#2008

# SOME RESULTS IN NONPARAMETRIC REGRESSION

by

Chih-Kang Chu

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics

Chapel Hill

1989

Approved by

Advisor

Reader

Reader

CHIH-KANG CHU. Some Results in Nonparametric Regression (Under the direction of J. S. MARRON.)


## ABSTRACT

For nonparametric regression, the two most popular methods for constructing kernel estimators, involving choosing weights by kernel evaluation or by subinterval integration, are compared for their asymptotic mean square errors. Their performance is quite different when the design points are serious departures from equal spacing, or when the design points are randomly chosen.

For choosing the bandwidth in nonparametric regression, the ordinary cross-validation provides poor bandwidth estimates when the observations are correlated. In the case of the short range dependent observations, the modified cross-validation is proposed. Based on the reduction of the asymptotic bias of the bandwidth estimates, the modified cross-validation would provide asymptotically optimal bandwidths. However, the modified cross-validated bandwidth suffers from a large amount of sample variability.

The partitioned cross-validation is applied to reduce the sample variability. However, the partitioned cross-validation provides asymptotically biased bandwidths. The two criteria are compared for their asymptotic mean square errors of the bandwidth estimates. In the simulation study, it is shown that the performance of the two criteria depends on the amount of sample variability.

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Steve Marron, for his constant support, enthusiastic guidance, and helpful criticism in all parts of the research presented in this dissertation. I also wish to thank the other members of my committee, Dr. Edward Carlstein, Dr. Gordon Simons, Dr. Walter Smith, and Dr. Young Truong for their valuable comments and suggestions.

Special thanks go to my fellow students for their warm encouragement and help during the years I spent in Chapel Hill. Thanks to Miguel Nakamura for his excellent knowledge about GAUSS and T3. Thanks to Mike Frey for his supports and helpful advice.

I am grateful to the Department of Statistics for providing a good study environment and financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER I

# INTRODUCTION OF NONPARAMETRIC REGRESSION

## 1.1 Introduction

In regression, one seeks to find a relationship between the response Y and the design point x. Frequently there is a functional relationship between the mean response of Y and the design point x. If the functional form is known except for some unknown parameters, then the regression is called a parametric regression. A parametric regression model is often expressed by

$$Y = m(x,\beta) + \epsilon.$$

Here m is some known function, $\beta$ is a vector of unknown parameters, and $\epsilon$ is a zero mean regression error introduced to account for measurement and all other possible errors. If the functional form is unknown but smooth, then the regression is called a nonparametric regression. A nonparametric regression model is expressed as

$$Y = m(x) + \epsilon.$$

In this case, m is a smooth unknown function, $\epsilon$ is a zero mean regression error introduced to account for measurement and all other possible errors.

When appropriate, parametric regression models have some definite advantages. For example, the corresponding inferential methods usually have nice efficiency properties. Also, the parameters generally have some physical meaning which makes them interpretable and of interest in their own right. If the assumed parametric model is grossly in error, the above advantages of the parametric approach will not be realized. Thus, there are few benefits from using a poorly specified parametric form of m. Parametric models are most appropriate when theory, past experience, or other sources are available that provide detailed knowledge about the process under study.

The result of a nonparametric regression is a curve fitted to a set of data. Since this is produced without assuming a parametric form of m, there is some loss in the interpretability and efficiency of estimators obtained in this fashion. In contrast to physics or engineering, it is not often appropriate to give a specific functional relationship between the mean response of Y and the design point x in biological systems. One reason is that modeling is more difficult for living organisms, a second is that it may be difficult to find one well-fitting parametric family of functions just for descriptive purposes.

Whenever there is no appropriate parametric model available, the data may provide information for the parametric regression. Thus, nonparametric regression may represent the final stage of data analysis or merely an exploratory or confirmatory step in the modeling process. See, for example, the monographs by Eubank (1988), Mueller (1988), and Haerdle (1988) for a large variety of interesting real data examples where applications of this method have yielded analyses essentially unobtainable by other techniques.

In Section 1.2 the two most popular methods for recovering the unknown regression function m(x) are discussed. In Section 1.3 we discuss cross-validation, a method for choosing the bandwidth in nonparametric regression. Partitioned cross-validation, a method for improving the rate of convergence of the cross-validated bandwidth, is briefly discussed in Section 1.4. Finally, Section 1.5 contains a brief summary of the results obtained.

## 1.2 Kernel Estimators

Nonparametric regression, a smoothing method, has been the focus of substantial theoretical work in recent years and also has been applied increasingly to real data. Typical examples of nonparametric regression from the medical field are: the fixed design case where the $x_i$ are times at which the observations $Y_i$ are made on a patient, and the random design case in which studies where $X_i$ and $Y_i$ are observed on the same patient; e.g. height and weight or blood pressure and cholesterol level.

The fixed design nonparametric regression model can be written as

$$Y_i = m(x_i) + \epsilon_i, \tag{1.1}$$

for $i = 1, 2, \ldots, n$. Here m is a smooth unknown regression function defined on the interval [0,1] (without loss of generality), $x_i$ are nonrandom design points with $0 \leq x_1 \leq x_2 \leq \ldots \leq x_n \leq 1$, $Y_i$ are noisy observations of the regression function m at the design points $x_i$, and $\epsilon_i$ are uncorrelated random variables with mean zero and finite variance $\sigma^2$. If $x_i$ are equally spaced, i.e. $x_i = i/n$, then (1.1) is called the equally spaced fixed design nonparametric regression model.

If the regression function m is believed to be smooth, then the observations at points $x_i$ near x should contain information about the value of m at x. Kernel estimators are local weighted averages of the response variables, i.e. using the weighted average of those $Y_i$ which have $x_i$ close to x to estimate m(x). The width of the neighborhood in which averaging is performed is called the bandwidth or smoothing parameter. The kernel function is a given function used to calculate the weights assigned to the observations. In this dissertation, the kernel function K is taken as a probability density function. The two most popular methods for constructing kernel estimators involve choosing weights either by kernel evaluation or subinterval integration. For other related kernel estimators, see, for example, Mack and Mueller (1989) and the monographs by Eubank (1988), Mueller (1988), and Haerdle (1988).

To construct kernel estimators, first Nadaraya (1964) and Watson (1964) proposed choosing weights by evaluating a kernel function at the design points and then dividing by the sum of the weights, so that they add up to one. Then Gasser and Mueller (1979) proposed choosing weights as the integrals of the kernel function on small subintervals which contain the design points. In both situations, given a kernel function K, a bandwidth h, and $K_h(\cdot) = h^{-1}K(\cdot/h)$, for $0 < x < 1$, the Nadaraya-Watson estimator is defined by

$$(1.2) \qquad \hat{m}_{NW}(x) = [n^{-1} \sum_{i=1}^{n} K_h(x-x_i)Y_i]/[n^{-1} \sum_{i=1}^{n} K_h(x-x_i)],$$

(if the denominator is zero, take $\hat{m}_{NW}(x) = 0$), and the Gasser-Mueller estimator is defined by

$$(1.3) \qquad \hat{m}_{GM}(x) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)\, dt\, Y_i,$$

where $s_0 = 0$, $s_n = 1$, and $x_i \le s_i \le x_{i+1}$, for $i = 1, 2, \ldots, n-1$. See Gasser and Mueller (1979) for a different choice of the kernel function.

If the design points $x_i$ are equally spaced, then the two estimators are almost equivalent to each other, as shown by the integral mean value theorem. The two estimators are equivalent in their asymptotic mean square error, almost sure convergence, $L_2$ convergence, rate of convergence, and asymptotic normality.

The random design nonparametric regression model is written as

$$Y_i = m(X_i) + \epsilon_i, \tag{1.4}$$

for i = 1, 2, ..., n. In this case, m, $Y_i$, and $\epsilon_i$ are assumed to have the same properties as they did in (1.1); only the $X_i$ are different. The $X_i$ are independent and identically distributed (IID) random variables with a density function f on the interval [0,1], and $X_i$ and $\epsilon_i$ are uncorrelated random variables.

For the random design case, $\hat{m}_{NW}(x)$ is expressed as

$$\hat{m}_{NW}(x) = [n^{-1} \sum_{i=1}^{n} K_h(x-X_i)Y_i]/[n^{-1} \sum_{i=1}^{n} K_h(x-X_i)] \tag{1.5}$$

(if the denominator is zero, take $\hat{m}_{NW}(x) = 0$), and $\hat{m}_{GM}(x)$ is expressed as

$$\hat{m}_{GM}(x) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)\, dt\, Y_{(i)}. \tag{1.6}$$

Here $Y_{(i)}$ are the observations corresponding to the i-th order statistic $X_{(i)}$ of the random design points. The integral points $s_i$ are taken as

$$s_i = \beta X_{(i)} + (1-\beta)X_{(i+1)}, \quad \beta \in [0,1],$$

for i = 1, 2, ..., n-1, and $s_0$ and $s_n$ are any reasonable choice, e.g. $s_0 = 0$ and $s_n = 1$. In the random design case, see Mack and Mueller (1989) for other kernel estimators.

A drawback of the Nadaraya-Watson estimator in the random design case is that the estimator becomes tricky to analyze from a technical view point. Indeed, for some high order kernel functions, Haerdle and Marron (1983) have shown that the moments of the Nadaraya-Watson estimator can fail to exist. The high order kernel function requires the value of the kernel function to be negative in some place (see Gasser and Mueller (1979)).

For the random design case, Devroye and Wagner (1980) showed the weak convergence of $\hat{m}_{NW}(x)$, and Schuster (1972) and Stute (1984) established an asymptotic normality of $\hat{m}_{NW}(x)$ under different conditions. Based on the existence of the continuous second derivative m'' of m and the continuous first derivative f' of f, Rosenblatt (1969) and Collomb (1981) gave an asymptotic mean square error of $\hat{m}_{NW}(x)$ based on

$$(1.7) \qquad \mathrm{Var}(\hat{m}_{NW}(x)) = n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}),$$

$$(1.8) \quad E(\hat{m}_{NW}(x)) = m(x) + h^2(m''/2 + m'f'/f)(x)\int u^2K + o(h^2) + O(n^{-1}h^{-1}).$$

Here and throughout this chapter, the notation $\int$ denotes $\int du$. Working under the same conditions as Rosenblatt (1969) and Collomb (1981), Jennen-Steinmetz and Gasser (1987) and Mack and Mueller (1989) gave an asymptotic mean square error of $\hat{m}_{GM}(x)$ based on

$$(1.9) \qquad \mathrm{Var}(\hat{m}_{GM}(x)) = 2n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}),$$

$$(1.10) \quad E(\hat{m}_{GM}(x)) = m(x) + (1/2)h^2m''(x)\int u^2K + o(h^2) + O(n^{-1}),$$

where the integral points of $\hat{m}_{GM}(x)$ are taken as the order statistics of the design points, i.e. $\beta$ is taken as 1 in (1.6).

Note that $\hat{m}_{GM}(x)$ has twice the variance of $\hat{m}_{NW}(x)$. Jennen–

Steinmetz and Gasser (1987) and Mack and Mueller (1989) mentioned that $\hat{m}_{NW}(x)$ is unattractive because its bias term has an extra factor $m'f'/f$. For the random design case, there is no overall conclusion about which kernel estimator is better in the sense of asymptotic mean square error.

## 1.3 Cross-validation

The magnitude of bandwidth controls the smoothness of the resulting estimate of m(x). In practice, we need to choose a value of the bandwidth and plug it into the kernel estimator. Figures 1.1 and 1.2 show a simulated regression setting with 100 equally spaced design points. The observations are generated by adding pseudo-random normal variables $\epsilon_i$, with mean 0 and variance $\sigma^2 = 0.0015^2$, to the regression curve $m(x) = x^3(1-x)^3$ for $0 \leq x \leq 1$. The kernel function is taken as the bi-weight kernel, $K(x) = (15/8)(1-4x^2)^2 I_{[-1/2,1/2]}(x)$. The goal of regression is to use the observations and the Nadaraya-Watson estimator (1.2) to recover the curve m(x). If the bandwidth is too small, i.e. the average is made only with a few observations, then the resulting estimate of m(x) is undersmooth (see Figure 1.1). On the other hand, if the bandwidth is too large, i.e. the average is made with many observations, then the resulting estimate of m(x) is oversmooth (see Figure 1.2). This is the essence of the smoothing problem. In the extreme case as the bandwidth approaches zero, then kernel estimators approach a $\delta$-function, i.e. the estimate of m(x) is $Y_i$ if x is $x_i$ for some i, otherwise the estimate is zero. This is because no design points are included in the sufficiently small neighbourhood of x if x is not a design point. On the other hand, if the bandwidth approaches

Figure 1.1: Plot of the regression function (dashed curve), the observations (solid squares), the Nadaraya-Watson estimate (solid curve), and the kernel function (dashed curve shown at bottom) for the sample size n = 200, the bandwidth h = 0.05, and the IID regression errors $N(0, 0.0015^2)$.

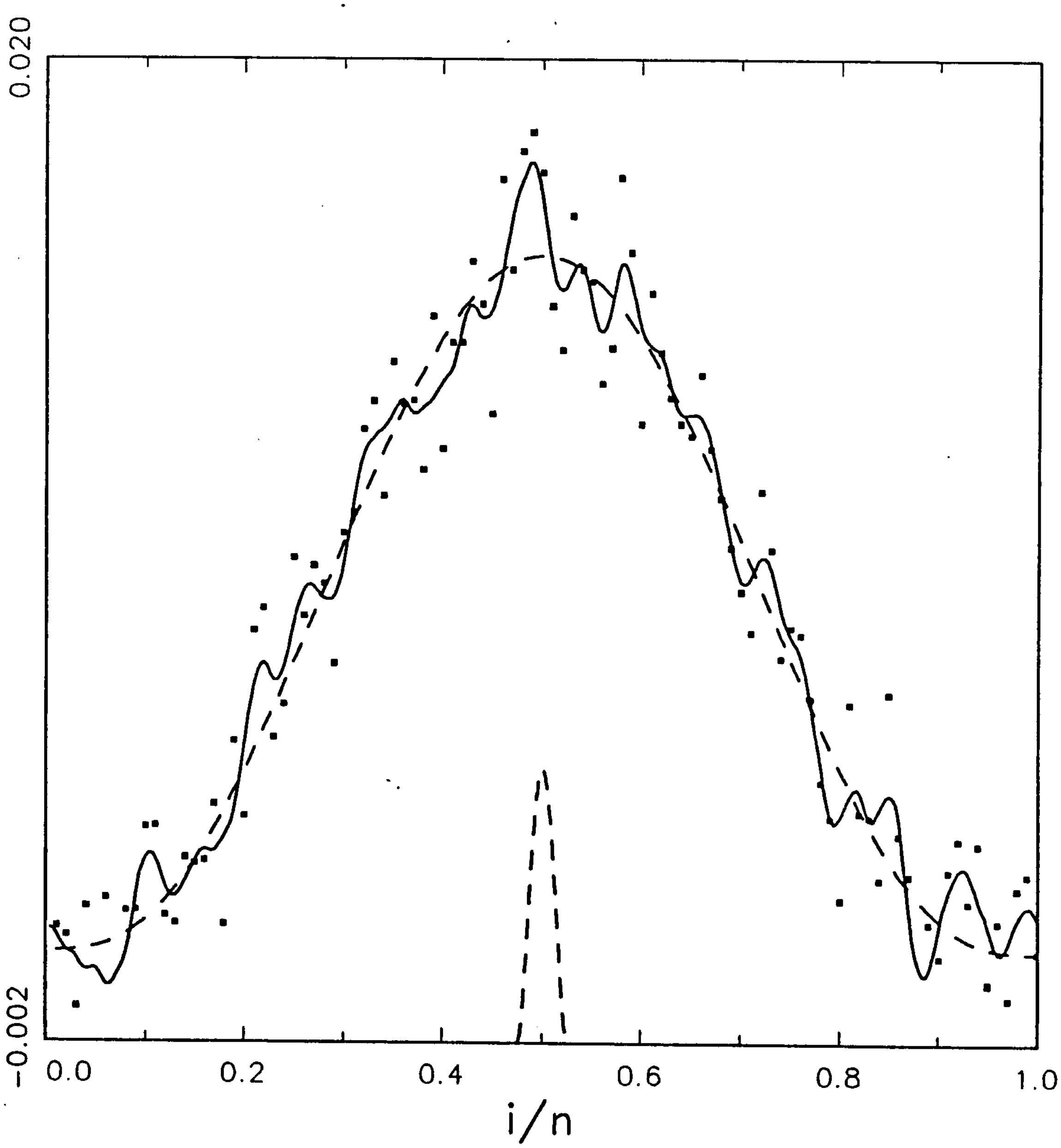
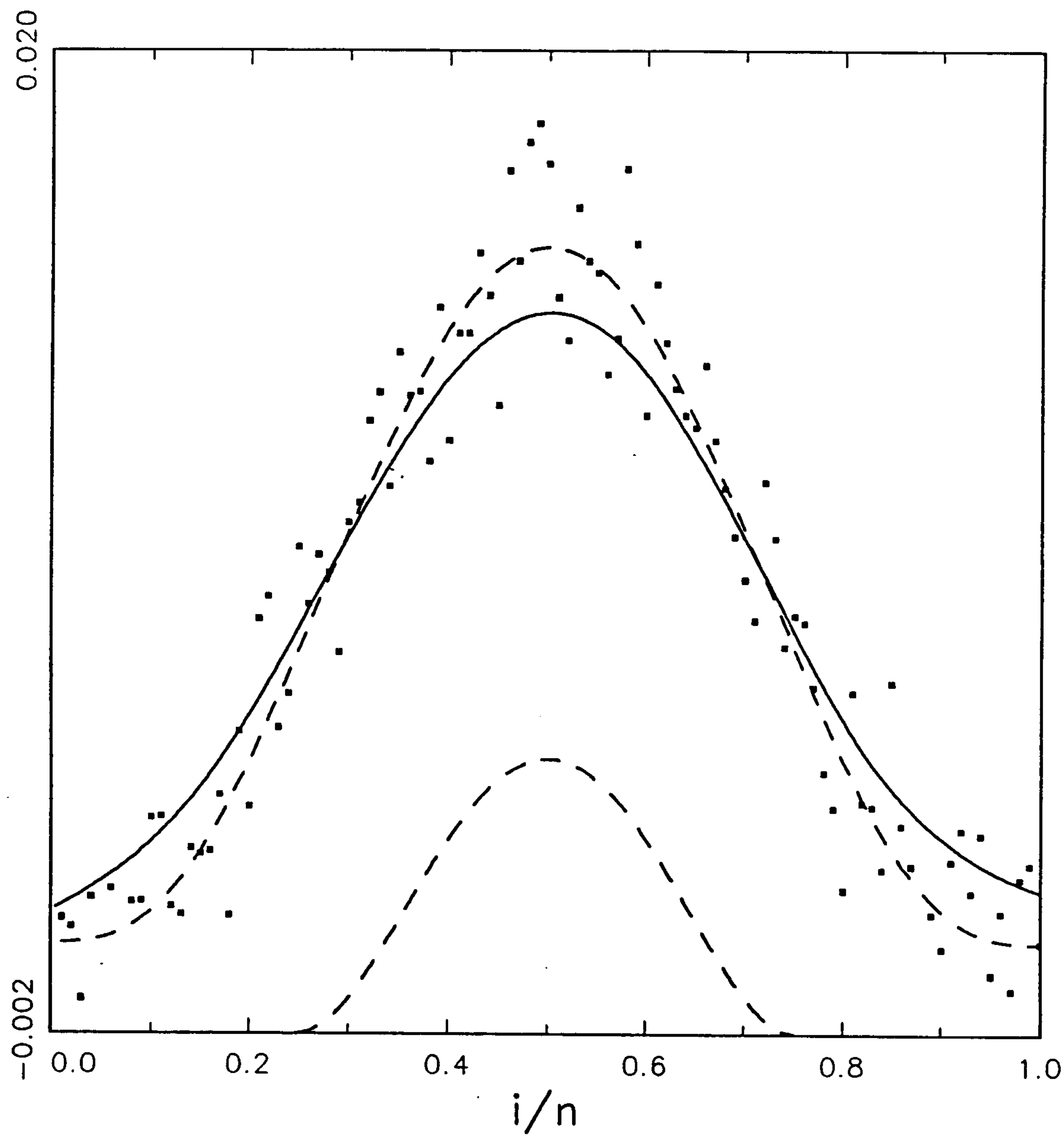


Figure 1.2: Plot of the regression function (dashed curve), the observations (solid squares), the Nadaraya-Watson estimate (solid curve), and the kernel function (dashed curve shown at bottom) for the sample size n = 200, the bandwidth h = 0.5, and the IID regression errors $N(0,0.0015^2)$.



infinity, then kernel estimators become an overall average of observations, i.e. the estimate of m(x) is $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ for all x. This is due to the fact that all design points are included in the sufficiently large neighbourhood of x for all x.

The two optimal bandwidths, $h_A$ and $h_M$, are taken as the minimizers of the average square error (ASE) and the mean average square error (MASE) respectively. The ASE function is defined by

$$(1.11) \qquad d_A(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}(x_j) - m(x_j)]^2 W(x_j),$$

where $\hat{m}(x_j)$ are kernel estimators of $m(x_j)$ for every j. The weight function W is introduced to allow elimination (or at least significant reduction) of boundary effects by taking W to be supported on a subinterval of the unit interval (see Gasser and Mueller (1979)). The MASE function is defined by

$$(1.12) \qquad d_M(h) = E(d_A(h)).$$

We call $h_A$ and $h_M$ the optimal bandwidths among all possible choices of bandwidths h because the estimated curve $\hat{m}(x)$, as derived by the bandwidth $h_A$, is the closest to the regression function m(x) for the data set at hand. Similarly, $h_M$ makes $\hat{m}(x)$ closest to m(x) according to the average over all possible data sets. In practice, the values of the two optimal bandwidths $h_A$ and $h_M$ are not available because these quantities depend on the unknown function m(x). Since the values of $h_A$ and $h_M$ can not be calculated, bandwidth selection methods have been designed with the purpose of providing bandwidths which approximate $h_A$ and $h_M$ in some mode. Most bandwidth selection methods are based on minimization of some functions of bandwidth h.

For the equally spaced fixed design nonparametric regression model (1.1), Clark (1975) introduced the cross-validation idea as the bandwidth-selection rule for kernel estimators. The essential idea is based upon the fact that the regression function $m(x)$ is the best mean square predictor of a new observation taken at x. This idea suggests choosing the bandwidth so that $\hat{m}(x)$ is a good predictor or, in other words, taking the bandwidth as the minimizer of the estimated prediction errors

$$EPE(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}(x_j) - Y_j]^2 W(x_j).$$

This criterion for choosing the minimizer has a problem, namely that it has a trivial minimum at $h = 0$. This is because, as h approaches zero, $\hat{m}(x)$ becomes a $\delta$-function. This problem can be viewed as a result of using the same data to both construct and assess the estimator. A reasonable approach to this problem is provided by replacing $\hat{m}(x_j)$ as $\hat{m}_j(x_j)$, where $\hat{m}_j(x_j)$ is the "leave-1-out" version of $\hat{m}(x_j)$, i.e. the observation $(x_j, Y_j)$ is left out in constructing $\hat{m}(x_j)$. To illustrate, for the Nadaraya-Watson estimator, $\hat{m}_j(x_j)$ is defined by

$$(1.13) \quad \hat{m}_j(x_j) = [(n-1)^{-1} \sum_{i:i\neq j} K_h(x_j-x_i)Y_i]/[(n-1)^{-1} \sum_{i:i\neq j} K_h(x_j-x_i)].$$

Hence, the cross-validation idea is to choose the bandwidth by minimizing the cross-validation score

$$(1.14) \qquad CV(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}_j(x_j) - Y_j]^2 W(x_j).$$

In the extreme case as the bandwidth h approaches zero, then $CV(h)$ approaches $n^{-1} \sum_{j=1}^{n} Y_j^2 W(x_j)$. On the other hand, if h approaches zero, then CV(h) approaches $n^{-1} \sum_{j=1}^{n} [\bar{Y} - Y_j]^2 W(x_j)$. The minimizer, $\hat{h}_{CV}$, of the CV(h) is called the cross-validated bandwidth. See Haerdle and Marron (1985) for a motivation of this criterion.

In the case of independent observations, there is another popular bandwidth-selection rule. Mallows' criterion is to choose the bandwidth by minimizing the risk function

$$(1.15) \qquad \hat{R}(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}(x_j) - Y_j]^2 - \sigma^2 + 2n^{-1}h^{-1}\sigma^2 K(0).$$

Here $\hat{R}(h)$ is an asymptotically unbiased estimate of $d_M(h)$ when $nh \longrightarrow \infty$. See Mallows (1973) for details. In practice, $\sigma^2$ is usually replaced by

$$\hat{\sigma}^2 = (2(n-1))^{-1} \sum_{j=1}^{n-1} [Y_{j+1} - Y_j]^2.$$

For bandwidth selection, Rice (1984) worked with the reciprocal of h and the equally spaced fixed, circular design Mallows' criterion. The equally spaced fixed, circular design means that the observations $Y_{j+kn}$ at the design points (j+kn)/n are the same as the observations $Y_j$ at the design points j/n for all integers k and j = 1, 2, ..., n. Finally, see Rice (1984) and Haerdle, Hall, and Marron (1988) for other bandwidth-selection rules.

After one bandwidth selection rule has been chosen, we can arrive at a bandwidth estimate $\hat{h}$. We need to measure how good the bandwidth estimate is. Suppose h is in some set $H_n \subset R_n^+$ which is a range of bandwidths of interest. Let the distance $d(\hat{m}_h, m)$ denote a given measure of accuracy for the kernel estimator $\hat{m}_h$, where the subscript h means that the kernel estimator $\hat{m}_h$ involves the bandwidth h. Following Shibata (1981), the bandwidth-selection rule $\hat{h}$ is said to be asymptotically optimal with respect to the measure d if

$$\lim_{n\to\infty} [d(\hat{m}_{\hat{h}}, m) / \inf_{h\in H_n} d(\hat{m}_h, m)] = 1$$

with probability one. Based on the equally spaced fixed design nonparametric regression model (1.1), Rice (1984) showed that, for the circular design case, Mallows' criterion provides weakly consistent estimates of $1/h_M$ with a relative rate $n^{-1/10}$. In the same case, Haerdle, Hall, and Marron (1988) showed that $\hat{h}_{CV}$, $h_A$, and $h_M$ approach zero at the same rate $n^{-1/5}$, and $\hat{h}_{CV}$ converges weakly to $h_A$ and $h_M$ with a rate $n^{-3/10}$. From Haerdle, Hall, and Marron (1988), it follows that $\hat{h}_{CV}$ converges to $h_A$ and $h_M$ with a relative rate $n^{-1/10}$. For the random design case, Haerdle and Marron (1985) showed that $\hat{h}_{CV}$ is an asymptotically optimal bandwidth estimator with respect to the measures ASE and MASE.

For the equally spaced fixed design nonparametric regression model (1.1), Haerdle, Hall, and Marron (1988) compared the cross-validation and some other bandwidth selection methods. They showed that the bandwidths derived by these bandwidth selection methods are asymptotically equivalent to one another. Hence, all properties described for the cross-validated bandwidth apply to these as well. See Marron (1988), a survey paper, for the properties of other bandwidth selectors.

## 1.4 Partitioned Cross-validation

In kernel density estimation, the cross-validated bandwidth also suffers an excruciatingly slow relative rate of convergence $n^{-1/10}$ to the optimal bandwidths (see Hall and Marron (1987)). For kernel density estimation, Marron (1988) proposed the partitioned cross-validation (PCV) criterion which improves the relative rate of convergence of the cross-validated bandwidth from $n^{-1/10}$ to $n^{-1/4}$. The idea of PCV could be applied to nonparametric regression and all other smoothing methods.

The basic idea of the PCV is to use extra smoothness on the regression function m(x) and split the observations into g subgroups by taking every g-th observation, calculating the cross-validation score for each subgroup separately and minimizing the average of these score functions. The resulting bandwidth is rescaled by the number of subgroups g to get the partitioned cross-validated bandwidth $\hat{h}_{PCV(g)}$. Then the asymptotic mean square error (AMSE) of $(\hat{h}_{PCV(g)}/h_M - 1)$ depends on parameters g and n, and other factors. For the PCV, we minimize the AMSE of $(\hat{h}_{PCV(g)}/h_M - 1)$ with respect to g and get a theoretically optimal value of g for the smallest value of the AMSE. One drawback to this approach is that the optimal value of g depends on the unknown regression function m(x). To address this drawback, we use the "plug-in" approach, in Section 6.3, to show how the value of g is arrived at from the simulated data sets.

## 1.5 Summary of Results

In Chapter 2, we compare the two kernel estimators $\hat{m}_{NW}(x)$ and $\hat{m}_{GM}(x)$ on the unequally spaced fixed and random design cases. In both cases we find different asymptotic representations for the asymptotic biases of $\hat{m}_{NW}(x)$. We also find that, in the random design case, the minimum asymptotic variance of $\hat{m}_{GM}(x)$ is arrived at by taking the integral points of $\hat{m}_{GM}(x)$ as the middle points of each consecutive two order statistics of the design points. In both cases, the asymptotic variances of $\hat{m}_{GM}(x)$ are greater than those of $\hat{m}_{NW}(x)$.

In Chapter 3, we explore the effect of dependent observations on the modified cross-validated bandwidth. When the observations are subjected to a short range dependence, the asymptotically optimal bandwidth is attainable by deleting enough points in the construction of the modified cross-validation score.

In Chapter 4, we derive an asymptotic normality of the modified cross-validated bandwidth when the observations are subjected to a short range dependence. The relative rate of convergence of the modified cross-validated bandwidth to $h_M$ is $n^{-1/10}$ which is the same as in the case of independent observations.

In Chapter 5, we use the partitioned cross-validation to improve the relative rate of convergence of the cross-validated bandwidth to $h_M$ when the observations are subjected to a short range dependence. Here the results are quite surprising in view of those from Marron (1988).

In Chapter 6, we do a simulation study to compare the performances between the modified cross-validated and the partitioned cross-validated bandwidths when the regression errors are an AR(1) process in time series analysis.

# CHAPTER II

# COMPARISON OF TWO KERNEL ESTIMATORS

## 2.1 Introduction

Kernel estimators are local weighted averages of the response variables. The two most popular methods for constructing kernel estimators, in the case of the equally spaced fixed design, involve choosing weights either by kernel evaluation or subinterval integration. If the design points are equally spaced, then the two estimators are very nearly the same as demonstrated by the integral mean value theorem. The purpose of this chapter is to compare the two estimators for the unequally spaced fixed design and the random design. Then, the estimator that proves more efficient will be chosen for further work.

Section 2.2 uses an example to illustrate the efficiency of the two kernel estimators in the case of the unequally spaced fixed design. A comparison of the two estimators for the random design is given in Section 2.3. An improved version of the Gasser-Mueller estimator is given in Section 2.4. Section 2.5 contains a discussion of our results. Finally, the proofs are given in Section 2.6.

## 2.2 A Comparison for the Unequally Spaced Fixed Design

Given the equally spaced fixed design nonparametric regression model (1.1), the Nadaraya-Watson estimator (1.2), and the Gasser-

Mueller estimator (1.3), we will construct an example of the unequally spaced fixed design to illustrate the point that the kernel estimator (1.3) is inefficient in the case of the unequally spaced fixed design.

For the Gasser-Mueller estimator $\hat{m}_{GM}(x)$, when looking at the weights $\int_{s_{i-1}}^{s_i} K_h(x-t)\,dt$ assigned to the observations $Y_i$, it becomes immediately clear that $Y_i$ will be downweighted if $s_{i-1}$ and $s_i$ are too close. We study this effect by starting with an equally spaced fixed design, and considering consecutive triples of design points. For each triple, we move the first and the third points toward the center. The amount of shift to the center can be parameterized by a value $\alpha \in [0,1]$. This results in the design points expressed as

$$x_i = \begin{cases} (3\ell+2-\alpha) \,/\, n & \text{if } i = 3\ell+1 \\ (3\ell+2) \,/\, n & \text{if } i = 3\ell+2, \\ (3\ell+2+\alpha) \,/\, n & \text{if } i = 3\ell+3 \end{cases} \tag{2.2.1}$$

for $\ell = 0, 1, \ldots, (n/3)-1$. Here, for convenience of notation, we assume that n is a multiple of 3. Note that $\alpha = 1$ is the usual equally spaced fixed design, while $\alpha = 0$ is a design which is also essentially equally spaced but which has three replications at each point. As $\alpha$ approaches zero, the weight given to the center observation is essentially zero. This is because the integral points, $s_{3\ell+1} \in [x_{3\ell+1}, x_{3\ell+2}]$ and $s_{3\ell+2} \in [x_{3\ell+2}, x_{3\ell+3}]$, of the center observation $Y_{3\ell+2}$ approach each other. In this case, $\hat{m}_{GM}(x)$ is making use of only 2/3 of the available observations, hence its inefficiency. However, in this case, the Nadaraya-Watson estimator $\hat{m}_{NW}(x)$ is nearly independent of $\alpha$. It is because the weights assigned to each triple of observations $Y_{3\ell+1}$, $Y_{3\ell+2}$, and $Y_{3\ell+3}$ are roughly equal to $n^{-1}K_h(x-x_{3\ell+2})$ as determined by the continuity of K and (2.6.1) given later. The extent of the inefficiency of the Gasser-Mueller estimator can be quantified by studying the asymptotic variance of the two kernel estimators.

We are not claiming this unequally spaced fixed design is important, however it is considered here because it provides a clear and simple illustration of the point being made. The effects described will provide an intuitive basis for understanding the causes of the similar effects in the more complicated random design case.

We now do an asymptotic analysis for the design points (2.2.1). Assume that the kernel function K is a probability density function with support contained in the interval [-1,1], and that the kernel function K and the regression function m are Hoelder continuous of order 1. The definition of the Hoelder continuity of order 1 of a function g is that there is a constant M such that

$$|g(s) - g(t)| \leq M \cdot |s-t|$$

for any s and t in the domain of function g. Then as $n \longrightarrow \infty$, $h \longrightarrow 0$, with $nh \longrightarrow \infty$, for $0 < x < 1$, it is shown in Section 2.6 that

$$\text{(2.2.2)} \qquad \text{Var}(\hat{m}_{NW}(x)) = n^{-1}h^{-1}\sigma^2\int K^2 + O(n^{-2}h^{-2}),$$

$$\text{(2.2.3)} \qquad \text{Var}(\hat{m}_{GM}(x)) = C(\alpha)n^{-1}h^{-1}\sigma^2\int K^2 + O(n^{-2}h^{-2}),$$

$$\text{(2.2.4)} \qquad \text{Bias}(\hat{m}_{NW}(x)) = \int K_h(x-t)(m(t)-m(x))\,dt + O(n^{-1}),$$

$$\text{(2.2.5)} \qquad \text{Bias}(\hat{m}_{GM}(x)) = \int K_h(x-t)(m(t)-m(x))\,dt + O(n^{-1}),$$

where $C(\alpha) = 1 + (1/2)(\alpha-1)^2$, $0 \leq \alpha \leq 1$. Here and throughout this chapter, the notation $\int$ denotes $\int du$.

Looking at (2.2.3), we observe that when $\alpha = 1$ (the minimizer of $C(\alpha)$), the two estimators have the same performance, which, as noted above, is expected from the integral mean value theorem. However, for any $\alpha \in [0,1)$, the variance of $\hat{m}_{GM}(x)$ is larger than that of $\hat{m}_{NW}(x)$. In the extreme case of $\alpha = 0$, note that $\hat{m}_{GM}(x)$ has 3/2 times the variance of $\hat{m}_{NW}(x)$. This is because $\hat{m}_{GM}(x)$ is only using 2/3 of the available data. Of course, if one really had three replications at each design point (as we have when $\alpha = 0$), then the obvious thing is to pool, i.e. work with the average of the observations at these points. One of the features of $\hat{m}_{NW}(x)$ is that it makes this adjustment automatically, as $\alpha \longrightarrow 0$. On the other hand, $\hat{m}_{GM}(x)$ has a disturbing tendency to delete observations. The biases of the two estimators are exactly the same in the asymptotic expression and rate of convergence. So this example suggests that the variance of $\hat{m}_{GM}(x)$ will be increased by the too close integral points.

The above results may be generalized in a straightforward manner to the case of forming clusters of k points, instead of the 3 points case. When this is done, (2.2.2), (2.2.4), and (2.2.5) remain the same, except (2.2.3) becomes

$$\text{(2.2.6)} \qquad \text{Var}(\hat{m}_{GM}(x)) = C_k(\alpha)n^{-1}h^{-1}\sigma^2\int K^2 + O(n^{-2}h^{-2}),$$

where $C_k(\alpha) = 1 + (1/2)(\alpha-1)^2(k-2)$, $0 \leq \alpha \leq 1$ and $k \geq 2$. Note that the downweighting effect of $\hat{m}_{GM}(x)$ can be magnified when $k \geq 3$, subject of course to the fact that these asymptotics describe only the situation where $nh >> k$.

## 2.3 A Comparison for the Random Design

In the random design case, it is clear that, by chance, some design points will be closer to their neighbours than others are. It is expected that the efficiency of the Gasser-Mueller estimator is affected by this situation. Given the random design nonparametric regression model (1.4), the Nadaraya-Watson estimator (1.5), and the Gasser-Mueller estimator (1.6), we shall do an asymptotic analysis of this point.

Using the integral mean value theorem and the Hoelder continuity of K, the ratios of the weights assigned to the observations $Y_{(i)}$ by the Nadaraya-Watson estimator $\hat{m}_{NW}(x)$ and the Gasser-Mueller estimator $\hat{m}_{GM}(x)$ are roughly equal to $(1/n) : D_i$, where

$$D_i = s_i - s_{i-1} = -\beta X_{(i-1)} - (1-2\beta)X_{(i)} + (1-\beta)X_{(i+1)}.$$

Figures 2.1 and 2.2 use the same 100 Uniform[0,1] pseudo random variables to show graphically how strong the variability of the integral points is in the random design case. Figure 2.1 shows the case $\beta = 1$ and Figure 2.2 shows the case $\beta = 1/2$. Note that the relative weights differ across the observations to a surprising degree with a substantial number of the points significantly downweighted, which means that the Gasser-Mueller estimator is making very inefficient use of the data. In view of (2.2.3), this effect is expected to give a dramatically increased variance of $\hat{m}_{GM}(x)$ with respect to $\hat{m}_{NW}(x)$ whose relative weights are roughly illustrated by the horizontal line. It is also expected that the variance of $\hat{m}_{GM}(x)$ at $\beta = 1/2$ is less than that at $\beta = 1$.

To make the above ideas more precise, we now start an asymptotic analysis. Assume that the kernel function K, the regression function

Figure 2.1: Plot of the relative weights assigned by the Nadaraya-Watson estimator (horizontal line) and the Gasser-Mueller estimator (vertical bars) to the observations for the case of $\beta = 1$.

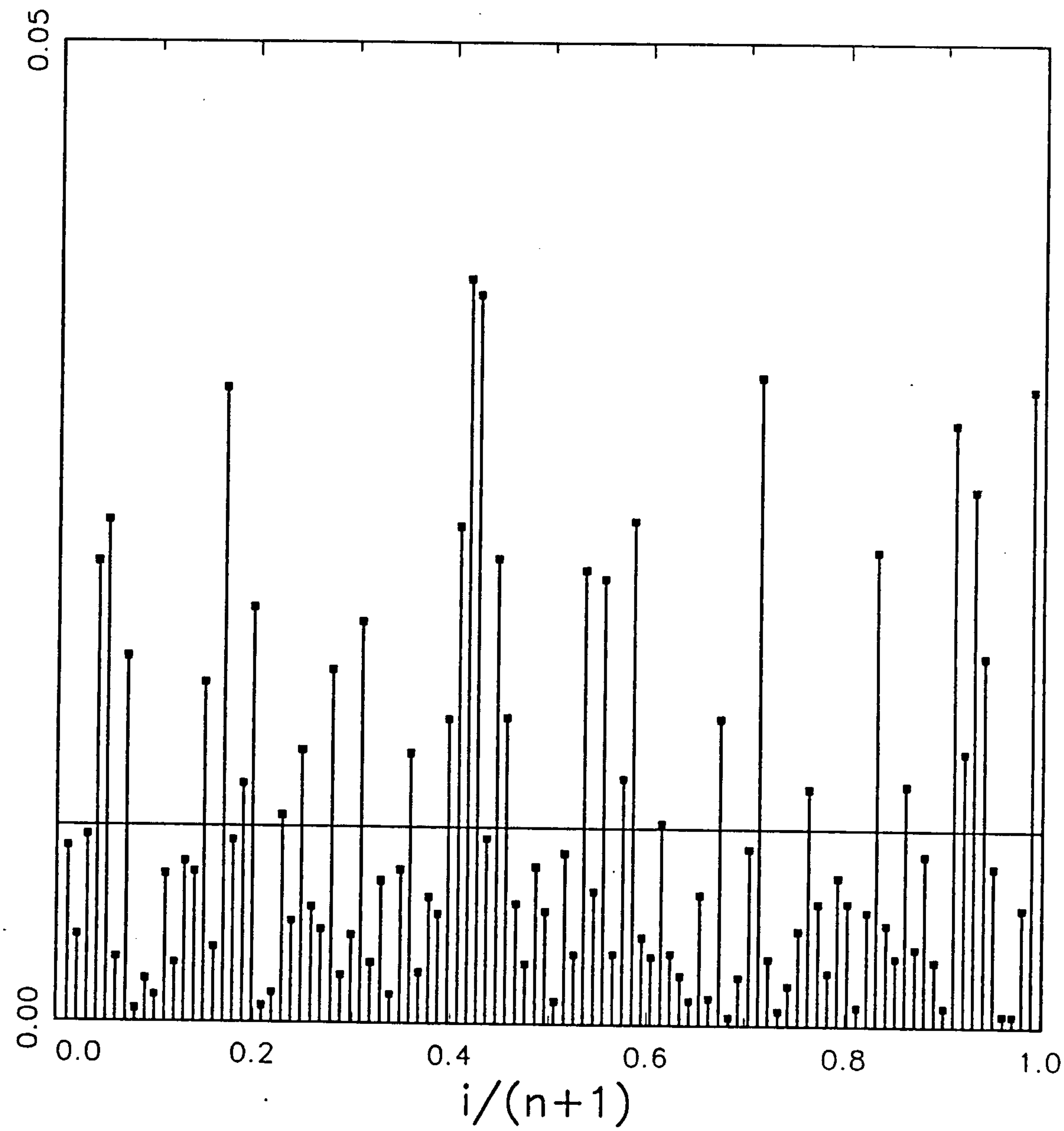
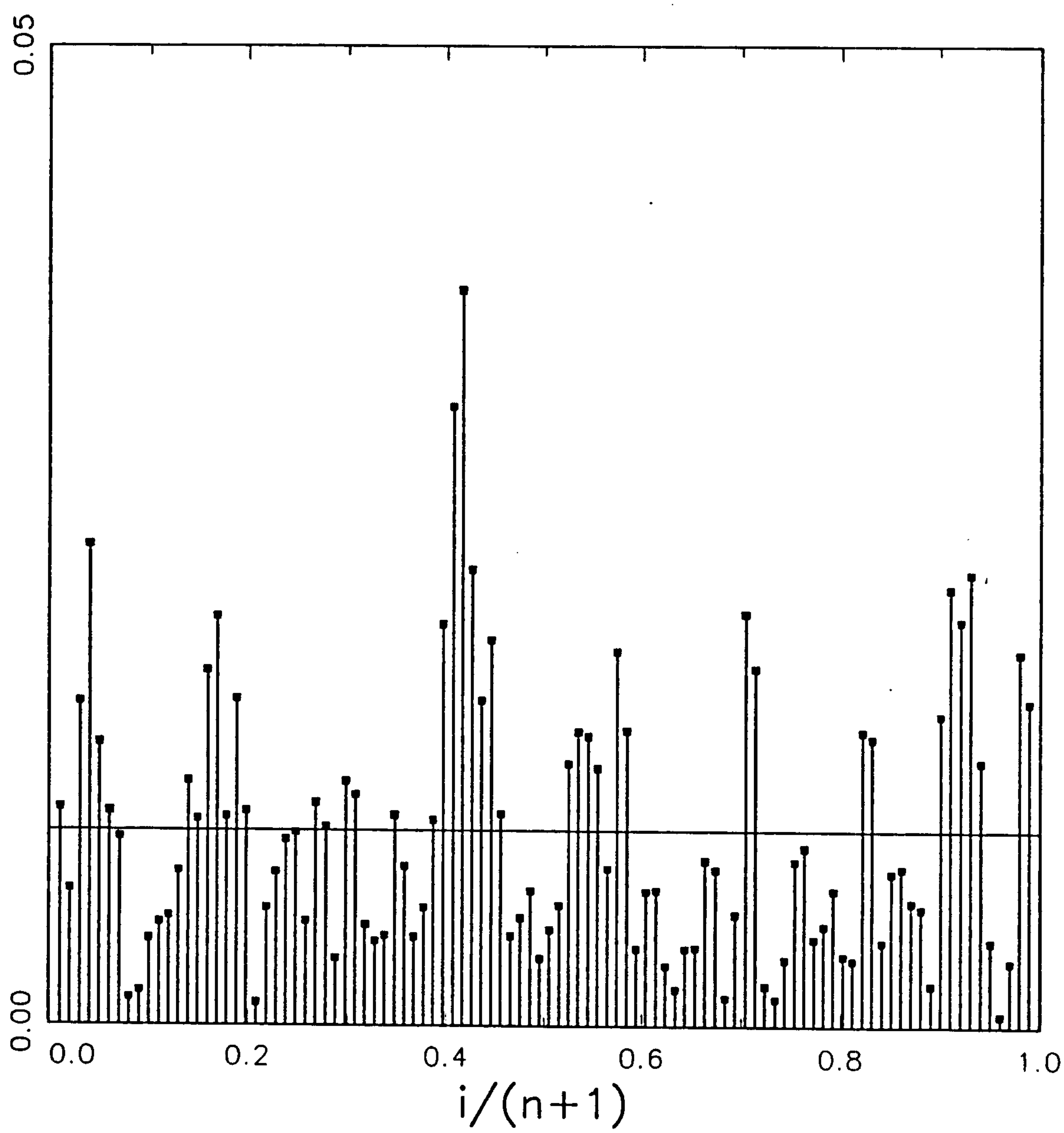


Figure 2.2: Plot of the relative weights assigned by the Nadaraya-Watson estimator (horizontal line) and the Gasser-Mueller estimator (vertical bars) to the observations for the case of $\beta = 1/2$.



m, and the density function f of the design points are Hoelder continuous of order 1, and that K is a probability density function with support contained in the interval [-1,1]. Without loss of generality, K has a positive lower bound $\ell_K$ on [-1/2,1/2] and f has a positive lower bound $\ell_f$ on [0,1]. This positive lower bound of f guarantees that the design has no holes on [0,1]. Then as $n \longrightarrow \infty$, with $n^{-1+\epsilon} < h < n^{-\epsilon}$, for some $\epsilon \in (0,1/2)$, and $0 < x < 1$, it is shown in Section 2.6 that

$$(2.3.1) \qquad \mathrm{Var}(\hat{m}_{NW}(x)) = n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}).$$

$$(2.3.2) \quad \mathrm{Var}(\hat{m}_{GM}(x)) = 2(1-\beta+\beta^2)n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}).$$

$$(2.3.3) \qquad \mathrm{Bias}(\hat{m}_{NW}(x)) = [\int K_h(x-t)(m(t)-m(x))f(t)dt] \,/\, [\int K_h(x-t)f(t)dt] + O(n^{-1}h).$$

$$(2.3.4) \qquad \mathrm{Bias}(\hat{m}_{GM}(x)) = \int K_h(x-t)(m(t)-m(x))dt + O(n^{-1}).$$

The above results are based on the assumption of the Hoelder continuity of m and f which is a little weaker than the continuity of the second derivative m" of m and the first derivative f' of f needed by Rosenblatt (1969), Collomb (1981), Jennen-Steinmetz and Gasser (1987), and Mack and Mueller (1989).

Given $\beta = 1/2$ in (2.3.2), the Gasser-Mueller estimator has the minimum asymptotic variance

$$1.5n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}).$$

Thus, the choice of $\beta = 1$, made in Jennen-Steinmetz and Gasser (1987) and Mack and Mueller (1989) seems inadvisable in practice. In the uniform random design case, both asymptotic biases have the same asymptotic expression $\int K_h(x-t)(m(t)-m(x))dt$. So, in the uniform random design case, we can conclude that $\hat{m}_{NW}(x)$ is better than $\hat{m}_{GM}(x)$ because $\hat{m}_{NW}(x)$ has a smaller value of asymptotic mean square error.

## 2.4 An Improved Gasser-Mueller Estimator

The results of Section 2.3 clearly show that the increased variance of $\hat{m}_{GM}(x)$ in the random design case is caused by the instability of the integral points $s_i$. A means of reducing this instability is to average together more order statistics in the definition of the integral points $s_i$. One means of doing this, given a positive integer k, is to define

$$(2.4.1) \qquad s_i = \sum_{j=1}^{2k} \beta_j X_{(i-k+j)},$$

for i = k, k+1, ..., n-k, where the weights $\beta_j$ are nonnegative and satisfy $\sum_{j=1}^{2k} \beta_j = 1$, and where $s_0, s_1, \ldots, s_{k-1}$ and $s_{n-k+1}, \ldots, s_n$ are defined in any reasonable manner (again the exact definition does not affect our main ideas).

Assumptions for an asymptotic analysis of the new version of $\hat{m}_{GM}(x)$ are the same as the assumptions given in Section 2.3. Then as $n \longrightarrow \infty$, $h \longrightarrow 0$, with $nh \longrightarrow \infty$, it is shown in Section 2.6 that

$$(2.4.2) \quad \mathrm{Var}(\hat{m}_{GM}(x)) = (1 + \sum_{j=1}^{2k} \beta_j^2)n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}),$$

$$(2.4.3) \qquad \mathrm{Bias}(\hat{m}_{GM}(x)) = \int K_h(x-t)(m(t)-m(x))\, dt + O(n^{-1}).$$

From (2.4.2), it is easy to see that the best choice of $\beta_j$ is $\beta_j = 1/(2k)$, for j = 1, 2, ..., 2k. In this case, the Gasser-Mueller estimator has the minimum asymptotic variance

$$\mathrm{Var}(\hat{m}_{GM}(x)) = (\frac{2k+1}{2k})n^{-1}h^{-1}(f(x))^{-1}\sigma^2\int K^2 + o(n^{-1}h^{-1}).$$

Thus the amount by which the variance of $\hat{m}_{NW}(x)$ improves over $\hat{m}_{GM}(x)$ can be reduced, if k is large. Of course, a practical limitation is that these asymptotics are only meaningful when nh >> k.

## 2.5 Discussion

Through the instability of the integral points $s_i$ of the Gasser-Mueller estimator $\hat{m}_{GM}(x)$, the observations with close neighbours will be downweighted. This effect causes the inefficiency of $\hat{m}_{GM}(x)$. When the observations are dependent, we expect that the minimum variance of $\hat{m}_{GM}(x)$ is arrived at by taking the middle points of each consecutive two order statistics of the design points as integral points. We also expect that, even in this case, the variance of $\hat{m}_{GM}(x)$ is still larger than that of $\hat{m}_{NW}(x)$.

In further analysis of (2.3.3) and (2.3.4), if m and f have three continuous derivatives at x and the kernel function K is a symmetric density function, then by Taylor's theorem, the asymptotic biases of the two kernel estimators are

$$(2.5.1)\quad \mathrm{Bias}(\hat{m}_{NW}(x)) = h^2(2f(x))^{-1}(m''f + 2m'f')(x)\int u^2K + O(h^3+n^{-1}h),$$

$$(2.5.2)\quad \mathrm{Bias}(\hat{m}_{GM}(x)) = (1/2)h^2m''(x)\int u^2K + O(h^3+n^{-1}).$$

Unfortunately, these expressions are not comparable. In some cases, one asymptotic bias will be bigger in magnitude (depending on the signs of m"(x)f(x) and m'(x)f'(x)), while in other cases, the other asymptotic bias will be bigger. The bias of $\hat{m}_{GM}(x)$ is certainly more simple and, at first glance, one might think it is more natural.

However a case can be made for the bias of $\hat{m}_{NW}(x)$ being the more natural one. In particular, when the data are being used efficiently, the design density f(x) is an important issue which should affect how well one can estimate m(x). The fact that f(x) nearly disappears in the bias of $\hat{m}_{GM}(x)$ can be considered an unattractive feature of that estimator. A completely different view of this is that, in order to make the design density essentially disappear from the bias, one must pay some price, which, in this case is increased variance.

