VARIANCE FUNCTION ESTIMATION

by

TITUS KITHANZE KIBUA

A dissertation submitted to the City University, London

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in the Department of Actuarial Science and Statistics

June 1995

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I thank God. Secondly, while I would probably go on for several pages to thank individually each of the kind souls without whom this dissertation would have never been completed, I would like especially to acknowledge those people whose help and understanding were of the highest order of magnitude. My most heartfelt appreciation goes to my advisor, Professor Henry Wynn. His vast expertise, availability for help and guidance, boundless enthusiasm and most importantly his patience were instrumental in the completion of my degree. Along the way, I had the opportunity to meet people whose help I will continue to cherish long after I have left City University. All the people at the Engineering Design Center, I thank you. To the British Council, thank you for providing financial assistance.

I cannot forget to thank the management and staff of the Alliance Club, the hostel where I stayed throughout my study. I thank St.Augustine, the church where I attended service every Sunday.

The understanding and support of my brother Dr. T. N. Kibua has been far above and beyond the call of duty; his phone calls and visit have been a lifeline for me throughout my stay here. The appreciation and gratitude I have for the love, encouragement and tolerance for my family during my study is simply inexpressible. A special debt is owed my wife, my two sons and a daughter for their constant encouraging letters, prayers and for maintaining a happy and cheerful home despite my frequent absence. Finally, to my father, brothers and sisters, thank you all for your corporation. I dedicate this dissertation to all of you.

Declaration of Power of Discretion

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

This thesis concerns the estimation of the variance function in regression data when the classical assumption of constant variance is violated. We have adopted the assumption that either the variance function is parametric or is unknown but smooth. The purpose in this thesis is to develop the techniques that are currently available.

The thesis contains two major parts. After an introduction chapter, Chapters 2 and 3 discuss the parametric approach for estimating variance functions. Chapter 2 reviews in depth a large and widely scattered literature, describes the specific procedures and provides an overview of the theory employed in estimating variance functions. Chapter 3 provides detailed empirical study of these procedures.

The second part of the thesis discusses the nonparametric approach for estimating variance functions. Chapter 4 describes in detail the techniques that are involved and studies these techniques empirically revealing that the use of sample standard deviations and absolute residuals may lead to better final variance estimates. One of the techniques associated with nonparametric approach is the determination of the amount of smoothing. Chapter 5 give some analytic theory particularly for bias, providing a new criterion for determining the amount of smoothing.

Finally, Chapter 6 applies both methods to real data.

Chapter 1

INTRODUCTION

1.1 Introduction

Classical regression theory is concerned with the study of the relationship between a variable $\boldsymbol{Y}$ and a p - variate vector of variables $\boldsymbol{X} = (X_1, \cdots, X_p)$. The vector $\boldsymbol{Y}$ is related to $\boldsymbol{X}$ through some functional relationship whose form is specified up to a finite set of unknown parameters. The estimation of these parameters becomes the primary problem. The variance function for a standard homoscedastic regression model is constant, however for heteroscedastic regression models, variance is not constant. Heteroscedastic regression models are accepted as appropriate in a wide variety of fields. Statisticians often look at residual plots to investigate whether the model is heteroscedastic or not and what the appropriate behaviour of the variance function is. However, such subjective judging by eye may be quite misleading.

The variance estimates are needed for better understanding of the variability in the data. Also, the better the estimates of the variances, the better the estimates of the regression parameters. In some applications, estimation of the variances is of independent interest or plays an important theoretical role in estimation of quantities other than the regression parameters. Thus, while much effort has been focused on the study of the estimation for the regression parameters, it is of increasing interest to investigate the estimation of the variance function as well.

In heteroscedastic regression models, the variances may themselves be determined by a regression function. There are two main approaches to the estimation of this variance function: the parametric approach which assumes that this function follows a parametric (linear or nonlinear) model and the nonparametric approach which only assumes that this function is smooth.

The purpose of this thesis is to compare parametric and nonparametric methods for estimating the variance function in heteroscedastic regression models.

1.2 Model

Consider a general heteroscedastic regression model for observable data y_i given by

$$y_i = f(x_i, \beta) + e_i, \qquad i = 1, ..., N \tag{1.1}$$

where, f is an unknown mean response function, e_i are uncorrelated errors with zero mean and variance σ_i^2, x_i is a p - vector of predictors, β is a $p \times 1$ regression parameter and N

is the total sample size. The heteroscedasticity represented by non constant σ_i may be regarded as of unknown form or may be modeled as a function of the independent variable x, known factors exogenous to the model and the regression parameters. The variance function may be completely known, specified up to additional unknown parameters or completely unknown.

1.3 Background

Before embarking on any technical details of the two approaches, we give a brief review of the literature.

In variance function estimation, we try to understand the structure of the variances as a function of predictors which might include the means. Thus variance function estimation is a form of regression and just as for regression on means, there are plotting techniques useful for understanding simple structure. Other methods first eliminate the location effects by forming the residuals after an appropriate fit to the means. One then computes the estimate of the variance function assuming that the residuals are the responses and the means are all known.

Simple models relating predictors, which might include modeling the logarithm of the variances as linear in the predictors, have been proposed. One might also hypothesize that the standard deviations or the variances follow a linear model in predictors. Another possibility is to model the inverses of the variances or standard deviations as linear in predictors. There are other natural models for the variances that might consist of two components, one constant and one depending on the mean. Alternatively, one can see the standard deviation modeled empirically as say a quadratic function of the predictors.

Methods for estimating the variance function in the parametric approach can be classified into four rough categories. (*i*) the mean and variance model holds (*ii*) a weighted regression of transformations of absolute residuals on their expected values assuming that the errors are independent and identically distributed (*iii*) likelihood techniques (*iv*) replicated responses at each value of the predictor and use the sample variances.

These methods will be discussed in full details in Chapters 2 and 3.

Under the nonparametric approach, residuals, sample variances and difference schemes provide the basic building blocks. They can be improved by making neighbouring points share information. These, together with the techniques for improvement will be pursued in Chapters 4 and 5.

Finally, Chapter 6 is devoted to seeing how both parametric and nonparametric approaches work in practice. The question of how far is the nonparametric from the parametric variance model is considered and some final remarks included.

Chapter 2

PARAMETRIC APPROACH

2.1 Introduction

The assumption is that the variances are not constant according to model (1.1). The problem of heterogeneity may be attacked directly by specifying models for both the mean and the variance and in particular a variance model with unknown parameters which must be estimated. There are many ways of doing this. In this chapter we will endeavor to rework in detail the most common methods for estimating variance functions in regression and in particular the estimation of these unknown parameters.

A general parametric model for the variance can be written as

$$\sigma_i^2 = \sigma^2 g^2(z_i, \mu_i(\beta), \theta) \tag{2.1}$$

where σ is an unknown scale parameter, g is the variance function, z_i is a known vector possibly containing x_i, $\mu_i(\beta) = f(x_i, \beta)$ and θ is an unknown $r \times 1$ vector of parameters.

There are a number of graphical techniques which can be used in choosing the model to be fitted by letting the data reveal themselves. The unweighted least squares residual plot is most widely used, see Weisberg (1985) and Carroll and Ruppert (1988, p. 29-51).

If $z_i = x_i$, the variance depends on the predictors. The variance can also depend on the known mean $\mu_i(\beta)$ or on the estimated mean response $\mu_i(\hat{\beta})$.

In practice as well as for theoretical investigations, g is taken to be known and to satisfy appropriate smoothness conditions. In a model such as (2.1), estimation of the variances essentially reduces to the estimation of θ , since β will be estimated routinely and the final estimates of β and θ may be used to obtain a final estimate of σ. Thus investigations of the properties of variance estimators for (2.1) focus on properties of estimators for θ. In some applications, estimation of θ is not the only problem of interest. In chemical and biological assay problems, issues of prediction and calibration arise. In such problems, the estimator of θ plays a central role. In radioimmunoassay, the statistical properties of prediction intervals and constructs such as the minimum detectable concentration are highly dependent on how one estimates θ . In engineering quality improvement applications, an important goal is to discover the sources of variability. This can be obtained directly from the variance function estimate. The foregoing discussion indicate that there are numerous practical situations in which the choice of the method for estimating the variance function will be important. In the case of model (2.1), the choice is defined by how we choose to estimate the variance function g, and in particular, θ.

Many of the methods for estimation of θ that have been proposed in the literature are (possibly weighted) regression methods based on functions of either absolute residuals from the current regression fit or in the case of replication at each design point, sample standard deviations. Still other methods are joint estimation methods based on assumptions about the underlying distributions in which (σ, β, θ) are in principle estimated simultaneously. Considered here are methods which are simple or in common use. These methods include:

(i)	Maximum Likelihood	(ii)	Pseudo Likelihood
(iii)	Weighted Squared Residuals	(iv)	Weighted Absolute Residuals
(v)	Logarithm Method	(vi)	Restricted Maximum Likelihood
(vii)	Modified Maximum Likelihood	$(viii)$	Extended Quasi Likelihood
(ix)	Rodbard	(x)	Sadler-Smith

These methods are considered in detail in Section (2.2). The author has reworked and consolidated all the basic formulae.

2.2 Procedures for Estimating θ

There follows a full description of the form of the specific procedures for estimating θ that have been outlined under Section (2.1). These procedures will be looked into in detail in the univariate regression case.

2.2.1 Pseudo Likelihood Procedure

Gong and Samaniego (1981) use the terminology pseudo likelihood. This procedure makes no distributional assumptions but relies only on the basic mean model (1.1) and variance model (2.1). However its efficiency can be diminished by deviations from normality. Pseudo likelihood estimates of θ are based on pretending that the regression parameter β is known. Then estimate θ by maximizing the likelihood, assuming normality.

Pretend the data are normal, then write the likelihood as

$$L(\beta,\sigma^2,\theta/\boldsymbol{Y}) = \prod_{i=1}^{N}[2\pi[\sigma^2 g^2(z_i,\mu_i(\beta),\theta)]]^{-\frac{1}{2}}\exp[-\frac{1}{2}\frac{[y_i - f(x_i,\beta)]^2}{\sigma^2 g^2(z_i,\mu_i(\beta),\theta)}]$$

$$= (2\pi)^{-\frac{N}{2}}[\sigma^2 g^2(z_i,\mu_i(\beta),\theta)]^{-\frac{N}{2}}\exp[-\frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i,\beta)]^2}{\sigma^2 g^2(z_i,\mu_i(\beta),\theta)}]$$

Let

$$\ell = \log L(\beta,\sigma^2,\theta/\boldsymbol{Y})$$

Then

$$\ell = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\sigma^2 g^2(z_i,\mu_i(\beta),\theta) - \frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i,\beta)]^2}{\sigma^2 g^2(z_i,\mu_i(\beta),\theta)}$$

Now

$$\frac{\partial\ell}{\partial\beta} = 0$$

gives

$$\frac{1}{\sigma^2}\frac{\sum_{i=1}^{N}[[y_i - f(x_i,\beta)]\frac{\partial}{\partial\beta}f(x_i,\beta)]}{g^2(z_i,\mu_i(\beta),\theta)} = 0$$

and

$$\frac{\sum_{i=1}^{N}\frac{\partial}{\partial\beta}f(x_i,\beta)[y_i - f(x_i,\beta)]}{g^2(z_i,\mu_i(\beta),\hat{\theta})} = 0 \qquad (2.2)$$

Equation (2.2) provides the estimate for the regression parameter β for some estimated value of θ. To obtain this estimated value of θ we maximize the loglikelihood $\ell(\theta,\sigma,\hat{\beta}/\boldsymbol{Y})$

Where

$$L(\hat{\beta},\theta,\sigma^2/\boldsymbol{Y}) = \prod_{i=1}^{N}[2\pi\sigma^2 g^2(z_i,\mu_i(\hat{\beta}),\theta)]^{-\frac{1}{2}}\exp[-\frac{1}{2}\frac{[y_i - f(x_i,\hat{\beta})]^2}{\sigma^2 g^2(z_i,\mu_i(\hat{\beta}),\theta)}]$$

$$= (2\pi)^{-\frac{N}{2}}(\sigma^2)^{-\frac{N}{2}}[g^2(z_i,\mu_i(\hat{\beta}),\theta]^{-\frac{N}{2}}\exp[-\frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i,\hat{\beta})]^2}{\sigma^2 g^2(z_i,\mu_i(\hat{\beta}),\theta)}]$$

and $\hat{\beta}$ is obtained from equation (2.2).

Now write

$$\ell' = \log L(\hat{\beta},\theta,\sigma^2/\boldsymbol{Y})$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log[g^2(z_i,\mu_i(\hat{\beta}),\theta)] - \frac{1}{2}\frac{\sum_{i=1}^{N}[y_i - f(x_i,\hat{\beta})]^2}{\sigma^2 g^2(z_i,\mu_i(\hat{\beta}),\theta)}$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log[g^2(z_i,\mu_i(\hat{\beta}),\theta)] - \frac{1}{2}\frac{\sum_{i=1}^{N} r_i^2}{\sigma^2}$$

where

$$r_i = \frac{y_i - f(x_i,\hat{\beta})}{g(z_i,\mu_i(\hat{\beta}),\theta)}$$

Then

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{r_i^2}{\sigma^4} = 0$$

yields

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{N} r_i^2}{N} \tag{2.3}$$

Next, let $\frac{\partial \ell}{\partial \theta} = 0$. Then we have

$$\frac{-Ng(z_i,\mu_i(\hat{\beta}),\theta)\frac{\partial}{\partial\theta_i}g(z_i,\mu_i(\hat{\beta}),\theta)}{g^2(z_i,\mu_i(\hat{\beta}),\theta)} + \frac{\sum_{i=1}^{N}[y_i - f(x_i,\hat{\beta})]^2\frac{\partial}{\partial\theta_i}g(z_i,\mu_i(\hat{\beta}),\theta)}{\hat{\sigma}^2 g^3(z_i,\mu_i(\hat{\beta}),\theta)} = 0$$

or

$$\frac{-N\frac{\partial}{\partial\theta_i}g(z_i,\mu_i(\hat{\beta}),\theta)}{g(z_i,\mu_i(\hat{\beta}),\theta)}+\frac{\sum_{i=1}^{N}[y_i-f(x_i,\hat{\beta})]^2\frac{\partial}{\partial\theta_i}g(z_i,\mu_i(\hat{\beta}),\theta)}{\hat{\sigma}^2g^3(z_i,\mu_i(\hat{\beta}),\theta)}=0$$

Simplifying gives

$$-Nv_{\theta_i}+\sum_{i=1}^{N}\frac{r_i^2v_{\theta_i}}{\hat{\sigma}^2}=0$$

or

$$\frac{-N\hat{\sigma}^2v_{\theta_i}+\sum_{i=1}^{N}r_i^2v_{\theta_i}}{\hat{\sigma}^2}=0$$

where

$$v_{\theta_i}=\frac{\frac{\partial}{\partial\theta_i}g(z_i,\mu_i(\hat{\beta}),\theta)}{g(z_i,\mu_i(\hat{\beta}),\theta)}$$

Then $\hat{\theta}$ is obtained by solving

$$\sum_{i=1}^{N}(r_i^2-\hat{\sigma}^2)v_{\theta_i}=0 \tag{2.4}$$

While this procedure does not at first appear in this form to be based on a regression using squared residuals, examination of the estimating equations (2.3) and (2.4) for estimating σ and θ respectively show that they have the form of equations for weighted squared residuals.

2.2.2 Restricted Maximum Likelihood Procedure

One objection to the pseudo likelihood procedure is that no compensation is made for the loss of degrees of freedom associated with the estimation of β. The restricted maximum likelihood procedure is obtained by modifying pseudo likelihood to account for the effect of "leverage" and for correcting the degrees of freedom loss. To obtain the adjustment define the hat matrix

$$\boldsymbol{H}_{N\times N} = \boldsymbol{X}_{\star}(\boldsymbol{X}_{\star}^{T}\boldsymbol{X}_{\star})^{-1}\boldsymbol{X}_{\star}^{T}$$

with diagonal elements h_{ii} where $\mathbf{X}_{\star}$ is the $N \times p$ matrix where ith row is the transpose of the column vector

$$\frac{\frac{\partial}{\partial\beta} f(x_i, \hat{\beta})}{g(z_i, \mu_i(\hat{\beta}), \hat{\theta})}$$

The diagonal elements h_{ii} are the leverage values. Then using these leverage values and changing the divisor of (2.3) to $N - p$ where p is the number of regression parameters, we solve for θ and σ equations

$$\sum_{i=1}^{N} \frac{[y_i - f(x_i, \hat{\beta})]^2}{\sigma^2 g^2(z_i, \mu_i(\hat{\beta}), \theta)} \begin{bmatrix} 1 \\ v_{\theta_i} \end{bmatrix} = \begin{bmatrix} N - p \\ \sum_{i=1}^{N}[v_{\theta_i}(1 - h_{ii})] \end{bmatrix}$$

obtaining

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{N} r_i^2}{N - p} \tag{2.5}$$

and

$$\sum_{i=1}^{N} \frac{[y_i - f(x_i, \hat{\beta})]^2}{\hat{\sigma^2} g^2(z_i, \mu_i(\hat{\beta}), \theta)} v_{\theta_i} = \sum_{i=1}^{N}[v_{\theta_i}(1 - h_{ii})]$$

Finally $\hat{\theta}$ is obtained from

$$\sum_{i=1}^{N} r_i^2 v_{\theta_i} - \hat{\sigma^2} \sum_{i=1}^{N}[v_{\theta_i}(1 - h_{ii})] = 0 \tag{2.6}$$

Where $\hat{\sigma^2}$ is as in equation (2.5) and r_i and v_{θ_i} are as defined in Subsection (2.2.1).

2.2.3 Least Squares on Squared Residuals Procedure

The motivating idea in this procedure is that the expectation of the squared residuals is approximately the variance. From equation (1.1), write the squared residuals as

$$[y_i - f(x_i, \hat{\beta}_\star)]^2$$

where $\hat{\beta}_\star$ is the current estimate of β. We consider a regression problem where the responses are squared residuals and the regression function is its approximate expectation $\sigma^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)$. Thus write

$$E[y_i - f(x_i, \hat{\beta}_\star)]^2 \simeq \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)$$

Then minimize

$$\sum_{i=1}^{N}[(y_i - f(x_i, \hat{\beta}_\star))^2 - \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2$$

in σ and θ. However, for normally distributed data the squared residuals are themselves heteroscedastic with variance approximately proportional to

$$\sigma^4 g^4(z_i, \mu_i(\beta), \theta)$$

Thus one is naturally led to generalized least squares, see Jobson and Fuller (1980). Thus, for generalized least squares, this suggests minimizing with respect to σ and θ the weighted least squares version

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2}{g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)}$$

obtaining

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \sigma^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})]\sigma g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})}{g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)} = 0 \tag{2.7}$$

and

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \hat{\sigma}^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)]g(z_i, \mu_i(\hat{\beta}_\star), \theta)\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}_\star), \theta)}{g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)} = 0 \tag{2.8}$$

where $\hat{\theta}_\star$ is the current estimate of θ. Now solve equations (2.7) and (2.8) to get $\hat{\sigma}$ and $\hat{\theta}$ respectively. Next, to account for the effect of leverage minimize

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \sigma^2(1 - h_{ii}) g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2}{[(1 - h_{ii})^2 g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]}$$

in σ and θ obtaining

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \sigma^2(1 - h_{ii})^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})]g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})}{[(1 - h_{ii})^2 g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]} = 0 \tag{2.9}$$

and

$$\sum_{i=1}^{N} \frac{[(y_i - f(x_i, \hat{\beta}_\star))^2 - \hat{\sigma}^2(1 - h_{ii})^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \theta)]g(z_i, \mu_i(\hat{\beta}_\star), \theta)\frac{\partial g(z_i, \mu_i(\hat{\beta}_\star), \theta)}{\partial \theta_i}}{[(1 - h_{ii})^2 g^4(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]} = 0 \tag{2.10}$$

with h_{ii} as defined under Subsection (2.2.2). Solving equations (2.9) and (2.10) gives $\hat{\sigma}$ and $\hat{\theta}$ respectively. This is done iteratively and the procedure is outlined in Chapter 3.

2.2.4 Least Squares on Absolute Residuals Procedure

In analogy to Subsection (2.2.3), we look at the expectation of the absolute residuals as being approximately the standard deviation $\sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)$ and write

$$E|y_i - f(x_i, \hat{\beta}_\star)| = \sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)$$

leading to the minimization of

$$\sum_{i=1}^{N}[|y_i - f(x_i, \hat{\beta}_\star)| - \sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2$$

with respect to σ and θ. However, since the residuals are to be appropriately weighted, Carroll and Ruppert (1988), suggest estimating θ by minimizing with respect to σ and θ the weighted version namely

$$\sum_{i=1}^{N}\frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2}{g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)}$$

This implies that, to find $\hat{\sigma}$ we solve

$$\sum_{i=1}^{N}\frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \sigma g(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})]g(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})}{g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)} = 0 \qquad (2.11)$$

and for $\hat{\theta}$ solve

$$\sum_{i=1}^{N} \frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \hat{\sigma} g(z_i, \mu_i(\hat{\beta}_\star), \theta)] \hat{\sigma} g(z_i, \mu_i(\hat{\beta}_\star), \theta) \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}_\star), \theta)}{g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)} = 0$$

This can be written as

$$\sum_{i=1}^{N} \frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \hat{\sigma} g(z_i, \mu_i(\hat{\beta}_\star), \theta)] g(z_i, \mu_i(\hat{\beta}_\star), \theta) \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}_\star), \theta)}{g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)} = 0 \qquad (2.12)$$

where $\hat{\sigma}$ is known from (2.11).

Next we modify this procedure to account for the effect of leverage by minimizing with respect to σ and θ

$$\sum_{i=1}^{N} \frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \sigma(1 - h_{ii}) g(z_i, \mu_i(\hat{\beta}_\star), \theta)]^2}{[(1 - h_{ii})^2 g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]}$$

Differentiate with respect to σ and θ respectively and equate to zero obtaining equations

$$\sum_{i=1}^{N} \frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \sigma(1 - h_{ii}) g(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})] g(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta})}{[(1 - h_{ii}) g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]} = 0 \qquad (2.13)$$

and

$$\sum_{i=1}^{N} \frac{[|y_i - f(x_i, \hat{\beta}_\star)| - \hat{\sigma}(1 - h_{ii}) g(z_i, \mu_i(\hat{\beta}_\star), \theta)] g(z_i, \mu_i(\hat{\beta}_\star), \theta) \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}_\star), \theta)}{[(1 - h_{ii}) g^2(z_i, \mu_i(\hat{\beta}_\star), \hat{\theta}_\star)]} = 0 \quad (2.14)$$

2.2.5 Modified Maximum Likelihood Procedure

This is the problem of estimating variances by pooling information from a large number of small samples. At each predictor value x_i, observe m_i replicated responses y_{ij}, $i = 1, ..., M$ and $j = 1, ..., m_i$. First consider the case of equal replications $m_i = m$ so that $M = Nm$ is the total number of observations. Raab (1981, p. 35) suggest the modification of the standard likelihood replacing the term $\sigma^{-\frac{m_i}{2}}$ by $\sigma^{-(\frac{m_i-1}{2})}$ and gives a number of justifications. We adopt this modified likelihood and write

$$L(\beta, \theta, \sigma^2/\boldsymbol{Y}) = \prod_{i=1}^{M}[2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta)]^{-(\frac{m-1}{2})}\exp[-\sum_{j=1}^{m}\frac{(y_{ij} - \mu_i(\beta))^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}]$$

Taking logarithm obtain

$$\ell = \log L(\beta, \theta, \sigma^2/\boldsymbol{Y}) = -(\frac{m-1}{2})\sum_{i=1}^{M}\log[2\pi\sigma^2 g^2(z_i, \mu_i(\beta), \theta)] - \sum_{i=1}^{M}\sum_{j=1}^{m}\frac{(y_{ij} - \mu_i(\beta))^2}{2\sigma^2 g^2(z_i, \mu_i(\beta), \theta)}$$

Now differentiate with respect to σ^2 and equate to zero, to obtain

$$\frac{\partial \ell}{\partial \sigma^2} = -(\frac{m-1}{2})\sum_{i=1}^{M}\frac{1}{\sigma^2} + \sum_{i=1}^{M}\sum_{j=1}^{m}\frac{(y_{ij} - \mu_i(\beta))^2}{2\sigma^4 g^2(z_i, \mu_i(\beta), \theta)}$$

giving

$$\frac{1}{2\sigma^2}\sum_{i=1}^{M}(m-1) = \sum_{i=1}^{M}\sum_{j=1}^{m}\frac{(y_{ij} - \mu_i(\beta))^2}{2\sigma^4 g^2(z_i, \mu_i(\beta), \theta)}$$

or

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{m}\frac{(y_{ij}-\mu_i(\hat{\beta}))^2}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})}}{\sum_{i=1}^{M}(m-1)} \tag{2.15}$$

Similarly, differentiate ℓ with respect to θ and equate to zero obtaining

$$\frac{\partial \ell}{\partial \theta} = -(m-1)\sum_{i=1}^{M} \frac{\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\beta), \theta)}{g(z_i, \mu_i(\beta), \theta)} + \sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\beta))^2 \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\beta), \theta)}{\sigma^2 g^3(z_i, \mu_i(\beta), \theta)}$$

or

$$\sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\hat{\beta}))^2 \frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{\hat{\sigma^2} g^3(z_i, \mu_i(\hat{\beta}), \theta)} - (m-1)\sum_{i=1}^{M} \frac{\frac{\partial}{\partial \theta_i} g(z_i, \mu_i(\hat{\beta}), \theta)}{g(z_i, \mu_i(\hat{\beta}), \theta)} = 0 \qquad (2.16)$$

Finally differentiate ℓ with respect to β and equate to zero getting

$$\frac{\partial \ell}{\partial \beta} = -(m-1)\sum_{i=1}^{M} \frac{\frac{\partial g(z_i, \mu_i(\beta), \theta)}{\partial \beta}}{g(z_i, \mu_i(\beta), \theta)} + \sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\beta))}{\sigma^2 g^2(z_i, \mu_i(\beta), \theta)} + \sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\beta))^2 \frac{\partial g(z_i, \mu_i(\beta), \theta)}{\partial \beta}}{\sigma^2 g^3(z_i, \mu_i(\beta), \theta)}$$

$$\Rightarrow \sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\beta))}{\hat{\sigma^2} g^2(z_i, \mu_i(\beta), \hat{\theta})} + \sum_{i=1}^{M}\sum_{j=1}^{m} \frac{(y_{ij} - \mu_i(\beta))^2 \frac{\partial g(z_i, \mu_i(\beta), \hat{\theta})}{\partial \beta}}{\hat{\sigma^2} g^3(z_i, \mu_i(\beta), \hat{\theta})} - (m-1)\sum_{i=1}^{M} \frac{\frac{\partial g(z_i, \mu_i(\beta), \hat{\theta})}{\partial \beta}}{g(z_i, \mu_i(\beta), \hat{\theta})} = 0 \qquad (2.17)$$

Now solve equations (2.15), (2.16) and (2.17) to obtain $\hat{\sigma^2}$, $\hat{\theta}$ and $\hat{\beta}$ respectively. Note that for the usual maximum likelihood estimate, σ is biased. It is made unbiased in this case by dividing the corrected sum of squares by the degrees of freedom rather than the sample size.

The unequal replication case is given in Subsection (3.2.6).

