

Time Series Analysis.

1. Intro into TS

Alexey Romanenko
alexromsput@gmail.com

FIVT MIPT, September 2017

Содержание

- 1 About this course
 - Intro
 - Course Plan
 - Time Series Forecasting Module

- 2 Time Series forecasting
 - Time Series
 - Example of Time Series
 - General components of Time Series

- 3 Exponential Smoothing models
 - Simple Exponential Smoothing
 - Trend and Seasonality ES models
 - Argumentation for ES models

Let me introduce: Alexey Romanenko

Education:

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

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

Advanced Machine Learning:

- Time Series Forecasting
- Reinforcement Learning
- Information Retrieval
- Recommendations Systems
- Guest Lectures (1-2)

Страница курса <https://ml-mipt.github.io/>

Date	Topic
06/09/2017	Intro in TS forecasting, Exponential Smoothing models
13/09/2017	ARIMAX and other autoregression models (ARCH, GARCH)
20/09/2017	Comparing of models, compositions, Aggregating Algorithm
27/09/2017	TSA in Retail: Hierarchial Forecasting

Date	Topic
06/09/2017	Intro in TS forecasting, Exponential Smoothing models
13/09/2017	ARIMAX and other autoregression models (ARCH, GARCH)
20/09/2017	Comparing of models, compositions, Aggregating Algorithm
27/09/2017	TSA in Retail: Hierarchial Forecasting

You will have to come through:

- 1 5 seminars (first you've already heard)
- 2 2 HW (for 1-2 hours)
- 3 1 Kaggle contest (for 14 days)

Date	Topic
06/09/2017	Intro in TS forecasting, Exponential Smoothing models
13/09/2017	ARIMAX and other autoregression models (ARCH, GARCH)
20/09/2017	Comparing of models, compositions, Aggregating Algorithm
27/09/2017	TSA in Retail: Hierarchial Forecasting

You will have to come through:

- 1 5 seminars (first you've already heard)
- 2 2 HW (for 1-2 hours)
- 3 1 Kaggle contest (for 14 days)

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

Time Series definition

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

Time series forecasting task — find function f_T :

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

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

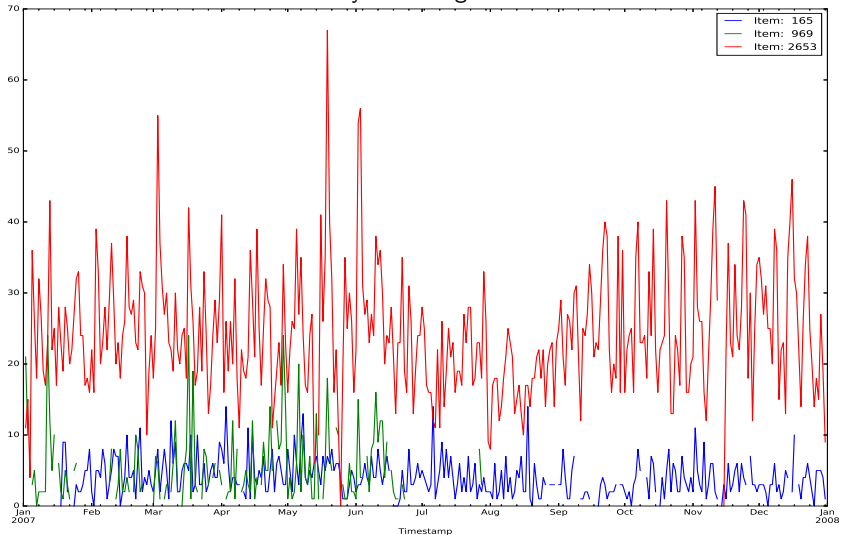
The river crested at 54.35 feet.

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

The forecasting interval was ± 9 feet

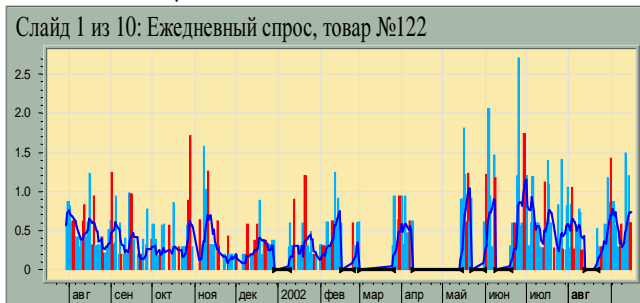
Time series in Retail

Daily sales of goods:



Time series in Retail

Specification of the TS in Retail



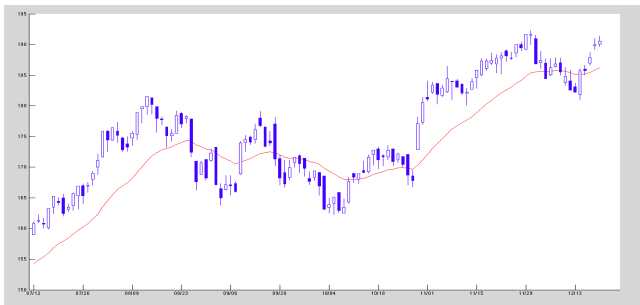
- $10^6 - 10^8$ TS to forecast at once
- missing in data
- out-of-stock (no sales by non-zero demand)
- dependence on promo-events, changes in price
- complex loss-function

