Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Michal Pešta

Izotonická regrese v Sobolevových prostorech

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce:

Mgr. Zdeněk Hlávka, Ph.D.

Studijní program:

Matematika

Studijní obor:

Pravděpodobnost, matematická statistika a ekonometrie

Studijní plán:

Matematická statistika

Charles University in Prague

Faculty of Mathematics and Physics

MASTER THESIS

Michal Pešta

Isotonic Regression in Sobolev Spaces

Department of Probability and Mathematical Statistics

Supervisor of master thesis:

Mgr. Zdeněk Hlávka, Ph.D.

Programme of Study:

Mathematics

Field of Study:

Probability, Mathematical Statistics and Econometrics

Curriculum:

Mathematical Statistics

Poděkování

Rád bych poděkoval Mgr. Zdeňku Hlávkovi, Ph.D. za jeho neocenitelnou pomoc při psaní této práce. Doufám, že mi pomůže i s psaním nějakých dalších prací ...

Také bych chtěl poděkovat všem učitelům Univerzity Karlovy v Praze, kteří mě uvedli do světa velké matematiky a matematické statistiky.

Čestné prohlášení

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 7. dubna, 2006

Michal Pešta

vlastnoruční podpis

Acknowledgements

I would like to thank Mgr. Zdeněk Hlávka, Ph.D. for his most valuable help in writing this thesis. Hope he will help me write another thesis sometimes later . . .

I would also like to thank to all my teachers from Charles University in Prague, Czech Republic for introducing me into the world of higher mathematics and mathematical statistics.

Statement of Honesty

I hereby declare that I have written this master thesis separately, independently and entirely with using quoted resources. I agree that the University Library shall make it available to borrowers under rules of the Library.

Prague, April 7, 2006

Michal Pešta

Signature

Contents

Abstract

Název práce: Izotonická regrese v Sobolevových prostorech
Autor: Michal Pešta
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Mgr. Zdeněk Hlávka, Ph.D.
e-mail vedoucího: `hlavka@karlin.mff.cuni.cz`
Abstrakt: Uvažme třídu neparametrických odhadů pro regresní modely založené na metodě nejmenších čtverců přes množiny dostatečně hladkých funkcí. Nejmenší čtverce dovolují uložení dodatečného omezení, izotonie, na neparametrické regresní odhady a jejich následné testování.

Odhady probíhají přes koule funkcí, které jsou prvky vhodných Sobolevových prostorů. Sobolevovy prostory jsou speciální typ Hilbertových prostorů, které umožňují projekci vzhledem k nejmenší čtvercům. Hilbertovskost nám umožňuje dělat projekci a tedy rozložit prostor do navzájem kolmých doplňků. Pak převedeme problém hledání nejlépe aproximující funkce v prostoru nekonečné dimenze na konečně-dimenzionální optimalizační problém.

Dokážeme, že koule funkcí v Sobolevových prostorech je omezená a má omezené i derivace vyššího řádu. To nám dovoluje odhadovat přes bohatou množinu funkcí s dostatečně malou metrickou entropií a použít zákony velkých čísel a centrální limitní věty.
Klíčová slova: izotonická regrese, Sobolevovy prostory, neparametrická, monotonie

Title: Isotonic Regression in Sobolev Spaces
Author: Michal Pešta
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Zdeněk Hlávka, Ph.D.
Supervisor's e-mail address: `hlavka@karlin.mff.cuni.cz`
Abstract: We propose a class of nonparametric estimators for the regression models based on least squares over the sets of sufficiently smooth functions. Least squares permit the imposition of additional constraint—isotonia—on nonparametric regression estimation and testing of this constraint.

The estimation takes place over the balls of functions which are elements of a suitable Sobolev space—special types of Hilbert spaces that facilitate calculation of the least squares projection. The Hilbertness is allowing us to take projections and hence to decompose spaces into mutually orthogonal complements. Then we transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimization problem.

We prove that the balls of functions in Sobolev space are bounded and have bounded higher order derivatives. It permits us to estimate over rich set of functions with sufficiently low metric entropy and apply Laws of Large Numbers and Central Limit Theorems.
Keywords: isotonic regression, Sobolev spaces, nonparametric, monotonicity

Dedication

This thesis is dedicated to my parents—Viera and Zdeněk—whose encouragement have meant to me so much during the pursuit of my graduate degree and the composition of the thesis.

Introduction

Motivation and Aim of Thesis

What is isotonia and Sobolev space? Let's try a heuristic description of the terms. Isotonia is non-negativity (non-positivity) of n-th derivative of function (e.g. monotonicity, convexity or concavity). Sobolev spaces are class of functions with smooth high order derivative. On the other hand, they are also spaces of integrable functions on a specific domain.

We think of the regression models based upon functions with specific features (quality). So we do not care about the formula of the regression function (nonparametric regression).

Main Steps

In first chapter we focus on Sobolev spaces and their main properties. We progressively define Sobolev norms, Sobolev spaces and afterwards Sobolev inner product. Declaration of the representors in Sobolev spaces provides us base elements for the functional representation. Finally we prove an important result for the boundedness of derivatives.

Second chapter concerns bare regression in Sobolev spaces. Basic and advanced models are defined. Very important theorem—transforming our problem from the infinite dimension into the finite one—is given. Optimizing algorithm and its properties are studied. We are also interested in Cross-Validation.

In third chapter we introduce isotonic regression. Our main imposition still remains the smoothness of the regression estimator. Completely different approach on the isotonic regression (but without smoothness demand) using Greatest Convex Minorant (GCM) can be found in Robertson et al. (1988). We extend our models by adding additional constraint—isotonia. Finally, we define various types of isotonia.

Fourth chapter surveys the asymptotic behavior of our estimators. Some prepositions for confidence intervals and tests of hypothesis are shown. Bootstrap techniques are also applied.

In the appendix we give a brief overview through the skeletons of algorithms, procedures and statistical computing described in the whole thesis. Parts of *R* source code are listed. The appendix provides lots of figures for better imagery. All the used important theorems and lemmas are also quoted in the end.

In the future, one can study and examine the relation between isotonic regression in Sobolev spaces and the location of zeros of regression functions (and their derivatives). We may subsequently dwell on its estimation. This could help us to estimate extremes of various functions.

Chapter 1

Sobolev Spaces

In this chapter, we give a brief overview on basic results of the theory of the Sobolev spaces. We assemble and prove necessary preliminaries and theorems for statistical regression in these spaces.

The crux of this chapter lies in Theorem 1.4.1 (Representors in Sobolev Space) from Bos and Yatchew (1997). I have continued in examining representors' properties and proved Theorem 1.4.2, which provides the way of construction of the representors and their exact form.

1.1 Impact of Lebesgue Integration Theory

We need to summarize some basic definitions and corollaries from Lebesgue integration theory.

Definition 1.1.1 (Domain). A connected Lebesgue-measurable (open or closed) subset Ω of an Euclidean space $\mathbb{R}^q$ with non-empty interior is called a domain.

From now on, we will denote Ω as a domain.

Definition 1.1.2 (Lebesgue Space). Consider a real-valued function on a given Lebesgue-measurable domain. Simply $f : \Omega \to \mathbb{R}$, $\Omega \in \mathfrak{M}_q(\lambda_q)$. The Lebesgue integral of function f is $\int_\Omega f(\mathbb{x})\mathrm{d}\lambda_q(\mathbb{x}) \equiv \int_\Omega f(\mathbb{x})\mathrm{d}\mathbb{x}$. Let

$$\|f\|_{L_p(\Omega)} := \begin{cases} \left(\int_\Omega f^p(\mathbb{x})\mathrm{d}\mathbb{x}\right)^{1/p} & \text{for } 1 \le p < \infty, \\ \inf\left\{C \ge 0 : |f| \le C \text{ a.e.}\right\} & \text{for } p = \infty. \end{cases} \tag{1.1.1}$$

We define a Lebesgue space by $L_p(\Omega) := \left\{f : \|f\|_{L_p(\Omega)} < \infty\right\}$, $1 \le p \le \infty$.

Definition 1.1.3 (Complete Space). A complete space Ξ is a metric space such that every Cauchy sequence $\{\xi_i\}_{i\in\mathbb{N}} \in \Xi$ has a limit $\xi \in \Xi$.

Completeness can be defined for a normed linear space, where the metric is induced by the respective norm.

Lemma 1.1.1. *Normed linear space is complete, if each convergent sequence $\{\xi_i\}_{i\in\mathbb{N}} \in \Xi$ has a limit $\xi \in \Xi$.*

Proof. Trivial. □

Definition 1.1.4 (Banach Space). Normed linear space $(\Xi, \|\cdot\|)$, complete with respect to the norm $\|\cdot\|$, is called a Banach space.

Corollary 1.1.2. *$L_p(\Omega)$, $1 \le p \le \infty$ is a Banach space.*

Proof. See Lukeš and Malý (1995). □

1.2 Sobolev Norms

We consider function $f : \Omega \to \mathbb{R}$ and denote by

$$D^{\boldsymbol{\alpha}} f(\mathbb{x}) := \frac{\partial^{|\boldsymbol{\alpha}|_1} f(\mathbb{x})}{\partial x_1^{\alpha_1} \dots \partial x_q^{\alpha_q}} \tag{1.2.2}$$

its partial derivatives of order $|\boldsymbol{\alpha}|_1$ for $\mathbb{x} \in \operatorname{int}(\Omega)(\equiv \Omega^\circ := \overline{\Omega}\backslash\partial\Omega)$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)' \in \mathbb{N}_0^q$ is a multiindex of the modulus $|\boldsymbol{\alpha}|_1 = \sum_{i=1}^q \alpha_i$.

Definition 1.2.1 (Spaces of Continuously Differentiable Functions). Let $m \in \mathbb{N}_0$. We define $\mathcal{C}^m(\Omega)$ space of m-times continuously differentiable scalar functions upon bounded domain Ω. Simply

$$\mathcal{C}^m(\Omega) := \left\{ f : \Omega \to \mathbb{R} \middle| D^{\boldsymbol{\alpha}} f \in \mathcal{C}^0(\Omega), |\boldsymbol{\alpha}|_\infty \le m \right\}, \tag{1.2.3}$$

where $\mathcal{C}^0(\Omega) \equiv \left\{ f : \Omega \to \mathbb{R} \middle| f \text{ continuous on } \Omega \right\}$ and $|\boldsymbol{\alpha}|_\infty = \max_{i=1,\dots,q} |\alpha_i|$.

In the space $\mathcal{C}^m(\Omega)$ we can define various norms

$$\|f\|_{\infty,\infty,\mathcal{C}^m(\Omega)} := \sum_{|\boldsymbol{\alpha}|_\infty \le m} \sup_{\mathbb{x}\in\operatorname{int}(\Omega)} |D^{\boldsymbol{\alpha}} f(\mathbb{x})|, \tag{1.2.4a}$$

$$\|f\|_{1,\infty,\mathcal{C}^m(\Omega)} := \max_{|\boldsymbol{\alpha}|_\infty \le m} \sup_{\mathbb{x}\in\operatorname{int}(\Omega)} |D^{\boldsymbol{\alpha}} f(\mathbb{x})| \tag{1.2.4b}$$

and then we also obtain corresponding normed linear spaces:

$$\left(\mathcal{C}^m(\Omega), \|\cdot\|_{\infty,\infty,\mathcal{C}^m(\Omega)}\right) \text{ and } \left(\mathcal{C}^m(\Omega), \|\cdot\|_{1,\infty,\mathcal{C}^m(\Omega)}\right). \tag{1.2.5}$$

Remark 1.2.1. $(\mathcal{C}^m(\Omega), \|\cdot\|_{\infty,\infty,\mathcal{C}^m(\Omega)})$ and $(\mathcal{C}^m(\Omega), \|\cdot\|_{1,\infty,\mathcal{C}^m(\Omega)})$ are both Banach spaces, because they are complete.

Definition 1.2.2 (Sobolev Norm). Let $f \in \mathcal{C}^m(\Omega) \cap L_p(\Omega)$. We introduce a Sobolev norm[1] $\|\cdot\|_{p,Sob,m}$:

$$\|f\|_{p,Sob,m} := \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \int_\Omega \left| D^{\boldsymbol{\alpha}} f(\mathbb{x}) \right|^p \mathrm{d}\mathbb{x} \right\}^{1/p}. \tag{1.2.6}$$

Proof. The correctness of Definition 1.2.2 needs to be shown. Let $f, g \in \mathcal{C}^m(\Omega) \cap L_p(\Omega)$, then the triangle inequality for the p-norms on $L_p(\Omega)$ and $l_p(\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}|_\infty \leq m\})$ implies

$$\|f + g\|_{p,Sob,m} = \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} f + D^{\boldsymbol{\alpha}} g\|^p_{L_p(\Omega)} \right\}^{1/p} \tag{1.2.7}$$

$$\leq \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \left[\|D^{\boldsymbol{\alpha}} f\|^p_{L_p(\Omega)} + \|D^{\boldsymbol{\alpha}} g\|^p_{L_p(\Omega)} \right] \right\}^{1/p} \tag{1.2.8}$$

$$\leq \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} f\|^p_{L_p(\Omega)} \right\}^{1/p} + \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} g\|^p_{L_p(\Omega)} \right\}^{1/p} \tag{1.2.9}$$

$$= \|f\|_{p,Sob,m} + \|g\|_{p,Sob,m}. \tag{1.2.10}$$

□

Definition 1.2.3 (Sobolev Space). A Sobolev space $\mathcal{W}^m_p(\Omega)$ is the completion of intersection of space $\mathcal{C}^m(\Omega)$ and space $L_p(\Omega)$ with respect to the Sobolev norm $\|\cdot\|_{p,Sob,m}$.

Remark 1.2.2. $\mathcal{C}^m(\Omega) \cap L_p(\Omega)$ is dense in $\mathcal{W}^m_p(\Omega)$ according to $\|\cdot\|_{p,Sob,m}$.

Definition 1.2.4 (Weak Derivative). Let $f \in L_1(\Omega)$ and $\boldsymbol{\alpha} \in \mathbb{N}^q_0$. The function f is said to have a weak derivative $D^{\boldsymbol{\alpha}}_w f$ if there exists a function $g \in L_1(\Omega)$ such that

$$\int_\Omega f(\mathbb{x}) D^{\boldsymbol{\alpha}} h(\mathbb{x}) \mathrm{d}\mathbb{x} = (-1)^{|\boldsymbol{\alpha}|_1} \int_\Omega g(\mathbb{x}) h(\mathbb{x}) \mathrm{d}\mathbb{x}, \quad h \in \mathcal{C}^\infty(\Omega). \tag{1.2.11}$$

Then we set $D^{\boldsymbol{\alpha}}_w f := g$.

[1] We could formally write $\|\cdot\|_{p,\infty,Sob,m}$, because the multiindex of modulus $|\boldsymbol{\alpha}|_\infty = \max_{i=1,\ldots,q} \alpha_i$ is taken with respect to maxim-norm.

Example 1.2.1. Let $q = 1$ and $\Omega = (-1, +1)$ The function $f(x) = |x|$, $x \in \Omega$ is not differentiable in the classical sense. However, it admits a weak derivative $D_w^1 f$ given by

$$D_w^1 f(x) = \begin{cases} -1\,, & x < 0, \\ whatever\,, & x = 0, \\ +1\,, & x > 0. \end{cases} \tag{1.2.12}$$

To show this, we break the interval $(-1, +1)$ into the two parts in which f is smooth and then integrate by parts.

Remark 1.2.3 *(General Definition of Sobolev Space).* A Sobolev space can be also defined in more general way:

$$\mathcal{W}_p^m(\Omega) := \{f \in L^p(\Omega) | D_w^{\boldsymbol{\alpha}} f \in L^p(\Omega), |\boldsymbol{\alpha}|_\infty \le m\}\,. \tag{1.2.13}$$

1.3 Hilbert Spaces

Definition 1.3.1 (Hilbert Space). A Hilbert space Ξ is a vector space with an inner product $\langle \cdot, \cdot \rangle$ such that the norm defined by $|\xi| = \sqrt{\langle \xi, \xi \rangle}$, $\xi \in \Xi$ turns Ξ into a complete metric space.

Corollary 1.3.1. *A Hilbert space is always a Banach space.*

Proof. Trivial. □

Example 1.3.1. $L_2(\Omega)$ is a Hilbert space with respect to the inner product $\langle f, g \rangle_{L_2(\Omega)} := \int_\Omega f(\mathbb{x}) g(\mathbb{x}) \mathrm{d}\mathbb{x}$.

Lemma 1.3.2. *If Δ is a subspace of Hilbert space Ξ, then it is also a Hilbert space.*

Proof. By definition Δ is closed under the norm $\|\cdot\|$ and that implies the completeness of $(\Delta, \|\cdot\|)$. □

Definition 1.3.2 (Sobolev Inner Product). Let $f, g \in \mathcal{W}_2^m(\Omega)$. We introduce a Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$:

$$\langle f, g \rangle_{Sob,m} := \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_\Omega D^{\boldsymbol{\alpha}} f(\mathbb{x}) D^{\boldsymbol{\alpha}} g(\mathbb{x}) \mathrm{d}\mathbb{x}. \tag{1.3.14}$$

Notation 1.3.1. We denote the Sobolev norm $\|\cdot\|_{2,Sob,m} := \|\cdot\|_{Sob,m}$ for simplicity.

Conjecture 1.3.3. The correctness of Definition 1.3.2 is guaranteed by the denseness of the space $\mathcal{C}^m(\Omega) \cap L_2(\Omega)$ in $\mathcal{W}_2^m(\Omega)$ (see Remark 1.2.2).

Remark 1.3.1. The Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ induces in $\mathcal{W}_2^m(\Omega)$ the Sobolev norm $\|\cdot\|_{2,Sob,m}$.

Proof. Each element of $\mathcal{W}_2^m(\Omega)$ is differentiable almost everywhere as many times as we need according to Remark 1.2.2. □

Definition 1.3.4. We denote the Sobolev space $\mathcal{H}^m(\Omega) := \mathcal{W}_2^m(\Omega)$.

Theorem 1.3.3. *$\mathcal{H}^m(\Omega)$ is a Hilbert space.*

Proof. It is straightforward to verify that $\mathcal{H}^m(\Omega)$ is a normed linear space. It is also complete by construction, so it is a Banach space. The inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ has been defined on $\mathcal{H}^m(\Omega)$, so it is a Hilbert space. □

The theory of the Sobolev spaces is enormous and more general than we have already presented. But our simplified theory that concerned the Sobolev spaces will be enough for our statistical needs. More information about the Sobolev spaces can be found in Adams (1975) or Maz'ja (1985).

1.4 Representors in Sobolev Space

$\mathcal{H}^m(\Omega)$ is a Hilbert space, so we can express $\mathcal{H}^m(\Omega)$ as a direct sum of subspaces that are orthogonal to each other and take the projections of elements of $\mathcal{H}^m(\Omega)$ into its subspaces. This property is very important in the regression.

Remark 1.4.1. From now on, let suppose $m \in \mathbb{N}$.

Notation 1.4.1. $\mathcal{Q}^q$ denotes closed unit cube in $\mathbb{R}^q$.

Remark 1.4.2. $\mathcal{Q}^q$ is a closed bounded domain.

Remark 1.4.3. Continuous function on a closed bounded subset $\Omega \subset \mathbb{R}^q$ is always Lebesgue-integrable. Hence $\mathcal{H}^m = \widehat{\mathcal{C}^m(\Omega) \cap L_p(\Omega)}^{\|\cdot\|_{Sob,m}} \equiv \mathcal{C}^m(\Omega)$, where $\widehat{\Xi}^{\|\cdot\|}$ denotes the completion of space Ξ with respect to the norm $\|\cdot\|$.

Theorem 1.4.1 (Representors in Sobolev Space). *For all $f \in \mathcal{H}^m(\mathcal{Q}^q)$, $\mathrm{a} \in \mathcal{Q}^q$ and $\mathrm{w} \in \mathbb{N}_0^q$, $|\mathrm{w}|_\infty \leq m-1$, there exists a function $\psi_{\mathrm{a}}^{\mathrm{w}}(\mathrm{x}) \in \mathcal{H}^m(\mathcal{Q}^q)$, s.t.*

$$\left\langle \psi_{\mathrm{a}}^{\mathrm{w}}, f \right\rangle_{Sob,m} = D^{\mathrm{w}} f(\mathrm{a}). \tag{1.4.15}$$

$\psi_{\mathbb{a}}^{\mathbb{w}}$ *is called a representor at the point* $\mathbb{a}$ *with the rank* $\mathbb{w}$. *Furthermore,* $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) = \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i)$ *for all* $\mathbb{x} \in \mathcal{Q}^q$*, where* $\psi_{a_i}^{w_i}(\cdot)$ *is the representor in the Sobolev space of functions of one variable on* $\mathcal{Q}^1$ *with inner product*

$$\begin{aligned} \frac{\partial^{w_i} f(\mathbb{a})}{dx_i^{w_i}} &= \left\langle \psi_{a_i}^{w_i}, f(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_q) \right\rangle_{Sob,m} \\ &= \sum_{\alpha=0}^{m} \int_{\mathcal{Q}^1} \frac{d^\alpha \psi_{a_i}^{w_i}(x_i)}{dx_i^\alpha} \frac{d^\alpha f(\mathbb{x})}{dx_i^\alpha} dx_i. \end{aligned} \tag{1.4.16}$$

Remark 1.4.4. Proof of this Theorem has been inspired by Bos and Yatchew (1997). But we implement some modifications, because the proof in Bos and Yatchew (1997) seems to be incorrect (there is a solution of differential equation with specific boundary conditions, which is wrong).

Proof. We divide the proof into two steps.

i) Construction of a representor $\psi_a (\equiv \psi_a^0)$

For simplicity, let's set $\mathcal{Q}^1 \equiv [0,1]$. We know that for functions of one variable we have

$$\langle g, h \rangle_{Sob,m} = \sum_{k=0}^{m} \int_{\mathcal{Q}^1} g^{(k)}(x) h^{(k)}(x) \mathrm{d}x, \tag{1.4.17}$$

so all we need to do is to construct a representor

$$\psi_a \in \mathcal{H}^m [0,1] \text{ s.t. } \langle \psi_a, f \rangle_{Sob,m} = f(a) \tag{1.4.18}$$

for all $f \in \mathcal{H}^m [0,1]$. It suffices to demonstrate the result for all $f \in \mathcal{C}^{2m}$ because of the denseness of $\mathcal{C}^{2m}$ (see Remark 1.2.2), hence assume that $f \in \mathcal{C}^{2m}$. This representor will be of the form:

$$\psi_a(x) = \begin{cases} L_a(x) & 0 \le x \le a, \\ R_a(x) & a \le x \le 1, \end{cases} \tag{1.4.19}$$

where $L_a(x) \in \mathcal{C}^{2m} [0,a]$ and $R_a(x) \in \mathcal{C}^{2m} [a,1]$. As $\psi_a \in \mathcal{H}^m [0,1]$, it suffices that $L_a^{(k)}(a) = R_a^{(k)}(a)$, $0 \le k \le m-1$. We get:

$$f(a) = \langle \psi_a, f \rangle_{Sob,m} = \int_0^a \sum_{k=0}^{m} L_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x + \int_a^1 \sum_{k=0}^{m} R_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x. \tag{1.4.20}$$

Integrating by parts, we have:

$$\begin{aligned}
&\sum_{k=0}^{m}\int_0^a L_a^{(k)}(x)f^{(k)}(x)\mathrm{d}x\\
&=\sum_{k=0}^{m}\left\{\sum_{j=0}^{k-1}(-1)^j\,L_a^{(k+j)}(x)f^{(k-j-1)}(x)\Big|_0^a+(-1)^k\int_0^a L_a^{(2k)}(x)f(x)\mathrm{d}x\right\}\\
&=\sum_{k=0}^{m}\sum_{j=0}^{k-1}(-1)^j\,L_a^{(k+j)}(x)f^{(k-j-1)}(x)\Big|_0^a+\int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x.
\end{aligned}\tag{1.4.21}$$

Let's try to substitute $i=k-j-1$ and rewrite it:

$$\begin{aligned}
&\sum_{k=0}^{m}\int_0^a L_a^{(k)}(x)f^{(k)}(x)\mathrm{d}x\\
&=\sum_{k=1}^{m}\sum_{i=0}^{k-1}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(x)f^{(i)}(x)\Big|_0^a\\
&\quad+\int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x\\
&=\sum_{i=0}^{m-1}\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(x)f^{(i)}(x)\Big|_0^a\\
&\quad+\int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x\\
&=\sum_{i=0}^{m-1}f^{(i)}(a)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(a)\right\}\\
&\quad-\sum_{i=0}^{m-1}f^{(i)}(0)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(0)\right\}\\
&\quad+\int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x.
\end{aligned}\tag{1.4.22}$$

Similarly:

$$\begin{aligned}
&\sum_{k=0}^{m} \int_a^1 R_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x \\
&= \sum_{i=0}^{m-1} f^{(i)}(1) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-i-1)}(1) \right\} \\
&\quad - \sum_{i=0}^{m-1} f^{(i)}(a) \left\{ \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-i-1)}(a) \right\} \\
&\quad + \int_a^1 \left\{ \sum_{k=0}^{m} (-1)^k R_a^{(2k)}(x) \right\} f(x) \mathrm{d}x.
\end{aligned} \tag{1.4.23}$$

These two results hold for all $f(x) \in \mathcal{C}^m[0,1]$. Thus we require that both L_a and R_a are the solutions of the constant coefficient differential equation

$$\sum_{k=0}^{m} (-1)^k \varphi^{(2k)}(x) = 0. \tag{1.4.24}$$

Boundary conditions are obtained by equality of the functional values of $L_a^{(i)}(x)$ and $R_a^{(i)}(x)$ at the point a and the coefficient comparison[2] of $f^{(i)}(0)$, $f^{(i)}(1)$ and $f^{(i)}(a)$:

$$r_a \in \mathcal{H}^m[0,1] \;\Rightarrow\; L_a^{(i)}(a) = R_a^{(i)}(a) \quad \ldots \; 0 \le i \le m-1, \tag{1.4.25}$$

$$f^{(i)}(0) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-i-1)}(0) = 0 \quad \ldots \; 0 \le i \le m-1, \tag{1.4.26}$$

$$f^{(i)}(1) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-i-1)}(1) = 0 \quad \ldots \; 0 \le i \le m-1, \tag{1.4.27}$$

$$f^{(i)}(a) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} \left\{ L_a^{(2k-i-1)}(a) - R_a^{(2k-i-1)}(a) \right\} = 0 \quad \ldots \; 1 \le i \le m-1, \tag{1.4.28}$$

$$f(a) \bowtie 1 \;\Rightarrow\; \sum_{k=1}^{m} (-1)^{k-1} \left\{ L_a^{(2k-1)}(a) - R_a^{(2k-1)}(a) \right\} = 1; \tag{1.4.29}$$

[2] $\varrho \bowtie \varsigma$ denotes that ϱ has a coefficient ς in a specific equation.

together $m + m + m + (m - 1) + 1 = 4m$ boundary conditions. To obtain the general solution of this differential equation we need to find the roots of its characteristic polynomial

$$P_m(\lambda) = \sum_{k=0}^{m} (-1)^k \lambda^{2k}. \tag{1.4.30}$$

Hence it follows

$$(1 + \lambda^2) P_m(\lambda) = 1 + (-1)^m \lambda^{2m+2}, \quad \lambda \neq \pm i. \tag{1.4.31}$$

Solving the last equation (1.4.31), we get characteristic roots

$$\lambda_k = e^{i\theta_k}, \tag{1.4.32}$$

where

$$\theta_k \in \begin{cases} \frac{(2k+1)\pi}{2m+2} & m \text{ even}, \quad k \in \{0, 1, \ldots, 2m+1\} \setminus \left\{\frac{m}{2}, \frac{3m+2}{2}\right\}, \\ \frac{k\pi}{m+1} & m \text{ odd}, \quad k \in \{0, 1, \ldots, 2m+1\} \setminus \left\{\frac{m+1}{2}, \frac{3m+3}{2}\right\}. \end{cases} \tag{1.4.33}$$

We have $(2m + 2) - 2 = 2m$ different complex roots together, but each has a pair that is conjugate with it. Thus if m is even then we have m complex conjugate roots with multiplicity one. We also have $2m$ base elements alike complex roots:

m even

$$\begin{aligned} \varphi_k(x) &= \exp\left\{\big(\Re(\lambda_k)\big)x\right\} \cos\left[\big(\Im(\lambda_k)\big)x\right], & (1.4.34) \\ & \quad k \in \{0, 1, \ldots, m\} \setminus \{m/2\}; \\ \varphi_{m+1+k}(x) &= \exp\left\{\big(\Re(\lambda_k)\big)x\right\} \sin\left[\big(\Im(\lambda_k)\big)x\right], & (1.4.35) \\ & \quad k \in \{0, 1, \ldots, m\} \setminus \{m/2\}. \end{aligned}$$

