ELECTRICITY CONSUMPTION PREDICTION

Jaromír Antoch

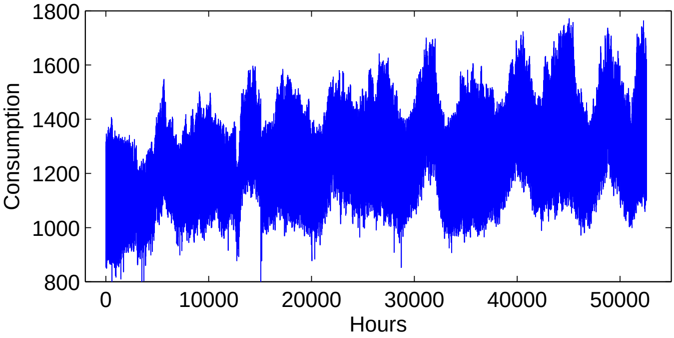
ENERGY DAYS 2025

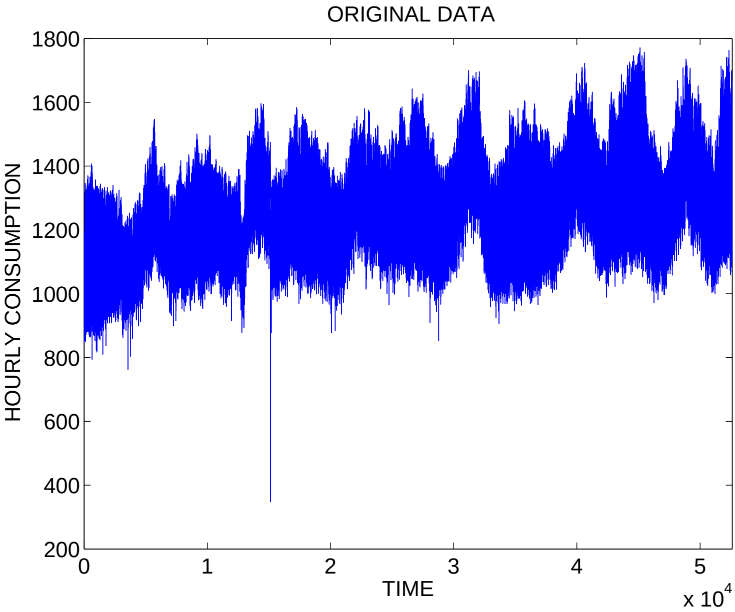
GOALS

- A 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 the effect of a functional variable on a functional response by means of functional linear regression models when the slope function is estimated with tensor product splines.
- The model serves to analyze a real data set concerning electricity consumption in Sardegna.
- Computational issues are addressed.

DATA

The model serves to analyze a real data set concerning electricity consumption in Sardegna. Data set consists of 52 584 values of electricity consumption collected every hour within the period from January 1, 2000, till December 31, 2005.

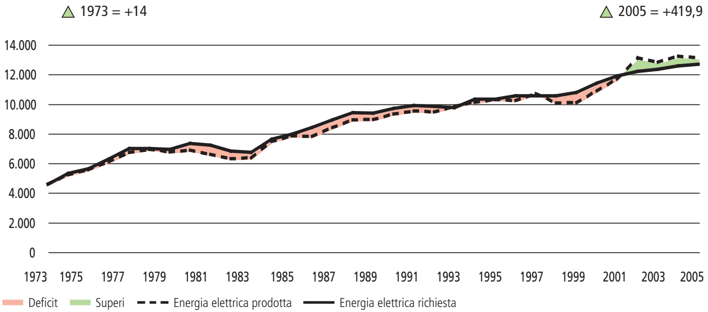



ORIGINAL DATA
1800
1600
1400
1200
1000
800
600
400
200
HOURLY CONSUMPTION
0
1
2
3
4
5
TIME
x 10^4

OFFICIAL DATA SARDEGNA

Energia richiesta

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



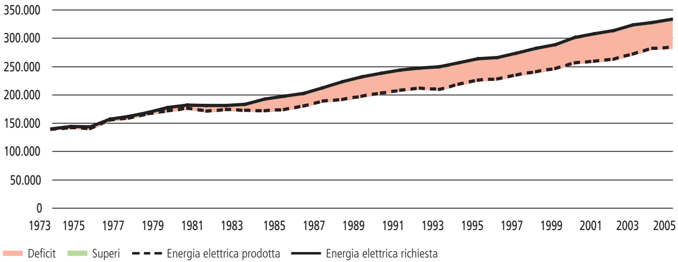
Consumi: complessivi 12.036,7 GWh; per abitante 7.286 kWh

