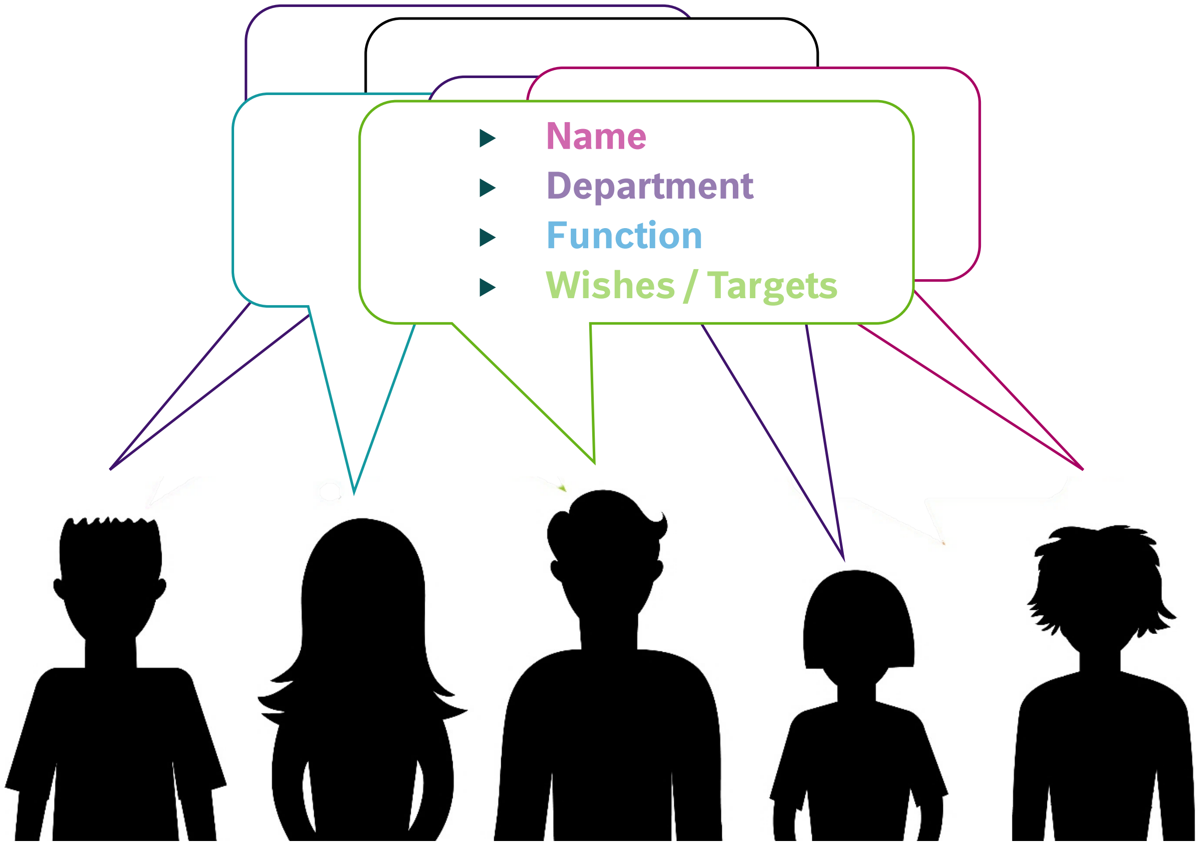
TIMESERIES FORECASTING (A)

ALEXANDROS PATELIS (CR/AEU2)

BOSCH

Timeseries Forecasting
Introduction of Participants and Trainer



Timeseries Forecasting
Introduction of Trainer: Patelis Alexandros (CR\AEU2)

- Timeseries Forecasting for Decision Support Systems
  - PhD, National Technical University of Athens, 1999

- ***Forecasting Working Fields***:
  - Stock Exchange
  - Banking
  - Insurance
  - Media & Advertisement
  - Telecommunications
  - Energy
  - Environment
  - Sales
  - Production
  - Customer Management
  - Travel & Tourism
  - Logistics
  - Real Estate
  - Medicine
  - Military

Timeseries Forecasting
Introduction of Participants

	Name	Department	Function	Wishes / Targets
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Timeseries Forecasting: Advanced
Agenda

1 Recap Beginners

2 Preparation & Analysis

3 Errors & Accuracy (More)

4 Decomposition

5 Theta Classical

6 Theta Alternative

7 Averaging

8 Arima

9 Multiple Regression

10 Limits of Predictability

11 Neural Networks

12 Special Topic Working Days Adj.

13 Special Topic Structural Changes

14 Special Topic Intermittent Demand

15 Use Case Inventory Demand

16 Use Case Energy Management

17 Use Case Social Media

18 More… ?

Timeseries Forecasting
Content

- What we will see today:
  - Some theory…
    - but not more than necessary!

  - Mathematical equations…
    - but only for completion!

$$S(\omega) = \frac{\alpha g^2}{\omega^5} e^{\left[-0.74\left\{\frac{\omega U_\omega 19.5}{g}\right\}^{-4}\right]}$$
$$= \frac{\alpha g^2}{\omega^5} \exp\left[-0.74\left\{\frac{\omega U_\omega 19.5}{g}\right\}^{-4}\right]$$

- But also:
  - Analytical examples 
  - Key messages & conclusions 
  - Best advices & rules of thumb 
  - Live examples with your participation 

BOSCH

1. RECAP (BEGINNERS)

2. TIMESERIE PREPARATION & ANALYSIS

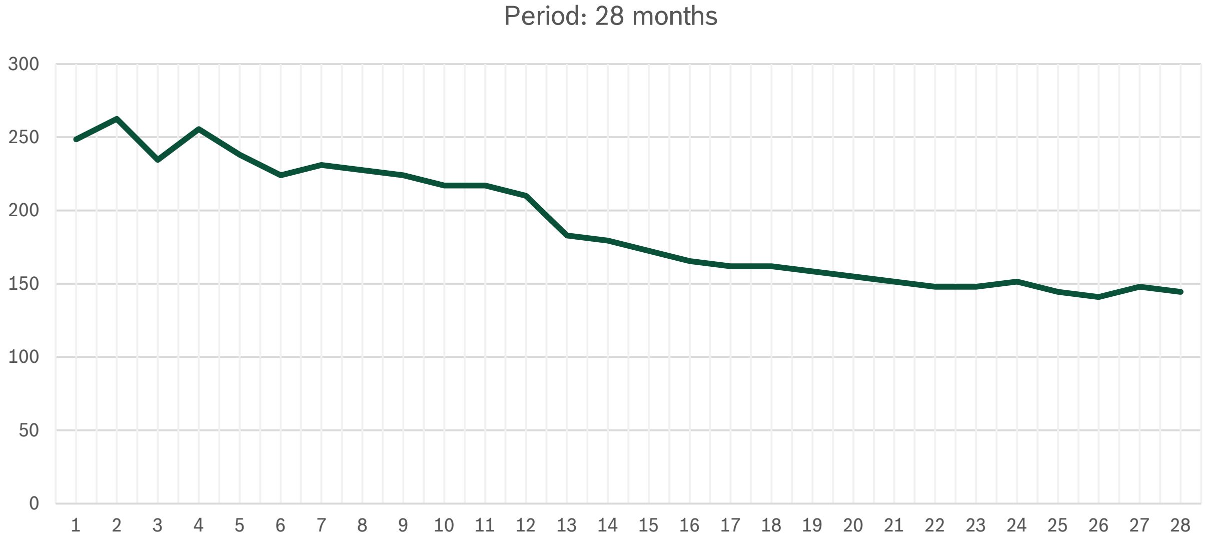
"By failing to prepare, you are preparing to fail"

Benjamin Franklin

BOSCH

Timeserie Preparation & Analysis
Volume of historical data

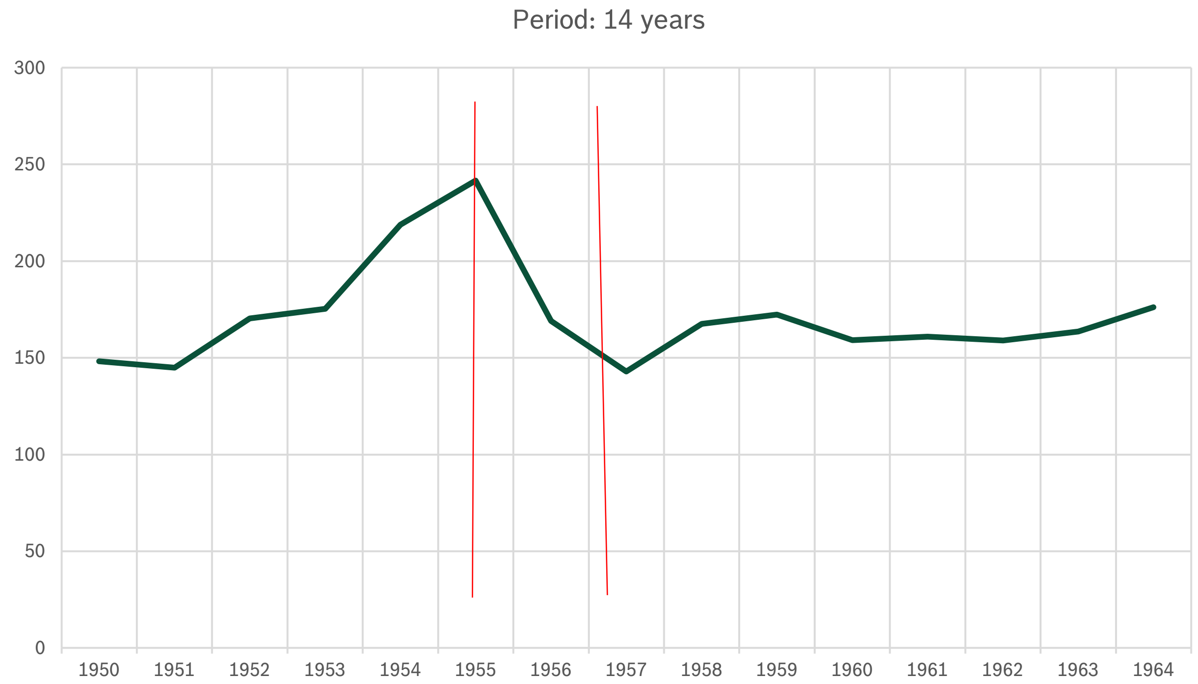
- In general, it is good to have ***as many historical data as possible***.
  - Not all historical data will be used in a method, but they will help to identify timeseries characteristics.
  - We have to ***act with caution***, when selecting historical data, since ***they will affect the results*** of our model selection and prediction results.
  - Example: Monthly prices, 28 months
    - ***Decreased trend***
    - R^2 = 0,974
    - Low error variance



Source: "Micro-Economics", Makridakis, Wheelwright & Hyndman, 1998).

Timeserie Preparation & Analysis
Volume of historical data

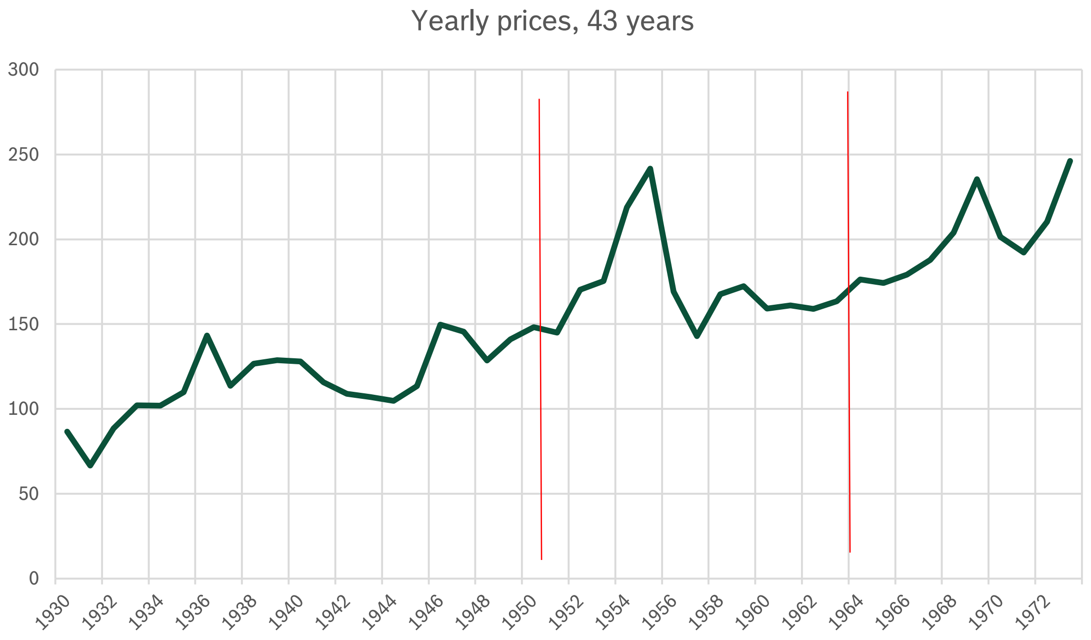
- Bigger picture: Yearly prices, 14 years
  - ***No decreased trend***, stable prices
  - $R^2 = 0{,}007$
  - Low error variance



Source: “Micro-Economics", Makridakis, Wheelwright & Hyndman, 1998).

Timeserie Preparation & Analysis
Volume of historical data

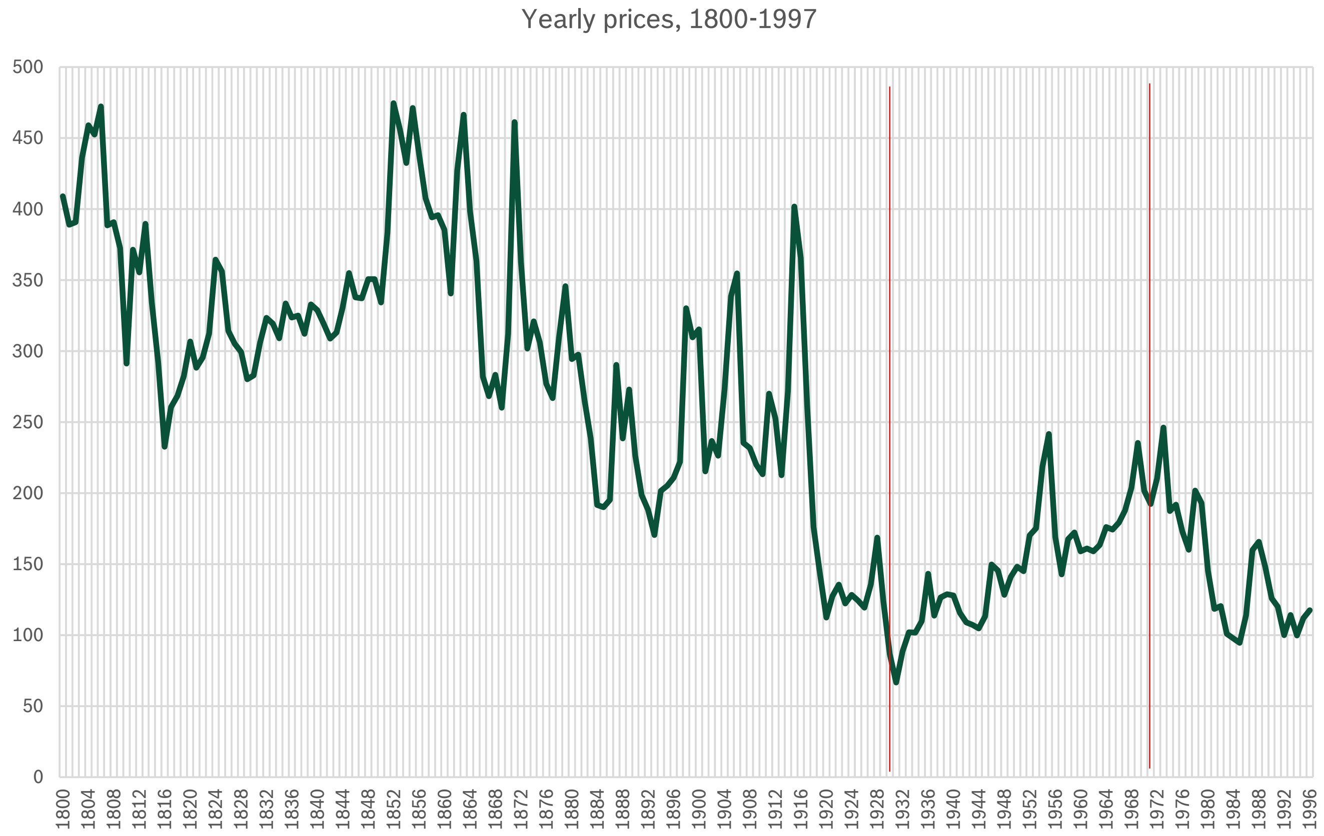
- Bigger picture: Yearly prices, 43 years
  - Now we observe ***increased trend***
  - R^2 = 0,743
  - Low error variance



Source: "Micro-Economics", Makridakis, Wheelwright & Hyndman, 1998).

Timeserie Preparation & Analysis
Volume of historical data

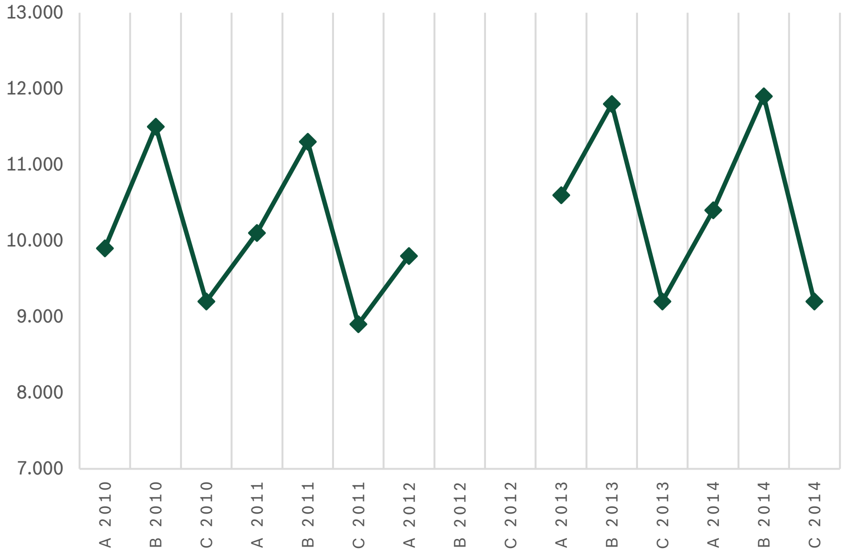
- The biggest picture: Yearly prices, almost 200 years
  - ***Decreased trend***
  - Cyclical behavior, with different duration
  - R^2 = 0,618
  - Low error variance



Source: "Micro-Economics", Makridakis, Wheelwright & Hyndman, 1998).

Timeserie Preparation & Analysis
Managing fault and missing values

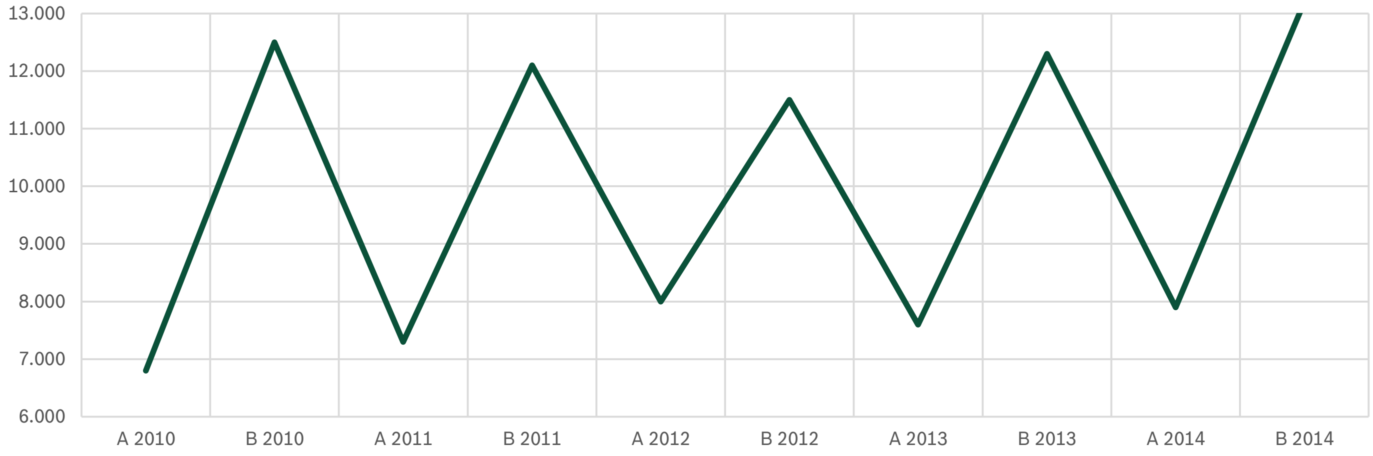
- Manage the missing values before applying a forecast method:
  - Try to find the missing values from ***other possible sources***.
  - Direct define the missing values if a safe ***judgmental estimation*** exists.
  - ***Statistically estimate*** the missing values.



Timeserie Preparation & Analysis
Non-Working Days Adjustments

An adjustment could be made, by taking into account ***the number of working days*** in each period.

- Define the number of working days in each period.
- Define the country and find the holidays.
- Adjust the data point values



Timeserie Preparation & Analysis
Outliers / Special Events & Actions

Level difference between current data of a timeserie and an initial level (or the level of the average value), could be the impact from a special event or action.

- Abnormal values (outliers)
- Level change (level shifts)

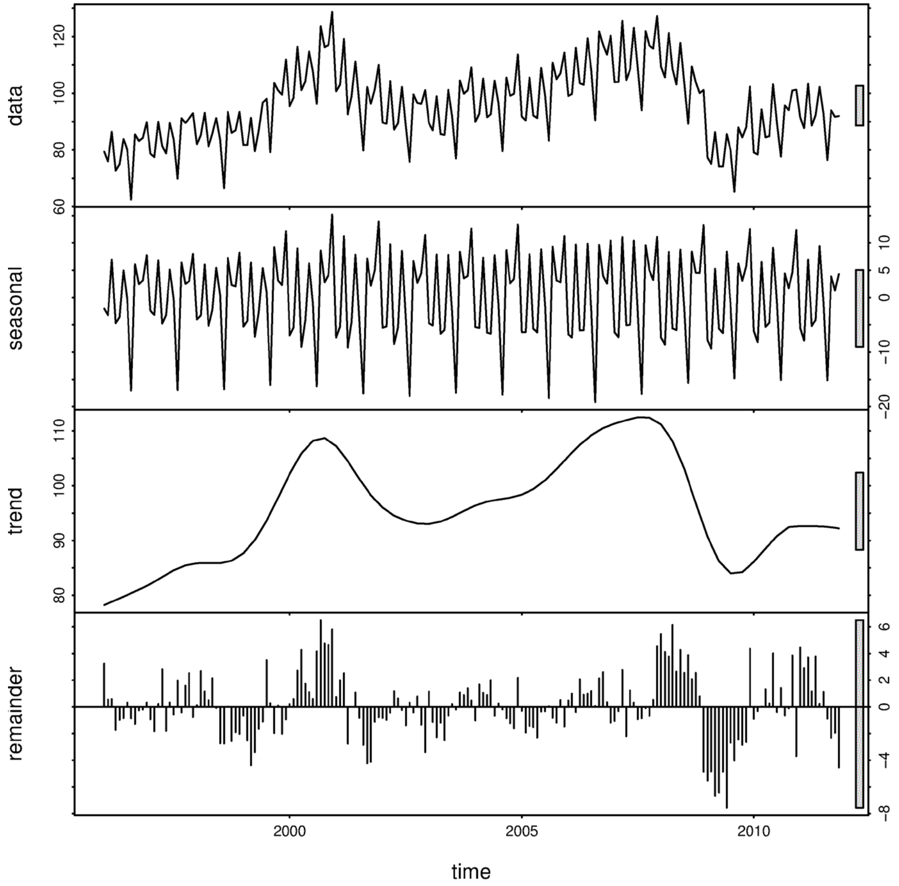
➢ Systematic recording & classification of impacts of various types of SEA

- could lead to more **accurate forecasts**, through appropriate critical interventions on statistical forecasts.

➢ These interventions will be based on **ratios** of past and upcoming events and actions.

Timeserie Preparation & Analysis
Estimate Components

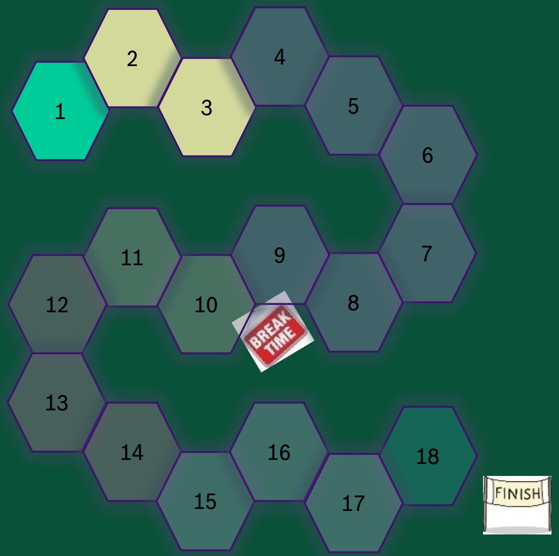
- Deconstructs a timeserie (Y) into the basic four parts (components):
  - ***Trend*** (T) component.
  - ***Cyclical*** (C) component.
  - ***Seasonal*** (S) component.
  - ***Random*** (R) or Irregular component.



BOSCH

Source: Forecasting – Principles and Practice

3. FORECASTING ERRORS & ACCURACY (MORE…)



"Some of the biggest, most sophisticated organizations in the world – which spend millions on forecasting – do not scientifically test the accuracy of the forecasting…"

Dan Gardner

Forecasting Errors & Accuracy (More)
Recap from basics

- ***Error*** (e): The difference between the actual value and the forecasted value, for a given time.

$$e_i = Y_i - F_i$$

- ***Mean Error (ME)***: The average error between the actual values and the forecasted values, for a given period.

$$ME = \frac{\sum_{i=1}^{n}(Y_i - F_i)}{n}$$

- ***Mean Absolute Error (MAE)***: The average absolute error for a given period

$$MAE = \frac{\sum_{i=1}^{n}|Y_i - F_i|}{n}$$

- ***Mean Squared Error (MSE)***: Average of all square errors

$$MSE = \frac{\sum_{i=1}^{n}(Y_i - F_i)^2}{n}$$

t	Y_t	$\widehat{Y_t}$	e_t	$\lvert e_t \rvert$	e_t^2
1	33	31	2	2	4
2	49	42	7	7	49
3	52	50	2	2	4
4	57	61	-4	4	16
5	78	73	5	5	25
6	83	85	-2	2	4
7	90	94	-4	4	16
8	112	103	9	9	81
9	118	115	3	3	9
10	116	124	-8	8	64
11		132			
12		141			
		sum	**10,00**	**46,00**	**272,00**
		Mean	**1,00**	**4,60**	**27,20**

$$e_1 = Y_1 - F_1 = 33 - 31 = \mathbf{2}$$

$$ME = \frac{1}{n} \times \sum_{i=1}^{n} e_1 = \frac{2 + 7 + \cdots - 8}{10} = \frac{10}{10} = \mathbf{1}$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} |e_1| = \frac{|2| + |7| + \cdots + |-8|}{10} = \frac{46}{10} = \mathbf{4,6}$$

$$MSE = \frac{1}{n} \times \sum_{i=1}^{n} (e_1)^2 = \frac{2^2 + 7^2 + \cdots + (-8)^2}{10} = \frac{272}{10} = \mathbf{27,2}$$

BOSCH

Forecasting Errors & Accuracy (More)
Error Types

- ***Root Mean Squared Error (RMSE)***: Root of the average of all square errors
  - $RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - F_i)^2}{n}}$
  - The use of RMSE is very common and it makes an excellent general purpose error metric for numerical forecasts.
- ***Mean Absolute Percentage Error (MAPE)***: the average absolute % error for each time period divided by actuals
  - $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - F_i}{Y_i}\right|100$
  - Most commonly used, but with zeros or near-zeros, it can give a distorted picture of error.

$$RMSE = \sqrt{MSE} = \mathbf{5,22}$$

$$MAPE = \frac{1}{n} \times \sum_{i=1}^{n} \frac{|e_i|}{Y_i} 100\% = \frac{\frac{|2|}{33} + \frac{|7|}{49} + \cdots + \frac{|-8|}{116}}{10} 100\% = \frac{0,619}{10} 100\% = \mathbf{6,19}\%$$

Forecasting Errors & Accuracy (More)
Error Types

t	Y_t	$\widehat{Y_t}$	e_t
1	33	31	2
2	49	42	7
3	52	50	2
4	57	61	-4
5	78	73	5
6	83	85	-2
7	90	94	-4
8	112	103	9
9	118	115	3
10	116	124	-8
11		132	
12		141	

- ***Symmetric Mean Absolute Percentage Error (sMAPE)***: Alternative to MAPE
  - $$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Y_i-F_i)}{\frac{(Y_i+F_i)}{2}}\right|100 = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{2(Y_i-F_i)}{(Y_i+F_i)}\right|100$$

$$sMAPE = \frac{1}{n}\times\sum_{i=1}^{n}\frac{2\times|e_i|}{Y_i}100\% = 2\,\frac{\frac{|2|}{33+31}+\frac{|7|}{49+42}+\cdots+\frac{|-8|}{116+124}}{10}100\% = 2\,\frac{0{,}317}{10}100\% = \mathbf{6{,}33\%}$$

- ***Mean Absolute Scaled Error (MASE)***:
  - $$MAsE = \frac{\frac{1}{n}\sum_{i=1}^{n}|Y_i-F_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i-Y_{i-1}|}$$
  - Tends to be the a standard when comparing forecasting accuracies.

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|e_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i-Y_{i-1}|} = \frac{\frac{|2|+|7|+\cdots+|-8|}{10}}{\frac{|49-33|+|52-49|+\cdots+|116-118|}{9}} = \frac{\frac{46}{10}}{\frac{87}{9}} = \mathbf{0{,}476}$$

Forecasting Errors & Accuracy (More)
Errors & Accuracy

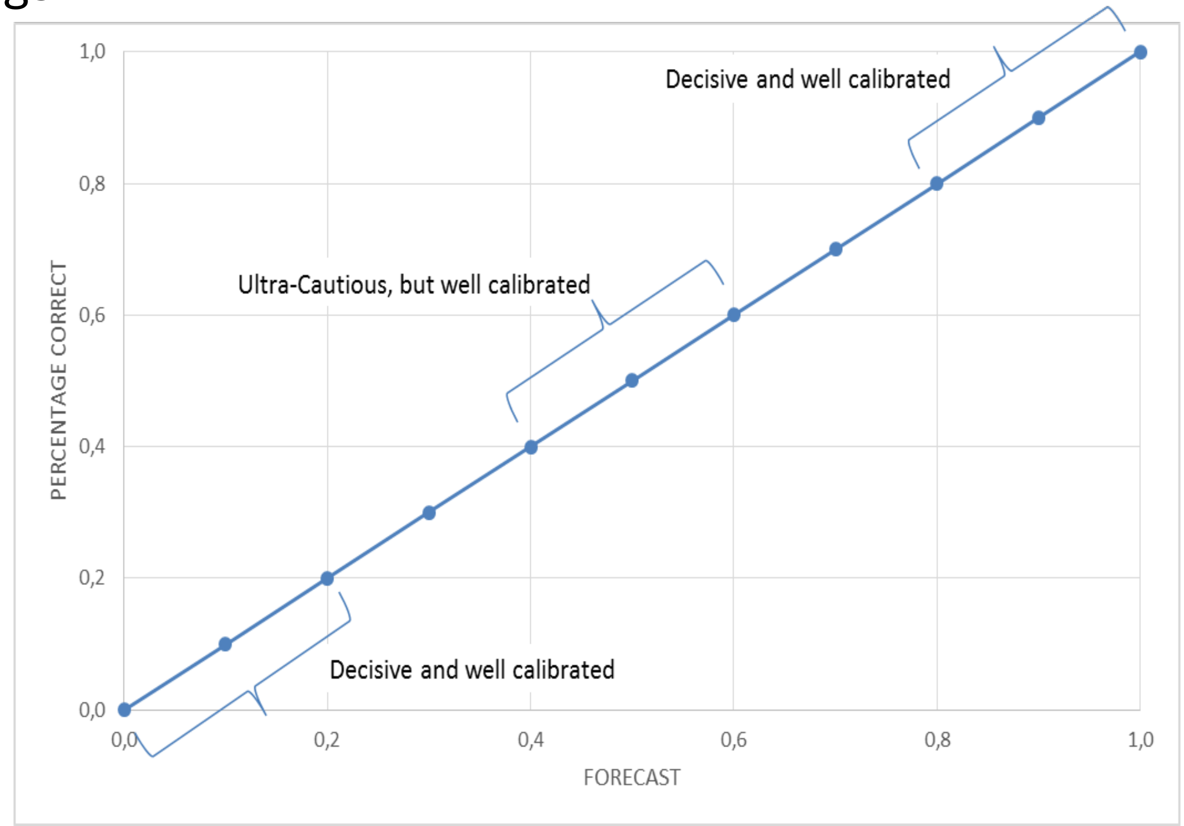
- In order to evaluate forecasts for accuracy, we have to be able to understand exactly what the forecast says. ***This is more difficult that it seems.***

- What is probability to happen?
  - A forecaster says that ***it is possible*** that something will happen.
  - What does “***possible***” means?
  - People were asked for their thoughts on the exact likelihood it would happen:
    - Answers ranged from 20% to 80% probability, illustrating that sometimes people have very different ideas about what something means.

Forecasting Errors & Accuracy (More)
Errors & Accuracy

- Avoid confusion: designate numerical meanings

Term	Meaning	Give or Take
Certain	100 %	-
Almost certain	93 %	6 %
Probable	75 %	12 %
Chances about even	50 %	10 %
Probably not	30 %	10 %
Almost certainly not	7 %	5 %
Impossible	0 %	-



Forecasting Errors & Accuracy (More)
Brier score

- Developed by Glenn W. Brier in 1950
- Measures the distance between what you forecast and what actually happened
- Lower is better!
  - Perfection: 0
  - A 50-50 call, or random guessing: 0.5
  - Worst: 2.0

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

- ***Depends on what it is being forecasted***

- Example: A yes/no question
  - Forecaster A: 0% will happen
  - Forecaster A: 50% will happen
  - Forecaster B: 80% will happen
  - Forecaster D: 100% will happen
- Actual: It happened

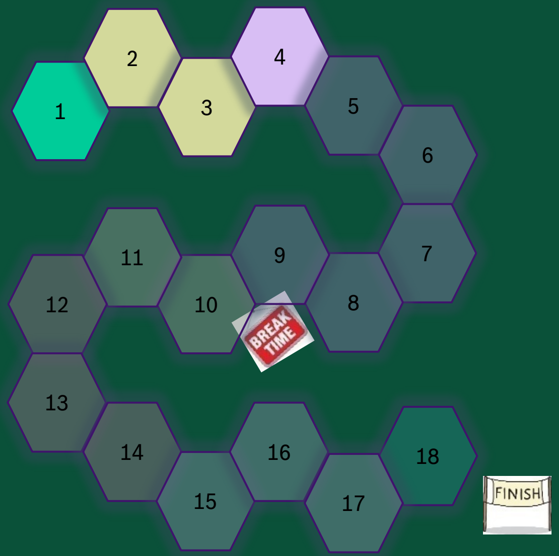
$$BS_A = (1-0.0)^2+(0-1.0)^2 = 2.00$$

$$BS_B = (1-0.5)^2+(0-0.5)^2 = 0.50$$

$$BS_C = (1-0.8)^2+(0-0.2)^2 = 0.08$$

$$BS_D = (1-1.0)^2+(0-0.0)^2 = 0.00$$

4. DECOMPOSITION



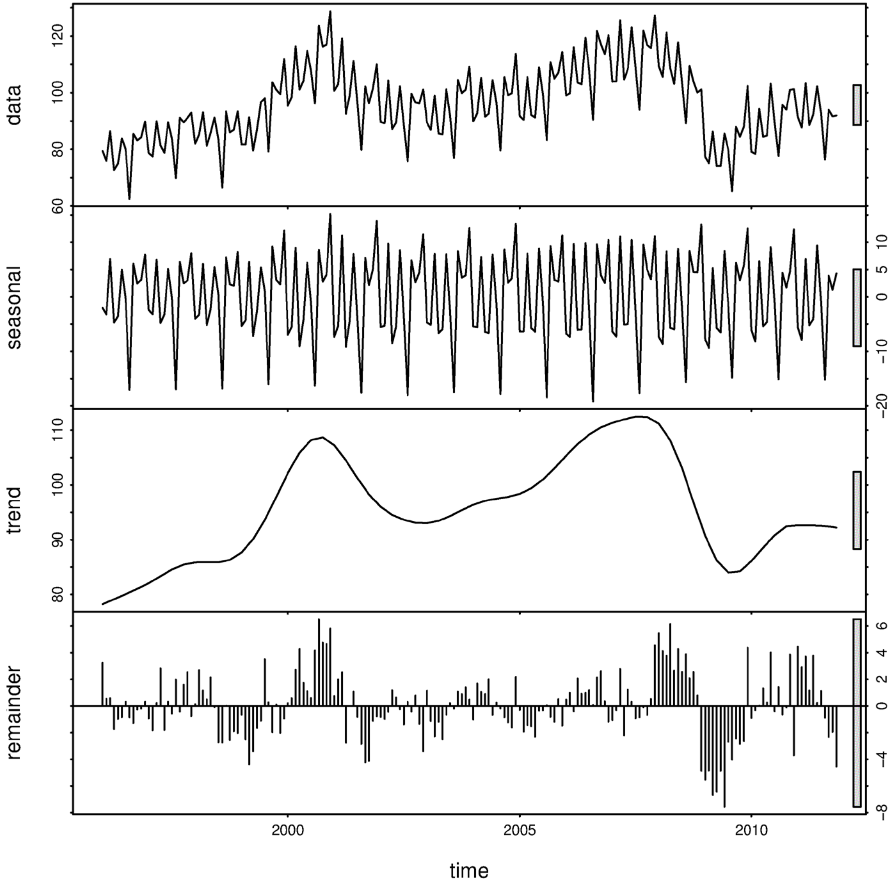
Decomposition
Definition

Decomposition:

- A statistical method that deconstructs a timeserie (Y) into four basic parts (components):
  - ***Trend*** (T) component.
  - ***Cyclical*** (C) component.
  - ***Seasonal*** (S) component.
  - ***Random*** (R) or Irregular component.

$$Y_t = f(S_t, T_t, C_t, R_t)$$

- Two forms of decomposition:
  - ***Additive*** $Y_t = S_t + T_t + C_t + R_t$
  - ***Multiplicative*** $Y_t = S_t \times T_t \times C_t \times R_t$



Source: Forecasting – Principles and Practice

BOSCH

Decomposition
Definition

Deseasonalization:

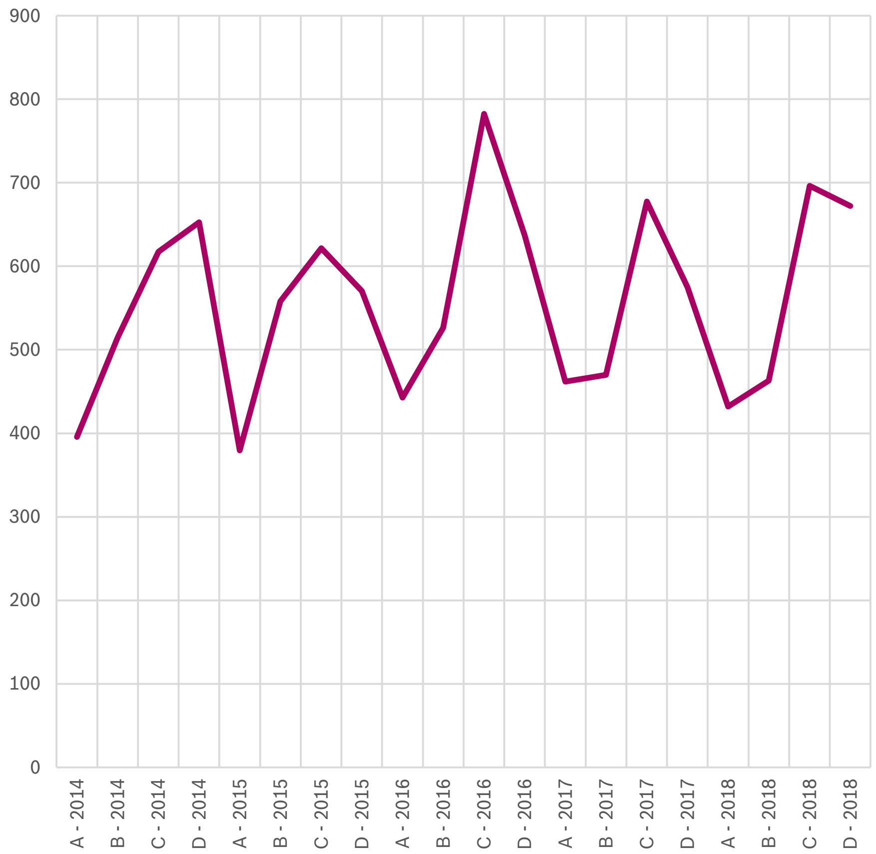
- A method for:
  - ***isolating the seasonal component*** from a timeserie, and
  - creating the deseasonalized version of the timeserie

- We must always apply forecasting methods on the deseasonalized timeseries.
  - This increases the reliability of the forecasts.

- Commonly used:
  - ***Classical Decomposition method***, based on Moving Averages

Decomposition
Observing seasonal effect

Data	
Year	**Value**
A - 2014	395,5
B - 2014	514,4
C - 2014	617,5
D - 2014	652,8
A - 2015	379,4
B - 2015	558,1
C - 2015	621,7
D - 2015	570,3
A - 2016	442,9
B - 2016	526,1
C - 2016	782,3
D - 2016	636,9
A - 2017	461,8
B - 2017	469,9
C - 2017	677,7
D - 2017	575,0
A - 2018	432,1
B - 2018	462,9
C - 2018	696,2
D - 2018	672,2
A - 2019	
B - 2019	
C - 2019	
D - 2019	



- What you can observe?
  - 20 data points for 5 years
  - Quarterly data
  - Trend ?
  - No “fault” or “missing” values
  - Quarters are affecting values

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Decomposition
Step 1

Step 1: ***Calculate Moving Average***

- MA(n): based on the length **n** of the seasonality:
  - example: MA(4) for quarterly timeserie, MA(12) for monthly timeserie.
  - Does not include seasonality, and
  - Contains very little or no randomness, since randomness is represented by random fluctuations that range around the average value of observations.
  - Gives a good estimate of the behaviour of timeseries Trend & Cyclical components.

$$MA_n = T \times C$$

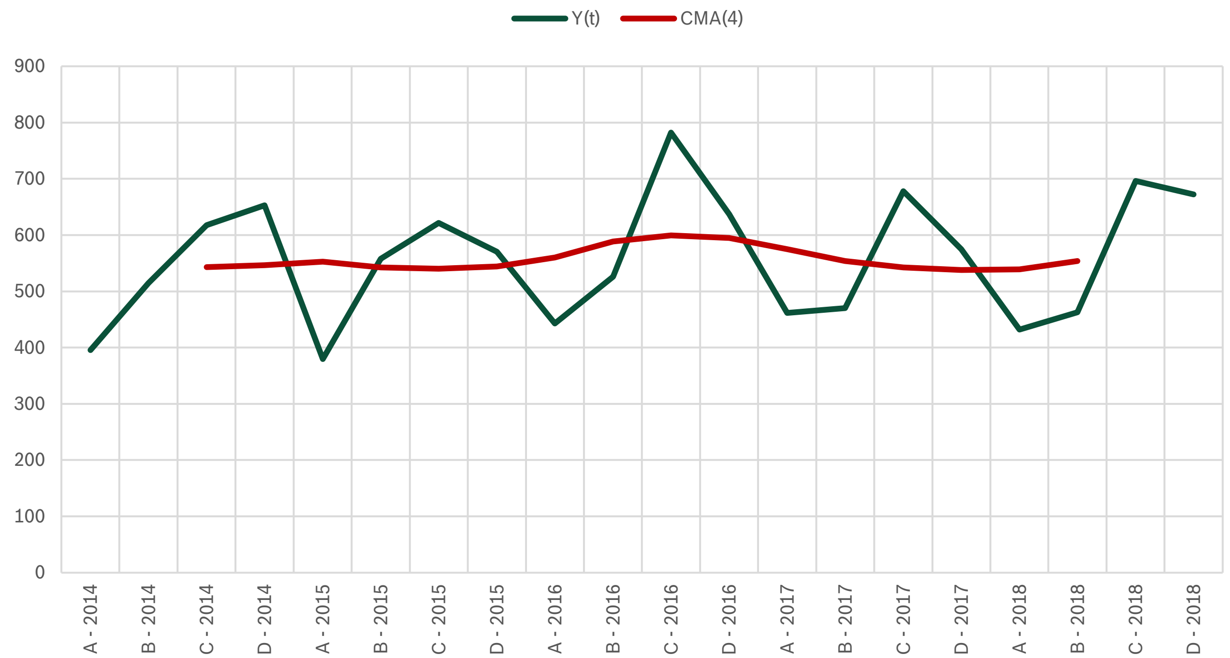
In most cases a SMA is used.

But, in the case when the seasonality length n is even (like in quarterly data with n = 4), then the usage of CMA is preferred.

Decomposition
Step 1

Step 1: Calculate Moving Average – Example

	Y(t)	CMA(4)
A - 2014	395,5	
B - 2014	514,4	
C - 2014	617,5	**543,0**
D - 2014	652,8	546,5
A - 2015	379,4	552,5
B - 2015	558,1	542,7
C - 2015	621,7	540,3
D - 2015	570,3	544,3
A - 2016	442,9	560,3
B - 2016	526,1	588,7
C - 2016	782,3	599,4
D - 2016	636,9	594,8
A - 2017	461,8	574,7
B - 2017	469,9	553,8
C - 2017	677,7	542,4
D - 2017	575,0	537,8
A - 2018	432,1	539,2
B - 2018	462,9	553,7
C - 2018	696,2	
D - 2018	672,2	



Decomposition
Step 2

Step 2: Initial Estimation of ***Seasonality***

- Divide the actual observations Y_t with the T*C values (that we estimate on the 1st step).
- With this way, we can have a first estimation of Seasonality, which has also randomness.

