Local Smoothers with Regularization

Hani Kabajah

"Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes"

"Printed with the assistance from the German Academic Exchange Service"

Local Smoothers with Regularization

Hani Kabajah

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. Gabriele Steidl

Datum der Disputation: 12. Mai 2010

D 386

Local Smoothers with Regularization

Hani Kabajah

Approved dissertation
by the Department of Mathematics
at the University of Kaiserslautern
for awarding the degree
Doctor of Natural Sciences
(Doctor rerum naturalium, Dr. rer. nat.)

First referee: Prof. Dr. Jürgen Franke
Second referee: Prof. Dr. Gabriele Steidl

Date of the dissertation oral defense: May 12, 2010

D 386

Abstract

Mrázek et al. [25] proposed a unified approach to curve estimation which combines localization and regularization. Franke et al. [10] used that approach to discuss the case of the regularized local least-squares (RLLS) estimate. In this thesis we will use the unified approach of Mrázek et al. to study some asymptotic properties of local smoothers with regularization. In particular, we shall discuss the Huber M-estimate and its limiting cases towards the L_2 and the L_1 cases. For the regularization part, we will use quadratic regularization. Then, we will define a more general class of regularization functions. Finally, we will do a Monte Carlo simulation study to compare different types of estimates.

Acknowledgments

First of all, I would like to thank Prof. Dr. Jürgen Franke for giving me the opportunity to work on this thesis, and for his continuous support and supervision.

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

Special thanks go to the German Academic Exchange Service (DAAD) for their financial support.

My thanks go to Nico Behrent for his continuous and fast support in computer issues.

I would like to thank Dr. Stephan Didas from the Department of Image Processing at the Fraunhofer Institute for Industrial and Financial Mathematics (ITWM) for the discussions regarding the practical aspects of the work.

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 (especially those with Dr. Joseph Tadjuidje).

Abbreviations

abbreviation	*meaning*
$\mathbb{N}$	The set of natural numbers: $\mathbb{N} = \{1, 2, 3, \dots\}$
$\mathbb{R}$	The set of real numbers: $\mathbb{R} = (-\infty, \infty)$
a_n, b_n	Real-valued sequences
$a_n = O(b_n)$	$\exists\ C > 0, N \in \mathbb{N}: \ \ \lvert a_n/b_n\rvert \leq C$ for every $n > N$
$a_n = o(b_n)$	$\lim_{n\to\infty} \lvert a_n/b_n\rvert = 0$
$a_n \sim b_n$	$\lim_{n\to\infty} \lvert a_n/b_n\rvert = 1$
$a_n \sim$ constant b_n	$\lim_{n\to\infty} \lvert a_n/b_n\rvert =$ constant $\neq 0$
X_n	Sequence of random variables
$X_n = O_{a.s.}(b_n)$	$\lvert X_n(\omega)\rvert/\lvert b_n\rvert < C_\omega$ for almost all ω where C_ω is a positive constant
$X_n = o_{a.s.}(b_n)$	$\lim_{n\to\infty} \lvert X_n\rvert/\lvert b_n\rvert = 0$ almost surely
$X_n = O_p(b_n)$	$\forall\delta > 0,\ \exists M > 0,\ N \in \mathbb{N}:\ \mathbb{P}\left(\lvert X_n\rvert/\lvert b_n\rvert > M\right) < \delta$ for every $n > N$
$X_n = o_p(b_n)$	$\lim_{n\to\infty} \mathbb{P}\left(\lvert X_n\rvert/\lvert b_n\rvert > \delta\right) = 0$ for every $\delta > 0$
$\text{supp}(f)$	Support of a function: $\text{supp}(f) = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$
$K(u)$	A kernel function
$K_h(u)$	A rescaled kernel function: $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$
Q_K	$= \int K^2(u)du$
S_K	$= \int K^3(u)du$
V_K	$= \int u^2 K(u)du$

$\mu(x)$	The regression function of the nonparametric model
μ_i	Shorthand writing for $\mu(x_i)$ where $x_i = i/N$
$\boldsymbol{\mu}$	$= (\mu(x_1), \ldots, \mu(x_N))^T = (\mu_1, \ldots, \mu_N)^T$
$\boldsymbol{\mu}''$	$= (\mu''(x_1), \ldots, \mu''(x_N))^T = (\mu''_1, \ldots, \mu''_N)^T$
$\hat{\mu}(x)$	The Priestley-Chao (PC) kernel estimate of $\mu(x)$
$\hat{\mu}_i$	Shorthand writing for $\hat{\mu}(x_i)$ where $x_i = i/N$
$\hat{\boldsymbol{\mu}}$	$= (\hat{\mu}_1, \ldots, \hat{\mu}_N)^T$
$\hat{\mu}_K(x, h)$	The PC-estimate with kernel K and bandwidth h
$\hat{\mu}_L(x, g)$	The PC-estimate with kernel L and bandwidth g
$\tilde{\mu}(x)$	The local Huber M-estimate (LHM-estimate) of $\mu(x)$
$\tilde{\mu}_i$	Shorthand writing for $\tilde{\mu}(x_i)$ where $x_i = i/N$
$\tilde{\boldsymbol{\mu}}$	$= (\tilde{\mu}_1, \ldots, \tilde{\mu}_N)^T$
$\tilde{\mu}_K(x, h)$	The LHM-estimate with kernel K and bandwidth h
$\tilde{\mu}_L(x, g)$	The LHM-estimate with kernel L and bandwidth g
$\hat{\mathbf{u}}_{\mathbf{PC}}$	Vector of PC-estimates at the grid points $x_i = i/N$
$\hat{\mathbf{u}}_{\mathbf{LS}}$	Vector of QRLLS-estimates at the grid points $x_i = i/N$
$\hat{\mathbf{u}}_{\mathbf{HM(c)}}$	Vector of QRLHM-estimates at the grid points $x_i = i/N$
$\hat{\mathbf{u}}_{\mathbf{LA}}$	Vector of QRLLA-estimates at the grid points $x_i = i/N$

Contents

List of Tables

Chapter 1

Estimation and Smoothing

In this chapter, we introduce the general idea of smoothing and, in particular, kernel smoothing. Then we introduce the general approach for image denoising developed by Mrázek et al. [25]. Based on this approach, we present some of the results obtained by Franke et al. [10] for the case of RLLS-estimates. Finally, we describe the problem we would like to discuss in detail here.

1.1 What is Smoothing?

Smoothing of a data set $\{(X_j, f_j) : j = 1, \ldots, N\}$ involves the approximation of the mean response curve μ in the regression relationship

$$f_j = \mu(X_j) + \varepsilon_j, \quad j = 1, \ldots, N. \tag{1.1}$$

The functional of interest could be the regression curve itself μ, certain derivatives of it or functions of derivatives such as extrema or inflection points. But we restrict the case here to estimating μ only.

If there are repeated observations at a fixed point $X = x$ estimation of $\mu(x)$ can be done by using just the average of the corresponding f-variables. However, in the majority of cases, repeated responses at a given x can not be obtained. In most studies of the regression relationship given by (1.1), there is just a single response variable f and a single predictor variable X which may be a vector in $\mathbb{R}^d$. In our study, we will consider only the case $d = 1$.

In the trivial case in which $\mu(x)$ is a constant, estimation of μ reduces to the point estimation of location, since an average over the response variable f yields an estimate of μ. In practical studies, it is unlikely that the regression curve is constant. Rather the assumed curve is modeled as a smooth continuous function of a particular structure which is "nearly constant" in a small neighbourhood of x.

A quite natural choice of the estimator of μ, denoted by $\hat{\mu}$, is the mean of the response variables near a point x. This (local average) should be constructed in such a way that it is defined only from observations in a small neighbourhood around x, since f-observations from points far away from x will have, in general, very different mean values. This local averaging procedure can be viewed as the basic idea of smoothing. More formally this procedure can be defined as

$$\hat{\mu}(x) = \frac{1}{N} \sum_{j=1}^{N} W_{Nj}(x) f_j \tag{1.2}$$

where $\{W_{Nj}(x)\}_{j=1}^{N}$ denotes a sequence of weights which depend on the whole vector $\{X_j\}_{j=1}^{N}$.

Smoothing methods are strictly or asymptotically of the form (1.2). The estimator of the regression function $\mu(x)$ (denoted by $\hat{\mu}(x)$, $\tilde{\mu}(x)$, etc.) is called a *smoother*.

Special attention has to be paid to the fact that smoothers average over observations with different mean values. The amount of averaging is controlled by the weight sequence $\{W_{Nj}(x)\}_{j=1}^{N}$ which is tuned by a smoothing parameter. The smoothing parameter regulates the size of the neighbourhood around x, and should be chosen in a way to balance over-smoothing and under-smoothing.

1.2 Kernel Smoothing

In this section we describe the basic idea of kernel smoothing and give some examples of kernel estimates.

For more details see Jennen-Steinmetz and Gasser [21] where they discuss the nonparametric regression estimation methods well-know up to 1988 (for example: the Priestley-Chao kernel estimate, the Nadaraya-Watson kernel estimate, the Gasser-Müller kernel estimate, the spline smoother, etc.).

We start by defining kernel functions.

Definition 1.1 *A kernel K is a bounded, continuous function on $\mathbb{R}$ satisfying $\int K(u)du = 1$.*

In estimating functions, a kernel usually has to satisfy the following

$$K(u) \geq 0, \qquad \int uK(u)du = 0, \qquad \int u^2 K(u)du < \infty. \tag{1.3}$$

These are the least assumptions imposed on kernels, further assumptions will be imposed later.

In the context of smoothing, we define

$$K_h(u) := \frac{1}{h} K\left(\frac{u}{h}\right)$$

where $h > 0$. K_h is called the rescaled kernel and the smoothing parameter h is called the bandwidth.

Definition 1.2 *Let m be a real-valued function then the support of m is defined as*

$$\text{supp}(m) = \overline{\{x \in \mathbb{R} : m(x) \neq 0\}}.$$

Moreover, if $\text{supp}(K) = [-1, +1]$, then $\text{supp}(K_h) = [-h, +h]$.

Example 1.3 (Some kernel functions) *a) Gauss kernel:*

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad u \in \mathbb{R}.$$

The support of this kernel is the whole real line.

b) *Epanechnikov (or Bartlett-Priestley) kernel:*

$$K(u) = \frac{3}{4}(1-u^2)^+, \quad u \in \mathbb{R},$$

where $u^+ = u\mathbb{1}_{\{u\geq 0\}}(u)$. *The support of this kernel is the interval* $[-1,1]$.

Regression Models

Now, we introduce two designs associated with the regression model (1.1). The first design is the *equidistant design* or the *deterministic equidistant design* model. This model arises when we observe a sample $(x_1, f_1), \ldots, (x_N, f_N)$ of data pairs which follows the regression model (1.1), where ε_j are independent identically distributed random variables with mean zero and variance σ^2, and the x_j come from an equidistant grid in the unit interval $[0,1]$. That is,

$$f_j = \mu(x_j) + \varepsilon_j, \quad \mathcal{L}(\varepsilon_j) = \text{i. i. d.}\,(0,\sigma^2), \quad x_j = \frac{j}{N}, \quad j = 1,\ldots,N. \tag{1.4}$$

The second design is the *stochastic design* or the *random design* model. This model arises when we observe a sample $(X_1, f_1), \ldots, (X_N, f_N)$ of data pairs which follows the regression model (1.1), where (conditional on $X_1,\ldots,X_N$) ε_j are independent identically distributed random variables with mean zero and variance σ^2. That is,

$$f_j = \mu(X_j) + \varepsilon_j, \quad \mathcal{L}(\varepsilon_j | X_1,\ldots,X_N) = \text{i. i. d.}\,(0,\sigma^2), \quad j = 1,\ldots,N. \tag{1.5}$$

In the stochastic design context $\mu(x) = \mathbb{E}\,(f|X=x)$ and $\sigma^2 = \operatorname{var}(f|X=x)$ are the conditional mean and variance of f given $X = x$. The density of $X_1,\ldots,X_N$ will be denoted by p.

Priestley-Chao and Nadaraya-Watson Kernel Estimates

Now, we introduce some estimates of the regression function μ.

Definition 1.4 *Let the model (1.5) hold. The Priestley-Chao kernel estimate of* $\mu : \mathbb{R} \to \mathbb{R}$ *with bandwidth* $h > 0$ *and some kernel* K *is defined as*

$$\begin{aligned}\hat{\mu}_K(x,h) :=& \sum_{j=1}^{N}(X_i - X_{i-1})K_h(x - X_j)f_j \\ =& \frac{1}{h}\sum_{j=1}^{N}(X_i - X_{i-1})K\left(\frac{x-X_j}{h}\right)f_j, \quad x \in \mathbb{R}.\end{aligned}$$

If the model (1.4) holds, then the Priestley-Chao kernel estimate of $\mu : [0,1] \to \mathbb{R}$ *is given by*

$$\hat{\mu}_K(x,h) := \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)f_j$$

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

In view of (1.2), the Priestley-Chao estimate under the equidistant design can be seen as a weighted average with weights

$$W_{Nj}(x)=K_h(x-x_j).$$

Definition 1.5 *Let the model (1.5) hold. The Rosenblatt-Parzen kernel density estimate of $p(x)$ with bandwidth $h>0$ and some kernel K is defined as*

$$\hat{p}_K(x,h):=\frac{1}{N}\sum_{j=1}^{N}K_h(x-X_j), \quad x\in\mathbb{R}.$$

Definition 1.6 *Let the model (1.5) hold. The Nadaraya-Watson kernel estimate of $\mu:\mathbb{R}\to\mathbb{R}$ with bandwidth $h>0$ and some kernel K is defined as*

$$\begin{aligned}\hat{\mu}_{NW}(x,h):=&\frac{\sum_{j=1}^{N}K_h(x-X_j)f_j}{\sum_{j=1}^{N}K_h(x-X_j)}\\ =&\frac{1}{\hat{p}_K(x,h)}\frac{1}{N}\sum_{j=1}^{N}K_h(x-X_j)f_j, \quad x\in\mathbb{R}.\end{aligned}$$

If the model (1.4) holds, then the Nadaraya-Watson kernel estimate of $\mu:[0,1]\to\mathbb{R}$ with bandwidth $h>0$ and some kernel K is given by

$$\begin{aligned}\hat{\mu}_{NW}(x,h):=&\frac{\sum_{j=1}^{N}K_h(x-x_j)f_j}{\sum_{j=1}^{N}K_h(x-x_j)}\\ =&\frac{\hat{\mu}_K(x,h)}{\hat{p}_K(x,h)}, \quad x\in[0,1].\end{aligned}$$

Like the Priestley-Chao estimate, the Nadaraya-Watson kernel estimate can also be seen as a weighted average with weights

$$W_{Nj}(x)=\frac{K_h(x-X_j)}{\hat{p}(x,h)}.$$

Under the equidistant design $\hat{p}_K(x,h)\to 1$ as $N\to\infty$ (cf. Lemma 1.9 below). Using this fact we can say that under the equidistant design the Priestley-Chao and the Nadaraya-Watson kernel estimates are asymptotically equivalent.

1.3 Spline Smoothing

Another well known method in nonparametric regression estimation is the method of spline smoothing. For example, under model (1.4), the *cubic spline estimator* $\hat{\mu}_{CS}(x, \lambda)$ is defined as the minimizer of

$$S_\lambda(g) = \frac{1}{N}\sum_{j=1}^{N}(f_j - g(x_j))^2 + \lambda \int_0^1 (g''(x))^2 dx \tag{1.6}$$

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 $\int_0^1 (g''(x))^2 dx$) and goodness of fit to the data (measured here by the least-squares). The larger the value of λ the smoother the estimate.

This form of spline smoothing is due to Schoenberg in 1964 and Reinsch in 1967. However, the idea of penalizing a measure of goodness of fit by a one for roughness was described by Whittaker in 1923.

In 1984, Silverman [30] showed that spline smoothers (which could be written as in (1.2) with weights $W_{Nk}^{\lambda}(x)$) are asymptotically equivalent to kernel estimates.

For more details about spline smoothers and further references see Silverman [30].

1.4 Robust M-Estimation

Under model (1.4), we can see the Nadaraya-Watson kernel estimate $\hat{\mu}_{NW}(x, h)$ as the solution of the following local least-squares minimization problem

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)(u - f_j)^2 = \min_{u\in\mathbb{R}}! \tag{1.7}$$

The Nadaraya-Watson kernel estimate and its asymptotically equivalent estimate, under model (1.4), the Priestley-Chao estimate, are optimal when the error terms are Gaussian. However, they are highly disturbed by outliers.

To get an estimate which is robust against outliers, Huber [16] proposed in 1964 using

$$\rho(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} \tag{1.8}$$

as a target function of the minimization problem instead of the quadratic function. For large c, ρ behaves like u^2 while for small c it behaves like $|u|$. For example, under model (1.4), the *local Huber M-estimate* $\tilde{\mu}(x, h)$ is defined as the solution of

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\rho(u - f_j) = \min_{u\in\mathbb{R}}! \tag{1.9}$$

For more details see Huber [16] (for M-estimates) and Härdle [13] (for local M-estimates).

1.5 Approach by Mrázek et al.

In this section we introduce the general approach for image denoising proposed by Mrázek et al. [25]. This approach covers most of the methods described above for nonparametric regression estimates and makes use of the penalization strategy to reduce over-smoothing when it occurs.

Let us assume there is an unknown (constant) signal u, and it is observed N-times. We obtain the noisy samples f_j, $j = 1, \ldots, N$, according to $f_j = u + \varepsilon_j$ where ε_j stands for the noise. If ε_j are zero-mean Gaussian (normal) random variables, one can estimate u by calculating the sample mean $\bar{u} = \frac{1}{N}\sum_{j=1}^{N} f_j$. The mean $\bar{u}$ is the maximum a posteriori (MAP) estimate of u, and minimizes the L_2 error $Q(u) = \sum_{j=1}^{N}(u - f_j)^2$.

In image analysis, the data (grey values) f_j are measured at positions (pixels) x_j, and we want to find a solution vector $\mathbf{u} = (u_j)_{j=1,\ldots,N}$ where each output value u_j belongs to the position x_j.

Mrázek et al. established a general approach for image denoising which combines localization and regularization. The localization effect comes from the weight functions introduced into the energy functional to be minimized, and the regularization effect is obtained by adding another smoothness penalizing term. The final energy functional to be minimized (with respect to $\mathbf{u}$) is:

$$\begin{aligned} Q(\mathbf{u}) &= Q_D(\mathbf{u}) + \frac{\lambda}{2} Q_S(\mathbf{u}) \\ &= \underbrace{\sum_{i,j=1}^{N} \underbrace{\Psi_D\big(|u_i - f_j|^2\big)}_{\text{tonal wt. func.}} \underbrace{w_D\big(|x_i - x_j|^2\big)}_{\text{spatial wt. func.}}}_{\text{Data Term}} + \underbrace{\frac{\lambda}{2} \sum_{i,j=1}^{N} \underbrace{\Psi_S\big(|u_i - u_j|^2\big)}_{\text{tonal wt. func.}} \underbrace{w_S\big(|x_i - x_j|^2\big)}_{\text{spatial wt. func.}}}_{\text{Smoothness Term}} . \end{aligned} \tag{1.10}$$

The *data loss function* or the *data tonal weight function* Ψ_D is a penalizing function measuring the fit of $\mathbf{u}$ to the observations $f_1, \ldots, f_N$, where the *smoothness loss function* or the *smoothness tonal weight function* Ψ_S is a penalizing function measuring the smoothness of the solution. The *data weight function* or the *data spatial weight function* w_D takes care of the localization effect in the data part, that is, the observations f_j whose corresponding x_j are closest to the point where we are making the estimation has more weight than other observations. Whereas the *smoothness weight function* or the *smoothness spatial weight function* w_S takes care of the localization effect in the smoothness part of the energy functional. The *tuning parameter* or the *regularization parameter* $\lambda \geq 0$ balances between fit and smoothness.

Example 1.7 *Under model (1.4), the general approach gives the following estimates for μ.*

1. *Least-squares estimate (the mean): $\Psi_D(s^2) = s^2$, $w_D(x^2) = 1$, and $\lambda = 0$. The solution is the vector*

$$\bar{\mathbf{f}} = \left(\frac{1}{N}\sum_{j=1}^{N} f_j, \ldots, \frac{1}{N}\sum_{j=1}^{N} f_j\right)^T = \big(\bar{f}, \ldots, \bar{f}\big)^T .$$

(We can see from here the importance of localization.)

2. *Least-absolute deviation estimate (the median):* $\Psi_D(s^2) = |s|$, $w_D(x^2) = 1$, *and* $\lambda = 0$. *The solution is the vector*

$$\tilde{\mathbf{f}} = \left(\tilde{f}, \ldots, \tilde{f}\right)^T,$$

where $\tilde{f}$ *is the sample median of the values* $f_1, \ldots, f_N$. *The solution is obtained by the so-called median minimizing property (for example, see [3]).*

3. *Local least-squares estimate (the Nadaraya-Watson kernel estimate):* $\Psi_D(s^2) = s^2$, $w_D(x^2) = K_h(x)$, *and* $\lambda = 0$. *The solution is the vector*

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{NW} &= \left(\frac{\sum_{j=1}^N K_h(x_1 - x_j) f_j}{\sum_{j=1}^N K_h(x_1 - x_j)}, \ldots, \frac{\sum_{j=1}^N K_h(x_N - x_j) f_j}{\sum_{j=1}^N K_h(x_N - x_j)}\right)^T \\ &= \left(\hat{\mu}_{NW}(x_1, h), \ldots, \hat{\mu}_{NW}(x_N, h)\right)^T.\end{aligned}$$

4. *Local Huber M-estimate:* $\Psi_D(s^2) = \rho(s)$, *where* ρ *is the Huber function given by* (1.8), $w_D(x^2) = K_h(x)$, *and* $\lambda = 0$. *The solution is the vector*

$$\begin{aligned}\tilde{\boldsymbol{\mu}} &= \left(\operatorname*{argmin}_{u_1 \in \mathbb{R}} \frac{1}{N}\sum_{j=1}^N K_h(x_1 - x_j)\rho(u_1 - f_j), \ldots, \operatorname*{argmin}_{u_N \in \mathbb{R}} \frac{1}{N}\sum_{j=1}^N K_h(x_N - x_j)\rho(u_N - f_j)\right)^T \\ &= \left(\tilde{\mu}(x_1, h), \ldots, \tilde{\mu}(x_N, h)\right)^T.\end{aligned}$$

1.6 Regularized Local Least-Squares Estimates

In this section we will have a look at the case of regularized local least-squares (RLLS) estimates discussed by Franke et al. [10]. All results presented in this section are due to Franke et al. [10], where complete proofs can be found.

Assuming model (1.4) the RLLS case is driven from the general approach by Mrázek et al. [25] by choosing

$$\Psi_D(s^2) = s^2, \quad w_D(x^2) = K_h(x), \quad \Psi_S(s^2) = s^2, \quad w_S(x^2) = L_g(x),$$

where the kernels K and L are standardized nonnegative, symmetric functions on $\mathbb{R}$ and the bandwidths $h, g > 0$ can be chosen to control the smoothness of the function estimate together with the balancing factor λ. Therefore, the RLLS minimization problem can be written as

$$\begin{aligned}Q(u_1, \ldots, u_N) = &\sum_{i,j=1}^N (u_i - f_j)^2 K_h(x_i - x_j) \\ &+ \frac{\lambda}{2}\sum_{i,j=1}^N (u_i - u_j)^2 L_g(x_i - x_j) = \min_{u_1, \ldots, u_N}!\end{aligned} \tag{1.11}$$

The solution here has an explicit representation in terms of the Priestley-Chao estimate. For convenience, the following notation for the values of the Priestley-Chao estimate at the grid points x_i, $i = 1, \dots, N$ will be used

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_N)^T \quad \text{with} \quad \hat{\mu}_i = \hat{\mu}_K(x_i, h), \quad i = 1, \dots, N.$$

Proposition 1.8 *Let $\hat{p}_L(x, g)$ be defined analogously to $\hat{p}_K(x, h)$ with L, g replacing K, h, and let $\hat{p}_\lambda(x, h, g) = \hat{p}_K(x, h) + \lambda \hat{p}_L(x, g)$. Let Λ denote the $N \times N$-matrix with entries $\Lambda_{i,j} = \frac{1}{N} L_g(x_i - x_j)$, and let $\hat{P}$ denote the $N \times N$-diagonal matrix with entries $\hat{P}_{ii} = \hat{p}_\lambda(x_i, h, g)$. Then, if $\hat{P} - \lambda\Lambda$ is invertible, the RLLS-estimate as the minimizer of (1.11) is given by*

$$\mathbf{u} = \left(\hat{P} - \lambda\Lambda\right)^{-1} \hat{\boldsymbol{\mu}}.$$

But in order to get the bias and variance terms as well as the asymptotic distribution of the estimate, some asymptotic expansion is needed. For that purpose Franke et al. impose the following assumptions,

(A1) a) K is a nonnegative, symmetric kernel function with compact support $[-1, 1]$.
- b) $\int K(u)du = 1$.
- c) K is Lipschitz continuous with Lipschitz constant C_K.
- d) $K(\pm 1) = 0$.
- e) $K \in \mathcal{C}^2(-1, +1)$ with bounded second derivative K''.
- f) K'' is Lipschitz continuous, and $K'(\pm 1) = 0$.

To make arguments simple, the discussion is restricted to the case where boundary effects are neglected, i.e. $x \in [h, 1-h]$ and $h > 0$. However, boundary effects vanish asymptotically since $h \to 0$ as $N \to \infty$.

Throughout the text we will use the following abbreviations

$$V_K = \int z^2 K(z)dz, \quad Q_K = \int K^2(z)dz.$$

In the case discussed here, $x_1, \dots, x_N$ are equidistant and behave similar to uniform random variables. In particular, $\hat{p}_K(x, h)$ converges to the density of the uniform distribution under the assumptions mentioned above. This is given in the following lemma.

Lemma 1.9 *Assuming (A1) a)-e) for the kernel K, we have for some constant $\alpha > 0$*

$$|1 - \hat{p}_K(x, h)| \le \frac{\alpha}{N^2 h^2} \quad \textit{for all} \quad x \in [h, 1-h].$$

To make notation easier we make the following definition.

Definition 1.10 (PC-iterated smoothers) *Set $\hat{\mu}_1(x, h, g) := \hat{\mu}_K(x, h)$, and recursively define the "iterated smoothers" as follows*

$$\hat{\mu}_{n+1}(x, h, g) := \frac{1}{N} \sum_{j=1}^{N} L_g(x - x_j) \hat{\mu}_n(x_j, h, g), \quad n \ge 1.$$

Using the recursion above

$$\hat{\mu}_{n+1}(x,h,g) = \frac{1}{N^n} \sum_{j_1,\dots,j_n} L_g(x - x_{j_1}) \dots L_g(x_{j_{n-1}} - x_{j_n}) \hat{\mu}_K(x_{j_n}, h).$$

The "iterated differences" are defined recursively as follows

$$\hat{\nu}_{n+1}(x,h,g) := \hat{\mu}_{n+1}(x,h,g) - \hat{\mu}_n(x,h,g), \quad n \geq 1.$$

To get an asymptotic approximation of the regularized local least-squares estimate u_i, Franke et al. first investigated the asymptotic properties of the iterated smoothers $\hat{\mu}_n(x,h,g)$, $n \geq 1$. For that purpose, they assumed the following

(A2) a) μ is twice continuously differentiable.

b) $\mu''(x)$ is Hölder continuous on $[0,1]$ with exponent β, i.e. for some $\beta > 0$, $H < \infty$ $|\mu''(x) - \mu''(y)| \leq H|x-y|^\beta$ for all $x, y \in [0,1]$.

These assumptions, along with the previous set of assumptions, will help us in getting the bias and the variance terms for each iterated smoother, and the covariance terms between any two iterated smoothers.

Proposition 1.11 *Assuming (A1) a)-e) and (A2), we have for the Priestley-Chao estimate $\hat{\mu}_K(x,h)$ (denoting $\hat{\mu}_K(x_i,h)$ by $\hat{\mu}_i$), for $N \to \infty$, $h \to 0$ such that $Nh \to \infty$,*

i) bias $\hat{\mu}_i = \mathbb{E}\,\hat{\mu}_i - \mu(x_i) = \frac{h^2}{2}\mu''(x_i)V_K + O\left(h^{2+\beta}\right) + O\left(\frac{1}{N^2h^2}\right)$ *uniformly in $x_i \in [h, 1-h]$.*

ii) var $\hat{\mu}_i = \mathbb{E}\,(\hat{\mu}_i - \mathbb{E}\,\hat{\mu}_i)^2 = \frac{\sigma^2}{Nh}Q_K + O\left(\frac{1}{N^3h^3}\right)$ *uniformly in $x_i \in [h, 1-h]$.*

iii) mse $\hat{\mu}_i = \mathbb{E}\,(\hat{\mu}_i - \mu(x_i))^2 = \frac{\sigma^2}{Nh}Q_K + \frac{h^4}{4}\{\mu''(x_i)\}^2 V_K^2 + O\left(h^{4+2\beta}\right) + O\left(\frac{1}{N^3h^3}\right)$ *uniformly in $x_i \in [h, 1-h]$. In particular*

$$\hat{\mu}_i - \mu(x_i) \xrightarrow{P} 0.$$

iv) cov $(\hat{\mu}_i, \hat{\mu}_k) = 0$ *if $|x_i - x_k| > 2h$, and*
cov $(\hat{\mu}_i, \hat{\mu}_k) = \frac{\sigma^2}{Nh} K * K\left(\frac{x_k - x_i}{h}\right) + O\left(\frac{1}{N^3h^3}\right)$ *uniformly in $x_i, x_k \in [h, 1-h]$, else, where $K * K$ denotes the convolution of K with itself.*

Theorem 1.12 *Let the model (1.4) hold. Let K, L satisfy (A1) a)-e) and let (A2) hold. Then we have for $N \to \infty$, $h, g, \lambda \to 0$ such that $Nh \to \infty$, $Ng \to \infty$*

$$u_i = (1-\theta) \sum_{k=0}^{t} \theta^k \hat{\mu}_{k+1}(x_i, h, g) + R_{N,i}, \tag{1.12}$$

where the remainder term satisfies uniformly in $\max(h,g) + tg \leq x_i \leq 1 - \max(h,g) - tg$

$$R_{N,i} = O_p(\lambda^{t+1}) + O_p\left(\frac{1}{N^2h^2}\right) + O_p\left(\frac{\lambda}{N^2g^2}\right), \quad \textit{and} \quad \theta = \frac{\lambda}{1+\lambda}.$$

Lemma 1.13 *Assume that K and L satisfy (A1) a)-f), and that μ satisfies (A2). Then, if $h, g \to 0$, $Ng^4, Nh^4 \to \infty$ for $N \to \infty$, we have for all $n \geq 1$ uniformly in $h + ng \leq x \leq 1 - (h + ng)$*

$$\mathbb{E}\,\hat{\nu}_{n+1}(x, h, g) = \text{ bias}\,\hat{\mu}_L(x, g) + o(g^2).$$

Theorem 1.14 *Let the model (1.4) hold. Let K, L satisfy (A1) a)-f), and let (A2) hold. Let, for $N \to \infty$, $h, g, \lambda \to 0$, such that $Nh^4, Ng^4 \to \infty$. Then, with t chosen as the smallest integer satisfying $\lambda^t = o(g^2)$, we have uniformly for all i satisfying $h + tg \leq x_i \leq 1 - (h + tg)$*

$$\begin{aligned}\text{bias}\, u_i &= \mathbb{E}\, u_i - \mu(x_i) \\ &= \text{ bias}\,\hat{\mu}_K(x_i, h) + \lambda\,\text{bias}\,\hat{\mu}_L(x_i, g) + o\left(\lambda g^2\right) + O\left(\frac{1}{N^2h^2}\right) \\ &= \frac{1}{2}\mu''(x_i)\left\{h^2 V_K + \lambda g^2 V_L\right\} + O(h^{\beta+2}) + o\left(\lambda g^2\right) + O\left(\frac{1}{N^2h^2}\right).\end{aligned}$$

Proposition 1.15 *Assume that K and L satisfy (A1) a)-e), and that μ satisfies (A2). Then, if $h, g \to 0$, $Ng, Nh \to \infty$ for $N \to \infty$, for all $n \geq m \geq 0$,*

$$\begin{aligned}&\text{cov}\left(\hat{\mu}_{m+1}(x, h, g), \hat{\mu}_{n+1}(\bar{x}, h, g)\right) \\ &= \frac{\sigma^2}{N} L_g^{*(n+m)} * K_h * K_h(x - \bar{x}) + O\left(\frac{1}{N^3h^3}\right) + O\left(\frac{1}{N^3g^2h \vee g}\right)\end{aligned}$$

uniformly in $ng + 2h \leq x, \bar{x} \leq 1 - (ng + 2h)$. In particular,

$$\begin{aligned}&\text{var}\,\hat{\mu}_{n+1}(x, h, g) \\ &= \frac{\sigma^2}{N} L_g^{*(2n)} * K_h * K_h(0) + O\left(\frac{1}{N^3h^3}\right) + O\left(\frac{1}{N^3g^2h \vee g}\right).\end{aligned}$$

We define the Fourier transforms of K, L as follows

$$\widehat{L}(\omega) := \int L(z)e^{-i\omega z}dz, \quad \widehat{K}(\omega) := \int K(z)e^{-i\omega z}dz. \tag{1.13}$$

Theorem 1.16 *Let the model (1.4) hold. Let K and L satisfy (A1) a)-e). Let μ satisfy (A2). For $N \to \infty$, let $h, g, \lambda \to 0$ such that $Nh^4 \to \infty$, $Ng^4 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^t = O\left(\frac{1}{N^2g^2}\right), \tag{1.14}$$

we have uniformly in all i satisfying $2h + tg \leq x_i \leq 1 - (2h + tg)$

$$\text{var}\, u_i = \frac{\sigma^2}{Nh} Q\left(\frac{g}{h}, \lambda\right)$$

$$+ O\left(\frac{1}{(Nh)^{5/2}}\right) + O\left(\frac{\lambda}{N^3 g^2 h \vee g}\right) + O\left(\frac{\lambda}{N^2 g^2}\right),$$

and

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

Proposition 1.17 *Assuming model (1.4) as well as (A1) a)-e) and (A2), we have for* $t \geq 1$ *and all* $0 < x < 1$,

$$\sqrt{N}\,\frac{\hat{\mu}_t(x,h,g) - \mathbb{E}\,\hat{\mu}_t(x,h,g)}{\sqrt{L_g^{*(2t-2)} * K_h * K_h(0)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

for $N \to \infty$, $h, g \to 0$ *such that* $Ng, Nh \to \infty$.

Theorem 1.18 *a) Under the assumptions of Theorem 1.16, we have for* $0 < x < 1$

$$\sqrt{Nh}\,\frac{u(x) - \mathbb{E}\,u(x)}{\sqrt{Q\left(\frac{g}{h},\lambda\right)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2) \quad \text{for} \quad N \to \infty$$

with $Q\left(\frac{g}{h},\lambda\right) = Q_K + O(\lambda)$.

b) If, additionally, the assumptions of Theorem 1.14 are satisfied

$$\operatorname{bias} u(x) = \mathbb{E}\,u(x) - \mu(x) = \frac{1}{2}\mu''(x)\left\{h^2 V_K + \lambda g^2 V_L\right\} + R'_N$$

with remainder $R'_N = O(h^{2+\beta}) + o(\lambda g^2) + O\left(\frac{1}{N^2h^2}\right)$ *uniformly in* $2h+tg \leq x \leq 1-2h-tg-\frac{1}{N}$.

Combining both parts of the theorem, we get

$$u(x) - \mu(x) \overset{\mathcal{L}}{\approx} \mathcal{N}\left(\frac{1}{2}\mu''(x)\left\{h^2V_K + \lambda g^2 V_L\right\}, \frac{\sigma^2}{Nh} Q\left(\frac{g}{h},\lambda\right)\right).$$

1.7 Regularized Local Huber M-Estimates

In this section we will give an outline of the next chapters. The goal of the next chapters is to consider the case of the *Quadratically Regularized Local Huber M-estimate.* Assuming model (1.4) the QRLHM case is derived from the general approach by Mrázek et al. [25] by choosing

$$\Psi_D(s^2) = \rho(s), \quad w_D(x^2) = K_h(x),$$

where ρ is the Huber function given by (1.8). We first study the case of quadratic regularization with a kernel weight, i.e.

$$\Psi_S(s^2) = s^2, \quad w_S(x^2) = L_g(x).$$

Later on we can see different choices for the loss function Ψ_S.

The kernels K and L are standardized nonnegative, symmetric functions on $\mathbb{R}$ and the bandwidths $h, g > 0$ can be chosen to control the smoothness of the function estimate together with the balancing factor λ. Therefore, the QRLHM minimization problem can be written as

$$\begin{aligned} Q(u_1, \ldots, u_N) = &\sum_{i,j=1}^{N} \rho(u_i - f_j) K_h(x_i - x_j) \\ &+ \frac{\lambda}{2} \sum_{i,j=1}^{N} \frac{1}{2}(u_i - u_j)^2 L_g(x_i - x_j) = \min_{u_1,\ldots,u_N}! \end{aligned} \tag{1.15}$$

and the solution is called the QRLHM-estimate.

The solution here does not have an explicit form like Proposition 1.8. For that reason we will try to get an approximation to the solution. This will be done by using a Taylor series expansion around the "Local Huber M-estimate" (the LHM-estimate is the QRLHM-estimate in the case $\lambda = 0$). The Taylor expansion used here is analogous to the Newton method for solving a system of equations, but we do not iterate here since the function of interest is at most quadratic.

To get the asymptotic properties (bias, variance, distribution) of the QRLHM-estimate we will establish some results similar to those in Franke et al. [10], which were presented in the previous section.

Chapter 2

Some Asymptotics of Local Huber M-Estimates (LHM-Estimates)

In this chapter we shall see some asymptotic properties of the M-estimates under the deterministic equidistant design model. In M-estimation many choices for the target function to be minimized are available. Our target function here is going to be the Huber function [16]. For localization, there are various choices as well, but we do the localization here using kernel weights.

Huber [16] provided some asymptotic properties of M-estimates without localization. Stützle and Mittal [33] gave some comments on the method as a generalization of kernel-type smoothers where they provided asymptotic rates for the bias and the variance. Härdle [13] and Fan et al. [7] provided some asymptotic properties of M-estimates under a different setting, where they considered the random design. Härdle and Gasser [14] showed some asymptotic properties of M-estimates under the fixed design but using the Gasser-Müller weights for localization. Chu et al. [4] considered M-estimates with kernel weights for localization but they used the kernel function as the tonal weight function. In their work they have assumed that the regression function to be estimated has four Lipschitz continuous derivatives. This assumption is relaxed here to two continuous derivatives where the second derivative is Hölder continuous.

According to Mrázek et al. terminology [25] the spatial weight function is the function responsible for localization and the tonal weight function is the function responsible for the quality of the estimate.

2.1 Setup of the Problem

We assume that our data (x_j, f_j), $j = 1, \ldots, N$, comes from the nonparametric regression model:

$$f_j = \mu(x_j) + \varepsilon_j, \quad j = 1, \ldots, N, \tag{2.1}$$

where $\varepsilon_j \sim$ i. i. d. $(0, \sigma^2)$, and $x_j = \frac{j}{N}$ form an equidistant grid in the unit interval $[0, 1]$.

We are interested here in investigating some of the asymptotic properties of the *local Huber M-estimate*, abbreviated as the LHM-estimate.

The LHM-estimate at a point x is denoted by $\tilde{\mu}(x)$, and it is defined implicitly as the solution of

$$\sum_{j=1}^{N} \rho(u - f_j) K_h(x - x_j) = \min_{u\in\mathbb{R}}! \tag{2.2}$$

or, equivalently, as the solution of

$$\sum_{j=1}^{N} \psi(u - f_j) K_h(x - x_j) = 0, \tag{2.3}$$

with respect to u, where ψ is the derivative of the Huber function [16],

$$\rho(u) = \begin{cases} \frac{1}{2}u^2, & |u| \le c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} \tag{2.4}$$

and K is a kernel function that is nonnegative and symmetric on $\mathbb{R}$, and the bandwidth $h > 0$. The equivalence between the two problems above is due to the convexity of ρ. To tell which kernel and bandwidth we are using we may also denote the LHM-estimate by $\tilde{\mu}_K(x, h)$ or $\tilde{\mu}(x, h)$.

The LHM-estimate belongs to a larger class of estimates known as the *M-smoothers*. An M-smoother at a point x is defined implicitly as the solution of

$$\sum_{j=1}^{N} \psi_D(u - f_j) K_h(x - x_j) = 0 \tag{2.5}$$

with respect to u, where $\psi_D : \mathbb{R} \to \mathbb{R}$ is a bounded, monotone, antisymmetric function.

Alternatively, the LHM-estimate could be obtained using the general approach for image denoising proposed by Mrázek et al. [25] with $\lambda = 0$. That is, by considering the problem:

$$Q(u_1, \ldots, u_N) = \sum_{i,j=1}^{N} \rho(u_i - f_j) K_h(x_i - x_j) = \min_{u_1,\ldots,u_N}! \tag{2.6}$$

The solution of the problem is a vector whose entries are the LHM-estimates at the grid points $x_j = j/N$, i.e. $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_N)^T = (\tilde{\mu}_K(x_1, h), \ldots, \tilde{\mu}_K(x_N, h))^T$. To get the LHM-estimate at any point x in the interval $[0, 1]$ we have to interpolate.

In the context of estimation with regularization we may call the *M-smoothers* without regularization *pure M-smoothers*.

For further analysis, it is useful to calculate the derivatives of ρ. The first two derivatives of ρ are

$$\rho'(u) = \begin{cases} c, & u > c, \\ u, & |u| \le c, \\ -c, & u < -c, \end{cases} \quad \text{and} \quad \rho''(u) = \begin{cases} 0, & |u| > c, \\ 1, & |u| < c, \\ \text{DNE}, & |u| = c, \end{cases} \tag{2.7}$$

while all higher derivatives

$$\rho^{(k)}(u) = \begin{cases} 0, & |u| \neq c, \\ \text{DNE}, & |u| = c, \end{cases} \quad \text{for all} \quad k \geq 3. \tag{2.8}$$

The term "DNE" stands for "does not exist".

Let us now define the indicator function as follows: Let $A \subseteq \mathbb{R}$ then

$$\mathbb{1}_A(u) = \begin{cases} 1, & u \in A, \\ 0, & u \notin A. \end{cases} \tag{2.9}$$

Using this definition we note that for all $u \in \mathbb{R} \setminus \{-c, c\}$

$$\rho''(u) = \mathbb{1}_{(-c,c)}(u) \qquad \text{and} \qquad \rho^{(k)}(u) = 0 \quad \text{for all} \quad k \geq 3. \tag{2.10}$$

Equivalently, we write

$$\rho''(u) = \mathbb{1}_{(-c,c)}(u) \quad a.s. \quad \text{and} \quad \rho^{(k)}(u) = 0 \quad a.s. \quad \text{for all} \quad k \geq 3,$$

where "a.s." stands for "almost surely with respect to the probability measure of ε_j".

However, choosing the Huber function as given by (2.4) means that $\rho(u) \stackrel{c\to\infty}{\longrightarrow} \frac{1}{2}u^2$, but $\rho(u) \stackrel{c\to 0}{\longrightarrow} 0$. Of course, this is undesired. It would be more interesting if the second limit tends to the absolute-value function instead of zero.

To get over this problem, we redefine the Huber function in the following manner,

$$\rho(u) := \begin{cases} \begin{cases} \frac{1}{2}u^2, & |u| \leq c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} & c \geq 1, \\ \\ \begin{cases} \frac{1}{2c}u^2, & |u| \leq c, \\ |u| - \frac{1}{2}c, & |u| > c, \end{cases} & c \leq 1. \end{cases} \tag{2.11}$$

We call the new ρ the modified Huber function. Using the modified Huber function it is now clear that $\rho(u) \stackrel{c\to\infty}{\longrightarrow} \frac{1}{2}u^2$, and $\rho(u) \stackrel{c\to 0}{\longrightarrow} |u|$. Hence, we capture both L_2 and L_1 cases as limit cases to our minimization problem.

