Автоматизация научных исследований в машинном обучении

Вадим Викторович Стрижов

Московский физико-технический институт

Осенний семестр 2018

1. Customer demand sales forecasting as the example of error function statement.
2. Periodical time series: the one-model autoregressive problem.
3. Non-periodical time series: the mixed-model autoregressive problem.
4. Some problems of regression and classification.

The goal is to state the problem of the time series event forecasting as the auto-regressive problem.

The set of retailers problems:

- custom inventory,
- calculation of optimal insurance stocks,
- consumer demand forecasting.

The initial problem statement

There given:

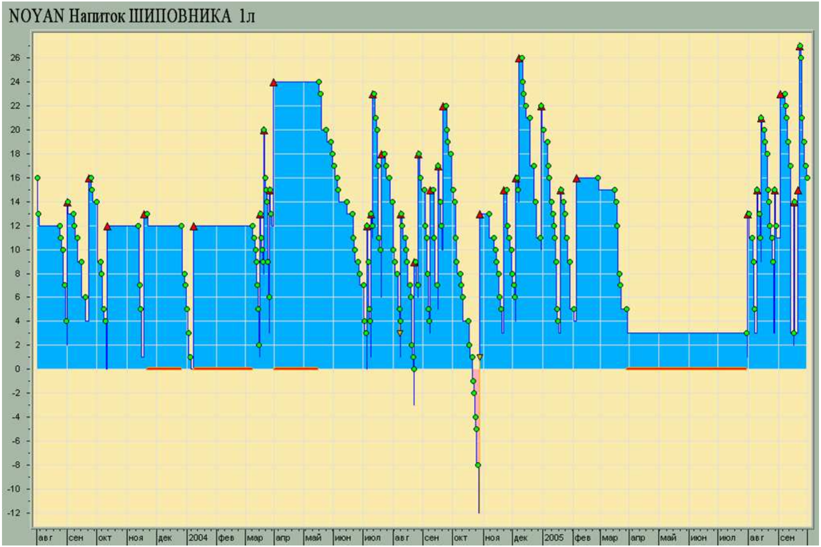
- time-scale,
- historical time series,
- additional time series;

the quality of forecasting:

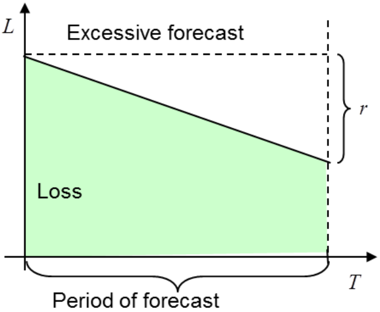
- minimum loss of money;

we must:

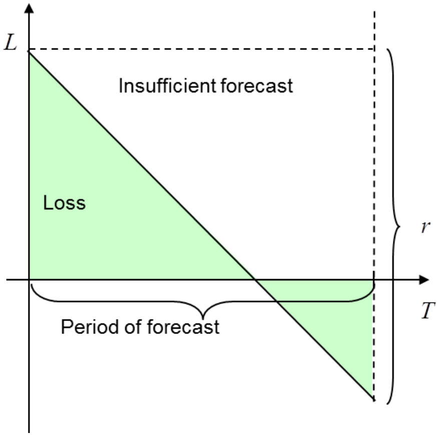
- forecast the time series.


NOYAN Напиток ШИПОВНИКА 1л
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
авг
сен
окт
ноя
дек
2004
фев
мар
апр
май
июн
июл
авг
сен
окт
ноя
дек
2005
фев
мар
апр
май
июн
июл
авг
сен

