Data Modeling II.

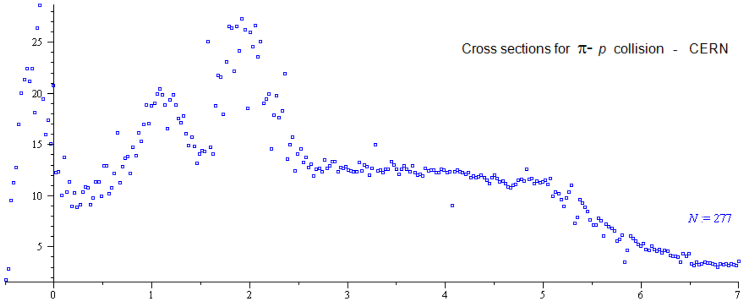
Csaba Török

UPJŠ, Košice, Slovakia
csaba.torok@upjs.sk

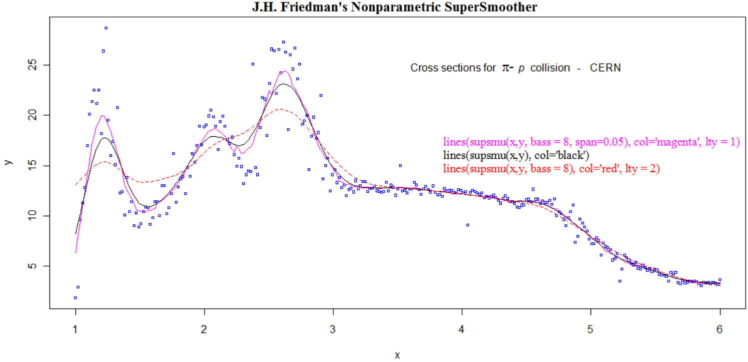
22-june-2010, Nový Smokovec

Content

1) Introduction - Our Motivation



Our Motivation



Key terms

- parametric piecewise smoothing
  - local and global models
- reference points
  - states
  - IZA reprezentation of polynomials
  - reparameterization
  - shared parameters

Key terms

- parametric piecewise smoothing
  - local and global models
- reference points: $\mathcal{R} = \{(x_0, y_0), (x_1, y_1), \ldots (x_{r-1}, y_{r-1})\}$
  - states: $y_0,\ y_1,\ \ldots, y_{r-1}$
  - IZA reprezentation of polynomials: $P(x) = I(x) + Z(x)A(x)$
  - reparameterization: $(a_0, a_1, \ldots, a_r, \ldots, a_p) \Rightarrow \mathcal{R} + (a_r, \ldots, a_p)$
  - shared parameters - within two segments

Key terms

- parametric piecewise smoothing
- reference points

2) IZA reprezentation of polynomials: P = I + ZA

Theorem

Assume $p \geq r \geq 2$. *Then any polynomial* $P_p(x)$ *can be expressed based on its any different r points* $\{[x_i,\ y_i],\ y_i = P_p(x_i),\ i = \overline{0, r-1}\}$ *as*

$$P_p(x) = I_{r-1}(x) + Z_r(x)A_{p-r,r}(x),$$

- $I_{r-1}(x) = \sum_{i=0}^{r-1} \Pi_i(x) y_i$ *is an incomplete interpolating polynomial, //*

$\Pi_i(x) = \prod_{v \in V_i} \frac{x-v}{x_i-v}, \quad V_i = V \backslash \{v_i\},\ \ V = \{x_0, x_1, \ldots, x_{r-1}\},\ i = \overline{0, r-1},$

- $Z_r(x) = \prod_{i=0}^{r-1} (x - x_i)$, - $A_{p-r,r}(x) = \mathbf{S}^T \cdot \alpha$,

$\mathbf{S} = (S_{0,r},\ S_{1,r}\ \ldots,\ S_{p-r,r})^T$, $\alpha = (a_r,\ a_{r+1}\ \ldots,\ a_p)^T$, $S_{0,r} = 1,\ r \geq 0,$

$S_{j,0} = x_0^j,\ j \geq 0, \quad S_{j,r} = S_{j,r-1} + x_r S_{j-1,r}, \quad j \geq 1,\ r \geq 1.$

IZA and reparameterization

Reparameterization in

$$P_4(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$$

IZA(4, 2):

$$\begin{aligned} P_4(x) &= \frac{x - v_1}{v_0 - v_1} P_4(v_0) + \frac{x - v_0}{v_0 - v_1} P_4(v_1) + \\ &+ (x - v_0)(x - v_1)\left[a_2 + (x + v_0 + v_1)\, a_3 + \left(x^2 + (v_0 + v_1)\, x + v_0{}^2 + v_1 (v_0 + v_1)\right) a_4\right] \end{aligned}$$