Mark that (2.4) and (2.11) differ only by a positive factor for $0 < c < 1$, such that our modification of the distance ρ makes no difference for the estimate which we get as a solution of a minimization problem like (2.6).

2.2 Assumptions and Notation

Throughout our work we assume that we are dealing with a kernel function that satisfies the following assumptions.

(A1) a) K is a nonnegative, symmetric kernel function with compact support $[-1, 1]$.

b) $\int K(u)du = 1$.

c) K is Lipschitz continuous with Lipschitz constant C_K.

Wals and Sewell [37] gave the following useful analytical tool that will enable us to interchange between Riemann integrals and Riemann sums.

Theorem 2.1 (Wals and Sewell, [37]) *Let $g(x)$ be continuous in the interval $[0,1]$ and possess there the modulus of continuity $\omega(g,\delta)$ in the sense that for values x and y in the interval $(0,1)$ the inequality $|x-y| \leq \delta$ implies $|g(x)-g(y)| \leq \omega(\delta)$. Then we have*

$$\left| \int_0^1 g(x)dx - \frac{1}{N}\sum_{k=1}^{N} g\left(\frac{k}{N}\right) \right| \leq \omega\left(g, \frac{1}{N}\right).$$

The modulus of continuity of a function g on an interval I is formally defined as

$$\omega(g,\delta) = \sup_{x,y\in I \ : \ |x-y|<\delta} |g(x)-g(y)|.$$

If g is Lipschitz continuous with Lipschitz constant M then

$$\omega(g,\delta) \leq M\delta.$$

Hence, the result holds true in particular for Lipschitz continuous functions.

Corollary 2.2 *Let $g(x)$ be Lipschitz continuous in the interval $[0,1]$ with Lipschitz constant M. Let $x_j = \frac{j}{N}$ for all $j = 1,\ldots,N$. Then we have*

$$\left| \int_0^1 g(x)dx - \frac{1}{N}\sum_{j=1}^{N} g(x_j) \right| \leq \frac{M}{N}.$$

A useful consequence of this corollary is that we can interchange the following sums with their Riemann integrals.

Lemma 2.3 *Let the kernel K satisfy (A1) a)-c). Let $x_j = \frac{j}{N}$ for $j = 1,\ldots,N$. For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then,*

$$(1) \quad \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j) = 1 + O\left(\frac{1}{Nh^2}\right)$$

$$(2) \quad \frac{1}{N^2}\sum_{j=1}^{N} K_h^2(x-x_j) = \frac{Q_K}{Nh} + O\left(\frac{1}{N^2h^3}\right) = \frac{Q_K}{Nh} + o\left(\frac{1}{Nh}\right)$$

$$(3) \quad \frac{1}{N^3}\sum_{j=1}^{N} K_h^3(x-x_j) = \frac{S_K}{N^2h^2} + O\left(\frac{1}{N^3h^4}\right) = \frac{S_K}{N^2h^2} + o\left(\frac{1}{N^2h^2}\right)$$

uniformly in $x \in [h, 1-h]$, *where*

$$Q_K = \int_{-1}^{1} K^2(u)du \quad \text{and} \quad S_K = \int_{-1}^{1} K^3(u)du.$$

Proof. The proof depends on the previous corollary. Equation (1) is clear since $g(\cdot) = K_h(x - \cdot)$ is Lipschitz continuous with Lipschitz constant C_K/h^2 and K integrates to 1 for all $x \in [h, 1-h]$. That is,

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j) = \int_0^1 K_h(x-y)dy + O\left(\frac{1}{Nh^2}\right) = \int_{\frac{-x}{h}}^{\frac{1-x}{h}} K(z)dz + O\left(\frac{1}{Nh^2}\right).$$

For all $x \in [h, 1-h]$ we have

$$\frac{-x}{h} \leq -1 \quad \text{and} \quad \frac{1-x}{h} \geq 1,$$

therefore

$$\int_{\frac{-x}{h}}^{\frac{1-x}{h}} K(z)dz = \int_{\frac{-x}{h}}^{-1} K(z)dz + \int_{-1}^{1} K(z)dz + \int_{1}^{\frac{1-x}{h}} K(z)dz = 1.$$

Similarly, equation (2) holds true since $g(\cdot) = K_h^2(x - \cdot)$ is Lipschitz continuous with Lipschitz constant of order $1/h^3$, and equation (3) holds true since $g(\cdot) = K_h^3(x - \cdot)$ is Lipschitz continuous with Lipschitz constant of order $1/h^4$. □

Lemma 2.4 *Let the kernel* K *satisfy (A1) a)-c). Let* $x_j = \frac{j}{N}$ *for* $j = 1, \ldots, N$*. For* $N \to \infty$*, let* $h \to 0$ *such that* $Nh^2 \to \infty$*. Then,*

$$(1) \quad \frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{x-x_j}{h}\right)\left(\frac{x_j - x}{h}\right) = O\left(\frac{1}{Nh^2}\right)$$

$$(2) \quad \frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{x-x_j}{h}\right)\left(\frac{x_j - x}{h}\right)^2 = V_K + O\left(\frac{1}{Nh^2}\right)$$

uniformly in $x \in [h, 1-h]$, *where*

$$V_K = \int z^2 K(z)dz.$$

Proof. The proof is similar to the proof of Lemma 2.3. Using (A1) a) we have $\int zK(z)dz = 0$. □

Remark 2.5 *From (A1) a)-c) we get that* V_K, Q_K *and* S_K *are bounded.*

Remark 2.6 *In view of the previous two lemmas we are going to consider only the case where* $x \in [h, 1-h]$*, that is, we are neglecting the performance of the estimate at the boundaries of the unit interval. However, we have* $h \to 0$ *as* $N \to \infty$*, that is, the boundaries are vanishing asymptotically.*

Discussing the boundary effects in $(0, h)$ and $(1-h, 1)$ would be possible as for common kernel estimates (compare, e.g., Härdle [12], Section 4.4), but we do not want to go into the rather technical details. We would rather concentrate on the main ideas.

We assume that the function we are estimating satisfies,

(A2) a) μ is twice continuously differentiable.

b) $\mu''(x)$ is Hölder continuous on $[0, 1]$ with exponent β, i.e. for some $\beta > 0$, $H < \infty$

$$|\mu''(x) - \mu''(y)| \leq H|x-y|^{\beta} \quad \forall\ x, y \in [0, 1].$$

These assumptions are just technical assumptions and they determine the least smoothness properties our estimate should have.

Remark 2.7 *From assumptions (A2) a)-b) we get that the function μ is Lipschitz continuous with Lipschitz constant C_μ. Therefore, for all $x \in [h, 1-h]$ and for all $x_j = j/N$ where $j = 1, \ldots, N$ such that $|x - x_j| \leq h$ we have*

$$|\mu(x) - \mu(x_j)| \leq C_\mu |x - x_j| \leq C_\mu h \overset{N\to\infty}{\longrightarrow} 0. \tag{2.12}$$

The error terms in the regression model are assumed to satisfy the following,

(E1) a) ε_j are continuous random variables with a probability density function $p_\varepsilon(x)$ that is symmetric around zero.

b) $p_\varepsilon(x)$ is a decreasing function of $|x|$.

c) $p_\varepsilon(x)$ is twice continuously differentiable.

We will denote the cumulative distribution function (CDF) of $\{\varepsilon_j\}_{j=1\ldots,N}$ by P_ε, that is

$$P_\varepsilon(x) = \int_{-\infty}^{x} dP_\varepsilon(t) = \int_{-\infty}^{x} p_\varepsilon(t)dt.$$

Using the definition of the modified Huber function, given by (2.11), and assumptions (E1) we are interested in getting the value of the following integrals,

$$\begin{aligned} &(1) \quad \int_{-\infty}^{\infty} \psi(u)p_\varepsilon(u)du, \\ &(2) \quad \int_{-\infty}^{\infty} \psi'(u)p_\varepsilon(u)du, \\ &(3) \quad \int_{-\infty}^{\infty} \psi^2(u)p_\varepsilon(u)du. \end{aligned} \tag{2.13}$$

Direct calculation shows that

$$\int_{-\infty}^{\infty} \psi(u)p_\varepsilon(u)du = 0,$$

and

$$\int_{-\infty}^{\infty} \psi'(u)p_\varepsilon(u)du = \begin{cases} \int\limits_{-c}^{c} p_\varepsilon(u)du, & c \geq 1, \\ \frac{1}{c}\int\limits_{-c}^{c} p_\varepsilon(u)du, & c < 1. \end{cases}$$

To make notation easier we make the following definition.

Definition 2.8 *For any fixed c we define*

$$\eta = \int_{-c}^{c} p_\varepsilon(u)du, \qquad \text{and} \qquad \eta_c = \begin{cases} \eta, & c \geq 1, \\ \frac{\eta}{c}, & c < 1. \end{cases}$$

Using this definition the value of the second integral of (2.13) is

$$\int_{-\infty}^{\infty} \psi'(u)p_\varepsilon(u)du = \eta_c.$$

The value of the third integral of (2.13) is calculated as follows

$$\begin{aligned}
\int_{-\infty}^{\infty} \psi^2(u)p_\varepsilon(u)du &= \int_{-\infty}^{-c} \psi^2(u)p_\varepsilon(u)du + \int_{-c}^{c} \psi^2(u)p_\varepsilon(u)du + \int_{c}^{\infty} \psi^2(u)p_\varepsilon(u)du \\
&= \begin{cases} \int_{-\infty}^{-c}(-c)^2 p_\varepsilon(u)du + \int_{-c}^{c}(u)^2 p_\varepsilon(u)du + \int_{c}^{\infty}(c)^2 p_\varepsilon(u)du, & c \geq 1, \\ \int_{-\infty}^{-c}(-1)^2 p_\varepsilon(u)du + \int_{-c}^{c}(u/c)^2 p_\varepsilon(u)du + \int_{c}^{\infty}(1)^2 p_\varepsilon(u)du, & c < 1, \end{cases} \\
&= \left\{ c^2\left(1 - \int_{-c}^{c} p_\varepsilon(u)du\right) + \int_{-c}^{c} u^2 p_\varepsilon(u)du \right\} \cdot \begin{cases} 1, & c \geq 1, \\ \frac{1}{c^2}, & c < 1, \end{cases} \\
&= \begin{cases} c^2(1-\eta) + \int_{-c}^{c} u^2 p_\varepsilon(u)du, & c \geq 1, \\ (1-\eta) + \frac{1}{c^2}\int_{-c}^{c} u^2 p_\varepsilon(u)du, & c < 1. \end{cases}
\end{aligned}$$

An interesting value obtained from the second and the third integrals of (2.13) is

$$\begin{aligned}
\frac{\int_{-\infty}^{\infty} \psi^2(u)p_\varepsilon(u)du}{\left(\int_{-\infty}^{\infty} \psi'(u)p_\varepsilon(u)du\right)^2} &= \begin{cases} \frac{c^2(1-\eta)}{\eta^2} + \frac{1}{\eta^2}\int\limits_{-c}^{c} y^2(y)p_\varepsilon(y)dy, & c \geq 1, \\ \frac{(1-\eta)}{(\eta/c)^2} + \frac{1}{c^2(\eta/c)^2}\int\limits_{-c}^{c} y^2(y)p_\varepsilon(y)dy, & c < 1, \end{cases} \\
&= \frac{c^2(1-\eta)}{\eta^2} + \frac{1}{\eta^2}\int_{-c}^{c} y^2 p_\varepsilon(y)dy.
\end{aligned}$$

As in the context of M-estimation without localization, the ratio calculated above, namely $\int \psi^2 dP_\varepsilon/(\int \psi' dP_\varepsilon)^2$, turns out to be the asymptotic variance.

Since we are going to use these integrals often in our analysis we make the following definition.

Definition 2.9 *For any fixed c we define*

$$\sigma_M^2 = \int_{-\infty}^{\infty} \psi^2(u) p_\varepsilon(u) du, \qquad \textit{and} \qquad \sigma_c^2 = \frac{\sigma_M^2}{\eta_c^2} = \frac{c^2(1-\eta)}{\eta^2} + \frac{1}{\eta^2} \int_{-c}^{c} y^2 p_\varepsilon(y) dy.$$

2.3 Consistency

In this section we want to show that the M-estimate (with kernel spatial weights) is consistent.

Proposition 2.10 *a) Let ψ be the derivative of the modified Huber function given by (2.11). Define*

$$H_N(x,s) = \frac{1}{N} \sum_{j=1}^{N} K_h(x - x_j) \psi(f_j - s).$$

Then $H_N(x,s)$ is non-increasing in s.

b) Moreover, let the model (2.1) hold. Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a). For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then, for all $x \in [h, 1-h]$

$$\operatorname{var} H_N(x) = \frac{Q_K}{Nh} \sigma_M^2 + o\left(\frac{1}{Nh}\right),$$

where $H_N(x) = H_N(x, \mu(x))$.

Proof. Consider $s_1, s_2 \in \mathbb{R}$. Since ψ is non-decreasing we get

$$\begin{aligned} s_1 < s_2 &\implies f_j - s_1 > f_j - s_2 \\ &\implies \psi(f_j - s_1) \geq \psi(f_j - s_2) \\ &\implies H_N(x, s_1) \geq H_N(x, s_2). \end{aligned}$$

Since $\varepsilon_1, \ldots, \varepsilon_N$ are independent, then for all $j, k = 1, \ldots, N$ such that $j \neq k$ we get

$$\begin{aligned} &\operatorname{cov}\{\psi(f_j - \mu(x)), \psi(f_k - \mu(x))\} \\ &= \operatorname{cov}\{\psi(\varepsilon_j + \mu(x_j) - \mu(x)), \psi(\varepsilon_k + \mu(x_k) - \mu(x))\} \\ &= \int\int \psi(y + \mu(x_j) - \mu(x))\ \psi(z + \mu(x_k) - \mu(x))\ p_{\varepsilon_j,\varepsilon_k}(y,z) dy dz \\ &\quad - \left(\int \psi(y + \mu(x_j) - \mu(x)) p_{\varepsilon_j}(y) dy\right) \cdot \left(\int \psi(z + \mu(x_k) - \mu(x)) p_{\varepsilon_k}(z) dz\right) \\ &= \int\int \psi(y + \mu(x_j) - \mu(x))\ \psi(z + \mu(x_k) - \mu(x))\ p_{\varepsilon_j}(y)\ p_{\varepsilon_k}(z) dy dz \\ &\quad - \left(\int \psi(y + \mu(x_j) - \mu(x)) p_{\varepsilon_j}(y) dy\right) \cdot \left(\int \psi(z + \mu(x_k) - \mu(x)) p_{\varepsilon_k}(z) dz\right) \\ &= 0. \end{aligned}$$

Also, since $\varepsilon_1, \dots, \varepsilon_N$ are identically distributed and using the continuity of ψ and the Lipschitz continuity of μ we have for every $|x - x_j| \le h$

$$\begin{aligned}
\operatorname{var} \psi(f_j - \mu(x)) &= \operatorname{var} \psi(\varepsilon_j + \mu(x_j) - \mu(x)) \\
&= \mathbb{E}\, \psi^2(\varepsilon_j + \mu(x_j) - \mu(x)) - (\mathbb{E}\, \psi(\varepsilon_j + \mu(x_j) - \mu(x)))^2 \\
&= \int \psi^2(u + \mu(x_j) - \mu(x)) dP_\varepsilon(u) - \left(\int \psi(u + \mu(x_j) - \mu(x)) dP_\varepsilon(u) \right)^2 \\
&= \int \psi^2(u) dP_\varepsilon(u) - \left(\int \psi(u) dP_\varepsilon(u) \right)^2 + o(1) \\
&= \int \psi^2(u) dP_\varepsilon(u) + o(1) \\
&= \sigma_M^2 + o(1).
\end{aligned}$$

From Lemma 2.3 we have

$$\begin{aligned}
\operatorname{var} H_N(x) &= \frac{1}{N^2} \sum_{j=1}^{N} K_h^2(x - x_j) \operatorname{var} \psi(\mu(x_j) - \mu(x) + \varepsilon_j) \\
&= \left\{ \frac{Q_K}{Nh} + o\left(\frac{1}{Nh} \right) \right\} \cdot \left\{ \sigma_M^2 + o(1) \right\} \\
&= \frac{Q_K}{Nh} \sigma_M^2 + o\left(\frac{1}{Nh} \right).
\end{aligned}$$

□

A useful tool for proving consistency of the LHM-estimate is the law of large numbers (for example, see [29]).

Theorem 2.11 (WLLN & SLLN) *Let Γ_j be a sequence of independent random variables with $\mathbb{E}\, \Gamma_j = \gamma_j$ and $\operatorname{var} \Gamma_j = v_j^2$.*

(Chebyshev) *If* $\sum_{j=1}^{N} v_j^2 = o(N^2)$ *then* $\frac{1}{N} \sum_{j=1}^{N} \Gamma_j - \frac{1}{N} \sum_{j=1}^{N} \gamma_j \xrightarrow{P} 0.$

(Kolmogorov) *If* $\sum_{j=1}^{N} \frac{1}{j^2} v_j^2$ *converges then* $\frac{1}{N} \sum_{j=1}^{N} \Gamma_j - \frac{1}{N} \sum_{j=1}^{N} \gamma_j \to 0$ *a.s.*

Now, we can prove that the LHM-estimate is consistent.

Theorem 2.12 (LHM Consistency) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy*

(E1) a)-b). For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then,

$$\tilde{\mu}(x) \xrightarrow{P} \mu(x)$$

for all $x \in [h, 1-h]$.

If additionally, $\sum_{j=1}^{N} \frac{1}{j^2} K_h^2(x - x_j)$ converges then,

$$\tilde{\mu}(x) \to \mu(x) \quad a.s.$$

Proof. To have consistency we need to show that

$$\mathbb{P}(\tilde{\mu}(x) - \mu(x) > \delta) \to 0 \quad \text{and} \quad \mathbb{P}(\tilde{\mu}(x) - \mu(x) < -\delta) \to 0 \quad \text{for all} \quad \delta > 0.$$

We use Chebyshev's LLN (Theorem 2.11) with

$$\Gamma_j = K_h(x - x_j)\psi(f_j - \mu(x)) = K_h(x - x_j)\psi(\mu(x_j) - \mu(x) + \varepsilon_j).$$

The random variables Γ_j are independent since ε_j are independent and

$$\mathbb{E}\,\Gamma_j = K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x)).$$

From the proof of Proposition 2.10

$$\begin{aligned} \frac{1}{N^2}\sum_{j=1}^{N} \operatorname{var}\Gamma_j &= \frac{1}{N^2}\sum_{j=1}^{N} K_h^2(x - x_j)\operatorname{var}\psi(\mu(x_j) - \mu(x) + \varepsilon_j) \\ &= \frac{Q_K}{Nh}\sigma_M^2 + o\left(\frac{1}{Nh}\right). \end{aligned}$$

Chebyshev's LLN (Theorem 2.11) implies that

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x) - \delta) - \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x) - \delta) \xrightarrow{P} 0,$$

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x)) - \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x)) \xrightarrow{P} 0,$$

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x) + \delta) - \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x) + \delta) \xrightarrow{P} 0.$$

Assuming (E1) a)-b) and that ψ is anti-symmetric around zero we get from Theorem 10.2 in [24] that

$$\mathbb{E}\,\rho(\varepsilon_1 - \delta) \quad \text{has a unique minimum at } \delta = 0,$$

where ρ is any even function. In particular that is true when ρ is the modified Huber function. Therefore,

$$\mathbb{E}\,\psi(\varepsilon_1 - \delta) < 0 \quad \text{and} \quad \mathbb{E}\,\psi(\varepsilon_1 + \delta) > 0 \quad \text{for all} \quad \delta > 0.$$

Using Lemma 2.3 and since ε_j have identical distributions, and ψ is continuous, the above three limits become

$$\begin{aligned} H_N(x, \mu(x) + \delta) &\xrightarrow{P} \mathbb{E}\,\psi(\varepsilon_1 - \delta) < 0, \\ H_N(x, \mu(x)) &\xrightarrow{P} \mathbb{E}\,\psi(\varepsilon_1) = 0, \\ H_N(x, \mu(x) - \delta) &\xrightarrow{P} \mathbb{E}\,\psi(\varepsilon_1 + \delta) > 0. \end{aligned}$$

Since H_N is non-increasing in the second argument (Proposition 2.10) we have

$$\tilde{\mu}(x) > \mu(x) + \delta \Longrightarrow H_N(x, \tilde{\mu}(x)) \leq H_N(x, \mu(x) + \delta) \Longleftrightarrow H_N(x, \mu(x) + \delta) \geq 0.$$

That is,

$$\lim_{N\to\infty} \mathbb{P}(\tilde{\mu}(x) > \mu(x) + \delta) \leq \lim_{N\to\infty} \mathbb{P}(H_N(x, \mu(x) + \delta) \geq 0),$$

and

$$\lim_{N\to\infty} \mathbb{P}(H_N(x, \mu(x) + \delta) \geq 0) = 0$$

since $H_N(x, \tilde{\mu}(x) + \delta) \xrightarrow{P} \mathbb{E}\,\psi(\varepsilon_1 - \delta)$ for all $\delta > 0$ and $\mathbb{E}\,\psi(\varepsilon_1 - \delta)$ is strictly less than zero.

Analogously,

$$\tilde{\mu}(x) < \mu(x) - \delta \Longrightarrow H_N(x, \tilde{\mu}(x)) \geq H_N(x, \mu(x) - \delta) \Longleftrightarrow H_N(x, \mu(x) - \delta) \leq 0.$$

$$\lim_{N\to\infty} \mathbb{P}(\tilde{\mu}(x) < \mu(x) - \delta) \leq \lim_{N\to\infty} \mathbb{P}(H_N(x, \mu(x) - \delta) \leq 0) = 0.$$

Hence,

$$\lim_{N\to\infty} \mathbb{P}(|\tilde{\mu}(x) - \mu(x)| > \delta) = 0.$$

□

2.4 Bias and Variance

Using the mean value theorem

$$\tilde{\mu}(x) - \mu(x) = \frac{H_N(x)}{D_N(x)} \quad a.s. \tag{2.14}$$

where $H_N(x)$ is defined as in Proposition 2.10, i.e.

$$H_N(x) = H_N(x, \mu(x)) = \frac{1}{N} \sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x)), \tag{2.15}$$

and $D_N(x)$ is defined as

$$D_N(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\psi'(f_j - \mu(x) + w_j), \tag{2.16}$$

where $|w_j| < |\tilde{\mu}(x) - \mu(x)|$. Note that (2.14) holds true only almost surely since ψ is only almost everywhere differentiable.

The variance of $H_N(x)$ is already given in Proposition 2.10. So, we will calculate the expected value of $H_N(x)$, then we will prove that $D_N(x)$ converges in probability to η_c. Consequently, we will see that the bias and the variance are given by

$$\operatorname{bias}\tilde{\mu}(x) = \frac{1}{\eta_c}\mathbb{E}\, H_N(x)\cdot(1+o(1)) \quad \text{and} \quad \operatorname{var}\tilde{\mu}(x) = \frac{1}{\eta_c^2}\operatorname{var} H_N(x)\cdot(1+o(1)),$$

if the bandwidth h is chosen appropriately.

Using this notation we can refer to the dominant part of $\frac{1}{\eta_c}\mathbb{E}\, H_N(x)$ as the "asymptotic bias term" and to the dominant term of $\frac{1}{\eta_c^2}\operatorname{var} H_N(x)$ as the "asymptotic variance term".

So, let us start by calculating the expected value of $H_N(x)$.

Proposition 2.13 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $Nh^3 \to \infty$. Then,*

$$\begin{aligned} B_N(x) &= \mathbb{E}\, H_N(x) = \mathbb{E}\, H_N(x, \mu(x)) \\ &= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)(\mu(x_j) - \mu(x))\eta_c + o(h^2) \\ &= \frac{1}{2}h^2\mu''(x)V_K\eta_c + o(h^2) \end{aligned}$$

uniformly in $x \in [h, 1-h]$.

Proof. Consider the case $c \geq 1$, the other case is completely analogous.

$$\begin{aligned} B_N(x) &= \mathbb{E}\, H_N(x) = \mathbb{E}\, H_N(x, \mu(x)) \\ &= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\mathbb{E}\,\psi(f_j - \mu(x)) \\ &= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\left\{\int_{\mathbb{R}} \psi(\mu(x_j) - \mu(x) + u)p_\varepsilon(u)du\right\} \\ &= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\left\{\int_{I_1}(-c)dP_\varepsilon(u) + \int_{I_2}(\mu(x_j) - \mu(x) + u)dP_\varepsilon(u) + \int_{I_3}(c)dP_\varepsilon(u)\right\} \end{aligned}$$

where

$$I_1 = (-\infty, -c + \mu(x) - \mu(x_j)],$$
$$I_2 = [-c + \mu(x) - \mu(x_j), c + \mu(x) - \mu(x_j)],$$
$$I_3 = [c + \mu(x) - \mu(x_j), \infty).$$

The goal now is to calculate the three integrals above. We will denote them by $\int_{I_1}$, $\int_{I_2}$, and $\int_{I_3}$ respectively. Using Remark 2.7 we can see that as N grows very largely, the intervals tend to the simple intervals $(-\infty, -c]$, $[-c, c]$, $[c, \infty)$.

We start with $\int_{I_1}$.

$$\int_{I_1} = \int_{I_1} (-c) dP_\varepsilon(u) = -c \int_{-\infty}^{-c+\mu(x)-\mu(x_j)} dP_\varepsilon(u) = -cP_\varepsilon(-c + \mu(x) - \mu(x_j)).$$

From Remark 2.7 we get that $|\mu(x) - \mu(x_j)| \overset{N\to\infty}{\longrightarrow} 0$ for all $x \in [h, 1-h]$ and all $x_j = j/N$ such that $|x - x_j| \le h$. Using this fact, we may expand P_ε around $-c$ as follows,

$$\begin{aligned} P_\varepsilon(-c + \mu(x) - \mu(x_j)) &= P_\varepsilon(-c) + (\mu(x) - \mu(x_j))p_\varepsilon(-c) \\ &\quad + \frac{1}{2}(\mu(x) - \mu(x_j))^2 p'_\varepsilon(-c) + o\left(|\mu(x) - \mu(x_j)|^2\right), \end{aligned}$$

then

$$\int_{I_1} = -cP_\varepsilon(-c) - c(\mu(x) - \mu(x_j))p_\varepsilon(-c) - \frac{c}{2}(\mu(x) - \mu(x_j))^2 p'_\varepsilon(-c) + o\left(|\mu(x) - \mu(x_j)|^2\right).$$

Analogously,

$$\int_{I_3} = \int_{I_3} (c) dP_\varepsilon(u) = c \int_{c+\mu(x)-\mu(x_j)}^{\infty} dP_\varepsilon(u) = c(1 - P_\varepsilon(c + \mu(x) - \mu(x_j))).$$

Now, we expand P_ε around c as follows,

$$\begin{aligned} P_\varepsilon(c + \mu(x) - \mu(x_j)) &= P_\varepsilon(c) + (\mu(x) - \mu(x_j))p_\varepsilon(c) \\ &\quad + \frac{1}{2}(\mu(x) - \mu(x_j))^2 p'_\varepsilon(c) + o\left(|\mu(x) - \mu(x_j)|^2\right), \end{aligned}$$

then

$$\int_{I_3} = c(1 - P_\varepsilon(c)) - c(\mu(x) - \mu(x_j))p_\varepsilon(c) - \frac{c}{2}(\mu(x) - \mu(x_j))^2 p'_\varepsilon(c) + o\left(|\mu(x) - \mu(x_j)|^2\right).$$

Summing up the integral over I_1 and I_3 we have

$$\begin{aligned} \int_{I_1 \cup I_3} &= c(1 - P_\varepsilon(c) - P_\varepsilon(-c)) - c(\mu(x) - \mu(x_j))[p_\varepsilon(c) + p_\varepsilon(-c)] \\ &\quad - \frac{c}{2}(\mu(x) - \mu(x_j))^2 [p'_\varepsilon(c) + p_\varepsilon(-c)] + o\left(|\mu(x) - \mu(x_j)|^2\right). \end{aligned}$$

By (E1) a), we get that: P_ε is a symmetric distribution function, p_ε is a symmetric function around zero, and p'_ε is an anti-symmetric function, i.e.

$$P_\varepsilon(-c) = 1 - P_\varepsilon(c), \quad p_\varepsilon(-c) = p_\varepsilon(c), \quad \text{and} \quad p'_\varepsilon(-c) = -p'_\varepsilon(c).$$

This reduces the integral over $I_1 \cup I_3$ to

$$\int_{I_1 \cup I_3} = \; - 2c(\mu(x) - \mu(x_j))p_\varepsilon(c) + o\left(|\mu(x) - \mu(x_j)|^2\right).$$

Now, we consider the integral over I_2,

$$\int_{I_2} = \int_{-c+\mu(x)-\mu(x_j)}^{c+\mu(x)-\mu(x_j)} (\mu(x_j) - \mu(x) + u)p_\varepsilon(u)du.$$

Substituting $z = \mu(x_j) - \mu(x) + u$,

$$\int_{I_2} = \int_{-c}^{c} z \; p_\varepsilon(z + \mu(x) - \mu(x_j))dz.$$

Since $\mu(x) - \mu(x_j)$ goes to zero as $N \to \infty$, we use (E1) c) to expand p_ε around z,

$$\begin{aligned}
\int_{I_2} &= \int_{-c}^{c} z \; \left\{ p_\varepsilon(z) + (\mu(x) - \mu(x_j))p'_\varepsilon(z) + \frac{1}{2}(\mu(x) - \mu(x_j))^2 p''_\varepsilon(z) + o(|\mu(x) - \mu(x_j)|^2) \right\} dz \\
&= \int_{-c}^{c} z \; p_\varepsilon(z)dz + (\mu(x) - \mu(x_j)) \int_{-c}^{c} z \; p'_\varepsilon(z)dz \\
&\quad + \frac{1}{2}(\mu(x) - \mu(x_j))^2 \int_{-c}^{c} z \; p''_\varepsilon(z)dz + o(|\mu(x) - \mu(x_j)|^2).
\end{aligned}$$

The symmetry of the density implies that

$$\int_{-c}^{c} z \; p_\varepsilon(z)dz = 0 \quad \text{and} \quad \int_{-c}^{c} z \; p''_\varepsilon(z)dz = 0.$$

This reduces the integral to,

$$\int_{I_2} = (\mu(x) - \mu(x_j)) \int_{-c}^{c} z \; p'_\varepsilon(z)dz + o(|\mu(x) - \mu(x_j)|^2).$$

Combining $\int_{I_2}$ with $\int_{I_1 \cup I_3}$,

$$\int_{\mathbb{R}} = (\mu(x) - \mu(x_j)) \left\{ -2cp_\varepsilon(c) + \int_{-c}^{c} z \; p'_\varepsilon(z)dz \right\} + o(|\mu(x) - \mu(x_j)|^2).$$

Using integration by parts and the symmetry of p_ε

$$\int_{c}^{c} z \; p'_\varepsilon(z)dz = zp_\varepsilon(z)\Big|_{-c}^{c} - \int_{-c}^{c} p_\varepsilon(z)dz = cp_\varepsilon(c) - (-c)p_\varepsilon(-c) - \eta = 2cp_\varepsilon(c) - \eta.$$

That is,

$$\int_{\mathbb{R}} = -\eta\ (\mu(x)-\mu(x_j)) + o(|\mu(x)-\mu(x_j)|^2).$$

Therefore,

$$B_N(x) = -\eta\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\ (\mu(x)-\mu(x_j)) + o(h^2).$$

Using assumption (A2) a), we expand $\mu(x_j)$ around x as follows,

$$\mu(x_j) = \mu(x) + \mu'(x)(x_j-x) + \frac{1}{2}\mu''(x)(x_j-x)^2 + o(|x_j-x|^2),$$

then,

$$\begin{aligned}
B_N(x) &= \eta\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\ \{\mu'(x)(x_j-x) + \frac{1}{2}\mu''(x)(x_j-x)^2\} + o(h^2) \\
&= \mu'(x)\eta\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)(x_j-x) + \frac{1}{2}\mu''(x)\eta\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)(x_j-x)^2 + o(h^2) \\
&= h\mu'(x)\eta\frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{x-x_j}{h}\right)\left(\frac{x_j-x}{h}\right) \\
&\quad + \frac{1}{2}h^2\mu''(x)\eta\frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{x-x_j}{h}\right)\left(\frac{x_j-x}{h}\right)^2 + o(h^2).
\end{aligned}$$

Using Lemma 2.4 we get that,

$$B_N(x) = \frac{1}{2}h^2\mu''(x)V_K\eta + O\left(\frac{1}{Nh}\right) + O\left(\frac{1}{N}\right) + o(h^2).$$

Assuming that $Nh^3 \to \infty$ combines the remainder terms to $o(h^2)$. □

To get the limit of $D_N(x)$ we will use the continuous mapping theorem (for example, see [1]) and Slutsky's theorem (for example, see [34]) which are going to be used repeatedly in the proofs.

Theorem 2.14 (Continuous Mapping Theorem) *Let m be a measurable function and let D_m be the set of discontinuity points of m.*

$$X_n \xrightarrow{\mathcal{L}} X, \quad \mathbb{P}(X\in D_m)=0 \quad \Longrightarrow \quad m(X_n) \xrightarrow{\mathcal{L}} m(X).$$

Theorem 2.15 (Slutsky's Theorem) *Let $X_N \xrightarrow{\mathcal{L}} X$ and let $Y_N \xrightarrow{P} a$, where a is a constant. Then*

(1) $X_N Y_N \xrightarrow{\mathcal{L}} aX.$

(2) $\frac{X_N}{Y_N} \xrightarrow{\mathcal{L}} \frac{X}{a}, \quad a \neq 0.$

(3) $X_N + Y_N \xrightarrow{\mathcal{L}} X + a.$

The continuous mapping theorem will help us first to prove the following lemma.

Lemma 2.16 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). Let $\varepsilon_j^* = f_j - \mu(x) + w_j$ where $|w_j| < |\tilde{\mu}(x) - \mu(x)|$. For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then, for all $x \in [h, 1-h]$, $x_j = j/N$, $j = 1, \ldots, N$ such that $|x - x_j| \leq h$ we have the following:*

If $c \geq 1$ then

$$(1) \quad \mathbb{E}\,\psi'(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \eta, \qquad (2) \quad \mathbb{E}\,\psi'^2(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \eta, \qquad (3) \quad \operatorname{var}\psi'(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \eta(1-\eta).$$

If $c < 1$ then

$$(4) \quad \mathbb{E}\,\psi'(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \frac{\eta}{c}, \qquad (5) \quad \mathbb{E}\,\psi'^2(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \frac{\eta}{c^2}, \qquad (6) \quad \operatorname{var}\psi'(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \frac{\eta(1-\eta)}{c^2}.$$

Proof. We will only prove part (1). The other parts follow directly.

From $f_j = \mu(x_j) + \varepsilon_j$ we have, $\varepsilon_j^* = \varepsilon_j + \mu(x_j) - \mu(x) + w_j$. Using the Lipschitz continuity of μ and the consistency of $\tilde{\mu}$ we have that $\varepsilon_j^* \xrightarrow{\mathcal{L}} \varepsilon_j$ for all $|x - x_j| \leq h$.

Consider the indicator function $m(\cdot) = \mathbb{1}_{(-c,c)}(\cdot)$ then m is measurable and $\mathbb{P}(\varepsilon_j \in D_m) = \mathbb{P}(\varepsilon_j \in \{-c, c\}) = 0$ for all j since the distribution of ε_j is assumed to be continuous.

Using the continuous mapping theorem (Theorem 2.14) we get,

$$\mathbb{1}_{(-c,c)}(\varepsilon_j^*) \xrightarrow{\mathcal{L}} \mathbb{1}_{(-c,c)}(\varepsilon_j)$$

By the definition of weak convergence we get that

$$\mathbb{E}\,b(\mathbb{1}_{(-c,c)}(\varepsilon_j^*)) \overset{N\to\infty}{\longrightarrow} \mathbb{E}\,b(\mathbb{1}_{(-c,c)}(\varepsilon_j)) \tag{2.17}$$

for every bounded and continuous function b.

That is true in particular if $b = \psi$ $(c \geq 1)$, i.e.

$$\mathbb{E}\,\psi(\mathbb{1}_{(-c,c)}(\varepsilon_j^*)) \overset{N\to\infty}{\longrightarrow} \mathbb{E}\,\psi(\mathbb{1}_{(-c,c)}(\varepsilon_j)). \tag{2.18}$$

But,

$$\psi \circ \mathbb{1}_{(-c,c)}(u) = \mathbb{1}_{(-c,c)}(u) \quad \forall\, u \in \mathbb{R},$$

then,

$$\mathbb{E}\,\mathbb{1}_{(-c,c)}(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \mathbb{E}\,\mathbb{1}_{(-c,c)}(\varepsilon_j). \tag{2.19}$$

Since $\psi' = \mathbb{1}_{(-c,c)}$ almost everywhere we get that,

$$\mathbb{E}\,\psi'(\varepsilon_j^*) \overset{N\to\infty}{\longrightarrow} \mathbb{E}\,\psi'(\varepsilon_j) = \int_{-c}^{c} p_\varepsilon(u)du = \eta. \tag{2.20}$$

□

Now using the previous lemma we get the limit of $D_N(x)$.

Proposition 2.17 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). Let*

$$D_N(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\psi'(f_j - \mu(x) + w_j),$$

where $|w_j| < |\tilde{\mu}(x) - \mu(x)|$. For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then, for all $x \in [h, 1-h]$ we have the following,

$$D_N(x) \overset{P}{\longrightarrow} \eta_c \quad \textit{and} \quad \frac{1}{D_N(x)} \overset{P}{\longrightarrow} \frac{1}{\eta_c}.$$

Proof. From Lemma 2.3 we have

$$\frac{1}{N^2}\sum_{j=1}^{N} K_h^2(x-x_j) = \frac{Q_K}{Nh} + o\left(\frac{1}{Nh}\right).$$

Lemma 2.16 implies that for all $|x - x_j| \le h$

$$\operatorname{var}\left(\psi'(f_j - \mu(x) + w_j)\right) = O(1).$$

Hence,

$$\frac{1}{N^2}\sum_{j=1}^{N} \operatorname{var}\left\{K_h(x-x_j)\psi'(f_j - \mu(x) + w_j)\right\} \to 0.$$

Then, by Chebyshev's LLN (Theorem 2.11)

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\psi'(f_j - \mu(x) + w_j) - \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\mathbb{E}\,\psi'(f_j - \mu(x) + w_j) \overset{P}{\longrightarrow} 0$$

Lemmas 2.3 and 2.16 imply

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\mathbb{E}\,\psi'(f_j - \mu(x) + w_j) \overset{N\to\infty}{\longrightarrow} \eta_c.$$

Using Slutsky's theorem completes the proof. □

Now, we need to prove that $D_N(x)$ is bounded from below away from zero. This is needed to show that the expected value of $1/D_N(x)$ converges to $1/\eta_c$. For the convergence of the expected values we will use the dominated convergence theorem (for example, see [34]).

Theorem 2.18 (Dominated Convergence Theorem) *Let $|X_n| \leq Y$ a.s., where Y is integrable. Then,*

$$X_n \xrightarrow{P} X \quad \Longrightarrow \quad \mathbb{E}\, X_n \xrightarrow{N\to\infty} \mathbb{E}\, X.$$

Lemma 2.19 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). Let*

$$D_N(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi'(f_j - \mu(x) + w_j),$$

where $|w_j| < |\tilde{\mu}(x) - \mu(x)|$. For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then,

$$\inf_{x\in[h,1-h]} D_N(x) \geq \frac{1}{2}\,\eta_c \quad a.s.$$

That is, there exists an $M > 0$ such that

$$\inf_{x\in[h,1-h]} D_N(x) \geq M \quad a.s. \quad \text{and} \quad \sup_{x\in[h,1-h]} \frac{1}{D_N(x)} \leq \frac{1}{M} \quad a.s.$$

Proof. Assume there exists an $x^* \in [h, 1-h]$ such that $D_N(x^*) < \frac{1}{2}\eta_c$, then

$$0 \leq \frac{1}{N}\sum_{j=1}^{N} K_h(x^* - x_j)\psi'(f_j - \mu(x^*) + w_j) < \frac{1}{2}\eta_c,$$

integrating with respect to the probability measure of ε_j

$$0 \leq \frac{1}{N}\sum_{j=1}^{N} K_h(x^* - x_j)\mathbb{E}\,\psi'(f_j - \mu(x^*) + w_j) < \frac{1}{2}\eta_c,$$

taking the limits as $N \to \infty$ (using the proof of Proposition 2.17)

$$0 \leq \eta_c \leq \frac{1}{2}\eta_c,$$

which is a contradiction to the fact that η_c is never zero.

Note that the proof also works if we choose any constant which is strictly less than η_c instead of $\frac{1}{2}\eta_c$. Hence, we can write the result as follows

$$\inf_{x\in[h,1-h]} D_N(x) \geq M \quad a.s. \quad \text{for some } M > 0$$

and

$$\sup_{x\in[h,1-h]} \frac{1}{D_N(x)} \leq \frac{1}{M} \quad a.s.$$

□

2.4.1 The Bias Term

Now, we give the bias term of the LHM-estimate.

Theorem 2.20 (LHM Bias) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $h \sim$ constant $N^{-1/5}$. Then,*

$$\operatorname{bias} \tilde{\mu}(x) = \frac{1}{2} h^2 \mu''(x) V_K + o(h^2),$$

for $x \in [h, 1-h]$.

Proof. From Propositions 2.10 and 2.13 and since h is chosen such that $h \sim$ constant $N^{-1/5}$ we get

$$\frac{1}{h^4}\,\mathbb{E}\,H_N^2(x) = \frac{1}{h^4}\,O\left(\frac{1}{Nh}\right) + \frac{1}{h^4}\,O(h^4) = O\left(\frac{1}{Nh^5}\right) + O(1) = O(1).$$

From Proposition 2.17, Lemma 2.19 and Slutsky's theorem we get

$$\left(\frac{1}{D_N(x)} - \frac{1}{\eta_c}\right)^2 \xrightarrow{P} 0 \quad \text{and} \quad \left(\frac{1}{D_N(x)} - \frac{1}{\eta_c}\right)^2 \leq \left(\frac{1}{M} + \frac{1}{\eta_c}\right)^2 \quad a.s.$$

therefore, the dominated convergence theorem yields

$$\mathbb{E}\left(\frac{1}{D_N(x)} - \frac{1}{\eta_c}\right)^2 \xrightarrow{N\to\infty} 0.$$

Using (2.14),

$$\begin{aligned}
\frac{\operatorname{bias} \tilde{\mu}(x) - \frac{1}{\eta_c}\mathbb{E}\,H_N(x)}{h^2} &= \frac{\mathbb{E}\,(\tilde{\mu}(x) - \mu(x)) - \frac{1}{\eta_c}\mathbb{E}\,H_N(x)}{h^2} \\
&= \frac{\mathbb{E}\,\frac{H_N(x)}{D_N(x)} - \frac{1}{\eta_c}\mathbb{E}\,H_N(x)}{h^2} \\
&= \frac{1}{h^2}\,\mathbb{E}\,H_N(x)\left(\frac{1}{D_N(x)} - \frac{1}{\eta_c}\right) \\
&\quad \text{(Using Cauchy-Schwarz inequality)} \\
&\leq \sqrt{\frac{1}{h^4}\mathbb{E}\,H_N^2(x)}\sqrt{\mathbb{E}\,\left(\frac{1}{D_N(x)} - \frac{1}{\eta_c}\right)^2}
\end{aligned}$$

$$= \sqrt{O(1)}\sqrt{o(1)} \overset{N\to\infty}{\longrightarrow} 0.$$

Therefore,

$$\operatorname{bias}\tilde{\mu}(x) = \frac{1}{\eta_c}\mathbb{E}\,H_N(x) + o\left(h^2\right) = \frac{1}{2}h^2\mu''(x)V_K + o(h^2).$$

□

2.4.2 The Variance Term

In this section we will show that,

$$Nh\operatorname{var}\tilde{\mu}(x) \overset{N\to\infty}{\longrightarrow} Q_K\sigma_c^2.$$

Before we see the proof we present the following lemma.

