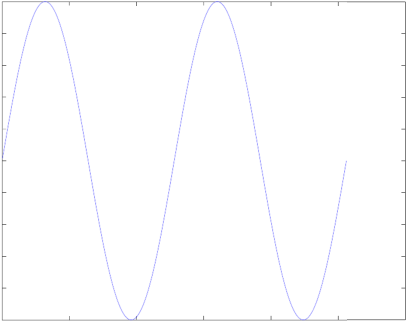
Introduction to Time Series Analysis

John Boland, University of South Australia

July 2022

What is it?

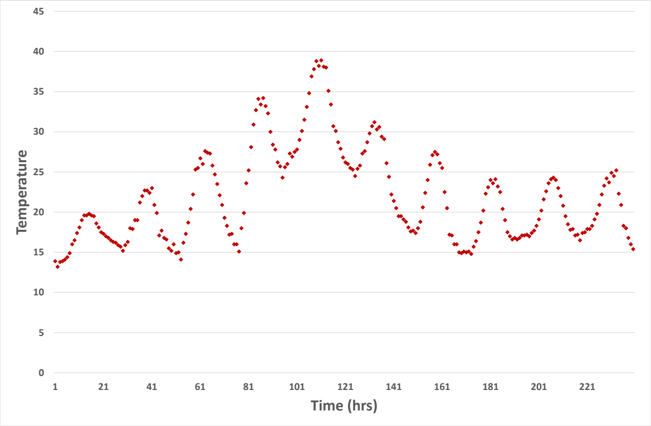
- A Time Series is a set of observations ordered in time.
- There are both *continuous* as well *discrete* time series.
- A example of a continuous time series is given in the following figure, an analogue signal.



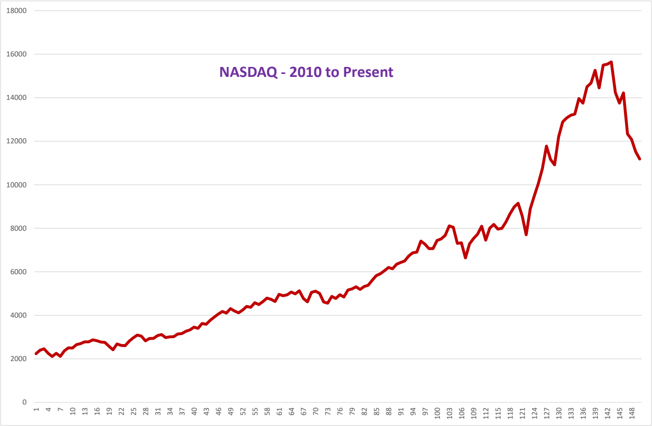
Our Focus

- In this course we will deal with only discrete time series, with observations $x_t, \; t = 1, 2, \ldots, n$.
- In fact, even if a time series were continuous one could still extract a discrete time series from it by taking readings at discrete times.
- This is called *sampling* a series.
- A common example of this is in the field of weather data time series. The ambient temperature is a continuous variable but we take readings of it only at discrete times, ie. we sample the series.

