2009

KERNEL REGRESSION SUBJECT TO INTERVAL-CENSORED RESPONSES AND QUALITATIVE CONSTRAINTS

Xiuli Kang

KERNEL REGRESSION SUBJECT TO INTERVAL-CENSORED RESPONSES AND QUALITATIVE CONSTRAINTS

(Spine title: Constrained and Interval-censored Kernel Regression)

(Thesis format: Monograph)

by

Xiuli Kang

Graduate Program
in
Statistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

THE SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

Dr. W.John Braun

Examiners

Dr. Reg J. Kulperger

Dr. David R. Bellhouse

Dr. Adam Metzler

Dr. Thierry Duchesne

The thesis by

Xiuli Kang

entitled:

KERNEL REGRESSION SUBJECT TO INTERVAL-CENSORED RESPONSES AND QUALITATIVE CONSTRAINTS

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date

Chair of the Thesis Examination Board

ABSTRACT

This thesis is concerned with problems arising when one wants to apply flexible nonparametric local regression models to data when there is additional qualitative information. It is also concerned with nonparametric regression problems involving interval-censored responses. These problems are studied via asymptotic theory where possible and by simulation.

Iterated conditional expectation methods and local likelihood estimation for nonparametric interval-censored regression are developed. Simulation results show that local likelihood estimation is often superior to local regression estimators when observations have been imputed using either interval midpoints or iterated conditional expectations when the censoring intervals are wide or of varying width. When the intervals are smaller and of fixed width, none of the imputation approaches dominate the others.

Constrained data sharpening for nonparametric regression is applied to new situations such as where constraints are defined by convexity, concavity, and in terms of differential operators. Data sharpening is compared with competing kernel methods in terms of bias, variance and MISE. It is proved that the constrained data sharpening estimator has the same rate of convergence as the constrained weighting estimator of Hall and Huang (2001). Also, penalized data sharpening is proposed as a new form of constrained data sharpening. The sharpened responses can be computed analytically which makes the method very convenient, both for studying theoretically and for applying practically.

KEYWORDS: interval-censored responses, iterated conditional expectation, local likelihood, data sharpening, qualitative constraint, penalized data sharpening.

CO-AUTHORSHIP STATEMENT

The work of Chapter 3 was done jointly with Dr.John Braun and Dr. James Stafford. All other material included in this thesis was completed under the supervision of Dr. John Braun.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my most sincere gratitude to my supervisor Dr. John Braun for his taking me as his student. Without his inspiring and valuable instructions, constant encouragement and patience, this study would not have been completed.

I am also grateful to the secretaries Jennifer Dungavell, Jane Bai and Lisa Smith in the department office for helping the department to run smoothly and for assisting me in many different ways.

Special thanks are due to my thesis examiners Dr. David Bellhouse, Dr. Reg Kulperger, Dr. Adam Metzler and Dr. Thierry Duchesne.

Sincere thanks are extended to Dr. Ian McLeod and Dr. David Stanford as the previous graduate chairs who helped me to figure things out.

I also want to thank Dr. Peter Hall for his help and encouragement.

I also wish to express my appreciation to Dr. Hao Yu, who instructed me to finish my master project and made many valuable suggestions and gave constructive advice during all my study here.

I am very much thankful to all my schoolmates for their help in various ways.

Above and beyond all, my heartfelt gratitude to my entire family. They helped me to concentrate on completing this thesis and supported me mentally during my study in Canada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

In the forests covering most of Canada, the timing of the first appearance of aspen leaves signifies the end of a short spring wildfire season that occurs after the snow melts. Before the leaves appear, the forest floor can become very dry creating favourable conditions for ignition and spread. After the leaves appear, the sun cannot dry the forest floor as quickly, and the young leaves act as a form of fire retardant.

From a fire management point of view, it is desirable to predict the time of flush in advance. Times of first leaf appearance (referred to as the "aspen flush dates") were recorded at 74 weather stations in Alberta. Monitoring times were sporadic so often the exact time of flush was not observed directly: the time of flush is interval-censored. Since daily temperature measurements are available at all 74 stations, and since higher temperatures would be expected to correlate with an earlier flush date, it is natural to ask whether flush date can be predicted in advance, given earlier temperature information.

The plot of aspen flush date (in number of days after January 1) versus average March temperature is presented in Figure 1.1 for one of the years of data. It is unreasonable to expect a strictly linear relationship between flush date and March temperature. Thus, we face a problem of estimating a nonlinear functional relationship. However, we have additional qualitative information: the flush date should decrease as the temperature increases. Thus, we aim to model the interval-censored flush date as a nonlinear function of the earlier temperature, subject to a monotonicity constraint.

In this thesis, we consider observations on two or possibly more variables, where it is assumed that one of the variables may be predicted from the others. In some cases, the response variable is not directly observed but rather is reported in the form of an interval which contains the true observation. The other variables are assumed to be completely observed without error. We will focus on the case of one predictor variable though our methods can be generalized to the cases of more than one predictor variable.

To this kind of data, we seek to fit the regression model

$$y_i = g(x_i) + \varepsilon_i, i = 1, 2, ..., n \tag{1.1}$$

where the ε_i's are independent and identically distributed, and have marginal density f_ε, and they are independent of the predictors x_i. When the y_i's are interval-censored, they are observed as $I_1, I_2, \ldots, I_n$, where $y_i \in I_i = [L_i, R_i]$. The x_i's are assumed exactly observed without measurement error. We assume further that $g(x)$ is a smooth function, possessing at least a second derivative.

Monotonicity is only one of many possible qualitative restrictions that could be imposed on a regression function. We will consider other possibilities as well, such as convexity, periodicity and so on.

Data sharpening is one way to impose such restrictions. It has not been studied thoroughly so one of our objectives is to study the properties of this methodology and to check its performance in a variety of situations.

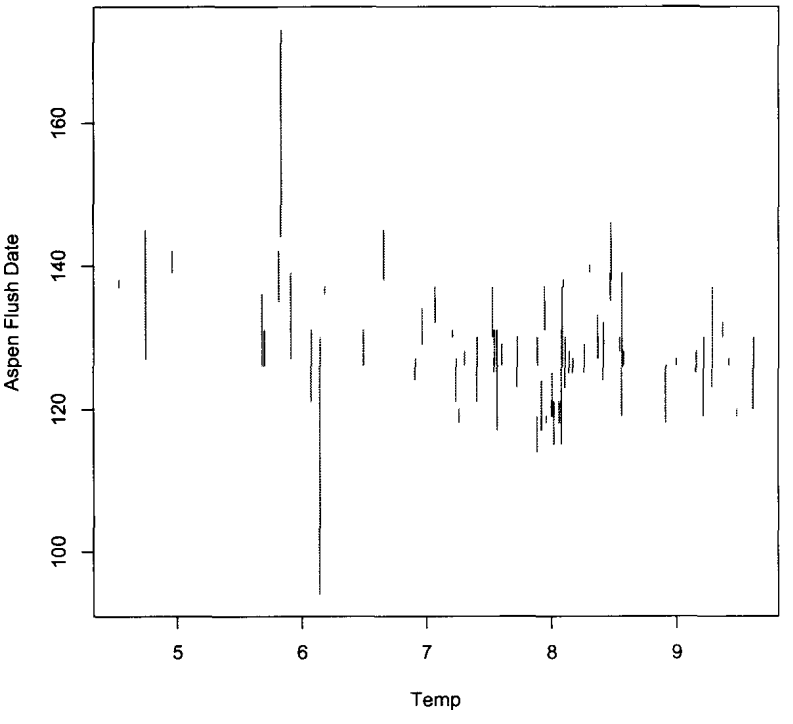

Figure 1.1: Aspen flush dates versus average March temperature (at 74 weather stations in Alberta, 2006). Vertical bars correspond to the time intervals in which flush is thought to have occurred.

The remainder of this thesis is constructed as follows. Chapter 2 provides a brief summary of background material on kernel-based nonparametric regression. It briefly surveys the literature on interval-censoring and on nonparametric regression with qualitative constraints. In Chapter 3, imputation via conditional expectation and local likelihood methods are contrasted with a midpoint imputation method for bivariate regression involving interval-censored responses. Although the methods can be extended in principle to higher order polynomials, our focus is on the local constant case. Chapter 4 develops a kernel regression method based on data sharpening which can impose many kinds of qualitative constraints and incorporate functional information. Simulation results show that imposing qualitative constraints can lead to substantial improvements in accuracy. We derive the asymptotic rate of convergence for this class of estimators. A reweighting procedure suggested in the literature for monotonic regression is also considered, adapted for use with more general forms of constraints and compared with the data sharpening procedure. In Chapter 5, we show that data sharpening can be used as a convenient way to subject kernel regression to roughness penalties as well as many other penalties such as may arise in functional data analysis. This form of data sharpening is highly tractable, allowing us to study its properties in a more direct way than in other forms of data sharpening. The final chapter summarizes the thesis and outlines a number of research questions that arise from the work.

Chapter 2

PRELIMINARIES AND BACKGROUND MATERIAL

2.1 Kernel Regression

In this thesis, we consider the nonparametric regression problem for bivariate data of the form (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ where it is assumed that

$$y_i = g(x_i) + \varepsilon_i.$$

The ε_i's are independent of each other and of the x_i's, and have mean 0 and variance σ^2. The function $g(x)$ is assumed to be smooth. The design points x_i may have been fixed in advance or generated randomly.

Kernel methods, such as local polynomial regression (e.g., Fan and Gijbels (1994)) are popular methods for estimating $g(x)$. Briefly, a polynomial model of degree p is fit to data in a neighbourhood of x using weighted least squares. Local regression coefficients, $\beta_x = [\beta_{x0}, \beta_{x1}, \ldots, \beta_{xp}]^T$, are chosen to minimize

$$\sum_{i=1}^{n} \left\{ y_i - \sum_{j=0}^{p} \beta_{xj}(x_i - x)^j \right\}^2 K_h(x_i - x) \tag{2.1}$$

for each value of x, where $K_h(x) = K(x/h)/h$. Here $K(x)$ is usually a symmetric probability density function, and h is the smoothing parameter – the bandwidth which determines the size of the neighbourhood of x. For example, the normal density with mean 0 and standard deviation h is often used as a kernel function, though there are other densities which provide slightly better performance.

The local least-squares solution is given by

$$\widehat{\beta}_x = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$$

where $\mathbf{y}$ is the n-vector with ith element y_i, $\mathbf{X}$ is an $n \times (p+1)$ matrix with (i,j)th element $(x_i - x)^j$, and $\mathbf{W}$ is the $n \times n$ diagonal matrix with (i,i)th element $K_h(x_i - x)$. The estimator for $g(x)$ is given by

$$\widehat{g}(x) = \widehat{\beta}_{0x}.$$

This technique originated in work by Nadaraya (1964) and Watson (1964) who considered the $p = 0$ (local constant) case. Fan & Gijbels (1996) advocate the use of $p = 1$ (local linear) or other odd positive integers. They also note the use of a binning procedure, which, when used with the Fast Fourier Transform (FFT), leads to fast, efficient computations of estimates. The methods studied in this thesis are amenable to this procedure.

A convenient kernel function is the normal or Gaussian kernel. Often, theoretical results are easier to obtain for compactly supported kernel functions, such as the biweight function which provides very similar performance to the normal kernel function.

Kernel regression estimators can also be derived from local likelihood considerations. We refer to Loader (1999) for details of the theory and methods. For regression with normally distributed errors, local log-likelihood modelling leads to maximization of

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - p(x_i))^2 K_h(x - x_i) \tag{2.2}$$

with respect to the coefficients of a polynomial $p(x)$; for example, when $p(x) = \beta_0$, maximizing (2.2) with respect to β_0 leads to the Nadaraya-Watson regression estimator

$$\hat{g}(x) = \hat{\beta}_0 = \frac{\sum K_h(x - x_i) y_i}{\sum K_h(x - x_i)}. \tag{2.3}$$

This entire class of estimators is linear in the responses, i.e.

$$\widehat{g}(x) = \sum_{i=1}^{n} A_i(x) y_i$$

for a sequence of functions $\{A_i(x)\}$ which are related to the kernel and are functionally

independent of the y's. For example, when $p = 0$,

$$A_i(x) = \frac{K_h(x - x_i)}{\sum_{j=1}^{n} K_h(x - x_j)}, \tag{2.4}$$

since in that case, it is well known that

$$\widehat{g}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{j=1}^{n} K_h(x - x_j)}.$$

Although the quality of the resulting estimate does not depend crucially on the kernel, choice of bandwidth is important (e.g. Wand and Jones (1995)). Several methods of bandwidth selection have been suggested, including cross-validation, advocated for example by Loader (1999). Others, including Wand and Jones (1995), have recommended a direct-plug-in approach which is an attempt to minimize the asymptotic mean integrated squared error (AMISE) of the estimator; this is done by "plugging-in" estimates of higher order derivatives of the regression function and an estimate of the error variance into an expression for the AMISE, and applying calculus. The selector of Ruppert *et al.* (1995) is an example; it is conveniently implemented in the R package *KernSmooth* (Wand and Ripley (2009)), and it will be employed at various points in this thesis.

A fixed bandwidth works best only when the second derivative of the regression function does not change rapidly. It is well-known that peaks in the regression function are usually underestimated (e.g. Choi et al, 2000). If the bandwidth is taken small enough to minimize this kind of bias, the resulting estimator often exhibits excessive fluctuations in regions where the regression function should be flat. Allowing the bandwidth to vary with x is one way to solve this problem, in principle, but it leads to the practical problem of determining this adaptive bandwidth. Doksum *et al.* (2000) provide a solution, but the numerical efficiency of the FFT is no longer available.

Choi *et al.* (2000) propose a procedure they refer to as "data sharpening". In fact, their paper describes a number of techniques for adjusting data before applying a Nadaraya-Watson estimator or local linear estimator. The bias of the resulting estimate is reduced to some extent. The first approach involves slightly shifting the design points (i.e., the values of the explanatory variable) so that they become more concentrated in places where the original design density had been relatively

low and more sparse where the concentration had been high. The second approach, which involves adjusting response variables rather than explanatory variables, reduces the bias of the Nadaraya-Watson estimator by an order of magnitude; and a third method, which perturbs both explanatory and response variables before substituting them back into the conventional Nadaraya-Watson estimator, achieves bias reductions of the same order as the second method, but in a different manner.

2.2 Interval-censoring

Linear regression with censored responses has been considered in the papers by Schmee and Hahn (1979), Miller (1976) and Buckley and James (1979).

The literature on nonparametric regression when responses are interval-censored is not extensive. A nonparametric maximum likelihood method has been employed by Rabinowitz *et al.* (1995) to linearly model log survival times with several covariates where the survival times were interval-censored and few assumptions were made about the error distribution.

A simple way of handling interval-censored responses is to replace each interval with its midpoint. A local regression procedure is then applied to these imputed responses as in the procedure at (2.3):

$$\widehat{g}(x) = \frac{\sum_i \frac{R_i+L_i}{2} K_h(x - x_i)}{\sum_i K_h(x - x_i)}. \tag{2.5}$$

Midpoint imputation is only reasonable when the intervals are fairly short. When the width of the interval is large, we may run into problems. Law and Brookmeyer (1992) show that the statistical properties of midpoint imputation depend strongly on the width of the interval between monitoring times. They note that midpoint imputation, used to estimate the regression parameter in a proportional hazards model, might result in a biased estimate if the intervals are wide and varied. Also the standard error of the estimator is underestimated since midpoint imputation assumes that the failure times are exactly known when in fact they are not (see Kim (2003)). Similar discussion on midpoint imputation approach can found on P.34-37 in Sun (2006).

Local likelihood methods for estimation of the hazard function with interval-censored responses have been developed by Betensky *et al.* (1999) and Betensky *et al.* (2002). Local likelihood methods also have been adapted for density estimation with

interval-censored data using conditional expectation and smoothed EM algorithm by Braun *et al.* (2005). The iteration is based on the fixed point equation

$$\hat{f}(x) = \frac{1}{n}\sum_{i-1}^{n} \mathrm{E}_{\widehat{f}}\left[\frac{1}{h}K\left(\frac{x-X}{h}\right)\middle| X \in I_i\right]. \tag{2.6}$$

Here, the random variables X have not been observed directly; instead the data are in the form of intervals: $I_1, \ldots, I_n$, each of which contains the true underlying random quantity.

2.3 Incorporating Qualitative Information

There is often interest in estimating the unknown mean function $g(x)$ not only with smoothness but also subject to a given set of qualitative constraints, such as monotonicity, convexity, and so on.

2.3.1 Constrained Weighting

Inspired by the biased-bootstrap techniques of Hall and Presnell (1999), Hall and Huang (2001) have proposed a smooth, monotonically constrained kernel-type estimator, which involves tilting the empirical distribution of the responses by the least possible amount, subject to enforcement of the constraint. We shall refer to this approach as the "constrained weighting" method.

The basic idea is to view the nonparametric regression estimator in the form

$$\widehat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} nA_i(x)y_i \tag{2.7}$$

Then the re-weighted form of (2.7) is

$$\widehat{g}(x|p) = \sum_{i=1}^{n} p_i nA_i(x)y_i, \tag{2.8}$$

where $p = (p_1, \cdots, p_n)$ is a probability distribution on the set $x_1, \cdots, x_n$. In order to impose the nondecreasing constraint, Hall and Huang (2001) choose $p = \hat{p}$ to minimize the distance $D(p)$ between p and the uniform distribution, $p_u = (1/n, \cdots, 1/n)$, subject to $\widehat{g}'(.|p) \geq 0$. The Kullback-Leibler divergence measure introduced by Cressie and Read (1984) has been used to compute to $D(p)$.

In order to make the problem easier and to reduce the restriction on the weights p_i's, the L_2 metric is considered by Racine and Parmeter (2009). Specifically, p is chosen to minimize

$$D(p) = \sum_{i=1}^{n} (p_i - \frac{1}{n})^2$$

subject to any number of constraints which take the form

$$\sum_{i=1}^{n} nB_i(x)y_i p_i - C(x) \geq 0 \tag{2.9}$$

or

$$\sum_{i=1}^{n} nB_i(x)y_i p_i - C(x) = 0. \tag{2.10}$$

Here $B_i(x)$ is defined as follows:

$$B_i(x) = \sum_{s \in \mathbf{S}} D_i(x) A_i^{(s)}(x),$$

where $\mathbf{S}$ is a finite set of nonnegative integers, $A_i^{(s)}(x)$ is the s-th derivative of the function $A_i(x)$ and $D_i(x)$ and C(x) are functions of x. The sum of the weights is 1, i.e., $\sum_{i=1}^{n} p_i = 1$. This leads to a quadratic programming problem with linear constraints. The probability vector p can be calculated using the `solve.QP()` function in the *quadprog* R library (Turlach and Weingessel (2007)).

2.3.2 Data Sharpening

The data sharpening procedure of Braun and Hall (2001) provides another way of imposing qualitative constraints. The data are perturbed in a minimal way so that the local regression estimator applied to the perturbed data satisfies the constraints at a grid of m points in the domain of the regression function.

To achieve monotonicity, one might impose the constraint $g'(x) \geq 0$. Let $y_i^\star, i = 1, 2, \ldots, n$ denote the perturbed responses, defined to minimize

$$\sum_{i=1}^{n} |y_i - y_i^\star|^p$$

subject to

$$\widehat{g}'(z_j) \geq 0 \quad j = 1, 2, \ldots, m$$

where $\widehat{g}(x) = \sum_{i=1} A_i(x) y_i^\star$. The z_j's may differ from the x_i's. In this chapter, the choice of z's has not been studied at all. This is addressed in Chapter 5.

This is a convex programming problem with linear constraints, thus, possessing a unique solution when $p > 1$. When $p = 2$, this is a quadratic programming problem which is straightforward to solve with publicly available software, even for fairly large values of n. For these reasons, we will focus on the quadratic case for the remainder of the thesis.

Data sharpening has the same goal as that of the constrained weighting scheme of Hall and Huang (2001); the data are pre-processed, so as to enhance the performance of a statistical procedure while preserving the advantages of relatively simple, low-order techniques, when substituting the new data into the conventional estimator.

The original idea behind data sharpening is due to Choi and Hall (1999) who were seeking to reduce bias in density estimation. The idea also has roots in the bias-bootstrap techniques of Hall and Presnell (1999) as does the approach of constrained weighting of Hall and Huang (2001). The difference between data sharpening and constrained weighting is that the former alters the data while keeping the weights associated with each point fixed, whereas the latter changes the weights associated with each data point, but keeps the data points fixed.

2.4 Other Methods and Tests

The earliest constrained nonparametric regression estimators subject to monotonicity is isotonic regression, proposed by Brunk (1955). The estimated curve is not smooth. A smooth version has been proposed by Mukerjee (1988) and Mammen (1991*a*). An alternative estimator, which is a slight modification of an estimator introduced by Friedman and Tibshirani (1984) was proposed by Mammen (1991*b*).

Dette *et al.* (2006) have suggested another method to impose monotonicity on a nonparametric regression function. This approach estimates the regression function by inverting a particular monotonic function.

Spline-based methods provide an alternative way to estimate the regression mean function $g(x)$ under the imposition of qualitative constraints. Some references here include Ramsay (1988), Dierckx (1980), Mammen (1991*b*), Mammen and Thomas-Agnan (1999) and Mayer (2008).

Statistical tests for the monotonicity of a regression function have also been proposed. This literature is reviewed briefly by Racine and Parmeter (2009). Testing the validity of other qualitative constraints has also been considered. For example, Yatchew and Hardle (2006) employ a residual-based test to check for monotonicity and convexity. The test of Yatchew and Bos (1997) is closest in spirit to the method we apply in the thesis, having the ability to test general smoothness constraints, though it is based on a series estimator instead of a kernel estimator.

Chapter 3

LOCAL REGRESSION WHEN THE RESPONSES ARE INTERVAL-CENSORED

3.1 Introduction

The purpose of the present chapter is to propose and study methods for estimating a regression function in the presence of interval-censoring. One of the proposals applies the iterated conditional expectation method of Braun *et al.* (2005) to intervals estimated to contain the regression errors. The other approach is to apply conditional expectation to the local likelihood directly, assuming a normal error distribution. Both approaches are compared with the midpoint method using simulation.

This chapter will proceed as follows. In Section 3.2, we will describe the approaches that we are studying. Section 3.3 is concerned with estimation of the residual variance. A simulation study will be carried out in Section 3.4, followed by a discussion of the results in Section 3.5. In Section 3.6, some illustrative examples will be provided.

3.2 Methods

As mentioned in Chapter 2, a simple method to handle interval-censored responses in nonparametric regression is midpoint imputation. However, it was pointed out there that under many circumstances, midpoint imputation can be very inaccurate, see Law and Brookmeyer (1992), Kim (2003)) and Sun (2006)

In this section, we will propose two methods for handling nonparametric regression problems where the responses are interval-censored. We will consider them in the context of the local constant (Nadaraya-Watson) regression estimator.

3.2.1 Iterated Conditional Expectation Imputation

Our first proposal is to impute interval-censored responses using $\mathrm{E}[Y_i|Y_i \in I_i]$. Applying the local constant regression estimator then leads to the form:

$$\widehat{g}(x) = \frac{\sum_i \widehat{\mathrm{E}}[Y_i|Y_i \in I_i]K_h(x - x_i)}{\sum_i K_h(x - x_i)}. \tag{3.1}$$

The conditional expectation must itself be estimated, because the conditional distribution of Y given $Y \in I_i$ is generally unknown. A previously obtained estimate $g(x)$ can be used to estimate $\mathrm{E}[Y_i|Y_i \in I_i]$, since

$$\begin{aligned} \mathrm{E}[Y_i|Y_i \in I_i] &= \frac{\int_{I_i} y f_\varepsilon(y - g(x_i))dy}{\int_{I_i} f_\varepsilon(y - g(x_i))dy} && (3.2) \\ &= g(x_i) + \frac{\int_{I_{\varepsilon_i}} z f_\varepsilon(z)dz}{\int_{I_{\varepsilon_i}} f_\varepsilon(z)dz} && (3.3) \\ &= g(x_i) + \mathrm{E}[\varepsilon_i|I_{\varepsilon_i}], && (3.4) \end{aligned}$$

where $I_{\varepsilon_i} = [L_i - g(x_i), R_i - g(x_i)]$.

We will consider two cases: one where specific parametric distribution assumptions on the error are made and one where a nonparametric estimate of the error density is used. Although we focus on the nonparametric assumption for $g(x)$, in the literature there are numerous examples of papers where $g(x)$ is parametric but no specific form for error density is assumed, e.g., Komrek *et al.* (2005), Yu *et al.* (2006) and Zeng *et al.* (2006), etc.

When the error density is assumed normal, the expectation of Y_i conditional on the censored interval I_i is

$$\mathrm{E}[Y_i|I_i] = g(x_i) - \sigma\frac{\phi(R_i') - \phi(L_i')}{\Phi(R_i') - \Phi(L_i')}, \tag{3.5}$$

where $I_i = [L_i, R_i]$, $L_i' = \frac{L_i - g(x_i)}{\sigma}$ and $R_i' = \frac{R_i - g(x_i)}{\sigma}$.

It is only necessary to estimate the dispersion parameter σ^2. Various methods are available for estimating σ; these are described and compared in the next section.

When we do not assume any parametric form for the errors, a local constant

regression estimate for $g(x)$ is based on the fact that

$$g(x) = \frac{\sum_i \frac{\int_{I_i} y f_\varepsilon(y-g(x_i))dy}{\int_{I_i} f_\varepsilon(y-g(x_i))dy} K_h(x-x_i)}{\sum_j K_h(x-x_j)}, \tag{3.6}$$

where the relation

$$f_\varepsilon(z) = \frac{1}{n}\sum_i \frac{\int_{I_{\varepsilon_i}} K_h(z-\omega) f_\varepsilon(\omega)d\omega}{\int_{I_{\varepsilon_i}} f_\varepsilon(\omega)d\omega} \tag{3.7}$$

leads to the local likelihood (constant) density estimate of the errors given by Braun *et al.* (2005).

The algorithm starts with an initial error density estimate and an initial regression function (chosen from interval midpoints, for example). Iterate the second relation (i.e. the density) to convergence at each iteration of the first relation. The specific algorithm we have employed is as follows:

1. Compute the interval midpoints of the Y_i's: y_i^0.

2. Compute the initial regression function estimate:

$$\hat{g}_o(x) = \frac{\sum_i K_h(x-x_i) y_i^0}{\sum_i K_h(x-x_i)}.$$

   Choose h, using a direct-plug-in method Ruppert *et al.* (1995), for example.

3. Estimate the noise density using the iterated conditional expectation procedure given by Braun *et al.* (2005).

   (a) The interval-censored errors are estimated as $I_{\varepsilon_i} = I_i - \hat{g}_o(x_i)$.

   (b) The error density is estimated by the fixed point of

$$\begin{aligned} \widehat{f}_\varepsilon(z) &= \frac{1}{n}\sum_{i=1}^n \widehat{\mathrm{E}}[K_h(z-\varepsilon_i)|\varepsilon_i \in I_{\varepsilon_i}]. \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\int_{I_{\varepsilon_i}} K_h(z-w)\widehat{f}_\varepsilon(w)dw}{\int_{I_{\varepsilon_i}} \widehat{f}_\varepsilon(w)dw}, \end{aligned}$$

   where the initial guess at $\widehat{f}_\varepsilon(w)$ is taken as a uniform density on the range of the observed intervals. In this step, the integrals over intervals are

computed using trapezoid rule, and direct plug-in methodology is used to select the bandwidth of a kernel density estimate.

4. Compute the conditional expectation, $\widehat{\mathrm{E}}[\varepsilon_i|I_{\varepsilon_i}] = \frac{\int_{I_{\varepsilon_i}} z\widehat{f}_\varepsilon(z)dz}{\int_{I_{\varepsilon_i}} \widehat{f}_\varepsilon(z)dz}$, of the ε_i's.

5. Set $\hat{\mathrm{E}}[y_i|I_i] = \hat{g}(x_i) + \widehat{\mathrm{E}}[\varepsilon_i|I_{\varepsilon_i}]$, where $\hat{g}(x_i) = \frac{\sum_i \widehat{\mathrm{E}}(y_i|I_i)K_h(x-x_i)}{\sum_i K_h(x-x_i)}$. The bandwidth h is recomputed using a direct plug-in approach. The expectation integral is computed using the `integrate()` function in R (Team (2008)).

6. Set $\hat{g}_0(x) = \hat{g}(x)$ and return to step (3) or stop when iteration has achieved convergence.

Note that the kernels do not have to be the same, nor the bandwidth. A density other than the uniform can be used for the initial guess at $\widehat{f}_\varepsilon(w)$, though its support should contain the range of the observed intervals.

3.2.2 Local Likelihood

An alternative procedure which does not require direct imputation can be obtained if we model the censoring mechanism directly in the local likelihood. Here, we are making explicit use of an assumption that the censoring mechanism is independent of the responses.

If we assume normal errors, then the log-likelihood function for model (1.1) is given by

$$l(\mu_x, \sigma^2) = \sum_{i=1}^{n} \log\left(\Phi(\frac{R_i - p_{x_i}}{\sigma}) - \Phi(\frac{L_i - p_{x_i}}{\sigma})\right) \tag{3.8}$$

when $g(x)$ can be modelled as a polynomial p_x. We can make this a local log-likelihood function by inserting a kernel:

$$\sum_{i=1}^{n} \log\left(\Phi(\frac{R_i - p_{x_i}}{\sigma}) - \Phi(\frac{L_i - p_{x_i}}{\sigma})\right) K_h(x - x_i). \tag{3.9}$$

Maximizing this expression with respect to the coefficients of the polynomial p_x gives the local polynomial estimator of $g(x)$. This estimator is consistent even if $g(x)$ is not a polynomial, but is sufficiently smooth.

When we set $p_x = \beta_0$, maximizing (3.9) with respect to β_0 and σ gives the local constant likelihood estimator for $g(x)$.

Comparing this method to that of Rabinowitz *et al.* (1995) we note that they allowed for non-normal errors, while we are allowing for a nonlinear regression function. By replacing the normal cdf above with an empirical cdf, we would obtain a generalization of their method, but that is beyond the scope of the present chapter.

Again, we select bandwidths using the direct plug-in approach of Ruppert *et al.* (1995) using interval midpoints as the responses.

3.3 Estimation of σ

A critical component of the parametric version of the iterated conditional expectation imputation approach is the estimation of σ. For this reason, as well as the fact that uncertainty estimates for the other estimators will require some form of standard deviation estimate, we devote the present section to a study of some of the possible estimators of σ.

We consider three estimators.

3.3.1 Method 1

The first was proposed in Schmee and Hahn (1979) for simple linear regression with interval-censored responses, $y_i \in [L_i, R_i]$. The estimate arises from an iteration on the imputed responses. Starting with an initial guess at the responses y_i, $i = 1, \ldots, n$, the intercept β_0, slope β_1, and residual variance σ^2 are repeatedly estimated leading to successive adjustments of the imputed responses as follows:

1. Set $L_i' = (L_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)/\widehat{\sigma}$ and
   $R_i' = (R_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)/\widehat{\sigma}$.
2. Set $c_i = \frac{\phi(R_i') - \phi(L_i')}{\Phi(R_i') - \Phi(L_i')}$.
3. Set $y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i - \widehat{\sigma} c_i$.

3.3.2 Method 2

We can develop a second estimator in the following way. Note that for simple regression with complete data, an unbiased estimator for σ^2 is given by

$$\frac{1}{n-2} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2$$

where $\widehat{\varepsilon}_i$ is the ith residual.

Replace $\widehat{\varepsilon}_i^2$ by $\mathrm{E}[\varepsilon_i^2 | Y_i \in I_i]$:

$$\widehat{\sigma^2} = \frac{1}{n-2} \sum_{i=1}^{n} \mathrm{E}[\varepsilon_i^2 | Y_i \in I_i]$$

We then use

$$\mathrm{E}[\varepsilon_i^2 | Y_i \in I_i] = \frac{\int_{L_i - \mathrm{E}[y_i]}^{R_i - \mathrm{E}[y_i]} \varepsilon^2 \phi(\varepsilon) d\varepsilon}{\int_{L_i - \mathrm{E}[y_i]}^{R_i - \mathrm{E}[y_i]} \phi(\varepsilon) d\varepsilon}$$

and integration by parts to obtain the relation

$$\widehat{\sigma^2} = \frac{\sigma^2}{n-2} \sum_{i=1}^{n} \left[1 - \frac{R_i' \phi(R_i') - L_i' \phi(L_i')}{\Phi(R_i') - \Phi(L_i')} \right]$$

and where, again, $L' = \frac{L - \widehat{y}_i}{\sigma}$ and $R' = \frac{R - \widehat{y}_i}{\sigma}$.

This last relation can be turned into a self-consistency algorithm or fixed-point iteration for an estimator, upon replacing all occurrences of σ with $\widehat{\sigma}$.

3.3.3 Method 3

Another self-consistency algorithm can be developed using the fact that

$$\mathrm{Var}(Y_i) = \mathrm{Var}(\mathrm{E}[Y_i | I_i]) + \mathrm{E}[\mathrm{Var}(Y_i | I_i)].$$

We thus consider an estimator of the form

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (\widehat{y}_i - \widehat{E}[\widehat{y}_i])^2 + \frac{1}{n} \sum_{i=1}^{n} \widehat{\sigma}_i^2.$$

where

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i - \widehat{\sigma} c_i$$

and

$$\widehat{E}[\widehat{y}_i] = \widehat{\beta}_0 + \widehat{\beta}_1 x_i - \frac{1}{n} \widehat{\sigma} \sum_{i=1}^{n} c_i$$

	Method	$n = 25$	$n = 50$	$n = 100$
$\alpha = .2$	1	.00341	.00231	.00182
	2	.00127	.00063	.00033
	3	.00139	.00066	.00034
$\alpha = .5$	1	.0142	.01159	.01010
	2	.0027	.00122	.00061
	3	.0034	.00139	.00064

Table 3.1: Mean-squared error estimates when estimating σ using the three methods applied to linear regression data where $\sigma = .2$

where, again,

$$c_i = \frac{\phi(R_i') - \phi(L_i')}{\Phi(R_i') - \Phi(L_i')}$$

and

$$\widehat{\sigma}_i^2 = \widehat{\sigma}^2 \left(1 - \left\{ \frac{\phi(R_i') - \phi(L_i')}{\Phi(R_i') - \Phi(L_i')} \right\}^2 - \frac{R_i'\phi(R_i') - L_i'\phi(L_i')}{\Phi(R_i') - \Phi(L_i')} \right).$$

The algorithm presented above is reminiscent of the approach used in generalized estimation equation (GEE) (see, Liang and Zeger (1986)) inference: there the variance/covariance parameters are nuisance parameters and are fixed so that mean parameters can be estimated, which yields new residuals that can be used to estimate new values of the covariance parameters, and these two processes are iterated until convergence.

3.3.4 Comparing the Methods by Simulation

In order to determine which of these approaches to use when estimating σ, we conducted a small simulation experiment in which 1000 bivariate samples each of size 25, 50 and 100 were generated with normal errors, an intercept of 0 and a slope of 1, i.e., $g(x) = x$. Two values of the error standard deviation were considered: .2 and 3.

We interval-censored the responses by adding and subtracting independent random exponential amounts to each response; mean values (α) for the exponential variates considered were .2 and .5 for the cases when $\sigma = .2$, and a mean value of .8 was considered when $\sigma = 3$. Results in the form of estimated mean-squared errors are given in Tables 3.1 and 3.2.

Based on these simulations, we observe that all three methods give broadly similar results, but that the second approach is consistently slightly better than either

	Method	$n = 25$	$n = 50$	$n = 100$
$\alpha = .8$	1	.2273	.1056	.05498
	2	.2105	.0994	.04968
	3	.2233	.0996	.05023

Table 3.2: Mean-squared error estimates when estimating σ using the three methods applied to linear regression data where $\sigma = 3$

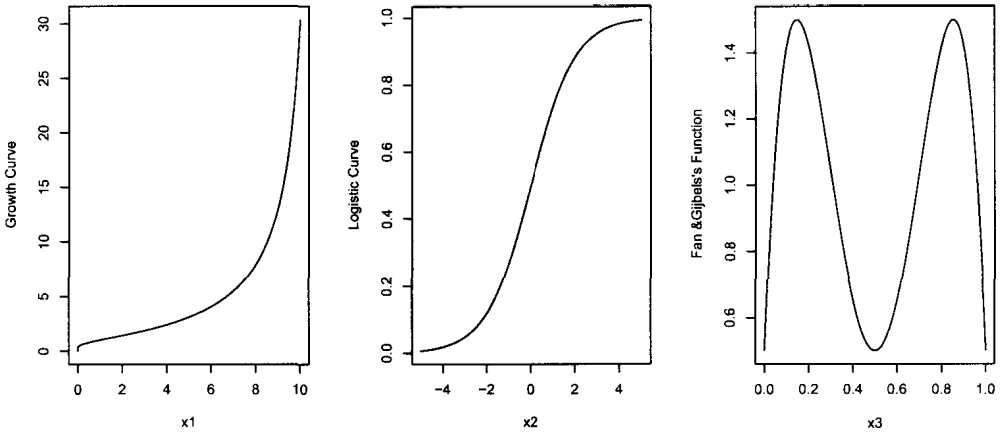

Figure 3.1: The three target functions considered in the simulation study

of the two approaches. Thus, we have adapted the second method to the kernel regression problem we are considering.

3.4 A Simulation Comparison of the Imputation Approaches

In this section, we describe a simulation study that was conducted in order to compare the four estimation methods described in Section 3.2. The four methods to be compared are midpoint imputation, normal-based and nonparametric iterated conditional expectation imputation and the local likelihood approach.

3.4.1 Study Designs

We considered the following target functions (see Figure 3.1) in our simulation studies:

1. Growth Curve: He and Shi (1998)

$$Y_i = \frac{1}{1 - .42(\log(x_i))} + 1.5\varepsilon_i,$$
$$X_i \sim \text{i.i.d.} \quad \text{Uniform}[0, 10], \quad \varepsilon_i \sim \text{i.i.d.} \quad N(0, 1),$$

where X_i and ε_i were independent.

2. Logistic Curve: He and Shi (1998)

$$Y_i = \frac{e^{x_i}}{1 + e^{x_i}} + 0.1\varepsilon_i,$$
$$X_i \sim \text{i.i.d.} \quad \text{Uniform}[-5, 5], \quad \varepsilon_i \sim \text{i.i.d.} \quad N(0, 1),$$

where X_i and ε_i were independent.

3. Polynomial Function: Fan and Gijbels (1994).

$$Y_i = 4.5 - 64X_i^2(1 - X_i)^2 - 16(X_i - .5)^2 + 0.25\varepsilon_i,$$
$$X_i \sim \text{i.i.d.} \quad \text{Uniform}[0, 1], \quad \varepsilon_i \sim \text{i.i.d.} \quad N(0, 1),$$

where X_i and ε_i were independent.

We considered two forms of censoring: one with exponentially distributed interval widths and one with fixed interval widths (for the growth curve target only).

For the exponential intervals, the upper and lower interval limits were obtained by evaluating Y_i in model (1.1) and adding and subtracting independent random variates from the exponential distribution with a specific rate:

$$L_i = Y_i - C_{i1}$$

and

$$R_i = Y_i + C_{i2}$$

where C_{i1} and C_{i2} are independent exponential random variables with rate λ. This form of interval would arise, for example, when evaluating current status at monitoring times which follow a stationary Poisson process. Note that small values of λ correspond to relatively wide censoring intervals.

Fixed interval limits were obtained by evaluating the product of w and the ceiling or floor function of Y_i/w, in which w is the fixed width of the interval:

$$L_i = [Y_i/w]w$$

$$R_i = [Y_i/w + 1]w.$$

We simulated 1000 sets of data for each target function. In each case, $n = 50$, 100 and 200 observations were simulated. The design points x and the responses y were obtained from the above three settings. For the growth curve function, the regression curve was estimated at 9 equally spaced grid points (not the same as the design points) on the interval $[1, 9]$. For the logistic curve function, the regression curve was estimated at 9 equally spaced grid points (not the same as the design points) on the interval $[-4, 4]$. For the polynomial function, the regression curve was estimated at 9 equally spaced grid points (not the same as the design points) on the interval $[0.1, 0.9]$. For the growth curve, we used values of $\lambda \in \{0.2, 1, 5\}$ for the exponentially distributed censoring mechanism, and we used $\lambda \in \{0.8, 2, 8\}$ for the logistic curve. For the polynomial target, we used a censoring mechanism which is similar in spirit to that used by Fan and Gijbels (1994). In particular, the censoring intervals were exponentially distributed with mean $\in \{(c(x)), 3c(x)/5, 3c(x)/10\}$, where

$$\begin{aligned} c(x) &= (1.25 - |4x - 1|), \quad \text{if} \quad 0 \leq x \leq 0.5, \\ &= (1.25 - |4x - 3|), \quad \text{if} \quad 0.5 \leq x \leq 1. \end{aligned}$$

For the growth curve, we also censored responses using the fixed-width mechanism, where the width w took values in $\{0.5, 2, 4\}$.

For each set of data, we calculated the local constant regression estimate using midpoint imputation (Page.7) , normal and nonparametric iterated conditional expectation imputation (Page.12-14), and the local likelihood approach (Page.15). All computations were done in R (Team (2008)).

3.4.2 Comparisons of Bias, Variance and MSE

The estimates of the bias, variance and the square root of the mean squared error ($\sqrt{MSE}$) were obtained at each grid point for each simulation run.

3.4.2.1 Variable Censoring Interval Widths

The results from exponential censoring are displayed in Tables 3.5 through 3.13. These tables show the square root of MSE, bias and variance estimates for the 9 selected points of each target function for each imputation method. The columns headed by the abbreviation MP contain results for the midpoint method, LLS (normal) refers to the normal-based iterated conditional expectation imputation method, LLS (NP) refers to the nonparametric iterated conditional expectation imputation method, and LMLE refers to the local likelihood method.

As would be expected, in all cases, the MSE decreases as λ increases. In the case of the growth curve and the polynomial function, we do not see a substantial decrease as the rate increases from its 2nd highest setting to its highest setting; this is likely due to the censoring intervals tending to be very small relative to the range of the target function. The MSE behaviour is similar for all methods when the intervals tend to be small.

Most of the decrease in MSE is attributable to a decrease in variance. The bias is fairly constant (and small) for different values of λ. In fact, the small bias is an indication that the bandwidth selector may be choosing bandwidth values that are somewhat smaller than optimal.

The MSE also decreases as a function of n for all imputation methods. In particular, we see that, for the growth curve and polynomial function, the rate of decrease in MSE is roughly proportional to $n^{-.8}$, while for the logistic curve, the rate of decrease is roughly proportional to n^{-1}.

The local likelihood method appears to give smaller values of MSE than the other methods in most situations. In the case of the logistic model, the local likelihood method is substantially better than any of the other methods. The only occasion where we see inferior MSE behaviour is at the boundaries with the polynomial function. The midpoint imputation method is almost always inferior and sometimes substantially so. The other two methods usually give intermediate values of MSE, giving slightly better results at the boundaries in the case of the polynomial function.

3.4.2.2 Fixed Censoring Interval Widths

The results for fixed width intervals are displayed in Tables 3.14 to 3.16. Here only the growth curve target function is considered.

Again, we see that MSE decreases with interval width for all methods, and it is roughly proportional to n^{-1}, for this range of sample sizes.

In this case, no method clearly dominates the others. MSE for the midpoint method can vary with x a lot, depending on the true value of the regression function. The quality of the local least-squares methods also depends to some extent on the value of the regression function. The local likelihood method appears to depend less on the regression function, but it can result in larger MSE values than the other methods.

3.5 Discussion

The simulation results show that the methods behave similarly when the interval widths are small. When the interval widths are large and varied, the local likelihood method outperforms other methods. All methods outperform the midpoint imputation approach in this case.

For large fixed-width intervals, midpoint imputation can outperform the other methods when the true mean is close to the mid point of the interval. Otherwise, the local likelihood method often behaves best. We will now examine the results a little more closely.

3.5.1 Asymptotic Bias and Variance for Midpoint Imputation

When the censoring intervals arise from a homogeneous Poisson monitoring process, an asymptotic analysis of the midpoint imputation method is possible. The imputed responses can be viewed as having additional error, as follows. Let the ith interval midpoint be given by

$$M_i = \frac{L_i + R_i}{2}$$

for $i = 1, 2, \ldots, n$. Then

$$M_i = Y_i + \frac{C_{i2} - C_{i1}}{2}$$

so that

$$\mathrm{E}[M_i] = \mathrm{E}[Y_i] = g(x_i)$$

and

$$\mathrm{Var}(M_i) = \sigma^2 + \frac{1}{2\lambda^2}.$$

Well-known results for the Nadaraya-Watson estimator with bandwidth h can then be applied to see that the asymptotic bias of the local constant regression esti-

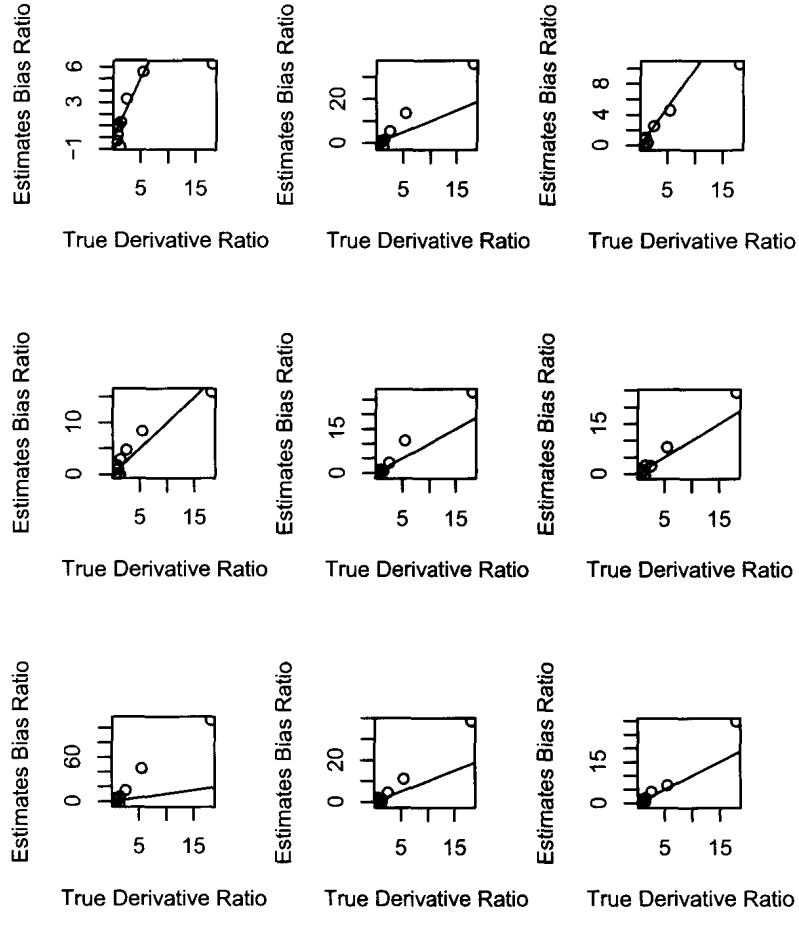

Figure 3.2: Bias ratios for the growth curve when the exponentially distributed censoring mechanism is operative. Sample size increases as one moves down the panels and mean censoring interval size decreases from left to right.

