Electricity Consumption Prediction with Functional Linear Regression

Jaromír Antoch, Luboš Prchal,
Maria Rosaria De Rosa and Pascal Sarda

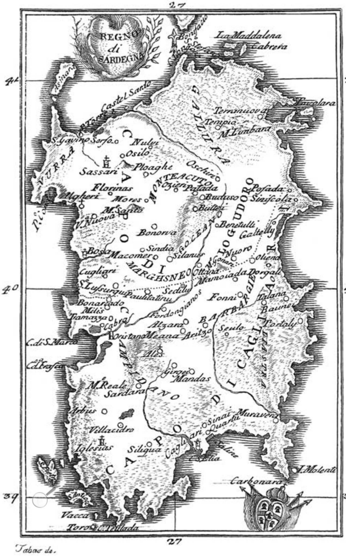
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

MODELLING SMART GRIDS 2015

enbis European Network for Business and Industrial Statistics

Goals

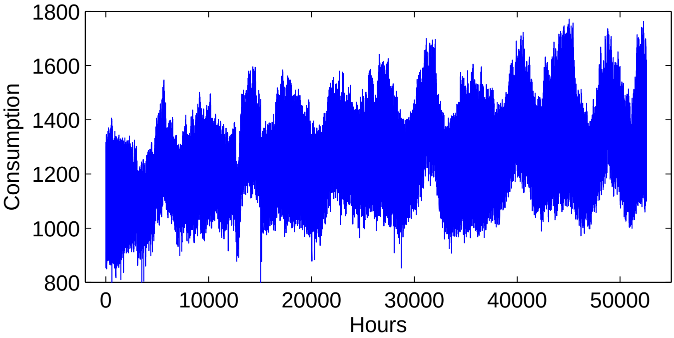
- Functional linear regression model linking observations of a functional response variable with measurements of an explanatory functional variable is considered.
- Our aim is to analyze effect of a functional variable on a functional response by means of functional linear regression models when slope function is estimated with tensor product splines.
- Model is applied to real data comprising electricity consumption of Sardinia 2000 – 2005.
- Computational issues are addressed.



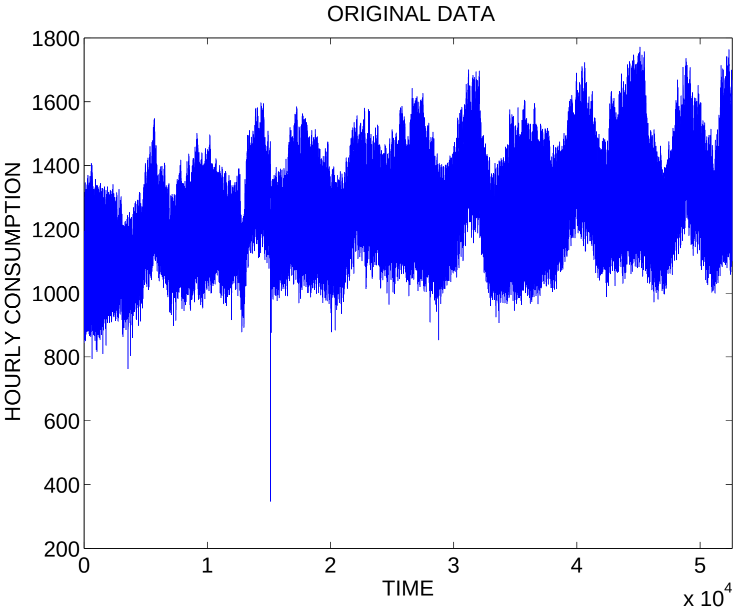
Data

Model serves to analyze real data set concerning electricity consumption of Sardinia.

Data set consists of 52 584 values of electricity consumption collected every hour within January 1, 2000 – December 31, 2005.



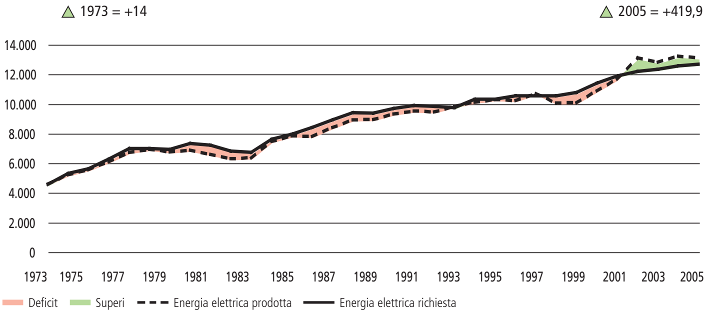
Data


ORIGINAL DATA
HOURLY CONSUMPTION
TIME
x 10^4
1800
1600
1400
1200
1000
800
600
400
200
0
1
2
3
4
5

Official data – Sardinia

Energia richiesta

Energia richiesta in Sardegna	GWh	12.611,6
△ Deficit (-) Superi (+) della produzione rispetto alla richiesta	GWh	+419,9
	%	3,3



