Local Smoothing Methods with Regularization in Nonparametric Regression Models

Mohammad Fawaz Kourabi

LOCAL SMOOTHING METHODS WITH REGULARIZATION IN NONPARAMETRIC REGRESSION MODELS

Mohammad Fawaz Kourabi

Vom Fachbereich Mathematik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

1. Gutachter: Prof. Dr. Jürgen Franke
2. Gutachter: Prof. Dr. Klaus Ritter

Datum der Disputation: 15. August 2011

D 386

To my parents, and Sawsan

Table of Contents

Abstract

Mrázek et al. [14] proposed a unified approach to curve estimation which combines localization and regularization. In this thesis we will use their approach to study some asymptotic properties of local smoothers with regularization. In Particular, we shall discuss the regularized local least squares (RLLS) estimate with correlated errors (more precisely with stationary time series errors), and then based on this approach we will discuss the case when the kernel function is dirac function and compare our smoother with the spline smoother. Finally, we will do some simulation study.

Acknowledgements

First I would like to thank Prof. Dr. Jürgen Franke, my supervisor, for his valuable suggestions and guidance throughout this work. I am also thankful to Mrs. Beata Siegler for her friendly assistance.

I would also like to thank Prof. Dr. Klaus Ritter for accepting to be the second referee of this work.

I am thankful to University of Damascus in Syria for providing me the financial support to get higher studies in Germany.

Of course, I am grateful to my parents for their prayers and *love*. Without them this work would never have come into existence (literally).

Words alone cannot express the thanks I owe to my wife Sawsan, for her encouragements and support that was a critical ingredient in providing me the strength and persistence necessary to complete my study.

Finally, I would like to thank the members of the statistics Group at the University of Kaiserslautern for the friendly atmosphere, and for the fruitful discussions.

Kaiserslauetrn — Mohammad Fawaz Kourabi
August 15, 2011

Chapter 1

Preliminaries

In this small chapter we will see the basic definitions of the Landau symbols in normed vector spaces. In mathematical statistics by letting the sample size N tend to infinity, we are confronted with sequences of vectors and matrices increasing in size. These sequences as they are do not form normed vector spaces and thus the mathematical tools do not apply immediately. A way out is to redefine these sequences in a way such that the resulting sequences form normed vector spaces. Thus, the mathematical tools hold without any problem.

1.1 Preliminary notations and definitions

Definition 1.1.1. Let $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ be sequences of real numbers. We define the following notations:

1. $a_n = O\left(b_n\right)$ if and only if $\left|\frac{a_n}{b_n}\right| \leq c$ for some $c > 0$ and all $n \in \mathbb{N}$.

2. $a_n = o\left(b_n\right)$ if and only if $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 0$.

3. $a_n \Theta b_n$ if and only if $a_n = O(b_n)$ and $b_n = O(a_n)$.

4. $a_n \sim b_n$ if and only if $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 1$ (or equivalently $a_n = b_n + o(b_n)$).

Remark 1.1.1. Let $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ be sequences of real numbers. Then

$$\frac{1}{2}(a_n + b_n) \leq a_n \vee b_n \leq a_n + b_n \Rightarrow (a_n \vee b_n)\,\Theta\,(a_n + b_n)$$

$$\frac{a_n b_n}{a_n + b_n} \leq a_n \wedge b_n \leq \frac{2a_n b_n}{a_n + b_n} \Rightarrow (a_n \wedge b_n)\,\Theta \frac{a_n b_n}{a_n + b_n}$$

where $x \wedge y := \min(x, y)$, $x \vee y := \max(x, y)$ for $x, y \in \mathbb{R}$.

Definition 1.1.2. Let $\{v_n\}_{n\in\mathbb{N}}$ be a sequence in a normed vector space $(V, \|\cdot\|)$, and $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers. We define the following notations:

1. $v_n = O(a_n)$ if and only if $\|v_n\| = O(a_n)$.

2. $v_n = o(a_n)$ if and only if $\|v_n\| = o(a_n)$.

Definition 1.1.3. Let $A_1, A_2, \ldots,\ B_1, B_2, \ldots,$ and A be real random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1. $A_n = O_p(B_n)$ iff $\forall\varepsilon > 0\ \exists M > 0$ such that $\mathbb{P}\left\{\left|\frac{A_n}{B_n}\right| > M\right\} < \varepsilon$, $\forall n \in \mathbb{N}$.

2. $A_n = o_p(B_n)$ iff $\forall\varepsilon > 0\ \lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{A_n}{B_n}\right| > \varepsilon\right\} = 0$.

3. $A_n \approx B_n$ iff $A_n = B_n + o_p(B_n)$.

4. convergence almost surely: $A_n \xrightarrow{a.s.} A$ iff $\mathbb{P}\{\lim_{n\to\infty} A_n = A\} = 1$.

5. convergence in probability: $A_n \xrightarrow{P} A$ iff $A_n - A = o_p(1)$, as $n \to \infty$.

6. convergence in rth mean: $A_n \xrightarrow{r} A$ iff $\mathbb{E}\,|A_n - A|^r = o(1)$, as $n \to \infty$ where $r > 0$.

7. convergence in distribution: $A_n \xrightarrow{\mathcal{L}} A$ iff $\mathbb{P}\{A_n < x\} \to \mathbb{P}\{A < x\} = F(x)$, at every point of continuity of $F(x)$ as $n \to \infty$.

Proposition 1.1.1. *Let $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$, $\{c_n\}_{n\in\mathbb{N}}$ be positive sequences of real numbers. Then*

i) $O(a_n) + O(b_n) = O(\max(a_n, b_n))$.

ii) if $c_n = O(a_n)$ and $c_n = O(b_n)$, then $c_n = O(\min(a_n, b_n))$.

iii) $O(a_n) O(b_n) = O(a_n b_n)$, $O(a_n b_n) = a_n O(b_n)$.

iv) $(O(a_n))^r = O(a_n^r)$ *for* $r > 0$.

v) $o(a_n) O(b_n) = o(a_n b_n)$.

vi) the statements i), ii), iii), iv) remain valid if O is everywhere replaced by o, O_p, or o_p.

vii) the statement v) remains valid if o, O is everywhere replaced by o_p, O_p respectively.

Proposition 1.1.2. *Let $X, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $r > 0$. The following implications hold:*

1. $X_n \xrightarrow{a.s.} X \;\Rightarrow\; X_n \xrightarrow{P} X$.

2. $X_n \xrightarrow{r} X \;\Rightarrow\; X_n \xrightarrow{P} X$.

3. $X_n \xrightarrow{P} X \;\Rightarrow\; X_n \xrightarrow{\mathcal{L}} X$.

4. $X_n \xrightarrow{P} X,\ |X_n| \le |Y|$ *a.s.* $\forall n \in \mathbb{N}$, *and* $\mathbb{E}|Y|^r < \infty \;\Rightarrow\; X_n \xrightarrow{r} X$.

Definition 1.1.4. A function $g : \mathbb{R} \to \mathbb{R}$ is called Lipschitz continuous on the interval $[a, b]$ with constant $L > 0$ if the following condition holds:

$$|g(x) - g(y)| \leq L\,|x - y| \;\text{ for all } x, y \in [a, b]$$

Definition 1.1.5. A function $g : \mathbb{R} \to \mathbb{R}$ is called Hölder continuous on the interval $[a, b]$ with constant $H > 0$ and exponent $0 < \beta < 1$ if the following condition holds:

$$|g(x) - g(y)| \leq H\,|x - y|^{\beta} \;\text{ for all } x, y \in [a, b]$$

Definition 1.1.6. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is a function. Then the support of f (written as $\operatorname{supp} f$), is the set

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R} \mid f(x) \neq 0\}}.$$

Definition 1.1.7. The following function is called Huber function with parameter c,

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } |u| \leq c; \\ c|u| - \frac{1}{2}c^2 & \text{if } |u| > c. \end{cases} \tag{1.1.1}$$

1.2 Building the normed vector space of Matrices

Consider the infinity-norm for vectors and its induced norm for matrices. That is,

$$\|\mathbf{u}\|_{\infty} := \sup_{1 \leq i \leq N} |u_i|$$

for all vectors $\mathbf{u} = (u_1, \ldots, u_N)^T \in \mathbb{R}^N$. And

$$\|A\|_{\infty} := \sup_{1 \leq i \leq N} \sum_{j=1}^{N} |A_{ij}|$$

for all matrices $A = (A_{ij})_{i,j=1,\ldots,N} \in \mathcal{M}^{N\times N}(\mathbb{R})$, where $\mathcal{M}^{N\times N}(\mathbb{R})$ is the set of all $N \times N$ matrices.

A technical problem we confront in our work is that the vectors and matrices we are dealing with come from $\mathbb{R}^N$ and $\mathcal{M}^{N\times N}(\mathbb{R})$ respectively. Whenever N is fixed there is no problem. But since we are interested in an asymptotic analysis we shall have $N \to \infty$. Then, for different sample sizes we have vectors and matrices of different sizes. This means that these vectors no longer form a vector space and same holds for the matrices. Thus, we have to define a new normed vector space in order to overcome this problem!

Consider the set $\mathcal{M}^\infty(\mathbb{R})$ of all real infinite-dimensional matrices $A = (A_{ij})_{i,j\in\mathbb{N}}$ satisfying

$$\sup_{i\in N}\sum_{j=1}^{\infty}|A_{ij}| < \infty$$

This can be turned into a vector space by defining vector addition as

$$(A_{ij})_{i,j\in\mathbb{N}} + (B_{ij})_{i,j\in\mathbb{N}} = (A_{ij} + B_{ij})_{i,j\in\mathbb{N}}$$

and the scalar multiplication as

$$\alpha(A_{ij})_{i,j\in\mathbb{N}} = (\alpha A_{ij})_{i,j\in\mathbb{N}}$$

The real-valued operation $\|\cdot\|_{\mathcal{M}^\infty(\mathbb{R})}$ defined by

$$\|A\|_{\mathcal{M}^\infty(\mathbb{R})} := \sup_{i\in\mathbb{N}}\sum_{j=1}^{\infty}|A_{ij}|$$

defines a norm on $\mathcal{M}^\infty(\mathbb{R})$. In fact, $\mathcal{M}^\infty(\mathbb{R})$ is a complete metric space with respect to this norm, and therefore is a Banach space.

The normed space $\left(\mathcal{M}^{N\times N}(\mathbb{R}), \|\cdot\|_\infty\right)$ is isometrically imbedded in $\Big(\mathcal{M}^\infty(\mathbb{R}), \|\cdot\|_{\mathcal{M}^\infty(\mathbb{R})}\Big)$ since the mapping

$$T_1 : \mathcal{M}^{N\times N}(\mathbb{R}) \to \mathcal{M}^\infty(\mathbb{R}) : A = (A_{ij})_{i,j\in\mathbb{N}} \mapsto A' = (A'_{ij})_{i,j\in\mathbb{N}}$$

where

$$A'_{ij} = \begin{cases} A_{ij} & \text{if } i,j \in \{1,\ldots,N\}\,; \\ 0 & \text{if } i,j \in \mathbb{N}\setminus\{1,\ldots,N\} \end{cases}$$

is isometric imbedding. Thus we can write $\mathcal{M}^{N\times N}(\mathbb{R}) \subset \mathcal{M}^{\infty}(\mathbb{R})$ and we have for $A \in \mathcal{M}^{N\times N}(\mathbb{R})$,

$$\|A\|_{\infty} = \|A\|_{\mathcal{M}^{\infty}(\mathbb{R})}$$

Similar for the N-dimensional vectors, the normed space $\left(\mathbb{R}^N, \|\cdot\|_{\infty}\right)$ is isometrically imbedded in the space $\left(\mathcal{M}^{\infty}(\mathbb{R}), \|\cdot\|_{\mathcal{M}^{\infty}(\mathbb{R})}\right)$ since the mapping

$$T_2 : \mathbb{R}^N \to \mathcal{M}^{\infty}(\mathbb{R}) : \mathbf{u} = (u_1,\ldots,u_N)^T \mapsto \mathbf{u}' = (u'_{ij})_{i,j\in\mathbb{N}}$$

where

$$u'_{ij} = \begin{cases} u_i & \text{if } i \in \{1,\ldots,N\} \text{ and } j = 1; \\ 0 & \text{otherwise} \end{cases}$$

is isometric imbedding. Thus we can write $\mathbb{R}^N \subset \mathcal{M}^{\infty}(\mathbb{R})$ and we have for $\mathbf{u} \in \mathbb{R}^N$,

$$\|\mathbf{u}\|_{\infty} = \|\mathbf{u}\|_{\mathcal{M}^{\infty}(\mathbb{R})}$$

Throughout our work we will use the norm $\|\cdot\|_{\mathcal{M}^{\infty}(\mathbb{R})}$ for matrices and vectors which is identical with the infinity norm on $\mathcal{M}^{N\times N}(\mathbb{R})$ and $\mathbb{R}^N$ and we write shortly $\|\cdot\|$ instead of $\|\cdot\|_{\mathcal{M}^{\infty}(\mathbb{R})}$.

Chapter 2

Estimation and smoothing

In this chapter, we introduce the general idea of smoothing and, in particular, kernel smoothing, and spline smoothing. Then we introduce the general approach for image denoising developed by Mrázek et al. [14]. Based on this approach we present some spacial cases.

2.1 Introduction

Statisticians realized that pure parametric thinking in curve estimations often does not meet the need for flexibility in data analysis and the development of hardware created the demand for theory of now computable nonparametric estimates. A regression curve describes a general relationship between an explanatory variable X and a response variable Y. If N data points $\{(X_i, Y_j)\}_{i=1}^{N}$ have been collected, the regression relationship can be modeled as

$$Y_j = m(X_j) + \varepsilon_j, \qquad j = 1, ..., N, \qquad X_j, Y_j \in \mathbb{R}$$

with unknown regression function m and observation errors ε_j. The task of approximating the mean function can be done essentially in two ways. The quite often used parametric approach and as alternative one could estimate m nonparametrically without reference to a specific form. A preselected parametric model might be too restrictive or of too low-dimension to fit unexpected features, whereas the nonparametric smoothing approach offers a flexible tool in analyzing unknown regression relationships. It has been observed that parametric models are too restrictive to give reasonable explanations of observed phenomena.

Definition 2.1.1. The general nonparametric regression model is defined as:

$$Y_j = m(X_j) + \varepsilon_j, \qquad j = 1, ..., N, \qquad X_j, Y_j \in \mathbb{R} \tag{2.1.1}$$

where m is unknown function and $\varepsilon_1, ..., \varepsilon_N$ denote zero-mean random variables with variance $\sigma_\varepsilon^2 < \infty$.

Definition 2.1.2. If $X_1, ..., X_N$ are assumed to be independent and identically distributed random variables and additionally independent of $\varepsilon_1, ..., \varepsilon_N$ then the model (2.1.1) is called stochastic design model:

$$\left.\begin{array}{l} Y_j = m(X_j) + \varepsilon_j, \qquad j = 1, ..., N \\ X_1, ..., X_N \text{ i.i.d. with density } p(x) \text{ and independent of } \varepsilon_1, ..., \varepsilon_N \end{array}\right\} \tag{2.1.2}$$

In this case we have

$$m(x) = \mathbb{E}(Y_j \mid X_j = x), \qquad j = 1, ..., N.$$

Definition 2.1.3. If $X_1, ..., X_N$ are fixed $X_j = x_j,\ j = 1, ..., N$ then the model (2.1.1) is called deterministic design model.

In many experiments the fixed points x_j are taken from an equidistant grid on an interval $[a, b]$; without loss of generality, it can be assumed that $[a, b] = [0, 1]$.

The deterministic equidistant model:

$$\left.\begin{aligned} &Y_j = m(x_j) + \varepsilon_j, \qquad j = 1, ..., N \\ &m : [0,1] \to \mathbb{R} \\ &x_j = \tfrac{j}{N}, \qquad j = 1, ..., N \end{aligned}\right\} \tag{2.1.3}$$

2.2 Local-average methods, basic idea of smoothing

Consider general nonparametric regression model (2.1.1). If m is believed to be smooth, then the observations at X_j near x should contain information about the value of m. Thus it should be possible to use something like a local average of the data near x to construct an estimator of $m(x)$.

For the trivial case, in which the regression curve $m(x)$ is a constant, estimation of m reduces to the point estimation of location, since an average over the response variable Y yields an estimate of m, i.e.

$$\text{if } m(x) = m_0, \text{ then } \frac{1}{N}\sum_{j=1}^{N} Y_j \xrightarrow{P} m_0 \text{ by law of large numbers.}$$

If the assumed curve is modelled as a smooth continuous function of a particular structure which is "nearly constant" in small neighborhoods around x, then estimate $m(x)$ by local average:

$$\frac{1}{N_x}\sum_{j=1}^{N} 1_{[x-h,x+h]}(X_j).Y_j,$$

where N_x = number of non-vanishing summands. This local averaging procedure can be viewed as the basic idea of smoothing. More formally, this procedure can be written in the general form:

$$\hat{m}(x) = \frac{1}{N} \sum_{j=1}^{N} W_{Nj}(x) \cdot Y_j, \tag{2.2.1}$$

where the weights $W_{Nj}(x)$ are large if $|x - X_j|$ small, and they may depend on all $X_1, X_2, ..., X_N$ simultaneously.

Definition 2.2.1. The regression estimator $\hat{m}(x)$ of the form (2.2.1) is called weighted local averages estimator or smoother.

2.3 Kernel methods

A conceptually simple approach to a representation of the weight sequence $\{W_{Nj}(x)\}_{j=1}^{N}$ is to describe the shape of weight function $W_{Nj}(x)$ by a density function with a scale parameter that adjusts the size and the form of the weight near x. It is quite common to refer to this shape function as a kernel K.

Definition 2.3.1. A function $K : \mathbb{R} \to \mathbb{R}$ is called a kernel if it is bounded, continuous on its support, and satisfies $\int\limits_{-\infty}^{+\infty} K(u)du = 1$.
The kernel function $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ where K is kernel and $h > 0$ is called the rescaled kernel with bandwidth h .

Some examples of kernels are:

- uniform kernel: $K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \leq 1; \\ 0 & \text{if } |u| > 1. \end{cases}$

- triangle kernel: $K(u) = \begin{cases} 1 - |u| & \text{if } |u| \leq 1; \\ 0 & \text{if } |u| > 1. \end{cases}$

- Epanechnikov kernel: $K\left(u\right)=\begin{cases}\frac{3}{4}\left(1-u^2\right) & \text{if } |u|\leq 1;\\ 0 & \text{if } |u|>1.\end{cases}$

- Gaussian Kernel: $K\left(u\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$.

Definition 2.3.2. Consider the deterministic equidistant design model (2.1.3). The Priestley-Chao (PC) kernel estimator of m with bandwidth (or smoothing parameter) $h>0$ is defined as:

$$\hat{m}_{PC}(x,h)=\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)Y_j,\qquad x\in[0,1]$$

where K is kernel function.

The corresponding weights $W_{Nj}\left(x\right)=K_h\left(x-x_j\right)$.

Definition 2.3.3. Consider the deterministic equidistant design model (2.1.3). The Gasser-Müller (GM) kernel estimator of m with bandwidth (or smoothing parameter) $h>0$ is defined as:

$$\hat{m}_{GM}(x,h)=\sum_{j=1}^{N}\int_{s_{j-1}}^{s_j}K_h(x-u)du\cdot Y_j,\qquad x\in[0,1]$$

where K is kernel function, $s_0=0$, $s_j=\frac{x_j+x_{j+1}}{2}$ for $j=1,...,N-1$ and $s_N=1$.

The corresponding weights $W_{Nj}\left(x\right)=N\int_{s_{j-1}}^{s_j}K_h(x-u)du\cdot Y_j$.

Definition 2.3.4. Consider the stochastic design model (2.1.2). The Rosenblatt-Parzen (RP) kernel density estimate of $p\left(x\right)$ with bandwidth $h>0$ is defined as:

$$\hat{p}\left(x,h\right)=\frac{1}{N}\sum_{j=1}^{N}K_h(x-X_j)$$

and the Nadaraya-Watson (NW) kernel estimate of m with bandwidth $h>0$ is defined as:

$$\hat{m}_{NW}(x,h)=\frac{1}{N}\sum_{j=1}^{N}K_h(x-X_j)Y_j/\hat{p}\left(x,h\right)$$

where K is kernel function.

The corresponding weights $W_{Nj}(x) = K_h(x - X_j)/\hat{p}(x,h)$.

Definition 2.3.5. (local and global error measures) Consider the deterministic equidistant design model (2.1.3) and $\hat{m}(\cdot)$ is an estimator of $m(\cdot)$.

1. The mean squared error of $\hat{m}(\cdot)$ at a point x is defined as:

$$\operatorname{mse}(\hat{m}(x)) = \mathbb{E}(\hat{m}(x) - m(x))^2 \tag{2.3.1}$$

2. The mean integrated squared error of $\hat{m}(\cdot)$ on the interval $[a,b]$ is defined as:

$$\operatorname{mise}(\hat{m}(\cdot)) = \mathbb{E}\int_a^b (\hat{m}(x) - m(x))^2 w(x)\,dx \tag{2.3.2}$$

   where w is a nonnegative weight function.

   The mise can be written as

$$\operatorname{mise}(\hat{m}(\cdot)) = \int_a^b \operatorname{mse}(\hat{m}(x))\, w(x)\,dx$$

The following theorem provides an asymptotic expansion of the mean-squared error and, as an immediate consequence, the consisting of the PC-estimate. Such results are well known in the literature, compare, e.g., the monograph by Härdle [1]. We want to stress the precise form of the error term as a function of N and h and, therefore, give a proof of the following version.

Theorem 2.3.1. *Assume the deterministic equidistant design model (2.1.3) with i.i.d. errors, $\hat{m}(x,h)$ is PC-estimate of $m(x)$, and*

1. *K is a nonnegative, symmetric kernel with compact support $[-1,1]$,*

2. *K is Lipschitz continuous on* $[-1, 1]$,

3. *m is twice continuously differentiable on* $[0, 1]$,

4. *m'' is Hölder continuous on* $[0, 1]$ *with exponent* $0 < \beta < 1$,

5. $N \to \infty$, $h \to 0$, $Nh^2 \to \infty$.

Then

(i) $\operatorname{bias}(\hat{m}(x,h)) = \mathbb{E}(\hat{m}(x,h)) - m(x) = \frac{h^2}{2} m''(x) V_K + O\left(h^{2+\beta}\right) + O\left(\frac{1}{Nh^2}\right)$ *uniformly in* $x \in [h, 1-h]$, *where*

$$V_K = \int_{-1}^{1} u^2 K(u) du$$

(ii) $\operatorname{Var}(\hat{m}(x,h)) = \frac{\sigma_\varepsilon^2}{Nh} Q_K + O\left(\frac{1}{N^2 h^3}\right)$ *uniformly in* $x \in [h, 1-h]$, *where*

$$Q_K = \int_{-1}^{1} K^2(u) du$$

$$\sigma_\varepsilon^2 = \operatorname{Var}(\varepsilon_j), \quad j = 1, ..., N$$

(iii) *the mean squared error of* $\hat{m}(x,h)$ *is*

$$\operatorname{mse}(\hat{m}(x,h)) = \operatorname{amse}(\hat{m}(x,h)) + O\left(h^{4+\beta}\right) + O\left(\frac{1}{N^2 h^4}\right) + O\left(\frac{1}{N}\right)$$

uniformly in $x \in [h, 1-h]$, *where the asymptotic mean squared error is given by*

$$\operatorname{amse}(\hat{m}(x,h)) = \frac{\sigma_\varepsilon^2}{Nh} Q_K + \frac{h^4}{4} \left(m''(x)\right)^2 V_K^2 \qquad (2.3.3)$$

In particular,

$$\hat{m}(x,h) \xrightarrow{P} m(x)$$

i.e. $\hat{m}(x,h)$ *is consistent estimator of* $m(x)$.

Two lemmas before the proof:

Lemma 2.3.2. *Let $g(y)$ be Lipschitz continuous on $[0,1]$ with Lipschitz constant C. Then,*

$$\left|\int_0^1 g(y)\,dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j)\right| \leq \frac{C}{N}$$

where $x_j = \frac{i}{N}$, $j = 1, ..., N$. i.e.

$$\frac{1}{N}\sum_{j=1}^{N} g(x_j) = \int_0^1 g(y)\,dy + O(\frac{1}{N}).$$

Proof. We have

$$\left|\int_0^1 g(y)\,dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j)\right| = \left|\sum_{j=1}^{N}\left\{\int_{x_{j-1}}^{x_j} g(y)\,dy - \frac{1}{N}g(x_j)\right\}\right|$$

By the mean-value theorem of integration there are $y_j \in [x_{j-1}, x_j]$, $j = 1, ..., N$, where $x_0 = 0$, such that:

$$\int_{x_{j-1}}^{x_j} g(y)\,dy = (x_j - x_{j-1})\,g(y_j) = \frac{1}{N}g(y_j)$$

Therefore,

$$\begin{aligned}\left|\int_0^1 g(y)\,dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j)\right| &= \left|\sum_{j=1}^{N}\frac{1}{N}\{g(y_j) - g(x_j)\}\right| \\ &\leq \frac{1}{N}\sum_{j=1}^{N}|g(y_j) - g(x_j)| \\ &\leq \frac{1}{N}\sum_{j=1}^{N} C\,|y_j - x_j| \\ &\leq \frac{C}{N}\sum_{j=1}^{N}\frac{1}{N} \leq \frac{C}{N}\end{aligned}$$

□

Lemma 2.3.3. *Assume a kernel K, and a function $m:[0,1]\to\mathbb{R}$ satisfying conditions 1.-4. of Theorem 2.3.1. Then uniformly for all $x\in[h,1-h]$, we have*

i) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\,m(x_j)=\int_{-1}^{1}K(u)\,m(x-hu)\;du+O\left(\frac{1}{Nh^2}\right).$

ii) $\displaystyle \frac{1}{N}\sum_{j=1}^{N}K_h^2(x-x_j)=\frac{1}{h}\int_{-1}^{1}K^2(u)\;du+O\left(\frac{1}{Nh^3}\right).$

as $N\to\infty$ s.t. $h\to 0$.

Proof. i) The function $g(y)=K_h(x-y)\,m(y)$ is Lipschitz continuous on $[0,1]$ with constant $\frac{C}{h^2}$ since,

$$\begin{aligned}
|K_h(x-y)\,m(y)-K_h(x-z)\,m(z)| &\leq |K_h(x-y)\,m(y)-K_h(x-y)\,m(z)|+\\
&\quad |K_h(x-y)\,m(z)-K_h(x-z)\,m(z)|\\
&\leq |K_h(x-y)|\,|m(y)-m(z)|+\\
&\quad |m(z)|\,|K_h(x-y)-K_h(x-z)|\\
&\overset{\exists\,\xi\in[0,1]}{\leq} \frac{A}{h}\,|m'(\xi)|\,|y-z|+\frac{MC_K}{h}\left|\frac{z-y}{h}\right|
\end{aligned}$$

$$\begin{aligned}
|K_h(x-y)\,m(y)-K_h(x-z)\,m(z)| &\leq \left[\frac{AM'}{h}+\frac{MC_K}{h^2}\right]|y-z|\\
&\overset{h<1}{\leq} \frac{AM'+MC_K}{h^2}\,|y-z|=\frac{C}{h^2}\,|y-z|
\end{aligned}$$

where $A=\sup_u K(u)$, $M=\sup_u m(u)$, $M'=\sup_u m'(u)$, $C=AM'+MC_K$. Therefore, we have by Lemma 2.3.2 with $g(y)=K_h(x-y)\,m(y)$,

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\,m(x_j) &= \int_0^1 K_h(x-y)\,m(y)\;dy+O\left(\frac{1}{Nh^2}\right)\\
&\overset{u=\frac{x-y}{h}}{=} \int_{\frac{x-1}{h}}^{\frac{x}{h}} K(u)\,m(x-hu)\;du+O\left(\frac{1}{Nh^2}\right)\\
&\overset{h\leq x\leq 1-h}{=} \int_{-1}^{1} K(u)\,m(x-hu)\;du+O\left(\frac{1}{Nh^2}\right)
\end{aligned}$$

ii) Using Lemma 2.3.2 with the function $g(y) = K_h^2(x-y)$ which is Lipschitz continuous on $[0,1]$ with constant $\frac{C'}{h^3}$,

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N} K_h^2(x-x_j) &= \int_0^1 K_h^2(x-y)\, dy + O\left(\frac{1}{Nh^3}\right) \\
&\overset{u=\frac{x-y}{h}}{=} \frac{1}{h}\int_{\frac{x-1}{h}}^{\frac{x}{h}} K^2(u)\, du + O\left(\frac{1}{Nh^3}\right) \\
&\overset{h\le x\le 1-h}{=} \frac{1}{h}\int_{-1}^{1} K^2(u)\, du + O\left(\frac{1}{Nh^3}\right)
\end{aligned}$$

□

Proof of Theorem 2.3.1. i) Using Lemma 2.3.3, we have

$$\begin{aligned}
\operatorname{bias}(\hat{m}(x,h)) &= \mathbb{E}(\hat{m}(x,h)) - m(x) \\
&= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, m(x_j) - m(x) \\
&= \int_{-1}^{1} K(u)\, m(x-hu)\, du - m(x) \underbrace{\int_{-1}^{1} K(u)\, du}_{=1} + O\left(\frac{1}{Nh^2}\right) \\
&= \int_{-1}^{1} K(u)\left[m(x-hu) - m(x)\right] du + O\left(\frac{1}{Nh^2}\right)
\end{aligned}$$

and using Taylor expansion we get for some $\theta \in [0,1]$

$$\begin{aligned}
\operatorname{bias}(\hat{m}(x,h)) &= \int_{-1}^{1} K(u)\left[m'(x)(-hu) + \frac{1}{2}m''(x-\theta hu)(hu)^2\right] du + O\left(\frac{1}{Nh^2}\right) \\
&= -hm'(x)\underbrace{\int_{-1}^{1} uK(u)\, du}_{=0} + \frac{1}{2}h^2 m''(x)\int_{-1}^{1} u^2 K(u)\, du \\
&\quad \underbrace{+\frac{1}{2}h^2\int_{-1}^{1} u^2 K(u)\left[m''(x-\theta hu) - m''(x)\right] du}_{=:r(x,h)} + O\left(\frac{1}{Nh^2}\right)
\end{aligned}$$

but $r(x,h) = O\left(h^{2+\beta}\right)$ since, uniformly for all $x \in [h, 1-h]$, we have

$$\begin{aligned}
|r(x,h)| &\leq \frac{1}{2}h^2 \int_{-1}^{1} u^2 K(u) \left|m''(x-\theta hu) - m''(x)\right| du \\
&\leq \frac{1}{2}h^2 \int_{-1}^{1} u^2 K(u) H \left|\theta hu\right|^{\beta} du \\
&\leq \frac{H}{2}h^{2+\beta} \int_{-1}^{1} \theta^{\beta} |u|^{2+\beta} K(u)\, du \\
&\leq \frac{H}{2}h^{2+\beta}
\end{aligned}$$

Therefore, $\text{bias}(\hat{m}(x,h)) = \frac{1}{2}h^2 m''(x) V_K + O\left(h^{2+\beta}\right) + O\left(\frac{1}{Nh^2}\right)$.

ii) Using Lemma 2.3.3,

$$\begin{aligned}
\text{Var}(\hat{m}(x,h)) &= \frac{1}{N^2}\sum_{j=1}^{N} K_h^2(x-x_j)\text{Var}(Y_j) = \frac{\sigma^2}{N^2}\sum_{j=1}^{N} K_h^2(x-x_j) \\
&= \frac{\sigma^2}{Nh}\int_{-1}^{1} K^2(u)du + O\left(\frac{1}{N^2h^3}\right)
\end{aligned}$$

iii) Combining the bias and variance expansion, we get for $N \to \infty$, $h \to 0$, $Nh^2 \to \infty$

$$\begin{aligned}
\text{mse}(\hat{m}(x,h)) &= \text{Var}(\hat{m}(x,h)) + (\text{bias}(\hat{m}(x,h)))^2 \\
&= \frac{\sigma^2}{Nh}Q_K + \frac{h^4}{4}(m''(x))^2 V_K^2 + O\left(h^{4+2\beta}\right) + O\left(\frac{1}{N^2h^4}\right) \\
&\quad + m''(x) O\left(h^{4+\beta}\right) + m''(x) O\left(\frac{1}{N}\right) + O\left(\frac{h^{\beta}}{N}\right) + O\left(\frac{1}{N^2h^3}\right)
\end{aligned}$$

As m'' is bounded, the assertion of iii) holds uniformly for all $x \in [h, 1-h]$. □

Corollary 2.3.4. *Assume the conditions of Theorem 2.3.1 and* $w(x)$ *is a continuous weight function with support* $[\delta, 1-\delta]$ *and* $w(x) > 0$ *for all* $x \in (\delta, 1-\delta)$ *and some* $\delta > 0$. *Then*

$$\text{mise}(\hat{m}(.,h)) = \text{amise}(\hat{m}(.,h)) + O\left(h^{4+\beta}\right) + O\left(\frac{1}{N^2h^4}\right) + O\left(\frac{1}{N}\right)$$

where the asymptotic mean integrated squared error are given by

$$\text{amise}(\hat{m}(\cdot,h)) = \frac{\sigma_\varepsilon^2}{Nh}Q_K \int w(x)\,dx + \frac{h^4}{4}V_K^2 \int (m''(x))^2 w(x)\,dx \tag{2.3.4}$$

2.3.1 Bandwidth selection

The choice of an appropriate bandwidth h plays a prominent role in nonparametric regression. For each nonparametric regression method, one has to choose how much to smooth for the given dataset. In this section we consider the deterministic equidistant design model (2.1.3) with i.i.d. errors and the PC-estimator $\hat{m}(\cdot, h)$ as an estimator of $m(\cdot)$. In the literature several bandwidth selection procedures have been presented that optimize quadratic error measures for the regression curve and its derivatives.