OFFICIAL DATA ITALY

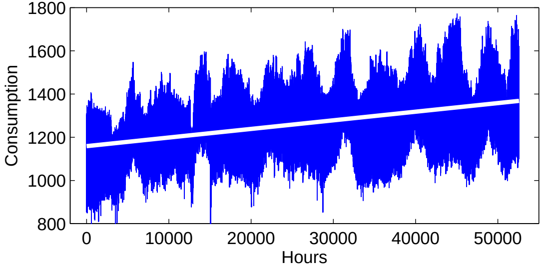
Energia richiesta

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

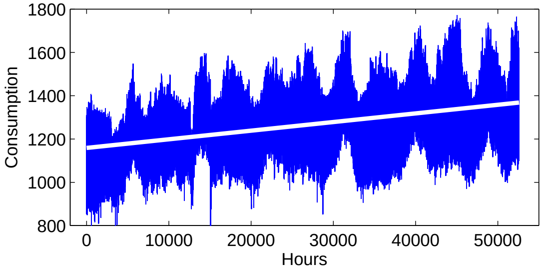
△ 1973 = -879 △ 2005 = -49.154,5

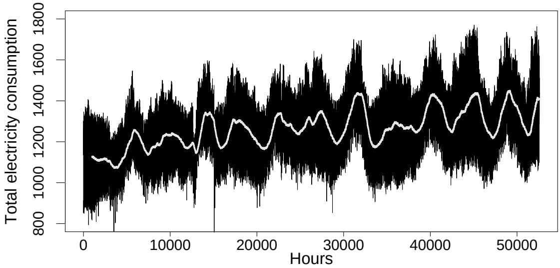


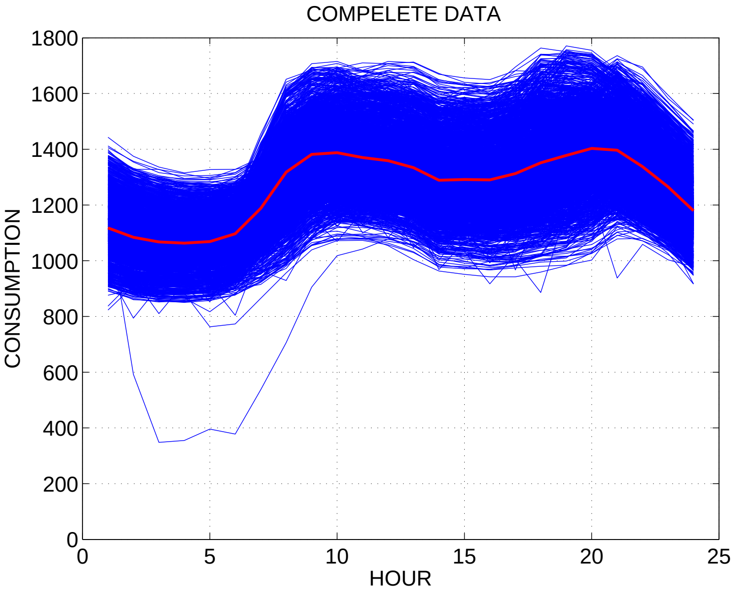
Consumi: complessivi 309.816,8 GWh; per abitante 5.286 kWh

