SFB 649 Discussion Paper 2006-069

Constrained General Regression in Pseudo-Sobolev Spaces with Application to Option Pricing

Zdeněk Hlávka*
Michal Pešta*

* Department of Probability and Mathematical Statistics,
Faculty of Mathematics and Physics,
Charles University in Prague,
Czech Republic

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Constrained General Regression in Pseudo-Sobolev Spaces with Application to Option Pricing*

Zdeněk Hlávka

Michal Pešta

Charles University in Prague

Faculty of Mathematics and Physics

Department of Probability and Mathematical Statistics

Sokolovská 83, 18675 Prague 8, Czech Republic

hlavka@karlin.mff.cuni.cz, tel. +420/221913284

pesta@karlin.mff.cuni.cz

This version: September 20, 2006

*We gratefully acknowledge the support of Deutsche Forschungsgemeinschaft, FEDC Guest Researcher Program for Young Researchers at SFB 649 "Economic Risk", and by MSM 0021620839 "Modern Mathematical Methods and Their Applications".

Abstract

State price density (SPD) contains important information concerning market expectations. In existing literature, a constrained estimator of the SPD is found by nonlinear least squares in a suitable Sobolev space. We improve the behavior of this estimator by implementing a covariance structure taking into account the time of the trade and by considering simultaneously both the observed Put and Call option prices.

Keywords and Phrases: isotonic regression, Sobolev spaces, monotonicity, multiple observations, covariance structure, option price

JEL classification: C10, C13, C14, C20, C88, G13

Let $Y_t(K,T)$ denote the price of a European Call with strike price K on day t and with expiry date T. The payoff at time T is given by $(S_T - K)_+ = \max(S_T - K, 0)$, where S_T denotes the price of the underlying asset at time T. The price of such an option may be expressed as the expected value of the payoff

$$Y_t(K,T) = \exp\{-r(T-t)\} \int_0^{+\infty} (S_T - K)_+ f(S_T) dS_T, \tag{1}$$

discounted by the known risk-free interest rate r. The expectation in (1) is evaluated with respect to the so-called State Price Density (SPD) $f(.)$. The SPD contains important information on the expectations of the market and its estimation is a statistical task of great practical interest (Jackwerth, 1999).

Similarly, we can express the price $Z_t(K,T)$ of the European Put with payoff $(K - S_T)_+$ as:

$$Z_t(K,T) = \exp\{-r(T-t)\} \int_0^{+\infty} (K - S_T)_+ f(S_T) dS_T. \tag{2}$$

In the following, the symbol $\mathbb{Z}$ denotes the vector of all Put option prices corresponding to a fixed date of expiry T observed on a given day t. Similarly, $\mathbb{Y}$ denotes a vector containing all Call option prices. The corresponding vectors of the strike prices for the Call and Put options are denoted by $\mathbb{x}_\alpha$ and $\mathbb{x}_\beta$, respectively.

Calculating the second derivative of (1) and (2) with respect to the strike price K, we can express the SPD as the second derivative of the European Call and Put option prices (Breeden and Litzenberger, 1978):

$$f(K) = \exp\{r(T-t)\} \frac{\partial^2 Y_t(K,T)}{\partial K^2} = \exp\{r(T-t)\} \frac{\partial^2 Z_t(K,T)}{\partial K^2}. \tag{3}$$

Both parametric and nonparametric approaches to SPD estimation are described in Jackwerth (1999). Nonparametric estimates of the SPD based on (3) are considered, among others, in Aït-Sahalia and Lo (2000);

Aït-Sahalia et al. (2001); Yatchew and Härdle (2005); Härdle and Hlávka (2006).

In this paper, we will generalize the nonlinear least squares method suggested in Yatchew and Härdle (2005) by including the covariance of the observed option prices suggested in Härdle and Hlávka (2006). The estimation of the SPD will be further improved by considering simultaneously both Put and Call option prices.

The investigation will be based on constrained (isotonic and convex) regression in pseudo-Sobolev spaces (Yatchew and Bos, 1997; Yatchew and Härdle, 2005). In Sections 1 and 2, we will describe the mathematical foundation of the method. In Section 3, we will discuss problems arising in the real life application on the observed option prices. The covariance structure suggested in Härdle and Hlávka (2006) is explained in Section 4. Finally, the SPD estimated from option prices on DAX are calculated in Section 5. The proofs of all theorems are given in Appendix B.

1 Pseudo-Sobolev Spaces

In this section, we give a brief overview on basic results of the theory of the Pseudo-Sobolev spaces. We assemble and prove necessary preliminaries and theorems for statistical regression in these spaces. The crux of this section lies in Theorem 1.2 (Representors in Pseudo-Sobolev Space) from Yatchew and Bos (1997). We have continued in examining representors' properties and proved Theorem 1.3, which provides the way of construction of the representors and their exact form.

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

$$D^{\boldsymbol{\alpha}} f(\mathbb{x}) := \frac{\partial^{|\boldsymbol{\alpha}|_1} f(\mathbb{x})}{\partial x_1^{\alpha_1} \dots \partial x_q^{\alpha_q}} \tag{4}$$

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

Definition 1.1 (Sobolev Norm). Let $f \in \mathcal{C}^m(\Omega) \cap \mathsf{L_p}(\Omega)$ (see Definitions A.2 and A.3). We introduce a Sobolev norm $\|\cdot\|_{p,Sob,m}$:

$$\|f\|_{p,Sob,m} := \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_\Omega \left| D^{\boldsymbol{\alpha}} f(\mathbb{x}) \right|^p \mathrm{d}\mathbb{x} \right\}^{1/p} . \tag{5}$$

We can write $\|\cdot\|_{p,\infty,Sob,m}$ since the multiindex of modulus $|\boldsymbol{\alpha}|_\infty = \max_{i=1,\dots,q} \alpha_i$ is taken with respect to maximum-norm.

Definition 1.2 (Pseudo-Sobolev Space). A Pseudo-Sobolev space of rank m, $\mathcal{W}_p^m(\Omega)$, is the completion of intersection of space $\mathcal{C}^m(\Omega)$ and space $\mathsf{L_p}(\Omega)$ with respect to the Sobolev norm $\|\cdot\|_{p,Sob,m}$.

Remark 1.1. $\mathcal{C}^m(\Omega) \cap \mathsf{L_p}(\Omega)$ is dense in $\mathcal{W}_p^m(\Omega)$ according to $\|\cdot\|_{p,Sob,m}$.

Definition 1.3 (Sobolev Inner Product). Let $f, g \in \mathcal{W}_2^m(\Omega)$. The Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ is defined as:

$$\langle f, g \rangle_{Sob,m} := \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \int_\Omega D^{\boldsymbol{\alpha}} f(\mathbb{x}) D^{\boldsymbol{\alpha}} g(\mathbb{x}) \mathrm{d}\mathbb{x}. \tag{6}$$

We denote the Sobolev norm $\|\cdot\|_{2,Sob,m} := \|\cdot\|_{Sob,m}$ for simplicity. The correctness of Definition 1.3 is guaranteed by the denseness of the space $\mathcal{C}^m(\Omega) \cap L_2(\Omega)$ in $\mathcal{W}_2^m(\Omega)$ (see Remark 1.1). The Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ induces in space $\mathcal{W}_2^m(\Omega)$ the Sobolev norm $\|\cdot\|_{2,Sob,m}$. We denote the Pseudo-Sobolev space $\mathcal{H}^m(\Omega) := \mathcal{W}_2^m(\Omega)$.

Theorem 1.1 (Hilbert Space). $\mathcal{H}^m(\Omega)$ *is a Hilbert space.*

The theory of the Sobolev spaces is vast and more general than we could have presented in this short introduction. However, our simplified theory of the Sobolev spaces is sufficient for our statistical needs. For more detailed information on Sobolev spaces we refer to Adams (1975).

1.1 Construction of Representors in Pseudo-Sobolev Space

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

In the following Theorem 1.2 we quote a representation theorem for Pseudo-Sobolev spaces derived in Yatchew and Bos (1997)—analogous to well-known Riesz Representation Theorem A.2. From now on, let us suppose that $m \in \mathbb{N}$. The symbol $\mathcal{Q}^q$ denotes closed unit cube in $\mathbb{R}^q$.

Theorem 1.2 (Representors in Pseudo-Sobolev Space). *For all* $f \in \mathcal{H}^m(\mathcal{Q}^q)$, $\mathbb{a} \in \mathcal{Q}^q$ *and* $\mathbb{w} \in \mathbb{N}_0^q$, $|\mathbb{w}|_\infty \leq m - 1$, *there exists a function* $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) \in \mathcal{H}^m(\mathcal{Q}^q)$, *s.t.*

$$\left\langle \psi_{\mathbb{a}}^{\mathbb{w}}, f \right\rangle_{Sob,m} = D^{\mathbb{w}} f(\mathbb{a}). \tag{7}$$

$\psi_{\mathbb{a}}^{\mathbb{w}}$ *is called a representor at the point* $\mathbb{a}$ *with the rank* $\mathbb{w}$. *Furthermore,*

$$\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) = \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i) \tag{8}$$

for all $\mathbb{x} \in \mathcal{Q}^q$, *where* $\psi_{a_i}^{w_i}(\cdot)$ *is the representor in the Pseudo-Sobolev space of functions of one variable on*

$\mathcal{Q}^1$ *with inner product*

$$\frac{\partial^{w_i} f(\mathbb{a})}{\partial x_i^{w_i}} = \left\langle \psi_{a_i}^{w_i}, f(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_q) \right\rangle_{Sob,m} = \sum_{\alpha=0}^{m} \int_{\mathcal{Q}^1} \frac{d^\alpha \psi_{a_i}^{w_i}(x_i)}{dx_i^\alpha} \frac{d^\alpha f(\mathbb{x})}{dx_i^\alpha} dx_i. \tag{9}$$

The proof given in Appendix B is using the idea of Yatchew and Bos (1997). In addition, we derive the exact form of the representor for Pseudo-Sobolev spaces $\mathcal{W}_p^m(\Omega)$ of both odd and even rank m.

In order to calculate the representor $\psi_a \equiv \psi_a^0$ of the function $f \in \mathcal{H}^m[0,1]$ (see (79)), we start with functions L_a and R_a defined in (96) and (97) where $a \in (0,1)$. The existence and uniqueness of the coefficients $\gamma_k(a)$ of the representor is demonstrated in the proof of Theorem 1.2. The coefficients $\gamma_k(a)$ are obtained as a solution of a system linear equations corresponding to the boundary conditions (85)–(89) of the differential equation (84).

Theorem 1.3 (Obtaining Coefficients $\gamma_k(a)$). *The coefficients $\gamma_k(a)$ of the representor ψ_a are the unique solution of the following $4m \times 4m$ system of linear equations:*

$$\begin{aligned}&\sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_k(a)\left(\varphi_k^{(m-j)}(0) + (-1)^j \varphi_k^{(m+j)}(0)\right)\\ &+ \sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{m+1+k}(a)\left(\varphi_{m+1+k}^{(m-j)}(0) + (-1)^j \varphi_{m+1+k}^{(m+j)}(0)\right) = 0, \quad j = 0, \dots, m-1;\end{aligned} \tag{10}$$

$$\begin{aligned}&\sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{2m+2+k}(a)\left(\varphi_k^{(m-j)}(1) + (-1)^j \varphi_k^{(m+j)}(1)\right)\\ &+ \sum_{\substack{k=0\\k\neq\kappa}}^{m} \gamma_{3m+3+k}(a)\left(\varphi_{m+1+k}^{(m-j)}(1) + (-1)^j \varphi_{m+1+k}^{(m+j)}(1)\right) = 0, \quad j = 0, \dots, m-1;\end{aligned} \tag{11}$$

$$\begin{aligned}&\sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_k(a) - \gamma_{2m+2+k}(a)\right)\varphi_k^{(j)}(a)\\ &+ \sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_{m+1+k}(a) - \gamma_{3m+3+k}(a)\right)\varphi_{m+1+k}^{(j)}(a) = 0, \quad j = 0, \dots, 2m-2;\end{aligned} \tag{12}$$

$$\begin{aligned}&\sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_k(a) - \gamma_{2m+2+k}(a)\right)\varphi_k^{(2m-1)}(a)\\ &+ \sum_{\substack{k=0\\k\neq\kappa}}^{m} \left(\gamma_{m+1+k}(a) - \gamma_{3m+3+k}(a)\right)\varphi_{m+1+k}^{(2m-1)}(a) = (-1)^{m-1};\end{aligned} \tag{13}$$

where

$$\kappa := \begin{cases} \frac{m}{2}, & m \text{ even}, \\ \frac{m+1}{2}, & m \text{ odd} \end{cases} \tag{14}$$

and φ_k *are defined in* (94a)–(95d).

The square system of the above system of $4m$ linear equations (10)–(13) can be written in a more illustrative way using matrix notation:

$$\underbrace{\left(\begin{array}{c|c} \begin{matrix}\varphi_k^{(m-j)}(0)\\ +(-1)^j\varphi_k^{(m+j)}(0)\end{matrix} & \emptyset \\ \hline \emptyset & \begin{matrix}\varphi_k^{(m-j)}(1)\\ +(-1)^j\varphi_k^{(m+j)}(1)\end{matrix} \\ \hline \varphi_k^{(j)}(a) & -\varphi_k^{(j)}(a) \\ \hline \varphi_k^{(2m-1)}(a) & -\varphi_k^{(2m-1)}(a) \end{array}\right)}_{\{\Gamma_{j,k}(a)\}} \underbrace{\left(\begin{array}{c} \gamma_0(a)\\ \vdots\\ \gamma_{\kappa-1}(a)\\ \gamma_{\kappa+1}(a)\\ \vdots\\ \gamma_m(a)\\ \gamma_{m+1}(a)\\ \vdots\\ \gamma_{m+\kappa}(a)\\ \gamma_{m+2+\kappa}(a)\\ \vdots\\ \gamma_{2m+1}(a)\\ \hline \gamma_{2m+2}(a)\\ \vdots\\ \gamma_{2m+1+\kappa}(a)\\ \gamma_{2m+3+\kappa}(a)\\ \vdots\\ \gamma_{3m+2}(a)\\ \gamma_{3m+3}(a)\\ \vdots\\ \gamma_{3m+2+\kappa}(a)\\ \gamma_{3m+4+\kappa}(a)\\ \vdots\\ \gamma_{4m+3}(a) \end{array}\right)}_{\{\gamma_k(a)\}} = \left(\begin{array}{c} 0\\ \vdots\\ 0\\ \hline 0\\ \vdots\\ 0\\ \hline 0\\ \vdots\\ 0\\ \hline (-1)^{m-1} \end{array}\right) \overbrace{\begin{array}{c} \scriptstyle 0\\ \vdots\\ \scriptstyle m-1\\ \scriptstyle 0\\ \vdots\\ \scriptstyle m-1\\ \scriptstyle 0\\ \vdots\\ \scriptstyle 2m-2\\ \scriptstyle 2m-1 \end{array}}^{j=} . \tag{15}$$

Hence, the coefficients $\gamma_k(a)$ are the solution of:

$$\boldsymbol{\gamma}(a) = (-1)^{m-1}\left[\{\boldsymbol{\Gamma}(a)\}^{-1}\right]_{\bullet,4m}. \tag{16}$$

