# Isotonic Regression in Sobolev Spaces


Charles University Prague
faculty of mathematics and physics

## Michal Pešta

Faculty of Mathematics and Physics, CHARLES UNIVERSITY IN PRAGUE, Department of Probability and Mathematical Statistics

michal.pesta@matfyz.cz — http://michal.pesta.matfyz.cz/



Robust 2006


## Abstract


We propose a class of nonparametric estimators for regression models based on least squares over sufficiently smooth sets of functions. These estimators permit the imposition of additional monotonicity and concavity constraints.

Estimation takes place over balls of functions which are elements of suitable Sobolev space. The Sobolev spaces are special types of Hilbert spaces that facilitate calculation of least square projection. The Hilbertness is allowing us to take projections and hence to decompose spaces into mutually orthogonal complements. We assemble and prove necessary preliminaries and theorems for statistical regression in these spaces. Then we transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimization problem.

In regression in Sobolev spaces, we have demanded only smoothness constraint on regression function $f \in \mathscr{F} = \left\{ f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|^2_{Sob,m} \leq L \right\}$. Now, our estimators should satisfy additional constraints. We therefore focus on the imposition of additional constraint —isotonia— on nonparametric regression estimation and testing of this constraint. Two basic types of isotonia (non-negativity or non-positivity of $n$-th derivative of regression curve) are monotonicity and concavity so we will concentrate mostly on them. We would like to estimate subject to $f \in \widetilde{\mathscr{F}} \subseteq \mathscr{F}$ where $\widetilde{\mathscr{F}}$ combines smoothness with further functional properties and to test $H_0 : f_0 \in \widetilde{\mathscr{F}}$.

We prove that balls of functions in Sobolev space are bounded and have bounded higher order derivatives. It permits us to estimate over rich set of functions with sufficiently low metric entropy and apply laws of large numbers and central limit theorems results. We also describe bootstrap techniques for isotonic regression in Sobolev spaces.


## 1 Sobolev Representors

A connected Lebesgue-measurable (open or closed) subset $\Omega$ of Euclidean space $\mathbb{R}^q$ with non-empty interior is called a domain. Consider a real-valued function on a given domain that is Lebesgue-measurable. Simply $f : \Omega \to \mathbb{R}$, $\Omega \in \mathfrak{M}_q(\lambda_q)$. We define the Lebesgue spaces by $L_p(\Omega) := \left\{ f : \|f\|_{L_p(\Omega)} < \infty \right\}$, $1 \leq p \leq \infty$, where $\|\cdot\|_{L_p(\Omega)}$ denotes $p$-th Lebesgue norm.

**Corollary 1.1.** *$L_p(\Omega)$, $1 \leq p \leq \infty$ is a Banach space. (See [Luk03].)*

We consider function $f : \Omega \to \mathbb{R}$ and denote by $D^{\vec{\alpha}} f(\mathbb{x}) := \frac{\partial^{|\vec{\alpha}|_1} f(\mathbb{x})}{\partial x_1^{\alpha_1} \ldots \partial x_q^{\alpha_q}}$ its partial derivatives of order $|\vec{\alpha}|_1$ for $\mathbb{x} \in \text{int}(\Omega) (\equiv \Omega^\circ := \overline{\Omega} \setminus \partial\Omega)$, where $\vec{\alpha} = (\alpha_1, \ldots, \alpha_q)^T \in \mathbb{N}_0^q$ is a multiindex of modulus $|\vec{\alpha}|_1 = \sum_{i=1}^q \alpha_i$. We define $\mathcal{C}^m(\Omega)$ space of $m$-times continuously differentiable scalar functions upon bounded domain $\Omega$.

**Definition 1.1** (Sobolev norm). Let $f \in \mathcal{C}^m(\Omega) \cap L_p(\Omega)$. We introduce Sobolev norm $\|f\|_{p,Sob,m} := \left\{ \sum_{|\vec{\alpha}|_\infty \leq m} \int_\Omega \left| D^{\vec{\alpha}} f(\mathbb{x}) \right|^p d\mathbb{x} \right\}^{1/p}$.

