COMP8811 Week 3 Day1

Time series

Neda Sakhaee

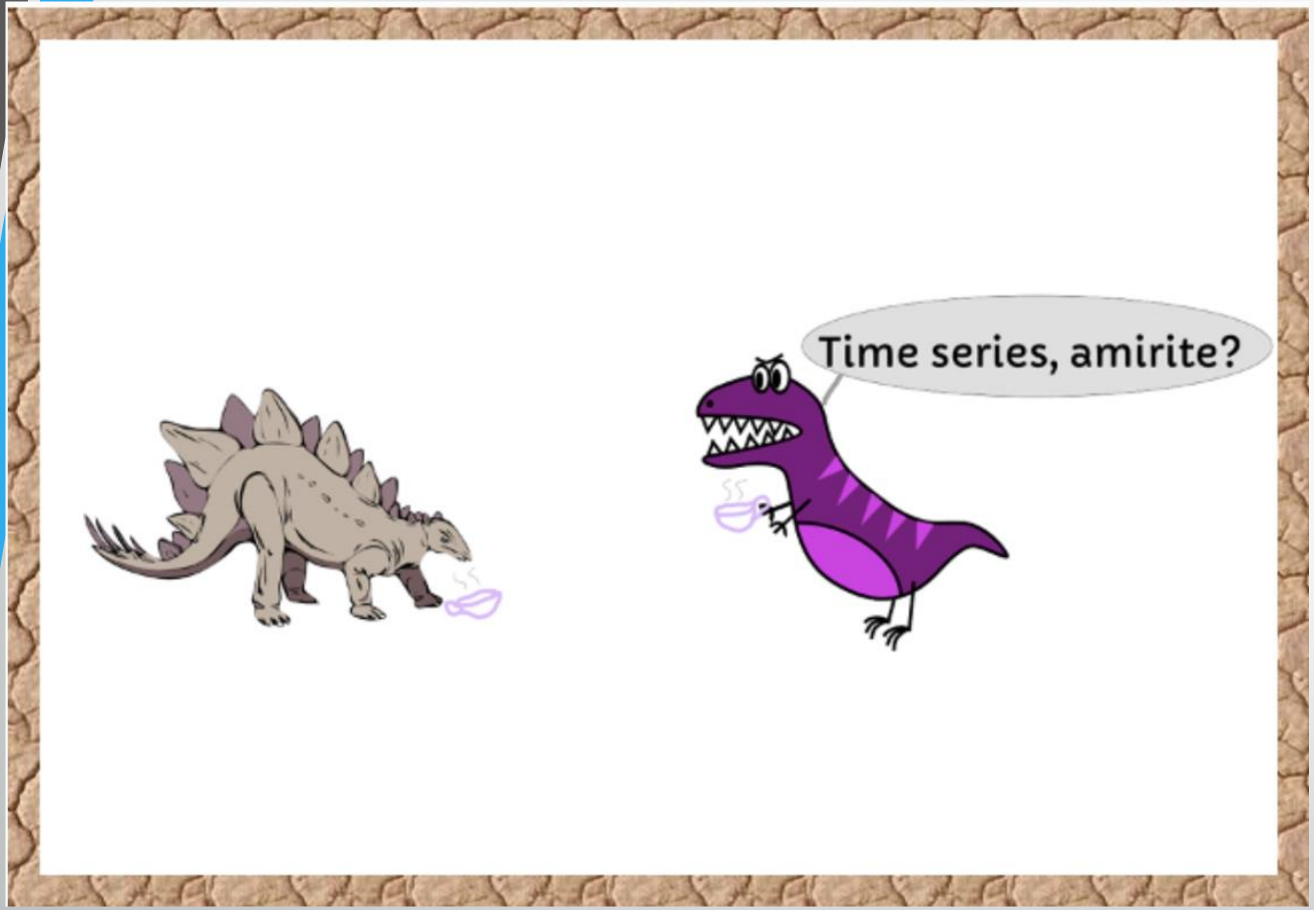
May 2024

Course Objectives

In this course, we are going to focus on the following learning objectives:

- 1. Basic time series concepts
- 2. Understand the decomposition model.
- 3. Build and fit various time series models to real world data sets.
- 4. Diagnose the fit of time series models through plots.
- 5. Forecast the future using fitted time series models.

By the end of this session, you will have a comprehensive introduction to basic time series concepts.



Time Series Data Overview

Time Series Data Overview

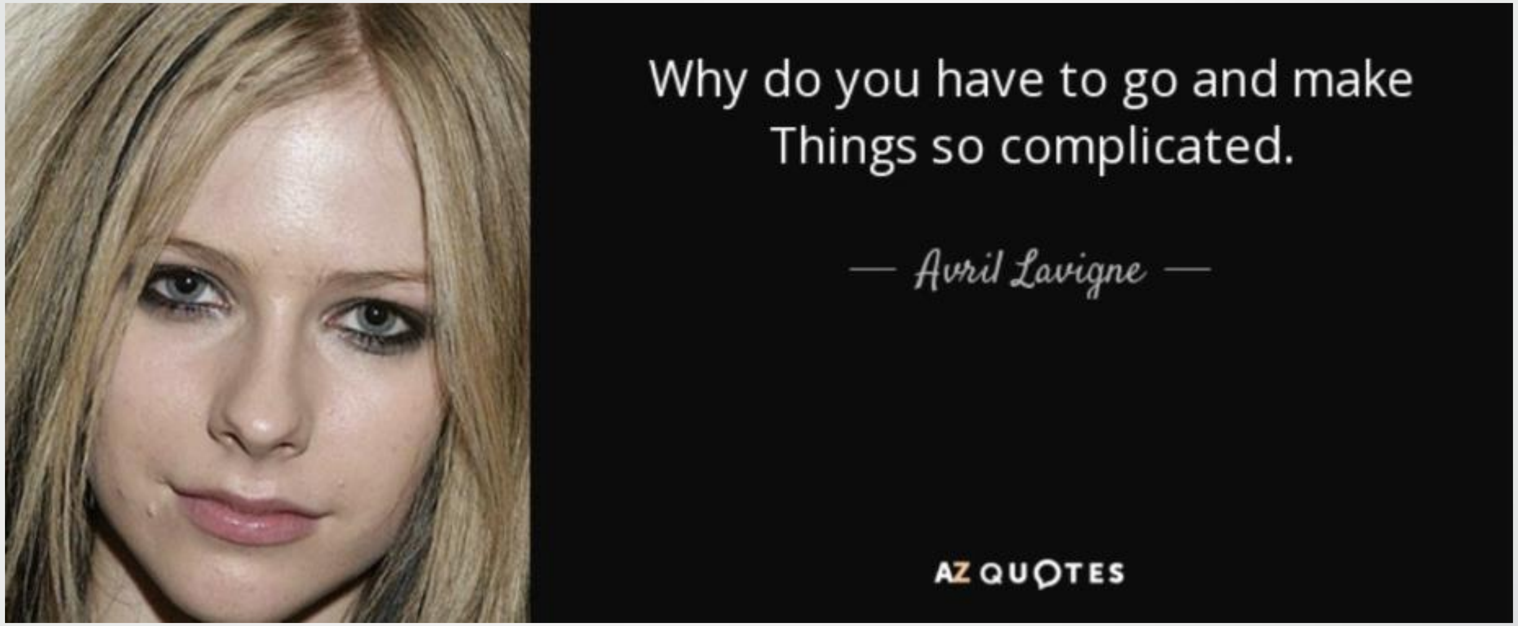
(Univariate) time series data is defined as sequence data over time:

$$Y_1, Y_2, \dots, Y_T$$

where t is the time period and Y_t is the value of the time series at a particular point.

Different time periods can be used for different applications, e.g. seconds, minutes, hours, days, weeks, months or years.

Examples: daily temperatures in Auckland, NZ election turnout by year, minute stock prices.



Why Time Series?

Why Time Series?

1. many forecasting tasks actually involve small samples which makes machine learning less effective
2. time series models are more interpretable and less black box than machine learning algorithms
3. time series appropriately accounts for forecasting uncertainty.

Prepare your RStudio

```
#load required r packages
library(IRdisplay)
library(magrittr)
library(tidyverse)
library(scales)
library(gridExtra)
library(forecast)
library(tseries)
library(ggthemes)
theme_set(theme_economist())
```

```
#load helper R functions
setwd("your working directory")
source("R Functions/compare_models_function.R")
source("R Functions/sim_random_walk_function.R")
print("Loading is completed")
```

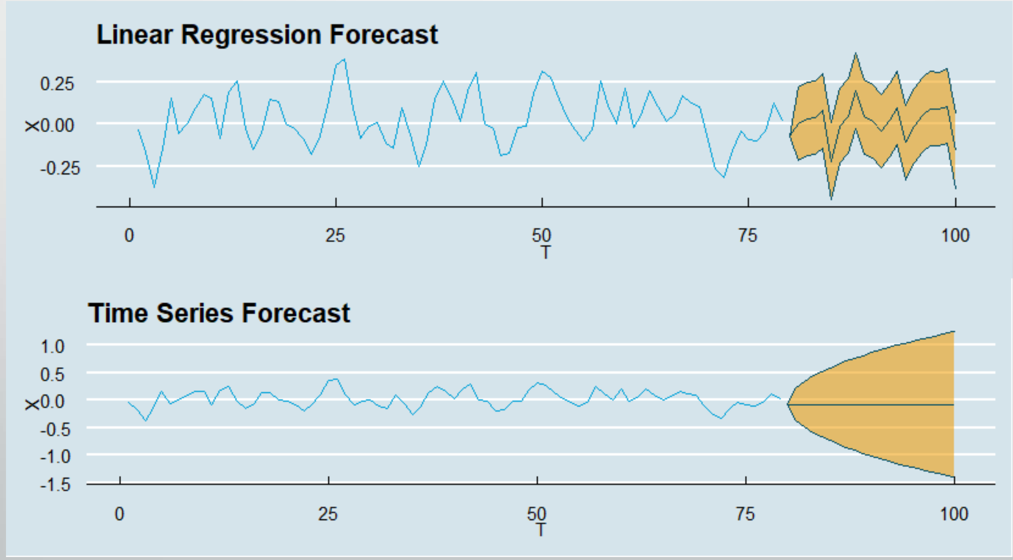
Why Time Series?