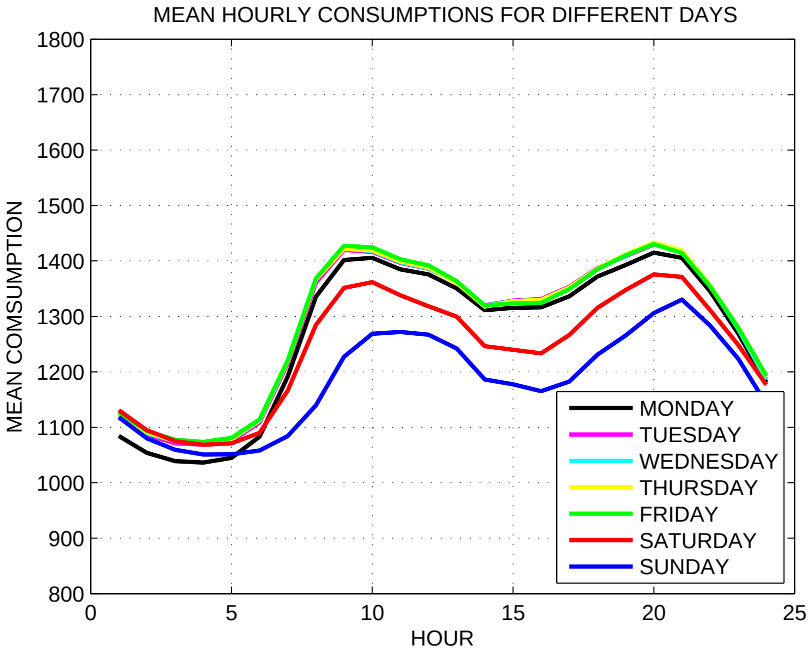


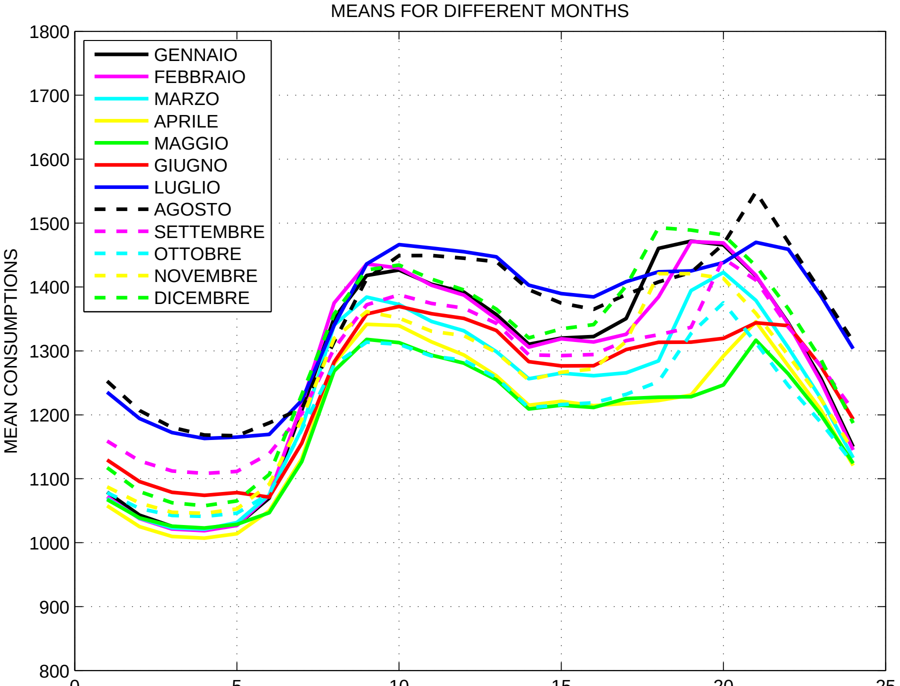
Consumption
1800
1600
1400
1200
1000
800
0
10000
20000
30000
40000
50000
Hours

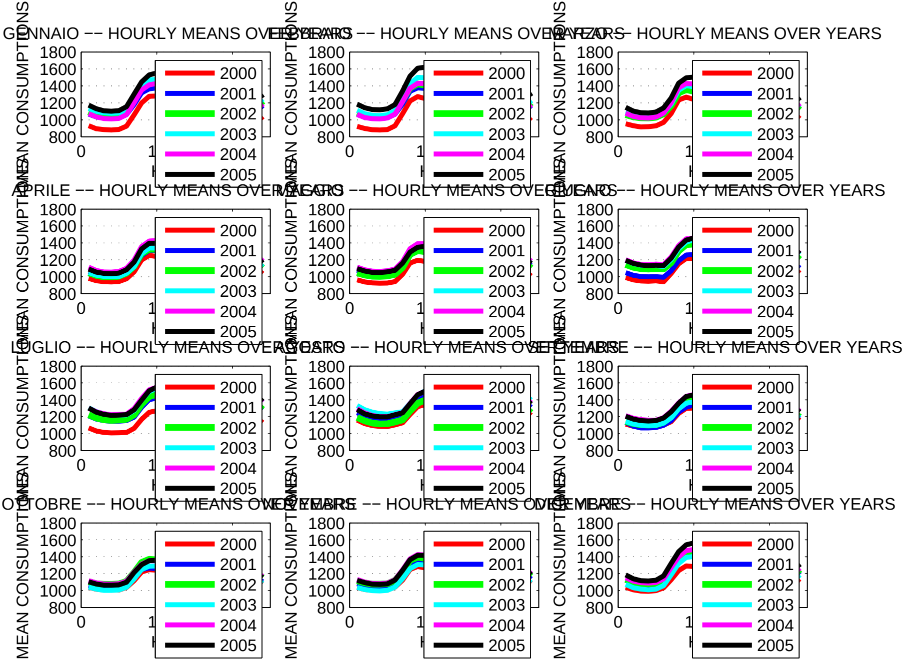

Consumption
Hours
1800
1600
1400
1200
1000
800
0
10000
20000
30000
40000
50000

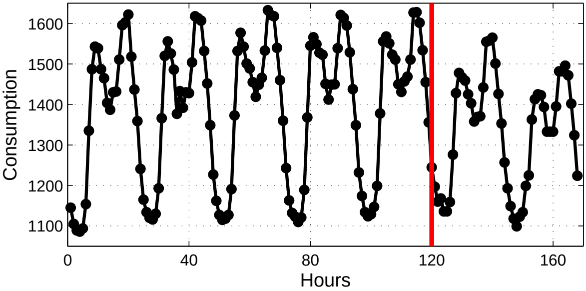

Total electricity consumption
Hours
800
1000
1200
1400
1600
1800
0
10000
20000
30000
40000
50000

