Nonparametric variable selection and FDA

Frédéric FERRATY

Toulouse Mathematics Institute

University of Toulouse - France

frederic.ferraty@math.univ-toulouse.fr

STAPH group
(*http://www.math.univ-toulouse.fr/staph)*

High-Dimensional data

Observe 1 random variable: $\boldsymbol{\mathcal{X}} \in \mathcal{F}$

Raw data:
$\{\boldsymbol{\mathcal{X}}_i\}_{i=1,\ldots,n}$ where $\boldsymbol{\mathcal{X}}_i = (\boldsymbol{\mathcal{X}}_{i\,1}, \boldsymbol{\mathcal{X}}_{i\,2}, \ldots, \boldsymbol{\mathcal{X}}_{i\,p})$

High-Dimensional data

Observe 1 random variable: $\mathcal{X} \in \mathcal{F}$

Raw data:
$\{\mathcal{X}_i\}_{i=1,\dots,n}$ where $\mathcal{X}_i = (\mathcal{X}_{i\,1}, \mathcal{X}_{i\,2}, \dots, \mathcal{X}_{i\,p})$

High-dimensional data $\Leftrightarrow$ $\boxed{p >> n}$

High-Dimensional settings

1) High-Dimensional Vector (HDV)

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \mathcal{X}_{i\,2}, \ldots, \mathcal{X}_{i\,p_n}) \in \mathcal{F}$$

with $dim(\mathcal{F}) = p_n \xrightarrow[n\to+\infty]{} +\infty$ and $p_n >> n$

High-Dimensional settings

1) High-Dimensional Vector (HDV)

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \mathcal{X}_{i\,2}, \ldots, \mathcal{X}_{i\,p_n}) \in \mathcal{F}$$

with $dim(\mathcal{F}) = p_n \xrightarrow[n \to +\infty]{} +\infty$ and $p_n >> n$

2) Functional Variable (FV)

$\mathcal{X}_i$ discretized realization of some continuous process:

$$\mathcal{X}_i = (\mathcal{X}_i(u_1), \mathcal{X}_i(u_2), \ldots, \mathcal{X}_i(u_p))$$

$\mathcal{X} = \{\mathcal{X}(u);\ u \in \mathcal{U}\} \ \in \ \mathcal{F}$: underlying process

with $dim(\mathcal{F}) = +\infty$

High-Dimensional approaches

Summary: $\mathcal{X}_1, \ldots, \mathcal{X}_i, \ldots, \mathcal{X}_n \sim \mathcal{X} \in \mathcal{F}$

$dim(\mathcal{F})$	$+\infty$ (FV)	$p_n >> n$ (HDV)
Data	collection of functions, surfaces, operators,.....	large matrices
Exemples	spectra, densities, radar waves, spatio-temporal processes,...	thousands gene expressions,...

But: raw data (i.e. what we observe) always discrete

Framework

Observe 2 random variables:

$\mathcal{X} \in \mathcal{F}$ (high-dimensional data)
$Y \in \mathbb{R}$

Framework

Observe 2 random variables:

$\mathcal{X} \in \mathcal{F}$ (high-dimensional data)
$Y \in \mathbb{R}$

Main issue:

- $\mathcal{X}$ covariate (i.e. input)
- Y response (i.e. output)

How estimating the relationship $\mathcal{X} \to Y$?

High-Dimension and model

Problem: no standard graphical tool for depicting the relationship $\mathcal{X} \to Y$

High-Dimension and model

Problem: no standard graphical tool for depicting the relationship $\mathcal{X} \rightarrow Y$

Consequence: moderate the introduction of a priori restrictive relationship assumption (i.e. linear model)

High-Dimension and model

Problem: no standard graphical tool for depicting the relationship $\mathcal{X} \to Y$

Consequence: moderate the introduction of a priori restrictive relationship assumption (i.e. linear model)

This motivates the use of nonparametric model:
$\longrightarrow$ only regularity constraint
$\longrightarrow$ flexible model (able to catch nonlinearities)

PLAN

I) NOnparametric VAriable Selection

→ NOVAS I ***coll. with P. Hall and P. Vieu***
→ NOVAS II ***coll. with P. Hall***

PLAN

I) NOnparametric VAriable Selection

$\rightarrow$ **NOVAS I** ***coll. with P. Hall and P. Vieu***
$\rightarrow$ **NOVAS II** ***coll. with P. Hall***

II) NOVAS: naive use for FDA

I NonPar. Variable Selection

I.1) Variable selection up to now

I.2) Nonparametric variable selection

I.3) Applications

Variable selection up to now

$i = 1, \ldots, n, \quad (\mathcal{X}_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \ldots, \mathcal{X}_{i\,p}) \text{ with } p >> n$$

Variable selection up to now

$i = 1, \ldots, n, \quad (\mathcal{X}_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \ldots, \mathcal{X}_{i\,p}) \quad \text{with} \quad p >> n$$

Sparse linear regression:

$$\exists \mathcal{J} \subset \{1, 2, \ldots, p\}, \quad \boxed{Y_i = \sum_{j \in \mathcal{J}} \beta_j \, X_{i\,j} + \varepsilon_i}$$

$\rightarrow$ only few covariates are linearly active

Variable selection up to now

Penalized L_1-regression: LASSO (Tibshirani, 1996)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \sum_{i=1}^{n} \left(\boldsymbol{Y_i} - \boldsymbol{\beta_j}\, \boldsymbol{X_{i\,j}}\right)^2 + \boldsymbol{\lambda} \underbrace{\sum_{j=1}^{p} |\boldsymbol{\beta_j}|}_{L_1-penalty} \right\}$$

Variable selection up to now

Penalized L_1-regression: LASSO (Tibshirani, 1996)

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} (Y_i - \beta_j \, X_{i\,j})^2 + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{L_1-penalty} \right\}$$

L_1-penalty sets most of $\widehat{\beta}_j$'s to zero: $\widehat{\mathcal{J}} = \left\{ j, \ \widehat{\beta}_j \neq 0 \right\}$

$$\Downarrow$$

$$\boxed{\widehat{Y}_i = \sum_{j \in \widehat{\mathcal{J}}} \widehat{\beta}_j \, X_{i\,j}}$$

Variable selection up to now

LASSO: by-products, refinements and extensions

Variable selection up to now

LASSO: by-products, refinements and extensions

- coordinate descent algorithm (Fu, 1998)
- SCAD (Fan and Li, 2001)
- LAR (Efron *et al.*, 2004)
- Elastic net (Zou and Hastie, 2005)
- Dantzig selector (Candès and Tao, 2007)
- Relaxed lasso (Meinshausen, 2007)
- Group lasso (Yuan and Lin, 2008)
- ......

overview $\longrightarrow$ Bülhman and van de Geer (2011)

NOnparam. VAriable Selection

$i = 1, \ldots, n, \quad (\mathcal{X}_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \ldots, \mathcal{X}_{i\,p}) \quad \text{with} \quad p >> n$$

NOnparam. VAriable Selection

$i = 1, \ldots, n, \quad (\mathcal{X}_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \ldots, \mathcal{X}_{i\,p}) \text{ with } p >> n$$

Sparse nonparametric regression:

$$\exists \mathcal{J} \subset \{1, 2, \ldots, p\}, \boxed{Y_i = \gamma_{\mathcal{J}}\left(X_i{}^{\mathcal{J}}\right) + \varepsilon_i}$$

- $X_i{}^{\mathcal{J}} = (\mathcal{X}_{i\,j},\, j \in \mathcal{J})$
- $\gamma_{\mathcal{J}}(\cdot)$: $\mathbb{R}^{|\mathcal{J}|} \rightarrow \mathbb{R}$ smooth function

NOnparam. VAriable Selection

$i = 1, \ldots, n, \quad (\mathcal{X}_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathcal{X}_i = (\mathcal{X}_{i\,1}, \ldots, \mathcal{X}_{i\,p}) \quad \text{with} \quad p >> n$$

Sparse nonparametric regression:

$$\exists \mathcal{J} \subset \{1, 2, \ldots, p\}, \boxed{Y_i = \gamma_{\mathcal{J}}\left(X_i{}^{\mathcal{J}}\right) + \varepsilon_i}$$

- $X_i{}^{\mathcal{J}} = (\mathcal{X}_{i\,j},\, j \in \mathcal{J})$
- $\gamma_{\mathcal{J}}(\cdot) : \mathbb{R}^{|\mathcal{J}|} \longrightarrow \mathbb{R}$ smooth function

$\longrightarrow$ linear assumption relaxed
$\longrightarrow$ only few covariates are nonparametrically active

NOnparam. VAriable Selection

- leave-one-out local linear regressor:

$$\underbrace{\sum_{k=1,\,k\neq i}^{n} \left(Y_k - \alpha - \langle \mathcal{X}_k - \chi, \beta\rangle\right)^2 K_h\left(\mathcal{X}_i^{\mathcal{J}} - \chi\right)}_{Q^{-i}(\alpha,\beta)}$$

$$\rightarrow \left(\widehat{\alpha}^{-i}(\chi), \widehat{\beta}^{-i}(\chi)\right) = \arg\min_{(\alpha,\beta)} Q^{-i}(\alpha,\beta)$$

$$\rightarrow \boxed{\widehat{\gamma}_{\mathcal{J}}^{-i}(\chi) = \widehat{\alpha}^{-i}(\chi)}$$

NOnparam. VAriable Selection

- leave-one-out local linear regressor:

$$\underbrace{\sum_{k=1,\, k\neq i}^{n} \left(Y_k - \alpha - \langle \mathcal{X}_k - \chi, \beta\rangle\right)^2 K_h\left(\mathcal{X}_i^{\mathcal{J}} - \chi\right)}_{Q^{-i}(\alpha,\beta)}$$

$$\rightarrow\ \left(\widehat{\alpha}^{-i}(\chi), \widehat{\beta}^{-i}(\chi)\right) = \arg\min_{(\alpha,\beta)} Q^{-i}(\alpha, \beta)$$

$$\rightarrow\ \boxed{\widehat{\gamma}_{\mathcal{J}}^{-i}(\chi) = \widehat{\alpha}^{-i}(\chi)}$$

- cross-validation subset criterion:

$$CV(\mathcal{J}) = \sum_{i=1}^{n} \left(Y_i - \widehat{\gamma}_{\mathcal{J}}^{-i}(\mathcal{X}_i)\right)^2$$

NOVAS I

Forward addition: $\mathcal{J}_0 = \phi$, $\overline{\mathcal{J}}_0 = \{1, \ldots, p\}$

NOVAS I

Forward addition: $\mathcal{J}_0 = \phi$, $\overline{\mathcal{J}}_0 = \{1, \dots, p\}$

• step 1: select the 1st most predictive covariate

$$\rightarrow \; j_1 \; = \; \arg\min_{j \in \overline{\mathcal{J}}_0} CV\left(\{j\}\right)$$

$$\rightarrow \; \mathcal{J}_1 = \{j_1\}\,,\, \overline{\mathcal{J}}_1 = \overline{\mathcal{J}}_0 \setminus \{j_1\}$$

NOVAS I

Forward addition: $\mathcal{J}_0 = \phi$, $\overline{\mathcal{J}}_0 = \{1, \ldots, p\}$

• step 1: select the 1st most predictive covariate

$\rightarrow\ j_1 \ =\ \arg\min_{j \in \overline{\mathcal{J}}_0} CV\left(\{j\}\right)$

$\rightarrow\ \mathcal{J}_1 = \{j_1\}\,,\ \overline{\mathcal{J}}_1 = \overline{\mathcal{J}}_0 \setminus \{j_1\}$

• step 2: select the 2nd most predictive covariate

$\rightarrow\ j_2 \ =\ \arg\min_{j \in \overline{\mathcal{J}}_1} CV\left(\mathcal{J}_1 \cup \{j\}\right)$

$\rightarrow\ \mathcal{J}_2 = \mathcal{J}_1 \cup \{j_2\}\,,\ \overline{\mathcal{J}}_2 = \overline{\mathcal{J}}_1 \setminus \{j_2\}$

NOVAS I

Forward addition: $\mathcal{J}_0 = \phi$, $\overline{\mathcal{J}}_0 = \{1, \ldots, p\}$

• step 1: select the 1st most predictive covariate

$$\rightarrow \; j_1 \; = \; \arg\min_{j \in \overline{\mathcal{J}}_0} CV\left(\{j\}\right)$$

$$\rightarrow \; \mathcal{J}_1 = \{j_1\}, \overline{\mathcal{J}}_1 = \overline{\mathcal{J}}_0 \setminus \{j_1\}$$

• step 2: select the 2nd most predictive covariate

$$\rightarrow \; j_2 \; = \; \arg\min_{j \in \overline{\mathcal{J}}_1} CV\left(\mathcal{J}_1 \cup \{j\}\right)$$

$$\rightarrow \; \mathcal{J}_2 = \mathcal{J}_1 \cup \{j_2\}, \overline{\mathcal{J}}_2 = \overline{\mathcal{J}}_1 \setminus \{j_2\}$$

• and so on . . . . . .

