Regression in Sobolev Spaces Using Total Least Squares

Michal Pešta‡

CHARLES UNIVERSITY IN PRAGUE

Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics

Idea & Motivation

- input data $[x_i, y_i]$, $i = 1, \ldots, n$
- observations $[\mathbf{x}, \mathbf{y}] \equiv \left[(x_1, \ldots, x_n)^\top, (y_1, \ldots, y_n)^\top\right]$ are considered to be measured with additive errors $[\boldsymbol{\delta}, \boldsymbol{\varepsilon}]$
- unobservable true values $[\mathbf{x}+\boldsymbol{\delta}, \mathbf{y}+\boldsymbol{\varepsilon}]$ satisfy an unknown functional relationship

$$y_i + \varepsilon_i = f(x_i + \delta_i), \quad i = 1, \ldots, n$$

Regression

- unknown function f is thought to be smooth
- searching for a suitable estimator $\hat{f}$ … misfit needs to be as small as possible

Mathematical Background & Statistical Setup

- we want a modelling technique to be applicable on various types (large number) of data $\Longrightarrow$ nonparametric approach
- smoothness of unknown function f needs to be ensured … but kernels, splines or wavelets can be too restrictive

$\Rightarrow$ fit a function from a general class of smooth functions $\rightsquigarrow$ Sobolev spaces

$$\left(\mathcal{H}^m, \|\cdot\|_{Sob,m}\right) := \left\{ g \in \mathrm{L}^2 : \|g\|_{Sob,m} := \left(\sum_{i=0}^{m} \int |g^{(i)}(t)|^2 \mathrm{d}t\right)^{1/2} < +\infty \right\}$$

- the better the fit, the wilder the function and vice versa, i.e.

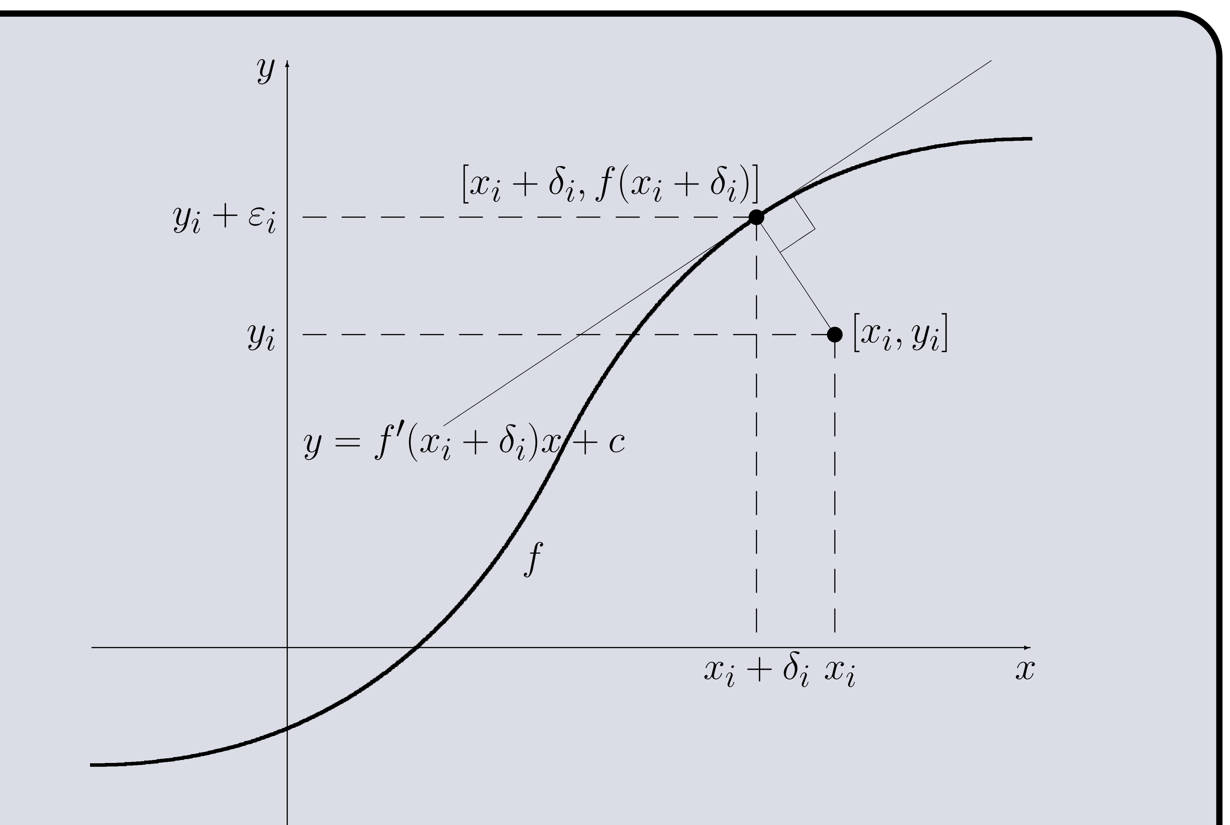
$$\frac{\text{small}}{\text{large}} \left\| \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2 \rightleftharpoons \frac{\text{large}}{\text{small}} \|f\|_{Sob,m}$$

