Local Polynomial M-estimation in Random Design Regression with Dependent Errors

By

Yixuan Liu

A thesis submitted to Macquarie University

for the degree of Master of Research

Department of Statistics

May 2018

MACQUARIE University
SYDNEY · AUSTRALIA

Typeset in $\LaTeX\,2_\varepsilon$.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Yixuan Liu

Acknowledgements

It is my honour to sincerely thank many people who have kindly assisted me while I wrote this thesis.

Firstly, I would like to thank my supervisor, Doctor Justin Wishart. His knowledge and suggestions opened the door of nonparametric statistics for me in which I have been obsessed. His patience and encouragement led me keeping going ahead, and his comments on this thesis brought the largest help to me.

I would like to thank the coordinator of MRes, Doctor Thomas Fung, who helped me with my project of research.

I would like to thank my fellow students in the Department of Statistics and the Department of Mathematics. Because of their help and company, I spent ten enjoyable months in my life. I appreciate the friendship we have built with each other.

I would like to thank Macquarie University and the Department of Statistics for their supporting on my research over these ten months.

I would like to specially thank the science fiction master, Issac Asimov. His compositions made me fall in love with science especially statistics, and his texts have always been inspiring me.

Finally, I would like to thank my parents and two special members in my family, Teddy and Lucky. Their supporting is the largest motivation to guide me.

Abstract

The random design nonparametric regression model with short-range dependent and long-range dependent errors is investigated. The asymptotic behaviour of the robust local polynomial M-estimator is investigated under two conditions.

Asymptotic results are established by decomposing the local polynomial estimator into two terms: a martingale term and a conditional expectation term. It is found that the local polynomial M-estimator is asymptotically normal when errors are short-range dependent. When the errors are long-range dependent, a more complex behaviour is observed that depends on the size of the bandwidth. If the bandwidth is small enough, the long-range dependent scenario is similar to the short-range dependent case. If the bandwidth is relatively large the asymptotic result is more intricate and the long-range dependent variables dominate. Moreover, the optimal bandwidth in the case of short-range dependence is determined.

Contents

Nomenclature

$\mathbb{R}$	Field of real numbers	2
$\mathbb{Z}$	Field of integers	7
$\mathbb{N}$	Natural numbers	13
$\xrightarrow{d}$	Converges in distribution	16
$\xrightarrow{P}$	Converges in probability	25
$x \vee y$	Maximum of x and y	28
∇	Gradient operator	10
$\lfloor \cdot \rfloor$	Floor function	43
$\|X\|_p^p$	The p^{th} moment of a random variable, X	13
$\mathcal{F}_j$	A filtration	9
f_∞	The density function of (Z, X)	10
f_1	The density function of (η, X)	10
f_Z	The density function of Z	10
f_X	The density function of X	10
f_η	The density function of η	10
$f_{i,Z}$	The density function of $Z_i - Z_{i,i-1}$	10

1
Introduction

Regression analysis is a very common issue in statistics. It explores the relationship between explanatory variables X and response variables Y, and seeks to explain how the change of X influences Y. Sometimes the relationship is not complex so parametric techniques can be applied; however, parametric techniques are not flexible and sometimes not suitable. Intuitively, a data-driven technique is required to complete the task, and the suitable technique is called nonparametric regression estimation. In this thesis, we focus on a specific method of nonparametric regression estimation, the local polynomial estimator, which has numerous advantages compared with the traditional local constant estimators, such as the Naradaya-Watson estimator and the Gasser-Müller estimator. We show that the asymptotic behaviour of a robust local polynomial estimator on a regression model with dependent errors is determined by the strength of dependence. When errors are short-range dependent, the asymptotic law is the same as the independent case. When errors are long memory, it is found that different bandwidths can result in different asymptotic behaviours. Before exhibiting our results, we shall introduce some useful concepts.

1.1 PARAMETRIC REGRESSION

The most common regression technique is linear regression and the aim of it is to fit a linear relationship between X and Y, and for those parts deviating from the line they are regarded as errors. Thus, the relationship can be modeled as

$$Y = a + bX + \varepsilon,$$

where ε is the error term and it is assumed that $\mathbb{E}\varepsilon = 0$. A very basic case of this model is that the conditional mean is linear:

$$m(x) =: \mathbb{E}(Y|X = x) = a + bx.$$

As a and b are unknown, the fitted model is constructed by estimating them with $\hat{a}$ and $\hat{b}$ via different approaches, the most common being least squares estimation.

However, the case that the conditional mean appears to be linear is not always guaranteed so a validation on the linearity needs to be conducted. To this end, diagnostic validation is conducted with a scatter plot of X against Y and a residual analysis. If there is no linearity in the scatter plot or a pattern in the residuals is observed, then linear regression is no longer suitable to fit the data. There are some techniques which can be used to fit a nonlinear regression and a superior one among them is called polynomial regression which assumes that the conditional mean has the structure of a polynomial. Consequently, polynomial regression satisfies

$$m(x) = a_0 + a_1 x + \cdots + a_p x^p, \quad a_i \in \mathbb{R}, i = 0, 1, 2, \cdots, p.$$

This polynomial structure is assumed to occur on the entire global domain of x. However, high order polynomial parametric models are fitted on a global scale in the data. That is, each a_i is fixed for all x in the domain. This has disadvantages where in nonlinear data the fitted curves are sensitive to the observed data, and higher order polynomial parametric models can overfit the trend in the data. Nonparametric regression methods are a superior alternative when the data is nonlinear with no obvious simple structure.

1.2 NONPARAMETRIC ESTIMATION METHODS

In this section, some key techniques for nonparametric regression estimation will be introduced and the local polynomial regression estimation will be discussed in detail.

1.2.1 Kernel estimators

It is difficult to obtain much information about the value of $m(x_0)$ from the y points with corresponding x far from x_0 if there is not a specific form of the function m. Intuitively, we can take a locally weighted average for the y variables to be the estimator, and kernel estimators are this class of estimators. Among numerous nonparametric regression estimators based on the kernel function, the Nadaraya-Watson estimator and the Gasser-Müller estimator are two very classic ones.

The Nadaraya-Watson estimator was proposed independently by Naradaya (1964) and Watson (1964). Given a kernel function $K : \mathbb{R} \to \mathbb{R}$, which is usually a symmetric probability density function, and a bivariate set of observations (X, Y), the Nadaraya-Watson estimator follows

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n K(\frac{X_i - x_0}{h}) Y_i}{\sum_{i=1}^n K(\frac{X_i - x_0}{h})},$$

where h is the bandwidth of the estimator. This estimator is regarded as the local constant estimator where $m(x_0)$ is estimated locally by a constant. Fan and Gijbels (1996) exhibit several drawbacks of the Nadaraya-Watson estimator. The first one is that it has a large bias especially when the regression function or the design density have a large derivative, and it is found that even when the true regression curve is linear the bias is still large. Another problem is that the estimator has zero minimax efficiency which is a benchmark to efficiency of linear estimation.

Due to the form of the Nadaraya-Watson estimator, it is not convenient to take derivatives of the estimator and to find out the asymptotic properties. Therefore, Gasser and Müller (1979) proposed an estimator called the Gasser-Müller estimator. Assume that the data are sorted by the variable X, and then the estimator follows

$$\hat{m}(x_0) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} \frac{1}{h} K\Big(\frac{u - x_0}{h}\Big) du \, Y_i,$$

with $s_i = (X_i + X_{i+1})/2$, $X_0 = -\infty$ and $X_{n+1} = +\infty$. The Gasser-Müller estimator is regarded as a local constant estimator as well. It should be pointed out that this estimator fits not only equispaced designs but also non-equispaced designs. The Gasser-Müller estimator resolves the problem of large bias which the Nadaraya-Watson estimator has; however, this improvement brings in another problem of increasing the variability. Moreover, it is found that there exists a large order of bias for both the Nadaraya-Watson estimator and the Gasser-Müller estimator when estimating a curve at a boundary region. More details of these issues are discussed by Fan and Gijbels (1996).

Kernel estimators are not what we are interested in so more details will not be given. The interested reader can check Härdle (1990) and Fan and Gijbels (1996).

1.2.2 Local polynomial estimator

Assuming a regression function m has at least $(p+1)$ derivatives, then Taylor's expansion can locally approximate the function by

$$m(u) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(u-x_0)^j \equiv \sum_{j=0}^{p} \beta_j (u-x_0)^j \tag{1.1}$$

for u in a neighbourhood of x_0. It is said that $m(u)$ is estimated locally by (1.1) with a simple polynomial model. Then, the locally weighted least squares polynomial regression is the solution to

$$\frac{1}{h}\sum_{i=1}^{n}\Big\{Y_i - \sum_{j=0}^{p}\beta_j (X_i - x_0)^j\Big\}^2 K\Big(\frac{X_i - x_0}{h}\Big), \tag{1.2}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Therefore, the estimators of β_j $(j = 0, 1, \cdots, p)$ are expected to minimize (1.2). The bandwidth h is a critical parameter in this method. The reason is that a too large bandwidth can produce a smoother estimate while paying the price of a high modeling bias; in contrast, choosing a too small bandwidth can lower the bias while causing high variability. Thus, a theoretical choice of the bandwidth is required. However, due to some unknown quantities, it is not easy to find the theoretical bandwidth, and then some practical techniques on bandwidths selection need to be explored. Moreover, the choice of another parameter, the order of the polynomial p, is also an important issue to be discussed. For a given bandwidth h, a higher order can decrease the bias by bringing in a high variance; oppositely, a lower order has low variability while the bias is large. A reasonable choice of the order will be given in the following discussion.

Compared with local constant estimators such as the Nadaraya-Watson estimator and the Gasser-Müller estimator, the local polynomial estimator has more advantages. First, the local polynomial estimator is more efficient in minimizing the mean squared error (MSE)

$$MSE = \{Bias[\hat{m}(x)]\}^2 + Var[\hat{m}(x)] \tag{1.3}$$

as well as the integrated mean squared error (MISE); in particular, when the order of the polynomial approximation p is odd, there exists a decrease in the bias without increasing in the variability compared to the case of $p-1$. In Fan and Gijbels (1996), they recommend using the lowest odd order, i.e. $p = 1$, or occasionally $p = 3$. Second, the local polynomial estimator adapts to both random designs and fixed designs. Third, boundary effects do not influence the local polynomial estimator which can adapt automatically to the estimation at boundaries. Furthermore, it turns out that local polynomial estimators are the best linear smoother both in the interior and at the boundary in terms of its high linear minimax efficiency and this method has a high overall minimax efficiency as well.

1.2.3 Other methods

There are some other popular nonparametric estimation methods, such as orthogonal series based methods, the spline smoothing and the logspline density estimator. Interested readers are referred to Fan and Gijbels (1996) and Wasserman (2006), and we will not display those methods here.

1.3 M-ESTIMATOR

It is found that the least squares criterion in (1.2) is sensitive to outliers or aberrant observations, so how to reduce the influence of outliers on estimation is an issue to overcome. Intuitively, some robust procedures are concerned. Huber (1964) proposed the Huber's loss function which is an M-type estimator, and then Huber (1973) introduced M-estimators in the context of regression the first time. Beaton and Tukey (1974) considered the application of M-estimation in the regression model and proposed a biweight (or bisquare) loss function. A procedure called locally weighted scatterplot smoothing (LOWESS) which employs an iterative reweighted least squares scheme to decrease the influence of outliers was introduced by Cleveland (1979). Cox (1983) investigated robustified smoothing splines, Fan and Gijbels (1996) applied the local polynomial estimator in a quantile regression and Fan and Jiang (1999) studied an M-type local polynomial estimator. The robust approach used in this thesis is the M-estimator so the following content will give the concept of the M-estimator.

Denote $\hat{\varepsilon}$ the estimator of errors ε, an estimator T_n is called M-estimator (or maximum likelihood type estimator) if it minimizes $\sum_{i=1}^{n} \rho(\hat{\varepsilon}_i; T_n)$ or satisfies $\sum_{i=1}^{n} \psi(\hat{\varepsilon}_i; T_n) = 0$, where ρ, a loss function, is called the *objective loss function* and ψ, the derivative of ρ, is called the *influence curve*. The objective loss function should satisfy the following properties:

Method	*Objective Loss Function*	*Influence Curve*
Huber	$\rho(\hat{\varepsilon}) = \begin{cases} \frac{1}{2}\hat{\varepsilon}^2, & \text{for } \lvert\hat{\varepsilon}\rvert \le k \\ k\lvert\hat{\varepsilon}\rvert - \frac{1}{2}k^2, & \text{for } \lvert\hat{\varepsilon}\rvert > k \end{cases}$	$\psi(\hat{\varepsilon}) = \begin{cases} -k, & \text{for } \hat{\varepsilon} < -k \\ \hat{\varepsilon}, & \text{for } -k \le \hat{\varepsilon} \le k \\ k, & \text{for } \hat{\varepsilon} > k \end{cases}$
Biweight	$\rho(\hat{\varepsilon}) = \begin{cases} \frac{k^2}{6}\{1 - [1 - (\frac{\hat{\varepsilon}}{k})^2]^3\}, & \text{for } \lvert\hat{\varepsilon}\rvert \le k \\ \frac{k^2}{6}, & \text{for } \lvert\hat{\varepsilon}\rvert > k \end{cases}$	$\psi(\hat{\varepsilon}) = \begin{cases} 0, & \text{for } \hat{\varepsilon} < -k \\ \hat{\varepsilon} - \frac{2\hat{\varepsilon}^3}{k^2} + \frac{\hat{\varepsilon}^5}{k^4}, & \text{for } -k \le \hat{\varepsilon} \le k \\ 0, & \text{for } \hat{\varepsilon} > k \end{cases}$

Table 1.1: Objective loss and influence function of Huber's and the biweight estimator.

- Nonnegativity, $\rho(\hat{\varepsilon}) \ge 0$
- Equal to zero at the origin: $\rho(\hat{\varepsilon}) = 0$
- Symmetricity: $\rho(\hat{\varepsilon}) = \rho(-\hat{\varepsilon})$
- Monotonicity: $\rho(\hat{\varepsilon}_i) \ge \rho(\hat{\varepsilon}_j)$ if $\lvert\hat{\varepsilon}_i\rvert > \lvert\hat{\varepsilon}_j\rvert$

The Huber's loss function and the biweight loss function we mentioned previously are two very common M-type objective loss functions. Table 1.1 shows the two functions and their respective influence curves. The value k in both the Huber's loss function and the biweight loss function is called the *tuning constant*, and smaller values of k produce more resistance to outliers at the cost of lower efficiency when the errors follow a normal distribution. A feature of the Huber's estimator is that as the residual $\hat{\varepsilon}$ departs from 0 the value of the objective loss function increases boundlessly; in contrast, there are no longer increases for the biweight function when $\lvert\hat{\varepsilon}\rvert > k$. There are several other robust methods, such as the L-estimator and the R-estimator. We recommend readers to go through Huber (1996) where elaborating details of these different methods are given.

Now, it is clear that by the M-estimator, (1.2) can be re-written as

$$\frac{1}{h}\sum_{i=1}^{n}\rho\Big\{Y_i - \sum_{j=0}^{p}\beta_j(X_i - x)^j\Big\}K\Big(\frac{X_i - x}{h}\Big).$$

Correspondingly, $\hat{\beta}_j$, the estimator of β_j, should satisfy

$$\frac{1}{h}\sum_{i=1}^{n}\psi\Big\{Y_i - \sum_{j=0}^{p}\hat{\beta}_j(X_i - x)^j\Big\}K\Big(\frac{X_i - x}{h}\Big)\mathbf{X} = \mathbf{0} \tag{1.4}$$

where $\mathbf{X} = [1, X_i - x, \ldots, (X_i - x)^p]^T$.

1.4 LONG MEMORY

Most of the methods we mentioned above mainly investigate independent and identical distribution (i.i.d.) cases. There are some authors focusing on dependent cases as well, such as Csörgo and Mielniczuk (1995a) using kernel estimators to fit a fixed-design nonparametric regression model with short-range dependent (SRD) errors and Cai and Ould-Saïd (2003) investigating stationary time series sequence under α-mixing conditions by using an M-type local polynomial estimator. As a special case of dependence, long memory (or long-range dependence) (LRD) has become of interest in recent years, and a number of articles are contributed to this case. Hall and Hart (1990) studied the convergence rate of the density estimator of a fixed-design regression function with LRD errors. Csörgo and Mielniczuk (1995b) and Csörgo and Mielniczuk (1999) respectively investigated the kernel density estimator with LRD variables and the Nadaraya-Watson estimator for a random-design regression model with LRD errors, and Wu and Mielniczuk (2002) and Mielniczuk and Wu (2004) extended the results. Beran et al. (2002) gave the optimal bandwidth of the robust local polynomial estimator for a fixed-design regression function with LRD errors. In this thesis, we focus on a random-design regression function with SRD and LRD errors respectively, and the following content in this section will give the concept of SRD and LRD for a linear process.