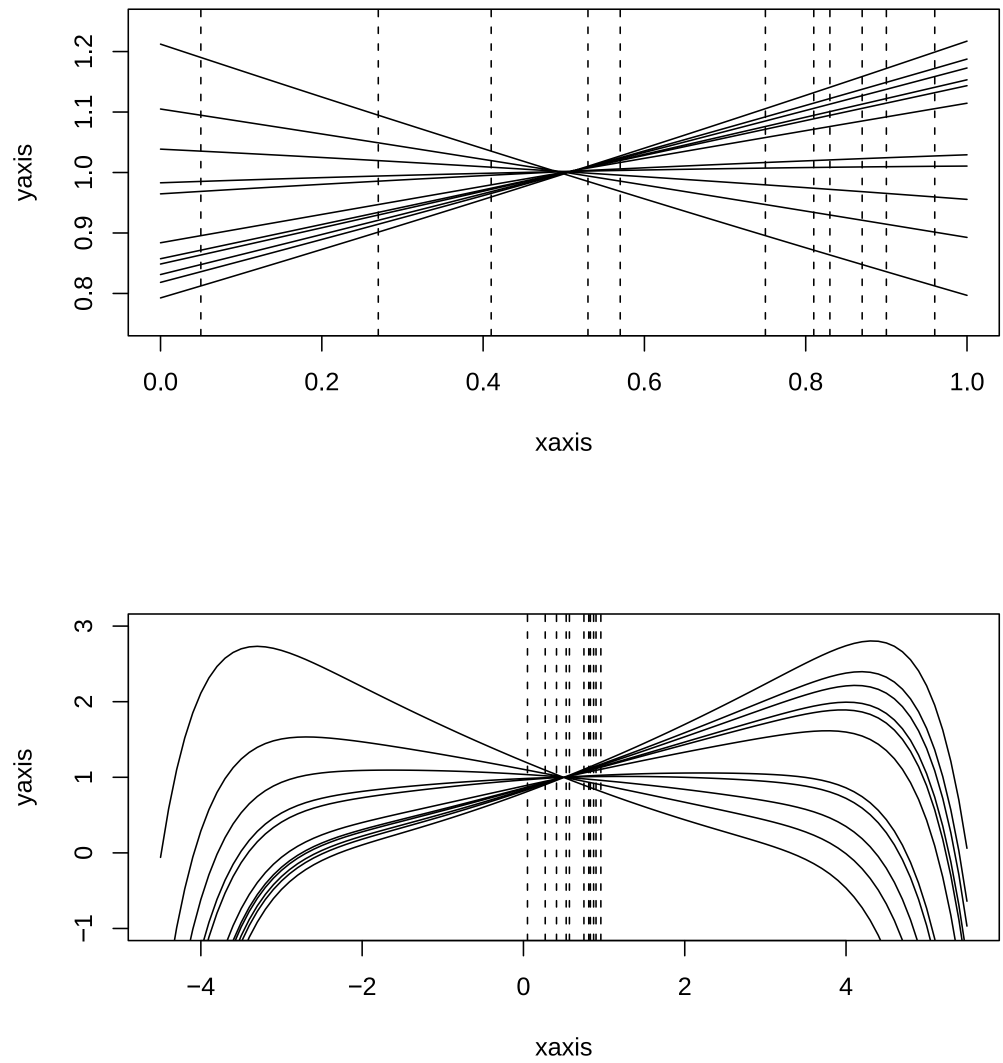

Figure 1: Representors in Pseudo-Sobolev space $\mathcal{H}^4[0,1]$ for data points $\mathbb{x} = (0.05, 0.27, 0.41, 0.53, 0.57, 0.75, 0.81, 0.83, 0.87, 0.9, 0.96)^\top$—dashed vertical lines. Zoomed view in the upper picture—interval $[0,1]$, reduced view in the lower picture—interval $[-4.5, +5.5]$.

Theorem 1.4 (Embedding). *The embedding $\mathcal{H}^m(\mathcal{Q}^q) \hookrightarrow \mathcal{C}^{m-1}(\mathcal{Q}^q)$ is compact.*

2 General Least Squares

Connection of features of L_2-spaces and $\mathcal{C}^m$-spaces can yield 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.

Definition 2.1 (General Single Equation Model). The weighted single equation model is

$$Y_i = f(\mathbb{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{17}$$

with these assumptions:

i) $\mathbb{x}_i$ are q-dimensional non-stochastic design points (knots);

ii) ε_i are random variables so that

- $\mathsf{E}\varepsilon_i = 0, \quad \forall i,$
- $\mathsf{cov}\,(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma_{ij}, & i \neq j, \\ \sigma_i^2 & i = j; \end{cases}$

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

Our setting is concerned with random variables $\{Y_i\}_{i=1}^n$, respectively $\{\varepsilon_i\}_{i=1}^n$. It is common terminology to refer to this setting as the *fixed design model*, which is concerned with controlled, non-stochastic variables $\{\mathbb{x}_i\}_{i=1}^n$.

From now on, we denote $\mathcal{H}^m \equiv \mathcal{H}^m(\mathcal{Q}^q)$, where $\mathcal{Q}^q$ is the unit cube in $\mathbb{R}^q$. We define the variance matrix $\mathbf{\Sigma} := (\sigma_{ij})_{i,j=1}^{n,n}$, where $\sigma_i^2 \equiv \sigma_{ii}$.

Our regression problem can be characterized by one of these ways:

a)
$$\min_{f \in \mathcal{H}^m} \frac{1}{n} \sum_{i=1}^n [Y_i - f(\mathbb{x}_i)]^2 \qquad \text{s.t.} \;\; \|f\|^2_{Sob,m} \leq L, \tag{18}$$

b)
$$\min_{f \in \mathcal{H}^m} \left\{ \frac{1}{n} \sum_{i=1}^n [Y_i - f(\mathbb{x}_i)]^2 + \chi \|f\|^2_{Sob,m} \right\}. \tag{19}$$

The Sobolev norm bound L and the smoothing parameter (bandwidth parameter) χ control the trade-off between the infidelity to the data and the roughness of the estimator.

Definition 2.2 (Penalizing Using General Least Squares). Optimizing using General Least Squares is

$$\min_{f \in \mathcal{H}^m} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi \left\| f \right\|_{Sob,m}^2 \tag{20}$$

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

Definition 2.3 (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 $\boldsymbol{\Psi}$ 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.

$$\Psi_{i,j} = \left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} = \psi_{\mathbb{x}_i}(\mathbb{x}_j) = \psi_{\mathbb{x}_j}(\mathbb{x}_i). \tag{21}$$

Theorem 2.1 (Infinite to Finite). *Let $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$, $\boldsymbol{\Sigma}$ an $n \times n$ positive definite and symmetric matrix and define*

$$\hat{\sigma}^2 = \min_{f \in \mathcal{H}^m} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi \left\| f \right\|_{Sob,m}^2, \tag{22}$$

$$s^2 = \min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right] + \chi \mathbb{c}^\top \boldsymbol{\Psi} \mathbb{c} \tag{23}$$

where $\mathbb{c}$ is an $n \times 1$ vector, $\boldsymbol{f}$ is declared in Definition 2.2 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 (22) *of the form*

$$\widehat{f} = \sum_{i=1}^{n} \hat{c}_i \psi_{\mathbb{x}_i} \tag{24}$$

where $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)^\top$ solves (23)*. The estimator $\widehat{f}$ is unique a.e.*

Theorem 2.1 transforms the infinite dimensional problem into a finite dimensional quadratic optimization problem. A similar theorem in Yatchew and Bos (1997) uses different penalization.

Corollary 2.2 (Form of the Regression Function Estimator). *The regression function estimator from*

Theorem 2.1 in one-dimensional case is:

$$\widehat{f}(x) = \begin{cases} \sum_{i=1}^{n} \hat{c}_i L_{x_i}(x), & 0 \leq x \leq x_1, \\ \vdots & \vdots \\ \sum_{i=j+1}^{n} \hat{c}_i L_{x_i}(x) + \sum_{i=1}^{j} \hat{c}_i R_{x_i}(x), & x_j < x \leq x_{j+1},\ j = 1, \ldots, n-1; \\ \vdots & \vdots \\ \sum_{i=1}^{n} \hat{c}_i R_{x_i}(x), & x_n < x \leq 1, \end{cases} \tag{25}$$

where $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)^\top$ *solves* (23) *and the functions* $L_{x_i}(x)$ *and* $R_{x_i}(x)$ *are defined in* (79).

Remark 2.1. Corollary 2.2 can be easily extended for a q-dimensional vector variable $\mathbb{x}$ if we recall how the representor $\psi_{\mathbb{a}}$ is produced in the proof of Theorem 1.2. We apply (79) on the form of each factor ψ_a of the product of representors $\psi_{\mathbb{a}}$ in (98). The only difference in (25) will be the number of cases. We will obtain $(n+1)^q$ decision conditions (vector $\mathbb{x}$ has q components) instead of actual number $n+1$ $(0 \leq x \leq x_1, \ldots, x_j < x \leq x_{j+1}, \ldots, x_n < x)$.

The form of the regression function estimator can be written alternatively:

$$\widehat{f}(x) = \sum_{j=1}^{n} \widehat{c_j} \sum_{k=1}^{2m} \exp\left[\Re\left(e^{i\theta_k}\right)x\right] \left\{ I_{[x \leq x_j]} \gamma_k(x_j) \cos\left[\Im\left(e^{i\theta_k}\right)x\right] + I_{[x > x_j]} \gamma_{2m+k}(x_j) \sin\left[\Im\left(e^{i\theta_k}\right)x\right] \right\}. \tag{26}$$

Note that $\widehat{f}$ is not estimated using goniometric splines neither kernel functions!

Lemma 2.3 (Symmetry of Representor Matrix). *Representor Matrix is symmetric.*

Theorem 2.4 (Positive Definiteness of Representor Matrix). *Representor Matrix is positive definite.*

In the linear model, the unknown coefficients are estimated using Least Squares. Gauss-Markov Theorem (Anděl, 2002) says that the estimate of these coefficients is the best linear unbiased estimate and underlies so-called "normal equations". The analogy can be found in our model.

Theorem 2.5 (Normal Equation for $\widehat{\mathbb{c}}$). *Assume General Single Equation Model 2.1. Let* $\mathbf{\Psi}$ *be a representor matrix. Then the vector* $\widehat{\mathbb{c}}$ *of coeficients of the unique minimizer* $\widehat{f}$ *from* (24) *is the unique solution with respect to* $\mathbb{c} = (c_1, \ldots, c_n)^\top$ *of the equation system*

$$\left(\mathbf{\Psi}\mathbf{\Sigma}^{-1}\mathbf{\Psi} + n\chi\mathbf{\Psi}\right)\mathbb{c} = \mathbf{\Psi}\mathbf{\Sigma}^{-1}\mathbb{Y} \tag{27}$$

for response vector $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$.

Remark 2.2 *(Hat Matrix).* We know that

$$\widehat{\mathbb{Y}} = \widehat{\boldsymbol{f}}(\mathbb{x}) = \boldsymbol{\Psi}\widehat{\mathbb{c}} \tag{28}$$

and from previous Normal Equation for $\widehat{\mathbb{c}}$ Theorem 2.5 we easily obtain the form of the projection "hat" matrix

$$\boldsymbol{\Lambda} := \boldsymbol{\Psi}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi} + n\chi\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1} \tag{29}$$

such that

$$\widehat{\mathbb{Y}} = \boldsymbol{\Lambda}\mathbb{Y}. \tag{30}$$

If the variance matrix $\boldsymbol{\Sigma}$ has full rank, its inverse matrix has also full rank and it can be decomposed using its square root matrix $\boldsymbol{\Xi}$ (formally proceeded by spectral decomposition).

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Xi}^{\top}\boldsymbol{\Xi}. \tag{31}$$

Notice that this square root matrix $\boldsymbol{\Xi}$ has full rank.

According to the Infinite to Finite Theorem 2.1 and Lagrange Multiplier Theorem it can be easily seen that there is a correspondence between Sobolev bound L and and smoothing parameter χ. It would be only a technical exercise to prove the 1–1 mapping between these two parameters (Pešta, 2006).

Theorem 2.6 (1–1 Mapping of Smoothing Parameters). *Let $L > 0$, $\boldsymbol{\Sigma}$ is positive definite and symmetric matrix and*

$$f^{*} = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^{\top}\boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] \quad s.t. \;\; \|f\|_{Sob,m}^2 \leq L \tag{32}$$

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

$$f^{*} = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^{\top}\boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi\|f\|_{Sob,m}^2 . \tag{33}$$

Hence, there exists a 1–1 mapping $\mathscr{L}\colon \mathbb{R}^{+} \to \mathbb{R}^{+}\colon L \mapsto \chi$.

Theorem 2.7 (Bijection Between the Smoothing Parameters). *Let $\chi > 0$, $\boldsymbol{\Sigma}$ is positive definite and symmetric matrix and*

$$f^{*} = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^{\top}\boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi\|f\|_{Sob,m}^2 \tag{34}$$

then there exists a unique $L > 0$ *such that*

$$f^* = \arg\min_{f \in \mathcal{H}^m} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] \quad s.t. \ \ \|f\|^2_{Sob,m} = L. \tag{35}$$

If ${\mathbb{c}^*}^\top \boldsymbol{\Psi} \mathbb{c}^* < L$, then we are talking about the interpolation not the approximation because $\mathbb{Y} = \boldsymbol{\Psi}\mathbb{c}^*$. This is very unusual case for a real statistical situation and our problem, too.

Theorem 2.8 (Asymptotic Behavior). *Suppose* $\tilde{\boldsymbol{\varepsilon}} := \boldsymbol{\Xi}\boldsymbol{\varepsilon}$ *is* $n \times 1$ *vector of i.i.d. random variables. Then*

$$\frac{1}{n} \left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] = \mathcal{O}_{\mathsf{P}}\left(n^{-\frac{2m}{2m+q}}\right), \quad n \to \infty. \tag{36}$$

2.1 Choice of the Smoothing Parameter

The smoothing parameter χ corresponds to the diameter of the set of functions over which the estimation takes place. Heuristically, for large bounds ($\equiv$ smaller χ), we obtain consistent but less efficient estimator. On the other hand, for smaller bounds, i.e., large χ, we obtain more efficient but inconsistent estimators.

Well-known selection method is minimization of the cross-validation criterion

$$\mathcal{CV}(L) = \frac{1}{n} \left[\mathbb{y} - \widehat{\mathbf{f}^*}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{y} - \widehat{\mathbf{f}^*}(\mathbb{x})\right] \tag{37}$$

where $\widehat{\mathbf{f}^*} = \left(\widehat{f}_{-1}, \ldots, \widehat{f}_{-n}\right)^\top$ is obtained by solving

$$\widehat{f}_{-i} = \arg\min_{f \in \mathcal{H}^m} \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left[\boldsymbol{\Xi}_{j,\bullet}\mathbb{y} - \boldsymbol{\Xi}_{j,\bullet}\mathbf{f}(\mathbb{x})\right]^2 + \chi \|f\|^2_{Sob,m}, \quad i = 1, \ldots, n, \tag{38}$$

where $\boldsymbol{\Xi}$ is the square root of the inverse of matrix $\boldsymbol{\Sigma}$ defined in (31).

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 χ (which correspodents to appropriate Sobolev bound L). Relationship between the data-fit and the smoothness of estimator is shown in Figure 2.

We can also use weighted version of cross-validation, called generalized cross-validation

$$\mathcal{GCV}(L) = \frac{\left[\mathbb{y} - \widehat{\mathbf{f}}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{y} - \widehat{\mathbf{f}}(\mathbb{x})\right]}{\mathsf{tr}\left(\mathbf{I} - \boldsymbol{\Lambda}\right)^2}. \tag{39}$$

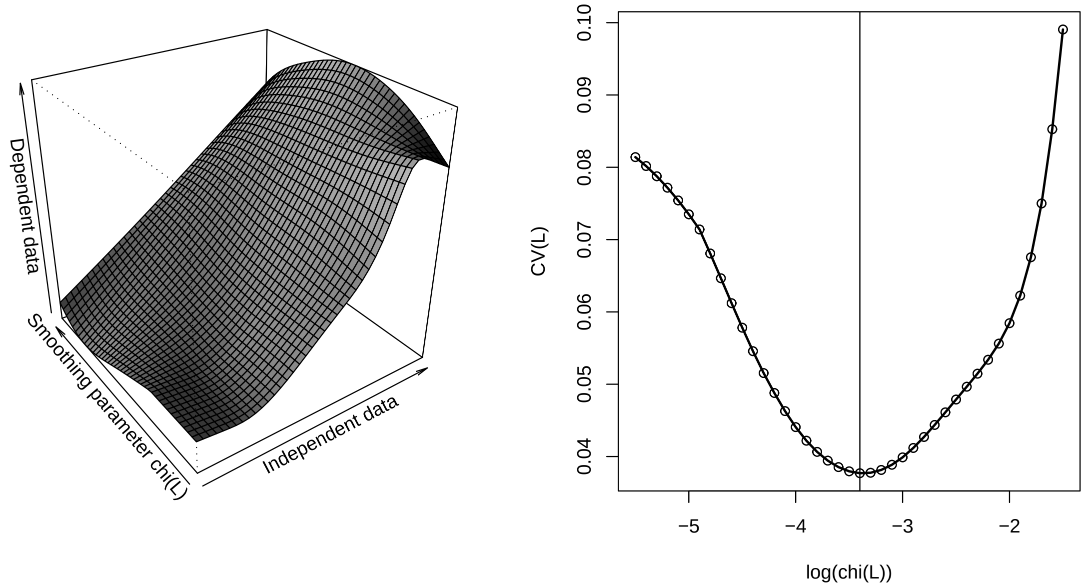

Figure 2: Left—changing monotone curve in $\mathcal{H}^2$ depending upon smoothing parameter. Right—optimal value of smoothing parameter according to Cross-Validation.