Definition 2.3.6. Let $d(h)$ be an performance criterion. We say that a bandwidth sequence h^* is asymptotically optimal w.r.t $d(h)$ if

$$\frac{d(h^*)}{\inf_{h \in H_N} d(h)} \xrightarrow{P} 1,$$

as $N \to \infty$, where H_N is the range of permissible bandwidth.

There are a number of alternative optimality criteria in use. Firstly, we may be interested in the quadratic loss of the estimator at a single point x, which is measured by the mean squared error $\operatorname{mse}(\hat{m}(x, h))$. Secondly, we may be only concerned with a global measure of performance. In this case we may consider the mean integrated squared error, $d_M(h) = \operatorname{mise}(\hat{m}(\cdot, h))$.

Corollary 2.3.5. *Under the conditions of Theorem 2.3.1,*

$$h_{\text{mse}}(x) = \left(\frac{\sigma_\varepsilon^2 Q_K}{(m''(x))^2 V_K^2} \right)^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{2.3.5}$$

is asymptotically optimal bandwidth at x with respect to mse.
If $h \sim N^{-\frac{1}{5}}$ then $\operatorname{mse}(\hat{m}(x, h)) \sim \operatorname{amse}(\hat{m}(x, h)) \sim \left(\sigma_\varepsilon^2 Q_K + \frac{1}{4}(m''(x))^2 V_K^2\right) N^{-\frac{4}{5}}$.

Corollary 2.3.6. *Under the conditions of Corollary 2.3.4, and also m'' does not vanish identically on $(\delta, 1-\delta)$,*

$$h_{\text{mise}} = \left(\frac{\sigma_\varepsilon^2 Q_K \int w(x)\, dx}{V_K^2 \int (m''(x))^2 w(x)\, dx} \right)^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{2.3.6}$$

is asymptotically optimal bandwidth with respect to mise.
If $h \sim N^{-\frac{1}{5}}$ then $\text{mise}(\hat{m}(\cdot, h)) \sim \text{amise}(\hat{m}(\cdot, h)) \sim \left(\sigma_\varepsilon^2 Q_K I_1 + \frac{1}{4} V_K^2 I_2\right) N^{-\frac{4}{5}}$*, where* $I_1 = \int w(x)\, dx,\ I_2 = \int (m''(x))^2 w(x)\, dx$.

As seen from either (2.3.5), (3.3.12) the asymptotically optimal bandwidth depends on m'' and σ_ε^2 which are unknown. As an equivalent alternative to mise we consider its in-sample version, the averaged squared error

$$d_A(h) = \text{ase}(\hat{m}(\cdot, h)) := \frac{1}{N} \sum_{i=1}^{N} (\hat{m}(x_i, h) - m(x_i))^2 w(x_i)$$

where $w(x)$ is nonnegative weight function.

Proposition 2.3.7 (Marron and Härdle, [1])**.**

$$\frac{\text{ase}(\hat{m}(\cdot, h))}{\text{mise}(\hat{m}(\cdot, h))} \xrightarrow{a.s.} 1$$

uniformly in $h \in \left[\frac{1}{N^{1-\delta}}, \frac{1}{N^\delta}\right]$ *for arbitrary* $0 < \delta < \frac{1}{2}$.

2.3.2 Leave-one-out cross-validation method

Cross-validation is a convenient method of global bandwidth choice for many problems and relies on the well established principle of out-of-sample predictive validation. Suppose that optimality with respect to $d_A(h)$ is the aim. We must first replace $d_A(h)$ by a computable approximation to it. A naive estimate would be to just replace the unknown values $m(x_j)$ by the observations Y_j at x_j:

$$\pi(h) = \frac{1}{N}\sum_{j=1}^{N}\left[Y_j - \hat{m}(x_j, h)\right]^2 w(x_j)$$

which is called the averaged squared prediction error. Unfortunately, $\pi(h)$ is a biased estimate of $d_A(h)$.

Figure 2.1 shows that $\hat{h}_1 = \operatorname{argmin}_h \pi(h)$ is strongly biased estimate of $\operatorname{argmin}_h d_A(h)$! The intuitive reason for the bias in $\pi(h)$ is that the observation Y_j is used (in $\hat{m}(x_j, h)$) to predict itself. There are several ways to find an unbiased estimate of ase, one of them is leave-one-out-technique. The simplest way to avoid this problem is to remove

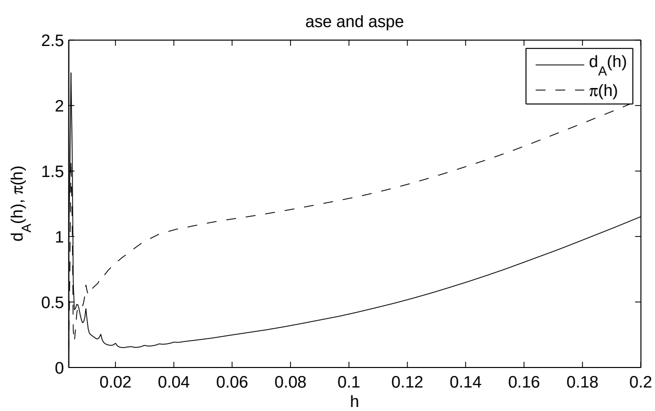

Figure 2.1: The averaged squared error and the averaged squared prediction error for a simulated data set $\{(x_i, Y_j)\}_{i=1,\ldots,N}$ generated from $Y_j = m(x_j) + \varepsilon_j$, $x_j = \frac{j}{N}$, $\varepsilon_1, \ldots, \varepsilon_N$ i.i.d. $\mathcal{N}(0,1)$, and $m(x) = 2 + 5x + \frac{25}{2\sqrt{2\pi}}\exp\left\{-98\left(x - \frac{1}{2}\right)^2\right\}$. The weight function was $w(x) = 1$.

the j-th observation. Consider the cross-validation function

$$CV(h) := \frac{1}{N}\sum_{j=1}^{N}\left[Y_j - \hat{m}_{-j}(x_j, h)\right]^2 w(x_j)$$

as an alternative estimate of $d_A(h)$ where

$$\hat{m}_{-j}(x, h) := \frac{1}{N}\sum_{\substack{i=1 \\ i\neq j}}^{N} K_h(x - x_i)Y_i, \quad j = 1, \ldots, N$$

is the leave-one-out estimate of $m(x)$. The cross-validatory selection of bandwidth is

$$\hat{h}_{CV} = \underset{h}{\operatorname{argmin}}\, CV(h)$$

The CV-function for a simulated data set is shown in Figure 2.2. We note that $\operatorname{argmin}_h CV(h) \simeq \operatorname{argmin}_h d_A(h)$.

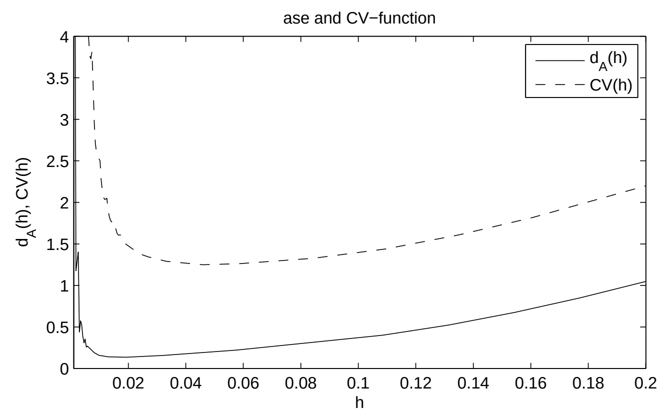

Figure 2.2: The CV-function and the averaged squared error for a simulated data set $\{(x_i, Y_j)\}_{i=1,\dots,N}$ generated from $Y_j = m(x_j) + \varepsilon_j$, $x_j = \frac{j}{N}$, $\varepsilon_1, \dots, \varepsilon_N$ i.i.d. $\mathcal{N}(0,1)$, and $m(x) = 2 + 5x + \frac{25}{2\sqrt{2\pi}} \exp\left\{-98\left(x - \frac{1}{2}\right)^2\right\}$. The weight function was $w(x) = 1$.

Theorem 2.3.8 (Marron and Härdle, [1]). *Under smoothness and other regularity assumptions on K, m and moment conditions on ε_j, $\hat{h}_{CV}$ is asymptotically optimal with respect to* mise.

2.4 Spline Smoothing

Another well known method in nonparametric regression estimation is the method of spline smoothing. For example, under model (2.1.3), the spline estimator $\hat{m}_{SP}(x, \lambda)$

is defined as the minimizer of

$$S_\lambda(g) = \frac{1}{N}\sum_{j=1}^{N}\left(Y_j - g\left(x_j\right)\right)^2 + \lambda\int_0^1 \left(g''\left(x\right)\right)^2 dx$$

over functions g which are twice continuously differentiable. The parameter $\lambda > 0$ is a smoothing parameter which controls the trade-off between smoothness (measured here by the total curvature $\int_0^1 \left(g''\left(x\right)\right)^2 dx$) and goodness of fit to the data (measured here by the least-squares). The larger the value of λ the smoother the estimate.

This idea of spline smoothing is due to Schoenberg [15] and Reinsch [16]. However, the idea of penalizing a measure of goodness of fit by a one for roughness was described already by Whittaker [17]. Such that this approach is called regularization in image analysis and other branches of mathematics.

In 1984, Silverman [18] showed that spline smoothers (which could be written as in (2.2.1) with weights $W_{Nj}^{(\lambda)}(x)$) are asymptotically equivalent to kernel estimates.

Theorem 2.4.1 (Silverman, [18]). *Assume the fixed design model (2.1.3). Then the spline smoother $\hat{m}_{SP}(x, \lambda)$ for (x_j, Y_j), $j = 1, \ldots, N$, can be written in the form*

$$\hat{m}_{SP}\left(x, \lambda\right) = \frac{1}{N}\sum W_{Nj}^{(\lambda)}(x) Y_j \text{ for } x \in [0, 1]$$

where, under the conditions $N \to \infty$, $\lambda \to 0$, $\lambda N^{1-\varepsilon} \to \infty$ for some $\varepsilon > 0$, the weights $W_{Nj}^{(\lambda)}(x)$ satisfy that

$$W_{Nj}^{(\lambda)}(x) \sim \frac{1}{h} K_{SP}\left(\frac{x - X_j}{h}\right) \text{ as } N \to \infty$$

uniformly for all x "not too close to the boundary", where

$$h = \sqrt[4]{\frac{\lambda}{N}},$$

and

$$K_{SP}(u) = \frac{1}{2} e^{\frac{-|u|}{\sqrt{2}}} \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right)$$

The kernel K_{SP} is symmetric with $\int u^2 K_{SP}(u)du = 0$.

2.5 Approach by Mrázek et al.

Pavel Mrázek et al. [14] established a general approach for image denoising which combines localization and regularization. For localization, we add a weight function to the target function to be minimized; we will see an example in problem (3.2.1), which give the NW kernel estimate as the solution. For regularization, we add a penalizing term which leads to extra tuning just like in spline smoothing. In this approach, we work under the "deterministic Equidistant Design" (2.1.3) which we recall here:

$$\left.\begin{array}{l} Y_j = m(x_j) + \varepsilon_j, \qquad j = 1, ..., N \\ m : [0,1] \to \mathbb{R} \\ x_j = \frac{j}{N}, \qquad j = 1, ..., N \end{array}\right\}$$

Also, we will have a higher freedom in choosing the structure of the target function:

$$Q(\mathbf{u}) = \sum_{i,j=1}^{N} \Psi_D\left(|u_i - Y_j|^2\right) \omega_D\left(|x_i - x_j|^2\right) + \frac{\lambda}{2} \sum_{i,j=1}^{N} \Psi_S\left(|u_i - u_j|^2\right) \omega_S\left(|x_i - x_j|^2\right)$$

where $\mathbf{u} = (u_1, \ldots, u_N)$.

Minimizing Q with respect to $\mathbf{u}$ shall give an estimate of $(m(x_1), \ldots, m(x_N))$, call it $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_1)$.

The Data Loss Function Ψ_D is a penalizing function measuring the fit of $\hat{u}_i$ to the observations $Y_1, \ldots, Y_N$, where the Smoothness Loss Function Ψ_S is penalizing function measuring the smoothness of the solution. The Data and the Smoothness Weighting

Functions ω_D, ω_S take care of the localization effect. The tuning parameter $\lambda \geq 0$ balances between fit and smoothness.

2.5.1 Some Special Cases

Let K, L be Kernel functions and $h, g > 0$. Then

1. If we set $\Psi_D\left(s^2\right) = s^2$, $\omega_D\left(x^2\right) = 1$, and $\lambda = 0$, then $\hat{u}_i$ is the Least-Squares estimate (the mean).

2. If we set $\Psi_D\left(s^2\right) = |s|$, $\omega_D\left(x^2\right) = 1$, and $\lambda = 0$, then $\hat{u}_i$ is the Least-Absolute Deviation estimate (the median). The solution is obtained by the so-called median minimizing property (for example, see [4]).

3. If we set $\Psi_D\left(s^2\right) = \rho(s)$, $\omega_D\left(x^2\right) = 1$, where ρ is the Huber loss function and $\lambda = 0$, then $\hat{u}_i$ is the Huber M-estimate.

4. If we set $\Psi_D\left(s^2\right) = s^2$, $\omega_D\left(x^2\right) = K_h(x)$, and $\lambda = 0$, then $\hat{u}_i$ is the Local Least-Squares estimate (The well known NW-estimate).

5. If we set $\Psi_D\left(s^2\right) = \rho(s)$, $\omega_D\left(x^2\right) = K_h(x)$, and $\lambda = 0$, where ρ is the Huber loss function, then $\hat{u}_i$ is the Local Huber M-estimate.

6. If we set $\Psi_D\left(s^2\right) = s^2$, $\omega_D\left(x^2\right) = K_h(x)$, $\Psi_S\left(s^2\right) = s^2$, $\omega_S\left(x^2\right) = L_g(x)$, and $\lambda \geq 0$, then $\hat{u}_i$ is the Regularized Local Least-Squares estimate.

7. If we set $\Psi_D\left(s^2\right) = \rho(s)$, $\omega_D\left(x^2\right) = K_h(x)$, $\Psi_S\left(s^2\right) = s^2$, $\omega_S\left(x^2\right) = L_g(x)$, and $\lambda \geq 0$, then $\hat{u}_i$ is the Regularized Local Huber M-estimate.

8. If we set $\Psi_D\left(s^2\right)=s^2$, $\omega_D\left(x^2\right)=\delta(x^2)$, where δ is the dirac function, $\Psi_S\left(s^2\right)=s^2$, $\omega_S\left(x^2\right)=L_g(x)$, and $\lambda\geq 0$, then $\hat{u}_i$ is a Quadratic Regularized Interpolation estimate.

9. If we set $\Psi_D\left(s^2\right)=s^2$, $\omega_D\left(x^2\right)=\delta(x^2)$, where δ is the dirac function, $\Psi_S\left(s^2\right)=\rho(s)$, $\omega_S\left(x^2\right)=L_g(x)$, and $\lambda\geq 0$, then $\hat{u}_i$ is a General Regularized Interpolation estimate.

The goal of the next chapter is to consider the case of the Regularized Local Least-Squares estimate and the errors ε_j, $j=1,\ldots,N$ in the model (2.1.3) will be assumed correlated. And in the fourth Chapter we will consider the case 8. where the data Weighting function ω_D is the dirac function

$$\delta(x)=\begin{cases}\infty & \text{if } x=0;\\ 0 & \text{if } x\neq 0.\end{cases}$$

.

Chapter 3

Regularized local least-squares estimator with time series errors

A vast literature now exists on using kernel smoothers to estimate regression functions nonparametrically. Practically most of this literature is based on the assumption that the observed data are uncorrelated. In essence this assumption implies that any observed trends, whether long or short term, are either deterministic in nature or simply anomalous chance occurrences. Such an implication is clearly undesirable. There are many settings, such as in time series analysis, where it is reasonable to model slowly varying trends deterministically, but to explain any other regular behavior in the data by means of a correlated model. Further more, kernel estimators are sometimes used to smooth data which result from processing uncorrelated data. Such processing can introduce correlation into the data to be smoothed. An example of the latter phenomenon occurs in the estimation of heteroscedasticity in regression (see [24]).

In this chapter we shall see some asymptotic properties of the Regularized Local Least-Squares estimate under the deterministic equidistant design model, where we assume that the errors come from a stationary time series. Firstly, we introduce some needed concepts from time series analysis and then we derive the mean-squared error asymptotic expansion. Finally, we introduce how to choose the smoothing parameters well to get a small mean-squared error.

3.1 Basic concepts from time series analysis

Definition 3.1.1. A stochastic process is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.1.2. The functions $\{X_\bullet(\omega), \omega \in \Omega\}$ on T are known as the realizations of the process $\{X_t, t \in T\}$.

Remark 3.1.1. We use the term time series to mean both the data and the process of which it is a realization.

Definition 3.1.3. A real-valued time series $\{X_t, t \in \mathbb{Z}\}$, is said to be stationary if

- $\mathbb{E}|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$,
- $\mathbb{E}X_t = \mu$ for all $t \in \mathbb{Z}$, and
- $\mathrm{Cov}(X_s, X_{s+t}) = r_t$ for all $s, t \in \mathbb{Z}$.

$\{r_t, t \in \mathbb{Z}\}$ is called the autocovariance sequence of $\{X_t\}$.

Remark 3.1.2. $\left\{\rho_t := \frac{r_t}{r_0}, t \in \mathbb{Z}\right\}$ is called the autocorrelation sequence of $\{X_t\}$.

Definition 3.1.4. A real-valued time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if

$$\mathcal{L}(X_{t_1+t}, ..., X_{t_n+t}) = \mathcal{L}(X_{t_1}, ..., X_{t_n})$$

for all $n \geq 1$, $t_1, ..., t_n, t \in \mathbb{Z}$.

Proposition 3.1.1 ([3]). *If $\{r_t, t \in \mathbb{Z}\}$ is the autocovariance sequence of a real-valued stationary time series $\{X_t, t \in \mathbb{Z}\}$, then*

1. $r_0 \geq 0$,
2. $|r_t| \leq r_0$ *for all* $t \in \mathbb{Z}$,
3. $r_t = r_{-t}$ *for all* $t \in \mathbb{Z}$.

Theorem 3.1.2 ([3]). *A sequence of real numbers $\{r_t, t \in \mathbb{Z}\}$ is the autocovariance sequence of a stationary time series $\{X_t, t \in \mathbb{Z}\}$ if and only if there exists a finite symmetric measure F on the measurable space $\left([-\pi, \pi], \mathcal{B}_{[-\pi,\pi]}\right)$, where $\mathcal{B}_{[-\pi,\pi]}$ is Borel-σ-algebra of $[-\pi, \pi]$, such that:*

$$r_t = \frac{1}{2\pi} \int\limits_{[-\pi,\pi]} e^{it\omega} F(d\omega) \text{ for all } t \in \mathbb{Z}$$

Definition 3.1.5. Let $\{r_t, t \in \mathbb{Z}\}$ be the autocovariance sequence of a real-valued stationary time series $\{X_t, t \in \mathbb{Z}\}$. Then the above measure F is called the spectral distribution of $\{X_t, t \in \mathbb{Z}\}$ and if F has a density f w.r.t. Lebesgue measure (i.e. f is nonnegative $\mathcal{B}_{[-\pi,\pi]}$-measurable function such that $F(B) = \int\limits_B f(\omega)\, d\omega$ for all $B \in \mathcal{B}_{[-\pi,\pi]}$) then f is called a spectral density of $\{X_t, t \in \mathbb{Z}\}$ and we can write

$$r_t = \frac{1}{2\pi} \int\limits_{[-\pi,\pi]} e^{it\omega} f(\omega)\, d\omega \text{ for all } t \in \mathbb{Z}$$

Corollary 3.1.3. *A spectral density of a real-valued stationary time series is nonnegative, integrable, and symmetric about zero.*

Theorem 3.1.4 ([3]). *Let* $\{r_t, t \in \mathbb{Z}\}$ *be the autocovariance sequence of a real-valued stationary time series* $\{X_t, t \in \mathbb{Z}\}$ *such that* $\sum_{t=-\infty}^{\infty} |r_t| < \infty$. *Then there exists a spectral density of* $\{X_t, t \in \mathbb{Z}\}$ *given by:*

$$f(\omega) = \sum_{s=-\infty}^{\infty} r_s e^{-is\omega}, \qquad \omega \in [-\pi, \pi]$$

in which case f *is called the spectral density of* $\{X_t, t \in \mathbb{Z}\}$.

Definition 3.1.6. A stochastic process $\{\varepsilon_t, t \in \mathbb{Z}\}$ with $\varepsilon_t,\ t \in \mathbb{Z}$, independent and identically distributed(i.i.d) is called strict white noise if

$$\mathbb{E}\varepsilon_t = 0 \text{ and } \operatorname{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$$

Definition 3.1.7. A stochastic process $\{\varepsilon_t, t \in \mathbb{Z}\}$ with $\varepsilon_t,\ t \in \mathbb{Z}$, uncorrelated is called white noise if

$$\mathbb{E}\varepsilon_t = 0 \text{ and } \operatorname{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty \text{ for all } t \in \mathbb{Z}$$

The strict white noise and the white noise are purely stochastic processes, but for observed data one other can assume that there exists a certain dependence among the data, in particular a linear dependence.

Definition 3.1.8. A stochastic process $\{X_t, t \in \mathbb{Z}\}$ is called an autoregressive process of order $p \geq 1$ [$AR(p)$-process] if

$$X_t = \sum_{k=1}^{p} \alpha_k X_{t-k} + \varepsilon_t, \qquad t \in \mathbb{Z} \tag{3.1.1}$$

where $\alpha_1, ..., \alpha_p \in \mathbb{R}$, $\alpha_p \neq 0$, and $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a white noise. $\{\varepsilon_t\}$ are called innovations.

Taking $p = 1$ we have an $AR(1)$-process with

$$X_t = \alpha X_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z}.$$

Definition 3.1.9. An $AR(p)$-process defined by the equations (3.1.1) is said to be causal if there exists a sequence of constants $\{\beta_k\}_{k\in\mathbb{Z}}$ s.t. $\sum_{k=0}^{\infty} |\beta_k| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \beta_k \varepsilon_{t-k}, \qquad t \in \mathbb{Z}.$$

Theorem 3.1.5 (Stationarity condition for autoregressive processes, [3])**.**

1. *For* $p \geq 0$ *there exists a causal stationary stochastic process satisfying (3.1.1) if and only if the generating polynomial* $A(z) := 1 - \alpha_1 z - ... - \alpha_p z^p,\ z \in \mathbb{C}$ *has no zeros in* $\{z \in \mathbb{C}, |z| \leq 1\}$.

2. *If the stationarity condition of 1. is satisfied, then every solution of (3.1.1) with arbitrary initial conditions:*

$$X_0 = x_0, ..., X_{-(p-1)} = x_{-(p-1)}$$

will be asymptotically stationary (with exponential rate).

Remark 3.1.3. (Algorithm for simulating a stationary $AR(p)$-process)

1. Fix $X_1, ..., X_p = 0$, generate i.i.d. $\varepsilon_1, \varepsilon_2, ...$ such that $\mathbb{E}\varepsilon_t = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$, for example $\mathcal{L}(\varepsilon_t) = \mathcal{N}(0, \sigma_\varepsilon^2)$.

2. Calculate recursively $X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \varepsilon_t$ for $t = p+1, ..., N+M$.

3. Throw away $X_1, ..., X_M$ and set

$$X_t^\star = X_{t+M} \text{ for } t = 1, ..., N$$

$X_1^\star, ..., X_N^\star$ is practically a realization of a stationary $AR(p)$-process.

3.2 Setup of the Problem

Assume that data $Y_1, ..., Y_N$ follow the deterministic equidistant design model (2.1.3):

$$\left.\begin{array}{ll} Y_j = m(x_j) + \varepsilon_j, & j = 1, ..., N \\ m : [0,1] \to \mathbb{R} & \\ x_j = \frac{j}{N}, & j = 1, ..., N \end{array}\right\}$$

where

Assumption E1: The errors $\varepsilon_j,\ j = 1, ..., N$ are part of a stationary time series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with $\mathbb{E}\left(\varepsilon_t\right) = 0$, $\mathrm{Var}\left(\varepsilon_t\right) = \sigma_\varepsilon^2 < \infty$ and autocovariances

$$r_t = \mathrm{Cov}\left(\varepsilon_s, \varepsilon_{t+s}\right), \qquad s, t \in \mathbb{Z}$$

i.e. the autocovariance sequence independent of N.

Let $u_j := m(x_j),\ j = 1, \ldots, N$ and $\mathbf{u} = (u_1, \ldots, u_N)^{\mathrm{T}}$. Let $\breve{u}_j$ denote the Nadaraya-Watson estimate of $m(x)$ at x_j, $j = 1, \ldots, N$ i.e.

$$\breve{u}_j := \frac{1}{N}\sum_{i=1}^{N} K_h\left(x_j - x_i\right) Y_i / p_K\left(x_j, h\right), \quad j = 1, \ldots, N$$

where K is kernel function on $\mathbb{R}$ and $p_K\left(x, h\right)$ given by:

$$p_K\left(x, h\right) = \frac{1}{N}\sum_{j=1}^{N} K_h\left(x - x_j\right), \quad x \in [0,1]$$

$\breve{\mathbf{u}} := (\breve{u}_1, \ldots, \breve{u}_N)$ as an estimate of the vector $\mathbf{u}$ can be interpreted as the solution of the following local least-squares problem:

$$\mathcal{D}\left(\mathbf{v}\right) := \sum_{i,j=1}^{N} \left(v_i - Y_j\right)^2 K_h\left(x_i - x_j\right) = \min! \tag{3.2.1}$$

where the minimization is w.r.t. $\mathbf{v} = (v_1, \ldots, v_N)^{\mathrm{T}}$. This can be shown from the fact that in the minimum

$$\frac{\partial}{\partial v_k}\sum_{i,j=1}^{N} \left(v_i - Y_j\right)^2 K_h\left(x_i - x_j\right) = 0, \quad k = 1, \ldots, N$$

have to be fulfilled.

To more tightly control the smoothness of the function estimate we will add a roughness penalty to the local least squares distance $\mathcal{D}(\mathbf{v})$ in (3.2.1) as the following definition.

Definition 3.2.1. Consider the deterministic equidistant design model (2.1.3). The regularized local least-squares estimate (RLLSE) $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_N)^{\mathrm{T}}$ of $\mathbf{u}$ is defined as the solution of the following regularized local least-squares problem:

$$Q(\mathbf{v}) := \sum_{i,j=1}^{N} (v_i - Y_j)^2 K_h(x_i - x_j) + \frac{\lambda}{2} \sum_{i,j=1}^{N} (v_i - v_j)^2 L_g(x_i - x_j) = \min_{\mathbf{v}}! \quad (3.2.2)$$

where $\lambda > 0$ a regularization parameter, K, L are kernel functions on $\mathbb{R}$ and K_h, L_g are the corresponding rescaled kernels w.r.t. the bandwidths $h > 0$, $g > 0$ respectively. The regularized local least-squares function estimate (RLLSFE) $\hat{m}(x, h, g, \lambda)$ of $m(x)$ can be defined as a function which interpolates the estimates $\hat{u}_1, \ldots, \hat{u}_N$ at $x_1, \ldots, x_N$:

$$\hat{m}(x_i, h, g, \lambda) = \hat{u}_i, \; i = 1, \ldots, N.$$

For convenience we define the regularized local least-squares function estimate as

$$\hat{m}(x, h, g, \lambda) = \hat{u}_i \text{ if } x_{i-1} < x \leq x_i, \; i = 1, \ldots, N$$

where $x_0 := 0$ and $\hat{m}(x_0, h, g, \lambda) = \hat{u}_1$.

The quantity $\mathcal{P}(\mathbf{v}) := \sum_{i,j=1}^{N} (v_i - v_j)^2 L_g(x_i - x_j)$ in the above definition is chosen as a particular roughness penalty. The tuning parameter λ determines the tradeoff between the two contradictory criteria. Putting λ very large puts high emphasis on minimizing $\mathcal{P}(\mathbf{v})$, forcing RLLSFE to become almost a straight line without regard to the local least squares distance $\mathcal{D}(\mathbf{v})$. Putting $\lambda = 0$ removes the influence of $\mathcal{P}(\mathbf{v})$

altogether and allows RLLSFE to become the Nadaraya-Watson function estimate. In general, the RLLSFE is "smoother" than the corresponding NW function estimate.

Let $p_\lambda(x,h,g) := p_K(x,h) + \lambda p_L(x,g)$, for $x \in [0,1]$ where $p_K(x,h)$ is defined as above and

$$p_L(x,g) = \frac{1}{N}\sum_{j=1}^{N} L_g(x - x_j), \quad x \in [0,1].$$

Let Λ denote the $N \times N$-matrix with entries $\Lambda_{ij} = \frac{1}{N} L_g(x_i - x_j)$ and P is $N \times N$-diagonal matrix with entries $P_{ii} = p_K(x_i,h) + \lambda p_L(x_i,g)$.

Suppose $\hat{m}_K(x,h)$, $\hat{m}_L(x,h)$ denote the Priestley-Chao estimates of $m(x)$ with the kernels K, L respectively. The following proposition explains that the regularized local least-squares estimate has an explicit representation in term of the Priestley-Chao estimate $\hat{\mathbf{u}}_{PC} := \left(\hat{u}_1^{PC}, \ldots, \hat{u}_N^{PC}\right)^{\mathrm{T}}$ of $\mathbf{u}$ where $\hat{u}_i^{PC} = \hat{m}_K(x_i,h)$, $i = 1, \ldots, N$.

