Data Mining
Time Series Forecasting

Jingpeng Li

What is a Time Series?

- A sequence of values or events where the next event is determined by the events that precede it
- The next step in a time series may be determined by 1 or more of the previous steps. The number of steps is known as the *order* of the time series

Usual Examples

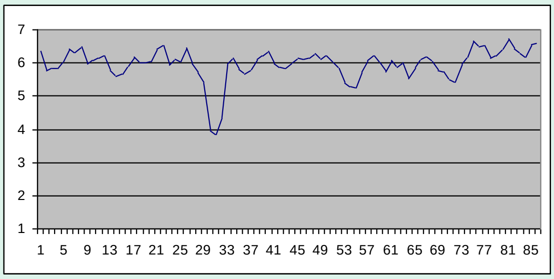
- Attempts at predicting stock price movements
- Models of machinery – often used for controlling such machinery
- Models of chemical processes
- Models of amino acid sequences in proteins

Anatomy of a Time Series

- A time series reflects the **process** being measured
- The process has certain **components** that affect its behaviour. It is important to think about the process that produces a time series when thinking about the data
- The next slides describe four different types of behaviour and how they are reflected in data
- Anything that produces data is a 'process'

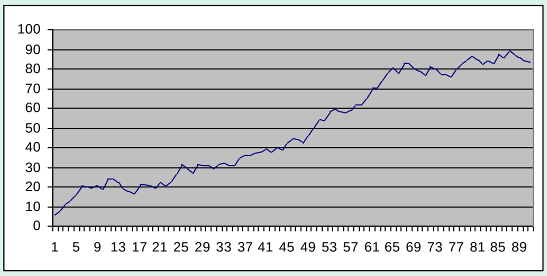
Level

- Level is simply the average value of the time series
- If the average level is the same throughout its length then the series is said to be 'stationary'
- A stationary system might get pushed off its level by a sudden shock, but it will return to this level quite quickly

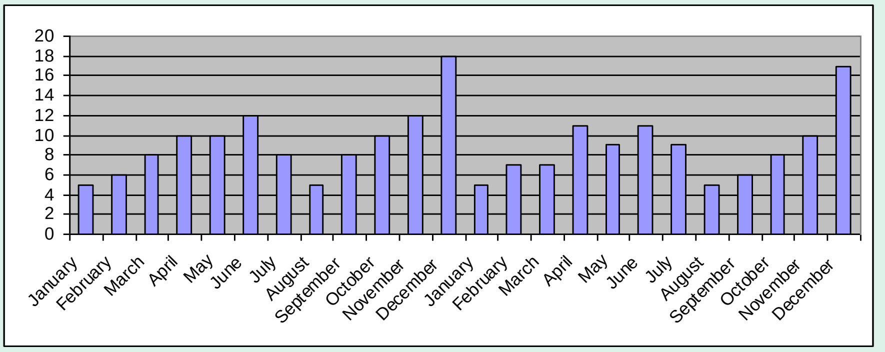


Trend

- A process that produces values that get continually larger (or smaller) over time is said to have a trend (or to be non-stationary)
- The average level for such data is of no use as the data will never be that value again
- Trend can be a function of time or previous values



Seasonality



Seasonality

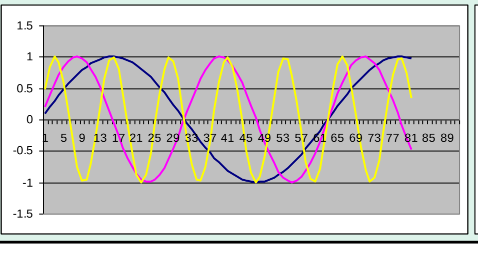
- When we think of seasons, we think of Spring, Summer, etc.
- In time series analysis, a season is any period of time that repeats through the data, e.g
  - Monday, Tuesday, Wednesday …
  - March, April, May …
  - 1pm, 2pm, 3pm …