2.2.6 Extended Quasi Likelihood Procedure

Wedderburn (1974) gives the definition of quasi likelihood while Nelder and Pregibon

(1987) discuss the extended quasi likelihood. When θ is known and the variance function has the form (2.1), quasi likelihood estimation of β is a form of iterated generalized least squares. The extended quasi likelihood method is a joint estimation scheme which attempts to extend the notion of quasi likelihood to include estimation of θ. The method is based on the assumption that the data arise from a class of distributions depending on θ and involves estimation of θ by minimizing with respect to β, θ and σ^2 the extended quasi likelihood

$$Q^+ = -\frac{1}{2}\sum_{i=1}^{N}[\log\{2\pi\sigma^2 g^2(z_i, y_i, \theta)\} - \frac{2}{\sigma^2}\int_{y_i}^{\mu_i(\beta)} \frac{y_i - u}{g^2(z_i, \mu_i(\beta), \theta)} du]$$

as in Davidian (1986, p. 16). If we differentiate Q^+ with respect to θ and σ^2 we obtain

$$\frac{\partial Q^+}{\partial \theta} = -\frac{1}{2}\sum_{i=1}^{N}[2\frac{\frac{\partial}{\partial\theta_i}g(z_i, y_i, \theta)}{g(z_i, y_i, \theta)} + \frac{4}{\sigma^2}\int_{y_i}^{\mu_i(\beta)}\{\frac{y_i - u}{g^3(z_i, \mu_i(\beta), \theta)}\}\{\frac{\partial}{\partial\theta_i}g(z_i, \mu_i(\beta), \theta)\}du]$$

and

$$\frac{\partial Q^+}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^{N}[\frac{1}{\sigma^2} + \frac{2}{\sigma^4}\int_{y_i}^{\mu_i(\beta)} \frac{y_i - u}{g^2(z_i, \mu_i(\beta), \theta)} du]$$

Equating the derivatives to zero gives $\hat{\theta}$ as the solution of

$$\sum_{i=1}^{N}[2\frac{\frac{\partial}{\partial\theta_i}g(z_i, y_i, \theta)}{g(z_i, y_i, \theta)} + \frac{4}{\hat{\sigma}^2}\int_{y_i}^{\mu_i(\hat{\beta})}\{\frac{y_i - u}{g^3(z_i, \mu_i(\hat{\beta}), \theta)}\}\{\frac{\partial}{\partial\theta_i}g(z_i, \mu_i(\hat{\beta}), \theta)\}du] = 0 \qquad (2.18)$$

and for $\hat{\sigma^2}$ we solve

$$\sum_{i=1}^{N}[\frac{1}{\sigma^2} + \frac{2}{\sigma^4}\int_{y_i}^{\mu_i(\hat{\beta})} \frac{y_i - u}{g^2(z_i, \mu_i(\hat{\beta}), \hat{\theta})} du] = 0 \qquad (2.19)$$

Similarly, differenting Q^+ with respect to β gives

$$\sum_{i=1}^{N}\left(\frac{y_i - \mu_i(\beta)}{g^2(z_i, \mu_i(\beta), \hat{\theta})}\right)\left(\frac{\partial \mu_i(\beta)}{\partial \beta}\right) = 0 \tag{2.20}$$

2.2.7 Logarithm of Absolute Residuals Procedure

This procedure exploits the fact that

$$E|y_i - f(x_i, \mu_i(\hat{\beta}_\star))| \simeq \sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)$$

and uses a two-step estimation process. The first step consists of taking the natural logarithms of the absolute residuals $|y_i - f(x_i, \mu_i(\hat{\beta}_\star))|$. These are then regressed on

$$\log\{\sigma g(z_i, \mu_i(\hat{\beta}_\star), \theta)\}$$

thereby yielding estimates of θ as the slope and $\log \sigma^2$ as the intercept. With the assumption that the errors are independent and identically distributed, this should be approximately a homoscedastic regression. A practical problem arises if one of the residuals is very near zero, in which case taking logarithms induces a rather large and artificial outlier. To avoid this potential difficulty for fitting the variance model, Carroll and Ruppert (1988) suggest that one might wish to delete a few of the smallest absolute residuals.

2.2.8 Rodbard and Frazier Procedure

This method uses replication as in the case of modified maximum likelihood and it is identical to the logarithm method. The idea is to avoid dependence on unweighted methods. Here the absolute residual is replaced by the sample standard deviation and $f(x_i, \hat{\beta}_\star)$ in the regression function is replaced by the sample mean $\bar{y}_i$. Thus the procedure is to regress the logarithm of the sample standard deviation on the logarithm of the sample mean.

2.2.9 Other Procedures

Two other procedures are as follows. First is the maximum likelihood procedure. The process here is the same as in the pseudo likelihood procedure. However instead of fixing β at the current value $\hat{\beta}$ and maximizing the likelihood function in θ, one maximizes the likelihood function jointly in β and θ. Maximum likelihood assumes that the variances do not depend on the mean. The second procedure is that of Sadler and Smith (1985). This is similar to the modified maximum likelihood procedure where one uses the sample mean $\bar{y}_i$ instead of μ_i.

2.3 Multiple Regression Model

The methodology discussed under the univariate regression model in Section (2.2) can be carried over to the multiple regression model without little alteration. For illustration, in this section we briefly discuss two methods; pseudo likelihood and restricted maximum likelihood. The other methods can be worked out in similar manner.

In multiple regression, several predictors are used to model a single response variable. The data can be depicted in an array as

$\mathbf{Y}$	$\mathbf{X_1}$	$\mathbf{X_2}$	$\cdots$	$\mathbf{X_p}$
y_1	x_{11}	x_{12}	$\cdots$	x_{1p}
y_2	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_N	x_{N1}	x_{N2}	$\cdots$	x_{Np}

The model is specified by a linear equation

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{X}_1 + \beta_2 \boldsymbol{X}_2 + \cdots + \beta_p \boldsymbol{X}_p + \boldsymbol{e} \tag{2.21}$$

and in matrix notation we write (2.21) as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{2.22}$$

The mean and variance models are written as

$$E(\boldsymbol{Y}) = \boldsymbol{\mu}(\boldsymbol{\beta}) = f(\boldsymbol{X}, \boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta}$$

and

$$\text{var}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{\Lambda} = \sigma^2 \boldsymbol{\Lambda}(\boldsymbol{Z}, \boldsymbol{\mu}(\boldsymbol{\beta}), \boldsymbol{\theta})$$

respectively. For simplicity we consider the uncorrelated case, thus $\boldsymbol{\Lambda}$ is a diagonal matrix. Now proceed to estimate the regression parameters as under Subsection (2.3.1).

2.3.1 Estimation of β

Now we assume normality and write the likelihood as

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Lambda}/\boldsymbol{Y}) = (2\pi\sigma^2)^{-\frac{N}{2}}|\boldsymbol{\Lambda}|^{-\frac{1}{2}}\exp\{-\frac{1}{2\sigma^2}[\boldsymbol{Y} - f(\boldsymbol{X}, \boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \boldsymbol{\beta})]\}$$

and let

$$\ell = \log L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Lambda}/\boldsymbol{Y})$$

$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\boldsymbol{\Lambda}| - \frac{1}{2\sigma^2}[\mathbf{Y} - f(\boldsymbol{X}, \boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}[\mathbf{Y} - f(\boldsymbol{X}, \boldsymbol{\beta})]$$

Then $\frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0$. The following sequence of steps derives the basic matrix version of the likelihood equations

$$-\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{Y}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y} - \boldsymbol{Y}^T\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X}, \boldsymbol{\beta}) - [f(\boldsymbol{X}, \boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}\boldsymbol{Y} + [f(\boldsymbol{X}, \boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X}, \boldsymbol{\beta})) = 0$$

which simplifies to

$$-\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{Y}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X}, \boldsymbol{\beta}) - \boldsymbol{Y}^T\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X}, \boldsymbol{\beta} + [f(\boldsymbol{X}, \boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X}, \boldsymbol{\beta})) = 0$$

or

$$-\frac{1}{2\sigma^2}(-2\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}\frac{\partial}{\partial\boldsymbol{\beta}}f(\boldsymbol{X},\boldsymbol{\beta}) + 2\boldsymbol{\Lambda}^{-1}f(\boldsymbol{X},\boldsymbol{\beta})\frac{\partial}{\partial\boldsymbol{\beta}}f(\boldsymbol{X},\boldsymbol{\beta})) = 0$$

Now factorize and obtain

$$\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}\frac{\partial}{\partial\boldsymbol{\beta}}f(\boldsymbol{X},\boldsymbol{\beta}) - \boldsymbol{\Lambda}^{-1}f(\boldsymbol{X},\boldsymbol{\beta})\frac{\partial}{\partial\boldsymbol{\beta}}f(\boldsymbol{X},\boldsymbol{\beta}) = 0$$

or

$$[\frac{\partial}{\partial\boldsymbol{\beta}}f(\boldsymbol{X},\boldsymbol{\beta})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X},\boldsymbol{\beta})] = 0$$

and finally

$$\boldsymbol{D}^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})] = 0 \tag{2.23}$$

where $\boldsymbol{D} = N \times p$ matrix of partial derivatives of $f(\boldsymbol{X},\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}^{-1}$ is the inverse of the variance matrix $\boldsymbol{\Lambda}(\boldsymbol{Z},\boldsymbol{\mu}(\boldsymbol{\beta}),\boldsymbol{\theta})$. Now in the special case when $\boldsymbol{\mu}(\boldsymbol{\beta}) = \boldsymbol{X\beta}$, (2.23) becomes $\boldsymbol{D}^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{X\beta}] = 0$ giving the standard form

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{D}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{D}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{Y} \tag{2.24}$$

To obtain matrix $\boldsymbol{\Lambda}^{-1}$ proceed as in Subsection (2.3.2).

2.3.2 Pseudo likelihood Estimation of θ

We maximize the loglikelihood $L(\boldsymbol{\theta}, \sigma^2, \hat{\boldsymbol{\beta}}/\boldsymbol{Y})$ where

$$L(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}, \sigma^2/\boldsymbol{Y}) = (2\pi)^{-\frac{N}{2}}(\sigma^2)^{-\frac{N}{2}}|\boldsymbol{\Lambda}|^{-\frac{1}{2}}\exp\{-\frac{1}{2\sigma^2}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]\}$$

Let

$$\ell = \log L(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}, \sigma^2/\boldsymbol{Y})$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2}\log|\boldsymbol{\Lambda}| - \frac{1}{2\sigma^2}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]$$

Then

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})] = 0$$

or

$$\frac{1}{2\sigma^4}\{-N\sigma^2 + [\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]\} = 0$$

giving

$$\hat{\sigma^2} = \frac{1}{N}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})] \tag{2.25}$$

Next, write $\frac{\partial \ell}{\partial \boldsymbol{\theta}} = 0$. Then use the following results of matrix differentiation (see for example Theil (1971))

$$\frac{\partial}{\partial \boldsymbol{X}}(\boldsymbol{a}^T\boldsymbol{X}^{-1}\boldsymbol{a}) = -\boldsymbol{X}^{-1}\boldsymbol{a}\boldsymbol{a}^T\boldsymbol{X}^{-1} \tag{2.26}$$

and

$$\frac{\partial}{\partial \boldsymbol{A}}\log|\boldsymbol{A}| = (\boldsymbol{A}^T)^{-1} \tag{2.27}$$

to obtain the partial derivative of ℓ with respect to θ. Thus

$$\frac{\partial \ell}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda} - \frac{1}{2\hat{\sigma}^2}[-\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})][\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda}] = 0$$

and

$$\frac{1}{\hat{\sigma}^2}[\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})][\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda}] - \boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda} = 0$$

giving

$$\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})][\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda} - \hat{\sigma}^2\boldsymbol{\Lambda}^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\Lambda} = 0$$

Thus to find $\hat{\theta}$ we solve the equations

$$\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - f(\boldsymbol{X},\hat{\boldsymbol{\beta}})][\boldsymbol{Y} - f(X,\hat{\boldsymbol{\beta}})]^T\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}} - \hat{\sigma}^2\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}} = 0$$

which can be written as

$$\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}][\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}]^T\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}} - \frac{1}{N}[\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}]\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}} = 0 \qquad (2.28)$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose elements are

$$\sum_{j=1}^{p}\sigma^2 g_i^2(z_{ij}, \mu_i(\beta), \theta), \qquad i = 1, ..., N$$

and $\dot{\boldsymbol{\Lambda}}$ is a matrix of partial derivatives of $\boldsymbol{\Lambda}$ with respect to $\boldsymbol{\theta}$.

2.3.3 Restricted Maximum Likelihood Estimation of θ

To correct leverage and degrees of freedom lost due to estimation of $\boldsymbol{\beta}$, let

$$[\boldsymbol{D}^T \boldsymbol{\Lambda}^{-1}]^T = \boldsymbol{A}$$

and compute the hat matrix $\boldsymbol{H}$ as $\boldsymbol{H} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$. Further, let $\bar{\boldsymbol{\Lambda}}_{N\times N}$ be the diagonal matrix whose elements are the main diagonal of $(\boldsymbol{I} - \boldsymbol{H})$, where $\boldsymbol{I}$ is an $N \times N$ identity matrix. Then, for $\hat{\boldsymbol{\theta}}$ solve

$$\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}][\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}]^T\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}} - \frac{1}{N-p}[\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}]^T\boldsymbol{\Lambda}^{-1}[\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}]\boldsymbol{\Lambda}^{-1}\dot{\boldsymbol{\Lambda}}\bar{\boldsymbol{\Lambda}} = 0 \quad (2.29)$$

where $\dot{\boldsymbol{\Lambda}}$ is as defined in equation (2.28).

2.4 Generalized Multivariate Regression Model

In this section we shall limit our study to the restricted maximum likelihood only since all the other methods will follow similar computation. Let $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_p$ be $N \times 1$ vectors representing N independent observations. We assume a linear model of the form

$$\boldsymbol{Y} = \mathbf{XB} + \mathbf{E} \quad (2.30)$$

where $\boldsymbol{Y}$ is a $N \times p$ matrix of response observations, $\boldsymbol{X}$ is a $N \times q$ matrix of predictors, $\boldsymbol{B}$ is a $q \times p$ matrix of unknown regression parameters and $\boldsymbol{E}$ is a $N \times p$ matrix of random variables. The parameter vectors $\boldsymbol{\beta}_i$ which are the components of $\boldsymbol{B}$ are specific to the chosen dependent variable $\boldsymbol{Y}_i$ and hence are different for each i, but the same $\boldsymbol{X}$ appears in all models

$$\boldsymbol{Y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{e}_i \tag{2.31}$$

This is because the same explanatory variables are being used in each separate predictive model. Hence the multiple regression model (2.31) is obtained by extracting the ith column of each matrix in the multivariate regression model (2.30). See that there are p univariate regression equations in (2.31) of the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i \tag{2.32}$$

where $\boldsymbol{Y}_i$ is a $N \times 1$ vector of observations, $\boldsymbol{X}_i$ is a $N \times q_i$ design matrix in the ith equation, $\boldsymbol{\beta}_i$ is a $q_i \times 1$ vector of unknown regression parameters and $\boldsymbol{e}_i$ is a $N \times 1$ vector of random variables. Note that if $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \cdots = \boldsymbol{X}_p$, then (2.32) reduces to (2.31). Now write (2.32) in compact and seemingly univariate form by defining $\ddot{\boldsymbol{Y}} = (\boldsymbol{Y}_1^T, \cdots, \boldsymbol{Y}_p^T)^T$, $\ddot{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_p)^T$, $\ddot{\boldsymbol{e}} = (\boldsymbol{e}_1^T, \cdots, \boldsymbol{e}_p^T)^T$ and $\ddot{\boldsymbol{X}}$ is a diagonal matrix whose elements are $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p$. This gives

$$\ddot{\boldsymbol{Y}} = \ddot{\boldsymbol{X}}\ddot{\boldsymbol{\beta}} + \ddot{\boldsymbol{e}} \tag{2.33}$$

where $\ddot{\boldsymbol{Y}}$ is a $Np \times 1$ matrix of observations, $\ddot{\boldsymbol{X}}$ is a $Np \times q$ block diagonal matrix of predictors, $\ddot{\boldsymbol{\beta}}$ is a $q \times 1$ matrix of regression parameters, $\ddot{\boldsymbol{e}}$ is a $Np \times 1$ matrix of random variables and $q = \sum_{i=1}^{p} q_i$. Observe that (2.33) looks exactly like (2.22) and therefore leads to

$$\hat{\ddot{\boldsymbol{\beta}}} = [\ddot{\boldsymbol{D}}^T \hat{\ddot{\boldsymbol{\Lambda}}}^{-1} \ddot{\boldsymbol{X}}]^{-1} \ddot{\boldsymbol{D}}^T \hat{\ddot{\boldsymbol{\Lambda}}}^{-1} \ddot{\boldsymbol{Y}} \tag{2.34}$$

as the equation for the regression parameters and

$$\ddot{\boldsymbol{\Lambda}}^{-1}[\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}\hat{\ddot{\boldsymbol{\beta}}}][\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}\hat{\ddot{\boldsymbol{\beta}}}]^T \ddot{\boldsymbol{\Lambda}}^{-1} \tilde{\boldsymbol{\Lambda}} - \frac{1}{Np-q}[\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}\hat{\ddot{\boldsymbol{\beta}}}]^T \ddot{\boldsymbol{\Lambda}}^{-1}[\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}\hat{\ddot{\boldsymbol{\beta}}}]\ddot{\boldsymbol{\Lambda}}^{-1}\tilde{\boldsymbol{\Lambda}}\bar{\ddot{\boldsymbol{\Lambda}}} = 0 \tag{2.35}$$

as the equation for $\hat{\ddot{\boldsymbol{\theta}}}$. Here $\ddot{\boldsymbol{\Lambda}} = Np \times Np$ diagonal matrix whose elements are

$$[\sum_{j=1}^{q_1} \sigma^2 g_i^2(z_{ij}^{(1)}, \mu_i^{(1)}(\beta), \theta^{(1)}), \cdots, \sum_{j=1}^{q_p} \sigma^2 g_i^2(z_{ij}^{(p)}, \mu_i^{(p)}(\beta), \theta^{(p)})], \qquad i = 1, \cdots, N$$

and $\tilde{\boldsymbol{\Lambda}}$ = matrix of partial derivatives of $\ddot{\boldsymbol{\Lambda}}$ with respect to $\ddot{\boldsymbol{\theta}}$ where $\ddot{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_p^T)^T$, $\ddot{\boldsymbol{D}} = Np \times q$ block diagonal matrix of partial derivatives of $f(\ddot{\boldsymbol{X}}, \ddot{\boldsymbol{\beta}})$ with respect to $\ddot{\boldsymbol{\beta}}$, $\bar{\ddot{\boldsymbol{\Lambda}}} = (\boldsymbol{I}_{Np} - \ddot{\boldsymbol{H}})$, $\ddot{\boldsymbol{H}} = \ddot{\boldsymbol{A}}(\ddot{\boldsymbol{A}}^T \ddot{\boldsymbol{A}})^{-1} \ddot{\boldsymbol{A}}^T$ and $\ddot{\boldsymbol{A}} = [\ddot{\boldsymbol{D}}^T \ddot{\boldsymbol{\Lambda}}^{-1}]^T$.

2.5 Conclusion

Different special cases of (2.1) have led to different procedures for estimation. All the basic equations have been rederived in a unifying framework.

Chapter 3

PERFORMANCE OF THE PARAMETRIC ESTIMATORS

3.1 Introduction

The procedures of Chapter 2 will be empirically studied. The objective is to have working experience with these procedures. In addition to the working assumptions such as normality there are other features that appear. For example the number of replications at each design point need not be constant. Also there is the problem of deleting a few absolute residuals that are very near zero in the case of logarithm of absolute residuals, for example choosing how many points to delete.

Using simulation, simple comparison is carried out for checking the similarity of these procedures if all conditions associated with them are met. For computational convenience, simple models for both mean and variance will be considered. Computation is carried out in S-PLUS programming language. The equations derived under Chapter 2 are general equations for any variance function encountered in practice. In order to carry out the simulations, these equations will be revisited once more to show the actual mathematical calculations involved for a simple model of univariate regression.

3.2 Working Model

Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \,, \qquad i = 1, \cdots, N \tag{3.1}$$

and take a simple variance function (for example, Carroll and Ruppert (1988, p. 65))

$$\mathrm{var}(y_i) = \sigma^2 g^2(x_i, \theta) = \sigma^2 (1 + \theta x_i^2)^2 \tag{3.2}$$

In this model $x = 0$ is the minimum of g. Now insert (3.1) and (3.2) into the equations of Chapter 2 as follows

3.2.1 Equation for $\hat{\beta}$

From equation (3.1), write

$$E(y_i) = \mu_i(\beta) = f(x_i, \beta) = \beta_0 + \beta_1 x_i$$

and together with (3.2) substitute into equation (2.2) to have

$$\sum_{i=1}^{N} \frac{\partial}{\partial \beta_j} f(x_i, \beta)[\frac{y_i - f(x_i, \beta)}{\sigma^2(1 + \hat{\theta} x_i^2)^2}] = 0 \,, \qquad j = 0, 1$$

giving

$$\sum_{i=1}^{N} \begin{bmatrix} \frac{\partial}{\partial \beta_0} f(x_i, \beta) \\ \frac{\partial}{\partial \beta_1} f(x_i, \beta) \end{bmatrix} [\frac{y_i - f(x_i, \beta)}{(1 + \hat{\theta} x_i^2)^2}] = 0$$

and

$$\sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_i \end{bmatrix} [\frac{y_i - f(x_i, \beta)}{(1 + \hat{\theta} x_i^2)^2}] = 0$$

Expand to obtain

$$\sum_{i=1}^{N} \frac{y_i - (\beta_0 + \beta_1 x_i)}{(1 + \hat{\theta} x_i^2)^2} = 0 \tag{3.3}$$

and

$$\sum_{i=1}^{N} \frac{y_i x_i - (\beta_0 x_i + \beta_1 x_i^2)}{(1 + \hat{\theta} x_i^2)^2} = 0 \tag{3.4}$$

Now write equation (3.3) as

$$\sum_{i=1}^{N} \{\frac{y_i}{(1 + \hat{\theta} x_i^2)^2}\} - \sum_{i=1}^{N} \{\frac{\beta_0 + \beta_1 x_i}{(1 + \hat{\theta} x_i^2)^2}\} = 0$$

or

$$\beta_0 \sum_{i=1}^{N} \{\frac{1}{(1 + \hat{\theta} x_i^2)^2}\} + \beta_1 \sum_{i=1}^{N} \{\frac{x_i}{(1 + \hat{\theta} x_i^2)^2}\} = \sum_{i=1}^{N} \{\frac{y_i}{(1 + \hat{\theta} x_i^2)^2}\} \tag{3.5}$$

Similarly, write equation (3.4) as

$$\beta_0 \sum_{i=1}^{N} \{ \frac{x_i}{(1+\hat{\theta} x_i^2)^2} \} + \beta_1 \sum_{i=1}^{N} \{ \frac{x_i^2}{(1+\hat{\theta} x_i^2)^2} \} = \sum_{i=1}^{N} \{ \frac{x_i y_i}{(1+\hat{\theta} x_i^2)^2} \} \tag{3.6}$$

In matrix form we have

$$\begin{bmatrix} \sum_{i=1}^{N} \{ \frac{1}{(1+\hat{\theta} x_i^2)^2} \} & \sum_{i=1}^{N} \{ \frac{x_i}{(1+\hat{\theta} x_i^2)^2} \} \\ \sum_{i=1}^{N} \{ \frac{x_i}{(1+\hat{\theta} x_i^2)^2} \} & \sum_{i=1}^{N} \{ \frac{x_i^2}{(1+\hat{\theta} x_i^2)^2} \} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \{ \frac{y_i}{(1+\hat{\theta} x_i^2)^2} \} \\ \sum_{i=1}^{N} \{ \frac{x_i y_i}{(1+\hat{\theta} x_i^2)^2} \} \end{bmatrix}$$

Next, solve to obtain

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \{ \frac{1}{(1+\hat{\theta} x_i^2)^2} \} & \sum_{i=1}^{N} \{ \frac{x_i}{(1+\hat{\theta} x_i^2)^2} \} \\ \sum_{i=1}^{N} \{ \frac{x_i}{(1+\hat{\theta} x_i^2)^2} \} & \sum_{i=1}^{N} \{ \frac{x_i^2}{(1+\hat{\theta} x_i^2)^2} \} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} \{ \frac{y_i}{(1+\hat{\theta} x_i^2)^2} \} \\ \sum_{i=1}^{N} \{ \frac{x_i y_i}{(1+\hat{\theta} x_i^2)^2} \} \end{bmatrix} \tag{3.7}$$

3.2.2 Pseudo Likelihood Equation for $\hat{\theta}$

From equation (2.4), obtain $\hat{\theta}$ as

$$\sum_{i=1}^{N} [\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\theta x_i^2}]^2 \frac{x_i^2}{1+\theta x_i^2} - \frac{1}{N} \sum_{i=1}^{N} [\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\theta x_i^2}]^2 \sum_{i=1}^{N} \frac{x_i^2}{1+\theta x_i^2} = 0 \tag{3.8}$$

and from equation (2.3) obtain

$$\hat{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} [\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\hat{\theta} x_i^2}]^2 \tag{3.9}$$

3.2.3 Restricted Maximum Likelihood Equation for $\hat{\theta}$

From equation (2.6), obtain $\hat{\theta}$ as

$$\sum_{i=1}^{N}[\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\theta x_i}]^2 \frac{x_i^2}{1+\theta x_i^2} - \frac{1}{N-2}\sum_{i=1}^{N}[\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\theta x_i^2}]^2 \sum_{i=1}^{N}\frac{x_i^2}{1+\theta x_i^2}(1-h_{ii}) = 0 \tag{3.10}$$

Where h_{ii} are the diagonal elements of the hat matrix defined in Subsection (2.2.2) and with obvious notation $\boldsymbol{X}_\star$ is the $N \times 2$ matrix given by

$$\boldsymbol{X}_\star = \begin{bmatrix} \frac{1}{1+\theta x_i^2} & \frac{x_i}{1+\theta x_i^2} \end{bmatrix}$$

Then from equation (2.5) we obtain

$$\hat{\sigma^2} = \frac{1}{N-2}\sum_{i=1}^{N}[\frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1+\hat{\theta} x_i^2}]^2 \tag{3.11}$$

3.2.4 Least Squares on Squared Residuals Equation for $\hat{\theta}$

Leverage Not Corrected

From equation (2.8), obtain $\hat{\theta}$ as

$$\sum_{i=1}^{N}\frac{\{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 - \sigma^2(1+\theta x_i^2)^2\}(1+\theta x_i^2)(x_i^2)}{(1+\hat{\theta}_\star x_i^2)^4} = 0 \tag{3.12}$$

and from equation (2.7) obtain σ^2 from equation

$$\sum_{i=1}^{N} \frac{\{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 - \sigma^2(1 + \hat{\theta} x_i^2)^2\}(1 + \hat{\theta} x_i^2)^2}{(1 + \hat{\theta}_\star x_i^2)^4} = 0$$

namely

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{N} \frac{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 (1 + \hat{\theta} x_i^2)^2}{(1 + \hat{\theta}_\star x_i^2)^4}}{\sum_{i=1}^{N} (\frac{1 + \hat{\theta} x_i^2}{1 + \hat{\theta}_\star x_i^2})^4} \tag{3.13}$$

Leverage Corrected

From equation (2.10), solve for $\hat{\theta}$ as

$$\sum_{i=1}^{N} \frac{\{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 - \sigma^2(1 - h_{ii})^2(1 + \theta x_i^2)^2\}(1 + \theta x_i^2)(x_i^2)}{(1 - h_{ii})^2(1 + \hat{\theta}_\star x_i^2)^4} = 0 \tag{3.14}$$

and from equation (2.9) obtain σ^2 as the solution of

$$\sum_{i=1}^{N} \frac{\{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 - \sigma^2(1 - h_{ii})^2(1 + \hat{\theta} x_i^2)^2\}(1 + \hat{\theta} x_i^2)^2}{(1 - h_{ii})^2(1 + \hat{\theta}_\star x_i^2)^4} = 0$$

or

$$\frac{\sum_{i=1}^{N} \frac{[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 (1 + \hat{\theta} x_i^2)^2}{(1 - h_{ii})^2 (1 + \hat{\theta}_\star x_i^2)^4}}{\sum_{i=1}^{N} (\frac{1 + \hat{\theta} x_i^2}{1 + \hat{\theta}_\star x_i^2})^4} \tag{3.15}$$

where $\{h_{ii}\}$ are as defined in Subsection (3.2.3).