Proposition 3.2.1 ([9])**.** *If the matrix $P - \lambda\Lambda$ is invertible, and the kernel L is symmetric about zero, then the regularized local least-squares estimate of $\mathbf{u}$ is given by*

$$\hat{\mathbf{u}} = (P - \lambda\Lambda)^{-1} \cdot \hat{\mathbf{u}}_{PC} \tag{3.2.3}$$

Proof. $\hat{\mathbf{u}} = \operatorname{argmin}_{\mathbf{v}} Q(\mathbf{v})$ has to fulfill

$$\begin{aligned} &\frac{\partial Q}{\partial \hat{u}_k}(\hat{\mathbf{u}}) = 0, \quad k = 1, \ldots, N \\ &\quad = 2\sum_{j=1}^{N} (\hat{u}_k - Y_j) K_h(x_k - x_j) + \lambda \sum_{j=1}^{N} (\hat{u}_k - \hat{u}_j) L_g(x_k - x_j) \\ &\qquad -\lambda \sum_{i=1}^{N} (\hat{u}_i - \hat{u}_k) L_g(x_i - x_k) \end{aligned}$$

This yields, by symmetry of the kernel L,

$$2\sum_{j=1}^{N} (\hat{u}_k - Y_j) K_h(x_k - x_j) + 2\lambda \sum_{j=1}^{N} (\hat{u}_k - \hat{u}_j) L_g(x_k - x_j) = 0$$

rearrange the formula

$$\hat{u}_k \left[\sum_{j=1}^{N} K_h \left(x_k - x_j\right) + \lambda \sum_{j=1}^{N} L_g \left(x_k - x_j\right) \right] - \lambda \sum_{j=1}^{N} \hat{u}_j L_g \left(x_k - x_j\right) = \sum_{j=1}^{N} K_h \left(x_k - x_j\right) Y_j$$

dividing by N and using the definition of p_K, p_L, and $\hat{m}_K$

$$\hat{u}_k \left[p_K \left(x_k, h\right) + \lambda p_L \left(x_k, g\right) \right] - \lambda \sum_{j=1}^{N} \hat{u}_j \Lambda_{k,j} = \hat{m}_K \left(x_k, h\right)$$

yielding

$$\hat{u}_k p_\lambda \left(x_k, h, g\right) - \lambda \sum_{j=1}^{N} \hat{u}_j \Lambda_{k,j} = \hat{u}_i^{PC}, \quad k = 1, \ldots, N$$

This implies

$$P\hat{\mathbf{u}} - \lambda \Lambda \hat{\mathbf{u}} = \hat{\mathbf{u}}_{PC}$$

$$\left(P - \lambda \Lambda\right) \hat{\mathbf{u}} = \hat{\mathbf{u}}_{PC}$$

and the assertion follows. □

Franke et al. [9] has investigated the asymptotic behavior of the mse of RLLSE where the errors were i.i.d. In the following we explore the asymptotic behavior of the mse of RLLSE where the errors $\varepsilon_1, \ldots, \varepsilon_N$ are correlated.

3.3 Mean squared error properties of RLLSE

For convenience, here we present some assumptions and notions concerning the kernel K, and L. The assumptions will be implicit throughout the reminder of this thesis.

Assumption K1: K is nonnegative and symmetric about zero.

Assumption K2: K has compact support $[-1, 1]$.

Assumption K3: K is Lipschitz continuous on $[-1, 1]$ with constant C_K.

The same assumptions L1-L3 regarding the kernel L are also assumed.

Usually, we need also the following additional assumptions concerning the kernel K:

Assumption K4: $K(\pm 1) = 0$.

Assumption K5: $K \in C^2(-1,1)$ with bounded second derivative K''.

Similar assumptions L4-L5 regarding the kernel L are also usually needed.

We could relax the assumptions of symmetry and compactness of the support of K, and of L, but we want to keep the arguments simple in this thesis. Due to the same reason, we mainly neglect boundary effects, which could be dealt with as in section 4.4 of Härdle [1], by restricting the attention to $x \in [\delta, 1-\delta]$ for some $\delta > 0$. Asymptotically, this will have no effect as we shall have $\delta \to 0$ for $N \to \infty$ anyhow. We also define two functionals of K:

1. $Q_K = \int\limits_{-1}^{1} K^2\,(u)du$
2. $V_K = \int\limits_{-1}^{1} u^2 K\,(u)du$

Regarding the regression function m we will assume:

Assumption M1: m is twice continuously differentiable on $[0,1]$.

Assumption M2: m'' is Hölder continuous on $[0,1]$ with constant H and exponent $0 < \beta < 1$.

It is necessary to make an assumption about the autocovariance sequence $\{r_t\}_{t\in\mathbb{Z}}$ of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, to ensure that summations over the correlation in the data remain bounded:

Assumption E2: $\sum\limits_{t=1}^{\infty} t\,|r_t| < \infty$.

This assumption, common in time series analysis, ensures that observations sufficiently far apart are essentially uncorrelated.

Under the assumption E2, the sum of the covariances, $\sum\limits_{t=1}^{\infty} r_t$, is well defined, since

$\sum_{t=1}^{\infty} |r_t| \leq \sum_{t=1}^{\infty} t\,|r_t| < \infty.$

In the following, let $f(\omega)$ denote the spectral density of the time series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ which exists as a consequence of assumption E2 (by theorem (3.1.4)). In particular, we have

$$f(0) = \sum_{t=-\infty}^{\infty} r_t = r_0 + 2\sum_{t=1}^{\infty} r_t = \sigma_\varepsilon^2 + 2R$$

where $R := \sum_{t=1}^{\infty} r_t$.

We state the following lemmas which will turn out to be useful later on.

Lemma 3.3.1. *Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a stationary time series. Under the Assumptions E1,E2, there is some constant $C < \infty$ independent of N such that:*

$$\sum_{i=1}^{N} \sum_{j\notin\{1,\ldots,N\}} |r_{i-j}| \leq C$$

i.e. $\sum_{i=1}^{N} \sum_{j\notin\{1,\ldots,N\}} |r_{i-j}| = O(1)$ *as* $N \to \infty$.

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{j\notin\{1,\ldots,N\}} |r_{i-j}| &= \sum_{i=1}^{N} \left(\sum_{j\leq 0} |r_{i-j}| + \sum_{j>N} |r_{i-j}| \right) \\
&= \sum_{i=1}^{N} \left(\sum_{k\geq i} |r_k| + \sum_{k<i-N} |r_k| \right) \\
&= \sum_{i=1}^{N} \left(\sum_{k=i}^{N} |r_k| + \sum_{k=N-i+1}^{N} |r_k| + 2\sum_{k>N} |r_k| \right) \\
&= \sum_{t=1}^{N} t\,|r_t| + \sum_{t=1}^{N} t\,|r_t| + 2N\sum_{t>N} |r_t| \\
&= 2\sum_{t=1}^{N} t\,|r_t| + 2\sum_{t>N} N\,|r_t| \\
&< 2\sum_{t=1}^{N} t\,|r_t| + 2\sum_{t>N} t\,|r_t| = 2\sum_{t=1}^{\infty} t\,|r_t| =: C < \infty.
\end{aligned}$$

□

Lemma 3.3.2. *Let $g \in C^2(a,b)$. Then for $x_i' = a + i(b-a)/M$, $i = 0, \ldots, M$,*

$$\left| \frac{1}{b-a} \int_a^b g(y)dy - \frac{1}{M} \sum_{i=0}^{M} g(x_i') \right| \leq \frac{(b-a)^2}{12M^2} \sup_{a<z<b} |g''(z)| + \frac{1}{2M} |g(b) + g(a)|$$

Proof. Let $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$ be the first two Bernoulli polynomials. Using integration by parts twice, we get for any function $f \in C^2(a,b)$

$$\begin{aligned}
\int_0^1 f(y)dy &= B_1(y)f(y)|_0^1 - \int_0^1 B_1(y)f'(y)dy \\
&= B_1(y)f(y)|_0^1 - \frac{1}{2} B_2(y)f'(y)|_0^1 + \frac{1}{2} \int_0^1 B_2(y)f''(y)dy \\
&= \frac{1}{2}[f(1) + f(0)] - \frac{1}{12}[f'(1) - f'(0)] + \frac{1}{2} \int_0^1 B_2(y)f''(y)dy \\
&= \frac{1}{2}[f(1) + f(0)] + \frac{1}{2} \int_0^1 \left(y^2 - y\right) f''(y)dy \\
&= \frac{1}{2}[f(1) + f(0)] - \frac{1}{2} f''(\xi) \int_0^1 \left(y - y^2\right) dy \\
&= \frac{1}{2}[f(1) + f(0)] - \frac{1}{12} f''(\xi) \qquad (*)
\end{aligned}$$

for some $\xi \in (0,1)$, using $x - x^2 \geq 0$ in $[0,1]$ and the mean-value theorem of integration.

Now, let $f_i(y) = g\left(a + \frac{b-a}{M}(y+i-1)\right)$, $i = 0, \ldots, M$, then we have

$$\begin{aligned}
\frac{1}{b-a}\int_a^b g(y)dy &= \frac{1}{M}\int_0^M g(a+\frac{b-a}{M}y)dy = \frac{1}{M}\sum_{i=1}^{M}\int_{i-1}^{i} g(a+\frac{b-a}{M}y)dy \\
&= \frac{1}{M}\sum_{i=1}^{M}\int_0^1 g\left(a+\frac{b-a}{M}(y+i-1)\right)dy = \frac{1}{M}\sum_{i=1}^{M}\int_0^1 f_i(y)dy \\
&\overset{(*)}{=} \frac{1}{2M}\sum_{i=1}^{M}[f_i(1)+f_i(0)] - \frac{1}{12M}\sum_{i=1}^{M} f_i''(\xi_i) \\
&= \frac{1}{2M}\sum_{i=1}^{M}[f_i(1)+f_{i-1}(1)] - \frac{1}{12M}\sum_{i=1}^{M} f_i''(\xi_i) \\
&= \frac{1}{M}\sum_{i=0}^{M} f_i(1) - \frac{1}{2M}[f_M(1)+f_0(1)] - \frac{1}{12M}\sum_{i=1}^{M} f_i''(\xi_i) \\
&= \frac{1}{M}\sum_{i=0}^{M} g(x_i') - \frac{1}{2M}[g(b)+g(a)] - \frac{(b-a)^2}{12M^3}\sum_{i=1}^{M} g''(\xi_i')
\end{aligned}$$

for some $\xi_i' = a + \frac{b-a}{M}(\xi_i + i - 1) \in (a,b)$, $\xi_i \in (0,1)$, $i = 1, \ldots, M$.

This yields

$$\left|\frac{1}{b-a}\int_a^b g(y)dy - \frac{1}{M}\sum_{i=1}^{M} g(x_i')\right| \le \frac{1}{2M}|g(b)+g(a)| + \frac{(b-a)^2}{12M^3}\sum_{i=1}^{M}|g''(\xi_i')|$$

and the assertion follows. □

Lemma 3.3.3. *Let g be a continuous function on $[a,b]$, twice continuously differentiable on (a,b), and vanishing outside (a,b) where $0 \le a < b \le 1$. Then we have*

$$\left|\int_0^1 g(y)\,dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j)\right| \le \frac{b-a}{12N^2}\sup_{z\in(a,b)}|g''(z)| + \frac{3}{N^2}\sup_{z\in(a,b)}|g'(z)|$$

where $x_j = \frac{j}{N}$, $j = 1, ..., N$.

Proof. We set $n_0 := \min\{n; x_n \ge a\}$, and define $n_k := n_0 + k$, for $k = 1, 2, \ldots, M$ where $M := \max\{k; x_{n_k} \le b\}$. Then we have $n_M = \max\{n; x_n \le b\}$, and $n_M - n_0 =$

$M \leq N(b-a)$. We apply Lemma 3.3.2 to $g(y)$ such that $x'_i = x_{n_0} + i(x_{n_M} - x_{n_0})/M = x_{n_0} + \frac{i}{N} = x_{n_0} + x_i = x_{n_0+i}$, $i = 0, \ldots, M$, and get

$$\left| \frac{1}{x_{n_M} - x_{n_0}} \int_{x_{n_0}}^{x_{n_M}} g(y)dy - \frac{1}{M} \sum_{i=0}^{M} g(x_{n_i}) \right| \leq \frac{(x_{n_M} - x_{n_0})^2}{12M^2} \sup_{x_{n_0} < z < x_{n_M}} |g''(z)| + \frac{1}{2M} |g(x_{n_0}) + g(x_{n_M})|$$

This gives

$$\left| \int_{x_{n_0}}^{x_{n_M}} g(y)dy - \frac{1}{N} \sum_{j=1}^{N} g(x_j) \right| \leq \frac{x_{n_M} - x_{n_0}}{12N^2} \sup_{a<z<b} |g''(z)| + \frac{1}{2N} |g(x_{n_0}) + g(x_{n_M})|$$

and

$$\left| \int_{a}^{b} g(y)dy - \frac{1}{N} \sum_{j=1}^{N} g(x_j) \right| \leq \frac{b-a}{12N^2} \sup_{a<z<b} |g''(z)| + \frac{1}{2N} |g(x_{n_0})| + \frac{1}{2N} |g(x_{n_M})| + \left| \int_{a}^{x_{n_0}} g(y)dy \right| + \left| \int_{x_{n_M}}^{b} g(y)dy \right| \qquad (3.3.1)$$

By the mean-value theorem and as $g(a) = 0$, $x_{n_0} - a \leq \frac{1}{N}$ we have for some $\eta \in [a, x_{n_0}]$

$$|g(x_{n_0})| = |g(x_{n_0}) - g(a)| = |x_{n_0} - a| \, |g'(\eta)| \leq \frac{1}{N} \sup_{a<z<b} |g'(z)|$$

and analogously, $|g(x_{n_M})| \leq \frac{1}{N} \sup_{a<z<b} |g'(z)|$.

By the mean-value theorem of integration and as $x_{n_0} - a \leq \frac{1}{N}$, $g(a) = 0$, we get for some $\xi \in [a, x_{n_0}]$, $\xi' \in [a, \xi]$

$$\begin{aligned} \left| \int_{a}^{x_{n_0}} g(y)dy \right| &= |x_{n_0} - a| \, |g(\xi)| \\ &\leq \frac{1}{N} |g(\xi) - g(a)| \\ &= \frac{1}{N} |\xi - a| \, |g'(\xi')| \\ &\leq \frac{1}{N^2} \sup_{a<z<b} |g'(z)| \end{aligned}$$

and analogously, $\left|\int_{x_{n_M}}^{b} g(y)dy\right| \leq \frac{1}{N^2}\sup_{a<z<b}|g'(z)|$. Therefore, we get from (3.3.1)

$$\left|\int_a^b g(y)dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j)\right| \leq \frac{b-a}{12N^2}\sup_{a<z<b}|g''(z)| + \frac{2}{2N^2}\sup_{a<z<b}|g'(z)| + \frac{2}{N^2}\sup_{a<z<b}|g'(z)| \tag{3.3.2}$$

and the assertion follows. □

For analyzing the mse of RLLS-estimate, we repeatedly consider

$$p_K(x,h) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)$$

Corollary 3.3.4. *Assuming K1-K5 for a kernel K, we have,*

$$|p_K(x,h) - 1| = O\left(\frac{1}{N^2h^2}\right) \text{ uniformly for all } x \in [h, 1-h]$$

as $N \to \infty$ s.t. $h \to 0$, and $Nh \to \infty$.

Proof. The assertion follows immediately by Lemma 3.3.3 with $g(y) = K_h(x-y)$, $a = x - h$, and $b = x + h$. □

Lemma 3.3.5. *Assuming K1-K5 for a kernel K, and M1-M2 for a function $m : [0,1] \to \mathbb{R}$, we have, as $N \to \infty$ s.t. $h \to 0$, $Nh \to \infty$,*

i) $\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, m(x_j) = \int_{-1}^{1} K(u)\, m(x - hu)\, du + O\left(\frac{1}{N^2h^2}\right)$ *uniformly for all $x \in [h, 1-h]$.*

ii) $\frac{1}{N}\sum_{j=1}^{N} K_h^2(x-x_j) = \frac{1}{h}\int_{-1}^{1} K^2(u)\, du + O\left(\frac{1}{N^2h^3}\right)$ *uniformly for all $x \in [h, 1-h]$.*

iii) $\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, K_h(x'-x_j) = \frac{1}{h}\int_{-1}^{1} K(u)\, K\left(u - \frac{x-x'}{h}\right) du + O\left(\frac{1}{N^2h^3}\right)$ *uniformly for all $x, x' \in [h, 1-h]$ with $|x - x'| \leq 2h$,*

and $\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, K_h(x'-x_j) = 0$ if $|x-x'| > 2h$.

Proof. i), ii) follow using Lemma 3.3.3 with $g(y) = K_h(x-y)\, m(y)$ for i) and $g(y) = K_h^2(x-y)$ for ii).

iii) Using Lemma 3.3.3 with $a = \max(x,x') - h$, $b = \min(x,x') + h$ and $g(y) = K_h(x-y)K_h(x'-y)$ we have for some constants B', B'',

$$\begin{aligned} \left| \int_0^1 g(y)\, dy - \frac{1}{N}\sum_{j=1}^{N} g(x_j) \right| &\leq \frac{2h - |x-x'|}{12N^2} \sup_{z\in(a,b)} |g''(z)| + \frac{3}{N^2} \sup_{z\in(a,b)} |g'(z)| \\ &\leq \frac{2h}{12N^2}\frac{B''}{h^4} + \frac{3}{N^2}\frac{B'}{h^3} = \frac{B''+18B'}{6}\frac{1}{N^2h^3} \end{aligned}$$

That gives

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, K_h(x'-x_j) &= \int_0^1 K_h(x-y)\, K_h(x'-y)\, dy + O\left(\frac{1}{N^2h^3}\right) \\ &= \frac{1}{h}\int_{-1}^{1} K(u)\, K\left(u - \frac{x-x'}{h}\right) du + O\left(\frac{1}{N^2h^3}\right) \end{aligned}$$

Now, if $|x-x'| > 2h$ then for all $j \in \{1,\ldots,N\}$ we have either $|x-x_j| > h$ or $|x_j - x'| > h$. Therefore the compactness of the support of K implies

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\, K_h(x'-x_j) = 0 \text{ if } |x-x'| > 2h.$$

□

The following proposition presents the asymptotic expansions for bias and variance of the PC-estimate when the errors are correlated. Note the important fact that the bias does not depend on the dependency structure of the errors, therefore results and proofs will be the same for independent and correlated case. We also add a more details statement about the asymptotic covariance of PC-estimates at different locations. A similar result has been shown by Halim [11] in the two-dimensional case where data are available on an equidistant grid in the plane. We give a proof never the less as we need detailed orders for the error terms.

Proposition 3.3.6. *Assume the deterministic equidistant design model (2.1.3), E1-E2, K1-K5, M1-M2, $\hat{m}_K(x,h)$ is PC-estimate of $m(x)$ and $N \to \infty$ s.t. $h \to 0$, and $Nh^4 \to \infty$, then uniformly for all $x, x' \in [h, 1-h]$ we have:*

(i) $\text{bias}\left(\hat{m}_K(x,h)\right) = \mathbb{E}\left(\hat{m}(x,h)\right) - m(x) = \frac{h^2}{2} m''(x) V_K + O\left(h^{2+\beta}\right)$.

(ii) $\text{Var}\left(\hat{m}_K(x,h)\right) = \frac{f(0)}{Nh} Q_K + O\left(\frac{1}{N^2h^2}\right)$ *where $f(\cdot)$ is the spectral density of the noise-process.*

(iii) $\text{mse}\left(\hat{m}_K(x,h)\right) = \frac{h^4}{4}\left(m''(x)\right)^2 V_K^2 + \frac{f(0)}{Nh} Q_K + O\left(h^{4+\beta}\right)$.

(iv) $\text{Cov}\left(\hat{m}_K(x,h), \hat{m}_K(x',h)\right) = \frac{f(0)}{Nh}(K*K)\left(\frac{x-x'}{h}\right) + O\left(\frac{1}{N^2h^2}\right)$.
In particular, if $|x-x'| > 2h$ we have $\text{Cov}\left(\hat{m}(x,h), \hat{m}(x',h)\right) = O\left(\frac{1}{N^2h^2}\right)$.

Proof. The assertion i) was proved in Theorem 2.3.1 for i.i.d. errors and it still holds for correlated errors but the remainder term here was modified to just $O\left(h^{2+\beta}\right)$ because we use here Lemma 3.3.5 instead of Lemma 2.3.3.

ii) We have:

$$
\begin{aligned}
\text{Var}\left(\hat{m}(x,h)\right) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_h(x-x_i) K_h(x-x_j) \underbrace{\text{Cov}(Y_i,Y_j)}_{=\text{Cov}(\varepsilon_i,\varepsilon_j)=r_{i-j}} \\
&= \underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_h(x-x_i)\left[K_h(x-x_j) - K_h(x-x_i)\right] r_{i-j}}_{=:V_1} \\
&\quad + \underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_h^2(x-x_i)\, r_{i-j}}_{=:V_2}
\end{aligned}
$$

For V_2 we have

$$
V_2 = \frac{1}{N^2}\sum_{i=1}^{N} K_h^2(x-x_i) \sum_{j=1}^{N} r_{i-j}
$$

but $\sum_{j=1}^{N} r_{i-j} = \sum_{j=-\infty}^{\infty} r_{i-j} - \sum_{j\notin\{1,\ldots,N\}} r_{i-j} = f(0) - \sum_{j\notin\{1,\ldots,N\}} r_{i-j}$ implies

$$V_2 = \underbrace{\frac{1}{N^2} f(0) \sum_{i=1}^{N} K_h^2 (x - x_i)}_{=:W_1} - \underbrace{\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j\notin\{1,\ldots,N\}} K_h^2 (x - x_i)\, r_{i-j}}_{=:W_2}$$

Note that $W_2 = O\left(\frac{1}{N^2 h^2}\right)$ since

$$\begin{aligned} |W_2| &\leq \frac{B_K}{N^2 h^2} \sum_{i=1}^{N} \sum_{j\notin\{1,\ldots,N\}} |r_{i-j}| \text{ where } B_K := \sup_u K^2(u) \\ &\overset{\text{Lemma } 3.3.1}{\leq} \frac{B_K C}{N^2 h^2} \end{aligned}$$

and using lemma (3.3.5) it can be shown

$$\begin{aligned} V_2 &= \frac{f(0)}{N} \left[\frac{1}{h} \int_{-1}^{1} K^2(u)\, du + O\left(\frac{1}{N^2 h^3}\right) \right] + O\left(\frac{1}{N^2 h^2}\right) \\ &= \frac{f(0)}{Nh} Q_K + O\left(\frac{1}{N^2 h^2}\right) \end{aligned}$$

Now, using that K is Lipschitz continuous with constant C_K, we have

$$\begin{aligned} |V_1| &\leq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_h(x - x_i) \left| K_h(x - x_j) - K_h(x - x_i) \right| |r_{i-j}| \\ &\leq \frac{C_K}{N^2 h^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_h(x - x_i) |x_i - x_j| \, |r_{i-j}| \\ &= \frac{C_K}{N^3 h^2} \sum_{i=1}^{N} K_h(x - x_i) \underbrace{\sum_{j=1}^{N} |i - j| \, |r_{i-j}|}_{\leq \sum_{t=-\infty}^{\infty} |t||r_t| = 2 \sum_{t=1}^{\infty} t|r_t| =: C < \infty} \\ &\leq \frac{C_K C}{N^2 h^2} p_K(x, h) \\ &\overset{\text{Corollary } 3.3.4}{=} O\left(\frac{1}{N^2 h^2}\right) \end{aligned}$$

This implies ii).

iii) Combining the bias and variance expansion, we get

$$\begin{aligned}\operatorname{mse}\left(\hat{m}\left(x,h\right)\right) &= \operatorname{Var}\left(\hat{m}\left(x,h\right)\right)+\left(\operatorname{bias}\left(\hat{m}\left(x,h\right)\right)\right)^2 \\ &= \frac{f\left(0\right)}{Nh}Q_K+\frac{h^4}{4}\left(m''\left(x\right)\right)^2V_K^2+O\left(h^{4+2\beta}\right)+m''\left(x\right)O\left(h^{4+\beta}\right) \\ &\quad +O\left(\frac{1}{N^2h^2}\right)\end{aligned}$$

As m'' is bounded, the assertion of iii) holds uniformly for all $x \in [h, 1-h]$.

iv)We have:

$$\begin{aligned}\operatorname{Cov}\left(\hat{m}\left(x,h\right),\hat{m}\left(x',h\right)\right) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_h\left(x-x_i\right)K_h\left(x'-x_j\right)\underbrace{\operatorname{Cov}\left(Y_i,Y_j\right)}_{=\operatorname{Cov}(\varepsilon_i,\varepsilon_j)=r_{i-j}} \\ &= \underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_h\left(x-x_i\right)\left[K_h\left(x'-x_j\right)-K_h\left(x'-x_i\right)\right]r_{i-j}}_{=:V_1} \\ &\quad +\underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_h\left(x-x_i\right)K_h\left(x'-x_i\right)r_{i-j}}_{=:V_2}\end{aligned}$$

For V_2 we have

$$V_2=\frac{1}{N^2}\sum_{i=1}^{N}K_h\left(x-x_i\right)K_h\left(x'-x_i\right)\sum_{j=1}^{N}r_{i-j}$$

but $\sum_{j=1}^{N}r_{i-j}=\sum_{j=-\infty}^{\infty}r_{i-j}-\sum_{j\notin\{1,\ldots,N\}}r_{i-j}=f\left(0\right)-\sum_{j\notin\{1,\ldots,N\}}r_{i-j}$ implies

$$V_2=\frac{1}{N^2}\sum_{i=1}^{N}K_h\left(x-x_i\right)K_h\left(x'-x_i\right)\left[f\left(0\right)-\sum_{j\notin\{1,\ldots,N\}}r_{i-j}\right]$$

$$V_2=\underbrace{\frac{1}{N^2}f\left(0\right)\sum_{i=1}^{N}K_h\left(x-x_i\right)K_h\left(x'-x_i\right)}_{=:W_1}-\underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\notin\{1,\ldots,N\}}K_h\left(x-x_i\right)K_h\left(x'-x_i\right)r_{i-j}}_{=:W_2}$$

Note that $W_2 = O\left(\frac{1}{N^2h^2}\right)$ since

$$\begin{aligned} |W_2| &\leq \frac{B_K^2}{N^2h^2}\sum_{i=1}^{N}\sum_{j\notin\{1,\dots,N\}}|r_{i-j}| \text{ where } B_K := \sup_u K(u) \\ &\overset{\text{Lemma } 3.3.1}{\leq} \frac{B_K^2 C}{N^2h^2} \end{aligned}$$

and using lemma (3.3.5) it can be shown

$$\begin{aligned} V_2 &= \frac{f(0)}{N}\left[\frac{1}{h}\int_{-1}^{1}K(u)\,K\left(u-\frac{x-x'}{h}\right)du + O\left(\frac{1}{N^2h^3}\right)\right] + O\left(\frac{1}{N^2h^2}\right) \\ &= \frac{f(0)}{Nh}K*K\left(\frac{x-x'}{h}\right) + O\left(\frac{1}{N^2h^2}\right) \end{aligned}$$

Now, using that K is Lipschitz continuous with constant C_K, we have

$$\begin{aligned} |V_1| &\leq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_h(x-x_i)\left|K_h(x'-x_j)-K_h(x'-x_i)\right||r_{i-j}| \\ &\leq \frac{C_K}{N^2h^2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_h(x-x_i)\,|x_i-x_j|\,|r_{i-j}| \\ &= \frac{C_K}{N^3h^2}\sum_{i=1}^{N}K_h(x-x_i)\underbrace{\sum_{j=1}^{N}|i-j|\,|r_{i-j}|}_{\leq \sum_{t=-\infty}^{\infty}|t||r_t| = 2\sum_{t=1}^{\infty}t|r_t| =: C < \infty} \\ &\leq \frac{C_K C}{N^2h^2}p_K(x,h) \\ &\overset{\text{Corollary } 3.3.4}{=} O\left(\frac{1}{N^2h^2}\right) \end{aligned}$$

This implies iv).

Lastly, if $|x-x'| > 2h$, then $V_2 = 0$ by the compactness of the support of K. Therefore $\text{Cov}(\hat{m}(x,h), \hat{m}(x',h)) = V_1 = O(\frac{1}{N^2h^2})$. □

For analyzing the mse of RLLSE we consider the crucial statistics

$$\begin{aligned}\hat{m}_n(x,h,g) &:= \frac{1}{N^n}\sum_{j_1,\ldots,j_n} L_g(x-x_{j_1})\,L_g(x_{j_1}-x_{j_2})\ldots L_g\left(x_{j_{n-1}}-x_{j_n}\right)\hat{m}_K(x_{j_n},h)\\ &= \frac{1}{N^{n+1}}\sum_{j_1,\ldots,j_n,k}\{L_g(x-x_{j_1})\,L_g(x_{j_1}-x_{j_2})\ldots L_g\left(x_{j_{n-1}}-x_{j_n}\right)\\ &\qquad\qquad \cdot K_h(x_{j_n}-x_k)\,Y_k\}\end{aligned}$$

for $n \geq 1$, $x \in [0,1]$.

This recursive sequence of explicitly defined estimates will be used to approximate the explicitly (as a solution of an extremum problem) defined RLLSE $\hat{m}(x; h, g, \lambda)$ later on. $\hat{m}_n$ is generated by repeatedly smoothing the PC-estimate $\hat{m}_K$ using the kernel L_g. Therefore, we could call them resmoothers. We need the following lemma which is a generalization of Lemma 3.3.3 to the multidimensional integrals.

Lemma 3.3.7. *Let D be an open convex set in $[0,1]^d$ and g be a continuous function on $\overline{D}$, twice continuously differentiable on D, and vanishing outside D. Then for each $\varepsilon > 1$ we have for N large enough,*

$$\left|\int\ldots\int_{[0,1]^d} g(y_1,\ldots,y_d)\,dy_1\ldots dy_d - \frac{1}{N^d}\sum_{i_1,\ldots,i_d=1}^{N} g(x_{i_1},\ldots,x_{i_d})\right|$$
$$\leq \frac{\varepsilon^d V}{12N^2}\sum_{i=1}^{d}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial^2 g}{\partial y_i^2}(y_1,\ldots,y_d)\right| + \frac{3\varepsilon^d}{N^2}\sum_{i=1}^{d} V_{y_i}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial g}{\partial y_i}(y_1,\ldots,y_d)\right|$$

where $x_j = \frac{j}{N}$; $j = 1,...,N$, $V = \int\ldots\int_D dy_1\ldots dy_d$,
$V_{y_i} = \int\ldots\int_{D_{y_i}} dy_1\ldots dy_{i-1}dy_{i+1}dy_d$, and
$D_{y_i} = \left\{(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_d)\in\mathbb{R}^{d-1}\,\middle|\,\exists\, y\in\mathbb{R} : (y_1,\ldots,y_{i-1},y,y_{i+1},\ldots,y_d)\in D\right\}$.

