Error Variance Estimation in Nonparametric Regression Models

By

Yousef Fayz M Alharbi

A thesis submitted to
The University of Birmingham
for the Degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
The University of Birmingham
April, 2013

Abstract

In this thesis, we take a fresh look at the error variance estimation in nonparametric regression models. The requirement for a suitable estimator of error variance in nonparametric regression models is well known and hence several estimators are suggested in the literature. We review these estimators and classify them into two types. Of these two types, one is difference-based estimators, whereas the other is obtained by smoothing the residual squares. We propose a new class of estimators which, in contrast to the existing estimators, is obtained by smoothing the product of residual and response variable. The properties of the new estimator are then studied in the settings of homoscedastic (variance is a constant) and heteroscedastic (variance is a function of x) nonparametric regression models.

In the current thesis, definitions of the new error variance estimators are provided in these two different settings. For these two proposed estimators, we carry out the mean square analysis and we then find their MSE-optimal bandwidth. We also study the asymptotic behaviour of the proposed estimators and we show that the asymptotic distributions in both settings are asymptotically normal distributions. We then conduct simulation studies to exhibit their finite sample performances.

Acknowledgements

I would like to thank my supervisor Dr. P.N. Patil for his great advices, help and guidance during this work and for giving me this topic.

I am grateful to Taibah University in Medinah and Saudi Arabia Government for granting me a scholarship to study a PhD degree at the University of Birmingham.

I am also very grateful to my family for their support and patient during my study at the University of Birmingham.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Regression analysis is one of the most widely used methodological tools in applied statistics. The main aim of regression analysis is to find a general relationship between a response variable and one or more predictor variables. For example, in the regression analysis with one predictor variable, the aim is to estimate the unknown mean of the response variable for a given value of the independent variable. For given data $(x_i\,,\,Y_i)$ $i = 1, 2, ...n$, which represent n values observed on the response Y, corresponding to the n values of the independent variable x, one can model the relation between x and Y as

$$Y_i = m(x_i) + \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \tag{1.1}$$

where ϵ_is denote the errors, which are assumed to be independent and identically distributed random variables while $m(x_i)$ represents the mean function, $E[Y_i|x_i]$.

The two most commonly used approaches to estimate the mean function in (1.1) are parametric and nonparametric methods. In the parametric approach, a specific functional form is assumed for the mean function. For instance, if we assume $m(x_i)$ to be a linear function, the model becomes

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \text{for} \;\; i = 1, 2, \cdots, n$$

where β_0 and β_1 are unknown parameters. The task here is to estimate these parameters. Theory for this model and for linear parametric models in general is well developed, and there are several text books one can refer to, such as Draper and Smith (1981) and Neter, Kutner, Nachtsheim and Wasserman (1996). In general, when the functional form of the relationship between the response and predictor variables complies with parametric assumptions, efficient inferential procedures are available. In addition, there are various software packages that facilitate the use of these data analytical tools. However, the important drawback of the parametric regression model is that, when the assumption of the functional form is not met, it can produce a high model bias for the mean function.

An alternative approach is to use nonparametric techniques to estimate the mean function. When functional form can not be assumed, this method can estimate the relationship between the response variable and the predictor variables. This approach is used often, because nonparametric techniques allow a regression curve to be estimated without making strong assumptions about the true shape of a regression function. In fact, the nonparametric regression models can be used to explore the nature of the functional form that one can use in parametric regression models. In other words, in nonparametric regression, we do not assume a functional form for the shape of the regression function. In nonparametric regression, although the first task is to estimate the mean function, estimation of error variance is also equally crucial because of the central role it plays in confidence bands for the mean function or tests of the hypothesis about the mean function. For the mean function estimation, several researchers have described ways to estimate it in nonparametric models, e.g. Gasser and Müller (1984), Müller (1987), Fan (1993) and Fan and Gijbels (1996). In this thesis, the focus will be on estimation of the error variance.

In this chapter, in section 1.2, we review the importance of the error variance estimation in regression models and briefly discuss the error variance estimation in parametric regression models. Since we will be addressing the issue of error variance estimation in two different settings of nonparametric regression, we first describe the two regression models in section 1.3. In the first setting, we assume error variance to be a constant, while in the other, error variance is assumed to vary with design points. For the former model, the literature review of the error variance estimation is given in section 1.4; for the latter model the relevant literature review is in section 1.5. It is observed that, in general, current procedures of error variance or variance function estimation in either settings use either residual-based or difference-based approach. In section 1.6, we propose a third approach to estimate the error variance, which uses some of the advantages of both residual- and difference-based approaches. In this section, we also point out the way in which the new estimator possesses the advantages of residual- and difference-based estimators. Finally, an outline of thesis is given in section 1.7.

1.2 Error Variance Estimation

Error variance estimation is one of the most important issues in regression models. The estimation of the error variance is essential to assess the variability of the estimated mean of Y_i given x_i. Therefore, the error variance plays an important role in regression analysis. For example, in the model (1.1), it is essential to know σ^2 to draw inferences about mean of Y_i and about regression coefficients; to assess the goodness of fit for the estimated mean function; to obtain a 95 % confidence interval for $m(x_i)$; and to predict a new Y for a given x. To sum up, almost in every inferential aspects, the knowledge of the error variance is essential.

In parametric regression models, the error variance can be a constant or a function of the independent variables. When the error variance is constant, it can be estimated by the ordinary least squared approach as follows. Suppose that

$$Y = X\beta + \epsilon$$

where X is an $(n \times p)$ matrix of the independent variables, β is an $(p \times 1)$ vector of the unknown parameters, which represent the regression coefficients, Y denotes an $(n \times 1)$ vector of the observations of the response variable and ϵ represents an $(n \times 1)$ matrix of random errors with zero mean and common variance σ^2. Then, the most commonly used estimator of the error variance based on the sum of squares of the residuals is given by

$$\hat{\sigma}_1^2 = \frac{Y^T Y - \hat{B}^T X^T Y}{n-p}, \tag{1.2}$$

where $\hat{B} = (X^T X)^{-1} X^T Y$. Note that if one denotes the fitted values by $\hat{Y} = X\hat{B}$, then $Y - \hat{Y} = \underline{e}$ is a vector of residuals and $\hat{\sigma}_1^2 = \frac{\underline{e}^T \underline{e}}{n-p}$.

In contrast, when the error variance is not a constant and varies with the levels of independent variables, the weighted least squares approach can be applied. The procedure of this approach is explained well in several text books such as Draper and Smith (1981). But before using a regression model with non-constant variance, one may assess the constancy of the variance using, for example, tests proposed by Levene (1960) or Breusch-Pagan (1979).

From the discussion related to the estimator $\hat{\sigma}_1^2$ in (1.2), it is clear that the basic idea in devising an estimator for error variance is to obtain residuals and then construct an estimator based on the sum of squares of the residuals. In nonparametric regression, a similar approach can be followed but first one needs to estimate the mean function to obtain residuals. The error variance estimation in nonparametric regression models is discussed in the following sections.

1.3 Error Variance Estimation in Nonparametric Regression Models

In nonparametric regression models, the error variance can be a constant or a function of independent variables as in the case of parametric regression models. In the case of constant error variance, all data points have the same error variance. We define the homoscedactic nonparametric regression model as

$$Y_i \ = \ m(x_i) \ + \ \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \tag{1.3}$$

where Y_i denotes the response variable, ϵ_is represent the errors, which are independent and identically distributed random variables with zero mean, $E(\epsilon_i) = 0$, and constant error variance σ^2. In this model, $m\,(x_i)$ represents the mean function $E[Y_i|x_i]$ and x_i s denotes the design points.

In contrast, when the variance of errors is a function of x_is, the variance changes as the x_is change. In other words, as the data points change, so does the error variance. In this case, we define the nonparametric regression model as

$$Y_i = m(x_i) + \sqrt{v(x_i)}\ \epsilon_i\,, \quad \text{for } \ i = 1, 2, \cdots, n, \tag{1.4}$$

where Y_i, x_i and $m\,(x_i)$ are the same as in the previous model (1.3), ϵ_is are independent random variables with zero mean and unit variance, while $v(x_i)$ denotes the variance function. Note that the above model is known as a heteroscedastic nonparametric regression model. In the next two sections, we review the literature on the error variance estimation in these two models.

1.4 Constant Error Variance Estimators in Nonparametric regression Models

In the literature, there are several estimators for the variance in the homoscedastic nonparametric regression model. In broad terms, these estimators can be classified into two classes: difference-based and residual-based estimators. For residual-based estimators, as noted earlier, one needs to estimate the mean function first. There are several approaches for estimating the mean function non-parametrically. However, the attention will be restricted to residual-based estimators, where mean function is estimated using either spline smoothing or kernel smoothing. At the end of this section, a comparison between the mean squared error of these estimators is drawn.

1.4.1 Residual-based Estimators Using Kernel Smoothing Method

As the name suggests, one is required to estimate the mean function first in order to obtain the residuals, and the residuals are then used to estimate the error variance. Hall and Marron (1990) have estimated the mean function by using a weighted average $\sum\limits_{j=1}^{n} w_{ij} Y_j$ where w_{ij}s are such that $\sum\limits_{j=1}^{n} w_{ij} = 1$ for each i. Thus, the ith residual is

$$\hat{e}_i = Y_i - \sum_{j=1}^{n} w_{ij} Y_j \quad \text{for} \quad i = 1, 2, ...n.$$

Then, their proposed residual-based estimator for the constant error variance is

$$\hat{\sigma}^2_{HM} = \frac{\sum\limits_{i=1}^{n} \left(Y_i - \sum\limits_{j=1}^{n} w_{ij} Y_j \right)^2}{\left(n - 2 \sum\limits_{i=1}^{n} w_{ii} + \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} w_{ij}^2 \right)}. \tag{1.5}$$

Observe that it is very similar to the standard error variance estimator, $\sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n-2}$, in the simple linear regression model with fitted value $\hat{Y}_i = \sum_{j=1}^{n} w_{ij} Y_j$ and divider $n-2$ replaced by $n - 2\sum_{i=1}^{n} w_{ii} + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^2$. The simplest form of w_{ij} is

$$w_{ij} = \frac{K(\frac{X_i - X_j}{h})}{\sum_{k=1}^{n} K(\frac{X_i - X_k}{h})}, \text{for } 1 \leq i, j \leq n,$$

where K is a kernel function and h is the parameter that controls the bandwidth of the kernel. If K is a function such that $K(u) > 0$, $\int K(u)\, du = 1$ and $\int u^2 K(u)\, du \neq 0$, K is a density function and is referred to as a second order kernel. The mean squared error or the integrated mean squared error of $\hat{m}(x_i) = \sum_{j=1}^{n} w_{ij} Y_j$ is heavily influenced by h. If h is large, the contribution of bias in the mean squared error becomes large, and if h is small, the contribution of variance becomes large in the second order. Thus, it is referred to as a smoothing parameter. Note that, the mean squared error of the estimates of the type $\hat{m}(x_i) = \sum_{j=1}^{n} w_{ij} Y_j$ can be improved by selecting a kernel function of rth order. The rth order kernel is defined as $\int K(u)\, du = 1$, $\int u^i K(u)\, du = 0$ for $i = 1, 2, ...r-1$ and $\int u^r K(u)\, du \neq 0$. For the detailed analysis of kernel-based estimators of mean regression function, see Hardle (1991). Hall and Marron showed that if the rth order kernel is used to estimate $m(x_i)$, then the mean squared error of the estimator in (1.5) is

$$MSE(\hat{\sigma}^2_{HM}) = n^{-1}\, var\,(\epsilon^2) + C_1\,(n^2\, h)^{-1} + C_2\, h^{4r} + o(n^2\, h)^{-1} + o(h^{4r})$$

where C_1 and C_2 are constants.

1.4.2 Residual-based Estimators Using The Spline Smoothing Method

There are several residual-based error variance estimators that use the spline smoothing method to estimate the mean function. In this subsection, we discuss some of the more important of these estimators as well as the concept of spline smoothing.

To estimate the mean function $m(x_i)$, Reinsch (1967) suggests estimating $m(x_i)$ by using the minimizer of the following least squared problem

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - m(x_i))^2 + \lambda \int_0^1 (m''(x_i))^2 \, dx$$

where λ is a parameter. This estimate of $m(x_i)$ is known as a cubic spline estimator. It is obvious that the performance of this estimator depends on the parameter λ, referred to as a smoothing parameter, and hence it is important to select λ appropriately. For the selection of λ , Wahba and Craven (1979) established the generalized cross-validation method. Further, Wahba (1990) proposed to select λ as the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{m}(x_i))^2 + \frac{2\sigma^2}{n} \; tr\, A(\lambda)$$

where $A(\lambda)$ is a $(n{\times}n)$ symmetric non-negative definite matrix and is such that $(\hat{m}(x))^T = A(\lambda)\; Y^T$. For a detailed discussion of this choice of λ and other properties, see Wahba (1990), Carter and Eagleson (1992) and Tong, Liu and Wang (2008). After finding the estimate of $m(x_i)$, the fitted values are

$$\hat{Y} = A(\lambda)\; Y.$$

Therefore, the residual sum of square is

$$RSS = Y^T\,(I - A(\lambda))^2\; Y.$$

Then an estimator of σ^2 is

$$\hat{\sigma_2^2} = \frac{Y^T\,(I - A(\lambda))^2\; Y}{tr\;(I - A(\lambda))} = \frac{RSS}{tr\;(I - A(\lambda))}.$$

As before, this estimator consists of the residual sum of squares divided by a normalizing factor. For details, see Wahba (1978). An alternative estimator has been considered by Buckley, Eagleson and Silverman (1988) and is defined as

$$\hat{\sigma}_3^2 = \frac{Y^T \left[I - A(\lambda)\right]^2 Y}{tr \, \left[(I - A(\lambda))^2\right]}.$$

Carter and Eagleson (1992) have shown that $\hat{\sigma}_3^2$ is an unbiased estimator for all λ. Another variant of Wahba estimator is defined by

$$\hat{\sigma}_4^2 = \frac{Y^T \left[I - A(\lambda)\right]^r Y}{tr \, \left[(I - A(\lambda))^r\right]}$$

and is studied by Thompson, Key and Titterington (1991). Here r is any integer instead of being two. When $r = 1$, this estimator has been studied by Ansely, Khon and Tharm (1990). When $r = 1$, the estimator is easier to find than other estimators of this type.

1.4.3 The Difference-based Estimators

The main advantage of the difference-based method is that the mean function estimation is not required. In this subsections, we describe some of the error variance estimators that use difference-based method.

The idea of the difference-based estimators is based on the the fact that if X_1 and X_2 are independent with same means and variances, then

$$E\left[\frac{(X_1 - X_2)^2}{2}\right] = \sigma^2. \tag{1.6}$$

Thus, if the regression function is assumed to be smooth, then for two consecutive observations in a small neighbourhood say, Y_i and Y_{i-1}, one expects $E\left[\frac{(Y_i - Y_{i-1})^2}{2}\right] \approx \sigma^2$.

Using this concept Rice (1984) has proposed the following estimator

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (Y_i - Y_{i-1})^2. \tag{1.7}$$

This is referred to as first-order difference-based estimator. By extending the idea of difference-based estimators to the second ordered differences, Gasser, Sroka and Jennen-Steinmetz (1986) proposed the following estimator

$$\hat{\sigma}_{GSJ}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} C_i^2 \, \hat{e}_i^2 \tag{1.8}$$

where $\hat{e}_i$ represents the difference between Y_i and the value at x_i of the line, which joins the two points (x_{i-1}, Y_{i-1}) and (x_{i+1}, Y_{i+1}), C_i^2s are selected such that $E(C_i^2 \, \hat{e}_i^2) = \sigma^2$ for all i when the mean function m is linear. When x_i's are equally spaced, Gasser $et\,al.$ (1986) show that the above estimator is reduced to

$$\hat{\sigma}_{GSJ}^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} (\frac{1}{2} Y_{i-1} - Y_i + \frac{1}{2} Y_{i+1})^2,$$

which is essentially the sum of squares of second ordered differences. Further, Buckley $et\ al.$ (1998) have shown that this estimator is essentially the Rice estimator of the second order. Gasser $et\,al.$ (1986) have also applied this estimator in nonlinear regression model. Lu (2012) extended the Gasser $et\,al.$ estimator to be used in complex surveys. For more details, see Lu (2012).

A difference-based estimator of rth order has been proposed by Hall, Kay and Titterington (1990). To estimate the error variance, they first order x_is such that $x_1 \leq x_2 \leq ...... \leq x_n$ and construct a sequence $\{d_k\}_{k=0}^{r}$ of real numbers such that

$$\sum_{k=0}^{r} d_k = 0\,, \ \ \sum_{k=0}^{r} d_k^2 = 1. \tag{1.9}$$

This sequence is referred to as sequence of the differences. Then, their error variance estimator is defined as

$$\hat{\sigma}^2_{HKT} = (n-r)^{-1} \sum_{j=1}^{n-r} \left(\sum_{k=0}^{r} d_k Y_{k+j} \right)^2 .$$

This estimator can be written as a quadratic form $Y^T DY/tr(D)$ where $D = D_1^T D_1$ and

$$\mathbf{D_1} = \begin{pmatrix} d_0 & \dots & d_r & 0 & \dots & 0 \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ 0 & \dots & 0 & d_0 & \dots & d_r \end{pmatrix} .$$

The condition in (1.9) is required to ensure that the above estimator is an unbiased estimator for σ^2. For more details, see for example Brown and Levine (2007).

If the mean function is smooth of pth order, Seifert, Gasser and Wolf (1993) have proved that it is not possible to find a difference-based estimator of pth order or less, which has a better mean squared error than $\hat{\sigma}^2_{GSJ}$. Therefore, they suggest an alternative estimator for the error variance. If we let ith pseudo-residual of order r to be

$$e_i = \sum_{k=0}^{r} d_{ik} Y_{i+k} \tag{1.10}$$

where

$$\sum_{k=0}^{r} d_{ik}^2 = \frac{1}{(n-r)} \quad \text{for } i = 1, ..., n-r. \tag{1.11}$$

Then, the pseudo-residuals are $e = C\,y,$ where matrix C is defined as

$$\mathbf{C} = \begin{pmatrix} d_{1,0} & \dots & & d_{1,r} & 0 & \dots & 0 \\ 0 & d_{2,0} & \dots & & d_{2,r} & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \dots & 0 & d_{(n-r),0} & & \dots & d_{(n-r),r} \end{pmatrix}.$$

Under the assumptions that x_is are equally spaced and that the mean function is smooth of pth order, the idea of Seifert, Gasser and Wolf estimator is to divide the differences of order r into some partition. To find these partitions, Seifert $et\,al.$ (1993) have defined the general divided differences of order $r = p + 1$ such that

$$\Delta^{(m,p)}y = \Delta D^{(p)}B^{(p)}.....D^{(1)}B^{(1)}y = \Delta\Delta^{(p)}y,$$

where Δ is $(n-r)\times(n-r+1)$ a bi-diagonal smoothing matrix such that

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & \delta_1 & & \\ & \ddots & \ddots & \\ & & 1 & \delta_1 \end{pmatrix},$$

$D^{(k)}$ is a $(n-k)\times(n-k)$ diagonal weight matrix such that

$$D^{(k)} = diag\left(\frac{1}{x_{i+k} - x_i}\right)_{i=1,(n-k)}$$

and $B^{(k)}$ is $(n-k)\times(n-k+1)$ a bi-diagonal matrix

$$\mathbf{B^{(k)}} = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}.$$

It should be noted that δ_1 is a weight and its optimal choice depends on the mean function, design points and sample size. After that, Seifert $et\,al.$ (1993) redefined the ith pseudo-residual by using weighted general divided differences as

$$e_i = w_i \; \Delta_i^{(m,p)} y$$

where $\Delta_i^{(m,p)}$ denotes the ith row of $\Delta^{(m,p)}$, w_i is a weight which is given to the ith row and the constraint (1.11) is satisfied. Then, the error variance estimator is

$$\hat{\sigma}^2_{SGW} = e^T e = y^T C^T C y = y^T A y$$

where $A = C^T C$.

Müller, Schick and Wefelmeyer (2003) propose another estimator for error variance. This estimator is a weighted estimator and it uses differences between any two different observations. First, x_is are assumed to be continuous and have a positive probability density function. The errors are assumed to have a finite fourth moment. Additionally, the mean function $m(x)$ is assumed to satisfy the Holder condition

$$\mid m(s) - m(t) \mid \leq C \mid s - t \mid^{\beta} \;, \quad s, t \in [0,1]$$

where C is a constant and β is a positive number less than one. Then, consider a symmetric and non-negative weight function w_{ij} such that

$$w_{ij} = \frac{1}{2\,h}(\frac{1}{\hat{g}_i} + \frac{1}{\hat{g}_j}) K\left(\frac{X_i - X_j}{h}\right) \quad \text{and} \quad \frac{1}{n(n-1)} \sum\sum_{i \neq j} w_{ij} = 1,$$

where $K(.)$ is a kernel function and

$$\hat{g}_k = \frac{1}{(n-1)\,h} \sum_{k \neq j} K\left(\frac{X_k - X_j}{h}\right) \;, \quad k = 1, 2, ...n.$$

Assume that the kernel function $K(.)$ is bounded, compactly supported in the interval $[-1, 1]$, symmetric and satisfies the conditions of the probability density function, and h is a suitable bandwidth. The drawback of this weight function is that it is not well defined when $\hat{g}_i = 0$. To solve this problem, Müller $et\,al.$ (2003) have suggested taking a bandwidth for which all $\hat{g}_i$s are all positive. Then, Müller $et\,al.$ (2003) defined their estimator for the error variance as

$$\hat{\sigma}^2_{MSW} = \frac{1}{n(n-1)} \sum_{i \neq j}\sum \frac{1}{2}(Y_i - Y_j)^2 w_{ij} \quad = \quad \frac{1}{2\sum\limits_{i \neq j}\sum w_{ij}} \sum_{i \neq j}\sum w_{ij}(Y_i - Y_j)^2.$$

Tong and Wang (2005) have proposed another estimator for the error variance. This estimator is developed by using the expectation of the Rice's estimator. First, Tong and Wang (2005) have shown that the Rice's estimator always has a positive bias. They suggest that a lag-k estimator of the Rice's estimator is defined as

$$\hat{\sigma}^2_R(k) = \frac{1}{2(n-k)} \sum_{i=k+1}^{n} (Y_i - Y_{i-k})^2, \quad k = 1, ...n-1.$$

Then, one can show that

$$E(\hat{\sigma}^2_R(k)) = \sigma^2 + Jd_k, \quad 1 \leq k \leq l \tag{1.12}$$

where l is a fixed number such that $l = o(n)$, J is the slope and d_k equals to $\frac{k^2}{n^2}$. Note that when $k = 1$, we can show that $J = \int\limits_0^1 \{g'(x)\}^2\, dx$. Therefore, Tong and Wang (2005) propose to estimate σ^2 by the intercept of the line described in (1.12). The equation (1.12) represents a simple linear regression model, where d_k is the independent variable. Let us define S_k as

$$S_k = \sum_{i=k+1}^{n} \frac{(Y_i - Y_{i-k})^2}{2(n-k)}, \quad 1 \leq k \leq l,$$

and a weight $w_k = (n-k)/N$, which is computed for an observation S_k where $N =$

$nl - \frac{l(l+1)}{2}$. Then, the following linear regression model is fitted

$$S_k = \alpha + \beta\, d_k + e_k, \quad k = 1, 2, ...l.$$

Then, $\hat{\sigma}^2_{TW}$ is the estimate of α obtained by minimising of the following weighted sum of squares with respect to α and β

$$\sum_{k=1}^{l} w_k (S_k - \alpha - \beta d_k)^2. \tag{1.13}$$

Thus, we obtain

$$\hat{\sigma}^2_{TW} = \hat{\alpha} = \bar{S_w} - \hat{\beta}\bar{d_w},$$

where $\bar{S_w} = \sum_{k=1}^{l} w_k S_k$, $\bar{d_w} = \sum_{k=1}^{l} w_k d_k$ and

$$\hat{\beta} = \frac{\sum_{k=1}^{l} w_k S_k (d_k - \bar{d_w})}{\sum_{k=1}^{l} w_k (d_k - \bar{d_w})^2}.$$

Park $et\,al.$ (2009) have used a local quadratic approximation approach to determine w_k and d_k in equation (1.13). Then, they estimated σ^2 using the same way of the Tong and Wang estimator (2005). For details, see Park $et\,al.$ (2009).

In the event $min\ \{\hat{g}_i = 0\}$, the Müller $et\,al.$ (2003) estimator is not well defined. To solve this problem, Tong $et\,al.$ (2008) have suggested another weight function of the form

$Y^T DY/tr(D)$ where

$$\mathbf{D} = \begin{pmatrix} \sum\limits_{j\neq 1} w_{1j} & -w_{21} & \dots & -w_{1n} \\ -w_{21} & \sum\limits_{j\neq 2} w_{2j} & \dots & -w_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ -w_{n1} & -w_{n2} & \dots & \sum\limits_{j\neq n} w_{nj} \end{pmatrix}$$

and w_{ij} are such that $w_{ij} = \frac{1}{h} K\left(\frac{x_i - X_j}{h}\right)$ where K is rth ordered kernel and satisfies the following conditions

$$\begin{cases} \int_{-1}^{1} K(u)du = 1 & \\ \int_{-1}^{1} u^i K(u)du = 0 & i = 1, ..., r-1 \\ \int_{-1}^{1} u^r K(u)du \neq 0 & i = r \\ \int_{-1}^{1} K^2(u)du < +\infty & \end{cases} .$$

This estimator could be used when x_is are equally spaced on [0,1] or when x_is are independent and identically distributed random variables with a density g on [0,1].

1.4.4 The Comparison of The Error Variance Estimators in Terms of The Mean Squared Error

Before comparing the mean squared error of the estimators discussed in the last three subsections, it should be noted that the mean squared error of the Hall and Marron estimator (1990) is

$$n^{-1}\, var\,(\epsilon^2) + C_1\, n^{-2}\, h^{-1} \;+\; C_2\, h^{4r} + o(n^2\, h)^{-1} \;+\; o(h^{4r})$$

where C_1 and C_2 are constants. For all other estimators, one can show that the mean squared error has the form

$$n^{-1}\, C_3 + n^{-2} h^{-1}\, C_4 + h^{2\,r}\, C_5 \,+ o(n^{-1}) + o\left(n^2 h\right)^{-1} + o(h^{2\,r})$$

where C_3, C_4 and C_5 are constants. It should be noted the constants C_is depend on the kernel and mean functions. From the above, it is clear that the Hall and Marron estimator (1990) has the smallest relative error. To define relative error, we suppose the optimal bandwidth to be $h \sim n^{-\alpha}$ where $0 < \alpha < 1$. Then, the size of relative error is defined as $n^{\alpha-1}$ in the following equality

$$\text{MSE} = n^{-1} \left[\text{constant} + \text{constant} \cdot n^{\alpha-1} \right].$$

So, the size of the relative error of the Hall and Marron estimator is of order $n^{-(4r-1)/(4r+1)}$ when the optimal bandwidth is chosen as $O\left(n^{-2/(4r+1)}\right)$. In contrast, none of the difference-based estimators achieves this size for their relative errors. Tong, Liu and Wang (2008) argue that this may be because the difference-based estimators do not require the estimation of the mean function. They have also noted that the size of relative errors do not imply a better performance in the finite sample properties. The residual-based estimators with their optimal bandwidths, such as Hall and Marron estimator, have achieved the following optimal rate in the first order

$$MSE(\sigma^2) = n^{-1}\, var(\epsilon^2) + o(n^{-1}). \tag{1.14}$$

Dette $et\,al.$ (1998) have shown that none of the fixed order difference-based estimators achieves this optimal rate. In contrast, estimators proposed by Müller $et\,al.$ (2003), Tong and Wang (2005) and Tong $et\,al.$ (2008) are not fixed order difference estimators, so they do achieve the optimal rate in (1.14).

1.5 Functional Variance Error Estimators

In this section, we review the literature on the error variance function estimators. As before, we classify the estimators of the error variance function into two classes: residual-based and difference-based estimators. The residual-based estimators of the error variance function are discussed in the next subsection and the difference-based estimators are discussed in the following subsection.

1.5.1 The Residual-Based Estimator for the Error Variance Function

The main idea behind these estimators is the same as the idea behind the residual-based estimators in the constant error variance model, except that now one has to account for the changes in variance as the design points x_is change. Thus, in this case, the interest is in the estimation of the function $v(x_i)$ (or $v(i/n)$, if x_is are equispaced design points in $[0\,,\,1]$). Hall and Carroll (1989) defined one of the first estimators for the error variance function v. To see how it works, assume that m and v are bounded functions, x_is are equidistant design points in the interval [0,1], the fourth moment of ϵ_i s are bounded and the mean function m has s_1 derivatives, whereas the variance function v has s_2 derivatives. So, the model (1.4) can be written as

$$Y_i = m(i/n) + \sqrt{v(i/n)}\,\epsilon_i\,, \quad 1 \leq i \leq n.$$

Also, assume that $0 < h \leq 1$, $\gamma \geq 0$ is an integer and $c_j = c_j(h,n),\ -\infty \leq j \leq +\infty$ are constants, which satisfy the following constraints

$$\begin{aligned} \mid c_j \mid \ &\leq\ Ch, \quad c_j = 0 \ \text{ for } \mid j \mid \geq Ch^{-1},\ \sum_j c_j = 1 \\ &\text{and } \sum_j j^i c_j = 0 \qquad \text{for } \ 1 \leq i \leq \gamma \end{aligned} \tag{1.15}$$

where the constant C does not depend on h. The c_js could be found for a smooth kernel function such that $c_j \ = \ h\,K(hj)$ where the function K satisfy

$$\int K(u)du = 1, \qquad \int u^j K(u)du = 0 \quad for \quad 1 \leq j \leq \gamma$$

and is compactly supported on [-1,1]. To estimate the mean function, select a series of constants $a_j \equiv c_j(h_1, s_1)$ such that the condition (1.15) is satisfied. Then, we can estimate the mean function at the point i/n by

$$\hat{m}(i/n) = \sum_j a_j Y_{i+j}, \quad 0 \leq i \leq n, \tag{1.16}$$

where Y_j is zero if $j < 0$ or $j > n$. So, the residual can be written as $e_i = Y_i - \hat{m}_i(i/n)$ for $i = 1, 2, ...n$, where $\hat{m}_i(i/n)$ is defined in (1.16). Then, construct $\hat{m}(x)$ for general $x \in [0, 1]$ by using linear interpolation technique on $\hat{m}(i/n)$. Now set $r_i = Y_i - m(i/n)$, then

$$r_i^2 = v(i/n) + v(i/n)\delta_i, \quad 1 \leq i \leq n$$

where $\delta_i^2 = \epsilon_i^2 - 1$, which has zero mean and δ_i^2s are independent and identically distributed random variables. Furthermore, set $r_i = 0$ if $i < 1$ *or* $i > n$. Now, to estimate the error variance function, find a sequence $b_j \equiv c_j(h_2, s_2)$ so that the condition (1.15) holds. Then, we can define an estimate of $v(i/n)$ to be

$$\tilde{v}(i/n) = \sum_j b_j r_{i+j}^2, \quad 1 \leq i \leq n.$$

Finally, construct $\tilde{v}(x)$ by using a linear interpolation technique on $\tilde{v}(i/n)$. Clearly, $\tilde{v}(x)$ is not a realistic estimator, since r_is are not known. Therefore, Hall and Carroll (1989) propose to estimate $v(x)$ by the following procedure. First, obtain $\hat{v}(i/n)$ by

$$\hat{v}(i/n) = \sum_j b_j e_{i+j}^2, \quad 1 \leq i \leq n$$

and then we can define an estimate of $v(x)$ to be $\hat{v}_1(x)$ as a linear interpolation of $\hat{v}(i/n)$. The properties of this estimator are discussed in Hall and Carroll (1989).

Ruppert, Wand, Holst and Hössjer (1997) and Fan and Yao (1998) have used local polynomial smoothing approach to estimate the error variance function. To define their estimators, let $X_p(x)$ be

$$\mathbf{X_p}(\mathbf{x}) = \begin{pmatrix} 1 & X_1 - x & \dots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & X_n - x & \dots & (X_n - x)^p \end{pmatrix}$$

and

$$W_h(x) = \ diag \left\{ K\left(\frac{X_1 - x}{h_1}\right), ..., K\left(\frac{X_n - x}{h_1}\right) \right\},$$

be the weight matrix where $K(.)$ is a kernel function that satisfies the condition of the probability density function and h_1 is a smoothing parameter. To estimate the mean function, we can define the p_1th degree local polynomial smoother matrix , $S_{p1,h1}$ whose (i,j)th entry such that

$$(S_{p1,h1})_{ij} = \zeta_1^T \{X_{p1}^T(X_i) W_{h1}(X_i) X_{p1}(X_i)\}^{-1} X_{p1}^T(X_i) W_{h1}(X_i) \zeta_j$$

where ζ_k is a column vector with zero everywhere except the kth position, which is one. Thus, the residuals are, $e = (I - S_{p1,h1})Y$. Then, Fan and Yao (1998) use local linear smoothing to estimate the error variance function, where $\hat{v}_2(x) = \hat{\alpha}$ and $\hat{\alpha}$ is obtained by solving the following minimising problem with respect to α and β

$$(\hat{\alpha}, \hat{\beta}) = arg \ \underset{\alpha,\beta}{min} \sum_{i=1}^{n} \{e_i^2 - \alpha - \beta(X_i - x)\}^2 \ K\left(\frac{X_i - x}{h_2}\right)$$

where $K(.)$ is a kernel function with bandwidth h_2.

Ruppert, Wand, Holst and Hössjer (1997) find the estimation of the conditional error variance function using local polynomial regression, such that

$$\hat{v}_3(x) = \hat{v}(x; p_1, h_1, p_2, h_3) = \frac{\zeta_1^T \{X_{p2}^T(X_i) W_{h3}(X_i) X_{p2}(X_i)\}^{-1} X_{p2}^T(X_i) W_{h3}(X_i) e^2}{1 + \zeta_1^T \{X_{p2}^T(X_i) W_{h3}(X_i) X_{p2}(X_i)\}^{-1} X_{p2}^T(X_i) W_{h3}(X_i) \Delta}$$

where $\Delta = diag(S_{p1,h1} \; S_{p1,h1}^T - 2S_{p1,h1})$, h_3 is an appropriate bandwidth for the kernel function and p_2 is the degree of the local polynomial, which is used in the estimation of the error variance function. Ruppert $et\,al.$ (1997) have shown that this estimator can be defined as

$$\hat{v}_3(x) = \frac{S_{p2,h3}\; e^2}{1 + S_{p2,h3}\Delta}.$$

However, Fan and Yao (1998) have proven that their estimator is asymptotically normal and when second-order kernel function is used, its mean squared error is

$$MSE(\hat{v}_2(x)) \;=\; n^{-1}\, h_2^{-1}\, C_1(x) \;+\; h_2^4\, C_2(x) \;+\; o(n^{-1}\, h_2^{-1}) \;+\; o(h_2^4) \tag{1.17}$$

where $C_1(x)$ and $C_2(x)$ are deterministic functions and $\hat{v}_2(x)$ represents the Fan and Yao estimator. The Ruppert $et\,al.$ estimator has the same form of the mean squared error in (1.17), but the deterministic functions might be different than that of the Fan and Yao estimator.

One of the drawbacks of using local polynomial regression in the estimation of the variance function is that the estimated variance function can be negative when the bandwidths are not selected appropriately. To avoid this drawback, Yu and Jones (2004) have proposed a local linear estimator such that the estimated variance function is always positive. For more details, see Yu and Jones (2004).

1.5.2 The Difference-Based Estimator for the Error Variance Function

As for the constant error variance estimators, when the variance is a function of x_is, the mean function is not required to be estimated in the difference-based estimators for the error variance function. The first estimator has been developed by Müller and Stadtmüller (1987). To see how it works, first, assume that in the model (1.4) x_is are equally spaced on $[0,1]$, the fourth moment of ϵ_is is bounded and $v(x)$ is Lipschitz continuous function with Lipschitz constant $\gamma \in (0,1]$. Thus, the error variance varies smoothly when the design points change. Then, we can estimate the local variance function $v(x)$ by using

$$\tilde{v}(x_i) = \tilde{\sigma}_i^2 = \left(\sum_{j=j_1}^{j_2} w_j Y_{j+i} \right)^2$$

where $x_i \in (0,1)$, $j_1 = -\lceil k/2 \rceil$, $j_2 = \lceil k/2 - 1/4 \rceil$ and $k \geq 2$ is a fixed integer. $\lceil b \rceil$ denotes the largest integer number $\leq b$. To ensure asymptotic unbiasedness of this estimator, it is necessary to have

$$\sum_{j=j_1}^{j_2} w_j = 0 \text{ and } \sum_{j=j_1}^{j_2} w_j^2 = 1.$$

However, Müller and Stadtmüller have shown that this estimator is not consistent. Therefore, Müller and Stadtmüller (1987) proposed modification to the above estimator; more specifically they used smoothing in the neighbourhood of $\tilde{v}(x_j)$

$$\hat{v}_4(x) = \frac{1}{h} \sum_{j=1}^{n} \int_{S_j - 1}^{S_j} K\left(\frac{x-u}{h} \right) du \;\; \tilde{v}(x_j)$$

where $S_j = \frac{x_j + x_{j+1}}{2}$ for $1 \leq j \leq n$, $S_o = 0$ and $S_n = 1$, K denotes a kernel function and the bandwidth h satisfies the following constraint

$$h \longrightarrow 0 \;, \quad nh \longrightarrow +\infty \quad as \; n \longrightarrow +\infty.$$

Müller and Stadtmüller (1987) have shown that the above estimator is uniformly consistent.

Brown and Levine (2007) have used a class of difference-based estimators for estimating the error variance function in the heteroscedastic nonparametric regression model. Here, x_is are equidistant points on $[0,1]$. To estimate the error variance function, first, define a pseudo-residual of order r to be

$$\Delta_i = \Delta_{r,i} = \sum_{k=0}^{r} d_k \; Y_{i+k-\lfloor r/2 \rfloor}, \qquad i = \lfloor r/2 \rfloor + 1, ..., n + \lfloor r/2 \rfloor - r$$

where $\lfloor a \rfloor$ represents the largest integer number that is less than a and the weight d_js are such that

$$\sum_{i=0}^{r} d_i = 0 \;\text{ and }\; \sum_{i=0}^{r} d_i^2 = 1.$$

Then, we can obtain the error variance function estimator $\hat{v}_5(x)$ using local polynomial smoothing of the squared pseudo-residual where $\hat{v}_5(x) = \hat{b}_0$ and $\hat{b}_0$ is such that

$$(\hat{b}_0, \hat{b}_1, ..., \hat{b}_p) = \underset{\hat{b}_0, \hat{b}_1, ..., \hat{b}_p}{arg\ min} \sum_{i=\lfloor r/2 \rfloor+1}^{n+\lfloor r/2 \rfloor - r} \left[\Delta_{r,i}^2 - \hat{b}_0 - \hat{b}_1(x - x_i) - ... - \hat{b}_p(x - x_i)^p \right]^2 \times K\left(\frac{x - x_i}{h}\right).$$

K is a kernel function that satisfies the standard conditions. That is, it is bounded, compactly supported and not identically equal to zero. Note that the kernel function in this estimator is used to account for variation in the variance when x_is change. Brown and Levine (2007) have shown that when a constant mean function has been used in the constant variance model, this estimator is unbiased and its mean squared error is stated in (1.17) where the deterministic functions $C_1(x)$ and $C_2(x)$ are different from that of Fan and Yao's estimator.

Wang, Brown, Cai and Levine (2008) have proposed an alternative estimator for the error variance function. First, assume that the mean function has α derivatives and the error variance function has β derivatives. Then, set $D_i = Y_i - Y_{i+1}$ for $i = 1, 2, ..., n-1$. Thus,

$$D_i = Y_i - Y_{i+1} = m(x_i) - m(x_{i+1}) + v^{1/2}(x_i)\epsilon_i - v^{1/2}(x_{i+1})\epsilon_{i+1} = \delta_i + \sqrt{2} v_i^{1/2} z_i$$

where $\delta_i = m(x_i) - m(x_{i+1})$, $v_i^{1/2} = \sqrt{\frac{1}{2}\,(\,v(x_i) - v(x_{i+1}))}$ and

$$z_i = (v(x_i) + v(x_{i+1}))^{-1/2}\left(v^{1/2}(x_i)\epsilon_i - v^{1/2}(x_{i+1})\epsilon_{i+1}\right).$$

Note that z_is have zero mean and unit variance. Then, assume that there is a kernel function $K(.)$ that satisfies

$$\int_{-1}^{1} K(t)\,dt = 1, \quad \int_{-1}^{1} t^i\,K(t)\;dt = 0 \quad \text{for } i = 1, 2, ..., \lfloor\beta\rfloor, \quad \int_{-1}^{1} K^2(t)\;dt < +\infty, \tag{1.18}$$

which is bounded and compactly supported on $[-1, 1]$. To avoid the boundary effect, define another kernel function $K_u(x)$ that satisfies the condition (1.18) for all $u \in [0, 1]$ and is compactly supported on $[-1, u]$. After that, for $i = 2, 3, ..., n-2$ and for any $x \in [0, 1]$, $0 < h < 1/2$, define the following weighted kernel function

$$K_i^h(x) = \begin{cases} \int_{(x_i+x_{i-1})/2}^{(x_i+x_{i+1})/2} \frac{1}{h} K\left(\frac{x-t}{h}\right)\;dt & \text{when } x \in (h, 1-h) \\ \int_{(x_i+x_{i-1})/2}^{(x_i+x_{i+1})/2} \frac{1}{h} K_u\left(\frac{x-t}{h}\right)\;dt & \text{when } x = u\,h \text{ for some } u \in [0,1] \\ \int_{(x_i+x_{i-1})/2}^{(x_i+x_{i+1})/2} \frac{1}{h} K_u\left(-\frac{x-t}{h}\right)\;dt & \text{when } x = 1 - u\,h \text{ for some } u \in [0,1] \end{cases}.$$

For $i = 1$, the integral is taken from 0 to $(x_1 + x_2)/2$. For $i = n-1$, the integral is taken from $(x_{n-1} + x_{n-2})/2$ to 1. Then, the error variance function $\hat{v}$ is estimated for any $0 \le x \le 1$ as

$$\hat{v}_6(x) = \frac{1}{2}\sum_{i=1}^{n-1} K_i^h\;D_i^2$$

where $\sum_{i=1}^{n-1} K_i^h = 1$. Wang, Brown, Cai and Levine (2008) provided numerical results, that show that when the mean function is very smooth, the discrete mean squared error of their estimator is smaller than that of the Fan and Yao estimator.

1.6 A New Class of The Error Variance Estimators

In this section, we propose a new class of the error variance estimators for nonparametric regression models in two settings. These settings are designed for constant variance and function variance models. Before we propose alternative estimators for the error variance or the error variance function, we make a note of variance estimation, when one has independent and identically distributed random variables $X_1, X_2, \cdots, X_n$ with mean μ and variance σ^2 with the fourth moment $< \infty$.

1.6.1 Estimation of Variance in Independent and Identically Distributed Random Sample

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with mean μ and variance σ^2 with a bounded fourth moment. The variance of this population can be estimated at least in the following three different ways

$$1)\ \hat{\sigma}^2_{d1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2; \tag{1.19}$$

$$2)\ \hat{\sigma}^2_{d2} = \sum_{i=1}^{n-1} \frac{(X_i - X_{i+1})^2}{2(n-1)}; \tag{1.20}$$

$$3)\ \hat{\sigma}^2_{d3} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n(n-1)} \sum \sum_{j \neq i} X_i X_j = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \frac{1}{n-1} \sum_{j \neq i} X_j \right) X_i. \tag{1.21}$$

Furthermore, all of the above estimators are unbiased estimators for σ^2. The variance of the above estimators, $var(\hat{\sigma}^2)$, is $\frac{2}{n-1}\sigma^4$ when the population is normally distributed.

1.6.2 Estimation of The Error Variance in Nonparametric Regression Models

In nonparametric regression models, all existing estimators for error variance belong to one of the following two classes. The first class is known as residual-based estimators, because of its dependence on the residual sums obtained from a nonparametric regression fit for the mean function. The idea of residual-based error variance estimators has been developed from the definitions (1.19) for the variance. Thus, in the constant error variance setting, the error variance can be defined as

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{m}(x_i))^2 .$$

As discussed in section 1.4, in this class of estimators, the estimation of the mean function is required. In its estimation using one of the nonparametric methods such as spline smoothing or kernel smoothing, all observations are used. Similarly, when the variance is a function of x_is, the form of the residual-based estimators for the error variance function is

$$\hat{v}(x) = \sum_{i=1}^{n} w_i(x) (Y_i - \hat{m}(x_i))^2 \tag{1.22}$$

where $w_i(x)$s are weight functions and $\hat{m}(x_i)$ is estimated as in the residual-based estimators for the constant error variance.

The second class of estimators is known as the difference-based estimators. The idea for these estimators comes from the definition (1.20). In the regression setting, the constant error variance can be estimated as

$$\hat{\sigma}^2 = \sum_{i=2}^{n} \frac{(Y_i - Y_{i-1})^2}{2(n-1)} .$$

Thus, the idea behind these estimators is the same as the difference idea in the time series analysis. That is, a trend is removed by the operation of differencing. This class of estimators is used widely because it is easy to implement and it does not require the estimation of the mean function. Furthermore, it often has small biases for small sample sizes as Dette, Munk and Wanger (1998) have noted. When the variance is a function of x_is, the above definition of the estimator is modified as

$$\hat{v}(x) = \sum_{i=2}^{n} w_i(x)\,(Y_i - Y_{i-1})^2.$$

where $w_i(x)$ is as in equation (1.22). It should be noted that the above discussion is for the simplest difference-based estimator. The ideas can be extended easily for the other difference-based estimators for the error variance.

The basic idea behind the estimators proposed below comes from a definition for the variance given in (1.21). Using this definition, a new class for the error variance estimators can be defined in the two different settings that are mentioned earlier. In the first setting of constant variance model in (1.3), where x_is are equidistant design points, $e_i = Y_i - \hat{m}(x_i)$ and the errors are independent and identically distributed random variables with zero mean and common variance σ^2, then the error variance can be estimated as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i\,Y_i. \tag{1.23}$$

In the above class of the error variance estimators, $e_i\,Y_i$s are averaged to estimate the error variance as opposed to averaging of e_i^2s as used in the residual-based estimators. To estimate error variance function $v(x)$ in the model in (1.4), a simple modification to (1.23) leads to estimators

$$\hat{v}(x) = \sum_{i=1}^{n} w_i(x)\,e_i\,Y_i. \tag{1.24}$$

Again, note that we are smoothing $e_i\,Y_i$ as opposed to smoothing e_i^2 which is used in the standard residual-based estimators of $v(x)$.

Clearly, the advantage difference-based estimators have over the residual-based estimators is that the estimation of the mean function is not required. However, at the same time, in difference-based estimators, order of the differences plays an important role and hence one needs to decide the order of difference. Also, the higher the order of differences, the more the lack of information on function $v(x)$ near the boundary which is not the case of the residual-based estimators.

Also, note that in estimators of the type defined in (1.23) and (1.24) or in any other residual-based estimators, to estimate the error variance function at point x_i, one uses the observation at x_i for both in order- to estimate the mean function and again through residual in order to estimate variance or variance function. Thus, our interest is to consider an estimator that does not use ith observation to estimate the mean function and is used only to estimate variance or variance function, and to investigate its properties. Therefore, rather than studying estimators defined in (1.23) and (1.24), we define, in the constant variance model in (1.3),

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{m}_{-i}(x_i))\, Y_i \tag{1.25}$$

and in the variance function model in (1.4),

$$\hat{v}(x) = \sum_{i=1}^{n} w_i(x)\ (Y_i - \hat{m}_{-i}(x_i))\, Y_i \tag{1.26}$$

where $\hat{m}_{-i}(x_i)$ denotes the estimate of mean function without using the ith observation and w_is are weight functions. To seek the answers to the questions of whether smoothing of e_iY_i has any advantage over smoothing e_i^2, and whether not using observation Y_i in the estimation of $m(x_i)$ has any advantage over using it $-$ we study estimators of the types in (1.25) and (1.26). Now, to estimate the mean function, we define the following weight function $w_{i\,j}$

$$w_{ij} = \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum\limits_{i \neq j} K\left(\frac{x_i - x_j}{h}\right)}, \qquad 1 \leq i\,,\, j\, \leq\, n \tag{1.27}$$

where h refers to a bandwidth parameter and $K(.)$ is a kernel function, satisfying the following assumptions

A1:

$$\int u^i\, K(u)du = \begin{cases} 1 & \text{for } i = 0; \\ 0 & \text{for } i = 1, ...r - 1; \\ \neq 0 & \text{for } i = r. \end{cases}$$

A2: It is bounded, symmetric around 0 and compactly supported on $[-1, 1]$.

We note that the weight function satisfies the constraint $\sum\limits_{j \neq i}^{n} w_{i\,j} = 1$ for each i . So, the mean function is estimated by $\hat{m}_{-i}(x_i) = \sum\limits_{j \neq i} w_{i\,j}\, Y_j$. It is clear that the ith observation is not used in the estimation of the mean function at point x_i. From (1.21) and (1.25), the new estimator of the constant error variance can be written as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j \neq i} w_{i\,j}\, Y_i\, Y_j. \tag{1.28}$$

In the variance function setting, assume the model in (1.4) where x_is are fixed design points and ϵ_is are independent and identically distributed with zero mean and unit variance. Thus, a new estimator for the error variance function can be defined as

$$\hat{v}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_2} K\left(\frac{x - x_i}{h_2}\right) \left\{ Y_i - \frac{1}{(n-1)h_1} \sum_{j \neq i} K\left(\frac{x_i - x_j}{h_1}\right) Y_j \right\} Y_i \tag{1.29}$$

where $K(.)$ is as in equation (1.28) and h_1 and h_2 are two different bandwidths. The bandwidth h_1 is used to estimate the mean function, while the other bandwidth is used to estimate the variance function. In the above estimator, to account for the changes in the variance as the design points x_is change, we use a kernel weight function.

1.7 Outlines of The Thesis

In this thesis, we first discuss the properties of variance estimator in (1.28) proposed in the setting of the constant variance regression model in (1.3). The similar investigation is then carried out for the variance function estimator in (1.29) where the regression model is given by (1.4). This thesis is organised as follows.

The theoretical properties of the estimator in (1.28) are studied in the second chapter. That is, the asymptotic mean square analysis for this estimator is carried out, and precise asymptotic expressions for mean and variance are obtained. A comparison between the mean squared error for the estimator in (1.28) and other estimators is also provided. We also note the effect of the bandwidth h on the mean squared error of this estimator. Then, the asymptotic distribution of the estimator in (1.28) is studied and shown to have asymptotically normal distribution. In the third chapter, the simulation studies are considered to exhibit the finite sample performance of the estimator in (1.28). In particular, the effects of the bandwidth selection and the different forms of the mean functions on the finite sample performance of the estimator in (1.28) are investigated.

In view of the advantages mentioned in section 1.6, the asymptotic properties of the estimator in (1.29) are investigated. Clearly, the estimator in (1.29) has two different bandwidths. The first one h_1 is used to estimate the mean function, whereas the other h_2 is used to estimate the variance function. Therefore, in chapter 4, we investigate the effect of the bandwidths h_1 and h_2 on the mean square analysis of the estimator in (1.29). In chapter 5, the asymptotic distribution of this estimator is studied and shown to be asymptotically normal distribution. As noted in section 1.6, one of the advantages of the estimator in (1.29) over the difference-based estimators is that it estimates the boundary of the variance function with smaller bias compared with that of the difference-based estimators. Hence, the finite sample properties of estimator in (1.29) are studied along with the difference-based estimators. Thus, we investigate the effect of the mean and variance functions and bandwidth selections on the finite sample performance of the estimator (1.29) in chapter 6. In the last chapter, the conclusion is drawn and future work is suggested.

Chapter 2

The Theoretical Properties of a New Estimator in the Setting Of Homoscedastic Nonparametric Regression Model

2.1 Introduction

In this chapter, the mean square analysis of the new estimator, proposed in section 1.6, is considered in the settings of the constant error variance model. We start with the following homoscedastic nonparametric regression model

$$Y_i = m(x_i) + \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \tag{2.1}$$

where $m(x_i)$ is the mean function $E(Y_i|x_i)$, Y_is represent the response variable and x_is denote the design points. The errors ϵ_is are assumed to be independent, identically distributed and random with zero mean and variance σ^2 and the fourth moment is bounded ($E(\epsilon^4) < \infty$).

Recall that the new estimator for the constant error variance is defined as

$$\hat{\sigma^2} \;=\; \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} w_{i\,j}\, Y_i\, Y_j \tag{2.2}$$

where $K(.)$ is a kernel function and $w_{i\,j}$s are defined in (1.27). If x_is are random with a density function $f(x)$, then we obtain

$$w_{ij} = \frac{K\left(\frac{x_i-x_j}{h}\right)}{\sum\limits_{i\neq j} K\left(\frac{x_i-x_j}{h}\right)} \approx K\left(\frac{x_i-x_j}{h}\right)/(n-1)\,h\,f(x_i).$$

Further, when $f(x)$ is the density function of the uniform distribution and x_is $\in [0,\,1]$, $w_{ij} \approx K\left(\frac{x_i-x_j}{h}\right)/(n-1)\,h$. So, the estimator in (2.2) can be written as

$$\hat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} Y_i^2 \;-\; \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h}\right) Y_i\, Y_j\,. \tag{2.3}$$

Another aim of this chapter is to derive the asymptotic distribution for the new estimator defined in (2.3).

This chapter is organised as follows. In section 2.2, the main results are stated, which provide the bias and variance of the estimator in (2.3) and its asymptotic distribution. Lemmas, which are later used to prove the main results of section 2.2, are stated in section 2.3. An outline of proof of the bias and the variance of the estimator in (2.3) is given in section 2.4, whereas a sketch of the proof of its asymptotic distribution is provided in section 2.5. As expected the bias and variance of the estimator in (2.3) depends on the bandwidth selection and thus the optimal bandwidth and its choice is discussed in section 2.6.

2.2 The Main Results

The following assumptions are made in addition to the assumptions $A1$ and $A2$ in section 1.6:

B1: $K'(u)$ exists for $u \in [-1, 1]$.

B2: x_is are equispaced design points in the interval [0,1] such that $x_i = i/n$ for $i = 1, 2, \cdots, n$.

$B2'$: The design points x_is are randomly chosen from the $U[0, 1]$ distribution.

B3: The mean function $m(x)$ is bounded, differentiable and has r-continuous derivatives where $r \geq 2$.

