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

This version: April 20, 2007

*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".

†Corresponding author: tel. +420/221913284, fax: +420/283073341, email: hlavka@karlin.mff.cuni.cz.

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

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

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) $h(.)$. 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:

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

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):

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

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, Wang and Yared (2001); Yatchew and Härdle (2006); Härdle and Hlávka (2006).

In this paper, we will generalize the nonlinear least squares method suggested in Yatchew and Härdle (2006) 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, 2006). In Sections 1 and 2, we will describe the mathematical foundation of the method. In Section 3, we 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, SPD estimates based on the observed DAX option prices are calculated in Section 5. The proofs of all theorems are given in Appendix A.

1 Pseudo-Sobolev Spaces

In this section, we give a brief overview of the basic results on the Pseudo-Sobolev spaces. We assemble and prove necessary preliminaries and some theorems for statistical regression in these spaces. The crux of this section lies in Theorem 1.1 (Representors in Pseudo-Sobolev Space) from Yatchew and Bos (1997). We examine representors' properties in more detail, see Theorem 1.2, providing both the exact form and the construction of the representors.

The symbol $\mathsf{L}_\mathsf{p}(\Omega)$ shall denote the *Lebesgue space* $\mathsf{L}_\mathsf{p}(\Omega) := \{f : \|f\|_{\mathsf{L}_\mathsf{p}(\Omega)} < \infty\}$, $1 \leq p \leq \infty$, where $\|f\|_{\mathsf{L}_\mathsf{p}(\Omega)} := (\int_\Omega f^p(\mathbb{x})\mathrm{d}\mathbb{x})^{1/p}$ for $1 \leq p < \infty$ and $\|f\|_{\mathsf{L}_\infty(\Omega)} := \inf\{C \geq 0 : |f| \leq C \text{ a.e.}\}$ for a measurable real-valued function $f : \Omega \to \mathbb{R}$ on a given Lebesgue-measurable domain Ω, i.e., a connected Lebesgue-measurable bounded subset of an Euclidean space $\mathbb{R}^q$ with non-empty interior.

The symbol $\mathcal{C}^m(\Omega)$, $m \in \mathbb{N}_0$ denotes the *space of m-times continuously differentiable scalar functions* upon bounded domain Ω, i.e., $\mathcal{C}^m(\Omega) := \left\{f : \Omega \to \mathbb{R} \middle| D^{\boldsymbol{\alpha}} f \in \mathcal{C}^0(\Omega), |\boldsymbol{\alpha}|_\infty \leq m\right\}$, where $|\boldsymbol{\alpha}|_\infty = \max_{i=1,\ldots,q} |\alpha_i|$.

1.1 Definition of Pseudo-Sobolev Space

Let us denote by $D^{\boldsymbol{\alpha}} f(\mathbb{x}) := \partial^{|\boldsymbol{\alpha}|_1} f(\mathbb{x}) / \partial x_1^{\alpha_1} \ldots \partial x_q^{\alpha_q}$ the partial derivative of the function $f : \Omega \to \mathbb{R}$ in $\mathbb{x} \in \operatorname{int}(\Omega) (\equiv \Omega^\circ := \overline{\Omega} \backslash \partial\Omega)$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^\top \in \mathbb{N}_0^q$ is a multiindex of the modulus $|\boldsymbol{\alpha}|_1 = \sum_{i=1}^q \alpha_i$.

Definition 1.1 (Sobolev Norm)**.** Let $f \in \mathcal{C}^m(\Omega) \cap \mathsf{L}_\mathsf{p}(\Omega)$. We introduce a Sobolev norm $\|\cdot\|_{p,Sob,m}$ as:

$$\|f\|_{p,Sob,m} := \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \leq m} \int_\Omega \left| D^{\boldsymbol{\alpha}} f(\mathbb{x}) \right|^p \mathrm{d}\mathbb{x} \right\}^{1/p}, \quad \text{where } |\boldsymbol{\alpha}|_\infty = \max_{i=1,\ldots,q} \alpha_i. \tag{4}$$

The triangle inequality for the Sobolev norm (4) follows easily from the triangle inequality for the p-norms on $\mathsf{L_p}(\Omega)$ and $l_p(\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}|_\infty \le m\})$. For any $f, g \in \mathcal{C}^m(\Omega) \cap \mathsf{L_p}(\Omega)$, we have that

$$\begin{aligned} \|f+g\|_{p,Sob,m} &= \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \le m} \|D^{\boldsymbol{\alpha}} f + D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)} \right\}^{1/p} \le \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \le m} \left[\|D^{\boldsymbol{\alpha}} f\|^p_{\mathsf{L_p}(\Omega)} + \|D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)} \right] \right\}^{1/p} \\ &\le \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \le m} \|D^{\boldsymbol{\alpha}} f\|^p_{\mathsf{L_p}(\Omega)} \right\}^{1/p} + \left\{ \sum_{|\boldsymbol{\alpha}|_\infty \le m} \|D^{\boldsymbol{\alpha}} g\|^p_{\mathsf{L_p}(\Omega)} \right\}^{1/p} = \|f\|_{p,Sob,m} + \|g\|_{p,Sob,m} . \end{aligned} \tag{5}$$

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

Definition 1.3 (Sobolev Inner Product)**.** Let $f, g \in \mathcal{W}^m_2(\Omega)$. The Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ is defined as: $\langle f, g \rangle_{Sob,m} := \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_\Omega D^{\boldsymbol{\alpha}} f(\mathbb{x}) D^{\boldsymbol{\alpha}} g(\mathbb{x}) \mathrm{d}\mathbb{x}$.

The correctness of Definition 1.3 is guaranteed by the denseness of the space $\mathcal{C}^m(\Omega) \cap L_2(\Omega)$ in $\mathcal{W}^m_2(\Omega)$, see Remark 1.1. The Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ induces the Sobolev norm $\|\cdot\|_{2,Sob,m}$ in $\mathcal{W}^m_2(\Omega)$ and we denote the Pseudo-Sobolev space $\mathcal{H}^m(\Omega) := \mathcal{W}^m_2(\Omega)$. For simplicity of notation, we denote the Sobolev norm $\|\cdot\|_{2,Sob,m} := \|\cdot\|_{Sob,m}$.

It is straightforward to verify that $\mathcal{H}^m(\Omega)$ is a normed linear space. By construction, $\mathcal{H}^m(\Omega)$ is complete and, hence, it is Banach space. Next, the inner product $\langle \cdot, \cdot \rangle_{Sob,m}$ has been defined on $\mathcal{H}^m(\Omega)$ and it follows that $\mathcal{H}^m(\Omega)$ is Hilbert space.

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

1.2 Construction of Representors in Pseudo-Sobolev Space

The Hilbert space $\mathcal{H}^m(\Omega)$ can be expressed as a direct sum of subspaces that are orthogonal to each other. For the nonparametric regression, see Section 2, it is very important that we can take projections of the elements of $\mathcal{H}^m(\Omega)$ into its subspaces.

The following Theorem 1.1 is the representation theorem for Pseudo-Sobolev spaces derived in Yatchew and Bos (1997), an analogy to the well-known Riesz Representation Theorem. From now on, we suppose that $m \in \mathbb{N}$. The symbol $\mathcal{Q}^q$ denotes the closed unit cube in $\mathbb{R}^q$.

Theorem 1.1 (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 representor $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) \in \mathcal{H}^m(\mathcal{Q}^q)$ at the point $\mathbb{a}$ with the rank $\mathbb{w}$ such that $\left\langle \psi_{\mathbb{a}}^{\mathbb{w}}, f \right\rangle_{Sob,m} = D^{\mathbb{w}} f(\mathbb{a})$. Furthermore, $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) = \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i)$ for all $\mathbb{x} \in \mathcal{Q}^q$, where $\psi_{a_i}^{w_i}(\cdot)$ is the representor in the Pseudo-Sobolev space of functions of one variable on $\mathcal{Q}^1$ with the inner product*

$$\frac{\partial^{w_i} f(\mathbb{a})}{\partial x_i^{w_i}} = \left\langle \psi_{a_i}^{w_i}, f(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_q) \right\rangle_{Sob,m} = \sum_{\alpha=0}^{m} \int_{\mathcal{Q}^1} \frac{d^\alpha \psi_{a_i}^{w_i}(x_i)}{dx_i^\alpha} \frac{d^\alpha f(\mathbb{x})}{dx_i^\alpha} dx_i. \tag{6}$$

The proof of Theorem 1.1 given in Appendix A is based on the ideas of Yatchew and Bos (1997). In addition, we derive the exact form of the representor for Pseudo-Sobolev spaces $\mathcal{W}_p^m(\Omega)$.

In order to derive the representor $\psi_a \equiv \psi_a^0$ of $f \in \mathcal{H}^m[0,1]$, we start with functions L_a and R_a defined in Appendix A in (51) and (54). The coefficients $\gamma_k(a)$ of the representor are obtained as the solution of a system linear equations corresponding to the boundary conditions (42)–(46) of the differential equation (41). The existence and uniqueness of the coefficients $\gamma_k(a)$ is shown in the proof of Theorem 1.1.

Theorem 1.2 (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:*

$$\sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \gamma_k(a) \left\{ \varphi_k^{(m-j)}(0) + (-1)^j \varphi_k^{(m+j)}(0) \right\} + \sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \gamma_{m+1+k}(a) \left\{ \varphi_{m+1+k}^{(m-j)}(0) + (-1)^j \varphi_{m+1+k}^{(m+j)}(0) \right\} = 0 \tag{7}$$

for $j = 0, \ldots, m-1$,

$$\sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \gamma_{2m+2+k}(a) \left\{ \varphi_k^{(m-j)}(1) + (-1)^j \varphi_k^{(m+j)}(1) \right\} + \sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \gamma_{3m+3+k}(a) \left\{ \varphi_{m+1+k}^{(m-j)}(1) + (-1)^j \varphi_{m+1+k}^{(m+j)}(1) \right\} = 0 \tag{8}$$

for $j = 0, \ldots, m-1$,

$$\sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \left\{ \gamma_k(a) - \gamma_{2m+2+k}(a) \right\} \varphi_k^{(j)}(a) + \sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \left\{ \gamma_{m+1+k}(a) - \gamma_{3m+3+k}(a) \right\} \varphi_{m+1+k}^{(j)}(a) = 0, \tag{9}$$

for $j = 0, \ldots, 2m-2$, and

$$\sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \left\{ \gamma_k(a) - \gamma_{2m+2+k}(a) \right\} \varphi_k^{(2m-1)}(a) + \sum_{\substack{k=0 \\ k\neq\kappa}}^{m} \left\{ \gamma_{m+1+k}(a) - \gamma_{3m+3+k}(a) \right\} \varphi_{m+1+k}^{(2m-1)}(a) = (-1)^{m-1}, \tag{10}$$

where κ is the integer part of $(m+1)/2$ and φ_k are defined in Appendix A in (49a)–(50d).

Let 0_n denote column vector of zeros of length n and $\boldsymbol{\gamma}(a)$ a column vector of the coefficients $\gamma_k(a)$,

$k = 1, \ldots, 4m+3$, appearing in equations (7)–(10) with nonzero coefficients, i.e.,

$$\begin{aligned}\boldsymbol{\gamma}(a) &= (\gamma_0(a), \ldots, \gamma_{\kappa-1}(a), \gamma_{\kappa+1}(a), \ldots, \gamma_{m+\kappa}(a), \gamma_{m+2+\kappa}(a), \ldots, \gamma_{2m+1}(a),\\ &\quad \gamma_{2m+2}(a), \gamma_{2m+1+\kappa}(a), \gamma_{2m+3+\kappa}(a), \ldots, \gamma_{3m+2+\kappa}(a), \gamma_{3m+4+\kappa}(a), \ldots, \gamma_{4m+3}(a))^\top\end{aligned}$$

The system of the $4m$ linear equations (7)–(10) can now be written in a more illustrative way:

$$\underbrace{\left(\begin{array}{c|c} \varphi_k^{(m-j)}(0) + (-1)^j \varphi_k^{(m+j)}(0) & 0_{m-1}0_{m-1}^\top \\ \hline 0_{m-1}0_{m-1}^\top & \varphi_k^{(m-j)}(1) + (-1)^j \varphi_k^{(m+j)}(1) \\ \hline \varphi_k^{(j)}(a) & -\varphi_k^{(j)}(a) \\ \hline \varphi_k^{(2m-1)}(a) & -\varphi_k^{(2m-1)}(a) \end{array}\right)}_{\{\Gamma_{j,k}(a)\}} \boldsymbol{\gamma}(\alpha) = \left(\begin{array}{c} 0_{m-1} \\ \hline 0_{m-1} \\ \hline 0_{2(m-1)} \\ \hline (-1)^{m-1} \end{array}\right). \qquad (11)$$

Hence, the coefficients can be expressed as $\boldsymbol{\gamma}(a) = (-1)^{m-1}\left[\{\boldsymbol{\Gamma}(a)\}^{-1}\right]_{\bullet,4m}$.