On the other hand if m is odd then we have $2m - 2$ different complex roots together (each has a pair that is conjugate with it) and two real roots. Two real roots are ± 1 and $m - 1$ complex conjugate roots have the multiplicity one. We also have $2(m - 1) + 2 = 2m$ base elements alike all roots, too. So these base elements are:

m odd

$$\begin{aligned}
\varphi_0(x) &= \exp\{x\}\,; && (1.4.36)\\
\varphi_k(x) &= \exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big], && (1.4.37)\\
& \quad k\in\{1,2,\ldots,m\}\setminus\{(m+1)/2\}\,; \\
\varphi_{m+1}(x) &= \exp\{-x\}\,; && (1.4.38)\\
\varphi_{m+1+k}(x) &= \exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big], && (1.4.39)\\
& \quad k\in\{1,2,\ldots,m\}\setminus\{(m+1)/2\}\,.
\end{aligned}$$

These vectors generate a subspace of $\mathcal{C}^m\,[0,1]$ that is the space of solutions of the differential equation (1.4.24). In this case, the general solution is given by the linear combination:

$$L_a(x)\begin{cases}
= \displaystyle\sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m}\gamma_k\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\
+ \displaystyle\sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m}\gamma_{m+1+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big] \\
\qquad\qquad \ldots m \text{ even}, \\
\\
= \gamma_0\exp\{x\} + \displaystyle\sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m}\gamma_k\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\
+ \gamma_{m+1}\exp\{-x\} + \displaystyle\sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m}\gamma_{m+1+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big] \\
\qquad\qquad \ldots m \text{ odd},
\end{cases} \tag{1.4.40}$$

$$R_a(x)\begin{cases} = \sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_{2m+2+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_{3m+3+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big] \\ \qquad\qquad\ldots m \text{ even,} \\ \\ = \gamma_{2m+2}\exp\{x\}+\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_{2m+2+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\gamma_{3m+3}\exp\{-x\}+\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_{3m+3+k}\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big] \\ \qquad\qquad\ldots m \text{ odd,}\end{cases} \tag{1.4.41}$$

where coefficients γ_k are arbitrary constants that satisfy the boundary conditions (1.4.25)–(1.4.29). It can be easily seen that we have obtained $4(m+1)-4=4m$ coefficients γ_k, because the first index of γ_k is 0 and the last one is $4m+3$. Thus we have $4m$ boundary conditions and $4m$ unknowns of γ_ks that lead us to the square $4m\times 4m$ system of the linear equations. Does ψ_a exist and is it unique? To show this, it suffices to prove that the only solution of the associated homogeneous system of linear equations is the zero vector. Suppose $L_a(x)$ and $R_a(x)$ are functions corresponding to the solution of the homogeneous system, because in linear system of equations (1.4.25)–(1.4.29) the right side has all zeros—coefficient of $f(a)$ in the last boundary condition is 0 instead of 1. Then, by the exactly the same integration by parts, it follows that $\langle\psi_a,f\rangle_{Sob,m}=0$ for all $f\in\mathcal{C}^m[0,1]$. Hence $\psi_a(x)$, $L_a(x)$ and $R_a(x)$ are zero almost everywhere and thus by the linear independence of base elements $\varphi_k(x)$, so we have unique γ_ks.

ii) Producing a representor $\psi_{\mathbb{a}}^{\mathbb{w}}$

Let's produce the representor $\psi_{\mathbb{a}}^{\mathbb{w}}$ by setting

$$\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x})=\prod_{i=1}^{q}\psi_{a_i}^{w_i}(x_i)\quad\text{for all } \mathbb{x}\in\mathcal{Q}^q, \tag{1.4.42}$$

where $\psi_{a_i}^{w_i}(x_i)$ is the representor at a_i in $\mathcal{H}^m\left(Q^1\right)$. We know that $\mathcal{C}^m$ is dense in $\mathcal{H}^m$, so it is sufficient to show the result for $f\in\mathcal{C}^m(\mathcal{Q}^q)$. For simplicity let's suppose $\mathcal{Q}^q \equiv [0,1]^q$. After rewriting the inner product and using Fubini theorem we have

$$\begin{aligned}
&\left\langle \psi_{\mathbb{a}}^{\mathbb{w}}, f \right\rangle_{Sob,m} = \left\langle \prod_{i=1}^{q} \psi_{a_i}^{w_i}, f \right\rangle_{Sob,m} \\
&= \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_{\mathcal{Q}^q} \frac{\partial^{\alpha_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{\alpha_q}} D^{\boldsymbol{\alpha}} f(\mathbb{x}) \mathrm{d}\mathbb{x} \\
&= \sum_{i_1,\ldots,i_q=0,\ldots,m} \int_{\mathcal{Q}^q} \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}\mathbb{x} \\
&= \sum_{i_1=0}^{m} \int_0^1 \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \left[\cdots \left[\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} . \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}x_q \right] \cdots \right] \mathrm{d}x_1.
\end{aligned} \tag{1.4.43}$$

According to Definition 1.3.2 and notation in (1.4.15) we can rewrite the centermost bracket

$$\begin{aligned}
&\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} . \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}x_q \\
&\qquad = \left\langle \psi_{a_q}^{w_q}, D^{(i_1,\ldots,i_{q-1})} f(x_1,\ldots,x_{i-1},\cdot) \right\rangle_{Sob,m} \\
&\qquad = D^{(i_1,\ldots,i_{q-1},w_q)} f(\mathbb{x}_{-q}, a_q).
\end{aligned} \tag{1.4.44}$$

Chain proceeding in this way we obtain the value for the whole expression to be equal to $D^{\mathbb{w}} f(\mathbb{a})$. □

1.4.1 Calculation of Representors

We will here just continue from the proof of Representor Theorem 1.4.1 on page 5.

We would like to calculate the representor $\psi_a \equiv \psi_a^0$ of the function $f \in \mathcal{H}^m[0,1]$ (see (1.4.19)). All we need to know are functions L_a and R_a in the form (1.4.40) and (1.4.41). The proof of Representor Theorem 1.4.1 gives us exactly λ_ks from (1.4.32) and also ensures us of the existence and the uniqueness of the coefficients γ_k. Unknowns γ_k can be obtained from the square system of linear equations that consists of the boundary conditions (1.4.25)–(1.4.29) for the solution of differential equation (1.4.24).

Definition 1.4.1. Given $f \in \mathcal{H}^m[0,1]$ and an appropriate representor $\psi_a \equiv \psi_a^0$ s.t. $\langle \psi_a, f \rangle_{Sob,m} = f(a)$. We call the coefficients γ_k from (1.4.40) and (1.4.41) as coefficients of the representor ψ_a.

Theorem 1.4.2 (Obtaining Coefficients γ_k). *Coefficients γ_k of the representor ψ_a are unique solution of $4m \times 4m$ system of linear equations*

$$\sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_k \left(\varphi_k^{(m-j)}(0) + (-1)^j \varphi_k^{(m+j)}(0)\right) + \sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{m+1+k} \left(\varphi_{m+1+k}^{(m-j)}(0) + (-1)^j \varphi_{m+1+k}^{(m+j)}(0)\right) = 0, \quad j = 0, \ldots, m-1 \tag{1.4.45a}$$

$$\sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{2m+2+k} \left(\varphi_k^{(m-j)}(1) + (-1)^j \varphi_k^{(m+j)}(1)\right) + \sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{3m+3+k} \left(\varphi_{m+1+k}^{(m-j)}(1) + (-1)^j \varphi_{m+1+k}^{(m+j)}(1)\right) = 0, \quad j = 0, \ldots, m-1 \tag{1.4.45b}$$

$$\sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_k - \gamma_{2m+2+k}\right) \varphi_k^{(j)}(a) + \sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_{m+1+k} - \gamma_{3m+3+k}\right) \varphi_{m+1+k}^{(j)}(a) = 0, \quad j = 0, \ldots, 2m-2, \tag{1.4.45c}$$

$$\sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_k - \gamma_{2m+2+k}\right) \varphi_k^{(2m-1)}(a) + \sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_{m+1+k} - \gamma_{3m+3+k}\right) \varphi_{m+1+k}^{(2m-1)}(a) = (-1)^{m-1}, \tag{1.4.45d}$$

where

$$\kappa := \begin{cases} \frac{m}{2}, & m \text{ even}, \\ \frac{m+1}{2}, & m \text{ odd} \end{cases} \tag{1.4.46}$$

and φ_k are defined in (1.4.34)–(1.4.39).

Proof. Existence and uniqueness of coefficients γ_ks has already been proved in the proof of Theorem 1.4.1.

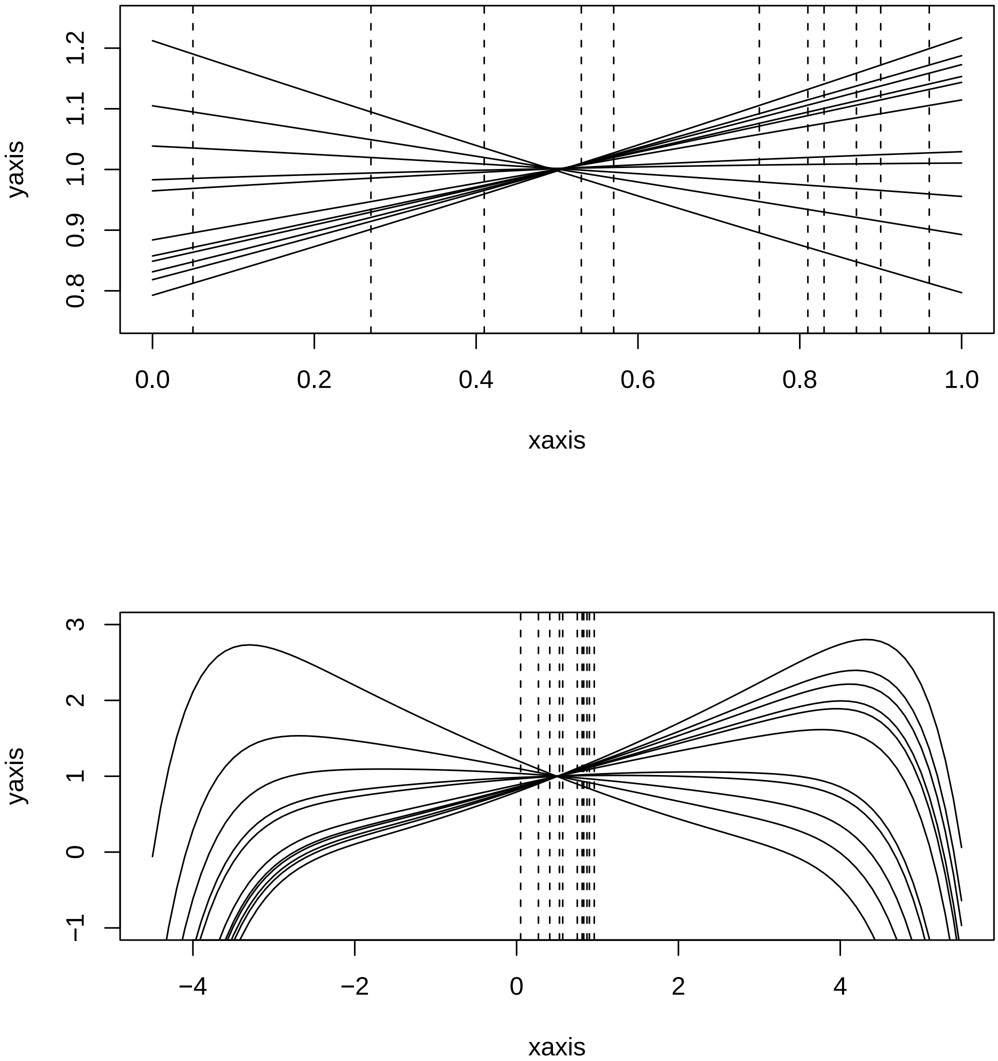

Figure 1.4.1: Representors in Sobolev space $\mathcal{H}^4[0,1]$ for data points $\mathbb{x} = (0.05, 0.27, 0.41, 0.53, 0.57, 0.75, 0.81, 0.83, 0.87, 0.9, 0.96)'$—dashed vertical lines. Zoomed view in the upper picture—interval $[0,1]$, reduced view in the lower picture—interval $[-4.5, +5.5]$.

Let's define

$$\Lambda_{a,I}^{(l)} := \begin{cases} L_a^{(l)}(0), & \text{for } I = L; \\ R_a^{(l)}(1), & \text{for } I = R; \\ L_a^{(l)}(a) - R_a^{(a)}(a), & \text{for } I = D. \end{cases} \tag{1.4.47}$$

From (1.4.26)–(1.4.29) we easily see

$$\sum_{k=i+1}^{m}(-1)^{k-i-1}\Lambda_{a,I}^{(2k-i-1)} = 0, \quad 0\le i\le m-1,\ I\in\{L,R,D\}\,, \quad \text{except } [i,I]=[0,D]\,; \tag{1.4.48}$$

$$\sum_{k=1}^{m}(-1)^{k-1}\Lambda_{a,D}^{(2k-1)} = 1. \tag{1.4.49}$$

If $m=1$ it directly follows from (1.4.48)–(1.4.49):

$$\Lambda_{a,I}^{(1)} = 0, \quad I\in\{L,R\}\,, \tag{1.4.50}$$

$$\Lambda_{a,D}^{(1)} = 1. \tag{1.4.51}$$

If $m=2$ it also directly follows from (1.4.48)–(1.4.49):

$$\Lambda_{a,I}^{(2)} = 0, \quad \forall I, \tag{1.4.52}$$

$$\Lambda_{a,I}^{(1)}-\Lambda_{a,I}^{(3)} = 0, \quad I\in\{L,R\}\,, \tag{1.4.53}$$

$$\Lambda_{a,D}^{(1)}-\Lambda_{a,D}^{(3)} = 1. \tag{1.4.54}$$

Suppose $m\ge 3$. We would like to prove this important step in our proof:

$$\Lambda_{a,I}^{(m-j)}+(-1)^{j}\Lambda_{a,I}^{(m+j)} = 0, \quad j=0,\ldots,m-2,\ \forall I, \tag{1.4.55}$$

$$\Lambda_{a,I}^{(1)}+(-1)^{m-1}\Lambda_{a,I}^{(2m-1)} = 0, \quad I\in\{L,R\}\,, \tag{1.4.56}$$

$$\Lambda_{a,D}^{(1)}+(-1)^{m-1}\Lambda_{a,D}^{(2m-1)} = 1, \tag{1.4.57}$$

where $j:=m-i-1$.
For $j=0$ is $i=m-1$ and from (1.4.48)–(1.4.49) we have straightforwardly

$$\Lambda_{a,I}^{(m)}=0, \quad \forall I, \tag{1.4.58}$$

which is correct according to (1.4.55). Consider $j=1$ and thus $i=m-2$. In a same way we get a correspondent result to (1.4.55)

$$\Lambda_{a,I}^{(m-1)}-\Lambda_{a,I}^{(m+1)}=0, \quad \forall I. \tag{1.4.59}$$

For $j=2$ and thus $i=m-3$ we have

$$\Lambda_{a,I}^{(m-2)}-\Lambda_{a,I}^{(m)}+\Lambda_{a,I}^{(m+2)}=0, \quad \forall I, \tag{1.4.60}$$

so it forces us to apply (1.4.58) and for $j=3$ and thus $i=m-4$ we have

$$\Lambda_{a,I}^{(m-3)}-\Lambda_{a,I}^{(m-1)}+\Lambda_{a,I}^{(m+1)}-\Lambda_{a,I}^{(m+3)}=0, \quad \forall I, \tag{1.4.61}$$

so we can apply (1.4.59). We could continue in this way finite times (formally we can proceed this by something like a finite double induction). We finish when $j = m - 1$. The last step ensures the correctness of (1.4.56) in case $I \in \{L, R\}$, eventually (1.4.57) in case $I = D$ instead of (1.4.55).
To finish this proof all we need to do is not to forget to think of (1.4.25). From (1.4.25) we obtain

$$\Lambda_{a,D}^{(j)} = 0, \quad j \in \{0, \ldots, m-1\}. \tag{1.4.62}$$

According to (1.4.55) for $I = D$ and (1.4.57) we further see:

$$\begin{aligned} \Lambda_{a,D}^{(j)} &= 0, \quad j \in \{m+1, \ldots, 2m-2\}; &(1.4.63)\\ \Lambda_{a,D}^{(2m-1)} &= (-1)^{m-1}. &(1.4.64) \end{aligned}$$

Alltogether we have obtained these $4m$ linear equations

$$\begin{aligned} \Lambda_{a,L}^{(m-j)} + (-1)^j \Lambda_{a,L}^{(m+j)} &= 0, \quad j = 0, \ldots, m-1, &(1.4.65)\\ \Lambda_{a,R}^{(m-j)} + (-1)^j \Lambda_{a,R}^{(m+j)} &= 0, \quad j = 0, \ldots, m-1, &(1.4.66)\\ \Lambda_{a,D}^{(j)} &= 0, \quad j = 0, \ldots, 2m-2, &(1.4.67)\\ \Lambda_{a,D}^{(2m-1)} &= (-1)^{m-1}, &(1.4.68) \end{aligned}$$

which after rewriting them using (1.4.47), (1.4.40)–(1.4.41) and (1.4.34)–(1.4.39) bring us to a close. □

Remark 1.4.5. The square system of $4m$ linear equations (1.4.45) can be written in the matrix notation in more illustrative way:

$$\underbrace{\left(\begin{array}{c|c} \begin{array}{c}\varphi_k^{(m-j)}(0)\\ +(-1)^j\varphi_k^{(m+j)}(0)\end{array} & \emptyset \\ \hline \emptyset & \begin{array}{c}\varphi_k^{(m-j)}(1)\\ +(-1)^j\varphi_k^{(m+j)}(1)\end{array} \\ \hline \varphi_k^{(j)}(a) & -\varphi_k^{(j)}(a) \\ \hline \varphi_k^{(2m-1)}(a) & -\varphi_k^{(2m-1)}(a) \end{array}\right)}_{\{\Gamma_{j,k}\}} \underbrace{\left(\begin{array}{c} \gamma_0\\ \vdots\\ \gamma_{\kappa-1}\\ \gamma_{\kappa+1}\\ \vdots\\ \gamma_m\\ \gamma_{m+1}\\ \vdots\\ \gamma_{m+\kappa}\\ \gamma_{m+2+\kappa}\\ \vdots\\ \gamma_{2m+1}\\ \hline \gamma_{2m+2}\\ \vdots\\ \gamma_{2m+1+\kappa}\\ \gamma_{2m+3+\kappa}\\ \vdots\\ \gamma_{3m+2}\\ \gamma_{3m+3}\\ \vdots\\ \gamma_{3m+2+\kappa}\\ \gamma_{3m+4+\kappa}\\ \vdots\\ \gamma_{4m+3} \end{array}\right)}_{\{\gamma_k\}} = \left(\begin{array}{c} 0\\ \vdots\\ 0\\ \hline 0\\ \vdots\\ 0\\ \hline 0\\ \vdots\\ 0\\ \hline (-1)^{m-1} \end{array}\right) \begin{array}{c} \overbrace{}^{j=}\\ 0\\ \vdots\\ \text{m-1}\\ 0\\ \vdots\\ \text{m-1}\\ 0\\ \vdots\\ \text{2m-2}\\ \text{2m-1} \end{array}. \tag{1.4.69}$$

Hence we obtain the coefficients γ_k as a solution of:

$$\boldsymbol{\gamma} = (-1)^{m-1}\left[\boldsymbol{\Gamma}^{-1}\right]_{\bullet,4m}. \tag{1.4.70}$$

Remark 1.4.6. We easily see from the proof of Theorem 1.4.1 and the proof of Theorem 1.4.2 that the form of representor ψ_{a} does not depend on exact closed cube $[c_1, d_1]\times\ldots\times[c_q, d_q] \in \mathbb{R}^q$. In Figure 1.4.1 we can see representors in Sobolev space $\mathcal{H}^4[0,1]$ for eleven data points.

1.5 Bounded Derivatives in Subnorm

Theorem 1.5.1 (Embedding). *The embedding* $\mathcal{H}^m(\mathcal{Q}^q) \hookrightarrow \mathcal{C}^{m-1}(\mathcal{Q}^q)$ *is compact.*

Remark 1.5.1. Proof of this Theorem can be found in Bos and Yatchew (1997). We only proceed some steps more properly for better understanding.

Proof. We divide the proof into two steps again.

i) Existence of the embedding

We will first show that $f \in \mathcal{H}^m(\mathcal{Q}^q)$ implies that $f \in \mathcal{C}^{m-1}(\mathcal{Q}^q)$. In one dimensional case just see Adams (1975).

Given $0 \leq w \leq m-1$ and $a \in \mathcal{Q}^1$, the mapping

$$\Phi\colon \mathcal{H}^m(\mathcal{Q}^1) \to \mathbb{R}^1\colon f \mapsto \frac{d^w f(a)}{\mathrm{d}x^w} \tag{1.5.71}$$

is a bounded linear functional on $\mathcal{H}^m(\mathcal{Q}^1)$. Bounded linear functional is continuous and then by the Riesz Representation Theorem C.1.3 there exists a representor $\psi_a^w \in \mathcal{H}^m(\mathcal{Q}^1)$ such that for all $f \in \mathcal{H}^m(\mathcal{Q}^1)$,

$$\langle \psi_a^w, f \rangle_{Sob,m} = \frac{d^w f(a)}{\mathrm{d}x^w}. \tag{1.5.72}$$

There is also $\rho > 0$ such that if $\|f\|_{Sob,m} \leq 1$ then $\|f\|_{\infty,\infty,m-1} \leq \rho$ (compact sets are bounded)[3]. Hence:

$$\exists_{\rho>0} \forall_{0\leq w\leq m-1} \forall_{a\in\mathcal{Q}^1} \forall_{f\in\mathcal{H}^m(\mathcal{Q}^1)} \ \colon \ \|f\|_{Sob,m} \leq 1 \Rightarrow \left| \langle \psi_a^w, f \rangle_{Sob,m} \right| \leq \rho. \tag{1.5.73}$$

Since $\psi_a^w \in \mathcal{H}^m(\mathcal{Q}^1)$ then

$$\left| \left\langle \psi_a^w, \frac{\psi_a^w}{\|\psi_a^w\|_{Sob,m}} \right\rangle_{Sob,m} \right| = \|\psi_a^w\|_{Sob,m} \leq \rho. \tag{1.5.74}$$

By the construction of the representors (see proof of Theorem 1.4.1) it can be easily seen that there exists a continuous mapping

$$\Upsilon\colon \mathcal{Q}^1 \to \mathcal{H}^m(\mathcal{Q}^1)\colon a \mapsto \psi_a^w \tag{1.5.75}$$

for $0 \leq w \leq m-1$.

[3] See notation (1.2.4a) in Definition 1.2.1.

Consider the several variable case and define $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) := \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i)$. For all $f \in \mathcal{H}^m(\mathcal{Q}^q)$ and $\mathbb{w}$ such that $|\mathbb{w}|_\infty \leq m-1$ by Theorem 1.4.1 $\langle \psi_{\mathbb{a}}^{\mathbb{w}} \rangle_{Sob,m} = D^{\mathbb{w}} f(\mathbb{a})$. It is straightforward to show that:

$$\|\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x})\|_{Sob,m} = \prod_{i=1}^{q} \left\|\psi_{a_i}^{w_i}(x_i)\right\|_{Sob,m}. \tag{1.5.76}$$

Let $f \in \mathcal{H}^m(\mathcal{Q}^q)$. By the definition there is a sequence $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{C}^m(\mathcal{Q}^q)$ such that $\|f - f_n\|_{Sob,m} \overset{n\to\infty}{\longrightarrow} 0$. We wish to show that $\{f_n\}_{n\in\mathbb{N}}$ is convergent in $\mathcal{C}^{m-1}(\mathcal{Q}^q)$ by showing that it is Cauchy in this space. But we know that $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in $\mathcal{H}^m(\mathcal{Q}^q)$. Hence, given $\epsilon > 0$, choose $n_0 \in \mathbb{N}$ large enough so that $n_1, n_2 > n_0$ implies $\|f_{n_1} - f_{n_2}\|_{Sob,m} < \frac{\epsilon}{m^q \rho^q}$. Then for $n_1, n_2 > n_0$ using notation (1.2.4a) from Definition 1.2.1, Theorem 1.4.1 and Cauchy-Schwartz C.1.2 we have:

$$\begin{aligned}
\|f_{n_1} - f_{n_2}\|_{\infty,\infty,m-1} &= \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \sup_{\mathbb{y}\in\mathcal{Q}^q} \left| D^{\boldsymbol{\alpha}} f_{n_1}(\mathbb{y}) - D^{\boldsymbol{\alpha}} f_{n_2}(\mathbb{y}) \right| \\
&= \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \sup_{\mathbb{y}\in\mathcal{Q}^q} \left| \left\langle \psi_{\mathbb{y}}^{\boldsymbol{\alpha}}, f_{n_1} - f_{n_2} \right\rangle_{Sob,m} \right| \\
&\leq \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \sup_{\mathbb{y}\in\mathcal{Q}^q} \left\|\psi_{\mathbb{y}}^{\boldsymbol{\alpha}}\right\|_{Sob,m} \left\|f_{n_1} - f_{n_2}\right\|_{Sob,m} \\
&< m^q \rho^q \frac{\epsilon}{m^q \rho^q} = \epsilon.
\end{aligned} \tag{1.5.77}$$

Do not forget that the last inequality follows from the relations (1.5.74) and (1.5.76) and that there are m^q elements in the summation (each α_i taking on values $0, \ldots, m-1$). Hence, we find out that $\{f_n\}_{n\in\mathbb{N}}$ is indeed Cauchy in $\mathcal{C}^{m-1}(\mathcal{Q}^q)$. But as $\mathcal{C}^{m-1}(\mathcal{Q}^q)$ is complete, there exists $g \in \mathcal{C}^{m-1}(\mathcal{Q}^q)$ such that $f_n \overset{\mathcal{C}^{m-1}(\mathcal{Q}^q)}{\longrightarrow} g$, that is, $D^w f_n \to D^w g$ uniformly for all $|\mathbb{w}|_\infty \leq m-1$. Then $D^w f_n \overset{L_2(\mathcal{Q}^q)}{\longrightarrow} D^w g$ for all $|\mathbb{w}|_\infty \leq m-1$, or in other words, $f_n \overset{\mathcal{H}^{m-1}(\mathcal{Q}^q)}{\longrightarrow} g$. But by the definition, $f_n \overset{\mathcal{H}^{m-1}(\mathcal{Q}^q)}{\longrightarrow} f$. Hence, by the uniqueness of limit, $f = g$, $f \in \mathcal{C}^{m-1}(\mathcal{Q}^q)$.

ii) Compactness of the embedding

To show the compactness, we proceed by induction on m. Consider $m = 1$. In this case, we must show that if $\{f_n\}_{n\in\mathbb{N}}$ is a bounded sequence in $\mathcal{H}^1(\mathcal{Q}^q)$ then there is a subsequence that converges in $\mathcal{C}^0(\mathcal{Q}^q)$. But by the Arzela-Ascoli Theorem C.1.1 it suffices to show that $\{f_n\}$ is equibounded and uniformly equicontinuous. We know that $\mathcal{Q}^q$ is a closed domain, hence $\{f_n\}$ is equibounded. If we proof that $\{f_n\}$ is equicontinuous then it will be also uniformly equicontinuous, because the same reason—$\mathcal{Q}^q$ is a closed domain. The equicontinuity follows from the fact

$$|f_n(\mathrm{a}) - f_n(\tilde{\mathrm{a}})| = \left|\langle \psi_{\mathrm{a}} - \psi_{\tilde{\mathrm{a}}}, f_n \rangle_{Sob,m}\right| \leq \|\psi_{\mathrm{a}} - \psi_{\tilde{\mathrm{a}}}\|_{Sob,m} \|f_n\|_{Sob,m} \quad (1.5.78)$$

and that ψ_{a} depends continuously on a.

