used, known as the bandwidth, determines the smoothness of the fit $\hat{f}$.

The decision to use splines or kernels is arbitrary, there is no consensus on which to choose. Some statisticians simply prefer kernels to splines or splines to kernels, perhaps through past experience or 'ease' of use. However, there are some clear cut advantages for each. For example, spline smoothing methods integrate into a mixed model smoothing approach seamlessly whereas kernels are often recommended when dealing with huge datasets of, say, a million observations. Kernels are also more popular in density estimation.

There are many papers which mention derivative estimates from kernel smoothing, however, similarly to spline derivative estimation, few which deal with the problems encountered in this thesis. Schwartz [52] and Wahba [60] give early but brief discussions on density derivative estimation. Stone [57] [58] investigates convergence of derivative estimators. Stoker [56] discusses the bias inherent in derivative estimates of a regression function. Müller et al. [37] discuss changing the bandwidth (the bandwidth in kernel smoothing is used in much the same way as the smoothing parameter in spline smoothing) in order to deliver better derivative estimates. Hastie & Loader [20] propose reducing bias in derivative estimation by using higher order polynomials in the fitting process. Fan & Gijbels [14] offer an adaptive bandwidth which "works out very neatly for derivative estimation". Xia [65] suggests using Gaussian kernels for derivative estimation because of their smoothness. Mack & Müller [33], Ruppert & Wand [49], Welsh [64], Lai & Chu [28], Huh & Carrierre [22] and Prewitt & Lohr [41] introduce their own methods for kernel smoothing which work well when estimating derivatives compared to basic kernel smoothing methods.

There is clearly more research for derivative estimation involving kernel smoothing than with splines. Comparisons between the two branches of smoothing, in terms of estimating derivatives are quite rare and a rigorous review of current methods across all smoothing approaches does not exist.

3.4 Chapter Summary

Finding derivatives as a by-product of a model used to fit a regression function seems to give reasonable estimates, which far outperform using the data alone, i.e. first and second order differencing. Several sophisticated models for estimating the regression function underlying observed noisy data have been introduced, with derivative estimates obtained as a by-product of these models. Simulations were used to reveal that smoothing with P-Splines or using semi-parametric regression achieves the best performance for estimates of first and second derivatives respectively. Some tuning of these smoothing methods has been attempted (altering λ) but this did not achieve much, if any, improvement. Adaptive smoothing methods have shown little to no improvement in derivative estimation over standard constant smoothing parameter techniques.

In order to discover the causes of the problems encountered in derivative estimation, an investigation into the effects of sample size, variability and method for selection of smoothing parameter are carried out. For this, P-Splines are chosen because they offer efficient computational properties and perform well in comparison to the other techniques. Moreover, the P-Spline model involves several choices which may affect the estimate $\hat{f}$ and so in the next Chapter a slight diversion is taken to provide guidelines on these P-Spline components when derivative estimation is of primary importance.

Chapter 4

Components of a P-Spline Model

P-Splines offer an intuitive, flexible and computationally efficient procedure for fitting curves to noisy data. Moreover, derivative estimates found using P-Spline smooths perform well in comparison to other spline smoothing methods (Section 3.2.5).

A common issue with the use of P-Splines is that the user often feels overwhelmed by the plethora of choices (i.e. λ, K, d and q) in the underlying mechanics of the model. In this Chapter the components of P-Spline smoothing are investigated to judge the sensitivity of derivative estimates to the alteration of these choices. Simulation studies into the effects of sample size, variability and the methods for selecting a smoothing parameter are included in an attempt to discern a path to improvement for derivative estimation.

4.1 Penalties, Knots and Basis Degree

Eilers & Marx [12] do not regard the choice of penalty order (d), number of knots (K) or B-Spline basis degree (q) as important to the performance of the P-Spline smoothing model. They recommend choosing $q = 3$, $d = 2$ and $K = min(40, n/5)$. Unfortunately the uninitiated may find these choices difficult to deal with.

4.1.1 Simulation Study into the Choices of P-Spline Components

Here the mechanics of the P-Spline model are varied and performance measured in order to examine the effect these choices have on the accuracy of an estimate.

A thousand replicates of $n = 50$ responses $y = f_i(x) + \epsilon$, $i = 1, \ldots, 6$ (from each of the six functions introduced in Chapter 1), were simulated with x uniform on $[0, 1]$, $\epsilon \sim N(0, \sigma^2)$ and $\sigma = \frac{1}{6} range(f_i)$. Each of the three parameters were varied and performance measured using the RMSED(1) and RMSED(2) for penalty order $d = 1, 2, 3, 4$, number of knots $K = 10, 20, 30, 40$ and B-Spline basis degree $q = 3, 5, 7, 9$.

4.1.1.1 Penalty Order

Recall the roughness penalty involved in the P-Spline framework is a penalty on the difference between adjacent coefficients, i.e.

$$\Delta\alpha_j = \alpha_{j+1} - \alpha_j$$

for $j = 1, \ldots, m - 1$. Eilers & Marx [12] recommend using a second order difference penalty

$$\Delta^2\alpha_j = \alpha_{j+2} - 2\alpha_{j+1} + \alpha_j.$$

Raising the order of the penalty forces more coefficients to 'hold hands', i.e. be closer together. Tables 4.1 and 4.2 summarise performance when using $d = 1, 2, 3, 4$ for derivative estimation.

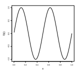
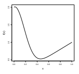
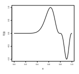
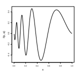
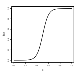
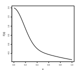
Function	$d = 1$	$d = 2$	$d = 3$	$d = 4$
 f_1	4.51(0.49)	5.15(0.50)	4.62(0.52)	3.64(0.40)
 f_2	0.58(0.27)	0.67(0.22)	0.54(0.19)	0.46(0.19)
 f_3	4.92(0.96)	5.22(1.05)	5.10(1.01)	4.80(0.83)
 f_4	0.62(0.11)	0.53(0.10)	0.50(0.16)	0.70(0.23)
 f_5	0.48(0.10)	0.56(0.09)	0.66(0.09)	0.67(0.08)
 f_6	1.42(1.19)	1.60(0.94)	1.50(0.73)	1.42(0.72)

Table 4.1: Mean RMSED(1) (standard deviation) comparing penalty orders.

Using a high order difference penalty, $d = 4$, results in the best performance in estimating f_1', f_2', f_3' and f_6' while the recommended $d = 2$ penalty order leads to the poorest performance for those four first derivatives.

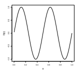
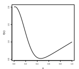
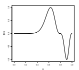
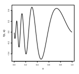
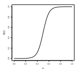
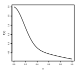
Function	$d = 1$	$d = 2$	$d = 3$	$d = 4$
 f_1	52.13(4.17)	58.85(4.79)	58.22(4.47)	75.83(11.28)
 f_2	18.22(10.64)	17.86(7.33)	14.44(5.92)	13.27(4.92)
 f_3	176.3(51.31)	177.7(47.19)	185.8(46.63)	194.2(47.27)
 f_4	19.63(20.65)	20.35(20.62)	21.26(20.30)	20.86(20.33)
 f_5	8.54(1.09)	9.87(1.22)	11.03(1.31)	12.34(1.42)
 f_6	38.00(44.83)	36.60(38.26)	30.04(27.14)	27.45(22.60)

Table 4.2: Mean RMSED(2) (standard deviation) comparing penalty orders.

When estimating second derivatives, $d = 1$ offers the lowest RMSED(2) for f_1, f_3, f_4 and f_5 but the highest RMSED(2) for f_2 and f_6! There is no apparent consistent evidence from these simulations for an optimum choice of penalty.

4.1.1.2 Knot Selection

Eilers & Marx [12] present little guidance on the choice of the number or placement of the knots which break up the data into local segments. Rather than allow the knots to dictate the modelling process, they recommend a modest number of knots $K = min(40, n/5)$ and use the smoothing parameter λ to control the smoothness of the fit. Moreover, several nice properties (such as the de Boor [9] formula (3.26)) exist when the knots are equally spaced along the range of x and so equally spacing knots is recommended. Tables 4.3 and 4.4 display RMSED(1) and RMSED(2) respectively at $K = 10, 20, 30, 40$ knots equally spaced across the range of x.

The recommended choice here at $n = 50$ would be $K = 10$ using $K = min(40, n/5)$. This choice leads to the best performance in estimating first derivative of f_2 and f_6 but also leads to the worst estimates of f_1', f_3', f_4' and f_5'. Choosing $K = 30$ is best for f_1', f_4' and f_5' and $K = 40$ offers the best estimates of f_3' but the worst for f_2' and f_6'. It is difficult to see a consistent pattern from these simulations and the results when estimating second derivatives do not resolve matters.

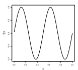
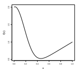
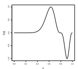
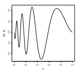
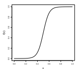
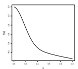
Function	$K = 10$	$K = 20$	$K = 30$	$K = 40$
 f_1	5.15(0.50)	1.55(0.32)	1.13(0.30)	1.25(0.32)
 f_2	0.67(0.22)	0.82(0.48)	0.90(0.71)	1.00(0.96)
 f_3	5.22(1.05)	4.23(0.82)	3.55(0.69)	3.18(0.64)
 f_4	0.53(0.10)	0.28(0.09)	0.27(0.09)	0.28(0.09)
 f_5	0.56(0.09)	0.27(0.06)	0.25(0.06)	0.27(0.07)
 f_6	1.60(0.94)	1.98(2.03)	2.45(2.70)	3.16(3.54)

Table 4.3: Mean RMSED(1) (standard deviation) comparing number of knots used.

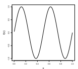
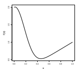
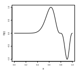
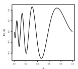
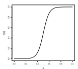
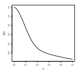
Function	$K = 10$	$K = 20$	$K = 30$	$K = 40$
 f_1	58.85(4.79)	22.42(6.72)	29.38(7.68)	41.05(11.78)
 f_2	17.86(7.33)	27.90(29.05)	38.35(54.24)	54.13(99.65)
 f_3	177.7(47.19)	148.9(40.43)	168.2(105.6)	193.1(163.7)
 f_4	20.35(20.62)	22.92(21.37)	25.71(25.11)	29.87(36.21)
 f_5	9.87(1.22)	5.42(0.10)	6.30(1.49)	7.34(2.70)
 f_6	36.60(38.26)	65.85(114.6)	104.2(198.2)	172.5(364.4)

Table 4.4: Mean RMSED(2) (standard deviation) comparing number of knots used.

Here $K = 10$ results in the highest mean RMSED(2) for f_1'' and f_5'' but the lowest in f_2'', f_4'' and f_6''. Using $K = 20$, which was not optimal for any of the first derivatives, leads to preferred second derivative estimates of f_1, f_3 and f_5. Similarly to the penalty order simulation, there is no consensus for an optimum number of knots. Therefore no reason to alter the recommendations of Eilers & Marx [12] is found.

4.1.1.3 Basis Degree

Local modelling has been introduced because local low order polynomials are as effective as global high order polynomials when joined together to give a smooth fit. Here a test of the performance in derivative estimation when changing the degree (q) of the B-Spline basis underlying the P-Spline model was performed. The RMSED(1) and RMSED(2) were measured for $q = 3, 5, 7, 9$. Eilers & Marx [12] recommend taking $q = 3$ for any smoothing situation.

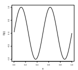
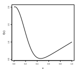
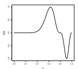
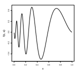
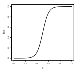
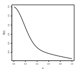
Function	$q = 3$	$q = 5$	$q = 7$	$q = 9$
 f_1	5.15(0.50)	5.65(0.52)	6.13(0.54)	6.57(0.55)
 f_2	0.67(0.22)	0.60(0.19)	0.55(0.17)	0.52(0.17)
 f_3	5.22(1.05)	5.34(1.08)	5.44(1.09)	5.54(1.10)
 f_4	0.53(0.10)	0.57(0.10)	0.60(0.11)	0.63(0.11)
 f_5	0.56(0.09)	0.61(0.09)	0.66(0.09)	0.70(0.09)
 f_6	1.60(0.94)	1.40(0.60)	1.35(0.55)	1.30(0.47)

Table 4.5: Mean RMSED(1) (standard deviation) comparing basis degree.

Table 4.5 summarises the performance in estimation of the first derivatives of $f_1, \ldots, f_6$. The recommended cubic B-Spline basis offers the best performance for first derivative estimates of f_1, f_3, f_4 and f_5. The functions f_2 and f_6 prove to have different needs and a high degree basis $q = 9$ is to be preferred here. There is no overwhelming evidence here to dispute the recommended degree.

Table 4.6 agrees wholeheartedly with Table 4.5 in terms of the values of q which achieve the premium performance for each function. However, as discussed in Section 2.2.3, it is important to take the degree to be at least $l + 2$ where $l = 0, 1, 2, \ldots$ is the highest derivative to be estimated. Thus $q = 3$ is not sensible when estimating second derivatives and should be discounted here. At the same time, changing the value of q seems to have the lowest influence among the components of the P-Spline model, and a value $q = 3$ seems to be quite reasonable unless derivatives of order $l \geq 2$ are required. In this situation it is recommended to take

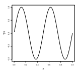
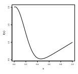
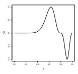
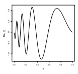
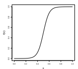
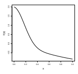
Function	$q = 3$	$q = 5$	$q = 7$	$q = 9$
 f_1	58.85(4.79)	64.82(4.63)	70.72(4.56)	76.18(4.60)
 f_2	17.86(7.33)	15.29(4.82)	13.59(2.82)	12.96(2.37)
 f_3	177.7(47.19)	180.9(42.14)	185.2(39.83)	187.2(38.61)
 f_4	20.35(20.62)	20.24(20.78)	20.40(20.68)	20.50(20.65)
 f_5	9.87(1.22)	10.57(1.29)	11.17(1.36)	11.66(1.42)
 f_6	36.60(38.26)	27.35(20.65)	24.28(14.58)	22.48(10.87)

Table 4.6: Mean RMSED(2) (standard deviation) comparing basis degree.

$q = l + 2$.

4.1.2 Discussion of the Effect of P-Spline Components

Figure 4.1 presents a matrix plot of estimates of f_1'.

There is little variation in estimates across penalty orders or basis degrees. Changing the number of knots seems to result in slight variation of estimates but no clear optimum level is evident. Simulations have been carried out to investigate whether any change to the recommended values of penalty order $d = 2$, number of knots $K = min(40, n/5)$, or basis degree $q = 3$ was necessary when dealing with derivative estimation. No decisive empirical evidence has been uncovered to stray from these recommendations when estimation of f' or f'' is of primary concern.

4.2 The Effect of Smoothing Parameter Selection Method

There are many automatic methods for smoothing parameter selection. Among these are the CV, GCV, AIC and BIC selection criteria described in Section 3.1.2. Kohn et al. [26] found evidence that CV and GCV are comparable when estimating f and f'. Indeed, the advantage of GCV is known to be not one of performance, but of computational efficiency. Shibata [54] and Hurvich et al. [23] find evidence that AIC can lead to undersmoothing of a (regression)

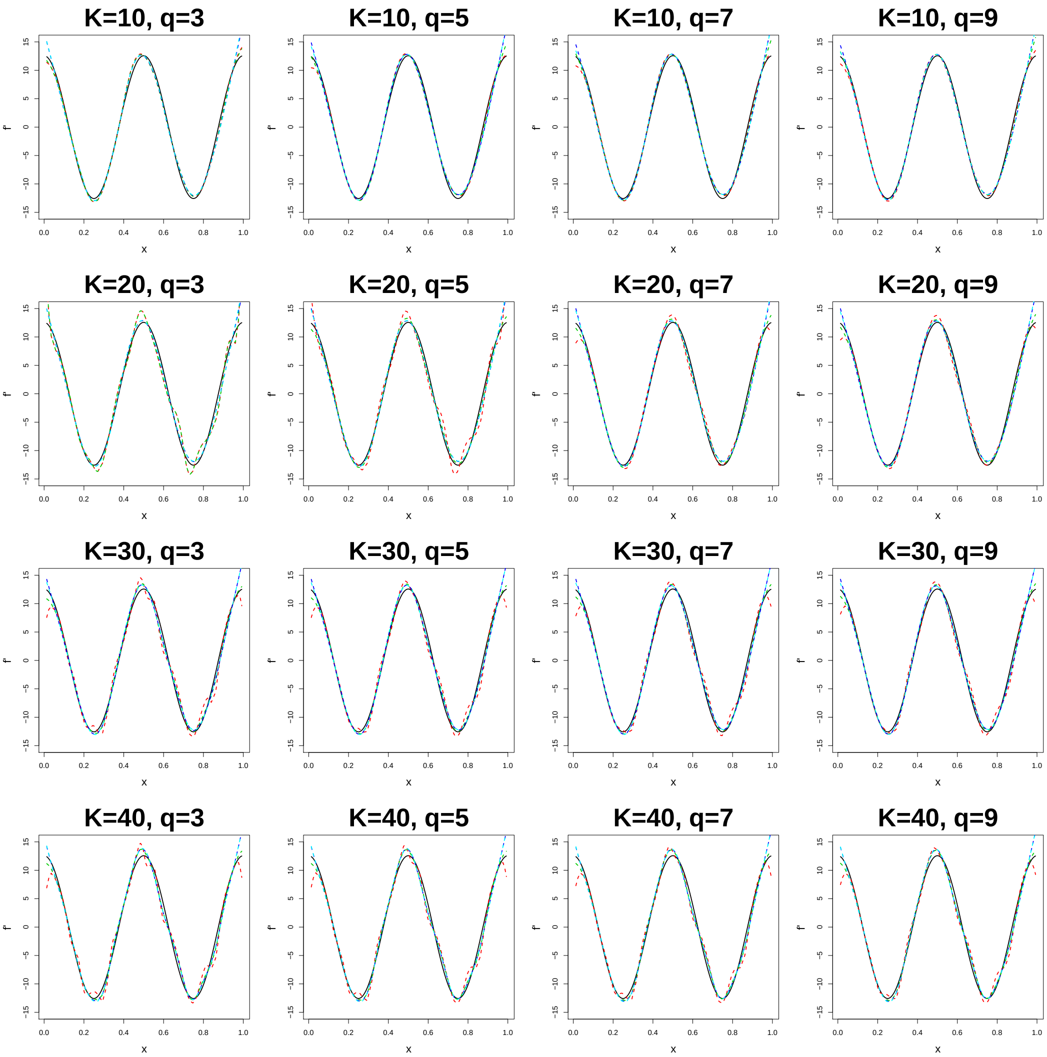


Figure 4.1: Matrix plot changing (left to right increasing q, top to bottom increasing K) the components of the P-Spline model with penalty order d =1, d =2, d =3 and d =4.

function.

4.2.1 Simulation Study into the Effect of Smoothing Parameter Selection Method

Here these selection methods are compared in terms of first and second derivative estimation performance using simulated data from the familiar six functions $f_1, \ldots, f_6$.

The four methods, namely CV, GCV, AIC and BIC, for selecting the smoothing parameter (λ) of a P-Spline fit were compared across the functions $f_1, \ldots, f_6$ using 1000 replicates of $n = 50$ responses from $y = f_i(x) + \epsilon$, $i = 1, \ldots, 6$, with $x \sim U[0, 1]$, $\epsilon \sim N(0, \sigma^2)$ and $\sigma = \frac{1}{6} range(f_i)$.

Results from the simulations into the smoothing parameter selection criteria are summarised in Tables 4.7 and 4.8. Consistently similar results are obtained when using CV or GCV, which is not surprising. AIC and BIC are outperformed in all but derivative estimation of data simulated from f_6.

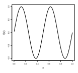
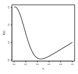
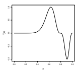
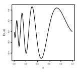
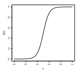
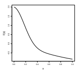
Function	CV	GCV	AIC	BIC
 f_1	5.15(0.50)	5.15(0.50)	5.15(0.50)	6.04(1.77)
 f_2	0.67(0.22)	0.68(0.22)	1.07(0.22)	1.07(0.22)
 f_3	5.22(1.05)	5.21(1.07)	6.04(1.05)	6.13(0.95)
 f_4	0.53(0.10)	0.53(0.10)	1.80(0.19)	1.80(0.19)
 f_5	0.56(0.09)	0.56(0.09)	1.47(0.14)	1.47(0.14)
 f_6	1.60(0.94)	1.63(1.00)	1.17(0.25)	1.17(0.25)

Table 4.7: Mean RMSED(1) (standard deviation) comparing smoothing parameter selection methods.

Again, when estimating second derivatives, the error resulting from CV and GCV λ selections is quite similar, with the CV selected λ offering the lower overall error in some functions. When estimating f_6'', both CV and GCV selected λ's lead to mean RMSED(2) double that obtained when using an AIC or BIC λ. Moreover, the variability of this error is higher than the mean! Given the similarities between f_2 and f_6 it is strange that such a difference in overall error is only evident in f_6. The CV/GCV selected λs lead to estimates that are evidently struggling to accurately describe the right hand tail of f_6 compared to the AIC/BIC selections.

4.2.1.1 Fudging λ Revisited

Altering the choice of smoothing parameter (λ) in an attempt to improve performance in derivative estimation has been discussed for the smoothing splines model in Section 3.2.1.3. No significant improvements in derivative estimation were found in this study although only a subjective (and mysterious) constant fudge value was used. Here an investigation into the merit of this idea is performed when a curve is fitted using P-Splines. Using simulated data one

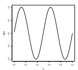
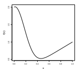
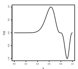
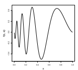
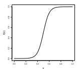
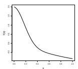
Function	CV	GCV	AIC	BIC
 f_1	58.85(4.79)	58.85(4.79)	58.85(4.79)	69.37(19.19)
 f_2	17.86(7.33)	18.69(9.46)	18.43(3.20)	18.43(3.20)
 f_3	177.7(47.19)	178.9(48.92)	181.8(30.14)	181.8(30.14)
 f_4	20.35(20.62)	20.47(20.73)	22.01(20.36)	22.01(20.36)
 f_5	9.87(1.22)	9.87(1.22)	15.45(2.02)	15.45(2.02)
 f_6	36.60(38.26)	38.96(43.48)	18.84(3.15)	18.84(3.15)

Table 4.8: Mean RMSED(2) (standard deviation) comparing smoothing parameter selection methods.

can fudge λ to the optimum value for derivative estimates and investigate whether there is a common fudge which may help estimates in general. Recall the reasoning behind this alteration is that λ is chosen based on optimising $\hat{f}$ and not the derivative, which may lead to considerable undersmoothing.

Using the three functions f_1, f_2 and f_3 the CV choice of λ was fudged until the minimum RMSED(1) and RMSED(2) were obtained. This was performed by searching over a ± 1 grid around the CV choice for λ until the values of λ which gave the lowest RMSED(1) and RMSED(2) were found. Data were simulated from each of f_1, f_2 and f_3, with x uniform on $[0, 1]$ and Gaussian error added with zero mean and standard deviation equal to one sixth the range of f_i, $i = 1, 2, 3$. A thousand replicates of $n = 50$ responses were simulated. The hope was that a common scalar transform of the CV choice could be found and used to improve derivative estimation.

Table 4.9 provides a summary of the effect on derivative estimates of fudging the smoothing parameter. The multiplier of λ is displayed since it better represents the change to λ. The simulations found that increasing the smoothing parameter improved the RMSED but when this was applied in 1000 replicates it resulted in a mean deterioration for estimating f_2', f_2'' and f_3''. There is no pattern to either the multipliers or the improvements in Table 4.9. In real life applications it is not possible to find the λ which gives the lowest RMSED$(\cdot)$ and therefore any

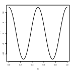
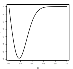
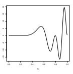
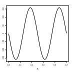
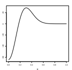
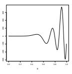
Function	Multiplier	% Improvement in RMSED
 f_1'	10.0	5.25
 f_2'	1.05	-1.47
 f_3'	18.0	0.36
 f_1''	2.30	6.62
 f_2''	1.05	-2.15
 f_3''	1.33	-0.02

Table 4.9: Results of using λ ‘fudge’ for derivative estimation of f_1, f_2 and f_3.

hope of using a ‘fudge’ relies on a common trait which is not evident here.

4.2.1.2 Discussion of Smoothing Parameter Selection Methods

Apart from the findings related to f_6 there is empirical evidence that using either a CV or GCV selected smoothing parameter works best when estimating derivatives using P-Splines in the context studied here. Moreover, fudging λ to improve derivative estimation does not lead to an ideal solution as no common multiplier for the CV selected λ was found in the scenarios considered here.

More smoothing is seemingly required in order to remove the undersmoothing and boundary effect issues present in derivative estimation. Unfortunately using a single constant (as opposed to spatially varying) smoothing parameter appears to be yet another dead end!

4.3 The Effect of Sample Size on Derivative Estimation

The motivating illustrations feature datasets of varying sample sizes, namely, 10, 31, 64 and 131. It would be expected that as sample size increases the resulting standard error decreases and the accuracy of estimates increases. The sample size may be controlled at the design phase and as such these simulations may be of interest in future studies (although the idea of power calculations for nonparametric smoothing is not well researched).

4.3.1 Simulation Study into the Effect of Sample Size on Derivative Estimation

The following simulation study was performed to investigate the effect of sample size on the performance of P-Spline derivative estimates. Three sample sizes of $n = 20, 50$ and 100 were used.

Data were simulated 1000 times from the six functions $f_1, \ldots, f_6$ under the situation $y = f_i(x) + \epsilon$, $i = 1, \ldots, 6$, with x uniform on $[0, 1]$. The error vectors were simulated from a Normal distribution with standard deviation σ equal to one sixth the range of f_i, $i = 1, \ldots, 6$. The P-Spline method was used to find first and second derivative estimates at sample sizes of $n = 20, 50$ and 100. Comparisons across the three sample sizes were made using the RMSED(1) and RMSED(2).

Figures 4.2, 4.3 and 4.4 display comparisons of the RMSED(1) for each of the six functions $f_1, \ldots, f_6$. The boxplots suggest that increasing sample size leads to better performance. Since there are more data pairs, there is more information for the P-Spline model, which leads to a better estimate of the underlying function, which, in turn, leads to improved derivative estimates. The variability displayed in the boxplots, however, does not appear to either increase or decrease consistently with increasing sample size. The range of RMSED(1) at $n = 50$ in the f_2 and f_6 scenarios is lower than at either of the other two sample sizes tested here.

Table 4.10 confirms the decrease in mean RMSED(1) as n increases, they also further allude to a strange relationship between the variability of RMSED(1) and sample size. The standard deviation of RMSED(1) at $n = 50$ is either the highest (f_1, f_3, f_4, f_5) or the lowest (f_2, f_6) among the three sample sizes chosen.

Table 4.11 reveals that the variability in RMSED(1) for f_2 and f_6 has caused the mean RMSED(2) to be lowest at $n = 50$. The difference in variability of RMSED(2) for both f_2 and f_6 is remarkably high, with standard deviations of 6 for $n = 50$, 24 for $n = 100$ and 100 for $n = 20$ for f_2. One would expect that as n increases, the variability in the error would decrease with each additional data pair but evidence here suggests otherwise.

Unusual relationships between n and the variance of RMSED($\cdot$) are evident for functions f_2 and f_6. These functions are alike in that they have a common e^{-16x^2} term, and also appear

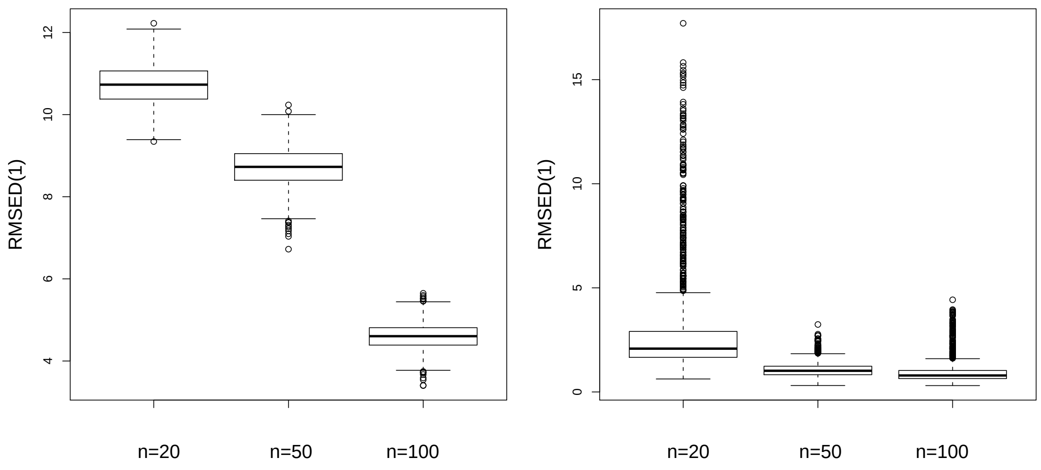


Figure 4.2: Boxplots of RMSED(1) for f_1(left plot) and f_2 (right plot) for comparisons of sample size.

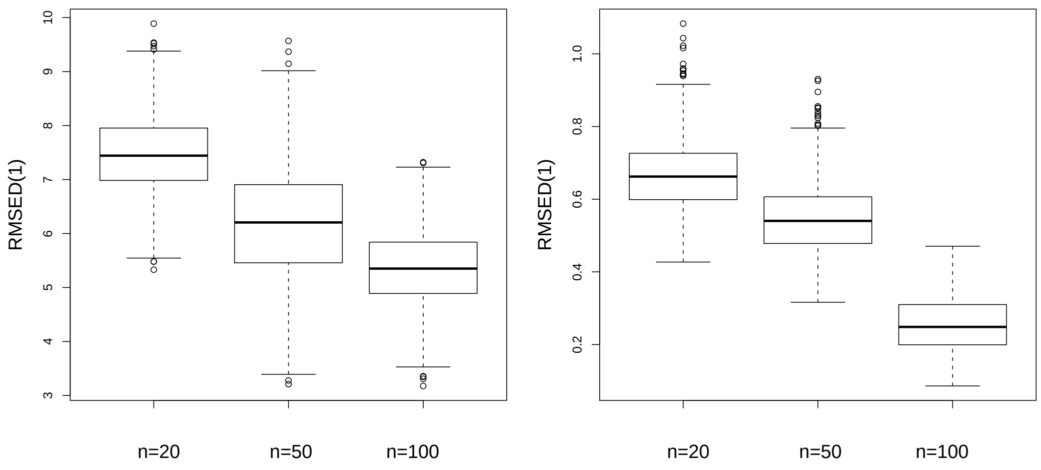


Figure 4.3: Boxplots of RMSED(1) for f_3(left plot) and f_4 (right plot) for comparisons of sample size.

quite similar from Figures 1.9 and 1.13. Figure 4.5 exhibits simulated fits from each of the three sample sizes using the same seed in R. Surprisingly, the P-Spline fit when $n = 100$ is the poorest of the three. It seems to miss the trough which is the minimum of f_2 and lose track of f_2 at the second tail.

Figure 4.6 displays derivative estimates at each of the three sample sizes, along with the actual first and second derivatives of f_2. None of the curves offer impressive derivative estimates

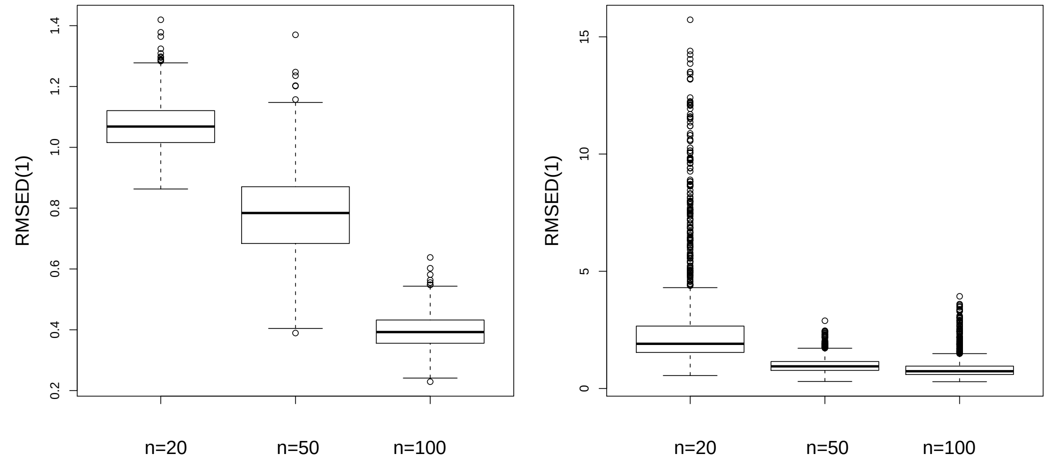


Figure 4.4: Boxplots of RMSED(1) for f_5(left plot) and f_6 (right plot) for comparisons of sample size.

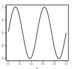
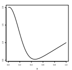
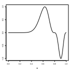
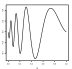
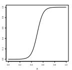
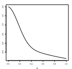
Function	$n = 20$	$n = 50$	$n = 100$
 f_1	10.72(0.50)	8.71(0.50)	4.60(0.33)
 f_2	3.22(2.91)	1.08(0.38)	1.01(0.66)
 f_3	7.46(0.73)	6.19(1.06)	5.35(0.70)
 f_4	0.66(0.10)	0.55(0.10)	0.26(0.07)
 f_5	1.07(0.08)	0.78(0.14)	0.39(0.06)
 f_6	2.93(2.62)	1.00(0.35)	0.92(0.59)

Table 4.10: Mean RMSED(1) (standard deviation) for comparisons of sample size.

but the estimates at $n = 50$ certainly seem to be the most smooth. There are boundary effect issues for each sample size.

4.4 The Effect of Variability on Derivative Estimation

The variability of the data is different in each of the motivating illustrations (Figure 4.7). In the Blood Lactate and Astronomical data there is very little variability such that fitting a curve

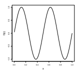
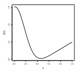
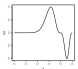
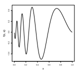
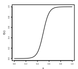
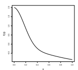
Function	$n = 20$	$n = 50$	$n = 100$
 f_1	155.20(5.51)	106.2(5.25)	54.00(3.73)
 f_2	53.62(99.63)	11.58(5.90)	16.69(23.86)
 f_3	200.0(44.48)	185.9(35.56)	188.6(35.64)
 f_4	32.71(39.51)	20.13(33.23)	19.99(27.83)
 f_5	12.66(1.07)	9.98(1.91)	7.12(0.90)
 f_6	48.77(89.92)	10.85(5.30)	15.40(21.49)

Table 4.11: Mean RMSED(2) (standard deviation) for comparisons of sample size.

