# Chapter 1: Introduction

## Types of data

Cross-sectional data: Measurements (either qualitative or quantitative) are taken on variables from observing a population at a given point in time (one point in time or a short period of time). A critical assumption when using cross-sectional data is that observations must be independent from one another.

Time series data: Measurements are taken on a variable at equally spaced time intervals (through time). The key point in time series data is that observations are not independent of one another.

Univariate Time series: measure a single variable through time.   
*Example: measuring total sales at Kmart through time*

Two key types of univariate time series:

1. Stationary time series: Constant mean and constant variation.
2. Non-stationary time series: Trend, Cycle and seasonality. May/may not have constant mean and variance.

Both stationary and non-stationary time series can contain autocorrelation and always have a random component.

Multivariate time series: measure several variables through time with each set of measurements taken in the same time period.   
*Example: measuring sales at Kmart within different departments through time*

Panel (Longitudinal) data: Measurements (either qualitative or quantitative) are taken on variables from observing a population at equally spaced time intervals (taken through time).   
Advantages of panel data: greater number of data points (increasing our degrees of freedom) and decreases any multi-collinearity problems among explanatory variables.   
*Example: measuring sales at Kmart within different departments and at different locations through time*

Time series vs Cross-sectional data? Time series we have a correlation pattern running through the data but in cross-sectional we have variables that are correlated with one another.

Multivariate series vs Panel data? Panel data is multivariate time series taken on different cross-sectional units.

Time is a dimension, ordered variable that can be recorded as integers from 1 to T.

## Aspects of time series

In order to model a time series that can be used for accurate predictions (assuming model conditions are met) 50 degrees of freedom are required.

* Every piece of independent information adds a degree of freedom
* Every parameter we estimate loses a degree of freedom

*Example: When we add a lagged response variable in our model to account for autocorrelation: we lose two degrees of freedom. We lose one degree as we lost an observation in our response variable and another for estimating the autocorrelation coefficient.*

Note the further we forecast into the future, the less accurate our forecasts will be. Prediction intervals will become wider the further into the future we forecast.

A good way to compare different time series models built:

* Remove observations from the end of the series (remove most recent observations)
* Fit the models using the remaining observations
* Then forecast the removed time period
* Compare forecasts to actual values and calculate RMSEP to numerically compare models
* Pick the best model and re-run on entire dataset then forecast future.

Independence problem in time series: **The future cannot influence the past, but the past can and often does influence the future.** This principle is called autocorrelation/serial correlation.   
The dependence on the past provides a pattern that allows us to build suitable models of time series.   
In stationary time series it is key as it the only pattern we can model in the series.

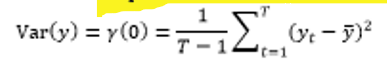
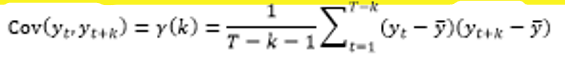
# Chapter 2: Preliminaries

## Autocorrelation

Autocorrelation is the ‘dependence on the past’. It is also called serial correlation or covariation.   
It is the key point; **The future cannot influence the past, but the past can, and often does, influence the future.’**

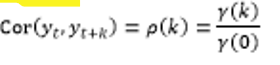
When dependence on the past occurs, we have a correlation structure within the series. We can use this correlation structure to explain some of the patterns in the series.

*Key metrics in time series:*

1. Variance of the observations
2. Autocovariance between observations k time periods apart – lag(k)

To calculate autocorrelation at lag(k): Calculate the average of the products of all possible deviations of observations from the mean that are k-time periods apart.

1. Autocorrelation = standardised autocovariance



Note: If we disregard autocorrelation when building our models:

* Overestimated R2 so we think we have a better fitting model than we actually do
* Estimated standard deviation of the residuals will be lower than actual error standard deviation
* Variance of the slope coefficient in the model will be underestimated
* T and F statistics will not be valid

**Detect autocorrelation: Durbin-Watson** test (hypothesis test where the H0: is no autocorrelation)

**Autocorrelation Function (ACF)** – a plot of the acf shows the estimated autocorrelation for each lag, independently of all other lags. This means if lags depend on one another this effect is ignored.

* Clustering of residuals -> positive autocorrelation
* Oscillation of residuals -> negative autocorrelation

**Partial Autocorrelation Function (PACF)** – a plot of the pacf shows the estimated autocorrelation for each lag accounting for any effects of dependent lags.