It would be nice to find some way to resolve this completely, say by finding some other methods to comprise the integral points of $\hat{m}_{GM}(x)$ in which one bias is bigger than the other, but we do not know how to do this. We do, however, speculate that if this can be done in a reasonable manner, the net result in terms of the asymptotic mean square error will be in favor of $\hat{m}_{NW}(x)$.

## 2.6 Proofs

Let c denote a generic constant.

Proofs of (2.2.2) and (2.2.4):

We first derive an asymptotic expression of the denominator of $\hat{m}_{NW}(x)$. Since the design points are clustered in groups of three, we rearrange the denominator of $\hat{m}_{NW}(x)$ as

$$n^{-1} \sum_{i=1}^{n} K_h(x-x_i) = A_1 + B_1 + C_1,$$

where

$$A_1 = \sum_{\ell=0}^{(n/3)-1} K_h(x-x_{3\ell+1}),$$
$$B_1 = \sum_{\ell=0}^{(n/3)-1} K_h(x-x_{3\ell+2}),$$
$$C_1 = \sum_{\ell=0}^{(n/3)-1} K_h(x-x_{3\ell+3}),$$

and where n is assumed to be a multiple of 3 for simplicity. Each of these summations is over on equally spaced grid, with width 3/n, so $A_1$, $B_1$, and $C_1$ are all the Riemann summation for

$$(1/3)\int K_h(x-t)\, dt + O(n^{-1}h^{-1}).$$

Hence, by the Hoelder continuity of K, we have

$$(2.6.1) \qquad n^{-1} \sum_{i=1}^{n} K_h(x-x_i) = 1 + O(n^{-1}h^{-1}).$$

Using (2.6.1), the variance and bias of $\hat{m}_{NW}(x)$ can be expressed as

$$\begin{aligned} Var(\hat{m}_{NW}(x)) &= Var[n^{-1} \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i \;/\; n^{-1} \sum_{i=1}^{n} K_h(x-x_i)] \\ &= \sigma^2 n^{-2} \sum_{i=1}^{n} K_h(x-x_i)^2\, (1 + O(n^{-1}h^{-1})), \end{aligned}$$

$$Bias(\hat{m}_{NW}(x)) = n^{-1} \sum_{i=1}^{n} K_h(x-x_i)(m(x_i)-m(x))(1 + O(n^{-1}h^{-1})).$$

The Riemann summation method used above show that asymptotically:

(2.6.2) $$n^{-2}\sum_{i=1}^{n} K_h(x-x_i)^2 = n^{-1}h^{-1}\int K^2 + O(n^{-2}h^{-2}).$$

(2.6.3) $$n^{-1}\sum_{i=1}^{n} K_h(x-x_i)(m(x_i)-m(x)) = \int K_h(x-t)(m(t)-m(x))\,dt + O(n^{-1}).$$

This completes the proofs of (2.2.2) and (2.2.4).

Proof of (2.2.5): It is immediate from equation (6) of Gasser and Mueller (1984).

Proof of (2.2.3):

Using the integral mean value theorem, the variance of $\hat{m}_{GM}(x)$ can be expressed as

$$\begin{aligned}\mathrm{Var}(\hat{m}_{GM}(x)) &= \mathrm{Var}[\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K_h(x-t)\,dt\,(m(x_i)+\epsilon_i)] \\ &= \sigma^2\sum_{i=1}^{n}[\int_{s_{i-1}}^{s_i} K_h(x-t)\,dt]^2 \\ &= \sigma^2\sum_{i=1}^{n}(s_i - s_{i-1})^2 K_h(x-t_i)^2,\end{aligned}$$

for some $t_i \in [s_{i-1}, s_i]$, $i = 1, 2, \ldots, n$. So, by the Hoelder continuity of K, we have

$$\begin{aligned}\mathrm{Var}(\hat{m}_{GM}(x)) &= \sigma^2\sum_{i=1}^{n}(s_i - s_{i-1})^2 K_h(x-x_i)^2 + O(n^{-2}h^{-2}) \\ &= \sigma^2\{\alpha^2 A_2 + ((3-\alpha)/2)^2 B_2 + ((3-\alpha)/2)^2 C_2\} + O(n^{-2}h^{-2}),\end{aligned}$$

where, for $\ell = 0, 1, \ldots, (n/3)-2$,

$$s_i - s_{i-1} = \begin{cases}\alpha/n & \text{if } i = 3\ell+2 \\ (3-\alpha)/2n & \text{if } i = 3\ell+3,\ 3\ell+4\end{cases},$$

and where

$$A_2 = n^{-2}\sum_{\ell=0}^{(n/3)-2} K_h(x-x_{3\ell+2})^2,$$

$$B_2 = n^{-2}\sum_{\ell=0}^{(n/3)-1} K_h(x-x_{3\ell+3})^2,$$

$$C_2 = n^{-2} \sum_{\ell=0}^{(n/3)-1} K_h(x-x_{3\ell+4})^2.$$

Since $A_2$, $B_2$, and $C_2$ all have the same asymptotic expression

$$(1/3)n^{-1}h^{-1}\int K^2 + O(n^{-2}h^{-2}),$$

which follows from (2.6.2), we conclude that $\mathrm{Var}(\hat{m}_{GM}(x))$ has an asymptotic expression

$$\sigma^2\{\alpha^2 + ((3-\alpha)/2)^2 + ((3/\alpha)/2)^2\}(1/3)n^{-1}h^{-1}\int K^2 + O(n^{-2}h^{-2})$$
$$= C(\alpha)n^{-1}h^{-1}\sigma^2\int K^2 + O(n^{-2}h^{-2}).$$

The proof of (2.2.3) is complete.

Proof of (2.2.6): The case of clustering k points can be easily derived by the same procedure as for the case of $k = 3$.

Proofs of (2.3.1) through (2.3.4):

For the Nadaraya-Watson estimator, before we use Taylor's theorem to handle the denominator, we need to make sure the denominator of $\hat{m}_{NW}(x)$ is bounded above zero, i.e. there are enough number of $X_i$ in the interval $[x-\frac{h}{2}, x+\frac{h}{2}]$, where $K \geq \ell_K$. For the Gasser-Mueller estimator, before we use the integral mean value theorem to approximate the integrals which are the weights assigned to the observations, we need to make sure the approximation error is negligible, i.e. there are not too many $X_i$ in the interval $[x-h, x+h]$.

Let N be the event that the number of $X_i$ in the interval $[x-\frac{h}{2}, x+\frac{h}{2}]$ is less than $C_N \cdot n \cdot \int_{x-h/2}^{x+h/2} f(t)dt$, where $C_N$ is a positive constant and $C_N \leq 1/4$, W the event that the number of $X_i$ in the interval $[x-\frac{h}{2}, x+\frac{h}{2}]$ is greater than $C_W \cdot n \cdot \int_{x-h/2}^{x+h/2} f(t)dt$ where $C_W$ is a constant and $C_W \geq e^1$, and G the event that the number of $X_i$ in the interval $[x-h, x+h]$ is greater than $C_G \cdot n \cdot \int_{x-h}^{x+h} f(t)dt$, where $C_G$ is a constant and $C_G \geq e^1$.

For deriving the probability of the events N, W, and G, we shall use (1) and (2) of the theorem given in Section 10.3.1 of Serfling (1980) which is described below:

Theorem (Chernoff). Let $T_1, \ldots, T_n$ be independent and identically distributed random variables with distribution F and put $S_n = \sum_{j=1}^{n} T_j$. Assume existence of the moment generating function $M_T(z) = E_F(e^{zT})$, z real, and put $b(\tau) = \inf_{z \in R} e^{-z\tau} M_T(z)$.

If $-\infty < \tau \leq E(T)$, then

(1) $$P(S_n \leq n\tau) \leq [b(\tau)]^n.$$

If $E(T) \leq \tau < \infty$, then

(2) $$P(S_n \geq n\tau) \leq [b(\tau)]^n.$$

Using (1) and (2) of the theorem, as $n \longrightarrow \infty$, we have

(2.6.4) $$P(N \cup W) \leq [1 - C_N \int_{x-h/2}^{x+h/2} f(t)dt]^n + [1 - \int_{x-h/2}^{x+h/2} f(t)dt]^n$$
$$\leq [1 - C_N \ell_f h]^n + [1 - \ell_f h]^n = O(\exp(-nh)),$$

(2.6.5) $$P(G) \leq [1 - \int_{x-h}^{x+h} f(t)dt]^n \leq [1 - \ell_f h]^n = O(\exp(-nh)),$$

where $\ell_f$ is a positive lower bound of f on [0,1]. The proofs of (2.6.4) and (2.6.5) are given at the end of this section.

Notation used to prove (2.3.1) and (2.3.3) include:

$$Z_n = n^{-1} \sum_{i=1}^{n} K_h(x-X_i),$$

$$M_n = n^{-1} \sum_{i=1}^{n} K_h(x-X_i)m(X_i),$$

$$B_n = n^{-1} \sum_{i=1}^{n} K_h(x-X_i)(m(X_i) - m(x)),$$

$$V_n = n^{-2} \sum_{i=1}^{n} K_h(x-X_i)^2,$$

$$\mu = \int K_h(x-t)f(t)\, dt,$$

$$\delta = \int K_h(x-t)m(t)f(t)\,dt,$$
$$\zeta = \int K_h(x-t)(m(t) - m(x))f(t)\,dt.$$

Because $X_i$ are IID random variables, and K, m, and f are Hoelder continuous, then $Z_n$, $M_n$, and $V_n$ have the following properties:

(a) Given the event $(NUW)^c$, the denominator $Z_n$ of $\hat{m}_{NW}(x)$ is bounded by $C_W \cdot f_{max} \cdot K_{max} \geq Z_n \geq C_N \cdot \ell_f \cdot \ell_K > 0$, since there are O(nh) terms in $\sum_{i=1}^{n} K_h(x-X_i)$.

(b)
$$E(Z_n) = \mu = f(x) + O(h),$$
$$E(M_n) = \delta = m(x)f(x) + O(h),$$
$$E(B_n) = \zeta = O(h),$$
$$E(V_n) = n^{-1}\int K_h(x-t)^2 f(t)\,dt = n^{-1}h^{-1}f(x)\int K^2 + O(n^{-1}).$$

(c)
$$E(Z_n^2) = n^{-1}\int K_h(x-t)^2 f(t)\,dt + (1-n^{-1})[\int K_h(x-t)f(t)\,dt]^2$$
$$= n^{-1}h^{-1}f(x)\int K^2 + \mu^2 + O(n^{-1}) = O(1),$$
$$E(M_n^2) = n^{-1}\int K_h(x-t)^2 m(t)^2 f(t)\,dt + (1-n^{-1})[\int K_h(x-t)m(t)f(t)\,dt]^2$$
$$= n^{-1}h^{-1}m(x)^2 f(x)\int K^2 + \delta^2 + O(n^{-1}) = O(1),$$
$$E(M_n Z_n) = n^{-1}\int K_h(x-t)^2 m(t)f(t)\,dt +$$
$$(1-n^{-1})[\int K_h(x-t)m(t)f(t)\,dt][\int K_h(x-t)f(t)\,dt]$$
$$= n^{-1}h^{-1}m(x)f(x)\int K^2 + \delta\mu + O(n^{-1}),$$
$$E(V_n^2) = O(n^{-2}h^{-2}).$$

(d) Given the event $(NUW)^c$, there are O(nh) terms in the $\sum_{i=1}^{n} K_h(x-X_i)$. Using Whittle's inequality (Whittle (1960)) as described in Section 3.5, we have, for each k = 1, 2, ....,

$$E((\mu - Z_n)^{2k} \mid (NUW)^c) = O(n^{-k}h^{-k+1}),$$
$$E((\delta - M_n)^{2k} \mid (NUW)^c) = O(n^{-k}h^{-k+1}),$$
$$E((\zeta - B_n)^{2k} \mid (NUW)^c) = O(n^{-k}h^{k+1}),$$
$$E(V_n^2 \mid (NUW)^c) = O(n^{-2}),$$

$$E(M_n^2 \mid (NUW)^c) = O(n^{-1} + h^2) = O(1).$$

For the proof of (2.3.3), using the conditional expectation, the bias of $\hat{m}_{NW}(x)$ can be expressed as

(2.6.6) $$\text{Bias}(\hat{m}_{NW}(x)) = E(B_n/Z_n)$$

$$= E(B_n/Z_n \mid NUW)P(NUW) + E(B_n/Z_n \mid (NUW)^c)P((NUW)^c).$$

Using $P(NUW) = O(\exp(-nh))$ and $|B_n/Z_n| \le 2|m|_{max}$, the first term of the right hand side in (2.6.6) is

$$E(B_n/Z_n \mid NUW)P(NUW) = O(\exp(-nh)).$$

Using property (a) and applying the first order Taylor's theorem to a reciprocal function, the second term of the right hand side in (2.6.6) can be expressed as

$$E(B_n/Z_n \mid (NUW)^c) = E(B_n(\mu^{-1} + (\mu - Z_n)\eta_n^{-2}) \mid (NUW)^c),$$

where $\eta_n$ is between $\mu$ and $Z_n$, i.e. $\eta_n^{-2} = O(1)$. Using properties (b) and (d), we have

$$|E(B_n(\mu-Z_n)\eta_n^{-2} \mid (NUW)^c)|$$
$$\le c|E((B_n-\zeta)(\mu-Z_n) \mid (NUW)^c) + \zeta E((\mu-Z_n) \mid (NUW)^c)|$$
$$\le c|E((B_n-\zeta)(\mu-Z_n) \mid (NUW)^c)|$$
$$\le c[E((B_n-\zeta)^2 \mid (NUW)^c)\cdot E((\mu-Z_n)^2 \mid (NUW)^c)]^{1/2}$$
$$= c[O(n^{-1}h^2)\cdot O(n^{-1})]^{1/2} = O(n^{-1}h).$$

Using the fact that

$$E(B_n\mu^{-1}) = \zeta/\mu$$
$$= E(B_n\mu^{-1} \mid NUW)P(NUW) + E(B_n\mu^{-1} \mid (NUW)^c)P((NUW)^c)$$
$$= O(\exp(-nh)) + E(B_n\mu^{-1} \mid (NUW)^c)P((NUW)^c),$$

then we have

$$E(B_n\mu^{-1} \mid (NUW)^c)P((NUW)^c) = \zeta/\mu + O(\exp(-nh)).$$

Then the bias of $\hat{m}_{NW}(x)$ can be expressed as

$$\text{Bias}(\hat{m}_{NW}(x)) = \zeta/\mu + O(n^{-1}h) + O(\exp(-nh))$$

$$= \int K_h(x-t)(m(t)-m(x))f(t)dt \Big/ \int K_h(x-t)f(t)dt + O(n^{-1}h).$$

The proof of (2.3.3) is complete.

For the proof of (2.3.1), using the fact that $X_i$ and $\epsilon_i$ are uncorrelated, the variance of $\hat{m}_{NW}(x)$ can be expressed as

(2.6.7) $$\mathrm{Var}(\hat{m}_{NW}(x)) = \sigma^2 E(V_n/(Z_n^2)) + \mathrm{Var}(M_n/Z_n).$$

Now we check the first term of the right hand side in (2.6.7). Using the conditional expectation, it can be expressed as

$$E(V_n/(Z_n^2)) = E(V_n/(Z_n^2) \mid NUW)P(NUW) + E(V_n/(Z_n^2) \mid (NUW)^c)P((NUW)^c).$$

Using $P(NUW) = O(\exp(-nh))$ and $|V_n/(Z_n^2)| \leq 1$, we have

$$E(V_n/(Z_n^2) \mid NUW)P(NUW) = O(\exp(-nh)).$$

Using property (a) and applying the first order Taylor's theorem to a reciprocal function, we have

$$\begin{aligned} &E(V_n/(Z_n^2) \mid (NUW)^c) \\ &= E(V_n(\mu^{-1} + (\mu-Z_n)\eta_n^{-2})^2 \mid (NUW)^c) \\ &= E(V_n\mu^{-2} + 2V_n\mu^{-1}(\mu-Z_n)\eta_n^{-2} + V_n(\mu-Z_n)^2\eta_n^{-4} \mid (NUW)^c) \end{aligned}$$

where $\eta_n$ is between $\mu$ and $Z_n$, i.e. $\eta_n^{-2}$ and $\eta_n^{-4}$ are $O(1)$. Using properties (b), (c), and (d), we have the following expressions:

$$\begin{aligned} &|E(V_n\mu^{-1}(\mu-Z_n)\eta_n^{-2} \mid (NUW)^c)| \\ &\leq c|E(V_n(\mu-Z_n) \mid (NUW)^c)| \\ &\leq c[E(V_n^2 \mid (NUW)^c)\cdot E((\mu-Z_n)^2 \mid (NUW)^c)]^{1/2} \\ &= c[O(n^{-2})\cdot O(n^{-1})]^{1/2} = o(n^{-1}h^{-1}), \end{aligned}$$

and

$$\begin{aligned} &|E(V_n(\mu-Z_n)^2\eta_n^{-4} \mid (NUW)^c)| \\ &\leq c|E(V_n(\mu-Z_n)^2 \mid (NUW)^c)| \\ &\leq c[E(V_n^2 \mid (NUW)^c)\cdot E((\mu-Z_n)^4 \mid (NUW)^c)]^{1/2} \\ &= c[O(n^{-2})\cdot O(n^{-2}h^{-1})]^{1/2} = o(n^{-1}h^{-1}). \end{aligned}$$

Thus, the first term of the right hand side in (2.6.7) becomes

$$E(V_n/(Z_n)^2) = E(V_n\mu^{-2} \mid (NUW)^c)P((NUW)^c) + o(n^{-1}h^{-1}).$$

Using property (b), we have

$$E(V_n\mu^{-2}) = n^{-1}h^{-1}f(x)\int K^2(f(x) + O(h))^{-2}$$
$$= n^{-1}h^{-1}(f(x))^{-1}\int K^2 + o(n^{-1}h^{-1}).$$

Using the conditional expectation, we have

$$E(V_n\mu^{-2}) = E(V_n\mu^{-2} \mid NUW)P(NUW) + E(V_n\mu^{-2} \mid (NUW)^c)P((NUW)^c)$$
$$= O(n^{-1}h^{-2}\cdot \exp(-nh)) + E(V_n\mu^{-2} \mid (NUW)^c)P((NUW)^c)$$
$$= o(n^{-1}h^{-1}) + E(V_n\mu^{-2} \mid (NUW)^c)P((NUW)^c).$$

Combining the two results, we have

$$E(V_n\mu^{-2}) \mid (NUW)^c)P((NUW)^c) = n^{-1}h^{-1}(f(x))^{-1}\int K^2 + o(n^{-1}h^{-1}),$$

i.e.

$$E(V_n/(Z_n^2)) = n^{-1}h^{-1}(f(x))^{-1}\int K^2 + o(n^{-1}h^{-1}).$$

Now we check the second term of the right hand side in (2.6.7). Let

$$L_n = M_n/Z_n - E(M_n/Z_n).$$

Using the conditional expectation, the second term of the right hand side in (2.6.7) can be expressed as

$$Var(M_n/Z_n) = E(L_n^2 \mid NUW)P(NUW) + E(L_n^2 \mid (NUW)^c)P((NUW)^c).$$

Using $P(NUW) = O(\exp(-nh))$ and $|L_n^2| \le (2|m|_{max})^2$, we have

$$E(L_n^2 \mid NUW)P(NUW) = O(\exp(-nh)).$$

Using property (a) and $E(M_n/Z_n) = \delta/\mu + O(n^{-1}h)$ as given in (2.6.6), and applying the second order Taylor's theorem to a reciprocal function, we have

$$E(L_n^2 \mid (NUW)^c)$$
$$= E([M_n(\mu^{-1} + \mu^{-2}(\mu-Z_n) + \eta_n^{-3}(\mu-Z_n)^2) - \delta/\mu + O(n^{-1}h)]^2 \mid (NUW)^c)$$
$$= E([A + B + C + D]^2 \mid (NUW)^c),$$

where $\eta_n$ is between $\mu$ and $Z_n$, i.e. $\eta_n^{-3} = O(1)$, and where

$$A = \mu^{-1}(M_n - (\delta/\mu)Z_n),$$
$$B = \mu^{-2}(M_n - \delta)(\mu - Z_n),$$
$$C = M_n\eta_n^{-3}(\mu - Z_n)^2,$$
$$D = O(n^{-1}h).$$

Using properties (b), (c), and (d), we have $E(B^2 \mid (N\cup W)^c)$, $E(C^2 \mid (N\cup W)^c)$, and $E(D^2 \mid (N\cup W)^c)$ are $o(n^{-1}h^{-1})$. The proof of (2.3.1) is complete when we show that $E(A^2 \mid (N\cup W)^c)$ is $o(n^{-1}h^{-1})$. Using the conditional expectation, $P(N\cup W) = O(\exp(-nh))$, $\delta/\mu = m(x) + O(h)$, $0 \leq A^2 < (K_{max}L_m)^2$ where $L_m$ is the constant of the Hoelder continuity of m, and $E(A^2 \mid N\cup W)P(N\cup W) = o(n^{-1}h^{-1})$, we have

$$\begin{aligned}
&E(A^2 \mid (N\cup W)^c) \\
&= E(A^2) + o(n^{-1}h^{-1}) \\
&= \mu^{-2}[E(M_n{}^2) - 2(\delta/\mu)E(M_nZ_n) + (\delta/\mu)^2E(Z_n{}^2)] + o(n^{-1}h^{-1}) \\
&= \mu^{-2}[n^{-1}h^{-1}\int K^2(m(x)^2f(x) - 2(\delta/\mu)m(x)f(x) + (\delta/\mu)^2f(x))] + o(n^{-1}h^{-1}) \\
&= \mu^{-2}n^{-1}h^{-1}\int K^2\cdot O(h) + o(n^{-1}h^{-1}) = o(n^{-1}h^{-1}).
\end{aligned}$$

The proof of (2.3.1) is complete.

Now we give notation and properties for the proofs of (2.3.2) and (2.3.4). Let

$$Q_i = F^{-1}(i/(n+1)),$$
$$d_i = X_{(i)} - X_{(i-1)},$$

for i = 2, 3, ..., n, where F is a distribution function of the design density f, and $F^{-1}(u) = \inf\{t \in R: F(t) \geq u\}$. Let

$$D_i = s_i - s_{i-1} = \beta d_i + (1-\beta)d_{i+1},$$

for i = 2, 3, ..., n-1.

We shall use the sampling distribution, the exponential-type probability inequality of Kolmogorov-Smirnov distance as given in Section 2.1.3 of Serfling (1980), and the Hoelder continuity of f, to derive a relationship between $f(x)$ and $f(Q_i)$, for each $X_{(i)} \in [x-h,x+h]$. If $X_{(i)}$ falls into the interval $[x-h,x+h]$, then (without loss of generality), we have

$$F_n(x-h) \leq i/(n+1) \leq F_n(x+h).$$

For all $n = 1, 2, \ldots,$ we have

$$P(\sup_{t\in R} |F_n(t) - F(t)| > d) \leq \exp(-2nd^2).$$

Taking $d = n^{-0.4}$, there is a probability $1-\exp(-2n^{0.2})$ that

$$F(x-h)-n^{-0.4} \leq F_n(x-h) \leq i/(n+1) \leq F_n(x+h) \leq F(x+h)+n^{-0.4}.$$

Thus we have, asymptotically,

$$F^{-1}(F(x-h)-n^{-0.4}) \leq Q_i \leq F^{-1}(F(x+h)+n^{-0.4}).$$

Because of the Hoelder continuity and the positive lower bound of f, we have

$$x-h+O(n^{-0.4}) \leq Q_i \leq x+h+O(n^{-0.4}),$$

which implies, by the Hoelder continuity of f,

$$f(Q_i) = f(x) + O(h + n^{-0.4}).$$

Using the results as given in Section 4.6 of David (1981) and the Hoelder continuity of f, we have the following properties:

(e) $E(d_i) = (nf(Q_i))^{-1} + O(n^{-2}),$

$$E(d_i d_j) = \begin{cases} 2(n^2 f(Q_i)^2)^{-1} + O(n^{-3}) & \text{if } i = j \\ (n^2 f(Q_i) f(Q_j))^{-1} + O(n^{-3}) & \text{if } i \neq j \end{cases},$$

$$E(D_i) = (nf(Q_i))^{-1} + O(n^{-2}),$$

$$E(D_i D_j) = \begin{cases} 2(1-\beta+\beta^2)(n^2 f(Q_i)^2)^{-1} + O(n^{-3}) & \text{if } i = j \\ 2(1+\beta-\beta^2)(n^2 f(Q_i)^2)^{-1} + O(n^{-3}) & \text{if } |i-j| = 1, \\ (n^2 f(Q_i)^2)^{-1} + O(n^{-3}) & \text{if } |i-j| \geq 2 \end{cases}$$

$E(D_i^k) = O(n^{-k})$, for $k = 2, 3, 4,$

for $i, j = 2, 3, \ldots, n-1$. The proof of property (e) are given at the end of this section. Whenever $X_{(i)}, X_{(j)} \in [x-h,x+h]$, using $f(Q_i) =$

$f(x) + O(h + n^{-0.4})$, property (e) can be modified as

(f) $E(d_i) = (nf(x))^{-1} + O(n^{-1}h + n^{-1.4})$,

$$E(d_i d_j) = \begin{cases} 2(n^2 f(x)^2)^{-1} + O(n^{-2}h + n^{-2.4}) & \text{if } i = j \\ (n^2 f(x)^2)^{-1} + O(n^{-2}h + n^{-2.4}) & \text{if } i \neq j \end{cases},$$

$E(D_i) = (nf(x))^{-1} + O(n^{-1}h + n^{-1.4})$,

$$E(D_i D_j) = \begin{cases} 2(1-\beta+\beta^2)(n^2 f(x)^2)^{-1} + O(n^{-2}h + n^{-2.4}) & \text{if } i = j \\ 2(1+\beta-\beta^2)(n^2 f(x)^2)^{-1} + O(n^{-2}h + n^{-2.4}) & \text{if } |i-j| = 1, \\ (n^2 f(x)^2)^{-1} + O(n^{-2}h + n^{-2.4}) & \text{if } |i-j| \geq 2 \end{cases}$$

for $i, j = 2, 3, \ldots, n-1$.

For the proof of (2.3.4), letting

$$I_n = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)(m(X_{(i)}) - m(t))\, dt,$$

then the bias of $\hat{m}_{GM}(x)$ can be expressed as

$$\begin{aligned} \text{Bias}(\hat{m}_{GM}(x)) &= E\left( \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)(m(X_{(i)}) - m(x))\, dt\right) \\ &= \int K_h(x-t)(m(t)-m(x))dt + E(I_n). \end{aligned}$$

Using the conditional expectation, $E(I_n)$ can be expressed as

$$E(I_n) = E(I_n \mid G)P(G) + E(I_n \mid G^c)P(G^c).$$

Using $P(G) = O(\exp(-nh))$ and $|I_n| \leq 2|m|_{max}$, we have

$$E(I_n \mid G)P(G) = O(\exp(-nh)).$$

Using the Hoelder continuity of m, the definition of the event $G^c$, and property (e), we have

$$\begin{aligned} |E(I_n \mid G^c)| &\leq E\left( \sum_{i=1}^{n} D_i^2 h^{-1} K_{max} L_m \mid G^c\right) \\ &\leq c\left(C_G n \int_{x-h}^{x+h} f(t)\, dt\right) n^{-2} h^{-1} = O(n^{-1}). \end{aligned}$$

Then the bias of $\hat{m}_{GM}(x)$ can be expressed as

$$\text{Bias}(\hat{m}_{GM}(x)) = \int K_h(x-t)(m(t) - m(x))dt + O(n^{-1}).$$

The proof of (2.3.4) is complete.

For the proof of (2.3.2), letting

$$W_n = \sum_{i=1}^{n} \left(\int_{s_{i-1}}^{s_i} K_h(x-t)dt\right)^2,$$

$$J_n = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)m(X_{(i)})dt,$$

then the variance of $\hat{m}_{GM}(x)$ can be expressed as

(2.6.8) $$\mathrm{Var}(\hat{m}_{GM}(x)) = \sigma^2 E(W_n) + \mathrm{Var}(J_n).$$

Now we check the first term of the right hand side in (2.6.8). Using the conditional expectation, $E(W_n)$ can be expressed as

$$E(W_n) = E(W_n \,|G)P(G) + E(W_n \,|G^c)P(G^c).$$

Using $P(G) = O(\exp(-nh))$ and $0 \le W_n \le 1$, we have

$$E(W_n \,|G)P(G) = O(\exp(-nh)).$$

Using the integral mean value theorem, the Hoelder continuity of f and K, and properties (e) and (f), we have the following expressions:

$E(W_n \,|G^c)$

$= E(\sum_{i=1}^{n} D_i^2 K_h(x-X_{(i)})^2 \,|G^c) + O(n^{-2}h^{-2}),$

since $|E(\sum_{i=1}^{n} \{[\int_{s_{i-1}}^{s_i} K_h(x-t)dt]^2 - D_i^2 K_h(x-X_{(i)})^2\} \,|G^c)|$

$\le |E(\sum_{i=1}^{n} D_i^3 h^{-3} L_K K_{max}^2 \,|G^c)|$ where $L_K$ is the constant of the Hoelder continuity of K

$\le c(C_G n\int_{x-h}^{x+h} f(t)\,dt)n^{-3}h^{-3} = O(n^{-2}h^{-2}),$

$= 2(1-\beta+\beta^2)(nf(x))^{-2}E(\sum_{i=1}^{n} K_h(x-X_{(i)})^2 \,|G^c) + O(n^{-1}+n^{-1.4}h^{-1}+n^{-2}h^{-2}),$

since $|E(\sum_{i=1}^{n} [D_i^2 - 2(1-\beta+\beta^2)(nf(x))^{-2}]K_h(x-X_{(i)})^2 \,|G^c)|$

$\le (C_G n\int_{x-h}^{x+h} f(t)\,dt)\cdot O(n^{-2}h+n^{-2.4})h^{-2}K_{max}^2 = O(n^{-1}+n^{-1.4}h^{-1}),$

$= 2(1-\beta+\beta^2)(f(x))^{-2}E(n^{-2}\sum_{i=1}^{n} K_h(x-X_i)^2 \,|G^c) + O(n^{-1}+n^{-1.4}h^{-1}+n^{-2}h^{-2}),$

$= 2(1-\beta+\beta^2)n^{-1}h^{-1}(f(x))^{-1}\int K^2 + o(n^{-1}h^{-1}),$

since $P(G) = O(\exp(-nh))$, and given the event G,

$$V_n = n^{-2} \sum_{i=1}^{n} K_h(x-X_i)^2 = O(n^{-1}h^{-2}) \text{ and}$$

$$E(V_n \mid G)P(G) = O(n^{-1}h^{-2} \cdot \exp(-nh)) = o(n^{-1}h^{-1}),$$

then we have

$$\begin{aligned} E(V_n \mid G^c) &= E(V_n) - E(V_n \mid G)P(G) \\ &= E(V_n) + o(n^{-1}h^{-1}) \\ &= n^{-1}h^{-1}f(x)\int K^2 + o(n^{-1}h^{-1}). \end{aligned}$$

So, we have

$$E(W_n) = 2(1-\beta+\beta^2)n^{-1}h^{-1}(f(x))^{-1}\int K^2 + o(n^{-1}h^{-1}).$$

Now we check the second term of the right hand side in (2.6.8). Using the conditional expectation, $Var(J_n)$ can be expressed as

$$Var(J_n) = Var(J_n \mid G)P(G) + Var(J_n \mid G^c)P(G^c).$$

Using $P(G) = O(\exp(-nh))$ and $|J_n| \le |m|_{max}$, we have

$$Var(J_n \mid G)P(G) = O(\exp(-nh)).$$

Using the Hoelder continuity of m, Minkowski's inequality, and property (f), we have the following expressions:

$$E(J_n) = E(\hat{m}_{GM}(x)) = \int K_h(x-t)m(t)\, dt + O(n^{-1}),$$

and

$$\begin{aligned} & Var(J_n \mid G^c) \\ &= E((J_n - E(J_n))^2 \mid G^c) \\ &= E([\sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)m(X_{(i)})dt - \int K_h(x-t)m(t)dt + O(n^{-1})]^2 \mid G^c) \\ &= E([\sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(x-t)(m(X_{(i)}) - m(t))dt + O(n^{-1})]^2 \mid G^c) \\ &\le E([\sum_{i=1}^{n} D_i^2 h^{-1} K_{max} L_m + O(n^{-1})]^2 \mid G^c) \\ &= E([\sum_{i=1}^{n} D_i^2 h^{-1} K_{max} L_m]^2 \mid G^c) + O(n^{-1})E([\sum_{i=1}^{n} D_i^2 h^{-1} K_{max} L_m] \mid G^c) + O(n^{-2}) \\ &\le c(C_G n\int_{x-h}^{x+h} f(t)\, dt \cdot n^{-2}h^{-1})^2 + O(n^{-1})(C_G n\int_{x-h}^{x+h} f(t)\, dt \cdot n^{-2}h^{-1}) + O(n^{-2}) \\ &= O(n^{-2}). \end{aligned}$$

The proof of (2.3.2) is complete.

Proofs of (2.4.2) and (2.4.3):

Using property (e) and

$$D_i = s_i - s_{i-1} = \sum_{j=1}^{2k} \beta_j (X_{(i-k+j)} - X_{(i-k+j-1)}) = \sum_{j=1}^{2k} \beta_j d_{i-k+j},$$

for $i = k+1, k+2, \ldots, n-k$, then we have

$$E(D_i^{\,2}) = (1 + \sum_{j=1}^{2k} \beta_j^{\,2})(nf(Q_i))^{-2} + O(n^{-3}).$$

Furthermore, if $X_{(i)} \in [x-h, x+h]$, then we have (without loss of generality)

$$E(D_i^{\,2}) = (1 + \sum_{j=1}^{2k} \beta_j^{\,2})\ (nf(x))^{-2} + O(n^{-2}h + n^{-2.4}),$$

$$E(D_i^{\,k}) = O(n^{-k}) \text{ for } k = 2, 3, 4.$$

Following the same steps as in the proof of (2.3.2) and (2.3.4), the proofs of (2.4.2) and (2.4.3) are complete.

Proofs of (2.6.4) and (2.6.5):

Since the event W is the same as the event G, the proof of P(G) will be given only. We shall use (1) of the theorem as given in section 10.3.1 of Serfling (1980) to prove $P(N) = O(\exp(-nh))$. Let

$$p = \int_{x-h/2}^{x+h/2} f(t)\, dt,$$

where $p > \ell_f \cdot h$. Let

$$T_j = I_{[x-h/2,\, x+h/2]}(X_j),$$

for $j = 1, 2, \ldots, n$. So, $T_j$ are IID binomial random variables with mean p and moment generating function $M_T(z) = 1-p+pe^z$, $z \in R$. The event N is equivalent to the event that

$$S_n = \sum_{j=1}^{n} T_j \le C_N \cdot n \cdot p = n\tau.$$

where the value of $\tau$ corresponding to (1) of the theorem is

$$\tau = C_N p < E(T) = p,$$

and where $0 < C_N \le 1/4$. We have, using c to express $C_N$,

(2.6.9) $$b(\tau) = \inf_{z \in R} e^{-z\tau} M_T(z) = c^{-cp}((1-p)/(1-cp))^{1-cp}.$$

The proof of (2.6.9) is given later.

Now we need to show $b(\tau) = b(cp) \le 1 - cp$. This is equivalent to show that

$$\ln (1-cp) - \ln (b(cp)) \ge 0.$$

Using Taylor's theorem, we have the following expressions:

$$\ln (1-cp) = -cp - \sum_{k=2}^{\infty} (c^k p^k / k),$$

$$\begin{aligned} \ln (b(cp)) &= \ln [((1-p)/(1-cp))^{1-cp}] \\ &= (1-cp)[\ln (1-p) - \ln (1-cp)] \\ &= -(1-c)p - \sum_{k=2}^{\infty} [\frac{c^k - c}{k-1} - \frac{c^k - 1}{k}] p^k . \end{aligned}$$

Thus, we have

$$\begin{aligned} &\ln (1-cp) - \ln (b(cp)) \\ &= [1 - 2c + c\cdot(\ln c)]p + \sum_{k=2}^{\infty} [\frac{k(1 - (1/k) - c - c^k) + 2c^k}{k(k-1)}] p^k . \end{aligned}$$

For the coefficient of p, $1 - 2c + c\cdot(\ln c)$ is decreasing on $c \in (0,1/4]$. The minimum value of the coefficient of p is occured at $c = 1/4$, i.e.

$$1-2c+c\cdot(\ln c) \ge 1-2c+c\cdot(\ln c) \Big|_{c=1/4} > 0.$$

So, the coefficient of p is positive for all $c \in (0,1/4]$. For each $k \ge 2$, the coefficient of $p^k$ is

$$\frac{k(1 - (1/k) - c - c^k) + 2c^k}{k(k-1)},$$

where

$$1 - (1/k) - c - c^k \ge 1 - (1/2) - (2c) \ge 0,$$

when $c \in (0, 1/4]$. So the coefficients of $p^k$ are positive for all $k \geq 1$. Then by (1) of the theorem, we have, asymptotically,

$$P(N) \leq [b(cp)]^n \leq [1-cp]^n \leq [1-c \cdot \ell_f \cdot h]^n = O(\exp(-nh)),$$

where $n^{-1+\epsilon} < h < n^{-\epsilon}$.

The proof of $P(N) = O(\exp(-nh))$ is complete.

We shall use (2) of the theorem as given in section 10.3.1 of Serfling (1980) to prove $P(G) = O(\exp(-nh))$. Let

$$p = \int_{x-h}^{x+h} f(t)dt,$$

where $p > 2\ell_f \cdot h$. Let

$$T_j = I_{[x-h, x+h]}(X_j),$$

for $j = 1, 2, \ldots, n$. So, $T_j$ are IID binomial random variables with mean p and moment generating function $M_T(z) = 1-p+pe^z$, $z \in R$. The event G is equivalent to the event that

$$S_n = \sum_{j=1}^{n} T_j \geq C_G \cdot n \cdot p = n\tau,$$

where the value of $\tau$ corresponding to (2) of the theorem is

$$\tau = C_G p > E(T) = p,$$

and where $C_G \geq e^1$. Through (2.6.9), we have, using c to express $C_G$,

$$b(\tau) = \inf_{z \in R} e^{-z\tau} M_T(z) = c^{-cp}((1-p)/(1-cp))^{1-cp}.$$

Now we need to show that

$$b(cp) < 1 - p.$$

If $p > 1/c$, then (2.6.5) is true by the fact that

$$P(S_n \geq npc) = P(S_n > n) = 0 < 1-p.$$

If $p = 1/c$, then (2.6.5) is true by the fact that

$$P(S_n \geq npc) = P(S_n = n) = p^n < 1-p.$$

If $0 < p < 1/c$, then we need to show that

$$\ln(1-p) - \ln(b(cp)) \geq 0.$$

Using Taylor's theorem, we have

$$\ln(1-p) - \ln(b(cp)) = [c(\ln c - 1)]p + \sum_{k=2}^{\infty} [\frac{c^k - ck}{k(k-1)}]p^k.$$

Since $c = C_G \geq e^1$, the coefficients of $p^k$ are positive for all $k \geq 1$. Then by (2) of the theorem, we have, asymptotically,

$$P(G) \leq [b(cp)]^n \leq [1-p]^n \leq [1-2\ell_f \cdot h]^n = O(\exp(-nh)),$$

where $n^{-1+\epsilon} < h < n^{-\epsilon}$.

The proof of (2.6.5) is complete.

Proof of (2.6.9):

The moment generating function of a binomial random variable T with mean p is $M_T(z) = 1-p+pe^z$, $z \in R$. The root of the first derivative of $e^{-zcp}(1-p+pe^z)$ is $\ln(c(1-p)/(1-cp))$. The second derivative of $e^{-zcp}(1-p+pe^z)$ is

$$e^{-zcp}p[c^2p(1-p) + (1-cp)^2e^z]$$

which is positive for all $z \in R$. So $(e^{-zcp}(1-p+pe^z))$ has the unique minimum value at $z = \ln(c(1-p)/(1-cp))$, i.e.

$$(e^{-zcp}(1-p+pe^z))\Big|_{e^z=c(1-p)/(1-cp)} = c^{-cp}((1-p)/(1-cp))^{1-cp}.$$

The proof (2.6.9) is complete.

Proof of property (e):

Let $U_1, U_2, \ldots, U_n$ be IID Uniform[0,1] random variables. Let $U_{(i)}$ denote the i-th order statistic of $U_i$. Let $p_i = i/(n+1)$, $q_i = 1-p_i$, $Q_i = F^{-1}(p_i)$, $Q_i' = Q_i'(p_i)$, etc. David (1981) has the following results (including his indices):

$$(4.6.2)\ X_{(i)} = Q_i + (U_{(i)} - p_i)Q_i' + (1/2)(U_{(i)} - p_i)^2Q_i'' + (1/6)(U_{(i)} - p_i)^3Q_i''' + \ldots.$$

$$(4.6.3)\quad E(X_{(i)}) = Q_i + p_i q_i (2(n+2))^{-1} Q_i'' + p_i q_i (n+2)^{-2}[(1/3)(q_i - p_i)Q_i''' + (1/8)p_i q_i Q_i''''] + O(n^{-3}),$$

$$(4.6.4)\quad \mathrm{Var}(X_{(i)}) = p_i q_i (n+2)^{-1}(Q_i')^2 + p_i q_i (n+2)^{-2}[2(q_i - p_i)Q_i' Q_i'' + p_i q_i (Q_i' Q_i''' + (1/2)(Q_i'')^2)] + O(n^{-3}),$$

$$(4.6.5)\quad \mathrm{Cov}(X_{(i)}, X_{(j)}) = p_i q_j (n+2)^{-1} Q_i' Q_j' + p_i q_j (n+2)^{-2}[(q_i - p_i)Q_i'' Q_j' + (q_j - p_j)Q_i' Q_j'' + (1/2)(p_i q_i Q_i''' Q_j' + p_j q_j Q_i' Q_j''' + p_i q_j Q_i'' Q_j'')] + O(n^{-3}),$$

for $i < j$.

Using (4.6.3) of David (1981), the expectation of $d_i$ can be expressed as

$$E(d_i) = E(X_{(i)} - X_{(i-1)}) = (Q_i - Q_{i-1}) + (2(n+2))^{-1}[p_i q_i Q_i'' - p_{i-1} q_{i-1} Q_{i-1}''] + O(n^{-2}).$$

Using the Hoelder continuity of f, through a straightforward calculation, we have obtained the following quantities:

$$Q_i - Q_{i-1} = (nf(Q_i))^{-1} + O(n^{-2}),$$

$$Q_{i-1}'' = Q_i'' + O(n^{-1}),$$

$$p_i q_i - p_{i-1} q_{i-1} = O(n^{-1}).$$

Then

$$E(d_i) = (nf(Q_i))^{-1} + O(n^{-2}).$$

To check $E(d_i d_j)$, using (4.6.3), (4.6.4), and (4.6.5) of David (1981), through a straightforward calculation, we have obtained the following quantities:

$$E(X_{(i)}^2) = Q_i^2 + p_i q_i (n+2)^{-1}[(Q_i')^2 + Q_i Q_i''] + p_i q_i (n+2)^{-2}[(q_i - p_i)((2/3)Q_i Q_i''' + 2Q_i' Q_i'') + p_i q_i ((3/4)(Q_i'')^2 + (1/4)Q_i Q_i'''' + Q_i' Q_i''')] + O(n^{-3}),$$

$$E(X_{(i)} X_{(j)}) = Q_i Q_j + (n+2)^{-1}[p_i q_j Q_i' Q_j' + (1/2)(p_i q_i Q_i'' Q_j + p_j q_j Q_i Q_j'')]$$

$$+ (n+2)^{-2}\{ \ (q_i - p_i)[(1/3)p_i q_i Q_i''' Q_j + p_i q_j Q_i'' Q_j']$$
$$+ (q_j - p_j)[(1/3)p_j q_j Q_i Q_j''' + p_i q_j Q_i' Q_j'']$$
$$+ p_i^2 q_i[(1/2)q_j Q_i''' Q_j' + (1/8)q_i Q_i'''' Q_j]$$
$$+ p_j q_j^2[(1/2)p_i Q_i' Q_j''' + (1/8)p_j Q_i Q_j'''']$$
$$+ [(1/2)p_i^2 q_j^2 Q_i'' Q_j'' + (1/4)p_i p_j q_i q_j Q_i'' Q_j''] \ \}$$
$$+ O(n^{-3}),$$

for $i < j$. Then

$$E(d_i^2) = E(X_{(i)} - X_{(i-1)})^2 = E(X_{(i)}^2) + E(X_{(i-1)}^2) - 2E(X_{(i-1)}X_{(i)})$$
$$= A + B + C + O(n^{-3}),$$

where

$$A = Q_i^2 + Q_{i-1}^2 - 2Q_i Q_{i-1},$$
$$B = (n+2)^{-1}\{ \ p_i q_i[(Q_i')^2 + Q_i Q_i''] + p_{i-1}q_{i-1}[(Q_{i-1}')^2 + Q_{i-1}Q_{i-1}'']$$
$$- 2[p_{i-1}q_i Q_{i-1}' Q_i' + p_{i-1}q_{i-1}Q_{i-1}'' Q_i + p_i q_i Q_{i-1} Q_i''] \ \},$$
$$C = (n+2)^{-2}\{ \ (q_i - p_i)p_i q_i[(2/3)Q_i Q_i''' + 2Q_i' Q_i'']$$
$$+ p_i^2 q_i^2[(3/4)(Q_i'')^2 + (1/4)Q_i Q_i'''' + Q_i' Q_i''']$$
$$+ (q_{i-1} - p_{i-1})p_{i-1}q_{i-1}[(2/3)Q_{i-1}Q_{i-1}''' + 2Q_{i-1}' Q_{i-1}'']$$
$$+ p_{i-1}^2 q_{i-1}^2[(3/4)(Q_{i-1}'')^2 + (1/4)Q_{i-1}Q_{i-1}'''' + Q_{i-1}' Q_{i-1}''']$$
$$+ (q_{i-1} - p_{i-1})p_{i-1}[(-2/3)q_{i-1}Q_{i-1}''' Q_i - 2q_i Q_{i-1}'' Q_i']$$
$$+ (q_i - p_i)q_i[(-2/3)p_i Q_{i-1} Q_i''' - 2p_{i-1}Q_{i-1}' Q_i'']$$
$$+ p_{i-1}^2 q_{i-1}[-q_i Q_{i-1}''' Q_i' - (1/4)q_{i-1}Q_{i-1}'''' Q_i]$$
$$+ p_i q_i^2[-p_{i-1}Q_{i-1}' Q_i''' - (1/4)p_i Q_{i-1} Q_i'''']$$
$$+ [-p_{i-1}^2 q_i^2 Q_{i-1}'' Q_i'' - (1/2)p_{i-1}p_i q_{i-1} q_i Q_{i-1}'' Q_i''].$$

Using the Hoelder continuity of $f$ to check each sum of coefficients of $QQ$, $QQ''$, $Q'Q'$, $QQ'''$, $Q'Q''$, $QQ''''$, $Q'Q'''$, and $Q''Q''$, through a straightforward calculation, we have obtained the following quantities:

$$A = (n^2 f^2(Q_i))^{-1} + O(n^{-3}),$$

$$B = n^{-2}(Q_i{}')^2 + O(n^{-3}) = (n^2 f^2(Q_i))^{-1} + O(n^{-3}).$$
$$C = O(n^{-3}).$$

The proof of $E(d_i^2)$ is complete.

Use the same method to check $E(d_i d_j)$ for $i < j$, and use (4.6.2) of David (1981) to check higher moments of $X_{(i)}$. Then the proof of property (e) is complete.

# CHAPTER III

# A DATA-DRIVEN BANDWIDTH ESTIMATOR

## 3.1 Introduction

The two optimal bandwidths, $h_A$ and $h_M$, are the minimizers of the average square error (ASE or $d_A(h)$) as given in (1.11) and the mean average square error (MASE or $d_M(h)$) as given in (1.12) respectively. Based on the independent observations, it has been shown that the cross-validated bandwidth $\hat{h}_{CV}$ as given in (1.14) is asymptotically optimal, i.e. $\hat{h}_{CV}/h_A$ or $\hat{h}_{CV}/h_M$ converge to 1 in some mode. This was established by Rice (1984) in the equally spaced fixed circular design context and by Haerdle and Marron (1985) in the random design setting.

However, Hart and Wehrly (1986) showed that the MSE criterion will be affected when the observations are dependent. Hart (1987) showed that, when the observations are positively correlated, the cross-validation criterion will produce rough kernel estimates. Chiu (1987) showed that Mallows' criterion does not select a proper bandwidth when the regression errors are subjected to the first order moving average process.

When the dependent observations are considered in the nonparametric regression, a first convenient choice of dependence structure for our analysis is the class of autoregressive-moving average (ARMA) processes in time series analysis. In the nonparametric sense, the ARMA parameters will not be estimated. The ARMA process will only be used to give the regression model a short range dependence structure.

In this chapter, the equally spaced fixed design nonparametric regression model (1.1) is used and the regression errors are taken as an unknown ARMA process. In Chapter 2, it was shown that the Nadaraya-Watson estimator $\hat{m}_{NW}(x)$ as given in (1.2) is more efficient than the Gasser-Mueller estimator $\hat{m}_{GM}(x)$ as given in (1.3). Subsequently, only $\hat{m}_{NW}(x)$ will be used to estimate the unknown regression function m(x).

Looking at (1.14), the cross-validation score CV(h) is affected by the dependence structure of the observations because of the dependence between $\hat{m}_j(x_j)$ as given in (1.13) and $Y_j$. Since the observations are assumed to suffer from a short range dependence, an immediate remedy for the cross-validation score (1.14) is to leave more than one point out when $\hat{m}_j(x_j)$ are constructed. This allows us to reduce the amount of dependence between $\hat{m}_j(x_j)$ and $Y_j$. We hope that the dependence effect on the cross-validation score is negligible whenever there are enough points deleted in the construction of $\hat{m}_j(x_j)$.

Let the new version of the cross-validation score be

(3.1.1) $$CV_\ell(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}_j(x_j) - Y_j]^2 W(x_j),$$

where the subscript $\ell$ means that $\hat{m}_j(x_j)$ are a "leave-$(2\ell+1)$-out" version of $\hat{m}(x_j)$ for all j, i.e.

(3.1.2) $$\hat{m}_j(x_j) = [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)Y_i] \,/\, [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)].$$

When $\ell = 0$, $CV_\ell(h)$ is the ordinary cross-validation score CV(h) as given in (1.14). The minimizer of $CV_\ell(h)$ is denoted by $\hat{h}_{CV(\ell)}$.

The idea of the modified cross-validation has been proposed and analyzed in Collomb (1985), Haerdle and Vieu (1987), and Vieu and Hart (1989) for the settings of strong mixing data. In these settings, the modified cross-validated bandwidth is asymptotically optimal when the value of $\ell$ is taken as $O(n^r)$ for some $r > 0$. For the correlated data, Chiu (1989) and Hart (1987) proposed methods to estimate the covariance function of the regression errors, and plugged the estimated covariance function into bandwidth selection methods. Chiu (1989) and Hart (1987) showed, in the simulation study, that these plug-in bandwidth selection methods would produce good bandwidths.

The purpose of this chapter is to study how the short range dependence among the observations affects the modified cross-validation criterion $CV_\ell(h)$ for each value of $\ell$. The definition and the properties of ARMA processes are given in Section 3.2. Section 3.3 describes the asymptotic behaviors of the optimal bandwidths $h_A$ and $h_M$. The asymptotic behaviors of the data-driven bandwidths $\hat{h}_{CV(\ell)}$ are described in Section 3.4. Section 3.5 contains a discussion of our results. Finally, the proofs are given in Section 3.6.

## 3.2 ARMA Processes

In this section, we only give the definition and the properties of ARMA processes which are quoted, including the indices, from Brockwell and Davis (1987):

(a) (Definition 3.1.2 (The ARMA(p,q) Process)) The process $\{\epsilon_j\}$ is said to be an ARMA(p,q) process if $\{\epsilon_j\}$ is stationary and if for every j,

$$\phi(B)\epsilon_j = \theta(B)e_j,$$

where $\{e_j\}$ is a sequence of uncorrelated random variables with mean zero and finite variance $\sigma^2$, $\phi$ and $\theta$ are the p-th and q-th degree polynomials

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p,$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q,$$

and B is the backward shift operator defined by

$$B^k\epsilon_j = \epsilon_{j-k},$$

$$B^k e_j = e_{j-k},$$

for all integers k.