Detailed information concerning a choice of the smoothing parameter χ can be found in Eubank (1999).

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

3 Application to Option Prices

In the regression in Pseudo-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, the estimators should underlie additional constraints. We focus on the imposition of additional constraint—such as isotonia—on the nonparametric regression estimator.

We estimate $f \in \widetilde{\mathscr{F}} \subseteq \mathscr{F}$ where $\widetilde{\mathscr{F}}$ combines smoothness with further functional properties. We consider properties such as monotonicity of particular derivatives of the function, i.e., monotonicity, convexity, etc.

The following discussion concerns only the one-dimensional case. From now on, we assume that $q = 1$.

Definition 3.1 (Derivative of Representor Matrix). Let $\psi_{x_1}, \ldots, \psi_{x_n}$ be the representors for function evaluation at $x_1, \ldots, x_n$, 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 $\mathbf{\Psi}^{(k)}$ be the k-th derivative of $n \times n$ representor matrix whose columns are equal to the k-th derivatives of the representors evaluated at $x_1, \ldots, x_n$; i.e.

$$\Psi^{(k)}_{i,j} = \psi^{(k)}_{x_j}(x_i), \quad i, j = 1, \ldots, n. \tag{40}$$

It is very important that the k-th derivative of the representor matrix is defined in a "column" way. In spite of Theorem 2.3, derivative of representor matrix needn't to be a symmetric one.

Definition 3.2 (Estimate of the Derivative). Define the estimate of the regression function derivative as the derivative of the regression function estimate, i.e.

$$\widehat{f^{(s)}} := \widehat{f}^{(s)}, \quad s \in \mathbb{N}. \tag{41}$$

Theorem 3.1 (Consistency of Estimator). *Suppose* $\tilde{\boldsymbol{\varepsilon}} := \boldsymbol{\Xi}\boldsymbol{\varepsilon}$ *is* $n \times 1$ *vector of i.i.d. random variables, the design points are equidistantly distributed on interval* $[a, b]$ *such that* $a = x_1 < \ldots < x_n = b$ *and* $\boldsymbol{\Sigma}$ *is positive definite covariance matrix of* $\boldsymbol{\varepsilon}$ *with its largest eigenvalue less or equal to a positive constant* $\vartheta > 0$ *for all* $n \in \mathbb{N}$. *Then for* $s = 0, \ldots, m-2$

$$\sup_{x \in [a,b]} \left| \widehat{f^{(s)}}(x) - f^{(s)}(x) \right| \xrightarrow[n \to \infty]{\mathsf{P}} 0. \tag{42}$$

Now we can show the relationship between the operator of derivative of the representor matrix and the isotonia, especially in the application to the Call and Put option properties.

3.1 Call and Put options

Let's have multiple Call and Put option prices for some strike prices. Suppose that the Call and the Put option prices are repeated observations at distinct fixed design points ϖ_i, $i = 1, \ldots, \omega$, called the strike price knots. The Call option prices represents $n_i \in \mathbb{N}_0$ responses Y_{i_k} for their strike prices $x_{i_k} \in \{\varpi_1, \ldots, \varpi_\omega\}$ in each strike price knot ϖ_i where $k = 1, \ldots, n_i$, $x_{i_1} = \ldots = x_{i_{n_i}}$, $\forall i$,

$$\sum_{i=1}^{\omega} I_{[n_i \geq 1]} = \omega_Y, \tag{43}$$

$$\sum_{i=1}^{\omega} n_i I_{[n_i \geq 1]} = n. \tag{44}$$

Similarly the Put option prices represents $m_j \in \mathbb{N}_0$ responses Z_{j_l} for their strike prices $x_{j_l} \in \{\varpi_1, \ldots, \varpi_\omega\}$ in each strike price knot ϖ_j where $l = 1, \ldots, n_j$, $x_{j_1} = \ldots = x_{j_{m_j}}$, $\forall j$,

$$\sum_{j=1}^{\omega} I_{[m_j \geq 1]} = \omega_Z, \quad (45)$$

$$\sum_{j=1}^{\omega} m_j I_{[m_j \geq 1]} = m. \quad (46)$$

Let $\boldsymbol{\Delta}$ be the connectivity $n \times \omega_Y$ matrix for Call option strike prices such that

$$\Delta_{ij} := \begin{cases} 1 & \text{if } x_i = \varpi_j, \\ 0 & \text{otherwise} \end{cases} \quad (47)$$

for

$$i \in \{\iota \,|\, 1 \leq \iota \leq n \,\&\, n_\iota \geq 1\}, \quad (48)$$

$$j \in \{\varsigma \,|\, 1 \leq \varsigma \leq \omega \,\&\, n_\varsigma \geq 1\} \quad (49)$$

and also let $\boldsymbol{\Theta}$ be the connectivity $m \times \omega_Z$ matrix for Put option strike prices such that

$$\Theta_{ij} := \begin{cases} 1 & \text{if } x_i = \varpi_j, \\ 0 & \text{otherwise} \end{cases} \quad (50)$$

for

$$i \in \{\iota \,|\, 1 \leq \iota \leq m \,\&\, m_\iota \geq 1\}, \quad (51)$$

$$j \in \{\varsigma \,|\, 1 \leq \varsigma \leq \omega \,\&\, m_\varsigma \geq 1\}. \quad (52)$$

Similar matrix has been already defined in Yatchew and Härdle (2005).

Definition 3.3 (Call and Put Option Model). Invoke the notation from the beginning of this section 3. The Call and Put option model is

$$Y_{i_k} = f(x_{i_k}) + \varepsilon_{i_k}, \qquad k = 1, \ldots, n, \quad \{i_1, \ldots, i_n\} \subseteq \{1, \ldots, \nu\}, \quad (53)$$

$$Z_{j_l} = g(x_{j_l}) + \epsilon_{j_l}, \qquad l = 1, \ldots, m, \quad \{j_1, \ldots, j_m\} \subseteq \{1, \ldots, \nu\} \quad (54)$$

with these assumptions:

i) $\{x_i\}_{i=1}^{\nu}$ are non-stochastic design points such that $x_i \in \{\varpi_1, \ldots, \varpi_\omega\}, \forall i$;

ii) ε_{i_k} and ϵ_{j_l} are random variables so that

- $\mathsf{E}\varepsilon_{i_k} = 0, \quad \forall k,$
- $\mathsf{E}\epsilon_{j_l} = 0, \quad \forall l,$
- $\mathsf{cov}\left(\varepsilon_{i_k}, \varepsilon_{i_l}\right) = \begin{cases} \xi_{i_k,i_l}, & k \neq l, \\ \xi_{i_k}^2 & k = l; \end{cases}$
- $\mathsf{cov}\left(\epsilon_{j_l}, \epsilon_{j_k}\right) = \begin{cases} \zeta_{j_l,j_k}, & l \neq k, \\ \zeta_{j_l}^2 & l = k; \end{cases}$
- $\mathsf{cov}\left(\varepsilon_{i_k}, \epsilon_{j_l}\right) = \sigma_{i_k,j_l}, \quad \forall k, l;$

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

The second derivatives of functions f and g have to be the same SPD. Hence, Infinite to Finite Theorem 2.1 provides a key to determining how to handle multiple (repeated) observations for our set-up in option prices model 3.3.

Theorem 3.2 (Call and Put Option Optimizing). *Invoke the assumptions from Call and Put Option Model 3.3. Define*

$$\begin{aligned} \hat{\sigma}^2 = \min_{f \in \mathcal{H}^p, g \in \mathcal{H}^p} & \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{f}\left(\mathbb{x}_\alpha\right) \\ \mathbf{g}\left(\mathbb{x}_\beta\right) \end{pmatrix}\right]^\top \\ & \boldsymbol{\Sigma}^{-1} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{f}\left(\mathbb{x}_\alpha\right) \\ \mathbf{g}\left(\mathbb{x}_\beta\right) \end{pmatrix}\right] \\ & + \chi \|f\|_{Sob,p}^2 + \theta \|g\|_{Sob,p}^2 \end{aligned} \tag{55}$$

subject to

$$-\mathbf{1} \leq \mathbf{f}'\left(\mathbb{x}_\alpha\right) \leq \mathbf{0}, \tag{56a}$$

$$\mathbf{0} \leq \mathbf{g}'\left(\mathbb{x}_\beta\right) \leq \mathbf{1}, \tag{56b}$$

$$\mathbf{f}''\left(\mathbb{x}_\alpha\right) \geq \mathbf{0}, \tag{56c}$$

$$\mathbf{g}''\left(\mathbb{x}_\beta\right) \geq \mathbf{0}, \tag{56d}$$

$$\mathbf{f}''\left(\mathbb{x}_\gamma\right) = \mathbf{g}''\left(\mathbb{x}_\gamma\right) \tag{56e}$$

and

$$s^2 = \min_{\mathbb{c}\in\mathbb{R}^{\omega_Y},\mathbb{d}\in\mathbb{R}^{\omega_Z}} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi} \end{pmatrix} \begin{pmatrix} \mathbb{c} \\ \mathbb{d} \end{pmatrix} \right]^\top \boldsymbol{\Sigma}^{-1} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi} \end{pmatrix} \begin{pmatrix} \mathbb{c} \\ \mathbb{d} \end{pmatrix} \right] + \chi\mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} + \theta\mathbb{d}^\top\boldsymbol{\Phi}\mathbb{d} \tag{57}$$

subject to

$$-\mathbf{1} \leq \boldsymbol{\Psi}^{(1)}\mathbb{c} \leq \mathbf{0}, \tag{58a}$$

$$\mathbf{0} \leq \boldsymbol{\Phi}^{(1)}\mathbb{d} \leq \mathbf{1}, \tag{58b}$$

$$\boldsymbol{\Psi}^{(2)}\mathbb{c} \geq \mathbf{0}, \tag{58c}$$

$$\boldsymbol{\Phi}^{(2)}\mathbb{d} \geq \mathbf{0}, \tag{58d}$$

$$\boldsymbol{\Psi}^{(2)}\mathbb{c}_\gamma = \boldsymbol{\Phi}^{(2)}\mathbb{d}_\gamma \tag{58e}$$

where $\chi > 0$, $\theta > 0$, $\boldsymbol{\Sigma}$ *is the* $(n+m)\times(n+m)$ *positive definite and symmetric matrix,* $\boldsymbol{\Delta}$ *is the connectivity* $n \times \omega_Y$ *matrix from* (47), $\boldsymbol{\Theta}$ *is the connectivity* $m \times \omega_Z$ *matrix from* (50), $\boldsymbol{\Psi}$ *is the* $\omega_Y \times \omega_Y$ *representor matrix at* $(x_\iota)^\top_{\iota\in\{\iota\,|\,n_\iota\geq 1\}}$, $\boldsymbol{\Phi}$ *is the* $\omega_Z \times \omega_Z$ *representor matrix at* $(x_\iota)^\top_{\iota\in\{\iota\,|\,m_\iota\geq 1\}}$, $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$, $\mathbb{Z} = (Z_1, \ldots, Z_m)^\top$, $\mathbf{f}(\mathbb{x}_\alpha) = (f(x_\iota))^\top_{\iota\in\{\iota\,|\,n_\iota\geq 1\}}$, $\mathbf{g}(\mathbb{x}_\beta) = (g(x_\iota))^\top_{\iota\in\{\iota\,|\,m_\iota\geq 1\}}$ *and* $\gamma := \alpha\cap\beta = (\iota\,|\,n_\iota \geq 1\,\&\,m_\iota \geq 1)^\top$ *is the vector of indices in increasing order. Then* $\hat{\sigma}^2 = s^2$. *Furthermore, there exists a solution to* (55) *with respect to* (56) *of the form*

$$\widehat{f} = \sum_{\{i\,|\,n_i\geq 1\}} \hat{c}_i\psi_{x_i}, \tag{59}$$

$$\widehat{g} = \sum_{\{j\,|\,m_j\geq 1\}} \hat{d}_j\phi_{x_j} \tag{60}$$

where $\widehat{\mathbb{c}} = (\hat{c}_i)^\top_{i\in\{i\,|\,n_i\geq 1\}}$ *and* $\widehat{\mathbb{d}} = (\hat{d}_j)^\top_{j\in\{j\,|\,m_j\geq 1\}}$ *solves* (57), ψ_{x_i} *is the representor at* x_i *for vector* $(x_\iota)^\top_{\iota\in\{\iota\,|\,n_\iota\geq 1\}}$ *and* ϕ_{x_j} *is the representor at* x_j *for vector* $(x_\iota)^\top_{\iota\in\{\iota\,|\,m_\iota\geq 1\}}$. *The estimators* $\widehat{f}$ *and* $\widehat{g}$ *are unique a.e.*

The structure of the $(n+m)\times(n+m)$ covariance matrix $\boldsymbol{\Sigma}$ of the random errors $(\varepsilon_1, \ldots, \varepsilon_n, \epsilon_1, \ldots, \epsilon_m)^\top$ will be investigated in Section 4. The minimization problem (57) under the constraints (58) can be implemented using GNU–R statistical software with function `pcls()` in the library `mgcv`.

4 Covariance Structure

Let us denote the vector of the true SPD in the ω distinct observed strike prices $\varpi_1, \ldots, \varpi_\omega$ as $f^{(2)} = (f^{(2)}(\varpi_1), \ldots, f^{(2)}(\varpi_\omega))^\top$. Assume that the expected values of the option prices given in (1) and (2) can be approximated by a linear combination of this discretized version of the SPD, i.e., we assume a linear model

$$Y_i = \alpha(x_i)^\top f^{(2)} + \varepsilon_i \tag{61}$$

for the Call option prices and

$$Z_j = \beta(x_j)^\top f^{(2)} + \varepsilon_j \tag{62}$$

for the Put option prices, $i = 1, \ldots, n$, $j = 1, \ldots, m$. We assume that the vectors of the coefficients $\alpha(x)$ and $\beta(x)$ depend only on the strike price x and can be interpreted as rows of a design matrices $\mathcal{X}_\alpha$ and $\mathcal{X}_\beta$, respectively. In the following, the state price density $f^{(2)}$ may depend on the time of the observation and the symbol $f_i^{(2)} = (f_i^{(2)}(\varpi_1), \ldots, f_i^{(2)}(\varpi_\omega))^\top$ will denote the true value of the SPD at the time of the i-th trade, $i = 1, \ldots, n + m$.

4.1 Constant SPD

Assuming that the random errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{n+m})^\top$ in the linear model

$$\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} = \begin{pmatrix} \mathcal{X}_\alpha \\ \mathcal{X}_\beta \end{pmatrix} f^{(2)} + \boldsymbol{\varepsilon}, \tag{63}$$

are independent and identically distributed, the model (63) for the i-th observation, corresponding to the strike price x_i, can be written as

$$\begin{aligned} Y_i &= \alpha(x_i)^\top f_i^{(2)} + \varepsilon_i \\ f_i^{(2)} &= f^{(2)} \end{aligned}$$

if the i_k-th observation is a Call option price or

$$\begin{aligned} Z_i &= \beta(x_i)^\top f_i^{(2)} + \varepsilon_i \\ f_i^{(2)} &= f^{(2)} \end{aligned}$$

if the i-th observations is a Put option price. Here, the estimated parameter (SPD) does not change during the observation period (one day).

This simplified model has been estimated in Yatchew and Härdle (2005) only for Call option prices.