As an example, let's look at the following data generating process known as a random walk: $X_t = X_{t-1} + \epsilon_t$

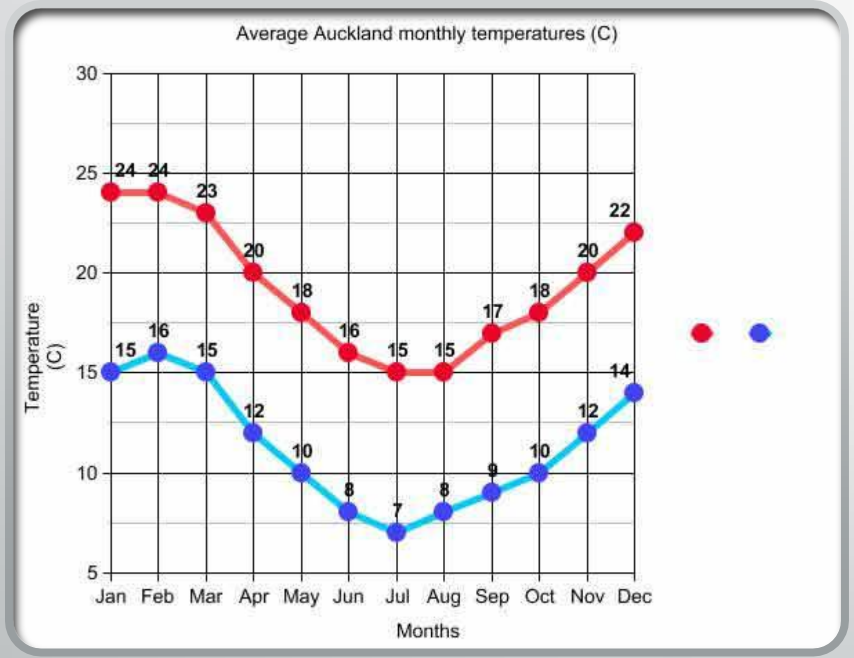
We can compare the forecasting performance of linear regression to that of a basic time series model known as an AR(1) model.

Open the file ***compare_models_function.R***, Run it. Then type:

compare.models(n=100)



Basics and Definitions



- A **time series** is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals.

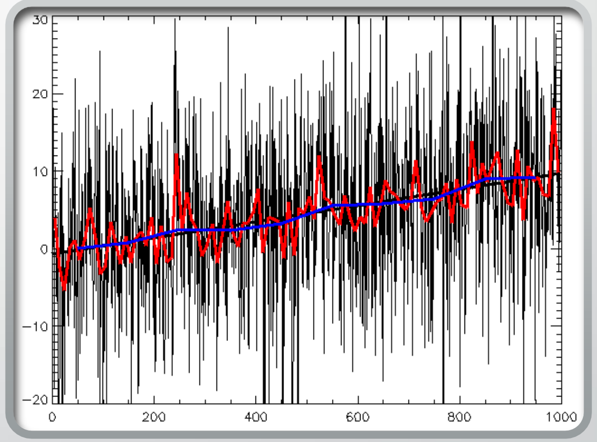
Basics and Definitions

- A time series is a sequence $\{y_t\}$ of values related to a quantity of interest that can be measured at specific time periods.
- Different time periods can be used for different applications, e.g. seconds, minutes, hours, days, weeks, months or years.

Basics and Definitions

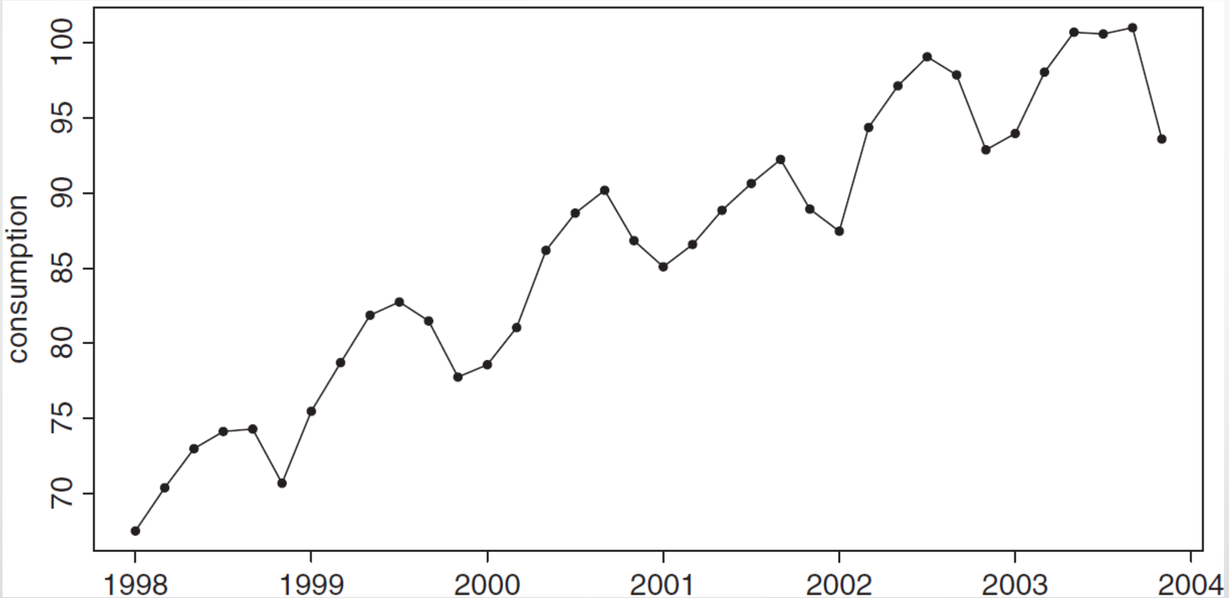
The aim of models, described in this lecture is to:

- Identifying regular patterns of an available time series
- Predicting the time series values in the future time periods



Basics and Definitions

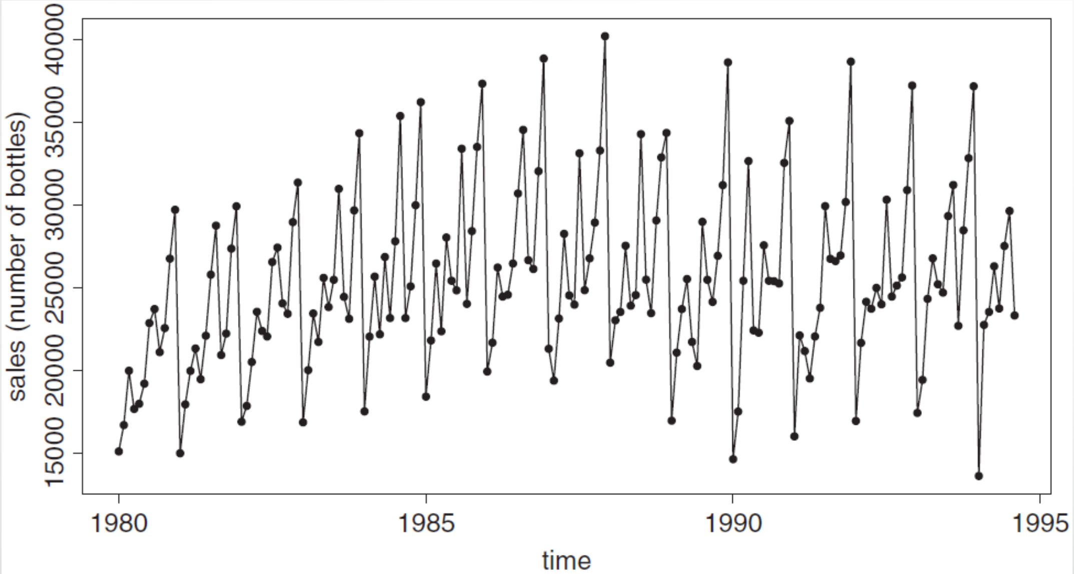
Example - Bimonthly electric assumption (M Watt Hours)in Italy



- Identify and state any particular patterns in this time series.
- Predict the consumption for the next two month.

Basics and Definitions

Monthly sales of wine in Australia (numbers of bottles)



- Identify and state any particular patterns in this time series.
- Predict the consumption for the next two month.

Basics and Definitions

- **Stochastic Nature of Time Series**
- Ideally, a model of the time series, *{Y_t}*, would be in the form of a stochastic process. In this situation main statistical parameters of the time series can be calculated as
    - Mean value
    - Variance (Second-order moment)

Basics and Definitions

- Assume that at time index N (current time) an actual time series is presented by a sequence of real numbers, measured at time indices 1 to k:
  - $y_1, y_2, y_3, \ldots, y_k$
- y_{k+1} and next values are not available because we cannot measure them before the actual process produces them.

Basics and Definitions

- A model for the time series can be developed (How? we discuss it later). This model could be perfect (exact) or imperfect. In general case, the model can re-calculate the time series sequence as
- **$f_1, f_2, f_3, \ldots, f_k, , f_{k+1}, \ldots,$**
- It can calculate future values as well because it's a mathematical model and causality constraints do not apply to it.

Basics and Definitions

The actual amount of money that you have spent daily on transporation from home to Uni (bus, fuel or etc) during the last 7 days (week) can be represented by a time series.

- Specify the actual time series: y_1, y_2,…, y_7 .
- State a set of rules (facts) that explain your transportation cost. This set forms a model for the actual time-series.
- Using the model, re-calculate the time series: f_1, f_2, …, f_7 .

