Time Series Analysis.
1. Intro into TS

Alexey Romanenko
alexromsput@gmail.com

Erevan, 2017

Содержание

Let me introduce: Alexey Romanenko

Education:

- Masters Degree at MIPT, 2011
- PhD thesis: Composition of Time Series Forecasting Algorithm at MIPT, 2017

Work experience:

- SAS Institute (software implementation), 2016 – now
- MIPT (teaching assistant), 2011 – now
- Svyaznoy (one of the largest cellphone retailer in Russia), 2010–2016
- Forecsys (Machine Learning software for business), 2008–2010

View full profile: `https://www.linkedin.com/in/alexromsput/`

Course Plan

Date	Topic
28/05/2017	Intro in TS forecasting
29/05/2017	Exponential Smoothing models
30/05/2017	ARIMA and ARIMAX
31/05/2017	Practice Day 1
01/06/2017	Accuracy of forecast, comparing of models
02/06/2017	ARCH, GARCH models
03/06/2017	Practice Day 2
04/06/2017	rest day
05/06/2017	Compositions, Aggregating Algorithm
06/06/2017	Overview of other approaches
07/06/2017	Overview of other approaches
08/06/2017	TSA in Retail: Hierarchial Forecasting, Demand Forming
09/06/2017	Practice Day 3

You will have to come through:

1. 3 practice days
2. 8 HW (for 1-2 hours)
3. 2 Labs (for 4-6 hours)
4. 1 Kaggle contest (for 7 days)

Results:
https://drive.google.com/open?id=11-lzxxR0aeBmEDGEW1GiIaP4X3fyCTHxTg4b0

Materials for this course:
https://drive.google.com/open?id=0B9Zs09o9XXqNV0JuQVNLWXRrMTA

You will have to come through:

1. 3 practice days
2. 8 HW (for 1-2 hours)
3. 2 Labs (for 4-6 hours)
4. 1 Kaggle contest (for 7 days)

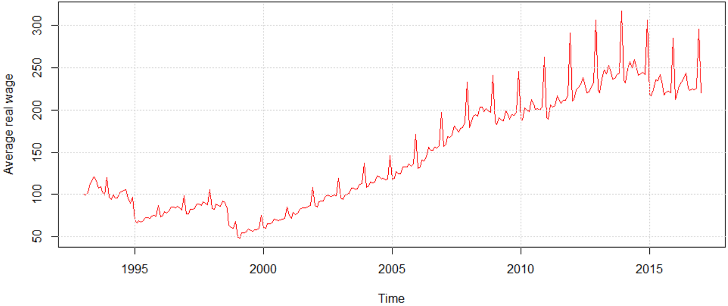
Results:
https://drive.google.com/open?id=11-lzxxR0aeBmEDGEW1GiIaP4X3fyCTHxTg4b0

Materials for this course:
https://drive.google.com/open?id=0B9Zs09o9XXqNV0JuQVNLWXRrMTA

- It is interesting!
- It is difficult!
- We will do it!

Time Series definition

Time series: $y_1, \dots, y_T, \dots, \quad y_t \in \mathbb{R}$, — a sequence of values of some variable, detected in a constant time interval.

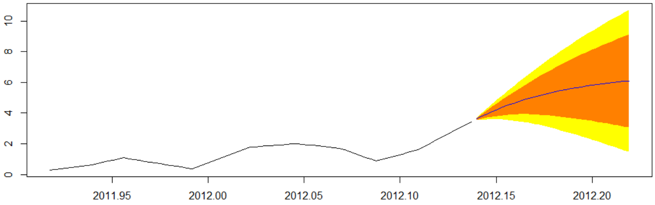


Time series forecasting task — find function f_T:

$$y_{T+d} \approx f_T\left(y_T, \dots, y_1, d\right) \equiv \hat{y}_{T+d|T},$$

where $d \in \{1, \dots, D\}$ — delay, D — horizon.