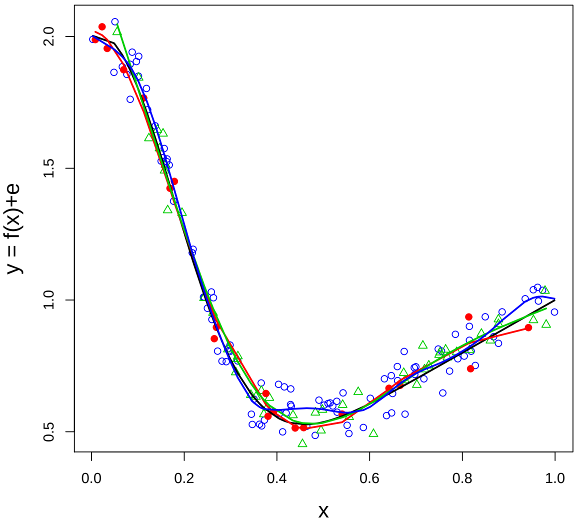


Figure 4.5: Fitted curves obtained by P-Splines for data simulated from f_2 under the same seed using $n = 20$ (red), $n = 50$ (green) and $n = 100$ (blue), with actual f_2 (black).

to the data by eye alone should be relatively accurate, whereas the Winter Nutrients study displays far more variability.

Certainly, in all real examples of noisy data, variability varies! This variation will effect derivative estimates, and one would expect that as variability increases so performance of derivative estimators should decrease.

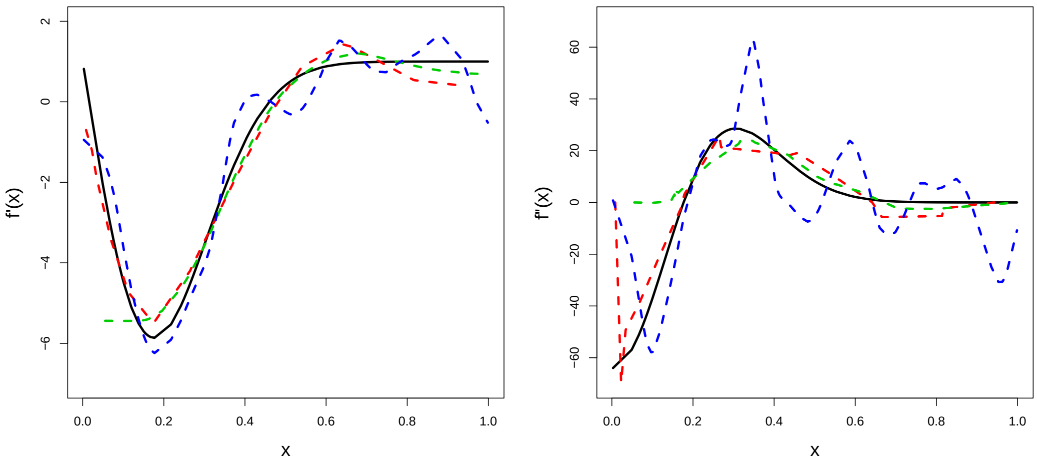


Figure 4.6: First (left) and second (right) derivative estimates using P-Spline for $n = 20$ (red), $n = 50$ (green) and $n = 100$ (blue), with actual f_2' and f_2'' (black).

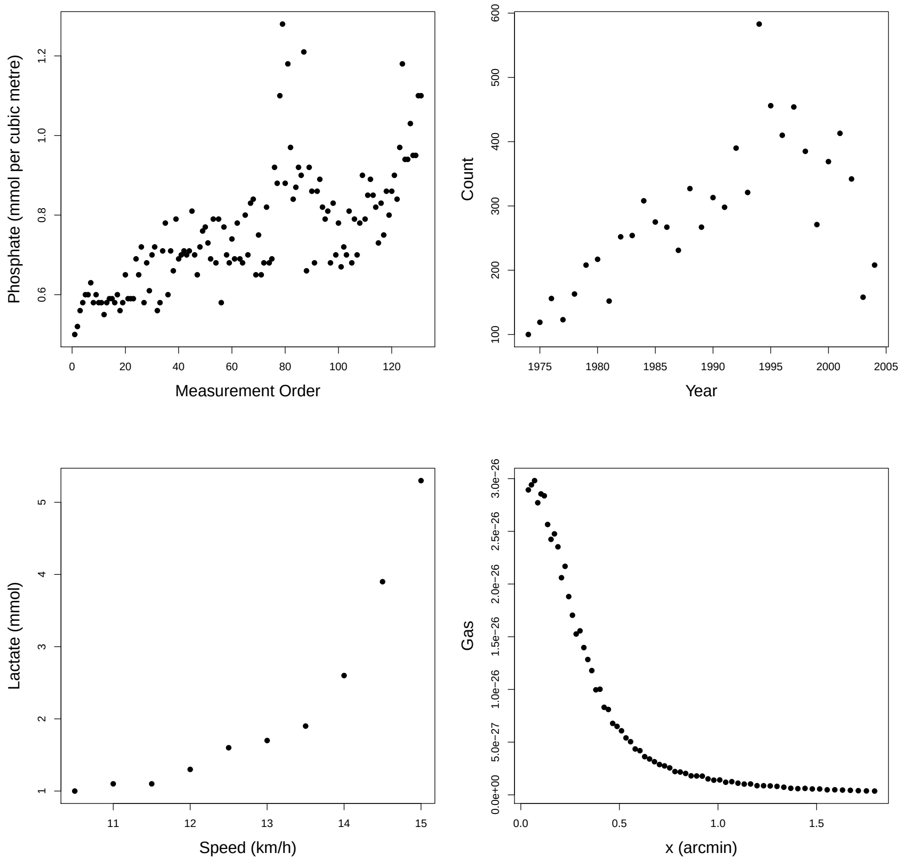


Figure 4.7: Differing levels of variability among motivating illustrations.

4.4.1 Simulation Study into the Effect of Variability on Derivative Estimation

A simulation study into the effects of changing the standard deviation of the error added to the simulated nonlinear functions follows.

Data were again simulated from the six functions f_1 to f_6 with x uniform on $[0,1]$ and $n = 50$. A thousand responses $y = f_i(x) + \epsilon$, with $i = 1,\ldots,6$ and $\epsilon \sim N(0,\sigma^2)$, at each of three values of σ were modelled using P-Splines and derivative estimates obtained using (3.26). The values of σ were selected to be $\frac{1}{3}range(f_i)$, $\frac{1}{6}range(f_i)$ and $\frac{1}{10}range(f_i)$ for $i = 1,\ldots,6$.

The boxplots in Figures 4.8, 4.9 and 4.10 display the RMSED(1) for this simulation study. As one would expect, increasing the variability decreases the performance of the P-Spline first derivative estimates. Similar to the sample size simulations, f_2 and f_6 stand apart from the others. Here there exists much more of a gulf in performance between high and moderate values of σ for f_2 and f_6.

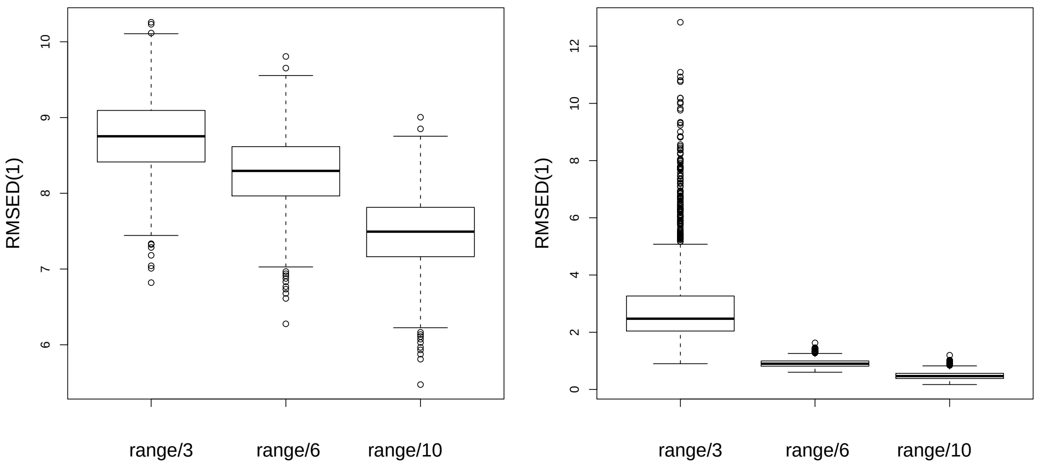


Figure 4.8: Boxplots of RMSED(1) for f_1(left plot) and f_2 (right plot) for comparisons of variability.

The performance in first and second derivative estimation of P-Splines with differing values of σ are summarised in Tables 4.12 and 4.13 respectively. The estimates display decreasing mean and variance of RMSED($\cdot$) as σ decreases. The variability in RMSED(1) is similar for functions f_1, f_3, f_4 and f_5 whereas there is quite a large drop with decreasing σ for f_2 and f_6.

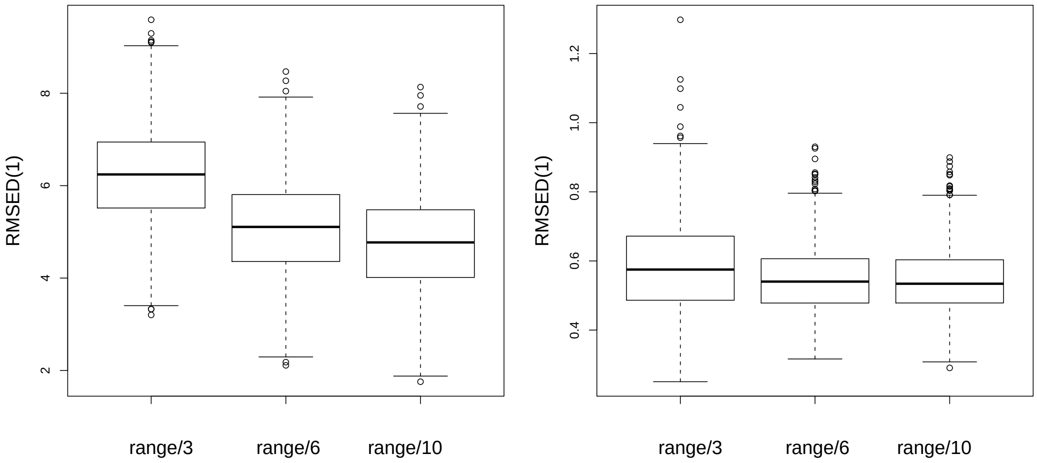


Figure 4.9: Boxplots of RMSED(1) for f_3(left plot) and f_4 (right plot) for comparisons of variability.

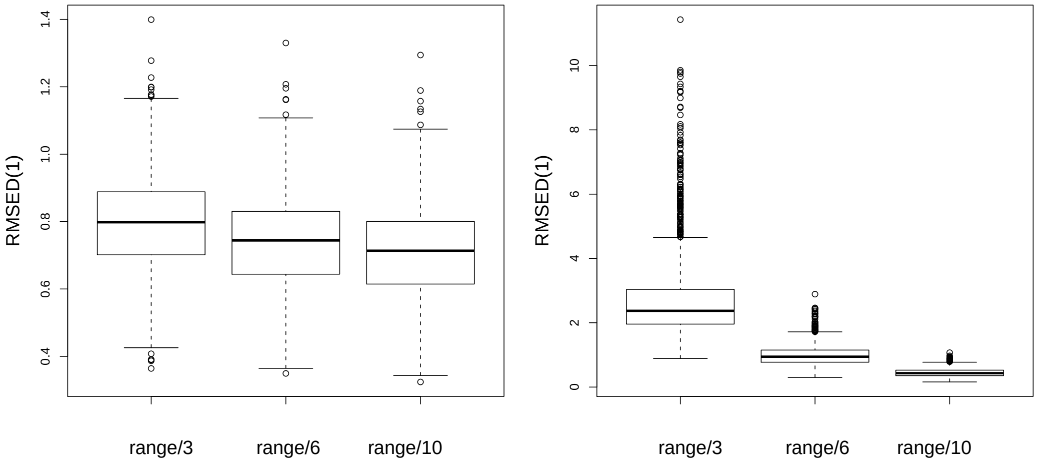


Figure 4.10: Boxplots of RMSED(1) for f_5(left plot) and f_6 (right plot) for comparisons of variability.

Similarly, the variability of RMSED(2) experiences a large decline only in f_2 and f_6. Odd behaviour of RMSED(1) and RMSED(2) has been observed for these two (similarly shaped) functions when altering sample size, and here they display a more dramatic decrease in error when decreasing variability.

As expected the performance of derivative estimation is inversely related to the variance of the observed data. With a low variance, more accurate fits of the underlying regression function

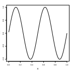
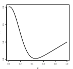
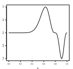
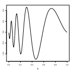
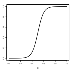
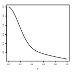
Function	$\sigma = \frac{1}{3}$range	$\sigma = \frac{1}{6}$range	$\sigma = \frac{1}{10}$range
 f_1	8.74(0.52)	8.28(0.50)	7.48(0.50)
 f_2	3.09(1.77)	1.51(0.38)	0.48(0.14)
 f_3	6.22(1.06)	5.09(1.06)	4.75(1.06)
 f_4	0.58(0.14)	0.55(0.10)	0.54(0.10)
 f_5	0.80(0.15)	0.74(0.14)	0.71(0.14)
 f_6	2.89(1.6)	1.00(0.35)	0.45(0.13)

Table 4.12: Mean RMSED(1) (standard deviation) for comparisons of variability.

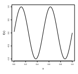
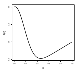
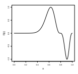
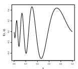
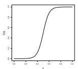
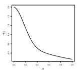
Function	$\sigma = \frac{1}{3}\times$range	$\sigma = \frac{1}{6}\times$range	$\sigma = \frac{1}{10}\times$range
 f_1	106.4(5.26)	103.2(5.25)	101.2(5.24)
 f_2	29.92(27.81)	11.58(5.90)	5.95(2.10)
 f_3	189.0(36.45)	180.9(35.56)	177.7(35.52)
 f_4	24.07(33.35)	20.13(33.23)	19.04(33.33)
 f_5	10.00(1.90)	9.00(1.91)	8.47(1.91)
 f_6	28.00(24.98)	10.85(5.30)	5.59(1.90)

Table 4.13: Mean RMSED(2) (standard deviation) for comparisons of variability.

are achieved, and thus better derivative estimates are obtained. In applications, one may be more confident about derivative estimates for data where a clear pattern is present rather than a highly variable nonlinear relationship.

4.5 Chapter Summary

P-Splines offer a flexible smoothing method in situations of nonlinear data and are easily applied to noncontinuous responses. However, in derivative estimation there is evidence of problems in undersmoothing and boundary effects. Investigations into P-Spline derivative estimates when varying modelling situations (sample size, variability) and the choices underlying the model (λ, d, K, q) have been carried out.

No evidence was found to alter the recommended values of penalty order, number of knots or basis degree given by Eilers & Marx [12]. Simulations have also given evidence that choosing λ by CV or GCV is to be preferred over AIC and BIC criteria and changing the value of λ using a constant 'fudge' offers little to no overall improvement in derivative estimation performance. However, despite finding no fault with these selection criteria it is still essential to check any fit visually before relying on automatic choices of λ.

Increasing sample size leads to better performance in derivative estimation, although this is not strictly the case for the variability in the precision of the estimator. Finally, as expected, the simulations suggest that the larger the variability of the observed data, the harder the problem of finding accurate derivative estimates becomes.

P-Spline smoothing has many advantages and has performed well in comparison with other derivative estimation techniques. Evidence has been found that the smoothing components of the P-Spline model (d, K and q) have little effect on derivative estimation. It seems that the model is most sensitive to change in the penalty term and corresponding smoothing parameter λ, although the method for selecting λ seems to have little impact. In the next Chapter the advantages of using P-Splines for derivative estimation are exploited with possible improvements in derivative estimation attempted using alternate penalisation.

Chapter 5

Derivative Estimation Using An Additive Penalty

The main issues in derivative estimation encountered in previous chapters relate to undersmoothing and boundary effects. To better estimate derivatives of the underlying function f additional smoothing is needed. Using penalised splines with a constant 'fudge' added to the smoothing parameter was found to offer no improvement (Section 3.2.1.3 and Section 4.2.1.1). More penalisation is needed, but perhaps in a different place!

5.1 An Additive Penalty Approach

Aldrin [2] and Belitz & Lang [4] introduced methods to include an additive penalty structure to a P-Spline model for increased sensitivity in smoothing. A secondary smoothing term has also been used for constrained smoothing by Bollearts et al. [5] when monotonicity of response is a scientific requirement. An example of an additive penalty P-Spline model with two penalty terms may be written

$$\sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda_1 \sum_{j=d_1+1}^{m} (\Delta^{(d_1)}\alpha_j)^2 + \lambda_2 \sum_{j=d_2+1}^{m} (\Delta^{(d_2)}\alpha_j)^2 \tag{5.1}$$

for $i = 1, \ldots, n$ observations, $j = 1, \ldots, m$ coefficients and difference penalty Δ with orders d_1 and d_2. The first two terms of (5.1) with $d_1 = 2$ is the standard P-Spline model. The additional $\lambda_2 \sum_{j=d_2+1}^{m} (\Delta^{(d_2)} \alpha_j)^2$ term allows for extra smoothing to ensure another coefficient is restricted to being similar to a neighbour.

The approach of using this additional penalty is extended to handle derivative estimation in Simpkin et al.[55]. The motivation being that extra penalty terms should allow for increased flexibility which is often required for derivative estimation. Using the additional smoothing term, more smoothing can be focused in areas which display undersmoothing, while still making sure other areas of the data are accurately described. The choice of where to place this extra smoothing adds yet more subjective choices to the (already choice laden) task of smoothing. Over the next few sections these choices shall be considered, with optimal selections determined.

5.2 Derivative Estimation with an Additive Penalty

The additive penalty model introduced in this thesis is a variant on the P-Spline model described in Chapter 4. It shares many similarities in terms of inference and estimation of coefficients but has derivative estimation as its primary aim.

5.2.1 Modelling a Continuous Response Variable

Consider the situation where an estimate of the underlying function f which describes the behaviour of a continuous response y is required, i.e.

$$y = f(x) + \epsilon$$

where ϵ is taken to be some *iid* Normal error with mean zero and constant variance σ^2. Using a B-Spline basis for f this can be rewritten

$$y = B\alpha + \epsilon$$

where B is the B-spline design matrix and α is the vector of coefficients. Minimising

$$\|y - B\alpha\|^2$$

will give the least squares estimate of α, which attempts to interpolate the data up to the number of knots chosen. Two additive roughness penalty terms are used to smooth out this attempted interpolation, i.e. minimise

$$\|y - B\alpha\|^2 + \lambda_1 \|D_{d_1}\alpha\|^2 + \lambda_2 \|D_{d_2}\alpha\|^2$$

and expand to get

$$y^T y - 2\alpha^T B^T y + \alpha^T (B^T B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})\alpha.$$

Taking the derivative with respect to α yields

$$-2B^T y + 2(B^T B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})\alpha.$$

Setting the derivative to zero and rearranging gives the equation for the fitted coefficients $\hat{\alpha}$ as follows

$$\hat{\alpha} = (B^T B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1} B^T y. \tag{5.2}$$

Derivatives are again obtained using (3.26). Since using the extra penalty term changes only the value of the α_j, the formula works as in the ordinary P-Spline model.

5.2.2 Modelling a Count Response Variable

In the case where the response variable y is in the form of counts, estimation of coefficients using the additive penalty is very similar to estimation using P-Splines. For the purposes of simplicity, only the case where the assumption of Poisson distributed counts has been made shall be considered here.

The response y generated by a Poisson distribution with mean rate μ and is modelled such that

$$ln(\mu) = B\alpha \tag{5.3}$$

where

$$\mu = e^{B\alpha}, \tag{5.4}$$

for B-Spline design matrix B and coefficients α.

where $j = 1, \dots, m$ is the number of coefficients to be fitted. The α_j are found using iteratively weighted least squares, i.e.

$$\hat{\alpha}_{t+1} = (B^T \hat{W}_t B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1} B^T \hat{W}_t \hat{z}_t \tag{5.5}$$

where

$$z = B\alpha + W^{-1}(y - \mu)$$

and

$$\hat{W} = diag(\hat{\mu})$$

Once the coefficients α_j have been estimated, derivatives are estimated using the de Boor formula (3.30) in Section 3.2.2.2.

5.3 Choosing Smoothing Parameters

An interesting problem motivated by the additive penalty model is the selection of multiple smoothing parameters λ_1 and λ_2. Similarly to P-Splines, a subjective choice is difficult to make since the derivative, and not the underlying function, is of primary concern. An automatic method for selection of the two smoothing parameters needs to be specified. In the following sections a cross validation method is suggested since, as has been seen in Chapter 4, choosing a smoothing parameter by GCV or CV give similar results with CV slightly better.

The first decision is whether λ_1 and λ_2 should be selected simultaneously or sequentially. A sequential approach is beneficial in terms of computational efficiency. A two dimensional grid search is far more exhaustive (n^2 calculations) than two one dimensional, sequential, searches ($2n$ calculations). Aldrin [2] found that a sequential approach is to be preferred when estimating the underlying function f, whether this holds for derivative estimation is investigated in the simulations which now follow.

5.3.1 Comparison of Sequential and Simultaneous Selection Methods

Sequential and simultaneous selection of (λ_1, λ_2) was assessed with a small simulation study, with RMSED(1) and RMSED(2) used to quantify performance in first and second derivative estimation respectively. Penalties on first and second order differences of coefficients were used i.e.

$$\sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda_1 \sum_{j=2}^{m}(\Delta^{(1)}\alpha_j)^2 + \lambda_2 \sum_{j=3}^{m}(\Delta^{(2)}\alpha_j)^2. \tag{5.6}$$

For ease of interpretation methods using an additional penalty term will be denoted as 'AP' and the order of the penalties used will be appended to 'AP'. For example, a model penalising first and second order differences of coefficients simultaneously shall be called APsim12, whereas the method penalising first and second order differences of coefficients sequentially shall be called APseq12 etc.

The methods were compared across the six functions $f_1, \ldots, f_6$ introduced in Section 1.5. A thousand samples of size $n = 50$ were simulated from each function with x uniformly distributed on $[0, 1]$ and Gaussian error $\epsilon \sim N(0, \sigma^2)$ at σ equal to one sixth the range of f added.

Figures 5.1 to 5.6 display boxplots of the RMSED(1) and RMSED(2) for both selection methods across the six functions. Tables 5.1 and 5.2 give the mean and standard deviation of the RMSED(1) and RMSED(2) for each function and method across 1000 replications. Figures 5.1 to 5.6 and Tables 5.1 and 5.2 suggest that the sequential approach outperforms the simultaneous selection method in terms of RMSED(1) in f_1, f_2, f_4 and f_6, with typically lower variability in its error for all but f_4. For second derivative estimates, the sequential selection method displays lower RMSED(2) for all functions, and also has lower variability in all but f_5.

A simultaneous method for selection of λ_1 and λ_2 will lead to the optimum amount of smoothing needed to best fit the data. It will therefore offer the same amount of smoothing as the standard P-Spline method, granted this smoothing is spread over the two penalty terms. For this reason, as has already been demonstrated in Section 3.2.5, the simultaneous selection approach will lead to undersmoothing in derivative estimates. The sequential method, on the other hand, will first find the optimum amount of smoothing to best fit the data, and then use

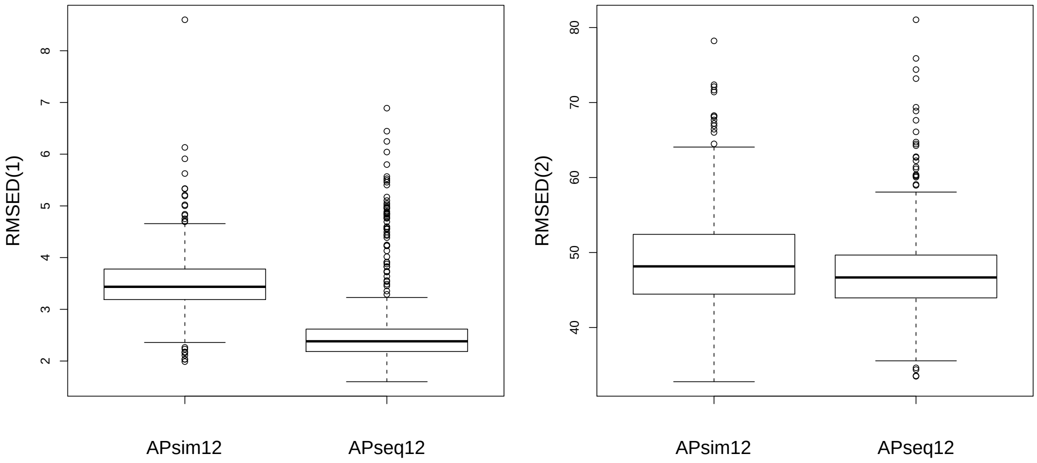


Figure 5.1: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_1 comparing smoothing parameter selection methods.

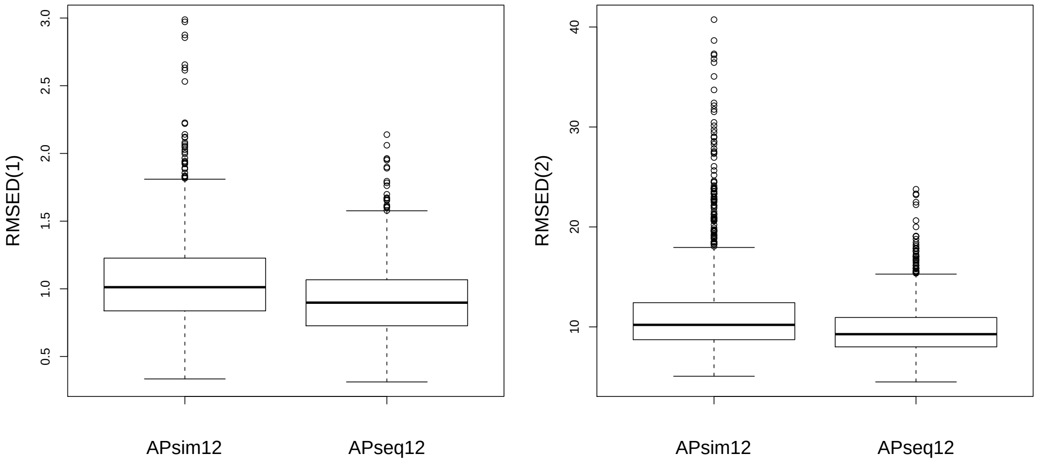


Figure 5.2: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_2 comparing smoothing parameter selection methods.

the extra penalty term to improve derivative estimates by oversmoothing in the fit to the data. It is also more computationally efficient to estimate the (λ_1, λ_2) sequentially.

As an illustrative example, Figure 5.7 displays data simulated from $f_2 = x + 2e^{-16x^2}$ with smooths using APsim12 and APseq12. Both methods are comparable when smoothing the data, with APseq12 there is evidence of oversmoothing at $x \approx 0.5$. For estimates of the derivative, this oversmoothing pays off. In Figure 5.8 undersmoothing is evident when using APsim12 in

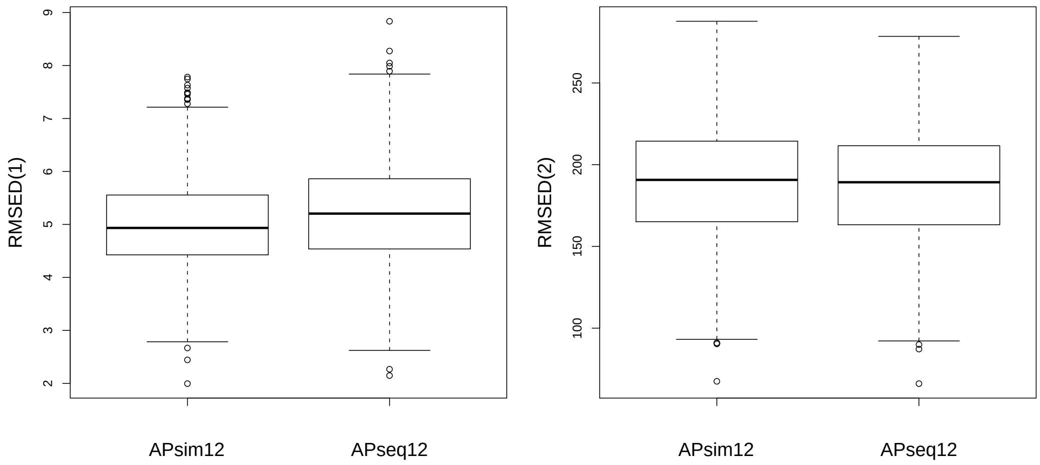


Figure 5.3: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_3 comparing smoothing parameter selection methods.

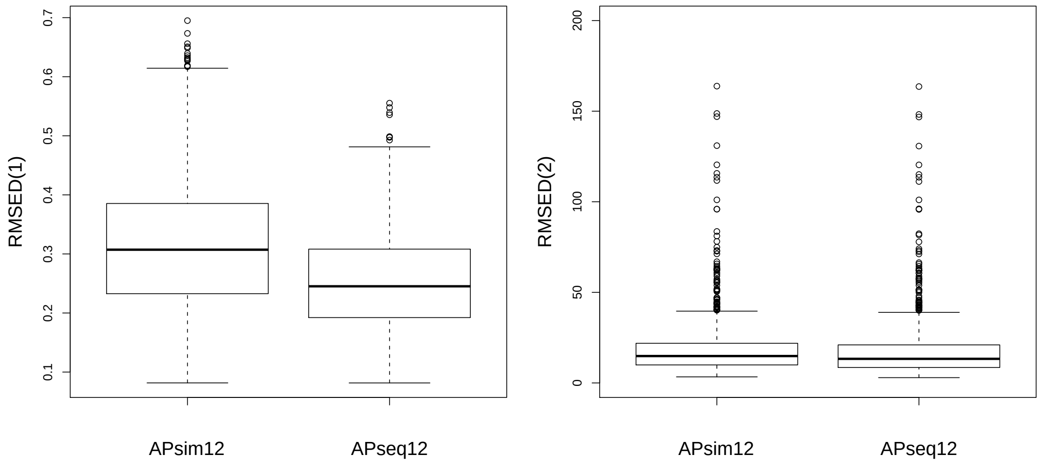


Figure 5.4: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_4 comparing smoothing parameter selection methods.

estimates of both the first and second derivative. APseq12 does a better job of capturing the derivatives of the underlying function explaining the observed data. Both techniques display boundary effects for each derivative at the left boundaries.

The results of this, albeit small, simulation study give enough evidence to eliminate the simultaneous approach to smoothing parameter selection for the additive penalty method when derivatives are of primary interest to an analysis. The simulations for the sequential method

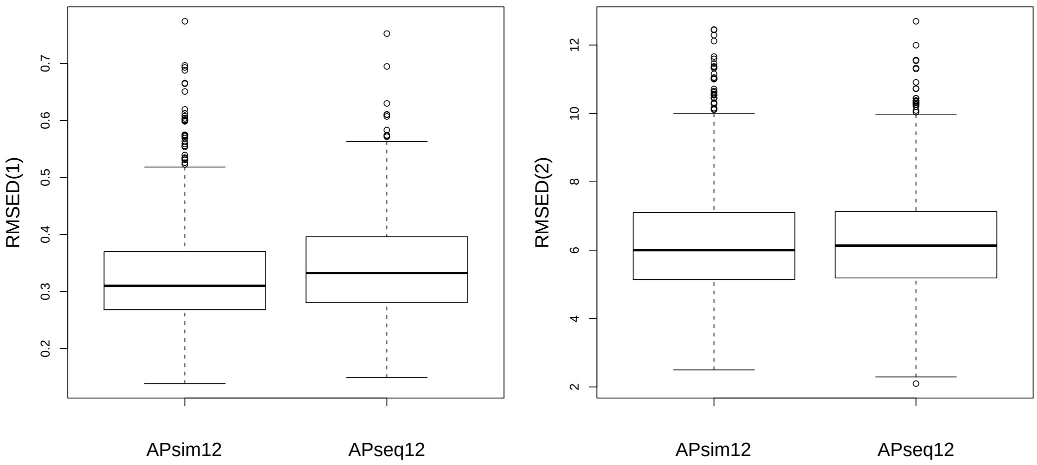


Figure 5.5: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_5 comparing smoothing parameter selection methods.

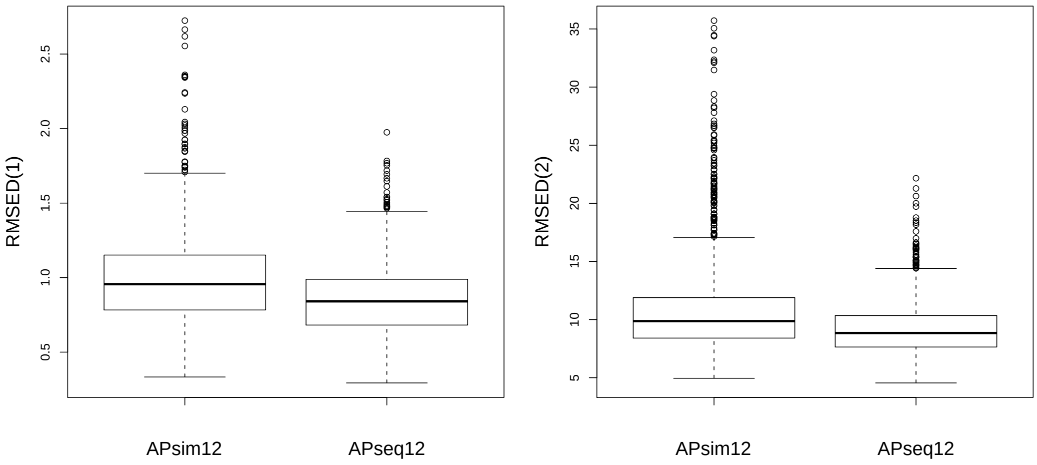


Figure 5.6: Boxplots of RMSED(1) (left plot) and RMSED(2) (right plot) of f_6 comparing smoothing parameter selection methods.

($\approx$ 1350 seconds) were roughly five times faster than those for the simulataneuous ($\approx$ 6900 seconds).

Attention can now be focused on determining the order of the penalties enforced on differences of coefficients. In the P-Spline literature little fuss is made regarding the choice of order of penalty, d. Whether this is the case for the extra penalty model is to be determined.

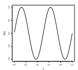
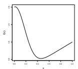
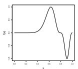
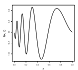
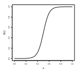
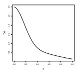
	APsim12	APseq12
 f_1	3.50(0.52)	2.50(0.62)
 f_2	1.06(0.35)	0.92(0.26)
 f_3	4.99(0.88)	5.20(0.98)
 f_4	0.32(0.11)	0.25(0.08)
 f_5	0.33(0.09)	0.34(0.09)
 f_6	1.00(0.33)	0.85(0.24)

Table 5.1: Mean RMSED(1) (standard deviation) comparing smoothing parameter selection methods.