Excessive forecast



Insufficient forecast

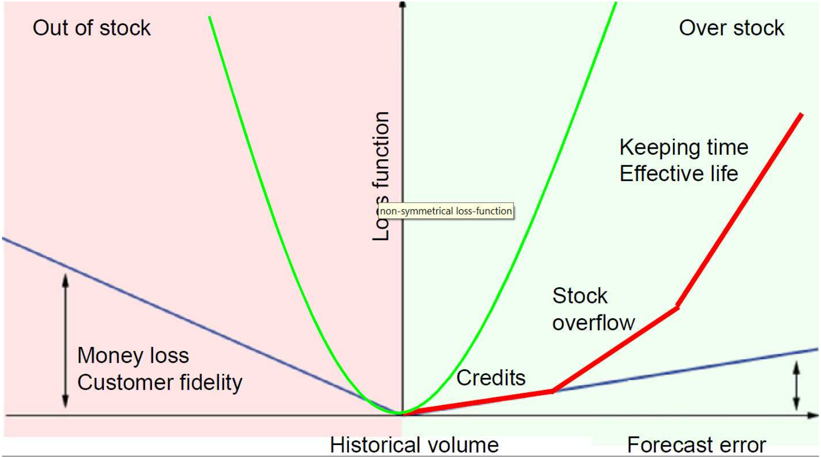


Loss function for the forecast error

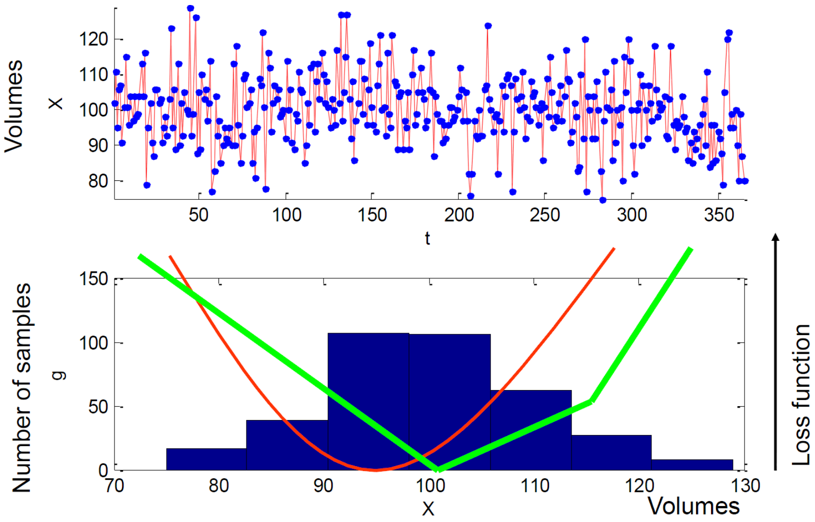
reflects the sales process and depends on the basis of the features of a particular trading network

- Symmetric quadratic function
- Module function
- Asymmetric function

Asymmetric loss function



- There is a historical time series of the volume off-takes (i.e. foodstuff).
- Let the time series be homoscedastic (its variance is the time-constant).
- Using the loss function one must forecast the next sample.



Volumes
x
120
110
100
90
80
50
100
150
200
250
300
350
t
Number of samples
g
150
100
50
0
70
80
90
100
110
120
130
x
Volumes
Loss function

Let there be given:

the historgam $H = \{X_i, g_i\}, i = 1, \dots, m;$

the loss function $L = L(Z, X);$

for example, $L = |Z - X|$ or $L = (Z - X)^2$.

The problem:

For given H and L, one must find the optimal forecast value $\tilde{X}$.

Solution:

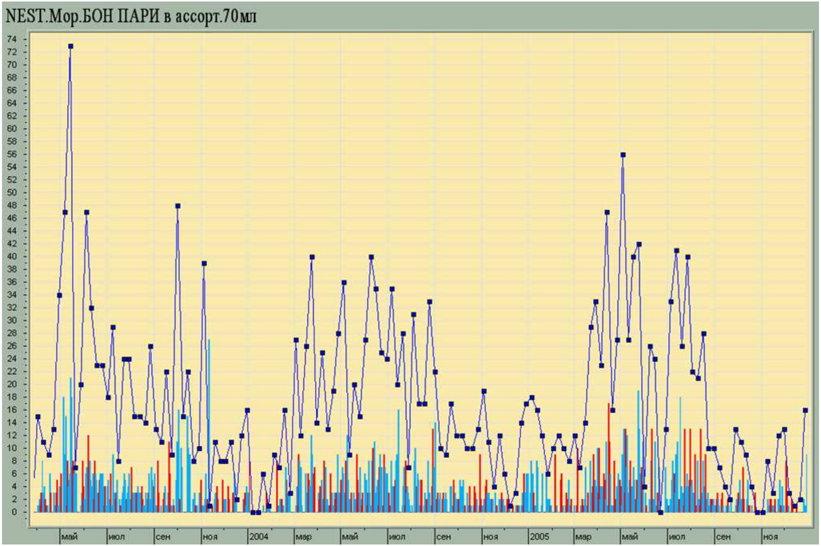
$$\tilde{X} = \arg \min_{Z \in \{X_1, \dots, X_m\}} \sum_{i=1}^{m} g_i L(Z, X_i).$$

