Time Series Analysis and Forecasting

Mohammad Alfan Alfian Riyadi (Data Scientist)
alfan@warungpintar.co
Meet - 1

About Me

- Education
  - Bachelor Degree Statistics (2012-2016)
  - Master Degree Statistics (2016-2018)
- Activity
  - Advisor DSI East Java Chapter
- Working Experience
  - Sept 2018 - Now

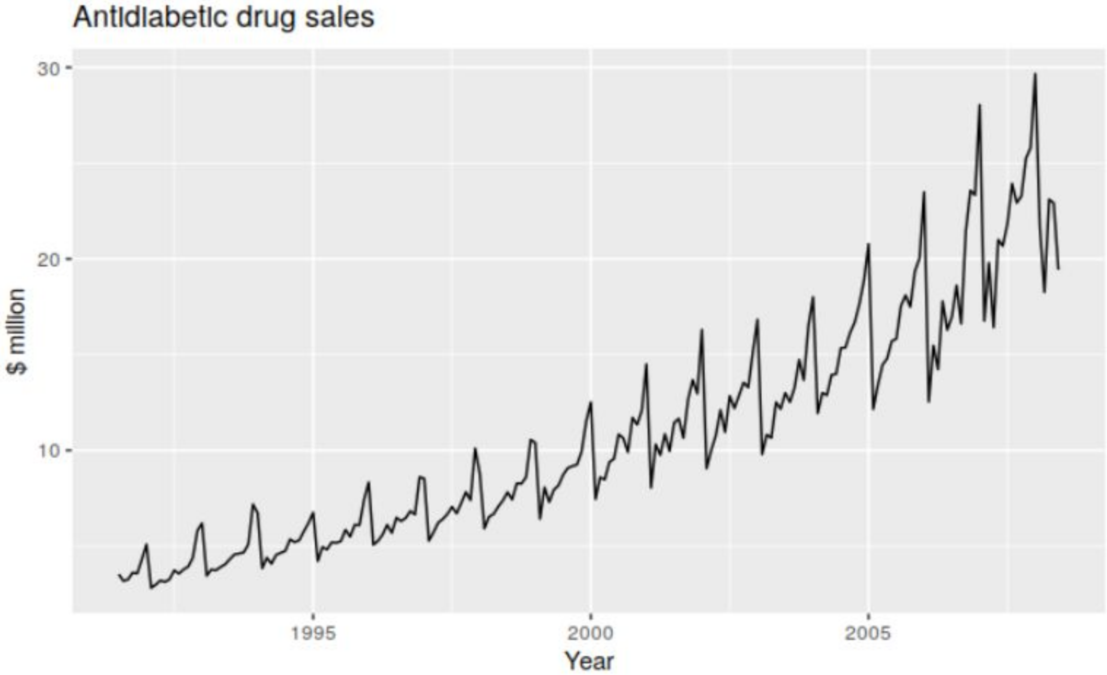
Outline

- Introduction
- Time Series Pattern
- Naive Model
- Moving Average Method
- Exponential Smoothing
- Time Series Regression
- Evaluating Forecast Accuracy

Outline

- **Introduction**
- Time Series Pattern
- Naive Model
- Moving Average Method
- Exponential Smoothing
- Time Series Regression
- Evaluating Forecast Accuracy

Introduction : Time Series Data



Source :