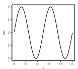
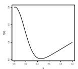
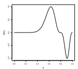
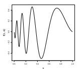
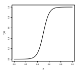
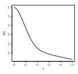
	APsim12	APseq12
 f_1	48.91(6.00)	47.10(5.01)
 f_2	11.57(4.98)	9.76(2.73)
 f_3	188.51(36.39)	186.83(35.23)
 f_4	20.58(33.18)	19.35(33.33)
 f_5	6.22(1.59)	6.21(1.56)
 f_6	11.20(4.75)	9.22(2.48)

Table 5.2: Mean RMSED(2) (standard deviation) comparing smoothing parameter selection methods.

5.4 Choosing Penalty Order

Now that a sequential approach to finding λ has been adopted, it remains to choose the order of the penalties imposed i.e. d_1 and d_2 in (5.1). Since it is recommended in the literature (Eilers & Marx [12]) to choose $d = 2$ for smoothing with P-Splines, this seems like a good place to begin. Here penalties on first, second and third order differences of coefficients are tested. Higher order difference penalties were not examined in much detail and perhaps this may be an area for further work. The order of selection for the sequential approach was found to have

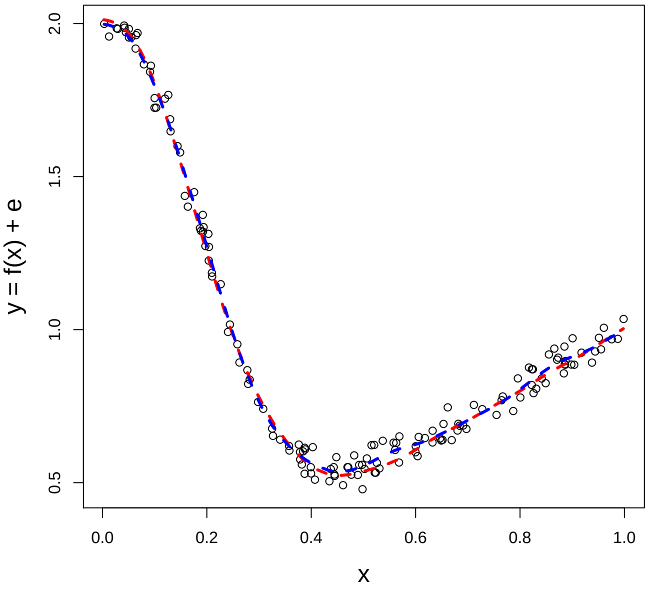


Figure 5.7: Plot of f_2 smoothed using APsim12 (blue) and APseq12 (red).

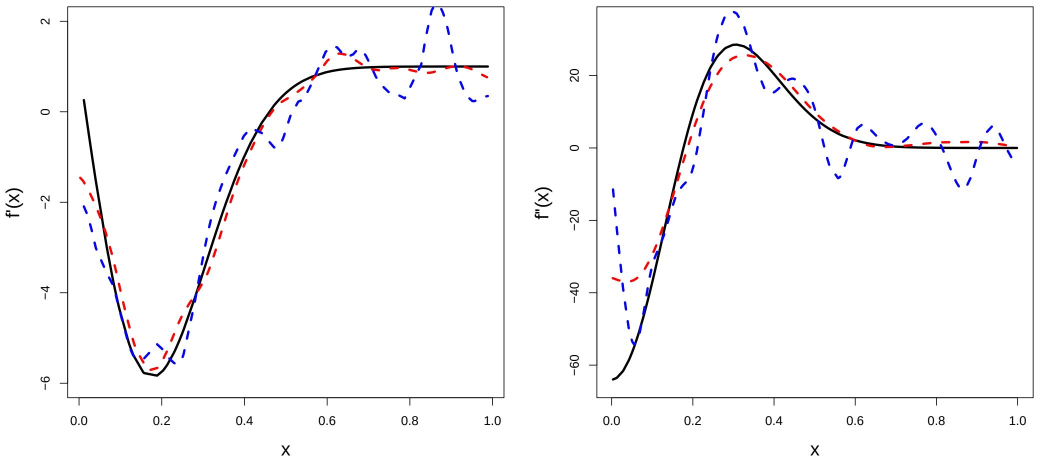


Figure 5.8: Plots of first (left) and second (right) derivative estimates of f_2 using APsim12 (blue) and APseq12 (red) with actual first and second derivative in black.

negligible impact on derivative estimates through a brief investigation.

5.4.1 Comparison of Penalty Orders for the Additive Penalty Model

Simulations were carried out to test which pairing of penalty orders is optimal for derivative estimation using a set of functions that mirror typical smoothing scenarios. The methods under examination being APseq12, APseq23 and APseq13. Once again the RMSED($\cdot$) was used as a measure of performance. The different approaches were tested across the six functions f_1 to f_6 from Section 1.5. The x values were simulated, as before, to be uniformly distributed on $[0, 1]$ with Gaussian error added at standard deviation equal to one sixth the range of f. One thousand samples of size $n = 50$ were taken in all comparisons.

Figures 5.9 to 5.11 show boxplots of the RMSED(1) for the six functions f_1 to f_6. Aside from the performance in f_1 the error produced using each penalty pairing is quite similar. It is possible that the difference in error for f_1 is caused by the fact that $f_1 \propto f_1''$ and APseq13 does not penalise second order differences. The pairing of second and third order penalties seems to result in more variability in the error of the derivative estimates with more outliers evident.

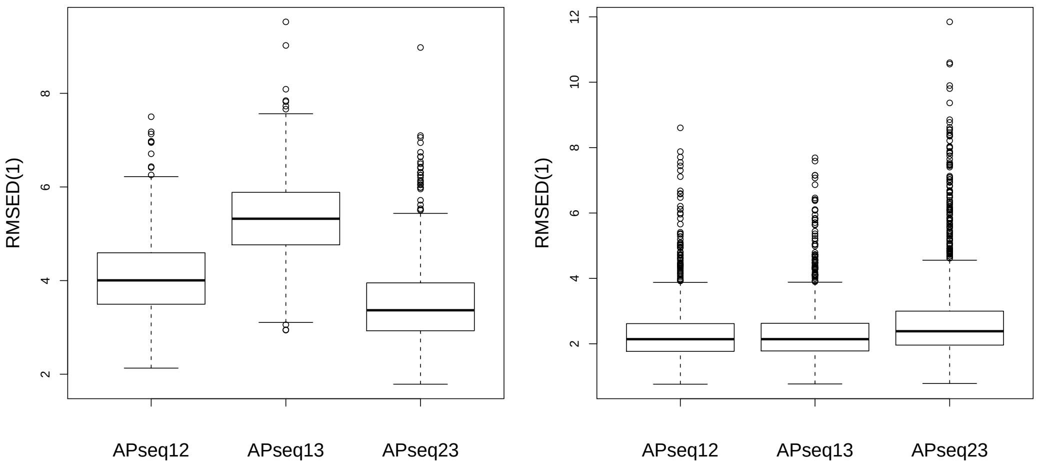


Figure 5.9: Boxplots of RMSED(1) for f_1 (left) and f_2 (right) comparing pairings of penalty orders.

Tables 5.3 and 5.4 show the similarity in goodness of fit for the three pairings of penalties. There is also little to choose from in terms of variability of the fitting process. It would appear that choice of order of difference penalty on the smoothing parameters has little bearing on the performance of an additive penalty method for derivative estimation.

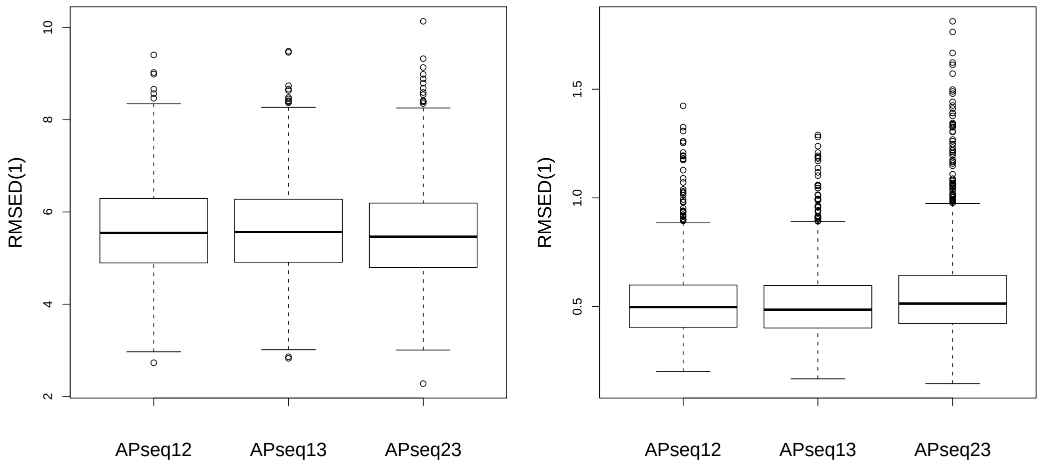


Figure 5.10: Boxplots of RMSED(1) for f_3 (left) and f_4 (right) comparing pairings of penalty orders.

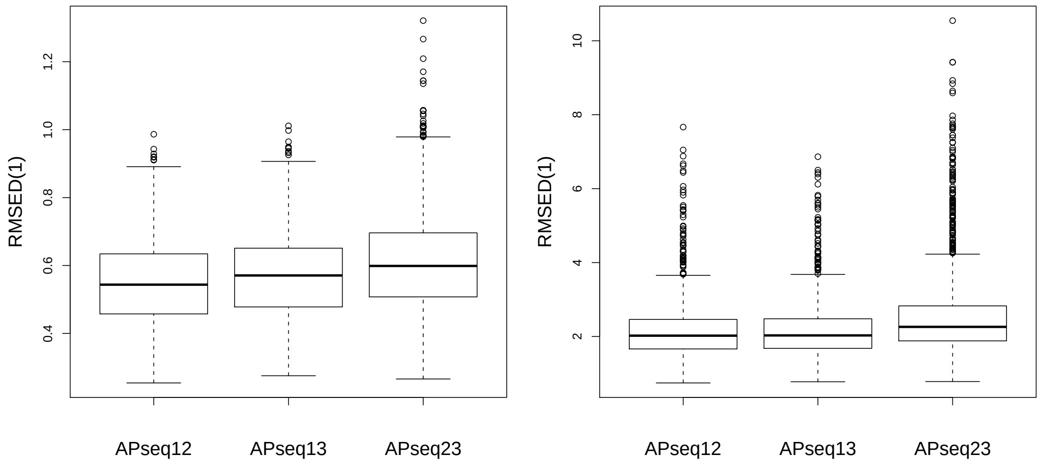


Figure 5.11: Boxplots of RMSED(1) for f_5 (left) and f_6 (right) comparing pairings of penalty orders.

It is quite clear from the boxplots that there is little to no difference in performance among the different penalty order pairings. This agrees with the literature (Eilers & Marx [12]) and results from Section 4.1.1.1, in which choosing the penalty order is seen as having little importance when compared to the choice of λ. In order to simplify comparisons with current derivative estimation techniques such as P-Splines and semiparametric smoothing the APseq12 method will be taken forward, since it has the lowest overall RMSED($\cdot$) across the six functions

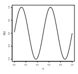
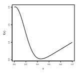
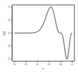
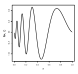
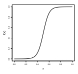
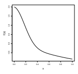
	APseq12	APseq13	APseq23
 f_1	4.08(0.82)	5.35(0.85)	3.53(0.86)
 f_2	2.34(0.97)	2.34(0.94)	2.86(1.54)
 f_3	5.61(1.04)	5.60(1.05)	5.54(1.07)
 f_4	0.52(0.17)	0.51(0.17)	0.57(0.24)
 f_5	0.55(0.13)	0.57(0.13)	0.61(0.15)
 f_6	2.19(0.88)	2.20(0.85)	2.67(1.38)

Table 5.3: Mean RMSED(1) (standard deviation) comparing pairings of penalty orders.

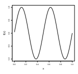
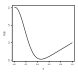
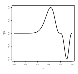
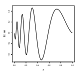
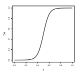
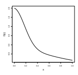
	APseq12	APseq13	APseq23
 f_1	53.7(7.18)	63.94(7.78)	55.65(9.78)
 f_2	20.3(11.18)	20.12(10.03)	27.10(24.75)
 f_3	188.6(35.70)	188.77(35.77)	189.27(36.51)
 f_4	21.7(32.94)	21.57(33.00)	23.47(33.19)
 f_5	7.9(1.73)	8.17(1.73)	8.72(2.39)
 f_6	19.2(10.01)	19.15(9.02)	25.56(22.23)

Table 5.4: Mean RMSED(2) (standard deviation) comparing pairings of penalty orders.

$f_1, \ldots, f_6$ and two derivatives in our simulations. For brevity, from this point forward APseq12 will be abbreviated as AP.

5.5 Comparison of the AP with P-Splines and Semiparametric Regression

It has been seen in Section 3.2.5 how modern spline smoothing techniques struggle in terms of estimating derivatives. Whether the additive penalty method improves estimation of these derivatives is investigated here.

5.5.1 Simulation Study for Comparison of the Additive Penalty with Spline Smoothing Methods

Once again simulated data are used to gauge performance. The methods under investigation are the AP, P-Splines and semiparametric regression since the latter two were seen to perform the best among available modern techniques in Chapter 3.

Testing was performed across the familiar six functions f_1 to f_6. For each of the functions, $n = 50$ observations were simulated 1000 times using x uniform on $[0, 1]$ and error $\epsilon \sim N(0, \sigma^2)$ added, with $\sigma = \frac{1}{3} range(f)$.

Figures 5.12 to 5.14 display boxplots of RMSED(1) for each function. The AP displays visibly lower RMSED(1) across all functions under consideration although in f_6 the semiparametric regression method seems competitive. There is a vast improvement present for derivative estimation of f_1 when using the AP over the other two approaches. Each technique does however result in many outliers for estimates of f_2 and f_6 which are quite similar in terms shape, i.e. smoothing difficultly.

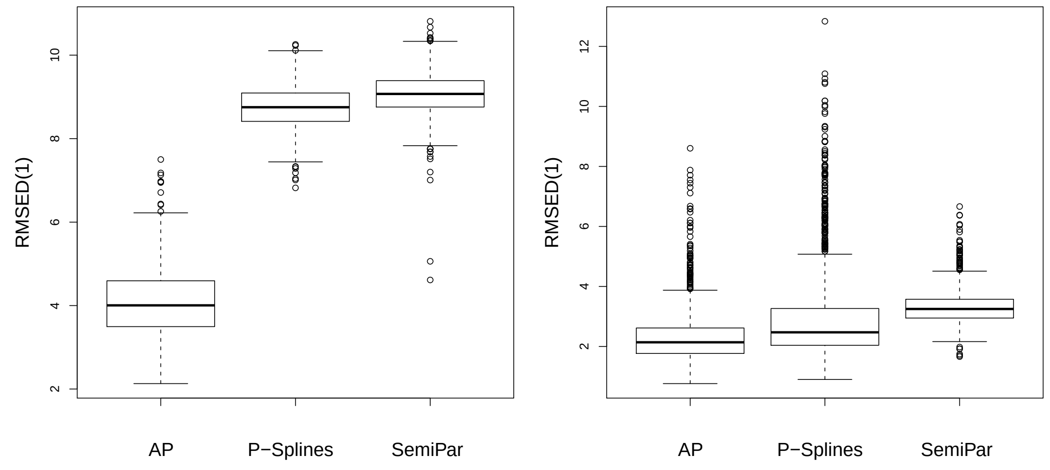


Figure 5.12: Boxplots of RMSED(1) for f_1 (left) and f_2 (right) comparing the AP approach with P-Splines and semiparametric regression.

Tables 5.5 and 5.6 highlight the improvements made by using an extra penalty term when

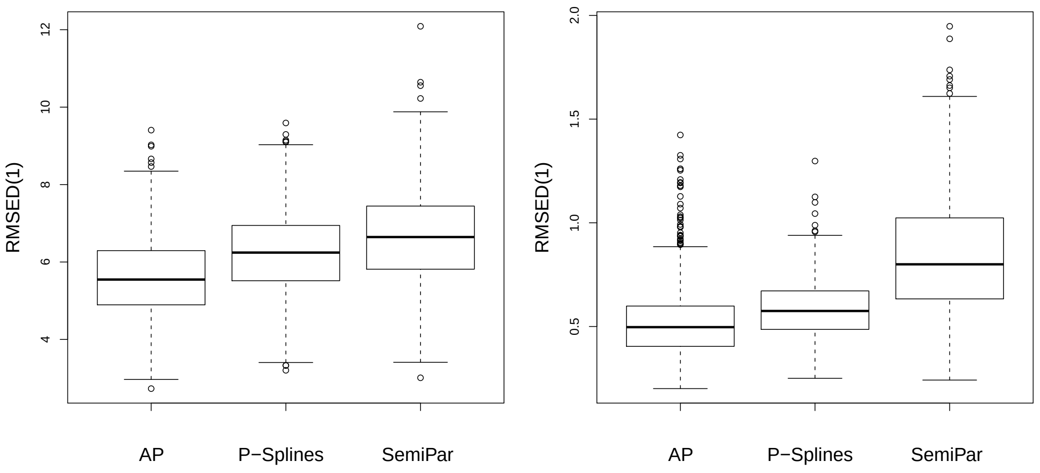


Figure 5.13: Boxplots of RMSED(1) for f_3 (left) and f_4 (right) comparing the AP approach with P-Splines and semiparametric regression.

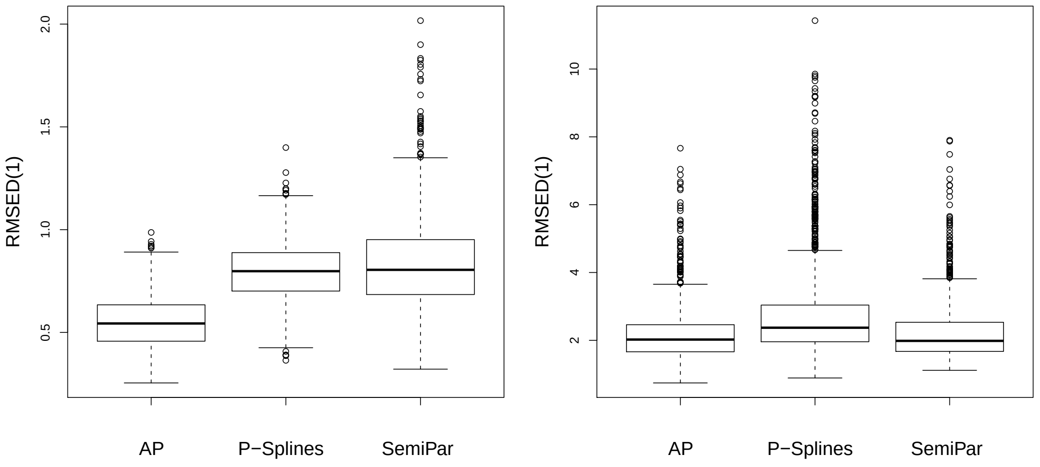


Figure 5.14: Boxplots of RMSED(1) for f_5 (left) and f_6 (right) comparing the AP approach with P-Splines and semiparametric regression.

searching for first and second derivative estimates. For each of the functions under investigation the AP method displays the lowest mean RMSED($\cdot$) and the lowest variance of RMSED($\cdot$) for f_3, f_5, f_6. The results of this simulation study give empirical evidence that using the additional penalty leads to better performance in derivative estimation.

In terms of RMSED($\cdot$), the results of this simulation study suggest that employing an extra additive penalty term to a P-Spline smoothing process benefits derivative estimation. By taking

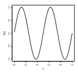
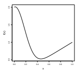
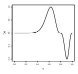
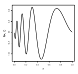
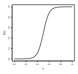
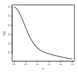
	AP	P-Splines	SemiPar
 f_1	4.08(0.82)	8.74(0.52)	9.06(0.53)
 f_2	2.34(0.97)	3.09(1.77)	3.34(0.61)
 f_3	5.61(1.04)	6.22(1.06)	6.64(1.22)
 f_4	0.52(0.17)	0.59(0.14)	0.84(0.28)
 f_5	0.55(0.13)	0.80(0.15)	0.85(0.24)
 f_6	2.19(0.88)	2.89(1.58)	2.26(0.93)

Table 5.5: Mean RMSED(1) (standard deviation) comparing the AP approach with P-Splines and semiparametric regression.

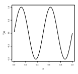
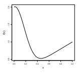
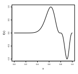
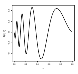
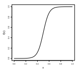
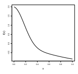
	AP	P-Splines	SemiPar
 f_1	53.7(7.18)	106.4(5.26)	112.0(6.40)
 f_2	20.3(11.18)	29.9(27.81)	25.3(5.28)
 f_3	188.6(35.70)	189.0(36.45)	193.4(37.86)
 f_4	21.7(32.94)	24.1(33.35)	28.4(32.48)
 f_5	7.9(1.73)	10.0(1.90)	14.5(3.72)
 f_6	19.2(10.01)	28.0(24.98)	27.1(10.31)

Table 5.6: Mean RMSED(2) (standard deviation) comparing the AP approach with P-Splines and semiparametric regression.

a closer look at the estimates returned for one of the functions, e.g. $f_1 = \sin(4\pi x)$, how this improvement is attained can be examined. Figure 5.15 shows smooths of f_1 using the three methods AP, P-Splines and semiparametric smoothing. One notices that the P-Spline (blue) dashed line seems to give the best smooth of the data, with both the AP and semiparametric methods slightly oversmoothing the behaviour of the data in the second crest and trough of the sine wave.

The actual first and second derivatives of f_1 (i.e. the gold standard) are displayed as the solid

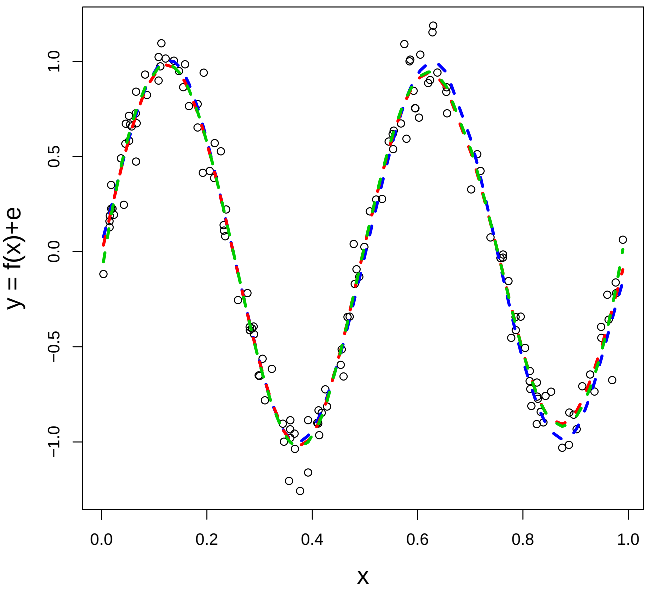


Figure 5.15: Plot of f_1 smoothed using AP (red), P-Splines (blue) and semiparametric (green) methods.

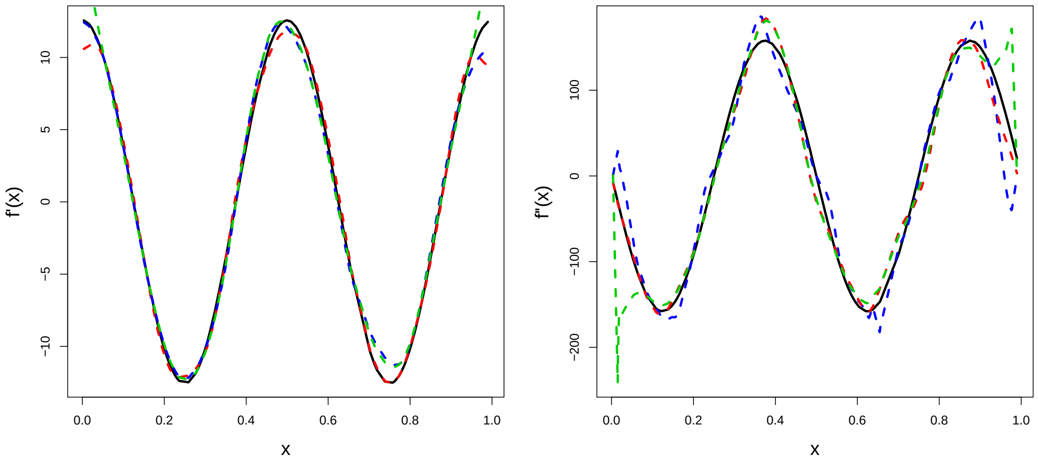


Figure 5.16: Plots of first (left plot) and second (right plot) derivative estimates of f_1 using AP (red), P-Splines (blue) and semiparametric (green) methods.

black curve in Figure 5.16. The coloured dashed lines show the efforts of the three techniques to capture these derivatives. Very similar estimates of the first derivative are made by all three methods, with the AP slightly more accurate in describing the second trough of the derivative of the sine wave. In the graph of the second derivative the AP method is clearly superior, with both the P-Spline and semiparametric fits exhibiting undersmoothing and boundary effects.

5.6 Comparison of the AP with P-Splines in the Motivating Datasets

Since it is apparent from the simulations that an improvement in derivative estimates can be achieved using the additive penalty, the motivating examples can now be revisited for comparison. It is likely, from what has been discussed in this Chapter, that the additive penalty method will display poorer estimates of the underlying function explaining the data but will offer improved derivative estimates.

5.6.1 Winter Nutrients Data

Figure 5.17 exhibits smooth fits to the 1990 NTRZ and Phosphate data by P-Splines and the AP method. Once more, the P-Spline approach leads to a more ‘wiggly’ fit for both Winter Nutrients because of the single smoothing penalty term. The goal here is to identify significant change in the trend of the data. Once more, this can be estimated using first derivative estimates.

Figure 5.18 shows the first derivative estimates for NTRZ and Phosphate data using P-Splines and the additive penalty approach. From the graphs, there is possibly a difference between the two methods in the time at which zero crossings occur. However, until variability bands for these derivative estimates have been developed, whether these zero crossings are ‘significant’, or merely due to sampling variation, cannot be confirmed.

5.6.2 Scottish Bird Count Data

The left panel of Figure 5.19 displays counts of Grey Plover from 1974 to 2004 along with smooth fits by the P-Spline and AP methods. Each of these smooths has used generalised smoothing to deal with the count data observed here. It is evident that the AP fit is the smoother of the two. This is expected due to the extra smoothing penalty enforced by the

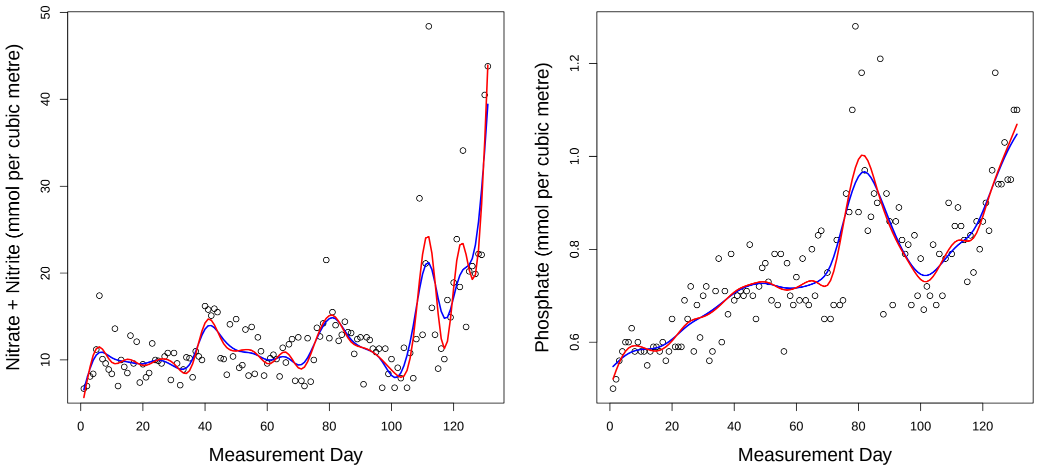


Figure 5.17: Plot of NTRZ (left) and Phosphate (right) data with smooth fits by the P-Spline (red) and AP (blue) methods.

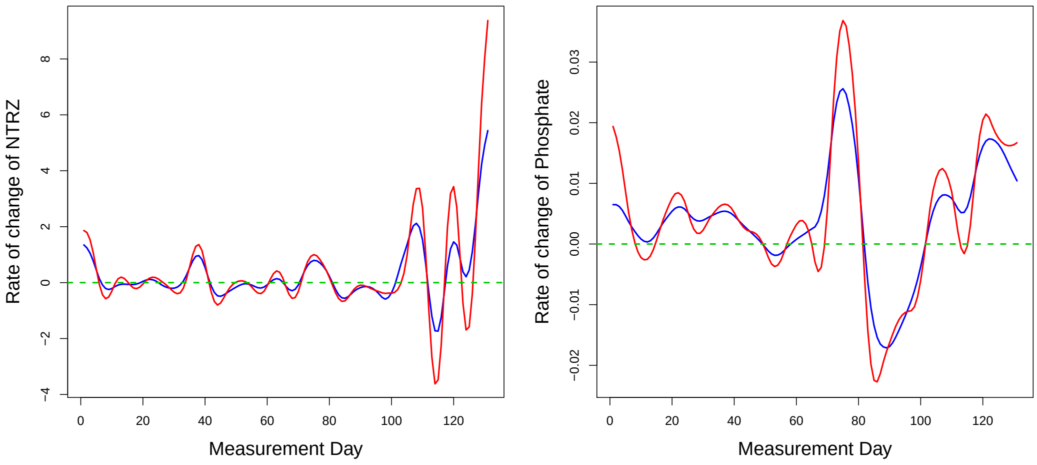


Figure 5.18: First derivative estimate of 1990 NTRZ (left) and Phosphate (right) with smooth fits by the P-Spline (red) and AP (blue) methods.

additive penalty fit. The goal of this study is to find the first instance of a decreasing count of Grey Plover and this can be estimated by looking for zero crossings of the first derivative estimate.

First derivative estimates of the Grey Plover data using P-Splines and the AP method can be seen in the right panel of Figure 5.19. There is little graphical difference between the first (and only) zero crossing of each smooth estimate, although whether or not these crossings are

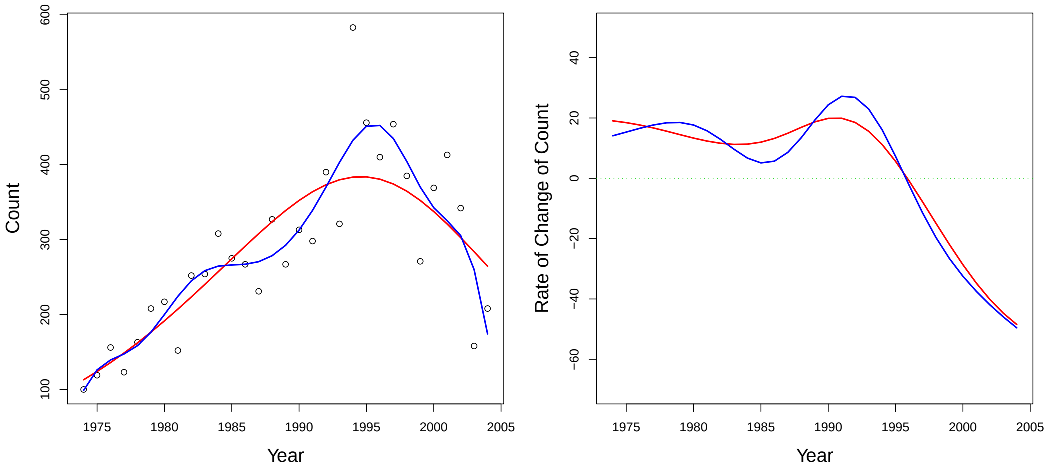


Figure 5.19: Grey Plover data with smooth fits (left) and first derivative estimates (right) by the P-Spline (blue) and AP12 (red) method.

significant is not known as variability bands have yet to be developed (see Chapter 6).

5.6.3 Blood Lactate Data

Figure 5.20 shows some difference between estimates of the lactate curve and its second derivative from the P-Spline and AP methods. The AP fit is more crude and the P-Spline estimate appears to be a better fit to the data. However, as has been seen previously (e.g. Section 2.2.4), this better fit of the underlying function does not always lead to a better derivative estimate.

The AP approach is oversmoothing here and this can be seen in Figure 5.20 to lead to a difference in derivative estimates. The AP smooth is far less 'wiggly' than the P-Spline method. Pertaining to the main question of interest in this study, the right panel of Figure 5.20 shows there to be a difference of roughly 1 km/h in the speed at which the maximum second derivative of the lactate function is estimated to occur when comparing the AP and P-Spline methods. This is quite a considerable difference given the range of speeds in question. From findings based on the simulations in this Chapter, it is anticipated that the additive penalty estimates are more accurate and therefore one would tend to believe the estimate given by the AP approach.

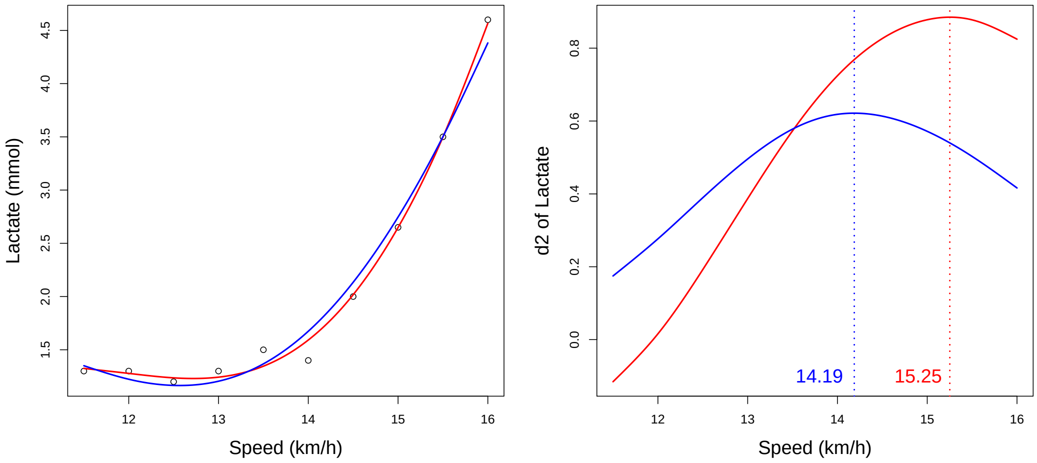


Figure 5.20: Plot of one individual's lactate data with smooth fits (left) and second derivative estimates (right) using P-Splines (red) and the AP (blue).

5.6.4 Astronomical Data

For the Astronomical data the main aim is to estimate ρ_{tot} (which is currently an unmeasurable quantity) using the equation

$$\frac{d}{dx}\frac{\frac{d}{dx}(\rho_{gas}kT\mu m_p)}{\rho_{gas}} = -4\pi G\rho_{tot} \tag{5.7}$$

where G is the universal gravitational constant, k is Boltzmann's constant, x (in arcmin) is a measure of distance, μ is the mean molecular weight in any cluster and m_p is the mass of the proton.

