ISCG8043 Adaptive Business Intelligence (ABI)

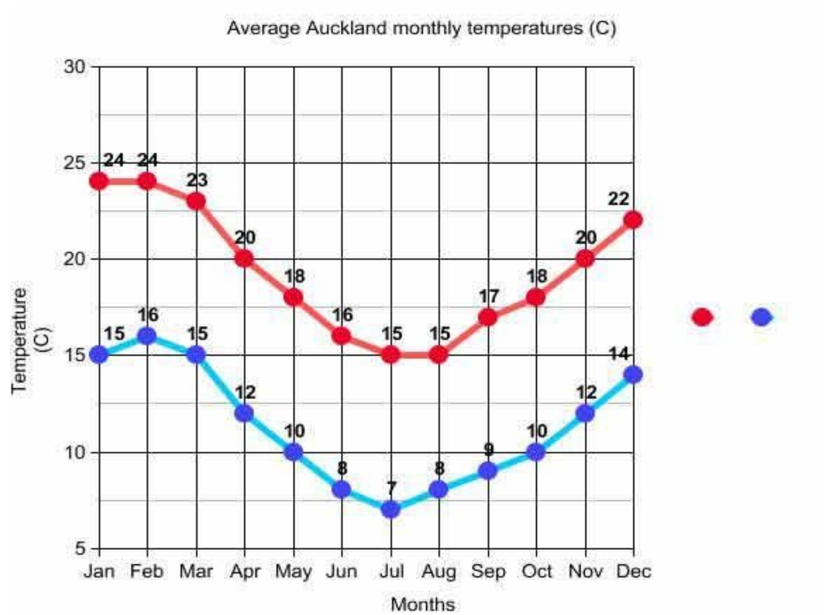
Time Series Analysis (Cont.)

Outline

- Basics and Definitions
- Evaluating Time Series Models
- Decomposition of Time Series

Basics and Definitions

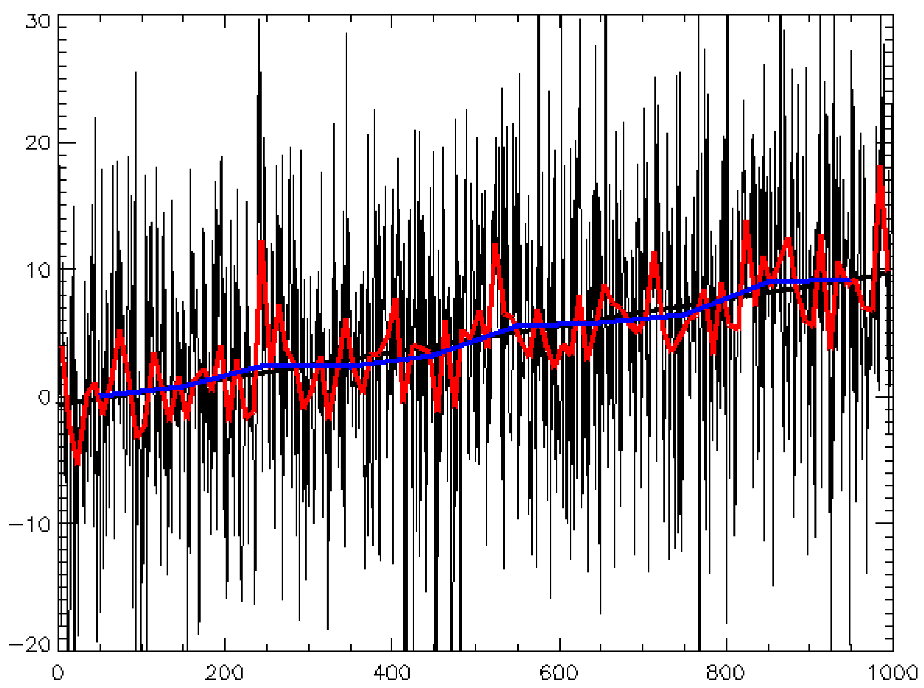
- A time series is a **sequence of data points (values)**, measured typically at
  successive points in time spaced at uniform time intervals.



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

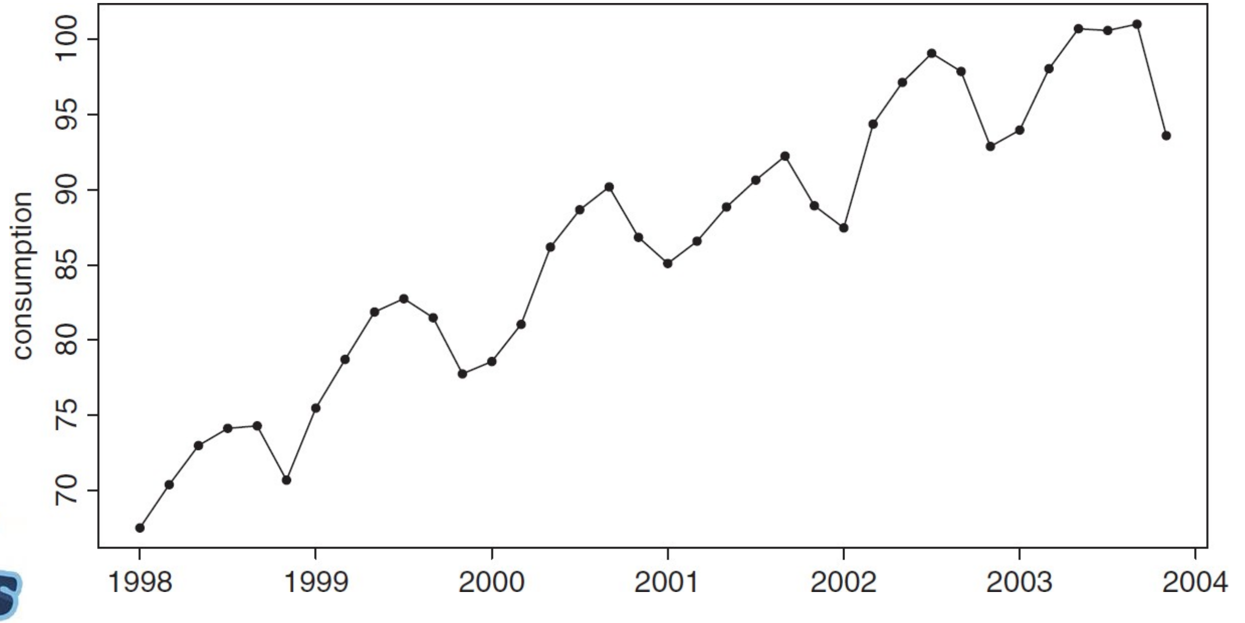
Basics and Definitions

- 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

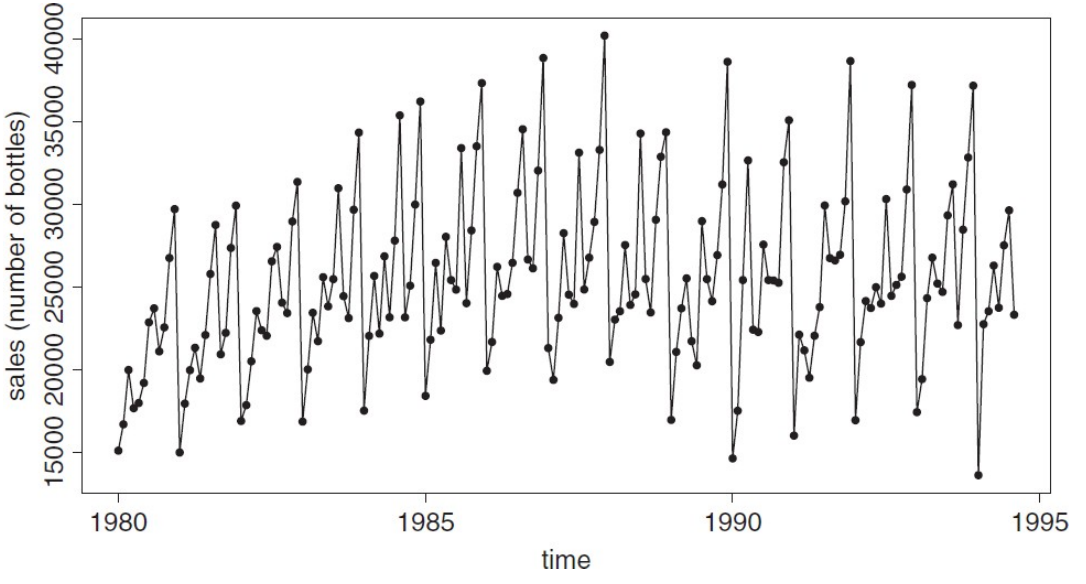
Bimonthly electric assumption (M Watt Hours)in Italy



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

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 months.