Proof. First, we remark that for d=1, the assertion follows immediately from Lemma 3.3.3. For proving the general case, we use induction and assume that the assertion holds for $d-1 \geq 1$. We have using Lemma 3.3.3

$$\left| \int \cdots \int_{[0,1]^d} g\left(y_1, \ldots, y_d\right) dy_1 \ldots dy_d - \frac{1}{N^d} \sum_{i_1, \ldots, i_d = 1}^{d} g(x_{i_1}, \ldots, x_{i_d}) \right|$$

$$= \left| \int \cdots \int_{[0,1]^d} g\left(y_1, \ldots, y_d\right) dy_1 \ldots dy_d - \frac{1}{N} \int \cdots \int_{[0,1]^{d-1}} \sum_{i_1=1}^{N} g\left(x_{i_1}, y_2, \ldots, y_d\right) dy_2 \ldots dy_d \right.$$

$$\left. + \frac{1}{N} \sum_{i_1=1}^{N} \int \cdots \int_{[0,1]^{d-1}} g\left(x_{i_1}, y_2, \ldots, y_d\right) dy_2 \ldots dy_d - \frac{1}{N^d} \sum_{i_1, \ldots, i_d = 1}^{d} g(x_{i_1}, \ldots, x_{i_d}) \right|$$

$$\leq \int \cdots \int_{[0,1]^{d-1}} \left| \int_0^1 g\left(y_1, \ldots, y_d\right) dy_1 - \frac{1}{N} \sum_{i_1=1}^{N} g\left(x_{i_1}, y_2, \ldots, y_d\right) \right| dy_2 \ldots dy_d$$

$$+ \frac{1}{N} \sum_{i_1=1}^{N} \left| \int \cdots \int_{[0,1]^{d-1}} g\left(x_{i_1}, y_2, \ldots, y_d\right) dy_2 \ldots dy_d - \frac{1}{N^{d-1}} \sum_{i_2, \ldots, i_d = 1}^{d} g(x_{i_1}, \ldots, x_{i_d}) \right|$$

$$\leq \int \cdots \int_{[0,1]^{d-1}} \left[\frac{b\left(y_2, \ldots, y_d\right) - a\left(y_2, \ldots, y_d\right)}{12N^2} \sup_{y_1 \in (a(y_2, \ldots, y_d), b(y_2, \ldots, y_d))} \left| \frac{\partial^2 g}{\partial y_1^2}\left(y_1, y_2, \ldots, y_d\right) \right| \right.$$

$$\left. + \frac{3}{N^2} \sup_{y_1 \in (a(y_2, \ldots, y_d), b(y_2, \ldots, y_d))} \left| \frac{\partial g}{\partial y_1}\left(y_1, y_2, \ldots, y_d\right) \right| \right] dy_2 \ldots dy_d$$

$$+ \frac{1}{N} \sum_{i_1=1}^{N} \left[\frac{\varepsilon^{d-1} V\left(x_{i_1}\right)}{12N^2} \sum_{i=2}^{d} \sup_{(y_2, \ldots, y_d) \in D(x_{i_1})} \left| \frac{\partial^2 g}{\partial y_i^2}\left(x_{i_1}, y_2, \ldots, y_d\right) \right| \right.$$

$$\left. + \frac{3\varepsilon^{d-1}}{N^2} \sum_{i=2}^{d} V_{y_i}\left(x_{i_1}\right) \sup_{(y_2, \ldots, y_d) \in D(x_{i_1})} \left| \frac{\partial g}{\partial y_i}\left(x_{i_1}, y_2, \ldots, y_d\right) \right| \right]$$

$$\leq \frac{1}{12N^2} \sup_{(y_1, \ldots, y_d) \in D} \left| \frac{\partial^2 g}{\partial y_1^2}\left(y_1, \ldots, y_d\right) \right| \int \cdots \int_{[0,1]^{d-1}} \left[b\left(y_2, \ldots, y_d\right) - a\left(y_2, \ldots, y_d\right) \right] dy_2 \ldots dy_d$$

$$+ \frac{3}{N^2} \int \cdots \int_{D_{y_1}} \sup_{y_1 \in (a(y_2, \ldots, y_d), b(y_2, \ldots, y_d))} \left| \frac{\partial g}{\partial y_1}\left(y_1, y_2, \ldots, y_d\right) \right| dy_2 \ldots dy_d$$

$$+ \frac{\varepsilon^{d-1}}{12N^2} \frac{1}{N} \sum_{i_1=1}^{N} V\left(x_{i_1}\right) \sum_{i=2}^{d} \sup_{(y_1, \ldots, y_d) \in D} \left| \frac{\partial^2 g}{\partial y_i^2}\left(y_1, \ldots, y_d\right) \right|$$

$$+ \frac{3\varepsilon^{d-1}}{N^3} \sum_{i_1=1}^{N} \sum_{i=2}^{d} V_{y_i}\left(x_{i_1}\right) \sup_{(y_1, \ldots, y_d) \in D} \left| \frac{\partial^2 g}{\partial y_i^2}\left(y_1, \ldots, y_d\right) \right|$$

where $a(y_2,\ldots,y_d)=\min\{y_1\in\mathbb{R}|(y_1,\ldots,y_d)\in D\}$,

$b(y_2,\ldots,y_d)=\max\{y_1\in\mathbb{R}|(y_1,\ldots,y_d)\in D\}$,

$D(x_{i_1})=\left\{(y_2,y_3,\ldots,y_d)\in\mathbb{R}^{d-1}\middle|\,(x_{i_1},y_2,\ldots,y_d)\in D\right\}$,

$V(x_{i_1})=\int\ldots\int_{D(x_{i_1})}dy_2\ldots dy_d$,

$D_{y_i}(x_{i_1})\;=\;\left\{(y_2,\ldots,y_{i-1},y_{i+1},\ldots,y_d)\in\mathbb{R}^{d-2}\middle|\,(x_{i_1},y_2,\ldots,y_{i-1},y_{i+1},\ldots,y_d)\in D_{y_i}\right\}$,

and

$D_{y_i}(x_{i_1})=\left\{(y_2,\ldots,y_{i-1},y_{i+1},\ldots,y_d)\in\mathbb{R}^{d-2}\middle|\,\exists\, y\in\mathbb{R}:(y_2,\ldots,y_{i-1},y,y_{i+1},\ldots,y_d)\in D(x_{i_1})\right\}$,

$D_{y_i}(x_{i_1})=\left\{(y_2,\ldots,y_{i-1},y_{i+1},\ldots,y_d)\in\mathbb{R}^{d-2}\middle|\,\exists\, y\in\mathbb{R}:(x_{i_1},y_2,\ldots,y_{i-1},y,y_{i+1},\ldots,y_d)\in D\right\}$,

$V_{y_i}(x_{i_1})=\int\ldots\int_{D_{y_i}(x_{i_1})}dy_2\ldots dy_{i-1}dy_{i+1}\ldots dy_d$.

Now as $\lim_{N\to\infty}\frac{1}{N}\sum_{i_1=1}^{N}V(x_{i_1})=V$ and $\lim_{N\to\infty}\frac{1}{N}\sum_{i_1=1}^{N}V_{y_i}(x_{i_1})=V_{y_i}$; $i=2,\ldots,d$, we get for N large enough $\frac{1}{N}\sum_{i_1=1}^{N}V(x_{i_1})<\varepsilon V$, $\frac{1}{N}\sum_{i_1=1}^{N}V_{y_i}(x_{i_1})<\varepsilon V_{y_i}$; $i=2,\ldots,d$, and

$$
\begin{aligned}
&\left|\int_{[0,1]^d}\!\!\ldots\int g(y_1,\ldots,y_d)\,dy_1\ldots dy_d-\frac{1}{N^d}\sum_{i_1,\ldots,i_d=1}^{d}g(x_{i_1},\ldots,x_{i_d})\right|\\
&\quad\le\frac{V}{12N^2}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial^2 g}{\partial y_1^2}(y_1,\ldots,y_d)\right|+\frac{3}{N^2}V_{y_1}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial g}{\partial y_1}(y_1,\ldots,y_d)\right|\\
&\qquad+\frac{\varepsilon^{d-1}\varepsilon V}{12N^2}\sum_{i=2}^{d}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial^2 g}{\partial y_i^2}(y_1,\ldots,y_d)\right|\\
&\qquad+\frac{3\varepsilon^{d-1}}{N^2}\sum_{i=2}^{d}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial^2 g}{\partial y_i^2}(y_1,\ldots,y_d)\right|\frac{1}{N}\sum_{i_1=1}^{N}V_{y_i}(x_{i_1})\\
&\quad\le\frac{\varepsilon^d V}{12N^2}\sum_{i=1}^{d}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial^2 g}{\partial y_i^2}(y_1,\ldots,y_d)\right|+\frac{3\varepsilon^d}{N^2}\sum_{i=1}^{d}V_{y_i}\sup_{(y_1,\ldots,y_d)\in D}\left|\frac{\partial g}{\partial y_i}(y_1,\ldots,y_d)\right|
\end{aligned}
$$

□

For convenience we will use the following notation:

$\mathcal{L}_g(\xi_1,\xi_2,\ldots,\xi_n):=L_g(\xi_1-\xi_2)L_g(\xi_2-\xi_3)\ldots L_g(\xi_{n-1}-\xi_n)$ for $\xi_1,\xi_2,\ldots,\xi_n\in[0,1]$

and $n \geq 2$.

Lemma 3.3.8. *Assuming $L1 - L5$ for a kernel L, we have for $n \geq 1$,*

$$\frac{1}{N^n} \sum_{j_1, j_2, \ldots, j_n = 1}^{N} \mathcal{L}_g\left(x, x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right) = 1 + O\left(\frac{1}{N^2 g^2}\right) \text{ uniformly for all } x \in [ng, 1-ng]$$

as $N \to \infty$ s.t. $g \to 0$, and $Ng \to \infty$.

Proof. First, we remark that for $n = 1$, the assertion has already been shown in Corollary 3.3.4. For proving the general case, we use the induction and assume that the assertion holds for $n \geq 1$. We have uniformly for all $x \in [(n+1)g, 1-(n+1)g]$,

$$\begin{aligned}
&\frac{1}{N^{n+1}} \sum_{j_1, j_2, \ldots, j_{n+1} = 1}^{N} \mathcal{L}_g\left(x, x_{j_1}, x_{j_2}, \ldots, x_{j_{n+1}}\right) \\
&= \frac{1}{N^n} \sum_{\substack{j_1 : \left|x - x_{j_1}\right| \leq g \\ j_2 : \left|x_{j_1} - x_{j_2}\right| \leq g \\ \vdots \\ j_n : \left|x_{j_{n-1}} - x_{j_n}\right| \leq g}} \mathcal{L}_g\left(x, x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right) \frac{1}{N} \sum_{j_{n+1} = 1}^{N} L_g\left(x_{j_n} - x_{j_{n+1}}\right) \\
&\overset{\text{Cor. 3.3.4}}{=} \frac{1}{N^n} \sum_{\substack{j_1 : \left|x - x_{j_1}\right| \leq g \\ j_2 : \left|x_{j_1} - x_{j_2}\right| \leq g \\ \vdots \\ j_n : \left|x_{j_{n-1}} - x_{j_n}\right| \leq g}} \mathcal{L}_g\left(x, x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right) \left[1 + O\left(\frac{1}{N^2 g^2}\right)\right] \\
&= \left[1 + O\left(\frac{1}{N^2 g^2}\right)\right] \frac{1}{N^n} \sum_{j_1, j_2, \ldots, j_n} \mathcal{L}_g\left(x, x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right) \\
&= \left[1 + O\left(\frac{1}{N^2 g^2}\right)\right] \left[1 + O\left(\frac{1}{N^2 g^2}\right)\right] \\
&= 1 + O\left(\frac{1}{N^2 g^2}\right)
\end{aligned}$$

□

Proposition 3.3.9. *Assume the deterministic equidistant design model (2.1.3), E1-E2, K1-K5, L1-L5, M1-M2, and $N \to \infty$, s.t. $h \to 0$, and $g \to 0$. Then for all $p \geq 1$ we have, uniformly for all $x \in [pg + h, 1 - (pg + h)]$,*

$$\mathbb{E}\left(\hat{m}_p\left(x, h, g\right)\right) = m(x) + \frac{1}{2} m''(x)\left(pV_L g^2 + V_K h^2\right) + O\left((g+h)^{2+\beta}\right) + O\left(\frac{(g+h)^2}{N^2 g^2 h^2}\right)$$

Proof. We have

$$\mathbb{E}\left(\hat{m}_p\left(x, h, g\right)\right) = \frac{1}{N^{p+1}} \sum_{i_1, i_2, \ldots, i_p, k} \mathcal{L}_g\left(x, x_{i_1}, \ldots, x_{i_p}\right) K_h\left(x_{i_p} - x_k\right) m\left(x_k\right)$$

We apply Lemma 3.3.7 to

$$g\left(u_1, \ldots, u_p, w\right) = \mathcal{L}_g\left(x, u_1, \ldots, u_p\right) K_h\left(u_p - w\right) m\left(w\right),$$

$d = p + 1$, and

$$D = \left\{ \left(u_1, \ldots, u_p, w\right) \in \mathbb{R}^{p+1} \middle| \, |x - u_1| < g, |u_i - u_{i+1}| < g; i = 1, \ldots, p-1, |u_p - w| < h \right\}$$

$$\left| \int \ldots \int_{[0,1]^d} g\left(u_1, \ldots, u_p, w\right) du_1 \ldots du_p dw - \frac{1}{N^d} \sum_{i_1, \ldots, i_d = 1}^{N} g(x_{i_1}, \ldots, x_{i_d}) \right|$$
$$\leq \frac{\varepsilon^d V}{12N^2} \left[\sum_{i=1}^{p} \sup_{y \in D} \left| \frac{\partial^2 g}{\partial u_i^2}\left(y\right) \right| + \sup_{y \in D} \left| \frac{\partial^2 g}{\partial w^2}\left(y\right) \right| \right]$$
$$+ \frac{3\varepsilon^d}{N^2} \left[\sum_{i=1}^{p} V_{u_i} \sup_{y \in D} \left| \frac{\partial g}{\partial u_i}\left(y\right) \right| + V_w \sup_{y \in D} \left| \frac{\partial g}{\partial w}\left(y\right) \right| \right]$$

The asymptotically order of $\sup_{y \in D} \left| \frac{\partial g}{\partial u_i}\left(y\right) \right|$:

$$\frac{\partial g}{\partial u_i}(y) = \left[-\left(L_g\right)'\left(u_{i-1} - u_i\right) L_g\left(u_i - u_{i+1}\right) + L_g\left(u_{i-1} - u_i\right)\left(L_g\right)'\left(u_i - u_{i+1}\right) \right]$$
$$\mathcal{L}_g\left(x, u_1, \ldots, u_{i-1}\right) \mathcal{L}_g\left(u_{i+1}, \ldots, u_p\right) K_h\left(u_p - w\right)$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial u_i}(y)\right| = O\left(\frac{1}{g^{p+1}h}\right);\ i=1,\ldots,p-1, \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_i^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial u_i^2}(y) = \big[(L_g)''(u_{i-1}-u_i)\,L_g(u_i-u_{i+1}) - (L_g)'(u_{i-1}-u_i)\,(L_g)'(u_i-u_{i+1})$$

$$-(L_g)'(u_{i-1}-u_i)\,(L_g)'(u_i-u_{i+1}) + L_g(u_{i-1}-u_i)\,(L_g)''(u_i-u_{i+1})\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_{i-1})\,\mathcal{L}_g(u_{i+1},\ldots,u_p)\,K_h(u_p-w)$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_i^2}(y)\right| = O\left(\frac{1}{g^{p+2}h}\right);\ i=1,\ldots,p-1, \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial g}{\partial u_p}(y)\right|$:

$$\frac{\partial g}{\partial u_p}(y) = \big[-(L_g)'(u_{p-1}-u_p)\,K_h(u_p-w) + L_g(u_{p-1}-u_p)\,(K_h)'(u_p-w)\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_{p-1})$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial u_p}(y)\right| = O\left(\frac{h+g}{g^{p+1}h^2}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_p^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial u_p^2}(y) = \big[(L_g)''(u_{p-1}-u_p)\,K_h(u_p-w) - (L_g)'(u_{p-1}-u_p)\,(K_h)'(u_p-w)$$

$$-(L_g)'(u_{p-1}-u_p)\,(K_h)'(u_p-w) + L_g(u_{p-1}-u_p)\,(K_h)''(u_p-w)\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_{p-1})$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_p^2}(y)\right| = O\left(\frac{(h+g)^2}{g^{p+2}h^3}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial g}{\partial w}(y)\right|$:

$$\frac{\partial g}{\partial w}(y) = \mathcal{L}_g(x,u_1,\ldots,u_p)\,\big[-(K_h)'(u_p-w)\,m(w) + K_h(u_p-w)\,m'(w)\big]$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial w}(y)\right| = O\left(\frac{1}{g^p h^2}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial w^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial w^2}(y) = \mathcal{L}_g(x, u_1, \ldots, u_p)\left[(K_h)''(u_p - w)\, m(w) - (K_h)'(u_p - w)\, m'(w) - (K_h)'(u_p - w)\, m'(w) + K_h(u_p - w)\, m''(w)\right]$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial w^2}(y)\right| = O\left(\frac{1}{g^p h^3}\right), \text{ uniformly for all } y \in D$$

Compute V:

$$V = \int \underset{D}{\ldots} \int du_1 \ldots du_p dw$$

$$D = \left\{ (u_1, \ldots, u_p, w) \in \mathbb{R}^{p+1} \middle| \, |x - u_1| < g, |u_i - u_{i+1}| < g; i = 1, \ldots, p-1, |u_p - w| < h \right\}$$

Using the following substitutions:

$$\begin{aligned} u_1 - x &= U_1, & u_1 &= x + U_1, \\ u_2 - u_1 &= U_2, & u_2 &= x + U_1 + U_2, \\ &\vdots & &\vdots \\ u_p - u_{p-1} &= U_p, & u_p &= x + U_1 + \ldots + U_p, \\ w - u_p &= W, & w &= x + U_1 + \ldots + U_p + W, \end{aligned} \tag{3.3.3}$$

we get

$$V = \int \underset{D'}{\ldots} \int \left| \frac{\partial (u_1, \ldots, u_p, w)}{\partial (U_1, \ldots, U_p, W)} \right| dU_1 \ldots dU_p dW = \int \underset{D'}{\ldots} \int dU_1 \ldots dU_p dW$$

where

$$D' = \left\{ (U_1, \ldots, U_p, W) \in \mathbb{R}^{p+1} \middle| \, |U_i| < g; i = 1, \ldots, p, |W| < h \right\}$$

That gives

$$V = (2g)^p (2h) = 2^{p+1} g^p h \tag{3.3.4}$$

Compute V_{u_i}

$$V_{u_i} = \int \dots \int_{D_{u_i}} du_1 \dots du_{i-1} du_{i+1} \dots du_p dw$$

where

$$\begin{aligned} D_{u_i} = \{ &(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_p, w) \in \mathbb{R}^p | \, \exists \, u_i \in \mathbb{R} : |x - u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h\} \\ = \{ &(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_p, w) \in \mathbb{R}^p | \, |x - u_1| < g, |u_1 - u_2| < g, \dots, \\ &|u_{i-2} - u_{i-1}| < g, |u_{i-1} - u_{i+1}| < 2g, |u_{i+1} - u_{i+2}| < g, \dots, |u_{p-1} - u_p| < g \; , \\ &|u_p - w| < h\} \end{aligned}$$

Analogously to the calculation of V we get

$$V_{u_i} = (2g)^{p-2} \, (4g) \, (2h) = 2^{p+1} g^{p-1} h, \;\; i = 1, \dots, p-1$$

Compute V_{u_p}:

$$\begin{aligned} D_{u_p} = \{ &(u_1, \dots, u_{p-1}, w) \in \mathbb{R}^p | \, \exists \, u_p \in \mathbb{R} : |x - u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h\} \\ = \{ &(u_1, \dots, u_{p-1}, w) \in \mathbb{R}^p | \, |x - u_1| < g, |u_1 - u_2| < g, \dots, \\ &|u_{p-2} - u_{p-1}| < g, |u_{p-1} - w| < g + h\} \end{aligned}$$

$$V_{u_p} = (2g)^{p-1} \, 2 \, (h + g) = 2^p g^{p-1} \, (h + g)$$

Compute V_{ω}:

$$\begin{aligned} D_w = \{ &(u_1, \dots, u_p) \in \mathbb{R}^p | \, \exists \, w \in \mathbb{R} : |x - u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h\} \\ = \{ &(u_1, \dots, u_{p-1}) \in \mathbb{R}^p | \, |x - u_1| < g, |u_j - u_{j+1}| < g; j = 1, \dots, p-1\} \end{aligned}$$

$$V_w = (2g)^p = 2^p g^p$$

Now,

$$\begin{aligned}
&\left| \underset{[0,1]^d}{\int \cdots \int} g\left(u_1, \ldots, u_p, w\right) du_1 \ldots du_p dw - \frac{1}{N^d} \sum_{i_1, \ldots, i_d = 1}^{N} g(x_{i_1}, \ldots, x_{i_d}) \right| \\
&\quad \leq \frac{2^{p+1} \varepsilon^d g^p h}{12 N^2} \left[(p-1)\, O\left(\frac{1}{g^{p+2} h}\right) + O\left(\frac{(h+g)^2}{g^{p+2} h^3}\right) + O\left(\frac{1}{g^p h^3}\right) \right] \\
&\qquad + \frac{3\varepsilon^d}{N^2} \left[(p-1)\, 2^{p+1} g^{p-1} h O\left(\frac{1}{g^{p+1} h}\right) + 2^p g^{p-1} (h+g)\, O\left(\frac{h+g}{g^{p+1} h^2}\right) + 2^p g^p O\left(\frac{1}{g^p h^2}\right) \right] \\
&\quad = \frac{2^{p+1} \varepsilon^d}{12 N^2} \left[O\left(\frac{1}{g^2}\right) + O\left(\frac{(h+g)^2}{g^2 h^2}\right) + O\left(\frac{1}{h^2}\right) \right] \\
&\qquad + \frac{3\varepsilon^d 2^p}{N^2} \left[O\left(\frac{1}{g^2}\right) + O\left(\frac{(h+g)^2}{g^2 h^2}\right) + O\left(\frac{1}{h^2}\right) \right] \\
&\quad = \frac{2^{p+1} \varepsilon^d}{12 N^2} O\left(\frac{(h+g)^2}{g^2 h^2}\right) + \frac{3\varepsilon^d 2^p}{N^2} O\left(\frac{(h+g)^2}{g^2 h^2}\right) = O\left(\frac{(h+g)^2}{N^2 g^2 h^2}\right)
\end{aligned}$$

That yields

$$\mathbb{E}\left(\hat{m}_p\left(x, h, g\right)\right) = \underbrace{\underset{[0,1]^d}{\int \cdots \int} \mathcal{L}\left(x, u_1, \ldots, u_p\right) K_h\left(u_p - w\right) m\left(w\right) du_1 \ldots du_p dw}_{=: I(x,h,g)} + O\left(\frac{(h+g)^2}{N^2 g^2 h^2}\right)$$

using the substitutions (3.3.3) we get

$$I\left(x, h, g\right) = \underset{[-1,1]^d}{\int \cdots \int} L_g\left(U_1\right) \ldots L_g\left(U_p\right) K_h\left(W\right) m\left(x + U_1 + \ldots + U_p + W\right) dU_1 \ldots dU_p dW$$

Note that the assumption $x \in [pg + h, 1 - (pg + h)]$ ensures that

$(x + U_1 + \ldots + U_p + W) \in [0, 1]$ in the last integral.

Now, using the Taylor expansion we get for some $\theta \in [0,1]$

$$
\begin{aligned}
I(x,h,g) = & \int_{[-1,1]^d} \dots \int L_g(U_1) \dots L_g(U_p) K_h(W) \Big[m(x) + m'(x)(U_1 + \dots + U_p + W) \\
& + \frac{1}{2}(U_1 + \dots + U_p + W)^2 m''(x + \theta(U_1 + \dots + U_p + W)) \Big] dU_1 \dots dU_p dW \\
= & m(x) + m'(x) \int_{[-1,1]^d} \dots \int L_g(U_1) \dots L_g(U_p) K_h(W)(U_1 + \dots + U_p + W) \\
& dU_1 \dots dU_p dW + \frac{1}{2} \int_{[-1,1]^d} \dots \int L_g(U_1) \dots L_g(U_p) K_h(W)(U_1 + \dots + U_p + W)^2 \\
& m''(x + \theta(U_1 + \dots + U_p + W)) dU_1 \dots dU_p dW \\
= & m(x) + \frac{1}{2} m''(x) \int_{[-1,1]^d} \dots \int L_g(U_1) \dots L_g(U_p) K_h(W)(U_1 + \dots + U_p + W)^2 \\
& dU_1 \dots dU_p dW + r_p(x,h,g) \\
= & m(x) + \frac{1}{2} m''(x) \left(pg^2 V_L + h^2 V_K \right) + r_p(x,h,g)
\end{aligned}
$$

where

$$
\begin{aligned}
r_p(x,h,g) = & \frac{1}{2} \int_{[-1,1]^d} \dots \int L_g(U_1) \dots L_g(U_p) K_h(W)(U_1 + \dots + U_p + W)^2 \\
& [m''(x + \theta(U_1 + \dots + U_p + W)) - m''(x)] dU_1 \dots dU_p dW
\end{aligned}
$$

Therefore

$$
\mathbb{E}(\hat{m}_p(x,h,g)) = m(x) + \frac{1}{2} m''(x) \left(pg^2 V_L + h^2 V_K \right) + r_p(x,h,g) + O\left(\frac{(h+g)^2}{N^2 g^2 h^2} \right)
$$

but $r_p(x,h,g) = O\left((g+h)^{2+\beta} \right)$ since, uniformly for all $x \in [pg+h, 1-(pg+h)]$, we have

$$
\begin{aligned}
|r_p(x,h,g)| \leq & \frac{1}{2} \int\limits_{[-1,1]^d} \ldots \int L_g(U_1) \ldots L_g(U_p) K_h(W) (U_1 + \ldots + U_p + W)^2 \\
& |m''(x + \theta(U_1 + \ldots + U_p + W)) - m''(x)| \, dU_1 \ldots dU_p dW \\
\leq & \frac{1}{2} \int\limits_{[-1,1]^d} \ldots \int L_g(U_1) \ldots L_g(U_p) K_h(W) (U_1 + \ldots + U_p + W)^2 H\theta^\beta \\
& |U_1 + \ldots + U_p + W|^\beta \, dU_1 \ldots dU_p dW \\
\leq & \frac{H}{2} \int\limits_{[-1,1]^d} \ldots \int \theta^\beta L_g(U_1) \ldots L_g(U_p) K_h(W) |U_1 + \ldots + U_p + W|^{2+\beta} \\
& dU_1 \ldots dU_p dW \\
\leq & \frac{H}{2} (pg + h)^{2+\beta} \int\limits_{[-1,1]^d} \ldots \int L_g(U_1) \ldots L_g(U_p) K_h(W) \, dU_1 \ldots dU_p dW \\
\leq & \frac{H}{2} p^{2+\beta} (g + h)^{2+\beta}
\end{aligned}
$$

and the assertion follows. □

For analyzing the variance of the RLLS-estimate we need first to analyze the covariances between the resmoothers $\hat{m}_p(x,h,g)$, $p \in \mathbb{N}$. For that purpose we use the notation L^{*p} for the p-fold convolution of L with itself.

Proposition 3.3.10. *Assume the deterministic equidistant design model (2.1.3), E1-E2, K1-K5, L1-L5, M1-M2, and $N \to \infty$, s.t. $h \to 0$, and $g \to 0$. Then for all $p, q \geq 0$ s.t. $p + q \geq 1$ we have, uniformly for all $x \in [h + pg, 1 - h - pg]$, and $\bar{x} \in [h + qg, 1 - h - qg]$,*

$$
\operatorname{Cov}\left(\hat{m}_p(x,h,g), \hat{m}_q(\bar{x},h,g)\right) = \frac{f(0)}{N} L_g^{*(p+q)} * K_h * K_h(x - \bar{x}) + O\left(\frac{1}{N^2 h^2}\right) + O\left(\frac{g+h}{N^3 g^2 h^2}\right)
$$

Proof.

$$
\begin{aligned}
&\mathrm{Cov}\left(\hat{m}_p(x,h,g), \hat{m}_q(\bar{x},h,g)\right) \\
&= \frac{1}{N^{p+q+2}} \sum_{i_1,i_2,\ldots,i_p,k} L_g(x-x_{i_1})L_g(x_{i_1}-x_{i_2})\ldots L_g(x_{i_{p-1}}-x_{i_p})K_h(x_{i_p}-x_k) \\
&\quad \sum_{j_1,j_2,\ldots,j_q,l} L_g(\bar{x}-x_{j_1})L_g(x_{j_1}-x_{j_2})\ldots L_g(x_{j_{q-1}}-x_{j_q})K_h(x_{j_q}-x_l)\mathrm{Cov}\left(Y_k,Y_l\right) \\
&= \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x,x_{i_1},x_{i_2},\ldots,x_{i_p}\right)\mathcal{L}_g\left(\bar{x},x_{j_1},x_{j_2},\ldots,x_{j_q}\right)K_h(x_{i_p}-x_k) \\
&\quad K_h(x_{j_q}-x_l)r_{k-l} \\
&= \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x,x_{i_1},x_{i_2},\ldots,x_{i_p}\right)\mathcal{L}_g\left(\bar{x},x_{j_1},x_{j_2},\ldots,x_{j_q}\right)K_h(x_{i_p}-x_k) \\
&\quad \left[K_h(x_{j_q}-x_l)-K_h(x_{j_q}-x_k)\right]r_{k-l} + \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x,x_{i_1},x_{i_2},\ldots,x_{i_p}\right) \\
&\quad \mathcal{L}_g\left(\bar{x},x_{j_1},x_{j_2},\ldots,x_{j_q}\right)K_h(x_{i_p}-x_k)K_h(x_{j_q}-x_k)r_{k-l} \\
&= V_1 + V_2
\end{aligned}
$$

For V_2 we have

$$
\begin{aligned}
V_2 &= \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x,x_{i_1},x_{i_2},\ldots,x_{i_p}\right)\mathcal{L}_g\left(\bar{x},x_{j_1},x_{j_2},\ldots,x_{j_q}\right) \\
&\qquad K_h(x_{i_p}-x_k)K_h(x_{j_q}-x_k)\sum_{l=1}^{N} r_{k-l}
\end{aligned}
$$

But $\sum_{l=1}^{N} r_{k-l} = \sum_{j=-\infty}^{\infty} r_{k-l} - \sum_{l\notin\{1,\ldots,N\}} r_{k-l} = f(0) - \sum_{l\notin\{1,\ldots,N\}} r_{k-l}$ implies

$$\begin{aligned} V_2 &= \frac{1}{N^{p+q+2}} f(0) \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) \\ &\quad \cdot K_h(x_{i_p} - x_k) K_h(x_{j_q} - x_k) - \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \sum_{l\notin\{1,\ldots,N\}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \\ &\quad \cdot \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) K_h(x_{j_q} - x_k) r_{k-l} \\ &= W_1 - W_2 \end{aligned}$$

Note that $W_2 = O\left(\frac{1}{N^2h^2}\right)$ since

$$\begin{aligned} |W_2| &\leq \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \sum_{l\notin\{1,\ldots,N\}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) \\ &\qquad K_h(x_{i_p} - x_k) K_h(x_{j_q} - x_k) \left|r_{k-l}\right| \\ &\leq \frac{B_K^2}{N^{p+q+2}h^2} \sum_{\substack{i_1,i_2,\ldots,i_p \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) \\ &\qquad \times \sum_{k=1}^{N} \sum_{l\notin\{1,\ldots,N\}} \left|r_{k-l}\right| \text{ where } B_K := \sup_u K(u) \\ &\overset{\text{Lemma } 3.3.1}{\leq} \frac{B_K^2 C}{N^2h^2} \frac{1}{N^p} \sum_{i_1,i_2,\ldots,i_p} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \frac{1}{N^q} \sum_{j_1,j_2,\ldots,j_q} \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) \\ &\overset{\text{Lemma } 3.3.8}{\leq} \frac{B_K^2 C'}{N^2h^2} \end{aligned}$$

$$\begin{aligned} V_1 &= \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) \\ &\qquad \left[K_h(x_{j_q} - x_l) - K_h(x_{j_q} - x_k)\right] r_{k-l} \end{aligned}$$

$$
\begin{aligned}
|V_1| &\le \frac{1}{N^{p+q+2}} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) \\
&\quad \left|K_h(x_{j_q} - x_l) - K_h(x_{j_q} - x_k)\right| |r_{k-l}| \\
&\le \frac{C_K}{N^{p+q+2}h^2} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q,l}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) \\
&\quad |x_k - x_l| \, |r_{k-l}| \\
&\le \frac{C_K}{N^{p+q+3}h^2} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) \\
&\quad \underbrace{\sum_{l=1}^{N} |k-l| \, |r_{k-l}|}_{\le \sum_{t=-\infty}^{\infty} |t||r_t| = 2\sum_{t=1}^{\infty} t|r_t| =: C < \infty} \\
&\le \frac{C_K C}{N^{p+q+3}h^2} \sum_{\substack{i_1,i_2,\ldots,i_p,k \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) K_h(x_{i_p} - x_k) \\
&\le \frac{C_K C}{N^{p+q+3}h^2} \sum_{\substack{i_1,i_2,\ldots,i_p \\ j_1,j_2,\ldots,j_q}} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right) \sum_{k=1}^{N} K_h(x_{i_p} - x_k) \\
&\le \frac{C_K C}{N^2 h^2} \frac{1}{N^p} \sum_{i_1,i_2,\ldots,i_p} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \frac{1}{N} \sum_{k=1}^{N} K_h(x_{i_p} - x_k) \\
&\quad \frac{1}{N^q} \sum_{j_1,j_2,\ldots,j_q} \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right)
\end{aligned}
$$