Forecasting interval, confidence of the forecast

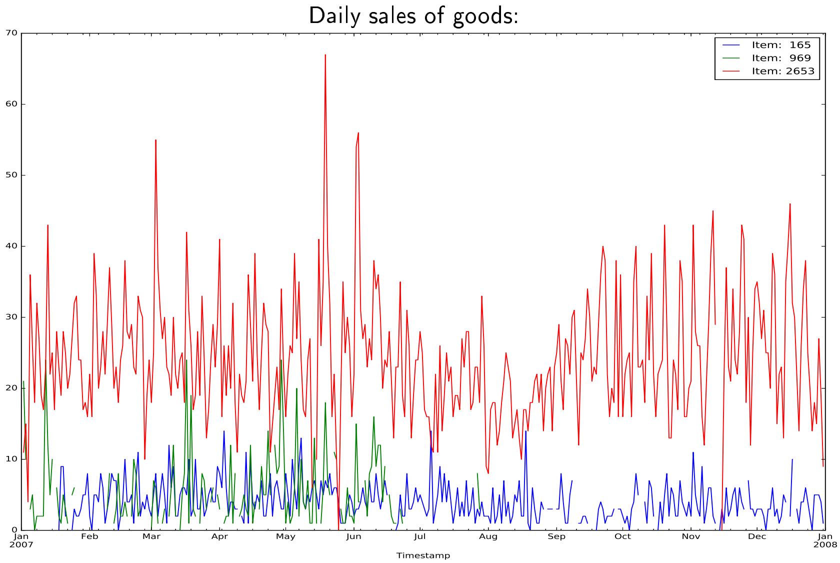


Example: April 1997, Grabd-Forks, ND, unexpected flood:
`https://www.youtube.com/watch?v=0iJUgddua-g`
The city was protected by dam of 51 feet;
The National Weather Service (NWS) had forecast that the river would crest at 49 feet
The river crested at 54.35 feet.

50000 citizens were evacuated, 75% buildings were damaged or destroyed, Property damage $\$3.5$ billion

The forecasting interval was ± 9 feet

Time series in Retail



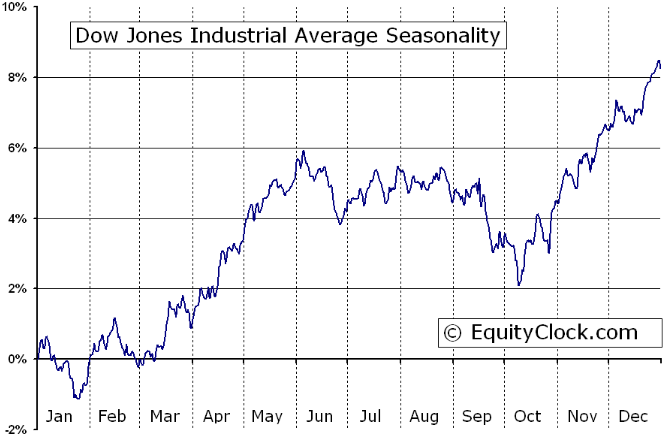
Time series in Retail

Specification of the TS in Retail

- $10^6 - 10^8$ TS to forecast at once
- missing in data
- out-of-stock (no sales by non-zero demand)
- dependence on promo-events, changes in price
- complex loss-function

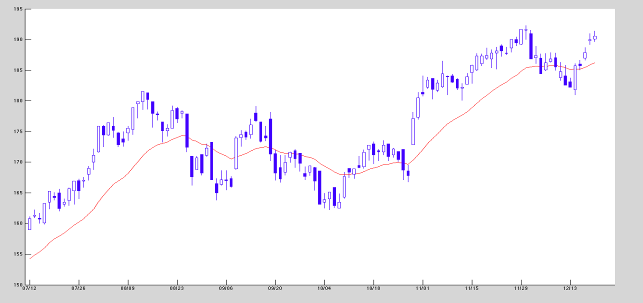
Time series in Finance

Jndex Daw-Jones:



Time series in Finance

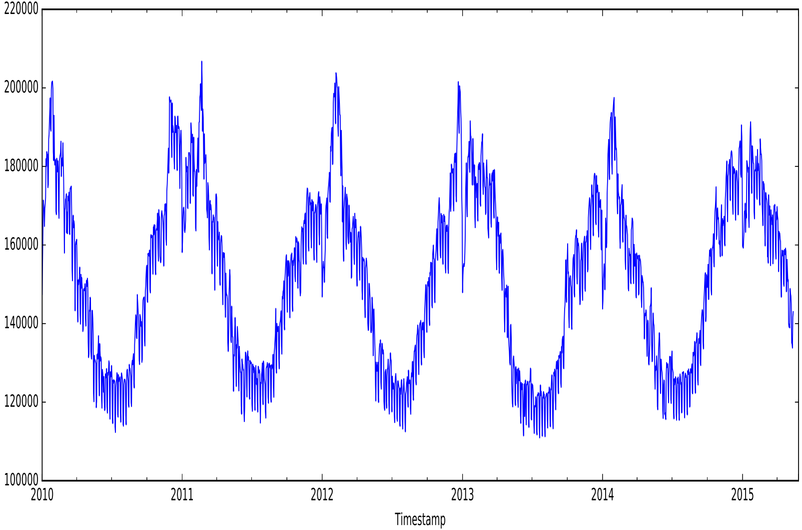
Specification of the TSA in Finance:



- high level of noise
- correlation with other financial index
- highly dependence on external events (big market events, politics)

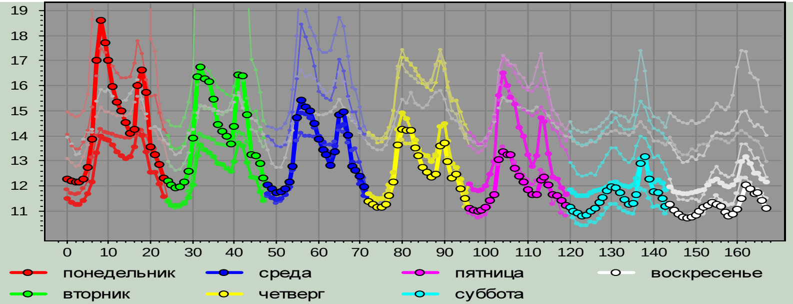
Time series in Industry: Consumption of Electricity

Volume of Consumption of Electricity:



Time series in Industry: Consumption of Electricity

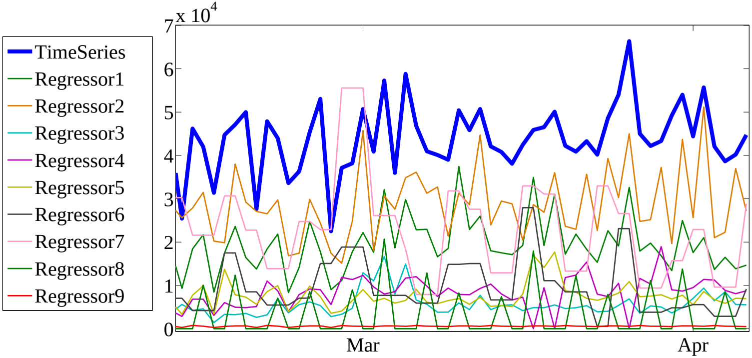
Specification of the TSA



- complex structure (yearly, weekly, daily – seasonality)
- dependence on temperature, price, calendar-events
- nonlinear dependence

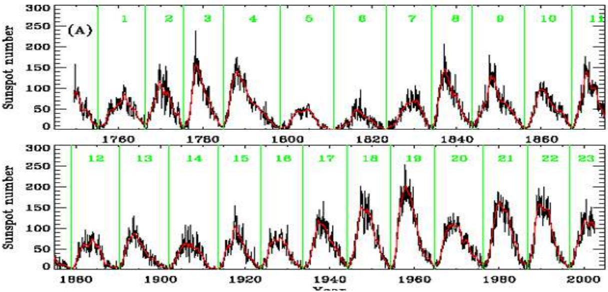
Time series in Industry: Manufacturing

Total man-hours in Warehouse:



- depends mainly on external factors
- can be described by clear physical model

Time series in Physics



- a-periodical changes
- complex physical model of dependence (Newton's laws, Kepler's laws, . . .)