(b) (Theorems 3.1.1 and 3.1.3) If $\{\epsilon_j\}$ is an ARMA(p,q) process for which the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros and $\phi(z)$ has no zeros on $|z| \le 1$, then $\{\epsilon_j\}$ has the unique solution, for every j,

$$\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i},$$

with $\sum_{i=0}^{\infty} |\psi_i| < \infty$, where the coefficients $\{\psi_i\}$ are determined by the relation

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i = \theta(z)/\phi(z), \quad |z| \le 1.$$

(c) (Theorem 3.2.1 and Example 3.2.3) If $\{\epsilon_j\}$ is an ARMA(p,q) process for which the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros and $\phi(z)$ has no zeros on $|z| \le 1$, then the autocovariance function $\gamma(\cdot)$ of $\{\epsilon_j\}$ is determined by

$$\gamma(k) = \mathrm{Cov}(\epsilon_j, \epsilon_{j+k}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k},$$

for every j and all integers k. Then we have

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \sigma^2 \left(\sum_{i=0}^{\infty} \psi_i\right)^2.$$

The implication of $\sum_{i=0}^{\infty} |\psi_i| < \infty$ is that $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$.

(d) (Exercise 3.11) If $\{\epsilon_j\}$ is an ARMA(p,q) process for which the polynomial $\phi(z)$ is not equal to zero on $|z| = 1$, then the autocovariance function $\gamma(\cdot)$ of $\{\epsilon_j\}$ is geometrically bounded, i.e. there exist constants $C > 0$ and $s \in (0,1)$, such that, for all integers k,

$$|\gamma(k)| \le C \cdot s^{|k|},$$

and hence that

$$\sum_{k=-\infty}^{\infty} |k\gamma(k)| < \infty.$$

(e) (Theorem 7.1.1) If $\{\epsilon_j\}$ is an ARMA(p,q) process for which the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros and $\phi(z)$ has no zeros on $|z| \le 1$, then, as $n \longrightarrow \infty$,

$$n^{-1}\mathrm{Var}\left(\sum_{j=1}^{n} \epsilon_j\right) = \sum_{|k|<n} (1-|k|/n)\gamma(k) \longrightarrow \sum_{k=-\infty}^{\infty} \gamma(k),$$

where $\gamma(\cdot)$ is the autocovariance function of $\{\epsilon_j\}$. This implies

$$\sum_{k=-\infty}^{\infty} \gamma(k) \ge 0.$$

### 3.3 Behaviors of the Two Optimal Bandwidths

In this section, the asymptotic behaviors of the two optimal bandwidths $h_A$ and $h_M$ are studied. In order to derive the asymptotic behaviors, using the equally spaced fixed design nonparametric regression model (1.1) and the Nadaraya-Watson estimator (1.2), we must impose the following assumptions:

(A.1) The regression function $m(x)$ supported on the interval $[0,1]$ has a uniformly continuous and square integrable second derivative $m''(x)$ on the interval $(0,1)$.

(A.2) The kernel function K is a symmetric probability density function with support contained in the interval $[-1,1]$. The second derivative $K''$ of K is Hoelder continuous of order 1.

(A.3) The weight function W is bounded, Hoelder continuous of order 1, and compactly supported on the interval $(0,1)$ with a nonempty interior.

(A.4) The regression errors $\epsilon_j$ are an unknown ARMA(p,q) process for which the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros, $\phi(z)$ has no zeros on $|z| \leq 1$, and $e_j$ are IID random variables with mean zero and all finite moments $\mu_k = E(e_1^k)$ for all positive integers k.

(A.5) The autocovariance function $\gamma(\cdot)$ of the regression errors $\epsilon_j$ has a positive sum, i.e. $0 < \sum_{k=-\infty}^{\infty} \gamma(k) < \infty$.

(A.6) The bandwidth h is chosen from the interval $H_n = [C^{-1}n^{-1+\delta}, Cn^{-\delta}]$ for $n = 1, 2, \ldots,$ constants $C > 0$ and $\delta \in (0,1/5)$.

(A.7) The number of observations deleted in the construction of the modified cross-validation score is $2\ell+1$, with $\ell \ll nh$. The total number of observations in this regression setting is n, with $n \longrightarrow \infty$.

Let the notation $X_n = o_u(v_n)$ mean that, as $n \longrightarrow \infty$, $|X_n/v_n| \longrightarrow 0$ almost surely, and uniformly on $H_n$ if $v_n$ involves h. As $n \longrightarrow \infty$, under the above assumptions, it is shown in Section 3.6 that $d_M(h)$ and $d_A(h)$ have the following asymptotic expressions:

$$d_M(h) = d_M^T(h) + o(n^{-1}h^{-1} + h^4), \tag{3.3.1}$$

$$d_A(h) = d_M^T(h) + o_u(n^{-1}h^{-1} + h^4), \tag{3.3.2}$$

where the superscript T represents a truncation and

$$d_M^T(h) = a_1 n^{-1}h^{-1} + b_1 h^4,$$

and where

$$a_1 = (\sum_{k=-\infty}^{\infty} \gamma(k)) \int K^2 \int W,$$

$$b_1 = (1/4)(\int u^2 K)^2 \int (m'')^2 W.$$

Here and throughout this chapter, the notation $\int$ denotes $\int du$. For the components of mean average square error, the terms $a_1 n^{-1}h^{-1}$ and $b_1 h^4$ represent the variance and bias square respectively. Through a straightforward calculation, $d_M^T(h)$ has the unique minimizer

$$h_M^T = C_0 n^{-1/5}, \tag{3.3.3}$$

where

$$C_0 = [a_1(4b_1)^{-1}]^{1/5} = [(\sum_{k=-\infty}^{\infty} \gamma(k)) \int K^2 \int W (\int u^2 K)^{-2} (\int (m'')^2 W)^{-1}]^{1/5}.$$

Here, if the coefficient $b_1$ is zero, then $d_M^T(h)$ has no minimizer. This implies that $d_M(h)$ and $d_A(h)$ have no minimizer asymptotically. Also, if h is zero, then $d_M^T(h)$ is $\gamma(0)\int W$. The implication of (3.3.1) and (3.3.2) is that $h_A$ and $h_M$ are of the same order $n^{-1/5}$ as $h_M^T$.

The rest of this section is devoted to deriving the relationships between $h_A$, $h_M$, and $h_M^T$. Let

$$D_M(h) = d_M(h) - d_M^T(h) = o(n^{-1}h^{-1} + h^4),$$

$$D_A(h) = d_A(h) - d_M^T(h) = o_u(n^{-1}h^{-1} + h^4).$$

Observe that

$$0 = d_M'(h_M) = d_M^{T\,\prime}(h_M) + D_M'(h_M) \tag{3.3.4}$$
$$= (h_M - h_M^T)d_M^{T\,\prime\prime}(h_M^{*}) + D_M'(h_M),$$

where $h_M^{*}$ lies inbetween $h_M$ and $h_M^T$, and

$$0 = d_A'(h_A) = d_M^{T\,\prime}(h_A) + D_A'(h_A) \tag{3.3.5}$$
$$= (h_A - h_M^T)d_M^{T\,\prime\prime}(h_A^{*}) + D_A'(h_A),$$

where $h_A^{*}$ lies inbetween $h_A$ and $h_M^T$. Since $h_A$, $h_M$, and $h_M^T$ are all of the same order $n^{-1/5}$, then $h_M^{*}$ and $h_A^{*}$ are of the order $n^{-1/5}$. Following the fact that $d_M^{T\,\prime\prime}(h) = O(n^{-1}h^{-3}+h^2)$, $D_M'(h) = o(n^{-1}h^{-2}+h^3)$, and $D_A'(h) = o_u(n^{-1}h^{-2}+h^3)$, we have

$$d_M^{T\,\prime\prime}(h_M^{*}) = O(n^{-2/5}),$$

$$D_M'(h_M) = o(n^{-3/5}),$$

$$d_M^{T\,\prime\prime}(h_A^{*}) = O_p(n^{-2/5}),$$

$$D_A'(h_A) = o_u(n^{-3/5}).$$

Thus, it is concluded that, as $n \longrightarrow \infty$, combining with (3.3.4), (3.3.5), and $h_M^T = O(n^{-1/5})$, we have the following results:

$$h_M \,/\, h_M^T \longrightarrow 1, \tag{3.3.6}$$

$$h_A \,/\, h_M \longrightarrow 1 \quad \text{a.s.} \tag{3.3.7}$$

## 3.4 Behaviors of the Modified Cross-validated Bandwidth

In this section, the asymptotic behavior of $\hat{h}_{CV(\ell)}$, the minimizer of $CV_\ell(h)$, is discussed for each $\ell \ll nh$. Through adding and subtracting the terms $\hat{m}(x_j)$ and $m(x_j)$, the modified cross-validation score $CV_\ell(h)$ can be expressed as

$$(3.4.1)\quad CV_\ell(h) = n^{-1} \sum_{j=1}^{n} \epsilon_j^2 W(x_j) + d_A(h) - 2Cross_\ell(h) + Remainder_\ell(h),$$

where

$$Cross_\ell(h) = n^{-1} \sum_{j=1}^{n} \epsilon_j[\hat{m}_j(x_j) - m(x_j)]W(x_j),$$

$$Remainder_\ell(h) = n^{-1} \sum_{j=1}^{n} [\hat{m}_j(x_j) - \hat{m}(x_j)][\hat{m}_j(x_j) + \hat{m}(x_j) - 2m(x_j)]W(x_j),$$

and where $\hat{m}_j(x_j)$ is a "leave-$(2\ell+1)$-out" version of $\hat{m}(x_j)$ as given in (3.1.2) for every j. For each $\ell \ll nh$, under the assumptions as given in Section 3.3, as $n \longrightarrow \infty$, it is shown in Section 3.6 that

$$(3.4.2)\qquad Remainder_\ell(h) = o_u(n^{-1}h^{-1} + h^4),$$

$$(3.4.3)\qquad E(Cross_\ell(h)) = 2n^{-1}h^{-1}\Big(\sum_{k>\ell} \gamma(k)\Big)K(0)\int W + O(\ell n^{-2}h^{-2}),$$

$$(3.4.4)\qquad Cross_\ell(h) = E(Cross_\ell(h)) + o_u(n^{-1}h^{-1} + h^4).$$

Looking at (3.4.1) through (3.4.4), we see that the effect of the dependent observations on the modified cross-validation criterion $CV_\ell(h)$ is determined asymptotically by the term $-2Cross_\ell(h)$ as given in (3.4.1). In the rest of this section, we shall study how this term affects the modified cross-validation criterion.

For each $\ell \ll nh$, let

(3.4.5) $$d^S_{M\ell}(h) = d_M(h) - 2E(\mathrm{Cross}_\ell(h)),$$

where the superscript S represents a subtraction and the subscript $\ell$ applies for the "leave-($2\ell$+1)-out" version of the modified cross-validation. Using (3.3.1), (3.3.2), and (3.4.1) through (3.4.4), we obtain the following asymptotic expressions for each $\ell \ll nh$:

(3.4.6) $$d^S_{M\ell}(h) = a^S_{1\ell}n^{-1}h^{-1} + b_1h^4 + o(n^{-1}h^{-1} + h^4),$$

(3.4.7) $$CV_\ell(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j^2 W(x_j) + d^S_{M\ell}(h) + o_u(n^{-1}h^{-1} + h^4)$$

$$= n^{-1}\sum_{j=1}^{n} \epsilon_j^2 W(x_j) + a^S_{1\ell}n^{-1}h^{-1} + b_1h^4 + o_u(n^{-1}h^{-1} + h^4),$$

where

$$a^S_{1\ell} = [(\sum_{k=-\infty}^{\infty} \gamma(k))\int K^2 - 4(\sum_{k>\ell} \gamma(k))K(0)]\int W.$$

If $a^S_{1\ell} > 0$, then, through a straightforward calculation, $a^S_{1\ell}n^{-1}h^{-1} + b_1h^4$ has the unique minimizer $C^S_{0\ell}n^{-1/5}$, where

$$C^S_{0\ell} = [a^S_{1\ell}(4b_1)^{-1}]^{1/5}$$

$$= [(\sum_{k=-\infty}^{\infty} \gamma(k)\int K^2 - 4\sum_{k>\ell} \gamma(k)K(0))\int W(\int u^2K)^{-2}(\int (m'')^2W)^{-1}]^{1/5}.$$

If $a^S_{1\ell}$ is negative or zero, then $a^S_{1\ell}n^{-1}h^{-1} + b_1h^4$ has no minimizer on $H_n$ asymptotically.

Let $h^S_{M\ell}$ be the minimizer of $d^S_{M\ell}(h)$. Since the first term of the right hand side in (3.4.7), $n^{-1}\sum_{j=1}^{n} \epsilon_j^2 W(x_j)$, is independent of h, and following the same arguments in (3.3.4) through (3.3.7), then we have the following asymptotic results for any $\ell \ll nh$ as $C^S_{0\ell} > 0$:

(3.4.8) $$h^S_{M\ell} = C^S_{0\ell} n^{-1/5}(1+o(1)).$$

(3.4.9) $$\hat{h}_{CV(\ell)} = C^S_{0\ell} n^{-1/5}(1+o_u(1)).$$

Finally, combining (3.3.6) with (3.4.9), we have, as $n \longrightarrow \infty$ and $C^S_{0\ell} > 0$,

(3.4.10) $$\hat{h}_{CV(\ell)} / h_M \longrightarrow (C^S_{0\ell} / C_0) \quad \text{a.s.}$$

where

$$(C^S_{0\ell} / C_0) = [1 - (4K(0)/\int K^2)(\sum_{k>\ell} \gamma(k) / \sum_{k=-\infty}^{\infty} \gamma(k))]^{1/5}.$$

3.5 Discussion

From (3.3.7), (3.4.7), (3.4.10), and the coefficient $a^S_{1\ell}$, we arrive at the following ideas about the effect of the dependence structure on the modified cross-validation criterion:

a. It moves $\hat{h}_{CV(\ell)}$ toward the right or the left of $h_M$ depending on the value of $\sum_{k>\ell} \gamma(k)$. If $\sum_{k>\ell} \gamma(k)$ is positive, then $\hat{h}_{CV(\ell)}$ is on the left side of $h_M$ and the value of the ratio $C^S_{0\ell}/C_0$ is between 0 and 1. If $\sum_{k>\ell} \gamma(k)$ is too large, i.e. $C^S_{0\ell} \leq 0$, then $CV_\ell(h)$ has no minimizer on $H_n$. If $\sum_{k>\ell} \gamma(k)$ is negative, then $\hat{h}_{CV(\ell)}$ is on the right side of $h_M$ and the value of the ratio $C^S_{0\ell}/C_0$ is greater than 1.

b. As the value of $\ell$ increases, i.e. $a^S_{1\ell}$ approaches $a_1$, then $\hat{h}_{CV(\ell)}$ approaches $h_A$ or $h_M$.

To illustrate the above ideas, we do a simulated regression. Figures 3.1, 3.2, 3.3, and 3.4 use the average of 100 curves of $CV_\ell(h)$ and $d_A(h)$ generated from the pseudo AR(1): $\epsilon_j = \phi\epsilon_{j-1} + e_j$, and MA(1):

Figure 3.1: Relationships among the modified cross-validated bandwidths (solid vertical lines) and the optimal bandwidth (dashed vertical line) for the average function and the positively correlated observations.

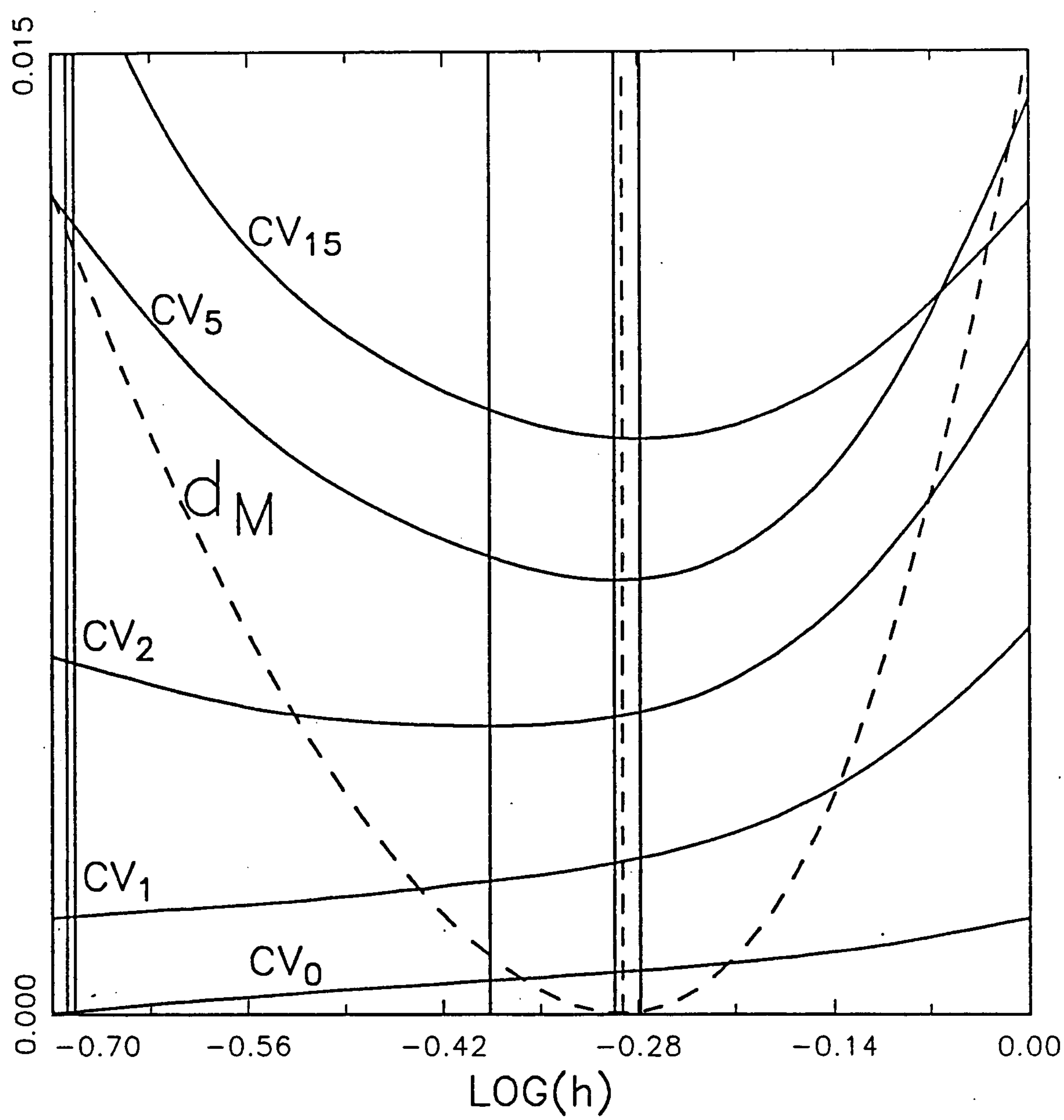


Figure 3.2: Relationships among the modified cross-validated bandwidths (solid vertical lines) and the optimal bandwidth (dashed vertical line) for the average function and the negatively correlated observations.

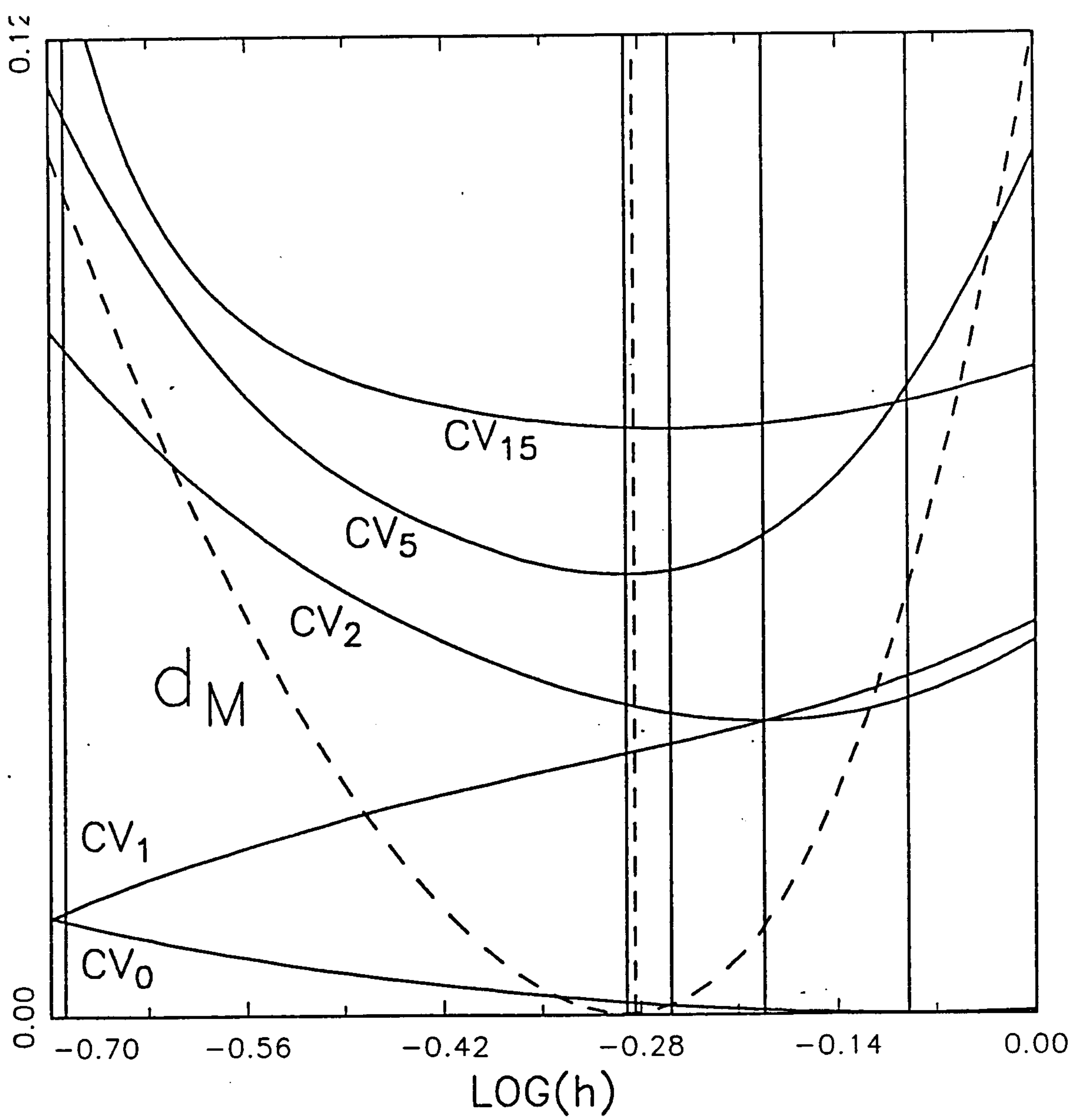


Figure 3.3: Relationships among the modified cross-validated bandwidths (solid vertical lines) and the optimal bandwidth (dashed vertical line) for the average function and the one step positively correlated observations.

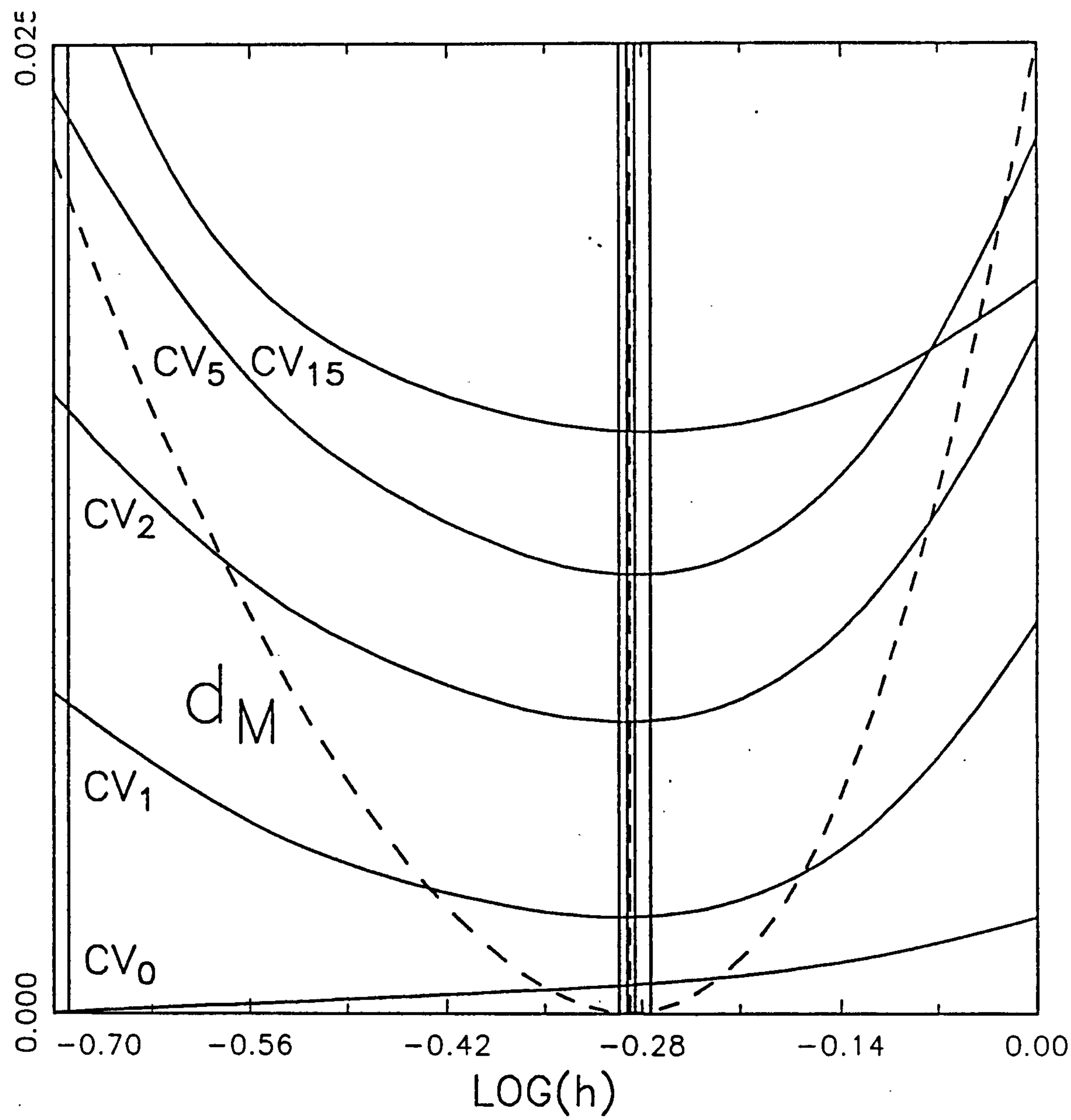
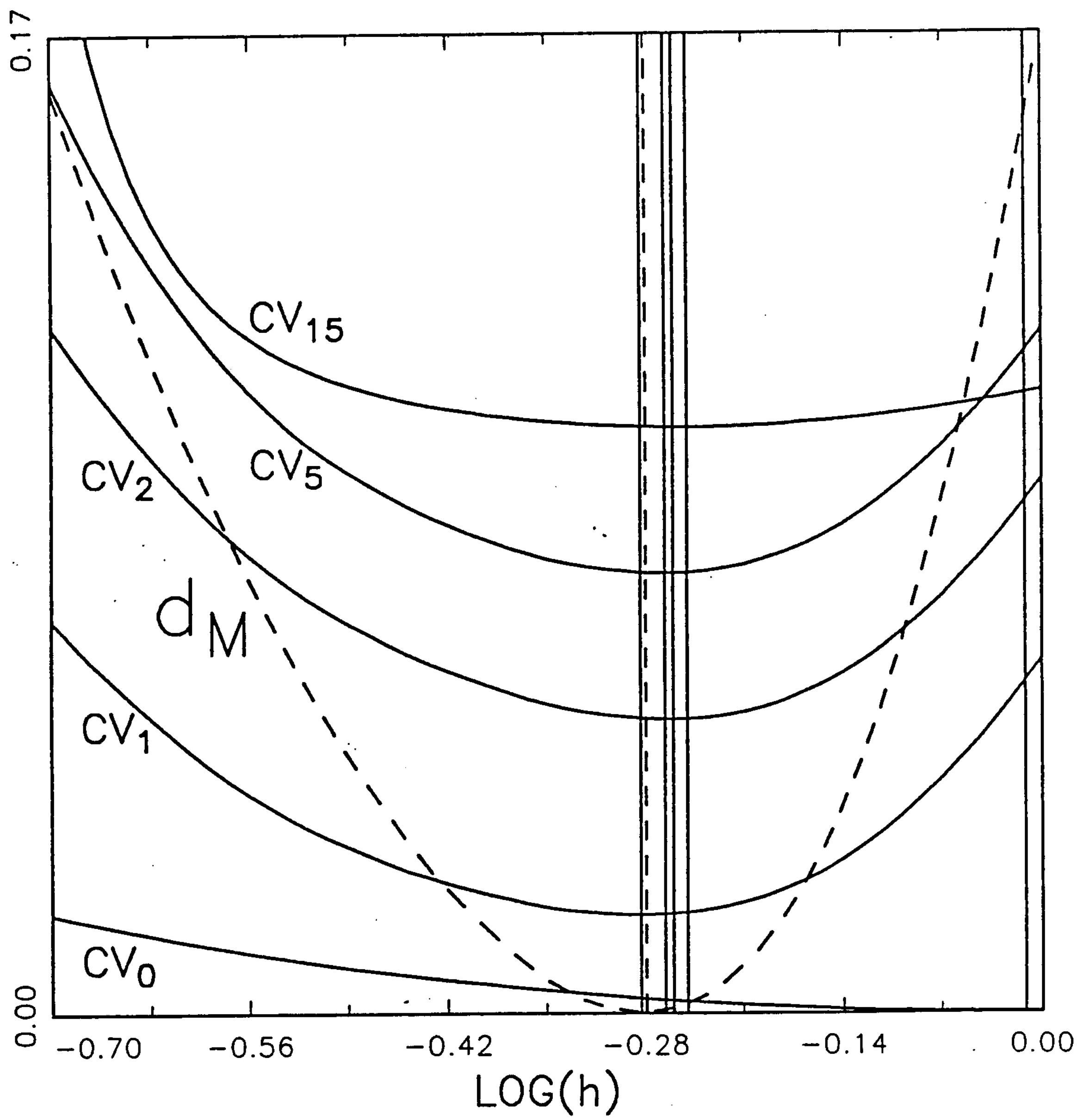


Figure 3.4: Relationships among the modified cross-validated bandwidths (solid vertical lines) and the optimal bandwidth (dashed vertical line) for the average function and the one step negatively correlated observations.



$\epsilon_j = \theta e_{j-1} + e_j$, where $e_j$ are IID $N(0,\sigma^2)$, processes of regression errors. The sample size is $n = 200$, and the values of $\ell$ are $\ell = 0$, 1, 2, 5, and 15. The regression function is given as $m(x) = x^3(1-x)^3$ for $0 \le x \le 1$. The kernel function is taken as the bi-weight kernel, $K(x) = (15/8)(1-4x^2)^2 I_{[-1/2,1/2]}(x)$. The weight function is taken as $W(x) = (5/3)I_{[1/5,4/5]}(x)$. The Nadaraya-Watson estimator (1.2) is applied to get kernel estimates. Figure 3.1 shows that the case of AR(1) regression errors with $\phi = 0.5$ and $\sigma^2 = 0.00885^2$, and Figure 3.2 shows $\phi = -0.5$ and $\sigma^2 = 0.02654^2$. Figure 3.3 shows that the case of MA(1) regression errors with $\theta = 2$ and $\sigma^2 = 0.005899^2$, and Figure 3.4 shows $\theta = -2$ and $\sigma^2 = 0.0177^2$. In the above cases, $\sigma^2$ are taken to make $h_M$ roughly 1/2. Vertical lines are the minimizers of the curves. The notation * denotes the intersection point where the vertical line and $CV_\ell(h)$ curve meets.

If the regression errors are a MA(q) process with an unknown order q, then the dependence effect is totally avoided whenever $\ell > q$, i.e. $\sum_{k>\ell} \gamma(k) = 0$. In practice, the MA(q) regression errors are detectable by looking at the sample autocorrelation function of $Y_j - Y_{j-1}$ which should vanish at lags greater than q+1. If the regression errors are an AR(1) process with parameter $\phi$, with $|\phi| < 1$, and finite variance $\sigma^2$, then

$$\left|\sum_{k>\ell} \gamma(k)\right| \le \sum_{k>\ell} |\gamma(k)| \le \sigma^2(1-\phi^2)^{-1}(1-|\phi|)^{-1}|\phi|^{\ell+1}.$$

We solve the inequality $|\phi|^\ell \le n^{-1/5}$ to get

$$\ell \ge (-1/5)(\log n)/(\log |\phi|).$$

If $|\phi|$ is not larger than 0.9, then taking $\ell = 4.5(\log n)$ is enough to make the quantity $\sum_{k>\ell} \gamma(k)$ be negligible. In practice, the AR(1) regression errors are also detectable by looking at the sample autocorrelation function of $Y_j - Y_{j-1}$ which would act as the usual AR(1) process at lags greater than or equal to 1.

For the general ARMA regression errors $\epsilon_j$, using the geometric boundedness of the autocovariance function $\gamma(\cdot)$ of $\epsilon_j$ (the property (d) as given in Section 3.2), we have

$$\left| \sum_{k>\ell} \gamma(k) \right| \le C \cdot s^{\ell+1}/(1-s).$$

This implies that $(C^S_{0\ell}/C_0)$ converges to 1 at a polynomial rate as $\ell \longrightarrow \infty$. Therefore, taking $\ell = O(\log n)$, the modified cross-validation criterion $CV_\ell(h)$ would provide asymptotically optimal bandwidths when the observations suffer from a short range dependence.

Since the optimal bandwidths $h_A$ and $h_M$ are of the same order $n^{-1/5}$, then the value of $\ell$ is allowed to be $o(n^{4/5})$ around the optimal bandwidths. This leaves a lot of room for the modified cross-validation criterion to handle the more complicated and the longer range dependence structure than the ARMA processes used in this chapter. See Granger and Joyeux (1980) and Cox (1984) for a detailed discussion of long range dependence processes.

The results above may be extended to the case where the regression errors $\epsilon_j$ are a linear process instead of only an ARMA process, i.e. $\epsilon_j$ are defined by

$$\epsilon_j = \sum_{i=-\infty}^{\infty} \psi_i e_{j-i},$$

for every j. The linear process $\epsilon_j$ still has the same results obtained in this chapter when the following assumptions hold:

a. The IID random variables $e_j$ have mean zero and all finite moments for every j.

b. The coefficients $\psi_i$ are real numbers such that

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty,$$

$$\sum_{k=-\infty}^{\infty} |k \cdot \gamma(k)| < \infty,$$

where $\gamma(k) = E(e_1^{\,2}) \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k}$ for all integers k.

c. When $\ell = o(n^{4/5})$, $|\sum_{k>\ell} \gamma(k)| = O(n^{-r})$, for some $r > 0$.

### 3.6 Proofs

Let c, $c_1$, and $c_2$ denote generic constants. Before we start the proofs, we need to introduce some properties which will be used in this section.

Theorem A (Section 1.4 of Serfling (1980)). Suppose that, as $n \longrightarrow \infty$, $X_n \longrightarrow X$ in distribution and the sequence $\{X_n{}^r\}$ is uniformly integrable, where $r > 0$. Then $E(|X|^r) < \infty$, $\lim_{n\to\infty} E(X_n{}^r) = E(X^r)$, and $\lim_{n\to\infty} E(|X_n|^r) = E(|X|^r)$.

Whittle's inequality (Theorem 2 of Whittle (1960)). Let

$$A = \sum_{i=1}^{n} a_i X_i,$$

$$B = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} X_i X_j,$$

and

$$\gamma_i(k) = [E(|X_i|^k)]^{1/k},$$

where $X_1, \ldots, X_n$ are independent random variables with mean zero, $a_i$, $b_{ij}$ are real numbers for $i, j = 1, 2, \ldots, n$. Then the following inequalities are valid

(i) $$E(|A|^k) \le 2^k C(k) (\sum_{i=1}^{n} a_i{}^2 \gamma_i(k)^2)^{k/2},$$

(ii) $$E(|B - E(B)|^k) \le 2^{3k} C(k) C(2k)^{1/2} (\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}{}^2 \gamma_i(2k)^2 \gamma_j(2k)^2)^{k/2},$$

where $C(k) = 2^{k/2} \Gamma(\frac{k+1}{2}) \pi^{-1/2}$, provided $k \ge 2$ and the right-hand members exist.

We first extend Whittle's inequality to infinite series. If $X_i$ are IID random variables with mean zero and all finite moments, and $a_i$

and $b_{ij}$ are real numbers such that $\sum_{i=-\infty}^{\infty} |a_i| < \infty$ and $\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |b_{ij}| < \infty$, then for all positive integers k, letting $A = \sum_{i=-\infty}^{\infty} a_i X_i$ and $B = \sum_{i=-\infty, i\neq j, j=-\infty}^{\infty} \sum^{\infty} b_{ij} X_i X_j$, then we have

$$E(A^{2k}) \le c_1 (\sum_{i=-\infty}^{\infty} a_i^2)^k, \tag{3.6.1}$$

$$E(B^{2k}) \le c_2 (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_{ij}^2)^k, \tag{3.6.2}$$

where $c_1$ and $c_2$ are constants involving k and moments of X.

Proofs of (3.6.1) and (3.6.2):

Because $\sum_{i=-\infty}^{\infty} |a_i|$ and $\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |b_{ij}|$ absolutely converge, we have, as $n \longrightarrow \infty$,

a. $$A_n = \sum_{i=-n}^{n} a_i X_i \longrightarrow A \text{ in probability,}$$

b. $$B_n = \sum_{i=-n, i\neq j, j=-n}^{n} \sum^{n} b_{ij} X_i X_j \longrightarrow B \text{ in probability.}$$

We first check that, for each k = 1, 2, ...,

1. $E(A_n^{2k}) < \infty$, $E(B_n^{2k}) < \infty$, for all n = 1, 2, ...,
2. $\{A_n^{2k}\}$ and $\{B_n^{2k}\}$ are uniformly integrable.

For the first one, using Whittle's inequality, we have

$$E(A_n^{2k}) \le c_1 (\sum_{i=-n}^{n} a_i^2)^k \le c_1 (\sum_{i=-\infty}^{\infty} a_i^2)^k \le c_1 (\sum_{i=-\infty}^{\infty} |a_i|)^{2k} < \infty,$$

$$E(B_n^{2k}) \le c_2 (\sum_{i=-n}^{n} \sum_{j=-n}^{n} b_{ij}^2)^k \le c_2 (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_{ij}^2)^k$$

$$\le c_2 (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |b_{ij}|)^{2k} < \infty.$$

For the second one, we have

$$\lim_{c\to\infty} \sup_n E(A_n^{2k} \cdot I_{[A_n^{2k} > c]}) \le \lim_{c\to\infty} \sup_n c^{-1} E(A_n^{4k}) = 0,$$

$$\lim_{c\to\infty} \sup_n E(B_n^{2k} \cdot I_{[B_n^{2k} > c]}) \le \lim_{c\to\infty} \sup_n c^{-1} E(B_n^{4k}) = 0,$$

since

$$E(A_n^{4k}) \le c_1 (\sum_{i=-\infty}^{\infty} |a_i|)^{4k} < \infty,$$

$$E(B_n^{4k}) \le c_2 (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |b_{ij}|)^{4k} < \infty.$$

Then, using the Theorem A in Section 1.4 of Serfling (1980) as described above, we have

$$E(A_n^{2k}) \longrightarrow E(A^{2k}) \le c_1 (\sum_{i=-\infty}^{\infty} a_i^2)^k,$$

$$E(B_n^{2k}) \longrightarrow E(B^{2k}) \le c_2 (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_{ij}^2)^k.$$

The proofs of (3.6.1) and (3.6.2) are complete.

If $\epsilon_j$ are a linear process defined by

$$\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i},$$

for $j = 1, 2, \ldots, n$, where $\psi_i$ are real numbers with $\sum_{i=0}^{\infty} |\psi_i| < \infty$, and $e_j$ are IID random variables with mean zero and all finite moments, then

(3.6.3) $$E(\sum_{j=1}^{n} \epsilon_j)^{2k} = O(n^k),$$

for all positive integers k. The equation (3.6.3) implies, $0 < x < 1$,

(3.6.4) $$n^{-1} \sum_{j=1}^{n} \epsilon_j = o_u(n^{-2/5}),$$

(3.6.5) $$n^{-1} \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i = o_u(n^{-2/5}h^{-2/5}).$$

Proofs of (3.6.3) through (3.6.5):

For each finite n, let

$$S = \sum_{j=1}^{n} \epsilon_j.$$

By Fubini's Theorem, absolute convergence of $\sum_{i=0}^{\infty} |\psi_i|$, and finite moments of $e_j$, we have

$$S = \sum_{j=1}^{n} \epsilon_j = \sum_{j=1}^{n} \sum_{i=0}^{\infty} \psi_i e_{j-i} = \sum_{i=0}^{\infty} \psi_i (\sum_{j=1}^{n} e_{j-i}) = \sum_{i=0}^{\infty} \psi_i X_i,$$

where, for each $i = 0, 1, \ldots,$

$$X_i = \sum_{j=1}^{n} e_{j-i},$$

and where $X_i$ satisfies, for all positive integers k,

$$E(X_i^{2k}) = E(X_1^{2k}) \le c_1 n^k,$$

by Whittle's inequality. Let, for each $\ell = 1, 2, \ldots,$

$$S_\ell = \sum_{i=0}^{\ell} \psi_i X_i.$$

Applying the absolute convergence of $\sum_{i=0}^{\infty} |\psi_i|$, we have, as $\ell \longrightarrow \infty$,

$$E(|S_\ell - S|) = E(|\sum_{i>\ell} \psi_i X_i|) \le \sum_{i>\ell} |\psi_i| E(|X_i|) = E(|X_1|) \sum_{i>\ell} |\psi_i| \longrightarrow 0,$$

i.e.

$$S_\ell \longrightarrow S \text{ in probability.}$$

Applying Minkowski's inequality and Whittle's inequality, for each finite $\ell$ and all positive integers k, we have

$$E(S_\ell^{2k}) \le E(X_1^{2k})(\sum_{i=0}^{\ell} |\psi_i|)^{2k} \le c_1 n^k (\sum_{i=0}^{\infty} |\psi_i|)^{2k} < \infty.$$

Using the finite moments of $S_\ell$, we have

$$\lim_{c\to\infty} \sup_{\ell} E(S_\ell^{2k} \cdot I_{[S_\ell^{2k} > c]}) \le \lim_{c\to\infty} \sup_{\ell} c^{-1} E(S_\ell^{4k}) = 0,$$

i.e.

$$\{S_\ell^{2k}\} \text{ is uniformly integrable.}$$

Then, using the Theorem A in Section 1.4 of Serfling (1980) as described above, we have, as $\ell \longrightarrow \infty$,

$$E(S_\ell^{2k}) \longrightarrow E(S^{2k}).$$

$$E(S^{2k}) \le c_1 n^k \left( \sum_{i=0}^{\infty} |\psi_i| \right)^{2k} = O(n^k).$$

The proof of (3.6.3) is complete.

For (3.6.4), use the Borel-Cantelli lemma, for any $\eta > 0$,

$$\sum_n P\left( \left| n^{2/5} \left( n^{-1} \sum_{j=1}^{n} \epsilon_j \right) \right| > \eta \right)$$
$$\le \sum_n \eta^{-2k} E\left( \left( n^{-3/5} \sum_{j=1}^{n} \epsilon_j \right)^{3k} \right)$$
$$\le \sum_n \eta^{-2k} c_1 n^{-k/5} < \infty,$$

if $k > 5$. The proof of (3.6.4) is complete.

For (3.6.5), by the definition of $K$ and $K_h(\cdot) = h^{-1}K(\cdot/h)$, there are $O(nh)$ terms in $n^{-1} \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i$. For $h \in H_n$, we have

$$C^{-1} n^{-1+\delta} \le n^{-1}h^{-1} \le Cn^{-\delta}.$$

Use the Borel-Cantelli lemma, for any $\eta > 0$,

$$\sum_n P\left( \left| (nh)^{2/5} \left( n^{-1} \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i \right) \right| > \eta \right)$$
$$\le \sum_n \eta^{-2k} c_1 (nh)^{-k/5}$$
$$\le \sum_n \eta^{-2k} c_1 n^{-\delta k/5} < \infty,$$

if $k > 5/\delta$. The proof of (3.6.5) is complete.

For any $\ell \ll nh$ and each $x_j$ with $W(x_j) \ne 0$, or $h < x_j < 1-h$, under the assumptions given in Section 3.3, we have the following asymptotic results:

(3.6.6) $$n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i) = 1 + O(n^{-1}h^{-1}),$$

(3.6.7) $$(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i) = 1 + O(\ell n^{-1}h^{-1}),$$

(3.6.8) $$b_j = \left[ n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)(m(x_i)-m(x_j)) \right] / \left[ n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i) \right]$$
$$= (1/2)h^2 m''(x_j) \int u^2 K + o(h^2).$$

$$(3.6.9)\quad b_{\ell j} = [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)(m(x_i)-m(x_j))] \Big/ [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)]$$

$$= b_j + o(h^2),$$

$$(3.6.10)\quad v_j = [n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i] \Big/ [n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)]$$

$$= n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i + o_u((nh)^{-7/5})$$

$$= o_u((nh)^{-2/5}),$$

$$(3.6.11)\quad v_{\ell j} = [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)\epsilon_i] \Big/ [(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)]$$

$$= v_j + o_u((nh)^{-4/5} + \ell^{3/5}(nh)^{-1}).$$

Proof of (3.6.9):

Here we only need to check

$$n^{-1} \sum_{i:|i-j|\le\ell} K_h(x_j-x_i)(m(x_i) - m(x_j)) = o(h^2).$$

Using the symmetry of K, the Hoelder continuity of K and m, and $\ell << nh$, we have

$$n^{-1} \sum_{i:|i-j|\le\ell} K_h(x_j-x_i)(m(x_i) - m(x_j))$$

$$= \int_{(j-\ell)/n}^{(j+\ell)/n} K_h(x_j-t)(m(t) - m(x_j))\, dt + O(\ell n^{-2}h^{-1})$$

$$= \int_{-\ell/nh}^{\ell/nh} K(u)(m(x_j - hu) - m(x_j))\, du + O(\ell n^{-2}h^{-1})$$

$$= (1/2)h^2(1+o(1))m''(x_j)\int_{-\ell/nh}^{\ell/nh} u^2K(u)\, du + O(\ell n^{-2}h^{-1})$$

$$= O(\ell n^{-1}h + \ell n^{-2}h^{-1}) = o(h^2).$$

The proof of (3.6.9) is complete.

Proof of (3.3.1):

Since $\hat{m}(x_j) - m(x_j) = b_j + v_j$, then $d_M(h)$ can be expressed as

$$d_M(h) = E(n^{-1}\sum_{j=1}^{n} (\hat{m}(x_j) - m(x_j))^2 W(x_j))$$
$$= E(n^{-1}\sum_{j=1}^{n} (b_j+v_j)^2 W(x_j))$$
$$= n^{-1}\sum_{j=1}^{n} b_j{}^2 W(x_j) + n^{-1}\sum_{j=1}^{n} E(v_j{}^2)W(x_j).$$

Using $b_j = (1/2)h^2 m''(x_j)\int u^2K + o(h^2)$ as given in (3.6.8) and the uniform continuity of m'', we have

(3.6.12)
$$n^{-1}\sum_{j=1}^{n} b_j{}^2 W(x_j)$$
$$= n^{-1}\sum_{j=1}^{n} [(1/2)h^2 m''(x_j)\int u^2K + o(h^2)]^2 W(x_j)$$
$$= (1/4)h^4(\int u^2K)^2 n^{-1}\sum_{j=1}^{n} (m''(x_j))^2 W(x_j) + o(h^4)$$
$$= (1/4)h^4(\int u^2K)^2\int (m'')^2 W + o(h^4).$$

Using $n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i) = 1 + O(n^{-1}h^{-1})$ as given in (3.6.6), we have

$$E(v_j{}^2) = n^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)\gamma(s-t)(1+O(n^{-1}h^{-1})).$$

Then, letting k=s-t and using $\sum_{k=-\infty}^{\infty} |k\gamma(k)| < \infty$, $K_h(x_j-x_s+x_k) = K_h(x_j-x_s) + O(|k|n^{-1}h^{-2})$, and O(nh) terms in $\sum_s$, we have

(3.6.13)
$$n^{-1}\sum_{j=1}^{n} E(v_j{}^2)W(x_j)$$
$$= n^{-3}\sum_{j=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)\gamma(s-t)W(x_j)(1+O(n^{-1}h^{-1}))$$
$$= n^{-3}\sum_{j=1}^{n}\sum_{|k|<n}\sum_{s} K_h(x_j-x_s)K_h(x_j-x_s+x_k)\gamma(k)W(x_j)(1+O(n^{-1}h^{-1}))$$
$$= n^{-1}\sum_{|k|<n} [n^{-1}\sum_{s} K_h(x_j-x_s)K_h(x_j-x_s+x_k)]\gamma(k)[n^{-1}\sum_{j=1}^{n} W(x_j)] (1+O(n^{-1}h^{-1}))$$
$$= n^{-1}\sum_{|k|<n} [n^{-1}\sum_{s} K_h(x_j-x_s)^2 + O(|k|n^{-1}h^{-2})]\gamma(k)[n^{-1}\sum_{j=1}^{n} W(x_j)] (1+O(n^{-1}h^{-1}))$$

$$= n^{-1} \sum_{|k|<n} [h^{-1}\int K^2 + O(|k|n^{-1}h^{-2})]\gamma(k)[\int W + O(n^{-1})](1+O(n^{-1}h^{-1}))$$
$$= n^{-1}h^{-1}\int K^2 \int W[\sum_{|k|<n} \gamma(k)] + O(n^{-2}h^{-2}).$$

The proof of (3.3.1) is complete.