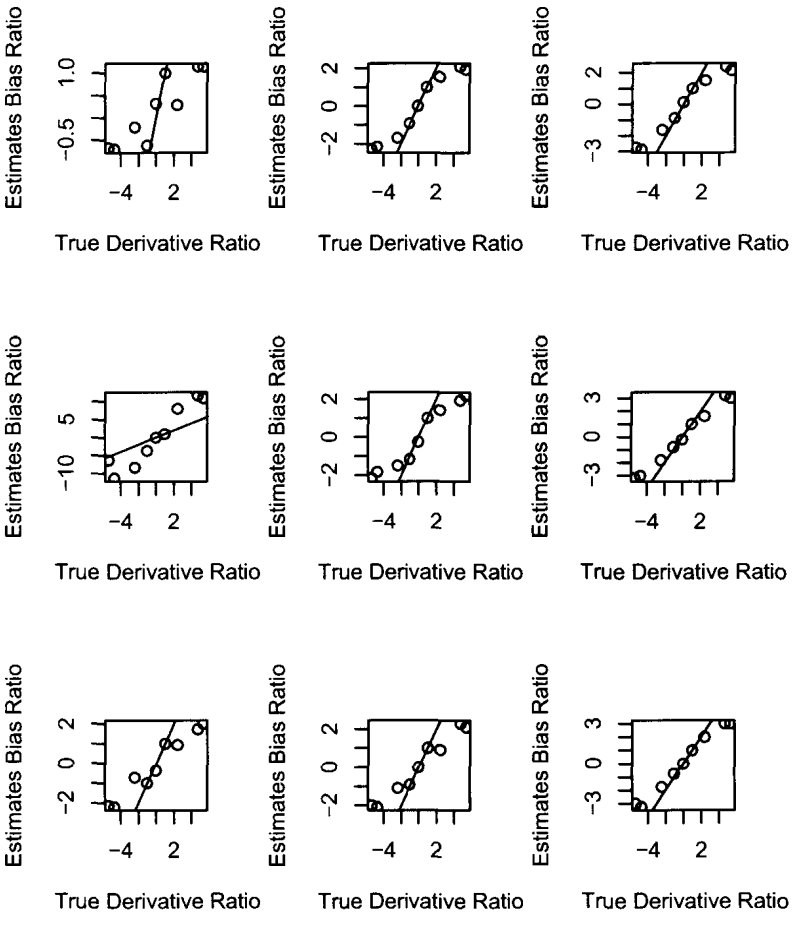

Figure 3.3: Bias ratios for the logistic curve when the exponentially distributed censoring mechanism is operative. Sample size increases as one moves down the panels and mean censoring interval size decreases from left to right.

Growth Curve at $x = 5$				
	Asymptotic	$\lambda = .2$	$\lambda = 1$	$\lambda = 5$
VR($n_1 = 50, n_2 = 100$)	0.574	0.573	0.583	0.562
VR($n_1 = 100, n_2 = 200$)	0.574	0.622	0.551	0.525
Logistic Curve at $x = 0$				
	Asymptotic	$\lambda = .8$	$\lambda = 2$	$\lambda = 8$
VR($n_1 = 50, n_2 = 100$)	0.574	0.545	0.583	0.667
VR($n_1 = 100, n_2 = 200$)	0.574	0.472	0.429	0.500
Polynomial Curve at $x = .5$				
	Asymptotic	$3/c(x)$	$5/c(x)$	$10/c(x)$
VR($n_1 = 50, n_2 = 100$)	0.574	0.533	0.500	0.556
VR($n_1 = 100, n_2 = 200$)	0.574	0.625	0.600	0.600

Table 3.3: Variance Ratios. Second column is the ratio of the theoretical asymptotic variances, and the third, fourth and fifth columns are the ratios of the variances from the simulations.

mator $\widehat{g}(x)$ applied to the interval midpoints is

$$B(\widehat{g}(x)) \propto \frac{h^2}{2} g''(x) \tag{3.10}$$

and the asymptotic variance is

$$\mathrm{Var}(\widehat{g}(x)) \propto \frac{\sigma^2 + 1/(2\lambda^2)}{nh}, \tag{3.11}$$

where the conditions that $h \to 0$, and $nh \to \infty$ as $n \to \infty$, and the kernel function K has compact support. The latter can be relaxed if more moments are assumed. See Wand and Jones (1995).

Thus, we have a theoretical argument for consistency of the midpoint imputation method when the response interval widths are exponentially distributed. In Figures 3.2, 3.3 and 3.4, we compare the theory with our simulations by plotting the ratio of the bias at x with the bias at 5 versus the ratio $g''(x)/g''(5)$. The ratio is used to remove the proportionality constant in (3.10). If the asymptotics are operative, the plotted points should appear near a straight line having slope 1. In fact, we see this in the cases where the mean interval size is very small (large λ) but not always when the mean interval size is larger. Thus, the asymptotic approximation to the bias sometimes requires a very large sample. We can also compare our empirical results with the asymptotic variance predicted by the above theory again computing ratios

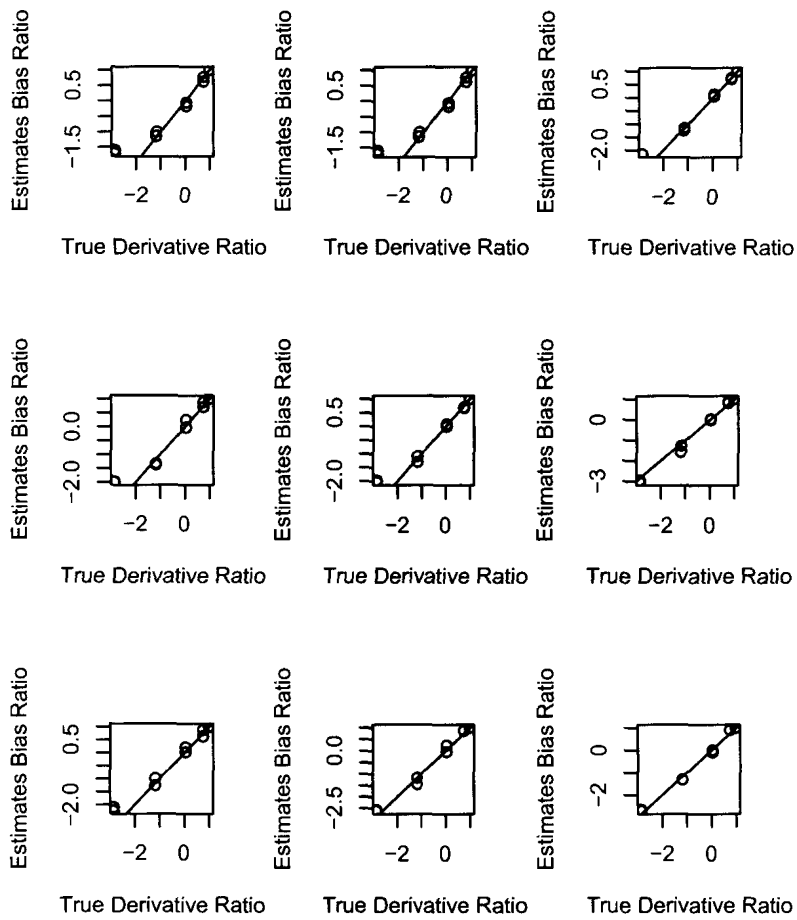

Figure 3.4: Bias ratios for the polynomial curve when the exponentially distributed censoring mechanism is operative. Sample size increases as one moves down the panels and mean censoring interval size decreases from left to right.

Growth Curve at $x = 5$

	Asymptotic	$n = 50$	$n = 100$	$n = 200$
$\text{VR}(\lambda_1 = 1, \lambda_2 = .2)$	0.068	0.269	0.274	0.243
$\text{VR}(\lambda_1 = 5, \lambda_2 = 1)$	0.736	0.871	0.840	0.799

Logistic Curve at $x = 0$

	Asymptotic	$n = 50$	$n = 100$	$n = 200$
$\text{VR}(\lambda_1 = 2, \lambda_2 = .8)$	0.059	0.182	0.194	0.176
$\text{VR}(\lambda_1 = 8, \lambda_2 = 2)$	0.039	0.250	0.286	0.333

Polynomial Curve at $x = .5$

	Asymptotic	$n = 50$	$n = 100$	$n = 200$
$\text{VR}(\lambda_1 = 5, \lambda_2 = 3)$	0.681	0.667	0.625	0.600
$\text{VR}(\lambda_1 = 10, \lambda_2 = 5)$	0.823	0.900	1.000	1.000

Table 3.4: Variance Ratios. Second column is the ratio of the theoretical asymptotic variances, and the third, fourth and fifth columns are the ratios of the variances from the simulations.

to remove the proportionality constants. In this case, we computed the ratios of the variances of the regression estimated at the midpoints of the sampling intervals only (e.g. $x = 5$ for the growth curve, etc.). The results are given in Tables 3.3 and 3.4.

The ratio of the asymptotic variances in Table 3.3 is calculated from

$$\text{VR}(n_1, n_2) = \frac{n_1^{4/5}}{n_2^{4/5}}.$$

This is displayed in the second column of the table. The third, fourth and fifth column of the table contain the ratios of the variances of the simulated estimates at sample sizes n_1 and n_2, while holding λ constant. With the exception of the logistic curve at the larger sample size and larger mean censoring interval size where we see a mild difference between the asymptotic approximation and the simulated value, the simulations match the asymptotics fairly well.

When the sample size is held constant, and λ is varied, the results are somewhat different as seen in Table 3.4. The ratio of the asymptotic variances in this table is calculated from

$$\text{VR}(\lambda_1, \lambda_2) = \frac{(\sigma^2 + 1/(2\lambda_1^2))^{8/5}}{(\sigma^2 + 1/(2\lambda_2^2))^{8/5}}.$$

The asymptotic ratio does not match the simulated ratio for the logistic curve at all

nor for the growth curve when the mean censoring interval size is large. Again, the larger intervals are having the effect of delaying the asymptotics, at least in these cases.

3.5.2 Estimating a Constant Mean from Fixed-Width Censoring Intervals

In order to more closely investigate the phenomenon whereby the MSE varies with the value of the regression function leading to the occasions where midpoint imputation can outperform the other methods, we considered the simpler situation of estimating the constant mean of a normally distributed population under censoring of both types considered above. We considered only the midpoint imputation method and the maximum likelihood method. The bias, variance and $\sqrt{MSE}$ are listed in Table 3.17 for the case where exponentially distributed intervals are censoring standard normal observations, for samples of size 50, 100 and 200. These results are consistent with the results observed above for estimation of nonlinear regression functions: the likelihood method is superior to the midpoint imputation method when the censoring intervals are of variable width.

For fixed width interval censoring applied to standard normal data, the results are given in Table 3.18. When the interval width is small, the three methods behave similarly. When the interval width is larger, the local likelihood method often performs better.

The results for different mean assumptions are listed in Table 3.19. When the interval width is 4, the local likelihood method has smaller variance than the midpoint based method at $\mu = 0$. However, at $\mu = 2$, the reverse is true. This is because of the way in which the data are coarsened: with a bin width of 4, the interval midpoint happens to coincide with the mean when $\mu = 2$. Thus, midpoint imputation will have excellent behaviour at that location, but only at that location.

When the interval width is 6, the local likelihood method has a smaller variance than the midpoint-based method at $\mu = 0$ and $\mu = 1$ and vice versa at $\mu = 2$ and $\mu = 3$. When the interval width is 8, the local likelihood method has a smaller variance than the midpoint-based method at $\mu = 0$ and $\mu = 7$, and vice versa at $\mu = 3$ and $\mu = 4$.

3.6 Applications

In this section, we illustrate the use of the imputation methods on several data sets, some of which have variable interval sizes and some which have fixed interval sizes.

3.6.1 Example: Aspen Flush Date

Figure 3.5 shows the regression function estimates that can be obtained when applying the four approaches to the aspen flush date data set which was described in the introduction to this thesis. The interval sizes are variable, so it is likely that the local likelihood method offers the most accurate curve. In order to see the imputed points more clearly, 3.6 shows a "close-up" view of part Figure 3.5. This shows that the conditional expectation imputation leads to different values from midpoint imputation.

3.6.2 Examples: Fire Rate versus Slope

Figures 3.7 and 3.8 are based on data collected from burning waxed paper experiments. In the first case, the burned area was measured as a function of time for a single fire which burned under very stable conditions (no wind, no slope, homogeneous fuel). This data set can be found at
`http://www.stats.uwo.ca/faculty/braun/datasets/firearea.R`

Because the flame often obscured the image, the true area could not be measured precisely, resulting in interval-censored responses. As can be seen in the figure, the response intervals are highly autocorrelated. This sets this example outside the scope of the simulation study conducted in the previous section.

Figure 3.8 is based on a set of 33 independently burned waxed paper fires at differing slopes. The rate of spread of the fire was obtained from a measure of the distance traveled by one edge of the fire in 2 seconds. The data can be found in Table 3.20. The interval widths are highly variable leading to some differences among the different imputed values. The local likelihood method is likely the most accurate rendering of the true regression curve.

3.6.3 Examples: Temperature

Two sets of temperature data were collected outside a London, Ontario house in late winter 2008 and in early summer 2007 using a DS1921G Thermochron iButton (see `http://www.embeddeddatasystems.com/page/EDS/CTGY/TCTC`). The first (plotted

in Figure 3.9) is two days of temperature data collected at hourly intervals and the second (plotted in Figure 3.10) is one day of temperature measurements collected every 3 minutes. The censoring intervals are all of width .5 degrees Celsius. These data sets can be found at
`http://www.stats.uwo.ca/faculty/braun/datasets/temperature1.R` and
`http://www.stats.uwo.ca/faculty/braun/datasets/temperature2.R`.

The three imputation approaches all yield similar results which is consistent with the observations made in the simulation study. The resulting local constant regression estimates are all very similar.

3.7 Conclusions

In this chapter, we have studied smoothing methods for interval-censored data and have shown that local likelihood estimation is often superior to local regression estimators where observations have been imputed using either interval midpoints or iterated conditional expectations when the censoring intervals are wide or of varying width. The fact that the midpoint imputation method is inferior to the other methods when the interval widths are variable but that it still enjoys consistency properties strengthens the argument in favor of the usefulness of the other methods, at least when the censoring intervals are variable.

When the intervals are smaller and of fixed width, none of the imputation approaches dominate the others.

We have only discussed local regression and local likelihood estimators with one covariate and with normal errors in this chapter. The methods can be extended to include multiple covariates. Replacing the normal cdf in the local likelihood method with a nonparametric estimate of the cdf as employed by Rabinowitz *et al.* (1995) is also worth investigating.

The example involving correlated fire area intervals is also motivation to study these methods more closely when the data exhibit serial dependence. Another form of dependence arises from the way in which data such as the aspen flush date data have been sampled. In order to handle multiple years of data, it will likely be necessary to use random effects models.

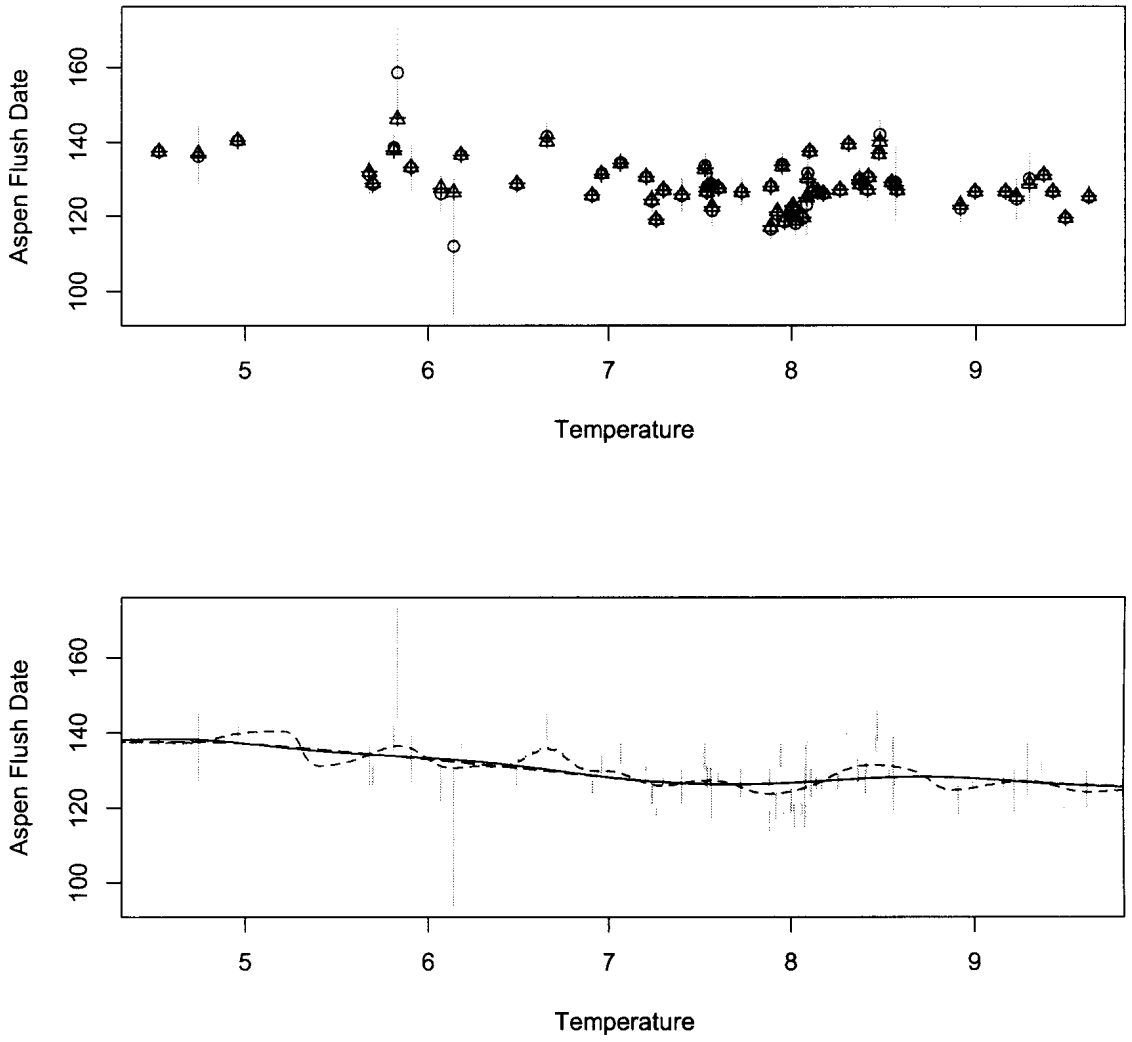

Figure 3.5: Top panel: flush date intervals, together with imputed values from three imputation methods studied in this chapter: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are from the iterated conditional expectation with a nonparametric estimate of the error distribution. Bottom panel: Local constant estimate of the regression curve for aspen flush date data using three imputation methods. Long dashed line is from midpoint imputation; dashed line is from iterated conditional expectation imputation under the normal error assumption; dotted line is from iterated conditional expectation with a nonparametric estimate of the error distribution; solid line is from local likelihood estimates.

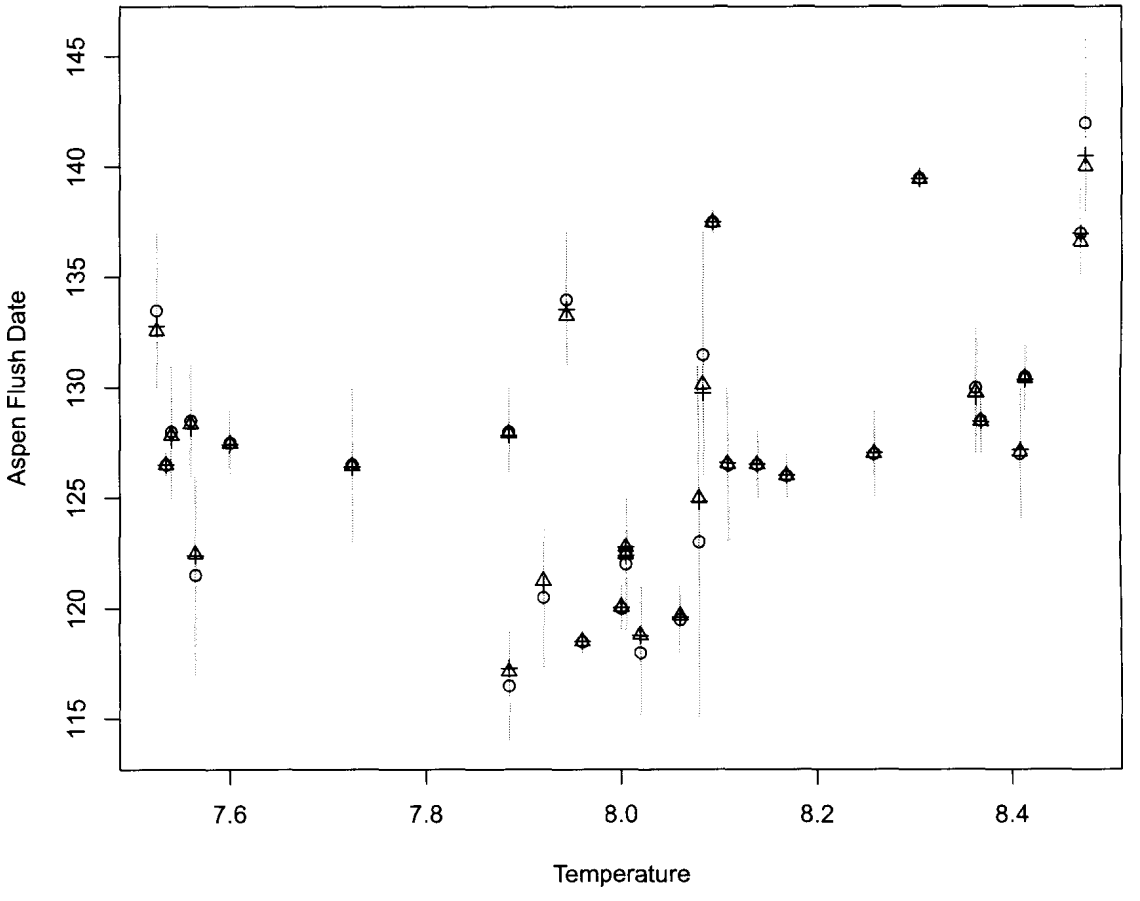

Figure 3.6: A "close-up" view of Figure 3.5 in the region of x between 7.5 and 8.5. Flush date intervals, together with imputed values from three imputation methods: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are from the iterated conditional expectation with a nonparametric estimate of the error distribution.

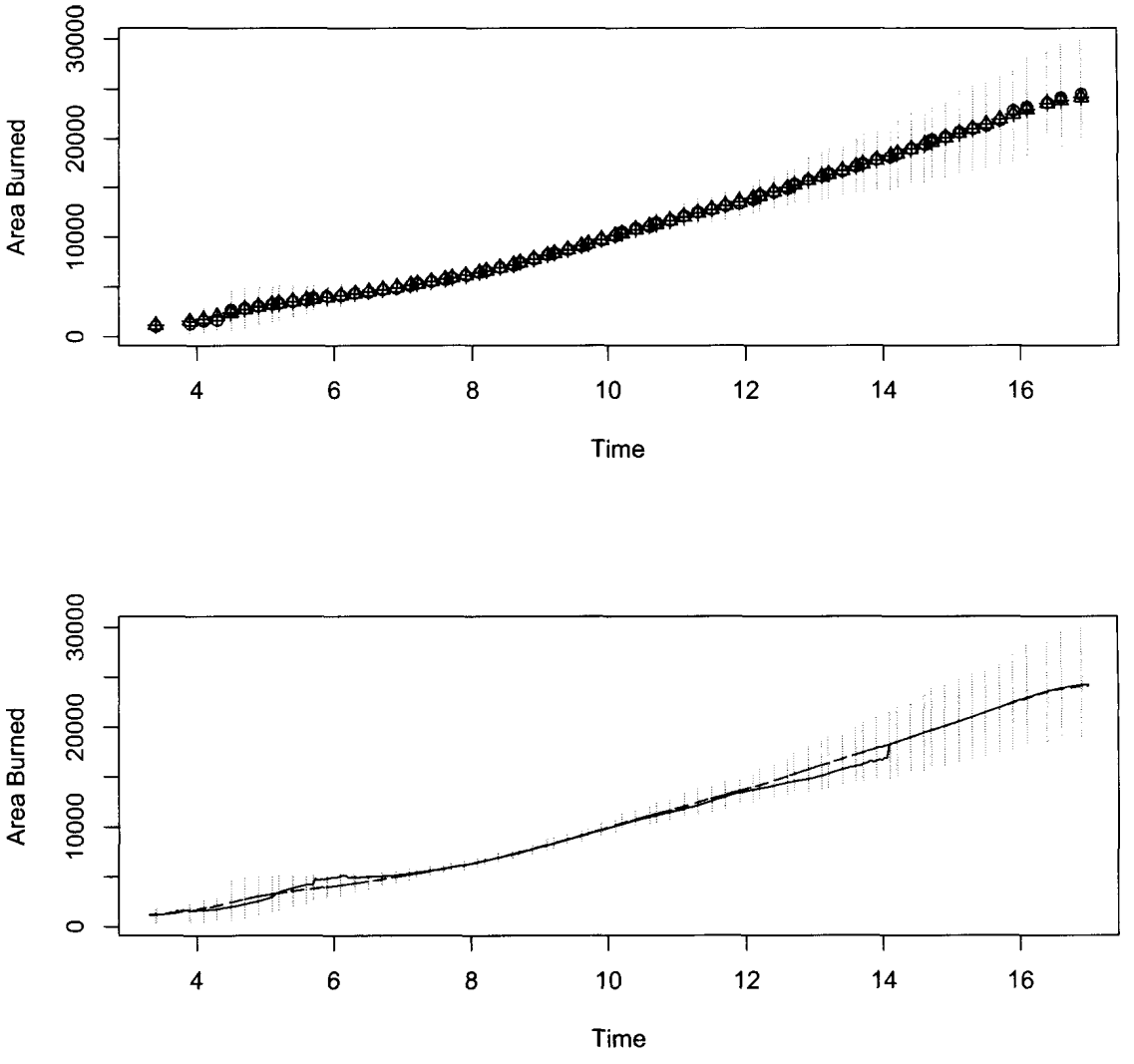

Figure 3.7: Top panel: Fire area intervals versus time, together with imputed values from three imputation methods studied in this chapter: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are imputed from the iterated conditional expectation with a nonparametric estimate of the error distribution. Bottom panel: Local constant estimate of the regression curve for aspen flush date data using three imputation methods. Long dashed line is from midpoint imputation; dashed line is from iterated conditional expectation imputation under the normal error assumption; dotted line is from iterated conditional expectation with a nonparametric estimate of the error distribution; solid line is from local likelihood estimates.

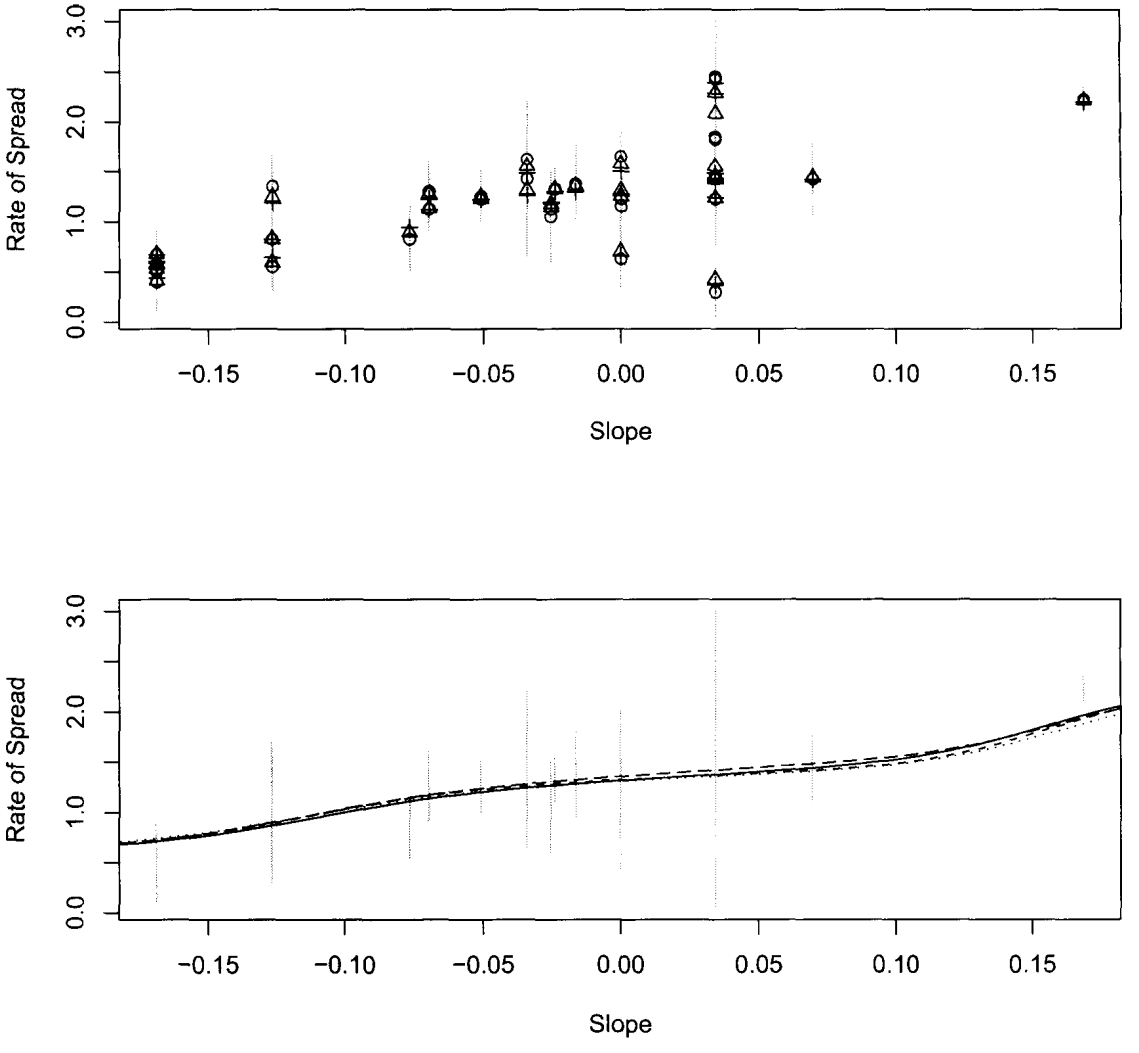

Figure 3.8: Top panel: rate of spread intervals, together with imputed values from three imputation methods studied in this chapter: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are imputed from the iterated conditional expectation with a nonparametric estimate of the error distribution. Bottom panel: Local constant estimate of the regression curve for fire rate data using three imputation methods. Long dashed line is from midpoint imputation; dashed line is from iterated conditional expectation imputation under the normal error assumption; dotted line is from iterated conditional expectation with a nonparametric estimate of the error distribution; solid line is from local likelihood estimates.

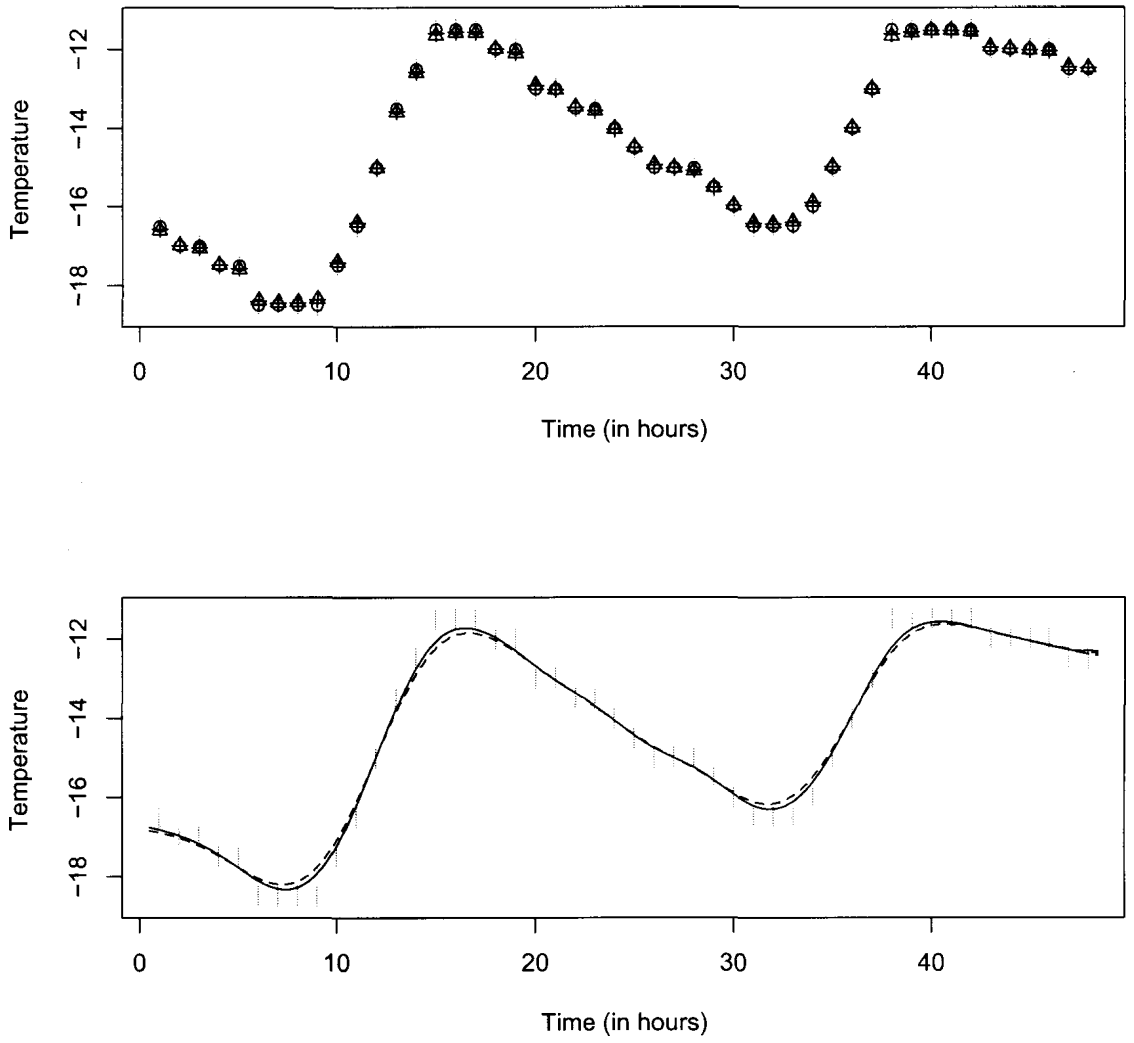

Figure 3.9: Top panel: Hourly temperature data collected in winter. Date intervals (in grey), together with imputed values from three imputation methods studied in this chapter: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are imputed from the iterated conditional expectation with a nonparametric estimate of the error distribution. Bottom panel: Local constant estimate of the regression curve for temperature data using three imputation methods. Long dashed line is from midpoint imputation; dashed line is from iterated conditional expectation imputation under the normal error assumption; dotted line is from iterated conditional expectation with a nonparametric estimate of the error distribution; solid line is from local likelihood estimates.

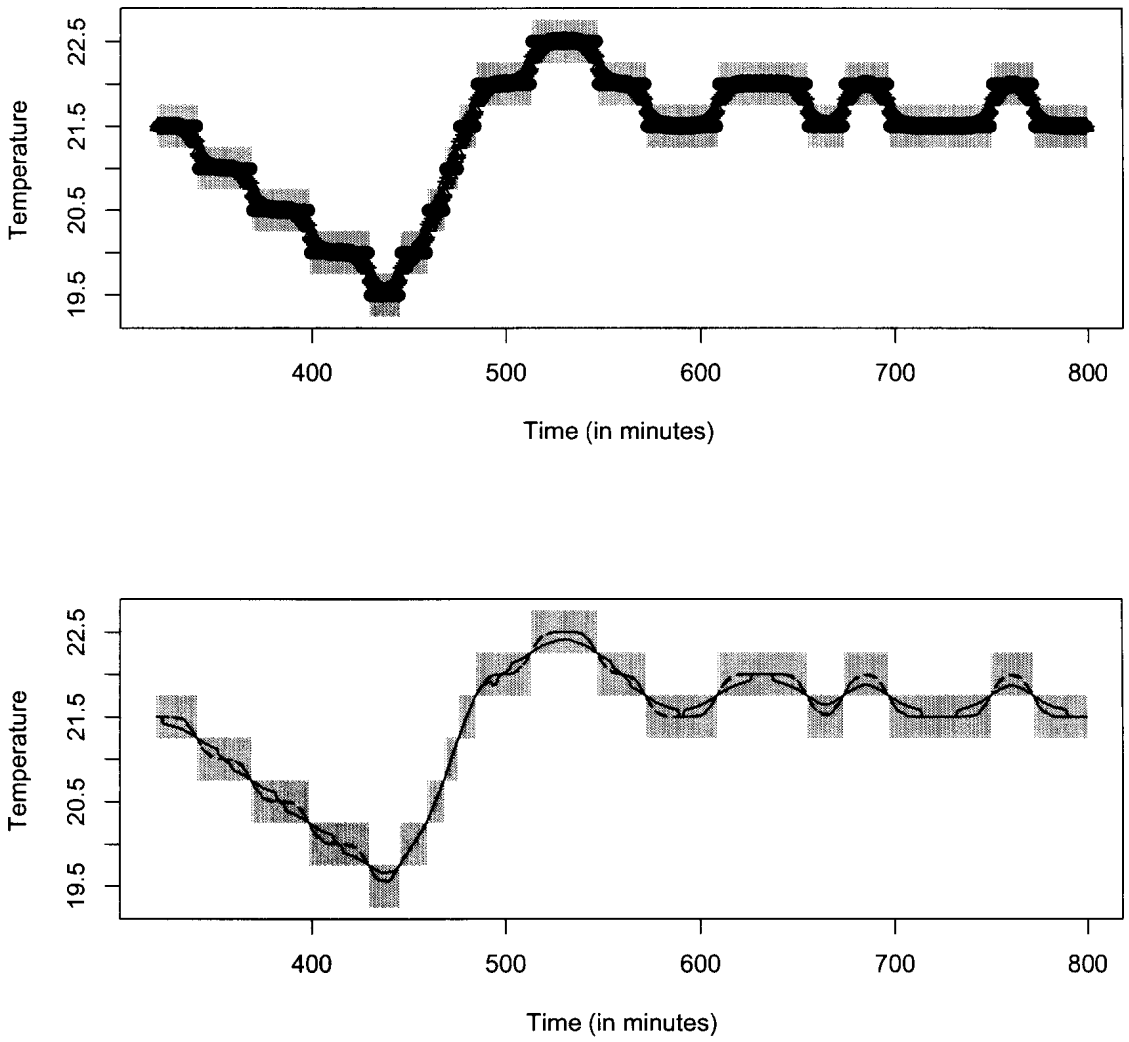

Figure 3.10: Top panel: Temperature data collected every 3 minutes. Intervals, together with imputed values from three imputation methods studied in this chapter: circles are from midpoint imputation; triangles are from iterated conditional expectation imputation under the normal error assumption; crosses are imputed from the iterated conditional expectation with a nonparametric estimate of the error distribution. Bottom panel: Local constant estimate of the regression curve for temperature data using three imputation methods. Long dashed line is from midpoint imputation; dashed line is from iterated conditional expectation imputation under the normal error assumption; dotted line is from iterated conditional expectation with a nonparametric estimate of the error distribution; solid line is from local likelihood estimates.

Table 3.5: $\sqrt{\text{MSE}}$ estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 50$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	1.287	1.060	1.039	0.856	0.638	0.627	0.627	0.626	0.613	0.613	0.613	0.613
2	1.411	1.210	0.976	0.948	0.777	0.645	0.628	0.629	0.629	0.603	0.603	0.603	0.603
3	1.857	1.235	0.988	0.951	0.782	0.655	0.640	0.643	0.642	0.604	0.604	0.604	0.604
4	2.394	1.275	1.010	0.983	0.785	0.627	0.612	0.615	0.618	0.598	0.597	0.598	0.598
5	3.086	1.247	1.002	0.971	0.789	0.646	0.626	0.629	0.626	0.605	0.604	0.604	0.604
6	4.041	1.312	1.077	1.043	0.823	0.626	0.611	0.612	0.620	0.575	0.575	0.575	0.575
7	5.473	1.311	1.067	1.039	0.882	0.696	0.688	0.690	0.689	0.626	0.626	0.626	0.626
8	7.897	1.388	1.253	1.208	1.148	0.789	0.789	0.787	0.782	0.766	0.768	0.768	0.767
9	12.959	2.045	2.050	2.005	1.972	1.456	1.495	1.485	1.444	1.392	1.395	1.394	1.386

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.290	0.230	0.201	0.109	0.109	0.088	0.082	0.061	0.044	0.041	0.041	0.039
-3	0.047	0.247	0.196	0.169	0.103	0.109	0.088	0.083	0.065	0.044	0.041	0.041	0.041
-2	0.119	0.242	0.197	0.170	0.112	0.115	0.095	0.087	0.072	0.049	0.048	0.048	0.047
-1	0.269	0.249	0.198	0.172	0.111	0.118	0.097	0.091	0.078	0.055	0.054	0.053	0.052
0	0.500	0.257	0.204	0.181	0.114	0.109	0.094	0.087	0.069	0.054	0.054	0.053	0.051
1	0.731	0.244	0.193	0.166	0.108	0.113	0.096	0.090	0.076	0.054	0.053	0.053	0.052
2	0.881	0.252	0.203	0.172	0.115	0.114	0.096	0.091	0.074	0.051	0.050	0.050	0.049
3	0.953	0.241	0.194	0.170	0.100	0.110	0.089	0.079	0.065	0.042	0.040	0.040	0.039
4	0.982	0.273	0.219	0.193	0.105	0.111	0.091	0.084	0.062	0.044	0.042	0.042	0.040

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	0.224	0.202	0.203	0.219	0.174	0.164	0.165	0.181	0.151	0.150	0.150	0.158
0.2	1.422	0.262	0.217	0.211	0.167	0.184	0.161	0.161	0.143	0.138	0.132	0.132	0.126
0.3	1.038	0.255	0.209	0.197	0.159	0.183	0.156	0.153	0.137	0.136	0.127	0.127	0.123
0.4	0.654	0.182	0.157	0.154	0.126	0.140	0.130	0.128	0.116	0.111	0.113	0.112	0.106
0.5	0.500	0.139	0.125	0.123	0.119	0.112	0.111	0.110	0.108	0.103	0.104	0.104	0.102
0.6	0.654	0.192	0.167	0.162	0.128	0.146	0.133	0.131	0.116	0.115	0.115	0.114	0.109
0.7	1.038	0.260	0.214	0.204	0.162	0.181	0.154	0.151	0.139	0.130	0.124	0.123	0.120
0.8	1.422	0.256	0.213	0.208	0.169	0.184	0.160	0.160	0.143	0.136	0.129	0.129	0.124
0.9	1.422	0.225	0.205	0.206	0.215	0.181	0.172	0.174	0.189	0.148	0.148	0.148	0.156

Table 3.6: Bias estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 50$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	-0.067	-0.048	-0.034	-0.033	0.027	0.029	0.027	0.028	0.013	0.013	0.012	0.013
2	1.411	0.021	0.015	0.018	0.020	0.001	0.005	0.002	0.007	-0.013	-0.013	-0.013	-0.013
3	1.857	-0.025	-0.019	-0.009	-0.009	0.007	0.013	0.009	0.014	0.043	0.043	0.043	0.043
4	2.394	0.079	0.066	0.066	0.046	0.010	0.009	0.008	0.009	0.001	0.001	0.001	0.000
5	3.086	0.086	0.078	0.085	0.091	0.016	0.017	0.014	0.018	0.053	0.053	0.053	0.053
6	4.041	0.112	0.120	0.122	0.127	0.022	0.027	0.025	0.024	0.015	0.016	0.015	0.015
7	5.473	0.282	0.257	0.261	0.260	0.087	0.097	0.094	0.090	0.134	0.134	0.133	0.134
8	7.897	0.478	0.475	0.457	0.521	0.218	0.239	0.232	0.222	0.241	0.243	0.242	0.240
9	12.959	0.535	0.591	0.596	0.669	0.574	0.584	0.582	0.577	0.556	0.556	0.556	0.544

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.034	0.029	0.029	0.020	0.013	0.012	0.012	0.012	0.008	0.008	0.008	0.009
-3	0.047	0.010	0.014	0.018	0.027	0.020	0.019	0.019	0.022	0.012	0.012	0.012	0.013
-2	0.119	0.039	0.039	0.038	0.043	0.027	0.027	0.026	0.034	0.019	0.019	0.018	0.020
-1	0.269	0.039	0.036	0.030	0.041	0.025	0.024	0.024	0.034	0.017	0.017	0.017	0.019
0	0.500	0.011	0.008	0.006	0.005	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001
1	0.731	-0.023	-0.023	-0.024	-0.036	-0.030	-0.029	-0.028	-0.034	-0.022	-0.020	-0.020	-0.022
2	0.881	-0.024	-0.026	-0.027	-0.040	-0.028	-0.027	-0.027	-0.035	-0.023	-0.022	-0.022	-0.024
3	0.953	-0.007	-0.013	-0.017	-0.026	-0.022	-0.021	-0.019	-0.026	-0.013	-0.013	-0.013	-0.013
4	0.982	-0.021	-0.019	-0.017	-0.017	-0.012	-0.012	-0.012	-0.015	-0.007	-0.008	-0.007	-0.008

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	-0.115	-0.114	-0.112	-0.121	-0.090	-0.087	-0.090	-0.117	-0.088	-0.088	-0.089	-0.101
0.2	1.422	-0.067	-0.075	-0.087	-0.082	-0.046	-0.053	-0.059	-0.051	-0.047	-0.050	-0.052	-0.044
0.3	1.038	-0.011	-0.010	-0.023	-0.039	-0.003	-0.004	-0.011	-0.023	0.005	0.004	0.002	-0.007
0.4	0.654	0.043	0.038	0.027	0.011	0.027	0.023	0.019	0.009	0.031	0.030	0.029	0.021
0.5	0.500	0.068	0.049	0.042	0.054	0.049	0.042	0.038	0.046	0.041	0.037	0.036	0.039
0.6	0.654	0.053	0.050	0.038	0.022	0.032	0.029	0.024	0.013	0.029	0.029	0.027	0.019
0.7	1.038	-0.003	-0.001	-0.016	-0.031	-0.005	-0.005	-0.012	-0.024	0.002	-0.000	-0.002	-0.011
0.8	1.422	-0.077	-0.081	-0.092	-0.087	-0.061	-0.064	-0.069	-0.069	-0.051	-0.055	-0.057	-0.047
0.9	1.422	-0.108	-0.109	-0.117	-0.154	-0.098	-0.096	-0.101	-0.127	-0.088	-0.090	-0.090	-0.103

