Machine Learning

Chapter 4: Forecasting

March 2021

Contents

comillas.edu

1

Introduction to Forecasting

Introduction

Forecasting

"Prediction is very difficult, especially if it's about the future."

- Nils Bohr, Nobel laureate in Physics

- **Forecast:** prediction of some future event
- Forecasting problems are classified as:
  - **Short-term**: a few time periods (days, weeks, months) into the future
  - **Medium-term**: from one to two years into the future
  - **Long-term**: many years
- Short and medium term forecasting are tipically based on identifying, modeling and extrapolating the **patterns** found in **historical data**.
- Long term forecasting is usually based on **expert knowledge** and fundamental models.

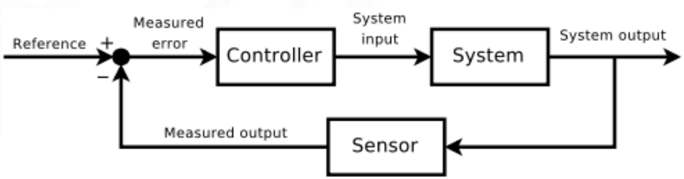
Objectives of Time Series Analysis

1. Describing the evolution of a time series.

2. Modelling the process that has generated the time series by means of a suitable statistical model.

3. Forecasting future values of the time series.

4. Control. Good forecasts enable the analyst to take actions so as to control a given process.



Electricity load forecasting (1)

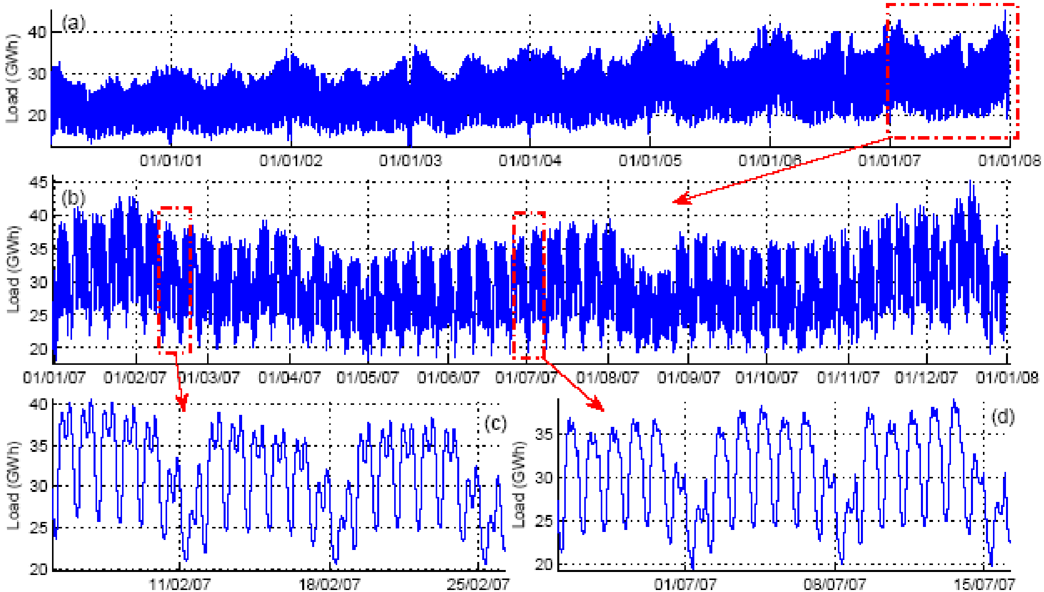
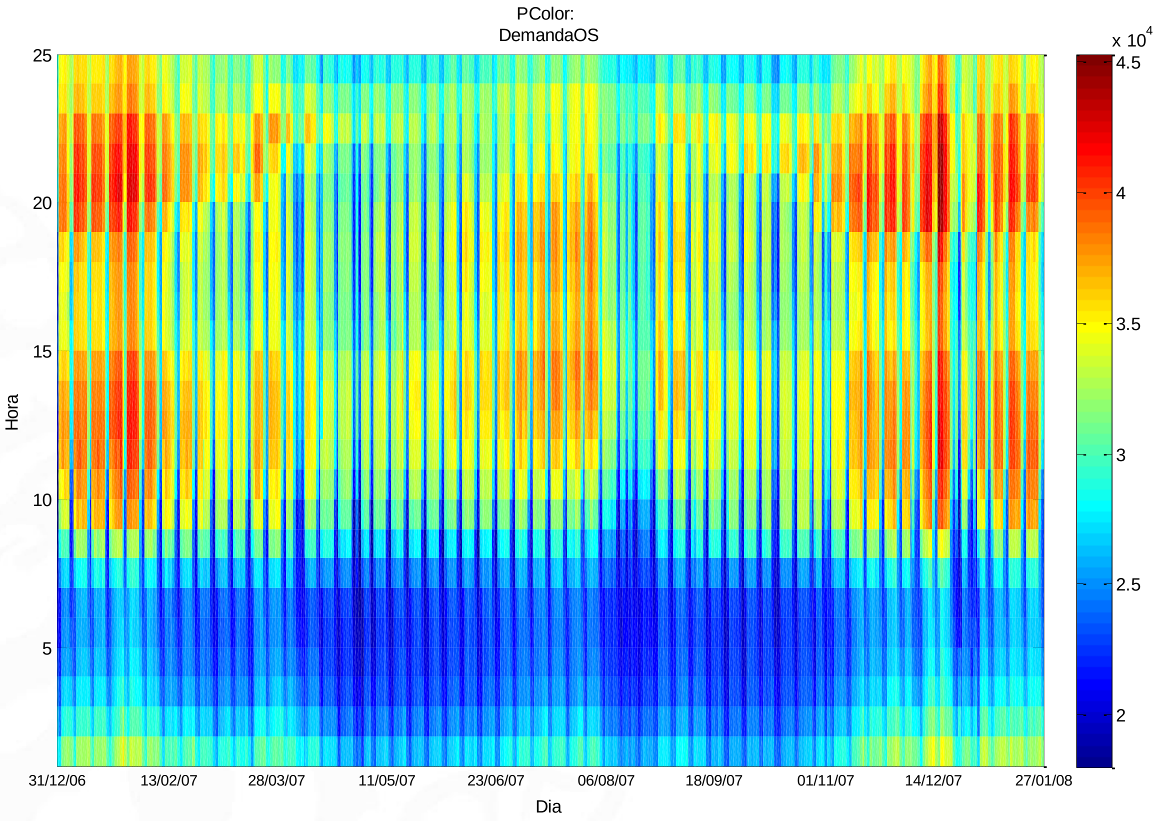


Fig. 1 Hourly electricity demand in Spain:(a)From January 1, 2000 to December 31, 2007; (b) From January 1, 2007 to December 31, 2007; (c) Three winter weeks (2007); (d) Three summer weeks (2007).

Source: Muñoz et al. (2010)

Introduction
Electricity load forecasting (1)



Introduction
Electricity load forecasting (2)

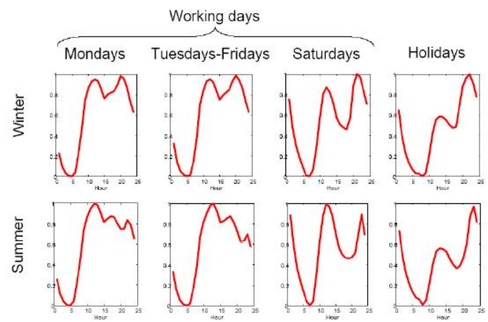


Fig. 2 Normalized intra-day load profiles for the Spanish electricity load.

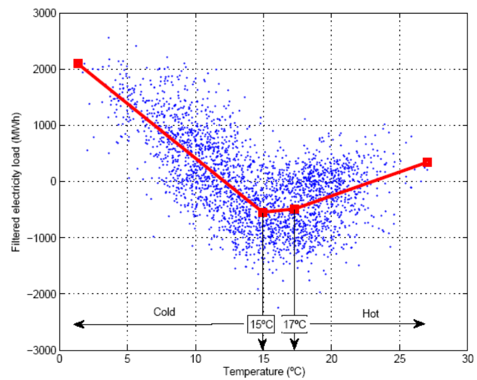


Fig. 3 Non-linear relationship between the (filtered) electricity load of a distribution area of the north of Spain and the average daily mean temperature of this region.

Source: Muñoz et al. (2010)

Forecasting methods

Forecasting techniques can be classified in two main groups:

- Quantitative methods:
  - Sufficient information about the past is available
  - This information can be set as numerical time series
  - We can assume that the future behavior of the process will be similar to the observed past behavior (continuity assumption).
- Qualitative methods:
  - Little or no quantitative information is available.
  - These methods are based on expert knowledge
  - Example: Delphi method

Quantitative methods

- 2 different types of models:
  - Explanatory models:

$$y = f(x_1, x_2, \ldots, x_n, noise)$$

  - Time Series models:

$$y(t) = f(y(t-1), y(t-2), \ldots, noise)$$

$$y(t) = f(y(t-1), y(t-2), \ldots, x(t), x(t-1), \ldots, noise)$$

- In both cases the observation is composed of two components:

$$y = \underbrace{pattern}_{forecast} + \underbrace{noise}_{uncertainty}$$

(additive noise)

- The objective of the modeling process is to separate both components, in order to use the pattern for forecasting and the observed noise for characterizing prediction errors.

2

Decomposition methods

Decomposition methods
Introduction

- Mathematical formulation:

$$y(t)=f(\tau(t), \sigma(t), \varepsilon(t))$$

where:

- $y(t)$: time series value at time t
- $\tau(t)$: trend cycle component at time t
- $\sigma(t)$: seasonal component at time t
- $\varepsilon(t)$: irregular (or remainder) component at time t

Decomposition methods
Example

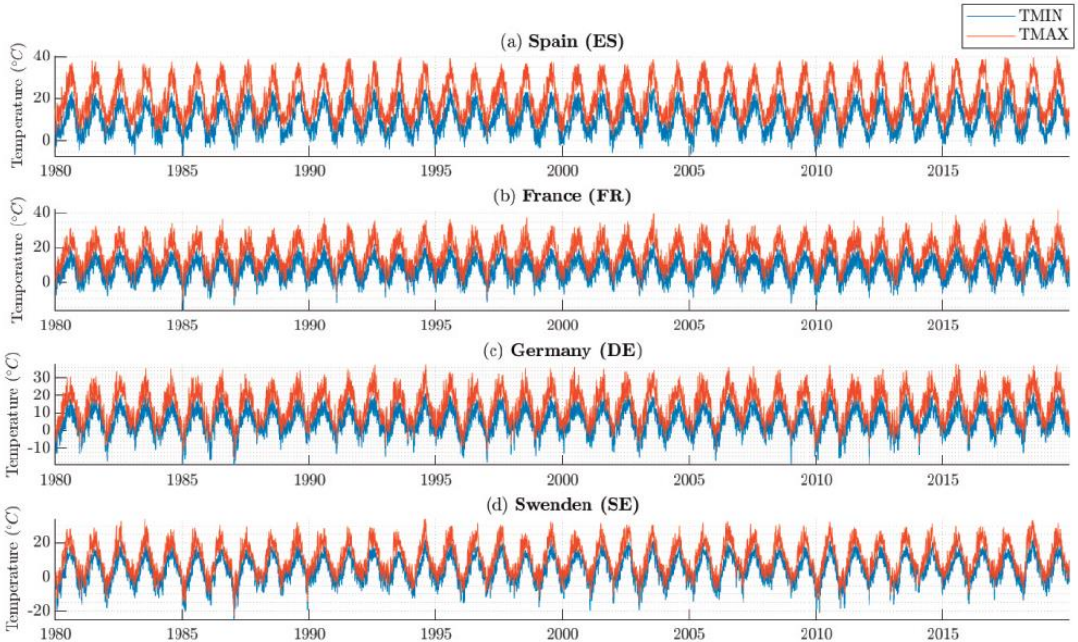


Figure 1. Minimum and maximum daily temperatures of four weather stations from Europe: (**a**) Madrid (Spain). (**b**) Paris (France). (**c**) Berlin (Germany). (**d**) Stockholm (Sweden).

"Analysis of the daily outdoor air temperatures for energy forecasting in the context of climate change". S. Moreno-Carbonell, E.F. Sánchez-Úbeda, A. Muñoz. Energies 2020, 13(7). (https://doi.org/10.3390/en13071569)

Decomposition methods
Example

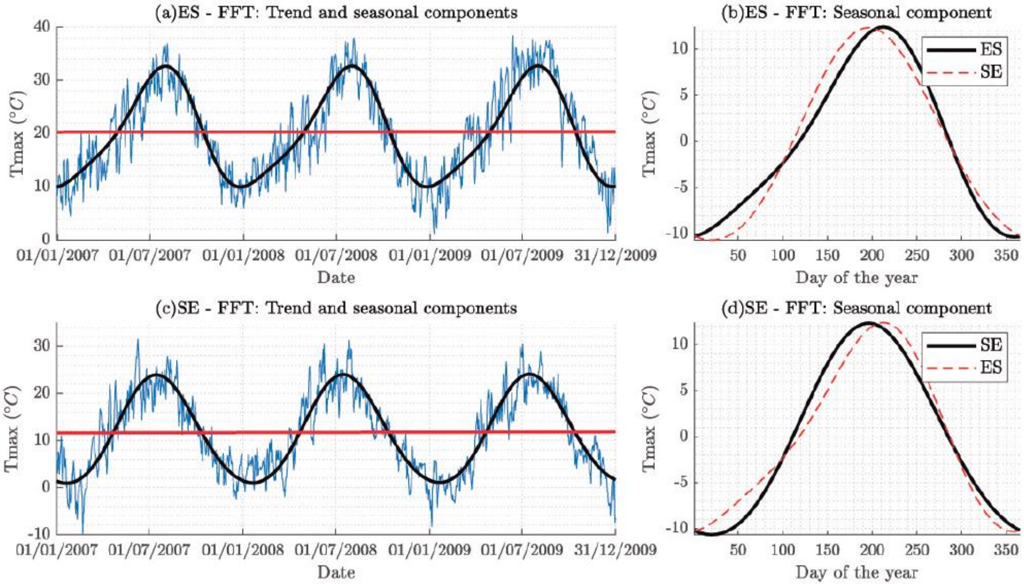


Figure 4. Trend and seasonal components estimated by the FFT model for (**a**) Spain, and (**c**) Sweden. (**b**) and (**d**) show the detail of the seasonal component estimated for each country.