IZA(4, 4):

$$\begin{aligned} P_4(x) &= \frac{(x - v_1)(x - v_2)(x - v_3)}{(v_0 - v_1)(v_0 - v_2)(v_0 - v_3)} P_4(v_0) + \frac{(x - v_0)(x - v_2)(x - v_3)}{(v_1 - v_0)(v_1 - v_2)(v_1 - v_3)} P_4(v_1) + \\ &+ \frac{(x - v_0)(x - v_1)(x - v_3)}{(v_2 - v_0)(v_2 - v_1)(v_2 - v_3)} P_4(v_2) + \frac{(x - v_0)(x - v_1)(x - v_2)}{(v_3 - v_0)(v_3 - v_1)(v_3 - v_2)} P_4(v_3) + \\ &+ (x - v_0)(x - v_1)(x - v_2)(x - v_3)\, a_4 \end{aligned}$$

IZA reprezentation and $S_{j,r}$:

$$P_p(x) = I_{r-1}(x) + Z_r(x)\ A_{p-r,r}(x),$$
$$= I_{r-1}(x) + Z_r(x)\ \mathbf{S}^T \cdot \alpha,$$

$$S_{0,r} = 1,\ r \geq 0,\ S_{j,0} = x_0^j,\ j \geq 0,$$

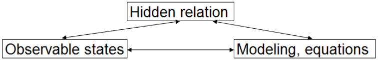
$$S_{j,r} = S_{j,r-1} + x_r S_{j-1,r}, \quad j \geq 1,\ r \geq 1.$$

$$S_{1,1} = S_{1,0} + x_1 S_{0,1} = x_0 + x_1,$$
$$S_{1,2} = S_{1,1} + x_2 S_{0,2} = x_0 + x_1 + x_2,$$
$$S_{2,1} = S_{2,0} + x_1 S_{1,1} = x_0^2 + x_1(x_0 + x_1),$$

$$S_{3,2} = S_{3,1} + x_2 S_{2,2} = (S_{3,0} + x_1 S_{2,1}) + x_2(S_{2,1} + x_2 S_{1,2}) =$$
$$= x_0^3 + x_1(x_0^2 + x_1(x_0 + x_1)) + x(x_0^2 + x_1(x_0 + x_1) + x(x_0 + x_1 + x)),$$

$$x_2 \equiv x.$$

IZA and the states



relation = states/interpolation + equation/approximation.

- IZA reprezentation - **reparameterization.**
- IZA reprezentation expresses the relation by explicit use of several observed **states** as reference points.
- IZA reprezentation can potentially leverage the **advantages** and avoid the drawbacks of interpolation and approximation.
- IZA reprezentation can express by appropriate **localization** of the reference points the connection between neighboring relations as smooth transition between two local approximants.

3) Two-part model - Problem statement

Over the intervals $[-1, 0]$ and $[0, 1]$ consider two polynomials

$$P_{\mathbf{a}}(x) \quad \text{and} \quad P_{\mathbf{b}}(x)$$

where $\mathbf{a} = (a_0, a_1, \ldots, a_p)^T$, $\mathbf{b} = (b_0, b_1, \ldots, b_q)^T$ and the degrees p, q are finite and $p, \ q \geq r \geq 2$.

To guarantee a smooth transition between the polynomials we require quasi spline conditions

$$\begin{aligned} P_{\mathbf{a}}(0) &= P_{\mathbf{b}}(0), \\ P_{\mathbf{a}}^{(j)}(0) &= P_{\mathbf{b}}^{(j)}(0) + o(\tau), \quad j = \overline{1, \ r-1}, \end{aligned} \tag{1}$$

i.e. $|P_{\mathbf{a}}^{(j)}(0) - P_{\mathbf{b}}^{(j)}(0)| < c_j \tau$, where τ is a small positive real number and $P^{(j)}$ denotes the j-th derivative of P.

Goal

For each M and N given observations from $[-1, 0]$ and $[0, 1]$

$$\{[x_{i,M}, \tilde{y}_{i,M}],\ x_{i,M} = -\frac{i}{M},\ \tilde{y}_{i,M} = P_{\mathbf{a}}(x_{i,M}) + \varepsilon^*_{i,M},\ i = \overline{1, M}\},$$

$$\{[x_{j,N}, \tilde{y}_{j,N}],\ x_{j,N} = \frac{j}{N},\ \tilde{y}_{j,N} = P_{\mathbf{b}}(x_{j,N}) + \varepsilon_{j,N},\ j = \overline{1, N}\},$$

$M, N \gg 1,\ M = \kappa N,\ 0 < \kappa < \infty$,
$\varepsilon^*, \varepsilon$ - uncorrelated.