Now suppose that the embedding $\mathcal{H}^m(\mathcal{Q}^q) \hookrightarrow \mathcal{C}^{m-1}(\mathcal{Q}^q)$ is compact for particular m and consider $m+1$. Let $\{f_n\}$ be a bounded sequence in $\mathcal{H}^{m+1}(\mathcal{Q}^q)$. Then $\{D^{\boldsymbol{\alpha}} f_n\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{H}^m(\mathcal{Q}^q)$ for each $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}|_\infty \leq 1$. Hence for each $|\boldsymbol{\alpha}|_\infty \leq 1$ there is a subsequence of $\{D^{\boldsymbol{\alpha}} f_n\}$ which converges in $\mathcal{C}^{m-1}(\mathcal{Q}^q)$, because in the compact space each sequence has a convergent subsequence. By passing to the subsequence $\{D^{\boldsymbol{\alpha}} f_{n_k}\}_{k\in\mathbb{N}}$ we have a sequence which converges in $\mathcal{C}^{m-1}(\mathcal{Q}^q)$ for all $|\boldsymbol{\alpha}|_\infty \leq 1$. That is, $\{f_{n_k}\}_{k\in\mathbb{N}}$ converges in $\mathcal{C}^{m-1+1}(\mathcal{Q}^q)$. □

Remark 1.5.2. The compactness of this embedding means that given a closed ball of functions in $\mathcal{H}^m(\mathcal{Q}^q)$ with respect to $\|\cdot\|_{Sob,m}$, its closure is compact in $\mathcal{C}^{m-1}(\mathcal{Q}^q)$. This result ensures us that functions in a bounded ball in $\mathcal{H}^m$ have all lower order derivatives bounded in supnorm.

Chapter 2

Regression in Sobolev Spaces

Connection of features of L_2-spaces and $\mathcal{C}^m$-spaces can grant us an interesting background for the nonparametric regression. L_2-spaces are special types of Hilbert spaces that facilitate the calculation of least square projection. On the other hand, we regard $\mathcal{C}^m$-spaces as one of the common classes of functions that we want to approximate the data with.

We propose some statistical models from Bos and Yatchew (1997) and Härdle and Yatchew (2003). According to the assumptions from these models, we choose least squares as an appropriate way of penalizing. The focus of this chapter is hidden in Theorem 2.2.1 (Infinite to Finite), again from Bos and Yatchew (1997). I have amended this theorem and shown the uniqueness of the regression estimator. Finally, I have generalized this theorem into the weighted regression case (Theorem 2.3.1). Then I have investigated the form of the regression estimator in the Sobolev space in Corollary 2.2.2. Hence I have also proved symmetry and positive definiteness of the representor matrix.

I have used Schur decomposition Theorem C.2.2 to solve optimizing problem with quadratic constraint (Theorem 2.2.5) and proved the existence and the uniqueness of its solution in Theorem 2.2.6 and Algorithm 2.2.1. The asymptotic behavior of the regression estimator (Theorem 2.2.7) is also investigated in Bos and Yatchew (1997).

In the end, we are concerned with Cross-Validation. I have proved the existence of 1–1 mapping between the Sobolev bound and the smoothing parameter (Theorem 2.6.1).

2.1 Single Equation Model

Definition 2.1.1 (Single Equation Model). The single equation model is

$$Y_i = f(\mathbb{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.1.1}$$

with these assumptions:

i) $\mathbb{X}_i$ are q-dimensional random vectors, i.i.d. with probability law $\mathcal{P}_x$ and density p_x bounded away from zero on the support $\mathcal{Q}^q$, the unit cube in $\mathbb{R}^q$;

ii) ε_i are i.i.d. random variables with probability law $\mathcal{P}_\varepsilon$ so that $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{V}\mathrm{ar}\,\varepsilon_i = \sigma_\varepsilon^2$ for all i; $\mathcal{P}_\varepsilon \in \mathscr{P}_\varepsilon$ a collection of probability laws with mean 0 and support contained in a bounded interval of $\mathbb{R}^1$; $\mathbb{X}_i$ and ε_i are independent;

iii) $f \in \mathscr{F}$, where $\mathscr{F}$ is a family of functions in the Sobolev space $\mathcal{H}^m(\mathcal{Q}^q)$ from $\mathbb{R}^q$ to $\mathbb{R}^1$, $m > \frac{q}{2}$, $\mathscr{F} = \left\{ f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|_{Sob,m}^2 \leq L \right\}$.

Remark 2.1.1. Our setting is concerned with independent, identically distributed random variables $\{(\mathbb{X}_i, Y_i)\}_{i=1}^n$. The regression curve can be also defined as

$$f(\mathbb{x}) = \mathbb{E}\left[Y | \mathbb{X} = \mathbb{x}\right]. \tag{2.1.2}$$

It is common terminology to refer to this setting as the *random design model*. By contrast, the *fixed design model* is concerned with controlled, non-stochastic variables $\mathbb{x}_i$. Unless otherwise indicated, we think of the random design model.

Notation 2.1.1. From this point now on we denote $\mathcal{H}^m \equiv \mathcal{H}^m(\mathcal{Q}^q)$, where $\mathcal{Q}^q$ is the unit cube in $\mathbb{R}^q$.

2.2 Least Squares

Least squares are the most typical and standard way of error penalization. Our regression problem can be characterized by one of these ways:

a)

$$\min_{f \in \mathcal{H}^m} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - f(\mathbb{x}_i)\right]^2 \quad \text{s.t.} \quad \|f\|_{Sob,m}^2 \leq L, \tag{2.2.3}$$

b)

$$\min_{f\in\mathcal{H}^m}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(\mathbb{x}_i)\right]^2 + \chi\left\|f\right\|^2_{Sob,m}\right\}. \tag{2.2.4}$$

The Sobolev norm bound L and also the smoothing parameter (bandwidth parameter) χ controlls the tradeoff between the infidelity to the data and roughness of the estimated solution.

Definition 2.2.1 (Representor Matrix). Let $\psi_{\mathbb{x}_1},\ldots,\psi_{\mathbb{x}_n}$ be the representors for function evaluation at $\mathbb{x}_1,\ldots,\mathbb{x}_n$ respectively, i.e. $\langle\psi_{\mathbb{x}_i}, f\rangle_{Sob,m} = f(\mathbb{x}_i)$ for all $f\in\mathcal{H}^m$, $i=1,\ldots,n$. Let $\mathbf{\Psi}$ be the $n\times n$ representor matrix whose columns (and rows) equal the representors evaluated at $\mathbb{x}_1,\ldots,\mathbb{x}_n$; i.e.

$$\Psi_{i,j} = \left\langle\psi_{\mathbb{x}_i},\psi_{\mathbb{x}_j}\right\rangle_{Sob,m} = \psi_{\mathbb{x}_i}(\mathbb{x}_j) = \psi_{\mathbb{x}_j}(\mathbb{x}_i). \tag{2.2.5}$$

Theorem 2.2.1 (Infinite to Finite). *Let $\mathbb{x}_1,\ldots,\mathbb{x}_n$ be $q\times 1$ vectors, $\mathbb{y} = (y_1,\ldots,y_n)'$ and define*

$$\hat{\sigma}^2 = \min_{f\in\mathcal{H}^m}\frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(\mathbb{x}_i)\right]^2 \quad s.t.\ \ \|f\|^2_{Sob,m}\leq L, \tag{2.2.6}$$

$$s^2 = \min_{\mathbb{c}\in\mathbb{R}^n}\frac{1}{n}\left[\mathbb{y}-\mathbf{\Psi}\mathbb{c}\right]'\left[\mathbb{y}-\mathbf{\Psi}\mathbb{c}\right] \quad s.t.\ \ \mathbb{c}'\mathbf{\Psi}\mathbb{c}\leq L \tag{2.2.7}$$

where $\mathbb{c}$ is an $n\times 1$ vector and $\mathbf{\Psi}$ is the representor matrix at $\mathbb{x}_1,\ldots,\mathbb{x}_n$. Then $\hat{\sigma}^2 = s^2$. Furthermore, there exists a solution to (2.2.6) *of the form*

$$\widehat{f} = \sum_{i=1}^{n}\hat{c}_i\psi_{\mathbb{x}_i} \tag{2.2.8}$$

where $\widehat{\mathbb{c}} = (\hat{c}_1,\ldots,\hat{c}_n)'$ solves (2.2.7). *The estimator $\widehat{f}$ is unique a.e.*

Remark 2.2.1. Proof of this Theorem was adopted from Bos and Yatchew (1997) with small modifications.

Proof. Let $M = span\left\{\psi_{\mathbb{x}_i} : i=1,\ldots,n\right\}$ and its orthogonal complement $M^{\perp} = \left\{h\in\mathcal{H}^m : \langle\psi_{\mathbb{x}_i}, h\rangle_{Sob,m} = 0,\ i=1,\ldots,n\right\}$. Representors exist by Theorem 1.4.1 and we can write the Sobolev space as a direct sum of its orthogonal subspaces, i.e. $\mathcal{H}^m = M\oplus M^{\perp}$ since $\mathcal{H}^m$ is a Hilbert space. Functions $h\in M^{\perp}$ take on the value zero at $\mathbb{x}_1,\ldots,\mathbb{x}_n$. Each $f\in\mathcal{H}^m$ can be written in form

$$f = \sum_{j=1}^{n}c_j\psi_{\mathbb{x}_j} + h, \quad h\in M^{\perp}. \tag{2.2.9}$$

Then

$$\begin{aligned}
&\sum_{i=1}^{n}[y_i - f(\mathbb{x}_i)]^2 = \sum_{i=1}^{n}\left[y_i - \left\langle \psi_{\mathbb{x}_i}, \sum_{j=1}^{n} c_j\psi_{\mathbb{x}_j} + h\right\rangle_{Sob,m}\right]^2 \\
&= \sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{n}\left\langle \psi_{\mathbb{x}_i}, c_j\psi_{\mathbb{x}_j}\right\rangle_{Sob,m}\right]^2 = \sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{n} c_j\left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j}\right\rangle_{Sob,m}\right]^2 \\
&= \sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{n}\Psi_{ij}c_j\right]^2 = [\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}]'[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}].
\end{aligned} \tag{2.2.10}$$

Note further that

$$\begin{aligned}
&\|f\|^2_{Sob,m} = \langle f, f\rangle_{Sob,m} \\
&= \left\langle \sum_{j=1}^{n} c_j\psi_{\mathbb{x}_j}, \sum_{j=1}^{n} c_j\psi_{\mathbb{x}_j}\right\rangle_{Sob,m} + \langle h, h\rangle_{Sob,m} = \mathbb{c}'\boldsymbol{\Psi}\mathbb{c} + \langle h, h\rangle_{Sob,m}.
\end{aligned} \tag{2.2.11}$$

Suppose that $f = \sum_{j=1}^{n} c_j\psi_{\mathbb{x}_j} + h$ minimizes

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f(\mathbb{x}_i)]^2 \text{ s.t. } \|f\|^2_{Sob,m} \leq L \tag{2.2.12}$$

then so does $f^* = f - h$ is zero at $\mathbb{x}_1, \ldots, \mathbb{x}_n$. Hence, there exists a function f^* minimizing the infinite dimensional optimizing problem that is a linear combination of the representors. Furthermore,

$$\|f^*\|^2_{Sob,m} \leq \|f^*\|^2_{Sob,m} + \|h\|^2_{Sob,m} = \|f\|^2_{Sob,m} \leq L. \tag{2.2.13}$$

We note also that $\|f^*\|^2_{Sob,m} = \mathbb{c}'\boldsymbol{\Psi}\mathbb{c}$. Finally, we observe that the following two problems are equivalent:

$$\min_{f^*} \frac{1}{n}\sum_{i=1}^{n}[y_i - f^*(\mathbb{x}_i)]^2 \quad \text{s.t.} \quad \|f^*\|^2_{Sob,m} \leq L,\, f^* \in M; \tag{2.2.14}$$

$$\min_{\mathbb{c}} \frac{1}{n}[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}]'[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}] \quad \text{s.t.} \quad \mathbb{c}'\boldsymbol{\Psi}\mathbb{c} \leq L. \tag{2.2.15}$$

Uniqueness is clear, since $\psi_{\mathbb{x}_i}$ are the base elements of M, $\left\|\widehat{f}\right\|^2_{Sob,m} = L$ and adding a function that is orthogonal to the spaces spanned by the representors will increase the norm. □

Remark 2.2.2. Theorem 2.2.1 transforms the infinite dimensional problem into a finite dimensional one—quadratic optimization problem.

Corollary 2.2.2 (Form of the Regression Function). *The regression function from Theorem 2.2.1 in one-dimensional case is of this form:*

$$\widehat{f}(x) = \begin{cases} \sum_{i=1}^{n} \hat{c}_i L_{x_i}(x), & 0 \leq x \leq x_1, \\ \vdots & \vdots \\ \sum_{i=j+1}^{n} \hat{c}_i L_{x_i}(x) + \sum_{i=1}^{j} \hat{c}_i R_{x_i}(x), & x_j < x \leq x_{j+1},\ j = 1, \ldots, n-1; \\ \vdots & \vdots \\ \sum_{i=1}^{n} \hat{c}_i R_{x_i}(x), & x_n < x \leq 1, \end{cases} \tag{2.2.16}$$

where $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)'$ *solves* (2.2.7) *and functions* $L_{x_i}(x)$ *and* $R_{x_i}(x)$ *are defined in* (1.4.19).

Proof. Trivial. It can be directly seen from the definition of form (1.4.19) of the representor and from (2.2.8). □

Remark 2.2.3. Corollary 2.2.2 can be easily extended for q-dimensional vector variable $\mathbb{x}$ if we realize how the representor $\psi_{\mathbb{a}}$ has been produced in the proof of Theorem 1.4.1. Then we just apply (1.4.19) on the form of each factor ψ_a of the product of representors $\psi_{\mathbb{a}}$ in (1.4.42). The only difference in (2.2.16) will be the number of cases. We will obtain $(n+1)^q$ decision conditions (vector $\mathbb{x}$ has q components) instead of actual number $n+1$ $(0 \leq x \leq x_1, \ldots, x_j < x \leq x_{j+1}, \ldots, x_n < x)$.

Remark 2.2.4. The form of the regression function can be written alternatively:

$$\begin{aligned} \widehat{f}(x) = \sum_{j=1}^{n} \widehat{c_j} \sum_{k=1}^{2m} \exp\left[\Re\left(e^{i\theta_k}\right)x\right] \\ \left\{ I_{[x \leq x_j]} \gamma_k \cos\left[\Im\left(e^{i\theta_k}\right)x\right] + I_{[x > x_j]} \gamma_{2m+k} \sin\left[\Im\left(e^{i\theta_k}\right)x\right] \right\}. \end{aligned} \tag{2.2.17}$$

It is very important to realize that $\widehat{f}$ is not estimated using goniometric splines neither kernel functions!

Lemma 2.2.3 (Symmetry of Representor Matrix). *Representor Matrix is symmetric.*

Proof. Trivial. Representor matrix is symmetric by Definition 2.2.1, because

$$\Psi_{i,j} = \left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} = \left\langle \psi_{\mathbb{x}_j}, \psi_{\mathbb{x}_i} \right\rangle_{Sob,m} = \Psi_{j,i}, \tag{2.2.18}$$

i.e. $\mathbf{\Psi} = \mathbf{\Psi}'$. □

Theorem 2.2.4 (Positive Definiteness of Representor Matrix). *Representor Matrix is positive definite.*

Proof. We proceed this proof only for one dimensional variable x. Extention into the multi-variable case is clearly simple (see Remark 2.2.3). For an arbitrary $\mathbb{c} \in \mathbb{R}^n$ we obtain

$$\begin{aligned} \mathbb{c}'\mathbf{\Psi}\mathbb{c} &= \sum_i c_i \sum_j \Psi_{ij} c_j = \sum_i \sum_j c_i \left\langle \psi_{x_i}, \psi_{x_j} \right\rangle_{Sob,m} c_j \\ &= \sum_i \sum_j \left\langle c_i\psi_{x_i}, c_j\psi_{x_j} \right\rangle_{Sob,m} = \left\langle \sum_i c_i\psi_{x_i}, \sum_j c_j\psi_{x_j} \right\rangle_{Sob,m} \\ &= \left\| \sum_i c_i \psi_{x_i} \right\|^2_{Sob,m} \geq 0. \end{aligned} \tag{2.2.19}$$

Hence $\mathbb{c}'\mathbf{\Psi}\mathbb{c} = 0$ iff $\sum_i c_i\psi_{x_i} = 0$ a.e. According to (1.4.40)–(1.4.41), (1.4.34)–(1.4.39) and (1.4.70) we have[1]

$$\psi_{x_i}(x) = \boldsymbol{\gamma}(x_i)'\boldsymbol{\varphi}(x) = (-1)^{m-1} \left[\mathbf{\Gamma}_{x_i}^{-1}\right]_{\bullet,4m}' \boldsymbol{\varphi}(x) \tag{2.2.20}$$

where $\boldsymbol{\varphi}(x)$ is vector which elements are linear independent base elements of space of the differential equation's (1.4.24) solutions, i.e. $\varphi_k(x)$ (see 1.4.34–

[1] If $x > x_i$ then $\boldsymbol{\gamma}(x_i) = (\gamma_0, \ldots, \gamma_{\kappa-1}, \gamma_{\kappa+1}, \ldots, \gamma_{m+\kappa}, \gamma_{m+2+\kappa}, \ldots, \gamma_{2m+1})'(x_i)$ else $\boldsymbol{\gamma}(x_i) = (\gamma_{2m+2}, \ldots, \gamma_{2m+1+\kappa}, \gamma_{2m+3+\kappa}, \ldots, \gamma_{3m+2+\kappa}, \gamma_{3m+4+\kappa}, \ldots, \gamma_{4m+3})'(x_i)$. Similarly with elements of vector $\left[\mathbf{\Gamma}_{x_i}^{-1}\right]_{\bullet,4m}$.

1.4.39). Thus linear independence of $\varphi_k(x)$ it follows that

$$\begin{aligned}
\sum_i c_i \psi_{x_i} &= (-1)^{m-1} \sum_i c_i \left[\boldsymbol{\Gamma}_{x_i}^{-1}\right]'_{\bullet,4m} \boldsymbol{\varphi} \\
&= (-1)^{m-1} \sum_i \sum_k c_i \left[\boldsymbol{\Gamma}_{x_i}^{-1}\right]_{4m,k} \varphi_k = 0 \quad \text{a.e.} \qquad (2.2.21) \\
&\Updownarrow \\
\varphi_k &= 0 \quad \text{a.e.} \quad k \in \{0,1,\ldots,2m+1\} \setminus \begin{cases} \left\{\frac{m}{2}, \frac{3m+2}{2}\right\} & m \text{ even}, \\ \left\{\frac{m+1}{2}, \frac{3m+3}{2}\right\} & m \text{ odd}; \end{cases} \qquad (2.2.22) \\
&\Downarrow \\
\psi_{x_i} &= 0 \quad \text{a.e.} \quad i = 1,\ldots,n. \qquad (2.2.23)
\end{aligned}$$

And $\psi_{x_i} = 0$ a.e. is a zero element of the space $\mathcal{H}^m$. □

Remark 2.2.5. Slightly different conception of representors has been described in Wahba (1992).

Theorem 2.2.5 (Optimizing with Constraint). *Optimizing problem with constraint*

$$\min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}]' [\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}] \quad s.t. \ \mathbb{c}'\boldsymbol{\Psi}\mathbb{c} \leq L, \qquad (2.2.24)$$

where $\boldsymbol{\Psi} > \mathbf{0}$ is a symmetric $n \times n$ matrix, $\mathbb{y}$ is an $n \times 1$ vector of constants and $L > 0$, has a solution

$$\widehat{\mathbb{c}} = \boldsymbol{\Phi}\widehat{\mathbb{d}} \qquad (2.2.25)$$

where $\boldsymbol{\Phi}$ is an orthogonal $n \times n$ matrix from Schur decomposition C.2.2

$$\boldsymbol{\Psi} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}' \qquad (2.2.26)$$

where

$$\begin{aligned}
\boldsymbol{\Lambda} &= diag\{\lambda_1,\ldots,\lambda_n\}, \qquad (2.2.27) \\
\lambda_i &> 0, \quad i = 1,\ldots,n, \qquad (2.2.28) \\
\mathbf{I} &= \boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Phi}' \qquad (2.2.29)
\end{aligned}$$

and $\widehat{\mathbb{d}} = (\hat{d}_1,\ldots,\hat{d}_n)'$ solves

$$\min_{\mathbb{d}\in\mathbb{R}^n} \frac{1}{n} \sum_{i=1}^n (\lambda_i d_i - z_i)^2 \quad s.t. \ \sum_{i=1}^n \lambda_i d_i^2 \leq L \qquad (2.2.30)$$

where $\mathbb{z} = (z_1,\ldots,z_n)' = \boldsymbol{\Phi}'\mathbb{y}$. Vector $\widehat{\mathbb{d}}$—solution of (2.2.30)*—always exists.*

Proof. From Schur decomposition C.2.2 we know that there are such matrices $\mathbf{\Phi}$ and $\mathbf{\Lambda}$, because according to Lemma 2.2.3 $\mathbf{\Psi}$ is symmetric. $\mathbf{\Psi}$ is positive definite—Theorem 2.2.4. Hence all the eigenvalues are positive, so $\lambda_i > 0$ for all i. Let $\mathbb{d} := \mathbf{\Phi}'\mathbb{c}$ and $\mathbb{z} := \mathbf{\Phi}'\mathbb{y}$. Try to calculate

$$\begin{aligned}\sum_{i=1}^{n} (\lambda_i d_i - z_i)^2 &\equiv \|\mathbf{\Lambda}\mathbb{d} - \mathbb{z}\|_2^2 = [\mathbf{\Lambda}\mathbb{d} - \mathbb{z}]' \mathbf{\Phi}'\mathbf{\Phi} [\mathbf{\Lambda}\mathbb{d} - \mathbb{z}] \\ &= \|\mathbf{\Phi}\mathbf{\Lambda}\mathbb{d} - \mathbf{\Phi}\mathbb{z}\|_2^2 = \|\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}'\mathbb{c} - \mathbf{\Phi}\mathbf{\Phi}'\mathbb{y}\|_2^2 \\ &= \|\mathbf{\Psi}\mathbb{c} - \mathbb{y}\|_2^2 \equiv [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}],\end{aligned} \tag{2.2.31}$$

where $\|\cdot\|_2$ denotes the classic Euclidean norm in $\mathbb{R}^n$ and

$$\begin{aligned}\sum_{i=1}^{n} \lambda_i d_i^2 &= \mathbb{d}'\mathbf{\Lambda}\mathbb{d} = \mathbb{d}'\mathbf{\Phi}'\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}'\mathbf{\Phi}\mathbb{d} = (\mathbf{\Phi}\mathbb{d})' (\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}') (\mathbf{\Phi}\mathbb{d}) \\ &= \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L.\end{aligned} \tag{2.2.32}$$

The existence of $\widehat{\mathbb{d}}$ is trivial, because the objective function $\frac{1}{n}\sum_{i=1}^{n} (\lambda_i d_i - z_i)^2$ is continuous in $\mathbb{d}$ and set $\mathcal{S} := \{\mathbb{d} \in \mathbb{R}^n : \sum_{i=1}^{n} \lambda_i d_i^2 \leq L\}$ is closed and bounded in $\mathbb{R}^n$, i.e. compact set. Finally $\widehat{\mathbb{c}} := \mathbf{\Phi}\widehat{\mathbb{d}}$. □

Remark 2.2.6. We could heighten the proposition of Theorem 2.2.5 and add the uniqueness of $\widehat{\mathbb{d}}$ and hence the uniqueness of $\widehat{\mathbb{c}}$. We show this in Theorem 2.2.6 and its proof. We should not add any restrictions to the assumptions of Theorem 2.2.5 to avoid pathological (degenerating) cases, because none can become (see proof of Theorem 2.2.6).

Remark 2.2.7. This optimizing problem in Theorem 2.2.5 can be also solved using SVD Theorem C.2.3 as it can be found in Golub and Loan (1996).

Theorem 2.2.6 (Minimizing Problem). *Let $L > 0$, $\lambda_i > 0$ for all $i = 1, \ldots, n$ and define $\mathcal{S} := \{\mathbb{d} \in \mathbb{R}^n : \sum_{i=1}^{n} \lambda_i d_i^2 \leq L\}$ a feasible set. Solution of*

$$\widehat{\mathbb{d}} = \underset{\mathbb{d}\in\mathbb{R}^n}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} (\lambda_i d_i - z_i)^2 \quad s.t. \quad \sum_{i=1}^{n} \lambda_i d_i^2 \leq L \tag{2.2.33}$$

can be obtained from Algorithm 2.2.1. This algorithm is finite and always provides a unique result.

Proof. Function $\frac{1}{n}\sum_{i=1}^{n} (\lambda_i d_i - z_i)^2$ is continuous in $\mathbb{R}^n$. It can take its local minimum at the point where all the first partial derivatives $\frac{\partial}{\partial d_i}$ are equal zero or lie on the boundary of the feasible set $\mathcal{S}$.

Algorithm 2.2.1 Quadratic Minimizing with Quadratic Constraint

Input: $\mathbb{z}$, $\boldsymbol{\lambda}$, L
Output: $\mathbb{d}$
Require: $L > 0$ and $\boldsymbol{\lambda} > \mathbf{0}$
if $\left(\frac{z_1}{\lambda_1}, \ldots, \frac{z_n}{\lambda_n}\right)' \in \mathcal{S}$ **then**
  $\hat{d}_i \leftarrow \frac{z_i}{\lambda_i} \quad \forall i$
else
  $\vartheta^* \leftarrow \text{solve}\left(\sum_{i=1}^n \frac{z_i^2}{(\lambda_i + n\vartheta)^2}\lambda_i = L\right)$
  $\hat{d}_i \leftarrow \frac{z_i}{\lambda_i + n\vartheta^*} \quad \forall i$
end if

Objective function takes its minimal value 0 only at one point $\left(\frac{z_1}{\lambda_1}, \ldots, \frac{z_n}{\lambda_n}\right)'$. It is exactly at the same point where all the first partial derivatives $\frac{\partial}{\partial z_i}$ are equal zero

$$\frac{\partial}{\partial d_i}\left[\frac{1}{n}\sum_{i=1}^{n}(\lambda_i d_i - z_i)^2\right] = 2\lambda_i(\lambda_i d_i - z_i) \stackrel{!}{=} 0 \quad \Rightarrow \quad d_i = \frac{z_i}{\lambda_i} \quad \forall i. \tag{2.2.34}$$

So if point $\left(\frac{z_1}{\lambda_1}, \ldots, \frac{z_n}{\lambda_n}\right)'$ belongs into the feasible set $\mathcal{S}$, we are done and this point is also one unique global minimum. If not, the local and also global minimum of our objective function must lie on the boundary $\partial\mathcal{S}$, because the feasible set $\mathcal{S}$ is a compact set. Then we need to solve

$$\min_{\mathbb{d}\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n}(\lambda_i d_i - z_i)^2 \quad \text{s.t.} \quad \sum_{i=1}^{n}\lambda_i d_i^2 = L \tag{2.2.35}$$

by using Lagrange multipliers. Define the Lagrange multiplier according to our objective function and feasible set

$$\mathcal{L}(\mathbb{d}, \vartheta) = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i d_i - z_i)^2 + \vartheta\left(\sum_{i=1}^{n}\lambda_i d_i^2 - L\right). \tag{2.2.36}$$

Let's find extreme of the Lagrange multiplier

$$\frac{\partial}{\partial d_i}\mathcal{L}(\mathbb{d}, \vartheta) = 2\frac{\lambda_i}{n}(\lambda_i d_i - z_i) + 2\vartheta\lambda_i d_i \stackrel{!}{=} 0 \quad \Rightarrow \quad d_i = \frac{z_i}{\lambda_i + n\vartheta} \quad \forall i. \tag{2.2.37}$$

To determine the Lagrange parameter we define

$$\xi(\vartheta) := \sum_{i=1}^{n}\lambda_i d_i^2(\vartheta) = \sum_{i=1}^{n}\frac{\lambda_i z_i^2}{(\lambda_i + n\vartheta)^2} \tag{2.2.38}$$

and seek the solution to $\xi(\vartheta) = L$. We know that

$$\left(\frac{z_1}{\lambda_1}, \ldots, \frac{z_n}{\lambda_n}\right)' \notin \mathcal{S}. \tag{2.2.39}$$

Thus

$$\xi(0) = \sum_{i=1}^{n} \lambda_i \left(\frac{z_i}{\lambda_i}\right)^2 > L. \tag{2.2.40}$$

Now $\xi(\vartheta)$ is monotone decreasing for $\vartheta > 0$, and $\sum_{i=1}^{n} \frac{z_i^2}{\lambda_i} > L$ therefore implies the existence of a unique positive ϑ^* for which $\xi(\vartheta^*) = L$. So we have only one unique ϑ^* which gives a unique $\hat{d}_i$ for all i (local and global minimum).
Finiteness of the algorithm is trivial, because it does not contain any loops. □

Remark 2.2.8. If the first condition in previous Algorithm 2.2.6 is true then our optimizing problem was only an interpolation. We cannot talk about approximation or estimation. This case is not important for us, because it does not turn up in the real situation.