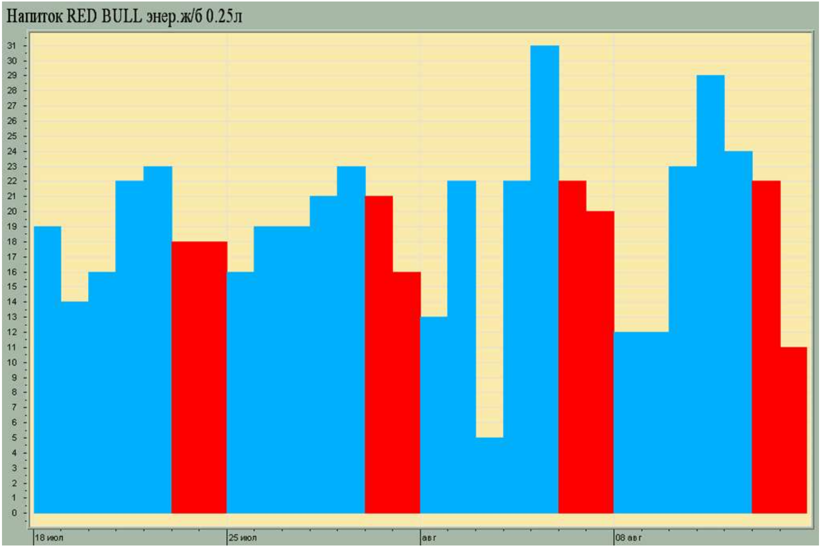
Result:

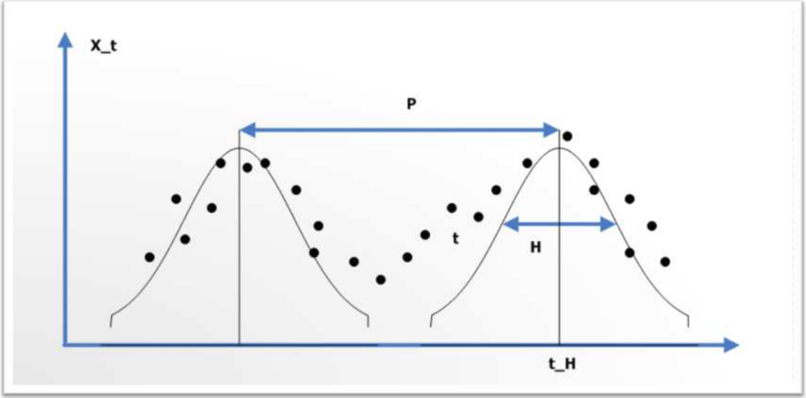
$\tilde{X}$ is the optimal forecast of the time series.

- There is a trend — total increase or decrease in sales volume,
- periodic component — week and year cycles,
- aperiodic component — promotional actions and holidays,
- life cycle of goods — mobile phones.

Year seasonality

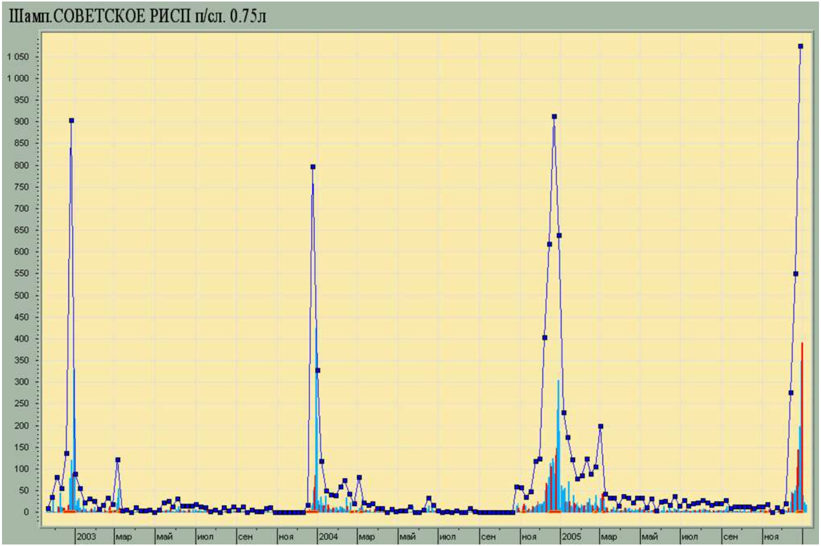



Напиток RED BULL энер.ж/б 0.25л
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
18 июл
25 июл
авг
08 авг