B4: $h \to 0$ such that $nh \to \infty$ as $n \to +\infty$.

Then the following theorem provides the bias and variance formulae of $\hat{\sigma^2}$.

Theorem **2.2.1.** *Suppose A1, A2, B1, B2, B3 and B4 are true and* $h \sim n^{-\alpha}$*, where* α *is positive number such that* $1/3 < \alpha < 1$*, then for* $\hat{\sigma^2}$ *in (2.3)*

$$(i)\, E(\hat{\sigma^2}) - \sigma^2 = h^r \cdot C_1 + o(h^r) + O(n^{-1}),$$

$$(ii)\, Var\,(\hat{\sigma^2}) = n^{-1}C_2 + n^{-2}\, h^{-1}C_3 + o(n^2\, h)^{-1},$$

where

$$C_1 = \frac{(-1)^r}{r!} \int_0^1 K(y) y^r \, dy \int_0^1 m(t)\, m^{(r)}(t) dt,$$

$$C_2 = \mu_4 - \sigma^4, \quad \& \quad \mu_r = E[(Y_i - m(x_i))^r] \quad \textit{and}$$

$$C_3 = 2\sigma^4 \int_0^1 K^2(y) dy + 4\sigma^2 \int_0^1 K^2(y) dy \int_0^1 m^2(x) dx.$$

Remark:

1) By using the results in the above theorem, the mean squared error can be written as

$$\begin{aligned} MSE(\hat{\sigma^2}) &= \left(E(\hat{\sigma}^2) - \sigma^2\right)^2 + Var\left(\hat{\sigma}^2\right) \\ &= n^{-1}\, C_2 + n^{-2}\, h^{-1}\, C_3 + h^{2\,r}\, C_1^2 + o\left(n^2 h\right)^{-1} + o(h^{2\,r}). \end{aligned} \tag{2.4}$$

Note that the bias is contributed by C_1, whereas the variance is contributed by C_2 and C_3. Clearly, when the bandwidth h is selected as stated in the above theorem,

$$MSE(\hat{\sigma^2}) \sim n^{-1} \cdot var(\epsilon^2)$$

Thus, the estimator in (2.3) achieves the same minimum mean squared error in the first order like some other estimators discussed in the literature for error variance, such as the estimators of Hall and Marron (1990), Müller, Schick and Wefelmeyer (2003) and Tong, Liu and Wang (2008).

2) In the second order, when the bandwidth of the estimator in (2.3) is chosen appropriately, it is expected to have a similar behaviour to that of the Hall and Marron estimator. Note that the detailed discussion is a referred to section 2.6.

Now, to find the asymptotic distribution of the new estimator, first note that the estimator in (2.3) can be expressed as

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} Z_i - \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right) U_i U_j \\ &- \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right) m(x_i)\, m(x_j) \end{aligned} \tag{2.5}$$

where $Z_i = Y_i^2 - b_i\,[Y_i - m(x_i)]$, $U_i = Y_i - m(x_i)$ and $b_i = \frac{2}{(n-1)\,h}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right) m(x_j)\,.$

The equation (2.5) consists of three terms, of which the first term is the sum of independent random variables, the second is a quadratic form and the third term is deterministic. The first term is $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Since Z_is are independent, the Lindeberg-Feller central limit theorem is used to derive its asymptotic distribution. The second term is a quadratic form in U_is. We note that $E[U_i] = 0$ and $E[\,U_i\,U_j|U_i] = 0$. Therefore, theorem 2.1 developed by De Jong (1987) is applied to derive its asymptotic distribution, which is given in the following theorem. Therefore, set $T_n(h) = \sum\sum_{1\le i<j\le n} T_{ij}$ and $T_{ij} = \frac{2}{n\,(n-1)\,h}\,K\left(\frac{x_i - x_j}{h}\right)\,U_i\,U_j$.

Theorem **2.2.2.** *Under the assumptions A1, A2, B1, B2, B3 and B4*

$$n\sqrt{h}\;T_n(h)\;\xrightarrow{d} N\left(0\,,\,2\,\sigma^4\int_0^1 K^2\,(v)\;dv\right).$$

The following corollary, which establishes the asymptotic normality of $\hat{\sigma}^2$, follows from the above theorem and the normality of the first term on the right hand side of equation (2.5).

Corollary **2.2.1.** *Under the assumptions A1, A2, B1, B2, B3 and B4 and* $h \sim n^{-\alpha}$*, where* α *is positive number such that* $\alpha < 1$

$$\sqrt{n}\,\left(\hat{\sigma}^2 - \sigma^2\right)\;\xrightarrow{d} N\left(0\,,\,\mu_4 - \sigma^4\right).$$

Remark:

1) Note that $\mu_4 - \sigma^4 = \mathrm{Var}(\epsilon^2)$ and so that the asymptotic distribution of the $\hat{\sigma^2}$ is exactly the same as the asymptotic distribution of the Hall and Marron estimator.

2) Note that if the assumption $B2'$ is used instead of the assumption $B2$, the results of Theorems 2.2.1 and 2.2.2 and Corollary 2.2.1 still hold true.

A sketch of the proof of Theorem 2.2.1 is given in section 2.4, whereas an outline of proofs of Theorem 2.2.2 and Corollary 2.2.1 are provided in section 2.5. In the following section, lemmas, that are used in the proofs of the above theorems and corollary, are stated.

2.3 Lemmas

All proofs of the following lemmas are given in Alharbi (2011).

Lemma **2.3.1.**

Suppose the assumptions A1,A2, B1, B2, B3 and B4 are satisfied. Let z *be a fixed number in the interval [0,1], then*

$$\frac{1}{n}\sum_{i=1}^{n} K^2\left(\frac{z-t_i}{h}\right)\ m(t_i) = \ h\ \int_0^1 K^2\,(y)\ \ m(z-h\,y)\ du + O(n^{-1}).$$

Lemma **2.3.2.**

Suppose the assumptions A1,A2, B1, B2, B3 and B4 are satisfied. Then,

$$\frac{1}{n^2}\sum\sum_{i\neq j} K\left(\frac{t_i-t_j}{h}\right)\ m(t_i)\ m(t_j) \quad = \quad h\ \int_0^1\int_0^1 K(y)m(t)m(t-hy)\ dt\,dy + O(n^{-1}\,h).$$

Lemma **2.3.3.**

Suppose the assumptions A1, A2, B1, B2, B3 and B4 hold. Then,

$$\frac{1}{n^3}\sum\sum_{i\neq j\neq k}\sum K^2\left(\frac{t_i-t_j}{h}\right) K^2\left(\frac{t_i-t_k}{h}\right)\ m(t_i)\ m(t_k) = h^2\ \int_0^1 m^2(x)\,(u)\ \ dx + O(h^2\,n^{-1}).$$

2.4 Proof of Theorem 2.2.1

First, to calculate the bias,

$$\begin{aligned} E(\hat{\sigma^2}) &= \frac{1}{n}\sum_{i=1}^{n} E(Y_i^2) - \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h}\right) E(Y_i)\,E(Y_j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{(m(x_i))^2+\sigma^2\} - \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h}\right) m(x_i)\,m(x_j). \end{aligned}$$

Therefore, we get

$$E(\hat{\sigma^2}) = \sigma^2 + \frac{1}{n}\sum_{i=1}^{n}(m(x_i))^2 - \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x_i - x_j}{h}\right) m(x_i)\, m(x_j). \quad (2.6)$$

Using the definition of Riemann integral, the second term is approximated by

$$\frac{1}{n}\sum_{i=1}^{n}(m(x_i))^2 = \int_0^1 m^2(t)\ dt + O(n^{-1}). \quad (2.7)$$

For the third term, since $m(x)$ is bounded and by applying Lemma 2.3.2, we obtain

$$\begin{aligned} &\frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x_i - x_j}{h}\right) m(x_i)m(x_j) \\ &= \int_0^1\int_0^1 K(y)m(t)m(t-hy)\ dt\, dy + O(n^{-1}). \end{aligned} \quad (2.8)$$

Using Taylor series expansion, the right hand side of equation (2.8) can be written as,

$$\begin{aligned} &\int_0^1\int_0^1 K(y)m(t)m(t-hy)\ dt\, dy + O(n^{-1}) \\ &= \int_0^1\int_0^1 K(y)m(t)\{m(t) - hy\, m'(t) + ... + \frac{h^r\,(-1)^r}{r!}\, y^r\ m^{(r)}(t) + o(h^r)\}\ dt\, dy \\ &+ O(n^{-1}) \\ &= \int_0^1 m^2(t)dt + \frac{h^r\,(-1)^r}{r!}\int_0^1 K(y)y^r\, dy \int_0^1 m(t)\, m^{(r)}(t)dt \\ &+ o(h^r) + O(n^{-1}) \end{aligned} \quad (2.9)$$

where $m^{(r)}(x)$ is the rth derivative of $m(x)$. By substituting (2.7) and (2.9) in equation (2.6), we obtain

$$\begin{aligned}
E(\hat{\sigma^2}) &= \sigma^2 + \frac{h^r\,(-1)^r}{r!} \int_0^1 K(y)y^r\,dy \int_0^1 m(t)m^{(r)}(t)dt + o(h^r) + O(n^{-1}) \\
&= \sigma^2 + h^r.C_1 + o(h^r) + O(n^{-1})
\end{aligned}$$

where $C_1 = \frac{(-1)^r}{r!} \int_0^1 K(y)y^r\,dy \int_0^1 m(t)\,m^{(r)}(t)dt$. This completes the proof of part (i) of Theorem 2.2.1.

In the view of the above theorem, the squared bias can be written as

$$\left(E(\hat{\sigma}^2) - \sigma^2\right)^2 = h^{2\,r}.C_1^2 + o(h^{2\,r}) + o(n^{-2}\,h^{-1}). \tag{2.10}$$

By computing $E(\hat{\sigma^2})^2$ and $(E(\hat{\sigma^2}))^2$, we get

$$\begin{aligned}
Var(\hat{\sigma^2}) &= E(\hat{\sigma^2})^2 - (E(\hat{\sigma^2}))^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left[\mu_4 - \sigma^4 + 4\mu_3 m(x_i) + 4\sigma^2 m^2(x_i)\right] \\
&- \frac{2}{n^2(n-1)h}\sum\sum_{i\neq j} K\left(\frac{x_i - x_j}{h}\right)\left[\mu_3 m(x_j) + \mu_3 m(x_i) + 4\sigma^2 m(x_i)m(x_j)\right] \\
&+ \frac{2}{n^2(n-1)^2h^2}\sum\sum_{i\neq j} K^2\left(\frac{x_i - x_j}{h}\right)\left[\sigma^4 + \sigma^2 m^2(x_i) + \sigma^2 m^2(x_j)\right] \\
&+ \frac{\sigma^2}{n^2(n-1)^2h^2}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x_i - x_j}{h}\right) K\left(\frac{x_i - x_k}{h}\right) m(x_j)m(x_k) \\
&+ \frac{\sigma^2}{n^2(n-1)^2h^2}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x_i - x_j}{h}\right) K\left(\frac{x_k - x_i}{h}\right) m(x_j)m(x_k) \\
&+ \frac{\sigma^2}{n^2(n-1)^2h^2}\sum\sum\sum_{i\neq k\neq d} K\left(\frac{x_i - x_k}{h}\right) K\left(\frac{x_k - x_d}{h}\right) m(x_i)m(x_d) \\
&+ \frac{\sigma^2}{n^2(n-1)^2h^2}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x_i - x_j}{h}\right) K\left(\frac{x_k - x_j}{h}\right) m(x_i)m(x_k) \qquad (2.11)
\end{aligned}$$

For details of computing $E(\hat{\sigma^2})^2$ and $(E(\hat{\sigma^2}))^2$, see Alharbi (2011). Now, the summations in the above equation will be approximated by the integration as follows. Using the definition of Riemann integral, the first term on the right hand side of equation (2.11) can be approximated as

$$\frac{1}{n^2}\sum_{i=1}^{n}\left[\mu_4 - \sigma^4 + 4\mu_3 m(x_i) + 4\sigma^2 m^2(x_i)\right]$$
$$= \frac{1}{n}\left[\mu_4 - \sigma^4 + 4\mu_3\int_0^1 m(u)du + 4\sigma^2\int_0^1 m^2(u)du\right] + o(n^{-2}\,h^{-1}). \qquad (2.12)$$

Since $m(x)$ is bounded, the lemmas in section 2.3 can be used to approximate the remaining terms on the right hand side of equation (2.11). By applying Lemma 2.3.2 and from the approximation of equation (2.8), the approximation of the second term on the right hand side of (2.11) may be written as follows,

$$\frac{-2}{n^2(n-1)h}\sum\sum_{i\neq j} K\left(\frac{x_i - x_j}{h}\right)\left[\mu_3 m(x_j) + \mu_3 m(x_i) + 4\sigma^2 m(x_i)m(x_j)\right]$$
$$= \frac{-2}{nh}\int_0^1\int_0^1 K\left(\frac{u-t}{h}\right)\left[\mu_3 m(u) + \mu_3 m(t) + 4\sigma^2 m(u)m(t)\right] du\,dt + O(n^{-2})$$
$$= \frac{-4}{n}\left[\mu_3\int_0^1 m(x)\,dx + 2\,\sigma^2\int_0^1 m^2(x)\,dx\right] + o(n^{-2}\,h^{-1}). \qquad (2.13)$$

For the third term on the right hand side of equation (2.11), again using Lemma 2.3.2, we obtain

$$\frac{2}{n^2\,h}\left[\sigma^4\int_0^1 K^2(y)dy + \sigma^2\int_0^1 K^2(y)dy\int_0^1 m^2(x)dx\right]$$
$$+\frac{2}{n^2\,h}\sigma^2\int_0^1 K^2(y)dy\int_0^1 m^2(x)dx + o(n^{-2}\,h^{-1}) + O(n^{-3}\,h^{-2}).$$

$$= \frac{2}{n^2\, h}\left[\sigma^4 \int_0^1 K^2(y)dy + 2\sigma^2 \int_0^1 K^2(y)dy \int_0^1 m^2(x)dx\right] + o\left(n^2\, h\right)^{-1}. \quad (2.14)$$

The 4^{th}, 5^{th}, 6^{th} and 7^{th} terms on the right hand side of equation (2.11) are approximated by using Lemma 2.3.3. The approximation of these terms are exactly the same. The approximation of each one of these terms is as follows

$$\frac{\sigma^2}{n} \int_0^1 m^2(x)dx + O(n^{-1}h^2) + O(n^{-2}).$$

The approximation of the last four terms on the right hand side of equation (2.11) is

$$\frac{4\sigma^2}{n} \int_0^1 m^2(x)dx + o(n^{-2}\, h^{-1}). \quad (2.15)$$

From the combination of equations (2.12)-(2.15), the variance of $\hat{\sigma^2}$ is

$$\begin{aligned} Var(\hat{\sigma^2}) &= \frac{1}{n}\left[\mu_4 - \sigma^4\right] \\ &+ \frac{2}{n^2\, h}\left[\sigma^4 \int_0^1 K^2(y)dy + 2\sigma^2 \int_0^1 K^2(y)dy \int_0^1 m^2(x)dx\right] \\ &+ o\left(n^2\, h\right)^{-1} \\ &= n^{-1}C_2 + n^{-2}\, h^{-1}C_3 + o(n^2\, h)^{-1} \qquad (2.16) \end{aligned}$$

where

$$\begin{aligned} C_2 &= \mu_4 - \sigma^4, \qquad \text{and} \\ C_3 &= 2\sigma^4 \int_0^1 K^2(y)dy + 4\sigma^2 \int_0^1 K^2(y)dy \int_0^1 m^2(x)dx. \qquad (2.17) \end{aligned}$$

Thus, the proof of Theorem 2.2.1 is completed. For more details, see Alharbi (2011).

2.5 Proofs of Theorem 2.2.2 and Corollary 2.2.1

Our main goal here is to find the asymptotic distribution of the estimator in (2.3). To prove Corollary 2.2.1, in subsection 2.5.1, we establish the asymptotic normality of the first term on the right hand side of equation (2.5) using the Lindeberg-Feller central limit theorem. An outline of the proof of Theorem 2.2.2 is presented in subsection 2.5.2, whereas the proof of Corollary 2.2.1 is given in subsection 2.5.3.

2.5.1 The Asymptotic Distribution of the First Term in Equation (2.5)

The aim in this subsection is to derive the asymptotic distribution of the first term on the right hand side of equation (2.5). First, observe that

$$\begin{aligned}\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i &= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - b_i\,[Y_i - m(x_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} W_i - \frac{1}{n}\sum_{i=1}^{n} V_i \end{aligned} \tag{2.18}$$

where $W_i = Y_i^2$, $\bar{W} = \frac{1}{n}\sum\limits_{i=1}^{n} W_i$, $V_i = b_i\,[Y_i - m(x_i)]$ and $b_i = \frac{2}{(n-1)\,h}\sum\limits_{j\neq i} K\left(\frac{x_i - x_j}{h}\right) m(x_j)\,.$ Note that W_is are independent random variables as are V_is. From the definition of W_i, it is not difficult to prove that

$$\begin{aligned} E\left(\bar{W}\right) &= \sigma^2 + \frac{1}{n}\sum_{i=1}^{n} m^2(x_i) \\ &= \sigma^2 + \int_0^1 m^2(u)\,du + O(n^{-1}) \\ &\longrightarrow \sigma^2 + \int_0^1 m^2(u)\,du, \quad \text{as } n \to \infty. \end{aligned} \tag{2.19}$$

It is clear from the definition of V_i that $E(V_i) = 0$. So, we obtain

$$E\left(\bar{Z}\right) = E\left(\bar{W}\right) = \sigma^2 + \int_0^1 m^2(u)\,du + O(n^{-1}). \tag{2.20}$$

To find the variance of the $\bar{Z}$, first note that

$$\text{Var}\,(W_i) = \mu_4 - \sigma^4 + 4\,\mu_3\, m(x_i) + 4\,\sigma^2\, m^2(x_i), \tag{2.21}$$

$$\text{Var}\,(V_i) = b_i^2\,\sigma^2 = 4\,\sigma^2\, m^2(x_i) + O(n^{-1}\,h^{-1}) \tag{2.22}$$

and

$$\text{Cov}\,(W_i\,,\,V_i) = 2\,\mu_3\, m(x_i) + 4\,\sigma^2\, m^2(x_i) + O(n^{-1}) + O(h^2). \tag{2.23}$$

Then by computing $E\left(\frac{1}{n}\sum\limits_{i=1}^{n} W_i\right)^2$ and $\left[E\left(\frac{1}{n}\sum\limits_{i=1}^{n} W_i\right)\right]^2$, we obtain

$$\text{Var}\left(\bar{W}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left[\mu_4 + 4\,\mu_3\, m(x_i) + 4\,\sigma^2\, m^2(x_i)\right] - \frac{\sigma^4}{n}.$$

Using the definition of Riemann integral, we get

$$\text{Var}\left(\bar{W}\right) = \frac{1}{n}\left[\mu_4 - \sigma^4 + 4\mu_3\int_0^1 m(u)\,du + 4\sigma^2\int_0^1 m^2(u)\,du\right] + o(n^{-1}). \tag{2.24}$$

Using a similar calculation, we can show that

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} V_i\right) = \frac{\sigma^2}{n^2}\sum_{i=1}^{n} b_i^2 \tag{2.25}$$

Thus, the right hand side of equation in (2.25) can be written as

$$
\begin{aligned}
&= \frac{4\,\sigma^2}{n^2\,(n-1)^2\,h^2} \sum\sum_{i \neq j} K^2\left(\frac{x_i - x_j}{h}\right)\, m^2(x_j) \\
&+ \frac{4\,\sigma^2}{n^2\,(n-1)^2\,h^2} \sum\sum\sum_{i \neq j \neq k} K\left(\frac{x_i - x_j}{h}\right)\, K\left(\frac{x_i - x_k}{h}\right)\, m(x_j)\, m(x_k).
\end{aligned}
$$

Using Lemmas 2.3.2 and 2.3.3, respectively, the approximation of the last two terms on the right hand side of the above equation is

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} V_i\right) &= \frac{4\,\sigma^2}{n\,(n-1)\,h} \int_0^1 K^2(y)\;dy \int_0^1 m^2(x)\;dx \,+\, o(n^{-2}\,h^{-1}) \\
&+ \frac{4\,\sigma^2}{n} \int_0^1 m^2(x)\;dx \,+\, o(n^{-1}) \\
&= \frac{4\,\sigma^2}{n} \int_0^1 m^2(x)\;dx \,+\, o(n^{-1}). \qquad (2.26)
\end{aligned}
$$

For the covariance between $\left(\frac{1}{n}\sum\limits_{i=1}^{n} W_i\right)$ and $\left(\frac{1}{n}\sum\limits_{i=1}^{n} V_i\right)$, one can show that

$$
\begin{aligned}
\mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\,,\, \frac{1}{n}\sum_{i=1}^{n} V_i\right) &= \frac{1}{n^2}\sum_{i=1}^{n} b_i\,\left[\,\mu_3 \,+\, 2\,\sigma^2\, m(x_i)\,\right] \\
&= \frac{2}{n^2\,(n-1)\,h} \sum\sum_{i \neq j} K\left(\frac{x_i - x_j}{h}\right) \\
&\times\; \left[\,\mu_3\, m(x_j) \,+\, 2\,\sigma^2\, m(x_i)\, m(x_j)\,\right].
\end{aligned}
$$

Since $m(x)$ is bounded and using Lemma 2.3.2, we obtain

$$
\mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} W_i,\, \frac{1}{n}\sum_{i=1}^{n} V_i\right) = \frac{2\,\mu_3}{n} \int_0^1 m(x)\,dx + \frac{4\,\sigma^2}{n} \int_0^1 m^2(x)\,dx + o(n^{-1}). \qquad (2.27)
$$

By using equations (2.24), (2.26) and (2.27), we can prove that

$$\begin{aligned}\text{Var}\left(\bar{Z}\right) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\right) + \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} V_i\right) - 2\,\text{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\,,\,\frac{1}{n}\sum_{i=1}^{n} V_i\right)\\ &= \frac{1}{n}\left[\mu_4 - \sigma^4\right] + o(n^{-1}). \qquad (2.28)\end{aligned}$$

For the details of the calculation of the variance of $\bar{Z}$, see Alharbi (2011).

To use the Lindeberg-Feller central limit theorem, it is required to show that the following condition holds,

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{i=1}^{n} E\left[(Z_i - E[Z_i])^2\, I\left[\,|\,Z_i - E[Z_i]\,| > \tau\, B_n\right]\right] = 0 \qquad (2.29)$$

where $B_n^2 = \sum_{i=1}^{n}\sigma_{Z_i}^2$, $\sigma_{Z_i}^2$ is the variance of Z_i and τ is a positive number. To verify the above condition, observe that

$$E\left[\,Z_i\,\right] = \sigma^2 + m^2(x_i), \qquad (2.30)$$

Using equations (2.21), (2.22) and (2.23), we can show that

$$\begin{aligned}\text{Var}\left(Z_i\right) &= \sigma_{Z_i}^2 = \mu_4 - \sigma^4 + O(n^{-1}\,h^{-1}) + O(h^2)\\ &\to \mu_4 - \sigma^4 \qquad \text{as} \quad h\to 0 \text{ and } n\,h \to \infty. \qquad (2.31)\end{aligned}$$

Clearly from the above equation, the variance of Z_i does not depend on i . So, we get

$$\begin{aligned}B_n^2 &= \sum_{i=1}^{n}\sigma_{Z_i}^2 = n\left(\mu_4 - \sigma^4\right) + O(h^{-1}) + O(n\,h^2)\\ &\to n\left(\mu_4 - \sigma^4\right) \qquad \text{as} \quad h\to 0 \text{ and } n \to \infty. \qquad (2.32)\end{aligned}$$

Also note that

$$\begin{aligned}
Z_i - E[Z_i] &= Y_i^2 - b_i\,(Y_i - m(x_i)) - \sigma^2 - m^2(x_i) \\
&= \left(Y_i - \frac{b_i}{2}\right)^2 - \frac{b_i^2}{4} + b_i\,m(x_i) - \sigma^2 - m^2(x_i) \\
&= \left(Y_i - m(x_i) + m(x_i) - \frac{b_i}{2}\right)^2 - \frac{b_i^2}{4} + b_i\,m(x_i) - \sigma^2 - m^2(x_i) \\
&= (Y_i - m(x_i))^2 - 2\,(Y_i - m(x_i))\left(m(x_i) - \frac{b_i}{2}\right) + \left(m(x_i) - \frac{b_i}{2}\right)^2 \\
&- \frac{b_i^2}{4} + b_i\,m(x_i) - \sigma^2 - m^2(x_i). \qquad (2.33)
\end{aligned}$$

In the last equability on the right hand side of (2.33), the first term is independent and identically distributed random variables. This term is clearly bounded, because its expected value is σ^2 and its variance is $\mu_4 - \sigma^4$. The second term on the right hand side of (2.33) converges to zero as $n \to \infty$, because $\left(m(x_i) - \frac{b_i}{2}\right) \to 0$ as $n \to \infty$. The remaining terms on the right hand side of (2.33) are constants and depend on the mean function. Since the mean function is bounded, these terms are also bounded. Using equation (2.32), it is obvious that

$$\tau\,B_n \to \infty, \qquad \text{as } n \to \infty.$$

Thus, $I\,[\,|\,Z - E[Z_i]\,\,| > \tau\,B_n\,]$ will be always zero as $n \to \infty$. This implies that

$$\frac{1}{B_n^2}\sum_{i=1}^{n} E\,\left[(\,Z_i - E[Z_i]\,)^2\,I\,[\,|\,Z - \mu_Z\,\,| > \tau\,B_n\,]\right] \longrightarrow 0\,.$$

Therefore, the condition (2.29) is satisfied. Therefore, using Lindeberg-Feller central limit theorem and equations (2.20) and (2.28), we obtain

$$\frac{\sqrt{n}\,\left(\bar{Z} - C_1\right)}{\sqrt{C_2}} \xrightarrow{d} N(0,1)$$

where $\xrightarrow{d}$ means convergence in distribution,

$$\begin{aligned} C_1 &= \sigma^2 + \int_0^1 m^2(u)\,du \quad \text{and} \\ C_2 &= \mu_4 - \sigma^4. \end{aligned}$$

That is,

$$\sqrt{n}\,\left(\bar{Z} - C_1\right) \xrightarrow{d} N\left(0, C_2\right). \tag{2.34}$$

2.5.2 Proof of Theorem 2.2.2

The aim of this subsection is to prove Theorem 2.2.2. Recall that

$$\begin{aligned} T_n(h) &= \frac{1}{n\,(n-1)\,h} \sum\sum_{i \neq j} K\left(\frac{x_i - x_j}{h}\right) U_i U_j \\ &= \sum\sum_{1 \leq i < j \leq n} T_{ij} \end{aligned}$$

where $U_i = Y_i - m(x_i)$ and $T_{ij} = \frac{2}{n\,(n-1)\,h}\, K\left(\frac{x_i - x_j}{h}\right) U_i\, U_j$.

To prove the asymptotic normality of $T_n(h)$, we will use theorem 2.1 developed by De Jong (1987). Below is the statement and its explanation of a statistics being ‘clean’.

Let $X_1, X_2, \ldots X_n$ be independent random variables and $W(n) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(X_i, X_j)$ where $W_{ij}(X_i, X_j)$ is a Borel function such that var $[W_{ij}(X_i, X_j)] = \sigma_{ij}^2$ is finite and

$$E(\, W_{ij}(X_i, X_j) | X_i) = E(\, W_{ij}(X_i, X_j) | X_j) = 0.$$

Definition 2.1 $W(n)$ is called ‘*clean*’, if the conditional expectation of W_{ij} vanish:

$$E\,[\, W_{ij} |\, X_i\,] = 0 \;\; \text{a.s. for all } i,\, j \leq n.$$

Theorem 2.1 *Let* $W(n)$ *be 'clean' with variance* $\sigma(n)^2$ *. Assume*

a) $\sigma(n)^{-2} \max\limits_{1 \le i \le n} \sum\limits_{1 \le j \le n} \sigma_{ij}^2 \to 0, \quad as\, n \to \infty,$

b) $\sigma(n)^{-4} E\,[\,W(n)^4\,] \to 3, \;\; as\, n \to \infty.$

Then

$$\sigma(n)^{-1}\, W(n) \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty.$$

For $1 \le i < j \le n$ and using the above definition, we obtain

$$\begin{aligned} E\,[\,T_{ij}|\,U_i\,] &= E\left[\frac{2}{n\,(n-1)\,h}\,K\left(\frac{x_i - x_j}{h}\right)\,U_i\,U_j\,|\,U_i\right] \\ &= \frac{2}{n\,(n-1)\,h}\,K\left(\frac{x_i - x_j}{h}\right)\,(\,Y_i - m(x_i)\,)\,E\,[\,Y_j - m(x_j)] \\ &= 0. \end{aligned}$$

Using similar arguments, it can be shown that $E\,[\,T_{ij}|\,U_j\,] = 0$. Hence, $T_n(h)$ is 'clean'. Therefore, the proof of Theorem 2.2.2 will be completed if the following conditions are satisfied

(D1) $\left(\sigma_{T_n}^2\,\right)^{-2} E\,[\,T_n^4(h)\,] \to 3, \;\; \text{as}\, n \to \infty,$

(D2) $\left(\sigma_{T_n}^2)\right)^{-1} \max\limits_{1 \le i \le n} \sum\limits_{1 \le j \le n} \sigma_{ij}^2 \to 0, \quad \text{as}\, n \to \infty,$

where σ_{ij}^2 is the variance of T_{ij}. It should be noted that the diagonal elements are not included in the definition of $T_n(h)$, but we can write

$$\begin{aligned} T_n(h) &= \frac{1}{n\,(n-1)\,h}\sum\sum_{i \ne j} K\left(\frac{x_i - x_j}{h}\right)\,U_iU_j \\ &= \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x_i - x_j}{h}\right)\,U_iU_j \\ &- \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n} K\,(0)\;U_i^2 \end{aligned} \tag{2.35}$$

Note that $K(0)$ is a constant and $\frac{1}{n}\sum_{i=1}^{n} U_i^2 \xrightarrow{p} \sigma^2$. Thus, $\frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n} U_i^2 \xrightarrow{p} 0$. Therefore, the right hand side of equation (2.35) can be written as

$$T_n(h) \quad = \quad \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x_i - x_j}{h}\right) U_i U_j \; + \; o_p\,(\,1\,).$$

So, the effect of not using the diagonal elements is of smaller order and Theorem 2.1 in De Jong (1986) can be used to show normality of $T_n(h)$. Also note that

$$\begin{aligned} \sigma_{ij}^2 \quad &= \quad E[\,T_{ij}\,]^2 \; - \; (\,E[\,T_{ij}\,]\,)^2 \\ &= \quad \frac{4}{n^2\,(n-1)^2\,h^2}\,K^2\left(\frac{x_i - x_j}{h}\right)\,E\,[\,U_i^2\,]\,E\,[\,U_j^2\,] \\ &= \quad \frac{4\,\sigma^4}{n^2\,(n-1)^2\,h^2}\,K^2\left(\frac{x_i - x_j}{h}\right). \end{aligned} \tag{2.36}$$

Now, to verify the conditions D1 and D2, first, second and fourth moments of $T_n(h)$ are needed. In Alharbi (2011), it is shown that

$$\begin{aligned} E\,(T_n(h)) \; &= E\left(\frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right)\,U_i\,U_j\right) \\ &= \; \frac{1}{n\,(n-1)\,h}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right) E[\,U_i]E[U_j] \; = \; 0, \end{aligned} \tag{2.37}$$

$$E\left[\,T_n^2(h)\,\right] \; = \; \frac{2\,\sigma^4}{n^2\,h}\int_0^1 K^2\,(v)\;dv \; + \; o(n^{-2}\,h^{-1}), \tag{2.38}$$

and

$$E\left[\,T_n^4(h)\,\right] \; = \; \frac{12\,\sigma^8}{n^4\,h^2}\left(\int_0^1 K^2\,(v)\;dv\right)^2 \; + \; o(n^{-4}\,h^{-2}). \tag{2.39}$$

Thus, we get

$$
\begin{aligned}
\text{Var}\ (\,T_n(h)\,) &= \sigma^2_{T_n} = E\left[T_n^2(h)\right] - \left[E(\,T_n(h)\,)\right]^2 \\
&= \frac{2\,\sigma^4}{n^2\,h}\int\limits_0^1 K^2\,(v)\ dv + o(n^{-2}\,h^{-1}). \qquad (2.40)
\end{aligned}
$$

For detailed calculation, see Alharbi (2011).

To check the condition (D1), using (2.40) and (2.39) and by the assumption B4, observe that

$$
\begin{aligned}
\left(\sigma^2_{T_n}\right)^{-2}\,E\left[T_n^4(h)\right] &= \left[\frac{2\,\sigma^4}{n^2\,h}\int\limits_0^1 K^2\,(v)\ dv + o(n^{-2}\,h^{-1})\right]^{-2} \\
&\times \left[\frac{12\,\sigma^8}{n^4\,h^2}\left(\int\limits_0^1 K^2\,(v)\ dv\right)^2 + o(n^{-4}\,h^{-2})\right] \\
&= 3 + o(1) \\
&\to 3. \qquad (2.41)
\end{aligned}
$$

To verify condition (D2), first note that using equation (2.36) and then Lemma 2.3.1, we obtain

$$
\begin{aligned}
\max_{1\le i\le n}\sum_{1\le j\le n}\sigma^2_{ij} &= \frac{4\,\sigma^4}{n^2\,(n-1)^2\,h^2}\max_{1\le i\le n}\sum_{1\le j\le n}K^2\left(\frac{x_i - x_j}{h}\right) \\
&= \frac{4\,\sigma^4}{n^3\,h^2}\max_{1\le i\le n}\left[\int\limits_{v=0}^1 K^2\left(\frac{x_i - v}{h}\right)\,dv + O(\,n^{-1})\right].
\end{aligned}
$$

By substituting $u = \frac{x_i - v}{h}$, the last equation becomes

$$\begin{aligned}
\max_{1\le i\le n} \sum_{1\le j\le n} \sigma_{ij}^2 &= \frac{4\,\sigma^4}{n^3\,h^2} \max_{1\le i\le n} \left[h \int\limits_{u=\max\{\frac{x_i-1}{h},-1\}}^{\min\{\frac{x_i}{h},1\}} K^2(u)\,du + O(n^{-1}) \right] \\
&\le \frac{4\,\sigma^4}{n^3\,h^2} \left[h \int_{-1}^{1} K^2(u)\,du + O(n^{-1}) \right] \\
&= O(n^{-3}\,h^{-1}). \qquad (2.42)
\end{aligned}$$

And finally, using equations (2.40) and (2.42), we get

$$\left(\sigma_{T_n}^2\right)^{-1} \max_{1\le i\le n} \sum_{1\le j\le n} \sigma_{ij}^2 \;\le\; \frac{O(n^{-3}\,h^{-1})}{O(n^{-2}\,h^{-1})} = O(n^{-1}) \to 0, \quad \text{as } n \to \infty. \quad (2.43)$$

Therefore by (2.41), (2.43) and using theorem 2.1 in De Jong (1987), it is easy to show that

$$\left(\sigma_{T_n}^2\right)^{\frac{-1}{2}} \; T_n(h) \;\xrightarrow{d}\; N\,(0\,,\,1)\,. \qquad (2.44)$$

That is,

$$n\,\sqrt{h}\;T_n(h) \;\xrightarrow{d}\; N\left(0\,,\,2\,\sigma^4 \int\limits_0^1 K^2(v)\,dv\right).$$

2.5.3 Proof of Corollary 2.2.1

First note that, using Lemma 2.3.2, the approximation of the third term on the right hand side of equation (2.5) is

$$\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j\ne i} K\left(\frac{x_i - x_j}{h}\right) m(x_i)\,m(x_j) = \int\limits_0^1 m^2(x)\,dx + O(h^r) + O(n^{-1}). \quad (2.45)$$

Equation (2.37) implies that $E(\bar{Z}) \cdot E(T_n(h)) = 0$. Again using equation (2.37) and the independence of Y_i, Y_j and Y_k and U_i, U_j and U_k for $i \ne j \ne k$, we can show that

$E(\ \bar{Z}\ \cdot\ T_n(h)\ )\ = 0$. This leads to

$$\text{Cov}\,(\ \bar{Z},\, T_n(h)) =\ E(\ \bar{Z}\ \cdot\ T_n(h)\ )\ -\ E(\ \bar{Z}\ )\cdot\ E(\,T_n(h))\ =\ 0$$

By using (2.34) and (2.44), respectively, $\bar{Z}$ and $T_n(h)$ may be written as

$$\bar{Z}\ =\ \frac{1}{\sqrt{n}}(\mu_4\ -\ \sigma^4)^{1/2}\,N_1\ +\ \sigma^2\ +\ \int_0^1 m^2(x)\,dx \tag{2.46}$$

and

$$T_n(h)\ =\ \frac{1}{n\,\sqrt{h}}\ \left[2\,\sigma^4 \int_0^1 K^2\,(v)\ dv\right]^{1/2} N_2 \tag{2.47}$$

where the random variables N_1 and N_2 are standard normal distribution. Then, by (2.45) and the above representation of $\bar{Z}$ and $T_n(h)$, we may express $\hat{\sigma}^2$ as

$$\begin{aligned}\hat{\sigma}^2\ \ &=\ \ \frac{1}{\sqrt{n}}(\mu_4\ -\ \sigma^4)^{1/2}\,N_1\ +\ \sigma^2\ +\ \int_0^1 m^2(x)\,dx\ +\ \frac{1}{n\,\sqrt{h}}\ \left[2\,\sigma^4 \int_0^1 K^2\,(v)\ dv\right]^{1/2} N_2\\ &-\ \ \int_0^1 m^2(x)\,dx\ +\ O(n^{-1})\ +\ O(\,h^r\,).\end{aligned}$$

So, we obtain

$$\left(\hat{\sigma}^2\ -\ \sigma^2\,\right) = \frac{1}{\sqrt{n}}(\mu_4 - \sigma^4)^{1/2} N_1 + \frac{1}{n\,\sqrt{h}}\ \left[2\,\sigma^4 \int_0^1 K^2\,(v)\ dv\right]^{1/2} N_2 + O(n^{-1}) + O(h^r\,).$$

Therefore, it is clear that

$$\begin{aligned}\sqrt{n}\ \left(\hat{\sigma}^2\ -\sigma^2\right)\ \ &=\ \ \left(\mu_4\ -\ \sigma^4\,\right)^{1/2}\,N_1\ +\ \left[(\,n\,h\,)^{-1}\,2\,\sigma^4 \int_0^1 K^2\,(v)\ dv\right]^{1/2} N_2\\ &+\ \ O(n^{-1/2}\,)\ +\ O(n^{1/2}\,h^r\,).\end{aligned}$$

This implies that

$$\sqrt{n}\,\left(\hat{\sigma}^2\ -\sigma^2\right)\ \xrightarrow{d} N\left(0\,,\,\mu_4\ -\ \sigma^4\right). \tag{2.48}$$

2.6 The Optimal Bandwidth Selection

It is obvious from Theorem 2.2.1 that the bias and variance of $\hat{\sigma}^2$ depend on the bandwidth h. If h is large, the bias is also large. However, if h is small, the variance increases in the second order. Thus, selection of an optimal bandwidth is vital for the efficient performance of the estimator in (2.3). In this section, we investigate analytically the optimal bandwidth of the estimator in (2.3).

From the previous section, the asymptotic mean squared error of the estimator in (2.3) under the assumptions A1, A2, B1, B2, B3 and B4 is

$$AMSE(\hat{\sigma}^2)\ \ \approx\ \ n^{-1}\,C_2 + n^{-2}\,h^{-1}\,C_3 + h^{2\,r}\,C_4$$

where C_1, C_2 and C_3 are as in Theorem 2.2.1 and $C_4\ =\ C_1^2$. To compute the asymptotic optimal bandwidth for the estimator in (2.3), it is necessary to solve the following equation

$$\frac{\partial\,(AMSE)}{\partial h} = 0.$$

Therefore, we obtain

$$\frac{\partial\,(AMSE)}{\partial h} = -n^{-2}h^{-2}\,C_3 +\ 2r\,h^{2\,r-1}\,C_4\ = 0. \tag{2.49}$$

From equation (2.49), it is obvious that

$$h_{\,opt} = \left(\frac{C_3}{2r\,C_4}\right)^{\frac{1}{2r+1}}\ \cdot\ n^{-2/2r+1}\,.$$

Thus, for the case of $r = 2$, $h_{opt} \sim n^{-2/5}$. It should be noted that the (MSE) optimal bandwidth for estimating the mean function is $n^{-1/5}$. However, here, $h \sim n^{-2/5}$. This means the estimator in (2.3) uses the estimate of the mean function, which has a smaller bias compared to the mean function estimator used in Hall and Marron estimator. This property is often desirable according to Wang, Brown, Cia and Levine (2008). The mean squared error corresponding to the asymptotic optimal bandwidth is

$$\begin{aligned} AMSE_{h_{opt}}(\hat{\sigma}^2) &= n^{-1} \cdot \left(\mu_4 - \sigma^4\right) + \left(\frac{2\,r}{(r!)^2}\right)^{\frac{1}{2r+1}} n^{(-4\,r/2r+1)}\, C_3^{(2\,r/2r+1)}\, C_5^{(1/2r+1)} \\ &+ \left(\frac{(r!)^2}{2\,r}\right)^{\frac{2\,r}{2r+1}} \cdot \frac{1}{(r!)^2}\, n^{(-4\,r/2r+1)}\, C_3^{(2\,r/2r+1)}\, C_5^{(1/2r+1)} \\ &= n^{-1}\left[\mu_4 - \sigma^4 + n^{(-2\,r+1)/(2r+1)} \cdot C_6\right] \end{aligned} \tag{2.50}$$

where

$$C_6 = \left[\left(\frac{2\,r}{(r!)^2}\right)^{\frac{1}{2r+1}} + \left(\frac{(r!)^2}{2\,r}\right)^{\frac{1}{2r+1}}\right] C_3^{(2\,r/2r+1)}\, C_5^{(1/2r+1)}.$$

For constants in C_6, it is noted that C_3 is as in Theorem 2.2.1 and

$$C_5 = \left(\int_0^1 K(y) y^r\, dy\right)^2 \left(\int_0^1 m(t)\, m^{(r)}(t) dt\right)^2.$$

One of the most important cases is for $r = 2$. The mean squared error in this case is

$$AMSE_{h_{opt}}(\hat{\sigma}^2) = n^{-1}\left\{\mu_4 - \sigma^4 + 1.25\, n^{-3/5}\, C_3^{4/5}\, C_5^{1/5}\right\},$$

where $C_5 = \left(\int_0^1 K(y) y^2\, dy\right)^2 \left(\int_0^1 m(t)\, m''(t) dt\right)^2$. In addition, we can note that

$$\begin{aligned} \text{Var}(\epsilon^2) &= E\left(\epsilon^4\right) - \left(E\left(\epsilon^2\right)\right)^2 \\ &= E\left(Y_i - m(x_i)\right)^4 - \left(E\left(Y_i - m(x_i)\right)^2\right)^2 = \mu_4 - \sigma^4. \end{aligned} \tag{2.51}$$

Hence, we obtain

$$
\begin{aligned}
AMSE_{h_{opt}}(\hat{\sigma}^2) &= n^{-1}\left\{\text{Var}(\epsilon^2) + 1.25\, n^{-3/5}\, C_3^{4/5}\, C_5^{1/5}\right\} \\
&= n^{-1}\ \text{var}(\epsilon^2) + o(n^{-1}).
\end{aligned}
$$

The relative error of the estimator in (2.3) is $n^{-3/5}$.

For the Hall and Marron estimator (1990) and when $r = 2$, one can show that the asymptotic optimal bandwidth is

$$
h_{HM} = \left(\frac{C_7}{C_8}\right)^{\frac{1}{9}} \cdot n^{-2/9} \tag{2.52}
$$

where

$$
C_7 = \sigma^4 \int_0^1 \left(K * K(u) - 2K(u)\right)^2\ du \quad \text{and} \tag{2.53}
$$

$$
C_8 = (2)^{-2}\left(\int_0^1 K(y)y^2\, dy\right)^4 \left(\int_0^1 \left(m''(t)\right)^2 dt\right)^2. \tag{2.54}
$$

Note that * denotes a convolution. Clearly, $n^{-2/9}$ compared with $n^{-2/5}$ is closer to $n^{-1/5}$, meaning that the bandwidth h required for the Hall and Marron estimator is close to the MSE-optimal bandwidth that one needs to estimate the mean function. So, the Hall and Marron estimator focuses on the estimation of the mean function so that the estimate of the mean is close to being MSE-optimal.

Before we comment on the mean square error of Hall and Marron estimator, we note that the asymptotic optimal bandwidth of the estimator in (2.3) is very close to being square of the asymptotic optimal bandwidth of the Hall and Marron estimator.

The asymptotic mean squared error of Hall and Marron estimator when $r = 2$ is

$$\begin{aligned} AMSE_{HM,\,h_{opt}} &= n^{-1}\left\{\text{var}(\epsilon^2) + (8)^{-7/9}\, n^{-7/9}\, C_9^{8/9}\, C_{10}^{2/9}\right\} \\ &= n^{-1}\,\text{var}(\epsilon^2) + o(\,n^{-1}) \end{aligned}$$

where $C_9 = 2\,C_7$ and $C_{10} = \frac{1}{4}\,C_8$. The relative error of the Hall and Marron estimator is $n^{-7/9}$. Therefore, in the first order, the optimal mean squared error rates for both of the estimators are the same. In the second order, it is obvious that the Hall and Marron estimator has smaller relative error than the estimator in (2.3). However, the smallest relative error does not lead to a better performance in the finite sample behaviour as Dette, Munk and Wanger (1998) and Tong, Liu and Wang (2008) have noted. To elaborate this further, consider the ratio of the MSEs of the new and Hall and Marron estimators by excluding the constants,

$$\frac{\text{MSE(Hall-Marron Estimator)}}{\text{MSE (New Estimator)}} = \frac{n^{-1}[1 + (n^{-7/9})]}{n^{-1}[1 + (n^{-3/5})]} = \frac{1 + (n^{-7/9})}{1 + (n^{-3/5})}.$$

The above ratio is approximately 0.97 for $n = 100$. This means that the difference in the second order of the mean squared error of the estimators has very little effect on the performance of these estimators. In fact, it means, in finite sample, the performance of the estimators considered is likely to be determined and dominated by the constants C_is involved in the MSE expressions. This aspect is illustrated in detail in chapter 3.

Although, in the difference-based estimators, one is not required to estimate the mean function explicitly, it estimates the mean implicitly with the smallest possible bias by taking bandwidth $h \sim n^{-1}$. To see this, observe that the first-order difference-based estimator is based on $(Y_i - Y_{i-1})^2$. So,

$$E\,(Y_i - Y_{i-1})^2 = 2\,\sigma^2 + [m(x_i) - m(x_{i-1})]^2 .$$

To have an estimate for σ^2 with smaller bias, $m(x_i) - m(x_{i-1})$ is required to be close to zero. Note that $(x_i - x_{i-1}) = n^{-1}$ and that $m(x_{i-1}) = m(x_i) + (x_i - x_{i-1})m'(x_i) + \cdots$ using the Taylor series expansion. This leads to the bin size of $(x_i - x_{i-1}) = n^{-1}$. That is, in general, the difference-based estimators use a smaller bandwidth than that of the new and the Hall and Marron estimators. The asymptotic optimal bandwidth of the new estimator in (2.3) is generally in the middle of the asymptotic optimal bandwidths of the difference-based estimators and the residuals-based estimators.

Chapter 3

Simulation Study: Finite Sample Behaviour

3.1 Introduction

In chapter 2, we studied the asymptotic properties of the new estimator in the setting of the homoscedastic nonparametric regression model. In this chapter, the main aim is to investigate the finite sample performance of the new estimator defined in (1.28) through simulation. In doing so, we will also verify that the asymptotic distribution of the new estimator is normal as proved in the last chapter. To exhibit the finite sample performance of the new estimator, we select a mean function and a bandwidth and then study the effect of these choices on the shape of the distribution of the new estimator. We repeat this for several different mean functions, each with different noise levels.

The general structure of the simulation studies is described in section 3.2. In section 3.3, in order to assess the finite sample performance of the new estimator, we consider different mean functions, each having different noise levels where the bandwidth is chosen appropriately. In particular, we choose six different mean functions and four levels of error variances.

From the mean square analysis in the second chapter, it is obvious that the performance of the new estimator depends on the bandwidth selection. As noted before, large bandwidth leads to a large bias and small bandwidth increases the variance in the second order. The effect of bandwidth selection through simulation is presented in section 3.4. To investigate this effect, the mean function and error variance are fixed, and bandwidth is allowed to vary. After that, the relation between bandwidth and the mean squared error is discussed in section 3.5, by plotting the logarithms of selected bandwidths against the logarithms of their mean squared errors.

3.2 The General Structure of the Simulation Studies

The model of the simulation studies in this chapter is $Y_i = m(X_i) + \epsilon_i$, where

C1. the design points $X_1, X_2, \cdots, X_n$ are independent and identically distributed uniform $[0,1]$ random variables,

C2. X_is are independent of ϵ_is and

C3. the errors $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independent and identically distributed random variables from Normal $N(0, \sigma^2)$.

For a given sample of size n, we first select randomly X_is and ϵ_is. Then Y_is are generated using the model $Y_i = m(X_i) + \epsilon_i$. The observed values of the new estimator $\hat{\sigma}^2_{NEW}$ are calculated using equation (1.28). The kernel function K is selected to be the standard normal probability density function. This procedure is replicated N times. Thus, for each chosen σ^2 and bandwidth h, there are N observed values of $\hat{\sigma}^2_{NEW}$. From these observed values, the kernel density estimator, $\frac{1}{N\,h_o}\sum_{i=1}^{N} W\left(\frac{u-u_i}{h_o}\right)$, is obtained where u_i is ith observed values of $\hat{\sigma}^2_{NEW}$. The kernel function used to smooth the observed values of $\hat{\sigma}^2_{NEW}$ is the standard normal density function with bandwidth $h_o = h_{opt,N} = 1.06\, S_N\, N^{\frac{-1}{5}}$

where S_N denotes the standard deviation of N observed values of $\hat{\sigma}^2_{NEW}$. For more details about the optimal bandwidth choice for the density function, see Silverman (1984), and Fan and Gijbels (1996). As noted in chapter 2, the new and the Hall and Marron estimators have the same asymptotic distribution, which was described in Corollary 2.2.1. For comparison, the asymptotic distributions are plotted in all considered cases in these simulation studies.

Remark:

In this simulation study, the uniform design has been used instead of equally spaced design. Note that since the distance from $X_{(i)}$ to $X_{(i+1)}$ remains approximately the same and roughly equals $\frac{1}{n}$, the simulation study results are valid for the equally spaced design.

3.3 The Effect of the Mean Function on the Finite Sample Performance of The New Estimator

To study the finite sample performance of the estimator in (1.28), six different mean functions are considered,

$$
\begin{aligned}
i)\ m_1(x) &= 1. \\
ii)\ m_2(x) &= 4.7 + 2.4\,x + 5\,x^2 + 4.3\,x^3. \\
iii)\ m_3(x) &= \left(3 + x + 4\,x^2 + 8\,x^4\right) \cdot I(x \leq 0.5) \\
&+ \left(5.875 - x - x^2 - x^3\right) \cdot I(x > 0.5). \\
iv)\ m_4(x) &= \exp\left(-2 - 4\,x - 5x^2 - 6x^3\right). \\
vi)\ m_5(x) &= \frac{4}{5}\sin\left(2\,\pi\,x\right). \\
vii)\ m_6(x) &= \frac{3}{4}\cos\left(10\,\pi\,x\right).
\end{aligned}
$$

The above mean functions are plotted in the figures (3.1) and (3.2). In particular, we select

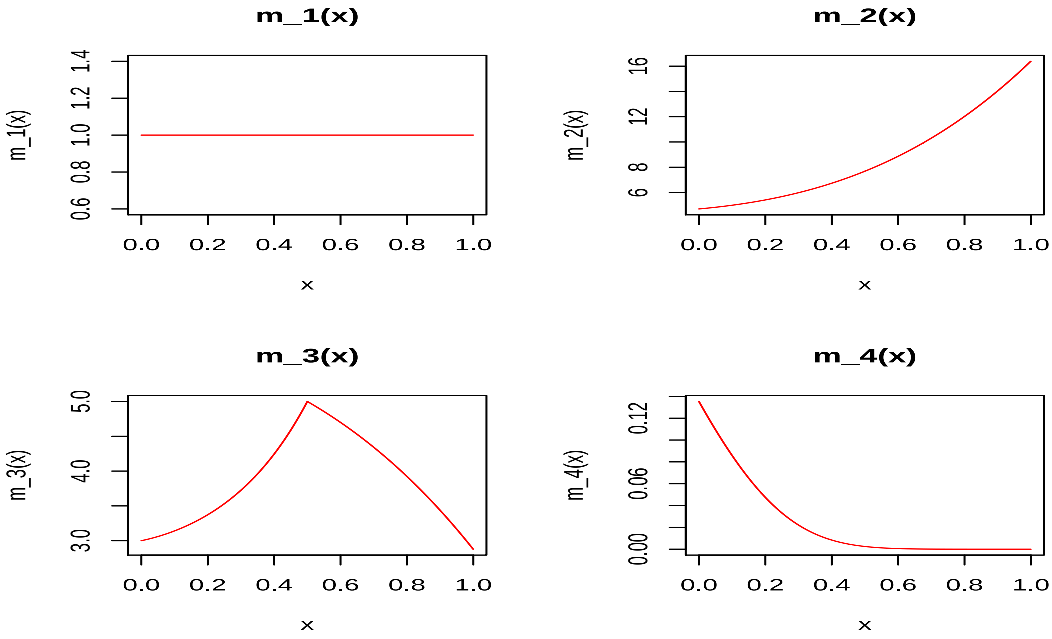

Figure 3.1: *The Plots of the Mean Functions* $m_1(x) - m_4(x)$.