Time series in Finance

Index Daw-Jones:

Time series in Finance

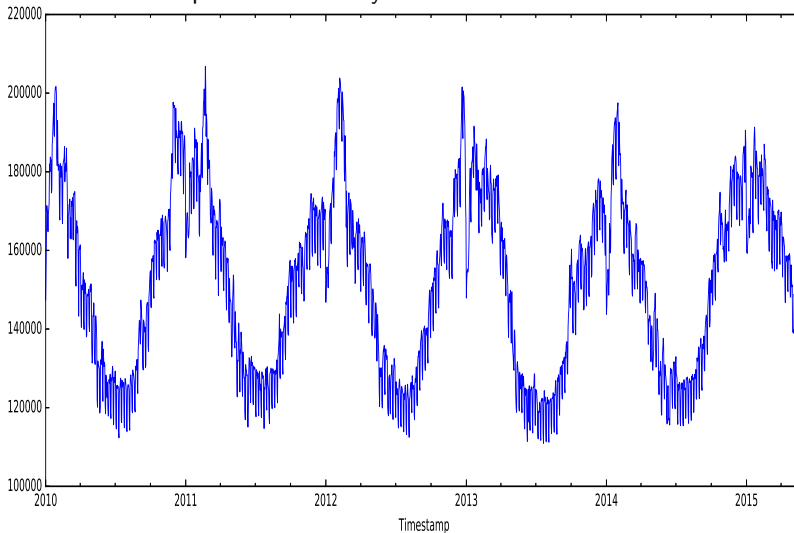
Specification of the TSA in Finance:



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

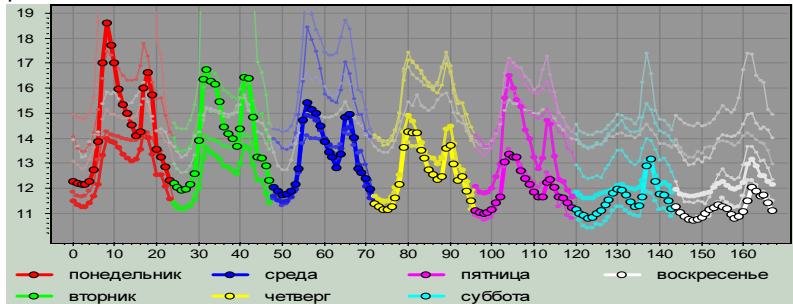
Time series in Industry: Consumption of Electricity

Volume of Consumption of Electricity:



Time series in Industry: Consumption of Electricity

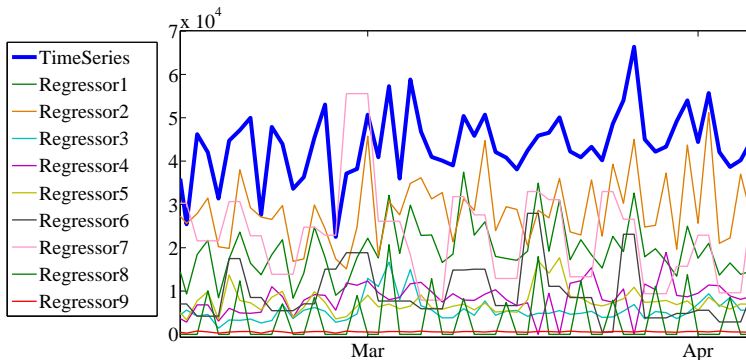
Specification of the TSA



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

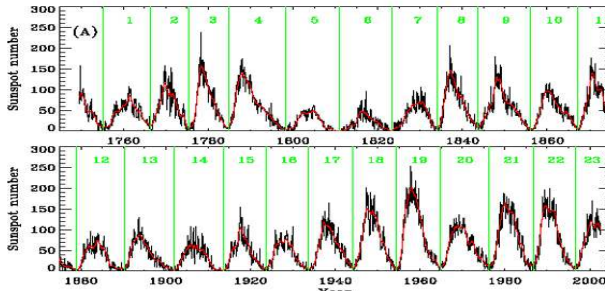
Time series in Industry: Manufacturing

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.