3.2.5 Least Squares on Absolute Residuals Equation for $\hat{\theta}$

Leverage Not Corrected

From equation (2.12), solve for $\hat{\theta}$ from

$$\sum_{i=1}^{N} \frac{\{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)| - \sigma(1 + \theta x_i^2)\}(1 + \theta x_i^2)(x_i^2)}{(1 + \hat{\theta}_\star x_i^2)^2} = 0 \tag{3.16}$$

and from equation (2.11), obtain σ from equation

$$\sum_{i=1}^{N} \frac{\{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)| - \sigma(1 + \hat{\theta} x_i^2)\}(1 + \hat{\theta} x_i^2)}{(1 + \hat{\theta}_\star x_i^2)^2} = 0$$

namely

$$\hat{\sigma} = \frac{\sum_{i=1}^{N} \frac{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)|(1 + \hat{\theta} x_i^2)}{(1 + \hat{\theta}_\star x i^2)^2}}{\sum_{i=1}^{N} \left(\frac{1 + \hat{\theta} x_i^2)}{(1 + \hat{\theta}_\star x_i^2)} \right)^2} \tag{3.17}$$

Leverage Corrected

From equation (2.14), solve for $\hat{\theta}$ as

$$\sum_{i=1}^{N} \frac{\{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)| - \sigma(1 - h_{ii})(1 + \theta x_i^2)\}(1 + \theta x_i^2)(x_i^2)}{{}^2(1 + \hat{\theta}_\star x_i^2)^2} = 0 \tag{3.18}$$

and from equation (2.13), obtain σ from equation

$$\sum_{i=1}^{N} \frac{\{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)| - \sigma(1 - h_{ii})(1 + \hat{\theta} x_i^2)\}(1 + \hat{\theta} x_i^2)}{(1 + \hat{\theta}_\star x_i^2)^2} = 0$$

which simplifies to

$$\hat{\sigma} = \frac{\sum_{i=1}^{N} \frac{|y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)|(1+\hat{\theta} x_i^2)}{(1-h_{ii})(1+\hat{\theta}_{\star} x_i^2)^2}}{\sum_{i=1}^{N} (\frac{(1+\hat{\theta} x_i^2)}{(1+\hat{\theta}_{\star} x_i^2)})^2} \tag{3.19}$$

where $\{h_{ii}\}$ are as defined in Subsection (3.2.3).

3.2.6 Modified Maximum Likelihood Equation for $\hat{\theta}$

From equation (2.15), obtain

$$\hat{\sigma^2} = \{\frac{1}{\sum_{i=1}^{M}(m-1)}\} \sum_{i=1}^{M} \frac{\sum_{j=1}^{m}[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{(1+\hat{\theta} x_i^2)^2} \tag{3.20}$$

and from equation (2.16), solve for $\hat{\theta}$ as

$$\sum_{i=1}^{M} \frac{\sum_{j=1}^{m}[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 x_i^2}{\hat{\sigma^2}(1+\theta x_i^2)^3} - (m-1) \sum_{i=1}^{M} \frac{x_i^2}{1+\theta x_i^2} = 0 \tag{3.21}$$

Then obtain the equation for $\hat{\beta}$ as per equation (2.17) which gives

$$\sum_{i=1}^{M} \frac{\sum_{j=1}^{m}[y_{ij} - (\beta_0 + \beta_1 x_i)]}{(1+\hat{\theta} x_i^2)^2} = 0$$

or

$$\sum_{i=1}^{M}\{\frac{\sum_{j=1}^{m} y_{ij}}{(1+\hat{\theta} x_i^2)^2}\} - \sum_{i=1}^{M}\{\frac{m(\beta_0 + \beta_1 x_i)}{(1+\hat{\theta} x_i^2)^2}\} = 0$$

and hence the first equation is given by

$$\beta_0 \sum_{i=1}^{M}\{\frac{1}{(1+\hat{\theta} x_i^2)^2}\} + \beta_1 \sum_{i=1}^{M}\{\frac{x_i}{(1+\hat{\theta} x_i^2)^2}\} = \frac{1}{m} \sum_{i=1}^{M}\{\frac{\sum_{j=1}^{m} y_{ij}}{(1+\hat{\theta} x_i^2)^2}\}$$

Similarly, write the second equation as

$$\beta_0 \sum_{i=1}^{M} \{\frac{x_i}{(1+\hat{\theta}x_i^2)^2}\} + \beta_1 \sum_{i=1}^{M} \{\frac{x_i^2}{(1+\hat{\theta}x_i^2)^2}\} = \frac{1}{m} \sum_{i=1}^{M} \{\frac{x_i \sum_{j=1}^{m} y_{ij}}{(1+\hat{\theta}x_i^2)^2}\}$$

and in matrix notation write

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M}\{\frac{1}{(1+\hat{\theta}x_i^2)^2}\} & \sum_{i=1}^{M}\{\frac{x_i}{(1+\hat{\theta}x_i^2)^2}\} \\ \sum_{i=1}^{M}\{\frac{x_i}{(1+\hat{\theta}x_i^2)^2}\} & \sum_{i=1}^{M}\{\frac{x_i^2}{(1+\hat{\theta}x_i^2)^2}\} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{M}\{\frac{\sum_{j=1}^{m} y_{ij}}{(1+\hat{\theta}x_i^2)^2}\} \\ \frac{1}{m}\sum_{i=1}^{M}\{\frac{x_i \sum_{j=1}^{m} y_{ij}}{(1+\hat{\theta}x_i^2)^2}\} \end{bmatrix}$$

Simplifying to

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M}\{\frac{1}{(1+\hat{\theta}x_i^2)^2}\} & \sum_{i=1}^{M}\{\frac{x_i}{(1+\hat{\theta}x_i^2)^2}\} \\ \sum_{i=1}^{M}\{\frac{x_i}{(1+\hat{\theta}x_i^2)^2}\} & \sum_{i=1}^{M}\{\frac{x_i^2}{(1+\hat{\theta}x_i^2)^2}\} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{M} \frac{\bar{y}_i}{(1+\hat{\theta}x_i^2)^2} \\ \sum_{i=1}^{M} \frac{x_i \bar{y}_i}{(1+\hat{\theta}x_i^2)^2} \end{bmatrix} \quad (3.22)$$

Notice that since μ_i is unknown it is replaced by $\bar{y}_i$. In the event of unequal replication, replace m by m_i in equations (3.20) and (3.21) obtaining

$$\hat{\sigma^2} = \{\frac{1}{\sum_{i=1}^{M}(m_i - 1)}\} \sum_{i=1}^{M} \frac{\sum_{j=1}^{m_i}[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{(1+\hat{\theta}x_i^2)^2} \quad (3.23)$$

and

$$\sum_{i=1}^{M} \frac{\sum_{j=1}^{m_i}[y_{ij} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 x_i^2}{\hat{\sigma^2}(1+\theta x_i^2)^3} - \sum_{i=1}^{M} \frac{(m_i - 1)x_i^2}{1+\theta x_i^2} = 0 \quad (3.24)$$

3.2.7 Extended Quasi Likelihood equation for $\hat{\theta}$

From equation (2.18) obtain $\hat{\theta}$ as

$$\sum_{i=1}^{N} \frac{\sigma^2 x_i^2(1+\theta x_i^2)^2 + 2y_i x_i^2(\hat{\beta}_0+\hat{\beta}_1 x_i) - 2y_i^2 x_i^2 - x_i^2(\hat{\beta}_0+\hat{\beta}_1 x_i)^2}{(1+\theta x_i^2)^3} = 0 \tag{3.25}$$

and from equation (2.19) obtain σ^2 from

$$\sum_{i=1}^{N} \frac{\sigma^2(1+\hat{\theta} x_i^2)^2 + 2y_i(\hat{\beta}_0+\hat{\beta}_1 x_i) - y_i^2 - (\hat{\beta}_0+\hat{\beta}_1 x_i)^2}{(1+\hat{\theta} x_i^2)^2} = 0$$

or

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{N} \frac{[y_i^2+(\hat{\beta}_0+\hat{\beta}_1 x_i)^2]}{(1+\hat{\theta} x_i^2)^2}}{\sum_{i=1}^{N} \frac{(1+\hat{\theta} x_i^2)^2+2y_i(\hat{\beta}_0+\hat{\beta}_1 x_i)}{(1+\hat{\theta} x_i^2)^2}} \tag{3.26}$$

3.2.8 Algorithm

We now give an iterative procedure for obtaining θ from the above formulated equations. The algorithm as presented here allows a general form for all the methods of estimating θ that have been considered. The procedure for the algorithm follow.

Step 1 Put the starting $\hat{\theta}_\star$ equal to zero and obtain the starting $\hat{\beta}_\star$. Notice that this is equivalent to starting off with the least squares estimate of $\hat{\beta}_\star$.

Step 2 Obtain an estimate of σ (closed form) from the starting $\hat{\beta}_\star$.

Step 3 Obtain the estimated $\hat{\theta}$ using the current $\hat{\beta}$ and $\hat{\sigma}$.

Step 4 Obtain new $\hat{\beta}$ using the current $\hat{\theta}$.

Step 5 Obtain new $\hat{\sigma}$ using the current $\hat{\theta}$ and $\hat{\beta}$.

Step 6 Repeat steps 3, 4 and 5.

Step 7 Stop when there is little change in both $\hat{\beta}$ and $\hat{\theta}$ say up to 10^{-8} accuracy.

We summarize the algorithm in Figure 3.1.

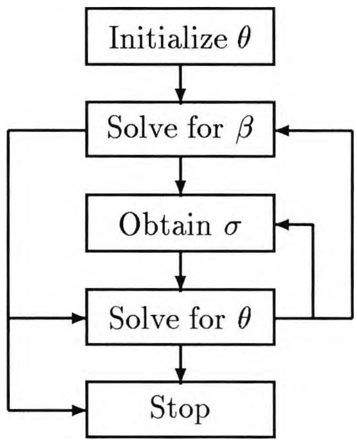

Figure 3.1: Simple Algorithm Flow Diagram

3.3 Simulations

We begin the exercise by obtaining simulated data as follows. In model (3.1), let

$$e_i = \epsilon_i \sigma(1 + \theta x_i^2)$$

and generate independent values of ϵ_i according to $\epsilon_i \sim N(0, 1)$. Further, let the $x_i's$ be equally spaced on the interval $[-1, 1]$ with 0.1 spacing. Thus the simulations are performed with a sample of size $N = 21$. Next, fix β_0 and β_1 to be 0 and 1 respectively. In model (3.2) let θ be 1 and σ^2 be 0.3. Finally, in the event of replicated procedures take $m = 2$. For each procedure 100 simulations were carried out and the bias, root mean squared error and root mean integrated squared error where computed as

$$BIAS(\hat{v}) = \frac{1}{100}\sum_{i=1}^{100} \hat{v}_i - v \qquad ,$$

$$RMSE(\hat{v}) = \sqrt{\frac{1}{100}\sum_{i=1}^{100}(\hat{v}_i - v)^2}$$

and

$$RMISE = \sqrt{\frac{1}{100}\sum_{i=1}^{100}\left\{\frac{1}{20001}\sum_{j=1}^{20001} d_j^2\right\}_i}$$

obtaining the results as in Section (3.4). Here v indicates θ, σ^2, β_0 and β_1 respectively, d_j is the difference between the fitted and the true function at the points $x_j = -1, \cdots, 1$ at intervals of 0.001 and subscript i denotes the ith simulation. In the logarithm of absolute residuals procedure, the problem of the residuals near zero was avoided by discarding all the simulations which portrayed such a problem.

3.4 Empirical Results

The numerical outputs are presented in four tables. Table 3.1 gives the averages, Table 3.2 gives the biases while Table 3.3 gives the root mean squared errors and Table 3.4 gives the root mean integrated squared errors for the various methods under study.

The abbreviations PLH, RML, SR1, SR2, AR1, AR2, LAR, EQL, MML, and ROD correspond respectively to pseudo likelihood, restricted maximum likelihood, squared residuals with leverage, squared residuals without leverage, absolute residuals with leverage, absolute residuals without leverage, logarithm of absolute residuals, extended quasi likelihood, modified maximum likelihood and Rodbard.

3.5 Discussion and Conclusion

Several different procedures were formulated to estimate the variance function with the emphasis on the estimation of θ. With the exception of Logarithm of Absolute Residuals (LAR), Extended Quasi Likelihood (EQL) and Rodbard (ROD) all the other procedures underestimate θ. We consider the absolute bias and rank these procedures according to the three criteria, absolute bias, root mean squared error and root mean integrated squared error as shown in Table 3.5. Absolute Residuals with leverage corrected (AR2) does best. However for RMSE Modified Maximum Likelihood (MML) seems to do well and for RMISE the best procedure is Pseudo Likelihood (PLH). Table 3.5 gives the ranking according to the three criteria. This clearly shows that although the differences are small Squared Residuals (SR1) and Absolute Residuals (AR2) are good at least for cases where replication is not available. Modified Maximum Likelihood (MML) could be recommended for replicated cases. Note that MML appear better because of the double sample size arising from the replication. Another observation is that the variances are larger than the biases for all the procedures as shown by Tables 3.2 and 3.6. No methods are terrible.

	$\hat{\theta}$	$\hat{\sigma^2}$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	0.83	0.32	0.02	0.95
RML	0.96	0.23	0.00	0.97
SR1	0.86	0.28	0.01	1.02
SR2	0.87	0.32	0.02	1.02
AR1	0.92	0.20	-0.01	1.01
AR2	0.98	0.23	0.02	1.02
LAR	1.14	0.38	0.01	0.99
EQL	1.16	0.27	-0.02	1.05
MML	0.93	0.49	0.01	0.97
ROD	1.15	0.34	0.02	0.96

Table 3.1: Average Estimates

	$\hat{\theta}$	$\hat{\sigma^2}$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	-0.17	0.02	0.02	-0.05
RML	-0.04	-0.07	0.00	-0.03
SR1	-0.14	-0.02	0.01	0.02
SR2	-0.13	0.02	0.02	0.02
AR1	-0.08	-0.10	-0.01	0.01
AR2	-0.02	-0.07	0.02	0.02
LAR	0.14	0.08	0.01	-0.01
EQL	0.16	-0.03	-0.02	0.05
MML	-0.07	0.19	0.01	-0.03
ROD	0.15	0.04	0.02	-0.04

Table 3.2: Biases

	$\hat{\theta}$	$\hat{\sigma^2}$	$\hat{\beta}_0$	$\hat{\beta}_1$
PLH	0.522	0.119	0.153	0.325
RML	0.568	0.191	0.183	0.311
SR1	0.474	0.111	0.131	0.309
SR2	0.565	0.137	0.160	0.288
AR1	0.530	0.120	0.139	0.326
AR2	0.510	0.122	0.133	0.283
LAR	0.481	0.118	0.122	0.316
EQL	0.490	0.104	0.158	0.326
MML	0.455	0.228	0.111	0.233
ROD	0.489	0.124	0.136	0.310

Table 3.3: Root Mean Square Errors

PLH	RML	SR1	SR2	AR1	AR2	LAR	EQL	MML	ROD
0.202	0.326	0.217	0.221	0.274	0.247	0.374	0.207	0.303	0.294

Table 3.4: Root Mean Integrated Square Errors

	$\hat{\theta}$		
	BIAS	RMSE	RMISE
PLH	9	7	1
RML	2	10	9
SR1	6	2	3
SR2	5	9	4
AR1	4	8	6
AR2	1	6	5
LAR	6	3	10
EQL	8	5	2
MML	3	1	8
ROD	7	4	7

Table 3.5: Ranking

	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$
PHL	0.244	0.014	0.025	0.103
RML	0.321	0.032	0.033	0.096
SR1	0.205	0.012	0.017	0.095
SR2	0.302	0.018	0.025	0.083
AR1	0.275	0.005	0.019	0.106
AR2	0.260	0.010	0.017	0.080
LAR	0.212	0.008	0.015	0.100
EQL	0.215	0.010	0.025	0.104
MML	0.202	0.016	0.012	0.054
ROD	0.217	0.014	0.018	0.095

Table 3.6: Variances

Chapter 4

NONPARAMETRIC APPROACH

4.1 Introduction

The nonparametric approach differs radically from the parametric approach. Here, the only assumption about the variance functions is that they are given by an unknown but smooth function of the design or the mean response. Consider for example the case where the variance function is known to be determined by the $x_i's$ and rewrite model (1.1) as

$$y_i = f(x_i) + e_i \,, \qquad i = 1, \ldots, N \tag{4.1}$$

We assume that the data are heteroscedastic and denote the possibly nonconstant variance function by $g(x_i)$, where $g(x_i)$ is unknown but assumed to be a smooth function. This chapter describes the techniques for carrying out the improvement of estimators based on data at a particular x_i.

There are two types of designs, the fixed design where the predictor variables $x_i's$ are equally spaced and the "random" design where spacing is haphazard. Here the fixed design is considered and only univariate data is studied. To investigate how to estimate the function $g(x_i)$ a foundation is necessary. We call this foundation the initial variance estimate.

4.2 Initial Variance Estimates

There are three categories of initial variance estimates which include; (i) residuals (ii) sample variance and sample standard deviation (iii) difference schemes. In all these initial variance estimates, the underlying idea is to remove the influence of the mean function first.

4.2.1 Residuals

Suppose the mean function $f(x_i)$ were known then we define the residuals as

$$r_i = y_i - f(x_i)\,, \quad i = 1, \cdots, N \tag{4.2}$$

Then take the square residuals as the initial estimate of $g(x_i)$. Modifying the squared residuals leads to some more initial variance estimates based on the squared residuals. One such modification is to take the logarithm of the squared residuals. Another possibility is to divide the squared residuals by some factor leading to studentized or standardized squared residuals as defined in Carroll and Ruppert (1988). We consider the cube root of the squared residuals as another possible modification. This particular modification has been frequently utilized in the diagnostics of heterogeneity.

Absolute residuals is another potential initial variance estimate. While squared residuals can be thought of as estimating the variances, absolute residuals can be thought of as estimating the standard deviations. The modification procedure applied on the squared residuals are carried over without alteration leading to some more initial variance estimates based on the absolute residuals.

4.2.2 Sample Variance and Standard Deviation

Suppose that at each point x_i there are m_i replication observations where $m_i \geq 2$. Then we write the regression model with one dimensional predictor as

$$y_{ij} = f(x_i) + e_{ij} \tag{4.3}$$

As an initial estimate of $g(x_i)$ take

$$s_i^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2 \tag{4.4}$$

and in analogy to the absolute residuals, take s_i as the initial estimates of the standard deviations.

4.2.3 Difference Schemes

These are weighted differences of the observations for points neighbouring a fixed x_i where the variance is to be estimated. The simplest example is the successive differences $y_{i+1} - y_i$. The general case considers weighted sums of any $m \geq 2$ observations where the weights are obtained through moving averages or some other criterion for example as described in Muller and Stadtmuller (1987). One general case is defined as

$$\tilde{g}_i = \sum_{j=m_i}^{m_2} w_j y_{i+j} \tag{4.5}$$

where w_j are weights such that $\sum_{j=m_1}^{m_2} w_j = 0$, $\sum_{j=m_1}^{m_2} w_j^2 = 1$, $m_1 = -[\frac{m}{2}]$, $m_2 = [\frac{m}{2} - \frac{1}{4}]$, $m \geq 2$ is a fixed integer and [.] denotes the integer part. Then we take $\tilde{g}_i^2$ as the initial variance estimates and $|\tilde{g}_i|$ as the initial estimates for the standard deviations.

Another difference scheme is motivated by the variance formula

$$\text{var}(y_i) = E[y_i - E(y_i)]^2 = E[y_i^2] - [E(y_i)]^2 \tag{4.6}$$

leading to the use of the squared observations or absolute observations as the initial variance estimates.

4.3 Improving the Initial Estimates

We denote by z_i the initial variance estimate at x_i. In order to obtain a better estimate, we smooth the neighbouring values of z_i. For this purpose, view the initial estimate z_i as measurement coming from the regression model

$$z_i = g(x_i) + \tilde{e}_i \tag{4.7}$$

where $g(x_i)$ is unknown but smooth in the sense that $g(.)$ has a continuous first derivative; if x_1 and x_2 are very close, so too should $g(x_1)$ and $g(x_2)$ and information about $g(x_1)$ can be obtained from data at x_2. Some of the best known smoothers are

1. Smoothing splines, Reinsch (1967), Rice and Rosenblatt (1981) and Wahba (1975).
2. K-nearest-neighbour smoother, Friedman (1975), Cover (1968) and Mack (1981).
3. Kernel smoothers with types Priestley-Chao (1972), Gasser- Muller (1979,1984), Nadaraya-Watson (1964), and Local Linear Regression smoother, Fan(1992, 1993).

Other smoothers have been considered in detail, Hastie and Tibshirani (1991). Kernel smoothers have been found to have minimax optimal properties, Gasser and Engel (1990).

4.4 Kernel Smoothing

We shall consider smoothing by the Gasser-Muller (1984) kernel smoother,

$$\hat{g}(x_i) = \frac{1}{b}\sum_{j=1}^{N}\int_{s_{j-1}}^{s_j} z_j K(\frac{x_i - u}{b})du \tag{4.8}$$

with $s_0 = x_1$, $s_N = x_N$, $s_j = \frac{x_j + x_{j+1}}{2}$ and $1 \leq j \leq N-1$. Here b is the bandwidth or smoothing parameter that controls the amount of smoothing and K denotes the kernel (weight) function with properties $K(x) \geq 0$, $\int_{-\infty}^{\infty} K(x) = 1$ and $K(-x) = K(x)$ for all x. The Gasser-Muller-type kernel smoother has been investigated and found to perform well relative to other kernel type smoothers, Jennen-Steinmetz and Gasser (1988), Mack and Muller (1988) and Hall and Wehrly (1991). To be able to apply the kernel smoother, b and K must be chosen.

4.5 Choice of Kernel

In the context of kernel smoothing, two types of kernels for smoothing functions have been studied; the minimum variance kernels which minimize the asymptotic variance and the optimal kernels which minimize the asymptotic integrated mean squared error, Gasser, Muller and Mammitzsch (1985) and Muller (1984).

An important characterization of a kernel is its degree of smoothness since it will typically be inherited by the smoothed curve. It is typical to use a μ times differentiable kernel, $\mu \geq 0$. In practice, $\mu = 1$ or $\mu = 2$ is recommended (see, Muller and Wang (1994)). Further, for carrying out the smoothing procedure, the kernel is required to satisfy two conditions; (i) $\int_{-1}^{1} K(x)dx = 1$, that is the kernel has support in the interval $[-1, 1]$ and (ii) the kernel is Lipschitz of order one in the interval $[-1, 1]$.

Definition 1 A continuous function f is said to satisfy a Lipschitz condition of order α if there exists positive constants m and α such that $|f(x_1) - f(x_2)| < m|x_1 - x_2|^{\alpha}$ for all x_1 and x_2 in the domain of f.

Remark Every continuous differentiable function f on a closed interval $[a, b]$ satisfies a

Lipschitz condition of order one on that interval.

Definition 2 A function f defined in an interval is called continuous in that interval if it is continuous at every point of the interval.

In this thesis the kernel function

$$K(x) = \frac{3}{4}(1 - x^2), \qquad |x| \leq 1 \tag{4.9}$$

is considered in detail. In addition to the above conditions there are two reasons why we chose to work with this kernel function. The first reason which is not very important is the computational convenience. The other reason is that this is an optimal kernel; optimal in the sense that it minimizes the asymptotic integrated mean square error as explained in Muller (1984).

4.6 Boundary Modification

There is one practical relevant problem that occurs when applying kernel smoothing. This is the boundary (edge) effects which is a phenomenon in which the bias of a smoother increases near the endpoints of the smoothed interval. To counteract this problem, we shall modify the kernel estimate near the boundaries by use of linear combination of estimates with several bandwidths, Rice (1984). Other techniques have been proposed to deal with problems of boundary or edge effects for nonparametric regression smoothing , Gasser and Muller (1979), Muller and Mammitzsch (1985), Rice and Rosenblatt (1983), Schuster

(1985), Silverman (1986), Cline and Hart (1991), Hall and Wehrly (1991), Muller (1991) and Muller and Wang (1994).

4.6.1 Modification Procedure

Define

$$B_L = \{x_i : x_1 \leq x_i < x_N\}$$

as the left boundary region

$$I = \{x_i : x_1 + b \leq x_i \leq x_N - b\}$$

as the interior where boundary effects do not occur and

$$B_R = \{x_i : x_N - b < x_i \leq x_N\}$$

as the right boundary region. In I use (4.8) without modification. Further, define

$$W_0(q_i) = \int_{-1}^{q_i} K(x)dx \ , \tag{4.10}$$

$$W_1(q_i) = \int_{-1}^{q_i} xK(x)dx \ , \tag{4.11}$$

$$R(q_i) = \frac{W_1(q_i)}{W_0(q_i)} \ ,$$

$$\alpha_i = 2 - q_i \ ,$$

and

$$\beta_i = \frac{R(q_i)}{\alpha_i R(\frac{q_i}{\alpha_i}) - R(q_i)} \tag{4.12}$$

then in B_L use

$$\dot{g}(x_i) = \hat{g}(x_i, b) + \beta_i[\hat{g}(x_i, b) - \hat{g}(x_i, b\alpha_i)] \tag{4.13}$$

with

$$q_i = \frac{x_i - x_1}{b}$$

and b is the bandwidth parameter used in the interior. Similarly, in B_R the procedure remains the same except that we integrate (4.10) and (4.11) from $-q_i$ to 1 where

$$q_i = \frac{x_N - x_i}{b}$$

and change the minus sign in (4.13) to plus sign. We therefore use in B_R the modified estimate

$$\dot{g}(x_i) = \hat{g}(x_i, b) + \beta_i[\hat{g}(x_i, b) + |\hat{g}(x_i, b\alpha_i)|] \tag{4.14}$$

4.7 Choice of the Bandwidth Parameter

The theory of bandwidth choice in nonparametric regression is developing fast. Most of the bandwidth selectors studied in the literature are based on the minimization of some function of b which is related to the residual sum of squares. Cao, Cuevas and Manteiga (1994) have given a critical up-to-date review of the main methods currently available.

However, no selector appears to be uniformly better. Considered here is a cross-validation version proposed by Hardle and Marron (1985) which has been shown to work relatively well, Hart and Wehrly (1992). The procedure is to minimize $\frac{1}{N}\sum_{i=1}^{N}[z_i - \hat{g}(x_i)]^2 w(x_i)$ with respect to b obtaining

$$\sum_{i=1}^{N} w(x_i)[\hat{g}(x_i) - z_i]\frac{d}{db}\hat{g}(x_i) = 0 \tag{4.15}$$

This is equivalent to the weighted cross-validation score criterion, Silverman (1985). The weights $w(x_i)$ are to be estimated through moving averages.

4.7.1 Iterative Estimation of the Weights

Obtain the estimates of the weights by use of moving averages procedure, Silverman (1985). Proceed as follows:

Step 1 Obtain the unweighted $\hat{g}(x_i)$

Step 2 Compute $r_j^2 = [z_i - \hat{g}(x_i)]^2$

Step 3 Compute $\hat{w}(x_i) = (n_i - m_i + 1)^{-1}\sum_{j=m_i}^{n_i} r_j^2$ with $m_i = \max(1, i-k)$ and $n_i = \min(N, i+k)$ for some fixed constant k. $k = 5$ was found to perform well.

Step 4 Substitute $\hat{w}(x_i)$ into (4.15) and obtain $\hat{b}$

Step 5 A new value of $\hat{g}(x_i)$ is obtained using $\hat{b}$

Step 6 Compute new weight estimates $\dot{w}(x_i) = (n_i - m_i + 1)^{-1}\hat{w}(x_i)\sum_{j=m_i}^{n_i} r_j^2$ where r_j are the new residuals.