Note 2.2.9. At this moment we finally find the procedure of determination of regression function $\widehat{f}$ (see Remark 2.2.4). The quadratic optimizing with constraints provides $\hat{c}_j$. Riesz Representation Theorem C.1.3 gives us θ_k. The boundary conditions of specific differential equation provides γ_k.

Theorem 2.2.7 (Asymptotic Behavior of Finite Optimizing Solution). *Let $\widehat{f}$ satisfy $s^2 = \min_{f\in\mathcal{H}^m} \frac{1}{n}\sum_{i=1}^{n} [Y_i - f(\mathbb{X}_i)]^2$. Suppose $f \in \mathcal{H}^m$, then:*

i) $s^2 \stackrel{a.s.}{\to} \sigma_\varepsilon^2, \quad n \to \infty,$

ii) $\frac{1}{n}\sum_{i=1}^{n} \left[\widehat{f}(\mathbb{X}_i) - f(\mathbb{X}_i)\right]^2 = O_p(n^{-r}), \quad$ *where* $r = \frac{2m}{2m+q}$,

iii) $n^{1/2}\left[s^2 - \sigma_\varepsilon^2\right] \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \mathbb{V}ar(\varepsilon^2)\right), \quad n \to \infty.$

Proof. Proof has been properly proceeded in Bos and Yatchew (1997). Powerful Kolmogorov-Tihomirov Theorem C.3.1 (see Kolmogorov and Tihomirov (1959)) and some important results from de Geer (1990) are applied. □

Remark 2.2.10. $\mathbb{V}\mathrm{ar}\,(\varepsilon^2)$ may be estimated consistently using the fourth order moments of the estimated residuals $\widehat{\varepsilon_i} = Y_i - \widehat{f}(\mathbb{X}_i)$ as we can find in Yatchew (2000).

2.3 Weighted Least Squares

For some reasons it is sometimes necessary to emphasize some observations, because some of them could be more important then the other ones. That's why we "weight" observation data.

Definition 2.3.1 (Penalizing Using Weighted Least Squares). Optimizing using Weighted Least Squares is

$$\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{y}-\boldsymbol{f}(\mathbb{x})\right]'\boldsymbol{\Lambda}\left[\mathbb{y}-\boldsymbol{f}(\mathbb{x})\right] \quad \text{s.t.} \quad \|f\|^2_{Sob,m}\leq L \tag{2.3.41}$$

where $\mathbb{x}$ is an $n\times 1$ vector of q-dimensional vector data points $\mathbb{x}_1,\ldots,\mathbb{x}_n$, $\boldsymbol{\Lambda}$ is an $n\times n$ positive definite matrix, $\mathbb{y}$ is an $n\times 1$ vector of constants, f is a real function of a real value, $\boldsymbol{f}(\mathbb{x})=(f(\mathbb{x}_1),\ldots,f(\mathbb{x}_n))'$ and $L>0$.

Theorem 2.3.1 (Weighted Infinite to Finite). *Let* $\mathbb{y}=(y_1,\ldots,y_n)'$, $\boldsymbol{\Lambda}$ *an* $n\times n$ *positive definite matrix and define*

$$\hat{\sigma}^2 = \min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{y}-\boldsymbol{f}(\mathbb{x})\right]'\boldsymbol{\Lambda}\left[\mathbb{y}-\boldsymbol{f}(\mathbb{x})\right] \quad s.t. \quad \|f\|^2_{Sob,m}\leq L, \tag{2.3.42}$$

$$s^2 = \min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n}\left[\mathbb{y}-\boldsymbol{\Psi}\mathbb{c}\right]'\boldsymbol{\Lambda}\left[\mathbb{y}-\boldsymbol{\Psi}\mathbb{c}\right] \quad s.t. \quad \mathbb{c}'\boldsymbol{\Psi}\mathbb{c}\leq L \tag{2.3.43}$$

where $\mathbb{c}$ *is an* $n\times 1$ *vector,* $\boldsymbol{f}$ *is declared in Definition 2.3.1 and* $\boldsymbol{\Psi}$ *is the representor matrix at* $\mathbb{x}_1,\ldots,\mathbb{x}_n$. *Then* $\hat{\sigma}^2=s^2$. *Furthermore, there exists a solution to* (2.3.42) *of the form*

$$\widehat{f}=\sum_{i=1}^{n}\hat{c}_i\psi_{\mathbb{x}_i} \tag{2.3.44}$$

where $\widehat{\mathbb{c}}=(\hat{c}_1,\ldots,\hat{c}_n)'$ *solves* (2.3.43). *The estimator* $\widehat{f}$ *is unique a.e.*

Proof. The idea of this proof is almost the same as in proof of Theorem 2.2.1. The only thing that we have to rewrite for correctness is this succession of equalities in (2.2.10):

$$
\begin{aligned}
&[\mathbb{y} - \boldsymbol{f}(\mathbb{x})]' \boldsymbol{\Lambda} [\mathbb{y} - \boldsymbol{f}(\mathbb{x})] \\
&= \left[\mathbb{y}_\bullet - \left\langle \psi_{\mathbb{x}_\bullet}, \sum_{j=1}^n c_j \psi_{x_j} + h \right\rangle_{Sob,m} \right]' \boldsymbol{\Lambda} \left[\mathbb{y}_\bullet - \left\langle \psi_{\mathbb{x}_\bullet}, \sum_{j=1}^n c_j \psi_{x_j} + h \right\rangle_{Sob,m} \right] \\
&= \left[\mathbb{y}_\bullet - \sum_{j=1}^n \left\langle \psi_{\mathbb{x}_\bullet}, c_j \psi_{x_j} \right\rangle_{Sob,m} \right]' \boldsymbol{\Lambda} \left[\mathbb{y}_\bullet - \sum_{j=1}^n \left\langle \psi_{\mathbb{x}_\bullet}, c_j \psi_{x_j} \right\rangle_{Sob,m} \right] \\
&= \left[\mathbb{y}_\bullet - \sum_{j=1}^n c_j \left\langle \psi_{\mathbb{x}_\bullet}, \psi_{x_j} \right\rangle_{Sob,m} \right]' \boldsymbol{\Lambda} \left[\mathbb{y}_\bullet - \sum_{j=1}^n c_j \left\langle \psi_{\mathbb{x}_\bullet}, \psi_{x_j} \right\rangle_{Sob,m} \right] \\
&= \left[\mathbb{y}_\bullet - \sum_{j=1}^n \Psi_{\bullet,j} c_j \right]' \boldsymbol{\Lambda} \left[\mathbb{y}_\bullet - \sum_{j=1}^n \Psi_{\bullet,j} c_j \right] = [\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}]' \boldsymbol{\Lambda} [\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}]
\end{aligned}
\tag{2.3.45}
$$

where for an arbitrary $g \in \mathcal{H}^m$

$$
\langle \psi_{\mathbb{x}_\bullet}, g \rangle_{Sob,m} = \Big(\langle \psi_{x_1}, g \rangle_{Sob,m}, \ldots, \langle \psi_{x_n}, g \rangle_{Sob,m} \Big)'. \tag{2.3.46}
$$

□

2.4 Multiple Observations

There are numbers of situations where multiple observations appear and we need to work with them. E.g. option price data often consist of multiple observations at a finite vector of strike prices (see Härdle and Yatchew (2003) for detailed set-up).

Let $\mathbb{X} = (X_1, \ldots, X_k)'$ be the vector of k distinct strike prices. We will assume that the vector $\mathbb{X}$ is in increasing order. Let $\sigma^2(X_1), \ldots, \sigma^2(X_k)$ be the residual variances at each of the distinct strike prices. Let $\boldsymbol{\Delta}$ be the $n \times k$ matrix such that

$$
\boldsymbol{\Delta}_{ij} := \begin{cases} 1 & \text{if } x_i = X_j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4.47}
$$

We may now rewrite our infinite optimizing problem as

$$
\min_{f \in \mathcal{H}^m} \frac{1}{n} [\mathbb{y} - \boldsymbol{\Delta}\boldsymbol{f}(\mathbb{X})]' \boldsymbol{\Sigma}^{-1} [\mathbb{y} - \boldsymbol{\Delta}\boldsymbol{f}(\mathbb{X})] \quad \text{s.t.} \quad \|f\|^2_{Sob,m} \leq L \tag{2.4.48}
$$

where $\mathbb{y}$ is an $n \times 1$ vector of strike prices, $\boldsymbol{\Sigma} > \mathbf{0}$ is a symmetric $n \times n$ matrix and $\boldsymbol{f}(\mathbb{X})$ is defined as in Definition 2.3.1, so it is also an $k \times 1$ vector of values. Nothing else that the representor matrix $\mathbf{\Psi}$ is in this case $k \times k$, the analogue to finite quadratic optimizing becomes:

$$\min_{\mathbb{c}\in\mathbb{R}^k} \frac{1}{n} [\mathbb{y} - \mathbf{\Delta\Psi}\mathbb{c}]' \mathbf{\Sigma}^{-1} [\mathbb{y} - \mathbf{\Delta\Psi}\mathbb{c}] \quad \text{s.t. } \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \tag{2.4.49}$$

where $\mathbb{c}$ is an $k \times 1$ vector of constants.

We can clearly see that this model is only a little extention of the Weighted Least Squares model declared in Section 2.3. The same properties remain for this Model with Multiple Observations, too. But we have to realize that the typical random design model cannot be appropriate for this set-up. We will discuss this model later in Theorem 4.1.2.

2.5 Multi-Equation Model

Definition 2.5.1 (Multi-Equation Model). The multi-equation model is

$$\mathbb{Y}_i = \mathbb{f}(\mathbb{X}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n \tag{2.5.50}$$

with these assumptions:

i) $\mathbb{X}_i$ are q-dimensional random vectors, i.i.d. with probability law $\mathcal{P}_x$ and density p_x bounded away from zero on the support $\mathcal{Q}^q$, the unit cube in $\mathbb{R}^q$;

ii) $\boldsymbol{\varepsilon}_i$ are p-dimensional i.i.d. random variables with probability law $\mathcal{P}_{\boldsymbol{\varepsilon}}$ so that $\mathbb{E}\boldsymbol{\varepsilon}_i = \mathbf{0}$ for all i and the covariance matrix is $\mathbb{V}\mathrm{ar}\,\boldsymbol{\varepsilon}_i = \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$; $\mathcal{P}_{\boldsymbol{\varepsilon}} \in \mathscr{P}_{\boldsymbol{\varepsilon}}$ a collection of probability laws with mean $\mathbf{0}$ and support contained in a bounded interval of $\mathbb{R}^q$; $\mathbb{X}_i$ and $\boldsymbol{\varepsilon}_i$ are independent;

iii) $\mathscr{F}$, a family of functions from $\mathbb{R}^q$ to $\mathbb{R}^p$, is a cross-product of the Sobolev balls

$$\mathscr{F} = \left\{ \mathbb{f} = (f_1, \ldots, f_p)' \,|\, f_j \in \mathcal{H}^m(\mathcal{Q}^q), \|f_j\|^2_{Sob,m} \leq L_j, j = 1, \ldots, p \right\} \tag{2.5.51}$$

where $m > \frac{q}{2}$.

Note 2.5.1. $\boldsymbol{f}$ and $\mathbb{f}$ do not mean the same! $\boldsymbol{f}$ is a vector of function values and $\mathbb{f}$ is a vector function.

Theorem 2.5.1 (Extensions to the Multi-Equation Settings). *Let* $\mathbb{x}_1, \ldots, \mathbb{x}_n$ *be* $q \times 1$ *vectors,* $\mathbb{y} = (y_1, \ldots, y_n)'$, $\mathbf{\Xi}$ *be a positive definite matrix* *and define*

$$\hat{\sigma}^2 = \min_{\mathbb{f}\in\mathcal{H}^m} \frac{1}{n}\sum_{i=1}^{n} [\mathbb{y}_i - \mathbb{f}(\mathbb{x}_i)]' \, \boldsymbol{\Xi} \, [\mathbb{y}_i - \mathbb{f}(\mathbb{x}_i)]$$
$$s.t. \quad \|f_j\|^2_{Sob,m} \le L_j, \quad j = 1,\dots,p \quad (2.5.52a)$$

$$s^2 = \min_{\boldsymbol{C}\in\mathbb{R}^{n\times p}} \frac{1}{n}\sum_{i=1}^{n} \left[\mathbb{y}_i - \boldsymbol{C}'\begin{pmatrix}\Psi_{i1}\\ \vdots \\ \Psi_{in}\end{pmatrix}\right]' \boldsymbol{\Xi} \left[\mathbb{y}_i - \boldsymbol{C}'\begin{pmatrix}\Psi_{i1}\\ \vdots \\ \Psi_{in}\end{pmatrix}\right]$$
$$s.t. \quad (C_{1j},\dots,C_{nj})\boldsymbol{\Psi}\begin{pmatrix}C_{1j}\\ \vdots \\ C_{nj}\end{pmatrix} \le L_j, \quad j = 1,\dots,p. \quad (2.5.52b)$$

where $\boldsymbol{C}$ is an $n\times p$ matrix and $\boldsymbol{\Psi}$ is the representor matrix of inner products of the $\psi_{\mathbb{x}_1},\dots,\psi_{\mathbb{x}_n}$. Then $\hat{\sigma}^2 = s^2$. Furthermore, there exists a solution of optimizing problem of the form $\widehat{\mathbb{f}} = (\widehat{f}_1,\dots,\widehat{f}_p)'$, $\widehat{f}_j = \sum_{i=1}^{n} \widehat{C_{ij}}\psi_{\mathbb{x}_i}$ for $j = 1,\dots,p$, where $\widehat{\boldsymbol{C}}$ solves finite dimensional problem (2.5.52b)*. The estimator $\widehat{\mathbb{f}}$ is unique a.e.*

Proof. This proof tends to be very similar to the proof of Theorem 2.2.1. Let $M = span\{\psi_{\mathbb{x}_i} : i = 1,\dots,n\}$ and its orthogonal complement $M^{\perp} = \left\{h \in \mathcal{H}^m : \langle\psi_{\mathbb{x}_i}, h\rangle_{Sob,m} = 0,\, i = 1,\dots,n\right\}$. Representors exist by Theorem 1.4.1 and we can write the Sobolev space as a direct sum of its orthogonal subspaces, i.e. $\mathcal{H}^m = M \oplus M^{\perp}$ since $\mathcal{H}^m$ is a Hilbert space. Functions $h \in M^{\perp}$ take on the value zero vector at $\mathbb{x}_1,\dots,\mathbb{x}_n$. Consider a vector function $\mathbb{f} = (f_1,\dots,f_p)'$. Each f_i can be written in the form

$$f_j = \sum_{i=1}^{n} C_{ij}\psi_{\mathbb{x}_i} + h_j, \quad h_j \in M^{\perp}, \quad j = 1,\dots,p. \quad (2.5.53)$$

Then

$$f_j(\mathbb{x}_k) = \left\langle \psi_{\mathbb{x}_k}, \sum_{i=1}^{n} C_{ij}\psi_{\mathbb{x}_i} + h_j \right\rangle_{Sob,m} = \sum_{i=1}^{n} \langle\psi_{\mathbb{x}_k}, C_{ij}\psi_{\mathbb{x}_i}\rangle_{Sob,m}$$
$$= \sum_{i=1}^{n} C_{ij}\langle\psi_{\mathbb{x}_k}, \psi_{\mathbb{x}_i}\rangle_{Sob,m} = \sum_{i=1}^{n} \Psi_{ki}C_{ij}, \quad k = 1,\dots,n, \quad j = 1,\dots,p. \quad (2.5.54)$$

Note further that

$$\begin{aligned}\|f_j\|^2_{Sob,m} &= \langle f_j, f_j\rangle_{Sob,m}\\ &= \left\langle \sum_{i=1}^n C_{ij}\psi_{\mathbb{x}_i}, \sum_{i=1}^n C_{ij}\psi_{\mathbb{x}_i}\right\rangle_{Sob,m} + \langle h_j, h_j\rangle_{Sob,m}\\ &= (C_{1j},\ldots,C_{nj})\boldsymbol{\Psi}\begin{pmatrix}C_{1j}\\ \vdots\\ C_{nj}\end{pmatrix} + \langle h_j, h_j\rangle_{Sob,m}, \quad j=1,\ldots,p.\end{aligned} \tag{2.5.55}$$

Suppose that $\mathbb{f}$ minimizes

$$\frac{1}{n}\sum_{i=1}^n [\mathbb{y}_i - \mathbb{f}(\mathbb{x}_i)]' \boldsymbol{\Xi} [\mathbb{y}_i - \mathbb{f}(\mathbb{x}_i)] \ \text{ s.t. } \|f_j\|^2_{Sob,m} \le L_j, \quad j=1,\ldots,p \tag{2.5.56}$$

then so does $\mathbb{f}^* = \mathbb{f} - \mathbb{h}$ is zero vector at $\mathbb{x}_1,\ldots,\mathbb{x}_n$ where

$$\mathbb{f} = \boldsymbol{C}'\begin{pmatrix}\psi_{\mathbb{x}_1}\\ \vdots\\ \psi_{\mathbb{x}_n}\end{pmatrix} + \begin{pmatrix}h_1\\ \vdots\\ h_p\end{pmatrix}, \qquad \mathbb{f}^* = \boldsymbol{C}'\begin{pmatrix}\psi_{\mathbb{x}_1}\\ \vdots\\ \psi_{\mathbb{x}_n}\end{pmatrix}. \tag{2.5.57}$$

Hence, there exists a function $\mathbb{f}^*$ minimizing the infinite dimensional optimizing problem that is a linear combination of the representors. Furthermore,

$$\left\|f_j^*\right\|^2_{Sob,m} \le \left\|f_j^*\right\|^2_{Sob,m} + \|h_j\|^2_{Sob,m} = \|f_j\|^2_{Sob,m} \le L_j, \quad j=1,\ldots,p. \tag{2.5.58}$$

We note also that $\left\|f_j^*\right\|^2_{Sob,m} = (C_{1j},\ldots,C_{nj})\boldsymbol{\Psi}(C_{1j},\ldots,C_{nj})'$. Finally, we observe that the following two problems are equivalent:

$$\min_{\mathbb{f}^*} \frac{1}{n}\sum_{i=1}^n [\mathbb{y}_i - \mathbb{f}^*(\mathbb{x}_i)]' \boldsymbol{\Xi} [\mathbb{y}_i - \mathbb{f}^*(\mathbb{x}_i)]$$

$$\text{s.t. } \left\|f_j^*\right\|^2_{Sob,m} \le L_j, \quad j=1,\ldots,p, \quad \mathbb{f}^* \in M; \tag{2.5.59a}$$

$$\min_{\boldsymbol{C}} \frac{1}{n}\sum_{i=1}^n \left[\mathbb{y}_i - \boldsymbol{C}'\begin{pmatrix}\Psi_{i1}\\ \vdots\\ \Psi_{in}\end{pmatrix}\right]' \boldsymbol{\Xi} \left[\mathbb{y}_i - \boldsymbol{C}'\begin{pmatrix}\Psi_{i1}\\ \vdots\\ \Psi_{in}\end{pmatrix}\right]$$

$$\text{s.t. } (C_{1j},\ldots,C_{nj})\boldsymbol{\Psi}\begin{pmatrix}C_{1j}\\ \vdots\\ C_{nj}\end{pmatrix} \le L_j, \quad j=1,\ldots,p. \tag{2.5.59b}$$

Uniqueness is clear, since $\psi_{\mathbb{x}_i}$ are base elements of M, $\left\|\widehat{\mathbb{f}}\right\|^2_{Sob,m} = L$ and adding a function that is orthogonal to the spaces spanned by the representors will increase the norm. □

2.6 Selection of Sobolev Norm Bounds

In nonparametric least squares the smoothing parameter L corresponds to the diameter of the set of functions over which the estimation takes place. Heuristically, the larger are the bounds (much larger than true norm), the less efficient estimators we obtain in spite of the fact that they will be consistent. On the other hand, the smaller bounds the more efficient estimators we have but inconsistent.

2.6.1 Cross-Validation

Definition 2.6.1 (Cross-Validation Function). We define the Cross-Validation function

$$\mathcal{CV}(L) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \widehat{f}_{-i}(\mathbb{x}_i)\right]^2 \tag{2.6.60}$$

where $\widehat{f}_{-i}$ is obtained by solving

$$\min_{f\in\mathcal{H}^m}\sum_{\substack{j=1\\ j\neq i}}^{n}\left[y_j - f(\mathbb{x}_j)\right]^2 \quad \text{s.t.} \quad \|f\|^2_{Sob,m} \leq L. \tag{2.6.61}$$

The idea of selection of the smoothing parameter by Cross-Validation is based on its ability to predict outside the sample. We omit the i-th observation from the estimation when the i-th observation is being predicted. Then we use the minimum of the Cross-Validation function $\mathcal{CV}$ to estimate the smoothing parameter (Sobolev bound L). Relationship between the data-fit and the smoothness of estimator is shown in Figure 2.6.1.

Theorem 2.6.1 (1–1 Mapping of Smoothing Parameter). *Let $L > 0$, $\mathbf{\Sigma}$ is positive definite and symmetric matrix and*

$$f^* = \arg\min_{f\in\mathcal{H}^m}\frac{1}{n}\left[\mathbb{y} - \boldsymbol{f}(\mathbb{x})\right]'\mathbf{\Sigma}^{-1}\left[\mathbb{y} - \boldsymbol{f}(\mathbb{x})\right] \quad s.t. \quad \|f\|^2_{Sob,m} = L \tag{2.6.62}$$

then there exists a unique $\chi > 0$ such that

$$f^* = \arg\min_{f\in\mathcal{H}^m}\frac{1}{n}\left[\mathbb{y} - \boldsymbol{f}(\mathbb{x})\right]'\mathbf{\Sigma}^{-1}\left[\mathbb{y} - \boldsymbol{f}(\mathbb{x})\right] + \chi\|f\|^2_{Sob,m}. \tag{2.6.63}$$

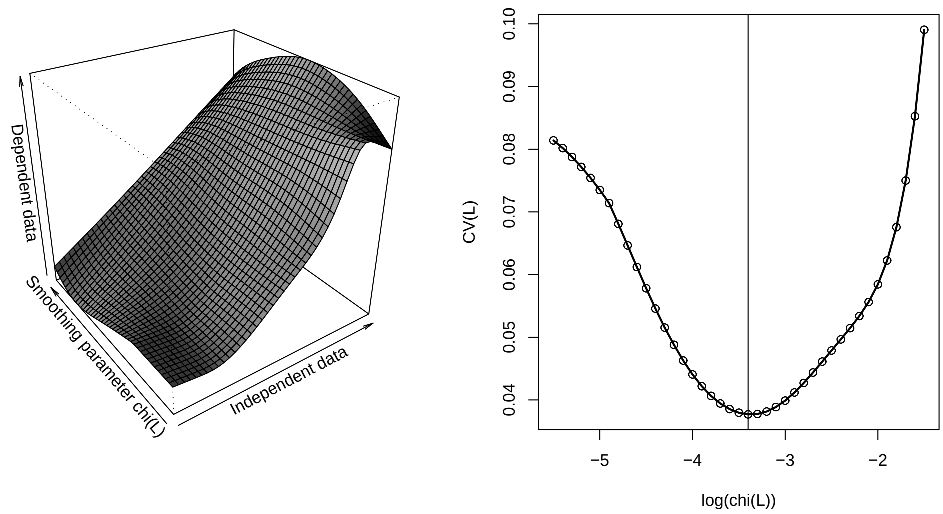

Figure 2.6.1: Left—changing monotone curve in $\mathcal{H}^2$ depending upon smoothing parameter. Right—optimal value of smoothing parameter according to Cross-Validation.

Proof. The solution of (2.6.62) always exists and is unique according to the proof of Theorem 2.3.1 (just alter some inequalities to equalities). From the same proof of Theorem 2.3.1 follows that finding f^*—optimizing (2.6.62)—is the same as searching optimal $\mathbb{c}^*$ such that

$$\mathbb{c}^* = \arg\min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right]' \mathbf{\Sigma}^{-1}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}'\mathbf{\Psi}\mathbb{c} = L \tag{2.6.64}$$

and again from the proof of Theorem 2.3.1 the existence and the uniqueness of $\mathbb{c}^*$ is guaranteed. Let's fix L. If $\mathbb{c}^{*\prime}\mathbf{\Psi}\mathbb{c}^* = L$, we can simply apply Lagrange Multiplier Theorem on our bond condition $\mathbb{c}'\mathbf{\Psi}\mathbb{c} = L$ using the Lagrange function

$$\mathcal{J}(\mathbb{c},\lambda) = \frac{1}{n}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right]' \mathbf{\Sigma}^{-1}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right] + \lambda\left(\mathbb{c}'\mathbf{\Psi}\mathbb{c} - L\right) \tag{2.6.65}$$

and it provides us a unique Lagrange multiplier χ. We do not care about $-\chi L$ because it does not depend on $\mathbb{c}$.

Quadratic form $\mathcal{J}(\cdot,\lambda)$ have to be positive definite according Lagrange Multiplier Theorem (we are minimizing $\mathcal{J}$). That implies $\chi > 0$. □

Remark 2.6.1. There exists a 1–1 mapping $\mathscr{X}\colon \mathbb{R}^+ \to \mathbb{R}^+\colon L \mapsto \chi$.

Remark 2.6.2. There has to be an equality not an inequality in condition of (2.6.62), because for the sharp inequality we are able to find lots of $\chi > 0$. If $\mathbb{c}^{*\prime}\mathbf{\Psi}\mathbb{c}^* < L$, then we are talking about the interpolation not the approximation because $\mathbb{y} = \mathbf{\Psi}\mathbb{c}^*$. This is very unusual case for a real statistical situation and our problem, too. See also Remark 2.2.8.

Proof. Global extreme of function

$$\mathcal{I}(\mathbb{c}) = \frac{1}{n}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right]' \mathbf{\Sigma}^{-1}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right] \tag{2.6.66}$$

is also its local extreme on open set $\mathcal{T} = \{\mathbb{c} \in \mathbb{R}^n | \mathbb{c}'\mathbf{\Psi}\mathbb{c} < L\}$. Hence all the partial derivatives need to be equal zero in the optimal point $\mathbb{c}^*$. Using the matrix notation of the partial derivatives, positive definiteness and symmetry of the matrices $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ we get:

$$\frac{\partial \mathcal{I}(\mathbb{c})}{\partial c_i} \stackrel{!}{=} 0 \quad \forall i \qquad \Longleftrightarrow \qquad \mathbb{y} = \mathbf{\Psi}\mathbb{c}. \tag{2.6.67}$$

□

Remark 2.6.3. At this moment we have shown that the penalization

a)
$$\min_{f \in \mathcal{H}^m} \frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(\mathbb{x}_i)\right]^2 \text{ s.t. } \|f\|^2_{Sob,m} \leq L, \tag{2.6.68}$$

b)
$$\min_{f \in \mathcal{H}^m} \left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(\mathbb{x}_i)\right]^2 + \chi\|f\|^2_{Sob,m}\right\} \tag{2.6.69}$$

have "almost" the same meaning.

2.6.2 Other Methods

Cross-Validation is basic Leave-One-Out method for choosing a smoothing parameter in the nonparametric regression. There are a lot of other methods based on penalizing functions or plug-in selectors. Specific types of "smoothing choosers"—such as Generalized Cross-Validation, Akaike's Information Criterion, Finite Prediction Error, Shibata's model selector or Rice's bandwidth selector—can be found in Härdle (1990).

Chapter 3

Isotonia

In regression in Sobolev spaces we have demanded only smoothness constraint on the regression function $f \in \mathscr{F} = \left\{ f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|^2_{Sob,m} \leq L \right\}$. Now our estimators should also underlie additional constraints. We therefore focus on the imposition of additional constraint—isotonia—on nonparametric regression estimation and testing of this constraint. Two basic types of isotonia are monotonicity and convexity/concavity so we will concentrate mostly on them.

We would like to estimate to the subject $f \in \widetilde{\mathscr{F}} \subseteq \mathscr{F}$ where $\widetilde{\mathscr{F}}$ combines smoothness with further functional properties and to test $H_0 : f \in \widetilde{\mathscr{F}}$.