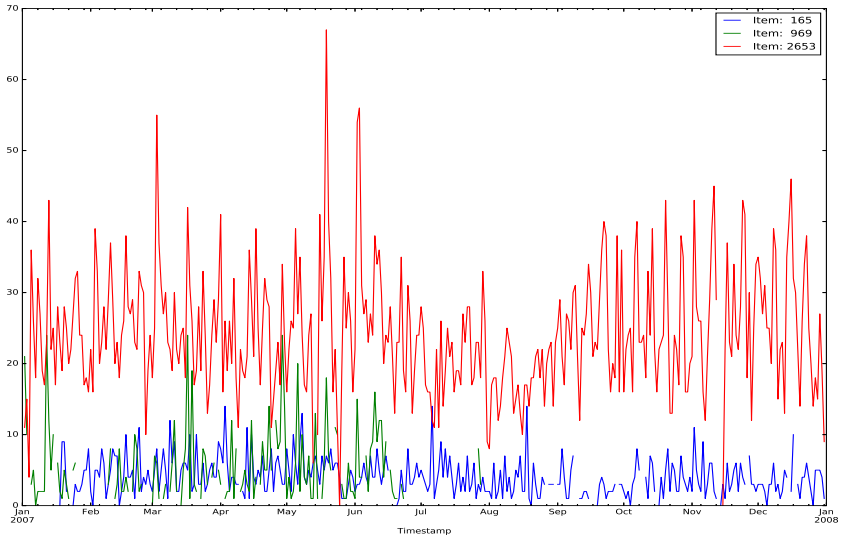
Components of Time Series

Components of Time Series

STL-decomposition:

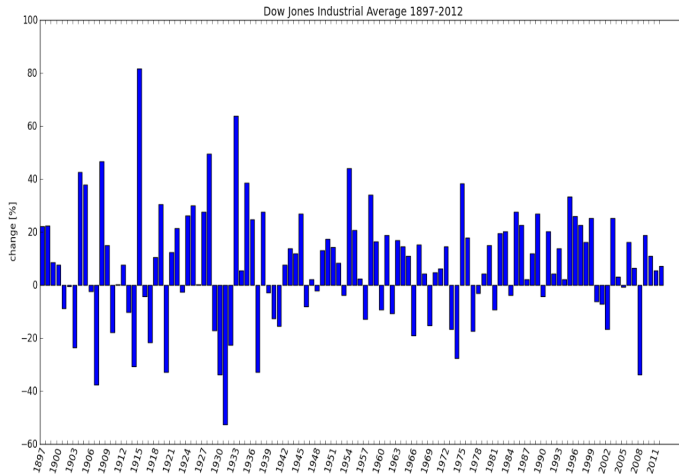
Time Series Model

Real time series of in Retail Chain:



Time Series Model

Index Dow-Jones:



Time Series Model

$y_0, y_1, \dots, y_t, \dots$ — is a time series, $y_i \in \mathbb{R}$

The model of time series:

$$y_t = l_t + \varepsilon_t$$

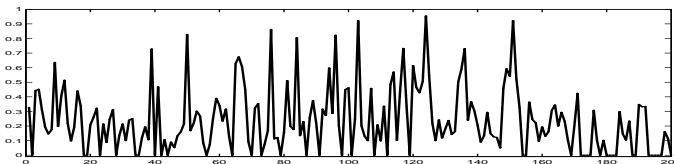
where l_t — level of time series (changing slowly),

ε_t — (unobserved) error component (noise),

Forecasting model:

$$\hat{y}_{t+d} = \hat{l}_t$$

where \hat{l}_t — an estimation of level,



Simple Exponential Smoothing

Weighted average with exponentially diminishing weights forecast:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots$$

$\alpha \uparrow 1 \Rightarrow$ greater weight to last points,

$\alpha \downarrow 0 \Rightarrow$ greater smoothing.

Time point	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	0.08192	0.05184	0.01536	0.00128
y_{T-5}	0.065536	0.031104	0.006144	0.000256

We find the optimal α^* using moving control:

$$Q(\alpha) = \sum_{t=T_0}^{T_1} (\hat{y}_t(\alpha) - y_t)^2 \rightarrow \min_{\alpha}$$

Empirical rules:

if $\alpha^* \in (0, 0.3)$ the series is stationary, ES works;

if $\alpha^* \in (0.3, 1)$ the series is non-stationary, we need a trend model.

Simple Exponential Smoothing

- The method suits forecasting of time series without trend and seasonality:

$$\hat{y}_{t+1|t} = l_t,$$
$$l_t = \alpha y_t + (1 - \alpha) l_{t-1} = \hat{y}_{t|t-1} + \alpha \cdot e_t.$$

where $e_t = y_t - \hat{y}_{t|t-1}$ — forecast error at time point t

Proof:

$$\hat{y}_{t+1} := \alpha y_t + (1 - \alpha) \hat{y}_t = \hat{y}_t + \alpha \cdot e_t$$

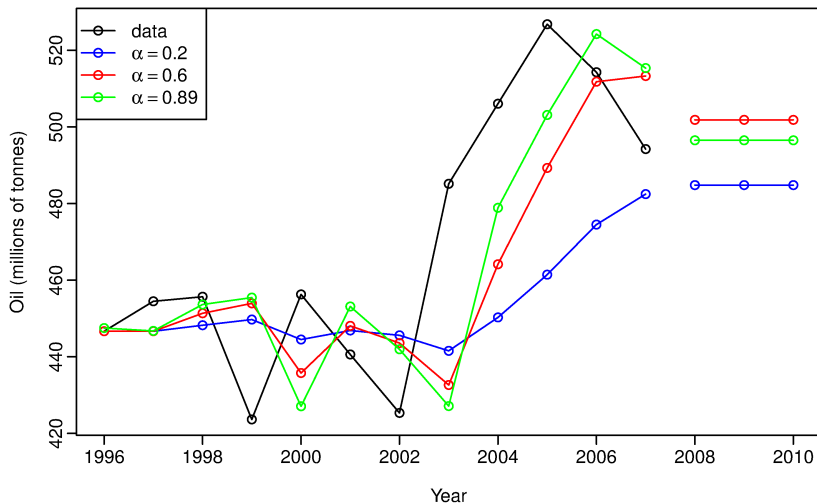
- The forecast depends on l_0 :

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha (1 - \alpha)^j y_{T-j} + (1 - \alpha)^T l_0.$$

We can take $l_0 = y_1$ or optimize it.