NOVAS I

- stop under necessary minimum gain:

$$\frac{CV\left(\mathcal{J}_l\right) - CV\left(\mathcal{J}_{l+1}\right)}{CV\left(\mathcal{J}_l\right)} \leq t \text{ and } \widehat{\mathcal{J}} = \mathcal{J}_l$$

NOVAS I

- stop under necessary minimum gain:

$$\frac{CV\left(\mathcal{J}_l\right) - CV\left(\mathcal{J}_{l+1}\right)}{CV\left(\mathcal{J}_l\right)} \leq t \text{ and } \widehat{\mathcal{J}} = \mathcal{J}_l$$

Remark:

$\rightarrow$ $PCV\left(\mathcal{J}\right) = CV\left(\mathcal{J}\right) \times \left(1 + c\,\frac{|\mathcal{J}|}{\log n}\right)$

$\rightarrow$ backward stage

$\rightarrow$ K-fold CV

NOVAS II

- step 1: most predictive covariates

  * compute the permutation $\widehat{\sigma}_1(1), \ldots, \widehat{\sigma}_1(p)$ such that

  $$CV\left(\mathcal{J}^1(1)\right) \leq \cdots \leq CV\left(\mathcal{J}^1(p)\right)$$

  with $\mathcal{J}^1(k) = \left\{j_{\widehat{\sigma}_1(k)}\right\}$

  * retain the p_1 most predictive singletons $\mathcal{J}^1(1),\ \mathcal{J}^1(2), \ldots,\ \mathcal{J}^1(p_1)$

NOVAS II

- step 2: most predictive pairs of covariates
  - * build all pairs

$$\begin{array}{ccc} \mathcal{J}_1 & = & \mathcal{J}^1(1) \cup \mathcal{J}^1(2), \\ \vdots & \vdots & \vdots \\ \mathcal{J}_{p_1-1} & = & \mathcal{J}^1(1) \cup \mathcal{J}^1(p_1), \\ \mathcal{J}_{p_1} & = & \mathcal{J}^1(2) \cup \mathcal{J}^1(3), \\ \vdots & \vdots & \vdots \end{array}$$

  - * compute the permutation $\widehat{\sigma}_2(1),\ \widehat{\sigma}_2(2),\ \ldots$ such that $CV\left(\mathcal{J}^2(1)\right) \leq CV\left(\mathcal{J}^2(2)\right) \leq \cdots$ with $\mathcal{J}^2(k) = \mathcal{J}_{\widehat{\sigma}_2(k)}$
  - * retain the $p_2 = p_1$ most predictive pairs

$$\mathcal{J}^2(1),\ \mathcal{J}^2(2), \ldots,\ \mathcal{J}^2(p_2)$$

NOVAS II

- step 3: most predictive subsets of covariates
  - * build all pairs
$$\begin{array}{ccc} \mathcal{J}_1 & = & \mathcal{J}^2(1) \cup \mathcal{J}^2(2), \\ \vdots & \vdots & \vdots \\ \mathcal{J}_{p_2-1} & = & \mathcal{J}^2(1) \cup \mathcal{J}^2(p_2), \\ \mathcal{J}_{p_2} & = & \mathcal{J}^2(2) \cup \mathcal{J}^2(3), \\ \vdots & \vdots & \vdots \end{array}$$
  - * compute the permutation $\widehat{\sigma}_3(1),\ \widehat{\sigma}_3(2),\ \ldots$ such that $CV\left(\mathcal{J}^3(1)\right) \leq CV\left(\mathcal{J}^3(2)\right) \leq \cdots$ with $\mathcal{J}^3(k) = \mathcal{J}_{\widehat{\sigma}_3(k)}$
  - * retain the p_3 most predictive subsets

$$\mathcal{J}^3(1),\ \mathcal{J}^3(2), \ldots,\ \mathcal{J}^3(p_3)$$

NOVAS II

- and so on . . . . . .

NOVAS II

- and so on . . . . . .

- stop under necessary minimum gain:

$$\frac{CV\left(\mathcal{J}^l(1)\right) - CV\left(\mathcal{J}^{l+1}(1)\right)}{CV\left(\mathcal{J}^l(1)\right)} \leq t \text{ and } \widehat{\mathcal{J}} = \mathcal{J}^l(1)$$

NOVAS II

- and so on . . . . . .

- stop under necessary minimum gain:

$$\frac{CV\left(\mathcal{J}^{l}(1)\right) - CV\left(\mathcal{J}^{l+1}(1)\right)}{CV\left(\mathcal{J}^{l}(1)\right)} \leq t \text{ and } \widehat{\mathcal{J}} = \mathcal{J}^{l}(1)$$

<u>*Remarks:*</u>

$$\rightarrow \; PCV\left(\mathcal{J}\right) = CV\left(\mathcal{J}\right) \times \left(1 + c\,\frac{|\mathcal{J}|}{\log n}\right)$$

$\rightarrow K$-fold CV

$$\rightarrow \; l \leq \left|\mathcal{J}^{l}(1)\right| \leq 2^{l-1}$$

NOVAS II

The numbers of retained subsets at each step (i.e. $p_1,\ p_2, \ldots$) may depend on the capability of our computational resources.

Consider the particular case

$$p_1 = O\left(\sqrt{p}\right),\ p_2 = O\left(\sqrt{p}\right), \ldots$$

- we look only at the top $O\left(\sqrt{p}\right)$ subsets
- searching over all pairs of $O\left(\sqrt{p}\right)$ subsets remains to rank $O\left(p\right)$ subsets which is not much more onerous than step 1.

Genomics data

$n = 64$ rats
3116 covariates = 3116 genes expressions

Genomics data

$n = 64$ rats
3116 covariates = 3116 genes expressions

9 clinical measurements (responses):

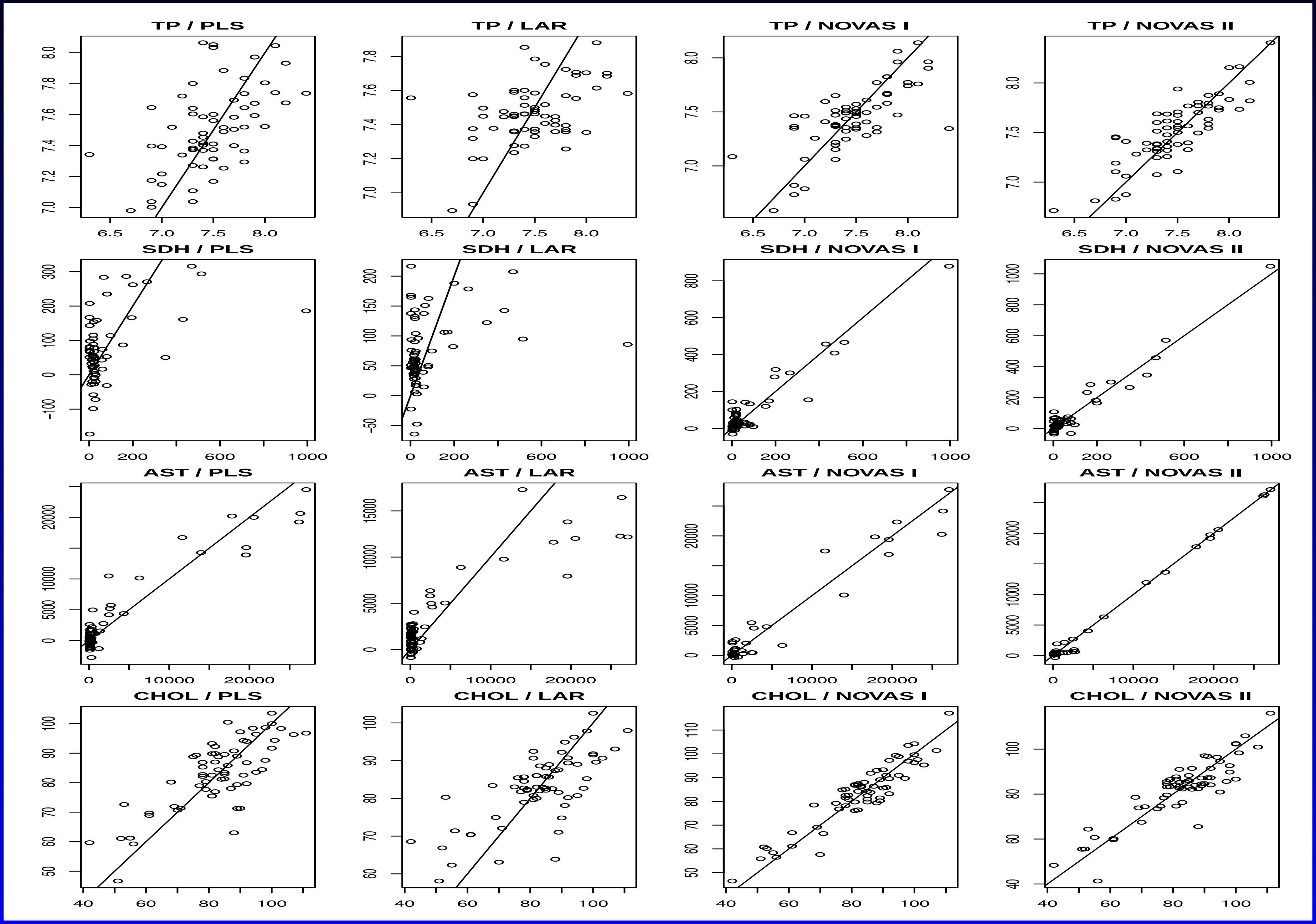
urea nitrogen (UN), total protein (TP), albumin (ALB), alanine aminotransferase (ALT), sorbitol dehydrogenase (SDH), aspartate aminotransferase (AST), alkaline phosphatase (ASP), total bile acids (TBA) and cholesterol (CHOL)

Performances (CV)