Step 7 The process is repeated until convergence occurs.

Other techniques have been proposed to estimate the weights, Rice(1994), Breiman and Meisel(1976), Gasser, Muller and Stadtmuller (1987).

4.8 Empirical Study

We now carry out our simulation study. The study is comparative aimed at finding the initial estimator that could lead to better estimate of $g(x_i)$. The simulation procedure is similar to that carried out in the parametric case study. The details of the study follow.

Consider a simple model

$$y_i = x_i + e_i \, , \qquad i = 1, \ldots, N \tag{4.16}$$

and take the true variance function as

$$\text{var}(y_i) = g(x_i) = 0.3(1 + x_i^2)^2 \tag{4.17}$$

Then in (4.16) let $e_i = \epsilon_i \sqrt{0.3}(1 + x_i^2)$ and generate independent values of ϵ_i according to $\epsilon_i \sim N(0, 1)$. Thus the true mean function is $E(y_i) = x_i$. Now let the $x_i's$ be equally spaced on the interval $[-1, 1]$ with 0.1 spacing. As in the case of parametric variance function estimation study, the simulations are performed with a sample of size $N = 21$. Table 4.1 gives the true means and variances respectively. Then for each case 100 simulations were carried out and computed the bias, root mean squared error and root mean integrated

	True Mean	True Variance
1.	-1.0	1.20
2.	-0.9	0.98
3.	-0.8	0.81
4.	-0.7	0.67
5.	-0.6	0.55
6.	-0.5	0.47
7.	-0.4	0.40
8.	-0.3	0.36
9.	-0.2	0.32
10.	-0.1	0.31
11.	0.0	0.30
12.	0.1	0.31
13.	0.2	0.32
14.	0.3	0.36
15.	0.4	0.40
16.	0.5	0.47
17.	0.6	0.55
18.	0.7	0.67
19.	0.8	0.81
20.	0.9	0.98
21.	1.0	1.20

Table 4.1: True Mean and True Variance

squared error as

$$BIAS[\hat{g}(x_i)] = \frac{1}{100}\sum_{i=1}^{100}\hat{g}(x_i) - g(x_i) \quad ,$$

$$RMSE[\hat{g}(x_i)] = \sqrt{\frac{1}{100}\sum_{i=1}^{100}[\hat{g}(x_i) - g(x_i)]^2}$$

and

$$RMISE = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left\{\frac{1}{21}\sum_{i=1}^{21}[\hat{g}(x_i) - g(x_i)]^2\right\}_j}$$

where subscript j denotes the jth simulation.

4.9 Results

The numerical outputs are presented in 13 tables. Tables 4.2 – 4.5 give the smooth variances. Tables 4.6 – 4.9 give the biases. Tables 4.10 – 4.13 give the root mean squared errors and Table 4.14 gives the root mean integrated squared errors. In these tables, r_i denote the ordinary residuals while $\tilde{r}_i$ denote the scaled residuals where the scales are obtained by moving averages. We denote by s_i the sample standard deviation where the number of replications is 2. Finally, $\tilde{g}_i$ denote the difference schemes defined in equation (4.5) and $\breve{g}_i$ indicate the difference schemes defined in equation (4.6).

4.10 Discussion and Comments

In the class of squared residuals the scaled squared residuals $\tilde{r}_i^2$ are good. They have low bias, low root mean squared error and low root mean integrated squared error. Absolute residuals show the same phenomenon offering $|\tilde{r}_i|$ as the best in this class. Generally, absolute residuals seem to provide more reasonable estimates than those obtained by the squared residuals. They portray a far much better symmetrical distribution behaviour. As a consequence, in the smoothing of residuals recommending the use of absolute residuals looks acceptable.

In the same manner, sample standard deviations have more symmetrical distribution than sample variances. As indicated by the biases, root mean squared errors and root mean integrated squared errors it seems that smoothing sample standard deviations is preferable to smoothing sample variances.

	r_i^2	$\log(r_i^2)$	$r_i^{\frac{2}{3}}$	$\tilde{r}_i^2$
1.	0.88	0.55	0.90	0.88
2.	0.70	0.49	0.77	0.78
3.	0.59	0.33	0.66	0.69
4.	0.51	0.30	0.57	0.61
5.	0.45	0.29	0.49	0.53
6.	0.43	0.24	0.43	0.47
7.	0.40	0.17	0.39	0.43
8.	0.40	0.17	0.35	0.39
9.	0.38	0.16	0.33	0.37
10.	0.39	0.15	0.32	0.35
11.	0.38	0.15	0.31	0.34
12.	0.38	0.16	0.31	0.34
13.	0.37	0.19	0.32	0.36
14.	0.39	0.23	0.35	0.38
15.	0.41	0.26	0.38	0.43
16.	0.44	0.31	0.44	0.48
17.	0.48	0.33	0.50	0.54
18.	0.58	0.34	0.56	0.61
19.	0.68	0.49	0.64	0.68
20.	0.72	0.50	0.73	0.75
21.	0.81	0.51	0.83	0.84

Table 4.2: Smooth Variance (Squared Residuals)

	$\|r_i\|$	$\log\|r_i\|$	$\|r_i\|^{\frac{2}{3}}$	$\|\tilde{r}_i\|$
1.	0.78	0.43	0.82	0.87
2.	0.67	0.35	0.71	0.75
3.	0.57	0.28	0.60	0.64
4.	0.48	0.23	0.51	0.54
5.	0.41	0.19	0.43	0.46
6.	0.35	0.16	0.37	0.39
7.	0.31	0.14	0.32	0.34
8.	0.27	0.13	0.29	0.31
9.	0.25	0.12	0.26	0.28
10.	0.23	0.11	0.25	0.26
11.	0.23	0.11	0.24	0.26
12.	0.23	0.10	0.25	0.26
13.	0.25	0.11	0.26	0.28
14.	0.27	0.12	0.29	0.30
15.	0.30	0.14	0.32	0.34
16.	0.35	0.17	0.37	0.39
17.	0.40	0.20	0.42	0.45
18.	0.47	0.23	0.49	0.52
19.	0.54	0.27	0.57	0.61
20.	0.63	0.31	0.66	0.70
21.	0.72	0.36	0.76	0.81

Table 4.3: Smooth Variance (Absolute Residuals)

	s_i^2	s_i
1.	1.02	0.95
2.	0.87	0.80
3.	0.72	0.67
4.	0.62	0.56
5.	0.52	0.46
6.	0.45	0.39
7.	0.41	0.32
8.	0.41	0.30
9.	0.39	0.28
10.	0.39	0.29
11.	0.38	0.29
12.	0.39	0.30
13.	0.38	0.30
14.	0.40	0.32
15.	0.41	0.34
16.	0.45	0.39
17.	0.50	0.46
18.	0.60	0.56
18.	0.72	0.67
20.	0.88	0.82
21.	1.05	0.94

Table 4.4: Smooth Variance (Sample Var and Std)

	$\tilde{q}_i^2$	$\|\tilde{q}_i\|$	$\check{q}_i^2$	$\|\check{q}_i\|$
1.	0.99	1.03	1.08	0.69
2.	0.79	0.85	0.88	0.55
3.	0.67	0.70	0.74	0.45
4.	0.57	0.58	0.64	0.37
5.	0.50	0.48	0.61	0.32
6.	0.47	0.40	0.60	0.28
7.	0.44	0.34	0.55	0.27
8.	0.43	0.30	0.53	0.26
9.	0.42	0.27	0.49	0.25
10.	0.43	0.25	0.49	0.25
11.	0.42	0.24	0.48	0.24
12.	0.43	0.24	0.49	0.23
13.	0.43	0.25	0.49	0.23
14.	0.46	0.27	0.53	0.23
15.	0.45	0.30	0.55	0.23
16.	0.51	0.35	0.61	0.25
17.	0.48	0.41	0.67	0.28
18.	0.57	0.49	0.78	0.33
19.	0.67	0.57	0.80	0.40
20.	0.81	0.68	0.95	0.51
21.	1.01	0.81	1.18	0.64

Table 4.5: Smooth Variance (Difference Schemes)

	r_i^2	$\log(r_i^2)$	$r_i^{\frac{2}{3}}$	$\tilde{r}_i^2$
1.	-0.32	-0.65	-0.30	-0.32
2.	-0.28	-0.49	-0.21	-0.20
3.	-0.22	-0.48	-0.15	-0.12
4.	-0.16	-0.48	-0.10	-0.06
5.	-0.10	-0.37	-0.06	-0.02
6.	-0.04	-0.26	-0.04	0.00
7.	0.00	-0.23	-0.01	0.03
8.	0.04	-0.23	-0.01	0.03
9.	0.06	-0.19	0.01	0.05
10.	0.08	-0.16	0.01	0.04
11.	0.08	-0.16	0.01	0.04
12.	0.07	-0.15	0.00	0.03
13.	0.05	-0.13	0.00	0.04
14.	0.03	-0.13	-0.01	0.02
15.	0.01	-0.14	-0.02	0.03
16.	-0.03	-0.16	-0.03	0.01
17.	-0.07	-0.22	-0.05	-0.01
18.	-0.09	-0.33	-0.11	-0.06
19.	-0.13	-0.32	-0.17	-0.13
20.	-0.26	-0.48	-0.25	-0.23
21.	-0.39	-0.69	-0.37	-0.36

Table 4.6: Bias (Squared Residuals)

	$\|r_i\|$	$\log\|r_i\|$	$\|r_i\|^{\frac{2}{3}}$	$\|\tilde{r}_i\|$
1.	-0.42	-0.77	-0.38	-0.33
2.	-0.31	-0.63	-0.27	-0.23
3.	-0.24	-0.53	-0.21	-0.17
4.	-0.19	-0.44	-0.16	-0.13
5.	-0.14	-0.36	-0.12	-0.09
6.	-0.12	-0.31	-0.10	-0.08
7.	-0.09	-0.26	-0.08	-0.06
8.	-0.09	-0.23	-0.07	-0.05
9.	-0.07	-0.20	-0.06	-0.04
10.	-0.08	-0.20	-0.06	-0.05
11.	-0.07	-0.19	-0.06	-0.04
12.	-0.08	-0.21	-0.06	-0.05
13.	-0.07	-0.21	-0.06	-0.04
14.	-0.09	-0.24	-0.07	-0.06
15.	-0.10	-0.26	-0.08	-0.06
16.	-0.12	-0.30	-0.10	-0.08
17.	-0.15	-0.35	-0.13	-0.10
18.	-0.20	-0.44	-0.18	-0.15
19.	-0.27	-0.54	-0.24	-0.20
20.	-0.35	-0.67	-0.32	-0.28
21.	-0.48	-0.84	-0.44	-0.39

Table 4.7: Bias (Absolute Residuals)

	s_i^2	s_i
1.	-0.18	-0.25
2.	-0.11	-0.18
3.	-0.09	-0.14
4.	-0.05	-0.11
5.	-0.03	-0.09
6.	-0.02	-0.08
7.	0.01	-0.08
8.	0.05	-0.06
9.	0.07	-0.04
10.	0.08	-0.02
11.	0.08	-0.01
12.	0.08	-0.01
13.	0.06	-0.02
14.	0.04	-0.04
15.	0.01	-0.06
16.	-0.02	-0.08
17.	-0.05	-0.09
18.	-0.07	-0.11
19.	-0.09	-0.14
20.	-0.10	-0.16
21.	-0.15	-0.26

Table 4.8: Bias (Sample Var and Std)

	$\tilde{q}_i^2$	$\lvert\tilde{q}_i\rvert$	$\check{q}_i^2$	$\lvert\check{q}_i\rvert$
1.	-0.21	-0.17	-0.12	-0.51
2.	-0.19	-0.13	-0.10	-0.43
3.	-0.14	-0.11	-0.07	-0.36
4.	-0.10	-0.09	-0.03	-0.30
5.	-0.05	-0.07	0.06	-0.23
6.	0.00	-0.07	0.13	-0.19
7.	0.04	-0.06	0.15	-0.13
8.	0.07	-0.06	0.17	-0.10
9.	0.10	-0.05	0.17	-0.07
10.	0.12	-0.06	0.18	-0.06
11.	0.12	-0.06	0.18	-0.06
12.	0.12	-0.07	0.18	-0.08
13.	0.11	-0.07	0.17	-0.09
14.	0.10	-0.09	0.17	-0.13
15.	0.05	-0.10	0.15	-0.17
16.	0.04	-0.12	0.14	-0.22
17.	-0.07	-0.14	0.12	-0.27
18.	-0.10	-0.18	0.11	-0.34
19.	-0.14	-0.24	-0.01	-0.41
20.	-0.17	-0.30	-0.03	-0.47
21.	-0.25	-0.39	-0.02	-0.56

Table 4.9: Bias (Difference Schemes)

	r_i^2	$\log(r_i^2)$	$r_i^{\frac{2}{3}}$	$\tilde{r}_i^2$
1.	0.81	0.90	0.66	0.65
2.	0.67	0.77	0.54	0.52
3.	0.55	0.58	0.45	0.44
4.	0.43	0.51	0.37	0.37
5.	0.32	0.44	0.31	0.30
6.	0.23	0.38	0.25	0.24
7.	0.17	0.34	0.21	0.20
8.	0.15	0.30	0.18	0.17
9.	0.14	0.29	0.16	0.16
10.	0.16	0.27	0.15	0.15
11.	0.16	0.26	0.15	0.15
12.	0.17	0.27	0.16	0.15
13.	0.17	0.29	0.18	0.17
14.	0.17	0.31	0.21	0.20
15.	0.20	0.36	0.25	0.24
16.	0.24	0.40	0.29	0.28
17.	0.30	0.46	0.33	0.31
18.	0.36	0.50	0.36	0.35
19.	0.43	0.54	0.39	0.38
20.	0.53	0.69	0.43	0.42
21.	0.64	0.83	0.51	0.50

Table 4.10: Root Mean Squared Error (Square Residuals)

	$\|r_i\|$	$\log\|r_i\|$	$\|r_i\|^{\frac{2}{3}}$	$\|\tilde{r}_i\|$
1.	0.72	1.15	0.69	0.64
2.	0.55	0.93	0.52	0.49
3.	0.43	0.77	0.41	0.38
4.	0.35	0.64	0.32	0.30
5.	0.26	0.52	0.25	0.23
6.	0.22	0.45	0.21	0.20
7.	0.17	0.37	0.17	0.16
8.	0.17	0.33	0.15	0.14
9.	0.14	0.29	0.13	0.13
10.	0.14	0.29	0.13	0.12
11.	0.14	0.28	0.13	0.12
12.	0.14	0.30	0.13	0.13
13.	0.15	0.30	0.14	0.13
14.	0.17	0.35	0.17	0.16
15.	0.20	0.38	0.19	0.18
16.	0.24	0.45	0.22	0.22
17.	0.28	0.52	0.26	0.25
18.	0.36	0.65	0.34	0.32
19.	0.45	0.79	0.42	0.39
20.	0.57	0.98	0.54	0.50
21.	0.75	1.22	0.70	0.65

Table 4.11: Root Mean Squared Error (Absolute Residuals)

	s_i^2	s_i
1.	0.76	0.64
2.	0.54	0.52
3.	0.41	0.44
4.	0.31	0.37
5.	0.26	0.30
6.	0.23	0.27
7.	0.21	0.22
8.	0.19	0.20
9.	0.16	0.17
10.	0.15	0.17
11.	0.14	0.15
12.	0.15	0.16
13.	0.17	0.16
14.	0.20	0.18
15.	0.23	0.20
16.	0.26	0.23
17.	0.29	0.26
18.	0.34	0.32
19.	0.41	0.40
20.	0.51	0.50
21.	0.65	0.64

Table 4.12: Root Mean Squared Error (Sample Var and Std)

	$\tilde{q}_i^2$	$\lvert\tilde{q}_i\rvert$	$\check{q}_i^2$	$\lvert\check{q}_i\rvert$
1.	0.97	0.72	0.91	0.92
2.	0.81	0.58	0.77	0.77
3.	0.68	0.50	0.64	0.64
4.	0.59	0.41	0.54	0.52
5.	0.50	0.34	0.42	0.40
6.	0.42	0.27	0.34	0.33
7.	0.34	0.22	0.29	0.26
8.	0.27	0.19	0.24	0.21
9.	0.24	0.18	0.22	0.17
10.	0.21	0.17	0.21	0.17
11.	0.20	0.16	0.21	0.16
12.	0.19	0.17	0.21	0.18
13.	0.21	0.18	0.23	0.20
14.	0.25	0.22	0.25	0.26
15.	0.30	0.25	0.29	0.30
16.	0.34	0.30	0.33	0.37
17.	0.38	0.34	0.37	0.45
18.	0.42	0.40	0.42	0.55
19.	0.47	0.48	0.48	0.65
20.	0.53	0.57	0.58	0.77
21.	0.60	0.71	0.72	0.92

Table 4.13: Root Mean Squared Error (Difference Schemes)

Squared Residuals	r_i^2 0.353	$\log(r_i^2)$ 0.670	$r_i^{\frac{2}{3}}$ 0.344	$\tilde{r}_i^2$ 0.334
Absolute Residuals	$\lvert r_i\rvert$ 0.368	$\log\lvert r_i\rvert$ 0.638	$\lvert r_i\rvert^{\frac{2}{3}}$ 0.345	$\lvert\tilde{r}_i\rvert$ 0.322
Sample Var & Std	s_i^2 0.321	s_i 0.318		
Difference Schemes	$\tilde{g}_i^2$ 0.434	$\lvert\tilde{g}_i\rvert$ 0.428	$\breve{g}_i^2$ 0.437	$\lvert\breve{g}_i\rvert$ 0.393

Table 4.14: Root Mean Integrated Squared Error

Now the problem is to compare the results of the residuals, sample variances, sample standard deviations and difference schemes in order to get a general perspective and draw some practical conclusions. In this respect a natural idea is to disregard the estimates based on squared residuals and sample variances since they are seen to be outperformed by the corresponding estimates based on absolute residuals and sample standard deviations respectively. Then scrutinizing the biases, the square root of the mean squared errors and the square root of the mean integrated squared errors, it is seen that smoothing sample standard deviations is almost always more efficient and less biased than smoothing either the absolute residuals or the difference schemes.

4.11 Conclusion

Based on these simulations, the results indicate that in practice where replication is involved, smoothing sample standard deviations might lead to better results. There are other situations where replication is not possible, then in such cases smoothing absolute residuals would serve as an alternative to sample standard deviations.

Chapter 5

SOME ASYMPTOTIC THEORY FOR VARIANCE FUNCTION SMOOTHING

5.1 Introduction

In smoothing functions, most authors assert that the choice of the kernel function is of relatively small importance. What matters is the choice of the smoothing parameter. However although the theory of selecting this parameter is widely expanding, there is yet no one particular method that is universally acceptable as the standard. Here, the contribution of both the bias and variance of the smoother to its mean squared error is investigated. The objective is to seek a standard criterion for selecting the smoothing parameter and to establish the validity of the claim that the choice of the kernel function does not matter. Both the homoscedastic and heteroscedastic regression cases are considered. Lengthy calculations are conducted on MAPLE. This is indicated in the text. The more elementary calculations will be checked by hand. These matters are explained in greater detail in the

Appendix together with a typical example.

5.2 Homoscedastic Regression

Consider the regression model

$$y_i = f(x_i) + e_i \quad , \quad i = 1, \cdots, N$$

We shall assume that the observations are independent and normally distributed with mean zero and constant variance σ^2. Our kernel smoother of $g(x_i)$ is

$$\hat{g}(x_i) = \frac{1}{b} \sum_{j=1}^{N} \int_{s_{j-1}}^{s_j} z_j K(\frac{x_i - u}{b}) du$$

where, $s_0 = x_1$, $s_N = x_N$, $s_j = \frac{x_j + x_{j+1}}{2}$ and $1 \leq j \leq N-1$. The kernel function is

$$K(x) = \frac{3}{4}(1 - x^2) \quad , \quad |x| \leq 1$$

and z_j are either the squared residuals, sample variance or difference schemes discussed in Chapter 4. We derive the bias, variance and mean square error associated with $\hat{g}(x_i)$ and investigate them leading to some simple criterion for selecting the parameter b. Now, write the expectation of $\hat{g}(x_i)$ as

$$E[\hat{g}(x_i)] = \frac{1}{b} \sum_{j=1}^{N} \int_{s_{j-1}}^{s_j} E(z_j) \frac{3}{4} [1 - (\frac{x_i - u}{b})^2] du \tag{5.1}$$

Then, due to our assumption, the z_j have a chi-squared distribution with one degree of freedom. Thus $E(z_j) = \frac{E(y_i^2)}{\sigma^2} = 1$. Therefore (5.1) becomes

$$E[\hat{g}(x_i)] = \frac{1}{b}\sum_{j=1}^{N}\int_{s_{j-1}}^{s_j}\frac{3\sigma^2}{4}[1-(\frac{x_i-u}{b})^2]du$$

$$= \frac{\sigma^2}{b}\sum_{j=1}^{N}[\frac{3}{4}\{u+\frac{b}{3}(\frac{x_i-u}{b})^3\}]_{s_{j-1}}^{s_j}$$

$$= \frac{3\sigma^2}{4b}\sum_{j=1}^{N}[\{s_j+\frac{b}{3}(\frac{x_i-s_j}{b})^3\}-\{s_{j-1}+\frac{b}{3}(\frac{x_i-s_{j-1}}{b})^3\}]$$

$$= \frac{3\sigma^2}{4b}\sum_{j=1}^{N}[s_j-s_{j-1}+\frac{1}{3b^2}\{(x_i-s_j)^3-(x_i-s_{j-1})^3\}] \quad (5.2)$$

Now from the relation $s_j = \frac{x_j+x_{j+1}}{2}$ we see that $s_j = x_j + \frac{1}{2}h$ where $h = x_{i+1} - x_i$. Similarly from $s_{j-1} = \frac{x_{j-1}+x_j}{2}$ see that $s_{j-1} = x_j - \frac{1}{2}h$. Therefore write (5.2) as

$$\frac{3\sigma^2}{4b}\sum_{j=1}^{N}[h+\frac{1}{3b^2}\{(x_i-x_j-\frac{1}{2}h)^3-(x_i-x_j+\frac{1}{2}h)^3\}] \quad (5.3)$$

For i fixed and $j = 1, \cdots, N$

$$x_i - x_j = (i-j)h$$

Therefore write (5.3) as

$$\frac{3\sigma^2}{4b}\sum_{j=1}^{N}[h + \frac{1}{3b^2}\{[(i-j)h - \frac{1}{2}h]^3 - [(i-j)h + \frac{1}{2}h]^3\}]$$

$$= \frac{3\sigma^2}{4b}\sum_{j=1}^{N}[h + \frac{h^3}{3b^2}\{(i-j-\frac{1}{2})^3 - (i-j+\frac{1}{2})^3\}]$$

$$= \frac{3\sigma^2}{4b}\sum_{j=1}^{N}[h - \frac{h^3}{12b^2}\{12(i-j)^2 + 1\}]$$

$$= \frac{3\sigma^2}{4b}\sum_{j=1}^{N}[h - \frac{h^3}{b^2}\{i^2 - 2ij + j^2 + \frac{1}{12}\}] \qquad (5.4)$$

Now put the sum inside to obtain

$$\frac{3\sigma^2}{4b}[Nh - \frac{h^3}{b^2}\{Ni^2 - 2i\sum_{j=1}^{N} j + \sum_{j=1}^{N} j^2 + \frac{N}{12}\}]$$

$$= \frac{3\sigma^2}{4b}[Nh - \frac{h^3}{b^2}\{Ni^2 - \frac{2iN(N+1)}{2} + \frac{N(N+1)(2N+1)}{6} + \frac{N}{12}\}]$$

$$= \frac{3\sigma^2}{4b}[Nh - \frac{h^3}{12b^2}\{12Ni^2 - 12iN(N+1) + 2N(N+1)(2N+1) + N\}]$$

$$= \sigma^2[\frac{Nh}{16b^3}(12b^2 - 12i^2h^2 + 12iNh^2 + 12ih^2 - 4N^2h^2 - 6Nh^2 - 3h^2)]$$

$$= \sigma^2[\frac{Nh}{16b^3}\{12b^2 - 4N^2h^2 + h^2(12i - 12i^2 - 3) + Nh^2(12i - 6)\}] \qquad (5.5)$$

Let $Nh = L$ implying that $h^2 = \frac{L^2}{N^2}$ and $Nh^2 = \frac{L^2}{N}$. Then the bias is $E[\hat{g}(x_i)] - \sigma^2$

$$= \sigma^2[\frac{L}{16b^3}\{12b^2 - 4L^2 + \frac{L^2(12i-6)}{N} + \frac{L^2(12i - 12i^2 - 3)}{N^2}\} - 1]$$

$$= \sigma^2[\frac{3L}{4b} - \frac{L^3}{4b^3} - 1 + \frac{L^3(12i-6)}{16b^3N} + \frac{L^3(12i - 12i^2 - 3)}{16b^3N^2}] \qquad (5.6)$$

Let $t = \frac{L}{b}$ and write (5.6) as

$$\sigma^2[\frac{3t}{4} - \frac{t^3}{4} - 1 + \frac{t^3(12i-6)}{16N} + \frac{t^3(12i - 12i^2 - 3)}{16N^2}]$$

Thus

$$\text{bias}[\hat{g}(x_i)] = \frac{\sigma^2}{16}[12t - 4t^3 - 16 + \frac{6t^3(2i-1)}{N} + \frac{3t^3(4i - 4i^2 - 1)}{N^2}] \qquad (5.7)$$

Next, write

$$\text{var}[\hat{g}(x_i)] = \frac{1}{b^2}\sum_{j=1}^{N}\text{var}(z_j)(\int_{s_{j-1}}^{s_j}\frac{3}{4}[1 - (\frac{x_i - u}{b})^2]du)^2$$

$$= \frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(\{s_j + \frac{b}{3}(\frac{x_i - s_j}{b})^3\} - \{s_{j-1} + \frac{b}{3}(\frac{x_i - s_{j-1}}{b})^3\})^2$$

$$= \frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(s_j - s_{j-1} + \frac{1}{3b^2}\{(x_i - s_j)^3 - (x_i - s_{j-1})^3\})^2$$

$$= \frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(h + \frac{1}{3b^2}\{(x_i - x_j - \frac{1}{2}h)^3 - (x_i - x_j + \frac{1}{2}h)^3\})^2 \qquad (5.8)$$

Then we use the relation $x_i - x_j = (i-j)h$ and simplify to obtain

$$\frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(h - \frac{h^3}{b^2}\{i^2 - 2ij + j^2 + \frac{1}{12}\})^2$$

$$= \frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(h^2 - \frac{2h^4}{b^2}\{i^2 - 2ij + j^2 + \frac{1}{12}\} + \frac{h^6}{b^4}\{i^2 - 2ij + j^2 + \frac{1}{12}\}^2)$$

$$= \frac{9\sigma^4}{8b^2}\sum_{j=1}^{N}(h^2 - \frac{2h^4}{b^2}\{i^2 - 2ij + j^2 + \frac{1}{12}\} + \frac{h^6}{b^4}\{i^4 - 4i^3j + 6i^2j^2 + \frac{i^2}{6} - 4ij^3 - \frac{ij}{3} + j^4 + \frac{j^2}{6} + \frac{1}{144}\})$$

$$= \frac{9\sigma^4}{8b^2}(Nh^2 - \frac{2h^4}{b^2}\{Ni^2 - 2i\sum_{j=1}^{N} j + \sum_{j=1}^{N} j^2 + \frac{N}{12}\} + \frac{h^6}{b^4}\{Ni^4 - 4i^3\sum_{j=1}^{N} j + 6i^2\sum_{j=1}^{N} j^2$$
$$+ \frac{Ni^2}{6} - 4i\sum_{j=1}^{N} j^3 - \frac{i}{3}\sum_{j=1}^{N} j + \sum_{j=1}^{N} j^4 + \frac{1}{6}\sum_{j=1}^{N} j^2 + \frac{N}{144}\})$$