Components of Time Series

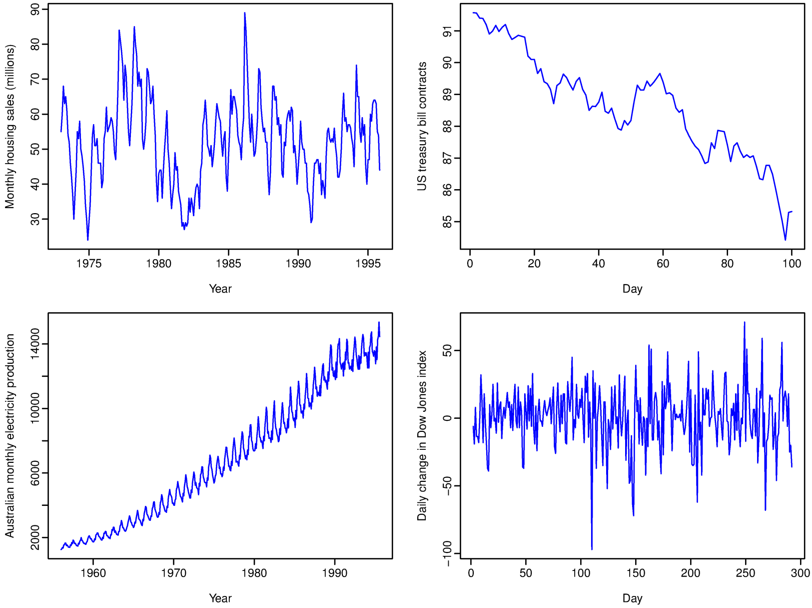
Level — average level of values;
Trend — monotonic long-term changes of Level;

Seasonality — periodical changes of values with constant period;

Cycle — changes in time series values (economical cycles, solar activity cycles).

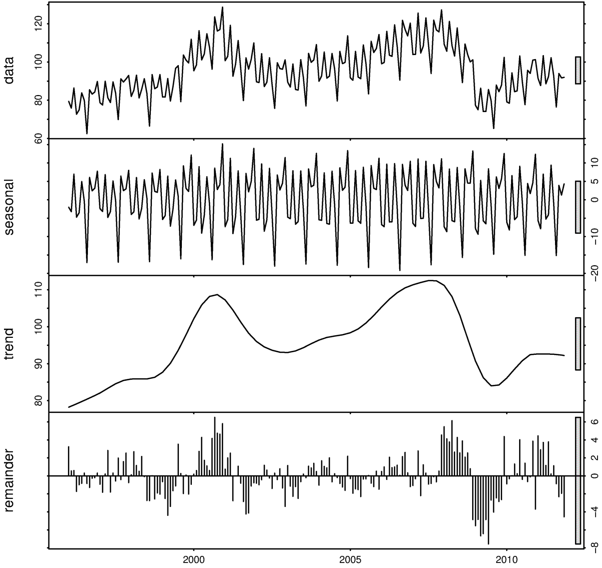
Error — random (unbiased) component of time series.

Components of Time Series



Components of Time Series

STL-decomposition:



Least Squared Method

- X — samples ($\mathbb{R}^n$); Y — answers ($\mathbb{R}$);
  $X^\ell = (x_i, y_i)_{i=1}^{\ell}$ — train samples;
  $y_i = y(x_i)$, $y\colon X \to Y$ — unknown function;
- $a(x) = f(x, \alpha)$ — regression model,
  $\alpha \in \mathbf{R}^p$ — weights of regressors.
- Least Squared Method:

$$Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} w_i \big(f(x_i, \alpha) - y_i\big)^2 \to \min_{\alpha},$$

where w_i — weight of i-th sample.

$Q(\alpha^*, X^\ell)$ — *loss function* (residual sum of squares, RSS).

Maximal Likelihood Approach

Probabilistic Model (with guassian errors):

$$y(x_i) = f(x_i, \alpha) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1, \dots, \ell.$$

Maximal Likelihood Method (MLE):

$$L(\varepsilon_1, \dots, \varepsilon_\ell | \alpha) = \prod_{i=1}^{\ell} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_i^2}\varepsilon_i^2\right) \to \max_{\alpha};$$

$$-\ln L(\varepsilon_1, \dots, \varepsilon_\ell | \alpha) = \mathsf{const}(\alpha) + \frac{1}{2}\sum_{i=1}^{\ell} \frac{1}{\sigma_i^2}\big(f(x_i, \alpha) - y_i\big)^2 \to \min_{\alpha};$$

Theorem

LS is equal to MLE for linear regression model and $w_i = \sigma_i^{-2}$.