2 General Least Squares

A combination of properties of the L_2 and $\mathcal{C}^m$ space yields an interesting background for nonparametric regression. The L_2 space is a special type of Hilbert space that facilitates the calculation of least squares projection. On the other hand, the $\mathcal{C}^m$ space contains classes of smooth (m-times continuously differentiable) functions suitable for nonparametric regression.

The general single equation model investigated in this section is:

$$Y_i = f(\mathbb{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \qquad (12)$$

where $\mathbb{x}_i$ are q-dimensional fixed design points (knots), ε_i are correlated random errors such that $\mathsf{E}\varepsilon_i = 0$ and $\mathsf{Var}\varepsilon = \boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,\ldots,n}$ with $\sigma_i^2 = \sigma_{ii} > 0$, and $f \in \mathscr{F}$, where $\mathscr{F}$ is a family of functions in the Pseudo-Sobolev space $\mathcal{H}^m(\mathcal{Q}^q)$ from $\mathbb{R}^q$ to $\mathbb{R}^1$, $m > \frac{q}{2}$, $\mathscr{F} = \left\{f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|_{Sob,m}^2 \le L\right\}$. From now on, we denote $\mathcal{H}^m \equiv \mathcal{H}^m(\mathcal{Q}^q)$.

In case of i.i.d. observations, the estimation of f is carried out in one of these ways:

$$\min_{f\in\mathcal{H}^m} \tfrac{1}{n}\sum_{i=1}^n \left[Y_i - f(\mathbb{x}_i)\right]^2 \text{ such that } \|f\|_{Sob,m}^2 \le L, \qquad (13)$$

$$\min_{f\in\mathcal{H}^m} \left\{\tfrac{1}{n}\sum_{i=1}^n \left[Y_i - f(\mathbb{x}_i)\right]^2 + \chi\|f\|_{Sob,m}^2\right\}. \qquad (14)$$

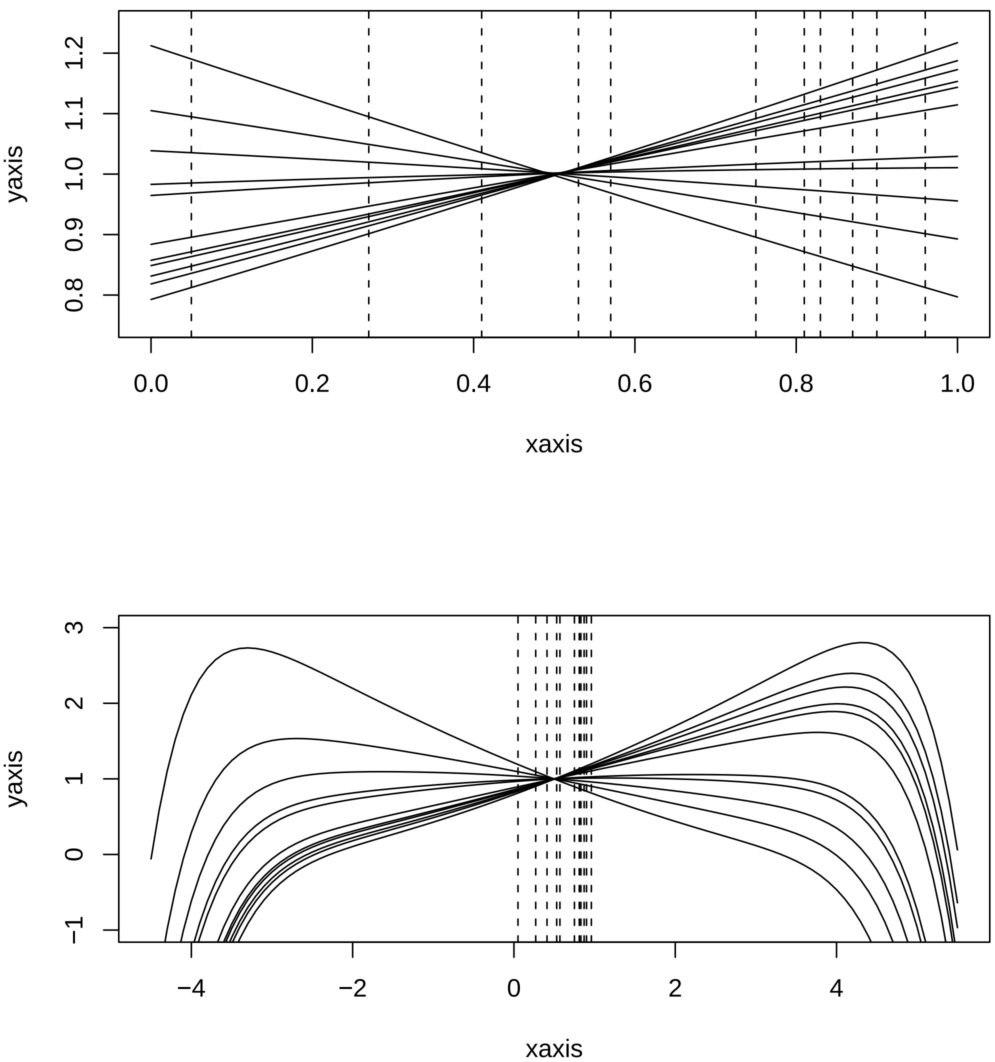

Figure 1: The full lines in both plots are the representors in the Pseudo-Sobolev space $\mathcal{H}^4\,[0,1]$ for data points $\mathbb{x} = (0.05, 0.27, 0.41, 0.53, 0.57, 0.75, 0.81, 0.83, 0.87, 0.9, 0.96)^\top$ marked by dashed vertical lines. The limits of the horizontal axis in the upper and the lower graph are $[0,1]$ and $[-4.5, +5.5]$, respectively.

The Sobolev norm bound L and the smoothing (or bandwidth) parameter χ control the trade-off between the infidelity to the data and the roughness of the estimator. For heteroscedastic and correlated data, we rewrite (14) as

$$\min_{f \in \mathcal{H}^m} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi \left\|f\right\|^2_{Sob,m} \tag{15}$$

where $\mathbb{x}$ is $(n \times q)$ matrix containing in its rows the q-dimensional design points $\mathbb{x}_1, \ldots, \mathbb{x}_n$, $\boldsymbol{\Sigma} > 0$ is $n \times n$ symmetric (variance) matrix, $\mathbb{Y}$ is $n \times 1$ vector of observations, $\boldsymbol{f}(\mathbb{x}) = (f(\mathbb{x}_1), \ldots, f(\mathbb{x}_n))^\top$, and $\chi > 0$.

Definition 2.1 (Representor Matrix)**.** Let $\psi_{\mathbb{x}_1}, \ldots, \psi_{\mathbb{x}_n}$ be the representors for function evaluation at $\mathbb{x}_1, \ldots, \mathbb{x}_n$, respectively. I.e., $\langle \psi_{\mathbb{x}_i}, f \rangle_{Sob,m} = f(\mathbb{x}_i)$ for all $f \in \mathcal{H}^m$, $i = 1, \ldots, n$. The representor matrix $\boldsymbol{\Psi}$ is the $(n \times n)$ matrix such that its columns and rows are the representors evaluated at $\mathbb{x}_1, \ldots, \mathbb{x}_n$, i.e., $\boldsymbol{\Psi} = (\Psi_{i,j})_{i,j=1,\ldots,n}$, where $\Psi_{i,j} = \left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} = \psi_{\mathbb{x}_i}(\mathbb{x}_j) = \psi_{\mathbb{x}_j}(\mathbb{x}_i)$.

Theorem 2.1 (Infinite to Finite)**.** *Assume that* $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$ *and* $\boldsymbol{\Sigma} > 0$ *is* $(n \times n)$ *symmetric matrix. Define*

$$\hat{\sigma}^2 = \min_{f \in \mathcal{H}^m} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi \left\|f\right\|^2_{Sob,m}, \tag{16}$$

$$s^2 = \min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right] + \chi \mathbb{c}^\top \boldsymbol{\Psi} \mathbb{c} \tag{17}$$

where $\mathbb{c}$ *is a* $(n \times 1)$ *vector,* $\boldsymbol{f}$ *is defined in (15), and* $\boldsymbol{\Psi}$ *is the representor matrix.*

Then $\hat{\sigma}^2 = s^2$*. Furthermore, there exists a solution to* (16) *of the form* $\widehat{f} = \sum_{i=1}^n \hat{c}_i \psi_{\mathbb{x}_i}$*, where* $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)^\top$ *solves* (17)*. The estimator* $\widehat{f}$ *is unique a.e.*

Theorem 2.1 transforms the infinite dimensional problem into a finite dimensional quadratic optimization problem. Similar result derived in Yatchew and Bos (1997) uses different penalization.

Corollary 2.2 (Form of the Regression Function Estimator)**.** *In one-dimensional case, the regression function estimator* $\widehat{f}$ *defined in Theorem 2.1 can be written as:*

$$\widehat{f}(x) = \begin{cases} \sum_{i=1}^{n} \hat{c}_i L_{x_i}(x), & 0 \le x \le x_1, \\ \sum_{i=j+1}^{n} \hat{c}_i L_{x_i}(x) + \sum_{i=1}^{j} \hat{c}_i R_{x_i}(x), & x_j < x \le x_{j+1},\ j = 1, \ldots, n-1; \\ \sum_{i=1}^{n} \hat{c}_i R_{x_i}(x), & x_n < x \le 1, \end{cases} \tag{18}$$

where $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)^\top$ *solves* (17) *and* $L_{x_i}(x)$ *and* $R_{x_i}(x)$ *are defined in* (37)*.*

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 1.1. We apply (37) on the form of each factor ψ_a of the product of representors $\psi_{\mathbb{a}}$ in (55). The only difference in (18) will be the number of cases. We will obtain $(n+1)^q$ decision conditions (vector $\mathbb{x}$ has q components) instead of actual number $n+1$ $(0 \leq x \leq x_1, \ldots, x_j < x \leq x_{j+1}, \ldots, x_n < x)$.

Alternatively, the regression function estimator $\widehat{f}$ can be written as:

$$\widehat{f}(x) = \sum_{j=1}^{n} \widehat{c_j} \sum_{k=1}^{2m} \exp\left[\Re\left(e^{i\theta_k}\right)x\right] \left\{ I_{[x \leq x_j]} \gamma_k(x_j) \cos\left[\Im\left(e^{i\theta_k}\right)x\right] + I_{[x > x_j]} \gamma_{2m+k}(x_j) \sin\left[\Im\left(e^{i\theta_k}\right)x\right] \right\}. \quad (19)$$

Note that the estimator $\widehat{f}$ is not calculated using trigonometric splines neither kernel functions!

Theorem 2.3 (Symmetry and Positive Definiteness of Representor Matrix)**.** *The representor matrix is symmetric and positive definite.*

In the linear model, the unknown coefficients are estimated using Least Squares. Gauss-Markov Theorem (Rao, 1973, Chapter 4) says that the Least Squares estimator is the best linear unbiased estimator and underlies the so-called normal equations. The normal equations for our model are derived in Theorem 2.4.

Theorem 2.4 (Normal Equations for $\widehat{\mathbb{c}}$)**.** *Let us consider the general single equation model (12). Let $\mathbb{Y}$ denote the response vector $(Y_1, \ldots, Y_n)^\top$ and $\boldsymbol{\Psi}$ the representor matrix. Then, the vector $\widehat{\mathbb{c}} = (c_1, \ldots, c_n)^\top$ of the coefficients of the minimizer $\widehat{f} = \sum_{i=1}^{n} \hat{c}_i \psi_{\mathbb{x}_i}$ derived in Theorem 2.1 is the unique solution of the system of equations $\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi} + n\chi\boldsymbol{\Psi}\right)\mathbb{c} = \boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\mathbb{Y}$.*

Remark 2.2 (Hat Matrix). The fitted values $\widehat{\mathbb{Y}}$ can be expressed as $\widehat{\mathbb{Y}} = \widehat{\boldsymbol{f}}(\mathbb{x}) = \boldsymbol{\Psi}\widehat{\mathbb{c}}$. From the normal equations for $\widehat{\mathbb{c}}$, see Theorem 2.4, we obtain the so-called hat matrix $\boldsymbol{\Lambda} := \boldsymbol{\Psi}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi} + n\chi\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}$ satisfying $\widehat{\mathbb{Y}} = \boldsymbol{\Lambda}\mathbb{Y}$.

