UNIVERSITY OF ABERDEEN

CELEBRATING 525 YEARS 1495 – 2020

ABERDEEN 2040

Time Series – 2

Data Mining & Visualisation

Lecture 26

2025

Today...

- ARIMA
- Variations of ARIMA

Auto Regressive Integrated Moving Average

ARIMA

ARIMA is a statistical method for analysing and forecasting time series data.

It is one of the more sophisticated, powerful, and popular approaches for doing so.

ARIMA

ARIMA is made up of three components:

- **Auto Regressive** component
- **Integrated** component
- **Moving Average** component

Auto Regressive Component

The **Auto Regressive** (AR) component determines how past values affect future values.

It does this by using a linear regression model to predict a given observation (point in time), using some number of previously observed values (lags).

We can define the number of *'lags'* that the model should consider through the p parameter.

Auto Regressive Component

The formula for AR is very similar to that of linear regression:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

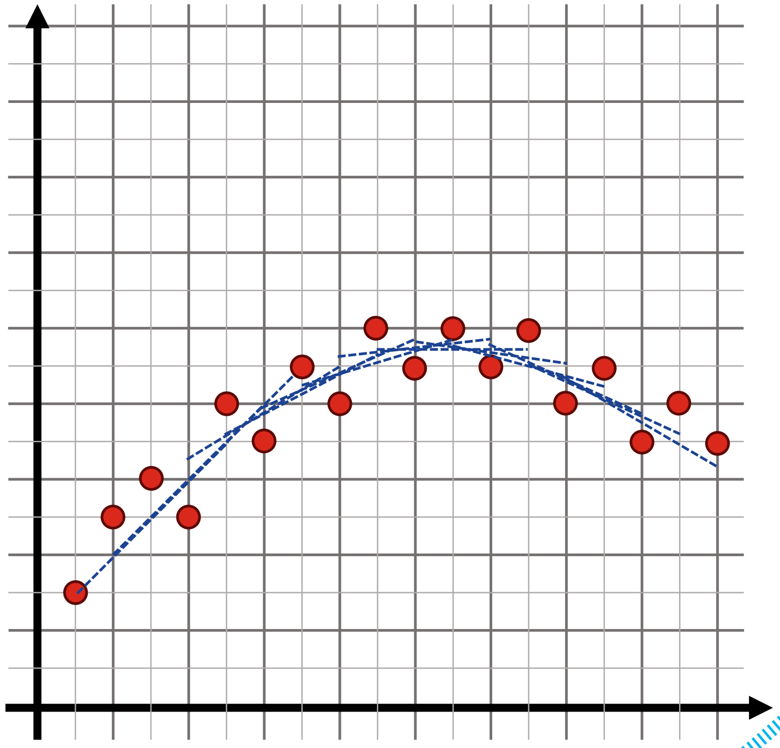
where:

- y_t is the observation at time t (that we are looking to predict)
- C is the AR model's intercept
- ϕ_i is the coefficient (weight) for observation i
- y_{t-i} is the observation at time t - i
- ϵ_t is our model's error

Note: This formula is not something you need to memorise.

Auto Regressive Component

We can think of this quite simply as a sliding window of linear models, with a fixed window size (i.e. the lag value).



Integrated Component

The **Integrated** component looks to make the time series data more stationary, through removing trends and seasonality.

It does this through ‘differencing’ the time series, by subtracting previous observations from the current one.

The d parameter determines the degree (order) of differencing that is involved.