The collected data consist of the T and ρ_{gas} variables. Thus it is necessary to obtain estimates of $\frac{d}{dx}(\rho_{gas}kT\mu m_p)$ and then of $\frac{d}{dx}\frac{\frac{d}{dx}(\rho_{gas}kT\mu m_p)}{\rho_{gas}}$. Figure 5.21 gives estimates of these by the P-Spline and AP methods. From 5.21 there seems to be little difference in the fits from each method. There is some difference in the right panel of Figure 5.21 and thus the estimated fitted values corresponding to ρ_{tot} will differ depending on the method which is used for derivative estimation. This will impact on astronomical models based on this estimate of ρ_{tot}.

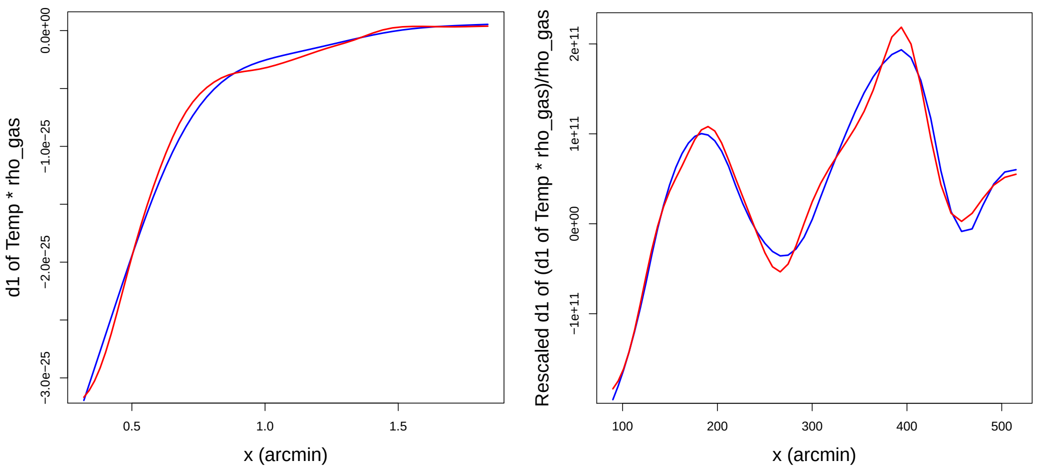


Figure 5.21: Plot of estimates of $\frac{d}{dx}(T \times \rho_{gas})$ (left) and $\frac{d}{dx}\left(\frac{\frac{d}{dx}(T \times \rho_{gas})}{\rho_{gas}}\right)$ after rescaling (right) using P-Splines (red) and an AP (blue).

5.7 Chapter Summary

Prior to this Chapter, considerable problems of undersmoothing and boundary effects in derivative estimation have been uncovered. An extra additive penalty included in the standard P-Spline model was proposed to attempt to remedy these concerns. This model necessitates the selection of two smoothing parameters (λ_1, λ_2), and so a decision on whether to use a sequential or simultaneous approach was needed. The sequential selection of (λ_1, λ_2) resulted in better derivative estimation performance and the simultaneous approach was eliminated.

Simulations varying the difference penalty order pairings (d_1, d_2) were carried out to test whether this choice had an effect on the error in derivative estimation. This simulation offered very similar results for each of the three pairings considered, although overall they suggest using first and second order difference penalties.

Proceeding with this form of additive penalty model, comparisons of precision in derivative estimation were made with P-Splines and semiparametric smoothing. The results of simulations suggest that the AP approach is to be preferred when estimating the derivatives of a (regression) function.

The AP and P-Spline methods were applied to the motivating datasets resulting in noticeably different estimates. Using the simulations as evidence it would seem that the AP estimates should be believed. However, these estimates still fall short in a number of categories when applying to the illustrative datasets. In the Winter Nutrient and Scottish Bird Count data it is necessary to determine whether zero crossings of the first derivative are indeed significant (i.e. represent a systematic change in the population of interest). Estimates of the standard error of these derivative estimates are now required so that variability/confidence bands can be constructed and this shall be the main focus of the next Chapter.

Chapter 6

Inference Using Derivative Estimates

Statistical inference is the process of drawing conclusions about a population parameter based on a sample statistic and its estimated standard error. These can be used for significance tests (tests that a parameter equals a certain value) or interval estimation (where a plausible range of values for the parameter are provided).

When estimating derivatives the same theory applies. The population parameter is the true derivative $f^{(l)}$, $l = 0, 1, 2, \ldots$ for which a point estimate $\hat{f}^{(l)}$ is calculated using, for example, spline smoothing. How to obtain this point estimate has been discussed and the previous Chapter provided empirical evidence that the additive penalty (AP) method offers improved performance for estimating the first and second derivatives of noisy data.

In this Chapter theory for estimating the standard error of derivative estimates is developed in order to create likely ranges around smooth estimates of the regression function and its derivatives. The motivating datasets introduce problems which cannot be solved purely through fitted values of a derivative estimate; features of the derivatives are often what is important. Several simulation studies to test the performance of variability bands and the accuracy of estimating certain features of a relationship between variables are carried out. The motivating examples from Chapter 1 are revisited with methods for estimating variability bands, and consequently areas of significant change, available.

6.1 Variability Bands

In statistical inference a confidence interval provides a likely range of values for a population parameter (e.g. μ) based on its point estimate (e.g. $\bar{x}$) and its estimated standard error $se(\bar{x})$. For large samples, a 95% confidence interval for μ is

$$\bar{x} \pm 1.96 se(\bar{x}).$$

In the smoothing/derivative estimation context where $f^{(l)}$, $l = 0, 1, 2, \ldots$, is the population parameter and $\hat{f}^{(l)}$ is its estimate this may be written

$$\hat{f}^{(l)} \pm 2se(\hat{f}^{(l)}) \tag{6.1}$$

and is called a pointwise confidence band. The value 2 (≈1.96) is taken since the asymptotic properties of interval estimation on which the value 1.96 is based are not well established for smoothing and using 2 is just a simple approximation. The confidence band provides a range of pointwise likely values for each observation of the true function/derivative $f^{(l)}$ but, as shall be seen, it is not a 95% confidence interval for the global parameter/function $f^{(l)}$. Thus the term variability band is used for (6.1) in order to disassociate from the well known interpretation of a confidence interval. These variability bands offer a somewhat plausible range of likely error of a smooth estimate of a function and its derivatives. Therefore they may be used to identify areas of significant change in observed data through comparisons of estimated first derivative variability bands with zero. Define significant increase of a response variable $y = f(x)$ when

$$f' - 2se[f'] > 0 \tag{6.2}$$

and significant decrease when

$$f' + 2se[f'] < 0. \tag{6.3}$$

Otherwise there is no significant change. The estimated standard error could potentially be used for significance testing that $f' = 0$ for example. However, it would be difficult to quantify the power of such a test since the asymptotics of the estimated standard errors are not well defined.

In the Scottish Bird Count and Winter Nutrients data, first derivative estimates can be used to test for evidence of a significant change in the respective response. The region representing significant change can be found by observing regions where both the upper or lower variability bands of a first derivative estimate lie above or below zero simultaneously (6.2, 6.3). In this section methods for calculating variability bands for each of P-Splines, AP and semiparametric smoothing shall be introduced or developed where necessary through estimating $se(f^{(l)})$.

6.1.1 Variability Bands for Semiparametric Regression

Variability bands have been developed for a semiparametric regression estimate and its derivatives. Ruppert, Wand and Carroll [48] describe how these bands are achieved and here that is summarised.

The package `SemiPar` gives the user the option to view variability bands for the smooth fit and derivative estimates but does not contain fitted values for these (nor for the fitted derivative estimates themselves). The Appendix of this thesis provides new code to extract fitted derivative estimates and their variability bands from the `spm` function. This code is useful as, in general, plots of the derivatives with variability bands alone are not adequate for answering some questions, such as the ones involved in the motivating illustrations. For example in the Winter Nutrients data, the analysis needs to report times at which a significant decrease or increase occurs. Simple graphical displays of first derivative estimates with bands may not be enough as reading from a graph is too subjective.

6.1.1.1 Variability Bands for a Semiparametric Regression fit $\hat{f}_{sp}$

Semiparametric regression is similar to the linear mixed model setup i.e.

$$y = X\beta + Zu + \epsilon \tag{6.4}$$

where X is the design matrix for the fixed parameters β, Z is the design matrix for the random effects $u \sim N(0, \sigma_u^2 I)$ and $\epsilon \sim N(0, \sigma_\epsilon^2 I)$. It has been seen in Chapter 3 that, where $y = f(x)+\epsilon$, taking, for example, three fixed parameters $\beta = (\beta_0, \beta_1, \beta_2)$ then fitted values are obtained by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{k=1}^{K} \hat{u}_k (x - \kappa_k)_+^2. \tag{6.5}$$

Similarly to the P-Spline method, the estimate of the observed y values are

$$\hat{y} = \hat{f}_{sp} = \theta^T y \tag{6.6}$$

where θ is a vector of length n turning observed y to fitted $\hat{y}$. The variance of y can then be estimated by

$$\begin{aligned} var(\hat{f}_{sp}) &= var(\theta^T y) \\ &= \theta^T var(y) \theta \\ &= \sigma_\epsilon^2 \theta^T \theta \end{aligned} \tag{6.7}$$

with the standard errors of y estimated as

$$se(\hat{f}_{sp}) = \hat{\sigma}_\epsilon \left\| \theta \right\|. \tag{6.8}$$

The variability bands for y can be estimated as

$$\hat{f}_{sp} \pm 2se(\hat{f}_{sp})$$

where 2 (≈ 1.96) is again chosen as an approximate 95% confidence coefficient. This approximate confidence coefficient is used throughout the rest of this Chapter for all methods. The $var(f_{sp})$ can be estimated by

$$var(\hat{f}_{sp}) = \sigma_\epsilon^2 C (C^T C + \frac{\sigma_\epsilon^2}{\sigma_u^2} D)^{-1} C^T \tag{6.9}$$

with

$$C = [X \ Z] \equiv [1 \ x \ x^2 \ (x - \kappa_1)_+^2 \ldots (x - \kappa_k)_+^2]$$

and

$$D = diag(0, 0, 0, 1, \dots, 1)$$

where the number of zeros beginning the diagonal of D is equal to the number of fixed parameters in the model.

6.1.1.2 Variability Bands for $\hat{f}'_{sp}$

Taking the derivative of $\hat{f}_{sp}$ gives

$$\hat{f}'_{sp} = \hat{\beta}_1 + 2\hat{\beta}_2 x + \sum_{k=1}^{K} 2\hat{u}_k (x - \kappa_k)_+. \tag{6.10}$$

Once more, variability bands are found by

$$\hat{f}'_{sp} \pm 2\sqrt{var(\hat{f}'_{sp})}$$

where

$$var(\hat{f}'_{sp}) = \sigma^2_\epsilon C'(C^T C + \frac{\sigma^2_\epsilon}{\sigma^2_u} D)^{-1} C'^T \tag{6.11}$$

and

$$C' = [X' \ Z'] \equiv [0 \ 1 \ 2x \ 2(x - \kappa_1)_+ \dots 2(x - \kappa_k)_+] \tag{6.12}$$

As an example, Figure 6.1 displays variability bands for the semiparametric regression estimate of the Phosphate data and the corresponding derivative estimate.

6.1.2 Variability Bands for P-Splines

Variability bands have not been developed for derivative estimates from a P-Spline fit in the literature and a new approach is now developed. Variability bands for a P-Spline fit appear in the literature (Marx & Eilers [35]) and the theory underlying these bands is summarised below.

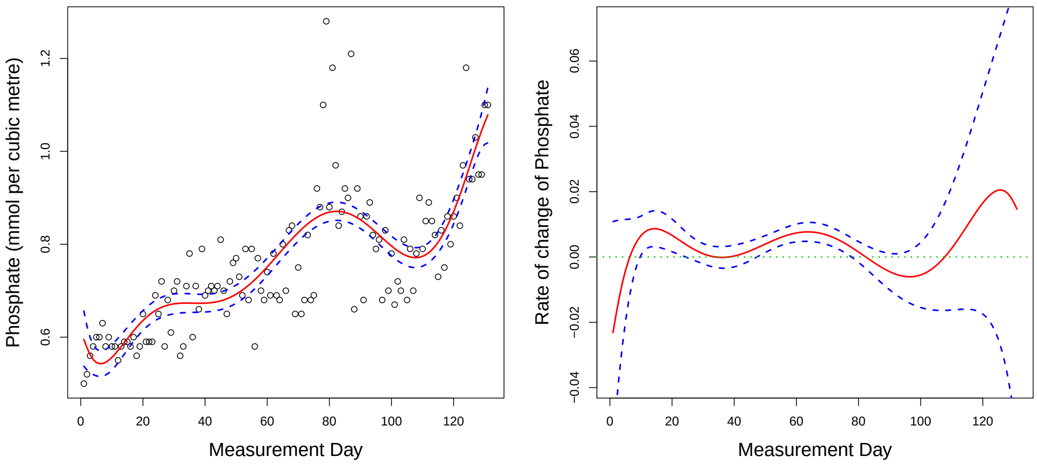


Figure 6.1: Fit (left) and first derivative estimate (right) of the Phosphate data using semi-parametric regression with variability bands.

6.1.2.1 Variability Bands for a P-Spline Fit $\hat{f}_p$

Consider the familiar setup $y = f_p(x) + \epsilon$. For a B-spline basis one may write

$$f_p = B\alpha$$

for some B-Spline matrix B and set of coefficients α. Then one may write

$$\hat{y} = \hat{f}_p = B\hat{\alpha} = Hy$$

where

$$H = B(B^T B + \lambda D^T D)^{-1} B^T$$

for some difference matrix D and smoothing parameter λ. Recall from Section 3.2.2 that a difference matrix, taking penalty order $d = 2$ and $m = 5$ coefficients for example, has the form

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Then

$$\begin{aligned} var(\hat{y}) &= var(Hy) \\ &= Hvar(y)H^T. \end{aligned} \tag{6.13}$$

The variance of y is σ^2 which is estimated by

$$\hat{\sigma}^2 = \frac{\|y - \hat{y}\|^2}{n - tr(H)} \tag{6.14}$$

since the effective dimension of the model is $tr(H)$ from (3.10). Variability bands, given by

$$\hat{f}_p \pm 2\sqrt{var(\hat{f}_p)},$$

can be approximated by

$$\hat{f}_p \pm 2\sqrt{diag(\hat{\sigma}^2 HH^T)}. \tag{6.15}$$

6.1.2.2 Variability Bands for a P-Spline First Derivative Estimate $\hat{f}'_p$

From (3.26) one may write

$$\hat{f}'_p = q(qh)^{-1}B'\Delta\hat{\alpha} \tag{6.16}$$

where h is the distance between adjacent knots, Δ is the difference operator, q is the degree of the B-spline basis and B' denotes the B-Spline design matrix for the 'derivative' B-Spline basis of degree $q - 1$. As before variability bands will be of the form

$$\hat{f}'_p \pm 2\sqrt{var(\hat{f}'_p)}.$$

Now $\hat{f}'_p$ is defined by (6.16), however methods for calculating $var(\hat{f}'_p)$ do not appear in the literature. The following is proposed:

$$\begin{aligned} var(\hat{f}'_p) &= var(H'y) \\ &= H'var(y)H'^T \\ &= \hat{\sigma}^2 H'H'^T \end{aligned} \tag{6.17}$$

where σ^2 is estimated by (6.14) and

$$H' = q(qh)^{-1}B'(B^TB + \lambda D^TD)^{-1}B'^T. \tag{6.18}$$

A problem of non-conformity arises here due to the fact that dropping the degree of a B-Spline basis means losing a column from the B-Spline design matrix B. Recall that a matrix B representing a degree q B-Spline basis is of the form

$$B = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \dots & B_m(x_1) \\ B_1(x_2) & B_2(x_2) & \dots & B_m(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & \dots & B_m(x_n) \end{pmatrix}$$

for $i = 1, \dots, n$ observations and $j = 1, \dots, m$ coefficients. The dimension of B depends on the number of observations and the number of basis functions, which itself depends on the number of knots and the degree of the basis, i.e. $m = K+q-1$ where K is the number of knots. Thus a $q-1$ B-Spline design matrix B' has dimension $n \times m'$ where $m' = K+(q-1)-1 = K+q-2 = m-1$ and

$$B' = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \dots & B_{m'}(x_1) \\ B_1(x_2) & B_2(x_2) & \dots & B_{m'}(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & \dots & B_{m'}(x_n) \end{pmatrix}.$$

Thus (6.18) involves multiplying an $n \times m'$ matrix B' by the $n \times m$ matrix $(B^TB+\lambda D^TD)^{-1}$, which is not possible. This can be rectified by attaching an extra column of zeros to the B' matrix. This may be appended to the first or last column of B'. Testing has shown negligible differences in results between this choice and here the case where a column of zeros are added to the end of B' is detailed. This new conformant B-Spline design matrix $B'_{(0)}$ is

$$B'_{(0)} = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \dots & B_{m'}(x_1) & 0 \\ B_1(x_2) & B_2(x_2) & \dots & B_{m'}(x_2) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & \dots & B_{m'}(x_n) & 0 \end{pmatrix}.$$

Similarly, when moving to higher order derivatives extra columns of zeros must be added to the end of the B matrix with the number of columns of zeros equal to the order of derivative computed. Variability bands can now be estimated using

$$\hat{f}'_p \pm 2\sqrt{diag(\hat{\sigma}^2 H' H'^T)}$$

with H' defined as

$$H' = q(qh)^{-1} B'_{(0)} (B^T B + \lambda D^T D)^{-1} B'^T_{(0)}. \tag{6.19}$$

Figure 6.2 exhibits variability bands for a P-Spline fit and corresponding derivative estimate of the Phosphate data.

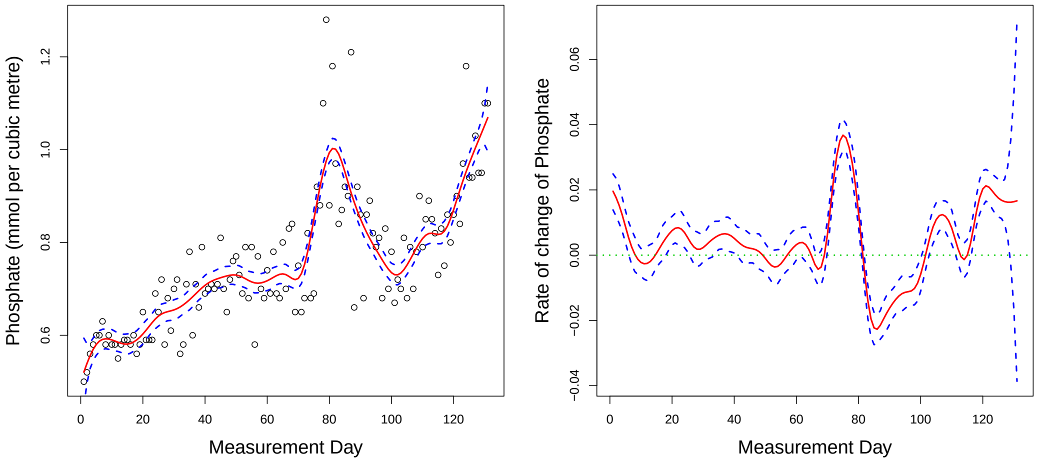


Figure 6.2: Fit (left) and first derivative estimate (right) of the Phosphate data using P-Splines with variability bands.

6.1.3 Variability Bands for the Additive Penalty Method

Variability bands for the additive penalty method are found in much the same way as in P-Splines with the extra smoothing/penalty term added to the P-Spline hat matrix resulting in the AP hat matrix H_{ap}. The non-conformity issue arises here again and the same amendment is made, i.e. adding a column of zeros to the B matrix for each order of derivative estimated.

6.1.3.1 Variability Bands for an AP fit $\hat{f}_{ap}$

Let

$$\hat{f}_{ap} = B\hat{\alpha} = H_{ap}y$$

where

$$H_{ap} = B(B^T B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1} B^T$$

with d_1 and d_2 the orders of difference penalty used in the smoothing process. Variability bands can be estimated by

$$\hat{f}_{ap} \pm 2\sqrt{var(\hat{f}_{ap})}$$

where

$$\begin{aligned} var(\hat{f}_{ap}) &= var(H_{ap}y) \\ &= H_{ap}var(y)H_{ap}^T \\ &= \sigma^2 H_{ap}H_{ap}^T \end{aligned}$$

with

$$\hat{\sigma}^2 = \frac{\|y - \hat{y}\|^2}{n - tr(H_{ap})}.$$

Then bands are obtained using

$$\hat{f}_{ap} \pm 2\sqrt{diag(\hat{\sigma}^2 H_{ap}H_{ap}^T)}.$$

6.1.3.2 Variability Bands for an AP First Derivative Estimate $\hat{f}'_{ap}$

The de Boor formula is used to obtain derivatives from AP estimates. Since it uses an equally spaced B-Spline basis

$$\hat{f}'_{ap} = q(qh)^{-1}B'\Delta\hat{\alpha} = H'_{ap}y$$

where h is the distance between adjacent knots, q is the degree of the B-Spline basis and

$$H'_{ap} = q(qh)^{-1}B'(B^TB + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1}B'^T.$$

Again the B' design matrix needs the addition of a column of zeros to the end of B' in order to conform. By calling this new design matrix $B'_{(0)}$ redefine

$$H'_{ap} = q(qh)^{-1}B'_{(0)}(B^TB + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1}B'^T_{(0)}.$$

The variance is estimated as follows

$$\begin{aligned} var(\hat{f}'_{ap}) &= var(H'_{ap}y) \\ &= H'_{ap}var(y)H'^T_{ap} \\ &= \hat{\sigma}^2 H'_{ap}H'^T_{ap}, \end{aligned}$$

and variability bands are given by

$$\hat{f}'_{ap} \pm 2\sqrt{diag(\hat{\sigma}^2 H'_{ap}H'^T_{ap})}$$

Figure 6.3 shows variability bands for a smooth estimate of the Phosphate data and corresponding derivative using the AP method.

6.1.4 Variability Bands Using the Bootstrap

Residual resampling is a useful computational approach for estimating variability bands in linear and nonlinear models. Consider the familiar model $y = f(x) + \epsilon$. Once $\hat{f}$ has been estimated, through P-Splines for example, a vector of residuals $e = (e_1, \dots, e_n)$ is obtained

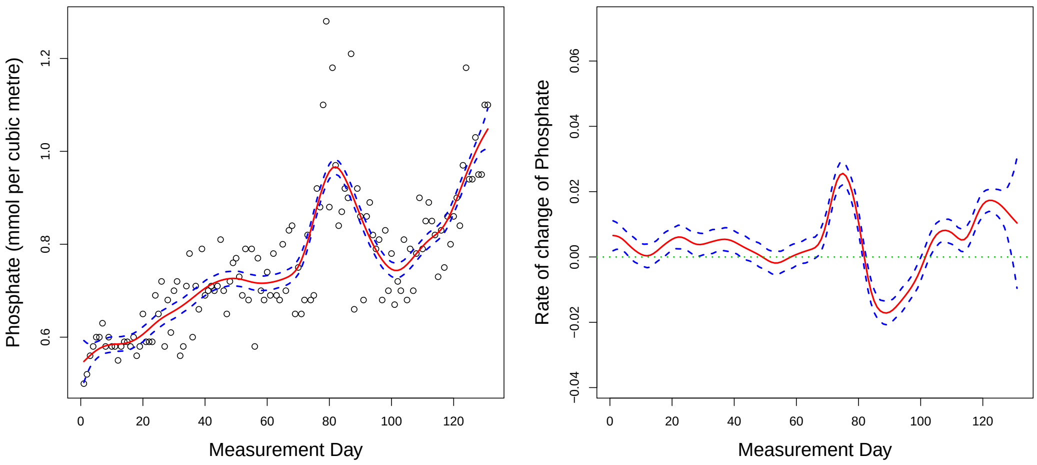


Figure 6.3: Fit (left) and first derivative estimate (right) of the Phosphate data using the AP with variability bands.

where $e_i = y_i - \hat{y}_i$ for $i = 1, \ldots, n$. Resampling residuals works by randomly resampling the vector of residuals e with replacement to form vectors e^* of length n. Now taking

$$y^* = \hat{y} + e^* \tag{6.20}$$

gives one resample. This can be repeated R times giving R resampled response vectors and smoothing each of these gives R fitted curves. Picking off the 2.5% and 97.5% percentiles of the $R \times n$ matrix of fitted values gives a data driven approach to creating pointwise variability bands for $\hat{f}$. Similarly, 2.5% and 97.5% percentiles of $R \times n$ matrices of first and second derivative estimates are taken as pointwise variability bands for $\hat{f}'$ and $\hat{f}''$. An example of such bands using the Phosphate data is given in Figure 6.4.

6.1.5 Comparison of Variability Band Estimators

Quantifying the performance of variability bands is not straightforward. There are two levels of performance which are vital; variability bands should offer an accurate description of the likely pointwise error across the range of the data whilst maintaining a respectable width, i.e. precision. Large band width implies low precision and vice versa. A very wide (i.e. imprecise) variability band is likely to contain the actual underlying function but will overestimate, and

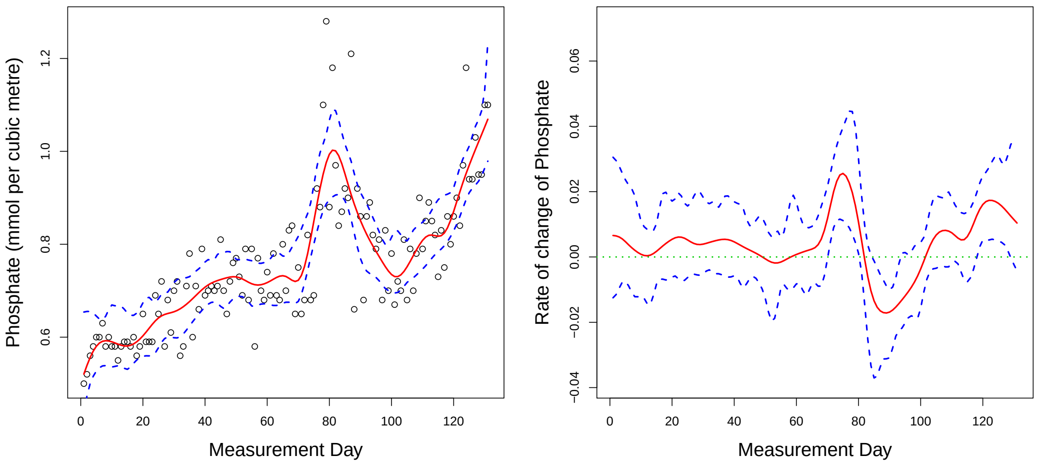


Figure 6.4: Fit (left) and first derivative estimate (right) of the Phosphate data using with bootstrap variability bands.

give little information about, the likely pointwise standard error. For example, Figure 6.5 displays data simulated from $f_2 = x + 2e^{-16x^2}$ with Gaussian error added. The right panel shows the true function which generated the data. An optimal band will contain the true function whilst maintaining the lowest width, i.e. good precision.

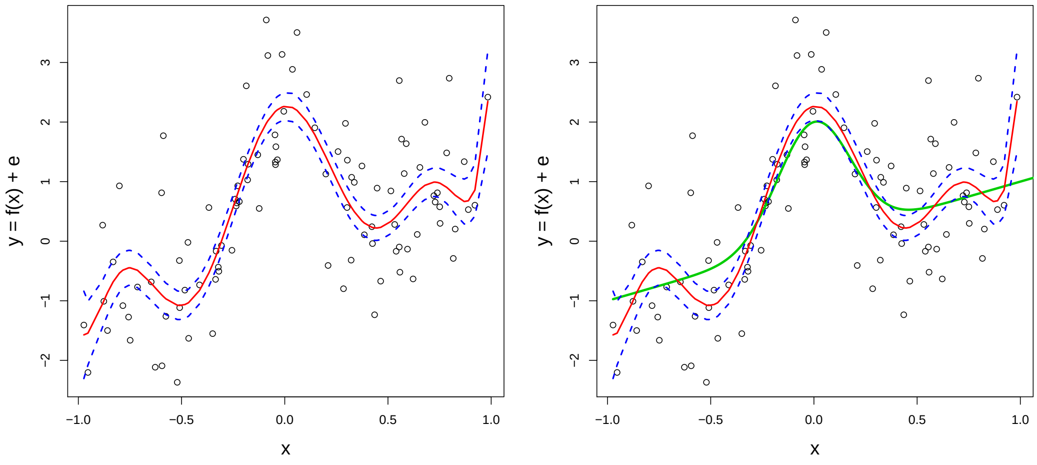


Figure 6.5: Data simulated from f_2 with semiparametric fit (red) and variability bands (blue). The right panel displays the true function (green).

One method for measuring the precision of each of the variability bands introduced in the previous section is as simple as finding the mean standard error across the range of x. Measuring the accuracy of the variability band, however, is more challenging. Optimum bands will fully contain the actual function or derivative whilst being as precise as possible (Figure 6.6).

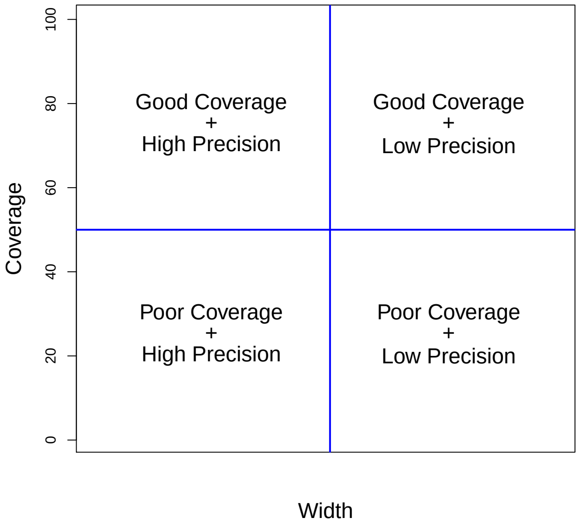

Coverage
0
20
40
60
80
100
Good Coverage + High Precision
Good Coverage + Low Precision
Poor Coverage + High Precision
Poor Coverage + Low Precision
Width

Figure 6.6: A representation of good variability band performance.

Good coverage is defined by the actual function or derivative lying inside their respective bands. In the simulation study that follows, the coverage of variability bands is measured in four ways. Firstly, as a measure of global coverage, the number of points of $f(x_i)$ for $i = 1, \ldots, n$ which lie inside $\hat{f} \pm 2se(\hat{f})$ (likewise for f' and f'') were counted for each of the 1000 samples. To measure pointwise coverage, the inclusion of three arbitrarily chosen observations in the variability bands produced by each of the four methods was recorded for each of 1000 simulations. Thirdly, the number of times the actual function f departed the variability bands was observed. Finally, a 'slope zero test' was conducted. This entailed first derivative estimates of a constant function f_c with error added being compared to $y' = 0$. Since f_c has zero slope, the line $y' = 0$ should be fully contained inside the variability bands for $\hat{f}'_c$. The number of points $f'_c(x_i)$, $i = 1, \ldots, n$ along the line $y' = 0$ contained in the bands was noted. This test is similar to a global test for zero slope used in simple linear regression. However, the size of this test does not represent the number of times in repeated sampling the entire line $y' = 0$ is not captured fully by the bands. Here it represents the number of points $f'_c(x_i)$ along the line $y' = 0$ in each sample which will not be captured by the estimated variability bands.

6.1.5.1 Strategy

The four variability band estimators, namely the additive penalty method (AP), P-Splines, semiparametric smoothing and bootstrap were tested on the function $f_2 = x + 2e^{-16x^2}$ with x uniform on $[-1, 1]$ and Normal error $\epsilon \sim N(0, \sigma^2)$ added with $\sigma = \frac{1}{3} range(f)$. One thousand samples of size $n = 100$ were simulated. Each of the four methods were measured for global and pointwise coverage for each derivative ($l = 0, 1, 2$), precision (i.e. mean standard error) and departures (i.e. the number of times that $f^{(l)}$ departed the range $\hat{f}^{(l)} \pm 2se[\hat{f}^{(l)}]$ for each $l = 0, 1, 2$). The bootstrap bands are estimated by first resampling the residuals from a P-Spline fit to the data 1000 times to obtain 1000 samples. For bands around $\hat{f}$ each sample is smoothed using P-Splines and the 2.5 and 97.5 percentiles of these smooth fits are taken as the estimated variability bands. For bands around $\hat{f}^{(l)}$ ($l = 1, 2$) each sample is smoothed using the AP and 2.5 and 97.5 percentiles of the derivative estimates are taken as the estimated variability bands.

6.1.5.2 Results

The findings from these simulations are summarised separately for coverage, precision and departures, in order to give a clearer picture of the results. Beginning with coverage, Figure 6.7 and Table 6.1 summarise the coverage percentages of f_2, f_2' and f_2'' (written f, f' and f'' for the remainder of these results) for each method. In terms of f, the semiparametric and bootstrap variability bands are evidently the best in terms of coverage. They double the performance of both the P-Spline and AP methods while halving (roughly) the variability of the coverage across the 1000 samples.

	f	f'	f''
P-Spline	48.7 (14.5)	59.8 (14.9)	64.6 (18.7)
AP	48.4 (14.3)	62.8 (14.4)	67.5 (16.5)
SemiPar	89.9 (7.2)	55.0 (9.2)	44.2 (10.7)
Bootstrap	91.4 (4.8)	94.1 (4.4)	95.3 (3.6)

Table 6.1: Mean (standard deviation) coverage percentage of f, f' and f''.