Definition 1.4.1. *A random sequence is said to be stationary if its joint probability distribution is invariant over time.*

Let $\{Z_n, n \in \mathbb{Z}\}$ be a stationary sequence and we assume that it follows a linear process

$$Z_i = \sum_{t=0}^{\infty} a_t \eta_{i-t} \tag{1.5}$$

where $\{\eta_i,\ i \in \mathbb{Z}\}$ is an i.i.d. sequence with mean of 0 and finite variance and coefficients a_i satisfy $\sum_{i=0}^{\infty} a_i^2 < \infty$. We also assume that $a_0 = 1$. Define

$$a_i = L(i) i^{-\alpha},$$

where $L(\cdot)$ is slowly varying for $i \to \infty$ and α is a self-similarity parameter, and a_i is said to be regularly varying at infinity with index α. It is found that the strength of dependence of the linear process depends on α.

Definition 1.4.2. *Z_n is said to be SRD if $\alpha > 1$, and in this case,*

$$\sum_{i=0}^{\infty} |a_i| < \infty$$

holds.

Definition 1.4.3. *Z_n is said to be LRD if $1/2 < \alpha < 1$, and in this case,*

$$\sum_{i=0}^{\infty} |a_i| = \infty$$

holds.

Moreover, there are three very useful results from Bojanic and Seneta (1973) and Seneta (1976) to bound the regularly varying sequences at the boundaries near zero and infinity:

i)

$$\sup_{i \geq n} i^{-p} L(i) \sim n^{-p} L(n), \quad \text{for } p > 0 \tag{1.6}$$

ii)

$$\lim_{n\to\infty} \frac{1}{n^{1+p} L(n)} \sum_{i=1}^{n} i^p L(i) = \frac{1}{1+p}, \quad \text{for } p > -1$$

iii)

$$\lim_{n\to\infty} \sum_{i=1}^{n} i^p L(i) = C, \quad \text{for } p < -1. \tag{1.7}$$

1.5 THESIS OUTLINE

This thesis is organized as follows. The notations and assumptions used throughout this thesis are listed in Chapter 2. Chapter 3 displays the main result via two theorems. Chapter 4 gives the conclusion and discusses the potential work of this thesis in the future, and our proofs to derive the result are provided in Chapter 5.

2

Notations and Assumptions

This chapter exhibits the notations and the assumptions used throughout this thesis. Section 2.1 gives the regression model and the corresponding local polynomial M-estimator. Section 2.2 exhibits the assumptions used to depict the asymptotic behaviour of the estimator.

2.1 NOTATIONS

Assume that $\{Y_i, X_i\}_{-\infty}^{\infty}$ are a stationary sequence, then we consider the regression model

$$Y_i = m(X_i) + Z_i, \tag{2.1}$$

where errors $\{Z_i\}_{i=1}^n$ are a stationary sequence and are independent of $\{X_i\}_{i=1}^n$, and $\{Z_i\}_{i=1}^n$ also satisfy $\mathbb{E}Z_i = 0$ almost surely. We assume $\{Z_i\}_{i=1}^n$ to be a linear process in the form of (1.5) with $a_i = L_Z(i)i^{-\alpha_Z}$ and $\{X_i\}_{i=1}^n$ to be i.i.d. satisfying $\mathbb{E}X_i = 0$ and $\mathbb{E}X^2 < \infty$.

Let $W_i = (Z_i, X_i)$ and $W_{i,j} = \mathbb{E}(W_i|\mathcal{F}_j)$, where $\mathcal{F}_j = \sigma(\ldots, \eta_{j-1}, X_{j-1}, \eta_j, X_j)$ is a filtration, and correspondingly, $Z_{i,j} = \mathbb{E}(Z_i|\mathcal{F}_j)$ and $X_{i,j} = \mathbb{E}(X_i|\mathcal{F}_j)$. Note that since $\{X_i\}_{i=1}^n$ are i.i.d., then $X_{i,j} = 0$ when $j < i$.

Denote f_1 the density function of (η, X) with the marginal density functions f_η and f_X respectively and f_∞ the density function of (Z, X) with the marginal density functions f_Z and f_X respectively. Due to the independence between $\{Z_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$, f_1 and f_∞ satisfy

$$f_1 = f_\eta \cdot f_X \quad \text{and} \quad f_\infty = f_Z \cdot f_X.$$

Denote $f_{i,Z}$ the density function of $Z_i - Z_{i,i-1} = \sum_{t=0}^{i-1} a_t \eta_{i-t}$.

Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$. Our purpose is to investigate the asymptotic behaviour of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$. Define

$$\Psi_n(\boldsymbol{\beta}) = \frac{1}{nh} \sum_{i=1}^n \psi \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right\} K\left(\frac{X_i - x_0}{h}\right) \mathbf{X},$$

so $\Psi_n(\boldsymbol{\beta})$ can be expanded at $\hat{\boldsymbol{\beta}}$ by Taylor's theorem to turn into

$$\Psi_n(\hat{\boldsymbol{\beta}}) = \Psi_n(\boldsymbol{\beta}) + \nabla \Psi_n(\boldsymbol{\beta}^*) \cdot (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{2.2}$$

where $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \dots, \beta_p^*)^T$ and each β_i^* is between β_i and $\hat{\beta}_i$, and

$$\nabla \Psi_n(\boldsymbol{\beta}^*) = \frac{1}{nh} \sum_{i=1}^n \psi' \left\{ Y_i - \sum_{j=0}^p \beta_j^* (X_i - x_0)^j \right\} K\left(\frac{X_i - x_0}{h}\right) \mathbf{X}\mathbf{X}^T$$

and ∇ denotes the gradient operator. Due to (1.4), it can be obtained that $\Psi_n(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, and then (2.2) indicates

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = -[\nabla \Psi_n(\boldsymbol{\beta}^*)]^{-1} \cdot \Psi_n(\boldsymbol{\beta}). \tag{2.3}$$

Define the remainder of $m(X_i)$ and its Taylor's expansion about x_0 with

$$R(X_i) = m(X_i) - \sum_{j=0}^p \frac{m^{(j)}(x_0)}{j!} (X_i - x_0)^j \equiv m(X_i) - \sum_{j=0}^p \beta_j (X_i - x_0)^j.$$

From (2.1), it is possible to decompose $\Psi_n(\boldsymbol{\beta})$ into

$$\begin{aligned}
\Psi_n(\boldsymbol{\beta}) &= \frac{1}{nh} \sum_{i=1}^n \psi \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right\} K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} \\
&= \frac{1}{nh} \sum_{i=1}^n \psi \left\{ Y_i - m(X_i) + m(X_i) - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right\} K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} \\
&= \frac{1}{nh} \sum_{i=1}^n \psi[Z_i + R(X_i)] K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} \\
&= \frac{1}{nh} \sum_{i=1}^n \psi(Z_i) K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} + \frac{1}{nh} \sum_{i=1}^n \psi'(Z_i) R(X_i) K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} \\
&\quad + \frac{1}{nh} \sum_{i=1}^n \Big\{ \psi\left[Z_i + R(X_i)\right] - \psi(Z_i) - \psi'(Z_i) R(X_i) \Big\} K\left(\frac{X_i - x_0}{h}\right) \mathbf{X} \\
&\equiv \Delta_{n,1} + \Delta_{n,2} + \Delta_{n,3}.
\end{aligned}$$

We also apply a decomposition in $\nabla\Psi_n(\boldsymbol{\beta}^*)$. Let

$$\delta_i = R(X_i) + \sum_{j=0}^{p}(\beta_j - \beta_j^*)(X_i - x_0)^j, \tag{2.4}$$

then $\nabla\Psi_n(\boldsymbol{\beta}^*)$ can be decomposed into

$$\begin{aligned}
\nabla\Psi_n(\boldsymbol{\beta}^*) &= \frac{1}{nh}\sum_{i=1}^{n}\psi'\left\{Y_i - \sum_{j=0}^{p}\beta_j^*(X_i - x_0)^j\right\}K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{nh}\sum_{i=1}^{n}\psi'\left\{Y_i - \sum_{j=0}^{p}\beta_j(X_i - x_0)^j + \sum_{j=0}^{p}(\beta_j - \beta_j^*)(X_i - x_0)^j\right\}K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{nh}\sum_{i=1}^{n}\psi'\left\{Y_i - m(X_i) + R(X_i) + \sum_{j=0}^{p}(\beta_j - \beta_j^*)(X_i - x_0)^j\right\}K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{nh}\sum_{i=1}^{n}\psi'(Z_i + \delta_i)K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{nh}\sum_{i=1}^{n}\psi'(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T + \frac{1}{nh}\sum_{i=1}^{n}\big[\psi'(Z_i + \delta_i) - \psi'(Z_i)\big]K\Big(\frac{X_i - x_0}{h}\Big)\mathbf{X}\mathbf{X}^T \\
&\equiv \Delta_{n,4} + \Delta_{n,5}.
\end{aligned}$$

Therefore, (2.3) is equivalent to

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} + (\Delta_{n,4} + \Delta_{n,5})^{-1}\Delta_{n,2} = -(\Delta_{n,4} + \Delta_{n,5})^{-1}(\Delta_{n,1} + \Delta_{n,3}). \tag{2.5}$$

Wu and Mielniczuk (2002) and Mielniczuk and Wu (2004) proposed a decomposition to study the behaviour of regression estimators with LRD errors, and we will extend this method to our robust local polynomial estimator. Let $\boldsymbol{\zeta}_{i,1} = \psi(Z_i)K((X_i - x_0)/h)\mathbf{X}$, then $\Delta_{n,1}$ can be decomposed into

$$\begin{aligned}
nh\Delta_{n,1} &= \sum_{i=1}^{n}\big[\boldsymbol{\zeta}_{i,1} - \mathbb{E}(\boldsymbol{\zeta}_{i,1}|\mathcal{F}_{i-1})\big] + \sum_{i=1}^{n}\mathbb{E}(\boldsymbol{\zeta}_{i,1}|\mathcal{F}_{i-1}) \\
&\equiv M_{n,1} + N_{n,1}.
\end{aligned}$$

We do a similar decomposition to $\Delta_{n,2}$, $\Delta_{n,4}$ and $\Delta_{n,5}$. Let $\boldsymbol{\zeta}_{i,2} = \psi'(Z_i)R(X_i)K((X_i - x_0)/h)\mathbf{X}$, $\boldsymbol{\zeta}_{i,4} = \psi'(Z_i)K((X_i - x_0)/h)\mathbf{X}\mathbf{X}^T$ and $\boldsymbol{\zeta}_{i,5} = \big[\psi'(Z_i + \delta_i) - \psi'(Z_i)\big]K((X_i - x_0)/h)\mathbf{X}\mathbf{X}^T$, then we have

$$\begin{aligned}
nh(\Delta_{n,k} - \mathbb{E}\Delta_{n,k}) &= \sum_{i=1}^{n}\big[\boldsymbol{\zeta}_{i,k} - \mathbb{E}(\boldsymbol{\zeta}_{i,k}|\mathcal{F}_{i-1})\big] + \sum_{i=1}^{n}\big[\mathbb{E}(\boldsymbol{\zeta}_{i,k}|\mathcal{F}_{i-1}) - h\mathbb{E}\Delta_{n,k}\big] \\
&\equiv M_{n,k} + \bar{N}_{n,k}
\end{aligned}$$

where $k = 2, 4$ and 5.

Definition 2.1.1. *On a probability space $(X, \mathcal{F}, P)$, a stochastic sequence $\{X_t\}_{-\infty}^{\infty}$ is said to be a martingale difference sequence if it satisfies*

$$\mathbb{E}|X_t| < \infty \quad \textit{and} \quad \mathbb{E}[X_t|\mathcal{F}_{t-1}] = 0 \quad a.s.$$

By the tower property of conditional expectations, it follows that each $M_{n,i}$ ($i = 1, 2, 4, 5$) forms a martingale difference sequence. Thus by the martingale central limit theorem, it can be shown that with a particular norming sequence, $M_{n,i}$ has a Gaussian limit (see Lemma 2). We find that the asymptotic behaviours of $\Delta_{n,2}$, $\Delta_{n,4}$ and $\Delta_{n,5}$ under SRD stay equivalent to the case of LRD (see Lemma 4). It is also found that $\Delta_{n,3}$ is negligible to $\Delta_{n,1}$ and the asymptotic law of $\hat{\boldsymbol{\beta}}$ is determined by $\Delta_{n,1}$. Further, when $\{Z_i\}_{i=1}^{n}$ are SRD, $M_{n,1}$ determines the asymptotic behaviour of the estimator; however, when $\{Z_i\}_{i=1}^{n}$ are LRD, there exists a dichotomous phenomenon: if the bandwidth h is small enough, then $M_{n,1}$ still determines the limit, and if h is large enough, then the asymptotic law is dominated by $N_{n,1}$.

Define

$$\Lambda_n^2 = 2n\Theta_{2n}^2 + \sum_{i=1}^{\infty}(\Theta_{n+i} - \Theta_i)^2,\ \Theta_n = \sum_{i=1}^{n}\theta_i,\ \theta_i = |a_{i-1}|\sqrt{A_{i-1}} \text{ and } A_i = \sum_{j=i}^{\infty}|a_j|^2,$$

it is found that Λ_n is a key sequence to depict the asymptotic behaviour of $\Delta_{n,1}$ (see Lemma 5). Mielniczuk and Wu (2004) finds that when $|a_i| = O(L_Z(i)i^{-\alpha_Z})$ for some $\alpha_Z > 1/2$, then $\Lambda_n = O\left[\sqrt{n}\sum_{i=1}^{2n} i^{1/2-2\alpha_Z}L^2(i) + n^{2-2\alpha_Z}L^2(n)\right]$. Thus, by (1.6)-(1.7), the order of Λ_n is able to be divided into

$$\Lambda_n = \begin{cases} O(\sqrt{n}), & \alpha_Z > \dfrac{3}{4} \\ O\left(n^{2-2\alpha_Z}L^2(n)\right), & \dfrac{1}{2} < \alpha_Z < \dfrac{3}{4}. \end{cases}$$

Let $\sigma_{n,Z}^2 = \mathbb{E}(\sum_{i=1}^{n} Z_i)^2$. By Mielniczuk and Wu (2004),

$$\sigma_{n,Z}^2 \sim C(\alpha_Z)n^{3-2\alpha_Z}L_Z^2(n),$$

where $C(\alpha_Z)$ is a constant depending on α_Z and $C(\alpha_Z) = O(1)$, so

$$\frac{\Lambda_n}{\sigma_{n,Z}} = o(1), \qquad \text{for } \frac{1}{2} < \alpha_Z < 1.$$

Moreover, Mielniczuk and Wu (2004) also point out that

$$\frac{\sum_{i=1}^{n} Z_i}{\sigma_{n,Z}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\mathcal{N}(0, 1)$ is a standard normal random variable.

2.2 ASSUMPTIONS

Now we list some conditions which are required in the proofs of the theorems.

Assumption 1. *The kernel function $K(\cdot)$ has a compact support $[-1, 1]$.*

Assumption 2. *f_1 is bounded and twice continuously differentiable with bounded derivatives.*

Assumption 3. *The regression function $m(\cdot)$ is at least $(p+2)$ times continuously differentiable at the given point x_0.*

Assumption 4. *The bandwidth h is positive and satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$.*

Assumption 5. *$\psi(\cdot)$ is continuous and is once differentiable almost everywhere, and $\mathbb{E}[\psi'(Z)]$ and $\mathbb{E}[\psi^2(Z)]$ are bounded. Furthermore, there exists a positive constant λ such that for $l = 0$ and 1, $\mathbb{E}[|\psi^{(l)}(Z)|^{\lambda}]$ is bounded.*

Assumption 6. *$\psi(\cdot)$ and $\psi'(\cdot)$ satisfy the following two conditions:*

i) $\mathbb{E}\left[\sup_{|r|\le\delta}\left|\psi'(Z+r) - \psi'(Z)\right| \Big| X = x\right] = o(1)$

ii) $\mathbb{E}\left[\sup_{|r|\le\delta}\left|\psi(Z+r) - \psi(Z) - \psi'(Z)r\right| \Big| X = x\right] = o(\delta)$

as $\delta \to 0$ uniformly in x in a neighbourhood of x_0

Assumption 7. *Let $R_i(z) = f_{i-1,Z}(z - Z_{i,1}) - f_{i-1,Z}(z - Z_{i,0}) + f'_{i-1,Z}(z - Z_{i,0})a_{i-1}\eta_1$. There exists a $C > 0$ such that for sufficiently large $i \in \mathbb{N}$,*

$$\sup_{z\in\mathbb{R}}\left\| \int_{\mathbb{R}} \psi(z)\left[f'_{i-1,Z}(z-\zeta) - f'_{i-1,Z}(z)\right]dz \right\| \le C\|\zeta\|$$

holds for $\zeta = Z_{i,0}$ and $Z_{i,1}$, and

$$\sup_{z\in\mathbb{R}}\left\| \int_{\mathbb{R}} \psi(z)R_i(z)dz \right\| \le Ca_{i-1}^2,$$

where $\|\cdot\| = (\mathbb{E}|\cdot|^2)^{1/2}$.