Source: Moreno-Carbonell, Sánchez-Úbeda, Muñoz, 2020

Decomposition methods
Example

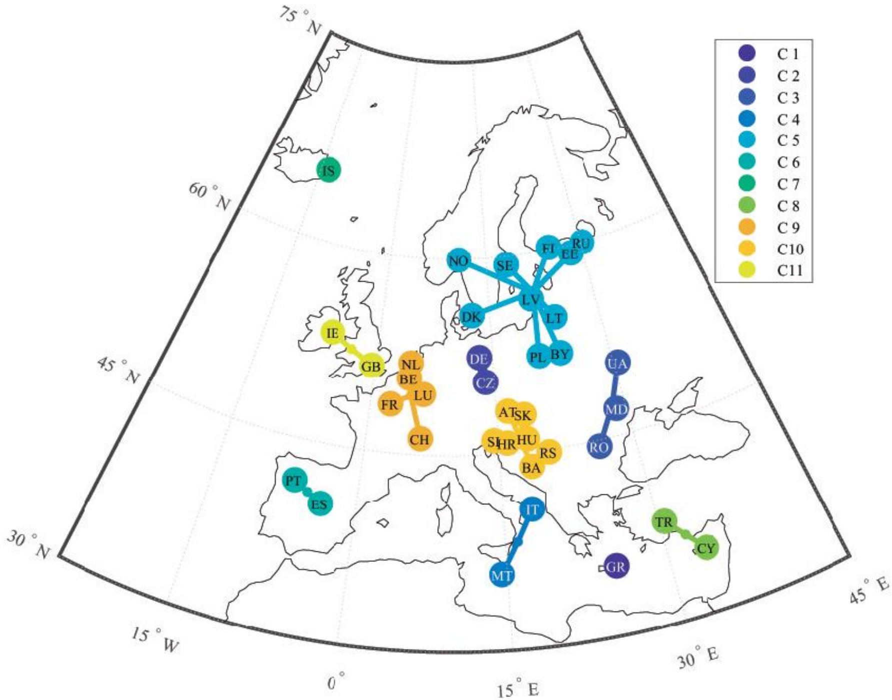


Figure 8. Location of the reference weather stations. The coloured clusters correspond to those formed using the dendrogram of Figure 7 (*Top*), based on the maximum temperature.

Source: Moreno-Carbonell, Sánchez-Úbeda, Muñoz, 2020

Decomposition methods
Example

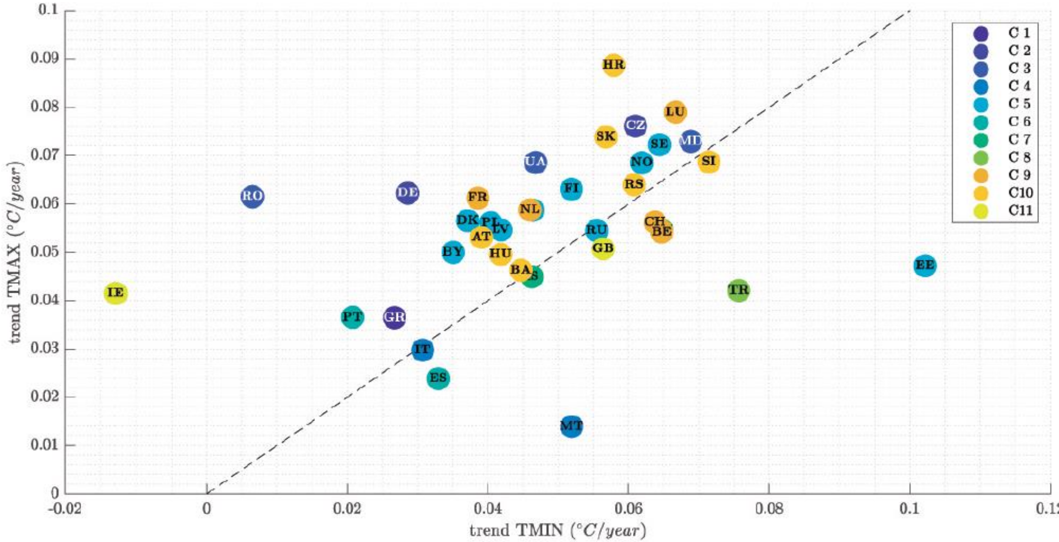


Figure 11. Scatterplot of the trends of the minimum and maximum temperatures. Each point represents a country, and colors indicate the cluster to which each point belongs according to Figure 7 (*bottom*). The black broken line represents the values of equal trend for minimum and maximum temperatures.

Source: Moreno-Carbonell, Sánchez-Úbeda, Muñoz, 2020

Decomposition methods
Example

(a) Annual increases in minimum temperatures (b) Annual increases in maximum temperatures

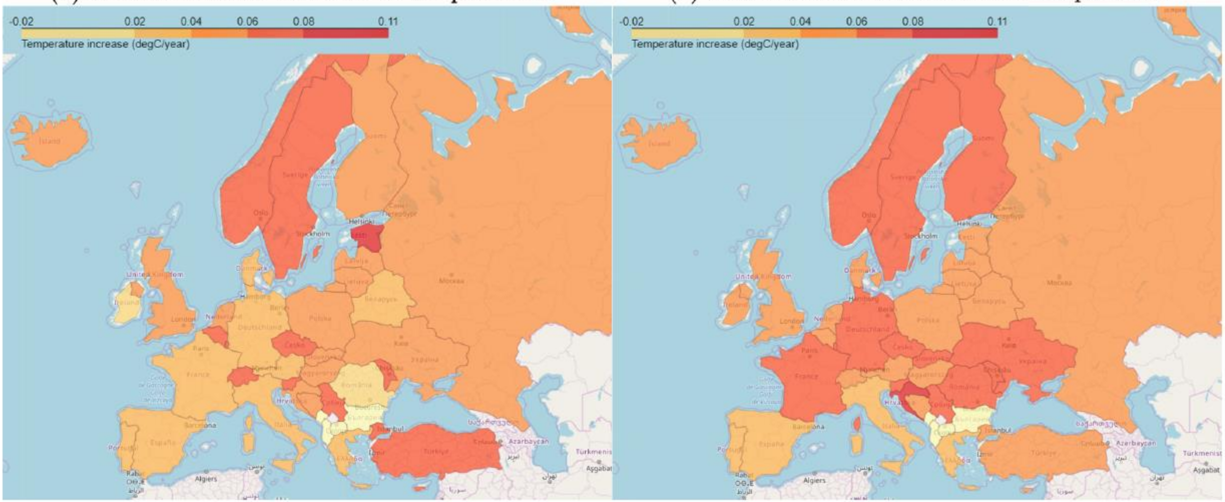


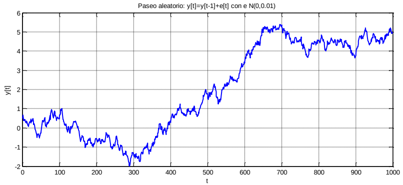
Figure 13. Annual increase of temperatures obtained from the trend component of the GAM for the **(a)** minimum and **(b)** maximum temperatures of the 37 countries.

Source: Moreno-Carbonell, Sánchez-Úbeda, Muñoz, 2020

Decomposition methods
General formulation

- **Additive Model**:

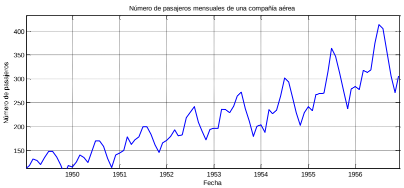
$$\boxed{y(t)=\tau(t)+\sigma(t)+\varepsilon(t)}$$



The additive model is appropiate if the magnitude of the seasonal fluctuations does nor vary with the level of the time series

- **Multiplicative Model**:

$$\boxed{y(t)=\tau(t) \times \sigma(t) \times \varepsilon(t)}$$



Multiplicative decomposition is more prevalent with economic series because most seasonal economic series do have seasonal variation which increases with the level of the series.

$$\Rightarrow \log(y(t))=\log(\tau(t))+\log(\sigma(t))+\log(\varepsilon(t))$$

Decomposition methods
General formulation

- **Pseudo-additive** decomposition: $\boxed{y(t) = \tau(t) \times (\sigma(t) + \varepsilon(t))}$

- **Example**: $D(t) = Pop(t) \times ConsPerCapita(t)$

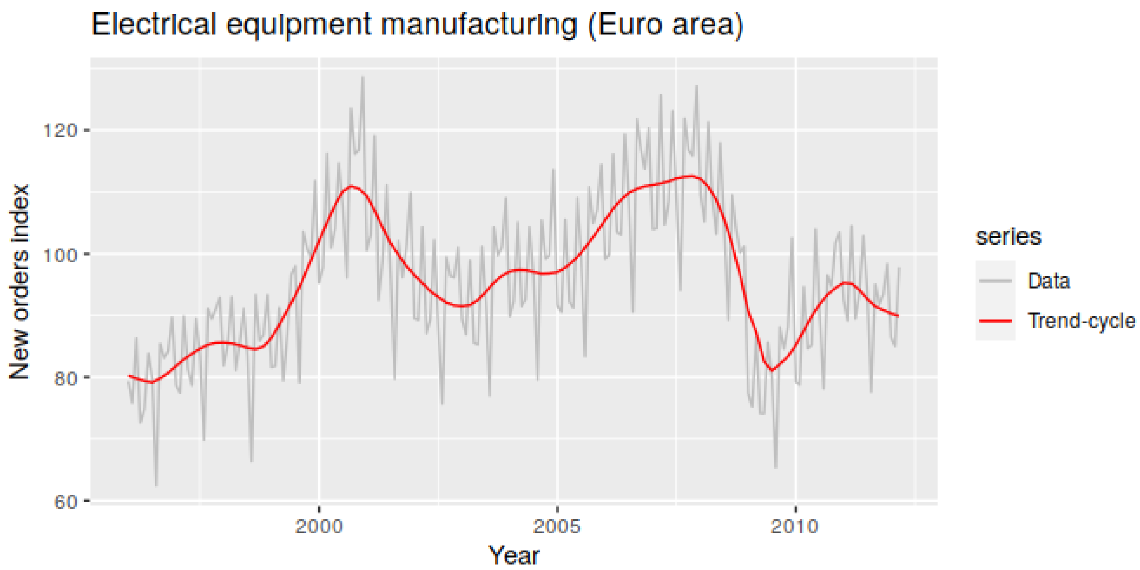
  with: $Pop(t) = \alpha t + \beta$

  $ConsPerCapita(t) = K + \Delta K sen(wt) + \varepsilon'(t)$

  Then: $$D(t) = \underbrace{K(\alpha t + \beta)}_{\tau(t)} \times (\underbrace{1 + \frac{\Delta K}{K} sen(wt)}_{\sigma(t)} + \underbrace{\frac{1}{K}\varepsilon'(t)}_{\varepsilon(t)})$$

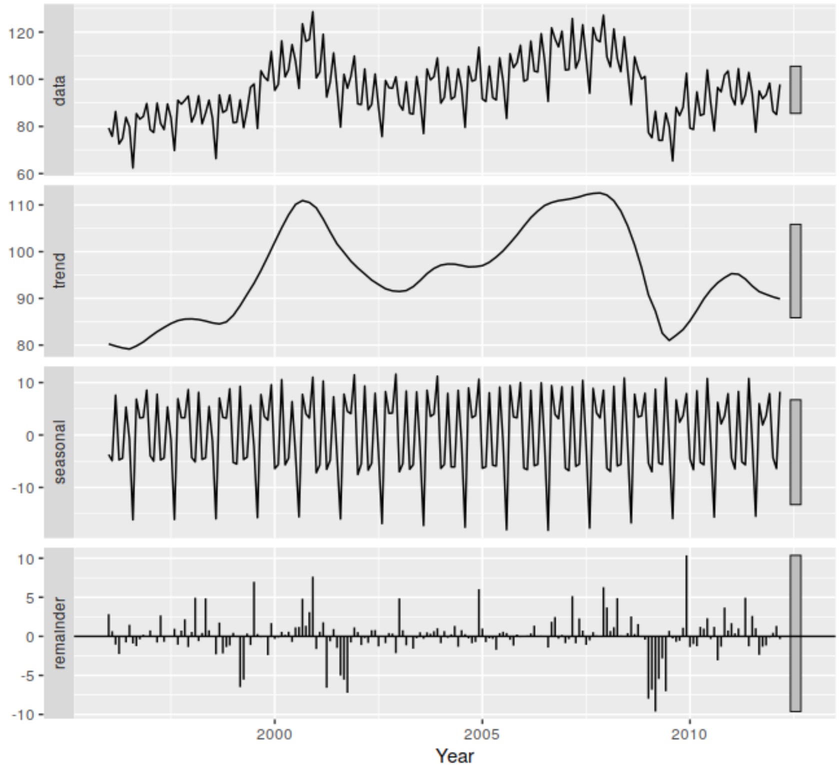
  $$= \tau(t) \times (\sigma(t) + \varepsilon(t))$$