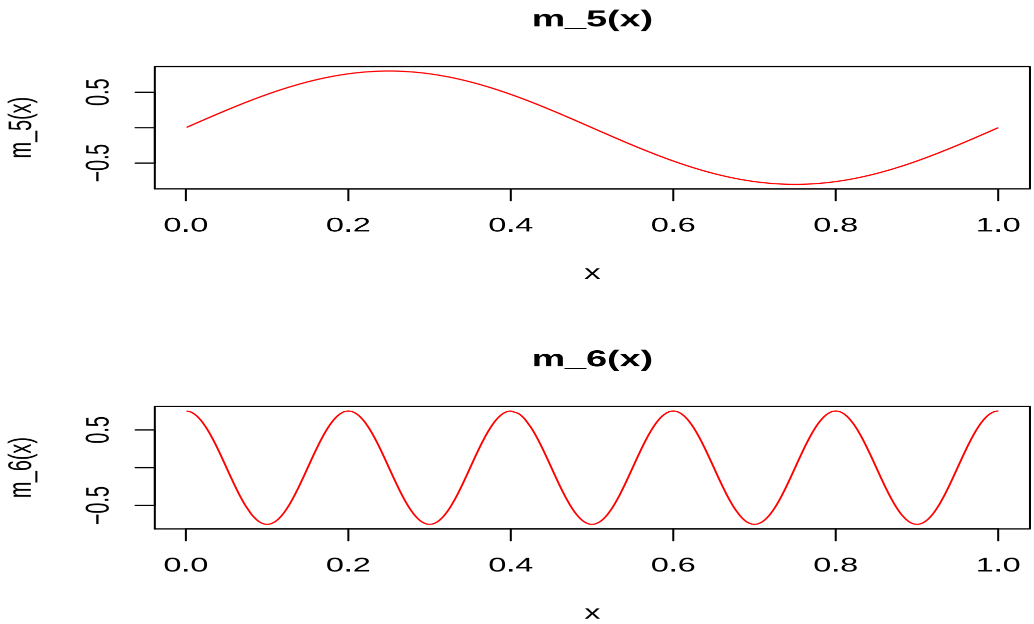

Figure 3.2: *The Plots of the Mean Functions* $m_5(x) - m_6(x)$.

three polynomial functions of different orders, an exponential function and two trigonometric functions. It is obvious that the shape of the first mean function is a line, whereas the shapes of the other mean functions are non-linear curves. It is noted that the last two mean function are periodic functions.

The error variances are chosen to be $\sigma^2 = 1,\ 4,\ 25$ and 100 in order to examine the effect of the size of the error variances on the finite sample performance of the estimator in (1.28). The aim of this section is to study the effect of the above mean functions on the mean and variance of the estimator in (1.28). Comparisons between the shape of the distributions of the estimator in (1.28), the Hall and Marron estimator (1990) and their asymptotic distribution are also presented for each mean function. It should be noted that the asymptotic performance of the estimator in (1.28) with bandwidth h^2 (in term of the mean squared error) is approximately equivalent to the asymptotic performance of the Hall and Marron estimator with bandwidth h. So, the bandwidth of the new estimator is selected to be square of that of the Hall and Marron estimator.

With the first mean function as stated above, our model is

$$Y_i = 1 + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n. \tag{3.1}$$

We first choose randomly a hundred ϵ_is and then Y_is are generated using the model (3.1). The bandwidths of the new estimator and the Hall and Marron estimator are taken as 0.16 and 0.4, respectively. Then, for the chosen bandwidths, $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ are computed where

$$\hat{\sigma}^2_{NEW} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} w_{ij}\, Y_i\, Y_j. \tag{3.2}$$

and $\hat{\sigma}^2_{HM}$ and w_{ij} are defined in equations (1.5) and (1.27), respectively. After that, we

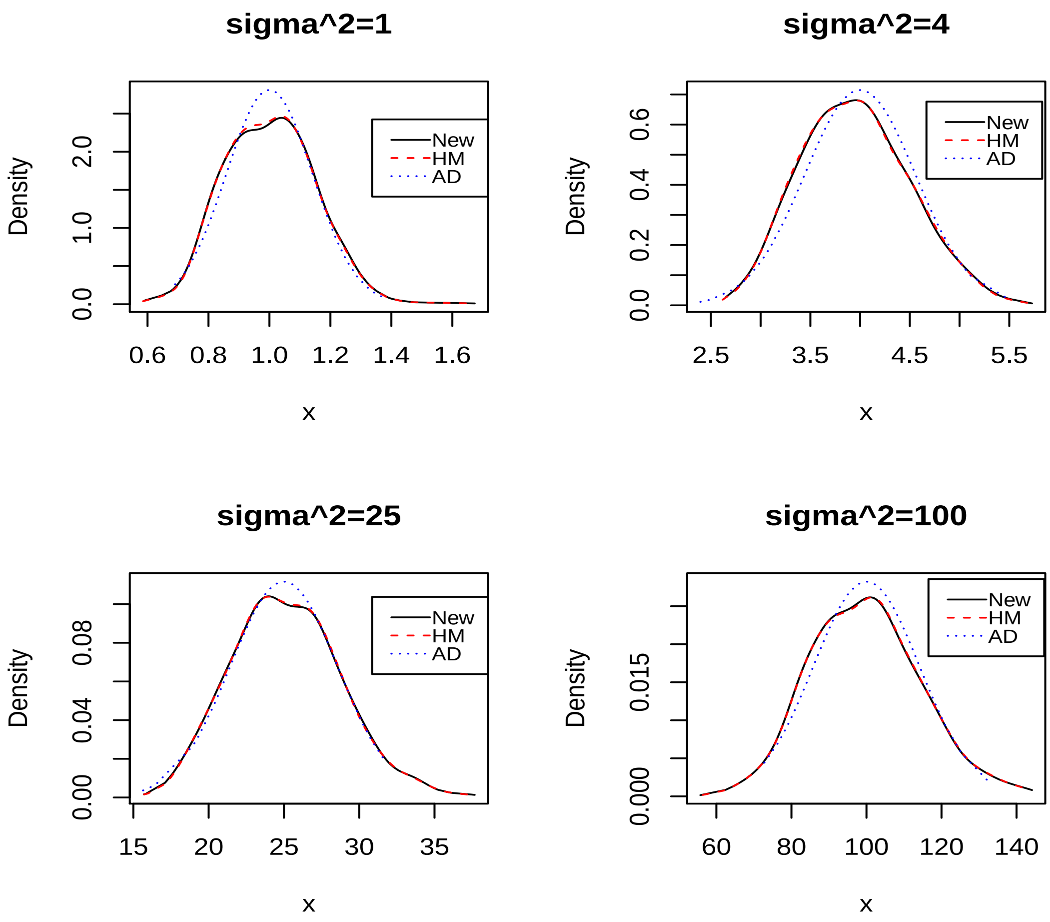

Figure 3.3: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted line) where* $m(x) = m_1(x)$.

replicate the above steps for $N = 1000$ times. Then, using $N = 1000$ observed values of $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$, the kernel density estimate of these estimators are plotted in the figure (3.3).

For the mean functions $m_2(x)$- $m_6(x)$, the same steps above are repeated where the models, the size of samples and the bandwidths for the kernel function in the estimators are specified for these mean functions as follows. For the $m_2(x)$, the model defines as

$$Y_i = 4.7 + 2.4\, x_i + 5\, x_i^2 + 4.3\, x_i^3 + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n.$$

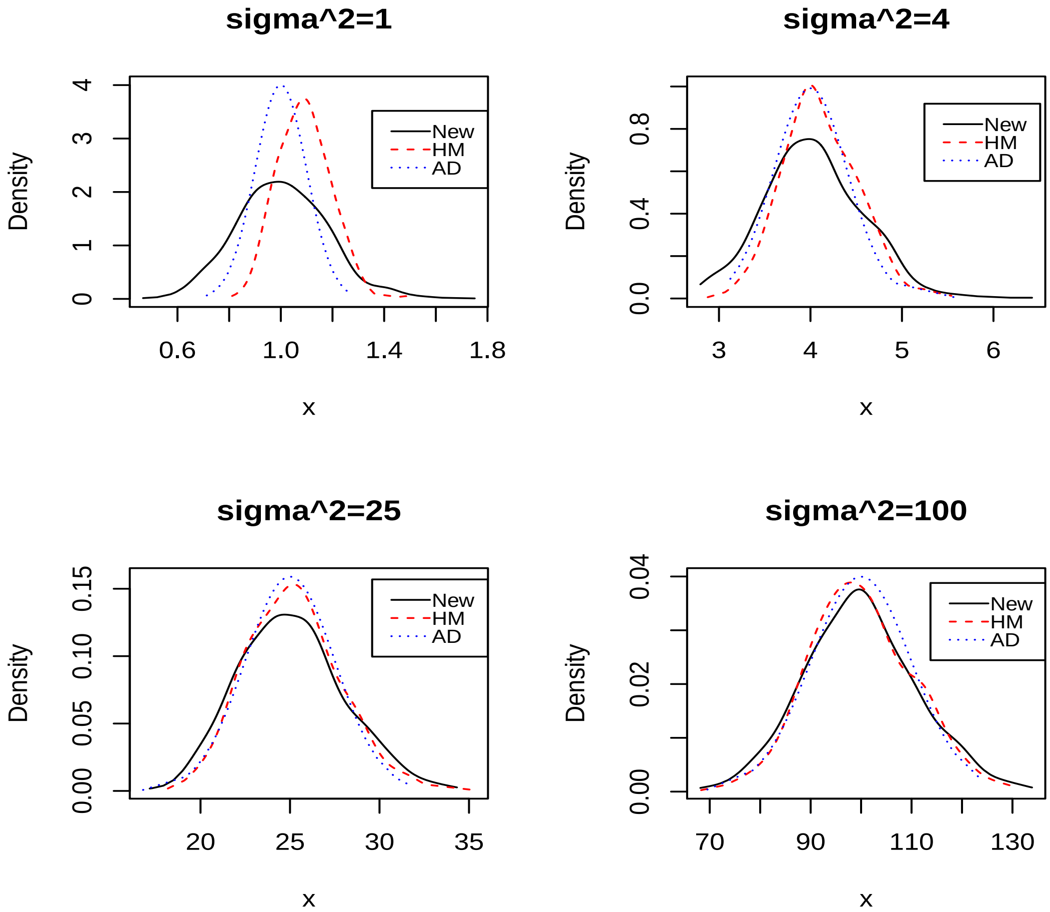

Figure 3.4: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted line) where* $m(x) = m_2(x)$.

In the figure (3.4), the sample size is $n = 200$. In all plots in this graph, the bandwidths are selected as 0.0064 and 0.08 for the estimator in (3.2) and the Hall and Marron estimator, respectively.

Using the third mean function $m_3(x)$, the model becomes

$$\begin{aligned} Y_i &= \left(3 + x_i + 4\,x_i^2 + 8\,x_i^4\right) \cdot I(x_i \leq 0.5) \\ &+ \left(5.875 - x_i - x_i^2 - x_i^3\right) \cdot I(\,x_i > 0.5) + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n. \end{aligned}$$

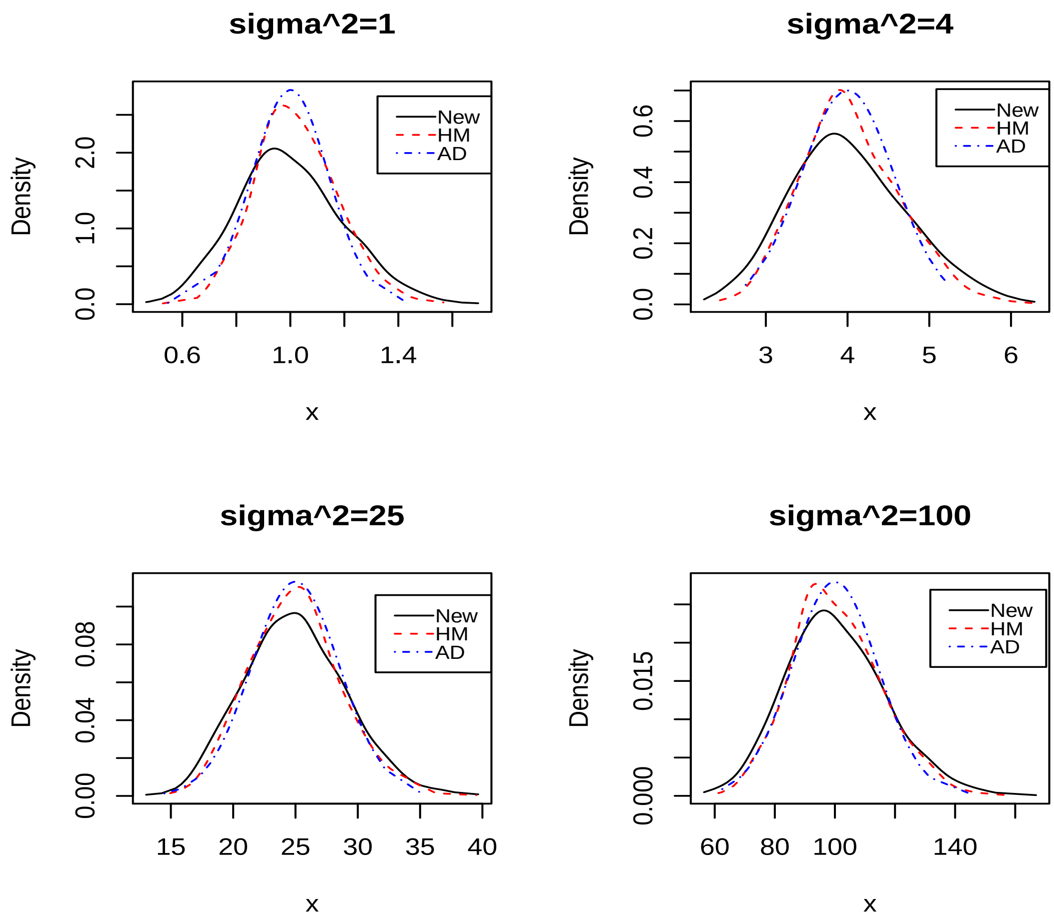

Figure 3.5: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted-dashed line) where* $m(x) = m_3(x)$.

The sample size for the figures (3.5)-(3.8) is chosen to be $n = 100$. The bandwidths of the kernel function in the new and the Hall and Marron estimators are taken as 0.01 and 0.1, respectively, in the figure (3.5).

Using the fourth mean function $m_4(x)$, the model becomes

$$Y_i = \exp\left(-2 - 4\,x_i - 5x_i^2 - 6x_i^3\right) + \epsilon_i \quad \text{for } i = 1, 2, ...n.$$

The bandwidths of the kernel function in the new and the Hall and Marron estimators are

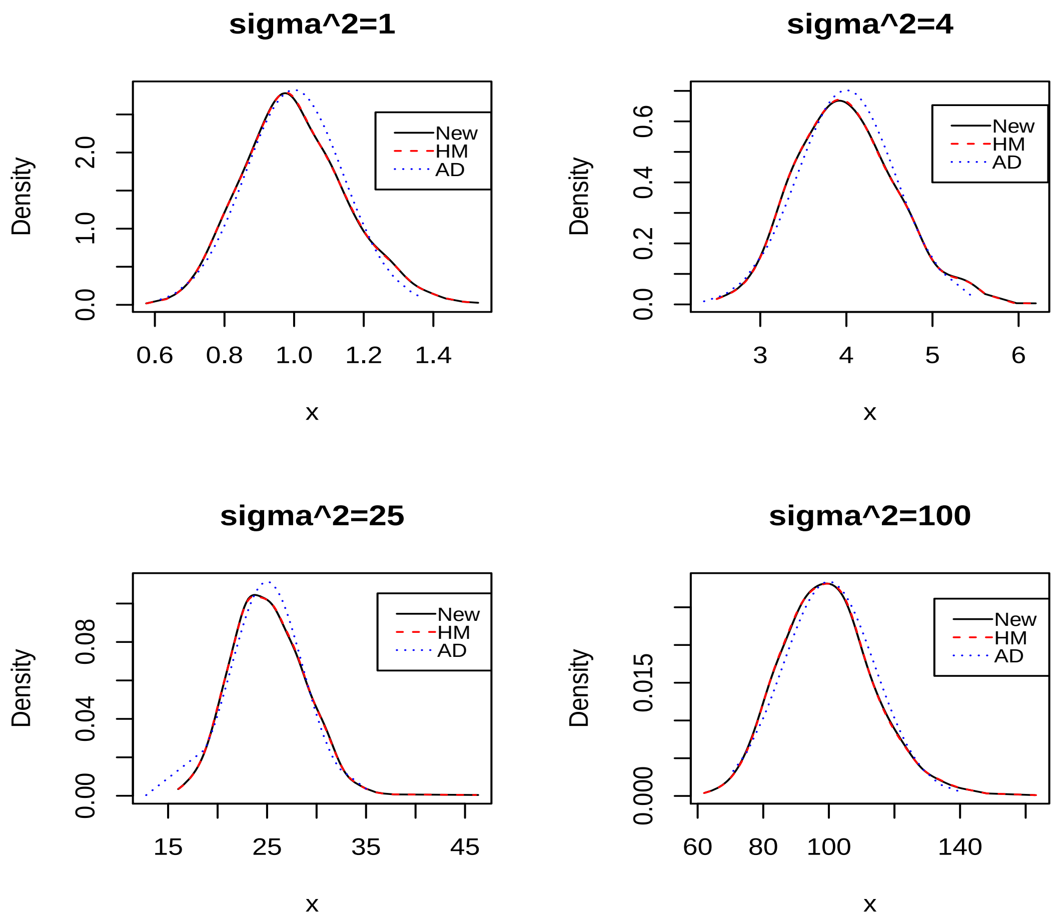

Figure 3.6: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted line) where* $m(x) = m_4(x)$.

taken as 0.25 and 0.5, respectively, in the figure (3.6).

The model using the $m_5(x)$ is

$$Y_i = \frac{4}{5} \sin(2\pi x_i) + \epsilon_i \quad \text{for } i = 1, 2, \ldots .n.$$

The model using the $m_6(x)$ can be defined as

$$Y_i = \frac{3}{4} \cos(10\pi x_i) + \epsilon_i \quad \text{for } i = 1, 2, \ldots .n.$$

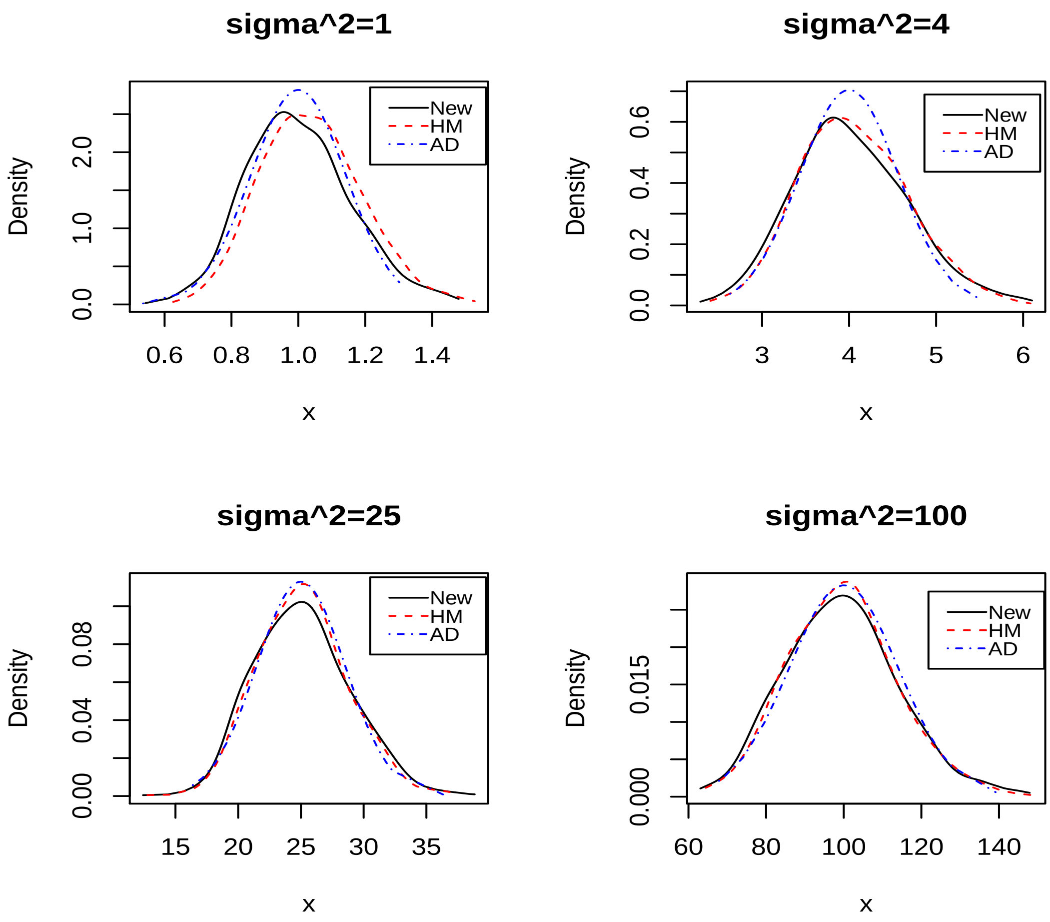

Figure 3.7: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted-dashed line) where* $m(x) = m_5(x)$.

For the figure (3.7), the bandwidths are chosen to be 0.0225 and 0.15 for the estimator in (3.2) and the Hall and Marron estimator, respectively. However, the bandwidths in the figure (3.8) are taken as 0.0144 and 0.12 for the same estimators, respectively. Note that, for the mean functions $m_2(x)$ to $m_6(x)$, the estimated kernel density function of the considered estimators is plotted in the figures (3.4)-(3.8), respectively, where the number of the replications is $N = 1000$.

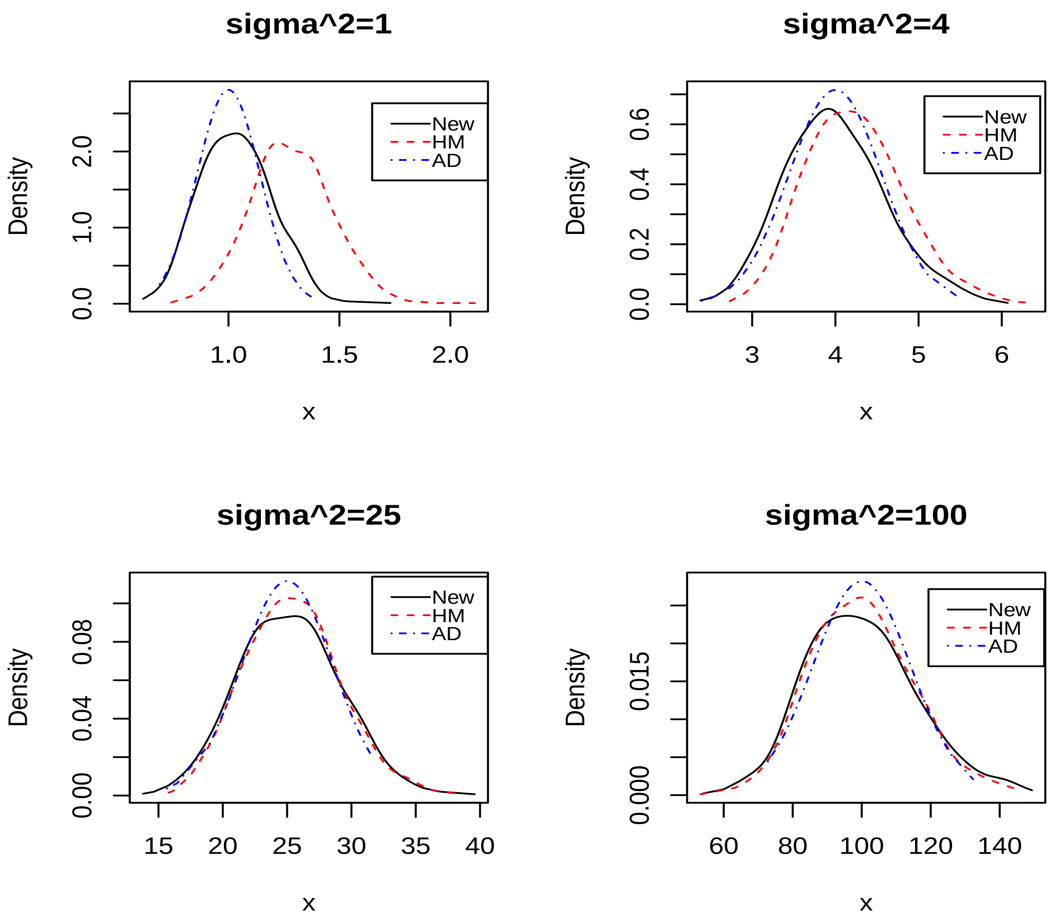

Figure 3.8: *The Comparison Between the Estimated Distributions of the New Estimator (solid line) and the H & M Estimator (dashed line) and their Asymptotic Distribution (dotted-dashed line) where* $m(x) = m_6(x)$.

3.3.1 Results:

Clearly, the figures (3.3) to (3.8) provide numerical verification of Corollary 2.2.1. But importantly, from the figures (3.3) and (3.6), it can be seen that means and variances of the $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ estimators are the same for all levels of σ^2 . In addition, we can conclude that the means and variances of the estimated distributions of the estimators are approximately the same for all levels of error variance in the figure (3.7).

From the figures (3.4), (3.5) and (3.8), for the small values of the error variances, the means of the $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ are approximately the same but the variance of the estimator in (3.2) is bigger than that of the Hall and Marron estimator. In the large values of the error variances, these two estimators have roughly the same means and variances. So, it is clear that as the error variance increases, the means and variances of these two estimators become close to each other.

For the mean functions $m_1(x)$, $m_4(x)$ and $m_5(x)$, the estimator in (3.2), the Hall and Marron estimator and their asymptotic distribution have approximately the same means and variances with a slight difference at the top of the density function curves. From figures (3.4), (3.5) and (3.8), it is obvious that the variance of the asymptotic distribution, which was described in Corollary 2.2.1, is different than that of the estimator in (3.2) and the Hall and Marron estimator when the error variance is small (for $\sigma^2 = 1$ and 4). This point will be discussed later in the next section. It should be noted that we studied the effect of some other mean functions. For details, see Alharbi (2011).

Remark:

1) The main R codes of all figures in this chapter are given in appendix A.

2) The numerical results of the simulation studies for figures (3.3) to (3.8) are given in appendix B.

3.3.2 Discussion:

To compare the mean squared errors of the estimator in (3.2) and the Hall and Marron estimator, we require $\int y^2 K(y)\, dy = 1$, $\int K^2(y)\, dy = 0.2821$ and $\int [K * K(y) - 2\, K(y)]^2\; dy = 0.40635$ where the kernel function K is the standard normal probability density function. It should be noted that, for ease in the calculation, all constants that involve kernel integration are obtained by integrating from $-\infty$ to $+\infty$ for all cases in the current chapter.

In constant and simple linear regression models, the new and the Hall and Marron estimators are unbiased estimators for σ^2 since the second derivative of the mean function in these models is zero. So, the means of the distributions of these estimators are expected to be approximately the same for all levels of σ^2. To study the effect of the constants C_1 and C_3 on the finite sample performance of the estimator in (3.2), we define the following mean function

$$m_7(x) = 0.2 + 0.4\,x + 0.004\,x^2 + 0.3\,x^3 + 0.02\,x^4 + 0.6\,x^5.$$

For $m_2(x)$ where $n = 200$, $h_{NEW} = 0.0064$ and $h_{HM} = 0.08$, the biases of the estimator in (3.2) and Hall and Marron estimator equal to 2.1×10^{-5} and 3.5×10^{-5}, respectively. For $m_7(x)$ where $n = 200$, $h_{NEW} = 0.04$ and $h_{HM} = 0.2$, the biases are 8.65×10^{-6} and 1.6×10^{-4}, respectively . So, it is obvious that the bias is negligible for both estimators in these two cases and is approximately the same. Note that, in these two cases, the bandwidth of the estimator in (3.2) is chosen as square of the bandwidth of the Hall and Marron estimator as described in the remark 2 of Theorem 2.2.1. Thus, the difference in the bias is due to the constants. For the detailed calculation of the bias, see Alharbi (2011).

From chapter 2, we know that the estimator in (3.2) and the Hall and Marron estimator have approximately the same variance in the first order. To compare the variance in the second order, we require to compute $E1 = n^{-2}\,h^{-1}\,C_3$ and $E2 = n^{-2}\,h^{-1}\,2\,C_7$ where C_3 is as in Theorem 2.2.1 and C_7 is defined in equation (2.50). Note that $\int_0^1 m_2^2(x)\,dx = 86$ and $\int_0^1 m_7^2(x)\,dx = 0.4513$. For $m_2(x)$ where $n = 200$, $h_{NEW} = 0.0064$, $h_{HM} = 0.08$ and $\sigma^2 = 1, 4, 25$ and 100, we obtain $E1 = 0.38, 1.6, 10.9$ and 60 and $E2 = 0.00025, 0.0004, 0.15$ and 2.5, respectively. For $m_7(x)$ where $n = 200$, $h_{NEW} = 0.04$,

$h_{HM} = 0.2$ and $\sigma^2 = 1, 4, 25$ and $100,$ we get $E1 = 0.00067, 0.007, 0.22$ and 3.5 and $E2 = 0.0001, 0.0016, 0.0625$ and $1,$ respectively. From comparisons of $E1$ and $E2$ for $m_2(x)$ and $m_7(x)$, it is obvious that the differences between $E1$ and $E2$ for $m_2(x)$ are larger than that of $m_7(x)$. Thus, these two estimators have approximately the same variances in the second order for $m_7(x)$. But importantly, the differences in the variances of the estimator in (3.2) and the Hall and Marron estimator are due to the differences in the constants. So, this difference becomes negligible as $n \to +\infty,\ h \to 0$ such that $n\,h \to +\infty$. From the discussion above and remark 2 of Theorem 2.2.1, it is clear that these two estimators expect to have a similar behaviour when $n \to +\infty,\ h \to 0$ such that $n\,h \to +\infty$.

In general, we found that when $\int_0^1 m^2(x)\ dx > 1$ (This term comes from the constant C_3) in the polynomial regression models of order ≥ 3 or when the mean function is a periodic function, the variances of the $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ estimators are expected to be different for small levels of σ^2. However, for large levels of σ^2, the means and the variances of the $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ estimators are nearly the same. In addition, when $\int_0^1 m^2(x)\ dx < 1$, the means and variances of both estimators are approximately the same for all levels of σ^2.

For the exponential mean function, it is obvious that the distributions of the two estimators are the same. This may happen because the exponent of a negative polynomial function is always between 0 and 1. In general, for any mean function of the type $m(x_i) = \exp(A)$ where A is a polynomial regression function and satisfy the constraint $A < 0$ for all x_is, the means and variances of the distributions of the estimator in (3.2) and the Hall and Marron estimator are approximately the same.

In all of the above cases, as $n \to +\infty,\ h \to 0$ such that $n\,h \to +\infty$, the distributions of the estimator in (3.2) and the Hall and Marron estimator are expected to be the same as their asymptotic distribution, which was described in Corollary 2.2.1. For small noise levels, when the mean function is a polynomial function of order ≥ 3 or a periodic function, the difference in the variances between the two considered estimators and their asymptotic distribution is due to the constants, and this difference becomes negligible as $n \to +\infty,\ h \to 0$ such that $n\,h \to +\infty$.

Conclusion:

From the discussion above, we conclude that these estimators have approximately the same means and variances for constant and simple linear regression models and for the exponent of negative polynomial mean functions. In the polynomial regression models of order $\geq$ 3, the estimators have a similar distributions for large error variance or when $\int_0^1 m^2(x)\ dx < 1$. However, when $\int_0^1 m^2(x)\ dx > 1$, the variances of these two estimators are different for small levels of σ^2. This difference is due to the constants. When the mean function is a periodic function for small levels of error variance, the variances of the estimated distributions of these estimators are also different, but the means are approximately the same. It should be noted that these results hold for appropriate choices of the bandwidths.

3.4 The Effect of the Bandwidth Selection

Bandwidth selection is one of the most important issues in the smoothing technique. Therefore, the effect of this choice on the finite sample performance of the new estimator is considered here. To find this effect, a model with a fixed mean function and error variance is assumed. Then, the bandwidth is allowed to vary. In addition, the number of replications in this simulation study is selected to be $N = 1000$ with samples of size $n = 200$. It should be noted that this simulation study has the same structure as in section 3.2.

We suppose that

$$Y_i = 1 + x_i + 0.7\,x_i^2 + 2\,x_i^3 + 1.5\,x_i^4 + 2\,x_i^5 + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n \qquad (3.3)$$

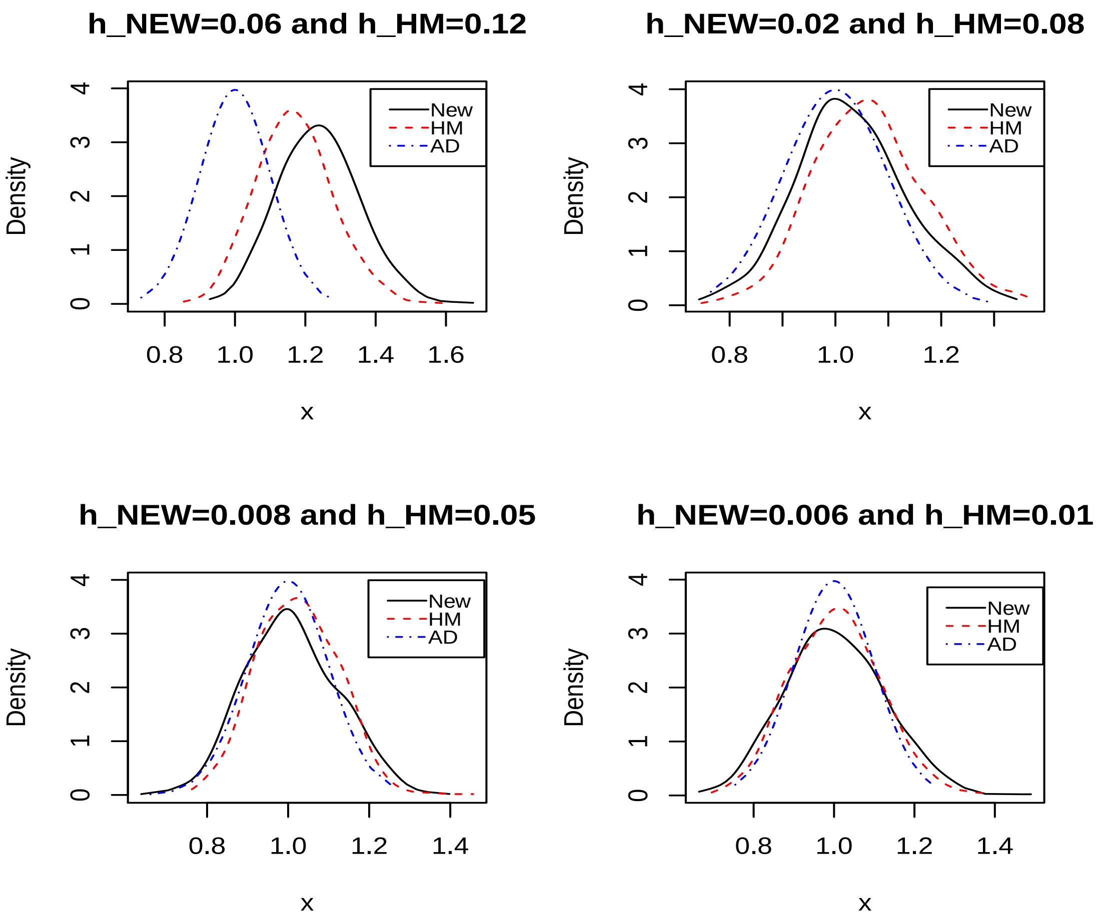

Figure 3.9: *The Comparison Between the Distributions of the New Estimator and the H & M Estimator Using a Model (3.3) with* $\sigma^2 = 1$.

where the assumptions C1, C2 and C3 are satisfied. For this model, two different levels of σ^2 are studied. In particular, σ^2 is chosen to be 1 and 36. The numerical results of the simulation studies in this section are presented in appendix B.

The figure (3.9) indicates the comparison between the estimated distributions of the estimator in (3.2) and the Hall and Marron estimator for various bandwidths, where $\sigma^2 = 1$. In particular, the bandwidths are taken as $h = 0.06,\ 0.02,\ 0.008$ and 0.006 for the estimator in (3.2) and $h = 0.12,\ 0.08,\ 0.05$ and 0.01 for the Hall and Marron estimator. In the figure (3.9), all chosen bandwidths are clearly given optimal results for the estimator in (3.2) except the bandwidth 0.06. As expected for the new estimator defined in (3.2), it is obvious that large bandwidth gives a small variance and large bias. However, small bandwidth gives large variance and small bias. For the Hall and Marron estimator, the bandwidths 0.12 and 0.08 indicate some bias in the estimation of σ^2. In addition, it is obvious that when an appropriate bandwidth is used, then these two estimators and their asymptotic distribution, which was stated in Corollary 2.2.1, have roughly the same means.

The second case for the model (3.3) is for $\sigma^2 = 36$. The figure (3.10) shows the comparisons between the estimated distributions of the same estimators in the figure (3.9) with $\sigma^2 = 36$ and a sample of selected bandwidths, chosen to be $h = 0.2, 0.1, 0.05$ and 0.01 for the estimator in (3.2) and $h = 0.4, 0.25, 0.1$ and 0.07 for the Hall and Marron estimator.

The conclusion from figure (3.10) shows that the bandwidths $h = 0.1, 0.05$ and 0.01 are nearly optimal for the estimator in (3.2) since the bandwidth 0.2 has given a slight bias result in the estimation of σ^2. On the other hand, for the Hall and Marron estimators, the last three of the chosen bandwidths are approximately optimal since the bandwidth $h = 0.4$ has given a bias estimation for $\sigma^2 = 36$. It is clear that if an appropriate bandwidth is used for the estimator in (3.2) and the Hall and Marron estimator, then these two estimators and their asymptotic distribution have approximately the same means and variances.

Conclusion: From the figures (3.9) and (3.10), the following conclusion can be drawn. There is clear influence of the bandwidth choice on the finite sample performance of the estimator in (3.2). For small error variance, the estimator in (3.2) has a narrow interval for the optimal bandwidths choice. In addition, the interval of the optimal bandwidth selection becomes wider for the estimator in (3.2) as the error variance increases. In the next section, these results are investigated through plotting the logarithms of various bandwidths versus the logarithms of their mean squared errors.

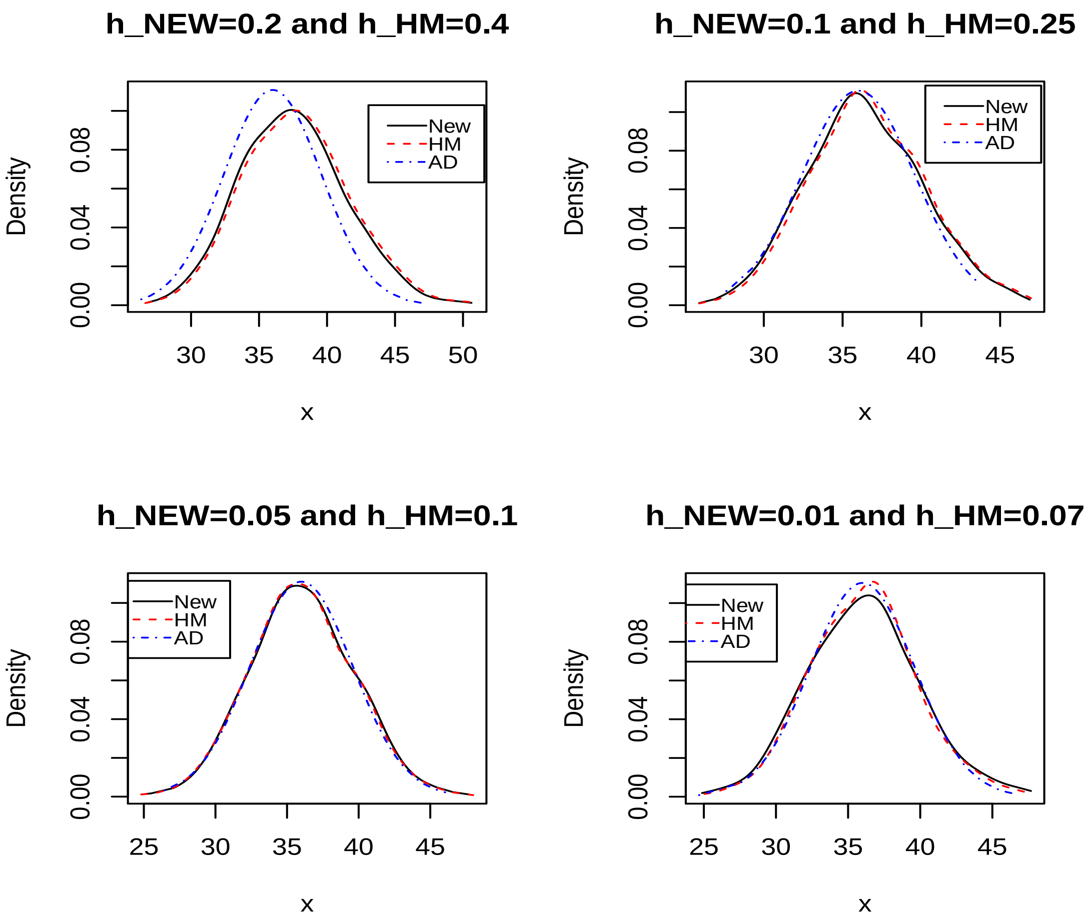

Figure 3.10: *The Comparison Between the Distributions of the New Estimator and the H & M Estimator Using a Model (3.3) with* $\sigma^2 = 36$.

3.5 The Relation Between Bandwidths and the Mean Squared Error

This section examines the differences between the mean squared error for the estimator in (3.2) and the Hall and Marron estimator for various bandwidths where the bandwidth of the estimator in (3.2) is selected as square of the bandwidth of the Hall and Marron estimator. Two different models are studied. In the first model, the bias is zero, which occurs when the second derivative of the mean function is zero, such as in the simple linear regression model. The bias in the second model is bigger than zero.

Assuming that

$$Y_i = 3 + 2\,x_i + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n \tag{3.4}$$

where the assumptions C1, C2 and C3 are satisfied. The aim is to plot a sample of the logarithms of selected bandwidths against the logarithms of their mean squared errors. In particular, the bandwidths for the Hall and Marron estimator are chosen to be from 0.00001 to 0.7, where the difference between $h_{(i)}$ and $h_{(i+1)}$ equals to 0.001. The squares of these bandwidths are used for the estimator in (3.2). The size of sample is assumed to be $n = 1000$. In addition, the kernel function in the $\hat{\sigma}^2_{NEW}$ and $\hat{\sigma}^2_{HM}$ estimators is the standard normal probability density function. Note that $\int_0^1 m^2(x)\,dx = 16.3334$.

The figure (3.11) shows the plot of the logarithms of various bandwidths versus the logarithms of the asymptotic mean squared errors of the estimator in (3.2) and the Hall and Marron estimator where the model (3.4) is used. From this figure, it can be concluded that there is a wide range of optimal bandwidth choices for this kind of models even when the true σ^2 is small. It is also noted that the logarithms of the asymptotic mean squared errors of these two estimators are approximately the same for all chosen levels of σ^2.

A polynomial regression model of order 3 is now studied. Under C1, C2 and C3, suppose that

$$Y_i = 8 + 3\,x_i + 4\,x_i^2 + 5\,x_i^3 + \epsilon_i \quad \text{for } i = 1, 2, \cdots, n. \tag{3.5}$$

To find the logarithms of the asymptotic mean squared errors of the considered estimators, it is important to note that $\int_0^1 m^2(x)\;dx = 121.53$, $\int_0^1 m''(x)\,m(x)\,dx = 245$ and $\int_0^1 [m''(x)]^2\;dx = 516$. From the graph (3.12), it can be concluded that the optimal bandwidths of the estimator in (3.2) for small values of error variances have a narrow interval of choices. However, it is

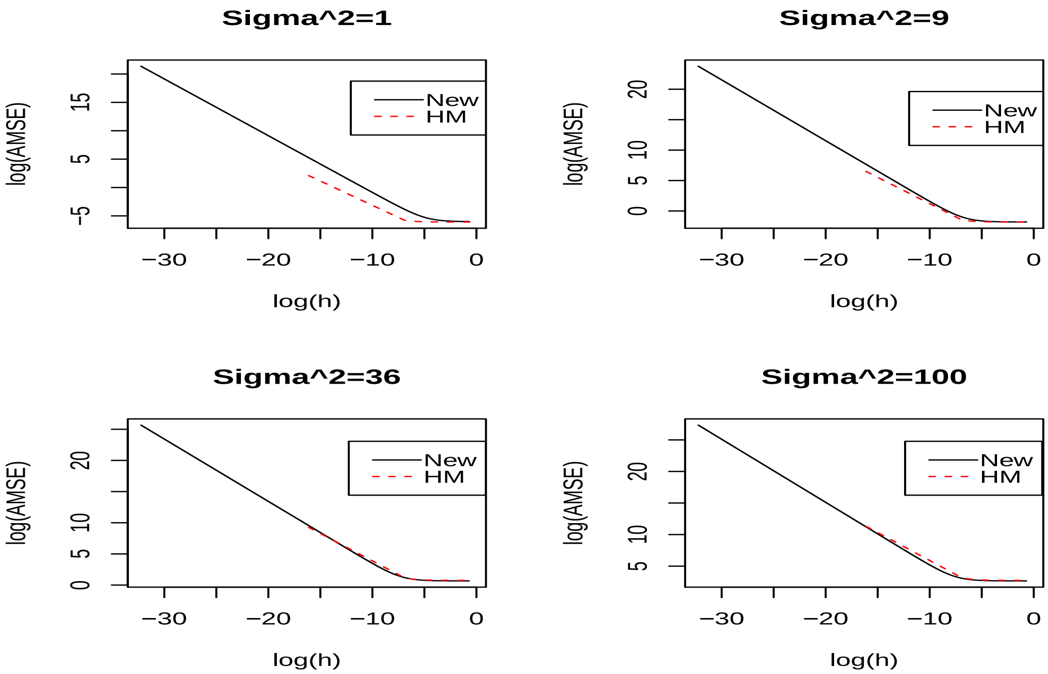

Figure 3.11: *The Plot of the Logarithms of Selected Bandwidths versus the Logarithms of the Asymptotic Mean Squared Error of* $\hat{\sigma}^2_{NEW}$ *(solid line) and* $\hat{\sigma}^2_{HM}$ *(dashed line) using Model (3.4).*

obvious that the interval of the optimal bandwidths increases as the the error variance rises. In addition, for all chosen noise levels, both of the estimators have approximately the same minimum of the asymptotic mean squared error.

For the polynomial regression models of order higher than 3 and when $\int_0^1 m^2(x)\, dx > 1$, models of different orders have been studied. The same comparisons between the logarithms of chosen bandwidths and the logarithms of their mean squared errors have been made by using these models. The conclusion of these comparisons is similar to the results of the model (3.5). So, the details are omitted.

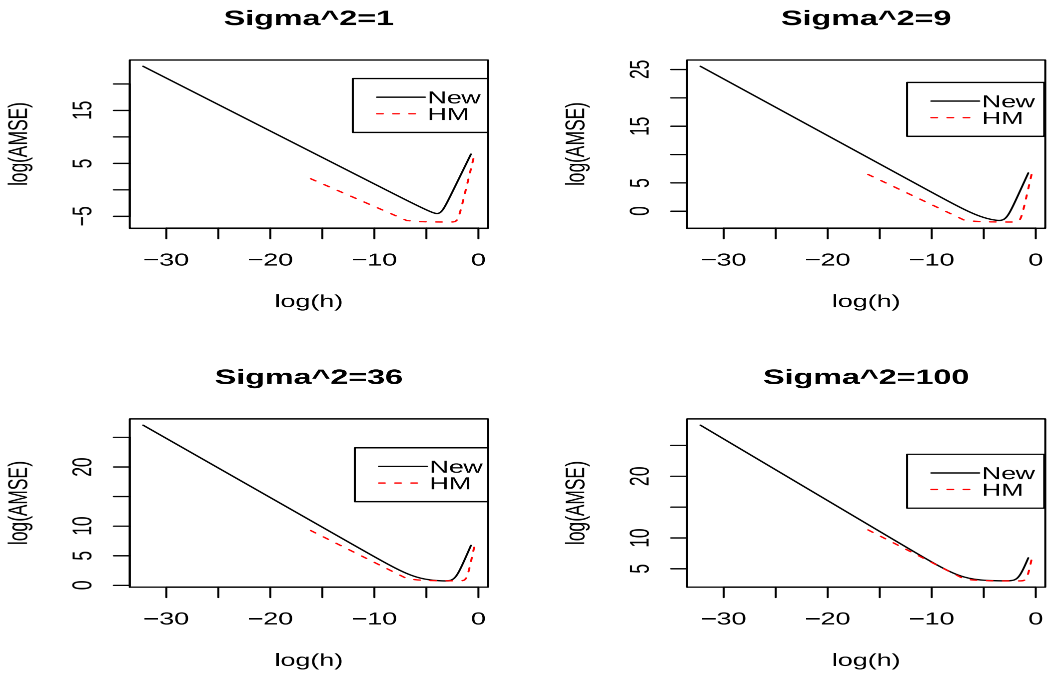

Figure 3.12: *The Plot of the Logarithms of Selected Bandwidths versus the Logarithms of the Asymptotic Mean Squared Error of* $\hat{\sigma}^2_{NEW}$ *(solid line) and* $\hat{\sigma}^2_{HM}$ *(dashed line) using Model (3.5).*

3.6 Summary

In the first part of this chapter, the effect of the mean function on the finite sample performance of the estimator in (3.2) is discussed. For the optimal bandwidth, we found that the means and variances of the estimator in (3.2), the Hall and Marron estimator and their asymptotic distribution are approximately the same for constant and simple linear regression models, the exponent of negative mean functions, and polynomial regression models when $\int_0^1 m^2(x)\ dx < 1$. When the mean function is a periodic function or a polynomial function where $\int_0^1 m^2(x)\ dx > 1$, the same previous result can also be drawn for large values of error variances. For small error variances, however the variances of the estimators differ. But importantly, this difference is due to the constants. The results in section 3.3 provide a numerical verification of normality of the estimator in (3.2).

The effect of the bandwidth selection is looked at in sections 3.4 and 3.5 in this chapter. We conclude that the bandwidth choice has clear influence on the finite sample performance of the estimator in (3.2). From section 3.5 and for the constant and simple linear regression models, we found that there is a wide range of the optimal bandwidth choices. In contrast, for polynomial regression models of order ≥ 3, when $\int_0^1 m^2(x)\, dx > 1$, the estimator in (3.2) has narrow interval of the optimal bandwidth choices for small error variances. Thus, the asymptotic performance of the estimator in (3.2) is affected by small variation in the bandwidth choices. However, the interval of optimal bandwidth choices increases as the error variance rises for both of the estimators. These results are supported by the results of simulation studies in section 3.4.

Chapter 4

The Mean Squared Error of a New Estimator for Functional Error Variance

4.1 Introduction

The error of variance can be defined in two different settings under consideration: the model (1.3) refers to constant variance (homoscedastic nonparametric regression) model and the model (1.4) refers to variance function (heteroscedastic nonparametric regression) model. So far we have discussed a new estimator of error variance in the setting of homoscedastic nonparametric regression model. In this chapter, we consider the following heteroscedastic nonparametric regression model

$$Y_i = m(x_i) + \sqrt{v(x_i)}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n \tag{4.1}$$

where $m(x_i)$ represents the unknown mean function $E(Y_i|x_i)$, Y_is denote the response variable, $v(x_i)$ represents the variance function and x_is denote the design points. The errors ϵ_is are assumed to be independent and identically distributed random variables with zero mean and unit variance and the fourth moment $\mu_4(x)$ is bounded where $\mu_r(x_i) = E\left[(Y_i - m(x_i))^r\right]$.

From section 1.6, a new estimator for the error variance function, $\hat{v}(x)$, is proposed where

$$\begin{aligned}
\hat{v}(x) &= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2 \\
&- \frac{1}{n\,(n-1)h_2\,h_1}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) Y_i Y_j \\
&= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\left[Y_i - \frac{1}{(n-1)\,h_1}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h_1}\right) Y_j\right] Y_i \\
&= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) e_i\, Y_i \qquad (4.2)
\end{aligned}$$

Note that $K(.)$ is a kernel function satisfying the assumptions $A1$ and $A2$ stated in section 1.6 and $e_i = Y_i - \frac{1}{(n-1)\,h_1}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h_1}\right) Y_j$. Thus, importantly it may be noted that we smooth $e_i\,Y_i$s as opposed to smooth e_i^2s in the residual-based estimators in order to estimate the error variance function. The bandwidth h_1 is used to estimate the mean function, whereas the bandwidth h_2 is used to estimate the variance function by way of smoothing $e_i\,Y_i$s. Our aim in the current chapter is to study the mean square error properties of the estimator given in (4.2). In particular, we will study the effect of the bandwidths h_1 and h_2 on the mean squared error of the estimator $\hat{v}(x)$.

Brown and Levine (2007) defined a class of difference-based estimators in the setting of the heteroscedastic nonparametric regression model. To compare the asymptotic mean squared error of the Brown and Levine estimator to that of the estimator in (4.2), we carry out the mean squared analysis of linear version of the Brown and Levine difference-based estimators. In this case, the Brown and Levine estimator can be defined as

$$\hat{v}_{BL}(x,1,h) = \frac{1}{n\,h}\sum_{i=1}^{n}\frac{\{\hat{S}_2(x,h) - \hat{S}_1(x,h)\,(x_i - x)\}\,K\left(\frac{x-x_i}{h}\right)\,\Delta_i^2}{\hat{S}_2(x,h)\,\hat{S}_0(x,h) - \hat{S}_1^2(x,h)} \qquad (4.3)$$

where $\Delta_i = \sum_{k=0}^{m} d_j \; y_{i+k-\lfloor m/2 \rfloor}$, for $i = \lfloor m/2 \rfloor + 1, ..., n + \lfloor m/2 \rfloor - m$; $\lfloor a \rfloor$ represents the largest integer number less than a; $\hat{S}_r(x,h) = \frac{1}{n} \sum_{i=2}^{n-1} (x_i - x)^r \, K_h(x_i - x)$; $K_h(u) = \frac{1}{h} K(\frac{u}{h})$; m denotes the order of differences and the difference sequence d_j is such that $\sum_{j=1}^{m} d_j = 0$ and $\sum_{j=1}^{m} d_j^2 = 1$. For more details, see Brown and Levine (2007).

This chapter is organised as follows. The main theorem, which gives the bias and the variance of the estimator in (4.2), is stated in section 4.2, whereas its proof is provided in section 4.3. In section 4.4, we give an outline of proof of the mean squared error of the Brown and Levine estimator stated in (4.3) where the order of differences is 2. Since the bias and the variance of $\hat{v}(x)$ depends on the bandwidths, we will briefly discuss the bandwidth selection and its optimal choice in section 4.5.

4.2 The Main Results

Our main aim here is to establish mean squared error properties of the estimator in (4.2). In addition to the assumptions $A1$ and $A2$ stated in section 1.6, we make the following assumptions:

E1: The kernel function K has r-continuous derivatives in $[-1, 1]$ where $r \geq 2$.