It was most convenient to evaluate this expression using maple because expressions like $\sum_{j=1}^{N} j^r$ can be evaluated exactly using the maple summation command. There is confidence that the command can be applied to the whole formula. Thus we obtain

$$\frac{9\sigma^4}{8b^2}(Nh^2-\frac{2h^4}{b^2}\{Ni^2-\frac{2iN(N+1)}{2}+\frac{N(N+1)(2N+1)}{6}+\frac{N}{12}\}$$
$$+\frac{h^6}{b^4}\{Ni^4-\frac{4i^3N(N+1)}{2}+\frac{6i^2N(N+1)(2N+1)}{6}+\frac{Ni^2}{6}-\frac{4iN^2(N+1)^2}{4}$$
$$-\frac{iN(N+1)}{6}+\frac{N(N+1)(2N+1)(3N^2+3N-1)}{30}+\frac{N(N+1)(2N+1)}{36}+\frac{N}{144}\})$$

$$=\frac{9\sigma^4}{8b^2}(Nh^2-\frac{2Nh^4}{b^2}\{i^2-i(N+1)+\frac{(N+1)(2N+1)}{6}+\frac{1}{12}\}$$
$$+\frac{Nh^6}{b^4}\{i^4-2i^3(N+1)+i^2(N+1)(2N+1)+\frac{i^2}{6}-iN(N+1)^2-\frac{i(N+1)}{6}$$
$$+\frac{(N+1)(2N+1)(3N^2+3N-1)}{30}+\frac{(N+1)(2N+1)}{36}+\frac{1}{144}\})$$

$$=\frac{9Nh^2\sigma^4}{8b^6}(b^4-2b^2h^2\{i^2-i(N+1)+\frac{(N+1)(2N+1)}{6}+\frac{1}{12}\}$$
$$+h^4\{i^4-2i^3(N+1)+i^2(N+1)(2N+1)+\frac{i^2}{6}-iN(N+1)^2-\frac{i(N+1)}{6}$$
$$+\frac{(N+1)(2N+1)(3N^2+3N-1)}{30}+\frac{(N+1)(2N+1)}{36}+\frac{1}{144}\})$$

$$=\frac{9Nh^2\sigma^4}{8b^6}(b^4-2b^2h^2i^2+2b^2h^2iN+2b^2h^2i-\frac{2b^2h^2N^2}{3}-b^2h^2N-\frac{b^2h^2}{3}-\frac{b^2h^2}{6}+h^4i^4$$
$$-2h^4i^3N-2h^4i^3+2h^4N^2i^2+3h^4Ni^2+h^4i^2+\frac{h^4i^2}{6}-h^4N^3i-2h^4N^2i-h^4iN$$
$$-\frac{h^4iN}{6}-\frac{h^4i}{6}+\frac{h^4N^4}{5}+\frac{h^4N^3}{2}+\frac{h^4N^2}{3}-\frac{h^4}{30}+\frac{h^4N^2}{18}+\frac{h^4N}{12}+\frac{h^4}{36}+\frac{h^4}{144})$$

$$
\begin{aligned}
= & \frac{9Nh^2\sigma^4}{8b^6}(b^4 - \frac{2b^2h^2N^2}{3} + h^2\{2b^2i - 2b^2i^2 - \frac{b^2}{3} - \frac{b^2}{6}\} + h^2N\{2b^2i - b^2\} \\
& +h^4\{i^4 - 2i^3 + i^2 + \frac{i^2}{6} - \frac{i}{6} - \frac{1}{30} + \frac{1}{36} + \frac{1}{144}\} + h^4N\{3i^2 - 2i^3 - i - \frac{i}{6} + \frac{1}{12}\} \\
& +h^4N^2\{2i^2 - 2i + \frac{1}{3} + \frac{1}{18}\} + h^4N^3\{\frac{1}{2} - i\} + \frac{h^4N^4}{5})
\end{aligned}
$$

Now use the relation $Nh = L$ and obtain

$$
\begin{aligned}
& \frac{9L^2\sigma^4}{8b^6N}(b^4 - \frac{2b^2L^2}{3} + \frac{L^2}{N^2}\{2b^2i - 2b^2i^2 - \frac{b^2}{3} - \frac{b^2}{6}\} + \frac{L^2}{N}\{2b^2i - b^2\} \\
+ & \frac{L^4}{N^4}\{i^4 - 2i^3 + i^2 + \frac{i^2}{6} - \frac{i}{6} - \frac{1}{30} + \frac{1}{36} + \frac{1}{144}\} + \frac{L^4}{N^3}\{3i^2 - 2i^3 - i - \frac{i}{6} + \frac{1}{12}\} \\
+ & \frac{L^4}{N^2}\{2i^2 - 2i + \frac{1}{3} + \frac{1}{18}\} + \frac{L^4}{N}\{\frac{1}{2} - i\} + \frac{L^4}{5})
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{9\sigma^4}{8N}(\frac{L^2}{b^2} - \frac{2L^4}{3b^4} + \frac{L^4}{b^4N^2}\{2i - 2i^2 - \frac{1}{3} - \frac{1}{6}\} + \frac{L^4}{b^4N}\{2i - 1\} \\
& +\frac{L^6}{b^6N^4}\{i^4 - 2i^3 + i^2 + \frac{i^2}{6} - \frac{i}{6} - \frac{1}{30} + \frac{1}{36} + \frac{1}{144}\} + \frac{L^6}{b^6N^3}\{3i^2 - 2i^3 - i - \frac{i}{6} + \frac{1}{12}\} \\
& +\frac{L^6}{b^6N^2}\{2i^2 - 2i + \frac{1}{3} + \frac{1}{18}\} + \frac{L^6}{b^6N}\{\frac{1}{2} - i\} + \frac{L^6}{5b^6})
\end{aligned}
$$

Substitute $t = \frac{L}{b}$ obtaining

$$
\begin{aligned}
& \frac{9\sigma^4}{8N}(t^2 - \frac{2}{3}t^4 + \frac{1}{5}t^6 + \frac{t^4}{N^2}\{2i - 2i^2 - \frac{1}{3} - \frac{1}{6}\} + \frac{t^4}{N}\{2i - 1\} \\
+ & \frac{t^6}{N^4}\{i^4 - 2i^3 + i^2 + \frac{i^2}{6} - \frac{i}{6} - \frac{1}{30} + \frac{1}{36} + \frac{1}{144}\} + \frac{t^6}{N^3}\{3i^2 - 2i^3 - i - \frac{i}{6} + \frac{1}{12}\} \\
+ & \frac{t^6}{N^2}\{2i^2 - 2i + \frac{1}{3} + \frac{1}{18}\} + \frac{t^6}{N}\{\frac{1}{2} - i\}
\end{aligned}
$$

$$= \frac{9\sigma^4}{8}[\frac{t^2(3t^4-10t^2+15)}{15N} + \frac{t^4(t^2(1-2i)+4i-2)}{2N^2}$$
$$+\frac{t^4(t^2(36i^2-36i+7)+9(4i-4i^2-1))}{18N^3} + \frac{t^6(36i^2-24i^3-24i+1)}{12N^4}$$
$$+\frac{t^6(720i^4-1440i^3+840i^2-120i+1)}{720N^5}]$$

Simplify on maple and obtain

$$\text{var}[\hat{g}(x_i)] = \frac{\sigma^4}{640}[\frac{48t^2(3t^4-10t^2+15)}{N} + \frac{360t^4(t^2(1-2i)+4i-2)}{N^2}$$
$$+ \quad \frac{40t^4(t^2(36i^2-36i+7)+9(4i-4i^2-1))}{N^3} + \frac{60t^6(36i^2-24i^3-24i+1)}{N^4}$$
$$+ \quad \frac{t^6(720i^4-1440i^3+840i^2-120i+1)}{N^5}] \tag{5.9}$$

Now square (5.7) and add to (5.9) obtaining the mean square error of $\hat{g}(x_i)$ as

$$\frac{\sigma^4}{1280}[80(3t-t^3-4)^2 + \frac{240t^3(3t-t^3-4)(2i-1)+96t^2(3t^4-10t^2+15)}{N}$$
$$+ \quad \frac{60t^3\{3t^2(2i-1)^2+2(3t-t^3-4)(4i-4i^2-1)\}+720t^4(t^2(1-2i)+4i-2)}{N^2}$$
$$+ \quad \frac{180t^6(2i-1)(4i-4i^2-1)+80t^4(t^2(36i^2-36i+7)+9(4i-4i^2-1))}{N^3}$$
$$+ \quad \frac{45t^6(4i-4i^2-1)^2+120t^6(36i^2-24i^3-24i+1)}{N^4}$$
$$+ \quad \frac{2t^6(720i^4-1440i^3+840i^2-120i+1)}{N^5}] \tag{5.10}$$

It is expected that at the middle of the smoothing region, all the three discrepancy measures, the bias, the variance and the mean squared error should be minimal. The center of the smoothing region is given by $i = \frac{N+1}{2}$. Substituting into (5.7), (5.9) and

(5.10) using maple respectively we obtain

$$\text{bias}[\hat{g}(x_{\frac{N+1}{2}})] = \frac{\sigma^2}{16}(12t - t^3 - 16) = -\frac{\sigma^2}{16}(t+4)(t-2)^2 \quad , \tag{5.11}$$

$$\text{var}[\hat{g}(x_{\frac{N+1}{2}})] = \frac{\sigma^4}{640}[\frac{3(3t^6 - 40t^4 + 240t^2)}{N} - \frac{410t^6}{N^3} + \frac{60t^6}{N^4} - \frac{74t^6}{N^5}] \tag{5.12}$$

and

$$MSE[\hat{g}(x_{\frac{N+1}{2}})] = \frac{\sigma^4}{1280}[5(12t - t^3 - 16)^2 + \frac{6(3t^6 - 40t^4 + 240t^2)}{N} - \frac{820t^6}{N^3} + \frac{120t^6}{N^4} - \frac{148t^6}{N^5}] \tag{5.13}$$

Now from (5.12) see that the variance is $O(\frac{1}{N})$ and vanishes as $N \to \infty$. However from (5.11) we see that the bias at $i = \frac{N+1}{2}$ does not depend on N at all. Equating the leading term of (5.11) to zero gives $t = 2$ and $L = 2b$. We investigate three more kernels obtaining the leading term of the asymptotic bias at the center of the smoothing region as in Table 5.1. Observe that equating the leading term to zero and solving gives $t = 2$. Thus the relation between $L = Nh$ and the bandwidth parameter b is same for all these kernels. Note that the pleasing factorizations in Table 5.1 formulae were discovered using the maple factorization command and checked by hand.

5.3 Heteroscedastic Regression

In this case the assumptions adopted in the homoscedastic regression case carry over without alteration. The only difference is that the data have nonconstant variance. In the derivation procedure the difference becomes visible in the summing stage. Thus (5.4) becomes

$$E[\hat{g}(x_i)] = \frac{3}{4b^3}\sum_{j=1}^{N} g(x_j)[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \tag{5.14}$$

Now assume that $g(x_i)$ is smooth and expand at x_i in a Taylor expansion to give

$$g(x_j) = g(x_i) + (x_j - x_i)g'(x_i) + \frac{(x_j - x_i)^2 g''(x_i)}{2} + \cdots$$

and from (5.14) approximate $E[\hat{g}(x_i)]$ by

$$\frac{3}{4b^3}\sum_{j=1}^{N}[g(x_i) + (x_j - x_i)g'(x_i) + \frac{(x_j - x_i)^2 g''(x_i)}{2}][b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}]$$

$$\begin{aligned} = & \frac{3}{4b^3}\sum_{j=1}^{N} g(x_i)[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \\ & + \frac{3}{4b^3}\sum_{j=1}^{N}(x_j - x_i)g'(x_i)[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \\ & + \frac{3}{8b^3}\sum_{j=1}^{N}(x_j - x_i)^2 g''(x_i)[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \end{aligned} \tag{5.15}$$

Then, since $x_i - x_j = (i - j)h$, (5.15) becomes

$$\begin{aligned} E[\hat{g}(x_i)] = & \frac{3g(x_i)}{4b^3}\sum_{j=1}^{N}[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \\ & - \frac{3g'(x_i)}{4b^3}\sum_{j=1}^{N}(i - j)h[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \end{aligned}$$

$$+\frac{3g''(x_i)}{8b^3}\sum_{j=1}^{N}(i-j)^2h^2[b^2h - h^3i^2 + 2h^3ij - h^3j^2 - \frac{h^3}{12}] \qquad (5.16)$$

The bias at x_i is $E[\hat{g}(x_i)] - g(x_i)$. Then the coefficient of $g(x_i)$ in (5.16) is exactly that of σ^2 in (5.7) namely

$$\frac{g(x_i)}{16}[12t - 4t^3 - 16 + \frac{6t^3(2i-1)}{N} + \frac{3t^3(4i-4i^2-1)}{N^2}] \qquad (5.17)$$

Similarly the second term of (5.16) becomes

$$\frac{g'(x_i)h^2N}{32b^3}[-24ib^2 + 24h^2i^3 - 36h^2i^2N - 36h^2i^2 + 24ih^2N^2 + 36ih^2N + 14ih^2$$
$$+ \quad 12Nb^2 + 12b^2 - 6h^2N^3 - 12h^2N^2 - 7h^2N - h^2]$$

$$= \frac{g'(x_i)L}{32}[-\frac{24iL}{bN} + \frac{24i^3L^3}{b^3N^3} - \frac{36i^2L^3}{b^3N^2} - \frac{36i^2L^3}{b^3N^3} + \frac{24iL^3}{b^3N} + \frac{36iL^3}{b^3N^2} + \frac{14iL^3}{b^3N^3} + \frac{12L}{b}$$
$$+\frac{12L}{bN} - \frac{6L^3}{b^3} - \frac{12L^3}{b^3N} - \frac{7L^3}{b^3N^2} - \frac{L^3}{b^3N^3}]$$

and simplifying further in maple it becomes

$$\frac{g'(x_i)L}{32}[6t(2-t^2) + \frac{12t(t-1)(t+1)(2i-1)}{N} - \frac{t^3[7+36i(i-1)]}{N^2}$$
$$+ \quad \frac{t^3(2i-1)(12i^2-12i+1)}{N^3}] \qquad (5.18)$$

Likewise, for the last term of (5.16) obtain

$$\frac{-g''(x_i)h^3N}{960b^3}[-360i^2b^2 + 360i^4h^2 + 390h^2i^2 - 720i^3h^2N - 720i^3h^2 + 720i^2h^2N^2$$
$$+ \quad 1080i^2h^2N + 360iNb^2 + 360ib^2 - 120N^2b^2 - 180Nb^2 - 60b^2 + 72h^2N^4 + 180h^2N^3$$
$$+ \quad 130h^2N^2 - 7h^2 - 360ih^2N^3 - 720ih^2N^2 - 390ih^2N - 30ih^2 + 15h^2N]$$

$$= \frac{-g''(x_i)L^2}{960}[\frac{-360i^2L}{bN^2} + \frac{360i^4L^3}{b^3N^4} + \frac{390i^2L^3}{b^3N^4} - \frac{720i^3L^3}{b^3N^3} - \frac{720i^3L^3}{b^3N^4} + \frac{720i^2L^3}{b^3N^2}$$
$$+\frac{1080i^2L^3}{b^3N^3} + \frac{360iL}{bN} + \frac{360iL}{bN^2} - \frac{120L}{b} - \frac{180L}{bN} - \frac{60L}{bN^2} + \frac{72L^3}{b^3} + \frac{180L^3}{b^3N}$$
$$+\frac{130L^3}{b^3N^2} - \frac{7L^3}{b^3N^4} - \frac{360iL^3}{b^3N} - \frac{720iL^3}{b^3N^2} - \frac{390iL^3}{b^3N^3} - \frac{30iL^3}{b^3N^4} + \frac{15L^3}{b^3N^3}]$$

and simplify in maple to get

$$\frac{-g''(x_i)L^2}{960}\{24t(3t^2-5) - \frac{180t(t-1)(t+1)(2i-1)}{N}$$
$$+ \quad \frac{10t[36i(2t^2-1)(i-1)+13t^2-6]}{N^2} - \frac{15t^3(2i-1)(24i^2-24i+1)}{N^3}$$
$$+ \quad \frac{t^3[30i(i-1)(12i^2-12i+1)-7]}{N^4}\} \tag{5.19}$$

Now sum (5.17), (5.18) and (5.19) to obtain

$$\text{bias}[\hat{g}(x_i)] = (\frac{3}{4}A + \frac{3}{8}B + \frac{1}{8}C)t - (\frac{1}{4}A + \frac{3}{16}B + \frac{3}{40}C)t^3 - A$$
$$+ \quad 6tZ(2i-1)[(\frac{1}{16}A + \frac{1}{16}B + \frac{1}{32}C)t^2 - \frac{1}{16}B - \frac{1}{32}C]$$
$$+ \quad tZ^2[\frac{1}{16}C - (\frac{3}{16}A + \frac{7}{32}B + \frac{13}{96}C)t^2 - 12i(i-1)\{(\frac{1}{16}A + \frac{3}{32}B + \frac{1}{16}C)t^2 - \frac{1}{32}C\}]$$

$$
\begin{aligned}
&+ \quad t^3Z^3(2i-1)[\frac{3}{8}i(i-1)(B+C)+\frac{1}{32}B+\frac{1}{64}C] \\
&+ \quad \frac{Ct^3Z^4}{960}[7-30i(i-1)(12i^2-12i+1)] \qquad (5.20)
\end{aligned}
$$

where $A = g(x_i)$, $B = g'(x_i)L$, $C = g''(x_i)L^2$ and $Z = \frac{1}{N}$. These substitutions were useful in allowing maple to aid the computation. Put $i = \frac{N+1}{2}$ in (5.20) and obtain

$$
(\frac{3}{4}A+\frac{1}{32}C)t-(\frac{1}{16}A+\frac{3}{640}C)t^3-A+\frac{tCZ^2}{384}(5t^2-12)-\frac{t^3CZ^4}{120} \qquad (5.21)
$$

Ignoring $O(\frac{1}{N})$ terms the coefficient of C is $\frac{t}{32}-\frac{3t^3}{640}$ which has a minimum at $\frac{2}{3}\sqrt{5}\approx 1.5$. Next, replace σ^4 in (5.8) by

$$
\begin{aligned}
g(x_j)^2 &= g(x_i)^2+(x_j-x_i)^2g'(x_i)^2+\frac{(x_j-x_i)^4g''(x_i)^2}{4}+2g(x_i)(x_j-x_i)g'(x_i) \\
&\quad +g(x_i)(x_j-x_i)^2g''(x_i)+(x_j-x_i)g'(x_i)(x_j-x_i)^2g''(x_i)
\end{aligned}
$$

$$
\begin{aligned}
&= g(x_i)^2+(i-j)^2h^2g'(x_i)^2+\frac{(i-j)^4h^4g''(x_i)^2}{4}-2(i-j)hg(x_i)g'(x_i) \\
&\quad +(i-j)^2h^2g(x_i)g''(x_i)-(i-j)^3h^3g'(x_i)g''(x_i)
\end{aligned}
$$

to obtain

$$
\begin{aligned}
\mathrm{var}[\hat{g}(x_i)] &= \frac{9g(x_i)^2}{8b^2}\sum_{j=1}^{N}[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\} \\
&\quad +\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{9h^2g'(x_i)^2}{8b^2}\sum_{j=1}^{N}(i-j)^2[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\}\\
&+\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}]\\
&+\frac{9h^4g''(x_i)^2}{32b^2}\sum_{j=1}^{N}(i-j)^4[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\}\\
&+\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}]\\
&-\frac{9hg(x_i)g'(x_i)}{4b^2}\sum_{j=1}^{N}(i-j)[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\}\\
&+\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}]\\
&+\frac{9h^2g(x_i)g''(x_i)}{8b^2}\sum_{j=1}^{N}(i-j)^2[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\}\\
&+\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}]\\
&-\frac{9h^3g'(x_i)g''(x_i)}{8b^2}\sum_{j=1}^{N}(i-j)^3[h^2-\frac{2h^4}{b^2}\{i^2-2ij+j^2+\frac{1}{12}\}\\
&+\frac{h^6}{b^4}\{i^4-4i^3j+6i^2j^2+\frac{i^2}{6}-4ij^3-\frac{ij}{3}+j^4+\frac{j^2}{6}+\frac{1}{144}\}] \qquad (5.22)
\end{aligned}
$$

Work out the first term of (5.22) to obtain

$$
\begin{aligned}
&\frac{g(x_i)^2h^2N}{640b^6}[720b^4+h^4-1440h^2i^2b^2-360h^2b^2+840h^4i^2+720h^4i^4-480h^2N^2b^2\\
&-\quad 720h^2Nb^2+144h^4N^4+360h^4N^3+280h^4N^2+1440h^2iNb^2+1440h^2ib^2\\
&-\quad 1440h^4i^3N-1440h^4i^3+1440h^4i^2N^2+2160h^4i^2N-720h^4iN^3-1440h^4iN^2\\
&-\quad 840h^4iN-120h^4i+60h^4N]
\end{aligned}
$$

These particularly difficult computations had proved prohibitively hard before maple was used for the remainder of this section. Simplify up to the term upon $\frac{1}{N^2}$ to get

$$\frac{g(x_i)^2}{640}[48t^2Z(15-10t^2+3t^4)-360t^4Z^2(t^2-2)(2i-1)] \tag{5.23}$$

The second term of (5.22) becomes

$$\begin{aligned}
&\frac{g'(x_i)^2h^4N}{26880b^6}[9870h^4iN+1176h^2b^2-2520h^2b^2N+5040b^4-90720h^4iN^4-30240ib^4 \\
+&\ 587h^4-30240h^4iN^5+15120Nh^4i^2+226800i^4h^4N+80640h^4i^4-9870h^4i^2 \\
+&\ 30240i^6h^4-10080h^4+16128h^4N^4+2520h^4N^3-3290h^4N^2+105h^4N \\
+&\ 15120h^4N^5+4320h^4N^6+10080b^4N^2+15120b^4N+30240i^2b^4-65520h^2i^2b^2 \\
-&\ 60480i^4h^2b^2-161280h^4i^3N+161280h^4i^2N^2-210h^4i-80640h^4iN^3-10080h^4iN^2 \\
-&\ 90720i^5h^4-90720i^5h^4N+151200h^4i^4N^2+90720h^4i^2N^4+226800h^4i^2N^3 \\
-&\ 151200h^4i^3N-302400h^4i^3N^2-21840h^2N^2-12096h^2b^2N^4-30240h^2b^2N^3 \\
-&\ 30240ib^4N+5040h^2ib^2+65520h^2ib^2N+120960i^3h^2b^2+120960i^3h^2b^2N \\
-&\ 120960h^2i^2b^2N^2-181440h^2i^2b^2N+60480h^2ib^2N^3+120960h^2ib^2N^2]
\end{aligned}$$

simplifying up to the term upon $\frac{1}{N^2}$ to

$$\frac{g'(x_i)^2L^2}{26880}[288t^2Z(35-42t^2+15t^4)-15120t^2Z^2(t-1)^2(t+1)^2(2i-1)] \tag{5.24}$$

The third term of (5.22) becomes

$$\frac{g''(x_i)^2 h^6 N}{26880b^6}[15120h^4N^7 + 2520h^4N^5 - 895h^4 - 1008b^4 - 57750h^4 + 20880h^4N^6$$

$$+ \quad 17850h^4i^2 + 105h^4N^3 + 146160h^4i^6 + 30240h^4i^8 - 11550h^4N^4$$

$$- \quad 420h^4i^3 - 282240h^4i^3N^5 + 120960h^4i^2N^6 - 30240h^4iN^7 + 282240h^4i^6N^2$$

$$+ \quad 423360h^4i^6N - 120960h^4i^7N - 120960h^4i^7 - 15120h^4i^5 + 423360h^4i^4N^4$$

$$+ \quad 1058400h^4i^4N^3 - 423360h^4i^5N^3 + 5950h^4N^2 + 3360h^4N^8 - 17850h^4iN$$

$$- \quad 846720h^4i^5N^2 - 438480h^4i^5N + 730800h^4i^4N^2 + 37800h^4i^4N - 846720h^4i^3N^4$$

$$- \quad 730800h^4i^3N^3 - 50400h^4i^3N^2 + 115500h^4i^3N + 37800h^4i^2N^3 - 115500h^4i^2N^2$$

$$+ \quad 630h^4i^2N + 423360h^4i^2N^5 + 438480h^4i^2N^4 - 120960h^4iN^6 - 146160h^4iN^5$$

$$- \quad 15120h^4iN^4 - 420h^4iN^2 + 57750h^4iN^3 - 2520h^2N^3b^2 - 61752h^2b^2$$

$$- \quad 31248h^2N^4b^2 + 8400h^2N^2b^2 + 6048N^4b^4 + 15120N^3b^4 + 10080N^2b^4$$

$$- \quad 30240h^2N^5b^2 - 8640h^2N^6b^2 + 30240i^4b^4 - 5040h^2ib^2 + 181440h^2i^5Nb^2$$

$$+ \quad 181440h^2i^5b^2 - 453600h^2i^4Nb^2 - 453600h^2i^4b^2 + 60480i^2N^2b^4 + 90720i^2Nb^4$$

$$+ \quad 30240i^2b^4 - 30240iN^3b^4 - 60480iN^2b^4 - 30240iNb^4 - 60480i^3Nb^4$$

$$- \quad 60480i^3b^4 + 302400h^2i^3N^3b^2 + 604800h^2i^3N^2b^2 + 312480h^2i^3Nb^2 + 10080h^2i^3b^2$$

$$- \quad 312480h^2i^2N^2b^2 - 15120h^2i^2Nb^2 + 25200h^2i^2b^2 - 181440h^2i^2N^4b^2 - 453600h^2i^2N^3b^2$$

$$+ \quad 60480h^2iN^5b^2 + 181440h^2iN^4b^2 + 156240h^2iN^3b^2 - 25200h^2iNb^2 + 10080h^2iN^2b^2]$$

which simplifies up to the term upon $\frac{1}{N^2}$ to

$$\frac{g''(x_i)^2L^4}{107520}[96t^2Z(63-90t^2+35t^4)-3024t^2Z^2\{10t^2-5t^4-5+i(10-20t^2+t^4)\}] \quad (5.25)$$

The fourth term of (5.22) becomes

$$-\frac{g(x_i)g'(x_i)h^3N}{640b^6}[720h^4i^4N-1920h^4i^3N^2+600h^2Nb^2-720h^2N^3b^2-720b^4+3h^4$$

$$+\ 1440h^2i^2b^2-480h^2N^2b^2-175h^4N-60h^4N^3+432h^4N^4-440h^4N^2$$

$$-\ 4200h^4i^2N-2880h^4i^2N^2-2880h^4iN^3+240h^4N^5+720Nb^4+600h^2b^2$$

$$-\ 1560h^4i^2+6112h^4i-720h^4i^4+1920h^4i^3-2080h^4iN^2-1920h^2ib^2$$

$$-\ 1152h^4iN^4+8640i^2h^2N^2b^2-1440h^2i^2Nb^2+1920h^2iN^2b^2]$$

simplifying up to the term upon $\frac{1}{N^2}$ to

$$-\frac{g(x_i)g'(x_i)L}{640}[240t^2Z(3-3t^2+t^4)-8t^2Z^2\{90+t^2(60-541t^2-240i+144it^2)\}] \quad (5.26)$$

Notice that the fifth term of (5.22) simplifies quickly to

$$\frac{g(x_i)g''(x_i)L^2}{26880}[288t^2Z(35-42t^2+15t^4)-15120t^2Z^2(t-1)^2(t+1)^2(2i-1)] \quad (5.27)$$