Basics and Definitions

- Stochastic Nature of Time Series
  - **Stochastic** = **Random**
  - 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:

$$\mathbf{y_1, y_2, y_3, \ldots, y_k}$$

- **$\mathbf{y_{k+1}}$ and next values are <u>not available</u>**
  - 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-calculates the time series sequence as

  $\mathbf{f_1, f_2, f_3, \ldots, f_k , , f_{k+1}, \ldots,}$ (*f is a forecasted value*)

- It **can calculate future values** as well because it's a mathematical model

Basics and Definitions

- The <u>amount of money</u> you have spent daily on transportation from home to Unitec (bus, fuel or etc) during the **last 7 days** (week) can be represented by **a time series**.
  - Specify the actual time series: $\mathbf{y_1, y_2, \ldots, y_7}$
  - 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: $f_1, f_2, \ldots, f_7$.
- You may do this exercise in Excel

Decomposition of Time Series

Decomposition

Decomposit
ion

Decomposition of Time Series

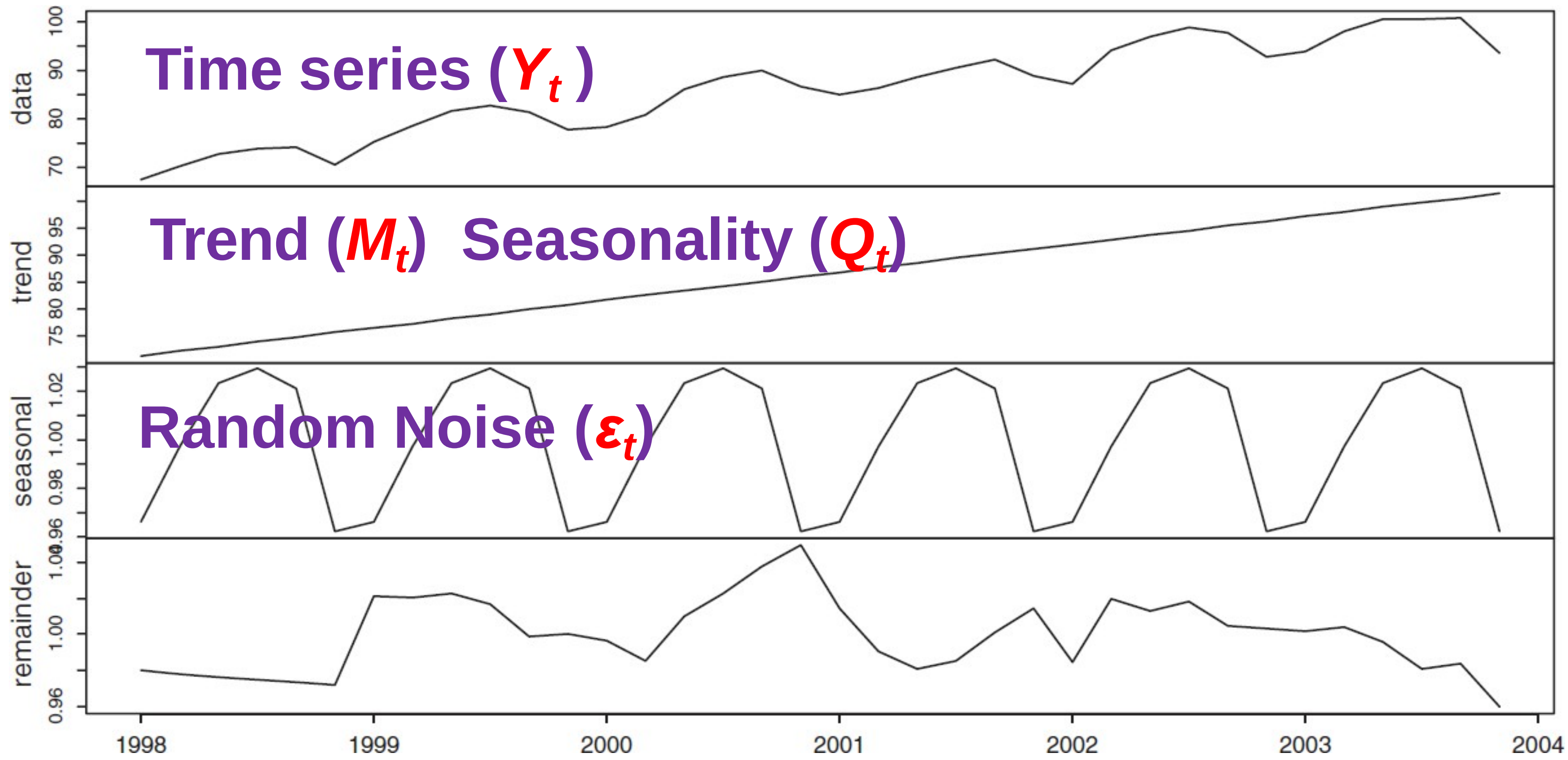
Time Series Components

- ☐ **Trend** M_t
- ☐ **Seasonality** Q_t
- ☐ **Random noise** ε_t

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

Decomposition of a time series

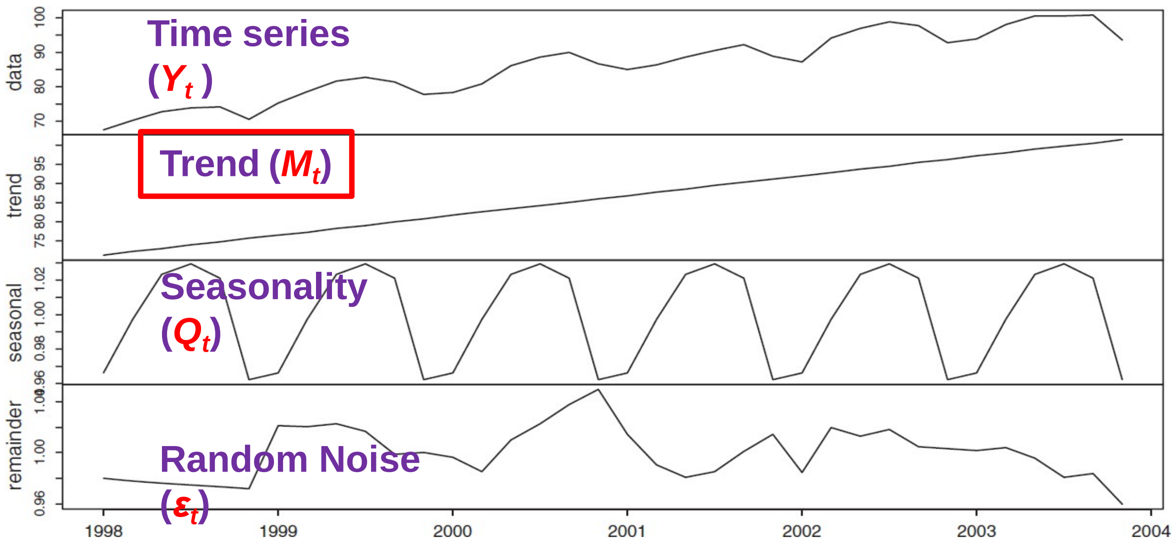
- For the power consumption example



Decomposition of Time Series

Trend (M_t)

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

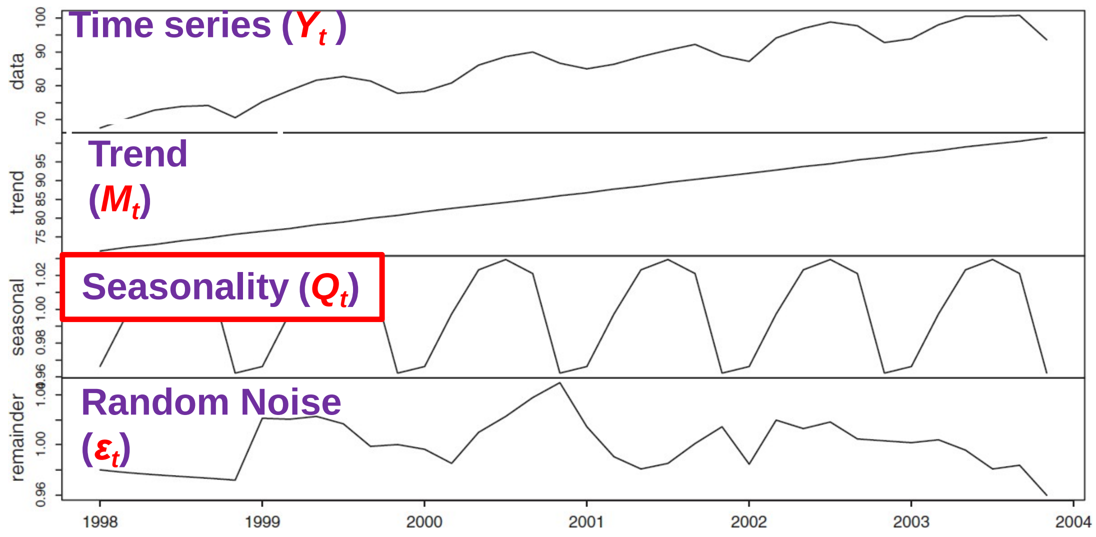


- The average behaviour of a time series over time
- It can be increasing, decreasing or stationary

Decomposition of Time Series

Seasonality (Q_t)

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

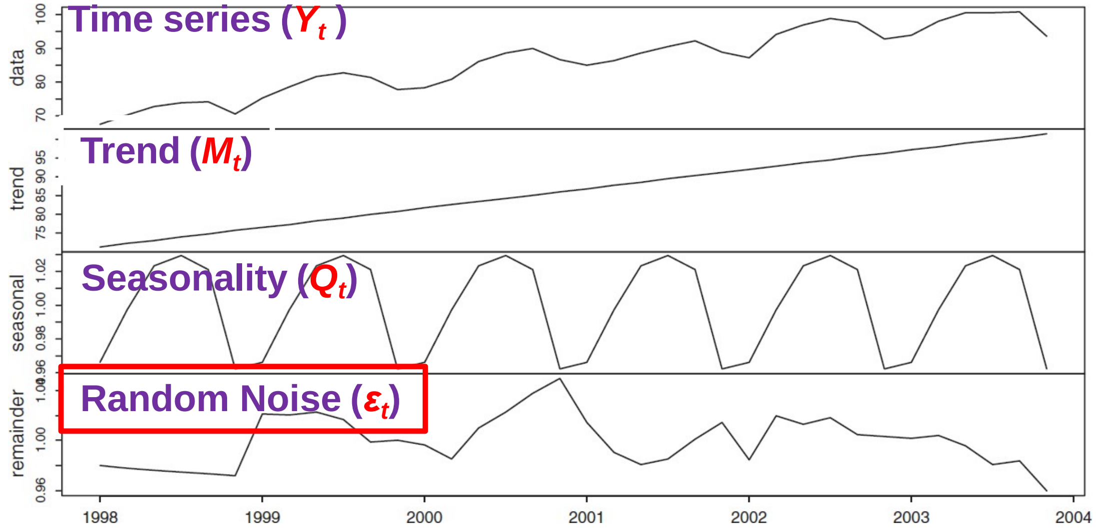


- Wavelike **short-term fluctuations** of regular frequency appear in the values of a time series.