Integrated Component

Differencing transforms our data, making it more stationary. The formula for first-order differencing (i.e. d = 1) is:

$$y'_t = y_t - y_{t-1}$$

where:

y'_t is the differenced observation at time t

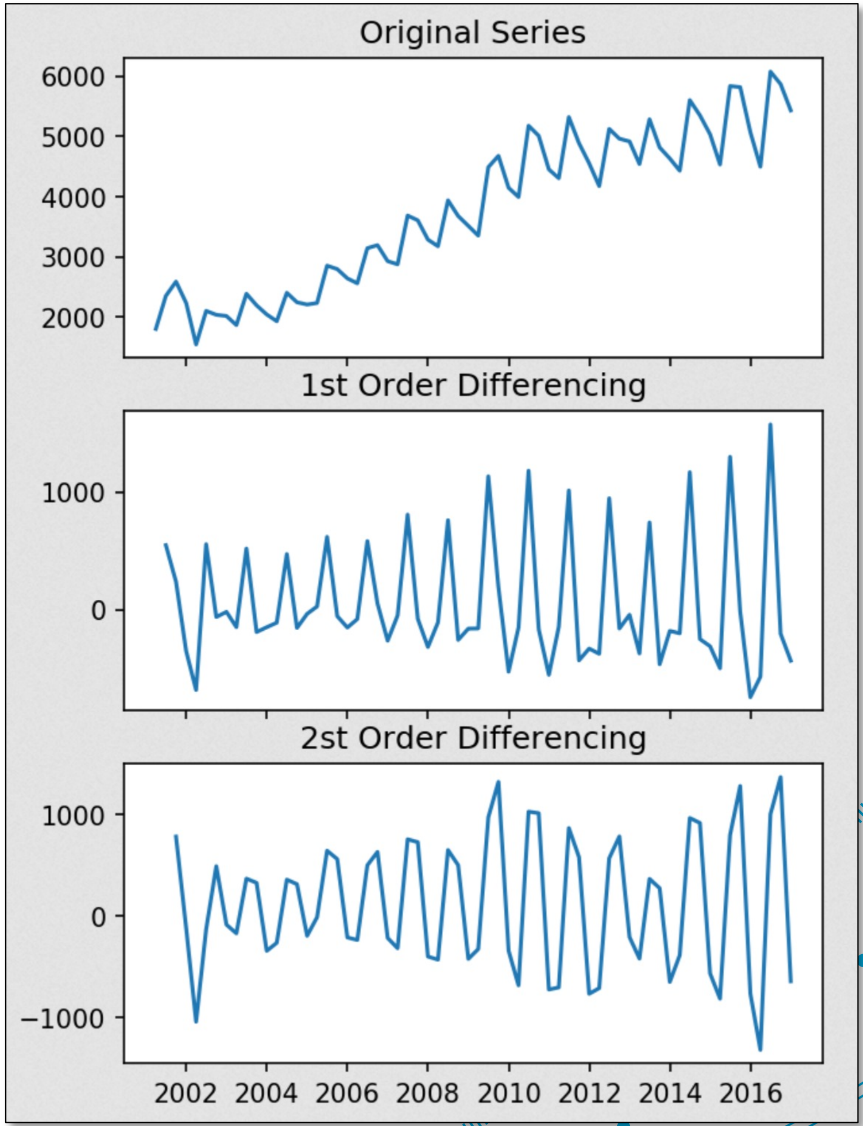
y_t is observation at time t

y_{t-i} is the observation at time - i

Note: This formula is not something you need to memorise.

Integrated Component

By applying differencing, we can remove the trends within our time series data.



Moving Average Component

The **Moving Average** (MA) component attempts to smooth out the time series by removing noise and errors.

It does this by predicting a given observation *using past forecast errors*, rather than the actual values (like AR does).

The q parameter determines the degree (order) of the moving average.

Moving Average Component

The formula for the MA component is:

$$y_t = C + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

where:

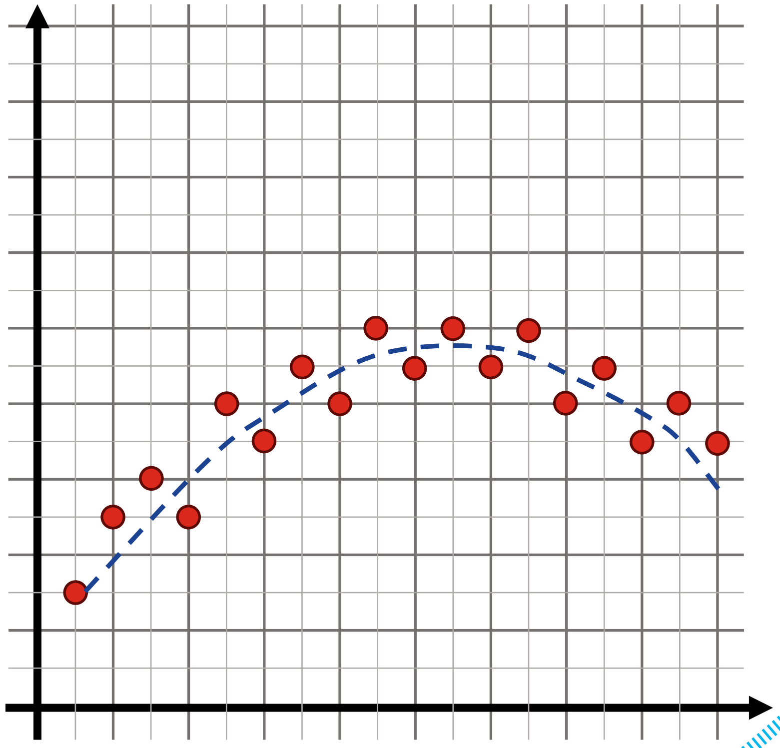
- y_t is the observation at time t (that we are looking to predict)
- C is the MA model's intercept
- ϵ_t is the current error
- θ_i is the coefficient (weight) for error i
- ϵ_{t-i} is the error at time t - i

Note: This formula is not something you need to memorise.

Moving Average Component

Again, our MA component uses a sliding window of linear models, this time focusing on a lag of the residual errors to predict the current observation.

In doing so, it smoothes the time series.



ARIMA – Bringing the Components Together

ARIMA combines these components into one method, with each addressing a different aspect of time series behaviour.

AR allows the model to capture long-term dependencies.

I stabilises the time series and handles non-stationarity.

MA accounts for natural variations and noise.

ARIMA – Bringing the Components Together

The formula for ARIMA is simply a combination of the components:

$$y'_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Shortened version:

$$y'_t = C + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

Note: These formulas are not something you need to memorise.