Finally the last term of (5.22) becomes

$$-\frac{g'(x_i)g''(x_i)h^5N}{512b^6}[-240i^4h^4 + 4032i^3h^4N^4 - 508i^3h^4 + 2112i^5h^4 + 576i^7h^4$$

$$+ \quad 2880i^4h^2Nb^2 - 5040i^4h^4N^3 - 48h^4N^4 + 127h^4N^3 - 2h^4N^2 - 288h^4N^6 - 352h^4N^5$$

$$- \quad 3840i^3h^2N^2b^2 - 6i^2h^4 - 41h^4N - 72h^4N^7 + 480i^3h^4N + 6048i^5h^4N - 6048i^2h^4N^4$$

$$+ \quad 2880i^4h^2b^2 - 5280i^2h^4N^3 - 480i^2h^4N^2 + 762i^2h^4N + 240ih^4N^3 - 508ih^4N^2 + 6ih^4N$$

$$+ \quad 2016h^4N^5 + 2112ih^4N^4 + 10080i^3h^4N^3 + 7040i^3h^4N^2 - 10080i^4h^4N^2 - 5280i^4h^4N$$

$$- \quad 2016i^2h^4N^5 + 576ih^4N^6 + 82ih^4 + 4032i^5h^4N^2 + 504h^2N^3b^2 + 48h^2N^2b^2$$

$$- \quad 72h^2Nb^2 + 192h^2N^5b^2 + 576h^2N^4b^2 - 144N^3b^4 - 288N^2b^4 - 144Nb^4 + 576i^3b^4$$

$$- \quad 5760i^3h^2Nb^2 - 2016i^3h^2b^2 - 1152i^5h^2b^2 + 576iN^2b^4 + 864iNb^4 + 288ib^4 - 864i^2Nb^4$$

$$- \quad 864i^2b^4 + 2880i^2h^2N^3b^2 + 5760i^2h^2N^2b^2 + 3024i^2h^2Nb^2 + 144i^2h^2b^2 - 2016ih^2N^2b^2$$

$$- \quad 144ih^2Nb^2 + 144ih^2b^2 - 1152ih^2N^4b^2 - 2880ih^2N^3b^2 - 2016i^6h^4Nb^2 - 2016i^6h^4b^2]$$

and simplifies up to the term upon $\frac{1}{N^2}$ to

$$-\frac{g'(x_i)g''(x_i)L^3}{512}[24t^2Z(8t^2 - 3t^4 - 6) + 288t^2Z^2(t-1)^2(t+1)^2(2i-1)] \tag{5.28}$$

Now sum (5.23), (5.24), (5.25), (5.26), (5.27) and (5.28) to get

$$\text{var}[\hat{g}(x_i)] = [\frac{3}{160}t^2(60A^2 + 20B^2 + 3C^2 - 60AB + 20AC + 15BC)$$

$$+ \quad \frac{3}{560}t^4(-140A^2 - 84B^2 - 15C^2 + 210AB - 84AC - 70BC)$$

$$
\begin{aligned}
&+ \quad \frac{1}{2240}t^6(504A^2+360B^2+70C^2-840AB+360AC+315BC)]Z \\
&+ \quad [\frac{9}{64}t^2(4B^2+C^2+8AB+4AC+4BC-8iB^2-2iC^2-8iAC-8iBC) \\
&- \quad \frac{3}{32}t^4(12A^2+12B^2+3C^2-8AB+12AC+12BC \\
&- \quad 2i(12A^2+12B^2+3C^2-16AB+12AC+12BC)) \\
&+ \quad \frac{9}{320}t^6(20A^2+20B^2+5C^2-24AB+20AC+20BC \\
&- \quad i(40A^2+40B^2+C^2-64AB+40AC+40BC))]Z^2 \qquad (5.29)
\end{aligned}
$$

with A, B, C and Z as defined under (5.20). Next, obtain the term up to $\frac{1}{N^2}$ of the square of (5.20) as

$$
\begin{aligned}
& A^2-\frac{1}{4}tA(6A+3B+16C)+\frac{1}{64}t^2(6A+3B+C)^2+\frac{1}{40}t^3A(20A+15B+6C) \\
- \quad & \frac{3}{320}t^4(40A^2+15B^2+2C^2+50AB+560AC+11BC)+\frac{1}{6400}t^6(20A+15B+6C)^2 \\
- \quad & \frac{3}{640}tZ(20t^3A-60tA+80A+6t^3C-10tC-30tB+15t^3B) \\
\times \quad & (t^2C-C+2t^2A-2B+2t^2B)(2i-1) \\
+ \quad & Z^2[-\frac{1}{8}At(1-6i+6i^2)+\frac{1}{256}t^2(24AC+46080B+36B^2+13C^2 \\
+ \quad & 12i(5C^2i+12ACi-18BCi+12B^2i-12B^2-5C^2-12AC)) \\
+ \quad & \frac{1}{48}t^3A(18A+21B+13C+36i(i-1)(2A+3B+2C)) \\
+ \quad & \frac{1}{1920}t^4(810AC+1440AB+885BC+540A^2+855B^2+218C^2 \\
+ \quad & 18i(120A^2+210B^2+56C^2+220AC+225BC+360AB)(i-1)) \\
+ \quad & \frac{1}{3840}t^6(861BC+908AC+1770AB+900A^2+855B^2+213C^2
\end{aligned}
$$

$$+\quad 36i(-100A^2 + 100A^2 i - 105B^2 + 105B^2 i - 27C^2 + 27C^2 i - 210AB$$

$$-\quad 112AC + 210ABi + 108BCi + 112ACi))]$$

and add it to (5.29) obtaining the mean squared error.

5.4 General Discussion and Recommendations

It is seen that the variance is $O(\frac{1}{N})$ so its contribution to the mean squared error is not as significant as that of the bias as N becomes large. Therefore it looks reasonable to minimize the bias. It depends on the unknown σ^2 in the homoscedastic regression case and per our approximation on the unknown $g(x_i)$, $g'(x_i)$ and $g''(x_i)$ in the heteroscedastic regression case. A suggestion to reduce the bias is as follows. First consider the asymptotic bias model (5.7) and observe that the bias is symmetric in t about $\frac{N+1}{2}$. Thus biases are equal at points $i = 1$ and $i = N$. Substitute $i = N$ into the leading term of equation (5.7) getting $\frac{1}{16}(12t - 4t^3 - 16 + \frac{6t^3}{N} - \frac{3t^3}{N^2})$. Ignoring the $O(\frac{1}{N})$ terms we have $\frac{1}{16}(12t - 4t^3 - 16)$ which is a minimum at $t = 1$. Then, recall from (5.11) that at the center our solution was $t = 2$. Therefore for the whole smoothing region, the good value of t should lie between 1 and 2. Now observe what happens to the bias as t varies between 1 and 2 in figure 5.1. Certainly the bias is seen to vary with t. As t increases from 1 to 2 the bias decreases at the center but at the same time it increases at the ends and vice versa. To balance this kind of phenomenon consider $i = \frac{1}{2}(\frac{N+1}{2} + 1) = \frac{N+3}{4}$ which is the center of one half of the interval. Substitute into (5.7) and minimize with respect to t after ignoring the $O(\frac{1}{N})$

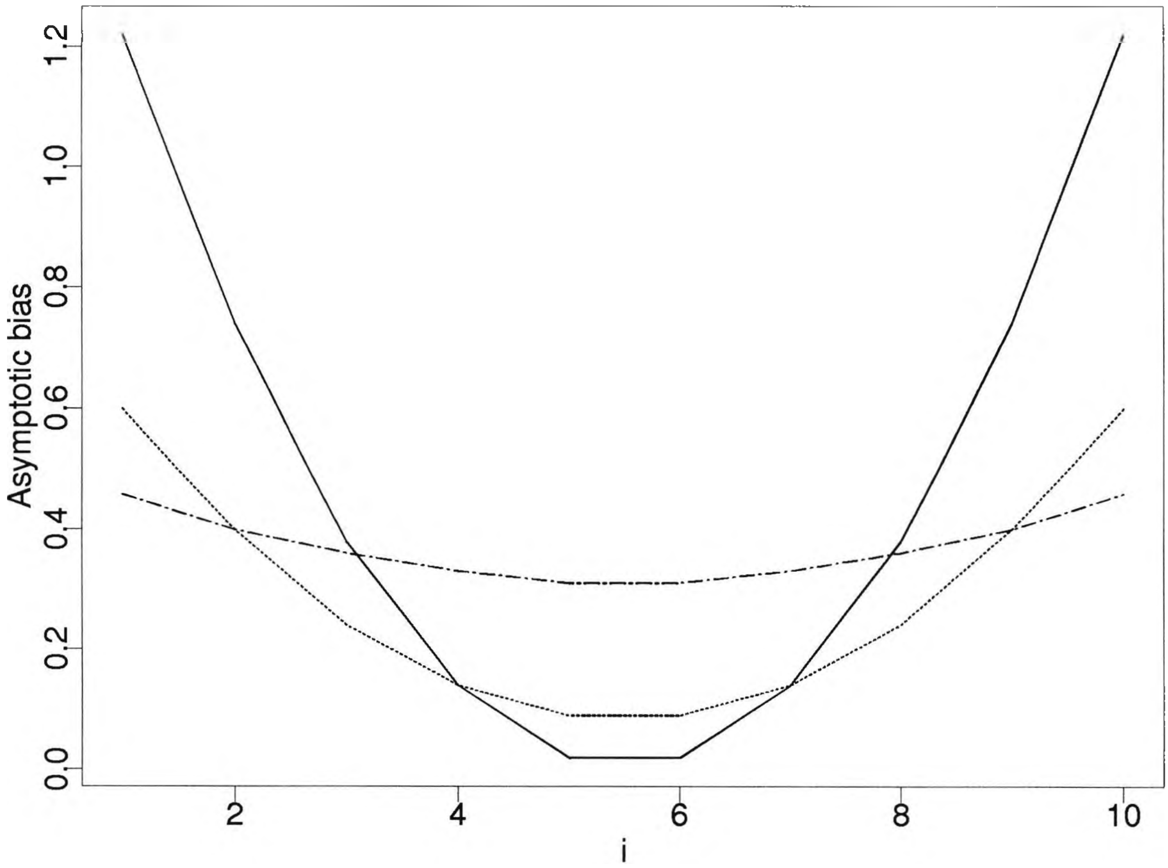

Figure 5.1: Absolute bias: $N = 10$, $t = 1$ (broken line) $t = 1.5$ (dotted line) and $t = 2$ (bold line).

terms to get $12 - \frac{21t^2}{4}$. Equate to zero and solve to obtain $t_{\frac{N+3}{4}} = \frac{4}{7}\sqrt{7} \approx 1.5$ and hence

$$b \approx \frac{2}{3}L \tag{5.30}$$

This suggests a global constant bandwidth for the homoscedastic regression case. It is an amount of smoothing which should stabilize the variation of the bias between the center and the ends of the data. It will ensure that the bias at the ends is more or less of the same order of magnitude as in the interior.

Another possible bias reduction criterion is to use local (varying) bandwidths. The idea here is to vary the bandwidth locally to smooth more at the ends where bias is very severe and to smooth less at the center where bias is minimal. To do this, use t_i instead of t to obtain b_i and see that when $t_N = 1$, $b_N = L$. This indicates that b is large at the ends of the data and therefore more smoothing there. Consequently, this implies that the bias is reduced at the ends while not affecting the bias at any other point at all. The amount of smoothing will keep reducing to the lowest value $b_{\frac{N+1}{2}} = \frac{1}{2}L$ when $t_{\frac{N+1}{2}} = 2$. Table 5.2 shows that asymptotically only three different bandwidths will be needed. Table 5.2 shows the asymptotic bias at different points together with the corresponding values of t that will minimize it at these points. Only three different values of t are obtained, one at the end points, another one at the center and finally one for all other points. Therefore for local bandwidth smoothing use $b = L$ to smooth at the ends, $b = \frac{1}{2}L$ to smooth at the center and $b = \frac{2}{3}L$ to smooth everywhere else. In the same way, substitute the corresponding values of i into the heteroscedastic regression asymptotic bias model (5.20). Minimize the contribution to bias with respect to t after ignoring the $O(\frac{1}{N})$ terms getting Table 5.3. Maple was especially useful in obtaining the interesting closed form solutions in this section. See that in this case t lies between $\frac{1}{3}\sqrt{5}$ and $\frac{2}{3}\sqrt{5}$ with the global bandwidth at

$$t_{\frac{N+3}{4}} = \frac{4}{183}\sqrt{2135} \approx 1$$

as

$$b \approx L \tag{5.31}$$

The major question that need to be answered here is: For what reason should one chose to consider the bias contribution term $g''(x_i)$ and not $g'(x_i)$ or $g(x_i)$? To answer this question, go back to equation (5.21) and see that the bias at $i = \frac{N+1}{2}$ does not depend on $g'(x_i)$. Further, from Tables 5.2 and 5.3 we can see the asymptotic contribution of the $g''(x_i)$ term. Also, recall that if the bias is considered at $g(x_i)$ it reduces to the homoscedastic regression case where the choice of kernel was found to have no significance effect at all, see Table 5.1.

Now consider the quartic kernel function

$$K(x) = \frac{15}{16}(1 - x^2)^2 \quad , \quad |x| \leq 1 \tag{5.32}$$

and obtain the leading term of the coefficient of $g''(x_i)$ in the asymptotic bias as

$$\begin{aligned}
& 3360t - 4032t^3 + 1440t^5 + tZ[5040 - 10080i + t^2(20160i - 10080) \\
+ \; & t^4(5040 - 10080i)] \\
+ \; & tZ^2[10080i^2 - 10080i + 1680 + t^2(40320i - 40320i^2 - 7280) \\
+ \; & t^4(30240i^2 - 30240i + 6048)] \\
+ \; & t^3Z^3[40320i^3 - 60480i^2 + 21840i - 840 \\
+ \; & t^2(2520 - 50400i^3 + 75600i^2 - 30240i)] \\
+ \; & t^3Z^4[392 - 20160i^4 + 40320i^3 - 21840i^2 + 1680i \\
+ \; & t^2(42 + 60480i^2 + 50400i^4 - 100800i^3 - 10080i)]
\end{aligned}$$

$$
\begin{aligned}
&+ \quad t^5 Z^5[63 - 126i - 60480i^3 - 30240i^5 + 75600i^4 + 15120i^2] \\
&+ \quad t^5 Z^6[10080i^6 - 30240i^5 + 30240i^4 - 10080i^3 + 126i^2 - 126i + 93] \qquad (5.33)
\end{aligned}
$$

where $Z = \frac{1}{N}$. Then substitute $i = N$ to obtain

$$
\begin{aligned}
& 3360t - 4032t^3 + 1440t^5 + Z(10080t^3 - 5040t - 5040t^5) \\
+ \quad & Z^2(1680t - 7280t^3 + 6048t^5) + Z^3(840t^3 - 2520t^5) \\
+ \quad & Z^4(42t^5 + 392t^3) - 63t^5 Z^5 + 93t^5 Z^6
\end{aligned}
$$

Ignore the $O(\frac{1}{N})$ terms and get $3360t - 4032t^3 + 1440t^5$. Minimizing with respect to t gives

$$
t_N = \frac{1}{15}\sqrt{3}\sqrt{63 + \sqrt{821}} \approx 1.1
$$

Similarly, substitute $i = \frac{N+1}{2}$ to obtain

$$
\begin{aligned}
& 1680t - 504t^3 + 45t^5 + Z^2(1400t^3 - 1680t - 189t^5) \\
+ \quad & Z^4(336t^5 - 896t^3) - 192t^5 Z^6
\end{aligned}
$$

Minimize with respect to t after ignoring the $O(\frac{1}{N})$ terms and get

$$
t_{\frac{N+1}{2}} = \frac{2}{15}\sqrt{3}\sqrt{63 + \sqrt{821}} \approx 2.2
$$

Finally, for $i = \frac{N+3}{4}$ minimize $1470t - \frac{3843}{4}t^3 + \frac{24615}{128}t^5$ with respect to t obtaining

$$t_{\frac{N+3}{4}} = \frac{4}{8205}\sqrt{7}\sqrt{1641}\sqrt{549 + \sqrt{224069}} \approx 1.7$$

Therefore, it is clear that a good choice of t for this kernel is between 1.1 and 2.2 which is around 1.7. Now compare the amount of smoothing of table 5.3 with the amount of smoothing offered by the quartic kernel function (5.32) and conclude that Table 5.3 gives oversmoothing.

Next, look at Table 5.3 and see that t_N is exactly the coefficient of kernel (4.9) while $t_{\frac{N+1}{2}}$ is twice this coefficient and $t_{\frac{N+3}{4}}$ is the mean of t_N and $t_{\frac{N+1}{2}}$ which is $\frac{3}{2}$ times this coefficient. However, for the quartic kernel (5.32) observe that t_N is exactly the reciprocal of its coefficient with $t_{\frac{N+1}{2}}$ and $t_{\frac{N+3}{4}}$ following immediately as above. Now, write kernel (4.9) and kernel (5.32) in a generalized form as

$$K(x) = \alpha(1 - x^2)^{\beta} \qquad (5.34)$$

then obtain the values of t using the same argument as above

$$t_N = \alpha^{-(-1)^{\beta}} \quad , \quad t_{\frac{N+3}{4}} = \frac{3}{2}\alpha^{-(-1)^{\beta}} \quad \text{and} \quad t_{\frac{N+1}{2}} = 2\alpha^{-(-1)^{\beta}} \qquad (5.35)$$

Kernels which are defined as in (5.34) are called optimal kernels. They minimize the integrated mean squared error, Gasser, Muller and Mammitzsch (1985) give detailed explanation. Another class of kernels for smoothing functions are those which minimize the asymptotic variance and are referred to as minimum variance kernels. An example is the kernel

$$K(x) = \frac{3}{8}(3 - 5x^2) \tag{5.36}$$

To be able to use formula (5.35), the coefficient of the minimum variance kernel is converted to the coefficient of an optimal kernel by multiplying the former by $\frac{k+1}{k}$, where k denotes the order of the kernel. The objective for this conversion is detailed in Gasser, Muller and Mammitzsch (1985). In this example, k is 2. For an illustration, obtain only the asymptotic bias contribution term $g''(x_i)$ at $i = \frac{N+1}{2}$ as $-36t + \frac{36}{N^2}t + 9t^3 - \frac{25}{N^2}t^3 + \frac{16}{N^4}t^3$. Ignore the $O(\frac{1}{N})$ terms and minimize with respect to t obtaining $t_{\frac{N+1}{2}} = \frac{2}{3}\sqrt{3} \approx 1.2$. Now use formula (5.35) to see that $t_{\frac{N+1}{2}} = 2(\frac{3}{2})(\frac{3}{8}) = 1.2$.

5.5 Conclusion Remarks

Given the asymptotic nature of all the arguments used in this chapter, it is established that the bandwidth parameter is clearly determined by the kernel function and the length of the smoothing region. The methodology developed here for the choice of this smoothing parameter is computationally simple. There are many reasons why this criterion of choice will be useful. It can in any case be used as a starting point for subsequent subjective adjustment for instance in boundary correction. To quote Silverman (1985, p. 5): "Scientists reporting or comparing their results will want to make reference to a standardized criterion". If the smoothing is to be used routinely on a large number of data sets or as part of a larger procedure then this criterion is essential. It does not require any consideration to prior assumptions. Today the theory of selecting the bandwidth parameter is expanding in several directions. However there is no single selection procedure widely acceptable as a standard to be included in the software packages.

Kernel	Asymptotic bias at $i = \frac{N+1}{2}$
$\frac{3}{8}(3-5x^2)$	$-\frac{\sigma^2}{32}(t-2)(5t^2+10t-16)$
$\frac{15}{16}(1-x^2)^2$	$\frac{\sigma^2}{256}(t-2)^3(3t^2+18t+32)$
$\frac{35}{32}(1-x^2)^3$	$-\frac{\sigma^2}{2048}(t-2)^4(5t^3+40t^2+116t+128)$

Table 5.1: Asymptotic bias at $i = \frac{N+1}{2}$ for various kernels

i	Asymptotic Bias	t_i	b_i
N	$12t - 4t^3 - 16 + \frac{6t^3}{N} - \frac{3t^3}{N^2}$	1	L
$\frac{N+1}{4}$	$12t - \frac{7t^3}{4} - 16 - \frac{6t^3}{4N} - \frac{3t^3}{4N^2}$	1.5	$\frac{2}{3}L$
$\frac{N+3}{4}$	$12t - \frac{7t^3}{4} - 16 + \frac{6t^3}{4N} - \frac{3t^3}{4N^2}$	1.5	$\frac{2}{3}L$
$\frac{3(N+1)}{4}$	$12t - \frac{7t^3}{4} - 16 - \frac{6t^3}{4N} - \frac{3t^3}{4N^2}$	1.5	$\frac{2}{3}L$
$\frac{N+1}{2}$	$12t - t^3 - 16$	2	$\frac{1}{2}L$

Table 5.2: Amount of smoothing in homoscedastic regression

i	Asymptotic Bias	t_i	b_i
N	$72t^3 - 120t$	$\frac{1}{3}\sqrt{5} \approx 0.75$	$\frac{4}{3}L$
$\frac{N+3}{4}$	$549t^3 - 1680t$	$\frac{4}{183}\sqrt{2135} \approx 1$	L
$\frac{N+1}{2}$	$9t^3 - 60t$	$\frac{2}{3}\sqrt{5} \approx 1.5$	$\frac{2}{3}L$

Table 5.3: Amount of smoothing in heteroscedastic regression

Chapter 6

APPLICATION TO SOME EXAMPLES

6.1 Introduction

Through artificially simulated data, the two approaches for estimating the variance function have been comprehensively studied in the preceding chapters. The conclusions arrived at in these chapters are investigated here through some real data situations. For this purpose, two examples are considered. A decisive procedure for checking the similarity or difference between the two estimated functions is described. Final remarks will be given regarding efficient and appropriate variance function estimation together with recommendations for some possible further extensions to this work.

6.2 Examples

Let us consider the first example.

6.2.1 Example 6.1 : Assay Data

The analysis of assay data has long been an important problem in clinical chemistry and the biological sciences. The most common method of analysis is to fit a nonlinear regression model to the data. This regression provides a way of exploring and presenting the relationship between the design variable and the response variable. It also gives predictions of observations yet to be made. Much recent work suggests that these data can be markedly heteroscedastic. Specifically, independent counts y_{ij} at concentrations x_i for $i = 1, \cdots, N$ and $j = 1, \cdots, m_i$ are made.

The data set used here is the assay experiment example (McCullagh and Nelder, 1989, p. 417). The data was originally obtained by Chen, Bliss and Robbins in 1942. The data show the relationship between heart mass and body mass for 149 male cats used in the experiment (Table 6.1). Note that for the 21st value of the body weight, only one heart weight observation is made. Therefore this value will be deleted in the analysis of this example.

Compute the sample means and sample standard deviations as in Table 6.2 and consider the parametric method of estimation first. A plot of the raw means appear reasonably linear. Take as the model for the means in these data

$$f(x_i, \beta) = \beta_0 + \beta_1[\text{body weight}]$$

In this case the x_i = body weight range from 1.7 to 3.8 with 0.1 spacing and $N = 22$. Since a plot of the raw standard deviations appear fairly quadratic, one might expect quadratic variation. Take as the model for the variances

$$\sigma_i^2 = \sigma^2(1 + \theta x_i^2)^2$$

Now apply the modified maximum likelihood method of estimation equations (3.22), (3.23) and (3.24). The estimated parameters are

$$\hat{\beta}_0 = -0.26 \quad \hat{\beta}_1 = 4.04 \quad \hat{\theta} = 11.5 \quad \hat{\sigma^2} = 0.0003$$

The complete set of the estimated means and standard deviations is listed in Table 6.2. The algorithm typically converged after some 20 iterations at an accuracy of 10^{-8}. Figure 6.1 displays the combined plots of the raw means together with the parametrically estimated and smoothed means. Similarly Figure 6.2 shows the combined plots of the raw variances together with the parametrically and nonparametrically estimated variance functions. We see that the parametric variance function fitted is monotonically increasing because it always has minimum at $x = 0$ and $x > 0$ for all x.

It is clear from this plot that differences at some points are visible. The question arises if this naked eye observation between the two estimated variance functions can suggest the use of a nonparametric instead of parametric method. One way to proceed is to find out how far the nonparametric is from the parametric model.

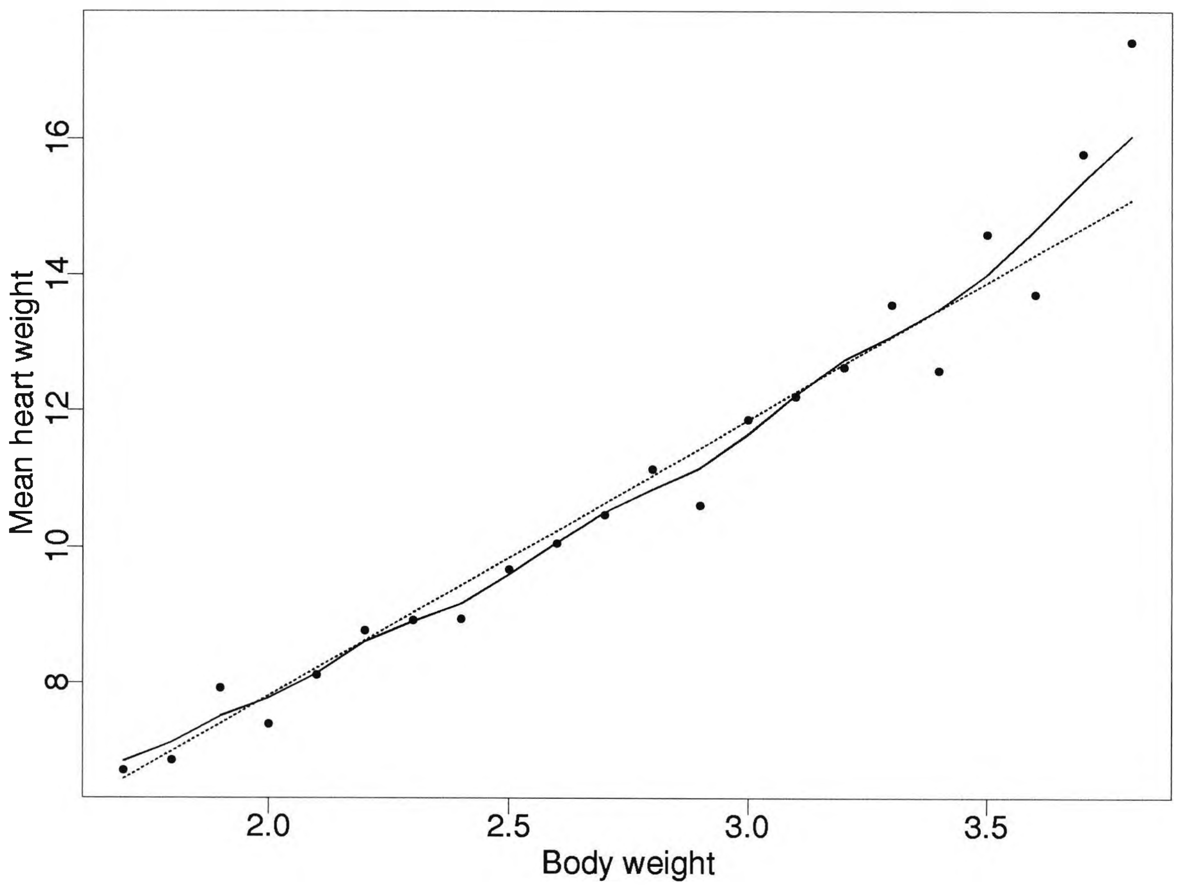

Figure 6.1: Mean plot: dots = raw mean, dotted line = parametric mean with parametric variance model $\sigma_i^2 = \sigma^2(1 + \theta x_i^2)^2$ and bold line = smooth mean. Body weight range from 1.7 to 3.8 with 0.1 spacing.