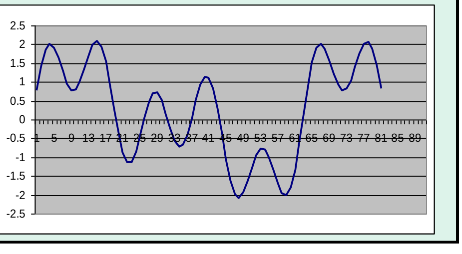
Seasonality

- All the seasons appear once in a single epoch, which depends on the scale of the season (days in a week, hours in a day …)
- Seasonality is always of a fixed and known period.
- Each season will have an impact of the data produced during that season
  - Sales may be much higher during December
  - Temperatures are higher in summer

Cycles

- Cycles may look similar to seasonality, but don't confuse them
- Cycles are the smooth undulations of a process (often a physical process)
- Cycles often add together to produce complex wave forms
  - Sound, other vibrations, images, etc



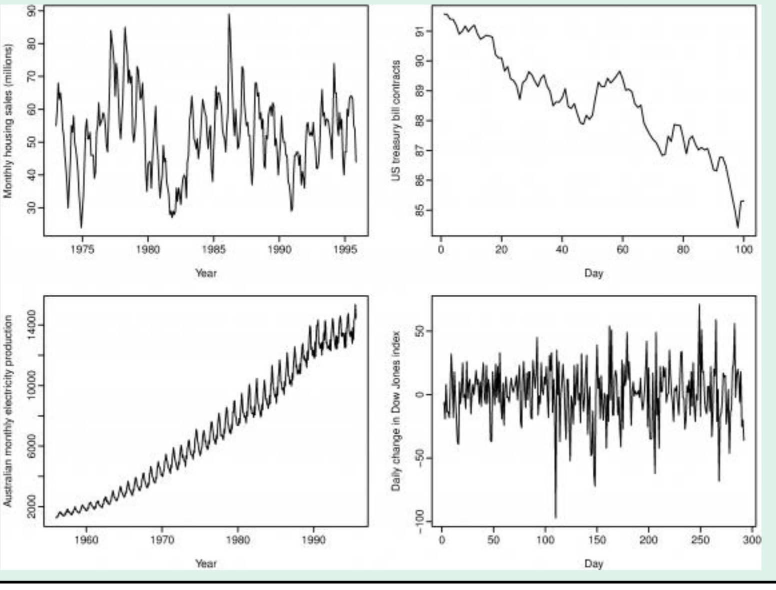


Cycles vs. Seasonality

- Cyclic pattern - the fluctuations are not of fixed period
- Seasonal pattern - the period is unchanging and associated with some aspect of the calendar
- The average length of cycles is longer than the length of a seasonal pattern
- The magnitude of cycles tends to be more variable than the magnitude of seasonal patterns.

11

4 Time Series with Different Patterns



4 Time Series with Different Patterns

- **The monthly housing sales (top left)** show strong seasonality within each year, as well as some strong cyclic behaviour with period about 6–10 years. There is no apparent trend in the data.
- **The US treasury bill contracts (top right)** show results from the Chicago market for 100 consecutive trading days in 1981. Here there is no seasonality, but an obvious downward trend.
- **The Australian monthly electricity production (bottom left)** shows a strong increasing trend, with strong seasonality. There is no evidence of any cyclic behaviour here.
- **The daily change in the Dow Jones index (bottom right)** has no trend, seasonality or cyclic behaviour. There are random fluctuations which do not appear to be very predictable, and no strong patterns that help with developing a forecasting model.

Techniques

- There are many techniques available for time series forecasting
- Different techniques are designed to use different components of a time series
- Using a technique designed to find trends on cyclic data will not work
- Data often contains more than one component and requires several techniques

© University of Stirling 2019

Techniques - Simple

- Predict that the next step will be
  - the same as the previous one
  - the average of the last few
  - a weighted average of the last few

© University of Stirling 2019

Techniques – Level

- A process that operates at a fixed level might never leave that level. Forecasting would be easy! The value is always the same
- Many processes ‘like’ to be at a certain level, but are pushed off it and then return
- You can use a **ARMA models** to predict how quickly the process moves back to its level after being pushed off it by a shock