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

**Definition 1.3** (Sobolev inner product). Let $f, g \in \mathcal{W}_2^m(\Omega)$. We introduce Sobolev inner product $\langle \cdot, \cdot \rangle_{Sob,m}$:

$$\langle f, g \rangle_{Sob,m} := \sum_{|\vec{\alpha}|_\infty \leq m} \int_\Omega D^{\vec{\alpha}} f(\mathbb{x}) D^{\vec{\alpha}} g(\mathbb{x}) d\mathbb{x}. \quad (1)$$

For simplicity we denote Sobolev norm $\|\cdot\|_{2,Sob,m} := \|\cdot\|_{Sob,m}$, Sobolev space $\mathcal{H}^m(\Omega) := \mathcal{W}_2^m(\Omega)$ and $\mathcal{Q}^q$ closed unit cube in $\mathbb{R}^q$.

**Theorem 1.2** (Hilbert space). *$\mathcal{H}^m(\Omega)$ is a Hilbert space. (See [Luk03].)*

**Theorem 1.3** (Representors in Sobolev space). *For all $f \in \mathcal{H}^m(\mathcal{Q}^q)$, $\mathbb{a} \in \mathcal{Q}^q$ and $\mathbb{w} \in \mathbb{N}_0^q$, $|\mathbb{w}|_\infty \leq m - 1$, there exists a function $\psi_{\mathbb{a}}^{\mathbb{w}}(\mathbb{x}) \in \mathcal{H}^m(\mathcal{Q}^q)$, s.t.*

$$\left\langle \psi_{\mathbb{a}}^{\mathbb{w}}, f \right\rangle_{Sob,m} = D^{\mathbb{w}} f(\mathbb{a}). \quad (2)$$

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

$$\frac{\partial^{w_i} f(\mathbb{a})}{dx_i^{w_i}} = \left\langle \psi_{a_i}^{w_i}, f(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_q) \right\rangle_{Sob,m}$$
$$= \sum_{\alpha=0}^m \int_{\mathcal{Q}^1} \frac{d^\alpha \psi_{a_i}^{w_i}(x_i)}{dx_i^\alpha} \frac{d^\alpha f(\mathbb{x})}{dx_i^\alpha} dx_i. \quad (3)$$

**Theorem 1.4.** *The embedding $\mathcal{H}^m(\mathcal{Q}^q) \hookrightarrow \mathcal{C}^{m-1}(\mathcal{Q}^q)$ is compact. (See [BY97].)*

Let's define $\mathcal{K} := \{0, 1, \ldots, 2m+1\} \setminus \{\kappa, m+1+\kappa\}$, where $\kappa := \begin{cases} \frac{m}{2} & 2 \mid m, \\ \frac{m+1}{2}, & 2 \nmid m \end{cases}$. From proof of Representors Theorem 1.3 we easily obtain that representor $\psi_a \in \mathcal{H}^m[0,1]$ s.t. $\langle \psi_a, f \rangle_{Sob,m} = f(a)$ for all $f \in \mathcal{H}^m[0,1]$ will be of the form $\psi_a(x) = \begin{cases} L_a(x) & 0 \leq x \leq a, \\ R_a(x) & a \leq x \leq 1, \end{cases}$ where

$$L_a(x) = \sum_{k \in \mathcal{K}} \gamma_k \varphi_k(x) \text{ and } R_a(x) = \sum_{k \in \mathcal{K}} \gamma_{2m+2+k} \varphi_k(x). \quad (4)$$

