Module 7. Time series analysis

Data Science & AI

Sabine De Vreese Lieven Smits Bert Van Vreckem

2023–2024

HO
GENT

Contents

HO
GENT

Learning goals

- Concepts, time series models
- Moving average
- Exponential smoothing

Remark

You should not memorize the formulas in this module, but you should understand them!

HO GENT

Time series and predictions

Time series and predictions

Time series

A time series is a sequence of observations of some variable over time.

- Monthly demand for milk
- Annual intake of students at HOGENT
- Price of a share or bond on the stock exchange (hourly, daily, ...)
- Number of HTTP requests per second for a website
- Evolution of disk usage on a backup server

HO
GENT

Time series and predictions

May decisions in business operations depend on a forecast of some quantity

- General development of future plans (investments, capacity ...)
- Budget planning to avoid shortcomings (operating budget, marketing budget ...)
- Procurement planning (e.g. storage capacity)
- Support for financial objectives
- Avoid uncertainty

HO
GENT

Time series and predictions

Time series are a **statistical** problem: observations vary with time

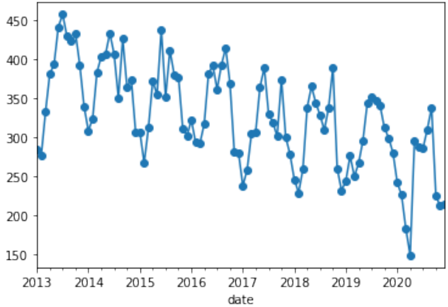
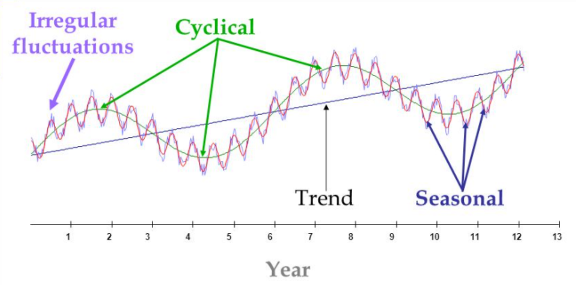


Figure: Number of heavily wounded in car accidents in Flanders.

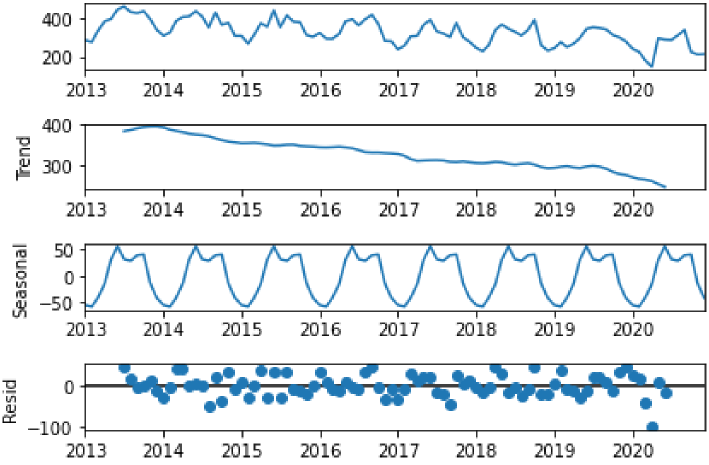
Time series components

- Level
- Trend
- Seasonal fluctuations
- Cyclic patterns
- Random noise (residuals)



HO GENT

Time series decomposition



HO
GENT

Time series models

HO
GENT

Mathematical model time series

The simplest model:

$$X_t = b + \varepsilon_t \tag{1}$$

- X_t: estimate for time series, at time t
- b: the level (a constant), based on **observations** x_t
- ε_t: random **noise**. We assume that $\varepsilon_t \sim Nor(\mu = 0; \sigma)$

HO
GENT

Mathematical model time series

We could also assume that there is a linear relationship:

$$X_t = b_0 + b_1 t + \varepsilon_t \tag{2}$$

with **level** b_0 and **trend** b_1.