ARMA Models

- ARMA is an acronym of **A**uto-**R**egressive **M**oving **A**verage

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

- The **AR** part models how previous values affect future ones

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

- The **MA** part models how the shock itself affects future values

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Techniques - Trend

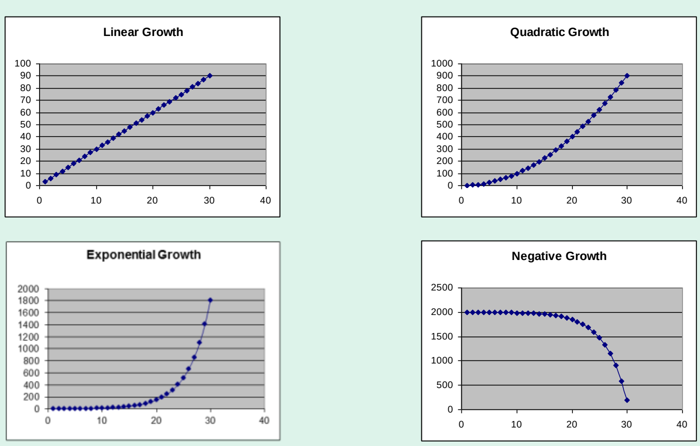
- You can extract trend from a time series with respect to the number of time steps since the series started
- Or with respect to the last value
- Either way, one technique is to use **Regression** to find the trend
- Another way to remove trend is to difference the series
- You may want to then remove the trend and see if there are other components in what is left

© University of Stirling 2019

ARIMA

- ARIMA is an acronym of Auto-Regressive Integrated Moving Average
- It is an extension of the ARMA model that incorporates trend
- Trend can be linear – growth by a constant factor or non-linear – the rate of growth changes over time too

Growth Rates



Techniques - Seasonality

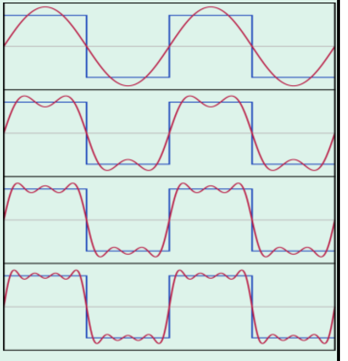
- Seasonal factors may be:
  - Additive (summer is usually 4 degrees warmer than winter)
  - Multiplicative (December sales are three times as high April sales)
- Seasonality must first be identified and then modelled

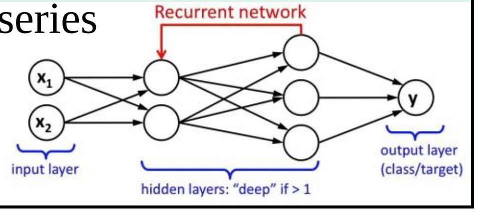
Auto-Correlation

- Auto-correlation is a method for finding the correlation between each value and the value before it
- You can also auto-correlate with the value two steps before, and then three, and so on
- Where ever you find a high correlation (say 12 steps) you should look for seasonal effects

Techniques - Cycles

- The Fourier Transform is a method for taking any signal and decomposing it into a set of sine waves
- Recurrent neural networks are good at finding cyclical components in a time series





Problems

- All techniques can ***appear*** to work even if the time series is random
- A predictable time series can look random to the eye
- Strict tests needed to establish whether or not predictions are better than guess work!
- Longer term trends are hard to capture

Time Intervals

- A system is said to be 'temporally dependent' if each step is predicted by previous ones
- Many time series need to be measured at fixed time intervals to make sense
- Many series are not measurable at fixed intervals and some don't depend on a fixed interval to be predictive

Certainty

- Unless a system is completely closed, future steps will be affected not only by previous steps but by outside forces too
- Some such forces will be measurable and can be included in a model
- Otherwise, these forces will appear as noise in the data and force you to qualify your predictions with some probability or confidence score
- Any part of the series that you cannot account for is called the 'Residual'. If the residual is not random, then you have missed something!