There is a noticeable trend when focussing on derivative estimation. Whereas the P-

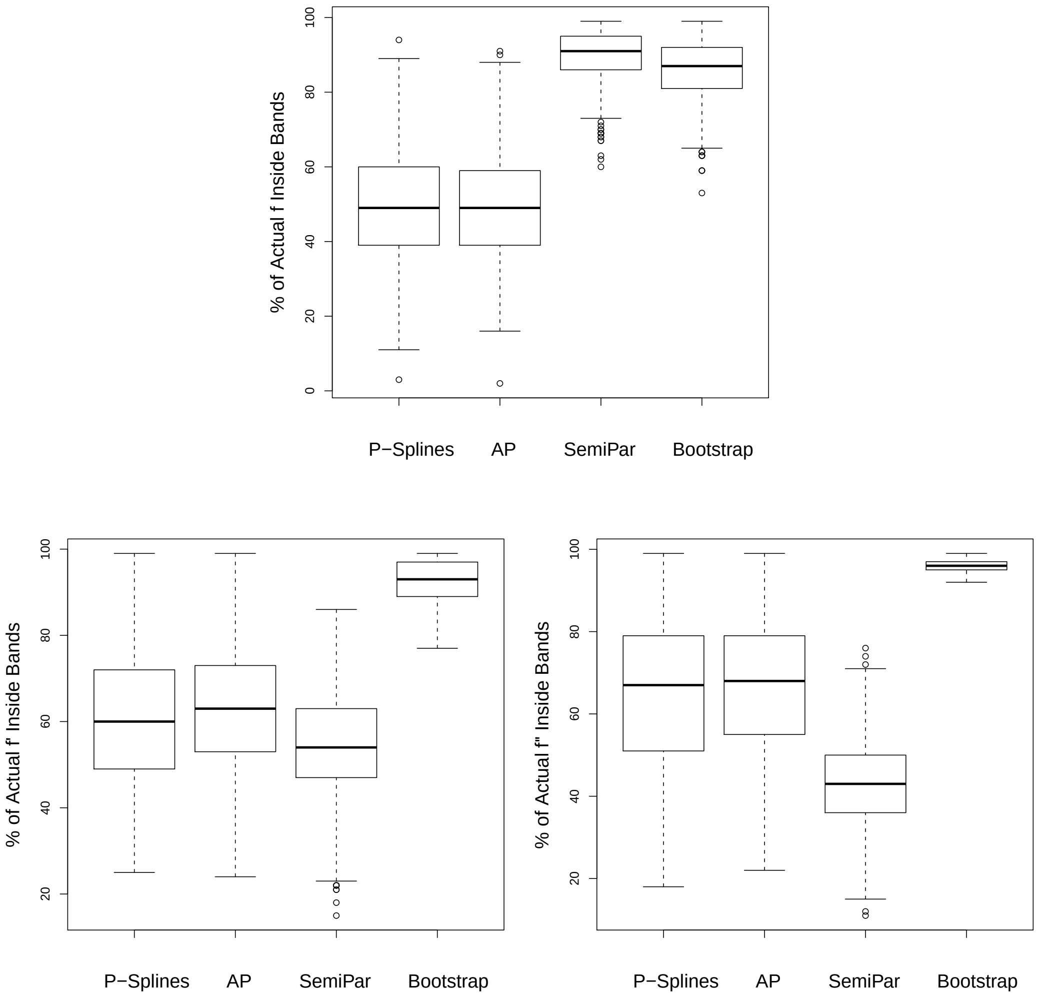


Figure 6.7: Percentage of f (top) and f' and f'' (bottom) contained within variability bands

Spline, AP and bootstrap bands significantly improve coverage, the semiparametric regression variability band coverage rapidly declines. For the first derivative estimates, the semiparametric bands perform the worst among the four methods. The bootstrap pointwise bands prove superior in capturing both f' and f'', whilst maintaining the lowest variability in coverage percentage. The AP and P-Spline bands offer similar results, with the AP method slightly ahead in mean and standard deviation of coverage for each of f, f' and f''.

The most striking observation from Table 6.1 is not any comparison between methods for calculating bands, but the actual coverage each technique offers. For derivative estimates each of the P-Spline, AP and semiparametric bands are achieving between 40% and 70% coverage of the actual derivatives. One should clearly tread carefully when making inferences about a population parameter based on these bands as they do not offer the classical interpretation of a confidence interval. Ruppert, Wand & Carroll [48] describe variability bands as acting as pointwise 95% confidence intervals at each x_i for $i = 1, \ldots, n$ observations assuming that $\hat{f}(x_i)$ is approximately normally distributed. For inferences about the entire curve, global or simultaneous variability bands are necessary.

The percentages in Tables 6.2, 6.3 and 6.4 refer to the proportion of inclusion of $f(x_i)$, $f'(x_i)$ and $f''(x_i)$ for arbitrarily chosen observations $i = 25, 50, 75$. These pointwise coverage measures should all be roughly to 95% given the definition of a variability band from Section 6.1.

	$f(x_{25})$	$f(x_{50})$	$f(x_{75})$
P-Spline	80.3	75.6	78.9
AP	81.4	72.8	79.1
SemiPar	93.2	91.8	92.7
Bootstrap	99	87	100

Table 6.2: Pointwise coverage percentage of $f(x_i)$ for $i = 25, 50, 75$.

The bootstrap variability bands are excellent in terms of pointwise coverage of the actual function f, with very few simulated variability bands missing the chosen observations. The three observations are captured by the AP and P-Spline variability bands 75 to 80% of the time for the function f. Therefore, these bands are unlikely to be 95% pointwise bands for the function f which would lead one to question the estimate of σ^2 from (6.14). Gasser [16] proposes a nonparametric estimate for the residual variance in nonlinear regression which could improve the performance of these variability bands in capturing the function f. The semiparametric variability bands achieve better pointwise coverage for the function f when compared to the AP and P-Spline approaches, with over 90% coverage. This is still slightly below the nominal 95% and perhaps a different estimate for the residual variance could be used.

	$f'(x_{25})$	$f'(x_{50})$	$f'(x_{75})$
P-Spline	89.2	87.2	91.6
AP	92.4	88.6	92.8
SemiPar	89.0	80.6	87.6
Bootstrap	100	93	100

Table 6.3: Pointwise coverage percentage of $f'(x_i)$ for $i = 25, 50, 75$.

Similarly to global coverage the bootstrap bands are vastly superior in terms of pointwise coverage for derivative estimation. For $i = 25, 75$, all variability bands captured the actual

	$f''(x_{25})$	$f''(x_{50})$	$f''(x_{75})$
P-Spline	90.9	85.8	90.1
AP	92.8	86.1	92.8
SemiPar	75.3	71.4	77.6
Bootstrap	100	97	100

Table 6.4: Pointwise coverage percentage of $f''(x_i)$ for $i = 25, 50, 75$.

pointwise first and second derivatives. This is not a good result, however, since the method for estimating these bands should offer 95% coverage. The bootstrap variability bands come closer to 95% coverage of $f'(x_{50})$ and $f''(x_{50})$ across the simulations here. The boundary effects, discussed earlier, could well be responsible for the over coverage when estimating the standard error closer to the tails. The AP and P-Spline bands underperform slightly in terms of derivative estimates, with roughly 90% pointwise coverage being achieved for both first and second derivative estimates. The pointwise coverage of the semiparametric variability bands deteriorates with each order of derivate taken. For first derivative estimates the pointwise coverage is still respectable, at between 80 and 90%. However, it dips below 80% for second derivative estimates, which is a serious concern for purposes of interpretation. Hence, there is empirical evidence here that the AP and P-Spline variability bands are superior to the recognised semiparametric bands in pointwise coverage of derivative estimates.

Moving to precision, Figure 6.8 and Table 6.5 present a summary of variability band width (not to be confused with bandwidth). Here, low width (i.e. high precision) is favourable as long as coverage percentage is maintained.

	f	f'	f''
P-Spline	0.10 (0.04)	1.07 (0.55)	12.7 (9.18)
AP	0.09 (0.02)	0.95 (0.29)	10.7 (4.31)
SemiPar	0.26 (0.02)	0.83 (0.08)	6.26 (0.73)
Bootstrap	0.54 (0.05)	5.02 (1.14)	80.2 (32.7)

Table 6.5: Mean (standard deviation) width of f, f' and f''.

The performance in coverage percentage for the bootstrap pointwise bands is obviously explained by the huge width of these bands, especially as attention shifts to derivative estimates. For estimates of f and f' the bootstrap bands are several times wider than those offered by P-Splines, the AP or semiparametric smoothing. In estimating bands around f'' the width of the bootstrap bands is much higher when compared to the other three, being over ten times

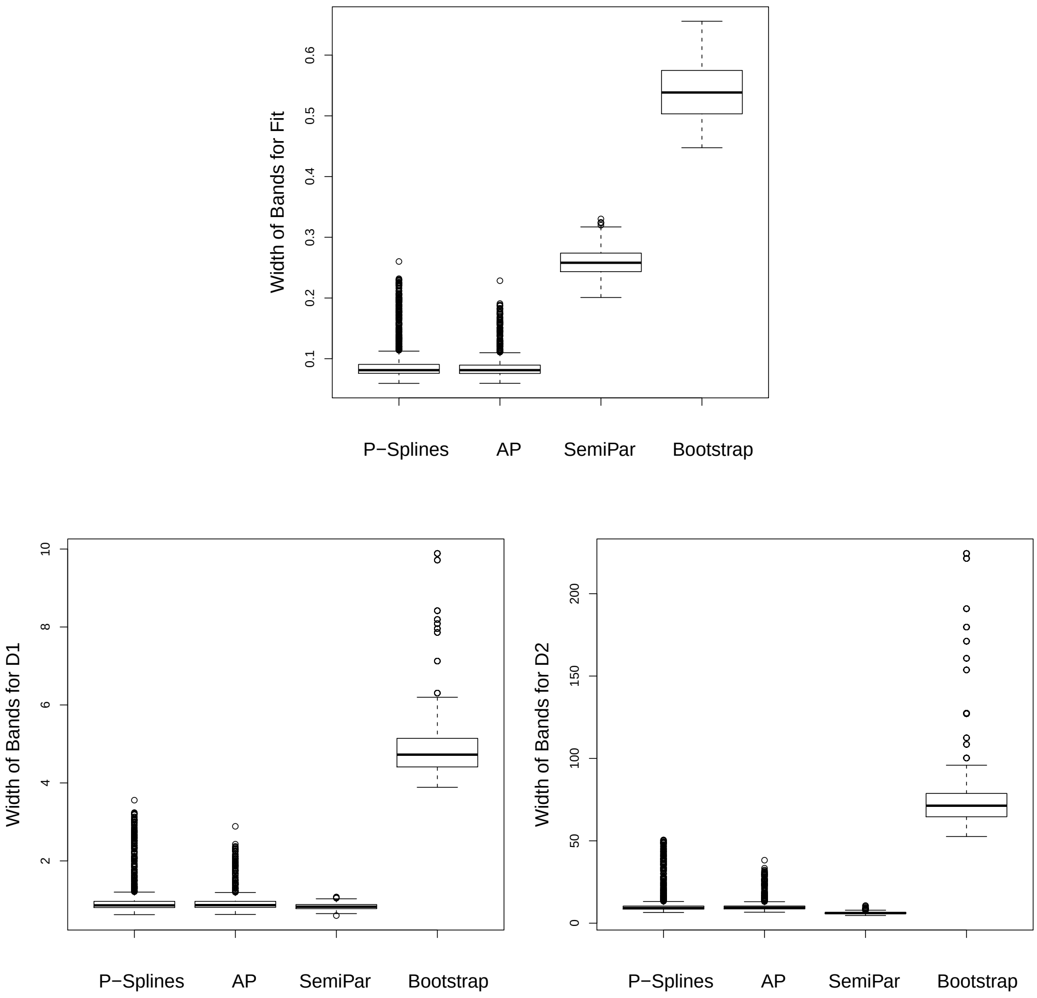


Figure 6.8: Width of variability bands for estimates of f (top), f' and f'' (bottom).

wider on average.

Similarly to coverage percentage, the AP and P-Spline bands are very similar in terms of width, with the P-Spline bands slightly wider on average and the AP bands more stable across the 1000 simulations from f. In estimating variability bands for the true function f the AP and P-Spline bands offer width three times lower than the semiparametric approach. This explains the results in Table 6.1 where the semiparametric smoothing variability bands outperformed those from the AP and P-Spline procedures.

Concentrating on derivative estimation, the three methods offer similar widths and, recalling Table 6.1, similar coverage percentages. To reiterate, the bootstrap bands here are roughly eight times wider than the other three bands on average. The variability of precision (as is evident by the mean to standard deviation ratio) seems to increase as one moves to higher order derivatives which is logical due to the sensitive nature of high order derivatives. One worrying factor discovered through viewing Figure 6.8 is the outliers in width of variability bands found using the AP and P-Splines. The semiparametric model does not show signs of this and offers far more stable variability bands. The bands constructed through the AP and P-Splines are clearly more susceptible to outliers.

Another measurement obtained from these simulations was the number of departures of f, f' and f'' from estimated variability bands. Ideally zero departures would be evident, although zero departures could mean that the function never enters the bands! This ambiguity is a facet of measuring coverage in this way. A low number of departures could mean that, say f', leaves the bands for most of its course and returns near the end. This is clearly not good performance and the simulations test for such a potential loophole in results (using a minimal coverage indicator). However, observing Figure 6.9 and Table 6.6 the results here concur with those found for coverage percentages in Table 6.1. The bootstrap bands offer excellent performance for estimates of f' and f'' and here prove superior in terms of the number of departures of f' and f''. The imprecise nature of the bootstrap bands has been shown to be the reason for this.

	f	f'	f''
P-Spline	4.8 (1.8)	5.4 (2.5)	5.4 (2.8)
AP	4.5 (1.4)	5.1 (2.0)	5.4 (2.3)
SemiPar	1.5 (0.8)	5.8 (1.1)	6.7 (1.3)
Bootstrap	2.8 (1.1)	2.3 (0.9)	1.8 (0.8)

Table 6.6: Mean (standard deviation) departures of f, f' and f'' from bands.

P-Splines and the AP offer a similar number of departures in the $n = 100$ observations, with f, f' and f'' leaving the bands on roughly five separate occasions. The semiparametric approach is comparable with the AP and P-Splines in departures of the actual first and second derivative of f whilst performing better in terms of departures of f itself.

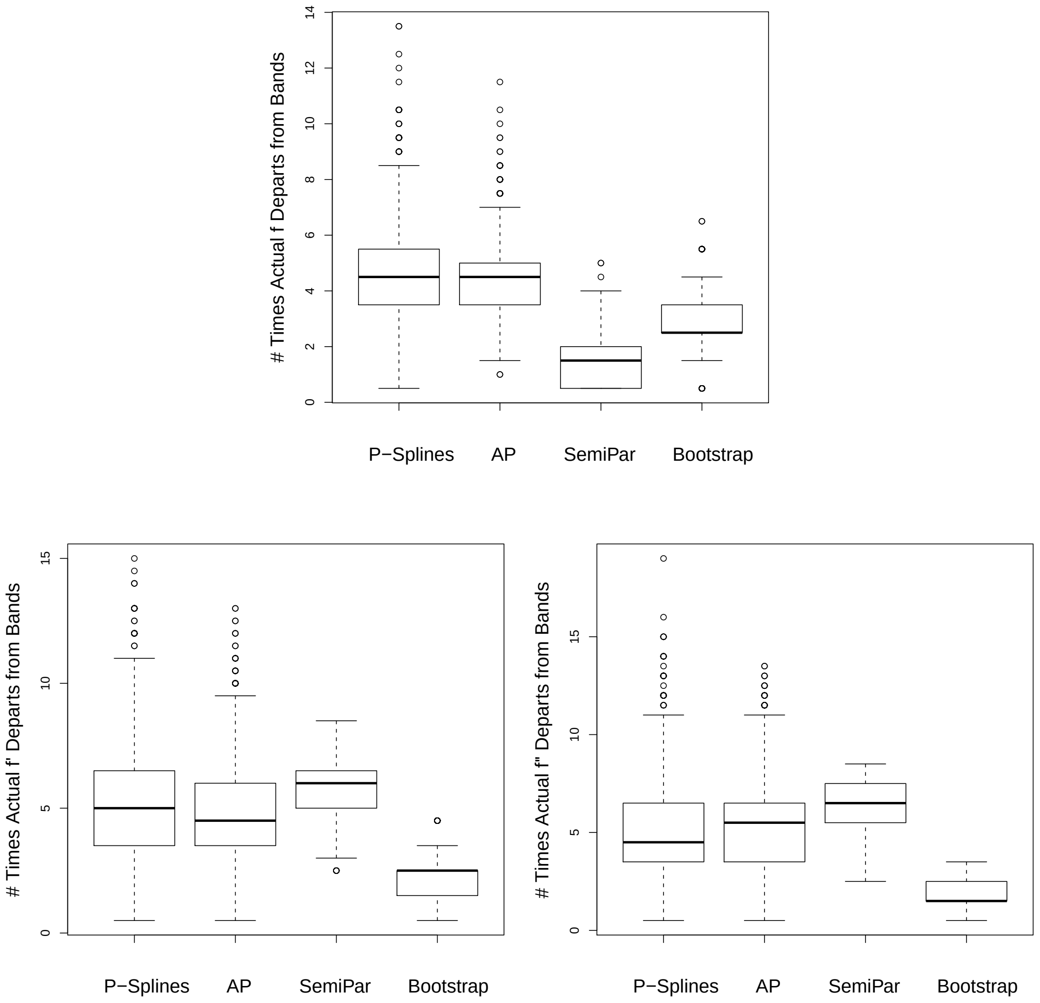


Figure 6.9: Number of departures in 100 observations of f (top), f' and f'' (bottom) from variability bands.

6.1.5.3 Discussion of Variability Band Performance

The performance of the four variability band estimators has been examined from an empirical perspective in terms of coverage, precision and number of departures. Figures 6.10, 6.11 and 6.12 display estimates and variability bands for f, f' and f'' by each of the four approaches. The superior coverage performance of the bootstrap bands has been discovered to be caused by large band width which offers little information about estimates of the function or its derivatives. The bottom left panels of Figures 6.10, 6.11 and 6.12 reaffirm that the width of these bands is much larger than in the other methods, although bands around f by semiparametric and bootstrap techniques are comparable.

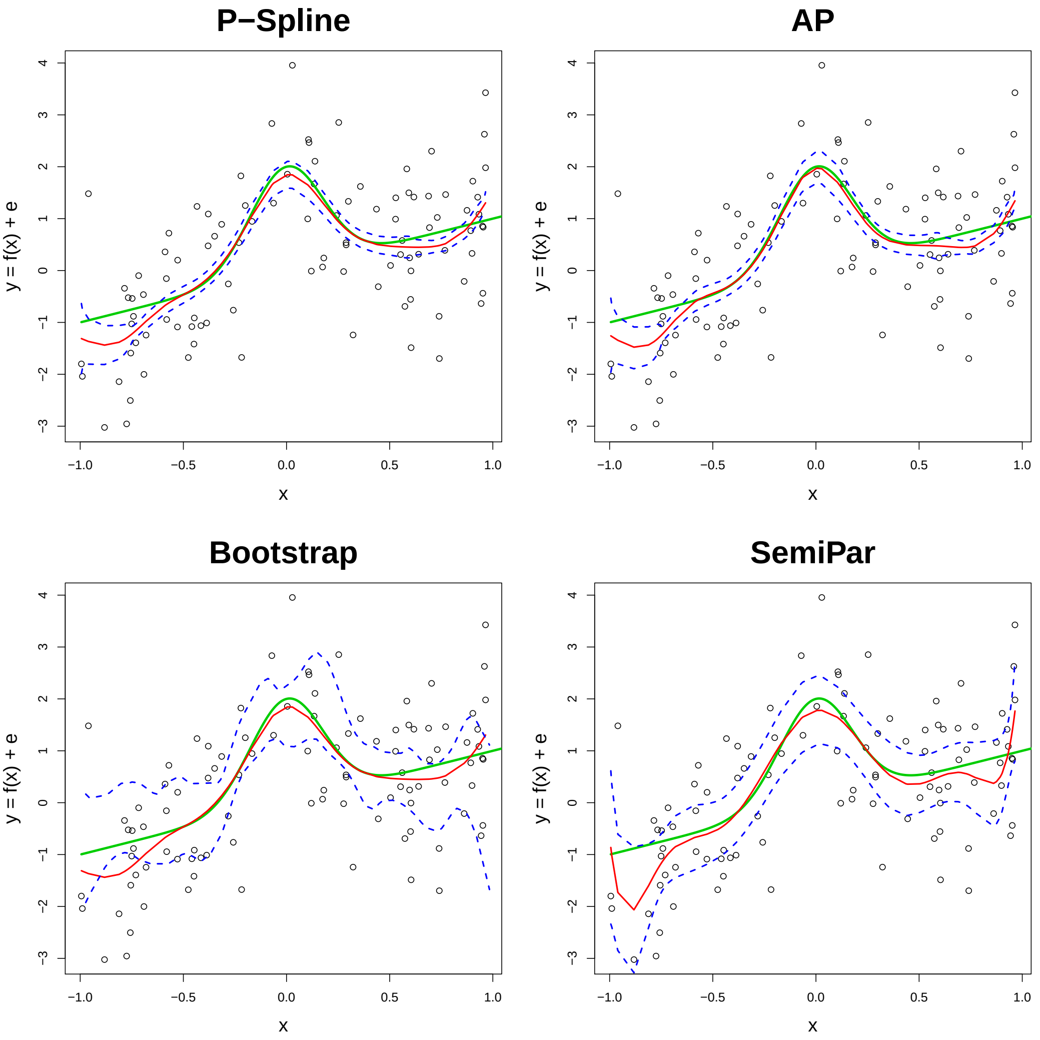


Figure 6.10: Actual f (green) with estimates of f (red) and variability bands (blue).

There is little difference between the P-Spline and AP variability bands in terms of coverage percentage from Table 6.1 and this can be seen clearly across estimates of f, f' and f''.

The semiparametric regression model has comparably wide variability bands for estimates of f and this leads to a good coverage percentage (Table 6.1). Moving to derivative estimates it is apparent that the poor coverage is caused by the substandard estimates of f' and f'' found using semiparametric smoothing and not by the variability bands themselves. Whereas the AP and P-Spline derivative estimates are reasonably accurate, the semiparametric regression derivative estimates are not. The bootstrap pointwise bands are excessively wide and offer limited information about the likely error of the estimate of f''.

Figure 6.13 gives a scatterplot of global coverage against precision of bands for f_2 and its

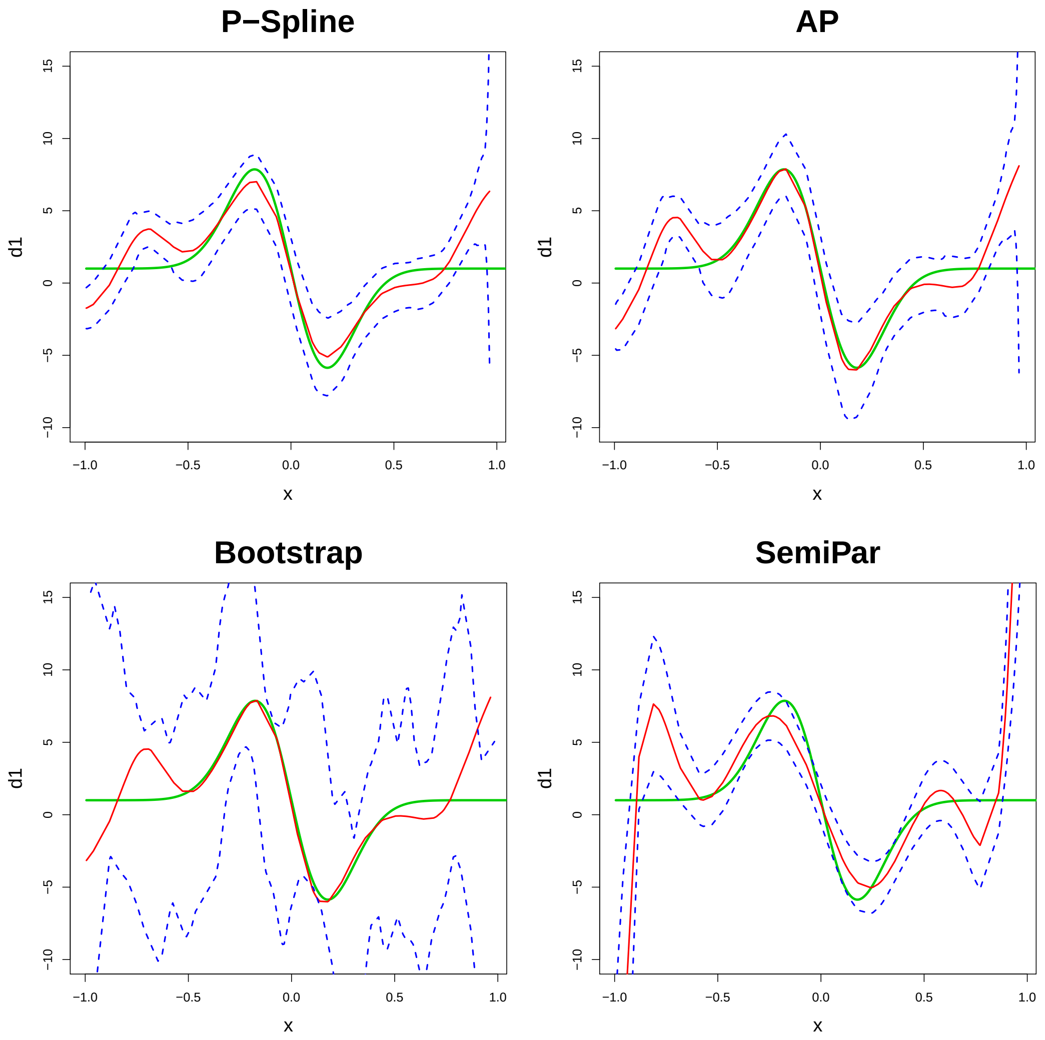


Figure 6.11: Actual f' (green) with estimates of f' (red) and variability bands (blue)

derivatives for each of the four methods. In derivative estimation the variability of width is larger for the P-spline bands than the semiparametric or AP but the reverse argument seems to hold for coverage. The bootstrap bands have excellent coverage but this is evidently due to a lack of precision.

6.1.5.4 Zero Slope Test

Another, more subtle, approach to measuring variability band performance is presented here through a test of zero slope.

Data from a function $f_c = 4$ with some error, $\epsilon \sim N(0, 1)$, added was simulated 1000 times and smoothed using each of the three smoothing techniques. Since $f'_c = 0$, each of

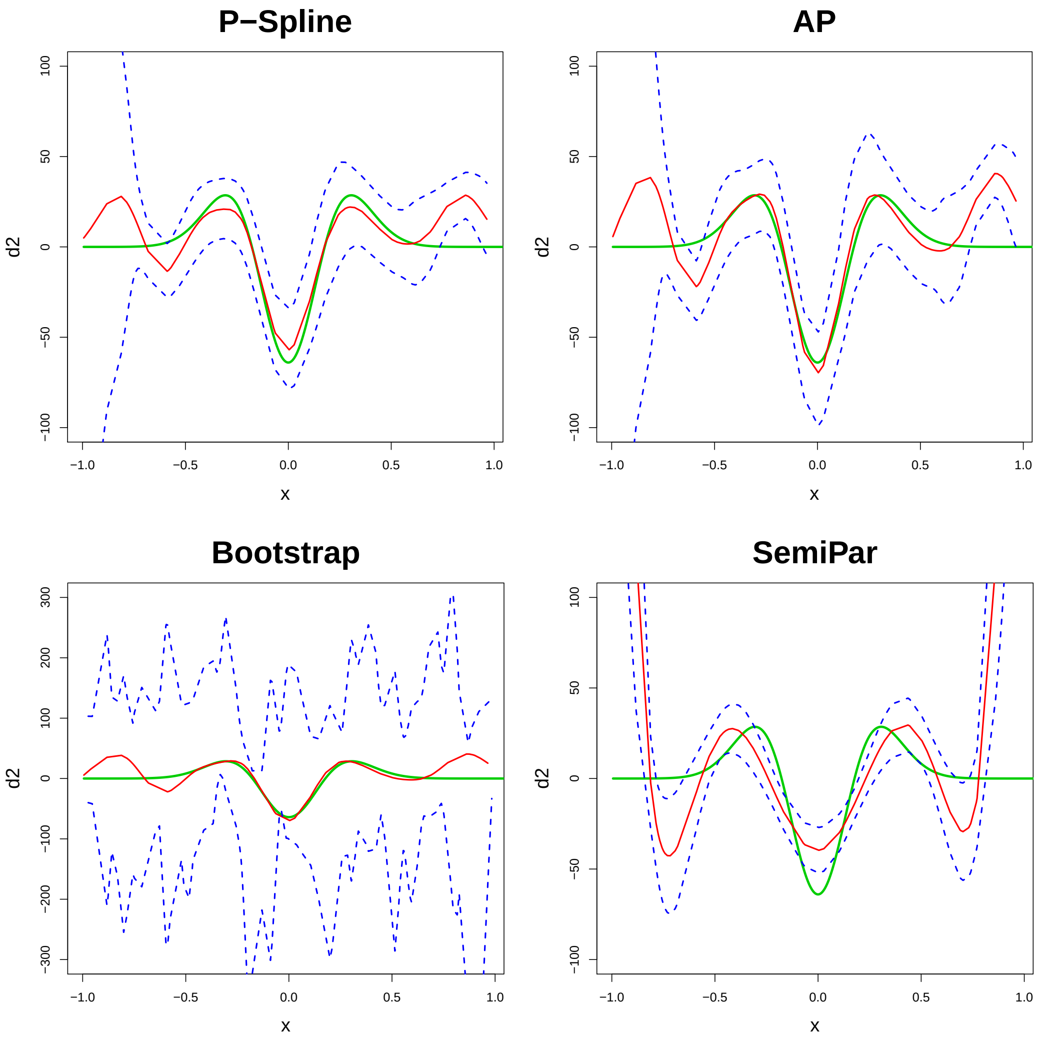


Figure 6.12: Actual f'' (green) with estimates of f'' (red) and variability bands (blue).

the derivative estimation techniques should find a flat and linear first derivative estimate at 0. Moreover, variability bands placed around these estimates should contain 0 for each x_i in $\hat{f}'_c(x_i) \pm 2se[\hat{f}'_c(x_i)]$, $i = 1, \ldots, n$. Each of the four variability band estimators were tested on this and the results are detailed here.

A large percentage of the zero line contained within the variability bands corresponds to good performance for this test. Figure 6.14 and Table 6.7 summarise the findings of the slope zero test.

The bootstrap bands performed perfectly by including the $n = 100$ observations on the line $y' = 0$ for 1000 simulations but, as has been discussed, has far wider bands than the other three approaches. The P-Spline and AP variability bands are once again comparable with the

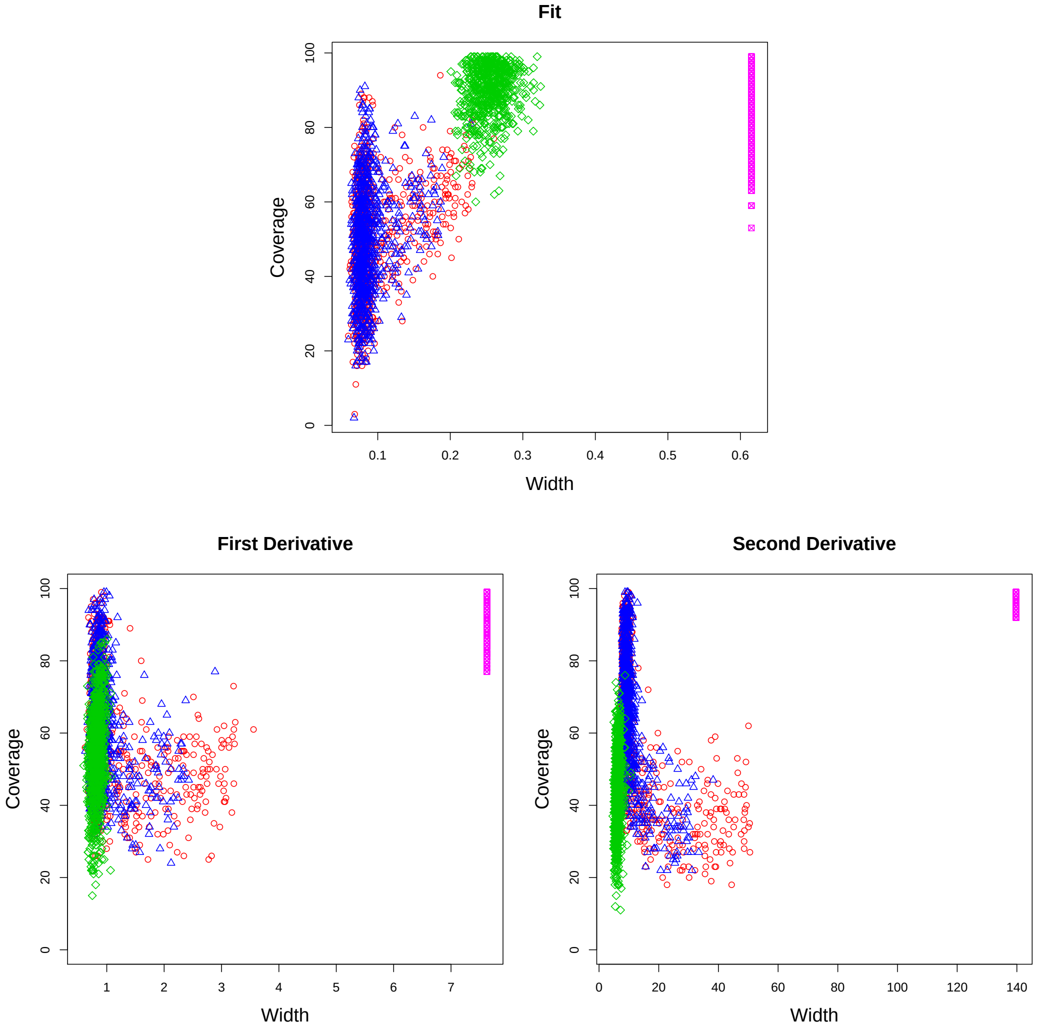


Figure 6.13: Coverage versus precision for the P-Splines (red), AP (blue), semiparametric (green) and bootstrap (purple) variability band estimators for the estimate of the function f_2 and its derivatives.

	% Coverage
P-Spline	95.1 (5.9)
AP	94.9 (6.8)
SemiPar	92.5 (12.2)
Bootstrap	100 (0)

Table 6.7: Mean Coverage Percentage of f_c' (i.e. $y = 0$) with standard deviation.

P-Spline bands slightly better in mean and standard deviation of coverage of the zero line. The semiparametric model is outperformed by both on each count and this agrees with findings for coverage percentages of f' from Table 6.1.

Figure 6.14 provides yet more evidence that variability bands found using the AP and P-

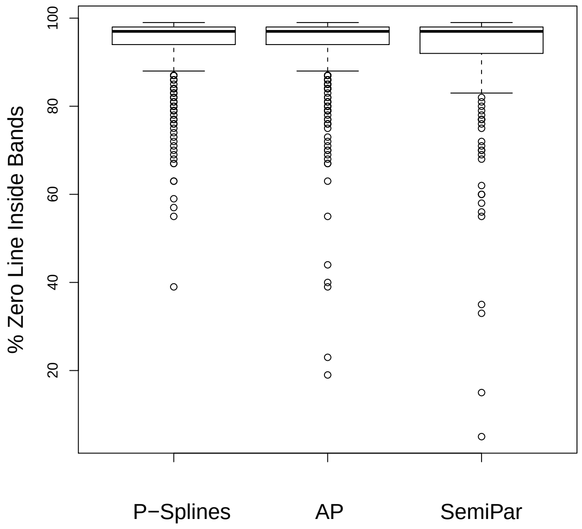


Figure 6.14: Percentage of zero line contained within variability bands.

Splines are not stable. Moreover, in terms of capturing the line $y' = 0$, the semiparametric method displays this instability. Sensitivity to large values of ϵ_i is problematic when estimating variability bands for derivative estimates.