Equation 1 and 2 are special cases of the **polynomial** case:

$$X_t = b_0 + b_1 t + b_2 t^2 + \cdots + b_n t^n + \varepsilon_t \tag{3}$$

HO
GENT

General expression time series

$$X_t = f(b_0, b_1, b_2, \dots, b_n, t) + \varepsilon_t \tag{4}$$

We make these assumptions:

- We consider two components of variability:
  - the mean of the predictions changes with time
  - the variations to this mean vary randomly
- The residuals of the model ($X_t - x_t$) have a constant variance in time (**homoscedastic**)

HO
GENT

Estimating the parameters

Make **predictions** based on the time series model:

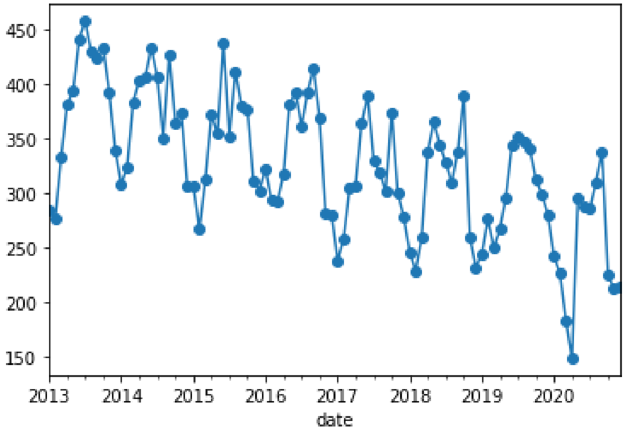
1. select the most suitable model
2. estimation for parameters $b_i (i : 1, \ldots, n)$ based on observations

The estimations $\hat{b_i}$ are selected so that they approximate the observed values as close as possible.

HO
GENT

Example

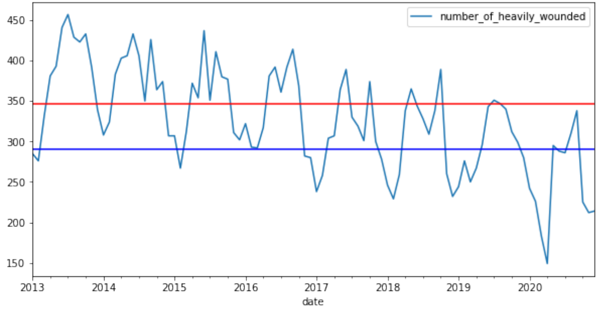
Number of heavily wounded in car accidents in Flanders



HO
GENT

Example: parameter estimation

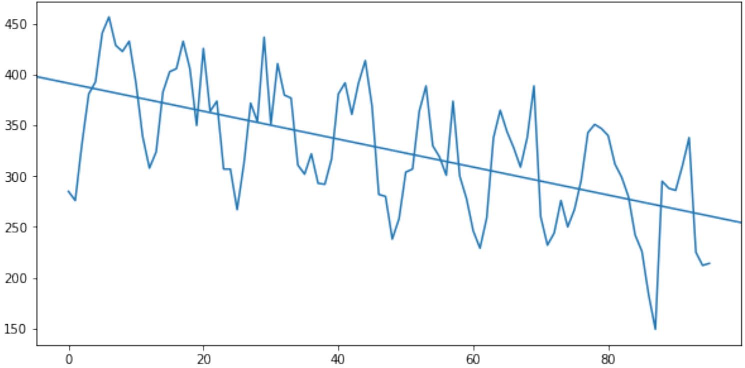
- We select the constant model of Equation 1
- You choose which observations you use to determine $\hat{b}$, e.g.
  - First 70 observations: $\hat{b} = \frac{1}{70}\sum_{t=1}^{70} x_t = 346.4$
  - Last 50 observations: $\hat{b} = \frac{1}{50}\sum_{t=47}^{96} x_t = 290.68$



HO GENT

Example: parameter estimation

If we want to model the observations with a linear function $X_t = b_0 + b_1 t + \varepsilon_t$, then we can use linear regression!



HO GENT

Moving average

HO
GENT

Moving average

Moving Average

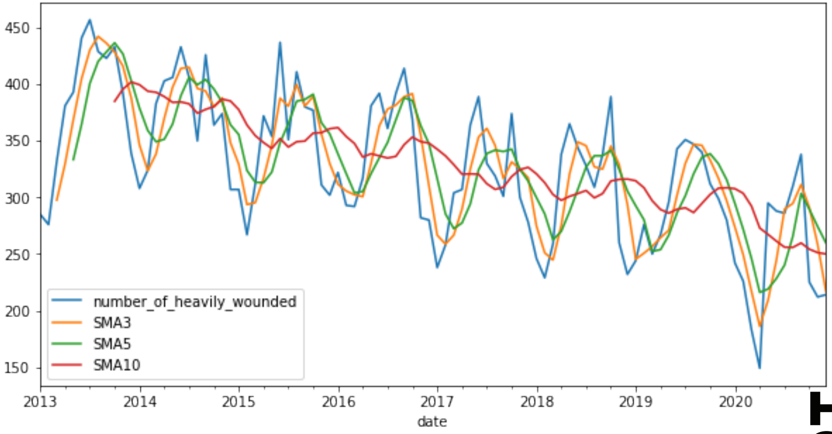
The moving average is a series of averages (means) of the last *m* observations