E2: x_is are equidistant design points in the interval [0,1] such that $x_i = i/n$ for $i = 1, 2, \cdots, n$.

E2′: The design points x_is are randomly chosen from the $U\,[0, 1]$ distribution.

E3: The mean function $m(x)$, the variance function $v(x)$, $\mu_3(x)$ and $\mu_4(x)$ are bounded, integrable, differentiable and have r-continuous derivatives.

E4: $n \to +\infty, h_1 \to 0$ and $h_2 \to 0$ such that $nh_1 \to \infty$, $nh_2 \to \infty$ and $\frac{h_1}{h_2} \to 0$.

The following theorem gives the bias and variance formulae of the estimator in (4.2).

Theorem **4.2.1.** *Suppose A1, A2, E1, E2, E3 and E4 are satisfied and* $h_2 \sim n^{-\alpha}$*, where* α

is positive constant such that $\alpha < 1$ *. Let* $r > 0$ *to be an even number, then*

$$(i)\, E(\hat{v}(x)) - v(x) = h_2^r \cdot C_1(x) + o(h_2^r) + O(n^{-1} h_2^{-1}),$$
$$(ii)\, Var\,(\,\hat{v}(x)\,) = n^{-1} h_2^{-1} C_2(x) + o(n\, h_2)^{-1},$$

where

$$\begin{aligned} C_1(x) &= \frac{1}{r!}\, v^{(r)}(x) \int y^r\, K\,(y)\; dy \;\; and \\ C_2(x) &= \big(\mu_4(x) - v^2(x)\big) \cdot \int K^2\;(t)\; dt. \end{aligned}$$

Remark:

1) By using the bias and the variance in the above theorem, the mean squared error of the estimator in (4.2) can be described as

$$\begin{aligned} MSE(\hat{v}(x)) &= (E(\hat{v}(x)) - v(x))^2 + Var\,(\hat{v}(x)) \\ &= h_2^{2\,r}\, C_1^2\,(x) + n^{-1} h_2^{-1}\, C_2(x) + o(n^{-1}\, h_2^{-1}) + o(h_2^{2\,r}). \end{aligned} \tag{4.4}$$

Note that the contributions of the bias and the variance in the above mean squared error depend on x through $C_1(x)$ and $C_2(x)$, respectively. Clearly from the above formula, the mean squared error of the estimator in (4.2) depends only on the bandwidth h_2. For the effect of the bandwidth h_1 on the mean squared error of this estimator, see the remark 2 below. In the above theorem, the bandwidth $h_2 \sim n^{-\alpha}$ where α is a positive constant such that $\alpha < 1$. Note that when $\alpha < 1/(2\,r + 1)$, the contribution of the variance in the mean squared error of the estimator (4.2) is larger than the contribution of the bias. On the other hand, when $1/(2\,r + 1) < \alpha < 1$, the opposite occurs. So, taking $\alpha = 1/(2\,r + 1)$ is balance between the squared bias and the variance of the estimator in (4.2) (the squared bias and variance trade-off).

2) From Theorem 4.2.1, it is obvious that the leading terms do not depend on the bandwidth

h_1. However, one can quantify the second order effect of the bandwidth h_1 on the mean squared error of $\hat{v}(x)$, but we do not pursue that here. Thus, estimate of the mean function does not have a first-order effect on the mean squared error of the estimator $\hat{v}(x)$.

3) The new estimator $\hat{v}(x)$ has approximately the same bias and variance as the residual-based local polynomial variance function estimator of Ruppert, Wand, Holset and Hössjer (1997) described in section 1.5, when r is an even number such that $r \geq 2$ and $x_i = i/n$, for $i = 1, 2, \cdots n$.

4) For the Brown and Levine estimator in (4.3), we can verify that the mean squared error of this estimator, when the second-order differences is used, equals to

$$MSE(v\hat{_{BL}}(x)) = n^{-1} h_2^{-1} C_2(x) + h_2^{2r} C_1^2(x) + o(n^{-1} h_2^{-1}) + o(h_2^{2r}) \qquad (4.5)$$

where $C_1(x)$ and $C_2(x)$ are defined in Theorem 4.2.1. Thus, in this case, the asymptotic mean squared error of the estimator in (4.2) and the Brown and Levine estimator in (4.3) are the same. In other words, these two estimators have the same mean squared error in the first order. An outline of proof of the above mean squared error is discussed in section 4.4.

5) If the assumption $E2'$ is used instead of E2, the results in Theorem 4.2.1 are still satisfied.

4.3 A Proof of Theorem 4.2.1

This section is divided into two subsections. In the first subsection, we state a lemma that is required later in the proof of Theorem 4.2.1. This is followed by the proof of the above theorem in the second subsection.

4.3.1 Lemma

Lemma **4.3.1.**

Suppose the assumptions A1,A2, E1, E2, E3 and E4 hold. Then,

(i)
$$\frac{1}{n^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right) K^2\left(\frac{x_i-x_j}{h_1}\right) m(x_j)$$
$$= h_1\, h_2\ m(x) \int K^2\,(t)\ dt\ + O(n^{-1}\, h_1).$$

(ii)
$$\frac{1}{n^3}\sum\sum_{i\neq j\neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_i-x_k}{h_1}\right) m(x_j)$$
$$= h_1^2\, h_2\ m(x) \int K^2\,(t)\ dt + o(h_1^2\, h_2).$$

(iii)
$$\frac{1}{n^3}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_k-x_i}{h_1}\right) m(x_j)$$
$$= h_1^2\, h_2\ m(x) \int K^2\,(t)\ dt\ +\ o(h_1^2\, h_2).$$

The proof of the above lemma is similar to the proof of Lemmas 2.3.1 and 3.4.1 in Alharbi (2011). Thus, the details are omitted here. For more details, see Alharbi (2011).

4.3.2 A Proof of Theorem 4.2.1

To prove part (i) in Theorem 4.2.1, we require to find

$$\begin{aligned} E(\hat{v}(x)) &= E\left(\frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\right) \\ &- E\left(\frac{1}{n(n-1)h_1\, h_2}\sum\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) Y_i Y_j\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) E(Y_i^2) \\
&- \frac{1}{n(n-1)h_2\,h_1}\sum\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) E(Y_i)E(Y_j) \\
&= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) \left(m^2(x_i) + v(x_i)\right) \\
&- \frac{1}{n(n-1)h_2\,h_1}\sum\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) m(x_i)m(x_j).
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
E(\hat{v}(x)) &= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) v(x_i) + \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) m^2(x_i) \\
&- \frac{1}{n(n-1)h_2\,h_1}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) m(x_i)m(x_j). \quad (4.6)
\end{aligned}
$$

Since the mean function $m(x)$ and the variance function $v(x)$ are bounded, Lemma 4.3.1 and Lemma 2.3.1 in section 2.3 can be used to approximate the summations in equation (4.6). That is, the first and second terms on the right hand side of equation (4.6) can be approximated as

$$
\begin{aligned}
&\frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) \left[v(x_i) + m^2(x_i)\right] \\
&\qquad = \frac{1}{h_2}\int K\left(\frac{x-u}{h_2}\right) \left[v(u) + m^2(u)\right] du + O(n^{-1}\,h_2^{-1}) \\
&\qquad = \int K(y) \left[v(t-h_2\,y) + (m(t-h_2\,y))^2\right] dy + O(n^{-1}\,h_2^{-1}). \quad (4.7)
\end{aligned}
$$

By using Taylor series expansion, the right hand side of equation (4.7) becomes

$$
\begin{aligned}
\text{RHS} &= \int K(y)\left[v(t) - h_2\,y\,v'(t) + \frac{h_2^2}{2!}\,y^2\,v''(t) - \cdots + \frac{(-1)^r h_2^r}{r!}\,y^r\,v^{(r)}(t) + o(h_2^r)\right] dy \\
&+ \int K(y)\left[m(t) - h_2 y m'(t) + \frac{h_2^2}{2!}\,y^2 m''(t) - \cdots + \frac{(-1)^r h_2^r}{r!} y^r m^{(r)}(t) + o(h_2^r)\right]^2 dy \\
&+ O(n^{-1}h_2^{-1})
\end{aligned}
$$

$$
\begin{aligned}
&= \quad v(t) + m^2(t) + \frac{h_2^r}{r!}\, v^{(r)}(x) \int y^r K(y)\, dy \\
&+ \quad h_2^r \int y^r K(y)\, dy \left[\frac{1}{0!\, r!}\, m^{(0)}(x)\, m^{(r)}(x) + \frac{1}{1!\,(r-1)!}\, m^{(1)}(x)\, m^{(r-1)}(x) \right. \\
&+ \quad \left. \cdots + \frac{1}{r!\, 0!}\, m^{(r)}(x)\, m^{(0)}(x) \right] + o(h_2^r) + O(n^{-1} h_2^{-1}), \qquad (4.8)
\end{aligned}
$$

where $v^{(r)}(x)$ is the rth derivatives of the variance function $v(x)$ and $m^{(r)}(x)$ denotes the rth derivatives of the mean function $m(x)$.

By applying the first part of Lemma 4.3.1 to the third term on the right hand side of equation (4.6), we obtain

$$
\begin{aligned}
&\frac{1}{n(n-1)h_2\, h_1} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{x - x_i}{h_2}\right) K\left(\frac{x_i - x_j}{h_1}\right) m(x_i) m(x_j) \\
&\quad = \frac{1}{h_1\, h_2} \int\int K\left(\frac{x-t}{h_2}\right) K\left(\frac{t-u}{h_1}\right) m(t) m(u)\, dt\, du + O(n^{-1} h_2^{-1}) \\
&\quad = m^2(x) + \frac{h_1^r}{r!}\, m(x)\, m^{(r)}(x) \int z^r K(z)\, dz \\
&\quad + h_2^r \int y^r K(y)\, dy \left[\frac{1}{0!\, r!}\, m^{(0)}(x)\, m^{(r)}(x) + \frac{1}{1!\,(r-1)!}\, m^{(1)}(x)\, m^{(r-1)}(x) \right. \\
&\quad + \left. \cdots + \frac{1}{r!\, 0!}\, m^{(r)}(x)\, m^{(0)}(x) \right] + o(h_2^r) + O(n^{-1} h_2^{-1}). \qquad (4.9)
\end{aligned}
$$

where we used substitutions $\frac{t-u}{h_1} = z,\ \frac{x-t}{h_2} = y$ to derive the last expression on the right hand side of equation (4.9). Using (4.8) and (4.9), equation (4.6) simplifies to

$$
\begin{aligned}
E(\hat{v}(x)) \quad &= \quad v(x) + \frac{h_2^r}{r!}\, v^{(r)}(x) \int y^r K(y)\, dy + \frac{h_1^r}{r!}\, m(x)\, m^{(r)}(x) \int z^r K(z)\, dz \\
&+ \quad o(h_2^r) + O(n^{-1} h_2^{-1}) \\
&= \quad v(x) + h_2^r\, C_1(x) + o(h_2^r) + O(n^{-1} h_2^{-1}),
\end{aligned}
$$

where $C_1(x) = \frac{1}{r!}\, v^{(r)}(x) \int y^r K(y)\, dy.$

Therefore, the squared bias is

$$(E(\hat{v}(x)) - v(x))^2 = h_2^{2r}.C_1^2(x) + o(h^{2r}) + O(n^{-2} h_2^{-2}). \tag{4.10}$$

This complete the proof of part (i) in Theorem 4.2.1.

Now, note that $Var(\hat{v}(x)) = E(\hat{v}(x))^2 - (E(\hat{v}(x)))^2$, which means to compute the variance, we need to know $E(Y_i^3)$ and $E(Y_i^4)$. Therefore, note that

$$\begin{aligned} E(Y_i^3) &= \mu_3(x_i) + 3\,v(x_i)\,m(x_i) + m^3(x_i), \\ E(Y_i^4) &= \mu_4(x_i) + 4\,\mu_3(x_i)\,m(x_i) + 6\,v(x_i)\,m^2(x_i) + m^4(x_i). \end{aligned}$$

To find $E(\hat{v}(x))^2$, we first consider $(\hat{v}(x))^2$ and express it as

$$(\hat{v}(x))^2 = \sum_{a=1}^{6} P_a$$

where

$$\begin{aligned} P_1 &= \frac{1}{n^2\,h_2^2} \sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right) Y_i^4, \\ P_2 &= \frac{1}{n^2\,h_2^2} \sum\sum_{i \neq k} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) Y_i^2 Y_k^2, \\ P_3 &= \frac{-2}{n^2\,(n-1)\,h_2^2\,h_1} \left(\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\right) \left\{\sum\sum_{i \neq j} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) Y_i Y_j\right\}, \\ P_4 &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j} K^2\left(\frac{x_i-x_j}{h_1}\right) K^2\left(\frac{x-x_i}{h_2}\right) Y_i^2 Y_j^2, \\ P_5 &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum\sum_{i \neq j \neq k} \left(K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) Y_i Y_j\right) \\ &\times \left(K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_k}{h_1}\right) Y_i Y_k\right) \end{aligned}$$

and

$$
\begin{aligned}
P_6 &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k}\left(\sum_{j\neq i}K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)Y_iY_j\right)\\
&\times \left(\sum_{d\neq k}K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_k-x_d}{h_1}\right)Y_kY_d\right).
\end{aligned}
$$

To analyse P_1+P_2, note that by the independence of Y_i and Y_k for $i\neq k$, we get

$$
\begin{aligned}
E(P_1+P_2) &= \frac{1}{n^2h_2^2}\sum_{i=1}^{n}K^2\left(\frac{x-x_i}{h_2}\right)E(Y_i^4)\\
&+ \frac{1}{n^2h^2}\sum\sum_{i\neq k}K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)E(Y_i^2)E(Y_k^2)\\
&= \frac{1}{n^2h_2^2}\sum_{i=1}^{n}K^2\left(\frac{x-x_i}{h_2}\right)\left[\mu_4(x_i)+4\mu_3(x_i)\,m(x_i)+6v(x_i)m^2(x_i)+m^4(x_i)\right]\\
&+ \frac{1}{n^2h_2^2}\sum\sum_{i\neq k}K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)\\
&\times \left[v(x_i)\,v(x_k)+v(x_k)m^2(x_i)+v(x_i)m^2(x_k)+m^2(x_i)\,m^2(x_k)\right]. \qquad (4.11)
\end{aligned}
$$

In case of P_3, since Y_i, Y_j and Y_k are independent for $i\neq j\neq k$, we have

$$
\begin{aligned}
E(P_3) &= \frac{-2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j}K^2\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)E(Y_i^3)E(Y_j)\\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq k}K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_k-x_i}{h_1}\right)E(Y_k^3)E(Y_i)\\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_k-x_j}{h_1}\right)\\
&\times E(Y_i^2)E(Y_j)E(Y_k)\\
&= \frac{-2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j}K^2\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \left[\mu_3(x_i)\,m(x_j)+3\,v(x_i)\,m(x_i)\,m(x_j)+m^3(x_i)\,m(x_j)\right]\\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq k}K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_k-x_i}{h_1}\right)\\
&\times \left[\mu_3(x_k)\,m(x_i)+3\,v(x_k)\,m(x_k)\,m(x_i)+m^3(x_k)\,m(x_i)\right]
\end{aligned}
$$

$$
\begin{aligned}
&- \quad \frac{2}{n^2\,(n-1)\,h_2^2\,h_1} \sum\sum_{i \neq j \neq k}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_k-x_j}{h_1}\right) \\
&\times \quad \left[v(x_i)\, m(x_j)\, m(x_k) + m^2(x_i)\, m(x_j)\, m(x_k) \right]. \qquad (4.12)
\end{aligned}
$$

For the term P_4, again Y_i and Y_j are independent since $i \neq j$. Thus, we obtain

$$
\begin{aligned}
E(P_4) &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j} K^2\left(\frac{x-x_i}{h_2}\right) K^2\left(\frac{x_i-x_j}{h_1}\right) E(Y_i^2) E(Y_j^2) \\
&= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j} K^2\left(\frac{x-x_i}{h_2}\right) K^2\left(\frac{x_i-x_j}{h_1}\right) \\
&\times \quad \left[v(x_i)\, v(x_j) + v(x_j) m^2(x_i) + v(x_i) m^2(x_j) + m^2(x_i) m^2(x_j) \right]. \qquad (4.13)
\end{aligned}
$$

Again using independence of Y_i, Y_j and Y_k for $i \neq j \neq k$, we get

$$
\begin{aligned}
E(P_5) &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j \neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times \quad E(Y_i^2) E(Y_j) E(Y_k) \\
&= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j \neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times \quad \left[v(x_i)\, m(x_j)\, m(x_k) + m^2(x_i)\, m(x_j)\, m(x_k) \right]. \qquad (4.14)
\end{aligned}
$$

Finally, by the independence of Y_i, Y_j, Y_k and Y_d for $i \neq j \neq k \neq d$, we obtain

$$
\begin{aligned}
E(P_6) &= \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq k} K^2\left(\frac{x_i-x_k}{h_1}\right) K\left(\frac{x-x_i}{h_2}\right) \\
&\times \quad K\left(\frac{x-x_k}{h_2}\right) E(Y_i^2) E(Y_k^2) \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i \neq j \neq k}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) \\
&\times \quad K\left(\frac{x_k-x_i}{h_1}\right) E(Y_i^2) E(Y_j) E(Y_k)
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k\neq d}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_k}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right)E(Y_k^2)E(Y_i)E(Y_d)\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_j}{h_1}\right)E(Y_j^2)E(Y_i)E(Y_k)\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k\neq j\neq d}\sum\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right)E(Y_i)E(Y_j)E(Y_k)E(Y_d)\\
&= \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum_{i\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K^2\left(\frac{x_i-x_k}{h_1}\right)\\
&\times \quad \left[v(x_i)\,v(x_k)+v(x_k)\,m^2(x_i)+v(x_i)\,m^2(x_k)+m^2(x_i)\,m^2(x_k)\right]\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_i}{h_1}\right)\left[v(x_i)\,m(x_j)m(x_k)+m^2(x_i)m(x_j)m(x_k)\ \right]\\
&+ \quad \frac{1}{n^2(n-1)^2h^4}\sum\sum_{i\neq k\neq d}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_k}{h_1}\right)K\left(\frac{x_k-x_d}{h_1}\right)\\
&\times \quad \left[v(x_k)\,m(x_i)m(x_d)+m^2(x_k)m(x_i)m(x_d)\ \right]\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)K\left(\frac{x_k-x_j}{h_1}\right)\\
&\times \quad \left[v(x_j)\,m(x_i)m(x_k)+m^2(x_j)m(x_i)m(x_k)\ \right]\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k\neq j\neq d}\sum\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right)m(x_i)m(x_j)m(x_k)m(x_d). \qquad (4.15)
\end{aligned}
$$

So, the expected value of $(\hat{v}(x))^2$ is

$$
E((\hat{v}(x))^2)=E[P_1+P_2]+E(P_3)+E(P_4)+E(P_5)+E(P_6). \qquad (4.16)
$$

where $E[P_1 + P_2]$ and $E[P_a]$, for $a = 3, \cdots 6,$ are given in equations (4.11)-(4.15), respectively.

To complete the computation of the variance, using equation (4.6), $(E(\hat{v}(x)))^2$ equals

$$
\begin{aligned}
(E(\hat{v}(x)))^2 &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,\left[\,v^2(x_i) + m^4(x_i) + 2\,v(x_i)\,m^2(x_i)\,\right] \\
&+ \frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,v(x_i)\,v(x_k) \\
&+ \frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,m^2(x_i)\,m^2(x_k) \\
&+ \frac{2}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,v(x_i)\,m^2(x_k) \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,v(x_i)\,m(x_i)\,m(x_j) \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times\ [v(x_k)\,m(x_i)\,m(x_k)\,] \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_k-x_j}{h_1}\right) \\
&\times\ [v(x_i)\,m(x_k)\,m(x_j)] \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,m^3(x_i)\,m(x_j) \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_k-x_i}{h_1}\right)\,m^3(x_i)\,m(x_k) \\
&- \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_k-x_j}{h_1}\right) \\
&\times\ [m^2(x_i)\,m(x_k)\,m(x_j)] \\
&+ \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,m^2(x_i)\,m^2(x_j) \\
&+ \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum\sum_{i\neq j\neq k} K^2\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times\ m^2(x_i)\,m(x_j)\,m(x_k)
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K^2\left(\frac{x_i-x_k}{h_1}\right) m^2(x_i) m^2(x_k) \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) \\
&\times \quad K\left(\frac{x_k-x_i}{h_1}\right) m^2(x_i)\, m(x_j)\, m(x_k) \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k\neq d}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right) m^2(x_k) m(x_i) m(x_d) \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) \\
&\times \quad K\left(\frac{x_k-x_j}{h_1}\right) m^2(x_j) m(x_i) m(x_k) \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k\neq j\neq d}\sum\sum K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) \\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right) m(x_i) m(x_k) m(x_j) m(x_d). \qquad (4.17)
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
Var(\hat{v}(x)) \quad &= \quad \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right) \left[\mu_4(x_i) +\, 4\mu_3(x_i)\, m(x_i) +\, 4v(x_i)\, m^2(x_i) - v^2(x_i)\right] \\
&- \quad \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq j} K\left(\frac{x_i-x_j}{h_1}\right) K^2\left(\frac{x-x_i}{h_2}\right) \\
&\times \quad [\mu_3(x_i)\, m(x_j) + 2\, v(x_i)\, m(x_i) m(x_j)] \\
&- \quad \frac{2}{n^2\,(n-1)\,h_2^2\,h_1}\sum\sum_{i\neq k} K\left(\frac{x_k-x_i}{h_1}\right) K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) \\
&\times \quad [\mu_3(x_i)\, m(x_k) + 2\, v(x_i)\, m(x_i) m(x_k)] \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq j} K^2\left(\frac{x_i-x_j}{h_1}\right) K^2\left(\frac{x-x_i}{h_2}\right) \\
&\times \quad \left[\, v(x_i)\, v(x_j)\, + v(x_i)\, m^2(x_j) + v(x_j)\, m^2(x_i)\right] \\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2}\sum\sum_{i\neq k} K^2\left(\frac{x_i-x_k}{h_1}\right) K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x-x_k}{h_2}\right) \\
&\times \quad \left[v(x_i)\, v(x_k)\, + v(x_i)\, m^2(x_k) + v(x_k)\, m^2(x_i)\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i\neq j\neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)K\left(\frac{x_i-x_k}{h_1}\right)\\
&\times \quad v(x_i)\,m(x_j)m(x_k)\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i\neq k\neq j}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_i}{h_1}\right)\,v(x_i)\,m(x_j)\,m(x_k)\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i\neq k\neq d}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_k}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_d}{h_1}\right)\,v(x_k)\,m(x_i)\,m(x_d)\\
&+ \quad \frac{1}{n^2\,(n-1)^2\,h_2^2\,h_1^2} \sum\sum_{i\neq j\neq k}\sum K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times \quad K\left(\frac{x_k-x_j}{h_1}\right)\,v(x_j)\,m(x_i)\,m(x_k).
\end{aligned}
\tag{4.18}
$$

As noted before, since the mean function $m(x)$, the variance function $v(x)$, $\mu_3(x)$ and $\mu_4(x)$ are bounded, Lemmas 2.3.1 and 4.3.1 are used to approximate the summations in the above equation. Using Lemma 2.3.1 and then Taylor expansion, the approximation of the first term on the right hand side of the above equation is

$$
\begin{aligned}
&\frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,\left[\mu_4(x_i)+\,4\mu_3(x_i)\,m(x_i)+\,4v(x_i)\,m^2(x_i)-v^2(x_i)\right]\\
&\qquad = \frac{1}{n\,h_2^2}\int K^2\left(\frac{x-u}{h}\right)\,[\mu_4(u)\,+4\,\mu_3(u)\,m(u)\\
&\qquad + \,4\,v(u)\,m^2(u)\,-\,v^2(u)\big]\,du\,+\,O(n^{-2}\,h_2^{-2}).\\
&\qquad = \frac{1}{n\,h_2}\int K^2\,(y)\,dy\,\left[\mu_4(x)+4\mu_3(x)\,m(x)+4\,v(x)\,m^2(x)-\,v^2(x)\right]\\
&\qquad + o(n^{-1}\,h_2^{-1}),
\end{aligned}
\tag{4.19}
$$

where the last expression on the right hand side of equation (4.19) is derived using the substitution $\frac{x-u}{h}=y$. By following the same steps of the approximation above and using Lemma 4.3.1-(i), we can show that the second, third, fourth and fifth terms on the right hand side of equation (4.18) can be approximated, respectively, as follows:

$$\frac{-2}{n\,h_2}\int K^2(t)\,dt\,\left[\mu_3(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]+\,o(n^{-1}\,h_2^{-1}),\tag{4.20}$$

$$\begin{aligned}
&\frac{-2}{n\,h_2}\int\int K(t)\,K(y)K(t-\frac{h_1}{h_2}y)\,dy\,dt\,\left[\mu_3(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]+\,o(n^{-1}\,h_2^{-1})\\
&\quad=\frac{-2}{n\,h_2}\int\int K(t)\,K(y)\left[K(t)\,-\,\frac{h_1}{h_2}\,y\,K'(t)\,+\,\frac{h_1^2}{2!\,h_2^2}\,y^2\,K''(t)\,+\,o(\frac{h_1^2}{h_2^2})\right]dy\,dt\\
&\quad\times\,\left[\mu_3(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]\,+\,o(n^{-1}\,h_2^{-1})\\
&\quad=\frac{-2}{n\,h_2}\int K^2(y)\;dy\,\left[\mu_3(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]\,+\,o(n^{-1}\,h_2^{-1}),
\end{aligned}\tag{4.21}$$

$$\begin{aligned}
&\frac{1}{n^2\,h_2\,h_1}\int\int K^2(t)\,K^2(y)\,dy\,dt\left[v^2(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]\\
&\qquad+\,o(n^{-2}\,h_2\,h_1)=\,o(n^{-1}\,h_2^{-1})
\end{aligned}\tag{4.22}$$

and

$$\begin{aligned}
&\frac{1}{n^2\,h_2\,h_1}\int\int K^2(t+\frac{h_1}{h_2}y)\,K(y)\,K(t)\,dy\,dt\left[v_2(x)\,m(x)\,+\,2\,v(x)\,m^2(x)\right]\\
&\qquad+o(n^{-2}\,h_2\,h_1)=\,o(n^{-1}\,h_2^{-1}).
\end{aligned}\tag{4.23}$$

Using Lemma 4.3.1-(ii) and then applying Taylor series expansion, the approximation of the sixth term on the right hand side of equation (4.18) is

$$\frac{1}{n\,h_2}\int K^2(t)\,dt\,v(x)\,m^2(x)\,+\,o(n^{-1}\,h_2^{-1}).\tag{4.24}$$

By applying Lemma 4.3.1-(iii), the approximation of each one of the last three terms on the right hand side of equation (4.18) equals to

$$\frac{1}{n\,h_2}\int K^2(t)\,dt\,v(x)\,m^2(x)\,+\,o(n^{-1}\,h_2^{-1}).\tag{4.25}$$

Thus, the approximation of the last three terms in (4.18) is

$$\frac{3}{n\,h_2}\int K^2(t)\,dt\,v(x)\,m^2(x)\;+\;o(n^{-1}\,h_2^{-1}). \tag{4.26}$$

Thus, using equations (4.19)-(4.24) and equation (4.26), the variance of $\hat{v}(x)$ is

$$\begin{aligned} Var(\hat{v}(x)) &= \frac{1}{n\,h_2}\int K^2\,(y)\;dy\;\left[\mu_4(x)\;-\;v^2(x)\right]\;+\;o(n^{-1}\,h_2^{-1}) \\ &= C_2(x)\,n^{-1}\,h_2^{-1}\;+\;o(n^{-1}\,h_2^{-1}), \end{aligned} \tag{4.27}$$

where

$$C_2(x)\;=\;\int K^2\,(y)\;dy\;\times\;\left(\mu_4(x)\;-\;v^2(x)\right).$$

Thus, part (ii) in Theorem 4.2.1 is proved.

4.4 An Outline of Proof of the Mean Squared Error of the Brown and Levine Estimator in (4.3)

The form of the mean squared error of the Brown and Levine estimators is known in the literature, which was described in (1.17), but the exact deterministic functions C_is are unknown. For that, in this section, the bias and the variance of the local linear version of the Brown and Levine estimator in (4.3) is obtained where the order of differences is 2. The assumptions A1 and A2 and the assumptions E1, E2, E3 and E4 stated in sections 1.6 and 4.2, respectively, are assumed to be satisfied in the following analysis. First, note that the Brown and Levine estimator in (4.3) can be written as

$$\hat{v}_{BL}(x,1,h)\;=\;e^T\;\left(X_x^T\,W_x\,X_x\right)^{-1}\;X_x^T\,W_x\;\underline{\Delta}^2 \tag{4.28}$$

where

$$e^T \;=\; [\,1\;0\,],$$

$$\underline{\Delta}^2 = \;\left[0\;\Delta_2^2\;\;\cdots\;\;\Delta_{n-1}^2\;\;0\right]^T,$$

$$X_x = \begin{pmatrix} 0 & 0 \\ 1 & x_2 - x \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_{n-1} - x \\ 0 & 0 \end{pmatrix}$$

and

$$W_x = \;diag\left\{0, K\left(\frac{x_2 - x}{h}\right), \cdots, K\left(\frac{x_{n-1} - x}{h}\right), \,0\right\}.$$

Thus, to find the bias, it is required to calculate

$$E\{\hat{v}_{BL}(x,1,h)\} \;=\; e^T\;\left(X_x^T\,W_x\,X_x\right)^{-1}\;X_x^T\,W_x\,E\left(\underline{\Delta}^2\right) \tag{4.29}$$

So, observe that

$$\begin{aligned} E\left(\Delta_i^2\right) \;&=\; \sum_{j=0}^{r} d_j^2\;\left[\,v(x_{i+j-\lfloor r/2\rfloor}) \,+\, m^2(x_{i+j-\lfloor r/2\rfloor})\,\right] \\ &+\; \sum\sum_{j\neq k} d_j\,d_k\,m(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+k-\lfloor r/2\rfloor}). \end{aligned}$$

And then using the Taylor series expansion, we obtain

$$m(x_{i+j-\lfloor r/2\rfloor}) \;=\; m(x) \,+\, (x -\, x_{i+j-\lfloor r/2\rfloor})\,m'(x) \,+\, \frac{1}{2}\,(x -\, x_{i+j-\lfloor r/2\rfloor})^2\,m''(x) \,+\,\cdots$$

and

$$v(x_{i+j-\lfloor r/2\rfloor}) \;=\; v(x) \,+\, (x -\, x_{i+j-\lfloor r/2\rfloor})\,v'(x) \,+\, \frac{1}{2}\,(x -\, x_{i+j-\lfloor r/2\rfloor})^2\,v''(x) \,+\,\cdots.$$

When $x \in [x_{i+j-\lfloor r/2 \rfloor -1},\ x_{i+j-\lfloor r/2 \rfloor}]$, the distance between x and $x_{i+j-\lfloor r/2 \rfloor}$ can be written as $h\, u_{ij}$. Thus, by putting $x - x_{i+j-\lfloor r/2 \rfloor} = h\, u_{ij}$, $x - x_{i+k-\lfloor r/2 \rfloor} = h\, u_{ik}$ and using the last two equations in above, we get

$$\begin{aligned} E\left(\Delta_i^2\right) &= \sum_{j=0}^{r} d_j^2 \left[v(x) + h\, u_{ij}\, v'(x) + \frac{1}{2} h^2\, u_{ij}^2\, v''(x) + m^2(x) + 2\, h\, u_{ij}\, m(x)\, m'(x) \right. \\ &+ \left. h^2\, u_{ij}^2 \left(m(x)\, m''(x) + (m'(x))^2 \right) \right] \\ &+ \sum_{j \neq k}\sum d_j\, d_k \left[m^2(x) + h\, u_{ij}\, m(x)\, m'(x) + h\, u_{ik}\, m(x)\, m'(x) + h^2\, u_{ij}\, u_{ik}\, (m'(x))^2 \right. \\ &+ \left. \frac{1}{2} h^2\, u_{ij}^2\, m(x)\, m''(x) + \frac{1}{2} h^2\, u_{ik}^2\, m(x)\, m''(x) \right] + o(h^2). \end{aligned}$$

For $r = 2$ (second order's difference), the optimal difference sequence, which gives the best possible performance of the difference-based estimators, is 0.809, -0.5 and $-\,0.309$ as stated in Hall, Kay and Titterington (1990). Using this sequence, one can show that

$$E\left(\Delta_i^2\right) = \sum_{j=0}^{2} d_j^2 \left[v(x) + h\, u_{ij}\, v'(x) + \frac{1}{2} h^2\, u_{ij}^2\, v''(x) \right] + o(h^2) + O(n^{-1}).$$

Now, observe that

$$X_x^T\, W_x\, X_x = \begin{bmatrix} \sum\limits_{i=2}^{n-1} K_h(x_i - x) & \sum\limits_{i=2}^{n-1} (x_i - x)\, K_h(x_i - x) \\ \sum\limits_{i=2}^{n-1} (x_i - x)\, K_h(x_i - x) & \sum\limits_{i=2}^{n-1} (x_i - x)^2\, K_h(x_i - x) \end{bmatrix}.$$

Then, by putting $\hat{S}_r(x,h) = \frac{1}{n} \sum\limits_{i=2}^{n-1} (x_i - x)^r\, K_h(x_i - x)$, we get

$$\left(n^{-1}\, X_x^T\, W_x\, X_x \right)^{-1} = \frac{1}{\hat{S_2}(x,h)\, \hat{S_0}(x,h) - \hat{S_1}^{\,2}(x,h)} \begin{bmatrix} \hat{S}_2(x,h) & -\,\hat{S}_1(x,h) \\ -\,\hat{S}_1(x,h) & \hat{S}_0(x,h) \end{bmatrix}.$$

Now, note that

$$n^{-1}\, X_x^T\, W_x \begin{bmatrix} 0 \\ (x_2 - x)^2 \\ \vdots \\ (x_{n-1}\, -\, x)^2 \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{S}_2(x,h) \\ \hat{S}_3(x,h) \end{bmatrix}$$

and

$$\hat{S}_r(x,h) \;=\; h^r \int u^r\, K(u)\; du \;+\; O\,(n^{-1}).$$

Since K is symmetric around zero, we can prove that

$$\frac{1}{2}\, v''(x)\, e^T\, \left(X_x^T\, W_x\, X_x\right)^{-1}\, X_x^T\, W_x \begin{bmatrix} 0 \\ (x_2 - x)^2 \\ \vdots \\ (x_{n-1}\, -\, x)^2 \\ 0 \end{bmatrix}$$

$$= \frac{1}{2}\, v''(x)\, h^2 \int u^2\, K(u)\; du \;+\; O\,(n^{-1}) \;+\; o(h^2).$$

So, we obtain

$$\begin{aligned} E\{\hat{v}_{BL}(x, 1, h)\} \;&=\; e^T\, \left(X_x^T\, W_x\, X_x\right)^{-1}\, X_x^T\, W_x\, X_x \begin{bmatrix} v(x) \\ v'(x) \end{bmatrix} \\ &\quad + \frac{1}{2}\, v''(x)\, h^2 \int u^2\, K(u)\; du \;+\; O\,(n^{-1}) \;+\; o(h^2) \\ &= v(x) \;+\; \frac{1}{2}\, v''(x)\, h^2 \int u^2\, K(u)\; du \;+\; O\,(n^{-1}) \;+\; o(h^2). \end{aligned}$$

The second order kernel function is clearly used to find the bias in the above calculation.

Note that if rth order kernel is used, it is easy to verify that

$$E\{\hat{v}_{BL}(x,1,h) \;=\; v(x) \;+\; \frac{1}{r!}\,v^{(r)}(x)\,h^r \int u^r\,K(u)\;du \;+\; O\left(n^{-1}\right) \;+\; o(h^r)$$

where r is an even number. To find the variance of the Brown and Levine estimator in (4.3), we require to compute

$$\begin{aligned}
\mathrm{Var}(\,\Delta_i^2\,) \;&=\; \sum_{j=0}^{r} d_j^4\,\left[\,\mu_4(x_{i+j-\lfloor r/2\rfloor}) \;+\; 4\,\mu_3(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+j-\lfloor r/2\rfloor})\right.\\
&+\; \left.4\,v(x_{i+j-\lfloor r/2\rfloor})\,m^2(x_{i+j-\lfloor r/2\rfloor}) \;-\; v^2(x_{i+j-\lfloor r/2\rfloor})\,\right]\\
&+\; 2\sum\sum_{j\neq k} d_j^2\,d_k^2\,\left[\,v(x_{i+j-\lfloor r/2\rfloor})\,v(x_{i+k-\lfloor r/2\rfloor})\right.\\
&+\; \left.v(x_{i+j-\lfloor r/2\rfloor})\,m^2(x_{i+k-\lfloor r/2\rfloor}) \;+\; v(x_{i+j-\lfloor r/2\rfloor})\,m^2(x_{i+k-\lfloor r/2\rfloor})\,\right]\\
&+\; 4\sum\sum_{j\neq k} d_j^3\,d_k\;\mu_3(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+k-\lfloor r/2\rfloor})\\
&+\; 8\sum\sum_{j\neq k} d_j^3\,d_k\;v(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+k-\lfloor r/2\rfloor})\\
&+\; 4\sum\sum\sum_{j\neq k\neq e} d_j^2\,d_k\,d_e\,v(x_{i+j-\lfloor r/2\rfloor})\,m(x_{i+k-\lfloor r/2\rfloor})\,m(x_{i+e-\lfloor r/2\rfloor}).
\end{aligned}$$

Using Taylor series expansion and the following sequence of differences $0.809,\ -0.5$ and $-\,0.309$, we obtain

$$\mathrm{Var}(\,\Delta_i^2\,) \;=\; \frac{1}{2}\,\left(\,\mu_4(x) \;+\; v^2(x)\,\right) \;+\; O\,(h).$$

Similarly, we can show that

$$\mathrm{Cov}(\,\Delta_i^2,\,\Delta_j^2\,) = \begin{cases} 0.185\,\mu_4(x) \;-\; 0.435\,v^2(x) & \text{for } j = i-1, i+1;\\ 0.0625\,(\,\mu_4(x) \;-\; v^2(x)\,) & \text{for } j = i-2, i+2;\\ 0 & \text{for } |i-j| >\; 2 \end{cases}$$

Also note that

$$n^{-1}\, X_x^T\, W_x\, V\, W_x\, X_x \;=\; \begin{bmatrix} h^{-1}\, C_5(x)\, R(k) \,+\, o(h^{-1}) & O(n^{-1}) \\ O(n^{-1}) & h\, C_5(x)\, \mu_2(k^2) \,+\, O(n^{-1}) \end{bmatrix}$$

where $R(k) \;=\; \int K^2(y)\, dy,\;\; \mu_2(k^2) \;=\; \int y^2\, K^2(y)\, dy,\;\; C_5(x) \;=\; \mu_4(x) \,-\, v^2(x),\;\; V_i =$ $\mathrm{Var}(\,\Delta_i^2\,),\; C_{i,j} \;=\; \mathrm{Cov}(\,\Delta_i^2,\, \Delta_j^2\,)$ and

$$\mathbf{V} = \begin{bmatrix} 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ \vdots & V_2 & C_{2,3} & C_{2,4} & 0 & \ldots & \ldots & 0 \\ \vdots & C_{3,2} & \ddots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & C_{4,2} & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & C_{n-3,n-1} & \vdots \\ \vdots & \vdots & 0 & \ddots & \ddots & \ddots & C_{n-2,n-1} & \vdots \\ \vdots & \vdots & \vdots & & C_{n-1,n-3} & C_{n-1,n-2} & V_{n-1} & \vdots \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \end{bmatrix}.$$

Now, we can show that

$$\begin{aligned} \mathrm{Var}\{\hat{v}_{BL}(x,1,h)\,\} \;&=\; e^T\, \left(X_x^T\, W_x\, X_x\right)^{-1}\, X_x^T\, W_x\, V\, W_x\, X_x\, \left(X_x^T\, W_x\, X_x\right)^{-1}\, e \\ &=\; n^{-1}\, h^{-1}\, C_5(x)\, R(k) \,+\, o(n^{-1}\, h^{-1}). \end{aligned}$$

4.5 The MSE-Optimal Bandwidth

It is obvious from Theorem 4.2.1 that the performance of the new estimator for error variance function depends on the bandwidth h_2. As in the usual smoothing problem, when h_2 is small, the variance is large and the bias is small. However, if h_2 is large, the opposite occurs. So, the optimal choice of the bandwidth h_2 will be obtained by balancing the squared bias and the variance of the estimator $\hat{v}(x)$. An analytical discussion on the optimal selection of the bandwidth h_2 is provided in this section.

It is clear from Theorem 4.2.1 that the asymptotic mean squared error of the estimator (4.2) is

$$AMSE(\hat{v}(x)) \quad = \quad n^{-1}\, h_2^{-1}\, C_2(x) \, + \, h^{2\,r}\, C_6(x)$$

where $C_1(x)$ and $C_2(x)$ are defined in Theorem 4.2.1 and $C_6(x) \; = \; C_1^2(x)$. To find the asymptotic optimal bandwidth of the estimator in (4.2), we minimise $AMSE(\hat{v}(x))$ with respect to h_2. For that, consider h_2 to be a solution of

$$\frac{\partial\,(AMSE(\hat{v}(x)))}{\partial h_2} = 0.$$

So, we obtain

$$\frac{\partial\,(AMSE(\hat{v}(x)))}{\partial h_2} \; = \; -n^{-1} h^{-2}\, C_2(x) + \, 2\, r\, h^{2\,r-1}\, C_6(x) \, = 0, \tag{4.30}$$

and hence, it is easy to verify that the asymptotic MSE-optimal choice of the bandwidth h_2 is

$$h_{2-opt} = \left(\frac{C_2(x)}{2\, r\, C_6(x)} \right)^{\frac{1}{2r+1}} \; \times \; n^{-1/2r+1}\,.$$

One of the most important cases is for $r = 2$ (second order kernel) since the kernel function in this case is a probability density function. So, when second order kenel function is used, the asymptotic MSE-optimal choice of h_2 is

$$h_{2-opt} \; = \; \left(\frac{C_2(x)}{4\, C_6(x)} \right)^{\frac{1}{5}} \; \times \; n^{-1/5} \; \sim \; n^{-1/5}\,.$$

Thus, one can obtain the general formula of the asymptotic mean squared error corresponding to this asymptotic optimal choice of h_2 as

$$
\begin{aligned}
AMSE_{h_{2-opt}}(\hat{v}(x)) &= (2\,r)^{\frac{1}{2r+1}}\; n^{(-2\,r/2r+1)}\, C_2^{(2\,r/2r+1)}(x)\, C_6^{(1/2r+1)}(x) \\
&+ \left(\frac{1}{2\,r}\right)^{\frac{2\,r}{2r+1}}\; n^{(-2\,r/2r+1)}\, C_2^{(2\,r/2r+1)}(x)\, C_6^{(1/2r+1)}(x) \\
&= n^{(-2\,r/2r+1)} \times C_7(x) \qquad (4.31)
\end{aligned}
$$

where

$$
C_7(x) = \left[(2\,r)^{\frac{1}{2r+1}} + \left(\frac{1}{2\,r}\right)^{\frac{2\,r}{2r+1}}\right]\; C_2^{(2\,r/2r+1)}(x)\, C_6^{(1/2r+1)}(x).
$$

Using equation (4.31), we can show that the asymptotic mean squared error using a second order kernel function is

$$
AMSE_{h_{2-opt}}(\hat{v}(x)) \;=\; n^{(-4/5)}\left[1.65 \times C_2^{(4/5)}(x)\, C_6^{(1/5)}(x)\right].
$$

Brown and Levine (2007) have shown that the asymptotic optimal choice of the bandwidth h for their estimator is approximately $n^{\frac{-1}{2\,r+1}}$ and the mean squared error corresponding to this choice of the bandwidth is

$$
MSE_{h_{opt}}(\hat{v}_{B\&L}(x)) \;=\; n^{(-2\,r/2r+1)} \cdot C_8(x) + o(\,n^{(-2\,r/2r+1)}\,) \qquad (4.32)
$$

where $C_8(x)$ is a bounded function which depends only on x. So, for the case of $r = 2$, the asymptotic MSE-optimal bandwidth is $h \sim n^{\frac{-1}{5}}$ and the asymptotic mean squared error is

$$
AMSE_{h_{opt}}(\hat{v}_{B\&L}(x)) \;=\; n^{(-4/5)} \times C_9(x)
$$

where $C_9(x)$ equals to $C_8(x)$ when $r = 2$.

From equations (4.31) and (4.32), the asymptotic mean squared errors of the new estimator for the error variance function and the Brown and Levine estimator are clearly of the same order. The difference is only in constants at the point of estimation. Thus, in the finite sample, when $C_7(x) < C_8(x)$, the new estimator $\hat{v}(x)$ has a smaller mean squared error than the Brown and Levine estimator. So, the new estimator is expected to perform better than the Brown and Levine estimator in this case. However, in the infinite sample (for large n), the mean squared errors of these two estimators converge to zero at the same rate.

From equation (4.5) and for the second order of differences, we can conclude that the asymptotic optimal bandwidth and its corresponding mean squared error of the local linear version of Brown and Levine estimator is the same as that of the estimator in (4.2).

Chapter 5

The Asymptotic Normality of a New Estimator for the Error Variance Function

5.1 Introduction

In chapter 4, we carried out the mean square analysis of the estimator of the variance function that we proposed in the setting of the heteroscedastic nonparametric regression model. The mean squared error allows to establish that the new estimator for estimating the variance function is consistent and asymptotically unbiased. However, if we were to conduct a hypothesis test or to find a confidence interval for the unknown variance function, it is essential that we know the asymptotic distribution of the estimator for the variance function. Thus, the aim in the current chapter is to derive the asymptotic distribution of the new estimator in the setting of the heteroscedastic nonparametric regression model.

The current chapter is organised as follows. We restate the model in section 5.2. We also state the main theorem of this chapter together with the descriptive outline of the way results are derived in later section. Then, we provide proofs for the main results in section 5.3.

5.2 The Main Results

Recall that our model is

$$Y_i = m(x_i) + \sqrt{v(x_i)}\,\epsilon_i, \quad \text{for } \; i = 1, 2, \cdots, n \tag{5.1}$$

where $m(x_i)$ represents the unknown mean function $E(Y_i|x_i)$, Y_is denote the response variable, $v(x_i)$ represents the variance function and x_is denote the design points. The errors ϵ_is are assumed to be independent and identically distributed random variables with zero mean and unit variance.

For simplicity, x_is are assumed to have a fixed design points in the interval [0,1] such that $x_i = i/n$ for $i = 1, 2, \cdots, n$. In this setting, the estimator for $v(x)$ defined in (4.2) is

$$\begin{aligned}\hat{v}(x) &= \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_2}\right) Y_i^2 \\ &- \frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x - x_i}{h_2}\right) K\left(\frac{x_i - x_j}{h_1}\right) Y_i\,Y_j\,.\end{aligned}$$

Now, by adding and subtracting the following expression

$$\frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x - x_i}{h_2}\right) K\left(\frac{x_i - x_j}{h_1}\right) \left[\,2\,m(x_j)\,Y_i \,+\, m(x_i)\,m(x_j)\,\right],$$

in the above definition for $\hat{v}(x)$ and then rearranging the terms, we express $\hat{v}(x)$ as

$$\hat{v}(x) = \sum_{k=1}^{3} S_k\,(h_1,\, h_2) \tag{5.2}$$

where

$$\begin{aligned} S_1\,(h_1,\,h_2) \;&=\; \frac{1}{n}\sum_{i=1}^{n} Q_i \quad \& \\ Q_i \;&=\; \frac{1}{h_2}\,K\left(\frac{x-x_i}{h_2}\right)\,\left[Y_i \;-\; \frac{2}{(n-1)\,h_1}\sum_{j\neq i}K\left(\frac{x_i-x_j}{h_1}\right)m(x_j)\right]\,Y_i\,, \end{aligned} \tag{5.3}$$

$$\begin{aligned} S_2\,(h_1,\,h_2) \;&=\; \frac{-1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,U_iU_j \quad \& \\ U_i \;&=\; Y_i \;-\; m(x_i)\,, \end{aligned} \tag{5.4}$$

$$S_3\,(h_1,\,h_2) = \frac{1}{n(n-1)h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,m(x_i)m(x_j). \tag{5.5}$$

The asymptotic behaviour of each one of the above three terms is studied separately, and then these terms are combined to establish the asymptotic normality of the estimator $\hat{v}(x)$.

Clearly, the third term $S_3\,(h_1,\,h_2)$ is a deterministic in nature. So, it can be approximated easily using the standard lemma stated in section 4.3.

The second term $S_2\,(h_1,\,h_2)$ is a quadratic form in U_is. For simplicity, set $S_2(h_1\,,\,h_2) = \sum\sum_{1\le i<j\le n} S_{ij}$ where

$$S_{ij} \;=\; \frac{2}{n\,(n-1)\,h_1\,h_2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,U_i\,U_j. \tag{5.6}$$

For this quadratic term, we will verify that $S_2\,(h_1,\,h_2) \;=\; o_p\,(\,S_1\,(h_1,\,h_2)\,)$.

The first term $S_1\,(h_1,\,h_2)$ is sum of independent random variables and its asymptotic normality is provided in the following theorem. Note that the assumptions A1, A2 and the assumptions E1, E2, E3 and E4 used in the following theorem and corollary are stated in sections 1.6 and 4.2, respectively.

Theorem **5.2.1.** *Under the assumptions A1, A2, E1, E2, E3 and E4*

$$\sqrt{n\,h_2}\ \left(S_1\,(h_1\,,\,h_2) \,-\, C_1(x) \right) \ \xrightarrow{d} N\left(0\,,\,\left(\mu_4(x)\,-\,v^2(x)\right)\ \int K^2\,(y)\;dy\right).$$

where $C_1(x) = v(x) \,-\, m^2(x)\,.$

The following corollary, which gives the asymptotic normality of the $\hat{v}(x)$, follows from the above theorem

Corollary **5.2.1.** *Suppose the assumptions A1, A2, E1, E2, E3 and E4 are satisfied. Let* $h_2 \sim n^{-\alpha}$*, where* α *is a positive constant such that* $\alpha < 1$*, then*

$$\sqrt{n\,h_2}\ \left(\hat{v}(x) \,-\, v(x)\right) \ \xrightarrow{d} N\left(0\,,\,\left(\mu_4(x)\,-\,v^2(x)\right)\ \int K^2\,(y)\;dy\right).$$

It should be noted that if the assumption $E2'$ stated in section 4.2 is used instead of the assumption E2, the results of Theorem 5.2.1 and Corollary 5.2.1 are still true.

5.3 Proofs

The main aim of this section is to provide proofs for Theorem 5.2.1 and Corollary 5.2.1. Since the random variables Q_is defined in (5.3) are independent, we use the Lindeberg-Feller central limit theorem to derive the asymptotic distribution of the term $S_1\,(h_1,\,h_2)$ in subsection 5.3.1. Then, we prove Corollary 5.2.1 in subsection 5.3.2. To prove this corollary, we first provide an approximation to the deterministic term $S_3\,(h_1,\,h_2)$. Then, we analyse the term $S_2\,(h_1,\,h_2)$ to show that $S_2\,(h_1,\,h_2)$ is $o_p\left(S_1\,(h_1,\,h_2)\right)$. Using these results and Theorem 5.2.1, we establish the asymptotic normality of the new estimator for $v(x)$ defined in (5.2).