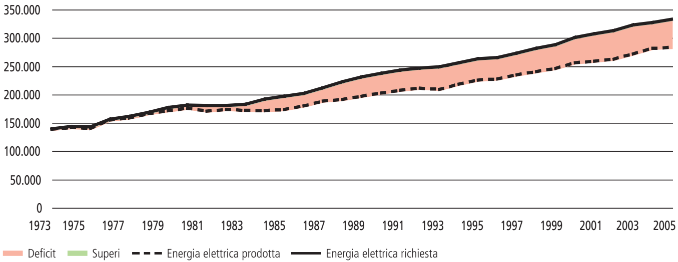
Consumi: complessivi 12.036,7 GWh; per abitante 7.286 kWh

Official data – Italy

Energia richiesta

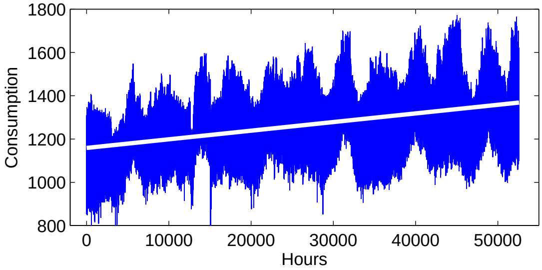
Energia richiesta in Italia	GWh	330.443,0
△ Deficit (-) Superi (+) della produzione rispetto alla richiesta	GWh	-49.154,5
	%	14,9

△ 1973 = -879 △ 2005 = -49.154,5

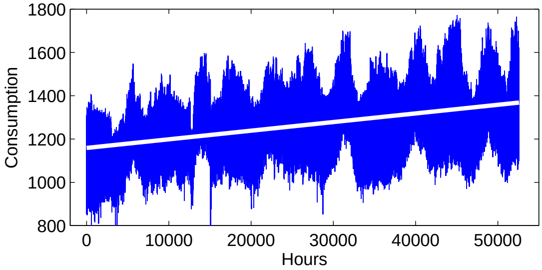


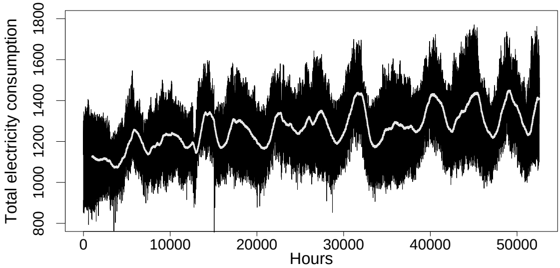
Consumi: complessivi 309.816,8 GWh; per abitante 5.286 kWh