As $x \in [h+pg, 1-h-pg]$, and $\bar{x} \in [h+qg, 1-h-qg]$, we have $x_{i_p} \in [h, 1-h]$ for all non-vanishing terms in the p-fold sum $\sum_{i_1,i_2,\ldots,i_p} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \sum_{k=1}^{N} K_h(x_{i_p} - x_k)$. Therefore using Lemma (3.3.4) we have

$$
\begin{aligned}
|V1| &\le \frac{C_K C}{N^2 h^2} \frac{1}{N^p} \sum_{i_1,i_2,\ldots,i_p} \mathcal{L}_g\left(x, x_{i_1}, x_{i_2}, \ldots, x_{i_p}\right) \left[1 + O\left(\frac{1}{N^2 h^2}\right)\right] \\
&\quad \frac{1}{N^q} \sum_{j_1,j_2,\ldots,j_q} \mathcal{L}_g\left(\bar{x}, x_{j_1}, x_{j_2}, \ldots, x_{j_q}\right)
\end{aligned}
$$

and using Lemma (3.3.8) gives

$$|V_1| \ \le \ \frac{C_K C}{N^2h^2}\left[1+O\left(\frac{1}{N^2g^2}\right)\right]^2\left[1+O\left(\frac{1}{N^2h^2}\right)\right]$$

This yields $V_1 = O\left(\frac{1}{N^2h^2}\right)$. Now,

$$\begin{aligned}&\mathrm{Cov}\left(\hat{m}_p(x,h,g),\hat{m}_q(\bar{x},h,g)\right)\\ &\quad = \frac{f(0)}{N^{p+q+2}}\sum_{\substack{i_1,i_2,\ldots,i_p,k\\ j_1,j_2,\ldots,j_q}}\mathcal{L}_g\left(x,x_{i_1},x_{i_2},\ldots,x_{i_p}\right)\mathcal{L}_g\left(\bar{x},x_{j_1},x_{j_2},\ldots,x_{j_q}\right)K_h(x_{i_p}-x_k)\\ &\qquad\times K_h(x_{j_q}-x_k)+O\left(\frac{1}{N^2h^2}\right)\end{aligned}$$

If $|x-\bar{x}| \ge (p+q)g+2h$ then $W_1 = 0$ and $\mathrm{Cov}\left(\hat{m}_p(x,h,g),\hat{m}_q(\bar{x},h,g)\right) = O\left(\frac{1}{N^2h^2}\right)$. For $|x-\bar{x}| < (p+q)g+2h$ we apply Lemma 3.3.7 to

$$g\left(u_1,\ldots,u_p,w,v_q,\ldots,v_1\right) = \mathcal{L}_g\left(x,u_1,\ldots,u_p\right)\mathcal{L}_g\left(\bar{x},v_1,\ldots,v_q\right)K_h\left(u_p-w\right)K_h\left(v_q-w\right),$$

$d = p+q+1$, and

$$\begin{aligned}D = \Big\{&\left(u_1,\ldots,u_p,w,v_q,\ldots,v_1\right)\in\mathbb{R}^{p+q+1}\Big|\,|x-u_1|<g, |u_i-u_{i+1}|<g; i=1,\ldots,p-1,\\ &|\bar{x}-v_1|<g, |v_j-v_{j+1}|<g; j=1,\ldots,q-1, |u_p-w|<h, |v_q-w|<h\}\end{aligned}$$

and we get

$$\begin{aligned}&\left|\int\!\!\cdots\!\!\int_{[0,1]^d} g\left(u_1,\ldots,u_p,w,v_q,\ldots,v_1\right)du_1\ldots du_p dw dv_q\ldots dv_1 - \frac{1}{N^d}\sum_{i_1,\ldots,i_d=1}^{N} g(x_{i_1},\ldots,x_{i_d})\right|\\ &\quad\le \frac{\varepsilon^d V}{12N^2}\left[\sum_{i=1}^{p}\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_i^2}(y)\right| + \sup_{y\in D}\left|\frac{\partial^2 g}{\partial w^2}(y)\right| + \sum_{i=1}^{q}\sup_{y\in D}\left|\frac{\partial^2 g}{\partial v_i^2}(y)\right|\right]\\ &\qquad + \frac{3\varepsilon^d}{N^2}\left[\sum_{i=1}^{p}V_{u_i}\sup_{y\in D}\left|\frac{\partial g}{\partial u_i}(y)\right| + V_w\sup_{y\in D}\left|\frac{\partial g}{\partial w}(y)\right| + \sum_{i=1}^{q}V_{v_i}\sup_{y\in D}\left|\frac{\partial g}{\partial v_i}(y)\right|\right]\end{aligned}$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial g}{\partial u_i}(y)\right|$:

$$\frac{\partial g}{\partial u_i}(y) = \left[-\left(L_g\right)'\left(u_{i-1}-u_i\right)L_g\left(u_i-u_{i+1}\right)+L_g\left(u_{i-1}-u_i\right)\left(L_g\right)'\left(u_i-u_{i+1}\right)\right]$$

$$\mathcal{L}_g\left(x,u_1,\ldots,u_{i-1}\right)\mathcal{L}_g\left(u_{i+1},\ldots,u_p\right)\mathcal{L}_g\left(\bar{x},v_1,\ldots,v_q\right)K_h\left(u_p-w\right)K_h\left(v_q-w\right)$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial u_i}(y)\right| = O\left(\frac{1}{g^{p+q+1}h^2}\right);\ i=1,\ldots,p-1, \text{ uniformly for all } y\in D$$

and analogously we have

$$\sup_{y\in D}\left|\frac{\partial g}{\partial v_i}(y)\right| = O\left(\frac{1}{g^{p+q+1}h^2}\right);\ i=1,\ldots,q-1, \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial g}{\partial u_p}(y)\right|$:

$$\frac{\partial g}{\partial u_p}(y) = \left[-\left(L_g\right)'\left(u_{p-1}-u_p\right)K_h\left(u_p-w\right)+L_g\left(u_{p-1}-u_p\right)\left(K_h\right)'\left(u_p-w\right)\right]$$

$$\mathcal{L}_g\left(x,u_1,\ldots,u_{p-1}\right)\mathcal{L}_g\left(\bar{x},v_1,\ldots,v_q\right)K_h\left(v_q-w\right)$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial u_p}(y)\right| = O\left(\frac{h+g}{g^{p+q+1}h^3}\right), \text{ uniformly for all } y\in D$$

and analogously we have

$$\sup_{y\in D}\left|\frac{\partial g}{\partial v_q}(y)\right| = O\left(\frac{h+g}{g^{p+q+1}h^3}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial g}{\partial w}(y)\right|$:

$$\frac{\partial g}{\partial w}(y) = \left[-\left(K_h\right)'\left(u_p-w\right)K_h\left(v_q-w\right)-K_h\left(u_p-w\right)\left(K_h\right)'\left(v_q-w\right)\right]$$

$$\mathcal{L}_g\left(x,u_1,\ldots,u_p\right)\mathcal{L}_g\left(\bar{x},v_1,\ldots,v_q\right)$$

$$\sup_{y\in D}\left|\frac{\partial g}{\partial w}(y)\right| = O\left(\frac{1}{g^{p+q}h^3}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_i^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial u_i^2}(y) = \big[(L_g)''(u_{i-1}-u_i)\,L_g(u_i-u_{i+1}) - (L_g)'(u_{i-1}-u_i)\,(L_g)'(u_i-u_{i+1})$$

$$- (L_g)'(u_{i-1}-u_i)\,(L_g)'(u_i-u_{i+1}) + L_g(u_{i-1}-u_i)\,(L_g)''(u_i-u_{i+1})\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_{i-1})\,\mathcal{L}_g(u_{i+1},\ldots,u_p)\,\mathcal{L}_g(\bar{x},v_1,\ldots,v_q)\,K_h(u_p-w)\,K_h(v_q-w)$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_i^2}(y)\right| = O\left(\frac{1}{g^{p+q+2}h^2}\right);\ i=1,\ldots,p-1, \text{ uniformly for all } y\in D$$

and analogously we have

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial v_i^2}(y)\right| = O\left(\frac{1}{g^{p+q+2}h^2}\right);\ i=1,\ldots,q-1, \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_p^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial u_p^2}(y) = \big[(L_g)''(u_{p-1}-u_p)\,K_h(u_p-w) - (L_g)'(u_{p-1}-u_p)\,(K_h)'(u_p-w)$$

$$- (L_g)'(u_{p-1}-u_p)\,(K_h)'(u_p-w) + L_g(u_{p-1}-u_p)\,(K_h)''(u_p-w)\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_{p-1})\,\mathcal{L}_g(\bar{x},v_1,\ldots,v_q)\,K_h(v_q-w)$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial u_p^2}(y)\right| = O\left(\frac{(h+g)^2}{g^{p+q+2}h^4}\right), \text{ uniformly for all } y\in D$$

and analogously we have

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial v_q^2}(y)\right| = O\left(\frac{(h+g)^2}{g^{p+q+2}h^4}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of $\sup_{y\in D}\left|\frac{\partial^2 g}{\partial w^2}(y)\right|$:

$$\frac{\partial^2 g}{\partial w^2}(y) = \big[(K_h)''(u_p-w)\,K_h(v_q-w) + (K_h)'(u_p-w)\,(K_h)'(v_q-w)$$

$$+ (K_h)'(u_p-w)\,(K_h)'(v_q-w) + K_h(u_p-w)\,(K_h)''(v_q-w)\big]$$

$$\mathcal{L}_g(x,u_1,\ldots,u_p)\,\mathcal{L}_g(\bar{x},v_1,\ldots,v_q)$$

$$\sup_{y\in D}\left|\frac{\partial^2 g}{\partial w^2}(y)\right| = O\left(\frac{1}{g^{p+q}h^4}\right), \text{ uniformly for all } y\in D$$

The asymptotically order of V:

$$V = \int \underset{D}{\ldots} \int du_1 \ldots du_p dw dv_q \ldots dv_1$$

Using the following substitutions

$$\begin{aligned} u_1 - x &= U_1, & v_q - w &= V_q,\\ u_2 - u_1 &= U_2, & v_{q-1} - v_q &= V_{q-1},\\ &\vdots & &\vdots\\ u_p - u_{p-1} &= U_p, & v_1 - v_2 &= V_1\\ w - u_p &= W, & & \end{aligned}$$

$$\begin{aligned} u_1 &= x + U_1, & v_q &= x + U_1 + \ldots + U_p + W + V_q,\\ u_2 &= x + U_1 + U_2, & v_{q-1} &= x - U_1 + \ldots + U_p + W + V_q + V_{q-1},\\ &\vdots & &\vdots\\ u_p &= x + U_1 + \ldots + U_p, & v_1 &= x + U_1 + \ldots + U_p + W + V_q + \ldots + V_1,\\ w &= x + U_1 + \ldots + U_p + W, & \bar{x} - v_1 &= \bar{x} - x - U_1 - \ldots - U_p - W - V_q - \ldots - V_1 \end{aligned}$$

we get

$$V = \int \underset{D'}{\ldots} \int dU_1 \ldots dU_p dW dV_q \ldots dV_1$$

where

$$D' = \Big\{(U_1,\ldots,U_p,W,V_q,\ldots,V_1) \in \mathbb{R}^{p+q+1}\Big|\, |U_i| < g; i = 1,\ldots,p, |W| < h, |V_q| < h,$$
$$|V_j| < g; j = 1,\ldots,q-1, |x - \bar{x} + U_1 + \ldots + U_p + W + V_q + \ldots + V_1| < g\}$$

That gives

$$V \leq (2g)^p\,(2h)^2\,(2g)^{q-1} = 2^{p+q+1}g^{p+q-1}h^2 \tag{3.3.5}$$

Otherwise if we use the substitutions

$$\begin{aligned} u_1 - x &= U_1, & v_1 - \bar{x} &= V_1, \\ u_2 - u_1 &= U_2, & v_2 - v_1 &= V_2, \\ &\vdots & &\vdots \\ u_p - u_{p-1} &= U_p, & v_q - v_{q-1} &= V_q \\ w - u_p &= W, \end{aligned}$$

$$\begin{aligned} u_1 &= x + U_1, & v_1 &= \bar{x} + V_1, \\ u_2 &= x + U_1 + U_2, & v_2 &= \bar{x} + V_1 + V_2, \\ &\vdots & &\vdots \\ u_p &= x + U_1 + \ldots + U_p, & v_q &= \bar{x} + V_1 + \ldots + V_q, \\ w &= x + U_1 + \ldots + U_p + W, & v_q - w &= \bar{x} + V_1 + \ldots + V_q - x - U_1 - \ldots - U_p - W \end{aligned}$$

then we have

$$V = \int \underset{D''}{\ldots} \int dU_1 \ldots dU_p dW dV_q \ldots dV_1$$

where

$$\begin{aligned} D'' = \Big\{ (U_1, \ldots, U_p, W, V_q, \ldots, V_1) \in \mathbb{R}^{p+q+1} \Big| \, &|U_i| < g; i = 1, \ldots, p, |W| < h, \\ &|V_j| < g; j = 1, \ldots, q, |x - \bar{x} + U_1 + \ldots + U_p + W - V_1 - \ldots - V_q| < g \} \end{aligned}$$

That yields

$$V \leq (2g)^p\,(2h)\,(2g)^q = 2^{p+q+1}g^{p+q}h \tag{3.3.6}$$

From the inequalities (3.3.5), (3.3.6) we get

$$V \leq 2^{p+q+1}\left(g^{p+q}h\right) \wedge \left(g^{p+q-1}h^2\right) = 2^{p+q+1}g^{p+q-1}h\left(g \wedge h\right)$$

$$V = O\left(g^{p+q-1}h\left(g \wedge h\right)\right)$$

The asymptotically order of V_{u_i}:

$$V_{u_i} = \underset{D_{u_i}}{\int \dots \int} du_1 \dots du_{i-1} du_{i+1} \dots du_p dw dv_q \dots dv_1$$

where

$$\begin{aligned}
D_{u_i} = \Big\{ & (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_p, w, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, \exists\, u_i \in \mathbb{R} : |x - u_1| < g, \\
& |u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h, |w - v_q| < h, \\
& |v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \} \\
= \Big\{ & (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_p, w, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, |x - u_1| < g, |u_1 - u_2| < g, \\
& \dots, |u_{i-2} - u_{i-1}| < g, |u_{i-1} - u_{i+1}| < g, |u_{i+1} - u_{i+2}| < g, \dots, |u_{p-1} - u_p| < g, \\
& |u_p - w| < h, |w - v_q| < h, |v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \}
\end{aligned}$$

Analogously to deriving the asymptotically order of V we get

$$V_{u_i} \leq 2^{p+q}\left(g^{p+q-1}h\right) \wedge \left(g^{p+q-2}h^2\right) = 2^{p+q}g^{p+q-2}h\left(g \wedge h\right)$$

$$V_{u_i} = O\left(g^{p+q-2}h\left(g \wedge h\right)\right),\ i = 1, \dots, p-1$$

and for V_{v_i} we have

$$V_{v_i} = O\left(g^{p+q-2}h\left(g \wedge h\right)\right),\ i = 1, \dots, q-1$$

The asymptotically order of V_{u_p}:

$$\begin{aligned}D_{u_p} = \Big\{ (u_1, \dots, u_{p-1}, w, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, &\exists\, u_p \in \mathbb{R} : |x-u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h, |w - v_q| < h, \\ &|v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \Big\} \\ = \Big\{ (u_1, \dots, u_{p-1}, w, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, &|x-u_1| < g, |u_1 - u_2| < g, \dots, \\ &|u_{p-2} - u_{p-1}| < g, |u_{p-1} - w| < g + h, |w - v_q| < h, \\ &|v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \Big\}\end{aligned}$$

Analogously to deriving the asymptotically order of V we get

$$V_{u_p} \le 2^{p+q} \left(g^{p+q-2} h(g+h) \right) \wedge \left(g^{p+q-1} h \right) \wedge \left(g^{p+q-1} (g+h) \right) = 2^{p+q} g^{p+q-1} h$$

$$V_{u_p} = O\left(g^{p+q-1} h \right)$$

and for V_{v_q} we have

$$V_{v_q} = O\left(g^{p+q-1} h \right)$$

The asymptotically order of V_w:

$$\begin{aligned}D_w = \Big\{ (u_1, \dots, u_p, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, &\exists\, w \in \mathbb{R} : |x-u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - w| < h, |w - v_q| < h, \\ &|v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \Big\} \\ = \Big\{ (u_1, \dots, u_{p-1}, v_q, \dots, v_1) \in \mathbb{R}^{p+q} \Big| \, &|x-u_1| < g, \\ &|u_j - u_{j+1}| < g; j = 1, \dots, p-1, |u_p - v_q| < 2h, \\ &|v_{j+1} - v_j| < g; j = 1, \dots, q-1, |v_1 - \bar{x}| < g \Big\}\end{aligned}$$

Analogously to deriving the asymptotically order of V we get

$$V_w \le 2^{p+q} \left(g^{p+q} \right) \wedge \left(g^{p+q-1} (2h) \right) = 2^{p+q} g^{p+q-1} \left(g \wedge (2h) \right)$$

$$V_w = O\left(g^{p+q-1}\left(g \wedge h\right)\right)$$

since $\frac{1}{2}(g \wedge (2h)) \leq g \wedge h \leq g \wedge (2h)$.

Now,

$$\left| \int \underset{[0,1]^d}{\dots} \int g\left(u_1, \dots, u_p, w, v_q, \dots, v_1\right) du_1 \dots du_p dw dv_q \dots dv_1 - \frac{1}{N^d} \sum_{i_1, \dots, i_d = 1}^{N} g(x_{i_1}, \dots, x_{i_d}) \right|$$

$$\begin{aligned}
&\leq \frac{\varepsilon^d O\left(g^{p+q-1} h\left(g \wedge h\right)\right)}{12N^2} \left[(p-1)O\left(\frac{1}{g^{p+q+2}h^2}\right) + O\left(\frac{(h+g)^2}{g^{p+q+2}h^4}\right) + O\left(\frac{1}{g^{p+q}h^4}\right)\right. \\
&\quad \left. + (q-1)O\left(\frac{1}{g^{p+q+2}h^2}\right) + O\left(\frac{(h+g)^2}{g^{p+q+2}h^4}\right)\right] \\
&\quad + \frac{3\varepsilon^d}{N^2}\left[(p-1)O\left(g^{p+q-2}h\left(g \wedge h\right)\right)O\left(\frac{1}{g^{p+q+1}h^2}\right) + O\left(g^{p+q-1}h\right)O\left(\frac{h+g}{g^{p+q+1}h^3}\right)\right. \\
&\quad + O\left(g^{p+q-1}\left(g \wedge h\right)\right)O\left(\frac{1}{g^{p+q}h^3}\right) + (q-1)O\left(g^{p+q-2}h\left(g \wedge h\right)\right)O\left(\frac{1}{g^{p+q+1}h^2}\right) + \\
&\quad \left. + O\left(g^{p+q-1}h\right)O\left(\frac{h+g}{g^{p+q+1}h^3}\right)\right] \\
&\leq \frac{\varepsilon^d}{12N^2} O\left(g^{p+q-1}h\left(g \wedge h\right)\right)\left[O\left(\frac{1}{g^{p+q+2}h^2}\right) + O\left(\frac{(h+g)^2}{g^{p+q+2}h^4}\right) + O\left(\frac{1}{g^{p+q}h^4}\right)\right] \\
&\quad + \frac{3\varepsilon^d}{N^2}\left[O\left(g^{p+q-2}h\left(g \wedge h\right)\right)O\left(\frac{1}{g^{p+q+1}h^2}\right) + O\left(g^{p+q-1}h\right)O\left(\frac{h+g}{g^{p+q+1}h^3}\right)\right. \\
&\quad \left. + O\left(g^{p+q-1}\left(g \wedge h\right)\right)O\left(\frac{1}{g^{p+q}h^3}\right)\right] \\
&\leq \frac{\varepsilon^d}{12N^2} O\left(g \wedge h\right)\left[O\left(\frac{1}{g^3 h}\right) + O\left(\frac{(h+g)^2}{g^3 h^3}\right) + O\left(\frac{1}{g h^3}\right)\right] \\
&\quad + \frac{3\varepsilon^d}{N^2}\left[O\left(\frac{g \wedge h}{g^3 h}\right) + O\left(\frac{h+g}{g^2 h^2}\right) + O\left(\frac{g \wedge h}{g h^3}\right)\right] \\
&\leq \frac{\varepsilon^d}{12N^2} O\left(g \wedge h\right) O\left(\frac{(h+g)^2}{g^3 h^3}\right) + \frac{3\varepsilon^d}{N^2} O\left(\frac{h+g}{g^2 h^2}\right) = O\left(\frac{(g \wedge h)(h+g)^2}{N^2 g^3 h^3}\right) \\
&\leq O\left(\frac{g+h}{N^2 g^2 h^2}\right)
\end{aligned}$$

$$
\begin{aligned}
&\operatorname{Cov}\left(\hat{m}_p(x,h,g),\hat{m}_q(\bar{x},h,g)\right)\\
&\quad=\frac{f(0)}{N^{p+q+2}}\sum_{\substack{i_1,i_2,\ldots,i_p,k\\ j_1,j_2,\ldots,j_q}} g\left(x_{i_1},x_{i_2},\ldots,x_{i_p},x_k,x_{j_1},x_{j_2},\ldots,x_{j_q}\right)+O\left(\frac{1}{N^2h^2}\right)\\
&\quad=\frac{f(0)}{N}\int\limits_{[0,1]^d}\ldots\int \mathcal{L}_g\left(x,u_1,\ldots,u_p\right)\mathcal{L}_g\left(\bar{x},v_1,\ldots,v_q\right)K_h\left(u_p-w\right)K_h\left(v_q-w\right)du_1\ldots\\
&\qquad\quad du_p dw dv_q\ldots dv_1+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)
\end{aligned}
$$

Using the substations:

$$
\begin{aligned}
u_1-x&=U_1, & v_q-w&=V_q,\\
u_2-u_1&=U_2, & v_{q-1}-v_q&=V_{q-1},\\
&\vdots & &\vdots\\
u_p-u_{p-1}&=U_p, & v_1-v_2&=V_1\\
w-u_p&=W,
\end{aligned}
$$

$$
\begin{aligned}
u_1&=x+U_1, & v_q&=x+U_1+\ldots+U_p+W+V_q,\\
u_2&=x+U_1+U_2, & v_{q-1}&=x-U_1+\ldots+U_p+W+V_q+V_{q-1},\\
&\vdots & &\vdots\\
u_p&=x+U_1+\ldots+U_p, & v_1&=x+U_1+\ldots+U_p+W+V_q+\ldots+V_1,\\
w&=x+U_1+\ldots+U_p+W, & \bar{x}-v_1&=\bar{x}-x-U_1-\ldots-U_p-W-V_q-\ldots-V_1
\end{aligned}
$$

with the Jacobian

$$
J=\frac{\partial\left(u_1,\ldots,u_p,w,v_q,\ldots,v_1\right)}{\partial\left(U_1,\ldots,U_p,W,V_q,\ldots,V_1\right)}=1
$$

we get

$$
\begin{aligned}
&\operatorname{Cov}\left(\hat{m}_p(x,h,g),\hat{m}_q(\bar{x},h,g)\right)\\
&\quad=\frac{f(0)}{N}\underset{[-1,1]^d}{\int\ldots\int} L_g\left(U_1\right)\ldots L_g\left(U_p\right)K_h\left(W\right)K_h\left(V_q\right)L_g\left(V_{q-1}\right)\ldots L_g\left(V_1\right)\\
&\qquad L_g\left(\bar{x}-x-U_1-\ldots-U_p-W-V_q-\ldots-V_1\right)dU_1\ldots dU_p dW dV_q\ldots dV_1\\
&\qquad+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)\\
&\quad=\frac{f(0)}{N}\underset{[-1,1]^{d-1}}{\int\ldots\int}\left[L_g\left(U_1\right)\ldots L_g\left(U_p\right)K_h\left(W\right)K_h\left(V_q\right)L_g\left(V_{q-1}\right)\ldots L_g\left(V_2\right)\int_{-1}^{1}L_g\left(V_1\right)\right.\\
&\qquad\left.L_g\left(\bar{x}-x-U_1-\ldots-U_p-W-V_q-\ldots-V_1\right)dV_1\right]dU_1\ldots dU_p dW dV_q\ldots dV_2\\
&\qquad+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)\\
&\quad=\frac{f(0)}{N}\underset{[-1,1]^{d-1}}{\int\ldots\int} L_g\left(U_1\right)\ldots L_g\left(U_p\right)K_h\left(W\right)K_h\left(V_q\right)L_g\left(V_{q-1}\right)\ldots L_g\left(V_2\right)\\
&\qquad L_g * L_g\left(\bar{x}-x-U_1-\ldots-U_p-W-V_q-\ldots V_2\right)dU_1\ldots dU_p dW dV_q\ldots dV_2\\
&\qquad+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)\\
&\quad=\frac{f(0)}{N}L_g^{*p} * K_h * K_h * L_g^{*q}\left(\bar{x}-x\right)+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)\\
&\quad=\frac{f(0)}{N}L_g^{*(p+q)}K_h * K_h\left(x-\bar{x}\right)+O\left(\frac{g+h}{N^3g^2h^2}\right)+O\left(\frac{1}{N^2h^2}\right)
\end{aligned}
$$

□

We need the following lemma which takes care of the boundary effects of p_L:

Lemma 3.3.11. *Assuming L1-L3 for a kernel L, we have for all* $x \in [0,1]$

$$\frac{1}{2}-\frac{C_L}{Ng^2}\leq p_L\left(x,g\right)\leq 1+\frac{C_L}{Ng^2}\ \textit{if}\ g\leq\frac{1}{2}$$

Proof. Using Lemma 2.3.2, we have for all $x \in [0,1]$

$$\left| p_L(x,g) - \int_0^1 L_g(x-y)\,dy \right| \le \frac{C_L}{Ng^2}$$

Now, $\int_0^1 L_g(x-y)\,dy = \int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du$ and by non-negativity of L, we get

$$\int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du \le \int_{-1}^{1} L(u)\,du = 1$$

By symmetry and non-negativity of L, we have

If $x \le g$, then

$$\int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du \overset{g\le\frac{1}{2}}{=} \int_{-1}^{\frac{x}{g}} L(u)\,du = 1 - \underbrace{\int_{\frac{x}{g}}^{1} L(u)\,du}_{\le\frac{1}{2}} \ge \frac{1}{2}$$

If $x \ge 1-g$, then

$$\int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du \overset{g\le\frac{1}{2}}{=} \int_{\frac{x-1}{g}}^{1} L(u)\,du = 1 - \int_{-1}^{\frac{x-1}{g}} L(u)\,du \ge \frac{1}{2}$$

If $g \le x \le 1-g$, then

$$\int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du = \int_{-1}^{1} L(u)\,du = 1$$

From that we conclude

$$\frac{1}{2} - \frac{C_L}{Ng^2} \le \int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du - \frac{C_L}{Ng^2} \le p_L(x,g) \le \int_{\frac{x-1}{g}}^{\frac{x}{g}} L(u)\,du + \frac{C_L}{Ng^2} \le 1 + \frac{C_L}{Ng^2}$$

for all $x \in [0,1]$. □

Now let us consider the matrices P, Λ defined before Proposition 3.2.1.

Lemma 3.3.12. *Under the assumptions L1-L3, the norm of the matrix* Λ *is bounded i.e.* $\|\Lambda\| = O(1)$ *as* $N \to \infty$ *s.t.* $g \to 0$, $Ng^2 \to \infty$.

Proof. Using the definitions of Λ and p_L, we have

$$\begin{aligned}
\|\Lambda\| &= \max_{1\le i\le N}\sum_{j=1}^{N}|\Lambda_{ij}| \\
&= \max_{1\le i\le N}\left[\frac{1}{N}\sum_{j=1}^{N}L_g\left(x_i-x_j\right)\right] \\
&= \max_{1\le i\le N}p_L\left(x_i,g\right) \\
&\overset{\exists i_N}{=} p_L\left(x_{i_N},g\right)
\end{aligned}$$

By lemma 3.3.11 this implies

$$\|\Lambda\|\le 1+\frac{C_L}{Ng^2}$$

and $\|\Lambda\|=O\left(1\right)$ as $N\to\infty$ s.t. $g\to 0$, $Ng^2\to\infty$. □

Lemma 3.3.13. *Assuming K1-K3, L1-L3, we have*

i) $\|P\|=O\left(1\right)$, $\|P^{-1}\|=O\left(1\right)$.

ii) If additionally K4-K5, L4-L5 are assumed. Then

$$\frac{1}{p_\lambda\left(x,h,g\right)}=\frac{1}{1+\lambda}+O\left(\frac{1}{N^2h^2}\right)+O\left(\frac{\lambda}{N^2g^2}\right)$$

uniformly for all $x\in\left[\max\left(h,g\right),1-\max\left(h,g\right)\right]$.

as $N\to\infty$ *s.t.* $h\to 0$, $g\to 0$, $\lambda\to 0$, $Nh^2\to\infty$, *and* $Ng^2\to\infty$.

Proof. i) Using the definition of P, p_λ we have

$$\begin{aligned}
\|P\| &= \max_{1\le i\le N}p_\lambda\left(x_i,h,g\right)=\max_{1\le i\le N}\left(p_K\left(x_i,h\right)+\lambda p_L\left(x_i,g\right)\right) \\
&\overset{\exists i_N}{=} p_K\left(x_{i_N},h\right)+\lambda p_L\left(x_{i_N},g\right)
\end{aligned}$$

Therefore, by lemma 3.3.11,

$$\|P\|\le 1+\lambda+\frac{C_K}{Nh^2}+\lambda\frac{C_L}{Ng^2}\underset{N\to\infty}{\longrightarrow}1$$

$$\left\|P^{-1}\right\| = \|P\|^{-1} \leq \frac{1}{\frac{1}{2} - \frac{C_K}{Nh^2} + \frac{\lambda}{2} - \lambda\frac{C_L}{Ng^2}} \underset{N\to\infty}{\longrightarrow} 2$$

and i) follows.

ii) For $x \in [\max(h,g), 1-\max(h,g)]$ we have,

$$\begin{aligned}
\left|\frac{1}{p_\lambda(x,h,g)} - \frac{1}{1+\lambda}\right| &= \left|\frac{1-p_K(x,h)+\lambda(1-p_L(x,g))}{(1+\lambda)(p_K(x,h)+\lambda p_L(x,g))}\right| \\
&\leq \frac{|1-p_K(x,h)| + \lambda|1-p_L(x,g)|}{(1+\lambda)(p_K(x,h)+\lambda p_L(x,g))} \\
&\overset{\text{Corollary } 3.3.4}{\underset{\text{Lemma } 3.3.11}{\leq}} \frac{\frac{C_K}{N^2h^2} + \lambda\frac{C_L}{N^2g^2}}{(1+\lambda)\left(\frac{1}{2}+\frac{\lambda}{2}-\frac{C_K}{Nh^2}-\lambda\frac{C_L}{Ng^2}\right)} \\
&\overset{\exists C>0}{\leq} \frac{1}{C}\left(\frac{C_K}{Nh^2}+\lambda\frac{C_L}{Ng^2}\right)
\end{aligned}$$

and ii) follows. □

Theorem 3.3.14. *Assume the deterministic equidistant design model (2.1.3), E1-E2, K1-K5, L1-L5, M1-M2, and $N\to\infty$, s.t. $h\to 0$, $g\to 0$, $\lambda\to 0$, and $N^2g^2h^2(g+h)^\beta\to\infty$. Assume that there is an integer number $t>0$ such that $N\lambda^{t+1} = o\left((g+h)^{2+\beta}\right)$, and $\lambda^{t+1} = o\left(\frac{1}{N^2h^2}\right)$. Then uniformly for all i satisfying $x_i \in [\max(h,g)+tg, 1-\max(h,g)-tg]$ we have,*

i) $\operatorname{bias}(\hat{u}_i) = \mathbb{E}(\hat{u}_i) - u_i = \frac{1}{2}\left[h^2V_K + \lambda g^2 V_L\right] m''(x_i) + O\left((g+h)^{2+\beta}\right)$.

ii) $\operatorname{Var}(\hat{u}_i) = \frac{f(0)}{Nh} Q\left(\frac{g}{h},\lambda\right) + O\left(\frac{1}{N^2h^2}\right)$
where $f(\cdot)$ is the spectral density of the noise-process, $\theta := \frac{\lambda}{1+\lambda}$ and

$$Q(b,\lambda) = \frac{1}{2\pi}\int\left(\frac{\widehat{K}(\omega)}{1+\lambda-\lambda\widehat{L}(b\omega)}\right)^2 d\omega$$

where $\widehat{K}(\omega) = \int K(z)e^{iwz}dz$, $\widehat{L}(\omega) = \int L(z)e^{iwz}dz$ are the Fourier transforms of K, L.