Table 3.7: Variance estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 50$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	1.652	1.121	1.078	0.732	0.407	0.392	0.392	0.392	0.376	0.375	0.375	0.376
2	1.411	1.465	0.953	0.899	0.603	0.416	0.394	0.396	0.395	0.364	0.363	0.364	0.364
3	1.857	1.525	0.975	0.905	0.612	0.429	0.409	0.413	0.412	0.363	0.363	0.363	0.363
4	2.394	1.619	1.016	0.962	0.615	0.393	0.375	0.378	0.382	0.357	0.356	0.357	0.357
5	3.086	1.548	0.998	0.936	0.615	0.417	0.392	0.395	0.392	0.363	0.362	0.363	0.362
6	4.041	1.709	1.146	1.073	0.661	0.392	0.373	0.373	0.383	0.331	0.331	0.331	0.331
7	5.473	1.641	1.073	1.012	0.709	0.477	0.464	0.467	0.466	0.374	0.374	0.374	0.374
8	7.897	1.697	1.345	1.252	1.047	0.575	0.5649	0.565	0.562	0.529	0.530	0.531	0.530
9	12.959	3.896	3.853	3.665	3.439	1.790	1.895	1.866	1.753	1.630	1.638	1.635	1.624

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.083	0.052	0.039	0.012	0.012	0.008	0.007	0.004	0.002	0.002	0.002	0.001
-3	0.047	0.061	0.038	0.028	0.009	0.012	0.007	0.006	0.004	0.002	0.002	0.002	0.001
-2	0.119	0.057	0.037	0.028	0.011	0.013	0.008	0.007	0.004	0.002	0.002	0.002	0.002
-1	0.269	0.061	0.038	0.029	0.011	0.013	0.009	0.008	0.005	0.003	0.003	0.003	0.002
0	0.500	0.066	0.041	0.033	0.013	0.012	0.009	0.008	0.005	0.003	0.003	0.003	0.003
1	0.731	0.059	0.037	0.027	0.010	0.012	0.008	0.007	0.005	0.002	0.002	0.002	0.002
2	0.881	0.063	0.040	0.029	0.012	0.012	0.008	0.008	0.004	0.002	0.002	0.002	0.002
3	0.953	0.058	0.038	0.029	0.009	0.012	0.007	0.006	0.004	0.002	0.001	0.001	0.001
4	0.982	0.074	0.047	0.037	0.011	0.012	0.008	0.007	0.004	0.002	0.002	0.002	0.002

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	0.037	0.028	0.026	0.023	0.022	0.019	0.019	0.019	0.015	0.015	0.015	0.015
0.2	1.422	0.064	0.041	0.037	0.021	0.032	0.023	0.022	0.018	0.017	0.015	0.015	0.014
0.3	1.038	0.065	0.043	0.038	0.024	0.033	0.024	0.023	0.018	0.018	0.016	0.016	0.015
0.4	0.654	0.031	0.023	0.023	0.016	0.019	0.016	0.016	0.013	0.011	0.012	0.012	0.011
0.5	0.500	0.015	0.013	0.013	0.011	0.010	0.011	0.011	0.010	0.009	0.010	0.009	0.009
0.6	0.654	0.034	0.025	0.025	0.016	0.020	0.017	0.017	0.013	0.012	0.012	0.012	0.012
0.7	1.038	0.068	0.046	0.041	0.025	0.033	0.024	0.023	0.019	0.017	0.015	0.015	0.014
0.8	1.422	0.060	0.039	0.035	0.021	0.030	0.021	0.021	0.017	0.016	0.014	0.014	0.013
0.9	1.422	0.039	0.030	0.029	0.022	0.023	0.020	0.020	0.020	0.014	0.014	0.014	0.014

Table 3.8: $\sqrt{\text{MSE}}$ estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 100$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	1.006	0.793	0.772	0.636	0.487	0.477	0.477	0.475	0.469	0.469	0.469	0.469
2	1.411	0.899	0.727	0.702	0.581	0.479	0.478	0.468	0.470	0.435	0.435	0.435	0.435
3	1.857	0.910	0.720	0.710	0.582	0.463	0.452	0.453	0.453	0.459	0.459	0.459	0.459
4	2.394	0.932	0.743	0.716	0.591	0.492	0.487	0.487	0.484	0.458	0.458	0.458	0.458
5	3.086	0.943	0.757	0.735	0.618	0.493	0.487	0.487	0.485	0.452	0.452	0.452	0.452
6	4.041	0.952	0.767	0.753	0.629	0.494	0.489	0.490	0.486	0.453	0.453	0.454	0.454
7	5.473	0.971	0.797	0.782	0.679	0.505	0.497	0.500	0.497	0.447	0.447	0.447	0.447
8	7.897	1.137	0.967	0.950	0.876	0.598	0.596	0.594	0.585	0.549	0.549	0.549	0.549
9	12.959	1.4777	1.400	1.375	1.427	1.035	1.058	1.053	1.032	0.944	0.944	0.945	0.938

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.207	0.167	0.141	0.071	0.083	0.065	0.059	0.043	0.031	0.029	0.029	0.029
-3	0.047	0.206	0.170	0.143	0.078	0.080	0.064	0.083	0.059	0.034	0.031	0.031	0.030
-2	0.119	0.185	0.150	0.131	0.080	0.087	0.071	0.087	0.065	0.038	0.036	0.036	0.036
-1	0.269	0.181	0.147	0.125	0.083	0.087	0.072	0.067	0.059	0.042	0.040	0.040	0.039
0	0.500	0.190	0.154	0.132	0.075	0.082	0.068	0.064	0.050	0.040	0.039	0.039	0.037
1	0.731	0.193	0.158	0.134	0.079	0.086	0.072	0.066	0.059	0.041	0.040	0.040	0.040
2	0.881	0.184	0.146	0.123	0.080	0.089	0.072	0.091	0.066	0.037	0.035	0.035	0.035
3	0.953	0.184	0.148	0.124	0.074	0.080	0.064	0.058	0.045	0.031	0.030	0.030	0.029
4	0.982	0.203	0.168	0.144	0.075	0.085	0.068	0.063	0.046	0.032	0.031	0.031	0.029

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	0.176	0.150	0.151	0.166	0.131	0.122	0.123	0.136	0.111	0.109	0.110	0.117
0.2	1.422	0.210	0.168	0.162	0.122	0.143	0.122	0.123	0.107	0.101	0.095	0.095	0.091
0.3	1.038	0.198	0.157	0.145	0.112	0.139	0.116	0.114	0.104	0.099	0.093	0.093	0.091
0.4	0.654	0.135	0.115	0.110	0.092	0.105	0.097	0.096	0.085	0.086	0.085	0.084	0.081
0.5	0.500	0.100	0.093	0.091	0.086	0.082	0.084	0.083	0.079	0.075	0.078	0.077	0.075
0.6	0.654	0.136	0.116	0.111	0.091	0.104	0.096	0.095	0.084	0.082	0.081	0.081	0.077
0.7	1.038	0.199	0.161	0.152	0.118	0.139	0.116	0.114	0.102	0.094	0.088	0.088	0.087
0.8	1.422	0.206	0.170	0.165	0.130	0.143	0.122	0.122	0.106	0.099	0.093	0.094	0.089
0.9	1.422	0.168	0.147	0.149	0.166	0.130	0.121	0.123	0.138	0.112	0.110	0.111	0.118

Table 3.9: Bias estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 100$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	-0.007	0.006	0.006	0.007	0.017	0.020	0.017	0.020	-0.013	-0.013	-0.013	-0.013
2	1.411	0.007	0.005	0.011	0.009	-0.016	-0.015	-0.016	-0.016	0.006	0.006	0.006	0.006
3	1.857	0.072	0.055	0.055	0.042	-0.000	-0.003	-0.006	-0.005	0.007	0.007	0.006	0.007
4	2.394	0.040	0.025	0.031	0.023	0.024	0.020	0.019	0.021	0.017	0.017	0.017	0.017
5	3.086	0.042	0.034	0.033	0.031	0.020	0.021	0.018	0.020	0.021	0.022	0.022	0.021
6	4.041	0.121	0.117	0.124	0.111	0.015	0.020	0.017	0.017	0.052	0.051	0.051	0.051
7	5.473	0.198	0.180	0.181	0.188	0.071	0.075	0.072	0.075	0.049	0.049	0.049	0.049
8	7.897	0.351	0.372	0.380	0.448	0.220	0.223	0.230	0.215	0.167	0.167	0.167	0.166
9	12.959	0.666	0.658	0.673	0.832	0.546	0.555	0.552	0.544	0.509	0.509	0.509	0.499

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.003	0.006	0.007	0.014	0.012	0.011	0.010	0.012	0.005	0.005	0.005	0.005
-3	0.047	0.024	0.022	0.019	0.023	0.017	0.018	0.017	0.021	0.008	0.008	0.007	0.008
-2	0.119	0.035	0.031	0.031	0.036	0.023	0.022	0.021	0.029	0.016	0.015	0.015	0.017
-1	0.269	0.033	0.031	0.030	0.038	0.026	0.023	0.022	0.030	0.015	0.014	0.014	0.016
0	0.500	0.000	0.000	0.000	0.001	-0.003	-0.003	-0.002	0.000	-0.001	-0.002	-0.002	-0.001
1	0.731	-0.019	-0.018	-0.017	-0.031	-0.026	-0.024	-0.023	-0.032	-0.016	-0.016	-0.016	-0.018
2	0.881	-0.034	-0.031	-0.032	-0.039	-0.022	-0.021	-0.020	-0.029	-0.015	-0.014	-0.014	-0.016
3	0.953	-0.025	-0.025	-0.023	-0.026	-0.018	-0.016	-0.015	-0.017	-0.009	-0.009	-0.009	-0.009
4	0.982	-0.011	-0.012	-0.014	-0.017	-0.014	-0.012	-0.011	-0.011	-0.004	-0.004	-0.004	-0.004

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	-0.086	-0.078	-0.086	-0.121	-0.075	-0.071	-0.074	-0.098	-0.070	-0.068	-0.069	-0.080
0.2	1.422	-0.058	-0.056	-0.066	-0.052	-0.040	-0.042	-0.048	-0.033	-0.029	-0.034	-0.035	-0.029
0.3	1.038	0.011	0.008	-0.004	-0.052	-0.000	-0.002	-0.008	0.019	0.001	0.000	-0.002	-0.009
0.4	0.654	0.031	0.027	0.017	0.007	0.025	0.023	0.019	0.010	0.019	0.019	0.018	0.012
0.5	0.500	0.044	0.031	0.025	0.037	0.037	0.027	0.025	0.032	0.023	0.019	0.019	0.022
0.6	0.654	0.039	0.033	0.024	0.011	0.026	0.023	0.019	0.009	0.020	0.019	0.017	0.012
0.7	1.038	-0.001	-0.003	-0.014	-0.027	0.003	-0.000	-0.005	-0.013	0.000	0.000	-0.002	-0.008
0.8	1.422	-0.060	-0.061	-0.071	-0.056	-0.048	-0.050	-0.055	-0.042	-0.036	-0.038	-0.040	-0.032
0.9	1.422	-0.087	-0.079	-0.086	-0.123	-0.073	-0.069	-0.073	-0.099	-0.068	-0.068	-0.067	-0.078

Table 3.10: Variance estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 100$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	1.012	0.629	0.595	0.405	0.237	0.227	0.228	0.225	0.220	0.220	0.220	0.220
2	1.411	0.808	0.529	0.492	0.338	0.229	0.219	0.220	0.218	0.189	0.189	0.189	0.189
3	1.857	0.822	0.515	0.501	0.337	0.215	0.205	0.205	0.205	0.211	0.211	0.211	0.211
4	2.394	0.867	0.551	0.511	0.242	0.236	0.237	0.234	0.382	0.210	0.210	0.210	0.210
5	3.086	0.887	0.571	0.539	0.381	0.243	0.236	0.237	0.235	0.204	0.203	0.203	0.204
6	4.041	0.892	0.574	0.551	0.384	0.244	0.238	0.239	0.236	0.203	0.203	0.203	0.203
7	5.473	0.903	0.603	0.579	0.425	0.250	0.241	0.245	0.241	0.198	0.197	0.198	0.198
8	7.897	1.170	0.797	0.758	0.566	0.309	0.305	0.304	0.296	0.274	0.274	0.274	0.274
9	12.959	1.739	1.526	1.439	1.345	0.772	0.812	0.804	0.770	0.631	0.634	0.634	0.632

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.043	0.028	0.020	0.005	0.007	0.004	0.003	0.002	0.001	0.001	0.001	0.001
-3	0.047	0.042	0.028	0.020	0.006	0.006	0.004	0.003	0.002	0.001	0.001	0.001	0.001
-2	0.119	0.033	0.022	0.016	0.005	0.007	0.005	0.004	0.002	0.001	0.001	0.001	0.001
-1	0.269	0.032	0.021	0.015	0.005	0.007	0.005	0.004	0.003	0.001	0.001	0.001	0.001
0	0.500	0.036	0.024	0.017	0.006	0.007	0.005	0.004	0.002	0.002	0.002	0.002	0.001
1	0.731	0.037	0.025	0.018	0.005	0.007	0.005	0.004	0.002	0.001	0.001	0.001	0.001
2	0.881	0.033	0.020	0.014	0.005	0.007	0.005	0.004	0.002	0.001	0.001	0.001	0.001
3	0.953	0.033	0.021	0.015	0.005	0.006	0.004	0.003	0.002	0.001	0.001	0.001	0.001
4	0.982	0.041	0.028	0.021	0.005	0.007	0.005	0.004	0.002	0.001	0.001	0.001	0.001

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	0.024	0.016	0.015	0.013	0.011	0.010	0.010	0.009	0.007	0.007	0.007	0.007
0.2	1.422	0.041	0.025	0.022	0.012	0.019	0.013	0.013	0.010	0.009	0.008	0.008	0.007
0.3	1.038	0.039	0.024	0.021	0.012	0.019	0.014	0.013	0.011	0.010	0.009	0.009	0.008
0.4	0.654	0.017	0.012	0.012	0.008	0.010	0.009	0.009	0.007	0.007	0.007	0.007	0.006
0.5	0.500	0.008	0.008	0.008	0.006	0.005	0.006	0.006	0.005	0.005	0.006	0.006	0.005
0.6	0.654	0.017	0.012	0.012	0.008	0.010	0.009	0.009	0.007	0.006	0.006	0.006	0.006
0.7	1.038	0.039	0.026	0.023	0.013	0.019	0.013	0.013	0.010	0.009	0.008	0.008	0.008
0.8	1.422	0.039	0.025	0.022	0.014	0.018	0.012	0.012	0.010	0.009	0.007	0.007	0.007
0.9	1.422	0.021	0.015	0.015	0.012	0.012	0.010	0.011	0.009	0.008	0.008	0.008	0.008

Table 3.11: $\sqrt{\text{MSE}}$ estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 200$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.727	0.570	0.551	0.446	0.370	0.363	0.364	0.361	0.330	0.330	0.330	0.330
2	1.411	0.722	0.563	0.546	0.444	0.362	0.353	0.354	0.354	0.346	0.347	0.346	0.346
3	1.857	0.721	0.556	0.538	0.430	0.381	0.372	0.373	0.373	0.330	0.330	0.330	0.330
4	2.394	0.726	0.571	0.558	0.445	0.374	0.370	0.371	0.370	0.341	0.341	0.341	0.341
5	3.086	0.743	0.590	0.565	0.468	0.366	0.363	0.363	0.363	0.327	0.327	0.327	0.327
6	4.041	0.698	0.549	0.536	0.441	0.362	0.354	0.354	0.352	0.349	0.349	0.349	0.349
7	5.473	0.715	0.577	0.560	0.473	0.386	0.381	0.381	0.379	0.342	0.343	0.343	0.343
8	7.897	0.802	0.667	0.648	0.591	0.403	0.398	0.398	0.393	0.383	0.383	0.383	0.383
9	12.959	1.163	1.067	1.050	1.149	0.725	0.747	0.744	0.725	0.703	0.705	0.704	0.699

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	SM1	SM2	SM3	SM4	SM1	SM2	SM3	SM4	SM1	2SM	SM3	SM4
-4	0.018	0.140	0.114	0.093	0.047	0.062	0.048	0.044	0.033	0.023	0.022	0.022	0.021
-3	0.047	0.133	0.110	0.093	0.052	0.058	0.046	0.043	0.032	0.025	0.023	0.023	0.022
-2	0.119	0.134	0.110	0.092	0.063	0.066	0.054	0.049	0.043	0.028	0.026	0.026	0.026
-1	0.269	0.129	0.105	0.088	0.062	0.066	0.054	0.050	0.043	0.031	0.029	0.029	0.029
0	0.500	0.129	0.106	0.088	0.051	0.059	0.050	0.046	0.036	0.027	0.026	0.027	0.025
1	0.731	0.128	0.106	0.090	0.062	0.062	0.052	0.048	0.045	0.030	0.029	0.028	0.028
2	0.881	0.133	0.110	0.092	0.060	0.063	0.051	0.047	0.042	0.028	0.026	0.026	0.026
3	0.953	0.130	0.106	0.089	0.051	0.058	0.046	0.042	0.033	0.025	0.022	0.022	0.022
4	0.982	0.131	0.106	0.088	0.049	0.059	0.047	0.043	0.033	0.025	0.024	0.023	0.023

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	SM1	SM2	SM3	SM4	SM1	SM2	SM3	SM4	SM1	2SM	SM3	SM4
0.1	1.422	0.129	0.112	0.116	0.137	0.103	0.095	0.097	0.112	0.083	0.081	0.082	0.088
0.2	1.422	0.148	0.117	0.115	0.087	0.105	0.085	0.086	0.073	0.072	0.068	0.069	0.066
0.3	1.038	0.151	0.120	0.113	0.088	0.105	0.084	0.083	0.076	0.071	0.067	0.067	0.066
0.4	0.654	0.106	0.088	0.084	0.067	0.078	0.071	0.070	0.063	0.064	0.063	0.063	0.061
0.5	0.500	0.077	0.071	0.070	0.067	0.061	0.062	0.061	0.060	0.058	0.060	0.060	0.058
0.6	0.654	0.110	0.084	0.081	0.066	0.077	0.071	0.070	0.063	0.063	0.063	0.063	0.060
0.7	1.038	0.147	0.115	0.109	0.086	0.102	0.083	0.081	0.074	0.071	0.065	0.065	0.065
0.8	1.422	0.156	0.125	0.123	0.090	0.107	0.090	0.091	0.079	0.073	0.068	0.069	0.064
0.9	1.422	0.131	0.110	0.114	0.134	0.104	0.095	0.097	0.111	0.084	0.082	0.083	0.089

Table 3.12: Bias estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 200$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	-0.021	-0.023	-0.007	-0.019	-0.001	-0.001	-0.003	-0.001	0.013	0.013	0.013	0.013
2	1.411	0.008	0.008	0.019	0.008	-0.015	-0.017	-0.018	-0.017	-0.009	-0.009	-0.009	-0.009
3	1.857	0.012	0.005	0.016	0.011	0.023	0.025	0.025	0.025	0.017	0.017	0.017	0.017
4	2.394	0.024	0.028	0.040	0.028	0.000	-0.002	-0.003	-0.001	0.020	0.020	0.020	0.020
5	3.086	0.006	0.010	0.018	0.015	0.012	0.012	0.011	0.012	0.015	0.015	0.015	0.015
6	4.041	0.038	0.052	0.063	0.058	0.015	0.017	0.015	0.016	0.028	0.028	0.028	0.028
7	5.473	0.089	0.085	0.092	0.092	0.054	0.058	0.057	0.057	0.062	0.062	0.064	0.065
8	7.897	0.269	0.261	0.263	0.291	0.132	0.132	0.131	0.127	0.098	0.099	0.099	0.098
9	12.959	0.662	0.624	0.623	0.836	0.461	0.474	0.471	0.466	0.445	0.446	0.445	0.437

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.014	0.012	0.008	0.011	0.011	0.009	0.009	0.010	0.004	0.004	0.004	0.004
-3	0.047	0.013	0.013	0.013	0.021	0.010	0.010	0.010	0.013	0.008	0.007	0.007	0.008
-2	0.119	0.024	0.023	0.025	0.034	0.025	0.022	0.021	0.025	0.012	0.012	0.012	0.013
-1	0.269	0.028	0.026	0.025	0.035	0.023	0.021	0.019	0.027	0.012	0.012	0.012	0.013
0	0.500	-0.005	-0.005	-0.005	-0.002	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.001
1	0.731	-0.030	-0.028	-0.027	-0.034	-0.022	-0.021	-0.019	-0.027	-0.012	-0.011	-0.011	-0.012
2	0.881	-0.031	-0.028	-0.024	-0.033	-0.023	-0.021	-0.021	-0.026	-0.013	-0.013	-0.013	-0.013
3	0.953	-0.010	-0.012	-0.013	-0.021	-0.012	-0.012	-0.011	-0.015	-0.007	-0.007	-0.007	-0.007
4	0.982	-0.014	-0.013	-0.011	-0.014	-0.010	-0.009	-0.009	-0.011	-0.003	-0.003	-0.002	-0.003

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	-0.080	-0.071	-0.078	-0.113	-0.064	-0.060	-0.064	-0.087	-0.053	-0.050	0.051	-0.061
0.2	1.422	-0.035	-0.037	-0.048	-0.036	-0.029	-0.031	-0.036	-0.024	-0.026	-0.028	-0.029	-0.022
0.3	1.038	0.001	-0.001	-0.012	-0.021	-0.001	-0.0001	-0.006	-0.013	0.000	0.000	-0.002	-0.007
0.4	0.654	0.023	0.021	0.012	0.004	0.022	0.019	0.015	0.007	0.018	0.017	0.016	0.012
0.5	0.500	0.036	0.025	0.020	0.031	0.025	0.018	0.017	0.022	0.020	0.018	0.017	0.019
0.6	0.654	0.032	0.027	0.018	0.008	0.022	0.020	0.016	0.009	0.018	0.017	0.016	0.012
0.7	1.038	0.008	0.006	-0.005	-0.016	0.006	0.003	-0.003	-0.012	-0.002	-0.003	-0.005	-0.010
0.8	1.422	-0.045	-0.045	-0.056	-0.040	-0.036	-0.036	-0.041	-0.028	-0.026	-0.027	-0.029	-0.021
0.9	1.422	-0.075	-0.065	-0.072	-0.109	-0.065	-0.058	-0.061	-0.085	-0.054	-0.052	-0.053	-0.063

Table 3.13: Variance estimates when estimating the three target curves where the responses are censored with exponential intervals ($n = 200$)

Target Function 1: Growth Curve (He and Shi (1998))

		$\lambda = 0.2$				$\lambda = 1$				$\lambda = 5$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.529	0.324	0.304	0.199	0.137	0.132	0.133	0.131	0.108	0.109	0.109	0.109
2	1.411	0.521	0.317	0.298	0.197	0.131	0.124	0.125	0.125	0.120	0.120	0.120	0.120
3	1.857	0.520	0.309	0.290	0.185	0.215	0.205	0.205	0.205	0.108	0.108	0.108	0.108
4	2.394	0.527	0.326	0.309	0.198	0.140	0.137	0.137	0.136	0.116	0.116	0.116	0.116
5	3.086	0.552	0.348	0.318	0.218	0.134	0.132	0.131	0.132	0.107	0.107	0.107	0.107
6	4.041	0.486	0.300	0.283	0.191	0.131	0.125	0.126	0.124	0.121	0.121	0.121	0.121
7	5.473	0.503	0.326	0.305	0.215	0.146	0.142	0.142	0.141	0.113	0.114	0.114	0.114
8	7.897	0.570	0.377	0.351	0.265	0.145	0.141	0.142	0.139	0.137	0.137	0.137	0.137
9	12.959	0.913	0.750	0.715	0.622	0.314	0.333	0.332	0.309	0.297	0.299	0.298	0.298

Target Function 2: Logistic Curve (He and Shi (1998))

		$\lambda = 0.8$				$\lambda = 2$				$\lambda = 8$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
-4	0.018	0.019	0.013	0.009	0.002	0.004	0.002	0.002	0.001	0.001	0.000	0.000	0.000
-3	0.047	0.017	0.012	0.008	0.002	0.003	0.002	0.002	0.001	0.001	0.000	0.000	0.000
-2	0.119	0.017	0.012	0.008	0.003	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.001
-1	0.269	0.016	0.010	0.007	0.003	0.004	0.003	0.002	0.001	0.001	0.001	0.001	0.001
0	0.500	0.017	0.011	0.008	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001
1	0.731	0.016	0.010	0.007	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001
2	0.881	0.017	0.011	0.008	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001
3	0.953	0.017	0.011	0.008	0.002	0.003	0.002	0.002	0.001	0.001	0.000	0.000	0.000
4	0.982	0.017	0.011	0.008	0.002	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001

Target Function 3: Function (Fan and Gijbels (1994))

		$\lambda = 3/c(x)$				$\lambda = 5/c(x)$				$\lambda = 10/c(x)$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
0.1	1.422	0.010	0.008	0.007	0.006	0.007	0.005	0.005	0.005	0.004	0.004	0.004	0.004
0.2	1.422	0.021	0.012	0.011	0.006	0.010	0.006	0.006	0.005	0.004	0.004	0.004	0.004
0.3	1.038	0.023	0.014	0.013	0.007	0.011	0.007	0.007	0.006	0.005	0.004	0.004	0.004
0.4	0.654	0.011	0.007	0.007	0.005	0.006	0.005	0.005	0.004	0.004	0.004	0.004	0.004
0.5	0.500	0.005	0.004	0.005	0.004	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003
0.6	0.654	0.009	0.006	0.006	0.004	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.003
0.7	1.038	0.021	0.013	0.012	0.007	0.010	0.007	0.007	0.005	0.005	0.004	0.004	0.004
0.8	1.422	0.022	0.014	0.012	0.006	0.010	0.007	0.007	0.005	0.005	0.004	0.004	0.004
0.9	1.422	0.012	0.008	0.008	0.006	0.007	0.006	0.006	0.005	0.004	0.004	0.004	0.004

Table 3.14: Bias, Variance and MSE Estimates when estimating the Growth Curve where the responses are censored with fixed width intervals ($n = 50$)

Bias

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.017	0.017	0.016	0.017	-0.002	-0.001	-0.004	-0.002	0.114	0.087	0.086	0.058
2	1.411	-0.011	-0.011	-0.011	-0.011	0.021	0.021	0.018	0.021	0.066	0.048	0.043	0.023
3	1.857	-0.014	-0.014	-0.014	-0.014	0.008	0.008	0.005	0.008	0.032	0.029	0.017	0.023
4	2.394	0.024	0.024	0.024	0.024	0.040	0.042	0.038	0.040	-0.039	-0.019	-0.038	-0.006
5	3.086	0.043	0.043	0.043	0.043	0.039	0.036	0.036	0.039	-0.023	0.008	-0.010	0.030
6	4.041	0.041	0.041	0.041	0.041	0.094	0.097	0.091	0.094	0.075	0.075	0.072	0.066
7	5.473	0.070	0.071	0.070	0.070	0.117	0.123	0.119	0.118	0.170	0.183	0.174	0.153
8	7.897	0.215	0.217	0.216	0.214	0.246	0.261	0.253	0.248	0.264	0.314	0.302	0.292
9	12.959	0.612	0.612	0.612	0.605	0.631	0.642	0.638	0.646	0.577	0.615	0.611	0.640

Variance

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.359	0.359	0.359	0.360	0.414	0.408	0.411	0.413	0.481	0.467	0.461	0.472
2	1.411	0.376	0.376	0.376	0.376	0.409	0.403	0.405	0.408	0.436	0.438	0.432	0.478
3	1.857	0.356	0.356	0.357	0.357	0.386	0.380	0.383	0.386	0.433	0.456	0.446	0.507
4	2.394	0.358	0.359	0.358	0.359	0.392	0.386	0.389	0.392	0.412	0.423	0.425	0.461
5	3.086	0.372	0.371	0.371	0.372	0.389	0.383	0.386	0.388	0.515	0.491	0.512	0.494
6	4.041	0.373	0.373	0.373	0.373	0.369	0.364	0.368	0.368	0.650	0.599	0.621	0.562
7	5.473	0.402	0.402	0.403	0.402	0.430	0.430	0.432	0.429	0.527	0.546	0.537	0.577
8	7.897	0.535	0.536	0.536	0.535	0.565	0.575	0.583	0.562	0.772	0.795	0.806	0.693
9	12.959	1.777	1.788	1.783	1.775	1.770	1.894	1.866	1.748	1.922	2.314	2.268	1.858

$\sqrt{MSE}$

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.600	0.559	0.600	0.600	0.643	0.639	0.641	0.643	0.703	0.689	0.685	0.690
2	1.411	0.613	0.613	0.613	0.613	0.645	0.640	0.635	0.637	0.664	0.663	0.659	0.692
3	1.857	0.597	0.597	0.597	0.597	0.622	0.617	0.619	0.621	0.659	0.676	0.668	0.713
4	2.394	0.599	0.599	0.599	0.599	0.627	0.622	0.625	0.627	0.643	0.651	0.653	0.679
5	3.086	0.611	0.611	0.611	0.611	0.625	0.620	0.622	0.624	0.718	0.701	0.716	0.703
6	4.041	0.612	0.612	0.612	0.612	0.615	0.611	0.613	0.614	0.810	0.777	0.791	0.753
7	5.473	0.638	0.638	0.638	0.638	0.666	0.667	0.668	0.666	0.746	0.762	0.753	0.775
8	7.897	0.762	0.763	0.763	0.762	0.791	0.802	0.804	0.790	0.917	0.945	0.947	0.882
9	12.959	1.467	1.471	1.469	1.463	1.473	1.519	1.507	1.471	1.501	1.641	1.626	1.506

Table 3.15: Bias, Variance and MSE Estimates when estimating the Growth Curve where the responses are censored with fixed width intervals ($n = 100$)

Bias

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	-0.015	0.015	0.015	0.015	-0.025	-0.025	-0.029	-0.025	0.068	0.039	0.044	0.017
2	1.411	-0.012	-0.012	-0.012	-0.012	-0.022	-0.022	-0.025	-0.023	0.024	-0.001	0.002	-0.023
3	1.857	0.014	0.014	0.014	0.014	0.016	0.016	0.011	0.016	0.034	0.031	0.024	0.028
4	2.394	0.030	0.030	0.030	0.030	0.016	0.016	0.012	0.016	-0.019	0.001	-0.016	0.016
5	3.086	0.037	0.037	0.037	0.037	0.046	0.048	0.043	0.046	-0.046	-0.015	-0.031	-0.006
6	4.041	0.011	0.011	0.011	0.011	0.055	0.057	0.053	0.055	0.018	0.019	0.016	0.014
7	5.473	0.065	0.066	0.065	0.065	0.088	0.093	0.088	0.089	0.114	0.111	0.108	0.086
8	7.897	0.168	0.168	0.168	0.168	0.243	0.259	0.250	0.245	0.176	0.203	0.196	0.190
9	12.959	0.520	0.520	0.520	0.511	0.643	0.656	0.653	0.657	0.613	0.644	0.640	0.658

Variance

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.195	0.195	0.195	0.195	0.409	0.403	0.407	0.409	0.327	0.319	0.315	0.320
2	1.411	0.198	0.197	0.197	0.198	0.374	0.367	0.371	0.373	0.264	0.274	0.267	0.291
3	1.857	0.199	0.199	0.199	0.199	0.371	0.365	0.370	0.371	0.270	0.289	0.280	0.314
4	2.394	0.192	0.192	0.192	0.192	0.381	0.375	0.379	0.381	0.227	0.246	0.246	0.264
5	3.086	0.201	0.201	0.201	0.201	0.421	0.417	0.420	0.421	0.313	0.306	0.313	0.308
6	4.041	0.215	0.215	0.215	0.215	0.404	0.398	0.401	0.403	0.346	0.317	0.329	0.295
7	5.473	0.192	0.192	0.192	0.192	0.434	0.433	0.436	0.433	0.284	0.304	0.296	0.313
8	7.897	0.246	0.246	0.246	0.246	0.567	0.584	0.586	0.564	0.274	0.274	0.274	0.274
9	12.959	0.670	0.674	0.673	0.672	1.790	1.932	1.883	1.755	0.751	0.872	0.863	0.738

$\sqrt{MSE}$

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.441	0.441	0.442	0.442	0.640	0.635	0.639	0.640	0.576	0.566	0.563	0.566
2	1.411	0.445	0.445	0.445	0.445	0.612	0.606	0.609	0.611	0.514	0.524	0.517	0.540
3	1.857	0.446	0.446	0.446	0.446	0.610	0.605	0.608	0.609	0.459	0.459	0.459	0.459
4	2.394	0.439	0.439	0.439	0.439	0.617	0.612	0.615	0.617	0.477	0.496	0.497	0.514
5	3.086	0.463	0.464	0.464	0.464	0.651	0.647	0.650	0.650	0.562	0.554	0.560	0.555
6	4.041	0.952	0.767	0.753	0.629	0.638	0.633	0.635	0.637	0.589	0.563	0.474	0.543
7	5.473	0.443	0.443	0.443	0.443	0.665	0.664	0.666	0.664	0.545	0.563	0.555	0.566
8	7.897	0.524	0.524	0.524	0.524	0.791	0.807	0.805	0.790	0.549	0.549	0.549	0.549
9	12.959	0.970	0.972	0.972	0.966	1.485	1.537	1.520	1.479	1.061	1.134	1.128	1.082

Table 3.16: Bias, Variance and MSE Estimates when estimating the Growth Curve where the responses are censored with fixed width intervals ($n = 200$)

Bias

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.002	0.002	0.002	0.002	-0.003	-0.003	-0.005	-0.003	0.013	0.013	0.013	0.013
2	1.411	-0.016	-0.016	-0.016	-0.016	0.002	0.002	0.001	0.002	-0.009	-0.009	-0.009	-0.009
3	1.857	-0.001	-0.002	-0.002	-0.002	0.002	0.002	0.001	0.002	0.017	0.017	0.017	0.017
4	2.394	0.010	0.010	0.010	0.010	-0.012	-0.011	-0.012	-0.012	0.020	0.020	0.020	0.020
5	3.086	0.011	0.011	0.011	0.011	0.018	0.018	0.017	0.018	0.015	0.015	0.015	0.015
6	4.041	0.031	0.031	0.030	0.031	0.017	0.018	0.017	0.017	0.028	0.028	0.028	0.028
7	5.473	0.062	0.062	0.062	0.062	0.027	0.029	0.028	0.027	0.062	0.062	0.064	0.065
8	7.897	0.205	0.206	0.206	0.205	0.114	0.121	0.119	0.116	0.098	0.099	0.099	0.098
9	12.959	0.576	0.576	0.576	0.568	0.426	0.426	0.439	0.426	0.445	0.446	0.445	0.437

Variance

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.212	0.212	0.212	0.212	0.131	0.130	0.131	0.131	0.108	0.109	0.109	0.109
2	1.411	0.212	0.212	0.212	0.212	0.124	0.123	0.123	0.124	0.120	0.120	0.120	0.120
3	1.857	0.205	0.205	0.205	0.205	0.128	0.127	0.128	0.128	0.108	0.108	0.108	0.108
4	2.394	0.206	0.206	0.206	0.206	0.139	0.138	0.138	0.139	0.116	0.116	0.116	0.116
5	3.086	0.198	0.198	0.198	0.198	0.136	0.135	0.135	0.136	0.107	0.107	0.107	0.107
6	4.041	0.191	0.192	0.193	0.191	0.123	0.122	0.122	0.123	0.121	0.121	0.121	0.121
7	5.473	0.213	0.213	0.213	0.213	0.131	0.130	0.131	0.131	0.113	0.114	0.114	0.114
8	7.897	0.254	0.254	0.254	0.254	0.144	0.146	0.146	0.143	0.137	0.137	0.137	0.137
9	12.959	0.687	0.691	0.690	0.687	0.314	0.335	0.333	0.313	0.297	0.299	0.298	0.298

$\sqrt{MSE}$

		$w = 0.5$				$w = 2$				$w = 4$			
x	g(x)	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE	MP	LLS (Normal)	LLS (NP)	LMLE
1	1.000	0.460	0.460	0.461	0.461	0.362	0.360	0.362	0.362	0.330	0.330	0.330	0.330
2	1.411	0.460	0.460	0.460	0.460	0.352	0.350	0.351	0.352	0.346	0.347	0.346	0.346
3	1.857	0.452	0.452	0.452	0.453	0.358	0.356	0.358	0.358	0.330	0.330	0.330	0.330
4	2.394	0.454	0.454	0.454	0.454	0.372	0.372	0.372	0.373	0.341	0.341	0.341	0.341
5	3.086	0.445	0.445	0.446	0.446	0.369	0.368	0.368	0.369	0.327	0.327	0.327	0.327
6	4.041	0.438	0.439	0.440	0.439	0.351	0.349	0.350	0.351	0.349	0.349	0.349	0.349
7	5.473	0.466	0.466	0.466	0.466	0.362	0.362	0.363	0.362	0.342	0.343	0.343	0.343
8	7.897	0.544	0.545	0.544	0.544	0.396	0.400	0.400	0.396	0.383	0.383	0.383	0.383
9	12.959	1.009	1.011	1.011	1.005	0.704	0.728	0.726	0.703	0.703	0.705	0.704	0.699

Table 3.17: Behaviour of midpoint imputation and maximum likelihood estimation on estimation of a constant mean with exponential intervals, for different sample sizes and exponential rates.

	$(n = 50, \mu = 0, \sigma = 1)$					
	$\lambda = 0.2$		$\lambda = 1$		$\lambda = 5$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	0.033	0.020	0.003	0.001	0.002	0.002
Variance	0.283	0.077	0.031	0.027	0.021	0.021
$\sqrt{\text{MSE}}$	0.533	0.278	0.176	0.163	0.144	0.143
	$(n = 100, \mu = 0, \sigma = 1)$					
	$\lambda = 0.2$		$\lambda = 1$		$\lambda = 5$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	-0.012	-0.002	-0.003	-0.003	0.003	0.003
Variance	0.128	0.036	0.014	0.013	0.011	0.011
$\sqrt{\text{MSE}}$	0.358	0.190	0.120	0.114	0.103	0.103
	$(n = 200, \mu = 0, \sigma = 1)$					
	$\lambda = 0.2$		$\lambda = 1$		$\lambda = 5$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	-0.005	0.001	-0.004	-0.003	0.003	0.003
Variance	0.066	0.017	0.007	0.006	0.005	0.005
$\sqrt{\text{MSE}}$	0.256	0.129	0.086	0.081	0.073	0.073

Table 3.18: Behaviour of midpoint imputation and maximum likelihood estimation on estimation of a constant mean with fixed width intervals, for different sample sizes and interval widths.

	$(n = 50, \mu = 0, \sigma = 1)$					
	$w = 0.5$		$w = 2$		$w = 4$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	0.001	0.001	0.007	0.005	0.017	0.006
Variance	0.020	0.020	0.028	0.027	0.077	0.011
$\sqrt{\text{MSE}}$	0.142	0.142	0.167	0.158	0.279	0.106
	$(n = 100, \mu = 0, \sigma = 1)$					
	$w = 0.5$		$w = 2$		$w = 4$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	0.006	0.006	0.005	0.004	-0.004	-0.002
Variance	0.011	0.011	0.013	0.012	0.039	0.006
$\sqrt{\text{MSE}}$	0.103	0.103	0.115	0.113	0.199	0.076
	$(n = 200, \mu = 0, \sigma = 1)$					
	$w = 0.5$		$w = 2$		$w = 4$	
Method	MP	MLE	MP	MLE	MP	MLE
Bias	-0.003	-0.003	0.001	0.001	0.002	0.001
Variance	0.005	0.005	0.007	0.006	0.020	0.003
$\sqrt{\text{MSE}}$	0.070	0.070	0.082	0.080	0.139	0.055

Table 3.19: Behaviour of midpoint imputation and maximum likelihood estimation on estimation of a constant mean with fixed width intervals, for different settings of the mean value and interval widths.

				$(n = 50,\ w = 4,\ \sigma = 1)$				
	$\mu = 0$		$\mu = 1$		$\mu = 2$		$\mu = 3$	
Method	MP	MLE	MP	MLE	MP	MLE	MP	MLE
Bias	-0.013	-0.005	0.360	-0.421	0.002	0.017	-0.372	0.415
Variance	0.072	0.011	0.045	0.043	0.015	0.566	0.044	0.045
$\sqrt{\text{MSE}}$	0.269	0.104	0.418	0.469	0.124	0.752	0.428	0.467
				$(n = 50,\ w = 6,\ \sigma = 1)$				
	$\mu = 0$		$\mu = 1$		$\mu = 2$		$\mu = 3$	
Method	MP	MLE	MP	MLE	MP	MLE	MP	MLE
Bias	0.006	0.002	1.038	-0.192	0.858	-0.198	-0.001	-0.008
Variance	0.193	0.028	0.092	0.021	0.016	0.600	0.002	0.306
$\sqrt{\text{MSE}}$	0.439	0.166	1.081	0.241	0.867	0.800	0.043	0.553
				$(n = 50,\ w = 8,\ \sigma = 1)$				
	$\mu = 0$		$\mu = 3$		$\mu = 4$		$\mu = 7$	
Method	MP	MLE	MP	MLE	MP	MLE	MP	MLE
Bias	-0.017	-0.006	0.990	0.867	0.000	0.000	-1.731	-0.080
Variance	0.314	0.045	0.002	0.266	0.000	0.018	0.156	0.038
$\sqrt{\text{MSE}}$	0.561	0.212	0.990	1.009	0.010	0.135	1.775	0.210

Burn No.	Slope	ROS (lower)	ROS (upper)
1	0.00	0.35	1.95
2	0.00	0.40	0.85
3	0.03	0.05	0.55
4	−0.03	0.65	2.20
5	0.03	1.90	3.00
6	0.03	1.10	1.35
7	0.03	2.15	2.70
8	0.03	1.10	2.60
9	0.03	0.95	1.95
10	−0.03	1.30	1.95
11	0.03	0.75	2.90
12	0.07	1.05	1.80
13	−0.07	1.00	1.60
14	−0.07	0.90	1.35
15	−0.02	0.95	1.80
16	−0.02	1.10	1.55
17	−0.03	0.75	1.50
18	−0.03	0.60	1.50
19	−0.05	1.10	1.35
20	−0.05	1.00	1.50
21	−0.08	0.50	1.15
22	−0.13	1.00	1.70
23	−0.13	0.65	1.00
24	−0.13	0.30	0.80
25	0.17	2.10	2.35
26	−0.17	0.50	0.85
27	−0.17	0.10	0.90
28	−0.17	0.40	0.75
29	−0.17	0.40	0.65
30	−0.17	0.25	0.55
31	0.00	0.90	1.55
32	0.00	0.90	1.55
33	0.00	1.30	2.00

Table 3.20: Rate of spread (ROS) measurements for a set of waxed paper fires at a variety of slopes. ROS (lower) denotes the left endpoint of the interval and ROS (upper) denotes the right endpoint of the interval containing the actual rate of spread.

Chapter 4

LOCAL REGRESSION SUBJECT TO CONSTRAINTS USING DATA SHARPENING

4.1 Introduction

Consider the regression model

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where the ε_i's are independent of each other and of the x_i's, and have mean 0 and variance σ^2. The mean function $g(x)$ is assumed to be smooth but also subject to a given condition or set of conditions, such as monotonicity, convexity, and so on.

Let

$$\widehat{g}(x) = \widehat{g}(x|\mathbf{y}) = \sum_{i=1}^{n} A_i(x) y_i \tag{4.2}$$

denote the kernel regression estimator for $g(x)$. For example, it can be a local constant, local linear or higher order local polynomial regression estimator. The form of $A_i(x)$ for the Nadaraya-Watson estimator was given at (2.4). Other estimators could also be considered as long as they can be written as a linear combination of the y's, the responses. The Priestley-Chao and Gasser-Müller estimators (see, e.g., Hall and Huang (2001)) fit into this category as do certain spline regression estimators.

The data-sharpened estimator of $g(x)$ is given by

$$\widehat{g}^\star(x) = \widehat{g}(x|\mathbf{y}^\star) = \sum_{i=1}^{n} A_i(x) y_i^\star, \tag{4.3}$$

where the $y_i^\star$'s are chosen to minimize the L_2 distance

$$D(\mathbf{y}^\star) = \sum_{i=1}^{n} (y_i - y_i^\star)^2$$

subject to any number of constraints which take the form

$$\mathbf{R}(x|\mathbf{y}^\star) = \sum_{i=1}^{n} B_i(x) y_i^\star - C(x) \geq 0 \tag{4.4}$$

or

$$\mathbf{R}(x|\mathbf{y}^\star) = \sum_{i=1}^{n} B_i(x) y_i^\star - C(x) = 0. \tag{4.5}$$

Here $B_i(x)$ is defined as follows:

$$B_i(x) = \sum_{s \in \mathbf{S}} D_i(x) A_i^{(s)}(x),$$

where $\mathbf{S}$ is a finite set of nonnegative integers, $A_i^{(s)}(x)$ is the s-th derivative of the function $A_i(x)$ and $D_i(x)$ is a function of x.

We restrict our attention to problems where the constraint on the estimated function can be expressed as linear functions of the responses.

If we further require that $\sum_{i=1}^{n} \frac{y_i^\star}{y_i} = n$, then we obtain an analogue to the constrained weighting method of Hall and Huang (2001) where it is required that $\sum_{i=1}^{n} p_i = 1$.

This is a quadratic programming problem with linear constraints. We can use the `solve.QP()` function in the *quadprog* library (Turlach and Weingessel (2007)), which makes the method very convenient. This estimator was first studied by Braun and Hall (2001), where most of the focus was on ensuring monotonicity of the regression function.

In this chapter, we will explore other possibilities and develop the theory. We will consider not only monotonicity and convexity constraints, but also point constraints, as well as functional and differential constraints, provided that these constraints are linear functions of the $\mathbf{y}'s$, i.e., they can be written in the form of (4.4) or (4.5). We will also make some comparisons with the constrained weighting approach proposed by Hall and Huang (2001); this partly meets a need identified by Racine and Parmeter (2009) since no one has compared these methods. We will refer to the Hall and Huang method as the HH method from now on.

The structure of the rest of the chapter is as follows. Section 4.2 is concerned with theoretical properties of the data sharpening method, including the problem of

existence of solutions and the rate of convergence of estimators to the true regression function. Simulation work is reviewed in Section 4.3, and the analysis of some real data sets will be discussed in Section 4.4.

4.2 Theoretical Properties

In this section we develop some theoretical properties of the estimator $\widehat{g}(x|\mathbf{y}^*)$. Much of this material follows Hall and Huang (2001) in their treatment of the HH method and gives parallel results for the class of data sharpening estimators we are studying.

In what follows, we assume $C(x) \equiv 0$ in the constraint on $\mathbf{R}(x)$. When $C(x)$ is not identically 0, analogous results can be obtained. Thus

$$\mathbf{R}(x|\mathbf{y}^\star) = \sum_{i=1}^{n} B_i(x) y_i^\star = \sum_{i=1}^{n} \left(\sum_{s \in \mathbf{S}} D_i(x) A_i^{(s)}(x) \right) y_i^\star.$$

The first result provides sufficient conditions for the existence of a solution to the data sharpening optimization problem.