4.2 Dependencies due to the time of the trade

Let us now assume that the observations are sorted according to the time of the trade $t_i \in (0,1)$ and denote by $\delta_i = t_i - t_{i-1} > 0$ the time between the $(i-1)$-st and the i-th trade.

The model described in Subsection 4.1 can now be generalized by moving the iid random errors ε_i to the SPD $f_i^{(2)}$ rather than to the observed call option price:

$$\begin{aligned} Y_i &= \alpha(x_i)^\top f_i^{(2)}, \\ f_i^{(2)} &= f_{i-1}^{(2)} + \delta_i^{1/2} \varepsilon_i. \end{aligned}$$

Expressing all observations in terms of the parameter $f_{n+1}^{(2)}$, corresponding to the "end of the day", it follows that the covariance of any two observed call option prices depends only on the time of the trade and their strike prices:

$$\begin{aligned} \mathrm{Cov}\{Y_{i-u}, Y_{i-v}\} &= \mathrm{Cov}(\alpha(x_{i-u})^\top f_{i-u}^{(2)}, \alpha(x_{i-v})^\top f_{i-v}^{(2)}) \\ &= \sigma^2 \alpha(x_{i-u})^\top \alpha(x_{i-v}) \sum_{m=1}^{\min(u,v)} \delta_{i+1-m}. \end{aligned} \tag{64}$$

Similarly, we obtain the covariances between the observed Put option prices:

$$\begin{aligned} \mathrm{Cov}\{Z_{i-u}, Z_{i-v}\} &= \mathrm{Cov}(\beta(x_{i-u})^\top f_{i-u}^{(2)}, \beta(x_{i-v})^\top f_{i-v}^{(2)}(k)) \\ &= \sigma^2 \beta(x_{i-u})^\top \beta(x_{i-v}) \sum_{l=1}^{\min(u,v)} \delta_{i+1-l}. \end{aligned} \tag{65}$$

and the covariance between the observed Put and Call option prices is:

$$\begin{aligned} \mathrm{Cov}\{Y_{i-u}, Z_{i-v}\} &= \mathrm{Cov}(\alpha_{x_{i-u}}^\top f_{i-u}^{(2)}, \beta(x_{i-v})^\top f_{i-v}^{(2)}(k)) \\ &= \sigma^2 \sum_{l=1}^{\min(u,v)} \delta_{i+1-l} \sum_{k=2}^{p-1} \alpha_{x_{i-u}}^\top \beta(x_{i-v}). \end{aligned} \tag{66}$$

Hence, the knowledge of the time of the trades allows us to estimate the covariance matrix of the observed option prices. Note that with this covariance structure we can estimate arbitrary future value of the SPD. It is natural that more recent observations are more important for the construction of the estimator and that observations corresponding to the same strike price and taken at approximately same time will be highly correlated.

5 DAX Option Prices

In this section, the theory developed in the previous sections is applied on real data set containing intra day Call and Put DAX option prices in year 1995. The data set, Eurex Deutsche Börse, was provided by the Financial and Economic Data Center (FEDC) at Humboldt-Universität zu Berlin in the framework of the SFB 649 Guest Researcher Program for Young Researchers.

In Figures 3 and 4, we present the analysis for the first two trading days in January 1995. On the first trading day, the time to expiry was $T - t = 0.05$ years, i.e., 18 days. Naturally, on the second trading day, the time to expiry was 17 days.

In both figures, the first two plots contain the fitted Put and Call option prices and the estimated SPD. Both smoothing parameters were chosen as 2×10^{-5} leading to a reasonably smooth SPD estimate in the upper right plot in Figures 3 and 4. Smaller values of the smoothing parameters would lead to a more variable and less smooth SPD estimates that would be difficult to interpret.

The second two plots in Figures 3 and 4 show ordinary residual plots separately for the observed Put and Call option prices. The size of each plotting symbol denotes the number of residuals lying in the respective area. The shape of the plotting symbols corresponds to the time of the trade, circles occurred in the morning, squares around the noon and the stars in the afternoon. We observe a strong heteroscedasticity and strong dependencies due to the time of the trade.

In the last two plots in Figures 3 and 4, we plot the same residuals transformed by Mahalanobis transformation, i.e., multiplied by the inverse square root of their assumed covariance matrix, see Section 4.2. This transformation removes most of the dependencies caused by the time of the trade. However, some outlying observations have now appeared. For example, for the Call options on the second day, plotted in Figure 4, we can see a very large positive and a very large negative residual at the same strike price 2050.

The outlying observations can be explained if we have a closer look at the original data set. In Table 1, we show the Call option prices, times of the trades, and the transformed residuals for all trades with the strike price $K = 2050$. The two observations with larg residuals, 358.7 and -342.2, occurred at approximately the same time, the time difference between them is approximately 0.13 hours, i.e., approximately five minutes. Simultaneously, the price difference of these two observations is quite large. Hence, the large correlation of these two very different prices leads to the large (suspicious) residuals appearing in the residual plot.

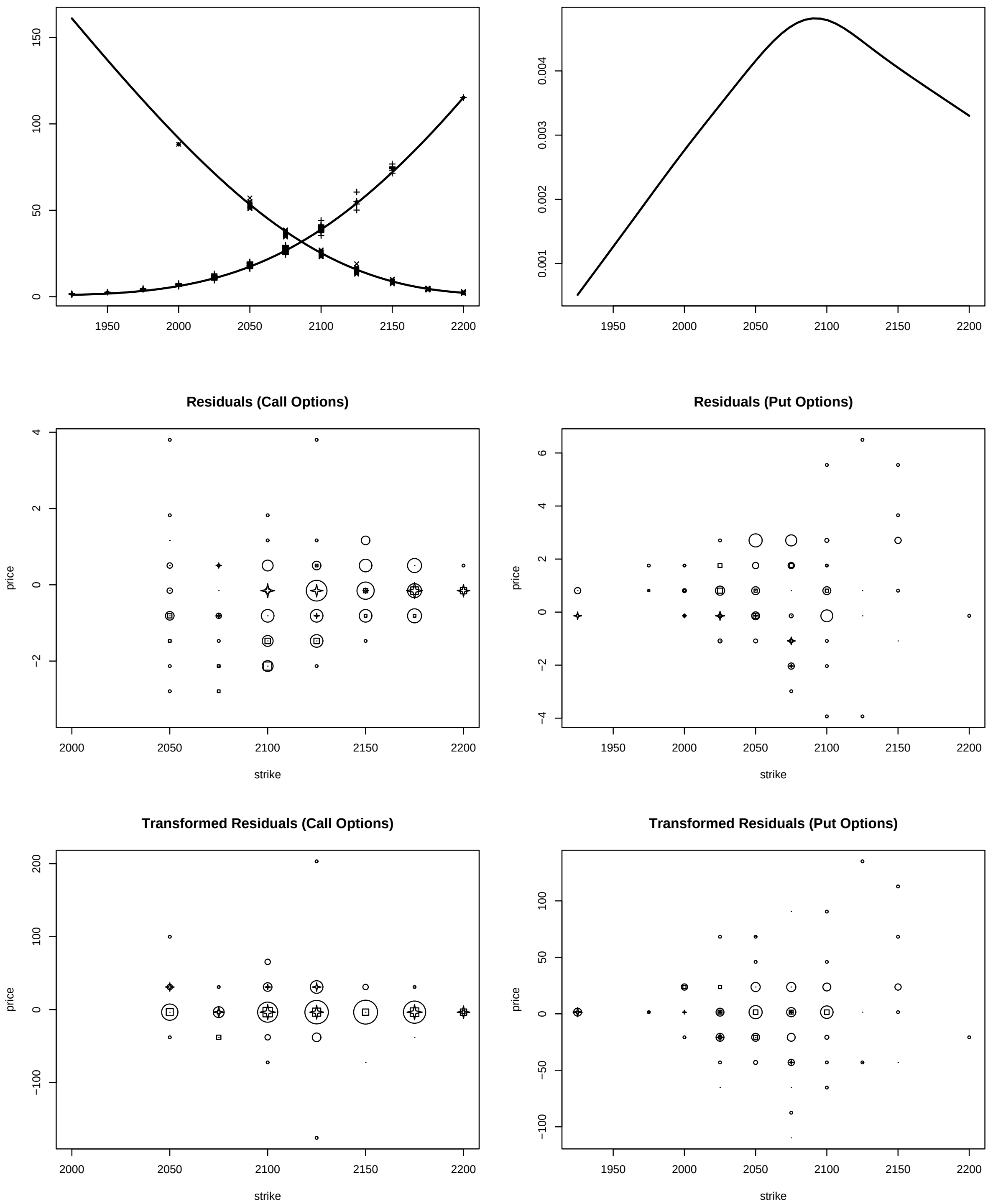

Figure 3: Estimates and residual plots on the 1st trading day in 1995 (January 2nd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and right hand side. The residuals plotted in the last two plots were corrected by the inverse square root of the covariance matrix.

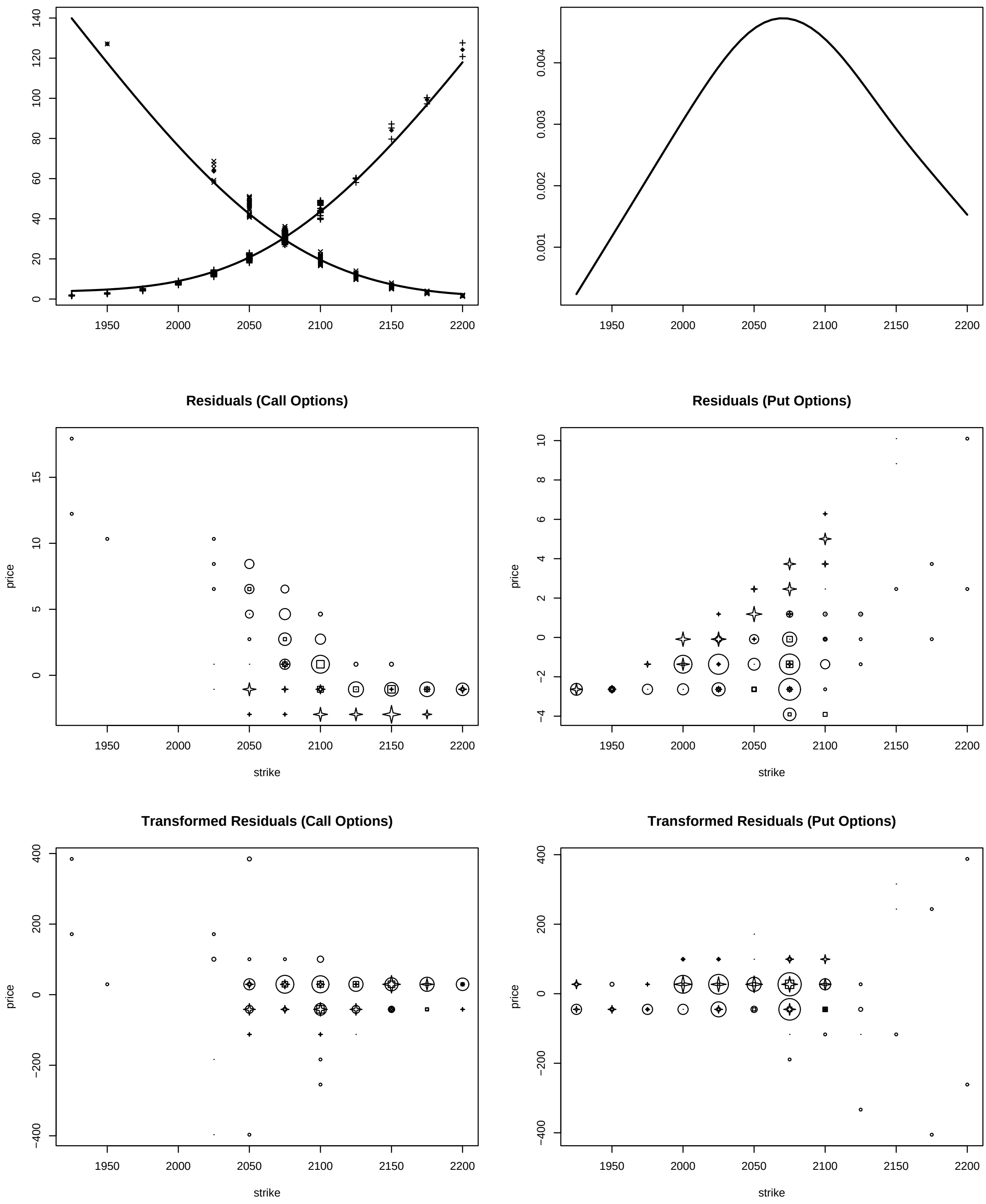

Figure 4: Estimates and residual plots on 2nd trading day in 1995 (January 3rd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and right hand side. The residuals plotted in the last two plots were corrected by the inverse square root of the covariance matrix.

Call price ($K = 2050$)	time (in hours)	transformed residual
50.62296	9.690	337.4
51.12417	9.702	73.2
50.62296	9.785	33.8
50.02150	9.807	6.5
48.11687	9.826	−10.3
46.61322	9.864	−11.5
47.31492	10.121	−6.9
48.11687	10.171	26.5
49.01906	10.306	24.3
49.01906	10.361	26.3
50.32223	10.534	358.7
46.61322	10.666	−342.2
47.61565	10.672	32.8
45.00932	11.187	−62.2
48.11687	11.690	28.2
45.10957	12.100	−72.6
48.11687	12.647	53.9
48.11687	12.766	13.3
48.11687	13.170	28.3
47.51541	14.205	11.2
44.10713	14.791	−4.8
42.10226	15.137	−34.1
42.10226	15.138	−93.4
40.99958	15.232	−32.4
41.60104	15.250	−14.2
42.10226	15.283	−2.4
42.10226	15.288	−87.6
40.69885	15.638	−31.2
41.60104	15.658	−48.9
42.60348	15.711	−46.6
42.10226	15.715	6.7
41.60104	15.796	−39.2
42.10226	15.914	−49.5

Table 1: Subset of observed prices of Call options on 2nd trading day in 1995 for strike price $K = 2050$, time of the trade in hours and residuals transformed by the Mahalanobis transformation. The fitted value for the strike price $K = 2050$ is $\hat{f}^{(2)}(2050) = 42.37$. This value can be interpreted as an estimate corresponding to 16:00 o'clock.