Our aim is to find approximants for the polynomials with the foregoing smooth conditions.

It must be underline that we will solve the question of smoothness of the two polynomials in their common point differently from splines thanks to reference points.

Two-part model

We consider a two-part **sequential** model

$$\begin{aligned} \tilde{y}_{i,M} &= P_{\mathbf{a}}(-i/M) + \varepsilon^*_{i,M}, & \overline{1,\ M}, \\ \tilde{y}_{i,N} &= I_{\mathcal{R}}(j/N) + \mathbf{w}_r^T(j/N)\cdot\beta + \varepsilon_{j,N}, & \overline{1,\ N}, \end{aligned} \tag{2}$$

- $I_{\mathcal{R}}(x)$ an incomplete interpolation
  $\mathcal{R} \equiv \mathcal{R}_{\mathbf{b}} = \{\,[x_j,\ y_j],\ \ y_j = P_{\mathbf{b}}(x_j),\ \ x_i \neq x_j \text{ for } i \neq j,\ \ i,j = \overline{0, r-1}\,\}.$
  $x_j = ?,\quad y_j = ?$
- $\mathbf{w}_r(\cdot)$ the vector of the basis functions
- $\beta = (b_r, b_{r+1}, \ldots, b_q)^T$.

Matrix form

Two-part model:
Schema R-I

$$\mathcal{R} \equiv \mathcal{R}_{\mathbf{a}} = \{\, [x_j,\ P_{\mathbf{a}}(x_j)] :\ \ x_j = -j\tau,\ \ j = \overline{0, r-1}\ \},$$

$$\begin{aligned} \tilde{\mathbf{Y}}^* &= \mathbf{X}^*\,\mathbf{a} + \varepsilon *, \\ \tilde{\mathbf{Y}} &= \mathbf{I}_{\mathcal{R}_{\mathbf{a}}} + \mathbf{W}\beta + \varepsilon. \end{aligned}$$

Estimaion - computation:
- $\widehat{\mathbf{a}}_M = (\mathbf{X}^{*\mathbf{T}}\mathbf{X}^*)^{-\mathbf{1}}\mathbf{X}^{*\mathbf{T}}\tilde{\mathbf{Y}}^*,$
- $\mathcal{R}_{\hat{\mathbf{a}}} = \{\, [x_j,\ P_{\hat{\mathbf{a}}}(x_j)] :\ \ x_j = -j\tau,\ \ j = \overline{0, r-1}\ \},$
- $\widehat{\beta}_N = (\mathbf{W}^{\mathbf{T}}\mathbf{W})^{-\mathbf{1}}\mathbf{W}^{\mathbf{T}}(\tilde{\mathbf{Y}} - \mathbf{I}_{\mathcal{R}_{\hat{\mathbf{a}}}}).$

4) Piecewise approximation - Model schemas

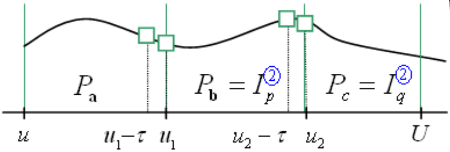
Thanks to IZA

- global piecewise smoothing
  - two-part scheme
    - sequential computation
    - simultaneously - shared parameters
  - three-part scheme
    - parallel computation

Model schemas

R - I

R - I - I - ... - I - I

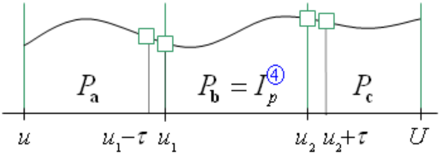


R-I-I, 2 x 2 reference points

R - I - R - I - R - ... - R - I - R

R - I - R - I - R - ... - R - I - R - I

I - R - I - R - ... - R - I - R



R-I-R, 1 x 4 reference points.