We examine general constrained submodel from Bos and Yatchew (1997). Then I have set up this model for isotonia and declared two main attitudes to isotonia—definite and indefinite.

3.1 Constrained Submodel

Definition 3.1.1 (Constrained Single Equation Model). Invoke the assumptions for the Single Equation Model 2.1.1 and add these assumptions:

iv) $\widetilde{\mathscr{F}} \subseteq \mathscr{F}$ is a closed set of functions such that the metric entropy $\log N(\delta; \mathscr{F}) \leq A\delta^{-\zeta}$ for some $A > 0, \zeta > 0$;

v) $\left\{ \widetilde{\mathscr{F}_n} \right\}_{n=1}^{\infty}$ is a descending sequence of the closed and possibly random sets of functions $\mathscr{F} \supseteq \widetilde{\mathscr{F}_1} \supseteq \ldots \supseteq \widetilde{\mathscr{F}_1} \supseteq \ldots \supseteq \widetilde{\mathscr{F}}$ such that $\bigcap_{n=1}^{\infty} \widetilde{\mathscr{F}_n} = \widetilde{\mathscr{F}}$ a.s. and $\log N(\delta; \widetilde{\mathscr{F}_n}) \leq A'\delta^{-\zeta}$, $n = 1, 2, \ldots$ for some $A' > 0$.

Remark 3.1.1. Metric entropy $H(\delta; \mathscr{F}) := \log N(\delta; \mathscr{F})$, where $N(\delta; \mathscr{F})$ defines the minimal number of points of a δ-net in $\mathscr{F}$ so here it denotes the minimum number of balls of radius δ in supnorm required to cover the set of functions $\mathscr{F}$.

Theorem 3.1.1 (Convergence of Constrained Estimation). *Let $\hat{f}$ satisfy $s^2 = \min_{f \in \mathcal{H}^m} \sum_i^n \frac{1}{n} \left[Y_i - f(\mathbb{X}_i)\right]^2$ s.t. $f \in \widetilde{\mathscr{F}}_n$. If $f \in \widetilde{\mathscr{F}}$ then the conclusion of Theorem 2.2.7 (Asymptotic Behavior of Finite Optimizing Solution) continue to hold with rate of convergence $\eta = \frac{2m}{2m+q}$. Suppose $f \notin \widetilde{\mathscr{F}}$, $\|f\|^2_{Sob,m}$ is finite and there exists a unique $\widetilde{f} \in \widetilde{\mathscr{F}}$ satisfying $\min_{g \in \widetilde{\mathscr{F}}} \int (f-g)^2 \, dP_x$. Then $s^2 \xrightarrow{a.s.} \sigma^2_\varepsilon + \int \left(f - \widetilde{f}\right)^2 dP_x$, $n \to \infty$.*

Proof. Proof has also been properly proceeded in Bos and Yatchew (1997). □

Definition 3.1.2 (Derivative of Representor Matrix). Let $\psi_{x_1}, \ldots, \psi_{x_n}$ be the representors for function evaluation at $x_1, \ldots, x_n$ respectively, i.e. $\langle \psi_{x_i}, f \rangle_{Sob,m} = f(x_i)$ for all $f \in \mathcal{H}^m(\mathcal{Q}^1)$, $i = 1, \ldots, n$. Let $\boldsymbol{\Psi}^{(1)}$ be the first derivative of $n \times n$ representor matrix whose columns are equal to the first derivatives of the representors evaluated at $x_1, \ldots, x_n$; i.e.

$$\Psi^{(1)}_{i,j} = \psi^{(1)}_{x_j}(x_i), \quad i,j = 1, \ldots, n. \tag{3.1.1}$$

Remark 3.1.2. It is very important that the derivative of the representor matrix is defined in a "column" way. In spite of Theorem 2.2.3, derivative of representor matrix needn't to be a symmetric one.

3.2 Monotonicity

Definition 3.2.1 (Definite Monotonicity). Optimizing with Smoothness and Definite Monotonicity Constraint is

$$\min_{\mathbb{c} \in \mathbb{R}^n} \quad \frac{1}{n} \left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right]' \left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right] \tag{3.2.2}$$

$$\text{s.t.} \quad \mathbb{c}' \boldsymbol{\Psi} \mathbb{c} \leq L \tag{3.2.3}$$

$$\boldsymbol{\Psi}^{(1)} \mathbb{c} \geq \mathbf{0} \tag{3.2.4}$$

where $\boldsymbol{\Psi}$ is an $n \times n$ representor matrix at the data points $x_1, \ldots, x_n$, $\boldsymbol{\Psi}^{(1)}$ is the first derivative of representor matrix at the data points $x_1, \ldots, x_n$, $\mathbb{y}$ is an $n \times 1$ vector of constants and $L > 0$.

Remark 3.2.1. We should call Definition 3.2.1 Minimizing with Smoothness and Definite Non-decreasing Constraint for correctness.

Conjecture 3.2.2. Definition 3.2.1 determines a set of functions

$$\widetilde{\mathscr{F}_n} := closure\left\{f \in \mathcal{H}^m(\mathcal{Q}^1) : \|f\|^2_{Sob,m} \leq L,\, f'(x_i) \geq 0,\, i = 1, \ldots, n\right\}. \tag{3.2.5}$$

Definition 3.2.3 (Indefinite Monotonicity). Optimizing with Smoothness and Indefinite Monotonicity Constraint is

$$\min_{\mathbb{c}\in\mathbb{R}^n} \quad \frac{1}{n}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right]'\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right] \tag{3.2.6}$$

$$\text{s.t.} \quad \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \tag{3.2.7}$$

$$\left[\mathbf{\Psi}\mathbb{c}\right]_i \leq \left[\mathbf{\Psi}\mathbb{c}\right]_j \quad \text{for } x_i \leq x_j, \quad i, j = 1, \ldots, n \tag{3.2.8}$$

where $\mathbf{\Psi}$ is an $n \times n$ representor matrix at the data points $x_1, \ldots, x_n$, $\mathbb{y}$ is an $n \times 1$ vector of constants and $L > 0$.

Remark 3.2.2. We should call Definition 3.2.3 Minimizing with Smoothness and Indefinite Non-decreasing Constraint for correctness.

Conjecture 3.2.4. Definition 3.2.3 determines a set of functions

$$\widetilde{\mathscr{F}_n} := closure \left\{f \in \mathcal{H}^m(\mathcal{Q}^1) : \|f\|^2_{Sob,m} \leq L,\, f(x_i) \leq f(x_j),\, x_i \leq x_j,\, i, j = 1, \ldots, n\right\}. \tag{3.2.9}$$

Remark 3.2.3. Versions 3.2.1 and 3.2.3 are slightly different. Using the Mean Value Theorem, the second definition ensures us that the estimated derivative is positive (non-negative) at some point between each pair of the consecutive points. The first one requires the estimated derivatives to be positive (non-negative) at the points $x_1, \ldots, x_n$. Neither procedure ensures that the estimated function is monotone everywhere in small samples but, as data accumulate, the smoothness requirement prevents non-monotonicity.

3.3 Convexity and Concavity

Definition 3.3.1 (Second Derivative of Representor Matrix). Let $\psi_{x_1}, \ldots, \psi_{x_n}$ be the representors for function evaluation at $x_1, \ldots, x_n$ respectively, i.e. $\langle \psi_{x_i}, f\rangle_{Sob,m} = f(x_i)$ for all $f \in \mathcal{H}^m(\mathcal{Q}^1)$, $i = 1, \ldots, n$. Let $\mathbf{\Psi}^{(2)}$ be the second derivative of $n \times n$ representor matrix whose columns are equal to the second derivatives of representors evaluated at $x_1, \ldots, x_n$; i.e.

$$\Psi^{(2)}_{i,j} = \psi^{(2)}_{x_j}(x_i), \quad i, j = 1, \ldots, n. \tag{3.3.10}$$

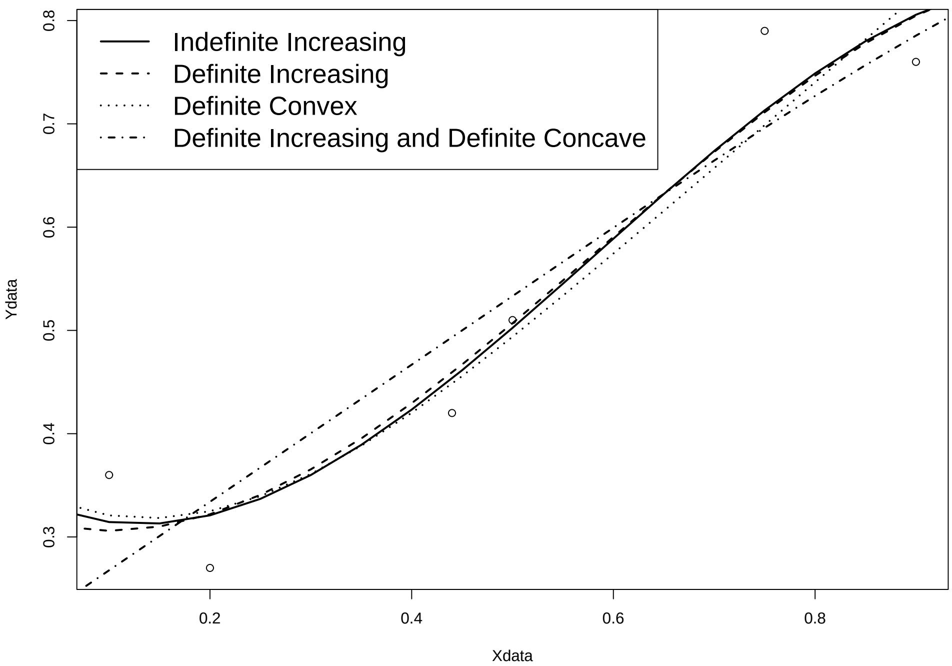

Figure 3.3.1: Various isotonic curves in $\mathcal{H}^4$. Data set is from Table 4.2.

Definition 3.3.2 (Definite Convexity). Optimizing with Smoothness and Definite Convexity Constraint is

$$\min_{\mathbb{c}\in\mathbb{R}^n} \quad \frac{1}{n}\left[\mathbb{y}-\boldsymbol{\Psi}\mathbb{c}\right]'\left[\mathbb{y}-\boldsymbol{\Psi}\mathbb{c}\right] \tag{3.3.11}$$

$$\text{s.t.} \quad \mathbb{c}'\boldsymbol{\Psi}\mathbb{c} \leq L \tag{3.3.12}$$

$$\boldsymbol{\Psi}^{(2)}\mathbb{c} \geq \mathbf{0} \tag{3.3.13}$$

where $\boldsymbol{\Psi}$ is an $n \times n$ representor matrix at the data points $x_1, \ldots, x_n$, $\boldsymbol{\Psi}^{(2)}$ is the second derivative of representor matrix at the data points $x_1, \ldots, x_n$, $\mathbb{y}$ is an $n \times 1$ vector of constants and $L > 0$.

Remark 3.3.1. Analogically we can also define Indefinite Convexity, Definite Concavity and Indefinite Concavity. In Figure 3.3.1 we can see various types of the isotonic estimators in the Sobolev space $\mathcal{H}^4$.

Chapter 4

Asymptotic Behavior

Our last chapter in theoretical approach concerns the asymptotic behavior of our estimators, e.g. its convergence in probability and in distribution.

We show asymptotic behavior based upon Strong Law of Large Numbers and Central Limit Theorem 4.1.2 (proved in Härdle and Yatchew (2003)). Then we examine the residual regression test for the regression with constraint from Bos and Yatchew (1997) and apply this test to the isotonia. Finally, we acquaint ourselves with bootstrap techniques from Yatchew (2003) and proceed to the construction of confidence intervals and test of isotonia in a bootstrap way.

4.1 Confidence Intervals

Theorem 4.1.1 (Consistency of Estimator). *Suppose* $Y_i = f(X_i) + \varepsilon_i$, *where* ε_i *are independently distributed with* $\mathbb{V}ar(\varepsilon_i) = \sigma_i^2$ *and* $0 < \sigma_i^2 < K$ *for some* K. X_i *are i.i.d. with continuous density* p_x *on* $[a, b]$ *bounded away from zero. Then*

$$\sup_{x \in [a,b]} \left| \widehat{f^{(s)}}(x) - f^{(s)}(x) \right| \xrightarrow{\mathcal{P}} 0, \quad n \to \infty \quad \textit{for} \quad s = 0, \ldots, m-2 \qquad (4.1.1)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left(f(X_i) - \widehat{f}(X_i) \right)^2 = \mathcal{O}_p \left(n^{-\frac{2m}{2m+1}} \right), \quad n \to \infty. \qquad (4.1.2)$$

Proof. Theorem 2.2.7 says that $\widehat{f}$ converges to f in mean squared error at the indicated rate of convergence. By Theorem 1.5.1, all the functions in the estimating set have derivatives up to order $m-1$ uniformly bounded in supnorm. If the first derivatives are uniformly bounded, the convergence in mean square implies the convergence of $\hat{f}$ to f in supnorm. In fact, if $(m-1)$-th derivatives are uniformly bounded, then this ensures that $\widehat{f^{(s)}}$ for $s \le m-2$ converges in supnorm. □

We do not discuss the random (stochastic) design model for a while. We propose a special type of the fixed design model in Theorem 4.1.2.

We focus on Multiple Observations Model in Section 2.4, defined in (2.4.48). Even if the number of distinct strike prices k does not increase, the call function can be estimated consistently at $X_1, \ldots, X_k$. This does not assure that the estimates of derivatives are estimated consistently. Indeed, no "nonparametric" estimator can consistently estimate derivatives at a data point without accumulation of observations in the neighborhood of the point.

Theorem 4.1.2 (Asymptotic Behavior Based upon Laws of Large Numbers and Central Limit Theorem). *Given data* $\{(x_i, Y_i)\}_{i=1}^n$ *where* $Y_i = f(x_i) + \varepsilon_i$*, the* ε_i *are independently distributed and* x_i *are sampled from a discrete distribution whose support is* $X_1, \ldots, X_k$ *with corresponding probabilities* $\pi_1, \ldots, \pi_k$*. Suppose that* f *lies strictly inside the ball of functions* $\|f\|^2_{Sob,m} < L$ *and* f *is strictly increasing and strictly convex. Let* $\bar{\boldsymbol{Y}}(\mathbb{X}) = \left(\bar{Y}_1(X_1), \ldots, \bar{Y}_k(X_k)\right)'$ *be the* k*-dimensional vector of average observations at the* k *values. Let* $\boldsymbol{\Pi}/n = \mathbb{V}ar(\bar{\boldsymbol{Y}}(\mathbb{X}))$ *be the* $k \times k$ *diagonal matrix of variances of the point means estimators, i.e.* $\Pi_{jj} = \sigma^2(X_j)/\pi_j$, $j = 1, \ldots, k$*. Then*

$$i)\quad \mathcal{P}\left[\widehat{\boldsymbol{f}}(\mathbb{X}) = \bar{\boldsymbol{Y}}(\mathbb{X})\right] \xrightarrow{n\to\infty} 1, \tag{4.1.3}$$

$$ii)\quad \widehat{\boldsymbol{f}}(\mathbb{X}) \xrightarrow[n\to\infty]{\mathcal{P}} \boldsymbol{f}(\mathbb{X}), \tag{4.1.4}$$

$$iii)\quad \widehat{\boldsymbol{f}^{(1)}}(\mathbb{X}) \xrightarrow[n\to\infty]{\mathcal{P}} \boldsymbol{f}^{(1)}(\mathbb{X}), \tag{4.1.5}$$

$$iv)\quad \widehat{\boldsymbol{f}^{(2)}}(\mathbb{X}) \xrightarrow[n\to\infty]{\mathcal{P}} \boldsymbol{f}^{(2)}(\mathbb{X}), \tag{4.1.6}$$

$$v)\quad \sqrt{n}\left(\widehat{\boldsymbol{f}}(\mathbb{X}) - \boldsymbol{f}(\mathbb{X})\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Pi}), \tag{4.1.7}$$

$$vi)\quad \sqrt{n}\,(\widehat{\mathbb{c}} - \mathbb{c}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\boldsymbol{\Psi}^{-1}), \tag{4.1.8}$$

$$vii)\quad \sqrt{n}\left(\widehat{\boldsymbol{f}^{(1)}}(\mathbb{X}) - \boldsymbol{f}^{(1)}(\mathbb{X})\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Psi}^{(1)}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{(1)}), \tag{4.1.9}$$

$$viii)\quad \sqrt{n}\left(\widehat{\boldsymbol{f}^{(2)}}(\mathbb{X}) - \boldsymbol{f}^{(2)}(\mathbb{X})\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\Psi}^{(2)}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{(2)}). \tag{4.1.10}$$

Remark 4.1.1. Proof of this Theorem can be found in Härdle and Yatchew (2003).

Proof. All we have to realize that $\mathbf{\Psi}^{-1}\bar{\boldsymbol{Y}}(\mathbb{X})$ solves the unconstrained optimizing problem (2.4.49):

$$\min_{\mathbb{c}\in\mathbb{R}^k} \frac{1}{n}\left[\mathbb{y}-\mathbf{\Delta\Psi}\mathbb{c}\right]'\mathbf{\Sigma}^{-1}\left[\mathbb{y}-\mathbf{\Delta\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}'\mathbf{\Psi}\mathbb{c}\leq L. \tag{4.1.11}$$

Now let $\widehat{\mathbb{c}}$ minimize (4.1.11) subject to (3.2.4) and (3.3.13). Since f is a linear combination of the representors

$$\hat{f}(x)=\sum_{j=1}^{k}\hat{c}_j\psi_{X_j}(x) \tag{4.1.12}$$

and its first two derivatives consistently estimate their true counterparts (see Theorem 4.1.1). We also know

$$\left[\widehat{\boldsymbol{f}}(\mathbb{X})\right]'\mathbf{\Psi}^{-1}\widehat{\boldsymbol{f}}(\mathbb{X})\xrightarrow{\mathcal{P}}\left[\boldsymbol{f}(\mathbb{X})\right]'\mathbf{\Psi}^{-1}\boldsymbol{f}(\mathbb{X})<L, \quad n\to\infty \tag{4.1.13}$$

and as sample size increases, the smoothness constraint become non-binding in probability. According to Strong Law of Large Numbers

$$\mathcal{P}\left[\widehat{\boldsymbol{f}}(\mathbb{X})=\bar{\boldsymbol{Y}}(\mathbb{X})\right]\xrightarrow{n\to\infty}1. \tag{4.1.14}$$

Using conventional Central Limit Theorems, we obtain

$$\sqrt{n}\left(\widehat{\boldsymbol{f}}(\mathbb{X})-\boldsymbol{f}(\mathbb{X})\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\mathbf{\Pi}), \quad n\to\infty \tag{4.1.15}$$

and equations (4.1.8), (4.1.9) and (4.1.10) follow immediately. □

4.2 Tests of Hypothesis

Our interest is also concentrated on testing the main property of our estimator, isotonia. Smoothness of our estimator is certain by its construction. The null hypothesis to be tested is that our regression function is isotonic:

$$H_0:\ f\in\widetilde{\mathscr{F}}=\{f\in\mathscr{F}|f \text{ is isotonic}\}, \tag{4.2.16}$$

while, without a specific alternative model, the alternative to be tested will be that the null is false:

$$H_1:\ f\in\mathscr{F}\backslash\widetilde{\mathscr{F}}. \tag{4.2.17}$$

Remark 4.2.1. Particular null hypothesis instead of the general one declared in (4.2.16) can be e.g.

$$H_0:\ f\in\widetilde{\mathscr{F}}=\{f\in\mathscr{F}|f \text{ is increasing \& convex}\}. \tag{4.2.18}$$

Let's focus only on the one-dimensional case (e.g. $q = 1$)—in higher dimension we can hardly define isotonic properties. We dwell on the residual regression test. The idea of our test is as follows. Under H_0, since $\mathbb{E}_\varepsilon\left[Y - \hat{f}(X)|X\right] = 0$, we have

$$\mathbb{E}_{\varepsilon,X}\left\{\left(Y - \hat{f}(X)\right)\mathbb{E}_\varepsilon\left[Y - \hat{f}(X)|X\right]p_x(X)\right\} = 0, \tag{4.2.19}$$

while under H_1, since $\mathbb{E}_\varepsilon\left[Y - \hat{f}(X)|X\right] = f(X) - \hat{f}(X)$, we have

$$\begin{aligned} & \mathbb{E}_{\varepsilon,X}\left\{\left(Y - \hat{f}(X)\right)\mathbb{E}_\varepsilon\left[Y - \hat{f}(X)|X\right]p_x(X)\right\} && (4.2.20)\\ = \; & \mathbb{E}_{\varepsilon,X}\left\{\left(\mathbb{E}_\varepsilon\left[Y - \hat{f}(X)|X\right]\right)^2 p_x(X)\right\} && (4.2.21)\\ = \; & \mathbb{E}_X\left[\left(f(X) - \widehat{f}(X)\right)^2 p_x(X)\right] && (4.2.22)\\ > \; & 0. && (4.2.23)\end{aligned}$$

The sharp inequality in (4.2.23) follows from the fact that under the alternative H_1 is $f \in \mathscr{F}\backslash\widetilde{\mathscr{F}}$, so $\mathcal{P}_X\left[f(X) \neq \hat{f}(X)\right] < 1$.

Define a special modification of a standard one-sample "second-order" U-statistic

$$U_n = \frac{1}{n(n-1)\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(Y_i - \widehat{f}(X_i)\right)\left(Y_j - \widehat{f}(X_j)\right)K\left(\frac{X_i - X_j}{\lambda}\right), \tag{4.2.24}$$

where $K(\cdot)$ is a kernel with the common smoothing parameter λ. We assume that the underlying univariate kernel is symmetric having support $[-1, +1]$. We can convert (4.2.24) into

$$U_n = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{f}(X_i)\right)\left[\frac{1}{(n-1)\lambda}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(Y_j - \widehat{f}(X_j)\right)K\left(\frac{X_i - X_j}{\lambda}\right)\right], \tag{4.2.25}$$

where the term in the square brackets may be thought of as an estimator of $\left(f(X_i) - \widetilde{f}(X_i)\right)p_X(X_i)$ such that $\widetilde{f}$ has been declared in Theorem 3.1.1.

We can also expand (4.2.25) into

$$\begin{aligned}
U_n &= U_{n,1} + U_{n,2} + U_{n,3} \\
&= \frac{1}{n(n-1)\lambda} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varepsilon_i \varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right) \\
&+ \frac{1}{n(n-1)\lambda} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(f(X_i) - \widehat{f}(X_i)\right)\left(f(X_j) - \widehat{f}(X_j)\right) K\left(\frac{X_i - X_j}{\lambda}\right) \\
&+ \frac{2}{n(n-1)\lambda} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varepsilon_i \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\frac{X_i - X_j}{\lambda}\right).
\end{aligned} \tag{4.2.26}$$

Theorem 4.2.1 (Test of Isotonia). *Suppose $f \in \widetilde{\mathscr{F}}$, $n\lambda \to \infty$ and $n^{1-\eta}\lambda^{1/2} \to 0$ where $\eta = \frac{2m}{2m+1}$ is the rate of the convergence of the restricted estimator. Then*

$$n\lambda^{1/2} U_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\sigma_\varepsilon^4 \int p^2(x)dx \int K^2(v)dv\right), \quad n \to \infty. \tag{4.2.27}$$

Let the estimated variance of U be given by:

$$\hat{\sigma}^2_{U_n} = \frac{2}{n^2(n-1)^2\lambda^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(Y_i - \widehat{f}(X_i)\right)^2 \left(Y_j - \widehat{f}(X_j)\right)^2 K^2\left(\frac{X_i - X_j}{\lambda}\right). \tag{4.2.28}$$

Then

$$n(n-1)\lambda\hat{\sigma}^2_{U_n} \xrightarrow{\mathcal{P}} 2\sigma_\varepsilon^4 \int p^2(x)dx \int K^2(v)dv, \quad n \to \infty. \tag{4.2.29}$$

Hence

$$\frac{U_n}{\hat{\sigma}_{U_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty. \tag{4.2.30}$$

Remark 4.2.2. We demonstrate an implementation of Theorem 4.2.1 in Algorithm 4.2.1 and also in Subsection A.6.1. Our set-up is the same as in Remark 4.2.1.

Remark 4.2.3. Proof of this Theorem 4.2.1 can be found in Bos and Yatchew (1997). We only proceed some steps differently or more properly for better understanding.

Proof. We divide the proof into five steps.

i) Variance of $U_{n,1}$
Note that

$$\mathbb{E}(U_{n,1}) = \mathbb{E}\left[\frac{1}{n(n-1)\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)\right] = 0 \tag{4.2.31}$$

according to the assumptions of independence in Definition 2.1.1 and $i \neq j$. Then $\mathbb{V}\mathrm{ar}\, U_{n,1} = \mathbb{E}U_{n,1}^2$. To obtain this, we need to calculate

$$\begin{aligned}
& \mathbb{E}\left[\frac{\varepsilon_i^2\varepsilon_j^2}{\lambda}K^2\left(\frac{X_i - X_j}{\lambda}\right)\right] && (4.2.32)\\
= \quad & \sigma_\varepsilon^4 \int_{\mathcal{Q}^n}\frac{1}{\lambda}K^2\left(\frac{x_i - x_j}{\lambda}\right)\prod_{k=1}^{n}p_x(x_k)\mathrm{d}\mathbb{x} && (4.2.33)\\
= \quad & \sigma_\varepsilon^4 \int_0^1\int_0^1 \frac{1}{\lambda}K^2\left(\frac{x_i - x_j}{\lambda}\right)p_x(x_i)p_x(x_j)\mathrm{d}x_i\mathrm{d}x_j && (4.2.34)\\
= \quad & \sigma_\varepsilon^4 \int_0^1 p_x(x_i)\left\{\int_{-x_j/\lambda}^{(1-x_j)/\lambda}K^2(v)p_x(\lambda v + x_j)\mathrm{d}v\right\}\mathrm{d}x_j && (4.2.35)\\
\xrightarrow{\lambda\to 0} \quad & \sigma_\varepsilon^4 \int_0^1 p_x^2(x_i)\mathrm{d}x_i\int_{-1}^{+1}K^2(v)\mathrm{d}v. && (4.2.36)
\end{aligned}$$

We have substituted $v := \frac{x_i - x_j}{\lambda}$ in (4.2.35) and thus $\mathrm{d}v = \frac{\mathrm{d}x_i}{\lambda}$. In (4.2.36) we have sent λ into 0, used Lebesgue's dominated convergence theorem and realized that $K(\cdot) = 0$ outside $[-1, +1]$. Hence

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\,(n\lambda^{1/2}U_{n,1})\\
&= \mathbb{E}\left[\frac{1}{(n-1)\lambda^{1/2}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)\right]^2\\
&= \frac{1}{(n-1)^2\lambda}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)\right]^2\\
&+ \frac{1}{(n-1)^2\lambda}\sum_{i=1}^{n}\sum_{\substack{k=1\\ k\neq i}}^{n}\mathbb{E}\left[\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)\right]\left[\sum_{\substack{j=1\\ j\neq k}}^{n}\varepsilon_k\varepsilon_j K\left(\frac{X_k - X_j}{\lambda}\right)\right]\\
&= \frac{1}{(n-1)^2\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}\left[\varepsilon_i^2\varepsilon_j^2 K^2\left(\frac{X_i - X_j}{\lambda}\right)\right]\\
&+ \frac{1}{(n-1)^2\lambda}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{\substack{k=1\\ k\neq i}}^{n}\varepsilon_i\varepsilon_k K\left(\frac{X_i - X_k}{\lambda}\right)\right]^2\\
&\overset{i\neq j}{=} \frac{2}{(n-1)^2}n(n-1)\mathbb{E}\left[\frac{\varepsilon_i^2\varepsilon_j^2}{\lambda}K^2\left(\frac{X_i - X_j}{\lambda}\right)\right]\\
&\xrightarrow{n\to\infty} 2\sigma_\varepsilon^4\int p_x^2\mathrm{d}x\int K^2(v)\mathrm{d}v,
\end{aligned}
\tag{4.2.37}
$$

because $\lambda \equiv \lambda_n$.

ii) Distribution of $U_{n,1}$

Let's define

$$
Z_{n,i} := \frac{\lambda^{1/2}}{\sigma_\varepsilon^2\sqrt{2\int p_x(x)\mathrm{d}x\int K^2(v)\mathrm{d}v}}\varepsilon_i\left[\frac{2}{(n-1)\lambda}\sum_{1\leq j<i}\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)\right]
\tag{4.2.38}
$$

and rewrite

$$
U_{n,1} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{2}{(n-1)\lambda}\sum_{1\leq j<i}\varepsilon_j K\left(\frac{X_i - X_j}{\lambda}\right)
\tag{4.2.39}
$$

where λ and $K\left(\frac{X_i - X_j}{\lambda}\right)$ are implicitly functions of n. Note that

$$\sum_{i=1}^{n} Z_{n,i} = \frac{n\lambda^{1/2} U_{n,1}}{\sigma_\varepsilon^2 \sqrt{2 \int p_x(x)\mathrm{d}x \int K^2(v)\mathrm{d}v}}. \qquad (4.2.40)$$

Due to the assumptions for Constrained Single Equation Model 3.1 in Definition 3.1.1, all the propositions in McLeish Theorem C.4.1 (see McLeish (1974)) are satisfied. There has been used terminology of Feller in Theorem C.4.1. In our set-up according to Definition 3.1.1, (C.4.12) means nothing else than the convergence of variances. Applying mentioned theorem, we conclude

$$\sum_{i=1}^{n} Z_{n,i} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty \qquad (4.2.41)$$

and hence

$$n\lambda^{1/2} U_{n,1} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\sigma_\varepsilon^4 \int p_x(x)\mathrm{d}x \int K^2(v)\mathrm{d}v\right), \quad n \to \infty. \qquad (4.2.42)$$

iii) Rate of Convergence of $U_{n,2}$ and $U_{n,3}$

Define $n \times n$ matrix $\boldsymbol{K}$ whose entries are $K_{ij} = K\left(\frac{X_i - X_j}{\lambda}\right)$, $i, j = 1, \ldots, n$. The matrix $\frac{\boldsymbol{K}}{n\lambda}$ when multiplied by a column vector of ones and evaluated in $\mathbb{x} = (x_1, \ldots, x_n)'$, yields the kernel estimate of the density p_x of $\mathbb{X} = (X_1, \ldots, X_n)'$ at $X_1 = x_1, \ldots, X_n = x_n$. The matrix is symmetric and may be decomposed according Eigen Decomposition Theorem C.2.1 as

$$\frac{\boldsymbol{K}}{n\lambda} = \boldsymbol{\Phi\Upsilon\Phi}' \qquad (4.2.43)$$

where $\boldsymbol{\Phi}$ is orthogonal and $\boldsymbol{\Upsilon}$ is the diagonal matrix of eigenvalues. For given n the matrix $\boldsymbol{K}$ and hence the eigenvalues are determined by the sequence $X_1, \ldots, X_n$. Under our assumptions for Constrained Single Equation Model 3.1 in Definition 3.1.1 for X_is, $\max|\Upsilon_{ii}|$ is bounded in probability. Similarly, if $\boldsymbol{K}^2$ is the matrix whose elements are the squares of the elements of $\boldsymbol{K}$, then $\frac{\boldsymbol{K}^2}{n\lambda}$ is symmetric with the largest eigenvalue bounded in probability.