5.3.1 Proof of Theorem 5.2.1

The main purpose of this subsection is to prove Theorem 5.2.1. To prove this theorem, we need to find the expected value and the variance of the term $S_1(h_1\,,\,h_2)$. First, observe that

$$\begin{aligned} S_1(h_1\,,\,h_2) &= \frac{1}{n}\sum_{i=1}^{n} Q_i = \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)[Y_i - d_i]\,Yi \\ &= \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2 - \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) d_i\,Y_i \qquad (5.7) \end{aligned}$$

where $d_i = \frac{2}{(n-1)\,h_1}\sum_{j\neq i} K\left(\frac{x_i-x_j}{h_1}\right)\, m(x_j)$. Thus, we obtain

$$\begin{aligned} E\,(\,S_1(h_1\,,\,h_2)\,) &= \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\left[\,v(x_i) + m^2(x_i)\,\right] \\ &- \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) d_i\, m(x_i). \qquad (5.8) \end{aligned}$$

Note that since the mean function $m(x)$ and the variance function $v(x)$ are bounded, lemmas in sections 2.3 and 4.3 can be used to approximate the summations in (5.8). Using Lemma 2.3.1, we note that

$$\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\left[\,v(x_i) + m^2(x_i)\,\right] = v(x) + m^2(x) + O(\,h_2^r) + O(n^{-1}\,h_2^{-1}). \qquad (5.9)$$

For the second term in equation (5.8), using Lemma 4.3.1-(i), we get

$$\begin{aligned} &\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) d_i m(x_i) \\ &\quad = \frac{2}{n(n-1)h_1h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) m(x_i)m(x_j) \\ &\quad = 2\,m^2(x) + O(\,h_2^r) + O(n^{-1}\,h_2^{-1}), \qquad (5.10) \end{aligned}$$

where the last equality follows from Lemma 4.3.1-(i). Thus, we obtain

$$\begin{aligned} E\,(\,S_1(h_1\,,\,h_2)\,) \quad &= \quad v(x)\,-\,m^2(x)\,+\,O(\,h_2^r)\,+\,O(n^{-1}\,h_2^{-1}) \\ &\longrightarrow \quad v(x)\,-\,m^2(x)\,, \end{aligned} \tag{5.11}$$

as $n \to \infty$ and $h_2 \to 0$ such that $n\,h_2 \to \infty$. To compute the variance of the term $S_1(h_1\,,\,h_2)$, first note that

$$\begin{aligned} E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\right)^2 &= E\left[\frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right) Y_i^4\right. \\ &+ \left.\frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\ K\left(\frac{x-x_k}{h_2}\right) Y_i^2\,Y_k^2\right] \\ &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\left[\mu_4(x_i)+4\mu_3(x_i)\,m(x_i)\,+\,6\,v(x_i)\,m^2(x_i)\,+\,m^4(x_i)\right] \\ &+ \frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\ K\left(\frac{x-x_k}{h_2}\right)\,\big[\,v(x_i)\,v(x_k)\,+\,v(x_k)\,m^2(x_i) \\ &+ v(x_i)\,m^2(x_k)\,+\,m^2(x_i)\,m^2(x_k)\,\big]\,, \end{aligned} \tag{5.12}$$

where $\mu_r(x_i) = E[(Y_i - m(x_i))^r]$. Also consider

$$\begin{aligned} \left(E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\right)\right)^2 &= \left[\frac{1}{n\;h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\big(\,v(x_i)\,+\,m^2(x_i)\big)\right]^2 \\ &= \frac{1}{n^2\;h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,\left[v^2(x_i)+\,2\,v(x_i)\,m^2(x_i)\,+\,m^4(x_i)\right] \\ &+ \frac{1}{n^2\;h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\ K\left(\frac{x-x_k}{h_2}\right)\big[\,v(x_i)\,v(x_k)\,+\,v(x_k)\,m^2(x_i) \\ &+ \;v(x_i)\,m^2(x_k)\,+\,m^2(x_i)\,m^2(x_k)\,\big]\,. \end{aligned} \tag{5.13}$$

Now, use of (5.12), (5.13) and standard algebra gives,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,Y_i^2\right) &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right) \\ &\times\,\left[\mu_4(x_i)+4\mu_3(x_i)\,m(x_i)\,+\,4\,v(x_i)\,m^2(x_i)\,-\,v^2(x_i)\right].\end{aligned}$$

Using Lemma 2.3.1 in the right hand side of the above equation, we get

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,Y_i^2\right) &= \frac{1}{n\,h_2}\int K^2\,(\,y)\;dy \\ \times\,\left[\mu_4(x)\,-\,v^2(x)\,+\,4\,\mu_3(x)m(x)\,+4\,v(x)\;m^2(x)\right] &+ o(n^{-1}\,h_2^{-1}).\end{aligned} \tag{5.14}$$

To calculate the variance of $\left(\frac{1}{n\,h_2}\sum\limits_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right)$, first note that

$$\begin{aligned}E\left[\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right]^2 &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,d_i^2\,E\left(Y_i^2\right) \\ &+\,\frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,d_i\,E\left(Y_i\right)\,d_k\,E\left(Y_k\right) \\ &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,d_i^2\,\left(v(x_i)\,+\,m^2(x_i)\right) \\ &+\,\frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,d_i\,d_k\,m(x_i)\,m(x_k).\end{aligned} \tag{5.15}$$

Also, consider that

$$\begin{aligned}\left(E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right)\right)^2 &= \left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,m(x_i)\right)^2 \\ &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)\,d_i^2\,m^2(x_i) \\ &+\,\frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,d_i\,d_k\,m(x_i)\,m(x_k).\end{aligned} \tag{5.16}$$

Therefore, using equations (5.15) and (5.16), we get

$$\begin{aligned}
\text{Var}\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) d_i\,Y_i\right) &= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right) d_i^2\,v(x_i) \\
&= \frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right) K^2\left(\frac{x_i-x_j}{h_1}\right) m^2(x_j)\,v(x_i) \\
&+ \frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j\neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_i-x_k}{h_1}\right) \\
&\times v(x_i)\,m(x_j)\,m(x_k). \qquad (5.17)
\end{aligned}$$

Now, using Lemma 4.3.1-(i) and then Taylor series expansion, the first term on the right hand side of the above equation is

$$\begin{aligned}
&\frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right) K^2\left(\frac{x_i-x_j}{h_1}\right) m^2(x_j)\,v(x_i) \\
&= \frac{4}{n^2\,h_1\,h_2}\left(\int K^2(y)\;dy\right)^2 m^2(x)\,v(x) + o(n^{-2}\;h_1^{-1}\,h_2^{-1}) \\
&= o(n^{-1}\,h_2^{-1}).
\end{aligned}$$

By applying Lemma 4.3.1-(ii) to the second term on the right hand side of equation (5.17), we obtain

$$\begin{aligned}
&\frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j\neq k}\sum K^2\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) \\
&\times K\left(\frac{x_i-x_k}{h_1}\right) v(x_i)\,m(x_j)\,m(x_k) \\
&= \frac{4}{n\,h_2}\int K^2(y)\;dy\;m^2(x)\,v(x) + o(n^{-1}\,h_2^{-1}).
\end{aligned}$$

Therefore, it is clear that

$$\text{Var}\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) d_i\,Y_i\right) = \frac{4}{n\,h_2}\int K^2(y)\,dy\;m^2(x)\,v(x) + o(n^{-1}\,h_2^{-1}). \qquad (5.18)$$

Now to find the covariance between $\left(\frac{1}{n\,h_2}\sum\limits_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,Y_i^2\right)$ and $\left(\frac{1}{n\,h_2}\sum\limits_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right)$, observe that

$$
\begin{aligned}
\text{Cov}&\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)Y_i^2\,,\;\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)d_i\,Y_i\right)\\
&= E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)Y_i^2\;\times\;\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)d_i\,Y_i\right)\\
&-\left[E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)Y_i^2\right)\;\times\;E\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)d_i\,Y_i\right)\right]\\
&= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)d_i\,E(\,Y_i^3\,)\\
&+\frac{1}{n^2\,h_2^2}\sum\sum_{i\neq k} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,d_k\,E(Y_i^2)\,E(Y_k)\\
&-\left[\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,[\,v(x_i)\,+\,m^2(x_i)\,]\right)\right.\\
&\times\left.\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)\,d_i\,m(x_i)\right)\right].
\end{aligned}
$$

Then, by substituting expected values of $Y_k,\; Y_k^2$ and Y_k^3 and then simplifying, we get

$$
\begin{aligned}
\text{Cov}&\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)Y_i^2\,,\;\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)d_i\,Y_i\right)\\
&= \frac{1}{n^2\,h_2^2}\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{h_2}\right)d_i\,[\,\mu_3(x_i)\,+\,2\,v(x_i)\,m(x_i)\,]\\
&= \frac{2}{n^2\,(n-1)\,h_1\,h_2^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\\
&\times\,[\,\mu_3(x_i)\,m(x_j)\,+\,2\,v(x_i)\,m(x_i)\,m(x_j)\,]\,.
\end{aligned}
$$

Then, finally using Lemma 4.3.1-(i), we obtain

$$
\text{Cov}\left(\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)Y_i^2\,,\;\frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)d_i\,Y_i\;\right)
$$

$$= \frac{2}{nh_2} \int K^2(y)\ dy\ \left[\mu_3(x)\, m(x) + 2\, v(x) m^2(x)\right] + o(n^{-1}\, h_2). \quad (5.19)$$

Then using equations (5.14), (5.18) and (5.19), we have

$$\begin{aligned}
\text{Var}\,(\, S_1(h_1\,,\, h_2)\,) &= \text{Var}\left(\frac{1}{n\, h_2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_2}\right)\, Y_i^2 \right) \\
&+ \text{Var}\left(\frac{1}{n\, h_2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_2}\right)\, d_i\, Y_i \right) \\
&- 2\,\text{Cov}\left(\frac{1}{n\, h_2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_2}\right)\, Y_i^2\,,\, \frac{1}{n\, h_2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_2}\right)\, d_i\, Y_i \right) \\
&= \frac{1}{n\, h_2} \int K^2\,(\,y)\ dy\ \left[\mu_4(x)\, -\, v^2(x)\,\right]\, +\, o(n^{-1}\, h_2^{-1}). \qquad (5.20)
\end{aligned}$$

To use the Lindeberg-Feller central limit theorem, it is necessary to verify that the following condition holds,

$$\lim_{n \to \infty} \frac{1}{D_n^2} \sum_{i=1}^{n} E\, \left[(\, Q_i\, -\, E[Q_i]\,)^2\, I\, [\,|\, Q_i\, -\, E[Q_i]\,| >\, \tau\, D_n\,]\right] = 0 \qquad (5.21)$$

where $D_n^2 \;=\; \sum_{i=1}^{n} V_{Q_i}^2,\ V_{Q_i}^2$ is the variance of $Q_i\,,$ and τ is a positive number. To show the above condition holds, observe that

$$\begin{aligned}
E\,[\,Q_i\,] &= \frac{1}{h_2}\, K\left(\frac{x - x_i}{h_2}\right)\, \left[v(x_i)\, +\, m^2(x_i)\, -\, d_i\, m(x_i)\,\right] \\
&= \frac{1}{h_2}\, K\left(\frac{x - x_i}{h_2}\right)\, \Bigg[v(x_i)\, +\, m^2(x_i) \\
&- \frac{2}{(n-1)\, h_1} \sum_{i \neq j} K\left(\frac{x_i - x_j}{h_1}\right)\, m(x_i)\, m(x_j)\,\Bigg].
\end{aligned}$$

Using Lemma 2.3.1, the last expression simplifies to

$$E\,[\,Q_i\,] = \frac{1}{h_2}\, K\left(\frac{x - x_i}{h_2}\right)\, \left[v(x_i)\, -\, m^2(x_i)\,\right]\, +\, O\left(\frac{h_1^r}{h_2}\right). \qquad (5.22)$$

To compute the variance of Q_i, note that

$$\begin{aligned} \text{Var}\left(\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\right) &= \frac{1}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) \left[\mu_4(x_i) - v^2(x_i) + 4\,\mu_3(x_i)\, m(x_i)\right. \\ &+ \left. 4\, v(x_i)\, m^2(x_i)\right], \end{aligned} \tag{5.23}$$

and

$$\begin{aligned} \text{Var}\left(\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) d_i\, Y_i\right) &= \frac{1}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) d_i^2\, v(x_i) \\ &= \frac{v(x_i)}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) \left[\frac{4}{(n-1)^2\, h_1^2}\left(\sum_{i\neq j} K^2\left(\frac{x_i-x_j}{h_1}\right) m^2(x_j)\right.\right. \\ &+ \left.\left. \sum_{i\neq j\neq k}\sum K\left(\frac{x_i-x_j}{h_1}\right) K\left(\frac{x_i-x_k}{h_1}\right) m(x_j)\, m(x_k)\right)\right]. \end{aligned}$$

Using Lemmas 2.3.1 and 4.3.1-(i), respectively, in the last term on the right hand side of the above equation, we obtain

$$\text{Var}\left(\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) d_i\, Y_i\right) = \frac{4}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) v(x_i)\, m^2(x_i) + O(n^{-1}\, h_1^{-1}\, h_2^{-2}) \tag{5.24}$$

To complete the computation of the variance of Q_i, we need to compute covariance between $\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) Y_i^2$ and $\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) d_i\, Y_i$. For that observe

$$\begin{aligned} &\text{Cov}\left(\frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) Y_i^2\,,\, \frac{1}{h_2} K\left(\frac{x-x_i}{h_2}\right) d_i\, Y_i\right) \\ &= \frac{1}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) \left[\, d_i\, \mu_3(x_i) + 2\, d_i\, v(x_i)\, m(x_i)\,\right] \\ &= \frac{1}{h_2^2} K^2\left(\frac{x-x_i}{h_2}\right) \left[\frac{2}{(n-1)h_1}\sum_{i\neq j} K\left(\frac{x_i-x_j}{h_1}\right) m(x_j)\, \mu_3(x_i)\right. \\ &+ \left.\frac{4}{(n-1)h_1}\sum_{i\neq j} K\left(\frac{x_i-x_j}{h_1}\right) m(x_j)\, v(x_i)\, m(x_i)\right]. \end{aligned}$$

Then, using Lemma 2.3.1 and Taylor series expansion, we get

$$
\begin{aligned}
&\text{Cov}\left(\frac{1}{h_2}\,K\left(\frac{x-x_i}{h_2}\right)\,Y_i^2\,,\,\frac{1}{h_2}\,K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right)\\
&\quad=\frac{2}{h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\left[\,\mu_3(x_i)\,m(x_i)\;+\;2\,v(x_i)\,m^2(x_i)\,\right]\;+\;O\left(\frac{h_1^r}{h_2^2}\right)\\
&\quad+O(n^{-1}\,h_2^{-2}).
\end{aligned}
\tag{5.25}
$$

Then, using equations (5.23), (5.24) and (5.25), we obtain

$$
\begin{aligned}
\text{Var}\,(\,Q_i\,)\quad&=\quad V_{Q_i}\;=\;\frac{1}{h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\;\left[\mu_4(x_i)\;-\;v^2(x_i)\right]\\
&+\quad O\left(\frac{h_1^r}{h_2^2}\right)\;+\;O(n^{-1}\,h_2^{-2}).\\
&\to\quad\frac{1}{h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\;\left[\mu_4(x_i)\;-\;v^2(x_i)\right],
\end{aligned}
\tag{5.26}
$$

as $h_1,\,h_2\;\to\;0$ and $n\,h_2\;\to\;\infty$. Clearly from the above equation, the variance of Q_i depends on x_i. Now, to verify that the condition (5.21) holds, we also require to compute D_n^2. So, we find

$$
D_n^2\;=\;\sum_{i=1}^{n}\,V_{Q_i}=\frac{1}{h_2^2}\sum_{i=1}^{n}\;K^2\left(\frac{x-x_i}{h_2}\right)\;\left[\mu_4(x_i)\;-\;v^2(x_i)\right].
$$

And then using Lemma 2.3.1 and Taylor series expansion, we obtain

$$
\begin{aligned}
D_n^2\quad&=\quad\frac{n}{h_2^2}\int\;K^2\left(\frac{x-u}{h_2}\right)\;\left[\mu_4(u)\;-\;v^2(u)\right]\;du\\
&=\quad\frac{n}{h_2}\int\;K^2\,(y)\;dy\;\left[\mu_4(x)\;-\;v^2(x)\right]\;+\;o\left(\frac{n}{h_2}\right).
\end{aligned}
\tag{5.27}
$$

Now, it should be noted that

$$
Q_i\;-\;E(Q_i)\quad=\quad\frac{1}{h_2}\,K^2\left(\frac{x-x_i}{h_2}\right)\left[Y_i^2\;-\;d_i\,Y_i\;-\;v(x_i)\;+\;m^2(x_i)\right].
\tag{5.28}
$$

Using equations (5.27) and (5.28), it is clear that

$$
\begin{aligned}
& I\,[\,|\,Q_i \,-\, E[Q_i]\,| > \tau\, D_n\,] \\
& \quad = \; I\left[\,|\,Q_i \,-\, E[Q_i]\,| > \tau \sqrt{\frac{n}{h_2} \int K^2\,(y)\; dy\; [\mu_4(x) \,-\, v^2(x)]}\,\right] \\
& \quad = \; I\left[\,h_2\;\;|\,Q_i \,-\, E[Q_i]\,| > \tau \sqrt{n\, h_2 \int K^2\,(y)\; dy\; [\mu_4(x) \,-\, v^2(x)]}\;\right] \\
& \quad \longrightarrow \; 0,
\end{aligned}
$$

because $h_2 \;\; |\, Q_i - E[Q_i]\,| = K^2 \left(\frac{x - x_i}{h_2}\right) [Y_i^2 \,-\, d_i\, Y_i \,-\, v(x_i) \,+\, m^2(x_i)\,]$ is a finite random variable. So, it is bounded. However, $\tau \sqrt{n\, h_2 \int K^2\,(y)\; dy\; [\mu_4(x) \,-\, v^2(x)]} \;\rightarrow\; \infty$ as $n\, h_2 \;\rightarrow\; \infty$. This means that the indicator function is always zero as n becomes large enough. This implies that

$$
\begin{aligned}
& \frac{1}{D_n^2} \sum_{i=1}^{n} E\; \big[(\,Q_i \,-\, E[Q_i]\,)^2\, I\,[\,|\,Q_i \,-\, E[Q_i]\,| > \tau\, D_n\,]\big] \\
& \qquad \longrightarrow \; 0.
\end{aligned}
\tag{5.29}
$$

So, the condition (5.21) is satisfied. Thus, using Lindeberg-Feller central limit theorem and equations (5.11) and (5.20), we obtain

$$
\frac{\sqrt{n\, h_2}\,(\,S_1(h_1\,,\, h_2) \,-\, C_1(x))}{\sqrt{C_2(x)}} \;\xrightarrow{d}\; N(0,1)
$$

where

$$
\begin{aligned}
C_1(x) \;&=\; v(x) \,-\, m^2(x) \quad \text{and} \\
C_2(x) \;&=\; \int K^2\,(\,y)\; dy\; \big[\mu_4(x) \,-\, v^2(x)\,\big]\,.
\end{aligned}
$$

That is,

$$
\sqrt{n\, h_2}\;\;(\,S_1(h_1\,,\, h_2) \,-\, C_1(x)) \;\xrightarrow{d}\; N\,(0, C_2(x))\,. \tag{5.30}
$$

Therefore, the proof of Theorem 5.2.1 is completed.

5.3.2 Proof of Corollary 5.2.1

To prove Corollary 5.2.1, we first require to approximate the term $S_3\,(h_1,\, h_2)$ defined in equation (5.5). Thus, using Lemma 4.3.1-(i) and then Taylor series expansion, the approximation of the term $S_3\,(h_1,\, h_2)$ is

$$\begin{aligned}
&\frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,m(x_i)\,m(x_j)\\
&\qquad = m^2(x)\,+\,O(n^{-1}\,h_2^{-1})\,+\,O(h_2^r)\\
&\qquad \longrightarrow m^2(x), \qquad (5.31)
\end{aligned}$$

as $n\to\infty$, h_1 and $h_2\to 0$ such that $n\,h_1\,h_2\to\infty$. Now, our aim is to find the expected value and the variance of the term $S_2(h_1\,,\,h_2)$. First, recall that

$$\begin{aligned}
S_2(h_1\,,\,h_2) &= \frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,U_iU_j\\
&= \sum\sum_{1\le i<j\le n} S_{ij}
\end{aligned}$$

where $U_i\,=\,Y_i\,-\,m(x_i)$ and S_{ij} is defined in equation (5.6).

Also note that U_i and U_j are independent for $i\neq j$. In addition, it should be noted that the first two moments of U_i are as follows

$$E[U_i] = E[Y_i] - m(x_i) = 0 \qquad (5.32)$$

and

$$E[U_i^2] = E\,[Y_i - m(x_i)]^2\,=\,v(x_i) \qquad (5.33)$$

Using equation (5.32) and by the independence of U_i and U_j for $i \neq j$, the expected value of $S_2(h_1\,,\,h_2)$ is

$$\begin{aligned} E\,(\,S_2(h_1\,,\,h_2)\,) &= E\left(\frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,U_i\,U_j\right) \\ &= \frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,E[U_i]E[U_j] \\ &= 0. \end{aligned} \tag{5.34}$$

Using (5.32) and (5.33) as well as the independence of U_i, U_j, U_k and U_d for $i \neq j \neq k \neq d$, one can see that

$$\begin{aligned} E[S_{ij}^2] &= E\left[\frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,U_i^2\,U_j^2\right] \\ &= \frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,E[\,U_i^2\,]\,E[\,U_j^2\,] \\ &= \frac{4\,v(x_i)\,v(x_j)}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K^2\left(\frac{x-x_i}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right), \end{aligned} \tag{5.35}$$

$$\begin{aligned} E[S_{ij}\,S_{ji}] &= E\left[\frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_j}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,U_i^2\,U_j^2\right] \\ &= \frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_j}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,E[\,U_i^2\,]\,E[\,U_j^2\,] \\ &= \frac{4\,v(x_i)\,v(x_j)}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_j}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right) \end{aligned} \tag{5.36}$$

and

$$\begin{aligned} E[\,S_{ij}\,S_{ki}\,] &= E\left[\frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\right. \\ &\times \left. K\left(\frac{x_k-x_i}{h_1}\right)\,U_i^2\,U_j\,U_k\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{4}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\,K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_k}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,K\left(\frac{x_k-x_i}{h_1}\right)\\
&\times \; E[\,U_i^2\,]\,E[\,U_j\,]\,E[\,U_k\,]\\
&= \; 0. \qquad (5.37)
\end{aligned}
$$

Similarly, we can show that

$$
E[\,S_{ij}\,S_{ik}\,] \,=\, E[\,S_{ij}\,S_{kj}\,] \,=\, E[\,S_{ij}\,S_{jd}\,] \,=\, E[\,S_{ij}\,S_{kd}\,] \,=\, 0. \qquad (5.38)
$$

Now, observe that

$$
\begin{aligned}
E\left[\,S_2^2(h_1\,,\,h_2)\,\right] &= E\left[\sum\sum_{1\le i<j\le n} S_{ij}\right]^2 = E\left[\frac{1}{2}\sum\sum_{i\neq j} S_{ij}\right]^2\\
&= \frac{1}{4}\left[\sum\sum_{i\neq j} E\,[\,S_{ij}^2\,] + \sum\sum_{i\neq j} E\,[\,S_{ij}\;S_{ji}] + \sum\sum\sum_{i\neq j\neq k} E\,[\,S_{ij}\;S_{ki}\,]\right.\\
&+ \;\sum\sum\sum_{i\neq j\neq k} E\,[\,S_{ij}\;S_{ik}\,] + \sum\sum\sum_{i\neq j\neq k} E\,[\,S_{ij}\,S_{kj}\,]\\
&+ \left.\;\sum\sum\sum_{i\neq j\neq d} E\,[\,S_{ij}\,S_{jd}\,] + \sum\sum\sum\sum_{i\neq j\neq k\neq d} E\,[\,S_{ij}\,S_{kd}\,]\right].
\end{aligned}
$$

Using (5.35), (5.36), (5.37) and (5.38), we get

$$
\begin{aligned}
E\left[\,S_2^2\,(h_1\,,\,h_2)\,\right] &= \frac{1}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)\,K^2\left(\frac{x_i-x_j}{h_1}\right)\,v(x_i)\,v(x_j)\\
&+ \;\frac{1}{n^2\,(n-1)^2\,h_1^2\,h_2^2}\sum\sum_{i\neq j} K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x-x_j}{h_2}\right)\\
&\times \; K^2\left(\frac{x_i-x_j}{h_1}\right)\,v(x_i)\,v(x_j).
\end{aligned}
$$

Since the variance function $v(x)$ is bounded, using Lemma 4.3.1-(i) and then using Taylor series expansion, one can show that

$$\begin{aligned}
E\left[\, S_2^2\,(h_1\,,\,h_2)\,\right] \;&=\; \frac{v^2(x)}{n^2\,h_1^2\;h_2^2}\,\left[\int\!\!\int K^2\left(\frac{x-u_1}{h_2}\right)\,K^2\left(\frac{u_1-u_2}{h_1}\right)\,du_1\,du_2\right.\\
&+\;\left.\int\!\!\int K\left(\frac{x-u_1}{h_2}\right)\,K\left(\frac{x-u_2}{h_2}\right)\,K^2\left(\frac{u_1-u_2}{h_1}\right)\,du_1\,du_2\,+\,O(n^{-1}\,h_1)\right]\\
&=\;\frac{2\,v^2(x)}{n^2\,h_1\;h_2}\int K^2\,(t)\;dt\int K^2\,(y)\;dy\,+\,o(n^{-2}\,h_1^{-1}\,h_2^{-1}). \qquad (5.39)
\end{aligned}$$

Note that $O(n^{-3}\,h_1^{-1}\,h_2^{-2}) \,=\, o(n^{-2}\,h_1^{-1}\,h_2^{-1})$ under the assumption E4. Then, using (5.34) and (5.39), we obtain

$$\begin{aligned}
\text{Var}\,(\,S_2(h_1\,,\,h_2)\,) \;&=\; E\left[\,S_2^2\,(h_1\,,\,h_2)\,\right]\,-\,[E(S_2(h_1\,,\,h_2)\,)]^2\\
&=\;\frac{2\,v^2(x)}{n^2\,h_1\,h_2}\left(\int K^2\,(t)\;dt\right)^2+o(n^{-2}\,h_1^{-1}\,h_2^{-1}). \qquad (5.40)
\end{aligned}$$

Now, our goal is to compute the covariance between $S_k(h_1\,,\,h_2)$'s. Since the term $S_3(h_1\,,\,h_2)$ is a deterministic function, we get

$$\text{Cov}\,[\,S_1(h_1\,,\,h_2),\,S_3(h_1\,,\,h_2)]\,=\,\text{Cov}\,[\,S_2(h_1\,,\,h_2),\,S_3(h_1\,,\,h_2)]\,=\,0.$$

Equation (5.34) implies that

$$E(\,S_1(h_1\,,\,h_2)\,)\cdot\,E(\,S_2(h_1\,,\,h_2)\,)\,=\,0.$$

Also note that, using the independence of Y_i, Y_j and Y_k for $i \neq j \neq k$, we obtain

$$\begin{aligned}
&\text{Cov}\,[\,S_1(h_1\,,\,h_2),\,S_2(h_1\,,\,h_2)] = \,E(\,S_1(h_1\,,\,h_2)\,\cdot\,S_2(h_1\,,\,h_2)\,)\\
&\quad-\,E(\,S_1(h_1\,,\,h_2)\,)\cdot\,E(\,S_2(h_1\,,\,h_2)\,)\\
&\quad=\,E\left[\left(\frac{1}{n\,h_2}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h_2}\right)\,Y_i^2\,-\,\frac{1}{n\,h_2}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h_2}\right)\,d_i\,Y_i\right)\right.\\
&\quad\times\,\left.\left(\frac{1}{n(n-1)h_1h_2}\sum\sum_{i\neq j}K\left(\frac{x-x_i}{h_2}\right)\,K\left(\frac{x_i-x_j}{h_1}\right)\,U_i\,U_j\right)\right]
\end{aligned}$$

$$= E\left[\frac{2}{n^2(n-1)h_1h_2^2}\sum\sum_{i\neq j} K^2\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\left(Y_i^3Y_j - Y_i^3m(x_j)\right.\right.$$

$$-Y_i^2m(x_i)Y_j + Y_i^2m(x_i)m(x_j) - d_iY_i^2Y_j + d_iY_i^2m(x_j)$$

$$\left.\left. + d_iY_i^2m(x_i)Y_j - d_iY_im(x_i)m(x_j)\right)\right]$$

$$+E\left[\frac{1}{n^2(n-1)h_1h_2^2}\sum\sum\sum_{i\neq j\neq k} K\left(\frac{x-x_i}{h_2}\right)K\left(\frac{x-x_k}{h_2}\right)K\left(\frac{x_i-x_j}{h_1}\right)\right.$$

$$\times\left(Y_k^2Y_iY_j - Y_k^2Y_im(x_j) - Y_k^2Y_jm(x_i) + Y_k^2m(x_i)m(x_j)\right.$$

$$\left.\left. -d_kY_kY_iY_j + d_kY_kY_im(x_j) + d_kY_kY_jm(x_i) - d_kY_km(x_i)m(x_j)\right)\right]$$

$$= 0.$$

Thus, the three terms $S_k(h_1, h_2)$'s are uncorrelated. Using equation (5.30), the term $S_1(h_1, h_2)$ can be written as

$$S_1(h_1, h_2) = \frac{1}{\sqrt{n h_2}}\left[(\mu_4(x) - v^2(x))\int K^2(y)\,dy\right]^{1/2} N_1 + v(x) - m^2(x) \quad (5.41)$$

where the random variable N_1 is standard normal distribution. From equations (5.34) and (5.40), it is clear that the expected value of the term $S_2(h_1, h_2)$ is zero and its variance is $o(n^{-1}h_2^{-1})$. Therefore, we obtain

$$\sqrt{n h_2}\,(\,S_2(h_1, h_2)\,) \xrightarrow{P} 0. \quad (5.42)$$

Thus, the effect of this term in the main distribution of $\hat{v}(x)$ is negligible compared to the effect of the term $S_1(h_1, h_2)$. Then, by equation (5.31) and the above representation of the term $S_1(h_1, h_2)$, the new estimator $\hat{v}(x)$ can be expressed as

$$\begin{aligned}\hat{v}(x) &= \frac{1}{\sqrt{n h_2}}((\mu_4(x) - v^2(x))\int K^2(y)\,dy)^{1/2}N_1 \\ &+ v(x) - m^2(x) + m^2(x) + O(n^{-1}h_2^{-1}) + O(h_2^r).\end{aligned}$$

Therefore, we get

$$\hat{v}(x) - v(x) \quad = \quad \frac{1}{\sqrt{n\, h_2}} \left[(\mu_4(x) - v^2(x)) \int K^2\,(y)\; dy \right]^{1/2} N_1 + O(n^{-1}\, h_2^{-1}) + O(h_2^r).$$

Thus, it is obvious that

$$\begin{aligned} \sqrt{n\, h_2}\; (\hat{v}(x) - v(x)\,) \quad &= \quad \left[(\mu_4(x) - v^2(x)) \int K^2\,(y)\; dy \right]^{1/2} N_1 \\ &+ \quad O(\, n^{-1/2}\, h_2^{-1/2}\,) + O(\, n^{1/2}\, h_2^{r+1/2}\,). \end{aligned}$$

This implies that

$$\sqrt{n\, h_2}\; (\hat{v}(x) - v(x)\,) \xrightarrow{d} N\left(0\,,\, (\mu_4(x) - v^2(x)) \int K^2\,(y)\; dy \right). \qquad (5.43)$$

Chapter 6

Simulation Studies: Finite Sample Behaviour of a New Estimator for the Error Variance Function

6.1 Introduction

In chapters 4 and 5, we carried out analytical study of the proposed estimator for the error variance function. Since the finite sample performance is one of the most important aspects in the assessment of the goodness of any estimator, in the this chapter, our interest is to investigate the finite sample performance of the new estimator in the setting of the heteroscedastic nonparametric regression model. So, we will investigate the effect of the mean function, the bandwidth used to estimate the mean function, and the bandwidth for estimating the variance function on the finite sample performance of the proposed variance function estimator. To study the effect of the mean function, we fix the variance function and we then use different mean functions to create the regression models. After that, we estimate the variance function in each case to explore the effect of the mean function where the bandwidths are chosen appropriately. We shall compare this estimated variance function with the true variance function in order to assess the goodness of the estimated variance function. We also provide plots of the variance and the mean squared error of $\hat{v}(x_i)$s in each case of the simulation studies.

A general structure to study the effect of the mean function and the bandwidths on the finite sample performance of the new estimator for the error variance function are given in section 6.2 . In section 6.3, we assess the effect of the mean function on the finite sample performance of the new estimator for the error variance function. For that, we select different mean functions and then each mean function is examined with several variance functions where the bandwidths are chosen appropriately. We particularly consider six different mean functions for each variance function we estimate.

The estimator for the error variance function considered here has two different bandwidths. The first one denoted by, h_1, is used to estimate the mean function, whereas the other, h_2, is used for estimating the variance function. Thus, the selections of these bandwidths have an effect on the finite sample performance of the new estimator. So, the influence of both bandwidths is studied in section 6.4. In particular, in subsection 6.4.1, the effect of the bandwidth h_1 on the finite sample performance of the new estimator is investigated. To study this effect, we fix the mean function, the variance function and the bandwidth h_2 and then the bandwidth h_1 is allowed to vary. After that, the effect of the bandwidth h_2 on the finite sample performance of the new estimator is studied in subsection 6.4.2 . To explore the effect of the bandwidth h_2 through simulation, we fix the mean function, the variance function and the bandwidth h_1, after which several values of the bandwidth h_2 are examined with these fixed choices.

We also employ the proposed estimator of the variance function in a real data set in section 6.5. The description of this real data set is given in subsection 6.5.1. Then, the estimation of the variance function is explained in subsection 6.5.2.

6.2 The General Structure of the Simulation Studies

In all simulation studies of this chapter, the regression model is $Y_i = m(x_i) + \sqrt{v(x_i)}\,\epsilon_i$, where

F1. the design points $x_1, x_2, \cdots, x_n$ are independent and identically distributed uniform $[0,1]$ random variables,

F2. x_is are independent of ϵ_is and

F3. the errors $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independent and identically distributed random variables from standard normal distribution.

As in the setting of the homoscedastic nonparametric regression model, for a given sample of size n, we first choose randomly x_is and ϵ_is. After that, sort x_is from the smallest to the the largest values and then Y_is are generated using the model $Y_i = m(x_i) + \sqrt{(v(x_i))}\,\epsilon_i$ for selected m and v functions. The observed values of the new estimator $\hat{v}_{new}(x)$ are computed using the following equation,

$$\begin{aligned} \hat{v}_{new}(x) &= \frac{1}{n\,h_2}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right) Y_i^2 \\ &- \frac{1}{n\,(n-1)\,h_1\,h_2}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{x-x_i}{h_2}\right) K\left(\frac{x_i-x_j}{h_1}\right) Y_i Y_j\,. \end{aligned} \tag{6.1}$$

where the bandwidths h_1 and h_2 are suitably selected. Note that the kernel function K is chosen to be the standard normal probability density function as was the case in the simulation studies of the constant variance model in chapter 3. This procedure is replicated N times. Therefore, for each selection of a variance function and a mean function and bandwidths h_1 and h_2, there are N observed values for each $\hat{v}_{new}(x_i)$s. Then, the mean of each $\hat{v}_{new}(x_i)$s is computed. Therefore, to present the estimated variance function, the mean values of $\hat{v}_{new}(x_i)$s are plotted versus the chosen x_is. The true variance function is plotted in all considered cases of the simulation studies in order to examine the goodness of the estimated functions. We also plot the corresponding variance and mean squared error versus x_is. The estimated variance functions by Brown and Levine estimators (2007) and their corresponding variance and mean squared error are also plotted in figures in sections 6.3.

Remarks:

1) The distance from $x_{(i)}$ to $x_{(i+1)}$ is approximately $O(n^{-1})$. Therefore, the results of the simulation studies are still valid for the equally spaced design points.

2) The main Matlab functions for all the figures in the current chapter are provided in appendix C.

3) As noted in section 2.1, it is obvious that

$$\begin{aligned} K\left(\frac{x_i - x_j}{h}\right)/(n-1)\,h &= K\left(\frac{x_i - x_j}{h}\right)/(n-1)\,h\,f(x_i) \\ &\approx \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum\limits_{i \neq j} K\left(\frac{x_i - x_j}{h}\right)} \end{aligned}$$

where $f(x_i)$ is the density function of the uniform distribution. Thus, it is easy to verify that the estimator (6.1) can be written as

$$\hat{v}_{new}(x) = \sum_{i=1}^{n} \frac{K\left(\frac{x-x_i}{h_2}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)} \left[Y_i^2 - \sum_{j \neq i} \frac{K\left(\frac{x_i - x_j}{h_1}\right) Y_i Y_j}{\sum\limits_{i \neq j} K\left(\frac{x_i - x_j}{h_1}\right)} \right]. \tag{6.2}$$

The above definition of the new estimator is used in all simulation studies in the current chapter. The main reason to use this definition is that the summation of weights given to each row is one. In other words, $\sum\limits_{j=1}^{n} w_{ij} = 1$ for each i where $w_{ij} = K\left(\frac{x-x_i}{h_2}\right) / \sum\limits_{i=1}^{n} K\left(\frac{x-x_i}{h_2}\right)$.

4) Recall that, the local linear version of the Brown and Levine estimator is

$$\hat{v}_{BL}(x,1,h) = \frac{1}{n\,h}\sum_{i=1}^{n} \frac{\{\hat{S}_2(x,h) - \hat{S}_1(x,h)\,(x_i - x)\}\,K\left(\frac{x-x_i}{h}\right)\,\Delta_i^2}{\hat{S}_2(x,h)\,\hat{S}_0(x,h) - \hat{S}_1^2(x,h)} \tag{6.3}$$

where $\Delta_i = \sum_{k=0}^{m} d_j\; y_{i+k-\lfloor m/2 \rfloor}$, for $i = \lfloor m/2 \rfloor + 1, ..., n + \lfloor m/2 \rfloor - m$; $\hat{S}_r(x,h) = \frac{1}{n}\sum_{i=2}^{n-1} (x_i - x)^r\, K_h(x_i - x)$; $\lfloor a \rfloor$ represents the largest integer number less than a and m denotes the order of differences, while d_js represents the difference sequence.

5) For the Brown and Levine estimator, we choose three orders of differences. In particular, these orders are $2,\, 4$ and 6. Note that we use the optimal difference sequence given in Hall, Kay and Titterington (1991) for the chosen orders. These sequences are given in the table (6.1).

Orders r	The optimal difference sequences $(d_0, \cdots d_r)$
2	(0.8090, −0.5, −0.309)
4	(0.2708, −0.0142, 0.6909, −0.4858, −0.4617)
6	(0.24, 0.03, −0.0342, 0.7738, −0.3587, −0.3038, −0.3472)

Table 6.1: The optimal difference sequences for the orders 2, 4 and 6

For all of the above optimal difference sequences, the following condition is satisfied,

$$\sum_{i=0}^{r} d_i = 0 \;\text{ and }\; \sum_{i=0}^{r} d_i^2 = 1. \tag{6.4}$$

In all simulation studies in sections 6.3 and 6.5, the local linear version of the Brown and Levine estimator is used with these three orders of differences. For more details, please see Hall, Kay and Titterington (1991) and Brown and Levine (2007).

6.3 The Finite Sample Performance of the Estimator in (6.2): The Effect of the Mean function

The main goal of this section is to study the effect of the mean function on the finite sample performance of the proposed estimator defined in (6.2). The description of the models and specific structure of the simulation studies are provided in subsection 6.3.1, whereas the main results and the discussion are given in subsections 6.3.2 and 6.3.3, respectively.

6.3.1 The Description of the Models and Specific Structure of the Simulation Studies

To study the effect of the mean function on the finite sample performance of the estimator in (6.2), we use the six mean functions, which were described in section 3.3. We have four different variance functions for each of the mean functions. So, in total, we generate data from 24 regression models. The chosen variance functions are:

$$\begin{aligned} i)\ v_1(x) &= 3 + 2x. \\ ii)\ v_2(x) &= 0.5\left(2 + 4x - 4x^2 + 3x^3\right). \\ iii)\ v_3(x) &= \exp(-4 - 5x^2). \\ iv)\ v_4(x) &= \mid 0.25\cos(\pi x) \mid . \end{aligned} \tag{6.5}$$

Note that the bandwidth h_1, which is used to estimate the mean function, is chosen to minimise the bias $E[\hat{m}(x)] - m(x)$, whereas the bandwidth h_2, which is used to estimate the variance function is selected such that the mean squared error of the variance function estimator is minimised. Thus, optimal choices of the bandwidths h_1 and h_2 are used in the simulation studies in this section. This point will be clarified in the following section. To illustrate the performance of the estimator in (6.2), we will provide plots of the estimated and the true variance functions. We also plot the estimated variance functions by the Brown and Levine estimators on the same graph to illustrate relative performance of the proposed estimator in (6.2) and the Brown and Levine estimators. Furthermore, we present the plots of the variance and the mean squared error of $\hat{v}(x_i)$s for both of the estimators.

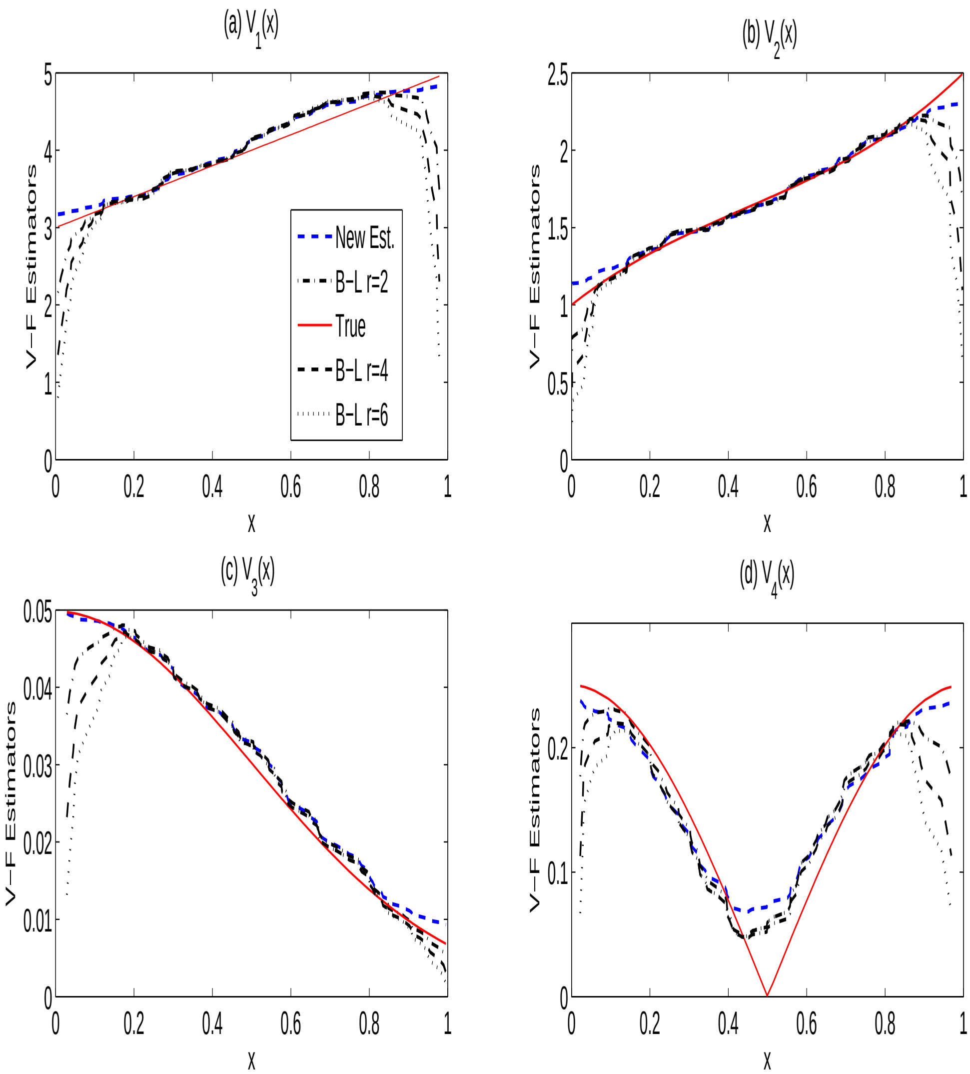

Figure 6.1: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_1(x)$ *, (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

To start with, consider the following regression model

$$\begin{aligned} Y_i &= m_1(X_i) + \sqrt{v_1(X_i)}\,\epsilon_i \\ &= 1 + \sqrt{3 + 2\,x_i}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n. \end{aligned} \tag{6.6}$$

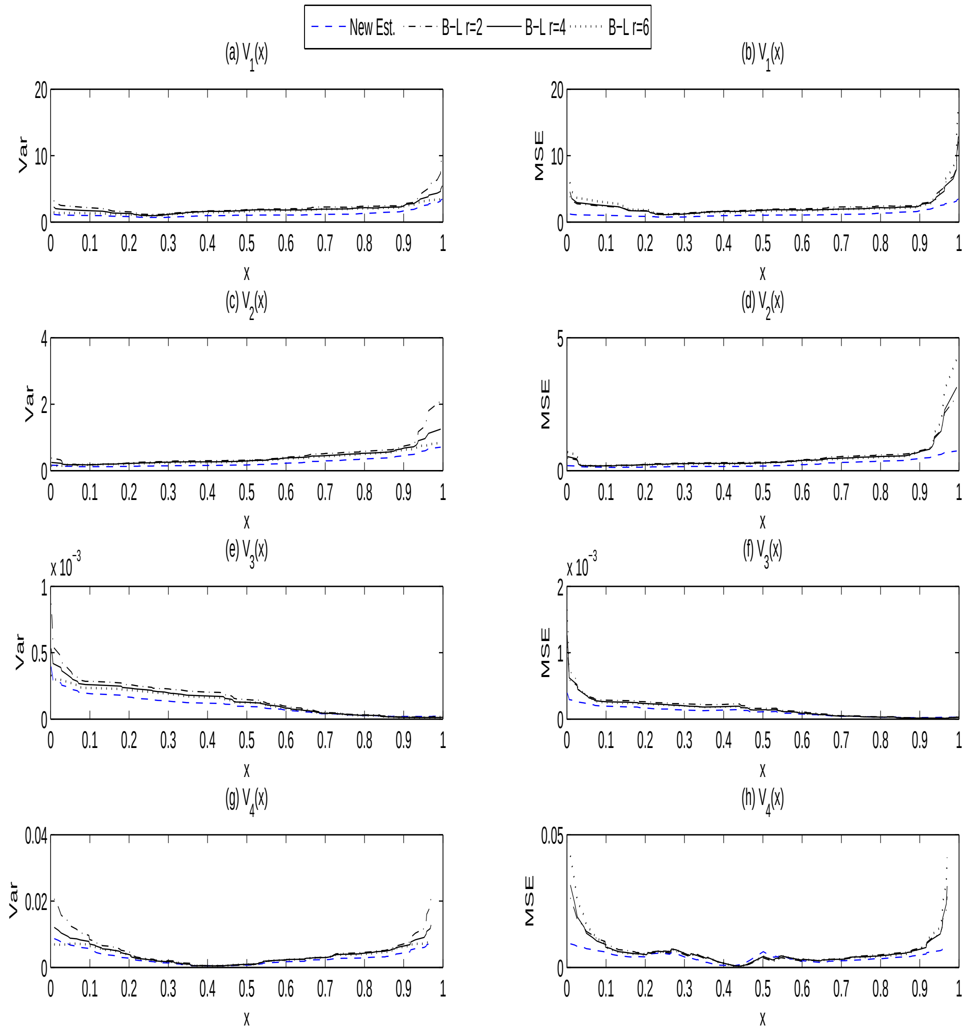

Figure 6.2: *The Variances and Mean Squared errors of the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.1), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

We first select randomly x_is from the uniform distribution $U[0,1]$ and ϵ_is from the $N(0,1)$ distribution, where the sample size is $n = 100$. After that, x_is are sorted into ascending order. Then, Y_is are generated using the model (6.6). The bandwidths h_1 and h_2 of the estimator in (6.2) are taken as 0.025 and 0.1, respectively. The bandwidth h_2 of the Brown and Levine estimators is chosen to be 0.06. Then, for the selected bandwidths, $\hat{\upsilon}_{New}(x)$s and $\hat{\upsilon}_{BL}(x)$s are calculated using the equations (6.2) and (6.3), respectively. The previous steps are then replicated for $N = 1000$ times. The mean of each of $\hat{v}_{New}(x)$s and $\hat{v}_{BL}(x)$s are computed. Using the means of $\hat{v}_{New}(x)$s and $\hat{v}_{BL}(x)$s, we plot the x_is versus the means of the $\hat{v}_{New}(x)$s and $\hat{v}_{BL}(x)$s in the plot (a) in the figure (6.1). For comparison, the true variance function defined in (i) in (6.5) is also plotted in the same figure. We also plot the variance and mean squared error of the estimators in the plots (a) and (b) in the figure (6.2). Using the same mean function, we repeat the above steps by taking $v(x)$ to be $v_2(x)$, $v_3(x)$ and $v_4(x)$. Thus, to produce the plots (b), (c) and (d) in the figure (6.1), we repeat the above steps for the following three models, respectively,

$$\begin{aligned}
(i)\, Y_i &= 1 + \sqrt{0.5\ (2\ + 4\, x_i\ -\ 4\, x_i^2\ +\ 3\, x_i^3\,)}\, \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n. \\
(ii)\, Y_i &= 1 + \sqrt{\exp\,(-4 -\ 5\, x_i^2)}\, \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n. \\
(iii)\, Y_i &= 1 + \sqrt{|\ 0.25 \cos\,(\pi\, x_i)\ |}\, \epsilon_i, \quad \text{for } i = 1, 2, \cdots, n.
\end{aligned}$$

Again, in all these three plots (b), (c) and (d) in the figure (6.1), the same bandwidth h_1 is used to estimate the mean function ($h_1 = 0.025$). The bandwidth h_2 of the estimator in (6.2) and the Brown and Levine estimators is taken to be 0.1 and 0.06, respectively. The corresponding variance and mean squared error of the estimators are presented in the plots (c)-(h) in the figure (6.2), respectively.

For the mean functions $m_2(x)$ to $m_6(x)$, the same above steps are repeated where the sample size is $n = 100$. The bandwidth h_2 of the Brown and Levine estimators is chosen to be 0.06 for the simulation studies in this section. The models and the chosen values of the bandwidths h_1 and h_2 for the estimator in (6.2) are specified in sequel. It should also be noted that the means of $\hat{v}_{New}(x)$s and $\hat{v}_{BL}(x)$s versus x_is are plotted in the figures (6.3), (6.5), (6.7), (6.9) and (6.11), respectively, where the number of the replications is also $N = 1000$ times. The corresponding variance and mean squared error of the estimators for the simulation studies in these figures are plotted in the figures (6.4), (6.6), (6.8), (6.10)

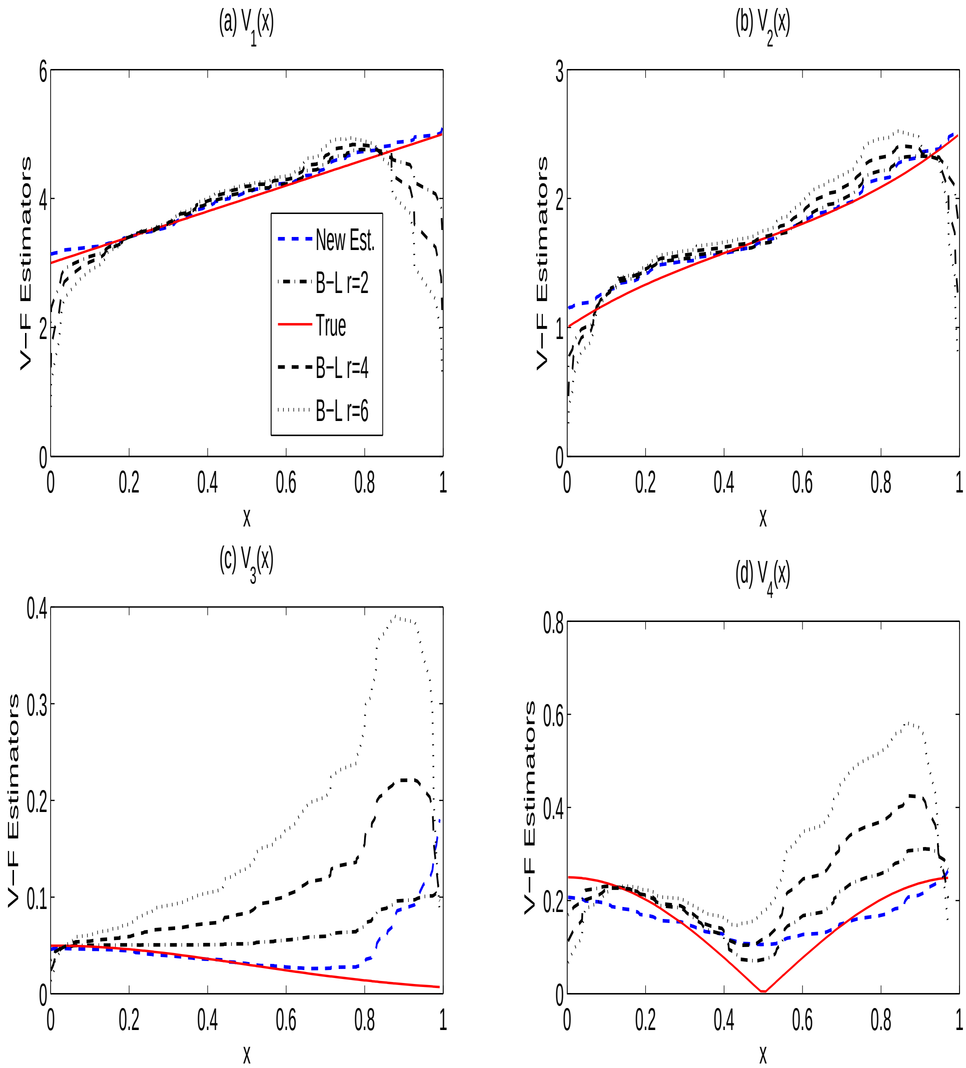

Figure 6.3: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_2(x)$, *(New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

and (6.12), respectively.