Basic trends

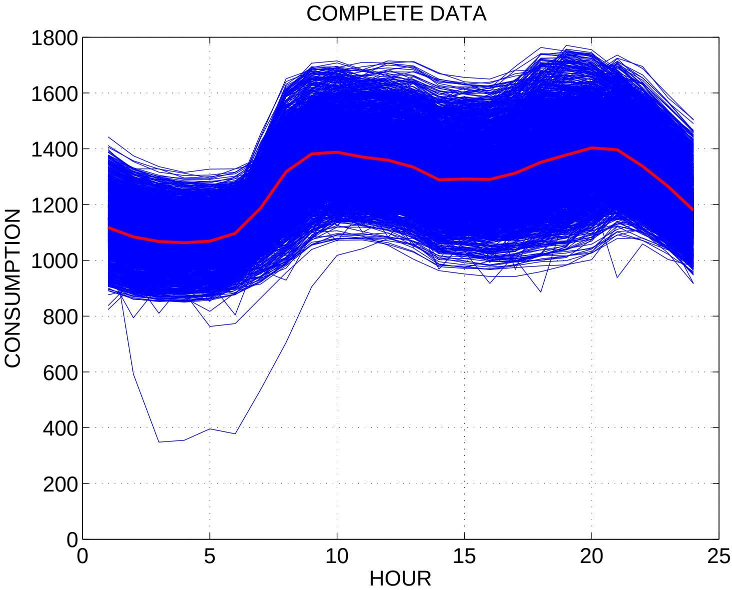


Basic trends

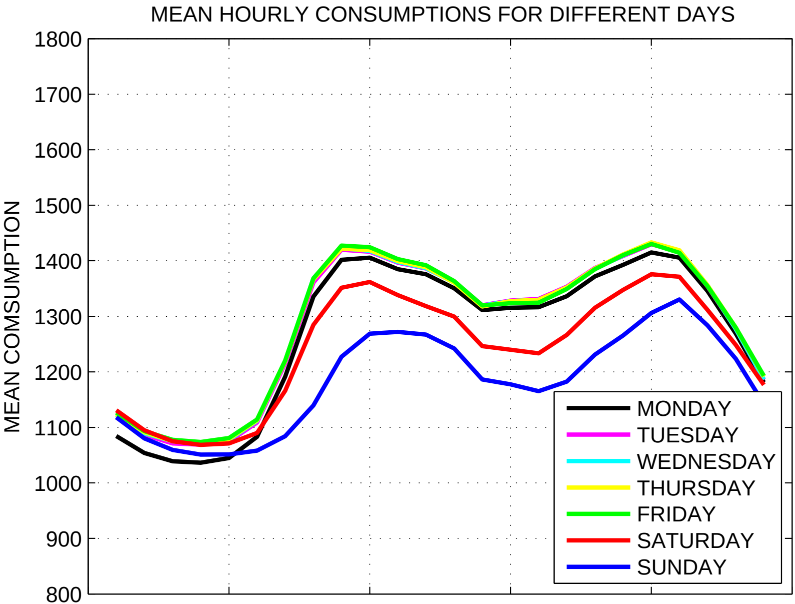




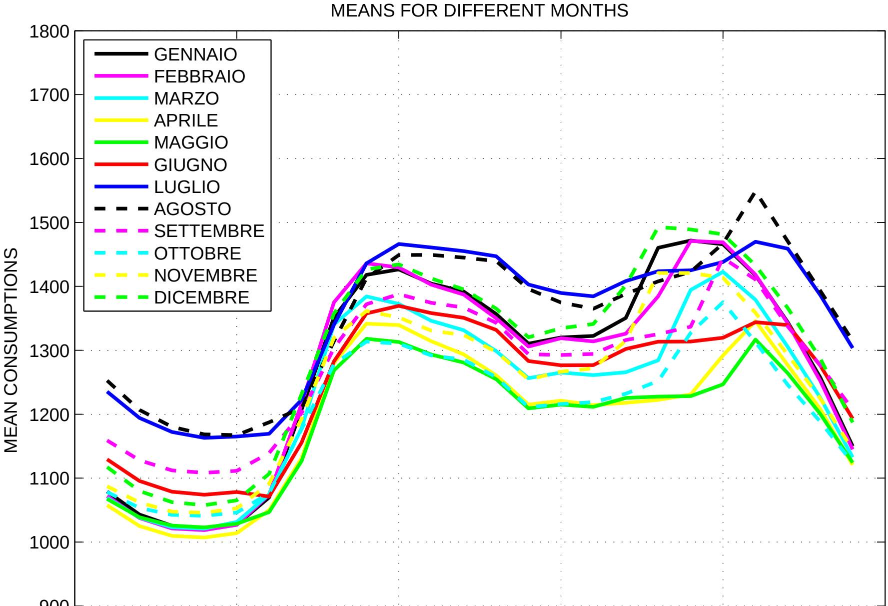
Consumptions for one day



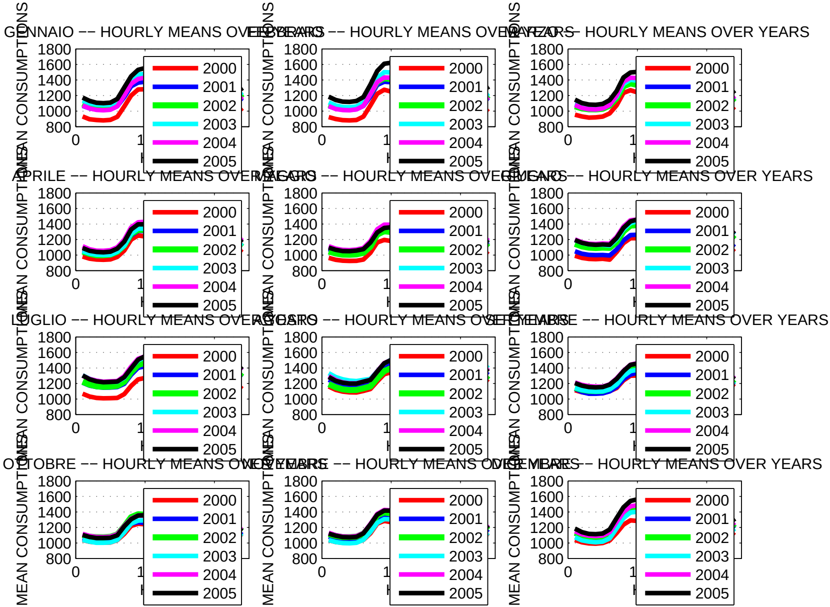
Mean consumption for individual days



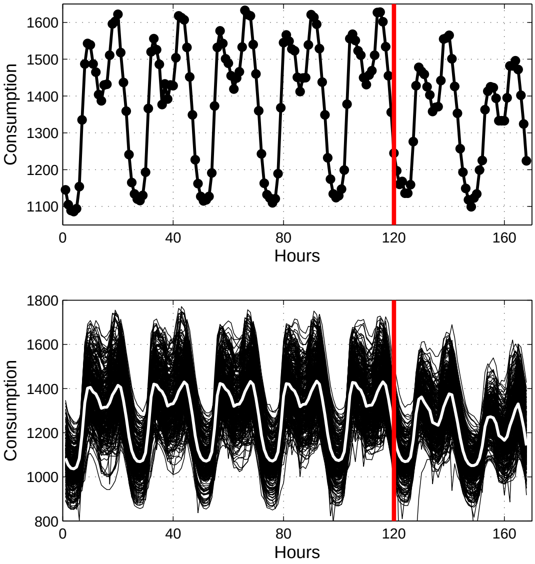
Mean consumption for individual months



Mean consumption : Individual months over years



Consumption for one week



Main tasks

- Main interest is predicting oncoming weekend and/or weekdays consumption curve if present weekdays consumption is known and functional predictor is curve of present weekdays consumption.