Hourly Temperature



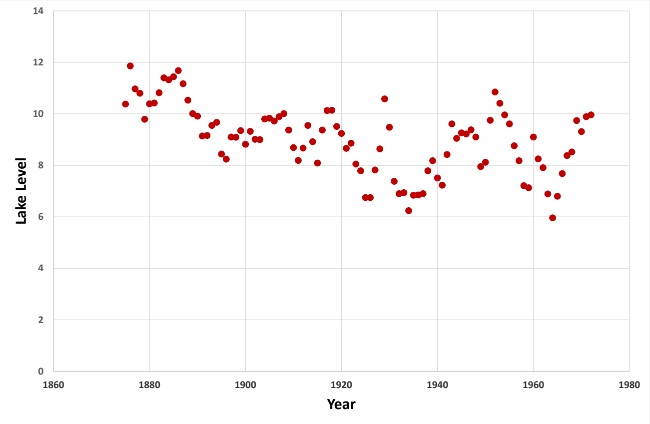
NASDAQ Index



Components of time series

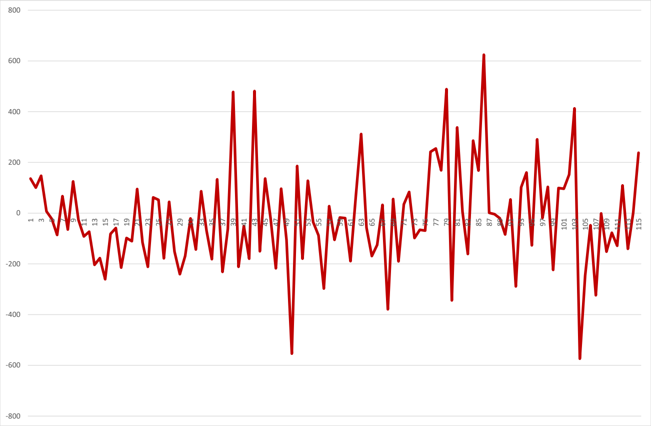
1. Trend
2. Seasonality - may be multiple seasons
3. Cycles
4. Stochastic variation

Trend



Lake Level
14
12
10
8
6
4
2
0
1860
1880
1900
1920
1940
1960
1980
Year

Stochastic variation



Motivation

- Description of the process
- Compaction - or model production
- Prediction of future behaviour
- Control

More detail

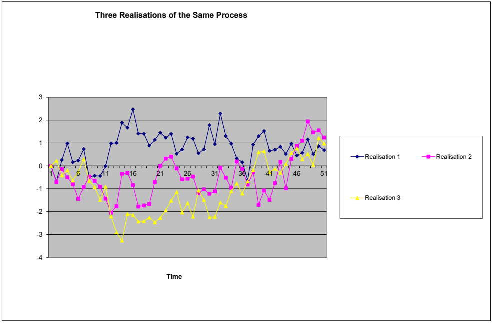
- When presented with data, plot it. That can give an indication of trend, seasonality, variability, maybe even outliers.
- Time Series Analysis will often allow us to better understand the physical mechanisms underlying some process. Examples of this include trends, seasonality but also autocorrelations and cross-correlations between variables which we will understand more about later.
- One of the reasons we analyse time series is to be able to either predict future values of the series or at least to be able to make some statistical statements about future possible values.
- When a time series is generated which measures the quality of a manufacturing process,the aim of the analysis may be to control the process. **Not our focus**.

Deterministic and Stochastic Processes

- We may want to form a model of the trajectory of a missile launched in a certain direction, with known velocity. If the exact calculations were possible the model would be entirely deterministic.
- However it is very rare that a phenomenon is totally deterministic due to any unknown factors that could be present and throw the missile off course, for example wind velocity.
- If a deterministic model cant be written, then it may be possible to derive a model that is used to calculate the probability of a future value lying between two specified limits. This model is called a stochastic model.

Series versus Process

- It should be noted here that we can make a distinction between the terms series and process.
- For instance, a stochastic series can be considered as a single realisation of a stochastic process.
- In the figure, we have three realisations or series generated by the same process.



Limitations

- It may be impossible to examine different actual realisations of a process.
- For example, one can't restart a weather series.
- It may be possible, and advantageous, to generate synthetic realisations of a process. For instance, if wanting to understand the bounds for output of a solar photovoltaic system in a particular location, one can build a model for solar irradiation, containing a stochastic component.
- Then one can generate many possible realisations of the process that the model represents, evaluate the performance of the PV system for each realisation, and then construct bounds for the performance. we will show how later.

Stationarity

- Essentially, we describe a series as stationary if there is no change in mean or variance over time, after any trend or seasonality is removed.
- Most of the probability theory of time series is concerned with stationary time series.
- One may remove the trend and seasonal variation from a set of data and then try to model the variation in the *residuals* by means of a stationary stochastic process.