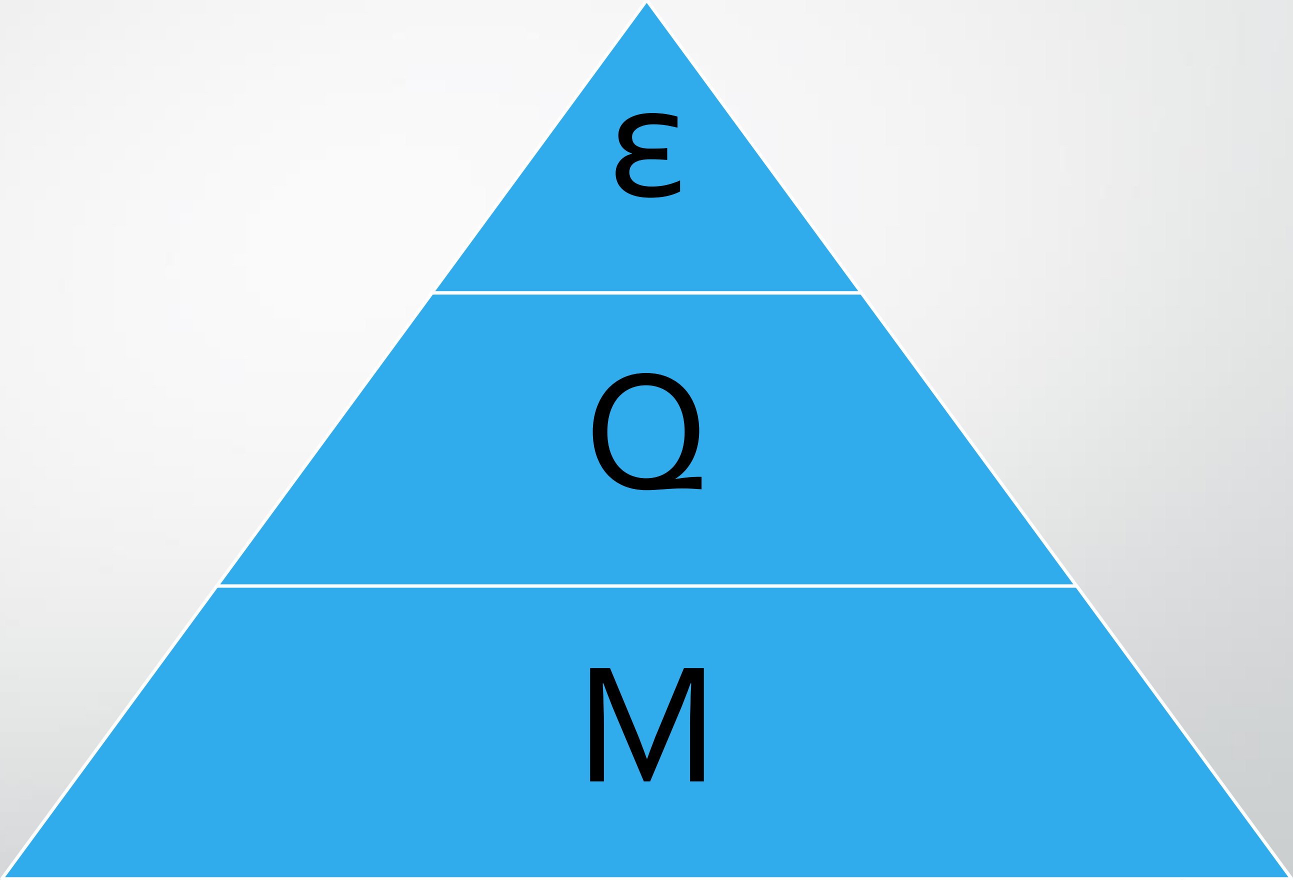
Decomposition of Time Series

Decomposition

Decomposition of Time Series

Time Series Components:

1. Trend
2. Seasonality
3. Random noise



$Y_t = g(M_t, Q_t, \varepsilon_t)$

Decomposition of Time Series

- A time series can be expressed as a combination of its components:

$$Y_t = g(M_t, Q_t, \varepsilon_t)$$

- where *g* represents an appropriate function to be selected.
- A multiplicative model is usually assumed, such as:

$$Y_t = M_t \times Q_t \times \varepsilon_t$$

- Alternatively, an additive model is assumed such as:

$$Y_t = M_t + Q_t + \varepsilon_t$$

Decomposition of Time Series

Trend:

- The average behaviour of a time series over time.
- It can be increasing, decreasing or stationary.
- Shown by M_t.

Decomposition of Time Series

Seasonality:

- The result of wavelike short-term fluctuations of regular frequency that appear in the values of a time series.
- Determined by the natural cycles by which demand develops, or by the seasonality of the products to which the time series refers.
- Shown by Q_t

Decomposition of Time Series

Random Noise (Residuals):

- Fluctuation component of a time series used to represent all irregular variations in the data that cannot be explained by the other components.
- Shown by $\{\varepsilon_t\}$
- In general, it is required that the time series $\{\varepsilon_t\}$, obtained from $\{y_t\}$ after the trend and the seasonality components have been identified and removed.
- Randomness or Noise or Residual is the random fluctuation or unpredictable change. This is something which we cannot guess which reflects any irregular fluctuation across every period.

Break

10 min

Decomposition of Time Series

Moving Average (an estimation of Trend):

$m_t(h) \approx M_t$

1. Shown by $m_t(h)$
2. The arithmetic mean of h consecutive observations of the time series $\{y_t\}$, such that the time index t belongs to the indices of the h averaged observations.
3. It is possible to compute different values of the moving average, depending on the position taken by the index t in the sequence of h contiguous observations used to calculate the average.

Decomposition of Time Series

Centred Moving Average:

1. The arithmetic mean of *h* observations such that *t* is the middle point of the set of periods corresponding to the observations.

2. Assuming that *h* is odd:

$$m_t(h) = \frac{y_{t+(h-1)/2} + y_{t+(h-1)/2-1} + \cdots + y_{t-(h-1)/2}}{h}$$

3. Assuming that h is even:

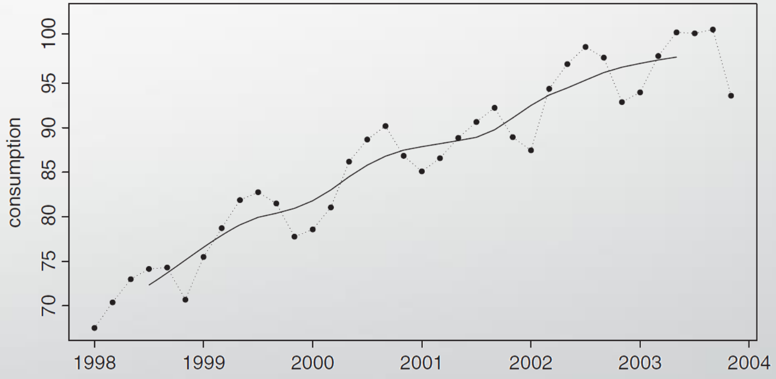
$$m_t(h) = \frac{y_{t+h/2} + y_{t+h/2-1} + \cdots + y_{t-h/2+1}}{2h} + \frac{y_{t+h/2-1} + y_{t+h/2-2} + \cdots + y_{t-h/2}}{2h}$$

Time Series : *Trend*

Exercise 1

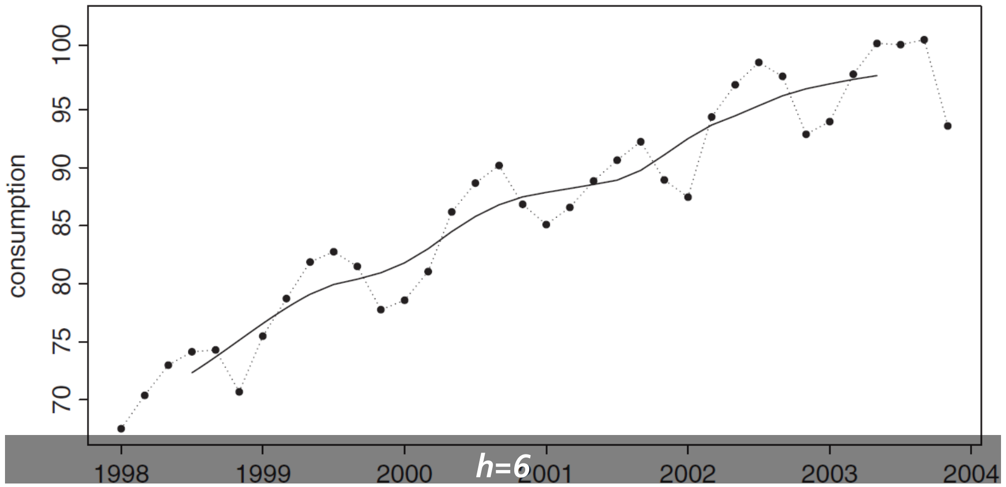
- **Open** the EXCEL sheet on the Moodle
- **Calculate** M_t for all periods (from 2000-2005) using h=6
- **Draw a graph** for M_t

$$m_t(h) = \left(y_{t-\left(\frac{h}{2}\right)} + 2\left(\sum_{i=-\left(\frac{h}{2}-1\right)}^{\left(\frac{h}{2}-1\right)} y_{t+i} \right) + y_{t+\left(\frac{h}{2}\right)} \right) / 2h$$

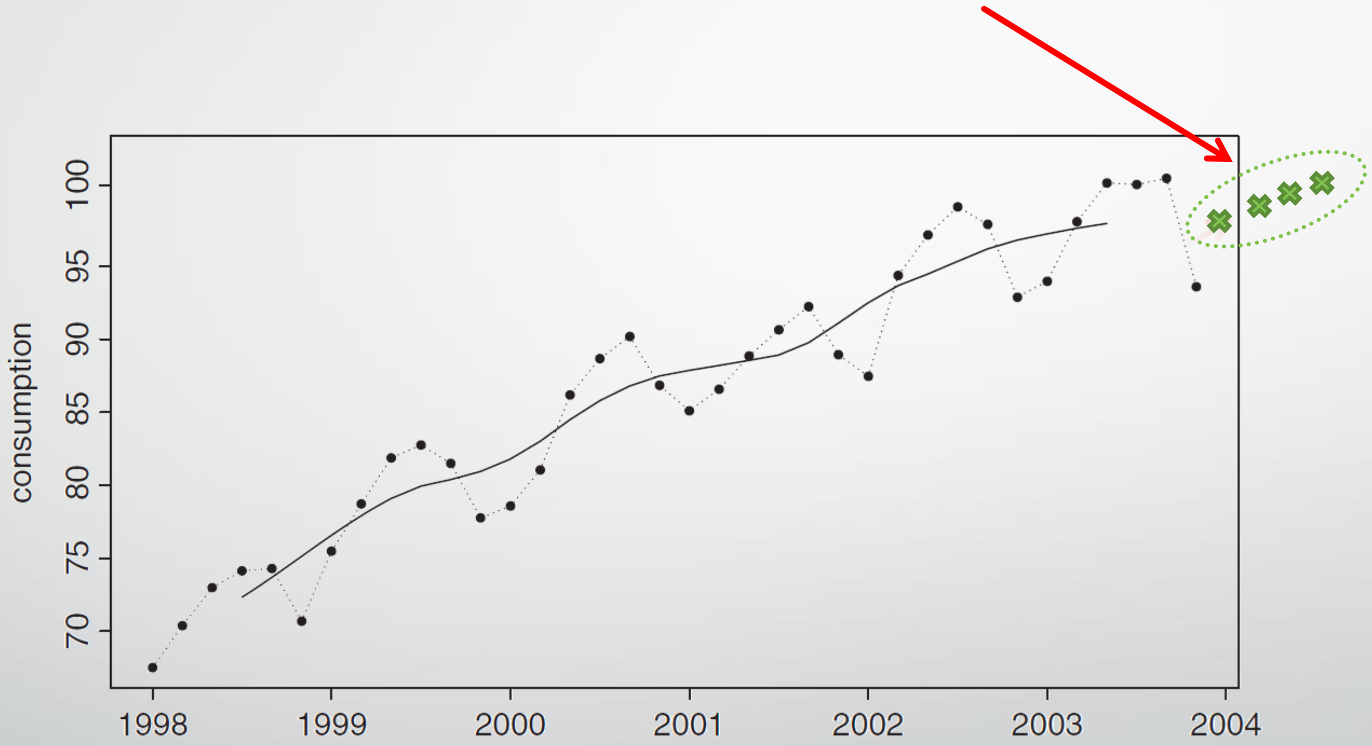


Decomposition of Time Series

- M_t for the power consumption example:



Time Series : *Trend*



- Can you obtain future values of the trend based on the **Moving Average** m_t ?