Theorem 4.2.1 *Assume that the set* $\{1, \cdots, n\}$ *contains an increasing sequence* $i_1, \cdots, i_r$ *with the properties:*

(a) $U_{i_1}, U_{i_2}, \ldots, U_{i_r}$ *is a sequence of (possibly extended) real numbers, and* $V_{i_1}, V_{i_2}, \ldots, V_{i_r}$ *is an increasing sequence of real numbers where, for each* $k \in \{1, \ldots, r\}$, *the differences* $V_{i_k} - U_{i_k}$ *are strictly positive (but may be infinite, if* $U_{i_k} = -\infty$*).*

(b) The function $B_{i_k}(x)$ *is strictly positive and continuous on* (U_{i_k}, V_{i_k}), *and vanishes on* $(-\infty, U_{i_k}]$.

(c) $[a, b] \subset \bigcup_{k=1}^{r} (U_{i_k}, V_{i_k})$.

(d) For $1 \leq i \leq n$, $B_i(x)$ *is continuous on* $(-\infty, \infty)$.

(e) Each $y_{i_k} > 0$.

Then there exists a vector $y^\star = (y_1^\star, \cdots, y_n^\star)$ *such that* $y_i^\star > 0, i = 1, \cdots, n$ *and*

$$\mathbf{R}(x|\mathbf{y}^\star) = \sum_{i=1}^{n} B_i(x) y_i^\star > 0$$

for all $x \in \mathcal{J}$.

PROOF. By properties (a), (b) and (c), there must be at least one i_k for which $U_{i_k} < a$. Let j_1 denote the largest such value. Clearly, $V_{j_1} = V_{i_k} > a$.

For $\ell = 1, 2, \ldots$, inductively define $j_{\ell+1}$ to be the largest i_k such that $U_{i_k} < V_{j_\ell}$, as long as $a < V_{j_\ell} \leq b$. Note that $j_{\ell+1} > j_\ell$.

Let $s = \#\{j_\ell : V_{j_\ell} \leq b\} + 1$. Thus, the sequence $\{j_1, \ldots, j_s\}$ is strictly increasing.

Set $\delta = \min_k(V_{j_k} - U_{j_{k+1}})$. Because r is finite, $\delta > 0$ and

$$M = \min(V_{j_1} - a, V_{j_s} - b, \delta) > 0.$$

Let $\epsilon \in (0, M)$ be given, and set $y^\star_{j_1} = y_{j_1}$. Then

$$B_{j_1}(x) y^\star_{j_1} > 0$$

for $x \in [a, V_{j_1})$, by strict positivity of B_{j_1} on (U_{j_1}, V_{j_1}) (which contains $[a, V_{j_1})$.

Now, take a sequence $y^\star_{j_1}, \cdots, y^\star_{j_l}$, where $l < s$, such that

$$\sum_{k \leq l} B_{j_k}(x) y^\star_{j_k} > 0$$

on $[a, V_{j_l} - \epsilon]$. Note the following three facts:

1. $U_{j_{l+1}} < V_{j_l} - \epsilon$, (by definition of j_{l+1} and the choice of ϵ)
2. $B_{j_{l+1}} > 0$ on $[U_{j_{l+1}}, V_{j_{l+1}}]$, (by property (a))
3. $B_{j_k}(x)$ is bounded away from $-\infty$ (there exists a constant c such that $-\infty < c < B_{i_k}(x)$), uniformly in $k \leq l$ and in $-\infty < x < \infty$, (by property (c)).

Therefore, we can choose $y^\star_{j_{l+1}}$ sufficiently large and positive so that

$$\sum_{k \leq l+1} B_{j_k}(x) y^\star_{j_k} > 0$$

on $[a, V_{j_{l+1}} - \epsilon]$. This can be achieved by the following arguments: because we have $\sum_{k \leq l} B_{j_k}(x) y^\star_{j_k} > 0$ on $[a, V_{j_l} - \epsilon]$ and $U_{j_{l+1}} < V_{j_l} - \epsilon$, $\sum_{k \leq l} B_{j_k}(x) y^\star_{j_k} > 0$ on $[a, U_{j_{l+1}}]$. From the third fact, there exists a constant c such that $-\infty < c < B_{i_k}(x)$ in $k \leq l$ and in $-\infty < x < \infty$. So even if $\sum_{k \leq l} B_{j_k}(x) y^\star_{j_k} \leq 0$ for some $x \in$

$(U_{j_{l+l}}, V_{j_{l+1}}]$, the number $y^{\star}_{j_{l+1}}$ can be chosen sufficiently large and positive such that $\sum_{k \leq l+1} B_{j_k}(x) y^{\star}_{j_k} > 0$.

Using mathematical induction on l, we have proved the existence of integers $j_1 < \cdots < j_s$ and values of $y^{\star}_{j_1}, \cdots, y^{*}_{j_s}$ such that $\{j_1, \cdots, j_s\} \subseteq \{1, \cdots, n\}$. These $y^{\star}_{j_k}$'s satisfy the condition that

$$\sum_k B_{j_k}(x) y^{\star}_{j_k} > 0$$

on $[a, V_{j_s} - \epsilon]$. By the definition of ϵ, the interval $[a, V_{j_s} - \epsilon]$ contains the interval $\mathcal{J} = [a, b]$. Since $\{j_1, \cdots, j_s\} \subseteq \{1, \cdots, n\}$, for each $i \notin \{j_1, \cdots, j_s\}$, we take $y^{\star}_i = 0$, yielding a sequence $y^{\star}_1, \cdots, y^{\star}_n$ which satisfies

$$\mathbf{R}(x|\mathbf{y}^{\star}) = \sum_i B_i(x) y^{\star}_i > 0.$$

By property (d), we can obtain the result even with all $y^{\star}_i$ positive, for each $i \notin \{j_1, \cdots, j_s\}$, as long as these values are sufficiently small. Thus, we can construct a sequence of positive numbers $y^{\star}_1, \cdots, y^{\star}_n$ such that

$$\mathbf{R}(x|\mathbf{y}^{\star}) = \sum_i B_i(x) y^{\star}_i > 0.$$

□

In the next theorem, we constrain $\widehat{g}(x|\mathbf{y}^{*})$ to be monotone increasing and consider the case where the bandwidth h is chosen in a way that would be "asymptotically optimal" for $\tilde{g}$. That is , we take $h = O(n^{-\frac{1}{5}})$.

This result will also depend on the following conditions.

1. The estimator is of Gasser-Müller, Nadaraya-Watson, Priestley-Chao or local linear type.
2. $\mathcal{J}$ and $\mathcal{F} = [c, d]$ are compact intervals with $\mathcal{J} \subseteq \mathcal{F}$ for the local linear estimator, and $\mathcal{J} \subseteq [c + \delta, d - \delta]$, for some $\delta > 0$, for the other 3 estimators.

3. The design points are either equally spaced on $\mathcal{J}$, or they are independent random variables coming from a distribution whose density $f(x)$ is continuous and nonvanishing on $\mathcal{F}$.
4. $K(x)$ is a symmetric, compactly supported with Hölder continuous derivative.
5. $\inf_{x\in\mathcal{J}} g(x) = B_1 > 0$.
6. $E(|\varepsilon_i|^t)$ is bounded, for some $t > 0$.
7. $\sum_{i=1}^{n} \frac{y_i^\star}{y_i} = n$

   Let $\tilde{g}(x) \equiv \widehat{g}(x|\mathbf{y})$ denote the unconstrained estimator.

Theorem 4.2.2 *Assume conditions 1-7 given above, and suppose $t \geq 4$ in condition 6, and the functions $f(x)$ and $g(x)$ have two continuous derivatives on $\mathcal{J} = [a, b]$, and $h = O(n^{-1/5})$ as $n \to \infty$. We take the distance $D = D(\mathbf{y}^*)$.*

(a) If $g'(x) > 0$ for all x in $\mathcal{J}$, then with probability 1, $\mathbf{y}^\star = \mathbf{y}$ for all sufficiently large n. Hence, $\widehat{g}(x|\mathbf{y}^\star) \equiv \tilde{g}(x)$ on $\mathcal{J}$, for all sufficiently large n.

(b) Suppose that $g'(x) > 0$ for all x in $\mathcal{J}$ except at a single point x_0 at one of the endpoints of the interval $\mathcal{J} = [a, b]$. Further suppose $g'(x_0) = 0$ and $g''(x_0) \neq 0$ at this single point x_0. Then

$$|\widehat{g}(x|\mathbf{y}^\star) - \tilde{g}(x)| = O_P(h^2)$$

uniformly on the interval $\mathcal{J}$.

(c) There exist random variables $\Delta = \Delta(n)$ and $Z_1 = Z_1(n) \geq 0$, satisfying $\Delta = O_P(h^{\frac{5}{2}})$ and $Z_1 = O_P(1)$, such that

$$\widehat{g}(x|\mathbf{y}^*) = (1 + \Delta)\tilde{g}(x)$$

uniformly in $x \in \mathcal{J}$ such that $|x - x_0| > Z_1 h$. This means that for a random variable $Z_2 = Z_2(n) > 0$ that satisfies $Z_2 = O_P(1)$, we have $y_i^\star = (1 + \Delta)y_i$ for all indices i such that both $|X_i - x_0| > Z_2 h$ and $A_i(x) \neq 0$ for some $x \in \mathcal{J}$.

Before proving the theorem, we need the following results from Hall and Huang (2001).

Proposition 4.2.3 *Under the conditions 1-6 stated above,*

$$\sup_{x\in\mathcal{J}} \sum_{i=1}^{n} A_i^{(j)}(x)^2 = O_P(n^{-1}h^{-2j-1}),$$

where $A_i^{(j)}(x)$ *is the* j*-th derivative of* $A_i(x)$.

This result is stated in step (ii) of the proof of Theorem 4.3 of Hall and Huang (2001).

Proposition 4.2.4 *Assume conditions 1-6 above. For any constant* $C_1 > 0$*, and for each* $\delta > 0$*, there exists* $C_2 = C_2(\delta)$ *such that for sufficiently large* n*,*

$$P\{\tilde{g}'(x) > C_1 g'(x) \text{ for all } x \geq a + C_2 h\} > 1 - \delta. \tag{4.6}$$

The proof of this result is embedded in the proof of Theorem 4.3 statement (v) in Hall and Huang (2001).

Proposition 4.2.5 *Assume conditions* $1-6$ *above and* $g'(x) > 0$ *for* $x \in [a, b]$ *and* $g'(a) = 0$*. Then the following results hold.*

1. *Given* $\delta > 0$*, we may choose* $C > 0$ *large enough so that for all sufficiently large* n*,*

$$P\{\tilde{g}'(x) > 3h, \text{ for all } x \in [a + Ch, b]\} > 1 - \frac{1}{3}\delta.$$

2. *Given both* C *and* δ *we may define a fixed compactly supported, twice differentiable function* L *to be linearly decreasing, at a sufficiently fast rate, on a sufficiently wide interval containing the origin, and returning sufficiently slowly to* 0 *on either side of the interval where it is decreasing, such that*

$$P\{\tilde{g}'(x) + hC_k L'(\frac{x_0 - x}{h}) > 2h \text{ for all } x \in [a, a + Ch]\} > 1 - \frac{1}{3}\delta$$

and

$$hC_k L'(\frac{x_0 - x}{h}) > -h \text{ for all } x \in [a + Ch, b],$$

where $C_k = \int yK'(y)dy$.

These results are embedded in the proof of Theorem 2.2 in the appendix of Racine and Parmeter (2009). We also need the following additional result.

Lemma 4.2.6 *Assume conditions 1-7 above and* $g'(x) > 0$ *for* $x \in [a, b]$ *and* $g'(a) = 0$. *For any* $\delta > 0$, *there exists a random vector* $\tilde{\mathbf{y}}^\star = \tilde{\mathbf{y}}^\star(\delta)$ *such that*

$$P\{\widehat{g}'(x|\tilde{\mathbf{y}}^\star) > 0 \text{ for all } x \in \mathcal{J}\} > 1 - \delta,$$

and $D(\tilde{\mathbf{y}}^\star) = O_P(1)$.

PROOF. Let

$$\tilde{y}_i^\star = y_i \left(1 + \Delta + g(X_i)^{-1} h^2 L(\frac{x_0 - X_i}{h})\right), \tag{4.7}$$

where L is the function defined by Prop 4.2.5 and Δ is defined by the equality $\sum_i \frac{\tilde{y}_i^\star}{y_i} = n$. Here Δ is a random variable which does not depend on i and is given by $\Delta = -\frac{h^2}{n} \sum_{i=1}^n \frac{L(\frac{x_0 - X_i}{h})}{g(X_i)}$. Then we compute

$$\begin{aligned}
\widehat{g}'(x|\tilde{\mathbf{y}}^\star) &= \sum_{i=1}^n A_i'(x) \tilde{y}_i^\star \\
&= \sum_{i=1}^n A_i'(x) y_i \left(1 + \Delta + g(X_i)^{-1} h^2 L(\frac{x_0 - X_i}{h})\right) \\
&= \sum_{i=1}^n A_i'(x) \left[g(X_i) + \varepsilon_i\right] \left(1 + \Delta + g(X_i)^{-1} h^2 L(\frac{x_0 - X_i}{h})\right) \\
&= (1 + \Delta) \tilde{g}'(x) + h^2 \Delta_1(x) + h^2 \Delta_2(x),
\end{aligned}$$

where

$$\Delta_1(x) = \sum_{i=1}^n L(\tfrac{x_0 - X_i}{h}) A_i'(x),$$

$$\Delta_2(x) = \sum_{i=1}^n g(X_i)^{-1} L(\tfrac{x_0 - x_i}{h}) A_i'(x) \varepsilon_i.$$

As in Hall and Huang (2001), it is true that

$$\sum_{i=1}^n g(X_i)^{-1} L(\frac{x_0 - X_i}{h}) = O_P(nh).$$

By definition of Δ, we obtain that $\sum_i \frac{\tilde{y}_i^\star}{y_i} = n$ for Δ, and

$$\Delta = O_P(h^3). \tag{4.8}$$

Similarly, it is true that $\sup_{x\in\mathcal{J}}|\tilde{g}(x)| = O_P(1)$.

Using the Cauchy-Schwarz, we have

$$\begin{aligned}
\sup_x \left| \sum_i g(X_i)^{-1} L(\frac{x_0 - x_i}{h}) A_i'(x)\varepsilon_i \right| &\leq \sup_x \sqrt{\sum_i g(X_i)^{-2} L^2(\frac{x_0 - x_i}{h})\varepsilon_i^2 \sum_i (A_i'(x))^2} \\
&= \sqrt{\sum_i g(X_i)^{-2} L^2(\frac{x_0 - x_i}{h})\varepsilon_i^2 O_P(n^{-1}h^{-3})} \\
&\longrightarrow \sqrt{E[g(X_i)^{-2} L^2(\frac{x_0 - x_i}{h})\varepsilon_i^2] O_P(h^{-3})} \\
&= \sqrt{E[g(X_i)^{-2} L^2(\frac{x_0 - x_i}{h})] E[\varepsilon_i^2] O_P(h^{-3})} \\
&= \sqrt{O_P(\frac{1}{n}\sum_i g^{-2}(X_i) L^2(\frac{x_0 - x_i}{h})) O_P(h^{-3})} \\
&= \sqrt{O_P(h) O_P(h^{-3})} \\
&= O_P(h^{-1})
\end{aligned} \tag{4.9}$$

Thus

$$\sup_{x\in\mathcal{J}}|\Delta_2(x)| = O_P(h^{-1}).$$

For $\Delta_1(x)$, we have

$$\begin{aligned}
\Delta_1(x) &= h^{-2}n^{-1}\sum_{i=1}^n K'(\frac{x - X_i}{h}) L(\frac{x_0 - x}{h} + \frac{x - X_i}{h}) \\
&= h^{-1}\int K'(y) L(\frac{x_0 - x}{h} + y)dy + O_P(1).
\end{aligned}$$

A Taylor expansion of L gives

$$L(\frac{x_0 - x}{h} + y) = L(\frac{x_0 - x}{h}) + yL'(\frac{x_0 - x}{h}) + O_P(1). \tag{4.10}$$

Since $\int K'(y)dy = 0$, substituting (4.10) into the formula of $\Delta_1(x)$, we get

$$\begin{aligned}\Delta_1(x) &= h^{-1}\int K'(y)[L(\frac{x_0-x}{h}) + yL'(\frac{x_0-x}{h}) + O_P(1)]dy + O_p(1)\\ &= h^{-1}L(\frac{x_0-x}{h})\int K'(y)dy + h^{-1}L'(\frac{x_0-x}{h})\int yK'(y)dy + h^{-1}\int K'(y)dy + O_P(1)\\ &= h^{-1}L'(\frac{x_0-x}{h})\int yK'(y)dy + O_p(h^{-1}) + O_p(1).\end{aligned}$$

Let $C_k = \int yK'(y)dy$, we get

$$\begin{aligned}\widehat{g}'(x|\tilde{y}^\star) &= \tilde{g}'(x) + h^2h^{-1}C_kL'(\frac{x_0-x}{h}) + O_P(h) + O_P(h^2) + \Delta\tilde{g}'(x) + h^2\Delta_2(x)\\ &= \tilde{g}'(x) + hC_kL'(\frac{x_0-x}{h}) + \Delta_3(x), \qquad (4.11)\end{aligned}$$

where $\Delta_3(x) = O_P(h) + O_P(h^2) + \Delta\tilde{g}'(x) + h^2\Delta_2(x) = O_P(h)$ uniformly in $x \in \mathcal{J}$.

Furthermore, for all C and $\delta > 0$ in Proposition 4.2.5 and all sufficiently large n,

$$P\{\Delta_3(x) \geq -h \text{ for all } x \in [a,b]\} > 1 - \frac{1}{3}\delta.$$

Combining (4.11), Proposition 4.2.5 and the above results on Δ_3, we see that

$$P\{\widehat{g}'(x|\mathbf{y}^\star) > 0, \text{ for all } x \in \mathcal{J}\} > 1 - \delta,$$

for sufficiently large n.

For the order of $D(\mathbf{y}^\star)$, we compute from (4.7):

$$\begin{aligned}
\sum_{i=1}^{n}(\tilde{y}_i^\star - y_i)^2 &= \sum_{i=1}^{n}\left(y_i\left(\Delta + g(X_i)^{-1}h^2L(\frac{x_0 - X_i}{h})\right)\right)^2 \\
&= \sum_{i=1}^{n}(g(X_i)+\varepsilon_i)^2\left(\Delta + g(X_i)^{-1}h^2L(\frac{x_0 - X_i}{h})\right)^2 \\
&= \Delta^2\sum_{i=1}^{n}g(X_i)^2 + \Delta\sum_{i=1}^{n}\varepsilon_i^2 + 2\Delta^2\sum_{i=1}^{n}g(X_i)\varepsilon_i \\
&+ 2\Delta h^2\sum_{i=1}^{n}g(X_i)L(\frac{x_0 - X_i}{h}) + 2\Delta h^2\sum_{i=1}^{n}g(X_i)^{-1}L(\frac{x_0 - X_i}{h})\varepsilon_i^2 \\
&+ 4\Delta h^2\sum_{i=1}^{n}L(\frac{x_0 - X_i}{h})\varepsilon_i + h^4\sum_{i=1}^{n}L(\frac{x_o - X_i}{h})^2 \\
&+ 2h^4\sum_{i=1}^{n}g(X_i)^{-1}L(\frac{x_0 - X_i}{h}^2)\varepsilon_i + h^4\sum_{i=1}^{n}g(X_i)^{-2}L(\frac{x_0 - X_i}{h}^2)\varepsilon_i^2 \\
&= O_P(1).
\end{aligned}$$

The above equality can be proved using Chebyshev's inequality and Condition 6 with $t = 4$. □

Proof of Theorem 4.2.2.

(a) As pointed out in the proof of Theorem 4.3(a) of Hall and Huang (2001), $\tilde{g}'(x) = g'(x) + o(1)$ uniformly on $\mathcal{I}$, with probability 1, i.e.,

$$P(\lim_{n\to\infty}\sup\left|\tilde{g}'(x) - g'(x)\right| = 0) = 1.$$

So for any $\epsilon > 0$, there exists a N such that when $n > N$

$$\sup_x\left|\tilde{g}_n'(x) - g'(x)\right| < \epsilon.$$

Since $g'(x)$ is continuous function, it has a minimum on the interval $\mathcal{J}$. Denote this as min. Thus $g'(x) \geq \min > 0$ for all $x \in \mathcal{J}$. Take $\epsilon < \min$, then we have

$$\tilde{g}_n'(x) > g'(x) - \epsilon \geq \min - \epsilon > 0,$$

when $n > N$ and for all $x \in \mathcal{J}$. Hence, the constraint is satisfied by the original data with probability 1, for large enough n. Therefore, $\mathbf{y}^\star = \mathbf{y}$, with probability 1, for large enough n.

(b) First from Lemma 4.2.6, we have

$$D(\mathbf{y}^\star) = O_P(1).$$

From Proposition 4.2.3, we have

$$\sup_{x\in\mathcal{I}} \left(\sum_{i=1}^{n} (A_i^{(j)}(x))^2 \right) = O_P(n^{-1}h^{-2j-1}), \quad j = 0, 1.$$

Hence, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left| \widehat{g}^{(j)}(x|\mathbf{y}^\star) - \tilde{g}^{(j)}(x) \right| &= \left| \sum_{i=1}^{n} A_i^{(j)}(x)(y_i^\star - y_i) \right| \\ &\leq \left(\sum_{i=1}^{n} (y_i^\star - y_i)^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^{n} (A_i^{(j)}(x))^2 \right)^{\frac{1}{2}} \\ &= O_P(h^{2-j}), \end{aligned} \tag{4.12}$$

where $j = 0, 1$ and (j) means that the j-th derivative. This is true uniformly in $x \in \mathcal{J}$.

Letting $j = 0$, we have $|\widehat{g}(x|\mathbf{y}^\star) - \tilde{g}(x)| = O_P(h^2)$, uniformly in the interval $\mathcal{J}$.

(c) Since $K(\cdot)$ is continuous with compact support, there exists a number M such that

$$A_i(x) = 0, \text{ if } |x - X_i| \geq Mh.$$

Suppose that η is a random variable such that $\eta = a + O_P(h)$ and

$$y_i^\star = (1 + \Delta)y_i, \text{ for all } i \text{ with } X_i \geq \eta,$$

where Δ is a random variable satisfying $-1 < \Delta < \infty$ and is not depending on i. So if $x \geq Mh + \eta$ we have $y_i^\star = (1+\Delta)y_i$ for each i such that $A_i(x) \neq 0$. Thus $\widehat{g}(x|\mathbf{y}^\star) = \sum_i A_i(x) y_i^\star$ means that

$$\widehat{g}(x|\mathbf{y}^\star) = (1+\Delta)\tilde{g}(x)$$

if $x \geq Mh + \eta$. Then from Proposition 4.2.4, we can find a C_2 such that for any $C_1 > 0$ and $\delta > 0$, the following is true:

$$\begin{aligned}
& P\{\widehat{g}'(x|\mathbf{y}^\star) > C_1(1+\Delta)g'(x) \text{ for all } x \geq \max(a + C_2h, Mh + \eta)\} \qquad (4.13)\\
= \; & P\{(1+\Delta)\tilde{g}'(x|\mathbf{y}) > C_1(1+\Delta)g'(x) \text{ for all } x \geq \max(a + C_2h, Mh + \eta)\}\\
= \; & P\{\tilde{g}'(x|\mathbf{y}) > C_1 g'(x) \text{ for all } x \geq \max(a + C_2h, Mh + \eta)\}\\
\geq \; & P(\tilde{g}'(x|\mathbf{y}) > C_1 g'(x) \text{ for all } x \geq a + C_2h)\}\\
> \; & 1 - \delta, \qquad (4.14)
\end{aligned}$$

i.e.,

$$P\{\widehat{g}'(x|\mathbf{y}^\star) > C_1(1+\Delta)g'(x) \text{ for all } x \geq \max(a + C_2h, Mh + \eta)\} > 1 - \delta. \quad (4.15)$$

From the results in the proof of (b) (4.12), we have

$$\sup_{x\in\mathcal{J}} |\widehat{g}'(x|\mathbf{y}^\star) - \tilde{g}'(x)| = O_P(h).$$

Then there exists a random variable Z_2 such that

$$Z_2h = \inf\{z > 0 : y_i^\star = (1+\Delta)y_i, \text{ for all } i \text{ such that } |x_0 - X_i| \geq z\}$$

and $Z_2 = O_P(1)$. Then we have $y_i^\star = (1+\Delta)y_i$ for each i such that $|x_0 - X_i| \geq Z_2h\}$.

Let

$$Z_1h = \inf\{z \geq 0 : \text{for all } x \in \mathcal{J}, \text{ for which } x \geq a+z, \; A_i(x) = 0 \text{ whenever } y_i^\star \neq (1+\Delta)y_i\}.$$

Then $Z_1 = O_P(1)$, and

$$\widehat{g}(x|y^\star) = (1+\Delta)\tilde{g}(x)$$

uniformly on $x \in \mathcal{J}$ such that $x \geq a + Z_1 h$. From the above argument, we have

$$\sum_{i=1}^{n}(y_i^\star - y_i)^2 = \sum((1+\Delta)y_i - y_i)^2 = \Delta^2 \sum_{i=1} y_i^2.$$

Since we have $D(\mathbf{y}^\star) = \sum_{i=1}^{n}(y_i^\star - y_i)^2 = O_P(1)$ and $\mathbf{y}$ has a finite second moment, we see that

$$\Delta^2 = O_P(n^{-1}).$$

Therefore,

$$\Delta = O_P(h^{\frac{5}{2}}).$$

□

4.3 A Simulation Study

In this section, we will apply the data sharpening method and competing kernel regression methods to simulated data for which the regression function is known to satisfy particular conditions. In particular, we will simulate from models which exhibit monotonicity, convexity, and concavity; later in the section, we consider models where the qualitative information comes in the form of functional or differential operators.

Comparisons among the estimation methods will be made on the basis of bias, variance, pointwise MSE and MISE (mean integrated squared error).

4.3.1 Model H: Monotonicity

For an example of monotonic regression, we consider the monotonic example studied by Hall and Huang (2001):

$$g(x) = -x^3 + 3x \qquad \text{(Model H)}$$

This function is monotonically increasing on the interval $[-.9, .9]$.

In order to make comparisons with their results, we have used the same simulation settings as they used. To do this, we simulated 250 samples of size $n = 50$.

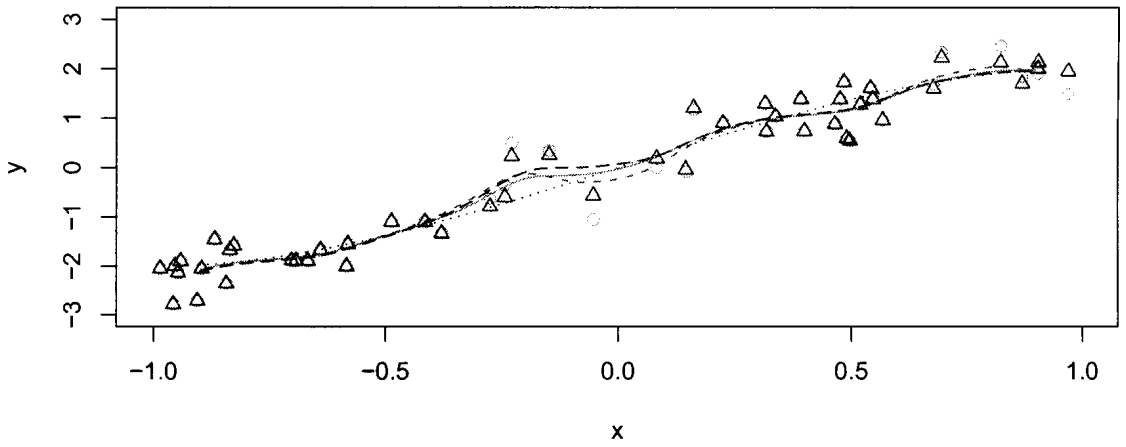

Figure 4.1: Constrained and unconstrained local linear estimates of the monotonic regression function (Model H, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), the HH estimate (long dashed curve), and the sharpened estimate (solid grey curve). Sharpened observations are represented by triangles.

The design points were randomly uniformly distributed on $[-1, 1]$. The errors were independent normal random variates with a standard deviation of 0.4.

For each simulated data set, we estimated the regression function using a local linear estimator without the monotonicity constraint as well as with the monotonicity constraint. To impose the constraint, we considered both the constrained weighting method of Hall and Huang (2001) (HH) and the data sharpening method (DS). In all cases, the biweight kernel function

$$K(x) = \frac{15}{16}(1 - x^2)^2 1_{\{|x| \leq 1\}}$$

was employed with bandwidth $h = 0.25$. As in Hall and Huang (2001), we took 100 equally spaced grid points on $[-.9, .9]$ when either imposing the constraints or plotting the curves.

Figure 4.1 demonstrates the estimates for a typical sample realization. It shows the simulated data points (open circles), sharpened data points (triangles), the true function (dotted curve), the local linear estimate (dashed curve), the data sharpened local linear estimate (solid curve) and the HH estimate (long dashed curve) under the monotonicity constraint. Most of the sharpened points coincide with original data so

only four unsharpened responses are visible.

The figure shows that the two constrained curves correct for the non-monotonicity displayed in the original estimator. The two constrained methods appear to be competitive with each other.

Figure 4.2 shows the weights used in the local linear HH estimator with constrained weighting. It also gives some summary information for the three estimators: pointwise squared bias, variance and mean squared error. If we compare with Figure 2 of Hall and Huang (2001), the most obvious feature is a marked decrease in the variance over most of $[-.9, .9]$. The bias of the constrained estimators is larger than for the unconstrained estimator. However, there is an overall smaller MSE; this is in agreement with the results of Hall and Huang (2001).

We also considered three versions of the bias-reduced data sharpening method of Choi *et al.* (2000) (to be denoted as CHR, from now on):

1. local constant regression, perturbing the design points only.

2. local linear regression, perturbing the design points and responses.

3. local linear regression, perturbing the responses only.

We did not include the performance results for these methods in Figure 4.2 in order to avoid clutter. However, we tabulated some summary information in Table 4.1 in order to make comparisons with the constrained estimators. MISE was computed for each method, as well as MSE values at the design points $x = -.1$ and $x = .7$.

The table shows that the CHR estimators tend to produce larger values of MISE than the unconstrained method. The constrained methods appear to give slightly better MISE and MSE performance than their unconstrained counterpart.

4.3.2 Models M1 and M2: Convexity Followed by Concavity

For this case, we aim to compare the unconstrained local linear estimators with the local linear estimator of two functions which are known to be convex in one region and concave in another. The regression functions we study are those that Mammen (1991*a*) considered when he proposed a constrained spline estimator to handle qualitative constraints:

$$g(x) = 15x(x - 0.5)(1 - x) \qquad \text{(Model M1)}$$

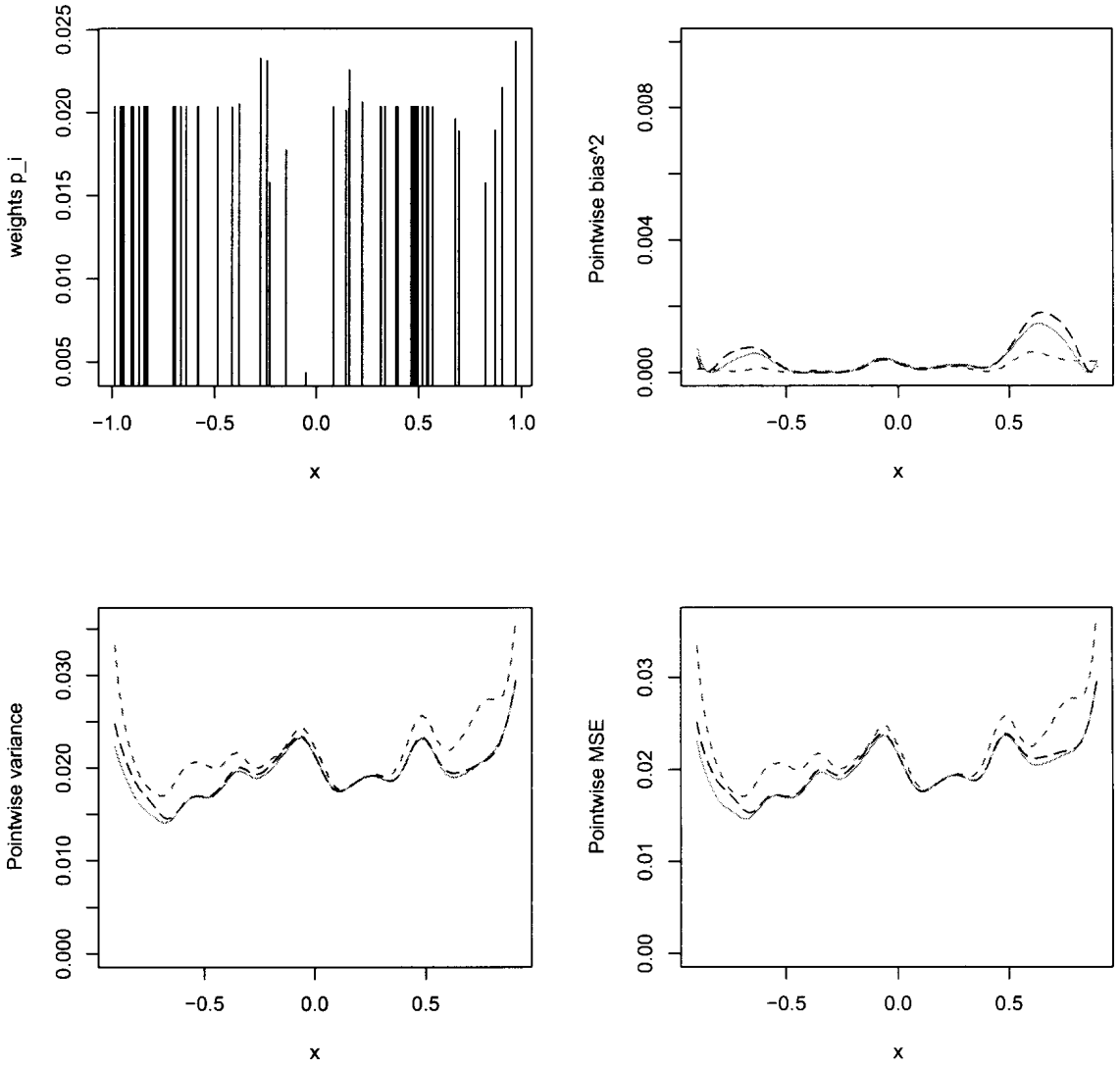

Figure 4.2: MSE Comparison of local linear estimators for Model H. Top left panel: HH weights for the responses pictured in Figure 4.1, plotted against the design points. The other panels give squared bias, variance and MSE of the unconstrained estimator (short dashed curve), HH estimator (long dashed curve), and data sharpened estimator (solid curve).

Table 4.1: MISE and MSE (at two design points) comparisons of local estimates of the monotonic regression function (Model H). CHR constant: local constant with perturbed design points; CHR linear-xy: local linear with perturbed design points and responses; CHR linear-y: local linear with perturbed responses; DS: data sharpening; HH: the method of Hall and Huang (2001).

Target Function: Monotone Function (Model H)			
		Mean squared error	
Estimator	MISE	at $x = -0.1$	at $x = 0.7$
local linear estimate	0.040	0.024	0.025
CHR constant	0.063	0.038	0.041
CHR linear-xy	0.061	0.037	0.038
CHR linear-y	0.056	0.035	0.035
DS	0.035	0.023	0.021
HH	0.036	0.023	0.022

and

$$g(x) = -\frac{10}{3}x + \frac{25}{3}(x - 0.3)_+ - \frac{25}{3}(x - 0.7)_+ \qquad \text{(Model M2)}$$

These models are referred to as $\mu_1(x)$ and $\mu_2(x)$ in Mammen (1991*a*). In order to compare the constrained data sharpening method with Mammen's estimator, we used his simulation settings. Thus, we simulated 1000 samples of size 200. Design points were taken to be equally spaced in $[0, 1]$. Two noise models were considered: both are normal with $\sigma = .1$ and with $\sigma = .5$.

Kernel regression estimates were calculated using the biweight kernel. The MISE-optimal bandwidth h was used in each case.

Data sharpening proceeded as follows. The sharpened responses were chosen as the minimizers of

$$\sum_{i=1}^{n} (y_i - y_i^\star)^2 \qquad (4.16)$$

subject to the constraints

$$\widehat{g}''(z|\mathbf{y}^\star) \geq 0, \qquad z < I$$

and

$$\widehat{g}''(z|\mathbf{y}^\star) \leq 0, \qquad z > I$$

where z was one of 401 equally spaced grid points in the interval $[0, 1]$. In order to select I, the above minimization was conducted for the 41 grid points in the neighbourhood of 0.5 (the true value). The grid point that gave the smallest value of the distance at (4.16) was taken as the estimate of I.

For a typical simulation run, constrained and unconstrained local linear estimates of $\mu_1(x)$ and $\mu_2(x)$ have been plotted in Figures 4.3 to 4.6. Figures 4.7, 4.8, 4.9 and 4.10 give the weights used in the HH estimator for typical simulated samples for each pair of cases under Models M1 and M2. The pointwise squared bias, variance and mean squared error are also plotted, in each case, for the sharpened constrained estimator, the HH estimator, and their unconstrained local linear estimator. Noteworthy are the difficulties with bias near .2 and .8, for Model M1 (the constrained estimators provide some reduction here). For Model M2, the discontinuities in the derivative at .3 and at .7 render the bias very large in all cases, and the constrained estimators are unable to cope with this any better than the unconstrained estimator. However, the constrained estimators are less biased in the region immediately surrounding these problem points. Variance is also smaller in this region, so the overall MSE is reduced almost everywhere when the constraints are imposed.

Again, we computed the CHR estimates, as described in Section 4.3.1, and we computed the HH estimate under the same constraint as was used for data sharpening, but performance measures on these estimators were not included in the above figures in order to avoid clutter.

Table (4.2) gives the MSE at $x = 0.5$ and $x = 0.8$ and the MISE over the interval $[0.1, 0.9]$ for all estimators under consideration, including the constrained, unconstrained, CHR and Mammen's estimator.

As we can see, the constrained data sharpening local linear estimator has the smallest MISE among all the estimators for the smooth function $\mu_1(x)$. The local linear HH estimator also behaves well but with a slightly larger MISE. Mammen's spline estimator has the smallest MISE among all the estimators for the broken-line curve $\mu_2(x)$ when $\sigma = 0.1$. This is not a surprising result, since his estimator is the only one that allows for discontinuities in the first derivative, among all other estimators considered here. When $\sigma = .5$, the spline estimator degrades because it tends to follow the trends in the noise too closely. As Mammen points out, even though the MISE behaviour of the local linear estimate is smaller, it will not necessarily respect the shape of the estimator. Here, again, the data sharpening estimator performs reasonably well in terms of accuracy at the larger noise level, but it is also successful in getting the shape right.

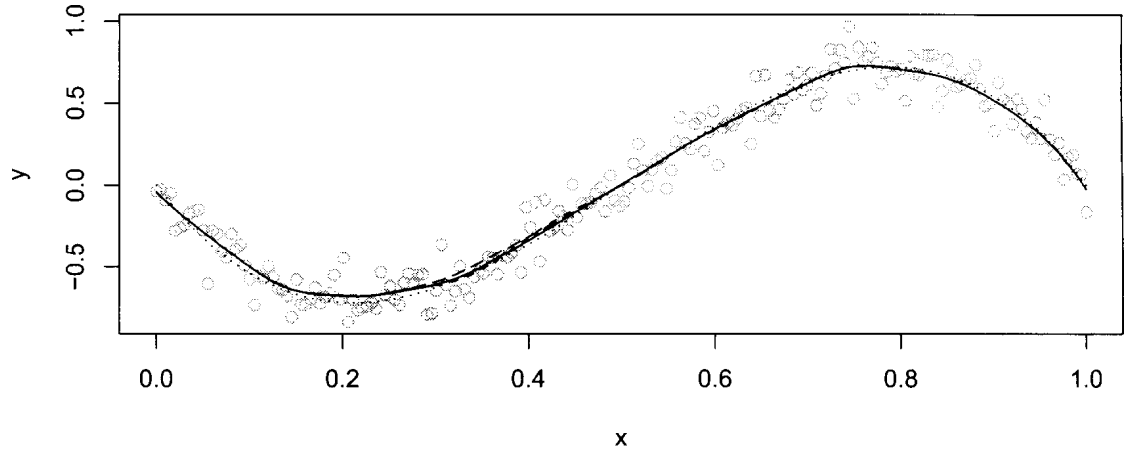

Figure 4.3: Constrained and unconstrained local linear estimates of the regression function (Model M1, $\sigma = .1$, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), the HH estimate (long dashed curve), and the sharpened estimate (solid curve).

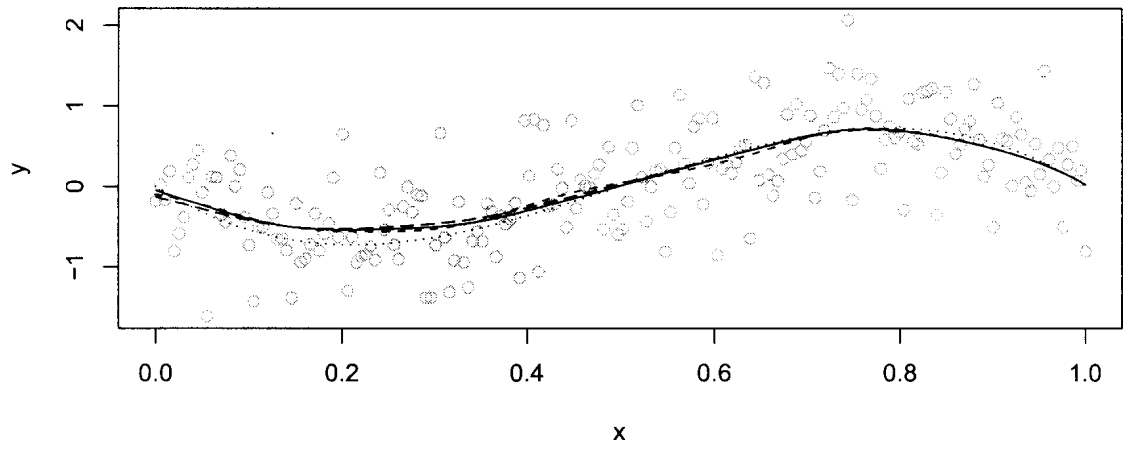

Figure 4.4: Constrained and unconstrained local linear estimates of the regression function (Model M1, $\sigma = .5$, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (dashed curve), the HH estimate (long dashed curve), and the sharpened estimate (solid curve).

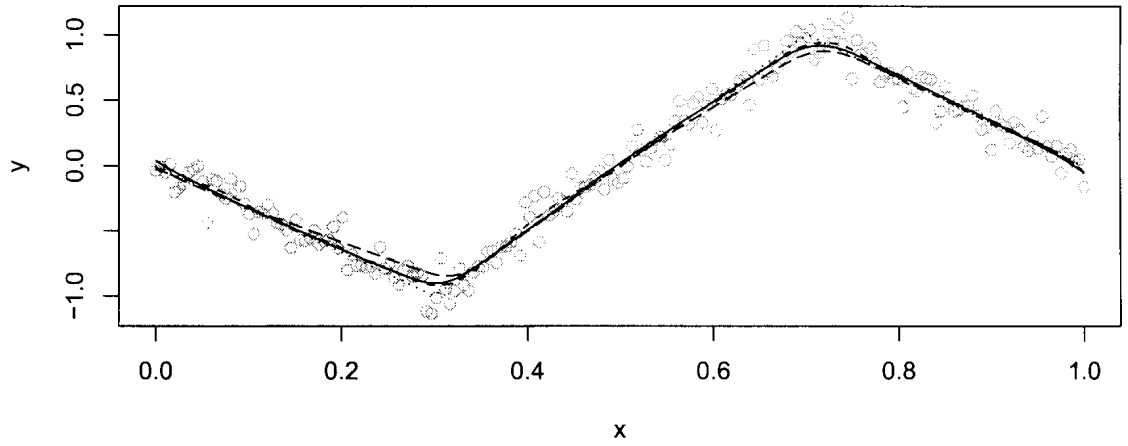

Figure 4.5: Constrained and unconstrained local linear estimates of the regression function (Model M2, $\sigma = .1$, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), the HH estimate (long dashed curve), and the sharpened estimate (solid curve).

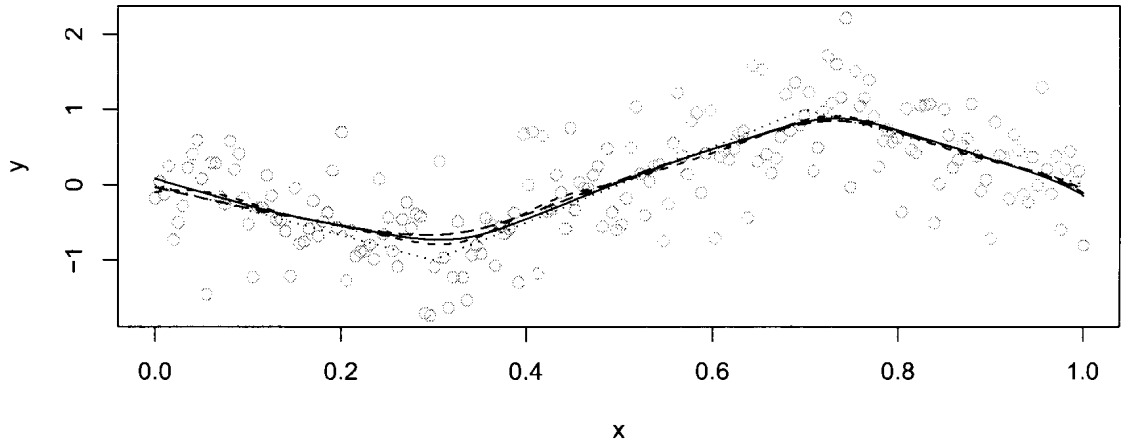

Figure 4.6: Constrained and unconstrained local linear estimates of the regression function (Model M2, $\sigma = .5$, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), the HH estimate (long dashed curve), and the sharpened estimate (solid curve).