The models, using the second mean function $m_2(x)$, are:

$$(i)\, Y_i \;=\; 4.7 \;+ 2.4\, x_i \,+\, 5\, x_i^2 \;+\; 4.3\, x_i^3 \,+\, \sqrt{3 \;+\; 2\, x_i}\, \epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n.$$

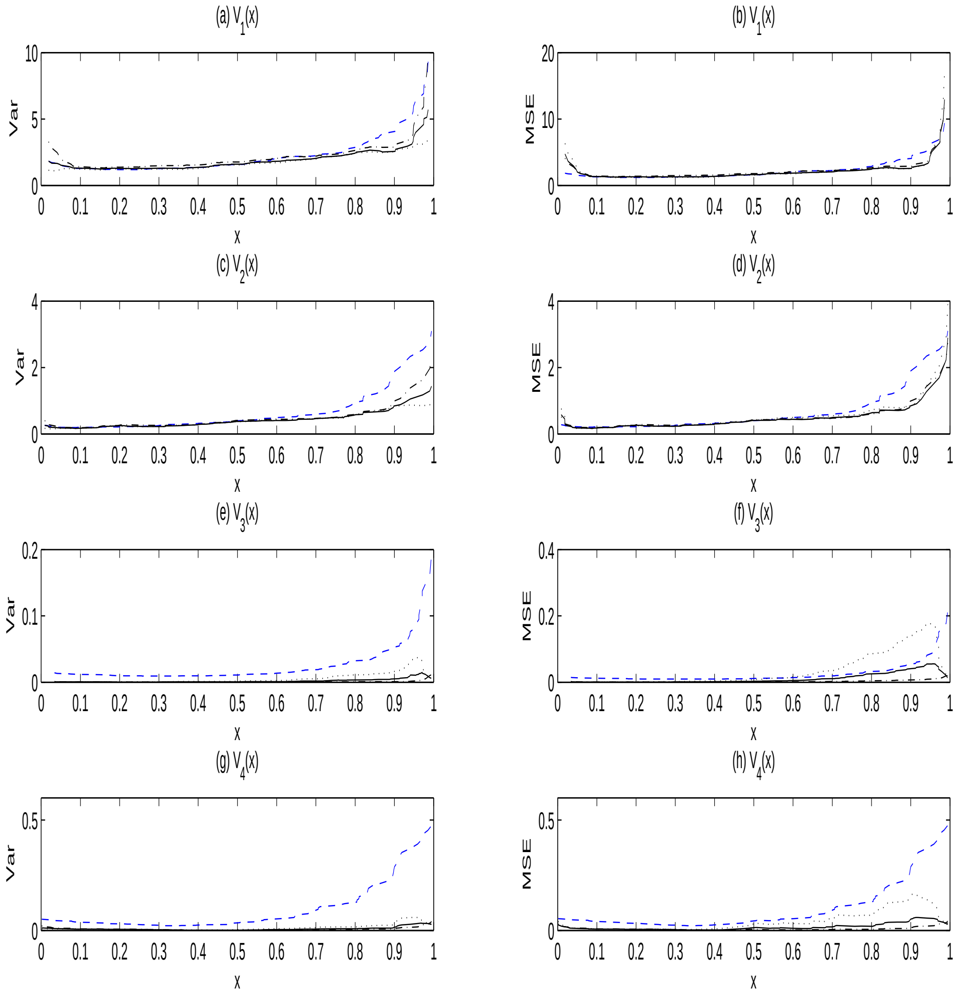

Figure 6.4: *The Variances and Mean Squared errors the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.3), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

$$(ii)\, Y_i \;=\; 4.7 \;+2.4\,x_i \;+\; 5\,x_i^2 \;+\; 4.3\,x_i^3$$

$$+\;\sqrt{0.5\;(2\;+4\,x_i\;-\;4\,x_i^2\;+\;3\,x_i^3\,)}\,\epsilon_i, \qquad \text{for}\;\; i=1,2,\cdots,n.$$

$$(iii)\, Y_i \;=\; 4.7 \;+2.4\,x_i \;+\; 5\,x_i^2 \;+\; 4.3\,x_i^3 \;+\; \sqrt{\exp\left(-4 -\; 5\,x_i^2\right)}\,\epsilon_i, \qquad \text{for}\;\; i=1,2,\cdots,n.$$

$$(iv)\, Y_i \;=\; 4.7 \;+2.4\,x_i \;+\; 5\,x_i^2 \;+\; 4.3\,x_i^3 \;+\; \sqrt{\mid\; 0.25\cos\left(\pi\,x_i\right)\;\mid}\,\epsilon_i, \qquad \text{for}\;\; i=1,2,\cdots,n.$$

The bandwidth h_1 of the estimator in (6.2) is chosen as 0.01 in the plots (a) and (b) in

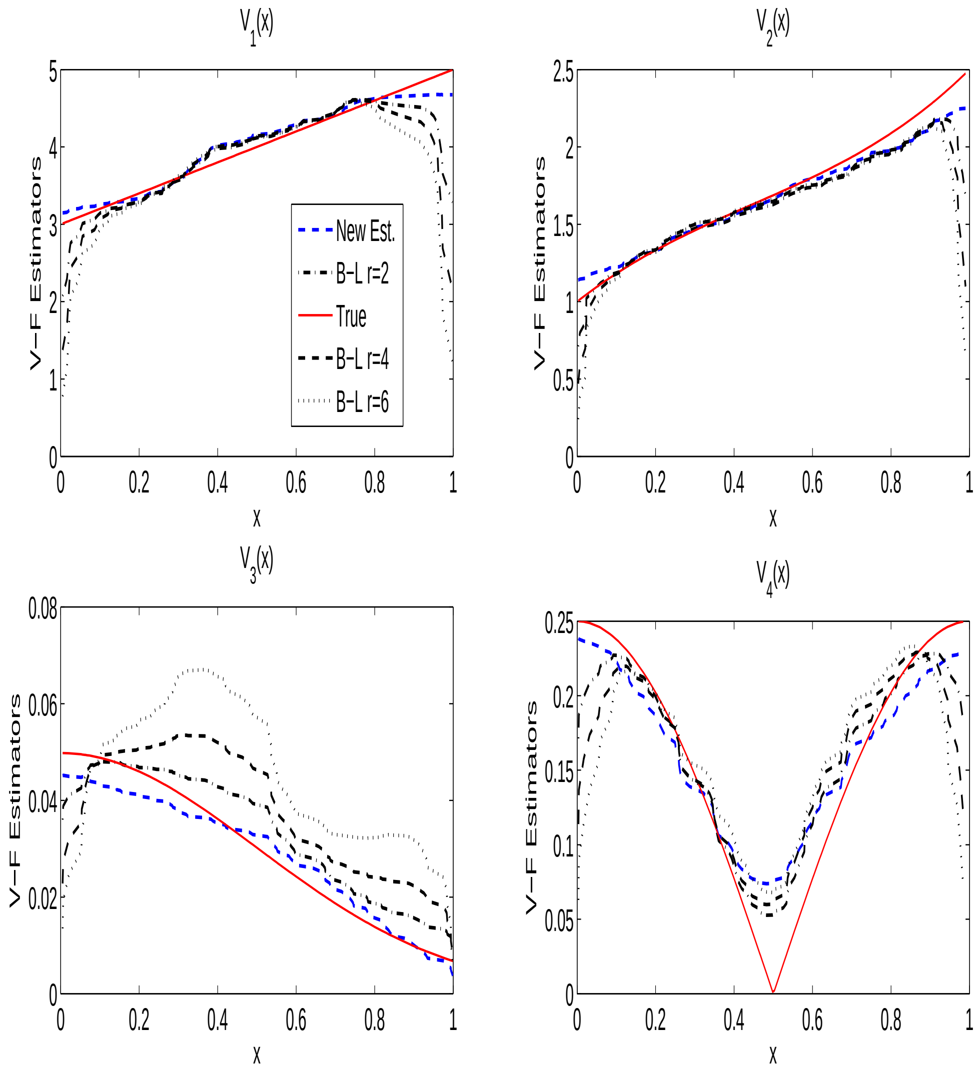

Figure 6.5: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_3(x)$*, (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

the figure (6.3) and to be 0.0032 in the plots (c) and (d) in the same figure. The bandwidth h_2 of the estimator in (6.2) is selected to be 0.1, 0.12, 0.15 and 0.2 for the plots (a), (b), (c) and (d) in the figure (6.3), respectively.

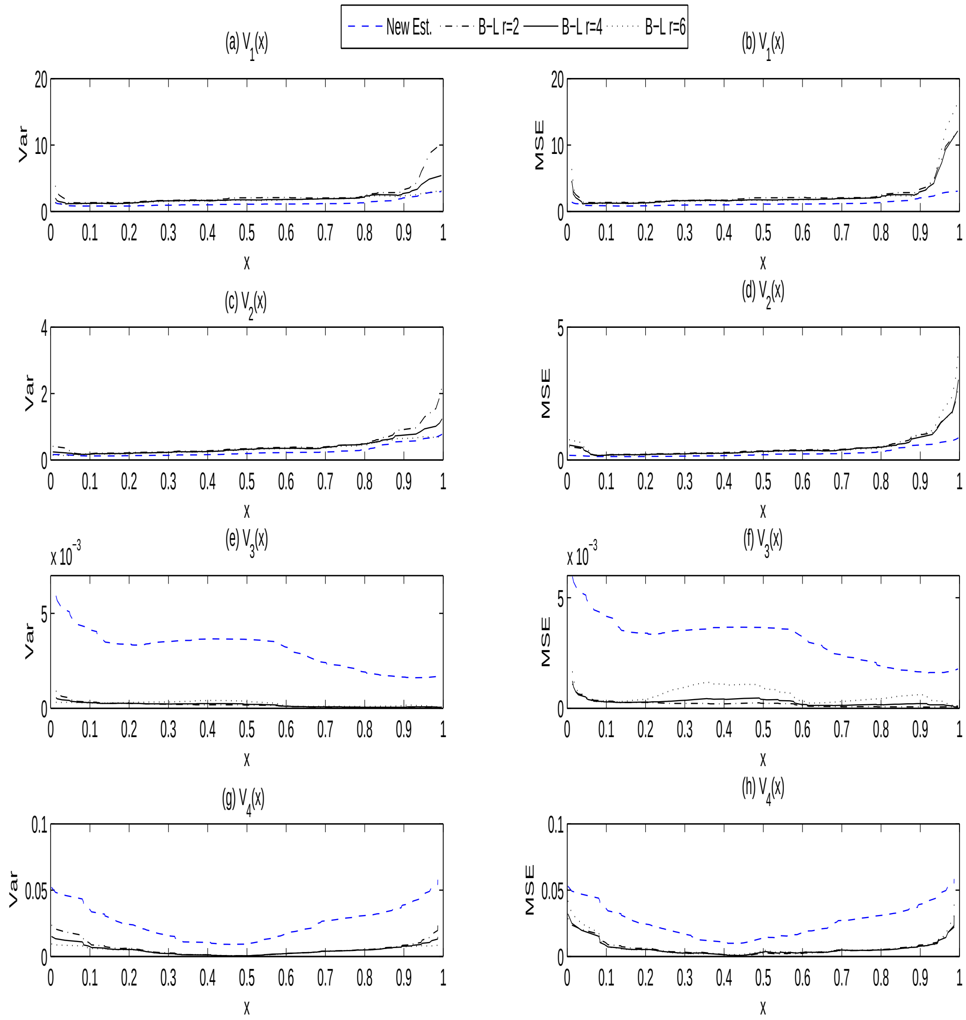

Figure 6.6: *The Variances and Mean Squared errors the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.5), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

For the third mean function $m_3(x)$, the models are:

$$\begin{aligned}
(i)\, Y_i &= \left(3 + x_i + 4\,x_i^2 + 8\,x_i^4\right) \cdot I(x_i \leq 0.5) + \left(5.875 - x_i - x_i^2 - x_i^3\right) \\
&\times\ I(\,x_i > 0.5) + \sqrt{3\ + 2\,x_i}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n. \\
(ii)\, Y_i &= \left(3 + x_i + 4\,x_i^2 + 8\,x_i^4\right) \cdot I(x_i \leq 0.5) + \left(5.875 - x_i - x_i^2 - x_i^3\right) \\
&\times\ I(\,x_i > 0.5) + \sqrt{0.5\ (2\ + 4\,x_i\ - 4\,x_i^2\ + 3\,x_i^3\,)}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n.
\end{aligned}$$

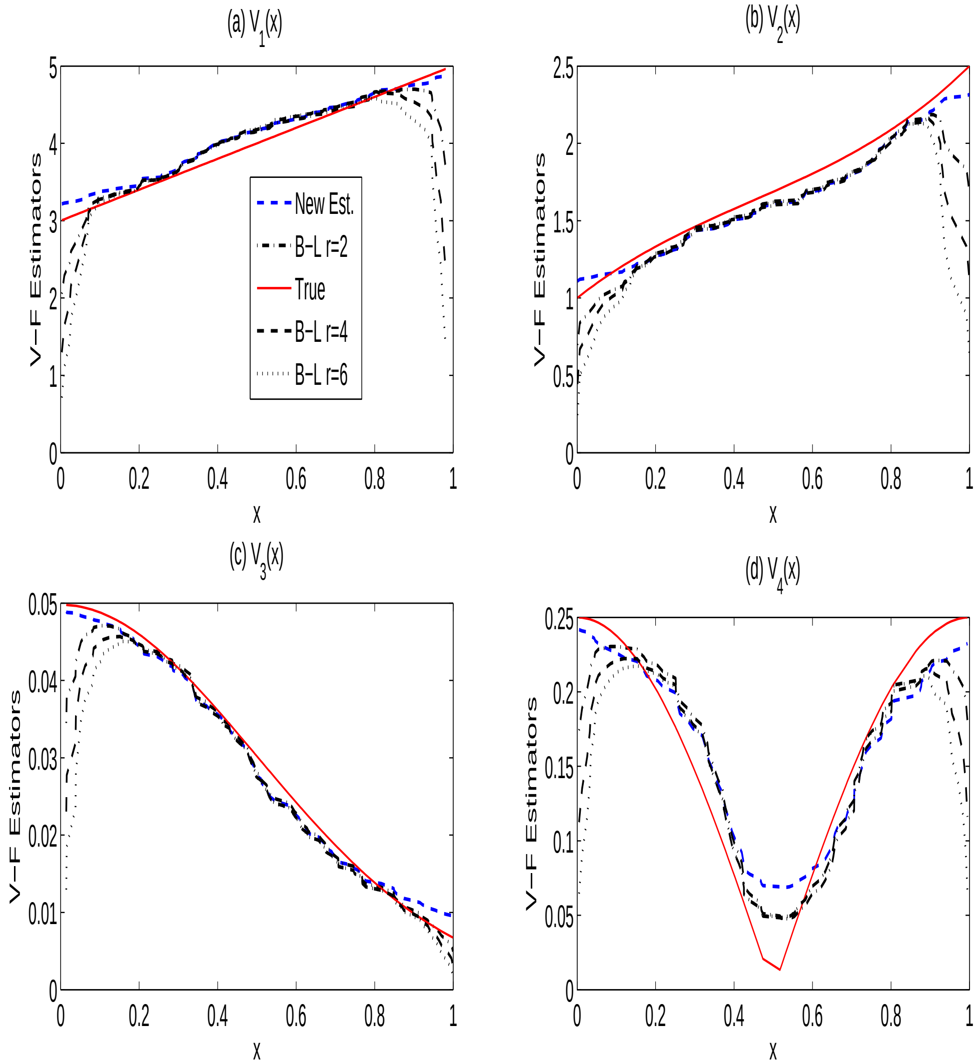

Figure 6.7: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_4(x)$ *, (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

$$
\begin{aligned}
(iii)\, Y_i \;&=\; \left(3 + x_i + 4\,x_i^2 + 8\,x_i^4\right)\,\cdot\, I(x_i \le 0.5) + \left(5.875 - x_i \,-\, x_i^2 \,-\, x_i^3\right) \\
&\times\;\; I(\,x_i > 0.5) \,+\, \sqrt{\exp\left(-4 - 5\,x_i^2\right)}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n. \\
(iv)\, Y_i \;&=\; \left(3 + x_i + 4\,x_i^2 + 8\,x_i^4\right)\,\cdot\, I(x_i \le 0.5) + \left(5.875 - x_i \,-\, x_i^2 \,-\, x_i^3\right) \\
&\times\;\; I(\,x_i > 0.5) \,+\, \sqrt{|\;0.25 \cos\left(\pi\, x_i\right)\;|}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n.
\end{aligned}
$$

The bandwidth h_1 of the estimator in (6.2) is taken as 0.025 in the plots (a) and (b) in

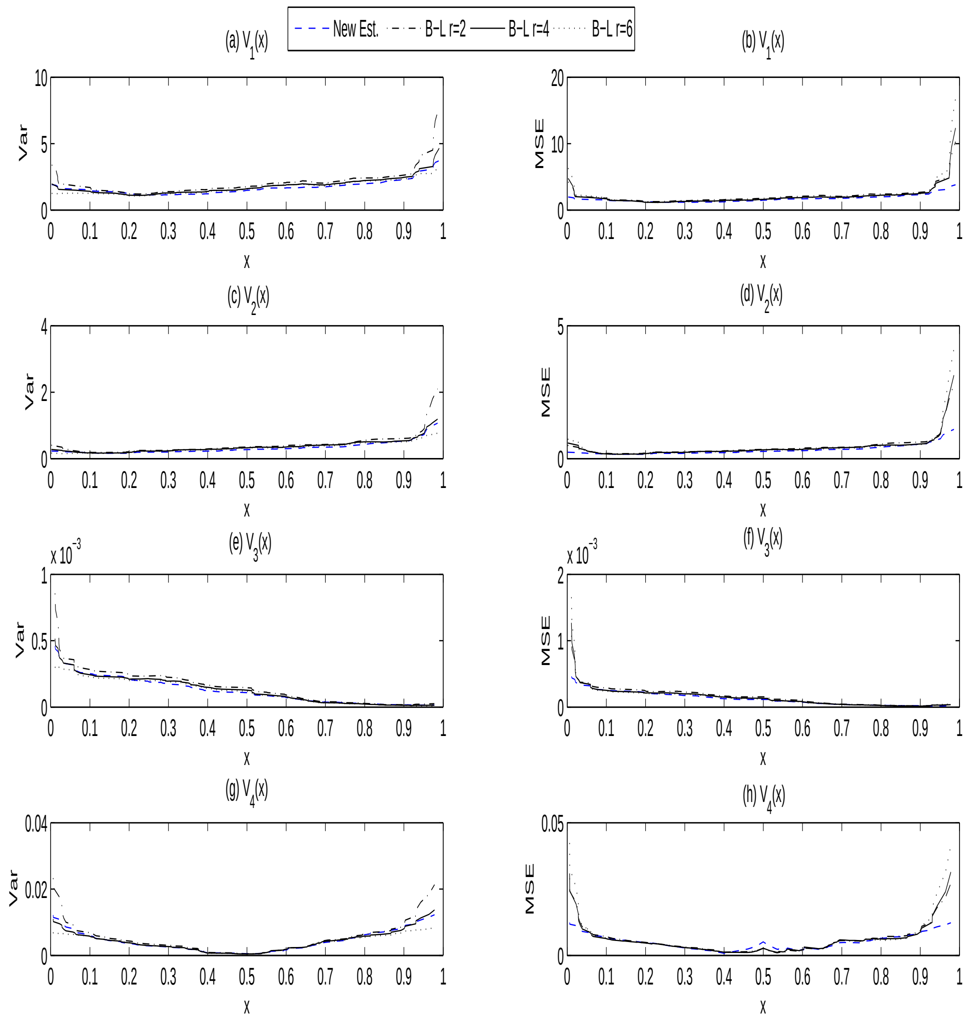

Figure 6.8: *The Variances and Mean Squared errors the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.7), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

the figure (6.5) and to be 0.0032 in the plots (c) and (d) in the same figure. The bandwidth h_2 of the estimator in (6.2), which is used to estimate the variance function, is chosen to be 0.1, 0.1, 0.12 and 0.1 for the plots (a), (b), (c) and (d) in the figure (6.5), respectively. Using the fourth mean function $m_4(x)$, the models become:

$$(i)\, Y_i \;=\; \exp\left(-2 - 4\,x_i - 5x_i^2 - 6x_i^3\right) + \sqrt{3 + 2\,x_i}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n.$$

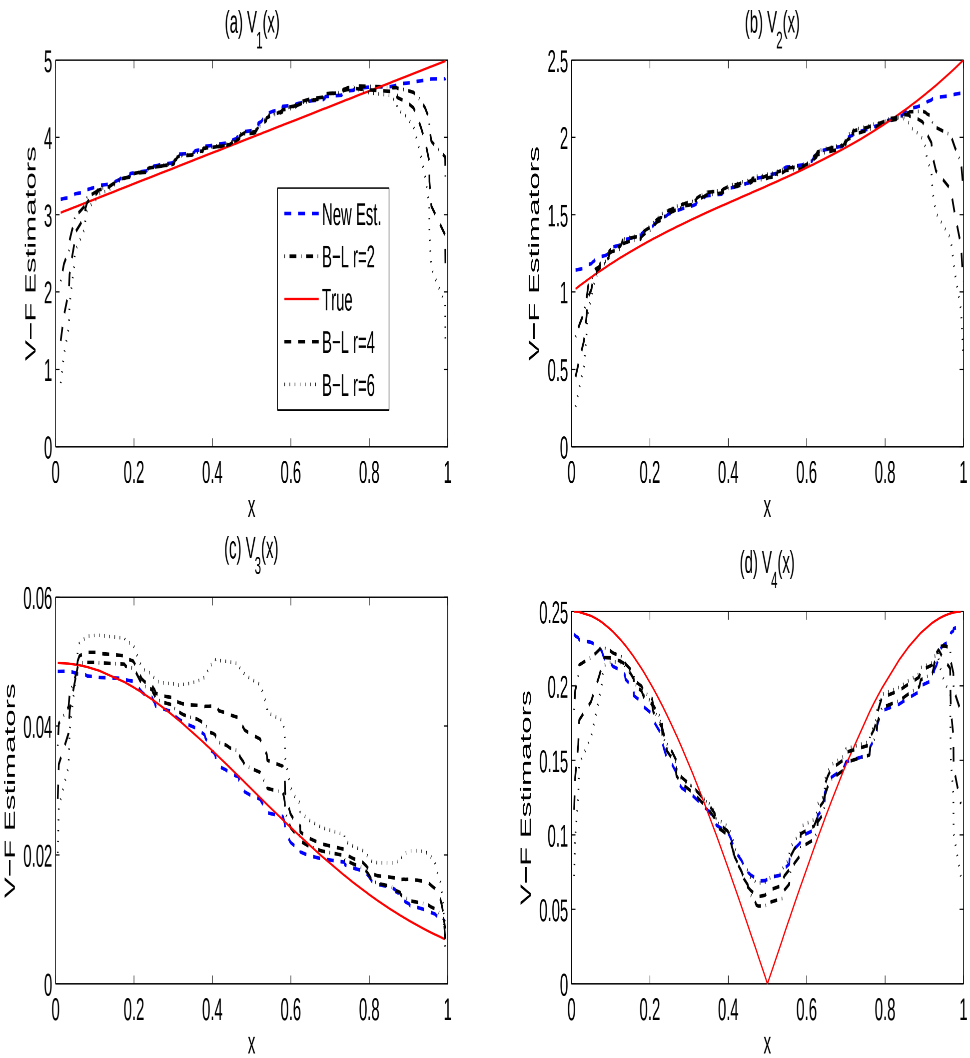

Figure 6.9: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_5(x)$, *(New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

$$
\begin{aligned}
(ii)\, Y_i &= \exp\left(-2 - 4\,x_i - 5x_i^2 - 6x_i^3\right) \\
&+ \sqrt{0.5\,\left(2 + 4\,x_i - 4\,x_i^2 + 3\,x_i^3\right)}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n. \\
(iii)\, Y_i &= \exp\left(-2 - 6\,x_i - 5x_i^2 - 6x_i^3\right) + \sqrt{\exp\left(-4 - 5\,x_i^2\right)}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n. \\
(iv)\, Y_i &= \exp\left(-2 - 6\,x_i - 5x_i^2 - 6x_i^3\right) + \sqrt{|\ 0.25\cos\left(\pi\,x_i\right)\ |}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n.
\end{aligned}
$$

For all plots in the figures (6.7), (6.9) and (6.11), the bandwidth h_1 for the estimator in

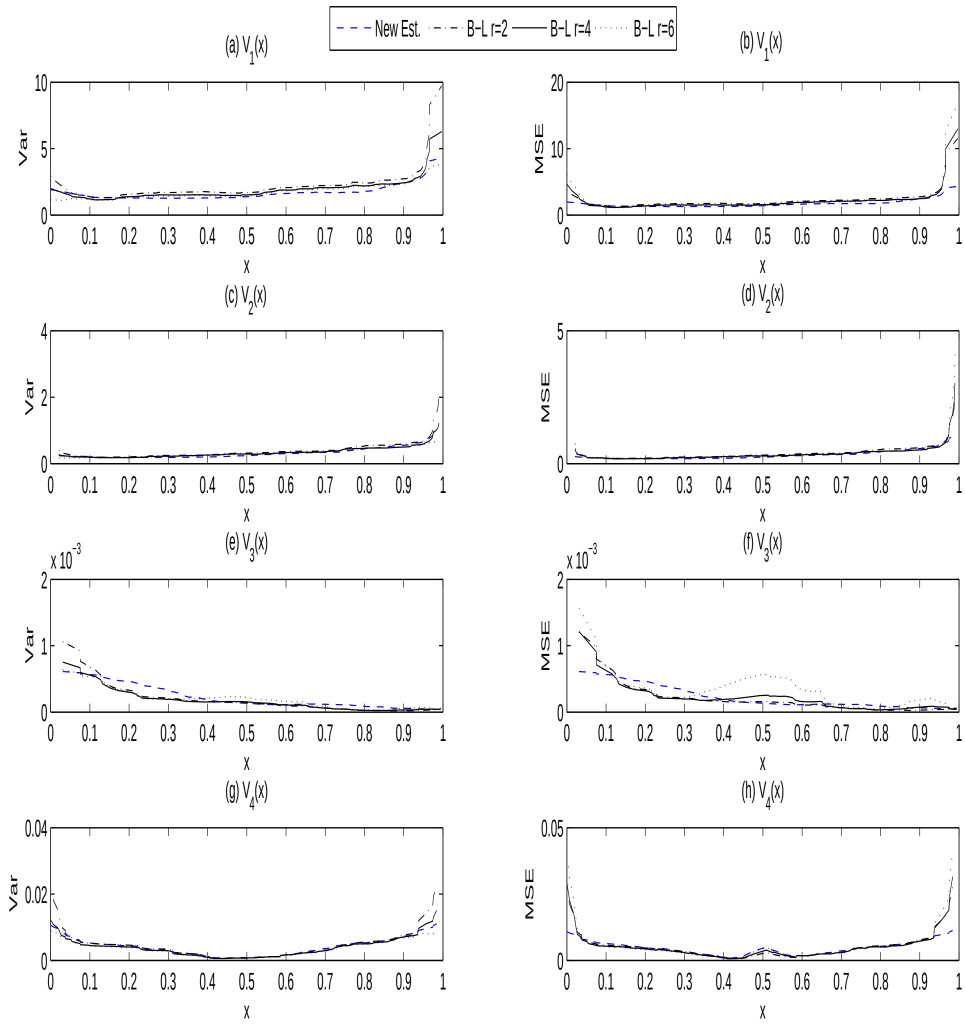

Figure 6.10: *The Variances and Mean Squared errors the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.9), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

(6.2) is 0.0032, whereas the bandwidth h_2 is selected to be 0.01. It should be noted that the models, which use to produce the figure (6.7), are described in the above four equations, whereas the models, using the mean function $m_5(x)$, are as follows:

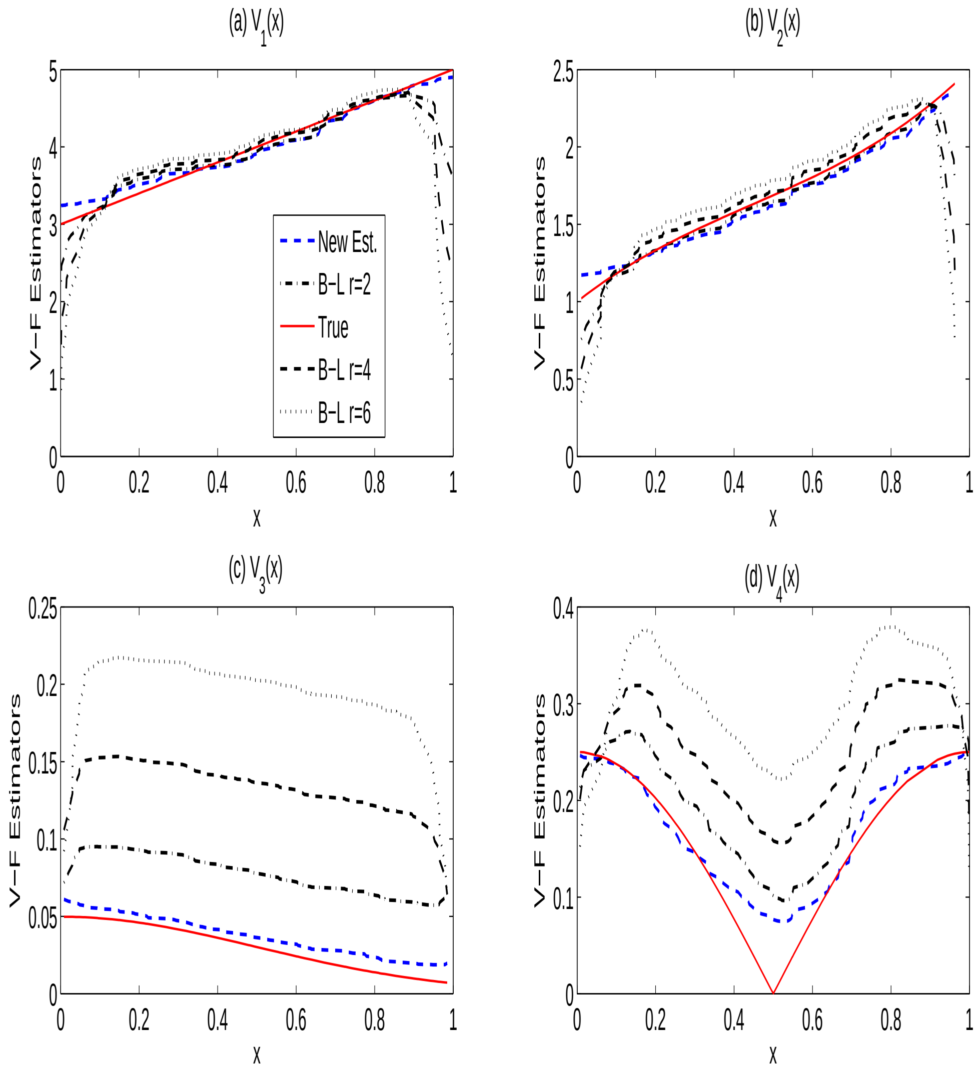

Figure 6.11: *The Comparison Between the Estimated Variance Functions by the New Estimator and the Brown and Levine Estimators where* $m(x) = m_6(x)$ *, (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

$$(i)\; Y_i \;=\; \frac{4}{5}\sin\left(2\,\pi\, x_i\right) \;+\; \sqrt{3 \;+\; 2\, x_i}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(ii)\; Y_i \;=\; \frac{4}{5}\sin\left(2\,\pi\, x_i\right) \;+\; \sqrt{0.5\;\left(2 \;+ 4\, x_i \;-\; 4\, x_i^2 \;+\; 3\, x_i^3\,\right)}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(iii)\; Y_i \;=\; \frac{4}{5}\sin\left(2\,\pi\, x_i\right) \;+\; \sqrt{\exp\left(-4 -\; 5\, x_i^2\right)}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(iv)\; Y_i \;=\; \frac{4}{5}\sin\left(2\,\pi\, x_i\right) \;+\; \sqrt{\mid\; 0.25\cos\left(\pi\, x_i\right)\;\mid}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

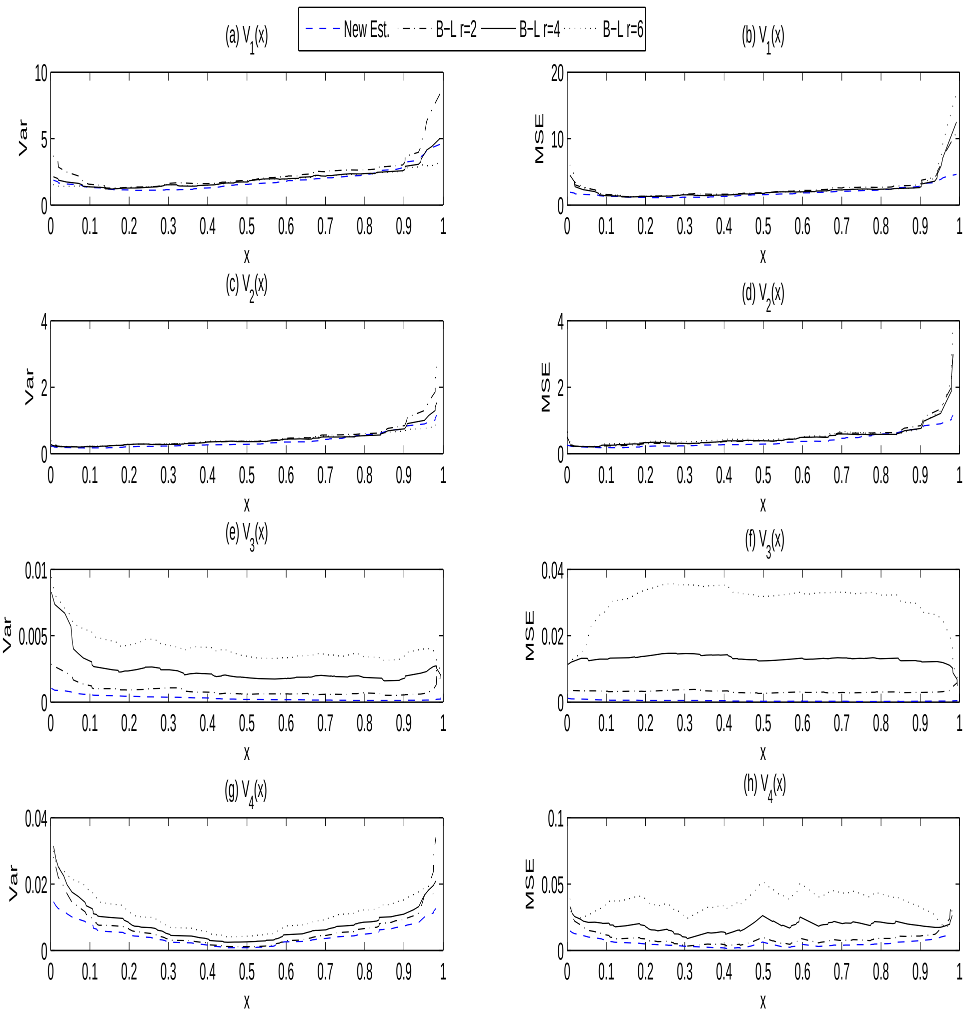

Figure 6.12: *The Variances and Mean Squared errors the New Estimator and the Brown and Levine Estimators for Simulation Studies in the Figure (6.11), (New estimator-Blue; Brown & Levine Estimators-Black; True-Red).*

The models using $m_6(x)$ are:

$$(i)\, Y_i \;=\; \frac{3}{4}\cos\left(10\,\pi\, x_i\right) \;+\; \sqrt{3 \;+\; 2\, x_i}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(ii)\, Y_i \;=\; \frac{3}{4}\cos\left(10\,\pi\, x_i\right) \;+\; \sqrt{0.5\;\left(2 \;+4\, x_i \;-\; 4\, x_i^2 \;+\; 3\, x_i^3\right)}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(iii)\, Y_i \;=\; \frac{3}{4}\cos\left(10\,\pi\, x_i\right) \;+\; \sqrt{\exp\left(-4 -\; 5\, x_i^2\right)}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

$$(iv)\, Y_i \;=\; \frac{3}{4}\cos\left(10\,\pi\, x_i\right) \;+\; \sqrt{\left|\;0.25\cos\left(\pi\, x_i\right)\;\right|}\,\epsilon_i, \qquad \text{for } \; i = 1, 2, \cdots, n.$$

6.3.2 Results

The main results from the figures (6.1)-(6.12) are described in this subsection. From the figures (6.1), (6.7), (6.9) and (6.11), in general, the performance of the estimator in (6.2) and the Brown and Levine estimators are approximately the same except near the boundary, which was expected. So, the Brown and Levine estimators do not perform as good in the boundary as the estimator in (6.2) since there is a clear lack of information near the boundary because of differencing. In fact, higher the order of differencing, worst is the performance of the Brown and Levine estimator near the boundary. In the figures (6.2), (6.8), (6.10) and (6.12), the variance and the mean squared error of the estimator in (6.2) are smaller than that of the Brown and Levine estimators.

In the plots (a) and (b) in the figures (6.3) and (6.5), the performances of the estimators are approximately the same in the interior points. The estimator in (6.2) has less bias in the boundary points than that of the Brown and Levine estimators. In the plots (c) and (d) in these two figures, the estimator in (6.2) may has less bias than that of the Brown and Levine estimators, but their variances and their mean squared errors are less than that of the estimator (6.2).

6.3.3 Discussions

This discussion is valid when the bandwidths for the estimator in (6.2) and the Brown and Levine estimators are chosen appropriately. Recall that the chosen values for the bandwidths h_1 and h_2 in subsection 6.3.1 are selected optimally for both estimators. When the variance function is a simple linear regression function, the bias of the estimated variance function by both estimators is expected to be zero since the second derivative of the variance function is zero. However, the bias is expected to be bigger than zero when the variance function is polynomial regression functions of order ≥ 3, exponential functions or trigonometric functions. In the figures (6.1), (6.3), (6.5), (6.7), (6.9) and (6.11), it should be noted that, in the boundary points, there is a clear bias in the estimated variance functions by the Brown and Levine estimators because of using differencing in the estimation of the variance function. This point will be clarified at the end of this subsection.

Clearly, when the mean function is a polynomial function, which has rth continuous derivatives where $r \leq 10$, its effect on the finite sample performance of the estimator in (6.2) is very clear. See for the example plots (f) and (h) in the figures (6.4) and (6.6). However, when the mean function is a smooth function, its effect on the performance of the estimator is less than that of a polynomial regression function of order $r \leq 10$, but this effect is smaller than that of the variance function. For the variance function, in general, when the variance function is a polynomial regression function of order smaller than 10, the estimator in (6.2) is expected to perform better than that of the Brown and Levine estimators in the boundary. However, the performances of the estimators are approximately the same in the interior points. On the other hand, when the mean and variance function are a smooth function, we can conclude that the estimator in (6.2) performs better than the Brown and Levine estimators. However, when the variance function is a smooth function and the mean function is a polynomial regression function of order smaller than 10, the performances of the estimator in (6.2) and the Brown and Levine estimators are approximately the same in the interior points, but the Brown and Levine estimators perform better than the estimator in (6.2) in the boundary in term of the mean squared error. In summary, when the mean function is a polynomial function of order ≥ 3, exponential function or trigonometric function and the variance function is a polynomial function of order ≥ 3, the performance of the estimator in (6.2) is better than that of the Brown and Levine estimators in the boundary. When the mean and variance function are smooth functions, the estimator in (6.2) performs better than the Brown and Levine estimators.

From Theorem 4.2.1, recall that the mean squared error of the estimator in (6.2) is

$$MSE(\hat{v}(x)) = h_2^{2r} C_1^2(x) + n^{-1} h_2^{-1} C_2(x) + o(n^{-1} h_2^{-1}) + o(h_2^{2r}), \tag{6.7}$$

where $C_1(x) = \frac{1}{r!} v^{(r)}(x) \int y^r K(y) \, dy$ and $C_2(x) = (\mu_4(x) - v^2(x)) \cdot \int K^2(t) \, dt$. Thus, it is obvious that the deterministic functions $C_1(x)$ and $C_2(x)$ depend on the variance function that we are estimating, whereas the mean function does not have a first order effect. Thus, its contribution in the the mean squared error of the estimator in (6.2) is expected to be negligible. In the finite sample, the results in this section support this conclusion. Note that the effect of the variance and mean functions on the finite sample performance of the estimator in (6.2) decreases as the size of the sample increases. From chapter 4, in the asymptotic analysis, we showed that the mean squared errors of the estimator in (6.2) and the Brown and Levine estimator in (6.3) are the same in the first order where the order of differences is 2. In the finite sample case, when the order of differences is 2, the difference in the boundary points between these two estimators are due to the lack of information in the difference-based method.

Generally, the performances of these two estimators are approximately the same in the interior points when the order of differences is 2. However, as the order of differences rises, it is clear that the lack of information in the boundary points increases and in some cases the bias also rises. It should be noted that when the optimal bandwidth of the Brown and Levine estimator is bigger than that of the estimator in (6.2), the Brown and Levine estimators might have a bigger bias than that of the estimator in (6.2). On the other hand, small choices of the bandwidth h_2 for the Brown and Levine estimator decrease the bias, but the variance rises and so does the lack of the information in the boundary points.

According to the literature, when the local polynomial fitting is used to estimate the mean function in the nonparametric regression models, the estimated mean function does not suffer from the boundary effect. So, we can use the same kernel in the boundary and interior points. In other words, the local polynomial estimators for the mean function adapt automatically when they estimate the boundary points. The same thing happens when the local polynomial fitting is used to estimate the density function. On the other hand, when the local polynomial fitting is used to estimate the error variance function in the difference-based method, the difference-based estimators suffer from the boundary bias. This bias in the boundary points is due to the lack of information when the differences' sequences are computed to estimate the error variance function.

Conclusion:

Generally, the performances of the estimators in the interior points are approximately the same except when the mean and variance function are trigonometric functions. In this case, the estimator in (6.2) performs better than the Brown and Levine estimators in the interior and boundary points. When the mean and variance are polynomial functions of order ≥ 3, the performance of the estimator in (6.2) is better than that of the Brown and Levine estimators in the boundary. However, when the mean function is a polynomial function of order ≥ 3 and the variance function is a smooth function, the opposite occurs.

Clearly, the effect of the variance function on the finite sample performance of the estimator in (6.2) is larger than the effect of the mean function in all cases in this section. Thus, these results support the conclusion of Theorem 4.2.1 that estimating mean function does not have a first order effect on the mean squared error of the estimator in (6.2). Note that when the mean function is a polynomial function of order ≥ 3, it has more influence on the finite sample performance of the estimator in (6.2) than when it is a smooth function.

6.4 The Effect of the Bandwidth Selection on the Performance of the Estimator in (6.2)

The bandwidth selection is one of the most important aspects in the error variance estimation since it plays a major role on the performance of the error variance function estimators. Furthermore, it is obvious from chapter 4 that, in the asymptotic analysis, the performance of the estimator in (6.2) depends on the bandwidths h_1 and h_2. In chapter 4, we also found that the effect of the bandwidth h_2 on the mean squared error of the estimator in (6.2) is more significant than the influence of the bandwidth h_1. In the current section, we study the effect of the bandwidth selection on the finite sample performance of the estimator in (6.2) through simulation. The effect of the bandwidth h_1 on the performance of the estimator in (6.2) is considered in subsection 6.4.1, whereas the influence of the bandwidth h_2 is studied in subsection 6.4.2.

6.4.1 The Effect of the Bandwidth h_1 on the Finite Sample Performance of the Estimator in (6.2)

The effect of the choices of the bandwidth h_1 on the finite sample performance of the estimator in (6.2) is studied in this subsection. To study this effect, the mean function, the variance function and the bandwidth h_2 are fixed. Then, the bandwidth h_1 is allowed to vary. Note that the bandwidth h_2 is selected to be balancing between the squared bias and the variance of the estimator in (6.2). The number of replications in the simulation studies is chosen to be $N = 1000$ of sample size $n = 100$. In particular, we study the influence of the h_1 on the finite sample performance of the estimator in (6.2) with three different mean functions and two variance functions. Note that the structure of the simulation studies in this

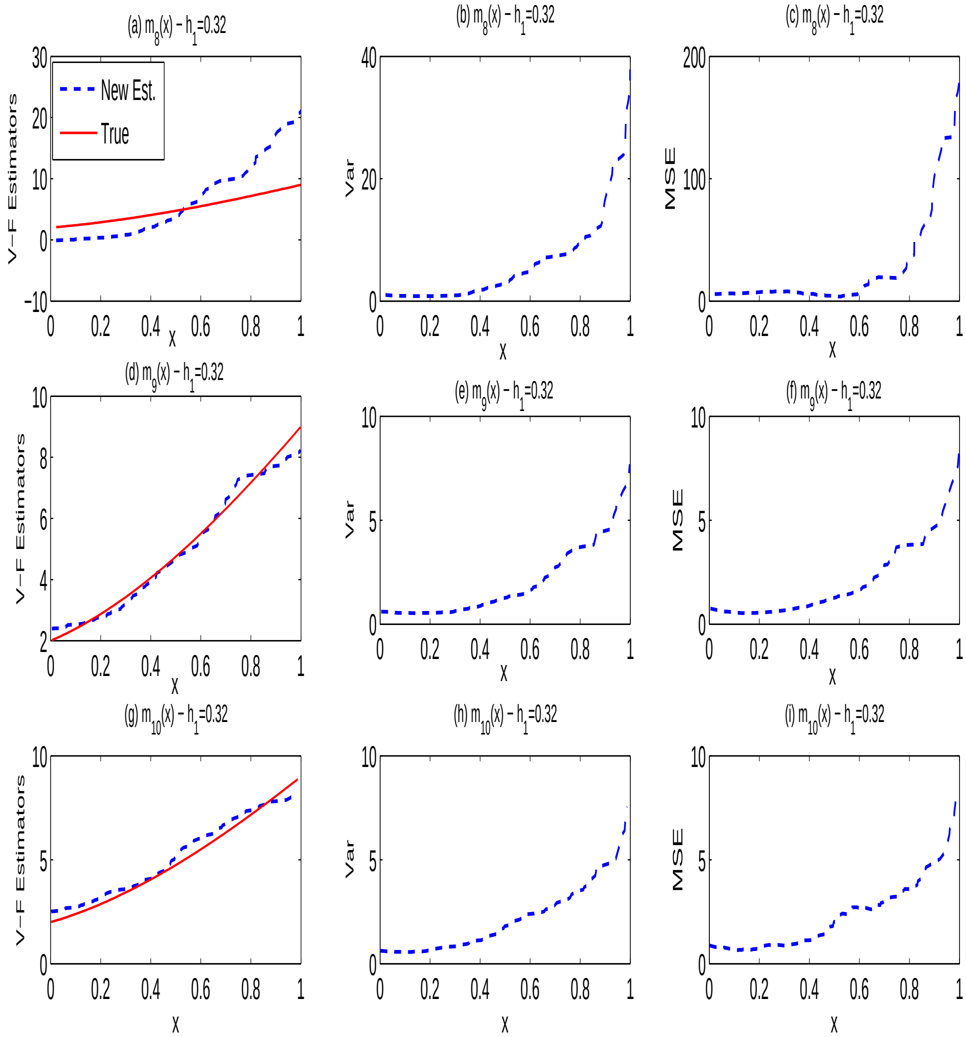

Figure 6.13: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$, $h_1 = 0.32$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

subsection is described in section 6.2. The chosen mean functions are:

$$
\begin{aligned}
i)\; m_8(x) &= 3 + 1.8\,x + 2.5\,x^2 + 1.5\,x^3. \\
ii)\; m_9(x) &= \exp(-2 - 5\,x - 2\,x^2). \\
iii)\; m_{10}(x) &= \frac{1}{2}\sin(25\,\pi\,x).
\end{aligned}
$$

It should be noted that the plots of the above three mean functions are similar to the plots of the mean functions $m_2(x)$, $m_4(x)$ and $m_6(x)$, respectively. Thus, their plots are omitted. However, the variance functions are selected to be

$$ii)\ v_5(x) = 2 + 3.5\,x + 4.5\,x^2 - x^3.$$

$$iv)\ v_6(x) = |\ 0.75 \sin(\pi\, x)\ |\,.$$

The aim of choosing different mean and variance functions, rather than that of the previous section is to assess the performance of the estimator in (6.2) with the largest possible number of the functions.

The models for the figures (6.13)-(6.16) are:

$$\begin{aligned}
(i)\ Y_i &= 3 + 1.8\,x_i + 2.5\,x_i^2 + 1.5\,x_i^3 \\
&\quad + \sqrt{2 + 3.5\,x_i + 4.5\,x_i^2 - x_i^3}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \\
(ii)\ Y_i &= \exp(-2 - 5\,x - 2\,x^2) \\
&\quad + \sqrt{2 + 3.5\,x_i + 4.5\,x_i^2 - x_i^3}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \\
(iii)\ Y_i &= \frac{1}{2}\sin(25\,\pi\,x) + \sqrt{2 + 3.5\,x_i + 4.5\,x_i^2 - x_i^3}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, n,
\end{aligned} \tag{6.8}$$

where the assumptions F1, F2 and F3 are satisfied. Using the model (i) in (6.8), we fix the bandwidth h_2 to be $h_2 = 0.1$ and the bandwidth h_1 is taken to be 0.32 . Then, using the same structure described in section 6.2 and the above models, the figure (6.13) is produced.

For the figures (6.14)-(6.16), the structure is as in the figure (6.13) where the bandwidth $h_1 = 0.063,\ 0.01$ and 0.0032, respectively. The bandwidth h_2 in these figures is fixed to be 0.1.

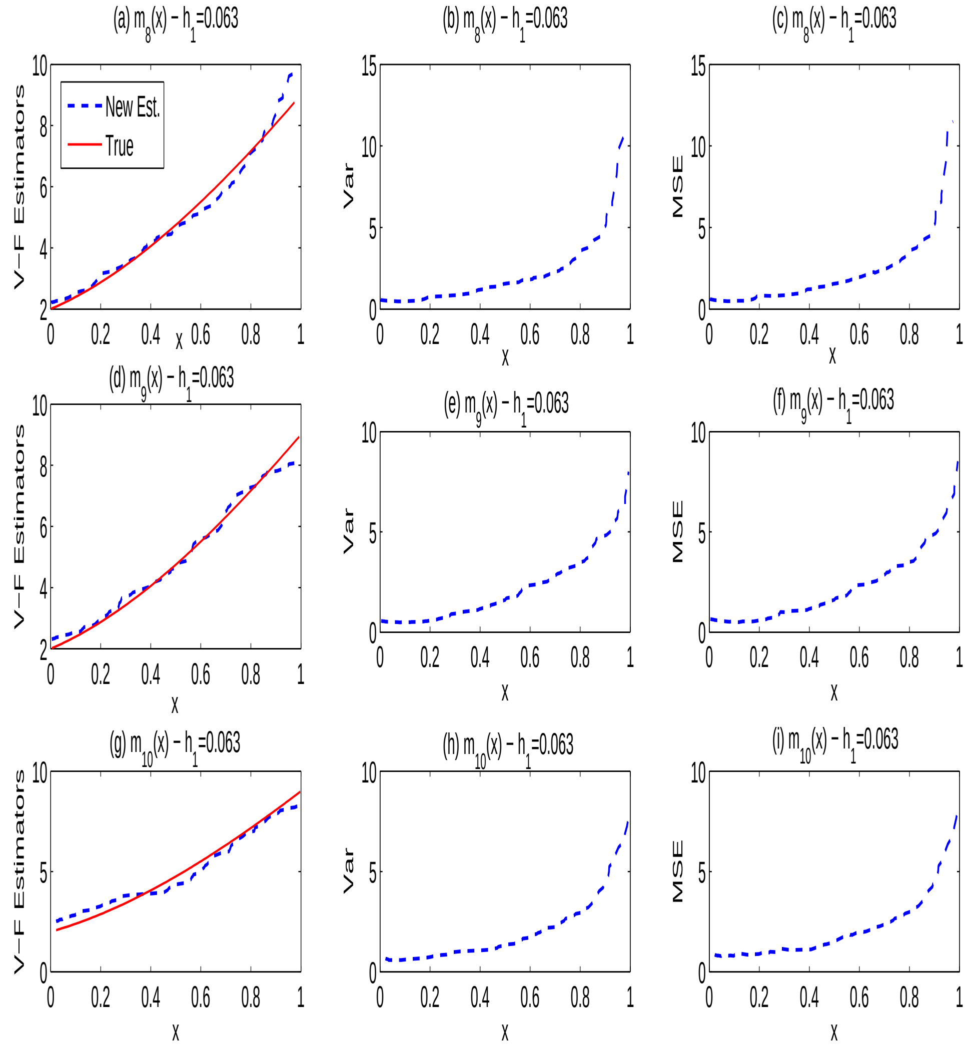

Figure 6.14: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$, $h_1 = 0.063$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

The models of the figure (6.17)-(6.20) can be defined as

$$
\begin{aligned}
(i)\ Y_i &= 3 + 1.8\, x_i + 2.5\, x_i^2 + 1.5\, x_i^3 + \sqrt{|\ 0.75 \sin(\pi\, x_i)\ |}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n,\\
(ii)\ Y_i &= \exp(-2 - 5\, x_i - 2\, x_i^2) + \sqrt{|\ 0.75 \sin(\pi\, x_i)\ |}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n,\\
(iii)\ Y_i &= \frac{1}{2}\sin(25\,\pi\, x) + \sqrt{|\ 0.75 \sin(\pi\, x_i)\ |}\,\epsilon_i, \qquad \text{for } i = 1, 2, \cdots, n, \qquad (6.9)
\end{aligned}
$$

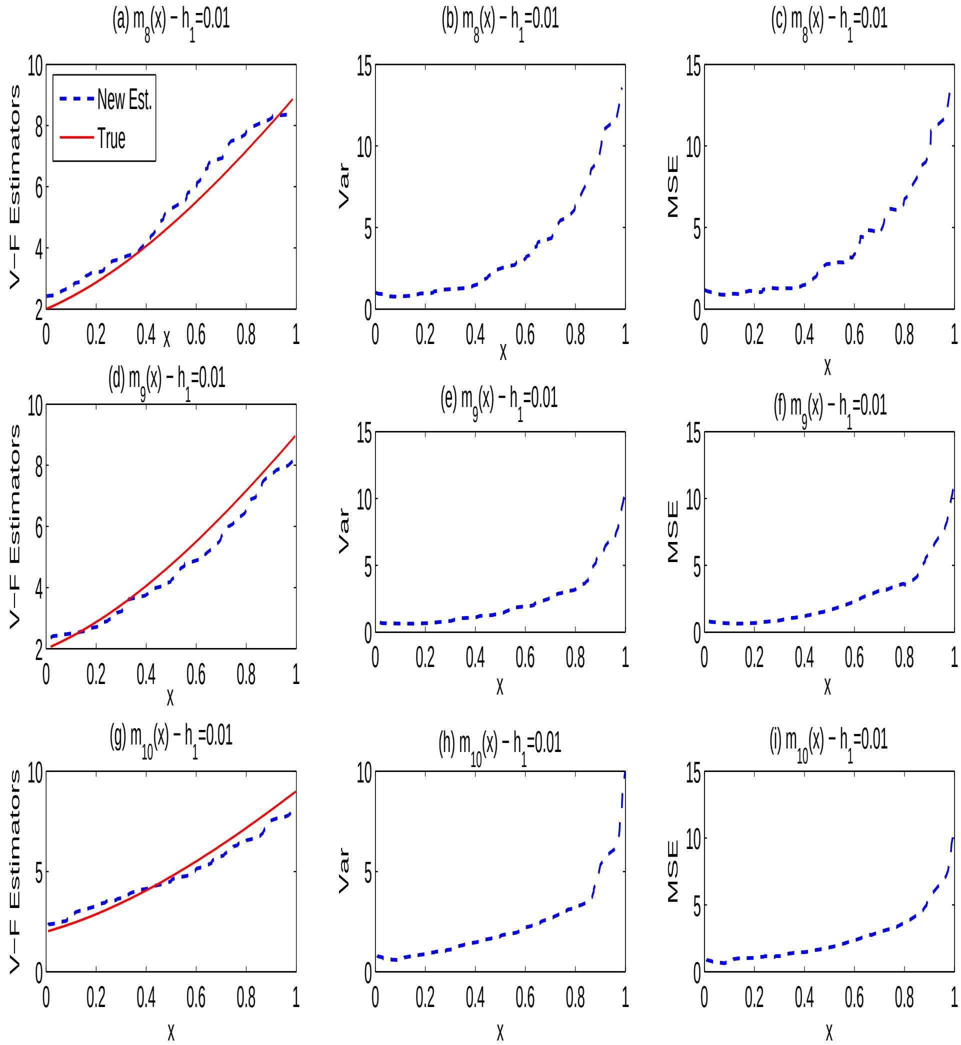

Figure 6.15: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$, $h_1 = 0.01$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

where the assumptions F1, F2 and F3 are true. The figures (6.17)-(6.20) are produced using the same structure as in section 6.2 and the models in (6.9). The bandwidth h_2 in these figures is fixed to be 0.1, whereas the bandwidth h_1 is selected to be 0.32, 0.063, 0.01 and 0.0032, respectively. Now, the results from figures (6.13) to (6.20) are given in the following sub-subsection.