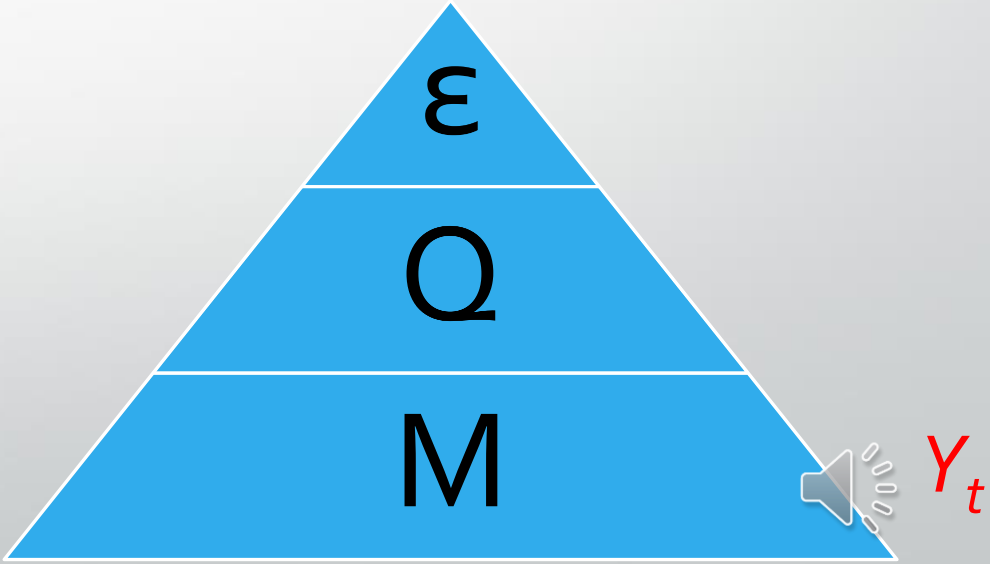
Decomposition of Time Series

Removal of the Trend Component

Now that we have identified $M_t \approx m_t(h)$, we can remove it from the time series.

What remained is another time series.

We show the new time series by B.



Decomposition of a time series

Removal of the Trend Component

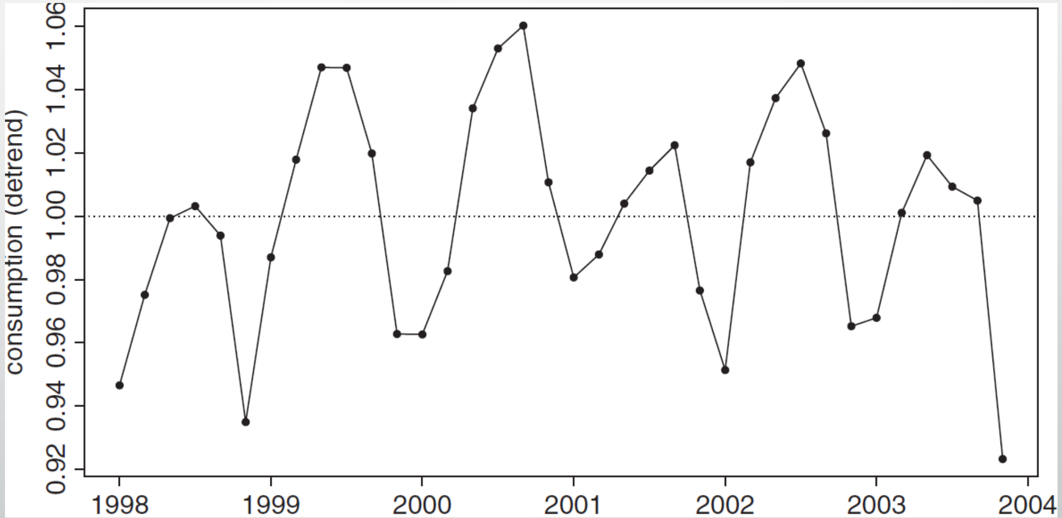
For a multiplicative model:

$$B_t = Q_t \varepsilon_t = \frac{Y_t}{M_t} \approx \frac{Y_t}{m_t(L)}$$

Time Series : *Seasonality*

Exercise 2

- **Use** the EXCEL sheet (with Exercise 1 completed)
- **Calculate** B_t for all periods (from 2000-2005)
- **Draw a graph** for B_t



Time Series : *Seasonality*

Seasonality component (Q_t)

Q_t = *the **mean value** of all* B_t

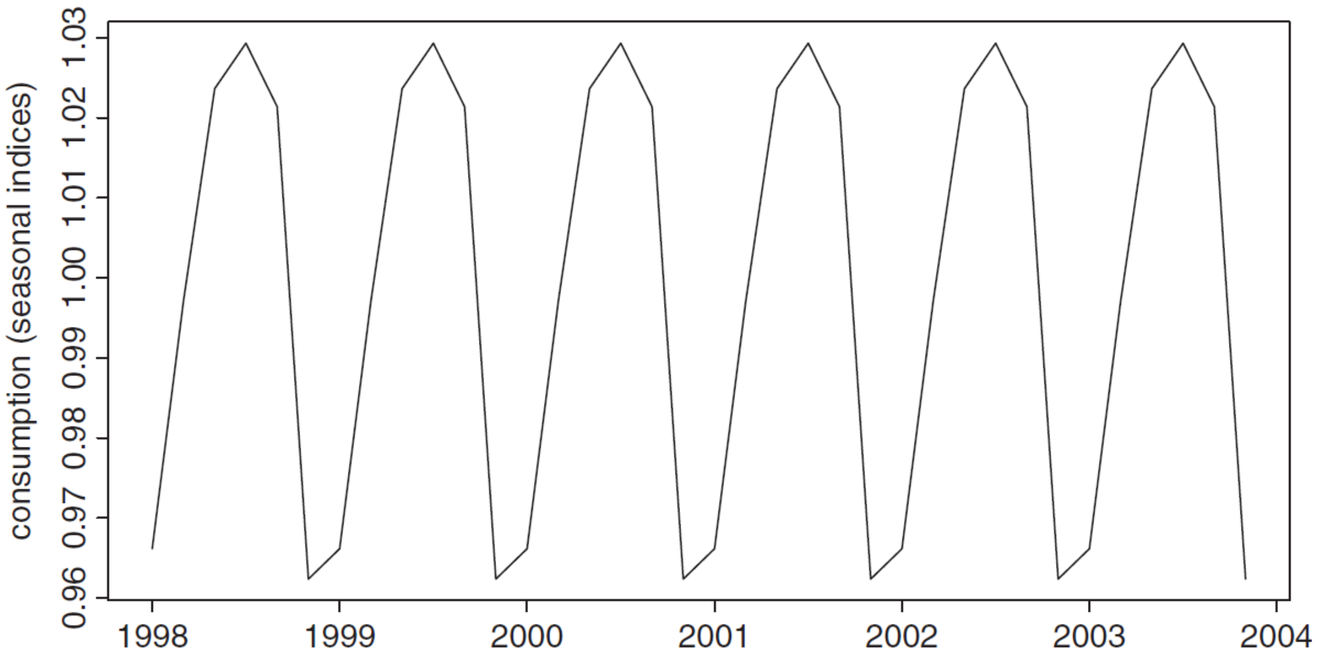
- We can write

$Q_t = \sum_{i=1}^{n} B_i \,/ n$

n is the number of observations at the same time t at different periods

Decomposition of a time series

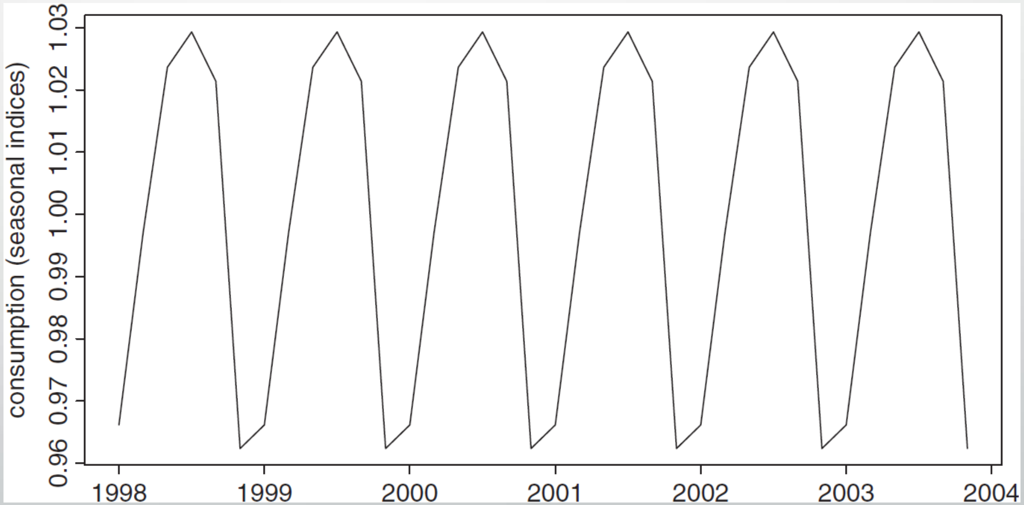
Seasonality for the power consumption example:



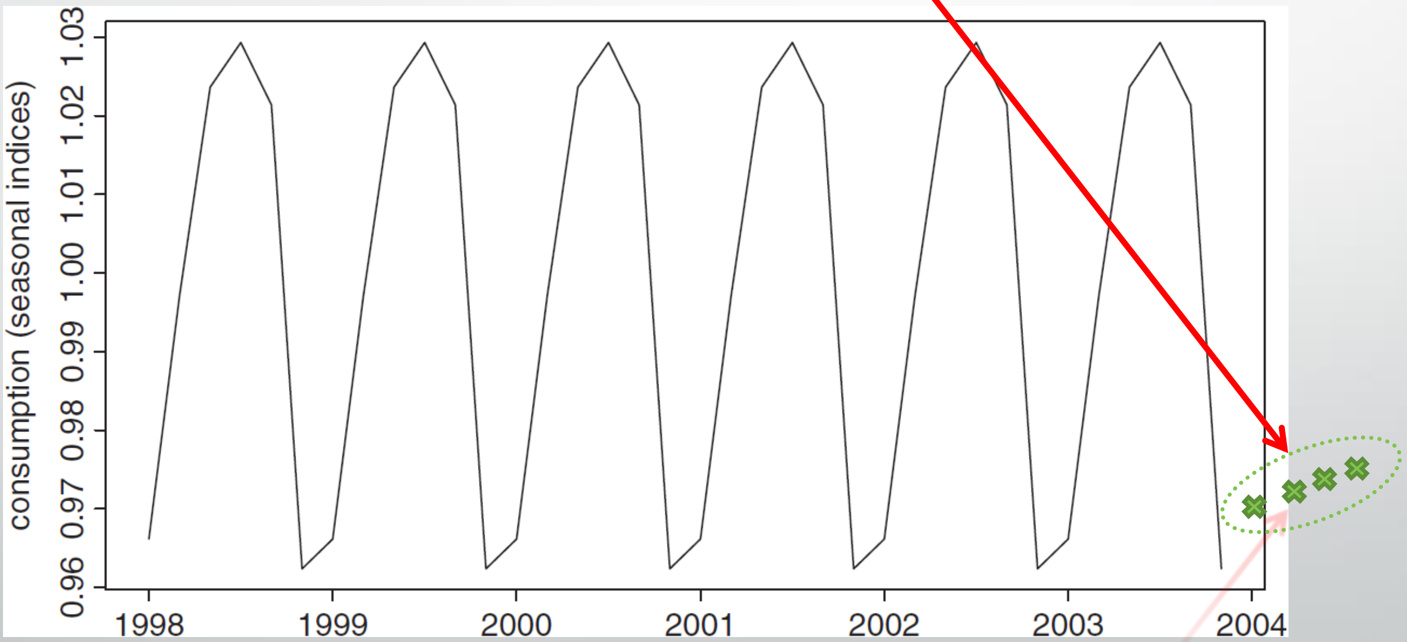
Time Series : *Seasonality Component*

Exercise 3

- **Use** the EXCEL sheet (with Exercise 1+2 completed)
- **Calculate** Q_t for **every month** during 2000-2005
- **Draw a graph** for Q_t



Time Series : *Seasonality*



- Can you obtain **future values** of the seasonality indices based on the **seasonality component**?