We also determine $\varphi_k(x)$ for <u>m even</u>

$$\varphi_k(x) = \exp\left\{ \left(Re(\lambda_k)\right) x \right\} \cos\left[ \left(Im(\lambda_k)\right) x \right], \quad k \in \{0\} \cup \mathcal{M}; \quad (5a)$$

$$\varphi_{m+1+k}(x) = \exp\left\{ \left(Re(\lambda_k)\right) x \right\} \sin\left[ \left(Im(\lambda_k)\right) x \right], \quad k \in \{0\} \cup \mathcal{M}; \quad (5b)$$

and for <u>m odd</u>

$$\varphi_0(x) = \exp\{x\}; \quad (6a)$$

$$\varphi_k(x) = \exp\left\{ \left(Re(\lambda_k)\right) x \right\} \cos\left[ \left(Im(\lambda_k)\right) x \right], \quad k \in \mathcal{M}; \quad (6b)$$

$$\varphi_{m+1}(x) = \exp\{-x\}; \quad (6c)$$

$$\varphi_{m+1+k}(x) = \exp\left\{ \left(Re(\lambda_k)\right) x \right\} \sin\left[ \left(Im(\lambda_k)\right) x \right], \quad k \in \mathcal{M}; \quad (6d)$$

where

$$\lambda_k = e^{i\theta_k}, \quad \theta_k \in \begin{cases} \frac{(2k+1)\pi}{2m+2}, & 2 \mid m, \ k \in \mathcal{K}, \\ \frac{k\pi}{m+1}, & 2 \nmid m, \ k \in \mathcal{K}. \end{cases} \quad (7)$$

and $\mathcal{M} := \{1, 2, \ldots, m\} \setminus \{\kappa\}$.

**Theorem 1.5** (Obtaining coefficients $\gamma_k$s). *Coefficients $\gamma_k$ of representor $\psi_a(x)$ are unique solution of $4m \times 4m$ system of linear equations*

$$\sum_{k \in \mathcal{K}} \gamma_k \left( \varphi_k^{(m-j)}(0) + (-1)^j \varphi_k^{(m+j)}(0) \right) = 0, \quad j = 0, \ldots, m-1; \quad (8a)$$

$$\sum_{k \in \mathcal{K}} \gamma_{2m+2+k} \left( \varphi_k^{(m-j)}(1) + (-1)^j \varphi_k^{(m+j)}(1) \right) = 0, \quad j = 0, \ldots, m-1; \quad (8b)$$

$$\sum_{k \in \mathcal{K}} (\gamma_k - \gamma_{2m+2+k}) \varphi_k^{(j)}(a) = 0, \quad j = 0, \ldots, 2m-2; \quad (8c)$$

$$\sum_{k \in \mathcal{K}} (\gamma_k - \gamma_{2m+2+k}) \varphi_k^{(2m-1)}(a) = (-1)^{(m-1)}; \quad (8d)$$

*where $\mathcal{K}$ is defined above and $\varphi_k$ in (5–6).*

## 2 Least Squares

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

$$Y_i = f_0(\mathbb{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n \quad (9)$$

with these assumptions:

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

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

iii) $f_0 \in \mathscr{F}$, where $\mathscr{F}$ is a family of functions in the Sobolev space $\mathcal{H}^m(\mathcal{Q}^q)$ from $\mathbb{R}^n$ to $\mathbb{R}^1$, $m > \frac{q}{2}$, $\mathscr{F} = \left\{ f \in \mathcal{H}^m(\mathcal{Q}^q) : \|f\|^2_{Sob,m} \leq L \right\}$.

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

- $\min_{f \in \mathcal{H}^m} \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i)]^2$ s.t. $\|f\|^2_{Sob,m} \leq L$,
- $\min_{f \in \mathcal{H}^m} \left\{ \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i)]^2 + \chi \|f\|^2_{Sob,m} \right\}$.

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

**Definition 2.2** (Representor Matrix). Let $\psi_{x_1}, \ldots, \psi_{x_n}$ be the representors for function evaluation at $x_1, \ldots, x_n$ respectively, i.e. $\langle \psi_{x_i}, f \rangle_{Sob,m} = f(x_i)$ for all $f \in \mathcal{H}^m$, $i = 1, \ldots, n$. Let $\mathbf{\Psi}$ be the $n \times n$ representor matrix whose columns (and rows) equal the representors evaluated at $x_1, \ldots, x_n$; i.e. $\Psi_{i,j} = \langle \psi_{x_i}, \psi_{x_j} \rangle_{Sob,m} = \psi_{x_i}(x_j) = \psi_{x_j}(x_i)$.