Using the Infinite to Finite Theorem 2.1 and Lagrange multipliers, a one-to-one correspondence between the Sobolev bound L and the smoothing parameter χ can be easily shown (Pešta, 2006). Formally, the relationship between the parameters L and χ is described in the following Theorems 2.5 and 2.6.

Theorem 2.5 (1–1 Mapping of Smoothing Parameters)**.** *Let $L > 0$, $\boldsymbol{\Sigma}$ is positive definite and symmetric matrix and*

$$f^* = \arg\min_{f \in \mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] \quad s.t. \;\; \|f\|_{Sob,m}^2 \leq L. \quad (20)$$

Then, there exists a unique $\chi \geq 0$ such that

$$f^* = \arg\min_{f \in \mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi\|f\|_{Sob,m}^2. \quad (21)$$

Theorem 2.6 (Bijection Between the Smoothing Parameters)**.** *Assume that* $\chi > 0$ *and* $\mathbf{\Sigma}$ *is positive definite and symmetric matrix and* $f^* = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right] + \chi \|f\|^2_{Sob,m}$. *Then, there exists a unique* $L > 0$ *such that* $f^* = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\mathbb{Y} - \boldsymbol{f}(\mathbb{x})\right]$ *s.t.* $\|f\|^2_{Sob,m} = L$.

If ${\mathbb{c}^*}^\top \mathbf{\Psi}\mathbb{c}^* < L$, then $\mathbb{Y} = \mathbf{\Psi}\mathbb{c}^* = \widehat{\mathbb{Y}}$ and the estimator just interpolates the observations.

Theorem 2.7 (Asymptotic Behavior)**.** *Suppose that* $\tilde{\varepsilon} := \mathbf{\Xi}\boldsymbol{\varepsilon}$ *is an* $(n \times 1)$ *vector of i.i.d. random variables. Then*

$$\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] = \mathcal{O}_\mathsf{P}\left(n^{-\frac{2m}{2m+q}}\right), \quad n \to \infty. \tag{22}$$

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.

A well-known selection method is based on the minimization of the cross-validation criterion $\mathcal{CV}(L) = \frac{1}{n}\left[\mathbb{y} - \widehat{\mathbf{f}^*}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\mathbb{y} - \widehat{\mathbf{f}^*}(\mathbb{x})\right]$, where $\widehat{\mathbf{f}^*} = \left(\widehat{f}_{-1}, \ldots, \widehat{f}_{-n}\right)^\top$ is the usual leave-one-out estimator obtained by solving $\widehat{f}_{-i} = \arg\min_{f\in\mathcal{H}^m} \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\left[\mathbf{\Xi}_{j,\bullet}\mathbb{y} - \mathbf{\Xi}_{j,\bullet}\mathbf{f}(\mathbb{x})\right]^2 + \chi\|f\|^2_{Sob,m}, \quad i = 1, \ldots, n$, where $\mathbf{\Xi}$ denotes the square root matrix of $\mathbf{\Sigma}^{-1}$. The smoothing parameter χ, which in-turn corresponds to unique Sobolev bound L, is chosen as the minimizer of the Cross-Validation function $\mathcal{CV}$. The relationship between the fit and the smoothness of the estimator is plotted in Figure 2.

Detailed information concerning the choice of the smoothing parameter χ can be found in Eubank (1999). Apart of the cross validation, there exist many other 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, are described, among others., in Härdle (1990).

3 Application to Option Prices

In Section 2, we have imposed only smoothness constraint on the estimated regression function $f \in \mathscr{F} = \left\{f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|^2_{Sob,m} \leq L\right\}$. However, in practice we often have a prior knowledge concerning the shape of the regression function. Hence, in this section we will focus on the inclusion of additional constraints, such as isotonia or convexity, in the nonparametric regression estimator.

Formally, we are interested in the estimation of $f \in \widetilde{\mathscr{F}} \subseteq \mathscr{F}$ where $\widetilde{\mathscr{F}}$ combines smoothness with further properties such as monotonicity of particular derivatives of the function. The following discussion concerns only the one-dimensional case ($q = 1$).

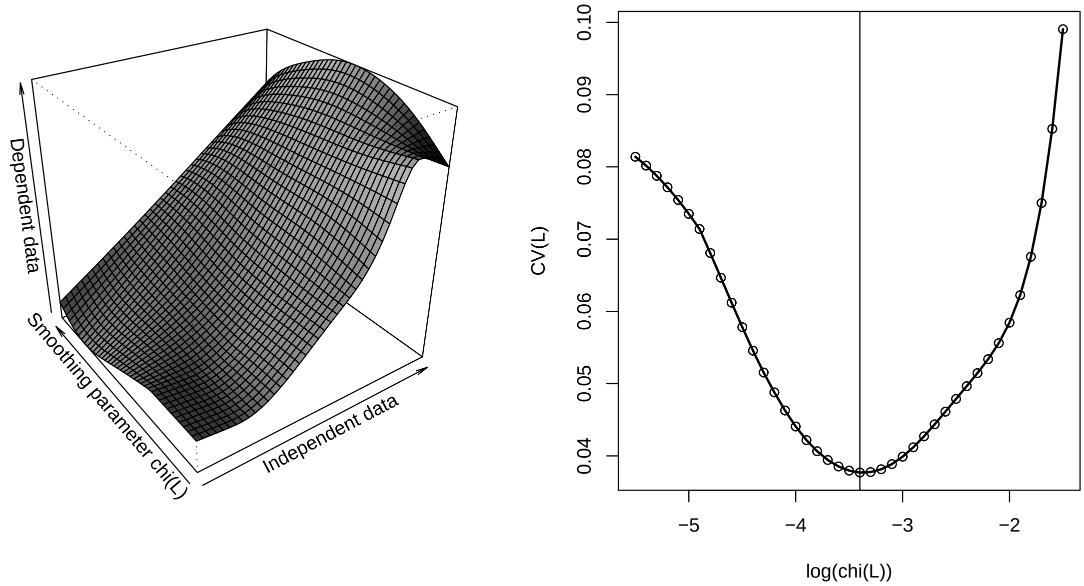

Figure 2: The left plot shows how the fitted curve in $\mathcal{H}^2$ changes depending on the smoothing parameter χ. The right plot displays the cross-validation criterion as function of χ. The optimal value of the smoothing parameter is marked by a vertical line.

Definition 3.1 (Derivative of the 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 $f \in \mathcal{H}^m(\mathcal{Q}^1)$, $i = 1, \ldots, n$. The k-th derivative of the representor matrix $\mathbf{\Psi}$ is the matrix $\mathbf{\Psi}^{(k)}$ whose columns are equal to the k-th derivatives of the representors evaluated at $x_1, \ldots, x_n$, i.e., $\Psi^{(k)}_{i,j} = \psi^{(k)}_{x_j}(x_i)$, $i, j = 1, \ldots, n$.

Contrary to Theorem 2.3, the derivatives of the representor matrix do not have to be symmetric.

Definition 3.2 (Estimate of the Derivative)**.** The estimate of the derivative of the regression function is defined as the derivative of the regression function estimate, i.e., $\widehat{f^{(s)}} := \widehat{f}^{(s)}, \quad s \in \mathbb{N}$.

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

Now we can show the relationship between the operator of derivative of the representor matrix and isotonia. We will concentrate mainly on the application to the Call and Put option properties.

3.1 Call and Put options

Suppose that Call and Put option prices are observed repeatedly for fixed distinct strike prices ϖ_i, $i = 1, \ldots, \omega$. The points ϖ_i are called the strike price knots.

In each strike price knot ϖ_i, we observe $n_i \in \mathbb{N}_0$ Call option prices Y_{i_k} with strikes $x_{i_k} = \varpi_i$, for $k = 1, \ldots, n_i$. We observe altogether $n = \sum_{i=1}^{\omega} n_i I_{[n_i \geq 1]}$ Call options in $\omega_Y = \sum_{i=1}^{\omega} I_{[n_i \geq 1]}$ distinct strike price knots. Similarly, in each strike price knot ϖ_j, $j = 1, \ldots, \omega$, we observe m_j Put option prices Z_{j_l} with strike prices $x_{j_l} = \varpi_j$, for $l = 1, \ldots, n_j$. In $\omega_Z = \sum_{j=1}^{\omega} I_{[m_j \geq 1]}$ distinct strike price knots, we observe $m = \sum_{j=1}^{\omega} m_j I_{[m_j \geq 1]}$ Put option prices.

Let us now denote by $\mathbb{Y}$ the vector of all observed Call option prices and by $\mathbb{x}_\alpha = (x_{\alpha,1}, \ldots, x_{\alpha,n})^\top$ the vector of the corresponding strike prices. Next, the symbol $\boldsymbol{\Delta} = (\Delta_{ij})_{i=1,\ldots,n; j=1\ldots,\omega_Y}$ denotes the connectivity matrix for Call option strike prices such that $\Delta_{ij} = I_{[x_{\alpha,i} = \varpi_j]}$. The symbol $\mathbb{Z}$ denotes the observed Put option prices. The vector $\mathbb{x}_\beta = (x_{\beta,1}, \ldots, x_{\beta,m})^\top$ of the strike prices corresponding to $\mathbb{Z}$ leads the connectivity matrix $\boldsymbol{\Theta} = (\Theta_{ij})_{i=1,\ldots,m; j=1\ldots,\omega_Z}$ for Put option prices defined as $\Theta_{ij} = I_{[x_{\beta,i} = \varpi_j]}$. Similar matrix has been already defined in Yatchew and Härdle (2006).

Our model for the observed Call and Put options prices can be written as:

$$
\begin{aligned}
Y_i &= f(x_{\alpha,i}) + \varepsilon_i, \text{ where } x_{\alpha,i} \in \{\varpi_1, \ldots, \varpi_\omega\} \text{ and } i = 1, \ldots, n, && (23)\\
Z_j &= g(x_{\beta,j}) + \epsilon_j, \text{ where } x_{\beta,j} \in \{\varpi_1, \ldots, \varpi_\omega\} \text{ and } j = 1, \ldots, m. && (24)
\end{aligned}
$$

under the assumptions:

i) ε_{i_k} and ϵ_{j_l} are random variables such that $\mathsf{E}\varepsilon_i = \mathsf{E}\epsilon_j = 0$, $\forall i, j$, $\mathsf{cov}\,(\varepsilon_i, \varepsilon_k) = \xi_{i,k}$, $\mathsf{cov}\,(\epsilon_j, \epsilon_l) = \zeta_{j,l}$, and $\mathsf{cov}\,(\varepsilon_i, \epsilon_j) = \sigma_{i,j}$. For simplicity, we will write $\xi_i^2 = \xi_{i,i}$ and $\zeta_j^2 = \zeta_{j,j}$.

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

We assume that the second derivatives of functions f and g have to be the same SPD, see equations (1)–(3) in the introduction. Theorem 2.1 allows to handle multiple (repeated) observations in our option prices setup (23)–(24).