$\Rightarrow$ find a reasonable compromise between misfit (Euclidean norm of the error vector) and smoothness (Sobolev norm of the estimator function) … choice of a smoothing parameter $\chi > 0$

$$\min_{f \in \mathcal{H}^m, \boldsymbol{\delta} \in \mathbb{R}^n, \boldsymbol{\varepsilon} \in \mathbb{R}^n} \left\{ \left\| \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2^2 + \chi \|f\|_{Sob,m}^2 \right\}, \quad \text{s.t. } \mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{f}(\mathbf{x} + \boldsymbol{\delta})$$

Optimizing

- Euclidean norm of errors $\rightsquigarrow$ orthogonal regression $\equiv$ Total Least Squares (TLS) approach
- statistical assumptions for so-called Errors-in-Variables (EIV) model:
  - rows of the errors $[\boldsymbol{\delta}, \boldsymbol{\varepsilon}]$ are iid with common zero mean and covariance matrix $\sigma^2 \mathbf{I}_2$ where $\sigma^2 > 0$ is unknown
  - no special distributional assumptions (i.e. no normality of errors)



Graphical Illustration

Estimator, Properties and Examples

- using Riesz representation theorem, Arzelà-Ascoli theorem and solving ODE one may easily derive so-called representor matrix $\boldsymbol{\Psi} \equiv \boldsymbol{\Psi}(\mathbf{x} + \boldsymbol{\delta})$ … see [2]

$$\min_{f \in \mathcal{H}^m, \boldsymbol{\delta} \in \mathbb{R}^n, \boldsymbol{\varepsilon} \in \mathbb{R}^n} \left\{ \left\| \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2^2 + \chi \|f\|_{Sob,m}^2 \right\}, \quad \text{s.t. } \mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{f}(\mathbf{x} + \boldsymbol{\delta})$$

$$\Updownarrow$$

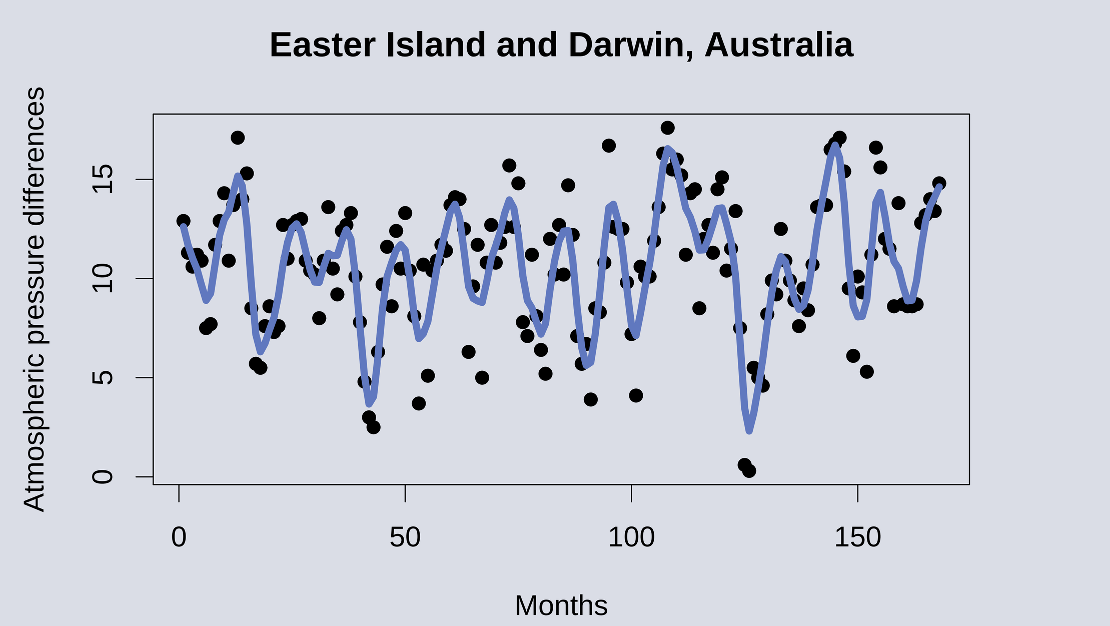
$$\min_{\mathbf{c} \in \mathbb{R}^n, \boldsymbol{\delta} \in \mathbb{R}^n} \left\{ \|\mathbf{y} - \boldsymbol{\Psi}\mathbf{c}\|_2^2 + \|\boldsymbol{\delta}\|_2^2 + \chi \mathbf{c}^\top \boldsymbol{\Psi} \mathbf{c} \right\}$$