**Theorem 2.1** (Infinite to Finite). *Let $\mathbb{y} = (y_1, \ldots, y_n)'$ and define*

$$\hat{\sigma}^2 = \min_{f \in \mathcal{H}^m} \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i)]^2 \text{ s.t. } \|f\|^2_{Sob,m} \leq L, \quad (10)$$

$$s^2 = \min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}] \text{ s.t. } \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad (11)$$

*where $\mathbb{c}$ is a $n \times 1$ vector and $\mathbf{\Psi}$ is the representor matrix. Then $\hat{\sigma}^2 = s^2$. Furthermore, there exists a solution of optimizing problem of the form $\widehat{f} = \sum_{i=1}^n \hat{c}_i \psi_{x_i}$, where $\widehat{\mathbb{c}} = (\hat{c}_1, \ldots, \hat{c}_n)'$ solves $\hat{\sigma}^2$. The estimator $\widehat{f}$ is unique a.s.*

**Lemma 2.2** (Symmetry of Representor Matrix). *Representor Matrix is symmetric.*

**Theorem 2.3** (Positive Definitness of Representor Matrix). *Representor Matrix is positive definite.*

**Theorem 2.4** (Asymptotic Behaviour of Finite Optimizing Solution). *Let $f$ satisfy $s^2 = \min \frac{1}{n} \sum_i^n (y_i - f(x_i))^2$ s.t. $f \in \mathscr{F}$. Suppose $f_0 \in \mathscr{F}$, then:*

*a) $s^2 \xrightarrow{a.s.} \sigma^2_{\varepsilon,0}$, $n \to \infty$;*

*b) $\frac{1}{n} \sum_i^n \left( \hat{f}(x_i) - f_0(x_i) \right)^2 = O_P(n^{-\eta})$ where $\eta = \frac{2m}{2m+q}$ and $n \to \infty$;*

*c) $n^{1/2} \left[ s^2 - \sigma^2_{\varepsilon,0} \right] \xrightarrow{\mathcal{D}} N\left(0, Var(\varepsilon^2)\right)$, $n \to \infty$.*

$\mathrm{Var}(\varepsilon^2)$ may be estimated consistently using fourth order moments of the estimated residuals $\hat{\varepsilon}_i = y_i - \hat{f}(x_i)$.

**Theorem 2.5** (Optimizing with Constraint). *Optimizing problem with constraint*

$$\min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}] \quad s.t. \ \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad (12)$$

*where $\mathbf{\Psi} > 0$ is a symmetric $n \times n$ matrix, $\mathbb{y}$ is an $n \times 1$ vector of constants and $L > 0$; has a solution $\widehat{\mathbb{c}} = \mathbf{\Phi}\widehat{\mathbb{d}}$, where $\mathbf{\Phi}$ is an orthogonal $n \times n$ matrix from Schur decomposition $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}'$, where*

$$\mathbf{\Lambda} = diag\{\lambda_1, \ldots, \lambda_n\}, \quad (13)$$
$$\lambda_i > 0, \quad i = 1, \ldots, n, \quad (14)$$
$$\mathbf{I} = \mathbf{\Phi}'\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}' \quad (15)$$

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

$$\min_{\mathbb{d} \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^n (\lambda_i d_i - z_i)^2 \quad s.t. \ \sum_{i=1}^n \lambda_i d_i^2 \leq L \quad (16)$$

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

<u>Multiple observations :</u> (See [HY03].)

- e.g. option price data often consist of multiple observations at a finite vector of strike prices
- let $\mathbb{X} = (X_1, \ldots, X_k)'$ be the vector of k distinct strike prices
- assume that the vector $\mathbb{X}$ is in increasing order
- let $\sigma^2(X_1), \ldots, \sigma^2(X_k)$ be the residual variances at each of the distinct strike prices

Let $\mathbf{\Delta}$ be the $n \times k$ matrix such that $\mathbf{\Delta}_{ij} := \begin{cases} 1 & \text{if } x_i = X_j \\ 0 & \text{otherwise} \end{cases}$.