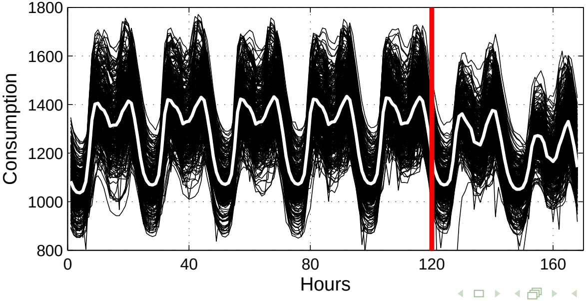

COMPELETE DATA
CONSUMPTION
HOUR
0
200
400
600
800
1000
1200
1400
1600
1800
0
5
10
15
20
25


MEAN HOURLY CONSUMPTIONS FOR DIFFERENT DAYS
MEAN COMSUMPTION
HOUR
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
0
5
10
15
20
25
MONDAY
TUESDAY
WEDNESDAY
THURSDAY
FRIDAY
SATURDAY
SUNDAY


MEANS FOR DIFFERENT MONTHS
MEAN CONSUMPTIONS
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
0
5
10
15
20
25
GENNAIO
FEBBRAIO
MARZO
APRILE
MAGGIO
GIUGNO
LUGLIO
AGOSTO
SETTEMBRE
OTTOBRE
NOVEMBRE
DICEMBRE


GENNAIO -- HOURLY MEANS OVER YEARS
-- HOURLY MEANS OVER YEARS
APRILE -- HOURLY MEANS OVER YEARS
GIUGNO -- HOURLY MEANS OVER YEARS
LUGLIO -- HOURLY MEANS OVER YEARS
SETTEMBRE -- HOURLY MEANS OVER YEARS
OTTOBRE -- HOURLY MEANS OVER YEARS
DICEMBRE -- HOURLY MEANS OVER YEARS
MEAN CONSUMPTIONS
1800
1600
1400
1200
1000
800
0
2000
2001
2002
2003
2004
2005


Consumption
1600
1500
1400
1300
1200
1100
0
40
80
120
160
Hours


Consumption
1800
1600
1400
1200
1000
800
0
40
80
120
160
Hours

DATA PREPARATION

- The complete data series has been cut into 307 weeks for which the weekdays (Monday to Friday) and the weekends (Saturday and Sunday) have been separated.
- The reason for such a separation leading to two sets of discretized electricity consumption curves rests in the fact that we observe important differences between weekdays and weekend consumptions.
- The main interest lies in predicting either oncoming weekend or oncoming weekdays consumption curves if present weekdays consumption is known. In both cases, the (functional) predictor is the (discretized) curve of present weekdays consumption.

Model

- 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 the same probability space and taking values in some functional spaces.
- We consider the separable real Hilbert spaces $L^2(I_1)$ and $L^2(I_2)$ of square integrable functions defined on the compact intervals $I_1 \subset \mathbb{R}$ and $I_2 \subset \mathbb{R}$, which are equipped with the standard inner products.
- We focus on the functional linear relation

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

$i = 1, \ldots, n$,where $\alpha(t) \in L^2(I_2)$ and $\beta(s,t) \in L^2(I_1 \times I_2)$ are unknown functional parameters and $\varepsilon_1(t), \ldots, \varepsilon_n(t)$ stand for a sample of i.i.d. centered random variables taking values in $L^2(I_2)$, $\varepsilon_i(t)$ and $X_i(s)$ being uncorrelated.

Model (cont)

- For a generic interval I, the set $L^2(I)$ is equipped with its usual inner product $\langle \phi, \psi \rangle = \int_I \phi(t)\psi(t)dt$, $\phi, \psi \in L^2(I)$ and the associated norm $\|\phi\| = \langle \phi, \phi \rangle^{1/2}$.
- In what follows we often omit arguments of the functional variables and parameters and simply write X_i, Y_i, ε_i and β instead of $\{X_i(s), s \in I_1\}$, $\{Y_i(t), t \in I_2\}$, $\{\varepsilon_i(t), t \in I_2\}$ and $\{\beta(s,t), s \in I_1, t \in I_2\}$, respectively.
- Notice that for our electricity data the models reads as

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

  - X_i's represent a weekdays curves
  - Y_i's represent a weekend curves, or a weekday curve in which case $Y_i = X_{i+1}$
  - Model (1) corresponds to an ARH(1)

Estimator of β

Let $\boldsymbol{B_j} = \left(B_{j1}, \ldots B_{jd_j}\right)', j = 1, 2,$ 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 $\mathcal{S}_{q_j k_j}(I_j)$.