Now, it is clear that 95% global or pointwise coverage is not achieved by any of the modelling based variability band estimators for any of f, f' or f''. Thus these bands are not 95% confidence bands in the sense of a 95% confidence interval. They are merely a representation of the likely error contained when using these smoothing or derivative estimation techniques.

6.1.6 Application of Variability Bands to the Motivating Illustrations

Variability bands for first derivative estimates allow significant changes in a response variable relative to an explanatory to be identified by comparison with $y' = 0$. The Winter Nutrients and Scottish Bird Count datasets require this type of analysis and in the following sections the four variability band estimators are compared in both illustrations.

6.1.6.1 Winter Nutrients Data

The Winter Nutrients data contain 131 salinity adjusted measurements of NTRZ (a nitrate and nitrite compound) and Phosphate. The main aim is to determine periods of time in which levels of the nutrients are increasing or decreasing. Here these times are estimated using first derivative estimates and corresponding variability bands.

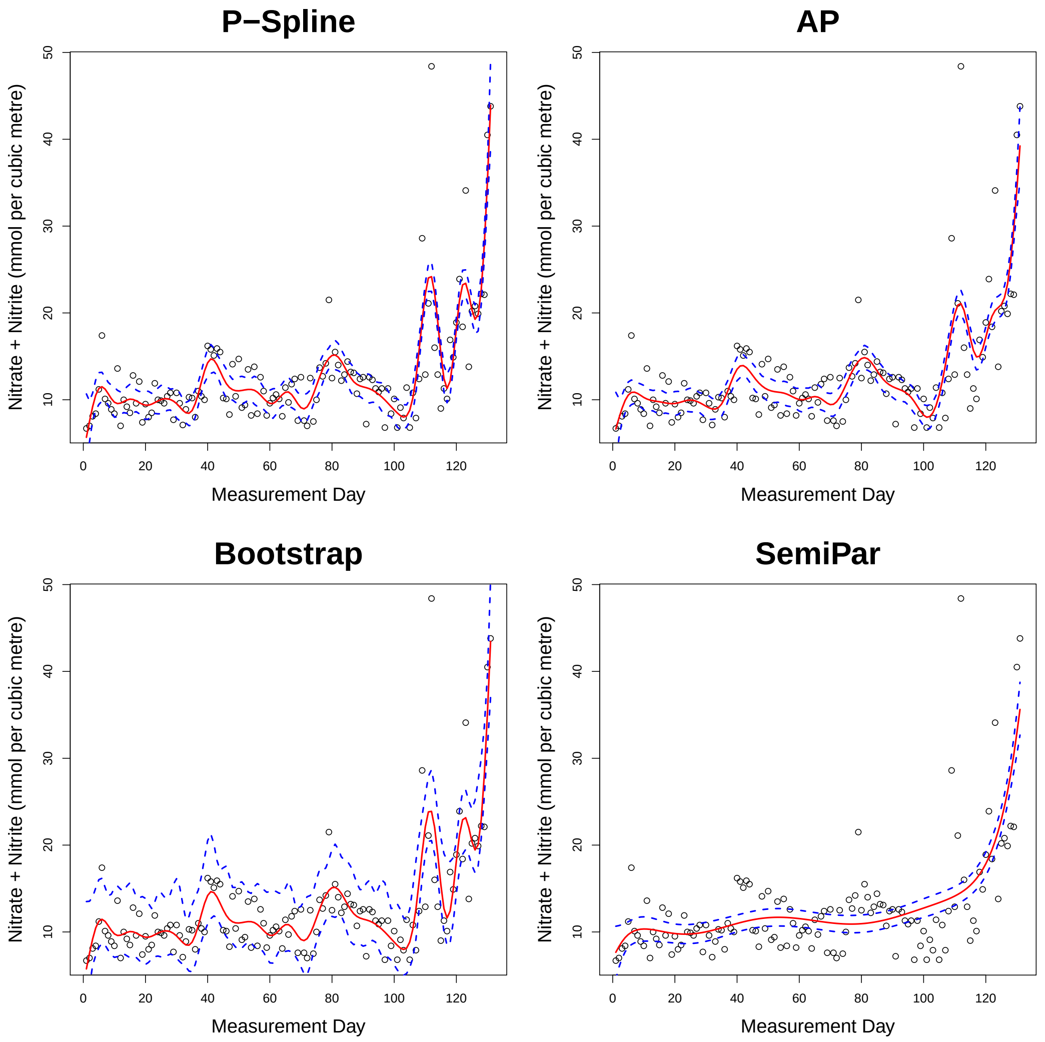


Figure 6.15: Smooth fits of NTRZ with variability bands.

In Figure 6.15 smooth fits to the NTRZ data are displayed. Once more the AP method is slightly more smooth than P-Splines owing to the extra smoothing from the additional penalty term, however the semiparametric fit is the smoothest of all. The bootstrap bands are found by resampling the residuals of the P-Spline smooth to the NTRZ data since the extra smoothing included in the AP model leads to an oversmooth fit to the data. The bootstrap bands for the first derivative estimates are obtained using the AP derivative estimation technique on the 1000 resamples found using the residuals of the P-Spline fit.

First derivative estimates are presented in Figure 6.16. The most striking observation from

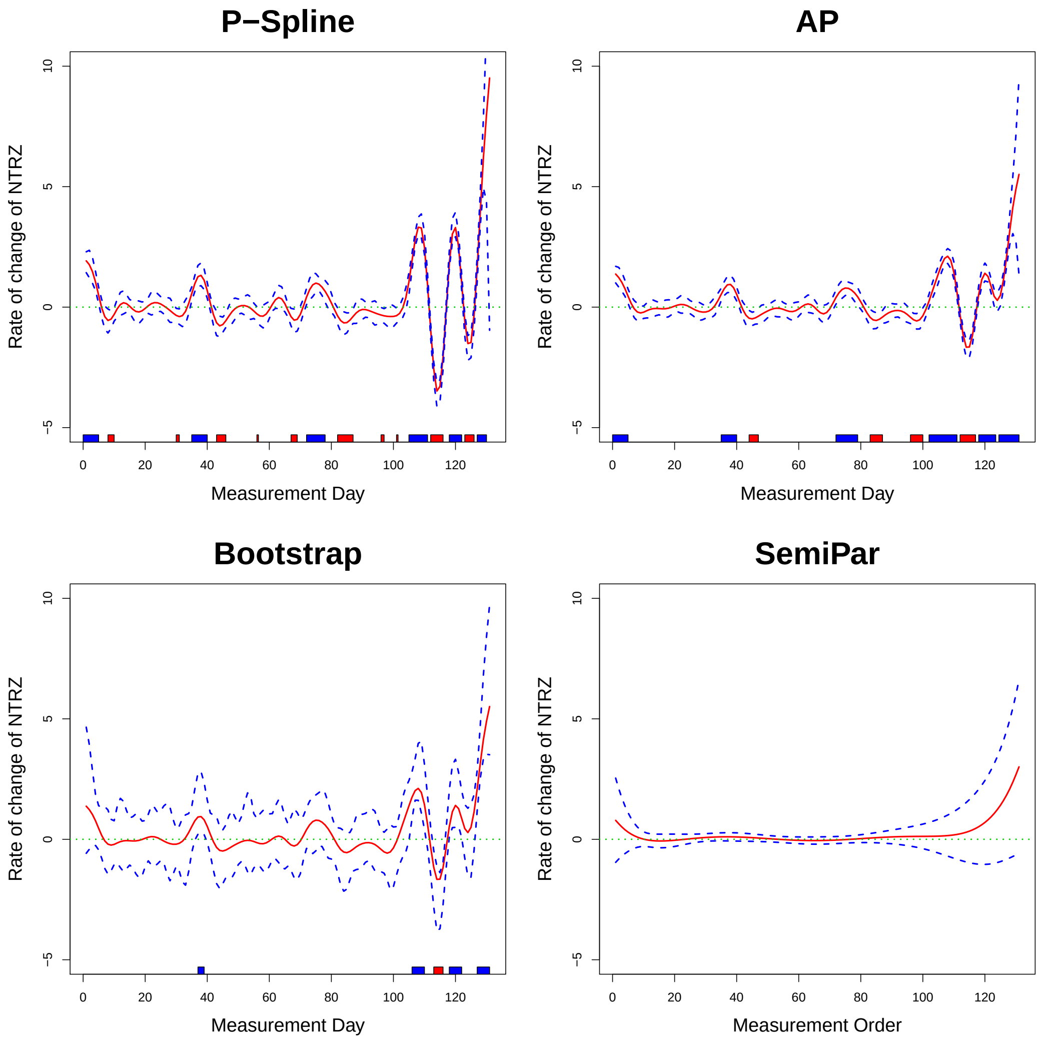


Figure 6.16: First derivative estimates of NTRZ with variability bands and highlighted regions significant decrease (red) and significant increase (blue).

these estimates is how flat and smooth the semiparametric derivative estimate is compared to the others. If this is to be believed there is no time during 1990/1991 at which NTRZ is significantly increasing or decreasing!

The other three plots offer an alternative view of events during this Winter. The AP and *P*-Spline bands appear to pick up similar regions of increase and decrease, with much fluctuation towards the right tail. The difference in estimated increasing/decreasing times seems to come from the end of the Winter, where the *P*-Spline fit is more erratic than the AP. Both display an almost sinusoidal effect, however the AP first derivative estimate barely crosses zero in the second trough (circa day 125) whereas the *P*-Spline derivative, together with its bands, dips considerably below zero. This agrees with all the previous discussions on comparisons of these two methods, where the AP fit is less volatile in derivative estimates due to the extra smoothing term.

The bootstrap bands, being the widest of all, fail to identify as many regions of significant change although any regions where these bands find significance are more likely to be true in the population. This could be thought of as analogous to a higher percentage confidence interval for a population parameter. If small Type I error is of primary importance to the researchers then using the bootstrap bands would be the recommended choice. In general it appears that there are more decreasing days than increasing, and that most of the time there is no change.

Tables 6.8 and 6.9 summarise the regions of significant change throughout 1990/1991 for each of the four methods. The semiparametric bands cannot recognise any regions of significant change owing to the severely oversmooth fit and subsequent derivative estimate. From Table 6.8 the only common area of significant decrease found by the other three method comes between days 113 and 116, this being the only such section determined by the bootstrap bands. It is evident that the P-Spline method determines many more periods of decrease than the AP. However, there are some sections of decline found by the AP and not P-Splines (days 47, 98-100 and 117). These days reflect the AP bands staying below zero for longer due to the extra smoothness inherent in the model.

	Periods of Decrease (Total)
P-Spline	8-10, 30-31, 43-46, 56, 67-69, 82-87, 96-97, 101, 112-116, 123-126 (31)
AP	44-47, 83-87, 96-100, 112-117 (20)
SemiPar	None identified
Bootstrap	113-116 (4)

Table 6.8: Estimated regions of significant decrease in NTRZ.

	Periods of Increase (Total)
P-Spline	1-5, 35-40, 72-78, 105-111, 118-122, 127-130 (34)
AP	1-5, 35-40, 72-79, 102-111, 118-123, 125-131 (42)
SemiPar	None identified
Bootstrap	37-39 106-110 118-122 127-131 (18)

Table 6.9: Estimated regions of significant increase in NTRZ.

Table 6.9 shows many similarities between periods of significant increase found by the P-Spline and AP. The bootstrap bands again find less regions due to their width and thus give more conservative estimates of regions of significant change.

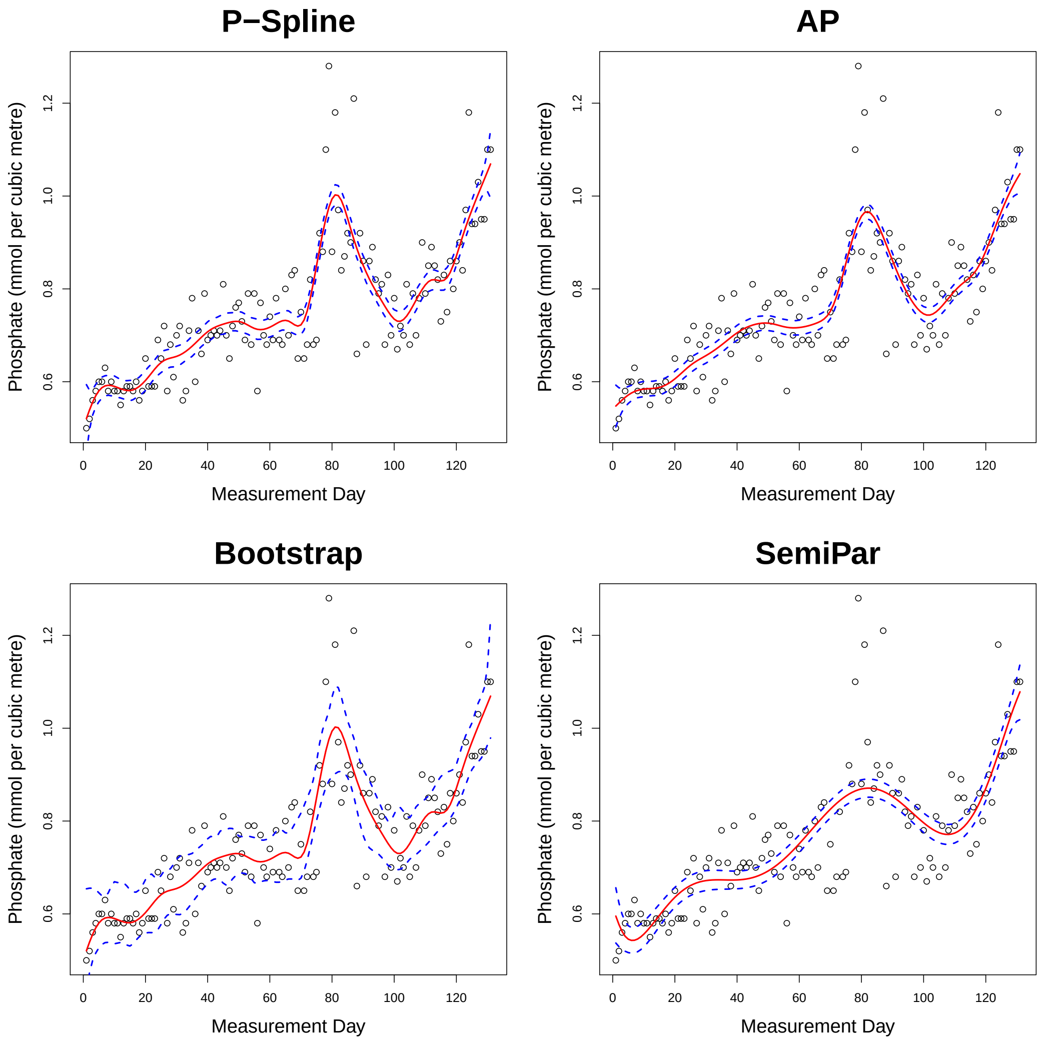


Figure 6.17: Smooth fits of Phosphate with variability bands.

Unsurprisingly, similar comparisons between methods can be drawn for the Phosphate data. Figure 6.17 reveals that the AP fit is smoother than that using P-Splines. The semiparametric fit is the smoothest of all and that the bootstrap pointwise bands are the widest. In Figure 6.18 this leads to more areas of significant change being identified by the AP and P-Spline approaches and, for instance, no periods of significant decrease being found by the semiparametric bands. Both large band width (bootstrap) and severe oversmoothing (semiparametric) lead to low power in finding significant change. For example, 102 days were found to display no significant contamination change using the bootstrap pointwise bands compared to 40 found by the AP bands. The AP derivative estimate detects more areas of significant increase than the P-Spline estimate due to more volatility in the P-Spline fit between days 20 and 40 (the AP derivative estimate together with its corresponding variability bands remain above zero for nearly all this period).

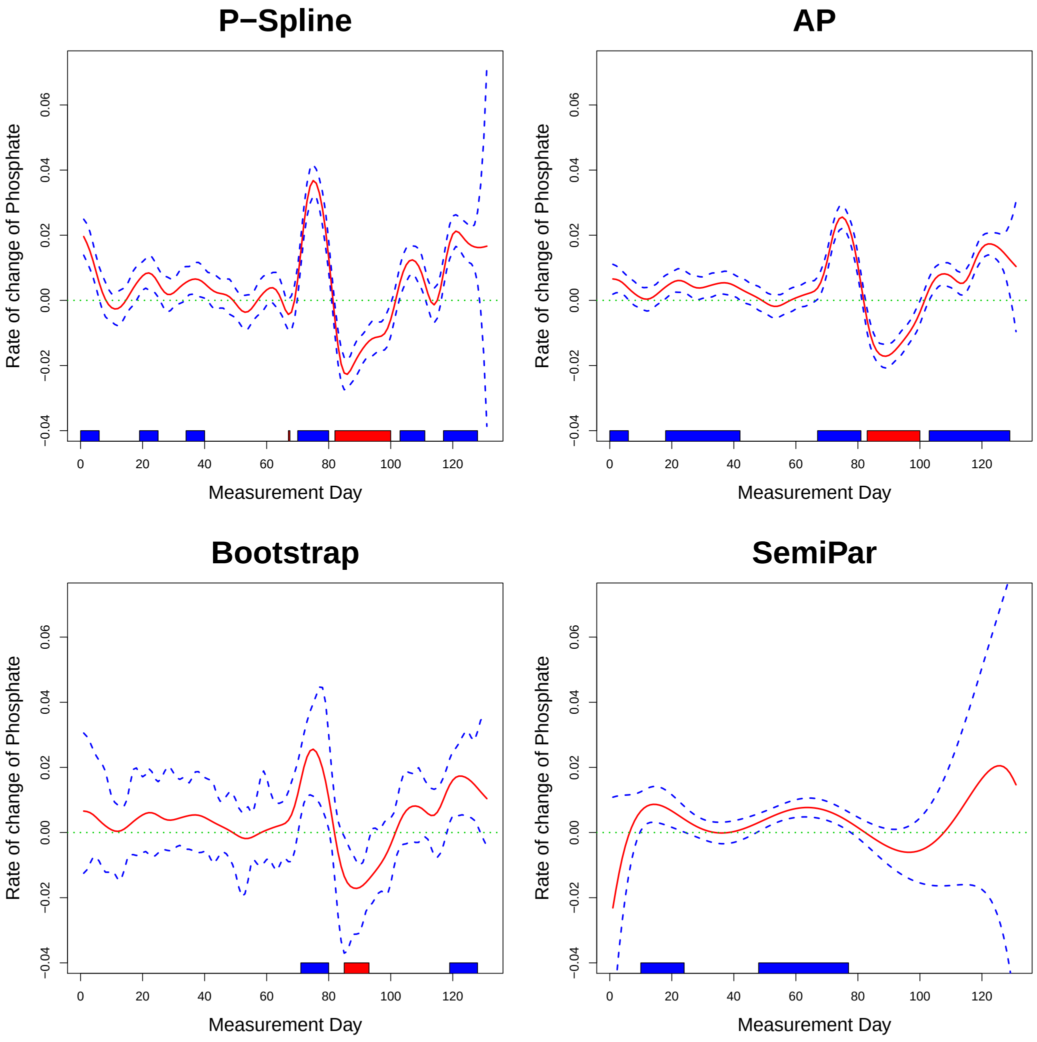


Figure 6.18: First derivative estimates of Phosphate with variability bands and highlighted regions significant decrease (red) and significant increase (blue).

Table 6.10 shows one common region of significant decrease determined by the P-Spline, AP and bootstrap methods (days 85-93). Evidence that the AP derivative estimate is more accurate than that found using P-Splines has been found (Section 5.5.1). Thus one would tend to believe these bands more given the similar performance of variability bands between the two (Section 6.1.5.2). Again if the researcher is interested in low Type I then 85-93 is an advisable estimate, otherwise one would report days 83-100 as having significant decrease.

	Periods of Decrease (Total)
P-Spline	67, 82-100 (20)
AP	83-100 (18)
SemiPar	None identified
Bootstrap	85-93 (9)

Table 6.10: Estimated regions of significant decrease in Phosphate.

Table 6.11 again shows marked similarities in regions of increase found by the P-Spline and AP techniques. For the first time the semiparametric bands manage to distinguish a period of significance and so an area of increase common to all four methods is between days 71 and 77. Even for the overtly careful researcher this represents an abundance of evidence that this feature exists in the population!

	Periods of Increase (Total)
P-Spline	1-6, 19-25, 34-40, 70-80, 103-111, 117-128 (52)
AP	1-6, 18-42, 67-81, 103-129 (73)
SemiPar	10-24, 48-77 (45)
Bootstrap	71-80, 119-128 (20)

Table 6.11: Estimated regions of significant increase in Phosphate.

6.1.6.2 Scottish Bird Count Data

Recall the Scottish Bird Count data consists of counts of Grey Plover (amongst other wetland bird species) measured annually for 31 years (1974 - 2004). The bottom left panel of Figure 6.19 is a P-Spline smooth of the Grey Plover data with pointwise bootstrap bands. Bootstrapping residuals from a generalised P-Spline smooth was used here.

The generalised AP and semiparametric fits are very similar in both the smooth estimates to the data and their variability bands. The bootstrap bands are wider than any of the other three, which is intuitive from the method by which these bands are calculated and agrees with simulation findings from Section 6.1.5.2. The AP smooth is much smoother than the P-Spline fit, which again agrees with the discussions from Chapter 5.

Figure 6.20 exhibits first derivative estimates from the Grey Plover data using the P-Spline, AP and semiparametric methods. Pointwise bootstrap bands of the AP first derivative estimate are shown in the bottom left panel of Figure 6.20.

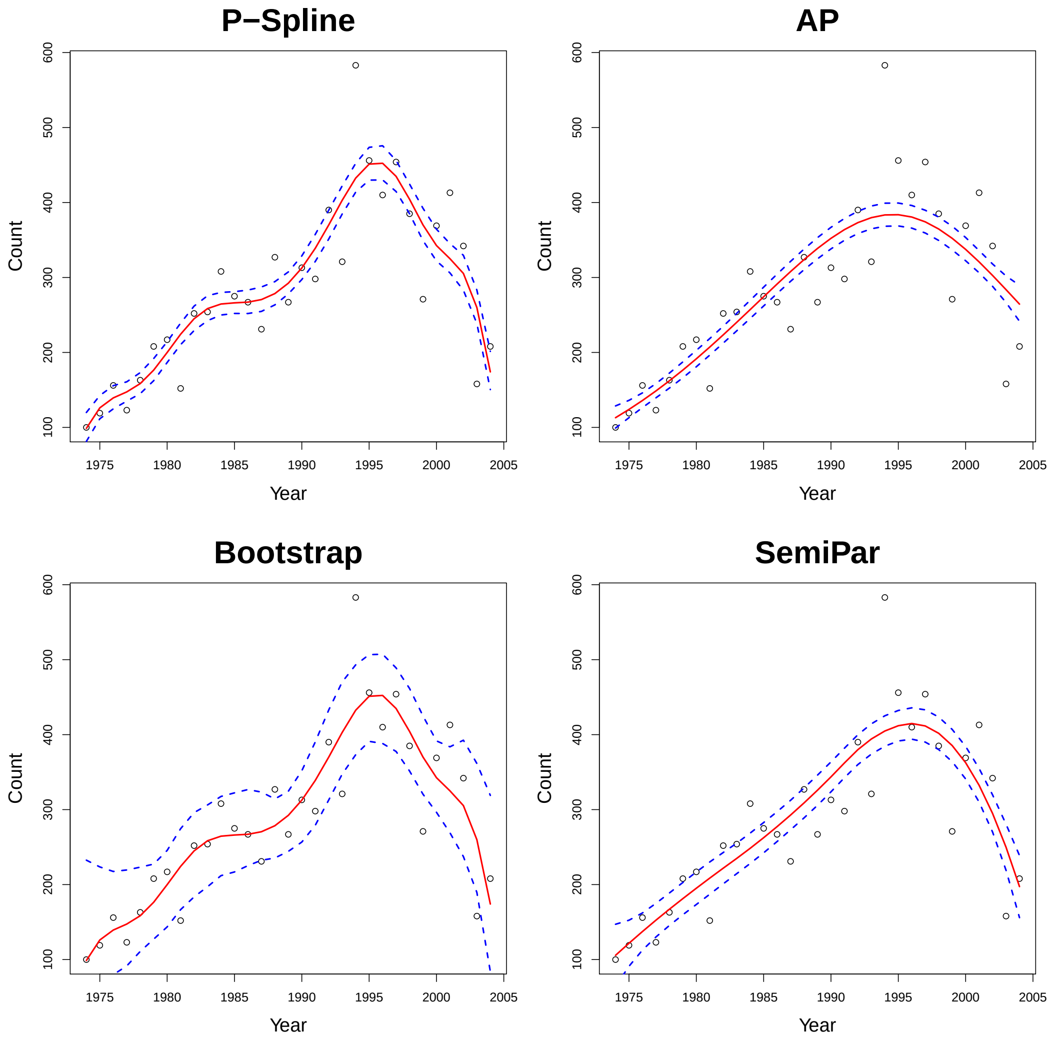


Figure 6.19: Smooth estimates of Grey Plover with variability bands. The semiparametric model is the only method which does not recognise that the response consists of counts.

The question of interest here is to investigate whether a significant decrease in count is evident and to estimate the time at which this decrease occurs. A first derivative estimate with fully positive variability bands represents evidence of an increasing count; a first derivative estimate with fully negative bands implies a decreasing count and if zero is contained inside the bands the count is neither significantly increasing or decreasing.

Each of the four estimates displayed here have similar features, with an increasing count up to the late 1990's followed by an eventual shift to a decreasing count somewhere around the turn of the millennium. Thus, each method agrees that a significant decrease in Grey Plover has occurred. The year where this decrease occurs is estimated by the point at which the upper

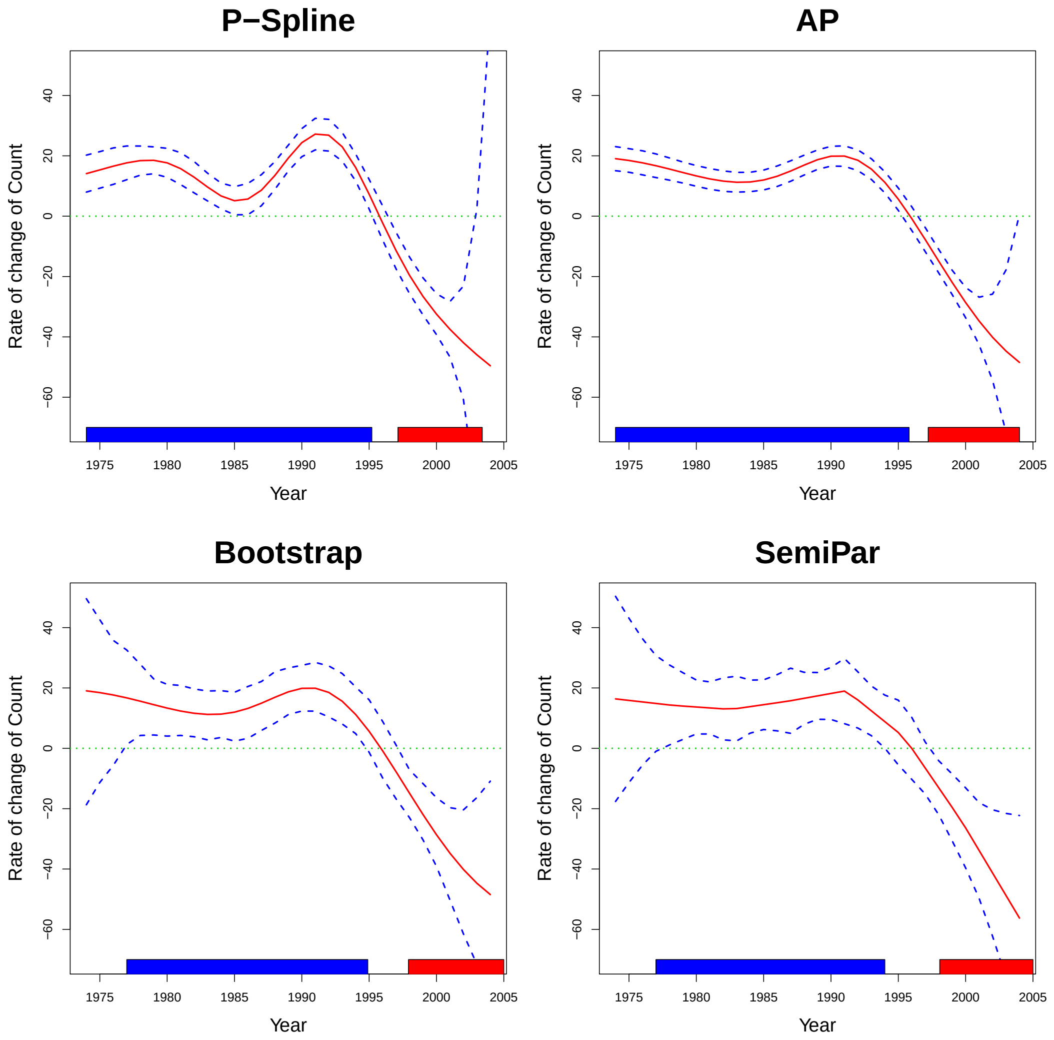


Figure 6.20: First derivative estimates of Grey Plover with variability bands and regions of significant increase (blue) and decrease (red) shown.

variability band crosses zero such that both bands are negative from (6.3). This procedure presents a more useful estimate to that found by merely observing the time at which the first derivative estimate itself crosses the zero line. Interestingly, the semiparametric and bootstrap bands share comparable precision, which opposes simulation findings where the bootstrap bands were by far the widest. The simple structure evident in this example could be responsible for this finding.

Table 6.12 gives four estimates for the time (in years) where a significant decrease in Grey Plover was identified. All four give similar results, the AP and P-Spline methods give evidence that a significant decrease in count began close to 1997 and the semiparametric and bootstrap

	$Year$
P-Spline	1997
AP	1997
SemiPar	1998
Bootstrap	1998

Table 6.12: Estimated year of significant decrease of Grey Plover count.

bands offer evidence that a significant decrease began closer to 1998. From simulations in Chapter 5, it is presumed that the AP estimate is the most accurate, although if low Type I error is of interest the bootstrap bands should be reported.

6.2 Estimating a Feature in Noisy Data

In the Blood Lactate illustration the question of interest concerns identifying a feature of the underlying lactate function which can be best explored through derivative estimation. The workload corresponding to the maximum second derivative of the Blood Lactate response to workload function is of interest. To test and compare the performance of the derivative estimation techniques in such problems, an overall measure of error, such as RMSED$(\cdot)$, is limited. Performance in estimating the features must also be assessed and this is a different problem when a feature and not the whole of f is of interest.

6.2.1 Simulation Study

A small simulation study was carried out to investigate which of the methods best estimated the x value ($xmax$) corresponding to a maximum second derivative. The function f_2 from Section 1.5 of Chapter 1 is used since it is similar, in having a unimodal second derivative, to a typical second derivative estimate of a lactate curve. An estimate of the magnitude of the maximum second derivative is not required in the Blood Lactate example but performance in estimating this value, i.e. $d2max$, is recorded here for reference.

A small sample size, $n = 15$, and Normal error ϵ with standard deviation equal to one sixth the range of f_2 was chosen. These choices of sample size and standard deviation were chosen to resemble typical Blood Lactate readings. Once again comparisons were made between the AP, P-Spline and semiparametric methods. One thousand responses $y = f_2 + \epsilon$ were simulated such that 1000 estimates of the $xmax$ and $d2max$ pair could be obtained through each of the AP, P-Spline and semiparametric methods. The corresponding absolute differences from the estimates to the actual $xmax$ and $d2max$, denoted $x.err$ and $d2.err$ respectively, were recorded to assess the performance of each of the three approaches.

Figure 6.21 displays boxplots for $x.err$ and $d2.err$ across 1000 simulations of f_2. The AP and P-Spline fits are roughly equal in ability to find the maximum of the second derivative and the x value at which this occurs. The semiparametric method is the poorest in terms of recognising this feature.

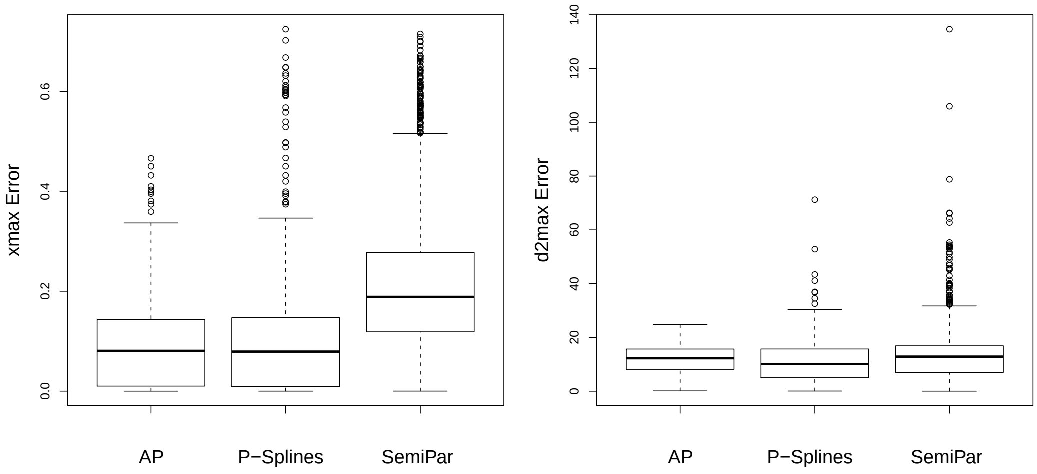


Figure 6.21: Boxplots of $x.err$ and $d2.err$ for f_2.

	Mean $x.err$ (sd)
AP	0.093(0.087)
P-Splines	0.104(0.118)
SemiPar	0.227(0.165)

Table 6.13: Performance in estimating the location of the maximum second derivative.

Table 6.13 provides evidence that the AP is to be preferred for estimating the x value corresponding to the location of the maximum of the second derivative, while Table 6.14 suggests the P-Spline fit gives the best estimate of the maximum of the second derivative itself (this is not required for the Blood Lactate illustration).

	Mean $d2.err$ (sd)
AP	11.77(5.126)
P-Splines	10.50(6.887)
SemiPar	13.56(10.87)

Table 6.14: Performance in estimating the maximum second derivative.