Types of Stationarity

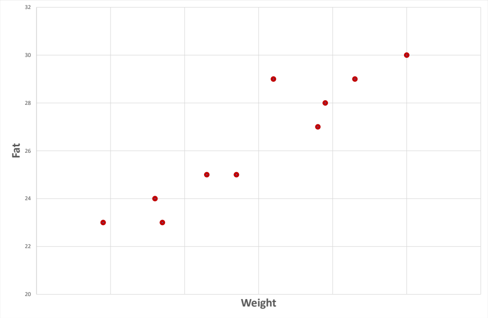
- **Weak stationarity**
  - $E[X_i] = \mu, \ \forall i$
  - $Cov[X_i, X_{i-k}] = \gamma_k$
- **Strong stationarity**
  - $(X_1, X_2, \ldots, X_m) \stackrel{d}{=} (X_{1+h}, X_{2+h}, \ldots, X_{m+h})$
- In this course, we will deal with series with weak stationarity, and even that will be relaxed at times - wait and see.

Trends

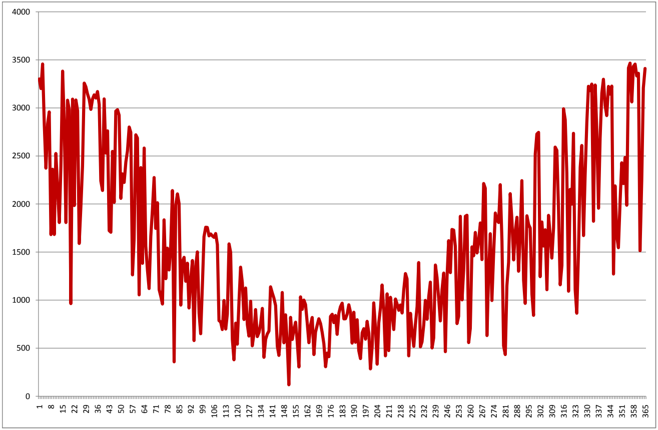
- Linear
- Polynomial
- Seasonal

Linear

An exercise physiologist used skinfold measurements to estimate the total body fat, expressed as a percentage of body weight, for 10 participants in a physical fitness program. The body fat percentages and the body weights are shown in the figure.

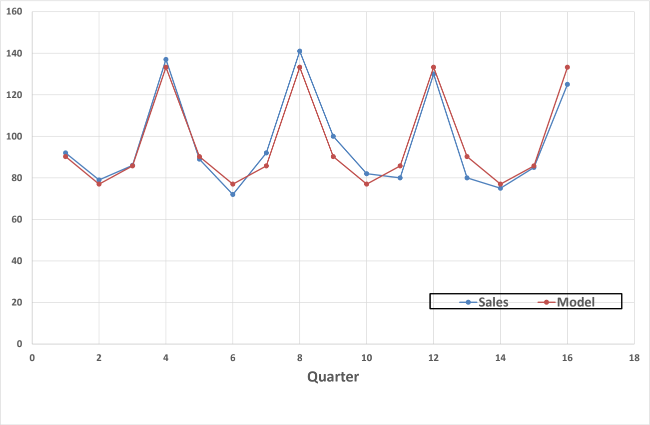


Seasonal



4000
3500
3000
2500
2000
1500
1000
500
0
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155 162 169 176 183 190 197 204 211 218 225 232 239 246 253 260 267 274 281 288 295 302 309 316 323 330 337 344 351 358 365

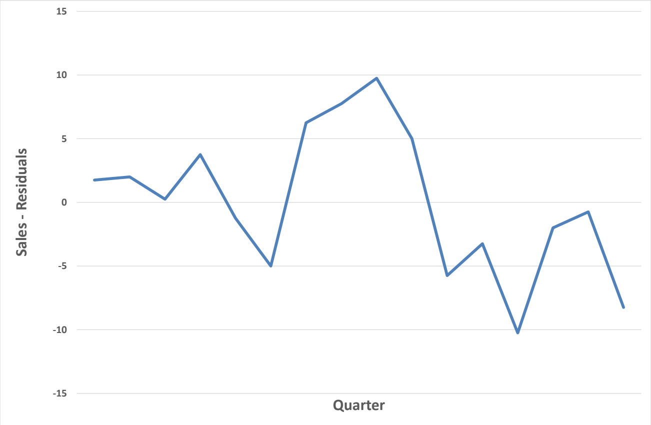
Jeans sales - quarterly data - plus model



How did I get the model?