We may now rewrite our infinite optimizing problem as

$$\min_{f \in \mathcal{H}^m} \frac{1}{n} [\mathbb{y} - \mathbf{\Delta} f(\mathbb{x})]' \mathbf{\Sigma}^{-1} [\mathbb{y} - \mathbf{\Delta} f(\mathbb{x})] \quad \text{s.t. } \|f\|^2_{Sob,m} \leq L \quad (17)$$

Nothing that the representor matrix $\mathbf{\Psi}$ is in this case $k \times k$, the analogue to finite quadratic optimizing becomes:

$$\min_{\mathbb{c} \in \mathbb{R}^k} \frac{1}{n} [\mathbb{y} - \mathbf{\Delta}\mathbf{\Psi}\mathbb{c}]' \mathbf{\Sigma}^{-1} [\mathbb{y} - \mathbf{\Delta}\mathbf{\Psi}\mathbb{c}] \quad \text{s.t. } \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad (18)$$

**Theorem 2.6** (Extentions to the Multi-Equation Settings Part). *Let $\mathbf{\Xi}$ be a positive definite matrix and define*

$$\hat{\sigma}^2 = \min_{f \in \mathcal{H}^m} \frac{1}{n} \sum_{i=1}^n [\mathbb{y}_i - f(\mathbb{x}_i)]' \mathbf{\Xi} [\mathbb{y}_i - f(\mathbb{x}_i)] \quad s.t. \ \|f_j\|^2_{Sob,m} \leq L_j, \quad j = 1, \ldots, p \quad (19a)$$

$$s^2 = \min_{\mathbb{C} \in \mathbb{R}^{n \times p}} \frac{1}{n} \sum_{i=1}^n \left[ \mathbb{y}_i - \mathbb{C}' \begin{pmatrix} \mathbf{\Psi}_{i1} \\ \vdots \\ \mathbf{\Psi}_{in} \end{pmatrix} \right]' \mathbf{\Xi} \left[ \mathbb{y}_i - \mathbb{C}' \begin{pmatrix} \mathbf{\Psi}_{i1} \\ \vdots \\ \mathbf{\Psi}_{in} \end{pmatrix} \right]$$
$$s.t. \ (C_{1i}, \ldots, C_{1n}) \mathbf{\Psi} \begin{pmatrix} C_{1i} \\ \vdots \\ C_{1n} \end{pmatrix} \leq L_j, \quad j = 1, \ldots, p \quad (19b)$$

*where $\mathbb{C}$ is a $n \times p$ matrix and $\mathbf{\Psi}$ is the representor matrix of inner products of the $\psi_{x_i}$. Then $\hat{\sigma}^2 = s^2$. Furthermore, there exists a solution of optimizing problem of the form $\widehat{\mathbb{f}} = (\widehat{f}_1, \ldots, \widehat{f}_p)'$, $\widehat{f}_i = \sum_{i=1}^n \hat{C}_{ji} \psi_{x_i}$ for $j = 1, \ldots, p$, where $\widehat{\mathbb{C}}$ solves $\hat{\sigma}^2$ finite dimensional problem. The estimator $\widehat{f}$ is unique a.s.*

<u>Choosing of the Smoothing Parameter :</u>

- in nonparametric least squares, the smoothing parameter $L$ corresponds to the diameter of the set of functions over which estimation takes place,
- the larger bounds (much larger than true norm) the less efficient estimators we obtain,
- the smaller bounds the more efficient estimators we have but inconsistent.