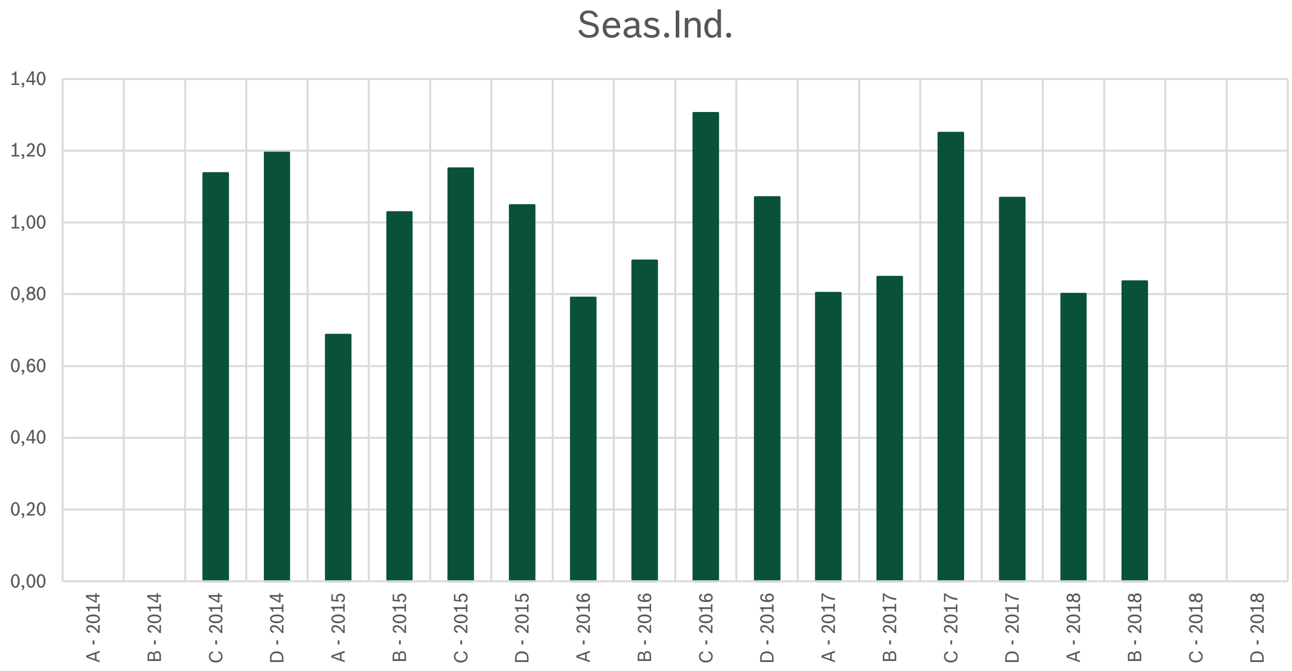
$$\frac{Y}{MA_n} = \frac{T \times C \times S \times R}{T \times C} = S \times R$$

BOSCH

Decomposition
Step 2

Step 2: Initial Estimation of Seasonality – Example

	Y(t)	CMA(4)	Seas.Ind.
A - 2014	395,5		
B - 2014	514,4		
C - 2014	617,5	543,0	1,137
D - 2014	652,8	546,5	1,195
A - 2015	379,4	552,5	0,687
B - 2015	558,1	542,7	1,028
C - 2015	621,7	540,3	1,151
D - 2015	570,3	544,3	1,048
A - 2016	442,9	560,3	0,790
B - 2016	526,1	588,7	0,894
C - 2016	782,3	599,4	1,305
D - 2016	636,9	594,8	1,071
A - 2017	461,8	574,7	0,804
B - 2017	469,9	553,8	0,848
C - 2017	677,7	542,4	1,249
D - 2017	575,0	537,8	1,069
A - 2018	432,1	539,2	0,801
B - 2018	462,9	553,7	0,836
C - 2018	696,2		
D - 2018	672,2		



BOSCH

Decomposition
Step 3

Step 3: ***Remove Randomness*** from Seasonality

- This is achieved by finding the ***average of the corresponding seasonality indices***.
  - For a monthly timeserie, we calculate the seasonal indice for Quarter A, by averaging all seasonality indices corresponding to January.
- Do we need normalization?
  - In some cases, the seasonality indices need to be normalized, so that their ***sum is equal to the seasonality*** length *n*.
  - If the timeserie contains ***significant randomness or outliers***, it is suggested to use the Intermediate Averages, for the calculation of seasonality indices. The difference lies in the ***non-use of maximum and minimum*** seasonality indice in the calculation, in order to stabilize them.

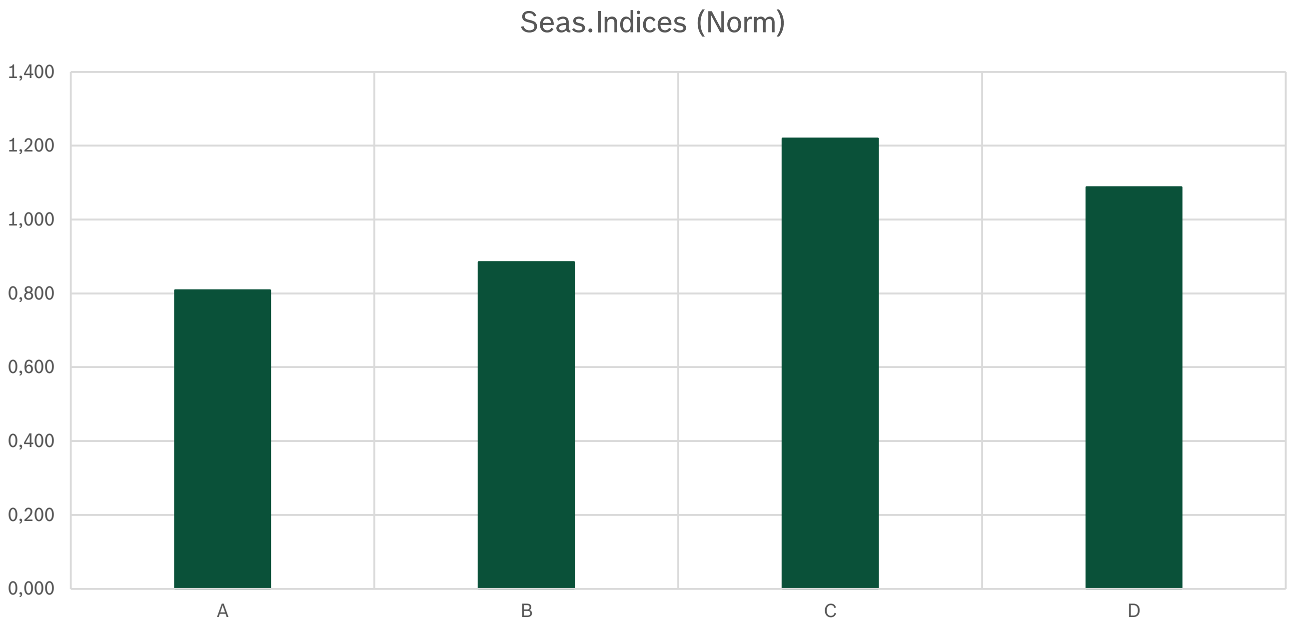
Decomposition
Step 3

Step 3: Remove Randomness from Seasonality – Example

	Seas.Ind.								
	2014	**2015**	**2016**	**2017**	**2018**	**min**	**max**	**Ind**	**Seas. Indices**
A		0,687	0,790	0,804	0,801	0,687	0,804	0,796	**0,809**
B		1,028	0,894	0,848	0,836	0,836	1,028	0,871	**0,885**
C	1,137	1,151	1,305	1,249		1,137	1,305	1,200	**1,219**
D	1,195	1,048	1,071	1,069		1,048	1,195	1,070	**1,087**
							sum	**3,937**	**4,000**

$$NF = \frac{\sum_{i=1}^{n} I_i}{n} = \frac{3{,}937}{4} = 0{,}9842$$

$$NI_i = \frac{I_i}{NF}$$



Decomposition
Step 4

Step 4: Estimate ***deseasonalized timeserie***

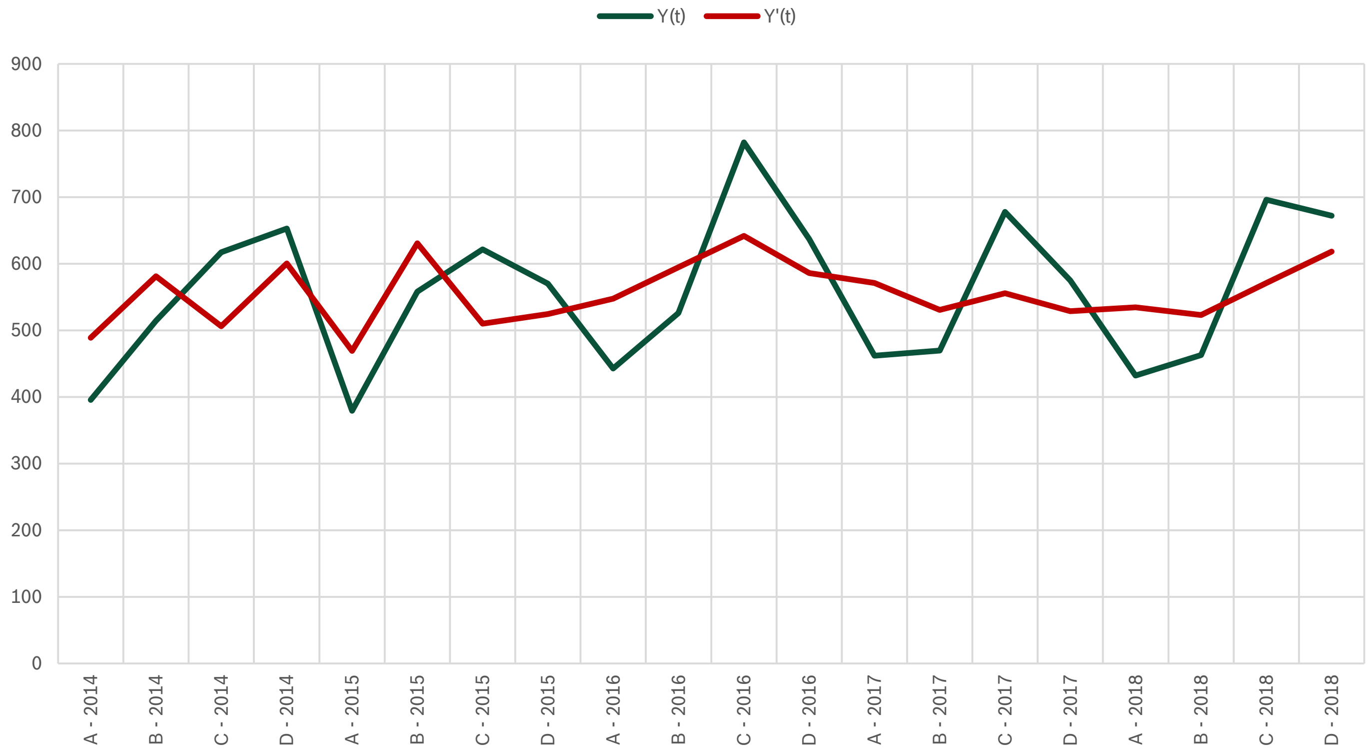
- Divide the actual observations with the seasonality indices, in order to estimate the deseasonalized timeserie.
- This timeserie now includes only Trend, Cycle and Random component.

$$\frac{Y}{S} = \frac{T \times C \times S \times R}{S} = T \times C \times R$$

Decomposition
Step 4

Step 4: Estimate deseasonalized Timeserie – Example

	Y(t)	Seas.Ind.	Y'(t)
A - 2014	395,5	0,809	489,1
B - 2014	514,4	0,885	581,3
C - 2014	617,5	1,219	506,5
D - 2014	652,8	1,087	600,5
A - 2015	379,4	0,809	469,2
B - 2015	558,1	0,885	630,6
C - 2015	621,7	1,219	509,9
D - 2015	570,3	1,087	524,6
A - 2016	442,9	0,809	547,7
B - 2016	526,1	0,885	594,5
C - 2016	782,3	1,219	641,6
D - 2016	636,9	1,087	585,8
A - 2017	461,8	0,809	571,1
B - 2017	469,9	0,885	531,0
C - 2017	677,7	1,219	555,8
D - 2017	575,0	1,087	528,9
A - 2018	432,1	0,809	534,4
B - 2018	462,9	0,885	523,1
C - 2018	696,2	1,219	571,0
D - 2018	672,2	1,087	618,3



BOSCH

Decomposition
Step 5

Step 5: ***Remove Randomness*** from deseasonalized Timeserie

- Calculate a ***Moving Average*** MA(3) or MA(6) of the deseasonalized timeserie.
- This new MA values are fairly smooth, they are an estimation of the Trend-Cycle component.
- For achieving optimal smoothing and elimination of randomness, it is recommended to use ***Double Moving Average DMA(3x3)***.

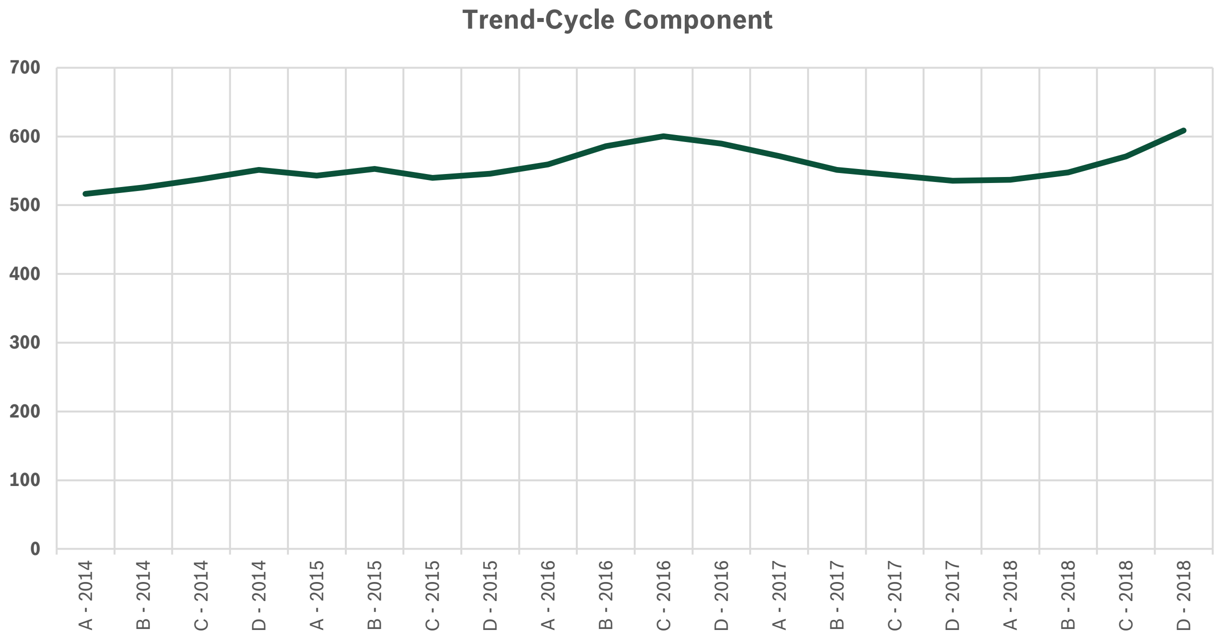
$$CMA(3x3) = T \times C$$

$$\frac{T \times C \times R}{CMA(3x3)} = \frac{T \times C \times R}{T \times C} = R$$

Decomposition
Step 5

Step 5: Remove Randomness from deseasonalized Timeserie – Example

	Y(t)	Seas.Ind.	Y'(t)	CMA(3)	CMA(3x3)
A - 2014	395,5	0,809	489,1		516,6
B - 2014	514,4	0,885	581,3	**525,6**	525,6
C - 2014	617,5	1,219	506,5	562,7	537,9
D - 2014	652,8	1,087	600,5	525,4	551,6
A - 2015	379,4	0,809	469,2	566,8	542,9
B - 2015	558,1	0,885	630,6	536,6	552,8
C - 2015	621,7	1,219	509,9	555,0	539,7
D - 2015	570,3	1,087	524,6	527,4	546,0
A - 2016	442,9	0,809	547,7	555,6	559,2
B - 2016	526,1	0,885	594,5	594,6	585,8
C - 2016	782,3	1,219	641,6	607,3	600,5
D - 2016	636,9	1,087	585,8	599,5	589,8
A - 2017	461,8	0,809	571,1	562,6	571,6
B - 2017	469,9	0,885	531,0	552,6	551,3
C - 2017	677,7	1,219	555,8	538,6	543,6
D - 2017	575,0	1,087	528,9	539,7	535,7
A - 2018	432,1	0,809	534,4	528,8	537,1
B - 2018	462,9	0,885	523,1	542,8	547,5
C - 2018	696,2	1,219	571,0	570,8	570,8
D - 2018	672,2	1,087	618,3		608,6



Decomposition
Step 6

Step 6: Calculate the ***Trend***

- The result of the previous step is a timeserie that includes only Trend and Cycle components.
- In case, that we need to separate these 2 components, we must find the Trend model which describes best the timeserie.
- If we assume ***Linear Trend***, then the calculation of the Trend component can be achieved by using simple linear regression (SLR) with a least square straight line.

$$T = a + \beta * t \qquad \beta = \frac{\frac{\sum_{i=1}^{n}(t_i * TC_i)}{n} - (\bar{t} - \overline{TC})}{\frac{\sum_{i=1}^{n} t_i^2}{n} - \overline{t^2}}, \qquad \alpha = \overline{TC} - (\beta * \bar{t})$$

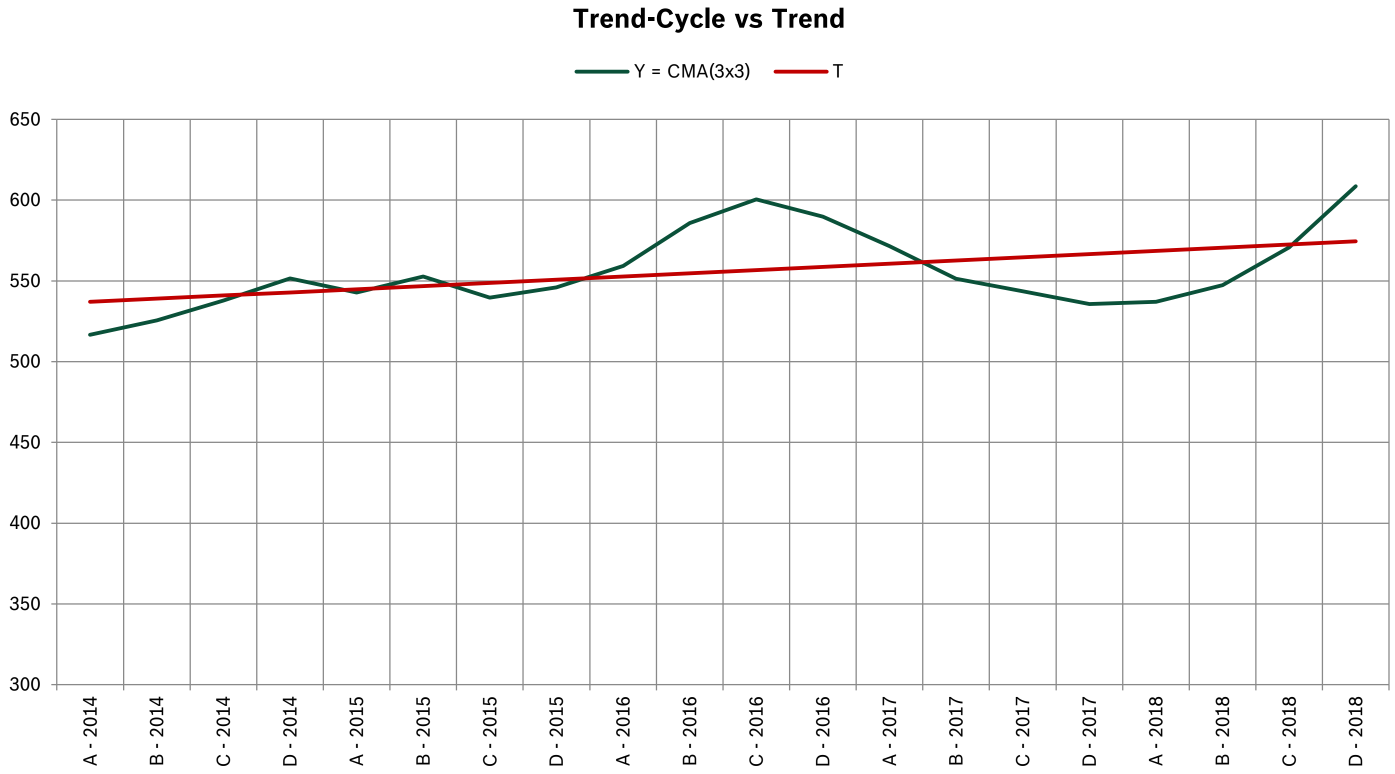
Decomposition
Step 6

Step 6: Calculate the Trend – Example

$$T = a + \beta \times t$$

$$a = 535{,}01$$

$$\beta = 1{,}974$$



Decomposition
Step 7

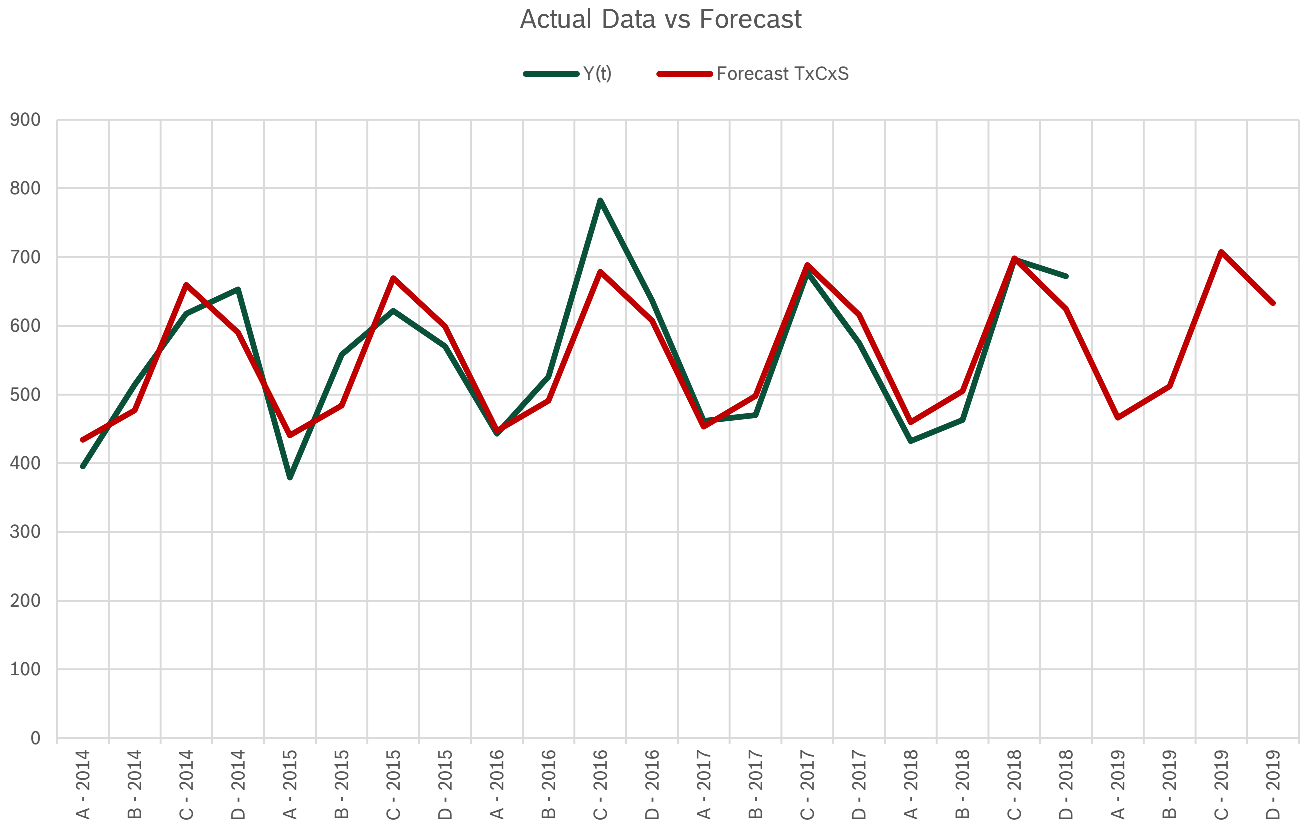
Step 7: ***Forecasting***

- Use the timeserie components for forecasting future values.
- Thus:
  - ***Extent the TxC into the future***, depending on the trend type
  - ***Multiply*** (or add) the ***seasonal*** component.

Decomposition
Step 7

Step 7: Forecasting – Example

X		Y(t)	Forecast TxC	S	Forecast TxCxS
1	A - 2014	395,5	537,0	0,809	434,2
2	B - 2014	514,4	539,0	0,885	477,0
3	C - 2014	617,5	540,9	1,219	659,5
4	D - 2014	652,8	542,9	1,087	590,2
5	A - 2015	379,4	544,9	0,809	440,6
6	B - 2015	558,1	546,9	0,885	484,0
7	C - 2015	621,7	548,8	1,219	669,2
8	D - 2015	570,3	550,8	1,087	598,8
9	A - 2016	442,9	552,8	0,809	447,0
10	B - 2016	526,1	554,7	0,885	490,9
11	C - 2016	782,3	556,7	1,219	678,8
12	D - 2016	636,9	558,7	1,087	607,4
13	A - 2017	461,8	560,7	0,809	453,4
14	B - 2017	469,9	562,6	0,885	497,9
15	C - 2017	677,7	564,6	1,219	688,4
16	D - 2017	575,0	566,6	1,087	616,0
17	A - 2018	432,1	568,6	0,809	459,7
18	B - 2018	462,9	570,5	0,885	504,9
19	C - 2018	696,2	572,5	1,219	698,0
20	D - 2018	672,2	574,5	1,087	624,6
21	A - 2019		576,5	0,809	466,1
22	B - 2019		578,4	0,885	511,9
23	C - 2019		580,4	1,219	707,7
24	D - 2019		582,4	1,087	633,1



Decomposition

What if no seasonality exists?

- In cases that the timeserie shows no significant seasonal behaviour, the ***steps 1-4 can be skipped***.
- We can evaluate this, by comparing:
  - The Autocorrelation of the data, with a delay period (k) equal to the number of the seasonality length (pos).
  - The autocorrelations of the data, with delay periods (k) till 1 period smaller than the seasonality length (pos).
- A timeserie has significant seasonal behaviour (with 90% confidence), only when:

$$\left|ACF_{pos}\right| > Limit$$

$$ACF_k = \frac{\sum_{i=1+k}^{n}[(Y_i - \bar{Y}) * (Y_{i-k} - \bar{Y})]}{\sum_{i=1}^{n}(Y_i - \bar{Y})}$$

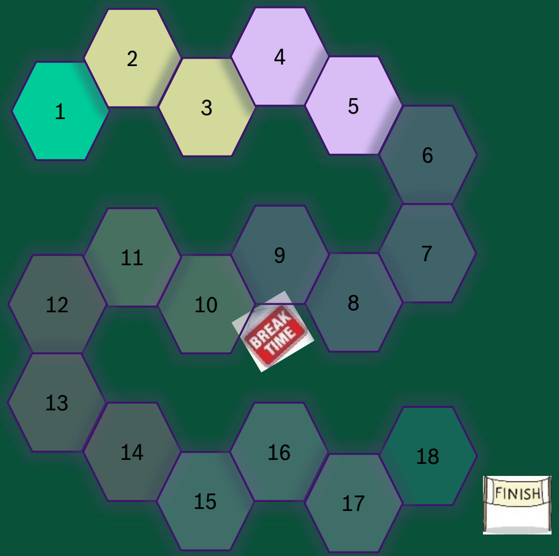
$$Limit = 1{,}645 \times \sqrt{\frac{1 + 2 * (ACF_1 + \sum_{i=2}^{pos-1} ACF_i^2)}{n}}$$

Decomposition
Practical Exercise

- Exercise steps:
  - Open file “**Forecasting Example - 15c - Exercise Decomposition (Quarterly).xlsx**”
  - Fill in your own Quarterly data
    - Random input
    - A real example
  - Observe the:
    - Calculated seasonal indexes
    - Cleaned values graph
  - All calculations are ready!

- Duration:
  - **5 minutes**
- Your Goal:
  - Get familiar with seasonal indexes
  - Observe how data are cleaned
- Who wants to present his exercise results?

5. THETA METHOD (CLASSICAL)



NTUA Greece - Forecasting & Strategy Unit (FSU)

Winner of the M3 Forecasting Competition (1999)

Held by INSEAD (Prof. S. Makridakis)

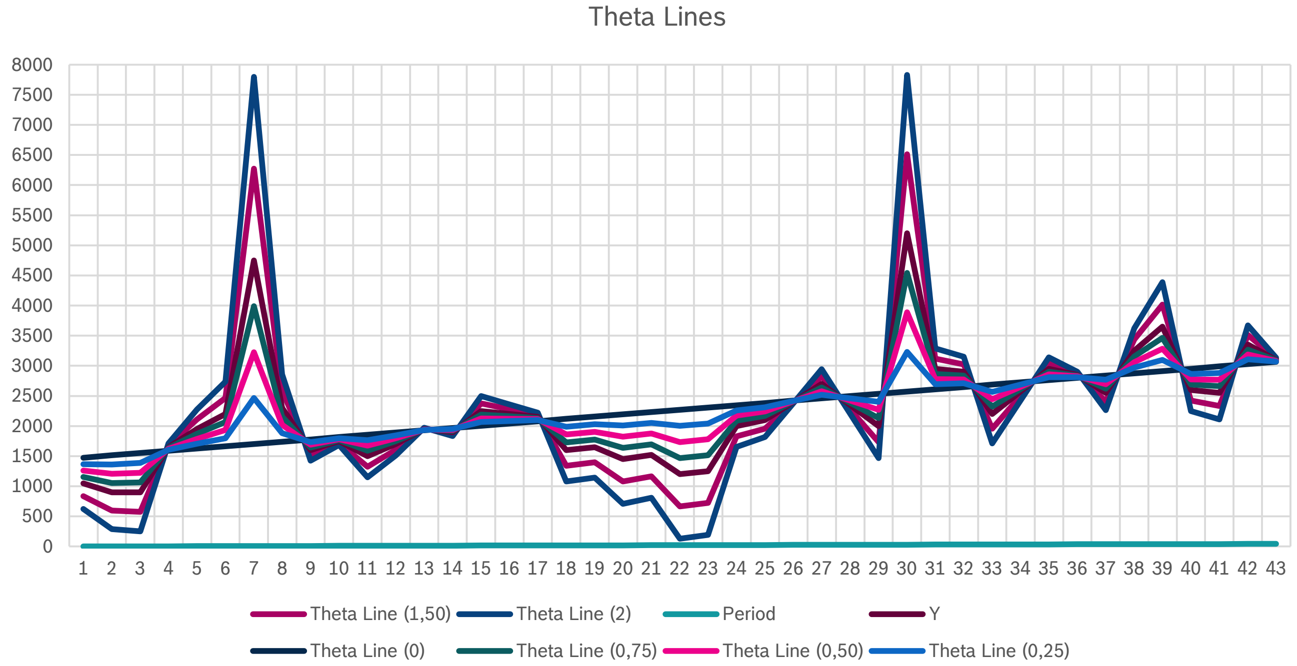
Theta Method
Model

- A new ***univariate forecasting method***, introduced in 1999 (*Assimakopoulos et. al.*), based:
  - on the decomposition model, and
  - on the concept of modifying ***the local curvatures*** of the timeseries through a coefficient "***Theta***" (the Greek letter θ).
  - The initial timeserie is spitted into ***2 or more*** Theta Lines, by using a θ parameter.
  - The resulting series, the "Theta-lines" maintain the mean and the slope of the original data, but not their local curvatures.
  - The ***basic characteristic*** of this method is:
    - a better approach of the long-term behaviour, or
    - the emphasis of short characteristics,

    depending on the value of the Θ parameter.

Theta Method
Model

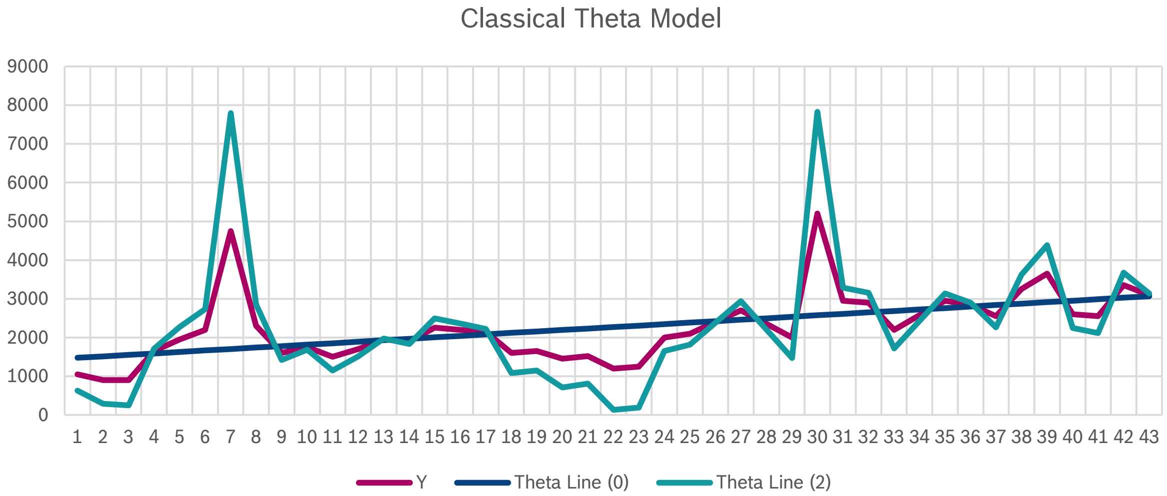
- The Theta method has introduced a different approach in the decomposition theory.
- The challenge for this method was ***to increase the degree of utilization*** of the useful information, which is hidden within the data ***before applying*** an extrapolative model.
- Acts like a magnifying lens through which the fluctuations of timeseries magnified or diminished.
- The linear combination of forecasts of the components, through this process, is more efficient.



Theta Method
Basic Model

- Simplest combination of 2 Theta lines, where:
  - Θ = 0 (straight line)
  - Θ = 2 (doubling of local curvatures)
- Each of these ***extends separately*** and their estimations are ***combined***.

➢ This combination has been used for estimating forecasts for the ***M3 forecasting competition*** (Makridakis et. al. 2000), and it has won the competition.



$$Y_t = w_0 * Y_t^0 + w_\theta * Y_t^\theta$$

$$Y_t = \frac{\theta - 1}{\theta} * Y_t^0 + \frac{1}{\theta} * Y_t^\theta$$

$$Y_t = \frac{1}{2} * Y_t^0 + \frac{1}{2} * Y_t^2$$

$$Y_t^\theta = \theta * Y_t + (1 - \theta) * Y_t^0$$

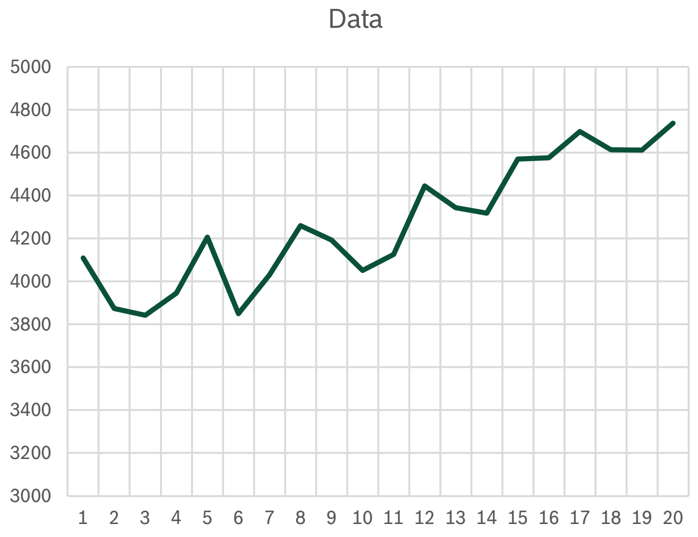
Theta Method
How to use

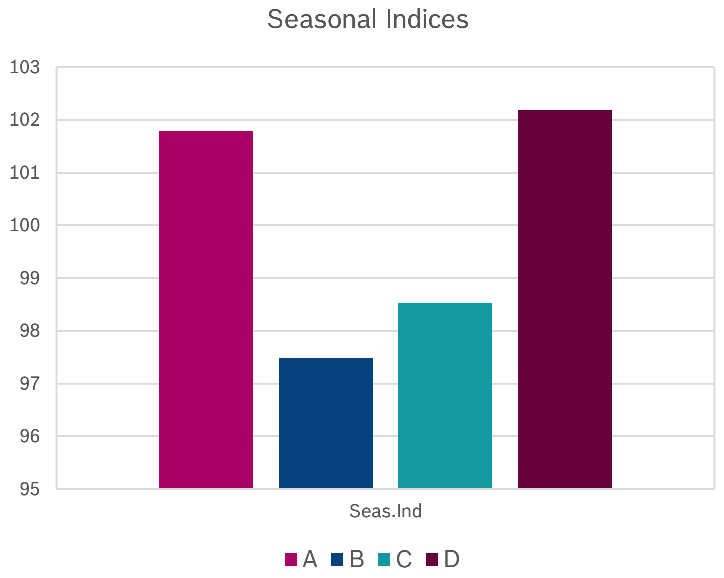
- Step 1: ***Deseasonalization***
  - Test timeserie for seasonal behaviour.
  - The timeserie is de-seasonalized, by using the classical decomposition method.
- Step 2: ***Decomposition & Theta lines***
  - The timeserie is decomposed into two Theta lines, for $\theta=0$ and $\theta=2$.
- Step 3: ***Forecast***
  - The first theta line ($\theta=0$) is extrapolated with simple linear regression (SLR) and the second line ($\theta=2$) with a selected forecasting model (example: simple exponential smoothing).
- Step 4: ***Synthesis***
  - The forecasts for the previous step are combined with equal weights, for estimating the final forecasts.
- Step 5: ***Seasonalization***
  - The final forecasts are seasonalized.

Theta Method
Example

▶ Step 1: ***Deseasonalization***

	Data	CMA(4)	SxRx100	Seas.Ind.	TxCxR
1	4.109			101,79	4.036,68
2	3.874			97,48	3.974,13
3	3.842	3.955,00	97,14	98,54	3.899,03
4	3.946	3.964,25	99,54	102,19	3.861,43
5	4.207	3.984,75	105,58	101,79	4.132,95
6	3.850	4.047,50	95,12	97,48	3.949,51
7	4.030	4.085,00	98,65	98,54	4.089,82
8	4.260	4.108,38	103,69	102,19	4.168,71
9	4.193	4.145,50	101,15	101,79	4.119,20
10	4.051	4.180,63	96,90	97,48	4.155,70
11	4.126	4.222,63	97,71	98,54	4.187,24
12	4.445	4.275,00	103,98	102,19	4.349,74
13	4.344	4.364,13	99,54	101,79	4.267,54
14	4.319	4.436,13	97,36	97,48	4.430,63
15	4.571	4.496,88	101,65	98,54	4.638,85
16	4.576	4.578,13	99,95	102,19	4.477,93
17	4.699	4.620,25	101,70	101,79	4.616,30
18	4.614	4.645,75	99,32	97,48	4.733,25
19	4.613			98,54	4.681,47
20	4.738			102,19	4.636,46





						min	max		average w/o min max	Seas.Ind
A		105,58	101,15	99,54	101,70	99,54	105,58		101,43	**101,79**
B		95,12	96,90	97,36	99,32	95,12	99,32		97,13	**97,48**
C	97,14	98,65	97,71	101,65		97,14	101,65		98,18	**98,54**
D	99,54	103,69	103,98	99,95		99,54	103,98		101,82	**102,19**
								sum	**398,555**	**400**
								Cor	**99,639%**	

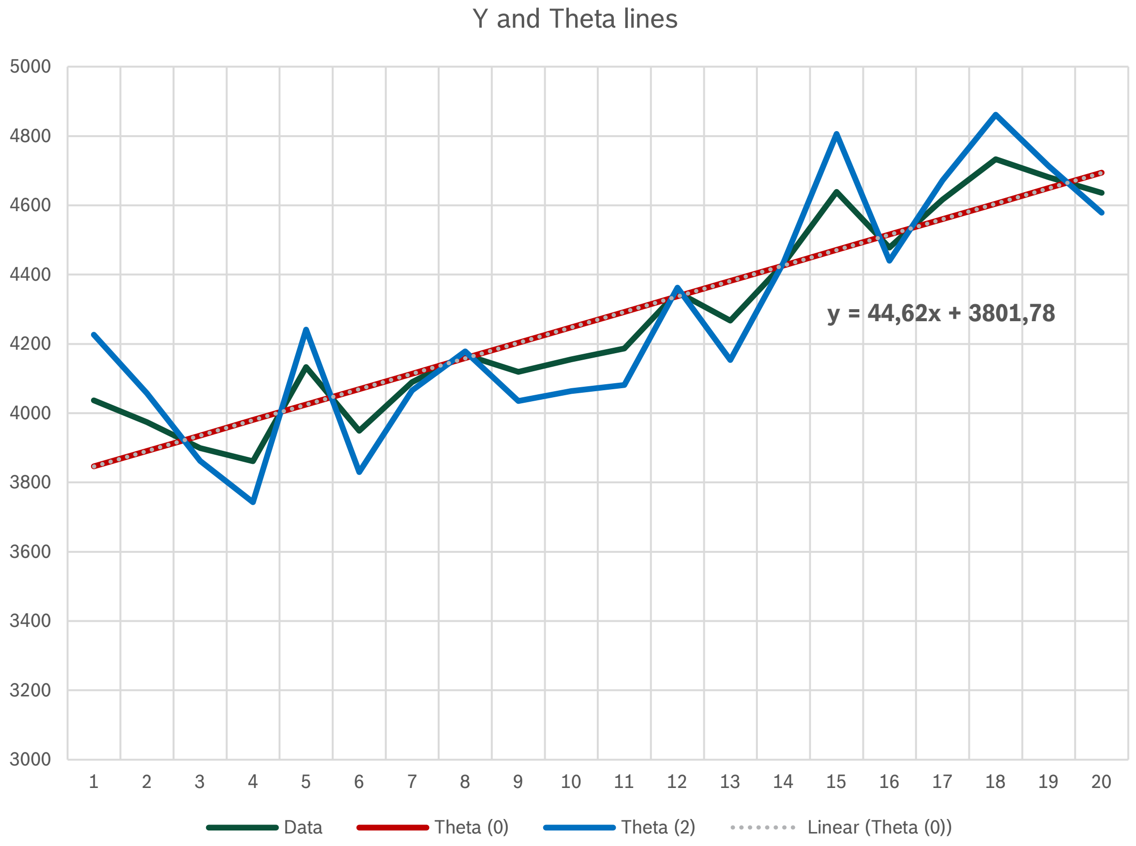
Theta Method
Example

$$Y_t^0 = 3801{,}78 + 44{,}62 * t$$

$$Y_t^2 = 2 * Y_t + (1-2) * Y_t^0 = 2 * Y_t - 1 * Y_t^0$$

▶ Step 2: ***Decomposition & Theta lines***

		Nominator			Denominator		
		A	B			SLR	
Period	Y_t	$X_t - \bar{X}$	$Y_t - \bar{Y}$	A x B	$(X_t - \bar{X})^2$	Theta (0)	Theta (2)
1	4036,68	-9,5	-233,65	2219,66	90,25	3846,41	4226,95
2	3974,13	-8,5	-296,20	2517,69	72,25	3891,03	4057,23
3	3899,03	-7,5	-371,30	2784,74	56,25	3935,65	3862,41
4	3861,43	-6,5	-408,90	2657,84	42,25	3980,28	3742,58
5	4132,95	-5,5	-137,38	755,58	30,25	4024,90	4241,00
6	3949,51	-4,5	-320,82	1443,68	20,25	4069,52	3829,50
7	4089,82	-3,5	-180,51	631,78	12,25	4114,15	4065,49
8	4168,71	-2,5	-101,62	254,05	6,25	4158,77	4178,65
9	4119,2	-1,5	-151,13	226,69	2,25	4203,39	4035,01
10	4155,7	-0,5	-114,63	57,31	0,25	4248,02	4063,38
11	4187,24	0,5	-83,09	-41,54	0,25	4292,64	4081,84
12	4349,74	1,5	79,41	119,12	2,25	4337,26	4362,22
13	4267,54	2,5	-2,79	-6,97	6,25	4381,89	4153,19
14	4430,63	3,5	160,30	561,06	12,25	4426,51	4434,75
15	4638,85	4,5	368,52	1658,35	20,25	4471,13	4806,57
16	4477,93	5,5	207,60	1141,81	30,25	4515,76	4440,10
17	4616,3	6,5	345,97	2248,81	42,25	4560,38	4672,22
18	4733,25	7,5	462,92	3471,91	56,25	4605,00	4861,50
19	4681,47	8,5	411,14	3494,70	72,25	4649,63	4713,31
20	4636,46	9,5	366,13	3478,25	90,25	4694,25	4578,67
				sum	**sum**		
				29674,52	**665,00**		
average			**SLR**	**Slope**	**44,62**		
10,5	4270,329			**constant**	**3801,78**		