We have $Q(b,\lambda) = Q_K + O(\lambda)$ and $\frac{Q_K}{1+2\lambda} \leq Q(b,\lambda) \leq Q_K$.

iii) The mean squared error of $\hat{u}_i$ *is*

$$\text{mse}(\hat{u}_i) = \text{amse}(\hat{u}_i) + R^*_{N,i}$$

where $R^*_{N,i} = O\left(\left(\lambda g^2 + h^2\right)(g+h)^{2+\beta}\right) + O\left((g+h)^{4+2\beta}\right) + O\left(\frac{1}{N^2h^2}\right)$ *and the asymptotic mean squared error is given by*

$$\text{amse}(\hat{u}_i) = \frac{1}{4}\left[h^2 V_K + \lambda g^2 V_L\right]^2 \left[m''(x_i)\right]^2 + \frac{f(0)}{Nh} Q\left(\frac{g}{h}, \lambda\right) \tag{3.3.7}$$

In particular,

$$\hat{u}_i \xrightarrow{P} u_i = m(x_i)$$

i.e. $\hat{u}_i$ *is consistent estimator of* $m(x_i)$.

Proof. Using Lemma 3.3.12, and Lemma 3.3.13 we see that $\|\lambda\Lambda P^{-1}\| \to 0$ as $\lambda \to 0$. Therefore we can expand the factor of $\hat{\mathbf{u}}_{PC}$ in the relation (3.2.3) as

$$\begin{aligned}
(P - \lambda\Lambda)^{-1} &= \left[\left(I - \lambda\Lambda P^{-1}\right)P\right]^{-1} \\
&= P^{-1}\left(I - \lambda\Lambda P^{-1}\right)^{-1} \\
&= P^{-1}\left[\sum_{n=0}^{t} \lambda^n \left(\Lambda P^{-1}\right)^n\right] + O\left(\lambda^{t+1}\left\|P^{-1}\right\|\left\|\left(\Lambda P^{-1}\right)^{t+1}\right\|\right)
\end{aligned}$$

where the reminder term is of order λ^{t+1} by Lemmas 3.3.12, 3.3.13. Therefore we have

$$\begin{aligned}
\mathbb{E}\hat{\mathbf{u}} = (P - \lambda\Lambda)^{-1}\mathbb{E}\hat{\mathbf{u}}_{PC} &= \left[P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n + O\left(\lambda^{t+1}\right)\right]\mathbb{E}\hat{\mathbf{u}}_{PC} \\
&= \left[P^{-1} + P^{-1}\sum_{n=1}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n + O\left(\lambda^{t+1}\right)\right]\mathbb{E}\hat{\mathbf{u}}_{PC}
\end{aligned} \tag{3.3.8}$$

But

$$\Lambda P^{-1} = \left(\sum_k \Lambda_{ik} P^{-1}_{kj}\right)_{ij} = \left(\frac{\Lambda_{ij}}{P_{jj}}\right)_{ij}$$

$$\left(\Lambda P^{-1}\right)^n = \left(\sum_{j_1,\ldots,j_{n-1}} \frac{\Lambda_{ij_1}}{P_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{P_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j}}{P_{jj}}\right)_{ij}$$

$$P^{-1}\sum_{n=1}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n = \left(P_{ii}^{-1}\sum_{n=1}^{t}\lambda^n\sum_{j_1,\ldots,j_{n-1}} \frac{\Lambda_{ij_1}}{P_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{P_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j}}{P_{jj}}\right)_{ij}$$

and, as $\mathbb{E}\left(m_K\left(x_i,h\right)\right) = m(x_i) + O\left(h^2\right)$ by Proposition 3.3.6, $\mathbb{E}\left(\hat{m}_K\left(x_i,h\right)\right)$ is bounded uniformly for all $i \in \{i; x_i \in [h, 1-h]\}$. Therefore formula (3.3.8) can, coordinate-wise, be written as

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \sum_k P_{ik}^{-1}\mathbb{E}\hat{u}_k^{PC} + \sum_{j_n} P_{ii}^{-1}\sum_{n=1}^{t}\lambda^n\sum_{j_1,\ldots,j_{n-1}} \frac{\Lambda_{ij_1}}{P_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{P_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j_n}}{P_{j_nj_n}}\mathbb{E}\hat{u}_{j_n}^{PC} \qquad (3.3.9)\\ &\quad + O\left(\lambda^{t+1}\right)\sum_{i=1}^{N}\mathbb{E}\hat{u}_i^{PC}\\ &= \frac{1}{P_{ii}}\mathbb{E}\hat{u}_i^{PC} + \frac{1}{P_{ii}}\sum_{n=1}^{t}\lambda^n\sum_{j_1,\ldots,j_n}\frac{\Lambda_{ij_1}}{P_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{P_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j_n}}{P_{j_nj_n}}\mathbb{E}\hat{u}_{j_n}^{PC} + O\left(\lambda^{t+1}\right)O\left(N\right)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \frac{1}{p_\lambda\left(x_i,h,g\right)}\sum_{n=1}^{t}\frac{\lambda^n}{N^n}\sum_{j_1,\ldots,j_n}\frac{L_g\left(x_i - x_{j_1}\right)}{p_\lambda\left(x_{j_1},h,g\right)}\frac{L_g\left(x_{j_1} - x_{j_2}\right)}{p_\lambda\left(x_{j_2},h,g\right)}\cdots\frac{L_g\left(x_{j_{n-1}} - x_{j_n}\right)}{p_\lambda\left(x_{j_n},h,g\right)}\\ &\quad \cdot\mathbb{E}\hat{u}_{j_n}^{PC} + \frac{1}{p_\lambda\left(x_i,h,g\right)}\mathbb{E}\hat{u}_i^{PC} + O\left(N\lambda^{t+1}\right) \qquad (3.3.10)\end{aligned}$$

As $\max(h,g) + tg \leq x_i \leq 1 - \max(h,g) - tg$, and L_g has support $[-g,g]$, we have $\max(h,g) \leq x_{j_n} \leq 1 - \max(h,g)$ for all j_n, $n = 1,\ldots,t$, corresponding to the non-vanishing terms in the n-fold sum of (3.3.10). Therefore, we may apply Lemma 3.3.13 to replace $p_\lambda\left(x_{j_n},h,g\right)$ by $1+\lambda$ in (3.3.10), and we get

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \frac{1}{1+\lambda}\sum_{n=1}^{t}\frac{\lambda^n}{N^n}\sum_{j_1,\ldots,j_n}\frac{L_g\left(x_i - x_{j_1}\right)}{1+\lambda}\frac{L_g\left(x_{j_1} - x_{j_2}\right)}{1+\lambda}\cdots\frac{L_g\left(x_{j_{n-1}} - x_{j_n}\right)}{1+\lambda}\mathbb{E}\hat{u}_{j_n}^{PC}\\ &\quad + \frac{1}{1+\lambda}\mathbb{E}\hat{u}_i^{PC} + O\left(N\lambda^{t+1}\right) + O\left(\frac{1}{N^2h^2}\right) + O\left(\frac{\lambda}{N^2g^2}\right)\end{aligned}$$

and using the definition of $\hat{m}_n$, $n \geq 0$ it can be shown

$$
\begin{aligned}
\mathbb{E}\hat{u}_i &= \frac{1}{1+\lambda}\sum_{n=1}^{t}\frac{\lambda^n}{(1+\lambda)^n}\frac{1}{N^n}\sum_{j_1,\ldots,j_n} L_g\left(x_i - x_{j_1}\right) L_g\left(x_{j_1} - x_{j_2}\right)\ldots L_g\left(x_{j_{n-1}} - x_{j_n}\right) \\
&\cdot \mathbb{E}\hat{u}_{j_n}^{PC} + \frac{1}{1+\lambda}\mathbb{E}\hat{u}_i^{PC} + R_{N,i} \\
&= \frac{1}{1+\lambda}\left[\sum_{n=1}^{t}\frac{\lambda^n}{(1+\lambda)^n}\mathbb{E}\hat{m}_n\left(x_i,h,g\right) + \mathbb{E}\hat{m}_K\left(x_i,h\right)\right] + R_{N,i} \\
&= (1-\theta)\sum_{n=0}^{t}\theta^n\mathbb{E}\hat{m}_n\left(x_i,h,g\right) + R_{N,i}
\end{aligned}
$$

where $\theta := \frac{\lambda}{1+\lambda}$ and $R_{N,i} = O\left(N\lambda^{t+1}\right) + O\left(\frac{1}{N^2h^2}\right) + O\left(\frac{\lambda}{N^2g^2}\right)$. Using Proposition 3.3.9 we get

$$
\begin{aligned}
\mathbb{E}\left(\hat{u}_i\right) &= (1-\theta)\sum_{n=0}^{t}\theta^n\left[m(x) + \frac{1}{2}m''(x)\left(nV_Lg^2 + V_Kh^2\right)\right. \\
&\left. + O\left((g+h)^{2+\beta}\right) + O\left(\frac{(g+h)^2}{N^2g^2h^2}\right)\right] + R_{N,i}
\end{aligned}
$$

But we have

$$
(1-\theta)\sum_{n=0}^{t}\theta^n = 1 - \theta^{t+1} = 1 + O\left(\theta^{t+1}\right) = 1 + O\left(\lambda^{t+1}\right)
$$

$$
(1-\theta)\sum_{n=0}^{t}n\theta^n = \frac{\theta - (t+1)\,\theta^{t+1} + t\theta^{t+2}}{1-\theta} = \lambda + O\left(\theta^{t+1}\right) = \lambda + O\left(\lambda^{t+1}\right)
$$

$$
\begin{aligned}
\mathbb{E}\left(\hat{u}_i\right) &= \left[1 + O\left(\lambda^{t+1}\right)\right]\left[m(x) + \frac{1}{2}m''(x)V_Kh^2 + O\left((g+h)^{2+\beta}\right) + O\left(\frac{(g+h)^2}{N^2g^2h^2}\right)\right] \\
&+ \left[\lambda + O\left(\lambda^{t+1}\right)\right]\frac{1}{2}m''(x)V_Lg^2 + R_{N,i} \\
&= m(x) + \frac{1}{2}m''(x)V_Kh^2 + \lambda\frac{1}{2}m''(x)V_Lg^2 + O\left((g+h)^{2+\beta}\right) + O\left(\frac{(g+h)^2}{N^2g^2h^2}\right) \\
&+ O\left(\lambda^{t+1}\right)\left[m(x) + \frac{1}{2}m''(x)V_Kh^2 + O\left((g+h)^{2+\beta}\right) + O\left(\frac{(g+h)^2}{N^2g^2h^2}\right)\right] \\
&+ O\left(\lambda^{t+1}\right)\frac{1}{2}m''(x)V_Lg^2 + R_{N,i}
\end{aligned}
$$

$$\begin{aligned}
\mathbb{E}\left(\hat{u}_i\right) &= m(x)+\frac{1}{2}m''(x)V_K h^2+\frac{1}{2}m''(x)V_L\lambda g^2+O\left((g+h)^{2+\beta}\right)+O\left(\frac{(g+h)^2}{N^2g^2h^2}\right)\\
&\quad +O\left(\lambda^{t+1}\right)+R_{N,i}\\
&= m(x)+\frac{1}{2}m''(x)V_K h^2+\frac{1}{2}m''(x)V_L\lambda g^2+O\left((g+h)^{2+\beta}\right)+O\left(\frac{(g+h)^2}{N^2g^2h^2}\right)\\
&\quad +O\left(\lambda^{t+1}\right)+O\left(N\lambda^{t+1}\right)+O\left(\frac{1}{N^2h^2}\right)+O\left(\frac{\lambda}{N^2g^2}\right)
\end{aligned}$$

As $N^2g^2h^2(g+h)^\beta\to\infty$ then we have

$$\mathbb{E}\left(\hat{u}_i\right)=m(x)+\frac{1}{2}m''(x)V_K h^2+\frac{1}{2}m''(x)V_L\lambda g^2+O\left(N\lambda^{t+1}\right)+O\left((g+h)^{2+\beta}\right)$$

And as $N\lambda^{t+1}=o\left((g+h)^{2+\beta}\right)$ then we get

$$\mathbb{E}\left(\hat{u}_i\right)=m(x)+\frac{1}{2}m''(x)V_K h^2+\frac{1}{2}m''(x)V_L\lambda g^2+O\left((g+h)^{2+\beta}\right)$$

$$\operatorname{bias}\left(\hat{u}_i\right)=\frac{h^2}{2}m''\left(x_i\right)V_K+\lambda\frac{g^2}{2}m''\left(x_i\right)V_L+O\left((g+h)^{2+\beta}\right)$$

and i) follows.

ii) We have

$$\begin{aligned}
\operatorname{Cov}\left(\hat{\mathbf{u}}\right) &= \left(P-\lambda\Lambda\right)^{-1}\operatorname{Cov}\left(\hat{\mathbf{u}}_{PC}\right)\left(P-\lambda\Lambda\right)^{-1}\\
&= \left[P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n+O\left(\lambda^{t+1}\right)\right]\operatorname{Cov}\left(\hat{\mathbf{u}}_{PC}\right)\\
&\quad\cdot\left[P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n+O\left(\lambda^{t+1}\right)\right]\\
&= \left[P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n\right]\operatorname{Cov}\left(\hat{\mathbf{u}}_{PC}\right)\left[P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n\right]+O\left(\lambda^{t+1}\right)
\end{aligned}$$

and the the variance of $\hat{u}_i$ is given by

$$\operatorname{Var}\left(\hat{u}_i\right)=\left(\left(P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n\right)\operatorname{Cov}\left(\hat{\mathbf{u}}_{PC}\right)\left(P^{-1}\sum_{n=0}^{t}\lambda^n\left(\Lambda P^{-1}\right)^n\right)\right)_{ii}+O\left(\lambda^{t+1}\right)$$

Now,

$$\left(\left(P^{-1}\sum_{n=0}^{t}\lambda^{n}\left(\Lambda P^{-1}\right)^{n}\right)\mathrm{Cov}\left(\hat{\mathbf{u}}_{PC}\right)\left(P^{-1}\sum_{n=0}^{t}\lambda^{n}\left(\Lambda P^{-1}\right)^{n}\right)\right)_{ii}$$

$$= \sum_{p=1}^{t}\sum_{q=1}^{t}\frac{1}{N^{p+q}}\sum_{k,i_1,\dots,i_{p-1}}\sum_{l,j_1,\dots,j_{q-1}}\frac{\lambda^{p+q}}{(1+\lambda)^{p+q+2}}\mathcal{L}_g\left(x_k, x_{i_1},\dots,x_{i_{p-1}},x_i\right)$$

$$\times \mathcal{L}_g\left(x_i, x_{j_1}, x_{j_2},\dots,x_{j_{q-1}},x_l\right)\mathrm{Cov}\left(\hat{m}_K\left(x_k,h\right),\hat{m}_K\left(x_l,h\right)\right)+O\left(\frac{1}{N^3h^3}\right)$$

$$+O\left(\frac{\lambda}{N^3g^2h}\right)$$

$$= (1-\theta)^2\sum_{p=1}^{t}\sum_{q=1}^{t}\theta^{p+q}\mathrm{Cov}\left(\hat{m}_p\left(x_i,h,g\right),\hat{m}_q\left(x_i,h,g\right)\right)+O\left(\frac{1}{N^3h^3}\right)$$

$$+O\left(\frac{\lambda}{N^3g^2h}\right)$$

$$= (1-\theta)^2\frac{f(0)}{N}\sum_{p=1}^{t}\sum_{q=1}^{t}\theta^{p+q}L_g^{*(p+q)}*K_h*K_h(0)+O\left(\frac{1}{N^2h^2}\right)+O\left(\frac{g+h}{N^3g^2h^2}\right)$$

$$+O\left(\frac{1}{N^3h^3}\right)+O\left(\frac{\lambda}{N^3g^2h}\right)$$

$$= (1-\theta)^2\frac{f(0)}{N}\sum_{p=1}^{t}\sum_{q=1}^{t}\theta^{p+q}L_g^{*(p+q)}*K_h*K_h(0)+O\left(\frac{1}{N^2h^2}\right)$$

$$\mathrm{Var}\left(\hat{u}_i\right) = (1-\theta)^2\frac{f(0)}{N}\sum_{p=1}^{t}\sum_{q=1}^{t}\theta^{p+q}L_g^{*(p+q)}*K_h*K_h(0)+O\left(\frac{1}{N^2h^2}\right)+O\left(\lambda^{t+1}\right)$$

As $\lambda^{t+1} = o\left(\frac{1}{N^2h^2}\right)$ then we get

$$\mathrm{Var}\left(\hat{u}_i\right) = (1-\theta)^2\frac{f(0)}{N}\sum_{p=1}^{t}\sum_{q=1}^{t}\theta^{p+q}L_g^{*(p+q)}*K_h*K_h(0)+O\left(\frac{1}{N^2h^2}\right)$$

As the Fourier transform of $L_g^{*(p+q)}*K_h*K_h(z)$ is $\widehat{L}_g^{p+q}(\omega)\,\widehat{K}_h^2(\omega)$, then we can apply the inverse Fourier transform (e.g. [5], Theorem 8.39), to get

$$\begin{aligned} L_g^{*(p+q)}*K_h*K_h(0) &= \frac{1}{2\pi}\int \widehat{L}_g^{p+q}(\omega)\,\widehat{K}_h^2(\omega)\,d\omega \\ &= \frac{1}{2\pi}\int \widehat{L}^{p+q}(g\omega)\,\widehat{K}^2(h\omega)\,d\omega \end{aligned}$$

Therefore

$$
\begin{aligned}
\operatorname{Var}(\hat{u}_i) &= (1-\theta)^2 \frac{f(0)}{N} \sum_{p=1}^{t} \sum_{q=1}^{t} \theta^{p+q} \frac{1}{2\pi} \int \widehat{L}^{p+q}(g\omega) \widehat{K}^2(h\omega)\, d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} (1-\theta)^2 \int \sum_{p=1}^{t} \sum_{q=1}^{t} \theta^{p+q} \widehat{L}^{p+q}(g\omega) \widehat{K}^2(h\omega)\, d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} \int \left((1-\theta) \sum_{p=0}^{t} \theta^{p} \widehat{L}^{p}(g\omega) \widehat{K}(h\omega)\right)^2 d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} \int \left((1-\theta) \sum_{p=0}^{t} \theta^{p} \widehat{L}^{p}(g\omega) \widehat{K}(h\omega)\right)^2 d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} \int \left((1-\theta) \frac{1-\theta^{t+1}\widehat{L}^{t+1}(g\omega)}{1-\theta\widehat{L}(g\omega)} \widehat{K}(h\omega)\right)^2 d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} \int \left((1-\theta) \frac{1}{1-\theta\widehat{L}(g\omega)} \widehat{K}(h\omega)\right)^2 d\,\omega + O\left(\theta^{t+1}\right) + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N} \int \left(\frac{\widehat{K}(h\omega)}{1+\lambda-\lambda\widehat{L}(g\omega)}\right)^2 d\,\omega + O\left(\frac{1}{N^2h^2}\right) \\
&= \frac{f(0)}{2\pi N h} \int \left(\frac{\widehat{K}(\omega)}{1+\lambda-\lambda\widehat{L}\left(\frac{g}{h}\omega\right)}\right)^2 d\,\omega + O\left(\frac{1}{N^2h^2}\right)
\end{aligned}
$$

$$
\operatorname{Var}(\hat{u}_i) = \frac{f(0)}{Nh} Q\left(\frac{g}{h}, \lambda\right) + O\left(\frac{1}{N^2h^2}\right)
$$

Now, as $\widehat{L}$ is uniformly bounded, and using the Parseval's relation we have

$$
\begin{aligned}
Q\left(\frac{g}{h}, \lambda\right) &= \frac{1}{2\pi} \int \left(\frac{\widehat{K}(\omega)}{1+\lambda-\lambda\widehat{L}\left(\frac{g}{h}\omega\right)}\right)^2 d\,\omega \\
&= \frac{1}{2\pi} \int \widehat{K}^2(\omega)\, d\,\omega + O(\lambda) = Q_K + O(\lambda)
\end{aligned}
$$

iii) Combining the bias and variance expansion, we get

$$\begin{aligned}
\operatorname{mse}(\hat{u}_i) &= (\operatorname{bias}(\hat{u}_i))^2 + \operatorname{Var}(\hat{u}_i) \\
&= \frac{1}{4}\left[h^2 V_K + \lambda g^2 V_L\right]^2 (m''(x_i))^2 + \left[h^2 V_K + \lambda g^2 V_L\right] m''(x_i) O\left((g+h)^{2+\beta}\right) \\
&\quad + O\left((g+h)^{4+2\beta}\right) + \frac{f(0)}{Nh} Q\left(\frac{g}{h}, \lambda\right) + O\left(\frac{1}{N^2 h^2}\right)
\end{aligned}$$

As m'' is bounded, the assertion of iii) holds uniformly for all i satisfying $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$. □

For sake of simplicity in choosing the order of the tuning parameters h, g, λ well for getting a small mean-squared estimation error we use the parameters h, b, λ instead of h, g, λ where $g = bh$. Therefore the asymptotic mean-squared error can be written as

$$\operatorname{amse}(\hat{u}_i; h, b, \lambda) = \frac{h^4}{4}\left(V_K + \lambda b^2 V_L\right)^2 (m''(x_i))^2 + \frac{f(0)}{Nh} Q(b, \lambda)$$

Corollary 3.3.15. *Under the conditions of Theorem 3.3.14 and* $m''(x_i) \neq 0$*,*

$$h_{\text{mse}}(x_i; b, \lambda) = \left(\frac{f(0) Q(b, \lambda)}{(m''(x_i))^2 \left(V_K + \lambda b^2 V_L\right)^2}\right)^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{3.3.11}$$

is asymptotically optimal bandwidth at x_i *with respect to* mse *where* b, λ *are fixed. If* $h = h_{\text{mse}}(x_i; b, \lambda)$ *then*

$$\operatorname{amse}(\hat{u}_i; h, b, \lambda) = \frac{5}{4}\Big[f(0) Q(b, \lambda)\Big]^{\frac{4}{5}} \left[(m''(x_i))^2 \left(V_K + \lambda b^2 V_L\right)^2\right]^{\frac{1}{5}} N^{-\frac{4}{5}}$$

Proof. In the minimum of $\operatorname{amse}(\hat{u}_i; h, b, \lambda)$

$$\frac{\partial}{\partial h}\operatorname{amse}(\hat{u}_i; h, b, \lambda) = h^3 \left(V_K + \lambda b^2 V_L\right)^2 (m''(x_i))^2 - \frac{f(0)}{Nh^2} Q(b, \lambda) = 0$$

have to be fulfilled. This yields (3.3.11). □

Remark 3.3.1. If the errors are i.i.d. $(0, \sigma_\varepsilon^2)$ then the variance of the RLLS-estimate:

$$\mathrm{Var}\left(\hat{u}_i\right) \sim \frac{\sigma_\varepsilon^2}{Nh} Q\left(\frac{g}{h}, \lambda\right) \text{ as } N \to \infty$$

since in this case $f(0)$ will be equal to σ_ε^2. And the asymptotic mean squared error will be given by

$$\mathrm{amse}\left(\hat{u}_i; h, b, \lambda\right) = \frac{h^4}{4}\left[V_K + \lambda b^2 V_L\right]^2 \left[m''(x_i)\right]^2 + \frac{\sigma_\varepsilon^2}{Nh} Q\left(b, \lambda\right)$$

where $b = \frac{g}{h}$. And therefore

$$h_{\mathrm{mse}}\left(x_i; b, \lambda\right) = \left(\frac{\sigma_\varepsilon^2 Q\left(b, \lambda\right)}{\left(m''\left(x_i\right)\right)^2 \left(V_K + \lambda b^2 V_L\right)^2}\right)^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{3.3.12}$$

will be, in the case of i.i.d. errors, the asymptotically optimal bandwidth at x with respect to mse where b, λ are fixed.

The effect of the errors correlation can easily be seen from (3.3.7) if we note that $f\left(0\right) = \sigma_\varepsilon^2 + 2R$ where $R = \sum_{t=1}^{\infty} r_t$. The presence of the the additional term R in the amse has important implications for the correct choice of bandwidth. If $R > 0$, implying that the errors correlation is (mostly) positive, then the variance of $\hat{u}_i$ will be larger than in the corresponding uncorrelated case. The amse is therefore minimized by a value for the bandwidth h that is larger than in the uncorrelated case. Conversely, if $R < 0$, the amse-optimal bandwidth is smaller than in the uncorrelated case.

In practice, we are much more frequently confronted with positively dependent errors, i. e. $R > 0$.

Chapter 4

The General Regularized Interpolation Estimates (GRI-Estimates)

4.1 Setup of the Problem

Assume that data $Y_1, ..., Y_N$ follow the deterministic equidistant design model (2.1.3):

$$\left.\begin{array}{l} Y_j = m(x_j) + \varepsilon_j, \qquad j = 1, ..., N \\ m : [0,1] \to \mathbb{R} \\ x_j = \frac{j}{N}, \qquad j = 1, ..., N \end{array}\right\}$$

where

Assumption E1: The errors $\varepsilon_j,\ j = 1, ..., N$ are part of a stationary time series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with $\mathbb{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$ and autocovariances

$$r_t = \mathrm{Cov}(\varepsilon_s, \varepsilon_{t+s}), \qquad s, t \in \mathbb{Z}$$

i.e. the autocovariance sequence independent of N.

Let $u_j := m(x_j),\ j = 1, \ldots, N$ and $\mathbf{u} = (u_1, \ldots, u_N)^{\mathrm{T}}$ We are interested in this chapter in getting an estimate of $\mathbf{u}$ using the general approach for image denoising proposed by Mrázek et. al. [14]. Using that approach we consider $\omega_D(x^2) = \delta(x^2)$ where δ is dirac function. Then we get the problem:

$$Q(\mathbf{u}) = \sum_{j=1}^{N} \rho_1(u_j - Y_j) + \frac{\lambda}{2} \sum_{i,j=1}^{N} \rho_2(u_i - u_j) L_g(x_i - x_j) = \min_{\mathbf{u}}! \qquad (4.1.1)$$

where the kernel L satisfies the assumptions L1-L5 from the previous chapter, the bandwidth $g \geq 0$, and the regularization parameter $\lambda \geq 0$. we assume that ρ_1, ρ_2 are symmetric twice differentiable functions with $\rho_1'(0) = \rho_2'(0) = 0$.

The solution $\hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{n}_N)^T$ of problem 4.1.1 estimates $\mathbf{u} = (u_1, \ldots, u_N)^T$ and is called *General regularized interpolation estimate*, abbreviated as the GRI-estimate.

$$\begin{aligned}
\frac{\partial Q}{\partial u_k}(\mathbf{u}) &= \rho_1'(u_k - Y_k) + \frac{\lambda}{2} \sum_{j=1}^{N} \rho_2'(u_k - u_j) L_g(x_k - x_j) - \frac{\lambda}{2} \sum_{i=1}^{N} \rho_2'(u_i - u_k) L_g(x_i - x_k) \\
&= \rho_1'(u_k - Y_k) + \frac{\lambda}{2} \sum_{j=1}^{N} \left[\rho_2'(u_k - u_j) - \rho_2'(u_j - u_k)\right] L_g(x_k - x_j) \\
&= \rho_1'(u_k - Y_k) + \frac{\lambda}{2} \sum_{j=1}^{N} \left[\rho_2'(u_k - u_j) + \rho_2'(u_k - u_j)\right] L_g(x_k - x_j)
\end{aligned}$$

by symmetry of ρ_2 as it implies the antisymmetry of ρ_2'. And that gives

$$\frac{\partial Q}{\partial u_k}(\mathbf{u}) = \rho_1'(u_k - Y_k) + \lambda \sum_{j=1}^{N} \rho_2'(u_k - u_j) L_g(x_k - x_j)$$

$$\frac{\partial^2 Q}{\partial u_k^2}(\mathbf{u}) = \rho_1''(u_k - Y_k) + \lambda \sum_{\substack{j=1 \\ j \neq k}}^{N} \rho_2''(u_k - u_j) L_g(x_k - x_j)$$

for $l \neq k$ we have:

$$\frac{\partial^2 Q}{\partial u_l \partial u_k}(\mathbf{u}) = -\lambda \rho_2''(u_k - u_l) L_g(x_k - x_l)$$

Notations:

$$Q_k(\mathbf{u}) := \frac{\partial Q}{\partial u_k}(\mathbf{u}), \quad Q_{kk}(\mathbf{u}) := \frac{\partial^2 Q}{\partial u_k^2}(\mathbf{u}), \quad Q_{kl}(\mathbf{u}) = \frac{\partial^2 Q}{\partial u_l \partial u_k}(\mathbf{u})$$

By the symmetry of ρ_2 and of the kernel L we have that $Q_{kl} = Q_{lk}$.