Our estimator $\widehat{\beta}$ of β is a bivariate spline

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

where

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

with a penalty parameter $\varrho > 0$ and the penalty term given by

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

Estimator of Θ

Using Kronecker product notation, we can write

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

where

$$\widehat{\boldsymbol{C}}_{\varrho} = \boldsymbol{P}_{\boldsymbol{2}}^{(0)'} \otimes \left(\widehat{\boldsymbol{C}} + \varrho \boldsymbol{P}_{\boldsymbol{1}}^{(\boldsymbol{m})}\right), \quad \boldsymbol{P}^{(\boldsymbol{m})} = \sum_{m_1=0}^{m-1} \binom{m}{m_1} \boldsymbol{P}_{\boldsymbol{2}}^{(\boldsymbol{m}-\boldsymbol{m}_1)'} \otimes \boldsymbol{P}_{\boldsymbol{1}}^{(\boldsymbol{m}_1)},$$

with

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

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

$$\boldsymbol{P}_{\boldsymbol{j}}^{(\boldsymbol{m}_1)} = (p^j_{kk'}) \in \mathbb{R}^{d_j} \times \mathbb{R}^{d_j}, \quad p^j_{kk'} = \langle B_{jk}^{(m_1)}, B_{jk'}^{(m_1)}\rangle, \ j = 1, 2.$$

Alternative 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

$$\widetilde{\text{Pen}}(m, \mathbf{\Theta}) = \int_{I_2}\int_{I_1} \left\{ \left[\boldsymbol{B}_1^{(m)\prime} \mathbf{\Theta} \boldsymbol{B}_2^{(0)} \right]^2 + \left[\boldsymbol{B}_1^{(0)\prime} \mathbf{\Theta} \boldsymbol{B}_2^{(m)} \right]^2 \right\} ds\, dt.$$

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

$$\widetilde{\mathbf{\Theta}} = -\left[\widehat{\boldsymbol{C}} + \varrho \boldsymbol{P}_1^{(m)} \right]^{-1} \boldsymbol{P}_1^{(0)} \widetilde{\boldsymbol{C}} \boldsymbol{P}_2^{(m)} \boldsymbol{P}_2^{(0)\,-1} + \widetilde{\boldsymbol{C}}, \tag{5}$$

with

$$\widetilde{\boldsymbol{C}} = \left[\widehat{\boldsymbol{C}} + \varrho \boldsymbol{P}_1^{(m)} \right]^{-1} \widehat{\boldsymbol{D}} \boldsymbol{P}_2^{(0)\,-1}.$$

- Approximating estimator of the functional parameter $\beta(s, t)$ is then

$$\widetilde{\beta}(s, t) = \boldsymbol{B}_1'(s) \widetilde{\mathbf{\Theta}} \boldsymbol{B}_2(t)$$

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

Estimation of intercept

- The intercept parameter α can be estimated either by:

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

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

Parameter Choice

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

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

- Orders q_j and m do not play important role compared to k_j and ρ $\Rightarrow$ it is usual to use $q_j = 3, 4$ and $m = 2$.
- It has been stressed in similar context that a reasonable strategy to choose k_j and ρ is to fix the number of knots k_j reasonably large, and thus both to prevent oversmoothing and to control degree of smoothness of the estimator with the parameter ρ.

Parameter Choice cont.

- In our practical and simulation experiments we have used value of k between 15 and 30, while ρ was chosen to minimize the leave-one-out cross-validation criterion

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

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

- An alternative and computationally faster criterion $\widetilde{\mathrm{cv}}(\varrho)$ can be used when replacing $\widehat{\beta}_i$ with $\widetilde{\beta}_i$. From our experience the approximating criterion provides in many cases a value close to the one obtained by minimizing the criterion (7). It can also be used to provide a pivotal parameter for the search of the minimizer of (7).