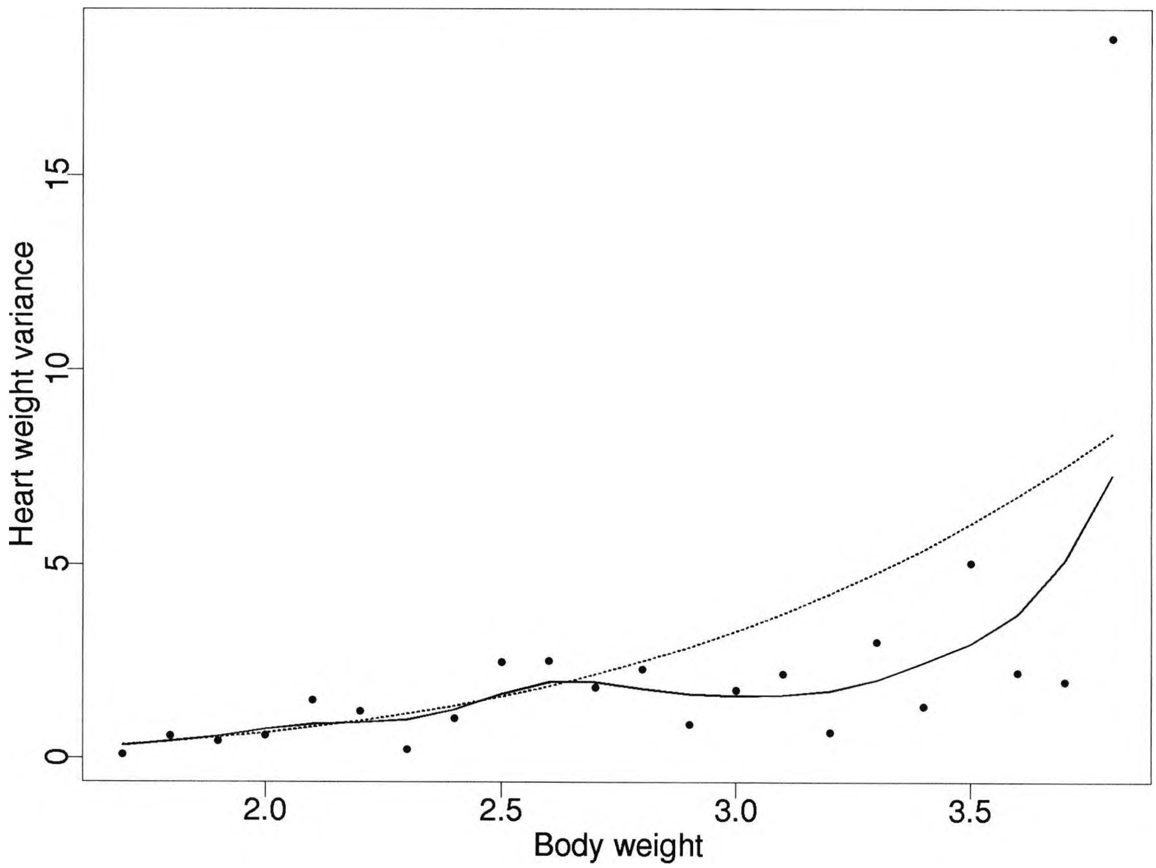

Figure 6.2: Variance plot: dots = raw variance, dotted curve = parametric variance with parametric variance model $\sigma_i^2 = \sigma^2(1 + \theta x_i^2)^2$ and bold curve = smooth variance. Body weight range from 1.7 to 3.8 with 0.1 spacing.

Measure the distance between them and use this as an informal test of

$$H_0 : \hat{g}_b(x_i) = g_{\hat{\theta}}(x_i)$$

against

$$H_1 : \hat{g}_b(x_i) \neq g_{\hat{\theta}}(x_i)$$

with subscript b and θ indicating nonparametric variance function and parametric variance function respectively. Compute the test statistic

$$T = \sum_{i=1}^{N}(\hat{g}_b(x_i) - g_{\hat{\theta}}(x_i))^2$$

and reject the null hypothesis H_0 in favour of the alternative hypothesis H_1 if $T \geq \chi^2_r$ with the critical rejection criterion value the $95th$ percentile of the χ^2_r distribution where r is the number of parameters in the model. Thus, since $T = 3.92$ and $\chi^2_1 = 3.84$ reject the null hypothesis at 0.05 level of significance.

The choice to use a squared deviation measure between the two fits is motivated by mathematical convenience. Certainly from a more data analytic point of view distances are more satisfactory which reflect similarities in the shape of the regression functions.

Next, investigate the effect of not specifying the parametric variance model correctly. Let a new parametric variance model perhaps with more parameters be

$$\sigma_i^2 = \sigma^2(\theta_1 + \theta_2 x_i + \theta_3 x_i^2)^2$$

and obtain the parameter estimates as

$$\hat{\beta}_0 = -0.25 \quad \hat{\beta}_1 = 4.04 \quad \hat{\theta}_1 = 0.01 \quad \hat{\theta}_2 = -0.05 \quad \hat{\theta}_3 = 11.5 \quad \hat{\sigma^2} = 0.0003$$

Full set of the estimated means and standard deviations is listed in Table 6.3 and for more visual clearity, Figure 6.3 display the plots of both smoothed and parametrically fitted variance functions. Once again the iteration procedure took slightly more cycles and hence longer time. Note that an arbitrary scale transformation on θ_1, θ_2 and θ_3 can be introduced so any single parameter can be fixed arbitrarily.

Now apply the test criterion and obtain $T = 3.68$ and $\chi^2_3 = 7.81$. Thus do not reject the null hypothesis at 0.05 level of significance. Notice the slight difference between the present and the former values of $\hat{\beta}_0$ leading to some slight difference in the estimated means. Observe column 4 entries 11,16,18,20,and 21 of both Tables 6.2 and 6.3 and notice that they are slightly different. This simply tells us that misspecification of the variance model can lead to a poor result.

6.2.2 Example 6.2

As another example, consider a moderately large data set. This particular data set was very kindly provided by the Engineering Design Centre at City University. Data is collected from an experiment in which an electronic circuit is repeatedly simulated at various input settings as part of a Robust Engineering Design (RED) experiment. The circuit used is an audio preamplifier with a frequency in the range of 20Hz to 20KHz. This type of

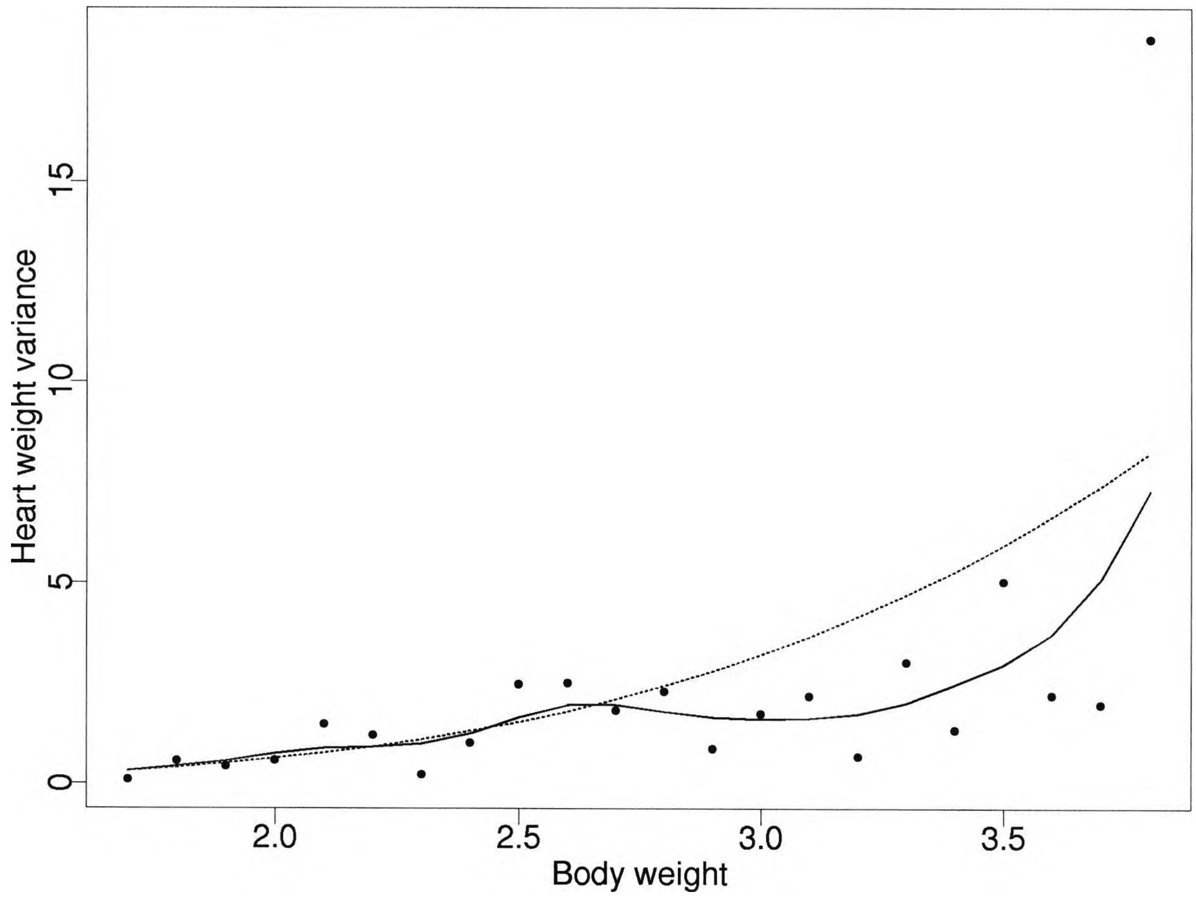

Figure 6.3: Variance plot: dots = raw variance, dotted curve = parametric variance with parametric variance model $\sigma_i^2 = \sigma^2(\theta_1 + \theta_2 x_i + \theta_3 x_i^2)^2$ and bold curve = smooth variance. Body weight range from 1.7 to 3.8 with 0.1 spacing.

Body wt.	Heart weight (gm.)												
1.7	6.5	7.0											
1.8	5.8	7.3	6.1	7.1	7.7	7.4							
1.9	8.1	9.1	8.0	7.2	7.3	8.0							
2.0	6.5	6.5	6.7	7.5	7.8	8.1	8.6	7.7					
2.1	10.1	7.0	7.2	8.1	8.3								
2.2	7.2	7.6	10.7	9.6	9.1	7.9	8.5	9.6	8.9				
2.3	9.6	9.6	8.5	8.8	8.2	9.2	8.7	8.9					
2.4	9.3	9.1	7.3	7.9	7.9	9.6	9.1	9.0	10.8	9.6			
2.5	8.8	12.7	8.6	12.7	9.3	7.9	11.0	8.8	9.3	8.2	8.7	10.4	9.6
2.6	10.5	8.3	9.4	7.7	11.5	9.4	13.6	10.1	10.9	9.6	9.9		
2.7	12.0	10.4	8.0	9.6	9.6	9.8	12.5	9.0	11.1	10.5	11.6	11.9	
2.8	10.0	12.0	13.5	13.3	9.1	10.2	11.4	10.1	10.9				
2.9	9.4	11.3	10.1	10.6	11.8								
3.0	13.3	10.0	13.8	10.6	12.4	12.7	10.4	11.6	12.2				
3.1	9.9	12.1	14.3	12.5	11.5	13.0							
3.2	11.6	13.6	12.3	13.0	13.5	11.9							
3.3	11.5	14.9	14.1	15.4	12.0								
3.4	14.4	12.2	12.8	11.2	12.4								
3.5	15.6	11.7	15.7	12.9	17.2								
3.6	14.8	13.3	15.0	11.8									
3.7	11.0												
3.8	14.8	16.8											
3.9	14.4	20.5											

Table 6.1: Relation in male cats of heart weight in gm. to body weight in Kg.

Bw	Raw Mean(Hw)	Smooth Mean	Par Mean	Raw std	Smooth std	Par std
x_i	$\bar{y}_i$	$\hat{\mu}_i$	$\hat{\mu}_i$	s_i	$\hat{\sigma}_i$	$\hat{\sigma}_i$
1.7	6.75	6.87	6.61	0.35	0.58	0.59
1.8	6.90	7.15	7.02	0.77	0.67	0.66
1.9	7.95	7.53	7.42	0.68	0.76	0.74
2.0	7.43	7.79	7.83	0.78	0.87	0.81
2.1	8.14	8.15	8.23	1.23	0.94	0.90
2.2	8.79	8.62	8.63	1.11	0.96	0.98
2.3	8.94	8.91	9.04	0.50	1.00	1.07
2.4	8.96	9.17	9.44	1.02	1.12	1.16
2.5	9.69	9.61	9.85	1.58	1.29	1.26
2.6	10.08	10.08	10.25	1.59	1.40	1.36
2.7	10.50	10.53	10.66	1.36	1.40	1.47
2.8	11.17	10.86	11.06	1.52	1.34	1.58
2.9	10.64	11.18	11.46	0.95	1.29	1.69
3.0	11.89	11.67	11.87	1.33	1.27	1.81
3.1	12.22	12.23	12.27	1.48	1.28	1.93
3.2	12.65	12.74	12.68	0.84	1.32	2.06
3.3	13.58	13.10	13.08	1.74	1.42	2.19
3.4	12.60	13.50	13.49	1.17	1.57	2.32
3.5	14.62	14.02	13.89	2.25	1.72	2.46
3.6	13.73	14.67	14.30	1.49	1.93	2.60
3.7	15.80	15.39	14.70	1.41	2.26	2.74
3.8	17.45	16.04	15.10	4.31	2.70	2.89

Table 6.2: Smooth and parametric numerical results with parametric model $\sigma_i^2 = \sigma^2(1 + \theta x_i^2)^2$.

Bw x_i	Raw Mean(Hw) $\bar{y}_i$	Smooth Mean $\hat{\mu}_i$	Par Mean $\hat{\mu}_i$	Raw std s_i	Smooth std $\hat{\sigma}_i$	Par std $\hat{\sigma}_i$
1.7	6.75	6.87	6.61	0.35	0.58	0.57
1.8	6.90	7.15	7.02	0.77	0.67	0.64
1.9	7.95	7.53	7.42	0.68	0.76	0.72
2.0	7.43	7.79	7.83	0.78	0.87	0.80
2.1	8.14	8.15	8.23	1.23	0.94	0.88
2.2	8.79	8.62	8.63	1.11	0.96	0.96
2.3	8.94	8.91	9.04	0.50	1.00	1.05
2.4	8.96	9.17	9.44	1.02	1.12	1.15
2.5	9.69	9.61	9.85	1.58	1.29	1.24
2.6	10.08	10.08	10.25	1.59	1.40	1.34
2.7	10.50	10.53	10.65	1.36	1.40	1.45
2.8	11.17	10.86	11.06	1.52	1.34	1.56
2.9	10.64	11.18	11.46	0.95	1.29	1.67
3.0	11.89	11.67	11.87	1.33	1.27	1.79
3.1	12.22	12.23	12.27	1.48	1.28	1.91
3.2	12.65	12.74	12.67	0.84	1.32	2.04
3.3	13.58	13.10	13.08	1.74	1.42	2.17
3.4	12.60	13.50	13.48	1.17	1.57	2.30
3.5	14.62	14.02	13.89	2.25	1.72	2.44
3.6	13.73	14.67	14.29	1.49	1.93	2.58
3.7	15.80	15.39	14.69	1.41	2.26	2.72
3.8	17.45	16.04	15.10	4.31	2.70	2.87

Table 6.3: Smooth and parametric numerical results with parametric model $\sigma_i^2 = \sigma^2(\theta_1 + \theta_2 x_i + \theta_3 x_i^2)^2$.

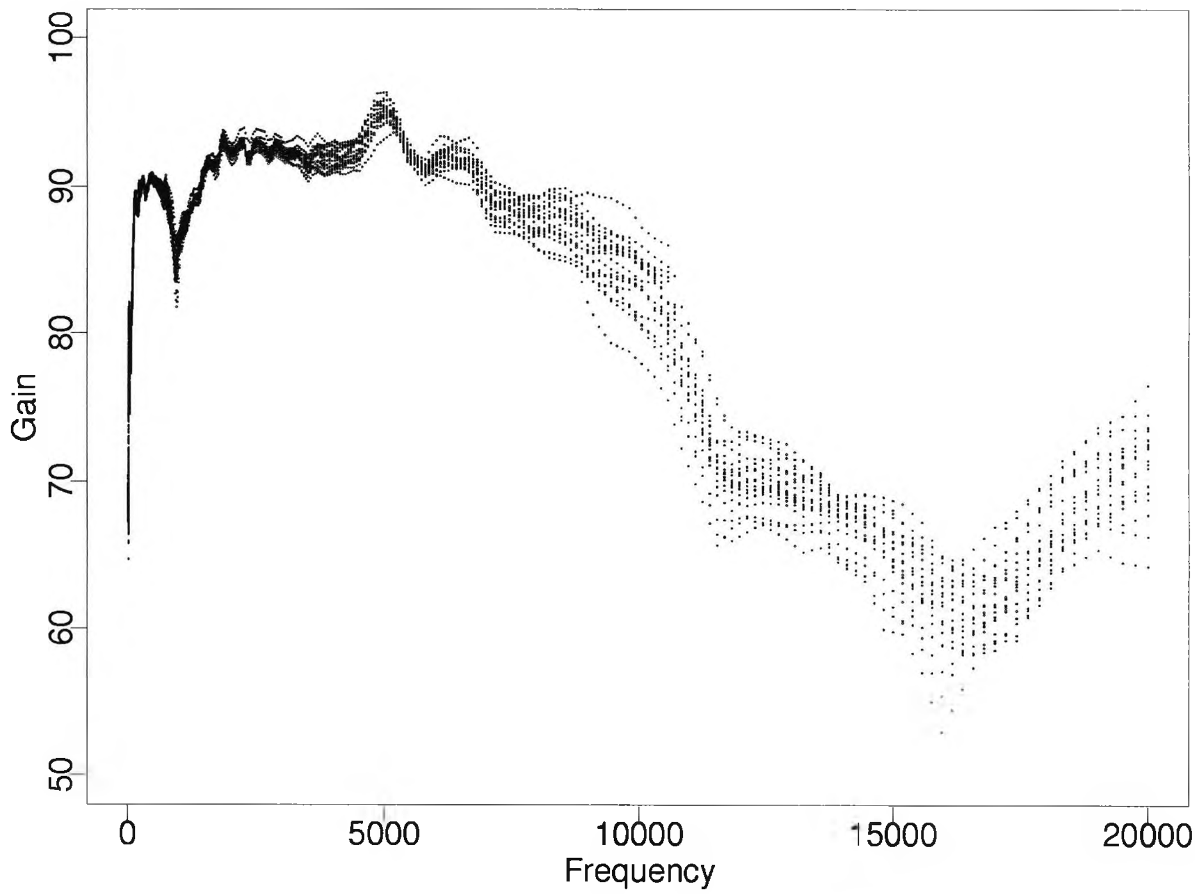

Figure 6.4: Scatter plot: Gain versus Frequency

experiment originates from physics where the predictor is a frequency and the response is the gain of an electronic circuit.

The data consist of 552 design points and for each of these design points, twenty observations are made. Since this data is very large only pictures will be used to display the results. Further only smoothing will be considered with the aim of showing what happens with the use of the various bandwidths discussed under Chapter 5. Figure 6.4 displays the scatter plot of these data. This gives a clear indication of the general pattern of the data. It makes the inhomogeneity in the variance very clear, specifically see how

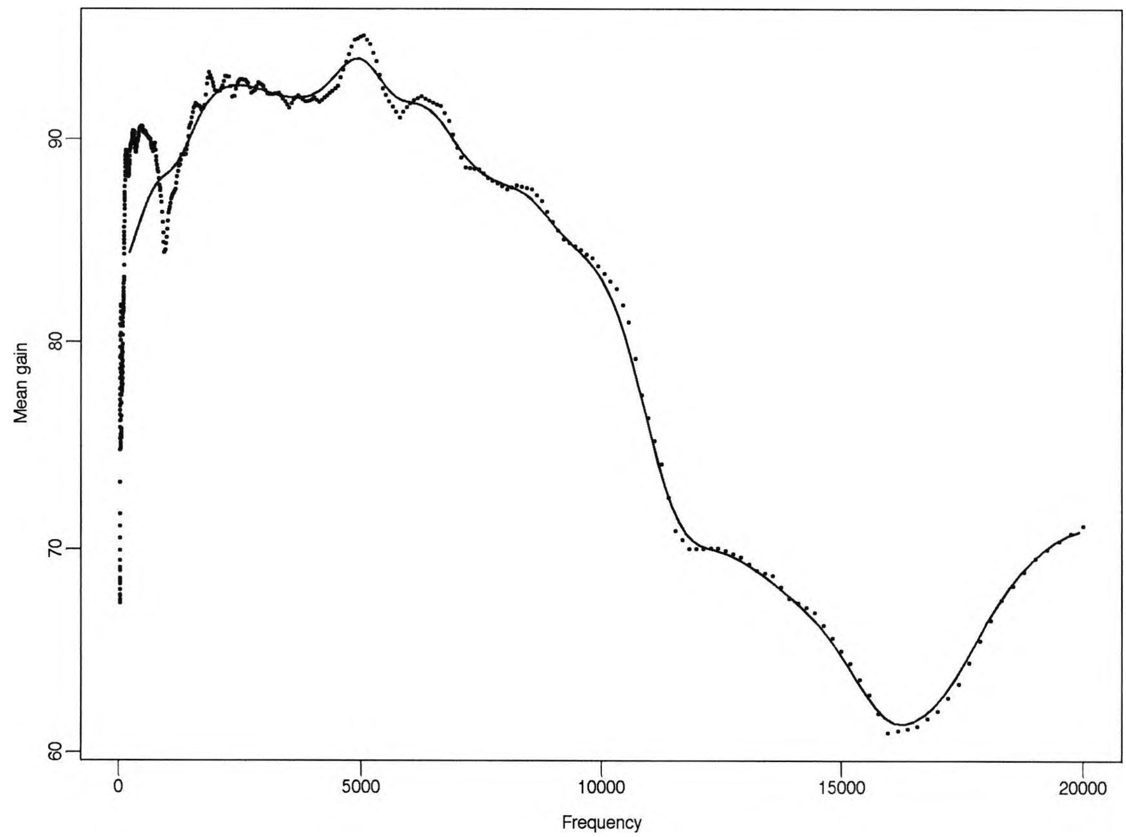

Figure 6.5: Smooth mean

the scatter spreads towards the right end. Obviously this indicates the data becoming more variable. Smooth the sample means getting the plot shown on Figure 6.5. Obtain the sample standard deviations and apply global smoothing giving the plot displayed in Figure 6.6 for the three bandwidths discussed under Chapter 5. Next apply local smoothing obtaining Figure 6.7.

6.3 Conclusion

Based on the experience of this work some advantages and disadvantages for both paramet-

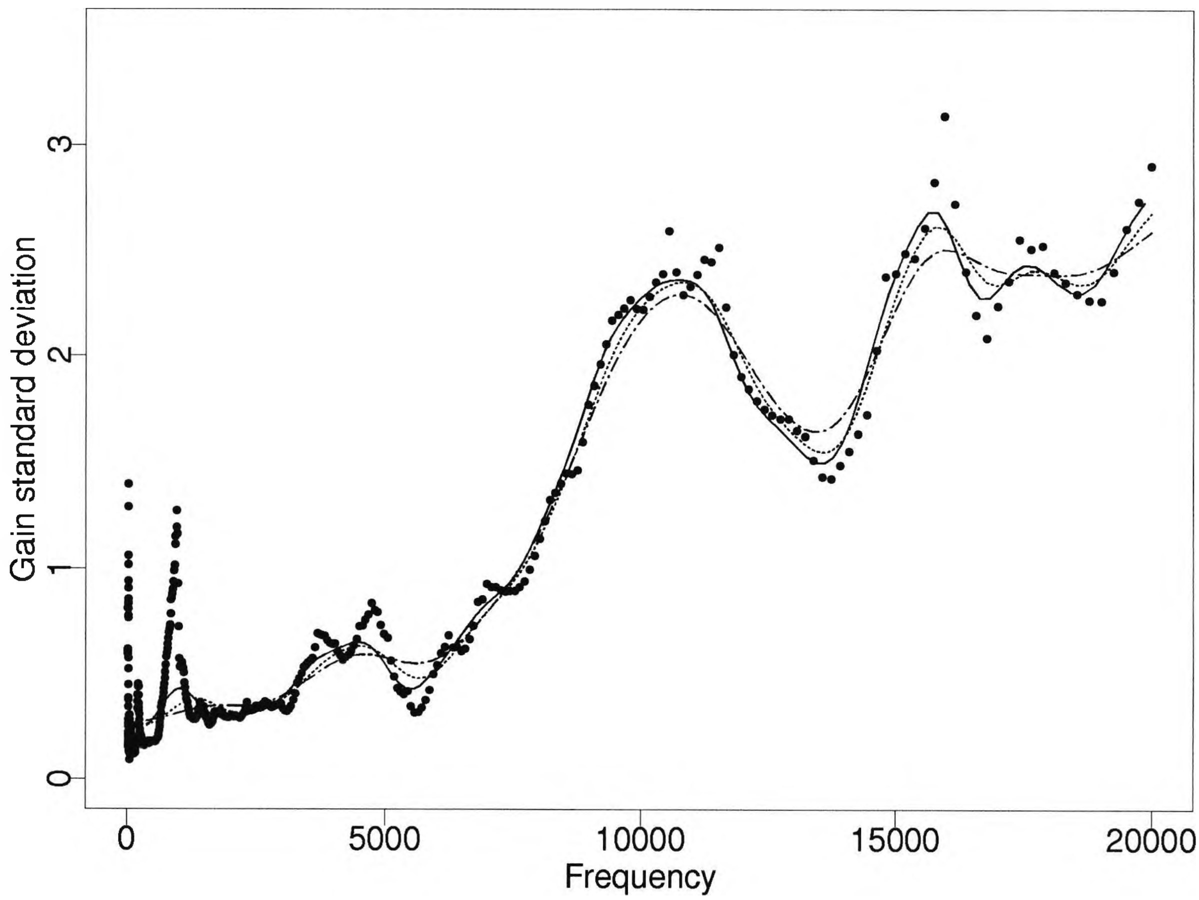

Figure 6.6: Global smooth: dots = raw standard deviation, bold curve = smooth standard deviation with big t, dotted curve = smooth standard deviation with medium t and broken curve = smooth standard deviation with small t.