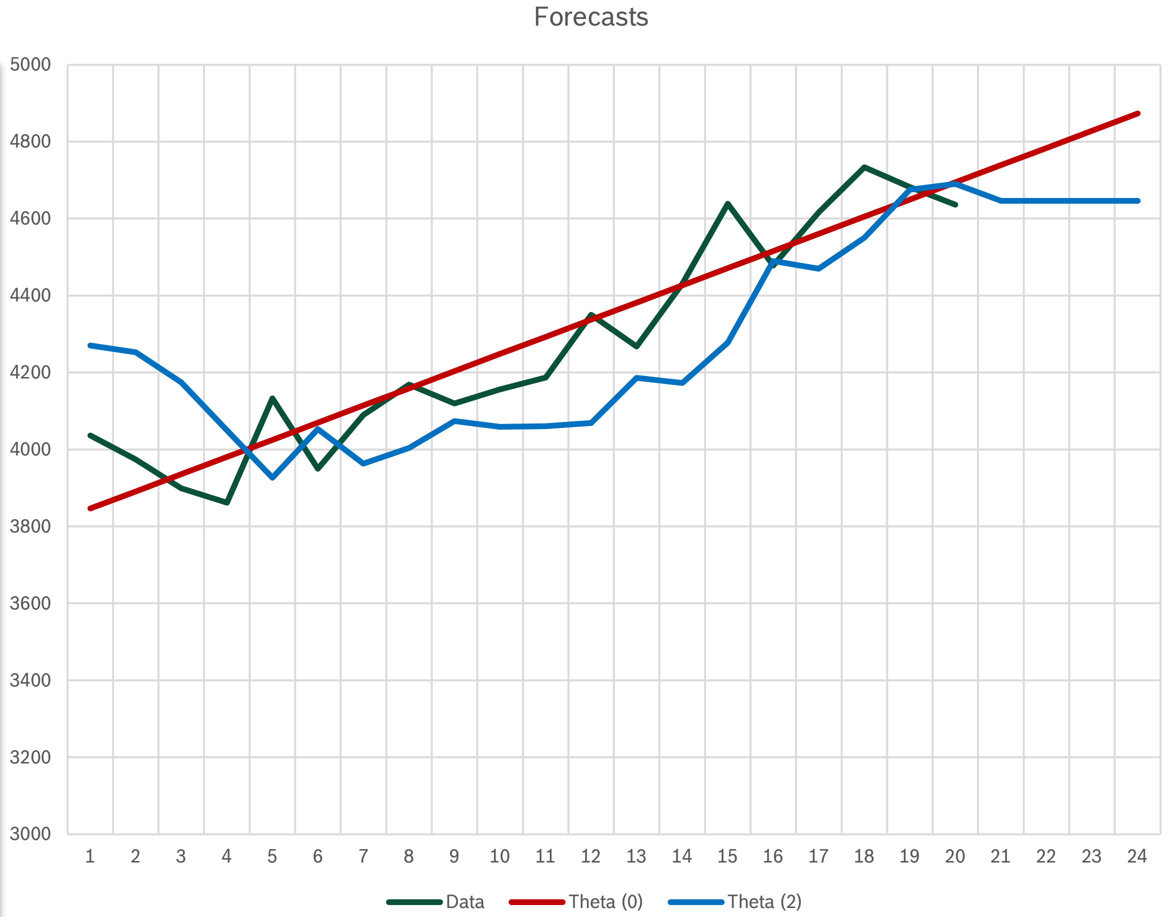


Theta Method
Example

▶ Step 3: ***Forecast***

		SLR
Period	**Data**	**Theta (0)**
1	4036,68	3846,41
2	3974,13	3891,03
3	3899,03	3935,65
4	3861,43	3980,28
5	4132,95	4024,90
6	3949,51	4069,52
7	4089,82	4114,15
8	4168,71	4158,77
9	4119,20	4203,39
10	4155,70	4248,02
11	4187,24	4292,64
12	4349,74	4337,26
13	4267,54	4381,89
14	4430,63	4426,51
15	4638,85	4471,13
16	4477,93	4515,76
17	4616,3	4560,38
18	4733,25	4605,00
19	4681,47	4649,63
20	4636,46	4694,25
21		4738,87
22		4783,50
23		4828,12
24		4872,74

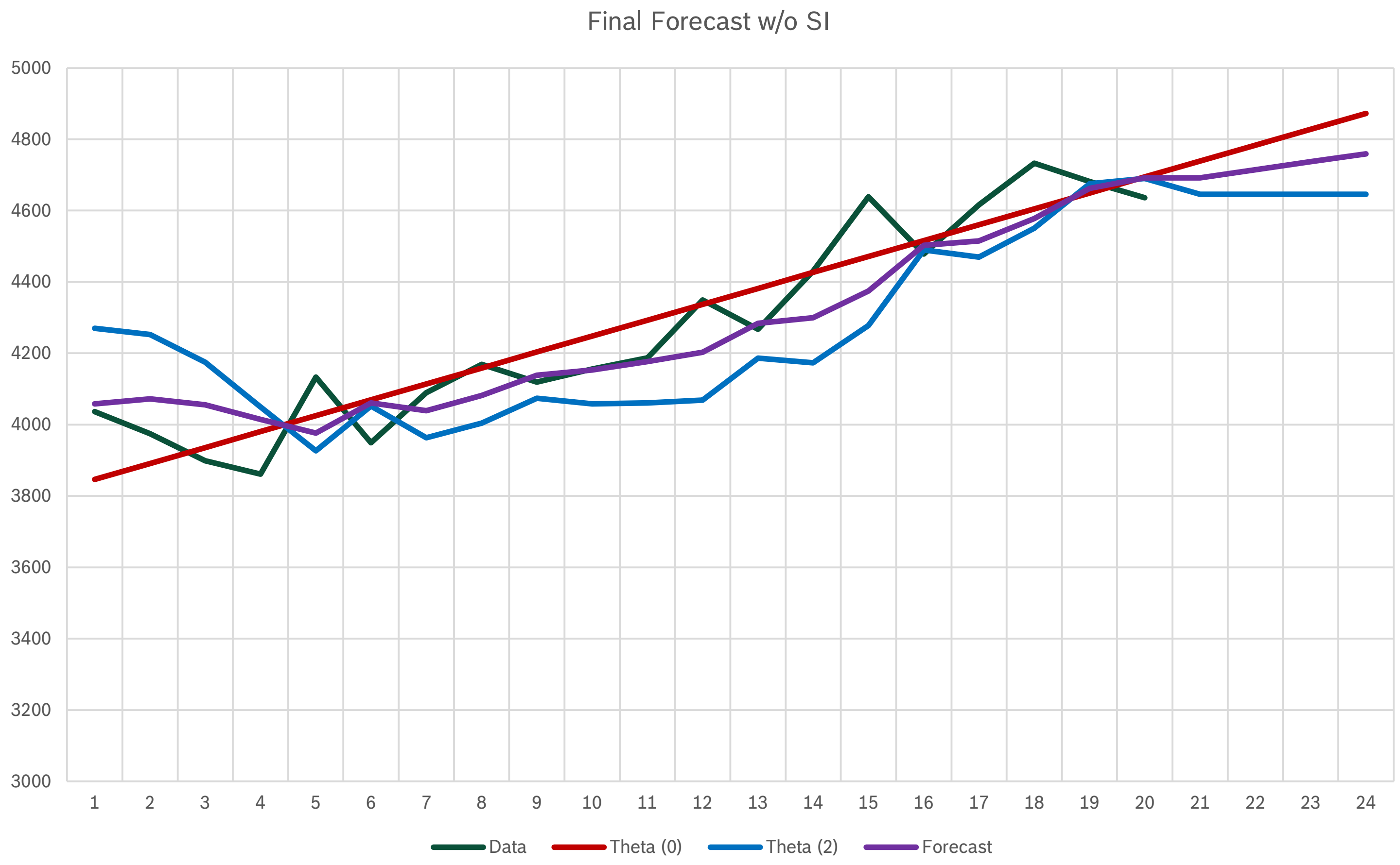
S(0) =	4270,33	α=0.4		
Period	**Theta (2)**	**Forecasts**	**Error e**	**Level S**
1	4226,95	4270,3	-43,4	4253,0
2	4057,23	4253,0	-195,7	4174,7
3	3862,41	4174,7	-312,3	4049,8
4	3742,58	4049,8	-307,2	3926,9
5	4241,00	3926,9	314,1	4052,5
6	3829,50	4052,5	-223,0	3963,3
7	4065,49	3963,3	102,2	4004,2
8	4178,65	4004,2	174,5	4074,0
9	4035,01	4074,0	-39,0	4058,4
10	4063,38	4058,4	5,0	4060,4
11	4081,84	4060,4	21,5	4069,0
12	4362,22	4069,0	293,2	4186,3
13	4153,19	4186,3	-33,1	4173,0
14	4434,75	4173,0	261,7	4277,7
15	4806,57	4277,7	528,8	4489,3
16	4440,10	4489,3	-49,2	4469,6
17	4672,22	4469,6	202,6	4550,6
18	4861,50	4550,6	310,9	4675,0
19	4713,31	4675,0	38,3	4690,3
20	4578,67	4690,3	-111,6	4645,7
21		4645,7		
22		4645,7		
23		4645,7		
24		4645,7		



Theta Method
Example

▶ Step 4: ***Synthesis***

Period	Data	Theta (0)	Theta (2)	Forecast
1	4036,68	3846,41	4270,33	4058,37
2	3974,13	3891,03	4252,98	4072,00
3	3899,03	3935,65	4174,68	4055,17
4	3861,43	3980,28	4049,77	4015,02
5	4132,95	4024,90	3926,90	3975,90
6	3949,51	4069,52	4052,54	4061,03
7	4089,82	4114,15	3963,32	4038,73
8	4168,71	4158,77	4004,19	4081,48
9	4119,2	4203,39	4073,97	4138,68
10	4155,7	4248,02	4058,39	4153,20
11	4187,24	4292,64	4060,39	4176,51
12	4349,74	4337,26	4068,97	4203,12
13	4267,54	4381,89	4186,27	4284,08
14	4430,63	4426,51	4173,04	4299,77
15	4638,85	4471,13	4277,72	4374,43
16	4477,93	4515,76	4489,26	4502,51
17	4616,3	4560,38	4469,60	4514,99
18	4733,25	4605,00	4550,65	4577,82
19	4681,47	4649,63	4674,99	4662,31
20	4636,46	4694,25	4690,32	4692,28
21		4738,87	4645,66	**4692,27**
22		4783,50	4645,66	**4714,58**
23		4828,12	4645,66	**4736,89**
24		4872,74	4645,66	**4759,20**

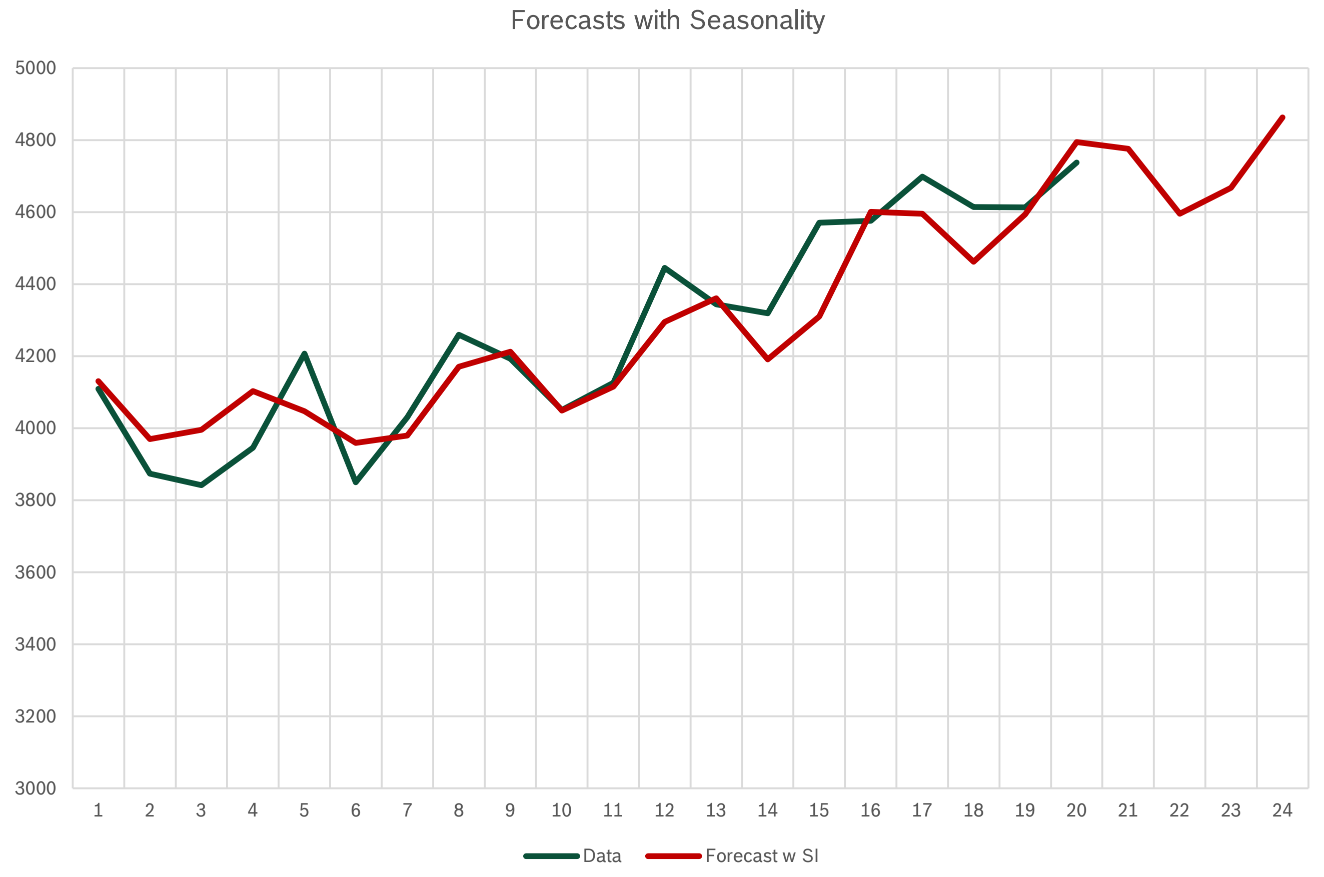


BOSCH

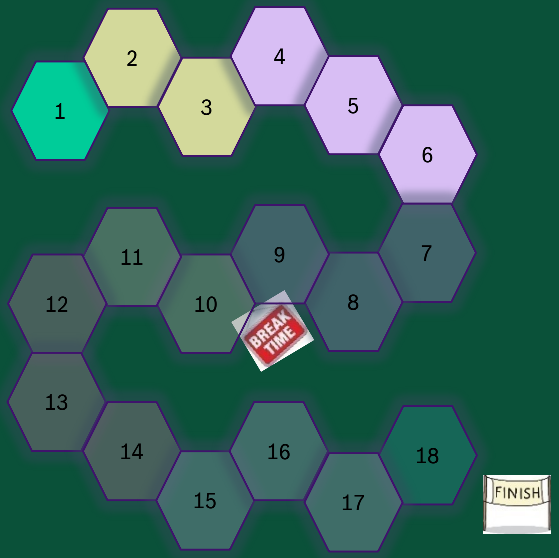
Theta Method
Example

▶ Step 5: ***Seasonalization***

Period	Data	Forecast w/o SI	Seas.Ind	Forecast w SI
1	4109	4058,37	101,79	4131,12
2	3874	4072,00	97,48	3969,48
3	3842	4055,17	98,54	3995,80
4	3946	4015,02	102,19	4102,92
5	4207	3975,90	101,79	4047,17
6	3850	4061,03	97,48	3958,78
7	4030	4038,73	98,54	3979,61
8	4260	4081,48	102,19	4170,83
9	4193	4138,68	101,79	4212,88
10	4051	4153,20	97,48	4048,63
11	4126	4176,51	98,54	4115,37
12	4445	4203,12	102,19	4295,13
13	4344	4284,08	101,79	4360,88
14	4319	4299,77	97,48	4191,51
15	4571	4374,43	98,54	4310,38
16	4576	4502,51	102,19	4601,08
17	4699	4514,99	101,79	4595,93
18	4614	4577,82	97,48	4462,56
19	4613	4662,31	98,54	4594,05
20	4738	4692,28	102,19	4795,01
21		4692,27	101,79	**4776,39**
22		4714,58	97,48	**4595,87**
23		4736,89	98,54	**4667,54**
24		4759,20	102,19	**4863,39**



6. ALTERNATIVE THETA METHODS

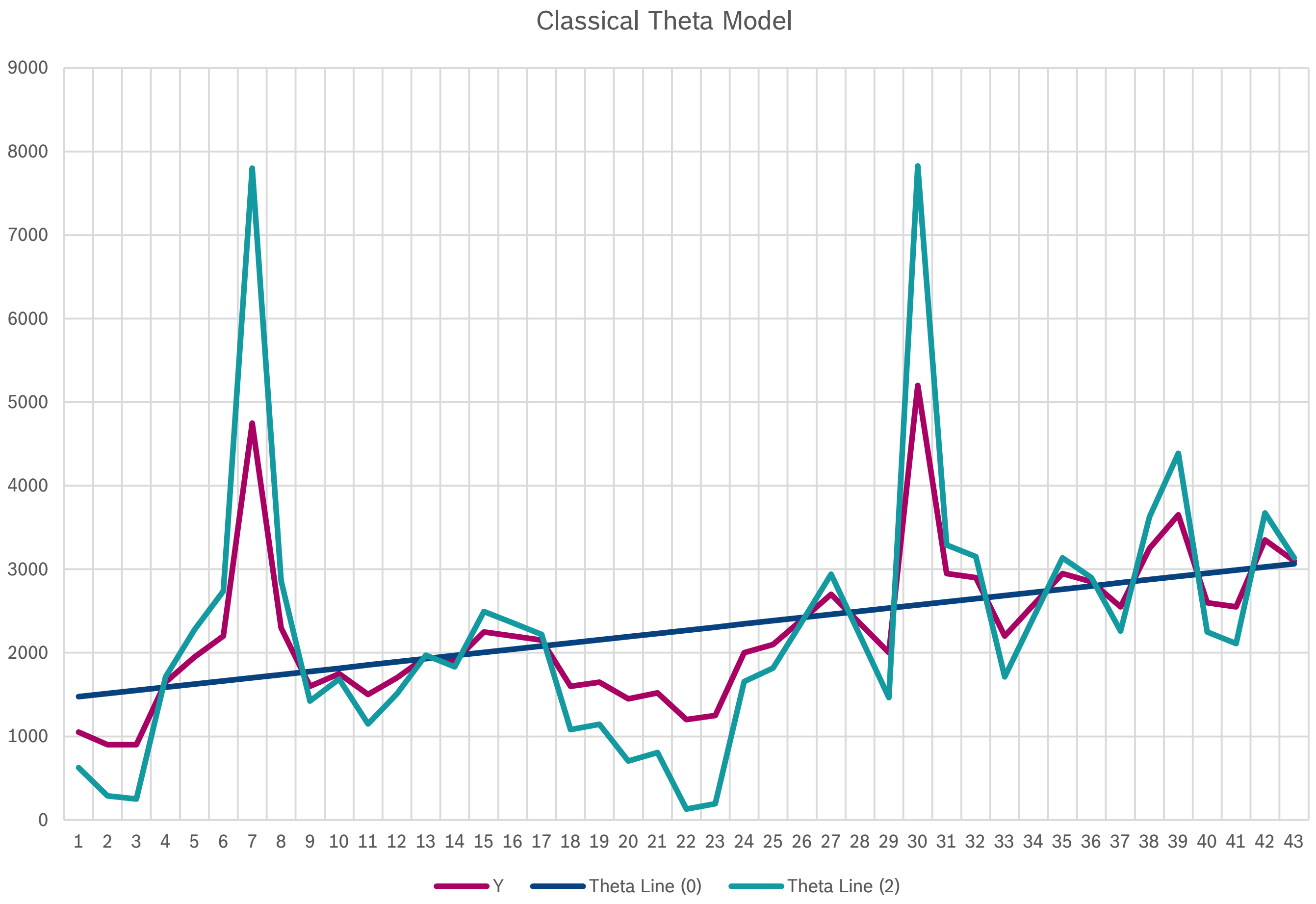


BOSCH

Alternative Theta Methods
Classical method overview

- Classical Theta method:
  - Multiplicative Seasonality
  - Additive relation of Theta lines
  - Two lines:
    - Θ = 0 (for Trend line)
    - Θ = 2 (for Level line)
  - Extrapolation:
    - Linear Regression (SLR) for trend line
    - Simple Exponential Smoothing (SES) for Level line

➢ ***Is this the limit?***



Alternative Theta Methods
Theta lines relation

- Trend / Level relation:
  - Additive relation

$$Y_t = \frac{\theta - 1}{\theta} * Y_t^0 + \frac{1}{\theta} * \mathrm{Y}_t^\theta$$

$$Y_t = \frac{1}{2} * Y_t^0 + \frac{1}{2} * \mathrm{Y}_t^2 \quad , when\ \theta = 2$$

  - Multiplicative relation

$$Y_t = \sqrt[\frac{\theta}{\theta-1}]{\mathrm{Y}_t^0} * \sqrt[\theta]{\mathrm{Y}_t^\theta}$$

$$Y_t = \sqrt[2]{Y_t^0} * \sqrt[2]{\mathrm{Y}_t^2} \quad , when\ \theta = 2$$

Alternative Theta Methods
Parameter θ selection

- Parameter θ:
  - Not always 2
  - Could be any number > 1

$$Y_t = \frac{1}{2} * Y_t^0 + \frac{1}{2} * \mathrm{Y}_t^\theta \quad , when\ \theta = 2$$

$$Y_t = \frac{0.75}{1.75} * Y_t^0 + \frac{1}{1.75} * \mathrm{Y}_t^\theta \quad , when\ \theta = 1.75$$

$$Y_t = \frac{2}{3} * Y_t^0 + \frac{1}{3} * \mathrm{Y}_t^\theta \quad , when\ \theta = 3$$

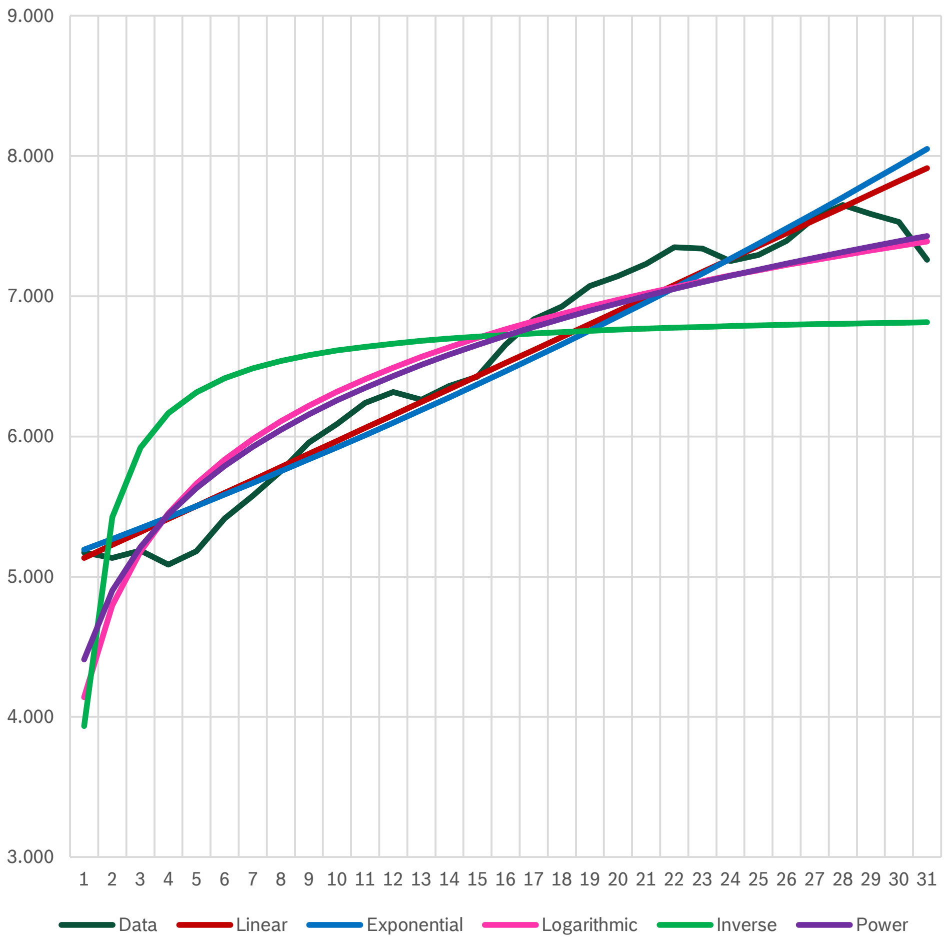
$$Y_t = \frac{0.50}{1.50} * Y_t^0 + \frac{1}{1.50} * \mathrm{Y}_t^\theta \quad , when\ \theta = 1.50$$

$$Y_t = \frac{3}{4} * Y_t^0 + \frac{1}{4} * \mathrm{Y}_t^\theta \quad , when\ \theta = 4$$

$$Y_t = \frac{0}{1.00} * Y_t^0 + \frac{1}{1.00} * \mathrm{Y}_t^\theta \quad , when\ \theta = 1.00$$

Alternative Theta Methods
Trend selection

- **Trend (Y^0):** Five different trends
  - Linear regression: $Y_t^0 = b + a \times t$
  - Exponential curve: $Y_t^0 = b \times e^{at}$, or $\log(Y_t^0) = \log b + at$
  - Logarithmic curve: $Y_t^0 = b + a \log t$
  - Inverse curve: $Y_t^0 = b + a\frac{1}{t}$
  - Power curve: $Y_t^0 = bt^a$, or $\log(Y_t^0) = \log b + a \log t$



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Alternative Theta Methods
Level extrapolation

- **Level(Y^{θ})**: Any method can be applied
  - Simple Exponential Smoothing (SES)

$$F_{t+1} = S_t \qquad S_t = S_{t-1} + (\boldsymbol{a} \times e_t) e_t = Y_t - F_t$$

  - Linear Exponential Smoothing (Holt)

$$F_{t+m} = S_t + m \times T_t$$

$$S_t = S_{t-1} + T_{t-1} + (\boldsymbol{a} \times e_t), \qquad T_t = T_{t-1} + (\boldsymbol{\beta} \times e_t) e_t = Y_t - F_t$$

  - Damped Exponential Smoothing (Damped)

$$F_{t+m} = S_t + \sum_{i=1}^{m} \varphi^i \times T_t$$

$$S_t = S_{t-1} + (\boldsymbol{\varphi} \times T_{t-1}) + (\boldsymbol{a} \times e_t), \qquad T_t = (\boldsymbol{\varphi} \times T_{t-1}) + (\boldsymbol{\beta} \times e_t) e_t = Y_t - F_t$$

Alternative Theta Methods
Considering nonlinear trends

- Generalize the method
  - Not only a classical theta method
- But, a group of methods:
  - 3 seasonality models (None, Additive, Multiplicative)
  - 2 trend / level models (Additive, Multiplicative)
  - Parameter θ
    - Greater than 1 to boost Trend, Less than 1 to boost Level
  - 5 different trends for Y^0 (Linear, Exponential, Logarithmic, Inverse, Power)
  - 3+ smoothing methods for Y^{θ} (+ Moving Averages methods)
    - Smoothing parameters can affect the performance of the model
    - Many different initialization can be selected and tested

More than 30+ alternative Theta methods

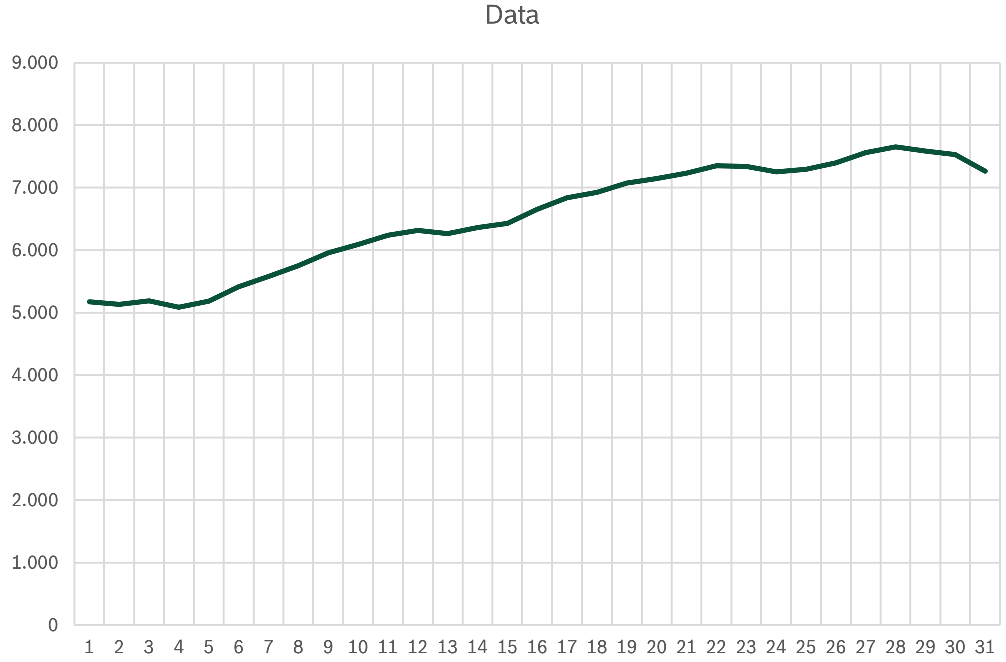
Alternative Theta Methods
Exponential Curve: Example

- Exponential curve:

$$Y_t^0 = b \times e^{at}, \text{ or } \log(Y_t^0) = \log b + at$$

$$Y_t = w_0 * Y_t^0 + w_\theta * Y_t^\theta$$

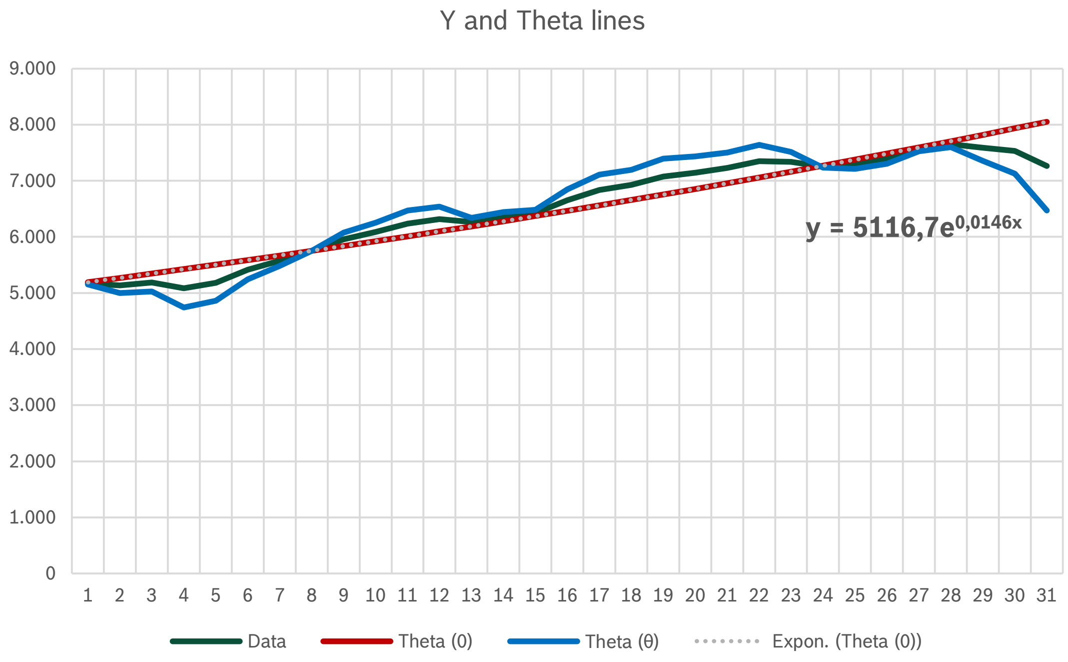
$$Y_t = \frac{\theta - 1}{\theta} * Y_t^0 + \frac{1}{\theta} * Y_t^\theta$$



Period	Data
1	5.172,10
2	5.133,50
3	5.186,90
4	5.084,60
5	5.182,00
6	5.414,30
7	5.576,20
8	5.752,90
9	5.955,20
10	6.087,80
11	6.238,90
12	6.317,20
13	6.262,70
14	6.361,00
15	6.427,40
16	6.654,90
17	6.835,40
18	6.925,50
19	7.073,50
20	7.144,00
21	7.230,60
22	7.349,60
23	7.339,20
24	7.250,80
25	7.294,60
26	7.393,90
27	7.560,90
28	7.651,40
29	7.587,30
30	7.530,50
31	7.261,10

Alternative Theta Methods
Exponential Curve

- Split into Theta lines:
  - Assuming no seasonality

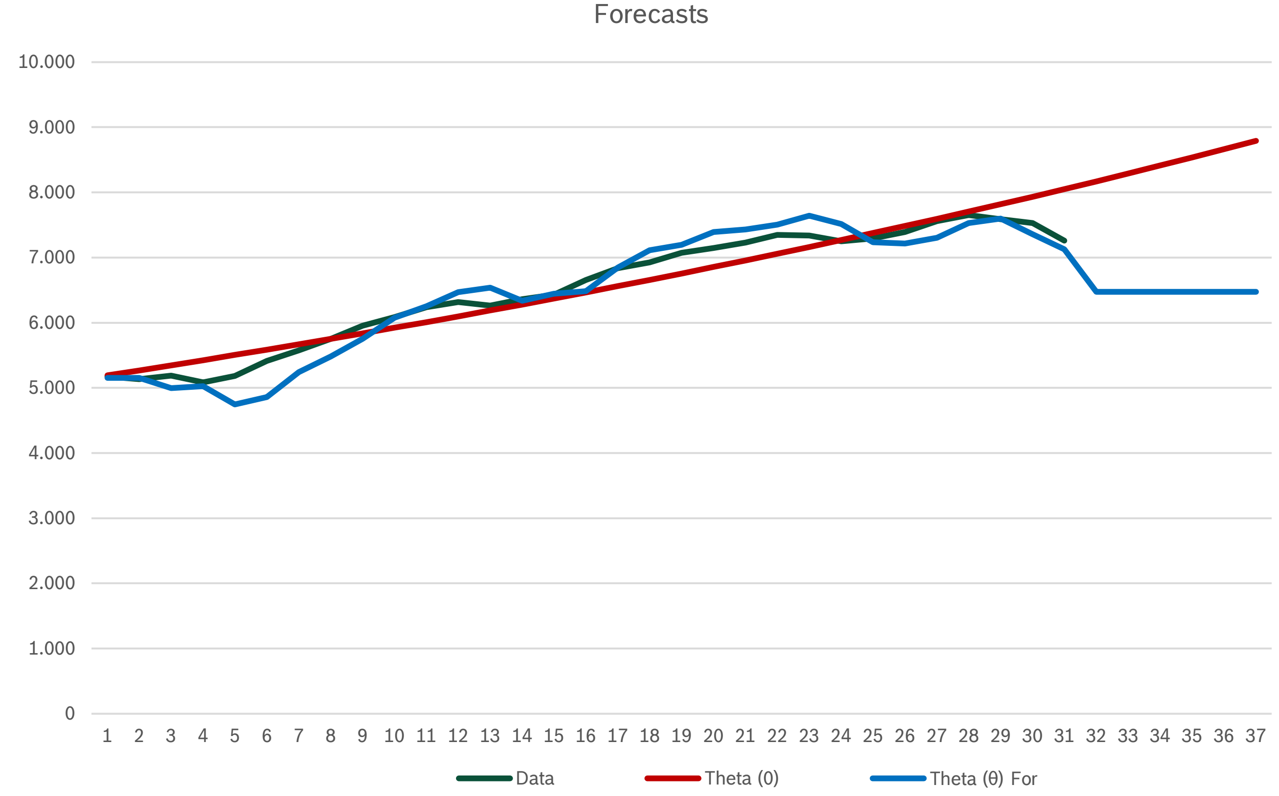


$$Y_t^0 = b \times e^{at}, \text{ or } \log(Y_t^0) = \log b + at$$

			Nominator			Denominator	THETA	2
			A	B	A x B	(X-Mean(X))^2	SLR	
Period	Data	log(Data)	X-meax(X)	Y=mean(Y)			Theta (0)	Theta (θ)
1	5.172,10	8,5510	-15	-0,2232	3,3476	225	5192,06	5152,14
2	5.133,50	8,5435	-14	-0,2307	3,2293	196	5268,53	4998,47
3	5.186,90	8,5539	-13	-0,2203	2,8641	169	5346,13	5027,67
4	5.084,60	8,5340	-12	-0,2402	2,8829	144	5424,88	4744,32
5	5.182,00	8,5529	-11	-0,2213	2,4339	121	5504,78	4859,22
6	5.414,30	8,5968	-10	-0,1774	1,7741	100	5585,86	5242,74
7	5.576,20	8,6263	-9	-0,1479	1,3315	81	5668,13	5484,27
8	5.752,90	8,6575	-8	-0,1168	0,9340	64	5751,62	5754,18
9	5.955,20	8,6920	-7	-0,0822	0,5753	49	5836,34	6074,06
10	6.087,80	8,7140	-6	-0,0602	0,3610	36	5922,30	6253,30
11	6.238,90	8,7386	-5	-0,0357	0,1783	25	6009,53	6468,27
12	6.317,20	8,7510	-4	-0,0232	0,0927	16	6098,05	6536,35
13	6.262,70	8,7424	-3	-0,0318	0,0955	9	6187,87	6337,53
14	6.361,00	8,7579	-2	-0,0163	0,0325	4	6279,01	6442,99
15	6.427,40	8,7683	-1	-0,0059	0,0059	1	6371,49	6483,31
16	6.654,90	8,8031	0	0,0289	0,0000	0	6465,34	6844,46
17	6.835,40	8,8299	1	0,0557	0,0557	1	6560,57	7110,23
18	6.925,50	8,8430	2	0,0688	0,1375	4	6657,20	7193,80
19	7.073,50	8,8641	3	0,0899	0,2697	9	6755,25	7391,75
20	7.144,00	8,8740	4	0,0998	0,3993	16	6854,75	7433,25
21	7.230,60	8,8861	5	0,1119	0,5593	25	6955,72	7505,48
22	7.349,60	8,9024	6	0,1282	0,7691	36	7058,17	7641,03
23	7.339,20	8,9010	7	0,1268	0,8874	49	7162,13	7516,27
24	7.250,80	8,8889	8	0,1147	0,9173	64	7267,62	7233,98
25	7.294,60	8,8949	9	0,1207	1,0861	81	7374,67	7214,53
26	7.393,90	8,9084	10	0,1342	1,3420	100	7483,29	7304,51
27	7.560,90	8,9307	11	0,1565	1,7219	121	7593,51	7528,29
28	7.651,40	8,9426	12	0,1684	2,0212	144	7705,36	7597,44
29	7.587,30	8,9342	13	0,1600	2,0803	169	7818,85	7355,75
30	7.530,50	8,9267	14	0,1525	2,1351	196	7934,02	7126,98
31	7.261,10	8,8903	15	0,1161	1,7411	225	8050,88	6471,32
average					sum	sum		
16	6.523,74	8,7742			36,26	2480,00		
						SLR		
						Slope	0,0146	
						constant	8,54	
							5116,69	

Alternative Theta Methods
Exponential Curve

- Extrapolate the lines:

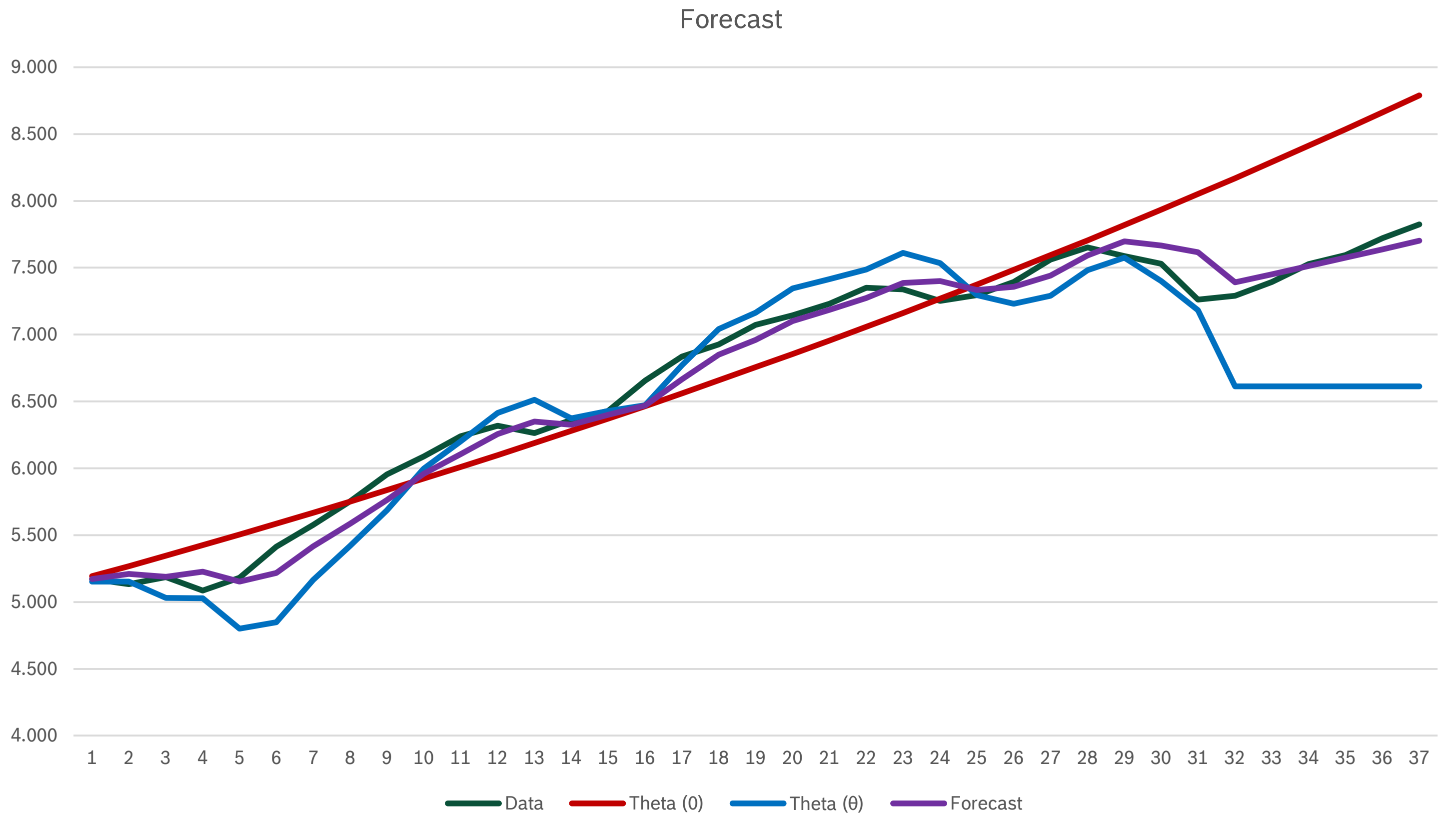


Forecast Theta(0) with SLR		Forecast Theta(θ) with SES				
		S(0) =	5.152,14	α		0,8
Period	**Theta (0)**	**Period**	**Theta (θ)**	**Theta (θ) For**	**Error e**	**Level S**
1	5192,06	1	5.152,14	5152,1	0,0	5152,1
2	5268,53	2	4.998,47	5152,1	-153,7	5029,2
3	5346,13	3	5.027,67	5029,2	-1,5	5028,0
4	5424,88	4	4.744,32	5028,0	-283,7	4801,1
5	5504,78	5	4.859,22	4801,1	58,2	4847,6
6	5585,86	6	5.242,74	4847,6	395,2	5163,7
7	5668,13	7	5.484,27	5163,7	320,6	5420,2
8	5751,62	8	5.754,18	5420,2	334,0	5687,4
9	5836,34	9	6.074,06	5687,4	386,7	5996,7
10	5922,30	10	6.253,30	5996,7	256,6	6202,0
11	6009,53	11	6.468,27	6202,0	266,3	6415,0
12	6098,05	12	6.536,35	6415,0	121,3	6512,1
13	6187,87	13	6.337,53	6512,1	-174,6	6372,4
14	6279,01	14	6.442,99	6372,4	70,5	6428,9
15	6371,49	15	6.483,31	6428,9	54,4	6472,4
16	6465,34	16	6.844,46	6472,4	372,0	6770,1
17	6560,57	17	7.110,23	6770,1	340,2	7042,2
18	6657,20	18	7.193,80	7042,2	151,6	7163,5
19	6755,25	19	7.391,75	7163,5	228,3	7346,1
20	6854,75	20	7.433,25	7346,1	87,2	7415,8
21	6955,72	21	7.505,48	7415,8	89,7	7487,5
22	7058,17	22	7.641,03	7487,5	153,5	7610,3
23	7162,13	23	7.516,27	7610,3	-94,1	7535,1
24	7267,62	24	7.233,98	7535,1	-301,1	7294,2
25	7374,67	25	7.214,53	7294,2	-79,7	7230,5
26	7483,29	26	7.304,51	7230,5	74,0	7289,7
27	7593,51	27	7.528,29	7289,7	238,6	7480,6
28	7705,36	28	7.597,44	7480,6	116,9	7574,1
29	7818,85	29	7.355,75	7574,1	-218,3	7399,4
30	7934,02	30	7.126,98	7399,4	-272,4	7181,5
31	8050,88	31	6.471,32	7181,5	-710,1	6613,4
32	8169,46	32		6613,4		6613,4
33	8289,79	33		6613,4		6613,4
34	8411,89	34		6613,4		6613,4
35	8535,79	35		6613,4		6613,4
36	8661,51	36		6613,4		6613,4
37	8789,09	37		6613,4		6613,4

Alternative Theta Methods

Exponential Curve

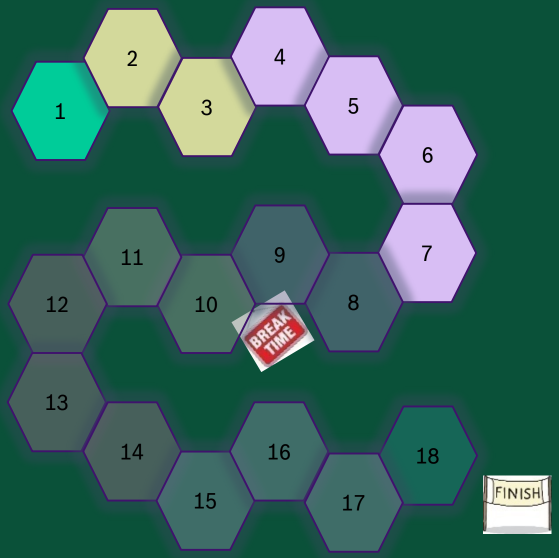
▶ Combine Forecasts & Estimate Error:



Period	Data	Theta (0)	Theta (θ)	Forecast	Error	AbsError
1	5.172,10	5192,06	5152,14	5172,10	0,00	0,000
2	5.133,50	5268,53	5152,14	5210,34	-76,84	76,837
3	5.186,90	5346,13	5029,20	5187,67	-0,77	0,768
4	5.084,60	5424,88	5027,98	5226,43	-141,83	141,825
5	5.182,00	5504,78	4801,05	5152,92	29,08	29,083
6	5.414,30	5585,86	4847,59	5216,72	197,58	197,576
7	5.576,20	5668,13	5163,71	5415,92	160,28	160,278
8	5.752,90	5751,62	5420,15	5585,89	167,01	167,012
9	5.955,20	5836,34	5687,37	5761,86	193,34	193,344
10	6.087,80	5922,30	5996,72	5959,51	128,29	128,287
11	6.238,90	6009,53	6201,98	6105,76	133,14	133,142
12	6.317,20	6098,05	6415,01	6256,53	60,67	60,671
13	6.262,70	6187,87	6512,08	6349,98	-87,28	87,275
14	6.361,00	6279,01	6372,44	6325,73	35,27	35,274
15	6.427,40	6371,49	6428,88	6400,19	27,21	27,213
16	6.654,90	6465,34	6472,42	6468,88	186,02	186,019
17	6.835,40	6560,57	6770,05	6665,31	170,09	170,089
18	6.925,50	6657,20	7042,20	6849,70	75,80	75,802
19	7.073,50	6755,25	7163,48	6959,37	114,13	114,133
20	7.144,00	6854,75	7346,09	7100,42	43,58	43,577
21	7.230,60	6955,72	7415,82	7185,77	44,83	44,833
22	7.349,60	7058,17	7487,55	7272,86	76,74	76,741
23	7.339,20	7162,13	7610,33	7386,23	-47,03	47,032
24	7.250,80	7267,62	7535,08	7401,35	-150,55	150,552
25	7.294,60	7374,67	7294,20	7334,43	-39,83	39,833
26	7.393,90	7483,29	7230,47	7356,88	37,02	37,022
27	7.560,90	7593,51	7289,70	7441,61	119,29	119,293
28	7.651,40	7705,36	7480,57	7592,96	58,44	58,436
29	7.587,30	7818,85	7574,07	7696,46	-109,16	109,159
30	7.530,50	7934,02	7399,41	7666,71	-136,21	136,214
31	7.261,10	8050,88	7181,47	7616,17	-355,07	355,073
32		8169,46	6613,35	7391,41		
33		8289,79	6613,35	7451,57		
34		8411,89	6613,35	7512,62		
35		8535,79	6613,35	7574,57		
36		8661,51	6613,35	7637,43		
37		8789,09	6613,35	7701,22		

BOSCH

7. AVERAGING



"Can anything beat the simple average?"