Resp.	PLSR	LAR	NOVAS I	NOVAS II $(p_k = O(\sqrt{p})$
UN	6.62	8.67	**2.62**	3.28
TP	0.104	0.117	0.072	**0.046**
ALB	0.0351	0.044	0.02	**0.016**
ALT	1236163	1709814	**46834**	67325
SDH	19736.38	21314.9	2669.6	**1478.2**
AST	5232362	9253432	2580580	**270448**
ALP	3097.81	3203.1	1225.4	**1070.6**
TBA	138.26	153.12	67.50	**39.82**
CHOL	76.91	94.96	**26.77**	41.43

PLSR=Partial Least Square Reg.
LAR=Least Angle Regression

Comparing predictions



Selected genes

Responses	Selected genes numbers
UN	1165 1421 1673 2570 1675
TP	1159 1970 2020 2173 2923 2927 2971
ALB	1038 1165 1992 2020 2105 2867 2931
ALT	1883 2770 1165 1888 2244 925 2902
SDH	764 1145 1624 1866 1940 1992 1996 2894
AST	1116 1193 1335 1826 1891 1909 2042 2057 2161 2197 2201
ALP	501 1064 1113 1395 1845 1848 2007 2385 2819
TBA	1891 1913 1916 1917 1954 2200 2205
CHOL	998 2687 2833 1848

LAR vs NOVAS

Resp.	Nb of genes LAR	Nb of genes NOVAS	Common gene(s)
UN	9 (8.67)	**5 (2.62)**	1165
TP	28 (0.117)	**7 (0.046)**	1159 2173
ALB	17 (0.044)	**7 (0.016)**	1165
ALT	52 (1709814)	**7 (46834)**	1883
SDH	14 (21315)	**8 (1478)**	764
AST	29 (9253432)	**8 (270448)**	1193 1335 1826 1909 2042 2197
ALP	27 (3283)	**9 (1071)**	501 1064 1395 2819
TBA	18 (153.12)	**7 (39.82)**	1913 1916
CHOL	11 (94.96)	**4 (26.77)**	$\emptyset$

Details of NOVAS I for SDH:

Steps	Selected genes numbers	cv
1	993	11177
2	993 1929	6814
3	993 1929 72	4291
4	993 1929 72 1868	3507
5	993 1929 72 1868 1765	2669

Details of NOVAS I for SDH:

Steps	Selected genes numbers	cv
1	993	11177
2	993 1929	6814
3	993 1929 72	4291
4	993 1929 72 1868	3507
5	993 1929 72 1868 1765	2669

Details of NOVAS II for SDH:

Steps	Selected genes numbers	cv
1	993	11177
2	764 1624	3987
3	764 1145 1624 1866	2403
4	764 1050 1145 1624 1940 1996	1588
5	764 1145 1624 1866 1940 1992 1996 2894	1478

NOVAS II: Asymptotics in brief

- $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$ i.i.d.
- $g(x_1, \dots, x_p) \stackrel{def}{=} \mathbb{E}\left(Y_i | \mathcal{X}_{i1} = x_1, \dots, \mathcal{X}_{ip} = x_p\right)$ depends on a finite subset $\{x_j;\ j \in \mathcal{J}\}$

Let $\widehat{\mathcal{J}}$ be the selected subset by NOVAS II; then, even if the number of covariates p_n diverges to infinity, it holds

$$\boxed{P\left(\widehat{\mathcal{J}} = \mathcal{J}\right) \stackrel{n \to +\infty}{\longrightarrow} 1}$$

II NOVAS: naive use for FDA

II.1) Predicting: Orange Juice data

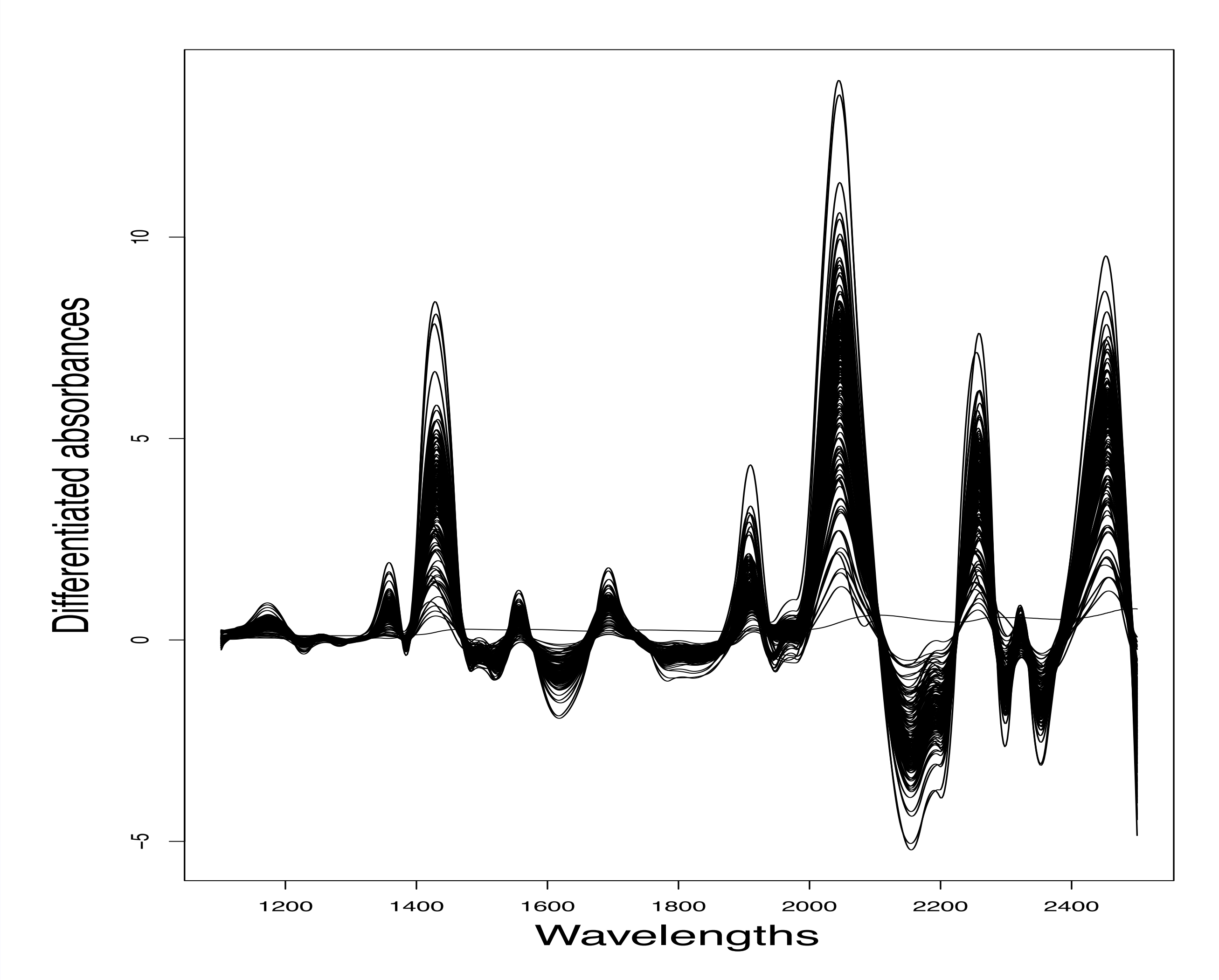
II.2) Forecasting:

- Heating-district data
- Standard & Poor’s Stock Price Index

II.3) Comparing with nonparametric functional regression

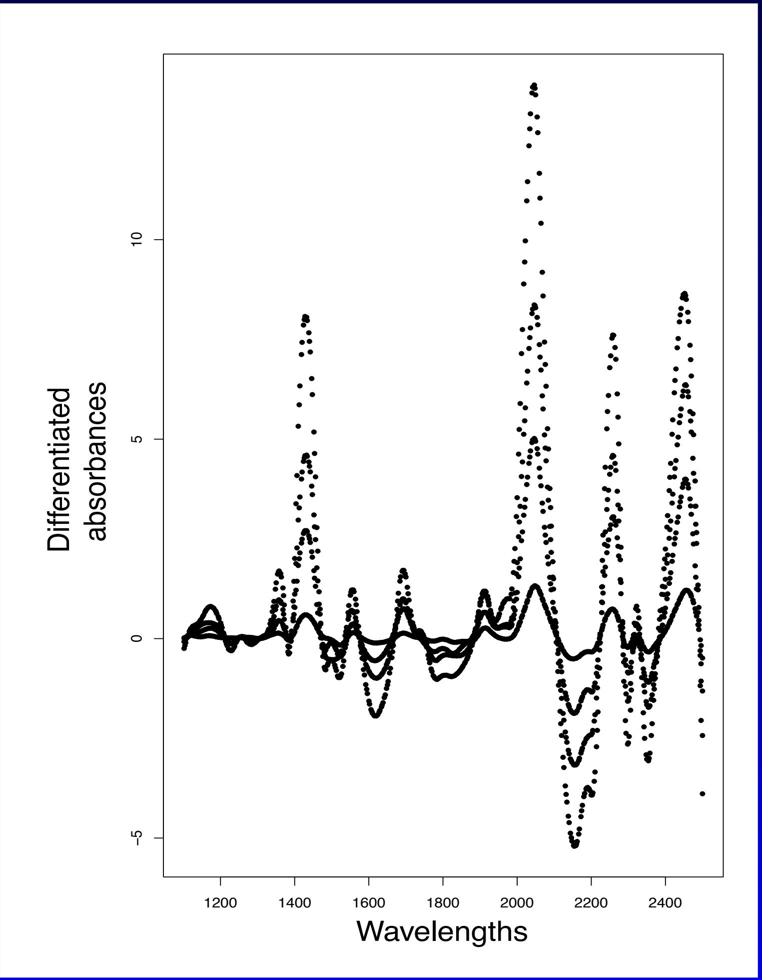
Orange Juice data

218 orange juice samples $\Rightarrow$ 218 NIR spectra
$\boldsymbol{\mathcal{X}}_1 = \{\boldsymbol{\mathcal{X}}_1(\tau);\ \tau \in [1102, 2500]\}, \ldots, \boldsymbol{\mathcal{X}}_{218}$



Raw data = discretized curves

Each spectrometric curve observed at 700 design points $\tau_1, \tau_2, \ldots, \tau_{700}$ in the range $[1102, 2500]$



$$\left.\begin{array}{c} \mathcal{X}_i(\tau_1) = \mathcal{X}_{i1} \\ \mathcal{X}_i(\tau_2) = \mathcal{X}_{i2} \\ \vdots \\ \mathcal{X}_i(\tau_{700}) = \mathcal{X}_{i700} \end{array}\right\} \quad 700 \text{ covariates}$$