*Residual Series and White noise*

When we build a regression model, we are going to have to check for autocorrelation in the Residual series to check the assumption of independence in the underlying errors.

Plot **time ordered residuals (Residual series) to assess independence**. If they are independent the series is known as a White Noise series. If the series is not independent, we need to extend our model and try to account for this dependence on the past.

**White noise**: zero mean and constant variance series with no autocorrelation. To satisfy the underlying model assumptions models require that residuals be a white noise series.



If we model all other patterns in a series and the residual series shows a pattern, then we have auto-correlation pattern in the series that has not been captured by the model.   
Note also when we build a model majority of the autocorrelation structure observed is removed by modelling the other components present.

Autocorrelation is important in Time series modelling?

* The past can influence the future so observations may not be independent of each other.
* If we do not model it our model assumption of independence is violated – this comes with a whole bunch of other problems

## Transformations – Variance stabilizing transforms

* Log transformation is to turn a multiplicative model (increasing seasonal variance through time) into an additive model (constant seasonal variance through time)
* Additive models are much easier to estimate than multiplicative models
* Other type of transform we could use is square-root – good for count data.

## Unusual observations

* We cannot delete an unusual observation from the series, as our observations have a specific order and the other matters.
* Note only complete the two methods below once you’ve tried to model all other components in the series.

Two approaches to dealing with unusual time series:

1. Dummying out an observation: Create a dummy variable that is 1 for the unusual observation and 0 otherwise. Essentially, we calculate an estimate for that observation so that it does not have undue influence in the final model.
   1. The coefficient of the dummy variable will be the value of the residual if we did not have the dummy variable.
   2. The value of the residual for the influential observation will be almost 0

When can we not dummy out a variable: We cannot dummy out a variable when we have autocorrelation. This is because the lagged response variable will still contain that unusual value.

1. Interpolation: Replace the unusual observation with something more reasonable.
   1. This is the method to replace unusual observations if the series has autocorrelation.

## Smoothing and Filtering Time Series

General idea: Suppress/remove certain features in a timer series in-order to build a model.

*Moving average:* Take a window of observations, average them and replace the original observations with the moving averages. The more observations we average over, the smoother our series.   
Common use for moving average is to remove the seasonal component from a time series.

**Drawback of using moving averages** is that she moving average series becomes shorter than the original time series.

Normal moving average only works for additive time series.

*Seasonal Trend Lowess (STL):*

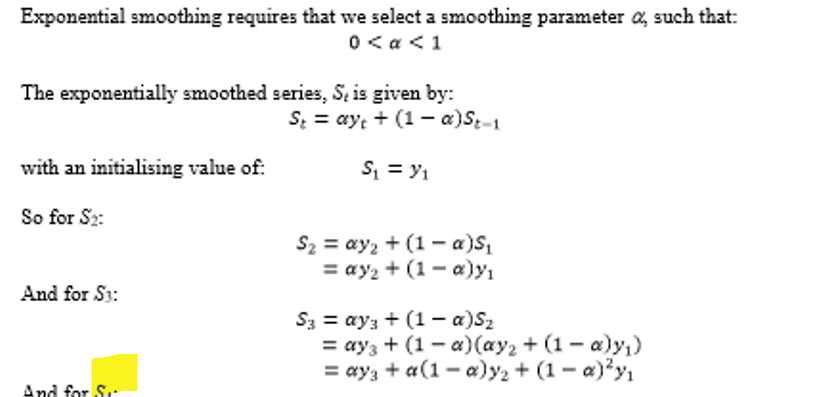
First smooth trend and cycle using a lowess smoother. The lowess smoother fits a local regression to a window of points and then uses the regression value in the middle of the window as the smoothed value. The regression is weighted such that more observations in the middle of the window are given higher weights than observations at the edges.

Secondly a second lowess smoother is used on each seasonal-sub series to find the seasonal estimates.

*Exponential smoothing:* uses a weighted average of all past values of the series to calculate a smoothed value. The weighted average weighs more recent observations higher than earlier ones.

Exponential smoothing is good as we do not lose any observations in the series and it places higher weight on the more recent observations.

Normal exponential smoothing – only use for stationary time series. Use Holt Winters for non-stationary series.



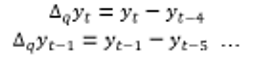
Note the smaller the value of the smoothing parameter (alpha) the smoother the resulting series.