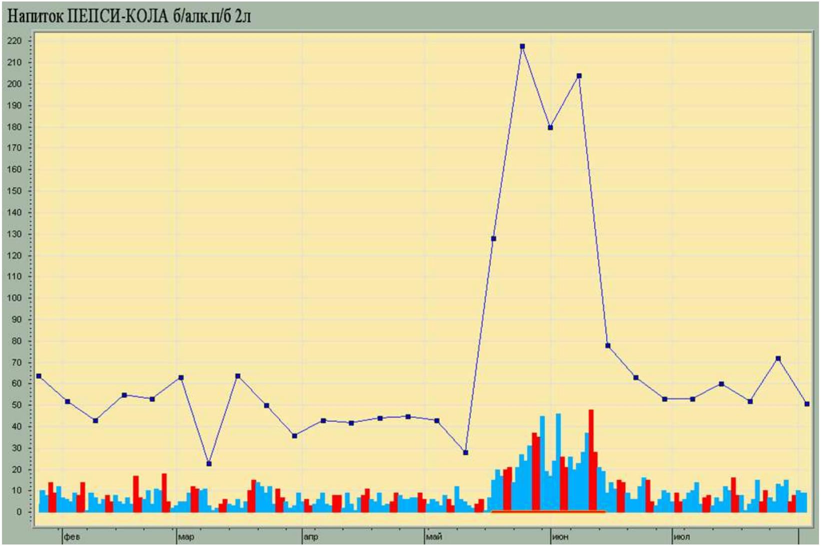


Weighting function includes neighborhood points to the sample set.

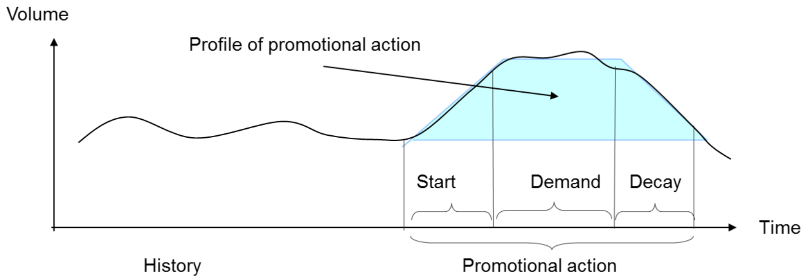
Holidays and week-ends



Promotional actions



Promotional profile extraction



- Hypothesis: the shape of the profile (excluding the profile parameters) does not depend of duration of the action.
- Problem: to forecast the customer demand during the promotional action.

Algorithmic composition

On forecasting one must consider:

- trend of time series,
- periodical components of time series,
- aperiodical components,

as well as the fact that the time series contain

- empty values — there is no information about the stock,
- empty values — new position on the stock,
- outliers and errors.

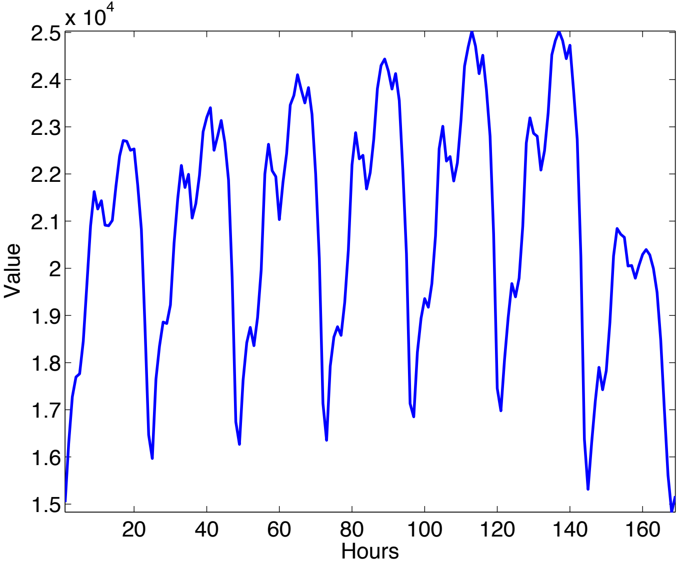
The periodic components of the multivariate time series

The time series:

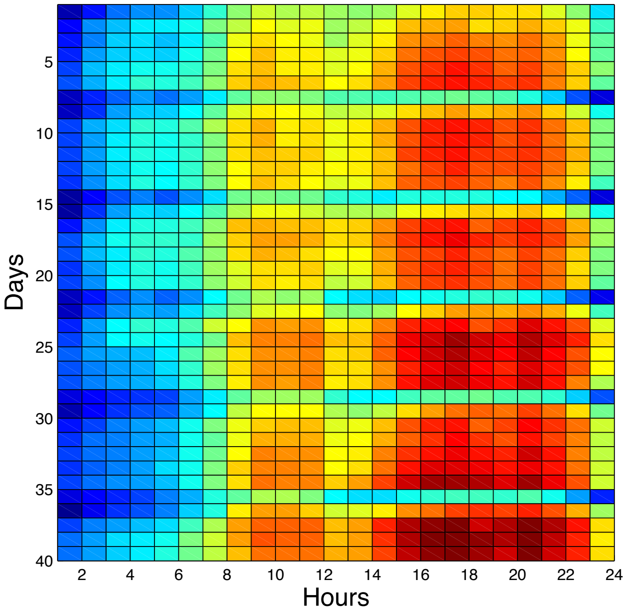
- energy price,
- consumption,
- daytime,
- temperature,
- humidity,
- wind force,
- holiday schedule.

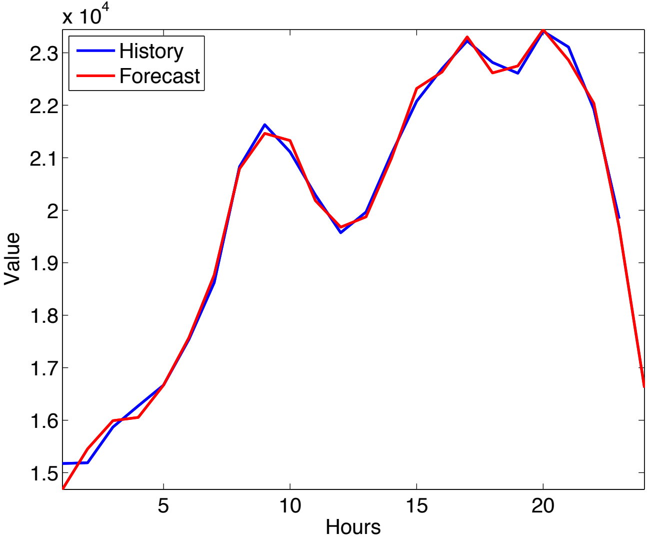
Periods:

- one year seasons (temperature, daytime),
- one week,
- one day (working day, week-end),
- a holiday,
- aperiodic events.


x 10^4
2.5
2.4
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
Value
20
40
60
80
100
120
140
160
Hours

The autoregressive matrix, five week-ends




x 10^4
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
Value
History
Forecast
5
10
15
20
Hours

Assume we have hourly data on price/consumption for three years. Then the matrix $\underset{(m+1)\times(n+1)}{X^*}$ is

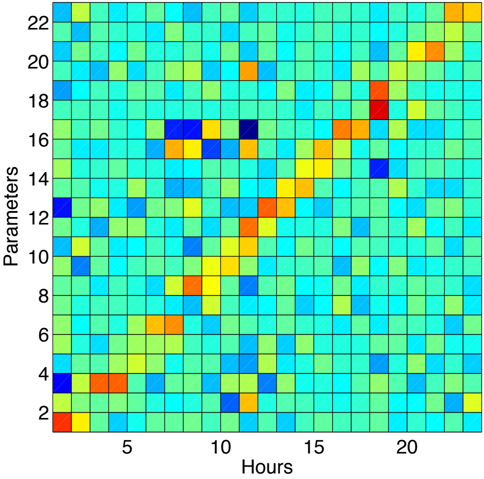
$$156 \times 168, \text{ in details: } 52\text{w} \cdot 3\text{y} \times 24\text{h} \cdot 7\text{d};$$

- for 6 time series the matrix X is 156×1008,
- for 4 primitive functions it is 156×4032,

$$m << n.$$

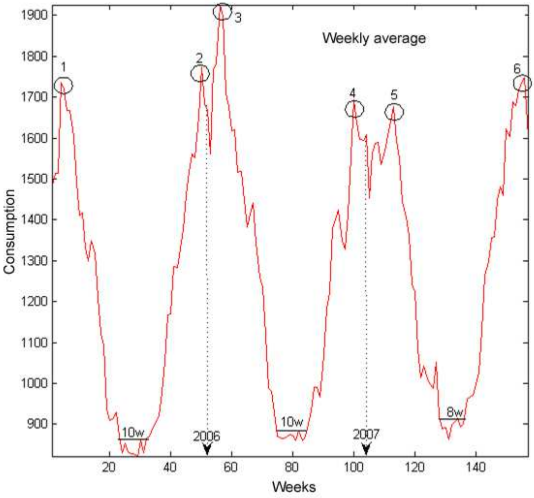
The autoregressive matrix could be considered as *ill-conditioned* and *multi-correlated*. The model selection procedure is required.