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

Model (23)–(24). Define

$$\hat{\sigma}^2 = \min_{f\in\mathcal{H}^p, g\in\mathcal{H}^p} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{f}(\mathbb{x}_\alpha) \\ \mathbf{g}(\mathbb{x}_\beta) \end{pmatrix}\right]^\top \mathbf{\Sigma}^{-1} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{f}(\mathbb{x}_\alpha) \\ \mathbf{g}(\mathbb{x}_\beta) \end{pmatrix}\right] + \chi \|f\|^2_{Sob,p} + \theta \|g\|^2_{Sob,p} \tag{25}$$

subject to

$$-\mathbf{1} \leq \mathbf{f}'(\mathbb{x}_\alpha) \leq \mathbf{0},\ \mathbf{0} \leq \mathbf{g}'(\mathbb{x}_\beta) \leq \mathbf{1},\ \mathbf{f}''(\mathbb{x}_\alpha) \geq \mathbf{0},\ \mathbf{g}''(\mathbb{x}_\beta) \geq \mathbf{0},\ \textit{and}\ \mathbf{f}''(\mathbb{x}_\gamma) = \mathbf{g}''(\mathbb{x}_\gamma) \tag{26}$$

and

$$s^2 = \min_{\mathbb{c}\in\mathbb{R}^{\omega_Y}, \mathbb{d}\in\mathbb{R}^{\omega_Z}} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi} \end{pmatrix} \begin{pmatrix} \mathbb{c} \\ \mathbb{d} \end{pmatrix}\right]^\top \times \mathbf{\Sigma}^{-1} \left[\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} - \begin{pmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi} \end{pmatrix} \begin{pmatrix} \mathbb{c} \\ \mathbb{d} \end{pmatrix}\right] + \chi \mathbb{c}^\top \mathbf{\Psi} \mathbb{c} + \theta \mathbb{d}^\top \mathbf{\Phi} \mathbb{d} \tag{27}$$

subject to

$$-\mathbf{1} \leq \mathbf{\Psi}^{(1)}\mathbb{c} \leq \mathbf{0},\ \mathbf{0} \leq \mathbf{\Phi}^{(1)}\mathbb{d} \leq \mathbf{1},\ \mathbf{\Psi}^{(2)}\mathbb{c} \geq \mathbf{0},\ \mathbf{\Phi}^{(2)}\mathbb{d} \geq \mathbf{0},\ \textit{and}\ \mathbf{\Psi}^{(2)}\mathbb{c}_\gamma = \mathbf{\Phi}^{(2)}\mathbb{d}_\gamma, \tag{28}$$

where $\chi > 0$, $\theta > 0$, $\mathbf{\Sigma}$ *is* $(n+m)\times(n+m)$ *positive definite and symmetric variance matrix,* $\mathbf{\Delta}$ *and* $\mathbf{\Theta}$ *are respectively the connectivity matrices for Call and Put options,* $\mathbf{\Psi}$ *is the* $\omega_Y \times \omega_Y$ *representor matrix at* $(x_\iota)^\top_{\iota\in\{\iota\,|\,n_\iota\geq 1\}}$, $\mathbf{\Phi}$ *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* (25) *with respect to* (26) *of the form*

$$\widehat{f} = \sum_{\{i\,|\,n_i\geq 1\}} \hat{c}_i \psi_{x_i} \ \textit{and}\ \widehat{g} = \sum_{\{j\,|\,m_j\geq 1\}} \hat{d}_j \phi_{x_j}, \tag{29}$$

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* (27), 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 $\mathbf{\Sigma}$ of the random errors $(\varepsilon_1,\ldots,\varepsilon_n,\epsilon_1,\ldots,\epsilon_m)^\top$ will be investigated in Section 4. The minimization problem (27) under the constraints (28) can be implemented using, e.g., 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 $h = (h(\varpi_1), \ldots, h(\varpi_\omega))^\top$. 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

$$Y_i = \alpha(x_i)^\top h + \varepsilon_i, \quad i = 1, \ldots, n, \text{ and } Z_j = \beta(x_j)^\top h + \epsilon_j, \quad j = 1, \ldots, m, \tag{30}$$

for the Call and Put option prices, respectively. 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 design matrices $\mathcal{X}_\alpha$ and $\mathcal{X}_\beta$ so that the observed option prices can be written as

$$\begin{pmatrix} \mathbb{Y} \\ \mathbb{Z} \end{pmatrix} = \begin{pmatrix} \mathcal{X}_\alpha \\ \mathcal{X}_\beta \end{pmatrix} h + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\epsilon} \end{pmatrix}. \tag{31}$$

In the following, the SPD may depend on the time of the observation and $h_k = (h_k(\varpi_1), \ldots, h_k(\varpi_\omega))^\top$ will denote the true value of the SPD at the time of the k-th trade, $k = 1, \ldots, n + m$.

4.1 Constant SPD

Assuming that the random errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{n+m})^\top$ in the linear model (31) are independent and identically distributed, the model (31) for the k-th observation, corresponding to the strike price x_k, can be written as

$$Y_i = \alpha(x_i)^\top h_k + \varepsilon_i, \text{ where } h_k = h, \tag{32}$$

if the k-th observation in the combined dataset is the i-th Call option price or

$$Z_j = \beta(x_j)^\top h_k + \epsilon_j \text{ where } h_k = h, \tag{33}$$

if the k-th observations in the combined dataset is the j-th Put option price.

Here, the SPD $h = h_1 = \cdots = h_{n+m}$ is constant in the observation period. This simplified model has been investigated in Yatchew and Härdle (2006) 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 $t_k \in (0,1)$ and denote by $\delta_k = t_k - t_{k-1} > 0$ the time between the $(k-1)$-st and the k-th trade. The model (32) can now be generalized by considering a different error structure:

$$\begin{aligned} Y_i &= \alpha(x_i)^\top h_k, \\ h_k &= h_{k-1} + \delta_k^{1/2} \varepsilon_k. \end{aligned}$$

Expressing all observations in terms of an artifical parameter $h = h_{n+m+1}$, corresponding to the time 1 that can be interpreted as, e.g., "end of the day", it follows that the covariance of any two observed call option prices depends only on their strike prices and on the time of the trade:

$$\begin{aligned} \text{Cov}\{Y_{i-u}, Y_{i-v}\} &= \text{Cov}(\alpha(x_{i-u})^\top h_{i-u}, \alpha(x_{i-v})^\top h_{i-v}) \\ &= \sigma^2 \alpha(x_{i-u})^\top \alpha(x_{i-v}) \sum_{m=1}^{\min(u,v)} \delta_{i+1-m}. \end{aligned} \tag{34}$$

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

$$\begin{aligned} \text{Cov}\{Z_{i-u}, Z_{i-v}\} &= \text{Cov}(\beta(x_{i-u})^\top h_{i-u}, \beta(x_{i-v})^\top h_{i-v}(k)) \\ &= \sigma^2 \beta(x_{i-u})^\top \beta(x_{i-v}) \sum_{l=1}^{\min(u,v)} \delta_{i+1-l}. \end{aligned} \tag{35}$$

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

$$\begin{aligned} \text{Cov}\{Y_{i-u}, Z_{i-v}\} &= \text{Cov}(\alpha_{x_{i-u}}^\top h_{i-u}, \beta(x_{i-v})^\top h_{i-v}(k)) \\ &= \sigma^2 \sum_{l=1}^{\min(u,v)} \delta_{i+1-l} \sum_{k=2}^{p-1} \alpha_{x_{i-u}}^\top \beta(x_{i-v}). \end{aligned} \tag{36}$$

Hence, the knowledge of the time of the trade allows us to approximate the covariance matrix of the observed option prices. Using this covariance structure, we can estimate arbitrary future value of the SPD. It is quite natural that more recent observations are more important for the construction of the estimator and that observations corresponding to the same strike price obtained 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 consisting of 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. The circles, squares and stars correspond, respectively, to morning, lunchtime and afternoon prices. Clearly, we observe both heteroscedasticity and strong dependency 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 large 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.

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 entire support of the SPD.

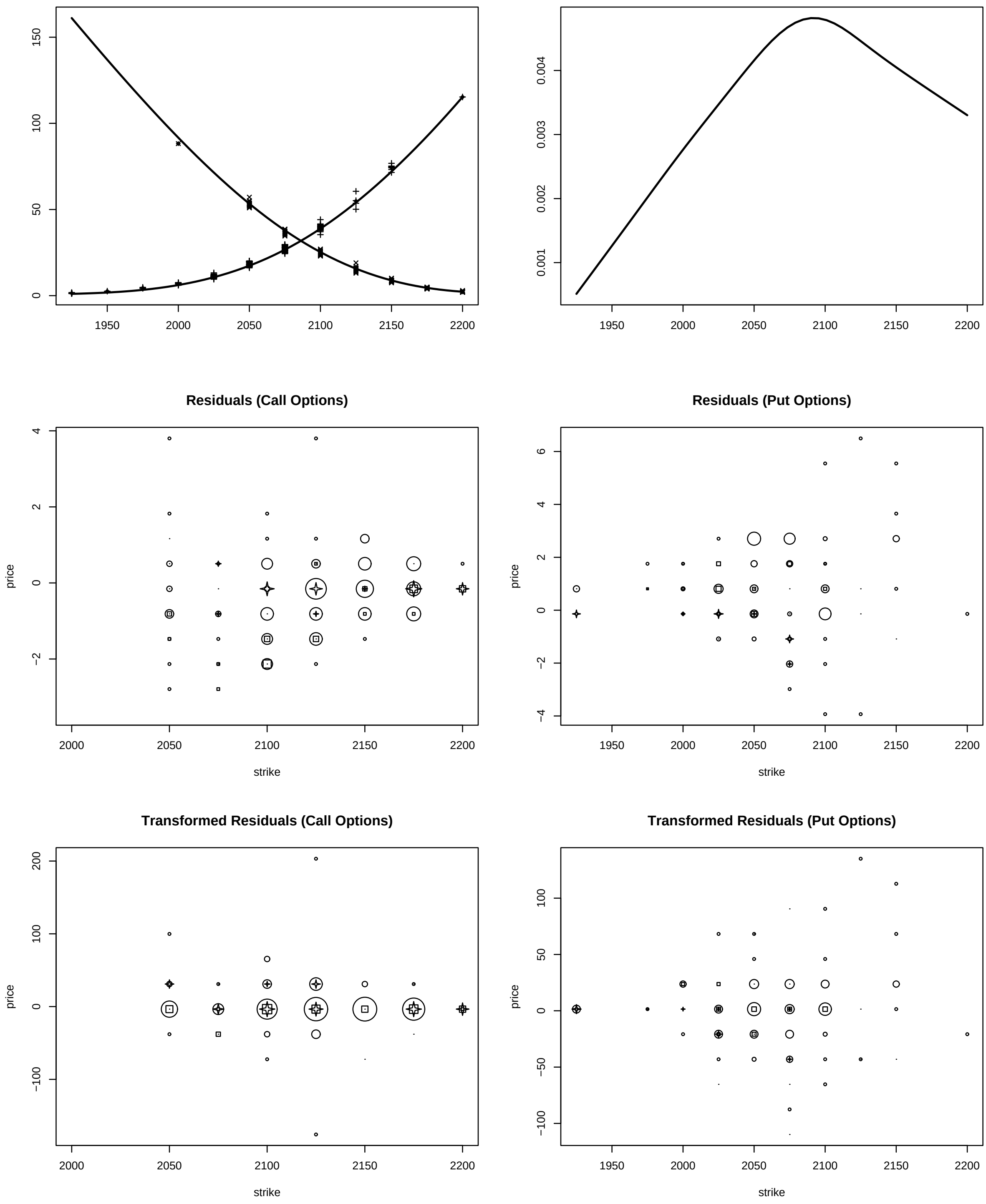

Figure 3: Estimates and residual plots on the 1st trading day in 1995 (January 2nd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and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

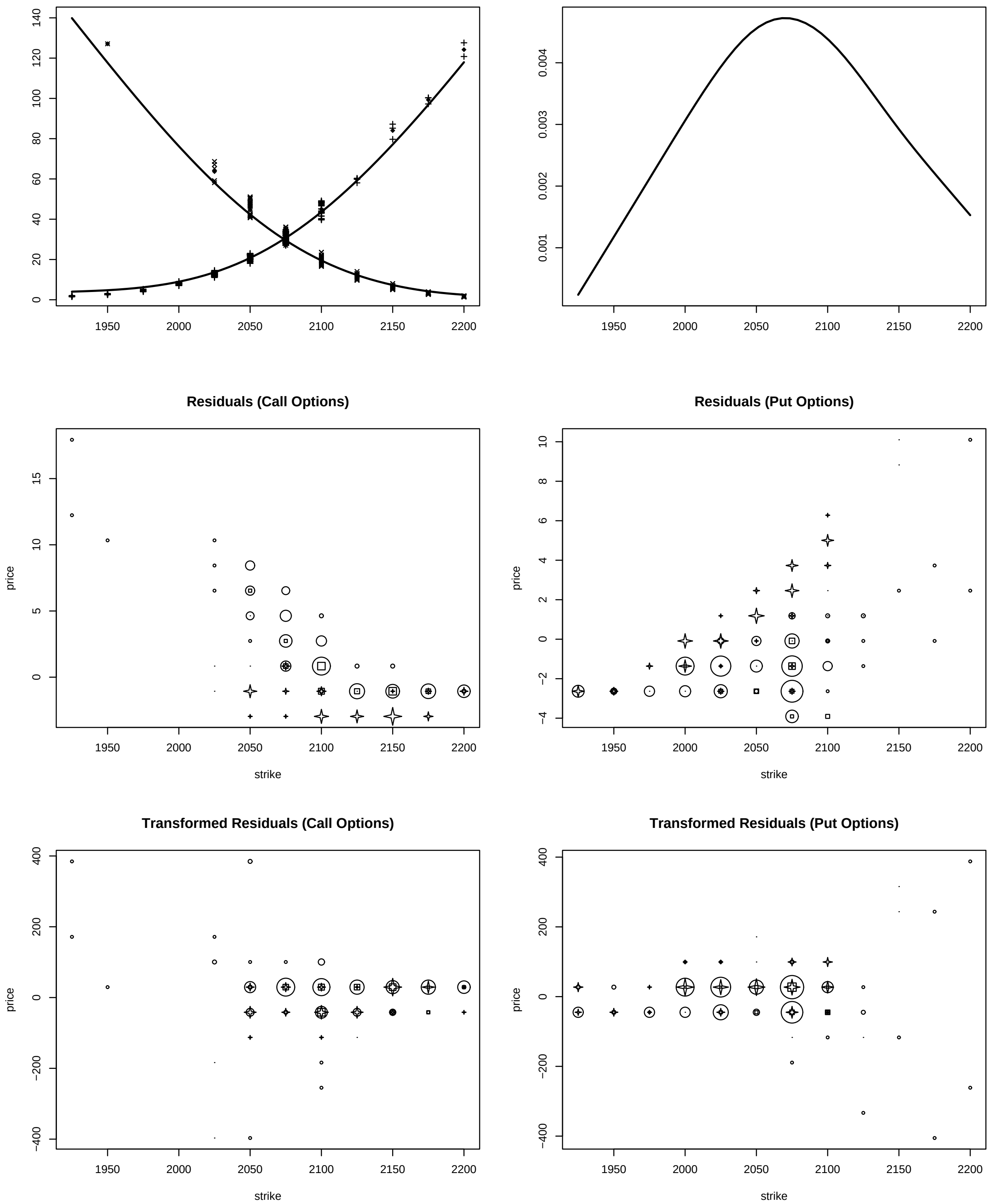

Figure 4: Estimates and residual plots on 2nd trading day in 1995 (January 3rd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and right 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.