Lemma 2.21 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N\to\infty$, let $h\to 0$ such that $h \sim$ constant $N^{-1/5}$. Then*

$$\mathbb{E}\,Nh\,H_N^2(x)\left(\frac{1}{D_N^2(x)} - \frac{1}{\eta_c^2}\right) = o(1)$$

for all $x\in[h,1-h]$.

Proof. We start with calculating $\mathbb{E}\,H_N^4(x)$.

$$\begin{aligned}\mathbb{E}\,H_N^4(x) &= \mathbb{E}\left[B_N(x) + H_N(x) - B_N(x)\right]^4 \\ &= \mathbb{E}\left[B_N^4(x) + 4B_N^3(x)(H_N(x)-B_N(x)) + 6B_N^2(x)(H_N(x)-B_N(x))^2 \right.\\ &\qquad \left. + 4B_N(x)(H_N(x)-B_N(x))^3 + (H_N(x)-B_N(x))^4\right] \\ &= B_N^4(x) + 4B_N^3(x)\mathbb{E}\left(H_N(x)-B_N(x)\right) + 6B_N^2(x)\mathbb{E}\left(H_N(x)-B_N(x)\right)^2 \\ &\qquad + 4B_N(x)\mathbb{E}\left(H_N(x)-B_N(x)\right)^3 + \mathbb{E}\left(H_N(x)-B_N(x)\right)^4.\end{aligned}$$

Using Propositions 2.10 and 2.13 and $h \sim$ constant $N^{-1/5}$

$$\begin{aligned}\mathbb{E}\,H_N^4(x) &= O(h^8) + O(h^6)\cdot 0 + O(h^4)\operatorname{var}H_N(x) \\ &\qquad + O(h^2)\mathbb{E}\left(H_N(x)-B_N(x)\right)^3 + \mathbb{E}\left(H_N(x)-B_N(x)\right)^4 \\ &= O(h^4)O\left(\frac{1}{Nh}\right) + O(h^2)\mathbb{E}\left(H_N(x)-B_N(x)\right)^3 + \mathbb{E}\left(H_N(x)-B_N(x)\right)^4.\end{aligned}$$

To get the rate of $\mathbb{E}\,H_N^4(x)$, we have to calculate

$$\mathbb{E}\left(H_N(x)-B_N(x)\right)^3 \quad \text{and} \quad \mathbb{E}\left(H_N(x)-B_N(x)\right)^4.$$

For making notation easier in this proof, we will write

$$\psi_i \quad \text{instead of} \quad \psi(f_i - \mu(x)) \quad \text{and} \quad \gamma_i \quad \text{instead of} \quad \mathbb{E}\,\psi(f_i-\mu(x)).$$

Then,

$$\begin{aligned}\mathbb{E}\left(H_N(x) - B_N(x)\right)^3 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} K_h(x-x_i)(\psi_i - \gamma_i)\right)^3 \\ &= \frac{1}{N^3}\sum_{i,j,k=1}^{N} K_h(x-x_i)K_h(x-x_j)K_h(x-x_k) \\ &\qquad \mathbb{E}\left(\psi_i-\gamma_i\right)(\psi_j-\gamma_j)(\psi_k-\gamma_k).\end{aligned}$$

Since $\{\varepsilon_j : j = 1, \ldots, N\}$ are independent and identically distributed we get,

$$\mathbb{E}\left(\psi_i-\gamma_i\right)(\psi_j-\gamma_j)(\psi_k-\gamma_k) = \begin{cases}\mathbb{E}\left(\psi_i-\gamma_i\right)^3, & i=j=k, \\ 0, & \text{else.}\end{cases}$$

By the symmetry of p_ε we get for all x, x_i such that $|x - x_i| \leq h$

$$\mathbb{E}\left(\psi_i-\gamma_i\right)^3 = \int_{-\infty}^{\infty} \psi^3(u)p_\varepsilon(u)du + o(1) = o(1).$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(H_N(x) - B_N(x)\right)^3 &= \frac{1}{N^3}\sum_{i=1}^{N} K_h^3(x-x_i)\mathbb{E}\left(\psi_i-\gamma_i\right)^3 \\ &= \left(\frac{S_K}{N^2h^2} + o\left(\frac{1}{N^2h^2}\right)\right)\cdot o(1) \\ &= o\left(\frac{1}{N^2h^2}\right).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}\left(H_N(x) - B_N(x)\right)^4 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} K_h(x-x_i)(\psi_i - \gamma_i)\right)^4 \\ &= \frac{1}{N^4}\sum_{i,j,k,\ell=1}^{N} K_h(x-x_i)K_h(x-x_j)K_h(x-x_k)K_h(x-x_\ell) \\ &\qquad \mathbb{E}\left(\psi_i-\gamma_i\right)(\psi_j-\gamma_j)(\psi_k-\gamma_k)(\psi_\ell-\gamma_\ell).\end{aligned}$$

Since $\{\varepsilon_j : j = 1, \dots, N\}$ are independent and identically distributed we get,

$$\mathbb{E}\,(\psi_i-\gamma_i)(\psi_j-\gamma_j)(\psi_k-\gamma_k)(\psi_\ell-\gamma_\ell) = \begin{cases} \mathbb{E}\,(\psi_i-\gamma_i)^4, & i=j=k=\ell,\\ \sigma_M^4 + o(1), & i=j \text{ and } k=\ell \text{ but } i\neq k,\\ \sigma_M^4 + o(1), & i=k \text{ and } j=\ell \text{ but } i\neq j,\\ \sigma_M^4 + o(1), & i=\ell \text{ and } j=k \text{ but } i\neq j,\\ 0, & \text{else.} \end{cases}$$

For all x, x_i such that $|x - x_i| \le h$

$$\mathbb{E}\,(\psi_i-\gamma_i)^4 = \int_{-\infty}^{\infty} \psi^4(u) p_\varepsilon(u) du + o(1) = O(1).$$

Therefore,

$$\begin{aligned} \mathbb{E}\,(H_N(x) - B_N(x))^4 &= \left(\frac{1}{N^4}\sum_{i=1}^N K_h^4(x-x_i)\right)\cdot O(1)\\ &\quad + 3\left[\left(\frac{1}{N^2}\sum_{i=1}^N K_h^2(x-x_i)\right)^2 - \frac{1}{N^4}\sum_{i=1}^N K_h^4(x-x_i)\right]\cdot O(1)\\ &= O\left(\frac{1}{N^3h^3}\right) + \left[O\left(\frac{1}{N^2h^2}\right) + O\left(\frac{1}{N^3h^3}\right)\right]\\ &= O\left(\frac{1}{N^2h^2}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}\,H_N^4(x) &= O(h^4)O\left(\frac{1}{Nh}\right) + O(h^2)o\left(\frac{1}{N^2h^2}\right) + O\left(\frac{1}{N^2h^2}\right)\\ &= O\left(\frac{1}{N^2h^2}\right). \end{aligned}$$

From Proposition 2.17, Lemma 2.19 and Slutsky's theorem we get

$$\left(\frac{1}{D_N^2(x)} - \frac{1}{\eta_c^2}\right)^2 \xrightarrow{P} 0 \quad \text{and} \quad \left(\frac{1}{D_N^2(x)} - \frac{1}{\eta_c^2}\right)^2 \le \left(\frac{1}{M^2} + \frac{1}{\eta_c^2}\right)^2 \quad a.s.$$

therefore, the dominated convergence theorem (Theorem 2.18) yields

$$\mathbb{E}\,\left(\frac{1}{D_N^2(x)} - \frac{1}{\eta_c^2}\right)^2 \xrightarrow{N\to\infty} 0.$$

From Cauchy-Schwarz inequality we get,

$$
\begin{aligned}
\mathbb{E}\, H_N^2(x)\left(\frac{1}{D_N^2(x)}-\frac{1}{\eta_c^2}\right) &\leq \sqrt{\mathbb{E}\, H_N^4(x)}\cdot\sqrt{\mathbb{E}\left(\frac{1}{D_N^2(x)}-\frac{1}{\eta_c^2}\right)^2}\\
&= \sqrt{O\left(\frac{1}{N^2h^2}\right)}\sqrt{o(1)}\\
&= o\left(\frac{1}{Nh}\right).
\end{aligned}
$$

□

Theorem 2.22 (LHM Variance) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N\to\infty$, let $h\to 0$ such that $h\sim$ constant $N^{-1/5}$. Then*

$$
\operatorname{var}\tilde{\mu}(x)=\frac{Q_K}{Nh}\sigma_c^2+o\left(\frac{1}{Nh}\right),
$$

for $x\in[h,1-h]$.

Proof. Using (2.14),

$$
\begin{aligned}
&\operatorname{var}\tilde{\mu}(x)-\operatorname{var}\left(\frac{H_N(x)}{\eta_c}\right)\\
&=\operatorname{var}\left(\frac{H_N(x)}{D_N(x)}\right)-\operatorname{var}\left(\frac{H_N(x)}{\eta_c}\right)\\
&=\mathbb{E}\left(\frac{H_N(x)}{D_N(x)}\right)^2-\left(\mathbb{E}\,\frac{H_N(x)}{D_N(x)}\right)^2-\mathbb{E}\left(\frac{H_N(x)}{\eta_c}\right)^2+\left(\mathbb{E}\,\frac{H_N(x)}{\eta_c}\right)^2\\
&\text{(using the proof of the Theorem 2.20)}\\
&=\mathbb{E}\, H_N^2(x)\left(\frac{1}{D_N^2(x)}-\frac{1}{\eta_c^2}\right)+\left(\mathbb{E}\,\frac{H_N(x)}{\eta_c}\right)^2-\left(\mathbb{E}\,\frac{H_N(x)}{\eta_c}+o(h^2)\right)^2\\
&\text{(using Proposition 2.13)}\\
&=\mathbb{E}\, H_N^2(x)\left(\frac{1}{D_N^2(x)}-\frac{1}{\eta_c^2}\right)+o(h^4)
\end{aligned}
$$

Therefore, if $h\sim$ constant $N^{-1/5}$ we get,

$$
Nh\left[\operatorname{var}\tilde{\mu}(x)-\operatorname{var}\left(\frac{H_N(x)}{\eta_c}\right)\right]=\mathbb{E}\,Nh\,H_N^2(x)\left(\frac{1}{D_N^2(x)}-\frac{1}{\eta_c^2}\right)+o(1).
$$

Using Lemma 2.21 completes the proof. □

2.5 Asymptotic Normality

In this section we will show that the LHM-estimate has an asymptotic normal distribution.
To do that we will use Lyapounov's CLT (for example, see [1]).

Theorem 2.23 (Lyapounov's CLT) *Let Γ_j be a sequence of independent random variables with $\mathbb{E}\,\Gamma_j = \gamma_j$ and $\operatorname{var}\Gamma_j = v_j^2$. Let also $s_N^2 = \sum_{j=1}^{N} v_j^2$. If for some $\delta > 0$, $\mathbb{E}\,|\Gamma_j|^{2+\delta} < \infty$ for all j and the Lyapounov's condition hold, i.e.*

$$\frac{1}{s_N^{2+\delta}} \sum_{j=1}^{N} \mathbb{E}\,|\Gamma_j - \gamma_j|^{2+\delta} \overset{N\to\infty}{\longrightarrow} 0,$$

then

$$Z_N = \frac{\sum_{j=1}^{N} (\Gamma_j - \gamma_j)}{s_N} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

Now, we present the asymptotic normality of the LHM-estimate. A usual assumption that has to be fulfilled is $\int |K(u)|^{2+\delta} du < \infty$ for some $\delta > 0$. This holds in our case by (A1) a)-c).

Theorem 2.24 (LHM Asymptotic Normality) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1).*

(a) For $N \to \infty$, let $h \to 0$ such that $h \sim$ constant $N^{-1/5}$, then we have

$$\sqrt{Nh} \left(\frac{\tilde{\mu}(x) - \mu(x) - \frac{1}{2} h^2 \mu''(x) V_K}{\sqrt{\sigma_c^2 Q_K}} \right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1)$$

for all $x \in [h, 1-h]$, where

$$\sigma_c^2 = \frac{\sigma_M^2}{\eta_c^2} = \frac{c^2(1-\eta)}{\eta^2} + \frac{1}{\eta^2} \int_{-c}^{c} y^2 p_\varepsilon(y) dy.$$

(b) Suppose that $h \to 0$ as $N \to \infty$ such that $Nh^5 \to 0$, then

$$\sqrt{Nh} \left(\frac{\tilde{\mu}(x) - \mu(x)}{\sqrt{\sigma_c^2 Q_K}} \right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

Proof. Using the mean value theorem, we get that

$$\tilde{\mu}(x) - \mu(x) = \frac{H_N(x)}{D_N(x)}, \quad a.s.$$

where $H_N(x) = H_N(x, \mu(x))$ and $D_N(x)$ are defined as above. The equation holds true only almost surely since ψ is only almost everywhere differentiable.

We decompose the term we are interested in as follows,

$$\begin{aligned}
&\sqrt{Nh}\left(\frac{\tilde{\mu}(x)-\mu(x)-\frac{1}{2}h^2\mu''(x)V_K}{\sqrt{Q_K\sigma_c^2}}\right)\\
&\overset{a.s.}{=} \frac{\frac{H_N(x)}{D_N(x)} - \frac{\frac{1}{2}h^2\mu''(x)V_K\eta_c}{\eta_c}}{\sqrt{\frac{Q_K}{Nh}\sigma_c^2}}\\
&= \frac{\frac{H_N(x)}{D_N(x)} - \frac{B_N(x)}{D_N(x)} + \frac{B_N(x)}{D_N(x)} - \frac{B_N(x)}{\eta_c} + \frac{B_N(x)}{\eta_c} - \frac{\frac{1}{2}h^2\mu''(x)V_K\eta_c}{\eta_c}}{\sqrt{\frac{Q_K}{Nh}\sigma_c^2}}\\
&= \frac{H_N(x)-B_N(x)}{\sqrt{\frac{Q_K}{Nh}\sigma_c^2\eta_c^2}}\cdot\frac{\eta_c}{D_N(x)} + \frac{B_N(x)\left(\frac{1}{D_N(x)}-\frac{1}{\eta_c}\right)}{\sqrt{\frac{Q_K}{Nh}\sigma_c^2}} + \frac{B_N(x)-\frac{1}{2}h^2\mu''(x)V_K\eta_c}{\sqrt{\frac{Q_K}{Nh}\sigma_c^2\eta_c^2}}.
\end{aligned}$$

Using Propositions 2.13 and 2.17 and the assumption $h \sim$ constant $N^{-1/5}$ we get,

$$\begin{aligned}
&\sqrt{Nh}\left(\frac{\tilde{\mu}(x)-\mu(x)-\frac{1}{2}h^2\mu''(x)V_K}{\sqrt{Q_K\sigma_c^2}}\right)\\
&\overset{a.s.}{=} \frac{H_N(x)-B_N(x)}{\sqrt{\frac{Q_K}{Nh}\sigma_M^2}}\cdot\underbrace{\frac{\eta_c}{D_N(x)}}_{\xrightarrow{P}1} + \underbrace{\frac{\sqrt{Nh}B_N(x)}{\sqrt{Q_K\sigma_c^2}}}_{=O(\sqrt{Nh}h^2)+o(\sqrt{Nh}h^2)=O(1)}\cdot\underbrace{\left(\frac{1}{D_N(x)}-\frac{1}{\eta_c}\right)}_{\xrightarrow{P}0} + \underbrace{\frac{o(h^2)}{\sqrt{\frac{Q_K}{Nh}\sigma_M^2}}}_{=o(\sqrt{Nh}h^2)=o(1)}.
\end{aligned}$$

To prove the asymptotic normality we have to show that

$$\frac{H_N(x)-B_N(x)}{\sqrt{\frac{Q_K}{Nh}\sigma_M^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2.21}$$

Part b) follows by Slutsky's theorem. ☐

Proposition 2.25 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $Nh \to \infty$, then we have*

$$\frac{H_N(x)-B_N(x)}{\sqrt{\frac{Q_K}{Nh}\sigma_M^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

for all $x \in [h, 1-h]$.

Proof. We will show the asymptotic normality of $H_N(x)$ using the Lyapounov's CLT (Theorem 2.23) and taking $\delta = 1$.

Using the notation of the CLT, define

$$\Gamma_j = \frac{1}{N} K_h(x - x_j)\psi(f_j - \mu(x)).$$

Then, Γ_j are independent due to the independence of ε_j and

$$\gamma_j = \mathbb{E}\,\Gamma_j = \frac{1}{N} K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x)), \qquad v_j^2 = \operatorname{var}\Gamma_j = \frac{1}{N^2} K_h^2(x - x_j)\operatorname{var}\psi(f_j - \mu(x)).$$

Using Proposition 2.10,

$$\begin{aligned} s_N^2 &= \sum_{j=1}^N v_j^2 = \sum_{j=1}^N \frac{1}{N^2} K_h^2(x - x_j)\operatorname{var}\psi(f_j - \mu(x)) \\ &= \operatorname{var} H_N(x) = \frac{Q_K}{Nh}\sigma_M^2 + o\left(\frac{1}{Nh}\right). \end{aligned}$$

From the definition of ψ we have $|\psi|^3 \le \max\{c^3, 1\}$ and thus,

$$\mathbb{E}\,|\Gamma_j|^3 = \mathbb{E}\left|\frac{1}{N} K_h(x - x_j)\psi(f_j - \mu(x))\right|^3 \le \frac{1}{N^3} K_h^3(x - x_j)\max\{c^3, 1\},$$

and this is bounded for every $j \in \{1, \dots, N\}$ since K is Lipschitz continuous on a compact support.

Moreover, for $|x - x_j| \le h$ we have

$$\begin{aligned} \mathbb{E}\,|\psi(f_j - \mu(x)) - \mathbb{E}\,\psi(f_j - \mu(x))|^3 &= \mathbb{E}\,|\psi(\varepsilon_j) - \mathbb{E}\,\psi(\varepsilon_j) + o(1)|^3, && (\psi \text{ is continuous}) \\ &= \mathbb{E}\,|\psi(\varepsilon_j)|^3 + o(1), && (|\cdot|^3 \text{ is continuous}) \\ &\le \max\{c^3, 1\} + o(1). && (|\psi|^3 \le \max\{c^3, 1\}) \end{aligned}$$

Now, we will use the above and Lemma 2.3 to show that the Lyapounov's condition holds,

$$\begin{aligned} 0 &\le \frac{1}{s_N^3} \sum_{j=1}^N \mathbb{E}\,|\Gamma_j - \gamma_j|^3 \\ &= \frac{1}{s_N^3} \sum_{j=1}^N \mathbb{E}\left|\frac{1}{N} K_h(x - x_j)\psi(f_j - \mu(x)) - \frac{1}{N} K_h(x - x_j)\mathbb{E}\,\psi(f_j - \mu(x))\right|^3 \\ &= \frac{1}{s_N^3}\frac{1}{N^3} \sum_{j=1}^N K_h^3(x - x_j)\mathbb{E}\,|\psi(f_j - \mu(x)) - \mathbb{E}\,\psi(f_j - \mu(x))|^3 \\ &\le \frac{1}{s_N^3}\frac{1}{N^3} \sum_{j=1}^N K_h^3(x - x_j)\left\{\max\{c^3, 1\} + o(1)\right\} \\ &= \left\{\frac{Q_K\sigma_M^2}{Nh} + o\left(\frac{1}{Nh}\right)\right\}^{-3/2} \left\{\frac{S_K}{N^2h^2} + o\left(\frac{1}{N^2h^2}\right)\right\} \left\{\max\{c^3, 1\} + o(1)\right\} \end{aligned}$$

$$= \left\{Q_K\sigma_M^2 + o\,(1)\right\}^{-3/2} \cdot \{S_K + o\,(1)\} \cdot \left\{\max\{c^3, 1\} + o(1)\right\} \cdot \left\{\frac{1}{\sqrt{Nh}}\right\}$$
$$\longrightarrow 0,$$

as $Nh \to \infty$. □

2.6 The L_2 and the L_1 Limiting Cases

It is also interesting to see the behavior of the LHM-estimate as $c \to 0$ (i.e. the least absolute deviation estimate, abbreviated as LAD-estimate) and as $c \to \infty$ (i.e. the Nadaraya-Watson estimate, abbreviated as NW-estimate).

Remark 2.26 *Since p_ε is a continuous density we have*

$$\lim_{c\to\infty} \eta_c = \lim_{c\to\infty} \int_{-c}^{c} p_\varepsilon(y)dy = 1. \qquad (L_2 \textit{ limiting case})$$

Let P_ε be the cumulative distribution function of $\{\varepsilon_j\}_{j=1\ldots,N}$ then

$$\begin{aligned}\lim_{c\to 0} \eta_c &= \lim_{c\to 0} \frac{1}{c}\int_{-c}^{c} p_\varepsilon(y)dy = \lim_{c\to 0} \frac{2}{c}\int_{0}^{c} p_\varepsilon(y)dy \\ &= 2\lim_{c\to 0} \frac{P_\varepsilon(c) - P_\varepsilon(0)}{c} = 2P'_\varepsilon(0) = 2p_\varepsilon(0). \qquad (L_1 \textit{ limiting case})\end{aligned}$$

Note that if p_ε is not symmetric around zero we will still have that $\eta_c \to 2p_\varepsilon(0)$ as $c \to 0$ since

$$\begin{aligned}\lim_{c\to 0} \eta_c &= \lim_{c\to 0} \frac{1}{c}\int_{-c}^{c} p_\varepsilon(y)dy = \lim_{c\to 0} \frac{P_\varepsilon(c) - P_\varepsilon(-c)}{c} \\ &= \lim_{c\to 0} \frac{P_\varepsilon(c) - P_\varepsilon(0)}{c} + \lim_{c\to 0} \frac{P_\varepsilon(0) - P_\varepsilon(-c)}{c} \\ &= \lim_{c\to 0} \frac{P_\varepsilon(c) - P_\varepsilon(0)}{c} + \lim_{c\to 0} \frac{P_\varepsilon(-c) - P_\varepsilon(0)}{-c} \\ &= P'_\varepsilon(0) + P'_\varepsilon(0) = 2p_\varepsilon(0).\end{aligned}$$

The assumption that p_ε is symmetric is not essential to have $\eta_c \to 2p_\varepsilon(0)$ as $c \to 0$, but makes work easier.

Corollary 2.27 *Let the model (2.1) hold. Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $h \sim$ constant $N^{-1/5}$, then for all $x \in [h, 1-h]$ we have the following,*

(a) the asymptotic distribution of the least-absolute deviation estimate is given by

$$\sqrt{Nh}\left(\tilde{\mu}_{LAD}(x) - \mu(x) - \frac{1}{2}h^2\mu''(x)V_K\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{Q_K}{4p_\varepsilon^2(0)}\right),$$

(b) and the asymptotic distribution of the Nadaraya-Watson kernel estimate is given by

$$\sqrt{Nh}\left(\tilde{\mu}_{NW}(x)-\mu(x)-\frac{1}{2}h^2\mu''(x)V_K\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\sigma^2 Q_K\right).$$

Proof. The proof follows from Theorem 2.24. The bias term has no dependence on c therefore it is the same in both cases. The variance in both cases could be obtained as a limit of σ_c^2 as c tends to zero (for the LAD-estimate) and as a limit of σ_c^2 as c tends to infinity (for the Nadaraya-Watson estimate).

From,

$$\lim_{c\to\infty}\eta_c=\lim_{c\to\infty}\int_{-c}^{c}p_\varepsilon(y)dy=\int_{-\infty}^{\infty}p_\varepsilon(y)dy=1,$$
$$\lim_{c\to 0}\eta_c=\lim_{c\to 0}\frac{1}{c}\int_{-c}^{c}p_\varepsilon(y)dy=2p_\varepsilon(0).$$

And,

$$\lim_{c\to\infty}\sigma_M^2=\lim_{c\to\infty}\left\{c^2(1-\eta)+\int_{-c}^{c}y^2p_\varepsilon(y)dy\right\}=\sigma^2,$$
$$\lim_{c\to 0}\sigma_M^2=\lim_{c\to 0}\left\{(1-\eta)+\frac{1}{c^2}\int_{-c}^{c}y^2p_\varepsilon(y)dy\right\}=1.$$

We get,

$$\lim_{c\to\infty}\sigma_c^2=\lim_{c\to\infty}\frac{\sigma_M^2}{\eta_c^2}=\frac{\sigma^2}{1}=\sigma^2,$$
$$\lim_{c\to 0}\sigma_c^2=\lim_{c\to 0}\frac{\sigma_M^2}{\eta_c^2}=\frac{1}{(2p_\varepsilon(0))^2}=\frac{1}{4p_\varepsilon^2(0)}.$$

□

2.7 Note on the Optimal Choice of the Bandwidth h

It has been stated in Section 2.4 that we can refer to the dominant part of $\frac{1}{\eta_c}\mathbb{E}\,H_N(x)$ as the "asymptotic bias term" and to the dominant term of $\frac{1}{\eta_c^2}\operatorname{var}H_N(x)$ as the "asymptotic variance term". That is,

$$\text{ABIAS}\,\tilde{\mu}(x)=\frac{1}{2}h^2\mu''(x)V_K\quad\text{and}\quad\text{AVAR}\,\tilde{\mu}(x)=\frac{Q_K}{Nh}\sigma_c^2.$$

Hence,

$$\text{AMSE}\,\tilde{\mu}(x)=\frac{Q_K}{Nh}\sigma_c^2+\frac{1}{4}h^4\left(\mu''(x)\right)^2V_K^2.$$

To get an optimal choice of h "locally" in the sense of minimal asymptotic mean-squared error we differentiate $\mathrm{AMSE}\,\tilde{\mu}(x)$ with respect to h.

$$\frac{\partial\,\mathrm{AMSE}\,\tilde{\mu}(x)}{\partial h} = -\frac{Q_K}{Nh^2}\sigma_c^2 + h^3\left(\mu''(x)\right)^2 V_K^2.$$

Setting the derivative equal to zero yields a "local asymptotically optimal bandwidth", i.e.

$$h_{opt}(x) = \left(\frac{Q_K\sigma_c^2}{\left(\mu''(x)\right)^2 V_K^2}\right)^{1/5} N^{-1/5}.$$

The result fits with the assumption $h \sim$ constant $N^{-1/5}$. This assumption was required to show that the bias and variance terms of the LHM-estimate have right convergence rates, and to show that the LHM-estimate has an asymptotic normal distribution.

Chapter 3

Uniform Consistency of the Local Huber M-Estimate

In Chapter 2 we have seen that the LHM-estimate is consistent. In this chapter, we will show, under the same assumptions, that the LHM-estimate is uniformly consistent.

Härdle and Luckhaus [15] have shown uniform consistency of the M-estimate under two settings. The first was under the random design and using a rescaled kernel function as the tonal weight function. While, the second was under the fixed design but using the Gasser-Müller weight function for localization.

Franke [9] has shown uniform consistency for the Priestley-Chao kernel estimate under the fixed design and using a rescaled kernel function as the tonal weight function.

Using the methods of Härdle and Luckhaus [15] and Franke [9], we will prove here the uniform consistency of the Huber M-estimate, under the fixed design, and using a rescaled kernel function as the tonal weight function.

3.1 Preliminaries

We recall from (2.14) that

$$\tilde{\mu}(x) - \mu(x) = \frac{H_N(x)}{D_N(x)} \quad a.s.$$

where

$$H_N(x) = H_N(x, \mu(x)) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x)),$$

and

$$D_N(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi'(f_j - \mu(x) + w_j),$$

where $|w_j| < |\tilde{\mu}(x) - \mu(x)|$.

We have seen in Lemma 2.19 that there exists an $M > 0$ such that

$$\inf_{x\in[h,1-h]} D_N(x) \geq M \quad a.s. \quad \text{and} \quad \sup_{x\in[h,1-h]} \frac{1}{D_N(x)} \leq \frac{1}{M} \quad a.s.$$

Using the above argument

$$\begin{aligned}\sup_{x\in[h,1-h]} |\tilde{\mu}(x) - \mu(x)| &\overset{a.s.}{=} \sup_{x\in[h,1-h]} \left|\frac{H_N(x)}{D_N(x)}\right| \\ &\leq \sup_{x\in[h,1-h]} |H_N(x)| \cdot \sup_{x\in[h,1-h]} \frac{1}{D_N(x)} \\ &\overset{a.s.}{\leq} \frac{1}{M} \sup_{x\in[h,1-h]} |H_N(x)| .\end{aligned} \tag{3.1}$$

So, our goal now is to study the behavior of $\sup_{x\in[h,1-h]} |H_N(x)|$.

3.2 The Uniform Behavior of $H_N(x)$

The methods used here are similar to those used by Härdle and Luckhaus [15] and Franke [9]. Again, we ignore the boundaries $[0, h)$ and $(1 - h, 1]$, i.e. we take $x \in [h, 1 - h]$.

We assume model (2.1) where

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

Now, we consider the following equidistant mesh points ξ_k in $[h, 1 - h]$

$$h \leq \xi_1 < \xi_2 < \cdots < \xi_{\ell_N} \leq 1 - h \quad \text{where } \ell_N \to \infty \text{ but } \ell_N \ll N.$$

Note that these mesh point differ from x_j.

Using the above setting we decompose $\sup_x |H_N(x)|$ as follows

$$\begin{aligned}&\sup_{x\in[h,1-h]} |H_N(x)| \\ &\leq \sup_{1\leq k\leq \ell_N} \sup_{|x-\xi_k|\leq \ell_N^{-1}} \left|\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x))\right| \\ &\leq \sup_{1\leq k\leq \ell_N} \sup_{|x-\xi_k|\leq \ell_N^{-1}} \left|\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(x)) - \frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(\xi_k))\right| \\ &\quad + \sup_{1\leq k\leq \ell_N} \sup_{|x-\xi_k|\leq \ell_N^{-1}} \left|\frac{1}{N}\sum_{j=1}^{N} K_h(x - x_j)\psi(f_j - \mu(\xi_k)) - \frac{1}{N}\sum_{j=1}^{N} K_h(\xi_k - x_j)\psi(f_j - \mu(\xi_k))\right| \\ &\quad + \sup_{1\leq k\leq \ell_N} \sup_{|x-\xi_k|\leq \ell_N^{-1}} \left|\frac{1}{N}\sum_{j=1}^{N} K_h(\xi_k - x_j)\psi(f_j - \mu(\xi_k)) - \frac{1}{N}\sum_{j=1}^{N} K_h(\xi_k - x_j)\mathbb{E}\,\psi(f_j - \mu(\xi_k))\right|\end{aligned}$$

$$+\sup_{1\le k\le \ell_N}\sup_{|x-\xi_k|\le \ell_N^{-1}}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)\mathbb{E}\,\psi(f_j-\mu(\xi_k))\right|$$
$$=: U_1(x)+U_2(x)+U_3(x)+U_4(x),$$

and for short, we will write

$$\sup_{x\in[h,1-h]}|H_N(x)|\le U_1(x)+U_2(x)+U_3(x)+U_4(x). \tag{3.2}$$

Now, we will study the behavior of $U_1(x)$, $U_2(x)$, $U_3(x)$, and $U_4(x)$.

Lemma 3.1 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). For $N\to\infty$, let $h\to 0$ such that $Nh^2\to\infty$. Then,*

$$U_1(x)=O\left(\frac{1}{\ell_N}\right).$$

Proof. Using the Lipschitz continuity of ψ and μ

$$\begin{aligned}|\psi(f_j-\mu(x))-\psi(f_j-\mu(\xi_k))| &\le C_\psi|\mu(x)-\mu(\xi_k)|\\ &\le C_\psi C_\mu|x-\xi_k|,\end{aligned}$$

then

$$\begin{aligned}U_1(x) &= \sup_{1\le k\le \ell_N}\sup_{|x-\xi_k|\le \ell_N^{-1}}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\psi(f_j-\mu(x))-\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\psi(f_j-\mu(\xi_k))\right|\\ &\le \sup_{1\le k\le \ell_N}\sup_{|x-\xi_k|\le \ell_N^{-1}}\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\,|\psi(f_j-\mu(x))-\psi(f_j-\mu(\xi_k))|\\ &\le C_\psi C_\mu\frac{1}{\ell_N}\sup_{h\le x\le 1-h}\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)=O\left(\frac{1}{\ell_N}\right),\end{aligned}$$

since

$$\sup_{h\le x\le 1-h}\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)=O\left(1\right)$$

by Lemma 2.3. □

Lemma 3.2 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). For $N\to\infty$, let $h\to 0$ such that $Nh^2\to\infty$. Then,*

$$U_2(x)=O\left(\frac{1}{Nh^2}\right).$$

Proof. Using Lemma 2.3 we have

$$\sup_{h\le x\le 1-h}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)-1\right|\le\frac{C_K}{Nh^2},\quad\text{and}\quad\sup_{h\le \xi_k\le 1-h}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)-1\right|\le\frac{C_K}{Nh^2}$$

for all $k=1,\ldots,\ell_N$. Since $|\psi(\cdot)|\le\max\{c,1\}$ we have

$$\begin{aligned}U_2(x)=&\sup_{1\le k\le\ell_N}\sup_{|x-\xi_k|\le\ell_N^{-1}}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)\psi(f_j-\mu(\xi_k))\right.\\&\qquad\left.-\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)\psi(f_j-\mu(\xi_k))\right|\\\le&\max\{c,1\}\sup_{1\le k\le\ell_N}\sup_{|x-\xi_k|\le\ell_N^{-1}}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(x-x_j)-\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)\right|\\\le&\max\{c,1\}\frac{2C_K}{Nh^2}=O\left(\frac{1}{Nh^2}\right),\end{aligned}$$

regardless of the choice of ℓ_N. □

Lemma 3.3 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a). For $N\to\infty$, let $h\to 0$ such that $Nh^2\to\infty$. Then,*

$$U_3(x)=O_p\left(\frac{1}{r_N}\right)$$

provided that

$$r_N\to\infty\qquad\textit{and}\qquad\frac{r_N^2\ell_N}{Nh}\quad\textit{is bounded.}$$

Proof. Note that x does not appear in $U_3(x)$. So, we will write U_3 instead. Then

$$U_3=\sup_{1\le k\le\ell_N}\left|\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)\psi(f_j-\mu(\xi_k))-\frac{1}{N}\sum_{j=1}^{N}K_h(\xi_k-x_j)\mathbb{E}\,\psi(f_j-\mu(\xi_k))\right|.$$

Using Chebyshev's inequality and Proposition 2.10, we have for any $\gamma>0$

$$\begin{aligned}\mathbb{P}\left(r_NU_3>\gamma\right)=&\,\mathbb{P}\left(r_N\sup_{1\le k\le\ell_N}|H_N(\xi_k)-\mathbb{E}\,H_N(\xi_k)|>\gamma\right)\\\le&\sum_{k=1}^{\ell_N}\mathbb{P}\left(r_N|H_N(\xi_k)-\mathbb{E}\,H_N(\xi_k)|>\gamma\right)\\\le&\sum_{k=1}^{\ell_N}\frac{r_N^2\operatorname{var}H_N(\xi_k)}{\gamma^2}\end{aligned}$$

$$= \frac{r_N^2}{\gamma^2} \sum_{k=1}^{\ell_N} O\left(\frac{1}{Nh}\right) = O\left(\frac{r_N^2 \ell_N}{Nh}\right).$$

Therefore,

$$U_3 = O_p\left(\frac{1}{r_N}\right) \qquad \text{provided that} \quad r_N \to \infty \quad \text{and} \quad \frac{r_N^2 \ell_N}{Nh} \quad \text{is bounded.}$$

□

Lemma 3.4 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $Nh^3 \to \infty$. Then,*

$$U_4(x) = O\left(h^2\right).$$

Proof. Also here x does not appear in $U_4(x)$. So, we will write U_4 instead. Then

$$U_4 = \sup_{1 \le k \le \ell_N} \left| \frac{1}{N} \sum_{j=1}^{N} K_h(\xi_k - x_j) \mathbb{E}\, \psi(f_j - \mu(\xi_k)) \right|.$$

From Proposition 2.13, we have

$$\begin{aligned} U_4 &= \sup_{1 \le k \le \ell_N} |B_N(\xi_k)| \\ &\le \frac{h^2}{2} V_K \eta_c \sup_{1 \le k \le \ell_N} |\mu''(\xi_k)| + o(h^2) = O\left(h^2\right). \end{aligned}$$

□

Collecting the previous four lemmas we get the following result.

Proposition 3.5 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1). For $N \to \infty$, let $h \to 0$ such that $Nh^3 \to \infty$. Moreover, assume that*

(1) $\ell_N \to \infty$ but $\ell_N \ll N$.

(2) $r_N \to \infty$ and $\frac{r_N^2 \ell_N}{Nh}$ is bounded.

(3) $\frac{r_N}{\ell_N}$ is bounded.

(4) $\frac{r_N}{Nh^2}$ is bounded.

(5) $r_N h^2$ is bounded.

Then,

$$r_N \sup_{x \in [h, 1-h]} |H_N(x)| = O_p(1).$$

In particular, if

$$\ell_N = \sqrt[3]{Nh} \quad \textit{and} \quad r_N = \sqrt[3]{Nh}, \quad \textit{provided that} \quad Nh^7 = O(1),$$

then

$$\sqrt[3]{Nh} \sup_{x\in[h,1-h]} |H_N(x)| = O_p(1).$$

Proof. Using (3.2) and Lemmas 3.1, 3.2, 3.3, and 3.4 we get

$$r_N \sup_{x\in[h,1-h]} |H_N(x)| = O\left(\frac{r_N}{\ell_N}\right) + O\left(\frac{r_N}{Nh^2}\right) + O_p(1) + O\left(r_N h^2\right).$$

Now, we have to choose the sequences r_N and ℓ_N such that they are of the same order and such that $\frac{r_N^2 \ell_N}{Nh}$ is bounded for $r_N \to \infty$, i.e. assumptions (2) and (3). Choosing $r_N = \ell_N = \sqrt[3]{Nh}$ will satisfy these assumptions and the assumptions (1) and (4) as well. Adding $Nh^7 = O(1)$ will satisfy the assumption (5). □

Corollary 3.6 *Let the assumptions of Proposition 3.5 hold. For* $N \to \infty$*, let* $h \to 0$ *such that* $h \sim$ *constant* $N^{-1/5}$*. Then,*

$$N^{4/15} \sup_{x\in[h,1-h]} |H_N(x)| = O_p(1).$$

3.3 Uniform Consistency

Using (3.1) and Proposition 3.5 we will prove now that the LHM-estimate is uniformly consistent.

Theorem 3.7 *Let the model (2.1) hold. Let* ρ *be the modified Huber function given by (2.11). Let* K *satisfy (A1) a)-c). Let* μ *satisfy (A2). Let* ε_j *satisfy (E1). For* $N \to \infty$*, let* $h \to 0$ *such that* $Nh^3 \to \infty$*. Moreover, assume that*

(1) $\ell_N \to \infty$ *but* $\ell_N \ll N$.

(2) $r_N \to \infty$ *and* $\frac{r_N^2 \ell_N}{Nh}$ *is bounded.*

(3) $\frac{r_N}{\ell_N}$ *is bounded.*

(4) $\frac{r_N}{Nh^2}$ *is bounded.*

(5) $r_N h^2$ *is bounded.*

Then,

$$r_N \sup_{x\in[h,1-h]} |\tilde{\mu}(x) - \mu(x)| = O_p(1).$$

In particular, if

$$\ell_N = \sqrt[3]{Nh} \quad \text{and} \quad r_N = \sqrt[3]{Nh}, \quad \text{provided that} \quad Nh^7 = O(1),$$

then

$$\sqrt[3]{Nh} \sup_{x\in[h,1-h]} |\tilde{\mu}(x) - \mu(x)| = O_p(1).$$

Proof. From (3.1) we have for all $\gamma > 0$

$$\mathbb{P}\left(r_N \sup_{x\in[h,1-h]} |\tilde{\mu}(x) - \mu(x)| > \gamma \right) \leq \mathbb{P}\left(r_N \sup_{x\in[h,1-h]} |H_N(x)| > M\gamma \right).$$

The rest follows from Proposition 3.5. □

Corollary 3.8 *Let the assumptions of Theorem 3.7 hold. For $N \to \infty$, let $h \to 0$ such that $h \sim$ constant $N^{-1/5}$. Then,*

$$N^{4/15} \sup_{x\in[h,1-h]} |\tilde{\mu}(x) - \mu(x)| = O_p\left(1\right).$$

Chapter 4

Mathematical Formalization of the Asymptotic Analysis

In this small chapter we will see the basic definitions of the Landau symbols in vector spaces. 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 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 vector spaces. Thus, the mathematical tools will hold without any problem.

4.1 Asymptotic Notations

Asymptotic order notations or asymptotic notations, for short, are powerful tools for writing asymptotic expressions in a compact way and they are very often used in approximation theory (for example see [19]). The asymptotic notations are "big-O", "little-o, "big-Omega", "little-omega", "big-Theta", and "in the order of". They have the following symbols $O, o, \Omega, \omega, \Theta$ and $\sim$ respectively.

In our work, we are only interested in the first two notations (big-O and little-o) which are sometimes called the "Landau symbols". We start with the definition of the Landau symbols for real-valued sequences and then we generalize the definition for vector-valued and matrix-valued sequences.

Definition 4.1 *Let $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ be sequences of real numbers. Then we have the following notations,*

1) $a_n = O(b_n)$ if and only if there exists an $M \in (0, \infty)$ such that $\frac{|a_n|}{|b_n|} \leq M$ for all $n \in \mathbb{N}$.

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

3) $a_n \sim b_n$ if and only if $\lim\limits_{n\to\infty} \frac{|a_n|}{|b_n|} \to 1$.

In the same manner, we can generalize the definition to vector-valued sequences. For example, we say a vector-valued sequence is big-O of a real-valued sequence if the norm of the vector-valued sequence is big-O of the real-valued sequence (same for little-o). More generally, we say that a vector-valued sequence is big-O of another vector-valued sequence if the norm of the first sequence is big-O of the norm of the second. In the same manner the Landau symbols are defined for matrix-valued sequences.

Definition 4.2 *Let $\{\mathbf{a}_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$ be $\mathbb{R}^p$-valued sequences, let $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ be $\mathbb{R}^{p\times q}$-valued sequences, and let $\{c_n\}_{n\in\mathbb{N}}$ be a real-valued sequence. Then,*

1) $\mathbf{a}_n = O(c_n)$ *if and only if there exists an* $M > 0$ *such that* $\frac{\|\mathbf{a}_n\|}{|c_n|} \leq M$ *for all* $n \in \mathbb{N}$.

2) $\mathbf{a}_n = o(c_n)$ *if and only if* $\lim\limits_{n\to\infty} \frac{\|\mathbf{a}_n\|}{|c_n|} \to 0$.

3) $\mathbf{a}_n = O(\|\mathbf{b}_n\|)$ *if and only if there exists an* $M > 0$ *such that* $\frac{\|\mathbf{a}_n\|}{\|\mathbf{b}_n\|} \leq M$ *for all* $n \in \mathbb{N}$.

4) $\mathbf{a}_n = o(\|\mathbf{b}_n\|)$ *if and only if* $\lim\limits_{n\to\infty} \frac{\|\mathbf{a}_n\|}{\|\mathbf{b}_n\|} \to 0$.

5) $A_n = O(c_n)$ *if and only if there exists an* $M > 0$ *such that* $\frac{\|A_n\|}{|c_n|} \leq M$ *for all* $n \in \mathbb{N}$.

6) $A_n = o(c_n)$ *if and only if* $\lim\limits_{n\to\infty} \frac{\|A_n\|}{|c_n|} \to 0$.

7) $A_n = O(\|B_n\|)$ *if and only if there exists an* $M > 0$ *such that* $\frac{\|A_n\|}{\|B_n\|} \leq M$ *for all* $n \in \mathbb{N}$.

8) $A_n = o(\|B_n\|)$ *if and only if* $\lim\limits_{n\to\infty} \frac{\|A_n\|}{\|B_n\|} \to 0$.

Throughout our work, we will consider the infinity-norm for vectors and its induced norm for matrices. That is,

$$\|\mathbf{a}\| := \|\mathbf{a}\|_\infty = \sup_{1\leq i\leq p} |a^{(i)}|$$

for all vectors $\mathbf{a} = \left(a^{(1)}, \ldots, a^{(p)}\right)^T \in \mathbb{R}^p$. And

$$\|A\| := \|A\|_\infty = \sup_{1\leq i\leq p} \sum_{j=1}^{q} |a^{(i,j)}|$$

for all matrices $A = \left(a^{(i,j)}\right)_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}} \in \mathbb{R}^{p\times q}$.