Michael Clements, Chief editor IJF

BOSCH

Averaging
Definition

- Averaging method: ***Combination of two or more*** simple statistical methods, in equal or unequal weights.
- The choice of methods to be involved (and their weights) are determined by:
  - the specific characteristics of each method.
  - the characteristics of the timeseries.
  - the ***forecasting horizon*** (plays a significant factor).

Averaging
Definition

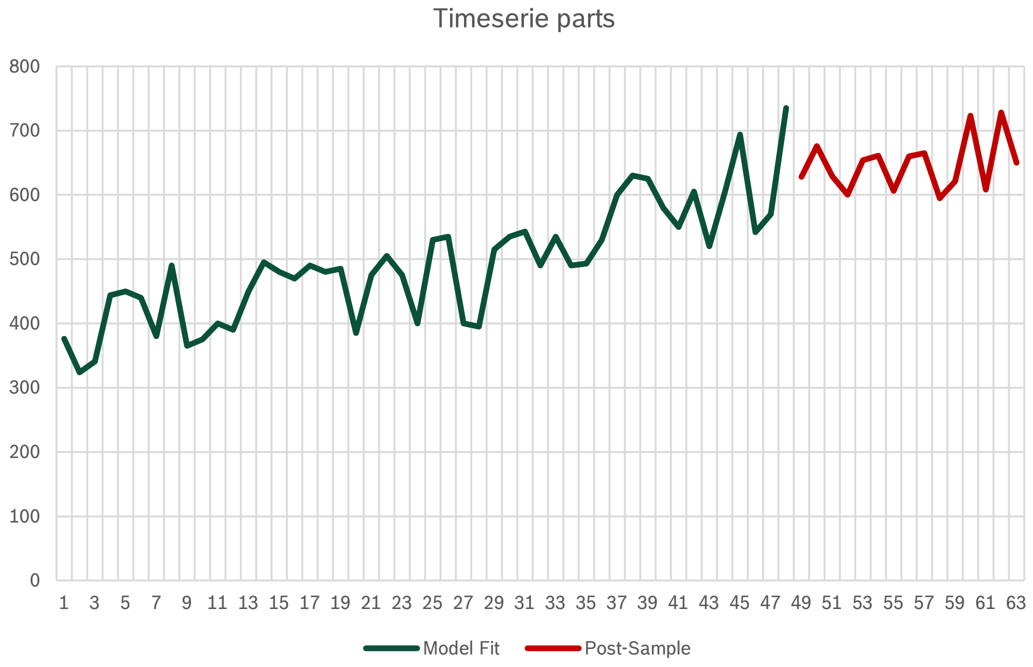
- If method X has smaller forecasting error E_x that the errors E_y of the method Y (in terms of MAPE or MSE), the combination of methods X and Y will ***not have as error the average*** of E_x and E_y.
- In general the combination of methods:
  - can lead to ***better forecasts*** and smaller forecasting errors.
  - allows us to be ***more accurate*** in our predictions.
  - ***reduces our uncertainty*** for the future, when we are not sure if the data will keep the same characteristics in the future.

- The benefits accrue because they allow the forecast to draw on a ***wider range of information***.
  - If X tends to be too high and Y to low, the combined forecast will tend to ***cancel out these biases***.

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

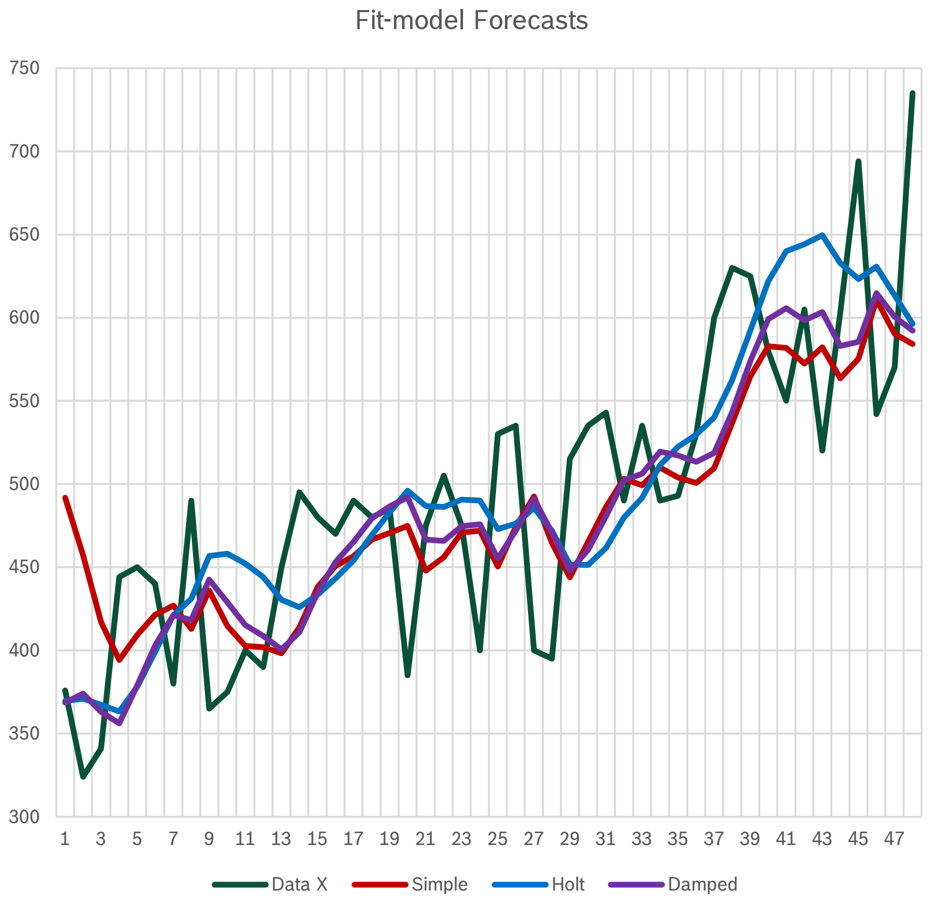
Averaging
Example

- We will estimate forecasts for a timeserie with:
  - 48 data points in model-fit, 15 data points in post-sample
  - With ***3 smoothing methods***:
    - Simple (SES)
    - Holt
    - Damped
  - With ***2 different set of weights***:
    - Simple weights, thus same weights for all 3 methods.
    - Advanced weights, thus the weights will be calculated with an inverse function of MAPE.



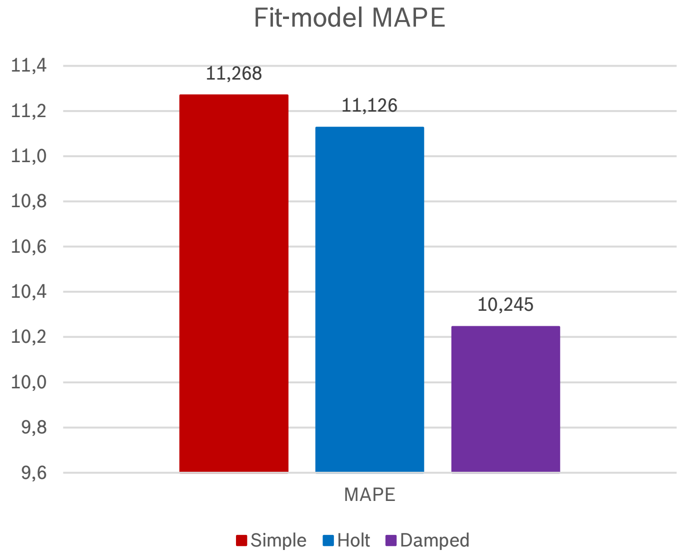
Averaging
Example

Time (t)	Data X	Simple	Holt	Damped
1	376	492	369,6	368,6
2	324	457	370,9	374,0
3	341	417	367,4	363,1
4	444	394	363,3	356,2
5	450	409	378,0	378,8
6	440	421	399,1	402,9
7	380	427	421,2	421,1
8	490	413	431,0	418,3
9	365	436	456,7	442,6
10	375	415	458,2	428,9
11	400	403	452,2	415,2
12	390	402	444,1	408,7
13	450	398	430,4	400,6
14	495	414	426,0	411,0
15	480	438	433,5	434,9
16	470	451	443,4	453,2
17	490	457	454,0	465,3
18	480	467	469,2	479,2
19	485	471	483,0	486,7
20	385	475	496,1	492,0
21	475	448	486,8	466,6
22	505	456	486,2	465,7
23	475	471	490,6	474,7
24	400	472	490,0	475,8
25	530	450	472,9	455,3
26	535	474	476,2	472,1
27	400	493	485,6	491,1
28	395	465	472,0	470,8
29	515	444	451,5	447,9
30	535	465	451,4	460,5
31	543	486	461,6	480,7
32	490	503	479,8	502,3
33	535	499	491,8	506,2
34	490	510	511,4	519,4
35	493	504	522,4	517,2
36	530	501	529,7	513,4
37	600	509	540,1	518,8
38	630	537	562,4	543,2
39	625	565	592,2	573,9
40	580	583	621,9	599,1
41	550	582	639,9	605,7
42	605	572	644,1	598,4
43	520	582	649,5	603,4
44	603	564	632,9	582,9
45	694	575	623,2	585,6
46	542	611	630,7	614,7
47	570	590	613,4	600,4
48	735	584	596,2	592,2



▶ Best method: ***Damped***

	Simple	Holt	Damped
MAPE	**11,268**	**11,126**	**10,245**

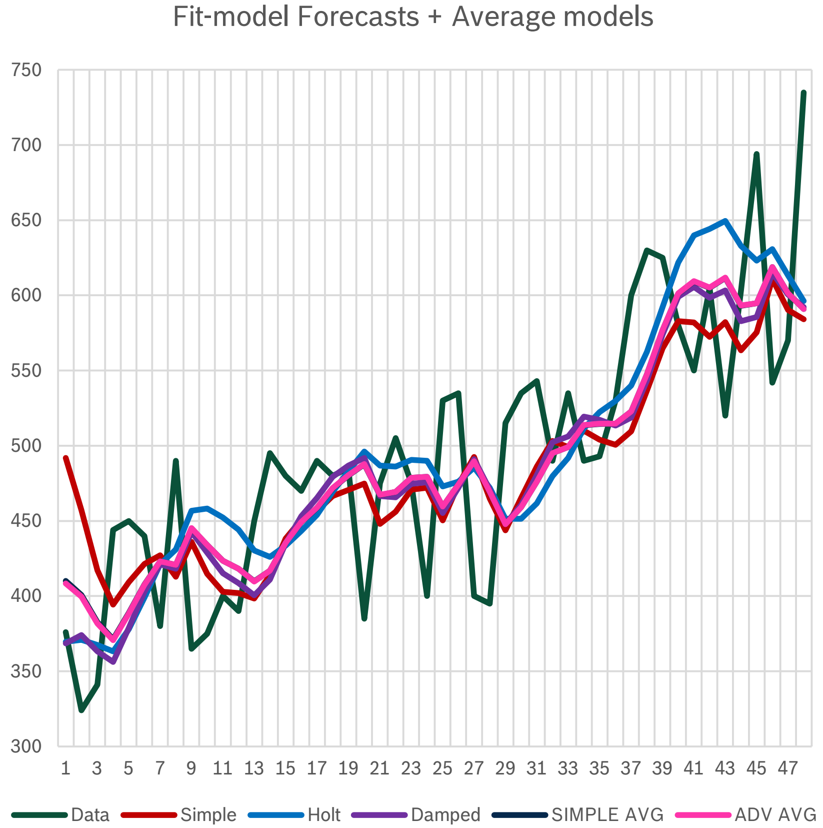


Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

BOSCH

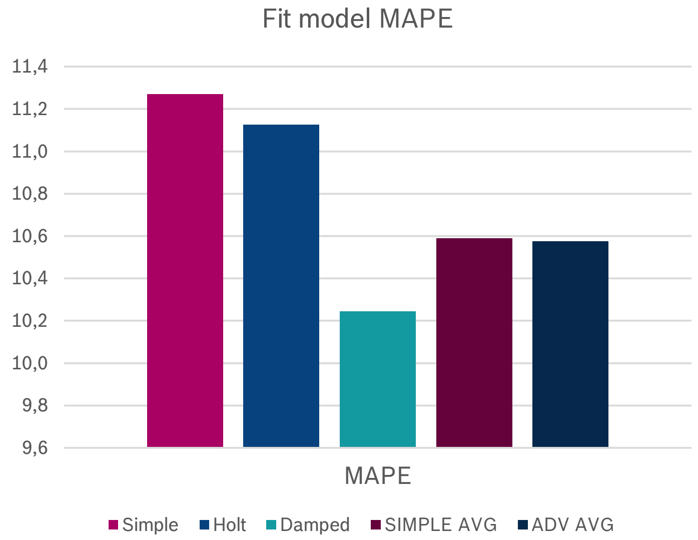
Averaging
Example

Time (t)	Data X	Simple	Holt	Damped	SIMPLE AVG	ADV AVG
1	376	491,875	369,6125	368,572	410	408,4
2	324	457,1125	370,89	374,02	401	399,6
3	341	417,1788	367,4	363,12	383	381,8
4	444	394,3251	363,3	356,22	371	370,8
5	450	409,2276	378	378,84	389	388,3
6	440	421,4593	399,1	402,86	408	407,6
7	380	427,0215	421,2	421,1	423	423,0
8	490	412,9151	431	418,26	421	420,8
9	365	436,0405	456,7	442,6	445	445,2
10	375	414,7284	458,2	428,86	434	434,0
11	400	402,8099	452,2	415,22	423	423,5
12	390	401,9669	444,1	408,68	418	418,2
13	450	398,3768	430,4	400,58	410	409,7
14	495	413,8638	426	410,98	417	416,9
15	480	438,2046	433,5	434,92	436	435,5
16	470	450,7433	443,4	453,18	449	449,2
17	490	456,5203	454	465,3	459	458,8
18	480	466,5642	469,2	479,24	472	471,9
19	485	470,5949	483	486,68	480	480,3
20	385	474,9165	496,1	491,98	488	487,9
21	475	447,9415	486,8	466,6	467	467,3
22	505	456,0591	486,2	465,74	469	469,4
23	475	470,7413	490,6	474,72	479	478,7
24	400	472,0189	490	475,76	479	479,3
25	530	450,4133	472,9	455,32	460	459,6
26	535	474,2893	476,2	472,06	474	474,1
27	400	492,5025	485,6	491,08	490	489,7
28	395	464,7517	472	470,82	469	469,3
29	515	443,8262	451,5	447,94	448	447,8
30	535	465,1784	451,4	460,52	459	459,0
31	543	486,1248	461,6	480,68	476	476,1
32	490	503,1874	479,8	502,3	495	495,2
33	535	499,2312	491,8	506,2	499	499,2
34	490	509,9618	511,4	519,44	514	513,8
35	493	503,9733	522,4	517,2	515	514,7
36	530	500,6813	529,7	513,36	515	514,7
37	600	509,4769	540,1	518,78	523	522,8
38	630	536,6338	562,4	543,16	547	547,4
39	625	564,6437	592,2	573,94	577	577,0
40	580	582,7506	621,9	599,08	601	601,4
41	550	581,9254	639,9	605,7	609	609,4
42	605	572,3478	644,1	598,44	605	605,2
43	520	582,1434	649,5	603,4	612	611,8
44	603	563,5004	632,9	582,94	593	593,2
45	694	575,3503	623,2	585,56	595	594,7
46	542	610,9452	630,7	614,72	619	618,8
47	570	590,2616	613,4	600,44	601	601,5
48	735	584,1831	596,2	592,16	591	590,9



- Best method: ***Damped***
  - 2nd best: ***Advanced Average***

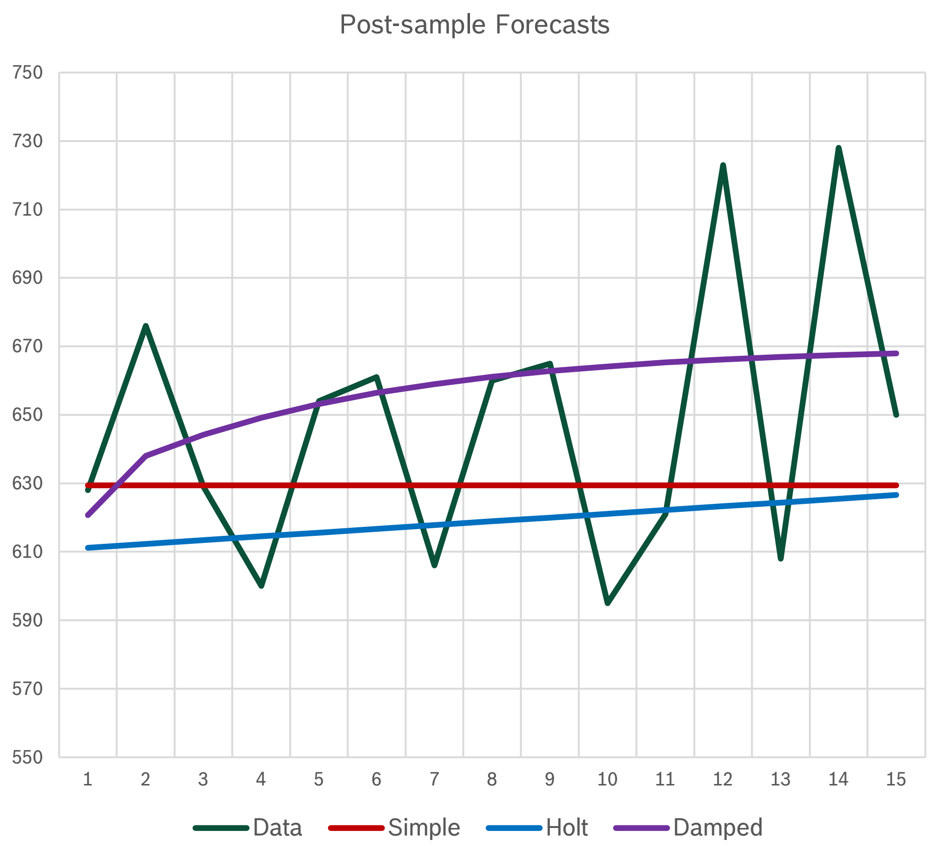
	Simple	Holt	Damped	SIMPLE AVG	ADV AVG
MAPE	11,268	11,126	10,245	10,589	10,575



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

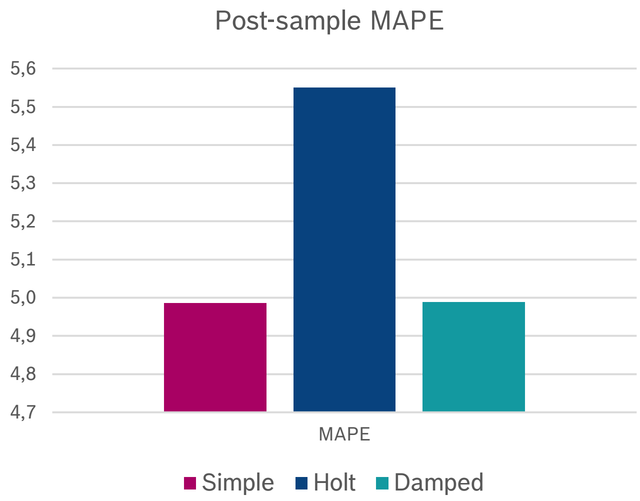
Averaging
Example

Time (t)	Data	Simple	Holt	Damped
49	628	629	611,2	630,3
50	676	629	612,3	638,0
51	629	629	613,4	644,2
52	600	629	614,5	649,2
53	654	629	615,6	653,2
54	661	629	616,7	656,4
55	606	629	617,8	659,0
56	660	629	618,9	661,1
57	665	629	620,0	662,8
58	595	629	621,1	664,2
59	621	629	622,2	665,3
60	723	629	623,3	666,2
61	608	629	624,4	666,9
62	728	629	625,5	667,5
63	650	629	626,6	668,0



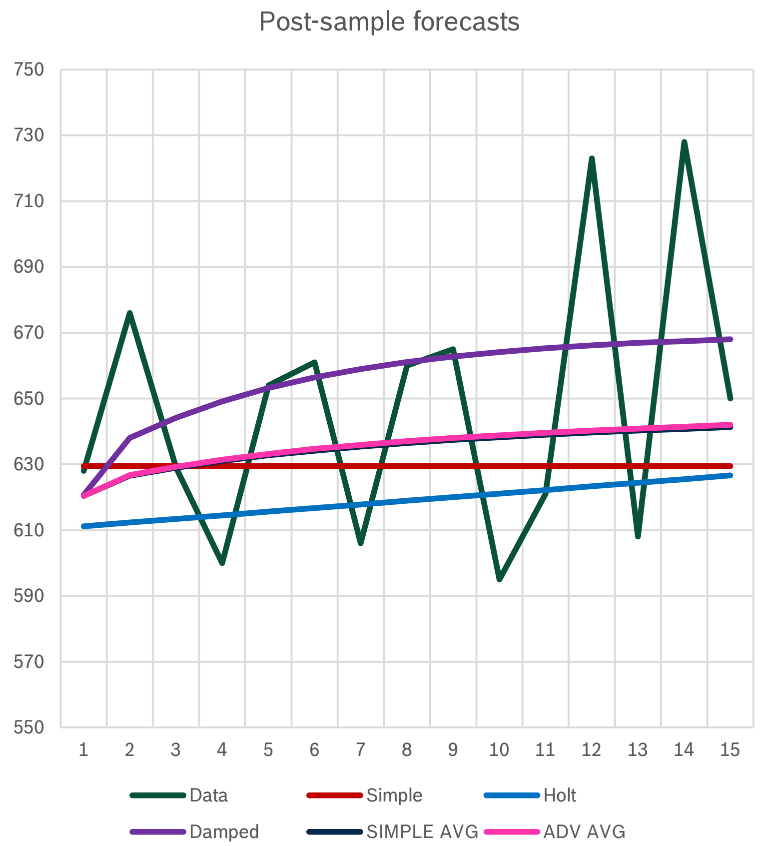
- Best method: ***Simple***

	Simple	Holt	Damped
MAPE	**4,986**	**5,550**	**4,989**



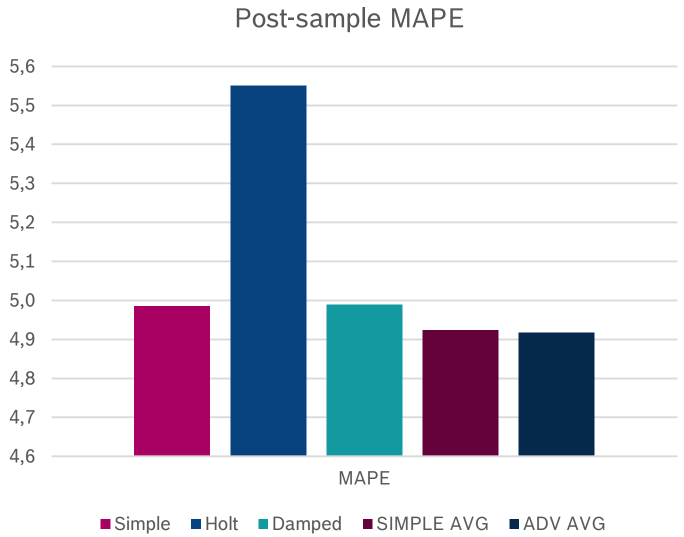
Averaging
Example

Time (t)	Data	Simple	Holt	Damped	SIMPLE AVG	ADV AVG
49	628	629,4	611,2	620,7	620,4	620,4
50	676	629,4	612,3	638,0	626,6	626,8
51	629	629,4	613,4	644,2	629,0	629,3
52	600	629,4	614,5	649,2	631,0	631,4
53	654	629,4	615,6	653,2	632,7	633,2
54	661	629,4	616,7	656,4	634,2	634,7
55	606	629,4	617,8	659,0	635,4	635,9
56	660	629,4	618,9	661,1	636,5	637,0
57	665	629,4	620,0	662,8	637,4	638,0
58	595	629,4	621,1	664,2	638,2	638,8
59	621	629,4	622,2	665,3	639,0	639,6
60	723	629,4	623,3	666,2	639,6	640,3
61	608	629,4	624,4	666,9	640,2	640,9
62	728	629,4	625,5	667,5	640,8	641,4
63	650	629,4	626,6	668,0	641,3	642,0



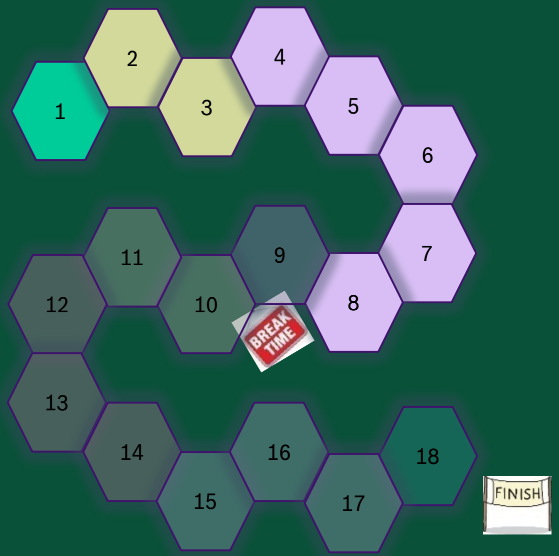
- Best method: ***Adv. Average***
  - 2nd best: ***Simple***

	Simple	Holt	Damped	SIMPLE AVG	ADV AVG
MAPE	**4,986**	**5,550**	**4,989**	**4,925**	**4,918**



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

8. ARIMA



BOSCH

ARIMA
Definition

- ARIMA: ***A**uto **R**egressive **I**ntegrated **M**oving **A**verage*
  - A stochastic model that tries to describe the ***evolution of a measure*** over time.
  - Introduced by ***Box & Jenkins*** (1971).
  - They can be applied in cases where data shows ***evidence of non-stationary***, where an initial differencing step (corresponding to the “integrated” part of the model) can be applied to remove the non-stationary.
- The model is generally referred to as an ***ARIMA (p, d, q)*** model where:
  - **p**, **d**, and **q** are non-negative integers that refer to the order of the ***autoregressive, integrated, and moving average parts*** of the model respectively.
    - If d = 0, then the data could be used without a differencing step, thus in a ARMA(p, q) model.

ARIMA
Equations

- ***AR(p): Autoregressive model***
  - A data point value depends on the previous data points ***values***.

$$Y_t = c + \varphi_1 \times \Upsilon_{t-1} + \varphi_2 \times \Upsilon_{t-2} + \cdots + \varphi_p \times \Upsilon_{t-p} + e_t,$$

where $\mathrm{c} = \bar{Y} \times \left(1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p\right)$

- ***MA(q): Moving Average model***
  - A data point value depends on the previous data points ***errors***.

$$Y_t = c - \theta_1 \times \mathrm{e}_{t-1} - \theta_2 \times \mathrm{e}_{t-2} - \cdots - -\theta_q \times \mathrm{e}_{t-q} + e_t$$

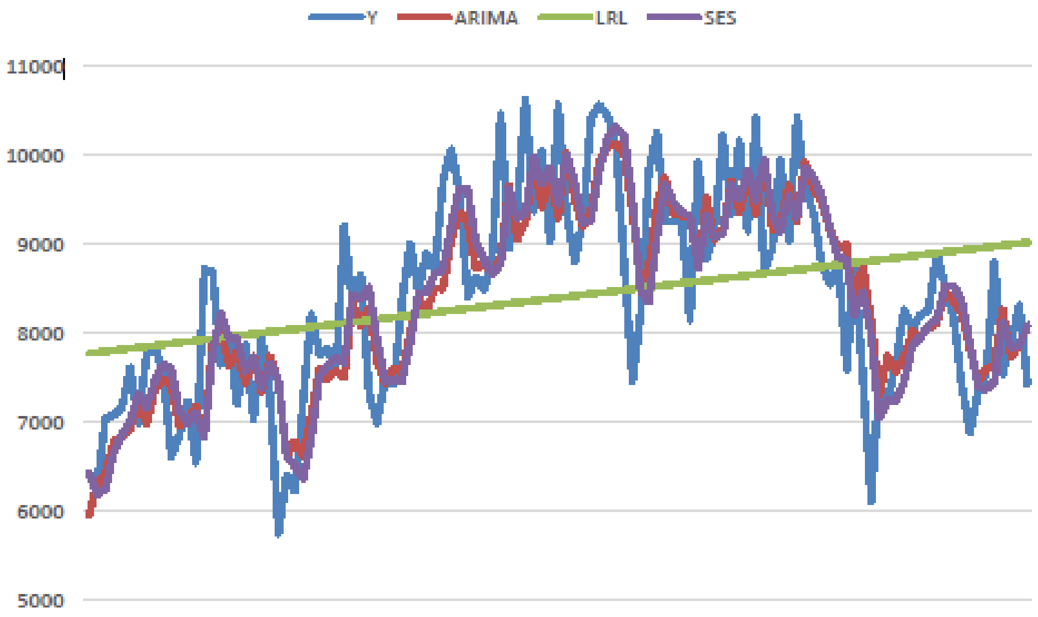
where $\mathrm{c} = \bar{Y}$

- ***ARMA(p, q)***

$$Y_t = c + \varphi_1 \times \Upsilon_{t-1} + \varphi_2 \times \Upsilon_{t-2} + \cdots + \varphi_p \times \Upsilon_{t-p} - \theta_1 \times \mathrm{e}_{t-1} - \theta_2 \times \mathrm{e}_{t-2} - \cdots - -\theta_q \times \mathrm{e}_{t-q} + e_t$$

ARIMA
Model

- They have a logic ***similar to regression (SLR) and smoothing (SES)*** methods, but they express each timeserie value as a linear sum of previous data points values.
  - ***SLR***: $\widehat{Y_{t+1}} = 7768 + 11 \times (t+1)$
  - ***SES***: $\widehat{Y_{t+1}} = \widehat{Y_t} + 0.48 \times e_t$
  - ***ARIMA***: $\widehat{Y_{t+1}} = 1.31 \times Y_{t-1} - 0.31 \times Y_{t-2} + 0.8 \times e_t$



ARIMA
Requirements & Limitations

- For using an ARIMA model:
  - The timeserie should be ***distinct***: data points in equal time frames.
  - The timeserie should be ***static***: mean value, variance, and autocorrelation must be static within the sample.
  - ***Short-term*** forecasts must be estimated.
  - ***Granularity***: Forecast depends on ***p*** historical observations which as known.
  - ***Stagnation***: ARMA models can ***not be applied to data with trend*** or to data with ***periodical fluctuations.***
  - ***Long-term*** forecasts are based in short-term forecast, therefore the forecast error increases significantly.

ARIMA
Model Selection

- How to select proper ARIMA(p, d, q) model: ***A four step approach***
  - ***Data Preparation*:** Remove seasonality and trend
    - By using integration, transformation, etc.
  - ***Recognition***: Detect potential representative models
    - by using statistical analysis, data autocorrelation, etc.
  - ***Evaluation***: Estimation of the p, d, q, parameters
    - by using expected likelihood, least square methods, etc.
  - ***Diagnostics***: Check the statistical significance of the parameters
    - by using t-test, error autocorrelation, etc.

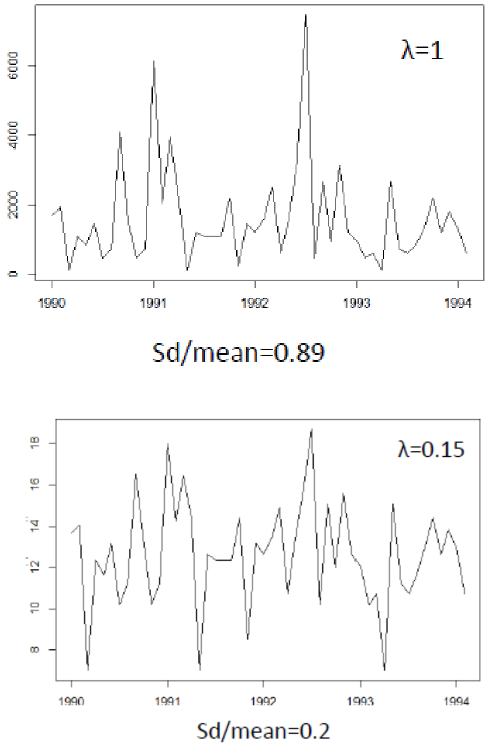
ARIMA
Data Preparation

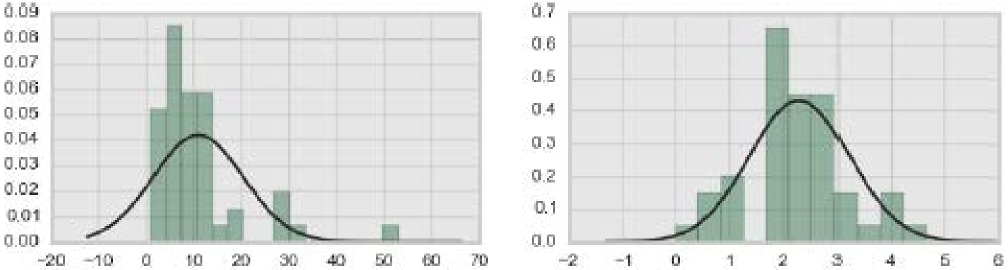
- ***Data transformations***:
  - Why? For limitation of variation and for removing abnormal values.
  - When? If timeserie has strong variance or the forecasting error is high.
- ***Data integration***:
  - Why? To remove (or to minimize) trend and seasonality.
  - When? If timeserie data have significant trend, or significant seasonality.

➢ ***We don’t need to deseasonalize*** the data (by using a decomposition method) before applying an ARIMA model. These models can handle seasonality though data transformations.

ARIMA
Data Preparation

- The ARIMA models can not interpret periodical variations.
  - forecasting accuracy is limited when the variance of the data is high (high level of noisy data).
  - In case of white noise, forecasting accuracy is not affected.
- For removing the noise, we need to ***normalize the data*** by using logarithm or other transformations
  - Example: Box-Cox transformation $x_t = \frac{{y_t}^{\lambda} - 1}{\lambda}$





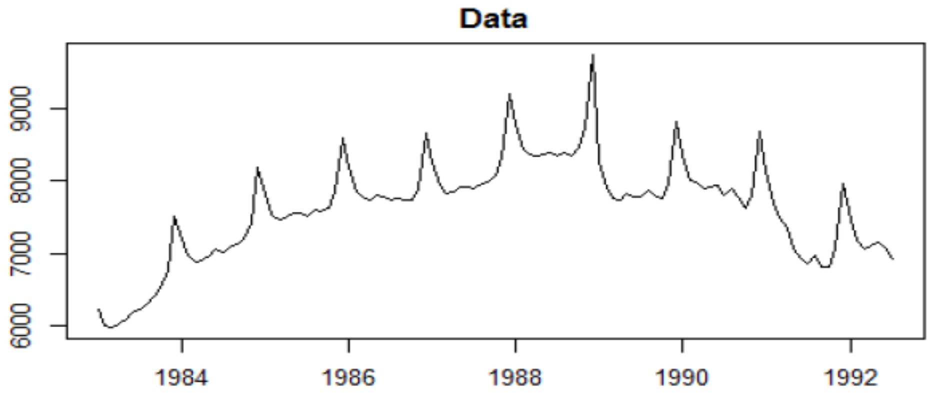
Marginal distribution: before & after applying Box-Cox transformation

ARIMA
Data Preparation

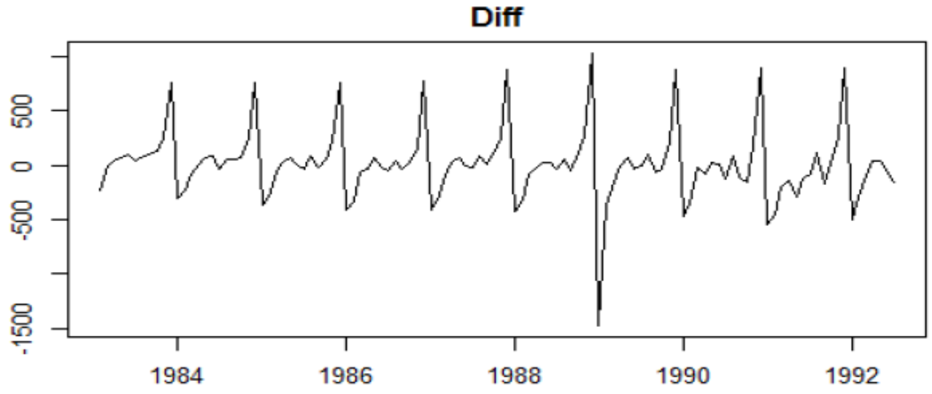
- The ***I(d)*** part of the ARIMA(p, d, q) model.
  - For removing a ***trend***, we create a new timeserie based on data point values differences
    - For a ***linear*** trend, we use ***1st level***, thus d=1: $Y_t{}' = Y_t - Y_{t-1}$
    - For a ***non-linear*** trend, we use ***2nd level***, thus d=2: $Y_t{}'' = Y_t{}' - Y_{t-1}{}' = Y_t - 2 \times Y_{t-1} - Y_{t-2}$
    - Parameter could be: $1 \leq d \leq n - 1$, where n is the number of timeseries' data points.

  - For removing strong ***seasonality***, we use seasonal differences:
    - For an m-seasonality: $Y_t{}' = Y_t - Y_{t-m}$

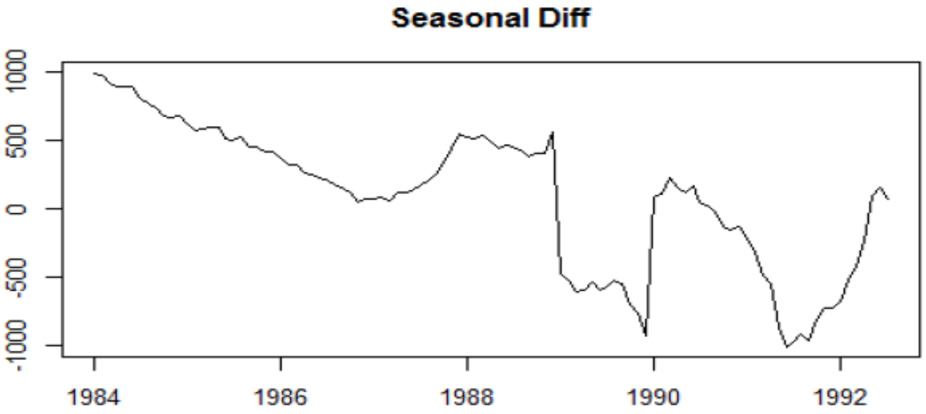
ARIMA
Data Preparation Example



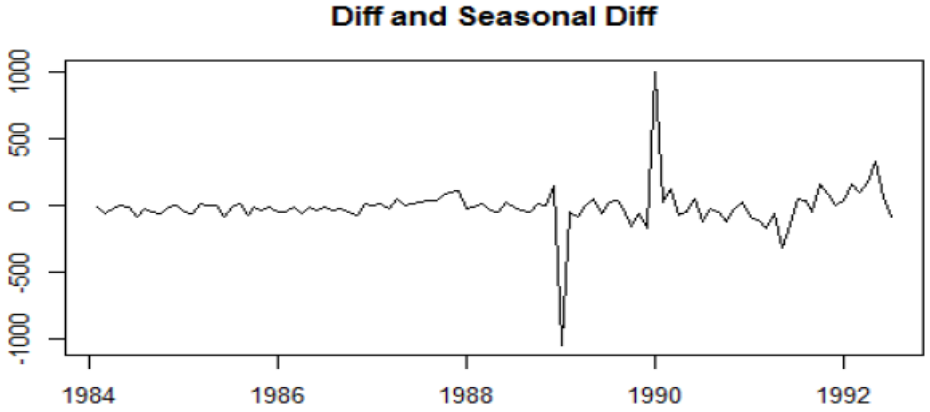
Data have significant trend and monthly seasonality



1st level integration can't remove seasonality → no stationary data



Seasonal integration can't remove trend → no stationary data



Solution: We must apply both ***1st level integration and seasonal integration***

ARIMA
Recognition

- ***ACF***: Indicates if a timeserie data point value depends from k past data points values

$$p_k = \frac{\sum_{t=k+1}^{T}(Y_t - \bar{Y}) \times (Y_{t-k} - \bar{Y})}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

- ***PACF***: Indicates if a timeserie data point value depends from a past data point values, without taking into account the intermediate data points.

$$\varphi_{11} = p_1, \varphi_{kk} = p_k - \sum_{j=1}^{k-1} \frac{\varphi_{k-1,j} \times p_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} \times p_j} \, when\ k = 2,3\ and\ \varphi_{k,j} = \varphi_{k-1,j} - \varphi_{k,k} \times \varphi_{k-1,k-j}\ when\ k = 3,4\ and\ j = 1,2, \ldots$$

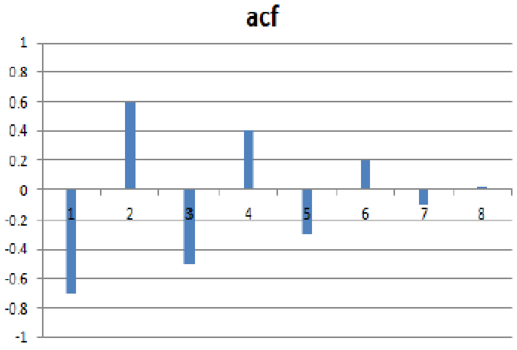
- Usually, we use them with ***k<3***. Greater values can lead to very complicated models without any gain in forecasting accuracy.
- These checks must be performed into stationary data, thus ***after applying integration***.

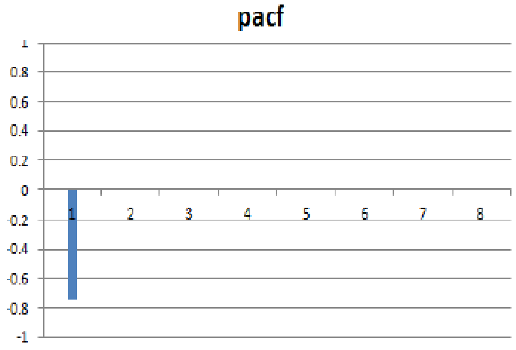
ARIMA
Recognition

$$Y_t = c + \varphi_1 \times Y_{t-1} + \varphi_2 \times Y_{t-2} + \cdots + \varphi_p \times Y_{t-p} + e_t$$

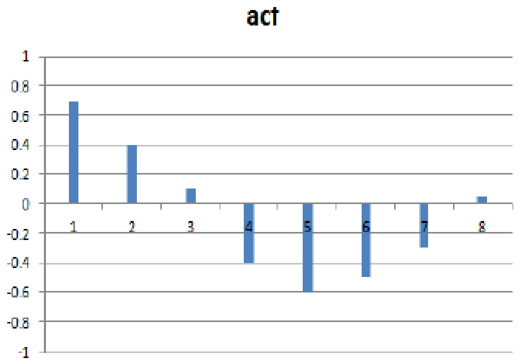
- For a stationary model ***AR(p):***
  - The ACF values must ***decrease*** (exponentially or sin) ***towards 0***.
  - The PACF values must be 0 ***immediately*** after p lag-periods.