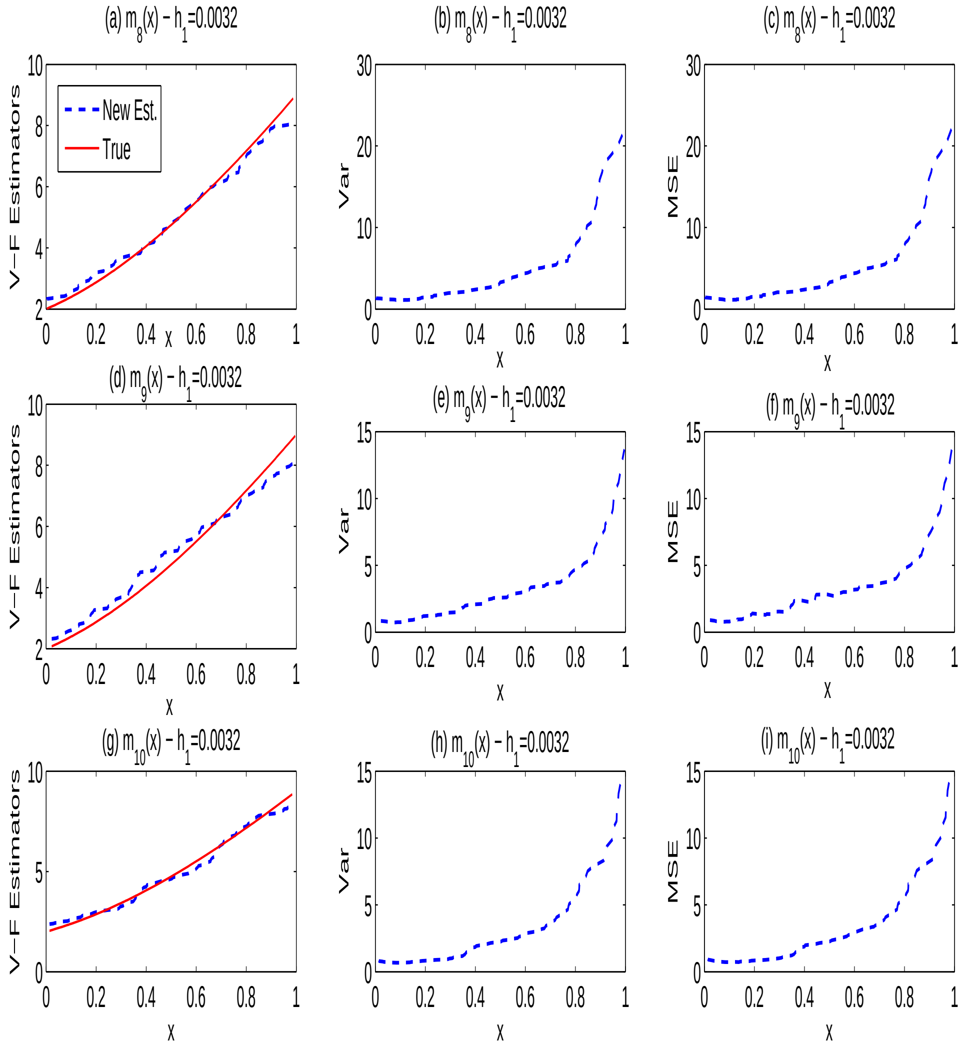

Figure 6.16: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$, $h_1 = 0.0032$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

Results:

In the figures (6.13)-(6.20), recall that the bandwidth h_2 is selected appropriately. In the figures (6.13), when the mean function is $m_8(x)$, there is a very little bias in the estimated variance function when x_is are less than 0.7. On the other hand, the bias of the estimated variance function is very large when x_is are bigger than 0.7. The mean squared error of the estimator in (6.2) is clearly high. For the remaining mean functions in the figure (6.13), there

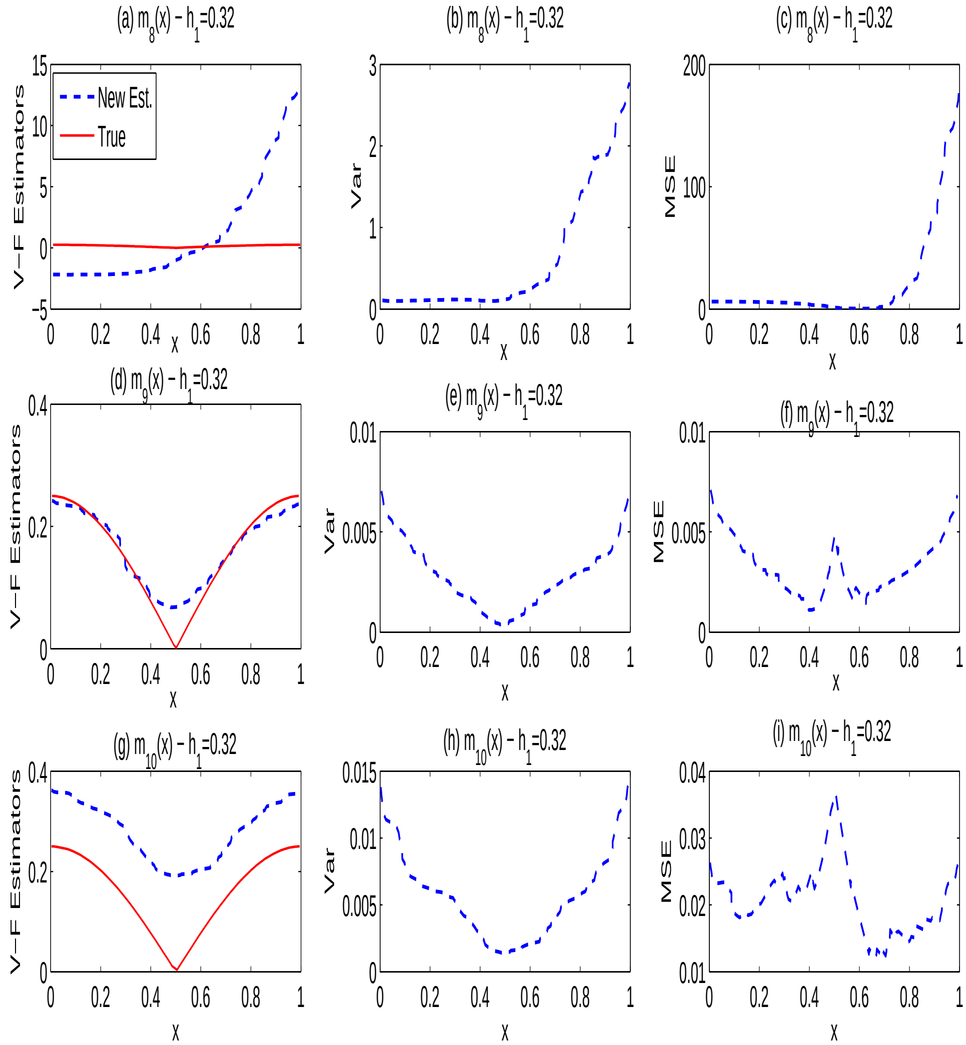

Figure 6.17: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.32$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

is a very little bias in the estimated variance function. In the figures (6.14)-(6.16), the estimated and true variance functions are approximately the same, while the variance increases slightly as the bandwidth h_1 becomes smaller. There is a little bias in the estimated variance function by the estimator in (6.2) when $h_1 = 0.0032$ and 0.01. However, for $h_1 = 0.063$, there is a clear bias in the estimated variance function when x_is are bigger than 0.8. The variance and the mean squared error in the figures (6.14)-(6.16) do not change significantly

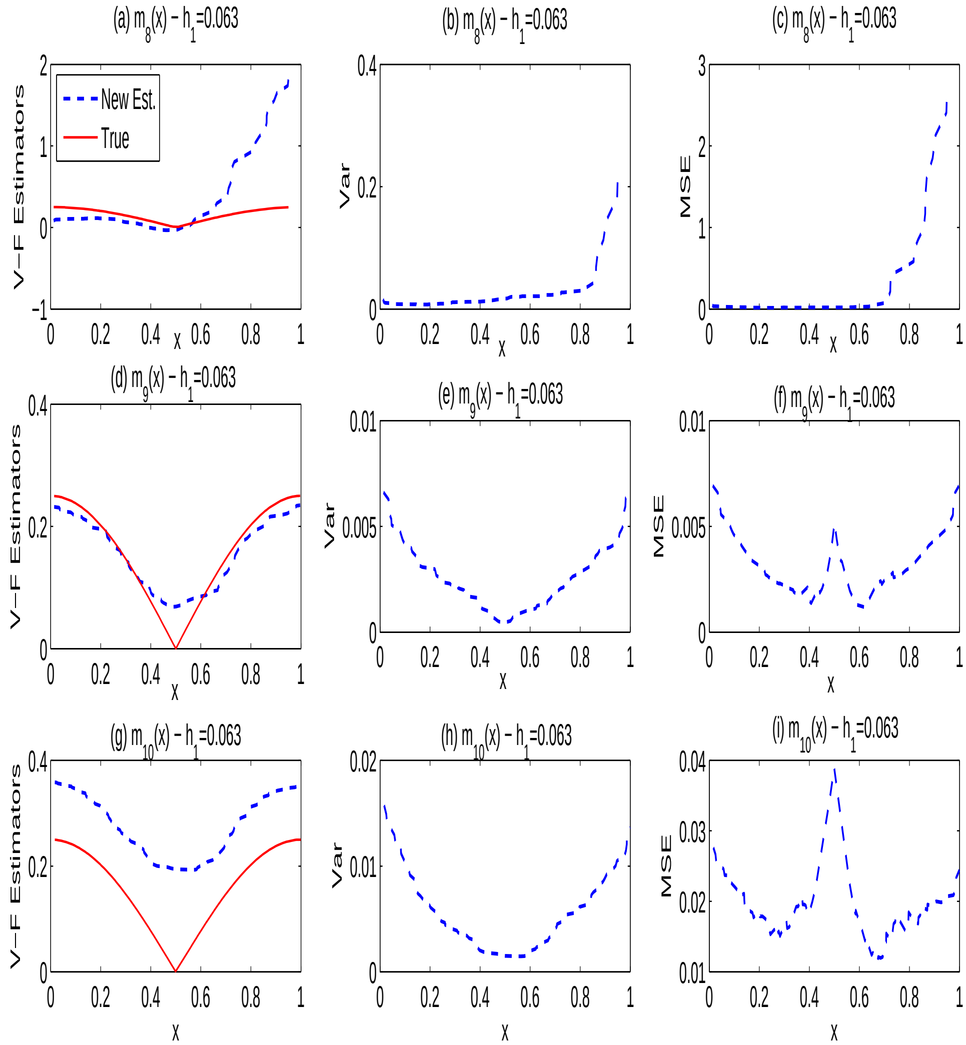

Figure 6.18: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.063$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

with the chosen values for the bandwidth h_1.

From the figures (6.17)-(6.20), when the bandwidth $h_1 = 0.32$ or 0.063 and the mean function is $m_8(x)$ or $m_{10}(x)$, the bias is obviously large. The appropriate choice of the bandwidth h_1, for these two mean functions, is approximately 0.0032. For all chosen values of the bandwidth h_1, the estimated and true variance functions are approximately the

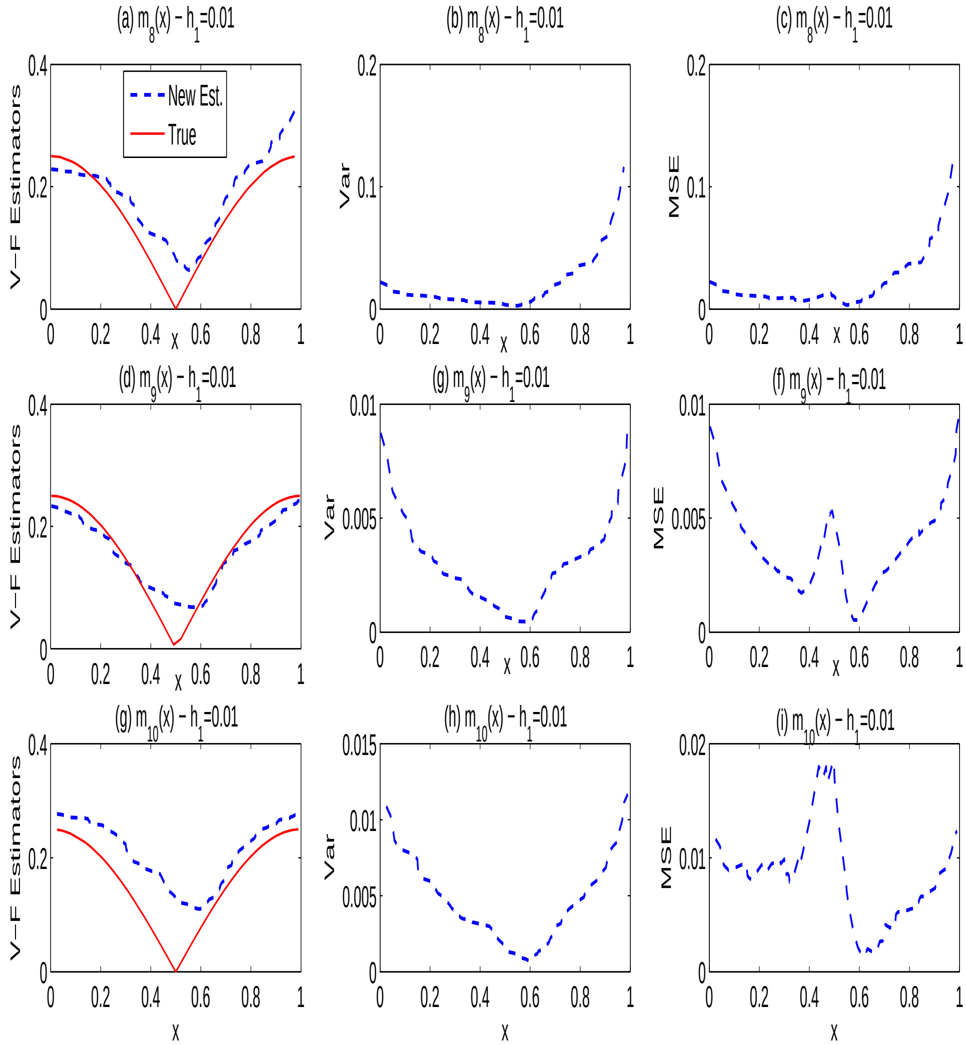

Figure 6.19: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.01$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

same when the mean function is $m_9(x)$. In the figures (6.17)-(6.20), the effect of the small selection of the bandwidth h_1 on the variance and the mean squared error of the estimator in (6.2) is very small. Generally, it is obvious from the figures (6.13)-(6.20) that the optimal finite sample performance of the estimator in (6.2) is obtained when $h_1 = 0.01$ or 0.0032.

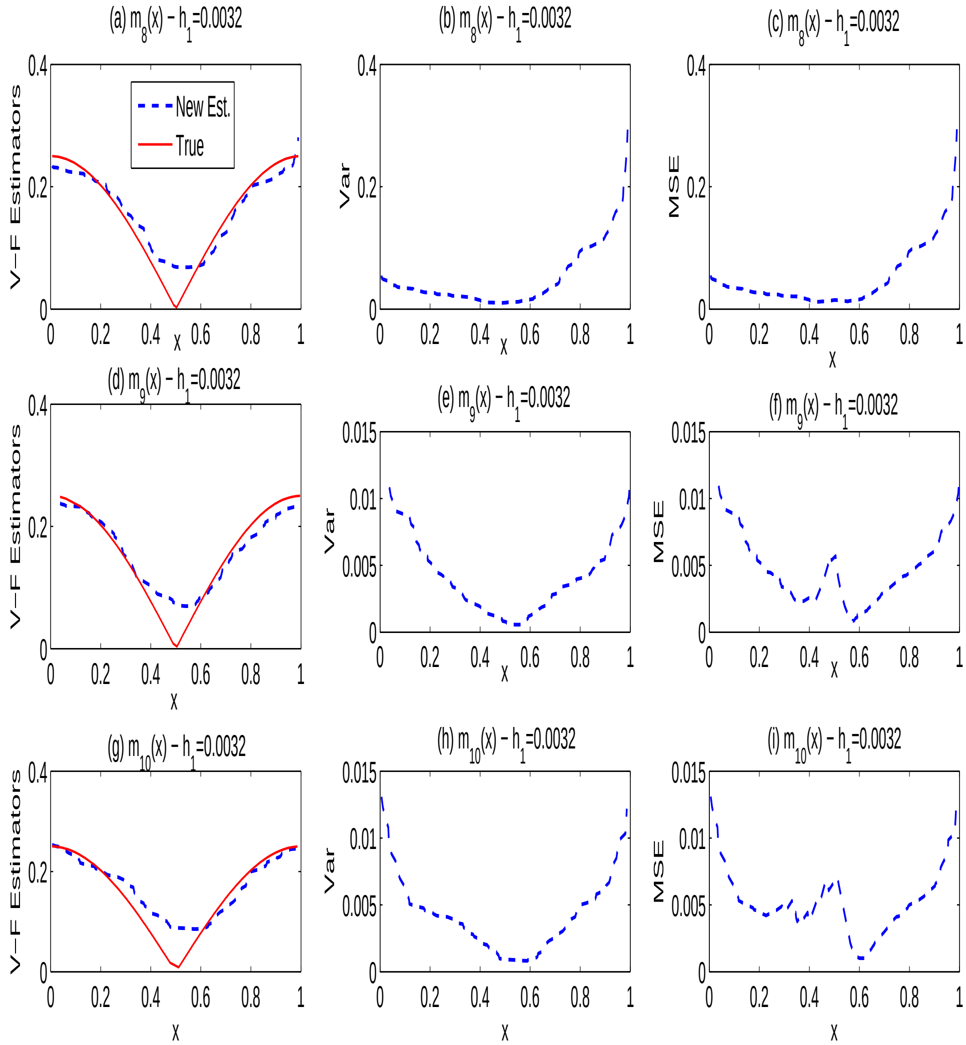

Figure 6.20: *The Effect of the Bandwidth* h_1 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.0032$ *and* $h_2 = 0.1$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

Discussions:

In this subsection, we study the bias in the estimated variance function using the estimator in (6.2) due to estimation of the mean function. So, the bandwidth h_2 of the estimator in (6.2) is supposed to be chosen optimally. From the results of the figures (6.13)-(6.20), we found that, when the mean function has rth continuous derivatives where $r \leq 10$ and the chosen value of the bandwidth h_1 is large, the bias of the estimated variance function is also very large. On the other hand, when h_1 is small, the bias is negligible as well as the mean squared error. So, we can conclude that the bandwidth h_1 should be chosen to minimise the bias of the estimated mean function.

However, when the mean function is a smooth function, we found that the effect of the bandwidth h_1 is negligible. Note that if the bandwidth h_1 is chosen to be too large, this will cause a bias in the estimated variance function using the estimator in (6.2) because the bias increases slightly as the chosen value for the bandwidth h_1 rises.

From the discussion above, we can conclude that small choices of the bandwidth h_1 are an appropriate selection for this bandwidth. In chapter 4, we found that the bandwidth h_1 does not have a first order effect. So, its contribution in the mean squared error of the estimator in (6.2) is negligible. The results in this subsection support this conclusion in the finite sample case. In addition to that, we found from the simulation studies in this subsection that any chosen value in the interval $(n^{-0.8}, n^{-1.3})$ is an appropriate choice of the bandwidth h_1 for the estimator in (6.2) for all considered models in this subsection. It should be noted that n refers to the chosen sample size, which is $n = 100$ for all simulation studies in this subsection. Note that if the bandwidth h_1 is chosen to be too small (less than $n^{-1.3}$), this will affect the finite sample performance of the estimator in (6.2) because the weight matrix in the estimation of the mean function becomes invalid. In other words, the summation of some rows in the weight matrix is zero instead of one, which is the correct summation.

Conclusion:

We can conclude that the effect of the bandwidth h_1 in the finite sample performance of the estimator in (6.2) is generally negligible for small choices of this bandwidth. However, large values for this bandwidth increase the bias of the estimated variance function and this bias is due to the bias in the estimation of the mean function. Thus, the bandwidth h_1

should be chosen to minimise the bias of the estimated mean function, $E\left[\hat{m}(x)\right] - m(x)$. These results support the results of Theorem 4.2.1 in chapter 4. The results of this subsection are only valid for optimal choices of the bandwidth h_2.

6.4.2 The Effect of the Bandwidth h_2 on the Finite Sample Performance of the Estimator in (6.2)

In the current subsection, we consider the influence of the bandwidth h_2 on the finite sample performance of the estimator in (6.2). To study this effect, we fix the mean function, the variance function and the bandwidth h_1 and then the simulation studies are generated for several values of the bandwidth h_2 with these fixed choices. The number of replications in this simulation studies is selected to be $N = 1000$ of sample size $n = 100$. As in the previous subsection, we study the effect of the bandwidth h_2 on the finite sample performance of the estimator in (6.2) with the same three mean functions and the same two variance functions. It should be noted that the simulation studies in this subsection have the same structure as described in section 6.2. An appropriate choice of the bandwidth h_1 is selected, which minimises the bias of the estimated mean function.

To find the effect of the bandwidth h_2, we start with the mean function $m_8(x)$ and the variance function $v_5(x)$. First, we select randomly x_is from the uniform $U[0,1]$ distribution and the ϵ_is from the standard normal distribution where the size of selected sample is $n = 100$. Then, we sort x_is into increasing order and then Y_is are generated using the model in (i) in equations (6.8). The observed values of the estimator in (6.2) are computed where the bandwidth $h_1 = 0.0032$ and the bandwidth $h_2 = 0.4$. Then, these steps are repeated for $N = 1000$ times. Then, the mean values of $\hat{v}_{New}(x_i)$s, their variance and their mean squared error are plotted versus the selected x_i in the plot (a), (b) and (c) in the figure (6.21), respectively. To draw the plots (d), (e) and (f) in the figure (6.21), the previous steps are repeated where the model in (ii) in equations (6.8) is used. The plots

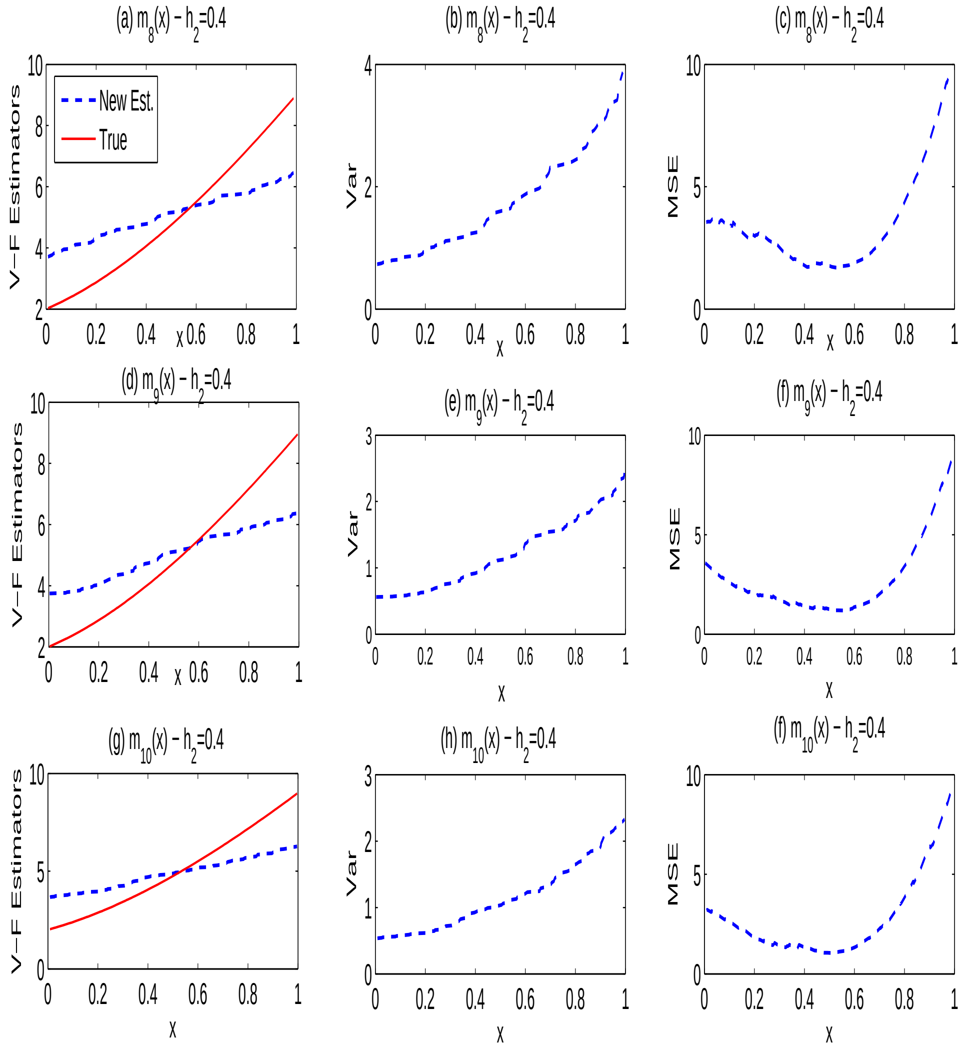

Figure 6.21: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$*,* $h_1 = 0.0032$ *and* $h_2 = 0.4$*, (Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

(g), (h) and (i) in the figure (6.21) are produced using the same steps where the model is in (iii) in equations (6.8).

For the figures (6.22)-(6.24), the same previous procedures are repeated where the Y_is are generated using the models in (6.8) and the bandwidth h_2 is chosen to be 0.15, 0.05 and 0.005, respectively. The figures (6.25)-(6.28) are produced using the same steps where the models

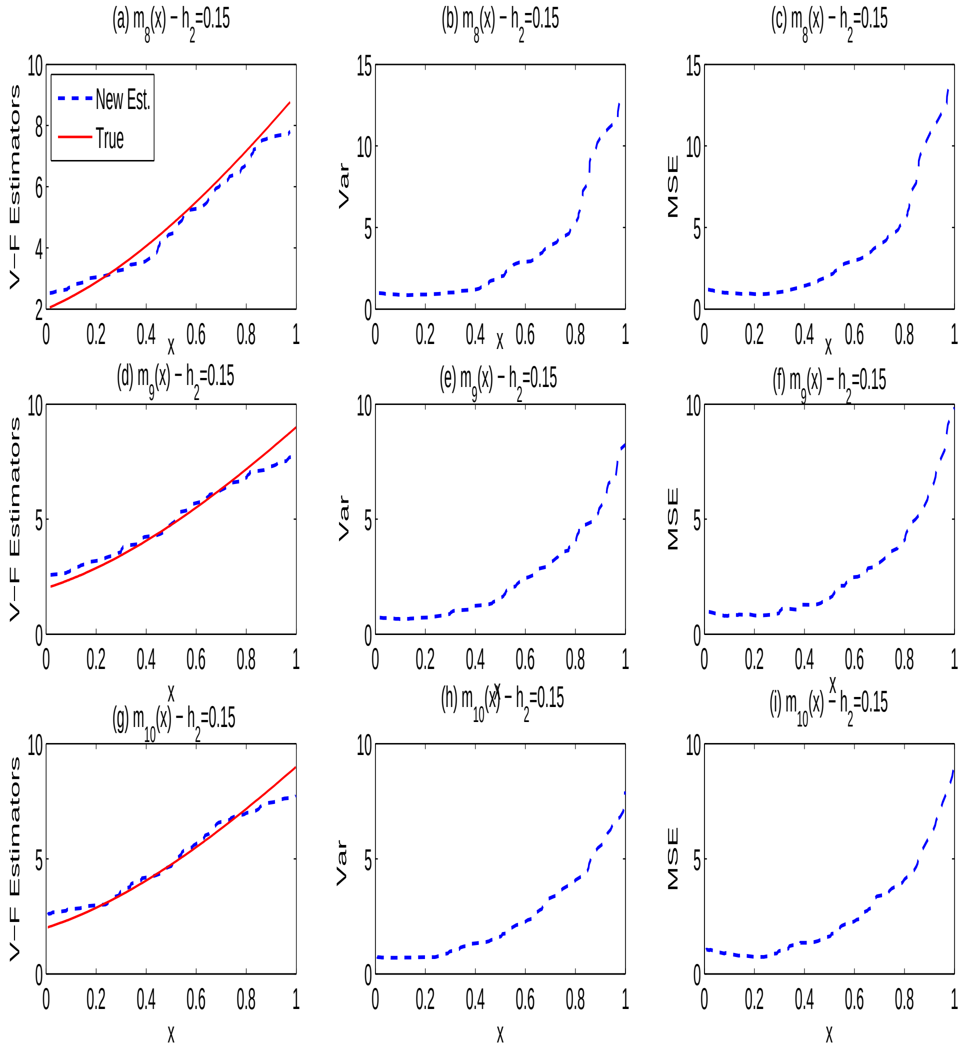

Figure 6.22: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$*,* $h_1 = 0.0032$ *and* $h_2 = 0.15$*, (Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

is in equations (6.9) and the bandwidth h_2 is taken as 0.4, 0.15, 0.05 and 0.005, respectively. In all these figures, the bandwidth h_1 is fixed to be 0.0032.

Results:

The main results in this subsection are described as follows. Recall that the bandwidth h_1 is chosen appropriately in the figures (6.21)-(6.28). It is obvious that the estimated variance

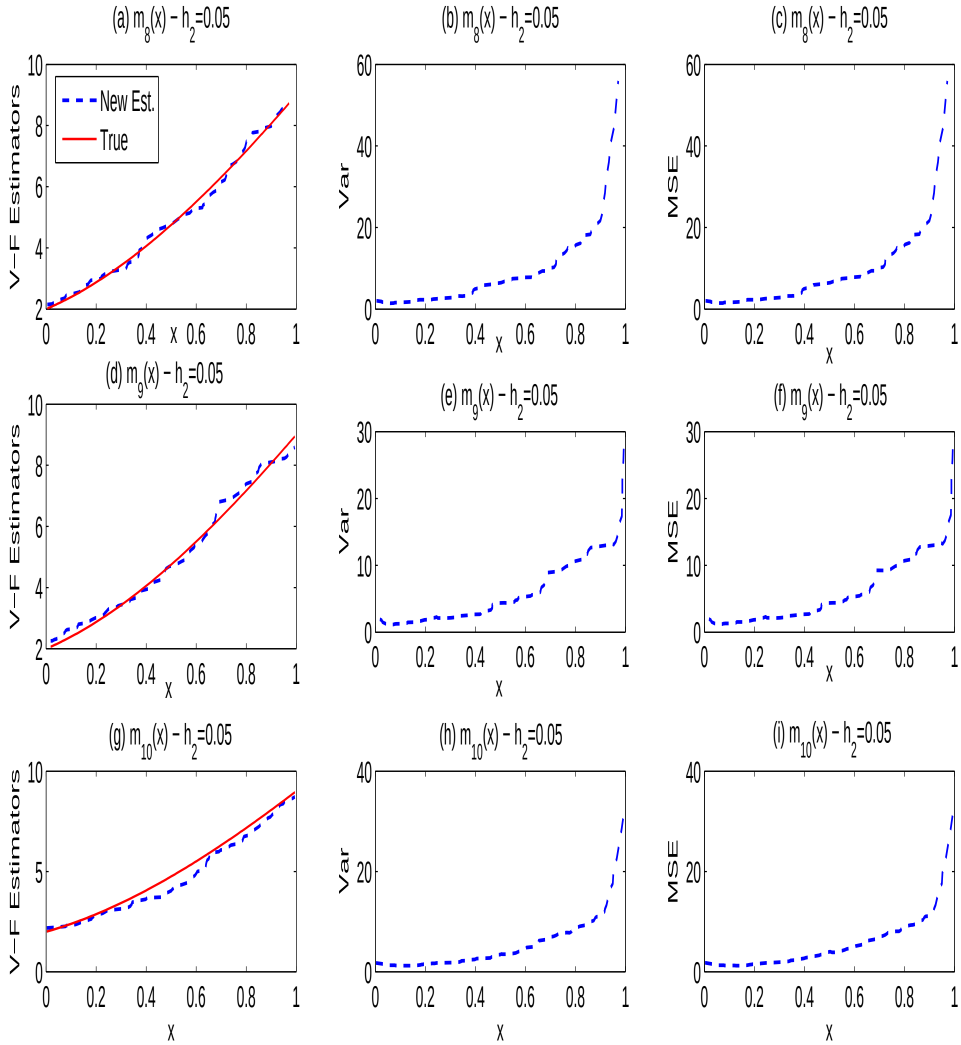

Figure 6.23: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$*,* $h_1 = 0.0032$ *and* $h_2 = 0.05$*, (Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

functions are biased in the figure (6.21), while the variance is small. On the other hand, in the figure (6.24), the bias is negligible, while the variance is very large. Hence, the mean squared error is also large. The estimated variance functions in the figure (6.22) are still a little biased, but its bias is smaller than that of the bandwidth $h_2 = 0.05$. Using the bandwidth $h_2 = 0.05$, the estimated and the true variance functions are approximately the same, but the variance is slightly high. Thus, from the figures (6.21)-(6.24), the optimal choice of the

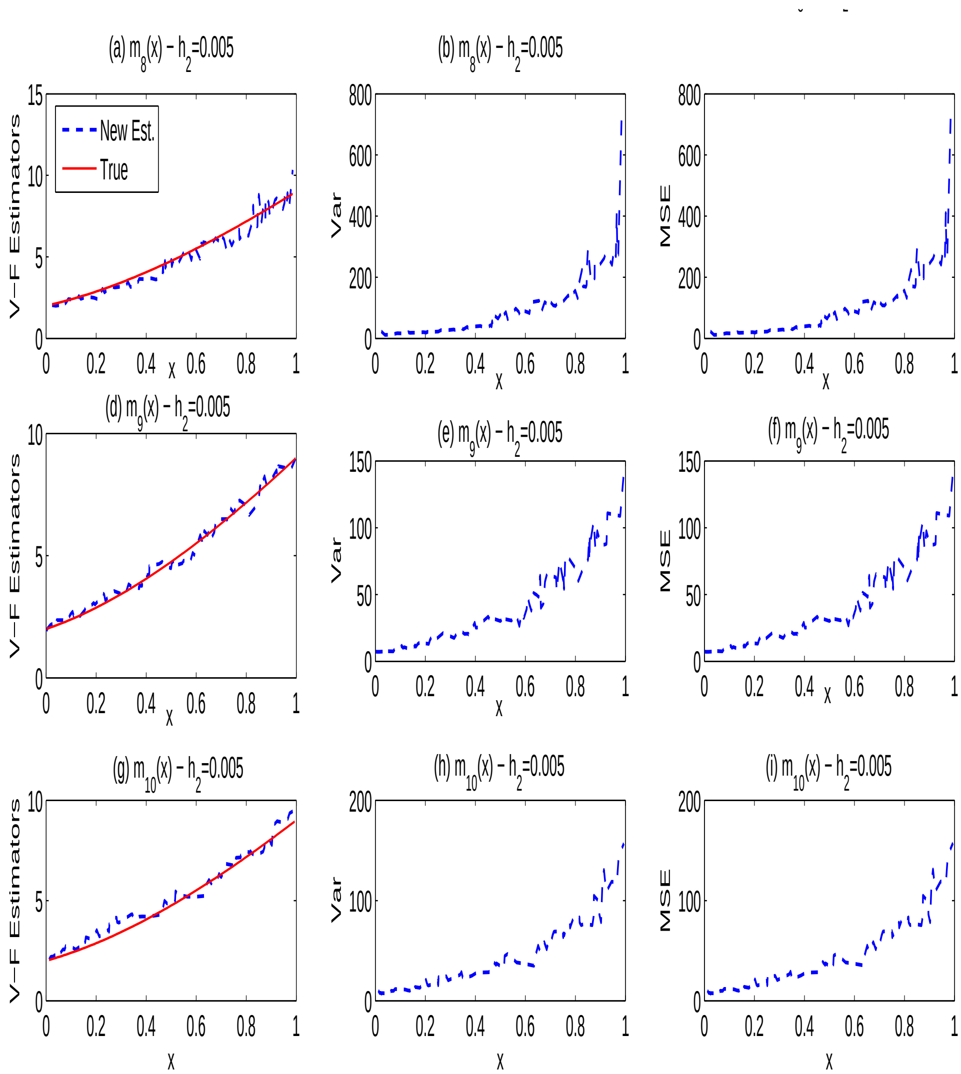

Figure 6.24: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_5(x)$, $h_1 = 0.0032$ *and* $h_2 = 0.005$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

bandwidth h_2 is approximately between 0.1 and 0.2, because it is balancing between the squared bias and the variance of the estimator in (6.2).

In the figures (6.25)-(6.28), the requirement of balancing between the squared bias and the variance is also very clear. When $h_2 = 0.4$, it is clear that the bias is large, whereas the variance is small. On the other hand, the variance is very high when $h_2 = 0.005$, while

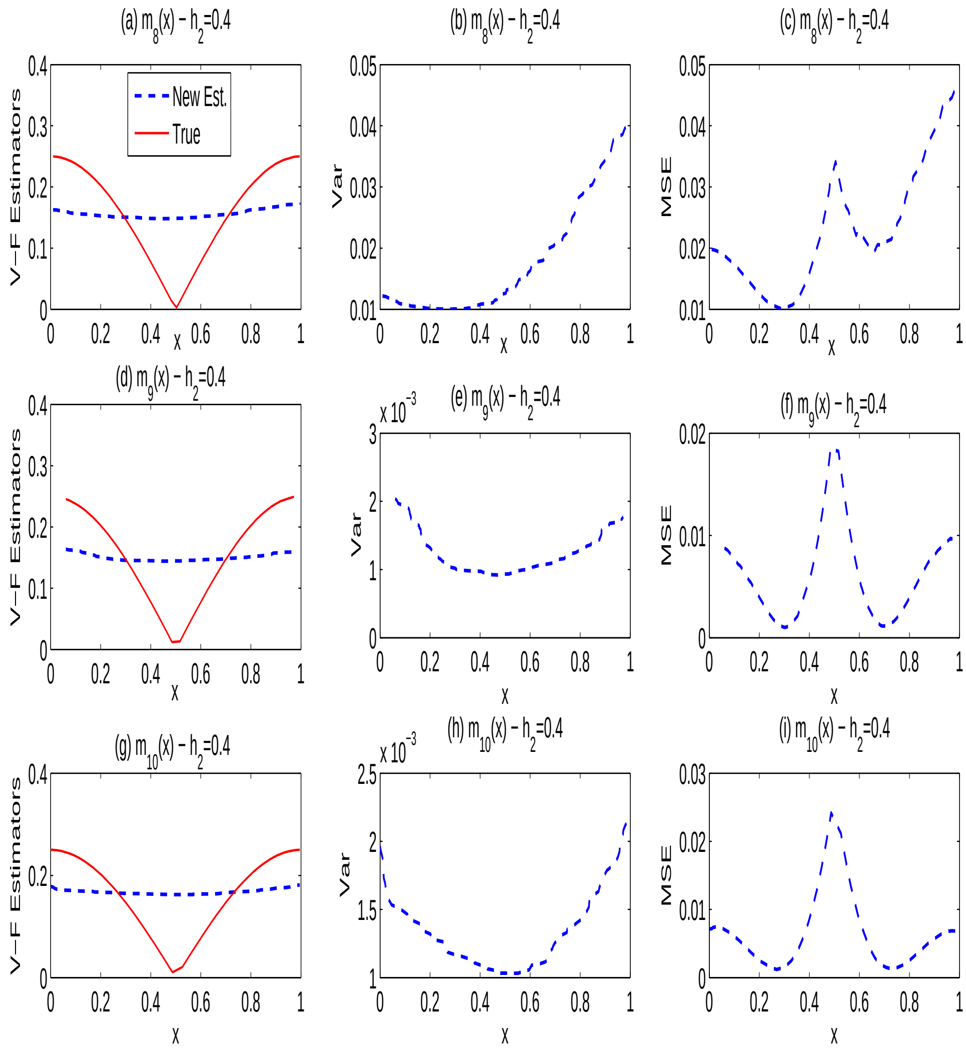

Figure 6.25: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.0032$ *and* $h_2 = 0.4$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

the bias is small. For the bandwidth $h_2 = 0.15$, the bias is slightly small, but the variance is slightly large. When the bandwidth $h = 0.05$, the bias is very small, but the variance is slightly larger than that when $h_2 = 0.15$. From the figures (6.25)-(6.28), it is evident that the optimal choice of the bandwidth h_2 is also between 0.1 and 0.2.

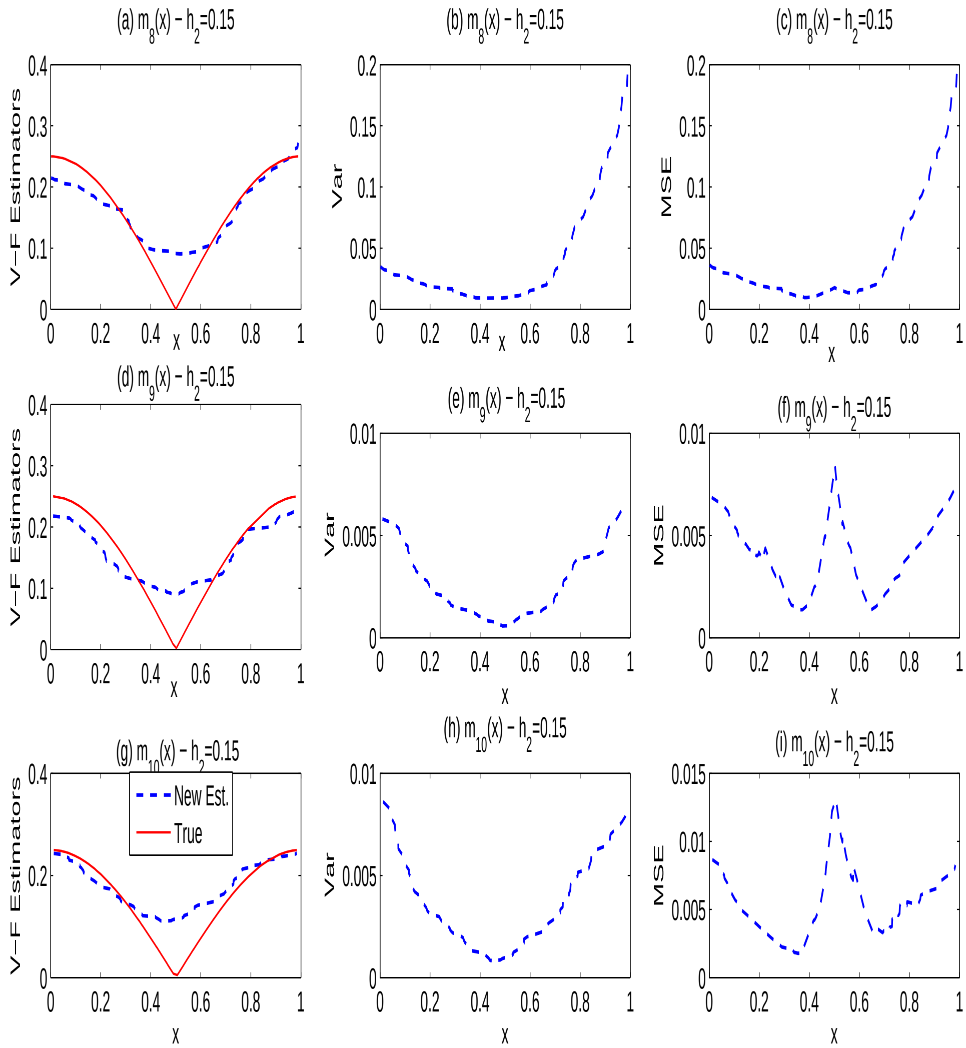

Figure 6.26: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$, $h_1 = 0.0032$ *and* $h_2 = 0.15$, *(Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

Discussions:

In this subsection, the effect of the bandwidth h_2 on the squared bias and the variance of the estimator in (6.2) is studied in the finite sample case through simulation. It should be noted that the bandwidth h_1 is selected appropriately in the figures (6.21)-(6.28). In these figures, for large values of the bandwidth h_2 and when the variance function is a polynomial regression function of order $r \leq 10$, the bias of the estimated variance function by

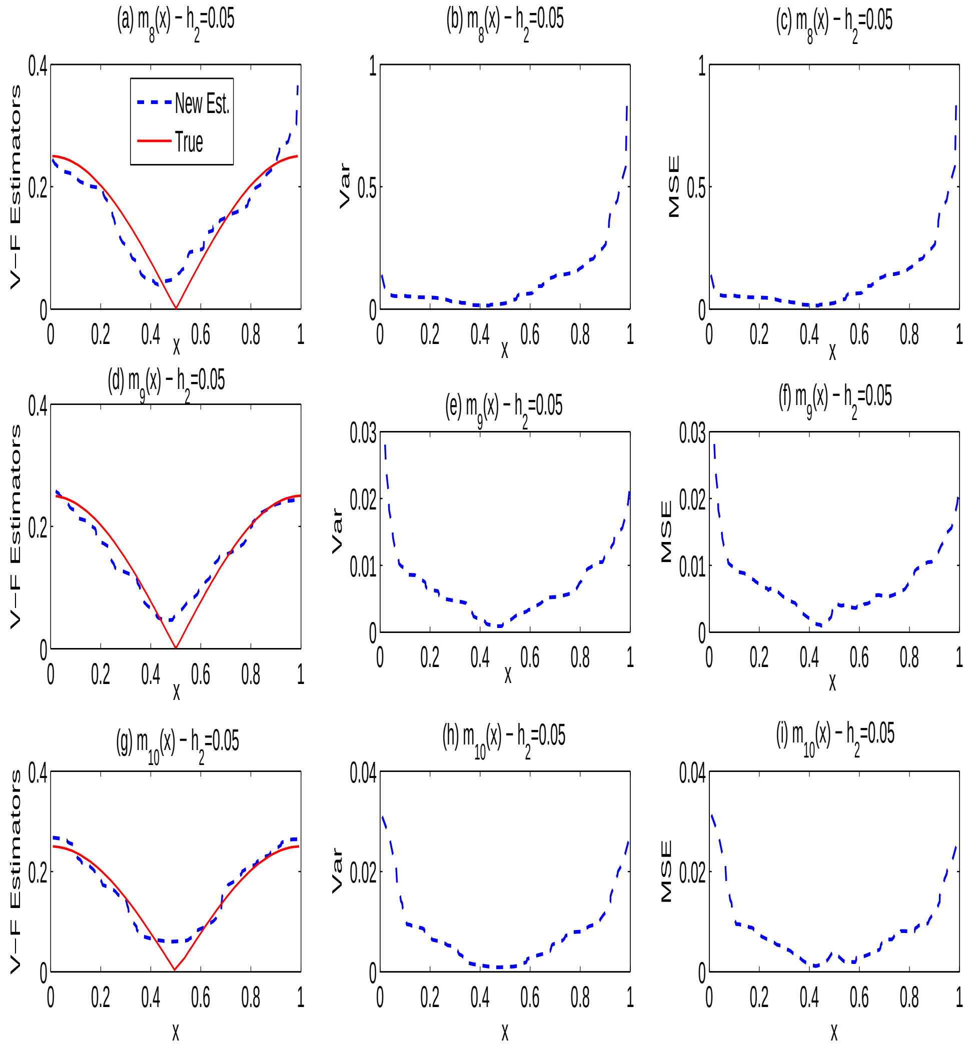

Figure 6.27: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$*,* $h_1 = 0.0032$ *and* $h_2 = 0.05$*, (Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

the estimator in (6.2) is large. This bias becomes smaller as the bandwidth h_2 decreases. On the other hand, when the bandwidth h_2 is small, the bias is clearly negligible, but the variance is large. Thus, when the variance function is a polynomial regression function of order $r \leq 10$, the optimal chosen value of the bandwidth h_2 should be balancing between the squared bias and the variance of the estimator in (6.2).

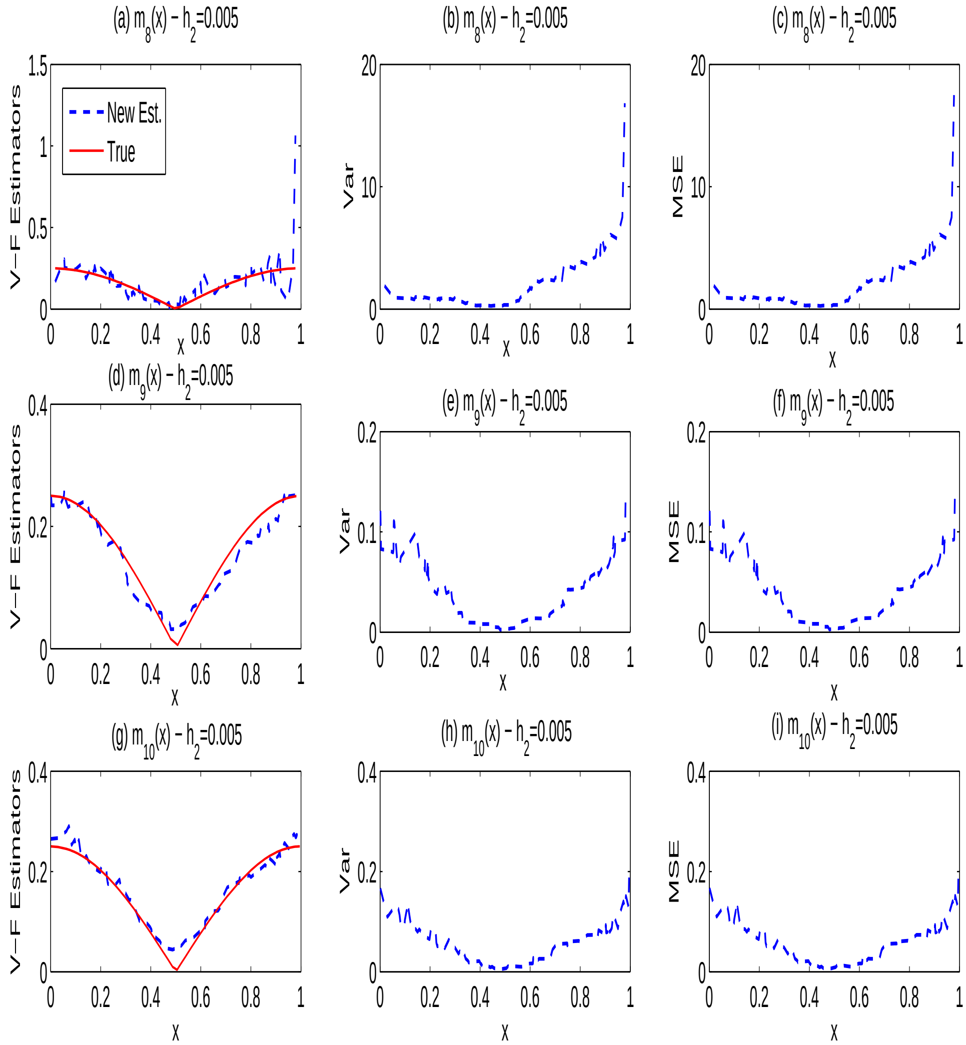

Figure 6.28: *The Effect of the Bandwidth* h_2 *on the Behaviour of the New Estimator where* $v(x) = v_6(x)$*,* $h_1 = 0.0032$ *and* $h_2 = 0.005$*, (Left: Estimated and True variance functions, Middle: the Variance and Right: The MSE).*

For a smooth variance functions, the bias in the estimated variance function by the estimator in (6.2) is very obvious when the bandwidth h_2 is large. Clearly, this bias becomes smaller as the bandwidth h_2 decreases. However, when the bandwidth h_2 is small, the variation is very clear in the estimated variance function. Thus, when the variance function is a smooth function, small choices of the bandwidth h_2 lead to a small bias and a large variance as in the usual smoothing problems. In contrast, large selections of the bandwidth

h_2 increase the bias and decrease the variance of the estimator in (6.2).

From the discussion above, the chosen mean or variance functions may have effect on the finite sample performance of the estimator in (6.2). However, this effect of the mean and variance functions can be minimised by choosing an optimal value for the bandwidth h_2. In addition, note that the bandwidth h_2 has range of the optimal selected values. If the chosen value is bigger than the maximum point in this range, this chosen value will cause bias in the estimated variance function by the proposed estimator in (6.2). On the other hand, when the selected value is less than the minimum point in this range, it will lead to large variance in the estimated variance function.

From Theorem 4.2.1, it is clear that the bandwidth h_2 has first order effects on the mean squared error of the estimator in (6.2). The results of the finite sample performance of the estimator in (6.2) in this subsection clearly support this conclusion. Thus, large values of the bandwidth h_2 lead to a large mean squared error for the estimator in (6.2) because they increase the bias of the estimated variance function by this estimator. However, small values of the bandwidth h_2 also increase the mean squared of the estimator in (6.2) because they rise the variance of this estimator. From the above discussion, we can conclude that the bandwidth h_2 should be chosen to be balancing between the squared bias and the variance of the estimator in (6.2). In addition, it is clear that the chosen value of the bandwidth h_2 plays important influence on the finite sample performance of the estimator in (6.2).

Conclusion:

We can conclude that the effect of the bandwidth h_2 on the finite sample performance of the estimator in (6.2) is generally very clear and plays important roles on the finite sample performance of this estimator . So, this bandwidth should be chosen to be balancing between the squared bias and the variance of the estimator in (6.2). This conclusion is applied for all chosen mean and variance functions. These results are only valid when the bandwidth h_1 is chosen appropriately.

6.5 The Age-Blood Pressure Data

The behaviour of the estimator in (6.2) on the real data set is one of the most important issues to assess the performance of this estimator. In the current section, we consider a real data set to test the performance of the estimator in (6.2). In the following subsection, this data is described, whereas the assessments of the performance of the estimator in (6.2) are looked at in subsection 6.5.2.

6.5.1 The Description of the Age-Blood Pressure Data

A brief description of the age-blood pressure data is given in this subsection. This data set is reported in the Applied Linear Statistical book by Neter, Kutner, Nachtsheim and Wasserman (1996). This data consists of two variables, which are the age and blood pressure of 54 women. The aim was to find the relationship between the age and blood pressure. For more details about this data, refer to Neter, Kutner, Nachtsheim and Wasserman (1996). The scatter plot of the age-blood pressure data is demonstrated in the figure (6.29). The age is counted by years, while the blood pressure is measured by millimetres of mercury (mmHG). From this figure, it is obvious that the variation in the blood pressure variable increases as the age rises. This suggests that the variance of the blood pressure variable is not constant. Thus, the relationship between the age and blood pressure variables can be found using the following heteroscedastic nonparametric regression model. Suppose that

$$Y_i = m(x_i) + \sqrt{v(x_i)}\,\epsilon_i, \quad \text{for } i = 1, 2, \cdots, 54 \tag{6.10}$$

where $m(x_i)$ represents the unknown mean function $E(Y_i|x_i)$, Y_is denote the blood pressure variable, $v(x_i)$ represents the unknown variance function at the point x_i and x_is denote the age variable. Note that the errors ϵ_is are assumed to be normally distributed random variables with zero mean and unit variance. In the following subsection, our aim is to estimate the unknown error variance function $v(x_i)$ using the estimator in (6.2).