  Model

$$Y_i(t) = \alpha(t) + \int_{I_1} X_i(s)\beta(s,t)\,ds + \varepsilon_i(t), \quad t \in I_2,\ i = 1, \dots, n$$

  Data

  - Functional predictors X_i's represent weekdays curves
  - Y_i's represent a weekend curves
    or a weekday curve in which case $Y_i = X_{i+1}$
  - Recall that model corresponds to ARH(1)
- Complete data series has been cut into 307 weeks Weekdays (Mo to Fri) and weekends (Sa to Su) separated (reason, leading to two sets of discretized electricity consumption curves, is fundamental difference between weekdays and weekend consumption).

Assumptions

- Data are observations of identically distributed random functional variables $\{X_i(s), Y_i(t), s \in I_1, t \in I_2\}$, $i = 1, \ldots, n$, defined on same probability space and taking values in some functional spaces.
- We consider separable real Hilbert spaces $L^2(I_1)$ and $L^2(I_2)$ of square integrable functions defined on compact intervals $I_1 \subset \mathbb{R}$ and $I_2 \subset \mathbb{R}$, equipped with standard inner products.
- We focus on functional linear relation

$$Y_i(t) = \alpha(t) + \int_{I_1} X_i(s)\beta(s,t)\,ds + \varepsilon_i(t), \quad t \in I_2, \; i = 1, \ldots, n$$

  - $\alpha(t) \in L^2(I_2)$ and $\beta(s,t) \in L^2(I_1 \times I_2)$ are unknown functional parameters
  - $\varepsilon_1(t), \ldots, \varepsilon_n(t)$ are i.i.d. centered random variables $\in L^2(I_2)$
  - $\varepsilon_i(t)$ and $X_i(s)$ are uncorrelated

Assumptions (cont.)

- For generic interval I set $L^2(I)$ is equipped with usual inner product $\langle \phi, \psi \rangle = \int_I \phi(t)\psi(t)dt$, $\phi, \psi \in L^2(I)$ and associated norm $\|\phi\| = \langle \phi, \phi \rangle^{1/2}$.
- We often omit arguments of functional variables and parameters and write X_i, Y_i, ε_i and β instead of $X_i(s)$, $Y_i(t)$, $\varepsilon_i(t)$ and $\beta(s, t)$
- Recall the model

$$Y_i(t) = \alpha(t) + \int_{I_1} X_i(s)\beta(s, t)\, ds + \varepsilon_i(t) \quad t \in I_2,\; i = 1, \ldots, n \quad (1)$$

  - X_i's represent a weekdays curves
  - Y_i's represent a weekend curves,
    or a weekday curve, in which case $Y_i = X_{i+1}$
  - $\alpha(t)$ and $\beta(s, t)$ are unknown functional parameters
  - Model (1) corresponds to an ARH(1)

Estimating β

Let $B_j = (B_{j1}, \ldots B_{jd_j})', j = 1, 2$ denote normalized B-splines basis of spline space $\mathcal{S}_{q_j k_j}(I_j)$ of degree q_j defined on interval I_j with $k_j - 1$ equidistant interior knots and $d_j = k_j + q_j$ being dimension of $\mathcal{S}_{q_j k_j}(I_j)$.

"Exact" estimator $\widehat{\beta}$ of β is bivariate spline

$$\widehat{\beta}(s,t) = \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} \widehat{\theta}_{kl} B_{1k}(s) B_{2l}(t) = \mathbf{B_1}'(s) \widehat{\boldsymbol{\Theta}} \mathbf{B_2}(t), \; s \in I_1, \; t \in I_2. \quad (2)$$

where

$$\widehat{\boldsymbol{\Theta}} = \mathop{\arg\min}_{\boldsymbol{\Theta} \in \mathbb{R}^{d_1 \times d_2}} \quad \frac{1}{n} \sum_{i=1}^{n} \left\| Y_i - \overline{Y} - \langle (X_i - \overline{X}), \mathbf{B_1}' \boldsymbol{\Theta} \mathbf{B_2} \rangle \right\|^2 + \varrho \, \mathrm{Pen}(m, \boldsymbol{\Theta}), \quad (3)$$

with penalty parameter $\varrho > 0$ and penalty term

$$\mathrm{Pen}(m, \boldsymbol{\Theta}) = \sum_{m_1=0}^{m} \frac{m!}{m_1!(m-m_1)!} \int_{I_2} \int_{I_1} \left[\frac{\partial^m}{\partial s^{m_1} \partial t^{m-m_1}} \mathbf{B_1}'(s) \boldsymbol{\Theta} \mathbf{B_2}(t) \right]^2 ds dt$$