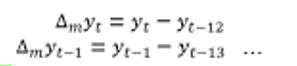
*Differencing:* remove trend/seasonality or cycles (RARELY) from a series.

* Remove trend: remove trend from a non-stationary time series turning it into a stationary time series of differences.
  + Differencing successive observations (first order differences) will remove any linear trend from the series.
  + Use the differenced values instead of the original values



* Remove seasonality
  + Differences between seasonal observations to remove seasonal component
  + *Example: quarterly data*



* + *Example: monthly data*

# Chapter 3 – Non-Stationary Time series

## Components of Non-Stationary Time Series

1. Trend
   1. Long-term slow smoothly changing component
   2. Increasing/decreasing over time and may be linear/non-linear
   3. Model trend by using Time (integer value ordered variable 1…T)
2. Cycle
   1. Recurrent irregular wave-like pattern
   2. Period and amplitude of cycles is neither fixed nor predictable
3. Seasonal
   1. Regular repeating pattern through the series
   2. Depends on calendar or clock
4. Random
   1. Most rapidly changing component
   2. Always present and present in both stationary and non-stationary
   3. Small variations about the trend, cycle and seasonal components
   4. Unpredictable but modelled as random observations from N(0, variance) distribution

## Trend Models

### Linear model



If we detect autocorrelation in the residual series – add a lagged response variable as an additional explanatory to fix:



### Quadratic model

If we detect non-linearity in our residual plots, then apply a quadratic model:



Easy to fit and forecast. Hard to describe behaviour of response variable over time in a meaningful way.

### Break in Trend/Change in slope model

Trends may have a break or change in slope at some point in time.   
Residual series has a ‘peak in it’ indicating change in slope.   
To model this, add an additional explanatory variable in the model. This dummy variable will be 0 before the break and 1 after the break.

### Growing/Decaying Trends

Model non-linear growth/decay trends.







If B1 > 1 then trend is growth.

If 0 <= B1 <= 1 the trend is decay.

Once we have figured out the trend of the series we need to determine if there is a seasonal component and how we are going to model it. Two methods:

1. Fit Holt-Winters (modified version of exponential smoothing)
2. De-seasonalize the series using smoothing techniques (Moving Average or STL) and then build seasonally adjusted models.

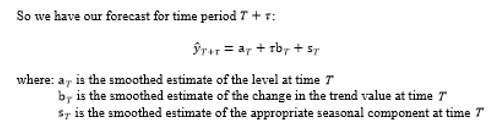
## Holt Winters

Holt Winters uses a modified version of exponential smoothing. It can be used for forecasting both stationary and non-stationary time series.

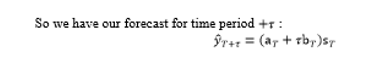
* Stationary time series: simply use exponential smoothing
  + To forecast exponentially smoothed stationary time series, we use the last value of the smooth for all future forecasts.
  + Essentially calculates the mean of the series and uses that as the forecast
* Non-Stationary time series
  + Advantage: no modelling assumptions to satisfy; do not need to worry about normality or autocorrelation
  + Drawback: Trend in the data is must be linear or reasonably linear
  + Two types of Holt Winters model
    - Additive Model: seasonal component is reasonably constant through time
    - Multiplicative Model: seasonal component is not constant through time

### Additive Holt Winters Model (3 smooth methods)

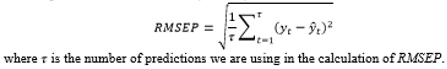
1. Level is smooth to give local average value for the series.
2. Trend is smoothed so the increase from period to period can be estimated.
3. Each seasonal sub-series is smoothed separately to give seasonal estimate for each season.
4. How to forecast (additive model)



### Holt Winters Multiplicative Model



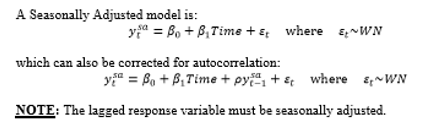
## RMSEP – Assess which model is better



RMSEP: root mean squared error of prediction

## Seasonally adjusted models

Estimate and remove the seasonal component from the series and then build a model using the de-seasonalized data. Then forecast and add back the seasonal estimates.



Huge note: we cannot dummy out observations in seasonally adjusted models.

Use Moving Average (MA) and Season-Trend-Lowess (STL) to extract seasonal estimates from models.