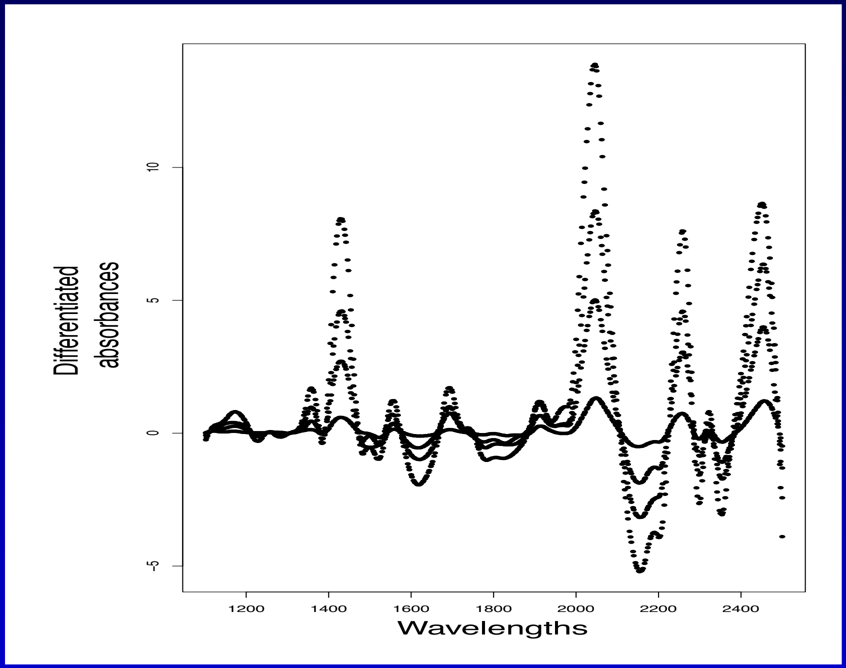
$$\mathcal{X}_i = (\mathcal{X}_{i1}, \ldots, \mathcal{X}_{i700}) \;=\; 700\text{-dimensioned vector}$$

Orange juice data

$\swarrow \quad \searrow$

$\mathcal{X}$ = Spec. vector $\qquad$ Y = Sucrose level (g/l)

$\swarrow \qquad\qquad \searrow$

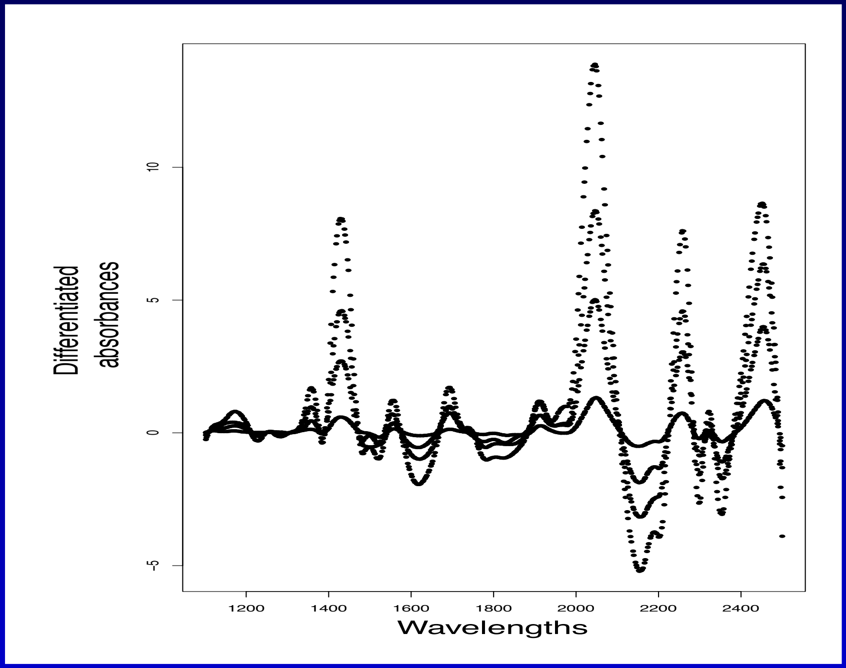


$\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{218}$

$Y_1, \ldots, Y_n$
(scalar responses)

Orange juice data

$\mathcal{X}$ = Spec. vector $\qquad$ Y = Sucrose level (g/l)

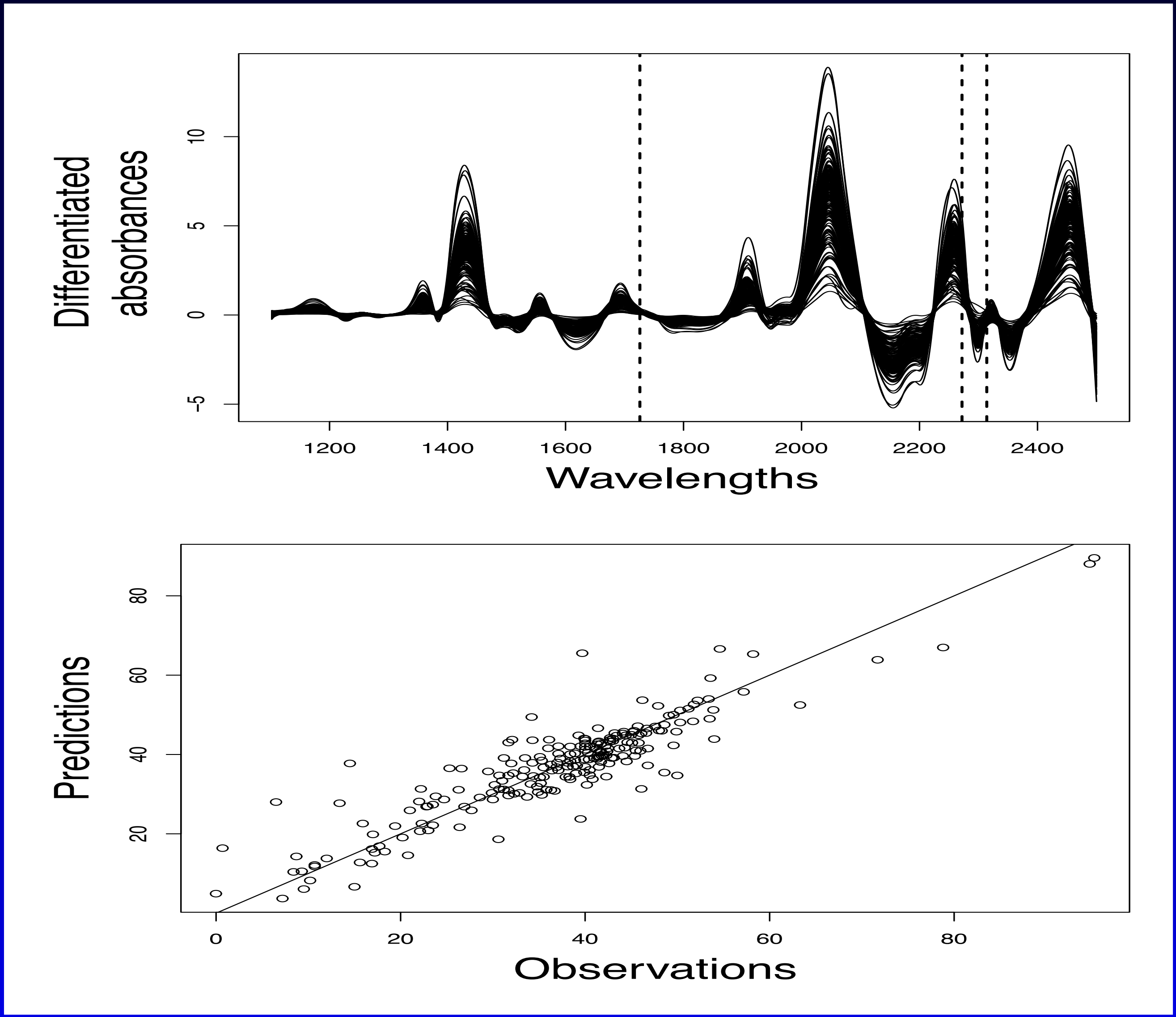


$\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, \dots, \boldsymbol{\mathcal{X}}_{218}$

$Y_1, \dots, Y_n$
(scalar responses)

Goal: predicting Y_i's from few selected covariates among $\boldsymbol{\mathcal{X}}_{i1}, \dots, \boldsymbol{\mathcal{X}}_{i700}$

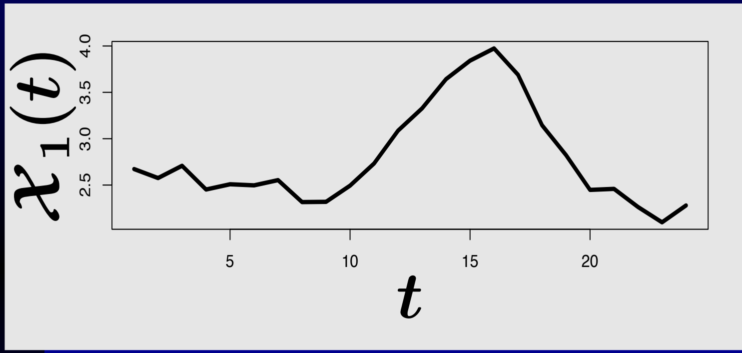
NOVAS selects only 3 covariates



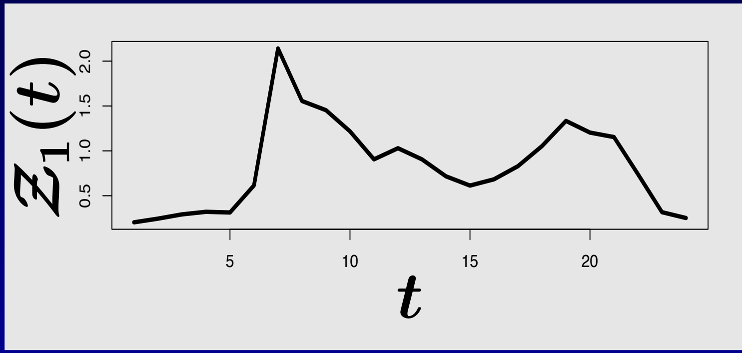
$$CV_{NOVAS} = 34.4$$

Heating-district data

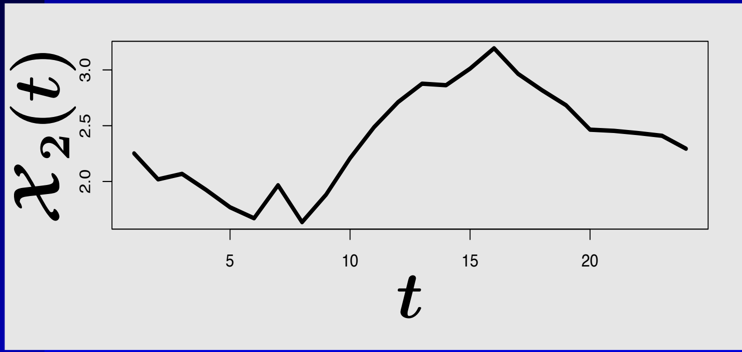
$$\mathcal{X}_1 = \{\mathcal{X}_1(t)\,;\, t \in [a,b]\}$$
$$= \textbf{temp. curve - day \#1}$$