Time Series : *De-seasonalized*

$Y_t = M_t \times Q_t \times \varepsilon_t$

Removal of the Season Component Q_t

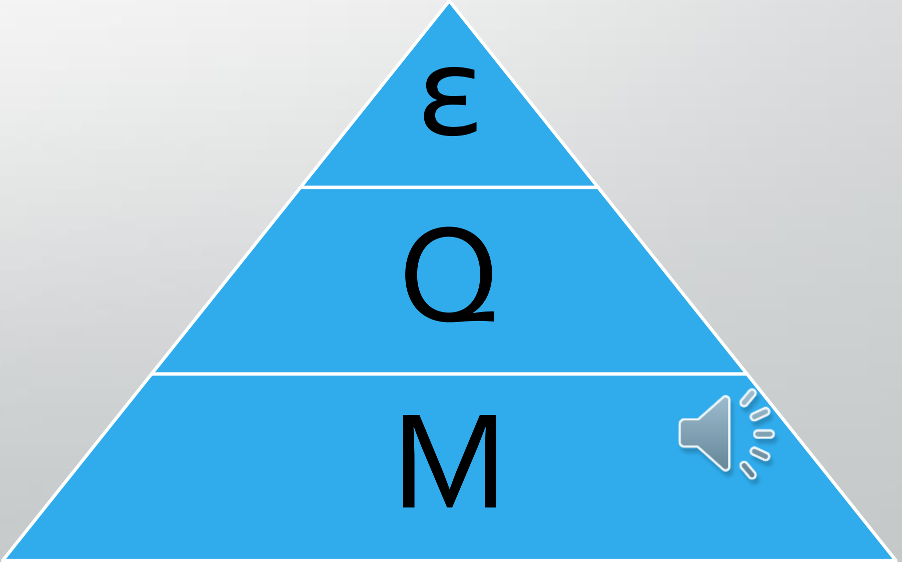
- We can remove Q_t from the time series
- What remained is another time series
- We show the new time series by **C**

Time Series : *De-seasonalized*

Removal of Seasonality component

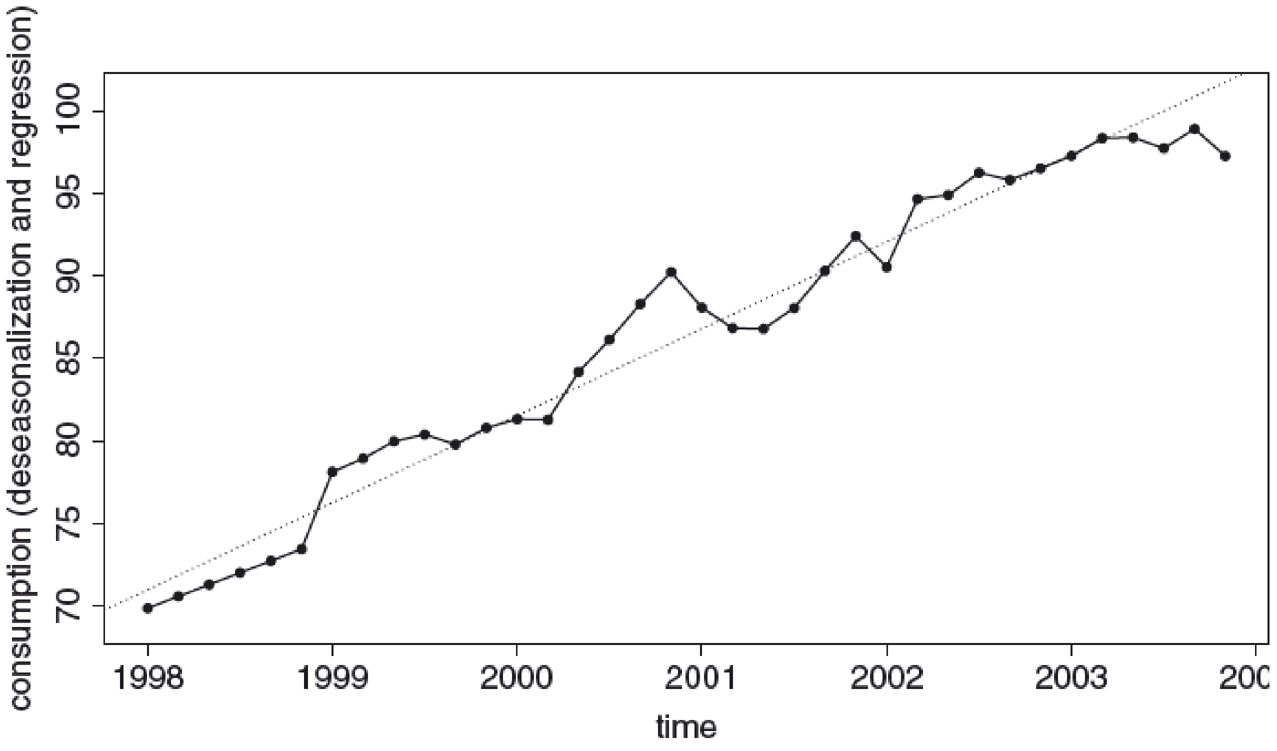
$C_t = M_t \times \varepsilon_t$
$= (Y_t/(Q_t \times \varepsilon_t)) \times \varepsilon_t$
$= Y_t/Q_t$

$Y_t = M_t \times Q_t \times \varepsilon_t$



Called de-seasonalization

So, we have $C_t = Y_t / Q_t$
Also … $C_t = M_t \times \varepsilon_t$



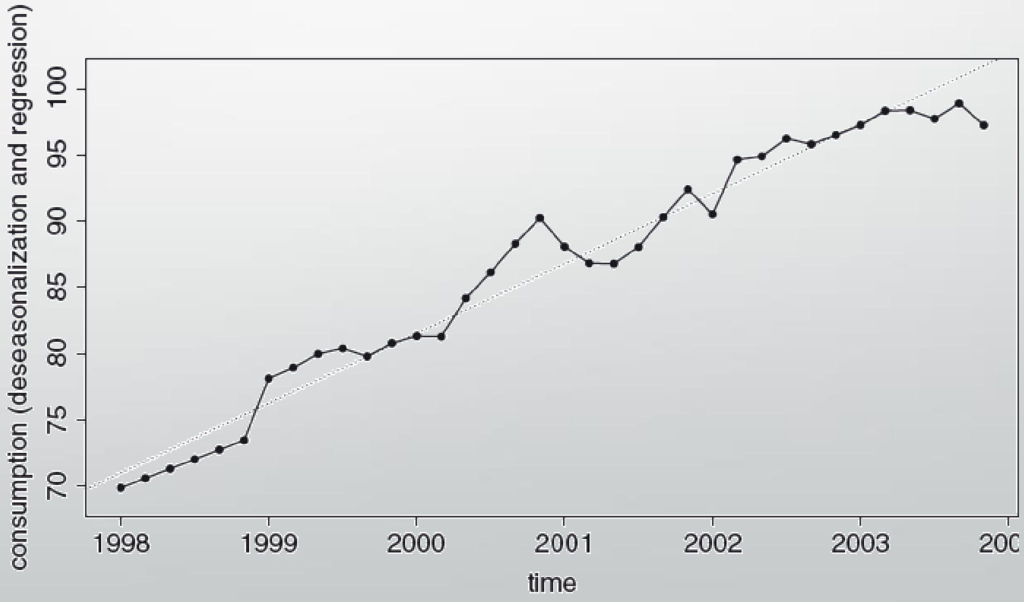
Decomposition of a time series

Deseasonalization for the power consumption example

Time Series : *De-seasonalized*

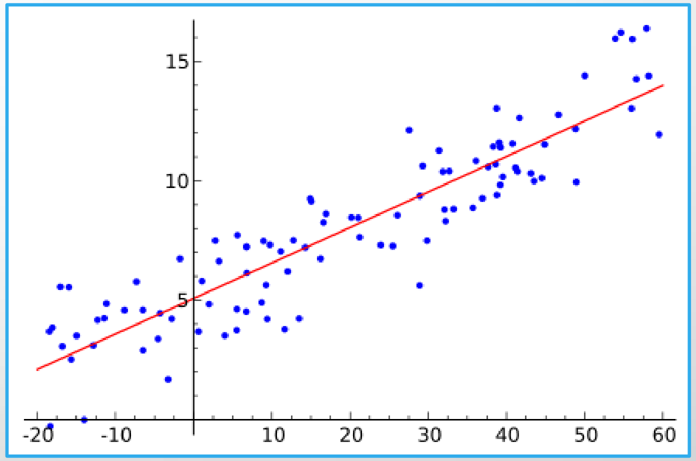
- **Use** the EXCEL sheet (with Exercise 1+2+3 completed)
- **Calculate** C_t for 2000-2005
- **Draw a graph** for C_t