- Notation: SMA
- Hide short-term fluctuations and show long-term trends
- Parameter m is the time window

$$SMA(t) = \sum_{i=k}^{t} \frac{x_i}{m} \tag{5}$$

with $k = t - m + 1$.

Example



HO
GENT

Example: “Golden cross”

Moving averages are used in the **technical analysis** of stock prices to discover trends:

Weighted moving average

- For SMA, the weights of the observations are equal
- For a weighted moving average (WMA), more recent observations gain relatively more weight
- A specific form of this is single exponential smoothing or the **exponential moving average** (EMA):

$$X_t = \alpha x_{t-1} + (1 - \alpha)X_{t-1} \tag{6}$$

with α the smoothing constant ($0 < \alpha < 1$), and $t \geq 3$

HO
GENT

Exponential smoothing

Equation 6 is only valid from $t = 3$. Hence, we need to choose a suitable value for X_2 ourselves. There are several options:

- $X_2 = x_1$
- $X_2 = \frac{1}{m}\sum_{i=1}^{m} x_i$ (so the mean of the first m observations)
- make X_2 equal to a specific objective
- …

HO
GENT

Exponential Smoothing

Example of calculation (α = 0.1)

t	1	2	3	4	...
x_t	8	10	11	13	...
X_t	—	8	8.2	8.68	...

$$X_1 = \text{undefined}$$
$$X_2 = x_1$$
$$X_3 = 0.1 \times 10 + 0.9 \times 8 = 8.2$$
$$X_4 = 0.1 \times 13 + 0.9 \times 8.2 = 8.68$$

HO GENT

Why “*exponential*”?

$$
\begin{aligned}
X_t &= \alpha x_{t-1} + (1-\alpha)X_{t-1} \\
&= \alpha x_{t-1} + (1-\alpha)\left[\alpha x_{t-2} + (1-\alpha)X_{t-2}\right] \\
&= \alpha x_{t-1} + \alpha(1-\alpha)x_{t-2} + (1-\alpha)^2 X_{t-2}
\end{aligned}
$$

or in general:

$$
= \alpha \sum_{i=1}^{t-2} (1-\alpha)^{i-1} x_{t-i} + (1-\alpha)^{t-i} X_{t-i}, t \geq 2
$$

In other words: older observations have an exponentially smaller weight.

HO GENT

Exponential smoothing

α	$(1-\alpha)$	$(1-\alpha)^2$	$(1-\alpha)^3$	$(1-\alpha)^4$
0.9	0.1	0.01	0.001	0.0001
0.5	0.5	0.25	0.125	0.062
0.1	0.9	0.81	0.729	0.6561

Table: Values for α and $(1-\alpha)^n$

The speed at which the old observations are "forgotten" depends on the value of α. For a value of α close to 1, old observations are quickly forgotten, whereas for α close to 0, this goes less fast.