In Lemma 1, we prove that if Assumption 1 holds then f_∞ has the same properties as f_1. Assumption 5 is useful in deriving the behaviours of $\Delta_{n,1}$, $\Delta_{n,2}$ and $\Delta_{n,4}$, and i) in Assumption 6 contribute to find the asymptotic law of $\Delta_{n,5}$. ii) in Assumption 6 is used to show that $\Delta_{n,3}$ is negligible to $\Delta_{n,1}$. Note that in Assumption 3, $m(\cdot)$ is assumed to be at least $(p + 2)$

times differentiable instead of $(p + 1)$ times, and the reason is that the approximation of $\Delta_{n,2}$ depends on the parity of p (see Lemma 4).

Assumption 7 is used in Lemma 5 to prove that $N_{n,1}$ can be split into a dominated term and a remainder. In fact, this assumption is a special case of C_4 in Mielniczuk and Wu (2004) where they investigate a model with errors based on two variables both following linear processes.

3

Results

In this chapter, the asymptotic behaviour of the SRD case is considered first, and then we discuss the LRD case. For LRD, it is found that the approximation behaves differently regarding to different bandwidths. Before exhibiting the theorems, some notations for simplicity are given. Define

$$\mu_l = \int_{\mathbb{R}} u^l K(u) du \quad \text{and} \quad \nu_l = \int_{\mathbb{R}} u^l K^2(u) du.$$

Further define the matrices $\mathbf{H} = \text{diag}(1, h, h^2, \cdots, h^p)$,

$$\boldsymbol{S}_p = \left(\mu_{j+k}\right)_{0 \le j,k \le p} = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix} \qquad \text{and} \qquad \widetilde{\mathbf{S}}_p = \left(\mu_{j+k+1}\right)_{0 \le j,k \le p}$$

$$\boldsymbol{S}_p^* = \left(\nu_{j+k}\right)_{0\le j,k\le p} = \begin{pmatrix} \nu_0 & \nu_1 & \cdots & \nu_p \\ \nu_1 & \nu_2 & \cdots & \nu_{p+1} \\ \vdots & \vdots & & \vdots \\ \nu_p & \nu_{p+1} & \cdots & \nu_{2p} \end{pmatrix} \quad \text{and} \quad \widetilde{\mathbf{S}}_p^* = \left(\nu_{j+k+1}\right)_{0\le j,k\le p}$$

3.1 Consistency

The existence of a consistent solution is established first before the convergence rate properties of those solutions is analysed in the SRD and LRD cases respectively.

Theorem 1. *If Assumptions 1 – 6 hold then the solution to the minimisation problem is consistent. That is,*

$$\mathbf{H}\left(\widehat{\beta} - \beta\right) \xrightarrow{p} 0.$$

3.2 SRD SEQUENCES

In the SRD case, it satisfies that $\alpha_Z > 1$, and it is found that $M_{n,1}$ dominates the asymptotic law.

Theorem 2. *If* $\sum_{i=n}^{\infty} |a_i| < \infty$ *and Assumptions 1-7 hold, then*

$$\sqrt{nh}\left[\mathbf{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - Bias(\mathbf{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}))\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\mathbb{E}[\psi^2(Z)]}{\mathbb{E}\left[\psi'(Z)\right]^2 f_X(x_0)} \boldsymbol{S}_p^{-1}\boldsymbol{S}_p^*\boldsymbol{S}_p^{-1}\right)$$

where the Bias term is,

$$\frac{h^{1+p}}{f_X(x_0)}\left(\boldsymbol{S}_p^{-1} - h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p + h\left(f_X'(x_0)\theta_{p+1} + f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right],$$

where $\theta_p = m^{(p)}(x_0)/p!$.

The theorem shows that for odd and even p, the only difference is the bias term, and the reason of this is that odd moments of $K(\cdot)$ are zero (see Lemma 4). Lemma 6 shows that the asymptotic distribution $N\left[0, \boldsymbol{\Sigma}_1(x_0)\right]$ is only determined by $M_{n,1}$. Moreover, the above result is equivalent to the asymptotic behaviour under the case of independence.

Based on the theorem, we can derive the optimal bandwidths for those two conditions. When p is odd, then the asymptotic mean squared error (AMSE) of $\hat{m}(x_0)$ is

$$\text{AMSE}(x_0) = \frac{h^{2+2p}\left[\mu_{1+p}m^{(p+1)}(x_0)\right]^2}{\left[(p+1)!\right]^2} + \frac{\boldsymbol{e}_1^T\boldsymbol{S}_p^{-1}\boldsymbol{S}_p^*\boldsymbol{S}_p^{-1}\boldsymbol{e}_1\mathbb{E}\left[\psi^2(Z)\right]}{nhf_X(x_0)\mathbb{E}^2\left[\psi'(Z)\right]},$$

where $\mathbb{E}^2(\cdot) \equiv [\mathbb{E}(\cdot)]^2$. Thus, the optimal bandwidth is

$$h_{opt} = \left\{\frac{\boldsymbol{e}_1^T\boldsymbol{S}_p^{-1}\boldsymbol{S}_p^*\boldsymbol{S}_p^{-1}\boldsymbol{e}_1\mathbb{E}\left[\psi^2(Z)\right]\cdot\left[(p+1)!\right]^2}{n(2+2p)f_X(x_0)\left[\mu_{1+p}m^{(p+1)}(x_0)\right]^2\mathbb{E}^2\left[\psi'(Z)\right]}\right\}^{\frac{1}{3+2p}}.$$

Note that when $p = 1$, the optimal bandwidth is of order $n^{-1/5}$ which is the optimal order of the kernel density estimator under independence.

3.3 LRD SEQUENCES

Assume that $h = n^{-\alpha_h}L_h(n)$, where $\alpha_h > 0$ and $L_h(n)$ is a slowly varying function. When $\alpha_h > 2 - 2\alpha_Z$, then $h^{1/2}\sigma_{n,Z} = o(n^{1/2})$, and we regard this case as small bandwidths. In this case, it is found that the asymptotic behaviour of the estimator is the same as the SRD case. When $0 < \alpha_h < 2/3 - (2/3)\alpha_Z$, then $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$, and this case is regarded as large bandwidths. In this case, it follows that the asymptotic behaviour is determined by $N_{n,1}$, then with a particular norming sequence the asymptotic behaviour of the estimator follows a different Gaussian distribution from the SRD case.

Theorem 3. *If $\sum_{i=n}^{\infty}|a_i| = \infty$ and Assumptions 1-7 hold, then*

i) *when $h^{1/2}\sigma_{n,Z} = o(n^{1/2})$, then the convergence result in Theorem 2 holds;*

ii) *when $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$, then,*

$$\frac{n}{\sigma_{n,Z}}\left[\mathbf{H}_*^{-1}\mathbf{H}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - Bias_{LRD}(\mathbf{H}_*^{-1}\mathbf{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}))\right] \xrightarrow{d} N\left(0, \frac{\boldsymbol{S}_p^{-1}\boldsymbol{c}_p^*(\boldsymbol{c}_p^*)^T\boldsymbol{S}_p^{-1}C_{LRD}^2}{\mathbb{E}\left[\psi'(Z)\right]^2 f_X^2(x_0)}\right)$$

where the Bias term is defined,

$$Bias_{LRD}(\mathbf{H}_*^{-1}\mathbf{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \mathbf{H}_*^{-1}\frac{h^{1+p}}{f_X(x_0)}\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p + h\left(f_X'(x_0)\theta_{p+1} + f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right]$$

where $\mathbf{H}_* = diag\left(h^{k(j)}\right)_{0\le j\le p}$ and $k(j)$ is 0 or 1 if j is even or odd respectively, the vector $\boldsymbol{c}_p^*$ is defined,

$$\boldsymbol{c}_p^* = \left(\mu_j f_X(x_0) + \mu_{j+1}f_X'(x_0)\right)_{0\le j\le p} = \boldsymbol{c}_p f_X(x_0) + \widetilde{\mathbf{c}}_p f_X'(x_0)$$

and the constant $C_{LRD} = \int_{\mathbb{R}} w\phi(w) \int_{\mathbb{R}} \psi(z) f_\eta(z - s_z w)\, dz\, dw$ *where* $s_Z = \|Z_{i,i-1}\|$.

The theorem implies that when errors are LRD, if the bandwidth is small enough, the asymptotic behaviour of $\hat{\boldsymbol{\beta}}$ is still determined by $M_{n,1}$ and is the same as the SRD case. While if the bandwidth is large enough then the estimator behaves differently. The reason is that with the norming sequence $n/\sigma_{n,Z}$, $M_{n,1}$ is negligible to $N_{n,1}$ resulting in that the asymptotic law is dominated by $N_{n,1}$ (see Lemma 6). This result is analogous to the the Nadaraya-Watson estimator in regression with LRD errors (see Mielniczuk and Wu (2004)). Furthermore, note that when the bandwidth is large, both the bias terms and the variances are different between odd and even p.

We can derive the corresponding AMSE of $\hat{m}(x_0)$ in the case of large bandwidths based on above results.The AMSE of $\hat{m}(x_0)$ is

$$\mathrm{AMSE}(x_0) = \frac{h^{2+2p}\left[\mu_{1+p} m^{(p+1)}(x_0)\right]^2}{\left[(p+1)!\right]^2} + \frac{\sigma_{n,Z}^2 C_{LRD}^2 \boldsymbol{e}_1^T \boldsymbol{S}_p^{-1} \boldsymbol{c}_p^* (\boldsymbol{c}_p^*)^T \boldsymbol{S}_p^{-1} \boldsymbol{e}_1}{n^2 \left(\mathbb{E}\left[\psi'(Z)\right]\right)^2}. \tag{3.1}$$

Note that $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$ implies $\sigma_{n,Z}^2/n > 1/h^3$, and this indicates that there does not exist the minimum value of the AMSE, so the optimal bandwidth in this case is not able to be derived. Since $0 < \alpha_h < 2/3 - (2/3)\alpha_Z$ indicates $h \in (n^{(2/3)\alpha_Z - 2/3} L_h(n), L_h(n))$, then a feasible approach to minimize the AMSE in (3.1) is choosing a sufficiently small h in the interval.

It should be pointed out that the asymptotic behaviour of $\hat{\boldsymbol{\beta}}$ in the case that $2/3-(2/3)\alpha_Z < \alpha_h < 2 - 2\alpha_Z$ is not exhibited here. In this case, $n^{1/2} = o(h^{1/2}\sigma_{n,Z})$ holds while $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$ does not hold. Mielniczuk and Wu (2004) show that when $n^{1/2} = o(h^{1/2}\sigma_{n,Z})$ the limit of the Nadaraya-Watson estimator is dominated by a conditional expectation term, and then follows a Gaussian distribution asymptotically; however, for the local polynomial M-estimator, we find that neither $M_{n,1}$ nor $N_{n,1}$ dominate the asymptotic law so the behaviour is more intricate. This is beyond the scope of this thesis and needs to be studied in the future work.

4

Conclusion

In this thesis, we extend the method proposed by Wu and Mielniczuk (2002) and Mielniczuk and Wu (2004) to the local polynomial M-estimator in random design regression with dependent errors. Consistent with similar results in the literature, it is found that the asymptotic behaviour of the estimator is determined by the strength of dependence. The estimator is decomposed into $\Delta_{n,k}$ ($k = 1, 2, 3, 4$ and 5) by Taylor's expansion, then we find that $\Delta_{n,1}$ determines the asymptotic behaviour. Further, $\Delta_{n,1}$ can be decomposed into a martingale term $M_{n,1}$ and a conditional expectation term $N_{n,1}$. When errors are SRD, we prove that $M_{n,1}$ dominates the asymptotic behaviour, and in terms of the martingale central limit theorem, with a norming sequence $\sqrt{nh}$ the estimator asymptotically follows a Gaussian distribution. When errors are LRD, the bandwidth h is found to be a factor which influences the asymptotic law: if the bandwidth is small enough, then the scenario is the same as the SRD case; however, if the bandwidth is large enough, then $N_{n,1}$ dominates the limit, and this results in a different Gaussian distribution where the norming sequence is $n/\sigma_{n,Z}$. Furthermore, we find that for both SRD and LRD cases, odd and even p behave differently. The reason of this is that if the kernel function $K(\cdot)$ is symmetric then its odd moments are zero. To avoid this issue, Taylor's theorem is used to obtain even moments of $K(\cdot)$ replacing odd moments.

For the SRD case, the optimal bandwidth minimizing the AMSE of $\hat{m}(x_0)$ is determined. If p is odd, then the optimal bandwidth is order of $n^{-1/(3+2p)}$; and when p is even, then the optimal bandwidth is order of $n^{-1/(5+2p)}$. This result is the same in the case of independent errors; particularly, when $p = 1$, the optimal bandwidth is order of $n^{-1/5}$ which is the optimal order of the kernel density estimator under independence. For the LRD case, the optimal bandwidth cannot be derived when $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$ since the corresponding AMSE does not have the minimum value. The only way to minimize the AMSE is choosing the sufficiently small bandwidth h in $(n^{(2/3)\alpha_Z - 2/3}L_h(n), L_h(n))$.

There are some potential extensions of this thesis. One extension is to investigate the asymptotic law for the case that $2/3 - (2/3)\alpha_Z < \alpha_h < 2 - 2\alpha_Z$. A reasonable conjecture of this case is that with a particular norming sequence both $M_{n,1}$ and $N_{n,1}$ contribute to the asymptotic behaviour, and the interval in which the optimal bandwidth locates in this case is able to be depicted. The other extension is to extend the regression model used in this thesis to a more general one. The general model

$$Y_i = m(X_i) + \varepsilon_i \tag{4.1}$$

assumes that errors $\varepsilon_i = G(Z_i, X_i)$ where $\{Z_i\}_{i=1}^n$ are latent variables and both $\{Z_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ are linear processes. Mielniczuk and Wu (2004) study the Nadaraya-Watson estimator on that model (4.1), and they find that the asymptotic behaviour depends on the strength of dependence of both $\{Z_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$. Therefore, we conjecture that the asymptotic behaviour of the local polynomial M-estimator is also determined by the strength of dependence of both $\{Z_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$. Moreover, some bandwidth selection techniques, such as the plug-in method, can be applied in practice to derive the optimal bandwidth based on our results.