Forecasting: Principles and Practice

Rob J Hyndman and George Athanasopoulos

Introduction : Time Series Data

Introduction : Time Series Data

Ahead of the announcement of the 2019 election results

Warung PINTAR

Introduction : Forecasting, Planning, and Goal

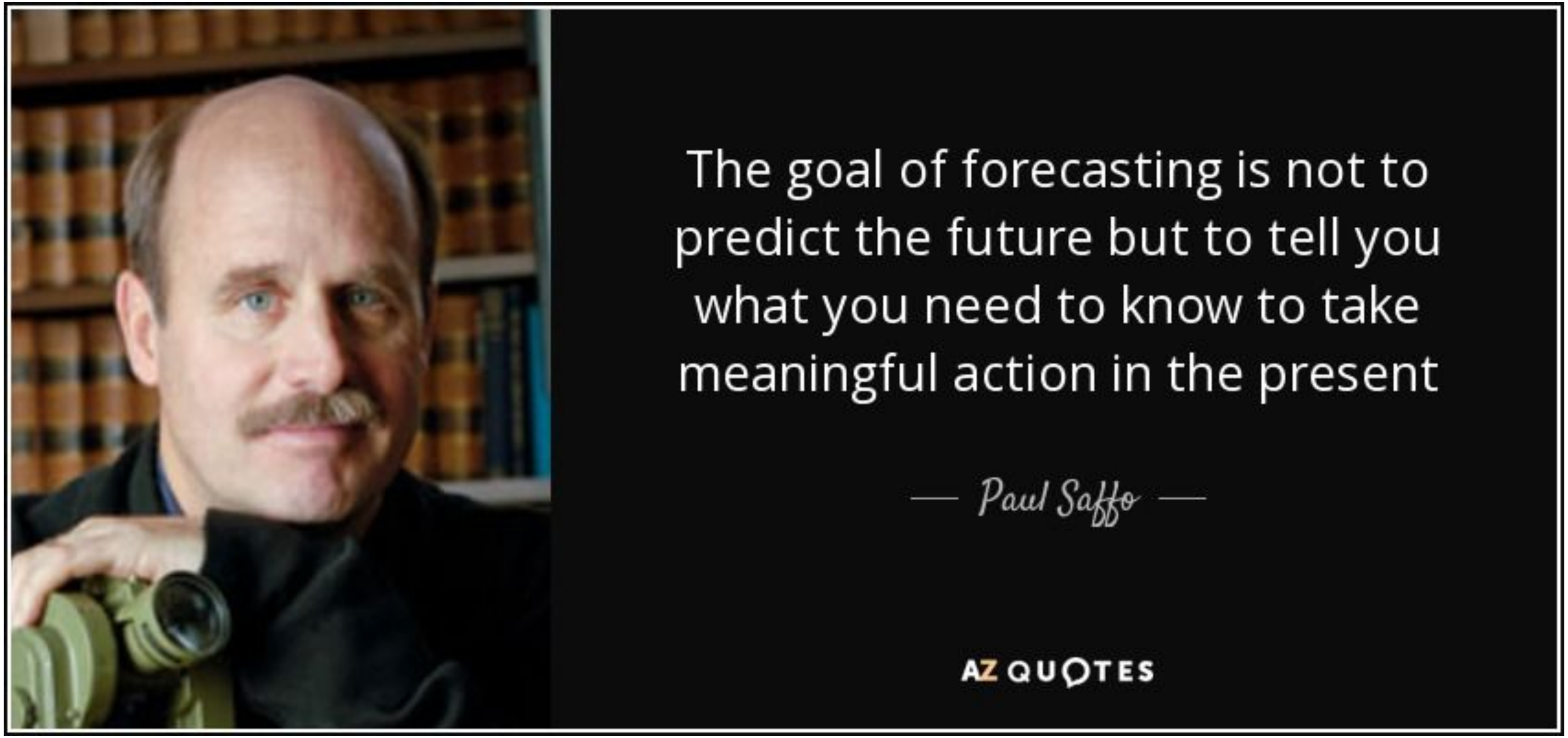
- **Forecasting** : is about **predicting the future as accurately as possible**, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.
- **Goals** : are what you **would like to have happen**. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.
- **Planning** : is a **response to forecasts and goals**. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

Source :
Forecasting: Principles and Practice
Rob J Hyndman and George Athanasopoulos

Introduction : Quotes

"Forecasting is the ART of saying what will happen, and then explaining why it didn't!"

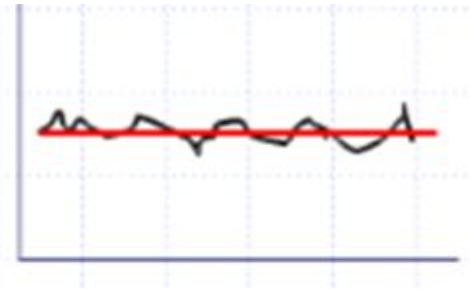
"A person who doesn't care about THE PAST
Is a person who doesn't have THE FUTURE"


The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present
— Paul Saffo —
AZ QUOTES

Outline

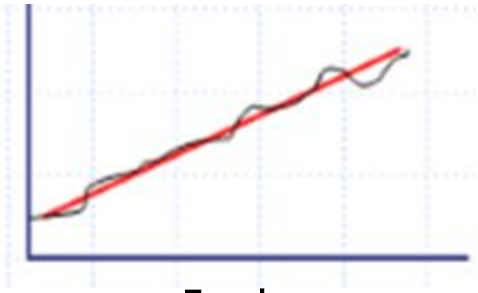
- Introduction
- **Time Series Pattern**
- Naive Model
- Moving Average Method
- Exponential Smoothing
- Time Series Regression
- Evaluating Forecast Accuracy