Let $\boldsymbol{f}(\mathbb{X}) - \widehat{\boldsymbol{f}}(\mathbb{X})$ be the $n \times 1$ vector with elements $f(X_i) - \widehat{f}(X_i)$. Since exists k such that

$$\mathcal{P}\left[\max|\Upsilon_{ii}| > k\right] \xrightarrow{n\to\infty} 0 \qquad (4.2.44)$$

we have with probability going to one:

$$\begin{aligned} U_{n,2} &= \frac{1}{n}\left(\boldsymbol{f}(\mathbb{X})-\widehat{\boldsymbol{f}}(\mathbb{X})\right)'\boldsymbol{\Phi\Upsilon\Phi}'\left(\boldsymbol{f}(\mathbb{X})-\widehat{\boldsymbol{f}}(\mathbb{X})\right) && (4.2.45)\\ &\leq \frac{1}{n}\left(\boldsymbol{f}(\mathbb{X})-\widehat{\boldsymbol{f}}(\mathbb{X})\right)'\boldsymbol{\Phi}k\boldsymbol{I}\boldsymbol{\Phi}'\left(\boldsymbol{f}(\mathbb{X})-\widehat{\boldsymbol{f}}(\mathbb{X})\right) && (4.2.46)\\ &= \frac{k}{n}\sum_{i=1}^{n}\left(f(X_i)-\widehat{f}(X_i)\right)^2 && (4.2.47)\end{aligned}$$

in which case,

$$U_{n,2} = \mathcal{O}_p\left(\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-\widehat{f}(X_i)\right)^2\right), \quad n\to\infty. \tag{4.2.48}$$

Using Kolmogorov-Tihomirov Theorem C.3.1 it can be shown that exists $A > 0$ such that for $\delta > 0$, we have $\log N(\delta;\mathscr{F}) < A\delta^{-1/m}$. Consequently applying Lemma 3.5 from de Geer (1990), we obtain that there exist positive constants C_0, K_0 such that for all $K > K_0$

$$\mathcal{P}\left[\sup_{\|g\|^2_{Sob,m}\leq L}\frac{\sqrt{n}\left|-\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i\left(f(X_i)-g(X_i)\right)\right|}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-g(X_i)\right)^2\right)^{\frac{1}{2}-\frac{1}{4m}}}\geq KA^{1/2}\right]\leq\exp\left\{-C_0K^2\right\}. \tag{4.2.49}$$

Since $f\in\mathscr{F}$ and $\widehat{f}$ minimizes the sum of squared residuals over $g\in\mathscr{F}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\widehat{f}(X_i)\right]^2 \leq \frac{1}{n}\sum_{i=1}^{n}\left[Y_i-g(X_i)\right]^2, \quad g\in\mathscr{F} \tag{4.2.50}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\left(f(X_i)-\widehat{f}(X_i)\right)+\varepsilon_i\right]^2 \leq \frac{1}{n}\sum_{i=1}^{n}\left[\left(f(X_i)-g(X_i)\right)+\varepsilon_i\right]^2, \quad g\in\mathscr{F}$$

$$\Downarrow \quad \text{realize that } f\in\mathscr{F}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-\widehat{f}(X_i)\right)^2 \leq -\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i\left(f(X_i)-\widehat{f}(X_i)\right). \tag{4.2.51}$$

Now combine (4.2.49) and (4.2.51) to obtain the result that $\forall K > K_0$

$$\mathcal{P}\left[\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-\widehat{f}(X_i)\right)^2\geq\left(\frac{K^2A}{n}\right)^{\frac{2m}{2m+1}}\right]\leq\exp\left\{-C_0K^2\right\}. \tag{4.2.52}$$

Thus

$$\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-\widehat{f}(X_i)\right)^2 = \mathcal{O}_p\left(n^{-\frac{2m}{2m+1}}\right), \quad n\to\infty. \tag{4.2.53}$$

Similarly,

$$U_{n,3} = \mathcal{O}_p\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left(f(X_i) - \widehat{f}(X_i)\right)\right) = \mathcal{O}_p\left(n^{-\frac{2m}{2m+1}}\right), \quad n \to \infty. \tag{4.2.54}$$

iv) Since by the assumptions, $n^{1-\eta}\lambda^{1/2} \to 0$, we have

$$n\lambda^{1/2}U_{n,2} \overset{\mathcal{P}}{\to} 0, \quad n \to \infty, \tag{4.2.55}$$

$$n\lambda^{1/2}U_{n,3} \overset{\mathcal{P}}{\to} 0, \quad n \to \infty, \tag{4.2.56}$$

in which case

$$n\lambda^{1/2}U_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, 2\sigma_\varepsilon^4 \int p_x(x)\mathrm{d}x \int K^2(v)\mathrm{d}v\right). \tag{4.2.57}$$

v) To show that

$$n(n-1)\lambda\hat{\sigma}^2_{U_n} \xrightarrow[n\to\infty]{\mathcal{P}} 2\sigma_\varepsilon^4 \int p_x(x)\mathrm{d}x \int K^2(v)\mathrm{d}v, \tag{4.2.58}$$

expand to obtain:

$$\begin{aligned} n(n-1)\lambda\hat{\sigma}^2_{U_n} &= \frac{2}{n(n-1)\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i^2\varepsilon_j^2 K^2\left(\frac{X_i - X_j}{\lambda}\right) \\ &+ \frac{2}{n(n-1)\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(f(X_i) - \widehat{f}(X_i)\right)^2\left(f(X_j) - \widehat{f}(X_j)\right)^2 K^2\left(\frac{X_i - X_j}{\lambda}\right) \\ &+ \frac{4}{n(n-1)\lambda}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varepsilon_i^2\left(f(X_j) - \widehat{f}(X_j)\right)^2 K^2\left(\frac{X_i - X_j}{\lambda}\right). \end{aligned} \tag{4.2.59}$$

The first term converges to $2\sigma_\varepsilon^4 \int p_x \int K^2$ from (4.2.36) and Strong Law of Large Numbers. Using arguments similar to part iii) above, the second and the third term converge to zero, in which case

$$n(n-1)\lambda\hat{\sigma}^2_{U_n} \xrightarrow[n\to\infty]{\mathcal{P}} 2\sigma_\varepsilon^4 \int p_x(x)\mathrm{d}x \int K^2(v)\mathrm{d}v. \tag{4.2.60}$$

Combining these results, we have

$$U_n/\hat{\sigma}_{U_n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{4.2.61}$$

□

Remark 4.2.4. If the null hypothesis is true, then the distribution of U_n (suitably standardized) is determined by the distribution of $U_{n,1}$, which is approximately normally distributed.

More elaborated approach that generalizes the above described has been examined in Zheng (1996), Yatchew (1998) and Fan and Li (1996).

Algorithm 4.2.1 Testing of Monotonicity and Convexity Based on Residuals

Input: m, α, $\{(x_i, y_i)\}_{i=1}^n$, χ, λ
Output: $U_n/\hat{\sigma}_{U_n}$ and **true**/**false**
Require: $\alpha \in (0,1)$, $\chi > 0$, $\lambda > 0$ *//χ could be also obtained using $\mathcal{CV}$*
Calculate representor matrix $\mathbf{\Psi}$ from (1.4.19).
Calculate $\widehat{\mathbb{c}}$ by solving (2.6.63) subject to (3.2.4) and (3.3.13). *//$\mathbf{\Sigma}$ can be identical matrix*
Perform the restricted regression of y on x from (2.2.8) to obtain $\widehat{f}(x)$.
for i **in** $1:n$ **do**
  $tmp \leftarrow |\boldsymbol{x} - x[i]|$
  $tmp \leftarrow (tmp \le \lambda)$
  $tmp\,[i] \leftarrow 0$
  $\boldsymbol{K}\,[i,] \leftarrow \frac{1}{2} tmp$
end for *//calculate the kernel matrix* $K\left(\frac{x_i - x_j}{\lambda}\right) = \frac{1}{2} I_{\{|x_i - x_j| \le \lambda\}}$
Determine U_n from (4.2.25).
Determine $\hat{\sigma}^2_{U_n}$ from (4.2.28).
Perform a one-sided test comparing $U_n/\hat{\sigma}_{U_n}$ on level $(1-\alpha) \times 100\%$ with the critical value from the $\mathcal{N}(0,1)$.

4.3 Bootstrap

The bootstrap is a resampling technique that prescribes taking "bootstrap samples" using the same random mechanism that generated the data.

In the case of the fixed design models with homoscedastic error structure, that are less general than the random ones, one may use only the estimated residuals (residuals resampling):

$$\hat{\varepsilon}_i = Y_i - \widehat{f}(X_i), \quad i = 1, \ldots, n. \tag{4.3.62}$$

Resampling them as $\left\{\hat{\varepsilon}_i^B\right\}_{i=1}^n$ gives bootstrap observations

$$Y_i^B = \widehat{f}(X_i) + \hat{\varepsilon}_i^B, \quad i = 1, \ldots, n. \tag{4.3.63}$$

We propose in our random design model only residuals resampling and we do not care about "naive bootstrap" $\left\{\left(\mathbb{X}_i^B, Y_i^B\right)\right\}_{i=1}^n$. Of course, (4.3.63) makes sense only if the error distribution does not depend on $\mathbb{X}$.

An alternative residual resampling methodology known as the "wild" or "external" bootstrap is useful particularly in heteroscedastic setting. Each bootstrap residual is drawn from the two-point distribution. So for each estimated residual $\hat{\varepsilon}_i = Y_i - \widehat{f}(X_i)$ one creates a two-point random distribution for a random variable (e.g. ω_i) with probabilities and properties as shown in Table 4.1. One then draws from this distribution to obtain $\hat{\varepsilon}_i^B$.

ω_i	$\mathcal{P}(\omega_i)$	$E(\omega_i)$	$E(\omega_i^2)$	$E(\omega_i^3)$
$\hat{\varepsilon}_i(1-\sqrt{5})/2$	$(5+\sqrt{5})/10$	0	$\widehat{\varepsilon}_i^2$	$\widehat{\varepsilon}_i^3$
$\hat{\varepsilon}_i(1+\sqrt{5})/2$	$(5-\sqrt{5})/10$	0	$\widehat{\varepsilon}_i^2$	$\widehat{\varepsilon}_i^3$

Table 4.1: A two-point distribution for "wild" or "external" bootstrap.

4.3.1 Bootstrap Confidence Intervals

Construction of confidence intervals and tests of hypothesis has been performed using the indicated asymptotic normal approximation. Alternatively, they may be implemented using the bootstrap as we describe below. Algorithm 4.3.1 outlines the implementation of the percentile bootstrap confidence interval at $f(x_0)$.

In Section B.1 and Section B.2 we practically construct various confidence intervals based on Theorem 4.1.2 and Algorithm 4.3.1. Implementation can be seen in Subsection A.5.1.

4.3.2 Bootstrap Tests of Hypothesis

Also tests of hypothesis may be implemented using the bootstrap procedures. Similar conclusion as in Theorem 4.2.1 can be induced for bootstrap.

Theorem 4.3.1. *Continuing with the assumptions of Theorem 4.2.1, holds*

$$U_n^B / \hat{\sigma}_{U_n}^B \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty. \tag{4.3.64}$$

Proof. This proof is similar to that of Theorem 4.2.1 and we can find it in Bos and Yatchew (1997). □

Implementation of Theorem 4.3.1 we can see in Algorithm 4.3.2, Example 4.3.1 and also in Subsection A.6.1.

Algorithm 4.3.1 Bootstrap Percentile Point-Wise Confidence Intervals for $\widehat{f}$, $\widehat{f^{(1)}}$ and $\widehat{f^{(2)}}$

Input: m, α, $\{(x_i, y_i)\}_{i=1}^n$, χ, b

Output: $\left[\widehat{f_{down}^{(s)}}(X_j), \widehat{f_{up}^{(s)}}(X_j)\right]$, $j = 1, \ldots, k$ and $s = 0, 1, 2$

Require: $\alpha \in (0,1)$, $\chi > 0$ *//χ could be also obtained using $\mathcal{CV}$*

$\boldsymbol{X} \leftarrow \boldsymbol{x}[1]$
for i **in** $\boldsymbol{x}$ **do**
  if $length(\boldsymbol{x}[\boldsymbol{x} = i]) = 0$ **then**
    $\boldsymbol{X} \leftarrow c(\boldsymbol{X}, i)$
  end if
end for *//determines values $\boldsymbol{X} = (X_1, \ldots, X_k)'$*
$k \leftarrow length(\boldsymbol{X})$
for i **in** $1 : n$ **do**
  for j **in** $1 : k$ **do**
    if $\boldsymbol{x}[i] = \boldsymbol{X}[j]$ **then**
      $\boldsymbol{\Delta}[i, j] = 1$
    else
      $\boldsymbol{\Delta}[i, j] = 0$
    end if
  end for
end for *//determines matrix $\boldsymbol{\Delta}$ according* (2.4.47)
Calculate $\widehat{\mathfrak{c}}$ from (2.4.49) applying amended Algorithm 2.2.1.
Perform the restricted regression of y on x from (2.2.8) to obtain $\widehat{f}(x)$.
Calculate the estimated residuals $\{\hat{\varepsilon}_i\}_{i=1}^n$ from (4.3.62).
for k **in** $1 : b$ **do**
  Construct a bootstrap data set $\left\{\left(x_i, y_i^B\right)\right\}_{i=1}^n$ from (4.3.63) where $\hat{\varepsilon}_i^B$ is obtained by sampling from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$ using the bootstrap (or "wild" bootstrap).
  Using the bootstrap data set obtain $\widehat{\mathfrak{c}}^B$ by solving (2.6.63).
  Calculate and save $\widehat{f}^B(X_j)$, $\left(\widehat{f^{(1)}}(X_j)\right)^B$ and $\left(\widehat{f^{(2)}}(X_j)\right)^B$.
end for *//b ... bootstrap range; repeat steps multiple times*
To obtain a $(1-\alpha) \times 100\%$ point-wise confidence intervals for $\widehat{f}$, $\widehat{f^{(1)}}$ and $\widehat{f^{(2)}}$ obtain $\alpha/2$ and $(1-\alpha/2)$ quantiles of the corresponding bootstrap estimates.

Algorithm 4.3.2 Bootstrap Residual Regression Test of Monotonicity and Convexity

Input: m, α, $\{(x_i, y_i)\}_{i=1}^n$, χ, λ, b
Output: $U_n^B/\hat{\sigma}_{U_n}^B$ and **true**/**false**
Require: $\alpha \in (0, 1)$, $\chi > 0$, $\lambda > 0$ *//χ could be also obtained using $\mathcal{CV}$*

Calculate representor matrix $\mathbf{\Psi}$ from (1.4.19).
Calculate $\widehat{\mathbb{c}}$ by solving (2.6.63) subject to (3.2.4) and (3.3.13) applying amended Algorithm 2.2.1.
Perform the restricted regression of y on x from (2.2.8) to obtain $\widehat{f}(x)$.
Calculate the estimated residuals $\{\hat{\varepsilon}_i\}_{i=1}^n$ from (4.3.62).
for i **in** $1 : n$ **do**
    $tmp \leftarrow |\boldsymbol{x} - x[i]|$
    $tmp \leftarrow (tmp \leq \lambda)$
    $tmp\,[i] \leftarrow 0$
    $\boldsymbol{K}\,[i,] \leftarrow \frac{1}{2} tmp$
end for *//calculate the kernel matrix* $K\left(\frac{x_i - x_j}{\lambda}\right) = \frac{1}{2} I_{\{|x_i - x_j| \leq \lambda\}}$
Determine U_n from (4.2.25).
Determine $\hat{\sigma}_{U_n}^2$ from (4.2.28).
for k **in** $1 : b$ **do**
    Construct a bootstrap data set $\left\{\left(x_i, y_i^B\right)\right\}_{i=1}^n$ from (4.3.63) where $\hat{\varepsilon}_i^B$ is obtained by sampling from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$ using the bootstrap (or "wild" bootstrap).
    Using the bootstrap data set, estimate the model under the null and calculate U_n^B and $\hat{\sigma}_{U_n}^B$.
    Calculate and save the standardized test statistic $U_n^B/\hat{\sigma}_{U_n}^B$.
    Define the bootstrap critical value for a $\alpha \times 100\%$ significance level test to be the $(1 - \alpha) \times 100$-th percentile of the $U_n^B/\hat{\sigma}_{U_n}^B$.
end for *//b . . . bootstrap range; repeat steps multiple times*
Compare $U_n/\hat{\sigma}_{U_n}$, the actual value of the statistic, with the bootstrap critical value.

Example 4.3.1 (Monotonicity and Convexity Residual Regression Test and its Bootstrap Versions). Given data set from Table 4.2.

Number of lines $\times 1000$	0.1	0.2	0.44	0.5	0.75	0.9
Time in seconds	0.35	0.28	0.42	0.51	0.79	0.77

Table 4.2: Translation time of the specific number of lines of this diploma thesis LaTeX 2_ε source code on my old *Pentium*® 120MHz with *FreeBSD 6.0* using *LaTeX 3.141592*.

i) Residual Regression Test
Follow Algorithm 4.2.1. Suppose order of Sobolev space $m = 4$. By solving (2.6.63) subject to (3.2.4) and (3.3.13), we easily obtain $\widehat{\mathbb{c}}$:

88.968305	-91.464147	-75.481021	8.381719	158.231554	-88.102874

We assume the kernel smoothing parameter $\lambda = 0.5590617$ (can be estimated as $n^{-1/5}[\max(\mathbb{x}) - \min(\mathbb{x})]$). Then we calculate vector $\left(U_n, \hat{\sigma}^2_{U_n}, U_n/\hat{\sigma}_{U_n}\right)'$ from (4.2.25) and (4.2.28) as we have implemented in Section A.6.1. For results see Table 4.3.

U_n	-8.956006×10^{-4}
$\hat{\sigma}^2_{U_n}$	3.468456×10^{-7}
$U_n/\hat{\sigma}_{U_n}$	-1.520710×10^{0}
$\Phi\left(\frac{U_n}{\hat{\sigma}_{U_n}}\right)$	6.416631×10^{-2}
$u(0.95)$	1.644854×10^{0}

Table 4.3: Results of asymptotic residual regression test of monotonicity and convexity. $\Phi(x)$ is distribution function of the standard normal distribution $\mathcal{N}(0,1)$ and $u(1-\alpha)$ is $(1-\alpha)\times 100\%$-quantile of standard normal distribution $\mathcal{N}(0,1)$.

Critical region for lower-tailed test is well-known interval $(-\infty, -1.645)$. Since $0.05 < 0.06416631$, with probability 95% we can say that we do not reject the null hypothesis and do not accept the alternative according to our asymptotic residual regression test and thus the observations from Table 4.2 can be considered as data with the increasing and convex character.

ii) Classical and Wild Bootstrap Version
Continue from part i) of this example and follow Algorithm 4.3.2. Hence we easily obtain results in Table 4.4:

Bootstrap resample number	$U_n^B / \hat{\sigma}_{U_n}^B$ classical	wild
10	-1.134556	-0.828338
20	-0.689638	-0.8126242
30	-1.050000	-0.6059807
40	-1.581657	-0.6981061
50	-1.649988	-1.153033
60	-0.5469642	-1.39981
70	-0.9886503	-0.5631235
80	-1.095642	-0.7864272
90	-0.9827762	-1.361833
100	-1.236391	-1.110179
mean	-0.97424928	-0.9559600
variance	0.08432592	0.0863544
95-th percentile	-0.56801890	-0.5081651
$\Phi(\cdot)$	0.1649664	0.1695462

Table 4.4: Results of classical and wild bootstrap residual regression test of monotonicity and convexity. $\Phi(x)$ is distribution function of the standard normal distribution $\mathcal{N}(0, 1)$.

Looking at the results, we have $0.05 \ll 0.1649664$ and $0.05 \ll 0.1695462$. Hence we obtain the same decision as in part i). With probability 95% we can say that we do not reject the null hypothesis and do not accept the alternative according to our “classical” and “wild” bootstrap residual regression test and thus the observations from Table 4.2 can be considered as data with the increasing and convex character.

Appendix A

Implementation into *R*

In this first chapter of the appendix we show the most important parts of *R* source code which has been used in this master thesis—mostly for generating figures that etch our theme of isotonic regression in the Sobolev spaces in.

There are lots of dropped hints in the *R* source code, which sometimes help to understand.

A.1 Calculations of Representors with *R*

A.1.1 Representor in Sobolev Space

To calculate the representor in the Sobolev space all we have to do is to follow Theorem 1.4.2. For better ideation we can look at (1.4.69).

We propose that some functions (procedures) have been already implemented, e.g. `derivsin()` which computes l-th derivative of function

$$\varphi_k(x) = \exp\left\{\left[\Re\left(\lambda_k\right)\right]x\right\}\sin\left\{\left[\Im\left(\lambda_k\right)\right]x\right\} \tag{A.1.1}$$

considering m-order Sobolev space or `relambda()` which computes the real part of complex number $e^{i\theta_k}$.

```
repre <- function(m=4,a=0.5) {
# computes coefficients of representors at point a
# creates Gamma matrix (4m x 4m)
# considering Sobolev space order=m and representor at point a from (0,1)
   dim <- 4*m
   Gamma <- matrix(0,nrow=dim,ncol=dim)
   kap <- kappa(m)
   # L_a^(m-j)(0)+(-1)^j*L_a^(m+j)(0)
   for (j in 0:(m-1)) {
      for (k in 0:(kap-1)) Gamma[j+1,k+1]
      <- derivcos(0,k,m,m-j)+((signum(j))*derivcos(0,k,m,m+j))
      for (k in (kap+1):m) Gamma[j+1,k]
      <- derivcos(0,k,m,m-j)+((signum(j))*derivcos(0,k,m,m+j))
```

```
      for (k in 0:(kap-1)) Gamma[j+1,m+k+1]
      <- derivsin(0,k,m,m-j)+((signum(j))*derivsin(0,k,m,m+j))
      for (k in (kap+1):m) Gamma[j+1,m+k]
      <- derivsin(0,k,m,m-j)+((signum(j))*derivsin(0,k,m,m+j))
      }
   # R_a^(m-j)(1)+(-1)^j*R_a^(m+j)(0)
   for (j in 0:(m-1)) {
      for (k in 0:(kap-1)) Gamma[m+1+j,2*m+1+k]
      <- derivcos(1,k,m,m-j)+((signum(j))*derivcos(1,k,m,m+j))
      for (k in (kap+1):m) Gamma[m+1+j,2*m+k]
      <- derivcos(1,k,m,m-j)+((signum(j))*derivcos(1,k,m,m+j))
      for (k in 0:(kap-1)) Gamma[m+1+j,3*m+1+k]
      <- derivsin(1,k,m,m-j)+((signum(j))*derivsin(1,k,m,m+j))
      for (k in (kap+1):m) Gamma[m+1+j,3*m+k]
      <- derivsin(1,k,m,m-j)+((signum(j))*derivsin(1,k,m,m+j))
      }
   # L_a^(j)(a)
   for (j in 0:(2*m-1)) {
      for (k in 0:(kap-1)) Gamma[2*m+1+j,k+1] <- derivcos(a,k,m,j)
      for (k in (kap+1):m) Gamma[2*m+1+j,k] <- derivcos(a,k,m,j)
      for (k in 0:(kap-1)) Gamma[2*m+1+j,m+k+1] <- derivsin(a,k,m,j)
      for (k in (kap+1):m) Gamma[2*m+1+j,m+k] <- derivsin(a,k,m,j)
      }
   # m odd ... real roots +-1
   if (m%%2 == 1) {
      for (j in 0:(m-1)) {
         Gamma[j+1,1] <- 1+signum(j) # 1=exp(0)
         Gamma[j+1,m+1] <- signum(m-j)+signum(j)*signum(m+j) # 1=exp(-0)
         Gamma[m+j+1,2*m+1] <- (1+signum(j))*exp(1)
         Gamma[m+j+1,3*m+1] <- exp(-1)*(signum(m-j)+signum(j)*signum(m+j))
         }
      for (j in 0:(2*m-1)) {
         Gamma[2*m+j+1,1] <- exp(a)
         Gamma[2*m+j+1,m+1] <- signum(j)*exp((-1)*a)
         }
      }
   # (-1)*R_a^(j)(a)
   for (j in (2*m+1):dim) for (k in (2*m+1):dim) Gamma[j,k] <- (-1)*Gamma[j,k-2*m]
   InvGamma <- solve(Gamma)
   last <- InvGamma[,dim]
   gmm <- last*signum(m-1)
   gmm
}
```