4.2 Building the Vector Spaces

A technical problem we confront in our asymptotic expansion is that the vectors and matrices we are dealing with come from $\mathbb{R}^N$ and $\mathbb{R}^{N\times N}$ 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 will have vectors and matrices of different sizes. This means that these vectors no longer form a vector space (same holds for the matrices). Thus, we have to check most of the analytical results for vector spaces before we can use them!

A way out is to define a mapping that maps our vectors which are of varying lengths into vectors that have the same length. For that purpose we define the extension mapping $\mathcal{E}^1$ which maps vectors from the space $\mathbb{R}^N$ into the space $\ell^\infty(\mathbb{R})$, where

$$\ell^\infty(\mathbb{R}) = \left\{ \mathbf{x} = (x^{(i)})_{i\in\mathbb{N}} \;:\; x^{(i)} \in \mathbb{R} \;;\; \|\mathbf{x}\|_{\ell^\infty(\mathbb{R})} := \sup_{i\in\mathbb{N}} |x^{(i)}| < \infty \right\}, \tag{4.1}$$

that is,

$$\begin{array}{cccccc} \mathcal{E}^1 & : & \big(\mathbb{R}^N, \|\cdot\|_\infty\big) & \longrightarrow & \big(\ell^\infty(\mathbb{R}), \|\cdot\|_{\ell^\infty(\mathbb{R})}\big) \\ & & \mathbf{x} = \big(x^{(1)}, \dots, x^{(N)}\big) & \longrightarrow & \mathbf{x}^* = \big(x^{(1)}, \dots, x^{(N)}, 0, 0, 0, \dots\big) \end{array} \tag{4.2}$$

The space $\ell^\infty(\mathbb{R})$ is now considered as the space of vectors both when N is fixed and when $N \to \infty$ and it is indeed a vector space.

In a similar manner we define a mapping that maps our $N \times N$ matrices into infinite "$\infty \times \infty$" or "$\mathbb{N} \times \mathbb{N}$" matrices as follows

$$\begin{array}{cccccc} \mathcal{E}^2 & : & \big(\mathbb{R}^{N\times N}, \|\cdot\|_\infty\big) & \longrightarrow & \big(\mathcal{M}^\infty(\mathbb{R}), \|\cdot\|_{\mathcal{M}^\infty(\mathbb{R})}\big) \\ & & A = \big(a^{(i,j)}\big)_{i,j=1,\dots,N} & \longrightarrow & A^* = \big(a^{*(i,j)}\big)_{i,j\in\mathbb{N}} \end{array} \tag{4.3}$$

where

$$a^{*(i,j)} = \begin{cases} a^{(i,j)}, & i,j \in \{1,\dots,N\}, \\ 0, & i,j \in \mathbb{N} \setminus \{1,\dots,N\}, \end{cases}$$

and the space $\mathcal{M}^\infty(\mathbb{R})$ is defined as follows

$$\mathcal{M}^\infty(\mathbb{R}) = \left\{ A = (a^{(i,j)})_{i,j\in\mathbb{N}} \;:\; a^{(i,j)} \in \mathbb{R} \;;\; \|A\|_{\mathcal{M}^\infty(\mathbb{R})} := \sup_{i\in\mathbb{N}} \sum_{j=1}^{\infty} |a^{(i,j)}| < \infty \right\}. \tag{4.4}$$

Similar to vectors, the space $\mathcal{M}^\infty(\mathbb{R})$ is now considered as the space of matrices both when N is fixed and when $N \to \infty$ and it is indeed a vector space.

The extension mappings $\mathcal{E}^1$ and $\mathcal{E}^2$ are defined in a way to result in vectors spaces and to preserve the norm of the vectors and matrices being mapped. That is

$$\|\mathbf{x}\|_\infty = \|\mathbf{x}^*\|_{\ell^\infty(\mathbb{R})} \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^N \tag{4.5}$$

where $\mathbf{x}^* = \mathcal{E}^1(\mathbf{x}) \in \ell^\infty(\mathbb{R})$, and

$$\|A\|_\infty = \|A^*\|_{\mathcal{M}^\infty(\mathbb{R})} \quad \text{for all} \quad A \in \mathbb{R}^{N\times N} \tag{4.6}$$

where $A^* = \mathcal{E}^2(A) \in \mathcal{M}^\infty(\mathbb{R})$.

Moreover, if we restrict ourselves to symmetric matrices, then we have

$$\mathcal{S}^\infty(\mathbb{R}) = \left\{ A = \big(a^{(i,j)}\big)_{i,j\in\mathbb{N}} \;:\; A \in \mathcal{M}^\infty(\mathbb{R}) \;;\; a^{(i,j)} = a^{(j,i)} \right\},$$

which is clearly a subspace of $\mathcal{M}^\infty(\mathbb{R})$.

In the following chapters, we will work with $\mathcal{S}^\infty(\mathbb{R})$.

4.3 A Note on the Built Vector Spaces

Throughout our work we will continue to write $\mathbf{x} \in \mathbb{R}^N$ for vectors and $A \in \mathbb{R}^{N\times N}$ for matrices. But for the precise mathematical writing we actually mean the $\mathcal{E}^1$ and $\mathcal{E}^2$ extended versions, i.e. $\mathbf{x}^* = \mathcal{E}^1(\mathbf{x}) \in \ell^\infty(\mathbb{R})$ and $A^* = \mathcal{E}^2(A) \in \mathcal{S}^\infty(\mathbb{R})$. Using this shorthand writing, we will also write $\|\mathbf{x}\|$ in stead of $\|\mathbf{x}^*\|_{\ell^\infty(\mathbb{R})}$ and $\|A\|$ in stead of $\|A^*\|_{\mathcal{M}^\infty(\mathbb{R})}$.

Chapter 5

Huber M-Estimates with Localization and Quadratic Regularization (QRLHM-Estimates)

In this chapter, we will consider the LHM-estimate by adding a regularization term to the problem of interest. We will use here the general approach by Mrázek et al. [25] which is described in Section 1.5. The data term will be the same as that already discussed in Chapter 2. For the smoothness term we use the quadratic function as the smoothness loss function. First, we consider an asymptotic expansion of the solution in vector form then in component form. The component form is of interest because it enables us to calculate the bias and variance terms more easily.

5.1 Setup of the Problem

As in Chapter 2 we assume we are given data (x_j, f_j), $j = 1, \ldots, N$ from the nonparametric regression model:

$$f_j = \mu(x_j) + \varepsilon_j, \quad j = 1, \ldots, N,$$

where $\varepsilon_j \sim$ i. i. d. $(0, \sigma^2)$, and $x_j = \frac{j}{N}$ from an equidistant grid in the unit interval $[0, 1]$.

We are interested in this chapter in getting an estimate of the function μ using the general approach for image denoising proposed by Mrázek et al. [25]. Using that approach we consider the problem:

$$\begin{aligned} Q(u_1, \ldots, u_N) = &\sum_{i,j=1}^{N} \rho(u_i - f_j) K_h(x_i - x_j) \\ &+ \frac{\lambda}{2} \sum_{i,j=1}^{N} \frac{1}{2} (u_i - u_j)^2 L_g(x_i - x_j) = \min_{u_1, \ldots, u_N}! \end{aligned} \tag{5.1}$$

where the kernels K and L are nonnegative, symmetric functions on $\mathbb{R}$, the bandwidths $h, g > 0$, and the regularization parameter $\lambda \geq 0$. The function ρ is the modified Huber function as given by (2.11).

The solution $\hat{\mathbf{u}} = (\hat{u}_1, ..., \hat{u}_N)^T$ of problem (5.1) estimates $\boldsymbol{\mu} = (\mu(x_1), ..., \mu(x_N))^T$ and is called the *quadratically regularized local Huber M-estimate*, abbreviated as the QRLHM-estimate.

A special case of the QRLHM-estimate is the *local Huber M-estimate* already discussed in Chapter 2. This could be seen by setting $\lambda = 0$ in problem (5.1).

5.2 A Rough Approximation

If we look at problem (5.1) we can see that it does not have an explicit solution due to the structure of the Huber function. But solving the problem is still possible. One way is to solve the problem numerically. An example of a numerical solution is given in the following theorem.

Theorem 5.1 *Let the model (2.1) hold. The solution of (5.1) can be approximated by the Priestley-Chao kernel estimate $\hat{\boldsymbol{\mu}}$ or by the LHM-estimate $\tilde{\boldsymbol{\mu}}$. That is,*

$$\hat{\mathbf{u}} \approx \hat{\boldsymbol{\mu}} - [\nabla^2 Q(\hat{\boldsymbol{\mu}})]^{-1} \nabla Q(\hat{\boldsymbol{\mu}}) \tag{5.2}$$

provided that $\nabla^2 Q(\hat{\boldsymbol{\mu}})$ is invertible, and

$$\hat{\mathbf{u}} \approx \tilde{\boldsymbol{\mu}} - [\nabla^2 Q(\tilde{\boldsymbol{\mu}})]^{-1} \nabla Q(\tilde{\boldsymbol{\mu}}) \tag{5.3}$$

provided that $\nabla^2 Q(\tilde{\boldsymbol{\mu}})$ is invertible.

Proof. The vector of Priestley-Chao kernel estimates with kernel K and bandwidth h is given at the grid points by

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}_K(x_1, h), \ldots, \hat{\mu}(x_N, h))^T = (\hat{\mu}_1, \ldots, \hat{\mu}_N)^T$$

and the vector of LHM-estimates at the grid points is given by

$$\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_K(x_1, h), \ldots, \tilde{\mu}(x_N, h))^T = (\tilde{\mu}_1, \ldots, \tilde{\mu}_N)^T .$$

The QRLHM-estimate $\hat{\mathbf{u}}$ solves (5.1) so $\nabla Q(\hat{\mathbf{u}}) = \mathbf{0}_{N\times 1}$. By a Taylor series expansion of ∇Q around $\hat{\boldsymbol{\mu}}$ we have

$$\begin{aligned}
\nabla Q(\hat{\mathbf{u}}) &= \nabla Q(\hat{\boldsymbol{\mu}}) + \nabla^2 Q(\hat{\boldsymbol{\mu}})(\hat{\mathbf{u}} - \hat{\boldsymbol{\mu}}) + o\left(\|\hat{\mathbf{u}} - \hat{\boldsymbol{\mu}}\|\right) \\
\nabla Q(\hat{\mathbf{u}}) &\approx \nabla Q(\hat{\boldsymbol{\mu}}) + \nabla^2 Q(\hat{\boldsymbol{\mu}})(\hat{\mathbf{u}} - \hat{\boldsymbol{\mu}}) \\
\mathbf{0}_{N\times 1} &\approx \nabla Q(\hat{\boldsymbol{\mu}}) + \nabla^2 Q(\hat{\boldsymbol{\mu}})(\hat{\mathbf{u}} - \hat{\boldsymbol{\mu}}) \\
\hat{\mathbf{u}} &\approx \hat{\boldsymbol{\mu}} - [\nabla^2 Q(\hat{\boldsymbol{\mu}})]^{-1} \nabla Q(\hat{\boldsymbol{\mu}}).
\end{aligned}$$

The approximation is exactly the same using $\tilde{\boldsymbol{\mu}}$. Both approximations rely on the facts that $\|\hat{\mathbf{u}} - \hat{\boldsymbol{\mu}}\|$ and $\|\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}}\|$ go to zero as $N \to \infty$. □

However, this is not really helpful if we are interested in studying the properties of the solution. For that purpose, we need a more precise result. To do that we will use assumptions (A1), (A2) and (E1), which are already stated in Chapter 2.

5.3 Notation: The Gradient and the Hessian

In this section we establish some results for the gradient vector and the Hessian matrix of the energy functional Q.

The gradient vector and its Jacobian are defined as follows,

$$\nabla Q(\mathbf{u}) = (Q_1(\mathbf{u}), \dots, Q_N(\mathbf{u}))^T \quad \text{and} \quad \nabla^2 Q(\mathbf{u}) = (Q_{i,\ell}(\mathbf{u}))_{i,\ell},$$

where $i, \ell = 1, \dots, N$. The Jacobian of the gradient vector is a matrix and is called the Hessian matrix. Q_i stands for the first partial derivative of the energy functional Q with respect to the i^{th} entry of the vector $\mathbf{u} \in \mathbb{R}^N$ and $Q_{i,\ell}$ stands for the second partial derivative of the energy functional Q with respect to the i^{th} and the ℓ^{th} entry of the vector $\mathbf{u} \in \mathbb{R}^N$.

That is,

$$Q_i(\mathbf{u}) := \frac{\partial Q(\mathbf{u})}{\partial u_i} = \sum_{j=1}^{N} \rho'(u_i - f_j) K_h(x_i - x_j) + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} (u_i - u_j) L_g(x_i - x_j),$$

$$Q_{i,i}(\mathbf{u}) := \frac{\partial^2 Q(\mathbf{u})}{\partial u_i^2} = \sum_{j=1}^{N} \rho''(u_i - f_j) K_h(x_i - x_j) + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} L_g(x_i - x_j),$$

$$Q_{i,\ell}(\mathbf{u}) := \frac{\partial^2 Q(\mathbf{u})}{\partial u_\ell \partial u_i} = -\lambda L_g(x_i - x_\ell), \quad \ell \neq i.$$

We define the vector

$$G := \frac{\nabla Q(\tilde{\boldsymbol{\mu}})}{N}, \tag{5.4}$$

and the matrix

$$J := \frac{\nabla^2 Q(\tilde{\boldsymbol{\mu}})}{N} = \tilde{P} - \lambda \Lambda, \tag{5.5}$$

where for $i, \ell = 1, \dots, N$

$$\tilde{P}_{i,\ell} = \begin{cases} \frac{1}{N} \sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N} \sum_{j=1}^{N} L_g(x_i - x_j), & i = \ell, \\ 0, & i \neq \ell, \end{cases} \tag{5.6}$$

and

$$\Lambda_{i,\ell} = \frac{1}{N} L_g(x_i - x_\ell). \tag{5.7}$$

Mark that for quadratic ρ, i.e. $\rho(u) = \frac{1}{2}u^2$, where $\rho''(u) = 1$, the matrix $\tilde{P}$ coincides with the matrix $\hat{P}$ of Proposition 1.8.

Throughout the text we will use the following abbreviations,

$$Q_i = Q_i(\tilde{\boldsymbol{\mu}}) = \left.\frac{\partial Q(\mathbf{u})}{\partial u_i}\right|_{\mathbf{u}=\tilde{\boldsymbol{\mu}}} \qquad \text{and} \qquad Q_{i,\ell} = Q_{i,\ell}(\tilde{\boldsymbol{\mu}}) = \left.\frac{\partial^2 Q(\mathbf{u})}{\partial u_\ell \partial u_i}\right|_{\mathbf{u}=\tilde{\boldsymbol{\mu}}},$$

to stand for the derivatives evaluated at the LHM-estimate. If the derivatives are evaluated at any other point it will be explicitly indicated.

5.4 Auxiliary Results

In this section we present some auxiliary results regarding the gradient vector and the above mentioned matrices.

We start with an asymptotic approximation to the gradient vector.

Lemma 5.2 *Let the model (2.1) hold. Assume (A1) a)-c) for the kernel L and assume (A2). For $N \to \infty$, let $g \to 0$ such that $Ng^2 \to \infty$. Then, we have*

$$\frac{1}{N}\sum_{j=1}^{N}\big(\mu(x_i) - \mu(x_j)\big)L_g(x_i - x_j) = o(1)$$

uniformly in $x_i \in [g, 1-g]$.

Proof. Let $x_i \in [g, 1-g]$ then using Lemma 2.3 and the Lipschitz continuity of μ (from assumptions (A2)) we have

$$\begin{aligned}\sup_{x_i}\left|\frac{1}{N}\sum_{j=1}^{N}\big(\mu(x_i) - \mu(x_j)\big)L_g(x_i - x_j)\right| &\le C_\mu g \sup_{x_i}\frac{1}{N}\sum_{j=1}^{N}L_g(x_i - x_j)\\ &= C_\mu g \cdot O(1)\\ &= o(1).\end{aligned}$$

□

Proposition 5.3 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$ and $g \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then,*

$$\frac{Q_i}{N} = o_p(\lambda)$$

uniformly in $x_i \in [\max(h, g), 1 - \max(h, g)]$.

Proof. Let $x_i \in [\max(h, g), 1-\max(h, g)]$ then using Corollary 3.8 and the previous lemma

$$\sup_{x_i}\left|\frac{1}{\lambda}\frac{Q_i}{N}\right| = \sup_{x_i}\left|\frac{1}{N}\sum_{j=1}^{N}\left(\tilde{\mu}_K(x_i, h) - \tilde{\mu}_K(x_j, h)\right)L_g(x_i - x_j)\right|$$

$$\leq \sup_{x_i} \left| \frac{1}{N} \sum_{j=1}^{N} \left(\tilde{\mu}_K(x_i,h) - \mu(x_i)\right) L_g(x_i - x_j) \right|$$
$$+ \sup_{x_i} \left| \frac{1}{N} \sum_{j=1}^{N} \left(\mu(x_i) - \mu(x_j)\right) L_g(x_i - x_j) \right|$$
$$+ \sup_{x_i} \left| \frac{1}{N} \sum_{j=1}^{N} \left(\mu(x_j) - \tilde{\mu}_K(x_j,h)\right) L_g(x_i - x_j) \right|$$
$$= o_p(1) + o(1) + o_p(1) = o_p(1).$$

□

Now, we write the result in vector form using the notation introduced in Chapter 4.

Proposition 5.4 (Norm of G) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to$ and $g \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then,*

$$\|G\| = o_p(\lambda).$$

for $x_1, \dots, x_N \in [\max(h,g), 1 - \max(h,g)]$.

Proof. Using the norm defined in (4.5) we have

$$\|G\| = \|G^*\|_{\ell^\infty(\mathbb{R})} = \sup_{i \in \mathbb{N}} |G_i^*| = \sup_{1 \leq i \leq N} |G_i| = \sup_{1 \leq i \leq N} \left| \frac{Q_i}{N} \right| = o_p(\lambda),$$

for $x_1, \dots, x_N \in [\max(h,g), 1 - \max(h,g)]$. □

We now give an approximation to the second derivative of the data term of the energy functional Q. The result is similar to Proposition 2.17.

Lemma 5.5 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernel K and assume (A2) and (E1) a)-b). Then, for $N \to \infty$, $h \to 0$ such that $Nh^2 \to \infty$ and for every $c \in (0, \infty)$ we have*

$$\frac{1}{N} \sum_{j=1}^{N} K_h(x_i - x_j) \rho''(\tilde{\mu}_i - f_j) \xrightarrow{P} \eta_c$$

for all $x_i \in [h, 1-h]$.

Proof. For $i = 1, \dots, N$ define

$$Y_i = \frac{1}{N} \sum_{j=1}^{N} K_h(x_i - x_j) \mathbb{1}_{(-c,c)}(\tilde{\mu}_i - f_j).$$

Using the facts that $\tilde{\mu}$ is consistent, μ is Lipschitz continuous, and that the indicator function is almost surely continuous with respect to the probability measure of ε_j, we get from the continuous mapping theorem (Theorem 2.14) that

$$\mathbb{1}_{(-c,c)}(\tilde{\mu}_i - f_j) \xrightarrow{\mathcal{L}} \mathbb{1}_{(-c,c)}(-\varepsilon_j) = \mathbb{1}_{(-c,c)}(\varepsilon_j),$$

for all $|x_i - x_j| \leq h$. Therefore, since the indicator function is bounded, we get from the definition of weak convergence that

$$\mathbb{E}\,\mathbb{1}_{(-c,c)}(\tilde{\mu}_i - f_j) \to \mathbb{E}\,\mathbb{1}_{(-c,c)}(\varepsilon_j) = \eta,$$

for all $|x_i - x_j| \leq h$. Using Lemma 2.3 we get

$$\mathbb{E}\,Y_i = \frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\mathbb{E}\,\mathbb{1}_{(-c,c)}(\tilde{\mu}_i - f_j) \to \eta.$$

To make notation easier we define $A_{ij} = (-c + \tilde{\mu}_i - \mu_j, c + \tilde{\mu}_i - \mu_j)$ for $i, j = 1, \ldots, N$. Then,

$$\operatorname{var} Y_i = \frac{1}{N^2}\sum_{j,k=1}^{N} K_h(x_i - x_j)K_h(x_i - x_k)\operatorname{cov}\left(\mathbb{1}_{A_{ij}}(\varepsilon_j), \mathbb{1}_{A_{ik}}(\varepsilon_k)\right).$$

Since $\varepsilon_1, \ldots, \varepsilon_N$ are independent, then for $j, k = 1, \ldots, N$ such that $j \neq k$ we get

$$\begin{aligned}\operatorname{cov}\left[\mathbb{1}_{A_{ij}}(\varepsilon_j), \mathbb{1}_{A_{ik}}(\varepsilon_k)\right] &= \mathbb{E}\left[\mathbb{1}_{A_{ij}}(\varepsilon_j)\cdot \mathbb{1}_{A_{ik}}(\varepsilon_k)\right] - \left[\mathbb{E}\,\mathbb{1}_{A_{ij}}(\varepsilon_j)\right]\cdot\left[\mathbb{E}\,\mathbb{1}_{A_{ik}}(\varepsilon_k)\right] \\ &= \mathbb{P}(\varepsilon_j \in A_{ij} \wedge \varepsilon_k \in A_{ik}) - \mathbb{P}(\varepsilon_j \in A_{ij})\cdot\mathbb{P}(\varepsilon_k \in A_{ik}) \\ &= \mathbb{P}(\varepsilon_j \in A_{ij})\cdot\mathbb{P}(\varepsilon_k \in A_{ik}) - \mathbb{P}(\varepsilon_j \in A_{ij})\cdot\mathbb{P}(\varepsilon_k \in A_{ik}) \\ &= 0,\end{aligned}$$

and

$$\operatorname{var}\left[\mathbb{1}_{A_{ij}}(\varepsilon_j)\right] = \begin{cases}\eta(1-\eta) & c \geq 1,\\ \frac{1}{c^2}\eta(1-\eta) & c < 1,\end{cases}$$

for all $i, j = 1, \ldots, N$ such that $|x_i - x_j| \leq h$. Using Lemma 2.3 we have

$$\operatorname{var} Y_i = \frac{1}{N^2}\sum_{j=1}^{N} K_h^2(x_i - x_j)\operatorname{var}\left[\mathbb{1}_{A_{ij}}(\varepsilon_j)\right] = O\left(\frac{1}{Nh}\right) \to 0.$$

Therefore, $Y_i \xrightarrow{L^2} \eta$ and hence in probability.

Since

$$\rho''(u) = \begin{cases}\mathbb{1}_{(-c,c)}(u), & c \geq 1,\\ \frac{1}{c}\mathbb{1}_{(-c,c)}(u), & c < 1,\end{cases}$$

almost surely with respect to the probability measure of ε_j, we get

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}_i - f_j) = \begin{cases} Y_i, & c \geq 1, \\ \frac{Y_i}{c}, & c < 1, \end{cases} \quad \text{a.s.}$$

and therefore,

$$\frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}_i - f_j) \xrightarrow{P} \eta_c.$$

□

Lemma 5.6 *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Let K satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then for any $c \in (0, \infty)$ we have*

$$(1) \sup_{h \leq x_i \leq 1-h} \left| \frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}(x_i) - f_j) - \eta_c \right| = O_{a.s.}(1).$$

$$(2) \inf_{h \leq x_i \leq 1-h} \frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}(x_i) - f_j) \geq \frac{\eta_c}{2} \ a.s.$$

$$(3) \sup_{h \leq x_i \leq 1-h} \left| \frac{1}{\frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}(x_i) - f_j)} - \frac{1}{\eta_c} \right| = O_{a.s.}(1).$$

Proof. The proof of the first relation is direct from Lemma 2.3 and the boundedness of the indicator function.

From the proof of Lemma 5.5

$$\mathbb{E}\,\frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}(x_i) - f_j) \xrightarrow{N\to\infty} \eta_c.$$

An argument similar to that used in the proof of Lemma 2.19 yields

$$\inf_{h \leq x_i \leq 1-h} \frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j)\rho''(\tilde{\mu}(x_i) - f_j) \geq \frac{\eta_c}{2} \ a.s.$$

The third relation follows directly from the second one. □

We now present some results regarding the norms of the matrices $\tilde{P}^{-1}$, $\tilde{P}^{-1}\Lambda$ and J^{-1}. These results will be helpful tools for getting a more precise form of the QRLHM-estimate.

Proposition 5.7 (Norm of $\tilde{P}^{-1}$) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and*

(E1) a)-b). For $N \to \infty$, let $h \to 0$, such that $Nh^2 \to \infty$ then we have

$$\|\tilde{P}^{-1}\| = O_{a.s.}(1)$$

for $x_1, \ldots, x_N \in [h, 1-h]$, $g \geq 0$, and $\lambda \geq 0$.

Proof. Since L is nonnegative and using Lemma 5.6 we have

$$\begin{aligned}
\|\tilde{P}^{-1}\| = & \sup_{1\leq i\leq N} \left| \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)} \right| \\
\leq & \sup_{1\leq i\leq N} \left| \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j)} \right| \\
\leq & \frac{2}{\eta_c} \ a.s.
\end{aligned}$$

for $x_1, \ldots, x_N \in [h, 1-h]$. □

Proposition 5.8 (Norm of $\tilde{P}^{-1}\Lambda$) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$ then we have*

$$\|\tilde{P}^{-1}\Lambda\| = O_{a.s.}(1), \quad \|\lambda\tilde{P}^{-1}\Lambda\| = o_{a.s.}(1), \quad \text{and} \quad \|\lambda\tilde{P}^{-1}\Lambda\| < 1 \ \ a.s.$$

for $x_1, \ldots, x_N \in [\max(h, g), 1 - \max(h, g)]$.

Proof. From Lemma 2.3 for the kernel L and Lemma 5.6 we get that

$$\begin{aligned}
\|\tilde{P}^{-1}\Lambda\| = & \sup_{1\leq i\leq N} \sum_{j=1}^{N} \frac{|\Lambda_{i,j}|}{|\tilde{P}_{i,i}|} \\
= & \sup_{1\leq i\leq N} \frac{\frac{1}{N}\sum_{j=1}^{N} L_g(x_i - x_j)}{\left|\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)\right|} \\
\overset{a.s.}{\leq} & \frac{2}{\eta_c} \sup_{1\leq i\leq N} \frac{1}{N}\sum_{j=1}^{N} L_g(x_i - x_j) = O(1)
\end{aligned}$$

for $x_1, \ldots, x_N \in [\max(h, g), 1 - \max(h, g)]$.

Therefore,

$$\|\tilde{P}^{-1}\Lambda\| = O_{a.s.}(1) \quad \text{and} \quad \|\lambda\tilde{P}^{-1}\Lambda\| \xrightarrow{a.s.} 0.$$

Since $\rho''(u) \geq 0$ for all $u \in \mathbb{R} \setminus \{-c, c\}$, we have for all $i = 1, \ldots, N$

$$\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) \geq 0$$

almost surely with respect to the probability measure of ε_j.

That is, for all i we have almost surely that

$$\frac{\frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)} \leq 1$$

Therefore, $\|\lambda\tilde{P}^{-1}\Lambda\| \leq 1$ almost surely. What is left to be shown is that $\|\lambda\tilde{P}^{-1}\Lambda\| \neq 1$ almost surely.

Consider,

$$\begin{aligned}
&\mathbb{P}\left(\|\lambda\tilde{P}^{-1}\Lambda\| = 1\right) \\
&= \mathbb{P}\left(\sup_{1\leq i\leq N} \frac{\frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)} = 1\right) \\
&= \mathbb{P}\left(\frac{\frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_k - x_j)}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_k - f_j) K_h(x_k - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_k - x_j)} = 1, \quad \text{for some } k\right) \\
&= \mathbb{P}\left(\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_k - f_j) K_h(x_k - x_j) = 0, \quad \text{for some } k\right) \\
&= 0,
\end{aligned}$$

since

$$\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_k - f_j) K_h(x_k - x_j) \geq \frac{1}{2}\eta_c \; a.s.$$

uniformly in $x_k \in [h, 1-h]$. □

Proposition 5.9 (Norm of J^{-1}) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and*

(E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$ then we have

$$\|J^{-1}\| = O_{a.s.}(1)$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$.

Proof. Using Neumann series and Propositions 5.7 and 5.8 we get

$$\begin{aligned}
\|J^{-1}\| = \|(\tilde{P} - \lambda\Lambda)^{-1}\| &\le \|(I_N - \lambda\tilde{P}^{-1}\Lambda)^{-1}\| \cdot \|\tilde{P}^{-1}\| \\
&\overset{a.s.}{\le} \frac{1}{1 - \|\lambda\tilde{P}^{-1}\Lambda\|} \cdot \|\tilde{P}^{-1}\| \\
&= \frac{1}{1 + o_{a.s.}(1)} \cdot O_{a.s.}(1) = O_{a.s.}(1).
\end{aligned}$$

□

5.5 Vector and Component Form of QRLHM-Estimates

Now, we can present more precise results regarding the asymptotic approximation of the QRLHM-estimate in terms of the LHM-estimate $\tilde{\boldsymbol{\mu}}$.

Theorem 5.10 (QRLHM Vector) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$. If $\nabla^2 Q(\tilde{\boldsymbol{\mu}})$ is invertible, then the solution of problem (5.1) is given by,*

$$\hat{\mathbf{u}} = \tilde{\boldsymbol{\mu}} - [\nabla^2 Q(\tilde{\boldsymbol{\mu}})]^{-1} \nabla Q(\tilde{\boldsymbol{\mu}}) + o_{a.s.}\left(\frac{1}{N}\right) \tag{5.8}$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$.

Proof. Using a Taylor series expansion of the gradient vector ∇Q around $\tilde{\boldsymbol{\mu}}$ or equivalently using the definition of the derivative for ∇Q as a function from $\mathbb{R}^N$ to $\mathbb{R}^N$, we have

$$\nabla Q(\hat{\mathbf{u}}) = \nabla Q(\tilde{\boldsymbol{\mu}}) + \nabla^2 Q(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}}) + o\left(\|\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}}\|\right),$$

that is,

$$\frac{\|\nabla Q(\hat{\mathbf{u}}) - \nabla Q(\tilde{\boldsymbol{\mu}}) - \nabla^2 Q(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}})\|}{\|\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}}\|} \to 0 \quad \text{as} \quad \hat{\mathbf{u}} \to \tilde{\boldsymbol{\mu}}.$$

Since $\hat{\mathbf{u}}$ is the solution of problem (5.1) then $\nabla Q(\hat{\mathbf{u}}) = \mathbf{0}_{N\times 1}$.

Letting $\lambda \to 0$ implies that

$$\|\hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}}\| \to 0.$$

Using Proposition 5.9 we have

$$
\begin{aligned}
N\left\|\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}}+J^{-1}G\right\| &= N\left\|J^{-1}\{J(\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}})+G\}\right\| \\
&\leq N\|J^{-1}\|\cdot\|J(\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}})+G\| \\
&= \|J^{-1}\|\cdot\|\nabla^2Q(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}})+\nabla Q(\tilde{\boldsymbol{\mu}})\| \\
&= \|J^{-1}\|\cdot\frac{\|\nabla^2Q(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}})+\nabla Q(\tilde{\boldsymbol{\mu}})\|}{\|\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}}\|}\cdot\|\hat{\mathbf{u}}-\tilde{\boldsymbol{\mu}}\| \\
&= O_{a.s.}(1)\cdot o(1)\cdot o(1)=o_{a.s.}(1),
\end{aligned}
$$

that is,

$$\hat{\mathbf{u}}=\tilde{\boldsymbol{\mu}}-[\nabla^2Q(\tilde{\boldsymbol{\mu}})]^{-1}\nabla Q(\tilde{\boldsymbol{\mu}})+o_{a.s.}\left(\frac{1}{N}\right).$$

□

To get the bias and variance terms it would be easier to have a componentwise version of the previous theorem.

Theorem 5.11 (QRLHM Component 1) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N\to\infty$, let $h\to 0$, $g\to 0$ and $\lambda\to 0$ such that $h\sim$ constant $N^{-1/5}$ and $Ng^2\to\infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1}=O\left(\frac{1}{N}\right), \tag{5.9}$$

we have uniformly in $x_i\in[\max(h,g)+tg, 1-\max(h,g)-tg]$, that:

$$
\begin{aligned}
\hat{u}_i=\tilde{\mu}_i-\frac{1}{\eta_c+\lambda}\frac{Q_i}{N}-\frac{1}{\eta_c+\lambda}\sum_{n=1}^{t}\left(\frac{\lambda}{\eta_c+\lambda}\right)^n\frac{1}{N^n}\sum_{j_1,\ldots,j_n}L_g(x_i-x_{j_1})\cdots \\
\cdots L_g(x_{j_{n-1}}-x_{j_n})\frac{Q_{j_n}}{N}+R_{N,i}
\end{aligned}
$$

where $\tilde{\mu}_i=\tilde{\mu}_K(x_i,h)$, $Q_i=\left.\frac{\partial Q(\mathbf{u})}{\partial u_i}\right|_{\mathbf{u}=\tilde{\boldsymbol{\mu}}}$, and $R_{N,i}=o_p(\lambda)$.

Proof. Using Neumann series and Propositions 5.7 and 5.8 we get

$$
\begin{aligned}
\frac{1}{\lambda^{t+1}}\left\|J^{-1}-\sum_{j=0}^{t}(\lambda\tilde{P}^{-1}\Lambda)^j\tilde{P}^{-1}\right\| &= \frac{1}{\lambda^{t+1}}\left\|(I-\lambda\tilde{P}^{-1}\Lambda)^{-1}\tilde{P}^{-1}-\sum_{j=0}^{t}(\lambda\tilde{P}^{-1}\Lambda)^j\tilde{P}^{-1}\right\| \\
&\overset{a.s.}{\leq} \frac{1}{\lambda^{t+1}}\left\|\sum_{j=0}^{\infty}(\lambda\tilde{P}^{-1}\Lambda)^j-\sum_{j=0}^{t}(\lambda\tilde{P}^{-1}\Lambda)^j\right\|\cdot\left\|\tilde{P}^{-1}\right\|
\end{aligned}
$$

$$\begin{aligned}
&\leq \frac{1}{\lambda^{t+1}} \|\tilde{P}^{-1}\| \sum_{j=t+1}^{\infty} \|\lambda \tilde{P}^{-1}\Lambda\|^j \\
&= \frac{1}{\lambda^{t+1}} \|\tilde{P}^{-1}\| \cdot \|\lambda \tilde{P}^{-1}\Lambda\|^{t+1} \sum_{j=0}^{\infty} \|\lambda \tilde{P}^{-1}\Lambda\|^j \\
&\overset{a.s.}{=} \|\tilde{P}^{-1}\| \cdot \|\tilde{P}^{-1}\Lambda\|^{t+1} \cdot \frac{1}{1 - \|\lambda \tilde{P}^{-1}\Lambda\|} \\
&= O_{a.s.}(1) \cdot O_{a.s.}(1) \cdot \frac{1}{1 + o_{a.s.}(1)} \\
&= O_{a.s.}(1),
\end{aligned}$$

that is,

$$J^{-1} = \sum_{j=0}^{t} (\lambda \tilde{P}^{-1}\Lambda)^j \tilde{P}^{-1} + O_{a.s.}(\lambda^{t+1}).$$

Using the previous expansion of J^{-1} and Proposition 5.4, we get from (5.8) that

$$\begin{aligned}
&N \left\| \hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}} + \left(\sum_{j=0}^{t} (\lambda \tilde{P}^{-1}\Lambda)^j \tilde{P}^{-1} \right) \cdot G \right\| \\
&\leq N \left\| \hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}} + J^{-1} \cdot G \right\| + N \left\| \left(\sum_{j=0}^{t} (\lambda \tilde{P}^{-1}\Lambda)^j \tilde{P}^{-1} \right) \cdot G - J^{-1} \cdot G \right\| \\
&\leq N \left\| \hat{\mathbf{u}} - \tilde{\boldsymbol{\mu}} + J^{-1} \cdot G \right\| + N \left\| \sum_{j=0}^{t} (\lambda \tilde{P}^{-1}\Lambda)^j \tilde{P}^{-1} - J^{-1} \right\| \cdot \|G\| \\
&= o_{a.s.}(1) + N \cdot O_{a.s.}(\lambda^{t+1}) \cdot o_p(\lambda) \\
&= o_p(1).
\end{aligned}$$

For the last step we have assumed that $N\lambda^{t+1} = O(1)$.

Therefore,

$$\hat{\mathbf{u}} = \tilde{\boldsymbol{\mu}} - \left(\sum_{j=0}^{t} (\lambda \tilde{P}^{-1}\Lambda)^j \tilde{P}^{-1} \right) \frac{\nabla Q(\tilde{\boldsymbol{\mu}})}{N} + o_p\left(\frac{1}{N}\right). \tag{5.10}$$

After calculating the matrices, the componentwise version is,

$$\begin{aligned}
\hat{u}_i = \tilde{\mu}_i - \frac{1}{\tilde{P}_{ii}} \frac{Q_i}{N} - \frac{1}{\tilde{P}_{ii}} \sum_{n=1}^{t} \frac{\lambda^n}{N^n} \sum_{j_1,\ldots,j_n} \frac{L_g(x_i - x_{j_1})}{\tilde{P}_{j_1 j_1}} \cdots \\
\cdots \frac{L_g(x_{j_{n-1}} - x_{j_n})}{\tilde{P}_{j_n j_n}} \frac{Q_{j_n}}{N} + o_p\left(\frac{1}{N}\right)
\end{aligned} \tag{5.11}$$

uniformly in $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$.

From Propositions 5.3 and 5.7 we have

$$\left| \frac{1}{\tilde{P}_{jj}} \frac{Q_j}{N} - \frac{1}{\eta_c + \lambda} \frac{Q_j}{N} \right| = \left| \frac{1}{\tilde{P}_{jj}} - \frac{1}{\eta_c + \lambda} \right| \cdot \left| \frac{Q_j}{N} \right| = O_{a.s.}(1) \cdot o_p(\lambda) = o_p(\lambda), \tag{5.12}$$

uniformly in $x_j \in [\max(h,g), 1 - \max(h,g)]$.

Finally, we use (5.12) to interchange $\frac{1}{\tilde{P}_{\bullet\bullet}} \frac{Q_\bullet}{N}$ with $\frac{1}{\eta_c+\lambda} \frac{Q_\bullet}{N}$ in (5.11). □

5.6 Bias and Variance of QRLHM-Estimates

Now, we will try to get a more compact form of the QRLHM-estimate which was given in Theorem 5.11. For that purpose we will define the LHM-iterated smoothers and the LHM-iterated differences (compare to Definition 1.10).

5.6.1 LHM-Iterated Smoothers (ILHM-Estimates)

Definition 5.12 (ILHM) *Set $\tilde{\mu}_1(x,h,g) := \tilde{\mu}_K(x,h)$, and recursively define the "LHM-iterated smoothers" as follows*

$$\tilde{\mu}_{n+1}(x,h,g) := \frac{1}{N} \sum_{j=1}^{N} L_g(x - x_j) \tilde{\mu}_n(x_j, h, g), \quad n \geq 1.$$

Using the recursion above

$$\tilde{\mu}_{n+1}(x,h,g) = \frac{1}{N^n} \sum_{j_1,\dots,j_n} L_g(x - x_{j_1}) \dots L_g(x_{j_{n-1}} - x_{j_n}) \tilde{\mu}_K(x_{j_n}, h).$$

The "LHM-iterated differences" are defined recursively as follows

$$\tilde{\nu}_{n+1}(x,h,g) := \tilde{\mu}_{n+1}(x,h,g) - \tilde{\mu}_n(x,h,g), \quad n \geq 1.$$

Getting the bias, variance and covariance terms of the iterated LHM-smoothers depends on Theorems 2.20 and 2.22. The calculations are very similar to the case of the PC-iterated smoothers discussed in [10], therefore the proofs are skipped.

Proposition 5.13 (ILHM Bias) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^4 \to \infty$. Then, we have for all $n \geq 1$ and for all $h + ng \leq x \leq 1 - (h + ng)$ that*

$$\operatorname{bias} \tilde{\mu}_n(x,h,g) = \left\{ \frac{1}{2} h^2 \mu''(x) V_K + o(h^2) \right\} + (n-1) \left\{ \frac{1}{2} g^2 \mu''(x) V_L + o(g^2) \right\}.$$

Consequently,

$$\mathbb{E}\, \tilde{\nu}_{n+1}(x,h,g) = \frac{1}{2} g^2 \mu''(x) V_L + o(g^2).$$

Proposition 5.14 (ILHM Variance) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, for all $n \geq m \geq 0$,*

$$\operatorname{cov}\left(\tilde{\mu}_{m+1}(x,h,g), \tilde{\mu}_{n+1}(\bar{x},h,g)\right) = \frac{\sigma_c^2}{N} L_g^{*(n+m)} * K_h * K_h(x-\bar{x}) + o\left(\frac{1}{Nh}\right)$$

for all $2h + ng \leq x, \bar{x} \leq 1 - (2h+ng)$. In particular,

$$\operatorname{var} \tilde{\mu}_{n+1}(x,h,g) = \frac{\sigma_c^2}{N} L_g^{*(2n)} * K_h * K_h(0) + o\left(\frac{1}{Nh}\right).$$

Using the LHM-iterated smoothers we will write the QRLHM-estimate in a compact form. This compact form will make the calculations of the bias and variance terms easier.

Theorem 5.15 (QRLHM Component 2) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have uniformly in $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$, that:

$$\hat{u}_i = \tilde{\mu}_i + \sum_{n=0}^{t} \delta^{n+1} \tilde{\nu}_{n+2}(x_i,h,g) + R_{N,i} \tag{5.13}$$

where $\delta = \frac{\lambda}{\eta_c + \lambda}$ and $R_{N,i} = o_p(\lambda)$. Moreover,

$$\hat{u}_i = (1-\delta) \sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x_i,h,g) + R_{N,i} \tag{5.14}$$

uniformly in $x_i \in [\max(h,g) + (t+1)g, 1 - \max(h,g) - (t+1)g]$.

Proof. From Corollary 3.8 and assumptions (A2) we have

$$\tilde{\mu}(x_i, h) = \mu(x_i) + o_p(1) = O_p(1). \tag{5.15}$$

uniformly in $x_i \in [h, 1-h]$. Moreover, from Lemma 2.3 we have

$$\left| \frac{Q_i}{N} + \lambda \tilde{\nu}(x_i,h,g) \right|$$

$$
\begin{aligned}
&= \left| \frac{1}{N}\sum_{j=1}^{N} \rho'(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N}(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) + \lambda \tilde{\nu}_2(x_i, h, g) \right| \\
&= \left| \frac{\lambda}{N}\sum_{j=1}^{N}(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} L_g(x_i - x_j)\tilde{\mu}_j - \lambda\tilde{\mu}_i \right| \\
&= \lambda \left|\tilde{\mu}(x_i, h)\right| \cdot \left| \frac{1}{N}\sum_{j=1}^{N} L_g(x_i - x_j) - 1 \right| \\
&= \lambda \cdot O_p(1) \cdot O\left(\frac{\lambda}{Ng^2}\right) = o_p\left(\frac{\lambda}{Ng^2}\right)
\end{aligned}
$$

uniformly in $x_i \in [\max(h, g), 1 - \max(h, g)]$.

Interchanging $Q_\bullet/N$ with $\lambda\hat{\nu}(x_\bullet, h, g)$ in the $\hat{u}_i$ obtained in Theorem 5.11 we have,

$$
\begin{aligned}
\hat{u}_i = \tilde{\mu}_i &+ \frac{\lambda}{\eta_c + \lambda}\tilde{\nu}_2(x_i, h, g) \\
&+ \frac{\lambda}{\eta_c + \lambda}\sum_{n=1}^{t}\frac{\delta^n}{N^n}\sum_{j_1,\ldots,j_n} L_g(x_i - x_{j_1}) \cdots \\
&\qquad \cdots L_g(x_{j_{n-1}} - x_{j_n})(\tilde{\mu}_2(x_{j_n}, h, g) - \tilde{\mu}_1(x_{j_n}, h, g)) \\
&+ o_p(\lambda),
\end{aligned}
$$

and hence (5.13).