Decomposition methods
Decomposition Chart



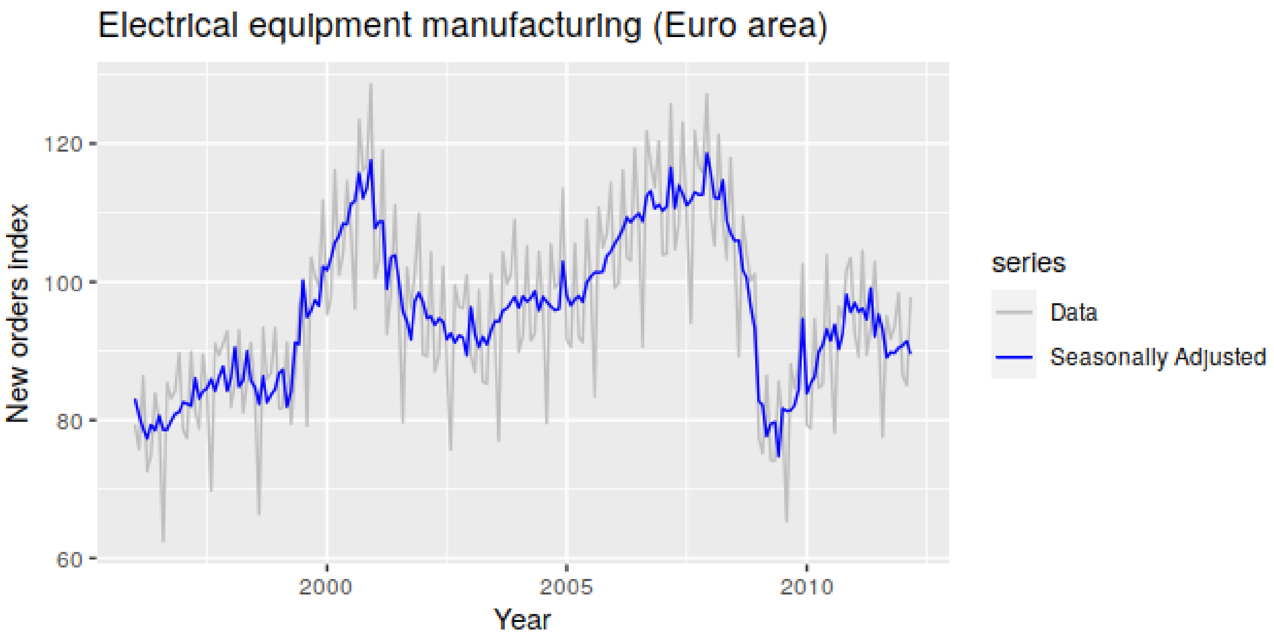
Source: *https://otexts.com/fpp3/*

Decomposition methods
Decomposition Chart



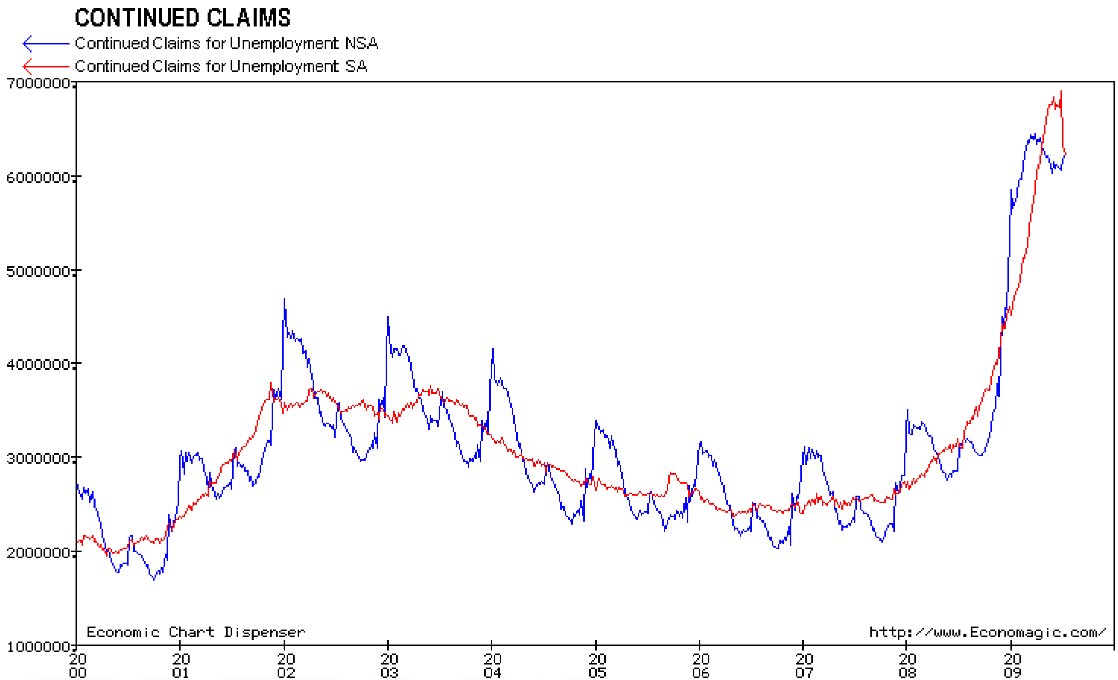
Source: https://otexts.com/fpp3/

Decomposition methods
Seasonal Adjustment



Source: https://otexts.com/fpp3/

Decomposition methods
Seasonal Adjustment



Decomposition methods
Additive Classical Decomposition

- **Additive** decomposition:

$$y(t) = \tau(t) + \sigma(t) + \varepsilon(t)$$

1) The trend-cycle is computed using a low-pass filter (centered *MA)*

2) The de-trended series is computed as:

$$y(t) - \tau(t) = \sigma(t) + \varepsilon(t)$$

3) The seasonal component, which is assumed to be constant from year to year, is estimated as the monthly average value of the de-trended series *r(t)*.

4) The irregular component is given by: $\varepsilon(t) = y(t) - \tau(t) - \sigma(t)$

Moving Averages

- **Moving Averages (MA)**: very simple smoothers that can be used to estimate the trend-cycle component.

- *MA_k*: Moving Average of order *k*, with *k* odd and *m=(k-1)/2*:

$$\tau(t) = \frac{1}{k} \sum_{j=-m}^{m} y(t+j)$$

- Moving Average of order 3:

$$\tau(t) = \frac{1}{3}\left(y(t-1) + y(t) + y(t+1)\right)$$

Decomposition methods
Moving Averages

- The order of the MA affects the smoothness of the resulting estimate:

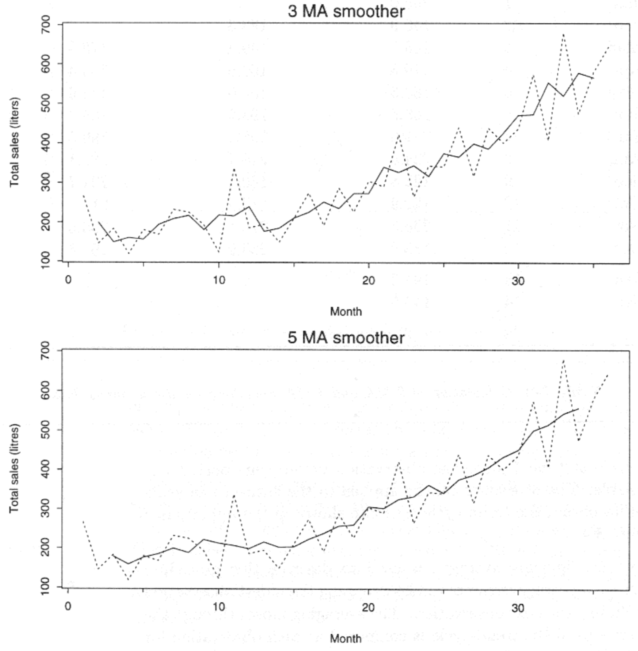


Figure 3-4: *3 MA and 5 MA smoothers for the shampoo data. The 3 MA smoother leaves too much randomness in the trend-cycle estimate. The 5 MA smoother is better, but the true trend-cycle is probably smoother still.*

Decomposition methods
Moving Averages

- It is not possible to estimate the trend-cycle close to the begining and end of the series.

- The *m* terms lost at the begining of the data are usually of little consequence, but those *m* terms lost in the end are critical.

- To overcome this problem a shorter length MA can be used in the begining and in the end. Example: MA_3

$$\tau(1) = \frac{1}{2}\left(y(1) + y(2)\right)$$

$$\tau(2) = \frac{1}{3}\left(y(1) + y(2) + y(3)\right)$$

$$\tau(3) = \frac{1}{3}\left(y(2) + y(3) + y(4)\right)$$

Decomposition methods
Centered Moving Averages

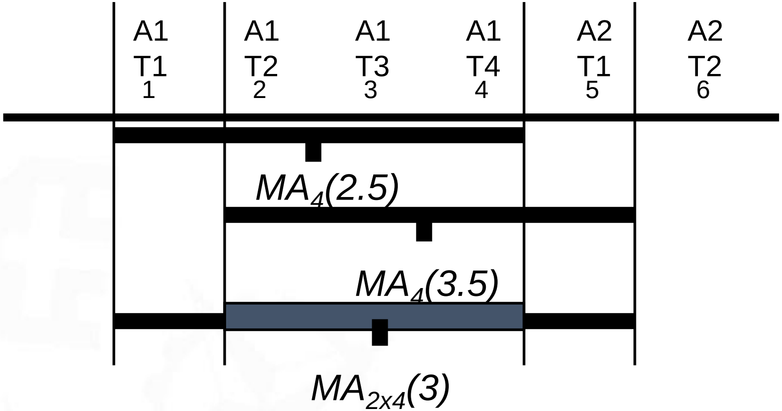
- The simple MA requires and odd number of observations to be included in the average.
- One possible solution if we wish to use a MA with *k=4*:

$$MA_4(2.5) = \tau(2.5) = \frac{1}{4}\left(y(1) + y(2) + y(3) + y(4)\right)$$

$$MA_4(3.5) = \tau(3.5) = \frac{1}{4}\left(y(2) + y(3) + y(4) + y(5)\right)$$

$$\begin{aligned} MA_{2x4}(3) &= \frac{\tau(2.5) + \tau(3.5)}{2} \\ &= \frac{1}{2}\left(\frac{y(1) + y(2) + y(3) + y(4)}{4} + \frac{y(2) + y(3) + y(4) + y(5)}{4}\right) \\ &= \frac{y(1) + 2y(2) + 2y(3) + 2y(4) + y(5)}{8} \end{aligned}$$

Decomposition methods
Centered Moving Averages



- The centered moving average of order 4 (MA_{2x4}) is equivalent to a weighted moving average of order 5.
- Very useful for estimating the trend-cycle in the presence of quaterly seasonality, as all quarters are given the same weight.
- Alternatives for quaterly data: MA_{2x8}, MA_{2x12}

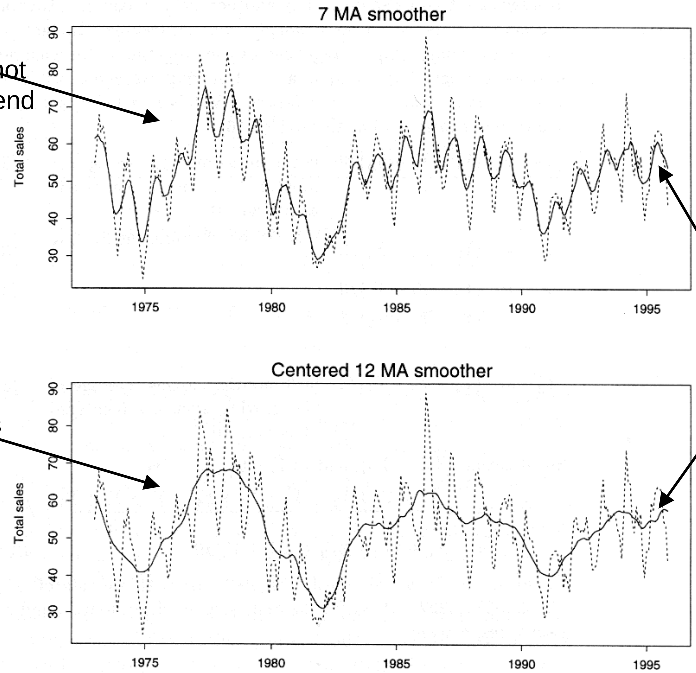
Decomposition methods
Centered Moving Averages

- In general, a MA_{2xk} is equivalent to a weighted MA of order $k+1$ and weights $1/k$ for all observations except for the first and the last observation in the average, which are assigned a weight of $1/(2k)$

- Example: $MA_{2x12} \Rightarrow$ 1/12 [0.5,1,1,1,1,1,1,1,1,1,1,1,0.5] could be used with monthly data, as it gives equal weights to all months.

Decomposition methods
Centered Moving Averages

The seasonal effect has not been removed from the trend



The seasonal effect has been removed

The trend-cycle at the end is contaminated by the seasonal component

Figure 3-5: *Moving averages applied to the housing sales data. The 7 MA tracks the seasonal variation whereas the 2×12 MA tracks the cycle without being contaminated by the seasonal variation.*

Decomposition methods
Weighted Moving Averages

- In general, a weighted (odd) *k*-point moving average can be written as:

$$\tau(t) = \sum_{j=-m}^{m} a_j y(t+j) \qquad \text{with } m=(k-1)/2$$

Name	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
3 MA	.333	.333										
5 MA	.200	.200	.200									
2 × 12 MA	.083	.083	.083	.083	.083	.083	.042					
3 × 3 MA	.333	.222	.111									
3 × 5 MA	.200	.200	.133	.067								
S15 MA	.231	.209	.144	.066	.009	-.016	-.019	-.009				
S21 MA	.171	.163	.134	.037	.051	.017	-.006	-.014	-.014	-.009	-.003	
H5 MA	.558	.294	-.073									
H9 MA	.330	.267	.119	-.010	-.041							
H13 MA	.240	.214	.147	.066	.000	-.028	-.019					
H23 MA	.148	.138	.122	.097	.068	.039	.013	-.005	-.015	-.016	-.011	-.004

S = Spencer's weighted moving average
H = Henderson's weighted moving average

Table 3-4: *Weight functions a_j for some common weighted moving averages.*

Decomposition methods
Multiplicative Classical Decomposition

- **Multiplicative** decomposition:

$$y(t) = \tau(t) \times \sigma(t) \times \varepsilon(t)$$

1) The trend-cycle is computed using a low-pass filter (centered *MA)*

2) The de-trended series is computed as the ratio:

$$r(t) = y(t)/\tau(t) = \sigma(t) \times \varepsilon(t)$$

3) The seasonal component, which is assumed to be constant from year to year, is estimated as the monthly average value of the de-trended series *r(t)*.

4) The irregular component is given by: $\varepsilon(t) = y(t) / (\tau(t) \times \sigma(t))$

Decomposition methods
X11/12/13-ARIMA

- X11, X12 and X13 ARIMA (Findley et al, 1997) are the most widely used variants of the Census II method developed by the U.S. Bureau of the Census. https://www.census.gov/srd/www/x13as/

- Census II decomposition is usually multiplicative, since most economic time series have seasonal variation which increases with the level of the series.

- It is an iterative procedure in which the decomposition is refined. The algorithm also minimizes the effect of outliers.

Decomposition methods X11/12/13-ARIMA

https://www.census.gov/srd/www/x13as/

U.S. Department of Commerce | Blogs | Index A-Z | Glossary | FAQs

Topics — Population, Economy | Geography — Maps, Products | Library — Infographics, Publications | Data — Tools, Developers | Surveys/Programs — Respond, Survey Data | Newsroom — News, Blogs | About Us — Our Research

You are here: Census.gov › Subjects A to Z › X-13ARIMA-SEATS

X-13ARIMA-SEATS Seasonal Adjustment Program

X-13ARIMA-SEATS Home | Win X-13 & X-13-Data | X-13-SAM | X-13-Graph | Win Genhol | SA Papers | References | Contact

Downloads:

- Download Win X-13
- Download X-13 (PC)
- Download X-13 (Linux)
- Download X-13-Data
- Download X-13-SAM
- Download X-13-Graph Batch
- Download X-13-Graph Java
- Download Win Genhol

Stay Informed About X-13ARIMA-SEATS

Announcements of program updates and other information related to X-13ARIMA-SEATS, Win X-13, and other programs are distributed through a moderated mailing list called **x12a-announce**.

Access this link to subscribe.

Contact Us

Contact X-13ARIMA-SEATS Support: Email us!
Call us: (301)763-1649

The X-13ARIMA-SEATS Seasonal Adjustment Program

X-13ARIMA-SEATS is a seasonal adjustment software produced, distributed, and maintained by the Census Bureau.

Features include:

- Extensive time series modeling and model selection capabilities for linear regression models with ARIMA errors (regARIMA models);
- The capability to generate ARIMA model-based seasonal adjustment using a version of the SEATS procedure originally developed by Victor Gómez and Agustín Maravall at the Bank of Spain as well as nonparametric adjustments from the X-11 procedure;
- Diagnostics of the quality and stability of the adjustments achieved under the options selected;
- The ability to efficiently process many series at once.

There are distributions of X-13ARIMA-SEATS for Windows® PC and Linux/Unix platforms. For Windows systems, the Census Bureau's **Win X-13** interface program is recommended, but it is also possible to execute the program using command line input. The **X-13-Data** program provides an interface between X-13-ARIMA and Excel users. **X-13-SAM** lets users quickly make changes to many X-13ARIMA-SEATS spec files at one time.

We are indebted to Agustín Maravall (formerly with the Bank of Spain), Gianluca Caporello, and contractors at the Bank of Spain for their collaboration and advice in the development of this software.

In addition, a SAS/Graph® program named **X-13-Graph** is available that allows users to generate useful graphical diagnostics from X-13ARIMA-SEATS (as well as X-12-ARIMA) output. It requires no knowledge of SAS®, and can generate more types of graphs than Win X-13. There is a Java version of the software available as well.