A.1.2 Representor Matrix

To produce a representor matrix is very simple from (1.4.40), (1.4.41) and Definition 2.2.1 using function `repre()`.

```
represmat <- function(m=4,xdata=c(0.1,0.5,0.2,0.44,0.9,0.75)) {
# computes representor matrix evaluated at the data points xdata
   n <- length(xdata)
   # creates an 4m x n matrix that has coefficients (gamma_k)s of representors in columns
   GMM <- matrix(nrow=4*m,ncol=n)
   for (i in 1:n) GMM[,i] <- t(repre(m,xdata[i]))
   # representor matrix Psi
   Psi <- matrix(nrow=n,ncol=n)
   for (i in 1:n) for (j in 1:n) {
   tmp <- 0
```

```
      # computes representor at x_i evaluated at x_j,
      # i.e. \psi_{x_i}(x_j)=\psi_{xdata[i]}(xdata[j])
      if (m%%2 == 0) {
         if (xdata[i]>xdata[j]) {
            # L_{xdata[i]}(xdata[j]) and m is even
            for (k in 0:(m/2-1)) tmp
            <- tmp+(GMM[k+1,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in (m/2+1):m) tmp
            <- tmp+(GMM[k,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in 0:(m/2-1)) tmp
            <- tmp+(GMM[m+k+1,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            for (k in (m/2+1):m) tmp
            <- tmp+(GMM[m+k,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            } else {
            # R_{xdata[i]}(xdata[j]) and m is even
            for (k in 0:(m/2-1)) tmp
            <- tmp+(GMM[2*m+k+1,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in (m/2+1):m) tmp
            <- tmp+(GMM[2*m+k,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in 0:(m/2-1)) tmp
            <- tmp+(GMM[3*m+k+1,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            for (k in (m/2+1):m) tmp
            <- tmp+(GMM[3*m+k,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            }
         } else {
         if (xdata[i]>xdata[j]) {
            # L_{xdata[i]}(xdata[j]) and m is odd
            tmp <- tmp+(GMM[1,i]*exp(xdata[j]))
            for (k in 1:((m+1)/2-1)) tmp
            <- tmp+(GMM[k+1,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in ((m+1)/2+1):m) tmp
            <- tmp+(GMM[k,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            tmp <- tmp+(GMM[m+1,i]*exp((-1)*xdata[j]))
            for (k in 1:((m+1)/2-1)) tmp
            <- tmp+(GMM[m+k+1,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            for (k in ((m+1)/2+1):m) tmp
            <- tmp+(GMM[m+k,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            } else {
            # R_{xdata[i]}(xdata[j]) and m is odd
            tmp <- tmp+(GMM[2*m+1,i]*exp(xdata[j]))
            for (k in 1:((m+1)/2-1)) tmp
            <- tmp+(GMM[2*m+k+1,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            for (k in ((m+1)/2+1):m) tmp
            <- tmp+(GMM[2*m+k,i]*exp(xdata[j]*relambda(k,m))*cos(xdata[j]*imlambda(k,m)))
            tmp <- tmp+(GMM[3*m+1,i]*exp((-1)*xdata[j]))
            for (k in 1:((m+1)/2-1)) tmp
            <- tmp+(GMM[3*m+k+1,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            for (k in ((m+1)/2+1):m) tmp
            <- tmp+(GMM[3*m+k,i]*exp(xdata[j]*relambda(k,m))*sin(xdata[j]*imlambda(k,m)))
            }
         }
      Psi[i,j] <- tmp
      }
   Psi
}
```

A.1.3 Plotting a Regression Curve in Sobolev Space

We need to plot a regression curve for many figures. To plot such a curve we just divide our predefined interval into equidistant sequence of points and evaluate the sum of representors multiplied by corresponding coefficients in these points. Finally we only connect evaluated points with the lines.

```
curve <- function(c,xdata,ydata,m) {
   n <-length(xdata)
   seq(0,1,by=0.05) -> xax # x-axis
   points <- length(xax) # number of points on x-axis
   yax <- rep(0,points) # prepared for y-axis data at this time only zeros
   for (k in (1:points)) {
      j <- length(xdata[xdata<xax[k]]) # how many xdata are less than xax[k]
      if (j == 0) for (i in (1:n)) yax[k] <- Lpart(m,xdata[i],xax[k])*c[i]+yax[k] else {
         if (j == n) for (i in (1:n)) yax[k] <- Rpart(m,xdata[i],xax[k])*c[i]+yax[k] else {
            for (i in ((j+1):n)) yax[k] <- Lpart(m,xdata[i],xax[k])*c[i]+yax[k]
            for (i in (1:j)) yax[k] <- Rpart(m,xdata[i],xax[k])*c[i]+yax[k]
            }
         }
      }
   lines(xax,yax)
}
```

A.2 Constrained Minimizing with *R*

A.2.1 Quadratic Minimizing in Finite Dimension

To estimate coefficients $\mathbb{c}$—necessary for the estimation of the regression function f according to Corollary 2.2.2 or Remark 2.2.4—we need to implement Theorem 2.2.5 and Algorithm 2.2.1. This procedure requires library `Matrix` for Schur decomposition of matrix.

```
estimator <- function(m=4,xdata,ydata,L=2,dif=0.05) {
# computes coefficients c of estimator \hat{f}
   n <- length(xdata)
   if (n != length(ydata)) write("ERROR in input data!",file="")
   # rearrange matched xdata vectors so that the first is in ascending order
   o <- order(xdata)
   rearranged <- rbind(xdata[o],ydata[o])
   xdata <- rearranged[1,]
   ydata <- rearranged[2,]
   PSI <- as(repremat(m,xdata), "dgeMatrix")
   schur.psi <- Schur(PSI)
   phi <- schur.psi$Z
   lambda <- schur.psi$WR
   z <- t(phi)%*%ydata
   # algorithm: ?is in a feasible set S or not?
   if (sum(z^2/lambda) <= L) d <- z/lambda else {
      # upper estimate of vartheta^*; maxlambda=min_i(lambda_i); maxz=max_i(z_i)
      maxlambda <- max(lambda)
      maxz <- max(abs(z))
      UP <- maxz*sqrt(maxlambda/(n*L))
      vartheta <- uniroot(function(x) sum(z^2*lambda/(lambda+n*x)^2)-L,low=0,up=UP)$root
```

```
      d <- (z/(lambda+n*vartheta))
      }
   c <- phi%*%d
   c
}
```

A.2.2 Ridge Regression

Above function `estimator()` (especially implementation of Algorithm 2.2.1) can be proceed similarly using preprogrammed functioned `optim()`. All we need to do is to take a different view of it—using ridge regression.

```
z <- t(phi)%*%ydata
objectivef <- function(unknown) {
   (1/n)*sum((lambda*unknown-z)^2)+chi*sum(lambda*(unknown^2))
}
gradient <- function(unknown) {
   (2*lambda/n)*(lambda*unknown-z)+2*chi*lambda*unknown
}
minimum <- optim((z/lambda),objectivef,gradient)
d <- minimum$par
```

A.3 Cross-Validation with *R*

A.3.1 Choosing the Smoothing Parameter

To choose the smoothing parameter χ, we plot a Cross-Validation function and then we find its minimum. To choose optimal χ is the same as choosing the Sobolev bound L according Theorem 2.6.1 and Remark 2.6.3.

```
CVchi <- function(m=2,xvalue,yvalue,chilower=-5.5,chiupper=-1.5,chidif=0.1) {
   n <- length(xvalue)
   if (n != length(yvalue)) write("ERROR in input data!",file="")
   # rearrange matched xvalue vectors so that the first is in ascending order
   o <- order(xvalue)
   rearranged <- rbind(xvalue[o],yvalue[o])
   xvalue <- rearranged[1,]
   yvalue <- rearranged[2,]
   chiveclog <- seq(chilower,chiupper,by=chidif)
   chivec <- 10^chiveclog
   ERvec <- numeric(length(chivec))
   for (s in (1:length(chivec))) {
      ER <- 0
      for (xr in xvalue) {
         xdata <- xvalue[xvalue!=xr]
         c <- estimator(m,xdata=xvalue[xvalue!=xr],ydata=yvalue[xvalue!=xr],chi=chivec[s])
         yr <- 0 # f_{-r} values at xvalue[-r] ... xr <-> yr
         j <- length(xdata[xdata<xr]) # how many xdata are less than xr
         if (j == 0) for (i in (1:(n-1))) yr <- Lpart(m,xdata[i],xr)*c[i]+yr else {
            if (j == (n-1)) for (i in (1:(n-1))) yr <- Rpart(m,xdata[i],xr)*c[i]+yr else {
               for (i in ((j+1):(n-1))) yr <- Lpart(m,xdata[i],xr)*c[i]+yr
               for (i in (1:j)) yr <- Rpart(m,xdata[i],xr)*c[i]+yr
               }
```

```
            }
         ER <- ER+(((yvalue[xvalue==xr])-yr)^2)
         }
      ERvec[s] <- ER
      }
   plot(chiveclog,ERvec,xlab=expression(log(chi(L))),ylab=expression(CV(L)))
   points(chiveclog,ERvec)
   lines(chiveclog,ERvec)
}
```

A.4 Isotonia with *R*

We would like to extend our setup for Constrained Minimizing with *R* in Section A.2. We want to add some isotonic properties.

A.4.1 Definite Non-decreasing

We implement model declared in Definition 3.2.1. This algorithm requires library `quadprog` for function `solve.QP()`. Procedure `derrepremat()` has been preprogrammed according to Definition 3.1.2 and similarly as shown procedure `repremat()`.

```
defmonotonic <- function(m=4,xdata,ydata,chi=0.0001) {
# chi ... smoothing parameter of ridge regression
   n <- length(xdata)
   PSI <- repremat(m,xdata)
   Dmat <- 2*(PSI%*%PSI/n+chi*PSI)
   dvec <- 2*(ydata%*%PSI)/n
   Amat <- derrepremat(m,xdata)
   bvec <- rep(0,n)
   minimum <- solve.QP(Dmat,dvec,Amat,bvec)
   c <- minimum$solution
   c
}
```

A.4.2 Indefinite Non-decreasing

This is similar case as above—implementation of the model declared in Definition 3.2.3.

```
Amat <- matrix(nrow=n,ncol=n-1)
for (i in 1:(n-1)) Amat[,i] <- PSI[,i+1]-PSI[,i]
bvec <- rep(0,n-1)
```

A.4.3 Definite Non-decreasing and Definite Convexity

Also similar as above—implementation of the combined model declared in Definition 3.2.1 and Definition 3.3.2.

```
Amat <- cbind(derrepremat(m,xdata),secderrepremat(m,xdata))
bvec <- rep(0,2*n)
```

A.4.4 Multiple Observations

Implementation of the model declared in Section 2.4.

```
multiple <- function(filedata="file.dat",m=4,chi=0.001) {
   data <- read.table(filedata,header=FALSE)
   xdata <- data[1,1]
   for (i in data[,1]) if (length(xdata[xdata==i])==0) xdata <- c(xdata,i)
   ydata <- data[,2]
   k <- length(xdata)
   n <- length(ydata)
   Delta <- matrix(nrow=n,ncol=k)
   for (i in 1:n) for (j in 1:k) if (xdata[j]==data[i,1]) {
      Delta[i,j] <- 1 else Delta[i,j] <- 0
      }
   PSI <- repremat(m,xdata)
   transDelta <- t(Delta)
   Dlt <- transDelta%*%Delta
   Delt <- PSI%*%Dlt
   Dmat <- 2*(Delt%*%PSI/n+chi*PSI)
   dvec <- 2*(PSI%*%(transDelta%*%ydata))/n
   Amat <- (-1)*derrepremat(m,xdata)
   bvec <- rep(0,k)
   minimum <- solve.QP(Dmat,dvec,Amat,bvec)
   c <- minimum$solution
   c
}
```

A.5 Confidence Intervals and Bootstrap with *R*

A.5.1 Calloption Prices

Let's construct an isotonic regression curve in m-order Sobolev space, which is non-increasing and convex. We can use function `pcls()` from library `mgcv` instead of function `solve.QP()` from library `quadprog`, because it is more complex and faster than the second one.

```
PSI <- repremat(m,xdata)
derPSI <- derrepremat(m,xdata)
M <- list(X=PSI,y=ydata,S=list(PSI),w=rep(1,n),Ain=SIG*t(derPSI),bin=rep(0,n),
C=matrix(0,0,0),off=0,sp=chi,p=solve(t(derPSI))%*%rep(epsilon,n))
c <- pcls(M)
```

We can also construct point-wise confidence intervals based upon classical bootstrap, wild bootstrap and CLT. See Theorem 4.1.2 and Algorithm 4.3.1.

```
# Classical Bootstrap Confidence Intervals
epsdgm <- ydata - PSI%*%c
epsdgm <- epsdgm - mean(epsdgm)
fdgm <- PSI%*%c
BootEstimates <- matrix(0,nrow=nboot,ncol=n)
for (iboot in 1:nboot) {
```

```
   epsboot <- sample(epsdgm,size=n,replace=T)
   yboot <- fdgm + epsboot
   MBoot <- list(X=PSI,y=yboot,S=list(PSI),w=rep(1,n),Ain=SIG*t(derPSI),bin=rep(0,n),
   C=matrix(0,0,0),off=0,sp=chi,p=solve(t(derPSI))%*%rep(epsilon,n))
   cBoot <- pcls(MBoot)
   BootEstimates[iboot,] <- PSI%*%cBoot
   }
BootCUp <- rep(0,n)
BootCLo <- rep(0,n)
for (i in 1:n) {
   BootCUp[i] <- quantile(BootEstimates[,i],prob=.975)
   BootCLo[i] <- quantile(BootEstimates[,i],prob=.025)
   }
# Bootstrap Wild Confidence Intervals
BootEstimates <- matrix(0,nrow=nboot,ncol=n)
for (iboot in 1:nboot) {
   randomBernoulli <- rbinom(n,1,((5+sqrt(5))/10))
   ranplusminus <- (1-sqrt(5))/2*randomBernoulli + (1+sqrt(5))/2*abs(randomBernoulli-1)
   epsboot <- epsdgm * ranplusminus
   yboot <- fdgm + epsboot
   MBoot <- list(X=PSI,y=yboot,S=list(PSI),w=rep(1,n),Ain=SIG*t(derPSI),bin=rep(0,n),
   C=matrix(0,0,0),off=0,sp=chi,p=solve(t(derPSI))%*%rep(epsilon,n))
   cBoot <- pcls(MBoot)
   BootEstimates[iboot,] <- PSI%*%cBoot
   }
BootCUp <- rep(0,n)
BootCLo <- rep(0,n)
for (i in 1:n) {
   BootCUp[i] <- quantile(BootEstimates[,i],prob=.975)
   BootCLo[i] <- quantile(BootEstimates[,i],prob=.025)
   }
# Asymptotic Confidence Intervals
sserr <- ((ydata-mean(ydata))^2)/(n-1)
CI <- 1.96*sqrt(sserr/n)
AsyCIup <- fdgm+CI
AsyCIlo <- fdgm-CI
```

A.6 Tests of Hypothesis with *R*

A.6.1 Test of Isotonia

Finally, let's proceed monotonicity and convexity residual regression test based on Theorem 4.2.1. This implementation follows from Algorithm 4.2.1.

```
Dmat <- 2*(PSI%*%PSI/n+chi*PSI)
dvec <- 2*(ydata%*%PSI)/n
Amat <- cbind(derrepremat(m,xdata),secderrepremat(m,xdata))
bvec <- rep(0,2*n)
minimum <- solve.QP(Dmat,dvec,Amat,bvec)
c <- minimum$solution
lambda <- n^(-0.2)*(max(xdata)-min(xdata))
K <- matrix(0,nrow=n,ncol=n)
for (i in (1:n)) {
   tmp <- abs(xdata-xdata[i])
   tmp <- (tmp<=lambda)
   tmp[i] <- 0 # exclude current observation from kernel matrix
   K[i,] <- 0.5*tmp
```

```
   }
eps <- ydata - PSI%*%c
U <- t(eps)%*%K%*%eps/(n*(n-1)*lambda)
sigma <- 2*t(eps^2)%*%(K^2)%*%(eps^2)/(n^2*(n-1)^2*lambda^2)
stat <- U/sqrt(sigma)
```

And its bootstrap version uses function `sample()`. It is based on similar Theorem 4.3.1 and follows from Algorithm 4.3.2.

```
fdata <- PSI%*%c
nboot <- 100
UBoot <- rep(0,nboot)
for (iboot in 1:nboot) {
   epsboot <- sample(eps,size=n,replace=T)
   yboot <- fdata + epsboot
   dvecboot <- 2*(t(yboot)%*%PSI)/n
   minimumboot <- solve.QP(Dmat,dvecboot,Amat,bvec)
   cboot <- minimumboot$solution
   epsboothat <- yboot - PSI%*%cboot
   U <- t(epsboothat)%*%K%*%epsboothat/(n*(n-1)*lambda)
   Omegahat <- 2*t(epsboothat^2)%*%K^2%*%epsboothat^2/(n^2*(n-1)^2*lambda^2)
   U <- U/sqrt(Omegahat)
   UBoot[iboot] <- U
   if (iboot%%10==0) cat("UBoot[iboot]=",UBoot[iboot],"\n")
   }
BootCrit <- quantile(UBoot,prob=c(.90,.95))
c(BootCrit,mean(UBoot),var(UBoot))
```

The practical performance of the residual regression test, its classical and wild bootstrap versions have been shown in Example 4.3.1.

A.7 Software

I have been using *GNU/GPL* statistical software *R*, versions 2.0.0–2.2.1. All the procedures, simulations and tests have been running on Unix-like operating systems—*FreeBSD 6.0* and *Ubuntu Linux 6.04*. In this way I would like to thank to all developers, testers and contributors of these *Open Source* operating systems and to the *R* Development Core Team.

Isotonic regression in Sobolev spaces has not been implemented into any statistical software yet. That's why I am preparing a package/library for *R*. At this moment, the source code of all used functions is placed on my homepage `http://michal.pesta.matfyz.cz/studium/` and can be downloaded from this site.

Appendix B

Figures

In this chapter we demonstrate up to now theory and algorithms. The regression estimation in the Sobolev space $\mathcal{H}^4$ with specific isotonic constraint is illustrated in the following figures.

We use simulated data sets as well as real data. The estimation of true function, its first and second derivative is proceeded. Isotonia of the first and the second order (monotonicity with convexity or concavity) is demanded in each case. We construct also point-wise confidence intervals—mostly on 95% probability level—for each estimate based upon asymptotic normality and bootstrap techniques—classical or wild. Number of bootstrap resample steps is always 1000, default for each estimation.

On each page, we can see a pair of figures with the same data, the estimator and point-wise confidence intervals. Lines are the only difference in the second figures that connect particular neighboring upper or lower ends of point-wise confidence intervals. This is done for better visualization, sometimes the first figure provides better information, sometimes the second one from the couple.

Asymptotic and bootstrap confidence intervals for the regression estimator of true function appear almost the same. But the noticeable sporadically striking difference occurs when we estimate derivatives. The bootstrap techniques provide much better confidence intervals.

B.1 Simulated Data

We simulate following data and estimate the true function and its derivatives:

$$\left\{x_i, \exp\left\{-x_i\right\} + \varepsilon_i\right\}_{i=1}^{14}, \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, .25), x_i \in [-2, 1] \text{ equidistant}, \tag{B.1.1}$$

$$\left\{x_i, \arctan(x_i) + \varepsilon_i\right\}_{i=1}^{8}, \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, .1), x_i \in [0, 2] \text{ equidistant}. \tag{B.1.2}$$

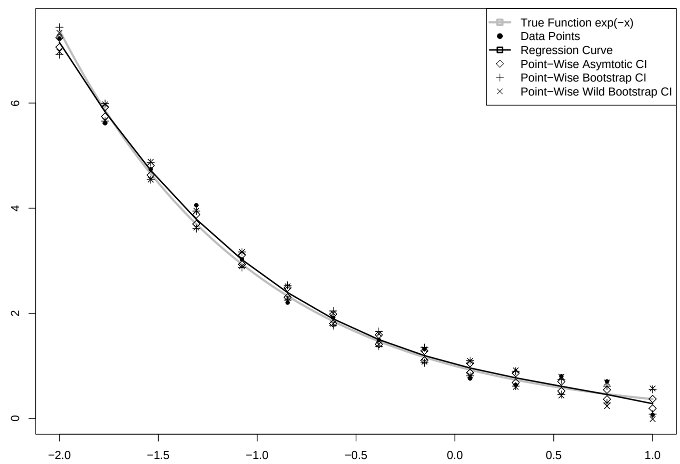

Figure B.1: Monotone and convex regression estimator (with various types of 95% point-wise confidence intervals) in $\mathcal{H}^4$ of simulated data $Y_i = \exp\{-x_i\} + \varepsilon_i,\ i = 1, \dots, 14$, where x_i are equidistributed on $[-2, 1]$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, .25)$.

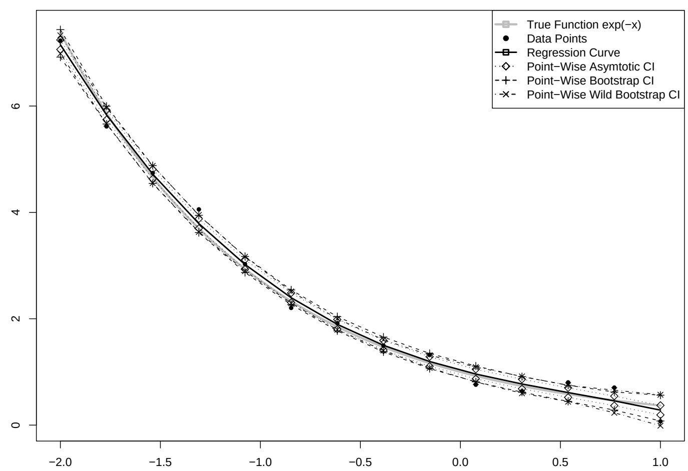

Figure B.2: Monotone and convex regression estimator with various types of 95% point-wise confidence intervals (connected by lines) in $\mathcal{H}^4$ of simulated data from Figure B.1.

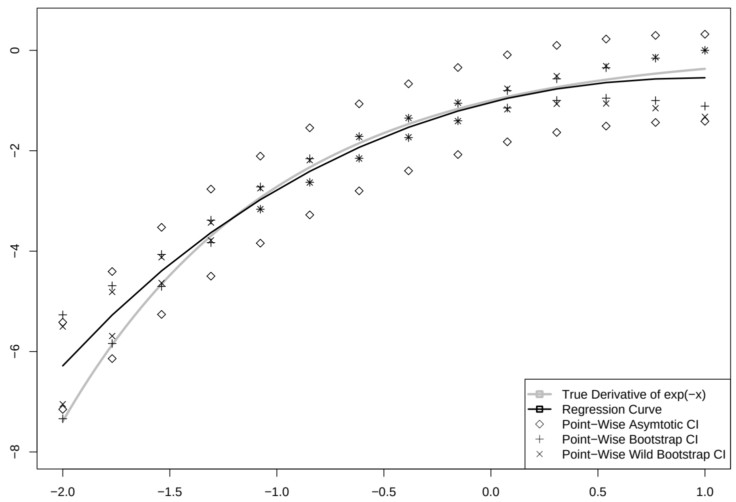

Figure B.3: Regression estimator of the first derivative of $\exp\{-x\}$ represented by simulated data from Figure B.1 (with various types of 95% point-wise confidence intervals). $[\exp\{-x\}]' = -\exp\{-x\}$.

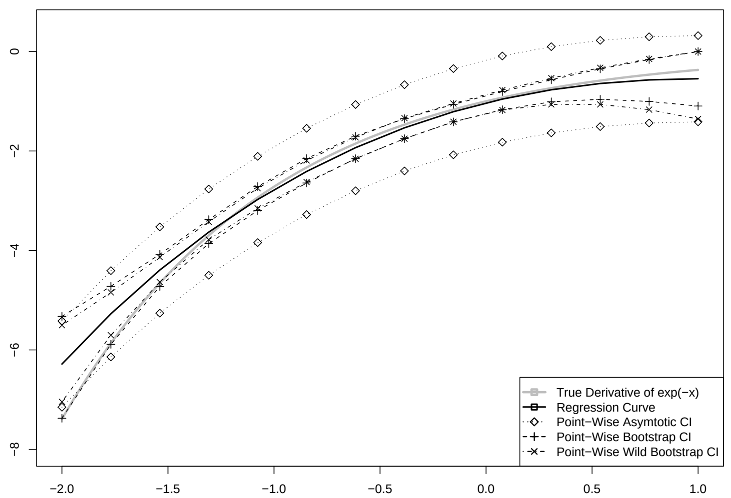

Figure B.4: Same as Figure B.3. 95% point-wise confidence intervals are connected by lines for better visualization.

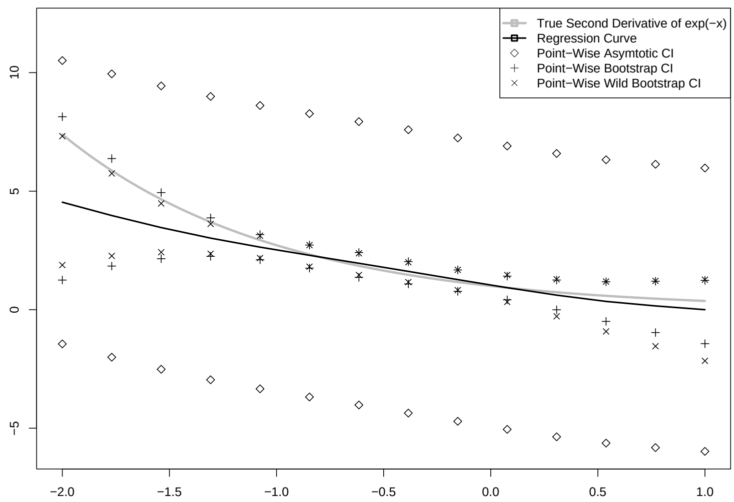

Figure B.5: Regression estimator of the second derivative of exp $\{-x\}$ represented by simulated data from Figure B.1 (with various types of 95% point-wise confidence intervals). $[\exp\{-x\}]'' = \exp\{-x\}$.

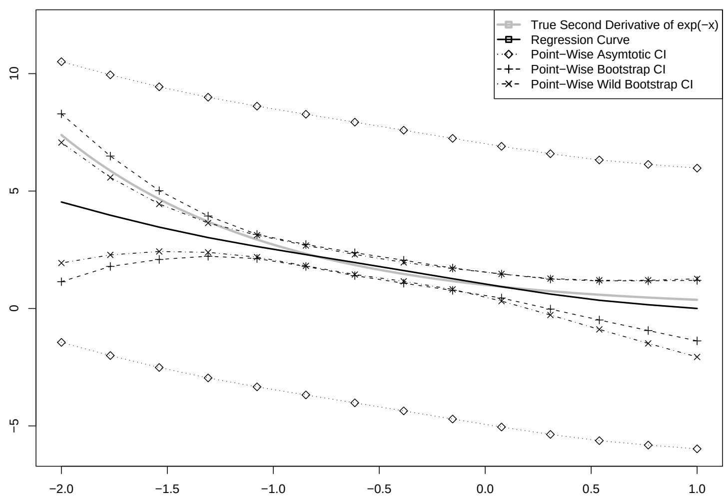

Figure B.6: Same as Figure B.5. 95% point-wise confidence intervals are connected by lines for better visualization.

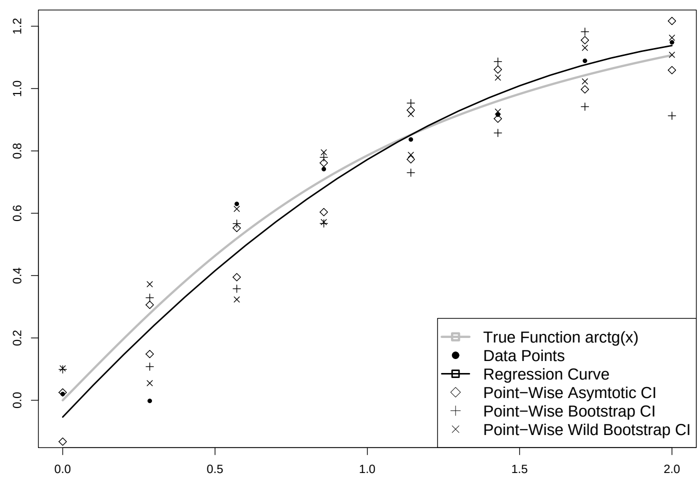

Figure B.7: Monotone and concave regression estimator (with various types of 95% point-wise confidence intervals) in $\mathcal{H}^4$ of simulated data $Y_i = \arctan(x_i) + \varepsilon_i$, $i = 1, \ldots, 8$, where x_i are equidistributed on $[0, 2]$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, .1)$.

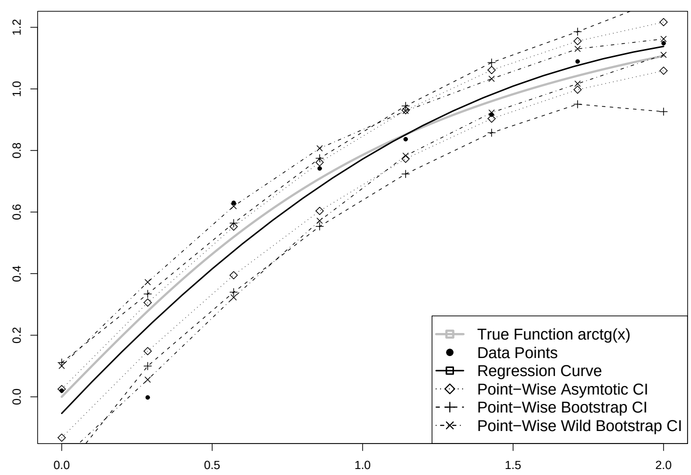

Figure B.8: Monotone and concave regression estimator with various types of 95% point-wise confidence intervals (connected by lines) in $\mathcal{H}^4$ of simulated data from Figure B.7.