The color shows the value of a parameter for each hour.

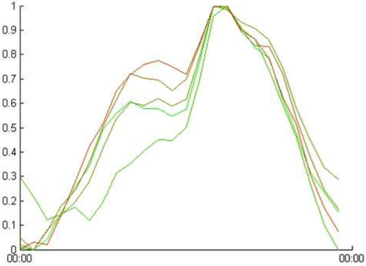


Estimate parameters $\mathbf{w}(\tau) = (X^\mathsf{T} X)^{-1} X^\mathsf{T} \mathbf{y}$, then calculate the sample $s(\tau) = \mathbf{w}^\mathsf{T}(\tau)\mathbf{x}_{m+1}$ for each τ of the next ($m+1$-th) period.

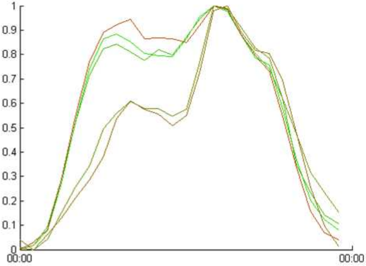
Structure of energy consumption



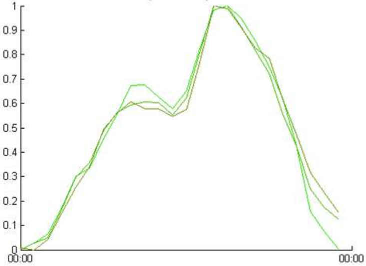
Similarity of daily consumption



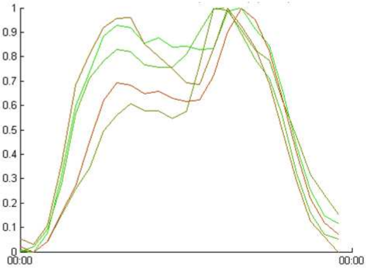
Five days of one week



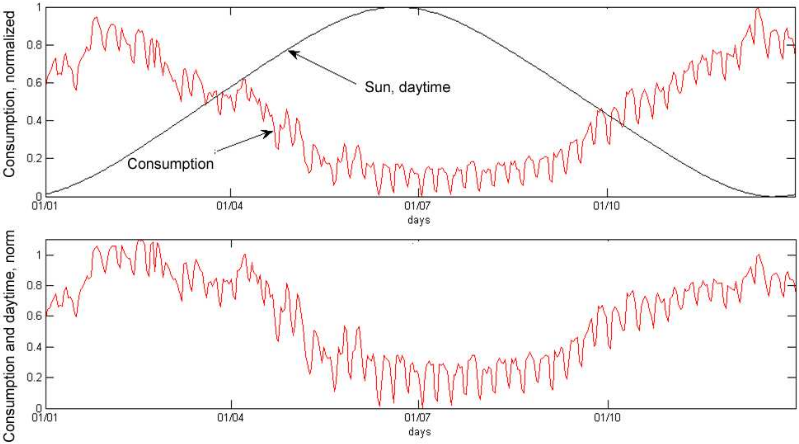
One day of five consequent weeks



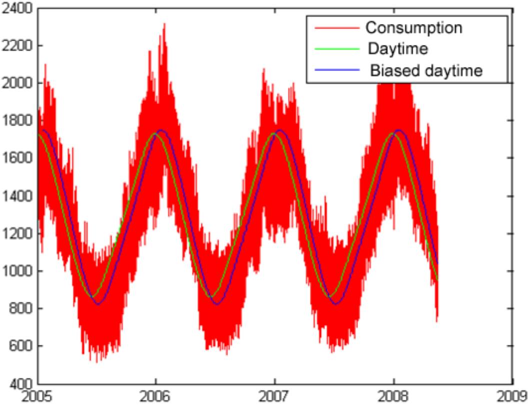
The same weekday and month of three years