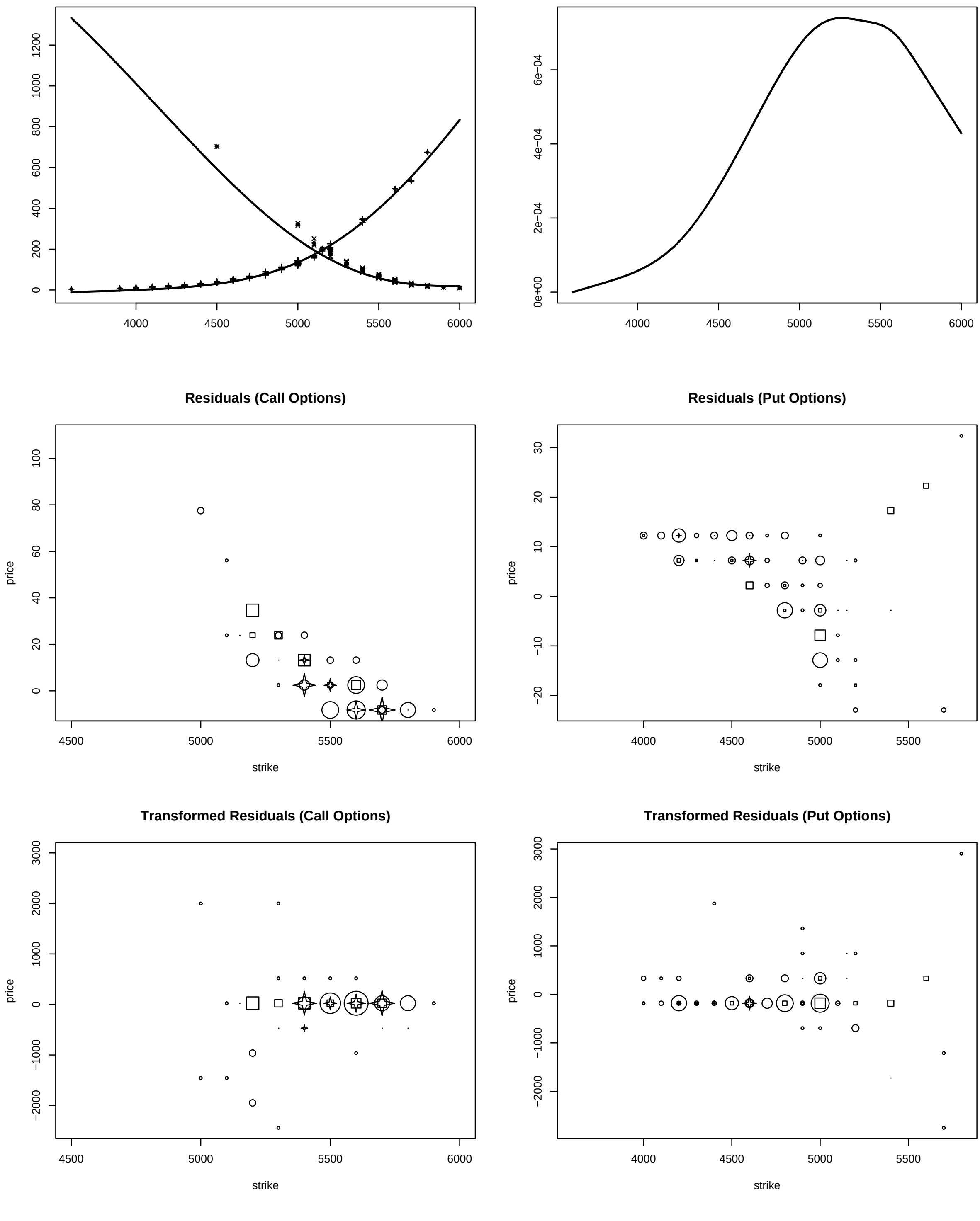

Figure 5: Estimates and residual plots on the 1st trading day in 2002 (January 2nd). The first plot shows fitted Call and Put option prices, the estimated SPD is plotted in the second plot. The remaining four graphics contain respectively residual plots for Call and Put option prices on the left and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

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 (2006). 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 (2006). The constrained general regression in pseudo-Sobolev spaces will certainly be very useful in various practical problems.

A Proofs

Proof of Theorem 1.1. We divide the proof into two steps. The proof follows closely the proof of Theorem 2.2 given in Yatchew and Bos (1997). However, we repeat it here since we need to introduce the notation needed for expressing the coefficients given in Theorem 1.2.

(i) Construction of a representor $\psi_a(\equiv \psi_a^0)$. For simplicity, let us set $\mathcal{Q}^1 \equiv [0,1]$. We know that for functions of one variable we have $\langle g,h\rangle_{Sob,m} = \sum_{k=0}^{m} \int_{\mathcal{Q}^1} g^{(k)}(x)h^{(k)}(x)\mathrm{d}x$. We are constructing a representor $\psi_a \in \mathcal{H}^m[0,1]$ such that $\langle \psi_a, f\rangle_{Sob,m} = f(a)$ 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. The representor is defined as:

$$\psi_a(x) = \begin{cases} L_a(x) & 0 \le x \le a, \\ R_a(x) & a \le x \le 1, \end{cases} \tag{37}$$

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

$$f(a) = \langle \psi_a, f \rangle_{Sob,m} = \int_0^a \sum_{k=0}^m L_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x + \int_a^1 \sum_{k=0}^m R_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x. \tag{38}$$

Integrating by parts and setting $i = k - j - 1$, we obtain:

$$\begin{aligned}
&\sum_{k=0}^m \int_0^a L_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x \\
&= \sum_{k=0}^m \left\{ \sum_{j=0}^{k-1} (-1)^j L_a^{(k+j)}(x) f^{(k-j-1)}(x) \Big|_0^a + (-1)^k \int_0^a L_a^{(2k)}(x) f(x) \mathrm{d}x \right\} \\
&= \sum_{k=0}^m \sum_{j=0}^{k-1} (-1)^j L_a^{(k+j)}(x) f^{(k-j-1)}(x) \Big|_0^a + \int_0^a \left\{ \sum_{k=0}^m (-1)^k L_a^{(2k)}(x) \right\} f(x) \mathrm{d}x \\
&= \sum_{k=1}^m \sum_{i=0}^{k-1} (-1)^{k-i-1} L_a^{(2k-i-1)}(x) f^{(i)}(x) \Big|_0^a + \int_0^a \left\{ \sum_{k=0}^m (-1)^k L_a^{(2k)}(x) \right\} f(x) \mathrm{d}x \\
&= \sum_{i=0}^{m-1} \sum_{k=i+1}^m (-1)^{k-i-1} L_a^{(2k-i-1)}(x) f^{(i)}(x) \Big|_0^a + \int_0^a \left\{ \sum_{k=0}^m (-1)^k L_a^{(2k)}(x) \right\} f(x) \mathrm{d}x \\
&= \sum_{i=0}^{m-1} f^{(i)}(a) \left\{ \sum_{k=i+1}^m (-1)^{k-i-1} L_a^{(2k-i-1)}(a) \right\} - \sum_{i=0}^{m-1} f^{(i)}(0) \left\{ \sum_{k=i+1}^m (-1)^{k-i-1} L_a^{(2k-i-1)}(0) \right\} \\
&\quad + \int_0^a \left\{ \sum_{k=0}^m (-1)^k L_a^{(2k)}(x) \right\} f(x) \mathrm{d}x
\end{aligned} \tag{39}$$

and, similarly,

$$\begin{aligned}
&\sum_{k=0}^m \int_a^1 R_a^{(k)}(x) f^{(k)}(x) \mathrm{d}x \\
&= \sum_{i=0}^{m-1} f^{(i)}(1) \left\{ \sum_{k=i+1}^m (-1)^{k-i-1} R_a^{(2k-i-1)}(1) \right\} - \sum_{i=0}^{m-1} f^{(i)}(a) \left\{ \sum_{k=i+1}^m (-1)^{k-i-1} R_a^{(2k-i-1)}(a) \right\} \\
&\quad + \int_a^1 \left\{ \sum_{k=0}^m (-1)^k R_a^{(2k)}(x) \right\} f(x) \mathrm{d}x.
\end{aligned} \tag{40}$$

This holds for all $f(x) \in \mathcal{C}^m[0,1]$. We require that both L_a and R_a are solutions of the constant coefficient differential equation

$$\sum_{k=0}^m (-1)^k \varphi_k^{(2k)}(x) = 0. \tag{41}$$

The boundary conditions are obtained by the equality of the functional values of $L_a^{(i)}(x)$ and $R_a^{(i)}(x)$ at a and the coefficient comparison of $f^{(i)}(0)$, $f^{(i)}(1)$ and $f^{(i)}(a)$, compare (38) to (39) and (40). Let $f^{(i)}(x) \bowtie c$

denote that the term $f^{(i)}(x)$ has the coefficient c in a certain equation. We can write:

$$r_a \in \mathcal{H}^m[0,1] \;\Rightarrow\; L_a^{(i)}(a) = R_a^{(i)}(a) \quad \ldots \quad 0 \le i \le m-1, \tag{42}$$

$$f^{(i)}(0) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} L_a^{(2k-i-1)}(0) = 0 \quad \ldots \quad 0 \le i \le m-1, \tag{43}$$

$$f^{(i)}(1) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} R_a^{(2k-i-1)}(1) = 0 \quad \ldots \quad 0 \le i \le m-1, \tag{44}$$

$$f^{(i)}(a) \bowtie 0 \;\Rightarrow\; \sum_{k=i+1}^{m} (-1)^{k-i-1} \left\{L_a^{(2k-i-1)}(a) - R_a^{(2k-i-1)}(a)\right\} = 0 \quad \ldots \quad 1 \le i \le m-1, \tag{45}$$

$$f(a) \bowtie 1 \;\Rightarrow\; \sum_{k=1}^{m} (-1)^{k-1} \left\{L_a^{(2k-1)}(a) - R_a^{(2k-1)}(a)\right\} = 1; \tag{46}$$

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 $P_m(\lambda) = \sum_{k=0}^{m} (-1)^k \lambda^{2k}$. Hence, it follows that

$$(1+\lambda^2)P_m(\lambda) = 1 + (-1)^m \lambda^{2m+2}, \quad \lambda \neq \pm i. \tag{47}$$

Solving (47), we get the characteristic roots $\lambda_k = e^{i\theta_k}$, where

$$\theta_k \in \begin{cases} \frac{(2k+1)\pi}{2m+2} & m \text{ even}, \quad k \in \{0,1,\ldots,2m+1\} \setminus \left\{\frac{m}{2}, \frac{3m+2}{2}\right\}, \\ \\ \frac{k\pi}{m+1} & m \text{ odd}, \quad k \in \{0,1,\ldots,2m+1\} \setminus \left\{\frac{m+1}{2}, \frac{3m+3}{2}\right\}. \end{cases} \tag{48}$$