HO
GENT

Example

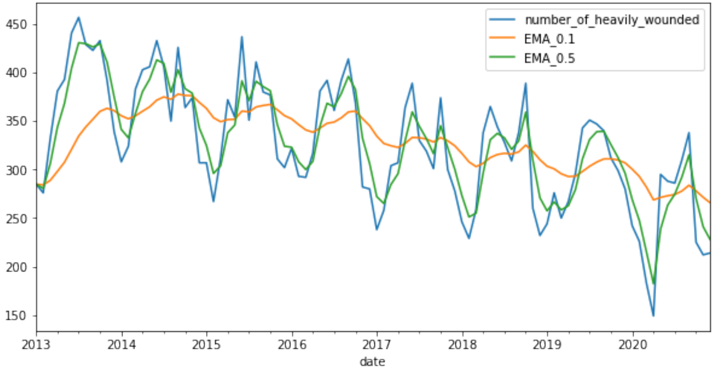
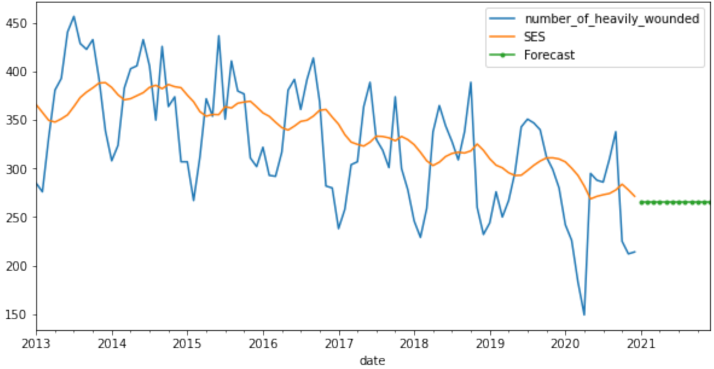


Figure: Single exponential smoothing with $\alpha = 0.1, 0.5$

Forecasting

As a forecast for time $t + m$ (m time units in the "future"), we always take the last estimate of the level:

$$F(t + m) = X_t$$



HO
GENT

Double exponential smoothing

Basic exponential smoothing does not work well if there is a trend in the data

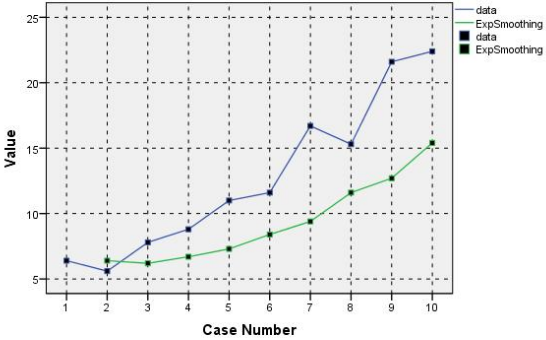


Figure: Exponential smoothing with a trend: the errors keep getting bigger

Double exponential smoothing

We add an additional term to model the trend. We use b_t for the estimation of the trend at time $t > 1$:

$$X_t = \alpha x_t + (1 - \alpha)(X_{t-1} + b_{t-1})$$
$$b_t = \beta(X_t - X_{t-1}) + (1 - \beta)b_{t-1}$$

with $0 < \alpha < 1$ and $0 < \beta < 1$

- b_t is an estimate for the slope of the trend line
- Added to the first equation to ensure that the trend is followed
- $X_t - X_{t-1}$ is positive or negative, this corresponds to an increasing/decreasing trend

HO GENT

Double exponential smoothing

Again, there are different options for selecting the initial values:

$$X_1 = x_1$$

$$b_1 = x_2 - x_1$$

$$b_1 = \frac{1}{3}\left[(x_2 - x_1) + (x_1 - x_2) + (x_4 - x_3)\right]$$

$$b_1 = \frac{x_n - x_1}{n - 1}$$

Predicting (forecasting)

To make a prediction (forecast) $F(t + 1)$ for time $t + 1$ we use:

$$F(t + 1) = X_t + b_t$$

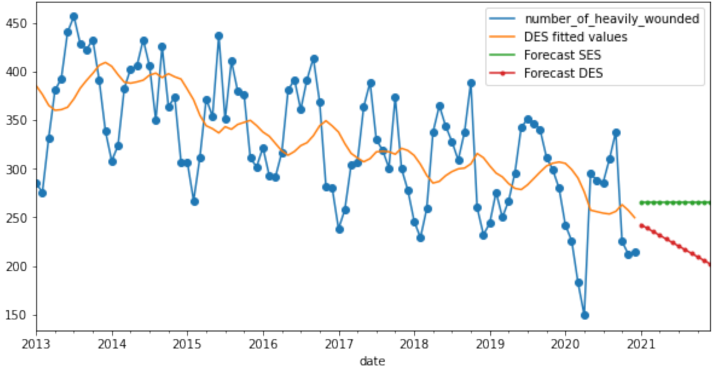
or in general for time $t + m$:

$$F(t + m) = X_t + mb_t$$

HO
GENT

Example

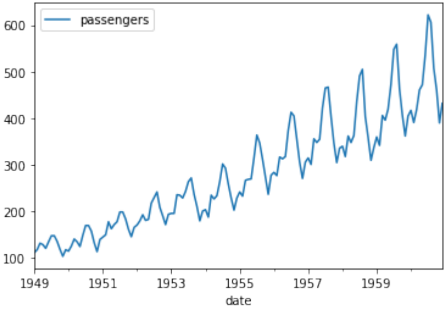
Comparison of Single/Double Exponential Smoothing forecasts



HO
GENT

Triple exponential smoothing

Some time series have recurring (seasonal) patterns, e.g.