Time Series Pattern : General



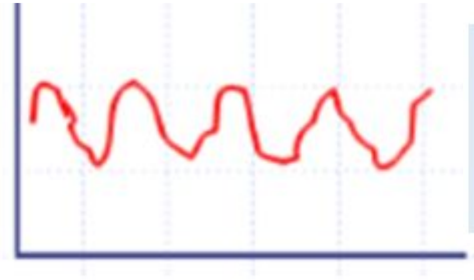
Stationary

- ✓ Naive Model
- ✓ Simple Exponential Smoothing
- ✓ Simple Average



Trend

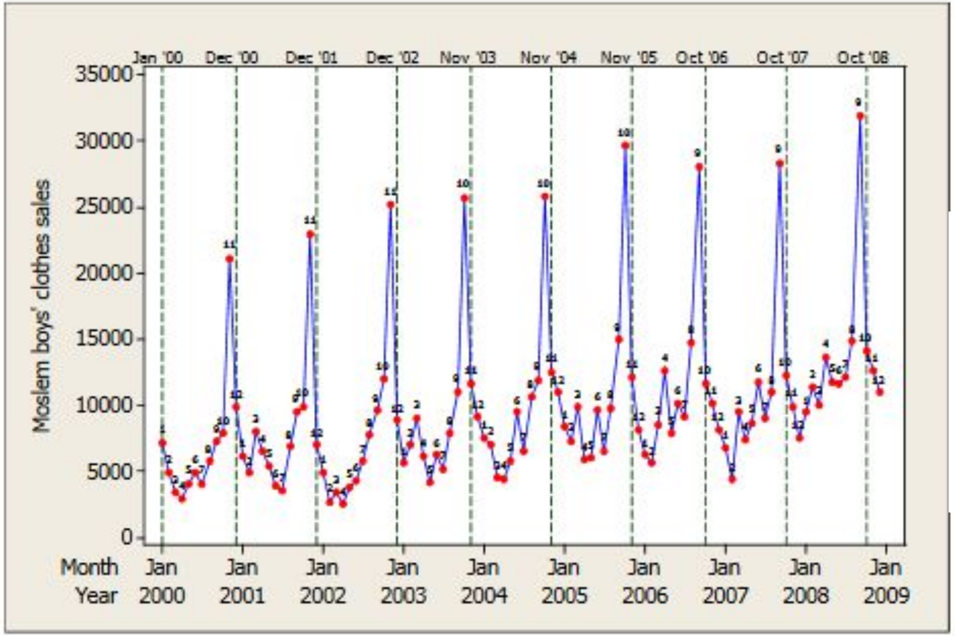
- ✓ Naive Model
- ✓ “Holt” Exponential Smoothing
- ✓ Double Moving Average
- ✓ Trend Analysis Time Series Regression



Seasonal

- ✓ Naive Model
- ✓ Seasonal Time Series Regression
- ✓ “Holt-Winters” Exponential Smoothing

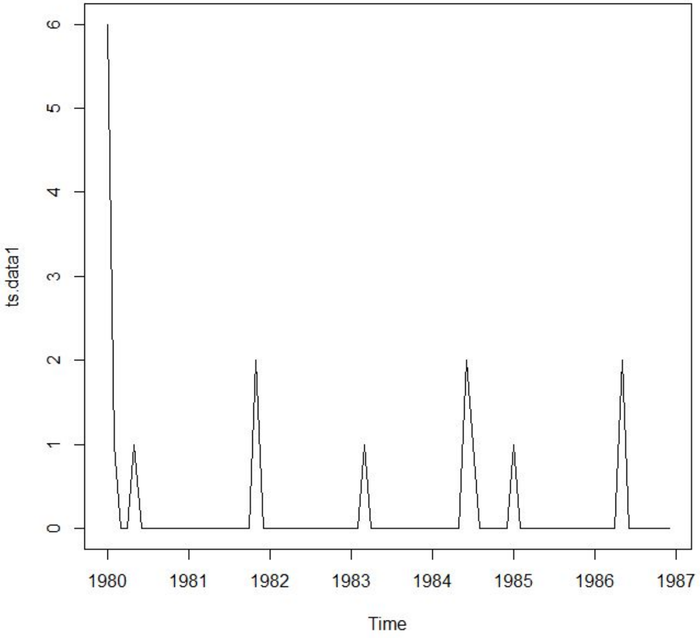
Time Series Pattern : Calendar Variation



Year	Date	Year	Date
2000	08-09 January	2004	14-15 November
	28-29 December	2005	03-04 November
2001	17-18 December	2006	23-24 October
2002	06-07 December	2007	12-13 October
2003	25-26 November	2008	01-02 October

Source :
Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect
Muhammad Hisyam Lee and Suhartono