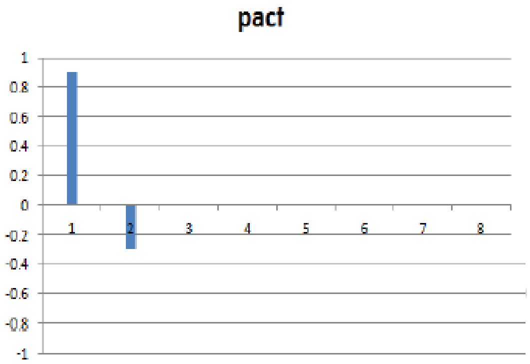
AR(1)





AR(2)





AR(1): $-1 < \varphi_1 < 1$

AR(2): $-1 < \varphi_2 < 1,\ \varphi_1 + \varphi_2 < 1, \varphi_2 - \varphi_1 < 1$

AR(p): $\varphi_1 + \varphi_2 + \cdots + \varphi_p < 1$

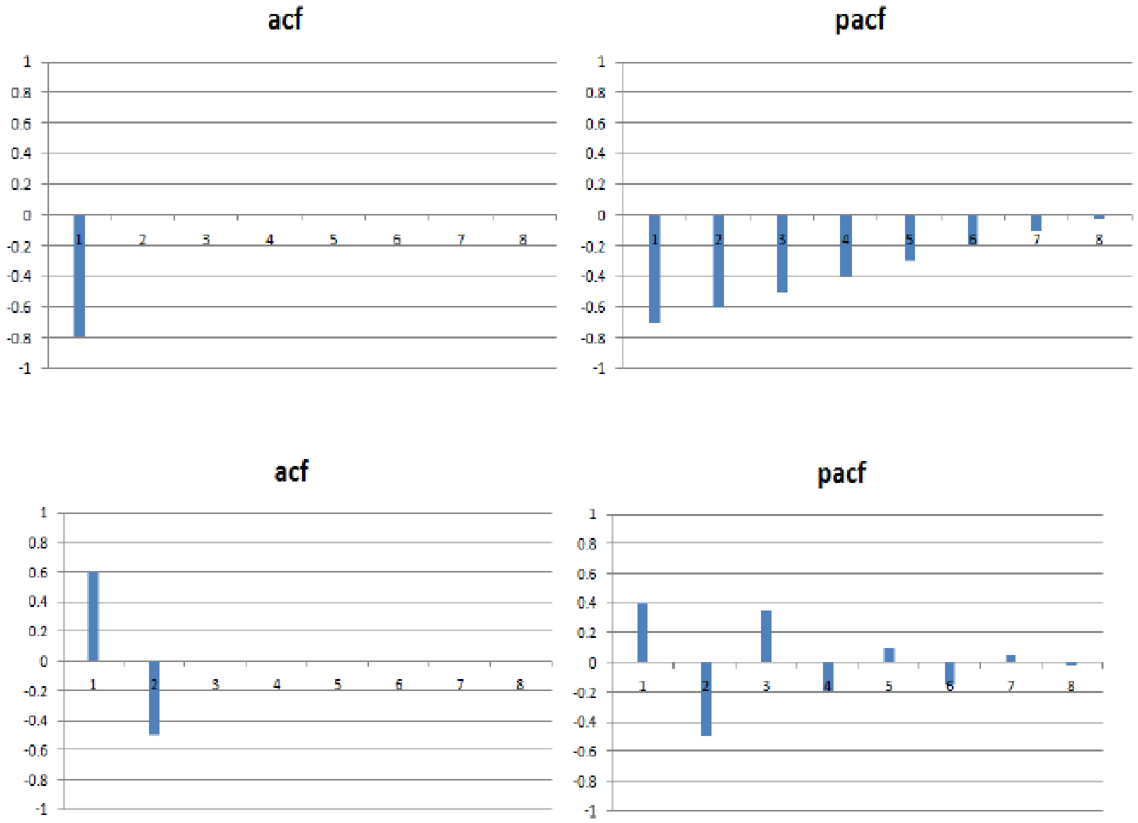
ARIMA
Recognition

$$Y_t = c - \theta_1 \times \mathrm{e}_{t-1} - \theta_2 \times \mathrm{e}_{t-2} - \cdots - -\theta_q \times \mathrm{e}_{t-q} + e_t$$

- For a stationary model ***MA(q):***
  - The ACF values must be 0 ***immediately*** after q lag-periods.
  - The PACF values must ***decrease*** (exponentially or sin) ***towards 0***.

MA(1)

MA(2)



MA(1): $-1 < \theta_1 < 1$

MA(2): $-1 < \theta_2 < 1,\ \theta_1 + \theta_2 > 1, \theta_1 - \theta_2 < 1$

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

ARIMA
Evaluation

- ***Likelihood***: the logarithm of the probability of the observed data

$$-2logL = n \times \left[log(2\pi) + 1 + log\left(\frac{RSS}{n}\right)\right]$$

- ***Akaike's Information Criterion (AIC)***:

$$AIC = -2logL + 2(p + q + k + 1)$$

$$AIC_c = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{n - p - q - k - 2}$$

- ***Bayesian Information Criterion (BIC)***:

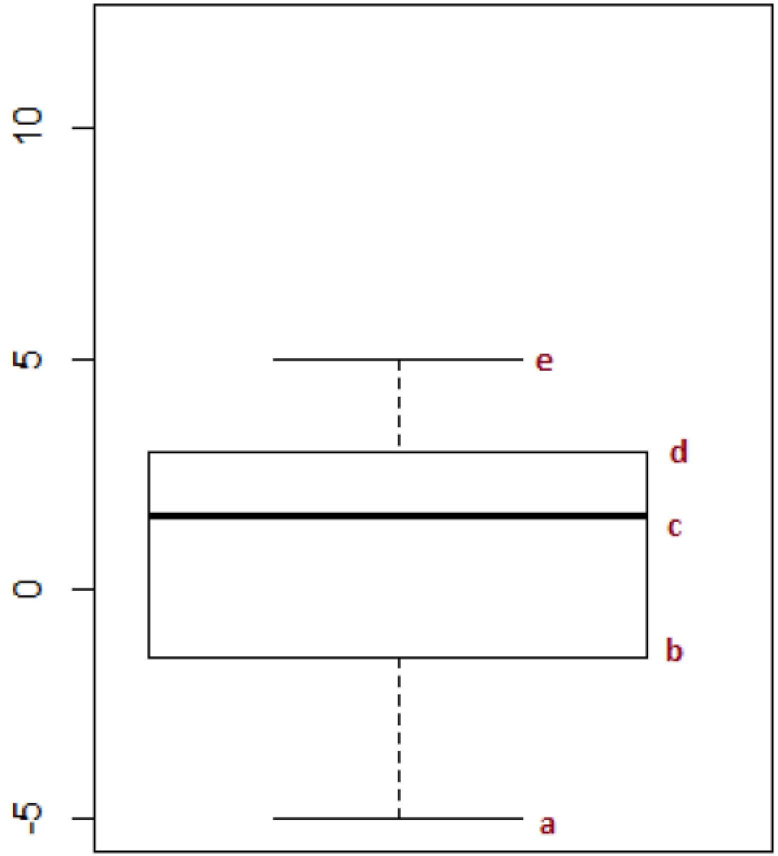
$$BIC = AIC + (logT - 2)(p + q + k + 1)$$

where n = number of observations, RSS=sum of squared errors, k=1 if the model has a constant (else 0)

➢ Good models are obtained by ***minimizing*** either the AIC, AIC_c or BIC.

ARIMA
Diagnostics

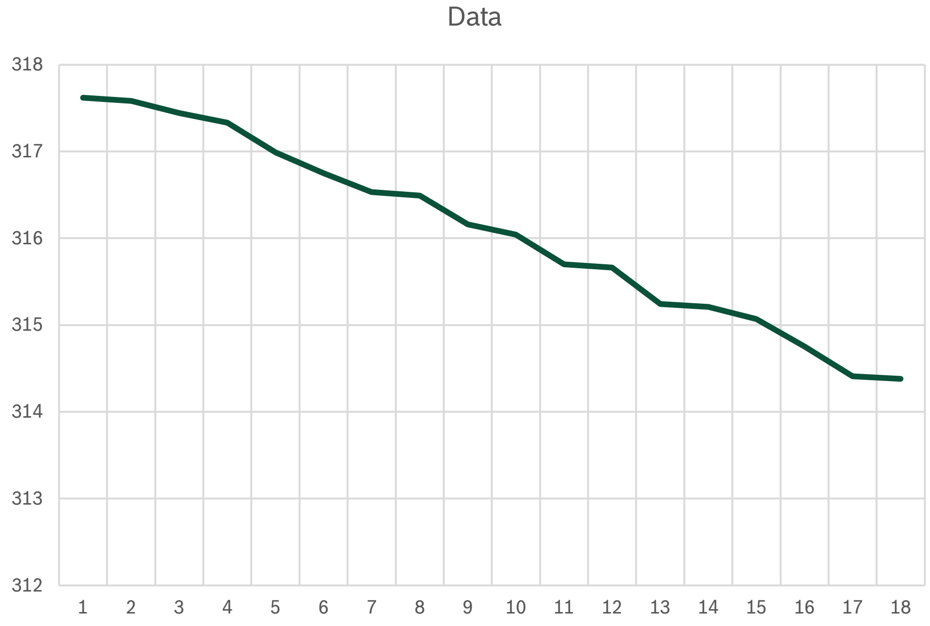
- ***Box plot***
  - Visualize ***error distribution***, where:
    - a = min
    - b = 25% of the observations
    - c = median
    - d = 25% of the observations
    - e = max
- Best forecasts when:
  - ***All distances are equal*** (a-b = b-c = c-d = d-e)
  - ***Median is zero*** (c = 0)

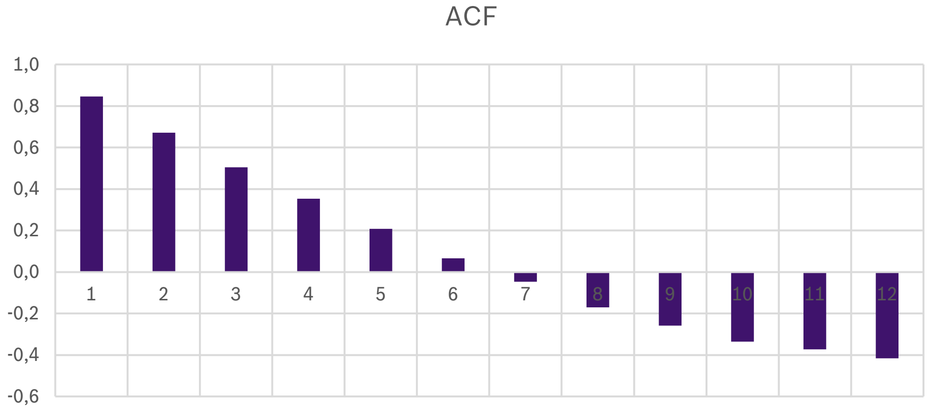


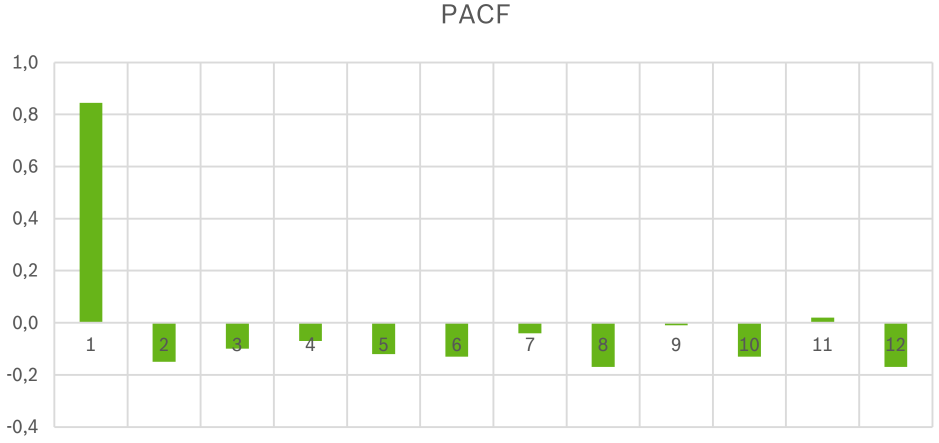
ARIMA
Example

- Estimate forecasts, for horizon = 3
  - Timeserie has a clear trend.

	Data
1	317,62
2	317,58
3	317,44
4	317,33
5	316,99
6	316,75
7	316,53
8	316,49
9	316,16
10	316,04
11	315,70
12	315,66
13	315,24
14	315,21
15	315,07
16	314,75
17	314,41
18	314,38



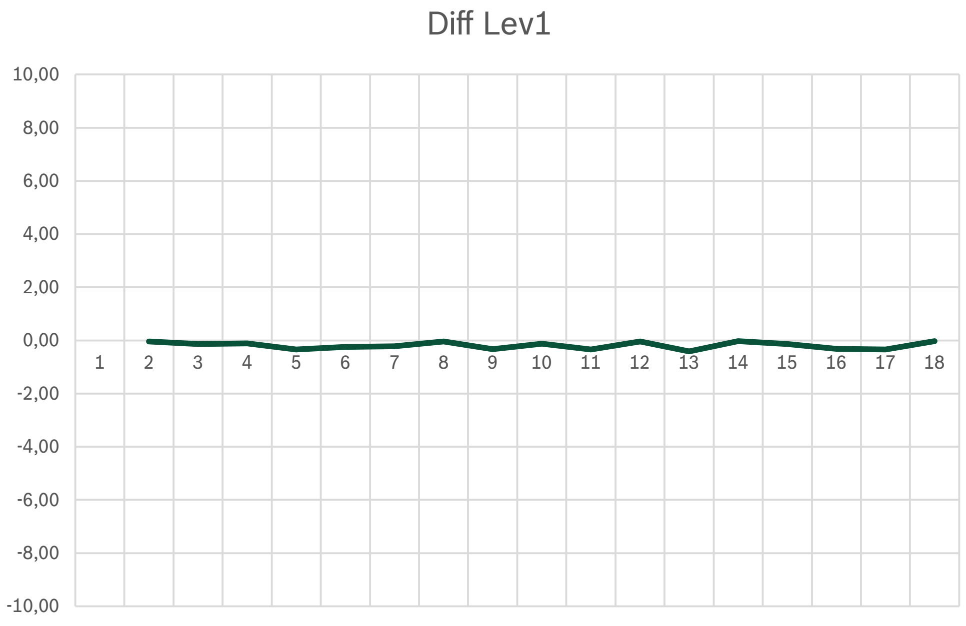




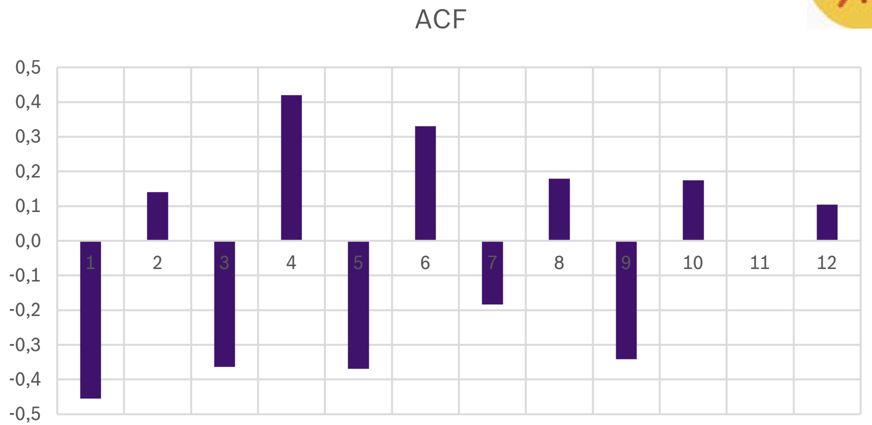
ARIMA
Example

- We need to do ***1st level integration***

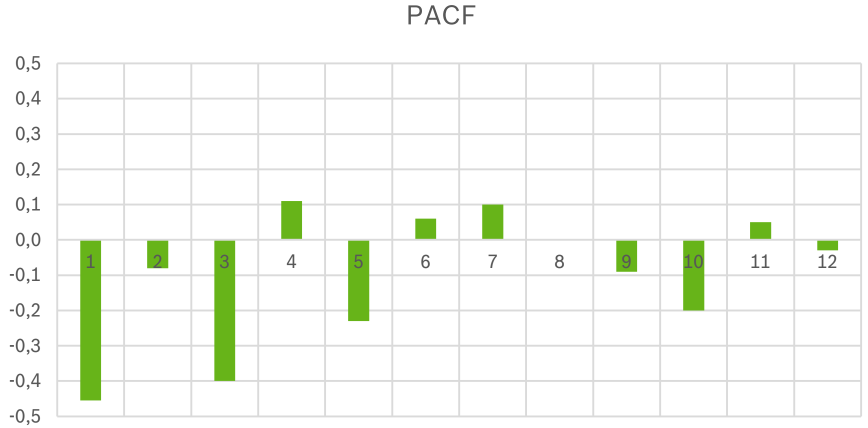
	Diff Lev1
1	
2	-0,04
3	-0,14
4	-0,11
5	-0,34
6	-0,24
7	-0,22
8	-0,04
9	-0,33
10	-0,12
11	-0,34
12	-0,04
13	-0,42
14	-0,03
15	-0,14
16	-0,32
17	-0,34
18	-0,03



- Data are ***stationary***, since average = -0,19



- ACF decreases (sin)



- PACF has large value for lag 1

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

ARIMA
Example

- Selected model: AR(1), over 1st level integration data. Thus ***ARIMA(1,1,0)***

			P1	
	Data	**Diff Lev1**	**Nom.**	**Denom.**
1	317,62		-	-
2	317,58	-0,04	-	0,023
3	317,44	-0,14	0,008	0,003
4	317,33	-0,11	0,004	0,006
5	316,99	-0,34	-0,012	0,022
6	316,75	-0,24	0,007	0,002
7	316,53	-0,22	0,001	0,001
8	316,49	-0,04	-0,004	0,023
9	316,16	-0,33	-0,021	0,019
10	316,04	-0,12	-0,010	0,005
11	315,70	-0,34	-0,011	0,022
12	315,66	-0,04	-0,022	0,023
13	315,24	-0,42	-0,035	0,053
14	315,21	-0,03	-0,037	0,026
15	315,07	-0,14	0,008	0,003
16	314,75	-0,32	-0,007	0,017
17	314,41	-0,34	0,019	0,022
18	314,38	-0,03	-0,024	0,026
sum			**-0,134**	**0,295**
average	**316,075**	**-0,191**	**P1=**	**-0,455**

$$\boldsymbol{\varphi_1} = p_1 = -0{,}455$$

$$\boldsymbol{c} = \bar{\acute{Y}} \times (1 - \varphi_1) = -0{,}277$$

$$\acute{Y}_t = c + \varphi_1 \times Y'_{t-1} + e_t$$

$$Y_t - Y_{t-1} = c + \varphi_1 \times (Y_{t-1} - Y_{t-2}) + e_t$$

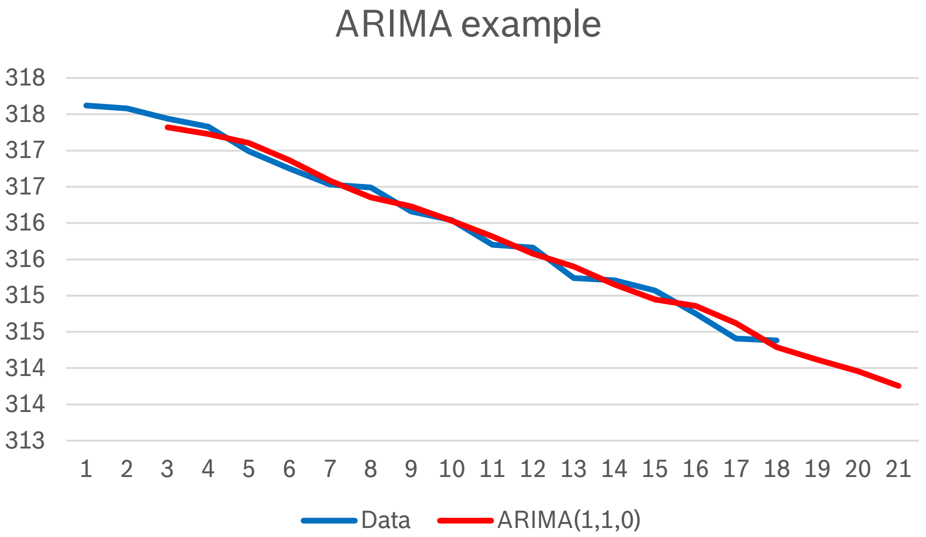
$$Y_t = c + \varphi_1 \times Y_{t-1} - \varphi_1 \times Y_{t-2} + e_t + Y_{t-1}$$

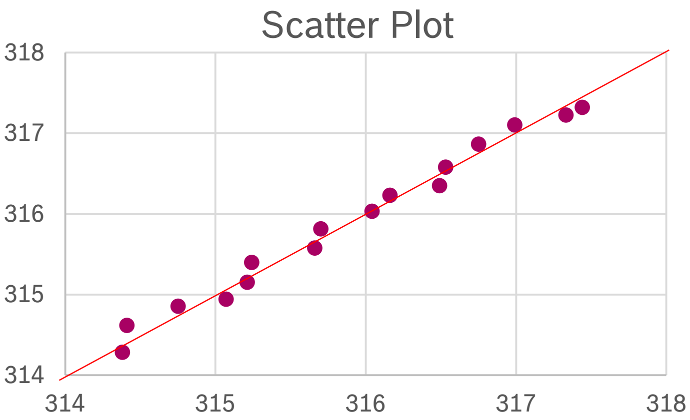
$$\boldsymbol{Y_t = -0,277 + 0,545 \times Y_{t-1} + 0,455 \times Y_{t-2} + e_t}$$

Forecasting Methods
Example

- Estimate forecasts

	Data	ARIMA(1,1,0)	Error
1	317,62		
2	317,58		
3	317,44	317,32	0,12
4	317,33	317,23	-0,02
5	316,99	317,10	-0,10
6	316,75	316,87	-0,02
7	316,53	316,58	-0,03
8	316,49	316,35	0,17
9	316,16	316,23	-0,24
10	316,04	316,03	0,25
11	315,70	315,82	-0,36
12	315,66	315,58	0,45
13	315,24	315,40	-0,61
14	315,21	315,15	0,66
15	315,07	314,95	-0,54
16	314,75	314,86	0,43
17	314,41	314,62	-0,64
18	314,38	314,29	0,73
19		314,12	
20		313,96	
21		313,75	

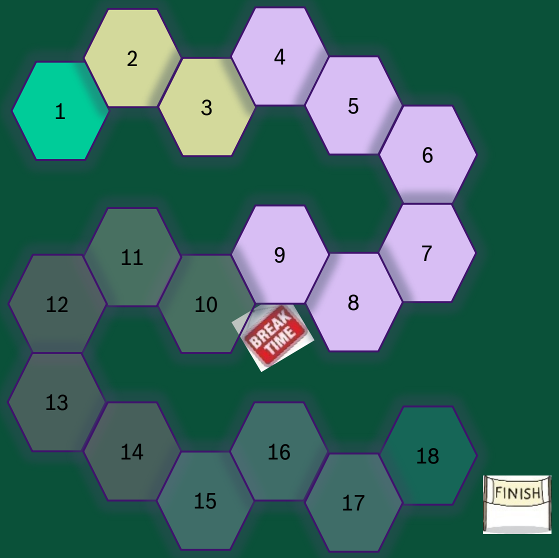




- Mean Error = 0,016
- Median error = -0,018

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

9. MULTIPLE REGRESSION



Multiple Regression
Definition

- ***Multiple Regression (MR)***: When several independent variables are required

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 + \cdots + b_k * X_k + e$$

- ***For 2 independent variables, we need to estimate 3 parameters***
  - calculate the partial derivatives for each one of the weights, assign the calculated derivatives equals with zero, and then solve a three-equation problem with three parameters.

$$Y_i = b_0 + b_1 * X_1 + b_2 * X_2 + e_i = \widehat{Y_i} + e_i$$

- For estimating the weights, we must minimize the error:

$$(b_0, b_1, b_2) \mid \min \left[\sum_{i=1}^{n} {e_i}^2 \right]$$

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(Y_i - \widehat{Y_i}\right)^2 = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 * X_{1,i} - b_2 * X_{2,i}\right)^2$$

- ***Important issue on multiple regression***:
  - the stability of the regression weights is depending on the **correlation** of the independent variables.
  - For two independent variables $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, the greater the correlations between them, the more unstable are the weights $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$.

Multiple Regression
How to define Predictors

- ***Dummy variables***
  - In most cases, the predictor variables are numerical.
  - But, what if a predictor is a ***categorical variable*** (true/false, yes/no)?
    - This predictor can be handled by creating a dummy variable taking values on 1 and 0.
    - If the predictor has more values, then the variable can be coded using several dummy variables.
- Example: use ***day of week*** as a predictor
  - ***6 dummy variables are needed***.
    - Last week day (Sunday) will be specified when all 6 dummy variables are set to zero.
    - Many beginners will try to add a seventh dummy variable for the seventh category.
    - This is known as the "***dummy variable trap***" because it will cause the regression to fail. There will be too many parameters to estimate. The general rule is to use ***one fewer dummy variables than categories***.

Multiple Regression
How to define Predictors

- ***Outliers***
  - If there is an outlier in the data, rather than omit it, it can be used as a dummy variable to remove its effect.
    - In this case, the dummy variable takes value one for that observation and zero everywhere else.
- ***Public holidays***
  - ***Static***: For daily data, the effect of public holidays can be accounted for by:
    - including a dummy variable predictor taking value one on public holidays and zero elsewhere.
  - ***Moving***: Easter is different from most holidays because it is not held on the same date each year and the effect can last for several days.
    - a dummy variable can be used with value one where any part of the holiday falls in the particular time period and zero otherwise.

Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Multiple Regression
How to define Predictors

- ***Working (trading) days***
  - The number of trading days in a month can vary considerably and can have a substantial effect on sales data.
    - To allow for this, the number of trading days in each month can be included as a predictor.
- **Intervention variables**
  - It is often necessary to model ***interventions*** that may have affected the variable to be forecast.
    - For example, competitor activity, advertising expenditure, industrial action, and so on, can all have an effect.
  - When the effect lasts ***only for one period***, a ***spike variable*** can be used.
    - Thus, a dummy variable taking value one in the period of the intervention and zero elsewhere.
    - A spike variable is equivalent to a dummy variable for handling an outlier.
  - When the effect is ***immediate and permanent*** (i.e. a level shift, the value of the series changes suddenly and permanently from the time of intervention), a ***step variable*** can be used.
    - A step variable takes value zero before the intervention and one from the time of intervention onwards.

BOSCH

Multiple Regression
Correlation Coefficient

- For MR case, we need to take into account the number of independent variables ***k*** and the number of observations ***n***. Thus, to use an ***"adjusted" coefficient***.

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{(Y_i - \bar{Y})^2}$$

$$\bar{R}^2 = 1 - (1 - R^2) * \frac{n-1}{n-k-1}$$

- This "adjusted" coefficient R^2 represents the percentage of the ***dispersion*** of the variable *Y*, which is justified by the independent variables.
  - the different ***(n-1)*** is the total degrees of freedom of the total variance of the model.
  - the ***(n-k-1)*** is the degrees of freedom of the interpreted variance.

Multiple Regression
Basic assumptions

1. The existence of a ***linear relationship*** between the dependent and the independent variables.
   - When this assumption is not valid, the independent variables must be transformed to other variables which have linear relationship with the dependent variable Y.
2. Constant ***variance of regression errors*** ("homoscedasticity", "heteroscedasticity").
   - This assumption indicates that the forecast errors should be stable throughout the range of observations.
3. The ***residual errors are independent*** of one another.
   - This means that the price of each residual error is independent of the values of the previous and the next.
   - When this assumption is not fulfilled, there is a serial correlation (or a auto-correlation) between consecutive values of residual errors. **Thus, a significant independent variable has been skipped!**.
     - Alternative ways of recognizing the independence of the residual errors is the graphical representation of their values and the calculation of the Durbin-Watson statistical index.

Multiple Regression
Basic assumptions

4. An important issue in MR is the probability of multi-collinearity.
   - The multi-collinearity arises ***when two or more independent variables are highly correlated***.
   - This is a frequent problem in financial and operational data, due to the high degree of correlation that exists between the different factors.
   - This fact should be taken into account, when selecting the independent variables and the data collection.
   - The goal is the use of independent variables which are not strongly correlated
     - a rule of thumb is that the correlation ***should not exceed the value 0.7 or less than -0.7***.

   - If the independent variables are strongly correlated, then they provide ***redundant information*** which does not improve the explanatory strength of the regression.

Multiple Regression
How NOT to select predictors

- ***Visual inspection***
  - A common approach is to plot the forecast variable against the specific predictor and to visual examine their relationship.
    - This is an approach that can lead to faults, because it is not always possible to see the relationship from a graph. It should be avoided.
- ***MLR on all predictors***
  - Another common approach is to perform multiple linear regression on all the predictors and discharged the predictors with ***p*** value ***greater than 0.05***.
    - This can also lead to faults and should be avoided, since **statistical significance does not always indicate predictive value**.
    - Also, ***p*** values can be misleading when two or more predictors are correlated with each other.

Multiple Regression
How NOT to select predictors

- ***R^2***
  - It is not a good approach to select predictors based on R^2 value.
    - If a model produces forecasts that are always 30% higher than the actual values, then R^2 is equal to 1 indicating a perfect correlation. But the forecasts are not close to the actual values.
    - Also, ***R^2 does not allow for degrees of freedom.*** Adding any variable tends to increase the R^2 value, even if the new variable is irrelevant.
- ***Mean Square Error (MSE)***
  - Another approach is to select predictors based on minimizing the mean of square errors (MSE) or sum of square errors (SSE).
    - But, minimizing MSE or SSE is equivalent to maximizing R^2 and **will always close the model with the most variables**.

Multiple Regression
How to select predictors

- ***Adjusted R^2***
  - An alternative approach for selecting predictors in the adjusted R^2

$$\bar{R}^2 = 1 - (1 - R^2) * \frac{n-1}{n-k-1}$$

where n is the number of observations and k is the number of predictors.

  - This is improving R^2, since it will no longer increase with each added predictor. The best set of predictors is the one which **maximizes** the value of adjusted R^2.
  - This approach works quite well as a method for selecting predictors, although it does ***tent sometimes*** to err on the side of selecting too many predictors.
- ***Variance of errors***
  - An equivalent approach for selecting predictors, in the minimization of the variance of the forecasts errors

$$\hat{\sigma}^2 = \frac{Sum\ of\ Squared\ Errors\ (SSE)}{n-k-1}$$

Multiple Regression
How to select predictors

- ***Cross-Validation***
  - a useful approach for examining the predictive ability of a forecast model. In general, leave-one-out cross-validation must follow the following steps:
    - Remove observation 1 for the data set, and fit the model using the remaining data. Then, forecast the value for observation 1 and estimate the ***cross-validation error*** (it is not the same as residual, since the 1st observation was not used for estimating forecasts).

$$e_1^* = Y_1 - \widehat{Y_1}$$

    - Remove the previous step for all observations of the data step, thus 1,2, ... n
    - Estimate the mean square error (MSE) from all cross-validation errors (CVe)
    - Repeat all previous steps for each model (for each predictors variables set).
  - ***The best model is the one with the minimum CVe.***

Multiple Regression
How to select predictors

- ***Akaike's Information Criterion (AIC)***
  - Another approach is Akaike's information criterion, which can be estimated as: $AIC = n * \log\left(\frac{SSE}{n}\right) + \ 2 * (k+2)$

    , where n is the number of observations and k the number of predictors in the model.
  - The *k+2* part of the equation is because there are k+2 parameters in the model (the k coefficients for the predictors, the intercept and the variance of the residuals).
  - ***The model with the minimum AIC value is often the best model for forecasting.***
- ***Corrected AIC_c***
  - For small amount of observations n, the AIC tends to select too many predictors.
  - So, a corrected AIC can be used: $AIC_C = AIC + \frac{2*(k+2)*(k+3)}{n-k-3}$
  - ***The model with the minimum AIC_c value is often the best model for forecasting.***

Multiple Regression
How to select predictors

- ***Schwarz Bayesian Information Criterion (SBIC)***
  - Another approach is the Schwarz Bayesian information criterion, which can be calculated as:

$$SBIC = n * \log\left(\frac{SSE}{n}\right) + (k+2) * \log n$$

  - ***As with AIC, the best model is the one with the minimum SBIC.***
  - The model chosen by SBIC is either the same as that chosen by AIC, or one with fewer predictors.
  - This is because SBIC penalizes the SSE (sum of squared errors) more heavily that AIC.
    - Many statisticians like to use SBIC because it has the feature that ***if there is a true underlying model, then with enough data the SBIC will select that model.***
    - However, in reality there is rarely if ever a true underlying model, and even if there was a true underlying model, selecting that model will not necessarily give the best forecasts (since the predictors values must also be forecasted, and these forecasts may not be accurate!).

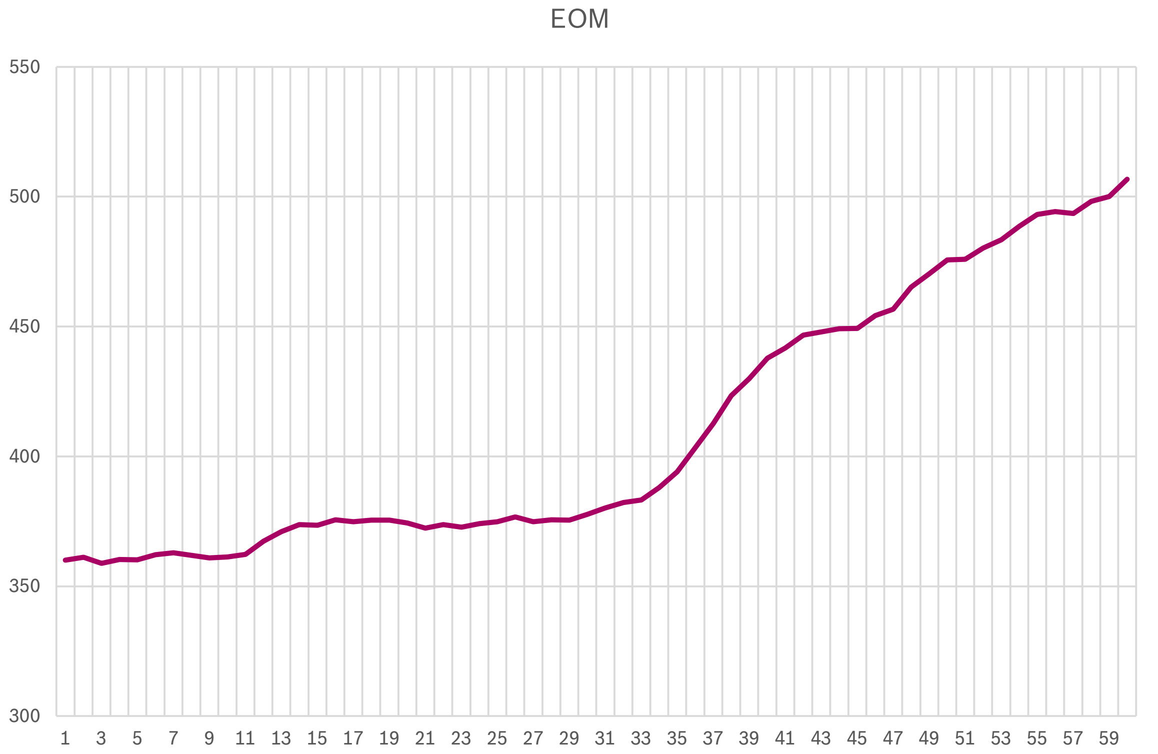
  - It must be mentioned that, SBIC ***does tent sometimes to err on the side of selecting too few predictors.***

BOSCH

Multiple Regression
Example

- ***Mutual savings bank deposit in a large metropolitan area***
  - Problem:
    - Monthly changes in deposits were getting smaller
    - monthly changes in withdrawals were getting bigger.
  - Develop a short-term model:
    - to forecast EOM changes.

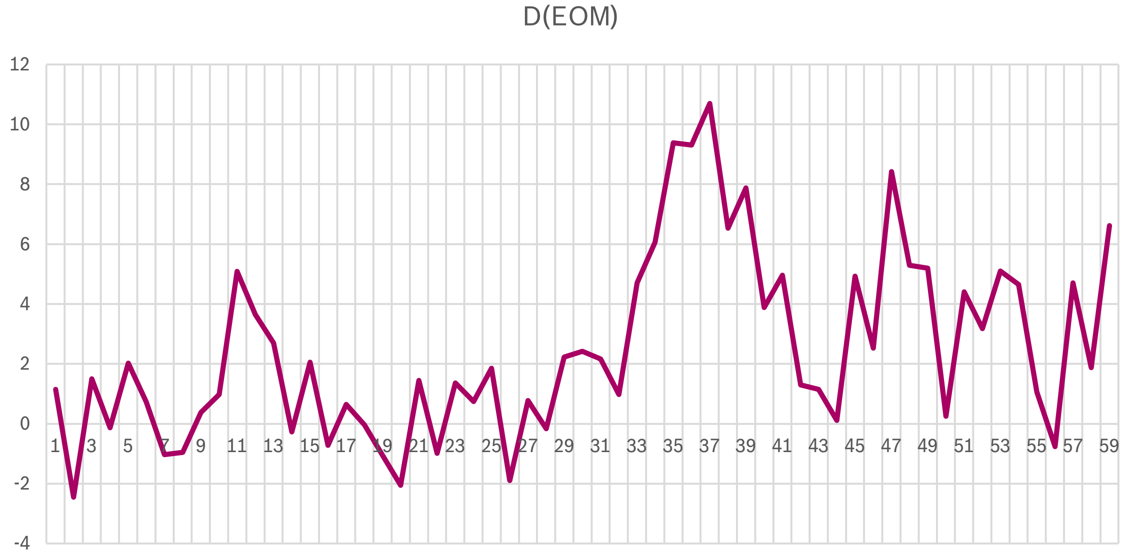
Months	EOM	Months	EOM	Months	EOM	Months	EOM	Months	EOM
1	360,071	13	371,031	25	374,880	37	412,727	49	470,408
2	361,217	14	373,734	26	376,735	38	423,417	50	475,600
3	358,774	15	373,463	27	374,841	39	429,948	51	475,857
4	360,271	16	375,518	28	375,622	40	437,821	52	480,259
5	360,139	17	374,804	29	375,461	41	441,703	53	483,432
6	362,164	18	375,457	30	377,694	42	446,663	54	488,536
7	362,901	19	375,423	31	380,119	43	447,964	55	493,182
8	361,878	20	374,365	32	382,288	44	449,118	56	494,242
9	360,922	21	372,314	33	383,270	45	449,234	57	493,484
10	361,307	22	373,765	34	387,978	46	454,162	58	498,186
11	362,290	23	372,776	35	394,041	47	456,692	59	500,064
12	367,382	24	374,134	36	403,423	48	465,117	60	506,684



Source: "Forecasting: Methods and Applications", Spyros Makridakis, Steven C. Wheelwright

Multiple Regression
Example

- The interest of the bank was the change in the EOM balance and so first differences of the EOM data are estimated.



- It is clear that the bank was facing a ***volatile situation*** in the last two years or so.
- The challenge is to forecast these rapidly changing EOM values.

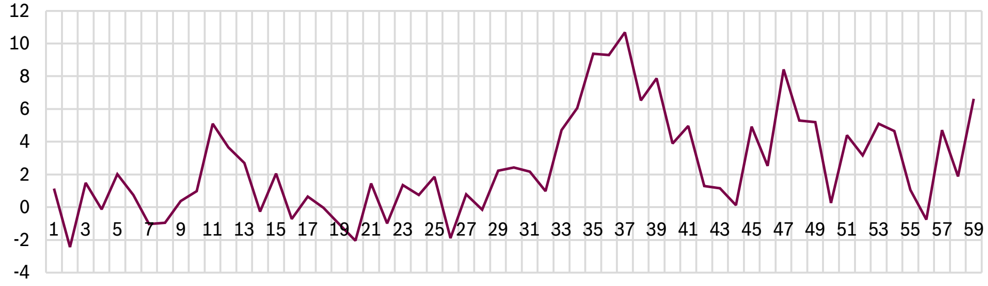
Multiple Regression
Example

- It was hypothesized that two variables ***had an influence*** on the EOM balance figures:
  - Composite AAA bond rates, and
  - The rates on U.S. Government 3-4 year bonds.
- ***Three explanatory*** variables (regressors):
  - X_1: the AAA bond rates, but they are now shown leading the D(EOM) values
  - X_2: the rates on 3-4 year government bonds, but they are shown leading the D(EOM) values by 1 month.
  - X_3: The first differences of the 3-4 year bond rates, and the timing for this variable coincides with that of the D(EOM) variable.

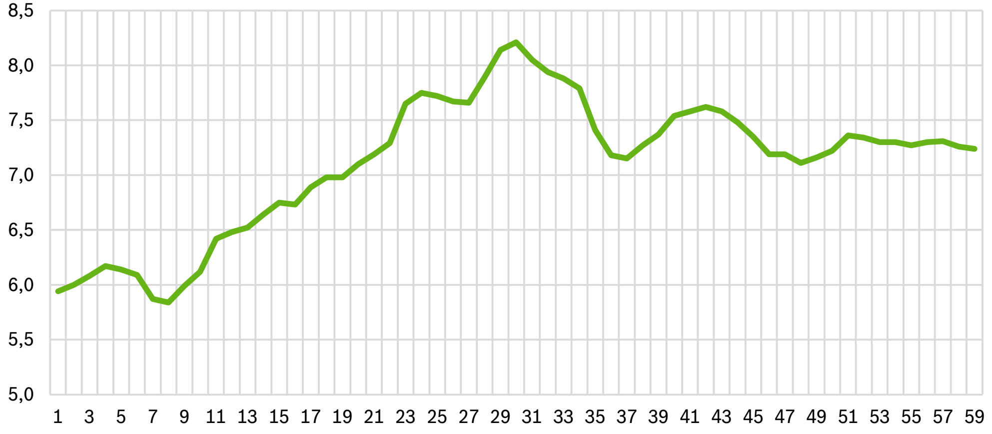
➢ ***D(EOM) is the forecast value***

Multiple Regression
Example

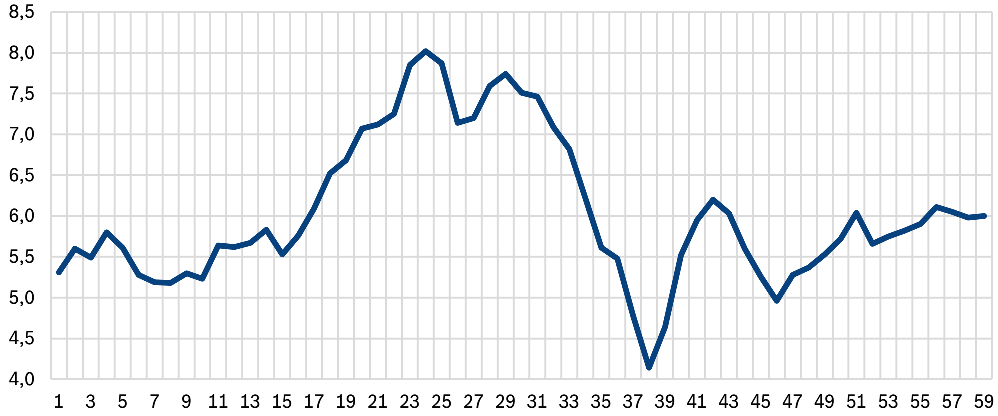
D(EOM)



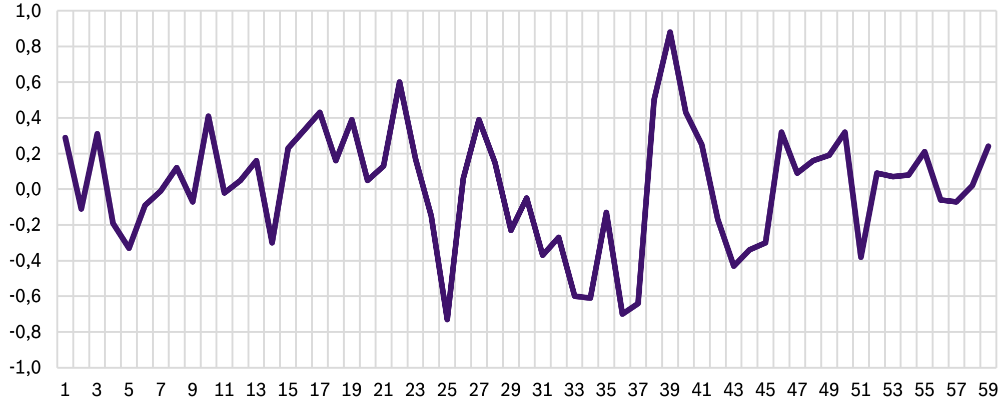
(AAA)



(3-4)



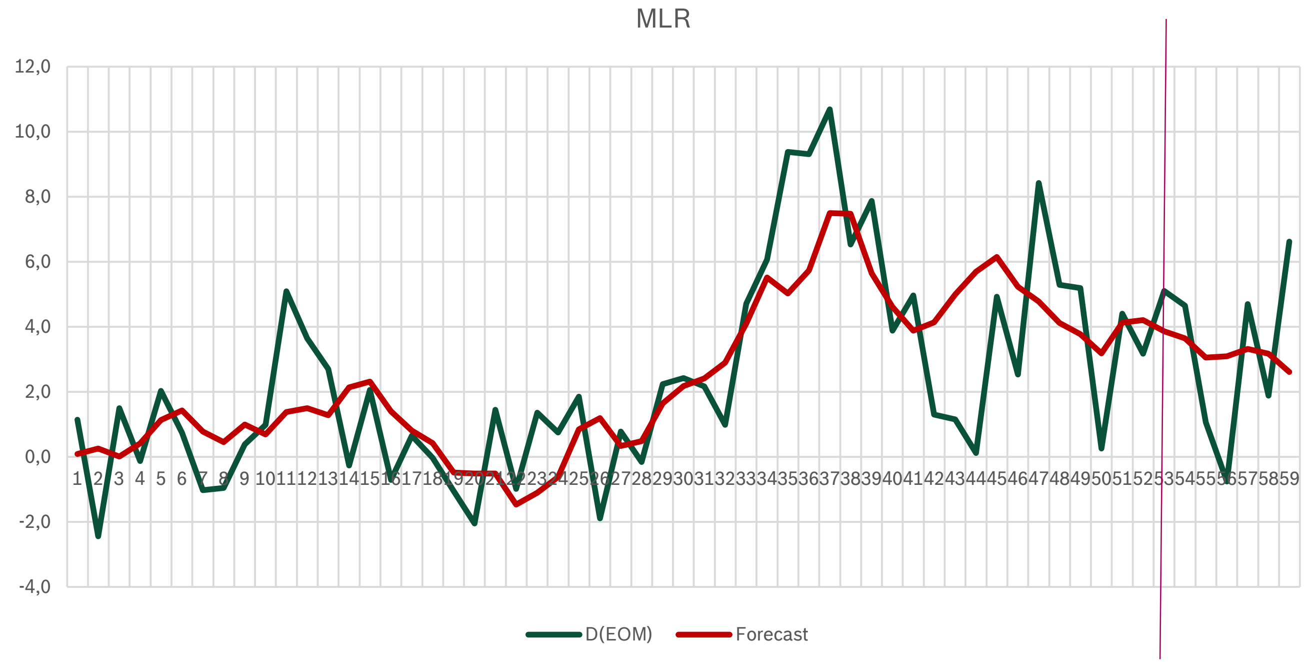
D(3-4)



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Multiple Regression
Example

- ***Model Estimation***: By using a least square error, we can estimate:
  - $Y = -4.3391 + 3.3722 * X_1 - 2.8316 * X_2 - 1.9648 * X_3$



	Y	X			Y'
Months	**D(EOM)**	**(AAA)**	**(3-4)**	**D(3-4)**	**Forecast**
1	1,146	5,940	5,310	0,290	0,086
2	-2,443	6,000	5,600	-0,110	0,253
3	1,497	6,080	5,490	0,310	0,009
4	-0,132	6,170	5,800	-0,190	0,417
5	2,025	6,140	5,610	-0,330	1,129
6	0,737	6,090	5,280	-0,090	1,424
7	-1,023	5,870	5,190	-0,010	0,779
8	-0,956	5,840	5,180	0,120	0,451
9	0,385	5,990	5,300	-0,070	0,990
10	0,983	6,120	5,230	0,410	0,684
11	5,092	6,420	5,640	-0,020	1,379
12	3,649	6,480	5,620	0,050	1,501
13	2,703	6,520	5,670	0,160	1,278
14	-0,271	6,640	5,830	-0,300	2,134
15	2,055	6,750	5,530	0,230	2,313
16	-0,714	6,730	5,760	0,330	1,397
17	0,653	6,890	6,090	0,430	0,806
45	4,928	7,350	5,260	-0,300	6,142
46	2,530	7,190	4,960	0,320	5,234
47	8,425	7,190	5,280	0,090	4,779
48	5,291	7,110	5,370	0,160	4,117
49	5,192	7,160	5,530	0,190	3,774
50	0,257	7,220	5,720	0,320	3,183
51	4,402	7,360	6,040	-0,380	4,124
52	3,173	7,340	5,660	0,090	4,209
53	5,104	7,300	5,750	0,070	3,859

Multiple Regression
Example

- ***Coefficient of Determination R^2*:**
  - For estimating the multiple regression coefficient, we can estimate the correlation between D(EOM) and Forecasts. Thus:
    - $R^2 = r_{Y\hat{Y}}{}^2 = 0.749^2 = \mathbf{0.561}$
  - Alternative:
    - $R^2 = \frac{\sum_{i=1}^{53}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{53}(Y_i - \bar{Y})^2} = \frac{explained\ SumSquaredDev}{total\ SumSquaredDev} = \frac{SSR}{SST} = \frac{280.389}{499.986} = 0.561$

	Y	Y'	SSE		SSR		SST	
Months	**D(EOM)**	**Forecast**	**Y' - Y**	**(Y' - Y)^2**	**Y' - AvgY**	**(Y' - AvgY)^2**	**Y-AvgY**	**(Y-AvgY)^2**
1	1,146	0,086	1,060	1,123	-2,338	5,465	-1,278	1,633
2	-2,443	0,253	-2,696	7,270	-2,171	4,712	-4,867	23,686
3	1,497	0,009	1,488	2,213	-2,415	5,830	-0,927	0,859
4	-0,132	0,417	-0,549	0,302	-2,006	4,026	-2,556	6,532
5	2,025	1,129	0,896	0,802	-1,295	1,676	-0,399	0,159
6	0,737	1,424	-0,687	0,471	-1,000	1,001	-1,687	2,846
7	-1,023	0,779	-1,802	3,248	-1,645	2,704	-3,447	11,881
8	-0,956	0,451	-1,407	1,980	-1,973	3,892	-3,380	11,424
9	0,385	0,990	-0,605	0,367	-1,433	2,055	-2,039	4,157
44	0,116	5,696	-5,580	31,137	3,272	10,707	-2,308	5,326
45	4,928	6,142	-1,214	1,473	3,718	13,823	2,504	6,271
46	2,530	5,234	-2,704	7,309	2,810	7,894	0,106	0,011
47	8,425	4,779	3,646	13,291	2,355	5,548	6,001	36,014
48	5,291	4,117	1,174	1,378	1,693	2,867	2,867	8,220
49	5,192	3,774	1,418	2,011	1,350	1,822	2,768	7,663
50	0,257	3,183	-2,926	8,560	0,759	0,576	-2,167	4,695
51	4,402	4,124	0,278	0,077	1,700	2,891	1,978	3,913
52	3,173	4,209	-1,036	1,074	1,785	3,187	0,749	0,561
53	5,104	3,859	1,245	1,551	1,435	2,059	2,680	7,183
SUM	128.465	128.465	0,004	219,605	-0,004	**280,389**	0,000	**499,986**
AVG	2,424							

- ➢ The sums and the average value of D(EOM) and Forecasts are the same.
- ➢ The sum of residuals (errors) is zero, as it will always be for fitting of linear regression models.
- ➢ Also, the coefficient of determination is highly significant.