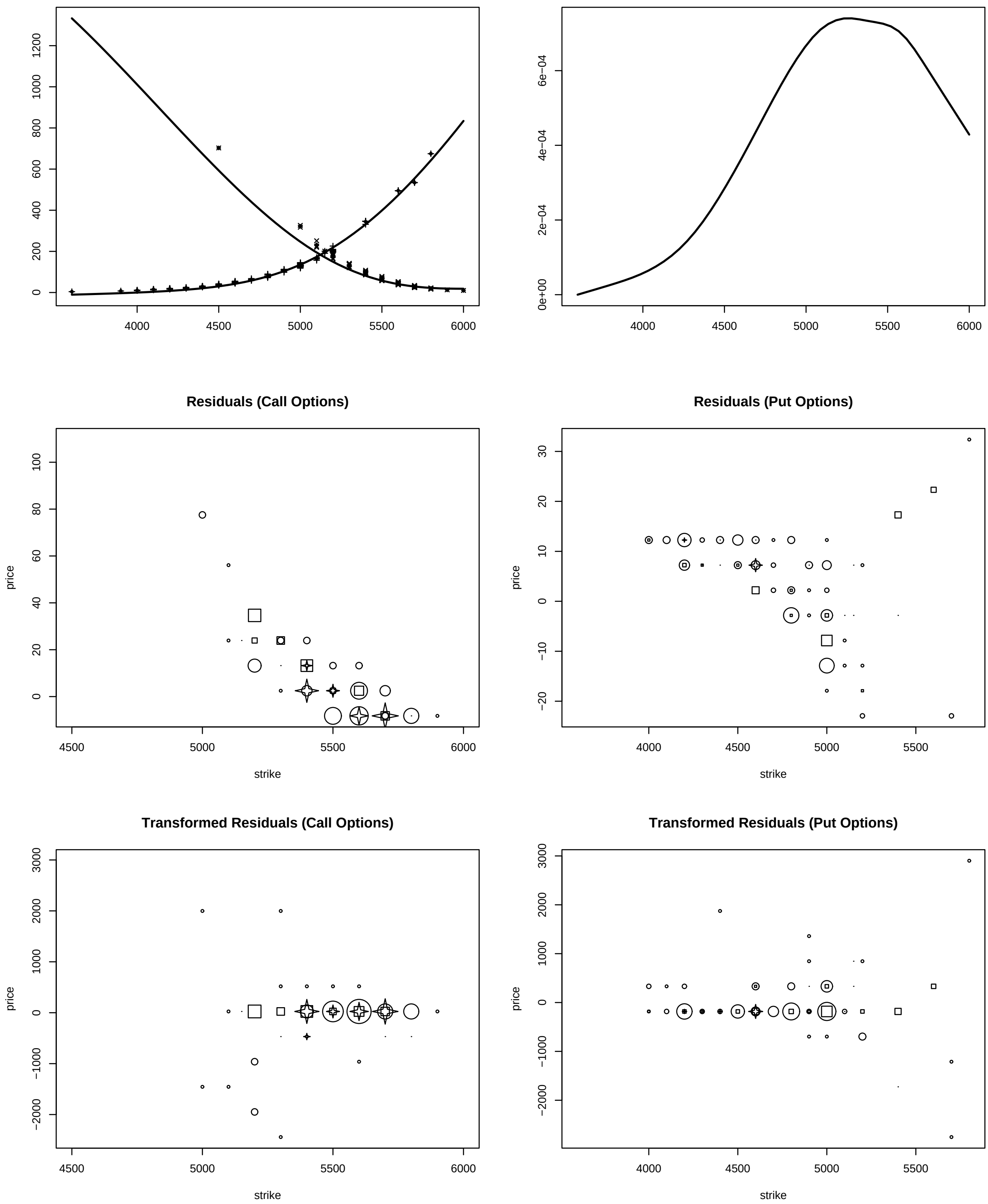

Figure 5: Estimates and residual plots on the 1st trading day in 2002 (January 2nd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and right hand side. The residuals plotted in the last two plots were corrected by the inverse square root of the covariance matrix.

An example of a more recent data set is plotted in Figure 5. In year 2002, the range of the traded strike prices was much wider than in 1995. The estimated SPD is plotted in the upper right plot. The estimate could be described as a unimodal probability density function with the right tail cut off. It seems that, especially on the right hand side, the traded strike prices do not cover the support of the SPD entirely.

The residual plots in Figure 5 look very similar to the residual plots in Figures 3 and 4. The residual analysis suggests that the simple model for the covariance structure presented in Section 4 is more appropriate for this estimation problem than the unrealistic iid assumptions. In practice, the traded strike prices do not cover the entire support of the SPD. Hence, our estimators recover only the central part of the SPD in Figures 3 and 4 or the left hand part of the SPD in Figure 5. Unfortunately, this implies that we cannot impose any conditions on the expected value of the SPD without additional distributional assumptions.

6 Conclusion

The mathematical foundation of the constrained regression in pseudo-Sobolev spaces is explained in Section 1, see also Yatchew and Bos (1997); Yatchew and Härdle (2005). In Section 2, we generalize the method to dependent observations and introduce the constrained general regression in pseudo-Sobolev spaces. The application of the method to the observed option prices is developed in Section 3. The resulting algorithm, using the covariance structure given in Section 4, see also Härdle and Hlávka (2006), is applied on a real data set in Section 5.

The main achievement of this paper is the simultaneous estimation of the SPD from both Put and Call option prices and the incorporation of the covariance structure in the nonparametric estimator that has been previously considered in Yatchew and Härdle (2005). The constrained general regression in pseudo-Sobolev spaces will certainly be very useful in various practical problems.

A Used definitions and theorems

A.1 Definitions

Definition A.1 (Domain). A connected Lebesgue-measurable (open or closed) bounded subset Ω of an Euclidean space $\mathbb{R}^q$ with non-empty interior is called a domain.

Definition A.2 (Lebesgue Space). Consider a measurable 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

$$\|f\|_{\mathsf{L}_\mathsf{p}(\Omega)} := \begin{cases} \left(\int_\Omega f^p(\mathbb{x})\mathrm{d}\mathbb{x}\right)^{1/p} & \text{for } 1 \le p < \infty, \\ \inf\left\{C \ge 0 : |f| \le C \text{ a.e.}\right\} & \text{for } p = \infty. \end{cases} \tag{67}$$

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

Definition A.3 (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

$$\mathcal{C}^m(\Omega) := \left\{f : \Omega \to \mathbb{R} \middle| D^{\boldsymbol{\alpha}} f \in \mathcal{C}^0(\Omega), |\boldsymbol{\alpha}|_\infty \le m\right\}, \tag{68}$$

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

Definition A.4 (General Definition of Sobolev Space). A Sobolev space can be also defined in more general way:

$$\mathcal{W}_p^m(\Omega) := \{f \in \mathsf{L}_\mathsf{p}(\Omega) | D_w^{\boldsymbol{\alpha}} f \in \mathsf{L}_\mathsf{p}(\Omega), |\boldsymbol{\alpha}|_\infty \le m\} \tag{69}$$

where D_w denotes an operator of the weak derivative (Maz'ja, 1985).

A.2 Theorems

Theorem A.1 (L_p Complete). *$\mathsf{L}_\mathsf{p}(\Omega)$, $1 \le p \le \infty$ is a Banach space.*

Theorem A.2 (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 $x \in \mathcal{H}$.*

Theorem A.3 (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*

$$H(\delta; \mathscr{F}) < A\delta^{-2\zeta}. \tag{70}$$

Theorem A.4 (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*

$$\mathbf{A}_{n\times n} = \mathbf{U}_{n\times n}\mathbf{W}_{n\times n}\mathbf{U}^\top_{n\times n}, \tag{71}$$

$$\mathbf{I}_{n\times n} = \mathbf{U}^\top_{n\times n}\mathbf{U}_{n\times n} = \mathbf{U}_{n\times n}\mathbf{U}^\top_{n\times n}. \tag{72}$$

Theorem A.5 (Cauchy-Schwartz Inequality). *If* $f \in \mathsf{L}_2(\Omega)$ *and* $f \in \mathsf{L}_2(\Omega)$, *then* $fg \in \mathsf{L}_1(\Omega)$ *and*

$$\int_\Omega |f(\mathbb{x})g(\mathbb{x})|\, d\mathbb{x} \leq \|f\|_{\mathsf{L}_2(\Omega)} \|g\|_{\mathsf{L}_2(\Omega)}. \tag{73}$$

Lemma A.6. *Suppose* $(f_n)_{n=1}^\infty$ *are non-negative Lipschitz functions on interval* $[a, b]$ *with a constant* $T > 0$ *for all* $n \in \mathbb{N}$. *If*

$$f_n \xrightarrow[n\to\infty]{\mathsf{L}_1} 0 \tag{74}$$

then

$$\|f_n\|_{\infty,[a,b]} := \sup_{x\in[a,b]} |f_n(x)| \xrightarrow[n\to\infty]{} 0. \tag{75}$$

B Proofs

Correctness of Definition 1.1. Let $f, g \in \mathcal{C}^m(\Omega) \cap \mathsf{L_p}(\Omega)$, the triangle inequality for the p-norms on $\mathsf{L_p}(\Omega)$ and $l_p\left(\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}|_\infty \leq m\}\right)$ implies

$$\begin{aligned}\|f+g\|_{p,Sob,m} &= \left\{\sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} f + D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)}\right\}^{1/p} \leq \left\{\sum_{|\boldsymbol{\alpha}|_\infty \leq m} \left[\|D^{\boldsymbol{\alpha}} f\|^p_{\mathsf{L_p}(\Omega)} + \|D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)}\right]\right\}^{1/p} \\ &\leq \left\{\sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} f\|^p_{\mathsf{L_p}(\Omega)}\right\}^{1/p} + \left\{\sum_{|\boldsymbol{\alpha}|_\infty \leq m} \|D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)}\right\}^{1/p} = \|f\|_{p,Sob,m} + \|g\|_{p,Sob,m}\,.\end{aligned} \tag{76}$$

□

Proof of Theorem 1.1. 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. □

Proof of Theorem 1.2. 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

$$\langle g, h\rangle_{Sob,m} = \sum_{k=0}^{m} \int_{\mathcal{Q}^1} g^{(k)}(x) h^{(k)}(x) \mathrm{d}x, \tag{77}$$

so all we need to do is to construct a representor

$$\psi_a \in \mathcal{H}^m[0,1] \text{ s.t. } \langle \psi_a, f\rangle_{Sob,m} = f(a) \tag{78}$$

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 1.1), hence assume that $f \in \mathcal{C}^{2m}$. This representor will be of the form:

$$\psi_a(x) = \begin{cases} L_a(x) & 0 \leq x \leq a, \\ R_a(x) & a \leq x \leq 1, \end{cases} \tag{79}$$

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 $L_a^{(k)}(a) = R_a^{(k)}(a)$, $0 \leq k \leq m-1$. We get:

$$f(a) = \langle \psi_a, f\rangle_{Sob,m} = \int_0^a \sum_{k=0}^{m} L_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x + \int_a^1 \sum_{k=0}^{m} R_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x. \tag{80}$$

Integrating by parts, we have:

$$\begin{aligned}\sum_{k=0}^{m}\int_0^a L_a^{(k)}(x)f^{(k)}(x)\mathrm{d}x &= \sum_{k=0}^{m}\left\{\sum_{j=0}^{k-1}(-1)^j\,L_a^{(k+j)}(x)f^{(k-j-1)}(x)\Big|_0^a + (-1)^k\int_0^a L_a^{(2k)}(x)f(x)\mathrm{d}x\right\}\\ &= \sum_{k=0}^{m}\sum_{j=0}^{k-1}(-1)^j\,L_a^{(k+j)}(x)f^{(k-j-1)}(x)\Big|_0^a + \int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x.\end{aligned} \tag{81}$$

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

$$\begin{aligned}\sum_{k=0}^{m}\int_0^a L_a^{(k)}(x)f^{(k)}(x)\mathrm{d}x &= \sum_{k=1}^{m}\sum_{i=0}^{k-1}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(x)f^{(i)}(x)\Big|_0^a + \int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x\\ &= \sum_{i=0}^{m-1}\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(x)f^{(i)}(x)\Big|_0^a + \int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x\\ &= \sum_{i=0}^{m-1}f^{(i)}(a)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(a)\right\}\\ &- \sum_{i=0}^{m-1}f^{(i)}(0)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,L_a^{(2k-i-1)}(0)\right\}\\ &+ \int_0^a\left\{\sum_{k=0}^{m}(-1)^k\,L_a^{(2k)}(x)\right\}f(x)\mathrm{d}x.\end{aligned} \tag{82}$$

Similarly:

$$\begin{aligned}\sum_{k=0}^{m}\int_a^1 R_a^{(k)}(x)f^{(k)}(x)\mathrm{d}x &= \sum_{i=0}^{m-1}f^{(i)}(1)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,R_a^{(2k-i-1)}(1)\right\}\\ &- \sum_{i=0}^{m-1}f^{(i)}(a)\left\{\sum_{k=i+1}^{m}(-1)^{k-i-1}\,R_a^{(2k-i-1)}(a)\right\}\\ &+ \int_a^1\left\{\sum_{k=0}^{m}(-1)^k\,R_a^{(2k)}(x)\right\}f(x)\mathrm{d}x.\end{aligned} \tag{83}$$

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

$$\sum_{k=0}^{m}(-1)^k\,\varphi^{(2k)}(x) = 0. \tag{84}$$

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[1] of $f^{(i)}(0)$, $f^{(i)}(1)$ and $f^{(i)}(a)$:

$$r_a \in \mathcal{H}^m[0,1] \quad \Rightarrow \quad L_a^{(i)}(a) = R_a^{(i)}(a) \quad \ldots \quad 0 \le i \le m-1, \tag{85}$$

$$f^{(i)}(0) \bowtie 0 \quad \Rightarrow \quad \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-i-1)}(0) = 0 \quad \ldots \quad 0 \le i \le m-1, \tag{86}$$

$$f^{(i)}(1) \bowtie 0 \quad \Rightarrow \quad \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-i-1)}(1) = 0 \quad \ldots \quad 0 \le i \le m-1, \tag{87}$$

$$f^{(i)}(a) \bowtie 0 \quad \Rightarrow \quad \sum_{k=i+1}^{m} (-1)^{k-i-1} \left\{ L_a^{(2k-i-1)}(a) - R_a^{(2k-i-1)}(a) \right\} = 0 \quad \ldots \quad 1 \le i \le m-1, \tag{88}$$

$$f(a) \bowtie 1 \quad \Rightarrow \quad \sum_{k=1}^{m} (-1)^{k-1} \left\{ L_a^{(2k-1)}(a) - R_a^{(2k-1)}(a) \right\} = 1; \tag{89}$$

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

$$P_m(\lambda) = \sum_{k=0}^{m} (-1)^k \lambda^{2k}. \tag{90}$$

Hence it follows

$$(1+\lambda^2) P_m(\lambda) = 1 + (-1)^m \lambda^{2m+2}, \quad \lambda \neq \pm i. \tag{91}$$

Solving the last equation (91), we get characteristic roots

$$\lambda_k = e^{i\theta_k}, \tag{92}$$

where

$$\theta_k \in \begin{cases} \frac{(2k+1)\pi}{2m+2} & m \text{ even}, \quad k \in \{0,1,\ldots,2m+1\} \setminus \left\{ \frac{m}{2}, \frac{3m+2}{2} \right\}, \\ \\ \frac{k\pi}{m+1} & m \text{ odd}, \quad k \in \{0,1,\ldots,2m+1\} \setminus \left\{ \frac{m+1}{2}, \frac{3m+3}{2} \right\}. \end{cases} \tag{93}$$

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

$$\varphi_k(x) = \exp\left\{ \big(\Re(\lambda_k)\big) x \right\} \cos\left[\big(\Im(\lambda_k)\big) x \right], \ k \in \{0,1,\ldots,m\} \setminus \left\{ \frac{m}{2} \right\}; \tag{94a}$$

$$\varphi_{m+1+k}(x) = \exp\left\{ \big(\Re(\lambda_k)\big) x \right\} \sin\left[\big(\Im(\lambda_k)\big) x \right], \ k \in \{0,1,\ldots,m\} \setminus \left\{ \frac{m}{2} \right\}. \tag{94b}$$

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

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

$$\varphi_0(x)=\exp\{x\}\,; \tag{95a}$$

$$\varphi_k(x)=\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big],\ k\in\{1,2,\ldots,m\}\setminus\left\{\frac{m+1}{2}\right\}; \tag{95b}$$

$$\varphi_{m+1}(x)=\exp\{-x\}\,; \tag{95c}$$

$$\varphi_{m+1+k}(x)=\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big],\ k\in\{1,2,\ldots,m\}\setminus\left\{\frac{m+1}{2}\right\}. \tag{95d}$$

These vectors generate a subspace of $\mathcal{C}^m\,[0,1]$ that is the space of solutions of the differential equation (84). In this case, the general solution is given by the linear combination:

$$L_a(x)\quad\begin{cases} =\displaystyle\sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_k(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\displaystyle\sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_{m+1+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big],\quad m\text{ even}; \\ =\gamma_0(a)\exp\{x\}+\displaystyle\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_k(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\gamma_{m+1}(a)\exp\{-x\}+\displaystyle\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_{m+1+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big],\quad m\text{ odd}; \end{cases} \tag{96}$$

$$R_a(x)\quad\begin{cases} =\displaystyle\sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_{2m+2+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\displaystyle\sum_{\substack{k=0\\k\neq\frac{m}{2}}}^{m}\gamma_{3m+3+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big],\quad m\text{ even}; \\ =\gamma_{2m+2}(a)\exp\{x\}+\displaystyle\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_{2m+2+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\cos\Big[\big(\Im(\lambda_k)\big)x\Big] \\ +\gamma_{3m+3}(a)\exp\{-x\}+\displaystyle\sum_{\substack{k=1\\k\neq\frac{m+1}{2}}}^{m}\gamma_{3m+3+k}(a)\exp\Big\{\big(\Re(\lambda_k)\big)x\Big\}\sin\Big[\big(\Im(\lambda_k)\big)x\Big],\quad m\text{ odd}; \end{cases} \tag{97}$$

where the coefficients $\gamma_k(a)$ are arbitrary constants that satisfy the boundary conditions (85)–(89). It can be easily seen that we have obtained $4(m+1)-4=4m$ coefficients $\gamma_k(a)$, because the first index of $\gamma_k(a)$ 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 (85)–(89) 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 coefficients $\gamma_k(a)$.