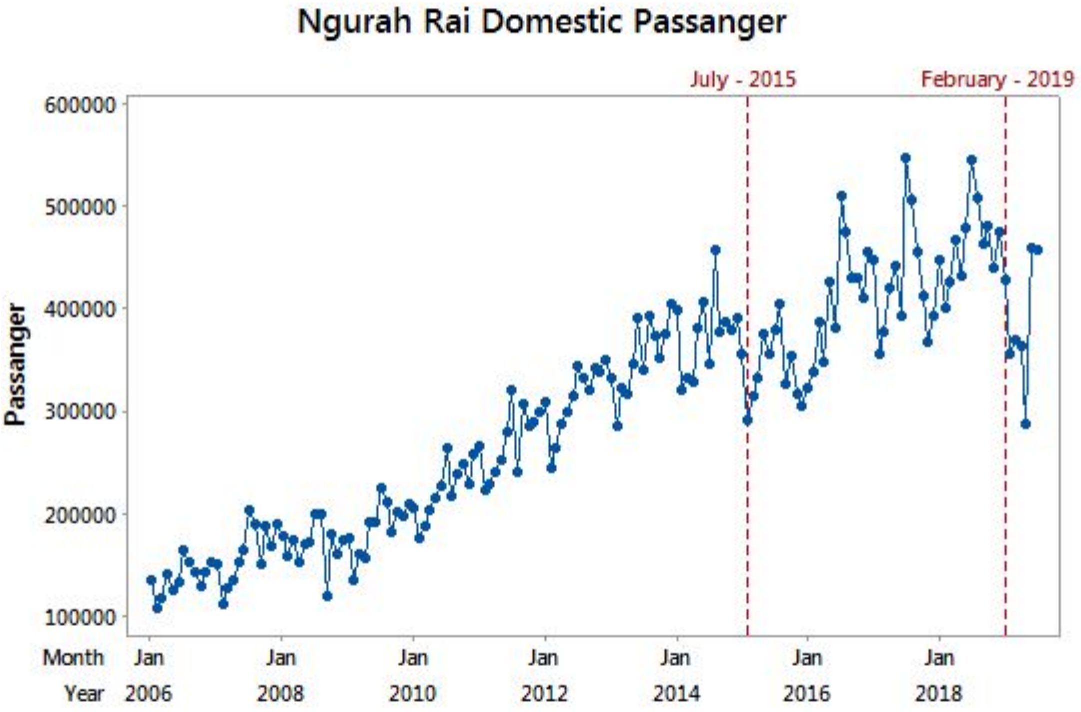
Time Series Pattern : Intermittent



Many zero values

Warung PINTAR

Time Series Pattern : Intervention Analysis



Gunung Raung Kembali Meletus, 3 Bandara Ditutup Lagi

Tanti Yulianingsih
22 Jul 2015, 15:24 WIB

Share
10.5k

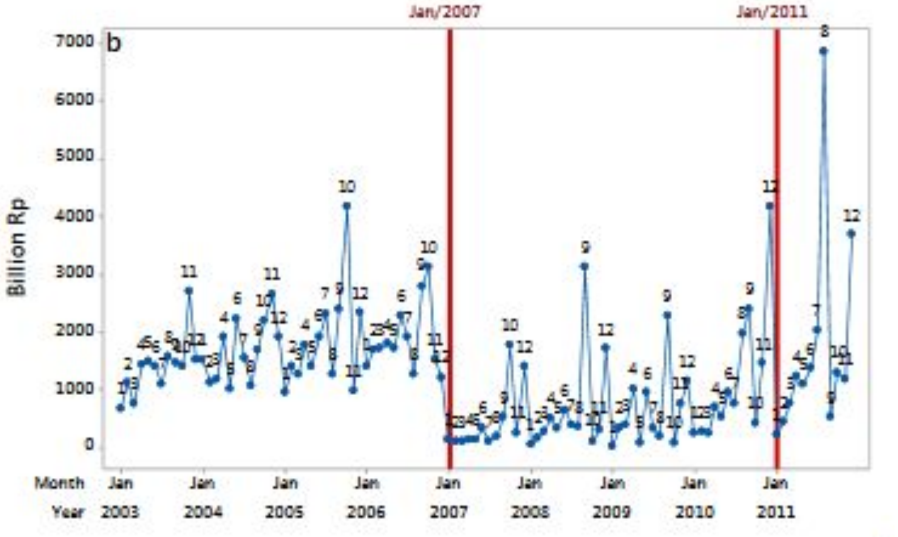
BERANDA / BERITA / MAKRO

Harga Tiket Naik, Jumlah Penumpang Pesawat Domestik Turun 15,5%

Penulis: Ihya Ulum Aldin
Editor: Desy Setyowati
1/4/2019, 15.33 WIB

Time Series Pattern : ?????

Outflow Rupiah at West Java



Journal of Physics: Conference Series

PAPER • OPEN ACCESS

GSTAR-SUR Modeling With Calendar Variations And Intervention To Forecast Outflow Of Currencies In Java Indonesia

M S Akbar[1,2], Setiawan[1], Suhartono[1], B N Ruchjana[4] and M A A Riyadi[3]

Outline

- Introduction
- Time Series Pattern
- **Naive Model**
- Moving Average Method
- Exponential Smoothing
- Time Series Regression
- Evaluating Forecast Accuracy