Multiple Regression
Example

- Correlation between variables

Correlation	D(EOM)	(AAA)	(3-4)	D(3-4)
D(EOM)	-	0,257	-0,391	-0,195
(AAA)	0,257	-	0,587	-0,204
(3-4)	-0,391	**0,587**	-	-0,201
D(3-4)	-0,195	-0,204	-0,201	-

  - None of the explanatory variables has a particularly high correlation with the Y value D(EOM).
  - The explanatory variables themselves do not correlate very highly.
  - We suspect no multicollinearity problem (since 0,587 is the biggest between X_i).

- The correlations of D(EOM) with X_i suggests that ***these three explanatory variables together will not be able to explain a lot of the variance in Y.***
  - They do combine to explain 56% (R^2), it is a significant contribution (F-test), and all three coefficients are significantly different from zero (t-test), but ***more can be done***.

BOSCH

Multiple Regression
Example

- What more can be done?
  - Since the bank data are monthly, we can use 11 dummy variables for the ***months***.

	Y	X			D										
Months	D(EOM)	(AAA)	(3-4)	D(3-4)	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
1	1,146	5,940	5,310	0,290	1	0	0	0	0	0	0	0	0	0	0
2	-2,443	6,000	5,600	-0,110	0	1	0	0	0	0	0	0	0	0	0
3	1,497	6,080	5,490	0,310	0	0	1	0	0	0	0	0	0	0	0
4	-0,132	6,170	5,800	-0,190	0	0	0	1	0	0	0	0	0	0	0
5	2,025	6,140	5,610	-0,330	0	0	0	0	1	0	0	0	0	0	0
6	0,737	6,090	5,280	-0,090	0	0	0	0	0	1	0	0	0	0	0
7	-1,023	5,870	5,190	-0,010	0	0	0	0	0	0	1	0	0	0	0
50	0,257	7,220	5,720	0,320	0	1	0	0	0	0	0	0	0	0	0
51	4,402	7,360	6,040	-0,380	0	0	1	0	0	0	0	0	0	0	0
52	3,173	7,340	5,660	0,090	0	0	0	1	0	0	0	0	0	0	0
53	5,104	7,300	5,750	0,070	0	0	0	0	1	0	0	0	0	0	0
54	4,646	7,300	5,820	0,080	0	0	0	0	0	1	0	0	0	0	0
55	1,060	7,270	5,900	0,210	0	0	0	0	0	0	1	0	0	0	0
56	-0,758	7,300	6,110	-0,060	0	0	0	0	0	0	0	1	0	0	0
57	4,702	7,310	6,050	-0,070	0	0	0	0	0	0	0	0	1	0	0
58	1,878	7,260	5,980	0,020	0	0	0	0	0	0	0	0	0	1	0
59	6,620	7,240	6,000	0,240	0	0	0	0	0	0	0	0	0	0	1

Multiple Regression
Example

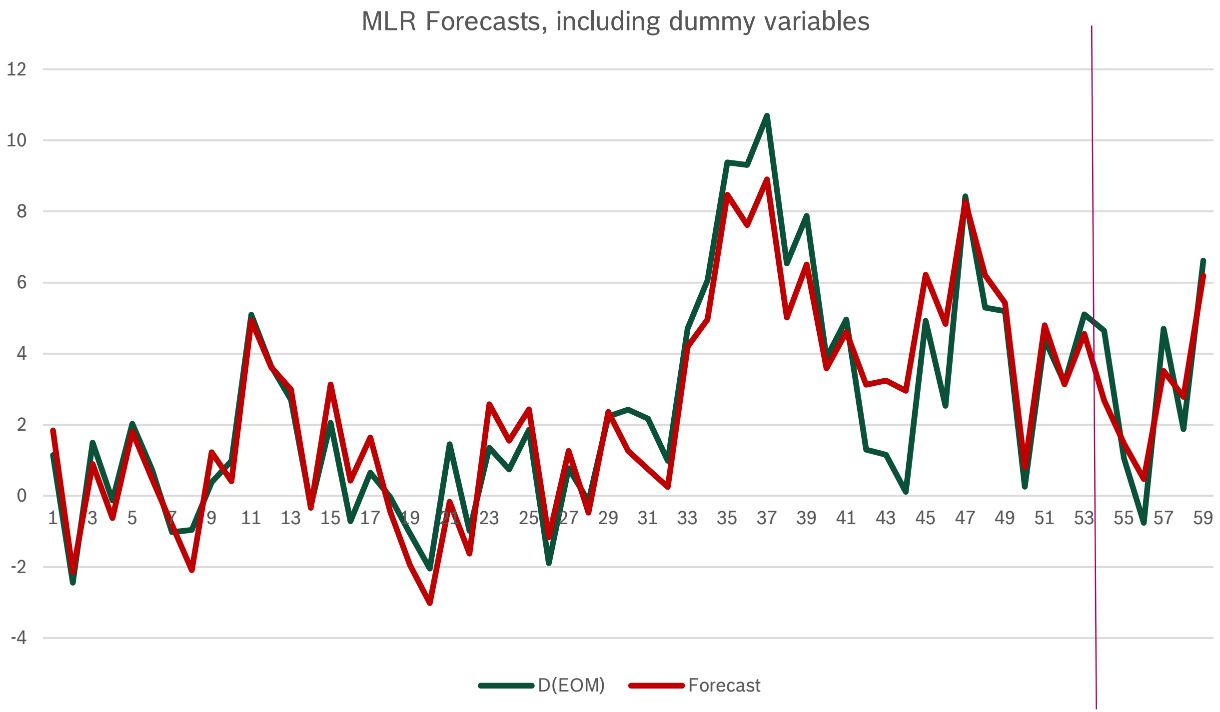
- If we run again a least square method (over the first 53 rows of data by using 3+11 explanatory variables), we can estimate that:

$$
\begin{aligned}
Y' = -2.1983 + 3.2988 * X_1 - 2.7524 * X_2 - 1.7308 * X_3 \\
- 0.4403 * D_1 - 4.5125 * D_2 - 1.3130 * D_3 - 3.1493 * D_4 - 1.3833 * D_5 - 3.0397 * D_6 \\
- 3.7102 * D_7 - 4.6966 * D_8 - 1.8684 * D_9 - 2.4762 * D_{10} + 1.4419 * D_{11}
\end{aligned}
$$

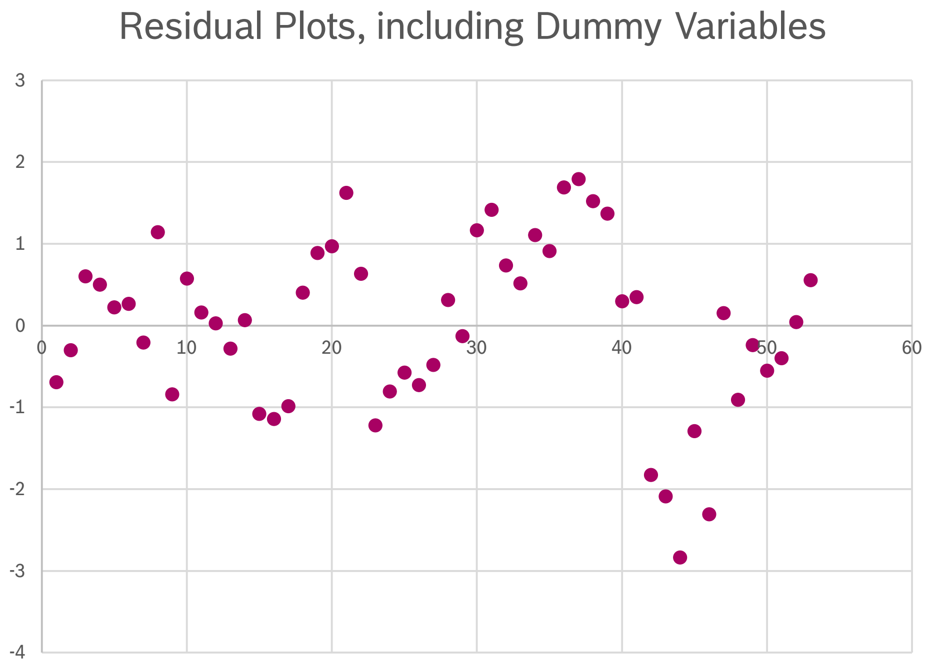
- Results:
  - ***R^2 = 0.887*** (previously 0.561)
    - The proportion of Y explained by regressing on these 14 explanatory variables is now 88.7%, instead on just 56.1%.
  - MSE = 1.49 (previously 4.48)
    - The mean squared error ***has dropped considerably***, from 4.48 to 1.49.

Multiple Regression
Example

▶ The forecasts are significantly better.



▶ The residuals are generally smaller, and with no indication of a seasonality.

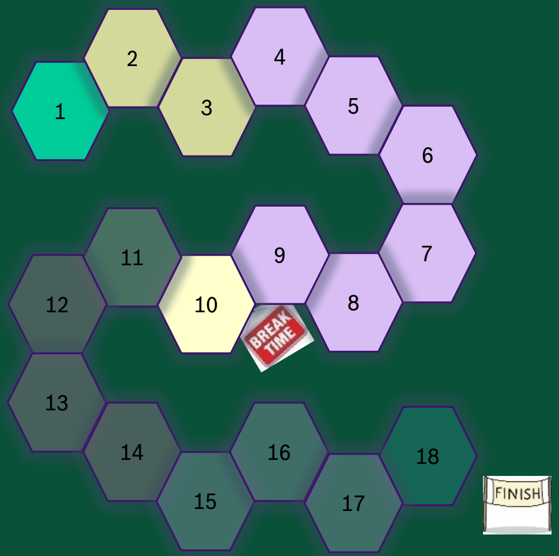


BOSCH

TIMESERIES FORECASTING (A)

BREAK TIME

10. LIMITS OF PREDICTABILITY



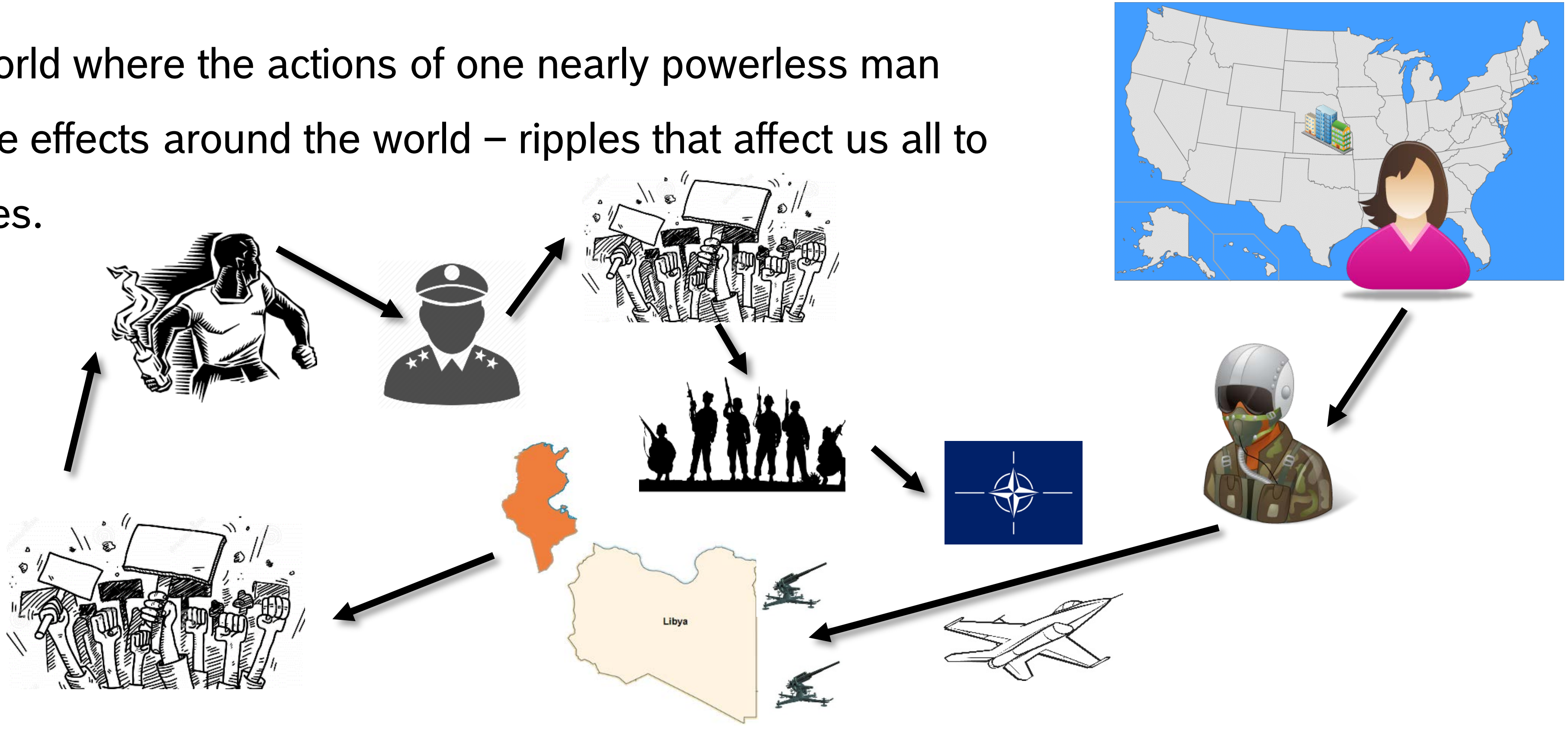
Limits of predictability
Introduction

- All techniques are ***extrapolative*** in nature.
  - They work well when the ***future is similar to the past***, or when changes happen to cancel out.
  - If established patterns or relationships change → no simple and reliable way to predict
- Complains about forecasting have increased over the years & the requests for forecasting have also increased.
  - Not a paradox!
  - In a turbulent environment with high uncertainty, ***the need for forecasts is great***.

Forecasting is not crystal-balling. Realistic expectations should be set.

Limits of predictability
Everything is related

- We live in a world where the actions of one nearly powerless man can have ripple effects around the world – ripples that affect us all to varying degrees.



It is misguided to think anyone can see very far into the future.

Limits of predictability
Factors that influence predictability

- ***Number of Items***: The larger the number of items involved, the more accurate the forecasts.
  - **Statistical law of large numbers**: the size of forecasting error, and therefore the accuracy, decreases as the number of items being forecast increases, and vice versa.
    - It is more accurate to predict the number of telephone calls arriving at a switching station during a five-minute interval that the number of personal computers sold on a certain day.

- ***Homogeneity of Data***: The more homogeneous the data, the more accurate the forecasts, and vice versa.
    - Data referring to a single region can predict seasonality more accurate that data covering many regions with varying weather patterns.

BOSCH

Limits of predictability
Factors that influence predictability

- ***Elasticity of Demand***: The more inelastic, the more accurate the forecasts.
    - – The demand for necessities can be forecast with a higher degree of accuracy than the demand for luxuries.
  - Related to the elasticity of demand is the influence of business cycles.
    - – Such cycles have the least impact on inelastic demand and the greatest impact on elastic demand.

- ***Competition***: The greater the competition, the more difficult it is to forecast.
  - Competitors can use the forecasts to change the course of future events and thus invalidate the forecasts.

BOSCH

Limits of predictability
Sources of errors & future uncertainty

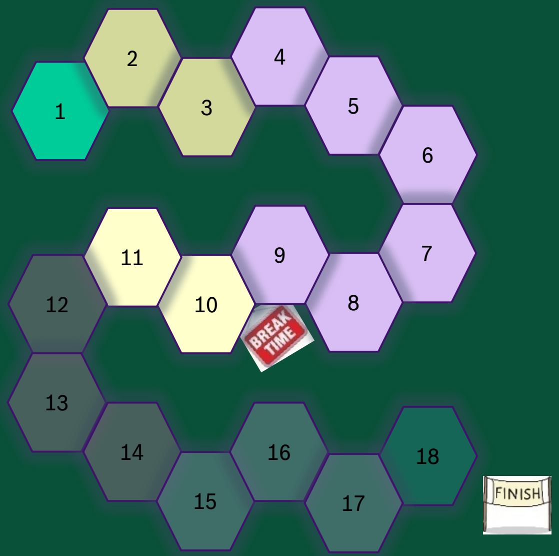
- ***Erroneous Identification of Patterns and Relationships***:
  - An illusory pattern or relationship might be identified when none really exists.
    - People often glimpse illusory correlation, while statistical models based on a small number of observations can "identify" a ***pattern that is not maintained*** over a longer period.
    - A relationship between two variables might be spurious, existing only because a ***third factor*** causes both variables to move in the same direction.
    - Patterns or relationships that exist might be ***incorrectly identified or ignored*** because insufficient information is available, or because reality is too complex to be understood or modelled with a limited number of variables.

  - Illusory or inappropriate identification ***can cause serious and non-random forecasting errors***, since the future could turn out to be very different from what was postulated by an erroneous pattern or relationship.

Limits of predictability
Sources of errors & future uncertainty

- ***Inexact Patterns or Imprecise Relationships***:
  - Although an average pattern or relationship can be identified, fluctuations around such an average exist in almost all cases.
    - The purpose of statistical modelling is to identify patterns or relationships in such a way as to make ***past fluctuations around the average as small and random*** as possible.
    - Whether or not this is a good strategy is questionable, but even if it is appropriate, it does not guarantee that future errors will be random or symmetric or that they will not exceed a certain magnitude.
- ***Changing Patterns or Relationships***:
  - Patterns or relationships may constantly ***changing over time*** in a way that is not predictable in the great majority of cases.
    - Changes in patterns or relationships can cause large persistent errors whose magnitude cannot be known in advance.
    - The size of such errors depends on the magnitude and duration of the change.

11. NEURAL NETWORKS

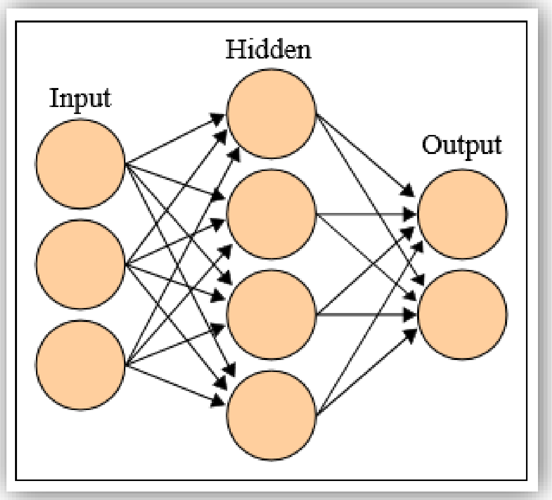


Neural Networks Introduction
Definition

- ***Neural Network***:
  - A mathematical model inspired by biological neural networks.
  - A neural network consists of an ***interconnected group of artificial neurons***, and it processes information using a connectionist approach to computation.
  - In Artificial Intelligence are also referred to as ***Machine Learning*** (ML) or ***artificial neural networks*** (ANNs)
    - these are essentially simple mathematical models defining a function or a distribution over or both and, but sometimes models are also intimately associated with a particular learning algorithm or learning rule.
    - A common use of the phrase ANN model really ***means the definition of a class of such functions*** (where members of the class are obtained by varying parameters, connection weights, or specifics of the architecture such as the number of neurons or their connectivity).

Neural Networks Introduction
Definition

- ***Network***: refers to the inter-connections between the neurons in the different layers of each system.
  - An example system has three layers.
    - The first layer has ***input neurons***, which send data via ***synapses*** to the second layer of neurons, and then via more synapses to the third layer of ***output neurons***.
    - More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons.
  - The synapses store parameters called "***weights***" that manipulate the data in the calculations.



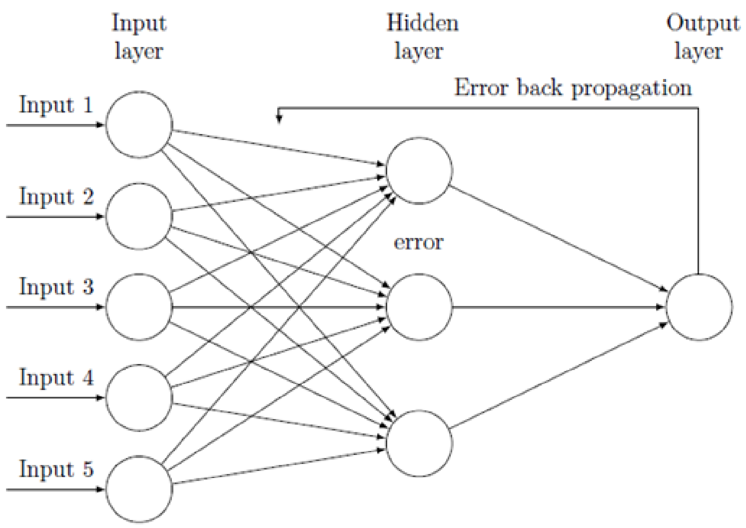
Neural Networks Introduction
Definition

- An ANN is typically defined by three types of parameters:
  - The ***interconnection*** pattern between different layers of neurons.
  - The ***learning process*** for updating the weights of the interconnections.
    - the process through which a neural network modifies itself to being able to produce a certain result with a given input.
  - The ***activation function*** that converts a neuron's weighted input to its output activation.
- Data:
  - NN is an adaptive system that changes its structure during a learning phase. Neural networks are used to model ***complex relationships between inputs and outputs*** or to find patterns in data.
  - They need a ***lot of data*** (in order to train the network) and they are usually ***time-consuming***.

Neural Networks Leaning
Definition

- ***Learning***:
  - Given a specific task to solve, and a class of functions F, learning means ***using a set of observations*** to find which ***solves the task*** in some optimal sense.
  - This entails defining a ***cost function*** such that, for the optimal solution, - i.e., no solution has a cost less than the cost of the optimal solution.
  - The ***cost function C*** is an important concept in learning, as it is a measure of how far away a particular solution is from an optimal solution to the problem to be solved.
  - Learning algorithms ***search through the solution space*** to find a function that has the smallest possible cost.

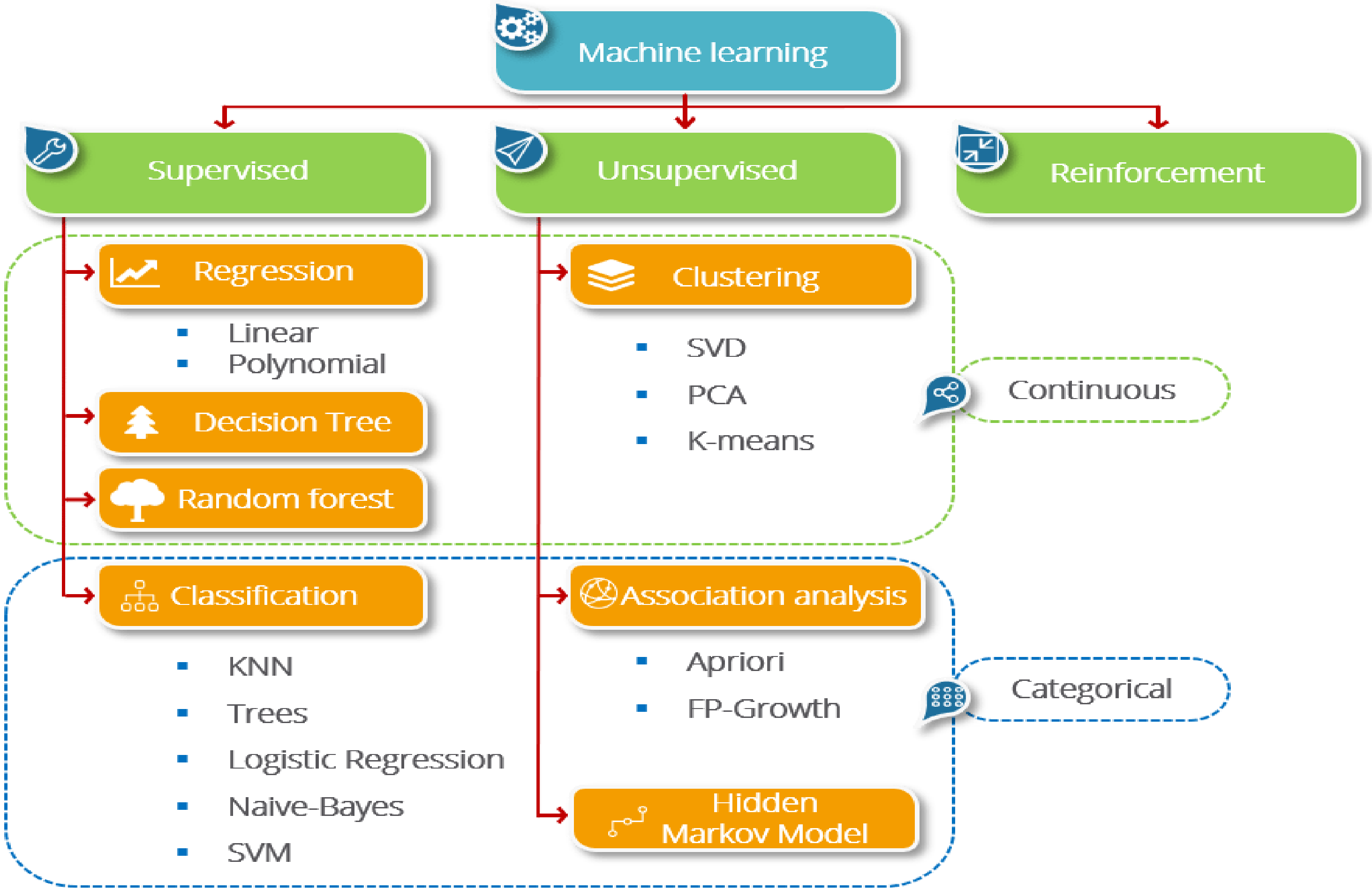
Neural Networks Leaning
Definition



- ***Learning***:
  - To start this process, the initial weights are ***chosen randomly***. Then the learning begins.
  - The neural network processes the data in the "training set" one at a time, using the weights and functions in the hidden layers, then compares the resulting outputs against the desired outputs.
- ***Cost (Errors)*** are then ***propagated back through the system***, causing the system to adjust the weights for application to the next data.
- This process occurs repeatedly as the weights are tweaked. During the learning phase of a network, the same set of data is processed many times as the connection weights are continually refined.

Source: Microsoft Social Technet

NN Categories & Types
Categories



 Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Source: Edureka: What is Machine Learning

Neural Networks & Timeseries
Introduction

- ML techniques have been ***gaining prominence*** over time as interest in AI has been rising.
- ML methods have been as alternatives to statistical ones
  - Only scant evidence is available about their performance in terms of accuracy and computational requirements.
- ***Traditional statistical methods are considerably more accurate*** that ML ones.
  - ML methods must become: more accurate, less time demanding, less of a black box.
  - Best fitted available data ***does not necessarily result*** in ore accurate predictions.

Source: *"The accuracy of Machine Learning (ML) Forecasting methods versus Statistical ones: Extending the results of the M3-Competition", Makridakis et. al., March 2018*

Neural Networks & Timeseries
Algorithmic Approach

- A different algorithmic approach is required.
- One of the most important properties an algorithm needs in order to be considered a time-series algorithm is the ability to ***extrapolate patterns outside of the domain of training data***.
- Many machine learning algorithms do not have this capability, as they tend to be restricted to a domain that is defined by training data.
  - They also lack **interpretability**, which is crucial to business leaders who want to make data-driven decisions.
- Therefore, they are not suited for time series, as the objective of time series is to project into the future.

Source: Towards Data Science

Neural Networks & Timeseries
Train & Retrain

- On NN we train a model, test it, retrain it if necessary until we have satisfactory results, and then evaluate it on a hold out data set. If we're satisfied with the performance, we then ***deploy it to production***.
  - Once in production, we score new data as it comes in. Eventually after a few months, we might want to update the model if a significant amount of new training data comes in.
  - Model training ***is a one time activity,*** or done at most at periodic intervals to maintain the model's performance.
    - Example: ***Classifying cat images***. The visual properties of cats are stable over time. Given enough data, the model we trained this week is good enough for the foreseeable future as well.

BOSCH
Source: Towards Data Science

Neural Networks & Timeseries
Train & Retrain

- On Statistical Methods ***this is not the case***!
  - We have to ***retrain our model every time*** we want to generate a new forecast, or new actuals come in.
  - The development data set and the production data set are not the same (real world).
  - We are actually retraining a new model from scratch every time we want to generate a new forecast.

➢ ***In practice***: deploying forecasting algorithms to production is very different from deploying NN models.

BOSCH

Source: Towards Data Science

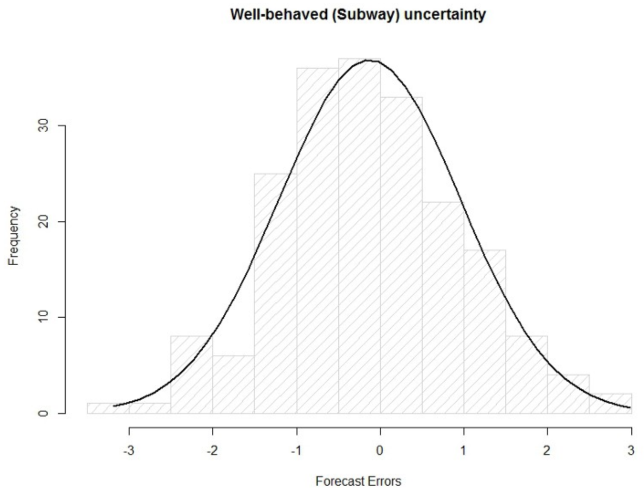
Neural Networks & Timeseries
Test/Train Split

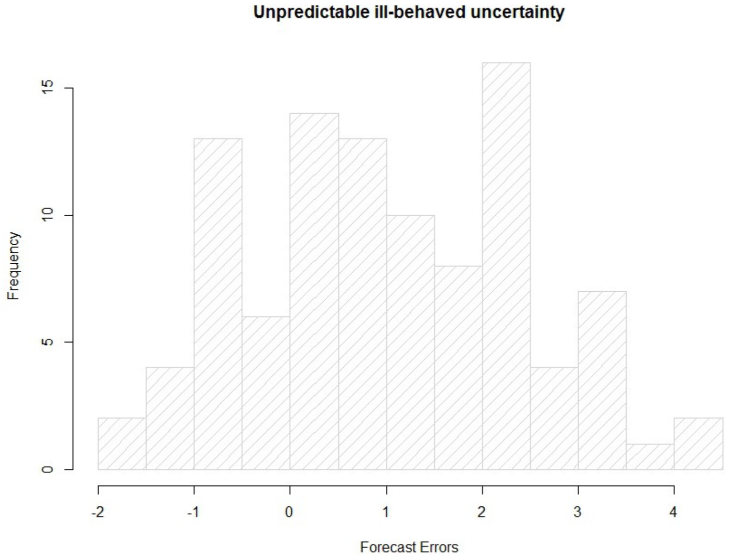
- NN require ***enough data*** to use (train & test) for building a model. But time series data is often very small compared to the data sets used in image processing or NLP.
  - Example: Two years of weekly data are 104 points, 10 years of quarterly data are 40 points.
- With data sets as small as this, we don’t have the luxury of setting aside 20% or 30% of the data for testing purposes.
- Cross validation can’t help, since it is not applicable at all.
  - It doesn’t make sense to try to forecast values for February using training data from January and March. That would amount to leakage.

BOSCH
Source: Towards Data Science

Neural Networks & Timeseries
Uncertainty

- We can reasonably expect that an NN (for image classification or for NLP problem) can eventually classify all new incoming examples ***accurately*** (given enough training data).
- This is usually ***not the case in business forecasting applications***, where forecasts are almost always going to be not accurate.
  - In demand forecasting and inventory applications, ***the uncertainty of forecast is crucial for the applications*** that consume the forecast.
  - The uncertainty of your forecast (represented by forecast intervals or by forecast quantiles) is what we will use to calculate safety stock.



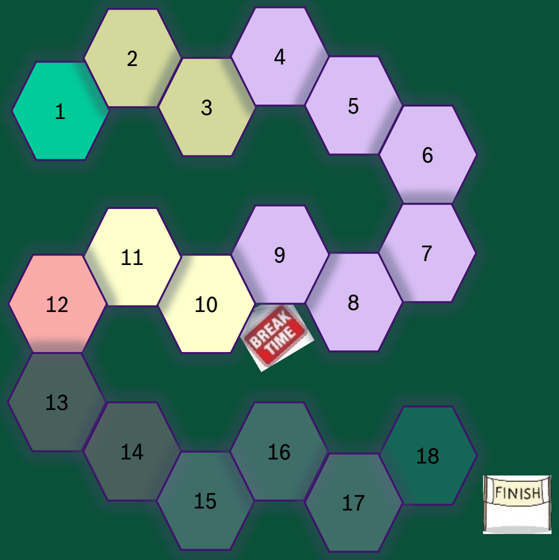


BOSCH

Source: Towards Data Science

12. SPECIAL TOPIC

WORKING DAYS ADJUSTMENT

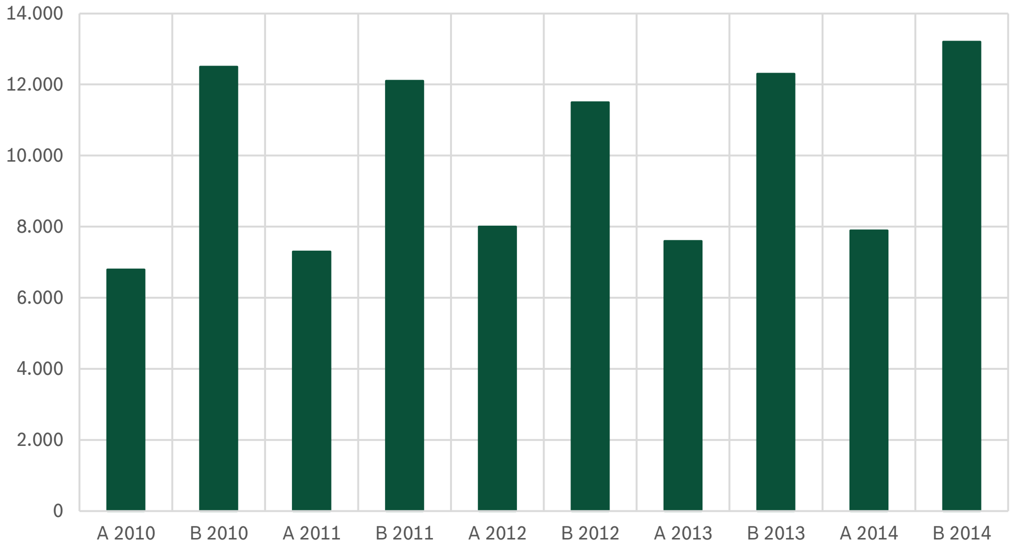


Working Days Adjustment
Definition

An adjustment could be made, by taking into account ***the number of working days*** in each period.

- Define the number of working days in each period.
- Define the country and find the holidays.
- Adjust the data point values, by using:

$$Y_t' = Y_t \times \frac{number\ of\ average\ trading\ days\ in\ all\ periods}{number\ of\ trading\ days\ in\ current\ period\ (t)}$$

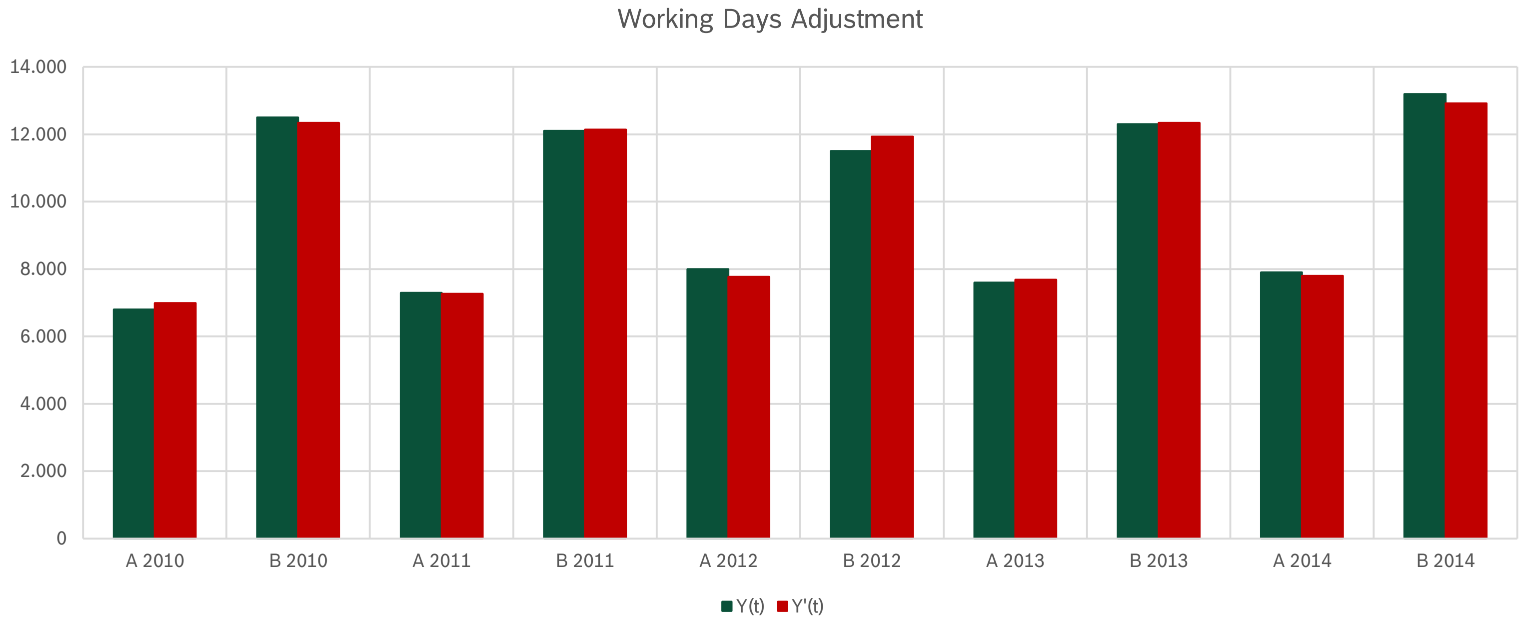


BOSCH

Working Days Adjustment
Applying Adjustment

Example:

	Y(t)	WD	Y'(t)
A 2010	6.800	118	6.996
B 2010	12.500	123	12.337
A 2011	7.300	122	7.264
B 2011	12.100	121	12.140
A 2012	8.000	125	7.770
B 2012	11.500	117	11.932
A 2013	7.600	120	7.689
B 2013	12.300	121	12.341
A 2014	7.900	123	7.797
B 2014	13.200	124	12.923
Average Working Days		**121,4**	

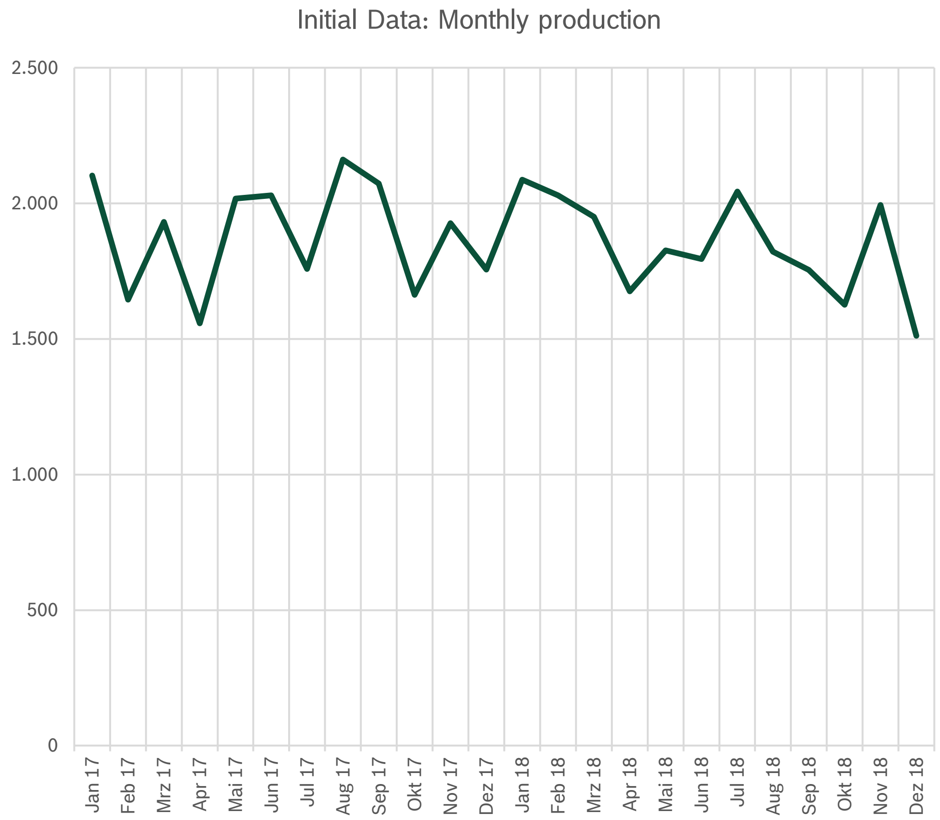


$$Y_1' = Y_1 \times \frac{Average}{WD1} = 6800 \times \frac{121{,}4}{118} = 6996$$

BOSCH

Working Days Adjustment
Example: Observe timeserie

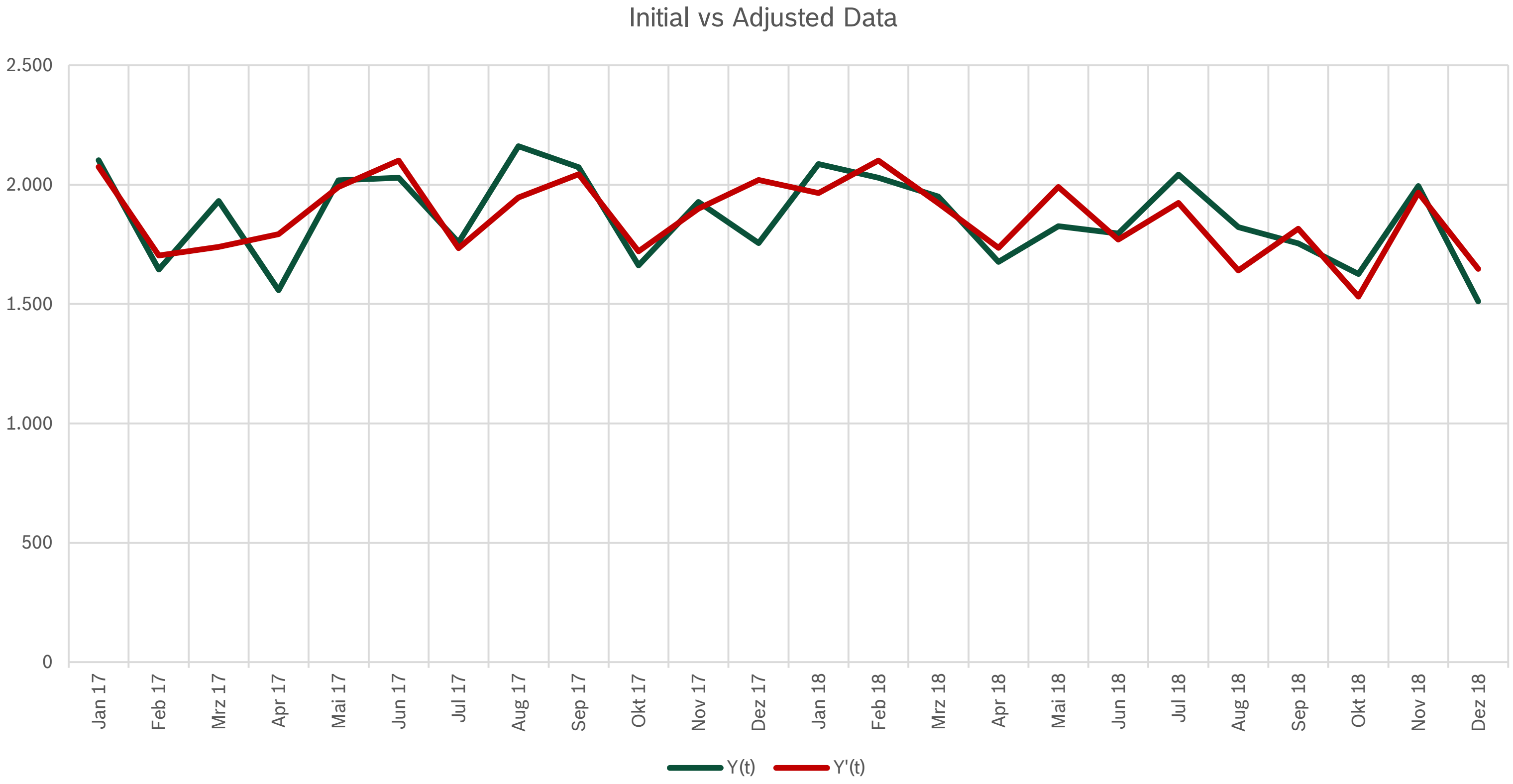
	Y(t)
Jan 17	2.103
Feb 17	1.645
Mrz 17	1.932
Apr 17	1.558
Mai 17	2.018
Jun 17	2.029
Jul 17	1.758
Aug 17	2.161
Sep 17	2.073
Okt 17	1.662
Nov 17	1.927
Dez 17	1.756
Jan 18	2.087
Feb 18	2.029
Mrz 18	1.950
Apr 18	1.676
Mai 18	1.826
Jun 18	1.795
Jul 18	2.043
Aug 18	1.822
Sep 18	1.754
Okt 18	1.626
Nov 18	1.994
Dez 18	1.511



- What you can observe?
  - 24 data points, Monthly data
  - No trend?
  - Production data
  - We need to adjust them on working days
    - Find working days per month for the specific country / region.