5

Proofs

This chapter gives the proofs of the theorems, and to this end, some useful lemmas are given. Before stating the lemmas, we discuss $R(X_i)$ first. By Assumption 3 and Taylor's thoerem with Lagrange remainders, due to the fact that $|X_i - x_0| < h$, then there exists a ξ_1 between X_i and x_0 satisfying

$$\begin{aligned}
R(X_i) =& m(X_i) - \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(X_i - x_0)^j \\
=& \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(X_i - x_0)^j + \frac{m^{(p+1)}(\xi_1)}{(p+1)!}(X_i - x_0)^{1+p} - \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(X_i - x_0)^j \\
=& \frac{m^{(p+1)}(x_0)}{(p+1)!}(X_i - x_0)^{1+p} + \frac{m^{(p+2)}(x_0)}{(p+2)!}(X_i - x_0)^{2+p} \\
& + \frac{1}{(p+2)!}(X_i - x_0)^{2+p}\left(m^{(p+2)}(\xi_1) - m^{(p+2)}(x_0)\right) \qquad (5.1)
\end{aligned}$$

and further, we get

$$\sup_{X_i : |X_i - x_0| < h} R^2(X_i) \leq O_p(h^{2+2p}).$$

Since all $\Delta_{n,k}$ ($k = 1, 2, 4$ and 5) are multivariate, for convenience, we only investigate the behaviour of the element of $\Delta_{n,k}$ throughout the lemmas. Let

$$
\begin{aligned}
\Delta_{n,1}^{j} &= \frac{1}{nh}\sum_{i=1}^{n}\psi(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j \\
&= \frac{1}{nh}\sum_{i=1}^{n}\Big\{\psi(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j-\mathbb{E}\Big[\psi(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j|\mathcal{F}_{i-1}\Big]\Big\} \\
&\quad+\frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\Big[\psi(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j|\mathcal{F}_{i-1}\Big] \\
&\equiv \frac{M_{n,1}^{j}}{nh}+\frac{N_{n,1}^{j}}{nh}.
\end{aligned}
\tag{5.2}
$$

Give the analogous definitions to $\Delta_{n,k}^{j}$ ($k=2,4$ and 5), so we get

$$
\Delta_{n,k}^{j}-\mathbb{E}\Delta_{n,k}^{j}=\frac{M_{n,k}^{j}}{nh}+\frac{\bar{N}_{n,k}^{j}}{nh}. \tag{5.3}
$$

Lemma 1. *If the density function f_1 of (η_i,γ_i) is Lipschitz continuous, then the density function f_∞ of (Z_i,X_i) satisfies*

$$
f_\infty(z,x)=\mathbb{E}f_1(z-Z_{i,i-1},x) \tag{5.4}
$$

and is Lipschitz continuous as well. Moreover, if Assumption 2 holds then f_∞ also has the same properties.

Proof of Lemma 1. Since $\{X_i\}_{i=1}^n$ are independent of $\{Z_i\}_{i=1}^n$, then to prove (5.4) is equivalent to showing that

$$
f_Z(z)=\mathbb{E}f_\eta(z-Z_{i,i-1})
$$

and f_Z is Lipschitz continuous. Assume that $C>0$ is the Lipschitz constant of f_η, then by Lipschitz continuity, we have $\mathbb{E}|f_\eta(z-Z_{i,i-1})-f_\eta(z)|\le\mathbb{E}\big(C|Z_{i,i-1}|\big)\le C\cdot\big[\mathbb{E}(Z_{i,i-1})^2\big]^{\frac{1}{2}}<\infty$. Thus, it is clear that $\mathbb{E}f_\eta(z-Z_{i,i-1})$ is finite. Denote F_Z and F_η the cumulative functions of Z_i and η_i respectively and $\tilde{f}_\eta$ the density function of $Z_{i,i-1}$. Obviously, since $a_0=1$, then F_Z is a convolution of F_η and $\tilde{f}_\eta$,

$$
F_Z(z)=\int_{\mathbb{R}}F_\eta(z-y)\tilde{f}_\eta(y)dy.
$$

Thus, for a $\xi > 0$, by Lagrange's mean value theorem,

$$\begin{aligned}
&|F_Z(z+\xi) - F_Z(z) - \xi \mathbb{E} f_\eta(z - Z_{i,i-1})| \\
&= |\int_{\mathbb{R}} F_\eta(z-y+\xi)\tilde{f}_\eta(y)dy - \int_{\mathbb{R}} F_\eta(z-y)\tilde{f}_\eta(y)dy - \xi \int_{\mathbb{R}} f_\eta(z-y)\tilde{f}_\eta(y)dy| \\
&\le \int_{\mathbb{R}} |F_\eta(z-y+\xi) - F_\eta(z-y) - \xi f_\eta(z-y)| \cdot \tilde{f}_\eta(y)dy \\
&= \int_{\mathbb{R}} |\xi f_\eta(z-y+\xi\delta) - \xi f_\eta(z-y)| \cdot \tilde{f}_\eta(y)dy \\
&\le C\xi^2,
\end{aligned}$$

where $-\xi < \delta < 0$. Hence, let $\xi \to 0$, then

$$\lim_{\xi \to 0} \left| \frac{F_Z(z+\xi) - F_Z(z)}{\xi} - \mathbb{E} f_\eta(z - Z_{i,i-1}) \right| = 0,$$

and this indicates $f_Z(z) = \mathbb{E} f_\eta(z - Z_{i,i-1})$. Thus, $|f_Z(z) - f_Z(y)| \le \mathbb{E}|f_\eta(z - Z_{i,i-1}) - f_\eta(y - Z_{i,i-1})| \le C|z-y|$, which implies the Lipschitz continuity of f_Z. Similarly, we are able to find that $f_Z'(z) = \mathbb{E} f_\eta'(z - Z_{i,i-1})$ and $f_Z''(z) = \mathbb{E} f_\eta''(z - Z_{i,i-1})$ then to show that f_Z is twice continuously differentiable with bounded derivatives, and further, proving that f_∞ satisfies the condition of Assumption 2. Details are omitted here. □

Lemma 2. *Let $\{Z_i\}_{i\in\mathbb{Z}}$ be a linear process and g a measureable function such that $\mathbb{E}\,|g(Z)| < \infty$ then,*

$$\mathbb{E}\left| \frac{1}{n}\sum_{i=1}^{n} \int_{\mathbb{R}} g(z) f_\eta(z - Z_{i,i-1})\, dz - \int_{\mathbb{R}} g(z) f_Z(z)\, dz \right| = o(1)$$

Proof of Lemma 2. The linear process $\{Z_i\}_{i\in\mathbb{Z}}$ is strictly stationary and ergodic (c.f. Section 2.1.1.2 of Beran et al. (2013)). This also implies that $\{Z_{i,i-1}\}_{i\in\mathbb{Z}}$ is also strictly stationary and ergodic. Define $W_i = \int_{\mathbb{R}} g(z) f_\eta(z - Z_{i,i-1})\, dz$, then by Theorem 1.3.3 of Taniguchi and Kakizawa (2000), $\{W_i\}_{i\in\mathbb{Z}}$ is stationary and ergodic. Further,

$$\begin{aligned}
\mathbb{E}\,|W| &\le |g(z)| \int_{\mathbb{R}} \mathbb{E} f_\eta(z - Z_{i,i-1})\, dz \\
&= \int_{\mathbb{R}} |g(z)| f_Z(z)\, dz < \infty.
\end{aligned}$$

The result of the Lemma therefore follows by the mean ergodic theorem, (c.f. Corollary 3.5.2 of Stout (1974)). □

Lemma 3. *If Assumption 5 holds, then $(nh^{1+2j})^{-1/2} M_{n,1}^j \to^d N[0, \sigma_{1,j}^2(x_0)]$, where*

$$\sigma_{1,j}^2(x_0) = \nu_{2j} f_X(x_0) \int_{\mathbb{R}} \psi^2(z) f_Z(z) dz$$

Proof of Lemma 3. Let

$$\zeta_{1,i,j} = \frac{1}{\sqrt{nh^{1+2j}}}\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j \quad \text{and} \quad \xi_{1,i,j} = \zeta_{1,i,j} - \mathbb{E}(\zeta_{1,i,j}|\mathcal{F}_{i-1}). \tag{5.5}$$

By the martingale central limit theorem, it is required to show that the following two conditions hold:

$$\sum_{i=1}^{n} \mathbb{E}(\xi^2_{1,i,j}|\mathcal{F}_{i-1}) \xrightarrow{P} \sigma^2_{1,j}(x_0) \tag{5.6}$$

and the Lindeberg condition

$$n\mathbb{E}[\xi^2_{1,i,j}\mathbf{I}_{[|\xi_{1,i,j}|>\varepsilon]}] = o(1), \tag{5.7}$$

where ε is any positive constant. Exploit the fact that $Z_{i,i-1} = O_p(1)$ and use a first order Taylor expansion of ψ to yield,

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j|\mathcal{F}_{i-1}\Big]\Big| \\
=&\Big|\iint_{\mathbb{R}^2}\psi(z + Z_{i,i-1})K\Big(\frac{x - x_0}{h}\Big)(x - x_0)^j f_1(z, x)dzdx\Big| \\
=&\Big|\iint_{\mathbb{R}^2}\psi(z + Z_{i,i-1})K\Big(\frac{x - x_0}{h}\Big)(x - x_0)^j f_X(x)f_\eta(z)dzdx\Big| \\
=&\Big|h^{1+j}\iint_{\mathbb{R}^2}\psi(z + Z_{i,i-1})K(u)u^j f_X(uh + x_0)f_\eta(z)dzdu\Big| \\
\le& h^{1+j}\int_{\mathbb{R}}\Big|K(u)u^j f_X(uh + x_0)\Big|du\cdot\Big\{|\mathbb{E}[\psi(\eta)]| + \big|Z_{i,i-1}\big|\,|\mathbb{E}[\psi'(\eta)]| \\
&\qquad + \big|Z_{i,i-1}\big|\int_{\mathbb{R}} f_\eta(z)\big|\psi'(z + \tau_{z,i}Z_{i,i-1}) - \psi'(z)\big|\,dz\Big\} \\
=& O_p(h^{1+j}),
\end{aligned} \tag{5.8}$$

where $|\tau_{z,i}| < 1$ forms the remainder of the Taylor expansion. The last line follows since the expectations are assumed bounded and ψ' is assumed to be Lipschitz. Therefore, we have the following result,

$$\begin{aligned}
&\Big|\sum_{i=1}^{n}\mathbb{E}[\zeta^2_{1,i,j}|\mathcal{F}_{i-1}] - \sum_{i=1}^{n}\mathbb{E}[\xi^2_{1,i,j}|\mathcal{F}_{i-1}]\Big| \\
\le& \sum_{i=1}^{n}\Big|\mathbb{E}[\zeta^2_{1,i,j}|\mathcal{F}_{i-1}] - \mathbb{E}[\xi^2_{1,i,j}|\mathcal{F}_{i-1}]\Big| \\
=& \sum_{i=1}^{n}\Big|\mathbb{E}[\zeta^2_{1,i,j}|\mathcal{F}_{i-1}] - \mathbb{E}[\zeta^2_{1,i,j}|\mathcal{F}_{i-1}] + 2\mathbb{E}^2[\zeta_{1,i,j}|\mathcal{F}_{i-1}] - \mathbb{E}^2[\zeta_{1,i,j}|\mathcal{F}_{i-1}]\Big| \\
=& \sum_{i=1}^{n}\mathbb{E}^2[\zeta_{1,i,j}|\mathcal{F}_{i-1}] \\
=& \sum_{i=1}^{n}\frac{1}{nh^{1+2j}}\mathbb{E}^2\Big[\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j|\mathcal{F}_{i-1}\Big] \le \frac{O_p(h^{2+2j})}{h^{1+2j}} = O_p(h) = o_p(1).
\end{aligned} \tag{5.9}$$

Thus, to show (5.6) is equivalent to showing

$$\sum_{i=1}^{n} \mathbb{E}(\zeta_{1,i,j}^2 | \mathcal{F}_{i-1}) \xrightarrow{P} \sigma_{1,j}^2(x_0). \tag{5.10}$$

By Lemma 1 in Brown (1971), (5.10) is equivalent to

$$\lim_{n\to\infty} \mathbb{E}\Big| \sum_{i=1}^{n} \mathbb{E}(\zeta_{1,i,j}^2 | \mathcal{F}_{i-1}) - \sigma_{1,j}^2(x_0) \Big| = 0. \tag{5.11}$$

Now, we prove (5.11). In view of the continuity of f_X, we have

$$\lim_{n\to\infty} \iint_{\mathbb{R}^2} \psi^2(z) K^2(u) u^{2j} f_\infty(z, uh + x_0) dz du = \nu_{2j} f_X(x_0) \int_{\mathbb{R}} \psi^2(z) f_Z(z) dz,$$

so it suffices to show $\lim_{n\to\infty} \mathbb{E}| \sum_{i=1}^n \mathbb{E}(\zeta_{1,i,j}^2 | \mathcal{F}_{i-1}) - \iint_{\mathbb{R}^2} \psi^2(z) K^2(u) u^{2j} f_\infty(z, uh + x_0) dz du| = 0$. Note that due to Lemma 2 and Assumption 5,

$$\begin{aligned}
&\mathbb{E}\Big| \sum_{i=1}^{n} \mathbb{E}[\zeta_{1,i,j}^2 | \mathcal{F}_{i-1}] - \iint_{\mathbb{R}^2} \psi^2(z) K^2(u) u^{2j} f_\infty(z, uh + x_0) dz du \Big| \\
&= \mathbb{E}\Big| \sum_{i=1}^{n} \frac{1}{nh^{1+2j}} \iint_{\mathbb{R}^2} \psi^2(z) K^2\Big(\frac{x - x_0}{h}\Big) (x - x_0)^{2j} f_1(z - Z_{i,i-1}, x) dz dx \\
&\quad - \iint_{\mathbb{R}^2} \psi^2(z) K^2(u) u^{2j} f_\infty(z, uh + x_0) dz du \Big| \\
&= \mathbb{E}\Big| \int_{\mathbb{R}} K^2(u) u^{2j} f_X(uh + x_0) du \int_{\mathbb{R}} \psi^2(z) \sum_{i=1}^{n} \frac{1}{n} \big(f_\eta(z - Z_{i,i-1}) - f_Z(z) \big)\, dz \Big| \\
&= f_X(x_0) \int_{\mathbb{R}} K^2(u) u^{2j} du \cdot \mathbb{E}\Big| \int_{\mathbb{R}} \psi^2(z) \sum_{i=1}^{n} \frac{1}{n} \big(f_\eta(z - Z_{i,i-1}) - f_Z(z) \big)\, dz \Big| \cdot (1 + O(h)) = o(1).
\end{aligned}$$

Therefore, (5.11) is proven. For the Lindeberg condition (5.7), we can use the Corollary 5.9.2 in Chow and Teicher (1997),

$$\begin{aligned}
n\mathbb{E}[\xi_{1,i,j}^2 \mathbf{I}_{[|\xi_{1,i,j}| > \varepsilon]}] &\le 4n\mathbb{E}[\zeta_{1,i,j}^2 \mathbf{I}_{[|\zeta_{1,i,j}| > \frac{\varepsilon}{2}]}] \\
&= 4 \iint_{|\psi(z) K(u) u^j| > \frac{\sqrt{nh}\varepsilon}{2}} \psi^2(z) K^2(u) u^{2j} f_\infty(z, uh + x_0) dz du.
\end{aligned}$$

Since $nh \to \infty$ as $n \to \infty$ and $K(\cdot)$ has a compact support, then we can see $n\mathbb{E}[\xi_{1,i,j}^2 \mathbf{I}_{[|\xi_{1,i,j}| > \varepsilon]}]$ $= o(1)$ almost everywhere. Thus, (5.7) holds. □

Note that $\|Z_{i,i-1}\|^2 = \|\eta\|^2 \sum_{t=1}^{\infty} a_t^2$,

$$\Big\| \sum_{i=1}^{n} Z_{i,i-1} \Big\| = \begin{cases} O(\sqrt{n}), & \text{under SRD;} \\ O(n^{\frac{3}{2} - \alpha_Z} L_Z(n)), & \text{under LRD.} \end{cases} \tag{5.12}$$

Lemma 4. *When Assumptions 5 and 6 hold, for both SRD and LRD cases, if $nh^3 \to \infty$, then we have the following three results:*

i)

$$\Delta_{n,2}^j = \begin{cases} h^{1+j+p}\mathbb{E}[\psi'(Z)]\mu_{1+j+p} f_X(x_0)\dfrac{m^{(p+1)}(x_0)}{(p+1)!} + o_p(h^{1+j+p}), & j+p \text{ is odd} \\ h^{2+j+p}\mathbb{E}[\psi'(Z)]\mu_{2+j+p}\left(f_X'(x_0)\dfrac{m^{(p+1)}(x_0)}{(p+1)!} + f_X(x_0)\dfrac{m^{(p+2)}(x_0)}{(p+2)}\right) + o_p(h^{2+j+p}), & j+p \text{ is even} \end{cases}$$

ii)

$$\Delta_{n,4}^j = \begin{cases} h^j \mu_j f_X(x_0)\mathbb{E}[\psi'(Z)] + o_p(h^j), & j \text{ is even} \\ h^{1+j}\mu_{1+j} f_X'(x_0)\mathbb{E}[\psi'(Z)] + o_p(h^{1+j}), & j \text{ is odd} \end{cases}$$

iii)

$$\Delta_{n,5}^j = o_p(h^{1+j}).$$

Proof of Lemma 4. i) By (5.3), we have $\Delta_{n,2}^j = \Delta_{n,2}^j - \mathbb{E}\Delta_{n,2}^j + \mathbb{E}\Delta_{n,2}^j = (nh)^{-1}M_{n,2}^j + (nh)^{-1}\bar{N}_{n,2}^j + \mathbb{E}\Delta_{n,2}^j$.