Exercise 4



Time Series : *De-seasonalized*

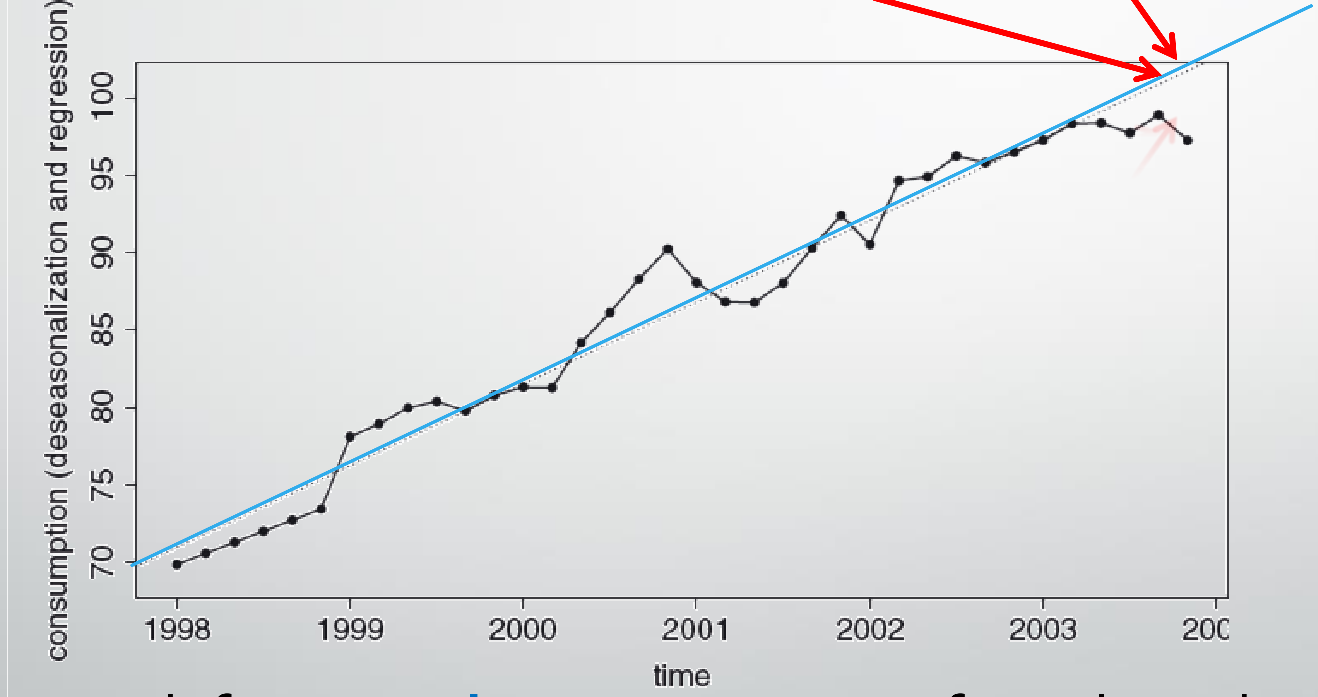
- At this stage we re-calculate the trend on the **De-seasonalized** C_t



- We look for **an analytical model** for the trend that *can be used for forecasting.*
  - **Linear regression** ?!?

Time Series : *De-seasonalized*

$$M_t = a + bt$$

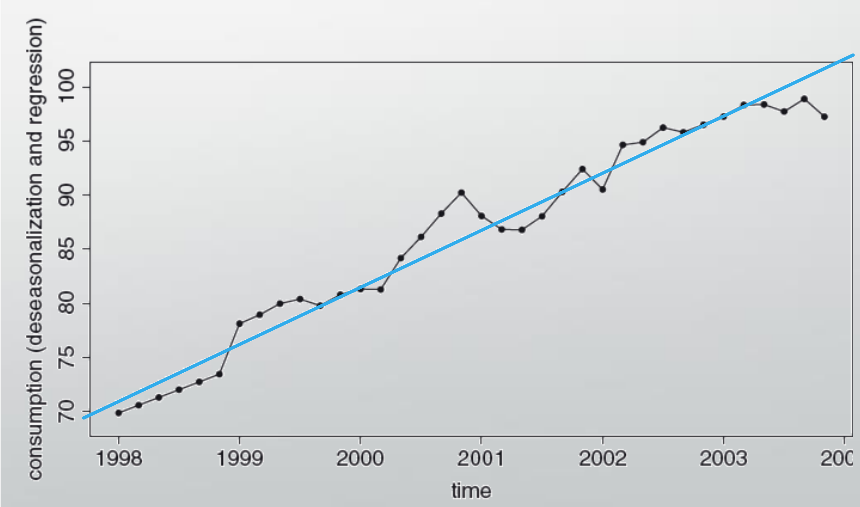


- We re-define **Trend component** (M_t) for C_t based on using **Linear Regression**

Exercise 5

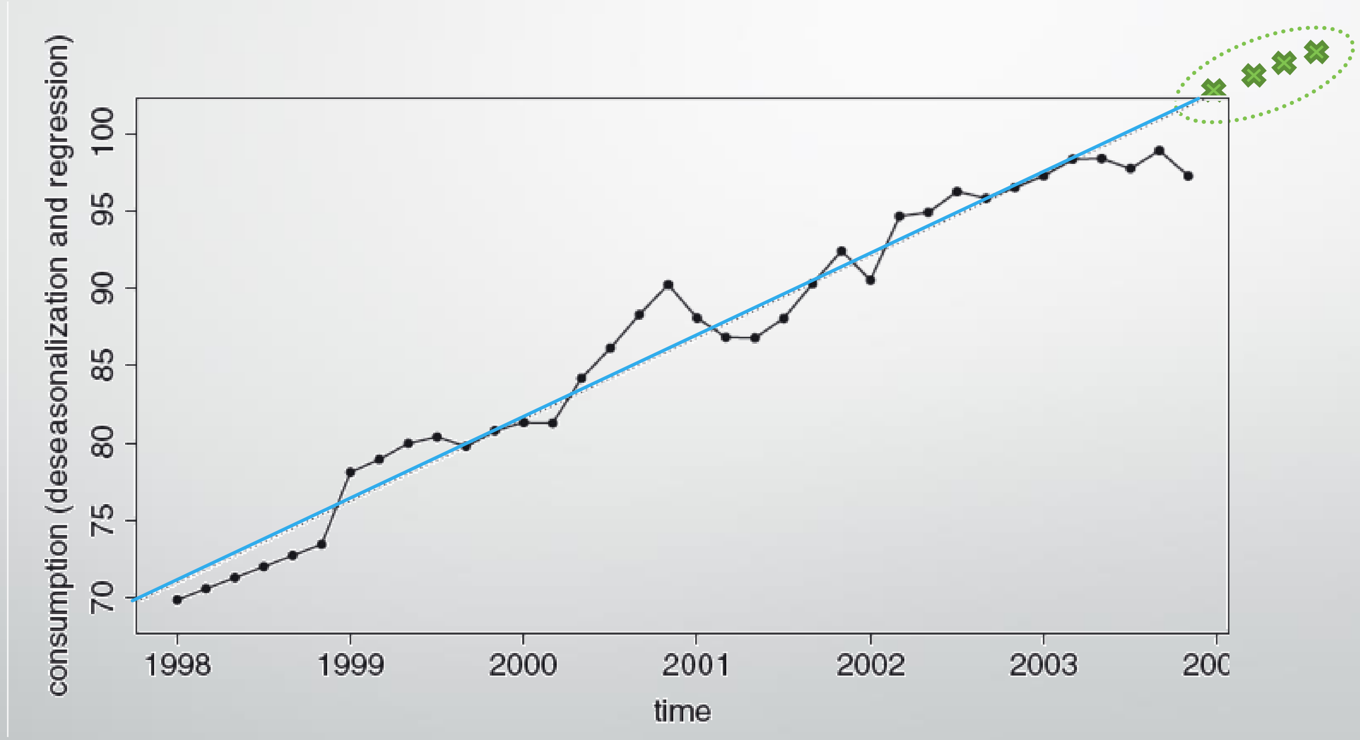
Time Series : *De-seasonalized*

- **Use** the EXCEL sheet (with Exercise 1+2+3+4 completed)
- **Calculate Trend M_t based on Linear Regression** for C_t during 2000-2005
- **Draw a graph** of M_t **and** C_t

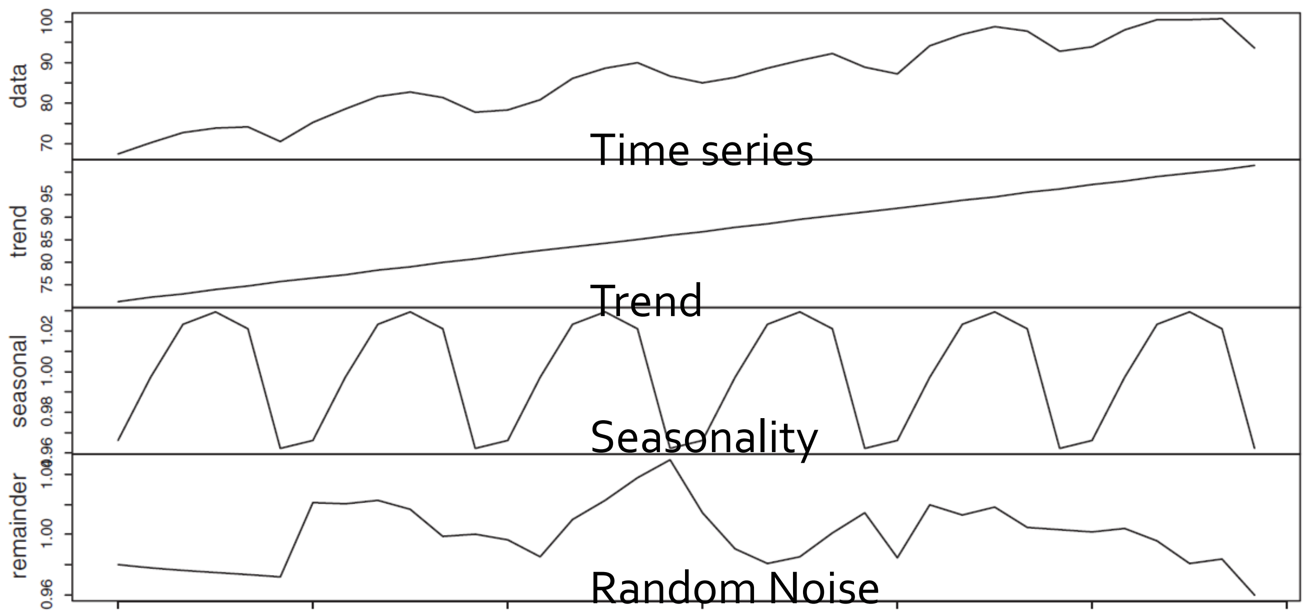


Time Series : *De-seasonalized*

- Can you obtain **future values** of the **seasonality indices** based on the **Trend M_t for De-seasonalized by using Linear Regression**?