Working Days Adjustment
Applying Adjustment

	Y(t)	WD	Y'(t)
Jan 17	2.103	21	2.074
Feb 17	1.645	20	1.703
Mrz 17	1.932	23	1.740
Apr 17	1.558	18	1.792
Mai 17	2.018	21	1.990
Jun 17	2.029	20	2.101
Jul 17	1.758	21	1.734
Aug 17	2.161	23	1.946
Sep 17	2.073	21	2.044
Okt 17	1.662	20	1.721
Nov 17	1.927	21	1.900
Dez 17	1.756	18	2.020
Jan 18	2.087	22	1.964
Feb 18	2.029	20	2.101
Mrz 18	1.950	21	1.923
Apr 18	1.676	20	1.735
Mai 18	1.826	19	1.990
Jun 18	1.795	21	1.770
Jul 18	2.043	22	1.923
Aug 18	1.822	23	1.640
Sep 18	1.754	20	1.816
Okt 18	1.626	22	1.531
Nov 18	1.994	21	1.966
Dez 18	1.511	19	1.647
Average Working Days		20,71	



$$Y_1' = Y_1 \times \frac{Average}{WD1} = 2103 \times \frac{20{,}71}{21} = 2074$$

Working Days Adjustment
Practical Exercise

- Exercise steps:
  - Open file “**Forecasting Example - 72c - Working Days Adjustments (Exercise).xlsx**”
  - Fill in your own Monthly data
    - Random input
    - A real example
  - Estimate adjusted data

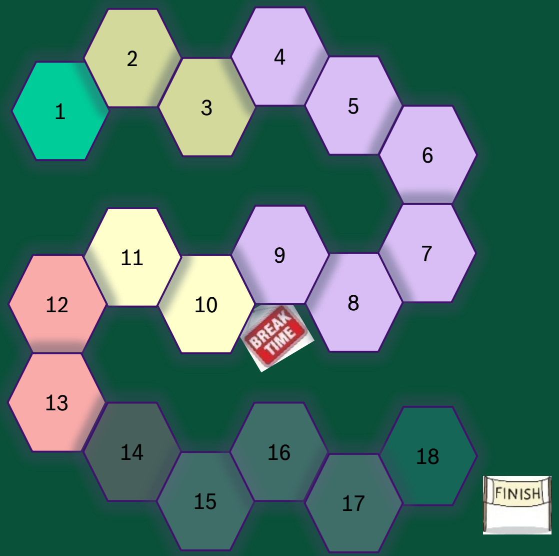
- Duration:
  - **5 minutes**

- Your Goal:
  - Get familiar with adjustment
  - Observe how values are estimated

- Who wants to present his exercise results?

13. SPECIAL TOPIC

STRUCTURAL CHANGE DETECTION



BOSCH

Structural Changes Detection
Definition

- ***Significant changes*** can affect a timeserie trend and values.
  - Example: population growth, change of demand of product or service, other products entering a market
  - This must be taken into account in all business activities, which are focusing ***in minimizing potential costs*** or ***maximizing profits*** by taking advantage of new opportunities.
- Companies face a emerging need to make decisions that will interrupt the normal production operations and will restrict existing resources and assets.
- ***Tracking of timeseries*** scope is to automatically ***detect significant changes*** in the pattern of a timeserie (for example sudden increases or decreases) as new data become available, ***as soon as possible***, and with the ***least possible false detections***.

Structural Changes Detection
Definition

- Important to use **automated tracking** applications with ***low cost***.
- These applications must label the timeseries correctly as “no-change” and “significant change”, in order to alert a company to make the necessary business actions.
- Timeseries tracking is directly linked to forecasting:
  - ***change detection is mainly based on the forecasting error***, which resulted from an extrapolation of the timeserie in the future.
    - Several surveys have studied the relationship between the accuracy of timeseries forecasting and the successfully detection of significant pattern changes in a timeserie.
  - tracking timeseries is (or it should be) a ***key component*** in business forecasting systems installed in a company, since business intervention required when simple predictive methods leads to large prediction errors.

Structural Changes Detection
Definition

- The main objective of the forecast is:
  - an estimate of future average value, taking into account the timeserie trend through appropriate methods and parameters and
  - assuming that within the forecast horizon a significant (structural) change in the data pattern will not take place.
- When a relatively ***large forecasting error*** is observed, or when a ***sequence of forecasting errors*** have the same direction (under or over forecasted), then the forecasts are biased.
  - This can be the result of a change in the timeserie pattern, which requires further analysis and control.
- The exponential smoothing methods will adjust their forecasts with a delay of one period, after the completion of the pattern change. But, the estimation errors in the interim periods are enough to detect the structural change on the timeserie data pattern.

Structural Changes Detection
Brown Tracking Method

- The ***Brown method*** of *k-periods* for a given period *n* is calculated as the absolute ration between:
  - the sum of forecasts errors of period *n* and the *k-1* previous periods, and
  - the exponential smoothing average deviation of the absolute errors at period *n*. *(β=smoothing parameter)*

$$Brown_k = \left|\frac{CUSUM_n^k}{MAD_n}\right| = \left|\frac{\sum_{t=n-k+1}^{n} e_t}{(1-\beta)^n * e_0 + \sum_{t=1}^{n} \beta * (1-\beta)^{n-t} * |e_t|}\right|,$$

  - e_0 is the initialization point for MAD and can be calculated by using the following formula:

$$e_0 = \frac{\sum_{t=1}^{6} |e_t|}{6}$$

Structural Changes Detection
Trigg Tracking Method

- **Trigg method** is calculated by replacing the CUSUM index in the numerator, with a sum of exponentially smoothing forecasting errors E_n.
  - the numerator exponential smoothing allows the index to apply lower weight to forecast errors, as the time lag between the current periods.

$$Trigg_k = \left|\frac{E_n}{MAD_n}\right| = \left|\frac{\sum_{t=1}^{n} \alpha * (1-\alpha)^{n-t} * e_t}{(1-\beta)^n * e_0 + \sum_{t=1}^{n} \beta * (1-\beta)^{n-t} * |e_t|}\right|$$

  - This change causes the pointer to reset automatically, after detecting the pattern change.

- It should be noted that the Trigg method with *a = 1* is identical to Brown method when *k = 1*.

Structural Changes Detection
Selecting smoothing parameter

- How to select the ***α*** and ***β*** parameters:
  - Gardner (1983) and Golden and Settle (1976) have originally proposed to use the same value, ***equal to the optimal smoothing parameter*** for the exponential smoothing method of constant level (SES).
  - McClain (1988) showed that the selection of different values for α and β has generally better results, suggesting ***lower values for β***.
  - In practice:

$$0.05 \leq \alpha \leq 1$$

$$0.05 \leq \beta \leq 0.5$$

$$\beta \leq \alpha$$

Structural Changes Detection
Tracking the change

- The detection of significant changes, according to Brown and Trigg methods, can be achieved for the period where the calculated index value ***exceeds a certain activation limit*** (threshold).
- The change of the threshold value implies a modification of the method sensitivity.
- In practice, the following activation rules are usually applied, depending on the selected tracking method:

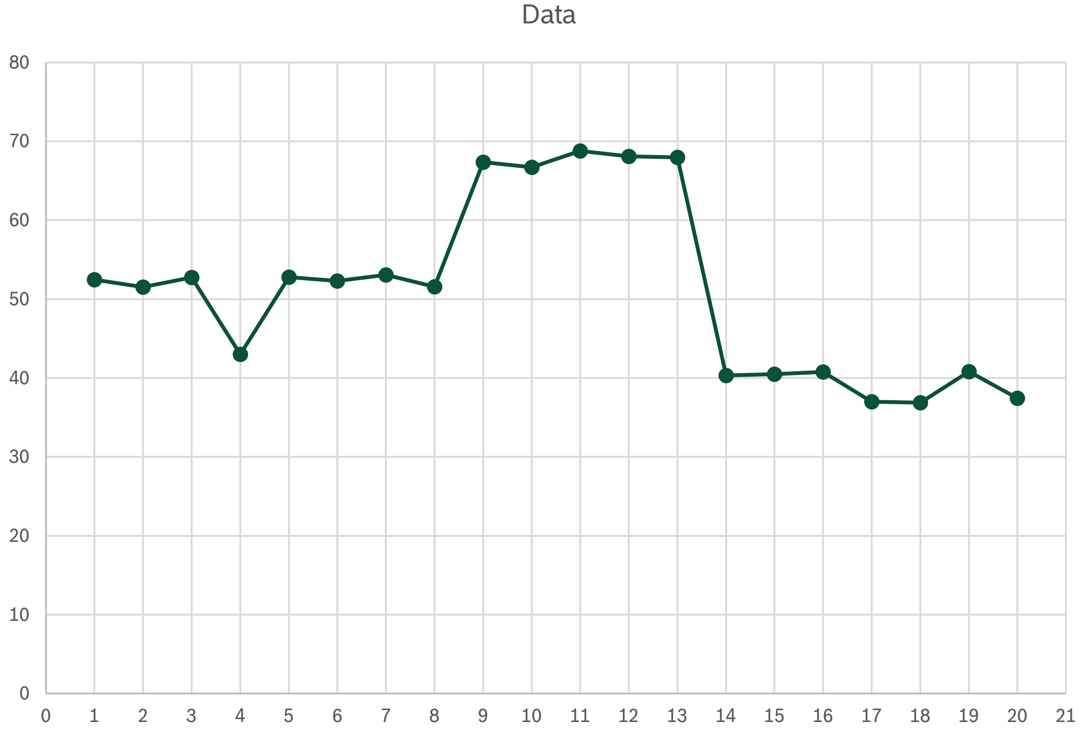
$$if\ Brown_k > 0.5\ then\ \text{"}Structural\ Change\ Detected\text{"}$$

$$if\ Trigg > 0.5\ then\ \text{"}Structural\ Change\ Detected\text{"}$$

Structural Changes Detection
Example

- Initial Data.

Time (t)	Data
1	52,474
2	51,543
3	52,749
4	43,000
5	52,806
6	52,298
7	53,092
8	51,583
9	67,362
10	66,731
11	68,770
12	68,079
13	67,968
14	40,321
15	40,491
16	40,790
17	36,985
18	36,897
19	40,801
20	37,441

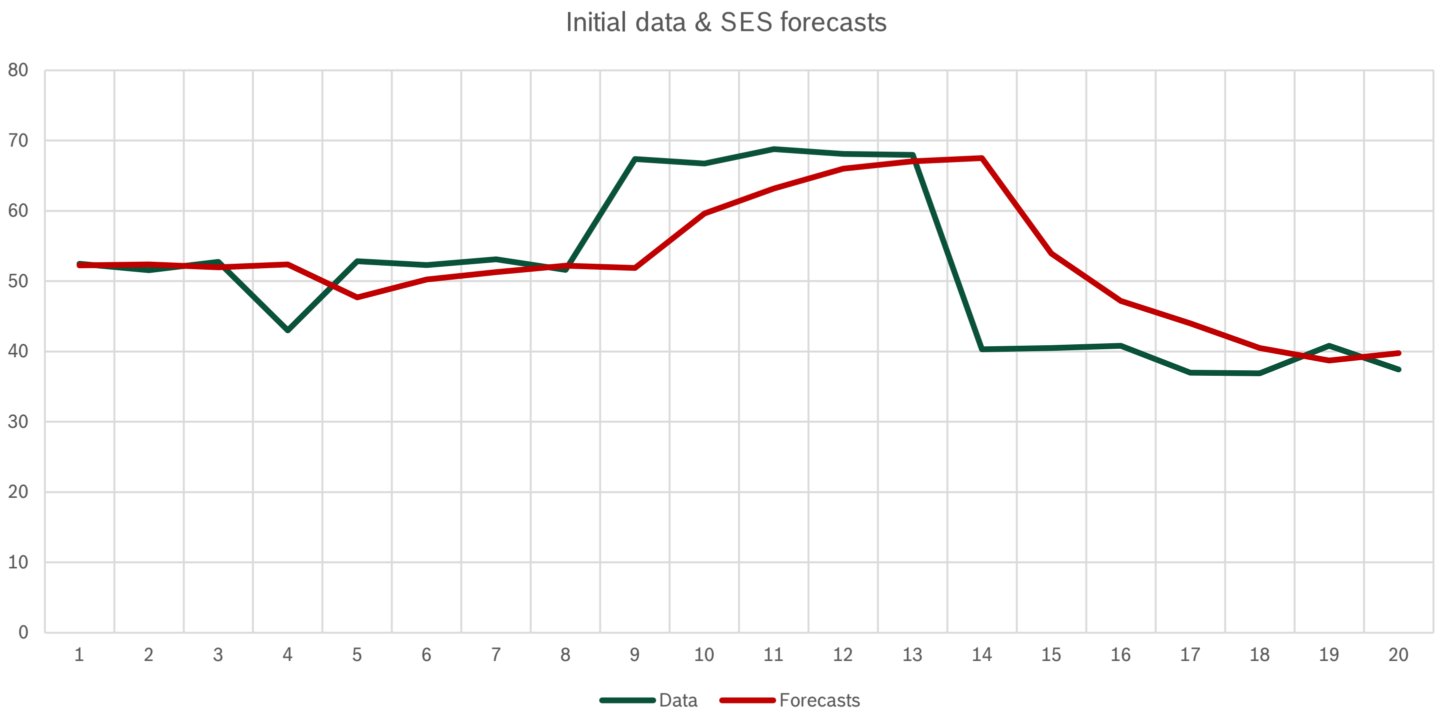


- What you can observe?
  - 20 data points
  - Big increase on t=9
    - The increase remains until t=13
  - Big decrease on t=14
    - The decrease remains until the end of the timeserie.

Structural Changes Detection
Example

- ***Step 1*****:** Apply simple exponential smoothing for estimating forecasts.

S(0) =	52,255	α=0.5		
Time (t)	**Data**	**Forecasts**	**Error e**	**Level S**
1	52,474	52,255	0,219	52,365
2	51,543	52,365	-0,822	51,954
3	52,749	51,954	0,795	52,351
4	43,000	52,351	-9,351	47,676
5	52,806	47,676	5,130	50,241
6	52,298	50,241	2,057	51,269
7	53,092	51,269	1,823	52,181
8	51,583	52,181	-0,598	51,882
9	67,362	51,882	15,480	59,622
10	66,731	59,622	7,109	63,176
11	68,770	63,176	5,594	65,973
12	68,079	65,973	2,106	67,026
13	67,968	67,026	0,942	67,497
14	40,321	67,497	-27,176	53,909
15	40,491	53,909	-13,418	47,200
16	40,790	47,200	-6,410	43,995
17	36,985	43,995	-7,010	40,490
18	36,897	40,490	-3,593	38,694
19	40,801	38,694	2,107	39,747
20	37,441	39,747	-2,306	38,594



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Structural Changes Detection
Example

▶ ***Step 2***: Estimate Trigg index.

				Trigg		
Time (t)	Data	Forecasts	Error e	Nominator	Denominator	Index
1	52,474	52,255	0,219	0,044	2,492	**0,018**
2	51,543	52,365	-0,822	-0,129	2,159	0,060
3	52,749	51,954	0,795	0,056	1,887	0,030
4	43,000	52,351	-9,351	-1,826	3,380	0,540
5	52,806	47,676	5,130	-0,435	3,730	0,117
6	52,298	50,241	2,057	0,064	3,395	0,019
7	53,092	51,269	1,823	0,415	3,081	0,135
8	51,583	52,181	-0,598	0,213	2,584	0,082
9	67,362	51,882	15,480	3,266	5,163	0,633
10	66,731	59,622	7,109	4,035	5,552	0,727
11	68,770	63,176	5,594	4,347	5,561	0,782
12	68,079	65,973	2,106	3,898	4,870	0,800
13	67,968	67,026	0,942	3,307	4,084	0,810
14	40,321	67,497	-27,176	-2,790	8,702	0,321
15	40,491	53,909	-13,418	-4,915	9,645	0,510
16	40,790	47,200	-6,410	-5,214	8,999	0,579
17	36,985	43,995	-7,010	-5,573	8,601	0,648
18	36,897	40,490	-3,593	-5,177	7,599	0,681
19	40,801	38,694	2,107	-3,720	6,501	0,572
20	37,441	39,747	-2,306	-3,438	5,662	0,607

$$Trigg_k = \left|\frac{E_n}{MAD_n}\right| = \left|\frac{\sum_{t=1}^{n} \alpha * (1-\alpha)^{n-t} * e_t}{(1-\beta)^n * e_0 + \sum_{t=1}^{n} \beta * (1-\beta)^{n-t} * |e_t|}\right|$$

$$e_0 = \frac{\sum_{t=1}^{6}|e_t|}{6} = 3{,}06$$

$$E_1 = \sum_{t=1}^{1} a \times (1-a)^{1-t} \times e_t = 0{,}2 \times (1-0{,}2)^{1-1} \times e_1 = 0{,}044$$

$$MAD_1 = (1-\beta)^1 \times e_0 + \sum_{t=1}^{1} \beta \times (1-\beta)^{1-t} \times |e_t| = 0{,}8 \times 3{,}06 + 0{,}2 \times (1-0{,}2)^0 \times |0{,}219| = 2{,}492$$

$$Trigg_1 = \left|\frac{E_1}{MAD_1}\right| = \left|\frac{0{,}044}{2{,}494}\right| = \mathbf{0{,}018}$$

$$E_2 = \sum_{t=1}^{2} a \times (1-a)^{2-t} \times e_t = 0{,}2 \times (1-0{,}2)^{2-1} \times e_1 + 0{,}2 \times (1-0{,}2)^{2-2} \times e_2 = -0{,}129$$

$$MAD_2 = (1-\beta)^2 \times e_0 + \sum_{t=1}^{2} \beta \times (1-\beta)^{2-t} \times |e_t| =$$

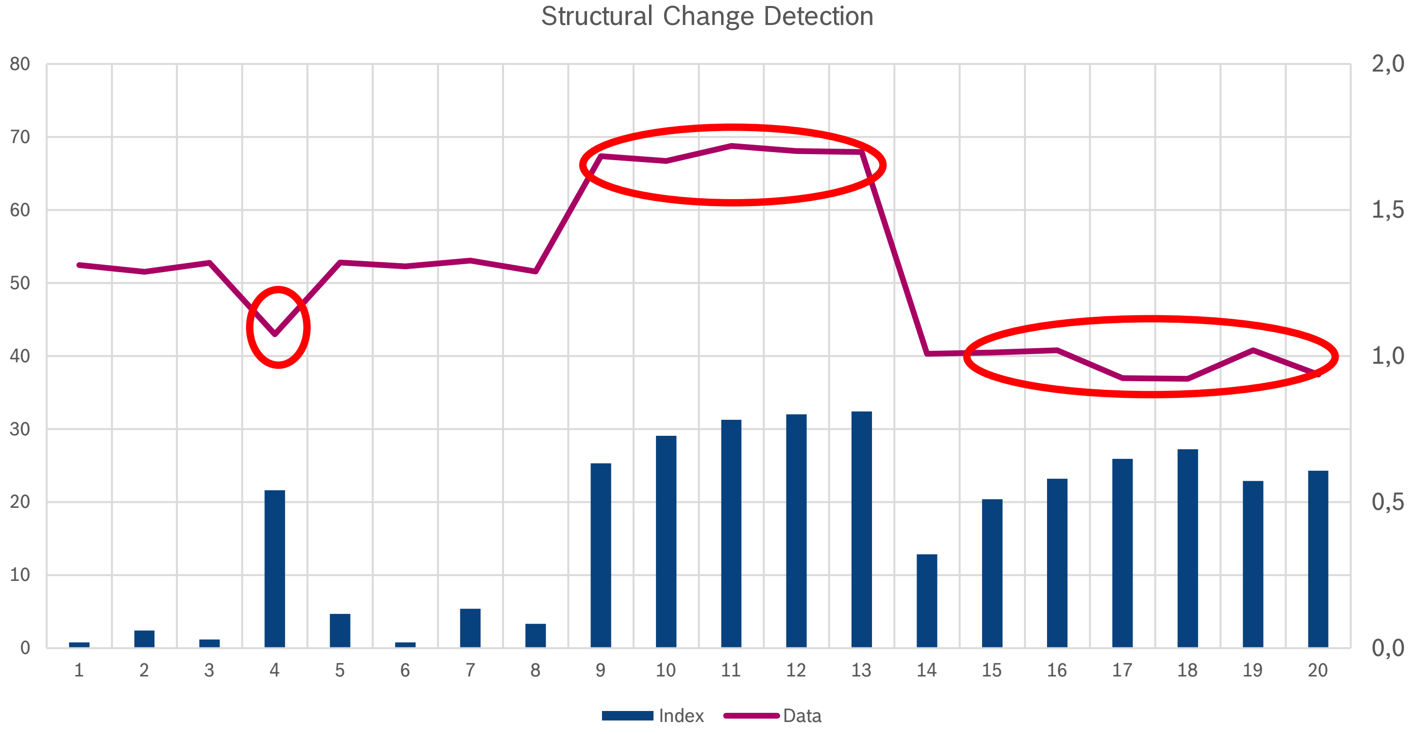
$$= 0{,}8^2 \times 3{,}06 + 0{,}2 \times (1-0{,}2)^1 \times |0{,}219| + 0{,}2 \times (1-0{,}2)^0 \times |-0{,}822| = 2.159$$

$$Trigg_2 = \left|\frac{E_2}{MAD_2}\right| = \left|\frac{-0{,}129}{2{,}159}\right| = \mathbf{0{,}060}$$

Structural Changes Detection
Example

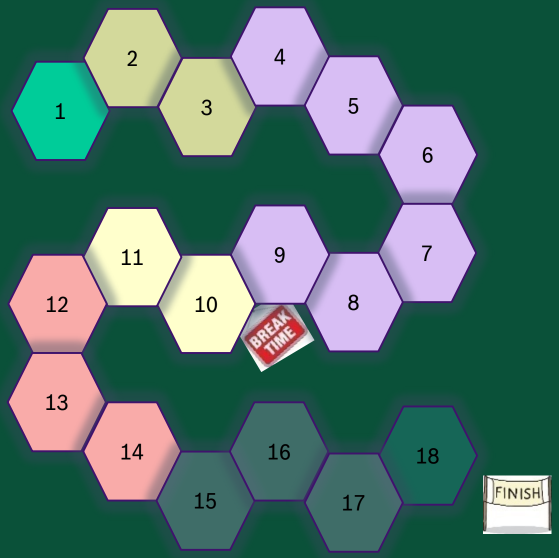
▶ ***Step 3***: Locate structural Changes (Trigg > 0,5)

Time (t)	Data	Index	Str.Change
1	52,474	0,018	
2	51,543	0,060	
3	52,749	0,030	
4	43,000	0,540	yes
5	52,806	0,117	
6	52,298	0,019	
7	53,092	0,135	
8	51,583	0,082	
9	67,362	0,633	yes
10	66,731	0,727	yes
11	68,770	0,782	yes
12	68,079	0,800	yes
13	67,968	0,810	yes
14	40,321	0,321	
15	40,491	0,510	yes
16	40,790	0,579	yes
17	36,985	0,648	yes
18	36,897	0,681	yes
19	40,801	0,572	yes
20	37,441	0,607	yes



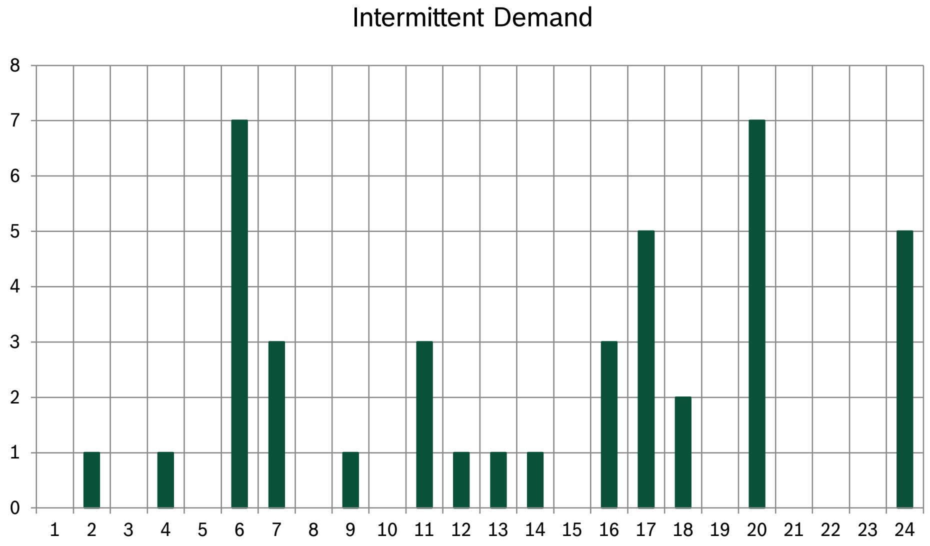
14. SPECIAL TOPIC

INTERMITTED DEMAND



Intermittent Demand
Introduction

- ***Intermittent (or sporadic) demand***: is a very common problem in business analysis.
  - It includes periods where ***demand is zero***. When demand occurs, the size varies significantly. Items with intermittent, or “slow-moving”, demand have many zero values interspersed with random spikes of non-zero demand.
- This irregular demand pattern is impossible to forecast with traditional, smoothing-based forecasting methods.
- The difficulty of forecasting lies:
  - not only in the ***discontinuity*** of demand, but
  - also in the ***large variation*** between two consecutive non-zero observations.



Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Intermittent Demand
Introduction

- This problem is especially prevalent in companies that ***manage large inventories*** of service/spare parts in industries such as:
  - aerospace, automotive, high tech, and electronics,
  - in MRO (Maintenance, Repair and Overhaul) organizations,
  - companies manufacturing high-priced capital goods,
  - stock keeping units and spare parts.
- In these businesses, as much as 70% of the parts and product items may have intermittent demand.
- The intermittent demand pattern makes it ***difficult*** to accurately ***estimate the safety stock*** and service level inventory requirements needed for successful supply chain planning.

Intermittent Demand
Introduction

- Firms make two common mistakes when forecasting intermittent demand:
  - they focus on estimates of per period demand when they should really focus on estimates of ***the inventory stocking requirements*** necessary to meet their desired service levels.
  - they use forecasting methods that are inappropriate for intermittent demand.
- Traditional forecasting methods fail because they try to identify recognizable patterns in the demand data, such as trend and seasonality. However, intermittent demand data don't exhibit such regular patterns and tend to be characterized by a preponderance of zero values.
- What to use?
  - Croston method
  - Syntetos & Boylan approximation - SBA
  - Aggregate-Disaggregate intermittent demand approach - ADIDA

Intermittent Demand
Croston

- ***Croston*** (1972): An alternative method, which takes into account both the ***volume of demand*** and the ***interval*** between non-zero observations.
- The logic behind the method, is to separate the initial timeserie into two (2) other timeseries:
  - the first is the ***intervals between the demands***, and
  - the other by the ***number of independent non-zero demands***.
- Croston estimated forecasts by independently applying simple exponential smoothing in:
  - non-zero values of the timeserie, and
  - in the intervals between non-zero values of the timeserie.

Intermittent Demand
Croston

- The two timeseries are extrapolated independently, using the ***constant level exponential smoothing*** method. In the literature, it is common to use the value of 0.05 as smoothing parameter.
- The final forecast is estimated by using the following formula:

$$DemandForecast = \frac{VolumeForecast}{Interval\ Forecast}$$

  - *Interval Forecast:* is the exponentially smoothed (or moving average) inter-demand interval, updated only if demand occurs in period.
  - *Volume Forecast:* is the exponentially smoothed (or moving average) size of demand, updated only if demand occurs in period.

➢ *Keep in mind: Other forecasting methods can also be used.*

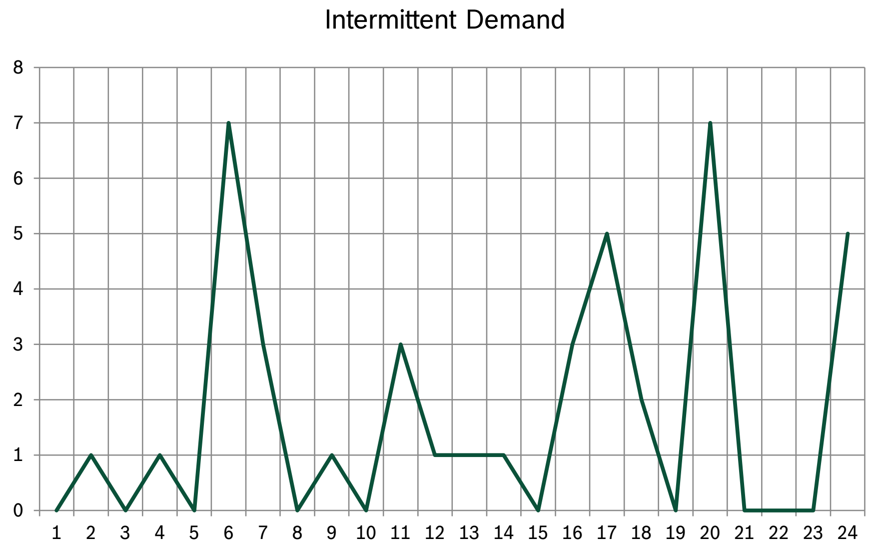
Intermittent Demand
Croston

- In general, intermittent demand can ***lead to increased levels*** of supply stocks, by making bias in the estimates of average demands.
- Croston however, argues that:
  - the approach of separate estimates for demand and time between demands, can lead to a reduction of prejudice (bias).
  - the additional component of frequency between the demands, allows the stock administrator to define inventory ***orders and costs with greater accuracy***, avoiding the over-stocking.

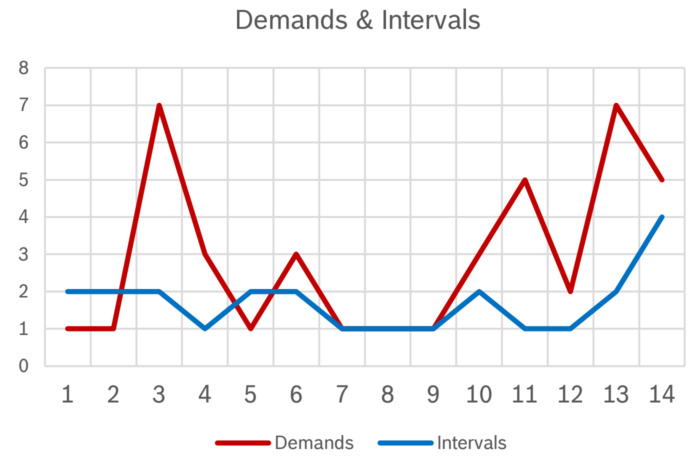
Intermittent Demand
Croston: Example

- Step 1: separate the initial timeserie into two (2) other timeseries

Time (t)	Intermittent Demand
1	0
2	1
3	0
4	1
5	0
6	7
7	3
8	0
9	1
10	0
11	3
12	1
13	1
14	1
15	0
16	3
17	5
18	2
19	0
20	7
21	0
22	0
23	0
24	5



Demands	Intervals
1	2
1	2
7	2
3	1
1	2
3	2
1	1
1	1
1	1
3	2
5	1
2	1
7	2
5	4

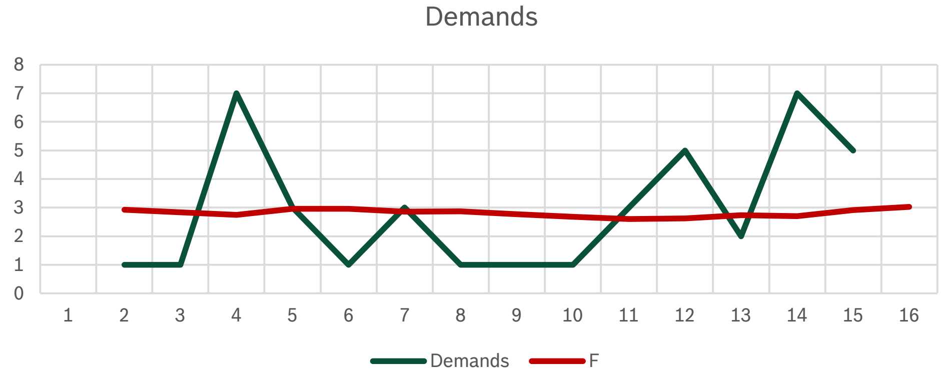


$S_0 = \overline{Demands}$

Intermittent Demand
Croston: Example

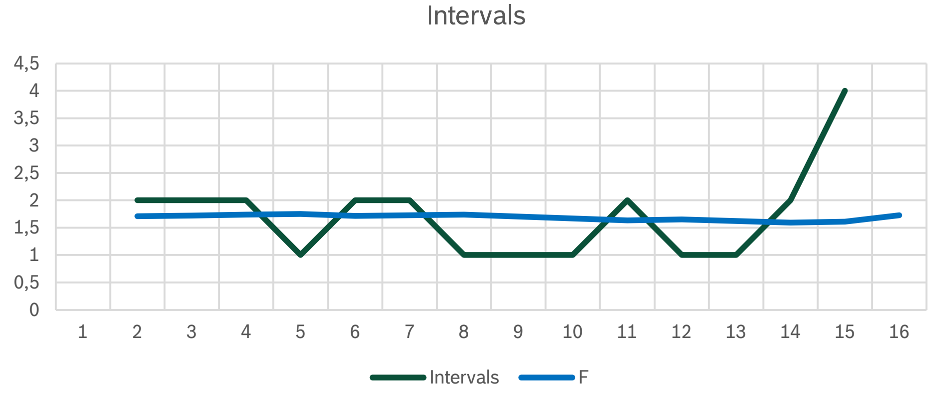
- Step 2: Use SES for extrapolation of each timeserie

Demands	F	e	S
			2,93
1	2,93	-1,93	2,83
1	2,83	-1,83	2,74
7	2,74	4,26	2,95
3	2,95	0,05	2,96
1	2,96	-1,96	2,86
3	2,86	0,14	2,87
1	2,87	-1,87	2,77
1	2,77	-1,77	2,68
1	2,68	-1,68	2,60
3	2,60	0,40	2,62
5	2,62	2,38	2,74
2	2,74	-0,74	2,70
7	2,70	4,30	2,92
5	2,92	2,08	3,02
	3,02		



$$S_0 = \overline{Demands}$$

Intervals	F	e	S
			1,71
2	1,71	0,29	1,72
2	1,72	0,28	1,74
2	1,74	0,26	1,75
1	1,75	-0,75	1,71
2	1,71	0,29	1,73
2	1,73	0,27	1,74
1	1,74	-0,74	1,70
1	1,70	-0,70	1,67
1	1,67	-0,67	1,64
2	1,64	0,36	1,65
1	1,65	-0,65	1,62
1	1,62	-0,62	1,59
2	1,59	0,41	1,61
4	1,61	2,39	1,73
	1,73		



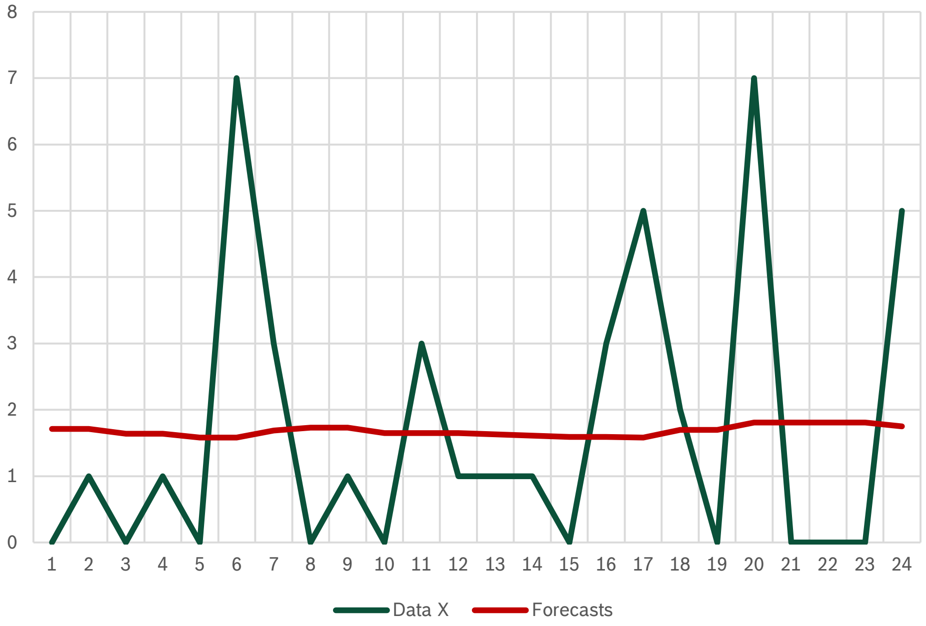
$$S_0 = \overline{Intervals}$$

Intermittent Demand
Croston: Example

- Step 3: Combine the forecasts

Forecasts		
Demands	**Intervals**	**Croston**
2,93	1,71	1,71
2,83	1,72	1,64
2,74	1,74	1,58
2,95	1,75	1,69
2,96	1,71	1,73
2,86	1,73	1,65
2,87	1,74	1,65
2,77	1,70	1,63
2,68	1,67	1,61
2,60	1,64	1,59
2,62	1,65	1,58
2,74	1,62	1,69
2,70	1,59	1,70
2,92	1,61	1,81
3,02	1,73	1,75

Time (t)	Data X	Forecasts
1	0	1,71
2	1	1,71
3	0	1,64
4	1	1,64
5	0	1,58
6	7	1,58
7	3	1,69
8	0	1,73
9	1	1,73
10	0	1,65
11	3	1,65
12	1	1,65
13	1	1,63
14	1	1,61
15	0	1,59
16	3	1,59
17	5	1,58
18	2	1,70
19	0	1,70
20	7	1,81
21	0	1,81
22	0	1,81
23	0	1,81
24	5	1,75



 Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

Intermittent Demand
Syntetos & Boylan

- Empirical evidence has shown that the gains from ***Croston implementation is worse than expected***, compared with simpler forecasting techniques. In some cases, even worse performance is observed.
- Syntetos & Boylan (2001) tried to identify the cause of this unexpected behavior.
  - they found that Croston method is ***positively biased***, thus has an optimistic trend in the forecasting results.
  - they managed to associate the ***level of optimistic trend with the smoothing parameter*** which used for the extrapolation of the two decomposed timeseries.
  - the maximum bias observed when the smoothing parameter has the maximum value of 1. In general, they observed large optimistic bias ***when the smoothing parameter is large***.
  - An empirical result is that Croston method must be used with a smoothing parameter ***not greater that 0.15***.

Intermittent Demand
Syntetos & Boylan

- The Syntetos and Boylan Approximation (SBA) is in fact a ***modification*** of the Croston method, where the forecasts are estimated using the formula:

$$F_{SBA} = \left(1 - \frac{a}{2}\right) \times \frac{VolumeForecast}{IntervalForecast}$$

Intermittent Demand
ADIDA

- ***Aggregate-Disaggregate Intermittent Demand Approach***: A process of non-overlapping aggregation of data in less-frequency periods.
  - in case of monthly data, a quarterly aggregation can be applied, setting the level of aggregation to 3 months.
- Advantages:
  - the discontinuity of data (due to the existence of zero values) may be reduced or eliminated.
  - the variation of the new timeserie is expected to be reduced, since a non-overlapping moving average for smoothing has been applied.
  - The proper definition of the level of aggregation will result a ***continuous demand timeserie***, without zero values, in which any forecasting technique can be applied in order to estimate forecasts for the aggregate level.

Intermittent Demand
ADIDA

- ***Four steps*** for using the ADIDA method:

1. Analysis of the initial timeserie, for selecting the appropriate level of aggregation.
2. Apply the aggregation on the initial data, for creating a new continuous timeserie.
3. Estimate forecasts, by using a predictive model.
4. Disaggregate forecasts, by using one of the following ways:
   - ***Equal weights***: Break the forecasts with similar weights. Suitable for timeseries with great randomness and no seasonal behavior.
   - ***Previous weights***: Break the forecasts by using the observed weights of the last m observations (where m is the aggregation level).
   - ***Average weights***: Break the forecasts by using the average weights (divide all observations into k groups of m observations each, where k*m is equal to the set of available observations). Indicated for timeseries with significant seasonality.

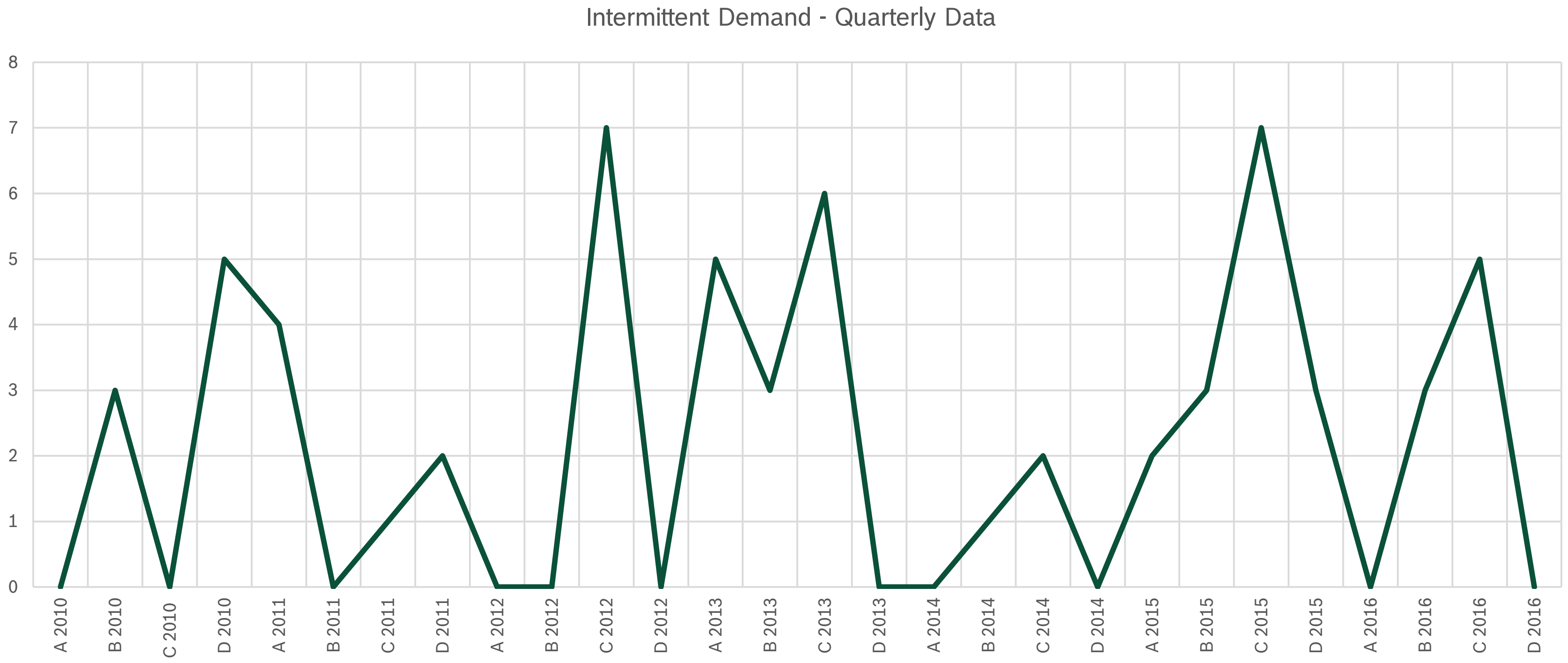
Intermittent Demand
ADIDA

- ***Characteristics***:
  - The non-overlapping data aggregation seems to be a ***promising approach*** for timeseries with interruptible demand, since forecasts in higher levels of aggregation are in general more accurate and have less variation that those in smaller aggregation levels.
  - It is common to use as aggregation level the ***number of forecasting horizon***.