Proof of (3.3.2):

Using (3.3.1), the proof of (3.3.2) is complete if we show that, as $n \longrightarrow \infty$,

(3.6.14) $$\sup_{h\in H_n} |(d_A(h) - d_M(h)) / d_M^T(h)| \longrightarrow 0 \quad \text{a.s.}$$

where

$$d_A(h) - d_M(h) = 2n^{-1} \sum_{j=1}^{n} b_j v_j W(x_j) + n^{-1} \sum_{j=1}^{n} (v_j{}^2 - E(v_j{}^2))W(x_j).$$

Using the Hoelder continuity of K and m, the compactness of the support of K, the boundedness of W, and $b_j(h) = O(h^2)$ as given in (3.6.8), we have the following expressions: for any $h$, $h_1 \in H_n$, with $h \leq h_1$,

$$v_j(h) = n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i(1+O(n^{-1}h^{-1})) = h^{-1}\cdot o_u(1),$$

$$v_j(h) - v_j(h_1) = n^{-1} \sum_{i=1}^{n} (K_h-K_{h_1})(x_j-x_i)\epsilon_i + o_u((nh)^{-7/5}+(nh_1)^{-7/5})$$
$$= (h^{-1} - h_1{}^{-1})\cdot o_p(1),$$

$$[v_j(h)^2 - E(v_j(h)^2)] - [v_j(h_1)^2 - E(v_j(h_1)^2)]$$
$$= n^{-2} \sum_{s=1}^{n} \sum_{t=1}^{n} [K_h(x_j-x_s)K_h(x_j-x_t) - K_{h_1}(x_j-x_s)K_{h_1}(x_j-x_t)]$$
$$(\epsilon_s\epsilon_t - \gamma(s-t)) + o_u((nh)^{-9/5})$$
$$= (h^{-2} - h_1{}^{-2})\cdot o_u(1).$$

Then, we have

$$n^{-1} \sum_{j=1}^{n} [b_j(h)v_j(h) - b_j(h_1)v_j(h_1)]W(x_j)$$

$$= n^{-1} \sum_{j=1}^{n} [b_j(h)(v_j(h) - v_j(h_1)) + (b_j(h) - b_j(h_1))v_j(h_1)]W(x_j)$$
$$= [h^2(h^{-1} - h_1^{-1}) + (h^2 - h_1^2)h_1^{-1}]\cdot o_u(1)$$
$$= h_1 |(h-h_1)/h| \cdot o_u(1).$$

and

$$n^{-1} \sum_{j=1}^{n} [(v_j(h)^2 - E(v_j(h)^2)) - (v_j(h_1)^2 - E(v_j(h_1)^2))]W(x_j)$$
$$= (h^{-2} - h_1^{-2})\cdot o_u(1)$$
$$= h^{-2}|(h-h_1)/h|\cdot o_u(1).$$

For any constant $r \geq 3$, $h$, $h_1 \in H_n = [C^{-1}n^{-1+\delta}, Cn^{-\delta}]$, with $h \leq h_1$, we have

$$\sup_{|(h-h_1)/h|\leq n^{-r}} |(d_A(h) - d_M(h)) - (d_A(h_1) - d_M(h_1))|$$
$$\leq \sup_{|(h-h_1)/h|\leq n^{-r}} |2n^{-1} \sum_{j=1}^{n} [b_j(h)v_j(h) - b_j(h_1)v_j(h_1)]W(x_j)| +$$
$$\sup_{|(h-h_1)/h|\leq n^{-r}} |n^{-1} \sum_{j=1}^{n} [(v_j(h)^2 - E(v_j(h)^2)) - (v_j(h_1)^2 - E(v_j(h_1)^2))]W(x_j)|$$
$$= \sup_{|(h-h_1)/h|\leq n^{-r}} (h_1 + h^{-2})|(h-h_1)/h|\cdot o_u(1)$$
$$= n^2 n^{-r}\cdot o_u(1) = o_u(n^{-1}).$$

Hence, it is sufficient to restrict the supremum in (3.6.14) to a set $H_n'$ which is a subset of $H_n$ so that $\#(H_n') \leq n^{r+1}$ and so that for any $h \in H_n$ there is an $h_1 \in H_n'$ with $|(h-h_1)/h| \leq n^{-r}$. Then we have

$$\sup_{h\in H_n} |d_A(h) - d_M(h)| \leq \sup_{h_1\in H_n'} |d_A(h_1) - d_M(h_1)| + \sup_{|(h-h_1)/h|\leq n^{-r}} |(d_A(h)-d_M(h)) - (d_A(h_1)-d_M(h_1))|$$
$$\leq \sup_{h_1\in H_n'} |d_A(h_1) - d_M(h_1)| + o_u(n^{-1}).$$

To verify (3.6.14), using the fact that $d_M^T(h) = O(n^{-1}h^{-1}+h^4)$ and the above Hoelder-continuity considerations, it is enough to show

$$(3.6.15)\qquad \sup_{h\in H_n'} \left|n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) \;/\; d_M^T(h)\right| \longrightarrow 0 \quad \text{a.s.}$$

$$(3.6.16)\qquad \sup_{h\in H_n'} \left|n^{-1}\sum_{j=1}^{n} (v_j^2 - E(v_j^2))W(x_j) \;/\; d_M^T(h)\right| \longrightarrow 0 \quad \text{a.s.}$$

To verify (3.6.15), given $\eta > 0$, for any $k = 1, 2, \ldots,$ we have

$$P\left(\sup_{h\in H_n'} \left|n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) \;/\; d_M^T(h)\right| > \eta\right)$$

$$\le \eta^{-2k}\cdot\#(H_n')\cdot \sup_{h\in H_n'} E\left(\left(n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) \;/\; d_M^T(h)\right)^{2k}\right).$$

The proof of (3.6.15) is complete when it is seen that there is a constant $\tau > 0$, so that for $k = 1, 2, \ldots,$ there are constants $c_k$ so that

$$(3.6.17)\qquad \sup_{h\in H_n} E\left(\left(n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) \;/\; d_M^T(h)\right)^{2k}\right) \le c_k n^{-\tau k}.$$

Using the Borel-Cantelli lemma, there is a sufficiently large k to make $(r+1-\tau k) < -1$, then for any given $\eta > 0$, we have

$$\sum_{n=1}^{\infty} P\left(\sup_{h\in H_n'} \left|n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) \;/\; d_M^T(h)\right| > \eta\right) \le c\sum_{n=1}^{\infty} n^{r+1-\tau k} < \infty.$$

Similarly, the proof of (3.6.16) is complete by showing that

$$(3.6.18)\qquad \sup_{h\in H_n} E\left(\left(n^{-1}\sum_{j=1}^{n} (v_j^2 - E(v_j^2))W(x_j) \;/\; d_M^T(h)\right)^{2k}\right) \le c_k n^{-\tau k}.$$

To check (3.6.17), using $n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i) = O(1)$ as given in (3.6.6), $b_j = O(h^2)$ as given in (3.6.8), and the boundedness of W, we have

$$n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) = n^{-2}\sum_{i=1}^{n}\left[\sum_{j=1}^{n} b_j K_h(x_j-x_i)W(x_j)(1+O(n^{-1}h^{-1}))\right]\epsilon_i$$

$$= O(n^{-1}h^2)\sum_{i=1}^{n} \epsilon_i.$$

Using $E((\sum_{i=1}^{n} \epsilon_i)^{2k}) = O(n^k)$ as given in (3.6.3), and $C^{-1}n^{-1+\delta} < h <$ $Cn^{-\delta}$, we have

$$E((n^{-1}h^2 \sum_{i=1}^{n} \epsilon_i \;/\; d_M^T(h))^{2k}) = O(h^k) \leq c_k n^{-\delta k}.$$

The proof of (3.6.17) is complete.

To check (3.6.18), express $\epsilon_s$, $\epsilon_t$, and $\gamma(s-t)$ by

$$\epsilon_s = \sum_{p=0}^{\infty} \psi_p e_{s-p},$$

$$\epsilon_t = \sum_{q=0}^{\infty} \psi_q e_{t-q},$$

$$\gamma(s-t) = \mu_2 \sum_{p=0}^{\infty} \psi_p \psi_{p-s+t},$$

for each s, t = 1, 2, ..., n. Using

$$n^{-1} \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i = o_u((nh)^{-2/5}) \text{ as given in (3.6.5)},$$

$$v_j = n^{-1} \sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i + o_u((nh)^{-7/5}) \text{ as given in (3.6.10)},$$

$$n^{-1} \sum_{j=1}^{n} E(v_j^2)W(x_j)$$

$$= n^{-3} \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)\gamma(s-t) + O(n^{-2}h^{-2})$$

as given in (3.6.13), then we have

$$n^{-1} \sum_{j=1}^{n} (v_j^2 - E(v_j^2))W(x_j)$$

$$= n^{-3} \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)(\epsilon_s\epsilon_t - \gamma(s-t))W(x_j) + o_u((nh)^{-9/5})$$

$$= A + B + o_u((nh)^{-9/5}),$$

where

$$A = n^{-3} \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t) \sum_{p=0}^{\infty} \psi_p\psi_{p-s+t}(e_{s-p}^2 - \mu_2)W(x_j),$$

$$B = n^{-3} \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)(\sum_{q\neq p-s+t} \sum \psi_p\psi_q e_{s-p}e_{t-q})W(x_j).$$

Then the inequality (3.6.18) is complete by showing that

(3.6.19) $$E(A^{2k}) = O(n^{-3k}h^{-2k}),$$

(3.6.20) $$E(B^{2k}) = O(n^{-2k}h^{-k}).$$

To check (3.6.19), letting $u = s-p$ and $\xi_u = e_{s-p}^2 - \mu_2$, express A as

$$A = \sum_{u=-\infty}^{\infty} a_u \xi_u,$$

where

$$a_u = n^{-3} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j)\psi_{s-u}\psi_{-u+t},$$

and $\xi_u$ are IID random variables with mean zero and all finite moments for all u. Using the extended Whittle's inequality as given in (3.6.1), we need to check $\sum_{u=-\infty}^{\infty} |a_u| < \infty$ first, then we have

$$E(A^{2k}) \leq c(\sum_{u=-\infty}^{\infty} a_u^2)^k.$$

Because of $n^{-1} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j) = O(h^{-1})$ for any s, t, $\sum_{u=-\infty}^{\infty} |\psi_{s-u}\psi_{-u+t}| \leq (\sum_{i=0}^{\infty} |\psi_i|)^2 < \infty$, and $n \cdot O(nh)$ terms in $\sum_{s=1}^{n} \sum_{t=1}^{n}$, then we have

$$\begin{aligned}\sum_{u=-\infty}^{\infty} |a_u| &\leq \sum_{u=-\infty}^{\infty} n^{-3} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j)|\psi_{s-u}\psi_{-u+t}| \\ &\leq c \sum_{u=-\infty}^{\infty} n^{-2}h^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} |\psi_{s-u}\psi_{-u+t}| \\ &\leq c \sum_{s=1}^{n} \sum_{t=1}^{n} n^{-2}h^{-1} < \infty.\end{aligned}$$

Now we calculate $\sum_{u=-\infty}^{\infty} a_u^2$. Because of $n^{-1} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j) = O(h^{-1})$ for any s, t, $\sum_{s=1}^{n} \psi_{s-u} = O(1)$, $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{-u+t}\psi_{-u+t'} = \gamma(t - t')$, $\sum_{t=1}^{n} \sum_{t'=1}^{n} \gamma(t-t') = O(n)$, then we have

$$\sum_{u=-\infty}^{\infty} a_u^2 = \sum_{u=-\infty}^{\infty} [n^{-3} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j)\psi_{s-u}\psi_{-u+t}]^2$$
$$\leq cn^{-4}h^{-2} \sum_{u=-\infty}^{\infty} [\sum_{s=1}^{n} \sum_{t=1}^{n} \psi_{s-u}\psi_{-u+t}]^2$$
$$\leq cn^{-4}h^{-2} \sum_{u=-\infty}^{\infty} \sum_{t=1}^{n} \sum_{t'=1}^{n} \psi_{-u+t}\psi_{-u+t'}$$
$$\leq cn^{-4}h^{-2} \sum_{t=1}^{n} \sum_{t'=1}^{n} \gamma(t-t') = O(n^{-3}h^{-2}).$$

The proof of (3.6.19) is complete.

To check (3.6.20), letting $u = s-p$, $v = t-q$, express B as

$$B = \sum_{u=-\infty, u\neq v, v=-\infty}^{\infty} \sum^{\infty} b_{uv}e_u e_v,$$

where

$$b_{uv} = n^{-3} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j)\psi_{s-u}\psi_{t-v}.$$

Using the extended Whittle's inequality as given in (3.6.2), we need to check $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}| < \infty$ first, then we have

$$E(B^{2k}) \leq c( \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2)^k.$$

Because of $n^{-1} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j) = O(h^{-1})$ for any s, t, $\sum_{u=-\infty}^{\infty} |\psi_{s-u}| < \infty$, $\sum_{v=-\infty}^{\infty} |\psi_{t-v}| < \infty$, and $n\cdot O(nh)$ terms in $\sum_{s=1}^{n} \sum_{t=1}^{n}$, then we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}|$$
$$\leq \sum_u \sum_v n^{-3} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j) |\psi_{s-u}\psi_{t-v}|$$
$$\leq c\sum_u \sum_v n^{-2}h^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} |\psi_{s-u}\psi_{t-v}|$$
$$\leq c \sum_{s=1}^{n} \sum_{t=1}^{n} n^{-2}h^{-1} < \infty.$$

Now we calculate $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2$. Because of, for any s, t,

$n^{-1}\sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j) = O(h^{-1})$, $\mu_2\sum_{u=-\infty}^{\infty} \psi_{s-u}\psi_{s'-u} = \gamma(s - s')$,

$\mu_2\sum_{v=-\infty}^{\infty} \psi_{t-v}\psi_{t'-v} = \gamma(t - t')$, $\sum_{s'=1}^{n} \gamma(s-s') = O(1)$, $\sum_{t'=1}^{n} \gamma(t-t') = O(1)$,

and $n\cdot O(nh)$ terms in $\sum_{s=1}^{n}\sum_{t=1}^{n}$, then we have

$$\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} b_{uv}^2 = \sum_u\sum_v [n^{-3}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{j=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)W(x_j)\psi_{s-u}\psi_{t-v}]^2$$

$$\le cn^{-4}h^{-2}\sum_u\sum_v [\sum_{s=1}^{n}\sum_{t=1}^{n} \psi_{s-u}\psi_{t-v}]^2$$

$$\le cn^{-4}h^{-2}\sum_u\sum_v\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{s'=1}^{n}\sum_{t'=1}^{n} \psi_{s-u}\psi_{t-v}\psi_{s'-u}\psi_{t'-v}$$

$$\le cn^{-4}h^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{s'=1}^{n}\sum_{t'=1}^{n} \gamma(s-s')\gamma(t-t')$$

$$\le c\sum_{s=1}^{n}\sum_{t=1}^{n} n^{-4}h^{-2} = O(n^{-2}h^{-1}).$$

The proof of (3.6.20) is complete, i.e. the proof of (3.3.2) is complete.

Proof of (3.4.2):

Since $\hat{m}(x_j)-m(x_j) = b_j + v_j$ and $\hat{m}_j(x_j)-m(x_j) = b_{\ell j} + v_{\ell j}$, then $\mathrm{Remainder}_\ell(h)$ can be expressed as

(3.6.21) $\quad \mathrm{Remainder}_\ell(h)$

$$= n^{-1}\sum_{j=1}^{n} [\hat{m}_j(x_j)-\hat{m}(x_j)][\hat{m}_j(x_j)+\hat{m}(x_j)-2m(x_j)]W(x_j)$$

$$= n^{-1}\sum_{j=1}^{n} [(b_{\ell j}-b_j) + (v_{\ell j}-v_j)][(b_{\ell j}+b_j) + (v_{\ell j}+v_j)]W(x_j)$$

$$= A + B + C,$$

where

$$A = n^{-1}\sum_{j=1}^{n} (b_{\ell j}-b_j)(b_{\ell j}+b_j)W(x_j),$$

$$B = n^{-1}\sum_{j=1}^{n} [(b_{\ell j}-b_j)(v_{\ell j}+v_j) + (v_{\ell j}-v_j)(b_{\ell j}+b_j)]W(x_j)$$

$$= 2n^{-1} \sum_{j=1}^{n} (v_{\ell j} b_{\ell j} - v_j b_j) W(x_j).$$

$$C = n^{-1} \sum_{j=1}^{n} (v_{\ell j} - v_j)(v_{\ell j} + v_j) W(x_j).$$

The proof of (3.4.2) is complete by showing that A, B, and C are all of the order $o_u(d_M^T(h))$ where $d_M^T(h) = a_1 n^{-1} h^{-1} + b_1 h^4$. To check $A = o_u(d_M^T(h))$, using $b_j = O(h^2)$ as given in (3.6.8) and $b_{\ell j} = b_j + o(h^2)$ as given in (3.6.9), we have

$$(b_{\ell j} - b_j)(b_{\ell j} + b_j) = b_{\ell j}^2 - b_j^2 = o(h^4).$$

Then, using the boundedness of W and $\ell = o(nh)$, we have, for any $h \in H_n$, as $n \longrightarrow \infty$,

$$A = o(h^4) = o(d_M^T(h)).$$

The proof of $A = o(d_M^T(h))$ is complete.

To check $B = o_u(d_M^T(h))$, following the same steps as in the proof of $n^{-1} \sum_{j=1}^{n} b_j v_j W(x_j) = o_u(d_M^T(h))$ as given in (3.6.15), then we have

$$n^{-1} \sum_{j=1}^{n} b_{\ell j} v_{\ell j} W(x_j)$$

$$= \sum_{i=1}^{n} [n^{-1}(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j - x_i) b_{\ell j} W(x_j)(1+O(\ell n^{-1} h^{-1}))]\epsilon_i$$

$$= O(n^{-1}h^2) \sum_{i=1}^{n} \epsilon_i = o_u(d_M^T(h)).$$

Thus, $n^{-1} \sum_{j=1}^{n} b_j v_j W(x_j)$ and $n^{-1} \sum_{j=1}^{n} b_{\ell j} v_{\ell j} W(x_j)$ are $o_u(d_M^T(h))$.

The proof of $B = o_u(d_M^T(h))$ is complete.

To check $C = o_u(d_M^T(h))$, express C as

$$C = n^{-1} \sum_{j=1}^{n} (v_{\ell j} - v_j)(v_{\ell j} + v_j) W(x_j)$$

$$= n^{-1} \sum_{j=1}^{n} (v_{\ell j}^2 - v_j^2) W(x_j)$$

$$= n^{-1} \sum_{j=1}^{n} (v_{\ell j}^2 - E(v_{\ell j}^2)) W(x_j) - n^{-1} \sum_{j=1}^{n} (v_j^2 - E(v_j^2)) W(x_j) +$$

$$n^{-1}\sum_{j=1}^{n} (E(v_{\ell j}^2) - E(v_j^2))W(x_j).$$

Following the proof of $n^{-1}\sum_{j=1}^{n} (v_j^2 - E(v_j^2))W(x_j) = o_u(d_M^T(h))$ as given in (3.6.16), then we have

$$n^{-1}\sum_{j=1}^{n} (v_{\ell j}^2 - E(v_{\ell j}^2))W(x_j) = o_u(d_M^T(h)).$$

Thus we have $C = o_u(d_M^T(h))$ by checking that

$$n^{-1}\sum_{j=1}^{n} (E(v_{\ell j}^2) - E(v_j^2))W(x_j) = o(d_M^T(h)).$$

Using, for any j, $K_h(x_j-x_s)K_h(x_j-x_s) = O(h^{-2})$, $\sum_{s=j-\ell}^{j+\ell} \sum_t \gamma(s-t) = O(\ell)$, $\sum_s \sum_{t=j-\ell}^{j+\ell} \gamma(s-t) = O(\ell)$, and the boundedness of W, then we have

$$n^{-1}\sum_{j=1}^{n} |E(v_j^2) - E(v_{\ell j}^2)|W(x_j)$$

$$\le n^{-3}\sum_{j=1}^{n}\sum_{s=j-\ell}^{j+\ell}\sum_t |K_h(x_j-x_s)K_h(x_j-x_t)\gamma(s-t)W(x_j)|(1+O(\ell n^{-1}h^{-1})) +$$

$$n^{-3}\sum_{j=1}^{n}\sum_s\sum_{t=j-\ell}^{j+\ell} |K_h(x_j-x_s)K_h(x_j-x_t)\gamma(s-t)W(x_j)|(1+O(\ell n^{-1}h^{-1}))$$

$$= n^{-3}\sum_{j=1}^{n} O(\ell h^{-2}) = O(\ell n^{-2}h^{-2}) = o(d_M(h)) \text{ for } \ell = o(nh).$$

The proof of (3.4.2) is completed.

Proof of (3.4.3):

Using $\hat{m}_j(x_j)-m(x_j) = b_{\ell j} + v_{\ell j}$, we have

$$\text{Cross}_\ell(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j(b_{\ell j}+v_{\ell j})W(x_j).$$

$$E(\text{Cross}_\ell(h)) = E(n^{-1}\sum_{j=1}^{n} \epsilon_j v_{\ell j}W(x_j)).$$

Using $(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i) = 1 + O(\ell n^{-1}h^{-1})$ as given in (3.6.7), $\sum_{k=-\infty}^{\infty} |k\gamma(k)| < \infty$, and letting $k = j - i$, then we have

$$
\begin{aligned}
&E(\text{Cross}_\ell(h)) \\
&= n^{-1}\sum_{j=1}^{n} W(x_j)(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)\gamma(i-j)(1+O(\ell n^{-1}h^{-1})) \\
&= n^{-1}(n-2\ell-1)^{-1}\sum_{|k|>\ell}\sum_{i} W(\tfrac{i+k}{n})K_h(\tfrac{k}{n})\gamma(k)(1+O(\ell n^{-1}h^{-1})) \\
&= n^{-1}\sum_{|k|>\ell} K_h(\tfrac{k}{n})\gamma(k)[(n-2\ell-1)^{-1}\sum_{i} W(\tfrac{i+k}{n})](1+O(\ell n^{-1}h^{-1})) \\
&= n^{-1}\sum_{|k|>\ell} [h^{-1}K(0) + |k|n^{-1}h^{-2}]\gamma(k)[\int W + O(|k|n^{-1})](1+O(\ell n^{-1}h^{-1})) \\
&= 2n^{-1}h^{-1}[\sum_{k>\ell}\gamma(k)]K(0)\int W + O(\ell n^{-2}h^{-2}).
\end{aligned}
$$

The proof of (3.4.3) is complete.

Proof of (3.4.4):

Since $\text{Cross}_\ell(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j(b_{\ell j} + v_{\ell j})W(x_j)$, we have

$$
\begin{aligned}
&\text{Cross}_\ell(h) - E(\text{Cross}_\ell(h)) \\
&= n^{-1}\sum_{j=1}^{n} \epsilon_j b_{\ell j}W(x_j) + n^{-1}\sum_{j=1}^{n} (\epsilon_j v_{\ell j} - E(\epsilon_j v_{\ell j}))W(x_j).
\end{aligned}
$$

Using $\epsilon_i = \sum_{p=0}^{\infty} \psi_p e_{i-p}$, $\epsilon_j = \sum_{q=0}^{\infty} \psi_q e_{j-q}$, $\gamma(i-j) = \mu_2\sum_{p=0}^{\infty} \psi_p\psi_{p-i+j}$, $(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i) = 1 + O(\ell n^{-1}h^{-1})$ as given in (3.6.7), then $\text{Cross}_\ell(h) - E(\text{Cross}_\ell(h))$ can be expressed as $A + B_1 + B_2$, where

$$A = n^{-1}\sum_{j=1}^{n} \epsilon_j b_{\ell j}W(x_j),$$

$$
\begin{aligned}
B_1 = n^{-1}(n-2\ell-1)^{-1}\sum_{j=1}^{n}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)\cdot \\
(\sum_{p=0}^{\infty} \psi_p\psi_{p-i+j}(e_{i-p}^{2} - \mu_2))W(x_j)(1 + O(\ell n^{-1}h^{-1})),
\end{aligned}
$$

$$
\begin{aligned}
B_2 = n^{-1}(n-2\ell-1)^{-1}\sum_{j=1}^{n}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)\cdot \\
(\sum_{q\neq p-i+j}\sum \psi_p\psi_q e_{i-p}e_{j-q})W(x_j)(1 + O(\ell n^{-1}h^{-1})).
\end{aligned}
$$

Following the same arguments as the proof of (3.3.2), it is enough to show that, for each $k = 1, 2, \ldots$, (3.6.22) $$E(A^{2k}) = O(n^{-k}h^{4k}),$$

(3.6.23) $$E(B_1^{2k}) = O(n^{-3k}h^{-2k}),$$

(3.6.24) $$E(B_2^{2k}) = O(n^{-2k}h^{-k}).$$

To check (3.6.22), expressing $\epsilon_j = \sum_{q=0}^{\infty} \psi_q e_{j-q}$ and letting $u = j-q$, then we have

$$A = n^{-1}\sum_{j=1}^{n} \epsilon_j b_{\ell j} W(x_j) = \sum_{u=-\infty}^{\infty} a_u e_u,$$

where

$$a_u = n^{-1}\sum_{j=1}^{n} b_{\ell j} W(x_j)\psi_{j-u}.$$

Using the extended Whittle's inequality as given in (3.6.1), we need to check $\sum_{u=-\infty}^{\infty} |a_u| < \infty$ first, then we have

$$E(A^{2k}) \le c(\sum_{u=-\infty}^{\infty} a_u^2)^k.$$

Using $b_{\ell j} = O(h^2)$ as given in (3.6.9), $\sum_{u=-\infty}^{\infty} |\psi_{j-u}| < \infty$, and the boundedness of W, we have

$$\sum_{u=-\infty}^{\infty} |a_u| \le c \sum_{u=-\infty}^{\infty} n^{-1}\sum_{j=1}^{n} h^2|\psi_{j-u}| \le ch^2 < \infty.$$

Now we calculate $\sum_{u=-\infty}^{\infty} a_u^2$. Using the boundedness of W,

$\mu_2 \sum_{u=-\infty}^{\infty} \psi_{j-u}\psi_{j'-u} = \gamma(j-j')$, $\sum_{j=1}^{n}\sum_{j'=1}^{n} |\gamma(j-j')| = O(n)$, then we have

$$\sum_{u=-\infty}^{\infty} a_u^2 = \sum_{u=-\infty}^{\infty} (n^{-1}\sum_{j=1}^{n} b_{\ell j}W(x_j)\psi_{j-u})^2 \le cn^{-2}h^4 \sum_{u=-\infty}^{\infty}\sum_{j=1}^{n}\sum_{j'=1}^{n} \psi_{j-u}\psi_{j'-u}$$

$$\le cn^{-2}h^4 \sum_{j=1}^{n}\sum_{j'=1}^{n} \gamma(j-j') \le cn^{-1}h^4.$$

The proof of (3.6.22) is complete.

To check (3.6.23), letting $u = i-p$ and $\xi_u = e_{i-p}^2 - \mu_2$, we have

$$B_1 = n^{-1}(n-2\ell-1)^{-1}\sum_{j=1}^{n}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)\cdot$$

$$(\sum_{p=0}^{\infty} \psi_p \psi_{p-i+j}(e_{i-p}^2 - \mu_2))W(x_j)(1 + O(\ell n^{-1}h^{-1})).$$

$$= \sum_{u=-\infty}^{\infty} b_u \xi_u,$$

where

$$b_u = n^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)\psi_{i-u}\psi_{-u+j}(1+O(\ell n^{-1}h^{-1}))$$

and $\xi_u$ are IID random variables with mean zero and all finite moments, for all u. Using the extended Whittle's inequality as given in (3.6.1), we need to check $\sum_{u=-\infty}^{\infty} |b_u| < \infty$ first, then we have

$$E(B_1^{2k}) \le c(\sum_{u=-\infty}^{\infty} b_u^2)^k.$$

Using the boundedness of W, $\sum_{u=-\infty}^{\infty} |\psi_{i-u}\psi_{-u+j}| \le (\sum_{i=0}^{\infty} |\psi_i|)^2 < \infty$, and $n\cdot O(nh)$ terms in $\sum_{j=1}^{n} \sum_i K_h(x_j-x_i)$, then we have

$$\sum_{u=-\infty}^{\infty} |b_u|$$

$$\le c \sum_u n^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)|\psi_{i-u}\psi_{-u+j}|$$

$$\le cn^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i) < \infty.$$

Now we calculate $\sum_{u=-\infty}^{\infty} b_u^2$. According to $\sum_i K_h(x_j-x_i)\psi_{i-u} = O(h^{-1})$, the boundedness of W, $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{-u+j}\psi_{-u+j'} = \gamma(j-j')$, and $\sum_{j=1}^{n} \sum_{j'=1}^{n} \gamma(j-j') = O(n)$, then we have

$$\sum_{u=-\infty}^{\infty} b_u^2 \le c \sum_{u=-\infty}^{\infty} (n^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)\psi_{i-u}\psi_{-u+j})^2$$

$$\le c \sum_{u=-\infty}^{\infty} (n^{-1}(n-2\ell-1)^{-1} \sum_{j=1}^{n} h^{-1}\psi_{-u+j})^2$$

$$\le cn^{-2}(n-2\ell-1)^{-2}h^{-2} \sum_{u=-\infty}^{\infty} \sum_{j=1}^{n} \sum_{j'=1}^{n} \psi_{-u+j}\psi_{-u+j'}.$$

$$\le cn^{-2}(n-2\ell-1)^{-2}h^{-2} \sum_{j=1}^{n} \sum_{j'=1}^{n} \gamma(j-j') = O(n^{-3}h^{-2}).$$

The proof of (3.6.23) is complete.

To check (3.6.24), letting $u = i-p$, $v = j-q$, we have

$$B_2 = n^{-1}(n-2\ell-1)^{-1} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)\cdot$$
$$\Big( \sum\sum_{q\neq p-i+j} \psi_p\psi_q e_{i-p}e_{j-q}\Big)W(x_j)(1 + O(\ell n^{-1}h^{-1}))$$
$$= \sum_{u=-\infty,u\neq v,}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}e_u e_v,$$

where

$$b_{uv} = n^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)\psi_{i-u}\psi_{j-v}(1+O(\ell n^{-1}h^{-1})).$$

Using the extended Whittle's inequality as given in (3.6.2), we need to check $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}| < \infty$, then we have

$$E(B_2^{2k}) \le \Big( \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2\Big)^k.$$

Using the boundedness of $W$, $\sum_u |\psi_{i-u}| < \infty$, $\sum_v |\psi_{j-v}| < \infty$, and $n\cdot O(nh)$ terms in $\sum_{j=1}^{n} \sum_i K_h(x_j-x_i)$, then we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}|$$
$$\le c \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} n^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)|\psi_{i-u}\psi_{j-v}|$$
$$\le cn^{-1} \sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j) < \infty.$$

Now we calculate $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2$. Using $K_h(x_j-x_i) = O(h^{-1})$,

$\mu_2 \sum_{u=-\infty}^{\infty} \psi_{i-u}\psi_{i'-u} = \gamma(i-i')$, $\mu_2 \sum_{v=-\infty}^{\infty} \psi_{j-v}\psi_{j'-v} = \gamma(j-j')$,

$\sum_{i'} \gamma(i-i') = O(1)$, $\sum_{j'} \gamma(j-j') = O(1)$, $n\cdot O(nh)$ terms in $\sum_{j=1}^{n} \sum_i$, and the boundedness of $W$, then we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^{2}$$

$$\leq c\sum_{u}\sum_{v} \Big(n^{-1}\sum_{j=1}^{n} (n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)W(x_j)\psi_{i-u}\psi_{j-v}\Big)^2$$

$$\leq cn^{-2}(n-2\ell-1)^{-2}h^{-2}\sum_{u}\sum_{v} \Big(\sum_{j=1}^{n}\sum_{i}\psi_{i-u}\psi_{j-v}\Big)^2$$

$$\leq cn^{-2}(n-2\ell-1)^{-2}h^{-2}\sum_{u}\sum_{v}\sum_{j=1}^{n}\sum_{i}\sum_{j'=1}^{n}\sum_{i'} \psi_{i-u}\psi_{j-v}\psi_{i'-u}\psi_{j'-v}$$

$$\leq cn^{-2}(n-2\ell-1)^{-2}h^{-2}\sum_{j=1}^{n}\sum_{i}\sum_{j'=1}^{n}\sum_{i'} \gamma(i-i')\gamma(j-j')$$

$$\leq cn^{-2}(n-2\ell-1)^{-2}h^{-2}\sum_{j=1}^{n}\sum_{i} O(1) = O(n^{-2}h^{-1}).$$

The proof of (3.6.24) is complete.

# CHAPTER IV

# THE PERFORMANCE OF THE MODIFIED CROSS-VALIDATED BANDWIDTH

## 4.1 Introduction

The two optimal bandwidths, $h_A$ and $h_M$, are the minimizers of the average square error (ASE or $d_A(h)$) as given in (1.11) and the mean average square error (MASE or $d_M(h)$) as given in (1.12) respectively. In Chapter 3, we showed that the modified cross-validation criterion, $CV_\ell(h)$ as given in (3.1.1), could provide asymptotically optimal bandwidths with respect to the measures ASE and MASE when the observations suffer from a short range dependence.

However, this property is asymptotic in character. In simulation studies, we often observed that $\hat{h}_{CV(\ell)}$, the minimizer of $CV_\ell(h)$, is not close to $h_A$ or $h_M$, no matter how large the value of $\ell$ is. The purpose of this chapter is to study the rates of convergence of $h_A/h_M$, $\hat{h}_{CV(\ell)}/h_M$, $d_M(h_A)/d_M(h_M)$, and $d_M(\hat{h}_{CV(\ell)})/d_M(h_M)$.

For the equally spaced fixed, circular design and independent observations, Rice (1984) showed that $1/\hat{h}_M$, the minimizer of Mallows' criterion as given in (1.15), converges to $1/h_M$ at a relative rate $n^{-1/10}$. For the equally spaced fixed design and the independent observations, the fact that $\hat{h}_{CV(0)}$ converges to $h_A$ and $h_M$ at a relative rate $n^{-1/10}$ and $d_A(\hat{h}_{CV(0)})$ converges to $d_A(h_A)$ and $d_A(h_M)$ at a relative rate $n^{-1/5}$ could be derived from Haerdle, Hall, and Marron (1988) through a straightforward calculation.

Section 4.2 gives the limiting distributions for $h_A/h_M$, $d_M(h_A)/d_M(h_M)$, $\hat{h}_{CV(\ell)}/h_M$, and $d_M(\hat{h}_{CV(\ell)})/d_M(h_M)$. Section 4.3 contains a discussion of our results. Finally, the proofs are given in Section 4.4.

## 4.2 Limiting Distributions for Bandwidths

The purpose of this section is to derive the limiting distributions for $h_A/h_M$, $d_M(h_A)/d_M(h_M)$, $\hat{h}_{CV(\ell)}/h_M$, and $d_M(\hat{h}_{CV(\ell)})/d_M(h_M)$. In Chapter 3, we showed that the bandwidths $h_A$, $h_M$, $h_M^T$ (minimizer of $d_M^T(h)$ as given in (3.3.1)), $h_{M\ell}^S$ (minimizer of $d_{M\ell}^S(h)$ as given in (3.4.5)), and $\hat{h}_{CV(\ell)}$ are all of the same order $n^{-1/5}$. Thus we may confine our attention to h within the interval $[an^{-1/5}, bn^{-1/5}]$ for arbitrarily small a and large b. In order to derive the limiting distributions, using the equally spaced fixed design nonparametric regression model (1.1) and the assumptions (A.1) through (A.7) as given in Section 3.2, we must also impose additional assumptions:

(A.8) The bandwidth h is chosen from the interval $H_n = [an^{-1/5}, bn^{-1/5}]$ for arbitrarily small a and large b and for n = 1, 2, ....

(A.9) The number of observations deleted in the construction of the modified cross-validation score is $2\ell+1$, with $\ell \ll n^{1/2}$. The total number of observations in this regression setting is n, with $n \longrightarrow \infty$.

Before we show the limiting distributions, we shall first repeat some notation which will be used later:

$$a_1 = \sum_{k=-\infty}^{\infty} \gamma(k) \int K^2 \int W,$$

$$b_1 = (1/4)\left(\int u^2 K\right)^2 \int (m'')^2 W,$$

$$d_M(h) = a_1 n^{-1}h^{-1} + b_1 h^4 + o(n^{-1}h^{-1} + h^4),$$

$$C_0 = [\sum_{k=-\infty}^{\infty} \gamma(k) \int K^2 \int W (\int u^2 K)^{-2} (\int (m'')^2 W)^{-1}]^{1/5},$$

$$C_{0\ell}^S = [(\sum_{k=-\infty}^{\infty} \gamma(k) \int K^2 - 4 \sum_{k>\ell} \gamma(k) K(0)) \int W (\int u^2 K)^{-2} (\int (m'')^2 W)^{-1}]^{1/5}.$$

Here and throughout this chapter, the notation $\int$ denotes $\int du$. As $n \to \infty$ and $C_{0\ell}^S > 0$, given the above assumptions, it is shown in Section 4.4 that

(4.2.1) $$n^{1/10}(h_A/h_M - 1) \Rightarrow N(0, (\sum_{k=-\infty}^{\infty} \gamma(k))^{1/5} \mathrm{Var}_{AM}),$$

(4.2.2) $$n^{1/5}[d_M(h_A)/d_M(h_M) - 1] \Rightarrow C_{AM} \cdot \chi_1^2,$$

(4.2.3) $$n^{1/10}(\hat{h}_{CV(\ell)}/h_M - C_{0\ell}^S/C_0)$$
$$\Rightarrow N(0, (C_{0\ell}^S/C_0)^{-7} (\sum_{k=-\infty}^{\infty} \gamma(k))^{1/5} \mathrm{Var}_M),$$

(4.2.4) $$n^{1/5}[d_M(\hat{h}_{CV(\ell)})/d_M(h_M) - (C_{0\ell}^S/C_0)^4] \Rightarrow C_{CM} \cdot \chi_1^2,$$

where

$$\mathrm{Var}_{AM} = (8/25)[\int (K*(K-L))^2 \int W^2 \int (m'')^2 W + 4 \int W \int K^2 \int (m'' W)^2] \Big/ [(\int K^2)^9 (\int W)^9 (\int u^2 K)^2 (\int (m'')^2 W)^6]^{1/5},$$

$$C_{AM} = 2(\sum_{k=-\infty}^{\infty} \gamma(k))^{1/5} \mathrm{Var}_{AM},$$

$$C_{0\ell}^S/C_0 = [1 - (4K(0)/\int K^2)(\sum_{k>\ell} \gamma(k) \Big/ \sum_{k=-\infty}^{\infty} \gamma(k))]^{1/5},$$

$$\mathrm{Var}_M = (8/25) \int (K*(K-L)-(K-L))^2 \int W^2 \Big/ [(\int K^2)^9 (\int W)^9 (\int u^2 K)^2 (\int (m'')^2 W)]^{1/5}.$$

$$C_{CM} = 2(C^S_{0\ell}/C_0)^{-5}(\sum_{k=-\infty}^{\infty} \gamma(k))^{1/5}\mathrm{Var}_M,$$

for $L(u) = -uK'(u)$ and $*$ meaning convolution.

Give an extra condition on the functions m and W:

(A.10) The fourth derivative $m^{(4)}$ of m and the first derivative W' of W are uniformly continuous.

Thus it is shown in Section 4.4 that, as $n \longrightarrow \infty$,

$$h_M = C_0 n^{-1/5} + B_0 n^{-3/5} + o(n^{-3/5}), \tag{4.2.5}$$

where

$$B_0 = (1/20)[(\sum_{k=-\infty}^{\infty} \gamma(k)\int K^2 \int W)^{3/5}(\int u^4 K)(\int (m^{(3)})^2 W + \int m'' m^{(3)} W')] \,/\, [(\int u^2 K)^{11}(\int (m'')^2 W)^8]^{1/5}.$$

It follows from (3.4.8), (3.4.9), and (4.2.5) that

$$\hat{h}_{CV(\ell)}/h_M = C^S_{0\ell}/C_0 + (B^S_{0\ell}/C_0)n^{-2/5} + o_u(n^{-2/5}), \tag{4.2.6}$$

where

$$B^S_{0\ell} = (1/20)[(\sum_{k=-\infty}^{\infty} \gamma(k)\int K^2 - 4\sum_{k>\ell} \gamma(k)K(0))^{3/5}(\int W)^{3/5}(\int u^4 K)\cdot (\int (m^{(3)})^2 W + \int m'' m^{(3)} W')] \,/\, [(\int u^2 K)^{11}(\int (m'')^2 W)^8]^{1/5}.$$

## 4.3 Discussion

In (4.2.3), the rate of convergence $n^{-1/10}$ of $\hat{h}_{CV(\ell)}/h_M$ is excruciatingly slow. In (4.2.4), the relative loss $d_M(\hat{h}_{CV(\ell)})/d_M(h_M)$ also has a slow rate of convergence $n^{-1/5}$. These explain, in practice, why the modified cross-validation criterion does not always provide good bandwidth estimates. In these two limiting distributions for

$\hat{h}_{CV(\ell)}$, we also see that the value of $\ell$ controls the amount of asymptotic biases and asymptotic variances, and is independent of the rates of convergence. Although the two rates of convergence for $\hat{h}_{CV(\ell)}$ seem, at first glance, to be excruciatingly slow, it should not be too disappointing because $h_A$ has the same rates of convergence as $\hat{h}_{CV(\ell)}$. The limiting distributions derived in this chapter are still true when the regression errors are a short range dependent linear process as defined in Section 3.5. The only difference in this case is that the range of allowable values of $\ell$ has been significantly reduced from $o(nh)$ to $o(n^{1/2})$.

In the rest of this section, we shall use two examples of MA(1) and AR(1) processes, the two easiest cases in time series analysis, of regression errors to illustrate how the asymptotic bias-squares and the asymptotic variances depend on the value of $\ell$. Note that while (4.2.3) and (4.2.4) quantify the sample noise or "variance" of the ratios $\hat{h}_{CV(\ell)}/h_M$ and $d_M(\hat{h}_{CV(\ell)})/d_M(h_M)$, $[(C^S_{0\ell}/C_0) + (B^S_{0\ell}/C_0)n^{-2/5} - 1]$ and $[(C^S_{0\ell}/C_0) - 1]$ can be thought as measuring a type of "bias" for these two ratios respectively. Looking at (4.2.3) and (4.2.4), (4.2.5), and (4.2.6), we have the following asymptotic mean square errors (AMSE) for $[\hat{h}_{CV(\ell)}/h_M - 1]$ and $[d_M(\hat{h}_{CV(\ell)})/d_M(h_M) - 1]$ as $n \longrightarrow \infty$ and $C^S_{0\ell} > 0$:

(4.3.1) $$\text{AMSE}(\hat{h}_{CV(\ell)}/h_M - 1)$$
$$= [C^S_{0\ell}/C_0 + (B^S_{0\ell}/C_0)n^{-2/5} - 1]^2 +$$
$$n^{-1/5}[C^S_{0\ell}/C_0]^{-7}\left(\sum_{k=-\infty}^{\infty} \gamma(k)\right)^{1/5}\text{Var}_M.$$

$$(4.3.2)\quad \mathrm{AMSE}(d_M(\hat{h}_{CV(\ell)})/d_M(h_M) - 1)$$
$$= [(C^S_{0\ell}/C_0)^4 - 1]^2 + 8n^{-2/5}(C^S_{0\ell}/C_0)^{-10}(\sum_{k=-\infty}^{\infty} \gamma(k))^{2/5}\mathrm{Var}_M^{\ 2}.$$

Now we begin to discuss theoretically how the asymptotic bias-squares and the asymptotic variances of these two AMSE depend on $\ell$. If the observations are independent, then $\ell$ has no effect on these two AMSE. This is because the ratio $C^S_{0\ell}/C_0$ equals 1 for any $\ell \geq 0$. However, for the dependent observations, the effect of $\ell$ on these two AMSE are different from the case of the independent observations. It is because the ratio $C^S_{0\ell}/C_0$ as given in (4.2.3) depends on $\ell$. For the MA(1) process of regression errors: $\epsilon_j = \theta e_{j-1} + e_j$ where $e_j$ are IID $N(0,\sigma^2)$ and $-1 < \theta \leq 1$, which has been discussed in Section 3.5, we have, through a straightforward calculation,

$$C^S_{0\ell}/C_0 = \begin{cases} [1 - (4K(0)/\int K^2)(\theta/(1+\theta)^2)]^{1/5} & \text{if } \ell = 0 \\ 1 & \text{if } \ell \geq 1 \end{cases}.$$

For each $\ell \geq 1$, the $\mathrm{AMSE}(\hat{h}_{CV(\ell)}/h_M - 1)$ becomes $n^{-1/5}(1+\theta)^{2/5}\sigma^{2/5}\mathrm{Var}_M$, and the $\mathrm{AMSE}(d_M(\hat{h}_{CV(\ell)})/d_M(h_M) - 1)$ becomes $8n^{-2/5}(1+\theta)^{4/5}\sigma^{4/5}\mathrm{Var}_M^{\ 2}$, where $\sigma^2$ is the MA(1) parameter. These are the same as in the case of the independent observations. This is because the "leave-$(2\ell+1)$-out" version of $\hat{m}_j(x_j)$ as given in (3.1.2) is independent of $Y_j$ for every j when $\ell \geq 1$ in the MA(1) case. If $\theta > 0$, then the asymptotic bias-squares and the asymptotic variances of these two AMSE at $\ell = 0$ are greater than those at $\ell = 1$. In this case, we prefer to take the value of $\ell$ as any positive integer, with $\ell \ll n^{1/2}$. If $\theta < 0$, then there are some values of K, n, $\theta$, and $\sigma^2$ for which the minimum values of these two AMSE might be at $\ell = 0$. This is because, for these two AMSE, the two asymptotic bias-squares decrease and the two asymptotic variances increase, as $\ell$ increases from 0 to 1. However, the value of $Var_M$ is not obtainable in practice since it involves the unknown regression function m(x). In Chapter 6, $Var_M$ is estimated by the plug-in methods in the simulated regression setting. As $n \longrightarrow \infty$, the two asymptotic bias-squares are dominant terms in these two AMSE since the asymptotic variances tend to 0. So, in this case, we conclude that, asymptotically, $\ell$ should be taken as any positive integer, with $\ell << n^{1/2}$.

When the regression errors are an AR(1) process: $\epsilon_j = \phi\epsilon_{j-1} + e_j$ where $e_j$ are IID $N(0,\sigma^2)$ and $|\phi| < 1$, we have, through a straightforward calculation,

$$C^S_{0\ell}/C_0 = [1 - (4K(0)/\int K^2)(\phi^{\ell+1}/(1+\phi))]^{1/5}.$$

If $\phi > 0$, then the asymptotic bias-squares and the asymptotic variances of these two AMSE decrease as $\ell$ increases. This is because the ratio $C^S_{0\ell}/C_0$ is between 0 and 1, and increases to 1 as $\ell$ increases. So, in this case, we prefer to take the value of $\ell$ as large as possible. Of course, $\ell$ is restricted to $\ell << n^{1/2}$. If $\phi < 0$, then the situations are discussed on the odd number of $\ell$ and the even number of $\ell$ separately. When $\ell$ increases as an odd number, the situation is the same as in the case of $\phi > 0$. When $\ell$ increases as an even number, the two asymptotic bias-squares decrease and the two asymptotic variances increase. Then the two AMSE are convex upward on the even numbers of $\ell$. We have a simulation study to show this situation in Chapter 6.

For the general ARMA processes of regression errors, the situation becomes more complicated than that of the MA(1) and AR(1) processes. In the general case, we only know that the ratio

$$C^S_{0\ell}/C_0 = [1 - (4K(0)/\int K^2)(\sum_{k>\ell}^{\infty} \gamma(k) \; / \sum_{k=-\infty}^{\infty} \gamma(k))]^{1/5}$$

converges to 1 at a polynomial rate as $\ell$ goes to infinity. Thus, asymptotically, taking large values of $\ell$ could result in smaller values of the two AMSE than in the case of $\ell = 0$.

## 4.4 Proofs

For the proofs of (4.2.1) through (4.2.4), we need the following notation:

$$C_1 = a_1 C_0^{-1} + b_1 C_0^{4},$$

$$C_2 = 2a_1 C_0^{-3} + 12b_1 C_0^{2},$$

$$C_{2\ell}^S = 2a_{1\ell}^S (C_{0\ell}^S)^{-3} + 12b_1 (C_{0\ell}^S)^2,$$

$$r_n = n^{-1}h^{-1} + h^4,$$

$$D(h) = d_A(h) - d_M(h), \quad D_1(h) = (-h/2)D'(h),$$

$$\delta(h) = -2(\mathrm{Cross}_\ell(h) - E(\mathrm{Cross}_\ell(h))), \quad \delta_1(h) = (h/2)\delta'(h),$$

where $\mathrm{Cross}_\ell(h)$ is defined in (3.4.1). Using the assumptions (A.1) through (A.9) as given above, it is shown later that $D_1(h)$ and $\delta_1(h)$ have the following asymptotic decompositions:

(4.4.1) $$D_1(h) = S_1(h) + S_2(h) + S_3(h) + o_p(n^{-9/10}),$$

(4.4.2) $$\delta_1(h) = T_1(h) + T_2(h) + o_p(n^{-9/10}),$$

where

$$S_1(h) = \sum_{s=1, s\neq t, t=1}^{n} \sum^{n} A_{st}(h)(\epsilon_s \epsilon_t - \gamma(s-t)),$$

$$S_2(h) = \sum_{s=1}^{n} B_s(h)\epsilon_s,$$

$$S_3(h) = \sum_{s=1}^{n} A_{ss}(h)(\epsilon_s^2 - \gamma(0)),$$

$$T_1(h) = \sum_{j=1}^{n} \sum_{|i-j|>\ell} D_{ij}(h)(\epsilon_i \epsilon_j - \gamma(i-j)),$$

$$T_2(h) = \sum_{j=1}^{n} E_j(h)\epsilon_j,$$

and where the coefficients $A_{st}(h)$, $B_s(h)$, $D_{ij}(h)$, and $E_j(h)$ are given in their proofs. Given the assumptions above, as $n \longrightarrow \infty$, it is shown later that

(4.4.3) $$S_3(\beta n^{-1/5}) = o_p(n^{-9/10}),$$

$$(4.4.4) \qquad n^{9/10}(S_1+S_2)(\beta n^{-1/5}) \Rightarrow N(0,\ V_S(\sum_{k=-\infty}^{\infty} \gamma(k),\beta)),$$

$$(4.4.5) \quad n^{9/10}(-S_1-S_2+T_1+T_2)(\beta n^{-1/5}) \Rightarrow N(0,\ V_{ST}(\sum_{k=-\infty}^{\infty} \gamma(k),\beta)),$$

for any $\beta \in [a,b]$ and where

$$V_S(u,v) = 2u^2v^{-1}\int(K*(K-L))^2\int W^2 + uv^4(\int u^2K)^2\int(m''W)^2,$$

$$V_{ST}(u,v) = 2u^2v^{-1}\int(K*(K-L) - (K-L))^2\int W^2.$$

Since $h_M = C_0n^{-1/5}(1+o(1))$ as given in (3.3.6), and $h_{M\ell}^S = C_{0\ell}^S n^{-1/5}(1+o(1))$ as given in (3.4.8), then (4.4.1) through (4.4.5) imply that, as $n \longrightarrow \infty$ and $C_{0\ell}^S > 0$,

$$(4.4.6) \qquad n^{7/10}D'(h_M) \Rightarrow N(0,\ 4C_0^{-2}\cdot V_S(\sum_{k=-\infty}^{\infty} \gamma(k),C_0)),$$

$$(4.4.7) \quad n^{7/10}(D'+\delta')(h_{M\ell}^S) \Rightarrow N(0,\ 4(C_{0\ell}^S)^{-2}V_{ST}(\sum_{k=-\infty}^{\infty} \gamma(k),C_{0\ell}^S)).$$

Under the assumptions given in Section 4.2, we have the following asymptotic properties (4.4.8) through (4.4.13) which are shown later:

For each positive integer k, there is a constant $C_4$, so that

$$(4.4.8) \qquad \sup_{h\in H_n} E[r_n^{-1}h^{1/2}(|D'(h)|^{2k} + |\delta'(h)|^{2k})] \leq C_4,$$

furthermore, there is an $\epsilon > 0$ and a constant $C_5$, so that

$$(4.4.9) \quad E[r_n^{-1}h^{1/2}(|D'(h) - D'(h_1)|^{2k} + |\delta'(h) - \delta'(h_1)|^{2k})] \leq C_5|(h-h_1)/h|^{\epsilon k},$$

whenever $h, h_1 \in H_n$, with $h \leq h_1$. For any $\rho \in (0, 1/10)$, we have

$$(4.4.10) \qquad \sup_{h\in H_n} [r_n^{-1}h^{1/2}(|D'(h)| + |\delta'(h)|)] = o_p(n^\rho),$$

furthermore, if $n^{1/5}h_1$ tends to a constant, then we have

$$(4.4.11) \qquad \sup_{|h-h_1|\leq n^{-3/10+\rho}} [r_n^{-1}h^{1/2}(|D'(h) - D'(h_1)| + |\delta'(h) - \delta'(h_1)|)] = o_p(1),$$

$$(4.4.12) \qquad |h_A - h_M| + |\hat{h}_{CV(\ell)} - h_{M\ell}^S| = o_p(n^{-3/10+\rho}).$$

For any $\ell \ll n^{1/2}$, we have

(4.4.13) $$\text{Remainder}_\ell'(h) = o_p(n^{-7/10}),$$

where $\text{Remainder}_\ell(h)$ is defined in (3.4.1).

The proof of (4.2.1) is based on the expansion

(4.4.14) $$0 = d_A'(h_A) = d_M'(h_A) + D'(h_A) = (h_A - h_M)d_M''(h^*) + D'(h_A),$$

where $h^*$ lies inbetween $h_A$ and $h_M$. Using $h_A = h_M(1+o_u(1))$ as given in (3.3.6), we have

(4.4.15) $$d_M''(h^*) = C_2 n^{-2/5}(1+o_u(1)).$$

Using (4.4.12), for any $\rho \in (0, 1/10)$, $h_A = h_M + o_p(n^{-3/10+\rho})$, and using (4.4.11) (with $h_1 = h_M = C_0 n^{-1/5}(1+o(1))$), we have

$$D'(h_A) = D'(h_M) + o_p(n^{-7/10}).$$

So (4.4.14) becomes

(4.4.16) $$0 = (h_A - h_M)C_2 n^{-2/5}(1+o_u(1)) + D'(h_M) + o_p(n^{-7/10}).$$

Combining (4.4.16) with $D'(h_M) = O_p(n^{-7/10})$ as given in (4.4.6), we have

(4.4.17) $$h_A - h_M = O_p(n^{-3/10}).$$

Multiplying (4.4.16) by $n^{7/10}$, and combining the result with (4.4.17) we have

(4.4.18) $$\begin{aligned} 0 &= n^{3/10}(h_A - h_M)C_2 + n^{7/10}D'(h_M) + o_p(1) \\ &= n^{1/10}(h_A/h_M - 1)C_0C_2 + n^{7/10}D'(h_M) + o_p(1), \end{aligned}$$

for $h_M = C_0 n^{-1/5}(1+o(1))$. Then combining (4.4.18) with (4.4.6), the proof of (4.2.1) is complete.