- Forecast turns out flat, i.e. $\hat{y}_{t+d|t} = \hat{y}_{t+1|t}$.

Simple Exponential Smoothing



Simple ES applied to data on oil production in Saudi Arabia (1996–2007).

Tracking Signal

Tracking signal [Trigg, 1964]

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t} \quad \begin{aligned} \hat{e}_{t+1} &:= \gamma e_t + (1 - \gamma)\hat{e}_t; \\ \tilde{e}_{t+1} &:= \gamma|e_t| + (1 - \gamma)\tilde{e}_t. \end{aligned}$$

Recommendation: $\gamma = 0.05 \dots 0.1$

Statistics adequacy test (at $\gamma \geq 0.1$, $t \rightarrow \infty$):

hypotheses $H_0: E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_{t+d} = 0$
is accepted at significance level δ if

$$|K_t| \leq 1.2\Phi_{1-\delta/2}\sqrt{1 - \gamma/(1 + \gamma)},$$

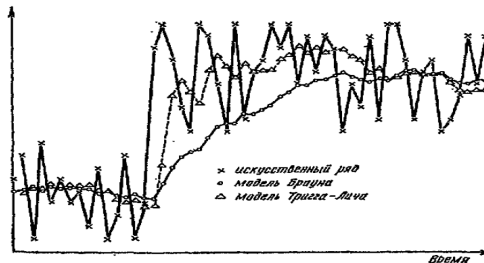
$\Phi_{1-\delta/2}$ — normal distribution quantile,

$\Phi_{1-\delta/2} = \Phi_{0.975} = 1.96$ at $\delta = 0.05$

Trigg-Leach Model [Trigg, Leach, 1967]

Problem: adaptive models adjust poorly to sharp changes of structure

Solution: $\alpha = |K_t|$

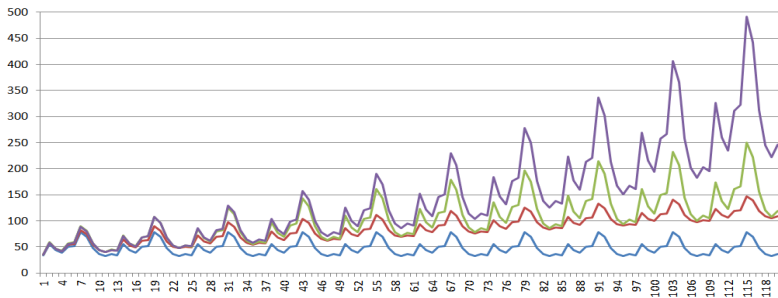


Drawbacks:

- 1) reacts poorly to single outliers; ($\alpha_t = |K_{t-1}|$)
- 2) requires fitting γ given recommended $\gamma = 0.05 \dots 0.1$.

Examples of Trend and Seasonality

Example: Combination of trend and seasonality (model data)



Ряд 1 — seasonality and no trend

Ряд 2 — linear trend, additive seasonality

Ряд 3 — linear trend, multiplicative seasonality

Ряд 4 — exponential trend, multiplicative seasonality

Holt Model = Linear Trend

Linear trend with no seasonality effect:

$$\hat{y}_{t+d} = l_t + b_t d,$$

where l_t , b_t — adaptive coefficients of linear trend

Recursive formula:

$$\begin{aligned} l_t &:= \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) = \hat{y}_t + \alpha e_t; \\ b_t &:= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} = b_{t-1} + \alpha\beta e_t. \end{aligned}$$

Particular case — Brown linear growth model:

$$\alpha = \alpha, \quad \beta = \alpha$$

Other Methods that Account Trend

Multiplicative linear (exponential) trend:

$$\begin{aligned}\hat{y}_{t+d|t} &= l_t b_t^d, \\ l_t &= \alpha y_t + (1 - \alpha) (l_{t-1} b_{t-1}), \\ b_t &= \beta \frac{l_t}{l_{t-1}} + (1 - \beta) b_{t-1}.\end{aligned}$$

$$\alpha, \beta \in [0, 1].$$

Other Methods that Account Trend

Additive damped trend:

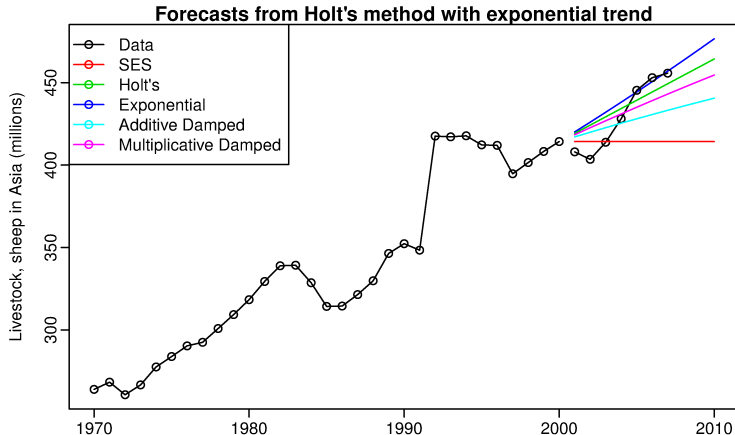
$$\begin{aligned}\hat{y}_{t+d|t} &= l_t + \left(\phi + \phi^2 + \cdots + \phi^d \right) b_t, \\ l_t &= \alpha y_t + (1 - \alpha) (l_{t-1} + \phi b_{t-1}), \\ b_t &= \beta (l_t - l_{t-1}) + (1 - \beta) \phi b_{t-1}.\end{aligned}$$

Multiplicative damped trend:

$$\begin{aligned}\hat{y}_{t+d|t} &= l_t b_t^{(\phi + \phi^2 + \cdots + \phi^d)}, \\ l_t &= \alpha y_t + (1 - \alpha) l_{t-1} b_{t-1}^\phi, \\ b_t &= \beta \frac{l_t}{l_{t-1}} + (1 - \beta) b_{t-1}^\phi.\end{aligned}$$

$$\alpha, \beta \in [0, 1], \quad \phi \in (0, 1).$$

Other Methods that Account Trend



Forecast of sheep population in Asia with regard for trend.

	SES	Holt's	Exponential	Additive damped	Multiplicative damped
α	1	0.98	0.98	0.99	0.98
β		0	0	0	0.00
ϕ				0.98	0.98

Winters Model = Multiplicative Seasonality

Multiplicative Seasonality of Period p :

$$\hat{y}_{t+d} = l_t \cdot s_{t-p+(d \bmod p)},$$

s_0, \dots, s_{p-1} — seasonality profile of period p .

Recursive formula:

$$\begin{aligned} l_t &:= \alpha(y_t/s_{t-p}) + (1 - \alpha)l_{t-1} = l_{t-1} + \alpha e_t/s_{t-p}; \\ s_t &:= \beta(y_t/l_t) + (1 - \beta)s_{t-p} = s_{t-p} + \beta(1 - \alpha)e_t/l_t. \end{aligned}$$

Proof of the last equation:

$$\begin{aligned} s_t &:= s_{t-p} + \beta(y_t/l_t - s_{t-p}) = s_{t-p} + \beta(y_t - s_{t-p}l_t)/l_t = \\ &= s_{t-p} + \beta(y_t - s_{t-p}(l_{t-1} + \alpha e_t/s_{t-p}))/l_t = s_{t-p} + \\ &+ \beta \left(\underbrace{y_t - s_{t-p}l_{t-1}}_{e_t} - \alpha e_t \right) / l_t \end{aligned}$$

Additive Seasonality ES Model

Additive seasonality with period of length p :

$$\hat{y}_{t+d|t} = l_t + s_{t-p+(d \bmod p)},$$

$$l_t = \alpha (y_t - s_{t-p}) + (1 - \alpha) (l_{t-1}) = \textcolor{red}{l_{t-1}} + \alpha e_t;$$

$$s_t = \gamma (y_t - l_{t-1}) + (1 - \gamma) s_{t-p} = \textcolor{red}{s_{t-p}} + \gamma(1 - \alpha)e_t.$$

Theil-Wage Model

Linear trend with additive seasonality of period s :

$$\hat{y}_{t+d} = (l_t + b_t d) + s_{t+(d \bmod s)-p}.$$

$l_t + b_t d$ — trend cleaned of seasonality,

s_0, \dots, s_{p-1} — seasonality profile of period p .

Recursive formula:

$$l_t := \alpha(y_t - s_{t-p}) + (1 - \alpha)(l_{t-1} + b_{t-1}) = l_{t-1} + b_{t-1} + \alpha e_t;$$

$$b_t := \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} = b_{t-1} + \alpha\beta e_t;$$

$$s_t := \gamma(y_t - l_t) + (1 - \gamma)s_{t-p} = s_{t-p} + \gamma(1 - \alpha)e_t.$$

Winters Model with Linear Trend

Multiplicative seasonality of period s with a linear trend:

$$\hat{y}_{t+d} = (l_t + b_t d) \cdot s_{t+(d \bmod p)-p},$$

$l_t + b_t d$ — trend cleaned of seasonality,

s_0, \dots, s_{p-1} — seasonality profile of period s .