Naive Model

All forecasts to be the value of the **last observation**

- **Stationary**

$$\hat{y}_{T+h|T} = y_T$$

- **Trend**

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right)$$

- **Seasonal**

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$$

Naive Model

All forecasts to be the value of the **last observation**

- **Stationary**

$$\hat{y}_{T+h|T} = y_T$$

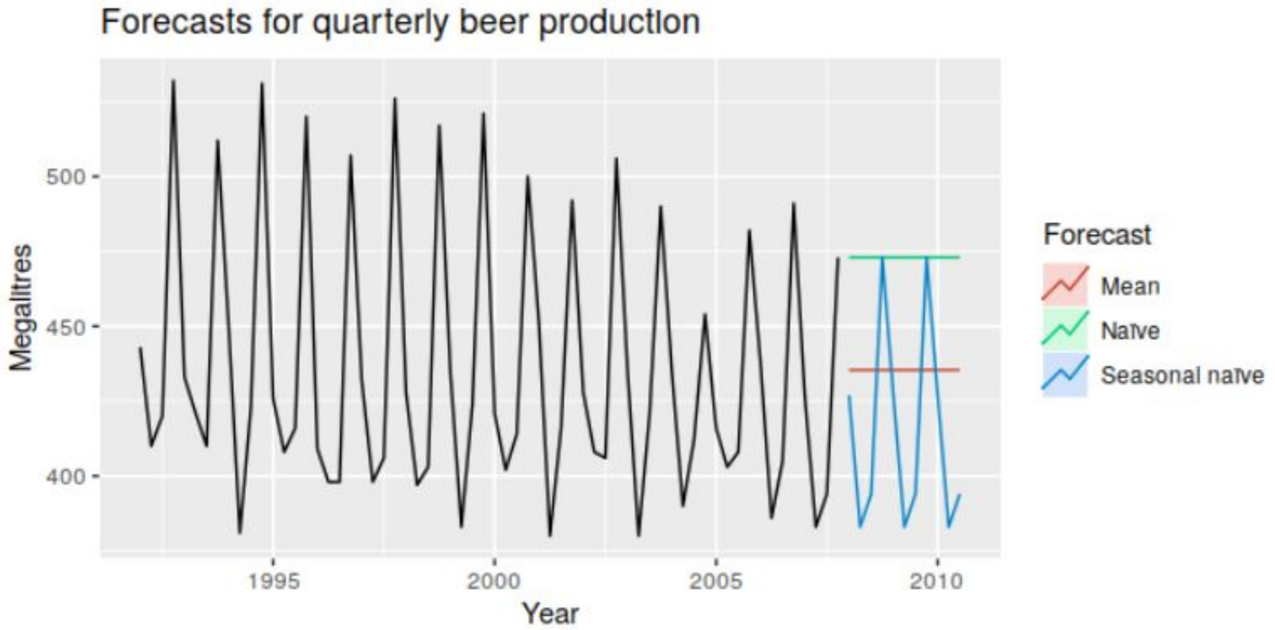
- **Trend**

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right)$$

- **Seasonal**

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$$

Naive Model



Source :
Forecasting: Principles and Practice
Rob J Hyndman and George Athanasopoulos

Outline

- Introduction
- Time Series Pattern
- Naive Model
- **Moving Average Method**
- Exponential Smoothing
- Time Series Regression
- Evaluating Forecast Accuracy

Moving Average Method

All future values are equal to **the average (or "mean")** of the historical data

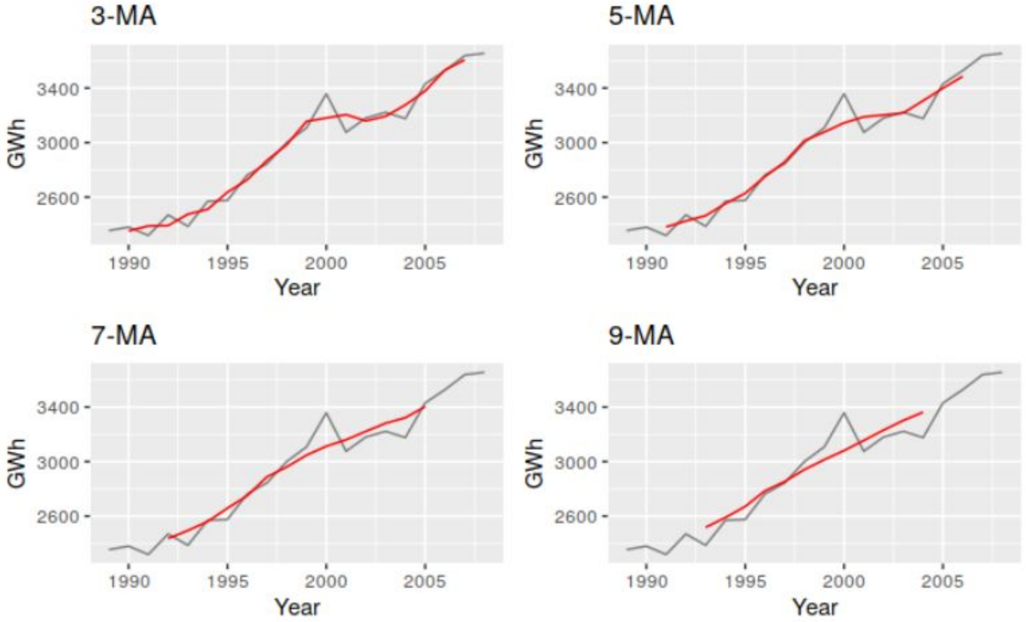
- **Simple Average**

$$\hat{Y}_{t+1} = \sum_{t=1}^{n} \frac{Y_t}{n}$$

- **Moving Average**

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}$$ where $$m = 2k + 1$$

Moving Average Method



Source :
Forecasting: Principles and Practice
Rob J Hyndman and George Athanasopoulos

Outline

- Introduction
- Time Series Pattern
- Naive Model
- Moving Average Method
- **Exponential Smoothing**
- Time Series Regression
- Evaluating Forecast Accuracy