For the limiting distribution of $d_M(h_A)/d_M(h_M)$, observe that

(4.4.19) $$[d_M(h_A) - d_M(h_M)]/d_M(h_M) = (1/2)(h_A - h_M)^2 d_M''(h^*)/d_M(h_M),$$

where $h^*$ lies inbetween $h_A$ and $h_M$. Given the facts that $d_M''(h^*) = C_2 n^{-2/5}(1+o_u(1))$ as stated in (4.4.15), $h_A - h_M = O_p(n^{-3/10})$ as stated in (4.4.17), and $d_M(h_M) = C_1 n^{-4/5}(1+o(1))$, then (4.4.19) becomes

$$(4.4.20)\quad d_M(h_A)/d_M(h_M) - 1 = (1/2)(h_A - h_M)^2 C_1^{-1} C_2 n^{2/5} + o_u(n^{-1/5}).$$

Then, combining (4.4.20), (4.4.18), and (4.4.6), the proof of (4.2.2) is complete.

The proofs of (4.2.3) and (4.2.4) are based on the expansion of $CV_\ell(h)$. According to the definitions of $D(h)$, $\delta(h)$, and $d^S_{M\ell}(h)$ as given in (3.4.5), then $CV_\ell(h)$ as given in (3.4.1) can be expressed as

$$(4.4.21)\qquad CV_\ell(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j^2 W(x_j) + d^S_{M\ell}(h) + (D + \delta)(h) + \text{Remainder}_\ell(h).$$

Then, the proofs of (4.2.3) and (4.2.4) are complete by taking the derivative of (4.4.21) and following the same steps of (4.4.14) through (4.4.20).

Proof of (4.2.5):

The proof of (4.2.5) is based on the proof of (3.3.1). Since the regression function $m(x)$ has a uniformly continuous fourth derivative and $h = O(n^{-1/5})$, then the Taylor expansion of $b_j$ as given in (3.6.8) becomes

$$\begin{aligned} b_j &= [n^{-1}\sum_{i=1}^{n} K_h(x_j - x_i)(m(x_i) - m(x_j))] \,/\, [n^{-1}\sum_{i=1}^{n} K_h(x_j - x_i)] \\ &= (1/2)h^2 m''(x_j)\int u^2 K + (1/24)h^4 m^{(4)}(x_j)\int u^4 K + o(h^4). \end{aligned}$$

Thus, the asymptotic expression of $d_M(h)$ as given in (3.6.12) and (3.6.13) becomes

$$(4.4.22)\qquad d_M(h) = a_1 n^{-1} h^{-1} + b_1 h^4 + b_2 h^6 + o(h^6),$$

where

$$\begin{aligned} a_1 &= \sum_{k=-\infty}^{\infty} \gamma(k) \int K^2 \int W, \\ b_1 &= (1/4)\left(\int u^2 K\right)^2 \int (m'')^2 W, \\ b_2 &= (-1/24)\int u^2 K \int u^4 K\left(\int (m^{(3)})^2 W + \int m'' m^{(3)} W'\right). \end{aligned}$$

Let $D_M(h) = d_M(h) - [a_1 n^{-1} h^{-1} + b_1 h^4] = b_2 h^6 + o(h^6)$. Using Taylor's theorem and $h_M = C_0 n^{-1/5}(1+o(1))$ as given in (3.3.6), where $C_0 = [a_1/(4b_1)]^{1/5}$, then we have

$$0 = d_M'(h_M) = -a_1 n^{-1} h_M^{-2} + 4b_1 h_M^3 + D_M'(h_M)$$
$$= (h_M - C_0 n^{-1/5})(2a_1 n^{-1}\tilde{h}^{-3} + 12b_1 \tilde{h}^2) + 6b_2 C_0^5 n^{-1} + o(n^{-1}),$$

where $\tilde{h}$ is inbetween $h_M$ and $C_0 n^{-1/5}$, i.e. $\tilde{h} = C_0 n^{-1/5}(1+o(1))$. Thus, solving the above inequality for $h_M - C_0 n^{-1/5}$, we have

$$h_M - C_0 n^{-1/5}$$
$$= [-6b_2 C_0^5 n^{-1} + o(n^{-1})] / [2a_1 n^{-1}\tilde{h}^{-3} + 12b_1 \tilde{h}^2]$$
$$= [-3a_1 b_2 (2b_1)^{-1} + o(n^{-1})] / [5a_1^{2/5}(4b_1)^{3/5} n^{-2/5}(1+o(1))]$$
$$= B_0 n^{-3/5}(1+o(1)),$$

where $B_0$ has been given in (4.2.5). The proof of (4.2.5) is complete.

The notation and properties used to prove (4.4.1) through (4.4.13) include:

(a) The constant c is a generic one.

(b) Given $L(u) = -uK'(u)$, $K_h(u) = h^{-1}K(u/h)$, $L_h(u) = h^{-1}L(u/h) = -uh^{-2}K'(u/h)$, then we have

$$(\partial/\partial h)K_h(u) = h^{-1}(L_h(u) - K_h(u)),$$
$$1 = \int K(u) = \int K_h(u) = \int L(u) = \int L_h(u),$$
$$0 = \int uK(u) = \int uK_h(u) = \int uL(u) = \int uL_h(u),$$
$$\int u^2 L(u) = 3\int u^2 K(u).$$

(c) For each $x_j$ with $W(x_j) \neq 0$, or $h < x_j < 1-h$, under the assumptions as given in Section 4.2, we have the following asymptotic results:

$$b_h(x_j) = n^{-1}\sum_{i=1}^{n} K_h(x_j - x_i)m(x_i) - m(x_j) = (1/2)h^2 m''(x_j)\int u^2 K + o(h^2),$$
$$c_h(x_j) = n^{-1}\sum_{i=1}^{n} L_h(x_j - x_i)m(x_i) - m(x_j) = (3/2)h^2 m''(x_j)\int u^2 K + o(h^2),$$

$$v_h(x_j) = n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i = o_u((nh)^{-2/5}),$$

$$(\partial/\partial h)(b_h(x_j)) = h^{-1}(c_h(x_j) - b_h(x_j)),$$

$$(\partial/\partial h)(v_h(x_j)) = h^{-1}[n^{-1}\sum_{i=1}^{n} L_h(x_j-x_i)\epsilon_i - v_h(x_j)],$$

$$b_j = [n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i)(m(x_i)-m(x_j))]/[n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i)]$$
$$= b_h(x_j) + O(n^{-1}h^{-1}),$$

$$b_{\ell j} = [(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)(m(x_i)-m(x_j))] \;/\; [(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)]$$
$$= b_j + O(\ell n^{-1}h + \ell n^{-2}h^{-1}) = b_j + o(h^2),$$

$$v_j = [n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i)\epsilon_i] \;/\; [n^{-1}\sum_{i=1}^{n} K_h(x_j-x_i)] = v_h(x_j) + o_u((nh)^{-7/5}),$$

$$v_{\ell j} = [(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)\epsilon_i] \;/\; [(n-2\ell-1)^{-1}\sum_{i:|i-j|>\ell} K_h(x_j-x_i)]$$
$$= v_j + o_u((nh)^{-4/5} + \ell^{3/5}(nh)^{-1}) = v_j + o((nh)^{-4/5}).$$

(d) For any i, j, s, t = 1, 2, ..., n, using the Riemann summation and the Hoelder continuity of K, L, and m, the coefficients given in (4.4.1) and (4.4.2) are

$$A_{st}(h) = n^{-3}\sum_{j=1}^{n} [K_h(x_j-x_s)K_h(x_j-x_t) - (1/2)K_h(x_j-x_s)L_h(x_j-x_t) - (1/2)L_h(x_j-x_s)K_h(x_j-x_t)]W(x_j)$$
$$= n^{-2}h^{-1}W(x_s)[K*(K-L)((s-t)/nh)] + O(n^{-2})$$
$$= O(n^{-2}h^{-1}),$$

$$B_s(h) = n^{-2}\sum_{j=1}^{n} (K_h(x_j-x_s)(2b_h(x_j) - c_h(x_j)) - L_h(x_j-x_s)b_h(x_j))W(x_j)$$
$$= (-1)n^{-1}h^2(\int u^2K)(m''W)(x_s) + o(n^{-1}h^2)$$
$$= O(n^{-1}h^2),$$

$$D_{ij}(h) = n^{-2}(K_h(x_j-x_i) - L_h(x_j-x_i))W(x_j) = O(n^{-2}h^{-1}),$$

$$E_j(h) = n^{-1}(b_h(x_j) - c_h(x_j))W(x_j)$$
$$= (-1)n^{-1}h^2\left(\int u^2K\right)(m''W)(x_j) + o(n^{-1}h^2)$$
$$= O(n^{-1}h^2).$$

Proof of (4.4.1):

Using the decomposition of $d_A(h) - d_M(h)$ as given in the proof of (3.6.14), then $D(h)$ can be expressed as

$$D(h) = d_A(h) - d_M(h)$$
$$= 2n^{-1}\sum_{j=1}^{n} b_j v_j W(x_j) + n^{-1}\sum_{j=1}^{n} (v_j^2 - E(v_j^2))W(x_j)$$
$$= 2n^{-1}\sum_{j=1}^{n} (b_h(x_j) + O(n^{-1}h^{-1}))(v_h(x_j) + o_u((nh)^{-7/5}))W(x_j) +$$
$$n^{-1}\sum_{j=1}^{n} (v_h(x_j)^2 - E(v_h(x_j)^2) + o_u((nh)^{-9/5})$$
$$= 2n^{-1}\sum_{j=1}^{n} b_h(x_j)v_h(x_j)W(x_j) +$$
$$n^{-3}\sum_{j=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} K_h(x_j-x_s)K_h(x_j-x_t)(\epsilon_s\epsilon_t - \gamma(s-t))W(x_j) + o_u((nh)^{-7/5}).$$

Using the derivatives of $K_h$, $b_h$, and $v_h$, we have

$$D_1(h) = (-h/2)D'(h) = S_1(h) + S_2(h) + S_3(h) + o_p((nh)^{-7/5}),$$

where

$$S_1(h) = \sum_{s=1, s\neq t,}^{n}\ \sum_{t=1}^{n} A_{st}(h)(\epsilon_s\epsilon_t - \gamma(s-t)),$$
$$S_2(h) = \sum_{s=1}^{n} B_s(h)\epsilon_s,$$
$$S_3(h) = \sum_{s=1}^{n} A_{ss}(h)(\epsilon_s^2 - \gamma(0)),$$

and where the definitions of $A_{st}(h)$ and $B_s(h)$ have been given above. The proof of (4.4.1) is complete.

Proof of (4.4.2):

Using the decomposition of $Cross_\ell(h) - E(Cross_\ell(h))$ as given in the proof of (3.4.4), then $\delta(h)$ can be expressed as

$$\delta(h) = -2(Cross_\ell(h) - E(Cross_\ell(h)))$$
$$= -2n^{-1} \sum_{j=1}^{n} (\epsilon_j v_{\ell j} - E(\epsilon_j v_{\ell j}))W(x_j) - 2n^{-1} \sum_{j=1}^{n} \epsilon_j b_{\ell j} W(x_j).$$

To check $T_1(h)$, using $(n-2\ell-1)^{-1} \sum_{i:|i-j|>\ell} K_h(x_j-x_i) = 1 + O(\ell n^{-1}h^{-1})$ as given in (3.6.7), then the first term in the decomposition of $\delta(h)$ becomes

$$-2n^{-1} \sum_{j=1}^{n} (\epsilon_j v_{\ell j} - E(\epsilon_j v_{\ell j}))W(x_j)$$
$$= -2n^{-2} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)(\epsilon_i\epsilon_j - \gamma(i-j))W(x_j)(1 + O(\ell n^{-1}h^{-1})).$$

Using (3.6.23) and (3.6.24), the second moment of the remainder term in the above equation is

$$E([n^{-2} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)(\epsilon_i\epsilon_j - \gamma(i-j))W(x_j)]^2) = O(n^{-2}h^{-1}).$$

This implies that the remainder term is $o_p(n^{-1}h^{-1/2})$. Using the derivative of $K_h$ as given above, we have

$$(h/2)(\partial/\partial h)(-2n^{-1} \sum_{j=1}^{n} (\epsilon_j v_{\ell j} - E(\epsilon_j v_{\ell j}))W(x_j))$$
$$= -h(\partial/\partial h)[n^{-2} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} K_h(x_j-x_i)(\epsilon_i\epsilon_j - \gamma(i-j))W(x_j)$$
$$+ o_p(n^{-1}h^{-1/2})]$$
$$= T_1(h) + o_p(n^{-1}h^{-1/2}),$$

where

$$T_1(h) = \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)(\epsilon_i\epsilon_j - \gamma(i-j)),$$
$$D_{ij}(h) = n^{-2}(K_h(x_j-x_i) - L_h(x_j-x_i))W(x_j).$$

To check $T_2(h)$, using $n^{-1} \sum_{j=1}^{n} \epsilon_j^2 = O_p(n^{-1/2})$ as given in (3.6.3), and $b_{\ell j} = b_j + o(h^2) = b_h(x_j) + o(h^2)$ as given in (3.6.9), then the second in the decomposition of $\delta(h)$ becomes

$$-2n^{-1}\sum_{j=1}^{n} \epsilon_j b_{\ell j} W(x_j) = -2n^{-1}\sum_{j=1}^{n} \epsilon_j (b_h(x_j) + o(h^2))W(x_j)$$

$$= -2n^{-1}\sum_{j=1}^{n} \epsilon_j b_h(x_j)W(x_j) + o_p(n^{-1/2}h^2).$$

Using the derivative of $b_h(x_j)$, we have

$$(h/2)(\partial/\partial h)(-2n^{-1}\sum_{j=1}^{n} \epsilon_j b_{\ell j} W(x_j))$$

$$= -h(\partial/\partial h)(n^{-1}\sum_{j=1}^{n} \epsilon_j b_h(x_j)W(x_j)) + o_p(n^{-1/2}h^2)$$

$$= T_2(h) + o_p(n^{-1/2}h^2),$$

where

$$T_2(h) = \sum_{j=1}^{n} E_j(h)\epsilon_j,$$

$$E_j(h) = n^{-1}(b_h(x_j) - c_h(x_j))W(x_j).$$

The proof of (4.4.2) is complete.

Proof of (4.4.3):

Using the linear expressions of $\epsilon_s$, $\epsilon_s^2$, and $\gamma(0)$:

$\epsilon_s = \sum_{i=0}^{\infty} \psi_i e_{s-i}$, $\epsilon_s^2 = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i\psi_j e_{s-i}e_{s-j}$, $\gamma(0) = \mu_2 \sum_{i=0}^{\infty} \psi_i^2$, then we have

$$\epsilon_s^2 - \gamma(0) = \sum_{i=0}^{\infty} \psi_i^2(e_{s-i}^2 - \mu_2) + \sum_{i=0,\, i\neq j,}^{\infty} \sum_{j=0}^{\infty} \psi_i\psi_j e_{s-i}e_{s-j}.$$

Letting $u=s-i$, $\xi_u = e_u^2 - \mu_2$, $v=s-j$, then $S_3(h)$ becomes

$$S_3(h) = \sum_{s=1}^{n} A_{ss}(h)(\epsilon_s^2 - \gamma(0))$$

$$= \sum_{s=1}^{n} A_{ss}(h) \sum_{i=0}^{\infty} \psi_i^2(e_{s-i}^2 - \mu_2) + \sum_{s=1}^{n} A_{ss}(h) \sum_{i=0,\, i\neq j,}^{\infty} \sum_{j=0}^{\infty} \psi_i\psi_j e_{s-i}e_{s-j}$$

$$= A + B,$$

where

$$A = \sum_{u=-\infty}^{\infty} a_u \xi_u ,$$

$$B = \sum_{u=-\infty, u \neq v, v=-\infty}^{\infty} \sum^{\infty} b_{uv} e_u e_v ,$$

and where

$$a_u = \sum_{s=1}^{n} A_{ss}(h) \psi_{s-u}^2 ,$$

$$b_{uv} = \sum_{s=1}^{n} A_{ss}(h) \psi_{s-u} \psi_{s-v} ,$$

and $\xi_u$ are IID random variables with mean zero and all finite moments.

So, $E(S_3(h)^2) = E(A^2) + E(B^2)$. Using the extended Whittle's inequality (3.6.1) to calculate $E(A^2)$, we need to check $\sum_{u=-\infty}^{\infty} |a_u| < \infty$ first, then $E(A^2) \leq \sum_{u=-\infty}^{\infty} a_u^2$. Since $A_{ss}(h) = O(n^{-2}h^{-1})$, and $\sum_{u=-\infty}^{\infty} \psi_{s-u}^2 < \infty$, we have

$$\sum_{u=-\infty}^{\infty} |a_u| = \sum_{u=-\infty}^{\infty} \left| \sum_{s=1}^{n} A_{ss}(h) \psi_{s-u}^2 \right| \leq c \sum_{u=-\infty}^{\infty} n^{-2} h^{-1} \sum_{s=1}^{n} \psi_{s-u}^2 = O(n^{-1}h^{-1}).$$

Using $A_{ss}(h) = O(n^{-2}h^{-1})$, $\sum_{u=-\infty}^{\infty} \psi_{s-u}^2 < \infty$, $\sum_{s'=1}^{n} \psi_{s'-u}^2 < \infty$, and $h = O(n^{-1/5})$, then we have

$$\begin{aligned}
\sum_{u=-\infty}^{\infty} a_u^2 &= \sum_{u=-\infty}^{\infty} \left[ \sum_{s=1}^{n} A_{ss}(h) \psi_{s-u}^2 \right]^2 \\
&\leq c \sum_{u=-\infty}^{\infty} \left( \sum_{s=1}^{n} \psi_{s-u}^2 \right)^2 n^{-4} h^{-2} \\
&\leq c \sum_{u=-\infty}^{\infty} \sum_{s=1}^{n} \sum_{s'=1}^{n} \psi_{s-u}^2 \psi_{s'-u}^2 n^{-4} h^{-2} \\
&= O(n^{-3}h^{-2}) = O(n^{-13/5}).
\end{aligned}$$

Using the extended Whittle's inequality (3.6.2) to calculate $E(B^2)$, we need to check $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}| < \infty$ first, then $E(B^2) \leq \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2$. Since $A_{ss}(h) = O(n^{-2}h^{-1})$, $\sum_{u=-\infty}^{\infty} |\psi_{s-u}| < \infty$, and

$\sum_{v=-\infty}^{\infty} |\psi_{s-v}| < \infty$, then we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}| \le c \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} n^{-2}h^{-1} \sum_{s=1}^{n} |\psi_{s-u}\psi_{s-v}| = O(n^{-1}h^{-1}).$$

Using $A_{ss}(h) = O(n^{-2}h^{-1})$, $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{s-u}\psi_{s'-u} = \mu_2 \sum_{v=-\infty}^{\infty} \psi_{s-v}\psi_{s'-v} =$ $\gamma(s-s')$, and $h = O(n^{-1/5})$, then we have

$$\begin{aligned}\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2 &\le c \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} n^{-4}h^{-2}[\sum_{s=1}^{n} \psi_{s-u}\psi_{s-v}]^2 \\ &\le c \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} n^{-4}h^{-2} \sum_{s=1}^{n} \sum_{s'=1}^{n} \psi_{s-u}\psi_{s-v}\psi_{s'-u}\psi_{s'-v} \\ &\le cn^{-4}h^{-2} \sum_{s=1}^{n} \sum_{s'=1}^{n} \gamma(s-s')^2 \\ &\le c \sum_{s=1}^{n} n^{-4}h^{-2} = O(n^{-3}h^{-2}) = O(n^{-13/5}).\end{aligned}$$

The proof of (4.4.3) is complete.

Proof of (4.4.4):

Using the property (b) of Section 3.2, then the ARMA process of regression errors $\epsilon_j$ can be uniquely expressed as

$$\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i},$$

where $\psi_i$ are real numbers with $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and $e_i$ are IID random variables with mean zero and all finite moments $\mu_k = E(e_1^k)$. For the proof of the asymptotic normality of $S_1+S_2$, we start from the case that $\epsilon_j$ are IID random variables, i.e. $\psi_i = 0$ for all $i > 0$. In this case, Haerdle, Hall, and Marron (1988) have shown that, as $n \longrightarrow \infty$, for any $\beta \in [a,b]$,

$$n^{9/10}(S_1+S_2)(\beta n^{-1/5}) \Rightarrow N(0, V_S(\gamma(0),\beta)),$$

where $\gamma(0) = Var(e_j)$. We now need to show that this asymptotic normality holds for the linear process $\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i}$.

Define, for j = 1, 2, ..., n, and $0 \leq q < n^{4/5}$,

$$\epsilon_{jq} = \sum_{i=0}^{q} \psi_i e_{j-i},$$

$$Y_{nq}(\beta) = n^{9/10}(S_{1q}+S_{2q})(\beta n^{-1/5}),$$

$$X_n(\beta) = n^{9/10}(S_1+S_2)(\beta n^{-1/5}) \text{ (with } \epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i}\text{)},$$

where $S_{1q}(h)$ and $S_{2q}(h)$ are $S_1(h)$ and $S_2(h)$ replaced $\epsilon_j$ by $\epsilon_{jq}$. First of all, the asymptotic normality holds for an MA(q) process with $q < n^{4/5}$, i.e. as $n \longrightarrow \infty$,

(4.4.4.1) $$Y_{nq}(\beta) \Rightarrow N(0, V_S(\mu_2(\sum_{i=0}^{q} \psi_i)^2, \beta)) \approx Y_q(\beta),$$

where the notation $A \approx B$ means A and B have the same distribution. The proof of (4.4.4.1) is given later.

To continue extending the asymptotic normality to the linear process given above, we shall use the Exercise 6.16 of Brockwell and Davis (1987):

(Exercise 6.16) Suppose that $X_n \approx N(\mu_n, v_n)$, where $\mu_n \longrightarrow \mu$, $v_n \longrightarrow v$ and $0 < v < \infty$, as $n \longrightarrow \infty$. Then $X_n \Rightarrow X$, where $X \approx N(\mu, v)$, as $n \longrightarrow \infty$.

Thus, we have, as $q \longrightarrow \infty$,

$$Y_q(\beta) \Rightarrow N(0, V_S(\mu_2(\sum_{i=0}^{\infty} \psi_i)^2, \beta)) \approx Y(\beta).$$

In order to claim that $X_n(\beta) \Rightarrow Y(\beta)$ as $n \longrightarrow \infty$, we shall use the Proposition 6.3.9 of Brockwell and Davis (1987):

(Proposition 6.3.9) Let $X_n$, n = 1, 2, ..., and $Y_{nq}$, q = 1, 2, ..., n = 1, 2, ..., be random p-vectors such that

(i) $Y_{nq} \Rightarrow Y_q$ as $n \longrightarrow \infty$ for each q = 1, 2, ...,

(ii) $Y_q \Rightarrow Y$ as $q \longrightarrow \infty$, and

(iii) $\lim_{q\to\infty} \limsup_{n\to\infty} P(|X_n - Y_{nq}| > \epsilon) = 0$ for every $\epsilon > 0$.

Then

$$X_n \Rightarrow Y \text{ as } n \longrightarrow \infty.$$

Now we only need to check the condition (iii) of the property described above to get $X_n(\beta) \Rightarrow Y(\beta)$. Using the following quantities which are shown later:

(4.4.4.2) $$\mathrm{Var}(S_1(h)) = 2n^{-2}h^{-1}(\sum_{k=-\infty}^{\infty} \gamma(k))^2\int(K*(K-L))^2\int W^2 + o(n^{-2}h^{-1}),$$

(4.4.4.3) $$\mathrm{Var}(S_2(h)) = n^{-1}h^4(\sum_{k=-\infty}^{\infty} \gamma(k))(\int u^2K)^2\int(m''W)^2 + o(n^{-1}h^4),$$

(4.4.4.4) $$\mathrm{Cov}(S_1(h),S_2(h)) = O(n^{-2}h),$$

where

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \mu_2(\sum_{i=0}^{\infty} \psi_i)^2,$$

then we have

$$\mathrm{Var}(X_n(\beta)) = V_S(\mu_2(\sum_{i=0}^{\infty} \psi_i)^2,\beta) + o(1).$$

For any q, we have

$$X_n(\beta) - Y_{nq}(\beta) = (X_n \,|\psi_i = 0,\ 0\leq i\leq q).$$

The variance of this term is

$$\mathrm{Var}(X_n(\beta) - Y_{nq}(\beta)) = V_S(\mu_2(\sum_{i>q} \psi_i)^2,\beta) + o(1).$$

Hence, using Chebychev's inequality, given any $\eta > 0$, we have

$$\begin{aligned}
&\lim_{q\to\infty} \limsup_{n\to\infty} P(|X_n(\beta) - Y_{nq}(\beta)| > \eta) \\
&\leq \lim_{q\to\infty} \limsup_{n\to\infty} \eta^{-2}\mathrm{Var}(X_n(\beta) - Y_{nq}(\beta)) \\
&\leq \lim_{q\to\infty} \eta^{-2}V_S(\mu_2(\sum_{i>q} \psi_i)^2,\beta) = 0.
\end{aligned}$$

The proof of (4.4.4) is complete.

For the proof of (4.4.4.1), using the linear expressions:

$\epsilon_{sq} = \sum_{i=0}^{q} \psi_i e_{s-i}$, $\epsilon_{tq} = \sum_{j=0}^{q} \psi_i e_{t-j}$, $\gamma(s-t) = \mu_2 \sum_{i=0}^{q} \psi_i \psi_{i-s+t}$, and

$\epsilon_{sq}\epsilon_{tq} - \gamma(s-t) = \sum_{i=0, j \neq i-s+t, j=0}^{q} \sum^{q} \psi_i \psi_j e_{s-i} e_{t-j} + \sum_{i=0}^{q} \psi_i \psi_{i-s+t}(e_{s-i}^2 - \mu_2)$, then $S_{1q}(h)$ can be decomposed into

$$S_{1q}(h) = \sum_{s=1, s \neq t, t=1}^{n} \sum^{n} A_{st}(h)(\epsilon_{sq}\epsilon_{tq} - \gamma(s-t)) = a_q(h) + b_q(h),$$

where

$$a_q(h) = \sum_{s=1, s \neq t, t=1}^{n} \sum^{n} A_{st}(h) \sum_{i=0, j \neq i-s+t, j=0}^{q} \sum^{q} \psi_i \psi_j e_{s-i} e_{t-j},$$

$$b_q(h) = \sum_{s=1, s \neq t, t=1}^{n} \sum^{n} A_{st}(h) \sum_{i=0}^{q} \psi_i \psi_{i-s+t}(e_{s-i}^2 - \mu_2).$$

Here, $a_q(h)$ can be further decomposed into

$$\begin{aligned} a_q(h) &= \sum_{s=1, s \neq t, t=1}^{n} \sum^{n} A_{st}(h) \sum_{i=0, j \neq i-s+t, j=0}^{q} \sum^{q} \psi_i \psi_j e_{s-i} e_{t-j} \\ &= \sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i \psi_j \sum_{s=1, s-t \neq 0, i-j, t=1}^{n} \sum^{n} A_{st}(h) e_{s-i} e_{t-j} \\ &= \sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i \psi_j \sum_{u,v=-q+1, u-v \neq 0, i-j}^{n} \sum^{n} A_{u+i, v+j}(h) e_u e_v \end{aligned}$$

by letting $u=s-i, v=t-j$,

$$= a_{1q}(h) + a_{2q}(h),$$

where

$$\begin{aligned} a_{1q}(h) &= \sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i \psi_j \sum_{u,v=-q+1, u-v \neq 0}^{n} \sum^{n} A_{u+i, v+j}(h) e_u e_v \\ &= \sum_{u,v=-q+1, u \neq v}^{n} \sum^{n} \left( \sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i \psi_j A_{u+i, v+j}(h) \right) e_u e_v, \end{aligned}$$

$$a_{2q}(h) = (-1) \sum_{k=-q}^{q} \left[ \sum_u \sum_i \psi_i \psi_{i-k} A_{u+i, u+i-2k}(h) e_u e_{u-k} \right].$$

Based on the assumptions given in Section 4.2, it is shown later that

(4.4.4.5) $\qquad b_q(h) = o_p(n^{-9/10})$ for all $h \in H_n$.

(4.4.4.6) $\qquad a_{2q}(h) = o_p(n^{-9/10})$ for all $h \in H_n$.

Now we rearrange $S_{2q}(h)$ as

$$S_{2q}(h) = \sum_{s=1}^{n} B_s(h)\epsilon_{jq}$$
$$= \sum_{s=1}^{n} B_s(h) \sum_{i=0}^{q} \psi_i e_{s-i}$$
$$= \sum_{u=-q+1}^{n} \left(\sum_{i=0}^{q} \psi_i B_{u+i}(h)\right)e_u \quad \text{by letting } u=s-i.$$

Thus, we have

$$S_{1q}(h) + S_{2q}(h)$$
$$= a_{1q}(h) + S_{2q}(h) + o_p(n^{-9/10})$$
$$= \sum_{u,v=-q+1, u\neq v}^{n} \left(\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j A_{u+i,v+j}(h)\right)e_u e_v + \sum_{u=-q+1}^{n} \left(\sum_{i=0}^{q} \psi_i B_{u+i}(h)\right)e_u + o_p(n^{-9/10}),$$

which is the same form as in the case of IID regression errors given in Haerdle, Hall, and Marron (1988). So, the asymptotic normality of $S_{1q}+S_{2q}$ is true. The expectation of $S_{1q}+S_{2q}$ is zero for any $h \in H_n$. Now we only need to calculate the variance of $S_{1q}+S_{2q}$. Using the following quantities which are shown later:

(4.4.4.7) $\mathrm{Var}(S_{1q}(h))$
$$= 2n^{-2}h^{-1}\left(\mu_2\left(\sum_{i=0}^{q} \psi_i\right)^2\right)^2 \int (K*(K-L))^2 \int W^2 + o(n^{-2}h^{-1}),$$

(4.4.4.8) $\mathrm{Var}(S_{2q}(h)) = n^{-1}h^4\mu_2\left(\sum_{i=0}^{q} \psi_i\right)^2\left(\int u^2K\right)^2\int (m''W)^2 + o(n^{-1}h^4),$

(4.4.4.9) $\mathrm{Cov}(S_{1q}(h), S_{2q}(h)) = O(n^{-2}h).$

Replacing h with $\beta n^{-1/5}$ in (4.4.4.7) through (4.4.4.9), then we have

$$\mathrm{Var}(n^{9/10}(S_{1q}+S_{2q})(\beta n^{-1/5})) = V_S\left(\mu_2\left(\sum_{i=0}^{q} \psi_i\right)^2, \beta\right).$$

The proof of (4.4.4.1) is complete.

For the proof of (4.4.4.5), changing of indices, then $b_q(h)$ can be expressed as

$$b_q(h) = \sum_{s=1, s\neq t, t=1}^{n} \sum^{n} A_{st}(h) \sum_{i=0}^{q} \psi_i\psi_{i-s+t}(e_{s-i}{}^2 - \mu_2)$$

$$= \sum_{s=1, s\neq t, t=1}^{n} \sum^{n} A_{st}(h) \sum_{u=-q+1}^{n} \psi_{s-u}\psi_{-u+t}\xi_u$$

by letting $u=s-i$, $\xi_u = e_u^2 - \mu_2$,

$$= \sum_{u=-q+1}^{n} \Big( \sum_{s=1, s\neq t, t=1}^{n} \sum^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t}\Big)\xi_u .$$

Using (1) of Whittle's inequality, independence between $\xi_u$, and $A_{st}(h) = O(n^{-2}h^{-1})$ for any s, t, then we have

$$E(b_q(h)^2) \le c \sum_{u=-q+1}^{n} \Big( \sum_{s=1, s\neq t, t=1}^{n} \sum^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t}\Big)^2$$

$$\le c \sum_{u=-q+1}^{n} n^{-4}h^{-2} = O(n^{-13/5}),$$

since $h = O(n^{-1/5})$. Then, we have $b_q(h) = o_p(n^{-9/10})$.

The proof of (4.4.4.5) is complete.

For the proof of (4.4.4.6), using $A_{st}(h) = O(n^{-2}n^{-1})$ for any s, t, $O(n)$ terms in $\sum_u$, $\sum_i \psi_i\psi_{i-k} = O(1)$, and $h = O(n^{-1/5})$, then we have

$$E(a_{2q}(h)^2) = E\Big(\sum_{k=-q}^{q} \sum_u \sum_i \psi_i\psi_{i-k}A_{u+i,u+i-2k}(h)e_u e_{u+k}\Big)^2$$

$$\le cn^{-4}h^{-2}E\Big(\sum_{k=-q}^{q} \sum_u \sum_{k'=-q}^{q} \sum_{u'} e_u e_{u+k} e_{u'} e_{u'+k'}\Big)$$

$$\le cn^{-4}h^{-2}n^{4/5}n^{4/5} = O(n^{-2})$$

since the subindices of the random variables must pairwisely equal, otherwise the expectations are zero. Thus, $a_{2q}(h) = o_p(n^{-1})$.

The proof of (4.4.4.6) is complete.

For the proof of (4.4.4.7), using the Hoelder continuity of K, L, and m, $n\cdot O(nh)$ terms in $\sum_{u,v=-q+1, u\neq v}^{n} \sum^{n}$, and the following quantities:

$$A_{st}(h) = n^{-2}h^{-1}W(x_s)(K*(K-L)((s-t)/nh)) + O(n^{-2}),$$

$$W(x_{u+i}) = W(x_u) + O(|i|/n) = W(x_u) + O(n^{-1}) \quad \text{for } |i| \le q,$$

$$K*(K-L)((u+i-v-j)/nh) = K*(K-L)((u-v)/nh) + O(|i-j|/nh)$$

$$= K*(K-L)((u-v)/nh) + O(n^{-1}h^{-1}) \quad \text{for } |i-j| \le q.$$

then we have

$$\mathrm{Var}(a_{1q}(h))$$

$$= \mathrm{Var}\Big(\sum_{u,v=-q+1,u\neq v}^{n} \sum^{n} \Big(\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j A_{u+i,v+j}(h)\Big)e_u e_v\Big)$$

$$= 2\mu_2^2 \sum_{u,v=-q+1,u\neq v}^{n} \sum^{n} \Big(\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j A_{u+i,v+j}(h)\Big)^2$$

since $A_{u+i,v+j}(h) = A_{v+j,u+i}(h)$

$$= 2\mu_2^2 \sum_{u,v=-q+1,u\neq v}^{n} \sum^{n} \Big[\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j n^{-2}h^{-1}W(x_{u+i})(K*(K-L)((u+i-v-j)/nh)) + O(n^{-2})\Big]^2$$

$$= 2\mu_2^2 \sum_{u,v=-q+1,u\neq v}^{n} \sum^{n} \Big[\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j n^{-2}h^{-1}W(x_u)(K*(K-L)((u-v)/nh)) + O(n^{-2})\Big]^2$$

$$= 2\mu_2^2 n^{-4}h^{-2}\Big(\sum_{i=0}^{q} \psi_i\Big)^4 \sum_{u,v=-q+1,u\neq v}^{n} \sum^{n} W(x_u)^2(K*(K-L)((u-v)/nh)) + O(n^{-2})$$

$$= 2\mu_2^2 n^{-2}h^{-1}\Big(\sum_{i=0}^{q} \psi_i\Big)^4 \int W^2 \int (K*(K-L))^2 + o(n^{-2}h^{-1}).$$

Since $S_{1q}(h) = a_{1q}(h) + a_{2q}(h) + b_q(h)$, $\mathrm{Var}(a_{1q}(h)) = O(n^{-2}h^{-1})$, and $E(b_q(h)^2)$ and $E(a_{2q}(h)^2)$ are $O(n^{-3}h^{-2})$ as given in the proof of (4.4.4.5) and (4.4.4.6), then we have $\mathrm{Var}(S_{1q}(h)) = \mathrm{Var}(a_{1q}(h)) + O(n^{-3}h^{-2})$. The proof of (4.4.4.7) is complete.

For the proof of (4.4.4.8), using the continuity of W and m'', and

$$B_s(h) = (-1)n^{-1}h^2\Big(\int u^2K\Big)(m''W)(x_s) + o(n^{-1}h^2) \quad \text{for all } s$$

$$(m''W)(x_{u+i}) = (m''W)(x_u) + o(1) \qquad \text{for } |i|\leq q,$$

then we have

$$\mathrm{Var}(S_{2q}(h))$$

$$= \mathrm{Var}\Big(\sum_{u=-q+1}^{n} \Big(\sum_{i=0}^{q} \psi_i B_{u+i}(h)\Big)e_u\Big)$$

$$= \mu_2 \sum_{u=-q+1}^{n} \Big[\sum_{i=0}^{q} \psi_i B_{u+i}(h)\Big]^2$$

$$= \mu_2 \sum_{u=-q+1}^{n} \Big[\sum_{i=0}^{q} \psi_i (-1)n^{-1}h^2\Big(\int u^2K\Big)(m''W)(x_{u+i}) + o(n^{-1}h^2)\Big]^2$$

$$= \mu_2 \sum_{u=-q+1}^{n} [\sum_{i=0}^{q} \psi_i(-1)n^{-1}h^2(\int u^2K)(m''W)(x_u) + o(n^{-1}h^2)]^2$$

$$= \mu_2(\sum_{i=0}^{q} \psi_i)^2 n^{-2}h^4(\int u^2K)^2 \sum_{u=-q+1}^{n} (m''W)(x_u)^2 + o(n^{-1}h^4)$$

$$= \mu_2(\sum_{i=0}^{q} \psi_i)^2 n^{-1}h^4(\int u^2K)^2 \int (m''W)^2 + o(n^{-1}h^4).$$

The proof of (4.4.4.8) is complete.

For the proof of (4.4.4.9), using $Var(S_{2q}(h)) = O(n^{-1}h^4)$, $E(a_{2q}(h)^2) = O(n^{-3}h^{-2})$, and $E(b_q(h)^2) = O(n^{-3}h^{-2})$, then we have

$$\begin{aligned} &Cov(S_{1q}(h), S_{2q}(h)) \\ &= Cov(a_{1q}(h)+a_{2q}(h)+b_q(h), S_{2q}(h)) \\ &= Cov(a_{1q}(h), S_{2q}(h)) + O(n^{-2}h) \\ &= E(a_{1q}(h) \cdot S_{2q}(h)) + O(n^{-2}h). \end{aligned}$$

Thus,

$$\begin{aligned} &E(a_{1q}(h) \cdot S_{2q}(h)) \\ &= \sum_{u,v=-q+1, u\neq v}^{n} \sum{}'^{n} \sum_{z=-q+1}^{n} (\sum_{i=0}^{q} \sum_{j=0}^{q} \psi_i\psi_j A_{u+i,v+j}(h))(\sum_{i=0}^{q} \psi_i B_{z+i}(h))E(e_u e_v e_z) \\ &= 0, \end{aligned}$$

since

$$E(e_u e_v e_z) = \begin{cases} \mu_3 & \text{if } u=v=z \\ 0 & \text{otherwise} \end{cases}.$$

The proof of (4.4.4.9) is complete.

For the proof of (4.4.4.2), using the linear expressions of $\epsilon_s$, $\epsilon_t$, and $\gamma(s-t)$: $\epsilon_s = \sum_{i=0}^{\infty} \psi_i e_{s-i}$, $\epsilon_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$, $\gamma(s-t) = \mu_2 \sum_{i=0}^{\infty} \psi_i \psi_{i-s+t}$, and letting $u=s-i$, $v=t-j$, $\xi_u = e_u^2 - \mu_2$, then $S_1(h)$ can be decomposed into $A + B$, where

$$A = \sum_{u=-\infty}^{\infty} a_u \xi_u,$$

$$B = \sum_{u=-\infty, u\neq v, v=-\infty}^{\infty} \sum^{\infty} b_{uv} e_u e_v,$$

and where

$$a_u = \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t},$$

$$b_{uv} = \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{t-v},$$

and $\xi_u$ are IID random variables with mean zero and all finite moments. Since A and B are uncorrelated, then $Var(S_1(h)) = E(A^2) + E(B^2)$.

Using the extended Whittle's inequality (3.6.1), we need to check $\sum_{u=-\infty}^{\infty} |a_u| < \infty$, then we have $E(A^2) \leq \sum_{u=-\infty}^{\infty} a_u^2$. Since $A_{st}(h) = O(n^{-2}h^{-1})$ for any s, t, and $\sum_{u=-\infty}^{\infty} |\psi_{s-u}\psi_{-u+t}| < \infty$, $O(n^2h)$ terms in $\sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n}$,

then we have

$$\sum_{u=-\infty}^{\infty} |a_u| = \sum_{u=-\infty}^{\infty} \Big| \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t}\Big|$$
$$\leq c \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} n^{-2}h^{-1} = O(1).$$

Since $A_{st}(h) = O(n^{-2}h^{-1})$ for any s, t, $\sum_{t=1}^{n} \psi_{-u+t} = O(1)$,

$\mu_2 \sum_{u=-\infty}^{\infty} \psi_{s-u}\psi_{s'-u} = \gamma(s-s')$, and $\sum_{s=1}^{n} \sum_{s'=1}^{n} \gamma(s-s') = O(n)$, then we have

$$\sum_{u=-\infty}^{\infty} a_u^2 = \sum_{u=-\infty}^{\infty} \Big[ \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t}\Big]^2$$
$$\leq c \sum_{u=-\infty}^{\infty} n^{-4}h^{-2}\Big( \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} \psi_{s-u}\psi_{-u+t}\Big)^2$$
$$\leq c \sum_{u=-\infty}^{\infty} n^{-4}h^{-2} \sum_{s=1}^{n} \sum_{s'=1}^{n} \psi_{s-u}\psi_{s'-u}$$
$$= O(n^{-3}h^{-2}) = o(n^{-2}h^{-1}).$$

Now we calculate $E(B^2)$. Since $b_{uv} = b_{vu}$, we have

$$E(B^2) = 2 \sum_{u=-\infty,u\neq v,}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2\mu_2^2 = 2\,\mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2 + o(n^{-2}h^{-1}),$$

where

$$\sum_{u=-\infty}^{\infty} b_{uu}^2 \leq c \sum_{u=-\infty}^{\infty} n^{-4}h^{-2}\Big( \sum_{s=1}^{n} \sum_{t=1}^{n} |\psi_{s-u}\psi_{t-u}|\Big)^2 = O(n^{-3}h^{-2}).$$

Using the the Riemann summation, changing of variable, and the Hoelder continuity of K, L, and W, then we have the following quantities:

$$A_{st}(h) = n^{-2}h^{-1}W(x_s)[K*(K-L)((s-t)/nh)] + O(n^{-2}),$$

$$A_{s-p,t-q}(h) = n^{-2}h^{-1}W(x_{s-p})[K*(K-L)((s-p-t+q)/nh)] + O(n^{-2})$$

$$= n^{-2}h^{-1}W(x_s)[K*(K-L)((s-t)/nh)] + O(n^{-2}+|p|n^{-3}h^{-1}+|p-q|n^{-3}h^{-2}).$$

Thus, we have

$$\mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2$$

$$= \mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} [\sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{t-v}]^2$$

$$= \mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} \sum_{s'=1,s'\neq t',}^{n} \sum_{t'=1}^{n} A_{st}(h)A_{s't'}(h)\psi_{s-u}\psi_{t-v}\psi_{s'-u}\psi_{t'-v}$$

$$= \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} \sum_{s'=1,s\neq t,}^{n} \sum_{t'=1}^{n} A_{st}(h)A_{s't'}(h)\gamma(s-s')\gamma(t-t')$$

for $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{s-u}\psi_{s'-u} = \gamma(s-s')$, $\mu_2 \sum_{v=-\infty}^{\infty} \psi_{t-v}\psi_{t'-v} = \gamma(t-t')$,

$$= \sum_{|p|<n} \sum_{|q|<n} \gamma(p)\gamma(q) \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)A_{s-p,t-q}(h)$$

by letting $p=s-s'$, $q=t-t'$,

$$= n^{-4}h^{-2}(\sum_{|p|<n} \gamma(p))^2 \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} W(x_s)^2(K*(K-L)((x_s-x_t)/h))^2 + o(n^{-2}h^{-1})$$

for $\sum_{p=-\infty}^{\infty} |p\gamma(p)| < \infty$, $\sum_{q=-\infty}^{\infty} |q\gamma(q)| < \infty$,

$$= n^{-2}h^{-1}(\sum_{|p|<n} \gamma(p))^2[n^{-1}\sum_{s} W(x_s)^2][n^{-1}\sum_{r} (K*(K-L)(rn^{-1}h^{-1}))^2] + o(n^{-2}h^{-1})$$

by letting $r=s-t$ and $u=((x_s-x_t)/h)$,

$$= n^{-2}h^{-1}(\sum_{|p|<n} \gamma(p))^2\int W^2\int(K*(K-L))^2 + o(n^{-2}h^{-1}).$$

The proof of (4.4.4.2) is complete.

For the proof of (4.4.4.3), using the linear expression of $\epsilon_s$, then $S_2(h)$ can be expressed as

$$S_2(h) = \sum_{u=-\infty}^{\infty} b_u e_u,$$

where

$$b_u = \sum_{s=1}^{n} B_s(h)\psi_{s-u}.$$

Using the Hoelder continuity of K, L, and W, the uniform continuity of m'', and the following quantities:

$$B_s(h) = (-1)n^{-1}h^2(\int u^2K)(m''W)(x_s) + o(n^{-1}h^2),$$

$$B_{s-p}(h) = (-1)n^{-1}h^2(\int u^2K)(m''W)(x_s) + o(|p|n^{-1}h^2),$$

$$(m''W)(x_s+hu) = (m''W)(x_s) + o(1),$$

then we have

$$\begin{aligned}
&\mathrm{Var}(S_2(h)) \\
&= \sum_{u=-\infty}^{\infty} b_u^2\mu_2 = \sum_{s=1}^{n}\sum_{s'=1}^{n} B_s(h)B_{s'}(h)\gamma(s-s') \\
&= \sum_{|p|<n} \gamma(p) \sum_{s=1}^{n} B_s(h)B_{s-p}(h) \qquad \text{by letting } p=s-s', \\
&= \sum_{|p|<n} \gamma(p)\, n^{-2}h^4(\int u^2K)^2 \sum_{s=1}^{n} (m''W)^2(x_s) + o(n^{-1}h^4) \\
&\qquad\qquad \text{for } \sum_{p=-\infty}^{\infty} |p\gamma(p)| < \infty, \\
&= n^{-1}h^4(\sum_{|p|<n} \gamma(p))(\int u^2K)^2\int(m''W)^2 + o(n^{-1}h^4).
\end{aligned}$$

The proof of (4.4.4.3) is complete.

For the proof of (4.4.4.4), because $S_1(h)$, $S_2(h)$, and $\epsilon_j$ have mean zero, we have

$$\begin{aligned}
&\mathrm{Cov}(S_1(h),S_2(h)) = E(S_1(h)\cdot S_2(h)) \\
&= E(\sum_{s=1,s\neq t,t=1}^{n}\sum^{n} A_{st}(h)(\epsilon_s\epsilon_t - \gamma(s-t))\cdot \sum_{u=1}^{n} B_u(h)\epsilon_u) \\
&= \sum_{s=1,s\neq t,t=1}^{n}\sum^{n}\sum_{u=1}^{n} A_{st}(h)B_u(h)\cdot E(\epsilon_s\epsilon_t\epsilon_u).
\end{aligned}$$

To calculate $E(\epsilon_s\epsilon_t\epsilon_u)$, using the linear expressions of $\epsilon_s$, $\epsilon_t$, and $\epsilon_u$:

$\epsilon_s = \sum_{p=0}^{\infty} \psi_p e_{s-p}$, $\epsilon_t = \sum_q \psi_{q-s+t} e_{s-q}$, $\epsilon_u = \sum_r \psi_{r-s+u} e_{s-r}$, then we have

$$E(\epsilon_s \epsilon_t \epsilon_u) = \sum_p \sum_q \sum_r \psi_p \psi_{q-s+t} \psi_{r-s+u} E(e_{s-p} e_{s-q} e_{s-r})$$
$$= \mu_3 \sum_{p=0}^{\infty} \psi_p \psi_{p-s+t} \psi_{p-s+u},$$

where

$$E(e_{s-p} e_{s-q} e_{s-r}) = \begin{cases} \mu_3 & \text{if } p = q = r \\ 0 & \text{otherwise} \end{cases},$$

for $e_i$ are IID random variables with mean zero and all finite moments $\mu_k = E(e_1^k)$ for all positive integers k.

Using $A_{st}(h) = O(n^{-2}h^{-1})$, $B_u(h) = O(n^{-1}h^2)$, $\sum_{u=1}^{n} \psi_{p-s+u} = O(1)$, $\mu_2 \sum_{p=0}^{\infty} \psi_p \psi_{p-s+t} = \gamma(s-t)$, and $\sum_{s=1}^{n} \sum_{t=1}^{n} \gamma(s-t) = O(n)$, then we have

$$\mathrm{Cov}(S_1(h), S_2(h)) = \sum_{s=1, s \neq t,}^{n} \sum_{t=1}^{n} \sum_{u=1}^{n} A_{st}(h) B_u(h) \mu_3 \sum_{p=0}^{\infty} \psi_p \psi_{p-s+t} \psi_{p-s+u}$$
$$\leq c n^{-3} h \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{u=1}^{n} \sum_{p=0}^{\infty} \psi_p \psi_{p-s+t} \psi_{p-s+u}$$
$$= O(n^{-2} h).$$

The proof of (4.4.4.4) is complete. Then the proof of (4.4.4) is complete.

Proof of (4.4.5):

The unique solution of the ARMA process of regression errors $\epsilon_j$ can be expressed as $\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i}$, where $\psi_i$ are real numbers with $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and $e_i$ are IID random variables with mean zero and all finite moments. For the proof of the asymptotic normality of $(-S_1-S_2+T_1+T_2)$, we shall use the same arguments as in the proof of (4.4.4). Starting from the case that $\epsilon_j$ are IID random variables i.e.

$\psi_i = 0$ for all $i > 0$, and $\ell = 0$, Haerdle, Hall, and Marron (1988) showed that

$$n^{9/10}(-S_1-S_2+T_1+T_2)(\beta n^{-1/5}) \Rightarrow N(0,\ V_{ST}(\gamma(0),\beta)),$$

where $\gamma(0) = Var(e_j)$. Using the same arguments in Haerdle, Hall, and Marron (1988), this asymptotic normality can be extended to the general case $\ell << nh$ for IID regression errors by showing that

(4.4.5.1) $\quad Var(T_1(h)) = 2n^{-2}h^{-1}\sigma^4\int(K-L)^2\int W^2 + o(n^{-2}h^{-1}).$

The proof of (4.4.5.1) is given later.

Following the same steps as in the proof of (4.4.4.4), the asymptotic normality of $(-S_1-S_2+T_1+T_2)$ holds for the MA(q) process with $q < n^{4/5}$. Then, using the Exercise 6.16 and the Proposition 6.3.9 of Brockwell and Davis (1987), the asymptotic normality of $(-S_1-S_2+T_1+T_2)$ holds for the linear process $\epsilon_j = \sum_{i=0}^{\infty} \psi_i e_{j-i}$. Thus, the proof of (4.4.5) is complete by showing the following quantities whose proofs are given later:

(4.4.5.2) $\quad Var(T_1(h)) = 2n^{-2}h^{-1}(\sum_{k=-\infty}^{\infty} \gamma(k))^2\int(K-L)^2\int W^2 + o(n^{-2}h^{-1}).$

(4.4.5.3) $\quad Var(T_2(h)) = n^{-1}h^4(\sum_{k=-\infty}^{\infty} \gamma(k))(\int u^2K)^2\int(m''W)^2 + o(n^{-1}h^4).$

(4.4.5.4) $\quad Cov(T_1(h),S_1(h)) = 2n^{-2}h^{-1}(\sum_{k=-\infty}^{\infty} \gamma(k))^2\int(K*(K-L)(K-L))\int W^2 + o(n^{-2}h^{-1}).$

(4.4.5.5) $Cov(T_2(h),S_2(h)) = n^{-1}h^4(\sum_{k=-\infty}^{\infty} \gamma(k))(\int u^2K)^2\int(m''W)^2 + o(n^{-1}h^4).$

(4.4.5.6) $\quad Cov(T_1(h),T_2(h)) + Cov(T_1(h),S_2(h)) + Cov(T_2(h),S_1(h)) = O(n^{-2}h),$

where

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \mu_2(\sum_{i=0}^{\infty} \psi_i)^2.$$

The proof of (4.4.5) is complete.