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

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

$$\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) = \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i) \quad \text{for all } \mathbb{x} \in \mathcal{Q}^q, \tag{98}$$

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

$$\begin{aligned}
\langle \psi_{\mathbb{a}}^{\mathbb{w}}, f\rangle_{Sob,m} &= \left\langle \prod_{i=1}^{q} \psi_{a_i}^{w_i}, f \right\rangle_{Sob,m} = \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_{\mathcal{Q}^q} \frac{\partial^{\alpha_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{\alpha_q}} D^{\boldsymbol{\alpha}} f(\mathbb{x}) \mathrm{d}\mathbb{x} \\
&= \sum_{i_1,\dots,i_q=0,\dots,m} \int_{\mathcal{Q}^q} \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} \frac{\partial^{i_1,\dots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\dots\partial x_q^{i_q}} \mathrm{d}\mathbb{x} \\
&= \sum_{i_1=0}^{m} \int_0^1 \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \left[\cdots \left[\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} \cdot \frac{\partial^{i_1,\dots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\dots\partial x_q^{i_q}} \mathrm{d}x_q \right] \cdots \right] \mathrm{d}x_1.
\end{aligned} \tag{99}$$

According to Definition 1.3 and notation in (7) we can rewrite the centermost bracket

$$\begin{aligned}
\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} \cdot \frac{\partial^{i_1,\dots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\dots\partial x_q^{i_q}} \mathrm{d}x_q &= \left\langle \psi_{a_q}^{w_q}, D^{(i_1,\dots,i_{q-1})} f(x_1,\dots,x_{i-1},\cdot)\right\rangle_{Sob,m} \\
&= D^{(i_1,\dots,i_{q-1},w_q)} f(\mathbb{x}_{-q}, a_q).
\end{aligned} \tag{100}$$

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

Proof of Theorem 1.3. Existence and uniqueness of coefficients $\gamma_k(a)$ has already been proved in the proof of Theorem 1.2.

Let's define

$$\Lambda_{a,I}^{(l)} := \begin{cases} L_a^{(l)}(0), & \text{for } I = L; \\ R_a^{(l)}(1), & \text{for } I = R; \\ L_a^{(l)}(a) - R_a^{(a)}(a), & \text{for } I = D. \end{cases} \tag{101}$$

From (86)–(89) we easily see

$$\sum_{k=i+1}^{m} (-1)^{k-i-1} \Lambda_{a,I}^{(2k-i-1)} = 0, \quad 0 \le i \le m-1,\ I \in \{L, R, D\}, \quad [i, I] \ne [0, D]; \tag{102}$$

$$\sum_{k=1}^{m} (-1)^{k-1} \Lambda_{a,D}^{(2k-1)} = 1. \tag{103}$$

If $m = 1$ it directly follows from (102)–(103):

$$\Lambda_{a,I}^{(1)} = 0, \quad I \in \{L, R\}, \tag{104}$$

$$\Lambda_{a,D}^{(1)} = 1. \tag{105}$$

If $m = 2$ it also directly follows from (102)–(103):

$$\Lambda_{a,I}^{(2)} = 0, \quad \forall I, \tag{106}$$

$$\Lambda_{a,I}^{(1)} - \Lambda_{a,I}^{(3)} = 0, \quad I \in \{L, R\}, \tag{107}$$

$$\Lambda_{a,D}^{(1)} - \Lambda_{a,D}^{(3)} = 1. \tag{108}$$

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

$$\Lambda_{a,I}^{(m-j)} + (-1)^j \Lambda_{a,I}^{(m+j)} = 0, \quad j = 0, \ldots, m-2,\ \forall I, \tag{109}$$

$$\Lambda_{a,I}^{(1)} + (-1)^{m-1} \Lambda_{a,I}^{(2m-1)} = 0, \quad I \in \{L, R\}, \tag{110}$$

$$\Lambda_{a,D}^{(1)} + (-1)^{m-1} \Lambda_{a,D}^{(2m-1)} = 1, \tag{111}$$

where $j := m - i - 1$.

For $j = 0$ is $i = m - 1$ and from (102)–(103) we have straightforwardly

$$\Lambda_{a,I}^{(m)} = 0, \quad \forall I, \tag{112}$$

which is correct according to (109). Consider $j = 1$ and thus $i = m - 2$. In the same way we get a correspondent result to (109)

$$\Lambda_{a,I}^{(m-1)} - \Lambda_{a,I}^{(m+1)} = 0, \quad \forall I. \tag{113}$$

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

$$\Lambda_{a,I}^{(m-2)} - \Lambda_{a,I}^{(m)} + \Lambda_{a,I}^{(m+2)} = 0, \quad \forall I, \tag{114}$$

so it forces us to apply (112) and for $j = 3$ and thus $i = m - 4$ we have

$$\Lambda_{a,I}^{(m-3)} - \Lambda_{a,I}^{(m-1)} + \Lambda_{a,I}^{(m+1)} - \Lambda_{a,I}^{(m+3)} = 0, \quad \forall I, \tag{115}$$

so we can apply (113). 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 (110) in case $I \in \{L, R\}$, eventually (111) in case $I = D$ instead of (109).
To finish this proof all we need to do is not to forget to think of (85). From (85) we obtain

$$\Lambda_{a,D}^{(j)} = 0, \quad j \in \{0, \ldots, m-1\}. \tag{116}$$

According to (109) for $I = D$ and (111) we further see:

$$\Lambda_{a,D}^{(j)} = 0, \quad j \in \{m+1, \ldots, 2m-2\}; \tag{117}$$

$$\Lambda_{a,D}^{(2m-1)} = (-1)^{m-1}. \tag{118}$$

Alltogether we have obtained these $4m$ linear equations

$$\Lambda_{a,L}^{(m-j)} + (-1)^j \Lambda_{a,L}^{(m+j)} = 0, \quad j = 0, \ldots, m-1, \tag{119}$$

$$\Lambda_{a,R}^{(m-j)} + (-1)^j \Lambda_{a,R}^{(m+j)} = 0, \quad j = 0, \ldots, m-1, \tag{120}$$

$$\Lambda_{a,D}^{(j)} = 0, \quad j = 0, \ldots, 2m-2, \tag{121}$$

$$\Lambda_{a,D}^{(2m-1)} = (-1)^{m-1}, \tag{122}$$

which after rewriting them using (101), (96)–(97) and (94a)–(95d) bring us to a close. □

Proof of Theorem 1.4. See Yatchew and Bos (1997). □

Proof of Theorem 2.1. Let $M = span\left\{\psi_{\mathbb{x}_i} : i = 1, \ldots, n\right\}$ and its orthogonal complement

$$M^{\perp} = \left\{ h \in \mathcal{H}^m : \langle \psi_{\mathbb{x}_i}, h \rangle_{Sob,m} = 0,\, i = 1, \ldots, n \right\}. \tag{123}$$

Representors exist by Theorem 1.2 and we can write the Pseudo-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

$$f = \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j} + h, \quad h \in M^{\perp}. \tag{124}$$

Then

$$\begin{aligned}
&[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})]^{\top} \boldsymbol{\Sigma}^{-1} [\mathbb{Y} - \boldsymbol{f}(\mathbb{x})] + \chi \|f\|^2_{Sob,m} \\
&= \left[\mathbb{Y}_{\bullet} - \left\langle \psi_{\mathbb{x}_{\bullet}}, \sum_{j=1}^{n} c_j \psi_{x_j} + h \right\rangle_{Sob,m} \right]^{\top} \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y}_{\bullet} - \left\langle \psi_{\mathbb{x}_{\bullet}}, \sum_{j=1}^{n} c_j \psi_{x_j} + h \right\rangle_{Sob,m} \right] + \chi \left\| \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j} + h \right\|^2_{Sob,m} \\
&= \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} \left\langle \psi_{\mathbb{x}_{\bullet}}, c_j \psi_{x_j} \right\rangle_{Sob,m} \right]^{\top} \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} \left\langle \psi_{\mathbb{x}_{\bullet}}, c_j \psi_{x_j} \right\rangle_{Sob,m} \right] + \chi \left\| \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j} \right\|^2_{Sob,m} + \chi \|h\|^2_{Sob,m} \\
&= \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} c_j \left\langle \psi_{\mathbb{x}_{\bullet}}, \psi_{x_j} \right\rangle_{Sob,m} \right]^{\top} \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} c_j \left\langle \psi_{\mathbb{x}_{\bullet}}, \psi_{x_j} \right\rangle_{Sob,m} \right] \\
&+ \chi \left\langle \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j}, \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} + \chi \|h\|^2_{Sob,m} \\
&= \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} \Psi_{\bullet,j} c_j \right]^{\top} \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y}_{\bullet} - \sum_{j=1}^{n} \Psi_{\bullet,j} c_j \right] + \chi \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \left\langle \psi_{\mathbb{x}_j}, \psi_{\mathbb{x}_k} \right\rangle_{Sob,m} c_k + \chi \|h\|^2_{Sob,m} \\
&= [\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}]^{\top} \boldsymbol{\Sigma}^{-1} [\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}] + \chi \mathbb{c}^{\top} \boldsymbol{\Psi} \mathbb{c} + \chi \|h\|^2_{Sob,m}
\end{aligned} \tag{125}$$

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

$$\langle \psi_{\mathbb{x}_{\bullet}}, g \rangle_{Sob,m} = \left(\langle \psi_{x_1}, g \rangle_{Sob,m}, \ldots, \langle \psi_{x_n}, g \rangle_{Sob,m} \right)^{\top}. \tag{126}$$

Hence, there exists a function f^* minimizing the infinite dimensional optimizing problem that is a linear combination of the representors. We note also that $\|f^*\|^2_{Sob,m} = \mathbb{c}^{\top} \boldsymbol{\Psi} \mathbb{c}$.

Uniqueness is clear, since $\psi_{\mathbb{x}_i}$ are the base elements of M, and adding a function that is orthogonal to the spaces spanned by the representors will increase the norm. □

Proof of Corollary 2.2. Trivial. It can be directly seen from the definition of form (79) of the representor and from (24). □

Proof of Lemma 2.3. Trivial. The representor matrix is symmetric by Definition 2.3, because

$$\Psi_{i,j} = \left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} = \left\langle \psi_{\mathbb{x}_j}, \psi_{\mathbb{x}_i} \right\rangle_{Sob,m} = \Psi_{j,i}, \tag{127}$$

i.e. $\mathbf{\Psi} = \mathbf{\Psi}^\top$. □

Proof of Theorem 2.4. We proceed this proof only for one dimensional variable x. Extention into the multi-variable case is clearly simple (see Remark 2.1). For an arbitrary $\mathbb{c} \in \mathbb{R}^n$ we obtain

$$\begin{aligned} \mathbb{c}^\top \mathbf{\Psi} \mathbb{c} &= \sum_i c_i \sum_j \Psi_{ij} c_j = \sum_i \sum_j c_i \left\langle \psi_{x_i}, \psi_{x_j} \right\rangle_{Sob,m} c_j = \sum_i \sum_j \left\langle c_i \psi_{x_i}, c_j \psi_{x_j} \right\rangle_{Sob,m} \\ &= \left\langle \sum_i c_i \psi_{x_i}, \sum_j c_j \psi_{x_j} \right\rangle_{Sob,m} = \left\| \sum_i c_i \psi_{x_i} \right\|^2_{Sob,m} \geq 0. \end{aligned} \tag{128}$$

Hence $\mathbb{c}^\top \mathbf{\Psi} \mathbb{c} = 0$ iff $\sum_i c_i \psi_{x_i} = 0$ a.e. According to (96)–(97), (94a)–(95d) and (16) we have[2]

$$\psi_{x_i}(x) = \boldsymbol{\gamma}(x_i)^\top \boldsymbol{\varphi}(x) = (-1)^{m-1} \left[\{\mathbf{\Gamma}(x_i)\}^{-1} \right]^\top_{\bullet,4m} \boldsymbol{\varphi}(x) \tag{129}$$

where $\boldsymbol{\varphi}(x)$ is vector which elements are linear independent base elements of space of the differential equation's (84) solutions, i.e. $\varphi_k(x)$ (see (94a)–(95d)). Thus linear independence of $\varphi_k(x)$ it follows that

$$\begin{aligned} \sum_i c_i \psi_{x_i} &= (-1)^{m-1} \sum_i c_i \left[\{\mathbf{\Gamma}(x_i)\}^{-1} \right]^\top_{\bullet,4m} \boldsymbol{\varphi} \\ &= (-1)^{m-1} \sum_i \sum_k c_i \left[\{\mathbf{\Gamma}(x_i)\}^{-1} \right]_{4m,k} \varphi_k = 0 \quad \text{a.e.} \end{aligned} \tag{130}$$

$$\Updownarrow$$

$$\varphi_k = 0 \quad \text{a.e.} \quad k \in \{0, 1, \ldots, 2m+1\} \setminus \begin{cases} \left\{ \frac{m}{2}, \frac{3m+2}{2} \right\} & m \text{ even}, \\ \left\{ \frac{m+1}{2}, \frac{3m+3}{2} \right\} & m \text{ odd}; \end{cases} \tag{131}$$

$$\Downarrow$$

$$\psi_{x_i} = 0 \quad \text{a.e.} \quad i = 1, \ldots, n. \tag{132}$$

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

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

Proof of Theorem 2.5. According to the Theorem 2.1, we want to find the minimum of function

$$\mathcal{L}(\mathbb{c}) := \frac{1}{n}\left[\mathbb{Y} - \mathbf{\Psi}\mathbb{c}\right]^\top \mathbf{\Sigma}^{-1}\left[\mathbb{Y} - \mathbf{\Psi}\mathbb{c}\right] + \chi\mathbb{c}^\top\mathbf{\Psi}\mathbb{c}. \tag{133}$$

Therefore the first partial derivatives of $\mathcal{L}(\mathbb{c})$ have to be equal zero at the minimum point $\widehat{\mathbb{c}}$:

$$\frac{\partial}{\partial c_i}\mathcal{L}(\mathbb{c}) \stackrel{!}{=} 0, \quad i = 1, \ldots, n. \tag{134}$$

If $\mathbf{\Sigma}^{-1} =: (\phi_{ij})_{i,j=1}^{n,n}$, we have

$$\begin{aligned} n\mathcal{L}(\mathbb{c}) &= \mathbb{Y}^\top\mathbf{\Sigma}^{-1}\mathbb{Y} - 2\mathbb{Y}^\top\mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbb{c} + \mathbb{c}^\top\mathbf{\Psi}\mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbb{c} + n\chi\mathbb{c}^\top\mathbf{\Psi}\mathbb{c} \\ &= \sum_{r=1}^n\sum_{s=1}^n Y_r\phi_{rs}Y_s - 2\sum_{r=1}^n\sum_{s=1}^n\sum_{t=1}^n Y_r\phi_{rs}\Psi_{st}c_t + \sum_{r=1}^n\sum_{s=1}^n\sum_{t=1}^n\sum_{u=1}^n c_r\Psi_{rs}\phi_{st}\Phi_{tu}c_u + n\chi\sum_{r=1}^n\sum_{s=1}^n c_r\Psi_{rs}c_s \end{aligned} \tag{135}$$

hence

$$\begin{aligned} 0 &\stackrel{!}{=} -2\sum_{r=1}^n\sum_{s=1}^n Y_r\phi_{rs}\Psi_{si} + 2\sum_{\substack{r=1\\ r\neq i}}^n\sum_{s=1}^n\sum_{t=1}^n c_r\Psi_{rs}\phi_{st}\Phi_{ti} + 2\sum_{r=1}^n\sum_{s=1}^n c_i\Psi_{is}\phi_{st}\Phi_{ti} + 2n\chi\sum_{\substack{r=1\\ r\neq i}}^n c_r\Psi_{ri} + 2n\chi c_i\Psi_{ii} \\ &= -2\mathbb{Y}^\top\mathbf{\Sigma}^{-1}\mathbf{\Psi}_{\bullet,i} + 2\mathbb{c}^\top\mathbf{\Psi}\mathbf{\Sigma}^{-1}\mathbf{\Psi}_{\bullet,i} + 2n\chi\mathbb{c}^\top\mathbf{\Psi}_{\bullet,i}, \quad i = 1, \ldots, n. \end{aligned} \tag{136}$$

Then we obtain our system of “normal” equations

$$\mathbb{c}^\top\left(\mathbf{\Psi}\mathbf{\Sigma}^{-1}\mathbf{\Psi}_{\bullet,i} + n\chi\mathbf{\Psi}_{\bullet,i}\right) = \mathbb{Y}^\top\mathbf{\Sigma}^{-1}\mathbf{\Psi}_{\bullet,i}, \quad i = 1, \ldots, n. \tag{137}$$

□

Proof of Theorem 2.6. The solution of (32) always exists and is unique according to the proof of Theorem 2.1. From the same proof of Theorem 2.1 follows that finding f^*—optimizing (32)—is the same as searching optimal $\mathbb{c}^*$ such that

$$\mathbb{c}^* = \arg\min_{\mathbb{c}\in\mathbb{R}^n}\frac{1}{n}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right]^\top\mathbf{\Sigma}^{-1}\left[\mathbb{y} - \mathbf{\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}^\top\mathbf{\Psi}\mathbb{c} \leq L \tag{138}$$

and again from the proof of Theorem 2.1 the existence and the uniqueness of $\mathbb{c}^*$ is guaranteed. Let's fix L.