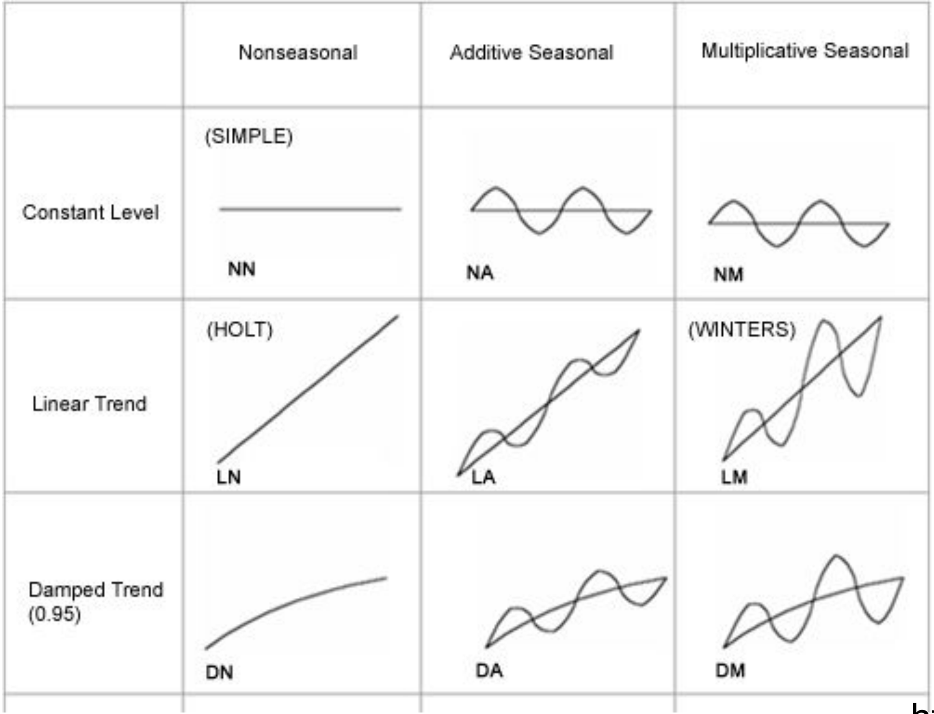
Exponential Smoothing

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

Short hand	Method
(N,N)	Simple exponential smoothing
(A,N)	Holt's linear method
(A_d,N)	Additive damped trend method
(A,A)	Additive Holt-Winters' method
(A,M)	Multiplicative Holt-Winters' method
(A_d,M)	Holt-Winters' damped method

Source :
Forecasting: Principles and Practice
Rob J Hyndman and George Athanasopoulos

Exponential Smoothing



http://www.forecastpro.com

Exponential Smoothing

Table 7.6: Formulas for recursive calculations and point forecasts. In each case, ℓ_t denotes the series level at time t, b_t denotes the slope at time t, s_t denotes the seasonal component of the series at time t, and m denotes the number of seasons in a year; α, β^*, γ and ϕ are smoothing parameters, $\phi_h = \phi + \phi^2 + \cdots + \phi^h$, and k is the integer part of $(h-1)/m$.

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A$_\text{d}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

Source :
Forecasting: Principles and Practice
Rob J Hyndman and George Athanasopoulos

Outline

- Introduction
- Time Series Pattern
- Naive Model
- Moving Average Method
- Exponential Smoothing
- **Time Series Regression**
- Evaluating Forecast Accuracy

Time Series Regression

- **Trend**

$$Y_t = a + b.t + error$$

- **Seasonal**

$$Y_t = a + b_1 D_1 + \cdots + b_{s-1} D_{s-1} + error$$

- **Trend + Seasonal**

$$Y_t = a + b.t + c_1 D_1 + \ldots + c_{s-1} D_{s-1} + error$$

Dummy Variables

Time Series Regression

Trend

Week	Sales
1	104.2257
2	108.6764
3	112.3969
4	116.6349
5	120.7849
6	124.8041
7	128.1536
8	132.6837
9	136.4367
10	140.2115
11	144.5022
12	148.6217
13	152.5882
14	156.1204
15	160.9185

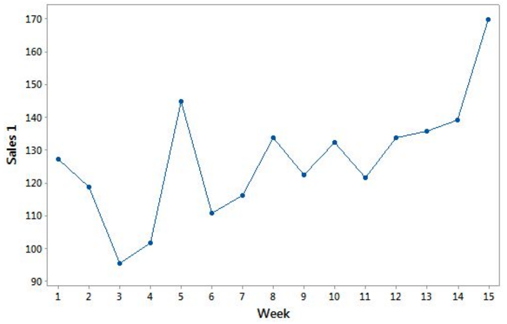
Seasonal

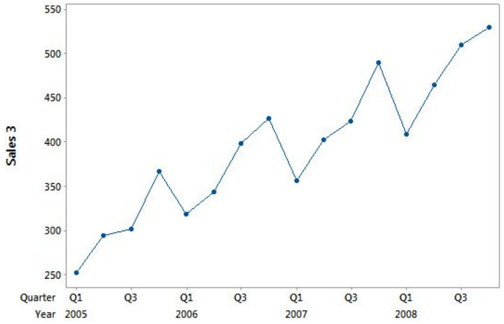
Year	Quarter	Sales
2005	1	217.4616
2005	2	235.2335
2005	3	284.8008
2005	4	305.6697
2006	1	229.8714
2006	2	252.0008
2006	3	269.8258
2006	4	297.2252
2007	1	209.2582
2007	2	263.998
2007	3	272.903
2007	4	293.4505
2008	1	205.8395
2008	2	254.2567
2008	3	277.8048
2008	4	300.5656

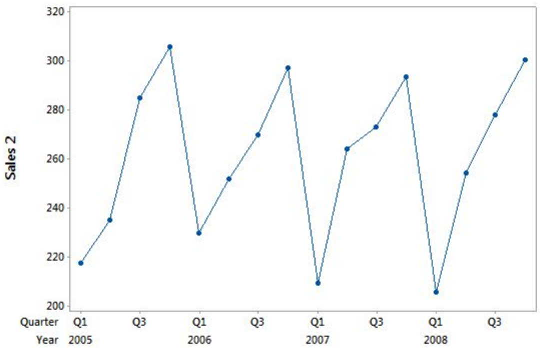
Seasonal + Trend

t	Year	Quarter	Sales
1	2005	1	252.4616
2	2005	2	294.2335
3	2005	3	301.8008
4	2005	4	367.6697
5	2006	1	318.8714
6	2006	2	344.0008
7	2006	3	398.8258
8	2006	4	427.2252
9	2007	1	357.2582
10	2007	2	402.998
11	2007	3	423.903
12	2007	4	490.4505
13	2008	1	408.8395
14	2008	2	465.2567
15	2008	3	509.8048
16	2008	4	530.5656