The gradient, then, is given by

$$\nabla Q(\mathbf{u}) = \left(Q_1(\mathbf{u}), \ldots, Q_N(\mathbf{u})\right)^T$$

and the Hassian by

$$\nabla^2 Q(\mathbf{u}) = \left(Q_{kl}(\mathbf{u})\right)_{1 \le k,l \le N}$$

A special case of the GRI-estimate is the *quadratic regularized interpolation estimate* where $\rho_1(u) = u^2$, $\rho_2(u) = u^2$. We will study this estimate in the following. The target function will be:

$$Q(\mathbf{u}) = \sum_{j=1}^{N} (u_j - Y_j)^2 + \frac{\lambda}{2} \sum_{i,j=1}^{N} (u_i - u_j)^2 L_g(x_i - x_j)$$

$$\begin{aligned}
\frac{\partial Q}{\partial u_k}(\mathbf{u}) &= u_k - Y_k + \lambda \sum_{j=1}^{N} (u_k - u_j) L_g(x_k - x_j) \\
&= u_k - Y_k + \lambda \sum_{j=1}^{N} u_k L_g(x_k - x_j) - \lambda \sum_{j=1}^{N} u_j L_g(x_k - x_j) \\
&= u_k - Y_k + N\lambda u_k p_L(x_k, g) - N\lambda \sum_{j=1}^{N} u_j \Lambda_{kj}
\end{aligned}$$

Let $\bar{\lambda} := N\lambda$ then

$$\nabla Q(\mathbf{u}) = \bar{P}\mathbf{u} - \bar{\lambda}\Lambda\mathbf{u} - \mathbf{Y}$$

where $\bar{P} := \operatorname{diag}\left(\bar{p}_\lambda(x_k, g)\right)_k$ s.t. $\bar{p}_\lambda(x, g) = 1 + \bar{\lambda} p_L(x, g)$.

$$\nabla Q(\mathbf{u}) = \left(\bar{P} - \bar{\lambda}\Lambda\right)\mathbf{u} - \mathbf{Y}$$

Now $\nabla Q(\hat{\mathbf{u}}) = 0$ gives the Regularized Interpolation Estimate

$$\hat{\mathbf{u}} = \left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1} \mathbf{Y} \tag{4.1.2}$$

4.2 The Bias

The Assumptions M1-M2 and E2 from the previous chapter will be assumed. Assume also that $g, \lambda \to 0$ as $N \to \infty$. If we assume $\bar{\lambda} = N\lambda \to 0$ we have similar to the RLLS-estimate

$$\begin{aligned}
\left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1} &= \left[\left(I - \bar{\lambda}\Lambda\bar{P}^{-1}\right)\bar{P}\right]^{-1} \\
&= \bar{P}^{-1}\left(I - \bar{\lambda}\Lambda\bar{P}^{-1}\right)^{-1} \\
&= \bar{P}^{-1}\left[\sum_{n=0}^{t} \bar{\lambda}^n \left(\Lambda\bar{P}^{-1}\right)^n\right] + O\left(\bar{\lambda}^{t+1}\left\|\bar{P}^{-1}\right\|\left\|\left(\Lambda\bar{P}^{-1}\right)^{t+1}\right\|\right)
\end{aligned}$$

where the remainder term is of order $\bar{\lambda}^{t+1}$ by Lemma 3.3.12, and by boundedness of $\left\|\bar{P}\right\|$, and $\left\|\bar{P}^{-1}\right\|$. Therefore we have

$$\begin{aligned}
\mathbb{E}\hat{\mathbf{u}} = \left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1}\mathbf{Y} &= \left[\bar{P}^{-1}\sum_{n=0}^{t} \bar{\lambda}^n \left(\Lambda\bar{P}^{-1}\right)^n + O\left(\bar{\lambda}^{t+1}\right)\right]\mathbb{E}\mathbf{Y} \\
&= \left[\bar{P}^{-1} + \bar{P}^{-1}\sum_{n=1}^{t} \bar{\lambda}^n \left(\bar{\lambda}\bar{P}^{-1}\right)^n + O\left(\bar{\lambda}^{t+1}\right)\right]\mathbb{E}\mathbf{Y} \qquad (4.2.1)
\end{aligned}$$

But

$$\Lambda\bar{P}^{-1} = \left(\sum_{k} \Lambda_{ik}\bar{P}^{-1}_{kj}\right)_{ij} = \left(\frac{\Lambda_{ij}}{\bar{P}_{jj}}\right)_{ij}$$

$$\left(\Lambda\bar{P}^{-1}\right)^n = \left(\sum_{j_1,\ldots,j_{n-1}} \frac{\Lambda_{ij_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j}}{\bar{P}_{jj}}\right)_{ij}$$

$$\bar{P}^{-1}\sum_{n=1}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n = \left(\bar{P}^{-1}_{ii}\sum_{n=1}^{t}\bar{\lambda}^n\sum_{j_1,\ldots,j_{n-1}} \frac{\Lambda_{ij_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{n-1}j}}{\bar{P}_{jj}}\right)_{ij}$$

and, as $\mathbb{E}Y_i = m(x_i)$, Y_i is bounded in probability uniformly for all $i \in \{1, \dots, N\}$. Therefore formula (4.2.1) can, coordinate-wise, be written as

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \sum_k \bar{P}_{ik}^{-1}\mathbb{E}Y_k + \sum_{j_n} \bar{P}_{ii}^{-1} \sum_{n=1}^{t} \lambda^n \sum_{j_1,\dots,j_{n-1}} \frac{\Lambda_{ij_1}}{\bar{P}_{j_1j_1}} \frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}} \cdots \frac{\Lambda_{j_{n-1}j_n}}{\bar{P}_{j_nj_n}} \mathbb{E}Y_{j_n} + O\left(N\bar{\lambda}^{t+1}\right) \\ &= \frac{1}{\bar{P}_{ii}}\mathbb{E}Y_i + \frac{1}{\bar{P}_{ii}} \sum_{n=1}^{t} \lambda^n \sum_{j_1,\dots,j_n} \frac{\Lambda_{ij_1}}{\bar{P}_{j_1j_1}} \frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}} \cdots \frac{\Lambda_{j_{n-1}j_n}}{\bar{P}_{j_nj_n}} \mathbb{E}Y_{j_n} + O\left(N\bar{\lambda}^{t+1}\right)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \frac{1}{\bar{p}_\lambda(x_i,g)} \sum_{n=1}^{t} \frac{\lambda^n}{N^n} \sum_{j_1,\dots,j_n} \frac{L_g(x_i - x_{j_1})}{\bar{p}_\lambda(x_{j_1},g)} \frac{L_g(x_{j_1} - x_{j_2})}{\bar{p}_\lambda(x_{j_2},g)} \cdots \frac{L_g\left(x_{j_{n-1}} - x_{j_n}\right)}{\bar{p}_\lambda(x_{j_n},g)} \mathbb{E}Y_{j_n} \\ &\quad + \frac{1}{\bar{p}_\lambda(x_i,g)}\mathbb{E}Y_i + O\left(N\bar{\lambda}^{t+1}\right) \qquad (4.2.2)\end{aligned}$$

where $\bar{p}_\lambda(x,g) := \bar{P}_{ii} = 1 + \bar{\lambda}p_L(x,g)$.

As $(t+1)g \le x_i \le 1-(t+1)g$, and L_g has support $[-g,g]$, we have $g \le x_{j_n} \le 1-g$ for all j_n, $n = 1, ..., t$, corresponding to the non-vanishing terms in the n-fold sum of (4.2.2). Therefore, we may apply Lemma 3.3.13 to replace $\bar{p}_\lambda(x_{j_n},g)$ by $1+\lambda$ in (4.2.2), and we get

$$\begin{aligned}\mathbb{E}\hat{u}_i &= \frac{1}{1+\bar{\lambda}} \sum_{n=1}^{t} \frac{\bar{\lambda}^n}{N^n} \sum_{j_1,\dots,j_n} \frac{L_g(x_i - x_{j_1})}{1+\bar{\lambda}} \frac{L_g(x_{j_1} - x_{j_2})}{1+\bar{\lambda}} \cdots \frac{L_g\left(x_{j_{n-1}} - x_{j_n}\right)}{1+\bar{\lambda}} \mathbb{E}Y_{j_n} \\ &\quad + \frac{1}{1+\bar{\lambda}}\mathbb{E}Y_i + R_N\end{aligned}$$

where $R_N = O\left(\bar{\lambda}^{t+1}\right) + O\left(\frac{\bar{\lambda}}{N^2g^2}\right)$.

Let

$$\hat{\mu}_n(x,g) := \frac{1}{N^n} \sum_{j_1,\dots,j_n} L_g(x - x_{j_1}) L_g(x_{j_1} - x_{j_2}) \dots L_g\left(x_{j_{n-1}} - x_{j_n}\right) Y_{j_n}, \text{ for } n \ge 1, x \in [0,1]$$

and $\hat{\mu}_0(x,g) := Y_{\mathfrak{i}(x)}$ where

$$\mathfrak{i}(x) = i; \text{ if } x_{i-1} \le x \le x_i \text{ for } x \in [0,1) \text{ and } \mathfrak{i}(1) = N$$

then we have

$$\begin{aligned}
\mathbb{E}\hat{u}_i &= \frac{1}{1+\bar{\lambda}}\sum_{n=1}^{t}\frac{\bar{\lambda}^n}{\left(1+\bar{\lambda}\right)^n}\frac{1}{N^n}\sum_{j_1,\ldots,j_n}L_g\left(x_i-x_{j_1}\right)L_g\left(x_{j_1}-x_{j_2}\right)\ldots L_g\left(x_{j_{n-1}}-x_{j_n}\right)\mathbb{E}Y_{j_n}\\
&\quad+\frac{1}{1+\bar{\lambda}}\mathbb{E}Y_i+R_N\\
&= \frac{1}{1+\bar{\lambda}}\left[\sum_{n=1}^{t}\frac{\bar{\lambda}^n}{\left(1+\bar{\lambda}\right)^n}\mathbb{E}\hat{\mu}_n\left(x_i,g\right)+\mathbb{E}\hat{\mu}_0\left(x_i,g\right)\right]+R_N\\
&= \left(1-\bar{\theta}\right)\sum_{n=0}^{t}\bar{\theta}^n\mathbb{E}\hat{\mu}_n\left(x_i,g\right)+R_N
\end{aligned}$$

where $\bar{\theta} := \frac{\bar{\lambda}}{1+\bar{\lambda}}$.

Proposition 4.2.1. *Assume the deterministic equidistant design model (2.1.3), E1-E2, L1-L5, M1-M2, and $N \to \infty$, s.t. $g \to 0$. Then for all $p \geq 1$ we have, uniformly for all $x \in [pg, 1-pg]$,*

$$\mathbb{E}\left(\hat{\mu}_p\left(x,g\right)\right) = m(x) + \frac{p}{2}m''(x)V_L g^2 + O\left(g^{2+\beta}\right) + O\left(\frac{1}{N^2g^2}\right)$$

Proof. The assertion can be proved by the same argument of the proof of Proposition 3.3.9. □

Now,

$$\mathbb{E}\hat{u}_i = \left(1-\bar{\theta}\right)\sum_{n=0}^{t}\bar{\theta}^n\mathbb{E}\hat{\mu}_n\left(x_i,g\right)+R_N$$

where $R_N = O\left(\bar{\lambda}^{t+1}\right) + O\left(\frac{\bar{\lambda}}{N^2g^2}\right)$. Using Proposition 4.2.1 we get

$$\mathbb{E}\left(\hat{u}_i\right) = \left(1-\bar{\theta}\right)m(x_i) + \left(1-\bar{\theta}\right)\sum_{n=1}^{t}\bar{\theta}^n\left[m(x_i)+\frac{n}{2}m''(x_i)V_L g^2 + O\left(g^{2+\beta}\right)+O\left(\frac{1}{N^2g^2}\right)\right]+R_N$$

But we have

$$\left(1-\bar{\theta}\right)\sum_{n=1}^{t}\bar{\theta}^n = \bar{\theta}-\bar{\theta}^{t+1} = O\left(\bar{\theta}\right) = O\left(\bar{\lambda}\right)$$

$$\left(1-\bar{\theta}\right)\sum_{n=1}^{t} n\bar{\theta}^{n} = \frac{\bar{\theta}-\left(t+1\right)\bar{\theta}^{t+1}+t\bar{\theta}^{t+2}}{1-\bar{\theta}} = \bar{\lambda}+O\left(\bar{\theta}^{t+1}\right) = \bar{\lambda}+O\left(\bar{\lambda}^{t+1}\right)$$

That gives

$$\begin{aligned}
\mathbb{E}\left(\hat{u}_i\right) &= \left(1-\bar{\theta}^{t+1}\right)m(x_i)+O\left(\bar{\lambda}\right)\left[O\left(g^{2+\beta}\right)+O\left(\frac{1}{N^2g^2}\right)\right] \\
&\quad +\left[\bar{\lambda}+O\left(\bar{\lambda}^{t+1}\right)\right]\frac{1}{2}m''(x_i)V_Lg^2+R_N \\
&= m(x_i)+\frac{1}{2}m''(x_i)V_L\bar{\lambda}g^2+O\left(\bar{\lambda}g^{2+\beta}\right)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right) \\
&\quad +O\left(\bar{\lambda}^{t+1}\right)\left[m(x_i)+\frac{1}{2}m''(x_i)V_Lg^2\right]+R_N \\
&= m(x_i)+\frac{1}{2}m''(x_i)V_L\bar{\lambda}g^2+O\left(\bar{\lambda}g^{2+\beta}\right)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)+O\left(\bar{\lambda}^{t+1}\right)+R_N \\
&= m(x_i)+\frac{1}{2}m''(x_i)V_L\bar{\lambda}g^2+O\left(\bar{\lambda}g^{2+\beta}\right)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)+O\left(N\bar{\lambda}^{t+1}\right)
\end{aligned}$$

If we assume $N^2g^{4+\beta}\to\infty$ then we have

$$\mathbb{E}\left(\hat{u}_i\right) = m(x_i)+\frac{1}{2}m''(x_i)V_L\bar{\lambda}g^2+O\left(\bar{\lambda}g^{2+\beta}\right)+O\left(N\bar{\lambda}^{t+1}\right)$$

and if we assume that $N\bar{\lambda}^{t+1}=o\left(\bar{\lambda}g^{2+\beta}\right)$ then we have

$$\mathbb{E}\left(\hat{u}_i\right) = m(x_i)+\frac{1}{2}m''(x_i)V_L\bar{\lambda}g^2+O\left(\bar{\lambda}g^{2+\beta}\right)$$

$$\text{bias}\left(\hat{u}_i\right) = \bar{\lambda}\,\frac{g^2}{2}m''(x_i)V_L+O\left(\bar{\lambda}g^{2+\beta}\right)$$

4.3 The Variance

Proposition 4.3.1. *Assume the deterministic equidistant design model (2.1.3), E1-E2, L1-L5, M1-M2, and $N\to\infty$, s.t. $g\to 0$. Then for all $p,q>0$ we have, uniformly for all $x\in[pg,1-pg]$, and $\bar{x}\in[qg,1-qg]$,*

$$\text{Cov}\left(\hat{\mu}_p(x,g),\hat{\mu}_q(\bar{x},g)\right) = \frac{f(0)}{N}L_g^{*(p+q)}(x-\bar{x})+O\left(\frac{1}{N^3g^3}\right)$$

and

$$\mathrm{Cov}\left(\hat{\mu}_p(x,g),\hat{\mu}_0(\bar{x},g)\right) = \frac{f(0)}{N} L_g^{*(p)}(x - x_{\mathrm{i}(\bar{x})}) + O\left(\frac{1}{N^3 g^3}\right)$$

Proof. The assertion can be proved by the same argument of the proof of Proposition 3.3.10. □

Now,

$$\begin{aligned}
\mathrm{Cov}\left(\hat{\mathbf{u}}\right) &= \left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1} \mathrm{Cov}\left(\mathbf{Y}\right)\left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1} \\
&= \left[\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n + O\left(\bar{\lambda}^{t+1}\right)\right]\mathrm{Cov}\left(\mathbf{Y}\right)\left[P^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n + O\left(\bar{\lambda}^{t+1}\right)\right] \\
&= \left[\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right]\mathrm{Cov}\left(\mathbf{Y}\right)\left[\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right] + O\left(\bar{\lambda}^{t+1}\right)
\end{aligned}$$

$$\mathrm{Var}\left(\hat{u}_i\right) = \left(\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)\mathrm{Cov}\left(\mathbf{Y}\right)\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)\right)_{ii} + O\left(\bar{\lambda}^{t+1}\right)$$

But,

$$\begin{aligned}
&\left(\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)\mathrm{Cov}\left(\mathbf{Y}\right)\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)\right)_{ii} \\
&\quad = \sum_{k,l}\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)_{i,k}\mathrm{Cov}\left(Y_k,Y_l\right)\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^n\left(\Lambda\bar{P}^{-1}\right)^n\right)_{l,i} \\
&\quad = \sum_{k,l}\bar{P}_{ik}^{-1}\mathrm{Cov}\left(Y_k,Y_l\right)\bar{P}_{li}^{-1} + \sum_{k,l}\bar{P}_{ik}^{-1}\mathrm{Cov}\left(Y_k,Y_l\right)\left[\bar{P}_{ll}^{-1}\sum_{q=1}^{t}\bar{\lambda}^q\sum_{j_1,\ldots,j_{q-1}}\frac{\Lambda_{lj_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{q-1}i}}{\bar{P}_{ii}}\right] \\
&\qquad + \sum_{k,l}\left[\bar{P}_{ii}^{-1}\sum_{p=1}^{t}\bar{\lambda}^p\sum_{i_1,\ldots,i_{p-1}}\frac{\Lambda_{ii_1}}{\bar{P}_{i_1i_1}}\frac{\Lambda_{i_1i_2}}{\bar{P}_{i_2i_2}}\cdots\frac{\Lambda_{i_{p-1}k}}{\bar{P}_{kk}}\right]\mathrm{Cov}\left(Y_k,Y_l\right)\bar{P}_{li}^{-1} \\
&\qquad + \sum_{k,l}\left[\bar{P}_{ii}^{-1}\sum_{p=1}^{t}\bar{\lambda}^p\sum_{i_1,\ldots,i_{p-1}}\frac{\Lambda_{ii_1}}{\bar{P}_{i_1i_1}}\frac{\Lambda_{i_1i_2}}{\bar{P}_{i_2i_2}}\cdots\frac{\Lambda_{i_{p-1}k}}{\bar{P}_{kk}}\right]\mathrm{Cov}\left(Y_k,Y_l\right) \\
&\qquad\quad \left[\bar{P}_{ll}^{-1}\sum_{q=1}^{t}\bar{\lambda}^q\sum_{j_1,\ldots,j_{q-1}}\frac{\Lambda_{lj_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{q-1}i}}{\bar{P}_{ii}}\right]
\end{aligned}$$

$$\left(\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^{n}\left(\Lambda\bar{P}^{-1}\right)^{n}\right)\mathrm{Cov}\left(\mathbf{Y}\right)\left(\bar{P}^{-1}\sum_{n=0}^{t}\bar{\lambda}^{n}\left(\Lambda\bar{P}^{-1}\right)^{n}\right)\right)_{ii}$$

$$= \bar{P}_{ii}^{-2}\mathrm{Cov}\left(Y_i,Y_i\right)+\sum_{l}\bar{P}_{ii}^{-1}\mathrm{Cov}\left(Y_i,Y_l\right)\left[\bar{P}_{ll}^{-1}\sum_{q=1}^{t}\bar{\lambda}^{q}\sum_{j_1,\ldots,j_{q-1}}\frac{\Lambda_{lj_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{q-1}i}}{\bar{P}_{ii}}\right]$$

$$+\sum_{k}\left[\bar{P}_{ii}^{-1}\sum_{p=1}^{t}\bar{\lambda}^{p}\sum_{i_1,\ldots,i_{p-1}}\frac{\Lambda_{ii_1}}{\bar{P}_{i_1i_1}}\frac{\Lambda_{i_1i_2}}{\bar{P}_{i_2i_2}}\cdots\frac{\Lambda_{i_{p-1}k}}{\bar{P}_{kk}}\right]\mathrm{Cov}\left(Y_k,Y_i\right)\bar{P}_{ii}^{-1}$$

$$+\sum_{k,l}\left[\bar{P}_{ii}^{-1}\sum_{p=1}^{t}\bar{\lambda}^{p}\sum_{i_1,\ldots,i_{p-1}}\frac{\Lambda_{ii_1}}{\bar{P}_{i_1i_1}}\frac{\Lambda_{i_1i_2}}{\bar{P}_{i_2i_2}}\cdots\frac{\Lambda_{i_{p-1}k}}{\bar{P}_{kk}}\right]\mathrm{Cov}\left(Y_k,Y_l\right)$$

$$\left[\bar{P}_{ll}^{-1}\sum_{q=1}^{t}\bar{\lambda}^{q}\sum_{j_1,\ldots,j_{q-1}}\frac{\Lambda_{lj_1}}{\bar{P}_{j_1j_1}}\frac{\Lambda_{j_1j_2}}{\bar{P}_{j_2j_2}}\cdots\frac{\Lambda_{j_{q-1}i}}{\bar{P}_{ii}}\right]$$

$$=\left(\frac{1}{1+\bar{\lambda}}+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)\right)^2\mathrm{Var}\left(Y_i\right)$$

$$+\sum_{q=1}^{t}\bar{\lambda}^{q}\left(\frac{1}{1+\bar{\lambda}}+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)\right)^{q+2}\sum_{l,j_1,\ldots,j_{q-1}}\frac{1}{N^q}\mathcal{L}_g\left(x_l,x_{j_1},x_{j_2},\ldots,x_{j_{q-1}},x_i\right)\mathrm{Cov}\left(Y_i,Y_l\right)$$

$$+\sum_{p=1}^{t}\bar{\lambda}^{p}\left(\frac{1}{1+\bar{\lambda}}+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)\right)^{p+2}\sum_{k,i_1,\ldots,i_{p-1}}\frac{1}{N^p}\mathcal{L}_g\left(x_i,x_{i_1},x_{i_2},\ldots,x_{i_{p-1}},x_k\right)\mathrm{Cov}\left(Y_k,Y_i\right)$$

$$+\sum_{p=1}^{t}\sum_{q=1}^{t}\bar{\lambda}^{p+q}\left(\frac{1}{1+\bar{\lambda}}+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)\right)^{p+q+2}\sum_{k,i_1,\ldots,i_{p-1}}\sum_{l,j_1,\ldots,j_{q-1}}\frac{1}{N^{p+q}}\mathcal{L}_g\left(x_i,x_{i_1},\ldots,x_{i_{p-1}},x_k\right)$$

$$\mathcal{L}_g\left(x_l,x_{j_1},\ldots,x_{j_{q-1}},x_i\right)\mathrm{Cov}\left(Y_k,Y_l\right)$$

$$=\left(1-\bar{\theta}\right)^2\sum_{p=0}^{t}\sum_{q=0}^{t}\bar{\theta}^{p+q}\mathrm{Cov}\left(\hat{\mu}_p\left(x_i,g\right),\hat{\mu}_q\left(x_i,g\right)\right)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)$$

$$=\left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2+\left(1-\bar{\theta}\right)^2\frac{f(0)}{N}\sum_{\substack{p,q=0\\p+q\neq 0}}^{t}\bar{\theta}^{p+q}L_g^{*(p+q)}(0)+O\left(\frac{1}{N^3g^3}\right)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)$$

$$=\left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2+\left(1-\bar{\theta}\right)^2\frac{f(0)}{N}\sum_{\substack{p,q=0\\p+q\neq 0}}^{t}\bar{\theta}^{p+q}L_g^{*(p+q)}(0)+O\left(\frac{\bar{\lambda}}{N^2g^2}\right)$$

Thus the variance is given by

$$\mathrm{Var}\left(\hat{u}_i\right) = \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\frac{f(0)}{N}\sum_{\substack{p,q=0\\p+q\neq 0}}^{t}\bar{\theta}^{p+q}L_g^{*(p+q)}(0) + S_N$$

where $S_N = O\left(\frac{\bar{\lambda}}{N^2g^2}\right) + O\left(\bar{\lambda}^{t+1}\right)$

Applying the inverse Fourier transform (e.g. [5], Theorem 8.39),

$$\begin{aligned} L_g^{*(p+q)}(0) &= \frac{1}{2\pi}\int \widehat{L}_g^{p+q}(\omega)\,d\,\omega \\ &= \frac{1}{2\pi}\int \widehat{L}^{p+q}(g\omega)\,d\,\omega \end{aligned}$$

Then the variance can be written as

$$\begin{aligned} \mathrm{Var}\left(\hat{u}_i\right) &= \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\frac{f(0)}{2\pi N}\sum_{\substack{p,q=0\\p+q\neq 0}}^{t}\bar{\theta}^{p+q}\int \widehat{L}^{p+q}(g\omega)\,d\,\omega + S_N \\ &= \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\frac{f(0)}{2\pi N}\int\sum_{\substack{p,q=0\\p+q\neq 0}}^{t}\bar{\theta}^{p+q}\widehat{L}^{p+q}(g\omega)\,d\,\omega + S_N \\ &= \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\frac{f(0)}{2\pi N}\int\left[\left(\sum_{p=0}^{t}\bar{\theta}^{p}\widehat{L}^{p}(g\omega)\right)^2 - 1\right]d\,\omega + S_N \\ &= \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\frac{f(0)}{2\pi N}\int\left[\left(\frac{1-\bar{\theta}^{t+1}\widehat{L}^{t+1}(g\omega)}{1-\bar{\theta}\widehat{L}(g\omega)}\right)^2 - 1\right]d\,\omega + S_N \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\left(\hat{u}_i\right) = \left(1-\bar{\theta}\right)^2\sigma_\varepsilon^2 + \left(1-\bar{\theta}\right)^2\bar{\theta}\frac{f(0)}{2\pi Ng}\int\frac{2\widehat{L}(\omega)}{\left(1-\bar{\theta}\widehat{L}(\omega)\right)^2}d\,\omega \\ + O\left(\frac{\bar{\theta}^2}{Ng}\right) + O\left(\frac{\bar{\lambda}}{N^2g^2}\right) + O\left(\bar{\lambda}^{t+1}\right) \end{aligned}$$

we note that:

$$\lim_{N\to\infty}\mathrm{Var}\left(\hat{u}_i\right) = \sigma_\varepsilon^2 \neq 0$$

i.e. the estimate is not consistent.

4.4 Is the QRIE Equivalent to a Spline Smoother?

Note that our estimate can be written in the form (2.2.1)

$$\hat{u}_i = \frac{1}{N}\sum_{j=1}^{N} W_{N,i,j}^{(g,\lambda)} \cdot Y_j$$

where $W_{N,i,j}^{(g,\lambda)} = N\left(\left(\bar{P} - \bar{\lambda}\Lambda\right)^{-1}\right)_{ij}$.

As the spline smoother, which relies on a quite similar approach as the QRIE, is asymptotically equivalent to a kernel estimate (Theorem 2.4.1), let us try to prove that our estimate is asymptotically equivalent to a spline smoother. so it should become asymptotically equivalent to a kernel estimate too.

Our estimator:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathbb{R}^N}{\operatorname{argmin}}\left\{\sum_{j=1}^{N}\left(u_j - Y_j\right)^2 + \frac{\lambda}{2}\sum_{i,j=1}^{N}\left(u_i - u_j\right)^2 L_g\left(x_i - x_j\right)\right\}$$

or equivalently

$$\hat{u}_i = \left(\underset{u\in C^2[0,1]}{\operatorname{argmin}}\left\{\sum_{j=1}^{N}\left(u(x_j) - Y_j\right)^2 + \frac{\lambda}{2}\sum_{i,j=1}^{N}\left(u(x_i) - u(x_j)\right)^2 L_g\left(x_i - x_j\right)\right\}\right)(x_i)$$

for $i = 1, \ldots, N$.

Lemma 4.4.1.

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\left(u(x_i) - u(x_j)\right)^2 L_g\left(x_i - x_j\right) = \int\int_{[0,1]^2}\left(u(x) - u(y)\right)^2 L_g\left(x - y\right)\, dxdy + O\left(\frac{1}{N^2 g^2}\right)$$

as $N \to \infty$, where u is a twice continuously differentiable function on $[0,1]$ and $x_i = \frac{i}{N}$,$i = 1, \ldots, N$.

Proof. The assertion follows from Lemma 3.3.7 with $g(x,y) = \left(u(x) - u(y)\right)^2 L_g\left(x - y\right)$ and $D = \left\{\,(x,y) \in [0,1]^2 \middle|\, |x - y| \leq g\right\}$.

$D_x = D_y = [0,1]$, $V = 2g - g^2$, $V_x = 1$, $V_y = 1$.
$\sup_{(x,y)\in D}\left|\frac{\partial g}{\partial x}(x,y)\right| = O\left(\frac{1}{g^2}\right)$, $\sup_{(x,y)\in D}\left|\frac{\partial^2 g}{\partial x^2}(x,y)\right| = O\left(\frac{1}{g^3}\right)$. □

From the last lemma we have

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\left(u(x_i)-u(x_j)\right)^2 L_g\left(x_i - x_j\right) \sim \iint_{[0,1]^2} \left(u(x)-u(y)\right)^2 L_g\left(x-y\right)\, dxdy$$

as $N \to \infty$.

The question now:

Is there exist for every kernel L a function $\varphi_L(g)$ satisfying that the integral $\int\int\left(u(x)-u(y)\right)^2 L_g\left(x-y\right)\, dxdy$ is asymptotically equivalent to $\varphi_L(g)\cdot\int\left(u''(x)\right)^2 dx$ as $g \to 0$ for all functions $u : [0,1] \to \mathbb{R}$ in $C^2[0,1]$?

Assume that we have a set of functions $u(x) = ax^2 + bx$, $a, b \in \mathbb{R}$. We have

$$\int_0^1 \left(u''(x)\right)^2\, dx = 4a^2$$

I will compute the integral:

$$\begin{aligned}
&\int_0^1\int_0^1 \left(u(x)-u(y)\right)^2 L_g\left(x-y\right)\, dxdy \\
&\quad= \int_0^1\int_0^1 \left(ax^2+bx-ay^2-by\right)^2 L_g\left(x-y\right)\, dxdy \\
&\quad= a^2\int_0^1\int_0^1 \left(x^2-y^2\right)^2 L_g\left(x-y\right)\, dxdy + 2ab\int_0^1\int_0^1 \left(x^2-y^2\right)\left(x-y\right) L_g\left(x-y\right)\, dxdy \\
&\qquad+ b^2\int_0^1\int_0^1 \left(x-y\right)^2 L_g\left(x-y\right)\, dxdy \\
&\quad= a^2 I_1 + 2ab I_2 + b^2 I_3
\end{aligned}$$

Compute I_1:

$$I_1 = \int_0^1 \int_0^1 \left(x^2 - y^2\right)^2 L_g\left(x-y\right)\,dxdy$$
$$= \int_0^1 \int_0^1 \left(x+y\right)^2 \left(x-y\right)^2 L_g\left(x-y\right)\,dxdy$$

using the substitution

$$r = x - y$$

$$t = x + y$$

$$D = \left\{(r,t) \in \mathbb{R}^2 \left| 0 \le t + r \le 2,\ 0 \le t - r \le 2\right.\right\}$$

we have

$$I_1 = \frac{1}{2}\iint_D t^2 r^2 L_g\left(r\right)\,drdt$$
$$= \int_0^1 \left(\int_r^{2+r} t^2 r^2 L_g\left(r\right) dt\right) dr$$
$$= \int_0^1 \left[\frac{t^3}{3}\right]_r^{2+r} r^2 L_g\left(r\right) dr$$
$$= \frac{1}{3}\int_0^1 \left(8 + 12r + 6r^2\right) r^2 L_g\left(r\right) dr$$
$$= \frac{8}{3}\int_0^1 r^2 L_g\left(r\right) dr + 4\int_0^1 r^3 L_g\left(r\right) dr + 2\int_0^1 r^4 L_g\left(r\right) dr$$
$$= \frac{4}{3} V_L g^2 + 2V_{L,3} g^3 + V_{L,4} g^4$$

where

$$V_{L,2} = V_L = 2\int_0^1 r^2 L\left(r\right) dr$$
$$V_{L,3} = 2\int_0^1 r^3 L\left(r\right) dr$$
$$V_{L,4} = 2\int_0^1 r^4 L\left(r\right) dr$$

Compute I_2:

$$\begin{aligned} I_2 &= \int_0^1 \int_0^1 (x+y)(x-y)^2 L_g(x-y)\, dxdy \\ &= \frac{1}{2} \iint_D tr^2 L_g(r)\, drdt \end{aligned}$$

$$\begin{aligned} I_2 &= \int_0^1 \left(\int_r^{2+r} tr^2 L_g(r)\, dt \right) dr \\ &= \int_0^1 \left[\frac{t^2}{2} \right]_r^{2+r} r^2 L_g(r)\, dr \\ &= \frac{1}{2} \int_0^1 (4+4r)\, r^2 L_g(r)\, dr \\ &= 2\int_0^1 r^2 L_g(r)\, dr + 2\int_0^1 r^3 L_g(r)\, dr \\ &= V_L g^2 + V_{L,3} g^3 \end{aligned}$$

Compute I_3:

$$I_3 = \int_0^1 \int_0^1 (x-y)^2 L_g(x-y)\, dxdy$$

using the substitution $r = x - y$ we get

$$\begin{aligned} I_3 &= \iint_{D'} r^2 L_g(r)\, drdy; \text{ where } D' = \left\{ (r,y) \in \mathbb{R}^2 \,|0 \le r+y \le 1,\ 0 \le y \le 1 \right\} \\ &= \int_{-1}^0 \left(\int_{-r}^1 r^2 L_g(r)\, dy \right) dr + \int_0^1 \left(\int_0^{1-r} r^2 L_g(r)\, dy \right) dr \\ &= \int_{-1}^0 (1+r)\, r^2 L_g(r)\, dr + \int_0^1 (1-r)\, r^2 L_g(r)\, dr \\ &= \int_0^1 (1-r)\, r^2 L_g(r)\, dr + \int_0^1 (1-r)\, r^2 L_g(r)\, dr \\ &= 2\int_0^1 r^2 L_g(r)\, dr - 2\int_0^1 r^3 L_g(r)\, dr \\ &= V_L g^2 - V_{L,3} g^3 \end{aligned}$$

Now,

$$\int_0^1 \int_0^1 (u(x)-u(y))^2 L_g(x-y)\, dxdy$$
$$= a^2 I_1 + 2abI_2 + b^2 I_3$$
$$= a^2\left(\frac{4}{3}V_L g^2 + 2V_{L,3}g^3 + V_{L,4}g^4\right) + 2ab\left(V_L g^2 + V_{L,3}g^3\right) + b^2\left(V_L g^2 - V_{L,3}g^3\right)$$

$$\int_0^1 \int_0^1 (u(x)-u(y))^2 L_g(x-y)\, dxdy$$
$$= \left(\frac{4}{3}a^2 + 2ab + b^2\right) V_L g^2 + \left(2a^2 + 2ab - b^2\right) V_{L,3}g^3 + a^2 V_{L,4}g^4$$
$$\sim \left(\frac{4}{3}a^2 + 2ab + b^2\right) V_L g^2 \text{ as } g \to 0 \text{ for all } a, b \in \mathbb{R}.$$

Now assume that there is a function $\varphi_L(g)$ satisfying that

$$\int\int (u(x)-u(y))^2 L_g(x-y)\, dxdy \sim \varphi_L(g) \cdot \int (u''(x))^2\, dx \text{ as } g \to 0$$

for all $a, b \in \mathbb{R}$.