For the proof of (4.4.5.1), since $D_{ij}(h) = D_{ji}(h)$ and $\epsilon_j$ are IID random variables with mean zero and all finite moments, then we have

$$\mathrm{Var}(T_1(h)) = 2\sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)^2\gamma(0)^2.$$

Letting k=j-i, then we have

$$\begin{aligned}
&2\sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)^2 \\
&= 2n^{-4}\sum_{j=1}^{n} \sum_{i:|i-j|>\ell} (K_h(x_j-x_i) - L_h(x_j-x_i))^2 W(x_j)^2 \\
&= 2n^{-4}\sum_{|k|>\ell} (K_h(k/n) - L_h(k/n))^2 \cdot \sum_i W(x_{k+i})^2 \\
&= 2n^{-2}\left[\int (K_h(t) - L_h(t))^2 dt + O(\ell n^{-1}h^{-2})\right]\left[\int W^2 + O(h)\right] \\
&= 2n^{-2}h^{-1}\int (K-L)^2 \int W^2 + o(n^{-2}h^{-1}).
\end{aligned}$$

The proof of (4.4.5.1) is complete.

For the proofs of (4.4.5.2) and (4.4.5.3), using the linear expressions of $\epsilon_i$, $\epsilon_j$, and $\gamma(i-j)$, then $\delta_1(h)$ can be decomposed into $T_{1A}(h) + T_{1B}(h) + T_2(h)$, where

$$T_{1A}(h) = \sum_{u=-\infty}^{\infty} a_u \xi_u,$$

$$T_{1B}(h) = \sum_{u=-\infty,\, u\neq v,}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv} e_u e_v,$$

$$T_2(h) = \sum_{u=-\infty}^{\infty} b_u e_u,$$

and where

$$a_u = \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)\psi_{i-u}\psi_{-u+j},$$

$$b_{uv} = \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)\psi_{i-u}\psi_{j-v},$$

$$b_u = \sum_{j=1}^{n} E_j(h)\psi_{j-u}.$$

Thus, we have

$$Var(T_1(h)) = E(T_{1A}(h)^2) + E(T_{1B}(h)^2).$$

Use the extended Whittle's inequality (3.6.1) to check that $E(T_{1A}(h)^2) = o(n^{-2}h^{-1})$. For $\sum_i |\psi_{i-u}| < \infty$ and $\sum_{u=-\infty}^{\infty} |\psi_{-u+j}| < \infty$, we have

$$\sum_{u=-\infty}^{\infty} |a_u| \le c \sum_{u=-\infty}^{\infty} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} n^{-2}h^{-1} |\psi_{i-u}\psi_{-u+j}| \le cn^{-1}h^{-1}.$$

Using $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{-u+j}\psi_{-u+j'} = \gamma(j-j')$ and $\sum_{j=1}^{n} \sum_{j'=1}^{n} \gamma(j-j') = O(n)$, we have

$$\begin{aligned}
E(T_{1A}(h)^2) &\le c \sum_{u=-\infty}^{\infty} a_u^2 \\
&= c \sum_{u=-\infty}^{\infty} \Big( \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)\psi_{i-u}\psi_{-u+j}\Big)^2 \\
&\le cn^{-4}h^{-2} \sum_{u=-\infty}^{\infty} \Big( \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \psi_{i-u}\psi_{-u+j}\Big)^2 \\
&\le cn^{-4}h^{-2} \sum_{u=-\infty}^{\infty} \sum_{j=1}^{n} \sum_{j'=1}^{n} \psi_{-u+j}\psi_{-u+j} \qquad \text{for } \sum_i |\psi_{i-u}| < \infty, \\
&\le cn^{-3}h^{-2} = o(n^{-2}h^{-1}).
\end{aligned}$$

To check $E(T_{1B}(h)^2)$, use the extended Whittle's inequality (3.6.2). Since $\sum_{u=-\infty}^{\infty} b_{uu}^2 = o(n^{-2}h^{-1})$ and $b_{uv} = b_{vu}$, then we have

$$\begin{aligned}
E(T_{1B}(h)^2) &= E\Big( \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}e_u e_v\Big)^2 + o(n^{-2}h^{-1}) \\
&= 2\mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2 + o(n^{-2}h^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
&\mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2 \\
&= \mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \Big( \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)\psi_{i-u}\psi_{j-v}\Big)^2 \\
&= \mu_2^2 \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{j'=1}^{n} \sum_{i':|i'-j'|>\ell} D_{ij}(h)D_{i'j'}(h)\psi_{i-u}\psi_{j-v}\psi_{i'-u}\psi_{j'-v} \\
&= \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{j'=1}^{n} \sum_{i':|i'-j'|>\ell} D_{ij}(h)D_{i'j'}(h)\gamma(i-i')\gamma(j-j')
\end{aligned}$$

for $\mu_2 \sum_{u=-\infty}^{\infty} \psi_{i-u}\psi_{i'-u} = \gamma(i-i')$ and $\mu_2 \sum_{v=-\infty}^{\infty} \psi_{j-v}\psi_{j'-v} = \gamma(j-j')$.

$$= \sum_{|p|<n} \sum_{|q|<n} \gamma(p)\gamma(q) \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)D_{i-q,j-p}(h)$$

by letting $p = j - j'$, $q = i - i'$,

$$= \sum_{|p|<n} \sum_{|q|<n} \gamma(p)\gamma(q) \sum_{|k|>\ell} \sum_{i} n^{-2}(K_h(x_k) - L_h(x_k))W(x_{i+k})\cdot$$
$$n^{-2}(K_h(x_{k-p+q}) - L_h(x_{k-p+q}))W(x_{k+i-p})$$

by letting $k=j-i$, and $D_{ij}(h)=n^{-2}(K_h-L_h)((x_j-x_i))W(x_j)$,

$$= (\sum_{|p|<n} \gamma(p))^2 \sum_{|k|>\ell} \sum_{i} n^{-4}(K_h(x_k) - L_h(x_k))^2 W(x_{i+k})^2 + O(n^{-3}h^{-1})$$

for $W(x_{k+i-p}) = W(x_{i+k}) + O(|p|/n)$, $\sum_{p=0}^{\infty} |p\gamma(p)| < \infty$,

$\sum_{q=0}^{\infty} |q\gamma(q)| < \infty$,

and $(K_h-L_h)(x_{k-p+q}) = (K_h-L_h)(x_k) + O(|p-q|n^{-1}h^{-1})$,

$$= (\sum_{|p|<n} \gamma(p))^2 n^{-2}[\int(K_h(t) - L_h(t))^2 dt + O(\ell n^{-1}h^{-2})][\int W^2 + O(h)] +$$
$$O(n^{-3}h^{-1})$$

$$= n^{-2}h^{-1}(\sum_{|p|<n} \gamma(p))^2 \int(K-L)^2 \int W^2 + o(n^{-2}h^{-1}).$$

The proof of (4.4.5.2) is complete.

For the proof of (4.4.5.3), through a straightforward calculation, we have

$$\mathrm{Var}(T_2(h))$$
$$= \mu_2(\sum_{u=-\infty}^{\infty} b_u^2)$$
$$= \mu_2 \sum_{u=-\infty}^{\infty} (\sum_{j=1}^{n} E_j(h)\psi_{j-u})^2$$
$$= \mu_2 \sum_{u=-\infty}^{\infty} \sum_{j=1}^{n} \sum_{j'=1}^{n} E_j(h)E_{j'}(h)\psi_{j-u}\psi_{j'-u}$$
$$= \sum_{j=1}^{n} \sum_{j'=1}^{n} E_j(h)E_{j'}(h)\gamma(j-j') \qquad \text{for } \mu_2 \sum_{u=-\infty}^{\infty} \psi_{j-u}\psi_{j'-u} = \gamma(j-j').$$
$$= \sum_{|p|<n} \gamma(p) \sum_{j=1}^{n} E_j(h)E_{j-p}(h) \qquad \text{by letting } p=j-j'.$$

where

$$E_j(h) = (-1)n^{-1}h^2\left(\int u^2K\right)(m''W)(x_j) + o(n^{-1}h^2),$$

$$E_{j-p}(h) = (-1)n^{-1}h^2\left(\int u^2K\right)(m''W)(x_j) + o(|p|n^{-1}h^2).$$

Thus

$$\begin{aligned}
&\mathrm{Var}(T_2(h)) \\
&= \Big(\sum_{|p|<n} \gamma(p)\Big)n^{-2}h^4\left(\int u^2K\right)^2 \sum_{j=1}^{n} (m''W)(x_j)^2 + o(n^{-1}h^4) \\
&= \Big(\sum_{|p|<n} \gamma(p)\Big)n^{-1}h^4\left(\int u^2K\right)^2\int(m''W)^2 + o(n^{-1}h^4).
\end{aligned}$$

The proof of (4.4.5.3) is complete.

For the proof of (4.4.5.4), follow the decompositions used in the proof of (4.4.4.1) and (4.4.5.2). Using the superscripts T and S denote the coefficients $b_{uv}$ used for $T_1(h)$ and $S_1(h)$ respectively, we have

$$\begin{aligned}
&\mathrm{Cov}(T_1(h),S_1(h)) \\
&= \mathrm{Cov}\Big(\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} b^T_{uv}e_ue_v,\ \sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} b^S_{uv}e_ue_v\Big) \\
&= 2\mu_2^2 \sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} b^T_{uv}b^S_{uv} \qquad \text{for } b^T_{uv}=b^T_{vu},\ b^S_{uv}=b^S_{vu}, \\
&= 2\mu_2^2 \sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} \Big[\sum_{j=1}^{n}\sum_{i:|i-j|>\ell} D_{ij}(h)\psi_{i-u}\psi_{j-v}\Big]\Big[\sum_{s=1,s\neq t,t=1}^{n}\sum^{n} A_{st}(h)\psi_{s-u}\psi_{t-v}\Big] \\
&= 2\sum_{|p|<n}\sum_{|q|<n} \gamma(p)\gamma(q)\sum_{j=1}^{n}\sum_{i:|i-j|>\ell} D_{ij}(h)A_{i-p,j-q}(h) \\
&\qquad \text{for } \mu_2\sum_{u=-\infty}^{\infty}\psi_{i-u}\psi_{s-u} = \gamma(i-s),\ \mu_2\sum_{v=-\infty}^{\infty}\psi_{j-v}\psi_{t-v} = \gamma(j-t), \\
&\qquad \text{and by letting } p=i-s,\ q=j-t, \\
&= 2\sum_{|p|<n}\sum_{|q|<n} \gamma(p)\gamma(q)\sum_{|k|>\ell} n^{-2}h^{-1}\sum_i (K-L)(x_k/h)W(x_{i+k})A_{i-p,k+i-q}(h) \\
&\qquad \text{by letting } k=j-i,\ D_{ij}(h)=n^{-2}(K_h-L_h)(x_j-x_i)W(x_j), \\
&= 2\Big(\sum_{|p|<n}\gamma(p)\Big)^2n^{-4}h^{-2}\sum_{|k|>\ell}\big((K-L)\cdot(K*(K-L))\big)(x_k/h)\cdot\sum_i W(x_i)^2 + O(n^{-2})
\end{aligned}$$

because in the proof of (4.4.4.1) we had

$$A_{i-p,k+i-q}(h) = n^{-2}h^{-1}W(x_{i-p})K*(K-L)(x_{k+p-q}/h) + O(n^{-2}),$$

$$K*(K-L)(x_{k+p-q}/h) = K*(K-L)(x_k/h) + O(|p-q|n^{-1}h^{-1}),$$

$$W(x_{i-p}) = W(x_i) + O(|p|/n), \text{ and } W(x_{i+k}) = W(x_i) + O(h),$$

$$= 2n^{-2}h^{-1}(\sum_{|p|<n} \gamma(p))^2 \int K*(K-L)\cdot(K-L)\int W^2 + o(n^{-2}h^{-1}).$$

The proof of (4.4.5.4) is complete.

For the proof of (4.4.5.5), following the same decompositions used in the proof of (4.4.5.2) and (4.4.4.3), where the superscripts T and S denote the coefficients for $T_2(h)$ and $S_2(h)$ respectively, we have

$$\mathrm{Cov}(T_2(h),S_2(h))$$

$$= \mathrm{Cov}(\sum_{u=-\infty}^{\infty} b_u^T e_u, \sum_{u=-\infty}^{\infty} b_u^S e_u)$$

$$= \mu_2 \sum_{u=-\infty}^{\infty} b_u^T b_u^S \qquad \text{since } e_u \text{ are IID random variables,}$$

$$= \mu_2 \sum_{u=-\infty}^{\infty} (\sum_{j=1}^{n} E_j(h)\psi_{j-u})(\sum_{s=1}^{n} B_s(h)\psi_{s-u})$$

$$= \sum_{j=1}^{n}\sum_{s=1}^{n} E_j(h)B_s(h)\gamma(j-s) \qquad \text{for } \mu_2 \sum_{u=-\infty}^{\infty} \psi_{j-u}\psi_{s-u} = \gamma(j-s)$$

$$= \sum_{|p|<n} \gamma(p) \sum_{s=1}^{n} E_{p+s}(h)B_s(h) \qquad \text{by letting } p=j-s,$$

$$= \sum_{|p|<n} \gamma(p)n^{-2}h^4(\int u^2K)^2 \sum_{s=1}^{n} (m''W)(x_s)^2 + o(n^{-1}h^4)$$

since in the proof of (4.4.4.2) we had

$$B_s(h) = -n^{-1}h^2\int u^2K(m''W)(x_s) + o(n^{-1}h^2) \text{ and}$$

$$E_{p+s}(h) = (-1)n^{-1}h^2\int u^2K(m''W)(x_s) + o(|p|n^{-1}h^2),$$

$$= n^{-1}h^4(\sum_{|p|<n} \gamma(p))(\int u^2K)^2\int(m''W)^2 + o(n^{-1}h^4).$$

The proof of (4.4.5.5) is complete.

For the proof of (4.4.5.6), following the proof of (4.4.4.4), we have similar results. Through a straightforward calculation, we have

$$\mathrm{Cov}(T_1(h),T_2(h))$$

$$= \mathrm{Cov}\Big(\sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)(\epsilon_i\epsilon_j-\gamma(i-j)),\ \sum_{k=1}^{n} E_k(h)\epsilon_k\Big)$$

$$= \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{k=1}^{n} D_{ij}(h)E_k(h)E(\epsilon_i\epsilon_j\epsilon_k)$$

where $E(\epsilon_i\epsilon_j\epsilon_k) = \mu_3 \sum_{p=0}^{\infty} \psi_p\psi_{p-i+j}\psi_{p-i+k}$.

$$= \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{k=1}^{n} D_{ij}(h)E_k(h)\mu_3 \sum_{p=0}^{\infty} \psi_p\psi_{p-i+j}\psi_{p-i+k}$$

where $D_{ij}(h) = O(n^{-2}h^{-1})$, $E_k(h) = O(n^{-1}h^2)$.

$$\le c \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} n^{-2}h^{-1}n^{-1}h^{2}\gamma(i-j) = O(n^{-2}h)$$

for $\sum_{i=1}^{n} \sum_{i:|i-j|>\ell} \gamma(i-j) = O(n)$.

Using the same reason as above, we have

$$\mathrm{Cov}(T_1(h),S_2(h))$$

$$= \mathrm{Cov}\Big(\sum_{j=1}^{n} \sum_{i:|i-j|>\ell} D_{ij}(h)(\epsilon_i\epsilon_j-\gamma(i-j)),\ \sum_{k=1}^{n} B_k(h)\epsilon_k\Big)$$

$$= \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{k=1}^{n} D_{ij}(h)B_k(h)\cdot E(\epsilon_i\epsilon_j\epsilon_k)$$

$$= \sum_{j=1}^{n} \sum_{i:|i-j|>\ell} \sum_{k=1}^{n} D_{ij}(h)B_k(h)\cdot \sum_{p=0}^{\infty} \psi_p\psi_{p-i+j}\psi_{p-i+k}$$

$= O(n^{-2}h)$ for $D_{ij}(h) = O(n^{-2}h^{-1})$ and $B_k(h) = O(n^{-1}h^2)$.

Using the same reason as above, we have

$$\mathrm{Cov}(T_2(h),S_1(h))$$

$$= \mathrm{Cov}\Big(\sum_{j=1}^{n} E_j(h)\epsilon_j,\ \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} A_{st}(h)(\epsilon_s\epsilon_t-\gamma(s-t))\Big)$$

$$= \sum_{j=1}^{n} \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} E_j(h)A_{st}(h)E(\epsilon_j\epsilon_s\epsilon_t)$$

$$= \sum_{j=1}^{n} \sum_{s=1,s\neq t,}^{n} \sum_{t=1}^{n} E_j(h)A_{st}(h) \sum_{p=0}^{\infty} \psi_p\psi_{p-i+s}\psi_{p-j+t}$$

$= O(n^{-2}h)$ for $E_j(h) = O(n^{-1}h^2)$, and $A_{st}(h) = O(n^{-2}h^{-1})$.

The proof of (4.4.5.6) is complete. So, the proof of (4.4.5) is complete.

Proofs of (4.4.8) through (4.4.12):

The proofs of the first parts of (4.4.8) through (4.4.13) are given first. Using the linear expressions of $\epsilon_s$, $\epsilon_t$, and $\gamma(s-t)$, we rearrange $D_1(h)$ as

$$D_1(h) = \sum_{s=1}^{n} \sum_{t=1}^{n} A_{st}(h)(\epsilon_s\epsilon_t - \gamma(s-t)) + S_2(h)$$
$$= S_{1A}(h) + S_{1B}(h) + S_2(h),$$

where

$$S_{1A}(h) = \sum_{u=-\infty}^{\infty} a_u\xi_u,$$
$$S_{1B}(h) = \sum_{u=-\infty,\, u\neq v,}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}e_ue_v,$$
$$S_2(h) = \sum_{u=-\infty}^{\infty} b_ue_u,$$

and where

$$a_u = \sum_{s=1}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{-u+t},$$
$$b_{uv} = \sum_{s=1}^{n} \sum_{t=1}^{n} A_{st}(h)\psi_{s-u}\psi_{t-v},$$
$$b_u = \sum_{s=1}^{n} B_s(h)\psi_{s-u}.$$

For the proof of (4.4.8), it is enough to check that, for any $h = O(n^{-1/5})$ and any positive integer k,

$$E(|r_n^{-1}h^{-1/2}S_{1A}(h)|^{2k}) \leq C_4,$$
$$E(|r_n^{-1}h^{-1/2}S_{1B}(h)|^{2k}) \leq C_4,$$
$$E(|r_n^{-1}h^{-1/2}S_2(h)|^{2k}) \leq C_4.$$

Since $r_n^{-1}h^{-1/2} = O(n^{9/10})$, it is equivalent to check that, for any $h = O(n^{-1/5})$, the expectaions of $S_{1A}(h)^{2k}$, $S_{1B}(h)^{2k}$, $S_2(h)^{2k}$ are all of order $O(n^{-9k/5})$. Using the extended Whittle's inequality (3.6.1) and

(3.6.2) to check the absolute convergence of $\sum_{u=-\infty}^{\infty} |a_u|$,
$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}|$, and $\sum_{u=-\infty}^{\infty} |b_u|$ first, then we have

$$E(S_{1A}(h)^{2k}) \leq (\sum_{u=-\infty}^{\infty} a_u^2)^k,$$

$$E(S_{1B}(h)^{2k}) \leq (\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2)^k,$$

$$E(S_2(h)^{2k}) \leq (\sum_{u=-\infty}^{\infty} b_u^2)^k,$$

which imply that we need to check, for $h = O(n^{-1/5})$, $\sum_{u=-\infty}^{\infty} a_u^2$, $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2$, and $\sum_{u=-\infty}^{\infty} b_u^2$ are all of the order $O(n^{-9/5})$.

In the proof of (4.4.4.3) through (4.4.4.4), we had

$$\sum_{u=-\infty}^{\infty} |a_u| = O(n^{-1}h^{-1}), \quad \sum_{u=-\infty}^{\infty} a_u^2 = O(n^{-3}h^{-2}) = O(n^{-9/5}),$$

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}| = O(1), \quad \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}^2 = O(n^{-2}h^{-1}) = O(n^{-9/5}),$$

$$\sum_{u=-\infty}^{\infty} |b_u| = O(h^2), \quad \sum_{u=-\infty}^{\infty} b_u^2 = O(n^{-1}h^4) = O(n^{-9/5}).$$

The proof of (4.4.8) is complete.

For the proof of (4.4.9), since $r_n^{-1}h^{-1/2} = O(n^{9/10})$, it is enough to check that, there is an $\epsilon > 0$, such that, for any $h, h_1 \in H_n$, with $h \leq h_1$, the expectations of $(S_{1A}(h) - S_{1A}(h_1))^{2k}$, $(S_{1B}(h) - S_{1B}(h_1))^{2k}$, and $(S_2(h) - S_2(h_1))^{2k}$ are all of the order $O(n^{-9k/5}|(h-h_1)/h|^{\epsilon k})$.
Using the absolutely convergent series $\sum_{u=-\infty}^{\infty} |a_u(h)|$, $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}(h)|$, and $\sum_{u=-\infty}^{\infty} |b_u(h)|$ as given in the proofs of (4.4.4.2) through (4.4.4.4), then we have the absolutely convergent series $\sum_{u=-\infty}^{\infty} |a_u(h) - a_u(h_1)|$, $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} |b_{uv}(h) - b_{uv}(h_1)|$, and

$\sum_{u=-\infty}^{\infty} |b_u(h) - b_u(h_1)|$. Using the extended Whittle's inequality (3.6.1) and (3.6.2), we have

$$E((S_{1A}(h) - S_{1A}(h_1))^{2k}) \le (\sum_{u=-\infty}^{\infty} (a_u(h) - a_u(h_1))^2)^k,$$

$$E((S_{1B}(h) - S_{1B}(h_1))^{2k}) \le (\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} (b_{uv}(h) - b_{uv}(h_1))^2)^k,$$

$$E((S_2(h) - S_2(h_1))^{2k}) \le (\sum_{u=-\infty}^{\infty} (b_u(h) - b_u(h_1))^2)^k.$$

These inequalities imply that we need to check $\sum_{u=-\infty}^{\infty} (a_u(h) - a_u(h_1))^2$, $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} (b_{uv}(h) - b_{uv}(h_1))^2$, $\sum_{u=-\infty}^{\infty} (b_u(h) - b_u(h_1))^2$ are all of the order $O(n^{-9/5}|(h-h_1)/h|^{\epsilon})$.

For $S_{1A}(h)$, using $A_{st}(h) = O(n^{-2}h^{-1})$, $\sum_{t=1}^{n} \psi_{-u+t} = O(1)$, and $\sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} \psi_{s-u})^2 = \sum_{u=-\infty}^{\infty} \sum_{s=1}^{n} \sum_{s'=1}^{n} \psi_{s-u}\psi_{s'-u} = \sum_{s=1}^{n} \sum_{s'=1}^{n} \gamma(s-s') = O(n)$, then we have

$$\begin{aligned}
&\sum_{u=-\infty}^{\infty} (a_u(h) - a_u(h_1))^2 \\
&= \sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} \sum_{t=1}^{n} (A_{st}(h) - A_{st}(h_1))\psi_{s-u}\psi_{-u+t})^2 \\
&\le c(n^{-2}h^{-1} - n^{-2}h_1^{-1})^2 \cdot \sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} \sum_{t=1}^{n} \psi_{s-u}\psi_{-u+t})^2 \\
&\le cn^{-4}(h^{-1} - h_1^{-1})^2 \cdot n \\
&\le cn^{-3}h_1^{-2}|(h-h_1)/h|^2 = O(n^{-13/5}|(h-h_1)/h|^2).
\end{aligned}$$

For $S_{1B}(h)$, we have

$$\begin{aligned}
&\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} (b_{uv}(h) - b_{uv}(h_1))^2 \\
&= \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} (b_{uv}(h)^2 - 2b_{uv}(h)b_{uv}(h_1) + b_{uv}(h)^2).
\end{aligned}$$

From the proof of (4.4.4.2), we have $\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}(h)^2 = O(n^{-2}h^{-1})$ and

$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}(h_1)^2 = O(n^{-2}h_1^{-1})$. Following the same steps as in the proof of (4.4.4.2), we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} b_{uv}(h)b_{uv}(h_1)$$
$$= n^{-2}h_1^{-1}(\sum_{|p|<n} \gamma(p))^2 \int W^2 \int K*(K-L)(t) \cdot K*(K-L)(th/h_1)\, dt + o(n^{-2}h_1^{-1})$$
$$= O(n^{-2}h_1^{-1}).$$

Thus, we have

$$\sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} (b_{uv}(h) - b_{uv}(h_1))^2$$
$$\leq c|n^{-2}h^{-1} - n^{-2}h_1^{-1}|$$
$$\leq cn^{-2}h_1^{-1}|(h-h_1)/h|$$
$$\leq cn^{-2}h_1^{-1}|(h-h_1)/h|^2 = O(n^{-9/5}|(h-h_1)/h|).$$

For $S_2(h)$, using $B_s(h) = O(n^{-1}h^2)$ and $\sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} \psi_{s-u})^2 = O(n)$, then we have

$$\sum_{u=-\infty}^{\infty} (b_u(h) - b_u(h_1))^2$$
$$= \sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} (B_s(h) - B_s(h_1))\psi_{s-u})^2$$
$$\leq c(n^{-1}h^2 - n^{-1}h_1^{\,2})^2 \sum_{u=-\infty}^{\infty} (\sum_{s=1}^{n} \psi_{s-u})^2$$
$$\leq cn^{-2}(h^2 - h_1^{\,2})^2 \cdot n$$
$$\leq cn^{-1}h^4((h^2 - h_1^{\,2})/h^2)^2$$
$$\leq cn^{-1}h^4((h+h_1)/h)^2((h-h_1)/h)^2 = O(n^{-9/5}|(h-h_1)/h|^2).$$

So, we take $\epsilon$ as 2. The proof of (4.4.9) is complete.

For the proof of (4.4.10), it is equivalent to show that

$$\sup_{h \in H_n} |r_n^{-1}h^{-1/2}D_1(h)| = o_p(n^{\rho}).$$

In the proof of (4.4.8), we have $E(S_{1A}(h)^2) = O(n^{-3}h^{-2})$, $E(S_{1B}(h)^2) = O(n^{-2}h^{-1})$, $E(S_2(h)^2) = O(n^{-1}h^4)$, from which, we have

$S_{1A}(h)$, $S_{1B}(h)$, $S_2(h)$ are $h^{-1}\cdot o_p(1)$. Thus, for any $h$, $h_1 \in H_n$, with $h \le h_1$, taking $r \ge 6/5$, then we have

$$\sup_{|(h-h_1)/h|\le n^{-r}} |D_1(h) - D_1(h_1)|$$

$$\le \sup_{|(h-h_1)/h|\le n^{-r}} (|S_{1A}(h)-S_{1A}(h_1)| + |S_{1B}(h)-S_{1B}(h_1)| + |S_2(h)-S_2(h_1)|)$$

$$\le \sup_{|(h-h_1)/h|\le n^{-r}} h^{-1}|(h-h_1)/h|\cdot o_p(1) = n^{-1/5}n^{-r}\cdot o_p(1) = o_p(n^{-1}),$$

which implies that

$$\sup_{h\in H_n} |D_1(h)| \le \sup_{h_1\in H_n'} |D_1(h_1)| + \sup_{|(h-h_1)/h|\le n^{-r}} |D_1(h) - D_1(h_1)|$$

$$\le \sup_{h_1\in H_n'} |D_1(h_1)| + o_p(n^{-1}).$$

Since $r_n^{-1}h^{-1/2} = O(n^{9/10})$ and $0 < \rho < 1/10$, the second term of the right hand side in the above inequality is $o_p(r_n h^{1/2} n^{\rho})$. Hence, it is sufficient to restrict the supremum in the statement of (4.4.10) to a set $H_n'$, which is a subset of $H_n$ so that $\#(H_n') \le n^{r+1}$ and so that for any $h \in H_n$, there is an $h_1 \in H_n'$ with $|(h-h_1)/h| \le n^{-r}$.

Using Bonferroni's inequality and (4.4.8), we have, for any $\eta > 0$, any positive integer k, as $n \longrightarrow \infty$,

$$P(\sup_{h\in H_n'} |r_n^{-1}h^{-1/2}n^{-\rho}D_1(h)| > \eta)$$

$$\le \#(H_n')\cdot\sup_{h\in H_n} E((\eta^{-1}r_n^{-1}h^{-1/2}n^{-\rho}D_1(h))^{2k})$$

$$\le cn^{r+1}\cdot\sup_{h\in H_n} E((\eta^{-1}r_n^{-1}h^{1/2}n^{-\rho}D'(h))^{2k})$$

$$\le cn^{r+1}\cdot n^{-2k\rho} \longrightarrow 0$$

by taking k sufficiently large to make $r+1-2k\rho < 0$.

The proof of (4.4.10) is complete.

For the proof of (4.4.11), it is equivalent to check

$$\sup_{|h-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h) - D_1(h_1))| = o_p(1).$$

Using the result of (4.4.10), through adding and subtracting $D_1(h')$, then we have

$$\sup_{|h-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h) - D_1(h_1))|$$

$$\le \sup_{|(h-h')/h|\le n^{-r}} |r_n^{-1}h^{-1/2}(D_1(h) - D_1(h'))| + \sup_{|h'-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h') - D_1(h_1))|$$

$$\le o_p(n^{-1/10}) + \sup_{|h'-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h') - D_1(h_1))|.$$

Thus we only need to check that

$$\sup_{|h'-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h') - D_1(h_1))| = o_p(1).$$

For $a < \lim n^{1/5}h_1 < b$ and for any $r > 1/10$, suppose that

$$h_1 - n^{-3/10+\rho} = t_0 < t_1 < t_1 < \ldots < t_p = h_1 + n^{-3/10+\rho},$$

where $t_i - t_{i-1} = n^{-(1/5)-r}$ for each i. To check

$$\sup_{|h'-h_1|\le n^{-3/10+\rho}} |r_n^{-1}h^{-1/2}(D_1(h') - D_1(h_1))| = o_p(1),$$

it suffices to check that

$$\sup_{(t_i,t_j)\in F} r_n^{-1}h^{-1/2}|D_1(t_i) - D_1(t_j)| = o_p(1),$$

where F is the set of all pairs $(t_i,t_j)$ with $0 < |t_i - t_j| \le n^{-3/10+\rho}$ and $0 \le i, j \le p$. The number of elements in F is of order $n^{2(r+\rho-1/10)}$.

Given any $\eta > 0$ and any positive integer k, using Bonferroni's inequality and (4.4.9), where $|(t_i-t_j)/t_i| \le n^{-1/10+\rho}$, as $n \longrightarrow \infty$, then we have

$$P(\sup_{(t_i,t_j)\in F} r_n^{-1}h^{-1/2}|D_1(t_i) - D_1(t_j)| > \eta)$$

$$\le \sup_{(t_i,t_j)\in F} E((\eta^{-1}r_n^{-1}h^{-1/2}|D_1(t_i) - D_1(t_j)|)^{2k})$$

$$\leq c \sup_{(t_i, t_j)\in F} E((\eta^{-1} r_n^{-1} h^{1/2} |D'(t_i) - D'(t_j)|)^{2k})$$

$$\leq cn^{2(r+\rho-1/10)} n^{(-1/10+\rho)\epsilon k} \longrightarrow 0,$$

by taking k sufficiently large to make $2(r+\rho-1/10) + (-1/10+\rho)\epsilon k < 0$. The proof of (4.4.11) is complete.

For the proof of (4.4.12), using (4.4.10), we have

$$o_p(n^{-7/10+\rho}) = D'(h_M) = d_A'(h_M) - d_M'(h_M) = d_A'(h_M).$$

Thus,

$$\begin{aligned} d_A'(h_M) &= d_A'(h_M) - d_A'(h_A) \\ &= (D'(h_M)+d_M'(h_M)) - (D'(h_A)+d_M'(h_A)) \\ &= d_M'(h_M) - d_M'(h_A) + o_p(n^{-7/10}). \end{aligned}$$

So, we have

$$o_p(n^{-7/10+\rho}) = d_M'(h_M) - d_M'(h_A) = (h_M - h_A)d_M''(h^*),$$

where $h^*$ lies inbetween $h_A$ and $h_M$. Combining this result with (4.4.15), we have

$$h_M - h_A = o_p(n^{-3/10+\rho}).$$

The proof of (4.4.12) is complete.

For the proofs of the second parts of (4.4.8) through (4.4.12), we shall use the decomposition of $\delta_1(h) = T_{1A}(h) + T_{1B}(h) + T_2(h)$ as given in the proof of (4.4.5.2). The orders of the coefficients of $T_{1A}(h)$, $T_{1B}(h)$, and $T_2(h)$ are the same as those of $S_{1A}(h)$, $S_{1B}(h)$, and $S_2(h)$ respectively. So the proofs of the second parts of (4.4.8), (4.4.10), (4.3.11) are exactly the same as the proofs of their first parts. Following the proof of (4.4.5.2), the proof of the second part of (4.4.9) is exactly the same as the proof of its first part.

For the proof of (4.4.12), we start from the first derivative of $CV_\ell(h)$ as given in (4.4.21). Using $\text{Remainder}_\ell'(h) = o_p(n^{-7/10})$ as given in (4.4.13), then

$$CV_\ell'(h) = d_{M\ell}^{S}{}'(h) + (D + \delta)'(h) + o_p(n^{-7/10}).$$

From (4.4.10) and (4.4.11), we have

$$\begin{aligned}
& o_p(n^{-7/10+\rho}) \\
&= (D + \delta)'(h_{M\ell}^{S}) = (D + \delta)'(h_{M\ell}^{S}) - CV_\ell'(\hat{h}_{CV(\ell)}) \\
&= (D + \delta)'(h_{M\ell}^{S}) - d_{M\ell}^{S}{}'(\hat{h}_{CV(\ell)}) - (D + \delta)'(\hat{h}_{CV(\ell)}) + o_p(n^{-7/10}) \\
&= D'(h_{M\ell}^{S}) - D'(\hat{h}_{CV(\ell)}) + \delta'(h_{M\ell}^{S}) - \delta'(\hat{h}_{CV(\ell)}) - d_{M\ell}^{S}{}'(\hat{h}_{CV(\ell)}) + o_p(n^{-7/10}) \\
&= -d_{M\ell}^{S}{}'(\hat{h}_{CV(\ell)}) + o_p(n^{-7/10}) \\
&= (h_{M\ell}^{S} - \hat{h}_{CV(\ell)}) d_{M\ell}^{S}{}''(h^*) + o_p(n^{-7/10}),
\end{aligned}$$

where $h^*$ lies inbetween $h_{M\ell}^{S}$ and $\hat{h}_{CV(\ell)}$. Using $\hat{h}_{CV(\ell)}/h_{M\ell}^{S} \longrightarrow 1$ a.s. as given in (3.4.9), we have

$$d_{M\ell}^{S}{}''(h^*) = C_{2\ell}^{S} n^{-2/5}(1+o_u(1)).$$

Thus, we have

$$o_p(n^{-7/10+\rho}) = (h_{M\ell}^{S} - \hat{h}_{CV(\ell)}) C_{2\ell}^{S} n^{-2/5}(1+o_u(1)),$$

which implies

$$h_{M\ell}^{S} - \hat{h}_{CV(\ell)} = o_p(n^{-3/10+\rho}).$$

The proof of the second part of (4.4.12) is complete.

Proof of (4.4.13):

Using the same decomposition as given in the proof of (3.4.2), then $\text{Remainder}_\ell(h)$ can be decomposed as $\text{Remainder}_\ell(h) = A + B + C$, where

$$A = n^{-1} \sum_{j=1}^{n} (b_{\ell j} - b_j)(b_{\ell j} + b_j) W(x_j),$$

$$B = 2n^{-1} \sum_{j=1}^{n} (v_{\ell j} b_{\ell j} - v_j b_j) W(x_j),$$

$$C = n^{-1} \sum_{j=1}^{n} (v_{\ell j} - v_j)(v_{\ell j} + v_j) W(x_j).$$

In the proof of (3.4.2), we have that, for term A,

$$A = O(\ell n^{-1} h^3 + \ell n^{-2} h).$$

Thus, the first derivative of A can be expressed as

$$(\partial/\partial h)(A) = O(\ell n^{-1}h^2 + \ell n^{-2}) = o_u(n^{-7/10})$$

for $\ell \ll n^{1/2}$ and $h = O(n^{-1/5})$. For term B, through subtracting and adding the term $v_j b_{\ell j}$, we have

$$B = 2n^{-1} \sum_{j=1}^{n} [(v_{\ell j} - v_j)b_{\ell j} + v_j(b_{\ell j} - b_j)]W(x_j),$$

where

$$(v_{\ell j} - v_j)b_{\ell j} = o_u(\ell^{3/5}n^{-1}h^{-1}\cdot h^2) = o_u(\ell^{3/5}n^{-1}h),$$
$$v_j(b_{\ell j} - b_j) = o_u((nh)^{-2/5})\cdot O(\ell n^{-1}h + \ell n^{-2}h^{-1}).$$

Thus, we have the first derivative

$$(\partial/\partial h)[(v_{\ell j} - v_j)b_{\ell j} + v_j(b_{\ell j} - b_j)]$$
$$= o_u(\ell^{3/5}n^{-1} + \ell n^{-7/5}h^{-2/5} + \ell n^{-12/5}h^{-12/5})$$
$$= o_u(n^{-7/10}) \quad \text{for } \ell \ll n^{1/2} \text{ and } h = O(n^{-1/5}).$$

This implies that the first derivative of B is $o_p(n^{-7/10})$. For term C, using the following quantities: $v_{\ell j} - v_j = o_u(\ell^{3/5}n^{-1}h^{-1})$, $v_{\ell j} + v_j = O_p((nh)^{-1/2})$, then we have

$$(\partial/\partial h)((v_{\ell j} - v_j)(v_{\ell j} + v_j))$$
$$= (\partial/\partial h)(o_u(\ell^{3/5}n^{-1}h^{-1}\cdot(nh)^{-1/2}))$$
$$= o_u(\ell^{3/5}n^{-1}) = o_u(n^{-7/10}) \quad \text{for } \ell \ll n^{1/2} \text{ and } h = O(n^{-1/5}).$$

This implies that the first derivative of C is $o_p(n^{-7/10})$.

The proof of (4.4.13) is complete.

CHAPTER V

# AN APPLICATION OF THE PARTITIONED CROSS-VALIDATION

## 5.1 Introduction

Recall from Chapters 3 and 4 that the optimal bandwidth, $h_M$, and the data-driven bandwidth, $\hat{h}_{CV(\ell)}$, are the minimizers of the mean average square error (MASE or $d_M(h)$) as given in (1.12), and the modified cross-validation score $CV_\ell(h)$ as given in (3.1.1) respectively. The ordinary cross-validation criterion CV(h) as given in (1.14) could not provide asymptotically optimal bandwidths because of the dependence between the "leave-1-out" version of $\hat{m}_j(x_j)$ as given in (1.13) and $Y_j$. In Chapters 3 and 4, we used the "leave-$(2\ell+1)$-out" version of $\hat{m}_j(x_j)$ as given in (3.1.2) to reduce the effect of dependence on bandwidth estimation. However, (4.2.3) and (4.2.4) showed that $\hat{h}_{CV(\ell)}$ suffers from slow relative rates of convergence to the optimal bandwidth $h_M$.

To improve the relative rate of convergence of the cross-validated bandwidth, Marron (1988) proposed the partitioned cross-validation (PCV) for kernel density estimation. In the PCV, the observations are split into g subgroups by taking every g-th observation. This means that the observations of each subgroup are distanced by g/n. Through this property, we could use the PCV to estimate bandwidth when the observations are dependent. This is because the dependence effect on the ordinary cross-validation score of each subgroup of observations is reduced as the value of g is increased. So, the PCV and the $CV_\ell(h)$ have a similar ability to alleviate the dependence effect on bandwidth estimation.

Section 5.2 gives some results for the ordinary cross-validated bandwidth of each subgroup of observations. The asymptotic normality of the partitioned cross-validated bandwidth is given in Section 5.3. A discussion of these results is given in Section 5.4. Finally, the proofs are given in Section 5.5.

## 5.2 Partitioned Cross-validation

The purpose of this section is to derive the asymptotic properties of the ordinary cross-validated bandwidth of each subgroup of observations. The number of observations in each subgroup is $\eta = n/g$. Here, for notational convenience, n is assumed to be a multiple of g. To derive the asymptotic properties, using the equally spaced fixed design nonparametric regression model (1.1), the assumptions (A.1) through (A.6) as given in Section 3.2, and the assumption (A.10) as given in Section 4.2, we must also impose the following assumptions:

(A.11) For the PCV, the number of subgroups is g. The number of observations in each subgroup is $\eta$, with $\eta \longrightarrow \infty$.

(A.12) For the PCV, the bandwidth h is chosen from the interval $H_{n,g} = [a\eta^{-1/5}, b\eta^{-1/5}]$ for arbitrarily small a and large b and for $\eta = 1, 2, \ldots$.

Before we show the asymptotic properties for the ordinary cross-validated bandwidth of each subgroup of observations, we shall first introduce notation. The notation, as defined in Chapters 3 and 4, with a subscript $\ell$ is simplified by ignoring the subscript $\ell$. Because, in this chapter, we only consider the ordinary cross-validated bandwidth, i.e. $\ell = 0$ in the modified cross-validation criterion. For example, $d_M^S(h)$ and $h_M^S$ now represent $d_{MO}^S(h)$ and $h_{MO}^S$ respectively. Secondly, a symbol with a subscript k means that the value is taken on the k-th subgroup of observations, for each $k = 1, 2, \ldots, g$. For example, $CV_k(h)$ denotes the ordinary cross-validation score CV(h) as given in (1.14) for the k-th subgroup of observations. Thirdly, the notation with a superscript * denotes the average of the values of the notation of all g subgroups of observations. For example, $CV^*(h)$ is defined by

$$CV^*(h) = g^{-1} \sum_{k=1}^{g} CV_k(h).$$

Finally, h denotes the minimizer of the function expressed by its subscript or superscript. For example, $\hat{h}_{CVk}$ and $\hat{h}_{CV}^*$ are the minimizers of $CV_k(h)$ and $CV^*(h)$ respectively.

Now we start to derive the MASE of each subgroup of observations. Given the assumptions above, the definition of $d_M(h)$, and the results of (4.2.5) and (4.4.22), we have, as $n \longrightarrow \infty$,

$$(5.2.1) \qquad d_M^*(h) = d_{M1}(h) = d_{M2}(h) = \ldots = d_{Mg}(h)$$

$$= a_{1g}\eta^{-1}h^{-1} + b_1h^4 + b_2h^6 + o(h^6),$$

which have these minimizers respectively,

$$(5.2.2) \quad h_M^* = h_{M1} = h_{M2} = \ldots = h_{Mg} = C_{0g}\eta^{-1/5} + B_{0g}\eta^{-3/5} + o(\eta^{-3/5}),$$

where

$$a_{1g} = (\sum_{k=-\infty}^{\infty} \gamma(gk))\int K^2 \int W,$$

$$b_1 = (1/4)(\int u^2K)^2 \int (m'')^2 W,$$

$$b_2 = (-1/24)\int u^2K\int u^4K(\int (m^{(3)})^2W+\int m''m^{(3)}W').$$

$$C_{0g} = [a_{1g}(4b_1)^{-1}]^{1/5} = [\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2\int W(\int u^2K)^{-2}(\int (m'')^2W)^{-1}]^{1/5}.$$

$$B_{0g} = (1/20)[(\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2\int W)^{3/5}(\int u^4K)(\int (m^{(3)})^2W+\int m''m^{(3)}W')] / [(\int u^2K)^{11}(\int (m'')^2W)^8]^{1/5}.$$

and where the subscript g denotes that the notation apply to the PCV with g subgroups of observations. Here and throughout this chapter, the notation $\int$ denotes $\int du$. The term $a_{1g}\eta^{-1}h^{-1}$ is an asymptotic representation of the variance of the MASE. Looking at coefficients $a_{1g}$ and $a_1$ as given in (3.3.1), the PCV would not reflect the dependence structure of the whole data set.

Now we start to derive the asymptotic properties of the ordinary cross-validated bandwidth. Given the assumptions above, the definition of $d_M^S(h)$ as given in (3.4.5), and the results (4.2.6) and (4.2.22), we have, as $n \longrightarrow \infty$,

$$d_M^{S*}(h) = d_{M1}^S(h) = d_{M2}^S(h) = \dots = d_{Mg}^S(h) = a_{1g}^S\eta^{-1}h^{-1} + b_1h^4 + b_2h^6 + o(h^6), \tag{5.2.3}$$

which have these minimizers respectively,

$$h_M^{S*} = h_{M1}^S = h_{M2}^S = \dots = h_{Mg}^S = C_{0g}^S\eta^{-1/5} + B_{0g}^S\eta^{-3/5} + o(\eta^{-3/5}), \tag{5.2.4}$$

where

$$a_{1g}^S = [\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2 - 4\sum_{k>0} \gamma(gk)K(0)]\int W.$$

$$C^S_{0g} = [a^S_{1g}(4b_1)^{-1}]^{1/5}$$

$$= [(\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2 - 4\sum_{k>0} \gamma(gk)K(0))\int W(\int u^2K)^{-2}(\int(m'')^2W)^{-1}]^{1/5},$$

$$B^S_{0g} = (1/20)[(\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2 - 4\sum_{k>0} \gamma(gk)K(0))^{3/5}(\int W)^{3/5}(\int u^4K)\cdot$$

$$(\int(m^{(3)})^2W + \int m''m^{(3)}W')] \,/\, [(\int u^2K)^{11}(\int(m'')^2W)^8]^{1/5},$$

and where the superscript S represents subtraction. Here, $a^S_{1g}$ must be positive, otherwise $d^{S*}_M(h)$, $d^S_{M1}(h)$, ..., $d^S_{Mg}(h)$ have no minimizers on $H_{n,g}$. Using (3.4.7), the ordinary cross-validation score of the k-th subgroup of observations $CV_k(h)$ has an asymptotic expression

(5.2.5) $$CV_k(h) = \eta^{-1}\sum_{j=0}^{\eta-1} \epsilon_{jg+k}^2W(x_{jg+k}) + d^S_{Mk}(h) + o_u(\eta^{-1}h^{-1}+h^4),$$

for each k = 1, 2, ..., g. The notation $X_n = o_u(v_n)$ denotes that $X_n/v_n$ converges to zero almost surely, and uniformly on $H_{n,g}$ if $v_n$ involves h. Using (3.4.9), we have, as $\eta \longrightarrow \infty$ and $C^S_{0g} > 0$,

(5.2.6) $$\hat{h}_{CVk}/h^S_{Mk} \longrightarrow 1 \text{ a.s.}$$

for every k. Taking the average of $CV_k(h)$, we have

(5.2.7) $$CV^*(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j^2W(x_j) + d^{S*}_M(h) + o_u(\eta^{-1}h^{-1}+h^4).$$

Then, by (3.4.7) and (3.4.9), we have, as $\eta \longrightarrow \infty$ and $C^S_{0g} > 0$,

(5.2.8) $$\hat{h}^*_{CV}/h^{S*}_M \longrightarrow 1 \text{ a.s.}$$

If $C^S_{0g}$ is negative or zero, then $CV^*(h)$ has no minimizer on $H_{n,g}$ asymptotically.

Thus having looked at these asymptotic relationships of the ordinary cross-validated bandwidths, we are now ready to define the partitioned cross-validated bandwidth. Since the optimal bandwidth $h_M$ is of the order $n^{-1/5}$ as given in (4.2.5) and $\hat{h}^*_{CV}$ is of the order $\eta^{-1/5} = n^{-1/5}g^{1/5}$ as given in (5.2.8), then the partitioned cross-validated bandwidth is defined as the rescaled $\hat{h}^*_{CV}$

$$(5.2.9) \qquad \hat{h}_{PCV(g)} = \hat{h}^*_{CV} g^{-1/5}.$$

Then, (5.2.4), (5.2.8), and (5.2.9) imply that

$$(5.2.10) \quad \hat{h}_{PCV(g)}/h_M = [C^S_{0g}/C_0] + [B^S_{0g}/C_0]n^{-2/5}g^{2/5} + o_u(n^{-2/5}g^{2/5}).$$

When g = 1, the partitioned cross-validated bandwidth is equal to the ordinary cross-validated bandwidth and the modified cross-validated bandwidth with $\ell = 0$, i.e. $\hat{h}_{PCV(1)} = \hat{h}_{CV} = \hat{h}_{CV(0)}$. However as $g \longrightarrow \infty$ with $g \ll n^{1/2}$, we have

$$C^S_{0g}/C_0 \longrightarrow [\gamma(0) \; / \sum_{k=-\infty}^{\infty} \gamma(k)]^{1/5}.$$

This means that the partitioned cross-validation criterion could not provide an asymptotically optimal bandwidth estimate for the measure MASE, no matter how large the vaule of g is.

## 5.3 Asymptotic Normality

The object of the PCV is to improve the rate of convergence of the cross-validated bandwidth. However, the modified cross-validation puts the emphasis on the reduction of the asymptotic bias of bandwidth estimates. In Section 5.2 we showed that the PCV could not provide asymptotically optimal bandwidths when the observations are dependent. It is still worth calculating the limiting distribution of $\hat{h}_{PCV(g)}/h_M$.

Based on the assumptions given in Section 5.2, it is shown in

Section 5.5 that, as $n \longrightarrow \infty$, $C^S_{0g} > 0$ and $g = o(n^{1/2})$,

$$(5.3.1) \qquad g^{2/5} n^{1/10} (\hat{h}_{PCV(g)}/h_M - C^S_{0g}/C_0)$$

$$\Rightarrow N(0,\ v_g (\sum_{k=-\infty}^{\infty} \gamma(k))^{-2/5} Var_M),$$

where

$$v_g = (\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \gamma(j)\gamma(j-ig))(\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2 - 4\sum_{k>0} \gamma(gk)K(0))^{-7/5},$$

$$Var_M = (8/25)\int (K*(K-L)-(K-L))^2 \int W^2 \ /\ [(\int K^2)^9 (\int W)^9 (\int u^2 K)^2 \int (m'')^2 W]^{1/5},$$

for $L(u) = -uK'(u)$ and $*$ meaning convolution. Based on (5.3.1), $d_M(\hat{h}_{PCV(g)})/d_M(h_M)$ has a noncentral $\chi_1^2$ distribution.

## 5.4 Discussion

First of all, the results obtained in this chapter are still true when the regression errors are a linear process as defined in Section 4.4. Pertaining to the PCV and the $CV_\ell(h)$, in (4.2.3) and (5.3.1), we know that $\hat{h}_{PCV(g)}$ has a faster relative rate of convergence than $\hat{h}_{CV(\ell)}$. However, as derived in (5.3.1), $\hat{h}_{PCV(g)}/h_M - 1$ suffers from an asymptotic bias $C^S_{0g}/C_0 - 1$ which converges to $(\gamma(0)\ /\ \sum_{k=-\infty}^{\infty} \gamma(k))^{1/5} - 1$ at a polynomial rate as $g \longrightarrow \infty$. This asymptotic bias is caused by the distance g/n among the observations of each subgroup. This means that the PCV could not reflect the dependence structure of the observations. An immediate remedy for this asymptotic bias is to split the observations into g subgroups by taking every g-th cluster. Each cluster is composed of $\zeta$ consecutive observations. Thus, the PCV would be able to reflect the dependence structure of the observations. Since the autocovariance function of ARMA regression errors is geometrically bounded (the property (d) of Section 3.2), it is enough to take the value of $\zeta$ as $c(\log n)$ for some constant c. A drawback to this approach is that it requires too many observations.