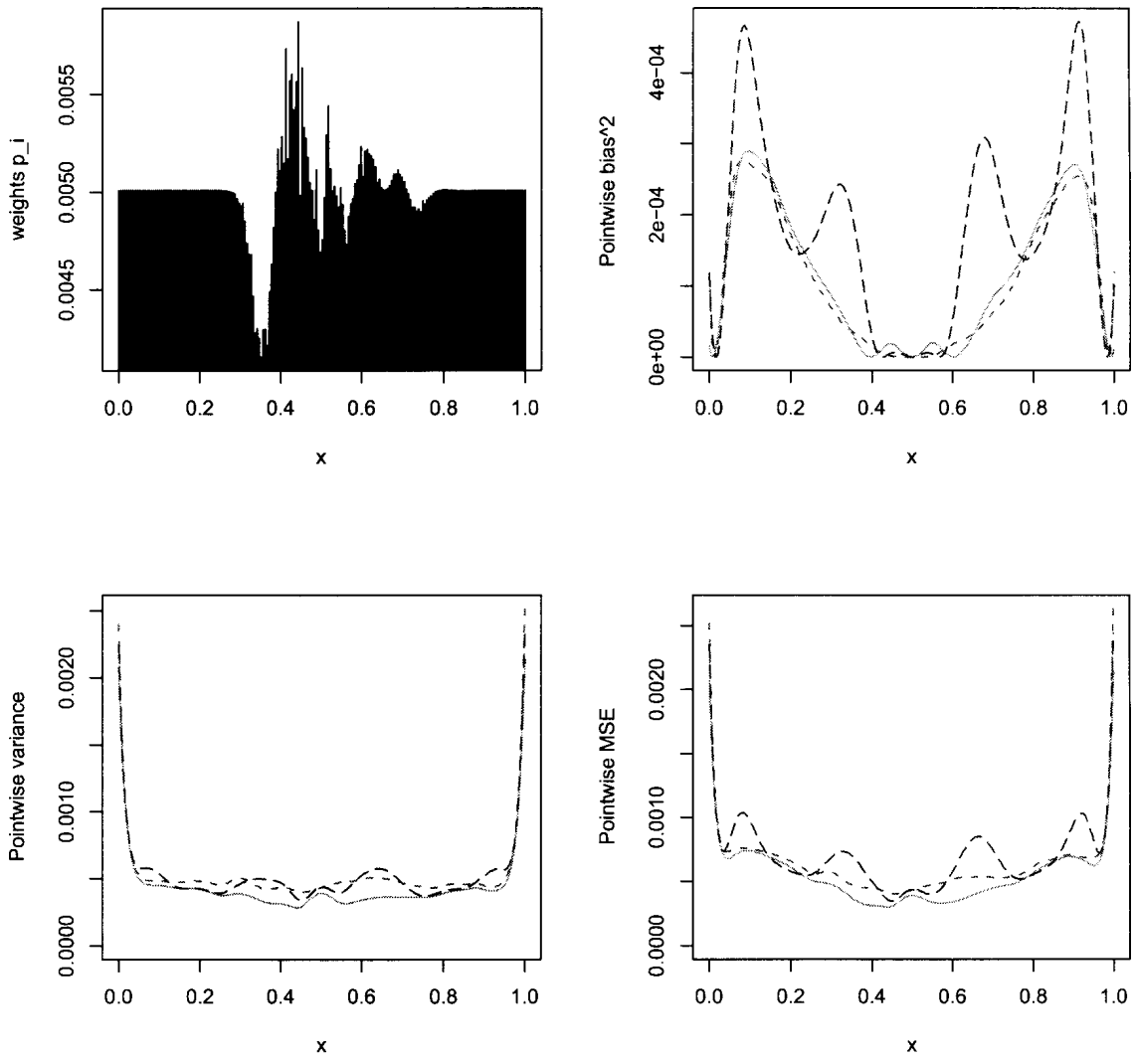

Figure 4.7: MSE Comparison of local linear estimators for Model M1 ($\sigma = .1$). Top left panel: HH weights for the responses pictured in Figure 4.3, plotted against the design points. The other panels give squared bias, variance and MSE of the unconstrained estimator (short dashed curve), HH estimator (long dashed curve), and data sharpened estimator (solid curve).

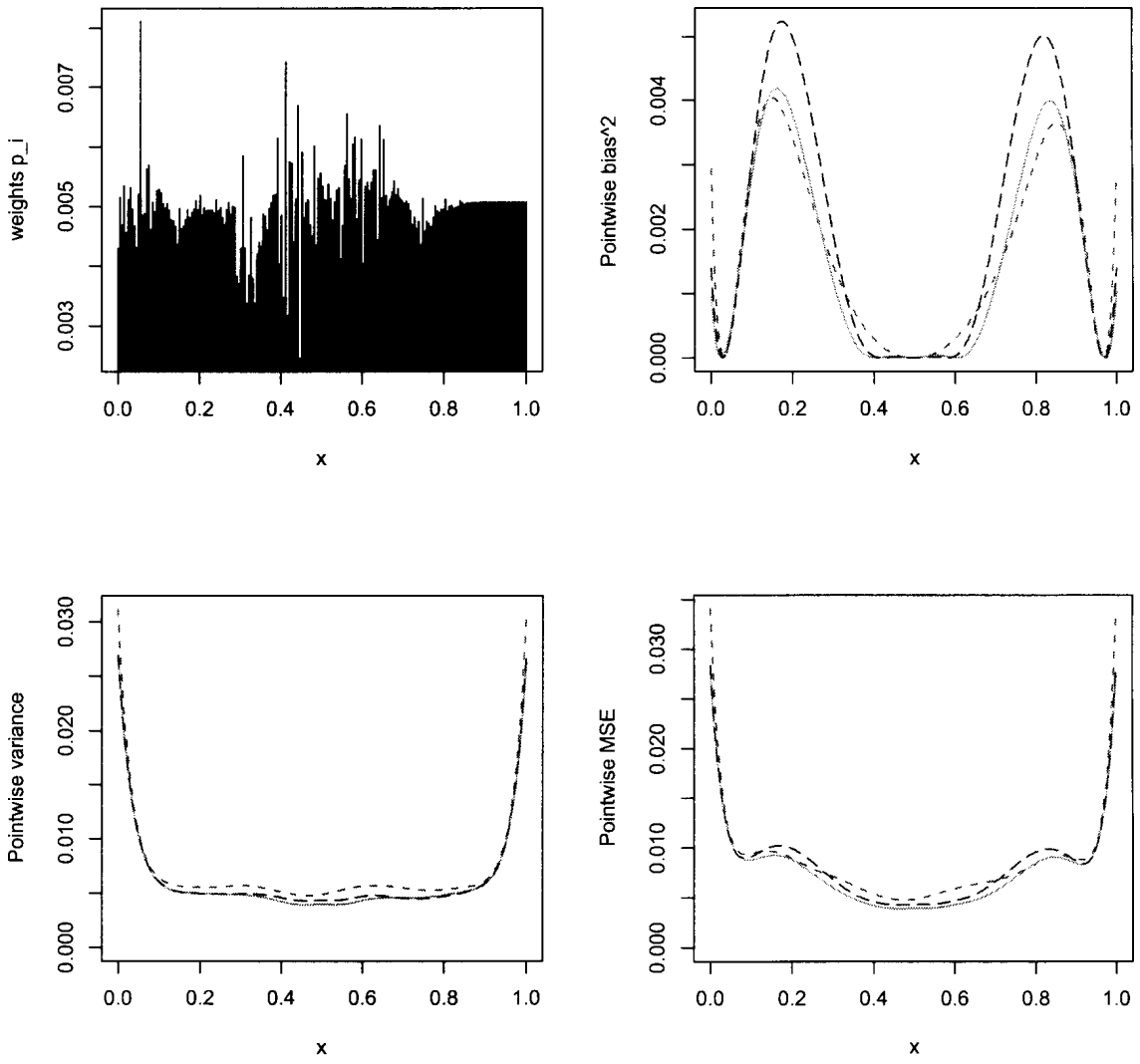

Figure 4.8: MSE Comparison of local linear estimators for Model M1 ($\sigma = .5$). Top left panel: HH weights for the responses pictured in Figure 4.4, plotted against the design points. The other panels give squared bias, variance and MSE of the unconstrained estimator (short dashed curve), HH estimator (long dashed curve), and data sharpened estimator (solid curve).

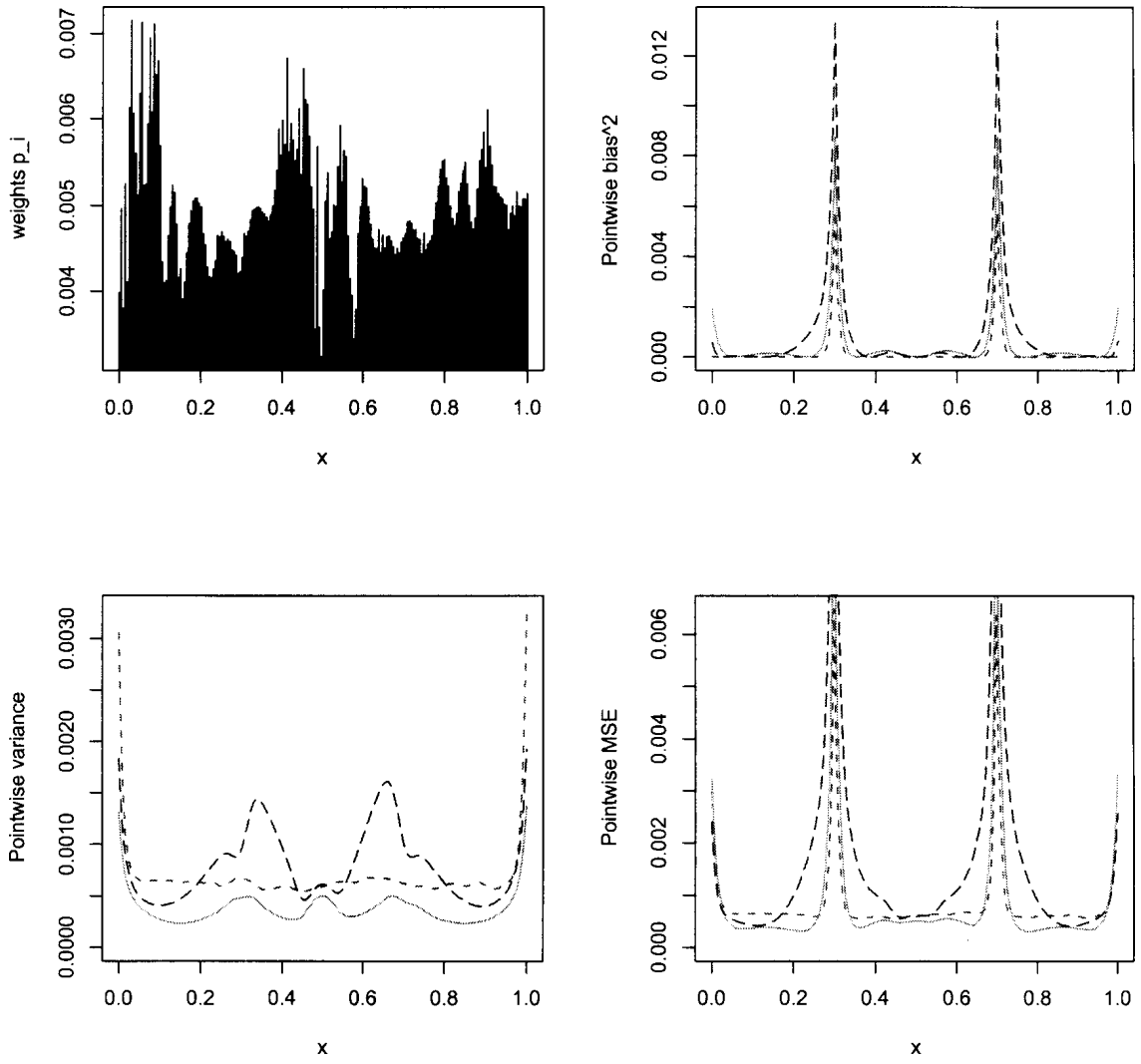

Figure 4.9: MSE Comparison of local linear estimators for Model M2 ($\sigma = .1$). Top left panel: HH weights for the responses pictured in Figure 4.5, plotted against the design points. The other panels give squared bias, variance and MSE of the unconstrained estimator (short dashed curve), HH estimator (long dashed curve), and data sharpened estimator (solid curve).

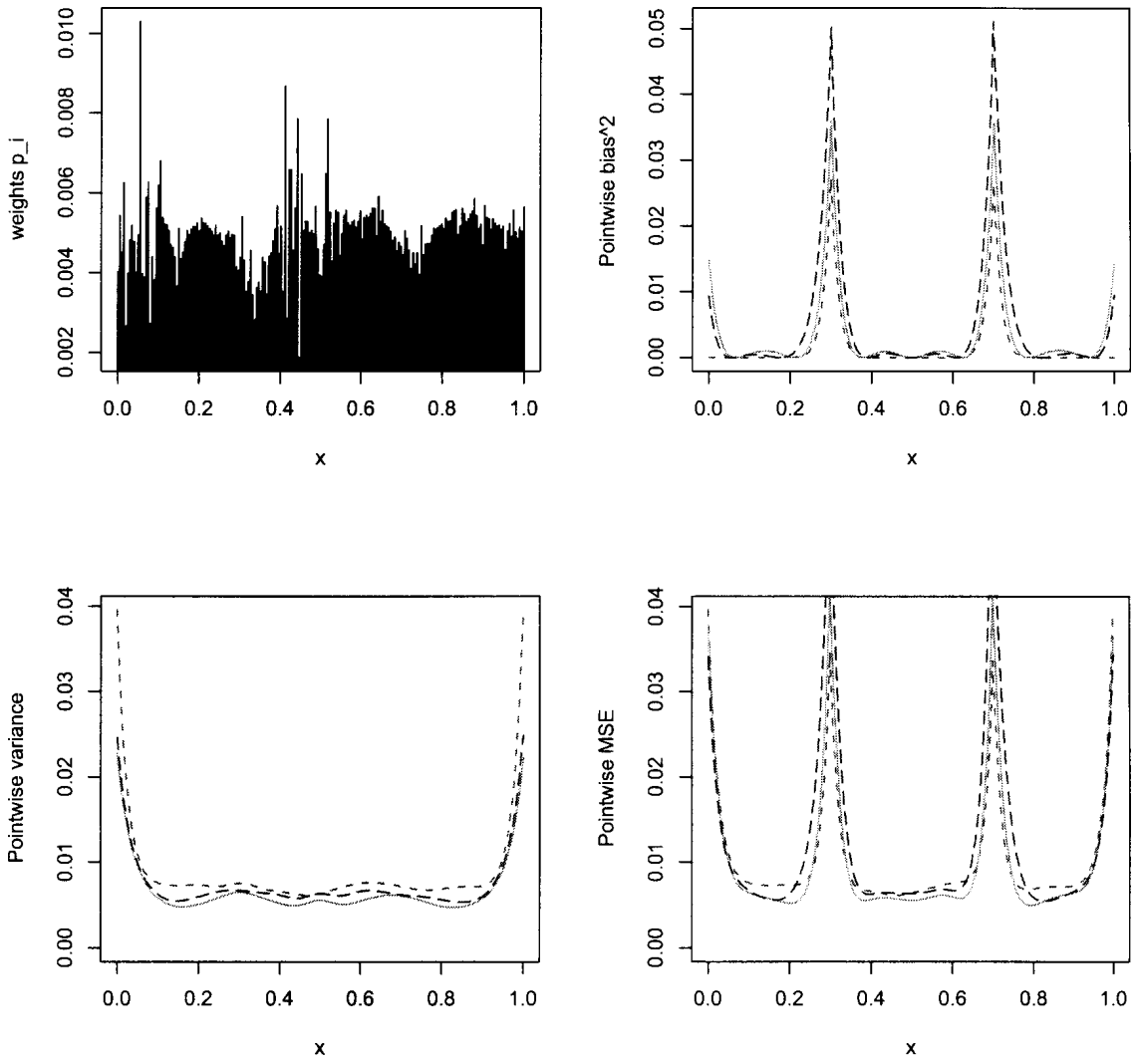

Figure 4.10: MSE Comparison of local linear estimators for Model M2 ($\sigma = .5$). Top left panel: HH weights for the responses pictured in Figure 4.6, plotted against the design points. The other panels give squared bias, variance and MSE of the unconstrained estimator (short dashed curve), HH estimator (long dashed curve), and data sharpened estimator (solid curve).

4.3.3 Incorporating Functional Information

In addition to imposing monotonic and convex/concave constraints, other constraints which incorporate qualitative functional information can be accommodated with data sharpening.

Sometimes it is known that the regression function is a solution to a differential equation. In many circumstances, the differential equation is not solvable analytically, and it may be that there are difficulties associated with the numerical solution as well. Furthermore, it is possible that the regression function is not exactly equal to a solution to the given differential equation, but that the differential equation has been set up as an approximation.

In such cases, it is desirable to have an estimation procedure which makes better use of sample information than is ordinarily the case when one solves a differential equation in practice. It is also desirable to have a regression method which would depend on the nature of the differential equation rather than the solution itself. In cases where the analytic solution is known, nonlinear least-squares presents itself as a natural way of fitting the solution to data; when the analytic solution is unknown or only an approximation to the true function, techniques of functional data analysis have been developed; these methods typically employ spline functions. One purpose of this section is to develop a parallel set of methods for kernel regression estimators. Data sharpening appears to be a natural way to achieve this goal.

This idea will be illustrated by two examples.

4.3.3.1 Model N: The Normal Distribution Function

On the basis of physical considerations, the temperature T of a long narrow object (e.g., a rod) at various locations x along the length of the object can be modeled as

$$T = g(x) + \varepsilon. \tag{4.17}$$

where $g(x)$ satisfies the differential equation

$$g''(x) = -\frac{xg'(x)}{\sigma^2}$$

Table 4.2: MISE and MSE (at two design points) comparisons of local estimates of the regression function (Models M1 and M2). CHR constant: local constant with perturbed design points; CHR linear-xy: local linear with perturbed design points and responses; CHR linear-y: local linear with perturbed responses; DS: data sharpening; HH: the method of Hall and Huang (2001).

Target Function: Models M1 and M2(Mammen (1991*b*))

					Mean squared error	
Estimator	h	σ	Model	MISE ($\times 10^2$)	at $x = 0.5$ ($\times 10^2$)	at $x = 0.8$ ($\times 10^2$)
linear spline		0.1	M1	0.063	0.091	0.065
local linear estimate		0.1	M1	0.044	0.043	0.061
CHR constant		0.1	M1	0.054	0.064	0.065
CHR linear-xy		0.1	M1	0.054	0.064	0.065
CHR linear-y		0.1	M1	0.053	0.064	0.065
DS		0.1	M1	0.038	0.039	0.057
HH		0.1	M1	0.049	0.044	0.055
linear spline		0.1	M2	0.059	0.056	0.035
local linear estimate		0.1	M2	0.085	0.059	0.060
CHR constant		0.1	M2	0.084	0.087	0.088
CHR linear-xy		0.1	M2	0.084	0.087	0.088
CHR linear-y		0.1	M2	0.084	0.087	0.088
DS		0.1	M2	0.092	0.050	0.033
HH		0.1	M2	0.170	0.061	0.090
linear spline		0.5	M1	0.98	1.22	0.96
local linear estimate		0.5	M1	0.57	0.49	0.84
CHR constant		0.5	M1	0.64	0.70	0.82
CHR linear-xy		0.5	M1	0.72	0.70	0.92
CHR linear-y		0.5	M1	0.63	0.70	0.78
DS		0.1	M1	0.50	0.39	0.85
HH		0.1	M1	0.56	0.43	0.96
linear spline		0.5	M2	1.24	1.23	0.83
local linear estimate		0.5	M2	0.69	0.64	0.70
CHR constant		0.5	M2	0.89	0.94	1.06
CHR linear-xy		0.5	M2	0.92	0.94	1.09
CHR linear-y		0.5	M2	0.88	0.94	1.01
DS		0.1	M2	0.73	0.56	0.50
HH		0.1	M2	0.97	0.64	0.59

with initial conditions $g(0) = 0$ and $g'(0) = \sqrt{\frac{2}{\pi}}\sigma$. The constant σ is related to physical properties of the rod, e.g, density and heat capacity, and rate of heat flow at a particular location at a particular time. Under some circumstances, it is imaginable that σ may be a known constant, or measurable to satisfactory precision.

The solution to this equation is given by

$$g(x) = 2\left(\Phi(\frac{x}{\sigma}) - \Phi(0)\right) \qquad \text{(Model N)}$$

where $\Phi(x)$ is the normal cumulative distribution function.

We have used this solution to simulate data from the model (4.17) in order to study the effectiveness of data sharpening in the estimation of functions such as $g(x)$ under a variety of constraint conditions. The possible constraints we considered are:

$$|g(x) - 2\left(\Phi(\frac{x}{\sigma}) - \Phi(0)\right)| < \epsilon \qquad \text{(Constraint N1)}$$

$$|g'(x) - \frac{2}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}| < \epsilon \qquad \text{(Constraint N2)}$$

and

$$|g''(x) + g'(x)\frac{x}{\sigma^2}| < \epsilon \qquad \text{(Constraint N3)}$$

where ϵ is a small value which is to be chosen in some manner.

For our simulation study, we assumed fixed design points which are equally spaced in the interval $[0, 10]$, and the standard deviation of the error term was 0.2. The parameter σ in function (Model N) was taken to be 2. 1000 samples of sizes 30, 50 and 100 were used in this study.

For each simulated data set, the following estimation procedures were applied: local linear estimation, the CHR bias reduction approach, constrained data sharpening, the HH weighting method and nonlinear least squares. The bandwidth was chosen with the direct plug-in selector `dpill` (see, Sheather and Jones (1986)) which is in the R package *KernSmooth* (Wand and Ripley (2009)). The normal kernel was used.

For the data sharpening and HH weighting estimators, each of the three constraint conditions listed above was tried, in turn. We chose the parameter ϵ using a cross-validation criterion.

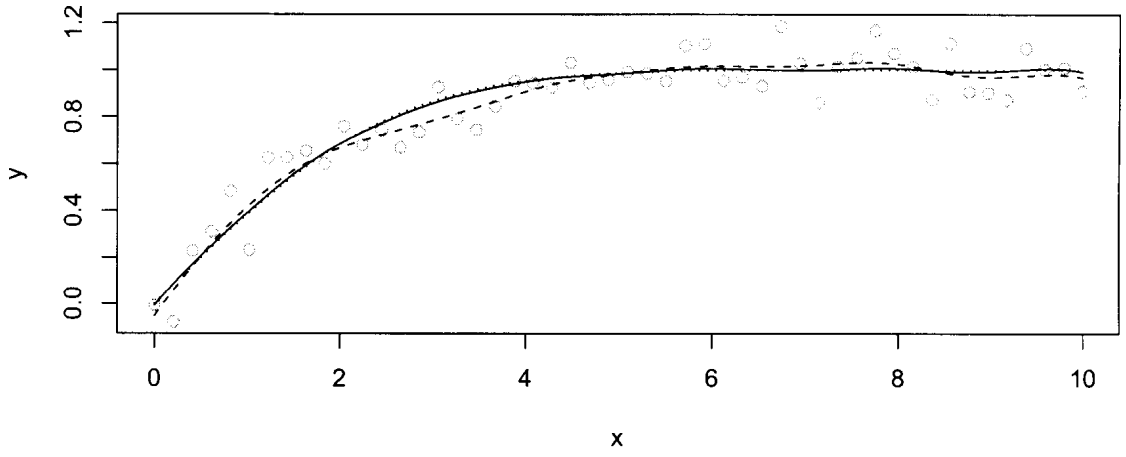

Figure 4.11: Constrained and unconstrained local linear estimates of the regression function (Model N, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid grey curve) subject to Constraint N1.

Typical simulation runs for the $n = 50$ case are plotted in Figures 4.11, 4.12, and 4.13. Figure 4.11 allows for comparison of the unconstrained local linear estimate with estimates subjected to Constraint N1. Figure 4.12 shows similar graphs but where Constraint N2 is operative, and Figure 4.13 involves Constraint N3.

Point-wise bias, variance and MSE is plotted in Figure 4.14 for the $n = 50$ case, for the unconstrained local linear estimator, together with the three versions of the data sharpened estimator and the three versions of the HH estimator. This plots show clearly that the qualitative information contained in the differential constraints (N2 and N3) is less useful than in the functional constraint (N1). However, these constraints are providing some improvement over the unconstrained estimator in terms of accuracy.

The same story is told in Table 4.3 where the three CHR methods (described in Section 4.3.1) are also being compared. None of the CHR methods appear to offer an increase in accuracy over unconstrained local linear estimation, while even the weakest constraint information (Constraint N3) provides an improvement. In that case, the HH procedure is best, while under Constraints N1 and N2, it is best to use data sharpening.

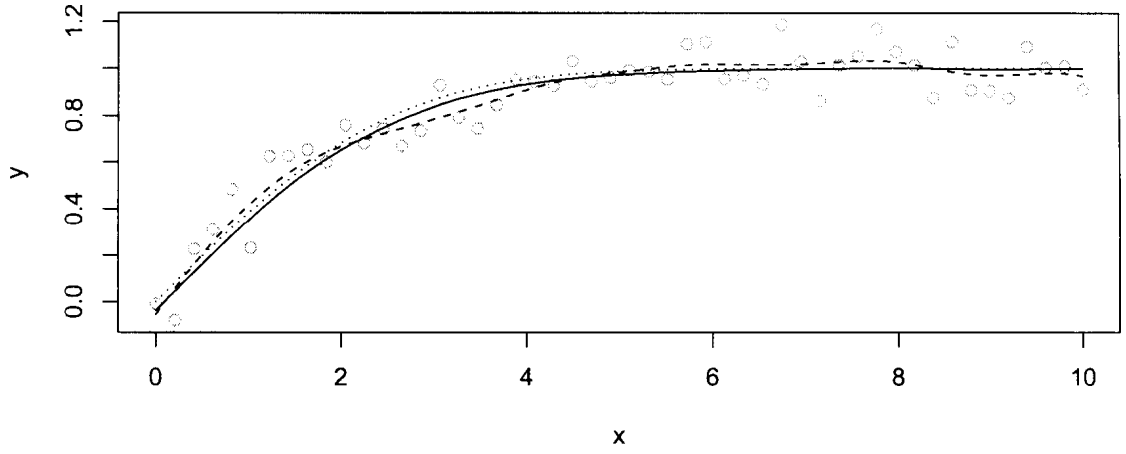

Figure 4.12: Constrained and unconstrained local linear estimates of the regression function (Model N, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid grey curve) subject to Constraint N2.

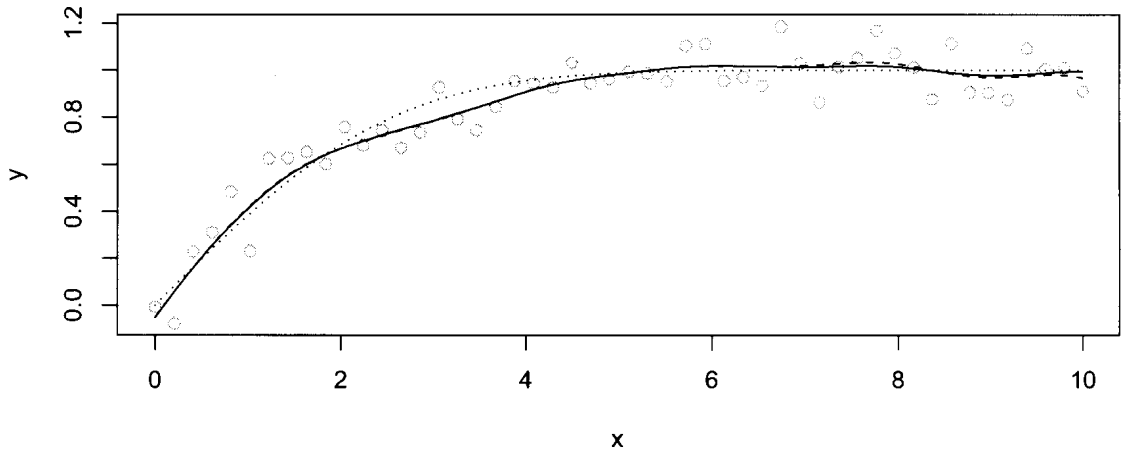

Figure 4.13: Constrained and unconstrained local linear estimates of the regression function (Model N, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid grey curve) subject to Constraint N3.

Table 4.3: MISE and MSE (at two design points) comparisons of local estimates of the regression function (Model N). CHR constant: local constant with perturbed design points; CHR linear-xy: local linear with perturbed design points and responses; CHR linear-y: local linear with perturbed responses; DS: data sharpening; HH: the method of Hall and Huang (2001).

Target Function: CDF of Normal Distribution

			Mean squared error	
Estimator	n	MISE	at $x = 5$	at $x = 8$
local linear estimate	30	0.015	0.001	0.001
CHR constant		0.020	0.002	0.002
CHR linear-xy		0.021	0.002	0.002
CHR linear-y		0.018	0.002	0.002
DS Case N1		0.003	0.000	0.000
HH Case N1		0.004	0.000	0.000
DS Case N2		0.007	0.001	0.001
HH Case N2		0.010	0.001	0.001
DS Case N3		0.013	0.001	0.001
HH Case N3		0.011	0.001	0.001
local linear estimate	50	0.010	0.001	0.001
CHR constant		0.014	0.001	0.001
CHR linear-xy		0.015	0.001	0.001
CHR linear-y		0.012	0.001	0.001
DS Case N1		0.002	0.000	0.000
HH Case N1		0.003	0.000	0.000
DS Case N2		0.004	0.000	0.000
HH Case N2		0.007	0.001	0.001
DS Case N3		0.009	0.001	0.001
HH Case N3		0.007	0.001	0.001
local linear estimate	100	0.006	0.001	0.001
CHR constant		0.008	0.001	0.001
CHR linear-xy		0.008	0.001	0.001
CHR linear-y		0.007	0.001	0.001
DS Case N1		0.001	0.000	0.000
HH Case N1		0.001	0.000	0.000
DS Case N2		0.002	0.000	0.000
HH Case N2		0.004	0.000	0.000
DS Case N3		0.005	0.000	0.000
HH Case N3		0.004	0.000	0.000

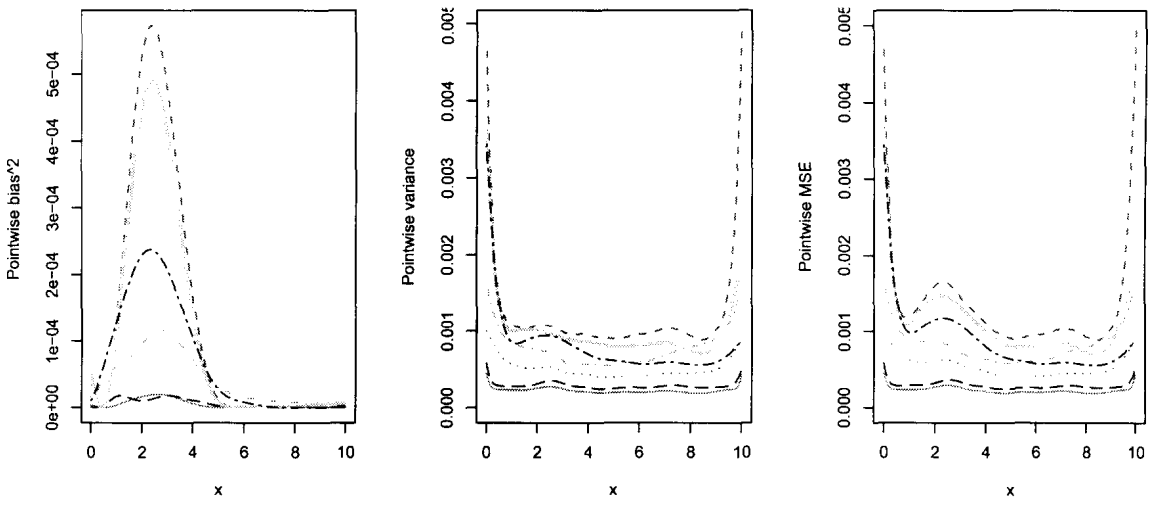

Figure 4.14: Performance measures for 7 local linear estimators applied to Model N. The panels give, respectively, pointwise squared bias, variance and mean squared error for the unconstrained estimator (dashed curve), data sharpening subject to Constraint N1 (solid curve), subject to Constraint N2 (dotted curve), subject to Constraint N3, (thick solid curve); the HH estimator subject to Constraint N1 (long dashed curve), subject to Constraint N2 (grey dot-dashed curve), and subject to Constraint N3 (black dot-dashed curve).

4.3.3.2 Model D: A Linear First Order Differential Equation

Assume the mean function $g(x)$ in model (4.1) is the solution of a differential equation of the form

$$g'(x) = s(x)g(x) + h(x) \tag{4.18}$$

where $g(0) = g_0$, and $s(x) = S'(x)$, for some sufficiently smooth function $S(x)$, and $s(x)$ and $h(x)$ are assumed to be known functions. Usually, $S(x)$ is unknown, though it can be obtained by integration, possibly numerically.

The solution to this differential equation is given by

$$g(x) = e^{S(x)} \int_0^x h(t)e^{-S(t)}dt + g_0 e^{S(x)}. \tag{4.19}$$

The integral which is explicitly written into equation (4.19) also needs to be computed, again, possibly numerically.

The constrained data sharpening method provides a possible alternative approach to estimating the function $g(x)$ when data are available. To illustrate the

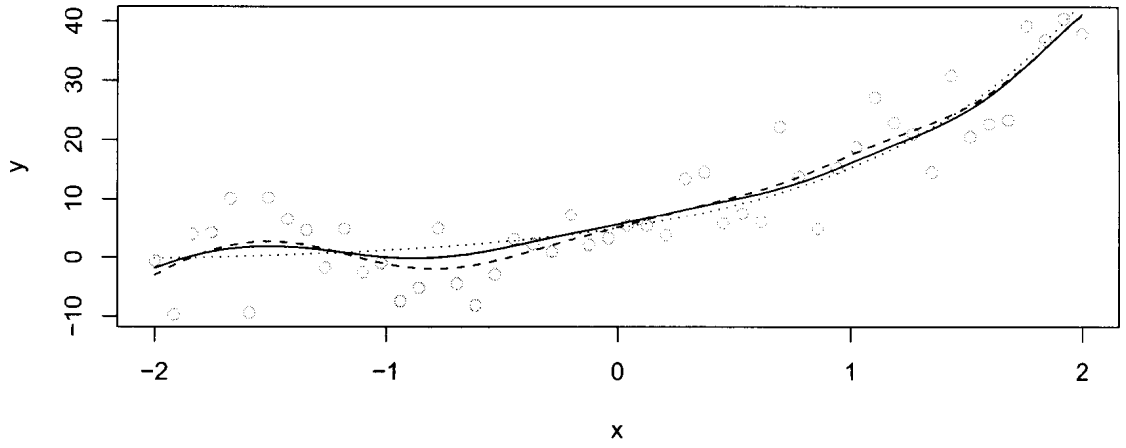

Figure 4.15: Constrained and unconstrained local linear estimates of the regression function (Model D, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid curve) subject to Constraint D1.

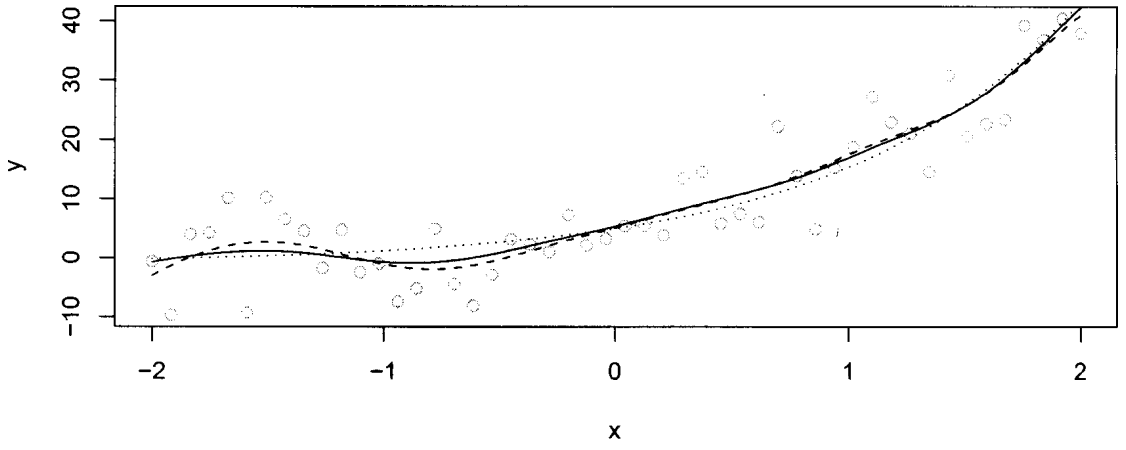

Figure 4.16: Constrained and unconstrained local linear estimates of the regression function (Model D, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid curve) subject to Constraint D2.

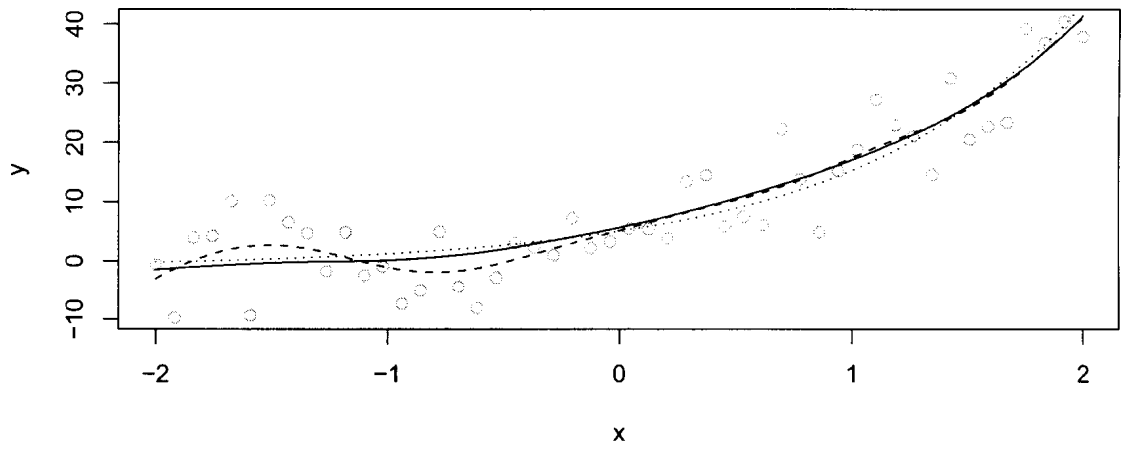

Figure 4.17: Constrained and unconstrained local linear estimates of the regression function (Model D, dotted curve) for a typical realization of simulated data (open circles). Pictured are the unconstrained estimate (short dashed curve), and the sharpened estimate (solid curve) subject to Constraint D3.

method, we consider a simple special case of (4.18):

$$g'(x) = \theta_1 x^{\theta_1 - 1} g(x) + \theta_2 \tag{4.20}$$

for given constants θ_1 and θ_2. In this case, we know that

$$g(x) = e^{x^{\theta_1}} \int_0^x \theta_2 e^{-t^{\theta_1}} dt + e^{x^{\theta_1}} g_0, \tag{4.21}$$

We will consider data coming from the particularly simple case where $\theta_1 = 1$ and $\theta_2 = 1$:

$$g(x) = e^x(1 + g_0) - 1. \tag{Model D}$$

We experimented with this model by running 1000 simulations of samples at sizes 30, 50 and 100, coming from

$$y = e^x(1 + g_0) - 1 + \varepsilon$$

where the noise was assumed to be normally distributed with standard deviation

$\sigma = 6$, and the initial condition was taken as $g_0 = 5$. The design points were taken to be equally spaced on the interval $[-2, 2]$.

For each data set we computed constrained and unconstrained local linear estimates of the regression function. The bandwidth was chosen with the direct plug-in selector `dpill` which is in the R package *KernSmooth* (Wand and Ripley (2009)). The normal kernel was used.

The following constraints were employed, in turn, in the data sharpening and HH procedures:

$$|g(x) - e^x(1 + g_0) + 1| < \epsilon, \qquad \text{(Constraint D1)}$$

$$|g'(x) - g(x) - 1| < \epsilon, \qquad \text{(Constraint D2)}$$

and

$$|g''(x) - g'(x)| < \epsilon \qquad \text{(Constraint D3)}$$

To employ Contraint D1, it is necessary to estimate g_0. This was done by regressing $y - e^x + 1$ on e^x. An alternative procedure would be to use nonlinear least squares. The parameter ϵ was selected using a cross-validation procedure.

Constrained and unconstrained local linear estimates of the regression function, under each constraint condition, are displayed in Figures 4.15 to 4.17, for typical samples coming from the simulation.

Figure 4.18 displays the pointwise squared bias, variance and MSE for the unconstrained and constrained estimators. In this case, the HH method is very close to the data sharpening method, so we have not included it here. We see that the qualitative information provided by each of the constraints leads to an increase in accuracy. Noteworthy is the fact that Contraint 2 is essentially as informative as Constraint 1. Constraint 3 does not depend on direct knowledge of the value of θ_2, so it is providing less information about $g(x)$ than the other two constraints.

Table (4.4) lists the MISE as well as MSE values at two design points. It also includes results for the three versions of the CHR bias-reduction method described in Section 4.3.1. Again, this general purpose bias reduction strategy does not give an improvement. The HH method and the data sharpening method both give a substantial improvement in MISE and MSE accuracy. The degree of improvement depends on the type of constraint imposed.

Table 4.4: MISE and MSE (at two design points) comparisons of local estimates of the regression function (Model D). CHR constant: local constant with perturbed design points; CHR linear-xy: local linear with perturbed design points and responses; CHR linear-y: local linear with perturbed responses; DS: data sharpening; HH: the method of Hall and Huang (2001).

Target Function: (Model D)				
			Mean squared error	
Estimator	n	MISE	at $x = 0$	at $x = 1.2$
local linear estimate	30	20.475	4.516	6.146
CHR constant		32.222	6.208	7.691
CHR linear-xy		28.700	6.214	8.333
CHR linear-y		25.450	6.180	6.804
DS Case D1		7.862	1.317	2.878
HH Case D1		8.382	1.538	2.970
DS Case D2		7.091	1.127	2.459
HH Case D2		7.791	1.312	2.509
DS Case D3		12.535	2.663	3.445
HH Case D3		13.362	2.893	3.556
local linear estimate	50	13.869	2.951	4.006
CHR constant		22.349	4.078	4.902
CHR linear-xy		19.791	4.080	5.593
CHR linear-y		17.479	4.064	4.419
DS Case D1		5.185	0.791	1.850
HH Case D1		5.249	0.846	1.778
DS Case D2		4.880	0.730	1.649
HH Case D2		4.993	0.750	1.857
DS Case D3		8.081	1.642	2.249
HH Case D3		8.373	1.683	2.566
local linear estimate	100	7.945	1.719	2.414
CHR constant		13.141	2.375	2.817
CHR linear-xy		11.299	2.375	3.194
CHR linear-y		9.967	2.374	2.547
DS Case D1		2.894	0.467	1.064
HH Case D1		2.890	0.508	0.982
DS Case D2		2.574	0.377	0.922
HH Case D2		2.877	0.440	0.987
DS Case D3		4.498	0.909	1.373
HH Case D3		4.984	0.900	1.450

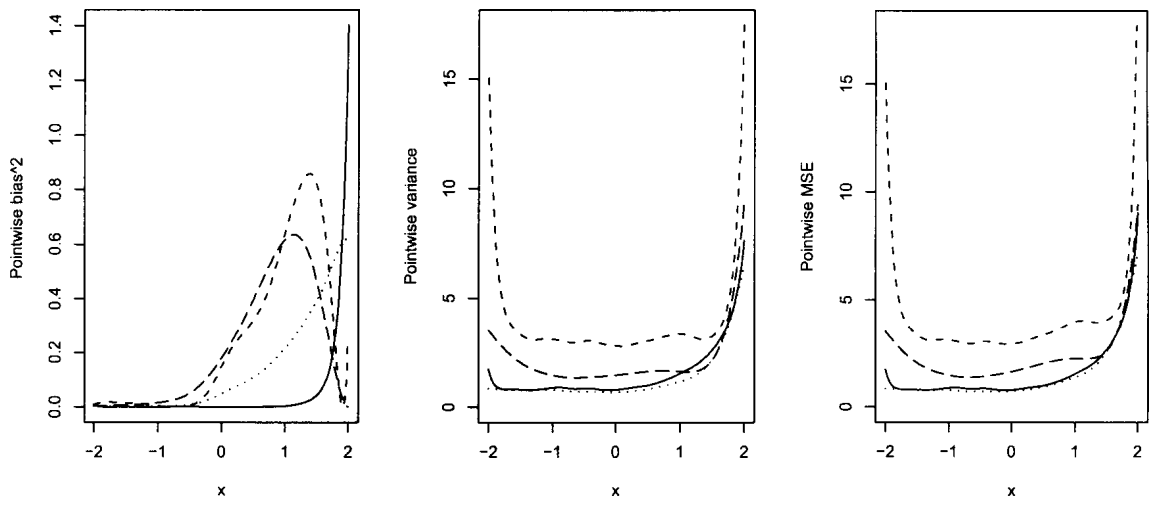

Figure 4.18: Performance measures for local linear estimators applied to Model D. The panels give, respectively, pointwise squared bias, variance and mean squared error for the unconstrained estimator (short dashed curve), data sharpening subject to Constraint D1 (solid curve), subject to Constraint D2 (dotted curve), and subject to Constraint D3 (long dashed curve).

4.4 Applications

In this section, we will illustrate the use of constrained local linear estimation on four data sets. These examples illustrate cases of non-decreasing monotonicity, non-increasing monotonicity, concavity and periodicity.

We will also compute point-wise bootstrap confidence bands for the regression function, and we will apply a test for the validity of the constraints. The test we will use is similar to one proposed by Hall *et al.* (2001) and further studied by Racine and Parmeter (2009).

The null hypothesis corresponds to the case that the true regression function satisfies the constraint. We use the objective function $D(\hat{y}^\star)$ as the test statistic, and we reject the null hypotheis if the observed value of $D(\hat{y}^\star)$ is too large. A bootstrap procedure enables us to compute the p-value for this test.

For the given data set $\{x_i, y_i\}$, $i = 1, \cdots, n$, we compute the sharpened responses $\hat{y}_i^\star$ and the corresponding regression estimates $\widehat{g}^\star(x_i)$ subject to the constraint. Sharpened residuals are defined as:

$$\hat{\varepsilon}_i^\star = \hat{y}_i^\star - \widehat{g}^\star(x_i)$$

We randomly resample the sharpened residuals to approximate a new sample of errors: $\{\hat{\varepsilon}_i^{\star b}\}$. These bootstrap residuals added to the estimated regression function in order to generate the resampled responses.

$$\hat{y}_i^b = \widehat{g}^\star(x_i) + \hat{\varepsilon}_i^{\star b}.$$

Because of the way in which the regression function was estimated, these responses are from a data set that satisfies the null hypothesis. This new data set can be sharpened as well, and the distance measure can again be calculated:

$$D(\hat{y}^{\star b}) = \sum_{i=1}^{n} (\hat{y}_i^b - \hat{y}_i^{\star b})^2.$$

Repeating this resampling process a large number of times allows us to estimate P_B, the proportion of the bootstrap resamples in which $D(\hat{y}^{\star b})$ exceeds $D(\hat{y}^\star)$.

4.4.1 Radiocarbon Data

These data were originally published by Pearson and Qua (1993), and a subset was analyzed by Bowman and Azzalini (1997) and Hall and Huang (2001). We use the same subset. The variables are radiocarbon age, predicted from a radiocarbon dating process, and the true calendar age.

Figure 4.19 shows the data. Overlaid as a dashed curve is the local linear estimate obtained with a normal kernel, using bandwidth $h = 30$, which was the one suggested by Bowman and Azzalini (1997). Also pictured is the local linear estimate based on data sharpening subject to the monotonicity constraint; the same bandwidth was used. The HH estimate is not plotted, since it is almost indistinguishable from the curve obtained by data sharpening.