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

m even

$$\varphi_k(x) = \exp\left\{\big(\Re(\lambda_k)\big)x\right\} \cos\left[\big(\Im(\lambda_k)\big)x\right],\ k \in \{0,1,\ldots,m\} \setminus \left\{\frac{m}{2}\right\}; \tag{49a}$$

$$\varphi_{m+1+k}(x) = \exp\left\{\big(\Re(\lambda_k)\big)x\right\} \sin\left[\big(\Im(\lambda_k)\big)x\right],\ k \in \{0,1,\ldots,m\} \setminus \left\{\frac{m}{2}\right\}. \tag{49b}$$

If m is odd, we have $2m-2$ different complex roots (each has a pair that is conjugate with it) and two real roots. The two real roots are ± 1. The $m-1$ complex conjugate roots have multiplicity one. We also have $2(m-1)+2 = 2m$ base elements alike all roots. These base elements are:

m odd

$$\varphi_0(x) = \exp\{x\}; \tag{50a}$$

$$\varphi_k(x) = \exp\left\{\big(\Re(\lambda_k)\big)x\right\}\cos\left[\big(\Im(\lambda_k)\big)x\right],\ k \in \{1,2,\ldots,m\} \setminus \left\{\frac{m+1}{2}\right\}; \tag{50b}$$

$$\varphi_{m+1}(x) = \exp\{-x\}; \tag{50c}$$

$$\varphi_{m+1+k}(x) = \exp\left\{\big(\Re(\lambda_k)\big)x\right\}\sin\left[\big(\Im(\lambda_k)\big)x\right],\ k \in \{1,2,\ldots,m\} \setminus \left\{\frac{m+1}{2}\right\}. \tag{50d}$$

These vectors generate the subspace of $\mathcal{C}^m[0,1]$ of solutions of the differential equation (41). The general solution is given by the linear combination:

$$\begin{aligned} L_a(x) &= \sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m} \gamma_k(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\cos\left[\big(\Im(\lambda_k)\big)x\right] \\ &\quad + \sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m} \gamma_{m+1+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\sin\left[\big(\Im(\lambda_k)\big)x\right], \quad \text{for } m \text{ even}; \end{aligned} \tag{51}$$

$$\begin{aligned} L_a(x) &= \gamma_0(a)\exp\{x\} + \sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m} \gamma_k(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\cos\left[\big(\Im(\lambda_k)\big)x\right] + \gamma_{m+1}(a)\exp\{-x\} \\ &\quad + \sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m} \gamma_{m+1+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\sin\left[\big(\Im(\lambda_k)\big)x\right], \quad \text{for } m \text{ odd}; \end{aligned} \tag{52}$$

$$\begin{aligned} R_a(x) &= \sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m} \gamma_{2m+2+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\cos\left[\big(\Im(\lambda_k)\big)x\right] \\ &\quad + \sum_{\substack{k=0\\ k\neq\frac{m}{2}}}^{m} \gamma_{3m+3+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\sin\left[\big(\Im(\lambda_k)\big)x\right], \quad \text{for } m \text{ even}; \end{aligned} \tag{53}$$

$$\begin{aligned} R_a(x) &= \gamma_{2m+2}(a)\exp\{x\} + \sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m} \gamma_{2m+2+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\cos\left[\big(\Im(\lambda_k)\big)x\right] + \gamma_{3m+3}(a)\exp\{-x\} \\ &\quad + \sum_{\substack{k=1\\ k\neq\frac{m+1}{2}}}^{m} \gamma_{3m+3+k}(a)\exp\left\{\big(\Re(\lambda_k)\big)x\right\}\sin\left[\big(\Im(\lambda_k)\big)x\right], \quad \text{for } m \text{ odd}; \end{aligned} \tag{54}$$

where the coefficients $\gamma_k(a)$ are arbitrary constants that satisfy the boundary conditions (42)–(46). 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 (42)–(46) 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, by the linear independence of the base elements $\varphi_k(x)$, we obtain the uniqueness of the coefficients $\gamma_k(a)$.

(ii) Producing a representor $\psi_{\mathbb{a}}^{\mathrm{w}}$. Let us define the representor $\psi_{\mathbb{a}}^{\mathrm{w}}$ by setting

$$\psi_{\mathbb{a}}^{\mathrm{w}}(\mathbb{x}) = \prod_{i=1}^{q} \psi_{a_i}^{w_i}(x_i) \quad \text{for all } \mathbb{x} \in \mathcal{Q}^q, \tag{55}$$

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

$$\begin{aligned}
\left\langle \psi_{\mathbb{a}}^{\mathrm{w}}, f\right\rangle_{Sob,m} &= \left\langle \prod_{i=1}^{q} \psi_{a_i}^{w_i}, f\right\rangle_{Sob,m} = \sum_{|\boldsymbol{\alpha}|_\infty \le m} \int_{\mathcal{Q}^q} \frac{\partial^{\alpha_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{\alpha_q}} D^{\boldsymbol{\alpha}} f(\mathbb{x}) \mathrm{d}\mathbb{x} \\
&= \sum_{i_1,\ldots,i_q=0,\ldots,m} \int_{\mathcal{Q}^q} \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}\mathbb{x} \\
&= \sum_{i_1=0}^{m} \int_0^1 \frac{\partial^{i_1}\psi_{a_1}^{w_1}(x_1)}{\partial x_1^{i_1}} \left[\cdots \left[\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} . \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}x_q \right] \cdots \right] \mathrm{d}x_1.
\end{aligned} \tag{56}$$

According to Definition 1.3 and notation in Definition 1.1, we can rewrite the center-most bracket

$$\begin{aligned}
\sum_{i_q=0}^{m} \int_0^1 \frac{\partial^{i_q}\psi_{a_q}^{w_q}(x_q)}{\partial x_q^{i_q}} . \frac{\partial^{i_1,\ldots,i_q} f(\mathbb{x})}{\partial x_1^{i_1}\ldots\partial x_q^{i_q}} \mathrm{d}x_q &= \left\langle \psi_{a_q}^{w_q}, D^{(i_1,\ldots,i_{q-1})} f(x_1,\ldots,x_{i-1},\cdot)\right\rangle_{Sob,m} \\
&= D^{(i_1,\ldots,i_{q-1},w_q)} f(\mathbb{x}_{-q}, a_q).
\end{aligned} \tag{57}$$

Proceeding sequentially in the same way, we obtain that the value of the above expression is $D^{\mathrm{w}} f(\mathbb{a})$. □

Proof of Theorem 1.2. Existence and uniqueness of coefficients $\gamma_k(a)$ has already been proved in the proof of Theorem 1.1. Let us define

$$\Lambda_{a,I}^{(l)} := \begin{cases} L_a^{(l)}(0), & \text{for } I = L; \\ R_a^{(l)}(1), & \text{for } I = R; \\ L_a^{(l)}(a) - R_a^{(a)}(a), & \text{for } I = D. \end{cases} \tag{58}$$

From the boundary conditions (43)–(46), we easily see that

$$\sum_{k=i+1}^{m} (-1)^{k-i-1}\Lambda_{a,I}^{(2k-i-1)} = 0, \quad 0 \le i \le m-1,\, I \in \{L,R,D\}\,, \quad [i,I] \ne [0,D]\,; \tag{59}$$

$$\sum_{k=1}^{m}(-1)^{k-1}\Lambda_{a,D}^{(2k-1)} = 1. \tag{60}$$

For $m=1$ it follows from (59)–(60) that:

$$\Lambda_{a,I}^{(1)} = 0, \quad I \in \{L,R\}\,, \tag{61}$$

$$\Lambda_{a,D}^{(1)} = 1. \tag{62}$$

For $m=2$, we have from (59)–(60):

$$\Lambda_{a,I}^{(2)} = 0, \quad \forall I, \tag{63}$$

$$\Lambda_{a,I}^{(1)} - \Lambda_{a,I}^{(3)} = 0, \quad I \in \{L,R\}\,, \tag{64}$$

$$\Lambda_{a,D}^{(1)} - \Lambda_{a,D}^{(3)} = 1. \tag{65}$$

Let us now suppose that $m \ge 3$. We would like to prove the next important step:

$$\Lambda_{a,I}^{(m-j)} + (-1)^j\Lambda_{a,I}^{(m+j)} = 0, \quad j=0,\dots,m-2,\, \forall I, \tag{66}$$

$$\Lambda_{a,I}^{(1)} + (-1)^{m-1}\Lambda_{a,I}^{(2m-1)} = 0, \quad I \in \{L,R\}\,, \tag{67}$$

$$\Lambda_{a,D}^{(1)} + (-1)^{m-1}\Lambda_{a,D}^{(2m-1)} = 1, \tag{68}$$

where $j := m-i-1$. For $j=0$, we obtain $i=m-1$ and (59)–(60) implies

$$\Lambda_{a,I}^{(m)} = 0, \quad \forall I, \tag{69}$$

which is correct according to (66). Consider $j=1$ and thus $i=m-2$. In the same way we get:

$$\Lambda_{a,I}^{(m-1)} - \Lambda_{a,I}^{(m+1)} = 0, \quad \forall I. \tag{70}$$

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

$$\Lambda_{a,I}^{(m-2)} - \Lambda_{a,I}^{(m)} + \Lambda_{a,I}^{(m+2)} = 0, \quad \forall I, \tag{71}$$

and we can use (69). For $j=3$ and thus $i=m-4$ we have

$$\Lambda_{a,I}^{(m-3)} - \Lambda_{a,I}^{(m-1)} + \Lambda_{a,I}^{(m+1)} - \Lambda_{a,I}^{(m+3)} = 0, \quad \forall I, \tag{72}$$

where we can apply (70). We can continue in this way until $j=m-1$. The last step ensures the correctness of (67) in case that $I \in \{L,R\}$, eventually (68) if $I=D$ instead of (66).

To finish the proof, we only need to keep in mind (42). From (42), it follows that

$$\Lambda_{a,D}^{(j)} = 0, \quad j \in \{0,\dots,m-1\}. \tag{73}$$

According to (66) for $I=D$ and (68), we further see:

$$\Lambda_{a,D}^{(j)} = 0, \quad j \in \{m+1,\dots,2m-2\}; \tag{74}$$

$$\Lambda_{a,D}^{(2m-1)} = (-1)^{m-1}. \tag{75}$$

Altogether we have obtained the following system of $4m$ linear equations:

$$\Lambda_{a,L}^{(m-j)} + (-1)^j \Lambda_{a,L}^{(m+j)} = 0, \quad j=0,\dots,m-1, \tag{76}$$

$$\Lambda_{a,R}^{(m-j)} + (-1)^j \Lambda_{a,R}^{(m+j)} = 0, \quad j=0,\dots,m-1, \tag{77}$$

$$\Lambda_{a,D}^{(j)} = 0, \quad j=0,\dots,2m-2, \tag{78}$$

$$\Lambda_{a,D}^{(2m-1)} = (-1)^{m-1}, \tag{79}$$

which, after rewriting them using (58), (51)–(54) and (49a)–(50d), bring us to a close. □

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

$$M^\perp = \left\{ h \in \mathcal{H}^m : \langle \psi_{\mathbb{x}_i}, h \rangle_{Sob,m} = 0,\, i=1,\dots,n \right\}. \tag{80}$$

Representors exist by Theorem 1.1 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 $\mathbb{x}_1,\dots,\mathbb{x}_n$. Each $f \in \mathcal{H}^m$ can be written as

$$f = \sum_{j=1}^{n} c_j \psi_{\mathbb{x}_j} + h, \quad h \in M^\perp. \tag{81}$$

Then,

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

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

$$
\langle \psi_{\mathbb{x}_\bullet}, g \rangle_{Sob,m} = \Big(\langle \psi_{x_1}, g \rangle_{Sob,m}, \ldots, \langle \psi_{x_n}, g \rangle_{Sob,m} \Big)^\top. \tag{82}
$$

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. It follows directly from (37) and from Theorem 2.1. □

Proof of Theorem 2.3. The representor matrix is symmetric by Definition 2.1 since

$$
\Psi_{i,j} = \left\langle \psi_{\mathbb{x}_i}, \psi_{\mathbb{x}_j} \right\rangle_{Sob,m} = \left\langle \psi_{\mathbb{x}_j}, \psi_{\mathbb{x}_i} \right\rangle_{Sob,m} = \Psi_{j,i}, \tag{83}
$$

i.e., $\boldsymbol{\Psi} = \boldsymbol{\Psi}^\top$.