Estimating Θ

Using Kronecker product notation, we can write

$$\operatorname{vec}\widehat{\boldsymbol{\Theta}} = \left[\widehat{C}_\varrho + \varrho P^{(m)}\right]^{-1} \operatorname{vec}\widehat{D}, \tag{4}$$

where

$$\widehat{C}_\varrho = P_{\mathbf{2}}^{(\mathbf{0})'} \otimes \left(\widehat{C} + \varrho P_{\mathbf{1}}^{(m)}\right), \quad P^{(m)} = \sum_{m_1=0}^{m-1} \binom{m}{m_1} P_{\mathbf{2}}^{(m-m_1)'} \otimes P_{\mathbf{1}}^{(m_1)},$$

with

$$\widehat{D} = (\widehat{d}_{kl}) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \qquad \widehat{d}_{kl} = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, B_{1k}\rangle \langle Y_i, B_{2l}\rangle,$$

$$\widehat{C} = (\widehat{c}_{kk'}) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}, \quad \widehat{c}_{kk'} = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, B_{1k}\rangle \langle X_i B_{1k'}\rangle,$$

$$P_j^{(m_1)} = (p^j_{kk'}) \in \mathbb{R}^{d_j} \times \mathbb{R}^{d_j}, \quad p^j_{kk'} = \langle B_{jk}^{(m_1)}, B_{jk'}^{(m_1)}\rangle, \ j = 1, 2$$

Approximative solution

- Alternatively one can approximate exact solution by a simpler matrix version $\widetilde{\mathbf{\Theta}}$ if $\text{Pen}(m, \mathbf{\Theta})$ in minimization task (3) is replaced by

$$\widetilde{\text{Pen}}(m, \mathbf{\Theta}) = \int_{I_2}\int_{I_1}\left\{\left[B_{\mathbf{1}}^{(m)\prime}\mathbf{\Theta} B_{\mathbf{2}}^{(0)}\right]^2 + \left[B_{\mathbf{1}}^{(0)\prime}\mathbf{\Theta} B_{\mathbf{2}}^{(m)}\right]^2\right\} ds\, dt.$$

Matrix of unknown parameters $\mathbf{\Theta}$ can be estimated as:

$$\widetilde{\mathbf{\Theta}} = -\left[\widehat{C} + \varrho P_{\mathbf{1}}^{(m)}\right]^{-1} P_{\mathbf{1}}^{(0)}\widetilde{C} P_{\mathbf{2}}^{(m)} P_{\mathbf{2}}^{(0)\,-1} + \widetilde{C}, \qquad (5)$$

with

$$\widetilde{C} = \left[\widehat{C} + \varrho P_{\mathbf{1}}^{(m)}\right]^{-1} \widehat{D} P_{\mathbf{2}}^{(0)\,-1}.$$

- **Approximative matrix estimator** $\widetilde{\beta}(s,t)$ of (functional parameter) $\beta(s,t)$ is

$$\widetilde{\beta}(s,t) = B_{\mathbf{1}}'(s)\widetilde{\mathbf{\Theta}} B_{\mathbf{2}}(t)$$

- Numerical calculations were performed using an algorithm discussed by Benner in Parallel Algorithms Appl. **17**, 2002.

Estimating intercept

Intercept parameter α can be estimated either by

$$\widehat{\alpha}(t) = \overline{Y}(t) - \int_{I_1} \widehat{\beta}(s,t)\overline{X}(s)\,ds, \qquad \forall t_2 \in I_2, \tag{6}$$

or approximated by $\widetilde{\alpha}(t)$ if $\widetilde{\beta}$ is used instead of $\widehat{\beta}$ in (6).

****

Recall the model

$$Y_i(t) = \alpha(t) + \int_{I_1} X_i(s)\beta(s,t)\,ds + \varepsilon_i(t) \quad t \in I_2,\ i = 1,\ldots,n$$

Choice of parameters

Numerical calculation of $\widehat{\beta}$ and $\widehat{\alpha}$ requires proper choice of several parameters:

1. Order q_j of splines
2. Order of derivatives m
3. Numbers of knots k_j
4. Penalization parameter ρ

Choice of parameters

Numerical calculation of $\widehat{\beta}$ and $\widehat{\alpha}$ requires proper choice of several parameters:

1. Order q_j of splines
2. Order of derivatives m
3. Numbers of knots k_j
4. Penalization parameter ρ

- Order of splines q_j and derivatives m do not play important role compared to k_j and ρ
  $\Rightarrow$ choice $q_j = 3, 4$ and $m = 2$ appeared appropriate
- Concerning number of knots k_j and penalization parameter ρ
  $\Rightarrow$ Reasonable strategy is to fix it large (to prevent oversmoothing) while controlling degree of smoothness of $\widehat{\beta}$ with ρ.
  General suggestion does not exists $\Rightarrow$ tuning is necessary

Choice of parameters (cont.)