The original data and the corresponding sharpened data are shown in the left panel of Figure 4.20. The weights p_i from the HH approach are shown in the right panel. It is clear that the sharpened data points that correspond to x_i's which are not in the immediate vicinity of dips and bumps in the unconstrained estimate, are coincidental with the original data points. This was predicted by arguments in Section 4.2. Note the pattern of the weights as well. Where the unconstrained estimate is non-

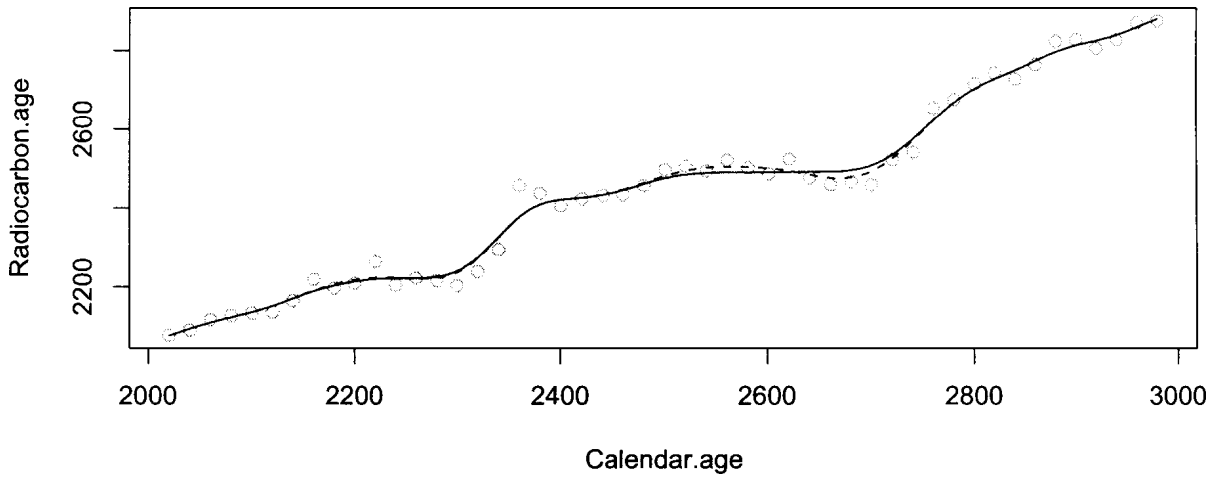

Figure 4.19: Subjecting Radiocarbon data to monotonicity. The constrained sharpening local linear estimate (solid line); the unconstrained local linear estimate(dashed line).

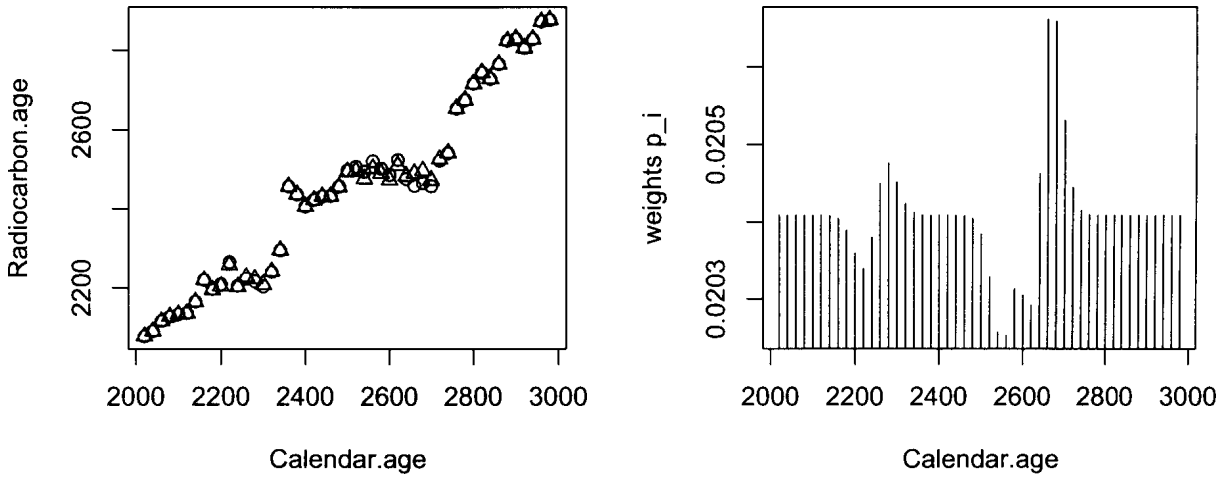

Figure 4.20: Subjecting Radiocarbon data to monotonicity. Left panel: original responses (open circles) and corresponding sharpened responses (triangles). Right panel: corresponding weights $\widehat{p}_i$.

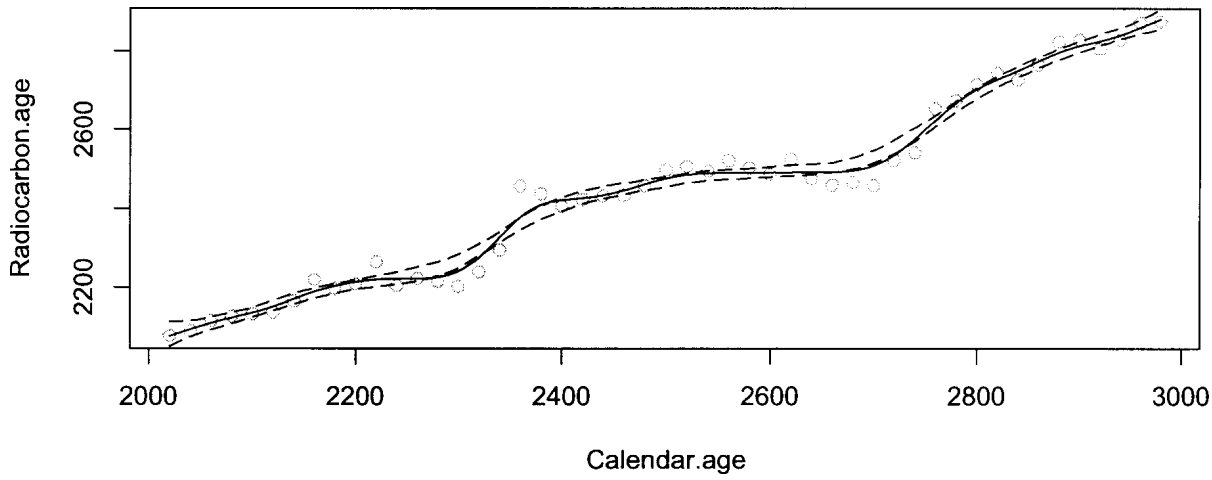

Figure 4.21: Bootstrap confidence bands of constrained sharpening estimate for Radiocarbon data. The constrained sharpening estimate (solid line) and corresponding bootstrap confidence bands(dashed line)

monotonic, the weights are adjusted accordingly: responses which were "too high" just near 2600 are downweighted, and responses around 2700 are given higher weight.

The bootstrap test for the validity of the constraints gives us a p-value of 0.02. This suggests that the constraint of monotonicity may not be appropriate for this data set. The decrease near the calendar age of 2700 might need to be investigated further.

Pointwise 95% confidence bands are shown in Figure 4.21.

4.4.2 Great Barrier Reef Survey Data

These data derive from a survey of fauna on the seabed between the coast of northern Queenland and the Great Barrier Reef. They were analyzed by Bowman and Azzalini (1997) and Hall and Huang (2001). The variables are longitude and score1; the latter variable is a score corresponding to the number of prawns caught at a particular location.

The data are plotted in Figure 4.22 together with the local linear estimate. A normal kernel with bandwidth $h = 0.1$ was used, as in Bowman and Azzalini (1997)

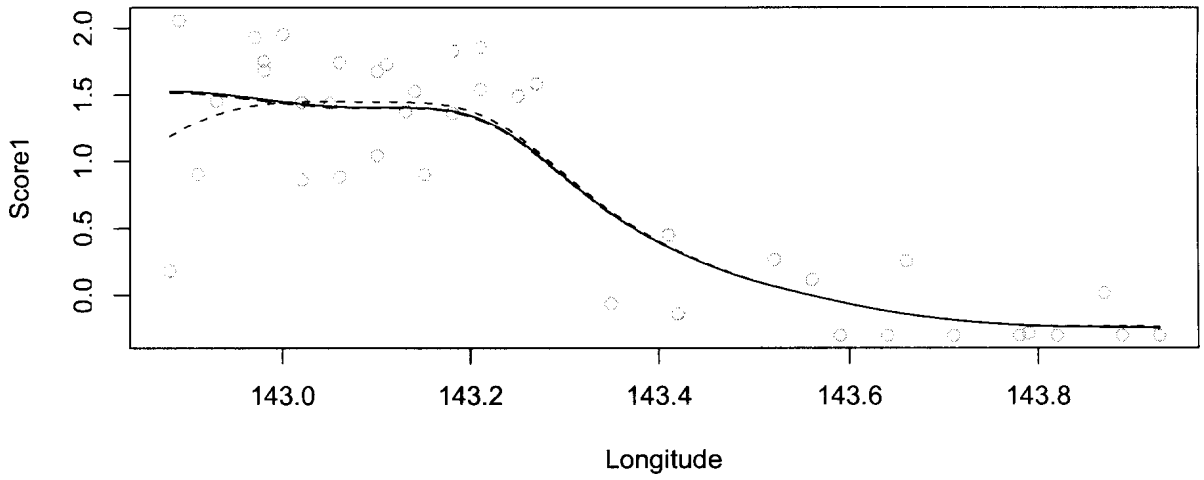

Figure 4.22: Subjecting Great Barrier Reef data to monotonicity. The constrained sharpening local linear estimate (solid line); the unconstrained local linear estimate(dashed line).

and Hall and Huang (2001). Data sharpening subject to the constraint that the regression function is non-increasing was also employed. The HH estimator gave an almost identical result so it is not pictured.

The original data and the corresponding sharpened data are shown in the left panel of Figure 4.23. The weights p_i from the HH approach are shown in the right panel. Where the unconstrained estimate is non-monotonic, i.e., for longitude below 143 , the sharpened data differ from the original and the weights are adjusted accordingly.

The bootstrap test for validity of the constraint gives us a p-value of 0.25. We also conducted this test using the HH method, and obtained a p-value of 0.64. We do not have evidence against the null hypothesis in this case.

The bootstrapped confidence bands of the fitted curve are shown in Figure 4.24.

4.4.3 Canadian High School Graduate Earnings Data

This Canadian cross-section wage data consists of a random sample taken from the 1971 Canadian Census Public Use Tapes for male individuals having 13 years of public

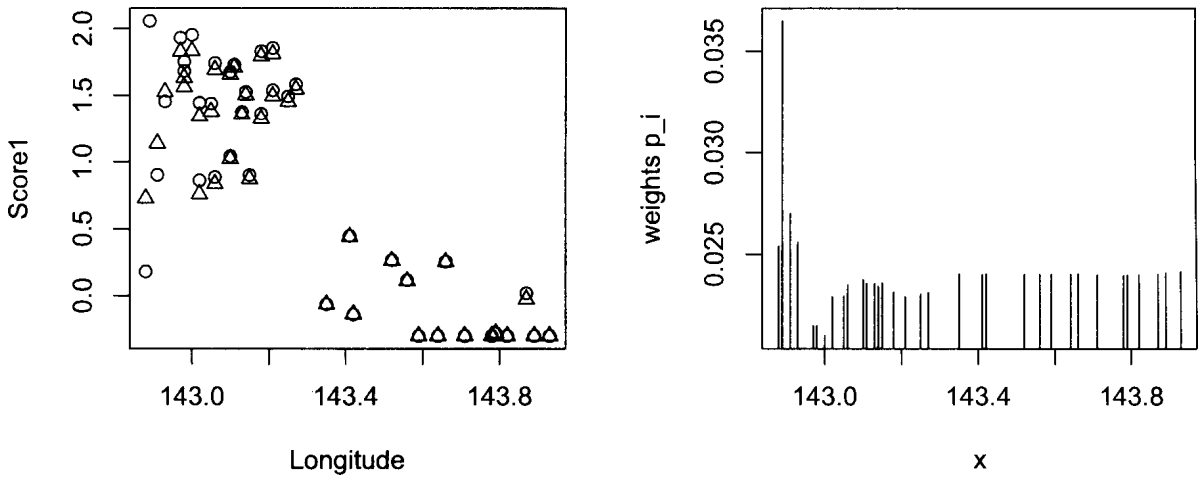

Figure 4.23: Subjecting Great Barrier Reef data to monotonicity. Left panel: original responses (open circles) and corresponding sharpened responses (triangles). Right panel: corresponding weights $\widehat{p}_i$.

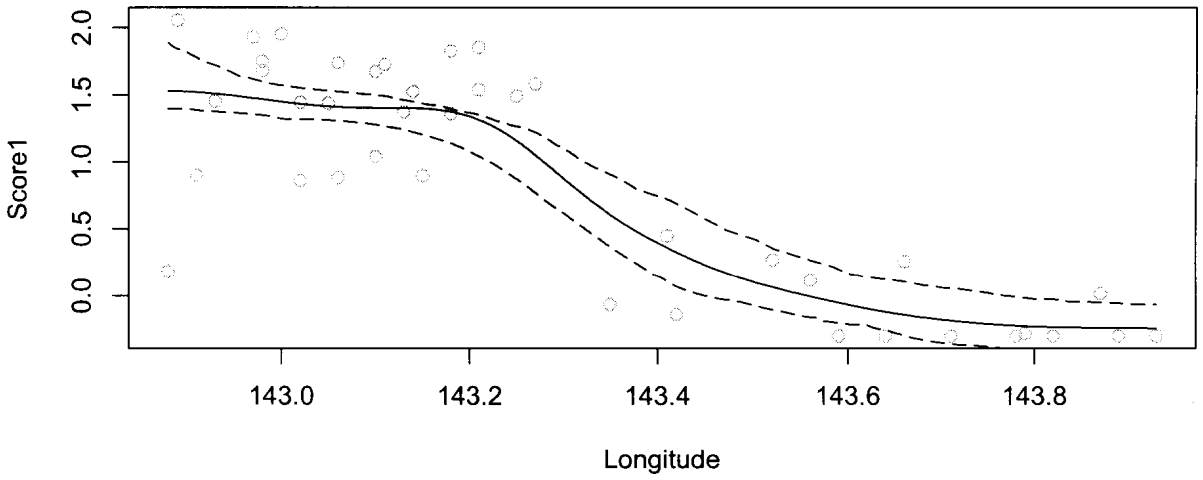

Figure 4.24: Bootstrap confidence bands of constrained sharpening estimate for Great Barrier Reef data. The constrained sharpening estimate (solid line) and corresponding bootstrap confidence bands(dashed line)

education. There are 205 observations in total. There are several studies based on this data set. Most of the early work focused on assuming a quadratic relation between earnings and age, e.g., Heckman and Polachek (1974). Recent work, such as Chu and Marron (1991) and Pagan and Ullah (1999) was based on nonparametric regression techniques.

Given that the human capital theory predicts concavity of the logarithm of a worker's wage as a function of age (potential work experience), Henderson and Parmeter (2009) fitted a concavity-restricted nonparametric age-earning profile using the HH approach.

We also fit a concavity-restricted local linear estimate using the data sharpening approach. A normal kernel was used together with a bandwidth selected using the `dpill` selector.

Figure 4.25 contains a plot of the data, together with the unconstrained and constrained local linear estimates. The HH estimate is indistinguishable from the data sharpened estimate.

As noted by Henderson and Parmeter (2009), the local linear estimates are quite different from what one would expect from a quadratic model, and the concave-restricted estimate does not have the 'dip' near age 40 which is visible in the unconstrained estimate. This is consistent with the core interpretation of the human capital theory.

The sharpened data and the weights used in the HH approach are plotted against the design points in the panels of Figure 4.26. The two concavity-restricted local linear estimate behave similarly.

The bootstrap hypothesis test for the concavity constraint gives a p-value of 0.40. If the HH method is used to compute the p-value, it comes out as 0.55. Thus we have no evidence against the concavity constraint.

The pointwise 95% confidence bands are shown in Figure 4.27.

4.5 Conclusions

In this chapter, we studied the constrained data sharpening technique of Braun and Hall (2001). This procedure provides a method for imposing may be expressed as linear combinations of the responses and where the distance metric is L_2.

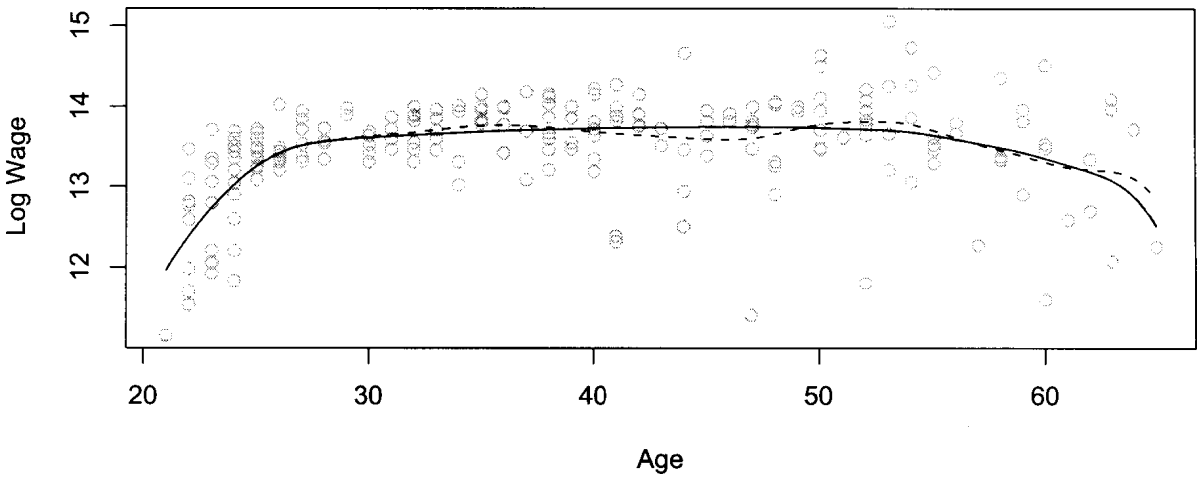

Figure 4.25: Subjecting Canadian High School Graduate Earnings data to a concavity constraint. The constrained sharpening local linear estimate (solid line); the unconstrained local linear estimate (dashed line).

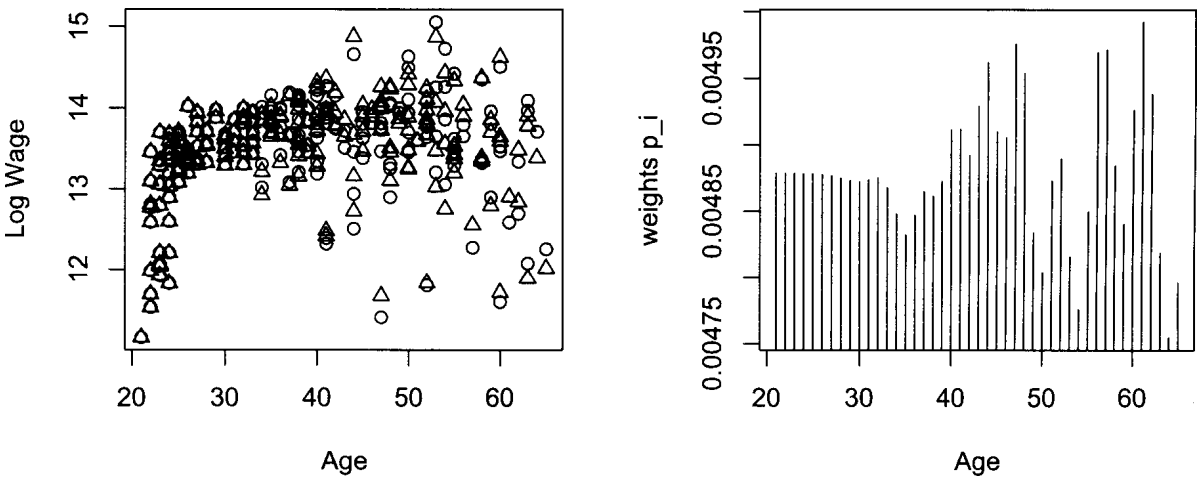

Figure 4.26: Subjecting Canadian High School Graduate Earnings data to a concavity constraint. Left panel: original responses (open circles) and corresponding sharpened responses (triangles). Right panel: corresponding weights $\widehat{p}_i$.

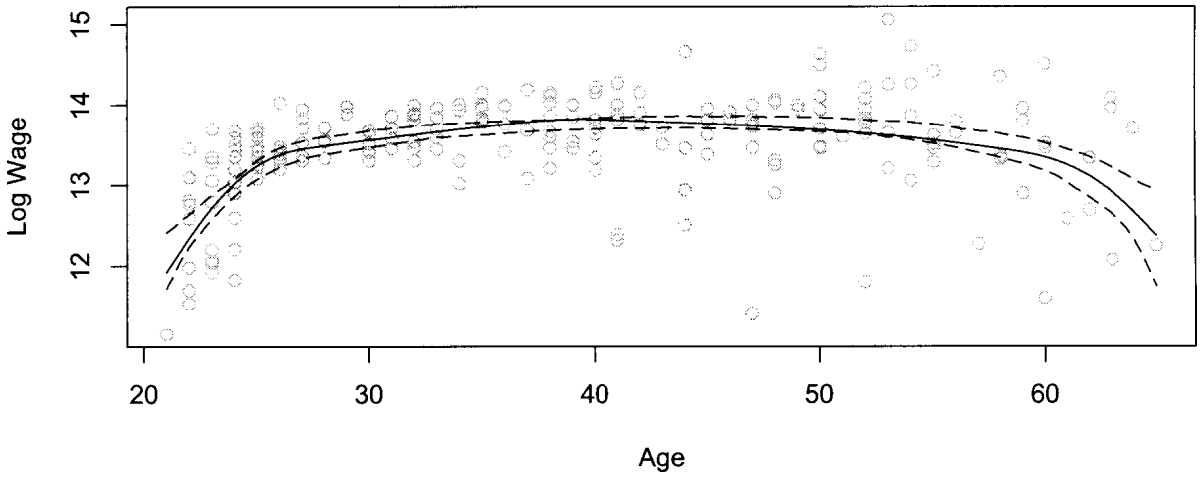

Figure 4.27: Bootstrap confidence bands of constrained sharpening estimate for Canadian High School Graduate Earnings data. The constrained sharpening estimate (solid line) and corresponding pointwise 95% bootstrap confidence bands(dashed line)

We also adapted the constrained weighting method of Hall and Huang (2001) so that it can handle the same kinds of constraints as the data sharpening method when using L_2 metric.

We derived the rate of convergence for data sharpening and showed that it is the same as for constrained weighting, under the monotonicity constraint. We studied the behaviour of the two methods under several other kinds of constraints by simulation. These results show that imposing qualitative constraints on kernel regression estimators can lead to substantial reductions in MISE, even when compared to the bias-reduced method of Choi *et al.* (2000).

Chapter 5

PENALIZED LOCAL REGRESSION

5.1 Introduction

Penalized likelihood methods have been employed in nonparametric regression to obtain improved estimates. For example, penalized splines (Eilers and Marx, 1996) generalize the smoothing spline, leading to greater flexibility, while continuing to reduce reliance on sophisticated knot selection. In this chapter, it is shown that kernel regression can be penalized as well, also leading to greater flexibility, while mitigating the need for adaptive bandwidth selection.

In the preceding chapter, constrained data sharpening was studied with conditions such as $-\varepsilon \leq \hat{g}''(x) \leq \varepsilon$ for all x in a specified set. One disadvantage of setting the constraint in this way is that it imposes a uniform amount of smoothness everywhere. Allowing ε to vary with x_j would provide more flexibility but at the expense of introducing an excessive number of tuning parameters to choose.

An alternative approach would be to constrain the sum of the squares of the second derivative:

$$\sum_{j=1}^{m} (g''(z_j))^2 \leq \varepsilon.$$

The required amount of smoothness is no longer determined in the same manner, and this form of constraint is not as rigid as the one described above. However, this constraint is no longer linear in the $y^\star$'s, so the computational simplicity enjoyed in the previous chapter is lost. We could employ sequential quadratic programming to the methods of the preceding chapter, or we could pursue an alternative approach.

In this chapter, we will pursue the alternative approach: we will propose penalized data sharpening for local regression. The outline of this chapter is as follows. In the next section we will describe data sharpening via a roughness penalty. We will illustrate the method with some preliminary simulation work in the next section. Next, we discuss the effects of penalty grid size. In Section 5.5, the exact bias and variance of the penalized local regression estimator is investigated. Specific examples will be studied with these tools. In Section 5.6, the relationship between penalized

local regression and smoothing splines is explored. In Section 5.7, other penalties besides the roughness penalty are discussed. In Section 8, the penalized local regression approach is illustrated with some examples involving real data.

5.2 Data Sharpening via a Roughness Penalty

Adjoining the constraint to the objective function as a penalty is a computationally simpler alternative, when $p = 2$: choose $y_i^\star$'s to minimize

$$\sum_{i=1}^{n} |y_i - y_i^\star|^2 + \lambda \sum_{j=1}^{m} (\widehat{g}''(z_j))^2,$$

where $z_1, z_2, \cdots, z_m$ are grid points in a given interval, which is usually a subset of the range of the design points. Again, $\widehat{g}(z) = \sum_{i=1} A_i(z) y_i^\star$. Now, $\lambda > 0$ is a tuning parameter. This form is analogous to one of the forms used in the penalized splines introduced by Eilers and Marx (1996). The connection to smoothing splines will be made explicit in the next section.

Let $\mathbf{A}$ denote the $n \times m$ matrix whose (i,j)th element is $A_i(z_j)$, and let $\mathbf{B}$ denote the corresponding matrix of second derivative values. Let $A(z)$ denote the column vector whose ith entry is $A_i(z)$, and let $B(z)$ denote the corresponding vector of second derivatives. Then

$$\widehat{g}(z) = A^{\mathbf{T}}(z)\mathbf{y}^\star$$

and

$$\widehat{g}''(z) = B^{\mathbf{T}}(z)\mathbf{y}^\star$$

The objective function for the data sharpening problem can then be written in vector-matrix notation as

$$(\mathbf{y} - \mathbf{y}^\star)^{\mathbf{T}}(\mathbf{y} - \mathbf{y}^\star) + \lambda \mathbf{y}^{\star\mathbf{T}}\mathbf{B}\mathbf{B}^{\mathbf{T}}\mathbf{y}^\star.$$

Minimizing this yields a unique solution, since $\mathbf{B}\mathbf{B}^{\mathbf{T}}$ is positive semidefinite.

Proposition 5.2.1 *Consider the optimization problem:*

$$\textit{Minimize} \quad (\mathbf{y} - \mathbf{y}^\star)^{\mathbf{T}}(\mathbf{y} - \mathbf{y}^\star) + \lambda \mathbf{y}^{\star\mathbf{T}}\mathbf{B}\mathbf{B}^{\mathbf{T}}\mathbf{y}^\star \tag{5.1}$$

with respect to $\mathbf{y}^\star$, *where* $\lambda > 0$ *and* B *is any* $n \times m$ *matrix.*

This problem has a unique solution.

Proof: A unique solution to the optimization problem exists if and only if the matrix $I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}}$ has an inverse.

Since $\mathbf{B}\mathbf{B}^{\mathbf{T}}$ is clearly positive semidefinite, any eigenvalue μ must be non-negative. Since the eigenvalues of $I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}}$ must be of the form $1 + \lambda\mu$, it follows that they must all be positive. Hence, $I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}}$ is invertible. □

The solution to the minimization problem is given by

$$\mathbf{y}^{\star} = (I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-1}\mathbf{y}. \tag{5.2}$$

Fitted values are given by

$$\widehat{\mathbf{y}} = \mathbf{A}^{\mathbf{T}}(I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-1}\mathbf{y},$$

and the estimated function is

$$\widehat{g}(x) = A(x)^{\mathbf{T}}(I + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-1}\mathbf{y}.$$

We remark that this is one of only a few instances of data sharpening subject to constraints that we are aware of where an explicit formula is available for the sharpened data. Under most circumstances, convex programming or sequential quadratic programming is required to obtain the sharpened data. Thus, we have a nontrivial circumstance here where it is relatively easy to study the properties of the sharpened data.

We have chosen to introduce the penalized data sharpening procedure with a specific type of penalty. However, many other penalties can be used, each of which will lead to an expression of the form (5.2), provided they are quadratic in $\mathbf{y}^{\star}$. For example, the penalty usually associated with smoothing splines could be imposed:

$$\sum_{i=1}^{n} |y_i - y_i^{\star}|^2 + \lambda \int_{x_1}^{x_n} (\widehat{g}''(z))^2 dz. \tag{5.3}$$

Again, (5.2) can be derived but now with the entries of the matrix $\mathbf{B}$ given by

$$B_{i,j} = \int_{x_1}^{x_n} A''(z - x_i)A''(z - x_j)dz, \quad i = 1, 2, \ldots, n, j = 1, 2, \ldots, n.$$

Constraints on the regression function $g(x)$ which take other forms can also be considered, though our main focus will be on the roughness penalty. The following list of operators on g is suggestive of what is possible, but not exhaustive:

- $L_1(x) = g'(x) - kg(x)$, for some constant k (i.e. exponential behaviour)
- $L_2(x) = g''(x) + k^2 g(x)$ (i.e. sinusoidal behaviour)
- $L_3(x) = g(x + k) - g(x)$ (i.e. periodic behaviour)

In all of these situations, we would propose to penalize functionals such as $\int L^2(x)dx$ or $\sum L^2(x_i)$. The data sharpening problems considered in this chapter are always of this form and lead to the optimization problem given at (5.1). Some of these other penalties will be considered briefly in Section 5.7.

This penalty method can be applied to local polynomial regression of any order, and we have done some experimenting with the method in local linear and local quadratic cases without any difficulty. However, we will focus our demonstrations on the local constant case, primarily because of its simplicity, but also because of its utility in situations where data are sparse (e.g. Choi *et al.* (2000)).

5.3 Preliminary Simulation Experiments

To attempt to gain some intuition of the penalized kernel regression method, a set of simulation experiments was conducted. Two sample sizes $n = 25$ and $n = 50$ were considered in order to get a sense of the consistency in the estimators.

The initial objective was to compare the performance of the penalized method with ordinary local regression using the mean integrated squared error (MISE) as the criterion. Since the shape of the estimated regression function may be of as much importance as its accuracy, the MISEs of the derivative of the estimates were also compared.

The target functions under consideration are

1. sine curve: $g(x) = \sin(x)$

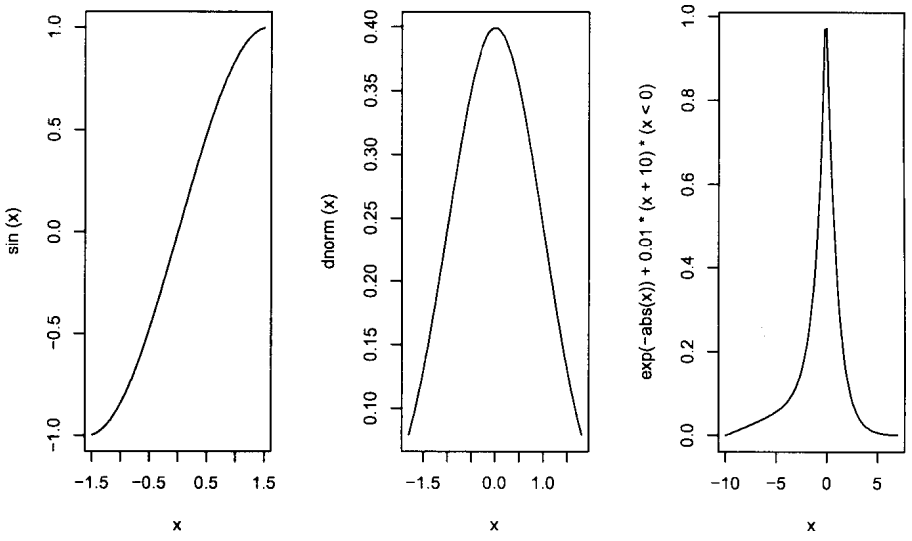

Figure 5.1: The target functions for the simulation study. From left to right: sine curve, standard normal density curve, asymmetric double exponential curve.

2. standard normal density curve: $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

3. asymmetric double exponential curve: $g(x) = e^{-|x|} + .01(x + 10)I_{(x<0)}$

These three functions are pictured in Figure 5.1.

Additive noise with expected value 0 and constant variance was simulated in each case. Error variances were .01, .0001, and .0025, respectively. Sample sizes of 25 and 50 were simulated for each target function. 500 simulation runs were conducted in each case.

Design points were equally spaced. For the first target function, design points were taken on the interval $[-1.5, 1.5]$. For the second target function, points were on the interval $[-2, 2]$ and for the last example, points were taken on the interval $[-10, 7]$.

To compare the penalized method with the unpenalized method without the influence of boundary effects, we studied the target functions on restricted domains. The first function was studied on $[-1.2, 1.2]$. The second function was studied on $[-1.8, 1.8]$, and the third function was studied on $[-9, 6]$. In each case, 401 equally spaced grid points were chosen within these intervals as locations where the target functions were estimated.

For each data set, local constant regression was employed, with and without penalty on the squared sum of second derivatives, using the normal kernel. In each case, a number of different bandwidths were used, and the penalty parameter λ was fixed but not optimized in any way.

5.3.1 Example 1: The Sine Curve

For the sine function, a typical realization is exhibited in the left panel of Figure 5.2 for the case where $n = 25$. Here, the bandwidth was taken to be $h = .2$ for both the penalized and unpenalized local constant regression estimates. The constraint on the second derivative was imposed by penalizing

$$\sum_{i=1}^{401} \left(\widehat{g}''(z_i)\right)^2$$

where $z_1, z_2, \ldots, z_{401}$ were taken to be equally spaced gridpoints in the range of the x values. The penalty parameter λ was set at .001. The right panel of 5.2 displays the corresponding derivatives of the estimates.

For the larger-scale simulation, the bandwidths considered were in the set $\{.1, .15, \ldots, .6\}$ and the penalty parameter λ was set to the value .001.

The MISE, as a function of bandwidth, is plotted in the left panel of Figure 5.3, for the case where $n = 25$. The right panel of the figure displays the MISE for the derivative of the estimate as a function of h. Figure 5.4 gives the analogous plots for the case where $n = 50$.

5.3.2 Example 2: The Standard Normal Density Curve

For the standard normal density function, a typical realization is exhibited in the left panel of Figure 5.5 for the case where $n = 25$. Here, the bandwidth was taken to be $h = .1$ for both the penalized and unpenalized local constant regression estimates. The constraint on the second derivative was imposed by penalizing

$$\sum_{i=1}^{401} \left(\widehat{g}''(z_i)\right)^2$$

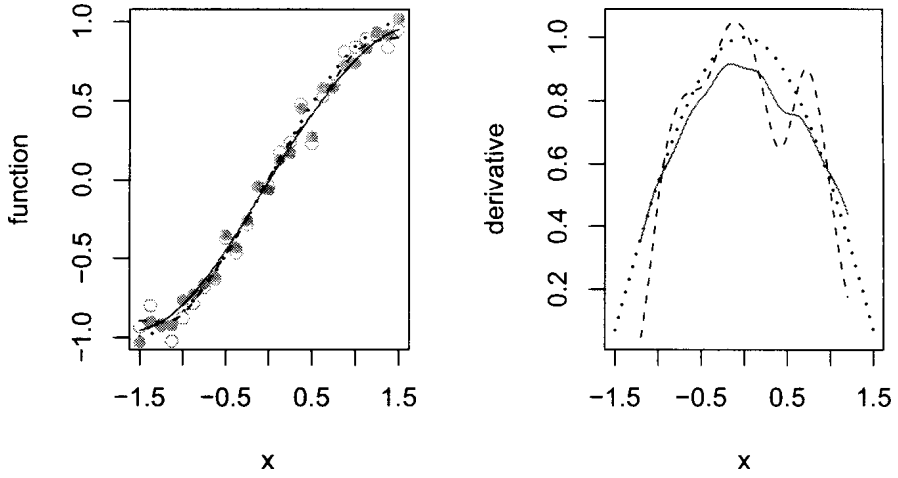

Figure 5.2: Left panel: sine function (dotted curve) estimates using penalized (solid curve) and unpenalized (dashed curve) local constant regression, based on a sample of size 25. The solid dots represent locations of the sharpened responses. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

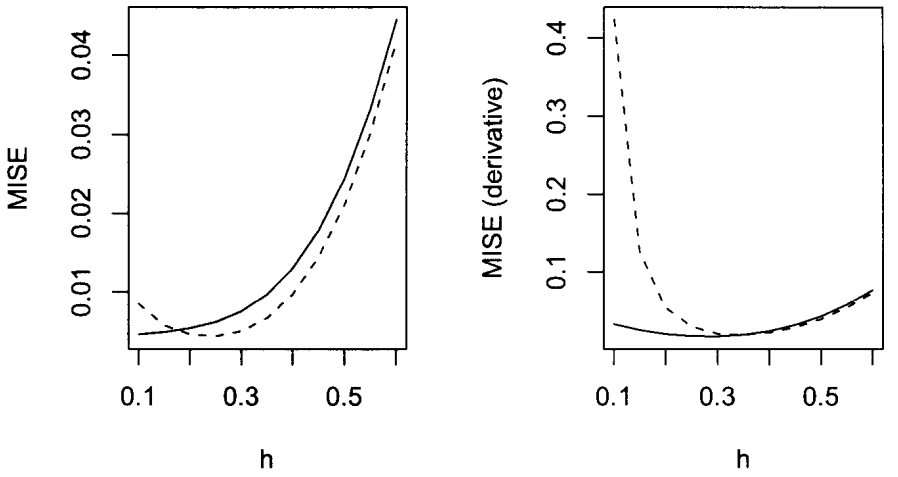

Figure 5.3: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of h, based on samples of size 25.

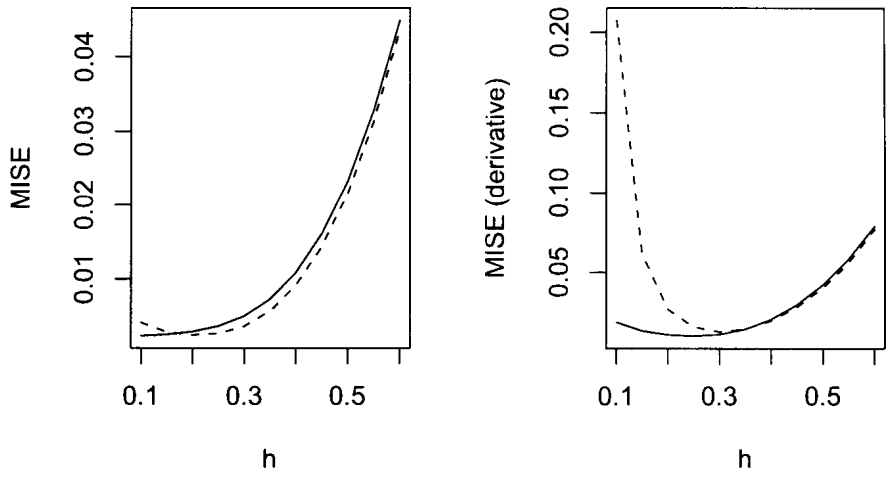

Figure 5.4: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of h, based on samples of size 50.

where $z_1, z_2, \ldots, z_{401}$ were taken to be equally spaced gridpoints in the range of the x values. The penalty parameter λ was set at .001. Again, the corresponding derivative estimates are supplied in the right panel of Figure 5.5.

For the larger-scale simulation, the bandwidths considered were in the set $\{.05, .07, \ldots, .25\}$ and the penalty parameter λ was set to the value .001.

The MISE, as a function of bandwidth, is plotted in the left panel of Figure 5.6, for the case where $n = 25$. The right panel of the figure displays the MISE for the derivative of the estimate as a function of h. Figure 5.7 gives the analogous plots for the case where $n = 50$.

5.3.3 Example 3: The Asymmetric Double Exponential Curve

For the asymmetric double exponential function, a typical realization is exhibited in the left panel of Figure 5.8 for the case where $n = 25$. Here, the bandwidth was taken to be $h = .25$ for both the penalized and unpenalized local constant regression

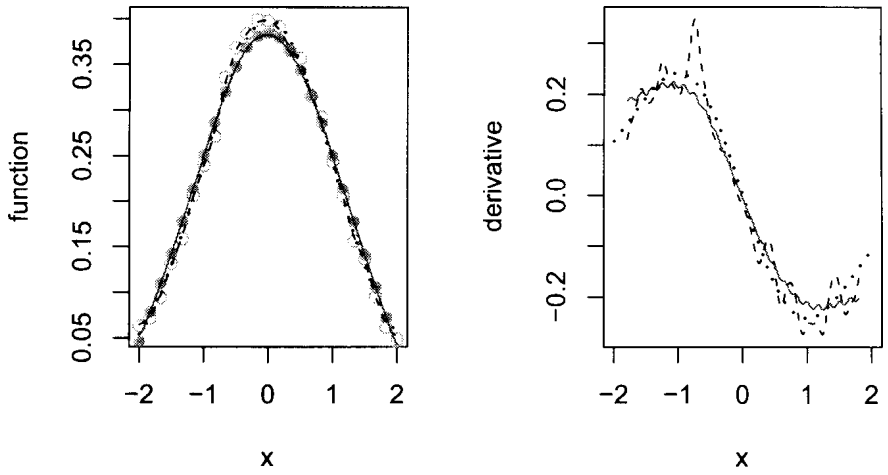

Figure 5.5: Left panel: standard normal density function (dotted curve) estimates using penalized (solid curve) and unpenalized (dashed curve) local constant regression, based on a sample of size 25. The solid dots represent locations of the sharpened responses. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

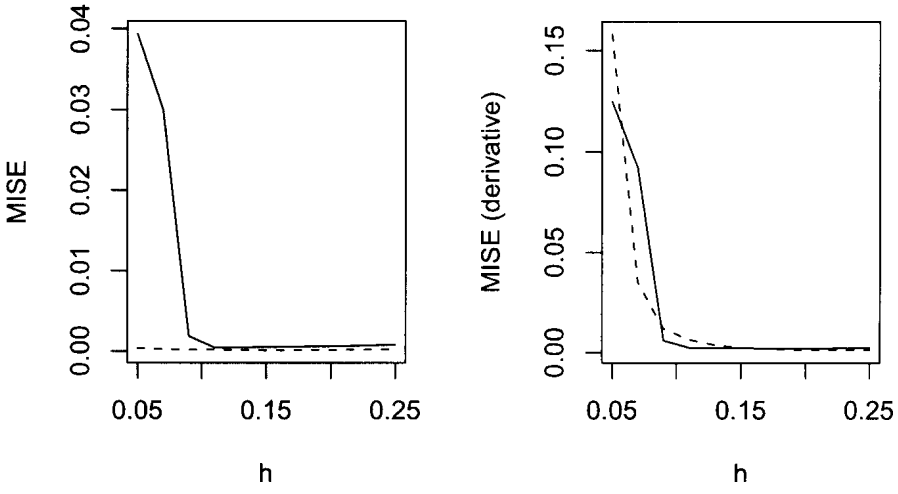

Figure 5.6: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the standard normal density function for various values of h, based on samples of size 25.

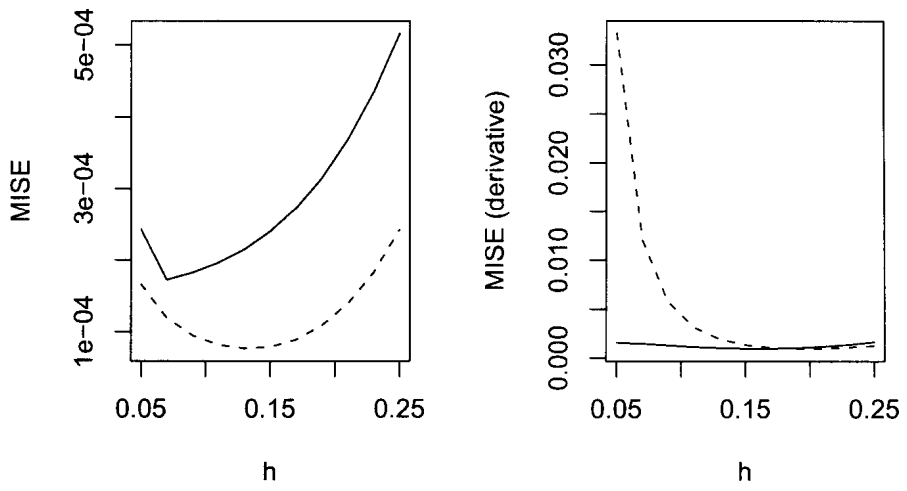

Figure 5.7: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the standard normal density function for various values of h, based on samples of size 50.

estimates. The constraint on the second derivative was imposed by penalizing

$$\sum_{i=1, i \notin \{200,201,\ldots,294\}}^{401} \left(\widehat{g}''(z_i)\right)^2$$

where $z_1, z_2, \ldots, z_{401}$ were taken to be equally spaced gridpoints in the range of the x values. The value of the penalty parameter was $\lambda = .1$. Note that in this case, the smoothness requirement was relaxed at the location near the peak of the curve. Again, the corresponding derivative estimates are supplied in the right panel of Figure 5.8.

For the larger-scale simulation, the bandwidths considered were in the set $\{.1, .15, \ldots, .6\}$ and the penalty parameter λ was set to the value .1.

The MISE, as a function of bandwidth, is plotted in the left panel of Figure 5.9, for the case where $n = 25$. The right panel of the figure displays the MISE for the derivative of the estimate as a function of h. Figure 5.10 gives the analogous plots for the case where $n = 50$.

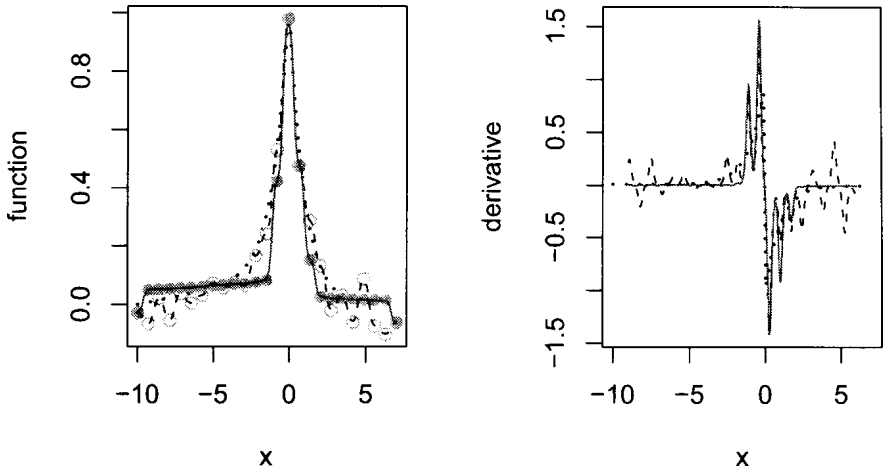

Figure 5.8: Left panel: double exponential function (dotted curve) estimates using penalized (solid curve) and unpenalized (dashed curve) local constant regression, based on samples of size 25. The solid dots represent locations of the sharpened responses. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

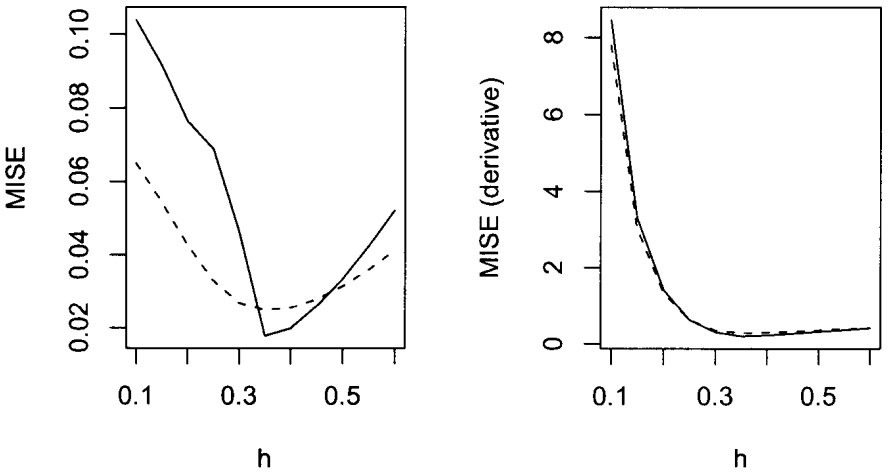

Figure 5.9: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the asymmetric double exponential function for various values of h, based on samples of size 25.