To get the second relation we rewrite the estimate as follows,

$$
\begin{aligned}
\hat{u}_i &= \tilde{\mu}_i + \sum_{n=0}^{t}\delta^{n+1}\tilde{\nu}_{n+2}(x_i, h, g) + o_p(\lambda) \\
&= \tilde{\mu}_1(x, h, g) + \sum_{n=0}^{t}\delta^{n+1}\tilde{\mu}_{n+2}(x_i, h, g) - \sum_{n=0}^{t}\delta^{n+1}\tilde{\mu}_{n+1}(x_i, h, g) + o_p(\lambda) \\
&= \sum_{n=-1}^{t-1}\delta^{n+1}\tilde{\mu}_{n+2}(x_i, h, g) + \delta^{t+1}\tilde{\mu}_{t+2}(x_i, h, g) - \delta\sum_{n=0}^{t}\delta^{n}\tilde{\mu}_{n+1}(x_i, h, g) + o_p(\lambda) \\
&= \sum_{k=0}^{t}\delta^{k}\tilde{\mu}_{k+1}(x_i, h, g) + \delta^{t+1}\tilde{\mu}_{t+2}(x_i, h, g) - \delta\sum_{n=0}^{t}\delta^{n}\tilde{\mu}_{n+1}(x_i, h, g) + o_p(\lambda) \\
\hat{u}_i &= (1 - \delta)\sum_{n=0}^{t}\delta^{n}\tilde{\mu}_{n+1}(x_i, h, g) + \delta^{t+1}\tilde{\mu}_{t+2}(x_i, h, g) + o_p(\lambda).
\end{aligned}
$$

From Corollary 3.8 and Lemmas 2.3 and 5.2 we have

$$
\tilde{\mu}_{t+2}(x_i, h, g) = \mu(x_i) + o_p(1) = O_p(1) \tag{5.16}
$$

uniformly in $x_i \in [\max(h,g)+(t+1)g, 1-\max(h,g)-(t+1)g]$. Therefore,

$$\hat{u}_i = (1-\delta)\sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x_i,h,g) + O_p(\lambda^{t+1}) + o_p(\lambda),$$

and hence (5.14). □

Important consequences of this result are the L_2 and L_1 limiting cases.

Corollary 5.16 *Let the assumptions of Theorem 5.15 hold.*

a) **The L_2 limiting case:** *If $c \to \infty$ then*

$$\hat{u}_i^{QRLLS} = \lim_{c\to\infty} \hat{u}_i^{QRLHM} = (1-\theta)\sum_{n=0}^{t} \theta^n \tilde{\mu}_{NW,n+1}(x_i,h,g) + o_p(\lambda),$$

for all $x_i \in [\max(h,g)+(t+1)g, 1-\max(h,g)-(t+1)g]$ where $\theta = \frac{\lambda}{1+\lambda}$.

b) **The L_1 limiting case:** *If $c \to 0$ and $p_\varepsilon(0) \neq 0$ then*

$$\hat{u}_i^{QRLLA} = \lim_{c\to 0} \hat{u}_i^{QRLHM} = (1-\zeta)\sum_{n=0}^{t} \zeta^n \tilde{\mu}_{LAD,n+1}(x_i,h,g) + o_p(\lambda),$$

for all $x_i \in [\max(h,g)+(t+1)g, 1-\max(h,g)-(t+1)g]$ where $\zeta = \frac{\lambda}{2p_\varepsilon(0)+\lambda}$.

Mark that part a) of the corollary above corresponds to result in Theorem 1.12.

5.6.2 The Bias Term

To get the bias term it will be easier to use the form obtained in (5.13).

Theorem 5.17 (QRLHM Bias) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^4 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have for all $x_i \in [\max(h,g)+(t+1)g, 1-\max(h,g)-(t+1)g]$, that:

$$\begin{aligned}
\operatorname{bias}\hat{u}_i &= \operatorname{bias}\tilde{\mu}_K(x_i,h) + \frac{\lambda}{\eta_c}\operatorname{bias}\tilde{\mu}_L(x_i,g) + o(\lambda) \\
&= \frac{1}{2}\mu''(x_i)\left\{h^2 V_K + \frac{\lambda}{\eta_c} g^2 V_L\right\} + o(h^2) + o(\lambda) \\
&= \frac{1}{2}h^2\mu''(x_i)V_K + o(h^2) + o(\lambda).
\end{aligned}$$

Proof. From Proposition 5.13 and (5.13) we have,

$$\begin{aligned}\text{bias}\,\hat{u}_i &= \text{bias}\,\tilde{\mu}_i + \sum_{n=0}^{t} \delta^{n+1} \mathbb{E}\,\tilde{\nu}_{n+2}(x_i,h,g) + \mathbb{E}\,R_{N,i} \\ &= \text{bias}\,\tilde{\mu}_i + \sum_{n=0}^{t} \delta^{n+1} \left(\text{bias}\,\tilde{\mu}_L(x_i,g) + o(g^2)\right) + \mathbb{E}\,R_{N,i} \\ &= \text{bias}\,\tilde{\mu}_i + \left(\text{bias}\,\tilde{\mu}_L(x_i,g) + o(g^2)\right) \sum_{n=0}^{t} \delta^{n+1} + \mathbb{E}\,R_{N,i}.\end{aligned}$$

Using the facts that $\delta = O(\lambda)$ and that $\lambda^{t+1} = O(\frac{1}{N})$ we get,

$$\sum_{k=0}^{t} \delta^{k+1} = \delta \frac{1-\delta^{t+1}}{1-\delta} = \frac{\lambda}{\eta_c} - \frac{\lambda}{\eta_c}\left(\frac{\lambda}{\eta_c+\lambda}\right)^{t+1} = \frac{\lambda}{\eta_c} + O\left(\lambda^{t+2}\right) = \frac{\lambda}{\eta_c} + o(\lambda). \tag{5.17}$$

Therefore,

$$\begin{aligned}\text{bias}\,\hat{u}_i &= \text{bias}\,\tilde{\mu}_K(x_i,h) + \frac{\lambda}{\eta_c}\,\text{bias}\,\tilde{\mu}_L(x_i,g) + o\left(\lambda g^2\right) + o(\lambda) \\ &= \frac{1}{2}\mu''(x_i)\left\{h^2 V_K + \frac{\lambda}{\eta_c} g^2 V_L\right\} + o(h^2) + o(\lambda).\end{aligned}$$

□

Remark 5.18 *Initially the bias seems to consists of two terms. One term of the bias is due to the data term of the minimization problem, which is* $\text{bias}\,\tilde{\mu}_K(x_i,h)$*. The other is due to the smoothness term of the minimization problem, that is* $\frac{\lambda}{\eta_c}\,\text{bias}\,\tilde{\mu}_L(x_i,g) + o(\lambda)$*.*

But since $g \to 0$ *we have* $\frac{\lambda}{\eta_c}\,\text{bias}\,\tilde{\mu}_L(x_i,g) \to 0$ *whatever* λ *is, that is, the contribution of the smoothness term is reduced to the term* $o(\lambda)$*.*

5.6.3 The Variance Term

We recall the Fourier transforms of K and L from (1.13), that is

$$\widehat{L}(\omega) = \int L(z)e^{-i\omega z}dz \quad \text{and} \quad \widehat{K}(\omega) = \int K(z)e^{-i\omega z}dz.$$

From (A1) K and L are symmetric and integrate to 1, hence $\widehat{K}$ and $\widehat{L}$ are real-valued and uniformly bounded in absolute value by 1.

Using Plancherel's theorem and the convolution theorem (for example see [26]) we get

$$L_g^{*(k+\ell)} * K_h * K_h(0) = \frac{1}{2\pi}\int \widehat{L}^{k+\ell}(g\omega)\widehat{K}^2(h\omega)d\omega. \tag{5.18}$$

To get the variance term it will be easier to use the form obtained in (5.14).

Theorem 5.19 (QRLHM Variance) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^4 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have for all $x_i \in [\max(2h, g) + (t+1)g, 1 - \max(2h, g) - (t+1)g]$, that:

$$\operatorname{var} \hat{u}_i = \frac{\sigma_c^2}{Nh} Q\left(\frac{g}{h}, \frac{\lambda}{\eta_c}\right) + o(\lambda^2) + o\left(\frac{\lambda}{\sqrt{Nh}}\right),$$

where

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

Proof. We write the QRLHM-estimate as in (5.14), i.e.

$$\hat{u}_i = (1-\delta) \sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x_i, h, g) + R_{N,i} = M_{N,i} + R_{N,i}.$$

From Cauchy-Schwarz inequality

$$\begin{aligned} \operatorname{var} \hat{u}_i &= \operatorname{var} M_{N,i} + \operatorname{var} R_{N,i} + 2 \operatorname{cov}(M_{N,i}, R_{N,i}) \\ &\le \operatorname{var} M_{N,i} + \operatorname{var} R_{N,i} + 2\sqrt{\operatorname{var} M_{N,i} \operatorname{var} R_{N,i}}. \end{aligned}$$

As $\lambda \to 0$ we get that $\delta \to 0$ and therefore

$$(1-\delta)^2 \sum_{k,\ell=0}^{t} \delta^{k+\ell} = (1-\delta^{t+1})^2 \to 1. \tag{5.19}$$

Using the above fact along with Proposition 5.14 and (5.18) we evaluate $\operatorname{var} M_{N,i}$,

$$\begin{aligned} \operatorname{var} M_{N,i} &= (1-\delta)^2 \sum_{k=0}^{t} \sum_{\ell=0}^{t} \delta^{k+\ell} \operatorname{cov}(\tilde{\mu}_{k+1}(x_i, h, g), \tilde{\mu}_{\ell+1}(x_i, h, g)) \\ &= (1-\delta)^2 \sum_{k,\ell=0}^{t} \delta^{k+\ell} \left\{ \frac{\sigma^2}{N} L_g^{*(k+\ell)} * K_h * K_h(0) + o\left(\frac{1}{Nh}\right) \right\} \\ &= \frac{\sigma^2}{2\pi N} (1-\delta)^2 \sum_{k,\ell=0}^{t} \delta^{k+\ell} \int \widehat{L}^{k+\ell}(g\omega) \widehat{K}^2(h\omega) d\omega + o\left(\frac{1}{Nh}\right) \end{aligned}$$

$$= \frac{\sigma^2}{2\pi N} \int (1-\delta)^2 \sum_{k,\ell=0}^{t} \delta^{k+\ell} \widehat{L}^{k+\ell}(g\omega) \widehat{K}^2(h\omega) d\omega + o\left(\frac{1}{Nh}\right)$$
$$= \frac{\sigma^2}{2\pi N} \int \left\{ (1-\delta) \sum_{k=0}^{t} \delta^k \widehat{L}^k(g\omega) \right\}^2 \widehat{K}^2(h\omega) d\omega + o\left(\frac{1}{Nh}\right).$$

So, we have to calculate the following term,

$$\begin{aligned}(1-\delta)\sum_{k=0}^{t}\delta^k\widehat{L}^k(g\omega) &= (1-\delta)\ \frac{1-\delta^{t+1}\widehat{L}^{t+1}(g\omega)}{1-\delta\widehat{L}(g\omega)} \\ &= \frac{1-\delta}{1-\delta\widehat{L}(g\omega)} + \frac{\delta^{t+2}\widehat{L}^{t+1}(g\omega)}{1-\delta\widehat{L}(g\omega)} \\ &= \frac{1-\delta}{1-\delta\widehat{L}(g\omega)} + O(\delta^{t+2}),\end{aligned}$$

substituting $\delta = \frac{\lambda}{\eta_c+\lambda}$ and using that $\delta = O(\lambda)$ and $N\lambda^{t+1} = O(1)$ we get,

$$\begin{aligned} &= \frac{1}{1+\frac{\lambda}{\eta_c}-\frac{\lambda}{\eta_c}\widehat{L}(g\omega)} + O(\lambda^{t+2}) \\ &= \frac{1}{1+\frac{\lambda}{\eta_c}-\frac{\lambda}{\eta_c}\widehat{L}(g\omega)} + O\left(\frac{\lambda}{N}\right).\end{aligned}$$

Then,

$$\begin{aligned}\operatorname{var} M_{N,i} &= \frac{\sigma^2}{2\pi N}\int\left\{\frac{1}{1+\frac{\lambda}{\eta_c}-\frac{\lambda}{\eta_c}\widehat{L}(g\omega)} + O\left(\frac{\lambda}{N}\right)\right\}^2 \widehat{K}^2(h\omega)d\omega + o\left(\frac{1}{Nh}\right) \\ &= \frac{\sigma^2}{2\pi N}\int\left\{\frac{\widehat{K}(h\omega)}{1+\frac{\lambda}{\eta_c}-\frac{\lambda}{\eta_c}\widehat{L}(g\omega)}\right\}^2 d\omega + \frac{Q_K}{Nh}O\left(\frac{\lambda}{N}\right) + o\left(\frac{1}{Nh}\right) \\ &= \frac{\sigma^2}{2\pi Nh}\int\left\{\frac{\widehat{K}(\omega)}{1+\frac{\lambda}{\eta_c}-\frac{\lambda}{\eta_c}\widehat{L}(\frac{g}{h}\omega)}\right\}^2 d\omega + o\left(\frac{1}{Nh}\right).\end{aligned}$$

Using the dominated convergence theorem, one can show that $\operatorname{var} R_{N,i} = o(\lambda^2)$, and we have already shown that $\operatorname{var} M_{N,i} = O\left(\frac{1}{Nh}\right)$. Therefore,

$$\operatorname{var} \hat{u}_i = \operatorname{var} M_{N,i} + o(\lambda^2) + o\left(\frac{\lambda}{\sqrt{Nh}}\right).$$

□

5.7 Consistency and Asymptotic Normality of QRLHM-Estimates

We have seen in the previous section that

$$\operatorname{mse} \hat{u}(x_i) \to 0$$

for all $x_i \in [\max(2h, g) + (t+1)g, 1 - \max(2h, g) - (t+1)g]$. Hence, we have pointwise consistency.

We have seen in Chapter 2 that the LHM-estimate has an asymptotic normal distribution, that is,

$$\sqrt{Nh}\left(\frac{\tilde{\mu}_K(x,h) - \mu(x) - \frac{1}{2}h^2\mu''(x)V_K}{\sqrt{\sigma_c^2 Q_K}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

for all $x \in [h, 1-h]$, where

$$\sigma_c^2 = \frac{\sigma_M^2}{\eta_c^2} = \frac{c^2(1-\eta)}{\eta^2} + \frac{1}{\eta^2}\int_{-c}^{c} y^2 p_\varepsilon(y)dy.$$

Our next goal is to prove that the QRLHM-estimate has the same asymptotic distribution.

From (5.14), the QRLHM-estimate can be written as

$$\hat{u}_i = (1-\delta)\sum_{n=0}^{t}\delta^n\tilde{\mu}_{n+1}(x_i, h, g) + o_p(\lambda).$$

for all $x_i \in [\max(2h, g) + (t+1)g, 1 - \max(2h, g) - (t+1)g]$.

We will use these two results to show that

$$\frac{\hat{u}(x_i) - \mathbb{E}\,\hat{u}(x_i)}{\sqrt{\operatorname{var}\hat{u}(x_i)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{for} \quad N \to \infty.$$

for all $x_i \in [\max(2h, g) + (t+1)g, 1 - \max(2h, g) - (t+1)g]$.

Theorem 5.20 (QRLHM Asymptotic Normality 1) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). Suppose that*

(1) $h \to 0$ as $N \to \infty$ such that $h \sim$ constant $N^{-1/5}$,

(2) $g \to 0$ as $N \to \infty$ such that $Ng^4 \to \infty$, and

(2) $\lambda \to 0$ as $N \to \infty$ such that $\lambda\sqrt{Nh} = O(1)$.

Then, with t chosen as the smallest integer satisfying

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have

$$\sqrt{Nh}\left(\frac{\hat{u}(x_i)-\mu(x_i)-\frac{1}{2}h^2\mu''(x_i)V_K-\frac{1}{2}\frac{\lambda}{\eta_c}g^2\mu''(x_i)V_L}{\sqrt{\sigma_c^2 Q\left(\frac{g}{h},\frac{\lambda}{\eta_c}\right)}}\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,1),$$

for all $x_i \in [\max(2h,g)+(t+1)g, 1-\max(2h,g)-(t+1)g]$.

Before proving the above theorem we need the following lemma.

Lemma 5.21 *Let the model (2.1) hold. Let* ρ *be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels* K *and* L *and assume (A2) and (E1). For* $N \to \infty$*, let* $h \to 0$*,* $g \to 0$ *and* $\lambda \to 0$ *such that* $h \sim$ *constant* $N^{-1/5}$ *and* $Ng^2 \to \infty$*. Then for all* $n \in \{1,...,t\}$

$$\delta\sqrt{Nh}\Big(\tilde{\mu}_{n+1}(x,h,g)-\mathbb{E}\,\tilde{\mu}_{n+1}(x,h,g)\Big)\xrightarrow{P}0,$$

for all $x \in [h+tg, 1-h-tg]$*, where* $\delta = \frac{\lambda}{\eta_c+\lambda}$.

Proof. We first prove the lemma for $n=1$. From the asymptotic normality of the LHM-estimate and Slutsky's theorem, we have the following,

$$\begin{aligned}
&\delta\sqrt{Nh}\Big(\tilde{\mu}_2(x,h,g)-\mathbb{E}\,\tilde{\mu}_2(x,h,g)\Big)\\
&=\frac{1}{N}\sum_{j=1}^{N}L_g(x-x_j)\ \delta\sqrt{Nh}\Big(\tilde{\mu}_K(x_j,h)-\mathbb{E}\,\tilde{\mu}_K(x_j,h)\Big)\\
&=\frac{1}{N}\sum_{j=1}^{N}L_g(x-x_j)\ o_p(1)=\left(1+O\left(\frac{1}{Ng^2}\right)\right)\cdot o_p(1)\\
&=o_p(1).
\end{aligned}$$

Now recursively for any $n \in \{2,...,t\}$ we have

$$\begin{aligned}
&\delta\sqrt{Nh}\Big(\tilde{\mu}_{n+1}(x,h,g)-\mathbb{E}\,\tilde{\mu}_{n+1}(x,h,g)\Big)\\
&=\frac{1}{N}\sum_{j=1}^{N}L_g(x-x_j)\ \delta\sqrt{Nh}\Big(\tilde{\mu}_n(x_j,h,g)-\mathbb{E}\,\tilde{\mu}_n(x_j,h,g)\Big)\\
&=\frac{1}{N}\sum_{j=1}^{N}L_g(x-x_j)\ o_p(1)=\left(1+O\left(\frac{1}{Ng^2}\right)\right)\cdot o_p(1)\\
&=o_p(1).
\end{aligned}$$

□

Proof of Theorem 5.20. The proof is based on the asymptotic normality of the LHM-estimate (Theorem 2.24), on (5.14) and on Slutsky's theorem.

From (5.14) and the assumption $\lambda\sqrt{Nh} = O(1)$

$$\hat{u}_i = (1-\delta)\sum_{n=0}^{t}\delta^n \tilde{\mu}_{n+1}(x_i,h,g) + o_p\left(\frac{1}{\sqrt{Nh}}\right).$$

Now using the previous lemma and Slutsky's theorem

$$\begin{aligned}
\sqrt{Nh}\left(\frac{\hat{u}_i - \mathbb{E}\,\hat{u}_i}{\sqrt{\sigma_c^2 Q_K}}\right) &= (1-\delta)\sum_{n=0}^{t}\delta^n\sqrt{Nh}\left(\frac{\tilde{\mu}_{n+1}(x_i,h,g) - \mathbb{E}\,\tilde{\mu}_{n+1}(x_i,h,g)}{\sqrt{\sigma_c^2 Q_K}}\right) + o_p(1)\\
&= (1-\delta)\sqrt{Nh}\left(\frac{\tilde{\mu}_K(x_i,h) - \mathbb{E}\,\tilde{\mu}_K(x_i,h)}{\sqrt{\sigma_c^2 Q_K}}\right)\\
&\quad + (1-\delta)\sum_{n=1}^{t}\delta^{n-1}o_p(1) + o_p(1)\\
&\xrightarrow{\mathcal{L}} \mathcal{N}(0,1).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sqrt{Nh}\left(\frac{\hat{u}(x_i) - \mu(x_i) - \frac{1}{2}h^2\mu''(x_i)V_K - \frac{1}{2}\frac{\lambda}{\eta_c}g^2\mu''(x_i)V_L}{\sqrt{\sigma_c^2 Q\left(\frac{g}{h},\frac{\lambda}{\eta_c}\right)}}\right)\\
&= \sqrt{Nh}\left(\frac{\hat{u}_i - \mathbb{E}\,\hat{u}_i + o(h^2) + o(\lambda)}{\sqrt{\sigma_c^2 Q_K}}\right)\cdot\frac{\sqrt{Q_K}}{\sqrt{Q\left(\frac{g}{h},\frac{\lambda}{\eta_c}\right)}}\\
&\xrightarrow{\mathcal{L}} \mathcal{N}(0,1),
\end{aligned}$$

since

$$\frac{\sqrt{Q_K}}{\sqrt{Q\left(\frac{g}{h},\frac{\lambda}{\eta_c}\right)}} \to 1.$$

□

5.8 The Optimal Choice of the Parameters

We present now some remarks regarding the optimal choice of h, g, λ, and t. These remarks are derived from the asymptotic normality result of the pointwise QRLHM-estimate.

5.8.1 The Optimal Choice of h

The choice of the local optimal h has already been discussed in Section 2.7, it was

$$\boxed{h_{opt}(x) \sim \text{constant } N^{-1/5}.}$$

5.8.2 The Optimal Choice of λ

For proving asymptotic normality we have assumed that

$$\lambda = O\left(\frac{1}{\sqrt{Nh}}\right).$$

This assumption is also necessary to have the term

$$\frac{\sigma_c^2}{Nh} Q\left(\frac{g}{h}, \frac{\lambda}{\eta_c}\right)$$

as the dominant part of the variance term.

Using this assumption and the local optimal choice of h we get

$$\boxed{\lambda_{opt}(x) \sim \text{constant } N^{-2/5}.}$$

5.8.3 The Optimal Choice of t

Throughout the work we have assumed that t should be chosen as the smallest positive integer satisfying

$$\lambda^{t+1} = O\left(\frac{1}{N}\right).$$

Using the local optimal choice of λ we get

$$\boxed{t_{opt} = 2.}$$

5.8.4 The Optimal Choice of g

Comparing the bandwidth g to the other bandwidth h we may have

$$(1)\ g = h, \quad (2)\ g < h, \quad (3)\ g > h.$$

Equivalently, we may consider the cases

$$(1)\ g \sim h, \quad (2)\ g = o(h), \quad (3)\ h = o(g).$$

From Theorem 5.19 we get

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

If we assume that $h = o(g)$ then $\widehat{L}\left(\frac{g}{h}\omega\right)$ diverges, which is undesirable. Hence, we concentrate on

$$g \sim h \quad \text{and} \quad g = o(h),$$

that is,

$$g_{opt}(x) \sim \text{constant } N^{-1/5} \quad \text{or} \quad g_{opt}(x) = o(N^{-1/5}).$$

Along with the assumption $Ng^4 \to \infty$ it is easier to take the first argument, i.e.

$$\boxed{g_{opt}(x) \sim \text{constant } N^{-1/5}.}$$

We have already seen that the contribution of the smoothness term to the bias is described by $o(\lambda)$. Similarly, we can see that the bias (which is due to the smoothness term) plays asymptotically no role in the asymptotic normality of the QRLHM-estimate if λ was chosen such that $\lambda\sqrt{Nh} = O(1)$, i.e.

$$\sqrt{Nh}\frac{1}{2}\frac{\lambda}{\eta_c}g^2\mu''(x_i)V_L \to 0, \quad \text{while} \quad \sqrt{Nh}\frac{1}{2}h^2\mu''(x_i)V_K \to \text{constant} \neq 0.$$

5.8.5 Rewriting the Asymptotic Normality Result

Using the remarks above we will now rewrite the asymptotic normality result as follows.

Theorem 5.22 (QRLHM Asymptotic Normality 2) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). Suppose that*

(1) $h \to 0$ as $N \to \infty$ such that $h \sim$ constant $N^{-1/5}$,

(2) $g \to 0$ as $N \to \infty$ such that $g \sim$ constant $N^{-1/5}$, and

(2) $\lambda \to 0$ as $N \to \infty$ such that $\lambda \sim$ constant $N^{-2/5}$.

Then, for all $x_i \in [5h, 1-5h]$ we have

$$\sqrt{Nh}\left(\frac{\hat{u}(x_i) - \mu(x_i) - \frac{1}{2}h^2\mu''(x_i)V_K}{\sqrt{\sigma_c^2 Q_K}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

5.9 Interpolating QRLHM-Estimates

So far, the QRLHM estimate was only evaluated for the grid points x_i, $i = 1, \ldots, N$. However, it would be more interesting if we could extend our estimate for any $x \in [0,1]$.

One way to do so is by interpolating the data set $\{(x_i, \hat{u}_i) : i = 1, \ldots, N\}$ in the following manner,

Piecewise constant interpolation:

$$\hat{u}(x) = \hat{u}_i \quad \text{for} \quad x_{i-1} < x \leq x_i, \quad i = 1, \ldots, N, \tag{5.20}$$

where $x_i = \frac{i}{N}$ for $i = 1, \ldots, N$ and $x_0 = 0$.

The piecewise constant interpolation is the simplest method of interpolating images. Of course there are methods which give better results but the main reason for choosing this interpolation scheme is that it makes further calculations easier.

5.10 Interpolated QRLHM-Estimates: Consistency and Asymptotic Normality

In this section we will see that under our assumptions the interpolated version of the QRLHM-estimate is consistent and asymptotically normal.

Asymptotic Normality

Using the interpolation scheme given above, for any $x \in (0, 1]$ we have

$$\hat{u}(x) = \hat{u}(x_k) \quad \text{for some } k \in \{1, \ldots, N\} \text{ such that } |x - x_k| \leq \frac{1}{N}.$$

Theorem 5.23 (QRLHM Asymptotic Normality 3) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). Suppose that*

(1) $h \to 0$ as $N \to \infty$ such that $h \sim$ constant $N^{-1/5}$,

(2) $g \to 0$ as $N \to \infty$ such that $g \sim$ constant $N^{-1/5}$, and

(3) $\lambda \to 0$ as $N \to \infty$ such that $\lambda \sim$ constant $N^{-2/5}$.

Then, for all $x \in [5h + \frac{1}{N}, 1 - 5h - \frac{1}{N}]$ we have

$$\sqrt{Nh} \left(\frac{\hat{u}(x) - \mu(x) - \frac{1}{2}h^2\mu''(x)V_K}{\sqrt{\sigma_c^2 Q_K}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Proof. Let $x \in [5h + \frac{1}{N}, 1 - 5h - \frac{1}{N}]$, then

$$\hat{u}(x) = \hat{u}(x_k) \quad \text{for some } k \in \{1, \ldots, N\} \text{ such that } |x - x_k| \leq \frac{1}{N},$$

i.e. $x_k \in [5h, 1 - 5h]$.

We decompose the quotient above as follows,

$$
\begin{aligned}
&\sqrt{Nh}\left(\frac{\hat{u}(x)-\mu(x)-\frac{1}{2}h^2\mu''(x)V_K}{\sqrt{\sigma_c^2 Q_K}}\right)\\
=&\sqrt{Nh}\left(\frac{\hat{u}(x_k)-\mu(x_k)-\frac{1}{2}h^2\mu''(x_k)V_K}{\sqrt{\sigma_c^2 Q_K}}\right)\\
&+\sqrt{Nh}\left(\frac{\mu(x_k)-\mu(x)}{\sqrt{\sigma_c^2 Q_K}}\right)\\
&+\frac{1}{2}\sqrt{Nh}\,h^2V_K\left(\frac{\mu''(x_k)-\mu''(x)}{\sqrt{\sigma_c^2 Q_K}}\right).
\end{aligned}
$$

From Theorem 5.22,

$$
\sqrt{Nh}\left(\frac{\hat{u}(x_k)-\mu(x_k)-\frac{1}{2}h^2\mu''(x_k)V_K}{\sqrt{\sigma_c^2 Q_K}}\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,1).
$$

From (A2) a)-b) we get that μ is Lipschitz continuous, hence

$$
\sqrt{Nh}\left|\frac{\mu(x_k)-\mu(x)}{\sqrt{\sigma_c^2 Q_K}}\right|\le\sqrt{Nh}\,\frac{C_\mu\,|x_k-x|}{\sqrt{\sigma_c^2 Q_K}}\le\sqrt{\frac{h}{N}}\,\frac{C_\mu}{\sqrt{\sigma_c^2 Q_K}}\overset{N\to\infty}{\longrightarrow}0.
$$

From (A2) b) (μ'' is Hölder continuous, $\beta>0$) and $h=O(N^{-1/5})$ we have

$$
\frac{1}{2}\sqrt{Nh}\,h^2V_K\left|\frac{\mu''(x_k)-\mu''(x)}{\sqrt{\sigma_c^2 Q_K}}\right|\le\frac{1}{2}\sqrt{Nh}\,h^2V_K\frac{H\,|x_k-x|^\beta}{\sqrt{\sigma_c^2 Q_K}}\le\frac{1}{2}\sqrt{Nh}\,h^2V_K\frac{HN^{-\beta}}{\sqrt{\sigma_c^2 Q_K}}\overset{N\to\infty}{\longrightarrow}0.
$$

Combining the three results above, using Slutsky's theorem completes the proof. □

Consistency

Using the previous theorem we may write

$$
\begin{aligned}
\hat{u}(x)&=\mu(x)+\frac{1}{2}h^2\mu''(x)V_K+O_p\left(\frac{1}{\sqrt{Nh}}\right)\\
&=\mu(x)+O(h^2)+O_p\left(\frac{1}{\sqrt{Nh}}\right),
\end{aligned}
$$

and hence,

$$
\hat{u}(x)\xrightarrow{P}\mu(x)
$$

for all $x\in[5h+\frac{1}{N},1-5h-\frac{1}{N}]$.

5.11 Interpolated QRLHM-Estimates: Uniform Consistency

In this section we will see that under our assumptions the interpolated version of the QRLHM-estimate is uniformly consistent.

Theorem 5.24 (QRLHM Interpolated) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have uniformly in $x \in [\max(h,g) + (t+1)g + \frac{1}{N}, 1 - \max(h,g) - (t+1)g - \frac{1}{N}]$, that:

$$\hat{u}(x) = (1-\delta)\sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x,h,g) + o_p(1) \tag{5.21}$$

where $\delta = \frac{\lambda}{\eta_c + \lambda}$.

Proof. Let $x \in [\max(h,g) + (t+1)g + \frac{1}{N}, 1 - \max(h,g) - (t+1)g - \frac{1}{N}]$, then

$$\hat{u}(x) = \hat{u}(x_k) \quad \text{for some } k \in \{1,\dots,N\} \text{ such that } |x - x_k| \le \frac{1}{N},$$

i.e. $x_k \in [\max(h,g) + (t+1)g, 1 - \max(h,g) - (t+1)g]$. Then, using (5.16) we have for all $n \in \{0,\dots,t\}$ that

$$\begin{aligned}
|\tilde{\mu}_{n+1}(x,h,g) - \tilde{\mu}_{n+1}(x_k,h,g)| \le\ & |\tilde{\mu}_{n+1}(x_k,h,g) - \mu(x_k)| \\
& + |\mu(x_k) - \mu(x)| \\
& + |\mu(x) - \tilde{\mu}_{n+1}(x,h,g)| \\
=\ & o_p(1) + o(1) + o_p(1),
\end{aligned}$$

uniformly in $x \in [\max(h,g) + (t+1)g + \frac{1}{N}, 1 - \max(h,g) - (t+1)g - \frac{1}{N}]$ and

$$\begin{aligned}
\hat{u}(x) = \hat{u}(x_k) &= (1-\delta)\sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x_k,h,g) + o_p(\lambda) \\
&= (1-\delta)\sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x,h,g) + o_p(1)
\end{aligned}$$

uniformly in x. □

Theorem 5.25 (QRLHM UC) *Let the model (2.1) hold. Let ρ be the modified Huber function given by (2.11). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have

$$\sup_{x \in \Delta} |\hat{u}(x) - \mu(x)| = o_p(1) \tag{5.22}$$

where $\Delta = [\max(h,g) + (t+1)g + \frac{1}{N}, 1 - \max(h,g) - (t+1)g - \frac{1}{N}]$.

Proof. Let $x \in \Delta$ then using Corollary 3.8

$$\begin{aligned}
\sup_x |\hat{u}(x) - \tilde{\mu}(x)| \leq & \sup_x \left| \hat{u}(x) - (1-\delta) \sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x,h,g) \right| \\
& + \sup_x \left| \tilde{\mu}(x) - (1-\delta) \sum_{n=0}^{t} \delta^n \tilde{\mu}_{n+1}(x,h,g) \right| \\
\leq & \; o_p(1) + \sup_x |\tilde{\mu}(x) - (1-\delta)\tilde{\mu}(x)| + \delta \sup_x \left| (1-\delta) \sum_{n=1}^{t} \delta^{n-1} \tilde{\mu}_{n+1}(x,h,g) \right| \\
= & \; o_p(1) + O_p(\delta) + O_p(\delta) \\
= & \; o_p(1).
\end{aligned}$$

Using Corollary 3.8 again

$$\begin{aligned}
\sup_x |\hat{u}(x) - \mu(x)| \leq & \sup_x |\hat{u}(x) - \tilde{\mu}(x)| + \sup_x |\tilde{\mu}(x) - \mu(x)| \\
= & \; o_p(1) + o_p(1) \\
= & \; o_p(1).
\end{aligned}$$

□

Chapter 6

Huber M-Estimates with Localization and Convex Regularization (CRLHM-Estimates)

In this chapter, we again consider M-smoothers with regularization. We go through the results we obtained in Chapter 5 modifying the smoothness loss function (i.e. the tonal weight function of the smoothness term) in (5.1) from a quadratic function to a convex function (which is not necessarily quadratic). Finally, we will see some examples from the convex regularization class of functions, including nonlinear regularization, and total variation regularization.

6.1 General Setup

We recall that our data (x_j, f_j), $j = 1, \ldots, N$, come from the nonparametric regression model:

$$f_j = \mu(x_j) + \varepsilon_j, \quad j = 1, \ldots, N,$$

where $\varepsilon_j \sim$ i. i. d. $(0, \sigma^2)$, and $x_j = \frac{j}{N}$ from an equidistant grid in the unit interval $[0, 1]$.

Using the general approach proposed by Mrázek et al. [25], we now consider the problem:

$$\begin{aligned} Q^S(u_1, \ldots, u_N) = & \sum_{i,j=1}^{N} \rho_D(u_i - f_j) K_h(x_i - x_j) \\ & + \frac{\lambda}{2} \sum_{i,j=1}^{N} \rho_S(u_i - u_j) L_g(x_i - x_j) = \min_{u_1, \ldots, u_N}! \end{aligned} \tag{6.1}$$

where the kernels K and L are nonnegative, symmetric functions on $\mathbb{R}$, the bandwidths $h, g > 0$, and the regularization parameter $\lambda \geq 0$.

Here, we use the modified Huber function (cf. Chapter 5) as ρ_D, i.e.

$$\rho_D(u) = \rho(u) := \begin{cases} \begin{cases} \frac{1}{2}u^2, & |u| \le c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} & c \ge 1, \\ \\ \begin{cases} \frac{1}{2c}u^2, & |u| \le c, \\ |u| - \frac{1}{2}c, & |u| > c, \end{cases} & c \le 1, \end{cases}$$

and we assume that the smoothness loss function ρ_S satisfies the following assumptions

$$(S1) \begin{cases} \text{a) } \rho_S \text{ is symmetric around zero.} \\ \text{b) } \rho_S \text{ is convex.} \\ \text{c) } \rho_S' \text{ is Lipschitz continuous with Lipschitz constant } C_{\rho_S'}. \\ \text{d) } \rho_S(0) = 0,\ \rho_S'(0) = 0 \text{ and } \rho_S''(0) = \text{constant} \neq 0. \end{cases}$$

Remark 6.1 *Mark that, since ρ_S is convex and continuous then its second derivative exists almost everywhere and is nonnegative (see Udrişte [35], page 73).*

Mark also that, since ρ_S' is Lipschitz continuous then it is almost everywhere differentiable and the derivative is bounded (where it exists).

The solution $\hat{\mathbf{u}}^S = (\hat{u}_1^S, \ldots, \hat{u}_N^S)^T$ of problem (6.1) estimates $\boldsymbol{\mu} = (\mu(x_1), \ldots, \mu(x_N))^T$ and is called the *convex regularized local Huber M-estimate*, abbreviated as the CRLHM-estimate. Special cases of the CRLHM-estimate are the LHM-estimate and QRLHM-estimate.

Indeed, all results in Chapter 5 that do not contain any treatment of the function ρ_S are still valid in the general context. Our goal in this chapter is to rewrite the results that were proven in the previous chapter (for $\rho_S(u) = u^2/2$) for a general ρ_S satisfying assumptions (S1).

6.2 Notation and Auxiliary Results

In this section we will rewrite some results regarding the gradient vector and the inverse of the Hessian matrix of the energy functional Q^S (which is Q using ρ_S as the tonal weight function).

The gradient vector and its Jacobian are written as

$$\nabla Q^S(\mathbf{u}) = (Q_1^S(\mathbf{u}), \ldots, Q_N^S(\mathbf{u}))^T \quad \text{and} \quad \nabla^2 Q^S(\mathbf{u}) = \left(Q_{i,\ell}^S(\mathbf{u})\right)_{i,\ell},$$

where

$$Q_i^S(\mathbf{u}) := \frac{\partial Q^S(\mathbf{u})}{\partial u_i} = \sum_{j=1}^{N} \rho_D'(u_i - f_j) K_h(x_i - x_j) + \lambda \sum_{j=1}^{N} \rho_S'(u_i - u_j) L_g(x_i - x_j),$$

$$Q_{ii}^S(\mathbf{u}) := \frac{\partial^2 Q(\mathbf{u})}{\partial u_i^2} = \sum_{j=1}^{N} \rho_D''(u_i - f_j) K_h(x_i - x_j) + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} \rho_S''(u_i - u_j) L_g(x_i - x_j),$$

$$Q_{\ell i}^S(\mathbf{u}) := \frac{\partial^2 Q(\mathbf{u})}{\partial u_\ell \partial u_i} = -\lambda \rho_S''(u_\ell - u_i) L_g(x_i - x_\ell), \quad \ell \neq i.$$

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

As in Chapter 5 we define

$$G^S := \frac{\nabla Q^S(\tilde{\boldsymbol{\mu}})}{N} \qquad \text{and} \qquad J^S := \frac{\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})}{N} = \tilde{P}^S - \lambda \Lambda^S \tag{6.2}$$

where for $i, \ell = 1, \ldots, N$

$$\tilde{P}_{i,\ell}^S = \begin{cases} \frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j), & i = \ell, \\ 0, & i \neq \ell, \end{cases} \tag{6.3}$$

and (using the symmetry of ρ_S'')

$$\Lambda_{i,\ell}^S = \frac{1}{N} \rho_S''(\tilde{\mu}_i - \tilde{\mu}_\ell) L_g(x_i - x_\ell). \tag{6.4}$$

Mark that using a quadratic smoothness loss function, i.e. $\rho_S(u) = \frac{1}{2}u^2$, the matrices $\tilde{P}^S$ and Λ^S coincide with the matrices $\tilde{P}$ and Λ of Section 5.3.

Throughout the text we will use the following abbreviations,

$$Q_i^S = Q_i^S(\tilde{\boldsymbol{\mu}}) = \left. \frac{\partial Q^S(\mathbf{u})}{\partial u_i} \right|_{\mathbf{u} = \tilde{\boldsymbol{\mu}}} \qquad \text{and} \qquad Q_{i,\ell}^S = Q_{i,\ell}^S(\tilde{\boldsymbol{\mu}}) = \left. \frac{\partial^2 Q^S(\mathbf{u})}{\partial u_\ell \partial u_i} \right|_{\mathbf{u} = \tilde{\boldsymbol{\mu}}}.$$

Before evaluating the norms of the gradient vector and the inverse of the Hessian matrix we present some useful lemmas. These lemmas will help us in interchanging expressions involving the general ρ_S with the already treated expressions (i.e. involving the quadratic ρ_S).

Lemma 6.2 *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11) and let ρ_S satisfy (S1). Let K and L satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$. Then,*

$$\frac{\lambda}{N} \sum_{j=1}^N \rho_S'(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) = \rho_S''(0) \frac{\lambda}{N} \sum_{j=1}^N (\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) + o_p(\lambda)$$

for all $x_i \in [\max(h, g), 1 - \max(h, g)]$, where $\lambda_S = \lambda \rho_S''(0)$.

Proof. From Theorem 2.12 we get that for all $x_i \in [h, 1-h]$

$$\tilde{\mu}_K(x_i, h) - \mu(x_i) = o_p(1). \tag{6.5}$$

That is, for all $x_i, x_j \in [h, 1-h]$

$$\tilde{\mu}_i - \tilde{\mu}_j = \mu_i - \mu_j + o_p(1), \tag{6.6}$$

and if $|x_i - x_j| \le g$ then assumptions (A2) imply that

$$\tilde{\mu}_i - \tilde{\mu}_j = O(g) + o_p(1) = o_p(1) \tag{6.7}$$

as $N \to \infty$, $h \to 0$, $g \to 0$ such that $Nh^2 \to \infty$.

Since ρ'_S is differentiable almost everywhere and $\rho''_S(0)$ exists, we expand $\rho'_S(\tilde{\mu}_i - \tilde{\mu}_j)$ around zero, then for all $x_i, x_j \in [h, 1-h]$ such that $|x_i - x_j| \le g$,

$$\begin{aligned}\rho'_S(\tilde{\mu}_i - \tilde{\mu}_j) &= \rho'_S(0) + (\tilde{\mu}_i - \tilde{\mu}_j)\rho''_S(0) + o_{a.s.}\left(\tilde{\mu}_i - \tilde{\mu}_j\right) \\ &= (\tilde{\mu}_i - \tilde{\mu}_j)\rho''_S(0) + o_p\left(1\right).\end{aligned}$$

Therefore, using Lemma 2.3 for the kernel L, we have

$$\begin{aligned}&\frac{\lambda}{N}\sum_{j=1}^{N}\rho'_S(\tilde{\mu}_i - \tilde{\mu}_j)L_g(x_i - x_j) \\ &= \rho''_S(0)\frac{\lambda}{N}\sum_{j=1}^{N}(\tilde{\mu}_i - \tilde{\mu}_j)L_g(x_i - x_j) + o_p(1)\frac{\lambda}{N}\sum_{j=1}^{N}L_g(x_i - x_j) \\ &= \rho''_S(0)\frac{\lambda}{N}\sum_{j=1}^{N}(\tilde{\mu}_i - \tilde{\mu}_j)L_g(x_i - x_j) + o_p(\lambda)\end{aligned}$$

as $N \to \infty$, $h \to 0$, $g \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$.

□

Proposition 6.3 (Norm of G^S) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11) and let ρ_S satisfy (S1). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$ and $g \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then,*

$$\|G^S\| = o_p(\lambda).$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$.

Proof. From the Lipschitz continuity of ρ'_S, and using the norm defined in (4.5) we have

$$\begin{aligned}\|G^S\| = \|G^{S^*}\|_{\ell^\infty(\mathbb{R})} = \sup_{i\in\mathbb{N}}|{G^{S^*}}_i| &= \sup_{1\le i\le N}\left|G_i^S\right| = \sup_{1\le i\le N}\left|\frac{Q_i^S}{N}\right| \\ &= \sup_{1\le i\le N}\left|\frac{\lambda}{N}\sum_{j=1}^{N}\rho'_S(\tilde{\mu}_i - \tilde{\mu}_j)L_g(x_i - x_j)\right| \\ &\le C_{\rho'_S}\sup_{1\le i\le N}\left|\frac{\lambda}{N}\sum_{j=1}^{N}|\tilde{\mu}_i - \tilde{\mu}_j|L_g(x_i - x_j)\right| \\ &\quad \text{(using Proposition 5.3)} \\ &= o_p(\lambda)\end{aligned}$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$. □

We now present some results regarding the norms of the matrices $(\tilde{P}^S)^{-1}$, $(\tilde{P}^S)^{-1}\Lambda^S$ and $(J^S)^{-1}$. These results will be helpful tools for getting the CRLHM-estimate.