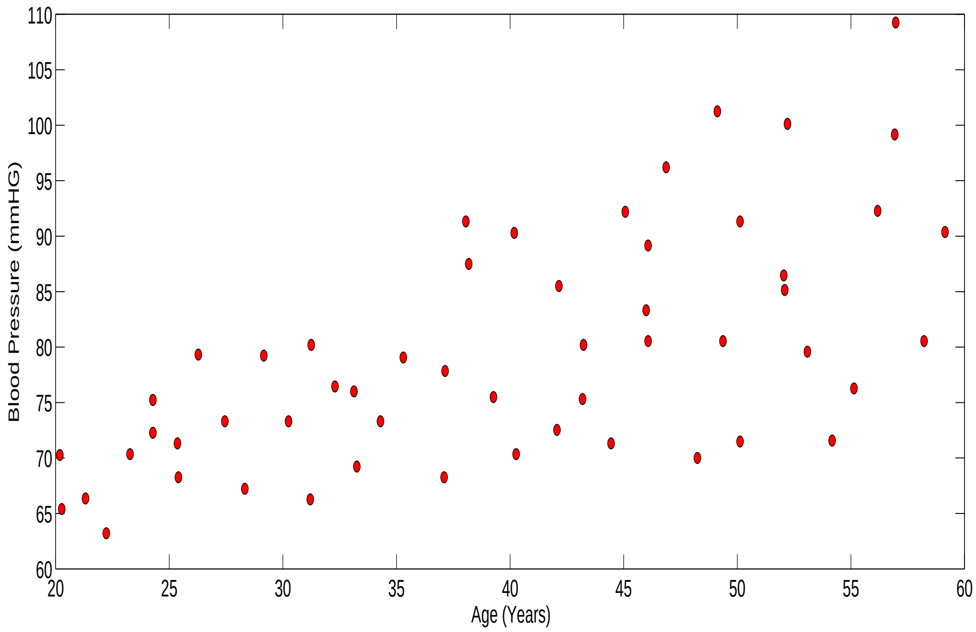

Figure 6.29: *The Scatter Plot of the Age-Blood Pressure Data*

6.5.2 The Estimation of the Error Variance Function in the Model (6.10)

In this subsection, we explain the way to estimate the unknown variance function in the model (6.10). First, the age variable is adjusted such that

$$Age_{adj} = age\,/\,(\text{Max}(age) + 1\,).$$

Using the adjusted age variable, the estimator in (6.2) is used to estimate the variance function where $h_1 = 0.034$, $h_2 = 0.1$ and Y_i denotes the blood pressure, whereas the design points x_is are the adjusted age points. Thus, the estimator in (6.2) produces a variance for every adjusted age point. To find the estimated variance function, the raw age variable is plotted versus the estimated variance points. To compare the performance of the estimator in (6.2) to that of the Brown and Levine estimators, the estimated variance functions by the Brown and Levine estimators are also plotted in the same figure for several differences' orders where the bandwidth $h_2 = 0.15$ and the orders of differences are 2, 4 and 6. These plots are demonstrated in the figure (6.30).

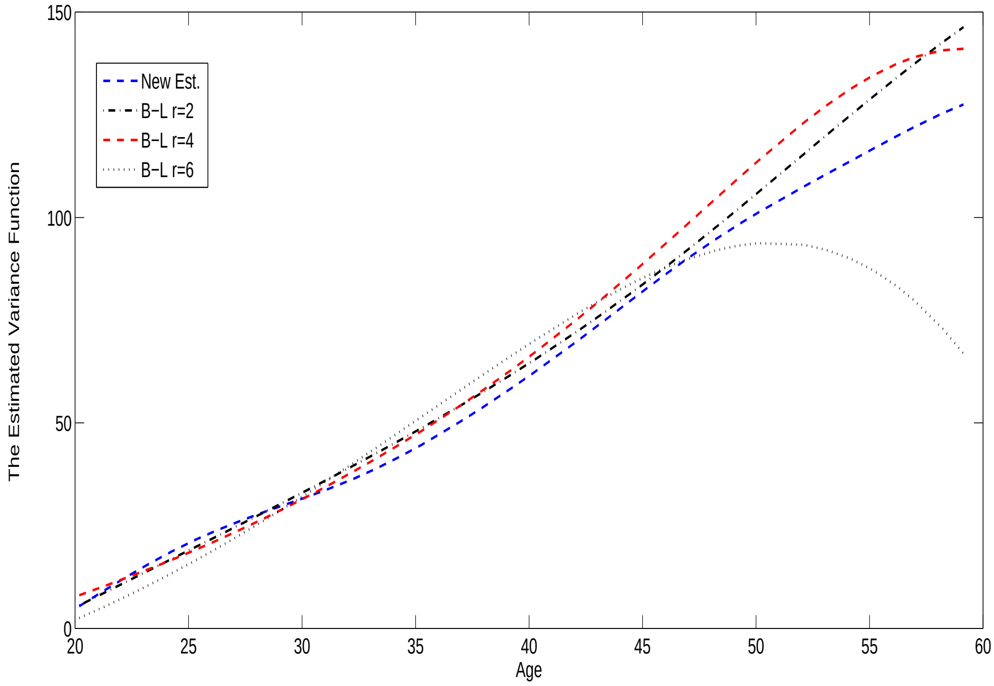

Figure 6.30: *The Estimated Variance Functions for the Age-Blood Pressure Data Using the Estimator in (6.2) and the Brown and Levine Estimators*

Remark:

If the age variable is not adjusted, the results are still the same, but suitable bandwidths should be chosen.

Clearly from figure (6.30), the estimated variance functions are approximately the same for all estimators until age 48. After age 48, the estimated variance functions are different. Obviously, in the figure (6.29), the variation in the blood pressure variable increases as the age rises. However, the estimated variance function by the Brown and Levine estimator decreases after age 47 when the order of differences is 6. This may be because of the lack of information in the boundary due to using differencing to estimate the variance function. Thus, this part of the estimated variance function by this estimator is unreliable. For the remaining estimators, we expect that their performances are reasonably good. It is not easy to assess this, because the true variance function is unknown. Note that the curve of the estimated variance function by the estimator in (6.2) for large ages is slightly lower than the curves of the estimated variance functions by the Brown and Levine estimators when the orders of differences are 2 and 4. Thus, we can conclude that the general performances of the estimator in (6.2) and the Brown and Levine estimators for second and fourth orders of differences are approximately the same.

6.6 Summary

In this chapter, we studied the effect of the mean function on the finite sample performance of the estimator in (6.2). We can conclude that the effect of the mean function is less than that of the variance function. Furthermore, the performance of the estimator in (6.2) in the interior points is better than, or the same as, that of the Brown and Levine estimators. In the boundary points, the performance of the estimator in (6.2) is better than that of the Brown and Levine estimators except when the mean function is a polynomial function of order ≥ 3 and the variance function is a smooth function. This lack of information in the boundary points of the Brown and Levine estimators increases as the order of differences rises.

For the bandwidths of the estimator in (6.2), we found that the bandwidth h_1 should be chosen to minimise the bias of the estimated mean function. On the other hand, the bandwidth h_2 should be selected to be balancing between the squared bias and the variance of the estimator in (6.2), because large choices of this bandwidth lead to a large bias in the estimated variance function. However, small choices of the bandwidth h_2 rise the variance of the estimated variance function by the estimator in (6.2). In the chosen real data set, we can conclude that the results for the considered estimators are approximately the same.

Chapter 7

Conclusion and Future Work

7.1 Introduction

In this thesis, the error variance estimation is considered in the settings of constant and functional variance nonparametric regression models. First, recall that to estimate the error variance, our proposal is to average $e_i\,Y_i$s as opposed to the averaging of e_i^2s, which is used in the residual-based estimators. That is, in the setting of constant variance model,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}(x_i)\right) Y_i \tag{7.1}$$

and when ith observation is not used in the estimation of the mean function, the above class of estimators becomes

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}_{-i}(x_i)\right) Y_i. \tag{7.2}$$

The class of estimators in (7.2) can be extended to be used when the error variance is a function of x_is. That is,

$$\hat{v}(x) = \sum_{i=1}^{n} w_i(x)\ \left(Y_i - \hat{m}_{-i}(x_i)\right) Y_i, \tag{7.3}$$

where $w_i(x)$ is a weight function.

Our interest is to seek answers to the questions $-$ about whether the smoothing of $e_i Y_i$ has any advantage over smoothing e_i^2 or using differencing, and whether not using observation Y_i in the estimation of $m(x_i)$ has any advantage over using it.

7.2 The Main Results

With respect to the mean squared analysis, the new estimator, which is defined in (1.28) and the residual-based estimators have a similar behaviour in the first order. In other words, the estimator in (1.28) and the Hall and Marron estimator have achieved the same optimal rate in the first order, which is

$$MSE(\hat{\sigma}^2) = n^{-1} \, var(\epsilon^2) + o(n^{-1}). \tag{7.4}$$

The above optimal rate is not achieved by the fixed order difference-based estimators. It was noted that the optimal bandwidth for the estimator considered here is approximately the square of the optimal bandwidth for the Hall and Marron estimator. For the second order kernel function, the optimal bandwidth to estimate the mean function with respect to the mean squared error is $n^{-1/5}$, whereas the optimal bandwidth of the estimator in (1.28) is $n^{-2/5}$, which means that it estimates the mean function with smaller bias. This property is often desirable as Wang, Brown, Cia and Levine (2008) have noted. In contrast, the Hall and Marron estimator estimates the mean function almost optimally since its optimal bandwidth is $n^{-2/9}$, which is closer to $n^{-1/5}$. The Hall and Marron estimator has smaller relative error in the second order than the estimator in (1.28). However, the relative error in the second order does not play a role in the finite sample behaviour as Dette, Munk and Wanger (1998) and Tong, Liu and Wang (2008) have noted. To put things in perspective, ignoring the constants, it was noted that

$$\frac{\text{MSE(Hall-Marron Estimator)}}{\text{MSE (New Estimator)}} = \frac{n^{-1}[1+(n^{-7/9})]}{n^{-1}[1+(n^{-3/5})]} = \frac{1+(n^{-7/9})}{1+(n^{-3/5})}.$$

The above ratio is approximately 0.97 for $n = 100$. In fact, it means, as observed in chapter 3, that the constants have more influence on the mean squared error in the finite sample behaviour than the second order.

In the finite sample, we investigated the performance of the estimator in (1.28) through the simulation studies. When the bandwidths of the estimator in (1.28) and that of the Hall and Marron estimator are chosen optimally, neither of the estimators is better than the other across all mean functions and different noise levels. Thus, for the small error variances, when the mean function is a periodic function or a polynomial function of order ≥ 3, the difference in the variances of the estimator in (1.28) and the Hall and Marron estimator is due to the constants.

To summarise, in the estimator in (1.28), which is based on the average of $e_i Y_i$, to get the best possible mean squared error for this estimator, the optimal bandwidth is approximately $n^{-2/5}$. This means that one is not required to estimate the mean function optimally. On the other hand, in the Hall and Marron estimator, based on the average of e_i^2, to get the smallest possible mean squared error, the optimal bandwidth is roughly $n^{-2/9}$. Thus, the mean function in the estimator in (1.28) is estimated with smaller bias compared to that of the mean function estimator used in the Hall and Marron estimator.

In the theoretical investigation, we showed the asymptotic normality of the distribution of the estimators in (1.28) and (1.29). We also proved that the estimator in (1.28) and the Hall and Marron estimator have the same asymptotic distribution. Also, the results of the simulation studies in chapter 3 provide a clear numerical verification of the normality of the estimator in (1.28). The asymptotic distribution of the estimator in (1.29) and that of the local linear version of the Brown and Levine estimator are approximately the same, where the order of differences is 2.

With respect to the mean square property, the estimator in (1.29) and the other error variance function estimators have a similar behaviour. To clarify, the form of the mean squared errors of the error variance function estimators, which include the estimator in (1.29), is

$$MSE(\hat{v}(x)) = h_2^{2r} C_1^2(x) + n^{-1} h_2^{-1} C_2(x) + o(n^{-1} h_2^{-1}) + o(h_2^{2r}).$$

where $C_1(x)$ and $C_2(x)$ are deterministic functions. So, the difference between these estimators, with respect to the mean squared error, is in the deterministic functions $C_1(x)$ and $C_2(x)$. We showed that the mean squared error of the estimator in (1.29) in the first order is the same as that of the local linear version of the Brown and Levine estimator (for the second order of differences only). Interestingly, the mean squared error of the estimator in (1.29) depends only on the bandwidth h_2, which is used to estimate the variance function. So, the bandwidth h_1 does not have a first order effect on the mean squared error of the estimator in (1.29). We also proved that the MSE-optimal selection of the bandwidth h_2 is approximately $n^{-1/2r+1}$. So, when the second order kernel is used, the MSE-optimal choice becomes $h_2 \sim n^{-1/5}$.

In the investigation of the finite sample performance of the error variance function estimators, we conclude that the estimator in (1.29) performs better than the Brown and Levine estimators in the boundary except when the mean function is a polynomial function of order ≥ 3 and the variance function is a smooth function. The performance of the estimator in (1.29) is approximately better than, or the same as, that of the Brown and Levine estimators in the interior points. We also found that the bandwidth h_1, which is used to estimate the mean function, should be chosen to minimise the bias of the estimated mean function, $E[\hat{m}(x)] - m(x)$. However, the bandwidth h_2 should be selected to be balancing between the squared bias and the variance of the estimator in (1.29). The effect of this bandwidth on the finite sample performance of the estimator in (1.29) is larger than that of the bandwidth h_1. In all considered cases in chapter 6, the effect of the variance function on the finite sample performance of the estimator (1.29) is larger than that of the mean function. To generalise this conclusion, we require to carry out more investigation regarding all different forms of the mean and variance functions. The results of the finite sample behaviour of the estimator in (1.29) support the conclusion of Theorem 4.2.1. In summary, one of the advantages of smoothing $e_i Y_i$s over using differencing is that there is no lack of information on the estimated variance function $\hat{v}(x)$ near the boundary. The estimation of the mean function does not have a first order effect on the mean squared error of the estimator in (1.29). So, the effect of using, or not using, observation Y_i in the estimation of the mean function is negligible.

7.3 Future Work

In the current thesis, the properties of the new estimators for the error variance in the settings of the constant and functional (when the error variance is a function of x_is) variance models are investigated. In particular, in these two settings, we carried out the asymptotic mean squared error analysis for the new estimators and we established their asymptotic normality. So, it will be of interest to study the finite sample performance of the data-based bandwidth selection methods in these two settings. In the current thesis, in both settings, we considered the univariate case. Therefore, it will be of interest to generalise the new estimators, which are defined in (1.28) and (1.29), to be used in a multivariate case. For that, consider the following multivariate nonparametric regression model where the error variance is a constant,

$$\underline{Y}_i = m(\underline{x}_i) + \underline{\epsilon}_i \qquad \text{for } i = 1, 2, \cdots n, \tag{7.5}$$

where m denotes the mean function $E(\underline{Y}_i|\underline{x}_i)$, $\underline{Y}_i$s represent the response variable, $\underline{x}_i$s denote the design points. The errors $\underline{\epsilon}_i$s are assumed to be independent, identically distributed and random with zero mean and variance σ^2. It should be noted that the index i is a d-dimensional index such that $i = (i_1, i_2, \cdots i_d)$ and $\underline{x}_i = (x_1, x_2, \cdots x_d)'$ where d denotes the number of the dimensions. For simplicity, the design points $\underline{x}_i$s are assumed to be an equispaced d-dimensional grid. In this case, each coordinate can be defined as $x_{i_k} = \frac{i_k}{n}$ where $i_k = 1, 2, \cdots n$ for $k = 1, 2, \cdots d$. Thus, the overall sample equals to $s = n^d$. Using the equispaced d-dimensional grid, the estimator of the error variance in the model (7.5) can be written as

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} \underline{Y}_i^2 - \frac{1}{n\,(n-d)\,h^d}\sum_{i=1}^{n}\sum_{j\neq i} K\left(\frac{\underline{x}_i - \underline{x}_j}{h}\right)\underline{Y}_i\underline{Y}_j \tag{7.6}$$

where $K(.)$ is a symmetric multivariate kernel function. It should be noted that the error variance σ^2 and the bandwidth h are assumed to be the same in all dimensions. When the error variance is a function of $\underline{x}_i$s, assume the following multivariate nonparametric regression model

$$\underline{Y}_i = m(\underline{x}_i) + \sqrt{v(\underline{x}_i)}\,\underline{\epsilon}_i \qquad \text{for } i = 1, 2, \cdots n, \tag{7.7}$$

where $\underline{Y}_i$s, m and $\underline{x}_i$s are as in the model (7.5), $\underline{\epsilon}_i$s are independent with zero mean and unit variance, $v(\underline{x}_i)$s represent the variance function and the absolute fourth moment of errors is bounded. Now, to estimate $v(\underline{x}_i)$ using a methodology similar to that employed in this thesis in the equispaced d-dimensional grid, we define a new estimator to be

$$\hat{v}(\underline{x}) = \frac{1}{n\,h_2^d}\sum_{i=1}^{n} K\left(\frac{\underline{x}-\underline{x}_i}{h_2}\right)\left\{\underline{Y}_i - \frac{1}{(n-d)h_1^d}\sum_{j\neq i} K\left(\frac{\underline{x}_i-\underline{x}_j}{h_1}\right)\underline{Y}_j\right\}\underline{Y}_i \tag{7.8}$$

where h_1 is used to estimate the mean function, whereas h_2 is used to estimate the variance function. These bandwidths are assumed to be the same in all dimensions.

For the estimators in (7.6) and (7.8), it will be of interest to investigate their asymptotic properties. In particular, the mean square analysis will be carried out. This analysis will be of help to find their MSE-optimal bandwidths. Also of interest will be the study of their asymptotic distributions. In addition, it will be of interest to study the finite sample properties of these estimators. So, one can investigate the effect of the mean functions and bandwidth selections on the finite sample performances of the estimators in (7.6) and (7.8). In addition, issues related to data-based bandwidth selection methods can be studied.

Appendix A

The R Commands of the Functions for the Figures (3.1) until (3.12)

To do the plots of the mean functions in the figure (3.1) and (3.2), the following R codes are used:

```
n=1000
x1=matrix(c(runif(n,0,1)),nrow=n)
x=matrix(c(sort(x1)),nrow=n)
b2=matrix(c(rep(1,n)),nrow=n)
fx1=b2
b3=matrix(c(rep(4.7,n)),nrow=n)
fx2=b3+(2.4 *x)+(5 *x^2)+(4.3 *x^3)
fx3=matrix(c(rep(0,n)))
for( i in 1:n){
if (x[i] <=0.5) fx3[i]=3+x[i]+(4*x[i]^2)+(8*x[i]^4)
else fx3[i]=5.875-x[i]-(x[i]^2)-(x[i]^3)}
fx4=1/(exp(2+4*x+5*x^2+6*x^3))
par(mfrow=c(2,2))
plot(x, fx1, col=c(2), type = "l", xlab = "x", ylab = "m_1(x)")
title(main="m_1(x)")
plot(x, fx2, col=c(2), type = "l", xlab = "x", ylab = "m_2(x)")
title(main="m_2(x)")
plot(x, fx3, col=c(2), type = "l", xlab = "x", ylab = "m_3(x)")
title(main="m_3(x)")
plot(x, fx4, col=c(2), type = "l", xlab = "x", ylab = "m_4(x)")
title(main="m_4(x)")
```

For the figure (3.2), we use the R codes below

```
n=1000
x1=matrix(c(runif(n,0,1)),nrow=n)
x=matrix(c(sort(x1)),nrow=n)
fx5= (4/5)*sin(2*pi*x)
```

```
fx6=(3/4)*cos(10*pi*x)
par(mfrow=c(2,1))
plot(x, fx5, col=c(2), type = "l", xlab = "x", ylab = "m_5(x)")
title(main="m_5(x)")
plot(x, fx6, col=c(2), type = "l", xlab = "x", ylab = "m_6(x)")
title(main="m_6(x)")
```

The following four functions are required in the functions for the figures (3.3) until (3.10). The first function is to compute $K\left(\frac{X_i - X_j}{h}\right)$. This function can be written as

```
xker1=function(x,n,h) {
tx=t(x)
xker=matrix(c(rep(0,n^2)),nrow=n)
for (i in 1:n){
for (j in 1:n)
xker[i,j]=dnorm((x[i,1] - tx[1,j])/h)}
xker}
```

The second function is to find $K\left(\frac{X_i - X_j}{h}\right) Y_j$, which is:

```
xkery1=function(x,y,n,h) {
tx=t(x)
ty=t(y)
xker=matrix(c(rep(0,n^2)),nrow=n)
ff=matrix(c(rep(0,n^2)),nrow=n)
for (i  in  1:n)
for (j in  1:n) {
xker[i,j]=dnorm((x[i,1] - tx[1,j])/h) }
for (i  in  1:n)
for (j  in  1:n){
ff[i,j]=xker[i,j]*ty[1,j] }
ff }
```

The third function is to calculate $\sum\limits_{j \neq i} w_{ij} Y_i Y_j$ where

$$w_{ij} = \frac{K\left(\frac{X_i - X_j}{h}\right)}{\sum\limits_{i \neq j} K\left(\frac{X_i - X_j}{h}\right)}, \qquad 1 \leq i\ , j\ \leq\ n.$$

This function can be demonstrated as

```
xkery2new1=function(x,y,n,h){
tx=t(x)
ty=t(y)
```

```
xker=matrix(c(rep(0,n^2)),nrow=n)
ppp=matrix(c(rep(0,n^2)),nrow=n)
for (i  in  1:n)
for (j  in  1:n){
xker[i,j]=dnorm((x[i,1] - tx[1,j])/h) }
xker1=xker - diag(dnorm(0),n,n)
ddd=xker1/(apply(xker1,1,sum))
for (i  in  1:n)
for (j  in  1:n){
ppp[i,j]=ddd[i,j]*ty[1,j]*y[i,1] }
ppp }
```

The fourth function is to estimate the density function using a kernel function, $f_h\hat{(}x) = \frac{1}{n\,h}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$. This function can be written as

```
kden=function(x,h, nn ){
tx=t(x)
ggg=matrix(c(rep(0,nn^2)),nrow=nn)
for (i  in  1:nn)
for(j  in  1:nn){
ggg[i,j]=dnorm((x[i,1]-tx[1,j])/h) }
kerden=(1/(nn * h))*(apply(ggg,1,sum))
kerden  }
```

The function for the figure (3.3) is as follows

```
conhm=function(h,n,nn,sig){
sigmane=matrix(c(rep(0,nn)))
sigmahm=matrix(c(rep(0,nn)))
b1=matrix(c(rep(1,n)),nrow=n)
for(k  in  1:nn) {
x=matrix(c(runif(n,0,1)),nrow=n)
y=b1+ matrix(c(rnorm(n,mean=0,sd=sig)),nrow=n)
yy1=xkery2new1(x=x,y=y,n=n,h=h^2)
sigmane[k]=((1/n)*sum(y^2))-((1/n)*(sum(apply(yy1,1,sum))))
y1=xker1(x=x,n=n,h=h)
y2=xkery1(x=x,y=y,n=n,h=h)
y11=y1/(apply(y1,1,sum))
y22=y2/(apply(y1,1,sum))
ysq=(y11)^2
y33=sum(((dnorm(0))/(apply(y1,1,sum))))
y44=sum(apply(ysq,1,sum))
```

```
y55=sum((y-(matrix(c(apply(y22,1,sum)),nrow=n)))^2)
sigmahm[k]=(1/(n-(2*y33)+y44))*y55  }
print(summary(sigmane))
print(summary(sigmahm))
print(var(sigmane))
print(var(sigmahm))
sigmanesort=matrix(c(sort(sigmane)),nrow=nn)
sigmahmsort=matrix(c(sort(sigmahm)),nrow=nn)
bw1=1.06*(sd(sigmane))*((nn)^(-1/5))
bw2=1.06*(sd(sigmahm))*((nn)^(-1/5))
d1=kden(x=sigmanesort,h=bw1,nn=nn)
d2=kden(x=sigmahmsort,h=bw2,nn=nn)
plot(range(sigmanesort,sigmahmsort), range(d1,d2),
 type = "n", xlab = "x", ylab = "Density")
lines(sigmanesort,d1, col = 2,lty=1)
lines(sigmahmsort,d2, col =3,lty=4
legend(locator(1), lty=1:2, col=2:3,
legend=c("New","HM"),xjust=1,yjust=0,x.intersp=0.1,y.intersp=0.8) }
par(mfrow=c(2,2))
conhm(h=0.1,n=100,nn=1000,sig=1)
conhm(h=0.1,n=100,nn=1000,sig=2)
conhm(h=0.1,n=100,nn=1000,sig=5)
conhm(h=0.1,n=100,nn=1000,sig=10)
```

The functions for the figures (3.3) until (3.10) are the same function above except the model and the last four lines. These should be modified as required for each figure. Thus, the details are omitted.

The following R commands is to do the plots in the figure (3.11):

```
msehc=function(x,y,n, sig2) {
h=seq(x,y, by=0.001)
h1=h^2
epsion=matrix(c(rnorm(n, mean=0, sd=sqrt(sig2))),nrow=n)
msenew =((sum(epsion^(4)))/n)*n^(-1)-(sig2)^2*n^(-1)
       +(((2*(sig2)^2*0.2821)
       +(4*sig2*0.2821*16.3333))*n^(-1)*(n-1)^(-1)*h^(-2))
msehm=as.numeric(var(epsion^2)*n^(-1))
       +(2*(sig2)^2*0.40635*n^(-2)*h^(-1))
plot(range(log(h1),log(h)), range(log(msenew),log(msehm)),
type = "n", xlab = "log(h1)", ylab = "log(AMSE)")
lines(log(h1),log(msenew), col = 1,lty=1)
lines(log(h),log(msehm), col =2,lty=2)
```

```
legend(locator(1), lty=1:2, col=1:2,
legend=c("New","HM"),xjust=1,yjust=0,x.intersp=0.1,y.intersp=0.8)}
par(mfrow=c(2,2))
msehc(x=0.00001,y=0.7,n=1000,sig2=1)
title(main="Sigma^2=1")
msehc(x=0.00001,y=0.7,n=1000,sig2=9)
title(main="Sigma^2=9")
msehc(x=0.00001,y=0.7,n=1000,sig2=36)
title(main="Sigma^2=36")
msehc(x=0.00001,y=0.7,n=1000,sig2=100)
title(main="Sigma^2=100")
```

To do the plots in the figure (3.12), the following R commands is used:

```
msehc=function(x,y,n, sig2) {
h=seq(x,y, by=0.001)
h1=h^2
epsion=matrix(c(rnorm(n, mean=0, sd=sqrt(sig2))),nrow=n)
msenew =(msenew=((sum(epsion^(4)))/n)*n^(-1)-(sig2)^2*n^(-1)
       +(((2*(sig2)^2*0.2821)
       +(4*sig2*0.2821*121.53))*n^(-1)*(n-1)^(-1)*h^(-2))
       +(h^(8)*15006.25)
msehm=as.numeric(var(epsion^(2)*n^(-1))
     +(2*(sig2)^2*0.40635*n^(-2)*h^(-1))+(h^(8)*16641)
plot(range(log(h1),log(h)), range(log(msenew),log(msehm)),
type = "n", xlab = "log(h1)", ylab = "log(AMSE)")
lines(log(h1),log(msenew), col = 1,lty=1)
lines(log(h),log(msehm), col =2,lty=2)
legend(locator(1), lty=1:2, col=1:2,
legend=c("New","HM"),xjust=1,yjust=0,x.intersp=0.1,y.intersp=0.8)}
par(mfrow=c(2,2))
msehc(x=0.00001,y=0.7,n=1000,sig2=1)
title(main="Sigma^2=1")
msehc(x=0.00001,y=0.7,n=1000,sig2=9)
title(main="Sigma^2=9")
msehc(x=0.00001,y=0.7,n=1000,sig2=36)
title(main="Sigma^2=36")
msehc(x=0.00001,y=0.7,n=1000,sig2=100)
title(main="Sigma^2=100")
```

Appendix B

The Summary statistics of the Simulation Studies in Chapter 3

The numerical results of the simulation studies for the figures (3.3) until (3.10) are, respectively, as follows,

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.5719	0.5746
1^{st} Qu.	0.8908	0.8927
Median	0.9855	0.9851
Mean	0.9983	0.9979
3^{rd} Qu.	1.0967	1.0987
Maximum	1.4905	1.4787
Variance	0.02332681	0.02287094

Table B.1: Simulation Results for the Figure (3.3) where $\sigma^2 = 1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	2.381	2.365
1^{st} Qu.	3.586	3.588
Median	3.993	3.988
Mean	4.015	4.016
3^{rd} Qu.	4.414	4.409
Maximum	5.889	5.923
Variance	0.3523456	0.3489922

Table B.2: Simulation Results for the Figure (3.3) where $\sigma^2 = 4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	15.60	15.76
1^{st} Qu.	22.56	22.52
Median	24.95	24.87
Mean	25.08	25.08
3^{rd} Qu.	27.45	27.45
Maximum	38.30	37.92
Variance	13.30174	13.08038

Table B.3: Simulation Results for the Figure (3.3) where $\sigma^2 = 25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	63.04	62.64
1^{st} Qu.	91.25	91.14
Median	100.04	99.96
Mean	100.78	100.78
3^{rd} Qu.	109.89	109.74
Maximum	150.58	149.43
Variance	197.6076	196.1564

Table B.4: Simulation Results for the Figure (3.3) where $\sigma^2 = 100$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.3429	0.7856
1^{st} Qu.	0.8762	1.0229
Median	1.0111	1.0913
Mean	1.0093	1.0968
3^{rd} Qu.	1.1276	1.1661
Maximum	1.626	1.4373
Variance	0.03338553	0.01117575

Table B.5: Simulation Results for the Figure (3.4) where $\sigma^2 = 1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	2.348	2.860
1^{st} Qu.	3.645	3.811
Median	3.979	4.089
Mean	4.012	4.097
3^{rd} Qu.	4.349	4.362
Maximum	5.807	5.434
Variance	0.2840502	0.1706614

Table B.6: Simulation Results for the Figure (3.4) where $\sigma^2 = 4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	16.45	18.06
1^{st} Qu.	23.00	23.38
Median	23.00	24.95
Mean	24.98	25.06
3^{rd} Qu.	26.73	26.61
Maximum	38.00	36.29
Variance	7.966907	5.910173

Table B.7: Simulation Results for the Figure (3.4) where $\sigma^2 = 25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	61.59	65.57
1^{st} Qu.	91.80	92.96
Median	99.34	99.58
Mean	99.78	99.95
3^{rd} Qu.	107.28	106.94
Maximum	144.28	140.69
Variance	133.8503	105.1202

Table B.8: Simulation Results for the Figure (3.4) where $\sigma^2 = 100$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.4662	0.528
1^{st} Qu.	0.8647	0.920
Median	0.9837	1.014
Mean	0.9993	1.024
3^{rd} Qu.	1.1258	1.122
Maximum	1.6955	1.567
Variance	0.03804316	0.02284399

Table B.9: Simulation Results for the Figure (3.5) where $\sigma^2 = 1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	2.244	2.439
1^{st} Qu.	3.504	3.629
Median	3.951	3.983
Mean	3.996	4.015
3^{rd} Qu.	4.479	4.254
Maximum	6.287	6.306
Variance	0.5124069	0.3650561

Table B.10: Simulation Results for the Figure (3.5) where $\sigma^2 = 4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	13.02	14.94
1^{st} Qu.	22.17	22.46
Median	24.91	24.99
Mean	24.97	24.98
3^{rd} Qu.	27.68	27.28
Maximum	39.66	39.51
Variance	16.48936	13.18847

Table B.11: Simulation Results for the Figure (3.5) where $\sigma^2 = 25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	56.45	61.22
1^{st} Qu.	88.83	90.48
Median	98.96	99.31
Mean	100.62	100.44
3^{rd} Qu.	110.68	110.03
Maximum	166.92	156.44
Variance	273.0497	214.5188

Table B.12: Simulation Results for the Figure (3.5) where $\sigma^2 = 100$.

Estimator	Proposed Estimator	Hall & Marron Estimator
Minimum	0.5669	0.5709
1^{st} Qu.	0.8947	0.8938
Median	0.9903	0.9893
Mean	0.9949	0.9950
3^{rd} Qu.	1.0877	1.0869
Maximum	1.4871	1.4878
Variance	0.0199973	0.019965

Table B.13: Simulation Results for Figure (3.6) where $\sigma^2 = 1$

Estimator	Proposed Estimator	Hall & Marron Estimator
Minimum	2.375	2.386
1^{st} Qu.	3.627	3.625
Median	3.974	3.978
Mean	4.015	4.014
3^{rd} Qu.	4.408	4.399
Maximum	6.492	6.479
Variance	0.339022	0.3384979

Table B.14: Simulation Results for Figure (3.6) where $\sigma^2 = 4$

Estimator	Proposed Estimator	Hall & Marron Estimator
Minimum	16.50	16.51
1^{st} Qu.	22.66	22.67
Median	24.96	24.92
Mean	25.16	25.16
3^{rd} Qu.	27.55	27.52
Maximum	38.63	38.43
Variance	12.7359	12.70057

Table B.15: Simulation Results for Figure (3.6) where $\sigma^2 = 25$

Estimator	Proposed Estimator	Hall & Marron Estimator
Minimum	53.39	54.26
1^{st} Qu.	90.47	90.43
Median	99.90	99.90
Mean	100.49	100.48
3^{rd} Qu.	109.94	109.80
Maximum	151.66	151.25
Variance	212.7284	211.2499

Table B.16: Simulation Results for Figure (3.6) where $\sigma^2 = 100$

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.5426	0.6257
1^{st} Qu.	0.8924	0.9344
Median	0.9903	1.0329
Mean	1.0019	1.0391
3^{rd} Qu.	1.0962	1.1353
Maximum	1.4796	1.5274
Variance	0.02410496	0.02265262

Table B.17: Simulation Results for the Figure (3.7) where $\sigma^2 = 1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	2.291	2.406
1^{st} Qu.	3.591	3.614
Median	3.961	4.021
Mean	4.027	4.059
3^{rd} Qu.	4.452	4.484
Maximum	6.102	6.085
Variance	0.4338757	0.3848115

Table B.18: Simulation Results for the Figure (3.7) where $\sigma^2 = 4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	12.42	12.77
1^{st} Qu.	22.31	22.59
Median	24.92	25.01
Mean	25.12	25.13
3^{rd} Qu.	27.48	27.35
Maximum	38.85	36.84
Variance	14.43093	12.60777

Table B.19: Simulation Results for the Figure (3.7) where $\sigma^2 = 25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	63.07	64.45
1^{st} Qu.	89.37	89.78
Median	99.17	99.54
Mean	99.5	99.66
3^{rd} Qu.	108.56	108.32
Maximum	147.91	148.51
Variance	213.8416	194.3858

Table B.20: Simulation Results for the Figure (3.7) where $\sigma^2 = 100$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.615	0.7415
1^{st} Qu.	0.913	1.1622
Median	1.032	1.2776
Mean	1.034	1.2845
3^{rd} Qu.	1.144	1.3992
Maximum	1.730	2.1096
Variance	0.02631763	0.03238408

Table B.21: Simulation Results for the Figure (3.8) where $\sigma^2 = 1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	2.377	2.731
1^{st} Qu.	3.562	3.821
Median	3.971	4.209
Mean	3.996	4.240
3^{rd} Qu.	4.381	4.615
Maximum	6.070	6.363
Variance	0.3649796	0.3393106

Table B.22: Simulation Results for the Figure (3.8) where $\sigma^2 = 4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	13.80	15.78
1^{st} Qu.	22.51	22.78
Median	25.08	25.31
Mean	25.22	25.44
3^{rd} Qu.	27.81	27.78
Maximum	39.58	38.02
Variance	15.46193	13.52381

Table B.23: Simulation Results for the Figure (3.8) where $\sigma^2 = 25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	53.87	53.24
1^{st} Qu.	88.37	89.58
Median	98.69	99.23
Mean	99.97	100.02
3^{rd} Qu.	109.84	109.61
Maximum	149.36	145.49
Variance	255.292	216.9286

Table B.24: Simulation Results for the Figure (3.8) where $\sigma^2 = 100$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.9282	0.8544
1^{st} Qu.	1.1557	1.0932
Median	1.2363	1.1669
Mean	1.2389	Mean :1.1695
3^{rd} Qu.	1.3129	1.2378
Maximum	1.6773	1.5907
Variance	0.01334013	0.01182779

Table B.25: Simulation Results for the Figure (3.9) where $\sigma^2 = 1,\ h_{NEW} = 0.06$ and $h_{HM} = 0.12$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.7422	0.7469
1^{st} Qu.	0.9614	0.9915
Median	1.0256	1.0601
Mean	1.0319	1.0634
3^{rd} Qu.	1.0987	1.1285
Maximum	1.3427	1.3698
Variance	0.01137822	0.01084738

Table B.26: Simulation Results for the Figure (3.9) where $\sigma^2 = 1,\ h_{NEW} = 0.02$ and $h_{HM} = 0.08$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.6374	0.7624
1^{st} Qu.	0.9277	0.9509
Median	1.0018	1.0209
Mean	1.0094	1.0233
3^{rd} Qu.	1.0874	1.0950
Maximum	1.3979	1.4564
Variance	0.0131534	0.00983333

Table B.27: Simulation Results for the Figure (3.9) where $\sigma^2 = 1,\ h_{NEW} = 0.008$ and $h_{HM} = 0.05$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	0.6642	0.6955
1^{st} Qu.	0.9189	0.9255
Median	0.9997	1.0029
Mean	1.0041	1.0043
3^{rd} Qu.	1.0867	1.08
Maximum	1.4899	1.3813
Variance	0.01533834	0.01220766

Table B.28: Simulation Results for the Figure (3.9) where $\sigma^2 = 1,\ h_{NEW} = 0.006$ and $h_{HM} = 0.01$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	26.84	26.64
1^{st} Qu.	34.76	34.97
Median	37.35	37.64
Mean	37.49	37.75
3^{rd} Qu.	39.96	40.27
Maximum	50.63	50.75
Variance	14.7104	14.90147

Table B.29: Simulation Results for the Figure (3.10) where $\sigma^2 = 36,\ h_{NEW} = 0.2$ and $h_{HM} = 0.4$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	25.89	25.90
1^{st} Qu.	33.86	34.10
Median	36.28	36.52
Mean	36.41	36.63
3^{rd} Qu.	38.95	39.26
Maximum	46.88	46.96
Variance	13.34199	13.19872

Table B.30: Simulation Results for the Figure (3.10) where $\sigma^2 = 36,\ h_{NEW} = 0.1$ and $h_{HM} = 0.25$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	25.17	24.81
1^{st} Qu.	33.77	33.67
Median	35.98	35.95
Mean	36.08	36.03
3^{rd} Qu.	38.45	38.39
Maximum	47.72	47.99
Variance	12.76928	12.68781

Table B.31: Simulation Results for the Figure (3.10) where $\sigma^2 = 36,\ h_{NEW} = 0.05$ and $h_{HM} = 0.1$.

Estimator	New Estimator	Hall & Marron Estimator
Minimum	24.87	25.04
1^{st} Qu.	33.39	33.69
Median	36.04	36.09
Mean	36.06	36.06
3^{rd} Qu.	38.48	38.37
Maximum	47.68	47.46
Variance	14.75015	13.21818

Table B.32: Simulation Results for the Figure (3.10) where $\sigma^2 = 36,\ h_{NEW} = 0.01$ and $h_{HM} = 0.07$.

Appendix C

The Matlab Functions for the Figures in Chapter 6

The main Matlab codes, which are used to produce the figures in chapter 6, are demonstrated in this appendix. The following function is used to define the mean and variance functions as required for each plot in the figures (6.1)-(6.28).

```
function [mx, rvx] = meanvar(x)
mx=2+(4*x)-(4*x.^2);
rvx=0.5*(2+(4*x)-(4*x.^2)+(3*x.^3));
end
```

In the above function, mx represents the mean function, whereas rvx denotes the variance function. The main Matlab function to produce the figures (6.1)-(6.28) is as follows where n denotes the sample size; nn represents the number of replications; h and h_2 refer to the bandwidths which is used to estimate the mean and variance functions for the new estimator, respectively and h_3 represents the bandwidth which is used to estimate the variance function for the Brown-Levine estimators:

```
function evfnbl(  n, nn , h,h2,h3)
ss2=zeros(nn,n);vx2=zeros(nn,n);
vx3=zeros(nn,n);vx4=zeros(nn,n);
for k=1:1:nn
x1=rand(1,n); x=sort(x1);
[mx,rvx]=meanvar(x)
y=mx +((sqrt(rvx)).*randn(1,n));
ty=y';
w=zeros(n,n);w1=zeros(n,n);
for i=1: 1:n
for j=1:1:n
a=x(j)-x(i);
if a==0
w(i,j)=0;
```

```
else
w(i,j)= (1/sqrt(2*pi))*exp(-0.5*(a/h).^2);
end
end
end
for i=1:n
w1(:,i)=w(:,i)./sum(w,2);
end
w2=zeros(n,n);
for i=1:n
for j=1:n
w2(i,j)=w1(i,j)*y(i)*ty(j);
end
end
w3=zeros(n,n);
for i=1:n
for j=1:n
w3(i,j)= (1/sqrt(2*pi))*exp(-0.5*(((x(i)-x(j))/h2).^2));
end
end
w33=zeros(n,n);
for i=1:n
w33(:,i)=w3(:,i)./sum(w3,2);
end
w4= (y.^2-(sum(w2,2))');
ss2(k,:)=w33*w4';
u=zeros(1,n);
for i=2:1:n-1
u(i)=sum((0.809*y(i-1))+(-0.5*y(i))+(-0.309*y(i+1)));
end
u1=u.^2;
u2=zeros(1,n);
for i=3:1:n-2
u2(i)=sum((0.2708*y(i-2))+(-0.0142*y(i-1))
+(0.6909*y(i))+(-0.4858*y(i+1))+(-0.4617*y(i+2)));
end
u12=u2.^2;
u3=zeros(1,n);
for i=4:1:n-3
u3(i)=sum((0.24*y(i-3))+(0.03*y(i-2))+(-0.0342*y(i-1))
+(0.7738*y(i))+(-0.3587*y(i+1))+(-0.3038*y(i+2))+(-0.3472*y(i+3)));
end
u13=u3.^2;
e=[1 0 ];
```

```
x11=ones(1,n); vx1=zeros(1,n);vx12=zeros(1,n); vx13=zeros(1,n);
for i=1:n
x2=x(i)-x; xxt=[ x11; x2];
w55=diag((1/sqrt(2*pi))*exp(-0.5*(((x(i)-x)/h3).^2)));
vx1(i)=e*(inv(xxt*w55*xxt'))*xxt*w55*u1';
vx12(i)=e*(inv(xxt*w55*xxt'))*xxt*w55*u12';
vx13(i)=e*(inv(xxt*w55*xxt'))*xxt*w55*u13';
end
vx2(k,:)=vx1;  vx3(k,:)=vx12; vx4(k,:)=vx13;
end
%To plot the estimated and true variance functions
plot(x,mean(ss2,1),'b --',x,mean(vx2,1),'k  -.', x,rvx,'r -',
x,mean(vx3,1),'k --', x ,mean(vx4,1),'k :')
end
```

For example, to draw the figure (6.1), the following Matlab codes are used where $meanvar$ function is changed every time as required:

```
subplot(221),
evfnbl(  100, 1000, 0.025,0.1,0.06);
subplot(222),
evfnbl(  100, 1000, 0.025,0.1,0.06);
subplot(223),
evfnbl(  100, 1000, 0.025,0.1,0.06);
subplot(224),
evfnbl(  100, 1000, 0.025,0.1,0.06);
legend('New Est.','B-L r=2', 'True', 'B-L r=4', 'B-L r=6')
```

To plot the variance and the MSE of the estimated variance function, we can save the data of the variance and the MSE every time, then we use the following Matlab codes:

```
plot(x,var(ss2,1),'b --',x,var(vx2,1),'k  -.',
x,var(vx3,1),'k -', x,var(vx4,1),'k :')
xlabel('x'),ylabel('Var'), title('V_{1}(x)')
plot(x,(mean(ss2,1)-rvx).^2 + var(ss2,1),'b --',
x,(mean(vx2,1)-rvx).^2 + var(vx2,1),'k  -.',
x,(mean(vx3,1)-rvx).^2+var(vx3,1),'k -',
x ,(mean(vx4,1)-rvx).^2 +var(vx4,1),'k :')
xlabel('x'),ylabel('MSE'), title('V_{1}(x)')
```

To produce the scatter plot of age-Blood Pressure data in the figure (6.29), we use the following codes:

```
xage=[20.1832495, 20.26676297, 21.31068133, 22.22932948,
```

```
23.27324784,24.27540946, 24.27540946, 25.36108455, 25.40284129,
26.27973271, 27.44892127, 28.32581269, 29.16094737, 30.24662246,
31.20702735, 31.24878409, 32.29270244, 33.12783713, 33.25310733,
34.29702569, 35.29918731, 37.09472689, 37.13648362, 38.05513178,
38.18040198, 39.26607707, 40.18472522,  40.26823869, 42.06377827,
42.14729174, 43.19121009, 43.23296683, 44.44391212, 45.07026314,
45.98891129, 46.07242476, 46.07242476, 46.86580271, 48.24377494,
49.12066636, 49.37120677, 50.12282798, 50.12282798, 52.04363776,
52.0853945, 52.2106647,  53.08755612, 54.17323121, 55.1336361,
56.17755446, 56.92917567, 56.97093241, 58.22363444, 59.14228259];
ybp=[70.26791809, 65.39419795, 66.35153584, 63.21843003,
70.35494881, 75.22866894, 72.26962457, 71.31228669, 68.2662116,
79.31911263, 73.31399317, 67.221843, 79.23208191, 73.31399317,
66.26450512, 80.1894198, 76.44709898, 76.01194539, 69.22354949,
73.31399317, 79.05802048, 68.2662116, 77.83959044, 91.32935154,
87.5, 75.48976109, 90.28498294, 70.35494881, 72.53071672,
85.49829352, 75.31569966, 80.1894198, 71.31228669, 92.1996587,
83.3225256, 89.15358362, 80.53754266, 96.20307167, 70.00682594,
101.2508532, 80.53754266, 91.32935154, 71.48634812, 86.4556314,
85.15017065, 100.1194539, 79.58020478, 71.57337884, 76.27303754,
92.28668942, 99.16211604, 109.2576792, 80.53754266,90.37201365];
plot(xage,ybp,'o')
```

To draw the figure (6.30), we use a similar codes to the $evfnbl$ function where $h_1 = 0.034$, $h_2 = h_3 = 0.15$, $n = 54$, $nn = 1$, $y = ybp$ and $x = xage$.

List of References

[1] Alharbi, Y. F. (2011). Error variance estimation in nonparametric regression models. University of Birmingham, MPhil dissertation.

[2] Altman, N. S. (1992). An introduction to kernel and nearest-neighbour nonparametric regression. Journal of the American Statistical Association 46, 175-185.

[3] Ansley, C. F., Khon, R. and Tharm, D. (1990). The estimation of the smoothing parametric and residual variance in nonparametric spline smoothing. In Proc. Bus. Econ. Sect., American Statistical Association.

[4] Ash, R. (1972). Real analysis and probability. London, Academic Press.

[5] Bock, M., Bowman, A. W . and Ismail, B. (2007). Estimation and inference for error variance in bivariate nonparametric regression. Journal of Statistics and Computing 17, 39-47.

[6] Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica 47, 1287-1294.

[7] Brown, L. D. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. The Annals of Statistics 35, 2219-2232.

[8] Buckley, M. J., Eagleson, G. K. and Silverman, B. W , (1988). The estimation of residual variance in nonparametric regression. Biometrika 75, 189-199.

[9] Buckley, M. J. and Eagleson, G. K. (1989). A graphical method for estimating the residual variance in nonparametric regression. Biometrika 76, 203-210.

[10] Cai, T., Levine, M. and Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. Journal of Multivariate Analysis 100, 126-136.

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

[12] Carroll, R. J. and Ruppert, D. (1988). Transforming and weighting in regression. London: Chapman & Hall.

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

[14] Chan, L. (1991). Optimal design for estimation of variance in nonparametric regression using first order difference. Biometrika 78, 926-929.

[15] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions (Estimating the correct degree of smoothing by the method of generalized cross-validation). Journal of Numerische Mathematik 31, 377-403.

[16] Dette, H., Munk, A. and Wanger, T. (1998). Estimating the variance in nonparametric regression- What is a reasonable choice?. Journal of the Royal Statistical Society . Series B (Statistical Methodology) 60, 751-764.

[17] Dette, H. and Pilz, K. (2009). On the estimation of a monotone conditional variance in nonparametric regression. Annals of the Institute of Statistical Mathematics 61, 111-141, Springer.

[18] Draper N.R. and Smith, H. (1981). Applied regression analysis. New York: John Willy & Sons.

[19] Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. Journal of the American Statistical Association 88, 1287-1301.

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

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

[22] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. The Annals of Statistics 21, 1900-1925.

[23] Fan, J. and Gijbels, I. (1996). Local polynomial modelling and its applications. London: Chapman & Hall.

[24] Fan, J. and Yao, Q. (1998). Efficient estimation for conditional variance function in stochastic regression. Biometrika 85, 645-660.

[25] Galambos, J. (1988). Advanced probability theory. New York and Basel, Marcel Dekker, INC.

[26] Gasser, T. and Müller, H. (1984). Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171-185.

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

[28] Gasser, T., Kneip, A. and Kohler, W. (1991). A flexible and fast method for automatic smoothing. Journal of the American Statistical Association 86, 643-652.

[29] Grimmett, G. and Welsh, D. (1986). Probability an introduction. Oxford, Clarendon Press.

[30] Hall, P. (1984). Central limit theorem for integrated square error of multivariate density estimators. Journal of Multivariate Analysis 14, 1-16.

[31] Hall, P. (1989). On convergence rates in nonparametric problems. Journal of International Statistical Review 57(1), 45-58.

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

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

[34] Hall, P. and Marron, J. (1987) Estimation of integrated squared density derivatives. Journal of Statistics & Probability Letters 6, 109-115.

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

[36] Härdle, W. (1990). Applied nonparametric regression. Cambridge: Cambridge University press.

[37] Härdle, W. (1991). Smoothing techniques with implementation in S. New York: Springer-Verlag.

[38] Johnson, N., Kotz, S. and Kemp, A. (1992). Univariate Discrete Distribution. New York, Jhon Wiley & Sons, INC.

[39] Jones, M., Marron, J. and Park, B. (1991). A simple root n bandwidth selector. The Annals of Statistics 19(4), 1919-1932.

[40] Jong, P. D. (1987). A central limit theorem for generalized quadratic forms. Probability Theory and Related Fields 75, 261-277.

[41] Kyung-Joon, C. and Schucany, W. R. (1998). Nonparametric kernel regression estimation near endpoints. Journal of Statistical Planning and Inference 66, 289-304.

[42] Levene, H. (1960). Robust test for the equality of variance, "in Contributions to Probability and Statistics", ed. I. Olkin. Palo alto, Calif: Stanford University Press, 278-292.

[43] Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimator in the nonparametric regression: possible approach. Journal of Computational Statistics & Data Analysis 50, 3405-3431.

[44] Lu, Y. (2012). Difference-based variance estimator for nonparametric regression in complex surveys. Journal of Statistical Computation and Simulation 1, 1-9, Taylor & Francis Group.

[45] Martinez, W. and Martinez, A. (2008). Computational statistics with MATLAB. London, Chapman & Hall/ CRC.

[46] Müller, H. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. The Annals of Statistics 15, 610-625.

[47] Müller, H. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. Journal of the American Statistical Association 82, 231-238.

[48] Müller, U. U., Schick, A. and Wefelmeyer, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. Statistics 37, 179-188.

[49] Munk, A., and Ruymgaart, F. (2002). Minimax rates for estimating variance and its derivatives in non-parametric regression. Australian & New Zealand Journal of Statistics 44, 479-488.

[50] Neter, J., Kutner, M., Nachtsheim, C. and Wasserman, W. (1996). Applied linear statistical models. Chicago: Irwin.

[51] Oh, J. (2003). Variance estimation of error in the regression model at a point. Journal of Applied Mathematics & Computing 13, 501-508.

[52] Park, C., Kim, I. and Lee, Y. (2012). Error variance estimation via least squares for small sample nonparametric regression. Journal of Statistical Planning and Inference 142, 2369-2385.

[53] Patil, P., Wells, M. and Marron J. (1994). Some heuristics of kernel based estimation for ratio function. Journal of Nonparametric Statistics 4(2), 203-209.

[54] Petrov, V. (1995). Limit theorems of probability theory: Sequences of independent random variables. Oxford, Clarendon Press.

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

[56] Rudemo (1982). Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics 9, 65-78.

[57] Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997). Local polynomial variance-function estimation. Journal of the Technometrics 39, 262-273.

[58] Seifert, B., Gasser, T. and Wolf, A. (1993). Nonparametric estimation of residual variance revisited. Biometrika 80, 373-383.

[59] Sheather, S. (2004). Density estimation. Journal of Stitiscal Science 19(4), 588-597.

[60] Silverman, B. W . (1986). Density estimation for statistics and data analysis. London: Chapman & Hall.

[61] Stone, C. J. (1977). Consistent nonparametric regression. The Annals of Statistics 5, 595-620.

[62] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. The Annals of Statistics 6, 1348-1360.

[63] Stone, J. (2004). Independent component analysis: a tutorial introduction. USA, Massachusetts Institute of Technology.

[64] Stuart, A. and Ord, J. K. (1994). Kendall's Advanced: Theory of Statistics (Distribution Theory). London, Hodder Arnold.

[65] Thompson, A. M., Kay, J. W. and Titterington, D. M. (1991). Noise estimation in signal restoration using regularization. Biometrika 78, 475-488.

[66] Tong, T., Liu, A. and Wang, Y. (2008). Relative errors of difference-based variance estimators in nonparametric regression. Communications in Statistics - Theory and Methods 37, 2890-2902.

[67] Tong, T. and Wang, Y. (2005). Estimating residual variance in nonparametric regression using least squares. Biometrika 92, 821-830.

[68] Venables, W. and Ripley (2002). Modern applied statistics with S. New York, Springer.

[69] Wand, M. and Jones, M. (1995). Kernel smoothing. London, Chapman & Hall.

[70] Wang, L., Brown, L. D., Cai, T. T. and Levine M. (2008). Effect of mean on variance function estimation in nonparametric regression. The Annals of Statistics 39, 646-664.

[71] Wang, L. (2008). Variance function estimation in nonparametric regression. USA, ProQuest LLC, UMI:3328698.

[72] Yu, K. and Jones, M. (2004). Likelihood based-local linear estimation of the conditional variance function. Journal of the American Statistical Association 99 (465), 139-144.