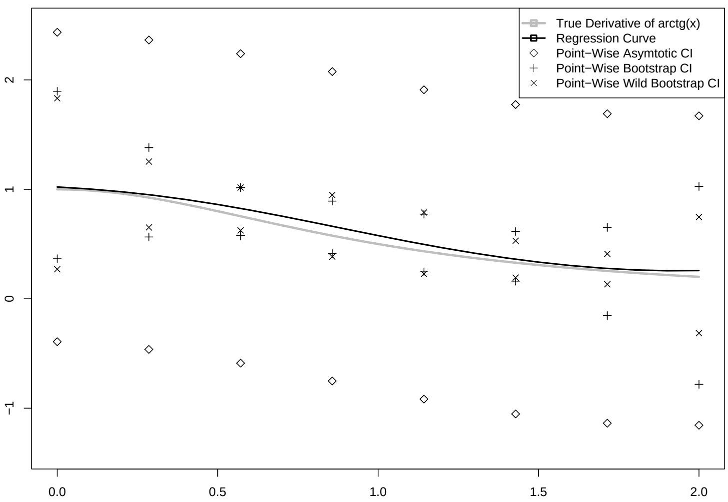

Figure B.9: Regression estimator of the first derivative of arctan(x) represented by simulated data from Figure B.1 (with various types of 95% pointwise confidence intervals). $[\arctan(x)]' = \frac{1}{1+x^2}$.

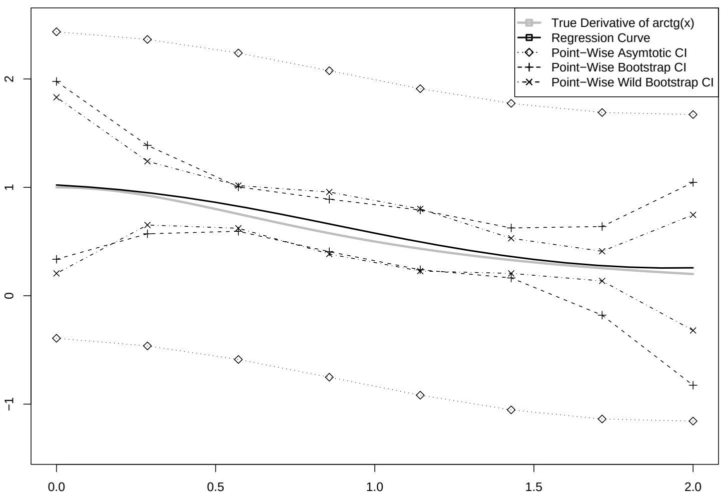

Figure B.10: Same as Figure B.9. 95% point-wise confidence intervals are connected by lines for better visualization.

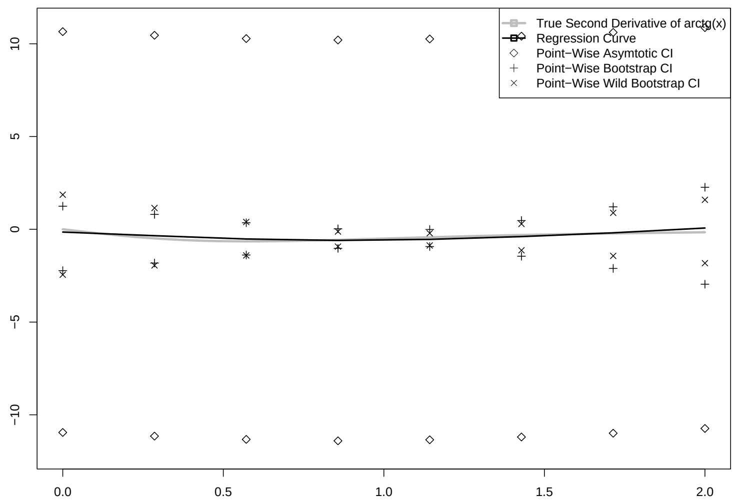

Figure B.11: Regression estimator of the second derivative of arctan(x) represented by simulated data from Figure B.1 (with various types of 95% point-wise confidence intervals). $[\arctan(x)]'' = -\frac{2x}{(1+x^2)^2}$.

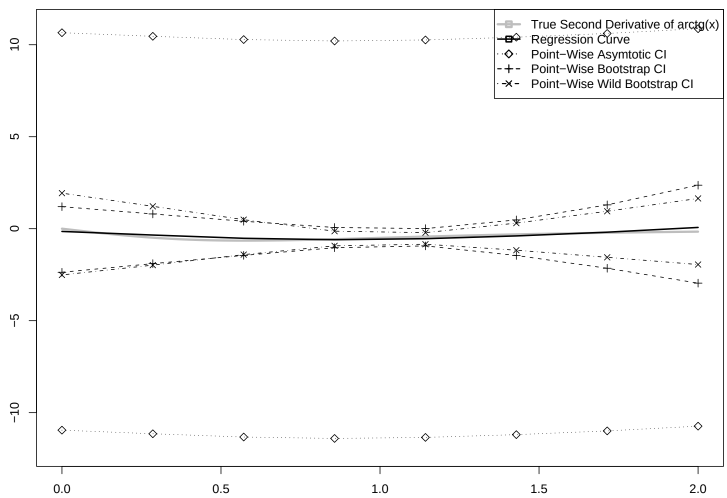

Figure B.12: Same as Figure B.11. 95% point-wise confidence intervals are connected by lines for better visualization.

B.2 Real Data – DAX Call Options

The German Stock Index (Deutscher Aktienindex—DAX) is a total return index of 30 selected German blue chip stocks traded on the Frankfurt Stock Exchange (Deutsche Börse AG). We have a data set of DAX call options, where independent variable is strike price and dependent variable is call option price adjusted by discount factor. From econometric theory (see Härdle and Hlávka (2005)) follows that option price—as a function of strike price—have to be decreasing and convex.

Our data set contains 561 prices of call options on DAX on January 1^{st}, 2001. We construct regression estimator with non-increasing and convex property for option price. Confidence intervals based on asymptotic normality and bootstrap are also constructed. Then we estimate first and second derivative of option price, but we do not show confidence intervals based on asymptotic normality in figures, because they are very wide and unreliable. It has two reasons—theoretical and computational. Theoretical reason is that variance in (4.1.9) and (4.1.10) is obviously huge, computational one is that matrix of representors $\mathbf{\Psi}$ can be ill-conditioned and hence we can obtain an inaccurate inverse matrix.

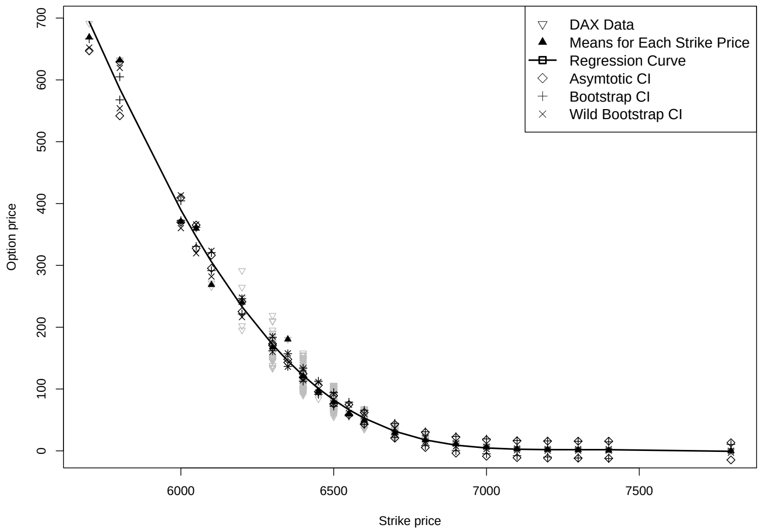

Figure B.13: DAX call options data—monotone (nonincreasing) and convex regression estimator in the Sobolev space $\mathcal{H}^4$ with various types of 95% point-wise confidence intervals.

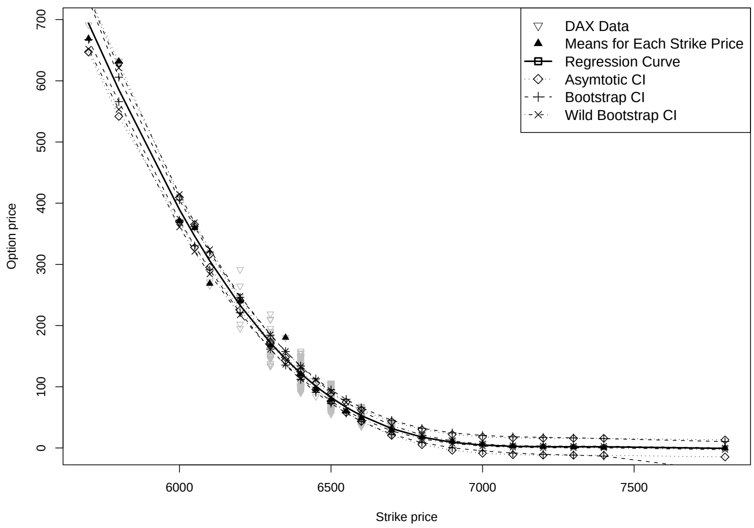

Figure B.14: Same as Figure B.13. 95% point-wise confidence intervals are connected by lines for better visualization.

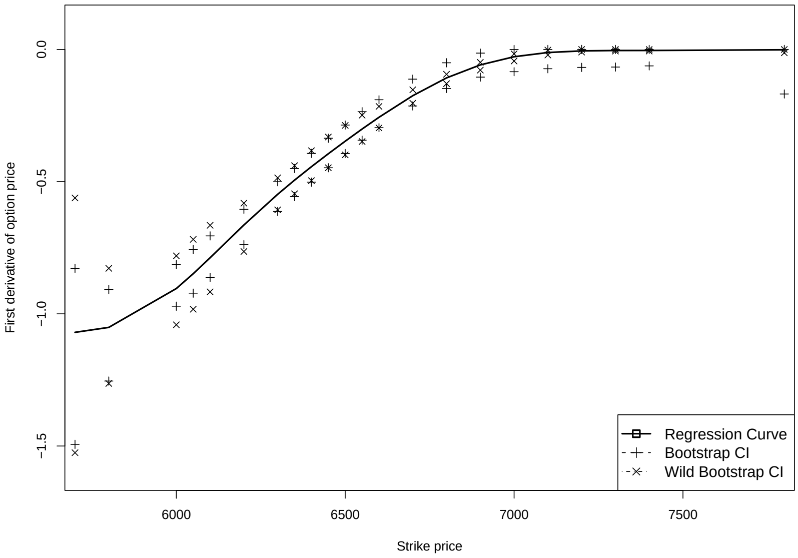

Figure B.15: Regression estimator of the first derivative of option price with various 95% point-wise confidence intervals.

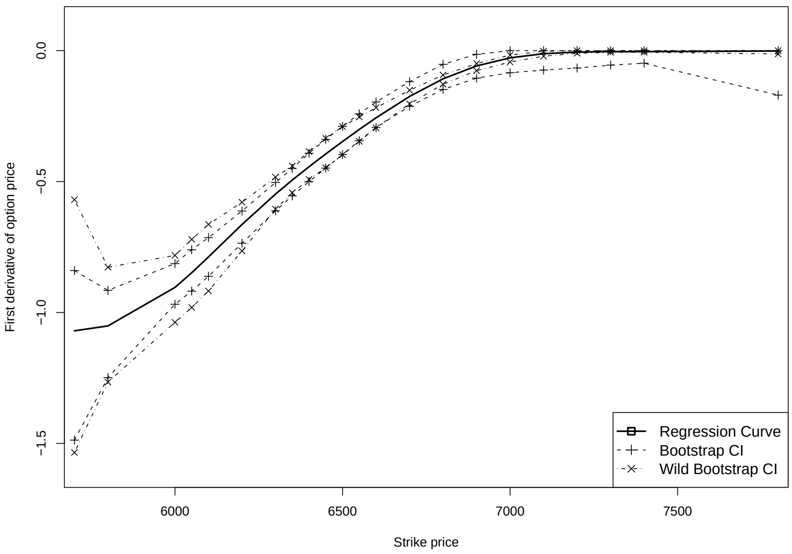

Figure B.16: Regression estimator of the first derivative of option price (95% point-wise confidence intervals are connected by lines for better visualization).

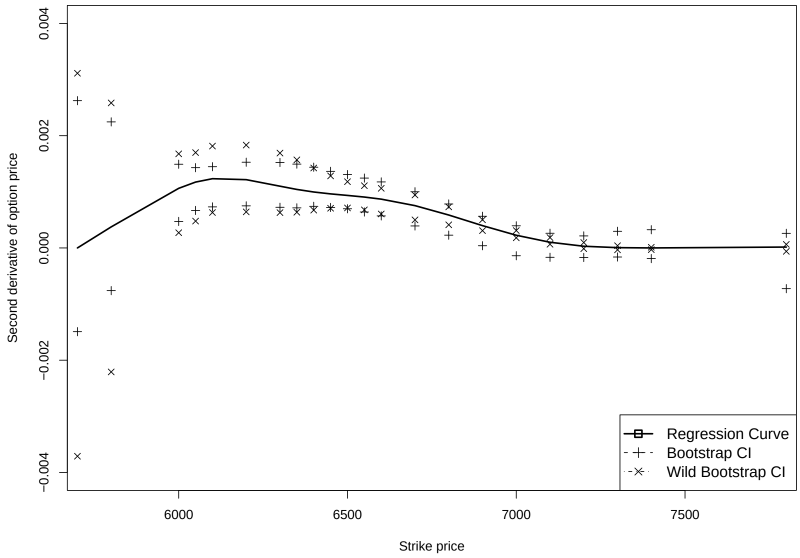

Figure B.17: Regression estimator of the second derivative of option price with various 95% point-wise confidence intervals.

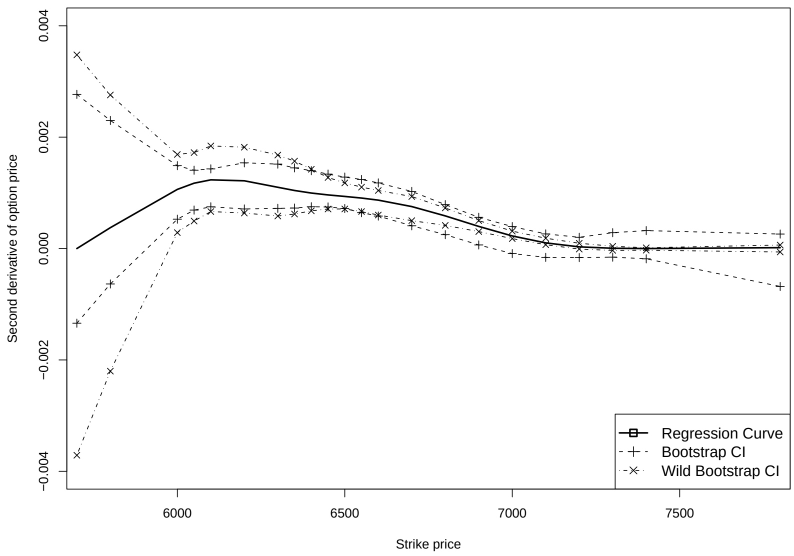

Figure B.18: Regression estimator of the second derivative of option price (95% point-wise confidence intervals are connected by lines for better visualization).

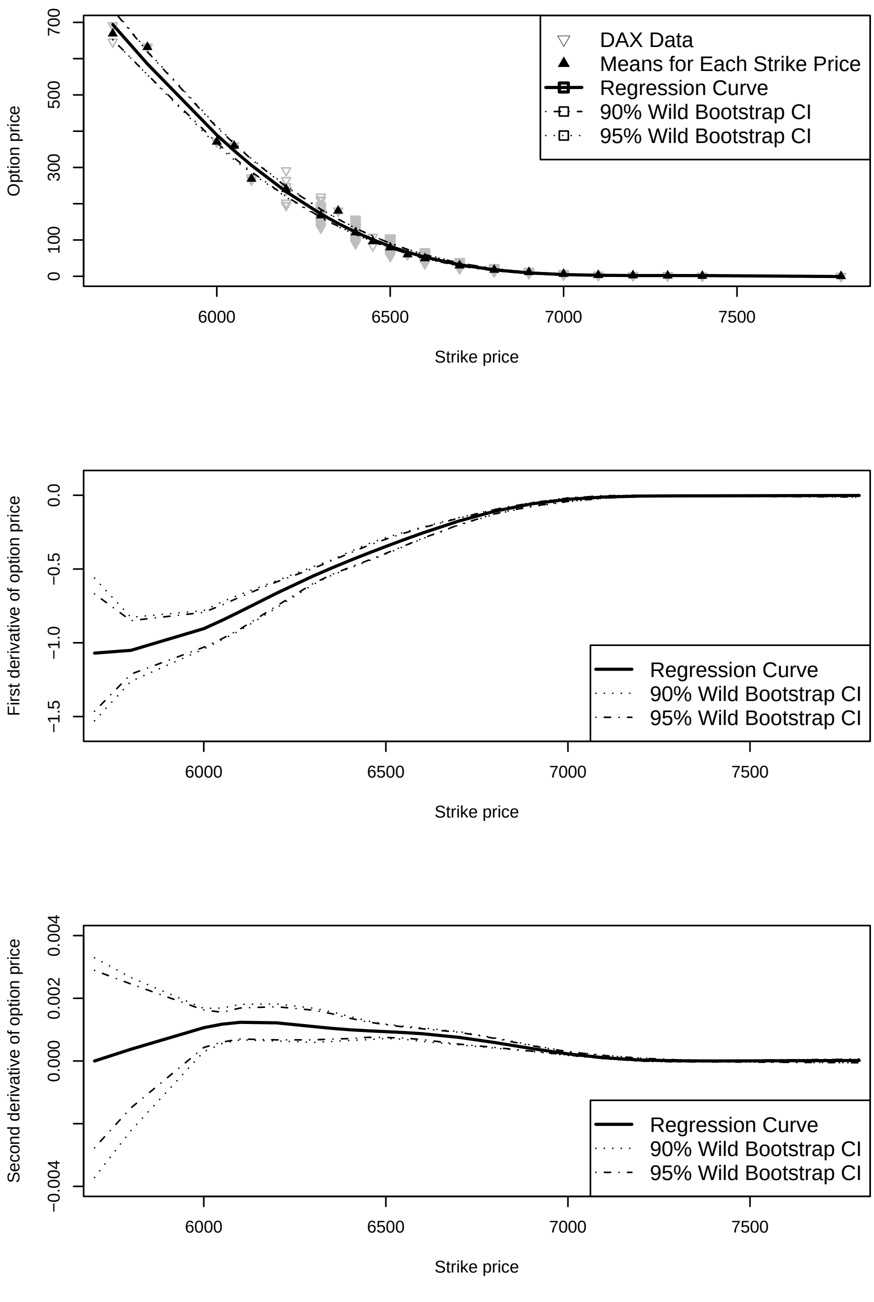

Figure B.19: Regression estimator of option price, the first and the second derivative of option price with 90% and 95% point-wise wild bootstrap confidence intervals.

Appendix C

Useful Theorems and Lemmas

Used theorems and lemmas are listed below. They are divided into the separate categories and then named in the alphabetical order.

C.1 Functional Analysis

Theorem C.1.1 (Arzela-Ascoli Theorem). *Let Ω be a bounded subset of $\mathbb{R}^n$ and $\{f_k\}_{k\in\mathbb{N}}$ a sequence of functions $f_k : \Omega \to \mathbb{R}^m$. If $\{f_k\}$ is equibounded and uniformly equicontinuous then there exists a uniformly convergent subsequence $\left\{f_{k_j}\right\}_{j\in\mathbb{N}}$.*

Theorem C.1.2 (Cauchy-Schwartz Inequality). *If $f \in L_2(\Omega)$ and $f \in L_2(\Omega)$, then $fg \in L_1(\Omega)$ and*

$$\int_\Omega |f(\mathbb{x})g(\mathbb{x})|\, d\mathbb{x} \leq \|f\|_{L_2(\Omega)} \|g\|_{L_2(\Omega)}. \tag{C.1.1}$$

Theorem C.1.3 (Riesz Representation Theorem). *For every continuous linear functional f on a Hilbert space $\mathcal{H}$, there is a unique $u \in \mathcal{H}$ such that $f(x) = \langle x, u\rangle$ for all $x \in \mathcal{H}$.*

C.2 Linear Algebra and Matrices

Theorem C.2.1 (Eigen Decomposition). *Let $\mathbf{P}_{n\times n}$ be a matrix of eigenvectors of a given square matrix $\mathbf{A}_{n\times n}$ and $\mathbf{W}_{n\times n}$ be a diagonal matrix with the corresponding eigenvalues on the diagonal. Then, as long as $\mathbf{P}_{n\times n}$ is a square matrix with full rank, $\mathbf{A}_{n\times n}$ can be written as an eigen decomposition*

$$\mathbf{A}_{n\times n} = \mathbf{P}_{n\times n}\mathbf{W}_{n\times n}\mathbf{P}^{-1}_{n\times n} \tag{C.2.2}$$

where $\mathbf{W}_{n\times n}$ *is a diagonal matrix. Furthermore, if* $\mathbf{A}_{n\times n}$ *is symmetric, then the columns of* $\mathbf{P}_{n\times n}$ *are orthogonal vectors. If* $\mathbf{P}_{n\times n}$ *is not a square matrix with full rank, then* $\mathbf{P}_{n\times n}$ *cannot have a matrix inverse and* $\mathbf{A}_{n\times n}$ *does not have an eigen decomposition.*

Theorem C.2.2 (Schur Decomposition). *Eigenvalues* $\lambda_1, \ldots, \lambda_n$ *of symmetric matrix* $\mathbf{A}_{n\times n}$ *are always real. Without losing of generality suppose that* $\lambda_1 \geq \ldots \geq \lambda_n$*. Let* $\mathbf{W}_{n\times n} = diag\{\lambda_1, \ldots, \lambda_n\}$*. Then there exists an orthogonal matrix* $\mathbf{U}_{n\times n}$ *such that*

$$\begin{aligned} \mathbf{A}_{n\times n} &= \mathbf{U}_{n\times n}\mathbf{W}_{n\times n}\mathbf{U}'_{n\times n}, && \text{(C.2.3)} \\ \mathbf{I}_{n\times n} &= \mathbf{U}'_{n\times n}\mathbf{U}_{n\times n} = \mathbf{U}_{n\times n}\mathbf{U}'_{n\times n}. && \text{(C.2.4)} \end{aligned}$$

Theorem C.2.3 (Singular Value Decomposition – SVD). *Let* $m \geq n$*. Any matrix* $\mathbf{A}_{m\times n}$ *can be written as the product of a column-orthogonal matrix* $\mathbf{U}_{m\times n}$*, a diagonal matrix with positive or zero elements* $\mathbf{W}_{n\times n}$*, and the transpose of an orthogonal matrix* $\mathbf{V}_{n\times n}$*:*

$$\mathbf{A}_{m\times n} = \mathbf{U}_{m\times n}\mathbf{W}_{n\times n}\mathbf{V}'_{n\times n} \tag{C.2.5}$$

where

$$\mathbf{W}_{n\times n} = \begin{pmatrix} w_1 & 0 & \ldots & 0 & 0 \\ 0 & w_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & w_{n-1} & 0 \\ 0 & 0 & \ldots & 0 & w_n \end{pmatrix} \tag{C.2.6}$$

and

$$\begin{aligned} w_1, \ldots, w_n &\geq 0, && \text{(C.2.7)} \\ \mathbf{U}'_{n\times m}\mathbf{U}_{m\times n} = \mathbf{V}'_{n\times n}\mathbf{V}_{n\times n} = \mathbf{V}_{n\times n}\mathbf{V}'_{n\times n} &= \mathbf{I}_{n\times n}. && \text{(C.2.8)} \end{aligned}$$

The diagonal elements of matrix $\mathbf{W}_{n\times n}$ *are the singular values of matrix* $\mathbf{A}_{m\times n}$ *and non-negative numbers.*

C.3 Information Theory

Theorem C.3.1 (Kolmogorov-Tihomirov). *Let* $\mathscr{F}$ *be a compact non-empty subset of a metric space. Then for all* $\delta > 0$ *exists* $A > 0$ *and* $0 < \zeta < 1$ *such that metric entropy*

$$H(\delta; \mathscr{F}) < A\delta^{-2\zeta}. \tag{C.3.9}$$

C.4 Probability

Theorem C.4.1 (McLeish). *Let* $\{X_{n,i}\}_{i=1}^{k_n}$ *be an array of random variables on the probability triple* $(\Omega, \mathscr{A}, \mathscr{P})$. *Let* $\{\mathscr{A}_{n,i}\}_{i=1}^{k_n}$ *be any triangular array of sub-sigma fields of* $\mathscr{A}$ *such that for each* n *and* $1 \leq i \leq k_n$, $X_{n,i}$ *is* $\mathscr{A}_{n,i}$*-measurable and* $\mathscr{A}_{n,i-1} \subset \mathscr{A}_{n,i}$. *Let* $X_{n,i}$ *be any array satisfying*

$$\text{i)} \quad \max_{1 \leq i \leq k_n} |X_{n,i}| \text{ is uniformly bounded in } L_2 \text{ norm}, \tag{C.4.10}$$

$$\text{ii)} \quad \max_{1 \leq i \leq k_n} |X_{n,i}| \xrightarrow{n\to\infty}_{\mathcal{P}} 0, \tag{C.4.11}$$

$$\text{iii)} \quad \sum_{i=1}^{k_n} X_{n,i}^2 \xrightarrow{n\to\infty}_{\mathcal{P}} 1, \tag{C.4.12}$$

$$\text{iv)} \quad \sum_{i=1}^{k_n} \mathbb{E}\left[X_{n,i} | \mathscr{A}_{n,i-1}\right] \xrightarrow{n\to\infty}_{\mathcal{P}} 0, \tag{C.4.13}$$

$$\text{v)} \quad \sum_{i=1}^{k_n} \mathbb{E}^2\left[X_{n,i} | \mathscr{A}_{n,i-1}\right] \xrightarrow{n\to\infty}_{\mathcal{P}} 0. \tag{C.4.14}$$

Then

$$\sum_{i=1}^{k_n} X_{n,i} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty. \tag{C.4.15}$$

List of Algorithms

List of Figures

List of Tables

Index

Bibliography

Adams, R. A. 1975. Sobolev Spaces. Academic Press, New York.

Bos, L. and Yatchew, A. J. 1997. Nonparametric least squares estimation and testing of economic models. *Journal of Quantitative Economics* 13:81–131.

de Geer, S. V. 1990. Estimating a regression function. *The Annals of Statistics* 18:907–924.

Fan, Y. and Li, Q. 1996. Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64:865–890.

Golub, G. H. and Loan, C. F. V. 1996. Matrix Computation. The Johns Hopkins University Press, Baltimore.

Härdle, W. 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Härdle, W. and Hlávka, Z. 2005. Dynamics of state price densities. SFB 649 "Economic Risk" Discussion Paper 2005–021, http://sfb649.wiwi.hu-berlin.de.

Härdle, W. and Yatchew, A. J. 2003. Nonparametric state price density estimation using constrained least squares and the bootstrap. This is a full unpublished entry, Humboldt-Universität zu Berlin and Toronto University, http://www.chass.utoronto.ca/˜yatchew/.

Kolmogorov, A. N. and Tihomirov, V. M. 1959. ε-entropy and ε-capacity of sets in function spaces. *Uspehi Matematičeskih Nauk* 14:3–86.

Lukeš, J. and Malý, J. 1995. Measure and Integral. MatfyzPress, Prague.

Maz'ja, V. G. 1985. Sobolev Spaces. Springer–Verlag, Berlin.

McLeish, D. L. 1974. Dependent central limit theorems and invariance principles. *The Annals of Probability* 2:620–628.

Robertson, T., Wright, F. T., and Dykstra, R. L. 1988. Order Restricted Statistical Inference. John Wiley & Sons, New York.

Wahba, G. 1992. Spline Models for Observation Data. SIAM, Philadelphia.

Yatchew, A. J. 1998. Nonparametric regression techniques in economics. *Journal of Economic Literature* 36:669–721.

Yatchew, A. J. 2000. Differencing methods in nonparametric regression. This is a full unpublished entry, Toronto University, http://www.chass.utoronto.ca/˜yatchew/.

Yatchew, A. J. 2003. Semiparametric Regression for the Applied Econometrician. Cambridge University Press, Cambridge.

Zheng, J. X. 1996. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75:263–289.