Linear Regression Matrix

$f_1(x), \dots, f_n(x)$ — regressors (features);
Model of Linear Regression:

$$f(x, \alpha) = \sum_{j=1}^{n} \alpha_j f_j(x), \qquad \alpha \in \mathbb{R}^n.$$

LR in matrix symbols:

$$\underset{\ell \times n}{F} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_n(x_\ell) \end{pmatrix}, \quad \underset{\ell \times 1}{y} = \begin{pmatrix} y_1 \\ \dots \\ y_\ell \end{pmatrix}, \quad \underset{n \times 1}{\alpha} = \begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_n \end{pmatrix}.$$

Loss Fucntion in matrix:

$$Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} \big(f(x_i, \alpha) - y_i\big)^2 = \|F\alpha - y\|^2 \to \min_{\alpha}.$$

Solution of linear LS

Necessary condition of minimum:

$$\frac{\partial Q}{\partial \alpha}(\alpha) = 2F^{\intercal}(F\alpha - y) = 0,$$

linear system for LS:

$$F^{\intercal}F\alpha = F^{\intercal}y,$$

where $\underset{n\times n}{F^{\intercal}F}$ — covariance matrix $f_1, \ldots, f_n$.

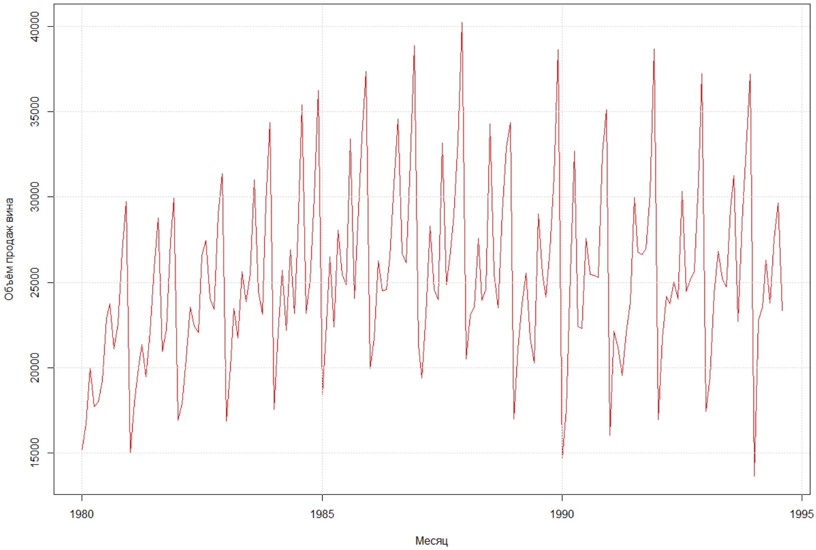
Solution:

$$\alpha^* = (F^{\intercal}F)^{-1}F^{\intercal}y = F^{+}y.$$

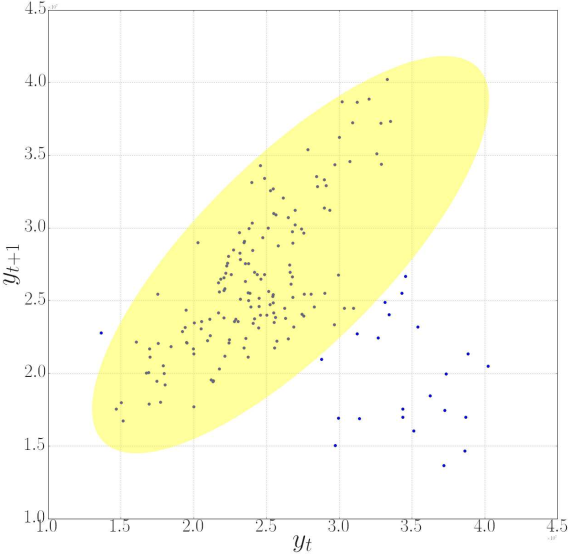
Loss functional: $Q(\alpha^*) = \|P_F y - y\|^2$,

where $P_F = FF^{+} = F(F^{\intercal}F)^{-1}F^{\intercal}$ — *projection matrix*.