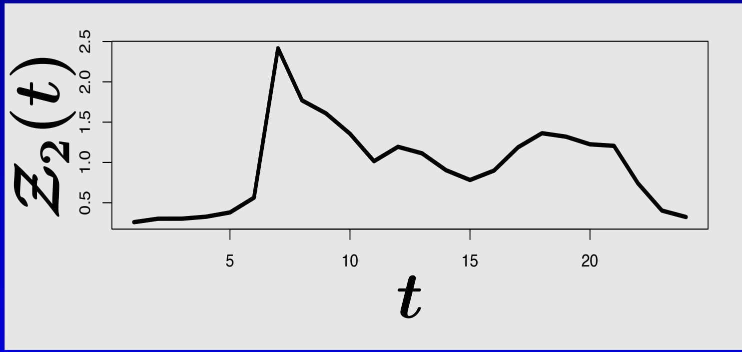
$$\mathcal{Z}_1 = \{\mathcal{Z}_1(t)\,;\, t \in [a,b]\}$$
$$= \textbf{load-demand curve - day \#1}$$



$$\mathcal{X}_2 = \textbf{temp. curve - day \#2}$$



$$\mathcal{Z}_2 = \textbf{load-demand curve - day \#2}$$

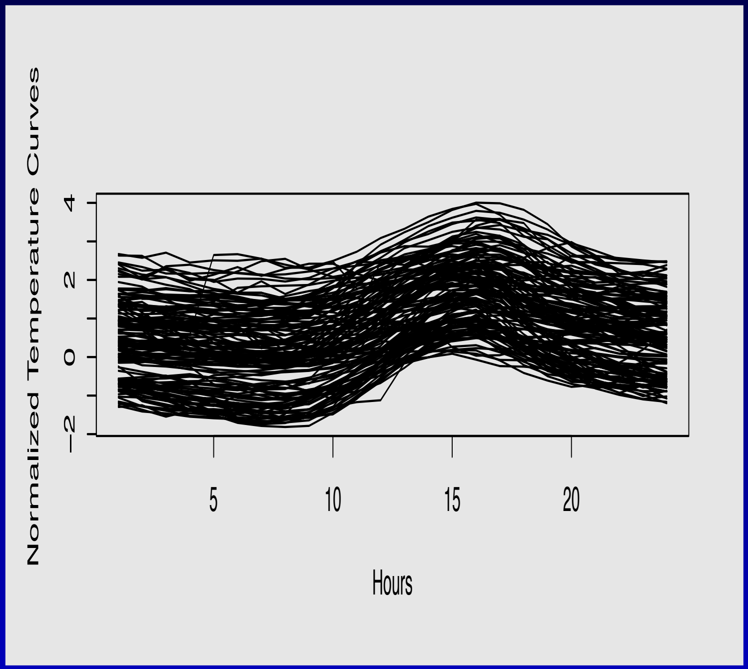


$\vdots$

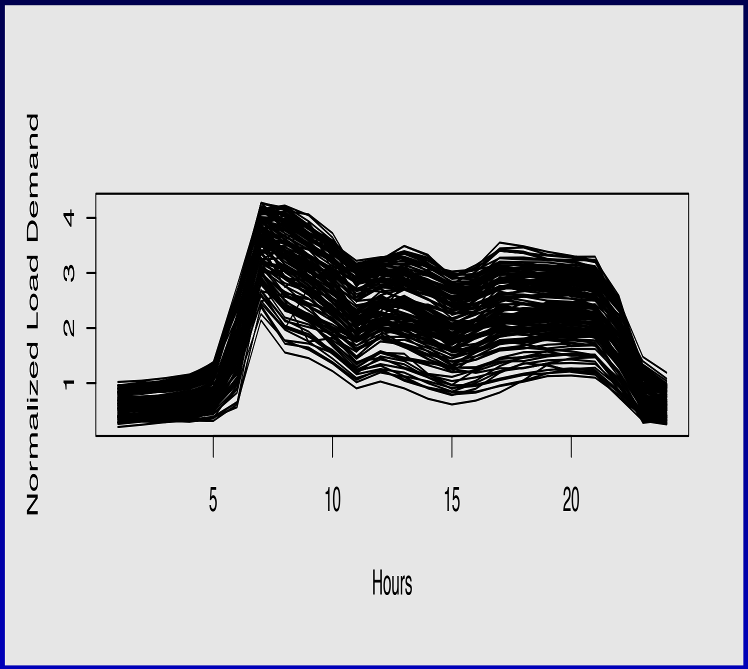
$\vdots$

Heating-district data

$\mathcal{X}$ = daily temp. curves ↓



$\mathcal{Z}$ = daily load curves ↓

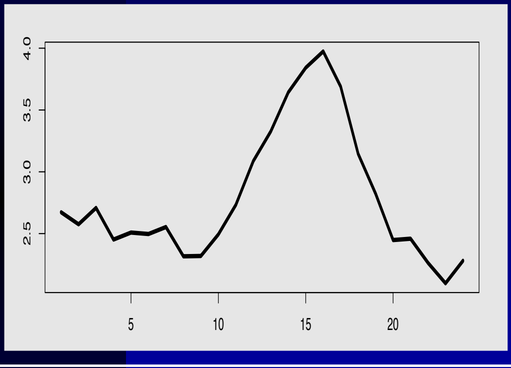


Heating-district pb

$(\mathcal{X}_i, \mathcal{Z}_i)_{i=1,\dots,n}$, n pairs of curves

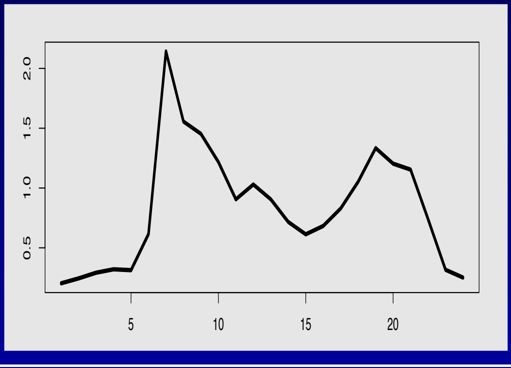
From temperature and load-demande curves at day n $(\mathcal{X}_n, \mathcal{Z}_n)$, can we predict $Y_n = \sup_t \mathcal{Z}_{n+1}(t)$?

$\mathcal{X}_1$



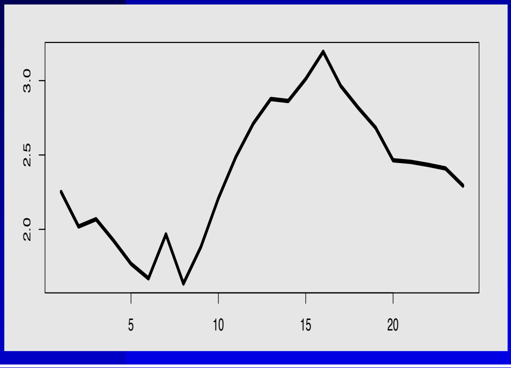
$+$

$\mathcal{Z}_1$



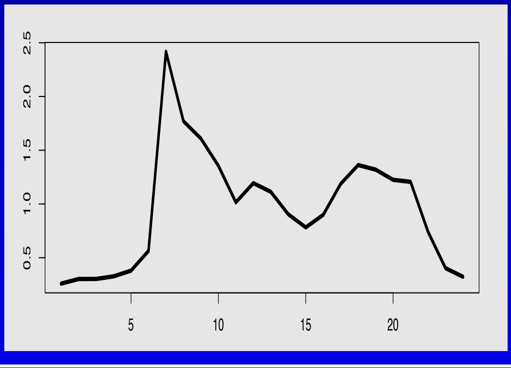
$\overset{?}{\rightarrow}$ $Y_1 = \sup_t \mathcal{Z}_2(t)$

$\mathcal{X}_2$



$+$

$\mathcal{Z}_2$



$\overset{?}{\rightarrow}$ $Y_2 = \sup_t \mathcal{Z}_3(t)$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$

NOVAS in action

Daily temperatures at day i	Daily load-demand at day i
$\mathcal{X}_i(t_1) = \mathcal{X}_{i1}$	$\mathcal{Z}_i(t_1) = \mathcal{Z}_{i1}$
$\mathcal{X}_i(t_2) = \mathcal{X}_{i2}$	$\mathcal{Z}_i(t_2) = \mathcal{Z}_{i2}$
$\vdots$	$\vdots$
$\mathcal{X}_i(t_{24}) = \mathcal{X}_{i24}$	$\mathcal{Z}_i(t_{24}) = \mathcal{Z}_{i24}$

Goal: explain nonparametrically Y_i (maximum of load-demand at day $i + 1$) by selecting the most predictive covariates among

$$\mathcal{X}_{i1}, \ldots, \mathcal{X}_{i24}, \mathcal{Z}_{i1}, \ldots, \mathcal{Z}_{i24}$$

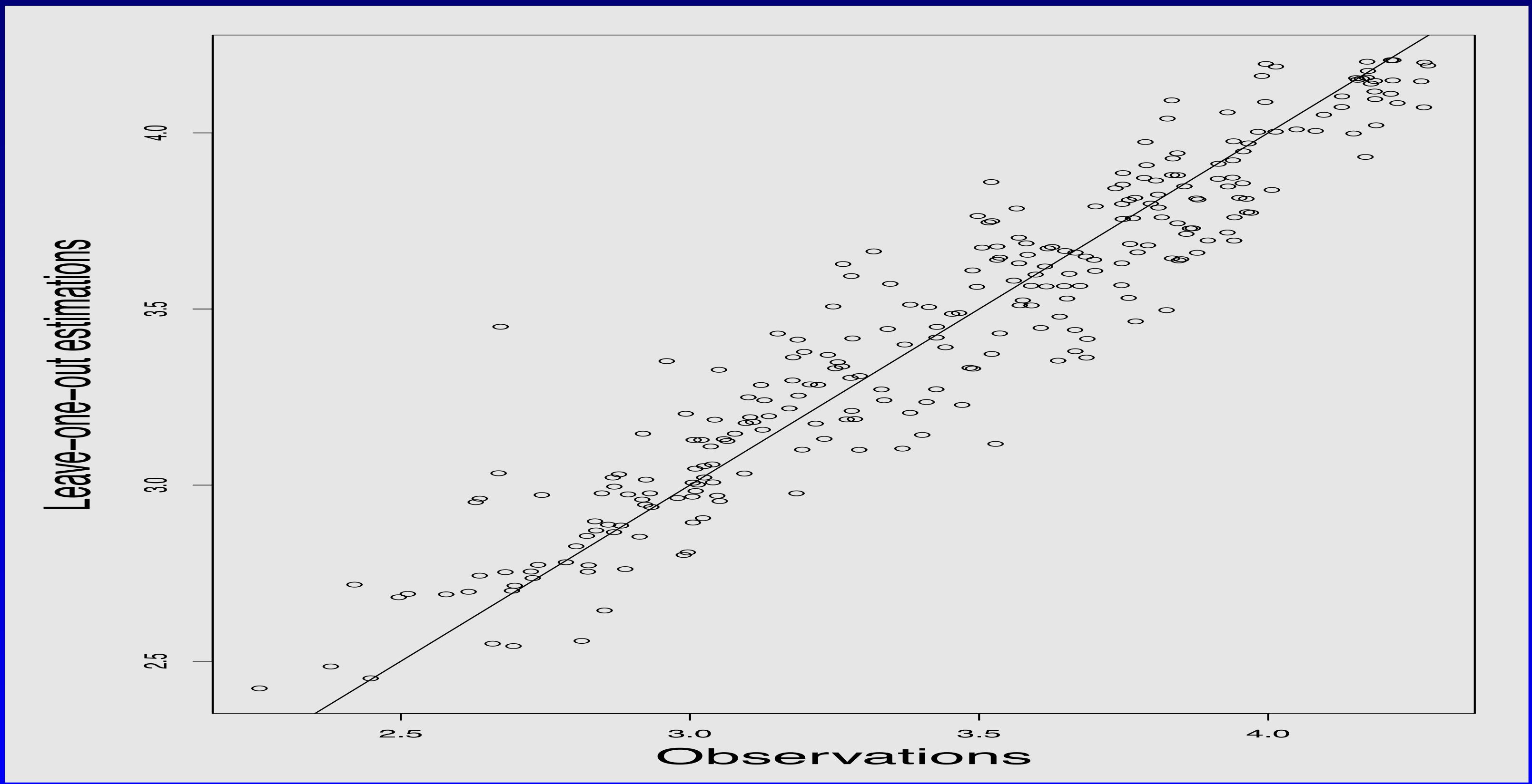
NOVAS results

NOVAS selects 2 covariates only:

$\rightarrow$ last temperature

$\rightarrow$ load-demand at 7am (maximum)

$\rightarrow$ $CV = 0.022\,(Var_n(Y) = 1)$

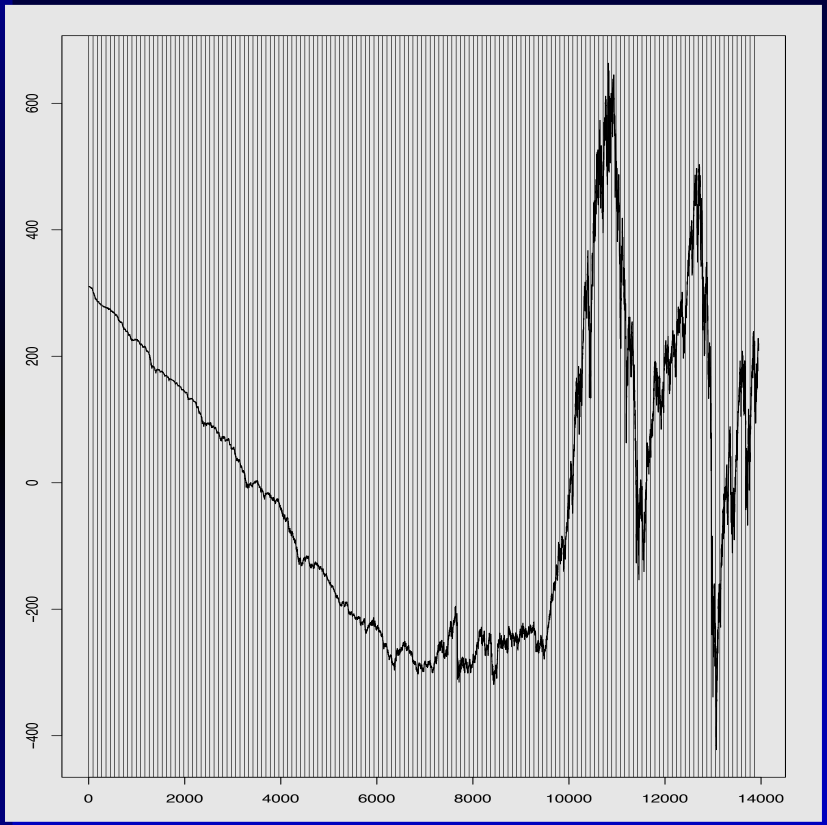


S&P500 Stock Price Index

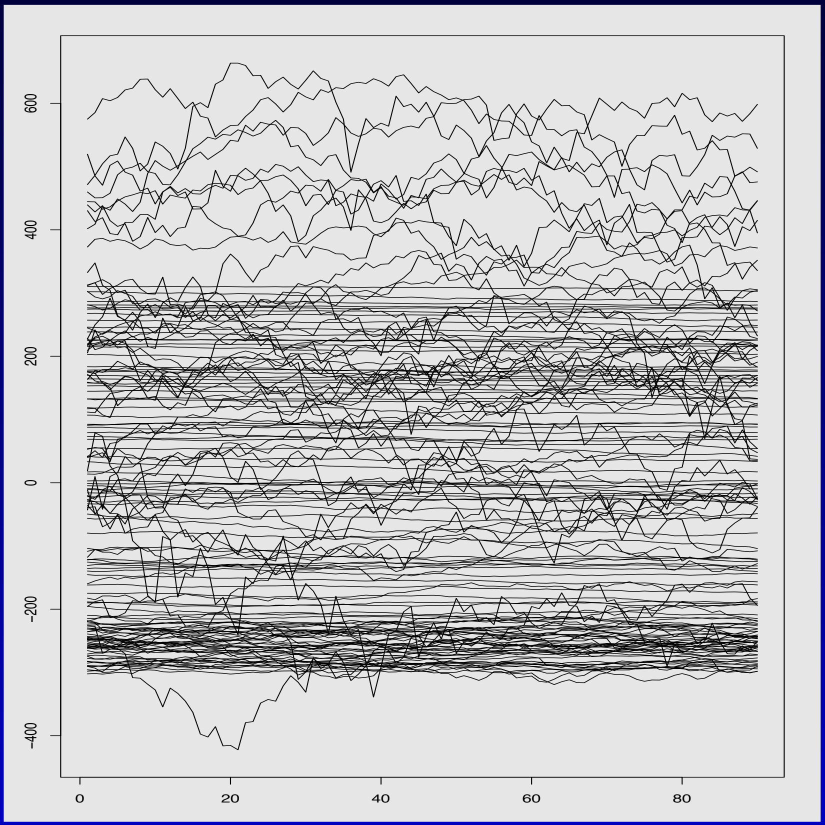
1957-03-08 $\longrightarrow$ 2012-08-30 (linear trend dropped)

S&P500 Stock Price Index

Cut into 155 $\times$ (90 days-time series):



$\rightarrow$



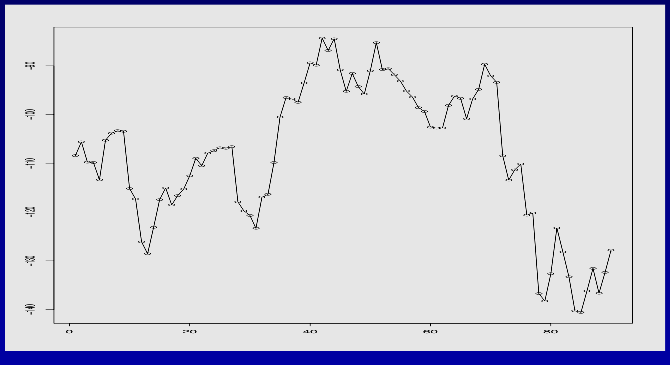
13950 days-time series

155 $\times$ (90 days-time series)

S&P forecasting pb

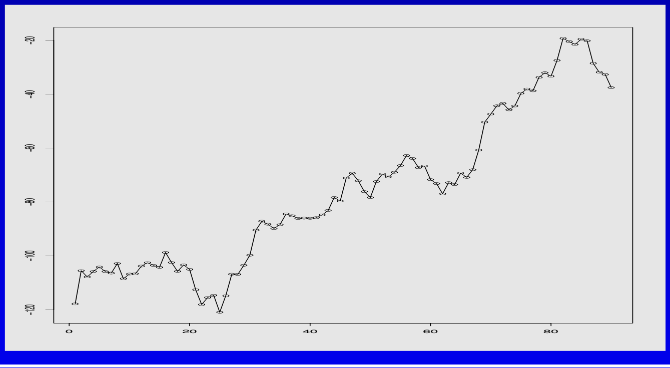
155 "sub" time series: $\mathcal{X}_i, \ldots, \mathcal{X}_{155}$; from the ith time series (90 days), can we predict the maximum reached during the next 90 days? (i.e. $Y_i = \max_t \mathcal{X}_{i+1}(t)$)

$$\mathcal{X}_i = (\mathcal{X}_{i1}, \ldots, \mathcal{X}_{i90})$$



$$\overset{?}{\rightarrow} \quad Y_i = \max_t \mathcal{X}_{i+1,j}$$

$$\mathcal{X}_{i+1} = (\mathcal{X}_{i+1,1}, \ldots, \mathcal{X}_{i+1,90})$$



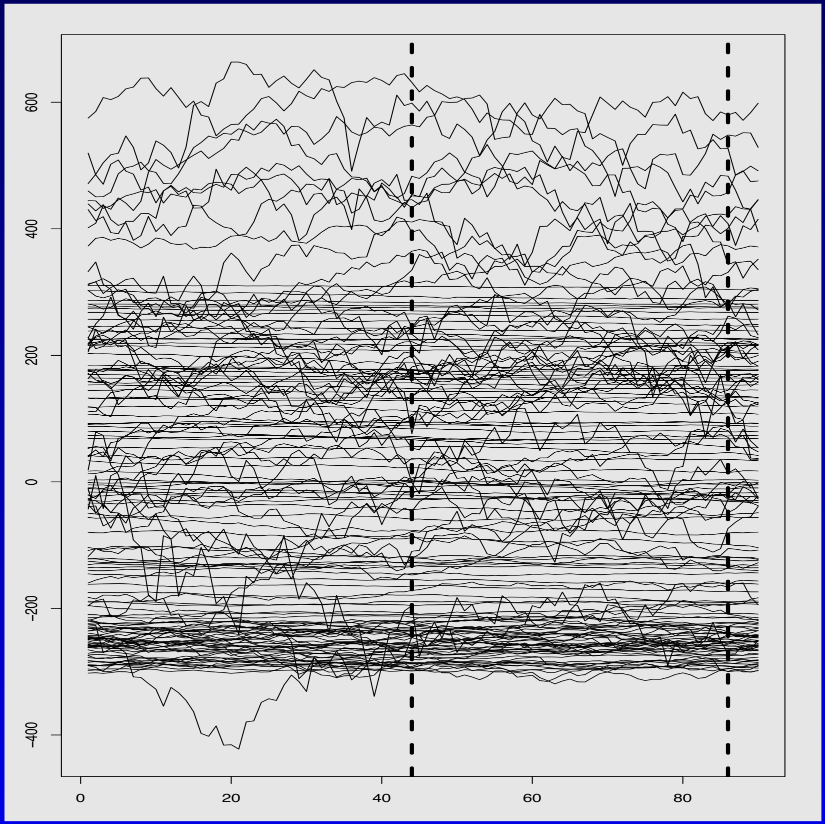
$$\overset{?}{\rightarrow} \quad Y_{i+1} = \max_j \mathcal{X}_{i+2,j}$$

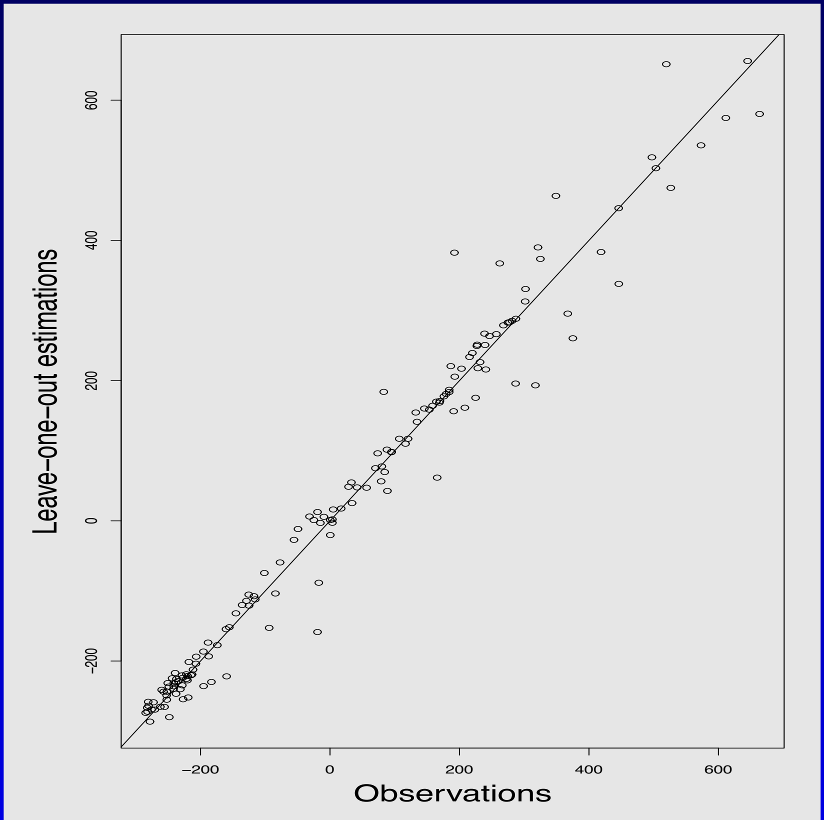
$\vdots \qquad \vdots \qquad \vdots$

S&P forecasting pb

Goal: explain nonparametrically Y_i (maximum reached during the period $i + 1$) by selecting the most predictive covariates among $\mathcal{X}_{i1}, \ldots, \mathcal{X}_{i90}$

NOVAS in action:





selection of 2 days only Predictions (CV=1587)

II.3) Comparing with Functional NonParametric Regression (FNPR)

FNPR model

$$\boxed{\boldsymbol{Y_i = r(\mathcal{X}_i) + \mathcal{E}_i,\ i = 1, \ldots n}}$$

$\mathcal{X}_i \in \mathcal{F}$: $\mathcal{F}$ semi-metric (infinite-dimensional) space

$Y_i \in \mathbb{R}$

$\mathbb{E}(\mathcal{E}_i | \mathcal{X}_i) = 0$

$r(.)$ smooth operator: $\mathcal{F} \rightarrow \mathbb{R}$
(continuous, Lipschitz,...)