Trend and seasonality elimination

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

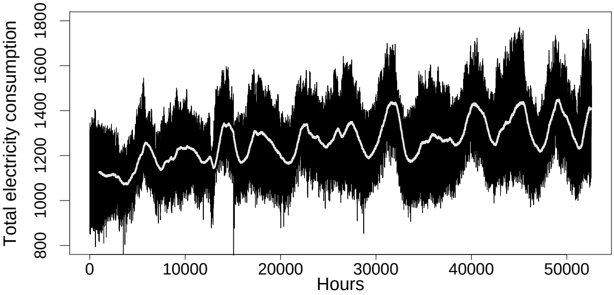
$$\widehat{\nu}_j = \widehat{\nu}(t_j) = \sum\nolimits_{k=j-h+1}^{j} \omega_k(j;h) Z_k \tag{8}$$

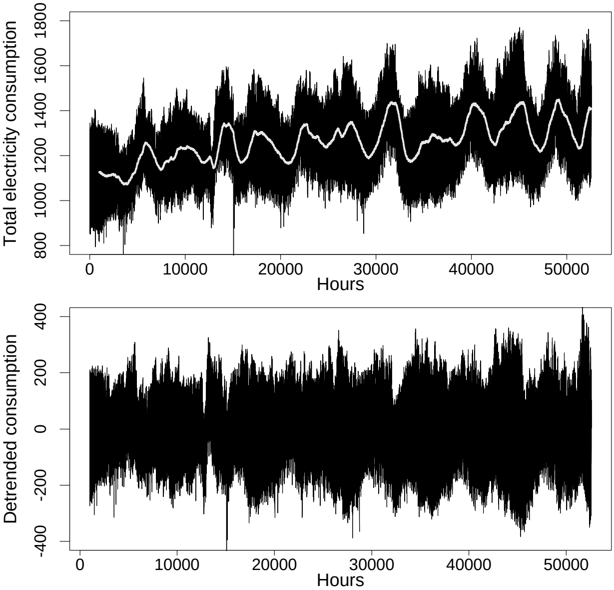
with Epanechnikov kernel

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

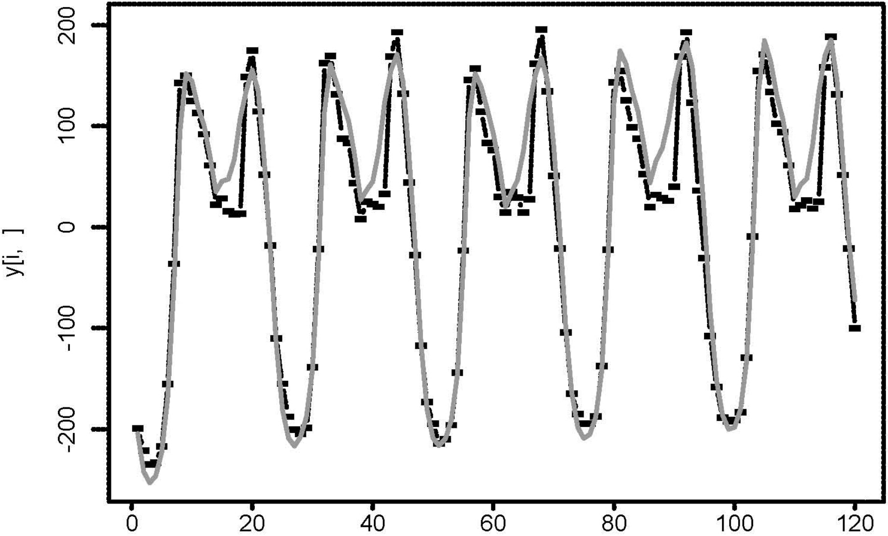
Why?

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

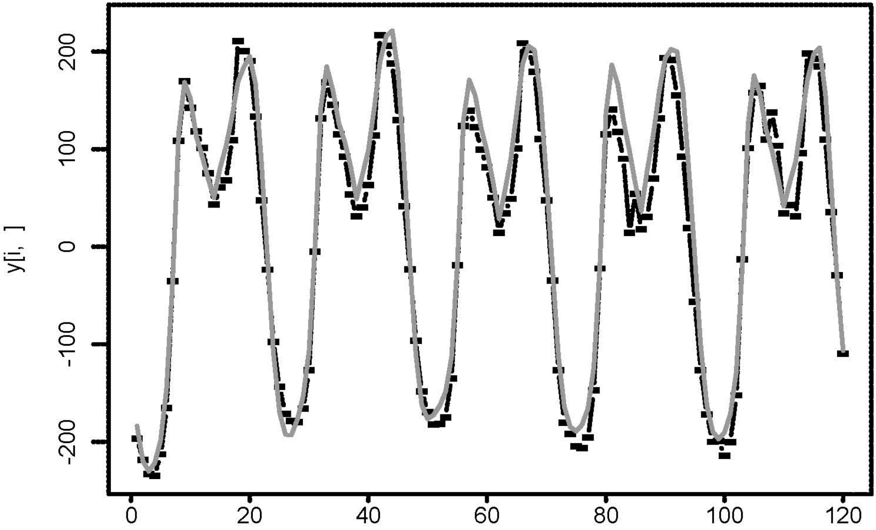

Total electricity consumption
800
1000
1200
1400
1600
1800
0
10000
20000
30000
40000
50000
Hours