If $\mathbb{c}^{*\top}\boldsymbol{\Psi}\mathbb{c}^* = L$, we can simply apply Lagrange Multiplier Theorem on our bond condition $\mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} = L$ using the Lagrange function

$$\mathcal{J}(\mathbb{c},\lambda) = \frac{1}{n}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right] + \lambda\left(\mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} - L\right) \tag{139}$$

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 $\chi > 0$.

If $\mathbb{c}^{*\top}\boldsymbol{\Psi}\mathbb{c}^* < L$, we just set $\chi = 0$ and we are done. □

Proof of Theorem 2.7. Trivial. Just apply Lagrange Multiplier Theorem. □

Proof of Theorem 2.8. Let's have fixed $\chi > 0$. Hence we have obtained unique $\widehat{f}$ and also $\widehat{\mathbb{c}}$ according to Theorem 2.1. Theorems 2.1 and 2.7 say that there esists a unique $L > 0$ such that $\widehat{\mathbb{c}}$ is also a unique solution of optimizing problem

$$\widehat{\mathbb{c}} = \arg\min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} = L. \tag{140}$$

Let's define

$$\widetilde{\mathbf{f}}(\mathbb{x}) := \boldsymbol{\Xi}\mathbf{f}(\mathbb{x}), \tag{141}$$

$$\widetilde{\mathbb{Y}} := \boldsymbol{\Xi}\mathbb{Y}, \tag{142}$$

$$\widehat{\widetilde{\mathbb{c}}} := \arg\min_{\widetilde{\mathbb{c}}\in\mathbb{R}^n} \frac{1}{n}\left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t.} \quad \widetilde{\mathbb{c}}^\top\boldsymbol{\Psi}\boldsymbol{\Xi}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\Psi}\widetilde{\mathbb{c}} \le L. \tag{143}$$

We can easily find out that

$$\widehat{\widetilde{\mathbb{c}}} = \boldsymbol{\Psi}^{-1}\boldsymbol{\Xi}\boldsymbol{\Psi}\widehat{\mathbb{c}} \tag{144}$$

and hence

$$\widehat{\widetilde{\mathbf{f}}}(\mathbb{x}) = \boldsymbol{\Xi}\widehat{\widetilde{\mathbb{c}}}. \tag{145}$$

Finally, there must exists $\widetilde{L} > 0$ such that

$$\widehat{\widetilde{\mathbb{c}}} = \arg\min_{\widetilde{\mathbb{c}}\in\mathbb{R}^n} \frac{1}{n}\left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t. } \widetilde{\mathbb{c}}^\top\boldsymbol{\Psi}\widetilde{\mathbb{c}} = \widetilde{L} \tag{146}$$

and hence this exactly same $\widehat{\widetilde{\mathbb{c}}}$ have to be a unique solution of optimizing problem

$$\widehat{\widetilde{\mathbb{c}}} = \arg\min_{\widetilde{\mathbb{c}} \in \mathbb{R}^n} \frac{1}{n} \left[\widetilde{\mathbb{Y}} - \mathbf{\Psi}\widetilde{\mathbb{c}}\right]^\top \mathbf{\Sigma}^{-1} \left[\widetilde{\mathbb{Y}} - \mathbf{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t. } \widetilde{\mathbb{c}}^\top \mathbf{\Psi}\widetilde{\mathbb{c}} \le \widetilde{L} \tag{147}$$

because $\mathbf{\Psi}$ is positive definite matrix ($\widetilde{\mathbb{c}}^\top \mathbf{\Psi}\widetilde{\mathbb{c}}$ is the volume of n-dimensional ellipsoid).

Now we think of model

$$\widetilde{Y}_i = \widetilde{f}(\mathbb{x}_i) + \widetilde{\varepsilon}_i, \quad \widetilde{\varepsilon}_i \sim i.i.d., \quad i = 1, \ldots, n \tag{148}$$

with least-squares estimator $\widehat{\widetilde{f}}$. Using Kolmogorov-Tihomirov Theorem A.3 it can be shown that exists $A > 0$ such that for $\delta > 0$, we have $\log N(\delta; \mathscr{F}) < A\delta^{-q/m}$. Consequently applying Lemma 3.5 from de Geer (1990), we obtain that there exist positive constants C_0, K_0 such that for all $K > K_0$

$$\mathsf{P}\left[\sup_{\|g\|^2_{Sob,m} \le \widetilde{L}} \frac{\sqrt{n}\left|-\frac{2}{n}\sum_{i=1}^n \widetilde{\varepsilon}_i \left(\widetilde{f}(x_i) - g(x_i)\right)\right|}{\left(\frac{1}{n}\sum_{i=1}^n \left(\widetilde{f}(x_i) - g(x_i)\right)^2\right)^{\frac{1}{2} - \frac{q}{4m}}} \ge KA^{1/2}\right] \le \exp\left\{-C_0K^2\right\}. \tag{149}$$

Since $\widetilde{f} \in \widetilde{\mathscr{F}} = \left\{g \in \mathcal{H}^m(\mathcal{Q}^q) : \|g\|^2_{Sob,m} \le \widetilde{L}\right\}$ and $\widehat{\widetilde{f}}$ minimizes the sum of squared residuals over $g \in \widetilde{\mathscr{F}}$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \left[\widetilde{Y}_i - \widehat{\widetilde{f}}(x_i)\right]^2 &\le \frac{1}{n}\sum_{i=1}^n \left[\widetilde{Y}_i - g(x_i)\right]^2, \quad g \in \widetilde{\mathscr{F}} \qquad (150) \\ \frac{1}{n}\sum_{i=1}^n \left[\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right) + \widetilde{\varepsilon}_i\right]^2 &\le \frac{1}{n}\sum_{i=1}^n \left[\left(\widetilde{f}(x_i) - g(x_i)\right) + \widetilde{\varepsilon}_i\right]^2, \quad g \in \widetilde{\mathscr{F}} \\ &\Downarrow \quad \text{realize that } \widetilde{f} \in \widetilde{\mathscr{F}} \\ \frac{1}{n}\sum_{i=1}^n \left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 &\le -\frac{2}{n}\sum_{i=1}^n \widetilde{\varepsilon}_i \left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right). \qquad (151) \end{aligned}$$

Now combine (149) and (151) to obtain the result that $\forall K > K_0$

$$\mathsf{P}\left[\frac{1}{n}\sum_{i=1}^n \left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 \ge \left(\frac{K^2A}{n}\right)^{\frac{2m}{2m+q}}\right] \le \exp\left\{-C_0K^2\right\}. \tag{152}$$

Thus

$$\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1} \left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] = \frac{1}{n}\sum_{i=1}^n \left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 = \mathcal{O}_{\mathsf{P}}\left(n^{-\frac{2m}{2m+q}}\right), \quad n \to \infty. \tag{153}$$

□

Proof of Lemma A.6. Reduction ad proof. Suppose that

$$\exists \epsilon > 0 \quad \forall n_0 \in \mathbb{N} \quad \exists n \geq n_0 \quad \exists x \in [a, b] \quad f_n(x) \geq \epsilon. \tag{154}$$

Then according to Lipschitz property of each $f_n \geq 0$ we have for fixed ϵ, n_0, n and $x \in [a, b]$ (drawing a picture could be helpful)

$$\begin{aligned}
\|f_n\|_{\mathsf{L}_1[a,b]} &= \int_a^b f_n(t)\mathrm{d}t \\
&\geq \min\left\{\frac{f_n(x)}{2}(x-a) + \frac{f_n(x)}{2}(b-x), \frac{f_n(x)}{2}(x-a) + \frac{f_n(x)}{2}\frac{f_n(x)}{T},\right. \\
&\quad \left.\frac{f_n(x)}{2}\frac{f_n(x)}{T} + \frac{f_n(x)}{2}(b-x), \frac{f_n(x)}{2}\frac{f_n(x)}{T} + \frac{f_n(x)}{2}\frac{f_n(x)}{T}\right\} \\
&\geq \min\left\{\frac{\epsilon}{2}(b-a), \frac{\epsilon}{2}(x-a) + \frac{\epsilon^2}{2T}, \frac{\epsilon^2}{2T} + \frac{\epsilon}{2}(b-x), \frac{\epsilon^2}{T}\right\} =: K > 0.
\end{aligned} \tag{155}$$

But K is a positive constant which does not depend on n. Hence this is an absurdum because it should hold (according to the assumption of this lemma)

$$\forall \delta > 0 \quad \exists n_1 \in \mathbb{N} \quad \forall n \geq n_1 \quad \|f_n\|_{\mathsf{L}_1[a,b]} < \delta. \tag{156}$$

□

Proof of Theorem 3.1. We divide the proof into two steps.

i) $s = 0$

The covariance matrix $\mathbf{\Sigma}$ is symmetric and positive definite with equibounded eigenvalues for all n. Hence it can be decomposed using Schur decomposition (Theorem A.4)

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Upsilon}\mathbf{\Gamma}^\top \tag{157}$$

where $\mathbf{\Gamma}$ is orthogonal, $\mathbf{\Upsilon}$ is diagonal (with eigenvalues on this diagonal) such that

$$0 < \Upsilon_{ii} \leq \vartheta \qquad i = 1, \ldots, n, \quad \forall n. \tag{158}$$

Hence

$$\mathbf{\Sigma}^{-1} = \mathbf{\Gamma} diag\left\{\Upsilon_1^{-1}, \ldots, \Upsilon_n^{-1}\right\}\mathbf{\Gamma}^\top. \tag{159}$$

Then

$$\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right] \geq \frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Gamma}\vartheta^{-1}\mathbf{I}\boldsymbol{\Gamma}^\top\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right] = \frac{1}{n\vartheta}\sum_{i=1}^{n}\left[\hat{f}(x_i)-f(x_i)\right]^2 \tag{160}$$

Let's define $h_n := \left|\hat{f}-f\right|$. We know $\left\|\hat{f}\right\|_{Sob,m}^2 \leq L$ for all n and $\left\|f\right\|_{Sob,m}^2 \leq L$. For every function $t \in \mathcal{H}^m[a,b]$ with $\left\|t\right\|_{Sob,m}^2 \leq L$ holds

$$\|t'\|_{\mathsf{L}_2[a,b]} \leq \|t\|_{Sob,1} \leq \|t\|_{Sob,m} \leq \sqrt{L}. \tag{161}$$

Then t has equibounded derivative and hence there exists a Lipschitz constant $T>0$ such that

$$|t(\xi)-t(\zeta)| < T\left|\xi-\zeta\right|, \quad \xi,\zeta \in [a,b]. \tag{162}$$

We easily see

$$\begin{aligned}\frac{|h_n(\xi)-h_n(\zeta)|}{|\xi-\zeta|} &= \frac{\left|\left|\hat{f}(\xi)-f(\xi)\right|-\left|\hat{f}(\zeta)-f(\zeta)\right|\right|}{|\xi-\zeta|} \leq \frac{\left|\left[\hat{f}(\xi)-f(\xi)\right]-\left[\hat{f}(\zeta)-f(\zeta)\right]\right|}{|\xi-\zeta|}\\ &\leq \frac{\left|\hat{f}(\xi)-\hat{f}(\zeta)\right|+\left|f(\xi)-f(\zeta)\right|}{|\xi-\zeta|} < 2T, \quad \xi,\zeta\in[a,b].\end{aligned} \tag{163}$$

Since h_n is T-Lipschitz function for all n and

$$\|h_n\|_{\mathsf{L}_2[a,b]} = \left\|\hat{f}-f\right\|_{\mathsf{L}_2[a,b]} \leq \left\|\hat{f}-f\right\|_{Sob,1} \leq \left\|\hat{f}-f\right\|_{Sob,m} \leq \left\|\hat{f}\right\|_{Sob,m}+\left\|f\right\|_{Sob,m} \leq 2\sqrt{L}, \quad \forall n, \tag{164}$$

we obtain that h_n is equibounded for all n with a positive constant M such that

$$\|h_n\|_{\infty,[a,b]} \leq M > 0, \quad \forall n. \tag{165}$$

Hence h_n^2 is also a Lipschitz function for all n, because for $\xi,\zeta\in[a,b]$

$$\frac{\left|h_n^2(\xi)-h_n^2(\zeta)\right|}{|\xi-\zeta|} = \frac{|h_n(\xi)-h_n(\zeta)|}{|\xi-\zeta|}\left[h_n(\xi)+h_n(\zeta)\right] \leq T\times 2\left\|h_n\right\|_{\infty,[a,b]} = 2MT =: U > 0, \quad \forall n. \tag{166}$$

Since h_n^2 is U-Lipschitz function for all n and design points $(x_i)_{i=1}^n$ are equidistantly distributed on $[a,b]$, we can write (drawing a picture could be helpful)

$$\begin{aligned}\int_a^b h_n^2(u)\mathrm{d}u &\leq \sum_{i=1}^{n-1} \frac{x_{i+1}-x_i}{2}\left\{h_n^2(x_i) + \left[h_n^2(x_i) + U(x_{i+1}-x_i)\right]\right\} \leq \frac{1}{2n}\left[2\sum_{i=1}^{n-1} h_n^2(x_i) + U(b-a)\right] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} h_n^2(x_i) + \frac{U(b-a)}{2n}.\end{aligned} \tag{167}$$

According to Theorem 2.8

$$\forall \epsilon > 0 \quad \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] > \epsilon\right\} \xrightarrow[n\to\infty]{} 0, \tag{168}$$

so it means

$$\forall \epsilon > 0 \quad \forall \delta > 0 \quad \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0 \quad \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] > \epsilon\right\} < \delta. \tag{169}$$

Let's fix an arbitrary $\epsilon > 0$ and $\delta > 0$. Hence fix

$$n_0 := \left\lceil \frac{U}{\epsilon^2} \right\rceil \tag{170}$$

and for all $n \geq n_0$ we can write

$$\begin{aligned}\delta &> \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] > \frac{\epsilon^2(b-a)}{2\vartheta}\right\} && \text{by (169)} \\ &\geq \mathsf{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{f}(x_i) - f(x_i)\right]^2 > \frac{\epsilon^2(b-a)}{2}\right\} && \text{by (160)} \\ &\geq \mathsf{P}\left\{\|h_n\|^2_{\mathsf{L}_2[a,b]} > \underbrace{\frac{\epsilon^2(b-a)}{2} + \frac{U(b-a)}{2n}}_{\tilde{\epsilon}}\right\} && \text{by (167)} \\ &\geq \mathsf{P}\left\{\|h_n\|_{\mathsf{L}_1[a,b]} > \frac{\sqrt{\tilde{\epsilon}}}{\|1\|_{\mathsf{L}_2[a,b]}}\right\} && \text{by Cauchy-Schwarz } A.5 \\ &\geq \mathsf{P}\left\{\|h_n\|_{\mathsf{L}_1[a,b]} > \epsilon\right\} && \text{by (170).}\end{aligned} \tag{171}$$

Thus

$$\|h_n\|_{\mathsf{L}_1[a,b]} \xrightarrow[n\to\infty]{\mathsf{P}} 0. \tag{172}$$

According to Lemma A.6 and the fact that the almost sure convergence implies convergence in probability, we have

$$\sup_{x\in[a,b]}\left|\widehat{f}(x)-f(x)\right| \xrightarrow[n\to\infty]{\mathsf{P}} 0. \tag{173}$$

ii) $s \geq 1$

If $m = 2$, we are done. Let $g_n := \widehat{f} - f$. According to the assumptions of our model, $g_n \in \mathcal{H}^m[a,b]$. By Theorem 1.4 (Embedding), all the functions in the estimating set have derivatives up to order $m-1$ uniformly bounded in supnorm. Then all the g_n'' are also bounded in supnorm ($m \geq 3$) and that implies the uniform boundedness of g_n'':

$$\exists M > 0 \quad \forall n \in \mathbb{N} \quad \|g_n''\|_{\infty,[a,b]} < M. \tag{174}$$

Let's have fixed $M > 0$. For any fixed $\epsilon > 0$, define $\tilde{\epsilon} := M\epsilon$ and there exists $n_0 \in \mathbb{N}$, such that $\forall n \geq n_0$: $[c_n, d_n] \subset [a,b]$ and

$$g_n'(c_n) = g_n'(d_n) = \tilde{\epsilon} \quad \& \quad g_n'(\xi) > \tilde{\epsilon},\, \xi \in (c_n, d_n) \tag{175}$$

because g_n' is continuous on $[c_n, d_n]$ (drawing a picture is helpful). If there does not exist such $[c_n, d_n]$, we are done.