In practical and simulation experiments we used:

- $15 \le k \le 30$
- $\widehat{\rho}$ using leave-one-out cross-validation criterion

$$\mathrm{cv}(\varrho) = \sum_{i=1}^{n} \int_{I_2} \left[Y_i(t) - \int_{I_1} \widehat{\beta}_i(s,t) X_i(s)\, ds \right]^2 dt \qquad (7)$$

  $\widehat{\beta}_i(s,t)$ is obtained from data with i-th pair (X_i, Y_i) omitted

- Alternative – computationally faster – estimate $\widetilde{\rho}$ is obtained replacing in (7) exact solution $\widehat{\beta}_i$ with approximative solution $\widetilde{\beta}_i$

Remark: According our experience approximative criterion provides in many cases estimate very close to the one obtained by minimizing (7)

Elimination of trend and seasonality

To eliminate (and estimate) trend(s), we used one-sided kernel smoother. Let $Z_1, \ldots, Z_N$ be discrete observations of underlying time-continuous process with $Z_j = Z(t_j)$. We estimated trend by

$$\widehat{\nu}_j \;=\; \widehat{\nu}(t_j) \;=\; \sum\nolimits_{k=j-h+1}^{j} \omega_k(j;h) Z_k \tag{8}$$

with Epanechnikov kernel

$$\omega_k(j;h) = \frac{1-(k-j)^2/h^2}{\sum_{l=j-h+1}^{j}\left(1-(k-j)^2/h^2\right)} \tag{9}$$

Elimination of trend and seasonality

To eliminate (and estimate) trend(s), we used one-sided kernel smoother. Let $Z_1, \ldots, Z_N$ be discrete observations of underlying time-continuous process with $Z_j = Z(t_j)$. We estimated trend by

$$\widehat{\nu}_j = \widehat{\nu}(t_j) = \sum_{k=j-h+1}^{j} \omega_k(j;h) Z_k \tag{8}$$

with Epanechnikov kernel

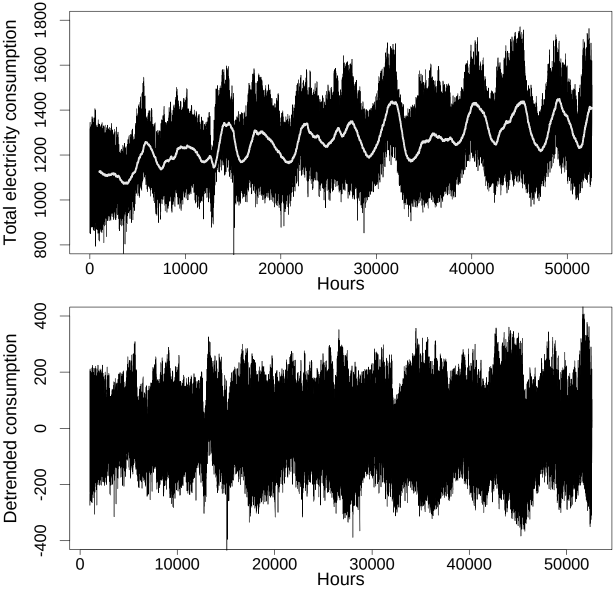
$$\omega_k(j;h) = \frac{1-(k-j)^2/h^2}{\sum_{l=j-h+1}^{j}\left(1-(k-j)^2/h^2\right)} \tag{9}$$

Why?

1. We essentially focus on functional data modelling
2. Kernel smoother is a well-known and intuitive nonparametric tool
3. Its performance can easily be controlled by smoothing parameter h

Remark: LOESS gives approximately the same results.

Data detrending



Detailed prediction

Nonparametric trend estimator can be extended to cover whole required time interval, e.g. $(T; T+48]$ for the weekend prediction, on a sufficiently fine time-grid.

- Let $t_1, \ldots, t_p$ denote time moments of interest. Then we :
- Start with $\widehat{Z}_{T+t_1}$.
- Add estimated $\widehat{Z}_{t_1}$ to the observed data
- Evaluate $\widehat{Z}_{T+t_2}$ profiting from the knowledge of $\widehat{Z}_{t_1}$.
- Recursively repeat trend estimation.

Generally

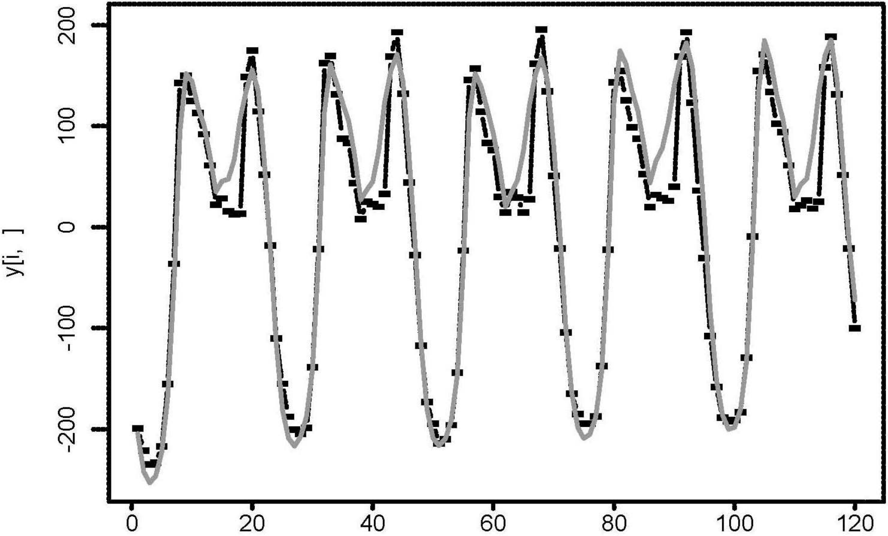
$$\widehat{Z}_{T+t_j} = \widehat{Y}(t_j \mid T) + \widetilde{\nu}(t_j \mid T)$$