Define the cross validation function $\mathcal{CV}(L) = \frac{1}{n} \sum_{i=1}^n \left[ y_i - \widehat{f}_{-i}(x_i) \right]^2$, where $\widehat{f}_{-i}$ is obtained by solving

$$\min_{f \in \mathcal{H}^m} \sum_{\substack{j=1 \\ j \neq i}}^n [y_j - f(x_j)]^2 \quad \text{s.t. } \|f\|^2_{Sob,m} \leq L. \quad (20)$$

**Theorem 2.7** (1–1 Mapping of Smoothing Parameter). *Let $L > 0$ and*

$$f^* = \arg \min_{f \in \mathcal{H}^m} \frac{1}{n} [\mathbb{y} - f(\mathbb{x})]' \mathbf{\Sigma}^{-1} [\mathbb{y} - f(\mathbb{x})] \quad s.t. \ \|f\|^2_{Sob,m} \leq L \quad (21)$$

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

$$f^* = \arg \min_{f \in \mathcal{H}^m} \frac{1}{n} [\mathbb{y} - f(\mathbb{x})]' \mathbf{\Sigma}^{-1} [\mathbb{y} - f(\mathbb{x})] + \chi \|f\|^2_{Sob,m}. \quad (22)$$

This is a 1–1 mapping $\mathscr{L} : \mathbb{R}^+ \to \mathbb{R}^+ : L \mapsto \chi$.

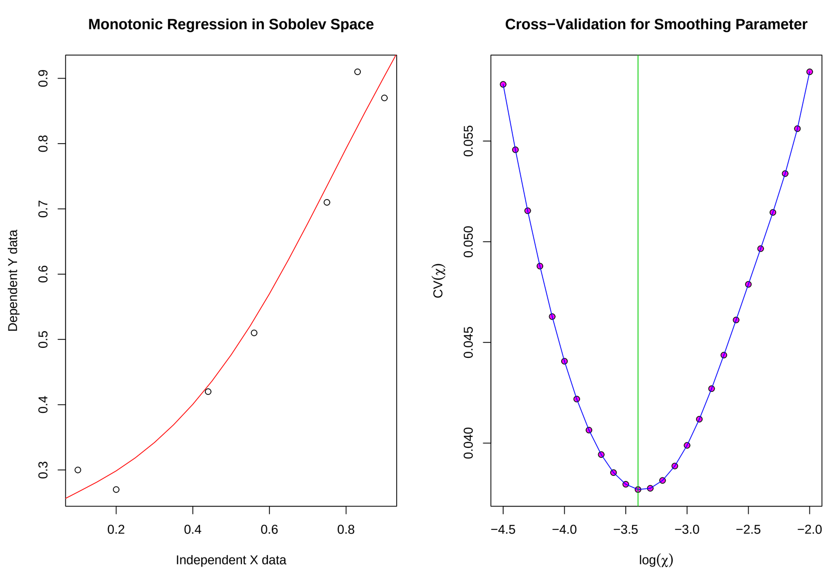


FIGURE 1: Monotonic Regression Curve in Sobolev Space of rank $m = 2$ for the best value of Smoothing Parameter $\chi$ (left) according to Cross-Validation function $\mathcal{CV}$ (right).

## 3 Isotonia

**Definition 3.1** (Constrained Single Equation Model). Invoke the assumptions for the Single Equation Model and add these assumptions:

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

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

Metric entropy $N(\delta; \mathscr{F})$ denotes the minimum number of balls of radius $\delta$ in supnorm required to cover the set of functions $\mathscr{F}$.

**Theorem 3.1** (Convergence of Constrained Estimation). *Let $\hat{f}$ satisfy $s^2 = \min_{f \in \mathcal{H}^m} \sum_i^n \frac{1}{n} (y_i - f(x_i))^2$ s.t. $f \in \widetilde{\mathscr{F}}_n$. If $f_0 \in \widetilde{\mathscr{F}}$ then the conclusion of theorem (Asymptotic Behaviour of Finite Optimizing Solution) continue to hold with rate of convergence $\eta = \frac{2m}{2m+q}$. Suppose $f \notin \widetilde{\mathscr{F}}$, $\|f_0\|^2_{Sob,m}$ is finite and there exists a unique $\widetilde{f}_0 \in \widetilde{\mathscr{F}}$ satisfying $\min_{f \in \widetilde{\mathscr{F}}} \int (f_0 - f)^2 dP_x$. Then $s^2 \xrightarrow{a.s.} \sigma^2_{\varepsilon,0} + \int \left( f_0 - \widetilde{f}_0 \right)^2 dP_x$, $n \to \infty$.*