$$\begin{aligned}
\mathbb{E}\Delta_{n,2}^j &= \frac{1}{nh}\sum_{i=1}^n \mathbb{E}\left[\psi'(Z_i)R(X_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^j\right] \\
&= \frac{1}{h}\iint_{\mathbb{R}^2} \psi'(z)R(x)K\left(\frac{x-x_0}{h}\right)(x-x_0)^j f_\infty(z,x)dzdx \\
&= \frac{1}{h}\int_{\mathbb{R}} R(x)K\left(\frac{x-x_0}{h}\right)(x-x_0)^j f_X(x)dx \int_{\mathbb{R}} \psi'(z)f_Z(z)dz \\
&= \frac{1}{h}\int_{\mathbb{R}} R(x)K\left(\frac{x-x_0}{h}\right)(x-x_0)^j f_X(x)dx \cdot \mathbb{E}[\psi'(Z)] \\
&= h^j \int_{\mathbb{R}} R(uh+x_0)K(u)u^j f_X(uh+x_0)du \cdot \mathbb{E}[\psi'(Z)] \\
&= h^{p+j+1}\mathbb{E}[\psi'(Z)]\int_{\mathbb{R}} K(u)u^{p+j+1}\left[\frac{m^{(p+1)}(x_0)}{(p+1)!} + \frac{m^{(p+2)}(x_0)}{(p+2)!}hu\right] f_X(x_0+hu)\,du \\
&\quad + \frac{h^{p+j+2}}{(p+2)!}\mathbb{E}[\psi'(Z)]\int_{\mathbb{R}} K(u)\left[m^{(p+2)}(x_0+\tau hu) - m^{(p+2)}(x_0)\right] f_X(x_0+hu)\,du \\
&= \frac{h^{p+j+1}\mathbb{E}[\psi'(Z)]}{(p+1)!}\left[\mu_{p+j+1}m^{(p+1)}(x_0)f_X(x_0) + h\mu_{p+j+2}\left(f_X'(x_0)m^{(p+1)}(x_0) + \frac{m^{(p+2)}(x_0)f_X(x_0)}{(p+2)}\right)\right] \\
&\quad + o(h^{p+j+2}) \qquad (5.13)
\end{aligned}$$

For $M_{n,2}^j$, we can show that $M_{n,2}^j = O_p(\sqrt{nh^{3+2j+2p}})$ by the method of Lemma 3. Let

$$\zeta_{2,i,j} = \frac{1}{\sqrt{nh^{3+j+p}}}\psi'(Z_i)R(X_i)K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^j \quad \text{and} \quad \xi_{2,i,j} = \zeta_{2,i,j} - \mathbb{E}(\zeta_{2,i,j}|\mathcal{F}_{i-1}).$$

Similarly to earlier in the proof of Lemma 3, exploit the Lipschitz smoothness in ψ'. Note from (5.13) that $\int_{\mathbb{R}} R(uh+x_0)K(u)u^j f_X(uh+x_0)du = O(h^{p+1})$

$$
\begin{aligned}
&\left|\mathbb{E}\left[\psi'(Z_i)R(X_i)K\left(\frac{X_i-x_0}{h}\right)(X_i-x_0)^j|\mathcal{F}_{i-1}\right]\right| \\
&= \left|\int_{\mathbb{R}} \psi'(z+Z_{i,i-1})f_\eta(z)dz \cdot h^{1+j}\int_{\mathbb{R}} R(x_0+hu)K(u)u^j f_X(uh+x_0)\,du\right| \\
&\le \left(|\mathbb{E}\psi'(\eta)| + \int_{\mathbb{R}} \left|\psi'(z+Z_{i,i-1})-\psi'(z)\right| f_\eta(z)\,dz\right) h^{1+j}\left|\int_{\mathbb{R}} R(x_0+hu)K(u)u^j f_X(uh+x_0)\,du\right| \\
&= \begin{cases} O_p(h^{2+j+p}), & p+j \text{ odd} \\ O_p(h^{3+j+p}), & p+j \text{ even} \end{cases}
\end{aligned}
$$

Hence $\sum_{i=1}^n \mathbb{E}^2(\zeta_{2,i,j}|\mathcal{F}_{i-1}) \le (nh^{3+2j+2p})^{-1}\sum_{i=1}^n O_p(h^{4+2j+2p}) = O_p(h) = o_p(1)$. So (5.9) and (5.10) imply that it suffices to show the convergence of the conditional variance.

$$
\begin{aligned}
&\mathbb{E}\left|\sum_{i=1}^n \mathbb{E}[\zeta_{2,i,j}^2|\mathcal{F}_{i-1}] - \int_{\mathbb{R}}[\psi'(z)]^2 f_Z(z)\,dz f_X(x_0)\nu_{2+2j+2p}\left(\frac{m^{(p+1)}(x_0)}{(p+1)!}\right)^2\right| \\
&\le \mathbb{E}\left|\frac{1}{n}\int_{\mathbb{R}}[\psi'(z)]^2\sum_{i=1}^n f_\eta(z-Z_{i,i-1})\,dz\left(\int_{\mathbb{R}}\left(\frac{m^{(p+1)}(x_0)}{(p+1)!}\right)^2 K^2(u)u^{2+2j+2p}f_X(x_0+hu)du + o(1)\right)\right. \\
&\qquad \left. - \int_{\mathbb{R}}[\psi'(z)]^2 f_Z(z)\,dz f_X(x_0)\nu_{2+2j+2p}\left(\frac{m^{(p+1)}(x_0)}{(p+1)!}\right)^2\right| \\
&= \mathbb{E}\left|\int_{\mathbb{R}}[\psi'(z)]^2\frac{1}{n}\sum_{i=1}^n\left(f_\eta(z-Z_{i,i-1}) - f_Z(z)\right)\,dz\left(\frac{m^{(p+1)}(x_0)}{(p+1)!}\right)^2 f_X(x_0)\nu_{2+2j+2p}(1+o(1))\right| \\
&= o(1).
\end{aligned}
$$

The proof of the Lindeberg condition follows the method of Lemma 3. Therefore, $M_{n,2}^j = O_p(\sqrt{nh^{3+2j+2p}}) \Rightarrow (nh^{2+j+p})^{-1}M_{n,2} = O_p((nh)^{-1/2}) = o_p(1)$. Use Lemma 2 and Assumption

6 to bound $\bar{N}_{n,2}$,

$$\begin{aligned}
\bar{N}_{n,2}^{j} &= \sum_{i=1}^{n}\left\{\mathbb{E}\left[\psi'(Z_i)R(X_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j|\mathcal{F}_{i-1}\right]-h\mathbb{E}\Delta_{n,2}^{j}\right\}\\
&= \sum_{i=1}^{n}\Bigg\{\iint_{\mathbb{R}^2}\psi'(z)R(x)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_1(z-Z_{i,i-1},x)dzdx\\
&\qquad -\iint_{\mathbb{R}^2}\psi'(z)R(x)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_\infty(z,x)dzdx\Bigg\}\\
&= \int_{\mathbb{R}}R(x)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)\,dx\int_{\mathbb{R}}\psi'(z)\sum_{i=1}^{n}\big[f_\eta(z-Z_{i,i-1})-f_Z(z)\big]dz\\
&= \int_{\mathbb{R}}R(x)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)\,dx\int_{\mathbb{R}}\psi'(z)\sum_{i=1}^{n}\big[f_\eta(z-Z_{i,i-1})-f_Z(z)\big]dz\\
&= h^{1+j}\int_{\mathbb{R}}R(x_0+hu)K(u)u^j f_X(x_0+hu)\,du\int_{\mathbb{R}}\psi'(z)\sum_{i=1}^{n}\big[f_\eta(z-Z_{i,i-1})-f_Z(z)\big]dz
\end{aligned}$$

Then by Lemma 2 and Assumption 5,

$$\mathbb{E}\left|\bar{N}_{n,2}^{j}\right| = \begin{cases} o(nh^{2+j+p}), & \text{when } j+p \text{ is odd;}\\ o(nh^{3+j+p}), & \text{when } j+p \text{ is even.}\end{cases}$$

This completes the proof of $\Delta_{n,2}$.

ii) By (5.3), it follows $\Delta_{n,4}^{j} = \Delta_{n,4}^{j} - \mathbb{E}\Delta_{n,4}^{j} + \mathbb{E}\Delta_{n,4}^{j} = (nh)^{-1}M_{n,4}^{j} + (nh)^{-1}\bar{N}_{n,4}^{j} + \mathbb{E}\Delta_{n,4}^{j}$. We first discuss $\mathbb{E}\Delta_{n,4}^{j}$,

$$\begin{aligned}
\mathbb{E}\Delta_{n,4}^{j} &= \frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left[\psi'(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\right]\\
&= \frac{1}{h}\iint_{\mathbb{R}^2}\psi'(z)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_\infty(z,x)dzdx\\
&= h^j\int_{\mathbb{R}}K(u)u^j f_X(uh+x_0)du\cdot\int_{\mathbb{R}}\psi'(z)f_Z(z)dz\\
&= h^j\mathbb{E}[\psi'(Z)]\left[\mu_j f_X(x_0)+h\mu_{j+1}f'_X(x_0)+o(h)\right].
\end{aligned}$$

We can obtain similar results from $\bar{N}^j_{n,4}$. Analogous to $\bar{N}^j_{n,2}$,

$$\begin{aligned}
\bar{N}^j_{n,4} &= \iint_{\mathbb{R}^2} \psi'(z)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)\left(\sum_{i=1}^n f_\eta(z-Z_{i,i-1})-f_Z(z)\right)dzdx \\
&= h^{1+j}\int_{\mathbb{R}} K(u)u^j f_X(uh+x_0)du\cdot\int_{\mathbb{R}}\psi'(z)\left(\sum_{i=1}^n f_\eta(z-Z_{i,i-1})-f_Z(z)\right)dz \\
&= h^{1+j}\left[\mu_j f_X(x_0)+h\mu_{j+1}f'_X(x_0)+o(h)\right]\cdot\int_{\mathbb{R}}\psi'(z)\left(\sum_{i=1}^n f_\eta(z-Z_{i,i-1})-f_Z(z)\right)dz,
\end{aligned}$$

Using Lemma 2 and Assumption 5,

$$\begin{aligned}
\mathbb{E}\left|\bar{N}^j_{n,4}\right| &= nh^{1+j}\left|\mu_j f_X(x_0)\right|\cdot\mathbb{E}\left|\int_{\mathbb{R}}\psi'(z)\left(\frac{1}{n}\sum_{i=1}^n f_\eta(z-Z_{i,i-1})-f_Z(z)\right)dz\right|\cdot[1+o(1)] \\
&= \begin{cases} o(nh^{1+j}), & \text{when } j \text{ is odd;} \\ o(nh^{2+j}), & \text{when } j \text{ is even.}\end{cases}
\end{aligned}$$

so by Markov's inequality,

$$\frac{\bar{N}^j_{n,4}}{nh} = \begin{cases} o_p(h^j), & \text{when } j \text{ is odd;} \\ o_p(h^{1+j}), & \text{when } j \text{ is even.}\end{cases}$$

Hence $(nh)^{-1}\bar{N}^j_{n,4} = o_p(h^k)$ where $k = j$ or $1+j$ for odd and even j respectively. Use the method of Lemma 3 to establish convergence of the martingale part,$(nh)^{-1}M^j_{n,4}$. Let

$$\zeta_{4,i,j} = \frac{1}{\sqrt{nh^{1+2j}}}\psi'(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j \quad\text{and}\quad \xi_{4,i,j} = \zeta_{4,i,j}-\mathbb{E}(\zeta_{4,i,j}|\mathcal{F}_{i-1}).$$

$$\begin{aligned}
&\left|\mathbb{E}\Big[\psi'(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j|\mathcal{F}_{i-1}\Big]\right| \\
&= \left|\int_{\mathbb{R}}\psi'(z+Z_{i,i-1})f_\eta(z)dz\,\cdot h^{1+j}\int_{\mathbb{R}}K(u)u^j f_X(uh+x_0)\,du\right| \\
&\le \left(|\mathbb{E}\psi'(\eta)|+\int_{\mathbb{R}}\left|\psi'(z+Z_{i,i-1})-\psi'(z)\right|f_\eta(z)\,dz\right)h^{1+j}\left|\int_{\mathbb{R}}K(u)u^j f_X(uh+x_0)\,du\right| \\
&= \begin{cases} O_p(h^{1+j}), & p+j \text{ odd} \\ O_p(h^{2+j}), & p+j \text{ even}\end{cases}
\end{aligned}$$

Hence $\sum_{i=1}^n \mathbb{E}^2(\zeta_{4,i,j}|\mathcal{F}_{i-1}) \le (nh^{1+2j})^{-1}\sum_{i=1}^n O_p(h^{2+2j}) = O_p(h) = o_p(1)$. So (5.9) and (5.10) imply that it suffices to show the convergence of the conditional variance.

$$
\begin{aligned}
&\mathbb{E}\Big|\sum_{i=1}^{n}\mathbb{E}[\zeta_{4,i,j}^2|\mathcal{F}_{i-1}] - \int_{\mathbb{R}}[\psi'(z)]^2 f_Z(z)\,dz f_X(x_0)\nu_{2j}\Big| \\
\le\ &\mathbb{E}\left|\frac{1}{n}\int_{\mathbb{R}}[\psi'(z)]^2\sum_{i=1}^{n} f_\eta(z-Z_{i,i-1})\,dz\left(\int_{\mathbb{R}}K^2(u)u^{2j}f_X(x_0+hu)du+o(1)\right)\right. \\
&\qquad \left. - \int_{\mathbb{R}}[\psi'(z)]^2 f_Z(z)\,dz f_X(x_0)\nu_{2j}\right| \\
=&\mathbb{E}\left|\int_{\mathbb{R}}[\psi'(z)]^2\frac{1}{n}\sum_{i=1}^{n}\left(f_\eta(z-Z_{i,i-1})-f_Z(z)\right)\,dz f_X(x_0)\nu_{2j}(1+o(1))\right| \\
&= o(1).
\end{aligned}
$$

The Lindeberg condition is verified by a similar method as earlier. Therefore, $M_{n,4}^j = O_p(\sqrt{nh^{1+2j}}) \Rightarrow (nh^{2+j})^{-1}M_{n,4} = O_p((nh^3)^{-1/2}) = o_p(1)$ which completes the proof for the convergene of $\Delta_{n,4}^j$.

iii) By (5.1), (2.4) indicates that for X_i satisfying $|X_i - x_0| < h$,

$$
\begin{aligned}
\sup_{i\in\{X_i:|X_i-x_0|<h\}}|\delta_i| &= \sup_{X_i:|X_i-x_0|<h}\Big|R(X_i)+\sum_{j=0}^{p}(\beta_j-\beta_j^*)(X_i-x_0)^j\Big| \\
&\le \sup_{X_i:|X_i-x_0|<h}\big|R(X_i)\big| + \sup_{X_i:|X_i-x_0|<h}\sum_{j=0}^{p}\big|\beta_j-\beta_j^*\big|\cdot\big|X_i-x_0\big|^j \\
&\le O_p(h^{1+p}) + \sup_{X_i:|X_i-x_0|<h}\sum_{j=0}^{p}\big|\beta_j-\beta_j^*\big|\cdot\big|X_i-x_0\big|^j.
\end{aligned}
$$

By Theorem 1, it is known that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Denote S_ζ the sphere centred at $\boldsymbol{\beta}_0$ with radius ζ. By the fact that $\big|X_i - x_0\big| < h$ and $\boldsymbol{\beta} \in S_\zeta$, we have

$$
|\delta_i| \le O_p(h^{1+p}) + (1+p)\zeta = O_p(h^{1+p}\vee\zeta) = o_p(1) \quad \text{as } \zeta \to 0.
$$

So by conditioning and exploiting Assumption 5 we can show that $\Delta_{n,5} = o_p(h^{1+j})$.