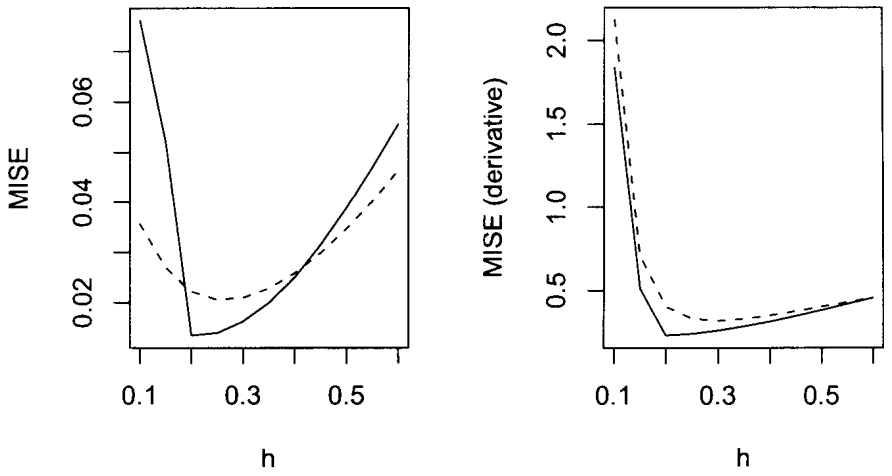

Figure 5.10: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the asymmetric double exponential function for various values of h, based on samples of size 50.

5.3.4 Observations

Based on this modest simulation study, it can be seen that the penalty method is not guaranteed to give better accuracy in the MISE sense than the unpenalized method. Choice of bandwidth is critical as usual. Choice of penalty parameter is also important.

The chosen penalty parameter values appear to have led to some over-smoothing. Evident in Figures 5.2 and 5.5 is a fairly substantial departure of the penalized curve from the true curve (and from the data themselves) near the boundaries of the domain of observation. This less satisfactory performance is reflected in the MISE values. As will be demonstrated in the next section, this difficulty is connected not only to the choice of penalty parameter but also to the choice of gridsize used in constructing the penalty function.

On the other hand, the MISE values for the derivatives of the penalized estimates can often be much smaller than for the corresponding unpenalized estimates, and in cases where they are larger, the difference is not substantial. Thus, applying the roughness penalty appears to improve accuracy in the estimation of the shape of the regression function, at a given bandwidth.

5.4 Effect of Penalty Gridsize

First, consider penalties on functions of the form $(Lg)(z_0)$, where L might be a differential operator or linear combination of differential operators. For example, $Lg = g''$. In other words, we are only constraining the function to behave in a particular way at the single point z_0; the penalty gridsize is 1.

In this case, the matrix $\mathbf{B}$ induced by the optimization problem at (5.1) reduces to a vector so the sharpened $\mathbf{y}^\star$ can be written as

$$\begin{aligned} \mathbf{y}^\star &= \left(\mathbf{I} - \frac{\lambda \mathbf{B}\mathbf{B}^\mathbf{T}}{1 + \lambda \mathbf{B}^\mathbf{T}\mathbf{B}}\right)\mathbf{y} \\ &= \mathbf{y} - \frac{\lambda \mathbf{B}\mathbf{B}^\mathbf{T}\mathbf{y}}{1 + \lambda \mathbf{B}^\mathbf{T}\mathbf{B}}. \end{aligned}$$

For large λ, the constraint will be approximately satisfied by the sharpened vector $\mathbf{y}^\star$. That is,

$$\lim_{\lambda \to \infty} \mathbf{B}^\mathbf{T}\mathbf{y}^\star = \lim_{\lambda \to \infty} \mathbf{B}^\mathbf{T}\mathbf{y} - \mathbf{B}^\mathbf{T}\frac{\lambda \mathbf{B}\mathbf{B}^\mathbf{T}\mathbf{y}}{1 + \lambda \mathbf{B}^\mathbf{T}B} = 0.$$

Thus, for large values of λ, the penalty is effectively imposing a linear restriction on $\mathbf{y}^\star$, resulting in a reduction in degrees of freedom by 1.

If the kernel function employed in the regression estimator has compact support, it is possible for multiple constraints to act independently, provided the gridpoints used to construct the penalty function are separated sufficiently. In other words, consider a penalty function of the form

$$\sum_{k=1}^{m} ((Lg)(z_k))^2$$

where $m < n$ and where the minimum distance between any pair of gridpoints z_k is greater than $2h$ and the support of the kernel function is assumed to be $[-h,\ h]$. The gridpoints are, of course, assumed to lie in the domain of the design points. In this case, the behaviour of $(Lg)(z_k)$ will be independent of $(Lg)(z_j)$ whenever $j \neq k$. Thus, the matrix $\mathbf{B}$ induced by the optimization problem at (5.1) will have

independent columns. For large λ, this results in an imposition of approximately m linear restrictions on $\mathbf{y}^\star$. Equivalently, $n - m$ eigenvalues of $\mathbf{BB^T}$ will be 0.

If the constraints are not sufficiently separated, or if a non-compact kernel is used, then the constraints will no longer act independently, and it is difficult to predict how many eigenvalues, if any, of $\mathbf{BB^T}$ will be 0. Because a function of the form $\sum_{i=1} A_i(x)y_i^\star$ might not have the flexibility required to exactly satisfy particular kinds of constraints (defined continuously in x; for example, $g''(x) = 0$, for all x in an interval), inconsistencies may arise.

This is likely what has led to some of the numerical problems with the original data sharpening method described in the previous chapter. It also leads to a situation where $\mathbf{BB^T}$ has no 0 eigenvalues. In this case, the only solution to the penalized optimization problem (for large λ) is $\mathbf{y}^\star = 0$.

However, a number of the eigenvalues of $\mathbf{BB^T}$ may be close to 0, if the functional form $\sum_{i=1} A_i(x)y_i^\star$ can provide a sufficiently good approximation to a function lying in the constrained region. These eigenvalues would be exactly 0, if the constraint could be exactly satisfied. The corresponding eigenvectors must then, at least approximately, satisfy the constraints.

There are at least three ways to resolve the difficulties described here. The first is to reduce the number of grid points used to develop the penalty function. In many situations, it will be sufficient to take the number of grid points at or just below the number of observed data points. This method will be the main focus of the rest of this chapter.

If a refined grid is required, then the above method will not be appropriate. This leads to a variant of the penalty method which would be applicable if the grid size is large and if large λ is desirable (i.e. there is a strong belief that the constraints are accurate, for example):

1. Calculate eigenvalues and corresponding eigenvectors of $\mathbf{BB^T}$.
2. Identify all "small" eigenvalues and their corresponding eigenvectors.
3. Set $\mathbf{y}^\star$ to be the linear combination of the selected eigenvectors which is closest to $\mathbf{y}$.

Remarks: Identifying the "small" eigenvalues in step 2 is subjective. Using the L_2 distance in step 3 leads to a linear regression problem: regress $\mathbf{y}$ on the selected eigenvectors. This technique has been applied to some simulated data sets with

reasonable success, but a rigourous examination of this method lies beyond the scope of this thesis.

Another way of handling this situation is to use higher order local polynomial regression. On the basis of some experiments we have conducted with local constant and local quadratic regression, we can assert that the improved accuracy in the local quadratic estimates of the second derivative allow for a substantial increase in the number of grid points where the penalty function can be imposed.

5.5 Bias and Variance

The expected value of $\mathbf{y}^\star$ is given by

$$E[\mathbf{y}^\star] = (\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^\mathbf{T})^{-1} E[\mathbf{y}] = (\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^\mathbf{T})^{-1} \mathbf{g}$$

where

$$\mathbf{g} = [g(x_1)\ g(x_2)\ \cdots\ g(x_n)]^\mathbf{T}.$$

Recall that B and λ do not depend on the responses. We can then find that the exact bias in the estimate of the regression function at a point x is

$$\text{Bias}(\widehat{g}(x)) = A^\mathbf{T}(x)(\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^\mathbf{T})^{-1}\mathbf{g} - g(x). \tag{5.4}$$

The variance can be calculated using

$$\text{Var}(\widehat{g}(x)) = \sigma^2 A^\mathbf{T}(x)(\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^\mathbf{T})^{-2} A(x). \tag{5.5}$$

These forms allow us to study the sharpening procedure without simulation. The MISE can be obtained by numerical integration of

$$\text{Var}(\widehat{g}(x)) + \text{Bias}^2(\widehat{g}(x))$$

over the domain containing the design points.

It is also possible to also study the properties of derivatives of the estimates as well. For example,

$$\widehat{g}'(x) = \frac{d}{dx} A^{\mathbf{T}}(x)(\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-1}\mathbf{y}$$

so

$$E[\widehat{g}'(x)] = \frac{d}{dx} A^{\mathbf{T}}(x)(\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-1}\mathbf{g}$$

and

$$\mathrm{Var}(\widehat{g}'(x)) = \sigma^2 \frac{d}{dx} A^{\mathbf{T}}(x)(\mathbf{I} + \lambda \mathbf{B}\mathbf{B}^{\mathbf{T}})^{-2} \frac{d}{dx} A(x).$$

Again, we can compute a numerical approximation to the MISE for the derivative of the estimated regression function using these expressions.

5.5.1 Revisiting the Simulation Experiments

Using the formulas obtained above, we can replace the simulation experiments conducted in Section 5.3 with exact calculations. When we do this, we find that the theory matches the simulation results almost perfectly. The exact results allow us to make more rapid progress in the study of particular situations.

We consider only the cases where $n = 25$ here, since the relative effect of the doubling of the sample size has already been seen.

5.5.1.1 The Sine Curve

For the first target function, we fixed the bandwidth at $h = .1$, because it appears to give good MISE performance, according to Figure 5.3. We computed approximate MISE values for the regression function estimate and its derivative using penalty parameter values in the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 12\}$.

Figure 5.11 displays the results. The MISE-optimal value of λ can be seen (on the left-hand panel) to be near .01. It can also be seen that the penalty function approach leads to a substantial reduction in MISE compared with the unpenalized local constant regression method. Derivative accuracy for the penalized estimator is also superior at this bandwidth.

However, this comparison is not quite fair. The bandwidth $h = .1$ is not optimal for the unpenalized estimator. If we use $h = .25$ instead, we find that the MISE for the unpenalized estimator is .00437 which is considerably closer to the MISE obtained for the penalized estimator using $h = .1$ and $\lambda = .01$: .00415. A bigger problem is

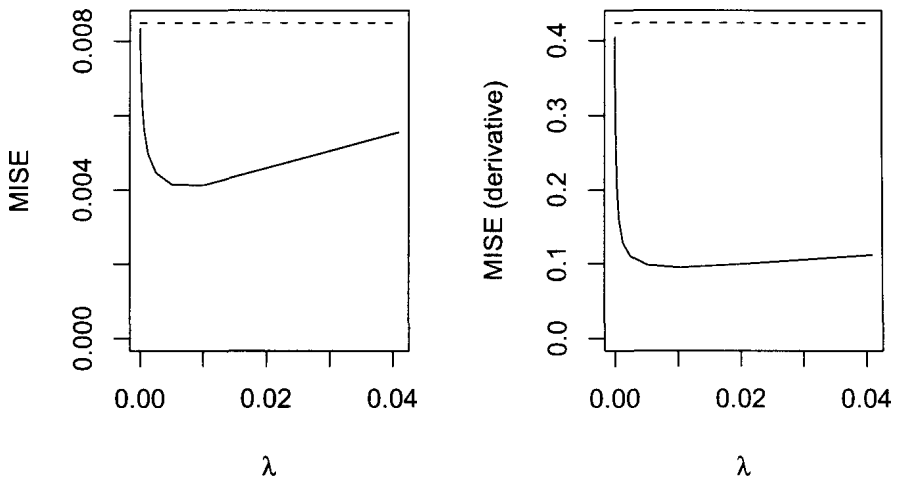

Figure 5.11: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of λ, based on samples of size 25, using $h = .1$.

the large size of the MISE of the derivative when using $h = .1$ and $\lambda = .01$ relative to the corresponding value when $h = .25$ and $\lambda = 0$: .1382 versus .0294. Thus, using the penalty method to improve accuracy in the estimated regression function can incur a loss of accuracy in the shape of the function estimate.

Figure 5.12 shows the regression estimate for a typical set of simulated data. Here, the smoothing parameters were taken as $\lambda = .01$ and $h = .1$. The derivative of the estimate is also displayed in the figure.

In order to test our theory (given in Section 5.4) regarding the effects of a large number of grid points in the penalty computation, we have computed the penalty in two ways: with 401 equally-spaced grid points in the interval $[-1.2, 1.2]$, and with 15 equally-spaced grid points in that same interval. The figure 5.12 shows a clear difference in the estimates, particularly in terms of the derivative; the larger number of grid points has resulted in a very over-smoothed curve.

Note that the 15 equally-spaced grid points were used in the computations that Figure 5.11 is based on.

We also computed the estimate using the bandwidth which is optimal for the unpenalized method but using a penalty parameter which optimizes the derivative of the estimate. We found that if $h = .25$ and $\lambda = .02048$, the MISE becomes .00467

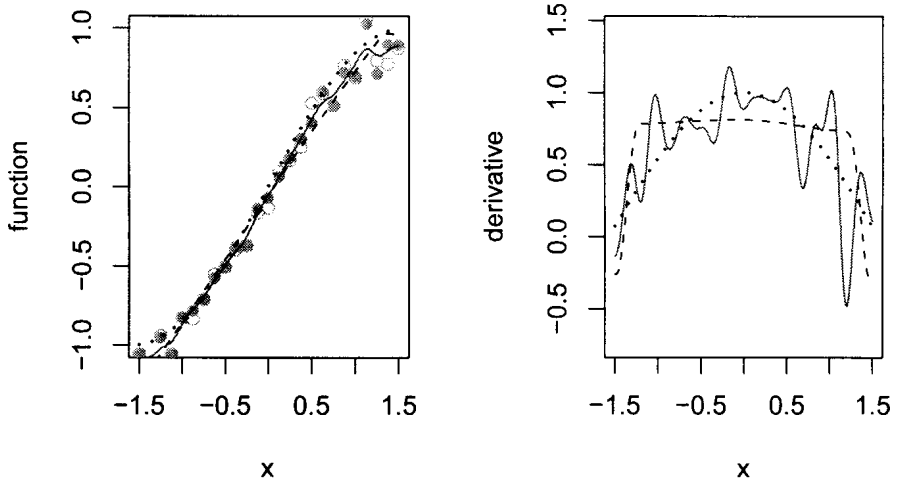

Figure 5.12: Left panel: sine function (dotted curve) estimates using penalized local constant regression with 401 grid points (dashed curve) and 15 grid points (solid curve) for the penalty. The solid dots represent locations of the sharpened responses. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

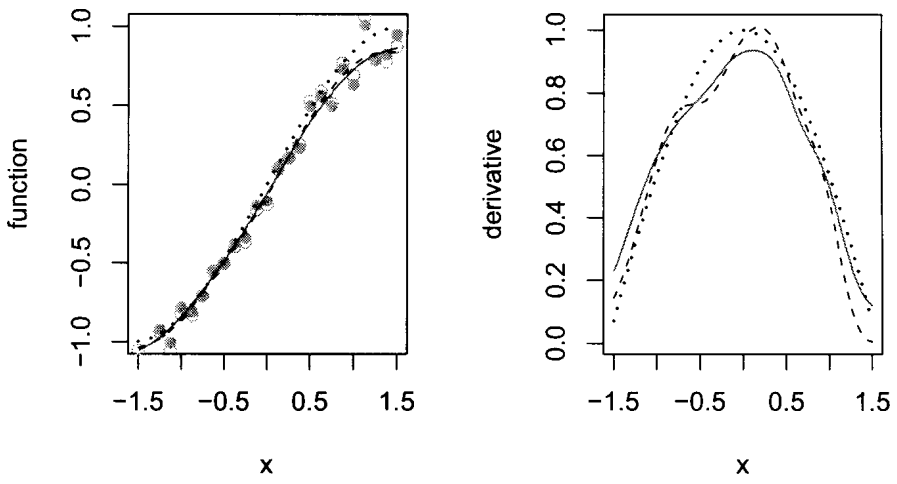

Figure 5.13: Left panel: sine function (dotted curve) estimates using unpenalized local constant regression (dashed curve) and penalized local constant regression with 15 grid points (solid curve) for the penalty. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

(which is slightly larger than the MISE for the unpenalized estimator), but the MISE of the derivative of the penalized estimator drops to .0156 (which is substantially lower than the MISE of the unpenalized estimator). Thus, in this case, we can obtain a substantially better shape at a slight expense in accuracy. The result is pictured in Figure 5.13.

5.5.1.2 The Normal Density Curve

For the second target function, we fixed the bandwidth at $h = .15$, because it appears to give good MISE performance, according to Figure 5.6. We computed approximate MISE values for the regression function estimate and its derivative using penalty parameter values in the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 10\}$. For the penalty function, 15 equally-spaced grid points were used in the interval $[-1.8, 1.8]$.

Figure 5.14 displays the results. The MISE-optimal value of λ can be seen (on the left-hand panel) to be near .001 (actually, .00128). It can also be seen that the penalty function approach leads to a reduction in MISE compared with the unpenalized local constant regression method. For the penalized method, the MISE is

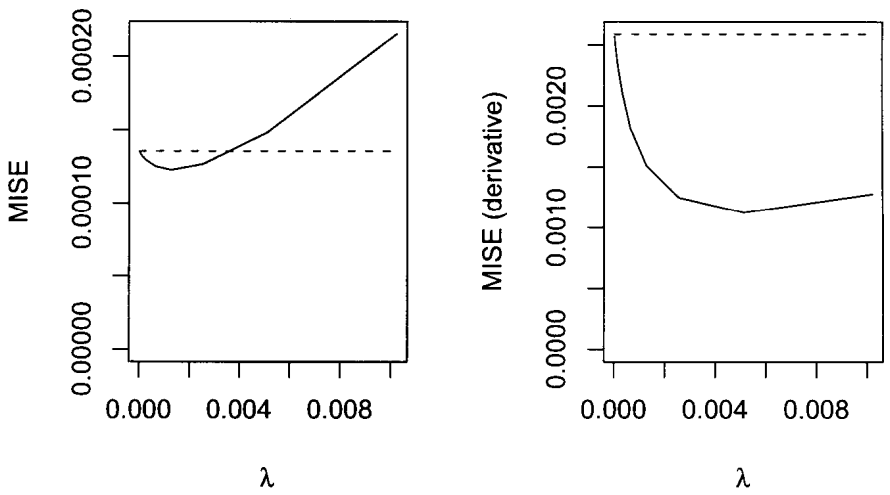

Figure 5.14: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the normal density function for various values of λ using $h = .15$.

.000123 compared with an MISE of .000135 for the unpenalized method.

Derivative accuracy for the penalized estimator is also improved at this bandwidth. For the penalized method, the MISE is .00112 compared with an MISE of .00127 for the unpenalized method.

Figure 5.15 shows the regression estimate for a typical set of simulated data. Here, the smoothing parameters were taken as $\lambda = .002$ and $h = .15$. The derivative of the estimate is also displayed in the figure. As another test of our theory (given in Section 5.4) regarding the effects of a large number of grid points in the penalty computation, we have computed the penalty in two ways: with 401 equally-spaced grid points in the interval $[-1.8, 1.8]$, and with 15 equally-spaced grid points in that same interval. Again, the figure shows a clear difference in the estimates, particularly in terms of the derivative; the larger number of grid points has resulted in a very over-smoothed curve.

5.5.1.3 The Asymmetric Double Exponential Curve

For the third target function, we fixed the bandwidth at $h = .35$, because it appears to give good MISE performance, according to Figure 5.9. We computed approximate

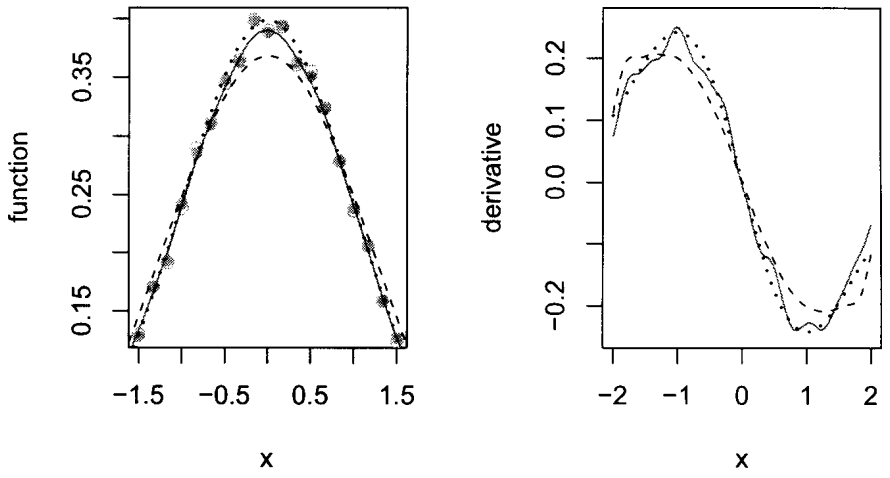

Figure 5.15: Left panel: normal density function (dotted curve) estimates using penalized local constant regression with 401 grid points (dashed curve) and 15 grid points (solid curve) for the penalty. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

MISE values for the regression function estimate and its derivative using penalty parameter values in the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 19\}$.

For the penalty function, 20 grid points were used, equally-spaced in the union of the intervals $[-9, -1.5]$ and $[2, 6]$. This (subjective) choice was made so that the obviously sharp peak would not be oversmoothed.

Figure 5.16 displays the results. The MISE-optimal value of λ can be seen (on the left-hand panel) to be near 3, though there does not appear to be a lot of sensitivity to this choice. It can also be clearly seen that the penalty function approach leads to a reduction in MISE compared with the unpenalized local constant regression method.

Derivative accuracy for the penalized estimator is also improved. For the penalized method, the MISE is .211 compared with an MISE of .283 for the unpenalized method.

Figure 5.17 shows the regression estimate for a typical set of simulated data. Here, the smoothing parameters were taken as $\lambda = 3$ and $h = .35$. The derivative of the estimate is also displayed in the figure. As a final test of our theory (given in Section 5.4) regarding the effects of a large number of grid points in the penalty

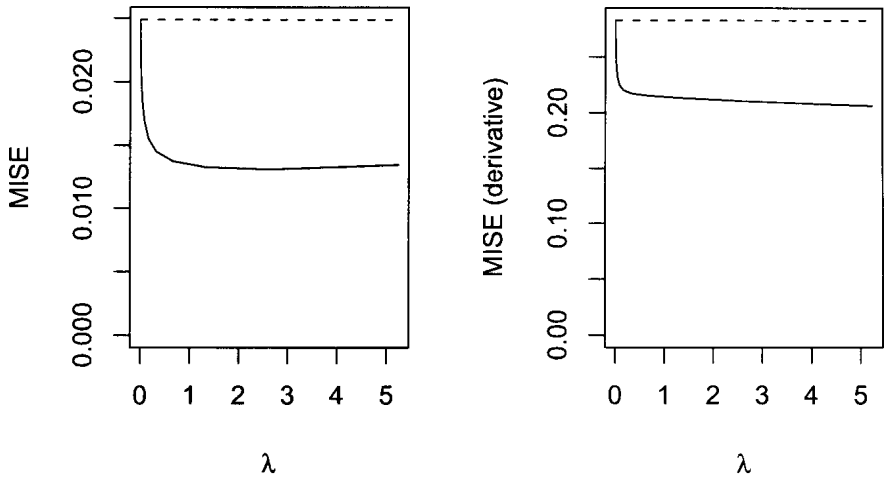

Figure 5.16: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the asymmetric double exponential function for various values of λ, using $h = .35$.

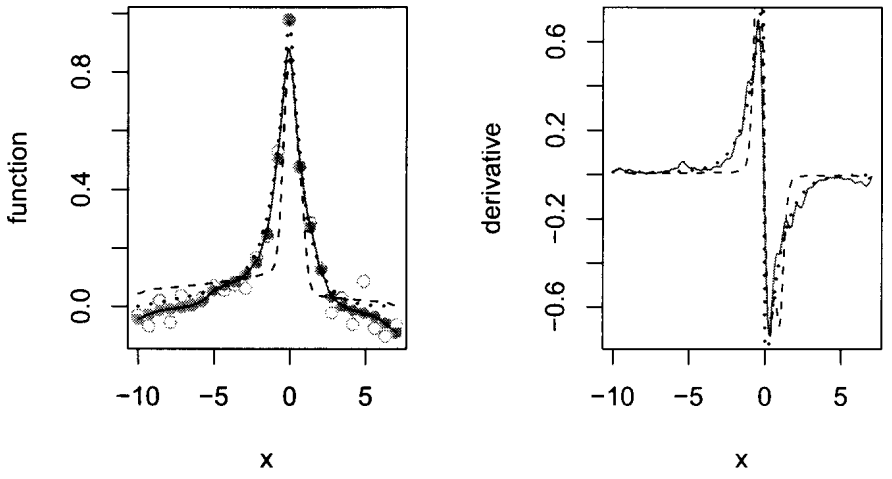

Figure 5.17: Left panel: asymmetric double exponential function (dotted curve) estimates using penalized local constant regression with 307 grid points (dashed curve) and 20 grid points (solid curve) for the penalty. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

computation, we have computed the penalty in two ways: with 307 grid points and with 20 grid points as described above. Again, the figure shows a clear difference in the estimates. Care needs to be taken when choosing the number of grid points for the penalty function.

5.6 Relationship with Smoothing Splines

The form of the relationship between $\mathbf{y}$ and $\mathbf{y}^\star$ expressed in (5.2) is strongly reminiscent of the relation between $\mathbf{y}$ and the associated smoothing spline $\widehat{g}_{ss}(x)$ (e.g. Chapter 2, Green and Silverman, 1994). Thus, we seek to understand the connection between $\mathbf{y}^\star$ and $\widehat{g}_{ss}(x)$.

To do this, let us return to the original smoothing spline problem which is to find the function $\widehat{s}(x)$, having continuous second derivative, which minimizes

$$\sum_{i=1}^{n} |y_i - s(x_i)|^2 + \lambda \int_{x_1}^{x_n} (s''(x))^2 dx.$$

As is pointed out by Green and Silverman (1994), the minimizer is a natural cubic spline whose knots are at the locations of the x's (assumed to be sorted).

We can view the smoothing spline form of penalized data sharpening given at (5.3) as choosing a function $h(x)$ to minimize

$$\sum_{i=1}^{n} |y_i - h(x_i)|^2 + \lambda \int_{x_1}^{x_n} (\widehat{g}''(x))^2 dx. \tag{5.6}$$

where $\widehat{g}(x) = \sum_{j=1}^{n} A_j(x) h(x_j)$. Here, we have associated the sharpened response $y_i^\star$ with the function value $h(x_i)$.

Now, let $\{\alpha_j(x)\}$ denote any sequence of n twice-continuously differentiable functions for which $\alpha_j(x_i) = A_j(x_i)$, for all i and j.

The minimal property of interpolating splines (de Boor (1978), Chapter XIV) tells us that if $\widehat{g}(x) = \sum_{j=1}^{n} \alpha_j(x_i) h(x_j)$, for any sequence $h(x_1), \ldots, h(x_n)$, then

$$\int_{x_1}^{x_n} (\widehat{g}''(x))^2 dx \geq \int_{x_1}^{x_n} (s''(x))^2$$

where $s(x)$ is a natural cubic spline with knots at $x_1, x_2, \ldots, x_n$ and where

$$s(x_i) = \sum_{j=1}^{n} A_j(x_i)h(x_j).$$

The optimal sequence of $h(x_i)$'s could be computed from the appropriate analogue of (5.2).

However, $\widehat{g}(x)$ is constrained to be $\sum_{j=1}^{n} A_j(x)h(x_j)$ where the functions $A_j(x)$ are induced by the kernel. Thus, the minimizer will not necessarily be the same; in fact, in most instances it will be different, since there are few circumstances where $\sum_{j=1}^{n} A_j(x)h(x_j)$ coincides with a cubic spline.

Thus, the estimator developed from penalized data sharpening bears a resemblance to a smoothing spline, but it is not necessarily equivalent.

5.7 Other Penalties

In this section, we give an indication as to how the penalty method might be useful in other situations where qualitative information about the regression function is available. Examples are given in which the regression function satisfies a differential equation of the form $g'' = -g$ and where the regression function is known to be periodic.

5.7.1 Penalizing $\sum(g''(x) + g(x))^2$

The first target function is well known to satisfy the differential equation

$$g'' = -g.$$

Often, an initial condition would be known as well, and this could easily be handled by our methodology, but we will suppose it is unknown in our current treatment.

Again, we will suppose we have a sample of size 25 taken at equally spaced design points on $[-1.5, 1.5]$. Instead of a smoothness penalty, we will now use the

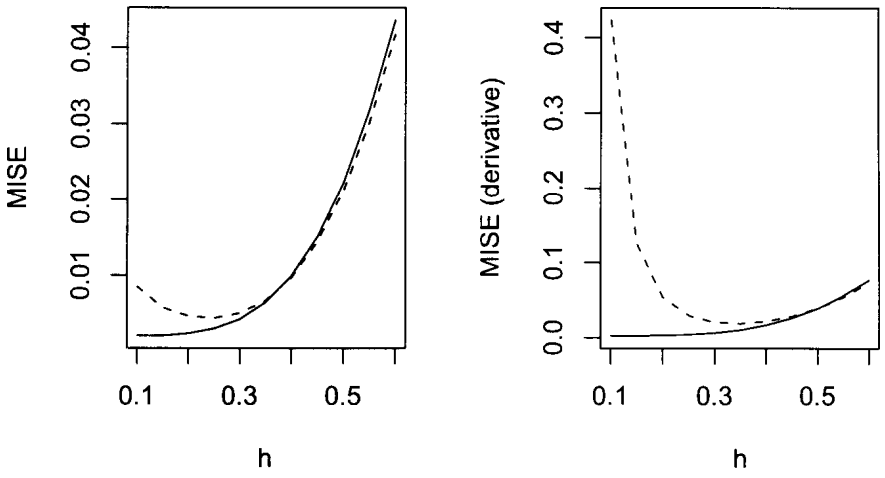

Figure 5.18: MISE values for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of h, based on samples of size 25.

differential equation to construct the penalty by choosing 30 equally spaced grid points on the interval $[-1.2, 1.2]$. That is, we penalize

$$\sum_{i=1}^{30} (\widehat{g}''(z_i) + \widehat{g}(z_i))^2 \tag{5.7}$$

where $z_1, \ldots, z_{30}$ represent the grid points.

Figure 5.18 shows MISE values for the penalized local constant regression estimator and the unpenalized local constant regression estimator for bandwidths $h \in \{.1, \ldots, .6\}$ and where $\lambda = .3$. From this plot, we see that $h = .1$ should give reasonable performance.

Figure 5.19 shows MISE values for penalized local regression with penalty parameter values taken from the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 16\}$ and where $h = .1$. From this plot, it appears that there is rapid improvement in MISE as λ increases away from 0, but that when $\lambda > .1$, there is not much further sensitivity to choice of λ.

Figure 5.20 displays penalized and unpenalized local constant regression estimates for a typical sample of size 25. The penalty parameter was taken to be 0.6,

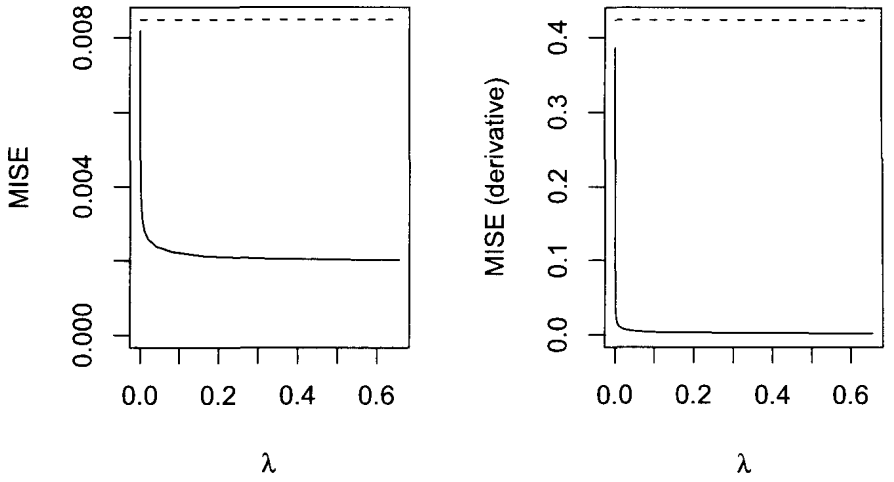

Figure 5.19: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of λ using $h = .1$.

and the bandwidth for both estimates was 0.1. The penalized curve matches the true curve almost perfectly on the constrained interval; similarly, the derivative is almost perfectly estimated on that interval.

For another example, we consider the same target function, but this time with design points taken from a larger domain: $[-3.5, 1.5]$. This time, we penalize a functional of the form (5.7) but this time with 40 grid points taken equally-spaced in the interval $[-3.2, 1.2]$.

Figure 5.21 shows MISE values for the penalized local constant regression estimator and the unpenalized local constant regression estimator for bandwidths $h \in \{.05, .1, \ldots, .6\}$ and where $\lambda = .05$. From this plot, we see that $h = .15$ should give reasonable performance.

Figure 5.22 shows MISE values for penalized local regression with penalty parameter values taken from the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 16\}$ and where $h = .15$. From this plot, it appears that there is rapid improvement in MISE as λ increases away from 0, but that when $\lambda > .4$, there is not much further sensitivity to choice of λ.

Figure 5.23 displays penalized and unpenalized local constant regression estimates for a typical sample of size 25. The penalty parameter was taken to be 0.6, and

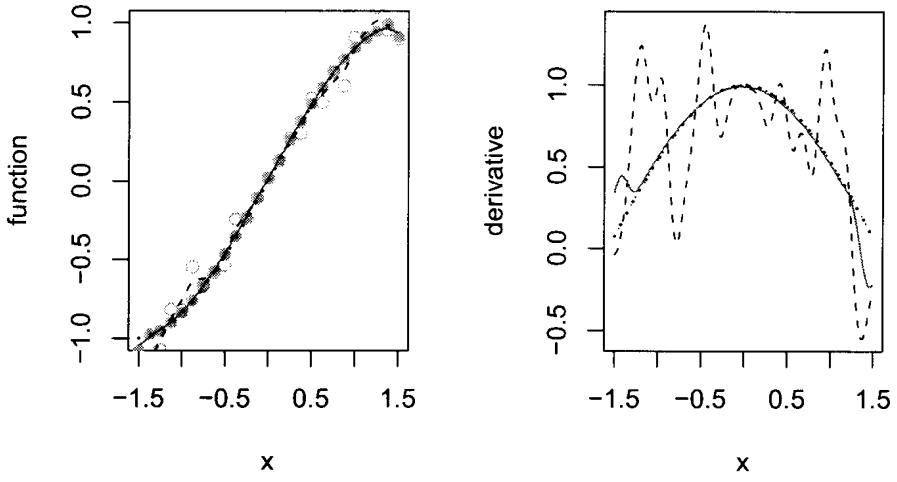

Figure 5.20: Left panel: sine function (dotted curve) estimates using penalized (solid curve) and unpenalized local regression(dashed curve). Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

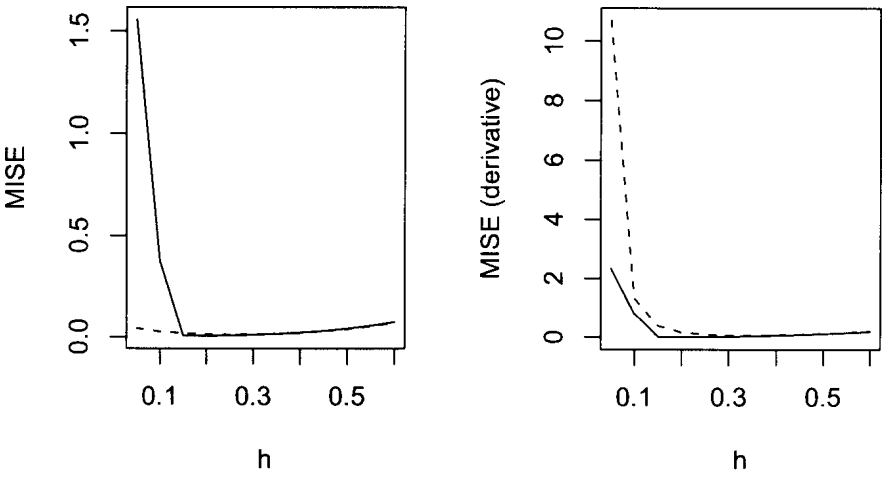

Figure 5.21: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function, over longer range, for various values of h, based on samples of size 25.

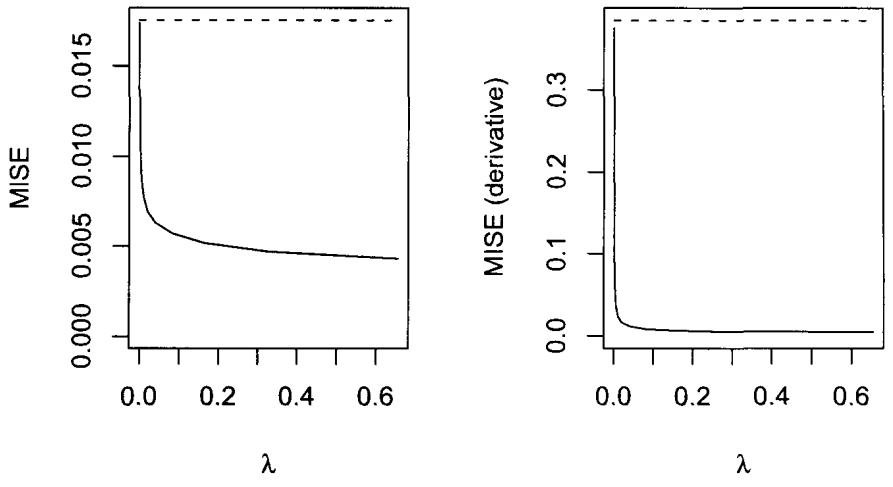

Figure 5.22: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function, over longer range, for various values of λ, using $h = .15$.

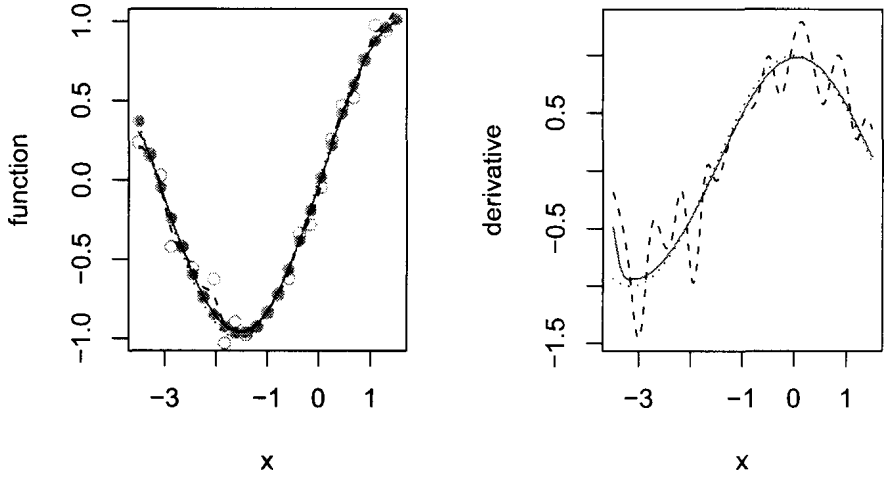

Figure 5.23: Left panel: sine function (dotted curve) estimates using penalized local constant regression (solid curve) for the penalty and unpenalized local regression(dashed curve). Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

the bandwidth for both estimates was 0.15. The penalized curve and its derivative match the corresponding true curves very well on the constrained interval.

5.7.2 Periodicity

The first target function is also well known to be periodic with period 2π:

$$g(x + 2\pi) = g(x).$$

This time, we will suppose we have a sample of size 200 taken at equally spaced design points on $[-10, 10]$. We will construct the penalty by choosing 150 equally spaced grid points on the interval $[-9.7, 9.7 - 2\pi]$. That is, we penalize

$$\sum_{i=1}^{150} (\widehat{g}(z_i + 2\pi) - \widehat{g}(z_i))^2 \tag{5.8}$$

where $z_1, \ldots, z_{150}$ represent the grid points.

Figure 5.24 shows MISE values for the penalized local constant regression estimator and the unpenalized local constant regression estimator for bandwidths $h \in \{.05, .1, \ldots, .6\}$ and where $\lambda = 1$. From this plot, we see that $h = .2$ should give reasonable performance.

Figure 5.25 shows MISE values for penalized local regression with penalty parameter values taken from the set $\{\lambda = .00001(2^j) : j = 0, 1, \ldots, 20\}$ and where $h = .2$. From this plot, it appears again that there is rapid improvement in MISE as λ increases away from 0, but that when $\lambda > 5$, there is not much further sensitivity to choice of λ.

Figure 5.26 displays penalized and unpenalized local constant regression estimates for a typical sample of size 200. The penalty parameter was taken to be 10, and the bandwidth for both estimates was 0.2.

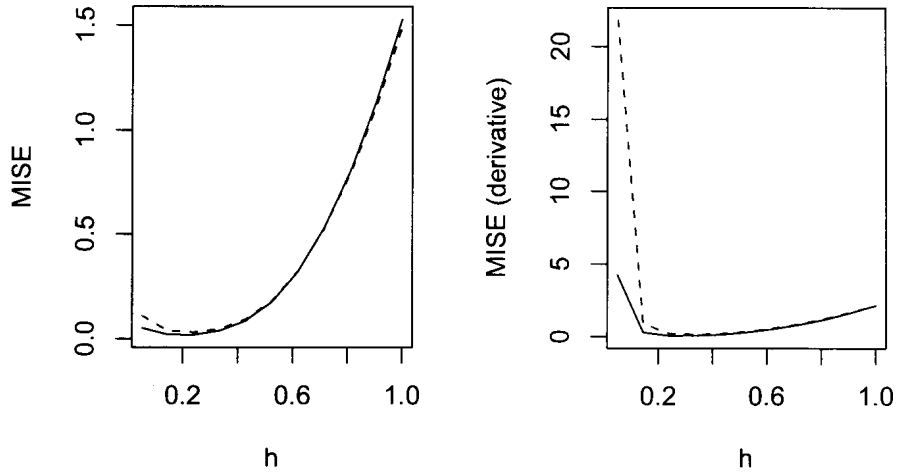

Figure 5.24: MISE estimates for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of h, based on samples of size 200.

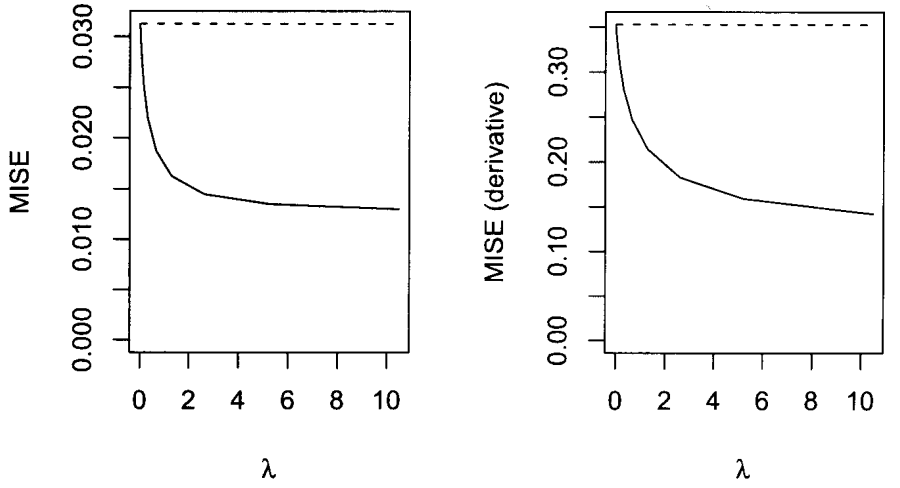

Figure 5.25: MISE for penalized (solid curve) and unpenalized (dashed curve) local constant estimates of the sine function for various values of λ, based on samples of size 200, using $h = .2$.

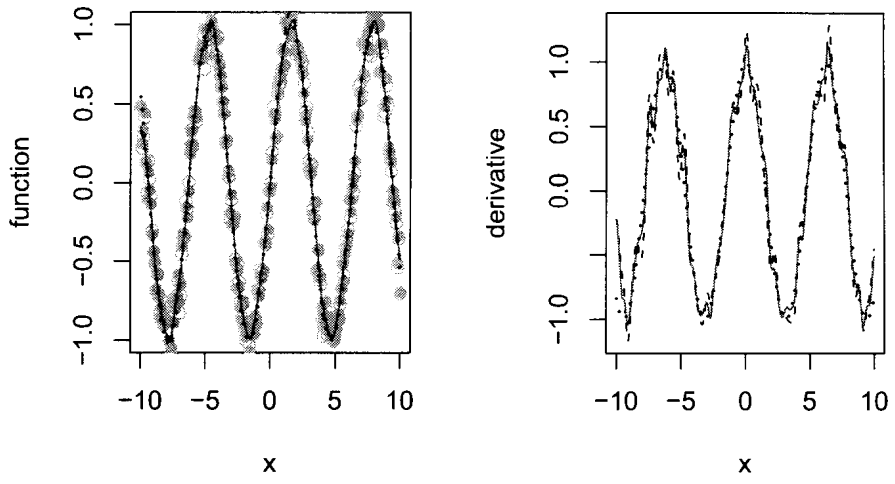

Figure 5.26: Left panel: sine function (dotted curve) estimates using penalized local constant regression with 401 grid points (dashed curve) and 150 grid points (solid curve) for the penalty, based on a sample of size 200. Right panel: derivatives of the corresponding estimates (solid and dashed curves) and true derivative (dotted curve).

5.8 Applications

5.8.1 Great Barrier Reef Survey Data

We again consider the Great Barrier Reef Survey data which featured in the paper of Hall and Huang (2001).