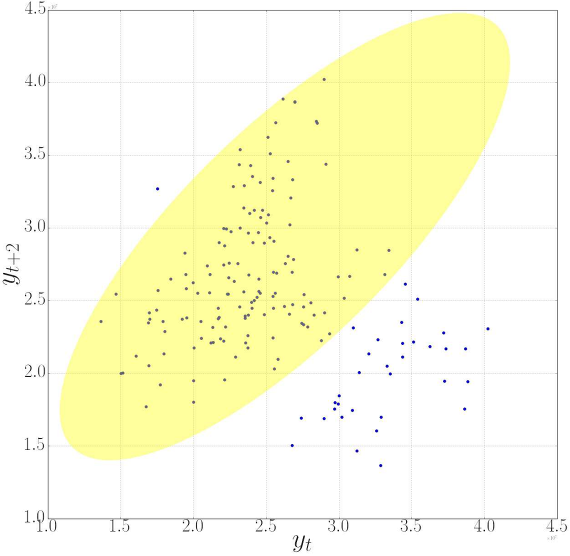
Sales of wine in Australia



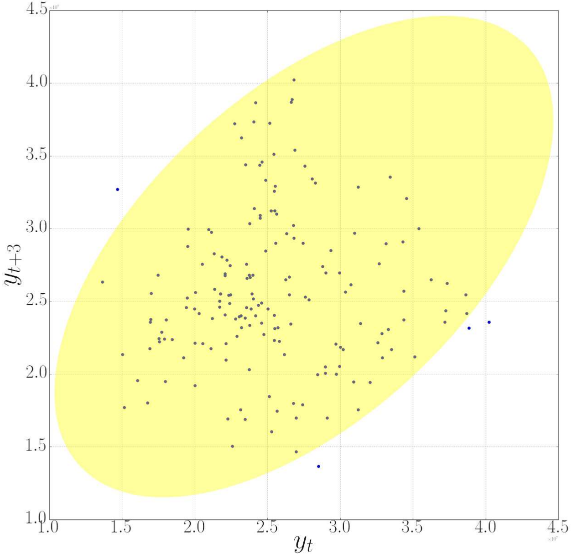
Dependence between sales for adjacent months



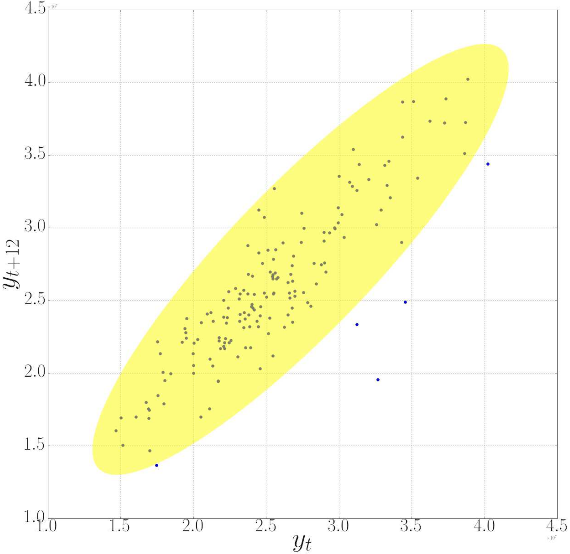
Dependence between sales for months with lag=2



Dependence between sales for months with lag=2



Dependence between sales for months with lag=12



Linear Autoregression for TimeSeries Forecasting

Regressors (features) — n previous points of time series:

$$\hat{y}_{t+1}(\alpha) = \sum_{j=1}^{n} \alpha_j y_{t-j+1}, \quad \alpha \in \mathbb{R}^n$$

Samples are $\ell = t - n + 1$ moments of time series:

$$\underset{\ell\times n}{F} = \begin{pmatrix} y_t & y_{t-1} & y_{t-2} & \dots & y_{t-n+1} \\ y_{t-1} & y_{t-2} & y_{t-3} & \dots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \dots & y_{t-n-1} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & y_{n-1} & y_{n-2} & \dots & y_1 \end{pmatrix}, \quad \underset{\ell\times 1}{y} = \begin{pmatrix} y_{t+1} \\ y_t \\ y_{t-1} \\ \dots \\ y_{n+1} \end{pmatrix}$$