**Definition 3.2** (Definite Monotonicity). Optimizing Problem with Smoothness and Definite Monotonicity Constraint is

$$\min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}] \quad (23)$$
$$\text{s.t.} \quad \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad \& \quad \mathbf{\Psi}^{(1)}\mathbb{c} \geq \vec{0} \quad (24)$$

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

We should call this Minimizing Problem with Smoothness and Definite Nondecreasing Constraint for correctness.
Definition determines set of functions

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

**Definition 3.3** (Indefinite Monotonicity). Optimizing Problem with Smoothness and Indefinite Monotonicity Constraint is

$$\min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}] \quad (25)$$
$$\text{s.t.} \quad \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad (26)$$
$$\& \ [\mathbf{\Psi}\mathbb{c}]_i \leq [\mathbf{\Psi}\mathbb{c}]_j \text{ for } x_i \leq x_j, \ i, j = 1, \ldots, n \quad (27)$$

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

**Definition 3.4** (Definite Convexity). Optimizing Problem with Smoothness and Definite Convexity Constraint is

$$\min_{\mathbb{c} \in \mathbb{R}^n} \frac{1}{n} [\mathbb{y} - \mathbf{\Psi}\mathbb{c}]' [\mathbb{y} - \mathbf{\Psi}\mathbb{c}] \quad (28)$$
$$\text{s.t.} \quad \mathbb{c}'\mathbf{\Psi}\mathbb{c} \leq L \quad \& \quad \mathbf{\Psi}^{(2)}\mathbb{c} \geq \vec{0} \quad (29)$$

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

Analogicaly we can also define Indefinite Convexity.

**Theorem 3.2** (Asymtotic Behaviour Based upon Laws of Large Numbers and Central Limit Theorem). *Consider* Constrained Model *and suppose that $f$ lies strictly inside the ball of functions $\|f\|^2_{Sob,m} < L$ and $f$ is strictly monotone increasing and strictly convex and is a linear combination of the representors $\psi_{X_1}, \ldots, \psi_{X_n}$. Let $\mathbf{\Pi}/n = Var(\mathbb{y}(\mathbb{X}))$. Then*

*i) $\hat{f}(\mathbb{X}) \xrightarrow{\mathcal{P}} f(\mathbb{X})$, $n \to \infty$,*

*ii) $\hat{f}^{(1)}(\mathbb{X}) \xrightarrow{\mathcal{P}} f^{(1)}(\mathbb{X})$, $n \to \infty$,*

*iii) $\hat{f}^{(2)}(\mathbb{X}) \xrightarrow{\mathcal{P}} f^{(2)}(\mathbb{X})$, $n \to \infty$,*

*iv) $n^{1/2} \left( \hat{f}(\mathbb{X}) - f(\mathbb{X}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{\Pi})$, $n \to \infty$,*

*v) $n^{1/2} (\hat{\mathbb{c}} - \mathbb{c}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{\Psi}^{-1}\mathbf{\Pi}\mathbf{\Psi}^{-1})$, $n \to \infty$,*

*vi) $n^{1/2} \left( \hat{f}^{(1)}(\mathbb{X}) - f^{(1)}(\mathbb{X}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{\Psi}^{(1)}\mathbf{\Psi}^{-1}\mathbf{\Pi}\mathbf{\Psi}^{-1}\mathbf{\Psi}^{(1)})$, $n \to \infty$,*

*vii) $n^{1/2} \left( \hat{f}^{(2)}(\mathbb{X}) - f^{(2)}(\mathbb{X}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{\Psi}^{(2)}\mathbf{\Psi}^{-1}\mathbf{\Pi}\mathbf{\Psi}^{-1}\mathbf{\Psi}^{(2)})$, $n \to \infty$.*