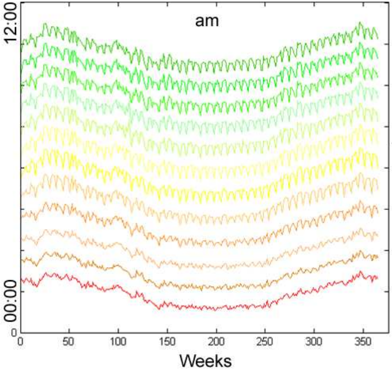
The same weekday of five months

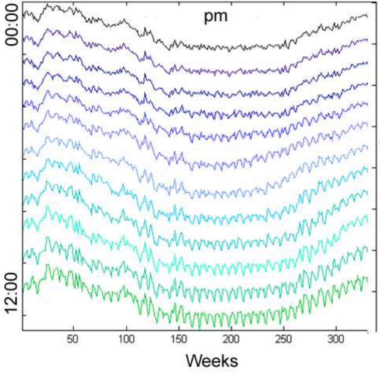

Consumption, normalized
1
0.8
0.6
0.4
0.2
0
01/01
01/04
01/07
01/10
days
Sun, daytime
Consumption
Consumption and daytime, norm
1
0.8
0.6
0.4
0.2
0
01/01
01/04
01/07
01/10
days