□

Before giving proofs of the theorems, it is necessary to point out that $\Delta_{n,3}$ is negligible to $\Delta_{n,1}$. This is since in terms of Assumption 5 which implies that $\mathbb{E}|\psi'(Z_i)|$ is bounded, we have

$$\begin{aligned}\mathbb{E}|\Delta_{n,1}^{j}| &\le \sum_{i=1}^{n}\frac{1}{nh}\mathbb{E}\Big|\psi(Z_i)K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\Big|\\ &\le \frac{1}{h}\int_{\mathbb{R}}\Big|K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j\Big|f_X(x)dx\cdot\mathbb{E}|\psi'(Z_i)|\\ &= h^j\int_{\mathbb{R}}|K(u)u^j|f_X(uh+x_0)du\cdot\mathbb{E}|\psi'(Z_i)| = O(h^j)\end{aligned}$$

which leads $\Delta_{n,1}^{j} = O_p(h^j) = o_p(1)$; however, for $\Delta_{n,3}^{j}$, by Assumption 6 and (5.1), we have

$$\begin{aligned}&\mathbb{E}|\Delta_{n,3}^{j}|\\ &\le \sum_{i=1}^{n}\frac{1}{nh}\mathbb{E}\Big|\big\{\psi\big[Z_i+R(X_i)\big]-\psi(Z_i)-\psi'(Z_i)R(X_i)\big\}K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\Big|\\ &\le \frac{1}{h}\mathbb{E}\Big|\big\{\psi\big[Z_i+R(X_i)\big]-\psi(Z_i)-\psi'(Z_i)R(X_i)\big\}K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\Big|\\ &= \frac{1}{h}\mathbb{E}\Big\{\Big|\psi\big[Z_i+R(X_i)\big]-\psi(Z_i)-\psi'(Z_i)R(X_i)\Big|\cdot\Big|K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\Big|\Big\}\\ &= \frac{1}{h}\mathbb{E}\Big\{\mathbb{E}\Big[\Big|\psi\big[Z_i+R(X_i)\big]-\psi(Z_i)-\psi'(Z_i)R(X_i)\Big|\,\Big|X\Big]\cdot\Big|K\Big(\frac{X_i-x_0}{h}\Big)(X_i-x_0)^j\Big|\Big\}\\ &\le \frac{1}{h}\int_{\mathbb{R}}\mathbb{E}\Big[\sup_{x:|x-x_0|<h}\Big|\psi\big[Z_i+R(X_i)\big]-\psi(Z_i)-\psi'(Z_i)R(X_i)\Big|\,\Big|X=x\Big]\\ &\qquad\cdot\Big|K\Big(\frac{X_i-x_0}{h}\Big)(x-x_0)^j\Big|f_X(x)dx\\ &\le h^j\int_{\mathbb{R}}|K(u)u^j|f_X(uh+x_0)du\cdot o(h^p) = o(h^{p+j})\end{aligned}$$

which leads $\Delta_{n,3}^{j} = o_p(h^{p+j}) = o_p(\Delta_{n,1}^{j})$. Therefore, we have $\Delta_{n,3} = o_p(\Delta_{n,1})$ and $\Delta_{n,3} = o_p(1)$.

Proof of Theorem 1. The proof of the Theorem follows an adapted method of Jiang and Mack (2001) and Cai and Ould-Saïd (2003). The equivalent minimisation problem can be expanded,

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \frac{1}{nh}\sum_{i=1}^{n}\rho(Y_i-\sum_{i=0}^{p}\beta_j(X_i-x_0)^j)K\left(\frac{X_i-x_0}{h}\right)\\ &= \ell(\boldsymbol{\beta}_0)+\nabla\ell(\boldsymbol{\beta}_0)\cdot(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\nabla^2\ell(\boldsymbol{\beta}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\end{aligned}$$

where the partial derivatives are defined with,

$$
\begin{aligned}
\nabla \ell(\boldsymbol{\beta}_0) &= -\frac{1}{nh}\sum_{i=1}^{n}\psi\left(Y_i - \sum_{i=0}^{p}\beta_j (X_i - x_0)^j\right) K\left(\frac{X_i - x_0}{h}\right)\mathbf{X} \\
&= -\frac{1}{nh}\sum_{i=1}^{n}\left[\psi\left(Z_i + R(X_i)\right) - \psi(Z_i) - \psi'(Z_i)R(X_i)\right] K\left(\frac{X_i - x_0}{h}\right)\mathbf{X} \\
&\quad - \frac{1}{nh}\sum_{i=1}^{n}\psi\left(Z_i\right) K\left(\frac{X_i - x_0}{h}\right)\mathbf{X} - \frac{1}{nh}\sum_{i=1}^{n}\psi'\left(Z_i\right) R(X_i) K\left(\frac{X_i - x_0}{h}\right)\mathbf{X} \\
\nabla^2 \ell(\boldsymbol{\beta}^*) &= \frac{1}{nh}\sum_{i=1}^{n}\psi'\left(Y_i - \sum_{i=0}^{p}\beta_j^* (X_i - x_0)^j\right)\mathbf{X}\mathbf{X}^T \\
&= \frac{1}{nh}\sum_{i=1}^{n}\left[\psi'\left(Z_i + \delta_i\right) - \psi'(Z_i)\right] K\left(\frac{X_i - x_0}{h}\right)\mathbf{X}\mathbf{X}^T \\
&\quad + \frac{1}{nh}\sum_{i=1}^{n}\psi'(Z_i) K\left(\frac{X_i - x_0}{h}\right)\mathbf{X}\mathbf{X}^T
\end{aligned}
$$

By adapting the proof of Lemma 2 it can be shown that $\nabla \ell(\beta_0) = o_p(1)$. For the second partial derivative, assume $\boldsymbol{\beta} \in S_\delta = \left\{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|| \le \delta\right\}$, which implies that $\boldsymbol{\beta}^*$ is also in a δ-neighbourhood of $\boldsymbol{\beta}_0$. Then,

$$
\sup_{i,|X_i - x_0| \le h} |\delta_i| \le \sup_{i,|X_i - x_0| \le h} |R(X_i)| + \sup_{i,|X_i - x_0| \le h} \sum_{j=0}^{p} (X_i - x_0)^j (\beta_j - \beta_j^*) \le C(h^{p+1} + (1+p)\delta) \longrightarrow 0
$$

as $h \longrightarrow 0$ and $\delta \longrightarrow 0$. Therefore by Assumption 6,

$$
\mathbb{E}\left| \frac{1}{nh}\sum_{i=1}^{n}\left[\psi'(Z_i + \delta_i) - \psi'(Z_i)\right] K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^{k+l} \right| = o(h^{k+l}) = o(1).
$$

On the other hand, an elementary calculation gives,

$$
\frac{1}{nh}\sum_{i=1}^{n}\psi'(Z_i) K\left(\frac{X_i - x_0}{h}\right)\mathbf{X}\mathbf{X}^T = f_X(x_0)\mathbb{E}\psi'(Z)\mathbf{T_1}\mathbf{H}(1 + o_p(1))
$$

The remainder of the proof follows an analagous argument as seen in Jiang and Mack (2001) and Cai and Ould-Saïd (2003) and is left to the reader. □

Define

$$\begin{aligned}
\tilde{N}_{n,1}^{j} &= -\sum_{i=1}^{n}\iint_{\mathbb{R}^2}\psi(z)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_{\infty}^{(1,0)}(z,x)dzdx\times Z_{i,i-1}\\
&= -\iint_{\mathbb{R}^2}\psi(z)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)f_Z'(z)dzdx\times\sum_{i=1}^{n}Z_{i,i-1}\\
&= \mathbb{E}\left[\psi'(Z)\right]\int_{\mathbb{R}}K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)dx\times\sum_{i=1}^{n}Z_{i,i-1}\\
&\equiv \alpha_{n,1}^{j}\sum_{i=1}^{n}Z_{i,i-1}.
\end{aligned}$$

Note that

$$\begin{aligned}
\alpha_{n,1}^{j} &= -\iint_{\mathbb{R}^2}\psi(z)K\Big(\frac{x-x_0}{h}\Big)(x-x_0)^j f_X(x)f_Z'(z)dzdx\\
&= -h^{1+j}\iint_{\mathbb{R}^2}\psi(z)K(u)u^j f_X(uh+x_0)f_Z'(z)dzdu\\
&= h^{1+j}\mathbb{E}\left[\psi'(z)\right]\left(\mu_j f_X(x_0)+h\mu_{j+1}f_X'(x_0)+o(h)\right),
\end{aligned}$$

Therefore, it follows,

$$\tilde{N}_{n,1}^{j} = h^{1+j}\mathbb{E}\left[\psi'(z)\right]\left(\mu_j f_X(x_0)+h\mu_{j+1}f_X'(x_0)+o(h)\right)\sum_{i=1}^{n}Z_{i,i-1}. \tag{5.14}$$

Lemma 5. *If Assumptions 2 and 7 hold, then*

$$\left\|N_{n,1}^{j}-\tilde{N}_{n,1}^{j}\right\| = O(h^{j+1}\Lambda_n)$$

Proof of Lemma 5. Define the functions,

$$\begin{aligned}
S_i'(x_0) &= -\int_{\mathbb{R}}\psi(z)f_X(x)f_Z'(z)\,dz, \quad S_i(x_0)=S_i'(x_0)Z_{i,i-1}\\
K_{h,j}(x) &= K(x/h)x^j, \quad T_i(x_0)=\int_{\mathbb{R}}\psi(z)f_\eta(z-Z_{i,i-1})f_X(x)\,dz.
\end{aligned}$$

Then the terms $N_{n,1}^{j}$ and $\tilde{N}_{n,1}^{j}$ can be rewritten as,

$$N_{n,1}^{j} = \sum_{i=1}^{n}(K_{h,j}*T_i)(x_0), \qquad \tilde{N}_{n,1}^{j} = \sum_{i=1}^{n}(K_{h,j}*S_i)(x_0),$$

where $*$ is the convolution operator, that is, $f*g$ is the convolution of functions f and g. Also consider that K is assumed bounded with compact support implies that for any $(\xi_t)_{t\in\mathbb{R}}$

$$\begin{aligned}
\mathbb{E}\left[\int_{\mathbb{R}}K_{h,j}(x-x_0)\xi_x,dx\right]^2 &\le \int_{\mathbb{R}}\int_{\mathbb{R}}|K_{h,j}(x-x_0)||K_{h,j}(y-x_0)|\mathbb{E}|\xi_x||\xi_y|\,dxdy\\
&\le Ch^{2j+2}\sup_{x:|x-x_0|\le h}\|\xi_x\|^2
\end{aligned}$$

The rest of the proof will use a modified approach of Lemmas 4 and 5 in Mielniczuk and Wu (2004). In the notation of Mielniczuk and Wu (2004), for a random variable ξ define the projection operator $\mathcal{P}_k\xi = \mathbb{E}[\xi|\mathcal{F}_k] - \mathbb{E}[\xi|\mathcal{F}_{k-1}]$. Then it follows that, $-\mathcal{P}_j(K_{h,j} * S_i)(x_0) = (K_{h,j} * S_i')(x_0)\mathcal{P}_j Z_{i,i-1}$. Also note that, $-\mathcal{P}_1(K_{h,j} * S_i)(x_0) = \int_{\mathbb{R}} K_{h,j}(x - x_0)\psi(z) f_z'(z) f_X(x)\, dxdz \times \mathcal{P}_1 Z_{i,i-1}$. Further, the projection $\mathcal{P}_k Z_{i,i-1} = a_{i-k}\eta_k$ and $\mathcal{P}_1 Z_{i,i-1} = a_{i-1}\eta_1$. The norm of the projection on the discrepancy is,

$$\|\mathcal{P}_1(K_{h,j} * (T_i - S_i))(x_0)\| = \left\| \int_{\mathbb{R}} K_{h,j}(x - x_0)\psi(z) f_X(x) \left[\mathcal{P}_1 f_\eta(z - Z_{i,i-1}) + f_Z'(z) c_{i-1}\eta_1\right] \right\|$$

Then note, Assumption 7 is a special case of C_4 in Mielniczuk and Wu (2004). Consequently, by applying the same equivalent steps in Lemma 4 and 5 of Mielniczuk and Wu (2004) the result follows. □

Lemma 6. *If Assumptions 1, 4 and 5 hold, then*

i) *for the SRD case,*

$$\sqrt{nh}\mathbf{H}^{-1}\Delta_{n,1} \xrightarrow{d} N\left[0, f_X(x_0)\mathbb{E}\left[\psi^2(Z)\right]\boldsymbol{S}_p^*\right]; \tag{5.15}$$

ii) *for the LRD case, when $h^{1/2}\sigma_{n,Z} = o(n^{1/2})$, then (5.15) holds; and when $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$ and $(\eta_i)_{i\in\mathbb{Z}}$ are Gaussian random variables, then*

$$\frac{n\widetilde{\mathbf{H}}^{-1}}{\sigma_{n,Z}}\Delta_{n,1} \xrightarrow{d} N\left[0, \boldsymbol{c}_p^*(\boldsymbol{c}_p^*)^T C_{LRD}^2\right] \tag{5.16}$$

where $C_{LRD} = \int_{\mathbb{R}} w\phi(w) \int_{\mathbb{R}} \psi(z) f_\eta(z - s_z w)\, dz\, dw$ and $\boldsymbol{c}_p^* = \boldsymbol{c}_p f_X(x_0) + \widetilde{\mathbf{c}}_p f_X'(x_0)$

Proof of Lemma 6. i) By the Crámer-Wold theorem, to prove (5.15) is equivalent to proving that for any real numbers $b_0, b_1, \cdots, b_p$ such that $\mathbf{b} = (b_0, b_1, \cdots, b_p)^T \neq \mathbf{0}$, $\Delta_{n,1}^j$ satisfies $\sum_{j=0}^p (nh^{1-2j})^{1/2} b_j \Delta_{n,1}^j \to^d N\left[0, f_X(x_0)\mathbb{E}\left[\psi^2(Z)\right]\mathbf{b}^T \boldsymbol{S}_p^* \mathbf{b}\right]$. Here, we do the same decomposition as (5.2) to $\sum_{j=0}^p (nh^{1-2j})^{1/2} b_j \Delta_{n,1}^j$,

$$\begin{aligned}
\sum_{j=0}^p b_j \sqrt{nh^{1-2j}}\Delta_{n,1}^j &= \frac{1}{nh}\sum_{i=1}^n \sum_{j=0}^p \psi(Z_i) K\Big(\frac{X_i - x_0}{h}\Big) b_j \sqrt{nh^{1-2j}} (X_i - x_0)^j \\
&= \frac{1}{nh}\Bigg\{\sum_{i=1}^n \sum_{j=0}^p \Bigg[\psi(Z_i) K\Big(\frac{X_i - x_0}{h}\Big) b_j \sqrt{nh^{1-2j}} (X_i - x_0)^j \\
&\quad - \mathbb{E}\Big(\psi(Z_i) K\Big(\frac{X_i - x_0}{h}\Big) b_j \sqrt{nh^{1-2j}} (X_i - x_0)^j \Big| \mathcal{F}_{i-1}\Big)\Bigg] \\
&\quad + \sum_{i=1}^n \sum_{j=0}^p \mathbb{E}\Big[\psi(Z_i) K\Big(\frac{X_i - x_0}{h}\Big) b_j \sqrt{nh^{1-2j}} (X_i - x_0)^j \Big| \mathcal{F}_{i-1}\Big]\Bigg\} \\
&\equiv \frac{1}{nh}(M_{n,1}^* + N_{n,1}^*).
\end{aligned}$$

First, we investigate the bahaviour of $(nh)^{-1}M^*_{n,1}$. We can show the asymptotic Gaussian limit in $(nh)^{-1}M^*_{n,1}$ via the method of Lemma 3. Let

$$\zeta^*_{1,i} = \frac{1}{nh}\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)\sum_{j=0}^{p} b_j\sqrt{nh^{1-2j}}(X_i - x_0)^j \quad \text{and} \quad \xi^*_{1,i} = \zeta^*_{1,i} - \mathbb{E}(\zeta^*_{1,i}|\mathcal{F}_{i-1}).$$

Observe that by (5.8),

$$\begin{aligned}
&\left|\mathbb{E}\Big[\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)\sum_{j=0}^{p} b_j\sqrt{nh^{1-2j}}(X_i - x_0)^j\Big|\mathcal{F}_{i-1}\Big]\right| \\
&= \left|\sum_{j=0}^{p} b_j\sqrt{nh^{1-2j}}\,\mathbb{E}\Big[\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\Big|\mathcal{F}_{i-1}\Big]\right| \\
&\le O_p(\sqrt{nh^3}),
\end{aligned}$$

and this indicates $\sum_{i=1}^{n}\mathbb{E}^2(\zeta^*_{1,i}|\mathcal{F}_{i-1}) \le (nh)^{-2}\sum_{i=1}^{n} O_p(nh^3) = O_p(h) = o_p(1)$. Furthermore, it is straightforward to check the Lindeberg condition by the same argument used previously in Lemma 3. Thus (5.9) implies that it suffices to show $\sum_{i=1}^{n}\mathbb{E}\big[(\zeta^*_{1,i})^2|\mathcal{F}_{i-1}\big] \to^p f_X(x_0)\mathbb{E}\big[\psi^2(Z)\big]\mathbf{b}^T\boldsymbol{S}^*_p\mathbf{b}$. Note that