with

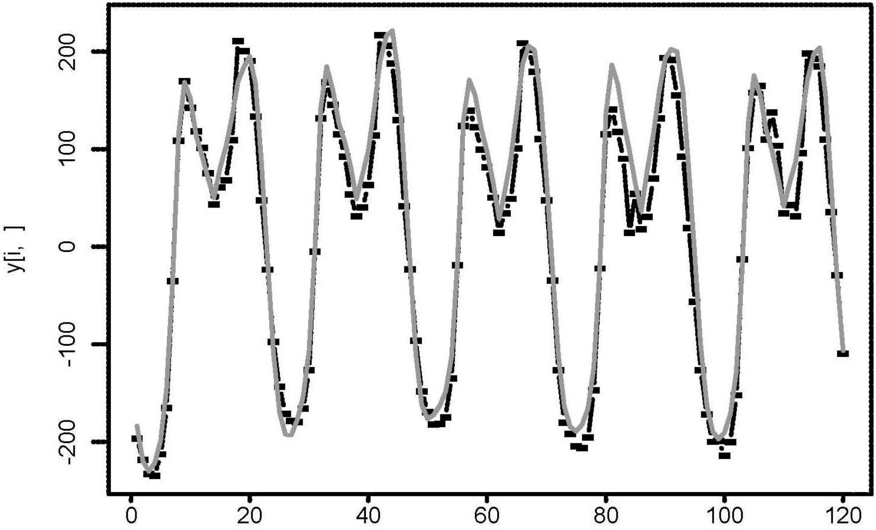
$$\widetilde{\nu}(t_j \mid T) = \sum_{k \in [T+t_j-h; T]} \widetilde{\omega}_k(T+t_j; h) Z_k + \sum_{k=1}^{j-1} \widetilde{\omega}_{T+t_k}(T+t_j; h) \widehat{Z}_{T+t_k}$$

Weights $\widetilde{\omega}$ are based on pooled sample $Z_1, \ldots, Z_T, \widehat{Z}_{T+t_1}, \ldots, \widehat{Z}_{T+t_{j-1}}$.