Proposition 6.4 (Norm of $(\tilde{P}^S)^{-1}$) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11) and let ρ_S satisfy (S1). Let K and L satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). For $N \to \infty$, let $h \to 0$ such that $Nh^2 \to \infty$. Then,*

$$\|(\tilde{P}^S)^{-1}\| = O_{a.s.}(1)$$

for $x_1, \dots, x_N \in [h, 1-h]$, $g \geq 0$, and $\lambda \geq 0$.

Proof. Since L is nonnegative and ρ_S'' is nonnegative almost everywhere

$$\begin{aligned}
\|(\tilde{P}^S)^{-1}\| &= \sup_{1\leq i\leq N} \left| \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)} \right| \\
&\overset{a.s.}{\leq} \sup_{1\leq i\leq N} \left| \frac{1}{\frac{1}{N}\sum_{j=1}^{N} \rho''(\tilde{\mu}_i - f_j) K_h(x_i - x_j)} \right| \\
&\quad \text{(Using Lemma 5.6)} \\
&\overset{a.s.}{\leq} \frac{2}{\eta_c}
\end{aligned}$$

for $x_1, \dots, x_N \in [h, 1-h]$. □

Proposition 6.5 (Norm of $(\tilde{P}^S)^{-1}\Lambda^S$) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11) and let ρ_S satisfy (S1). Let K and L satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$. Then,*

$$\|(\tilde{P}^S)^{-1}\Lambda^S\| = O_{a.s.}(1), \quad \|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| = o_{a.s.}(1), \quad \textit{and} \quad \|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| < 1 \;\; a.s.$$

for $x_1, \dots, x_N \in [\max(h, g), 1 - \max(h, g)]$.

Proof. Since L is nonnegative and ρ_S'' is nonnegative almost everywhere, and using Lemma 5.6 we get that

$$\begin{aligned}
\|(\tilde{P}^S)^{-1}\Lambda^S\| &= \sup_{1\leq i\leq N} \sum_{j=1}^{N} \frac{|\Lambda^S_{i,j}|}{|\tilde{P}^S_{i,i}|} \\
&= \sup_{1\leq i\leq N} \frac{\frac{1}{N}\sum_{j=1}^{N} \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)}{\left|\frac{1}{N}\sum_{j=1}^{N} \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)\right|}
\end{aligned}$$

$$\overset{a.s.}{\leq} \frac{2}{\eta_c} \left(\sup_u \rho_S''(u) \right) \cdot \left(\sup_{1 \leq i \leq N} \frac{1}{N} \sum_{j=1}^N L_g(x_i - x_j) \right)$$

(Using Remark 6.1 and Lemma 2.3)

$$= O_{a.s.}(1)$$

for $x_1, \ldots, x_N \in [\max(h, g), 1 - \max(h, g)]$.

Therefore,

$$\|(\tilde{P}^S)^{-1}\Lambda^S\| = O_{a.s.}(1) \quad \text{and} \quad \|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| \xrightarrow{a.s.} 0.$$

Since $\rho_D''(u) \geq 0$ for all $u \in \mathbb{R} \setminus \{-c, c\}$, we have for all $i = 1, \ldots, N$,

$$\frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) \geq 0$$

almost surely with respect to the probability measure of ε_j.

That is, for all i we have almost surely that

$$\frac{\frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)}{\frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)} \leq 1.$$

Therefore, $\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| \leq 1$ almost surely. What is left to be shown is that $\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| \neq 1$ almost surely.

Consider,

$$\begin{aligned}
&\mathbb{P}\left(\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\| = 1 \right) \\
&= \mathbb{P}\left(\sup_{1 \leq i \leq N} \frac{\frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)}{\frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j)} = 1 \right) \\
&= \mathbb{P}\left(\frac{\frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_k - \tilde{\mu}_j) L_g(x_k - x_j)}{\frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_k - f_j) K_h(x_k - x_j) + \frac{\lambda}{N} \sum_{j=1}^N \rho_S''(\tilde{\mu}_k - \tilde{\mu}_j) L_g(x_k - x_j)} = 1, \quad \text{for some } k \right) \\
&= \mathbb{P}\left(\frac{1}{N} \sum_{j=1}^N \rho_D''(\tilde{\mu}_k - f_j) K_h(x_k - x_j) = 0, \quad \text{for some } k \right) \\
&= 0,
\end{aligned}$$

since

$$\frac{1}{N}\sum_{j=1}^{N}\rho_D''(\tilde{\mu}_k - f_j)K_h(x_k - x_j) \geq \frac{1}{2}\eta_c \; a.s.$$

uniformly in $x_k \in [h, 1-h]$. □

Proposition 6.6 (Norm of $(J^S)^{-1}$) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11) and let ρ_S satisfy (S1). Let K and L satisfy (A1) a)-c). Let μ satisfy (A2). Let ε_j satisfy (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$. Then*

$$\|(J^S)^{-1}\| = O_{a.s.}(1)$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$.

Proof. Using Neumann series and Propositions 6.4 and 6.5 we get

$$\begin{aligned}
\|(J^S)^{-1}\| &= \|(\tilde{P}^S - \lambda\Lambda^S)^{-1}\| \leq \|(I_N - \lambda(\tilde{P}^S)^{-1}\Lambda^S)^{-1}\| \cdot \|(\tilde{P}^S)^{-1}\| \\
&\overset{a.s.}{\leq} \frac{1}{1 - \|\lambda(\tilde{P}^S)^{-1}\Lambda^S\|} \cdot \|(\tilde{P}^S)^{-1}\| \\
&= \frac{1}{1 + o_{a.s.}(1)} \cdot O_{a.s.}(1) = O_{a.s.}(1).
\end{aligned}$$

□

6.3 Vector and Component Form of CRLHM-Estimates

The following theorem is very similar to Theorem 5.10.

Theorem 6.7 (CRLHM Vector) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11). Let ρ_S satisfy (S1). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1) a)-b). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $Nh^2 \to \infty$ and $Ng^2 \to \infty$. If $\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})$ is invertible, then the solution of problem (6.1) is given by,*

$$\hat{\mathbf{u}}^S = \tilde{\boldsymbol{\mu}} - [\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})]^{-1}\nabla Q^S(\tilde{\boldsymbol{\mu}}) + o_{a.s.}\left(\frac{1}{N}\right) \tag{6.8}$$

for $x_1, \ldots, x_N \in [\max(h,g), 1-\max(h,g)]$.

Proof. Using a Taylor series expansion of the gradient vector ∇Q^S around $\tilde{\boldsymbol{\mu}}$ or equivalently using the definition of the derivative for ∇Q^S as a function from $\mathbb{R}^N$ to $\mathbb{R}^N$, we have

$$\nabla Q^S(\hat{\mathbf{u}}) = \nabla Q^S(\tilde{\boldsymbol{\mu}}) + \nabla^2 Q^S(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}) + o\left(\|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}\|\right),$$

that is,

$$\frac{\|\nabla Q^S(\hat{\mathbf{u}}) - \nabla Q^S(\tilde{\boldsymbol{\mu}}) - \nabla^2 Q^S(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}})\|}{\|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}\|} \to 0 \quad \text{as} \quad \hat{\mathbf{u}}^S \to \tilde{\boldsymbol{\mu}}.$$

Since $\hat{\mathbf{u}}^S$ is the solution of problem (6.1) then $\nabla Q^S(\hat{\mathbf{u}}^S) = \mathbf{0}_{N\times 1}$.

Letting $\lambda \to 0$ implies that

$$\|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}\| \to 0.$$

Using Proposition 6.6 we have

$$\begin{aligned}
N\left\|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}} + (J^S)^{-1}G^S\right\| &= N\left\|(J^S)^{-1}\{J^S(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}) + G^S\}\right\| \\
&\leq N\|(J^S)^{-1}\| \cdot \|J^S(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}) + G^S\| \\
&= \|(J^S)^{-1}\| \cdot \|\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}) + \nabla Q^S(\tilde{\boldsymbol{\mu}})\| \\
&= \|(J^S)^{-1}\| \cdot \frac{\|\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})(\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}) + \nabla Q^S(\tilde{\boldsymbol{\mu}})\|}{\|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}\|} \cdot \|\hat{\mathbf{u}}^S - \tilde{\boldsymbol{\mu}}\| \\
&= O_{a.s.}(1) \cdot o(1) \cdot o(1) = o_{a.s.}(1),
\end{aligned}$$

that is,

$$\hat{\mathbf{u}}^S = \tilde{\boldsymbol{\mu}} - [\nabla^2 Q^S(\tilde{\boldsymbol{\mu}})]^{-1}\nabla Q^S(\tilde{\boldsymbol{\mu}}) + o_{a.s.}\left(\frac{1}{N}\right).$$

□

The following theorem is very similar to Theorem 5.11.

Theorem 6.8 (CRLHM Component 1) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11). Let ρ_S satisfy (S1). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right), \tag{6.9}$$

we have uniformly in $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$, that:

$$\hat{u}_i^S = \tilde{\mu}_i - \frac{1}{\eta_c + \lambda_S}\frac{Q_i^S}{N} - \frac{1}{\eta_c + \lambda_S}\sum_{n=1}^{t}\left(\frac{\lambda}{\eta_c + \lambda_S}\right)^n \frac{1}{N^n}\sum_{j_1,\dots,j_n} L_g(x_i - x_{j_1})\cdots$$
$$\cdots L_g(x_{j_{n-1}} - x_{j_n})\frac{Q_{j_n}^S}{N} + R_{N,i}$$

where $\tilde{\mu}_i = \tilde{\mu}_K(x_i, h)$, $Q_i^S = \left.\frac{\partial Q^S(\mathbf{u})}{\partial u_i}\right|_{\mathbf{u}=\tilde{\boldsymbol{\mu}}}$, $\lambda_S = \lambda\rho_S''(0)$, and $R_{N,i} = o_p(\lambda)$.

Proof. Using Neumann series and Propositions 6.4 and 6.5 we get

$$
\begin{aligned}
&\frac{1}{\lambda^{t+1}}\left\|(J^S)^{-1}-\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}\right\| \\
&= \frac{1}{\lambda^{t+1}}\left\|(I-\lambda(\tilde{P}^S)^{-1}\Lambda^S)^{-1}(\tilde{P}^S)^{-1}-\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}\right\| \\
&\overset{a.s.}{\leq} \frac{1}{\lambda^{t+1}}\left\|\sum_{j=0}^{\infty}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j-\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j\right\|\cdot\left\|(\tilde{P}^S)^{-1}\right\| \\
&\leq \frac{1}{\lambda^{t+1}}\ \|(\tilde{P}^S)^{-1}\|\sum_{j=t+1}^{\infty}\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\|^j \\
&= \frac{1}{\lambda^{t+1}}\ \|(\tilde{P}^S)^{-1}\|\cdot\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\|^{t+1}\sum_{j=0}^{\infty}\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\|^j \\
&= \|(\tilde{P}^S)^{-1}\|\cdot\|(\tilde{P}^S)^{-1}\Lambda^S\|^{t+1}\cdot\frac{1}{1-\|\lambda(\tilde{P}^S)^{-1}\Lambda^S\|} \\
&= O_{a.s.}(1)\cdot O_{a.s.}(1)\cdot\frac{1}{1+o_{a.s.}(1)} \\
&= O_{a.s.}(1),
\end{aligned}
$$

that is,

$$(J^S)^{-1}=\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}+O_{a.s.}(\lambda^{t+1}).$$

Using the previous expansion of $(J^S)^{-1}$ and Proposition 6.3 we get from (6.8) that

$$
\begin{aligned}
&N\left\|\hat{\mathbf{u}}^S-\tilde{\boldsymbol{\mu}}+\left(\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}\right)\cdot G^S\right\| \\
&\leq N\left\|\hat{\mathbf{u}}^S-\tilde{\boldsymbol{\mu}}+(J^S)^{-1}\cdot G^S\right\|+N\left\|\left(\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}\right)\cdot G^S-(J^S)^{-1}\cdot G^S\right\| \\
&\leq N\left\|\hat{\mathbf{u}}^S-\tilde{\boldsymbol{\mu}}+(J^S)^{-1}\cdot G^S\right\|+N\left\|\sum_{j=0}^{t}(\lambda(\tilde{P}^S)^{-1}\Lambda^S)^j(\tilde{P}^S)^{-1}-(J^S)^{-1}\right\|\cdot\|G^S\| \\
&= o_{a.s.}(1)+N\cdot O_{a.s.}(\lambda^{t+1})\cdot o_{a.s.}(\lambda) \\
&= o_{a.s.}(1)
\end{aligned}
$$

provided that $N\lambda^{t+1}=O(1)$.

Therefore,

$$\hat{\mathbf{u}}^S = \tilde{\boldsymbol{\mu}} - \left(\sum_{j=0}^{t} (\lambda (\tilde{P}^S)^{-1} \Lambda^S)^j (\tilde{P}^S)^{-1}\right) \frac{\nabla Q^S(\tilde{\boldsymbol{\mu}})}{N} + o_{a.s.}\left(\frac{1}{N}\right). \tag{6.10}$$

After calculating the matrices, the componentwise version is,

$$\begin{aligned} \hat{u}_i^S = \tilde{\mu}_i - \frac{1}{\tilde{P}_{ii}^S}\frac{Q_i^S}{N} - \frac{1}{\tilde{P}_{ii}^S}\sum_{n=1}^{t}\frac{\lambda^n}{N^n}\sum_{j_1,\dots,j_n}\frac{L_g(x_i - x_{j_1})}{\tilde{P}_{j_1 j_1}^S}\cdots \\ \cdots\frac{L_g(x_{j_{n-1}} - x_{j_n})}{\tilde{P}_{j_n j_n}^S}\frac{Q_{j_n}^S}{N} + o_p\left(\frac{1}{N}\right). \end{aligned} \tag{6.11}$$

uniformly in $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$.

From Propositions 6.3 and 6.4 we have

$$\left|\frac{1}{\tilde{P}_{jj}^S}\frac{Q_j^S}{N} - \frac{1}{\eta_c + \lambda_S}\frac{Q_j^S}{N}\right| = \left|\frac{1}{\tilde{P}_{jj}^S} - \frac{1}{\eta_c + \lambda_S}\right| \cdot \left|\frac{Q_j^S}{N}\right| = O_{a.s.}(1)\cdot o_p(\lambda) = o_p(\lambda), \tag{6.12}$$

uniformly in $x_j \in [\max(h,g), 1 - \max(h,g)]$.

Finally, we use (6.12) to interchange $\frac{1}{\tilde{P}_{\bullet\bullet}^S}\frac{Q_\bullet^S}{N}$ with $\frac{1}{\eta_c+\lambda_S}\frac{Q_\bullet^S}{N}$ in (6.11). □

As in Chapter 5 we will use the LHM-iterated smoothers to write the CRLHM-estimate in a compact form as in Theorem 5.15.

Theorem 6.9 (CRLHM Component 2) *Let the model (2.1) hold. Let ρ_D be the modified Huber function given by (2.11). Let ρ_S satisfy (S1). Assume (A1) a)-c) for the kernels K and L and assume (A2) and (E1). For $N \to \infty$, let $h \to 0$, $g \to 0$ and $\lambda \to 0$ such that $h \sim$ constant $N^{-1/5}$ and $Ng^2 \to \infty$. Then, with t chosen as the smallest integer satisfying*

$$\lambda^{t+1} = O\left(\frac{1}{N}\right),$$

we have for all $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$, that:

$$\hat{u}_i^S = \tilde{\mu}_i + \sum_{n=0}^{t}\frac{\delta_S^{n+1}}{(\rho_S''(0))^n}\tilde{\nu}_{n+2}(x_i,h,g) + R_{N,i} \tag{6.13}$$

where $\delta_S = \frac{\lambda_S}{\eta_c+\lambda_S}$ and $R_{N,i} = o_p(\lambda)$. Moreover, if we standardize ρ_S such that $\rho_S''(0) = 1$, then

$$\hat{u}_i^S = (1-\delta)\sum_{n=0}^{t}\delta^n\tilde{\mu}_{n+1}(x_i,h,g) + R_{N,i} \tag{6.14}$$

for all $x_i \in [\max(h,g) + (t+1)g, 1 - \max(h,g) - (t+1)g]$.

In view of Theorem 5.15 and under the above assumptions we have

$$\hat{u}_i^S = \hat{u}_i + o_p(\lambda) \tag{6.15}$$

for all $x_i \in [\max(h,g) + tg, 1 - \max(h,g) - tg]$.

Proof. From Lemmas 2.3 and 6.2 we can rewrite Q_i^S/N as follows

$$\begin{aligned}
\frac{Q_i^S}{N} &= \frac{1}{N}\sum_{j=1}^{N} \rho_D'(\tilde{\mu}_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{N}\sum_{j=1}^{N} \rho_S'(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) \\
&= \rho_S''(0)\frac{\lambda}{N}\sum_{j=1}^{N}(\tilde{\mu}_i - \tilde{\mu}_j) L_g(x_i - x_j) + o_p(\lambda) \\
&= \frac{\lambda_S}{N}\sum_{j=1}^{N} L_g(x_i - x_j)\tilde{\mu}_i - \frac{\lambda_S}{N}\sum_{j=1}^{N} L_g(x_i - x_j)\tilde{\mu}_j + o_p(\lambda) \\
&= -\lambda_S \left\{\tilde{\mu}_2(x_i,h,g) - \tilde{\mu}_K(x_i,h)\right\} + O\left(\frac{\lambda}{Ng^2}\right) + o_p(\lambda) \\
&= -\lambda_S \tilde{\nu}_2(x_i,h,g) + o_p(\lambda)
\end{aligned}$$

for all $x_i \in [\max(h,g), 1 - \max(h,g)]$.

Interchanging $Q_\bullet^S/N$ with $\lambda_S \hat{\nu}(x_\bullet, h, g)$ in the $\hat{u}_i$ obtained in Theorem 6.8 we have,

$$\begin{aligned}
\hat{u}_i^S = \tilde{\mu}_i &+ \frac{\lambda_S}{\eta_c + \lambda_S}\tilde{\nu}_2(x_i,h,g) \\
&+ \frac{\lambda_S}{\eta_c + \lambda_S}\sum_{n=1}^{t}\left(\frac{\lambda}{\eta_c + \lambda_S}\right)^n \frac{1}{N^n}\sum_{j_1,\dots,j_n} L_g(x_i - x_{j_1})\cdots \\
&\qquad \cdots L_g(x_{j_{n-1}} - x_{j_n})(\tilde{\mu}_2(x_{j_n},h,g) - \tilde{\mu}_1(x_{j_n},h,g)) \\
&+ o_p(\lambda) \\
= \tilde{\mu}_i &+ \frac{\lambda_S}{\eta_c + \lambda_S}\tilde{\nu}_2(x_i,h,g) \\
&+ \frac{\lambda_S}{\eta_c + \lambda_S}\sum_{n=1}^{t}\left(\frac{\lambda}{\eta_c + \lambda_S}\right)^n (\tilde{\mu}_{n+2}(x_i,h,g) - \tilde{\mu}_{n+1}(x_i,h,g)) \\
&+ o_p(\lambda) \\
= \tilde{\mu}_i &+ \delta_S \tilde{\nu}_2(x_i,h,g) + \sum_{n=1}^{t}\frac{\delta_S^{n+1}}{(\rho_S''(0))^n}\tilde{\nu}_{n+2}(x_i,h,g) + o_p(\lambda),
\end{aligned}$$

and hence (6.13). Moreover, if we have $\rho_S''(0) = 1$ then

$$\hat{u}_i^S = \tilde{\mu}_i + \sum_{n=0}^{t} \delta^{n+1} \tilde{\nu}_{n+2}(x_i, h, g) + R_{N,i}.$$

This expression is exactly the same as the expression obtained for the QRLHM-estimate in (5.13). Therefore, (6.14) follows automatically and

$$\hat{u}_i^S = \hat{u}_i + o_p(\lambda)$$

for all $x_i \in [\max(h, g), 1 - \max(h, g)]$. □

6.4 Conclusion

Important consequences of this result are the L_2 and L_1 limiting cases.

Corollary 6.10 *Let the assumptions of Theorem 6.9 hold and let $\rho_S''(0) = 1$.*

a) **The L_2 limiting case:** *If $c \to \infty$ then*

$$\hat{u}_i^{CRLLS} = \lim_{c \to \infty} \hat{u}_i^S = (1 - \theta) \sum_{n=0}^{t} \theta^n \tilde{\mu}_{NW,n+1}(x_i, h, g) + o_p(\lambda),$$

for all $x_i \in [\max(h, g) + (t + 1)g, 1 - \max(h, g) - (t + 1)g]$ where $\theta = \frac{\lambda}{1+\lambda}$.

b) **The L_1 limiting case:** *If $c \to 0$ such that $p_\varepsilon(0) \neq 0$ then*

$$\hat{u}_i^{CRLLA} = \lim_{c \to 0} \hat{u}_i^S = (1 - \zeta) \sum_{n=0}^{t} \zeta^n \tilde{\mu}_{LAD,n+1}(x_i, h, g) + o_p(\lambda),$$

for all $x_i \in [\max(h, g) + (t + 1)g, 1 - \max(h, g) - (t + 1)g]$ where $\zeta = \frac{\lambda}{2p_\varepsilon(0)+\lambda}$.

Remark 6.11 *From Theorem 6.9 we see that using any function ρ_S that satisfies (S1) will give an estimate that is equivalent to the estimate obtained using the quadratic ρ_S.*

In particular (if we standardize $\rho_S''(0) = 1$)

$$\hat{u}_i^{CRLHM} = \hat{u}_i^{QRLHM} + o_p(\lambda).$$

Assumptions (S1) define a class of functions that both the quadratic function and the Huber function belong to. In other words, it make no difference (asymptotically) if we are using quadratic regularization or Huber M-type regularization.

6.5 Examples

In practice and in addition to the data smoothing criterion, there are a number of functions which are usually used as tonal weight function. These functions define different regularization methods. Examples are: quadratic regularization, LAD regularization, nonlinear diffusion regularization (using the Perona-Malik function), nonlinear regularization, and total variation regularization.

In this section, we are interested in checking if these functions belong to the class of functions defined by assumptions (S1).

We define the class of function $\mathcal{G}^{(S1)}$ as follows,

$$\rho_S \in \mathcal{G}^{(S1)} \quad \Longleftrightarrow \quad \rho_S \text{ satisfies (S1).} \tag{6.16}$$

The functions used in the examples below are taken from Table 1 in [5].

6.5.1 Quadratic Regularization(Q-R)

It can be easily checked that $\rho_S(z) = z^2/2$ belong to $\mathcal{G}^{(S1)}$.

6.5.2 Least-Absolute Deviation Regularization (LAD-R)

To get LAD regularization, we use the modified Huber functions as the tonal weight function in the smoothness term and and let c go to zero. It can be easily shown that the modified Huber function belong to $\mathcal{G}^{(S1)}$.

6.5.3 Nonlinear Diffusion Regularization (ND-R)

The term nonlinear diffusion regularization stands for using the Perona-Malik function as the tonal weight function in the smoothness term. There are two versions of the Perona-Malik function. We will have a look at both and see if they belong to $\mathcal{G}^{(S1)}$.

The first version of the Perona-Malik function is given by,

$$\boxed{\Gamma_1(u) = \gamma^2 \log\left(1 + \frac{u^2}{\gamma^2}\right), \quad \gamma > 0.} \tag{6.17}$$

The function is symmetric around zero $\Rightarrow$ (S1) a).

The first three derivatives are given by

$$\Gamma_1'(u) = \frac{2u}{1 + \frac{u^2}{\gamma^2}}, \quad \Gamma_1''(u) = \frac{2 - \frac{2u^2}{\gamma^2}}{\left(1 + \frac{u^2}{\gamma^2}\right)^2}, \quad \text{and} \quad \Gamma_1'''(u) = \frac{-\frac{4u}{\gamma^2}\left(3 - \frac{u^2}{\gamma^2}\right)}{\left(1 + \frac{u^2}{\gamma^2}\right)^3}.$$

Using the first derivative test for Γ_1'' we get that

$$-0.25 \leq \Gamma_1''(u) \leq 2 \quad \text{for all } u \in \mathbb{R}, \tag{6.18}$$

where

$$\Gamma_1''(\pm\sqrt{3}\gamma) = -0.25 \quad \text{and} \quad \Gamma_1''(0) = 2. \tag{6.19}$$

Therefore, Γ_1 is not convex and hence $\Gamma_1 \notin \mathcal{G}^{(S1)}$.

The second version of the Perona-Malik function is given by,

$$\boxed{\Gamma_2(u) = \gamma^2 \left(1 - e^{-\frac{u^2}{\gamma^2}}\right), \quad \gamma > 0.} \tag{6.20}$$

The function is symmetric around zero $\Rightarrow$ (S1) a).

The first three derivatives are given by

$$\Gamma_2'(u) = 2ue^{-\frac{u^2}{\gamma^2}}, \quad \Gamma_2''(u) = \left(2 - \frac{4u^2}{\gamma^2}\right) e^{-\frac{u^2}{\gamma^2}}, \quad \text{and} \quad \Gamma_2'''(u) = -\frac{2}{\gamma^2} u \left(6 - \frac{4u^2}{\gamma^2}\right) e^{-\frac{u^2}{\gamma^2}}.$$

The first derivative test for Γ_2'' implies that

$$-4e^{-3/2} \le \Gamma_2''(u) \le 2 \quad \text{for all } u \in \mathbb{R}, \tag{6.21}$$

where

$$\Gamma_2''(\pm\sqrt{1.5}\gamma) = -4e^{-3/2} \approx -0.8925 \quad \text{and} \quad \Gamma_1''(0) = 2. \tag{6.22}$$

Like Γ_1, $\Gamma_2 \notin \mathcal{G}^{(S1)}$ since Γ_2 is not convex.

That is, we cannot say that nonlinear diffusion regularization (with both versions of the Perona-Malik function) is asymptotically equivalent to quadratic regularization.

6.5.4 Nonlinear Regularization (N-R) & Total Variation Regularization (TV-R)

The term nonlinear regularization stands for using

$$\boxed{\Gamma_3(u) = 2\gamma^2 \left(\sqrt{1 + \frac{u^2}{\gamma^2}} - 1\right), \quad \gamma > 0,} \tag{6.23}$$

as the tonal weight function in the smoothness term. While total variation regularization stands for using

$$\boxed{\Gamma_4(u) = 2\left(\sqrt{u^2 + \gamma^2} - \gamma^2\right), \quad \gamma > 0,} \tag{6.24}$$

instead.

But a simple calculation shows that

$$\boxed{\Gamma_4(u) = \frac{1}{\gamma}\Gamma_3(u).} \tag{6.25}$$

That is, if the scaling parameter γ is fixed and non-zero, then both functions are equivalent as regularization functions.

So, we will have a look to see if Γ_3 belongs to $\mathcal{G}^{(S1)}$ and consequently Γ_4.

From the u^2 ingredient, Γ_3 is symmetric around zero $\Rightarrow$ (S1) a).

The first three derivatives of Γ_3 are given by

$$\Gamma_3'(u) = 2u\left(1+\frac{u^2}{\gamma^2}\right)^{-1/2},$$
$$\Gamma_3''(u) = \frac{-2u^2}{\gamma^2}\left(1+\frac{u^2}{\gamma^2}\right)^{-3/2} + 2\left(1+\frac{u^2}{\gamma^2}\right)^{-1/2},$$
$$\Gamma_3'''(u) = \frac{6u^3}{\gamma^4}\left(1+\frac{u^2}{\gamma^2}\right)^{-5/2} - \frac{6u}{\gamma^2}\left(1+\frac{u^2}{\gamma^2}\right)^{-3/2}.$$

Using the first derivative test for Γ_3'' we get that

$$0 < \Gamma_3''(u) \leq 2 \quad \text{for all } u \in \mathbb{R}. \tag{6.26}$$

Hence, Γ_3 is convex and Γ_3' is Lipschitz continuous. That is, Γ_3 fulfills (S1) b) and c).

Finally,

$$\Gamma_3(0) = 0, \quad \Gamma_3'(0) = 0, \quad \Gamma_3''(0) = 2 \neq 0, \tag{6.27}$$

hence (S1) d) is fulfilled.

Therefore, $\Gamma_3 \in \mathcal{G}^{(S1)}$, and hence $\Gamma_4 \in \mathcal{G}^{(S1)}$.

That is, using nonlinear or total variation regularization is asymptotically equivalent to using quadratic regularization.

6.5.5 Conclusion on Examples

From the above examples we have the following remarks.

Remark 6.12 (Huber-R, N-R, and TV-R) *From the above arguments we can see that using the Huber function or Γ_3 or Γ_4 as ρ_S will give an estimate that is equivalent to the estimate obtained using the quadratic ρ_S.*

Remark 6.13 (ND-R: PM I and II) *Functions Γ_1 and Γ_2 do not belong to $\mathcal{G}^{(S1)}$. That is, we cannot say that nonlinear diffusion regularization (with both versions of the Perona-Malik function) is asymptotically equivalent to quadratic regularization.*

Chapter 7

Simulation Study

In this chapter we do a Monte Carlo simulation study to have an idea about the performance of several proposed solutions of problem (5.1). In this chapter, we do not aim to provide an algorithm for solving problems like (5.1), such algorithms are given in [5] and [22]. What we are really interested in here, is to make a comparison among different smoothers which solve (5.1).

7.1 General Setup and Notation

We generate data $\{(x_i, f_j) : j = 1, ..., N\}$ using the nonparametric regression model

$$f_j = \mu(x_j) + \varepsilon_j, \quad j = 1, \dots, N, \tag{7.1}$$

where $x_j = \frac{j}{N}$ are from an equidistant grid in the unit interval $[0, 1]$.

The regression function is

$$\mu(x_j) = \sin(2\pi x_j) \tag{7.2}$$

and the error terms ε_j follow different laws.

Using the general approach proposed by Mrázek et al. we consider

$$Q(u_1, \dots, u_N) = \sum_{i,j=1}^{N} \rho(u_i - f_j) K_h(x_i - x_j) + \frac{\lambda}{2} \sum_{i,j=1}^{N} \frac{1}{2}(u_i - u_j)^2 L_g(x_i - x_j) = \min_{u_1,\dots,u_N}!$$

We use the probability density function of the standard normal distribution

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

for the kernels K and L.

We use different choices for ρ, namely,

$$\rho_{LS}(u) = \frac{1}{2}u^2, \qquad \rho_{HM}(u) = \begin{cases} \begin{cases} \frac{1}{2}u^2, & |u| \leq c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases} & c \geq 1, \\ \\ \begin{cases} \frac{1}{2c}u^2, & |u| \leq c, \\ |u| - \frac{1}{2}c, & |u| > c, \end{cases} & c \leq 1, \end{cases} \qquad \text{and} \qquad \rho_{LA}(u) = |x|.$$

we will denote the solutions obtained from solving these problems by $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$, and $\hat{\mathbf{u}}_{\mathbf{LA}}$ respectively. If $\lambda = 0$ then the solution of the problem using the quadratic function as the data loss function is the Nadaraya-Watson kernel estimate which is asymptotically equivalent to the Priestley-Chao kernel estimate $\hat{\mathbf{u}}_{\mathbf{PC}} = \hat{\boldsymbol{\mu}}$.

Implementation

The Implementation was done using MATLAB. $\hat{\mathbf{u}}_{\mathbf{PC}}$ was calculated directly since it has an explicit form. The other estimates ($\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$, and $\hat{\mathbf{u}}_{\mathbf{LA}}$) were calculated be minimizing the energy functional Q, using the MATLAB function `fminunc`. Different values for the parameters h, g, λ, and c were considered.

The calculations were repeated M times; each time giving the solutions $\hat{\mathbf{u}}_{\mathbf{PC}}^{(k)}$, $\hat{\mathbf{u}}_{\mathbf{LS}}^{(k)}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}^{(k)}$, and $\hat{\mathbf{u}}_{\mathbf{LA}}^{(k)}$ where $k = 1, \ldots, M$. Then the average was taken, i.e. in the tables below

$$\hat{\mathbf{u}}_* = \frac{1}{M}\sum_{k=1}^{M} \hat{\mathbf{u}}_*^{(j)} \quad \text{where} \quad * \in \{\mathbf{PC}, \mathbf{LS}, \mathbf{HM(c)}, \mathbf{LA}\}.$$

As a measure of quality for these estimates we will use the empirical root mean-squared error and the empirical mean absolute deviation.

Definition 7.1 *Let $T = (T_1, \ldots, T_N)$ be an estimate for the parameter $\theta = (\theta_1, \ldots, \theta_N)$ then the empirical root mean-squared error (ermse) and the empirical mean absolute deviation (emad) are defined as follows,*

$$ermse(T) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - \theta_i)^2}, \quad emad(T) = \frac{1}{N}\sum_{i=1}^{N}|T_i - \theta_i|.$$

Also, we will provide values for the *maximum absolute deviation (maxad)*

$$maxad(T) = \max_{1 \leq i \leq N} |T_i - \theta_i|.$$

Types of Error Terms

In the simulations done here we consider two ways to study the performance of our proposed estimates. First, we consider three different types of errors,

(1) Pure normal error terms with $\mathbb{E}\,\varepsilon_j = 0$ and $\operatorname{var}\varepsilon_j = 0.04$,

$$\mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2)$$

that is

$$p_\varepsilon(x) = \varphi_{0,0.04}(x).$$

(2) Mixed (or contaminated) normal error terms with $\mathbb{E}\,\varepsilon_j = 0$ and $\operatorname{var}\varepsilon_j = 3.232$,

$$\mathcal{L}(\varepsilon_j) = \mathcal{B}(1, 0.8)\mathcal{N}(0, 0.2^2) + \mathcal{B}(1, 0.2)\mathcal{N}(0, 4^2)$$

that is,

$$p_\varepsilon(x) = 0.8\,\varphi_{0,0.04}(x) + 0.2\,\varphi_{0,16}(x).$$

(3) Double exponential (or Laplace) error terms with $\mathbb{E}\,\varepsilon_j = 0$ and $\operatorname{var}\varepsilon_j = 0.08$,

$$\mathcal{L}(\varepsilon_j) = \text{Laplace}(0, 0.2)$$

that is,

$$p_\varepsilon(x) = \frac{5}{2}\mathrm{e}^{-5|x|}.$$

Then, we add outliers to the regression function itself and use normal error terms, i.e.

(4) Single outlier with normal error terms,

$$\mu(x) = \begin{cases} \sin(2\pi x), & x \neq \frac{1}{2} \\ 1, & x = \frac{1}{2}. \end{cases} \quad \text{and} \quad \mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2).$$

(5) Two outliers with normal error terms,

$$\mu(x) = \begin{cases} \sin(2\pi x), & x \neq \frac{1}{4}, \frac{1}{2}, \\ -1, & x = \frac{1}{4}, \\ 1, & x = \frac{1}{2}. \end{cases} \quad \text{and} \quad \mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2).$$

In this way, we will be able to see the robustness properties of the M-estimates and the LA-estimates.

7.2 Results 1: Pure Normal Error Terms

The error terms here follow the law $\mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2)$, i.e.

```
x = [0:1/N:1]';
e = normrnd(0,0.2,length(x),1);
```

In the following we summarize the results obtained for $N = 100$ and $M = 100$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.050943	0.057118	0.057115	0.056659	0.054807	0.071585
0.03	0.01	0.1	0.048243	0.052190	0.052189	0.051515	0.049269	0.054301
0.03	0.01	2	0.048802	0.058498	0.058497	0.056078	0.053311	0.057599
0.03	0.01	40	0.050839	0.097679	0.097679	0.083246	0.073063	0.073971
0.01	0.005	20	0.040687	0.048018	0.048018	0.043411	0.041613	0.044305
0.01	0.005	40	0.037191	0.050480	0.050480	0.040870	0.037347	0.038798
0.015	0.005	40	0.037770	0.056041	0.056041	0.047909	0.041794	0.043470
0.015	0.005	80	0.040025	0.069836	0.069836	0.058016	0.049473	0.050854

Table 7.1: Pure normal error terms: ermse for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.022986	0.021726	0.021722	0.021065	0.018611	0.043629
0.03	0.01	0.1	0.019968	0.017793	0.017790	0.017183	0.015176	0.017856
0.03	0.01	2	0.020386	0.022062	0.022062	0.019941	0.016878	0.019605
0.03	0.01	40	0.021661	0.075522	0.075522	0.051429	0.036414	0.037433
0.01	0.005	20	0.013889	0.015313	0.015313	0.012965	0.012517	0.014336
0.01	0.005	40	0.011357	0.017503	0.017503	0.011038	0.009547	0.010205
0.015	0.005	40	0.011806	0.020688	0.020688	0.014219	0.010987	0.012043
0.015	0.005	80	0.012498	0.035391	0.035391	0.022226	0.015086	0.015887

Table 7.2: Pure normal error terms: emad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.082638	0.147336	0.147334	0.145902	0.147825	0.169853
0.03	0.01	0.1	0.068887	0.124539	0.124537	0.122330	0.118778	0.147371
0.03	0.01	2	0.075555	0.160505	0.160505	0.148463	0.137546	0.159834
0.03	0.01	40	0.079041	0.322278	0.322278	0.264427	0.228605	0.233755
0.01	0.005	20	0.039909	0.114493	0.114494	0.092762	0.074331	0.090366
0.01	0.005	40	0.030475	0.121022	0.121024	0.086823	0.066230	0.073348
0.015	0.005	40	0.038565	0.150667	0.150667	0.117505	0.088399	0.096181
0.015	0.005	80	0.042069	0.202010	0.202010	0.156136	0.122143	0.126962

Table 7.3: Pure normal error terms: maxad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

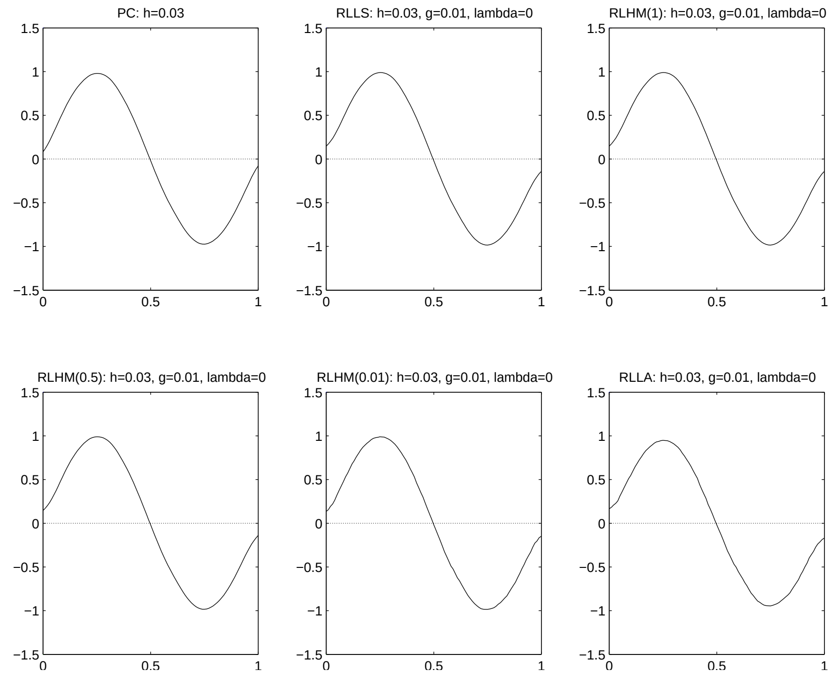

Figure 7.1: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0)$.

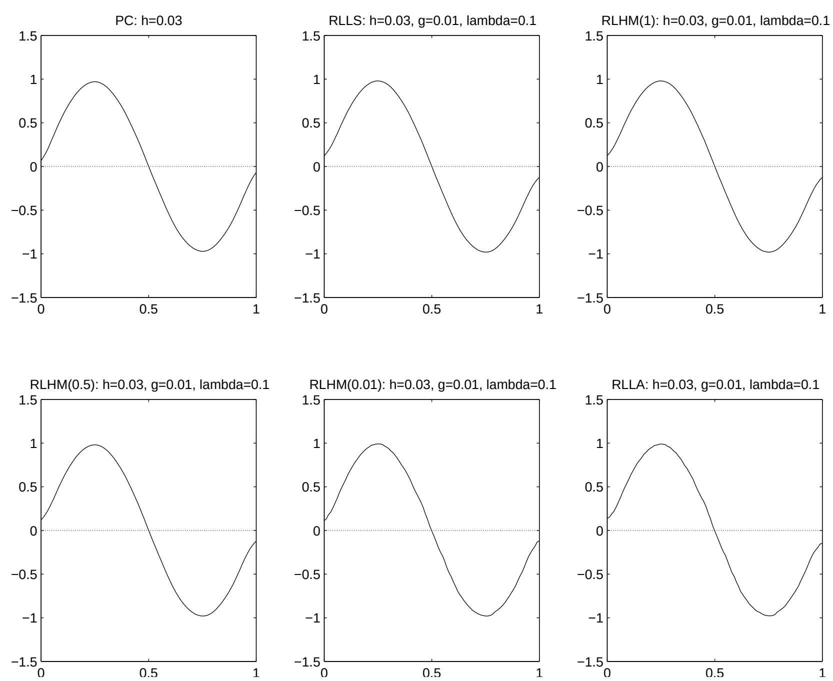

Figure 7.2: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0.1)$.

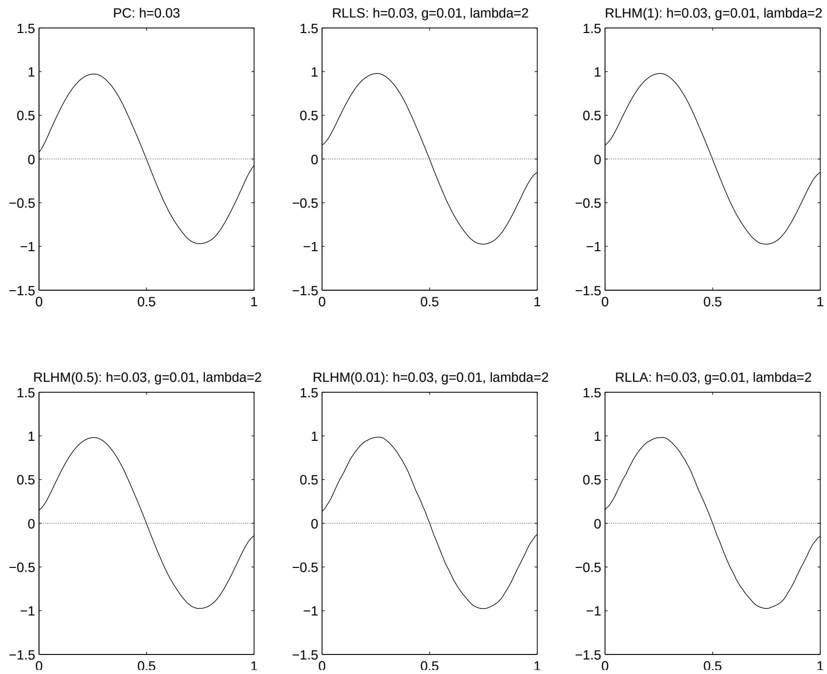

Figure 7.3: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 2)$.

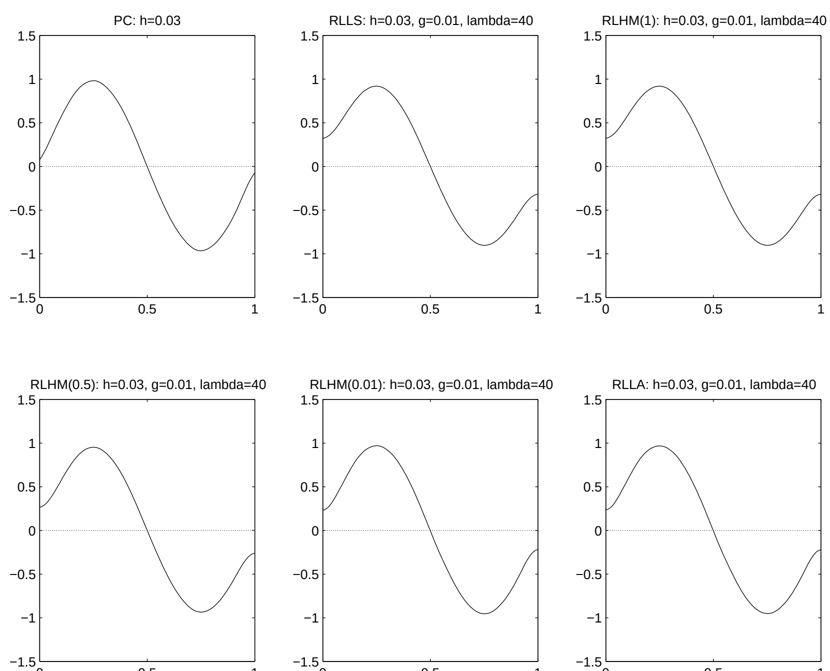

Figure 7.4: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 40)$.