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\big[(\zeta^*_{1,i})^2|\mathcal{F}_{i-1}\big] \\
&= \sum_{i=1}^{n}\sum_{j=0}^{p}\mathbb{E}\Big[\frac{b_j^2}{nh^{1+2j}}\psi^2(Z_i)K^2\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\Big|\mathcal{F}_{i-1}\Big] \\
&\qquad + \sum_{i=1}^{n}\sum_{\substack{k,t=0\\k\ne t}}^{p}\mathbb{E}\Big[\frac{b_k b_t}{nh^{1+k+t}}\psi^2(Z_i)K^2\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^{k+t}\Big|\mathcal{F}_{i-1}\Big] \\
&= \sum_{i=1}^{n}\sum_{j=0}^{p} b_j^2\mathbb{E}(\zeta^2_{1,i,j}|\mathcal{F}_{i-1}) + \sum_{i=1}^{n}\sum_{\substack{k,t=0\\k\ne t}}^{p} b_k b_t\mathbb{E}\Big(\zeta^2_{1,i,\frac{k+t}{2}}|\mathcal{F}_{i-1}\Big),
\end{aligned}$$

where $\zeta^2_{1,i,j}$ ($j = 0, 1, \cdots, p$ and $(k+t)/2$, and $k, t = 0, 1, \cdots, p$ and $k \ne t$) are defined in (5.5). Therefore, by Lemma 3, we have

$$\begin{aligned}
\sum_{i=1}^{n}\mathbb{E}\big[(\zeta^*_{1,i})^2|\mathcal{F}_{i-1}\big] &\xrightarrow{P} \sum_{j=0}^{p} b_j^2\sigma_{1,j}(x_0)^2 + \sum_{\substack{k,t=0\\k\ne t}}^{p} b_k b_t\sigma^2_{1,\frac{k+t}{2}}(x_0) \\
&= f_X(x_0)\mathbb{E}\big[\psi^2(Z)\big]\mathbf{b}^T\boldsymbol{S}^*_p\mathbf{b},
\end{aligned}$$

and further, $(nh)^{-1}M^*_{n,1} \to^d N\big[0, f_X(x_0)\mathbb{E}\big[\psi^2(Z)\big]\mathbf{b}^T\boldsymbol{S}^*_p\mathbf{b}\big]$. Now we study the behaviour of $(nh)^{-1}N^*_{n,1}$. Note that $(nh)^{-1}N^*_{n,1} = \sum_{j=0}^{p}(nh^{1+2j})^{-1/2}b_j N^j_{n,1} = \sum_{j=0}^{p}(nh^{1+2j})^{-1/2}b_j(\tilde{N}^j_{n,1} +$

$N^j_{n,1} - \tilde{N}^j_{n,1}$), so it is supposed to derive the behaviours of $\tilde{N}^j_{n,1}$ and the remainder, $N^j_{n,1} - \tilde{N}^j_{n,1}$. In (5.14), it is known that $\alpha^j_{n,1} = O(h^{1+j})$ when j is even and $\alpha^j_{n,1} = O(h^{2+j})$ when j is odd. So it follows

$$\begin{aligned}\mathbb{E}\left|\tilde{N}^j_{n,1}\right| = \mathbb{E}\left|\alpha^j_{n,1}\sum_{i=1}^n Z_{i,i-1}\right| &= \left|\alpha^j_{n,1}\right| \cdot \mathbb{E}\left|\sum_{i=1}^n Z_{i,i-1}\right| \\ &\le \left|\alpha^j_{n,1}\right| \cdot \left\|\sum_{i=1}^n Z_{i,i-1}\right\| \\ &= \begin{cases} O(n^{\frac{1}{2}}h^{1+j}), & j \text{ is even} \\ O(n^{\frac{1}{2}}h^{2+j}), & j \text{ is odd.}\end{cases}\end{aligned}$$

Thus, for both even and odd j, $\mathbb{E}|(nh^{1+2j})^{-1/2}\tilde{N}^j_{n,1}| = o(1)$, and then by Markov's inequality, it follows that $(nh^{1+2j})^{-1/2}\tilde{N}^j_{n,1} = o_p(1) \Rightarrow \sum_{j=0}^p (nh^{1+2j})^{-1/2} b_j \tilde{N}^j_{n,1} = o_p(1)$. Similarly, for the remainder term, by Lemma 5,

$$\sum_{j=0}^p (nh^{1+2j})^{-1/2} b_j \left(N^j_{n,1} - \tilde{N}^j_{n,1}\right) = O_p(h^{1/2}) = o_p(1)$$

Therefore, $\sum_{j=0}^p (nh^{1-2j})^{-1/2} b_j \Delta^j_{n,1} \to^d N\left[0, f_X(x_0)\mathbb{E}\left[\psi^2(Z)\right]\mathbf{b}^T \boldsymbol{S}^*_p \mathbf{b}\right]$, and then (5.15) is proven.

ii) First, we consider the case that $h^{1/2}\sigma_{n,Z} = o(n^{1/2})$. In terms of part i), it follows that $\sum_{i=1}^n (nh^{1-2j})^{1/2} b_j \Delta^j_{n,1} = (nh)^{-1}(M^*_{n,1} + N^*_{n,1})$. The behaviour of $(nh)^{-1}M^*_{n,1}$ is the same as the SRD case. For $N^*_{n,1}$, we still investigate $\tilde{N}^j_{n,1}$ and the remainders respectively. Note that (5.12) implies $\sum_{i=1}^n Z_{i,i-1} = O_p(\sigma_{n,Z})$ under the LRD condition, so by (5.14)

$$\begin{aligned}\frac{\tilde{N}^j_{n,1}}{\sqrt{nh^{1+2j}}} &= \frac{\alpha^j_{n,1}\sum_{i=1}^n Z_{i,i-1}}{\sqrt{nh^{1+2j}}} \\ &= \frac{\alpha^j_{n,1}}{h^{1+j}} \cdot \frac{\sum_{i=1}^n Z_{i,i-1}}{\sigma_{n,Z}} \cdot \frac{h^{1+j}\sigma_{n,Z}}{\sqrt{nh^{1+2j}}} \\ &= \frac{\alpha^j_{n,1}}{h^{1+j}} \cdot \frac{\sum_{i=1}^n Z_{i,i-1}}{\sigma_{n,Z}} \cdot \frac{h^{\frac{1}{2}}\sigma_{n,Z}}{n^{\frac{1}{2}}} \\ &= \begin{cases} o_p(1), & j \text{ is even} \\ o_p(h), & j \text{ is odd.}\end{cases}\end{aligned}$$

This indicates that for both even and odd j, $(nh^{1+2j})^{-1/2}\tilde{N}^j_{n,1} = o_p(1)$ holds. Hence, $\sum_{j=0}^p (nh^{1+2j})^{-1/2} b_j \tilde{N}^j_{n,1} = o_p(1)$. By Lemma 5 and (5.12)

$$\sum_{j=0}^p (nh^{1+2j})^{-1/2} b_j \left(N^j_{n,1} - \tilde{N}^j_{n,1}\right) = O_p(h^{1/2}\Lambda_n n^{-1/2}) = o_p(1).$$

Indeed, for $\alpha_Z > 3/4$, $\Lambda_n = O(n^{1/2})$ and the result is obvious. On the other hand, for $\alpha_Z \in (1/2, 3/4)$, note that $\Lambda_n = O(\sigma^2_{n,Z}/n)$ and,

$$h^{1/2}\Lambda_n n^{-1/2} = \frac{h^{1/2}\sigma_{n,Z}}{n^{1/2}} \cdot \frac{\sigma_{n,Z}}{n} = o(1),$$

when $\alpha_Z \in (1/2, 3/4)$. Therefore, we have $(nh)^{-1}N^*_{n,1} = o_p(1)$. Thus, (5.15) follows $\sum_{j=0}^{p}(nh^{1-2j})^{-1/2} b_j \Delta^j_{n,1} \to^d N\left[0, f_X(x_0)\mathbb{E}\left[\psi^2(Z)\right]\mathbf{b}^T \boldsymbol{S}^*_p \mathbf{b}\right]$.

Now, consider the case of $n^{1/2} = o(h^{3/2}\sigma_{n,Z})$. Recall, $\Delta^j_{n,1} = (nh)^{-1}(M^j_{n,1} + N^j_{n,1})$. From earlier, it is known that $M^j_{n,1} = O_p(\sqrt{nh^{1+2j}})$ which implies that $\sigma^{-1}_{n,Z} h^{-1-j-k(j)} M^j_{n,1} = O_p(n^{1/2}/(h^{1/2+k(j)}\sigma_{n,Z})) = o_p(1)$ for both even and odd j. Consider the dependent part $N^j_{n,1}$ when $Z_{i,i-1}$ is a Gaussian LRD variable. Evaluating the term directly,

$$\begin{aligned}
N^j_{n,1} &= \sum_{i=1}^{n} \mathbb{E}\left[\psi(Z_i) K\left(\frac{X_i - x_0}{h}\right)(X_i - x_0)^j \middle| \mathcal{F}_{i-1}\right] \\
&= h^{1+j} \sum_{i=1}^{n} \int_{\mathbb{R}} \psi(z) f_\eta(z - Z_{i,i-1})\, dz \int_{\mathbb{R}} K(u) u^j f_X(x_0 + hu)\, dx \\
&= h^{1+j} \sum_{i=1}^{n} \left(\mu_j f_X(x_0) + h\mu_{j+1} f'_X(x_0) + o(h)\right) \int_{\mathbb{R}} \psi(z) f_\eta(z - Z_{i,i-1})\, dz \\
&= h^{1+j} \sum_{i=1}^{n} \left(\mu_j f_X(x_0) + h\mu_{j+1} f'_X(x_0)\right) \int_{\mathbb{R}} \psi(z) f_\eta(z - Z_{i,i-1})\, dz\, (1 + o(1)) \\
&= h^{1+j+k(j)} \sum_{i=1}^{n} C_j(x_0) \int_{\mathbb{R}} \psi(z) f_\eta(z - Z_{i,i-1})\, dz\, (1 + o(1)) \\
&\equiv h^{1+j+k(j)} \sum_{i=1}^{n} G_j(W_i, x_0)(1 + o(1))
\end{aligned}$$

where the functions

$$k(j) = \begin{cases} 0, & \text{if } j \text{ even;} \\ 1, & \text{if } j \text{ odd.} \end{cases} \quad \text{and} \quad c_j(x_0) = \begin{cases} \mu_j f_X(x_0), & \text{if } j \text{ even;} \\ \mu_{j+1} f'_X(x_0), & \text{if } j \text{ odd.} \end{cases}$$

Also, by Jensen's inequality, $\mathbb{E}G_j(W_i, x_0)^2 < \infty$. So by Corollary 5.1 in Taqqu (1975), it is sufficient to check the Hermite rank of the summand in $N^j_{n,1}$ to establish convergence of a non-central limit theorem. $G_j(W_i, x_o) = \int_{\mathbb{R}} \psi(z) f_\eta(z - s_Z W_i)\, dz \left(\mu_j f_X(x_0) + h\mu_{j+1} f'_X(x_0)\right)$ and $W_i \sim \mathcal{N}(0, 1)$ and $s_z = \sqrt{\mathbb{E}Z^2_{i,i-1}}$. To evaluate the Hermite rank, re-express the integrand in terms of a convolution. By assumption, η is a Gaussian variable, therefore f_η is a Gaussian density and thus an even function. Therefore, we can write,

$$\begin{aligned}\int_{\mathbb{R}} \psi(z) f_\eta(z - s_z w)\, dz &= \int_{\mathbb{R}} \psi(z) f_\eta(s_z(w - z/s_z))\, dz \\ &= s_z \int_{\mathbb{R}} \psi(s_z u) f_\eta(s_z(w-u))\, dz \\ &= s_z(\psi_{s_z} * f_{\eta,s_z})(w)\end{aligned}$$

where $\psi_{s_z}(u) = \psi(s_Z u)$ and $f_{\eta,s_z}(u) = f_\eta(s_z u)$. Also, by assumption ψ is an odd function which implie ψ_{s_z} is an odd function. Since f_η is even is also implies f_{η,s_z} is even. Therefore, the convolution $\psi_{s_z} * f_{\eta,s_z}$ is an odd function. The proof is omitted but can be shown directly or via the use of Fourier transforms. Furthermore, $s_z(\psi_{s_z} * f_{\eta,s_z})(0) = \mathbb{E}\psi(\eta) = 0$. Define $C_{1,j}(x_0) = \mathbb{E}\left[W_i G_j(W_i, x_0)\right]$. Then checking the Hermite rank, consider the following,

$$\frac{C_{1,j}(x_0)}{C_j(x_0)} = \frac{1}{C_j(x_0)}\mathbb{E}\left[W_i G_j(W_i, x_0)\right] = s_z \int_{\mathbb{R}} w\phi(w)(\psi_{s_z} * f_{\eta,s_z})(w)\, dw = \int_{\mathbb{R}} q(w)\, dw > 0$$

The last inequality holds since the integrand $q(w) = s_z w\phi(w)(\psi_{s_z} * f_{\eta,s_z})(w)$ is even function with $q(w) = 0$. The function $q(w)$ is even since it is a product of two odd functions, w and $(\psi_{s_z} * f_{\eta,s_z})(w)$; and one even function $\phi(w)$.

Thus, by Corollary 5.1 of Taqqu (1975), we have,

$$\frac{1}{h^{1+j+k(j)}\sigma_{n,Z}} N^j_{n,1} \xrightarrow{d} N(0, C^2_{1,j}(x_0))$$

To complete the result, consider now the multivariate case with, $N_{n,1}$. Define $\widetilde{\mathbf{H}} = \left(h^{j+k(j)}\right)_{0\le j\le p}$ and $\mathbf{H}_* = \left(h^{k(j)}\right)_{0\le j\le p}$ yielding $\widetilde{\mathbf{H}} = \mathbf{H}\cdot\mathbf{H}_*$. Then by the Crámer-Wold device, for any $\boldsymbol{b} \neq \mathbf{0}$,

$$\begin{aligned}\frac{\boldsymbol{b}^T\widetilde{\mathbf{H}}^{-1}N_{n,1}}{h\sigma_{n,Z}} &= \frac{1}{\sigma_{n,Z}}\sum_{i=1}^{n}(\boldsymbol{b}^T\boldsymbol{c}_p f_X(x_0) + \boldsymbol{b}^T\widetilde{\mathbf{c}}_p f'_X(x_0)) \int_{\mathbb{R}} \psi(z) f_\eta(z - Z_{i,i-1})\, dz(1 + o(1)) \\ &\equiv \frac{1}{\sigma_{n,Z}}\sum_{i=1}^{n} G(W_i, x_0)(1 + o(1)).\end{aligned}$$

For convenience define, $\boldsymbol{c}^*_p = (\boldsymbol{b}^T\boldsymbol{c}_p f_X(x_0) + \boldsymbol{b}^T\widetilde{\mathbf{c}}_p f'_X(x_0))$. Then with an equivalent argument to the $N^j_{n,1}$ case it follows, that $\mathbb{E}W_i G(W_i, x_0)/(\boldsymbol{b}^T\boldsymbol{c}^*_p) = \int_{\mathbb{R}} q(w)\, dw > 0$ and consequently,

$$\frac{\boldsymbol{b}^T\widetilde{\mathbf{H}}^{-1}N_{n,1}}{h\sigma_{n,Z}} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{b}^T\boldsymbol{c}^*_p(\boldsymbol{c}^*_p)^T\boldsymbol{b})$$

and the result is proven. □

Before giving proofs of the theorems, it is necessary to point out that $\Delta_{n,3}$ is negligible to $\Delta_{n,1}$. This is since in terms of Assumption 5 which implies that $\mathbb{E}|\psi'(Z_i)|$ is bounded, we have

$$\begin{aligned}\mathbb{E}|\Delta_{n,1}^j| &\leq \sum_{i=1}^n \frac{1}{nh}\mathbb{E}\left|\psi(Z_i)K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\right| \\ &\leq \frac{1}{h}\int_{\mathbb{R}}\left|K\Big(\frac{x - x_0}{h}\Big)(x - x_0)^j\right| f_X(x)dx \cdot \mathbb{E}|\psi'(Z_i)| \\ &= h^j \int_{\mathbb{R}} |K(u)u^j| f_X(uh + x_0)du \cdot \mathbb{E}|\psi'(Z_i)| = O(h^j)\end{aligned}$$

which leads $\Delta_{n,1}^j = O_p(h^j)$; however, for $\Delta_{n,3}^j$, by Assumption 6 and (5.1), we have

$$\begin{aligned}&\mathbb{E}|\Delta_{n,3}^j| \\ &\leq \sum_{i=1}^n \frac{1}{nh}\mathbb{E}\left|\{\psi[Z_i + R(X_i)] - \psi(Z_i) - \psi'(Z_i)R(X_i)\}K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\right| \\ &\leq \frac{1}{h}\mathbb{E}\left|\{\psi[Z_i + R(X_i)] - \psi(Z_i) - \psi'(Z_i)R(X_i)\}K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\right| \\ &= \frac{1}{h}\mathbb{E}\left\{\left|\psi[Z_i + R(X_i)] - \psi(Z_i) - \psi'(Z_i)R(X_i)\right| \cdot \left|K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\right|\right\} \\ &= \frac{1}{h}\mathbb{E}\left\{\mathbb{E}\left[\left|\psi[Z_i + R(X_i)] - \psi(Z_i) - \psi'(Z_i)R(X_i)\right| \Big| X\right] \cdot \left|K\Big(\frac{X_i - x_0}{h}\Big)(X_i - x_0)^j\right|\right\} \\ &\leq \frac{1}{h}\int_{\mathbb{R}} \mathbb{E}\left[\sup_{x:|x-x_0|<h}\left|\psi[Z_i + R(X_i)] - \psi(Z_i) - \psi'(Z_i)R(X_i)\right| \Big| X = x\right] \\ &\qquad \cdot \left|K\Big(\frac{X_i - x_0}{h}\Big)(x - x_0)^j\right| f_X(x)dx \\ &\leq h^j \int_{\mathbb{R}} |K(u)u^j| f_X(uh + x_0)du \cdot o(h^p) = o(h^{p+j})\end{aligned}$$

which leads $\Delta_{n,3}^j = o_p(h^{p+j}) = o_p(\Delta_{n,1}^j)$. Therefore, we have $\Delta_{n,3} = o_p(\Delta_{n,1})$.