Infinite Dimension Into Finite

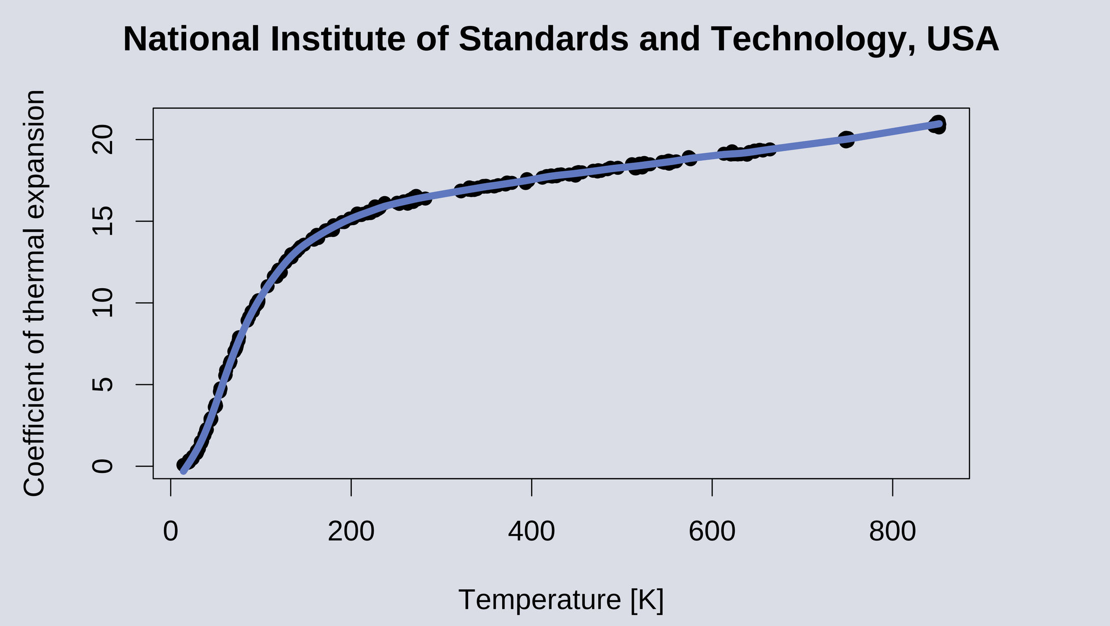
- obtaining solution $\hat{\mathbf{c}} \rightsquigarrow$ always exists a unique estimator $\hat{f}$ … relatively easy to compute, but complicated formulas
- consistency (assumption: variables "spread out" fast enough)

$$\sup_t |\hat{f}(t) - f(t)| \xrightarrow{\mathsf{P}} 0, \quad n \to \infty$$

- moreover, if the distribution of the rows of $[\boldsymbol{\delta}, \boldsymbol{\varepsilon}]$ possesses finite fourth moment $\rightsquigarrow$ asymptotic normality



El Niño – Southern Oscillation



Thermal Expansion of Copper

Remarks

- assumption of common variance σ^2 is not restrictive $\rightsquigarrow$ Scaled TLS with scaling parameter $\gamma > 0$ (see [1])

$$\min_{f \in \mathcal{H}^m, \boldsymbol{\delta} \in \mathbb{R}^n, \boldsymbol{\varepsilon} \in \mathbb{R}^n} \left\{ \|\boldsymbol{\delta}\|_2^2 + \gamma \|\boldsymbol{\varepsilon}\|_2^2 + \chi \|f\|_{Sob,m}^2 \right\}, \quad \text{s.t. } \mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{f}(\mathbf{x} + \boldsymbol{\delta})$$

- this method works without a prior knowledge of functional relation or error distribution; extendable into multivariate case
- References:

[1] Paige, C. C. and Strakoš, Z.: Core problems in linear algebraic systems, *SIAM J. on Matrix Analysis and Applications*, 27, 861–875, 2006.

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

Acknowledgments: The present work was supported by the Grant Agency of the Czech Republic (grant 201/05/H007).

‡ pesta@karlin.mff.cuni.cz, 2008