Recursive formula:

$$l_t := \alpha(y_t/s_{t-p}) + (1 - \alpha)(l_{t-1} + b_{t-1}) = l_{t-1} + b_{t-1} + \alpha e_t/s_{t-p};$$

$$b_t := \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} = b_{t-1} + \alpha\beta e_t/s_{t-p};$$

$$s_t := \gamma(y_t/l_t) + (1 - \gamma)s_{t-p} = s_{t-p} + \gamma(1 - \alpha)e_t/l_t.$$

Winters Model with exponential trend

Multiplicative trend model exponential trend:

$$\hat{y}_{t+d} = l_t(r_t)^d \cdot s_{t+(d \bmod p)-p},$$

$l_t(r_t)^d$ — exponential trend without seasonality,

s_0, \dots, s_{p-1} — seasonal trend p .

Recurrent version:

$$l_t := \alpha(y_t/s_{t-p}) + (1 - \alpha)l_{t-1}r_{t-1} = l_{t-1}r_{t-1} + \alpha e_t/s_{t-1};$$

$$r_t := \beta(l_t/l_{t-1}) + (1 - \beta)r_{t-1} = r_{t-1} + \alpha\beta e_t/st - 1;$$

$$s_t := \gamma(y_t/l_t) + (1 - \gamma)s_{t-p} = s_{t-p} + \gamma(1 - \alpha)e_t/l_t.$$

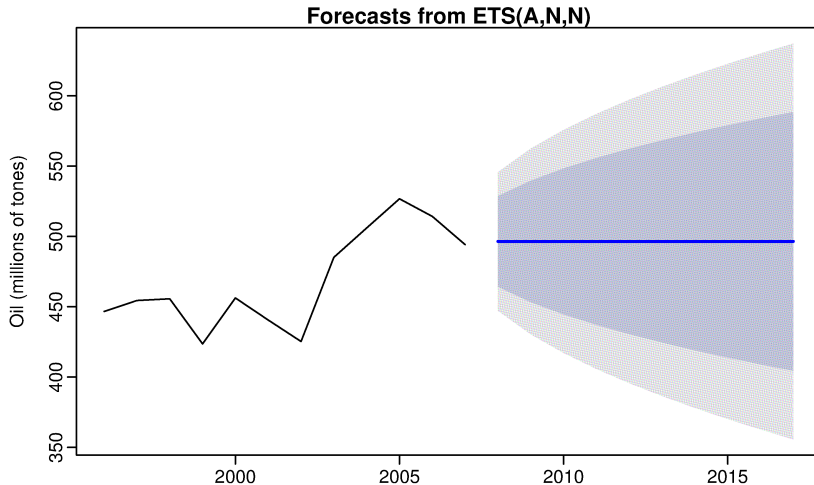
ES Models

	Seasonality		
Trend	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
Ad (Additive damped)	(Ad,N)	(Ad,A)	(Ad,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
Md (Multiplicative damped)	(Md,N)	(Md,A)	(Md,M)

We may additionally suggest an additive (A) or a multiplicative (M) error (the type of error does not influence single-value prediction). Multiplicative error is suitable only for strictly positive time series.

The final model may be written as $ESM(\cdot, \cdot, \cdot)$.

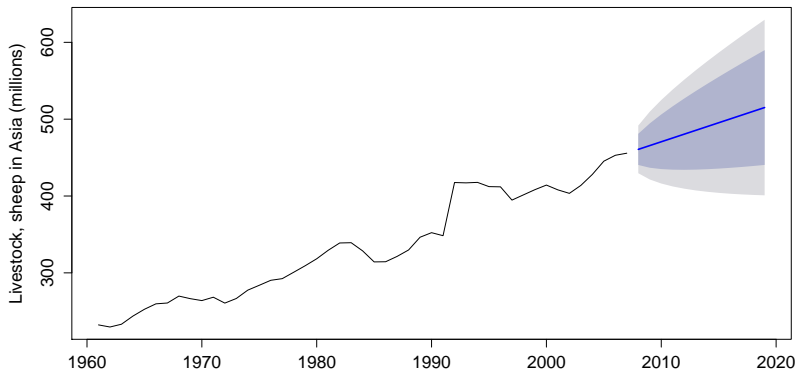
Examples of Forecast



For the data on oil production in Saudi Arabia function ESM selects simple ES.

Examples of Forecast

Forecasts from ETS(M,A,N)



For the sheep population in Asia function ESM selects the model with multiplicative error and an additive linear trend.

Examples of Forecast

For the quantity of nights spent by tourists in Australia function ESM selects a model with multiplicative error, seasonality and a damped trend.