- Empirical results had showed that:
  - there is probably an "optimal" level of aggregation, while the definition of the level of aggregation equal to the lead time plus a period leads to very good results regarding the predictive accuracy.

Intermittent Demand
ADIDA: Example

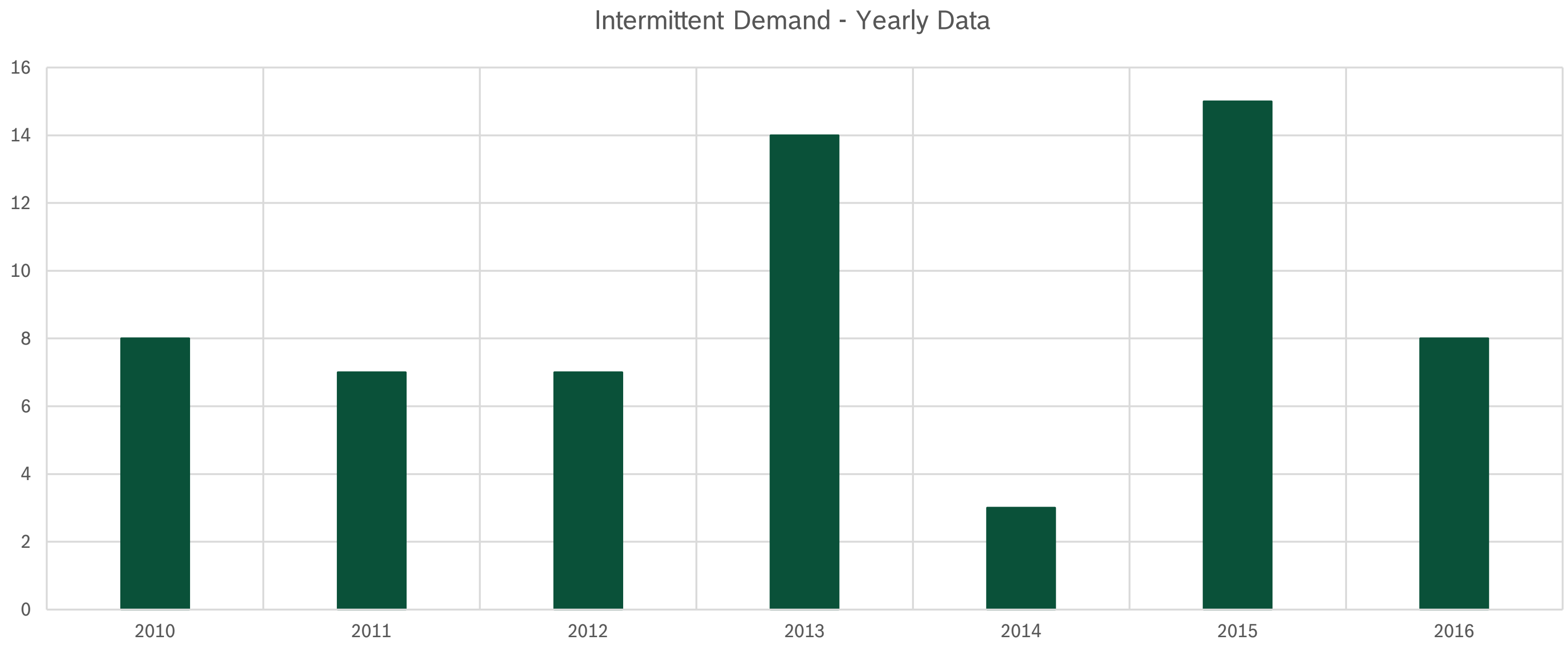
- Step 1: Level of aggregation selection.



Time (t)	Intermittent Demand
A 2010	0
B 2010	3
C 2010	0
D 2010	5
A 2011	4
B 2011	0
C 2011	1
D 2011	2
A 2012	0
B 2012	0
C 2012	7
D 2012	0
A 2013	5
B 2013	3
C 2013	6
D 2013	0
A 2014	0
B 2014	1
C 2014	2
D 2014	0
A 2015	2
B 2015	3
C 2015	7
D 2015	3
A 2016	0
B 2016	3
C 2016	5
D 2016	0

Intermittent Demand
ADIDA: Example

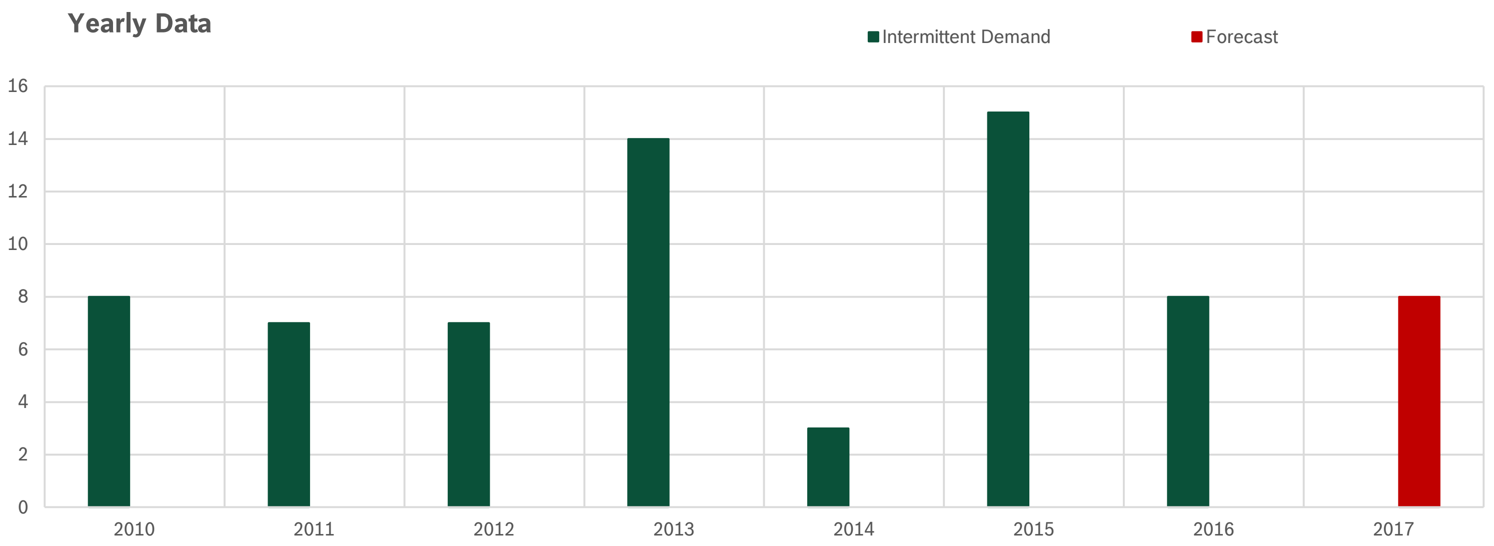
- Step 2: Apply aggregation.



Time (t)	Intermittent Demand
2010	8
2011	7
2012	7
2013	14
2014	3
2015	15
2016	8

Intermittent Demand
ADIDA: Example

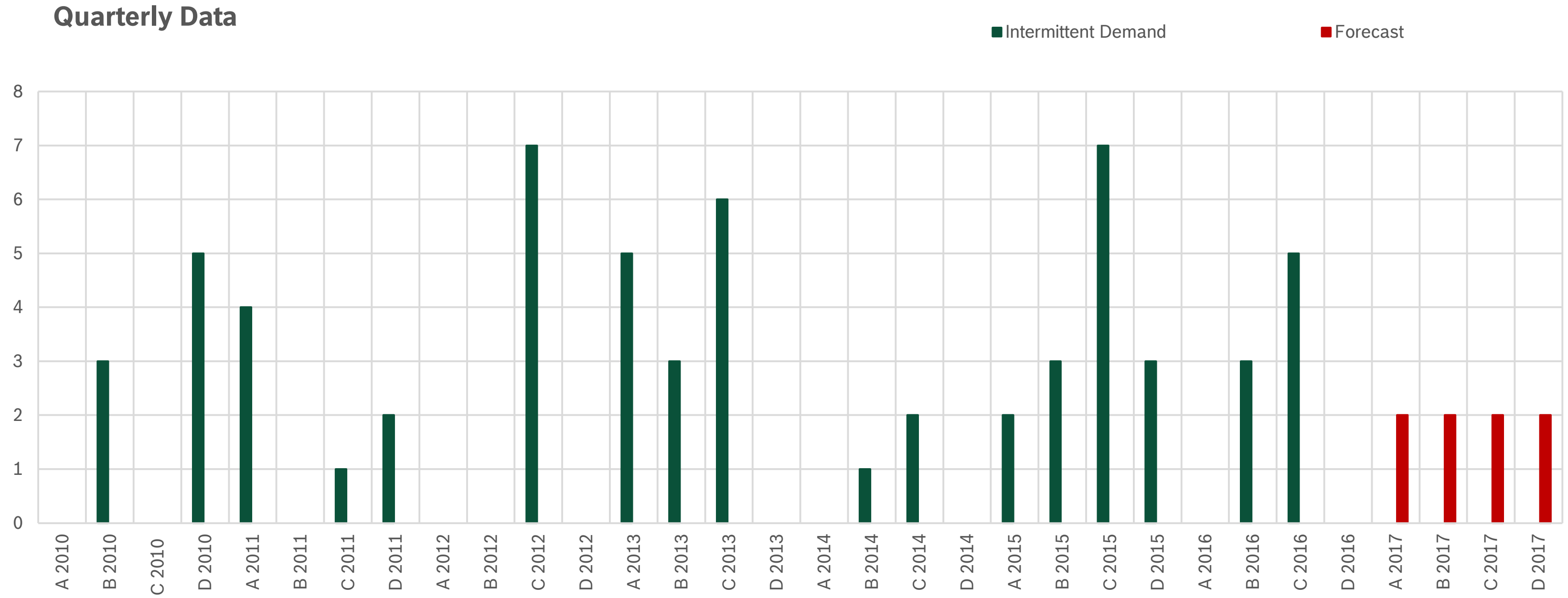
- Step 3: Estimate forecasts.
  - example: Naïve method



Time (t)	Intermittent Demand	Forecast
2010	8	
2011	7	
2012	7	
2013	14	
2014	3	
2015	15	
2016	8	
2017		8

Intermittent Demand
ADIDA: Example

- Step 4: Disaggregate forecasts.
  - example: by equal weights



Time (t)	Intermittent Demand	Forecast
A 2010	0	
B 2010	3	
C 2010	0	
D 2010	5	
A 2011	4	
B 2011	0	
C 2011	1	
D 2011	2	
A 2012	0	
B 2012	0	
C 2012	7	
D 2012	0	
A 2013	5	
B 2013	3	
C 2013	6	
D 2013	0	
A 2014	0	
B 2014	1	
C 2014	2	
D 2014	0	
A 2015	2	
B 2015	3	
C 2015	7	
D 2015	3	
A 2016	0	
B 2016	3	
C 2016	5	
D 2016	0	
A 2017		2
B 2017		2
C 2017		2
D 2017		2

BOSCH

Intermittent Demand
ADIDA: Example

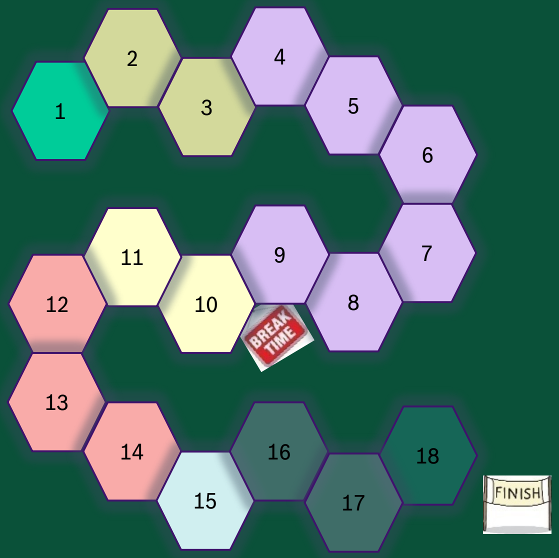
- Step 1: Level of aggregation selection.
- Step 2: Apply aggregation.
- Step 3: Estimate forecasts.
  - example: Naïve method
- Step 4: Disaggregate forecasts.
  - example: by equal weights

 Internal | Patelis Alexandros - CR/AEU2 | 2019-07-01

BOSCH

15. USE-CASE

INVENTORY DEMAND FORECASTING



BOSCH

Use-Case
Inventory demand forecasting

- ***Scope***: A grocery store needs to forecast demand for perishable items
  - Too many → discard valuable product
  - Too few → run out of stock
  - It a challenging task to ***optimise***.

- Two ways of defining a model:
  - By reducing the amount of overstocked items
  - By increasing the amount of understocked items.

BOSCH

Source: A machine learning approach to Inventory demand forecasting

Use-Case
Inventory demand forecasting

- ***Overstocked*** items:
  - A direct loss since expired items must be discarded.
  - A 10% less rate, will result 54.750 € per year.
- Similar issues:
  - Clothing store that overstocks winter coats
    - Need for bigger storage rooms.
    - Reduce limited store space.
    - Must sell coats at discount or loss.

Grocery Chain		
Stores	10	10
Products per Store	10	10
Items per product	100	100
Discard rate	**5,0%**	**4,5%**
Average cost of product	3,00 €	3,00 €
Daily loss	1.500,00 €	1.350,00 €
Annual loss	547.500,00 €	492.750,00 €

Use-Case
Inventory demand forecasting

- ***Understock Items***:
  - A more severe issue.
  - Loss of sales.
  - Lack of product may drive customers to a competitor.

- Similar issue:
  - ***understaffing a call-center*** on a busy day can result in long wait times and poor service.

Use-Case
Inventory demand forecasting

- ***Step 1***: Data
  - Collect, interpret, and analyse historical data.
  - Types of data:
    - Products on stock per day/week.
    - Sales of products per day/week.
    - Additional data (such as: store opening hours, promotion activities, weather data, holidays, etc.)
  - Volume of data:
    - At least 2 to 3 years of data, in order to monitor possible seasonal trends.

Use-Case
Inventory demand forecasting

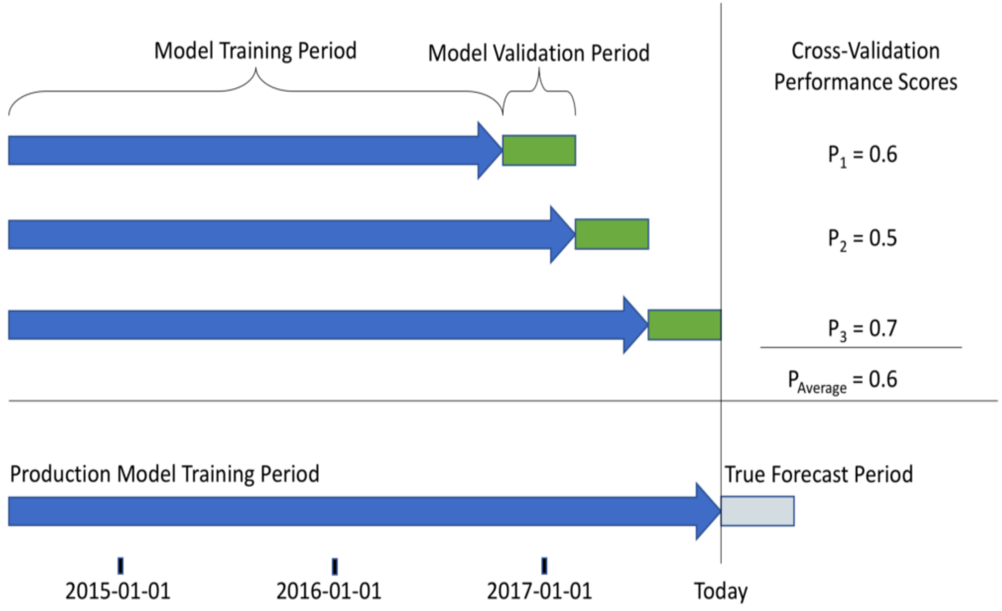
- ***Step 2***: Define the model
  - What to forecast:
    - Depends on the goal of the business (sales?, minimum overstocked? maximum understocked?).
    - In our case: ***Forecast daily sales for each product in each store***.
  - How:
    - Select a statistical method to use.
    - In our case: Exponential Smoothing.
  - Evaluation metric:
    - There is no best evaluation metric.
    - Again, it depends on the business.
    - In our case: Root Mean Square Error (RMSE).

Use-Case
Inventory demand forecasting

- ***Step 3***: Define a benchmark
  - We need a baseline score, in order to evaluate our model.
  - Common benchmark:
    - using ***Naive*** (product sales next week will be the same as last week), or
    - using ***Seasonal Naïve*** (product sales next month with be as the same month last year).
    - using ***Judgmental*** (ask each store manager to make predictions).
  - In our case: Benchmark = 0.8

Use-Case
Inventory demand forecasting

- ***Step 4***: Validate the proposed model
  - Let's assume that the historical data are from 01.01.2015 until 31.12.2017.
  - Train the model ***using a data subset***, and predict sales for the next period (week/month). Estimate RMSE(1).
  - Do the same with ***N*** different data subsets.
  - Estimate the average RMSE.
  - Example: ***Model performance = 0.6***
- We can validate more models.



- If the benchmark is better that the model then:
  - Our model is not appropriate, or
  - The store managers are pretty good!

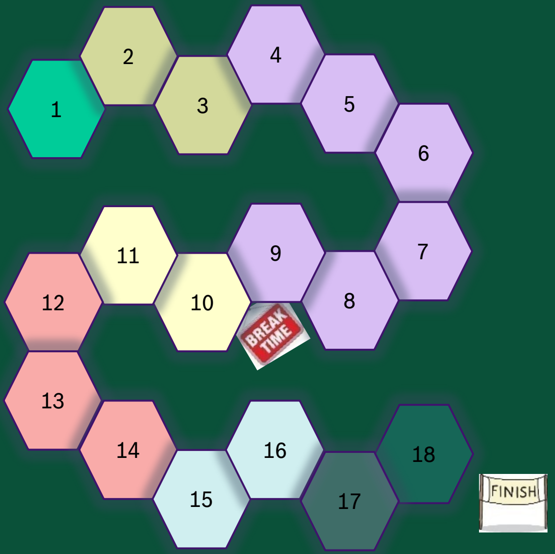
BOSCH

Use-Case
Inventory demand forecasting

- ***Step 5***: Finalize
  - Implement the best performing model into production.
  - Monitor the real-time performance frequently.
    - Use the model for real future forecasts.
    - Check the forecasts for “strange” values (too high, too low, negative, etc.).
    - ***Check if it performs as good as in the evaluation***.

16. USE-CASE

INTELLIGENT ENERGY MANAGEMENT IN BUILDINGS



Use-Case
Intelligent Energy Management in Buildings

- **Scope**: a DSS for the energy managers of buildings, which can assist them in setting indoor temperature set point, based on the feedback received by the occupants.
  - ***optimize energy use***, which is achieved though the efficient management of the heating and cooling systems
    - energy savings and thermal comfort are not always compatible.
    - indoor conditions shaped are not the ideal ones from the occupants' perspective (municipal building).
    - Current results: case-dependent, lead to relative time-independent suggestions (might not always be appropriate).
  - ***Set-point management*** should be dynamically applied in buildings, aiming both at:
    - Creating acceptable comfort levels for occupants by grasping their thermal sensations feeling, and
    - Achieving energy consumption reduction, leading to energy and cost savings.
  - We need a system which effectively captures the "real" sensation of the users and correlates it with the "predicted" one to detect a range of accepted set point temperatures.

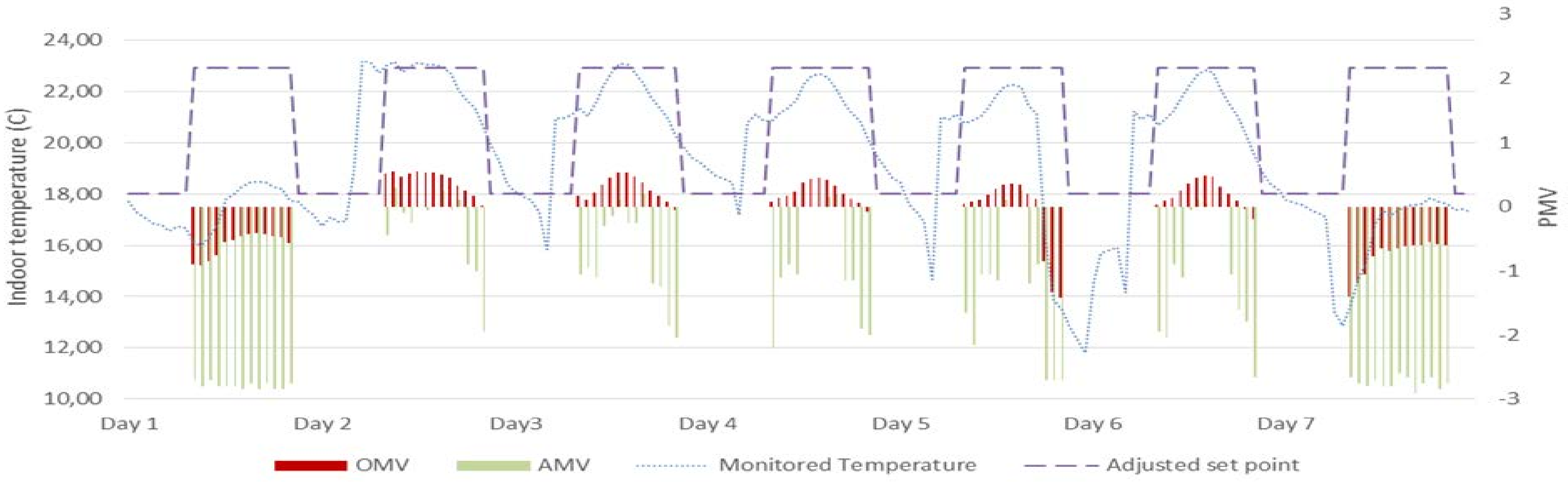
Use-Case
Intelligent Energy Management in Buildings

- **Metrics**:
  - Predicted Mean Vote (PMV) index, a 7-point thermal sensation scale.
    - can be calculated (EN ISO 7730:2005) by using predictions for the indoor conditions of a building.
  - Observed Mean Vote (OMV) index.
    - can be calculated using monitored values of temperature and humidity, as measured by building sensors in real-time.
  - Actual Mean Vote (AMV) though Thermal Comfort Validator (TCV).
    - derived directly by the TCV, (http://validator.optimussmartcity.eu), a web application, where occupants are encouraged to submit structured feedback on their thermal sensation.

BOSCH

Use-Case
Intelligent Energy Management in Buildings

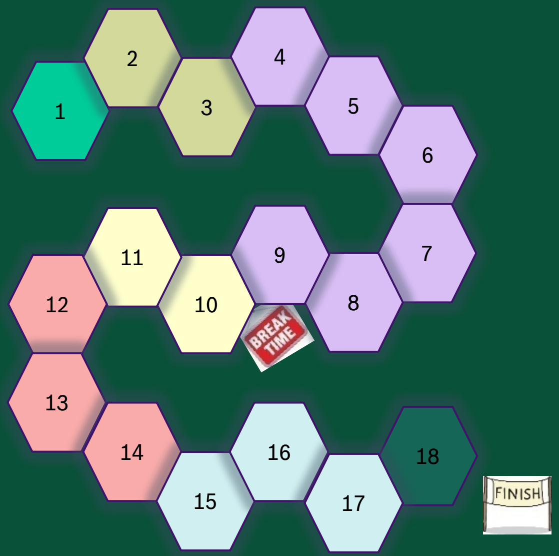
- Visualization of the ***real indoor temperature*** monitored, the ***set point temperature suggested***, as well as the OMV/AMV values estimated for the operating hours of the Sant-Cugat Town Hall (Spain)



Source: *"Decision Support System for Intelligent Energy Management in Buildings using the Thermal Comfort Model", Marinakis et. al., International Journal of Computational Intelligence Systems, Vol. 10 (2017) 882–893*

17. USE-CASE

THE FORECASTING POWER OF SOCIAL MEDIA

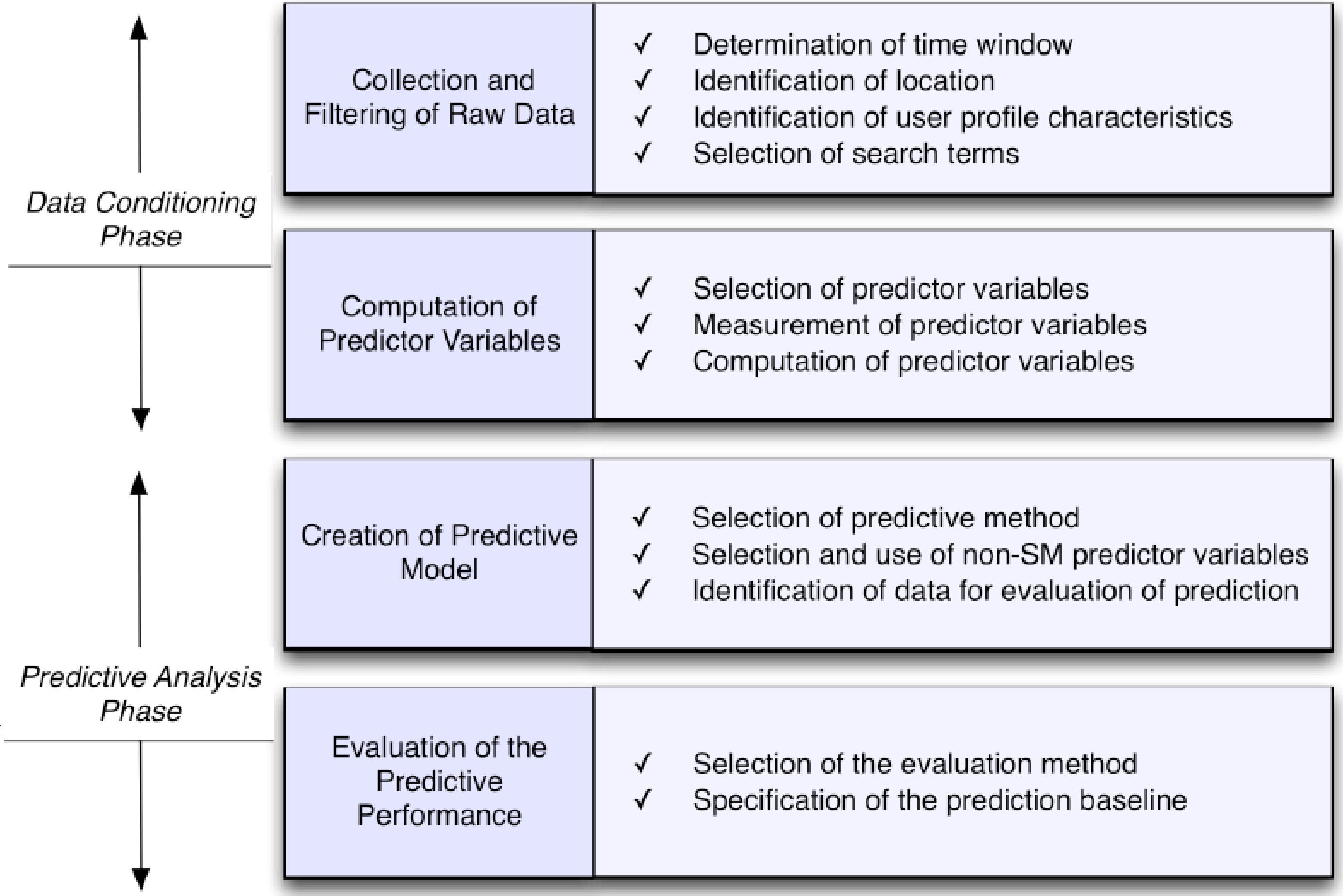


Use-Case
The forecasting power of social media

- Use of ***Social Media (SM) data*** for estimating forecasts in various areas, such as:
  - Product sales,
  - Stock market volatility,
  - Disease outbreaks,
  - Elections outcome,
  - etc.
- SM data includes: textual content, rating scores, like or dislike indicators, web search queries, tags and profile information.
- SM data incorporates personal opinions, thoughts and behaviors making it a vital component of the Web and a fertile ground for a variety of business and research endeavours.

Use-Case
The forecasting power of social media

- **Data Conditioning phase:**
  - Transformation of noisy raw data into ***high quality data*** that are structured based on some predictor variables.
    - When? Where? Who? What?
- **Predictive Analysis phase:**
  - Creation and evaluation of a ***forecasting model*** that enables estimating outcome from a new set of observations



Use-Case
The forecasting power of social media

- ***SM-Predictor variables***:
  - ***Volume-related*** variables: they measure the amount of SM data
    - Examples: number of tweets, number of reviews, number of queries etc.
  - ***Sentiment-related*** variables: they measure the sentiment expressed through the data.
    - Examples: bullishness index, review valence, review rating, etc.
  - ***Profile characteristics*** of online users:
    - Examples: Facebook friends, number of followers of users that posted a tweet, total posts, the location of the reviewer, in-degree.
- ***Non SM-Predictor variables***:
  - Past values of phenomenon, demographics, budget, etc.

Use-Case
The forecasting power of social media

- ***Forecasting method***:
  - Most common in literature: Linear Regression.
  - Other methods: Markov models, neural networks, support vector machine, Causality models

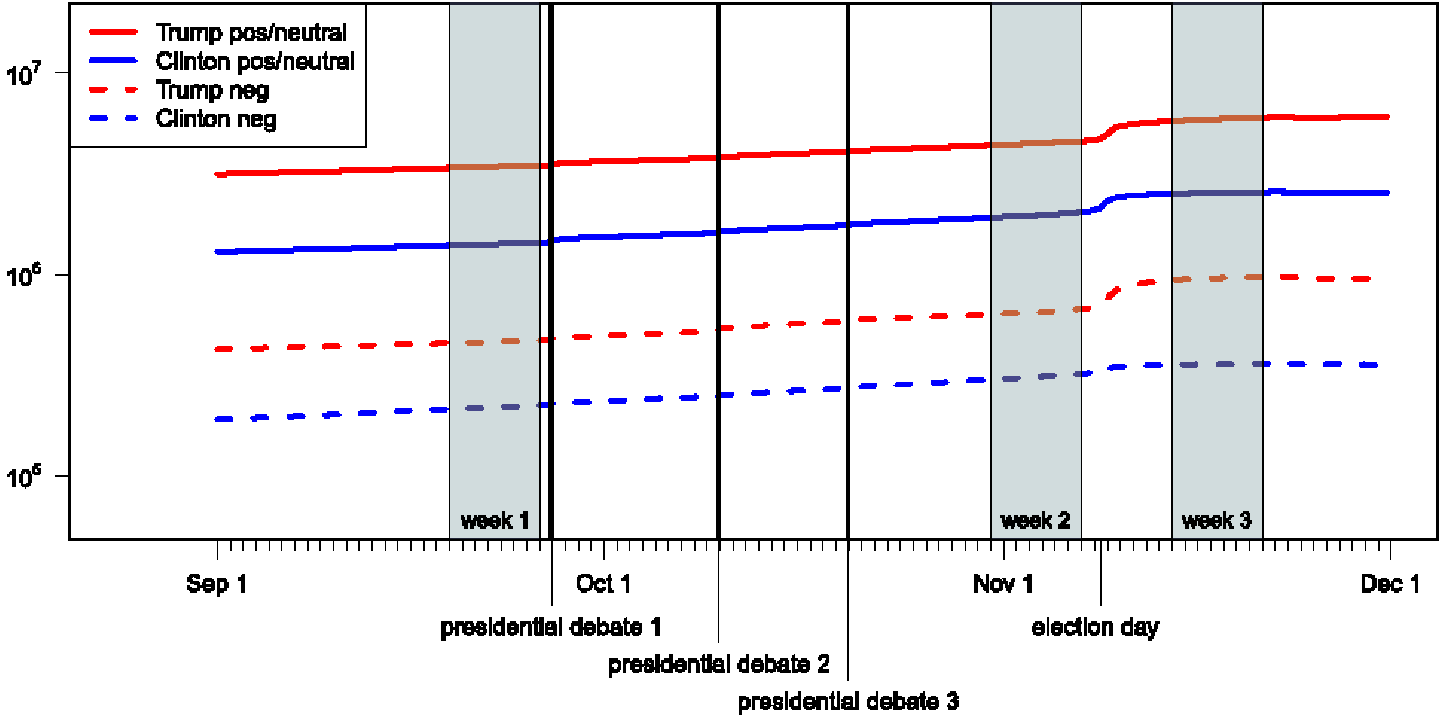
- ***Don't confuse forecasting and explanation!***
  - Forecasting power refers to the ability of predicting new observations accurately, while
  - Explanatory power to the strength of association indicated by a statistical model.
  - *"A statistically significant effect or relationship **does not guarantee** high predictive power, because the precision or magnitude of the causal effect might not be sufficient for obtaining levels of predictive accuracy that are practically meaningful".*

Source: *"Understanding the Predictive power of Social Media", Kalampokis et. al., Internet Research, Vol. 23 (2013)*

Use-Case
The forecasting power of social media: Example

- The 2016 US presidential election:
  - **Analyze the time of media posting related to Clinton and Trump**.
  - 4 timeseries:
    - Clinton vs. Trump,
    - supporters vs. opponents.
  - Instagram posts:
    - Users are 60% between 18-35
    - 16 hashtags included

Number of media, pos/neutral and negative hashtags (logarithmic scale).

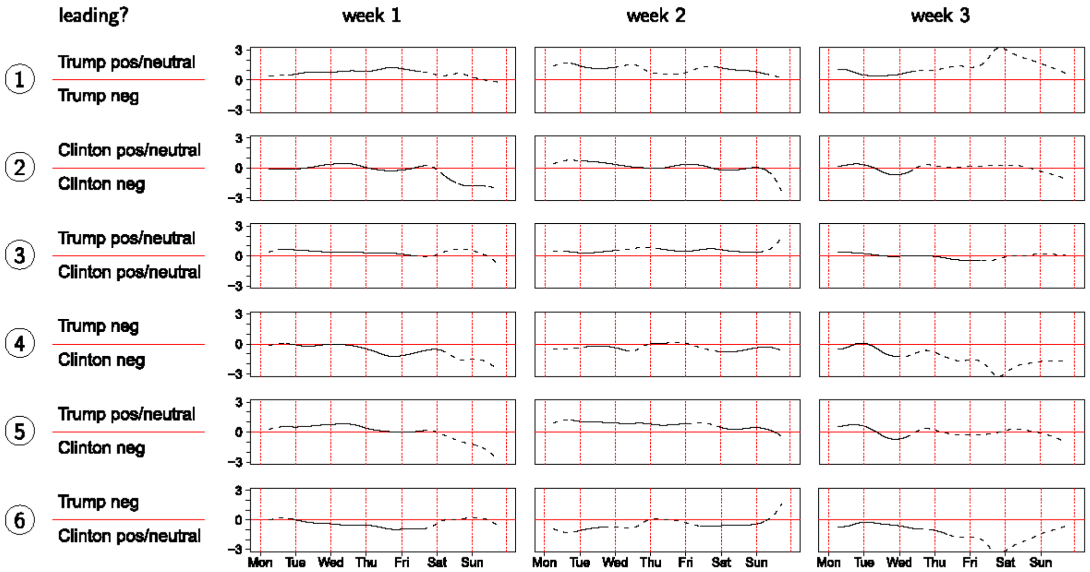


***Source**: "The 2016 US presidential election and media on Instagram: Who was in the lead?", Schmidbauer et. al., Computers in Human Behavior, Vol. 81 (2018), pp 148-160*

Use-Case
The forecasting power of social media: Example

- Instagram media postings in favor of, or neutral towards, Trump was ***massively higher*** than any other category.
- Clinton-related postings rose significantly in number, but were still behind on the election day.

Time series of phase differences (in hours) at period 12 h.

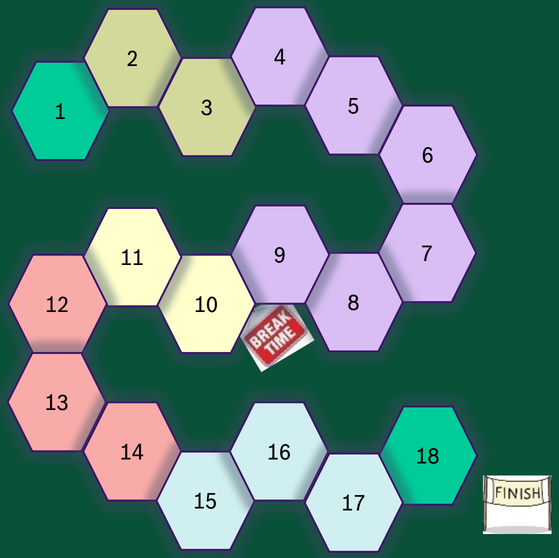


➢ ***Lesson learned*****:** Monitoring media uploads on Instagram and the analysis of the upload behavior could provide a real-time barometer of public opinion and sentiments for policymakers.

__*Source*__*: "The 2016 US presidential election and media on Instagram: Who was in the lead?", Schmidbauer et. al., Computers in Human Behavior, Vol. 81 (2018), pp 148-160*

BOSCH

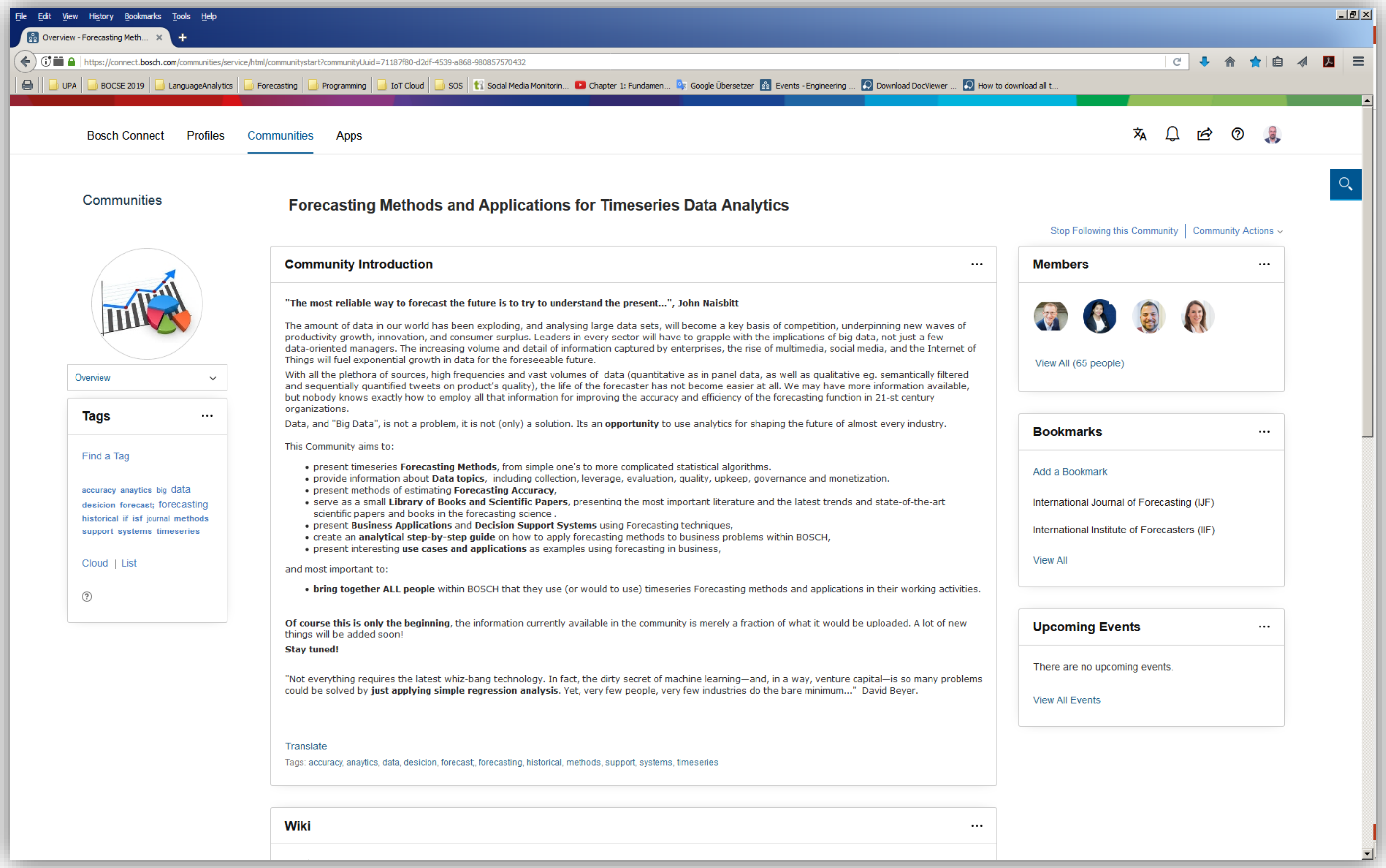
18. MORE?



PROPOSED LINKS & LITERATURE

Timeseries Forecasting: Methods & Applications

Bosch Community



http://bos.ch/Cfd

BOSCH

Timeseries Forecasting – Data Analytics & Quantitative Methods
Proposed Links & Literature

- K. Nikolopoulos, D. Thomakos (2019): *“Forecasting with the Theta Method: Theory and Applications”*
- R. Hyndman, G. Athanasopoulos (2013): “*Forecasting: Principles and Practice”*
- J. Ord, R. Fildes (2012): “*Principles of Business Forecasting”*
- T. Ruud, S. Babangida (2009): *“On the bias of Croston’s forecasting method*”, European Journal of Operational Research, Vol.194, pp.177-183.
- J. Hanke, A. Reitsch (2008): “*Business Forecasting”*
- K. Nikolopoulos, P. Goodwin, A. Patelis, V. Assimakopoulos (2007): *“Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches”*, European Journal of Operational Research, Vol.180, pp.354-368.
- G. Rowe (2007): *“A guide to Delphi”,* International Journal of Applied Forecasting 8, 11-16.
- P. Goodwin, R. Fildes, M. Lawrence, K. Nikolopoulos (2007): *“The process of using a forecasting support system”*, International Journal of Forecasting (IJF), Vol.23, pp.391-404.
- K. C. Green, J. S. Armstrong (2007): *“Structured analogies for forecasting”*, International Journal of Forecasting 23(3), 365–376.
- R. Hyndman, A. Koehler (2006): *“Another look at measures of forecast accuracy”*, IJF, Vol.22, pp.679-688.
- E. Gardner (2006): *“Exponential smoothing: The state of the art – Part II”*, IJF, Vol.22, pp.637-666.

Timeseries Forecasting – Data Analytics & Quantitative Methods
Proposed Links & Literature

- J. De Gooijer, R. Hyndman (2006): *“25 years of time series forecasting”*, IJF, Vol.22, pp.443-473.
- M. Lawrence, P. Goodwin, M. O’Connor, D. Önkal (2006): “*Judgmental forecasting: A review of progress over the last 25 years”,* International Journal of Forecasting 22(3), 493–518.
- Synetos, J. Boylan (2005): *“The accuracy of intermittent demand estimates”*, IJF, Vol.21, pp.303-314.
- Synetos, J. Boylan, J. Croston (2005): *“On the categorization of demand patterns”*, Journal of Operational Research Society, Vol.56, pp.495-503.
- R. Buehler, D. Messervey, D. Griffin (2005), *“Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation?”*, Organizational Behavior and Human Decision Processes 97(1), 47–63.
- R.G. Brown (2004), *“Smoothing, Forecasting and Prediction of Discrete Time Series”*, Courier Corporation ,Technology & Engineering.
- Hyndman & Billah (2003): *“Theta: SES with drift?”*
- R. Hyndman, A. Koehler, R. Snyder, S. Grose (2002): *“A state space framework for automatic forecasting using exponential smoothing methods”*, IJF, Vol.18, pp.439-454.
- V. Assimakopoulos, K. Nikolopoulos (2000): *“The Theta model: A decomposition approach to forecasting”*, IJF, Vol.16, pp.521-530.
- S. Makridakis, M. Hibon (2000): *“The M3-competition: Results, conclusions and implications”*, IJF, Vol.16 (special issue), pp.451-476.

BOSCH

Timeseries Forecasting – Data Analytics & Quantitative Methods
Proposed Links & Literature

- G. Rowe. G. Wright (1999): *"The Delphi technique as a forecasting tool: issues and analysis"*, IJF, Vol.15, pp.353-375.
- S. Makridakis, S. Wheelwright, R. Hyndman (1998): "*Forecasting: Methods and Applications*"
- Assimakopoulos V. (1995): *"A successive filtering technique for identifying long-term trends"*, IJF Vol14, 35-43.
- J. D. Hamilton (1994): "*Timeseries Analysis*"
- J. Armstrong, F. Collopy (1992): *"Error Measures for generalizing about forecasting methods: Empirical comparisons"*, IJF, Vol.8, pp.69-80.
- S. Makridakis (1990): *"Sliding Simulation: A new approach to time series forecasting"*, Management Science, Vol.36, pp.505-512.
- S. Wheelwright, S. Makridakis (1985): "*Forecasting methods for Management*"
- S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, R. Winkler (1982): *"The accuracy of extrapolation (time series) methods: Results of a forecasting competition"*, IJF, Vol.1, pp 111-153.
- H. Kahn, A.J. Wiener (1967): "*The use of scenarios*", Hudson Institute
- http://www.forecasters.org
- http://www.sciencedirect.com
- http://www.elsevier.com

THANK YOU

"The mind is not a vessel to be filled, but a fire to be kindled."

Plutarch, Greek Historian