- Determined by the **natural cycles** by which demand develops, or by the **seasonality of the products** to which the time series refers

Decomposition of Time Series

Random Noise (ε_t)



- **All irregular variations** in the data that **cannot** be explained by the other components

Decomposition of Time Series

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

$$\boxed{Y_t = g(M_t, Q_t, \varepsilon_t)}$$

- where *g* represents a function to be selected
  - A **muliplicative model** is

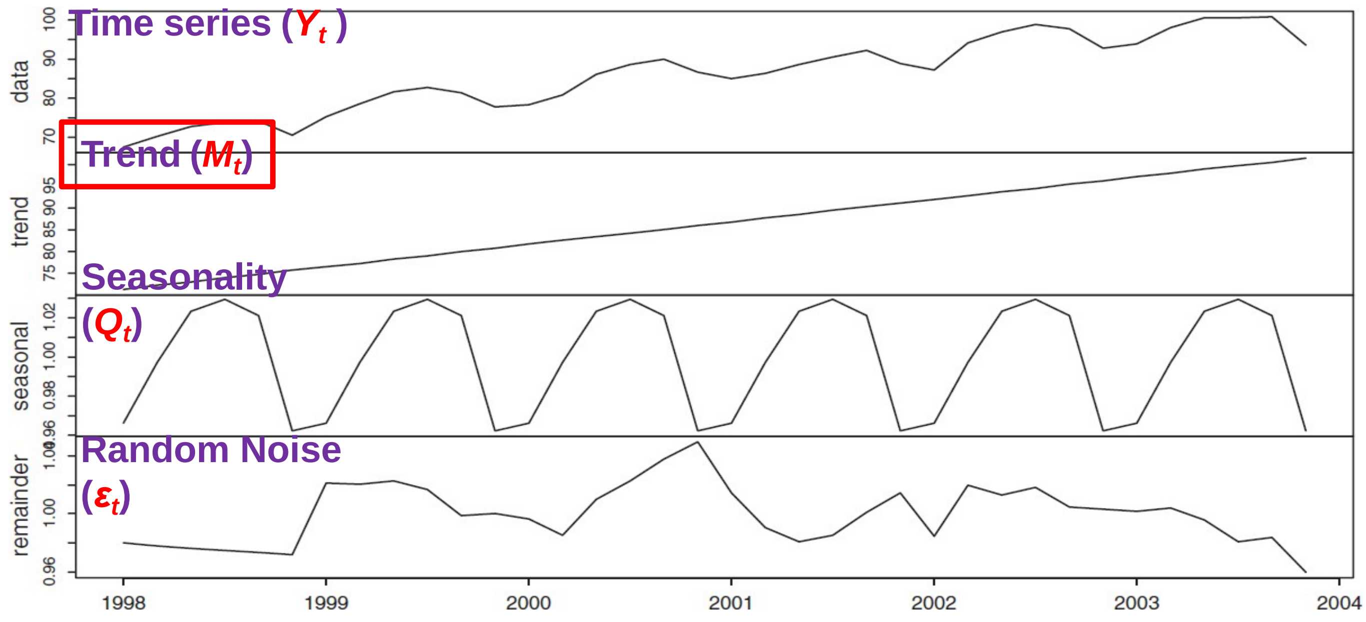
$$\boxed{Y_t = M_t \, x \, Q_t \, x \, \varepsilon_t}$$

  - Alternatively, an **additive model** is

$$\boxed{Y_t = M_t + Q_t + \varepsilon_t}$$

Time Series : *Trend*

- What is **Trend**?



- How do we find the **Trend** ?

Time Series : *Trend*

Moving Average (an estimation of Trend M_t)

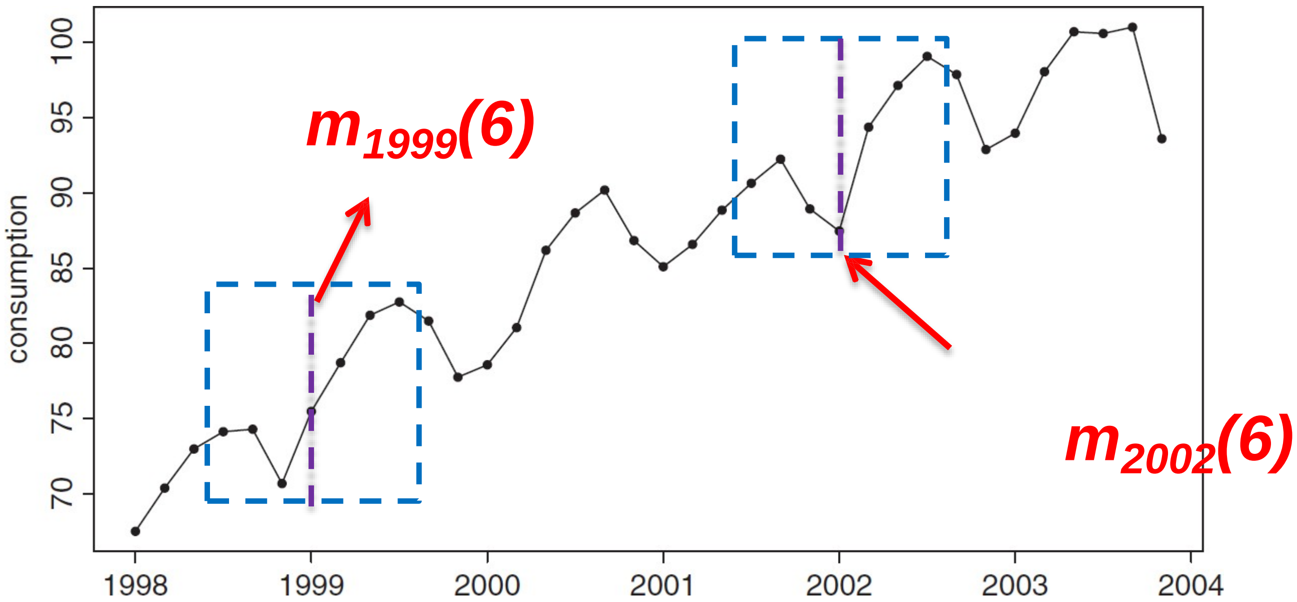
$$\boxed{m_t(h) \approx M_t}$$

- 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.
- There are various methods of calculating **Moving Average**
  - **Different methods give different results (**http://en.wikipedia.org/wiki/Moving_average**)**
  - **For example, here we learn "Centred Moving Average"**

Time Series : *Trend*

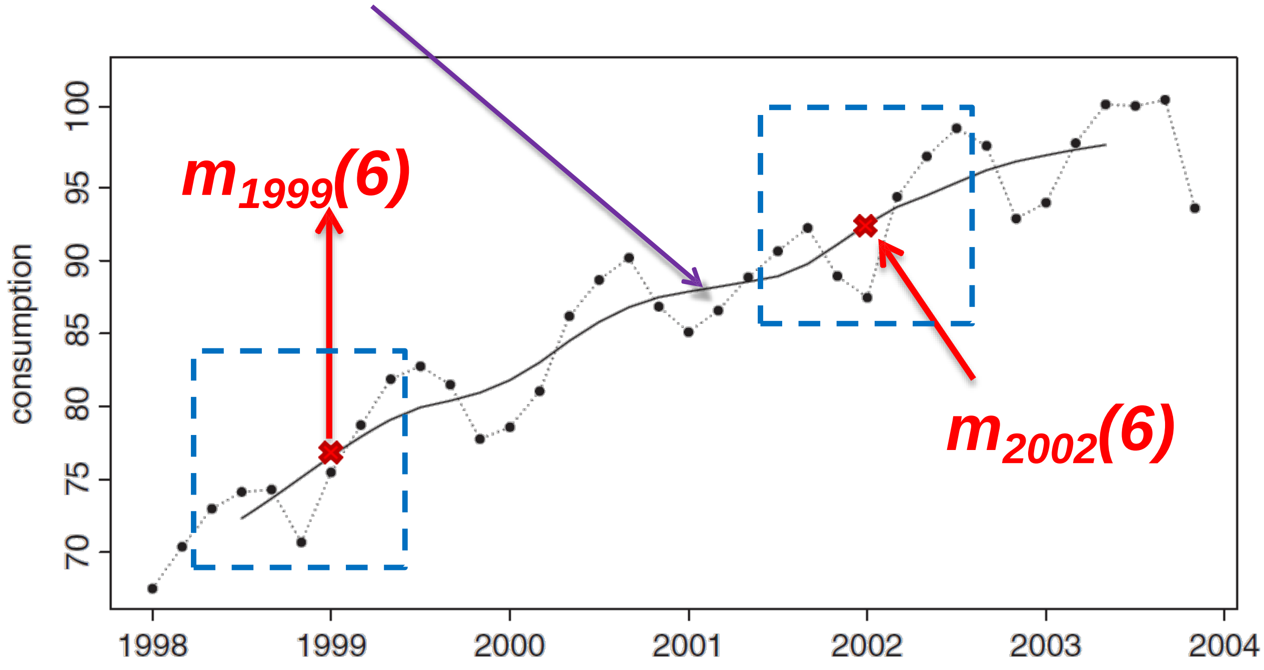
$m_t(h) \approx M_t$ based on Centred Moving Average

- For example, we assume h=6 for 6 *consecutive observations*



Time Series : *Trend*

- If we calculate $m_t(6)$ from **1998 – 2004**, we can draw a **trend line** (M_t)



Time Series : *Trend*

Centred Moving Average $m_t(h)$

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

- If *h* is **even**:

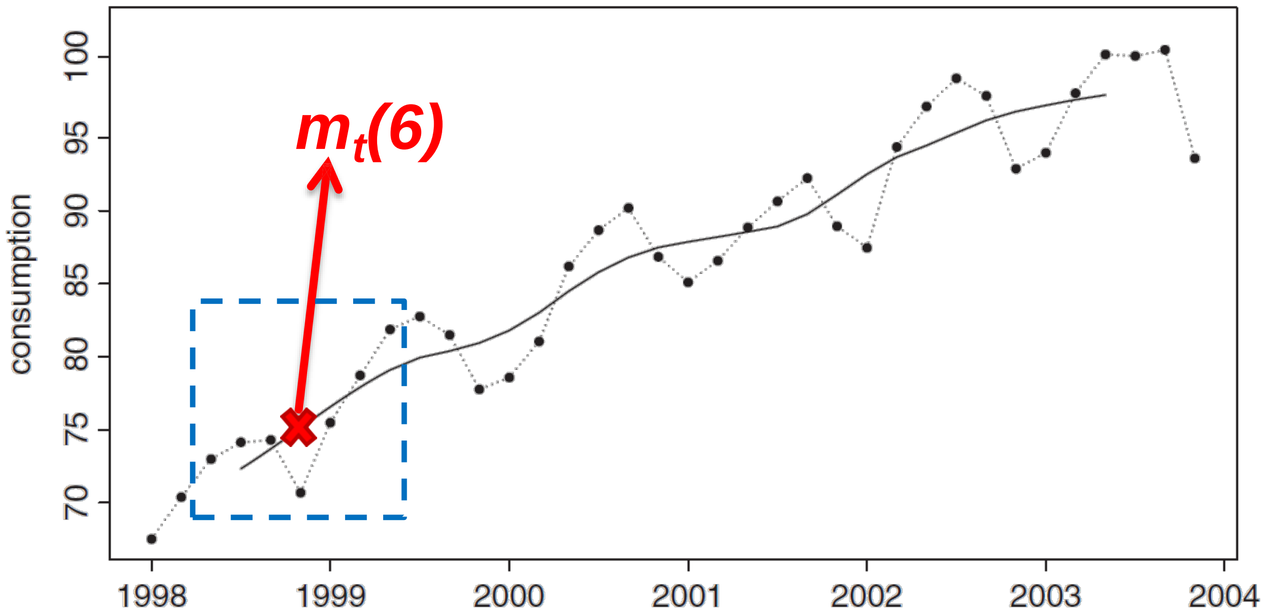
$$m_t(h) = \left(y_{t-\left(\frac{h}{2}\right)} + 2\left(\sum_{i=-\left(\frac{h}{2}-1\right)}^{\left(\frac{h}{2}-1\right)} y_{t+i} \right) + y_{t+\left(\frac{h}{2}\right)} \right) / 2h$$

- If *h* is **odd**:

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

Time Series : *Trend*

- **Let's do this one $m_t(6)$**

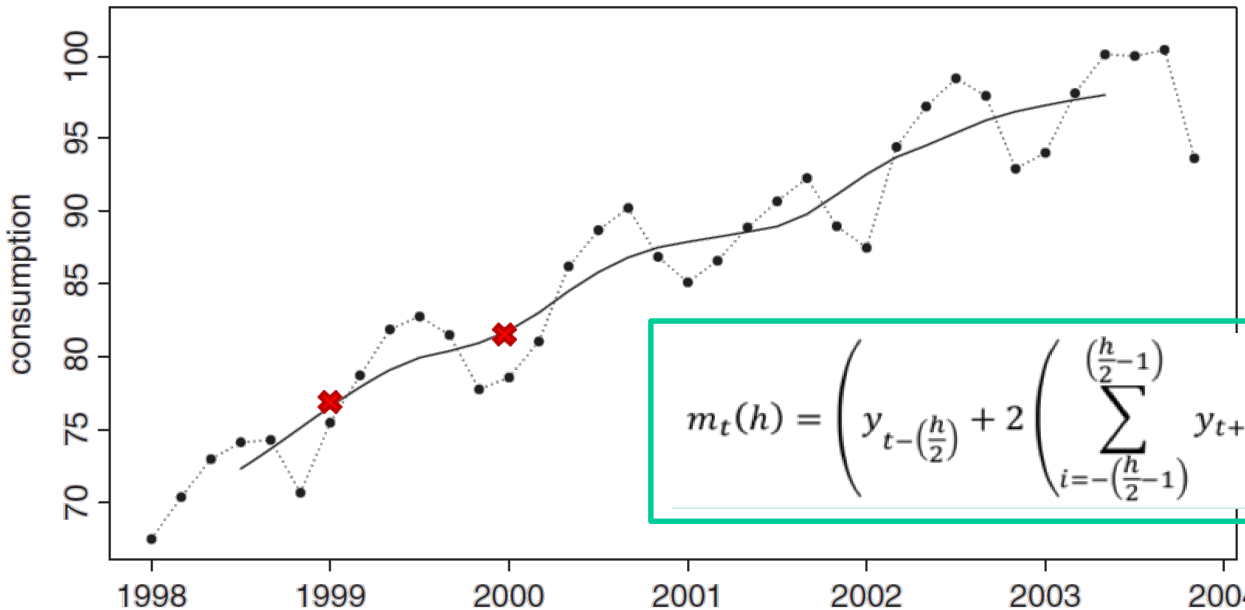


- $m_t(6) = (y_{t+3} + 2(y_{t+2} + y_{t+1} + y_t + y_{t-1} + y_{t-2}) + y_{t-3}) / \ 2 \times 6$

$$= (83+2(79+75+70+74+74)+73) \ / \ 12 \approx \mathbf{75}$$

Time Series : *Trend*

- **Try more, calculate M_t for $m_{1999}(6)$ and $m_{2000}(6)$**



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

- $m_{1999}(6)$ $= (y_{t+3} + 2(y_{t+2} + y_{t+1} + y_t + y_{t-1} + y_{t-2}) + y_{t-3})/2\text{x}6$

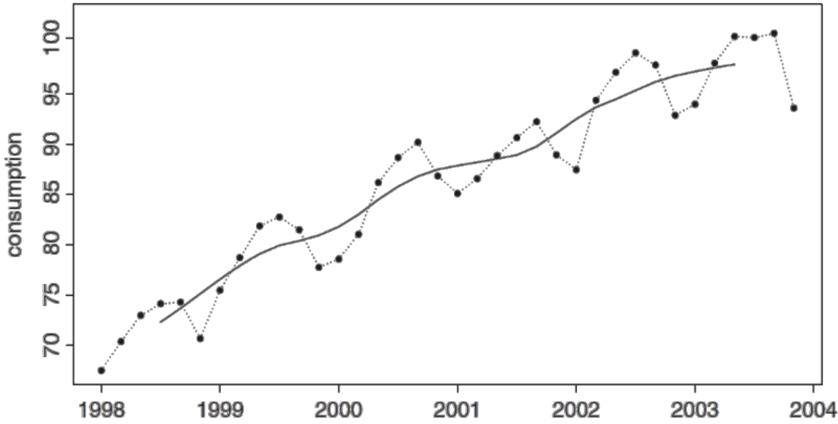
- $m_{2000}(6)$ $=$

Time Series : *Trend*

Exercise 1

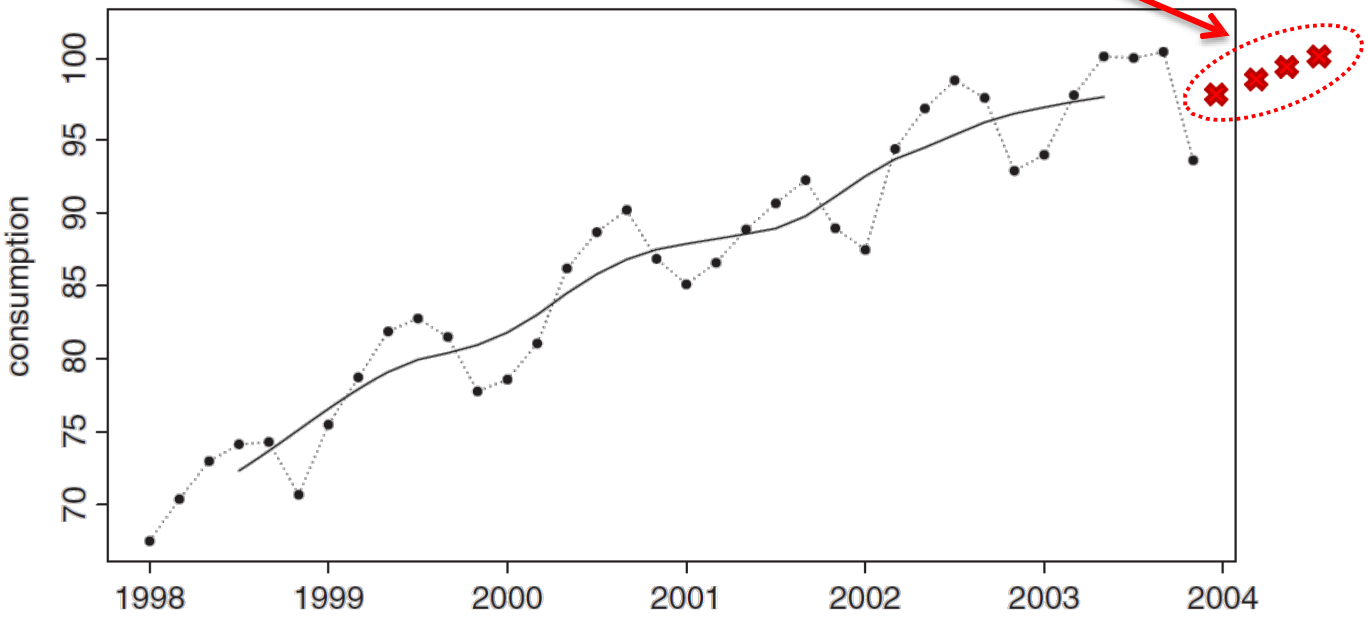
- **Open** the EXCEL sheet on the Moodle
- **Calculate** M_t for all periods (from 2000-2005) using h=6
- **Draw a graph** for M_t