Kernel estimate

NonParametric regression Model:

$$\boxed{Y_i \;=\; r(\mathcal{X}_i) + \mathcal{E}_i,\; i = 1, \ldots, n}$$

with $\underline{\mathcal{X}_i \in \{\mathcal{F},\, d(\cdot,\cdot)\}}$ and $\underline{Y_i \in \mathbb{R}}$.

Kernel estimate:

$$\widehat{r}_h(\chi) \;=\; \frac{\sum_{i=1}^n Y_i\, K\left(h^{-1} d(\chi, \mathcal{X}_i)\right)}{\sum_{i=1}^n K\left(h^{-1} d(\chi, \mathcal{X}_i)\right)}$$

$\rightarrow$ $d(.,.)$ =(pseudo) metric on the function space $\mathcal{F}$
$\rightarrow$ $K(.)$ = weight (kernel) function
$\rightarrow$ h = bandwidth (smoothing parameter)

About FNPR

- convergences of $\widehat{r}_h(.)$ (pointwise, uniform, asymptotic normality,...)
- bootstrap and confidence bands for $r(.)$
- optimal choice of h
- kNN estimator
- extension to functional response ($Y \in \mathcal{H}, \mathcal{B}, ...$)
- $\ldots$

FNPR vs NOVAS

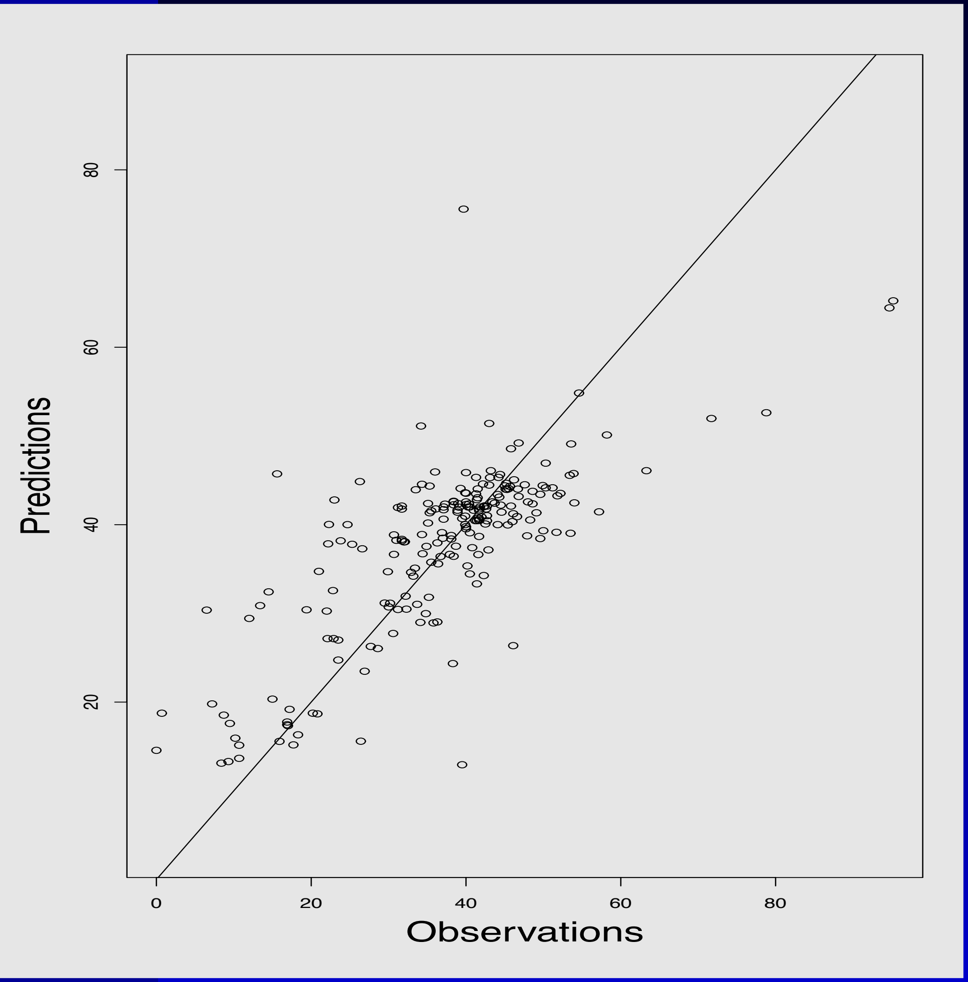
Main difference:

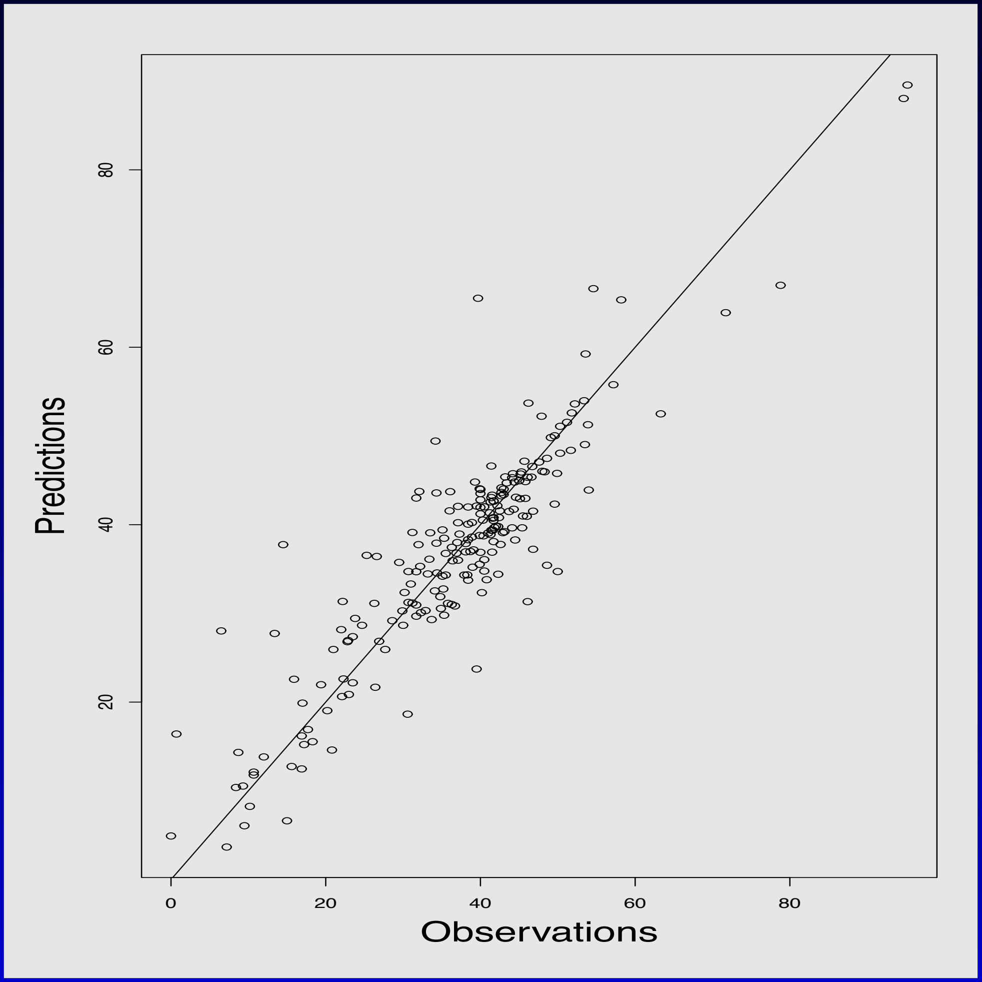
FNPR uses nonparametrically the whole curve/path

whereas

NOVAS uses nonparametrically few points of the curve/path

FNPR vs NOVAS: Orange juice data





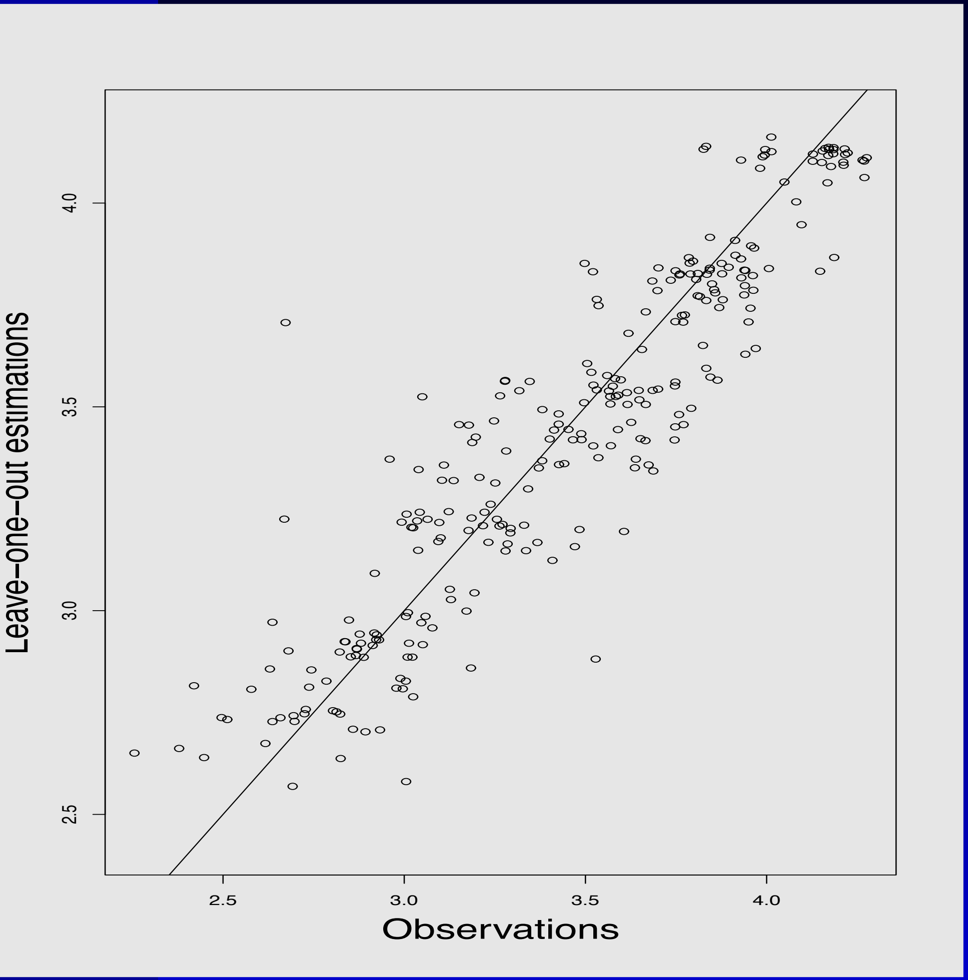
FNPR (CV=75.6)