Biased and original daytime to fit consumption over years

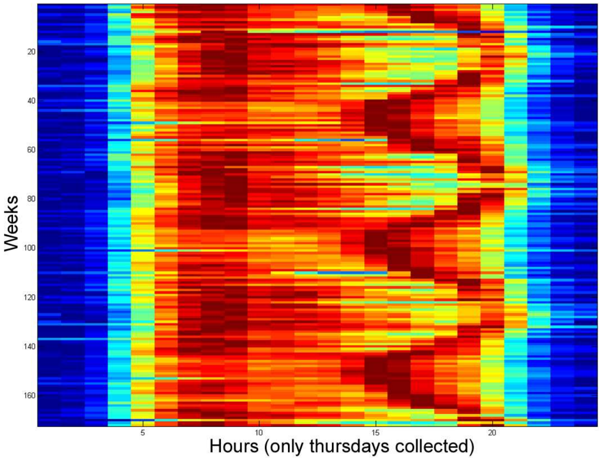


One-hour line, day-by-day during a year: autoregressive analysis

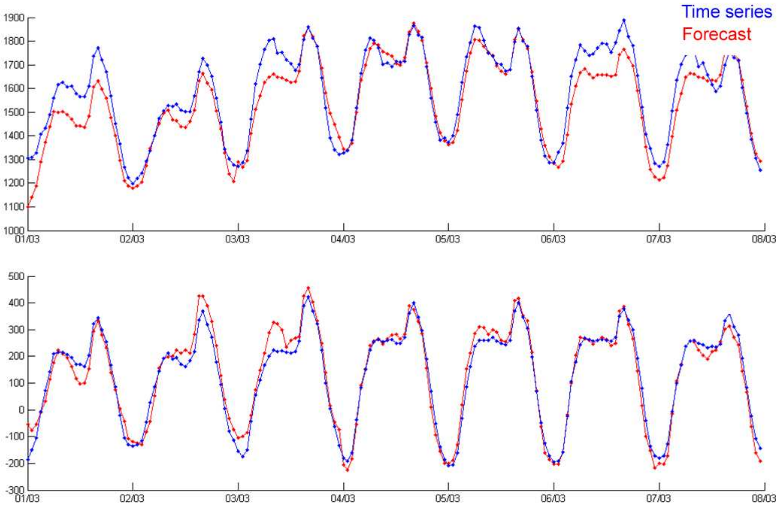


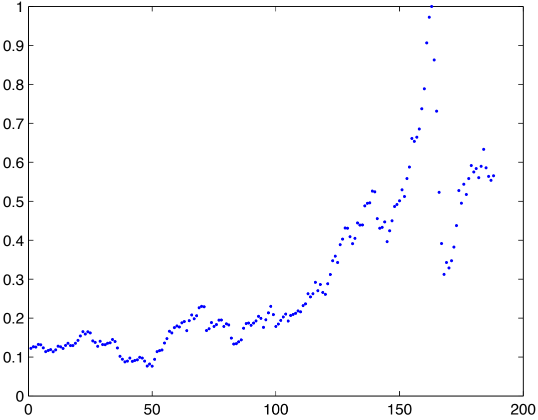


Daily similarity



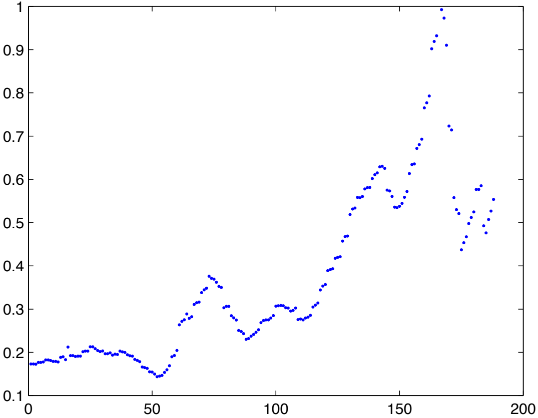
Energy consumption one-week forecast





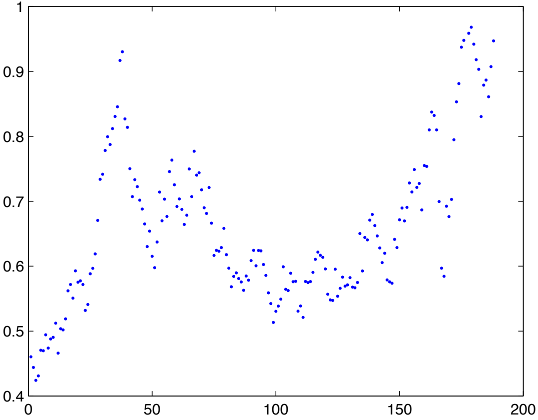
The World Bank. Global Economic Monitor 2010.

http://data.worldbank.org/data-catalog/global-economic-monitor.

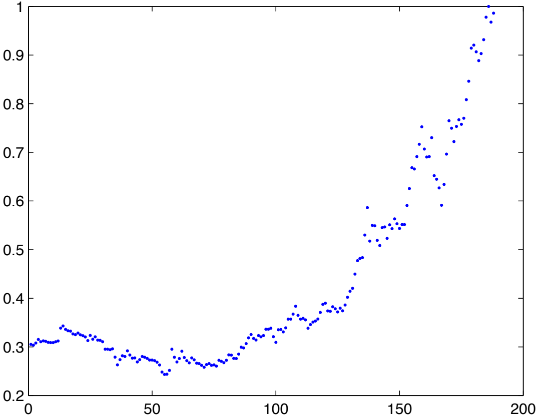


The World Bank. Global Economic Monitor 2010.

http://data.worldbank.org/data-catalog/global-economic-monitor.



The World Bank. Global Economic Monitor 2010.
http://data.worldbank.org/data-catalog/global-economic-monitor.



The World Bank. Global Economic Monitor 2010.

http://data.worldbank.org/data-catalog/global-economic-monitor.

There are N time series of length T (denote the element $a_{n,t} \in \mathcal{M}$) as the matrix

$$A = \begin{matrix} a_{1,1} & a_{1,2} & \dots & a_{1,T-1} & a_{1,T} & a_{1,T+1} \\ a_{2,1} & a_{2,2} & \dots & a_{2,T-1} & a_{2,T} & a_{2,T+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{N,1} & a_{N,2} & \dots & a_{N,T-1} & a_{N,T} & a_{N,T+1} \end{matrix}$$

Denote Δ the time-lag and for the time series $[a_{1,t}], t \in \{\Delta+1, \dots, T\}$ form the matrix

$$A_t = \begin{matrix} a_{1,t-\Delta} & a_{1,t-\Delta+1} & \dots & a_{1,t-2} & a_{1,t-1} \\ a_{2,t-\Delta} & a_{2,t-\Delta+1} & \dots & a_{2,t-2} & a_{2,t-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{N,t-\Delta} & a_{N,t-\Delta+1} & \dots & a_{N,t-2} & a_{N,t-1} \end{matrix}$$

and vectorize it to obtain the sample $\mathbf{x}_t$

$$\mathbf{x}_t = [a_{1,t-\Delta}, a_{2,t-\Delta}, \dots, a_{N,t-\Delta}, a_{1,t-\Delta+1}, \dots, a_{N,t-1}]^T.$$

Set $y_t \equiv a_{1,t}$.

Introduce the data set $D = (X, \mathbf{y})$, where

$$X = \begin{bmatrix} \mathbf{x}_1^\mathsf{T} \\ \mathbf{x}_2^\mathsf{T} \\ \vdots \\ \mathbf{x}_{T-\Delta}^\mathsf{T} \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-\Delta} \end{bmatrix}.$$

Treat this classification problem as the logistic regression

$$\mathbf{f}(\mathbf{w}, X) = \frac{1}{1 + \exp(-X\mathbf{w})} \to \mathbf{y}$$

or as another classification problem.
The error function

$$S(\mathbf{w}) = \sum_{i \in \mathcal{I}} y_i \ln f(\mathbf{w}, \mathbf{x}_i) + (1 - y_i) \ln (1 - f(\mathbf{w}, \mathbf{x}_i)).$$