We give the proof of positive definiteness of the representor matrix only for one dimensional variable x. The extention into the multivariate case is straightforward, see Remark 2.1. For an arbitrary $\mathbb{c} \in \mathbb{R}^n$, we obtain

$$
\begin{aligned}
\mathbb{c}^\top \mathbf{\Psi} \mathbb{c} &= \sum_i c_i \sum_j \Psi_{ij} c_j = \sum_i \sum_j c_i \left\langle \psi_{x_i}, \psi_{x_j} \right\rangle_{Sob,m} c_j = \sum_i \sum_j \left\langle c_i \psi_{x_i}, c_j \psi_{x_j} \right\rangle_{Sob,m} \\
&= \left\langle \sum_i c_i \psi_{x_i}, \sum_j c_j \psi_{x_j} \right\rangle_{Sob,m} = \left\| \sum_i c_i \psi_{x_i} \right\|^2_{Sob,m} \geq 0.
\end{aligned} \tag{84}
$$

Hence $\mathbb{c}^\top \mathbf{\Psi} \mathbb{c} = 0$ iff $\sum_i c_i \psi_{x_i} = 0$ a.e.

For $x > x_i$, we define $\boldsymbol{\gamma}(x_i) = \left(\gamma_0, \ldots, \gamma_{\kappa-1}, \gamma_{\kappa+1}, \ldots, \gamma_{m+\kappa}, \gamma_{m+2+\kappa}, \ldots, \gamma_{2m+1}\right)^\top (x_i)$. Otherwise, $\boldsymbol{\gamma}(x_i) = \left(\gamma_{2m+2}, \ldots, \gamma_{2m+1+\kappa}, \gamma_{2m+3+\kappa}, \ldots, \gamma_{3m+2+\kappa}, \gamma_{3m+4+\kappa}, \ldots, \gamma_{4m+3}\right)^\top (x_i)$. Similarly, we will work with elements of the vector $\left[\{\mathbf{\Gamma}(x_i)\}^{-1}\right]_{\bullet,4m}$. According to (51)–(54), (49a)–(50d) and (11), we have

$$
\psi_{x_i}(x) = \boldsymbol{\gamma}(x_i)^\top \boldsymbol{\varphi}(x) = (-1)^{m-1} \left[\{\mathbf{\Gamma}(x_i)\}^{-1}\right]^\top_{\bullet,4m} \boldsymbol{\varphi}(x) \tag{85}
$$

where $\boldsymbol{\varphi}(x)$ is vector containing the base elements of the space of the solutions of the differential equation (41), i.e., $\varphi_k(x)$ (see (49a)–(50d)). From the linear independence of $\varphi_k(x)$ it follows that

$$
\begin{aligned}
\sum_i c_i \psi_{x_i} &= (-1)^{m-1} \sum_i c_i \left[\{\mathbf{\Gamma}(x_i)\}^{-1}\right]^\top_{\bullet,4m} \boldsymbol{\varphi} \\
&= (-1)^{m-1} \sum_i \sum_k c_i \left[\{\mathbf{\Gamma}(x_i)\}^{-1}\right]_{4m,k} \varphi_k = 0 \quad \text{a.e.} && (86) \\
&\Updownarrow \\
\varphi_k &= 0 \quad \text{a.e.} \quad k \in \{0, 1, \ldots, 2m+1\} \setminus \begin{cases} \left\{\frac{m}{2}, \frac{3m+2}{2}\right\} & m \text{ even}, \\ \left\{\frac{m+1}{2}, \frac{3m+3}{2}\right\} & m \text{ odd}; \end{cases} && (87) \\
&\Downarrow \\
\psi_{x_i} &= 0 \quad \text{a.e.} \quad i = 1, \ldots, n. && (88)
\end{aligned}
$$

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

Proof of Theorem 2.4. According to the Theorem 2.1, we want to minimize the function

$$
\mathcal{L}(\mathbb{c}) := \frac{1}{n} \left[\mathbb{Y} - \mathbf{\Psi}\mathbb{c}\right]^\top \mathbf{\Sigma}^{-1} \left[\mathbb{Y} - \mathbf{\Psi}\mathbb{c}\right] + \chi \mathbb{c}^\top \mathbf{\Psi} \mathbb{c}. \tag{89}
$$

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

$$
\frac{\partial}{\partial c_i} \mathcal{L}(\mathbb{c}) \stackrel{!}{=} 0, \quad i = 1, \ldots, n. \tag{90}
$$

Denoting $\boldsymbol{\Sigma}^{-1} =: (\phi_{ij})_{i,j=1}^{n,n}$, we can write:

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

and, hence,

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

Then we obtain our system of the normal equations

$$
\mathbb{c}^\top\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_{\bullet,i} + n\chi\boldsymbol{\Psi}_{\bullet,i}\right) = \mathbb{Y}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_{\bullet,i}, \quad i = 1,\ldots,n. \tag{91}
$$

□

Proof of Theorem 2.5. The solution of (20) 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 (20)—is the same as searching optimal $\mathbb{c}^*$ such that

$$
\mathbb{c}^* = \arg\min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} \le L \tag{92}
$$

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 multipliers on the condition $\mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} = L$ using the Lagrange function

$$
\mathcal{J}(\mathbb{c},\lambda) = \frac{1}{n}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1}\left[\mathbb{y} - \boldsymbol{\Psi}\mathbb{c}\right] + \chi\left(\mathbb{c}^\top\boldsymbol{\Psi}\mathbb{c} - L\right) \tag{93}
$$

and it provides a unique Lagrange multiplier χ. The term $-\chi L$ can be omitted as it does not depend on $\mathbb{c}$.

The quadratic form $\mathcal{J}(\cdot,\lambda)$ has to be positive definite according to 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 2.6. The proof is an easy application of Lagrange multipliers. □

Proof of Theorem 2.7. 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 2.6 say that there exists a unique $L > 0$ such that $\widehat{\mathbb{c}}$ is also a unique solution of optimizing problem

$$\widehat{\mathbb{c}} = \arg\min_{\mathbb{c}\in\mathbb{R}^n} \frac{1}{n} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\mathbb{Y} - \boldsymbol{\Psi}\mathbb{c}\right] \quad \text{s.t. } \mathbb{c}^\top \boldsymbol{\Psi}\mathbb{c} = L. \tag{94}$$

Let's define

$$\widetilde{\mathbf{f}}(\mathbb{x}) := \boldsymbol{\Xi}\mathbf{f}(\mathbb{x}), \tag{95}$$

$$\widetilde{\mathbb{Y}} := \boldsymbol{\Xi}\mathbb{Y}, \tag{96}$$

$$\widehat{\widetilde{\mathbb{c}}} := \arg\min_{\widetilde{\mathbb{c}}\in\mathbb{R}^n} \frac{1}{n} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t.} \quad \widetilde{\mathbb{c}}^\top \boldsymbol{\Psi}\boldsymbol{\Xi}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\Xi}^{-1}\boldsymbol{\Psi}\widetilde{\mathbb{c}} \leq L. \tag{97}$$

We can easily find out that

$$\widehat{\widetilde{\mathbb{c}}} = \boldsymbol{\Psi}^{-1}\boldsymbol{\Xi}\boldsymbol{\Psi}\widehat{\mathbb{c}} \tag{98}$$

and hence

$$\widehat{\widetilde{\mathbf{f}}}(\mathbb{x}) = \boldsymbol{\Xi}\widehat{\widetilde{\mathbb{c}}}. \tag{99}$$

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

$$\widehat{\widetilde{\mathbb{c}}} = \arg\min_{\widetilde{\mathbb{c}}\in\mathbb{R}^n} \frac{1}{n} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t. } \widetilde{\mathbb{c}}^\top \boldsymbol{\Psi}\widetilde{\mathbb{c}} = \widetilde{L} \tag{100}$$

and hence this $\widehat{\widetilde{\mathbb{c}}}$ has to be a unique solution of the optimizing problem

$$\widehat{\widetilde{\mathbb{c}}} = \arg\min_{\widetilde{\mathbb{c}}\in\mathbb{R}^n} \frac{1}{n} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right]^\top \boldsymbol{\Sigma}^{-1} \left[\widetilde{\mathbb{Y}} - \boldsymbol{\Psi}\widetilde{\mathbb{c}}\right] \quad \text{s.t. } \widetilde{\mathbb{c}}^\top \boldsymbol{\Psi}\widetilde{\mathbb{c}} \leq \widetilde{L} \tag{101}$$

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

Now we think of model

$$\widetilde{Y}_i = \widetilde{f}(\mathbb{x}_i) + \widetilde{\varepsilon}_i, \quad \widetilde{\varepsilon}_i \sim i.i.d., \quad i = 1, \ldots, n \tag{102}$$

with least-squares estimator $\widehat{\widetilde{f}}$. As in the proof of Lemma 1 in Yatchew and Bos (1997), using Kolmogorov and Tihomirov (1959), it can be shown that there exists $A > 0$ such that for $\delta > 0$, we have $\log N(\delta; \mathscr{F}) < A\delta^{-q/m}$, where $N(\delta; \mathscr{F})$ denotes the minimum number of balls of radius δ in sup-norm required to cover the set of functions $\mathscr{F}$. Consequently, applying Van de Geer (1990, Lemma 3.5), we obtain that there exist positive constants C_0, K_0 such that for all $K > K_0$

$$\mathsf{P}\left[\sup_{\|g\|^2_{Sob,m}\leq\widetilde{L}} \frac{\sqrt{n}\left|-\frac{2}{n}\sum_{i=1}^n \widetilde{\varepsilon}_i\left(\widetilde{f}(x_i)-g(x_i)\right)\right|}{\left(\frac{1}{n}\sum_{i=1}^n\left(\widetilde{f}(x_i)-g(x_i)\right)^2\right)^{\frac{1}{2}-\frac{q}{4m}}} \geq KA^{1/2}\right] \leq \exp\left\{-C_0K^2\right\}. \tag{103}$$

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

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n\left[\widetilde{Y}_i - \widehat{\widetilde{f}}(x_i)\right]^2 &\leq \frac{1}{n}\sum_{i=1}^n\left[\widetilde{Y}_i - g(x_i)\right]^2, \quad g \in \widetilde{\mathscr{F}} && (104)\\
\frac{1}{n}\sum_{i=1}^n\left[\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right) + \widetilde{\varepsilon}_i\right]^2 &\leq \frac{1}{n}\sum_{i=1}^n\left[\left(\widetilde{f}(x_i) - g(x_i)\right) + \widetilde{\varepsilon}_i\right]^2, \quad g \in \widetilde{\mathscr{F}} &&\\
&\Downarrow \quad \text{realize that } \widetilde{f} \in \widetilde{\mathscr{F}} &&\\
\frac{1}{n}\sum_{i=1}^n\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 &\leq -\frac{2}{n}\sum_{i=1}^n\widetilde{\varepsilon}_i\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right). && (105)
\end{aligned}$$

Now combine (103) and (105) to obtain the result that $\forall K > K_0$

$$\mathsf{P}\left[\frac{1}{n}\sum_{i=1}^n\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 \geq \left(\frac{K^2A}{n}\right)^{\frac{2m}{2m+q}}\right] \leq \exp\left\{-C_0K^2\right\}. \tag{106}$$

Thus

$$\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] = \frac{1}{n}\sum_{i=1}^n\left(\widetilde{f}(x_i) - \widehat{\widetilde{f}}(x_i)\right)^2 = \mathcal{O}_{\mathsf{P}}\left(n^{-\frac{2m}{2m+q}}\right), \quad n \to \infty. \tag{107}$$

□

Lemma A.1. *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 $f_n \xrightarrow[n\to\infty]{\mathsf{L}_1} 0$ then $\|f_n\|_{\infty,[a,b]} := \sup_{x\in[a,b]}|f_n(x)| \xrightarrow[n\to\infty]{} 0$.*

Proof of Lemma A.1. Suppose that

$$\exists\epsilon > 0 \quad \forall n_0 \in \mathbb{N} \quad \exists n \geq n_0 \quad \exists x \in [a, b] \quad f_n(x) \geq \epsilon. \tag{108}$$

Then according to Lipschitz property of each $f_n \geq 0$ we have for fixed ϵ, n_0, n and $x \in [a,b]$ that

$$\begin{aligned}
\|f_n\|_{\mathsf{L}_1[a,b]} &= \int_a^b f_n(t)\mathrm{d}t \\
&\geq \min\left\{ \frac{f_n(x)}{2}(x-a) + \frac{f_n(x)}{2}(b-x), \frac{f_n(x)}{2}(x-a) + \frac{f_n(x)}{2}\frac{f_n(x)}{T}, \right. \\
&\quad \left. \frac{f_n(x)}{2}\frac{f_n(x)}{T} + \frac{f_n(x)}{2}(b-x), \frac{f_n(x)}{2}\frac{f_n(x)}{T} + \frac{f_n(x)}{2}\frac{f_n(x)}{T} \right\} \\
&\geq \min\left\{ \frac{\epsilon}{2}(b-a), \frac{\epsilon}{2}(x-a) + \frac{\epsilon^2}{2T}, \frac{\epsilon^2}{2T} + \frac{\epsilon}{2}(b-x), \frac{\epsilon^2}{T} \right\} =: K > 0.
\end{aligned} \tag{109}$$