6.3 Inverse Prediction

To make inferences when estimating a feature of the data it is required to calculate the estimated standard error for the purpose of performing interval estimation or significance testing. Methods for calculating the standard error of an estimate $\hat{f}$ have been provided (Section 6.1). However, there are circumstances in which the prediction of a value of explanatory variable (e.g. x_{inv}) which gave rise to a certain value of response is needed, i.e. the inverse prediction problem. For example in the Blood Lactate data the estimated speed corresponding to the maximum second derivative value is simply a point estimate of the true population parameter. For inference it is necessary to calculate the standard error underlying this estimate. A simple empirical approach of residual resampling is now described to find a range of likely values for the true x_{inv}. The residuals e_i, $i = 1, \ldots, n$ are taken from the fit, i.e.

$$e_i = y_i - \hat{y}_i.$$

The bootstrap technique of resampling $e = (e_1, \ldots, e_n)$ with replacement leads to R 'new' observed samples $y^{(r)}$, $r = 1, \ldots, R$. These are calculated by

$$y^{(r)} = y + e^{(r)}$$

where $e^{(r)}$ is the rth resample of the residuals. For each $y^{(r)}$ the required feature is estimated and the corresponding value of explanatory (x^r_{inv}) is obtained. In this manner a distribution of R x_{inv}'s is obtained. From this empirical distribution a bootstrap interval estimate can be constructed by taking the required percentiles from this distribution.

6.3.1 Application of Inverse Prediction in the Blood Lactate Dataset

In the Blood Lactate example it is necessary to estimate the speed (x_{inv}) at which the maximum second derivative of the underlying lactate function (D2LMax) occurs. Evidence that the AP approach performs better than the P-Spline and semiparametric methods in this task has been found (Section 6.2). Using residual resampling the left panel of Figure 6.22 displays $R = 1000$ second derivative estimates along with the sample estimate in bold.

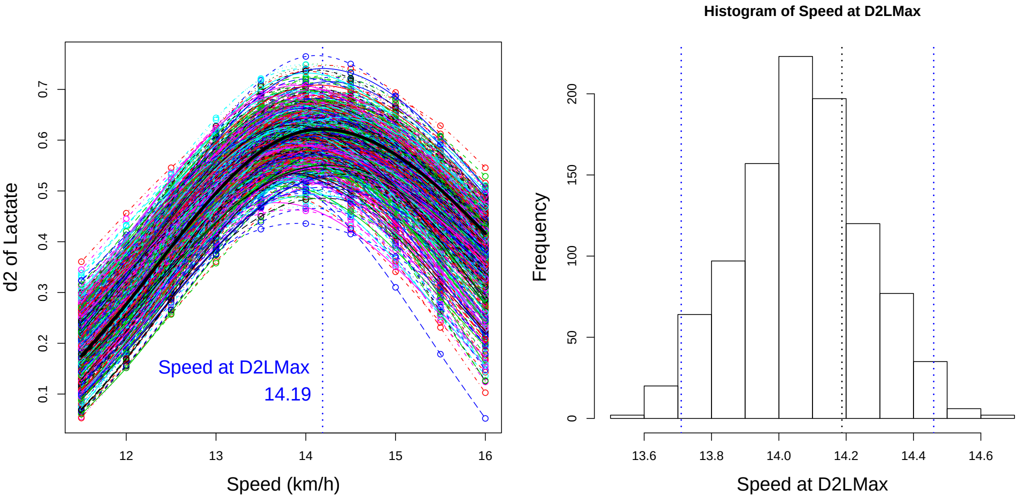


Figure 6.22: Left: A thousand second derivative estimates of the Blood Lactate data found using resampling (multicoloured) with original sample estimate (bold). Right: Histogram of 1000 x_{inv}'s obtained through resampling residuals with sample estimate (dashed black line) and 95% bootstrap interval estimate (blue dashed lines).

For each of these R D2LMax's a corresponding speed is obtained. A histogram of these speeds is exhibited in the right panel of Figure 6.22. The interval estimate was found to be (13.71, 14.46) km/h using the 2.5% and 97.5% percentiles of the empirical distribution of x_{inv}. This leads to an estimated range of likely values for the speed at which the true maximum second derivative of lactate occurs.

6.4 Chapter Summary

The need for variability bands is motivated by the requirement to estimate periods of significant change in a response variable over time or relative to some other explanatory variable. In this

Chapter, methods for estimating the variance of derivative estimates of semiparametric models have been summarised and new methods for variance calculation of P-Spline and AP derivative estimates have been developed. These variance estimates have been used to create approximate variability bands for each of the three methods, with a further resampling method for variability bands introduced.

These four approaches were tested across a large simulation study for coverage and precision. Also tested were the number of departures of a true function (f_2) and its derivatives from variability bands and the ability to correctly estimate a slope of zero from simulated data. The AP and P-Spline variability bands performed best overall in terms of coverage and precision when estimating derivatives, although the semiparametric bands offer the best coverage and precision when fitting a curve to observed data. Varying results were found between the four methods when applied to the Winter Nutrients and Scottish Bird Count examples and a discussion of which to believe based on the simulation results was provided.

The performance of the P-Spline, AP and semiparametric approaches to estimating a maximum second derivative and its location was assessed. This type of analysis is needed for the Blood Lactate study and evidence that the AP method is better at estimating the location of the maximum second derivative was found. Residual resampling was used to construct an empirical interval estimate in an inverse prediction problem.

A discussion of the main findings and conclusion of this thesis will now follow. Some interesting research problems which could stem from this research are discussed. The motivating datasets will also be revisited for a final time with concluding remarks provided.

Chapter 7

Conclusion and Further Work

The main aims of this research were to review current methods for derivative estimation and to attempt to find an improved method to estimate derivatives in situations where a nonlinear relationship exists. The goals laid out in the Introduction were as follows:

- To provide a comprehensive review of derivative estimation for noisy data.
- To outline the challenges faced in obtaining accurate derivative estimates when a nonlinear relationship between explanatory and response exists.
- To compare the performance of current methods for derivative estimation.
- To develop an approach to derivative estimation that improves on current methods.
- To establish suitable variance estimates for these estimators in order to produce reliable variability bands.

This thesis has summarised approaches to estimating derivatives using sequential differencing, linear models and several spline smoothing methods. Issues of instability and lack of flexibility were cause to reject the use of sequential differences and linear models respectively. Using local modelling via spline smoothing to first fit a curve to estimate the underlying function f and then to estimate derivatives as a by-product of this estimate was shown to improve on simpler estimation methods.

Through simulation, evidence suggested that the P-Spline and semiparametric regression techniques were to be preferred, in terms of first and second derivative estimation respectively, compared to other spline smoothing methods. Unfortunately, problems with boundary effects and concerns regarding undersmoothing suggested that more smoothing was necessary when attempting derivative estimation. Hence, it was decided to introduce an extra additive penalty to the P-Spline framework in order to penalise another feature of the estimate to the underlying function. This motivation came from evidence that any single constant smoothing parameter could not resolve the problems encountered in derivative estimation. The additive penalty (AP) method was shown empirically to achieve more accurate first and second derivative estimates than both P-Splines and semiparametric regression across a range of smoothing scenarios.

Methods to estimate the standard error of derivative estimates were developed for the AP and P-Splines. These estimated standard errors were then used to build pointwise variability bands around derivative estimates. When these bands were compared through simulation, they were found to have better (global and pointwise) coverage and precision for derivative estimation when compared to the recognised variability bands from semiparametric regression. However, concerns about the interpretation of these bands were raised since in the simulation study, none of the AP, P-Spline or semiparametric variability bands offered 95% pointwise coverage. A possible reason for this is the estimate of σ^2 which is based on an assumption of normality. This could be amended using a nonparametric estimate such as that introduced by Gasser [16]. The AP and P-Spline bands still outperform the recognised semiparametric bands in terms of global and pointwise coverage and do achieve respectable performance for derivative estimation. Therefore the use of these bands, in particular those of the AP method, in determining regions of significant change for example, is recommended.

7.1 Summary

Chapter 1 introduced the four main motivating illustrations for which derivative estimation played a crucial role. Each context had a slightly different question of interest. The Winter Nutrients illustration, for example, involved obtaining first derivative estimates for two nutrients (Phosphate and NTRZ) over time in order to find evidence of significant increase or decrease. A similar analysis was needed for the Scottish Bird Count study but as the response variable represented counts a different modelling approach was required. Accurate second derivative estimates were required for the Blood Lactate data such that the workload corresponding to the maximum second derivative could be identified. In the Astronomical example it was required to build a response from a convolution of two exploratory variables which involved derivatives. In Chapter 1 the six functions on which simulations throughout the thesis are based were introduced with an explanation of the pretext behind the choice of each provided.

Chapters 2 and 3 reviewed and compared several derivative estimation techniques which are commonly employed in the literature. These ranged from basic methods such as sequential differences and linear models to more complex models of spatially adaptive smoothing. In between the popular techniques of smoothing splines, mixed model smoothing and P-Splines were discussed. Many simulations were carried out to compare the performance in derivative estimation of these approaches. It was found that sequential differences are far too volatile to handle noisy data and that the lack of flexibility to deal with local change rules out the use of high order polynomial regression. The mixed model smoothers offer stability of estimates and P-Splines offer good performance as well as computational efficiency. No evidence was found that altering the chosen smoothing parameter, either by way of a constant 'fudge' or by letting it vary spatially, improved performance in derivative estimation.

P-Spline derivative estimation was thoroughly examined in Chapter 4 through yet more simulations. Under investigation were varying sample size, error variance and the selection method for the smoothing parameter. It was suggested that CV be used to select λ and that increasing n and decreasing σ resulted in increased precision of derivative estimates. No grounds were found to alter the Eilers & Marx recommended choices for basis degree, number of knots or penalty order when derivative estimation was of primary concern.

The proposed additive penalty approach to derivative estimation was introduced in Chapter 5 and extended to the Poisson case in order to model the Scottish Bird Count data. The additive penalty requires several extra choices for implementation, namely the use of a either a simultaneous or sequential approach to the selection of multiple smoothing parameters and then to the choice of order of difference penalties to employ. These matters were investigated through simulations and evidence was found to proceed with a sequential method using difference penalties of order 1 and 2. This additive penalty model was then compared to semiparametric smoothing and P-Splines, resulting in improved derivative estimates across the six simulated functions.

The Winter Nutrients and Scottish Bird Count examples require tests of the rate of change in levels of contaminants and bird count respectively. Derivative estimates alone are incapable of answering these questions. In Chapter 6 variability bands were developed for both P-Spline and additive penalty derivative estimates. These were then compared to semiparametric bands in a large simulation study which tested coverage and precision. Variability bands found by residual resampling were also included in the study but these bands are generally imprecise. The comparisons of the other three variability bands indicated that the semiparametric bands were to be preferred when estimating the underlying function f but that P-Splines and the additive penalty were superior for derivative estimation. The P-Spline and AP bands offer very similar results in terms of coverage, precision and departures. The Blood Lactate illustration requires a point estimate of the explanatory variable for a particular feature of the second derivative of the response. A brief simulation was performed to test which of P-Splines, semiparametric smoothing and AP were best able to handle this task with the AP found to offer the most precision. Finally a simple data driven approach to obtaining a likely range for an explanatory value at which a certain feature of a response occurs was given.

7.2 Motivating Examples

The four illustrations from Chapter 1 have been constantly updated through each Chapter in this thesis. A final visit to each is now provided using all of the results which have been gained throughout this work.

7.2.1 Winter Nutrients Data

The Winter Nutrients dataset contains salinity adjusted measurements of the nutrients NTRZ and Phosphate taken in the Irish Sea in the Winter of 1990/1991. Of interest to researchers is the rate of change in the level of nutrients over this Winter. More specifically, the primary goal is to evaluate the periods of significant decrease and significant increase in levels. A period of significant increase is defined as one in which both variability bands of the first derivative estimate are positive, significant decrease where bands are negative and no significant change where 0 is contained within the bands. The data, which are displayed in Figure 7.1, involve a nonlinear relationship between the measurement order and level of nutrient.

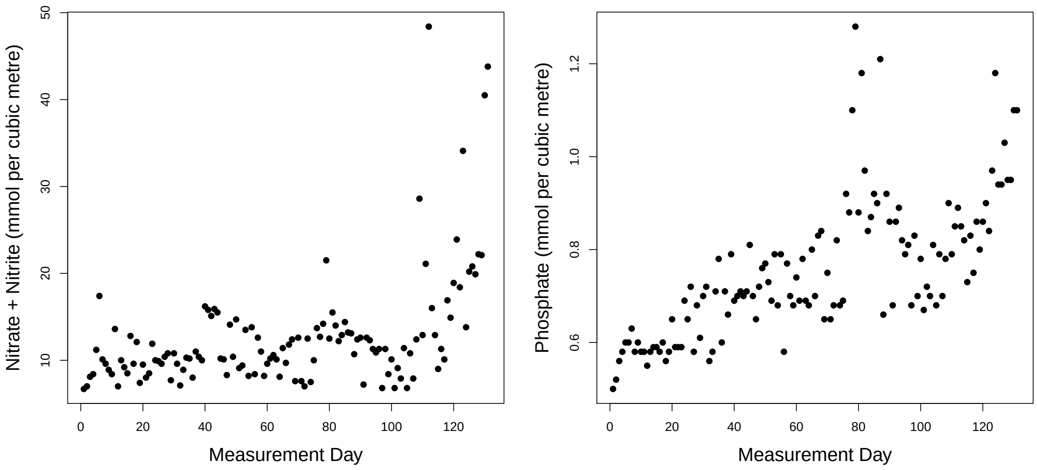


Figure 7.1: Winter Nutrients data.

In order to estimate the regions of significant change it is required to find accurate first derivative estimates from each which include variability bands. The AP method has been seen to offer improved first derivative estimates in simulations carried out in Chapter 5 as well as respectable variability bands in terms of precision and coverage in Chapter 6. These methods applied to both the NTRZ and Phosphate data are shown in Figure 7.2.

Tables 7.1 and 7.2 summarises the regions of significant increase and significant decrease for both nutrients.

Nutrient	Periods of Significant Decrease
NTRZ	44-47, 83-87, 96-100, 112-117 (20)
Phosphate	83-100 (18)

Table 7.1: Estimated regions of significant decrease for the Winter Nutrients example.

Nutrient	Periods of Significant Increase
NTRZ	1-5, 35-40, 72-79, 102-111, 118-123, 125-131 (42)
Phosphate	1-6, 18-42, 67-81, 103-129 (73)

Table 7.2: Estimated regions of significant increase for the Winter Nutrients example.

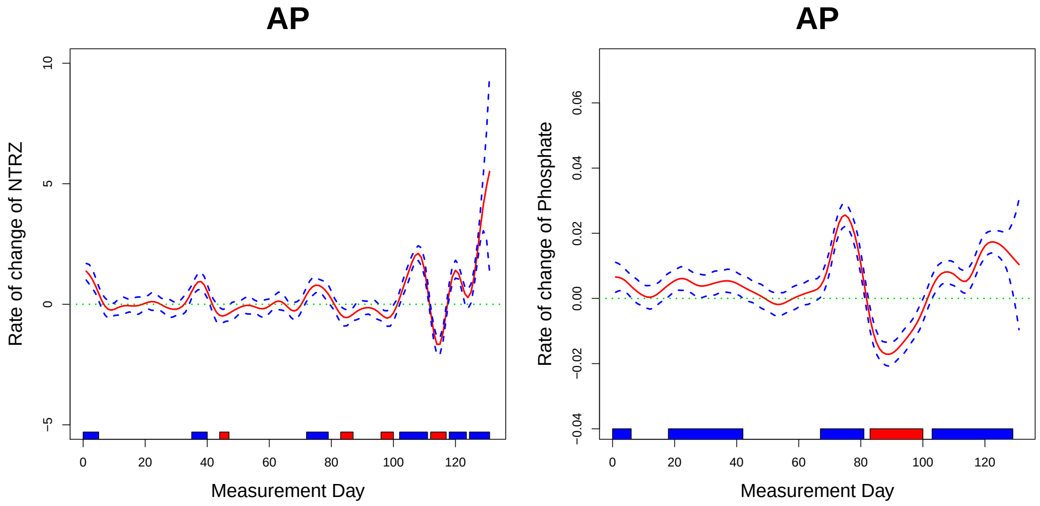


Figure 7.2: First derivative estimates with variability bands for the Winter Nutrients data. The blue boxes indicate areas of significant increase and the red boxes refer to areas of significant decrease

Since the simulations suggest the AP gives the most accurate estimates for rate of change among methods under examination, one can be relatively confident in the results presented in Tables 7.1 and 7.2. It seems that in 1990 Phosphate has longer overall periods of increase than decrease and that NTRZ has more decreasing periods than increasing.

7.2.2 Scottish Bird Count Data

The Scottish Bird Count example contains 31 annual counts of 11 bird species. The main goal is to estimate whether a significant decrease in bird count occurs and, if so, at what year this decrease begins. Similarly to the Winter Nutrients data, a first derivative estimate with accurate variability bands is required in order to answer this question. The Grey Plover counts are displayed in Figure 7.3.

One important difference in this study is that the response variable is one of counts rather than of continuous observations. Generalised smoothing using the AP method has been introduced in Chapter 5. By incorporating the assumptions of generalised smoothing one can be more confident of the accuracy of the first derivative estimates obtained. First derivative estimates of the Grey Plover counts along with variability bands are displayed in the right panel of Figure 7.3. A significant decrease can be seen when the bands are both negative, as happens

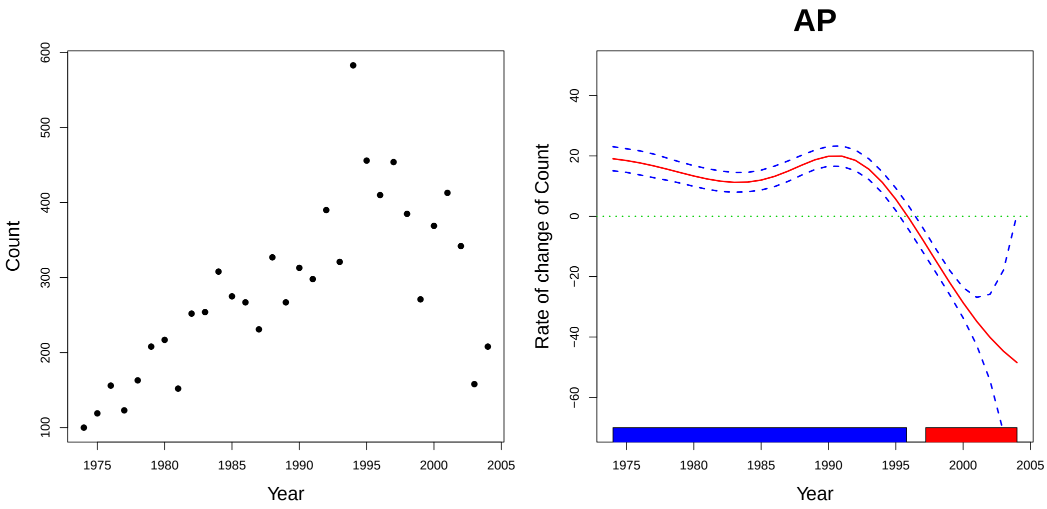


Figure 7.3: Counts of Grey Plover (left) with first derivative estimate and variability bands (right) using the AP method. The blue boxes indicate areas of significant increase and the red boxes refer to areas of significant decrease

here at roughly 1997.

7.2.3 Blood Lactate Data

The Blood Lactate illustration consists of lactate measurements taken from 23 elite athletes at several incremental workloads on a treadmill. The primary aim here is to achieve an objective measure of endurance. One such marker is the D2LMax (Newell et al. [39]) where the speed corresponding to the maximum of the second derivative of the underlying lactate function is taken. The data for one athlete are displayed in Figure 7.4.

Once again there is a nonlinear relationship between the explanatory variable (speed) and response (blood lactate). In comparison to the Winter Nutrients example, and to an extent in the Scottish Bird Count data, there is less noise present in this case. It is required to obtain a second derivative estimate from the observed data, the right panel of Figure 7.4 exhibits one such estimate using the AP method described in Chapter 5. Given the evidence in Chapter 5 the speed at the maximum second derivative for this athlete is likely to be close to 14.19 km/h with the 95% bootstrap interval between 13.71km/h and 14.46 km/h.

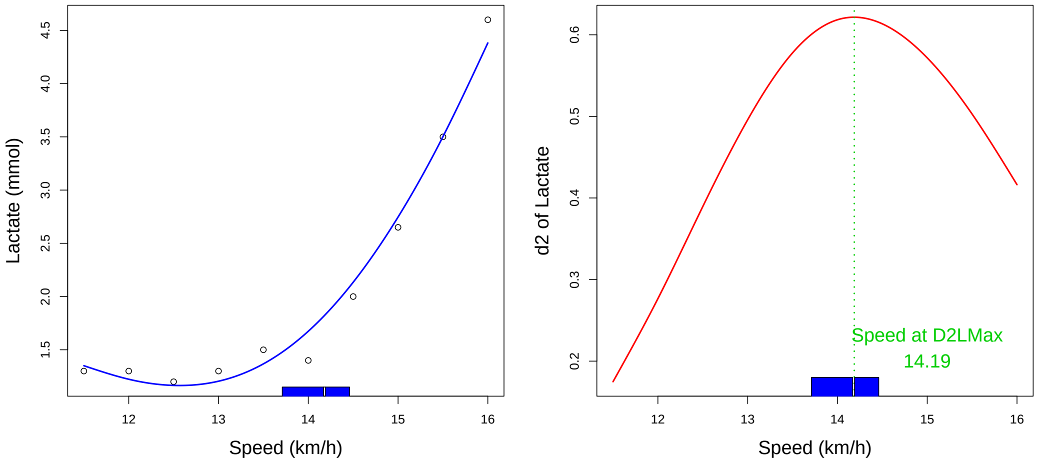


Figure 7.4: Blood lactate measured on one athlete at several incremental speeds with 95% bootstrap interval (blue box) for speed at D2LMax (left) and second derivative estimate using the AP method with 95% bootstrap interval (blue box) for speed at D2LMax (right).

7.2.4 Astronomical Data

The Astronomical dataset comprises of two main variables gas (ρ_{gas}) and temperature (T) measured at several distances in arcmin from the centre of galaxy cluster A1995. It is the primary goal of this research to estimate a variable ρ_{tot} found by

$$\frac{d}{dx}\frac{\frac{d}{dx}(\rho_{gas}kT\mu m_p)}{\rho_{gas}} = -4\pi G\rho_{tot} \tag{7.1}$$

where x (in arcmin) is a measure of distance, G is the universal gravitational constant, k is Boltzmann's constant, μ is the mean molecular weight in any cluster and m_p is the mass of the proton. The gas and temperature profiles are shown in Figure 7.5.

A nonlinear relationship exists for both profiles and another issue is that gas is measured 64 times over a range (0.32, 1.84) of arcmin while temperature is measured 8 times over a range of (0.29, 1.81) arcmin. The equation in (7.1) requires these two variables to be multiplied. The temperature is first smoothed and estimated at the arcmin for gas over a range of (0.32, 1.84) arcmin. The smoothing curve using P-Splines is shown in the top panel of Figure 7.6.

The next stage in building the response is to multiply T by ρ_{gas} and obtain a first derivative estimate using the AP method, these are shown in the bottom row of Figure 7.6. To complete

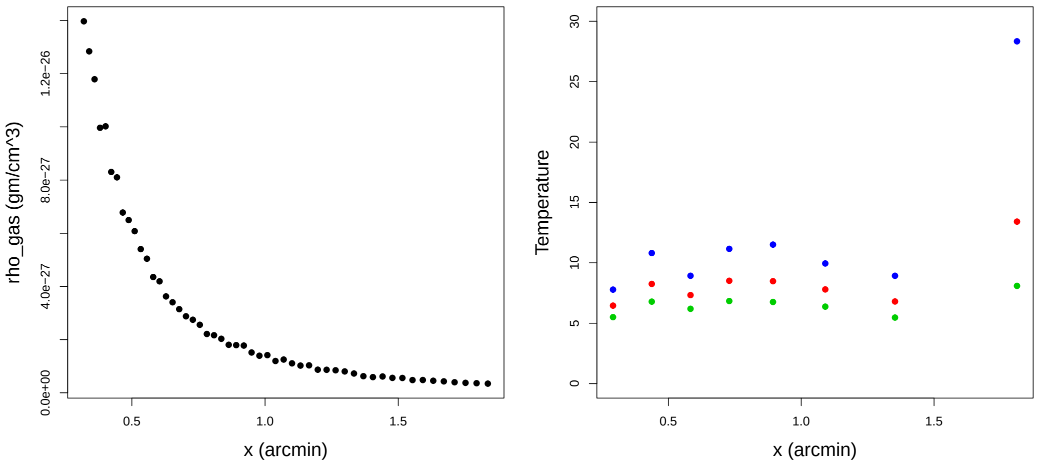


Figure 7.5: Gas and temperature profiles for the Astronomical data.

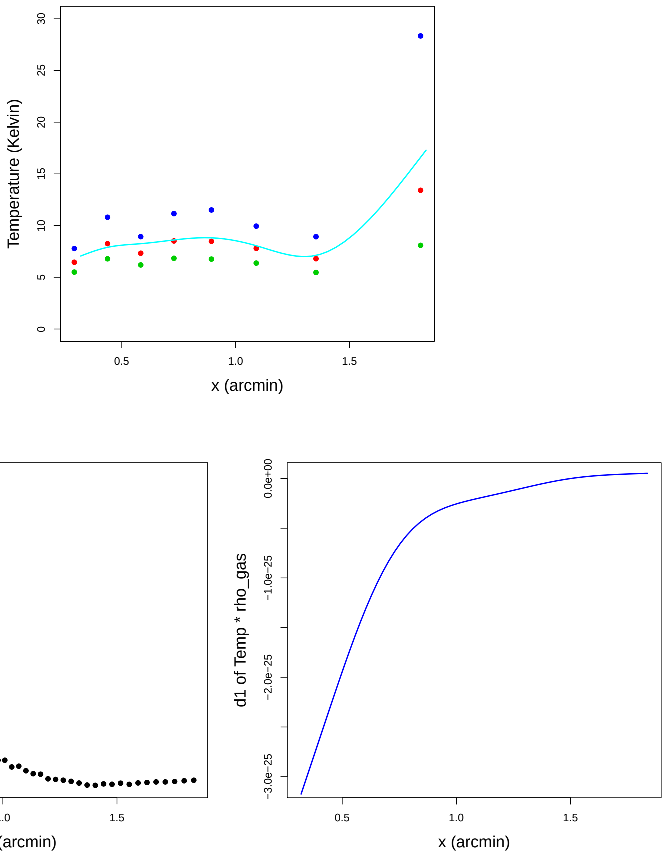


Figure 7.6: Top: Temperature smoothed and estimated at the arcmin for gas over a range of (0.32, 1.81) arcmin. Bottom: $T\rho_{gas}$ (left panel) with first derivative estimate (right panel).

the building of ρ_{tot}, this derivative is divided by ρ_{gas}, a first derivative is again obtained by AP and then is rescaled accordingly. The final estimate for ρ_{tot} is plotted in Figure 7.7.

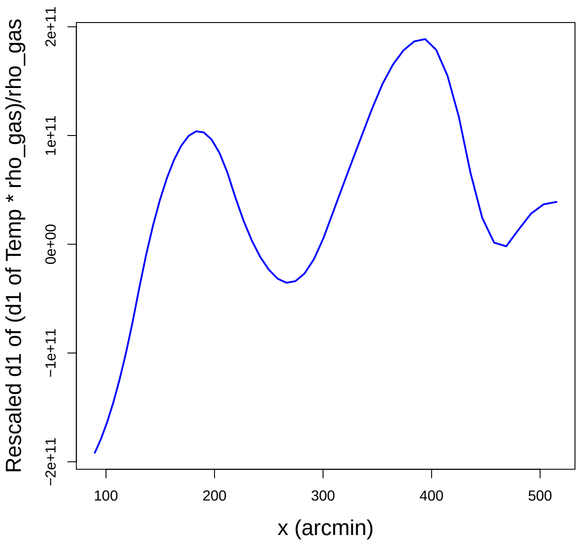


Figure 7.7: Estimated ρ_{tot}.

7.3 Review of Simulation Studies Performed

During the course of this research, simulation studies were performed to compare methods of derivative estimation. These were generally performed by simulating data from the six functions $f_1, \ldots, f_6$ (Section 1.5) with x uniformly distributed and some Gaussian error added with constant variance equal to a fraction of the range of $f(x)$. Here a review of the main results of these extensive simulations is presented such that clear recommendations for a potential future researcher can be outlined.

The simulation studies from Chapters 2 and 3 reveal that P-Splines or semiparametric regression should be used for derivative estimation when compared to other well-known techniques such as linear models and smoothing splines as well as the adaptive methods summarised in Section 3.2.4. In Chapter 4 no evidence was found to change any recommended choices of knots, penalty or or basis degree of the P-Spline fitting procedure. Evidence was discovered that using CV/GCV to choose λ for derivative estimation was to be preferred to AIC/BIC methods. Once the AP method was introduced in Chapter 5, simulations were first performed to eliminate some potential choices underlying the approach. Firstly, for reasons of accuracy in derivative estimation and computational efficiency, it was decided to select the smoothing parameters sequentially. Choosing different pairings of penalty orders was demonstrated to have negligible impact on the performance of the AP for derivative estimation. Finally the AP was compared to P-Splines and semiparametric regression and clear empirical evidence that the AP should be used for derivative estimation was found across each function $f_1, \ldots, f_6$. It is therefore the recommendation of this research that an additive penalty P-Spline approach, with penalty orders $d_1 = 1$ and $d_2 = 2$ should be used when derivative estimation is the main aim of an analysis.

7.4 Further Work

While evidence that the additive penalty method offers improved performance in derivative estimation has been seen it is also clear that these estimates themselves are far from perfect. Issues of boundary effects and inability to handle heteroscedatic data still linger. Further research into derivative estimation methods could well find an approach which offers yet more improvement, however, derivative estimation is doubtless a difficult problem. Here some extensions and potential future research topics in derivative estimation are summarised.

7.4.1 Derivative Estimation on a Surface

The most obvious extension of the additive penalty method is to two dimensional relationships where surface fitting is required in place of curve fitting. Most of the smoothing techniques described in Chapter 3 have been extended to two dimensions in the literature, although software packages are not widely available. The rate of change of a surface is, although more complex, as essential in some contexts as derivative estimation for explanatory/response data. One nice example can be found in Durban et al. [11] where the relationship between age, year and deaths from a mortality database is studied using a two dimensional Poisson P-Spline model. It is clear that the rate of change of this relationship would be of interest to people in the life insurance industry.

7.4.2 Mixed Model Derivative Estimation

Evidence that an additive penalty, when applied to the standard P-Spline model, offers improved derivative estimates has been provided in Chapter 5. No attempt was made to modify the mixed model approach to handle derivative estimation. It is a matter for future research whether this would lead to improvements in derivative estimation. Using an additive penalty approach to mixed model smoothing could lead to improvements, since evidence that the mixed model smoothers already offer more stable derivative estimates was found in Section 3.2.5.

7.4.3 Bayesian Derivative Estimation

Bayesian approaches have not been considered in this research (although the adaptive mixed model technique from Section 3.2.4.2 does specify a distribution for the shrinkage penalties in order to have a spatially varying fit). Bayesian P-Splines have appeared in the literature (Lang [29]), yet no application of these methods to derivative estimation has been attempted. Using Bayesian methods has the potential to aid the problems in derivative estimation, through the choice of a suitable prior for example, but research into this has yet to be explored.

7.4.4 Correlated Error Structure

The simulations which have been used to compare derivative estimation methods have used constant variance in the errors added to simulated data. In real life situations this is not always the case. Alterations to the P-Spline model to take into account other variance/covariance structures have been made by Currie & Durban [8]. Further testing into scenarios of correlated error structures and how the different derivative estimation methods cope with this is required. Moreover, the AP derivative estimation theory developed as part of this research should be expanded to handle problems with a correlated error structure. For instance, it is quite likely that the Blood Lactate example contains a response in which correlated error exists since the measurements are taken on the same player at incremental workloads.

7.4.5 SiZer for *P*-Splines and the AP

Chaudhuri & Marron [6] introduced an exploratory tool known as SiZer to test whether certain features of a relationship ‘really exist’. The method is based on kernel smoothing and this is extended to incorporate smoothing splines in Marron & Zhang [34]. The method works by varying the bandwidth to get a ‘family’ of smooth fits of the data. For each bandwidth at each x, a confidence interval for the derivative is constructed. The SiZer map is then colour coded where this interval is fully above zero (blue), below zero (red) and contains zero (purple). In situations where modes are indistinguishable from background noise the map is coloured grey. An example of a SiZer map is given in Figure 7.8, the white curves show effective window widths for each bandwidth.

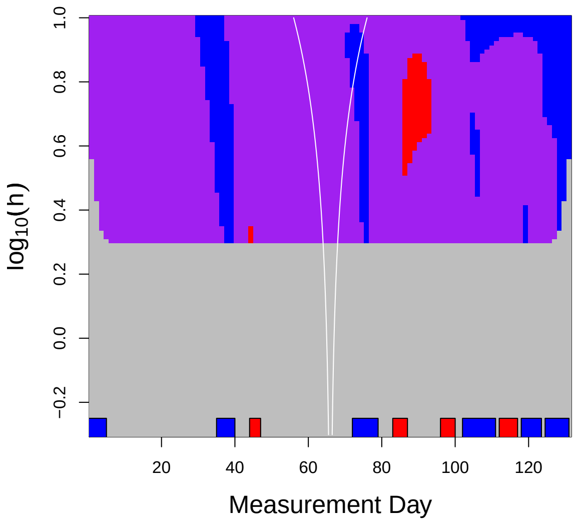


Figure 7.8: SiZer map for the NTRZ data with regions found to have significant decrease (red), increase (blue) or no change (purple). The boxes represent regions found by the AP to have significant decrease (red) or increase (blue) in contamination.

Both the Winter Nutrients and Scottish Bird Count datasets use significant zero crossings of the derivative as their primary goal. As such these SiZer plots are of immense use to researchers in these studies. The SiZer method could be updated to use the methods discussed in this thesis, which offer improvements in derivative estimation over smoothing splines.

7.4.6 Derivative Estimation for Generalized Linear Models

The Poisson P-Spline model discussed in Section 3.2.2.2 has been extended to handle derivative estimation. Generalised linear smoothing has been discussed in the literature, for instance, a Binomial P-Spline model appears in Eilers & Marx [12]. Derivative estimation needs to be extended to the case where

$$g(\mu) = \eta = B\alpha$$

for $\mu = \mathbb{E}(y)$ such that the changes in any exponential family response variable relative to some explanatory can be modelled sufficiently.

7.4.7 Comparisons with Kernel Derivative Estimates

A brief discussion of the available kernel derivative estimation literature was provided in Section 3.3. No extensive comparison into the performance in derivative estimation of kernel and spline methods appears in the literature. For practical analysis, it would be of great interest to understand which set of methods works better in specific situations (e.g. sample sizes, variances).

7.4.8 Multiple Sample Problems

In the Scottish Bird Data the count of Grey Plover has been analysed using derivative estimation. However, as was mentioned in Section 1.2, the Grey Plover is but one of eleven species contained in this dataset. It would be useful to compare across these eleven species to find whether there exist common or, alternatively, distinct features of the rate of change of the counts of the bird species. In general, the extension of derivative estimation to multiple sample problems should be addressed such that comparisons of estimates and error between samples can be accurately described.