Data model

- X — samples (\mathbb{R}^n); Y — answers (\mathbb{R});
 $X^\ell = (x_i, y_i)_{i=1}^\ell$ — train samples;
 $f(x, y)$ — joint distribution data come from;
- Data model:

$$y_i = m(x_i) + \varepsilon_i,$$

$m(x)$ — unknown function;

ε_i — error (random variable)

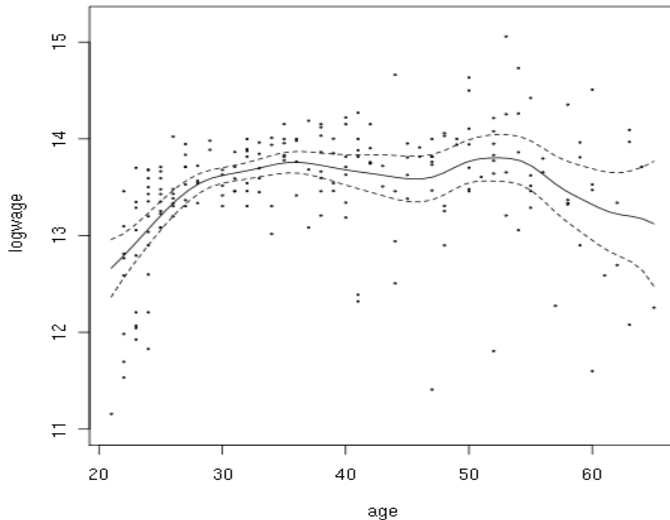
$$\mathbb{E}(\varepsilon_i) = 0, \quad \mathbb{D}(\varepsilon_i) = \sigma_i^2, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \forall i \neq j$$

- estimator (according to bayes optimal decision rule)

$$a(x) = \hat{m}(x) = \mathbb{E}[Y|X = x] = \int y f(y, x) dy = \frac{\int y f(x, y) dy}{\int f(x, y) dy}$$

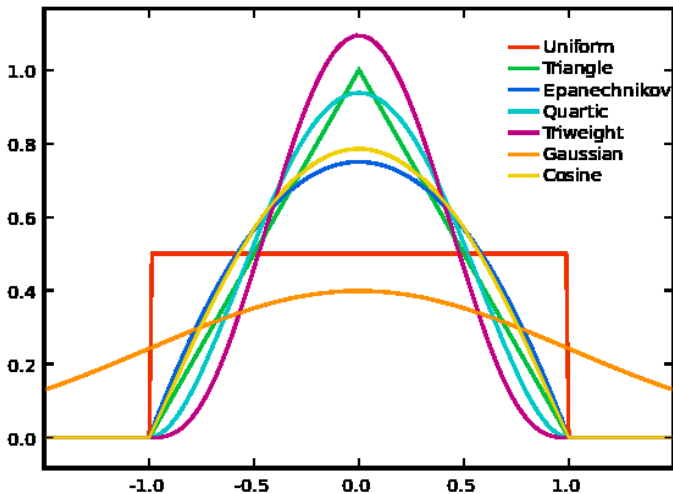
We just need to estimate numerator and denominator.

Data Examples



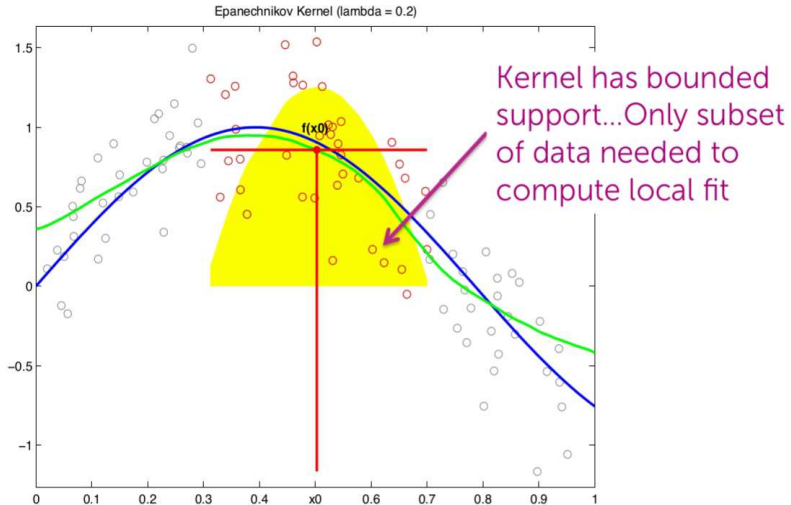
Kernel estimator

Kernel is non-negative real-valued integrable even (symmetric) function



Kernel estimator

Kernel Estimator:



Kernel estimator

Density estimator:

$$\hat{f}(x, y) = \frac{1}{\ell h_x h_y} \sum_{i=1}^{\ell} K\left(\frac{\rho(x, x_i)}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right) = \frac{1}{\ell} \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) K_{h_y}(y - y_i);$$

Estimation of numerator:

$$\begin{aligned} \int y \hat{f}(x, y) dy &= \frac{1}{\ell} \int y \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) K_{h_y}(y - y_i) dy = \\ &= \frac{1}{\ell} \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) \underbrace{\int y K_{h_y}(y - y_i) dy}_{y_i} = \frac{1}{\ell} \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) y_i; \end{aligned}$$

Estimation of denominator:

$$\begin{aligned} \int \hat{f}(x, y) dy &= \frac{1}{\ell} \int \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) K_{h_y}(y - y_i) dy = \\ &= \frac{1}{\ell} \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) \underbrace{\int K_{h_y}(y - y_i) dy}_1 = \frac{1}{\ell} \sum_{i=1}^{\ell} K_{h_x}(\rho(x, x_i)) = \hat{f}(x); \end{aligned}$$