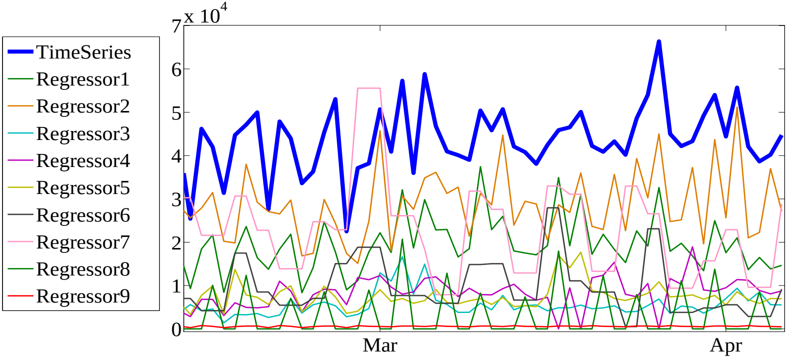
Loss Functional:

$$Q_t(\alpha, X^\ell) = \sum_{i=n+1}^{t+1} \left(\hat{y}_i(\alpha) - y_i\right)^2 = \|Fw - y\|^2 \to \min_{\alpha}$$

See example of LR for TS forecasting in `1_intro.ipnb`
Questions: which problems you can meet when you use LR?

Regression model of TS

$y_0, y_1, \dots, y_t, \dots$ — time series,
$\vec{x}_0, \vec{x}_1, \dots, \vec{x}_t, \dots$ — regressors, $\vec{x}_t = (x_{1,t}, \dots, x_{n,t}) \in \mathbb{R}^n$
$\hat{y}_{t+d}(\alpha) = f_t\left(\vec{x}_{t+d}; \vec{\alpha_t}\right)$ — regression model of TS,

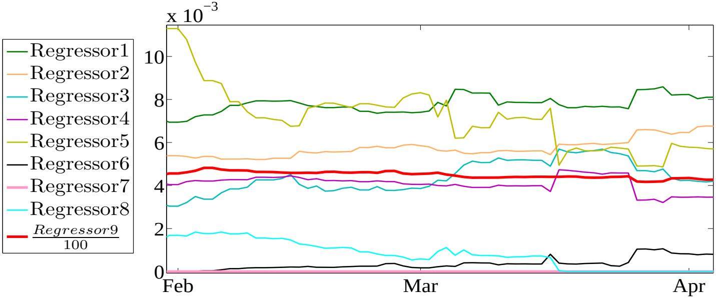


Question: How can we make LR model adaptive?

LAWR: Local Adaptation of weights with regularization

At each step t weights are found as LS with reweighting samples and smoothing with weights at the moment $t-1$:

$$\begin{cases} \left[\sum_{t=0}^{T} \beta^{(T-t)} \left(\sum_{j=1}^{k} \alpha_j x_{j,t} - y_t\right)^2\right] + \lambda \sum_{j=1}^{k} \left(\alpha_j - \alpha_{j,T-1}\right)^2 \to \min_{\alpha_j, j=1,\dots,k} \\ \sum_{j=1}^{k} \alpha_j \geq 0. \end{cases}$$



Conclusion

- time series differs in business regions
- time series forecasting problem has specific

- LR is the common approach to build the forecast
- LR can be benchmark solution of TS forecasting

- LR is not good enough:

Conclusion

- time series differs in business regions
- time series forecasting problem has specific

- LR is the common approach to build the forecast
- LR can be benchmark solution of TS forecasting

- LR is not good enough:
  1. last sample and first sample in train set have the same weight
  2. LR needs a lot of machine resources

Conclusion

- time series differs in business regions
- time series forecasting problem has specific properties for each business

- LR is the common approach to build the forecast
- LR can be used as benchmark of TS forecast

- LR is not good enough:
  1. last sample and first sample in train set have the same weight
  2. LR needs a lot of machine resources

- there are adaptive variations of LR to add adaptation and stabilization