The bandwidth, $h = .0414$, was chosen using the `dpill()` function from the *KernSmooth* package (Wand and Ripley (2009)) in R. The dashed curve in Figure 5.27 represents the local constant regression estimate using this bandwidth.

The solid curve is the result of a penalized local constant estimate. The penalty term was based on the sum of squared second derivative estimates. Thus, the goal was to increase the smoothness in the estimated curve.

The penalty parameter, $\lambda = .00008$, was chosen to minimize the numerical approximation to the mean integrated squared error based on the MISE formula described in Section 5.5. However, because this formula is based on the true regression function, a pilot estimate was needed. We used the local constant estimate for this purpose. Minimizing this approximate MISE with respect to λ proceeded by grid

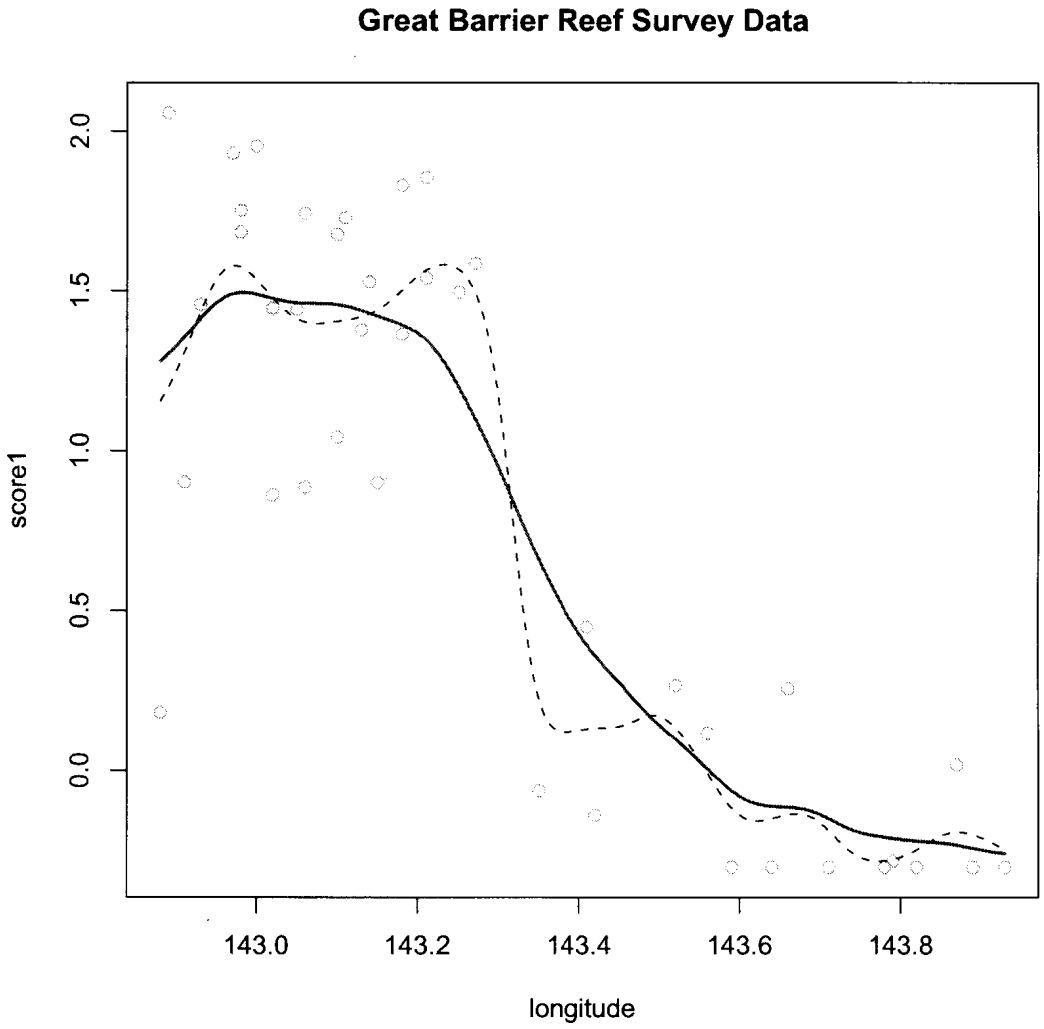

Figure 5.27: The Score 1 measurements from the Great Barrier Reef Survey plotted against longitude. The dashed curve represents the local constant regression estimate and the solid curve represents the penalized sharpened local constant regression estimate.

search. In other words, the expression

$$\widehat{\text{Var}}(\widehat{g}(x)) + \widehat{\text{Bias}}^2(\widehat{g}(x))$$

was integrated numerically (using a trapezoid rule), where the variance and bias estimates were obtained from formulas (5.4) and (5.5) with the unknown function $g(x)$ replaced by its local constant estimate $\widehat{g}(x)$.

5.8.2 Titanium Heat Data

Figure 5.28 shows the titanium heat data (e.g. de Boor (1978)) together with local constant and penalized local constant curve estimates. The bandwidth chosen here is one-third of value selected by the `dpill()` function. Larger values result in too much bias at the peak of the curve when estimated using the local constant estimator.

Again, the goal was to increase smoothness, using the sum of the squares of the second derivative, but the penalty was imposed only at values remote from the peak. In other words, no constraint was imposed for values of temperature between 890 and 900 degrees.

The rough approximation to the mean integrated squared error described in section 5.8.1 was minimized to obtain a value for the penalty parameter. This gave the value $\lambda = 8000$, and the resulting curve is plotted in Figure 5.28. Both estimated curves reach the peak satisfactorily, and there is not a great deal of difference between the two curves otherwise. However, it can be noted that the penalized curve does not interpolate the data to the same extent as the unpenalized local constant curve.

5.8.3 Winnipeg Temperature Data

Figure 5.29 shows daily maximum temperature data for the city of Winnipeg, Manitoba for the years 1960, 1961 and 1962. The data exhibit an obvious seasonality, and it may be reasonable to suppose that there is a period of 365.25 days. This leads us to model this data set using

$$y = g(t) + \varepsilon$$

where $g(t + 365.25) = g(t)$.

The bandwidth, $h = 13.8$, was chosen using the `dpill()` function from the *KernSmooth* package (Wand and Ripley (2009)) in R. The dashed curve in Figure

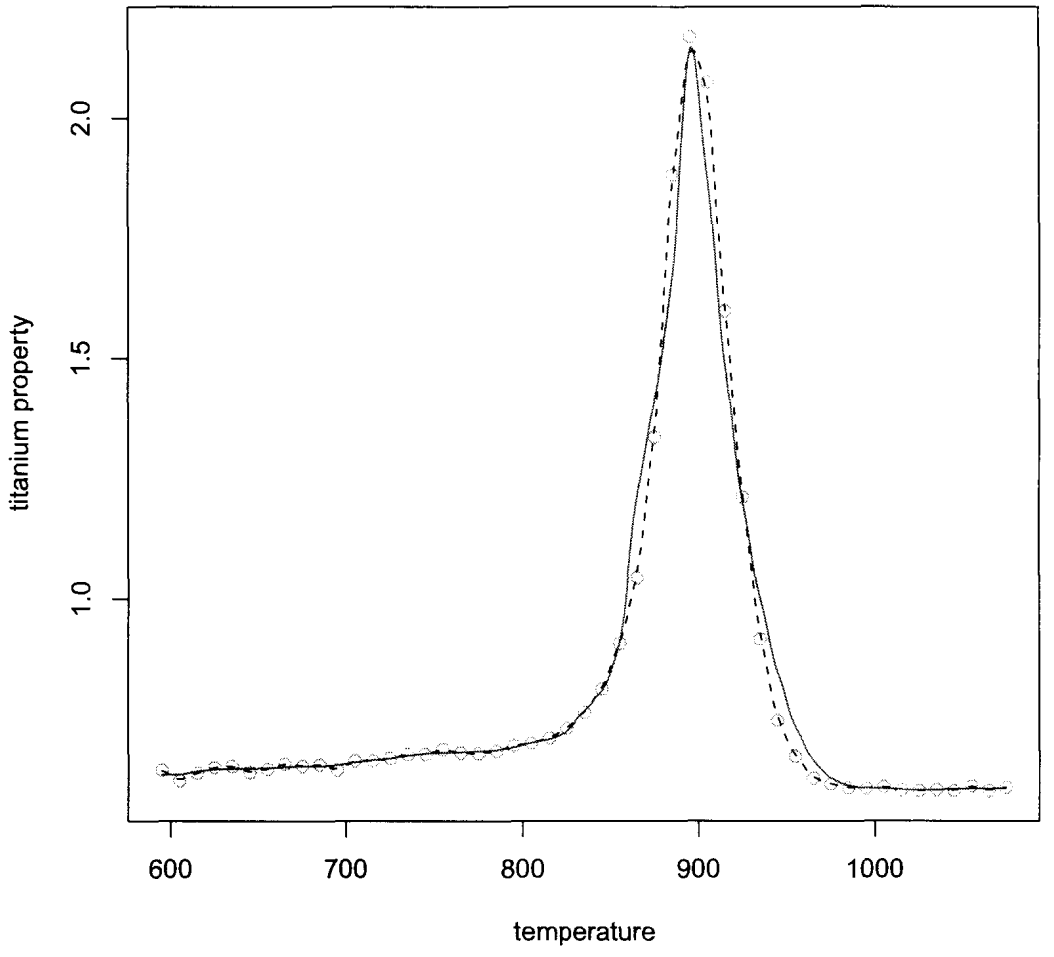

Figure 5.28: Titanium heat data. The dashed curve represents the local constant regression estimate with $h = 4.72$, and the solid curve represents the penalized sharpened local constant regression estimate using the same bandwidth and $\lambda = 8000$.

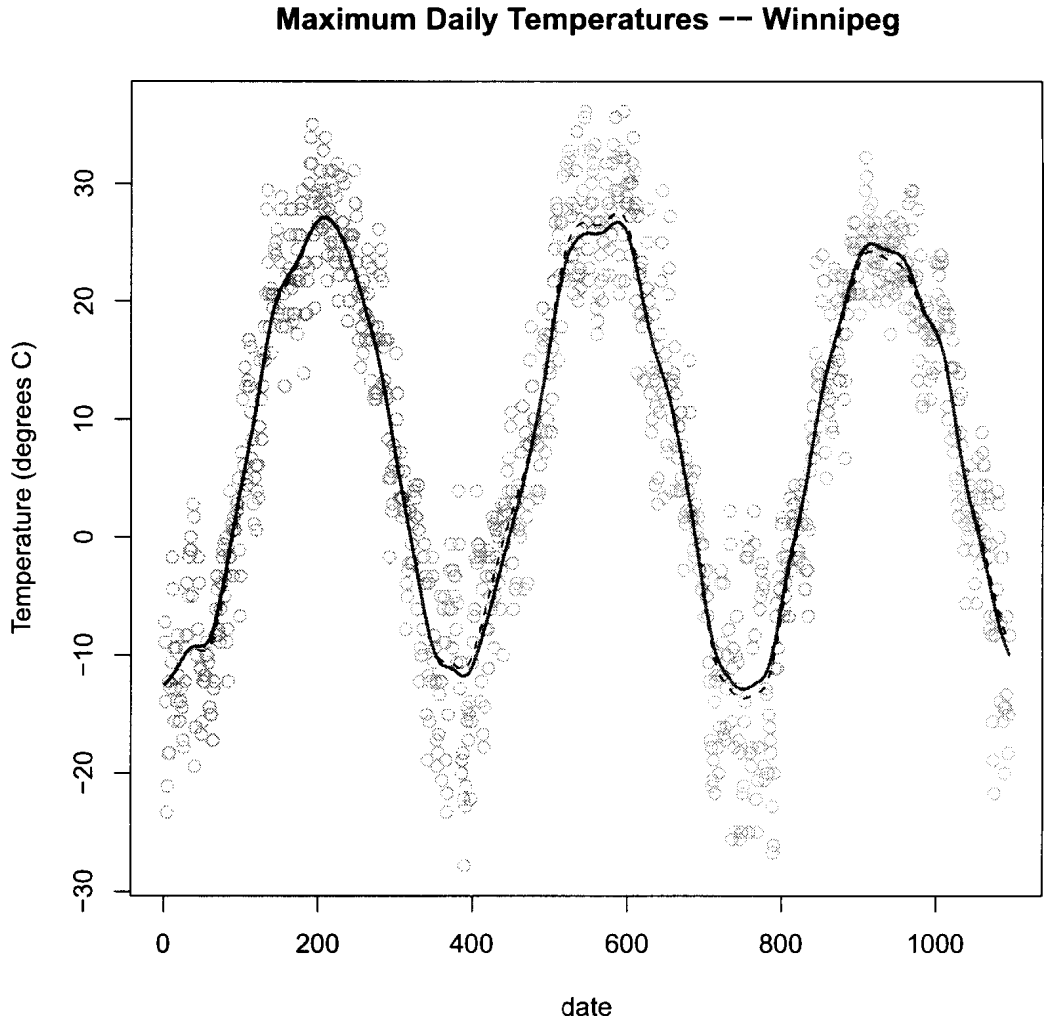

Figure 5.29: Winnipeg daily maximum temperature data for 1960, 1961 and 1962. The dashed curve represents the local constant regression estimate and the solid curve represents the penalized sharpened local constant regression estimate.

5.29 represents the local constant regression estimate using this bandwidth. There are more bumps in this curve than we should expect. This is mainly due to autocorrelation in the data which is well known to adversely affect bandwidth selector performance (Wand and Jones (1995)).

The solid curve is the result of a penalized local constant estimate. The penalty parameter, $\lambda = 3$, was chosen to minimize the rough approximation to the mean integrated squared error described in Section 5.8.1

The method for choosing λ may be biased downwards somewhat. This observation is based on a very limited number of simulation experiments, in which the target function was a sine function (with period chosen to resemble the temperature data in this example) and 1096 independent observations were simulated. In all of these cases, the selected value of λ based on the estimated regression function was less than the value based on the true regression curve. Another problem is autocorrelation in the measured temperature values. Another simulation experiment revealed that λ will be substantially underestimated if there is autocorrelation of the type observed here (i.e. approximately autoregressive with order 1, with a first lag autocorrelation near 0.6).

Figure 5.30 shows what happens when a substantially larger penalty parameter is applied to the Winnipeg temperature data. The resulting curve is an improvement, since it reveals the periodicity in the data more strongly, and some of the anomalous bumps still visible when using the smaller value of λ have been removed.

5.9 Conclusions

In this chapter, we studied an alternative constrained sharpening local regression estimator by imposing the constraint as a penalty. This penalized kernel regression estimator can easily handle constraints which are quadratic in the responses y. Such constraints cannot be handled with the original constrained data sharpening using ordinary quadratic programming, and are thus less tractable.

Furthermore, the fact that we can express penalized sharpened responses analytically has allowed us to investigate finite sample properties for specific cases. Simulation and theoretical results demonstrate that the penalized local regression estimator can lead to an improvement over the local regression estimator. Although not exploited in this thesis, the explicit bias and variance formulas could be used to add confidence limits to penalized curve estimates.

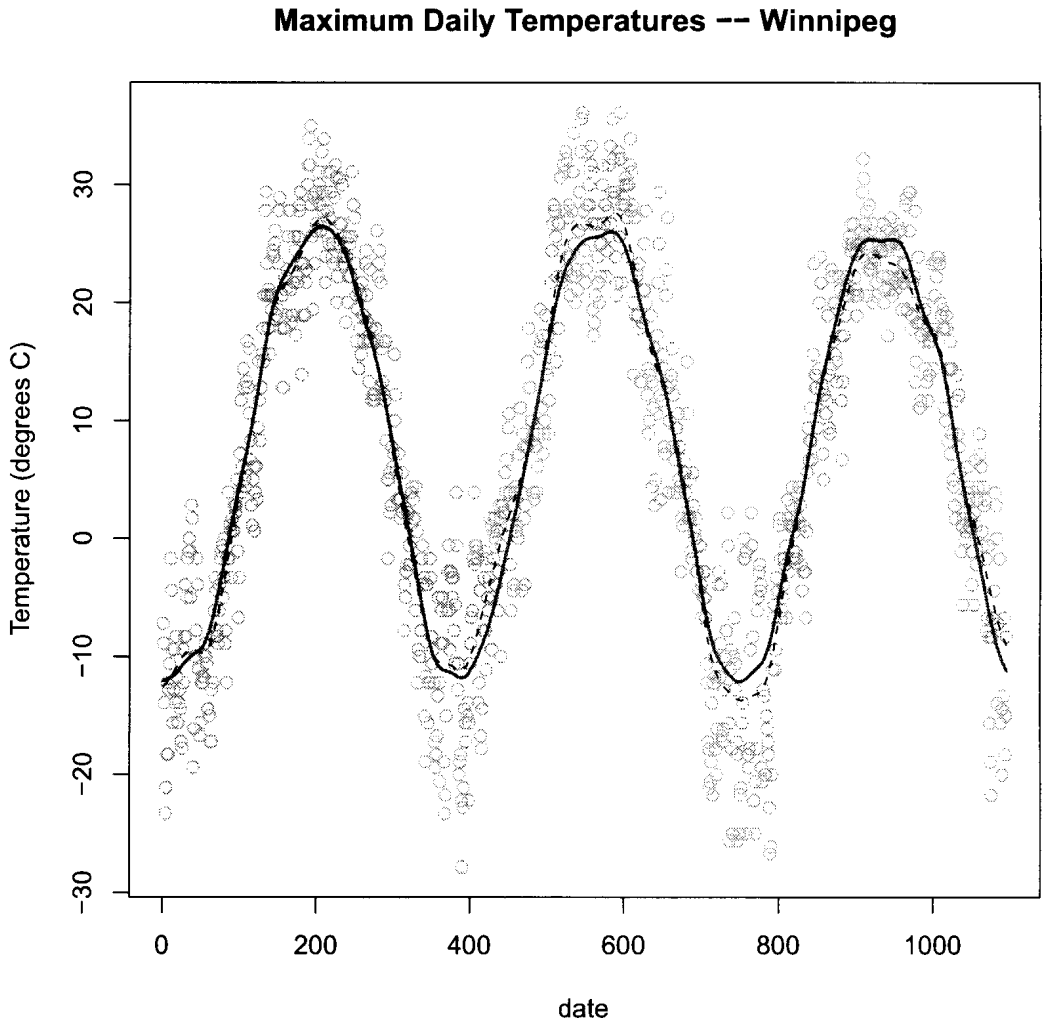

Figure 5.30: Winnipeg daily maximum temperature data for 1960, 1961 and 1962. The dashed curve represents the local constant regression estimate and the solid curve represents the penalized sharpened local constant regression estimate using the same bandwidth as before and $\lambda = 30$.

Although much of the focus of the chapter was on roughness penalties, leading to a form of smoothing kernel estimator, it was also shown that other penalties can be considered. The data examples exhibit some of the flexibility of the method as well.

Chapter 6

CONCLUSIONS AND FURTHER WORK

6.1 Contributions of this thesis

This thesis has been concerned with problems arising when one wants to apply nonparametric local regression subject to qualitative constraints and subject to interval-censored responses. These problems were studied via asymptotic theory and by simulation.

In Chapter 3, iterated conditional expectation methods and local likelihood for nonparametric interval-censored regression estimators were developed. Estimators for the error variance were proposed and compared with previously suggested estimators from the literature. Simulation results showed that local likelihood estimation is often superior to local regression estimators where observations have been imputed using either interval midpoints or iterated conditional expectations when the censoring intervals are wide or of varying width. The fact that the midpoint imputation method is inferior to the other methods when the interval widths are variable but that it still enjoys consistency properties strengthens the argument in favor of the usefulness of the other methods, at least when the censoring intervals are variable. When the intervals are smaller and of fixed width, none of the imputation approaches dominate the others.

In Chapter 4, constrained data sharpening for nonparametric regression was applied to new situations such as where constraints are defined in terms of differential operators. Data sharpening was compared with competing kernel methods in terms of bias, variance and MISE. We proved that the constrained data sharpening estimator has the same rate of convergence as the constrained weighting estimator of Hall and Huang (2001).

In Chapter 5, penalized data sharpening was proposed as a new form of constrained data sharpening. The sharpened responses can be computed analytically which makes the method very convenient. It is also very easy to study performance in terms of bias and variance. It is also possible to understand why the method does not perform as well as might have been expected. Also, a new approach based on

selecting appropriate linear combinations of eigenvectors of a matrix derived from the constraints has been introduced, though not yet studied systematically.

6.2 Research Questions Arising from the Thesis

This thesis has explored interval-censored regression estimators and data sharpening estimators in detail. However, a number of questions emerge from this thesis.

We have only discussed local regression and local likelihood estimators with one covariate. With few exceptions, we have only considered normal errors with constant variance. The methods considered can be extended to include multiple covariates with normal and non-normal errors. Replacing the normal cdf in the local likelihood method for interval-censored data with a nonparametric estimate of the cdf as employed by Rabinowitz *et al.* (1995) is worth investigating. Meanwhile, the simulation study of Chapter 3 does not treat the midpoint imputation and nonparametric iterated conditional expectation imputation methods fairly, since these two methods have to work without the normality assumption that benefits other two methods. It would have been interesting to see if the conclusions remain the same when the ε_i follows logistic or smallest extreme value distributions. Also since most parametric regression models that fit well in survival analysis are log-linear, it would have been nice to have a simulation where the log of the response follows model (1.1).

There is scope for much more study of the use of data sharpening to handle differential constraints. In each of the examples considered in Chapter 4, the regression function is the analytic solution to a differential equation. This has made it possible to check the accuracy of the methods being applied, but this does not mean it is not possible to apply the techniques in situations where only the differential equation itself is available. Furthermore, only linear differential equations were considered here (with one exception), in order that the data sharpening method can proceed using quadratic programming. Using nonlinear programming methods, it is possible, in principle, to handle constraints derived from a larger class of differential equations. As was seen in Chapter 4, and again in Chapter 4, the data sharpening method is not guaranteed to be accurate in all circumstances; it sometimes requires considerable care to implement properly. This will likely continue to be the case when nonlinear differential constraints are operative; thus, the study of this problem may be large enough to generate a whole new thesis.

Data sharpening may be a potential way to mitigate the well-known "curse of dimensionality" associated with higher-dimensional kernel smoothing problems.

Handling a non-constant variance would also be of interest, particularly if there is qualitative information available about the nature of the variance as a function of the regressor or response. The question of testing the validity of constraints was touched on briefly in this thesis, but is an area that requires much more extensive study.

The techniques presented in this thesis can also be applied to other nonparametric regression estimators, such as regression splines, including the kernel spline regression estimator proposed by Braun and Huang (2005). In addition, the convergence rate of the data sharpening estimator can be obtained under the more general constraints (4.4) or (4.5).

The example involving correlated fire area intervals provides motivation to study these methods more closely when the data exhibit serial dependence. Another form of dependence arises from the way in which data such as the aspen flush data have been sampled. In order to handle multiple years of data, it will likely be necessary to use random effects models.

The penalty method should be studied more systematically for higher order local polynomial regression. The bandwidth and penalty parameter selectors that were proposed in Chapter 5 need to be studied systematically as well. Furthermore, it may be possible to provide confidence limits for penalized kernel regression estimates using the explicit pointwise bias and variance formulas given in the thesis.

We have also studied, briefly, how combining the data sharpening method of Choi *et al.* (2000) with other data sharpening methods can affect estimation performance. Incorporating qualitative information into interval-censored regression problems is another way in which combining two or more of these types of procedures may help to improve estimation performance. As a final example, we conclude the thesis with an illustration as to how this may be done, and show how we can address the aspen flush problem stated in the opening chapter of the thesis.

6.2.1 Constrained Data Sharpening for Interval-censored Responses

When the responses are interval-censored, incorporating qualitative information into the local polynomial estimator is possible. In this example, we show how monotonicity can be incorporated.

When the responses Y_i are interval-censored and observed as intervals I_i's, a local constant estimator for $g(x)$ is

$$\widehat{g}(x) = \frac{\sum_i K_h(x - x_i)\widehat{\mathrm{E}}(Y_i | Y_i \in I_i)}{\sum_i K_h(x - x_i)}, \tag{6.1}$$

where

$$\widehat{\mathrm{E}}(Y_i|Y_i \in I_i) = \widehat{g}(x_i) + \frac{\int_{I_{\varepsilon_i}} x\widehat{f}_\varepsilon(x)dx}{\int_{I_{\varepsilon_i}} \widehat{f}_\varepsilon(x)dx},$$

and $I_{\varepsilon_i} = [L_i - \widehat{g}(x_i), R_i - \widehat{g}(x_i)]$.

We can incorporate monotonicity into this estimator using a form of data sharpening.

$$\widehat{g}^*(x) = \frac{\sum_i K_h(x - x_i)Y_i^*}{\sum_i K_h(x - x_i)}, \tag{6.2}$$

where the Y_i^* are sharpened versions of $\widehat{\mathrm{E}}[Y_i|Y_i \in I_i]$ and can be obtained by minimizing

$$\sum_i (Y_i^* - \widehat{\mathrm{E}}[Y_i|I_i])^2, \text{subject to } \widehat{g}'(x) \geq 0. \tag{6.3}$$

The details of the algorithm are:

(1) Compute the interval midpoints of the Y_i's: y_i^0.

(2) Compute the initial regression function

$$\hat{g}_o(x) = \frac{\sum_i K_h(x - x_i)y_i^0}{\sum_i K_h(x - x_i)}.$$

Choose h using a DPI approach, for example.

(3) Estimate the noise density using iterated conditional expectation as in Chapter 3. The interval-censored errors are estimated as $I_{\varepsilon_i} = I_i - \widehat{g}_o(x_i)$ and the error density is estimated by the fixed point of

$$\widehat{f}_\varepsilon(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_{I_{\varepsilon_i}} K_h(z - w)\widehat{f}_\varepsilon(w)dw}{\int_{I_{\varepsilon_i}} \widehat{f}_\varepsilon(w)dw}.$$

(4) Compute the conditional expectation, $\widehat{\mathrm{E}}[\varepsilon_i|I_{\varepsilon_i}]$, of the ε's.

(5) Set $\widehat{\mathrm{E}}[y_i|I_i] = \widehat{g}(x_i) + \widehat{\mathrm{E}}[\varepsilon_i|I_{\varepsilon_i}]$, where $\widehat{g}(x_i) = \frac{\sum_i K_h(x-x_i)\widehat{\mathrm{E}}(y_i|I_i)}{\sum_i K_h(x-x_i)}$.

(6) Minimize $\sum_i (y_i^* - \widehat{\mathrm{E}}(Y_i|I_i))^2$ subject to $\widehat{g}'(x) \geq 0$.

(7) Set $\widehat{g}_0(x) = \widehat{g}(x)$ and return to step (2) or stop when the iteration has achieved convergence.

Although the convergence of this iteration has not been proven, our experience with it indicates that under typical conditions, convergence is rapid: one or two iterations are sometimes all that is required.

6.2.2 A Final Simulation Study

In order to check whether constrained data sharpening can improve estimation when responses are interval-censored, we conducted a final simulation study.

We drew 1000 samples $(X_i, Y_i), 1 \leq i \leq n$ of size $n = 30$ from the regression model with

$$g(x) = \frac{e^{5x}}{1 + e^{5x}}$$

with equally spaced design points on $[-1, 1]$. The error distribution was normal with standard deviation 0.2. The response intervals were generated according to

$$L_i = Y_i - \exp(n, 10)$$

and

$$R_i = Y_i + \exp(n, 10).$$

For each simulated sample, the following estimators were applied:

1. local linear regression.
2. the bias-reduced data sharpening method of Choi *et al.* (2000) applied to
   (a) local constant regression, perturbing the design points only.
   (b) local linear regression, perturbing the design points and responses.
   (c) local linear regression, perturbing the responses only.
3. constrained data sharpening under the Cases listed below.

For the data sharpened local linear estimates, we imposed three different constraints:

Case 1: $g(x)$ is non-negative and monotonically decreasing.

Case 2: $g(x)$ is non-negative, and convex for $x < m_1$ and concave for $x > m_1$.

Case 3: $g(x)$ satisfies the conditions of Cases 1 and 2 and $g(1) = 1$.

Figures 6.1 to 6.3 show typical realizations from this simulation. Figure 6.1 shows the result of fitting the data with the unconstrained interval-censored local linear estimator as well as the constrained estimator based on the non-negativity and monotonicity assumptions. Figure 6.2 compares the unconstrained estimate with the

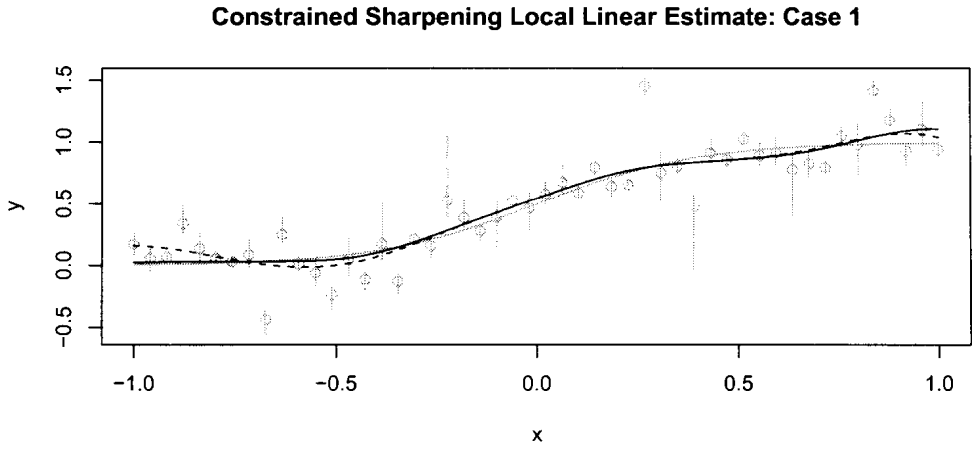

Figure 6.1: Local linear estimates for interval-censored responses subject to constraints. True curve (solid grey curve), unconstrained local linear estimate (dashed black curve) and constrained (Case 1) estimate (solid black curve).

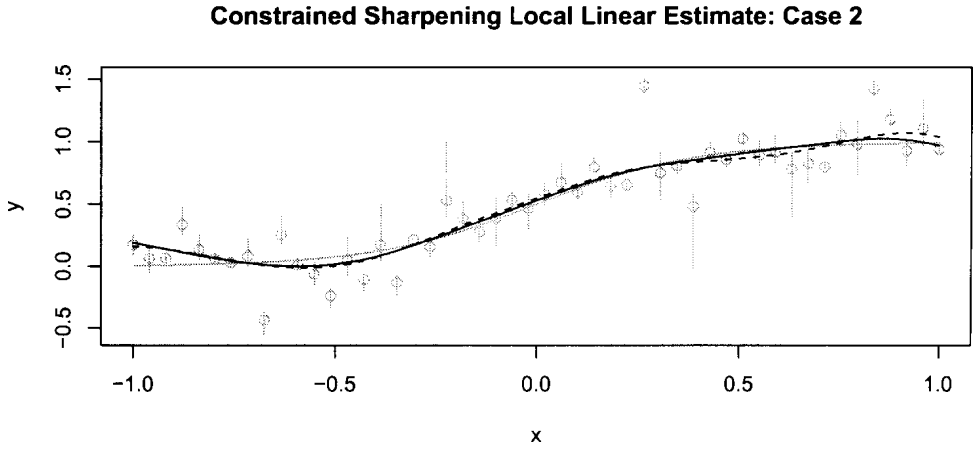

Figure 6.2: Local linear estimate for interval-censored responses subject to constraints. True curve (solid grey curve), unconstrained local linear estimate (dashed black curve), constrained (Case 2) estimate (solid black curve).

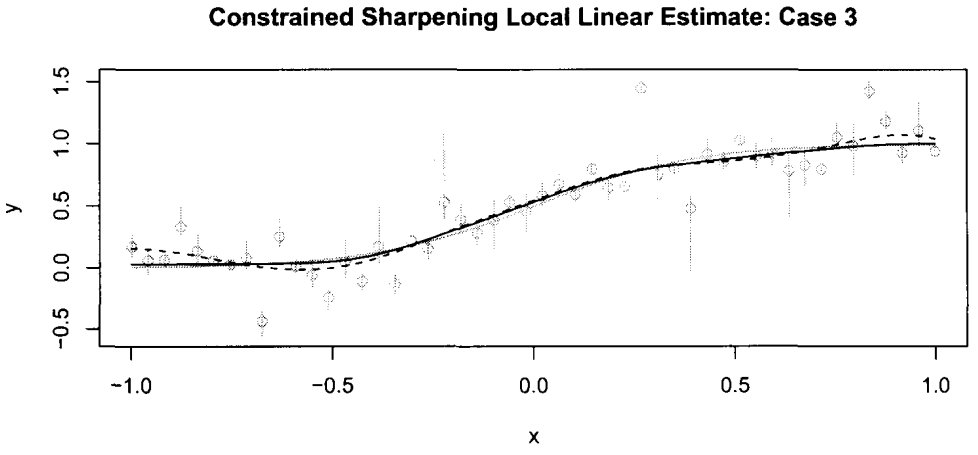

Figure 6.3: Local linear estimates for interval-censored responses subject to constraints. True curve (solid grey curve), unconstrained local linear estimate (dashed curve), constrained (Case 3) estimate (solid black curve).

constrained estimate under the further assumption of convexity. Figure 6.3 compares the unconstrained estimate with the constrained estimate under the additional assumption that $g(1) = 1$.

The MISE was computed as well as the MSE at the points $x = 0$ and $x = 0.6$, for each method. The results are given in Table 6.1.

Table 6.1 shows that imposing qualitative constraints leads to substantial reductions in MISE. For the third constraint, the point constraint $g(1) = 1$ gives fairly accurate information about the function (the true value of $g(1)$ is 0.9933). The MISE for the third case is the smallest among the three kinds of constraints.

The pointwise squared bias, variance and MSE are depicted in Figure 6.4.

6.2.3 Application to the Aspen Flush Data

The aspen flush data introduced at the beginning of this thesis will be investigated again here. The local linear estimate subject to a monotonically decreasing constraint is applied. The unconstrained and data sharpened local linear estimates are depicted in Figure 6.2.3. The same direct plug-in bandwidth $h = 0.126$ is used for these two estimates.

The figure shows that the local linear estimator incorporating the monotonicity information is smooth and fits the data set well compared to the unconstrained local linear estimator.

Table 6.1: MISE and MSE (at two design points) comparisons of kernel regression estimates when responses are interval-censored. CHR constant: local constant with perturbed design points; CHR linear-xy: local linear with perturbed design points and responses; CHR linear-y: local linear with perturbed responses; DS Cases 1-3: data sharpening subject to Constraint Cases 1-3.

Target Function: Logistic Curve

Optimal bandwidth	n	MISE ($\times 10^2$)	Mean squared error at $x = 0$ ($\times 10^2$)	Mean squared error at $x = 0.6$ ($\times 10^2$)
local linear estimate	30	1.112	0.468	0.564
CHR constant		1.309	0.564	0.660
CHR linear-xy		1.230	0.470	0.658
CHR linear-y		1.160	0.511	0.605
DS Case 1		0.987	0.455	0.508
DS Case 2		0.923	0.423	0.554
DS Case 3		0.773	0.397	0.404
local linear estimate	50	0.746	0.364	0.385
CHR constant		0.808	0.380	0.413
CHR linear-xy		0.786	0.344	0.431
CHR linear-y		0.717	0.346	0.380
DS Case 1		0.636	0.354	0.330
DS Case 2		0.609	0.307	0.383
DS Case 3		0.502	0.301	0.271
local linear estimate	100	0.452	0.193	0.226
CHR constant		0.430	0.176	0.213
CHR linear-xy		0.425	0.158	0.217
CHR linear-y		0.399	0.167	0.200
DS Case 1		0.375	0.189	0.197
DS Case 2		0.365	0.176	0.225
DS Case 3		0.305	0.172	0.173

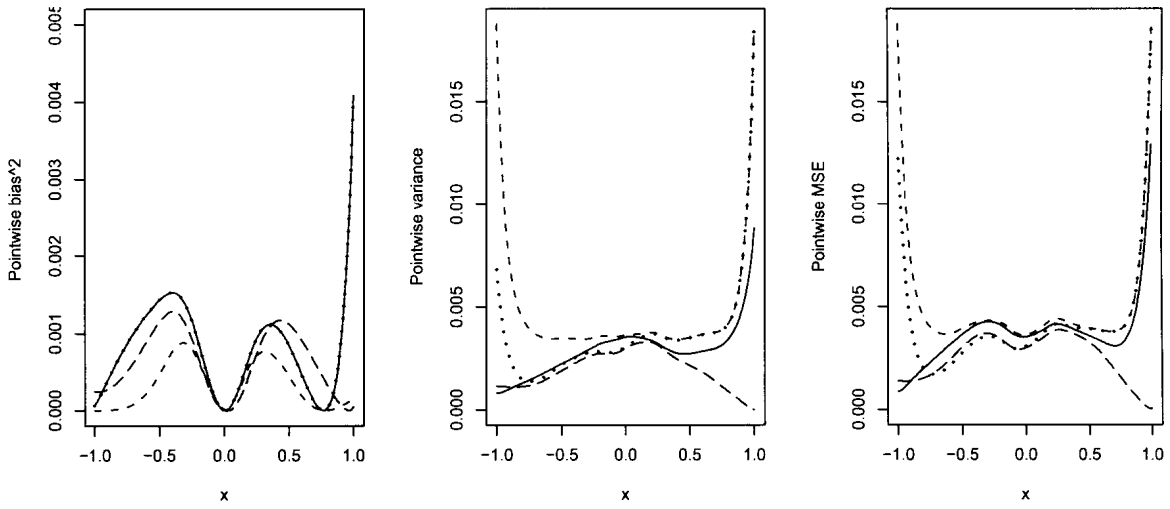

Figure 6.4: Local linear estimates for interval-censored responses subject to constraints. True curve (solid grey curve), unconstrained local linear estimate (dashed curve) and constrained (Case 3) estimate (solid black curve).

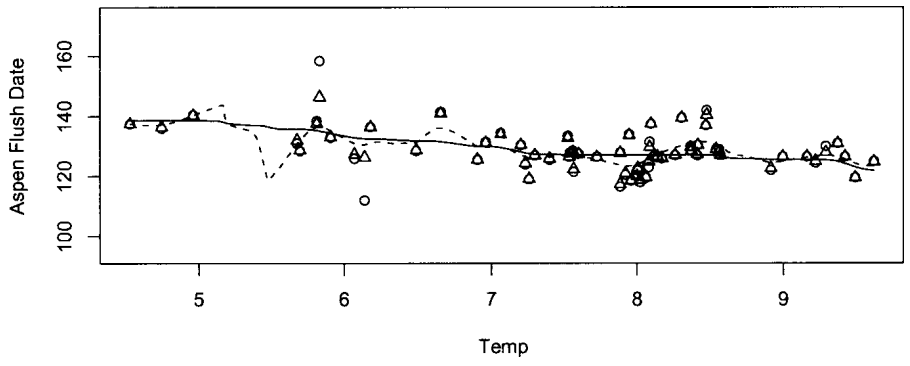

Figure 6.5: Local linear estimates for aspen flush data. Unsharpened estimate (dashed line) and monotonically decreasing estimate (solid line).

BIBLIOGRAPHY

Betensky, R., Lindsey, J., Ryan, L. and Wand, M. (1999). Local em estimation of the hazard function for interval-censored data. *Biometrics* **55**, 238–245.

Betensky, R., Lindsey, J., Ryan, L. and Wand, M. (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine* **21**, 263–275.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford UNIV Press.

Braun, W., Duchesne, T. and Stafford, J. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics* **33**, 39–60.

Braun, W. and Hall, P. (2001). Data sharpening for nonparametric inference subject to constraints. *Journal of Computational and Graphical Statistics* **10**, 786–806.

Braun, W. and Huang, L. (2005). Kernel spline regression. *The Canadian Journal of Statistics* **33**, 259–278.

Brunk, H. (1955). Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics* **26**, 607–716.

Buckley, J. and James, L. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika* **86**, 941–947.

Choi, E., Hall, P. and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *The Annals of Statistics* **28**, 1339–1355.

Chu, C. and Marron, J. (1991). Choosing a kernel regression estimator. *Statistical Science* **6**, 404–436.

Cressie, N. and Read, T. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B* **46**, 440–464.

de Boor, C. (1978). *A Practical Guide to Splines.* Springer.

Dette, H., Neumeyer, N. and Pliz, K. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12**, 469–490.

Dierckx, P. (1980). An algorithm for cubic spline fitting with convexity constraints. *Computing* **24**, 349–371.

Doksum, K., Peterson, D. and Samarov, A. (2000). On variable bandwidth selection in local polynomial regression. *Journal Of The Royal Statistical Society Series B* **62**, 431–448.

Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of American Statistical Association* **89**, 560–570.

Friedman, J. and Tibshirani, R. (1984). The monotone smoothing of scatter plots. *Technometrics* **26**, 243–350.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics* **29**, 624–647.

Hall, P., Huang, L.-S., Gifford, J. and Gijbels, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statistics* **10**, 592–614.

Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society: Series B* **61**, 143–158.

He, X. and Shi, P. (1998). Monotone b-spline smoothing. *Journal of American Statistical Association* **93**, 643–650.

Heckman, J. and Polachek, S. (1974). Empirical evidence of functional form of the earnings-schooliing relationship. *Journal of the American Statistical Association* **69**, 350–354.

Henderson, D. and Parmeter, C. (2009). Imposing economic constraints in nonparametric regression: survey, implementation and extension. *Discussion Paper* **0**, 0.

Kim, J. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. R. Statist. Soc. B.* **65**, 489–502.

Komrek, A., Lesaffre, E. and J., H. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics* **14**, 726–745.

Law, G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine* **11**, 1569–1578.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Loader, C. (1999). *Local Regression and Likelihood.* Springer.

Mammen, E. (1991*a*). Estimating a smooth monotone regression function. *The Annals of Statistics* **19**, 724–740.

Mammen, E. (1991*b*). Nonparametric regression under qualitative smoothness assunptions. *The Annals of Statistics* **19**, 741–759.

Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics* **26**, 239–252.

Mayer, M. (2008). Inference using shape-restricted regression splines. *The annals of Applied Statistics* **2**, 1013–1033.

Miller, R. (1976). Least squares regression with censored data. *Biometrika* **63**, 449–464.

Mukerjee, H. (1988). Monotone nonparametric regression. *Annals of Statistics* **16**, 741–750.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Pearson, G. and Qua, F. (1993). High precision^{14} measurement of irish oaks to show the natural^{14} variations from ad 1840-5000 bc: a correction. *Radiocarbon* **35**, 105–113.

Rabinowitz, D., Tsiatis, A. and Aragon, J. (1995). Regression with interval-censored data. *Biometrika* **82**, 501–513.

Racine, J. and Parmeter, C. (2009). Constrainted nonparametric kernel regression: estimation and inference. *preprint* **0**, 0.

Ramsay, J. (1988). Monotone regression splines in action (with comments). *Statistical Science* **3**, 425–461.

Ruppert, D., Sheather, S. and Wand, M. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* **90**, 1257–1270.

Schmee, J. and Hahn, G. (1979). A simple method for regression analysis with censored data. *Technometrics* **21**, 417–432.

Sheather, S. J. and Jones, M. C. (1986). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society-Series B* **53**, 683–690.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

Team, R. D. C. (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.

Turlach, B. and Weingessel, A. (2007). *quadprog: Functions to solve Quadratic Programming Problems*. R package version 1.4-11.

Wand, M. and Jones, M. (1995). *Kernel Smoothing.* Chapman and Hall.

Wand, M. and Ripley, B. (2009). *KernSmooth: Functions for kernel smoothing for Wand & Jones (1995).* R package version 2.23-2.

Yatchew, A. and Bos, L. (1997). Nonparametric regression and testing in economic models. *Journal of Qualitative Economics* **13**, 81–131.

Yatchew, A. and Hardle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics* **133**, 579–599.

Yu, Q., Wong, Y. C. and Kong, F. (2006). Consistency of the semi-parametric mle in linear regression models with interval-censored data. *Scandinavian Journal of Statistics* **33**, 367–378.

Zeng, D., Cai, J. and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica* **16**, 287–302.

Appendix A

Derivation of the Likelihood under Two Censoring Mechanisms

Proof of (3.8).
Since X_{1i} and X_{2i} are exponentially distributed with rate λ, the probability density function of R_i and L_i conditional on y are given respectively by

$$f_{R_i|\mathbf{y}}(r) = \lambda e^{-\lambda(r-y)}, \text{for y} \leq r$$

and

$$f_{L_i|\mathbf{y}}(r) = \lambda e^{-\lambda(y-l)}, \text{for y} \geq r.$$

In addition, $\mathbf{y} \sim N(\mu, \sigma^2)$. Thus the joint probability density function of $L_i, R_i, \mathbf{y}$ is

$$\begin{aligned} f_{L_i,R_i,\mathbf{y}}(l,r,y) &= f_{L_i,R_i|y}(l,r) f_{\mathbf{y}}(y) \\ &= f_{L_i|\mathbf{y}}(l) f_{R_i|y}(r) f_{\mathbf{y}}(y), \end{aligned} \tag{A.1}$$

since the random variables L_i and R_i are independently distributed conditional on y. Thus, we have

$$f_{L_i,R_i,\mathbf{y}}(l,r,y) = \frac{\lambda^2 e^{-\lambda(r-l)} e^{-(y-\mu)^2/2\sigma^2}}{\sqrt{2\pi}},$$

where $l \leq y \leq r$. And the marginal joint probability density function of L_i and R_i is

$$\begin{aligned} f_{L_i,R_i}(l,r) &= \lambda^2 e^{-\lambda(r-l)} \int_l^r \frac{e^{-(y-\mu)^2/2\sigma^2}}{\sqrt{2\pi}} dy \\ &= \lambda^2 e^{-\lambda(r-l)} \left(\Phi(\frac{r-\mu}{\sigma}) - \Phi(\frac{l-\mu}{\sigma}) \right), \end{aligned} \tag{A.2}$$

where $r \geq l$.

So the likelihood function is

$$L(\lambda, \mu, \sigma^2) = \prod_{i=1}^{n} \lambda^2 e^{-\lambda(r_i - l_i)} \left(\Phi(\frac{r_i - \mu}{\sigma}) - \Phi(\frac{l_i - \mu}{\sigma}) \right). \tag{A.3}$$

where $r_i \geq l_i$, for $i = 1, \ldots, n$ and the log-likelihood function is

$$l(\lambda, \mu, \sigma^2) = \sum_{i=1}^{n} \log \lambda^2 - \lambda \sum_{i=1}^{n} (r_i - l_i) + \sum_{i=1}^{n} \log \left(\Phi(\frac{r_i - \mu}{\sigma}) - \Phi(\frac{l_i - \mu}{\sigma}) \right). \tag{A.4}$$

Thus removing the terms which do not involve the unknown parameters μ and σ, the target log-likelihood function is given in (3.8).

For the fixed width interval, we have

$$l(\mu, \sigma^2) = \prod_{i=1}^{n} P(l_i \leq \mathbf{y}_i \leq r_i), \tag{A.5}$$

where l_i and r_i are the observations,not random variables and y_i is normally distributed with mean μ and variance σ. So the log-likelihood function in this case can be easily obtained as (3.8).