Recent News

- Version 1.0 of X-13-SAM was released on March 24, 2017.
- Updated versions of several software packages were released on March 10, 2017:
  - PC and Linux/Unix versions of the latest release of X-13ARIMA-SEATS (Build 39 of Version 1.1, notes on the release available on the site);
  - Version 2.5 of Win X-13, the windows interface to X-13ARIMA-SEATS.
- Papers were added to the Seasonal Adjustment Papers website on February 17, 2017.
- Updated versions of several software packages were released on March 1, 2016:
  - PC and Linux/Unix versions of the latest release of X-13ARIMA-SEATS (Build 26 of Version 1.1, notes on the release available on the site);
  - Version 2.4 of Win X-13, the windows interface to X-13ARIMA-SEATS.
- Licensing information and disclaimer for the various software products has been updated (March 1, 2016).
- Papers were added to the Seasonal Adjustment Papers website on August 7, 2015.
- Updated versions of several software packages were released on April 2, 2015:
  - PC and Linux/Unix versions of the latest release of X-13ARIMA-SEATS (Build 19 of Version 1.1, notes on the release available on the site);
  - Version 2.3 of Win X-13, the windows interface to X-13ARIMA-SEATS;
  - Version 2.0 of X-13-Data;

Decomposition methods
Forecasting and Decomposition

- Forecasts based directly on a decomposition are performed by extending each of the components of the series.

- In practice it rarely works well:
  - The trend-cycle is the most difficult component to forecast. It is sometimes proposed to be modeled by a simple function as a straight line, but such models are rarely adequate.

  - The seasonal component for future years can be based on the seasonal component from the last full period of data. But if the seasonal component is changing over time, this will be unlikely to be adequate.

  - The irregular component may be forecast as zero (for additive decomposition) or one (for multiplicative decomposition). This assumes that the irregular component is serially uncorrelated, which is not often the case.

$\Rightarrow$Decomposition methods$\equiv$ Exploratory methods

Decomposition methods
R implementation

- Classical additive and multiplicative decomposition can be computed with function `decompose()`

```
#Additive decomposition
dec_add <- decompose(fdata_ts,
           type="additive")
autoplot(dec_add)  #Plot components
```

```
#Multiplicative decomposition
dec_mult <- decompose(fdata_ts,
           type=“multiplicative")
autoplot(dec_mult)  #Plot components
```

- For SEATS decomposition, use package *seasonal* and function `seas().`

```
#SEATS decomposition
Library(seasonal)
dec_seas <- seas(fdata_ts)
autoplot(dec_seas)  #Plot components
```

- The package *seasonalview* and view() function is a graphical tool for choosing a seasonal adjustment model.
- Use `seasonal()`, `trendcycle()` and `remainder()` functions to extract the individual components.
- Use `seasadj()` to compute the seasonally adjusted time series.

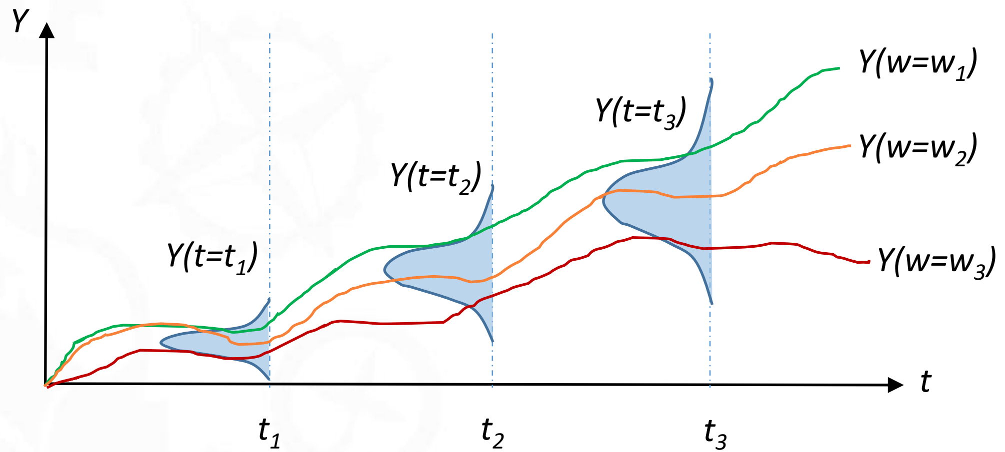
3

Stochastic processes

Fundamental concepts
Time series and Stochastic Processes

- A **stochastic process** *Y(w,t)* is a family of time indexed random variables.

belongs to a sample space *belongs to an index set*

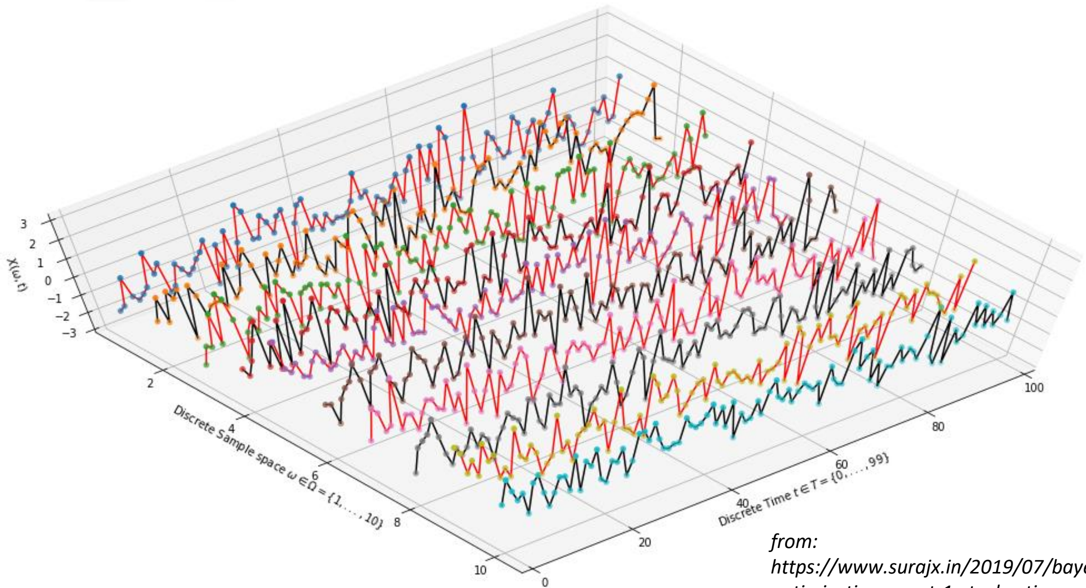


Fundamental concepts
Time series and Stochastic Processes

- A **stochastic process** *Y(w,t)* is a family of time indexed random variables.

belongs to a sample space

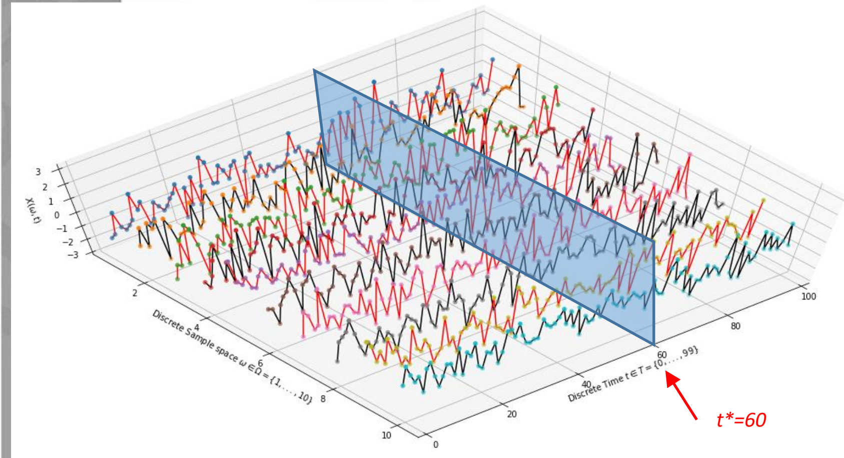
belongs to an index set

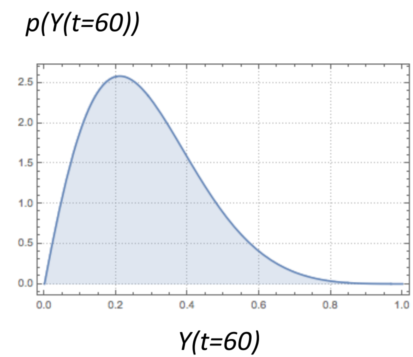


from: https://www.surajx.in/2019/07/bayesian-optimization---part-1-stochastic-processes/

Fundamental concepts
Time series and Stochastic Processes

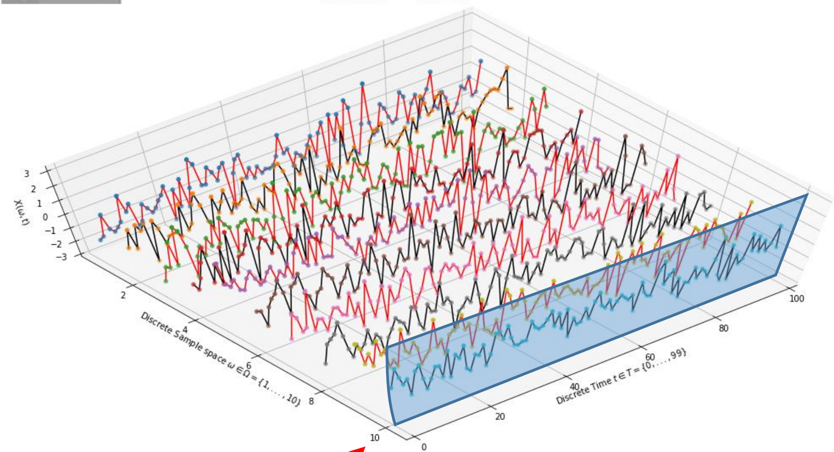
- For a fixed *t**, *Y(w,t*)* is a **random variable.**



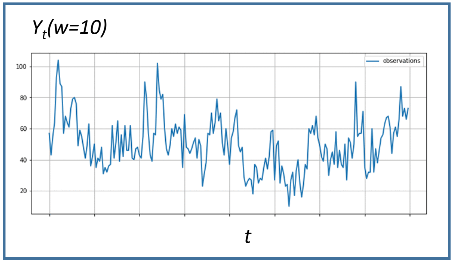


Fundamental concepts
Time series and Stochastic Processes

- For a given *w**, *Y(w*,t)* is called a sample function or **realization** of the stochastic process.

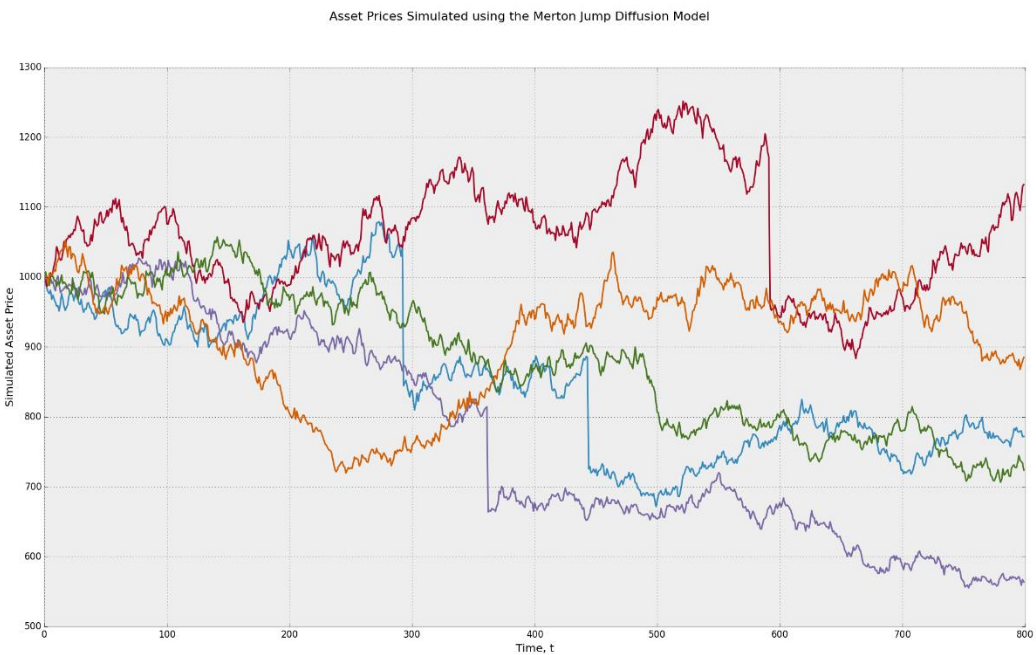


w=10*



- The population of all possible realizations is called the **ensemble**.

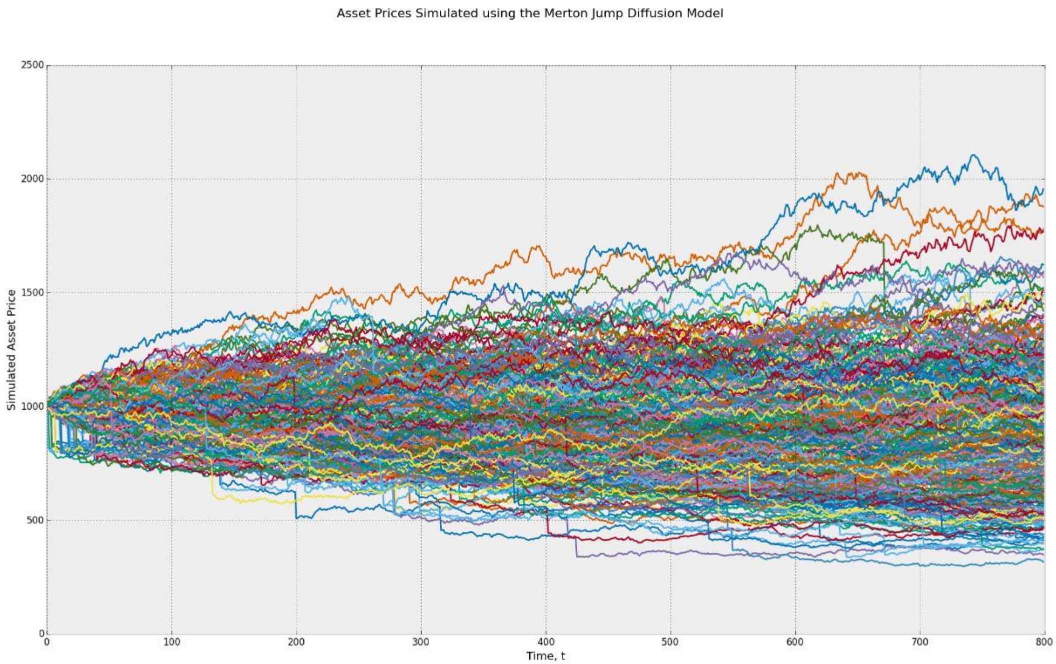
Fundamental concepts
Example



5 realizations

from http://www.turingfinance.com/random-walks-down-wall-street-stochastic-processes-in-python/