Total electricity consumption
800
1000
1200
1400
1600
1800
0
10000
20000
30000
40000
50000
Hours
Detrended consumption
-400
-200
0
200
400
0
10000
20000
30000
40000
50000
Hours

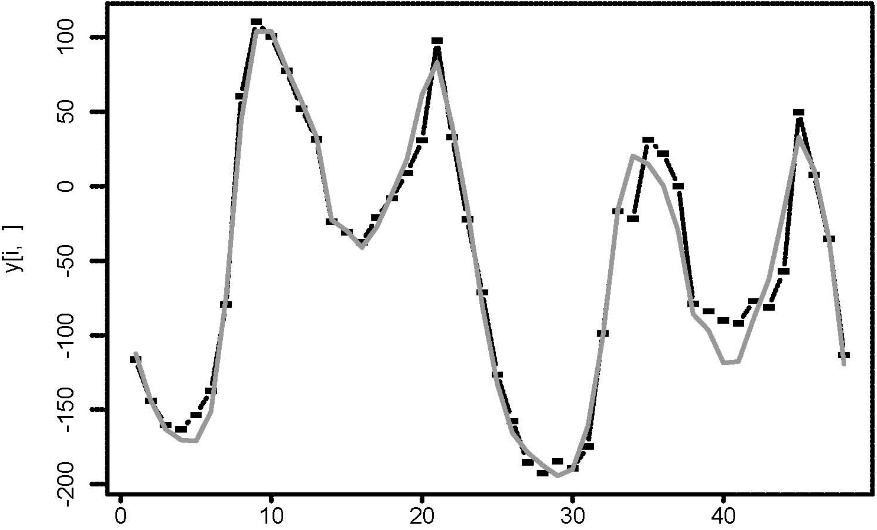
EXAMPLE OF PREDICTION



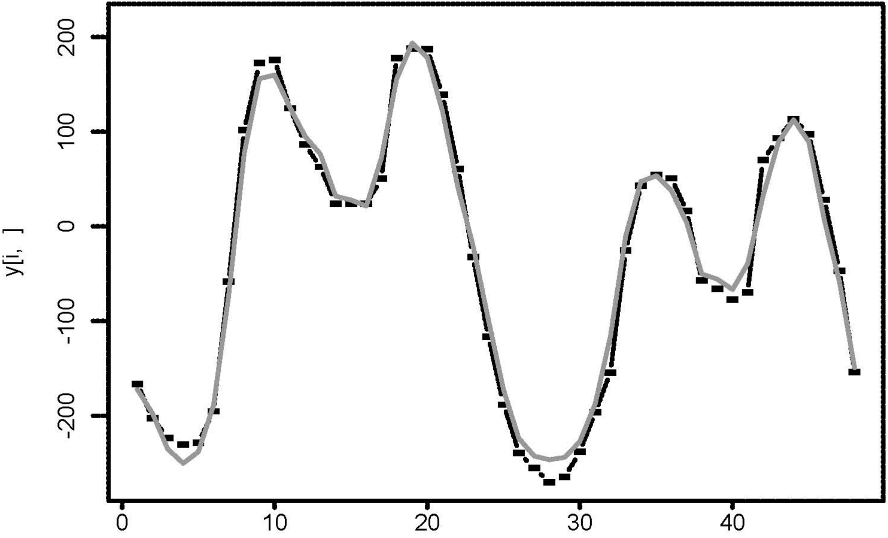
EXAMPLE OF PREDICTION



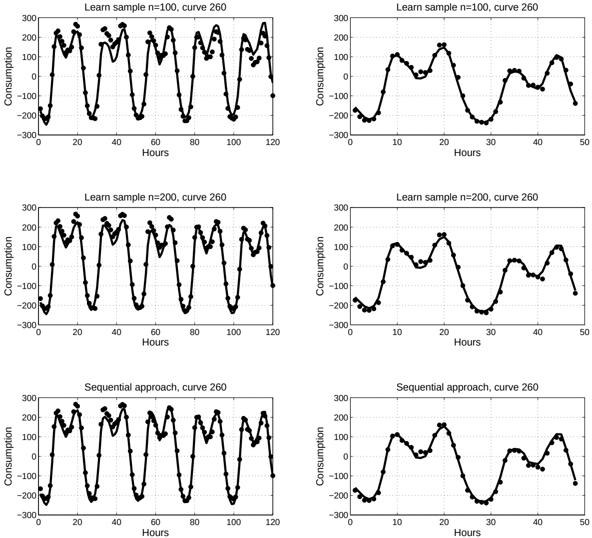
EXAMPLE OF PREDICTION



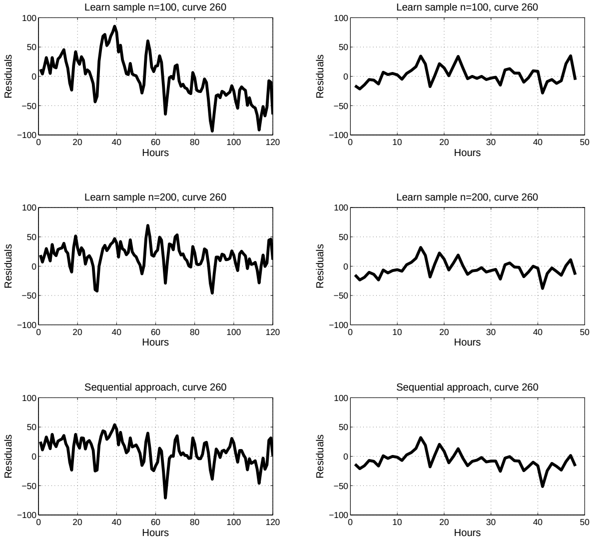
EXAMPLE OF PREDICTION



Prediction of detrended data



Residuals of detrended data



Prediction of data

Nonparametric trend estimator can be easily extended to cover whole required time interval, e.g. $(T; T+48]$ for the weekend prediction, on a sufficiently fine time-grid. Indeed, let $t_1, \ldots, t_p$ denote time moments of interest, e.g. the hours. Then we start with $\widehat{Z}_{T+t_1}$ as above, add the estimated $\widehat{Z}_{t_1}$ to the observed data, evaluate $\widehat{Z}_{T+t_2}$ profiting from the knowledge of $\widehat{Z}_{t_1}$ and recursively repeat the trend estimation. Generally,