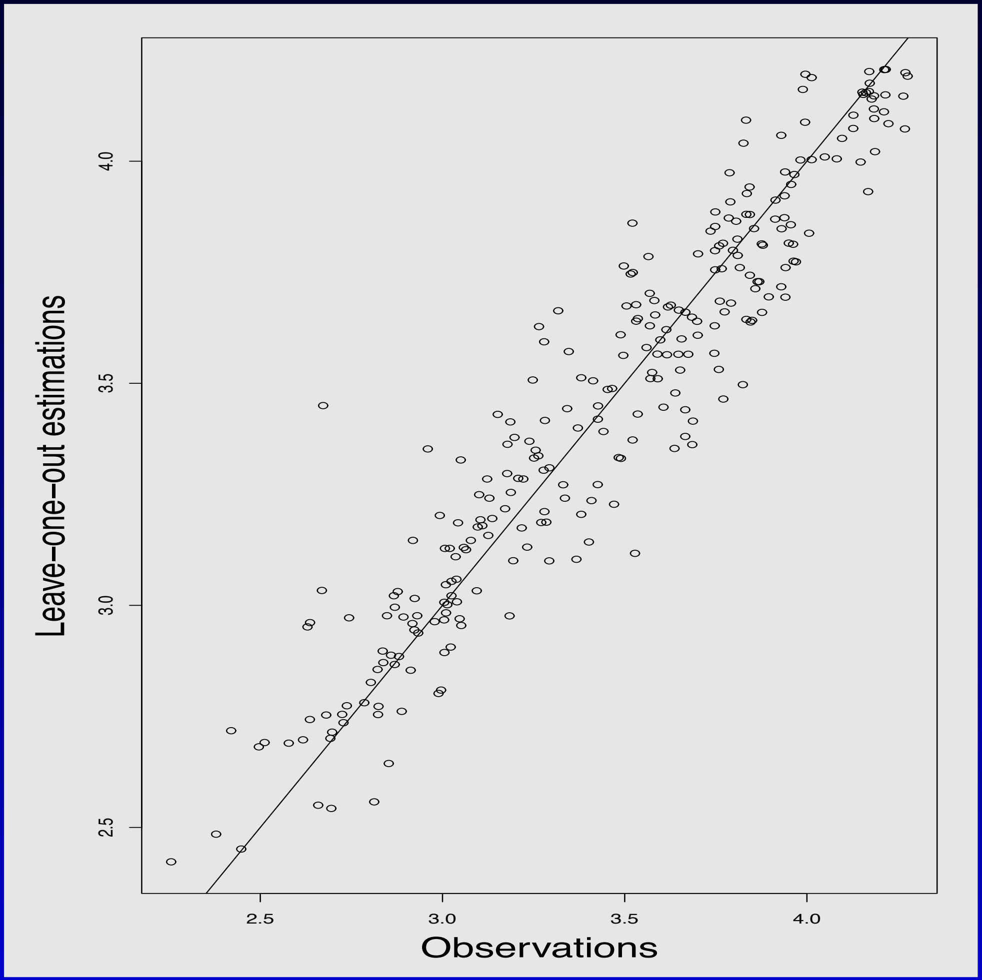
$$Y_i = r(\mathcal{X}_i) + \epsilon_i$$

NOVAS (CV=32.3)

3 (over 700) selected design points

FNPR vs NOVAS: Heating-district data





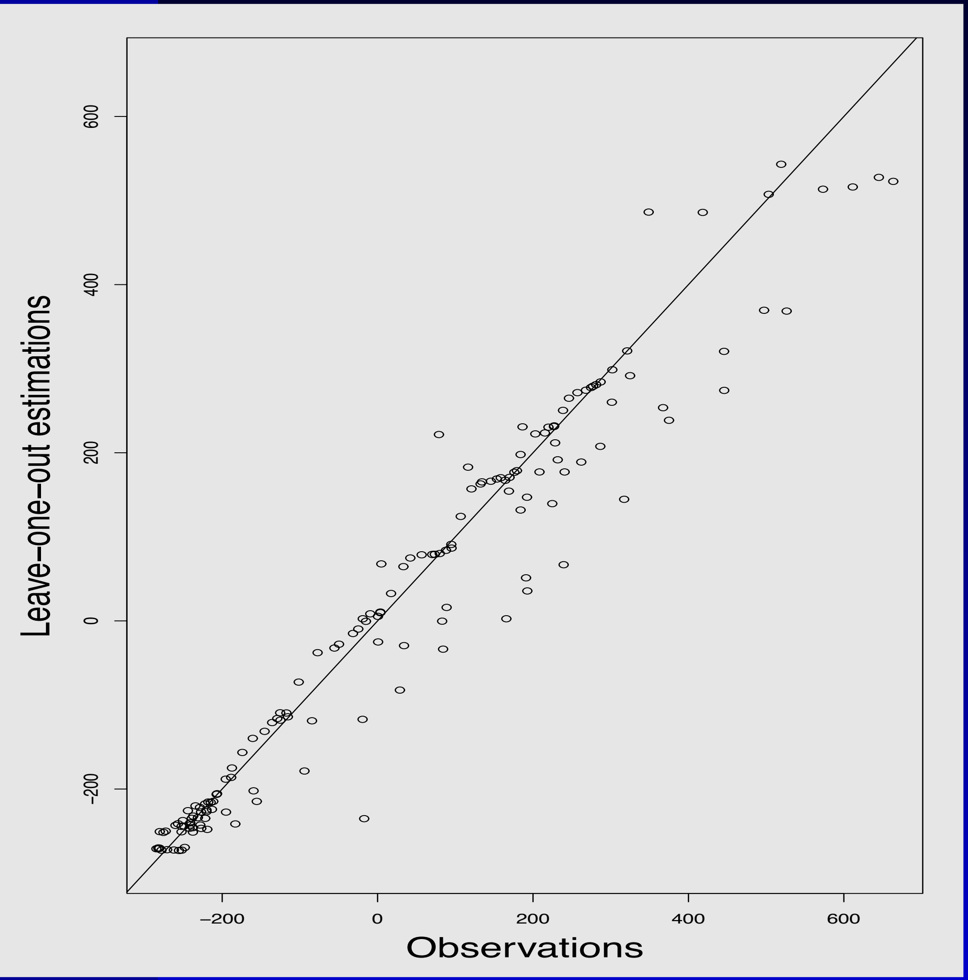
FNPR (CV=0.037)

$$Y_i = r(\mathcal{X}_i, \mathcal{Z}_i) + \epsilon_i$$

NOVAS (CV=0.022)

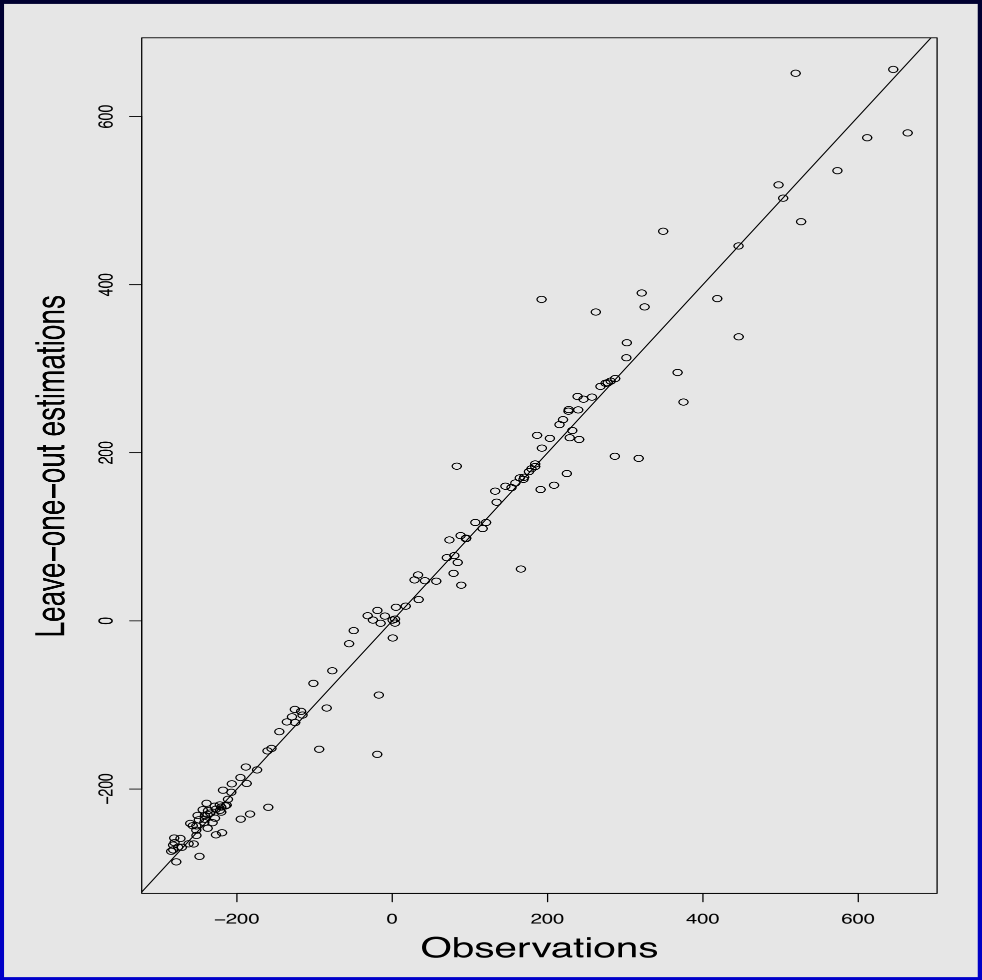
3 (over 48) selected design points

FNPR vs NOVAS: S&P500 data



FNPR (CV=3406)

$$Y_i = r(\mathcal{X}_i) + \epsilon_i$$



NOVAS (CV=1587)

2 (over 90) selected design points

Conclusion

- FD may mix continuous and pointwise structures
- NOVAS may be a useful complementary FDA tool
- Towards "expert" use:

  $\longrightarrow$ nonparam. selection of components scores

  $\longrightarrow$ nonparam. selection of functional covariates (replace the standard nonparametric regression estimator by a functional one)

  $\longrightarrow \cdots$
- Nonparametric statistics in high-dimensional setting is promising!

Main references

- F. Ferraty, P. Hall. An Algorithm for nonlinear, nonparametric model choice and prediction. (*submitted work*)
- F. Ferraty, P. Hall, P. Vieu (2010). Most predictive design points for functional data predictors. *Biometrika*, **97**, 807-824.
- F. Ferraty, P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. http://www.math.univ-toulouse.fr/staph/npfda

Main references

- F. Ferraty, P. Hall. An Algorithm for nonlinear, nonparametric model choice and prediction. (*submitted work*)
- F. Ferraty, P. Hall, P. Vieu (2010). Most predictive design points for functional data predictors. *Biometrika*, **97**, 807-824.
- F. Ferraty, P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. http://www.math.univ-toulouse.fr/staph/npfda

Thanks for your attention!!