Looking at (5.3.1), as $n \longrightarrow \infty$, $C^S_{0g} > 0$, and $g \ll n^{1/2}$, the asymptotic mean square error (AMSE) of $\hat{h}_{PCV(g)}/h_M - 1$ is

(5.4.1) $$\mathrm{AMSE}(\hat{h}_{PCV(g)}/h_M - 1)$$

$$= (C^S_{0g}/C_0 + B^S_{0g}/C_0 \cdot n^{-2/5} g^{2/5} - 1)^2 + g^{-4/5} n^{-1/5} v_g (\sum_{k=-\infty}^{\infty} \gamma(k))^{-2/5} \mathrm{Var}_M.$$

Since the autocovariance function of an ARMA process is geometrically bounded, then the coefficients in (5.4.1) have the following polynomial rates of convergence, as $g \longrightarrow \infty$ and $g = O(\log n)$:

$$C^S_{0g}/C_0 - 1 \longrightarrow A,$$

$$B^S_{0g}/C_0 \longrightarrow B,$$

$$v_g (\sum_{k=-\infty}^{\infty} \gamma(k))^{-2/5} \mathrm{Var}_M \longrightarrow V,$$

where

$$A = [\gamma(0) \;/\; \sum_{k=-\infty}^{\infty} \gamma(k)]^{1/5} - 1,$$

$$B = [\gamma(0)^3 \;/\; \sum_{k=-\infty}^{\infty} \gamma(k)]^{1/5} \cdot b,$$

$$V = (\sum_{k=-\infty}^{\infty} \gamma(k)^2)[(\sum_{k=-\infty}^{\infty} \gamma(k))^2 \gamma(0)^7]^{-1/5} \cdot v,$$

and where

$$b = (1/20)[(\int K^2 \int W)^{2/5} (\int u^4 K)(\int (m^{(3)})^2 W + \int m'' m^{(3)} W')] \;/$$
$$[(\int u^2 K)^9 (\int (m'')^2 W)^7]^{1/5},$$

$$V = [\mathrm{Var}_M \,/\, (\int K^2)^{7/5}].$$

Thus, as $g = O(\log n)$, this AMSE can be expressed as

(5.4.2) $$\mathrm{AMSE}(\hat{h}_{PCV(g)}/h_M - 1) = [A + Bn^{-2/5}g^{2/5}]^2 + Vn^{-1/5}g^{-4/5},$$

where B and V are positive quantities. If A is negative, then this AMSE decreases as g increases. This is because, asymptotically, A is the dominant term in the asymptotic bias. If A is positive, then this AMSE may be minimized over g giving the minimizer

(5.4.3) $$g_0 = [V^5 n \,/\, (A^5 B^5)]^{1/6} - (5/6)[V^7 \,/\, (A^{13} Bn)]^{1/6} + o(n^{-1/6}).$$

Note that while this result is of theoretical interest, the practical applicability of it is somewhat hampered by the fact that it depends on the unknowns, $\gamma(0)$, $\sum_{k=-\infty}^{\infty} \gamma(k)$, $\int (m^{(3)})^2 W$, and $\int (m'')^2 W$. For this choice of $g_0$, the asymptotic expression of (5.4.2) implies that

(5.4.4) $$[\hat{h}_{PCV(g_0)}/h_M] - [\gamma(0) \,/\, \sum_{k=-\infty}^{\infty} \gamma(k)]^{1/5} \sim n^{-1/6}$$

which provides an improvement over (4.2.3). Here the notation ~ means that the random variable on the left side has an asymptotic normal distribution, as $n \longrightarrow \infty$, when normalized by the sequence on the right side. In Section 6.3, we do a simulation study to show the performance of $\hat{h}_{PCV(g)}$ when the value of $g_0$ is decided by the plug-in approach.

We shall use two examples of MA(1) and AR(1) processes of regression errors, which have been given in Sections 3.5 and 4.3, to discuss how the asymptotic bias-square and the asymptotic variance depend on the value of g. For the MA(1) process of regression errors, we have, through a straightforward calculation,

$$C^S_{0g}/C_0 = \begin{cases} [1 - (4K(0)/\int K^2)\theta(1+\theta)^{-2}]^{1/5} & \text{if } g = 1 \\ [(1+\theta^2)(1+\theta)^{-2}]^{1/5} & \text{if } g \geq 2 \end{cases}.$$

$$B^S_{0g}/C_0 = \begin{cases} (1+\theta)^{4/5}[1-(4K(0)/\int K^2)\theta(1+\theta)^{-2}]^{3/5}\cdot\sigma^{4/5}b & \text{if } g = 1 \\ (1+\theta^2)^{3/5}(1+\theta)^{-2/5}\cdot\sigma^{4/5}b & \text{if } g \geq 2 \end{cases}.$$

$$v_g = \begin{cases} (1+4\theta+6\theta^2+4\theta^3+\theta^4)\sigma^{6/5}((1+\theta)^2\int K^2 - 4\theta K(0))^{-7/5} & \text{if } g = 1 \\ (1+6\theta^2+\theta^4)\sigma^{6/5}((1+\theta^2)\int K^2)^{-7/5} & \text{if } g = 2, \\ (1+4\theta^2+\theta^4)\sigma^{6/5}((1+\theta^2)\int K^2)^{-7/5} & \text{if } g \geq 3 \end{cases}$$

where $\theta$ and $\sigma^2$ are the MA(1) parameters, and $\theta \neq -1$. For any sensible choice of kernel function K, the maximum value of K is assumed at 0, then we have

$$\int K^2 < K(0)\int K = K(0) < 2K(0).$$

Looking at the above coefficients $C^S_{0g}/C_0$, $B^S_{0g}/C_0$, and $v_g$, the effect of the dependent observations on the asymptotic bias-square and the asymptotic variance becomes constant when $g \geq 3$. This is because the observations of each subgroup are independent. However, the minimum value of the AMSE could possibly be at the large value of g. This is because the asymptotic variance decreases with a rate $g^{-4/5}$, and the asymptotic bias-square increases or decreases with a rate $g^{2/5}$. For example, if $\theta > 0$ and $C^S_{0g} > 0$, then

$$0 < C^S_{01}/C_0 < C^S_{0g}/C_0 < 1 \quad \text{for all } g \geq 2,$$

$$0 < B^S_{01}/C_0 < B^S_{0g}/C_0 \quad \text{for all } g \geq 2,$$

$$v_g < v_2 < v_1 \quad \text{for all } g \geq 3.$$

In this case, the minimum value of the AMSE might be at any $g \geq 1$ for the different combinations of the value of the factors n, K, b, v, $\theta$, and $\sigma^2$. If $\theta < 0$ and $C^S_{0g} > 0$, then

$$1 < C^S_{0g}/C_0 < C^S_{01}/C_0 \quad \text{for all } g \geq 2,$$

$$0 < B^S_{0g}/C_0 < B^S_{01}/C_0 \quad \text{for all } g \geq 2,$$

$$v_1 < v_g < v_2 \quad \text{for all } g \geq 3.$$

In this case, there are some values of K, n, and $Var_M$ for which the minimum value of the AMSE might be at $g = 1$. However, these are rare cases. Generally, in the MA(1) case, we prefer to take the value of g as large as possible no matter what the values of $\theta$ and $\sigma^2$ are. Of course, the value of g is restricted to $g << n^{1/2}$.

For the AR(1) process of regression errors, we have, through a straightforward calculation,

$$C^S_{0g}/C_0 = [(1-\phi)(1+\phi)^{-1}r_{1g}]^{1/5},$$

$$B^S_{0g}/C_0 = [\sigma^4(1-\phi)^{-1}(1+\phi)^{-3}]^{1/5}r_{1g}{}^{3/5}\cdot b,$$

$$v_g = [\sigma^6(1-\phi^2)^{-3}(\int K^2)^{-7}]^{1/5}r_{2g}{}^{-7/5},$$

where

$$r_{1g} = 1 - [(4K(0)/\int K^2) - 2]\phi^g(1-\phi^g)^{-1},$$

$$r_{2g} = (1+\phi^2)(1-\phi^2)^{-1}(1+\phi^g)(1-\phi^g)^{-1} + 2g\phi^g(1-\phi^g)^{-2},$$

and where the meaning of the notation b has been given above. If $\phi > 0$ and $C^S_{0g} > 0$, then $r_{1g}$ increases to 1 and $r_{2g}$ decreases (when $g \geq -(\log \phi)^{-1}$), as g increases. This implies both of the asymptotic bias-square and the asymptotic variance decrease. In this case, the minimum value of the AMSE is at the large value of g, with $g \ll n^{1/2}$. If $\phi < 0$ and $C^S_{0g} > 0$, then the cases of even and odd are discussed separately. When g increases as an even number, the results are the same as in the case of $\phi > 0$. This is because $\phi^g = (\phi^2)^k$ for the even number g and some k. When g increases as an odd number, using (5.4.3), then the minimum value of the AMSE is at

$$g_0 = \{[(1-\phi)^{2/5}(1+\phi^2)(1-\phi^2)^{-1}v] \ / \ [\sigma^{2/5}((\frac{1-\phi}{1+\phi})^{1/5} - 1)b]\}^{5/6} \cdot n^{1/6}.$$

## 5.5 Proofs

Proof of (5.3.1):

We shall give the outline of the proof of (5.3.1) first. The details are given later. Let $t_n = \eta^{-1}h^{-1} + h^4$. The meaning of the notation with a superscript * or a subscript k has been described in Section 5.2. Based on the assumptions as given in Section 5.2, as $\eta \longrightarrow \infty$, $C^S_{0g} > 0$, and $g << n^{1/2}$, it is shown later that

(5.5.1) $$S^*_3(\beta\eta^{-1/5}) = o_p(g^{-1/2}\eta^{-9/10}),$$

(5.5.2) $$g^{1/2}\eta^{9/10}(-S^*_1-S^*_2+T^*_1+T^*_2)(\beta\eta^{-1/5}) \Rightarrow N(0,P_{ST}(g,\beta)),$$

for any $\beta \in [a,b]$, where

$$P_{ST}(g,\beta) = 2\beta^{-1}(\sum_{p=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\gamma(t)\gamma(t-pg))\int W^2\int(K*(K-L) - (K-L))^2.$$

The implications of (5.5.1) and (5.5.2) are, as $\eta \longrightarrow \infty$, for any $\beta \in [a,b]$, and $h^{S*}_M = C^S_{0g}\eta^{-1/5}(1+o(1))$,

(5.5.3) $$g^{1/2}\eta^{7/10}(D^*+\delta^*)'(\beta\eta^{-1/5}) \Rightarrow N(0,4\beta^{-2}P_{ST}(g,\beta)),$$

(5.5.4) $$g^{1/2}\eta^{7/10}(D^*+\delta^*)'(h^{S*}_M) \Rightarrow N(0,4(C^S_{0g})^{-2}P_{ST}(g,C^S_{0g})).$$

Based on the assumptions given above, we have the following asymptotic properties (5.5.5) through (5.5.10) which are shown later:

For each positive integer k, there is a constant $C_4$, so that

(5.5.5) $$\sup_{h\in H_{n,g}} E(|g^{1/2}t_n^{-1}h^{1/2}(D^*+\delta^*)'(h)|^{2k}) \le C_4,$$

furthermore, there is an $\epsilon > 0$ and a constant $C_5$, so that

(5.5.6) $$E(|g^{1/2}t_n^{-1}h^{1/2}[(D^*+\delta^*)'(h) - (D^*+\delta^*)'(h_1)]|^{2k}) \le C_5|(h-h_1)/h|^{\epsilon k},$$

whenever $h, h_1 \in H_{n,g}$ with $h \le h_1$. For any $\rho \in (0,1/10)$, we have

(5.5.7) $$\sup_{h\in H_{n,g}} |g^{1/2}t_n^{-1}h^{1/2}(D^*+\delta^*)'(h)| = o_p(n^{\rho}),$$

furthermore, if $\eta^{1/5}h_1$ tends to a constant, then we have

(5.5.8) $\sup_{|h-h_1|\le\eta^{-3/10+\rho}} |g^{1/2}t_n^{-1}h^{1/2}[(D^*+\delta^*)'(h) - (D^*+\delta^*)'(h_1)]|$

$= o_p(1)$,

(5.5.9) $|\hat{h}^*_{CV} - h^{S*}_M| = o_p(g^{-1/2}\eta^{-3/10+\rho})$.

Finally, we have

(5.5.10) $\text{Remainder}^{*\prime}(h) = o_u(g^{-1/2}\eta^{-7/10})$.

Now we derive a limiting distribution for $\hat{h}_{PCV(g)}/h_M$. According to (4.4.21), the average of all ordinary cross-validation scores $CV^*(h)$ can be expressed as

(5.5.11) $$CV^*(h) = n^{-1}\sum_{j=1}^{n} \epsilon_j^2 W(x_j) + d^{S*}_M(h) + D^*(h) + \delta^*(h) + \text{Remainder}^*(h).$$

Taking $h = \hat{h}^*_{CV}$ in the first derivative of (5.5.11), we have

(5.5.12) $$0 = CV^{*\prime}(\hat{h}^*_{CV}) = (d^{S*}_M + D^* + \delta^* + \text{Remainder}^*)'(\hat{h}^*_{CV})$$
$$= (\hat{h}^*_{CV} - h^{S*}_M)d^{S*\prime\prime}_M(h^*) + (D^*+\delta^*)'(\hat{h}^*_{CV}) + \text{Remainder}^{*\prime}(\hat{h}^*_{CV}),$$

where $h^*$ lies inbetween $\hat{h}^*_{CV}$ and $h^{S*}_M$. Using $\hat{h}^*_{CV} = C^S_{0g}\eta^{-1/5}(1+o_u(1))$ as given in (5.2.4), and (5.2.8), we have

$$d^{S*\prime\prime}_M(h^*) = C^S_{2g}\eta^{-2/5}(1+o_u(1)),$$

where

$$C^S_{2g} = 2a^S_{1g}(C^S_{0g})^{-3} + 12b_1(C^S_{0g})^2$$
$$= 5[(\sum_{k=-\infty}^{\infty} \gamma(gk)\int K^2 - 4\sum_{k>0} \gamma(gk)K(0))^2(\int W)^2(\int u^2K)^6(\int (m'')^2W)^3]^{1/5}.$$

Using (5.5.9), for any $\rho \in (0,1/10)$, $\hat{h}^*_{CV} = h^{S*}_M + o_p(g^{-1/2}\eta^{-3/10+\rho})$, and using (5.5.8) (with $h_1 = h^{S*}_M = C^S_{0g}\eta^{-1/5}(1+o(1))$), we have

$$(D^*+\delta^*)'(\hat{h}^*_{CV}) = (D^*+\delta^*)'(h^{S*}_M) + o_p(g^{-1/2}\eta^{-7/10}).$$

Combining this result with (5.5.10), then (5.5.12) becomes

(5.5.13) $$0 = (\hat{h}^*_{CV} - h^{S*}_M)C^S_{2g}\eta^{-2/5}(1+o_u(1)) + (D^*+\delta^*)'(h^{S*}_M) + o_p(g^{-1/2}\eta^{-7/10}).$$

Next, combining (5.5.13) with $(D^*+\delta^*)'(h^{S*}_M) = O_p(g^{-1/2}\eta^{-7/10})$ as given in (5.5.4), then we have

(5.5.14) $$\hat{h}^*_{CV} - h^{S*}_M = O_p(g^{-1/2}\eta^{-3/10}).$$

Multiplying (5.5.13) by $g^{1/2}\eta^{7/10}$ and combining the result with (5.5.14), we have

(5.5.15) $$0 = g^{1/2}\eta^{3/10}(\hat{h}^*_{CV} - h^{S*}_M)C^S_{2g} + g^{1/2}\eta^{7/10}(D^*+\delta^*)'(h^{S*}_M) + o_p(1)$$
$$= g^{2/5}\eta^{1/10}[\hat{h}_{PCV(g)}/h_M - C^S_{0g}/C_0]C_0C^S_{2g} + g^{1/2}\eta^{7/10}(D^*+\delta^*)'(h^{S*}_M) + o_p(1),$$

where $\hat{h}_{PCV(g)} = \hat{h}^*_{CV}g^{-1/5}$ as given in (5.2.9), $h_M = C_0n^{-1/5}(1+o(1))$ as given in (3.3.6), and $h^{S*}_M = C^S_{0g}\eta^{-1/5}(1+o(1))$ as given in (5.2.4). Then, combining (5.5.15) and (5.5.4), the proof of (5.3.1) is complete.

The notation and the properties used to prove (5.5.1) through (5.5.10) are the same as (a), (b), (c), and (d) given in Section 4.4.

Proof of (5.5.1):

Using the expression of $A_{st}(h)$ as given in property (d) of Section 4.4, for each k = 1, 2, ..., g, we have

$$S_{3k}(h) = \sum_{s=0}^{\eta-1} A_{sg+k,sg+k}(h)\epsilon_{sg+k},$$

where

$$A_{sg+k,sg+k}(h) = \eta^{-2}h^{-1}(1+o(1))\int K(K-L)\cdot W(x_{sg+k}).$$

Thus

$$S^*_3(h) = g^{-1}\sum_{k=1}^{g} S_{3k}(h) = g^{-1}\eta^{-2}h^{-1}(1+o(1))\int K(K-L)\sum_{j=1}^{n} W(x_j)\epsilon_j.$$

Through a straightforward calculation, the second moment of $S^*_3(h)$ is

$$E(S^*_3(h)^2) = g^{-2}\eta^{-4}h^{-2}(1+o(1))(\int K(K-L))^2\cdot E(\sum_{j=1}^{n} W(x_j)\epsilon_j)^2$$
$$= O(g^{-2}\eta^{-4}h^{-2}\cdot n) = O(g^{-1}\eta^{-3}h^{-2}),$$

for $E(\sum_{j=1}^{n}\epsilon_j)^2 = O(n)$ as given in (3.6.3), $g \ll n^{1/2}$, and the boundedness of W. This implies $S^*_3(h) = o_p(g^{-1/2}\eta^{-9/10})$. The proof of

(5.5.1) is complete.

Proof of (5.5.2):

Using the expression of $B_s(h)$ as given in property (d) of Section 4.4, for each $k = 1, 2, \ldots, g$, we have

$$S_{2k}(h) = \sum_{s=0}^{\eta-1} B_{sg+k}(h)\epsilon_{sg+k},$$

where

$$B_{sg+k}(h) = -\eta^{-1}h^2(1+o(1))(\int u^2K)(m''W)(x_{sg+k}).$$

Thus

$$(5.5.16) \quad S_2^*(h) = g^{-1}\sum_{k=1}^{g} S_{2k}(h)$$
$$= -g^{-1}\eta^{-1}h^2(1+o(1))(\int u^2K)\sum_{j=1}^{n} (m''W)(x_j)\epsilon_j.$$

Using the same reason as $S_2^*(h)$, we have

$$(5.5.17) \quad T_2^*(h) = -g^{-1}\eta^{-1}h^2(1+o(1))(\int u^2K)\sum_{j=1}^{n} (m''W)(x_j)\epsilon_j.$$

Using the boundedness of $m''W$ and $\sum_{j=1}^{n} \epsilon_j = O_p(n^{-1/2})$ as given (3.6.3), we have

$$(5.5.18) \quad -S_2^*(h)+T_2^*(h) = o(1)\cdot O_p(g^{-1}\eta^{-1}h^2n^{1/2}) = o_p(g^{-1}\eta^{-9/10}).$$

For the expression of $S_1^*(h)$, using the $A_{st}(h)$ as given in property (d) of Section 4.4, we have

$$S_{1k}(h) = \sum_{s=0, s\neq t,}^{\eta-1} \sum_{t=0}^{\eta-1} A_{sg+k, sg+k}(h)(\epsilon_{sg+k}\epsilon_{tg+k} - \gamma(sg-tg)),$$

where

$$A_{sg+k, tg+k}(h) = \eta^{-2}h^{-1}(1+o(1))K*(K-L)(\frac{sg-tg}{nh})W(\frac{sg+k}{n}).$$

Thus

$$(5.5.19) \quad S_1^*(h) = g^{-1}\sum_{k=1}^{g} S_{1k}(h)$$

$$= g^{-1}\eta^{-2}h^{-1}(1+o(1)) \sum_{k=1}^{g} \sum_{s=0,s\neq t,t=0}^{\eta-1} \sum^{\eta-1} K*(K-L)(\frac{sg-tg}{nh})W(\frac{sg+k}{n})\cdot$$

$$(\epsilon_{sg+k}\epsilon_{tg+k} - \gamma(sg-tg))$$

$$= g^{-1}\eta^{-2}h^{-1}(1+o(1)) \sum_{p\neq 0} \sum_{s=1,s-t=pg,t=1}^{n} \sum^{n} K*(K-L)(\frac{pg}{nh})W(\frac{s}{n})(\epsilon_s\epsilon_t - \gamma(pg)).$$

Using the same reason as $S_1^*(h)$, we have

$$(5.5.20) \quad T_1^*(h) = g^{-1}\eta^{-2}h^{-1}(1+o(1)) \sum_{p\neq 0} \sum_{i=1,i-j=pg,j=1}^{n} \sum^{n} (K-L)(\frac{pg}{nh})W(\frac{i}{n})\cdot$$

$$(\epsilon_i\epsilon_j - \gamma(pg)).$$

Using (3.6.3), we have

$$(-S_1^*-S_2^*+T_1^*+T_2^*)(h)$$

$$= g^{-1}\eta^{-2}h^{-1} \sum_{p\neq 0} \sum_{s=1,s-t=pg,t=1}^{n} \sum^{n} (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\cdot$$

$$(\epsilon_s\epsilon_t - \gamma(pg)) + o_p(g^{-1/2}\eta^{-9/10})$$

$$= U + o_p(g^{-1/2}\eta^{-9/10}),$$

where

$$U = g^{-1}\eta^{-2}h^{-1} \sum_{p\neq 0} \sum_{s} (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})(\epsilon_s\epsilon_{s-pg} - \gamma(pg)).$$

Thus, using the same arguments as in the proof of (4.4.5), then the asymptotic normality of (5.5.2) is true. Then, the proof of (5.5.2) is complete by checking that its asymptotic variance is correct.

Using the linear expressions of $\epsilon_s$, $\epsilon_{s-pg}$, and $\gamma(pg)$:

$$\epsilon_s = \sum_{i=0}^{\infty} \psi_i e_{s-i}, \quad \epsilon_{s-pg} = \sum_{j=0}^{\infty} \psi_j e_{s-pg-j}, \quad \gamma(pg) = \mu_2 \sum_{i=0}^{\infty} \psi_i\psi_{i-pg}, \text{ and}$$

$$\epsilon_s\epsilon_{s-pg} - \gamma(pg) = \sum_{i=0}^{\infty} \psi_i\psi_{i-pg}(e_{s-i}{}^2 - \mu_2) + \sum_{j\neq i-pg}\sum \psi_i\psi_j e_{s-i}e_{s-pg-j},$$

then U can be decomposed into $U_A + U_B$, where

$$U_A = \sum_u a_u\xi_u,$$

$$U_B = \sum_{u\neq v}\sum b_{uv}e_ue_v,$$

and have

$$a_u = g^{-1}\eta^{-2}h^{-1}\sum_{p\neq 0}\sum_s (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\psi_{s-u}\psi_{s-pg-u},$$

$$b_{uv} = g^{-1}\eta^{-2}h^{-1}\sum_{p\neq 0}\sum_s (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\psi_{s-u}\psi_{s-pg-v},$$

for letting u=s-i, v=s-pg-j, $\xi_u = e_u^2 - \mu_2$. Thus we have

$$Var(U) = E(U_A^2) + E(U_B^2).$$

To calculate $E(U_A^2)$, using the boundedness of W and $(-K*(K-L)+(K-L))$, $\sum_{p\neq 0}\psi_{s-pg-u} = O(1)$ for any u, O(n) terms in $\sum_s$, $\sum_s\sum_{s'}\gamma(s-s') = O(n)$, and $\sum_u \psi_{s-u}\psi_{s'-u} = \mu_2^{-1}\gamma(s-s')$, then we have

$$E(U_A^2) \le c\sum_u a_u^2$$
$$\le c\sum_u [g^{-1}\eta^{-2}h^{-1}\sum_{p\neq 0}\sum_s (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\psi_{s-u}\psi_{s-pg-u}]^2$$
$$\le cg^{-2}\eta^{-4}h^{-2}\sum_u[\sum_s \psi_{s-u}]^2 = cg^{-2}\eta^{-4}h^{-2}\sum_u\sum_s\sum_{s'}\psi_{s-u}\psi_{s'-u}$$
$$\le cg^{-2}\eta^{-4}h^{-2}\cdot n = O(g^{-1}\eta^{-3}h^{-2}) = o(g^{-1}\eta^{-9/5}).$$

To calculate $E(U_B^2)$, using $b_{uv}=b_{vu}$ for any u, v, $\mu_2\sum_u \psi_{s-u}\psi_{s'-u} = \gamma(s-s')$, $\mu_2\sum_v \psi_{s-pg-v}\psi_{s'-p'g-v} = \gamma(s-s'-pg-p'g)$, then we have

$$E(U_B^2) = E(\sum_{u\neq v}\sum b_{uv}e_ue_v)^2 = 2\mu_2^2\sum_{u\neq v}\sum b_{uv}^2$$
$$= 2\mu_2^2\sum_{u\neq v}\sum[g^{-1}\eta^{-2}h^{-1}\sum_{p\neq 0}\sum_s (-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\psi_{s-u}\psi_{s-pg-v}]^2$$
$$= 2\mu_2^2 g^{-2}\eta^{-4}h^{-2}\sum_{u\neq v}\sum\sum_{p\neq 0}\sum_s\sum_{p'\neq 0}\sum_{s'}(-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\cdot$$
$$(-K*(K-L)+(K-L))(\frac{p'g}{nh})W(\frac{s'}{n})\psi_{s-u}\psi_{s-pg-v}\psi_{s'-u}\psi_{s'-p'g-v}$$
$$= 2g^{-2}\eta^{-4}h^{-2}\sum_{p\neq 0}\sum_s\sum_{p'\neq 0}\sum_{s'}(-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\cdot$$
$$(-K*(K-L)+(K-L))(\frac{p'g}{nh})W(\frac{s'}{n})\gamma(s-s')\gamma(s-s'-pg-p'g)$$
$$= 2g^{-2}\eta^{-4}h^{-2}\sum_{p\neq 0}\sum_s\sum_i\sum_j(-K*(K-L)+(K-L))(\frac{pg}{nh})W(\frac{s}{n})\cdot$$
$$(-K*(K-L)+(K-L))(\frac{pg-ig}{nh})W(\frac{s-j}{n})\gamma(j)\gamma(j-ig)$$

by letting i=p-p' and j=s-s',

$$= 2g^{-2}\eta^{-4}h^{-2}\sum_{p\neq 0}\sum_s\sum_i\sum_j(-K*(K-L)+(K-L))^2(\frac{pg}{nh})W(\frac{s}{n})^2\gamma(j)\gamma(j-ig) +$$

$$O(g^{-2}\eta^{-3}h^{-2}).$$

Since $(-K*(K-L)+(K-L))(\frac{pg-ig}{nh}) = (-K*(K-L)+(K-L))(\frac{pg}{nh}) + O(\frac{ig}{nh})$, $W(\frac{s-j}{n}) = W(\frac{s}{n}) + O(\frac{j}{n})$, $\sum_{j=-\infty}^{\infty} |j\gamma(j)| = O(1)$, $\sum_{i=-\infty}^{\infty} |ig\gamma(j-ig)| = O(|j|)$, the boundedness of $W$ and $(-K*(K-L)+(K-L))$, $O(\eta h)$ terms is $\sum_{p\neq 0}$, and $O(n)$ terms in $\sum_s$, then we have

$$E(U_B^2)$$
$$= 2g^{-1}\eta^{-2}h^{-1}\int(K*(K-L)-(K-L))^2\int W^2[\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\gamma(j)\gamma(j-ig)] + o(g^{-1}\eta^{-2}h^{-1}).$$

The proof of (5.5.2) is complete.

For the proofs of (5.5.5) through (5.5.10), using the expressions of $S_1^*(h)$, $S_2^*(h)$, $T_1^*(h)$, and $T_2^*(h)$ as given in (5.5.16) through (5.5.20), and following the same arguments given in the proofs of (4.4.8) through (4.4.13), through a straightforward calculation, then the proofs of (5.5.5) through (5.5.10) are complete.

CHAPTER VI

# A SIMULATION STUDY

## 6.1 Introduction

To investigate the practical implications of asymptotic results for bandwidth estimates presented in Chapters 3, 4, and 5, an empirical study is carried out. In this section, we shall describe the simulated regression setting. The regression model is taken as the equally spaced fixed design nonparametric regression model (1.1). The underlying regression function is taken to be $m(x) = x^3(1-x)^3$ for $0 \leq x \leq 1$. This function has the nice effect of allowing a circular design to eliminate boundary effects and has a uniformly continuous fourth derivative, a condition needed by (4.2.5). The regression errors are taken to be a first order autoregressive process, AR(1), i.e. $\epsilon_1 \approx N(0, \sigma^2/(1-\phi^2))$ and $\epsilon_i = \phi\epsilon_{i-1} + e_i$, for $i = 2, 3, \ldots, n$, where $e_i$ are independent and identically distributed (IID) $N(0, \sigma^2)$, and $\phi$ and $\sigma^2$ are the AR(1) parameters. The notation $A \approx B$ means that A and B have the same distribution. The observations are generated by adding the regression values and the regression errors together, i.e. $Y_i = m(x_i) + \epsilon_i$, for $i = 1, 2, \ldots, n$. Five combinations of $\phi$ and $\sigma$ ($\phi = 0$, $\sigma = 0.0177$; 0.6, 0.0071; 0.6, 0.0018; -0.6, 0.0283; -0.6, 0.0029) and the sample size $n = 200$ are investigated. Here the values of $\sigma$ make $h_M$ correspond roughly 1/5, 1/4, or 1/2. The kernel function is $K(x) = (15/8)(1-4x^2)^2 I_{[-1/2, 1/2]}(x)$. The weight function is $W(x) =$

$(5/3)I_{[1/5,4/5]}(x)$. The Nadaraya-Watson estimator (1.2) is applied to get the kernel estimates. The same functions K and m were also used in Rice (1984) and Haerdle, Hall, and Marron (1988). For each combination of $\phi$ and $\sigma$, 1000 independent sets of data are generated. All the normal random variables are generated by the function RNDNS in GAUSS where the seed is taken as 123.

For the modified cross-validation criterion $CV_{\ell}(h)$ as given in (3.1.1), the values of $\ell$ are 0, 1, 2, ..., 14. Recall that $2\ell+1$ observations are deleted in the construction of the cross-validated estimators. For the partitioned cross-validation criterion PCV as given in (5.2.10), the values of g are 1, 2, ..., 15. Recall that the observations are split into g subgroups by taking every g-th observation. The score function $d_M(h)$ as given in (1.12) is approximated by the average of the 1000 values of $d_A(h)$ as given in (1.11). The approximate expected value of the cross-validated bandwidth $\hat{h}_{CV(\ell)}$ is approximated by the minimizer of the average of the 1000 values of $CV_{\ell}(h)$ for each $\ell$. The approximate expected value of the partitioned cross-validated bandwidth $\hat{h}_{PCV(g)}$ is approximated by the product of $g^{-1/5}$ and the minimizer of the average of the 1000 values of $CV^*(h)$ for each g, where $CV^*(h)$ was defined in Section 5.2. The values of the score functions $d_A(h)$, $CV_{\ell}(h)$, and $CV^*(h)$ are calculated on an equally spaced logarithmic grid of 11 values. The endpoints of the grid are different for the different settings, and chosen to contain essentially all the bandwidths of interest. The minimizers, $h_A$, $h_M$, $\hat{h}_{CV(\ell)}$, and $\hat{h}^*_{CV}$ of the score functions, $d_A(h)$, $d_M(h)$, $CV_{\ell}(h)$, and $CV^*(h)$ respectively, are calculated. The partitioned cross-validation bandwidth $\hat{h}_{PCV(g)}$ was defined as $g^{-1/5}\hat{h}^*_{CV}$.

After evaluation on the grid, a one step interpolation improvement is done, with the results taken as the selected bandwidths. If the score functions had multiple minimizers on the grid, the algorithm chooses the smaller of them (this choice is made arbitrarily). The approximate expected values of $h_M$, $\hat{h}_{CV(\ell)}$, and $\hat{h}_{PCV(g)}$ could also be calculated by the formulas $C_0 n^{-1/5} + B_0 n^{-3/5}$ as given in (4.2.5), $C^S_{0\ell} n^{-1/5} + B^S_{0\ell} n^{-3/5}$ as given in (4.2.6), and $C^S_{0g} n^{-1/5} + B^S_{0g} n^{-3/5} g^{3/5}$ as given in (5.2.4) respectively. The coefficients B and C are derived by the following quantities:

$$\int K^2 = 10/7, \qquad \int u^2 K = 1/28, \qquad \int u^4 K \sim 0.002976, \qquad \int W^2 = 5/3,$$

$$\int (m'')^2 W \sim 0.05612, \qquad \int (m^{(3)})^2 W \sim 4.22126, \qquad \int m'' m^{(3)} W' = 0,$$

$$\int (K-L)^2 = 10/7, \quad \int (K*(K-L))^2 \sim 0.2079, \quad \int (K*(K-L) - (K-L))^2 \sim 0.7629,$$

where the notation $a \sim b$ denotes that b approximates a with the accuracy to the last decimal of b.

To show the structure of the population of bandwidth ratios, we use a kernel density estimate with the same kernel function

$$K(x) = (15/8)(1-4x^2)^2 I_{[-1/2,1/2]}(x)$$

as for the nonparametric regression, where $\hat{h}$ denotes $h_A$, $\hat{h}_{CV(\ell)}$, and $\hat{h}_{PCV(g)}$. The bandwidth h is taken as $1000^{-1/5}$. Then the density estimates are calculated by $(1/1000) \sum_{i=1}^{1000} K_h(x-X_i)$, where $X_i$ denotes the value of $\hat{h}/h_M$ of the i-th data set and $K_h(\cdot)$ denotes $h^{-1}K(\cdot/h)$.

Section 6.2 gives our simulation results. Section 6.3 gives the estimated values of the optimal value of the subgroups as given in (5.4.3).

## 6.2 Simulation Results

In this section, we shall give the sample mean square errors (MSE) and the kernel density estimates for the population of bandwidth ratios for each combination of $\phi$ and $\sigma^2$. The sample variances, the sample bias-squares, and the MSE of the bandwidth estimates $\hat{h}/h_M$ are summarized in Tables 6.1 through 6.5, where $\hat{h}$ denotes $h_A$, $\hat{h}_{CV(\ell)}$, and $\hat{h}_{PCV(g)}$. The bias-square of the bandwidth estimates is derived as the square of the average of the 1000 values of $\hat{h}/h_M - 1$. The MSE is the sum of the variance and the bias-square. Table 6.1 presents the results for the independent observations where $\sigma^2 = 0.0177^2$ and $h_M$ roughly equals 1/2. Table 6.2 contains the results for the positively correlated observations with a large amount of sample variability, where $\phi = 0.6$, $\sigma^2 = 0.0071^2$, and $h_M$ roughly equals 1/2. Table 6.3 gives the results for the positively correlated observations with a small amount of sample variability, where $\phi = 0.6$, $\sigma^2 = 0.0018^2$, and $h_M$ roughly equals 1/4. Table 6.4 contains the results for the negatively correlated observations with a large amount of sample variability, where $\phi = -0.6$, $\sigma^2 = 0.0283^2$, and $h_M$ roughly equals 1/2. Finally, Table 6.5 gives the results for the negatively correlated observations with a small amount of sample variability, where $\phi = -0.6$, $\sigma^2 = 0.0029^2$, and $h_M$ roughly equals 1/5.

Looking at Table 6.1, the case of the independent observations, the bias-squares of $\hat{h}_{CV(\ell)}/h_M$ and $\hat{h}_{PCV(g)}/h_M$ are in the same amount for all values of $\ell$ and g. As g increases, the variance of $\hat{h}_{PCV(g)}/h_M$ decreases. However, the variance of $\hat{h}_{CV(\ell)}/h_M$ keeps at the same amount as $\ell$ increases. In this case, the PCV is preferred to the $CV_\ell(h)$

Next we shall discuss the contents of the tables with the

Table 6.1: The sample mean square errors (MSE) of the ratios of the modified cross-validated bandwidths and the partitioned cross-validated bandwidths to the optimal bandwidth for the independent observations.

| Ratios | | Variance | Bias-square | MSE |
|---|---|---|---|---|
| $h_A/h_M$ | | 0.066793 | 0.000915 | 0.067707 |
| $\hat{h}_{CV(\ell)}/h_M$ | $\ell$ Value | | | |
| | 0 | 0.141572 | 0.000382 | 0.141954 |
| | 1 | 0.143701 | 0.000023 | 0.143724 |
| | 2 | 0.147578 | 0.000599 | 0.148177 |
| | 3 | 0.153700 | 0.001122 | 0.154822 |
| | 4 | 0.155854 | 0.002106 | 0.157960 |
| | 5 | 0.162344 | 0.004151 | 0.166495 |
| | 6 | 0.159727 | 0.004408 | 0.164135 |
| | 7 | 0.155042 | 0.003674 | 0.158716 |
| | 8 | 0.155480 | 0.003543 | 0.159023 |
| | 9 | 0.152513 | 0.003024 | 0.155536 |
| | 10 | 0.147163 | 0.002397 | 0.149560 |
| | 11 | 0.149591 | 0.003348 | 0.152939 |
| | 12 | 0.150184 | 0.001787 | 0.151971 |
| | 13 | 0.143045 | 0.000368 | 0.143413 |
| | 14 | 0.140391 | 0.000324 | 0.140714 |
| $\hat{h}_{PCV(g)}/h_M$ | g Value | | | |
| | 1 | 0.142141 | 0.000472 | 0.142613 |
| | 2 | 0.122478 | 0.000189 | 0.122667 |
| | 3 | 0.104857 | 0.001186 | 0.106043 |
| | 4 | 0.090875 | 0.000703 | 0.091578 |
| | 5 | 0.082542 | 0.000901 | 0.083443 |
| | 6 | 0.069547 | 0.001303 | 0.070850 |
| | 7 | 0.065137 | 0.000890 | 0.066027 |
| | 8 | 0.052551 | 0.001325 | 0.053876 |
| | 9 | 0.046748 | 0.001629 | 0.048377 |
| | 10 | 0.044378 | 0.000631 | 0.045009 |
| | 11 | 0.034843 | 0.001059 | 0.035902 |
| | 12 | 0.040050 | 0.000105 | 0.040155 |
| | 13 | 0.032656 | 0.000096 | 0.032753 |
| | 14 | 0.026781 | 0.000121 | 0.026902 |
| | 15 | 0.025813 | 0.000053 | 0.025866 |

Table 6.2: The sample mean square errors (MSE) of the ratios of the modified cross-validated bandwidths and the partitioned cross-validated bandwidths to the optimal bandwidth for the positively correlated observations with a large amount of sample variability.

| Ratios | | Variance | Bias-square | MSE |
|---|---|---|---|---|
| $h_A/h_M$ | | 0.067989 | 0.000839 | 0.068828 |
| $\hat{h}_{CV(\ell)}/h_M$ | $\ell$ Value | | | |
| | 0 | 0.015217 | 0.340901 | 0.356118 |
| | 1 | 0.094015 | 0.157793 | 0.251808 |
| | 2 | 0.129529 | 0.066091 | 0.195620 |
| | 3 | 0.142950 | 0.035485 | 0.178436 |
| | 4 | 0.144861 | 0.022526 | 0.167387 |
| | 5 | 0.150511 | 0.016509 | 0.167020 |
| | 6 | 0.152457 | 0.013700 | 0.166157 |
| | 7 | 0.153662 | 0.011340 | 0.165001 |
| | 8 | 0.152041 | 0.010281 | 0.162322 |
| | 9 | 0.154379 | 0.009622 | 0.164000 |
| | 10 | 0.148788 | 0.008326 | 0.157113 |
| | 11 | 0.148136 | 0.008082 | 0.156217 |
| | 12 | 0.145901 | 0.007954 | 0.153855 |
| | 13 | 0.142961 | 0.005565 | 0.148526 |
| | 14 | 0.143305 | 0.004422 | 0.147727 |
| $\hat{h}_{PCV(g)}/h_M$ | g Value | | | |
| | 1 | 0.014969 | 0.342923 | 0.357892 |
| | 2 | 0.051444 | 0.285233 | 0.336677 |
| | 3 | 0.079858 | 0.186568 | 0.266425 |
| | 4 | 0.082228 | 0.127764 | 0.209992 |
| | 5 | 0.075802 | 0.091330 | 0.167132 |
| | 6 | 0.072712 | 0.071990 | 0.144702 |
| | 7 | 0.065812 | 0.059694 | 0.125507 |
| | 8 | 0.063662 | 0.055051 | 0.118712 |
| | 9 | 0.059589 | 0.049922 | 0.109511 |
| | 10 | 0.056709 | 0.050466 | 0.107175 |
| | 11 | 0.054767 | 0.049093 | 0.103862 |
| | 12 | 0.054734 | 0.046613 | 0.101347 |
| | 13 | 0.052097 | 0.046844 | 0.098941 |
| | 14 | 0.046694 | 0.048023 | 0.094718 |
| | 15 | 0.045878 | 0.048859 | 0.094738 |

Table 6.3: The sample mean square errors (MSE) of the ratios of the modified cross-validated bandwidths and the partitioned cross-validated bandwidths to the optimal bandwidth for the positively correlated observations with a small amount of sample variability.

| Ratios | | Variance | Bias-square | MSE |
|---|---|---|---|---|
| $h_A/h_M$ | | 0.057907 | 0.000035 | 0.057942 |
| $\hat{h}_{CV(\ell)}/h_M$ | $\ell$ Value | | | |
| | 0 | 0.001158 | 0.424437 | 0.425595 |
| | 1 | 0.041539 | 0.279507 | 0.321046 |
| | 2 | 0.074666 | 0.144123 | 0.218790 |
| | 3 | 0.084220 | 0.081520 | 0.165740 |
| | 4 | 0.086808 | 0.052977 | 0.139785 |
| | 5 | 0.084461 | 0.039172 | 0.123633 |
| | 6 | 0.079699 | 0.031801 | 0.111500 |
| | 7 | 0.073712 | 0.027187 | 0.100900 |
| | 8 | 0.069808 | 0.025106 | 0.094913 |
| | 9 | 0.061553 | 0.018699 | 0.080252 |
| | 10 | 0.057085 | 0.015261 | 0.072346 |
| | 11 | 0.053950 | 0.012969 | 0.066919 |
| | 12 | 0.049806 | 0.011853 | 0.061659 |
| | 13 | 0.045128 | 0.009669 | 0.054797 |
| | 14 | 0.041218 | 0.007396 | 0.048614 |
| $\hat{h}_{PCV(g)}/h_M$ | g Value | | | |
| | 1 | 0.001068 | 0.418798 | 0.419865 |
| | 2 | 0.022676 | 0.378389 | 0.401064 |
| | 3 | 0.044861 | 0.241548 | 0.286410 |
| | 4 | 0.043792 | 0.152653 | 0.196446 |
| | 5 | 0.037051 | 0.107216 | 0.144268 |
| | 6 | 0.030271 | 0.085374 | 0.115645 |
| | 7 | 0.028279 | 0.077010 | 0.105289 |
| | 8 | 0.025144 | 0.070429 | 0.095573 |
| | 9 | 0.026751 | 0.068851 | 0.095602 |
| | 10 | 0.022914 | 0.066056 | 0.088970 |
| | 11 | 0.022240 | 0.062975 | 0.085215 |
| | 12 | 0.022638 | 0.066291 | 0.088929 |
| | 13 | 0.023159 | 0.068660 | 0.091819 |
| | 14 | 0.023357 | 0.066338 | 0.089695 |
| | 15 | 0.028926 | 0.074854 | 0.103780 |

Table 6.4: The sample mean square errors (MSE) of the ratios of the modified cross-validated bandwidths and the partitioned cross-validated bandwidths to the optimal bandwidth for the negatively correlated observations with a large amount of sample variability.

| Ratios | | Variance | Bias-square | MSE |
|---|---|---|---|---|
| $h_A/h_M$ | | 0.066622 | 0.000923 | 0.067545 |
| $\hat{h}_{CV(\ell)}/h_M$ | $\ell$ Value | | | |
| | 0 | 0.053738 | 0.480186 | 0.533925 |
| | 1 | 0.008799 | 0.369738 | 0.378538 |
| | 2 | 0.115370 | 0.167648 | 0.283019 |
| | 3 | 0.141201 | 0.132474 | 0.273675 |
| | 4 | 0.181069 | 0.023440 | 0.204509 |
| | 5 | 0.202484 | 0.026484 | 0.228968 |
| | 6 | 0.201588 | 0.000285 | 0.201873 |
| | 7 | 0.206957 | 0.008039 | 0.214996 |
| | 8 | 0.203060 | 0.000264 | 0.203324 |
| | 9 | 0.201329 | 0.002119 | 0.203449 |
| | 10 | 0.196773 | 0.000693 | 0.197466 |
| | 11 | 0.190999 | 0.001972 | 0.192973 |
| | 12 | 0.196088 | 0.000592 | 0.196680 |
| | 13 | 0.194194 | 0.000100 | 0.194294 |
| | 14 | 0.194244 | 0.000125 | 0.194368 |
| $\hat{h}_{PCV(g)}/h_M$ | g Value | | | |
| | 1 | 0.054103 | 0.499512 | 0.553615 |
| | 2 | 0.196045 | 0.147679 | 0.343724 |
| | 3 | 0.038372 | 0.185286 | 0.223658 |
| | 4 | 0.211217 | 0.004263 | 0.215481 |
| | 5 | 0.064431 | 0.054817 | 0.119249 |
| | 6 | 0.130354 | 0.002965 | 0.133319 |
| | 7 | 0.076233 | 0.017172 | 0.093404 |
| | 8 | 0.089382 | 0.004054 | 0.093436 |
| | 9 | 0.061838 | 0.007784 | 0.069622 |
| | 10 | 0.068736 | 0.002063 | 0.070798 |
| | 11 | 0.042000 | 0.005920 | 0.047920 |
| | 12 | 0.049226 | 0.001418 | 0.050644 |
| | 13 | 0.042437 | 0.001181 | 0.043618 |
| | 14 | 0.034366 | 0.001194 | 0.035560 |
| | 15 | 0.037513 | 0.000116 | 0.037630 |

Table 6.5: The sample mean square errors (MSE) of the ratios of the modified cross-validated bandwidths and the partitioned cross-validated bandwidths to the optimal bandwidth for the negatively correlated observations with a small amount of sample variability.

| Ratios | | Variance | Bias-square | MSE |
|---|---|---|---|---|
| $h_A/h_M$ | | 0.041937 | 0.000077 | 0.042014 |
| $\hat{h}_{CV(\ell)}/h_M$ | $\ell$ Value | | | |
| | 0 | 0.008722 | 0.333118 | 0.341840 |
| | 1 | 0.002672 | 0.246337 | 0.249008 |
| | 2 | 0.027080 | 0.103456 | 0.130537 |
| | 3 | 0.068115 | 0.082290 | 0.150405 |
| | 4 | 0.060550 | 0.016934 | 0.077483 |
| | 5 | 0.080834 | 0.006755 | 0.087589 |
| | 6 | 0.066284 | 0.004151 | 0.070435 |
| | 7 | 0.067686 | 0.000001 | 0.067687 |
| | 8 | 0.056633 | 0.005088 | 0.061721 |
| | 9 | 0.048331 | 0.004142 | 0.052473 |
| | 10 | 0.042212 | 0.011381 | 0.053593 |
| | 11 | 0.035176 | 0.011860 | 0.047036 |
| | 12 | 0.033291 | 0.021996 | 0.055287 |
| | 13 | 0.027471 | 0.028349 | 0.055820 |
| | 14 | 0.062268 | 0.039337 | 0.065606 |
| $\hat{h}_{PCV(g)}/h_M$ | g Value | | | |
| | 1 | 0.008956 | 0.366789 | 0.375746 |
| | 2 | 0.077615 | 0.173606 | 0.251221 |
| | 3 | 0.027660 | 0.253039 | 0.280670 |
| | 4 | 0.159741 | 0.001283 | 0.161025 |
| | 5 | 0.070814 | 0.148202 | 0.219017 |
| | 6 | 0.109601 | 0.045872 | 0.155472 |
| | 7 | 0.079275 | 0.107663 | 0.186938 |
| | 8 | 0.097323 | 0.067464 | 0.164787 |
| | 9 | 0.082095 | 0.102081 | 0.184175 |
| | 10 | 0.081028 | 0.089276 | 0.170304 |
| | 11 | 0.067858 | 0.099434 | 0.167292 |
| | 12 | 0.072630 | 0.094587 | 0.167218 |
| | 13 | 0.078190 | 0.094887 | 0.173077 |
| | 14 | 0.053247 | 0.088721 | 0.141969 |
| | 15 | 0.065906 | 0.093108 | 0.159014 |

dependent observations. When the observations are positively correlated or negatively correlated, the ordinary cross-validation always produces too small or too large bandwidth estimates respectively as mentioned in Chiu (1989a) and Hart(1987). In our simulated regression setting, the $CV_\ell(h)$ and the PCV still suffer from the dependence effect for the small values of $\ell$ and g, e.g. $\ell \leq 2$ and $g \leq 3$. Thus, the bandwidth estimates would be close to the right or the left bound of the bandwidth selection interval for the small values of $\ell$ and g. This implies that the bandwidth estimates would have small variances when $\ell$ and g are small.

In Tables 6.2 and 6.3, we give the results for the combinations of the AR(1) parameters $\phi = 0.6$, and $\sigma^2 = 0.0071^2$ and $\sigma^2 = 0.0018^2$. In these two cases, the bias-square of $\hat{h}_{CV(\ell)}/h_M$ decreases to 0 as $\ell$ increases. However, the bias-square of $\hat{h}_{PCV(g)}/h_M$ decreases to a nonzero constant. This is because the PCV could not reflect the actual dependence structure of the observations and the bias-square of $\hat{h}_{PCV(g)}$ depends on $\sigma^2$ as given in (5.4.2). In contrast to the bias-square, the variance of $\hat{h}_{CV(\ell)}/h_M$ stays the same for all $\ell$ and the variance of $\hat{h}_{PCV(g)}/h_M$ decreases monotonely as g increases. Thus, the MSE of $\hat{h}_{CV(\ell)}/h_M$ decreases for the decreasing of its bias-square, and the MSE of $\hat{h}_{PCV(g)}/h_M$ decreases for the decreasing of its variance.

In Table 6.2, variance is the dominant term in MSE. Thus using the PCV to reduce the variance of the bandwidth estimates would result in a smaller value of MSE than using the $CV_\ell(h)$ to reduce the bias-square of the bandwidth estimates. On the other hand, in Table 6.3, bias-square is the dominant term in MSE. In this case, using the $CV_\ell(h)$ to reduce the bias-square of the bandwidth estimates is better than using the PCV to reduce the variance of the bandwidth estimates.

The choice between the $CV_{\ell}(h)$ and the PCV should be made on the basis of which component, variance or bias-square, is the dominant term in MSE. In Section 6.3, we use the plug-in approach to estimate the unknown quantities for the regression setting. This method could be used to estimate the variance and the bias-square of $\hat{h}_{CV(\ell)}/h_M$ and $\hat{h}_{PCV(g)}/h_M$.

In Tables 6.4 and 6.5, we give the results for the combinations of the AR(1) parameters $\phi = -0.6$, and $\sigma^2 = 0.0283^2$ and $\sigma^2 = 0.0029^2$. The variance and the bias-square of $\hat{h}_{PCV(g)}/h_M$ decreases along even g's and odd g's separately. This is because $\phi^g = \phi^{2k}$ where $g = 2k$ for the even number of g and some k as discussed in Section 5.4. The conclusions for Tables 6.4 and 6.5 are the same as those for Tables 6.2 and 6.3.

Kernel density estimates of the ratios $h_A/h_M$, $\hat{h}_{CV(\ell)}/h_M$, $\hat{h}_{PCV(g)}/h_M$ are given in Figures 6.1 through 6.10, corresponding to the combinations of $\phi$ and $\sigma^2$ as given in Tables 6.1 through 6.5. For the density estimates of $\hat{h}_{CV(\ell)}/h_M$, the integers 1, 2, 3, and 4 correspond to the curves for the values of $\ell = 1, 6, 10$, and 14 respectively. For the density estimates of $\hat{h}_{PCV(g)}/h_M$, the integers 1, 2, 3, and 4 correspond to the curves for the values of $g = 3, 4, 9$, and 15 respectively. Here the values of $\ell$ and g are chosen arbitrarily. In all cases, the character A corresponds to the curve for $h_A/h_M$. The vertical lines are the ratios of the approximate expected values of $h_A$, $\hat{h}_{CV(\ell)}$, and $\hat{h}_{PCV(g)}$, to that of $h_M$, for the corresponding values of $\ell$ and g. Here the approximate expected values of $h_A$ and $\hat{h}_{CV(\ell)}$ are the minimizers of the average of 1000 curves $d_A(h)$ and $CV_{\ell}(h)$ respectively. Also the approximate expected value of $\hat{h}_{PCV(g)}$ is the multiplication of

Figure 6.1: Kernel density estimates of the ratios of the modified cross-validated bandwidths to the optimal bandwidth for the independent observations.