Fundamental concepts
Example

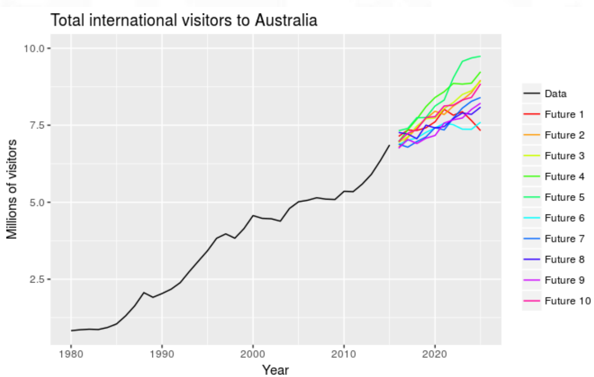


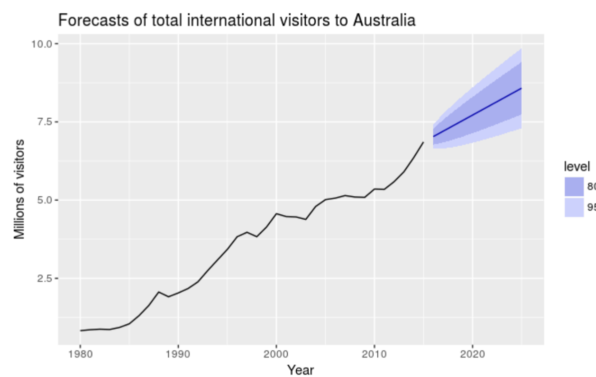
500 realizations

from http://www.turingfinance.com/random-walks-down-wall-street-stochastic-processes-in-python/

Fundamental concepts
Time Series

- A **time series** is a collection of **observations** made sequentially through time.
- Formal definition: a time series is the **realization** of a discrete **stochastic process**





Source Hyndman et al. (2017)

Means, Variances and Covariances

- For a stochastic process *{y[t], t=0,±1, ±2...}*, the **mean function** is defined by:

$$\mu_t = E(y[t])$$

- The **autocovariance function** $\gamma_{t,s}$ is defined as:

$$\gamma_{t,s} = Cov(y[t], y[s]) = E((y[t] - \mu_t)(y[s] - \mu_s))$$

- The **autocorrelation function** $\rho_{t,s}$ is defined as:

$$\rho_{t,s} = Corr(y[t], y[s]) = \frac{Cov(y[t], y[s])}{\sqrt{Var(y[t])Var(y[s])}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

Means, Variances and Covariances

- Properties:

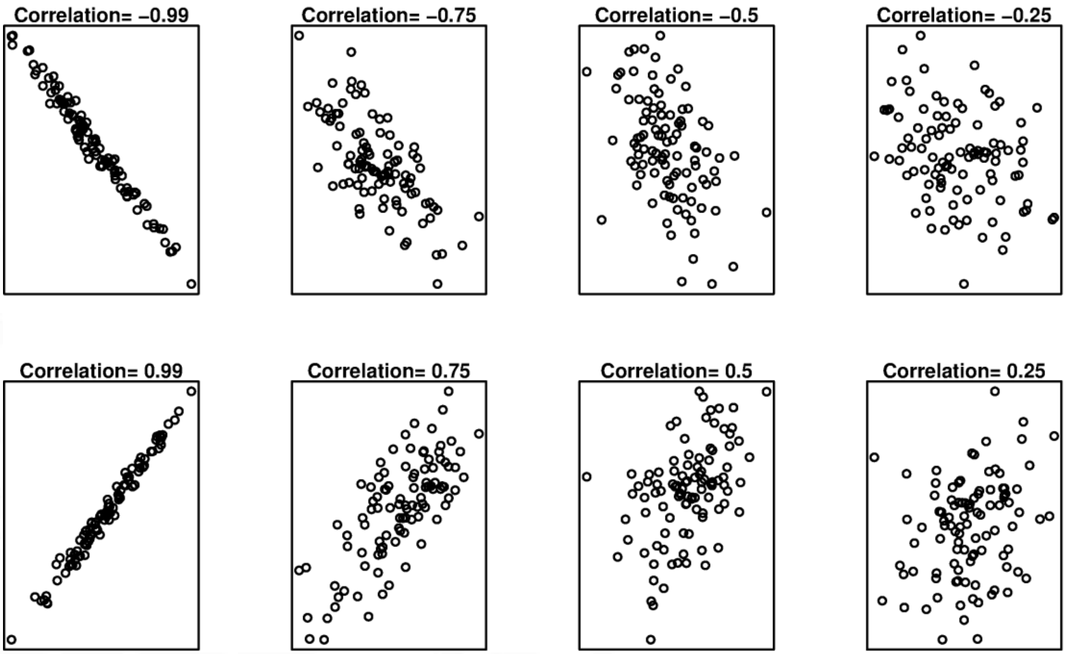
$$\boxed{\gamma_{t,t} = Var(y[t])} \qquad \rho_{t,t} = 1$$

$$\gamma_{t,s} = \gamma_{s,t} \qquad \rho_{t,s} = \rho_{s,t}$$

$$|\gamma_{t,s}| \leq \sqrt{\gamma_{t,t}\,\gamma_{s,s}} \qquad \boxed{|\rho_{t,s}| \leq 1}$$

- If $\rho_{t,s} = 0$ we say that $y[t]$ and $y[s]$ are uncorrelated

Fundamental concepts
Correlation

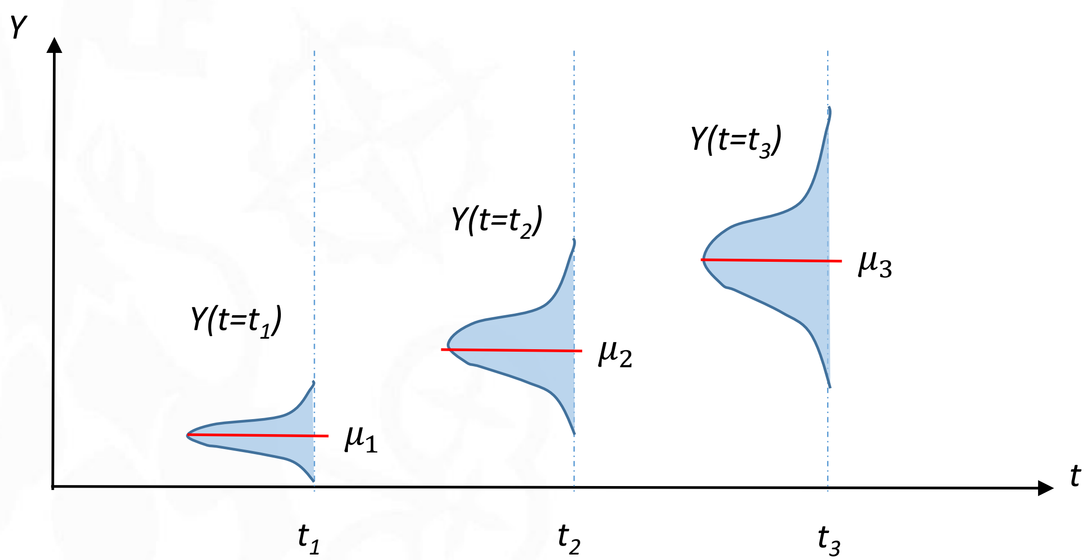


From [Hyndman & Athanasopoulos, 2013]

Fundamental concepts
Means, Variances and Covariances

$$\mu_t = E(y[t])$$

$$\gamma_{t,s} = Cov(y[t], y[s]) = E((y[t] - \mu_t)(y[s] - \mu_s))$$



Fundamental concepts
Stationary Processes

- A process is said to be **stationary** when the properties of the underlying process do not vary with time

- We say that a process **is stationary in the strict sense**, when to make the same shift in the timing of all the variables of any finite joint distribution, the distribution does not vary:

$$p(y[t_1], \dots, y[t_N]) = p(y[t_1 + k], \dots, y[t_N + k]) \ \ \forall k$$

for all finite set $\{y[t_1],\dots,y[t_N]\}$

$$\Longrightarrow \begin{cases} p(y[t]) = p(y[t+k]) \quad \forall t, k \\ p(y[t], y[s]) = p(y[t+k], y[s+k]) \quad \forall t, s, k \\ \dots \end{cases}$$

Fundamental concepts
Stationary Processes

- A process is said to be **first order stationary** in distribution if:

$$p(y[t]) = p(y[t+k]) \quad \forall k$$

$$\Rightarrow E(y[t]) = \mu \quad \forall t$$

- A process is said to be **second order** or **wide sense stationary** if it satisfies:

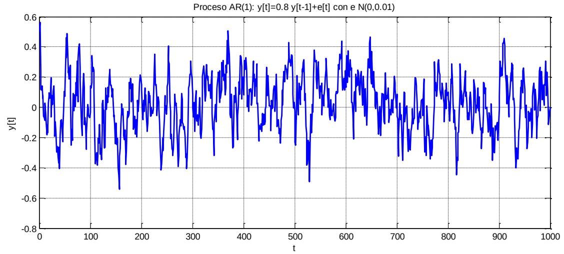
$$E(y[t]) = \mu \quad \forall t$$

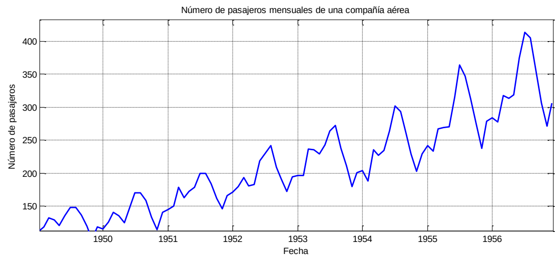
$$E\left((y[t]-\mu)^2\right) = \sigma^2 < \infty \quad \forall t$$

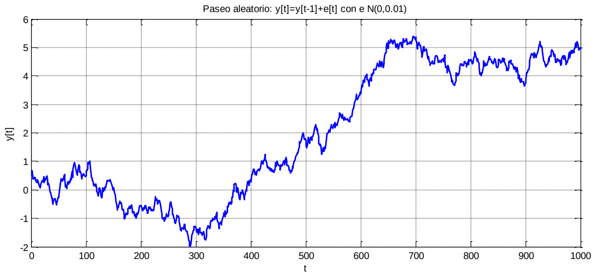
$$E\left((y[t+k]-\mu)(y[t]-\mu)\right) = \gamma_k \quad \forall t$$

Fundamental concepts
Stationary Processes

- Examples:







Fundamental concepts
Stationary Processes

- When the process is stationary, its first and second order moments can be estimated from **only one realization** of the process:
  - Mean:

$$\mu = E\left[y[t]\right] \rightarrow \hat{\mu} = \frac{1}{N}\sum_{t=1}^{N} y[t]$$

  - Autocovariance:

$$\gamma_k = E\left[(y[t+k]-\mu)(y[t]-\mu)\right]$$

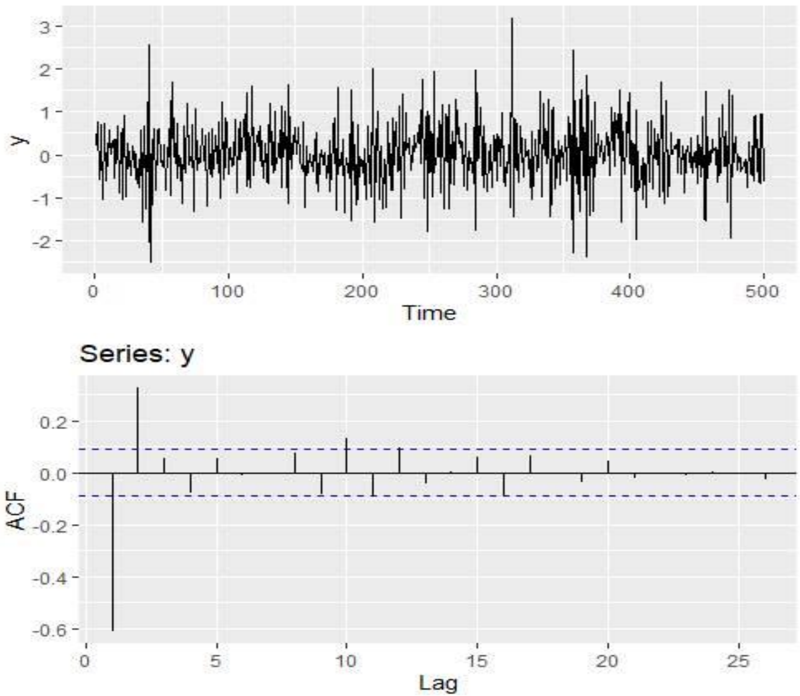
$$\hat{\gamma}_k = \frac{1}{N}\sum_{t=1}^{N-k}(y[t+k]-\hat{\mu})(y[t]-\hat{\mu})$$

  - Autocorrelation:

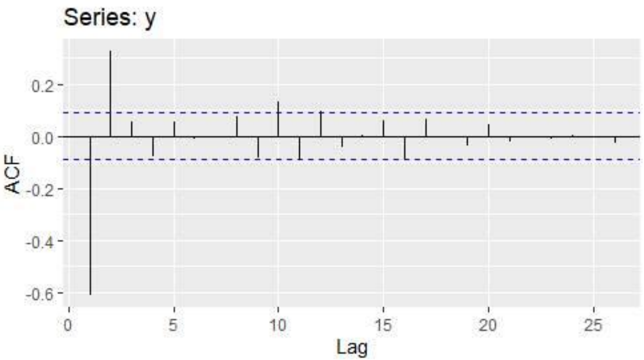
$$\rho_k = \frac{\gamma_k}{\gamma_0} \rightarrow \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{N-k}(y[t+k]-\bar{y})(y[t]-\bar{y})}{\sum_{t=1}^{N}(y[t]-\bar{y})^2}$$

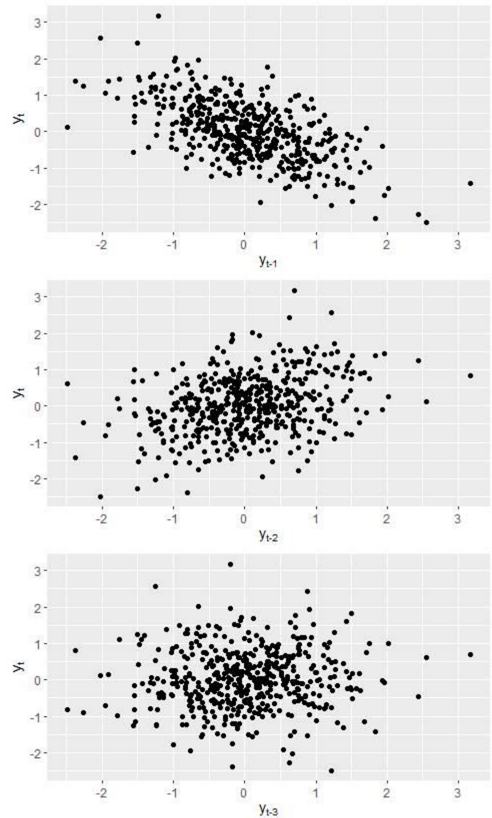
Fundamental concepts
Stationary Processes

- Correlogram $=\{\hat{\rho}_k\}$ for $k=1,\ldots$ (not recommended for $k>N/4$)



Fundamental concepts
Stationary Processes





t	y(t)	y(t-1)	y(t-2)	y(t-1)
1	0.1531	NA	NA	NA
2	-0.4074	0.1531		NA
3	0.0383	-0.4074	0.1531	NA
4	-0.0557	0.0383	-0.4074	0.1531
5	0.0899	-0.0557	0.0383	-0.4074
6	-0.0641	0.0899	-0.0557	0.0383
7	0.4666	-0.0641	0.0899	-0.0557
8	0.1722	0.4666	-0.0641	0.0899
9	0.0705	0.1722	0.4666	-0.0641
10	-0.0123	0.0705	0.1722	0.4666

Fundamental concepts
Measures of forecast accuracy

- R-squared statistic: $$R^2 = 1 - \frac{\sum_{t=1}^{N}(y[t] - \hat{y}[t])^2}{\sum_{t=1}^{N}(y[t] - \bar{y})^2}$$

- Adjusted R-squared statistic: $$R_{Adj}^2 = 1 - \frac{\sum_{t=1}^{N}(y[t] - \hat{y}[t])^2/(N-M)}{\sum_{t=1}^{N}(y[t] - \bar{y})^2/(N-1)}$$

- Akaike Information Criterion: $$AIC = N\, ln(\hat{\sigma}_{\varepsilon}^2) + 2M$$

- Schwarz Bayesian Criterion: $$SBC = N\, ln(\hat{\sigma}_{\varepsilon}^2) + M\, ln(N)$$

where *M* is the number of parameters

Fundamental concepts
Measures of forecast accuracy

Forecast error obtained from a benchmark (random walk -> last observation)

Scale-dependent measures

Mean Square Error (MSE) $= \text{mean}(e_t^2)$

Root Mean Square Error (RMSE) $= \sqrt{\text{MSE}}$

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

Median Absolute Error (MdAE) $= \text{median}(|e_t|)$.