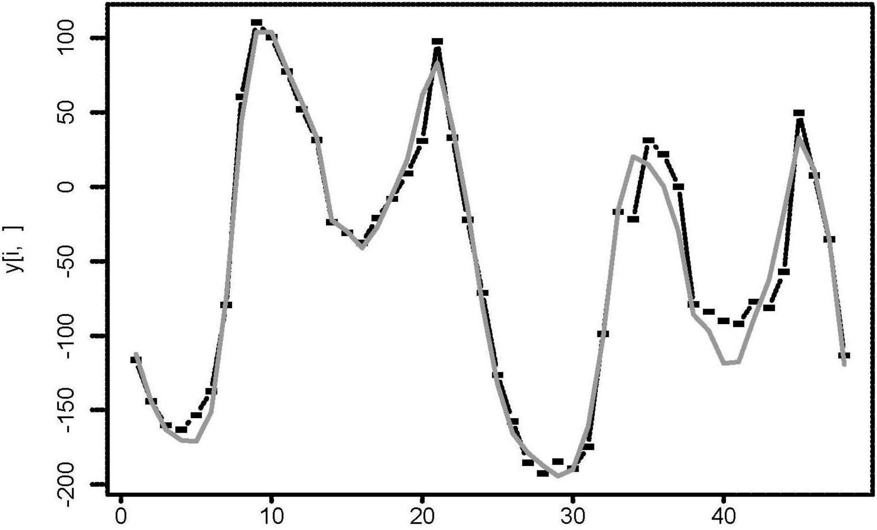
Example of prediction



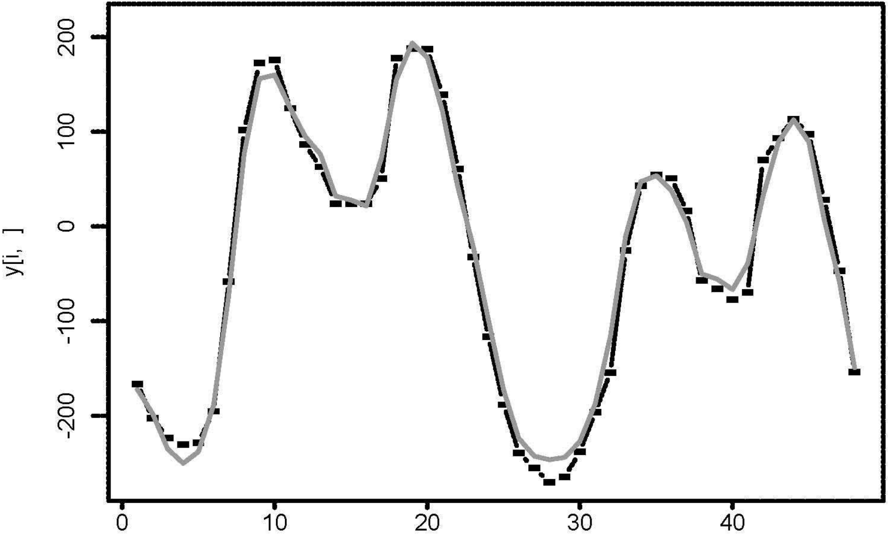
Example of prediction



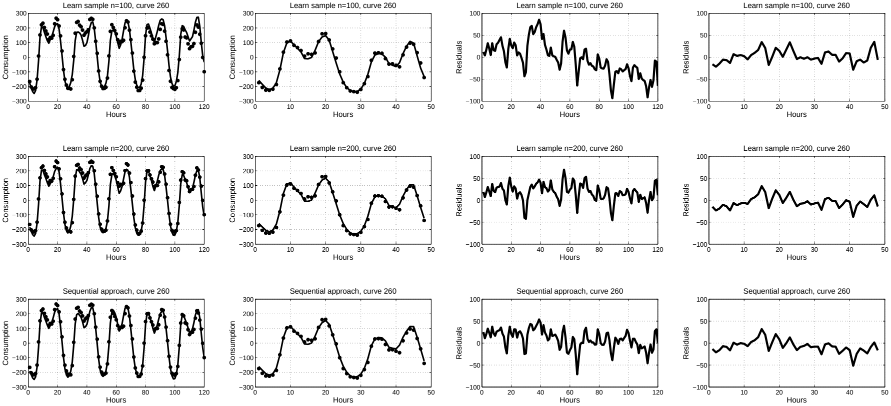
Example of prediction



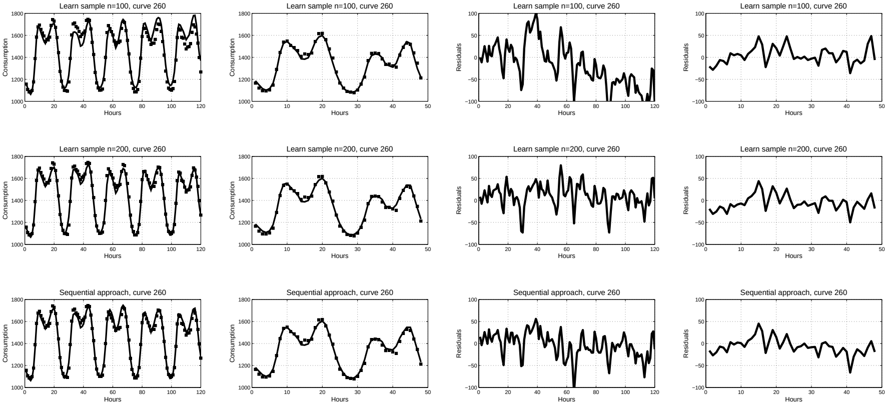
Example of prediction



Prediction of detrended data (► details)



Prediction of electricity data (▶ details)



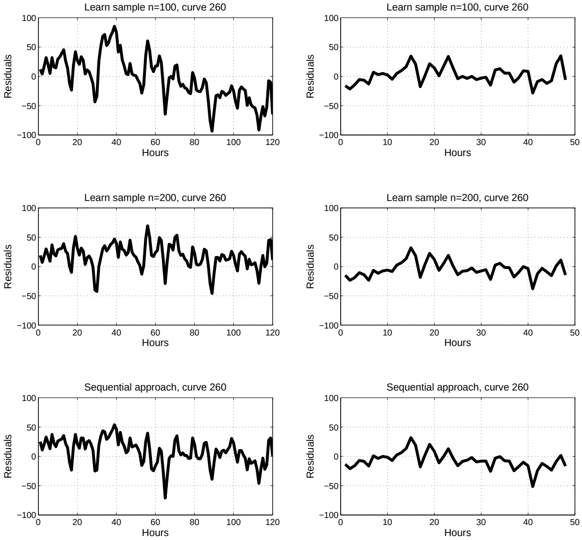
Conclusions

- Approximating matrix solution is competitive with the exact estimator and, as concerns data fitting, behaves satisfactorily.
- If one primarily focuses on the functional parameter estimation, the exact solution should be preferred as it is more stable as concerns tuning parameters of the method.
- Matrix approach can still be used throughout the cross-validation procedure at least as the pivot parameter, whose neighborhood is then seek throughout by the exact method.
- In many situations a very small number of knots is sufficient to obtain good estimators. As the matrix method behaves well and is fast, it is worth performing estimation for several knot setups – eventually a kind of cross-validation can be used for the knots as well.
- Interesting is also the case of errors-in-variables due to, e.g., not exact predictor registering, for which a presmoothing of the curves or functional total least squares might be involved.

THANKS

Residuals of detrended data

Return back



Residuals for predicted electricity data

Return back