Formula of Nadaraya–Watson

As consequence:

Formula (*kernel smoothing*) Nadaraya–Watson:

$$a_h(x; X^\ell) = \frac{\sum_{i=1}^{\ell} y_i w_i(x)}{\sum_{i=1}^{\ell} w_i(x)} = \frac{\sum_{i=1}^{\ell} y_i K\left(\frac{\rho(x, x_i)}{h_x}\right)}{\sum_{i=1}^{\ell} K\left(\frac{\rho(x, x_i)}{h_x}\right)}.$$

where $w_i(x) = K\left(\frac{\rho(x, x_i)}{h}\right)$ — weights of x_i with respect to x ;
 $K(r)$ — Kernel, bounded, smooth; h — window of smoothing.

We could get the same solution as follows: find as constant $a(x) = \alpha$ in neighborhood of $x \in X$:

$$Q(\alpha; X^\ell) = \sum_{i=1}^{\ell} w_i(x) (\alpha - y_i)^2 \rightarrow \min_{\alpha \in \mathbb{R}};$$

Argumentation of formula Nadaraya-Watson

Theorem

If next conditions are met:

1) *sample $X^\ell = (x_i, y_i)_{i=1}^\ell$ is generated from a joint distribution $f(x, y)$;*

2) *Kernel $K(r)$ is bounded: $\int_0^\infty K(r) dr < \infty$, $\lim_{r \rightarrow \infty} rK(r) = 0$;*

3) *$E(y|x)$ has limited second moment:*

$$E(y^2|x) = \int_Y y^2 p(y|x) dy < \infty \text{ for all } x \in X;$$

4) *sequence h_ℓ decreases but not fast:*

$$\lim_{\ell \rightarrow \infty} h_\ell = 0, \quad \lim_{\ell \rightarrow \infty} \ell h_\ell = \infty.$$

Then there is convergence in probability:

$$a_{h_\ell}(x; X^\ell) \xrightarrow{P} E(y|x) \text{ in all point } x \in X,$$

and $E(y|x)$, $f(x)$ и $D(y|x)$ are continuous functions and $f(x) > 0$.

Simple Exponential Smoothing

Simple regression model for TS forecasting — constant $\hat{y}_{t+1} = c$,

Kernel function $K(\rho(y_t, y_\tau)) = \alpha^{t-\tau}$

In accordance with formula Nadaraya–Watson:

$$c \equiv \hat{y}_{t+1} = \frac{\sum_{i=0}^t \alpha^i y_{t-i}}{\sum_{i=0}^t \alpha^i}$$

Note, we can obtain the same result by solving optimization problem:

$$F(c) = \sum_{t=0}^T \alpha^{T-t} (y_t - c)^2 \rightarrow \min_c, \quad \alpha \in (0, 1).$$

Proof: just find solution of the equation $\frac{\partial F(c)}{\partial c} = 0$.

Simple Exponential Smoothing

Write the same for \hat{y}_t and use approximation $\sum_{i=0}^t \alpha^i \approx \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}$,
get

$$\hat{y}_{t+1} = \hat{y}_t \alpha + (1 - \alpha)y_t$$

Replace $\alpha = 1 - \alpha \Rightarrow$ Simple ES:

$$\hat{y}_{t+1} := \hat{y}_t + \alpha(y_t - \hat{y}_t) = \alpha y_t + (1 - \alpha)\hat{y}_t,$$

Corollary

Simple ES gives the optimal solution if loss function looks like

$$Loss(T) = \sum_{t=1}^t \alpha^{T-t} (y_t - \hat{y}_t)^2,$$

where α is the same as in SES.

Other theoretical Properties of Simple ES

The model of time series:

$$y_t = l_t + \varepsilon_t$$

where l_t — level of time series (changing slowly),

ε_t — (unobserved) error component (noise),

$$E(\varepsilon_i) = 0, \quad D(\varepsilon_i) = \sigma^2 < \infty \quad \forall i, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$$

Error of forecast:

$$e_t = y_t - \hat{y}_t$$

Math expectation of squared loss of SES:

$$E((e_t)^2) \approx D(\varepsilon)(1 + (t-1)\alpha^2)$$

Conclusion

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

Conclusion

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

Pro ES:

- ES is very simple model
- very useful for short TS or simple TS
- can be easily modified for different components of TS
- Most of ES models imply quiet a clear interpretation

Conclusion

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

Pro ES:

- ES is very simple model
- very useful for short TS or simple TS
- can be easily modified for different components of TS
- Most of ES models imply quiet a clear interpretation

Cons ES:

- ES does not take into account independent variables
- heuristic method (there is no theoretical guaranties about it's work)
- Forecast of ES depends on initialization (l_0)

Literature

Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. — OTexts,
<https://www.otexts.org/book/fpp>

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов.
Финансы и статистика, 2003, <http://www.arshinov74.ru/files/files/3.pdf>