Measures based on percentage errors $\quad p_t = 100 e_t / Y_t$.

Mean Absolute Percentage Error (MAPE)

$= \text{mean}(|p_t|)$

Median Absolute Percentage Error (MdAPE)

$= \text{median}(|p_t|)$

Root Mean Square Percentage Error (RMSPE)

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

Root Median Square Percentage Error (RMdSPE)

$= \sqrt{\text{median}(p_t^2)}$

Symmetric Mean Absolute Percentage Error (sMAPE)

$= \text{mean}(200|Y_t - F_t|/(Y_t + F_t))$

Symmetric Median Absolute Percentage Error (sMdAPE)

$= \text{median}(200|Y_t - F_t|/(Y_t + F_t))$

Measures based on relative errors $\quad r_t = e_t / e_t^*$

Mean Relative Absolute Error (MRAE) $= \text{mean}(|r_t|)$

Median Relative Absolute Error (MdRAE)

$= \text{median}(|r_t|)$

Geometric Mean Relative Absoluate Error (GMRAE)

$= \text{gmean}(|r_t|)$

Relative measures

RelMAE $= \text{MAE}/\text{MAE}_b$.

MAE from the benchmark

Scaled errors

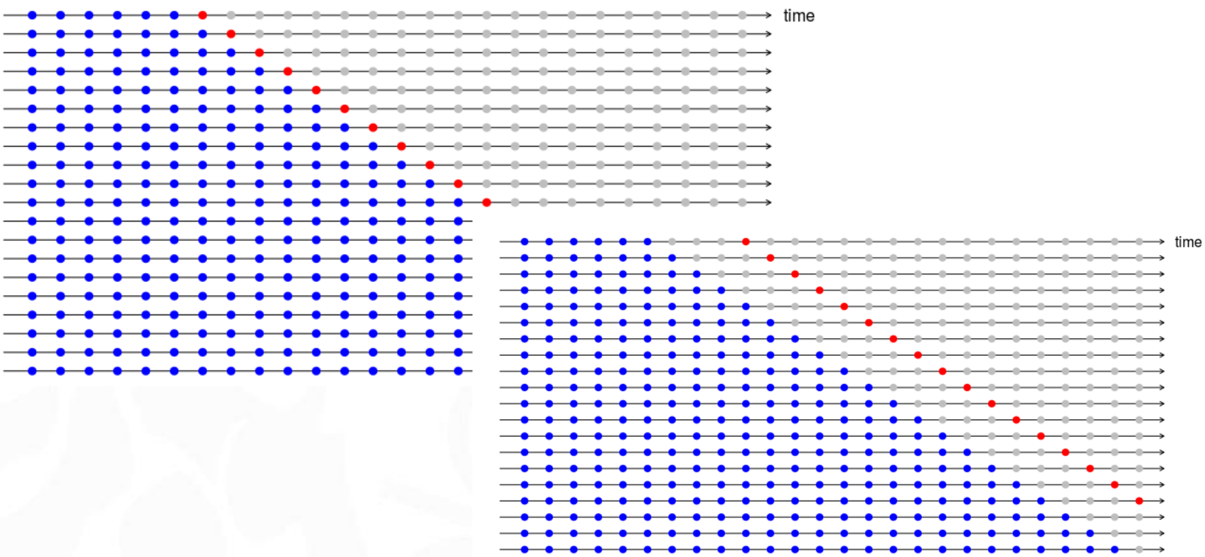
$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$$

MASE $= \text{mean}(|q_t|)$.

“Another look at measures of forecast accuracy”
Rob J. Hyndman and Anne B. Koehler
International Journal of Forecasting
Volume 22, Issue 4, October-December 2006, Pages 679-688

Measures of forecast accuracy

- Cross-validation methods:
  - Training set ("in-sample") → parameter optimization
  - Validation set ("out-of-sample") → measuring the generalization capabilities of the model



Source: https://robjhyndman.com/hyndsight/tscv/

Mathematical transformations and adjustments

- It will often be necessary to transform and/or adjust the series under study to fulfill the model assumptions (constant level, constant variance, normally distributed, ...)

- Lets analyze the following transformations:
  - Mathematical transformations
  - Calendar Adjustments
  - Adjustments for inflation and population growth

Mathematical transformations and adjustments

- Mathematical transformations:
  - Square root (*): $$w(t) = \sqrt{y(t)}$$
  - Cubic root: $$w(t) = \sqrt[3]{y(t)}$$
  - Logarithm (*): $$w(t) = log(y(t))$$
  - Inverse: $$w(t) = -\frac{1}{y(t)}$$

*A constant is added to *y(t)* to ensure it is greater than 1

Fundamental concepts
Mathematical transformations and adjustments

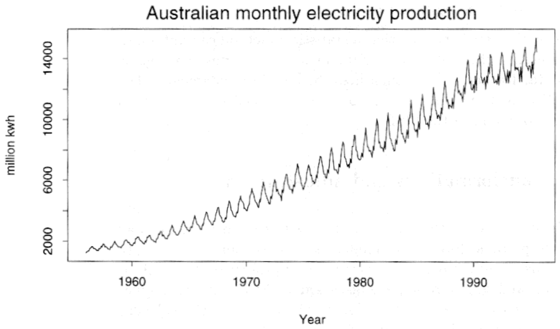


Figure 2-10: *Monthly Australian electricity production from January 1956 to August 1995. Note the increasing variation as the level of the series increases.*

See Makridakis et al. (1998)

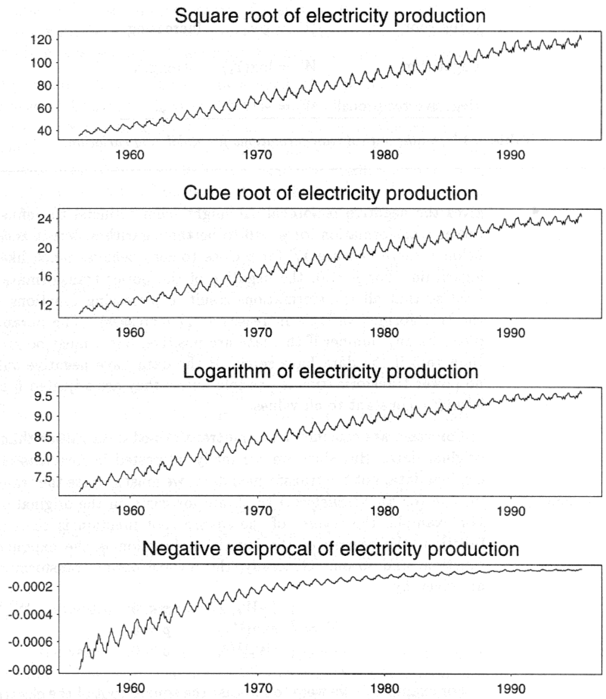


Figure 2-11: *Transformations of the electricity production data. Of these, either a square root, cube root, or a log transformation could be used to stabilize the variation to form a series which has variation approximately constant over the series.*

Mathematical transformations and adjustments

- The above transformations can be generalized in the form proposed by Box & Cox:

$$w(t) = \begin{cases} \log(y(t)), & \lambda = 0 \\ \dfrac{\left(y(t)^{\lambda} - 1\right)}{\lambda}, & \lambda \neq 0 \end{cases}$$

```
lambda <- BoxCox.lambda(y)
y_transf <-BoxCox(y,lambda)
```

see https://otexts.org/fpp2/transformations.html

- In practice, the square root and the logarithm are the transformations most frequently used.

Mathematical transformations and adjustments

- Calendar adjustments:
  - **Month duration**: the differences from month to month can reach

$$\frac{31-28}{31} \approx 10\%$$

  - This effect can be corrected for monthly forecasts by using:

$$w(t) = y(t)\frac{\text{mean number of days per month}}{\text{number of days in month } t}$$

$$= y(t)\frac{365.25/12}{\text{number of days in month } t}$$

```
Function monthdays(ts) can be used in R
to obtain the number of days for each
month in the time series.
```

Fundamental concepts
Mathematical transformations and adjustments

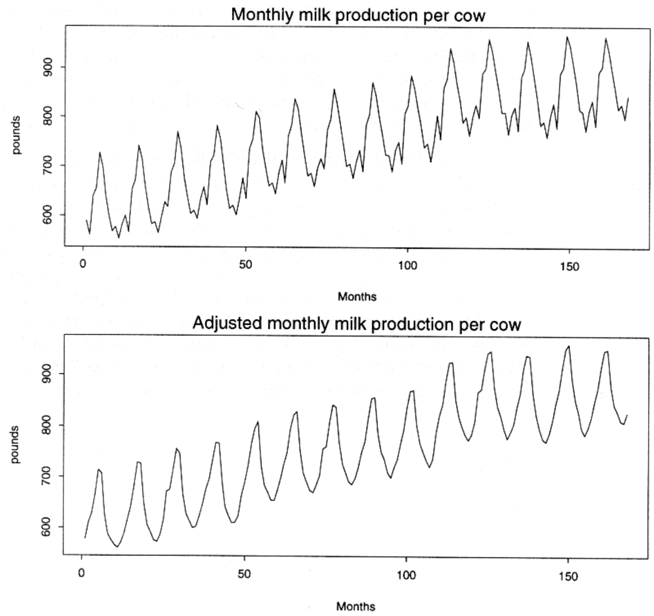


Figure 2-12: *Monthly milk production per cow over 14 years (Source: Cryer, 1986). The second graph shows the data adjusted for the length of month. This yields a simpler pattern enabling better forecasts and easier identification of unusual observations.*

See Makridakis et al. (1998)

Mathematical transformations and adjustments

- Calendar adjustments:
  - **Holidays**: the number of holidays per month is very different from month to month.

    If it is possible to classify the working days and holidays, and all holidays have the same effect, the following adjustment simplifies the problem of forecasting monthly time series:

$$w(t) = y(t) \frac{\text{mean number of holidays per month}}{\text{number of holidays in month } t}$$

Mathematical transformations and adjustments

- Adjustments for inflation and population growth:
  - **Inflation**: it is necessary to take it into account when predicting prices. For that purpose prices are referred to the same date.
  - **Population growth**: it is necessary to take it into account when predicting series as the number of users of public transport. If demographic studies are available, it is preferable to normalize the series and predict the proportion of users.

4

Exponential Smoothing methods

Exponential smoothing mehtods
Simple Exponential Smoothing

- Formulation: $\hat{y}(t+1) = \hat{y}(t) + \alpha(y(t) - \hat{y}(t))$

  where α is a constant between 0 and 1.

- Comments:
  - Not suitable for time series with trend
  - Analogy with proportional control
  - Requires very low storage space $\Rightarrow$ suitable when the number of series is very high

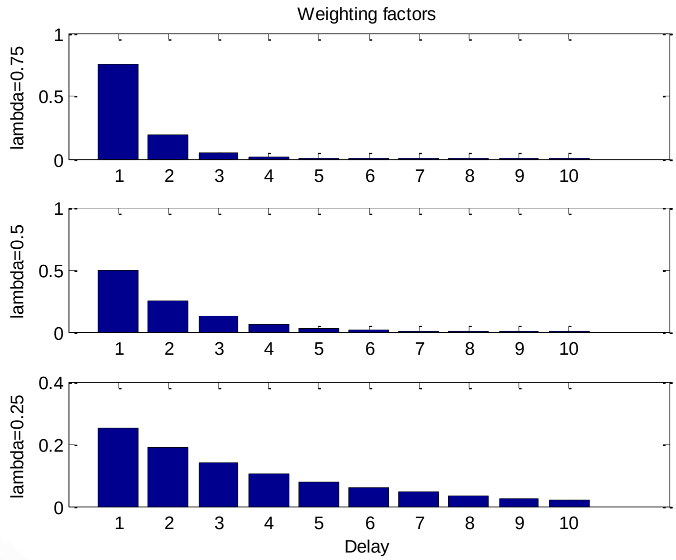
- The above expression can be put in the weighted average form:

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

Exponential smoothing mehtods
Simple Exponential Smoothing

- Developing the previous expression:

$$\begin{aligned}\hat{y}(t+1) = \alpha y(t) &+ \\ \alpha(1-\alpha) y(t-1) &+ \\ \alpha(1-\alpha)^2 y(t-2) &+ \\ \alpha(1-\alpha)^3 y(t-3) &+ \\ \ldots\end{aligned}$$



$$\hat{y}(t+1) = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j y(t-j) + (1-\alpha)^t L(0)$$

$\Rightarrow$Geometric progression

Decomposition methods
Simple Exponential Smoothing

- Component form:

  - Smoothing equation: $L(t) = \alpha y(t) + (1-\alpha)L(t-1)$

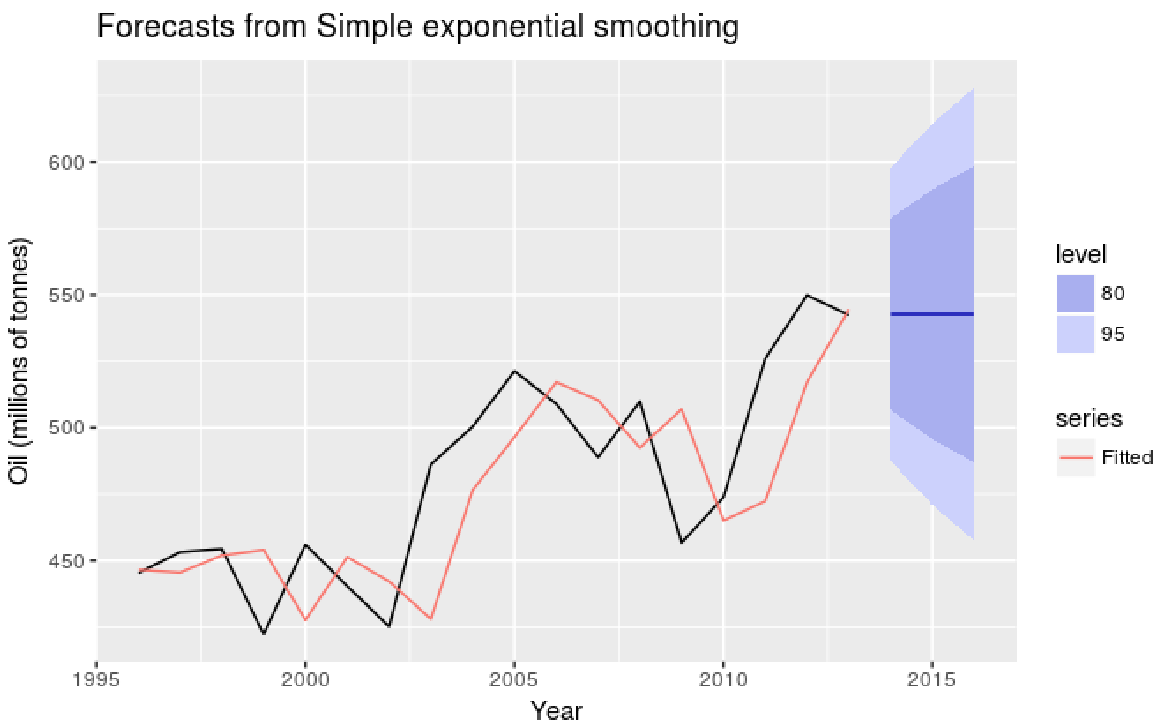
  - Forecast equation: $\hat{y}(t+1) = L(t)$

- Multi-horizon forecast: $\hat{y}(t+h\,|\,t) = L(t)$

  $\rightarrow$ "*Flat forecast*"

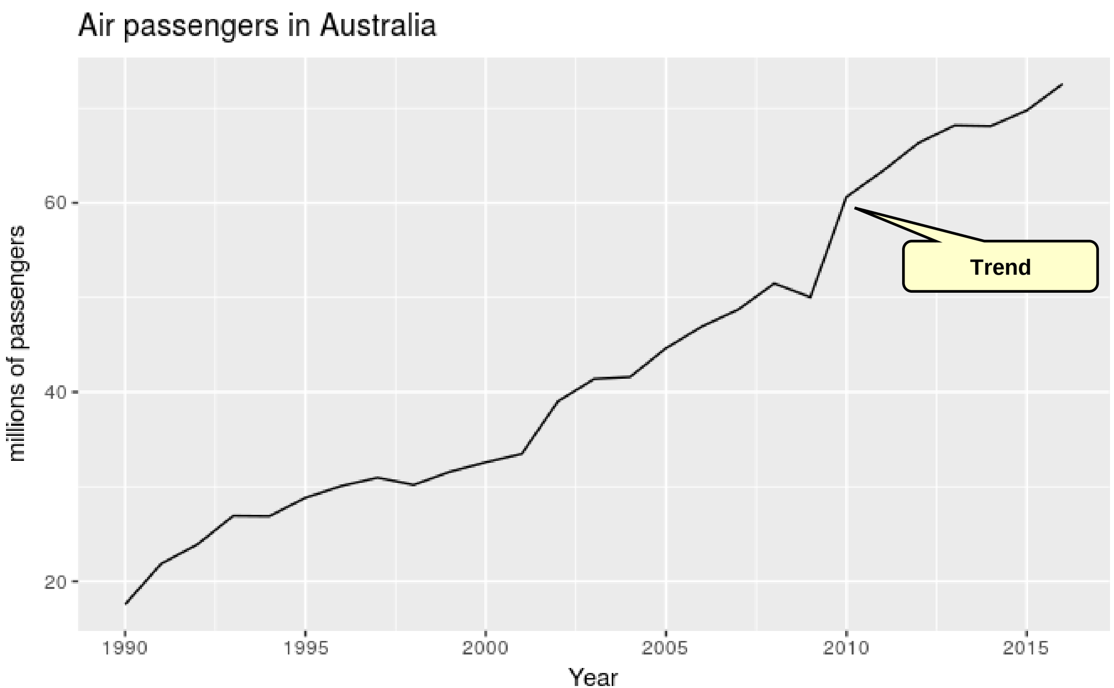
- Optimization: we need to select *L(0)* and α

Exponential smoothing mehtods
Simple Exponential Smoothing



Source Hyndman et al. (2017)

Exponential smoothing methods
Holt’s Linear Trend method



Source Hyndman et al. (2017)

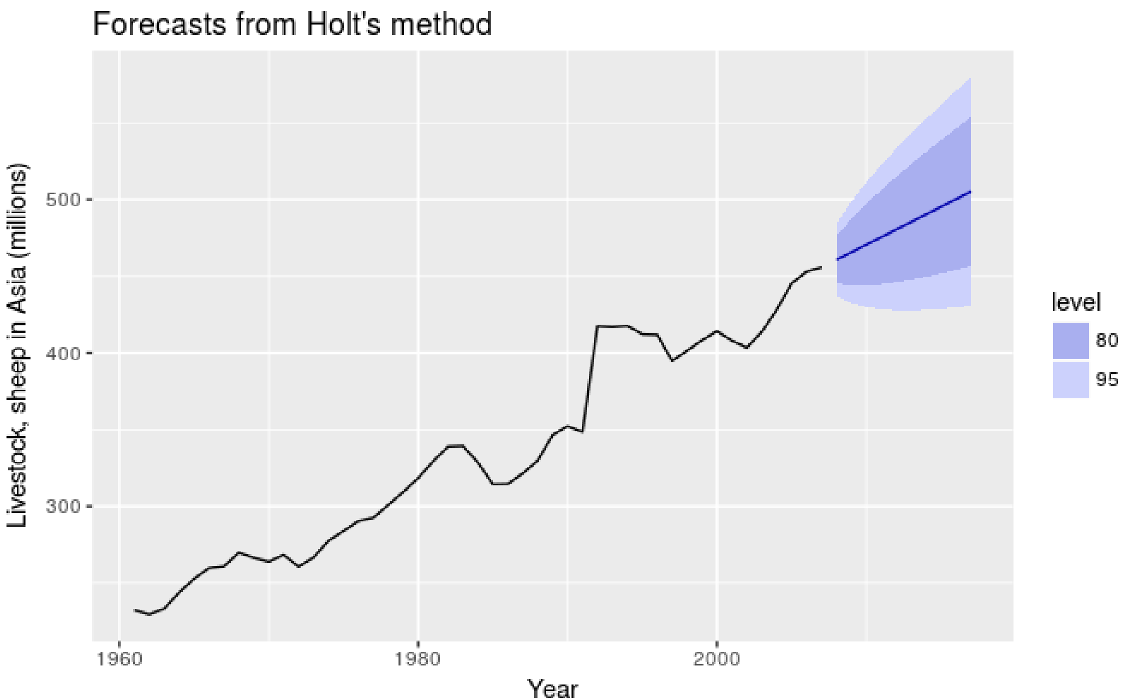
Exponential smoothing methods
Holt's Linear Trend method

- Given the constants α and β between 0 y 1, the Holt's linear trend model is given by:

  - Level: $$L(t) = \alpha y(t) + (1-\alpha)(L(t-1) + T(t-1))$$
  - Trend: $$T(t) = \beta(L(t) - L(t-1)) + (1-\beta)T(t-1)$$
  - Forecast: $$\hat{y}(t+m) = L(t) + T(t)\,m$$

- *L(t)* is a weighted average of observation *y(t)* and the one-step-ahead training forecast for time *t*, given by *(L(t−1)+T(t−1))*.
- *T(t)* is a weighted average of the estimated trend at time *t* based on *(L(t)-L(t-1))* and *T(t−1)*, the previous estimate of the trend.

Exponential smoothing methods
Holt's Linear Trend method

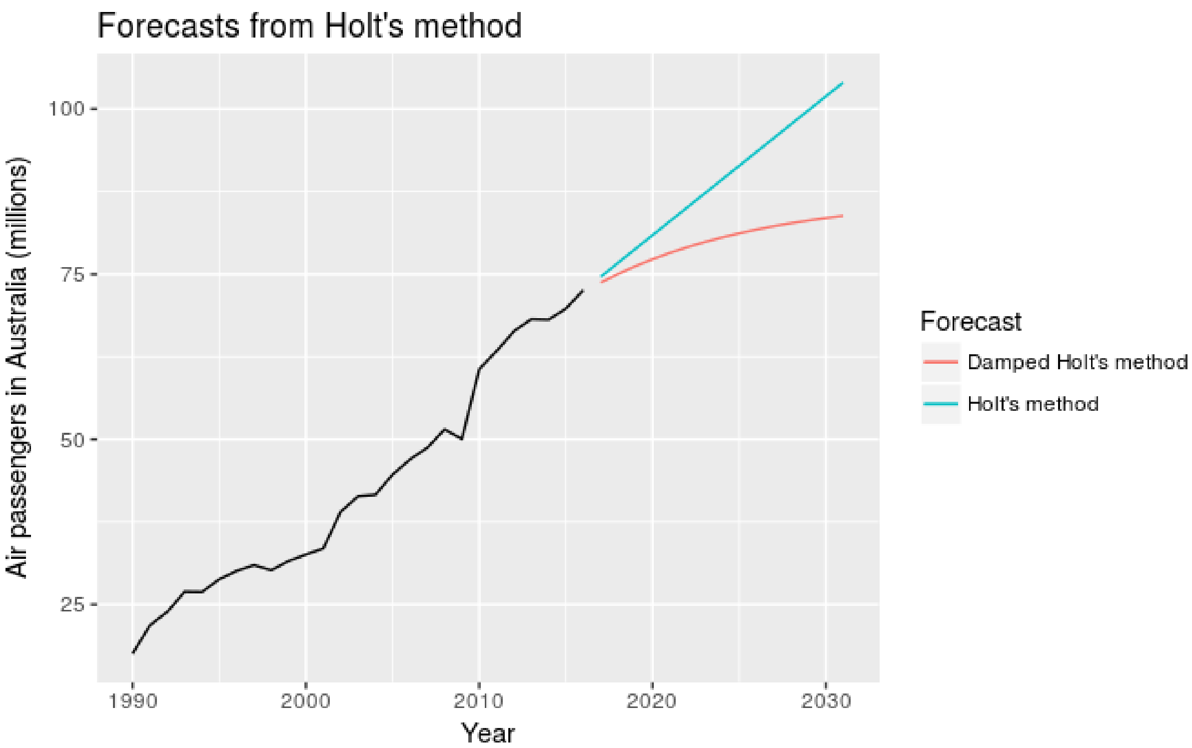


Source Hyndman et al. (2017)

Exponential smoothing methods
Damped Trend method

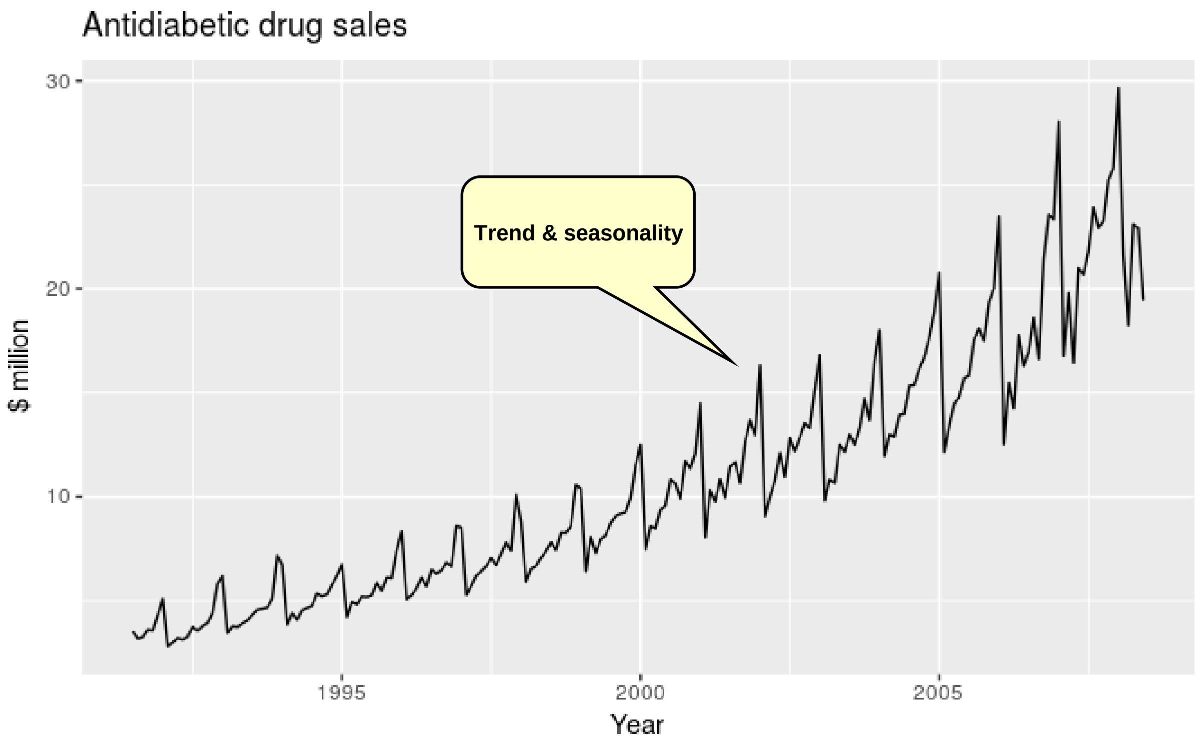
- Given the constants α, β and ϕ between 0 y 1, the damped linear trend model is given by:
  - Level: $L(t) = \alpha y(t) + (1-\alpha)(L(t-1) + \phi T(t-1))$
  - Trend: $T(t) = \beta(L(t) - L(t-1)) + (1-\beta)\phi T(t-1)$
  - Forecast: $\hat{y}(t+m) = L(t) + (\phi + \phi^2 + \ldots + \phi^m)T(t)$
- In practice, ϕ is rarely less than 0.8

Exponential smoothing methods
Damped Trend method



Source Hyndman et al. (2017)

Exponential smoothing methods
Holt-Winters Exponential Smoothing



Source Hyndman et al. (2017)

Exponential smoothing methods
Holt-Winters Exponential Smoothing

- Given the constants α, β, and γ between 0 y 1, the additive model is given by:
  - Level: $L(t) = \alpha(y(t)-S(t-s)) + (1-\alpha)(L(t-1)+T(T-1))$
  - Trend: $T(t) = \beta(L(t)-L(t-1)) + (1-\beta)T(t-1)$
  - Seasonality: $S(t) = \gamma(y(t)-L(t)) + (1-\gamma)S(t-s)$
  - Forecast: $\hat{y}(t+m) = L(t) + T(t)\, m + S(t-s+m)$

- The level equation shows a weighted average between the **seasonally adjusted observation** $(y(t)-S(t-s))$ and the non-seasonal forecast $(L(t-1)+T(t-1))$ for time t.
- The trend equation is identical to Holt's linear method.
- The seasonal equation shows a weighted average between the **current seasonal index**, $(y(t)-L(t))$, and the **seasonal index of the same season last year**.