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



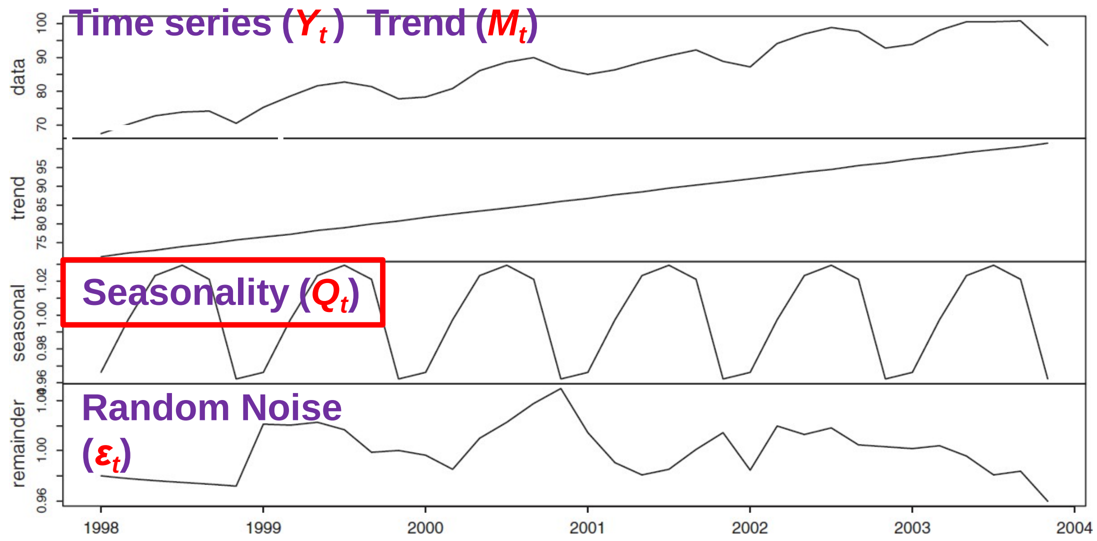
Time Series : *Trend*

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



Time Series : *Seasonality*

- What is **Seasonality**?



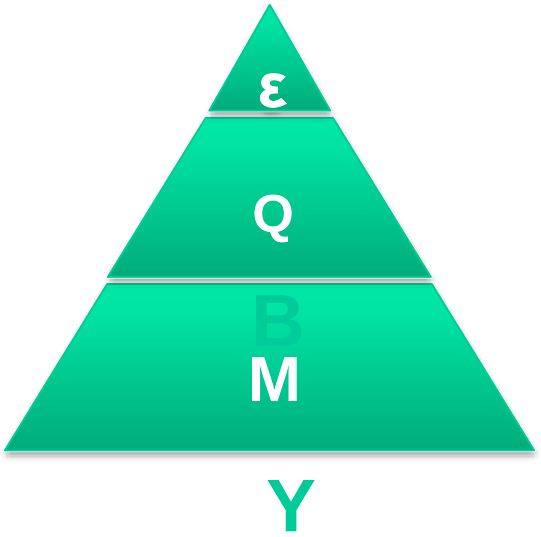
- How do we find the **Seasonality**?

Time Series : *Seasonality*

Removal of the Trend Component M_t

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

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



Time Series : *Seasonality*

Removal of the Trend Component M_t

- For a multiplicative time series model

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

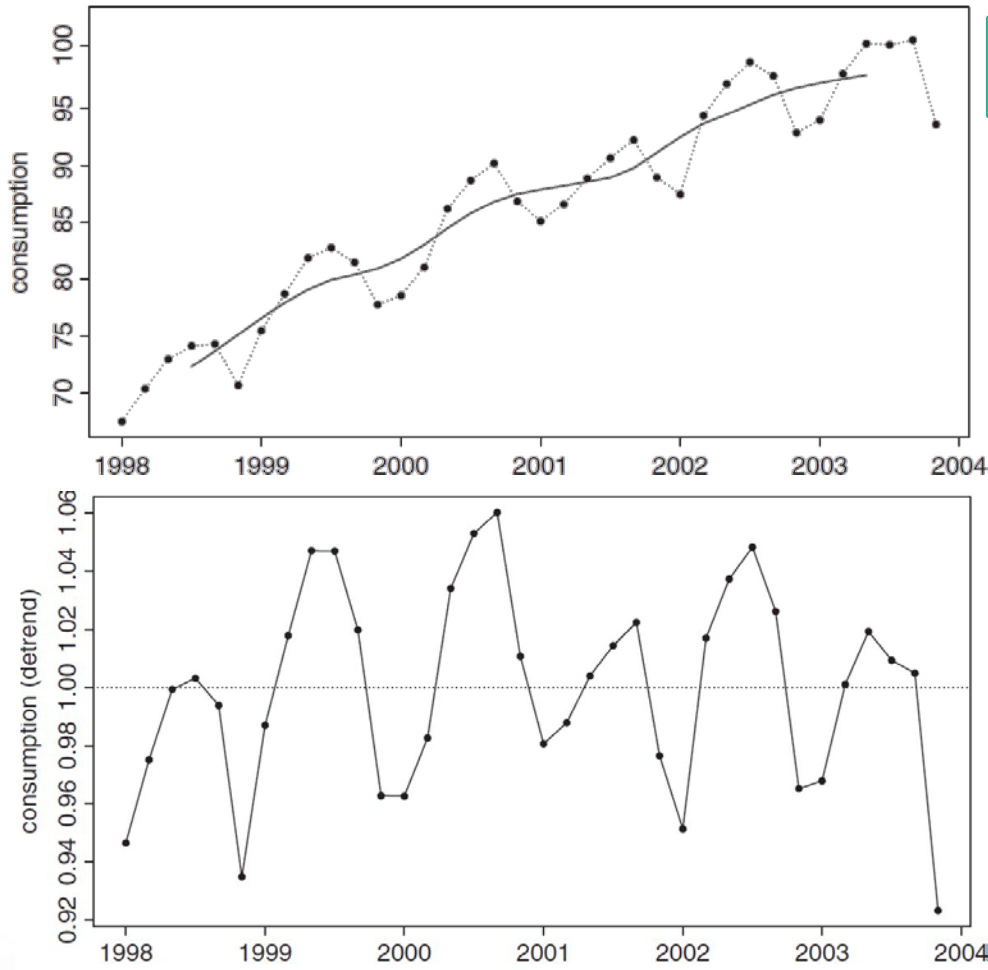
- In a time series B_t (with M_t removed), we have

$$\begin{aligned} B_t &= Q_t \times \varepsilon_t \\ &= Y_t / M_t \\ &\approx Y_t / m_t(h) \end{aligned}$$

Later, B_t is used to calculate Seasonality

Time Series : *Seasonality*

- B_t **for the power consumption example**



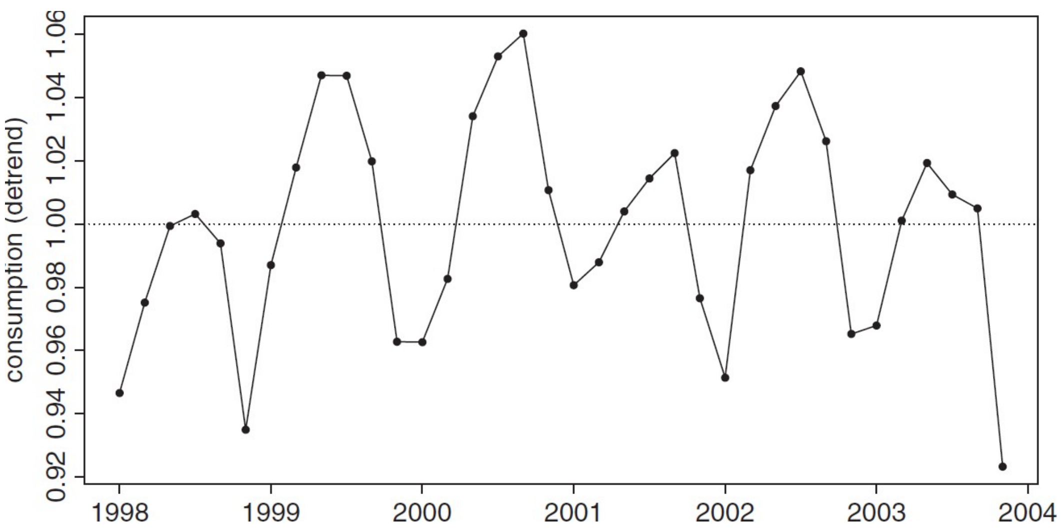
$M_t \approx m_t(h)$

$B_t \approx Y_t / m_t(h)$

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

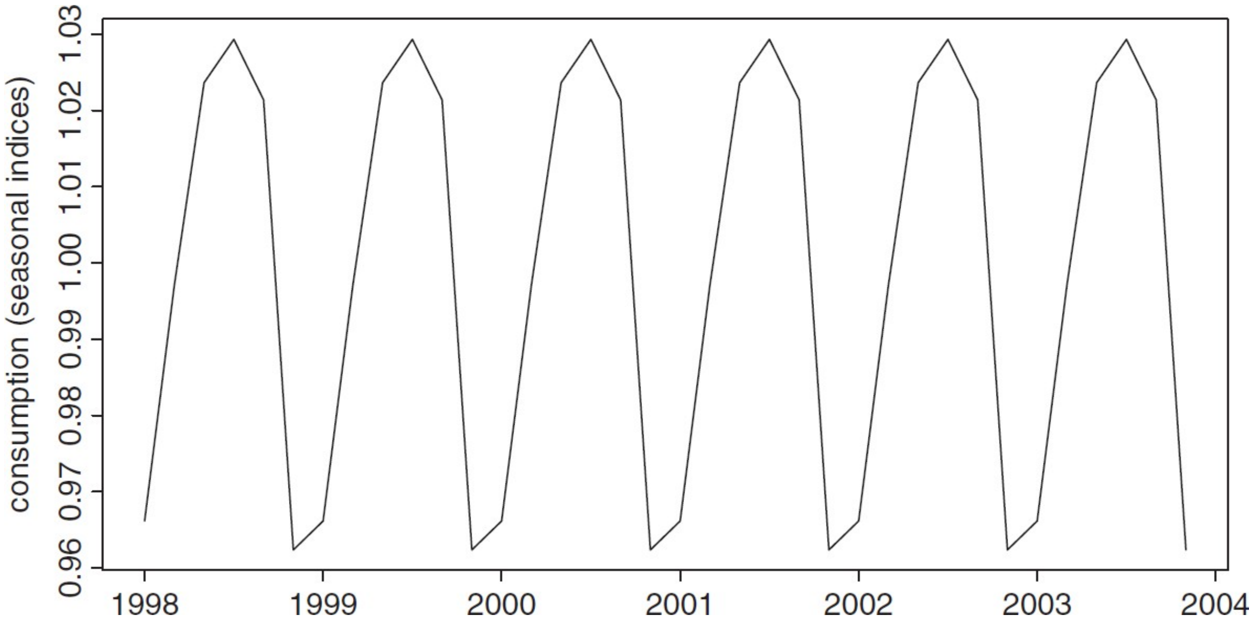
$$Q_t = \sum_{i=1}^{n} B_i \,/ n$$

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

Time Series : *Seasonality*

Seasonality (Q_t) for the power consumption example:

- Based on **monthly season**

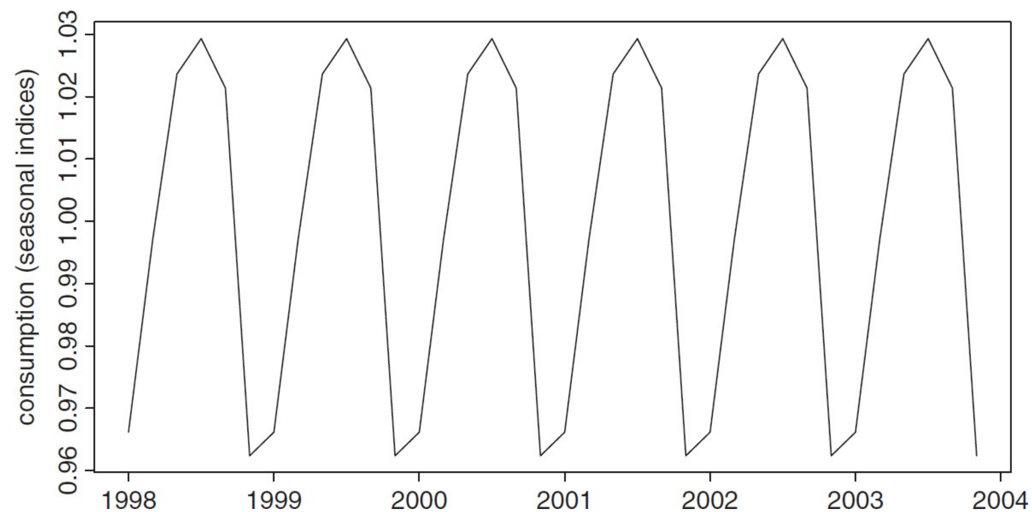


- Notice the **same month** of every year has **<u>the same</u>** Q_t

Time Series : *Seasonality*

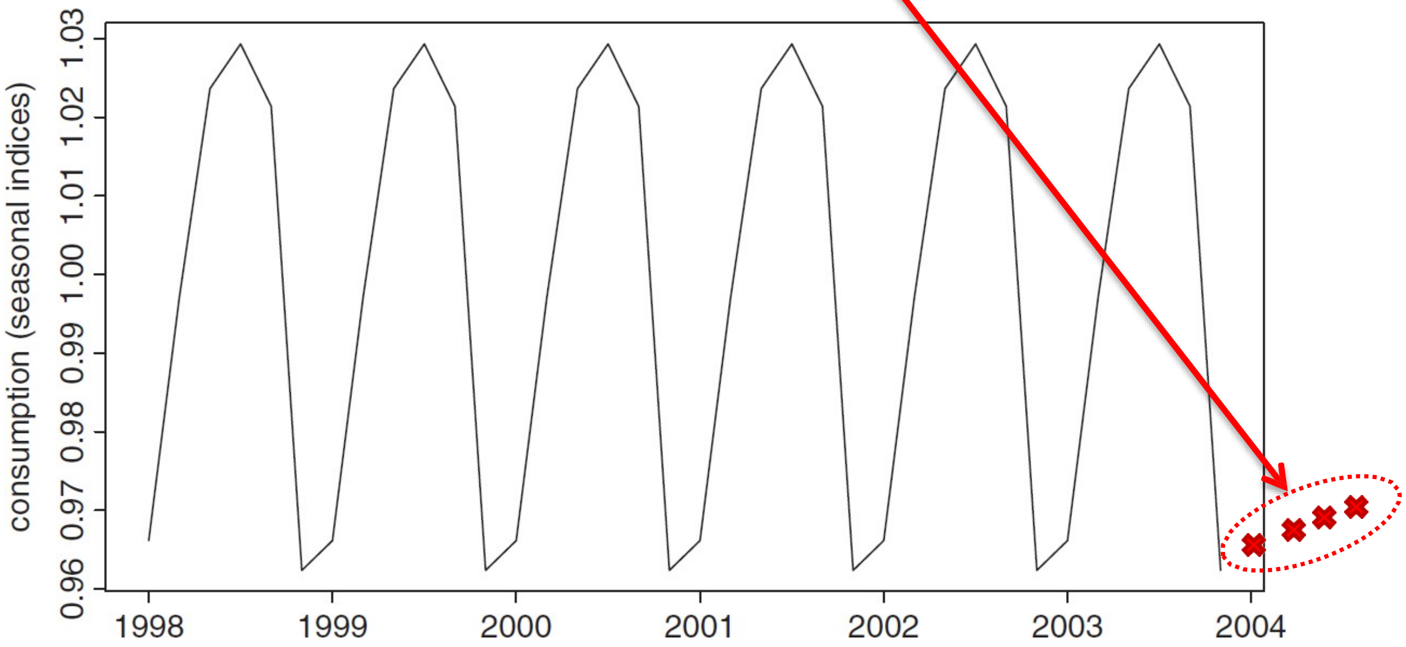
Exercise 3

- **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**?

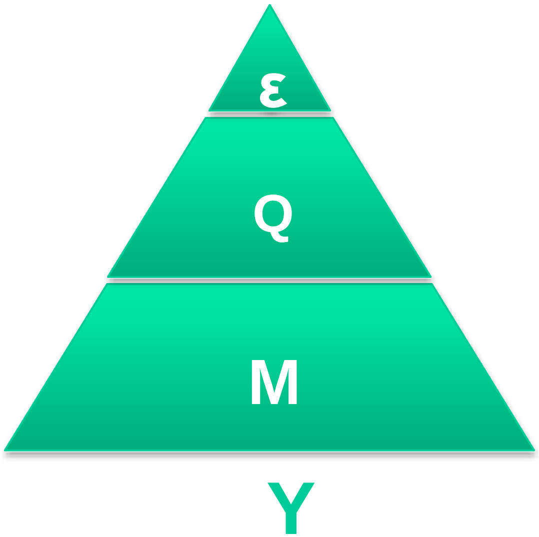


Time Series : *De-seasonalized*

Removal of the Season Component Q_t

$Y_t = M_t \, x \, Q_t \, x \, \varepsilon_t$

- 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

- **Called de-seasonalization**

$$
\begin{aligned}
C_t &= M_t \times \varepsilon_t \\
&= (Y_t / (Q_t \times \varepsilon_t)) \times \varepsilon_t \\
&= \mathbf{Y_t / Q_t}
\end{aligned}
$$

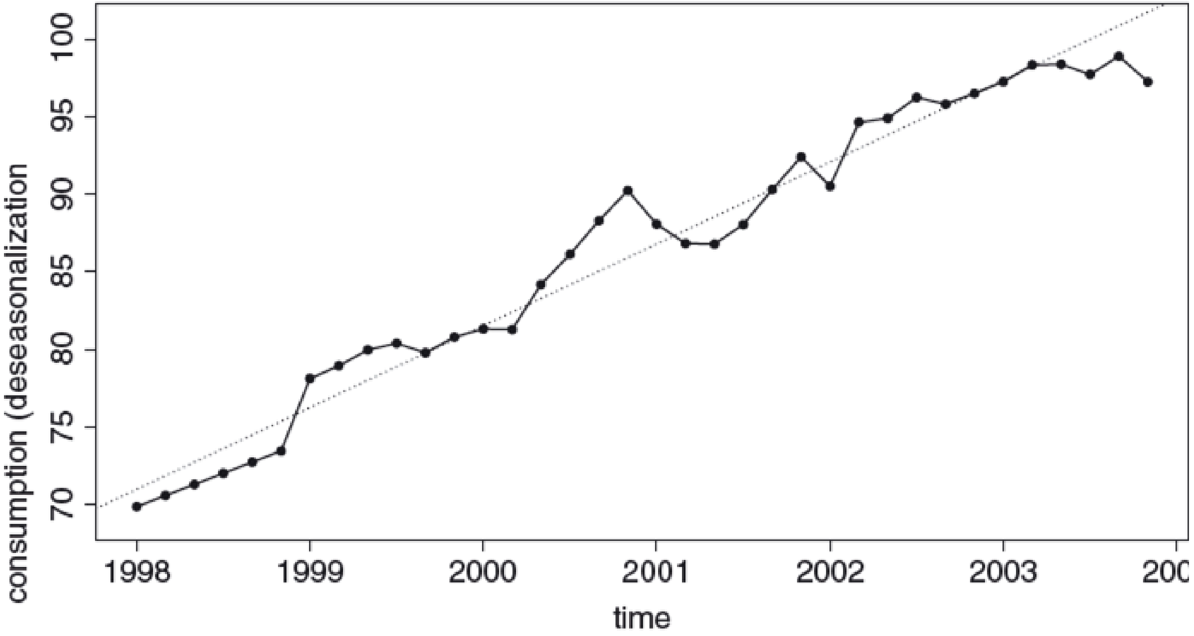
$$Y_t = M_t \times Q_t \times \varepsilon_t$$