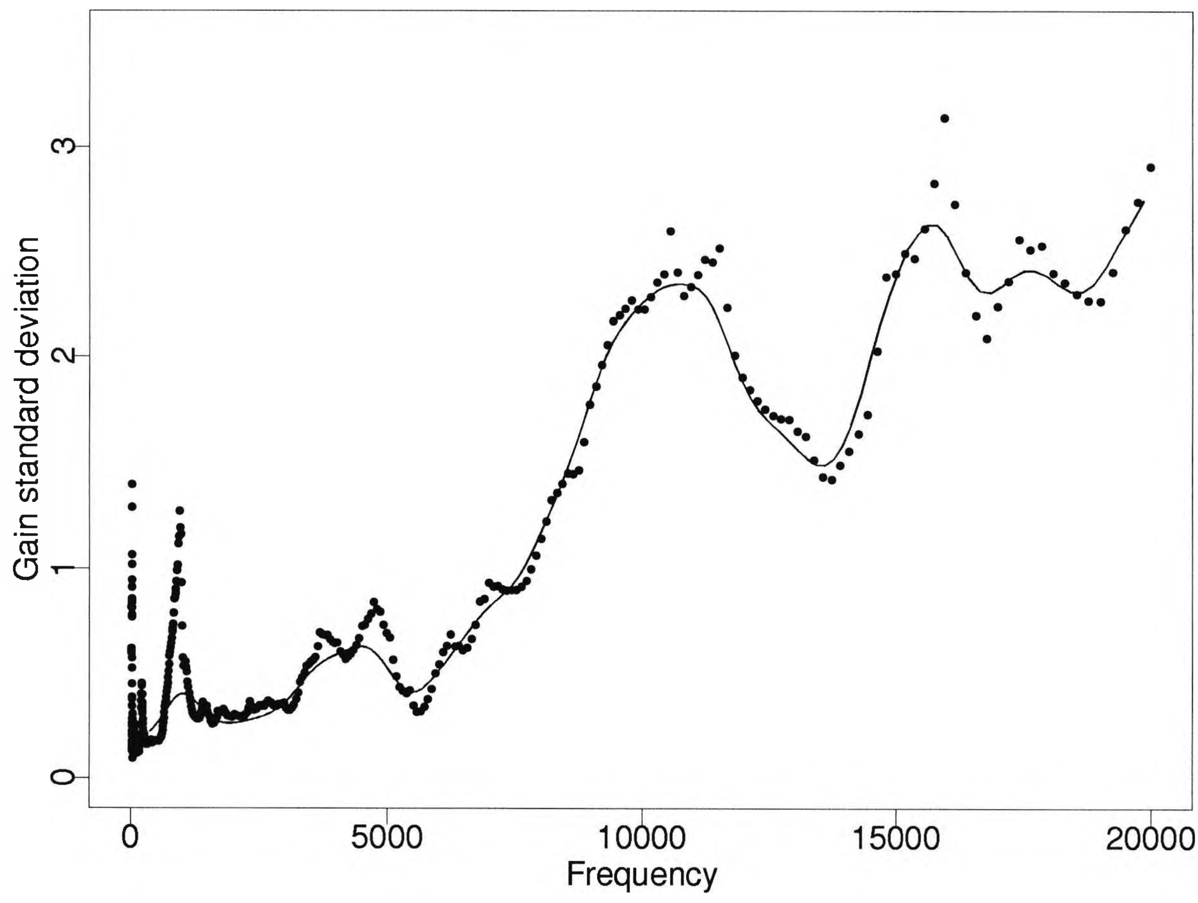

Figure 6.7: Local smooth: dots = raw standard deviation and bold curve = smooth standard deviation with the combination of the three t's.

ric and nonparametric approaches for estimating the variance function have been established. It has been found that parametric methods have small biases but large variances. Smoothing is directly opposite, the biases are substantially large relative to the variances. However a bias reduction can be carried out. Notice again here the use of the word reduction and not eradication. Complete eradication of the bias remain a major unresolved problem. After reducing the bias, then both approaches are worthwhile. However, although parametric procedures are easier to understand, they can be poor if there is misspecification of the model to be fitted.

6.4 Some Open Problems

Obvious cases that could be of interest for further research emerged during the process of this exercise. A brief summary of some of these is listed here.

1. For purposes of simplicity, this work have required the case of equal spacing (fixed design). A valuable next step would be to determine just what would happen with the departure from this requirement to the case of nonequal spacing (random design).

2. The smoothing procedure developed here can be extended to multivariate data.

3. There is need to investigate the possibility of incorporating boundary kernels into the theory of Chapter 5 and see if this can offer any further bias reduction.

4. Another idea that may require some exploration is the possibility of relating a parametric model to the bias model developed in Chapter 5 to see if this relationship can

lead to a further reduction of the bias.

Appendix

The most lengthy computations in Chapter 5 were carried on MAPLE. The more elementary calculations were checked by hand. Because of the discrete nature of the summations for the equally spaced data, it can be assumed that maple gives the correct solutions for all formulae. An example is the formula $\sum_{j=1}^{N} j^r$, where $r \geq 1$ is an integer. The command

$$\mathrm{sum}(j\hat{\ }r, j = 1..N);$$

computed the definite sum over the given range $1 \cdots N$. The command

$$\mathrm{sum}(j, j = 1 \cdots N);$$

gives the solution $\frac{1}{2}(N+1)^2 - \frac{1}{2}N - \frac{1}{2}$. Combination of commands would then give a simplified solution. For example, to make this summation solution look neat we would use the command simplify. Thus

$$\mathrm{simplify}(\mathrm{sum}(j, j = 1 \cdots N));$$

would give the solution $\frac{1}{2}N^2 + \frac{1}{2}N$. This solution can then be made neater by a third command, say, factorize. Therefore

$$\mathrm{factor}(\mathrm{simplify}(\mathrm{sum}(j, j = 1 \cdots N)));$$

gives the solution $\frac{1}{2}N(N+1)$. For the minimization procedure of the kind following (5.33) we used the commands diff and solve giving us the real and complex roots. For example,

$$\text{diff}(12*t-4*t\hat{}3-16,t);$$

would give the solution $12-12t^2$. Using the commands solve and diff as

$$\text{solve}(\text{diff}(12*t-4*t\hat{}3-16,t));$$

would then give the real and the complex roots as 1, $-\frac{1}{2}+\frac{1}{2}I3^{\frac{1}{2}}$ and $-\frac{1}{2}-\frac{1}{2}I3^{\frac{1}{2}}$. The solve command was particularly useful in obtaining the closed form solutions in Tables 5.2 and 5.3. The following worked example is very typical of the way maple is used in Chapter 5.

Example

In this example we show in full details how MAPLE was used to move from the last part of equation (5.16) to equation (5.19).

1. MAPLE: Steps $m1$ to $m3$ simplifies the term under the summation sign. Steps $m4$ to $m8$ brings the summation sign inside. Step $m9$ finalizes the expansion of (5.16). Note that $A = g''(x_i)$.

2. HAND: The important substitutions of L for Nh and $t = \frac{L}{b}$ is done by hand in steps hand_1 and hand_2 for clarity.

3. MAPLE: After substitution is done, the term inside the bracket of step hand_2 is computed and simplified in steps $m10$ to $m23$. In step $m23$, the coefficients of $\frac{1}{N^4}$ and $\frac{1}{N^2}$ respectively can be simplified further as in steps $m24$ and $m25$. Then the required compact solution is obtained. The factor command was of considerable benefit here.

MAPLE

$> m1 := \text{expand}((i - j)\hat{}2);$

$$m1 := i^2 - 2ij + j^2$$

$> m2 := (b\hat{}2 * h - h\hat{}3 * i\hat{}2 + 2 * h\hat{}3 * i * j - h\hat{}3 * j\hat{}2 - h\hat{}3/12);$

$$m2 := b^2h - h^3i^2 + 2h^3ij - h^3j^2 - 1/12h^3$$

$> m3 := \text{expand}(m1 * m2);$

$$m3 := i^2b^2h - i^4h^3 + 4i^3h^3j - 6i^2h^3j^2 - 1/12h^3i^2 - 2ijb^2h$$
$$+4ij^3h^3 + 1/6h^3ij + j^2b^2h - j^4h^3 - 1/12h^3j^2$$

$> m4 := (N*i\hat{\ }2*b\hat{\ }2*h - N*i\hat{\ }4*h\hat{\ }3 + 4*i\hat{\ }3*h\hat{\ }3*\mathrm{sum}(j, j = 1..N));$

$$m4 := Ni^2b^2h - Ni^4h^3 + 4i^3h^3(1/2(N+1)^2 - 1/2N - 1/2)$$

$> m5 := (-6*i\hat{\ }2*h\hat{\ }3*\mathrm{sum}(j\hat{\ }2, j = 1..N) - N*1/12*h\hat{\ }3*i\hat{\ }2);$

$$m5 := -6i^2h^3(1/3(N+1)^3 - 1/2(N+1)^2 + 1/6N + 1/6) - 1/12Nh^3i^2$$

$> m6 := (-2*i*b\hat{\ }2*h*\mathrm{sum}(j, j = 1..N) + 4*i*h\hat{\ }3*\mathrm{sum}(j\hat{\ }3, j = 1..N));$

$$m6 := -2ib^2h(1/2(N+1)^2 - 1/2N - 1/2)$$
$$+4ih^3(1/4(N+1)^4 - 1/2(N+1)^3 + 1/4(N+1)^2)$$

$> m7 := (1/6*h\hat{\ }3*i*\mathrm{sum}(j, j = 1..N) + b\hat{\ }2*h*\mathrm{sum}(j\hat{\ }2, j = 1..N));$

$$m7 := 1/6h^3i(1/2(N+1)^2 - 1/2N - 1/2)$$
$$+b^2h(1/3(N+1)^3 - 1/2(N+1)^2 + 1/6N + 1/6)$$

$> m8 := (-h\hat{\ }3*\mathrm{sum}(j\hat{\ }4, j = 1..N) - 1/12*h\hat{\ }3*\mathrm{sum}(j\hat{\ }2, j = 1..N));$

$$m8 := -h^3(1/5(N+1)^5 - 1/2(N+1)^4 + 1/3(N+1)^3 - 1/30N - 1/30)$$
$$-1/12h^3(1/3(N+1)^3 - 1/2(N+1)^2 + 1/6N + 1/6)$$

$> m9 := (3 * A * h\hat{}2/(8 * b\hat{}3) * \mathrm{factor}(m4 + m5 + m6 + m7 + m8));$

$$m9 := -1/960 A h^3 N(-720h^2i^3N + 360h^2i^4 - 360i^2b^2 + 390h^2i^2 + 72h^2N^4 + 180h^2N^3 + 130h^2N^2 - 7h^2 + 15h^2N - 720h^2i^3 + 720h^2i^2N^2 + 1080h^2Ni^2 + 360ib^2N + 360ib^2 - 360h^2iN^3 - 720h^2iN^2 - 390h^2iN - 30h^2i - 120b^2N^2 - 180b^2N - 60b^2)/b^3$$

HAND

$$\mathrm{hand}_1 = \frac{-AL^2}{960}[-\frac{360i^2L}{bN^2} + \frac{360i^4L^3}{b^3N^4} + \frac{390i^2L^3}{b^3N^4} - \frac{720i^3L^3}{b^3N^3} - \frac{720i^3L^3}{b^3N^4} + \frac{720i^2L^3}{b^3N^2} + \frac{1080i^2L^3}{b^3N^3} + \frac{360iL}{bN} + \frac{360iL}{bN^2} - \frac{120L}{b} - \frac{180L}{bN} - \frac{60L}{bN^2} + \frac{72L^3}{b^3} + \frac{180L^3}{b^3N} + \frac{130L^3}{b^3N^2} - \frac{7L^3}{b^3N^4} - \frac{360iL^3}{b^3N} - \frac{720iL^3}{b^3N^2} - \frac{390iL^3}{b^3N^3} - \frac{30iL^3}{b^3N^4} + \frac{15L^3}{b^3N^3}]$$

$$\mathrm{hand}_2 = \frac{-AL^2}{960}[-\frac{360i^2t}{N^2} + \frac{360i^4t^3}{N^4} + \frac{390i^2t^3}{N^4} - \frac{720i^3t^3}{N^3} - \frac{720i^3t^3}{N^4} + \frac{720i^2t^3}{N^2} + \frac{1080i^2t^3}{N^3} + \frac{360it}{N} + \frac{360it}{N^2} - 120t - \frac{180t}{N} - \frac{60t}{N^2} + 72t^3 + \frac{180t^3}{N} + \frac{130t^3}{N^2} - \frac{7t^3}{N^4} - \frac{360it^3}{N} - \frac{720it^3}{N^2} - \frac{390it^3}{N^3} - \frac{30it^3}{N^4} + \frac{15t^3}{N^3}$$

MAPLE

$> m10 := (-360 * i\hat{}2 * t/N\hat{}2 + 360 * i\hat{}4 * t\hat{}3/N\hat{}4 + 390 * i\hat{}2 * t\hat{}3/N\hat{}4);$

$$m10 := -360\frac{i^2t}{N^2} + 360\frac{i^4t^3}{N^4} + 390\frac{i^2t^3}{N^4}$$

> $m11 := (-720*i\hat{\ }3*t\hat{\ }3/N\hat{\ }3 - 720*i\hat{\ }3*t\hat{\ }3/N\hat{\ }4 + 720*i\hat{\ }2*t\hat{\ }3/N\hat{\ }2);$

$$m11 := -720\frac{i^3t^3}{N^3} - 720\frac{i^3t^3}{N^4} + 720\frac{i^2t^3}{N^2}$$

> $m12 := (1080*i\hat{\ }2*t\hat{\ }3/N\hat{\ }3 + 360*i*t/N + 360*i*t/N\hat{\ }2 - 120*t);$

$$m12 := 1080\frac{i^2t^3}{N^3} + 360\frac{it}{N} + 360\frac{it}{N^2} - 120t$$

> $m13 := (-180*t/N - 60*t/N\hat{\ }2 + 72*t\hat{\ }3 + 180*t\hat{\ }3/N);$

$$m13 := -180t/N - 60\frac{t}{N^2} + 72t^3 + 180\frac{t^3}{N}$$

> $m14 := (130*t\hat{\ }3/N\hat{\ }2 - 7*t\hat{\ }3/N\hat{\ }4 - 360*i*t\hat{\ }3/N - 720*i*t\hat{\ }3/N\hat{\ }2);$

$$m14 := 130\frac{t^3}{N^2} - 7\frac{t^3}{N^4} - 360\frac{it^3}{N} - 720\frac{it^3}{N^2}$$

> $m15 := (-390*i*t\hat{\ }3/N\hat{\ }3 - 30*i*t\hat{\ }3/N\hat{\ }4 + 15*t\hat{\ }3/N\hat{\ }3);$

$$m15 := -390\frac{it^3}{N^3} - 30\frac{it^3}{N^4} + 15\frac{t^3}{N^3}$$

> $m16 := (m10 + m11 + m12 + m13 + m14 + m15);$

$$m16 := -360\frac{i^2t}{N^2} + 360\frac{i^4t^3}{N^4} + 390\frac{i^2t^3}{N^4} - 720\frac{i^3t^3}{N^3} - 720\frac{i^3t^3}{N^4} + 720\frac{i^2t^3}{N^2}$$

$$+1080\frac{i^2t^3}{N^3} + 360\frac{it}{N} + 360\frac{it}{N^2} - 120t - 180t/N - 60\frac{t}{N^2} + 72t^3$$
$$+180\frac{t^3}{N} + 130\frac{t^3}{N^2} - 7\frac{t^3}{N^4} - 360\frac{it^3}{N} - 720\frac{it^3}{N^2} - 390\frac{it^3}{N^3} - 30\frac{it^3}{N^4}$$
$$+15\frac{t^3}{N^3}$$

> $m17 := \text{series}(m16, N);$

$$m17 := (-720i^3t^3 + 360i^4t^3 + 390i^2t^3 - 7t^3 - 30it^3)N^{-4}$$
$$+(-390it^3 - 720i^3t^3 + 15t^3 + 1080i^2t^3)N^{-3}$$
$$+(-360i^2t + 360it - 720it^3 + 130t^3 + 720i^2t^3 - 60t)N^{-2}$$
$$+(-360it^3 + 180t^3 + 360it - 180t)N^{-1} + (-120t + 72t^3)$$

> $m18 := ((-720 * i\hat{}3 * t\hat{}3 + 360 * i\hat{}4 * t\hat{}3 + 390 * i\hat{}2 * t\hat{}3 - 7 * t\hat{}3 - 30 * i * t\hat{}3)/N\hat{}4);$

$$m18 := \frac{-720i^3t^3 + 360i^4t^3 + 390i^2t^3 - 7t^3 - 30it^3}{N^4}$$

> $m19 := ((-390 * i * t\hat{}3 - 720 * i\hat{}3 * t\hat{}3 + 15 * t\hat{}3 + 1080 * i\hat{}2 * t\hat{}3)/N\hat{}3);$

$$m19 := \frac{-390it^3 - 720i^3t^3 + 15t^3 + 1080i^2t^3}{N^3}$$

> $m20 := ((-360 * i\hat{}2 * t + 360 * i * t - 720 * i * t\hat{}3 + 130 * t\hat{}3 + 720 * i\hat{}2 * t\hat{}3 - 60 * t)/N\hat{}2);$

$$m20 := \frac{-360i^2t + 360it - 720it^3 + 130t^3 + 720i^2t^3 - 60t}{N^2}$$

$> m21 := ((-360 * i * t\hat{}3 + 180 * t\hat{}3 + 360 * i * t - 180 * t)/N);$

$$m21 := \frac{-360it^3 + 180t^3 + 360it - 180t}{N}$$

$> m22 := (-120 * t + 72 * t\hat{}3);$

$$m22 := -120t + 72t^3$$

$> m23 := (\text{factor}(m18) + \text{factor}(m19) + \text{factor}(m20) + \text{factor}(m21) + \text{factor}(m22));$

$$m23 := \frac{t^3(-720i^3 + 360i^4 + 390i^2 - 7 - 30i)}{N^4} - 15\frac{t^3(2i-1)(24i^2 - 24i + 1)}{N^3} + 10\frac{t(-36i^2 + 36i - 72it^2 + 13t^2 + 72i^2t^2 - 6)}{N^2} - 180\frac{t(t-1)(t+1)(2i-1)}{N} + 24t(-5 + 3t^2)$$

$> m24 := (t\hat{}3 * \text{factor}(-720 * i\hat{}3 + 360 * i\hat{}4 + 390 * i\hat{}2 - 30 * i) - 7));$

$$m24 := t^3(30i(i-1)(12i^2 - 12i + 1) - 7)$$

$> m25 := (10 * t * (\text{factor}(-36 * i\hat{}2 + 36 * i - 72 * i * t\hat{}2 + 72 * i\hat{}2 * t\hat{}2) + 13 * t\hat{}2 - 6));$

$$m25 := 10t(36i(-1 + 2t^2)(i-1) + 13t^2 - 6)$$

$$
\begin{aligned}
\text{solution} &= \frac{-AL^2}{960}\{24t(3t^2-5) - \frac{180t(t-1)(t+1)(2i-1)}{N} \\
&+ \frac{10t[36i(2t^2-1)(i-1)+13t^2-6]}{N^2} - \frac{15t^3(2i-1)(24i^2-24i+1)}{N^3} \\
&+ \frac{t^3[30i(i-1)(12i^2-12i+1)-7]}{N^4}\}
\end{aligned}
$$

Bibliography

[1] Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics* 6, 365-370.

[2] Breiman, L. and Meisel, W. S. (1976). General estimates of the intrinsic variability of data in nonlinear regression models. *Journal of the American Statistical Association* 71, 301-307.

[3] Buckley, M. J., Eagleson, G. K. and Silverman, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* 75, 189-200.

[4] Cao, R., Cuevas, A. and Manteiga, W. G. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis* 17, 153-176.

[5] Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Annals of Statistics* 10, 1224-1233.

[6] Carroll, R. J. and Ruppert, D. (1982a). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *Journal of the American*

Statistical Association 77, 878-882.

[7] Carroll, R. J. and Ruppert, D. (1982b). Robust estimation in heteroscedastic linear models. *Annals of Statistics* 10, 429-441.

[8] Carroll, R. J. and Ruppert, D. (1985). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association* 79, 321-328.

[9] Carroll, R. J. and Ruppert, D. (1987). Diagnostics and robust estimation when transforming the regression model and the response. *Technometrics* 29, 287-299.

[10] Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression.* Chapman and Hall, London.

[11] Carter C. K. and Eagleson, G. K. (1992). A comparison of variance estimators in nonparametric regression. *Journal of the Royal Statistical Society* B 54, 773-780.

[12] Cline, D. B. H. and Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivative. *Statistics* 22, 69-84.

[13] Cohen, M., Dalal, S. R. and Tukey, J. W. (1993). Robust, smoothly heterogeneous variance regression. *Applied Statistics* 42, 339-353.

[14] Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* 70, 1-10.

[15] Cook, R. D. and Weisberg, S. (1992). *Residuals and Influence in Regression.* Chapman and Hall, London.

[16] Cover, T. M. (1968). Estimation by the nearest neighbor rule. *IEEE Transactions on Information theory* 14, 50-55.

[17] Davidian, M. (1986). Variance function estimation in heteroscedastic regression models. Ph.D. Dissertation, University of North Carolina at Chapel Hill.

[18] Davidian, M. (1990). Estimation of variance functions in assays with possibly unequal replication and nonnormal data. *Biometrika* 77, 43-54.

[19] Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* 82, 1079-1091.

[20] Davidian, M. and Carroll, R. J. (1988). An analysis of extended quasi-likelihood. *Journal of the Royal Statistical Society.* B 50.

[21] Dent, W. T. and Hildreth, C. (1977). Maximum likelihood estimation in random coefficient models. *journal of the American Statistical Association* 72, 69-72.

[22] Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* 65, 457-481.

[23] Egy, D. and Lahiri, K. (1979). On maximum likelihood estimation of functional forms and heteroscedasticity. *Economics letters* 2, 155-160.

[24] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998-1004.

[25] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics* 21, 196-216.

[26] Finney, D. J. (1964). *Statistical Methods in Biological Assay.* London: Griffin.

[27] Finney, D. J. (1976). Radioligand assay. *Biometrics* 32, 721-740.

[28] Firth, D. (1987). On the efficiency of quasi-likelihood estimation. *Biometrika* 74, 233-246.

[29] Friedman, J. H., Baskett, F. and Shustek, L. J. (1975). An algorithm for finding nearest neighbours. *IEEE Transactions on computers* C-24, 1000-1006.

[30] Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Annals of Statistics* 6, 1149-1158.

[31] Gasser, T. and Engel, J. (1990). The choice of weights in kernel regression estimation. *Biometrika* 77, 377-381.

[32] Gasser, T. and Muller, H. G. (1979). Kernel estimation of regression functions", *in smoothing techniques for curve estimation* (Lecture notes in Mathematics 757), eds. T. Gasser and M. Rosenblatt, Berlin Springer-Verlag.

[33] Gasser, T. and Muller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandnavian Journal of Statistics* 11, 197-211.

[34] Gasser, T., Muller, H. G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society* B 47, 238-252.

[35] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73, 625-633.

[36] Giltinan, D. M., Carroll, R. J. and Ruppert, D. (1986). Some new methods for weighted regression when there are possible outliers. *Technometrics* 28, 219-230.

[37] Glejser, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical association* 64, 316-323.

[38] Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics* 9, 861-869.

[39] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London.

[40] Haerdle, W. (1990). *Applied Nonparametric Regression*, Cambridge, U.K., Cambridge University Press.

[41] Haerdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926-1947.

[42] Haerdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics* 13, 1465-1481.

[43] Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society* B 51, 3-14.

[44] Hall, P. and Titterington, D. M. (1987). Common structure of techniques for choosing smoothing parameters in regression problems. *Journal of the Royal Statistical Society* B 49, 184-198.

[45] Hall, P., Kay, J. W. and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* 77, 521-528.

[46] Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* 77, 415-419.

[47] Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association* 86, 665-672.

[48] Harrison, M. J. and McCabe, B. P. M. (1979). A test for heteroscedasticity based on least squares residuals. *Journal of the American Statistical Association* 74, 494-500.

[49] Hart, J. D. and Wehrly, T. E. (1992). Kernel regression when the boundary region is large, with an application to testing the adequacy of polynomial models. *Journal of the American Statistical Association* 87, 1018-1024.

[50] Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 79, 302-308.

[51] Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrics* 44, 461-465.

[52] Hastie, T. J. and Tibshirani, R. J. (1991). *Generalized additive models.* Chapman and Hall, London.

[53] Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association* 63, 584-595.

[54] Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with non constant, unknown error variances: sampling experiments with least squares and maximum likelihood estimators. *Biometrics* 24, 607-626.

[55] Jacquez, J. A. and Norusis, M. (1973). Sampling experiments on the estimation of parameters in heteroscedastic linear regression. *Biometrics* 29, 771-780.

[56] Jennen-Steinmetz, C. and Gasser, T. (1988). A unifying approach to nonparametric regression estimation. *Journal of the American Statistical Association* 83, 1084-1089.

[57] Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633-643.

[58] Jobson, J. D. and Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Association* 75, 176-181.

[59] Jolly, L. B. W. (1961). *Summation of Series.* Dover Publications, inc. New York.

[60] Mack, Y. P. (1981). Local properties of k-nn regression estimates. *SIAM Journal of Algebraic and Discrete Methods* 2, 311-323.

[61] Mack, Y. P. and Muller, H. G. (1988). Convolution type estimators for nonparametric regression. *Statistics and Probability Letters* 7, 229-239.

[62] McCullagh, P. (1983). Quasi likelihood functions. *Annals of Statistics* 11, 59-67.

[63] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* Chapman and Hall, London.

[64] Muller, H. G. (1984). Smooth optimum kernel estimators of regression curves, densities and modes. *Annals of Statistics* 12, 766-774.

[65] Muller, H. G. (1991). Smooth optimum kernel estimators near end points. *Biometrika* 78, 521-530.

[66] Muller, H. G. (1993). On variance function estimation with quadratic forms. *Journal of Statistical Planning and Inference* 35, 213-231.

[67] Muller, H. G. and Stadtmuller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics* 15, 610-625.

[68] Muller, H. G. and Stadtmuller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Annals of Statistics* 15, 182-201.

[69] Muller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* 50, 61-76.

[70] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications* 9, 141-142.

[71] Nelder, J. A. and Pregibon, D. (1987). An extended quasi likelihood function. *Biometrika* 74, 221-232.

[72] Neumann, M. H. (1994). Fully data driven nonparametric variance estimators. *Statistics* 25, 189-212.

[73] Patefield, W. M. (1977). On the maximized likelihood function. *Sankhya* B 39, 92-96.

[74] Pierce, D. A. and Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association* 81, 977-986.

[75] Priestley, M. B. and Chao, M. T. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society* B 34, 385-392.

[76] Pritchard, D. J., Downie, J. and Bacon, D. W. (1977). Further consideration of heteroscedasticity in fitting kinetic models. *Technometrics* 19, 227-236.

[77] Raab, G. M. (1981). Estimation of a variance function with application to radioimmunoassay. *Applied Statistics* 30, 32-40.

[78] Rao, J. N. K. and Subrahmanian, K. (1971). Combining independent estimators in linear regression with unequal variances. *Biometrics* 27, 271-290.

[79] Reinsch, C. (1967). Smoothing by spline functions. *Numerische Mathematik* 10, 177-183.

[80] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics* 12, 1215-1230.

[81] Rice, J. (1984). Boundary modification for kernel regression. *Communications in Statistics* Part A - Theory and Methods, 13, 893-900.

[82] Rice, J. and Rosenblatt, M. (1981). Integrated mean square error of a smoothing spline. *Journal of Approximation Theory* 33, 353-369.

[83] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Annals of Statistics* 11, 141-156.

[84] Robinson, P. M. (1986). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* 55, 875-892.

[85] Rodbard, D. and Frazier, G. R. (1975). Statistical analysis of radioligand assay data. *Methods of Enzymology* 37, 3-22.

[86] Rothenberg, T. J. (1984). Approximate normality of generalized least squares estimates. *Econometrics* 52, 811-825.

[87] Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 77, 828-838.

[88] Sadler, W. A. and Smith, M. H. (1985). Estimation of the response error relationship in immunoassay. *Clinical Chemistry* 31/11, 1802-1805.

[89] Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics* Part A - Theory and Methods, 14, 1123-1136.

[90] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *Journal of the Royal Statistical Society* B 47, 1-50.

[91] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[92] Theil, H. (1971). *Principles of Econometrics*, New York: John Wiley and Sons.

[93] Wahba, G. (1975). Smoothing noisy data by spline functions. *Numerische Mathematik* 24, 383-394.

[94] Watson, G. S. (1964). Smooth regression analysis. *Sankhya* A 26, 359-372.

[95] Wedderburn, R. W. M. (1974). Quasi likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* 61, 439-447.

[96] Weisberg, S. (1985). *Applied Linear Regression.* Wiley, New York.

[97] Wolter, K. M. and Fuller, W. A. (1982). Estimation of non linear errors in variables models. *Annals of Statistics* 10, 539-548.