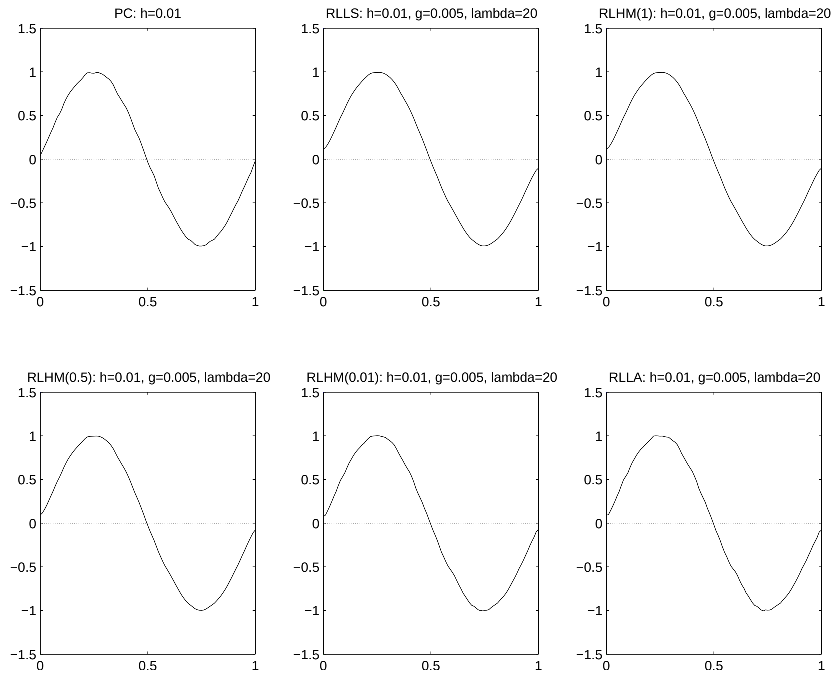

Figure 7.5: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 20)$.

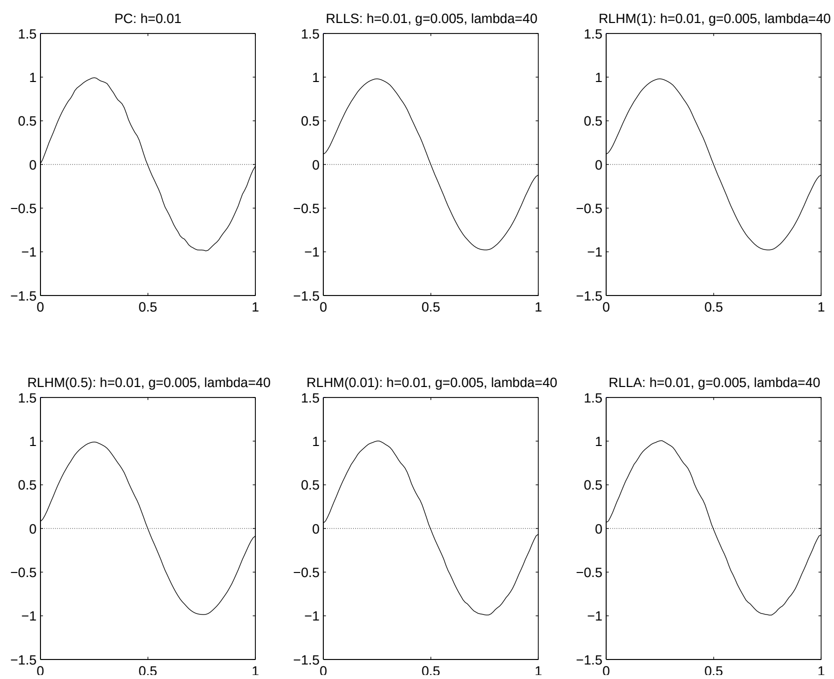

Figure 7.6: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 40)$.

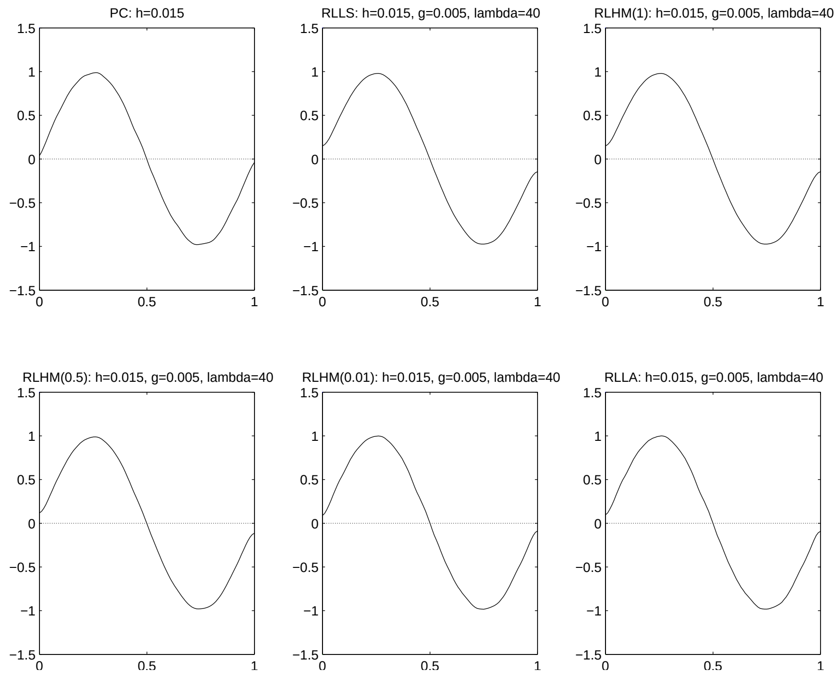

Figure 7.7: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 40)$.

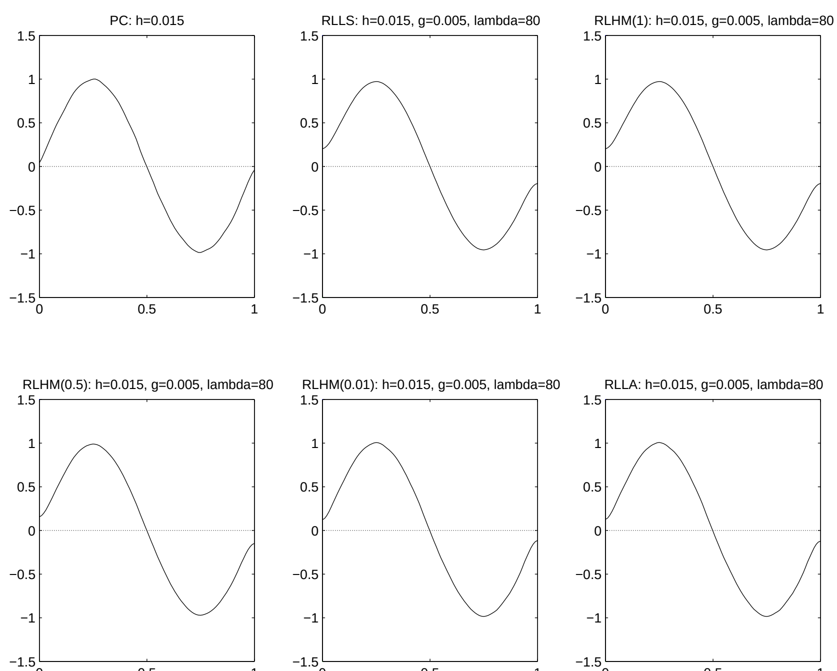

Figure 7.8: Pure normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 80)$.

From the previous tables and figures we note the following,

1) The ermse of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were approximately equal.

2) The emad of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were approximately equal.

3) The maxad of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were approximately equal.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{LS}}$ is quit close to $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$.

4) The estimate with the lowest ermse was $\hat{\mathbf{u}}_{\mathbf{PC}}$ and then $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

5) The estimate with the lowest emad was $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ for different λ. But, $\hat{\mathbf{u}}_{\mathbf{PC}}$ got smaller emad values when λ was very big.

6) The smallest maxad was of $\hat{\mathbf{u}}_{\mathbf{PC}}$ then of $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{PC}}$ is best estimate here. Another good estimate would be $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

7.3 Results 2: Mixed Normal Error Terms

The error terms here follow the law $\mathcal{L}(\varepsilon_j) = \mathcal{B}(1, 0.8)\mathcal{N}(0, 0.2^2) + \mathcal{B}(1, 0.2)\mathcal{N}(0, 4^2)$, i.e.

```
x = [0:1/N:1]';
b = binornd(1,0.8,length(x),1);
e = b.*normrnd(0,0.2,length(x),1) + (1-b).*normrnd(0,4,length(x),1);
```

In the following we summarize the results obtained for $N = 100$ and $M = 100$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.074307	0.075227	0.058036	0.054972	0.052522	0.064782
0.03	0.01	0.1	0.075126	0.077162	0.054044	0.050466	0.047962	0.052022
0.03	0.01	2	0.087388	0.095469	0.064980	0.059038	0.056165	0.059770
0.03	0.01	40	0.073115	0.106002	0.103856	0.088082	0.077742	0.080670
0.01	0.005	20	0.093787	0.078070	0.054820	0.046327	0.040784	0.042541
0.01	0.005	40	0.103375	0.083676	0.055600	0.047574	0.043383	0.044613
0.015	0.005	40	0.098545	0.091438	0.065016	0.054462	0.048369	0.049275
0.015	0.005	80	0.090448	0.077505	0.068145	0.057923	0.051090	0.052001

Table 7.4: Mixed normal error terms: ermse for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.047879	0.049508	0.025214	0.021426	0.019562	0.034841
0.03	0.01	0.1	0.046102	0.049789	0.023985	0.019027	0.014871	0.017445
0.03	0.01	2	0.059276	0.064902	0.028925	0.022794	0.018963	0.021285
0.03	0.01	40	0.044213	0.094155	0.088858	0.060037	0.042203	0.048765
0.01	0.005	20	0.070629	0.052082	0.023938	0.016306	0.012088	0.013089
0.01	0.005	40	0.086919	0.051087	0.022428	0.015334	0.013321	0.013757
0.015	0.005	40	0.077407	0.069110	0.031228	0.021258	0.016146	0.016787
0.015	0.005	80	0.067256	0.045419	0.032590	0.021427	0.015975	0.016908

Table 7.5: Mixed normal error terms: emad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.104916	0.131995	0.146678	0.141179	0.132028	0.169280
0.03	0.01	0.1	0.113405	0.106609	0.120900	0.120215	0.113394	0.138078
0.03	0.01	2	0.192004	0.360224	0.194811	0.165734	0.166709	0.182641
0.03	0.01	40	0.115513	0.362166	0.353351	0.291899	0.256088	0.261158
0.01	0.005	20	0.261608	0.152482	0.117242	0.095657	0.069779	0.079382
0.01	0.005	40	0.257557	0.192903	0.131664	0.106623	0.090481	0.098800
0.015	0.005	40	0.288690	0.225952	0.166206	0.131101	0.111024	0.115408
0.015	0.005	80	0.168083	0.263322	0.204932	0.165205	0.133298	0.137049

Table 7.6: Mixed normal error terms: maxad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

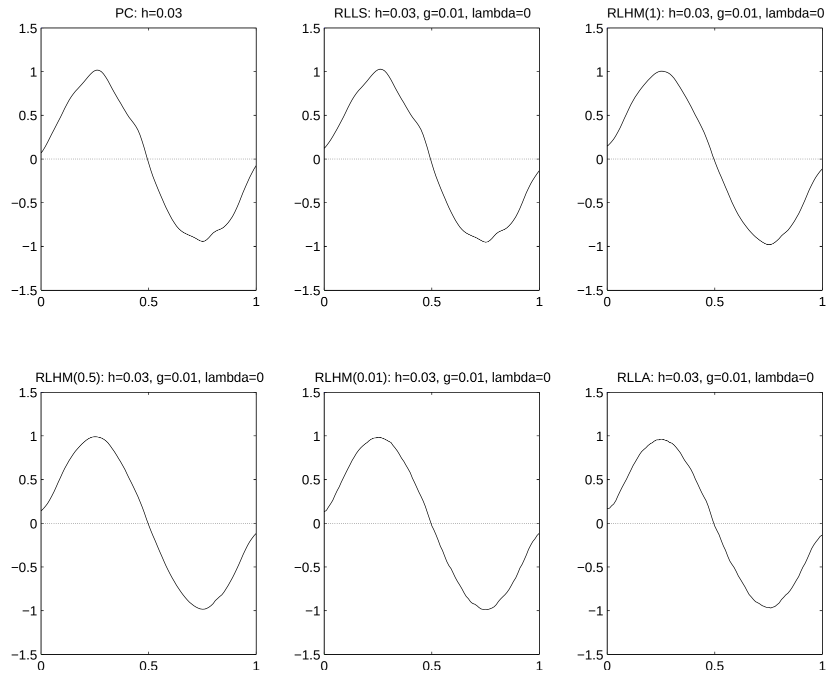

Figure 7.9: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0)$.

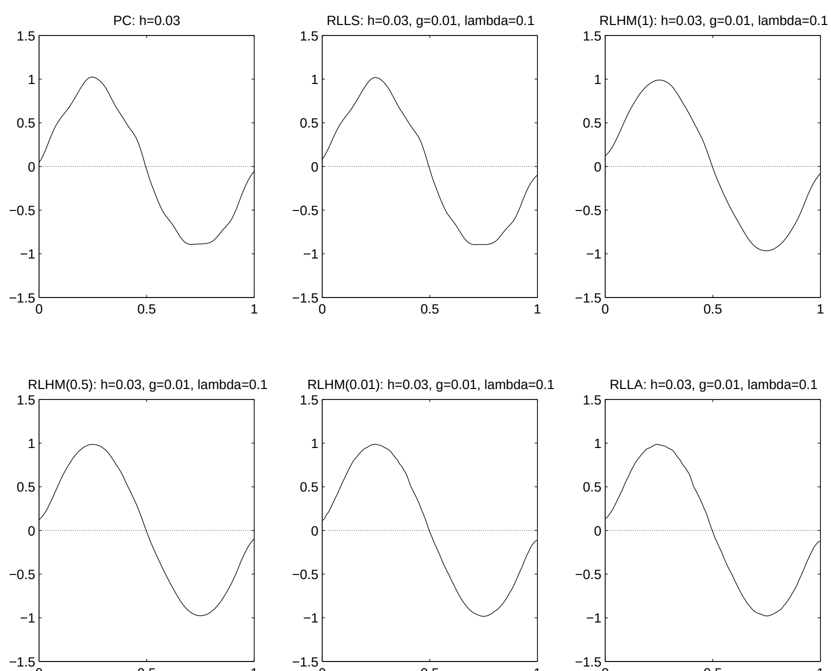

Figure 7.10: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0.1)$.

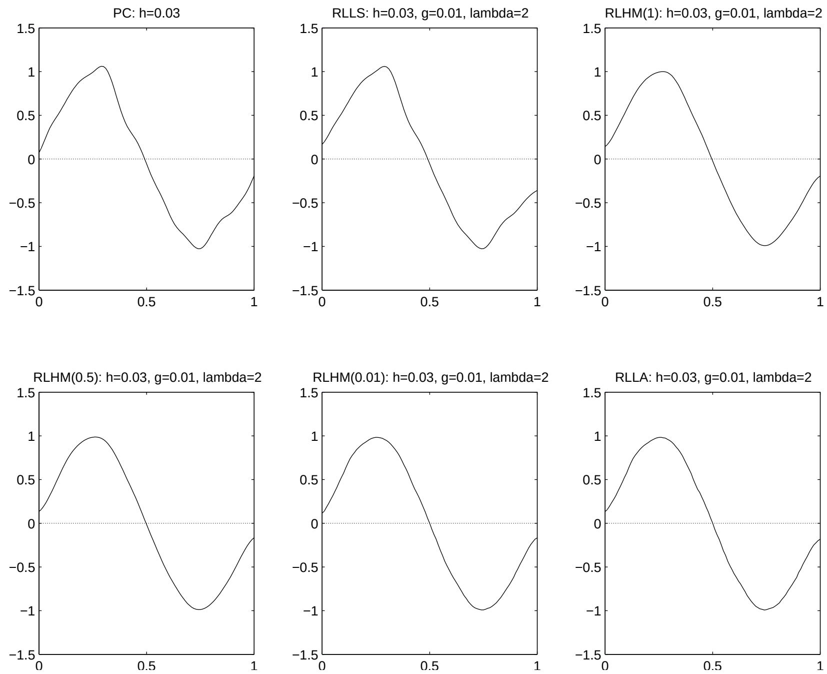

Figure 7.11: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 2)$.

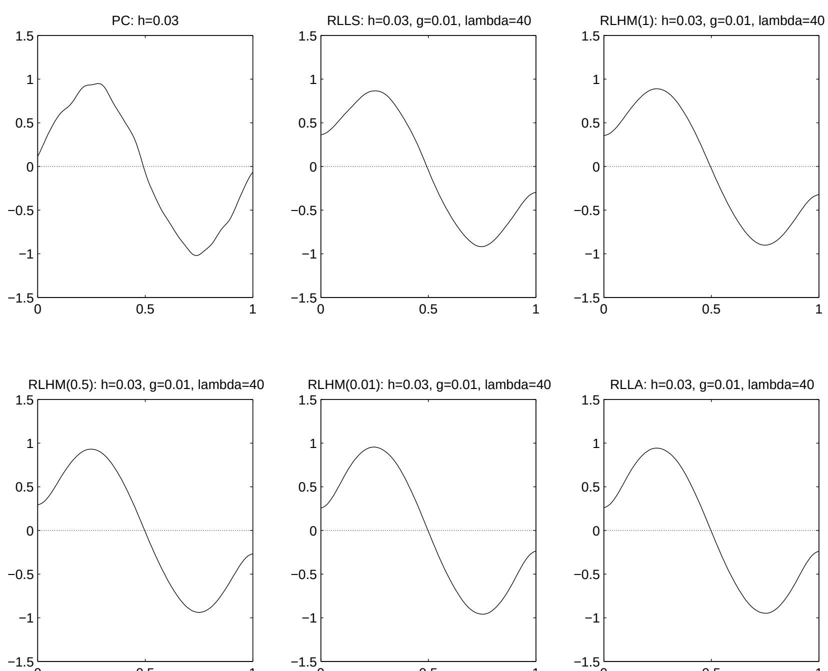

Figure 7.12: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 40)$.

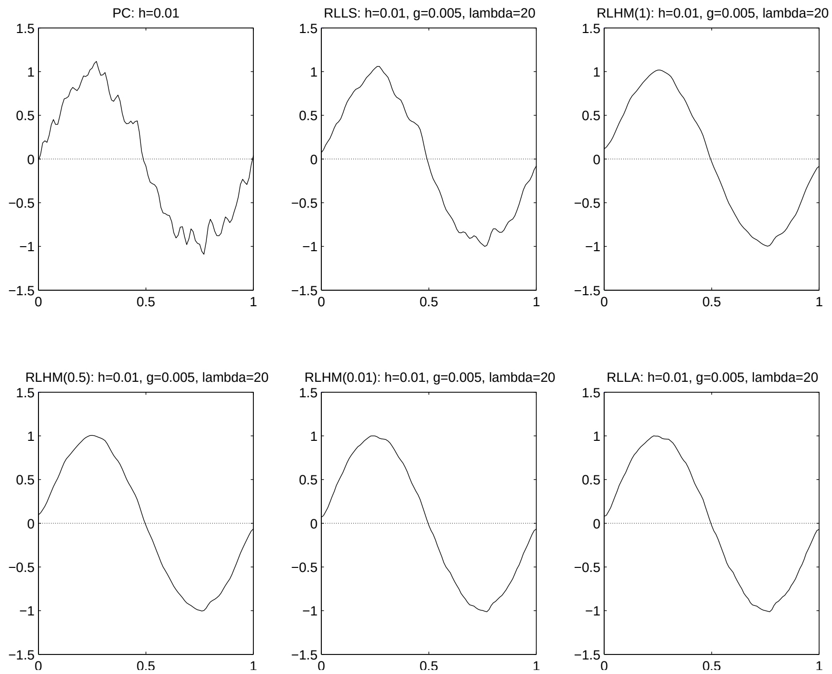

Figure 7.13: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 20)$.

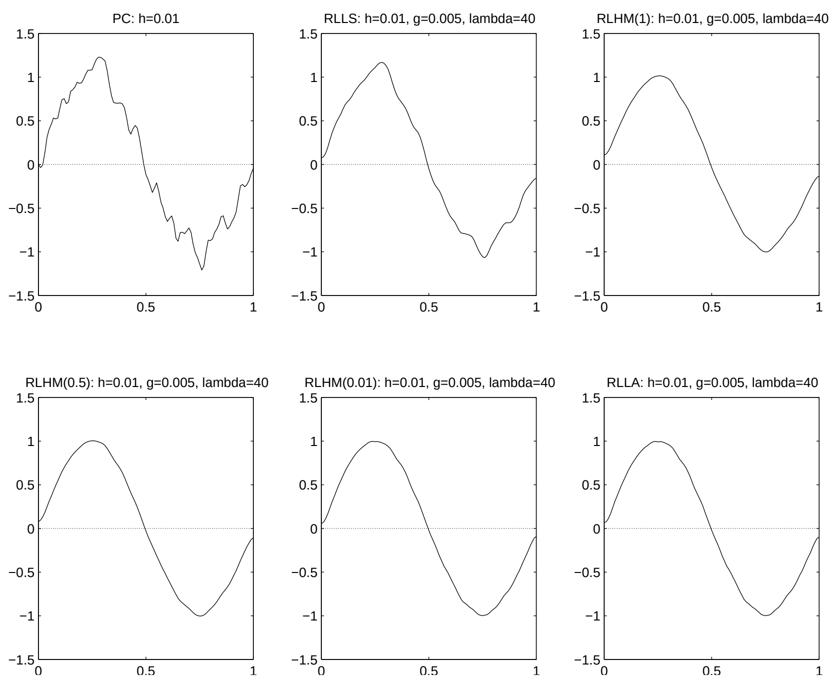

Figure 7.14: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 40)$.

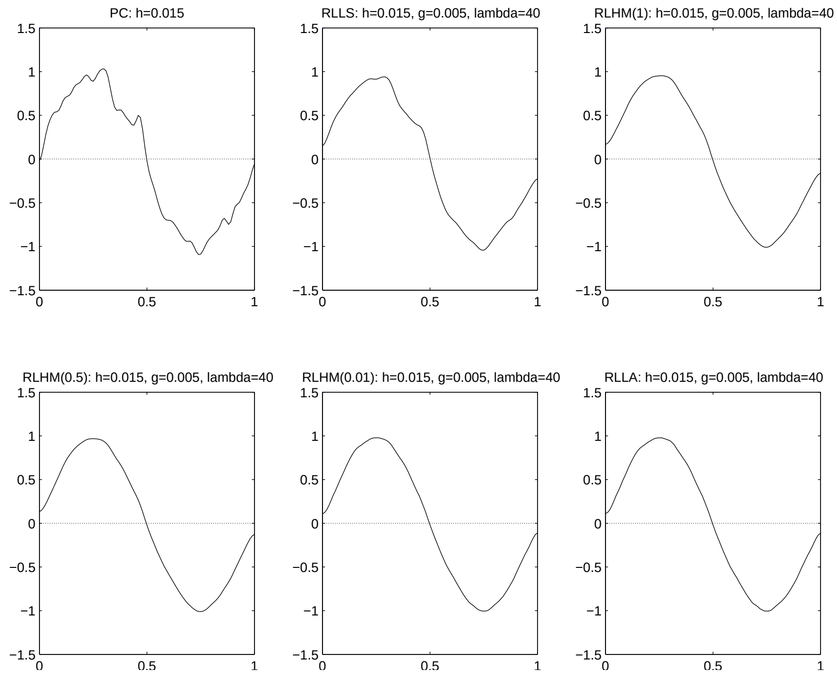

Figure 7.15: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 40)$.

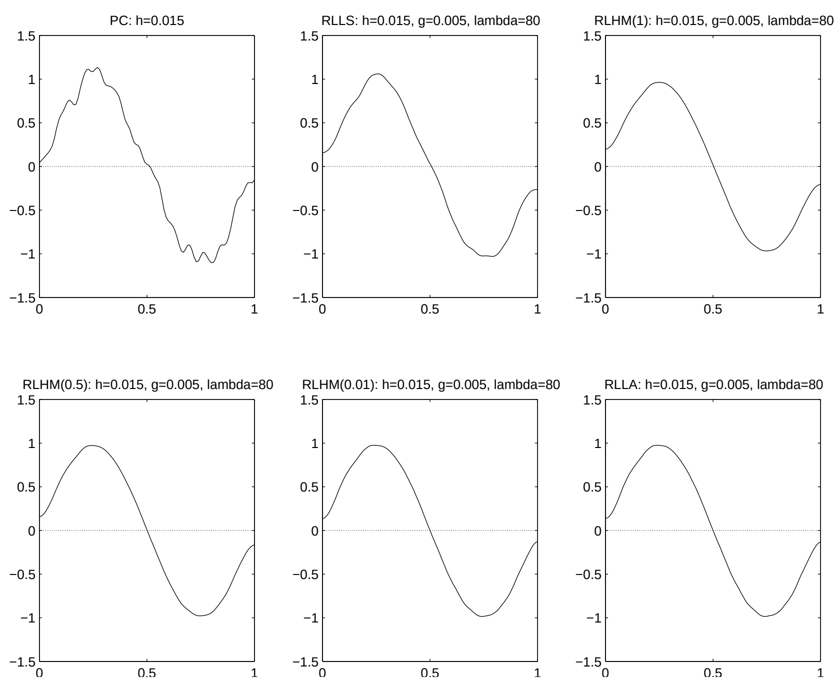

Figure 7.16: Mixed normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 80)$.

From the previous tables and figures we note the following,

1) The ermse of $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were significantly different.

2) The emad of $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were significantly different

3) The maxad of $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were significantly different.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{LS}}$ is different from $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ here. $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ is better here.

4) The estimate with the lowest ermse was $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ and then $\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$ or $\hat{\mathbf{u}}_{\mathbf{LA}}$.

5) The estimate with the lowest emad was $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ then $\hat{\mathbf{u}}_{\mathbf{LA}}$.

6) The smallest maxad was of different estimates for different parameters.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ is the best estimate here. Another good estimate would be $\hat{\mathbf{u}}_{\mathbf{LA}}$ or $\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$.

7.4 Results 3: Double Exponential Error Terms

The error terms here follow the law $\mathcal{L}(\varepsilon_j) = \text{Laplace}(0.2)$, i.e.

```
x = [0:1/N:1]';
b = binornd(1,0.5,length(x),1);
e = (2*b-1).*exprnd(1/5,length(x),1);
```

In the following we summarize the results obtained for $N = 100$ and $M = 100$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.049798	0.05642	0.056205	0.054529	0.053081	0.088469
0.03	0.01	0.1	0.047191	0.054945	0.0546	0.053533	0.05163	0.056168
0.03	0.01	2	0.051128	0.060755	0.060663	0.058167	0.05598	0.059931
0.03	0.01	40	0.046536	0.09528	0.09544	0.082155	0.072019	0.072892
0.01	0.005	20	0.042242	0.0466	0.046456	0.041379	0.040922	0.043555
0.01	0.005	40	0.039473	0.053183	0.052945	0.045212	0.04118	0.042695
0.015	0.005	40	0.04246	0.058419	0.058406	0.050779	0.046894	0.048542
0.015	0.005	80	0.038771	0.066039	0.066137	0.054727	0.048404	0.049699

Table 7.7: Double exponential error terms: ermse for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.0202	0.019623	0.019516	0.018589	0.017791	0.070262
0.03	0.01	0.1	0.018198	0.016912	0.016394	0.01554	0.015544	0.018155
0.03	0.01	2	0.021766	0.023568	0.023396	0.021303	0.018538	0.021076
0.03	0.01	40	0.017941	0.071036	0.071389	0.049465	0.033775	0.034718
0.01	0.005	20	0.01464	0.014081	0.013985	0.012216	0.012369	0.014019
0.01	0.005	40	0.012252	0.016486	0.01609	0.010953	0.010774	0.011302
0.015	0.005	40	0.014653	0.022253	0.022181	0.016517	0.013722	0.014777
0.015	0.005	80	0.012595	0.030741	0.030968	0.019521	0.013843	0.01486

Table 7.8: Double exponential error terms: emad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.080986	0.14439	0.14392	0.13679	0.1465	0.15733
0.03	0.01	0.1	0.085084	0.15315	0.15296	0.14931	0.14007	0.16721
0.03	0.01	2	0.085583	0.17842	0.17803	0.16324	0.15875	0.17687
0.03	0.01	40	0.068263	0.31611	0.31729	0.26404	0.22468	0.22896
0.01	0.005	20	0.051542	0.10538	0.10535	0.078236	0.075993	0.087367
0.01	0.005	40	0.057715	0.15091	0.15097	0.12263	0.10636	0.11458
0.015	0.005	40	0.04795	0.16816	0.16799	0.13342	0.12371	0.12956
0.015	0.005	80	0.034213	0.18776	0.18847	0.14289	0.12355	0.12688

Table 7.9: Double exponential error terms: maxad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

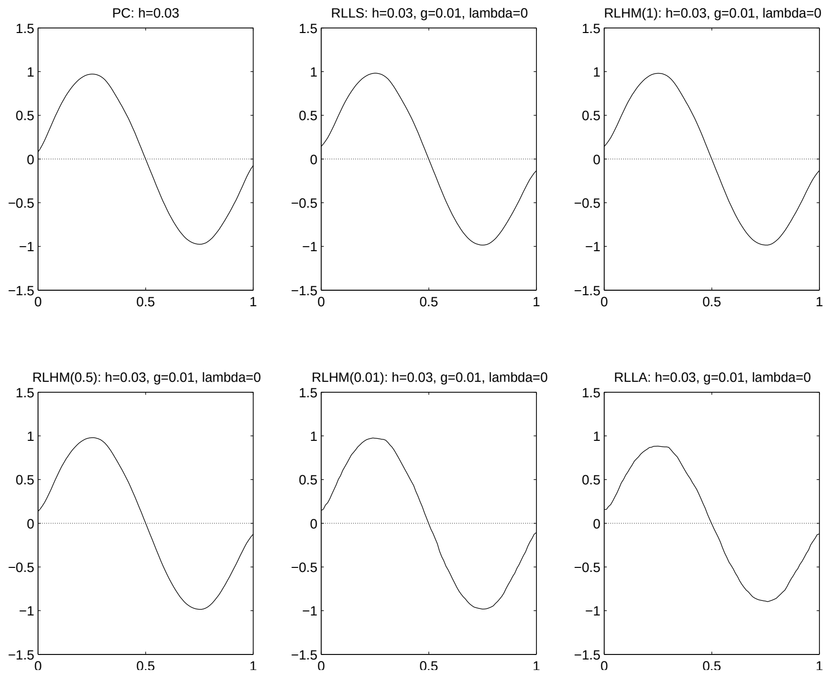

Figure 7.17: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0)$.

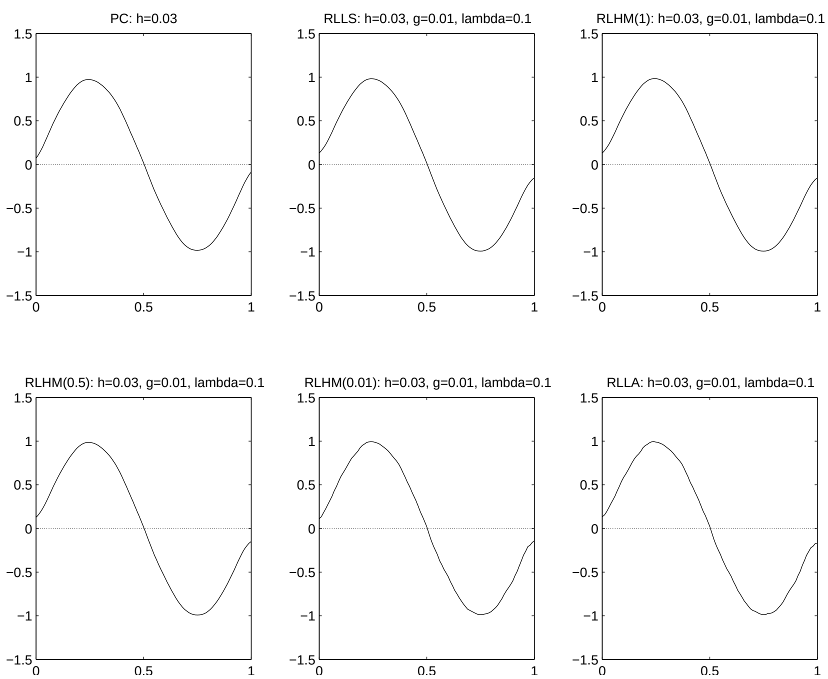

Figure 7.18: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0.1)$.

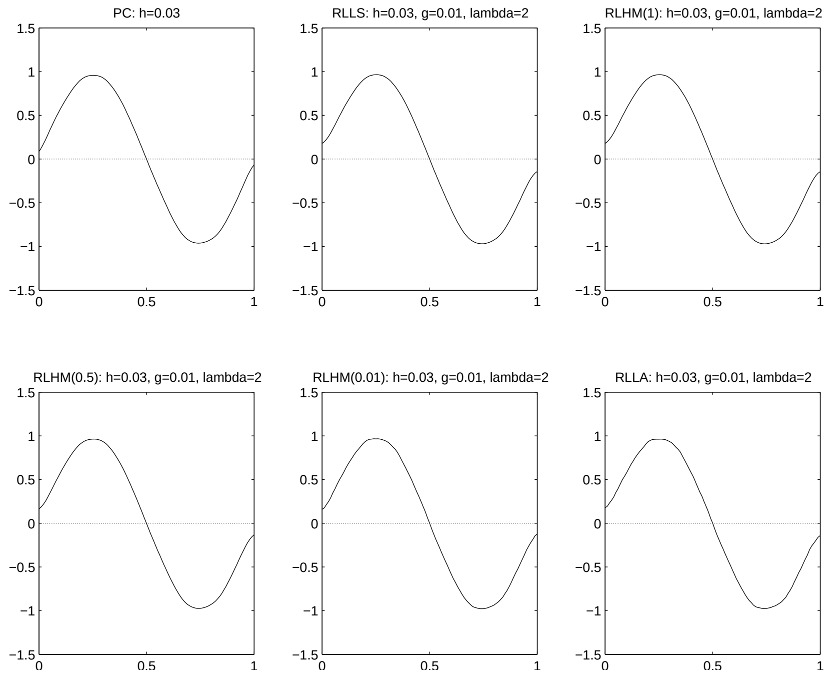

Figure 7.19: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 2)$.

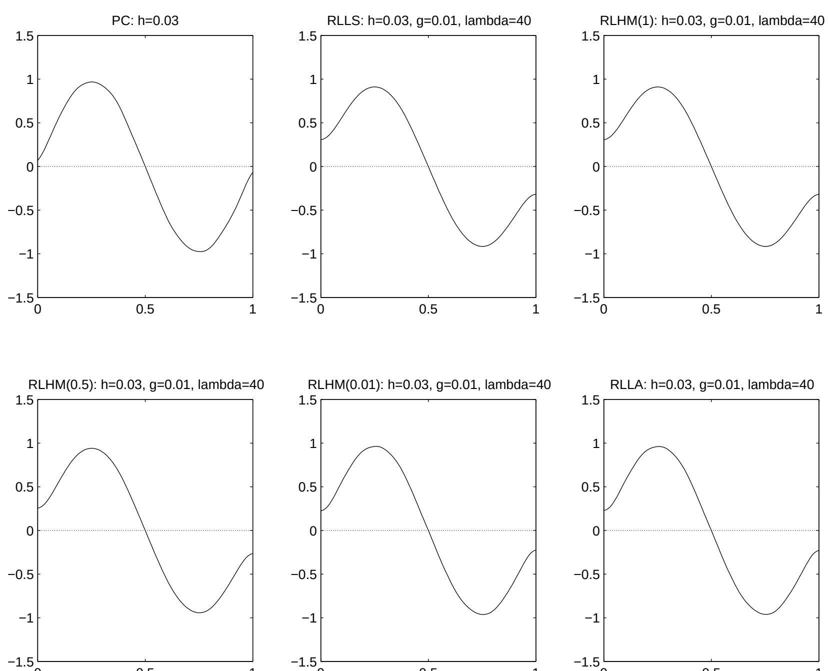

Figure 7.20: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 40)$.

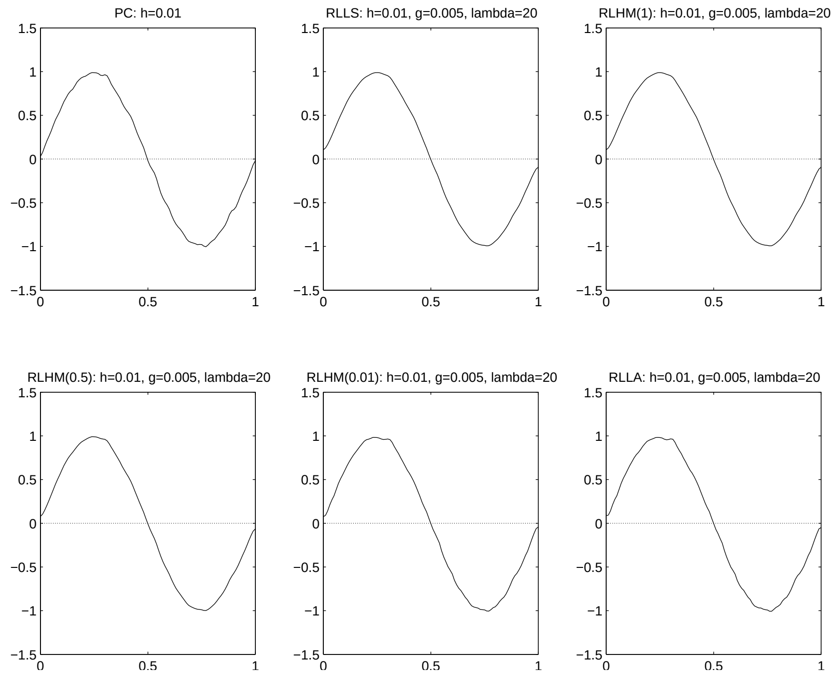

Figure 7.21: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 20)$.

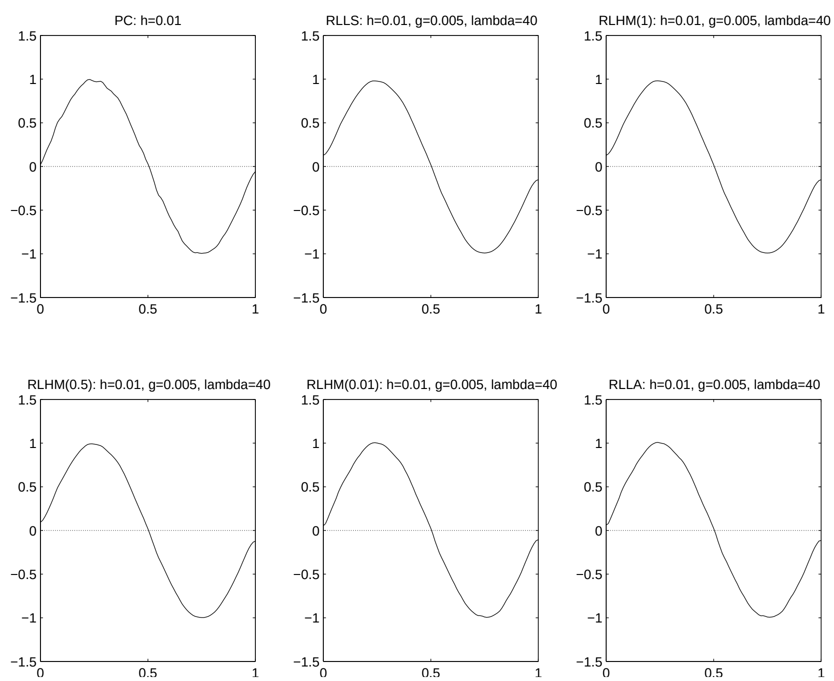

Figure 7.22: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 40)$.

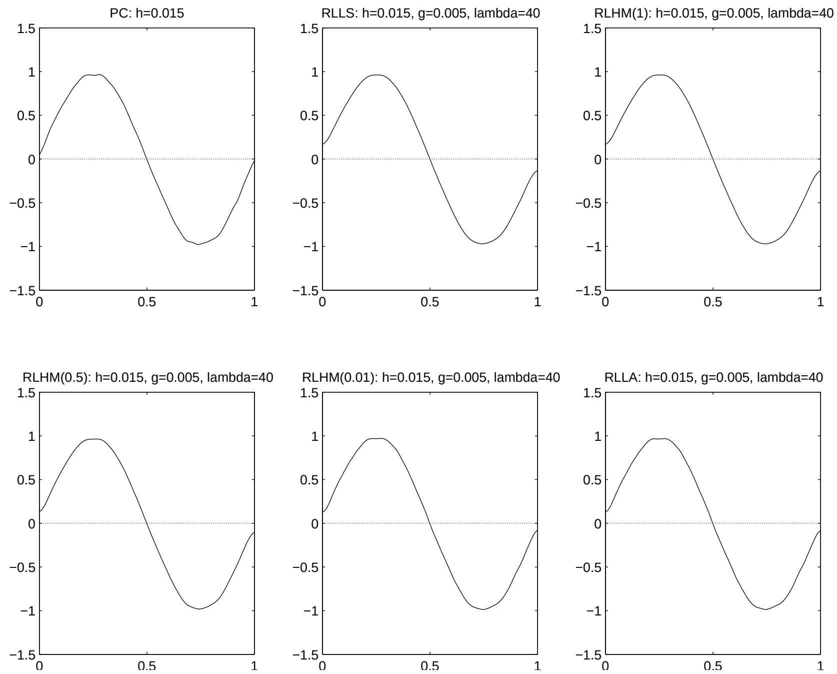

Figure 7.23: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 40)$.

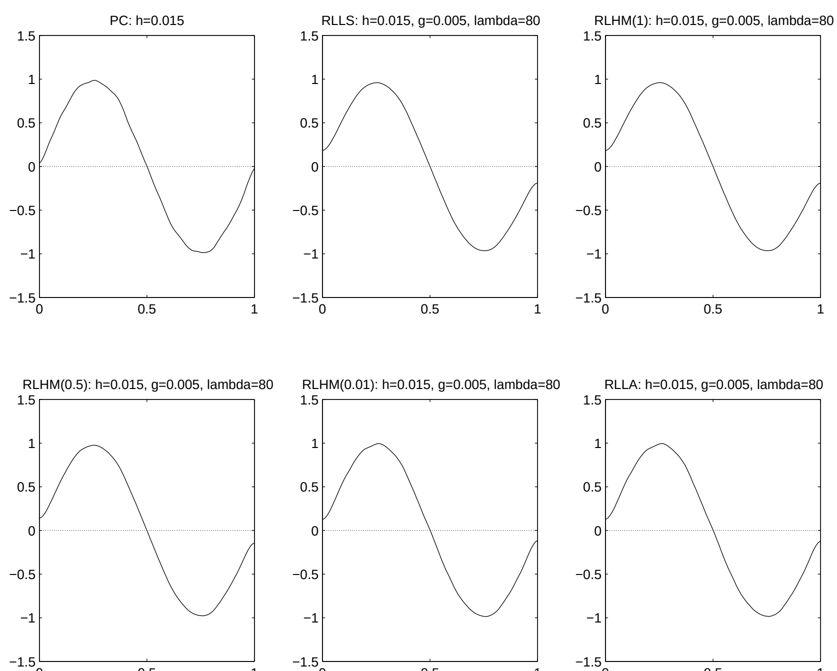

Figure 7.24: Double exponential error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 80)$.