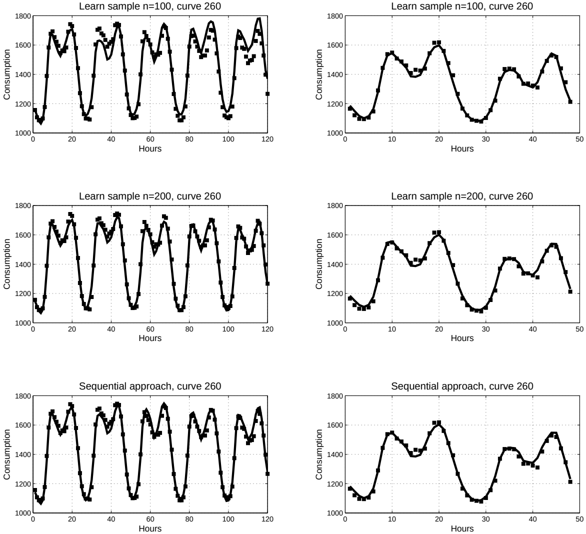
$$\widehat{Z}_{T+t_j} = \widehat{Y}(t_j \mid T) + \widetilde{\nu}(t_j \mid T)$$

with

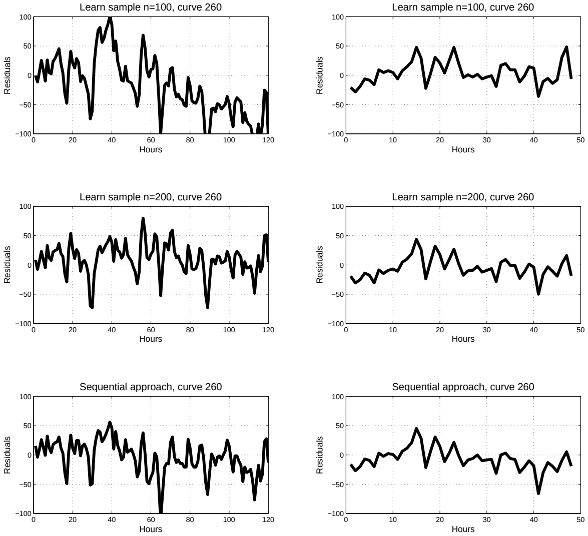
$$\widetilde{\nu}(t_j \mid T) = \sum_{k \in [T+t_j-h; T]} \widetilde{\omega}_k(T+t_j; h) Z_k + \sum_{k=1}^{j-1} \widetilde{\omega}_{T+t_k}(T+t_j; h) \widehat{Z}_{T+t_k}, \tag{10}$$

where the weights $\widetilde{\omega}$ are obtained analogously to (9) but are based on the pooled sample $Z_1, \ldots, Z_T, \widehat{Z}_{T+t_1}, \ldots, \widehat{Z}_{T+t_{j-1}}$.

Prediction of electricity data



Residuals for predicted data



Conclusions from simulations

- 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.
- The matrix approach, however, can still be used throughout the cross-validation procedure at least as the pivot parameter, whose neighborhood is then seek throught 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.

TERNA PREDICTION