Exponential smoothing methods

Holt-Winters Exponential Smoothing

- Given the constants $\alpha,$ $\beta,$ and γ between 0 y 1, the multiplicative seasonality model is given by:

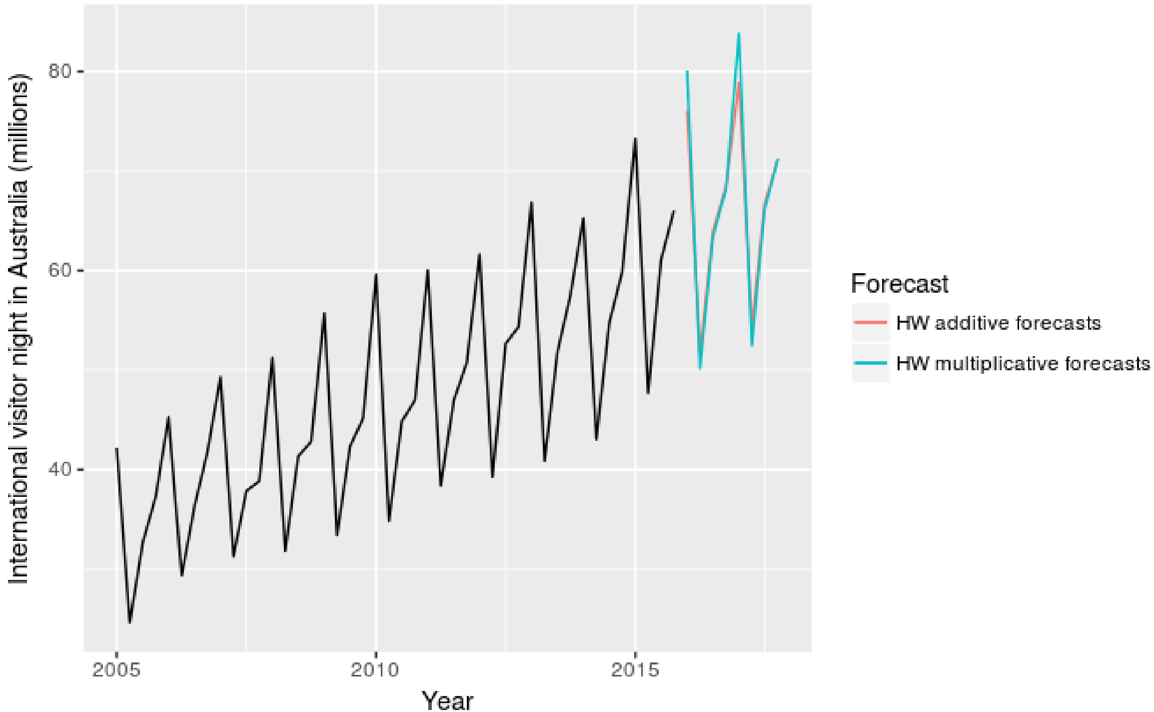
  - Level: $$L(t) = \alpha \frac{y(t)}{S(t-s)} + (1-\alpha)(L(t-1) + T(t-1))$$

  - Trend: $$T(t) = \beta(L(t) - L(t-1)) + (1-\beta)T(t-1)$$

  - Seasonality: $$S(t) = \gamma \frac{y(t)}{L(t)} + (1-\gamma)S(t-s)$$

  - Forecast: $$\hat{y}(t+m) = \left(L(t) + T(t)\, m\right) S(t-s+m)$$

Exponential smoothing methods
Holt-Winters Exponential Smoothing



Exponential smoothing methods
Holt-Winters' damped method

- Given the constants $\alpha, \ \beta, \gamma$ and ϕ between 0 y 1, the Holt-Winters method with a damped trend and multiplicative seasonality is given by:
  - Level: $L(t) = \alpha \dfrac{y(t)}{S(t-s)} + (1-\alpha)(L(t-1) + \phi T(t-1))$
  - Trend: $T(t) = \beta(L(t) - L(t-1)) + (1-\beta)\phi T(t-1)$
  - Seasonality: $S(t) = \gamma \dfrac{y(t)}{L(t)} + (1-\gamma)S(t-s)$
  - Forecast: $\hat{y}(t+m) = \left(L(t) + (\phi + \phi^2 + \ldots + \phi^m)T(t)\right)S(t-s+m)$

Trend	Seasonality		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	$S_t=\alpha X_t+(1-\alpha)S_{t-1}$ $\hat{X}_t(m)=S_t$	$S_t=\alpha(X_t-I_{t-p})+(1-\alpha)S_{t-1}$ $I_t=\delta(X_t-S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_t+I_{t-p+m}$	$S_t=\alpha(X_t/I_{t-p})+(1-\alpha)S_{t-1}$ $I_t=\delta(X_t/S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_tI_{t-p+m}$
	$S_t=S_{t-1}+\alpha e_t$ $\hat{X}_t(m)=S_t$	$S_t=S_{t-1}+\alpha e_t$ $I_t=I_{t-p}+\delta(1-\alpha)e_t$ $\hat{X}_t(m)=S_t+I_{t-p+m}$	$S_t=S_{t-1}+\alpha e_t/I_{t-p}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m)=S_tI_{t-p+m}$
A (Additive)	$S_t=\alpha X_t+(1-\alpha)(S_{t-1}+T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)T_{t-1}$ $\hat{X}_t(m)=S_t+mT_t$	$S_t=\alpha(X_t-I_{t-p})+(1-\alpha)(S_{t-1}+T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)T_{t-1}$ $I_t=\delta(X_t-S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_t+mT_t+I_{t-p+m}$	$S_t=\alpha(X_t/I_{t-p})+(1-\alpha)(S_{t-1}+T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)T_{t-1}$ $I_t=\delta(X_t/S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=(S_t+mT_t)I_{t-p+m}$
	$S_t=S_{t-1}+T_{t-1}+\alpha e_t$ $T_t=T_{t-1}+\alpha\gamma e_t$ $\hat{X}_t(m)=S_t+mT_t$	$S_t=S_{t-1}+T_{t-1}+\alpha e_t$ $T_t=T_{t-1}+\alpha\gamma e_t$ $I_t=I_{t-p}+\delta(1-\alpha)e_t$ $\hat{X}_t(m)=S_t+mT_t+I_{t-p+m}$	$S_t=S_{t-1}+T_{t-1}+\alpha e_t/I_{t-p}$ $T_t=T_{t-1}+\alpha\gamma e_t/I_{t-p}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m)=(S_t+mT_t)I_{t-p+m}$
DA (Damped additive)	$S_t=\alpha X_t+(1-\alpha)(S_{t-1}+\phi T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)\phi T_{t-1}$ $\hat{X}_t(m)=S_t+\sum_{i=1}^{m}\phi^i T_t$	$S_t=\alpha(X_t-I_{t-p})+(1-\alpha)(S_{t-1}+\phi T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)\phi T_{t-1}$ $I_t=\delta(X_t-S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_t+\sum_{i=1}^{m}\phi^i T_t+I_{t-p+m}$	$S_t=\alpha(X_t/I_{t-p})+(1-\alpha)(S_{t-1}+\phi T_{t-1})$ $T_t=\gamma(S_t-S_{t-1})+(1-\gamma)\phi T_{t-1}$ $I_t=\delta(X_t/S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=\left(S_t+\sum_{i=1}^{m}\phi^i T_t\right)I_{t-p+m}$
	$S_t=S_{t-1}+\phi T_{t-1}+\alpha e_t$ $T_t=\phi T_{t-1}+\alpha\gamma e_t$ $\hat{X}_t(m)=S_t+\sum_{i=1}^{m}\phi^i T_t$	$S_t=S_{t-1}+\phi T_{t-1}+\alpha e_t$ $T_t=\phi T_{t-1}+\alpha\gamma e_t$ $I_t=I_{t-p}+\delta(1-\alpha)e_t$ $\hat{X}_t(m)=S_t+\sum_{i=1}^{m}\phi^i T_t+I_{t-p+m}$	$S_t=S_{t-1}+\phi T_{t-1}+\alpha e_t/I_{t-p}$ $T_t=\phi T_{t-1}+\alpha\gamma e_t/I_{t-p}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m)=\left(S_t+\sum_{i=1}^{m}\phi^i T_t\right)I_{t-p+m}$

M (Multiplicative)	$S_t=\alpha X_t+(1-\alpha)(S_{t-1}R_{t-1})$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}$ $\hat{X}_t(m)=S_t\, R_t^m$	$S_t=\alpha(X_t-I_{t-p})+(1-\alpha)S_{t-1}R_{t-1}$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}$ $I_t=\delta(X_t-S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_tR_t^m+I_{t-p+m}$	$S_t=\alpha(X_t/I_{t-p})+(1-\alpha)S_{t-1}R_{t-1}$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}$ $I_t=\delta(X_t/S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=(S_tR_t^m)I_{t-p+m}$
	$S_t=S_{t-1}R_{t-1}+\alpha e_t$ $R_t=R_{t-1}+\alpha\gamma e_t/S_{t-1}$ $\hat{X}_t(m)=S_tR_t^m$	$S_t=S_{t-1}R_{t-1}+\alpha e_t$ $R_t=R_{t-1}+\alpha\gamma e_t/S_{t-1}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t$ $\hat{X}_t(m)=S_tR_t^m+I_{t-p+m}$	$S_t=S_{t-1}R_{t-1}+\alpha e_t/I_{t-p}$ $R_t=R_{t-1}+(\alpha\gamma e_t/S_{t-1})/I_{t-p}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m)=(S_tR_t^m)I_{t-p+m}$
DM (Damped multiplicative)	$S_t=\alpha X_t+(1-\alpha)(S_{t-1}R_{t-1}^{\phi})$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}^{\phi}$ $\hat{X}_t(m)=S_tR_t^{\sum_{i=1}^{m}\phi^i}$	$S_t=\alpha(X_t-I_{t-p})+(1-\alpha)S_{t-1}R_{t-1}^{\phi}$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}^{\phi}$ $I_t=\delta(X_t-S_t)+(1-\delta)I_{t-p}$ $\hat{X}_t(m)=S_tR_t^{\sum_{i=1}^{m}\phi^i}+I_{t-p+m}$	$S_t=\alpha(X_t/I_{t-p})+(1-\alpha)(S_{t-1}R_{t-1}^{\phi})$ $R_t=\gamma(S_t/S_{t-1})+(1-\gamma)R_{t-1}^{\phi}$ $I_t=\delta(X_t/S_t)+(1-\delta)I_{t-1}$ $\hat{X}_t(m)=\left(S_tR_t^{\sum_{i=1}^{m}\phi^i}\right)I_{t-p+m}$
	$S_t=S_{t-1}R_{t-1}^{\phi}+\alpha e_t$ $R_t=R_{t-1}^{\phi}+\alpha\gamma e_t/S_{t-1}$ $\hat{X}_t(m)=S_tR_t^{\sum_{i=1}^{m}\phi^i}$	$S_t=S_{t-1}R_{t-1}^{\phi}+\alpha e_t$ $R_t=R_{t-1}^{\phi}+\alpha\gamma e_t/S_{t-1}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t$ $\hat{X}_t(m)=S_tR_t^{\sum_{i=1}^{m}\phi^i}+I_{t-p+m}$	$S_t=S_{t-1}R_{t-1}^{\phi}+\alpha e_t/I_{t-p}$ $R_t=R_{t-1}^{\phi}+(\alpha\gamma e_t/S_{t-1})/I_{t-p}$ $I_t=I_{t-p}+\delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m)=\left(S_tR_t^{\sum_{i=1}^{m}\phi^i}\right)I_{t-p+m}$

E.S. Gardner Jr. / International Journal of Forecasting 22 (2006) 637–666

Exponential smoothing methods
R implementation

- Three main models are implemented:
  - **Simple ES**

```
ES.forecast <- ses(fdata_ts_train, h=12)
```

  - **Holt's method**. It can be dampled if desired.

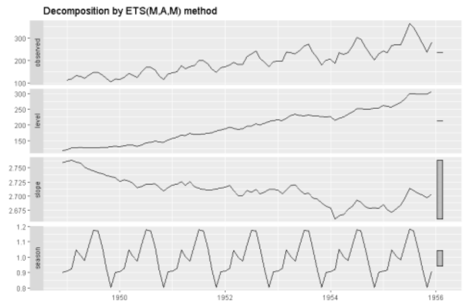
```
ES.forecast <- holt(fdata_ts_train, damped = TRUE, h=12)
```

  - **Holt-Winters' seasonal ES**. It can be damped if desired.

```
ES.forecast <- hw(fdata_ts_train, damped = FALSE, h=12, seasonal="multiplicative")
```

- These functions estimate the corresponding parameters and produce a forecast for the specified horizon.
- Components and forecast can be plotted with `autoplot`.

```
#fitted coefficients and model performance
summary(ses.forecast)
#plot forecasts and confidence intervals
autoplot(ses.forecast)
#plot components
autoplot(ses.forecast$model)
```



5

Bibliography

comillas.edu

Bibliography

- "Forecasting Principles and Practice". Rob J Hyndman and George Athanasopoulos. 3rd Edition. Otexts. 2017 (http://otexts.org/fpp3/)
- **"**Time series analysis. Univariate and Multivariate Methods**".** William W.S. Wei. 2nd edition. Pearson Addison Wesley. 2006.
- "Análisis de series temporales". Daniel Peña. Alianza Editorial. 2005
- "Time Series Analysis: Forecasting & Control". 3ª Ed. G. Box, G.M. Jenkins, G. Reinsel. Prentice Hall. 1994
- "Principles of business forecasting". Keith Ord, Robert Fildes & Nikolaos Kourentzes. 2nd Edition. Wessex Press Inc. 2017
- "Time Series Analysis". James D. Hamilton. Princeton University Press. 1994
- "Short term Forecasting in Energy. A Guided Tour". A. Muñoz, E.F. Sánchez-Úbeda, A. Cruz, J. Marín. *Handbook of Power Systems II*. Editors Pardalos, P.M.; Rebennack, S.; Pereira, M.V.F y Iliadis, N.A.. Ed. Springer. Berlin, Germany, 2010
- "Analysis of the daily outdoor air temperatures for energy forecasting in the context of climate change". S. Moreno-Carbonell, E.F. Sánchez-Úbeda, A. Muñoz. Energies 2020, 13(7). (https://doi.org/10.3390/en13071569)

Alberto Aguilera 23, E-28015 Madrid - Tel: +34 91 542 2800 - http://www.iit.comillas.edu