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

Time Series : *De-seasonalized*

De-seasonalization for the power consumption example

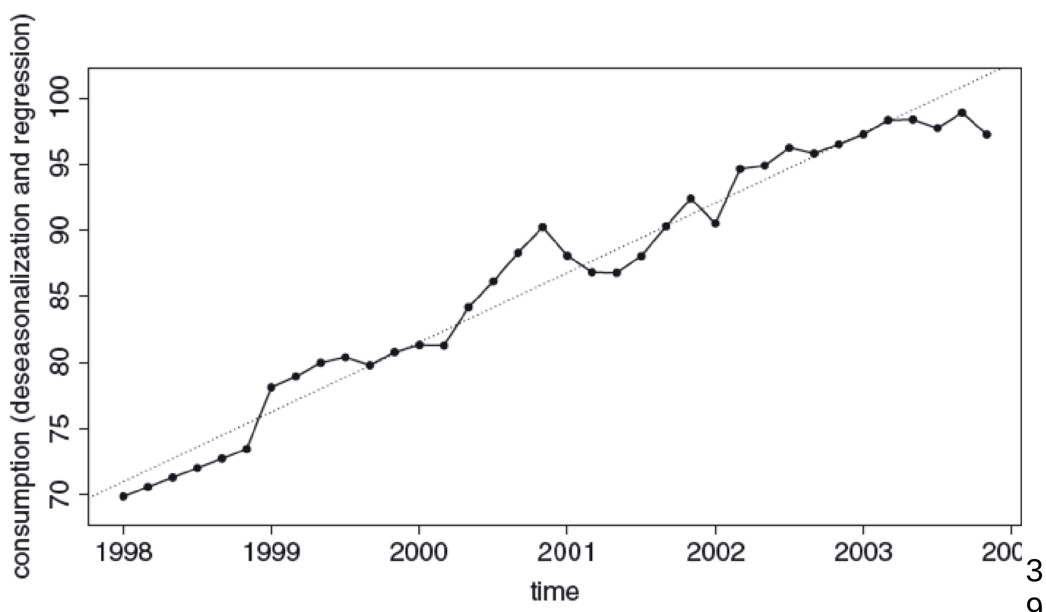
$C_t = Y_t / Q_t$



Time Series : *De-seasonalized*

Exercise 4

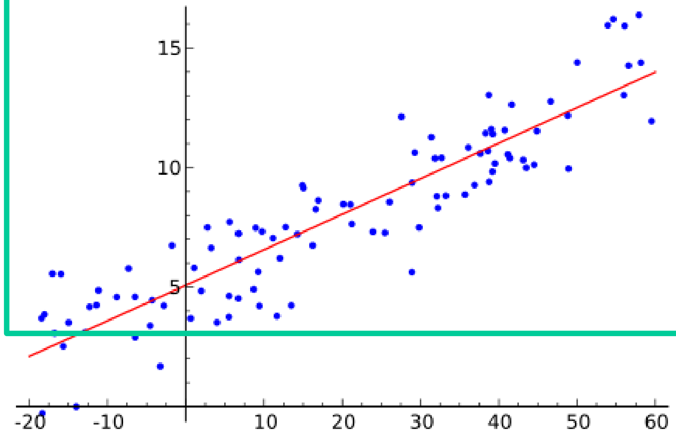
- **Use** the EXCEL sheet (with Exercise 1+2+3 completed)
- **Calculate** C_t for 2000-2005
- **Draw a graph** for C_t



Time Series : *De-seasonalized*

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

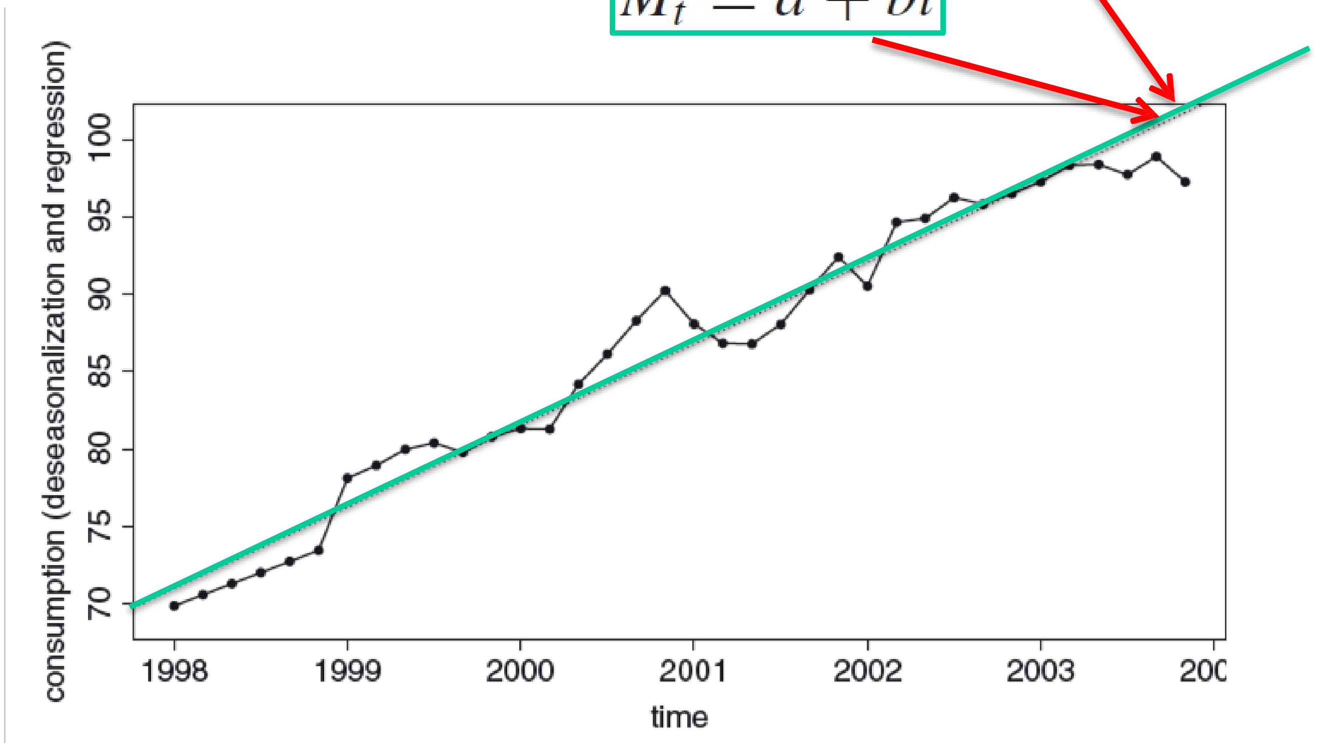
- We look for **an analytical model** for the trend that *can be used for forecasting*
  - **Linear regression** ?!?