Time Series Regression







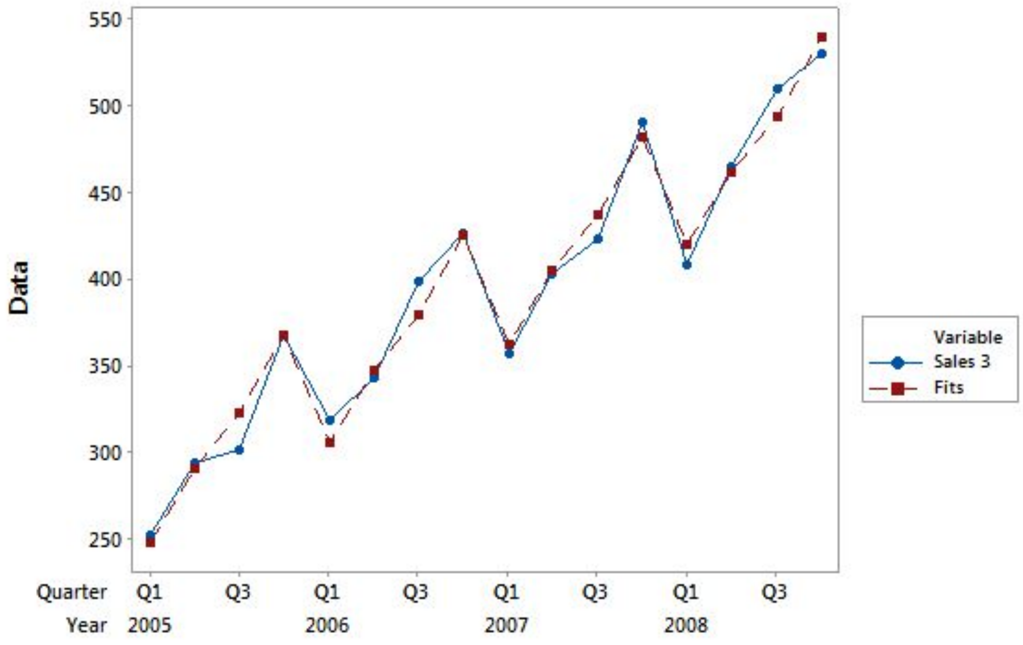
Time Series Regression

Seasonal + Trend

t	Year	Quarter	Sales
1	2005	1	252.4616
2	2005	2	294.2335
3	2005	3	301.8008
4	2005	4	367.6697
5	2006	1	318.8714
6	2006	2	344.0008
7	2006	3	398.8258
8	2006	4	427.2252
9	2007	1	357.2582
10	2007	2	402.998
11	2007	3	423.903
12	2007	4	490.4505
13	2008	1	408.8395
14	2008	2	465.2567
15	2008	3	509.8048
16	2008	4	530.5656

t	Year	Quarter	Sales	Quarter_1	Quarter_2	Quarter_3
1	2005	1	252.4616	1	0	0
2	2005	2	294.2335	0	1	0
3	2005	3	301.8008	0	0	1
4	2005	4	367.6697	0	0	0
5	2006	1	318.8714	1	0	0
6	2006	2	344.0008	0	1	0
7	2006	3	398.8258	0	0	1
8	2006	4	427.2252	0	0	0
9	2007	1	357.2582	1	0	0
10	2007	2	402.998	0	1	0
11	2007	3	423.903	0	0	1
12	2007	4	490.4505	0	0	0
13	2008	1	408.8395	1	0	0
14	2008	2	465.2567	0	1	0
15	2008	3	509.8048	0	0	1
16	2008	4	530.5656	0	0	0

Time Series Regression



Outline

- Introduction
- Time Series Pattern
- Naive Model
- Moving Average Method
- Exponential Smoothing
- Time Series Regression
- **Evaluating Forecast Accuracy**

Evaluation Forecast Accuracy

Forecast Error

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

Mean Absolute Error

$$\text{MAE} = \text{mean}(|e_t|)$$

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\text{mean}(e_t^2)}$$

Mean Absolute Percentage Error

$$\text{MAPE} = \text{mean}(|p_t|)$$ Where $p_t = 100e_t/y_t$

Warung PINTAR

Evaluation Forecast Accuracy

ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

international journal of forecasting

www.elsevier.com/locate/ijforecast

The M3-Competition: results, conclusions and implications

Spyros Makridakis, Michèle Hibon*

INSEAD, Boulevard de Constance, 77305 Fontainebleau, France

Abstract

This paper describes the M3-Competition, the latest of the M-Competitions. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results/conclusions with those of the previous two M-Competitions as well as with those of other major empirical studies. Finally, the implications of these results and conclusions are considered, their consequences for both the theory and practice of forecasting are explored and directions for future research are contemplated.