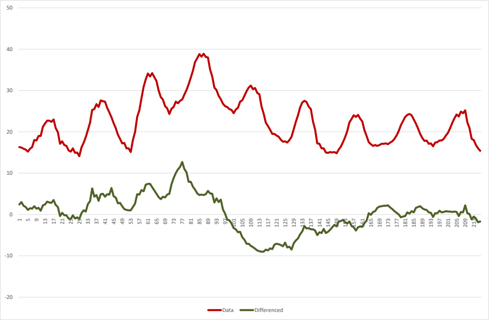
$$\hat{x}_i = x_i - \frac{1}{n}\sum_{j=0}^{n-1} x_{i+4j}$$

Residuals



Another way - Differencing - hourly temperature data

$$\nabla_d x_i = x_i - x_{i-d} = x_i - x_{i-24} = (1 - B^d)x_i$$



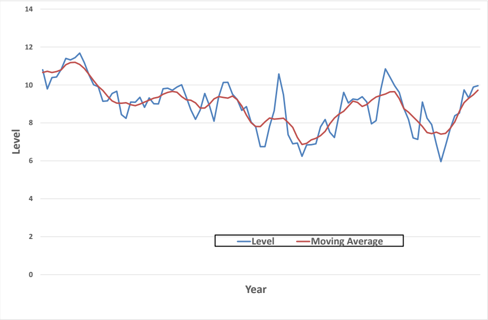
B is the backwards shift operator defined by $Bx_i = x_{i-1}$.

Smoothing

- We noted with the jeans data that by removing the seasonality, we could see more of the long term trend.
- In non-seasonal data, we can sometimes get a better picture of overall behaviour by smoothing the data.
- This is very useful when the data is highly fluctuating.

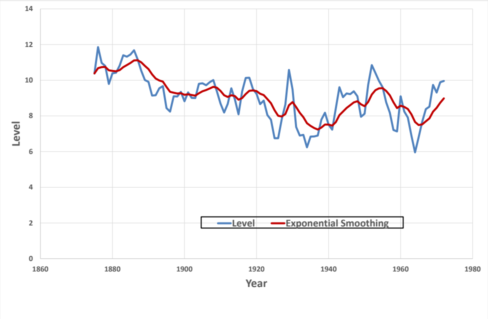
Moving Average

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t+q}, \ q \leq t \leq n-q.$$



Exponential Smoothing

$$\hat{m}_t = \alpha x_t + (1-\alpha) x_{t-1},\ t \geq 2,\ 0 < \alpha < 1.$$



Notes

- Note that if one took a higher value of q in the moving average, then there would be a greater degree of smoothing until obviously if one took q such that the moving average covered the whole series we would have one number, the mean of the series.
- Since the moving average at time t is calculated using values symmetric about time t, it is not of use for forecasting future values of the series from past values
- As we will see later, exponential smoothing can be used for forecasting.

Why is it called Exponential Smoothing?

For $t \geq 2$,

$$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j x_{t-j} + (1-\alpha)^{t-1} x_1$$

This is a weighted average of $x_t, x_{t-1}, x_{t-2}, \ldots$ with weights decreasing exponentially.

Linear Filters

- A linear filter converts one time series X_t into another by $Y_t = \sum_{j=-\infty}^{\infty} a_j X_t$.
- The a_j are a set of weights.
- A *low pass* filter removes the high frequency variations - giving a smoothed version.
- A *high pass* filter removes trends, seasonality and so on to give the high frequency fluctuations.

Next Workshop

- We will look in depth at how to deal with smoothly varying seasonality.