But K is a positive constant which does not depend on n and its existence would contradict the assumptions of this lemma, i.e., $\forall \delta > 0 \quad \exists n_1 \in \mathbb{N} \quad \forall n \geq n_1 \quad \|f_n\|_{\mathsf{L}_1[a,b]} < \delta$. □

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

(i) $s = 0$**.** The covariance matrix $\boldsymbol{\Sigma}$ is symmetric and positive definite with equibounded eigenvalues for all n. Hence it can be decomposed using Schur decomposition: $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Upsilon}\boldsymbol{\Gamma}^\top$, where $\boldsymbol{\Gamma}$ is orthogonal, $\boldsymbol{\Upsilon}$ is diagonal (with eigenvalues on this diagonal) such that $0 < \Upsilon_{ii} \leq \vartheta$, $i = 1, \ldots, n$, $\forall n$. Hence $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Gamma} diag\left\{\Upsilon_1^{-1}, \ldots, \Upsilon_n^{-1}\right\}\boldsymbol{\Gamma}^\top$. Then

$$\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] \geq \frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Gamma}\vartheta^{-1}\mathbf{I}\boldsymbol{\Gamma}^\top\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] = \frac{1}{n\vartheta}\sum_{i=1}^n \left[\hat{f}(x_i) - f(x_i)\right]^2 \tag{110}$$

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 $\left\|t\right\|_{Sob,m}^2 \leq L$ it holds that

$$\|t'\|_{\mathsf{L}_2[a,b]} \leq \|t\|_{Sob,1} \leq \|t\|_{Sob,m} \leq \sqrt{L}. \tag{111}$$

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

$$|t(\xi) - t(\zeta)| < T\left|\xi - \zeta\right|, \quad \xi, \zeta \in [a,b]. \tag{112}$$

We easily see

$$\begin{aligned}
\frac{|h_n(\xi) - h_n(\zeta)|}{|\xi - \zeta|} &= \frac{\left|\left|\hat{f}(\xi) - f(\xi)\right| - \left|\hat{f}(\zeta) - f(\zeta)\right|\right|}{|\xi - \zeta|} \leq \frac{\left|\left[\hat{f}(\xi) - f(\xi)\right] - \left[\hat{f}(\zeta) - f(\zeta)\right]\right|}{|\xi - \zeta|} \\
&\leq \frac{\left|\hat{f}(\xi) - \hat{f}(\zeta)\right| + \left|f(\xi) - f(\zeta)\right|}{|\xi - \zeta|} < 2T, \quad \xi, \zeta \in [a,b].
\end{aligned} \tag{113}$$

Since h_n is T-Lipschitz function for all n and

$$\|h_n\|_{\mathsf{L}_2[a,b]} = \left\|\hat{f} - f\right\|_{\mathsf{L}_2[a,b]} \leq \left\|\hat{f} - f\right\|_{Sob,1} \leq \left\|\hat{f} - f\right\|_{Sob,m} \leq \left\|\hat{f}\right\|_{Sob,m} + \left\|f\right\|_{Sob,m} \leq 2\sqrt{L}, \quad \forall n, \tag{114}$$

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

$$\|h_n\|_{\infty,[a,b]} \leq M > 0, \quad \forall n. \tag{115}$$

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

$$\frac{\left|h_n^2(\xi) - h_n^2(\zeta)\right|}{|\xi - \zeta|} = \frac{|h_n(\xi) - h_n(\zeta)|}{|\xi - \zeta|}\left[h_n(\xi) + h_n(\zeta)\right] \leq T \times 2\left\|h_n\right\|_{\infty,[a,b]} = 2MT =: U > 0, \quad \forall n. \tag{116}$$

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 that

$$\begin{aligned}\int_a^b h_n^2(u)\mathrm{d}u &\leq \sum_{i=1}^{n-1} \frac{x_{i+1} - x_i}{2}\left\{h_n^2(x_i) + \left[h_n^2(x_i) + U(x_{i+1} - x_i)\right]\right\} \leq \frac{1}{2n}\left[2\sum_{i=1}^{n-1} h_n^2(x_i) + U(b-a)\right] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} h_n^2(x_i) + \frac{U(b-a)}{2n}.\end{aligned} \tag{117}$$

According to Theorem 2.7

$$\forall \epsilon > 0 \quad \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] > \epsilon\right\} \xrightarrow[n\to\infty]{} 0, \tag{118}$$

so it means

$$\forall \epsilon > 0 \quad \forall \delta > 0 \quad \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0 \quad \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right]^\top \mathbf{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x}) - \mathbf{f}(\mathbb{x})\right] > \epsilon\right\} < \delta. \tag{119}$$

Let's fix an arbitrary $\epsilon > 0$ and $\delta > 0$. Next, we fix

$$n_0 := \left\lceil \frac{U}{\epsilon^2} \right\rceil \tag{120}$$

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

$$\begin{aligned}
\delta > \mathsf{P}\left\{\frac{1}{n}\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right]^\top \boldsymbol{\Sigma}^{-1}\left[\widehat{\mathbf{f}}(\mathbb{x})-\mathbf{f}(\mathbb{x})\right] > \frac{\epsilon^2(b-a)}{2\vartheta}\right\} & \quad \text{by (119)}\\
\geq \mathsf{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{f}(x_i)-f(x_i)\right]^2 > \frac{\epsilon^2(b-a)}{2}\right\} & \quad \text{by (110)}\\
\geq \mathsf{P}\left\{\|h_n\|^2_{\mathsf{L}_2[a,b]} > \underbrace{\frac{\epsilon^2(b-a)}{2}+\frac{U(b-a)}{2n}}_{\tilde{\epsilon}}\right\} & \quad \text{by (117)}\\
\geq \mathsf{P}\left\{\|h_n\|_{\mathsf{L}_1[a,b]} > \frac{\sqrt{\tilde{\epsilon}}}{\|1\|_{\mathsf{L}_2[a,b]}}\right\} & \quad \text{by Cauchy-Schwarz inequality}\\
\geq \mathsf{P}\left\{\|h_n\|_{\mathsf{L}_1[a,b]} > \epsilon\right\} & \quad \text{by (120).}
\end{aligned} \tag{121}$$

Thus

$$\|h_n\|_{\mathsf{L}_1[a,b]} \xrightarrow[n\to\infty]{\mathsf{P}} 0. \tag{122}$$

According to Lemma A.1 and the fact that the almost sure convergence implies convergence in probability, we have

$$\sup_{x\in[a,b]}\left|\widehat{f}(x)-f(x)\right| \xrightarrow[n\to\infty]{\mathsf{P}} 0. \tag{123}$$

(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 Yatchew and Bos (1997, Theorem 2.3), all functions in the estimating set have derivatives up to order $m-1$ uniformly bounded in sup-norm. Then, all the g_n'' are also bounded in sup-norm ($m \geq 3$) and this implies the uniform boundedness of g_n'':

$$\exists M > 0 \quad \forall n \in \mathbb{N} \quad \|g_n''\|_{\infty,[a,b]} < M. \tag{124}$$

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

$$g_n'(c_n) = g_n'(d_n) = \tilde{\epsilon} \quad \& \quad g_n'(\xi) > \tilde{\epsilon},\, \xi \in (c_n, d_n) \tag{125}$$

because g_n' is continuous on $[c_n, d_n]$ (drawing a picture is helpful). If such $[c_n, d_n]$ does not exist, the proof is finished.

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

$$|\tilde{\epsilon}(d_n - c_n)| \leq \left|\int_{c_n}^{d_n} g_n'(\xi)\mathrm{d}\xi\right| = |g_n(d_n) - g_n(c_n)| \leq 2\epsilon^2 \tag{126}$$

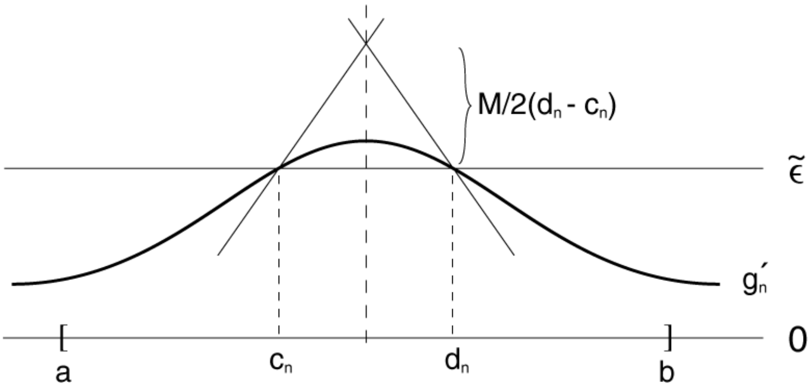

Figure 6: Uniform convergence of g_n'.

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

$$|d_n - c_n| \leq \frac{2\epsilon}{M}. \tag{127}$$

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

$$|g_n'(x)| \leq \left|\tilde{\epsilon} + M\frac{d_n - c_n}{2}\right| \leq M\epsilon + M\frac{\epsilon}{M} \leq \epsilon(M+1). \tag{128}$$

We can continue in this way finitely times (formally we can proceed 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 sup-norm. Finally, we have to realize that convergence almost sure implies convergence in probability and each convergent sequence in probability has a subsequence that converges almost sure. □

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

$$f = \sum_{\{i \mid n_i \geq 1\}} c_i \psi_{x_i} + h_f, \quad h_f \in \{span\{\psi_{\mathtt{x}_i} : n_i \geq 1\}\}^{\perp}, \tag{129}$$

$$g = \sum_{\{j \mid m_j \geq 1\}} d_j \phi_{x_j} + h_g, \quad h_g \in \left\{span\left\{\phi_{\mathtt{x}_j} : m_j \geq 1\right\}\right\}^{\perp}. \tag{130}$$

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

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

We can proceed in the same way also for $n < \iota \leq n + m$.

Finally, it remains to rewrite the constraints using (6) from Theorem 1.1:

$$f'(x_\iota) = \left\langle \psi_{x_\iota}, \sum_{\{i \,|\, n_i \geq 1\}} c_i \psi'_{x_i} + h_f \right\rangle_{Sob,m} = \left[\mathbf{\Psi}^{(1)} \mathbb{c}\right]_\iota \quad \forall \iota : n_\iota \geq 1. \tag{131}$$

Similarly, we obtain

$$g'(x_\iota) = \left[\mathbf{\Phi}^{(1)} \mathbb{d}\right]_\iota \quad \forall \iota : m_\iota \geq 1, \tag{132}$$

$$f''(x_\iota) = \left[\mathbf{\Psi}^{(2)} \mathbb{c}\right]_\iota \quad \forall \iota : n_\iota \geq 1, \tag{133}$$

$$g''(x_\iota) = \left[\mathbf{\Phi}^{(2)} \mathbb{d}\right]_\iota \quad \forall \iota : m_\iota \geq 1. \tag{134}$$

□

References

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

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

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

Breeden, Douglas T., and Robert H. Litzenberger. (1978). "Prices of state-contingent claims implicit in option prices." *Journal of Business* 51, 621–651.

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

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

Härdle, Wolfgang, and Zdeněk Hlávka. (2006). "Dynamics of state price densities." SFB 649 Discussion Paper 2005-021, Humboldt-Universität zu Berlin; submitted to Journal of Econometrics.

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

Kolmogorov, Andrei N., and Vladimir M. Tihomirov. (1959). "ε-entropy and ε-capacity of Sets in Function Spaces." *Uspehi Matematičeskih Nauk* 14(2), 3–86. English transl.: *A*mer. Math. Soc. Transl.: Series 2 (1961) 17, 277–364.

Pešta, Michal. (2006). Isotonic Regression in Sobolev Spaces. Master's thesis Charles University in Prague Czech Republic: .

Rao, C. Radhakrishna. (1973). *Linear statistical inference and its applications. 2nd ed.* John Wiley & Sons.

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

Yatchew, Adonis, and Len Bos. (1997). "Nonparametric Least Squares Estimation and Testing of Economic Models." *Journal of Quantitative Economics* 13, 81–131.

Yatchew, Adonis, and Wolfgang Härdle. (2006). "Nonparametric state price density estimation using constrained least squares and the bootstrap." *Journal of Econometrics* 133(2), 579–599.