From the previous tables and figures we note the following,

1) The ermse of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were roughly equal.

2) The emad of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were roughly equal.

3) The maxad of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were roughly equal.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{LS}}$ is roughly close to $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ here.

4) The estimate with the lowest ermse was $\hat{\mathbf{u}}_{\mathbf{PC}}$ and then $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

5) There was no single estimate with the lowest emad, but the following estimates had lowest emad for different parameters: $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$, $\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$, $\hat{\mathbf{u}}_{\mathbf{LA}}$, $\hat{\mathbf{u}}_{\mathbf{PC}}$.

6) The smallest maxad was of $\hat{\mathbf{u}}_{\mathbf{PC}}$ then of $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

$\Rightarrow$ $\hat{\mathbf{u}}_{\mathbf{PC}}$ is the best estimate here. Another good estimate would be $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

7.5 Results 4: Single Outlier with Normal Error Terms

We consider here one added outlier to the regression function with normal error terms, i.e.

$$\mu(x) = \begin{cases} \sin(2\pi x), & x \neq \frac{1}{2} \\ 1, & x = \frac{1}{2}. \end{cases} \quad \text{and} \quad \mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2). \tag{7.3}$$

In the following we summarize the results obtained for $N = 100$ and $M = 100$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.097426	0.098437	0.098415	0.099643	0.100686	0.105917
0.03	0.01	0.1	0.097050	0.097595	0.097506	0.098444	0.099575	0.104732
0.03	0.01	2	0.096956	0.098904	0.098891	0.099467	0.099619	0.104185
0.03	0.01	40	0.097387	0.116126	0.116139	0.108822	0.105290	0.105597
0.01	0.005	20	0.085106	0.093536	0.093599	0.094023	0.096499	0.099855
0.01	0.005	40	0.084504	0.095519	0.095561	0.095001	0.096946	0.098283
0.015	0.005	40	0.090848	0.097333	0.097392	0.096971	0.097499	0.098360
0.015	0.005	80	0.090946	0.101907	0.102011	0.099491	0.099368	0.099994

Table 7.10: Single outlier with normal error terms: ermse for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.035959	0.034612	0.034027	0.031425	0.028785	0.033579
0.03	0.01	0.1	0.036708	0.034799	0.034265	0.031330	0.027805	0.032337
0.03	0.01	2	0.035660	0.037356	0.036758	0.032170	0.028301	0.031412
0.03	0.01	40	0.037948	0.089055	0.088419	0.063287	0.047434	0.048186
0.01	0.005	20	0.023925	0.026906	0.026815	0.023047	0.021872	0.023457
0.01	0.005	40	0.022837	0.033402	0.033290	0.024354	0.020379	0.021232
0.015	0.005	40	0.025589	0.035400	0.035164	0.026332	0.021887	0.022794
0.015	0.005	80	0.026501	0.050962	0.050549	0.035330	0.025755	0.026257

Table 7.11: Single outlier with normal error terms: emad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.888892	0.887781	0.890514	0.933186	0.964727	1.052822
0.03	0.01	0.1	0.867638	0.866935	0.869182	0.913597	0.950190	1.041784
0.03	0.01	2	0.873609	0.883505	0.887120	0.926911	0.944263	1.029417
0.03	0.01	40	0.871161	0.928062	0.933691	0.952868	0.969148	0.971174
0.01	0.005	20	0.638993	0.807845	0.810697	0.849510	0.912877	0.976986
0.01	0.005	40	0.617095	0.829165	0.831622	0.865071	0.924888	0.950046
0.015	0.005	40	0.743505	0.857798	0.861459	0.897424	0.928662	0.944095
0.015	0.005	80	0.741230	0.882595	0.889051	0.918522	0.954366	0.965194

Table 7.12: Single outlier with normal error terms: maxad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

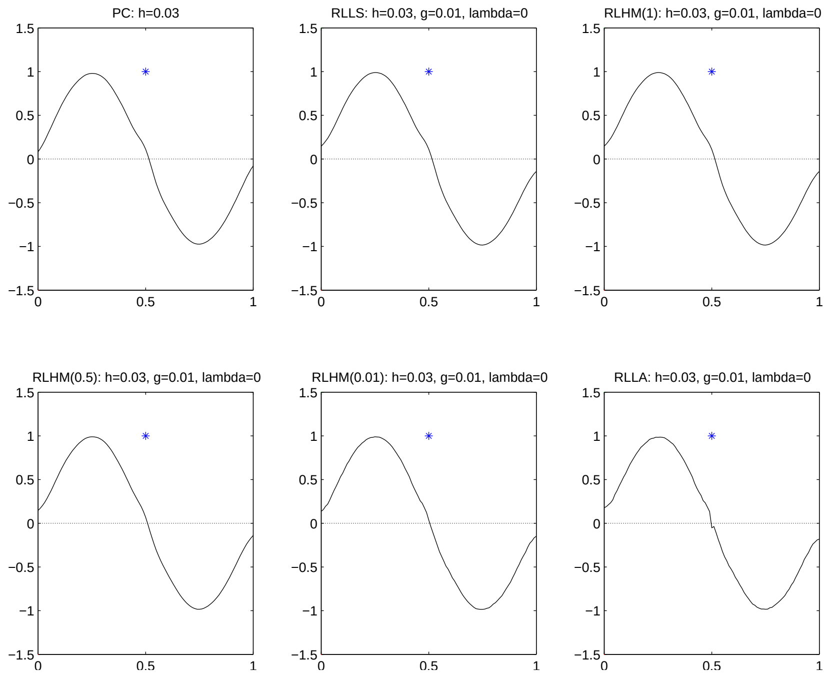

Figure 7.25: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0)$.

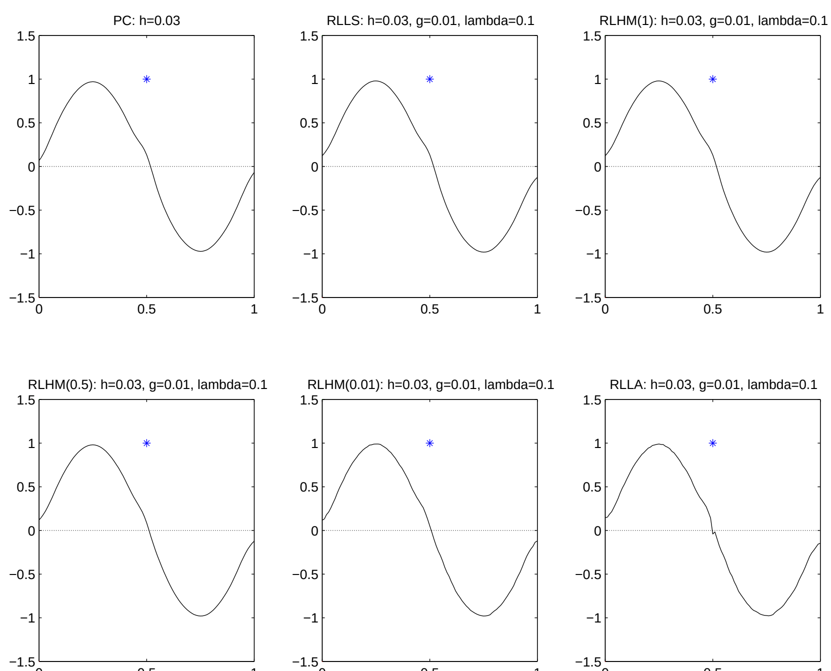

Figure 7.26: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0.1)$.

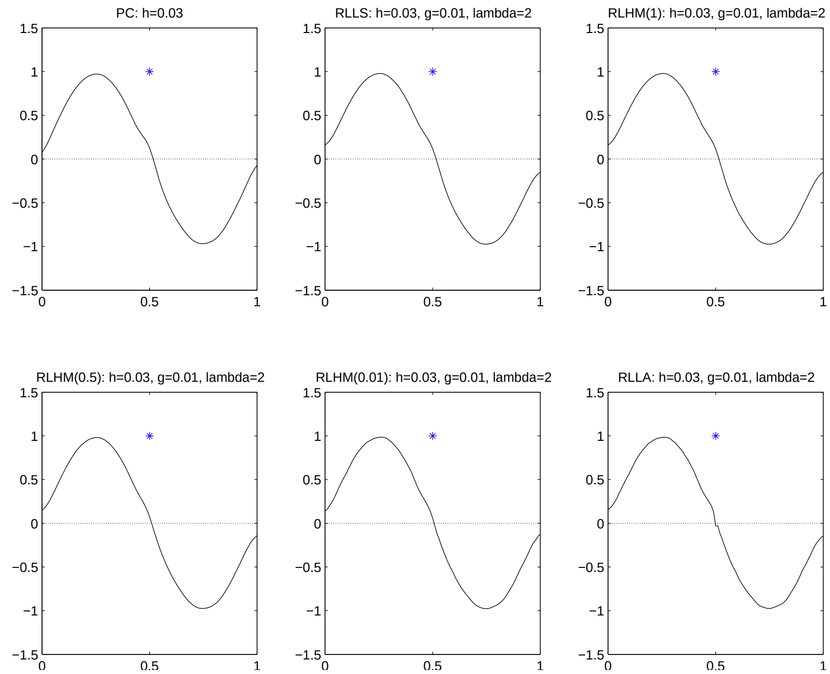

Figure 7.27: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 2)$.

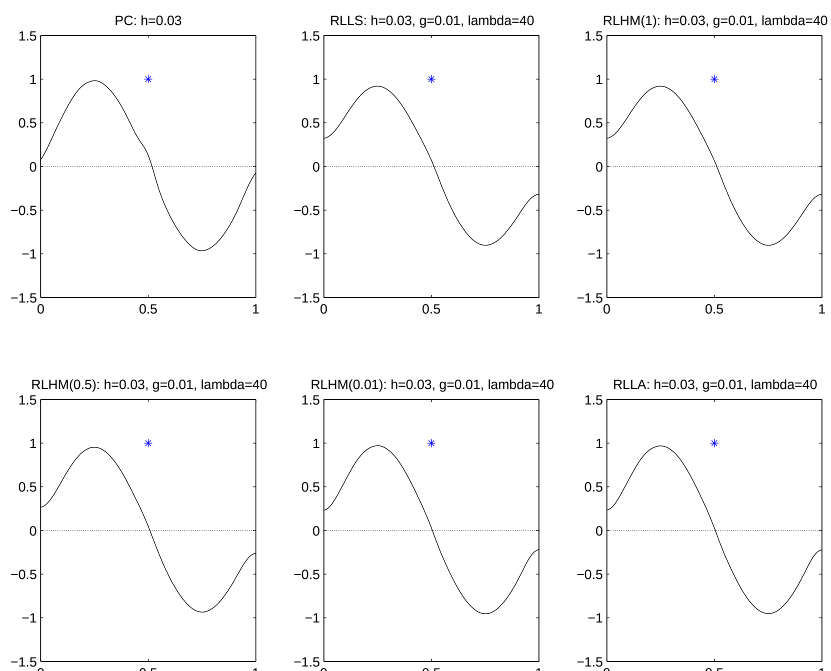

Figure 7.28: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 40)$.

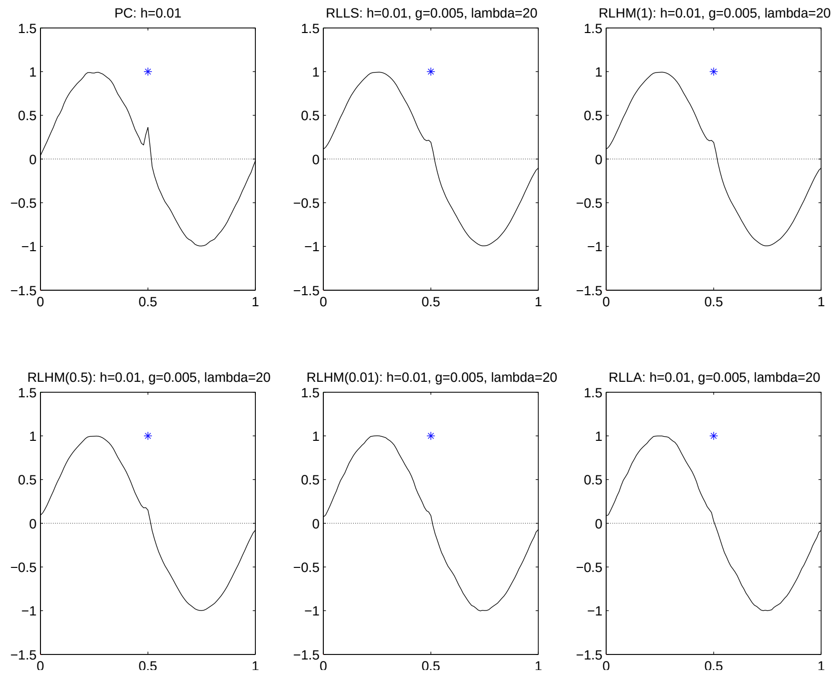

Figure 7.29: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 20)$.

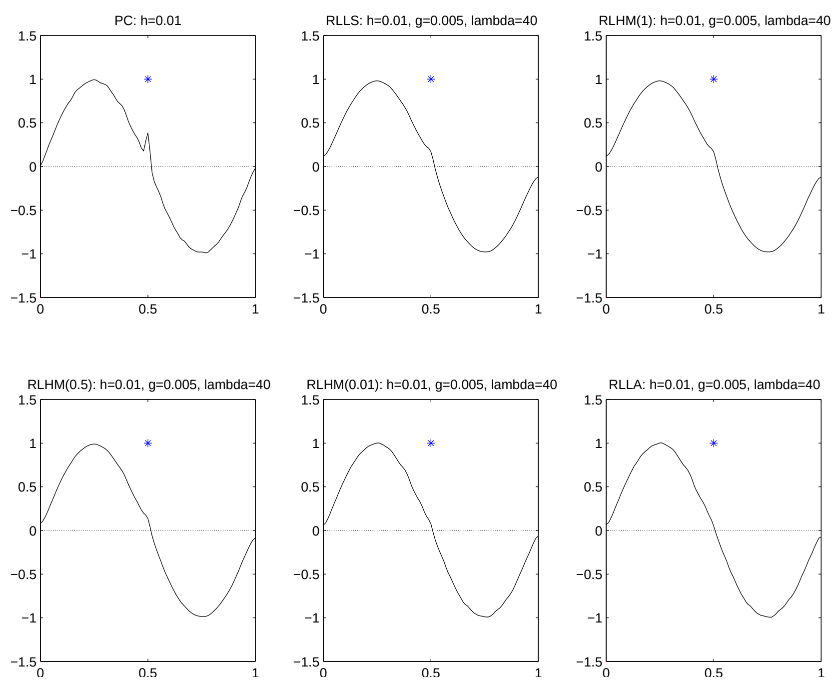

Figure 7.30: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 40)$.

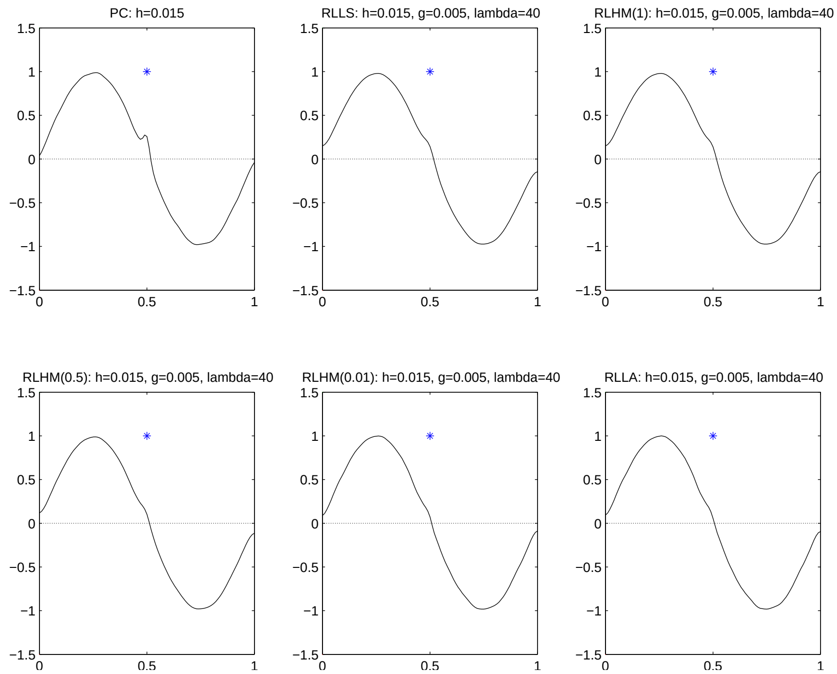

Figure 7.31: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 40)$.

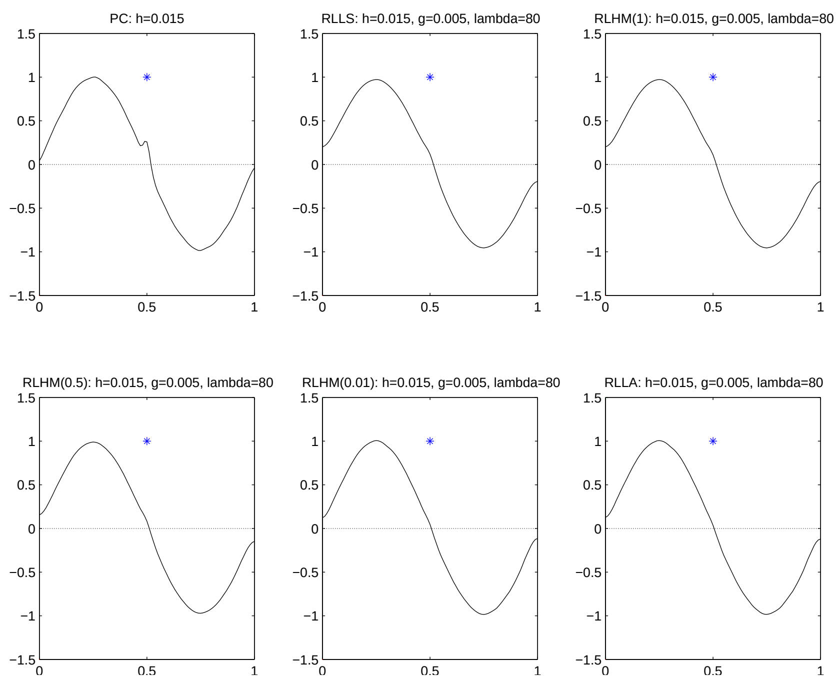

Figure 7.32: Single outlier with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 80)$.

From the previous tables and figures we note the following,

1) The ermse of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were approximately equal.

2) The emad of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were approximately equal.

⋆ The values of the maxad here are not significant since the outlier is included in the calculation.

⇒ $\hat{\mathbf{u}}_{\mathbf{LS}}$ is quit close to $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ here.

3) The estimate with the lowest ermse was $\hat{\mathbf{u}}_{\mathbf{PC}}$.

4) The estimate with the lowest emad was $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$. Other good estimates are $\hat{\mathbf{u}}_{\mathbf{LA}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$.

⋆ If we look at the graphs of this case we can see that the ermse is not a good measure of quality. The good measure here is the emad, which is more robust against outliers.

⇒ $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ is the best estimate here. Other good estimates are $\hat{\mathbf{u}}_{\mathbf{LA}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$.

⋆ The performance of the estimates at the outlier was best for $\hat{\mathbf{u}}_{\mathbf{LA}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

⋆ But from the graphs we can see that there is a bit of over resistance in the $\hat{\mathbf{u}}_{\mathbf{LA}}$.

7.6 Results 5: Two Outliers with Normal Error Terms

Now we look on the case of the two added outliers with normal error terms, i.e.

$$\mu(x) = \begin{cases} \sin(2\pi x), & x \neq \frac{1}{4}, \frac{1}{2}, \\ -1, & x = \frac{1}{4}, \\ 1, & x = \frac{1}{2}, \end{cases} \quad \text{and} \quad \mathcal{L}(\varepsilon_j) = \mathcal{N}(0, 0.2^2). \tag{7.4}$$

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.143261	0.143569	0.145061	0.147086	0.148404	0.151865
0.03	0.01	0.1	0.143164	0.143353	0.144625	0.146448	0.148022	0.151901
0.03	0.01	2	0.143108	0.144164	0.145367	0.146864	0.148011	0.151957
0.03	0.01	40	0.143313	0.152352	0.152356	0.150063	0.149499	0.149748
0.01	0.005	20	0.125126	0.137437	0.140406	0.142830	0.145690	0.148880
0.01	0.005	40	0.125316	0.140305	0.142497	0.143977	0.146476	0.147611
0.015	0.005	40	0.134864	0.142219	0.143960	0.145274	0.146865	0.147760
0.015	0.005	80	0.135035	0.145024	0.146211	0.146486	0.147515	0.148144

Table 7.13: Two outliers with normal error terms: ermse for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	0.070075	0.069070	0.061881	0.055351	0.050866	0.055169
0.03	0.01	0.1	0.070643	0.069096	0.062127	0.055250	0.050153	0.055490
0.03	0.01	2	0.069638	0.072035	0.064862	0.056200	0.050289	0.054856
0.03	0.01	40	0.072128	0.123526	0.116576	0.087804	0.069820	0.070570
0.01	0.005	20	0.047466	0.057616	0.053050	0.045178	0.042267	0.043798
0.01	0.005	40	0.046696	0.066049	0.060655	0.047626	0.041457	0.042654
0.015	0.005	40	0.054534	0.069030	0.062907	0.049932	0.043309	0.044961
0.015	0.005	80	0.055714	0.085515	0.079097	0.059691	0.048080	0.048785

Table 7.14: Two outliers with normal error terms: emad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

h	g	λ	$\hat{\mathbf{u}}_{\mathbf{PC}}$	$\hat{\mathbf{u}}_{\mathbf{LS}}$	$\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.50)}}$	$\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$	$\hat{\mathbf{u}}_{\mathbf{LA}}$
0.03	0.01	0	1.715730	1.722887	1.837174	1.913304	1.951053	2.001648
0.03	0.01	0.1	1.706654	1.714863	1.825518	1.901139	1.948587	2.015806
0.03	0.01	2	1.708818	1.733983	1.837511	1.905009	1.948729	2.024419
0.03	0.01	40	1.718885	1.777737	1.833276	1.895933	1.933945	1.936566
0.01	0.005	20	1.193683	1.557656	1.725950	1.827757	1.903856	1.983818
0.01	0.005	40	1.202828	1.633212	1.771623	1.853353	1.926373	1.952365
0.015	0.005	40	1.457113	1.682031	1.804032	1.877666	1.933273	1.953382
0.015	0.005	80	1.473244	1.733758	1.827582	1.891827	1.935679	1.949680

Table 7.15: Two outliers with normal error terms: maxad for $\hat{\mathbf{u}}_{\mathbf{PC}}$, $\hat{\mathbf{u}}_{\mathbf{LS}}$, $\hat{\mathbf{u}}_{\mathbf{HM(c)}}$ and $\hat{\mathbf{u}}_{\mathbf{LA}}$.

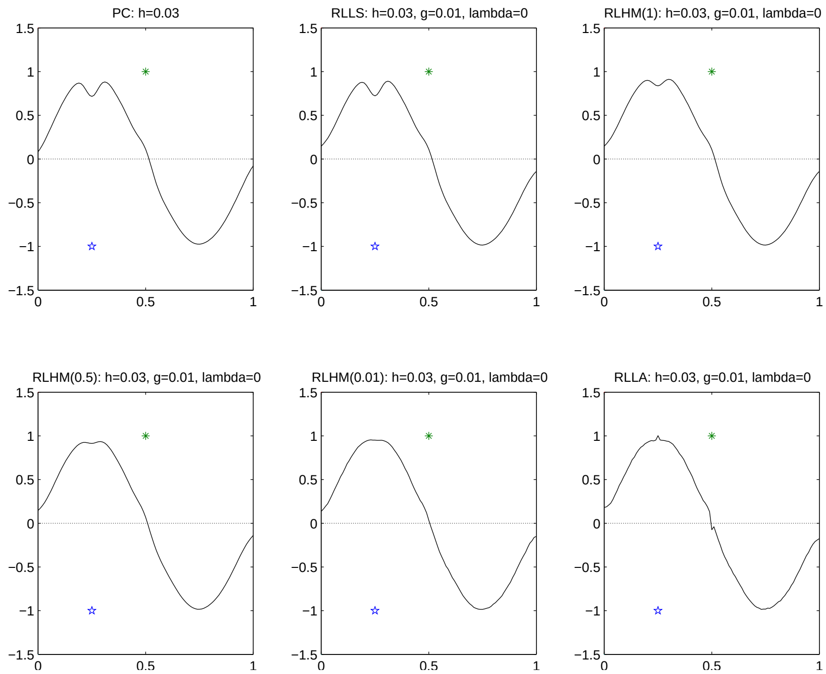

Figure 7.33: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0)$.

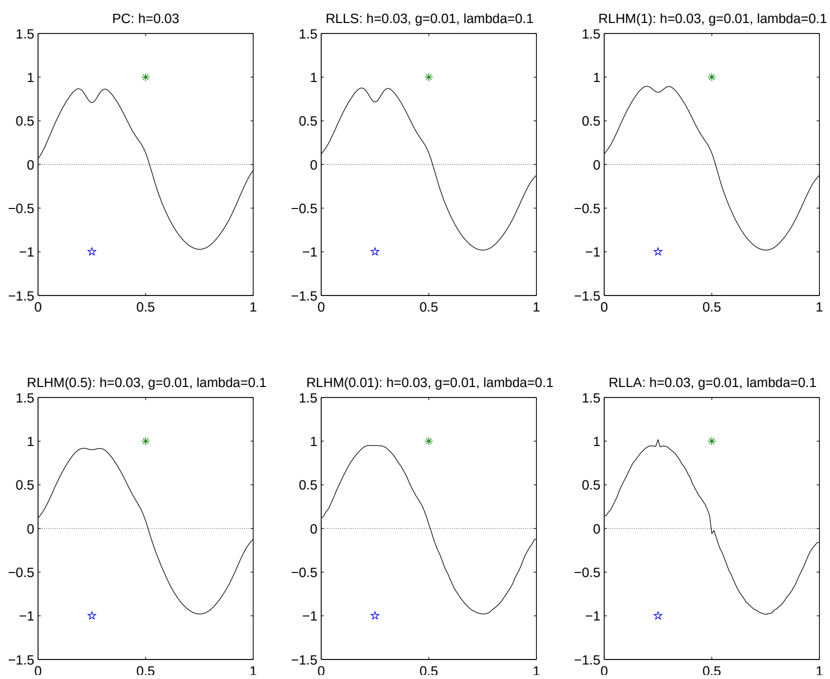

Figure 7.34: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 0.1)$.

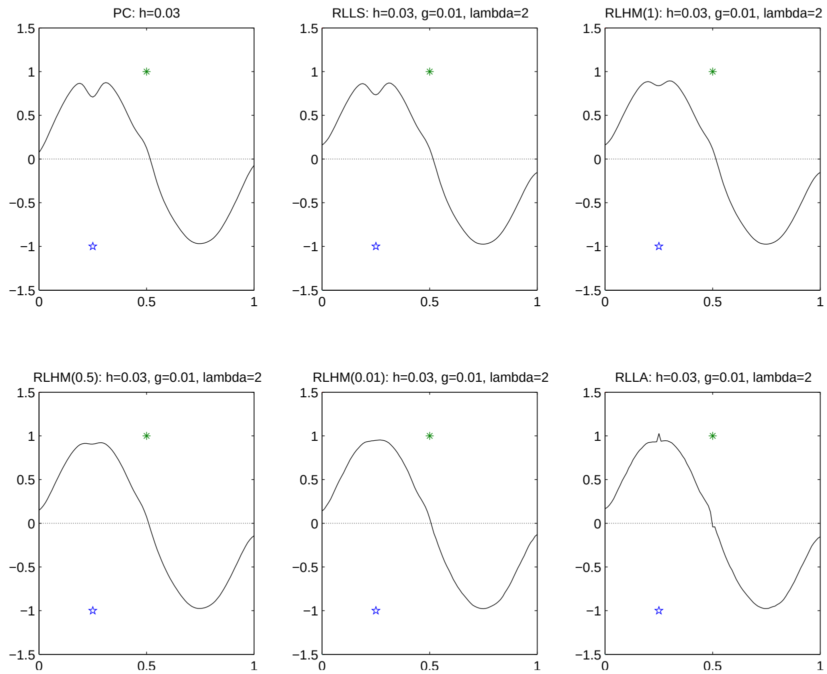

Figure 7.35: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 2)$.

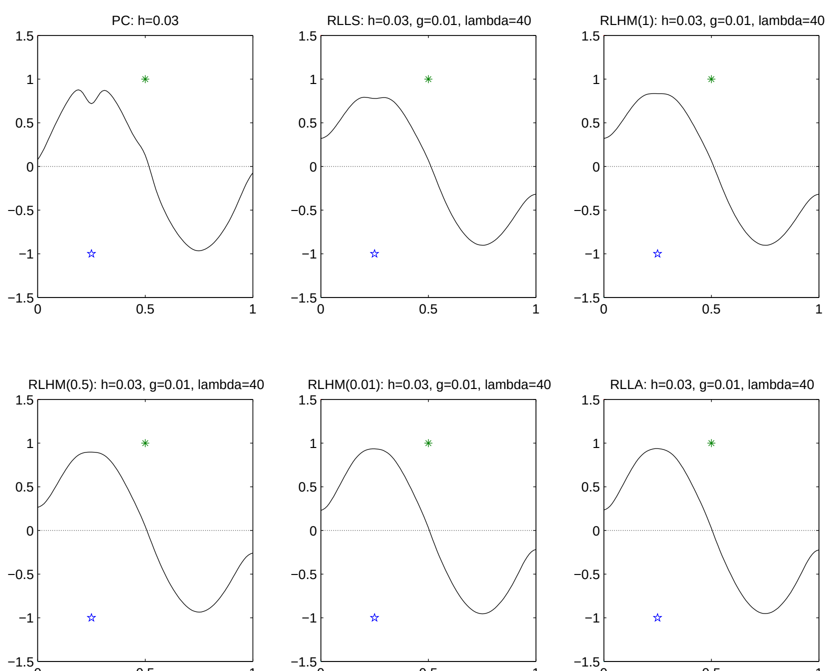

Figure 7.36: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.03, 0.01, 40)$.

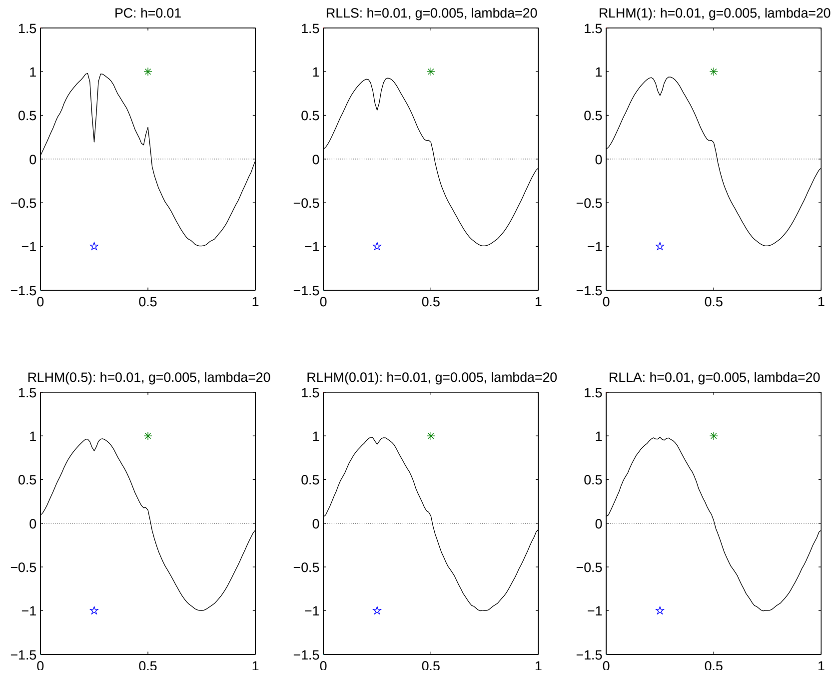

Figure 7.37: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 20)$.

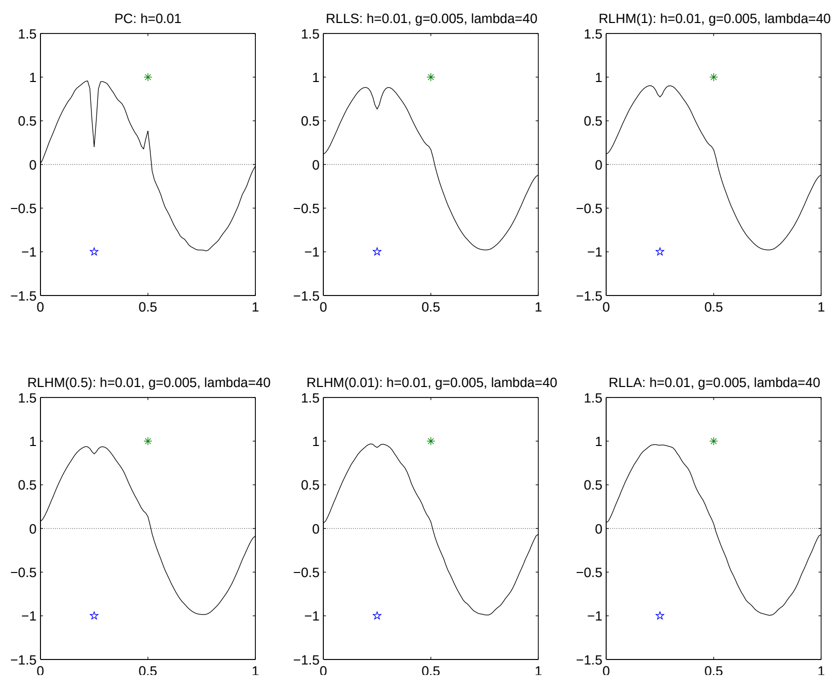

Figure 7.38: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.01, 0.005, 40)$.

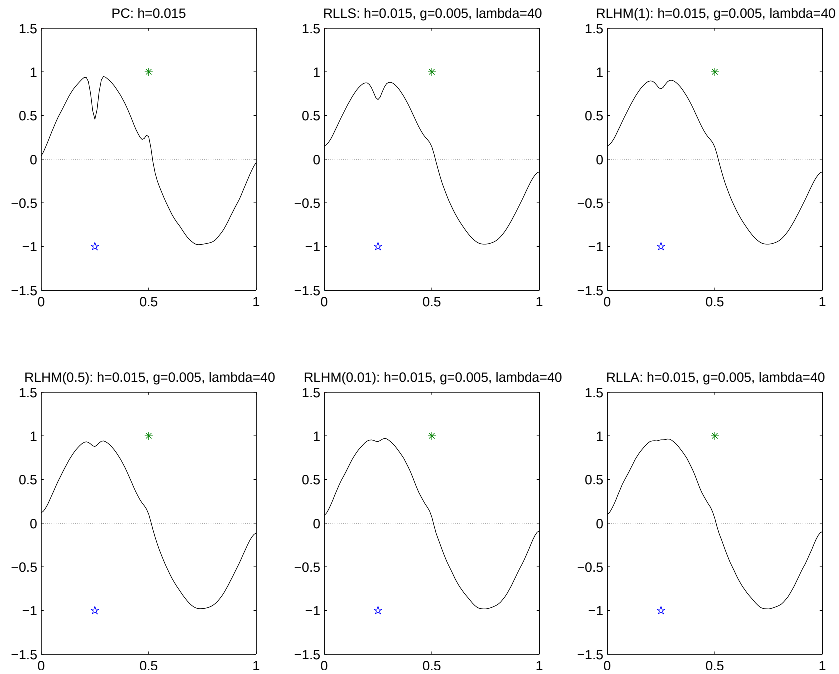

Figure 7.39: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 40)$.

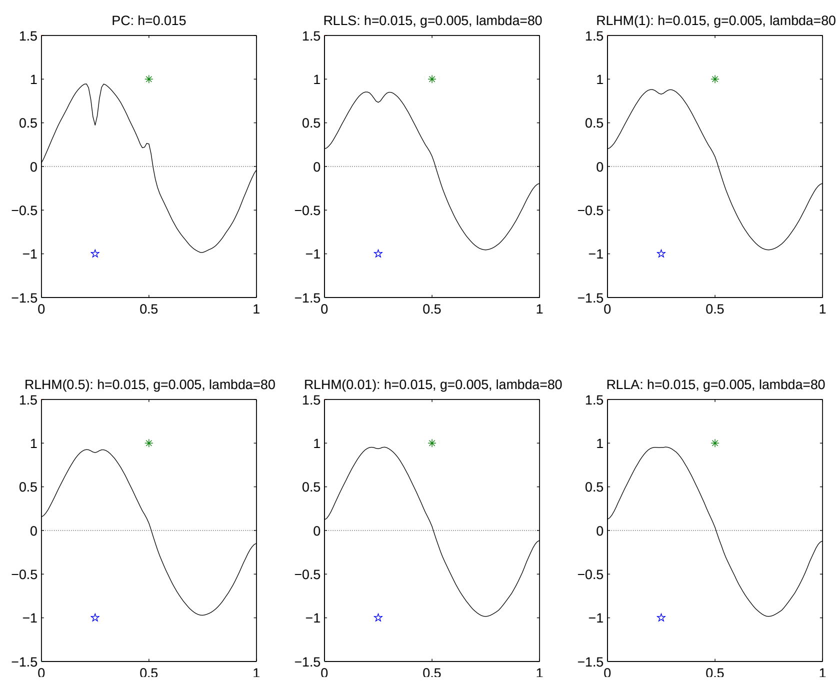

Figure 7.40: Two outliers with normal error terms: $(N, M, h, g, \lambda) = (100, 100, 0.015, 0.005, 80)$.

From the previous tables and figures we note the following,

1) The ermse of both $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were roughly equal.

2) The emad of $\hat{\mathbf{u}}_{\mathbf{LS}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ were significantly different.

⋆ The values of the maxad here are not significant since the outliers were included in the calculation.

⇒ $\hat{\mathbf{u}}_{\mathbf{LS}}$ is different from $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$. $\hat{\mathbf{u}}_{\mathbf{HM(1.00)}}$ is better here.

3) The estimate with the lowest ermse was $\hat{\mathbf{u}}_{\mathbf{PC}}$ then $\hat{\mathbf{u}}_{\mathbf{LS}}$.

4) The estimate with the lowest emad was $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ then $\hat{\mathbf{u}}_{\mathbf{LA}}$.

⋆ Again, if we look at the graphs of this case we can see that the ermse is not a good measure of quality. The good measure here is the emad.

⇒ $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ is the best estimate here. Another good estimate is $\hat{\mathbf{u}}_{\mathbf{LA}}$.

⋆ The performance of the estimates at the outlier was best for $\hat{\mathbf{u}}_{\mathbf{LA}}$ and $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$.

⋆ But from the graphs we can see that there is a bit of over resistance in the $\hat{\mathbf{u}}_{\mathbf{LA}}$ and that $\hat{\mathbf{u}}_{\mathbf{HM(0.01)}}$ is slightly affected by the extreme outlier.

Bibliography

[1] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[2] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1987.

[3] G. Casella and R. L. Berger. *Statistical Inference*. Second edition, Duxbury, Pacific Grove, California, 2002.

[4] C. K. Chu, I. K. Glad, F. Godtliebsen, and J. S. Marron. Edge preserving smoothers for image processing (with discussion). *Journal of the American Statistical Association*, 93(442):526–556, 1998.

[5] S. Didas, P. Mrázek, and J. Weickert. Energy-based image simplification with nonlocal data and smoothness terms. In A. Iske and J. Levesley, editors, *Algorithms for Approximation*. Springer, Heidelberg, 2006.

[6] D. Estep. *Practical Analysis in One Variable*. Springer-Verlag, New York, 2002.

[7] J. Fan, T.-C. Hu, and Y. K. Truong. Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, 21(4):433–446, 1994.

[8] W. Fleming. *Functions of Several Variables*. Second edition, Springer-Verlag, New York, 1977.

[9] J. Franke. Uniform consistency of regularized local least-squares estimates and the bootstrap. Personal communication, Department of Mathematics, University of Kaiserslautern, 2006.

[10] J. Franke, J. Tadjuidje Kamgaing, S. Didas, and J. Weickert. Some asymptotics for local least-squares regression with regularization. Report in Wirtschaftsmathematik, No. 107, Department of Mathematics, University of Kaiserslautern, 2008.

[11] F. R. Hampel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.

[12] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.

[13] W. Härdle. Robust regression function estimation. *Journal of Multivariate Analysis*, 14:169–180, 1984.

[14] W. Härdle and T. Gasser. Robust non-parametric function fitting. *Journal of the Royal Statistical Society: Series B*, 46(1):42–51, 1984.

[15] W. Härdle and S. Luckhaus. Uniform consistency of a class of regression function estimators. *The Annals of Statistics*, 12(2):612–623, 1984.

[16] P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[17] P. J. Huber. Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5):799–821, 1973.

[18] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[19] J. K. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific, Singapore, 2001.

[20] J. Jacod and P. Protter. *Probability Essentials*. Second edition (corrected second printing), Springer-Verlag, Berlin and Heidelberg, 2004.

[21] C. Jennen-Steinmetz and T. Gasser. A unifying approach to nonparametric regression estimation. *Journal of the American Statistical Association*, 83(404):1084–1089, 1988.

[22] T. C. M. Lee and H.-S. Oh. Robust penalized regression spline fitting with application to additive mixed modeling. *Computational Statistics*, 22:159–171, 2007.

[23] I. J. Maddox. *Elements of Functional Analysis*. Second edition, Cambridge University Press, Cambridge, 1988.

[24] R. A. Maronna, R. D. Martin, and V. J. Yohai. *Robust Statistics: Theory and Methods*. Reprinted with corrections. Wiley, Chichester, 2006.

[25] 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.

[26] A. Pinkus and S. Zafrany. *Fourier Series and Integral Transforms*. Cambridge University Press, Cambridge, 1997.

[27] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[28] R. T. Seeley. *Calculus of Several Variables: An Introduction*. Scott, Foresman and Co., Glenview, Illinois, 1970.

[29] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.

[30] B. W. Silverman. Spline smoothing: the equivalent variable kernel method. *The Annals of Statistics*, 12(3):898–916, 1984.

[31] R. G. Staudte and S. J. Sheather. *Robust Estimation and Testing*. Wiley, New York, 1990.

[32] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Second edition, Springer-Verlag, New York, 1993.

[33] W. Stützle and Y. Mittal. Some comments on the asymptotic behavior of robust smoothers. In T. Gasser and M. Rosenblatt, editors, *Smoothing Techniques for Curve Estimation, Proceedings of a Workshop held in Heidelberg, April 24, 1979*. Lecture Notes in Mathematics, Vol. 757, Springer-Verlag, Berlin, Heidelberg, and New York, 1979.

[34] M. Taniguchi and Y. Kakizawa. *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York, 2000.

[35] C. Udrişte. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers, Dordrecht, 1994.

[36] A. W. van der Vaart. *Asymptotic Statistics*. Eighth printing, Cambridge University Press, Cambridge, 2007.

[37] J. L. Wals and W. E. Sewell. Note on degree of approximation to an integral by Riemann sums. *The American Mathematical Monthly*, 44(3):155–160, 1937.

[38] M. P. Wand and M. C. Jones. *Kernel Smoothing*. First edition, Chapman and Hall, London, 1995.

[39] D. Werner. *Funktionalanalysis (Functional Analysis, in German)*. Sixth corrected edition, Springer-Verlag, Berlin and Heidelberg, 2007.

Curriculum Vitae

1999	High School Certificate, Palestine
2000–2004	Studies in Mathematics and Education at Birzeit University, Palestine
2004	Bachelor of Science in Mathematics
2004	Teaching Diploma
2005–2007	Studies in Mathematics at the University of Kaiserslautern
2007	Master of Science in Mathematics with the thesis *M–smoothers with regularization*
2007–2010	PhD student at the Mathematics Department of the University of Kaiserslautern