Evaluation Forecast Accuracy

- Statistically sophisticated or **complex methods** do not necessarily provide **more accurate forecasts** than **simpler ones**.
- The **relative ranking of the performance** of the various methods varies according to the **accuracy measure** being used.
- The accuracy when various methods are being **combined** outperforms, on average, the **individual methods being combined** and does very well in comparison to other methods.
- The **accuracy** of the various methods depends upon the **length of the forecasting horizon** involved.

Evaluation Forecast Accuracy

International Journal of Forecasting 34 (2018) 802–808

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The M4 Competition: Results, findings, conclusion and way forward

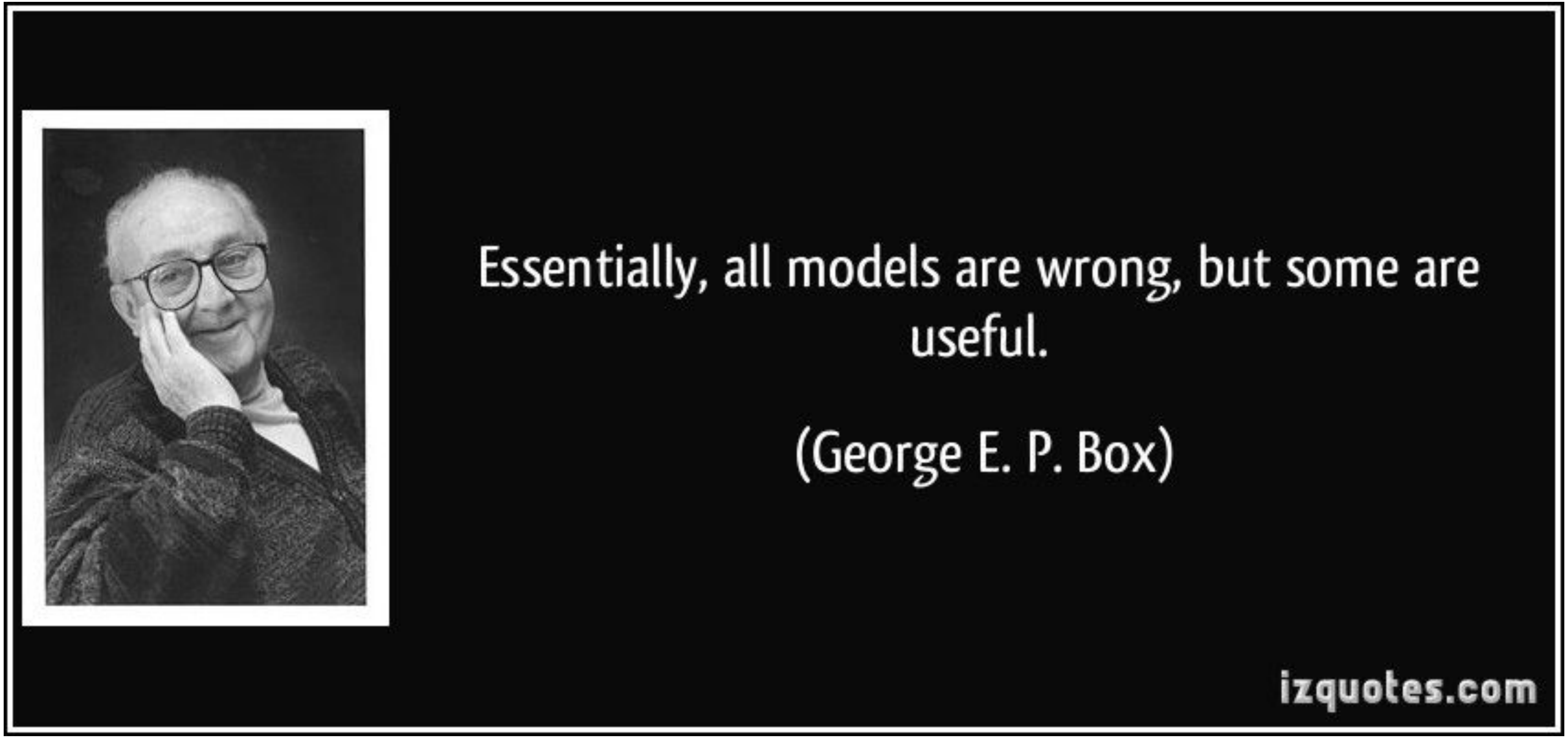
Spyros Makridakis [a,b,*], Evangelos Spiliotis [c], Vassilios Assimakopoulos [c]

[a] *University of Nicosia, Nicosia, Cyprus*
[b] *Institute For the Future (IFF), Nicosia, Cyprus*
[c] *Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, Greece*

Evaluation Forecast Accuracy

- Out Of the **17 most accurate methods**, 12 were "combinations" of mostly **statistical approaches**.
- The biggest surprise was a **"hybrid" approach** that utilized both **statistical and ML features**. This method's average sMAPE was close to 10% more accurate than the combination benchmark used to compare the submitted methods.
- The second most accurate method was a **combination** of seven **statistical methods and one ML one**, with the weights for the averaging being calculated by a ML algorithm that was trained to minimize the forecasting.
- The two most accurate methods also achieved an amazing success in specifying the 95% prediction intervals correctly.
- The **six pure ML methods** performed **poorly,** with none of them being more accurate than the combination benchmark and only one being more accurate than Naïve2.

Warung PINTAR



Let’s discuss!