Triple exponential smoothing

Or Holt-Winter's Method

Notation:

- L: length of the seasonal cycle (number of time units)
- c_t: term that models the seasonal variations
- γ (gamma): smoothing factor for the seasonal variation

$$X_t = \alpha \frac{x_t}{c_{t-L}} + (1-\alpha)(X_{t-1} + b_{t-1})$$ Smoothing

$$b_t = \beta(X_t - X_{t-1}) + (1-\beta)b_{t-1}$$ Trend smoothing

$$c_t = \gamma \frac{x_t}{X_t} + (1-\gamma)c_{t-L}$$ Seasonal smoothing

HO
GENT

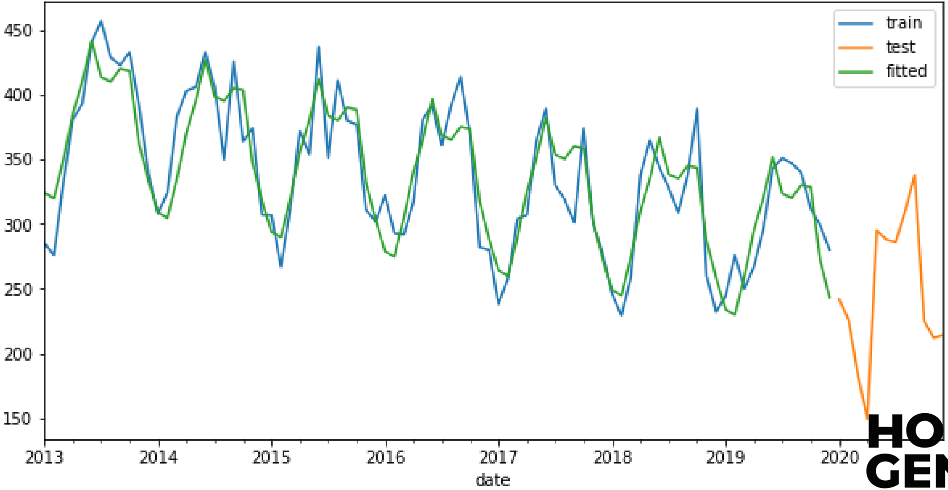
Triple exponential smoothing

Prediction at time $t + m$:

$$F_{t+m} = (X_t + mb_t)c_{t-L+m}$$

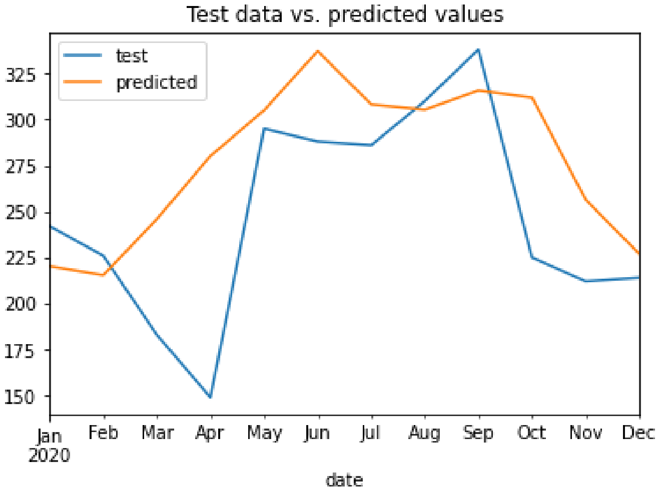
HO
GENT

Example



HO GENT

Quality of a time series model



HO
GENT

Quality of a time series model

Compare forecast results with actual observations, when they become available:

- Mean absolute error: $MAE = \frac{1}{m} \sum_{i=t+1}^{t+m} \left| x_i - F_i \right|$
- Mean squared error $MSE = \frac{1}{m} \sum_{i=t+1}^{t+m} (x_i - F_i)^2$

If square root of MSE is well below standard deviation over all observations, you have a good model!

HO
GENT