(3D data - chess board)

Shared parameters - simultaneous computation

Two-part scheme with **shared parameters**

$$\tilde{y}_{i,M} = I_R(x_i) + w(x_{i,M})\alpha + \varepsilon^*, \quad i = \overline{1,M},$$
$$\tilde{y}_{i,N} = I_R(x_i) + w(x_{i,N})\beta + \varepsilon, \quad i = \overline{1,N},$$

where $R = \{[v_i, P_i],\ v_i = u_1 - i\tau,\ P_i = P(v_i),\ i = 0,1\}$, $\alpha = (a_2, a_3)^T$, $\beta = (b_2, b_3)^T$. Parameters

$$\boxed{P_0,\ P_1 \text{ and } \alpha,\ \beta}$$

are estimated by LS from $\boxed{\tilde{Y}_1 = X_1 \cdot (a_2, a_3, P_0, P_1, b_2, b_3)^T}$,

where $\tilde{Y}_1 = \begin{pmatrix} \tilde{Y}^* \\ \tilde{Y} \end{pmatrix}$,

$$X_1 = \begin{pmatrix} f_1(x_{1,M}) & f_2(x_{1,M}) & f_3(x_{1,M}) & f_4(x_{1,M}) & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_1(x_{M,M}) & f_2(x_{M,M}) & f_3(x_{M,M}) & f_4(x_{M,M}) & 0 & 0 \\ 0 & 0 & f_3(x_{1,N}) & f_4(x_{1,N}) & f_5(x_{1,N}) & f_6(x_{1,N}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & f_3(x_{N,N}) & f_4(x_{N,N}) & f_5(x_{N,N}) & f_6(x_{N,N}) \end{pmatrix}$$

$$f_1(x) \equiv f_5(x) = (x - v_0)(x - v_1),$$
$$f_2(x) \equiv f_6(x) = (x - v_0)(x - v_1)(v_0 + v_1 + x),$$
$$f_3(x) = \frac{(x - v_1)}{(v_0 - v_1)}, \quad f_4(x) = \frac{(x - v_0)}{(v_1 - v_0)}.$$

Knot detection

The sum of squared residuals $y_i - \hat{y}_i$

$$SSE = (\widetilde{P} - \widehat{P})^T(\widetilde{P} - \widehat{P})$$

The estimate of that variance

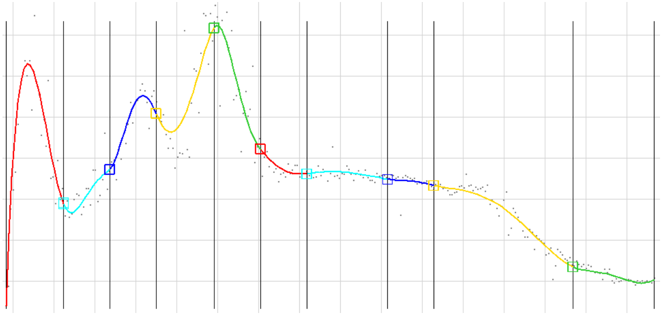
$$\widehat{\sigma^2} = \frac{(\widetilde{P} - \widehat{P})^T(\widetilde{P} - \widehat{P})}{n - k}$$

Penalization function

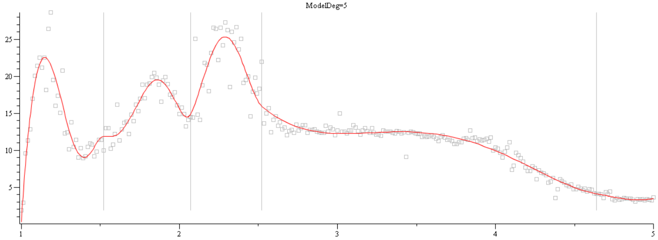
$$Aka_k = \frac{\left(\widetilde{P} - \widehat{P}\right)^T \left(\widetilde{P} - \widehat{P}\right) \left(1 + kn^{-0.25}\right)}{n - k} \xrightarrow{k} min$$

For choosing the maximal (right) interval for $P_3(x)$:
$SSE_3 > SSE_4$ - if estimating $P_3(x)$
$Aka_3 < Aka_4$ - at the beginning
$Aka_3 > Aka_4 \quad \Rightarrow$ stop and the previous interval is choosen.

- Manual knot detection/selection
- Sequential computation - two-part schema



- Automatic knot detection
- Simultaneous computation



5) Properties of two-part model - (matrix form)

Two-part model:
Schema R-I

$$\mathcal{R} \equiv \mathcal{R}_{\mathbf{a}} = \{\, [x_j,\ P_{\mathbf{a}}(x_j)] :\ \ x_j = -j\tau,\ \ j = \overline{0, r-1}\ \},$$

$$\begin{aligned} \tilde{\mathbf{Y}}^* &= \mathbf{X}^*\,\mathbf{a} + \varepsilon *, \\ \tilde{\mathbf{Y}} &= \mathbf{I}_{\mathcal{R}_{\mathbf{a}}} + \mathbf{W}\beta + \varepsilon. \end{aligned}$$

Assumption 1:
- $\widehat{\mathbf{a}}_M = (\mathbf{X}^{*\mathbf{T}}\mathbf{X}^*)^{-1}\mathbf{X}^{*\mathbf{T}}\tilde{\mathbf{Y}}^*$,
- $\mathcal{R}_{\hat{\mathbf{a}}} = \{\, [x_j,\ P_{\hat{\mathbf{a}}}(x_j)] :\ \ x_j = -j\tau,\ \ j = \overline{0, r-1}\ \}$,
- $\widehat{\beta}_N = (\mathbf{W}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathbf{T}}(\tilde{\mathbf{Y}} - \mathbf{I}_{\mathcal{R}_{\hat{\mathbf{a}}}})$.