<u>Bootstrap :</u>

- Clasic Bootstrap
  - construct a bootstrap data set $(x_1, y_1^B), \ldots, (x_n, y_n^B)$, where $y_i^B = \hat{f}(x_i) + \hat{\varepsilon}_i^B$.
- An alternative way — "Wild" or "External" Bootstrap (see [Yat03])
  - for each estimated residual $\hat{\varepsilon}_i = y_i - \hat{f}(x_i)$ one creates a two–point distribution for random variable $\omega_i$,

| $\omega_i$ | $Prob(\omega_i)$ | $E(\omega_i)$ | $E(\omega_i^2)$ | $E(\omega_i^3)$ |
|---|---|---|---|---|
| $\hat{\varepsilon}_i(1-\sqrt{5})/2$ | $(5+\sqrt{5})/10$ | 0 | $\varepsilon_i^2$ | $\varepsilon_i^3$ |
| $\hat{\varepsilon}_i(1+\sqrt{5})/2$ | $(5-\sqrt{5})/10$ | 0 | $\varepsilon_i^2$ | $\varepsilon_i^3$ |

  - one then draws from this distribution to obtain $\hat{\varepsilon}_i^B$.

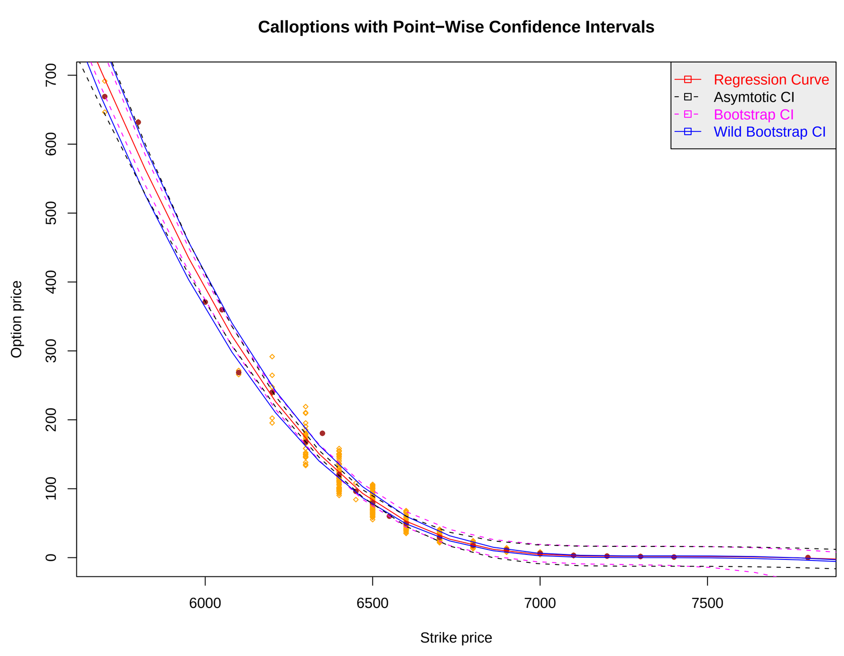


FIGURE 2: DAX Calloptions Data – Monotonic (Decreasing) and Convex Regression Curve in Sobolev Space of rank $m = 4$ with various types of 95% Confidence Intervals.

## 4 Acknowledgement


I would like to express my thanks to Mgr. Zdeněk Hlávka, Ph.D. for his generous support, valuable comments and help. My attendance on Robust 2006 was supported by grant ČSOB.


## References


[BY97] Len Bos and Adonis J. Yatchew. Nonparametric least squares estimation and testing of economic models. *Journal of Quantitative Economics*, 13:81–131, 1997.

[HY03] W. Härdle and A. J. Yatchew. Nonparametric state price density estimation using constrained least squares and the bootstrap. *University of Toronto*, 2003.

[Luk03] Jaroslav Lukeš. *Zápisky z funkcionální analýzy.* Karolinum, Prague, 2003.

[Yat03] Adonis J. Yatchew. *Semiparametric Regression for the Applied Econometrician.* Cambridge University Press, December 2003.