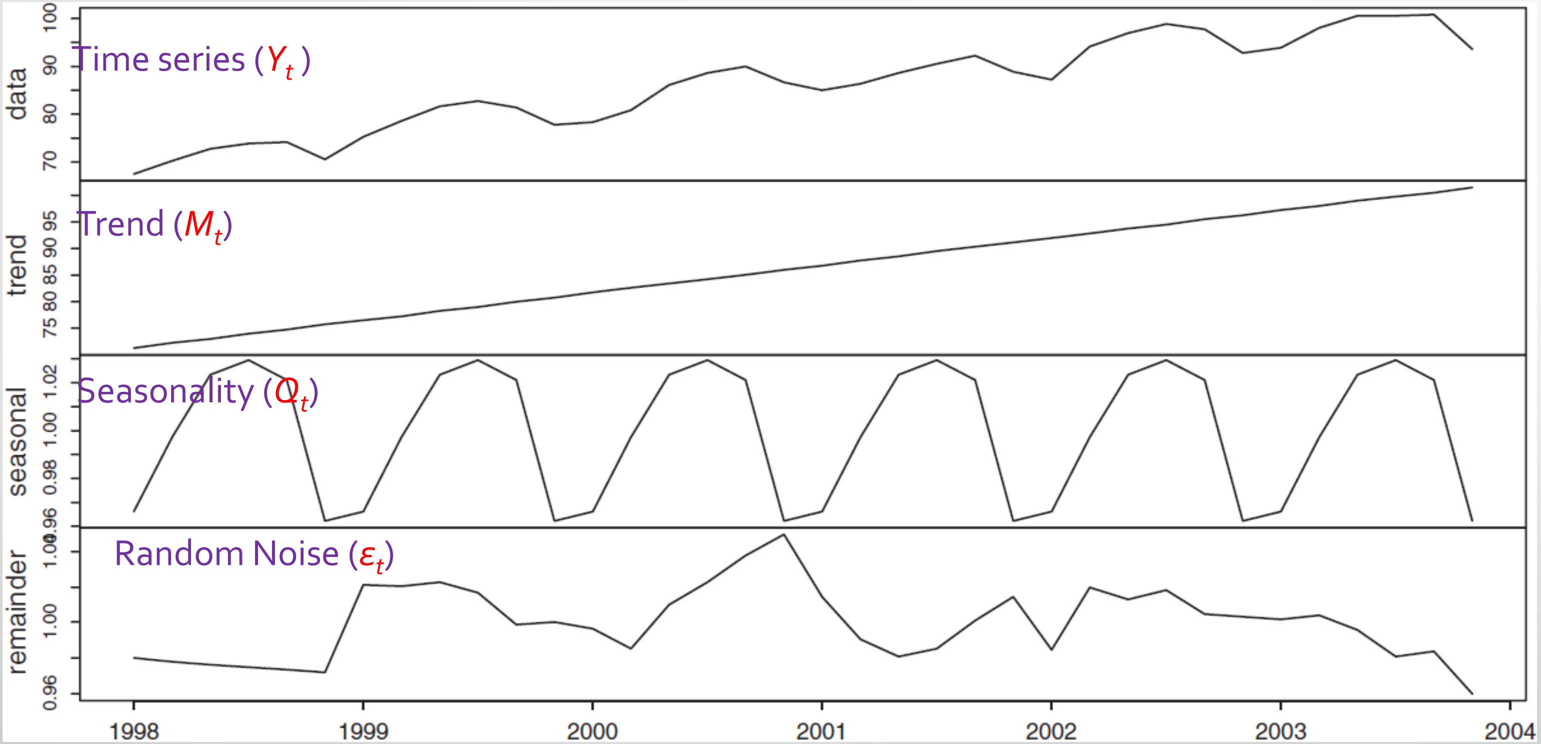
Yes, we can use it as a prediction model



Decomposition of a time series

For the power consumption example

Time Series : *Random noise*



- What is **Random Noise**?
- How do we find the **Random Noise**?

Time Series : *Random noise*

- **Random Noise** (ε_t)

$$Y_t = M_t \times Q_t \times \varepsilon_t$$

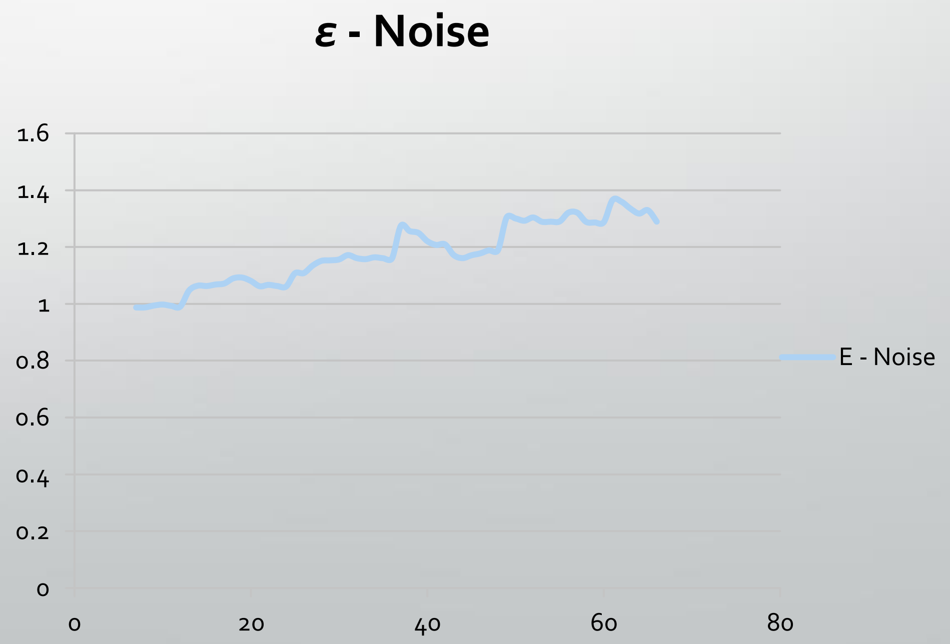
$$\varepsilon_t = Y_t/(M_t \times Q_t)$$

- *where* M_t *is a* ***Trend component*** *for* ***De-seasonalized by using Linear Regression***

Exercise 6

Time Series : *Random noise*

- **Use** the EXCEL sheet (with Exercise 1+2+3+4+5 completed)
- **Calculate** **Random Noise** ε_t **for Trend** M_t during 1998-2004
- **Draw a graph** of ε_t



Time Series : *Summary*

$Y_t = M_t \times Q_t \times \varepsilon_t$

Time series components:

- **Trend M_t based on Moving Average**

$$m_t(h) = \frac{y_{t+h/2} + y_{t+h/2-1} + \cdots + y_{t-h/2+1}}{2h} + \frac{y_{t+h/2-1} + y_{t+h/2-2} + \cdots + y_{t-h/2}}{2h}$$

$$m_t(h) = \frac{y_{t+(h-1)/2} + y_{t+(h-1)/2-1} + \cdots + y_{t-(h-1)/2}}{h}$$

- **Seasonality Q_t – By Trend Removal on Moving Average**

$$Q_t = \sum_{i=1}^{n} B_i \,/n$$
$$B_t \approx Y_t/m_t(h)$$

- **De-seasonalized C_t + Linear Regression on C_t ➔ *Prediction model***

$$C_t = Y_t / Q_t$$

- **Random noise ε_t**

$$\varepsilon_t = Y_t/(M_t \times Q_t)$$

Decomposition of a time series

Now we have a mathematical expression (regression) for the trend, and a periodic set of indices for the seasonality. Thus, we can express the time series based on the following decomposition:

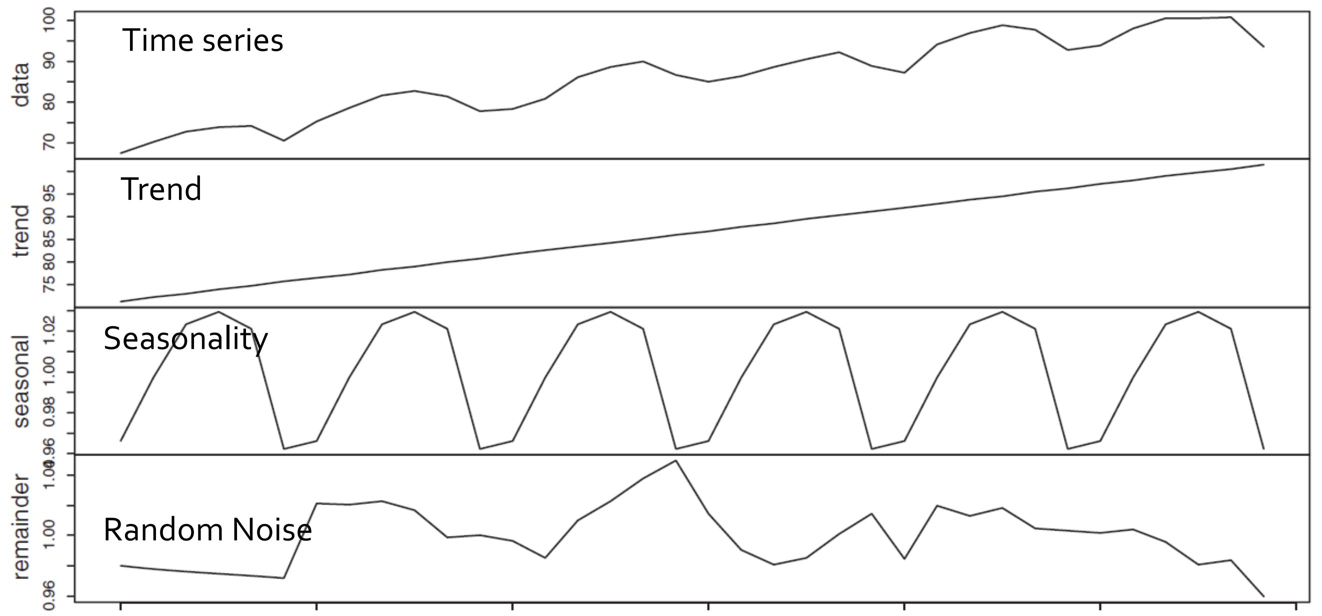
$$f_t = M_t \times Q_{l(t)}$$

We can also forecast future values by

$$f_{t+i} = M_{t+i} \times Q_{l(t+i)}$$

Decomposition of a time series

- For the power consumption example



Evaluating Time Series Models

- Yes, we have a prediction model by using **Trend** (M_t)
  - that utilizes **Linear regression** to predict **the future values** … **That's great!!☺**

- But…. ***How*** to do we **evaluate the accuracy** of the **model**?