Properties - parameter $\hat{\beta}_N$

Lemma

Under the assumption 1

a) $\hat{\beta}_N$ *is unbiased*

$$E\,\hat{\beta}_N = \beta,$$

b) Let $\mathbf{O} = \mathbf{O}_{l\times(p+1)} = \mathbf{W}^T\mathbf{J}$. *Then*

$$\begin{aligned} cov\,\hat{\beta}_N &= \sigma_*^2\,(\mathbf{W}^T\mathbf{W})^{-1}\,\mathbf{O}\,({\mathbf{X}^*}^T\mathbf{X}^*)^{-1}\,\mathbf{O}^T\,(\mathbf{W}^T\mathbf{W})^{-1} + \\ &+ \sigma^2(\mathbf{W}^T\mathbf{W})^{-1}, \end{aligned}$$

c) $\hat{\beta}_N$ *is consistent,*

$$\hat{\beta}_N \xrightarrow{P} \beta \quad as \quad N \to \infty.$$

The elements of $\mathbf{W}^T\mathbf{W}$, $\mathbf{W}^T\mathbf{J}$ and ${\mathbf{X}^*}^T\mathbf{X}^*$ have order N, i.e.
$\mathbf{W}^T\mathbf{W} = \{O_{i,j}(N)\}_{i=\overline{1,l};\, j=\overline{1,l}}, \quad \ldots$

Properties - parameter $\hat{\beta}_N$

Notation 1. Let for $N \to \infty$

$$\frac{\mathbf{X}^{*T}\mathbf{X}^{*}}{N} \to \kappa\mathbf{U}, \quad \frac{\mathbf{W}^{T}\mathbf{J}}{N} \to \mathbf{c}, \quad \frac{\mathbf{W}^{T}\mathbf{W}}{N} \to \mathbf{V}.$$

Asymptotic normality:

Theorem

Under assumption 1 and notation 1

$$\left(\sqrt{N}(\hat{\beta}_N - \beta)\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \ \frac{\sigma_*^2}{\kappa}\mathbf{V}^{-1}\,\mathbf{c}\,\mathbf{U}^{-1}\,\mathbf{c}^T\,\mathbf{V}^{-1} + \sigma^2\,\mathbf{V}^{-1}\right)$$

when $N \to \infty$.

Properties - approximant $\hat{Y} \equiv \hat{Y}_N$

$$\begin{aligned}\hat{\mathbf{Y}} &= \hat{\mathbf{I}} + \mathbf{W}\hat{\beta} = \\ &= \hat{\mathbf{I}} + \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T(\tilde{\mathbf{Y}} - \hat{\mathbf{I}}).\end{aligned}$$

Lemma

a) $\hat{\mathbf{Y}}$ is unbiased

$$E\hat{\mathbf{Y}} = \mathbf{Y},$$

b) Let $\mathcal{I}$ be the identical matrix and $\mathbf{H} \equiv \mathbf{H}_{N \times N} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$. Then

$$cov\,\hat{\mathbf{Y}} = \sigma_*^2(\mathcal{I} - \mathbf{H})\mathbf{J}({\mathbf{X}^*}^T\mathbf{X}^*)^{-1}\mathbf{J}^T(\mathcal{I} - \mathbf{H}) + \sigma^2 H.$$

c) The mean of the residual sum-of-squares is

$$E\,(\tilde{\mathbf{Y}} - \hat{\mathbf{Y}})^T \cdot (\tilde{\mathbf{Y}} - \hat{\mathbf{Y}}) = \sigma_*^2\ tr\left((\mathcal{I} - \mathbf{H})\mathbf{J}({\mathbf{X}^*}^T\mathbf{X}^*)^{-1}\mathbf{J}^T\right) + \sigma^2(N - q + r - 1).$$

Summary and conclusions

Parametric Piecewise Smoothing

- local and global models
- reference points
  - IZA reprezentation of polynomials
  - reparameterization
- knot detection
- principles
  - states
  - shared parameters
- numerical impact - system of equations with fewer parameters

Many thanks