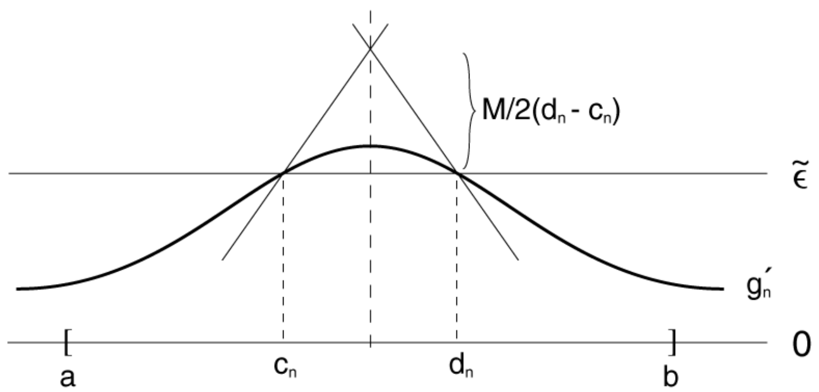

Figure 6: Uniform convergence of g_n'.

Otherwise there exists $n_1 \geq n_0$ such that $\forall n \geq n_1$ holds:

$$|\tilde{\epsilon}(d_n - c_n)| \leq \left|\int_{c_n}^{d_n} g_n'(\xi)\mathrm{d}\xi\right| = |g_n(d_n) - g_n(c_n)| \leq 2\epsilon^2 \tag{176}$$

because $g_n \xrightarrow{n\to\infty} 0$ uniformly in supnorm on interval $[a,b]$. Hence

$$|d_n - c_n| \leq \frac{2\epsilon}{M}. \tag{177}$$

Uniform boundedness of g_n'' implies Lipschitz property (see Figure 6):

$$|g_n'(x)| \leq \left|\tilde{\epsilon} + M\frac{d_n - c_n}{2}\right| \leq M\epsilon + M\frac{\epsilon}{M} \leq \epsilon(M+1). \tag{178}$$

We could continue in this way finitely times (formally we can proceed this by something like a finite induction). In fact, if $(m-1)$-th derivatives are uniformly bounded ($g_n \in \mathcal{H}^m[a,b]$), then this ensures that $\widehat{f}^{(s)}$ for $s \leq m-2$ converges in supnorm. Finally, we have to realize that convergence almost sure implies convergence in probability and each convergent sequence in probability has a convergent subsequence that converges almost sure. □

Proof of Theorem 3.2. The proof is very similar to the proof of Infinite to Finite Theorem 2.1. We use the same argumentation. Each $f, g \in \mathcal{H}^m$ can be written in form

$$f = \sum_{\{i\,|\,n_i \geq 1\}} c_i \psi_{x_i} + h_f, \quad h_f \in \{span\{\psi_{\mathbb{x}_i} : n_i \geq 1\}\}^{\perp}, \tag{179}$$

$$g = \sum_{\{j\,|\,m_j \geq 1\}} d_j \phi_{x_j} + h_g, \quad h_g \in \left\{span\left\{\phi_{\mathbb{x}_j} : m_j \geq 1\right\}\right\}^{\perp}. \tag{180}$$

For $1 \leq \iota \leq n$, we easily note that

$$\begin{aligned}
&\left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{f}(\mathbb{x}_\alpha) \\ \mathbf{g}(\mathbb{x}_\beta) \end{pmatrix}\right]_\iota \\
&= Y_\iota - \left\{\sum_{\{i\,|\,n_i \geq 1\}} \Delta_{\iota i} f(x_i) + \sum_{\{i\,|\,m_i \geq 1\}} \Theta_{\iota i} g(x_i)\right\} \\
&= Y_\iota - \sum_{\{i\,|\,n_i \geq 1\}} \Delta_{\iota i} \left\langle \psi_{x_i}, \sum_{\{j\,|\,n_j \geq 1\}} c_j \psi_{x_j} + h_f \right\rangle_{Sob,m} \\
&- \sum_{\{i\,|\,m_i \geq 1\}} \Theta_{\iota i} \left\langle \phi_{x_i}, \sum_{\{j\,|\,m_j \geq 1\}} d_j \phi_{x_j} + h_g \right\rangle_{Sob,m} \\
&= Y_\iota - \sum_{\{i\,|\,n_i \geq 1\}} \Delta_{\iota i} \sum_{\{j\,|\,n_j \geq 1\}} \Psi_{ij} c_j - \sum_{\{i\,|\,m_i \geq 1\}} \Theta_{\iota i} \sum_{\{j\,|\,m_j \geq 1\}} \Phi_{ij} d_j \\
&= \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi} \end{pmatrix} \begin{pmatrix} \mathbb{c} \\ \mathbb{d} \end{pmatrix}\right]_\iota.
\end{aligned} \tag{181}$$

Analogically for $n < \iota \leq n+m$.

Finally, we have only to rewrite the constraints using (9) from Theorem 1.2:

$$f'(x_\iota) = \left\langle \psi_{x_\iota}, \sum_{\{i \,|\, n_i \geq 1\}} c_i \psi'_{x_i} + h_f \right\rangle_{Sob,m} = \left[\mathbf{\Psi}^{(1)} \mathbb{c} \right]_\iota \quad \forall \iota : n_\iota \geq 1. \tag{182}$$

We analogically obtain

$$g'(x_\iota) = \left[\mathbf{\Phi}^{(1)} \mathbb{d} \right]_\iota \quad \forall \iota : m_\iota \geq 1, \tag{183}$$

$$f''(x_\iota) = \left[\mathbf{\Psi}^{(2)} \mathbb{c} \right]_\iota \quad \forall \iota : n_\iota \geq 1, \tag{184}$$

$$g''(x_\iota) = \left[\mathbf{\Phi}^{(2)} \mathbb{d} \right]_\iota \quad \forall \iota : m_\iota \geq 1. \tag{185}$$

□

References

Adams, R. A. (1975). *Sobolev Spaces.* New York: Academic Press.

Aït-Sahalia, Y. and A. W. Lo (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics 94*(1–2), 9–51.

Aït-Sahalia, Y., Y. Wang, and F. Yared (2001). Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics 102*(1), 67–110.

Anděl, J. (2002). *Základy matematické statistiky.* Prague: MatfyzPress. Preprint in Czech.

Breeden, D. and R. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business 51*, 621–651.

de Geer, S. V. (1990). Estimating a regression function. *The Annals of Statistics 18*(2), 907–924.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing.* New York: Marcel Dekker, Inc.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Härdle, W. and Z. Hlávka (2006). Dynamics of state price densities. submitted to Journal of Econometrics.

Jackwerth, J. (1999). Option-implied risk-neutral distributions and implied binomial trees: a literature review. *Journal of Derivatives 7*, 66–82.

Maz'ja, V. G. (1985). *Sobolev Spaces.* Berlin: Springer-Verlag.

Pešta, M. (2006). Isotonic regression in sobolev spaces. Master's thesis, Charles University in Prague, Czech Republic.

Yatchew, A. and W. Härdle (2005). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*. doi:10.1016/j.jeconom.2005.06.031.

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

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
004 "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.
005 "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.
006 "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
007 "Robust utility maximization in a stochastic factor model" by Daniel Hernández–Hernández and Alexander Schied, January 2006.
008 "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.
009 "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.
010 "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, Jauary 2006.
011 "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.
012 "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.
013 "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.
014 "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.
015 "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.
016 "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.
017 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, March 2006.
018 "The Bologna Process: How student mobility affects multi-cultural skills and educational quality" by Lydia Mechtenberg and Roland Strausz, March 2006.
019 "Cheap Talk in the Classroom" by Lydia Mechtenberg, March 2006.
020 "Time Dependent Relative Risk Aversion" by Enzo Giacomini, Michael Handel and Wolfgang Härdle, March 2006.
021 "Finite Sample Properties of Impulse Response Intervals in SVECMs with Long-Run Identifying Restrictions" by Ralf Brüggemann, March 2006.
022 "Barrier Option Hedging under Constraints: A Viscosity Approach" by Imen Bentahar and Bruno Bouchard, March 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

023 "How Far Are We From The Slippery Slope? The Laffer Curve Revisited" by Mathias Trabandt and Harald Uhlig, April 2006.
024 "e-Learning Statistics – A Selective Review" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, April 2006.
025 "Macroeconomic Regime Switches and Speculative Attacks" by Bartosz Maćkowiak, April 2006.
026 "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets" by Bartosz Maćkowiak, April 2006.
027 "Institutional Competition, Political Process and Holdup" by Bruno Deffains and Dominique Demougin, April 2006.
028 "Technological Choice under Organizational Diseconomies of Scale" by Dominique Demougin and Anja Schöttner, April 2006.
029 "Tail Conditional Expectation for vector-valued Risks" by Imen Bentahar, April 2006.
030 "Approximate Solutions to Dynamic Models – Linear Methods" by Harald Uhlig, April 2006.
031 "Exploratory Graphics of a Financial Dataset" by Antony Unwin, Martin Theus and Wolfgang Härdle, April 2006.
032 "When did the 2001 recession *really* start?" by Jörg Polzehl, Vladimir Spokoiny and Cătălin Stărică, April 2006.
033 "Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power" by Jörg Polzehl and Vladimir Spokoiny, April 2006.
034 "Spectral calibration of exponential Lévy Models [1]" by Denis Belomestny and Markus Reiß, April 2006.
035 "Spectral calibration of exponential Lévy Models [2]" by Denis Belomestny and Markus Reiß, April 2006.
036 "Spatial aggregation of local likelihood estimates with applications to classification" by Denis Belomestny and Vladimir Spokoiny, April 2006.
037 "A jump-diffusion Libor model and its robust calibration" by Denis Belomestny and John Schoenmakers, April 2006.
038 "Adaptive Simulation Algorithms for Pricing American and Bermudan Options by Local Analysis of Financial Market" by Denis Belomestny and Grigori N. Milstein, April 2006.
039 "Macroeconomic Integration in Asia Pacific: Common Stochastic Trends and Business Cycle Coherence" by Enzo Weber, May 2006.
040 "In Search of Non-Gaussian Components of a High-Dimensional Distribution" by Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny and Klaus-Robert Müller, May 2006.
041 "Forward and reverse representations for Markov chains" by Grigori N. Milstein, John G. M. Schoenmakers and Vladimir Spokoiny, May 2006.
042 "Discussion of 'The Source of Historical Economic Fluctuations: An Analysis using Long-Run Restrictions' by Neville Francis and Valerie A. Ramey" by Harald Uhlig, May 2006.
043 "An Iteration Procedure for Solving Integral Equations Related to Optimal Stopping Problems" by Denis Belomestny and Pavel V. Gapeev, May 2006.
044 "East Germany's Wage Gap: A non-parametric decomposition based on establishment characteristics" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
045 "Firm Specific Wage Spread in Germany - Decomposition of regional differences in inter firm wage dispersion" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

046 "Produktdiversifizierung: Haben die ostdeutschen Unternehmen den Anschluss an den Westen geschafft? – Eine vergleichende Analyse mit Mikrodaten der amtlichen Statistik" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
047 "The Division of Ownership in New Ventures" by Dominique Demougin and Oliver Fabel, June 2006.
048 "The Anglo-German Industrial Productivity Paradox, 1895-1938: A Restatement and a Possible Resolution" by Albrecht Ritschl, May 2006.
049 "The Influence of Information Costs on the Integration of Financial Markets: Northern Europe, 1350-1560" by Oliver Volckart, May 2006.
050 "Robust Econometrics" by Pavel Čížek and Wolfgang Härdle, June 2006.
051 "Regression methods in pricing American and Bermudan options using consumption processes" by Denis Belomestny, Grigori N. Milstein and Vladimir Spokoiny, July 2006.
052 "Forecasting the Term Structure of Variance Swaps" by Kai Detlefsen and Wolfgang Härdle, July 2006.
053 "Governance: Who Controls Matters" by Bruno Deffains and Dominique Demougin, July 2006.
054 "On the Coexistence of Banks and Markets" by Hans Gersbach and Harald Uhlig, August 2006.
055 "Reassessing Intergenerational Mobility in Germany and the United States: The Impact of Differences in Lifecycle Earnings Patterns" by Thorsten Vogel, September 2006.
056 "The Euro and the Transatlantic Capital Market Leadership: A Recursive Cointegration Analysis" by Enzo Weber, September 2006.
057 "Discounted Optimal Stopping for Maxima in Diffusion Models with Finite Horizon" by Pavel V. Gapeev, September 2006.
058 "Perpetual Barrier Options in Jump-Diffusion Models" by Pavel V. Gapeev, September 2006.
059 "Discounted Optimal Stopping for Maxima of some Jump-Diffusion Processes" by Pavel V. Gapeev, September 2006.
060 "On Maximal Inequalities for some Jump Processes" by Pavel V. Gapeev, September 2006.
061 "A Control Approach to Robust Utility Maximization with Logarithmic Utility and Time-Consistent Penalties" by Daniel Hernández–Hernández and Alexander Schied, September 2006.
062 "On the Difficulty to Design Arabic E-learning System in Statistics" by Taleb Ahmad, Wolfgang Härdle and Julius Mungo, September 2006.
063 "Robust Optimization of Consumption with Random Endowment" by Wiebke Wittmüß, September 2006.
064 "Common and Uncommon Sources of Growth in Asia Pacific" by Enzo Weber, September 2006.
065 "Forecasting Euro-Area Variables with German Pre-EMU Data" by Ralf Brüggemann, Helmut Lütkepohl and Massimiliano Marcellino, September 2006.
066 "Pension Systems and the Allocation of Macroeconomic Risk" by Lans Bovenberg and Harald Uhlig, September 2006.
067 "Testing for the Cointegrating Rank of a VAR Process with Level Shift and Trend Break" by Carsten Trenkler, Pentti Saikkonen and Helmut Lütkepohl, September 2006.
068 "Integral Options in Models with Jumps" by Pavel V. Gapeev, September 2006.
069 "Constrained General Regression in Pseudo-Sobolev Spaces with Application to Option Pricing" by Zdeněk Hlávka and Michal Pešta, September 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".