Evaluating Time Series Models

- **Evaluation Indicators:**
- Distortion measures
- Dispersion measures
- Tracking signal

Evaluating Time Series Models

Distortion Measures

Given the observations y_t of a time series and the corresponding forecasts f_t , the prediction error is defined as:

$$e_t = y_t - f_t$$

and the percentage prediction error as:

$$e_t^P = \frac{y_t - f_t}{y_t} \times 100$$

Evaluating Time Series Models

Distortion Measures

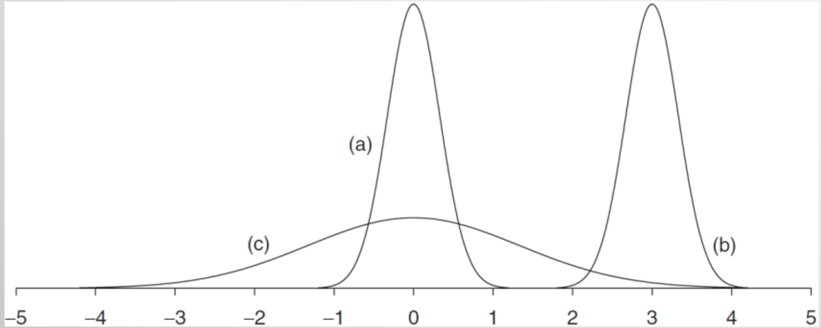
At time index *k* (current time), the mean error (*ME*) is defined as the mean of prediction errors:

$$\text{ME} = \frac{\sum_{t=1}^{k} e_t}{k} = \frac{\sum_{t=1}^{k} (y_t - f_t)}{k}$$

and the mean percentage error (*MPE*) as:

$$\text{MPE} = \frac{\sum_{t=1}^{k} e_t^P}{k}$$

Evaluating Time Series Models



errors density for three models

- Sometimes distortion measures are insufficient to evaluate models because positive and negative prediction errors can cancel out each others, resulting in small distortion indices foe inaccurate models.

Evaluating Time Series Models

Dispersion Measures (1 of 3)

At time index *k* (current time), the mean absolute deviation (*MAD*) is defined as:

$$\text{MAD} = \frac{\sum_{t=1}^{k} |e_t|}{k} = \frac{\sum_{t=1}^{k} |y_t - f_t|}{k}$$

and the mean absolute percentage error as:

$$\text{MAPE} = \frac{\sum_{t=1}^{k} |e_t^P|}{k}$$

Evaluating Time Series Models

Dispersion Measures (2 of 3)

The most popular dispersion measure is the mean square error (*MSE*), defined as:

$$\text{MSE} = \frac{\sum_{t=1}^{k} e_t^2}{k} = \frac{\sum_{t=1}^{k} (y_t - f_t)^2}{k}$$

Unlike *MAD*, *MSE* is a differentiable function.

Evaluating Time Series Models

Dispersion Measures (3 of 3)

The problem with *MSE* is that it amplifies the effect of the largest error, so that the standard deviation of errors (*SDE*) is introduced:

$$\text{SDE} = \sqrt{\frac{\sum_{t=1}^{k} e_t^2}{k}} = \sqrt{\frac{\sum_{t=1}^{k} (y_t - f_t)^2}{k}}$$

Exercise

Rental dataset

20 min

Break

10 min

Prepare your RStudio

```
#load required r packages
library(IRdisplay)
library(magrittr)
library(tidyverse)
library(scales)
library(gridExtra)
library(forecast)
library(tseries)
library(ggthemes)
theme_set(theme_economist())

#load helper R functions
setwd("your working directory")
source("R Functions/compare_models_function.R")
source("R Functions/sim_random_walk_function.R")
print("Loading is completed")
```

Decomposition method practice in R

- Read data:

  ur = read.csv("Mass Monthly Unemployment Rate.csv")

DATE	MAURN
1976-01-01	11.6
1976-02-01	11.3
1976-03-01	10.9
1976-04-01	9.9
1976-05-01	9.4

- Create a time series from the data:

  ur.ts = ts(ur$MAURN, frequency = 12)

ur.ts	Time-Series [1:529] from 1 to 45: 11.6 11.…

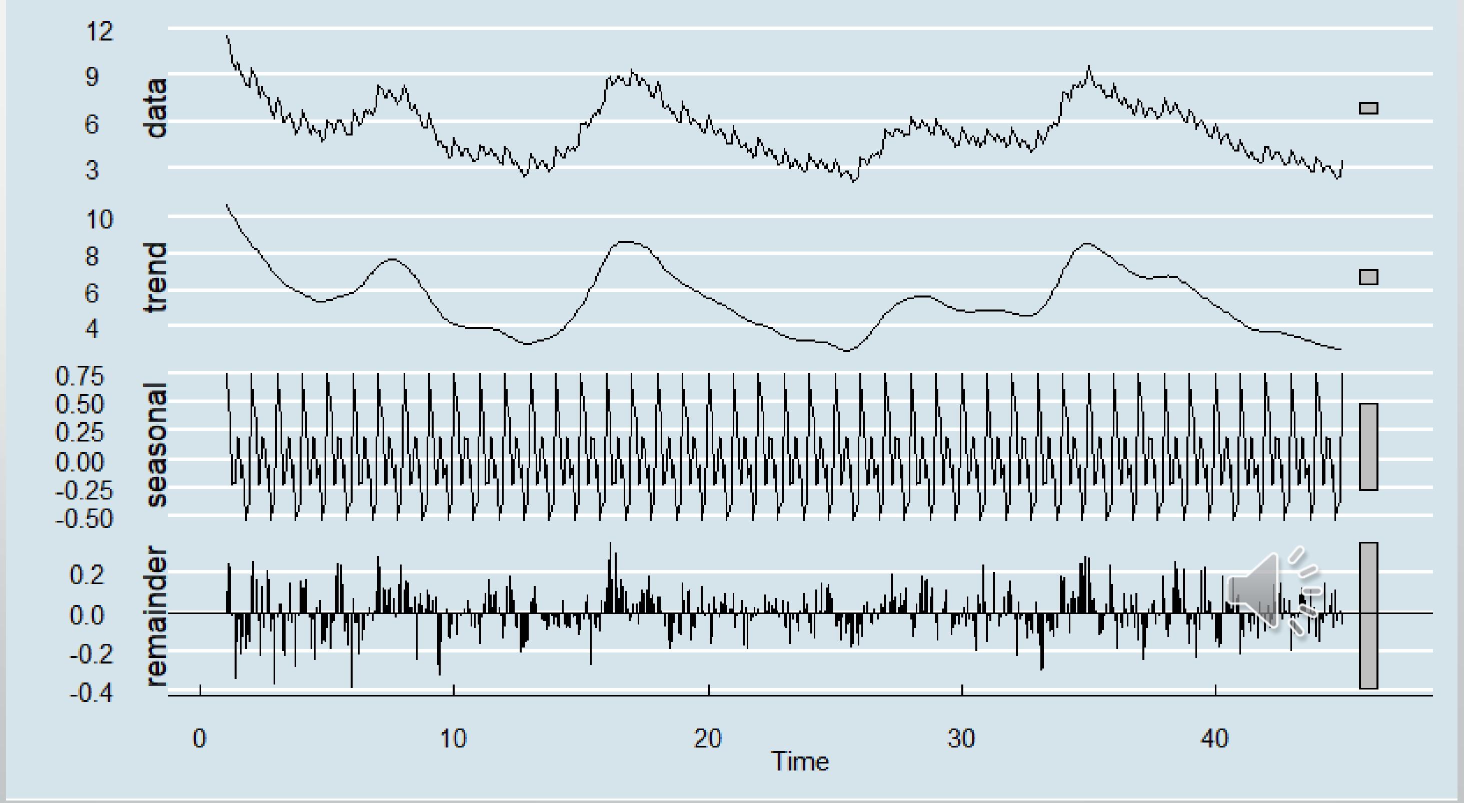
Decomposition method practice in R

- Run a decomposition model (Seasonal Trend Loess)

stl.model = stl(ur.ts,s.window = "periodic")

- Plot the model:

autoplot(stl.model)



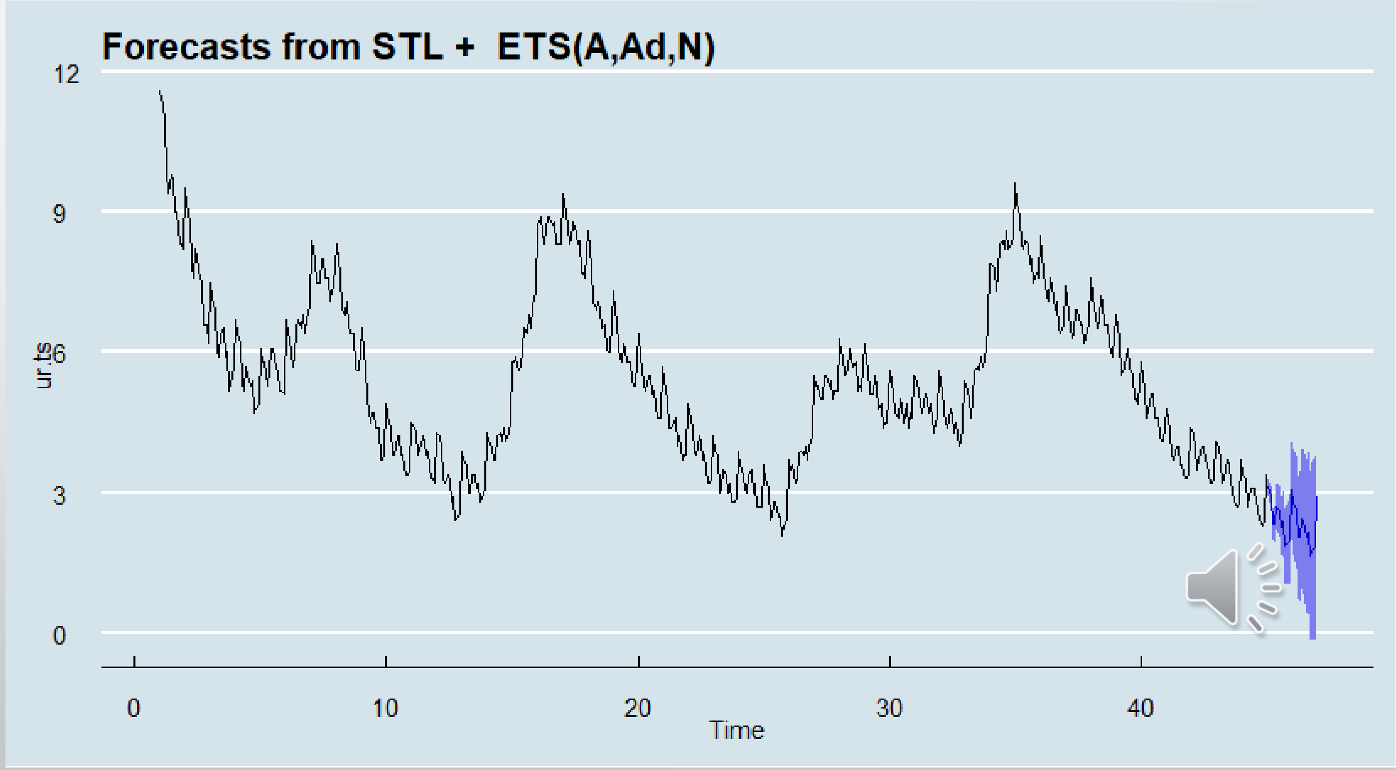
Decomposition method practice in R

- Forecast using the generated time series model:

stl.forecast = forecast(stl.model, h=24, level=80)

- Plot the forecast:

autoplot(stl.forecast)



Exercise

Power Consumption dataset with R

20 min

Questions?