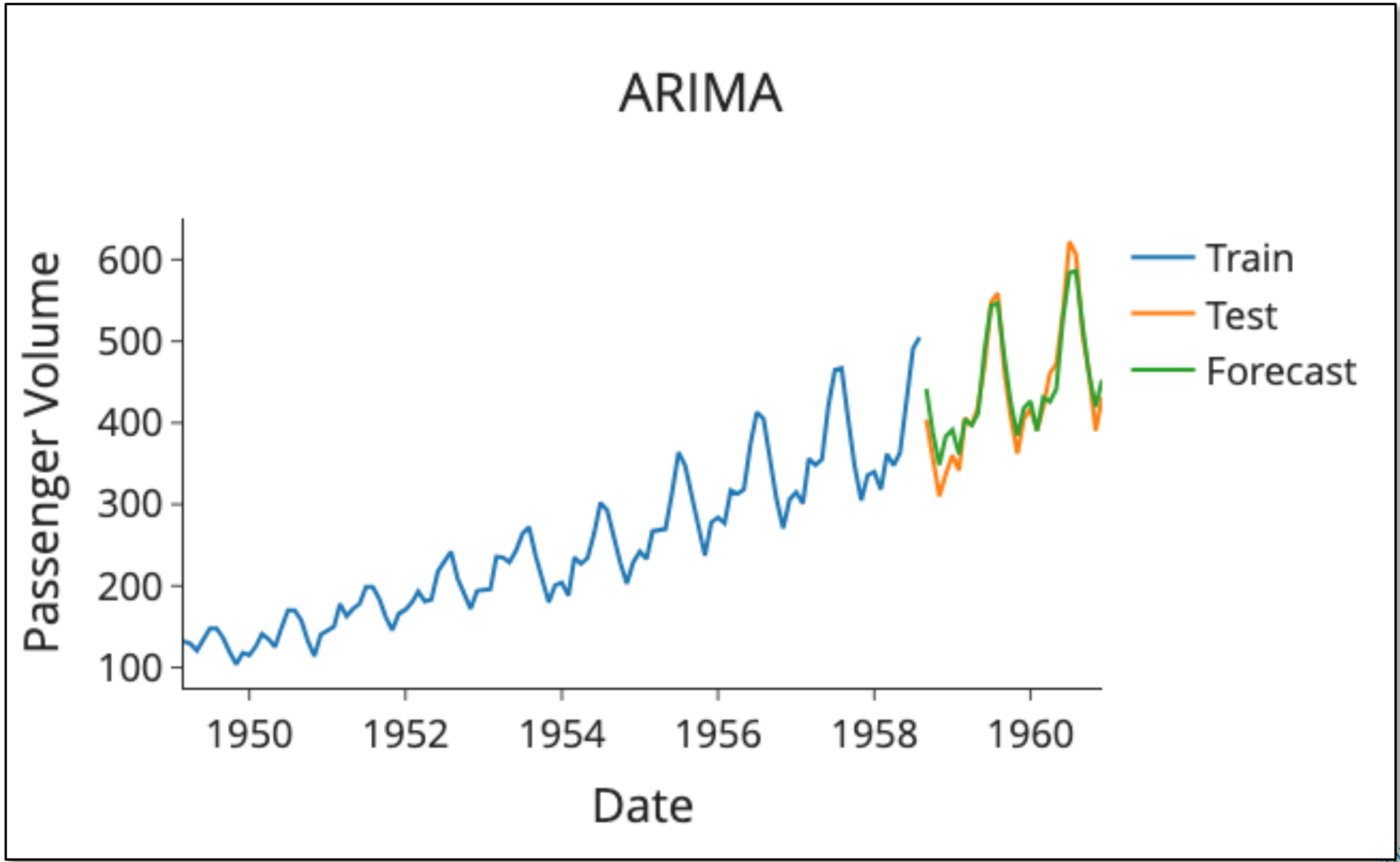
ARIMA – Bringing the Components Together

Given that ARIMA is simply a combination of these three components, it still relies on the following parameters:

p to determine the number of *'lags'* to be considered.
d to determine the order of differencing that is involved.
q to determine the order of the moving average.

It is for this reason that it is sometimes referred to as the *ARIMA(p, d, q) model*.

ARIMA – Forecasting



Source: https://tinyurl.com/arima-example

ARIMA – Advantages

ARIMA is a technique with an incredible level of **versatility**, in that it can be applied across diverse time series scenarios.

It is also a relatively **simple** approach to time series analysis and forecasting, which provides a degree of **interpretability**.

Due to its Integrated component, it can also handle (a degree of) non-stationarity reasonably well.

ARIMA – Disadvantages

However, due to its nature, it assumes a degree of **linearity** within the time series, which isn’t often the case in reality.

As a result, it is relatively **inept at handling seasonality** within time series data.

It also **fails to take into consideration external factors**, as it assumes that all variations in the data can be explained by past observations.

Variations of ARIMA

Variations of ARIMA

Over the years, a few variations to ARIMA have emerged.

These typically address one or more of ARIMA's shortcomings.

A few of these include…

SARIMA

Seasonal ARIMA (**SARIMA**) is one such variation, which explicitly accounts for seasonality within time series data.

It does this by extending the parameters, p, d, and q, to also include seasonal components: P, D, and Q, as well as s to indicate the length of seasonal cycles.

ARIMAX & SARIMAX

ARIMA *with Exogenous Variables* (**ARIMAX**) attempts to take into account the influence of external factors into the analysis.

This might include things like weather, economic trends, or other factors that may affect a time series without being present within past observations.

There is also **SARIMAX**, which combines SARIMA and ARIMAX, accounting for both seasonality and external factors.

VARIMA

Vector ARIMA (**VARIMA**) introduces the ability to deal with multivariate time series data.

As such, we can use it to understand and factor in the relationship between multiple time series.