Then we have

$$\left(\frac{4}{3}a^2 + 2ab + b^2\right) V_L g^2 \sim \varphi_L(g)\left(4a^2\right) \text{ as } g \to 0 \text{ for all } a, b \in \mathbb{R}$$

and that is contradiction. i.e. for all functions $u \in C^2[0,1]$, there is no function $\varphi_L(g)$ s.t.

$$\int\int (u(x)-u(y))^2 L_g(x-y)\, dxdy \sim \varphi_L(g) \cdot \int (u''(x))^2\, dx \text{ as } g \to 0$$

and we can conclude that our estimator is not asymptotically equivalent to the spline smoother.

Notation:

Note that in my example

$$\int (u'(x))^2\, dx = \frac{4}{3}a^2 + 2ab + b^2$$

i.e.

$$\int\int (u(x)-u(y))^2 L_g(x-y)\,dxdy \sim V_L g^2 \int (u'(x))^2\,dx \text{ as } g\to 0 \text{ for all } a,b\in\mathbb{R}$$

Chapter 5

Some simulation results

The finite sample properties of the RLLSE have been studied by simulation. After describing the simulation design we shall illustrate the results for some typical situations. Noise processes and regression functions varied across simulations. As noise processes we took processes normed to variance $\sigma_\varepsilon^2 = 1$:

- strict white noise (SWN),
- $AR(1)$-processes with coefficients $-0.9, 0.9$,
- $AR(3)$-process with coefficients $(0.2, 0.3, 0.4)$,

which were simulated by using the algorithm in remark 3.1.3.

As regression functions we mainly used

- $m_1(x) = a_1 + a_2 x + a_3 \varphi(a_4(x - a_5)), x \in [0,1]$ with $a_1 = 2$, $a_2 = 5$, $a_3 = \frac{25}{2}$, $a_4 = 14$, and $a_5 = \frac{1}{2}$, where φ is the standard normal probability density function, and
- $m_2(x) = b_1 \sin(2\pi b_2 x), x \in [0,1]$ with $b_1 = 1$, $b_2 = 3$.

The sample size N was 200.

As a kernel functions K, L we used the Epanechnikov kernel for computing the PC- and RLLS-estimate. The averaged squared error

$$\text{ase}\left(\hat{m}\left(\cdot\right)\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{m}\left(x_i\right) - m\left(x_i\right)\right)^2 w(x_i)$$

is used here to evaluate the estimates, where the weight function was chosen to be $w(x) = 1_{[0.1,0.9]}(x)$. The entries $\text{ase}_{RLLS}(h, g, \lambda)$, $\text{ase}_{PC}(h)$ in the table 5.1 denote the averaged squared errors of RLLS- and PC-estimate respectively and h_{ase} denotes the sample optimal bandwidth for PC-estimate, optimal with respect to $\text{ase}_{PC}(h)$.

noise	m	h	g	λ	$\text{ase}_{RLLS}(h, g, \lambda)$	$\text{ase}_{PC}(h)$	$\text{ase}_{PC}(h_{\text{ase}})$
SWN	m_2	0.1127	0.02	0.5	0.0235	0.0180	0.0178
SWN	m_2	0.1127	0.3	0.5	0.0502	0.0180	0.0178
SWN	m_2	0.02	0.02	40	0.0261	0.0989	0.0178
SWN	m_2	0.1063	0.02	40	0.0354	0.0178	0.0178
$AR(1)$	m_2	0.1888	0.04	0.4	0.0905	0.0341	0.0341
$AR(1)$	m_2	0.1888	0.4	0.4	0.1003	0.0341	0.0341
$AR(1)$	m_2	0.0221	0.043	40	0.1043	0.3387	0.0341
$AR(1)$	m_2	0.1888	0.02	40	0.0918	0.0341	0.0341
$AR(3)$	m_1	0.145	0.36	0.08	0.1865	0.1926	0.1926
$AR(3)$	m_1	0.06	0.2	0.95	0.1580	0.4978	0.1926
$AR(1)$	m_1	0.03	0.05	5	0.0703	0.0056	0.0046
$AR(1)$	m_1	0.01	0.01	5	0.0066	0.1486	0.0046

Table 5.1: Various values of the parameters h, g, λ, and the averaged squared errors of RLLS- and PC-estimate obtained in the simulation study.

In the following, we present the graphical results recorded in the Table 5.1.

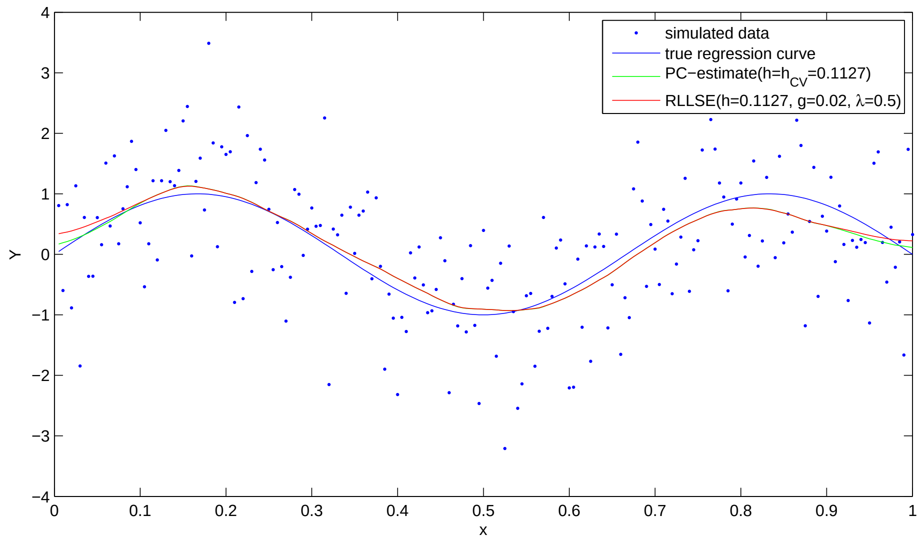

Figure 5.1: RLLSE with $h = 0.1127$, $g = 0.02$, $\lambda = 0.5$, and PC-estimate with bandwidth $h = 0.1127$ selected by using cross-validation for 200 simulated data points with i.i.d.$\mathcal{N}(0,1)$ errors where the true regression function was $m_2(x)$.

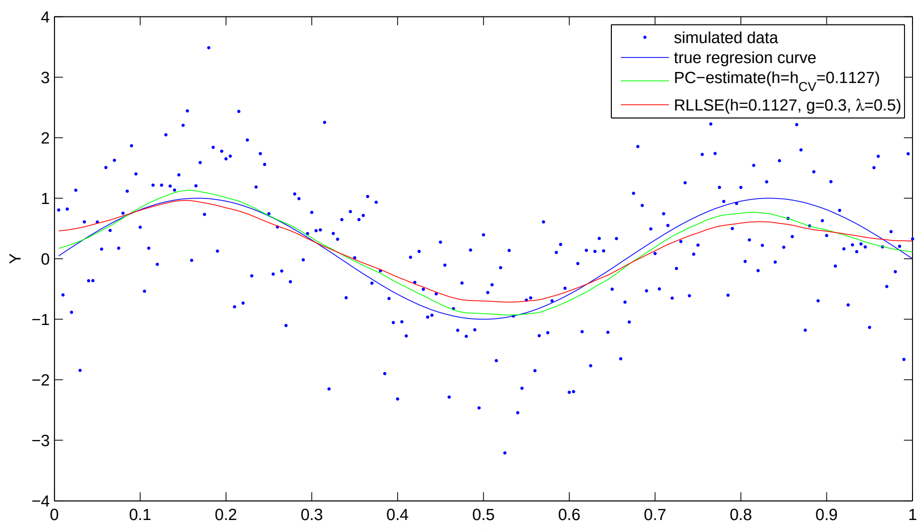

Figure 5.2: RLLSE with $h = 0.1127$, $g = 0.3$, $\lambda = 0.5$, and PC-estimate with bandwidth $h = 0.1127$ selected by using cross-validation for 200 simulated data points with i.i.d.$\mathcal{N}(0,1)$ errors where the true regression function was $m_2(x)$.

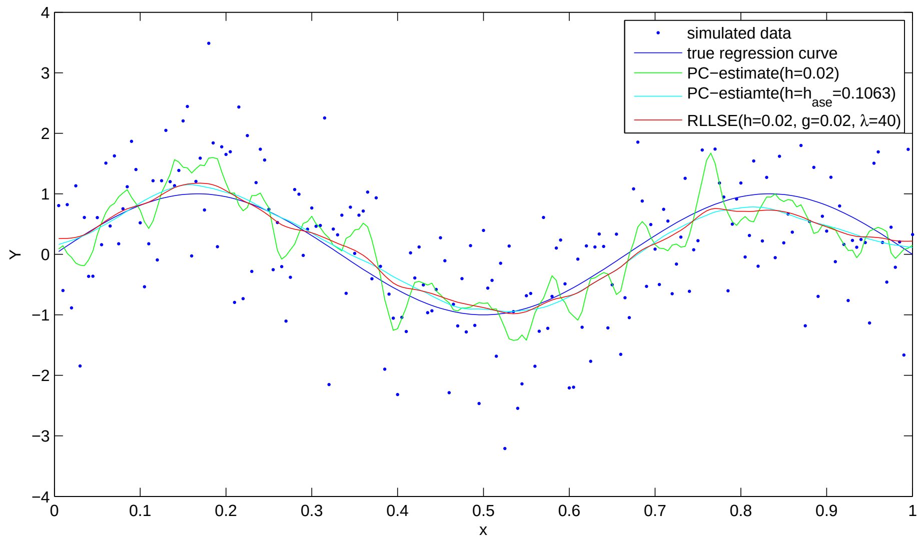

Figure 5.3: RLLSE with $h = 0.02$, $g = 0.02$, $\lambda = 40$, PC-estimate with bandwidth $h = 0.02$, and best PC-estimate for 200 simulated data points with i.i.d.$\mathcal{N}(0, 1)$ errors where the true regression function was $m_2(x)$.

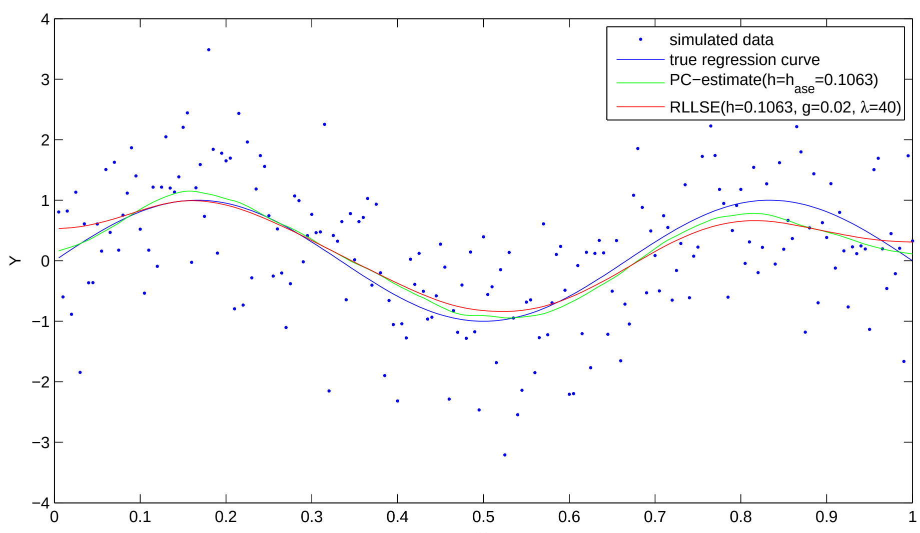

Figure 5.4: RLLSE with $h = 0.1063$, $g = 0.02$, $\lambda = 40$ and best PC-estimate for 200 simulated data points with i.i.d.$\mathcal{N}(0, 1)$ errors where the true regression function was $m_2(x)$.

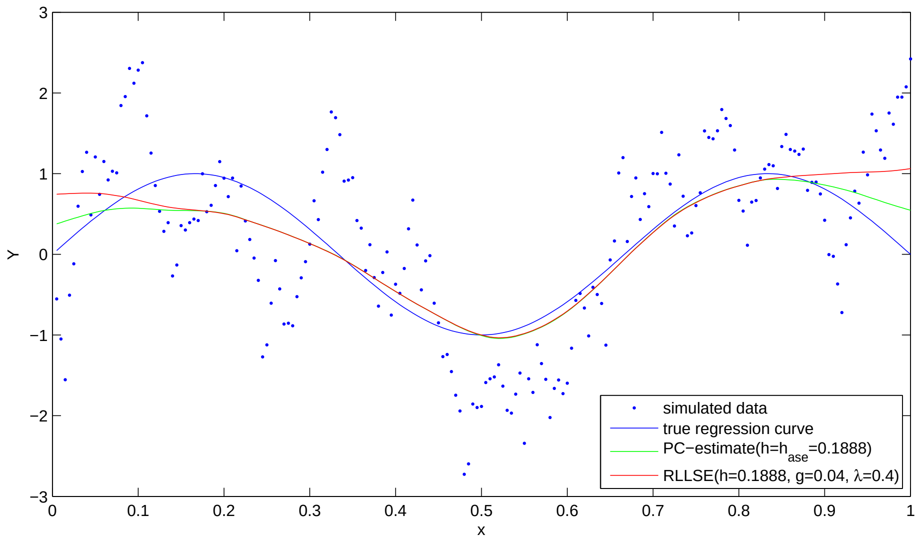

Figure 5.5: RLLSE with $h = 0.1888$, $g = 0.04$, $\lambda = 0.4$ and best PC-estimate for 200 simulated data points with AR(1) errors with coefficient $\alpha = 0.9$ where the true regression function was $m_2(x)$.

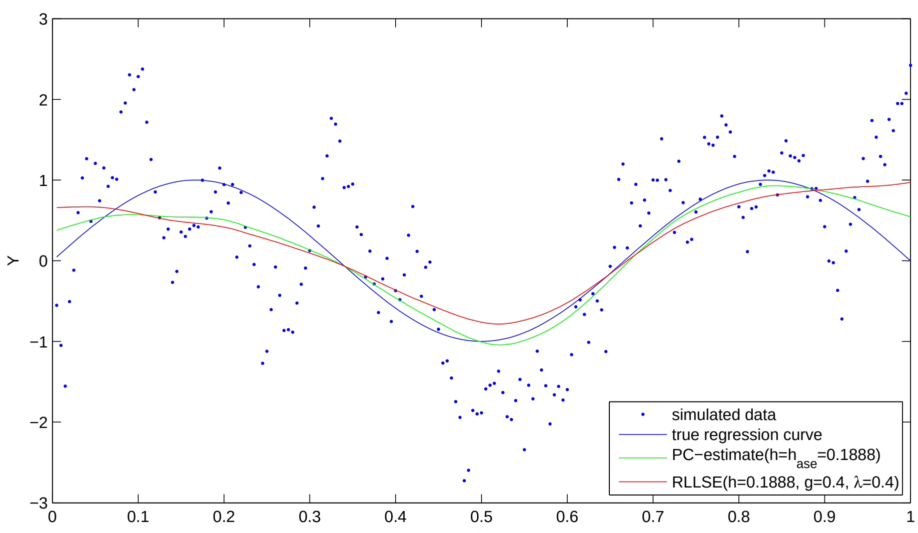

Figure 5.6: RLLSE with $h = 0.1888$, $g = 0.4$, $\lambda = 0.4$ and best PC-estimate for 200 simulated data points with AR(1) errors with coefficient $\alpha = 0.9$ where the true regression function was $m_2(x)$.

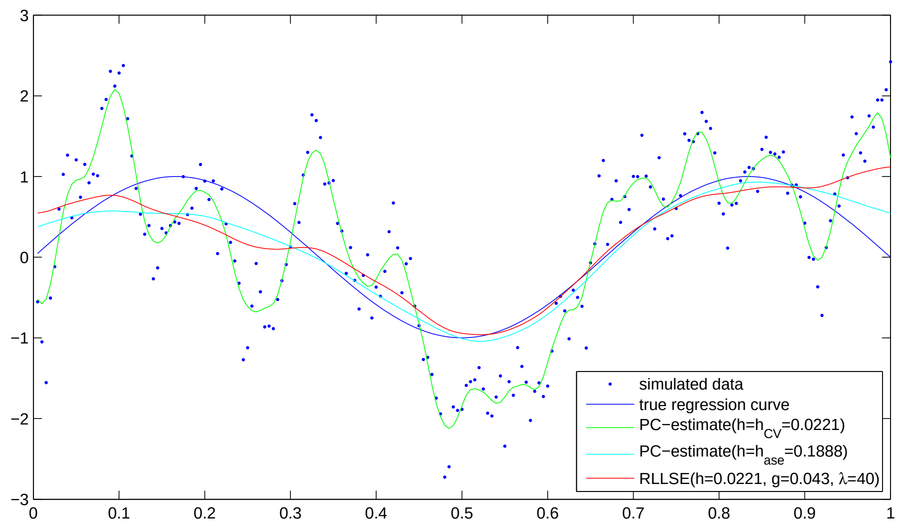

Figure 5.7: RLLSE with $h = 0.0221$, $g = 0.043$, $\lambda = 40$, PC-estimate with bandwidth $h = 0.0221$ selected by using cross validation, and best PC-estimate for 200 simulated data points with AR(1) errors with coefficient $\alpha = 0.9$ where the true regression function was $m_2(x)$.

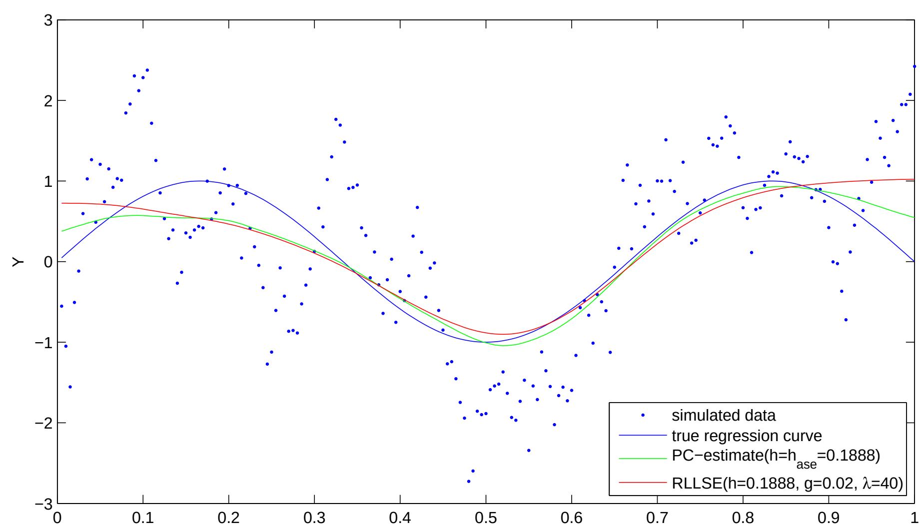

Figure 5.8: RLLSE with $h = 0.1888$, $g = 0.02$, $\lambda = 40$ and best PC-estimate for 200 simulated data points with AR(1) errors with coefficient $\alpha = 0.9$ where the true regression function was $m_2(x)$.

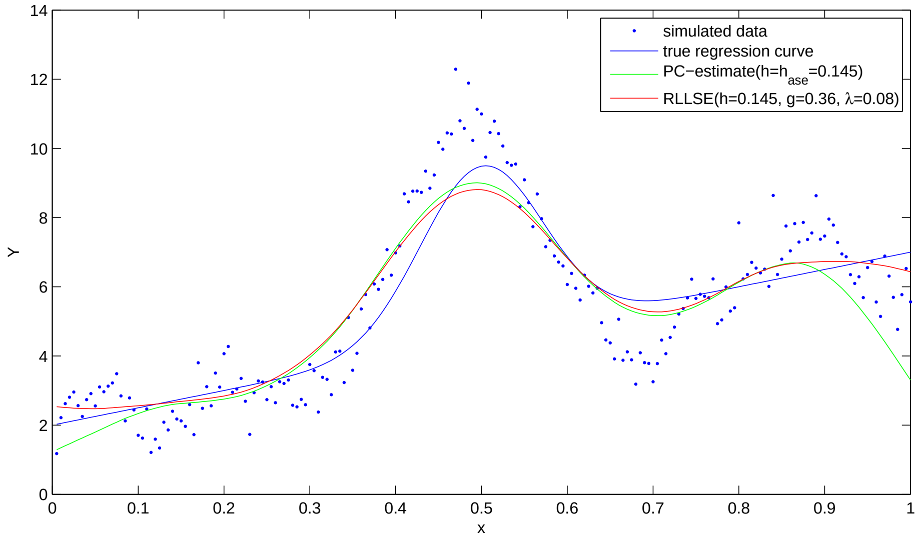

Figure 5.9: RLLSE with $h = 0.145$, $g = 0.36$, $\lambda = 0.08$ and best PC-estimate for 200 simulated data points with AR(3) errors with coefficients $\alpha_1 = 0.2$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, where the true regression function was $m_1(x)$.

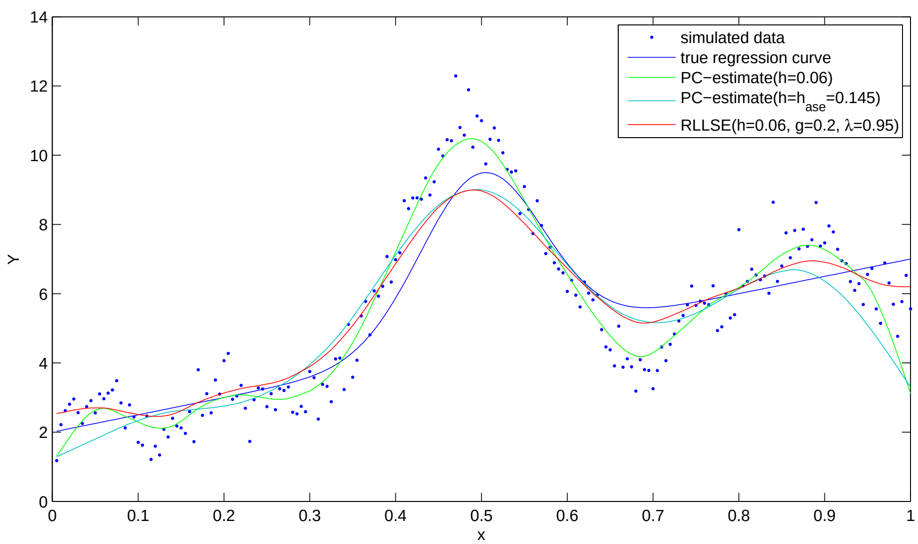

Figure 5.10: RLLSE with $h = 0.06$, $g = 0.2$, $\lambda = 0.95$, PC-estimate with bandwidth $h = 0.06$, and best PC-estimate for 200 simulated data points with AR(3) errors with coefficients $\alpha_1 = 0.2$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, where the true regression function was $m_1(x)$.

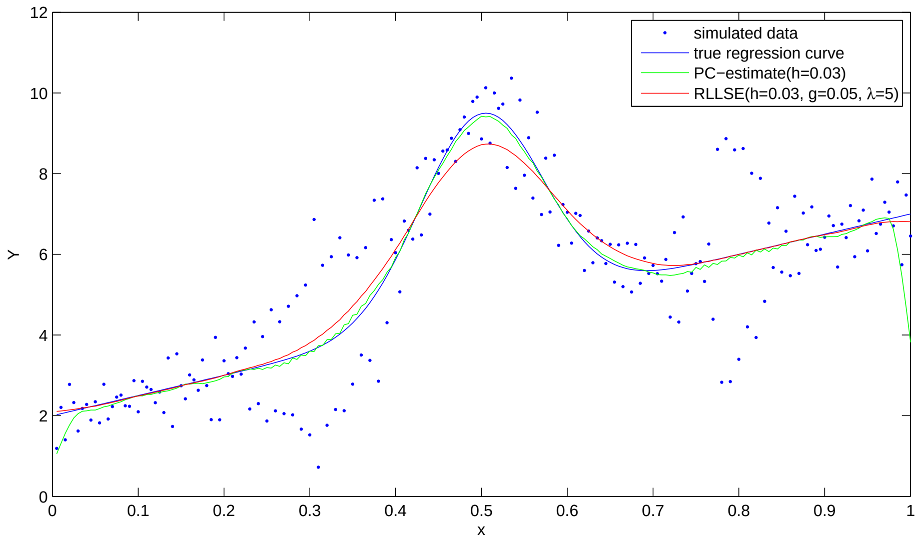

Figure 5.11: RLLSE with $h = 0.03$, $g = 0.05$, $\lambda = 5$, and PC-estimate with bandwidth $h = 0.03$ for 200 simulated data points with AR(1) errors with coefficient $\alpha = -0.9$ where the true regression function was $m_1(x)$.

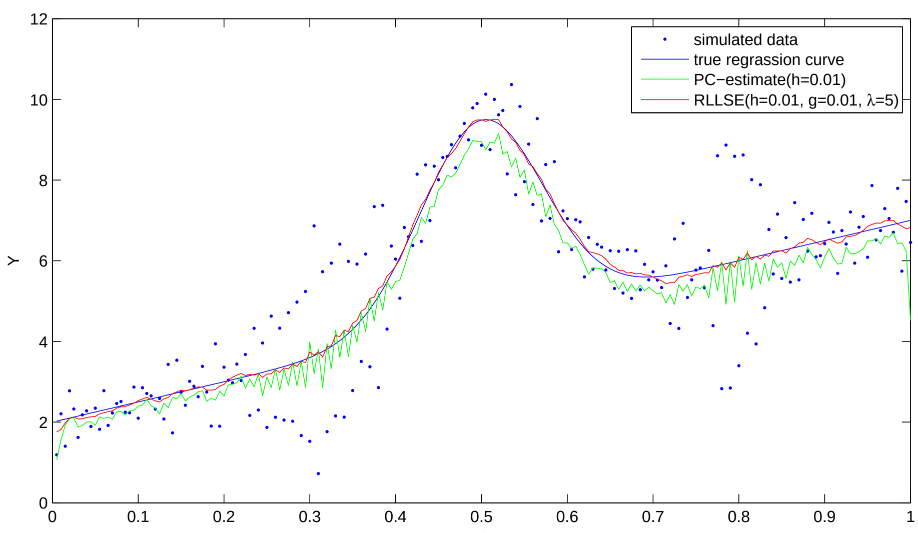

Figure 5.12: RLLSE with $h = 0.01$, $g = 0.01$, $\lambda = 5$ and PC-estimate with bandwidth $h = 0.01$ for 200 simulated data points with AR(1) errors with coefficient $\alpha = -0.9$ where the true regression function was $m_1(x)$.

Bibliography

[1] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.

[2] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, Hall, London, 1980.

[3] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Second Edition. Springer Series in Statistics, Springer 1996.

[4] G. Casella and R. L. Berger. *Statistical Inference*. Second edition. Duxbury, Pacific Grove, California, 2002.

[5] L. Breiman. *Probability*. Addison-Wesley, Reading, 1968.

[6] W. Fleming. *Functions of Several Variables*. Second edition. Springer-Verlag, New York, 1977.

[7] D. Werner. *Funktionalanalysis (Functional Analysis, in German)*. Sixth corrected edition. Springer-Verlag, Berlin and Heidelberg, 2007.

[8] T. Gasser and H.-G. Müller. *Kernel Estimation of Regression Functions*, in Smoothing Techniques for Curve Estimation, Lecture Notes in Math. 757 Berlin: Springer-Verlag, pp. 23-68, 1979.

[9] J. Franke, J. Tadjuidje, J. Weickert, and S. Didas. *Some asymptotics for local least-squares regression with regularization*. Report in Wirtschaftsmathematik, No. 107, Department of mathematics, University of Kaiserslautern, 2007.

[10] J. Franke and S. Halim. *Image denoising with local adaptive smoothing in the presence of correlated noise*. Preprint, Department of mathematics, University of Kaiserslautern, Germany.

[11] S. Halim. *Spatially adaptive detection of local disturbances in time series and stochastic processes on the integer lattice* $\mathbb{Z}^2$. PhD thesis, Universität Kaiserslautern-Germany, Department of Mathematics, 2005.

[12] V. Friederichs. *Local Smoothing Methods in Image Processing*. PhD thesis, Universität Kaiserslautern-Germany, Department of Mathematics, 1999.

[13] M. B. Priestley and M. T. Chao. *Nonparametric function fitting*. J. Royal Stat. Soc., Ser. B, 34: 385-392, 1972.

[14] P. Mrázek, J. Weickert, and A. Bruhn. *On robust estimation and smoothing with spatial and tonal kernels*. In R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors, *Geometric Properties from Incomplete Data*. Springer, Dordrecht, 2006.

[15] I. J. Schoenberg. *Spline functions and the problem of graduation*. Proc. Nat. Acad. Sci. (USA), 52: 947-950, 1964.

[16] C. H. Reinsch. Smoothing by spline functions. Numer. Math., 10: 177-183, 1967.

[17] E. Whittaker. On a new method of graduation. Proc. Edinburgh Math. Soc. 41: 63-75, 1923.

[18] B. W. Silverman. *Spline smoothing: the equivalent variable kernel method*. The Annals of Statistics, 12(3): 898-916, 1984.

[19] J. Franke and W. Härdle. *On bootstrapping kernel spectrum estimates*. Ann. Statist. 20: 121-145, 1992.

[20] E. Herrmann, T. Gasser, and A. Kneip. *Choice of bandwidth for kernel regression when residuals are correlated*. Biometrika, 79,4: 783-795, 1992.

[21] N.S. Altman. *Kernel smoothing with correlated errors*. J. Am. Statist. Assoc. 85: 749-759, 1990.

[22] J. D. Hart. *Kernel regression estimation with time series errors*. J. R. Statist. Soc. B 53: 173-188, 1991.

[23] J. W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial & Applied Mathematics,U.S., Berkeley, California, 1997.

[24] H. G. Müller and U. Stadtmüller. *Estimation of heteroscedasticity in regression analysis*. Ann. Statist. Ass., 15: 610-625, 1987.

[25] M. Taniguchi and Y. Kakizawa. *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York, 2000.

[26] A. W. van der Vaart. *Asymptotic Statistics*. Eighth printing. Cambridge University Press, Cambridge, 2007.

[27] M. P. Wand and M. C. Jones. *Kernel Smoothing*. First edition. Chapman and Hall, London, 1995.

[28] J. K. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific, Singapore, 2001.

[29] J. Jacod and P. Protter. *Probability Essentials*. Second edition (corrected second printing). Springer-Verlag, Berlin and Heidelberg, 2004.

[30] I. J. Maddox. *Elements of Functional Analysis*. Second edition. Cambridge University Press, Cambridge, 1988.

[31] A. Pinkus and S. Zafrany. *Fourier Series and Integral Transforms*. Cambridge University Press, Cambridge, 1997.

Curriculum Vitae

1996	High School Certificate, Syria
1996-2001	Studies in Mathematics and Informatics at the University of Damascus, Syria
2001	Bachelor of Science in Mathematics "section Informatics"
2001-2002	Studies in Mathematics and Informatics at the University of Damascus
2002	Diploma in Mathematics and Informatics, University of Damascus, Syria
2004-2007	Studies in Mathematics at the University of Kaiserslautern
2007	Master of Science in Mathematics specialization "mathematical statistics", University of Kaiserslautern, Germany
2007-2011	PhD student at the Mathematics Department of the University of Kaiserslautern, Germany