Proof of Theorem 2. By Lemma 4, it follows $\Delta_{n,4} = \big(h^{j+k}(\mu_{j+k}f_X(x_0) + h\mu_{j+k+1}f'_X(x_0) + o_p(h))\big)_{1\leq j,k\leq p}$ and $\Delta_{n,5} = \big(o_p(h^{j+k})\big)$. Further, define $\boldsymbol{S}_p = (\mu_{j+k})_{0\leq j,k\leq p}$ and $\widetilde{\mathbf{S}}_p = (\mu_{j+k+1})_{0\leq j,k\leq p}$. Therefore, it follows that there exists a matrix $A_n = \big(a_{j,k}(n)\big)_{0\leq j,k\leq p}$ such that $a_{j,k}(n) \xrightarrow{p} 0$ and,

$$\Delta_{n,4} + \Delta_{n,5} = \mathbb{E}\left[\psi'(Z)\right]\mathbf{H}\left(f_X(x_0)\boldsymbol{S}_p + hf'_X(x_0)\widetilde{\mathbf{S}}_p + A_n\right)\mathbf{H} \tag{5.17}$$

For convenience, define $B_n = f_X(x_0)\boldsymbol{S}_p + hf'_X(x_0)\widetilde{\mathbf{S}}_p$. This has inverse,

$$\begin{aligned}B_n^{-1} &= \left(f_X(x_0)\boldsymbol{S}_p + hf'_X(x_0)\widetilde{\mathbf{S}}_p\right)^{-1} \\ &= \frac{\boldsymbol{S}_p^{-1}}{f_X(x_0)} - h\frac{f'_X(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X^2(x_0)} + O_p(h^2) \\ &= \frac{1}{f_X(x_0)}\left(\boldsymbol{S}_p^{-1} - h\frac{f'_X(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right) + O_p(h^2)\end{aligned}$$

and this leads to the inverse,

$$\begin{aligned}
\left(\Delta_{n,4}+\Delta_{n,5}\right)^{-1} &= \frac{1}{\mathbb{E}\left[\psi'(Z)\right]}\left[\mathbf{H}\left(B_n+A_n\right)\mathbf{H}\right]^{-1} \\
&= \frac{1}{\mathbb{E}\left[\psi'(Z)\right]}\mathbf{H}^{-1}\left(I+B_n^{-1}A_n\right)^{-1}B_n^{-1}\mathbf{H}^{-1} \\
&= \frac{1}{\mathbb{E}\left[\psi'(Z)\right]}\mathbf{H}^{-1}\left(B_n^{-1}-B_n^{-1}A_nB_n^{-1}\right)\mathbf{H}^{-1} \\
&= \frac{1}{\mathbb{E}\left[\psi'(Z)\right]}\mathbf{H}^{-1}\left(B_n^{-1}-B_n^{-1}A_nB_n^{-1}\right)\mathbf{H}^{-1} \\
&= \frac{1}{\mathbb{E}\left[\psi'(Z)\right]f_X(x_0)}\mathbf{H}^{-1}\left(\boldsymbol{S}_p^{-1}-h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\mathbf{H}^{-1}(1+o_p(1)) \qquad (5.18)
\end{aligned}$$

Define $c_p=\left(\mu_j\right)_{0\le j\le p}$ and $\widetilde{c}_p=\left(\mu_{1+j}\right)_{0\le j\le p}$ then Lemma 4 also yields

$$\Delta_{n,2}=\begin{cases}
\mathbb{E}\left[\psi'(Z)\right]\begin{pmatrix}
h^{1+p}\mu_{1+p}f_X(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!} \\
h^{3+p}\mu_{3+p}\left(f_X'(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!}+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right) \\
h^{3+p}\mu_{3+p}f_X(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!} \\
\vdots \\
h^{2+2p}\mu_{2+2p}\left(f_X'(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!}+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right)
\end{pmatrix}[1+o_p(1)], & p \text{ is odd} \\
\mathbb{E}\left[\psi'(Z)\right]\begin{pmatrix}
h^{2+p}\mu_{2+p}\left(f_X'(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!}+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right) \\
h^{2+p}\mu_{2+p}f_X(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!} \\
h^{4+p}\mu_{4+p}\left(f_X'(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!}+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right) \\
\vdots \\
h^{2+2p}\mu_{2+2p}\left(f_X'(x_0)\frac{m^{(p+1)}(x_0)}{(p+1)!}+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right)
\end{pmatrix}[1+o_p(1)], & p \text{ is even}
\end{cases}$$

$$\begin{aligned}
&= \frac{h^{1+p}\mathbb{E}[\psi'(Z)]\mathbf{H}}{(p+1)!}\left[f_X(x_0)m^{(p+1)}(x_0)\boldsymbol{c}_p+h\left(f_X'(x_0)m^{(p+1)}(x_0)+f_X(x_0)\frac{m^{(p+2)}(x_0)}{(p+2)}\right)\widetilde{\mathbf{c}}_p\right] \\
&\qquad \times[1+o_p(1)] \qquad (5.19)
\end{aligned}$$

Denote, $\theta_{p+1}=m^{(p+1)}(x_0)/(p+1)!$. Thus, by (5.17) and (5.24),

$$\begin{aligned}
&\left(\Delta_{n,4}+\Delta_{n,5}\right)^{-1}\cdot\Delta_{n,2} \\
&= \frac{h^{1+p}}{f_X(x_0)}\mathbf{H}^{-1}\left(\boldsymbol{S}_p^{-1}-h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p+h\left(f_X'(x_0)\theta_{p+1}+f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right] \\
&\qquad (5.20)
\end{aligned}$$

By (2.5), we have

$$\sqrt{nh}\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}+\left(\Delta_{n,4}+\Delta_{n,5}\right)^{-1}\cdot\Delta_{n,2}\right]=-\sqrt{nh}\left(\Delta_{n,4}+\Delta_{n,5}\right)^{-1}\left(\Delta_{n,1}+\Delta_{n,3}\right). \qquad (5.21)$$

Since $\Delta_{n,3}$ is negligible to $\Delta_{n,1}$, then we can find

$$\sqrt{nh}(\Delta_{n,1}+\Delta_{n,3}) = \sqrt{nh}\Delta_{n,1}[1+o_p(1)]$$

By (5.15) in Lemma 6, we have

$$\sqrt{nh}\mathbf{H}^{-1}\Delta_{n,1} \xrightarrow{d} N\left[0, f_X(x_0)\mathbb{E}\left[\psi^2(Z)\right]\boldsymbol{S}_p^*\right]. \tag{5.22}$$

Therefore, by the above and (5.18),

$$\begin{aligned}&\sqrt{nh}(\Delta_{n,4}+\Delta_{n,5})^{-1}(\Delta_{n,1}+\Delta_{n,3})\\&\quad=\frac{1}{\mathbb{E}\left[\psi'(Z)\right]f_X(x_0)}\mathbf{H}^{-1}\left(\boldsymbol{S}_p^{-1}-h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\sqrt{nh}\mathbf{H}^{-1}\Delta_{n,1}(1+o_p(1)),\end{aligned} \tag{5.23}$$

and then (5.20)-(5.23) leads to

$$\begin{aligned}&\sqrt{nh}\left[\mathbf{H}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\frac{h^{1+p}}{f_X(x_0)}\left(\boldsymbol{S}_p^{-1}-h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p+h\left(f_X'(x_0)\theta_{p+1}+f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right]\right]\\&\qquad\xrightarrow{d}\mathcal{N}\left(0,\frac{\mathbb{E}\left[\psi^2(Z)\right]}{\mathbb{E}\left[\psi'(Z)\right]^2f_X(x_0)}\boldsymbol{S}_p^{-1}\boldsymbol{S}_p^*\boldsymbol{S}_p^{-1}\right)\end{aligned}$$

□

Proof of Theorem 3. i) Same as the proof of Theorem 2, omitted here.

ii) Again, decompose the estimator in a similar to way as used in the proof of Theorem 2. Note that the matrices $\boldsymbol{S}_p$ are symmetric and therefore diagonalisable implying that $\boldsymbol{S}_p^{-1}$ is also symmetric and diagonalisable. From (5.20),

$$\begin{aligned}&\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\mathbf{H}(\Delta_{n,4}+\Delta_{n,5})^{-1}\Delta_{n,2}\\&\quad=\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\frac{h^{1+p}}{f_X(x_0)}\left(\boldsymbol{S}_p^{-1}-h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p+h\left(f_X'(x_0)\theta_{p+1}+f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right]\\&\quad=\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\frac{h^{1+p}}{f_X(x_0)}\left[f_X(x_0)\theta_{p+1}\boldsymbol{c}_p+h\left(f_X'(x_0)\theta_{p+1}+f_X(x_0)\theta_{p+2}\right)\widetilde{\mathbf{c}}_p\right](1+o(1))\\&\quad=Bias_{LRD}(\mathbf{H}_*^{-1}\mathbf{H}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}))(1+o(1))\end{aligned} \tag{5.24}$$

Note that two diagonalisable matrices are commutative. Therefore, from (5.18)

$$
\begin{aligned}
&\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\mathbf{H}(\Delta_{n,4}+\Delta_{n,5})^{-1} \\
&\quad= \frac{1}{\mathbb{E}\left[\psi'(Z)\right] f_X(x_0)}\mathbf{H}_*^{-1}\mathbf{H}\mathbf{H}^{-1}\left(\boldsymbol{S}_p^{-1} - h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\mathbf{H}^{-1}(1+o_p(1)) \\
&\quad= \frac{1}{\mathbb{E}\left[\psi'(Z)\right] f_X(x_0)}\left(\boldsymbol{S}_p^{-1} - h\frac{f_X'(x_0)\boldsymbol{S}_p^{-1}\widetilde{\mathbf{S}}_p\boldsymbol{S}_p^{-1}}{f_X(x_0)}\right)\mathbf{H}_*^{-1}\mathbf{H}^{-1}(1+o_p(1)) \\
&\quad= \frac{1}{\mathbb{E}\left[\psi'(Z)\right] f_X(x_0)}\boldsymbol{S}_p^{-1}\widetilde{\mathbf{H}}^{-1}(1+o_p(1))
\end{aligned} \tag{5.25}
$$

The last two lines follow since both $\mathbf{H}_*^{-1}$ and $\boldsymbol{S}_p^{-1}$ are symmetric and diagonalisable and therefore commute. Recall $\Delta_{n,3} = o_p(\Delta_{n,1})$ then from (2.5), the estimator has decomposition,

$$
\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\mathbf{H}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} + (\Delta_{n,4}+\Delta_{n,5})^{-1}\Delta_{n,2}\right) = -\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\mathbf{H}(\Delta_{n,4}+\Delta_{n,5})^{-1}\Delta_{n,1}(1+o_p(1))
$$

Then apply (5.24) and (5.25) to yield,

$$
\frac{n}{\sigma_{n,Z}}\mathbf{H}_*^{-1}\mathbf{H}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - Bias_{LRD}(\mathbf{H}_*^{-1}\mathbf{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = -\frac{n}{\sigma_{n,Z}\mathbb{E}\left[\psi'(Z)\right] f_X(x_0)}\boldsymbol{S}_p^{-1}\widetilde{\mathbf{H}}^{-1}\Delta_{n,1}(1+o_p(1))
$$

Then by (5.16) in Lemma 6 we have,

$$
\frac{n\boldsymbol{S}_p^{-1}\widetilde{\mathbf{H}}^{-1}}{\sigma_{n,Z}\mathbb{E}\left[\psi'(Z)\right] f_X(x_0)}\Delta_{n,1} \xrightarrow{d} N\left(0, \frac{\boldsymbol{S}_p^{-1}\boldsymbol{c}_p^*(\boldsymbol{c}_p^*)^T\boldsymbol{S}_p^{-1}C_{LRD}^2}{\mathbb{E}\left[\psi'(Z)\right]^2 f_X^2(x_0)}\right)
$$

□

References

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16:147–192.

Beran, J., Feng, Y., Chosh, S., and Sibbertsen, P. (2002). On robust local polynomial estimation with long-memory errors. *Int. J. Forecasting*, 18:227–241.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Springer.

Bojanic, R. and Seneta, E. (1973). A unified theory of regularly varying sequences. *Math. Z*, 134:91–106.

Brown, B. (1971). Martingale central limit theorems. *Ann. Math. Statist.*, 42:59–66.

Cai, Z. and Ould-Saïd, E. (2003). Local m-estimator for nonparametric time series. *Statist. Probab. Lett.*, 65:433–449.

Chow, Y. and Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York, 3rd edition.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 74:829–836.

Cox, D. D. (1983). Asymptotics for m-type smoothing spline. *Ann. Statist.*, 11:530–551.

Csörgo, S. and Mielniczuk, J. (1995a). Close short-range dependent sums and regression estimation. *Acta Sci. Math. (Szeged)*, 60:177–196.

Csörgo, S. and Mielniczuk, J. (1995b). Density estimation under long-range dependence. *Ann. Statist.*, 23:990–999.

Csörgo, S. and Mielniczuk, J. (1999). Random-design regression under long-range dependent errors. *Bernoulli*, 5:209–224.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability. Chapman & Hall, New York.

Fan, J. and Jiang, J. (1999). Variable bandwidth and one-step local m-estimator. *Sci. China Ser. A*, 29:1–15.

Gasser, T. and Müller, H.-G. (1979). Kernal estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, volume 757 of *Lecture Notes in Mathematics*, pages 23–68, New York. Springer-Verlag.

Hall, P. and Hart, J. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, 36:339–351.

Härdle, W. (1990). *Smoothing techniques with implementation in S*. Springer Series in Statistics. Springer-Verlag, New York.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, 35:73–101.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *Ann. Statist.*, 1:733–821.

Huber, P. J. (1996). *Robust statistical procedures*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 2nd edition.

Jiang, J. and Mack, Y. P. (2001). Robust local polynomial regression for dependent data. *Statistica Sinica*, 11(3):705–722.

Mielniczuk, J. and Wu, W. B. (2004). On random-design model with dependent errors. *Statist. Sinica*, 14:1105–1126.

Naradaya, E. A. (1964). On estimating regression. *Theory Prob. Appl.*, 9:141–142.

Seneta, E. (1976). *Regularly Varying Functions*. Springer, Berlin.

Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press, London.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer Series in Statistics. Springer New York, New York, NY.

Taqqu, M. S. (1975). Weak convergence to fractional brownian motion and to the rosenblatt process. *Advances in Applied Probability*, 7(2):249.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Texts in Statistics. Springer, New York.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372.

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.*, 30:1441–1459.