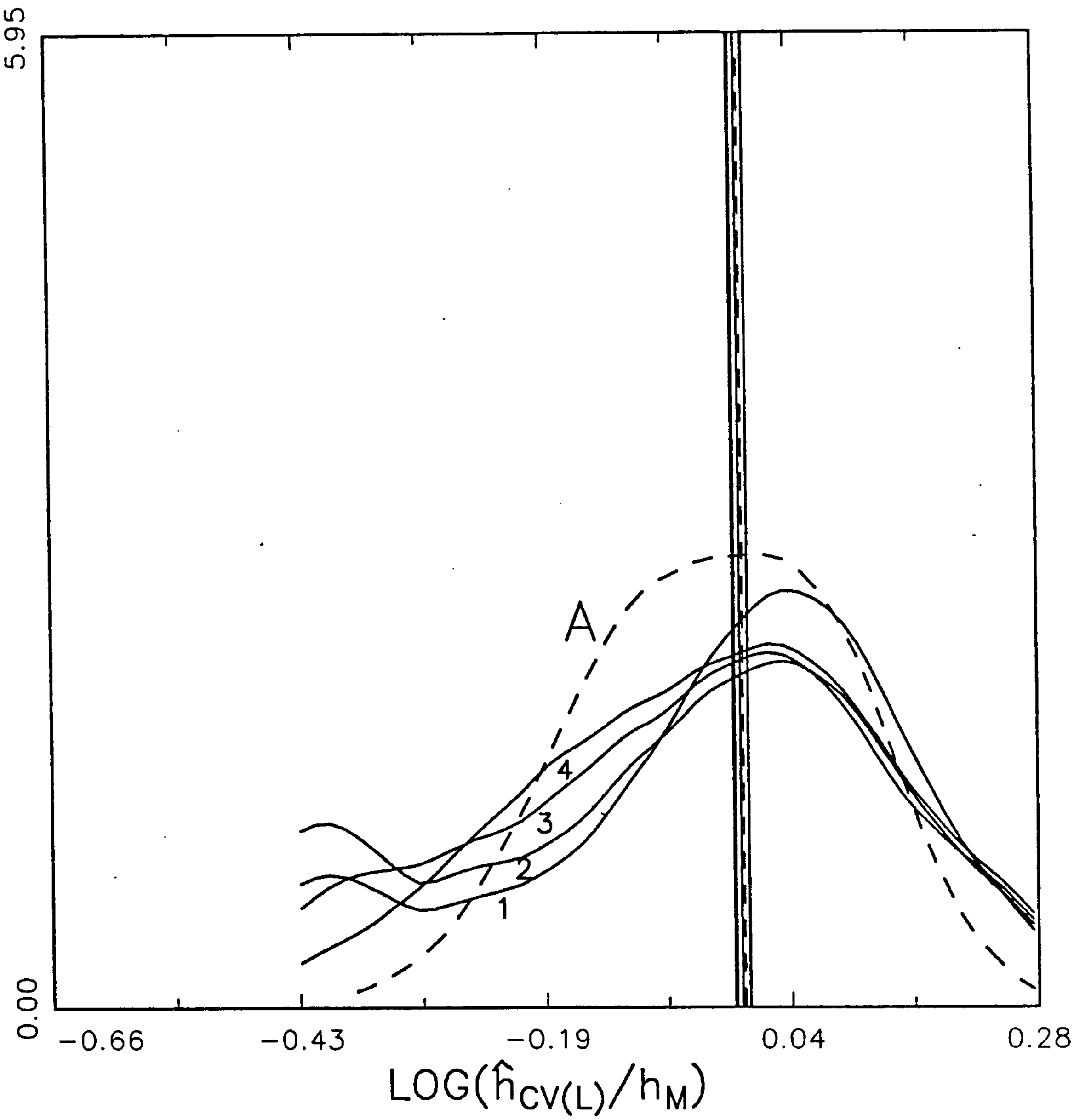


Figure 6.2: Kernel density estimates of the ratios of the partitioned cross-validated bandwidths to the optimal bandwidth for the independent observations.

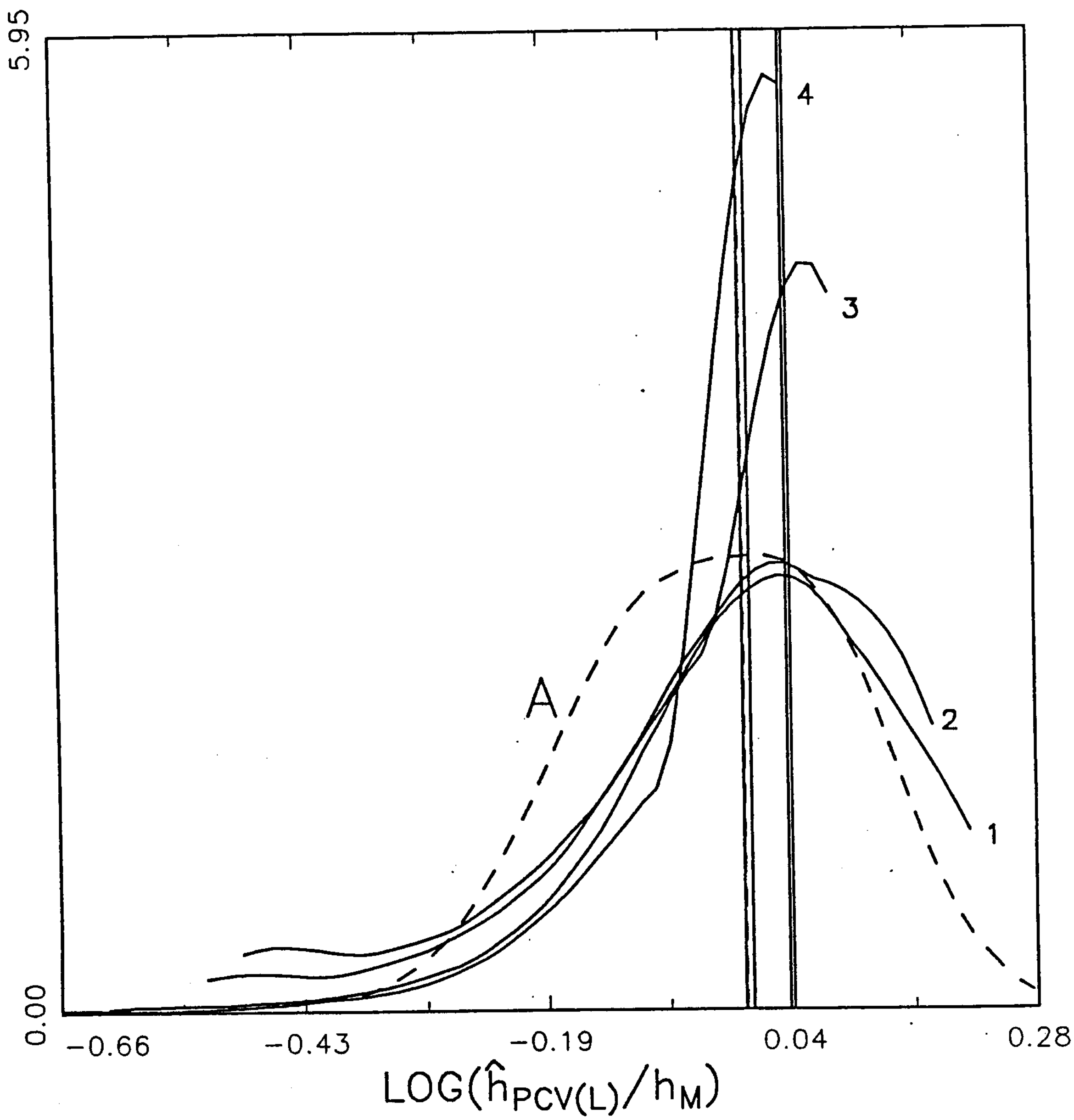


Figure 6.3: Kernel density estimates of the ratios of the modified cross-validated bandwidths to the optimal bandwidth for the positively correlated observations with a large amount of sample variability.

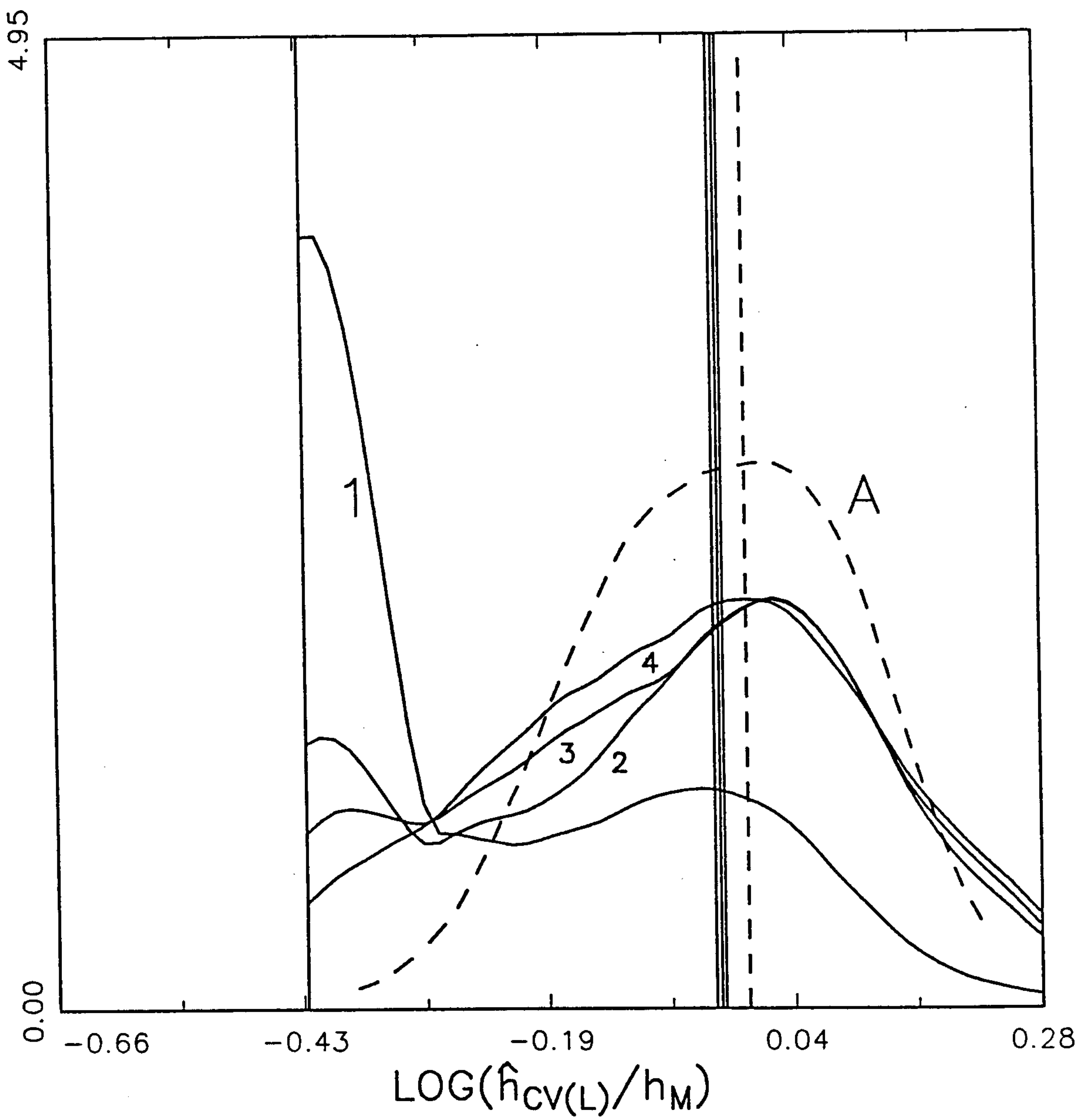


Figure 6.4: Kernel density estimates of the ratios of the partitioned cross-validated bandiwdths to the optimal bandwidth for the positively correlated observations with a large amount of sample variability.

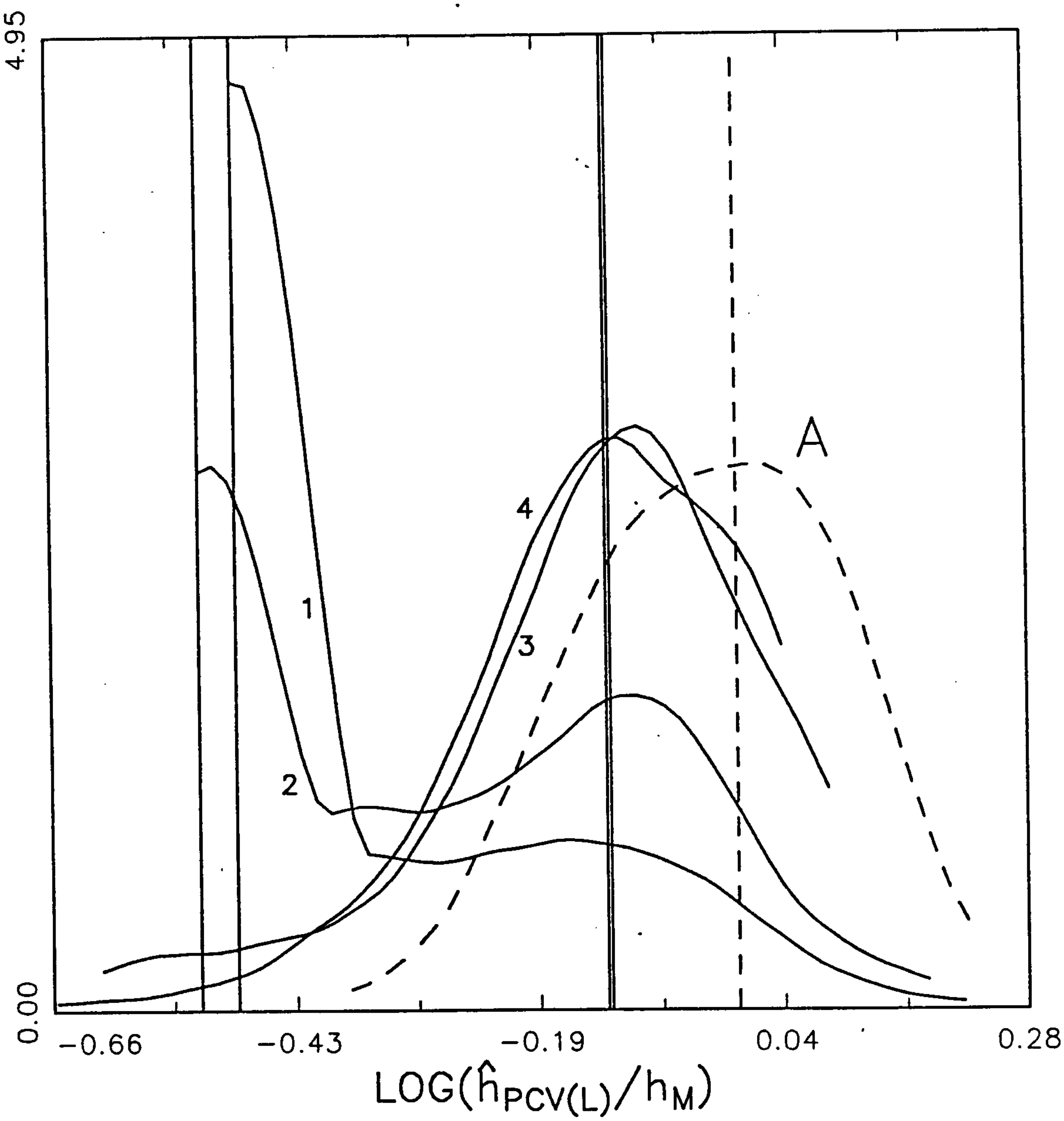


Figure 6.5: Kernel density estimates of the ratios of the modified cross-validated bandwidths to the optimal bandwidth for the positively correlated observations with a small amount of sample variability.

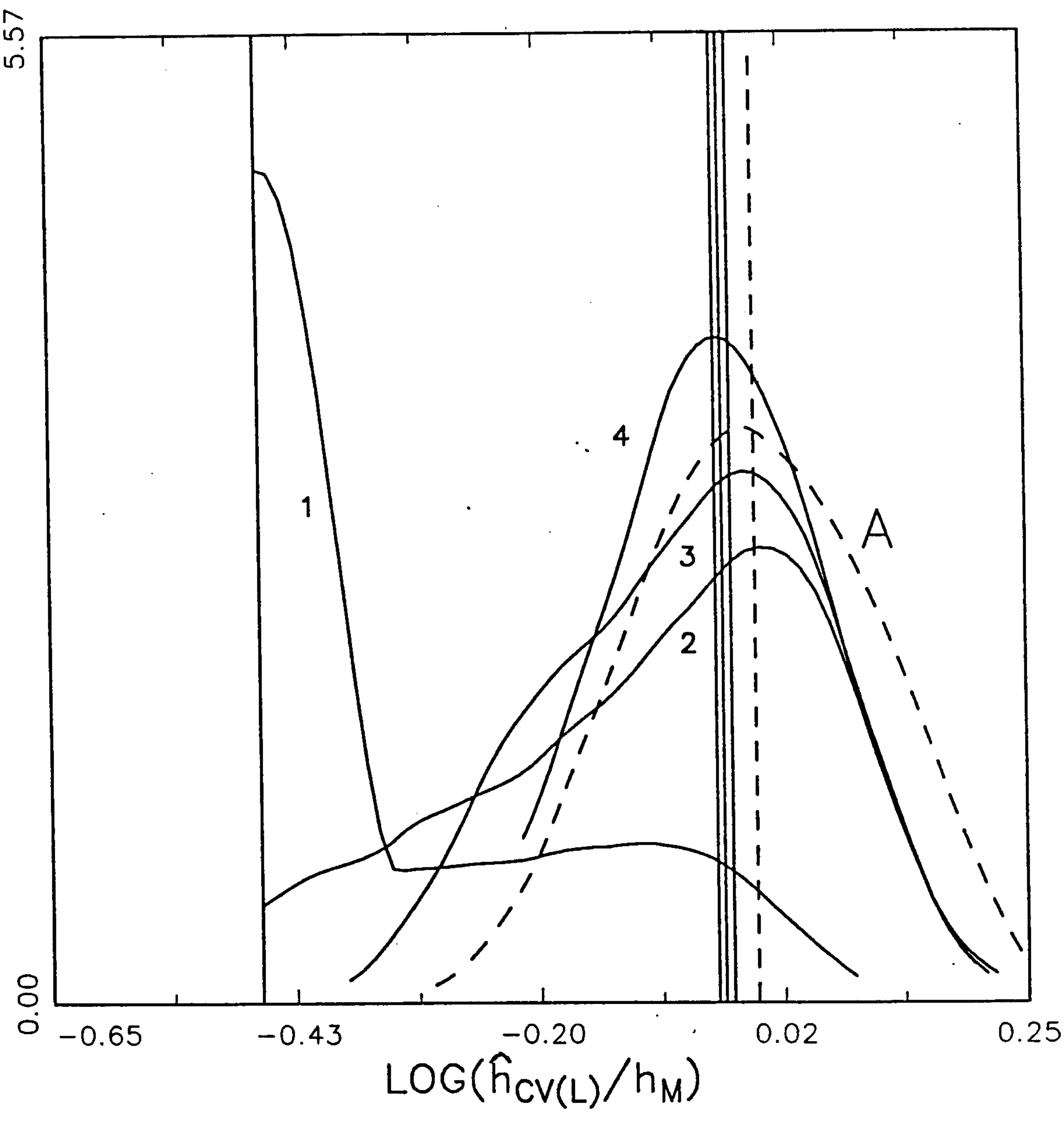


Figure 6.6: Kernel density estimates of the ratios of the partitioned cross-validated bandiwdths to the optimal bandwidth for the positively correlated observations with a small amount of sample variability.

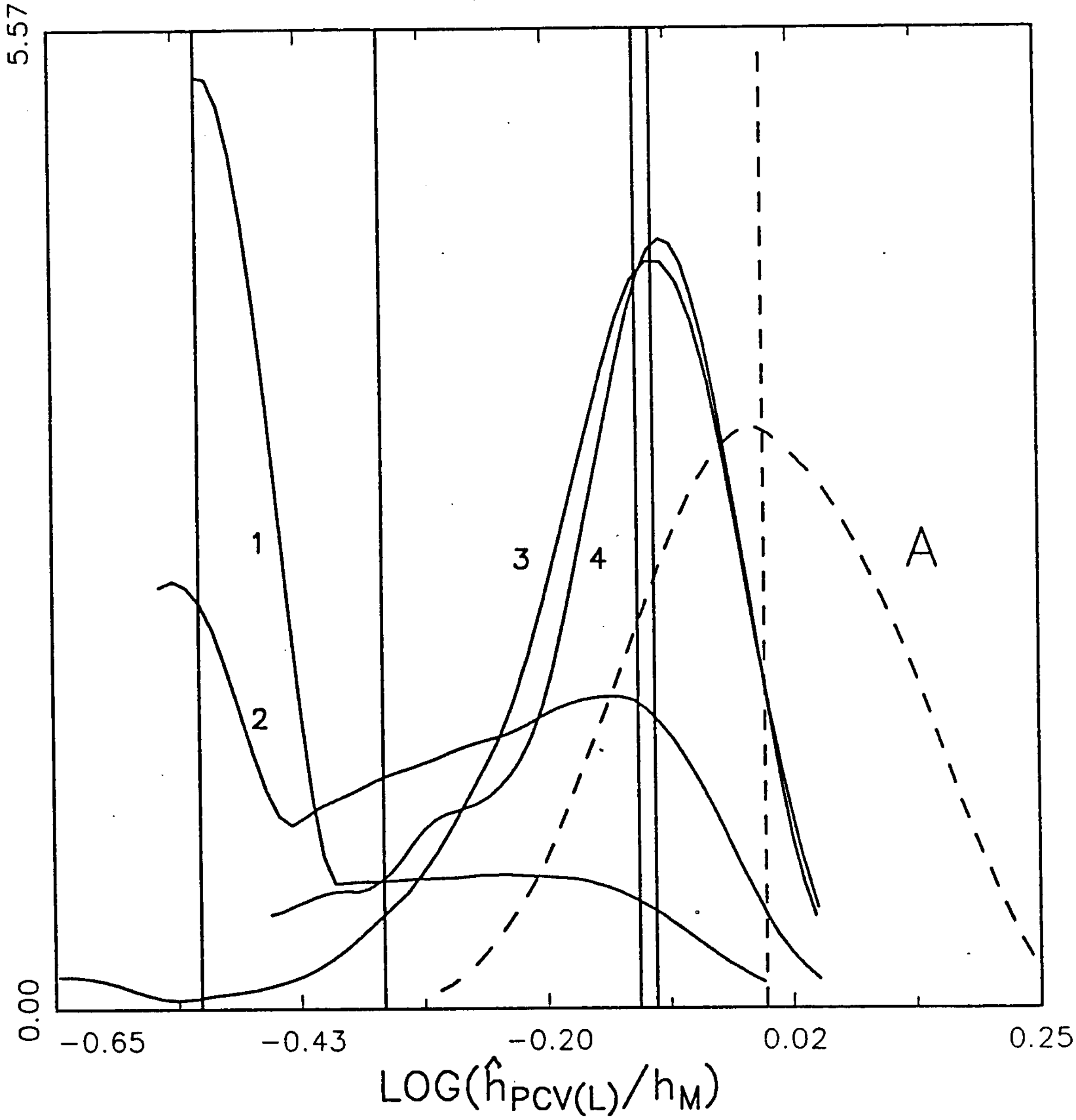


Figure 6.7: Kernel density estimates of the ratios of the modified cross-validated bandwidths to the optimal bandwidth for the negatively correlated observations with a large amount of sample variability.

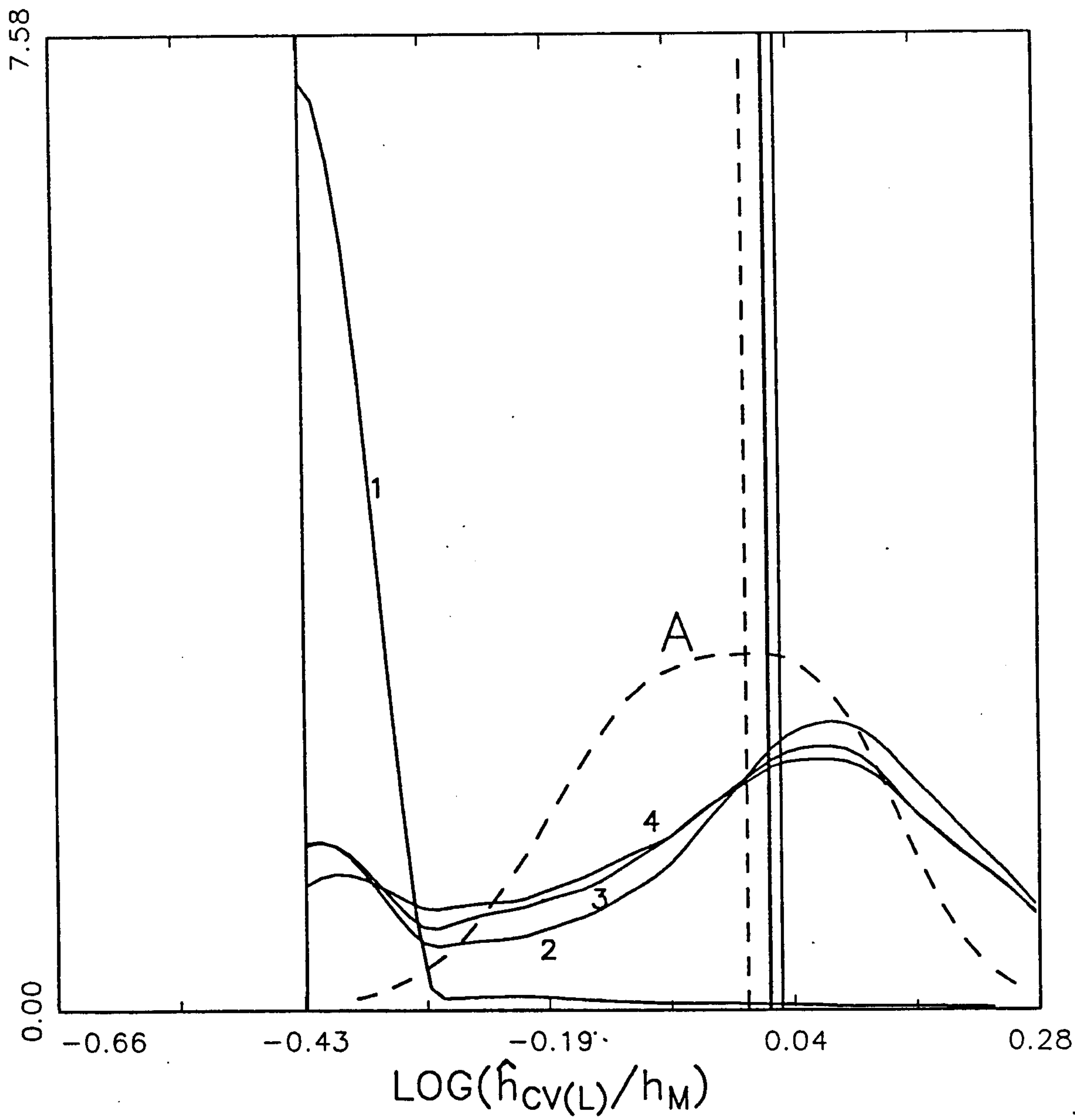


Figure 6.8: Kernel density estimates of the ratios of the partitioned cross-validated bandiwdths to the optimal bandwidth for the negatively correlated observations with a large amount of sample variability.

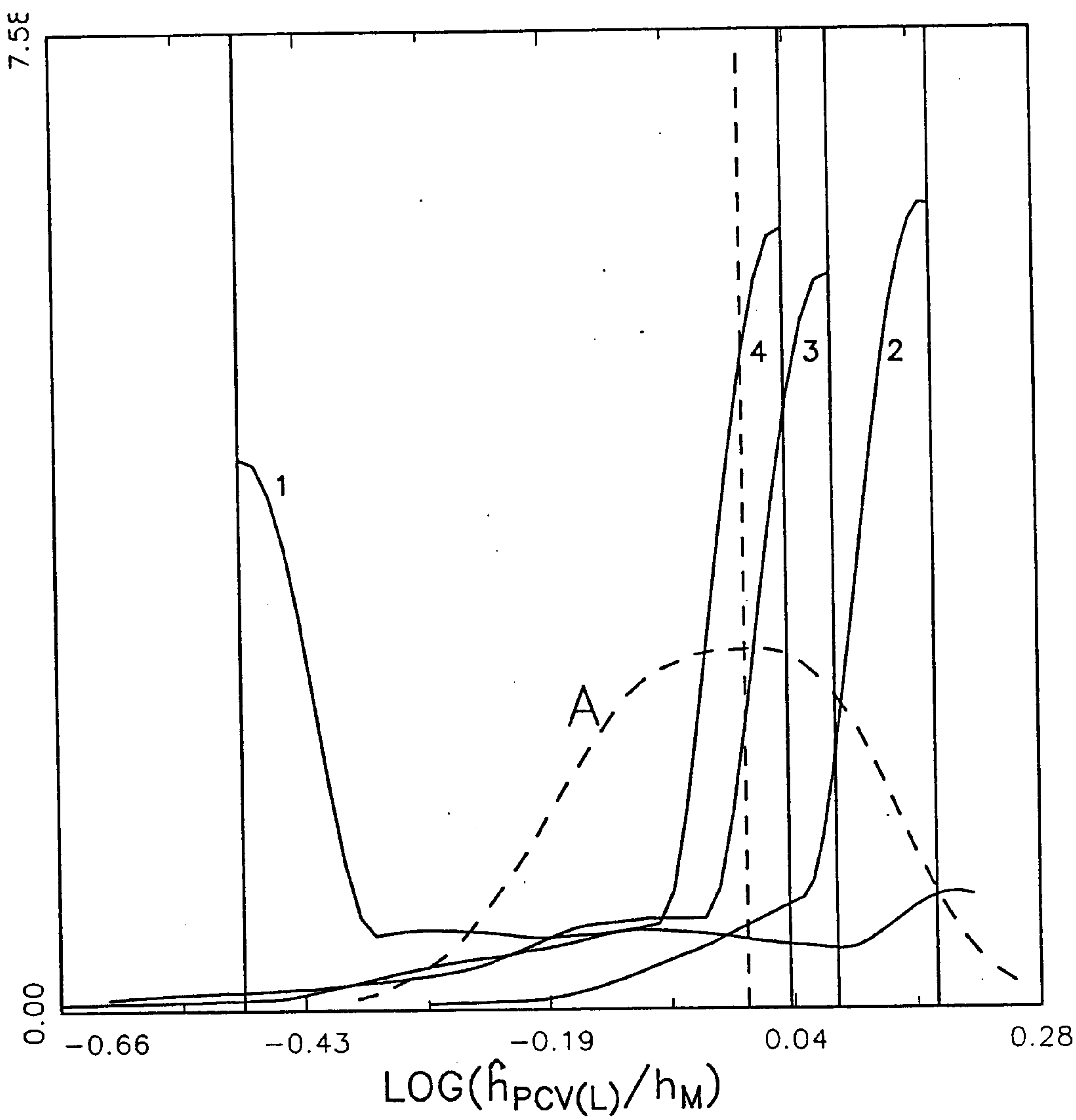


Figure 6.9: Kernel density estimates of the ratios of the modified cross-validated bandwidths to the optimal bandwidth for the negatively correlated observations with a small amount of sample variability.

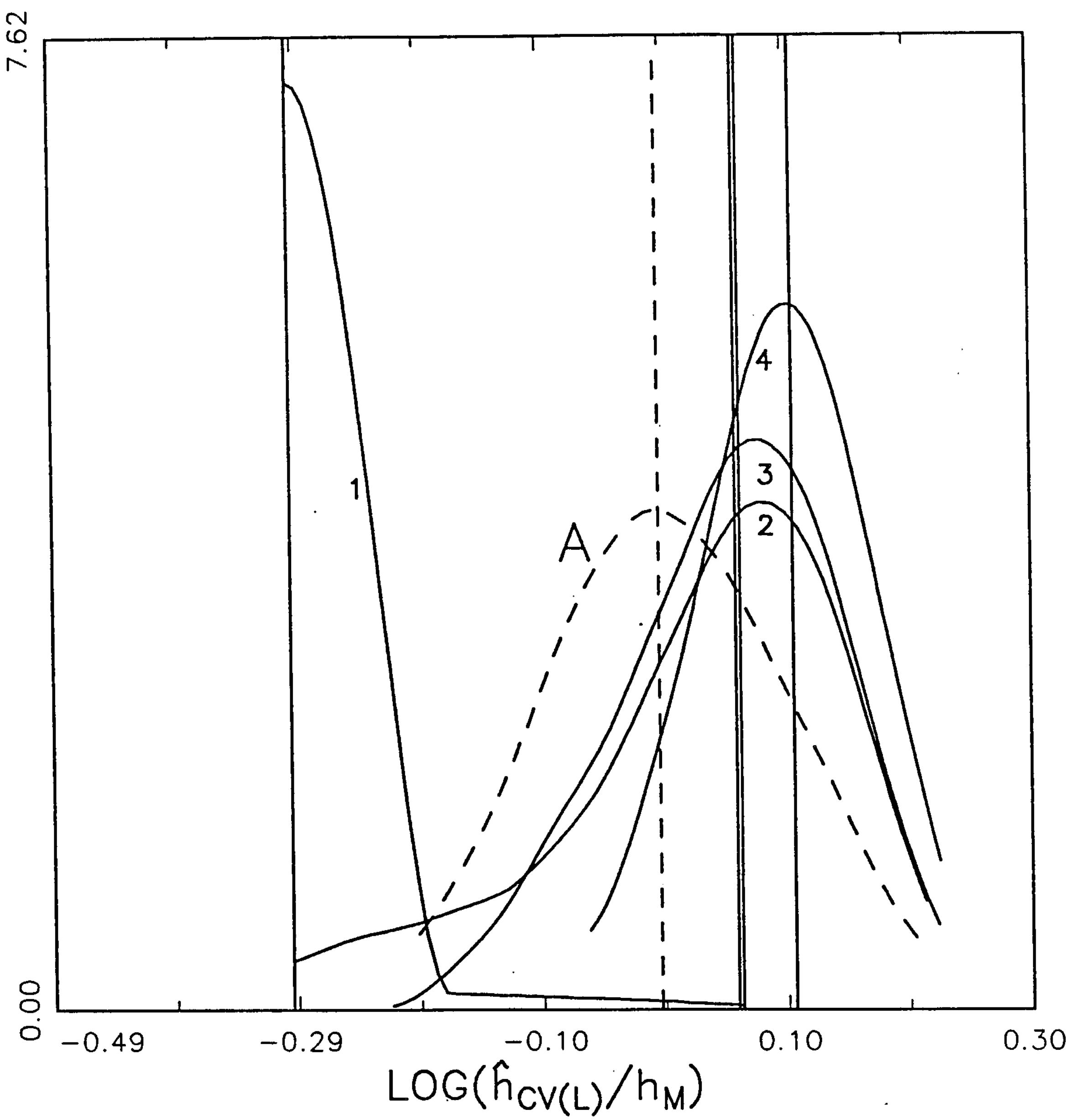
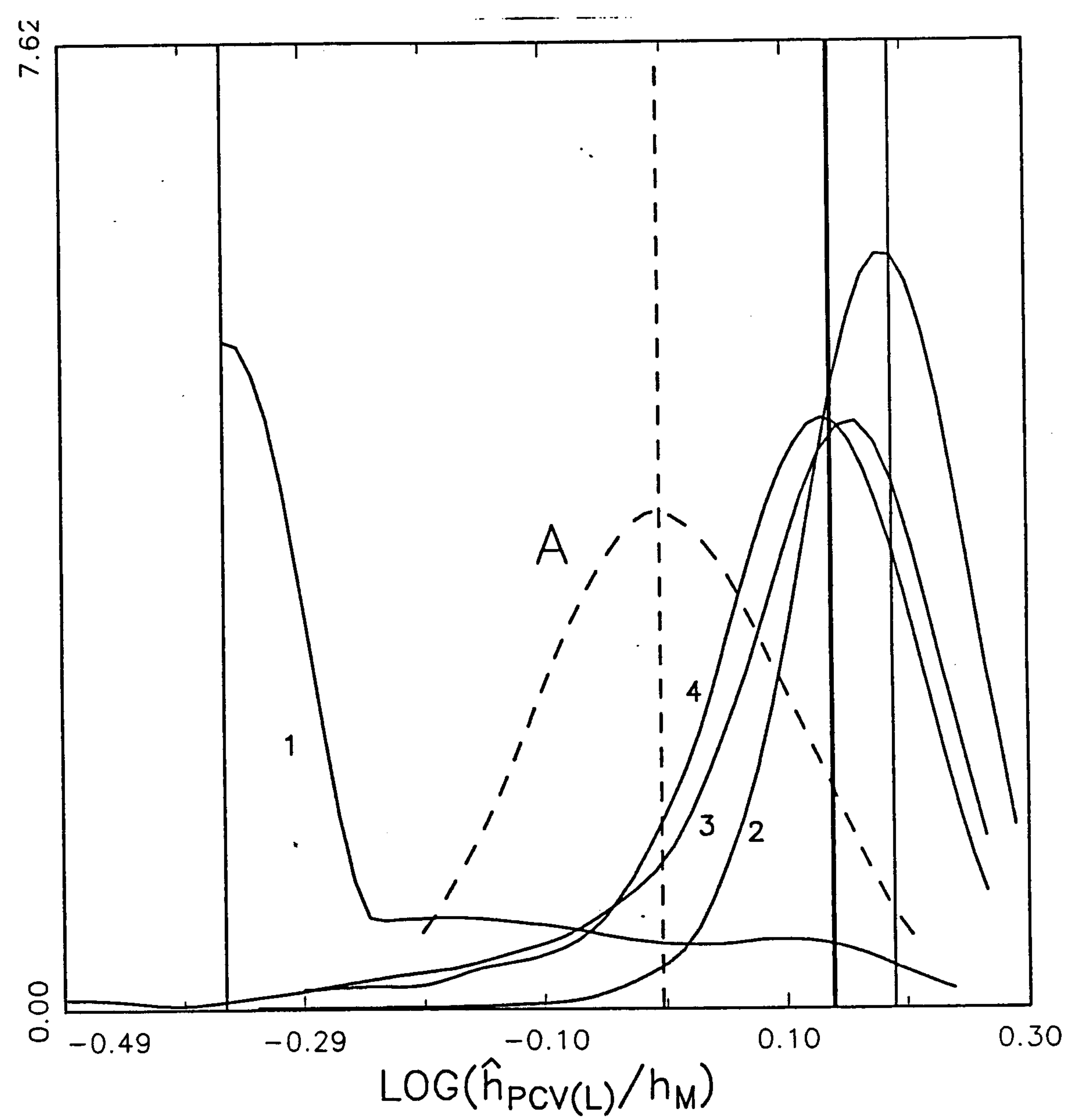


Figure 6.10: Kernel density estimates of the ratios of the partitioned cross-validated bandiwdths to the optimal bandwidth for the negatively correlated observations with a small amount of sample variability.



$g^{-1/5}$ and the minimizer of the average of 1000 curves $CV^*(h)$. The peak on the left or the right side of the density estimate is caused by the dependence effect and the bandwidth selection interval. If there still exists strong dependence effect for any value of $\ell$ or g, then the bandwidth estimate would be on the left or the right bound of the bandwidth selection interval.

Observe that the density estimates of $h_A/h_M$, $\hat{h}_{CV(\ell)}/h_M$, and $\hat{h}_{PCV(g)}/h_M$ show quite clearly that there is substantial skewness and kurtosis despite the limiting normal distribution as given in (4.2.1), (4.2.3), and (5.3.1) respectively. This indicates that, for each $\ell$ and g,

a. The ratios of bandwidths have very slow rates of convergence.

b. There may be some other limiting distributions which provide a better fit than the asymptotic normality. For example, a limiting exponential distribution and a limiting $\chi^2$ distribution with some parameters are possible. For the equally spaced fixed, circular design Mallows' criterion, and independent observations, Chiu (1989b) established that the bandwidth estimator has a limiting $\chi^2$ distribution.

For the dependent observations, the bias of the partitioned cross-validated bandwidths could be understood by looking at Figures 6.4, 6.6, 6.8, and 6.10. For all values of g, the kernel density estimates of $\hat{h}_{PCV(g)}/h_M$ are not close to that of $h_A/h_M$. The curves marked by the numbers 3 and 4, corresponding to the large values of g, are very close. This means that as g increases, the density estimate of $\hat{h}_{PCV(g)}/h_M$ converges together, but not close to that of $h_A/h_M$.

The results given in Table 6.2 and Figures 6.3 and 6.4 were derived from the same regression setting. In the analysis of the results given in Table 6.2, we concluded that the PCV was better than the $CV_\ell(h)$. By looking at the kernel density estimates, we can also decide which one, the PCV or the $CV_\ell(h)$, is better. Looking at Figures 6.3 and 6.4, the area of the curve for $h_A/h_M$ not covered by $\hat{h}_{CV(\ell)}/h_M$ is larger than that by $\hat{h}_{PCV(g)}/h_M$. Thus, In this case, the PCV is better than the $CV_\ell(h)$. On the other hand, Figures 6.5 and 6.6, which were derived from the same regression setting, shows that the $CV_\ell(h)$ is better than the PCV. It is because that the area of the curve for $h_A/h_M$ not covered by $\hat{h}_{CV(\ell)}/h_M$ is smaller than that by $\hat{h}_{PCV(g)}/h_M$. This conclusion is the same as the analysis of the results given in Table 6.3.

### 6.3 Estimation of the Optimal Number of Subgroups for the PCV

In the analysis of the asymptotic mean square error of $\hat{h}_{PCV(g)}/h_M$ as given in (5.4.1), there exists a theoretically optimal number of subgroups $g_0$ as given in (5.4.3) for the negatively correlated observations. However, $g_0$ involves some unknown quantities. In this section, we shall use the plug-in methods to estimate the regression function and the autocovariance function $\gamma(\cdot)$ of the regression errors.

Since the regression function is assumed to be smooth, then we may use a polynomial to approximate the regression function. Given the parametric model $m(x) = a + bx + cx^2 + dx^3 + ex^4$, and the weight function $W(x) = (5/3)I_{[1/5,4/5]}(x)$, through a straightforward calculation, we have the following quantities:

$$\int (m'')^2 W = 4c^2 + 12cd + 13.44ce + 10.08d^2 + 24.48de + 15.71328e^2,$$

$$\int (m^{(3)})^2 W = 36d^2 + 144dc + 161.28c^2,$$

$$\int m'' m^{(3)} W' = 0,$$

where the notation $\int$ denotes $\int dx$. Then the ordinary least square estimates of the regression parameters are plugged into the above equations.

For the estimation of the autocovariance function $\gamma(\cdot)$ of the regression errors, there are two different approaches available. Solo (1981) showed that the ordinary least square estimates of the regression parameters are strongly consistent if the regression errors are a stationary process with a bounded spectral density. The spectral density of an ARMA process as defined in Section 3.2 is bounded if the $|\phi(z)| \geq r$ for some $r > 0$ on $|z| = 1$. Thus the residuals from the ordinary least square fit may be used to estimate the autocovariance function $\gamma(\cdot)$ of the regression errors. For the second method, given the AR(p) process of the regression errors, Truong (1989) showed that the estimator of the AR(p) parameters converges to the real values weakly with a rate of convergence $n^{-1/2}$. Specifically, the semiparametric regression model given in Truong (1989) is

$$Y_i = m(x_i) + \epsilon_i,$$

for $i = 1, 2, \ldots, n$. Here, the design points $x_i$ equal $i/n$. The regression function $m$ is Hoelder continuous of order 1. The regression errors $\epsilon_i$ are an AR(p) process and defined as

$$\epsilon_i = e_i + \phi_1 \epsilon_{i-1} + \phi_2 \epsilon_{i-2} + \cdots + \phi_p \epsilon_{i-p},$$

where $e_i$ are IID $N(0,\sigma^2)$ and $1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0$ on $|z| \leq 1$. Given the uniform kernel function $K(x) = (1/2)I_{[-1,1]}(x)$, the

Nadaraya-Watson estimator (1.2), and the bandwidth $h \sim (n^{-1}\log n)^{1/3}$, then

$$n^{1/2}(\hat{\Phi} - \Phi) \Rightarrow N(0, \sigma^2 \Gamma_p^{-1}),$$

where $\Phi = (\phi_1, \phi_2, \ldots, \phi_p)'$ and $\Gamma_p = [\text{Cov}(\epsilon_i, \epsilon_j)]$ for all i, j = 1, 2, ..., p. Here $\hat{\Phi}$ is an estimator of $\Phi$ obtained by regression $\hat{\epsilon}_j$ on $\hat{\epsilon}_{j-1}$, $\hat{\epsilon}_{j-2}$, ..., $\hat{\epsilon}_{j-p}$, where $\hat{\epsilon}_j$ are defined as $Y_j - \hat{m}(x_j)$ for all j. The notation $a_n \sim b_n$ means that $a_n/b_n$ is bounded away from zero and infinity. A weakness of this approach is that $e_i$ are assumed to be $N(0, \sigma^2)$.

In the rest of this section, we shall give the results for estimation of $g_0$, the optimal number of subgroups of the PCV. The simulated regression setting is the same as the one described in Section 6.1. The regression errors are an AR(1) process. The combination of $\phi$ and $\sigma^2$ is the same as the one given in Table 6.5. We do not use the combination of $\phi$ and $\sigma^2$ as given in Table 6.4. This is because the value of MSE in Table 6.4 decreases monotonely. The pseudo normal random variables are generated by the function RNDNS in GAUSS where the seed is given as 456. One thousand sets of data are generated. The average and the standard deviation of the estimated values of $g_0$ fitted by the AR(1) process of regression errors are given in Table 6.6. The symbol * in Table 6.6 denotes that the two methods use the same value. Here the bandwidth h is taken as $(3/10)(n^{-1}\log n)^{1/3}$. The two methods of Solo (1981) and Truong (1989) use the same estimates of $\int(m'')^2 W$ and $\int(m^{(3)})^2 W$. The value of $g_0$ is calculated by

$$[V^5 n/(A^5 B^5)]^{1/6} - (5/6)[V^7/(A^{13} Bn)]^{1/6}$$

as given in (5.4.3). The quantities V, A, and B were given in (5.4.3).

Table 6.6: The average and the standard deviation of the estimated values of the theoretically optimal number of subgroups derived by the methods given in Solo (1981) and Truong (1989), and fitted by the AR(1) process of the regression errors. The notation * denotes these two methods use the same data.

| | $\int (m'')^2 W$ | $\int (m''')^2 W$ | $\phi$ | $\gamma(0)$ | A | B | V | $g_0$ |
|---|---|---|---|---|---|---|---|---|
| Real Value | 0.056 | 4.221 | -0.6 | 0.0000128 | 0.24 | 0.32 | 0.34 | 6.9 |
| Average of the above quantity | | | | | | | | |
| Solo | 0.047 | 2.895 | -0.59 | 0.0000129 | 0.32 | 0.21 | 0.36 | 8.1 |
| Truong | * | * | -0.60 | 0.0000128 | 0.33 | 0.21 | 0.38 | 8.2 |
| Standard deviation of the above quantity | | | | | | | | |
| Solo | 0.003 | 0.396 | 0.06 | 0.0000018 | 0.05 | 0.02 | 0.09 | 0.6 |
| Truong | * | * | 0.06 | 0.0000019 | 0.05 | 0.02 | 0.10 | 0.7 |

Looking at Table 6.5, the minimum value of the MSE is at $g = 6$. Using one step interpolation improvement, the minimizer of the MSE is 6.25. In Table 6.6, the theoretical value of $g_0$ is 6.9. These two methods provide good estimates for the AR(1) parameters $\phi$ and $\gamma(0)$. However, the fitted polynomial does not provide good estimates for the integration of the derivatives of the regression function. This means that we need other methods which give good estimates of the derivatives of the regression function.

## BIBLIOGRAPHY


Brockwell, P. J. and Davis, R. A. (1987). Times Series: Theory and Methods. Springer Series in Statistics, Springer-Verlag, New York.

Chiu, S. T. (1987). Estimating the parameters of the noise spectrum for a time series with trend: with application to bandwidth selection for nonparametric regression. To appear.

Chiu, S. T. (1989a). Bandwidth selection for kernel estimation with correlated noise. To appear in Statistics and Probability Letters.

Chiu, S. T. (1989b). On the asymptotic distributions of bandwidth estimates. To appear in Annals of Statistics.

Chu, C. K. and Marron, J. S. (1988). Comparison of kernel regression estimators. North Carolina Institute of Statistics, Mimeo Series No. 1754.

Clark, R. M. (1975). A calibration curve for radiocarbon data. Antiquity, 49, 251-266.

Collomb, G. (1981). Estimation non-param'etrique de la regression: revue bibliographique. International Statistical Review, 49, 75-93.

Collomb, G. (1985). Non parametric time series analysis and prediction: uniform almost sure convergence of the window and K-NN autoregressive estimates. Statistics, 16, 297-307.

Collomb, G. (1985). Nonparametric regression: an up-to-date bibliography. Statistics, 16, 309-324.

Cox, D. R. (1984). Long-range dependence: a review. In: Statistics: An Appraisal, H. A. David and H. T. David, eds., 55-74. Iowa State Univ. Press.

David, H. A. (1981). Order Statistics, Wiley, New York.

Devroye, L. P. and Wagner, T. J. (1980). On the $L_1$ convergence of kernel estimators of regression functions with applications is discrimination, Z. Wahrsh. vrew. Gebiete, 51, 15-25.

Eubank, R. L. (1988). Spline Smoothing and Nonparametric Regression. Marcel Dekker Inc., New York.

Gasser, T. and Mueller, H. G. (1979). Kernel estimation of regression functions. In Smoothing techniques for curve estimation, Lecture Notes in Math., No. 757, 23-68, Spring-Verlag, New York.

Gasser, T. and Mueller, H. G. (1984). Estimating regression functions and their derivatives by the kernel methods. Scandinavian Journal of Statistics, 11, 171-185.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series and fractional differencing. Journal of Time Series Analysis, 1, 15-30.

Haerdle, W. (1988). Applied nonparametric regression, unpublished monograph.

Haerdle, W., Hall, P., and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? Journal of the American Statistical Association, 83, 86-101.

Haerdle, W. and Marron, J. S. (1983). The nonexistence of moments of some kernel regression estimators. North Carolina Institute of Statistics, Memio Series No. 1537.

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

Haerdle, W. and Vieu, P. (1987). Non parametric kernel regression function estimation for $\phi$-mixing observations, Part I: Optimal squared error estimation. To appear.

Hall, P. (1984). Central limit theorem for integrated squared error of multivariate nonparametric density estimators. Journal of Multivariate Analysis, 14, 1-16.

Hall, P. and Marron, J. S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. Theory of Probability and Related Fields, 74, 567-581.

Hart, J. D. (1987). Kernel regression estimation with time series errors. To appear.

Hart, J. D. and Wehrly, T. E. (1986). Kernel regression using repeated measurements data. Journal of the American Statistical Association, 81, 1080-1088.

Jennen-Steinmetz, C. and Gasser, T. (1987). A unifying approach to nonparametric regression estimation. Journal of the American Statistical Association, 83, 1084-1089.

Mack, Y. P. and Mueller, H. G. (1989). Convolution type estimators for nonparametric regression. Statistics and Probability Letters, 7, 229-239.

Mallows, C. L. (1973). Some comments on $C_p$. Technometrics, 15, 661-675.

Marron, J. S. (1987). Partitioned cross-validation. Econometric Reviews, 6, 271-284.

Marron, J. S. (1988). Automatic smoothing parameter selection: A survey. Empirical Economics, 13, 187-208.

Mueller, H. G. (1988). Nonparametric Analysis of Longitudinal Data, Lecture Notes in Statistics, No. 46, Spring-Verlag, Berlin.

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

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

Rosenblatt, M. (1969). Conditional probability density and regression estimates. In: Multivariate Analysis II, P. R. Krishnaiah, ed., 25-31, Wiley, New York.

Schuster, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Annals of Mathematical Statistics, 43, 84-88.

Serfling, R. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.

Shibata, R. (1981). An optimal selection of regression variables. Biometrika, 68, 45-54.

Solo, V. (1981). Strong consistency of least squares estimators in regression with correlated disturbances. The Annals of Statistics, 9, 689-693.

Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. Annals of Statistics, 12, 917-926.

Truong, Y. K. (1989). Nonparametric curve estimation with time series errors. To appear.

Vieu, P. and Hart, J. (1989). Nonparametric regression under dependence: A class of asymptotically optimal data-driven bandwidths. To appear.

Wahba, G. and Wold, S. (1975). A completely automatic french curve: fitting spline functions by cross-validation. Communications in Statistics, 4, 1-17.

Watson, G. S. (1964). Smooth regression analysis. Sankhya, Series A, 26, 359-372.

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. Theory of Probability and its Applications, 5, 302-305.