Time Series : *De-seasonalized*

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

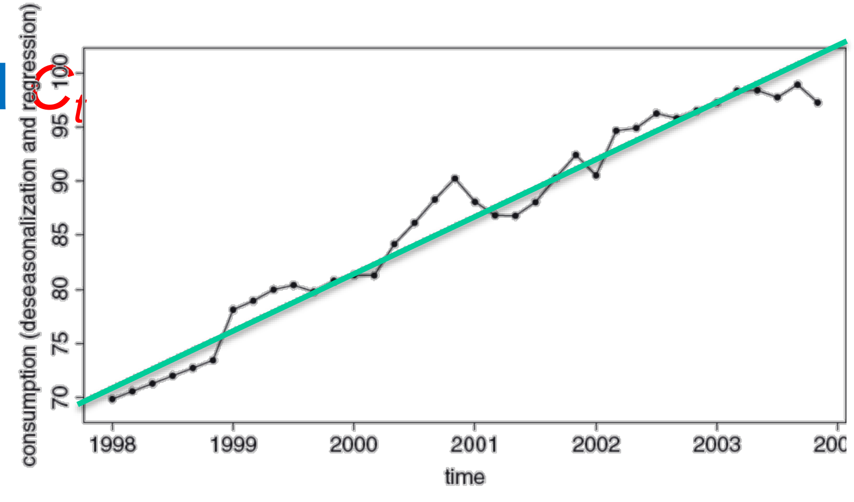
$$\boxed{M_t = a + bt}$$



Time Series : *De-seasonalized*

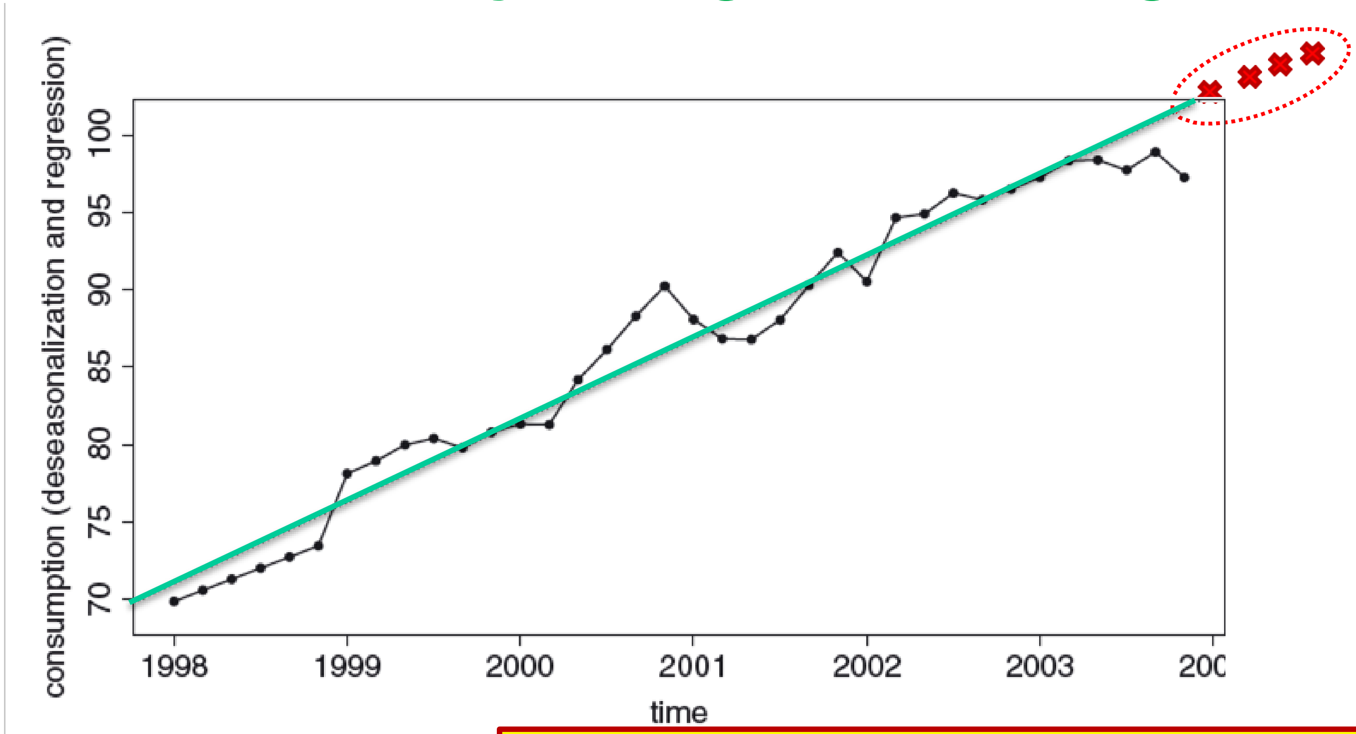
Exercise 5

- **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*

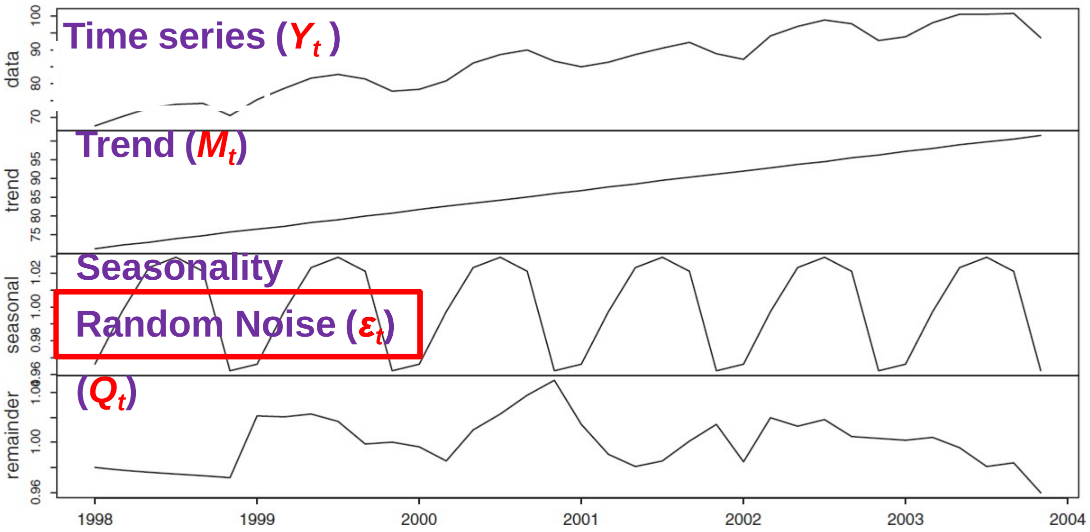
- Can you obtain **future values** of the **seasonality indices** based on the **Trend M_t for De-seasonalized by using Linear Regression**?



Yes, we can use it as a prediction model

Time Series : *Random noise*

- What is **Random Noise**?



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

Time Series : *Random noise*

- **Random Noise** (ε_t)

$$\boxed{Y_t = M_t \times Q_t \times \varepsilon_t}$$

$$\boxed{\varepsilon_t = Y_t / (M_t \times Q_t)}$$

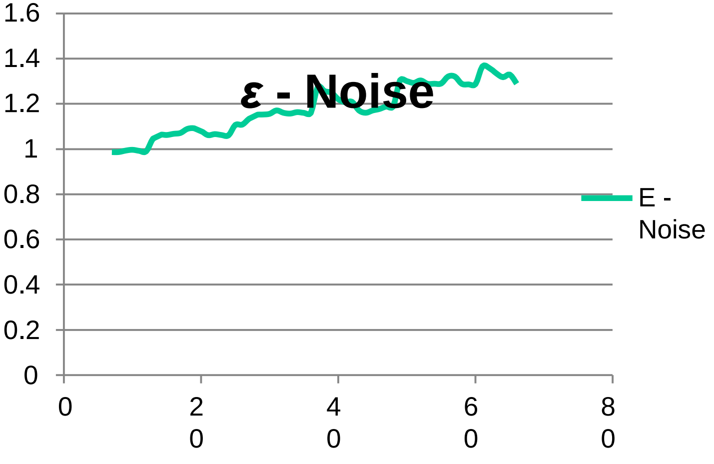
- *where M_t is a **Trend component** for **De-seasonalized by using Linear Regression***

Time Series : *Random noise*

CLASSROOM ACTVITIES

Exercise 6

- **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*

- **Time series components** $Y_t = M_t \times Q_t \times \varepsilon_t$
- **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**

$$Q_t = \mathrm{sum}(B_t)/n$$
$$B_t \approx Y_t/m_t(h)$$

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

$$C_t = Y_t / Q_t$$

- **Random**

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

Unitec NEW ZEALAND Te Whare Wānanga o Wairaka

Evaluating Time Series Models

- Yes, we have a <u>prediction model</u> by using **Trend** (M_t)
  - ☐ that utilizes **Linear regression** to <u>predict</u> **the future values** …
    That's g r e a t ! ! ☺

- But…. ***How*** to do we <u>**evaluate the accuracy**</u> of the **model**?

Evaluating - Distortion measures

- Given the observations y_t of a time series and the corresponding forecasts f_t, the **Prediction Error** is defined as

$$\boxed{e_t = y_t - f_t}$$

- and the **Percentage Prediction Error** as

$$\boxed{\boldsymbol{e_t(\%)}} \quad \boxed{e_t^P = \frac{y_t - f_t}{y_t} \times 100}$$

Evaluating - Distortion measures

Exercise 7

- **Use** the EXCEL sheet (with Exercise 1+2+3+4+5+6 completed)

- **Calculate** e_t **and** e_t^P (%) **of Trend** M_t **based on Linear Regression** for C_t during 2000-2005

$$\boxed{e_t = y_t - f_t}$$

Evaluating - Tracking Signal

Tracking Signal

- The tracking signal TS_k is an index that is useful in the **monitoring stage of forecasting process**

- The formulae is:

$$TS_k = \frac{\sum_{t=1}^{k} e_t}{\sum_{t=1}^{k} |e_t|}$$

$$\boxed{e_t = y_t - f_t}$$

Evaluating - Tracking Signal

- TS takes values in the interval **[-1,1].**
  - **If *TS* is close to -1** the prediction model is distorted by **excess**,
    - the prediction is *greater* than the actual value.
  - **If *TS* is close to 1** the prediction model is distorted by **defect**,
    - the prediction is *smaller* than the actual value.
  - **If *TS* is close to 0**
    - the prediction model has substantial **lack of distortion**.

Unitec NEW ZEALAND Te Whare Wānanga o Wairaka

Evaluating - Tracking Signal

- **In practice**, a variant of ***TS*** is used $TS_k = \frac{\sum_{t=1}^{k} e_t}{\sum_{t=1}^{k} |e_t|}$
  - for **reducing computational complexity**

$$TS_k = \left| \frac{D_k}{G_k} \right|$$

$$\boxed{e_t = y_t - f_t}$$

$$D_k = \beta e_k + (1-\beta) D_{k-1}$$

$$G_k = \beta |e_k| + (1-\beta) G_{k-1}$$

$0<\beta<1$ is a parameter

- It is in the interval of **[0,1]**

Evaluating - Tracking Signal

- **Example, using TS with $\beta = 0.2$**

$$TS_k = \left| \frac{D_k}{G_k} \right|$$

$$D_k = \beta e_k + (1-\beta) D_{k-1}$$
$$G_k = \beta |e_k| + (1-\beta) G_{k-1}$$

t	Y_t	F_t	e_t	D_k	G_k	TS_k
1	2	2.5	-0.5	(0.2*-0.5) + (1-0.2)*1	(0.2*|-0.5|) + (1-0.2)*1	|**0.7**/**0.9**| = 0.78
2	3	3.5	-0.5	(0.2*-0.5) + (0.8*0.7)	(0.2*|-0.5|) + (0.8*0.9)	|**0.46**/**0.82**| = 0.56
3	4	3.6	0.4	(0.2*0.4) + (0.8***0.46**)	(0.2*|0.4|) + (0.8***0.82**)	|**0.448**/**0.736**| = 0.61
4	5	4.2	0.8	(0.2*0.8) + (0.8***0.448**)	(0.2*|0.8|) + (0.8***0.736**)	|0.518/0.7488| = 0.69

Evaluating - Tracking Signal

Exercise 8

- **Use** the EXCEL sheet (with Exercise 1+2+3+4+5+6+7 completed)

- **Calculate TS_k with β=0.5 of Trend M_t based on Linear Regression** for C_t during 2000-2005

$$TS_k = \left| \frac{D_k}{G_k} \right|$$

$$\boxed{e_t = y_t - f_t}$$

$$D_k = \beta e_k + (1 - \beta) D_{k-1}$$

$$G_k = \beta |e_k| + (1 - \beta) G_{k-1}$$

Evaluating Time Series Models

Tracking Signal

- *TS* is **ideally close to 0**

- The actual *TS* at time *k* is calculated and it is then compared with a **threshold** (U).
  - If **TS>U**, it means that the **Trend (Prediction) model** should be revised