7.4.9 The Inverse Prediction Problem

In the Blood Lactate example it is required to estimate the speed corresponding to the maximum second derivative of the underlying lactate function. Several derivative estimation methods introduced in this thesis are easily capable of obtaining these estimates (although results may vary!). An empirical interval estimate giving a range of likely values for this speed was discussed in Section 6.3. This rather crude method could be improved on by using asymptotic properties of a smooth fit such that true 95% confidence intervals may be obtained for a certain value of explanatory variable at which a feature of the response occurs.

7.4.10 Derivative Estimation for Spatial Data

P-Splines are used to smooth spatial data in Lee & Durban [30] such that the AP could easily be extended to handle this type of analysis. However, for the case of a spatial derivative estimation problem, little research exists. This is unfortunate since this is a clear area of application for derivative estimation. In the Marine dataset, where levels of two nutrients are measured from the same place over time, repeated measurements at different positions across the Irish Sea would lead to a spatio-temporal model involving derivative estimation.

Bibliography

[1] Akaike, H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika 60*, 2 (1973), 255–265.

[2] Aldrin, M. Improved predictions penalizing both slope and curvature in additive models. *Journal of Computational Statistics and Data Analysis 50*, 2 (2006), 267–284.

[3] Beaver, W. L., Wasserman, K., and Whipp, B. J. Improved detection of lactate threshold during exercise using a log-log transformation. *Journal of Applied Physiology 59*, 6 (1985), 1936–1940.

[4] Belitz, C., and Lang, S. Complex additive penalties for generalized structured additive regression. *Proceedings of the 23rd IWSM* (2008), 115–120, Utrecht.

[5] Bollaerts, K., Eilers, P. H. C., and van Mechelen, I. Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology 59* (2006), 451–469.

[6] Chaudhuri, P., and Marron, J. S. Sizer for exploration of structures in curves. *Journal of the American Statistical Association 94*, 447 (1999), 807–823.

[7] Craven, P., and Wahba, G. Smoothing noisy data with spline functions. *Numerische Mathematik 31* (1978), 377–403.

[8] Currie, I., and Durban, M. Flexible smoothing with P-splines: a unified approach. *Statistical Modelling 2*, 4 (2002), 333–349.

[9] de Boor, C. *A Practical Guide to Splines.* Springer, New York, 1978.

[10] Donoho, D., and Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association 90* (1995), 1200–1224.

[11] Durban, M., Currie, I., and Eilers, P. H. C. Using P-splines to smooth two dimensional poisson data. *Proceedings of the 17th IWSM* (2002), 207–214, Crete.

[12] Eilers, P. H. C., and Marx, B. Flexible smoothing with B-splines and penalties. *Statistical Science 11*, 2 (1996), 89–121.

[13] Eilers, P. H. C., and Marx, B. Splines, knots and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics 2* (2010), 637–653.

[14] Fan, J., and Gijbels, I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 57*, 2 (1995), 371–394.

[15] Friedman, J. H. Multivariate adaptive regression splines. *The Annals of Statistics 19*, 1 (1991), 1–67.

[16] Gasser, T., Sroka, L., and Jennen-Steinmetz, C. Residual variance and residual pattern in nonlinear regression. *Biometrika 73*, 3 (1986), 625–633.

[17] Green, P. J., and Silverman, B. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, London, 1994.

[18] Härdle, W. *Applied Nonparametric Regression.* Cambridge University Press, Boston, 1990.

[19] Härdle, W., and Bowman, A. W. Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association 83*, 401 (1988), 102–110.

[20] Hastie, T., and Loader, C. Local regression: automatic kernel carpentry. *Statistical Science 8*, 2 (1993), 120–129.

[21] Heckman, N. E., and Ramsay, J. O. Penalized regression with model-based penalties. *The Canadian Journal of Statistics 28*, 2 (2000), 241–258.

[22] Huh, J., and Carrière, K. C. Estimation of regression functions with a discontinuity in a derivative with local polynomial fits. *Statistics and Probability Letters 56*, 3 (2002), 329–343.

[23] Hurvich, C. M., Simonoff, J. S., and Tsai, C. L. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*, 2 (1998), 271–293.

[24] Hutchinson, M. F., and de Hoog, F. R. Smoothing noisy data with spline functions. *Numerische Mathematik 47* (1985), 99–106.

[25] Kendall, M. G. A new measure of rank correlation. *Biometrika 30*, 1-2 (1938), 81–93.

[26] Kohn, R., Ansley, C. F., and Tharm, D. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association 86*, 416 (1991), 1042–1050.

[27] Krivobokova, T., Crainiceanu, C. M., and Kauermann, G. Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics 17*, 1 (2008), 1–20.

[28] Lai, C. J., and Chu, C. K. Triple smoothing estimation of the regression function and its derivatives in nonparametric regression. *Journal of Statistical Planning and Inference 98*, 1-2 (2001), 157–175.

[29] Lang, S., and Brezger, A. Bayesian P-splines. *Journal of Computational and Graphical Statistics 13*, 1 (2004), 183–212.

[30] Lee, D. J., and Durbán, M. Smooth-car mixed models for spatial count data. *Journal of Computational Statistics and Data Analysis 53*, 8 (2009), 2968–2979.

[31] Lundberg, M. A., Hughson, R. L., Weisiger, K. H., Jones, R. H., and Swanson, G. D. Computerized estimation of lactate threshold. *Computers and Biomedical Research 19*, 5 (1986), 481–486.

[32] Luo, Z., and Wahba, G. Hybrid adaptive splines. *Journal of the American Statistical Association 92*, 437 (1997), 107–116.

[33] Mack, Y. P., and Müller, H. G. Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā: The Indian Journal of Statistics, Series A 51*, 1 (1989), 59–72.

[34] Marron, J. S., and Zhang, J. Sizer for smoothing splines. *Computational Statistics 20* (2005), 481–502.

[35] Marx, B., and Eilers, P. H. C. Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis 28* (1998), 193–209.

[36] McCulloch, C. E., and Searle, S. R. *Generalized, Linear, and Mixed Models*. Wiley, New York, 2001.

[37] Müller, H. G., Stadtmüller, U., and Schmitt, T. Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika 74*, 4 (1987), 743–749.

[38] Newell, J., and Einbeck, J. A comparative study of nonparametric derivative estimators. *Proceedings of the 22nd IWSM* (2007), 453–456, Barcelona.

[39] Newell, J., Einbeck, J., Madden, N., and McMillan, K. Model free endurance markers based on the second derivative of blood lactate curves. *Proceedings of the 20th IWSM* (2005), 357–364, Sydney.

[40] Newell, J., McMillan, K., Grant, S., and McCabe, G. Using functional data analysis to summarise and interpret lactate curves. *Computers in Biology and Medicine 36* (2006), 262–275.

[41] Prewitt, K., and Lohr, S. Bandwidth selection in local polynomial regression using eigenvalues. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*, 1 (2006), 135–154.

[42] Ramsay, J. O. Derivative estimation. *StatLib S News* (Thu, 12 March 1998).

[43] Ramsay, J. O., and Li, X. Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*, 2 (1998), 351–363.

[44] Ramsay, J. O., and Silverman, B. *Functional Data Analysis, Second Edition.* Springer, New York, 2006.

[45] Reinsch, C. H. Smoothing by spline functions. *Numerische Mathematik 10* (1967), 77–83.

[46] Ruppert, D. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics 11*, 4 (2002), 735–757.

[47] Ruppert, D., and Carroll, R. J. Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics 42*, 2 (2000), 205–223.

[48] Ruppert, D., Carroll, R. J., and Wand, M. P. *Semiparametric Regression.* Cambridge University Press, Cambridge, 2003.

[49] Ruppert, D., and Wand, M. P. Multivariate locally weighted least squares regression. *The Annals of Statistics 22*, 3 (1994), 1346–1370.

[50] Sangalli, L., Secchi, P., Vantini, S., and Veneziani, A. Efficient estimation of three-dimensional curves and their derivatives by free knot regression splines, applied to the analysis of inner carotid artery centrelines. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 58*, 3 (2009), 285–306.

[51] Schoenberg, I. J. Contribution to the problem of approximation of equidistant data by analytic functions. *Quarterly Applied Mathematics 4* (1946), 45–99, 112–141.

[52] Schwartz, S. C. Estimation of probability density by an orthogonal series. *The Annals of Mathematical Statistics 38*, 4 (1967), 1261–1265.

[53] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics 6*, 2 (1978), 461–464.

[54] Shibata, R. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika 63*, 1 (1976), 117–126.

[55] Simpkin, A., Eilers, P. H. C., Gampe, J., and Newell, J. An additive penalty approach to derivative estimation of noisy data. *Proceedings of the 24th IWSM* (2009), 351–358, Ithaca.

[56] Stoker, T. M. Smoothing bias in density derivative estimation. *Journal of the American Statistical Association 88*, 423 (1993), 855–863.

[57] Stone, C. J. Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics 8*, 6 (1980), 1348–1360.

[58] Stone, C. J. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics 10*, 4 (1982), 1040–1053.

[59] Tukey, J. W. *Exploratory data analysis.* Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass., 1977.

[60] Wahba, G. A polynomial algorithm for density estimation. *The Annals of Mathematical Statistics 42*, 6 (1971), 1870–1886.

[61] Wahba, G., and Wold, S. A completely automatic french curve: Fitting spline functions by cross-validation. *Communications in Statistics 4* (1975), 1–17.

[62] Wahba, G., and Wold, S. Periodic splines for spectral density estimation: The use of cross validation for determining the degree of smoothing. *Communications in Statistics 4*, 2 (1975), 25–141.

[63] Wand, M. P. Smoothing and mixed models. *Computational Statistics 18* (2003), 223–249.

[64] Welsh, A. H. Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica 6* (1996), 347–366.

[65] Xia, Y. Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*, 4 (1998), 797–811.

[66] Zhou, S., and Wolfe, D. A. On derivative estimation in spline regression. *Statistica Sinica 10* (2000), 93–108.

Appendix A

R Code

A.1 Additive Penalty Method

The ‘AP’ function fits the additive penalty P-Spline model to data with one explanatory and one continuous (`family = 2`) or count (`family = 1`) response. The smoothing parameters may be inputed directly or chosen using CV, GCV, AIC or BIC. The output provides first and second derivative estimates as well as variability bands for the fit and derivative estimates. The routine includes functions for creating the B-Spline and difference matrices B and D respectively and also has calls to fitting and update functions which are listed below.

```
AP <-
function (x, y, offset, w, family = 2, nseg = min(40,
floor(length(x)/5)), bdeg = 3, pord = 2,
lambda1 = NULL, lambda2 = NULL, df = NULL,
method = 1, control = list(), se = 2)
{
    # other components
    if (missing(w)) {
        w <- rep(1, length(y))
    }
    if (missing(offset)) {
        offset <- rep(0, length(y))
    }
    m      <- length(y)
    FAM    <- family
    wei    <- w
    MET    <- method
    MON    <- F
    TOL1   <- 1e-06
    TOL2   <- 0.002
    MAX.IT <- 50
    TOL    <-  TOL1
    xl     <- min(x)
    xr     <- max(x)
    xmax   <- xr + 0.01 * (xr - xl)
    xmin   <- xl - 0.01 * (xr - xl)
```

```
# some functions
ndiff <- function(n, d = 1) {
if (d == 1)
{D <- diff(diag(n))}
else
{D <- diff(ndiff(n, d - 1))}
 D}

tpower <- function(x, t, p)
(x - t) ^ p * (x > t)

bbase <- function(x, xl, xr, nseg, bdeg){
dx    <- (xr - xl) / nseg
knots <- seq(xl - bdeg * dx, xr + bdeg * dx,
         by = dx)
P     <- outer(x, knots, tpower, bdeg)
n     <- dim(P)[2]
D     <- ndiff(n, bdeg + 1) / (gamma(bdeg + 1)
         * dx ^ bdeg)
B     <- (-1) ^ (bdeg + 1) * P %*% t(D)
B }

# set up B and D's
B      <- bbase(x, xmin, xmax, nseg, bdeg)
nb     <- ncol(B)
zeros1 <- rep(0, nb - pord)
zeros2 <- rep(0, nb - (pord+1))
D.1    <- diff(diag(nb), diff=pord)
D.2    <- diff(diag(nb), diff=pord+1)

# count response
if (FAM == 1) {
# initialise
y[is.na(y)] <- 0
eta0        <- log(y + 1)
mu0         <- exp(eta0 + offset)
w0          <- wei * mu0
z0          <- wei * ((y - mu0)/mu0 + eta0)
P1          <- sqrt(1e+08) * D.1
P2          <- sqrt(1e+08) * D.2
fit0        <- lsfit(rbind(B, P1, P2), c(z0,
               zeros1, zeros2), wt = c(w0,
               (zeros1 + 1), zeros2 + 1),
               intercept = F)
a.init      <- fit0$coef

# Method 1 or 2, optimise lambdas by AIC or BIC
if (MET == 1 | MET == 2) {
    by.lambda <- length(seq(0.0001, 1,
                 by = TOL2))
    opt.ic1   <- function(X) {
        FIT   <- APpois(x = x, y = y, offset =
                 offset, wei = wei, zeros1 =
                 zeros1, zeros2 = zeros2,
                 B = B, lambda1 = X, lambda2
                 = 0, D.1 = D.1, D.2 = D.2,
                 a.init = a.init, MON = MON,
                 TOL = TOL1,
            MAX.IT = MAX.IT)
        return(ifelse(MET == 2, FIT$aic,
               FIT$bic))
    }
    lambda1.hat <- cleversearch(fn = opt.ic1,
                   lower = 0.0001, upper = 1,
                   ngrid = by.lambda, logscale
                   = F, verbose = FALSE)[[1]]

    opt.ic2 <- function(X) {
        FIT <- APpois(x = x, y = y, offset =
               offset, wei = wei, zeros1 =
               zeros1, zeros2 = zeros2, B = B,
               lambda1 = lambda1.hat,
               lambda2 = X, D.1 = D.1, D.2 = D.2,
               a.init = a.init, MON = MON,
               TOL = TOL1,
            MAX.IT = MAX.IT)
        return(ifelse(MET == 2, FIT$aic, FIT$bic))
    }
    lambda2.hat <- cleversearch(fn = opt.ic2,
                   lower = 0.0001, upper = 1,
                   ngrid = by.lambda, logscale = F,
                   verbose = FALSE)[[1]]
    FIT <- APpois(x = x, y = y, offset = offset,
                  wei = wei, zeros1 = zeros1,
                  zeros2 = zeros2, B = B,
                  lambda1 = lambda1.hat,
                  lambda2 = lambda2.hat,
                  D.1 = D.1, D.2 = D.2, a.init
                  = a.init, MON = MON, TOL = TOL1,
        MAX.IT = MAX.IT)
}

# Method 3, given lambdas
if (MET == 3) {
    lambda1.hat <- lambda1
    lambda2.hat <- lambda2
    FIT <- APpois(x = x, y = y, offset = offset,
           wei = wei, zeros1 = zeros1, zeros2 =
           zeros2, B = B, lambda1 = lambda1.hat,
           lambda2 = lambda2.hat, D.1 = D.1, D.2
           = D.2, a.init = a.init, MON = MON,
           TOL = TOL1, MAX.IT = MAX.IT)
}

# fill
aic            <- FIT$aic
bic            <- FIT$bic
df             <- FIT$df
dev            <- FIT$dev
coef           <- FIT$a
h              <- FIT$h
eta.hat        <- B %*% coef
y.hat          <- exp(eta.hat + offset)
lambda1.hat    <- lambda1.hat
lambda2.hat    <- lambda2.hat
sigma          <- sqrt(sum((y - y.hat)^2)/(m - sum(h)))
q              <- bdeg
P1             <- sqrt(lambda1.hat)*D.1
P2             <- sqrt(lambda2.hat)*D.2
W              <- matrix(0, nrow = m, ncol = m)
for(i in 1:m)
{
W[i,i]     <- exp(eta.hat)[i]
}

# error bands
u          <- seq(xl, xr, length = m)
Bu         <- bbase(u, xmin, xmax, nseg, q)
Covb       <- solve(t(B) %*% W %*% B + t(P1) %*% P1
              + t(P2) %*% P2)
Covz       <- sigma^2 * Bu %*% Covb %*% t(Bu) %*% W
              %*% Bu %*% Covb %*% t(Bu)
seb        <- se * sqrt(diag(Covz))

# first derivative
b.d1      <-  bbase(x, xmin, xmax, nseg, q - 1)
vec       <-  as.vector(coef)
dseg      <-  (xmax-xmin)/nseg
temp1     <-  cbind(vec[2:length(vec)],
              c(vec[2:length(vec)-1]))
d1.coef   <-  temp1[,1] - temp1[,2]
d1.eta    <-  q * (q * dseg)^(-1) * b.d1 %*%
              d1.coef
d1        <-  d1.eta*exp(eta.hat + offset)
```

```
# d1 error bands
u        <-  seq(xl, xr, length = m)
Bd1      <-  bbase(u, xmax, xmin, nseg, q - 1)
dummy    <- cbind(Bd1, rep(0, m))
Covb     <- solve(t(B) %*% W %*% B + t(P1) %*%
            P1 + t(P2) %*% P2)
Covz     <- sigma^2 * q^2/((q*dseg)^2) * dummy
            %*% Covb %*% t(dummy) %*% W %*% dummy
            %*% Covb %*% t(dummy)
seb.d1   <- se * sqrt(diag(Covz))

# second derivative
b.d2     <-  bbase(x, xmin, xmax, nseg, q - 2)
temp2    <-  cbind(vec[3:length(vec)],
             c(vec[3:length(vec)-1]),
             c(vec[3:length(vec)-2]))
d2.coef  <-  temp2[,1] - 2 * temp2[,2] + temp2[,3]
d2.eta   <-  (q - 1) * (dseg * (q - 1))^(-1) * q
             * (q * dseg)^(-1) * b.d2 %*% d2.coef
d2       <-  exp(d2.eta + offset)

# d2 error bands
u        <-  seq(xl, xr, length = m)
Bd2      <-  bbase(u, xmin, xmax, nseg, q - 2)
dummy    <- cbind(Bd2, rep(0, m), rep(0, m))
Covz     <- ((q - 1) * (dseg * (q - 1))^(-1) * q *
            (q*dseg)^(-1))^2 * dummy %*% Covb %*%
            t(dummy) %*% W %*% dummy %*% Covb %*%
            t(dummy)
seb.d2   <- se * sqrt(diag(Covz))

object <- list(call = call, n = m, aic = aic,
          bic = bic, lev = h, df = df, dev = dev,
          lambda1 = lambda1.hat, lambda2 =
          lambda2.hat, nseg = nseg, bdeg = bdeg,
          pord = pord, x = x, y = y, offset =
          as.vector(offset), w = as.vector(wei),
          y.hat = y.hat, linear.predictors =
          as.vector(eta.hat), coefficients =
          as.vector(coef), d1 = as.vector(d1),
          d2 = as.vector(d2), se = se, seb = seb,
          seb.d1 = seb.d1, seb.d2 = seb.d2)

class(object) <- "AP"
return(object)
}
```

```
# continuous response
if(FAM == 2)
{
# initialise
y[is.na(y)] <- 0
mu0         <- mean(y)
P1          <- sqrt(1e+08)*D.1
P2          <- sqrt(1e+08)*D.2
fit0        <- lsfit(rbind(B, P1, P2), c(y, zeros1,
               zeros2), wt = c(wei, (zeros1 + 1),
               (zeros2 + 1)), intercept = F)
a.init      <- fit0$coef

# Method 1, optimise lambdas by CV
if(MET==1){
    by.lambda <- length(seq(0.001,1,by=TOL2))
    opt.cv1 <- function(X){
        FIT <- APnorm(x=x, y=y, wei=wei, zeros1=
               zeros1, zeros2=zeros2, B=B,
               lambda1=X, lambda2=0, D.1=D.1,
               D.2=D.2, a.init=a.init, MON=MON,
               TOL=TOL1, MAX.IT=MAX.IT)
        return(FIT$cv)
    }
    lambda1.hat <- cleversearch(fn=opt.cv1,
                   lower=0.001, upper=1, ngrid=
                   by.lambda, logscale=F,
                   verbose=F)[[1]]

    opt.cv2 <- function(X){
        FIT <- APnorm(x=x, y=y, wei=wei, zeros1=
               zeros1, zeros2=zeros2, B=B, lambda1
               =lambda1.hat, lambda2=X, D.1=D.1,
               D.2=D.2, a.init=a.init, MON=MON,
               TOL=TOL1, MAX.IT=MAX.IT)
        return(FIT$cv)
    }
    lambda2.hat <- cleversearch(fn=opt.cv2,
                   lower=0.01, upper=1, ngrid=
                   by.lambda, logscale=F,
                   verbose=F)[[1]]
    FIT <- APnorm(x=x, y=y, wei=wei, zeros1=
           zeros1, zeros2=zeros2, B=B,
           lambda1=lambda1.hat, lambda2=
           lambda2.hat, D.1=D.1, D.2=D.2,
           a.init=a.init, MON=MON, TOL=TOL1,
           MAX.IT=MAX.IT)
}
```

```
    # Method 2, optimise lambdas by GCV
    if(MET==1){
        by.lambda <- length(seq(0.001,1,by=TOL2))
        opt.cv1 <- function(X){
            FIT <- APnorm(x=x, y=y, wei=wei, zeros1
                   =zeros1, zeros2=zeros2, B=B,
                   lambda1=X, lambda2=0, D.1=D.1,
                   D.2=D.2, a.init=a.init, MON=MON,
                   TOL=TOL1, MAX.IT=MAX.IT)
            return(FIT$gcv)
        }
        lambda1.hat <- cleversearch(fn=opt.cv1,
                       lower=0.001, upper=1, ngrid=
                       by.lambda, logscale=F,
                       verbose=F)[[1]]

        opt.cv2 <- function(X){
            FIT <- APnorm(x=x, y=y, wei=wei, zeros1
                   =zeros1, zeros2=zeros2, B=B,
                   lambda1=lambda1.hat, lambda2=X,
                   D.1=D.1, D.2=D.2, a.init=a.init,
                   MON=MON, TOL=TOL1, MAX.IT=MAX.IT)
            return(FIT$gcv)
        }
        lambda2.hat <- cleversearch(fn=opt.cv2, lower
                       =0.01, upper=1, ngrid=by.lambda,
                       logscale=F, verbose=F)[[1]]
        FIT <- APnorm(x=x, y=y, wei=wei, zeros1=zeros1,
               zeros2=zeros2, B=B,  lambda1=lambda1.hat,
               lambda2=lambda2.hat, D.1=D.1, D.2=D.2,
               a.init=a.init, MON=MON, TOL=TOL1,
               MAX.IT=MAX.IT)
    }


# Method 3, given lambdas
    if(MET==3){
        lambda1.hat <- lambda1
        lambda2.hat <- lambda2
          FIT <- APnorm(x=x, y=y, wei=wei, zeros1
                 =zeros1, zeros2=zeros2, B=B,
                 lambda1=lambda1.hat, lambda2
                 =lambda2.hat, D.1=D.1, D.2=D.2,
                 a.init=a.init, MON=MON, TOL=TOL1,
                 MAX.IT=MAX.IT)
    }
```

```
# fill
coef        <- FIT$a
cv          <- FIT$cv
gcv         <- FIT$gcv
df          <- FIT$df
dev         <- FIT$dev
h           <- FIT$h
y.hat       <- B%*%coef
lambda1.hat <- lambda1.hat
lambda2.hat <- lambda2.hat
sigma       <- sqrt(sum((y - y.hat)^2)/(m
               - sum(h)))
q           <- bdeg
P1          <- sqrt(lambda1.hat)*D.1
P2          <- sqrt(lambda2.hat)*D.2

# error bands
u         <- seq(xl, xr, length = m)
Bu        <- bbase(u, xmin, xmax, nseg, q)
Covb      <- solve(t(B) %*% B + t(P1) %*%
             P1 + t(P2) %*% P2)
Covz      <- sigma ^ 2 * (Bu %*% Covb %*%
             t(Bu))^2
seb       <- se * sqrt(diag(Covz))

# first derivative
b.d1     <-  bbase(x, xmin, xmax, nseg, q - 1)
beta     <-  as.vector(coef)
mu       <-  B %*% beta
vec      <-  beta
dseg     <-  (xmax-xmin)/nseg
temp1    <-  cbind(vec[2:length(vec)],
             c(vec[2:length(vec)-1]))
d1.coef  <-  temp1[,1] - temp1[,2]
d1       <-  q * (q * dseg)^(-1) * b.d1
             %*% d1.coef

# d1 error bands
u        <-  seq(xl, xr, length = m)
Bu       <-  bbase(u, xmax, xmin, nseg, q - 1)
dummy   <- cbind(Bu, rep(0, m))
Covb    <- solve(t(B) %*% B + t(P1) %*% P1 +
           t(P2) %*% P2)
Covz    <- sigma ^ 2 * q^2/((q*dseg)^2) * (dummy
           %*% Covb %*% t(dummy))^2
seb.d1  <- se * sqrt(diag(Covz))

# second derivative
```

```
    b.d2      <-  bbase(x, xmin, xmax, nseg, q - 2)
    temp2     <-  cbind(vec[3:length(vec)],
                  c(vec[3:length(vec)-1]),
                  c(vec[3:length(vec)-2]))
    d2.coef   <-  temp2[,1] - 2 * temp2[,2] + temp2[,3]
    d2        <-  (q - 1) * (dseg * (q - 1))^(-1) * q
                  * (q * dseg)^(-1) * b.d2 %*% d2.coef

    # d2 error bands
    u         <-  seq(xl, xr, length = m)
    Bu        <-  bbase(u, xmin, xmax, nseg, q - 2)
    dummy     <- cbind(Bu, rep(0, m), rep(0, m))
    Covb      <- solve(t(B) %*% B + t(P1) %*% P1 +
                 t(P2) %*% P2)
    Covz      <- sigma ^ 2 * ((q - 1) * (dseg *
                 (q - 1))^(-1) * q * (q*dseg)^(-1))^2
                 * (dummy %*% Covb %*% t(dummy))^2
    seb.d2    <- se * sqrt(diag(Covz))

    object    <- list(call = call, n = m, cv = cv,
                      gcv = gcv, lev = h,  df = df,
                      dev = dev, lambda1 = lambda1.hat,
                      lambda2 = lambda2.hat,
                      nseg = nseg, bdeg = bdeg,
                      pord = pord, x = x, y = y,
                      offset = offset,
                      w = as.vector(wei), y.hat
                      = y.hat,
                      coefficients = as.vector(coef),
                      sigma = sigma, d1 = d1, d2 = d2,
                      se = se, seb = seb,
                      seb.d1 =  seb.d1,
                      seb.d2 = seb.d2)
    class(object) <- "AP"
    return(object)
    }
}
```

The fitting function for a continuous response is APnorm:

```
APnorm <-
function(x, y, wei, zeros1, zeros2, B, lambda1,
         lambda2, D.1, D.2, a.init, MON,
         TOL, MAX.IT)
{
    w  <- wei
    # penalty
    P1 <- sqrt(lambda1)*D.1
    P2 <- sqrt(lambda2)*D.2
# initialise
tol <- 1
i <- 0
a <- a.init
a.old <- 10
# run
while(tol > TOL && i < MAX.IT){
        i <- i+1
        # update the coefficients
        a <- APnorm.update(x=x, y=y, wei=wei,
             B=B, P1=P1, P2=P2, zeros1=zeros1,
             zeros2=zeros2, a=a)
        # compute tol
        tol <- max(abs(a - a.old)/abs(a))
        # replace old coefficients
        a.old <- a
    }
if(i > (MAX.IT-1)) {
    warning(paste("Parameter estimates did NOT
            converge in", MAX.IT, "iterations.
            Increase MAX.IT in control."))
}
# fit
fit <- lsfit(rbind(B, P1, P2), c(y, zeros1,
       zeros2), wt = c(wei, (zeros1 + 1),
       (zeros2 + 1)), intercept = F)
a   <- fit$coef
mu  <- B%*%a
# diagonal of the hat-matrix
h <- hat(fit$qr)[1:length(w)]
# effective dimension
df <- sum(h)
# dev
dev <- sum(fit$residuals^2)
# cv
r <- (y - mu)/(1 - h)
cv <- sqrt((sum(r^2))/m)
# gcv
g <- (y - mu)/(1 - ((1/m))*sum(h))
gcv <- sqrt((sum(g^2))/m)
# aic
aic <- dev + 2 * df
# bic
bic <- dev + log(length(y)) * df
# fill
return(list(a=a, h=h, df=df, cv=cv, gcv=gcv,
       aic=aic, bic=bic, dev=dev))
}
```

The update function for continuous response is `APnorm.update`:

```
APnorm.update <-
function(x, y, wei, B, P1, P2, zeros1, zeros2, a)
{
    # expected values
    mu <- B%*%a
    # weights
    w <- wei
    # working response
    z <- wei*((y - mu)/mu)
    # fit
    fit <- lsfit(rbind(B, P1, P2), c(y, zeros1,
           zeros2), wt = c(wei, (zeros1 + 1),
           (zeros2 + 1)), intercept = F)
    # coefficients
    a <- matrix(fit$coef, ncol = 1)
    return(a)
}
```

The fitting function for count response is `APpois`:

```
APpois <-
function(x, y, offset, wei, zeros1, zeros2, B,
         lambda1, lambda2, D.1, D.2, a.init, MON,
         TOL, MAX.IT){
    # penalty
    P1 <- sqrt(lambda1)*D.1
    P2 <- sqrt(lambda2)*D.2
    # initialise
    tol <- 1
    i <- 0
    a <- a.init
    a.old <- 10
    # run
    while(tol > TOL && i < MAX.IT){
            i <- i+1
            # update the coefficients
            a <- APpois.update(x=x, y=y, offset =
                 offset, wei=wei, B=B, P1=P1, P2=P2,
                 zeros1=zeros1, zeros2=zeros2, a=a)
            # compute tol
            tol <- max(abs(a - a.old)/abs(a))
            # replace old coefficients
            a.old <- a
        }
    if(i > (MAX.IT-1)) {
        warning(paste("Parameter estimates did NOT
                converge in", MAX.IT, "iterations.
                Increase MAX.IT in control."))
    }
    # fit
    eta <- B%*%a
    mu <- exp(eta + offset)
    w <- wei*mu
    z <- wei*((y - mu)/mu + eta)
    fit <- lsfit(rbind(B, P1, P2), c(z, zeros1,
           zeros2), wt = c(wei, (zeros1 + 1),
           (zeros2 + 1)), intercept = F)
    a <- fit$coef
    # diagonal of hat matrix
    h <- hat(fit$qr)[1:length(w)]
    # effective dimension
    df <- sum(h)
    # deviance
    dev <- sum(fit$residuals^2)
    # cv
    r <- (y - mu)/(1 - h)
    cv <- sqrt((sum(r^2))/m)
    # gcv
    g <- (y - mu)/(1 - ((1/m))*sum(h))
    gcv <- sqrt((sum(g^2))/m)
    # aic
    aic <- dev + 2 * df
    # bic
    bic <- dev + log(length(y)) * df
    # fill
    return(list(a=a, h=h, df=df, cv=cv, gcv=gcv,
           aic=aic, bic=bic, dev=dev))
}
```

The update function for count response is `APpois.update`:

```
APpois.update <-
function(x, y, wei, offset, B, P1, P2, zeros1,
         zeros2, a)
{
    # expected values
    eta <- B%*%a
    mu <- exp(eta + offset)
    # weights
    w <- wei*mu
    # working response
    z <- wei*((y - mu)/mu + eta)
    # fit
    fit <- lsfit(rbind(B, P1, P2), c(z, zeros1,
           zeros2),  wt = c(wei, (zeros1 + 1),
           (zeros2 + 1)), intercept = F)
    # coefficents
    a <- matrix(fit$coef, ncol = 1)
    return(a)
}
```

A.2 Extracting Derivative Estimates from an spm Object

The spm routine in the **SemiPar** library of R allows the user to fit a semiparametric regression model to noisy data. Using the plot function one may plot the fit as well as derivative estimates along with estimated variability bands. However, fitted values for the derivative estimates and for the estimated standard errors must be extracted. The following code is for an example of data pairs (x_i, y_i), $i = 1, \ldots, n$ using a truncated polynomial basis of degree $p = 8$.

```
# fit
fit  <-  spm(y~f(x, cv = T, basis="trunc.poly", degree=8))
yhat    <-  fit$fit$fitted
cov.mat <-  fit$aux$cov.mat

# get ese of fit
dm     <-  dim(fit$fit$data)
theta  <-  c(fit$fit$coef$fixed, fit$fit$coef$random)
C.mat  <-  as.data.frame(fit$fit$data)[,-c(1,dm[2])]
X      <-  C.mat[,1:length(fit$fit$coef$fixed)]
Z.mat  <-  as.data.frame(fit$fit$data)[,-1]
Z      <-  Z.mat[,(length(fit$fit$coef$fixed)+1):(dm[2]-2)]
C      <-  as.matrix(cbind(X, Z))
se     <-  as.vector(sqrt(diag(C%*%cov.mat%*%t(C))))


# get fitted d1 and ese of d1
X.drv1      <-  X[,-length(fit$fit$coef$fixed)]
X.drv1[,2]  <-  2*X.drv1[,2]
X.drv1[,3]  <-  3*X.drv1[,3]
X.drv1[,4]  <-  4*X.drv1[,4]
X.drv1[,5]  <-  5*X.drv1[,5]
X.drv1[,6]  <-  6*X.drv1[,6]
X.drv1[,7]  <-  7*X.drv1[,7]
X.drv1[,8]  <-  8*X.drv1[,8]
X.drv1      <-  cbind(rep(0,length(x)), X.drv1)
Z.drv1      <-  8*Z^(7/8)
C.drv1      <-  as.matrix(cbind(X.drv1,Z.drv1))
drv1        <-  C.drv1%*%theta
se.d1  <- as.vector(sqrt(diag(C.drv1%*%cov.mat%*%t(C.drv1))))

# get fitted d2 and ese of d2
X.drv2      <-  X[,-c(length(fit$fit$coef$fixed)-1,
                length(fit$fit$coef$fixed))]
X.drv2[,1]  <-  2*X[,1]
X.drv2[,2]  <-  6*X[,2]
```

```
X.drv2[,3]  <-  12*X[,3]
X.drv2[,4]  <-  20*X[,4]
X.drv2[,5]  <-  30*X[,5]
X.drv2[,6]  <-  42*X[,6]
X.drv2[,7]  <-  56*X[,7]
X.drv2      <-  cbind(rep(0,length(x)), rep(0, length(x)),
                X.drv2)
Z.drv2      <-  7*Z.drv1^(6/7)
C.drv2      <-  as.matrix(cbind(X.drv2, Z.drv2))
drv2        <-  C.drv2%*%theta
se.d2       <-  as.vector(sqrt(diag(C.drv2%*%cov.mat%*%
                t(C.drv2))))
```