Google Cloud

Time Series Modeling

Introduction

What is a Time Series

Time Series Concepts

Time Series Modeling

Lab Objectives

What is a Time Series

What are some of the basic concepts in Time Series

How do we analyze Time Series data to predict future values from past values

Agenda

Concepts in Time Series

Time series terminology

AR - Auto regressive

MA- Moving Average

Putting it all together

ARIMA Model

Concepts in Time Series

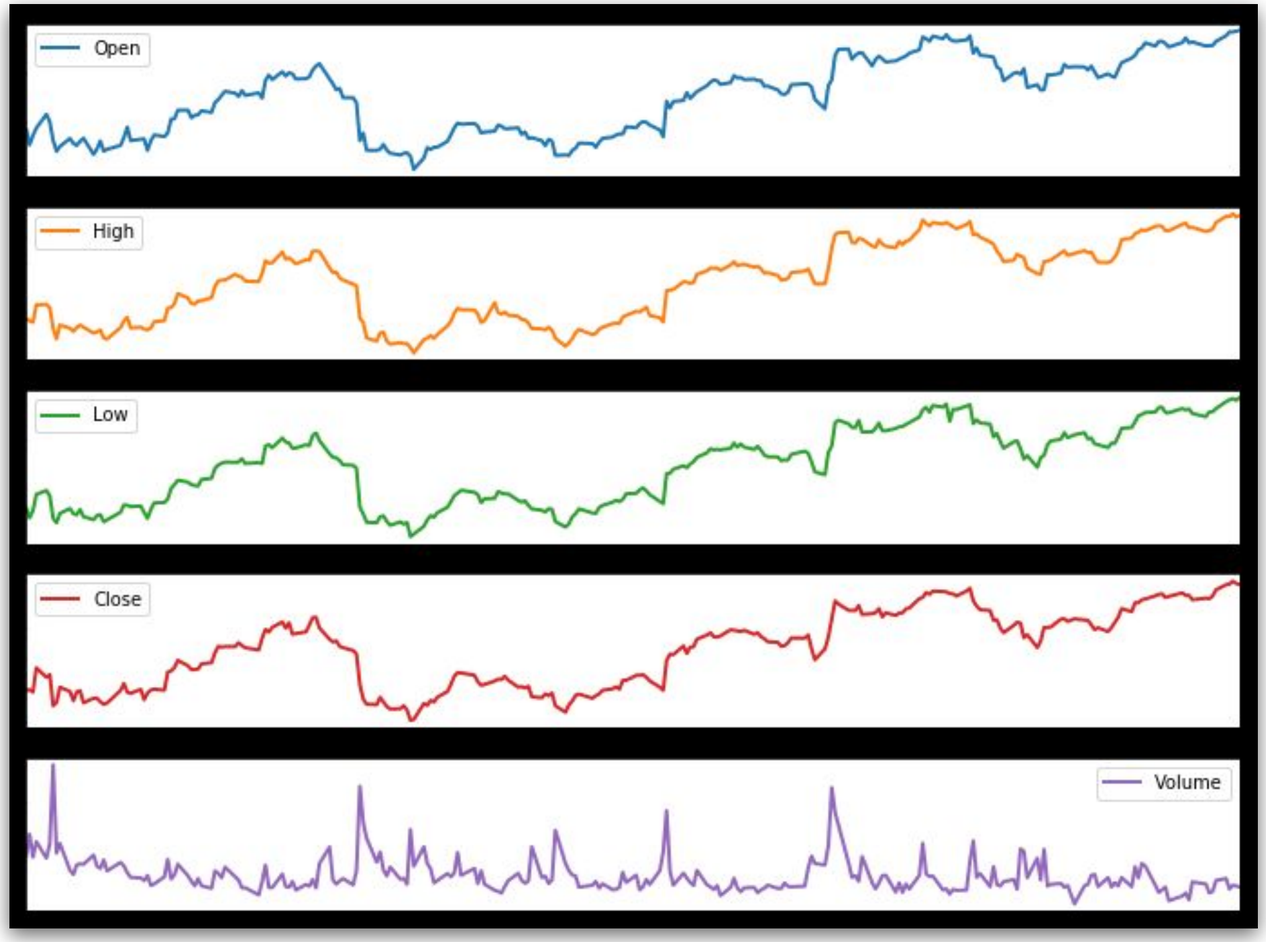
What is a Time Series?

A time series is a series of data points indexed in time order.

Most commonly, a time series is

a sequence of snapshots of a process taken at successive equally spaced points in time.

Thus it is a sequence of discrete-time data.



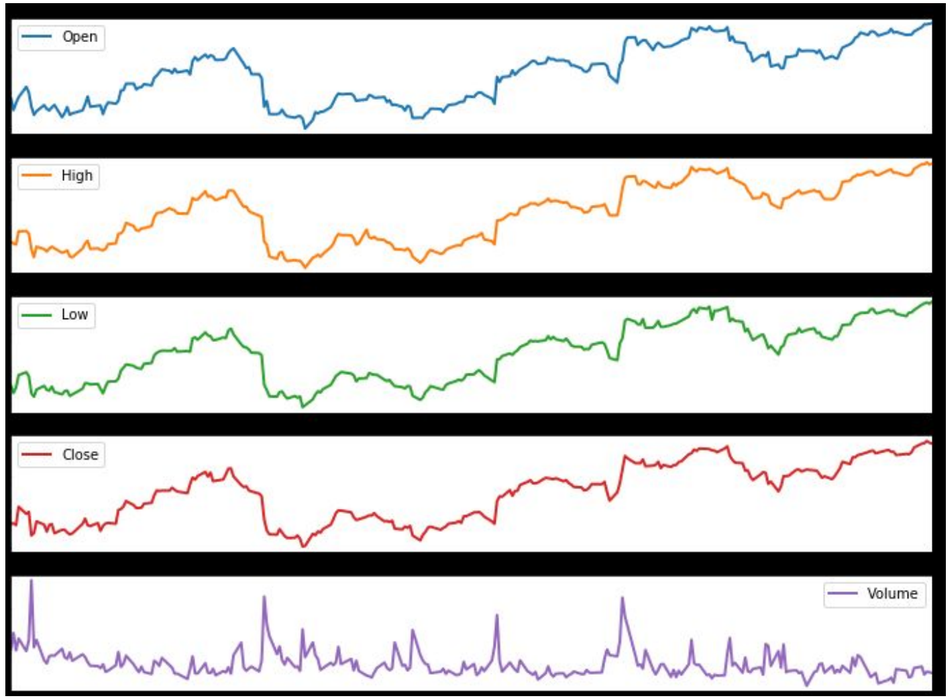
Time Series

What is a Time Series?

A time series is a series of data points indexed in time order.

Most commonly, a time series is a sequence of snapshots of a process taken at successive equally spaced points in time.

Thus it is a sequence of discrete-time data.



Agenda

Concepts in Time Series

Time series terminology

AR - Auto regressive

MA- Moving Average

Putting it all together

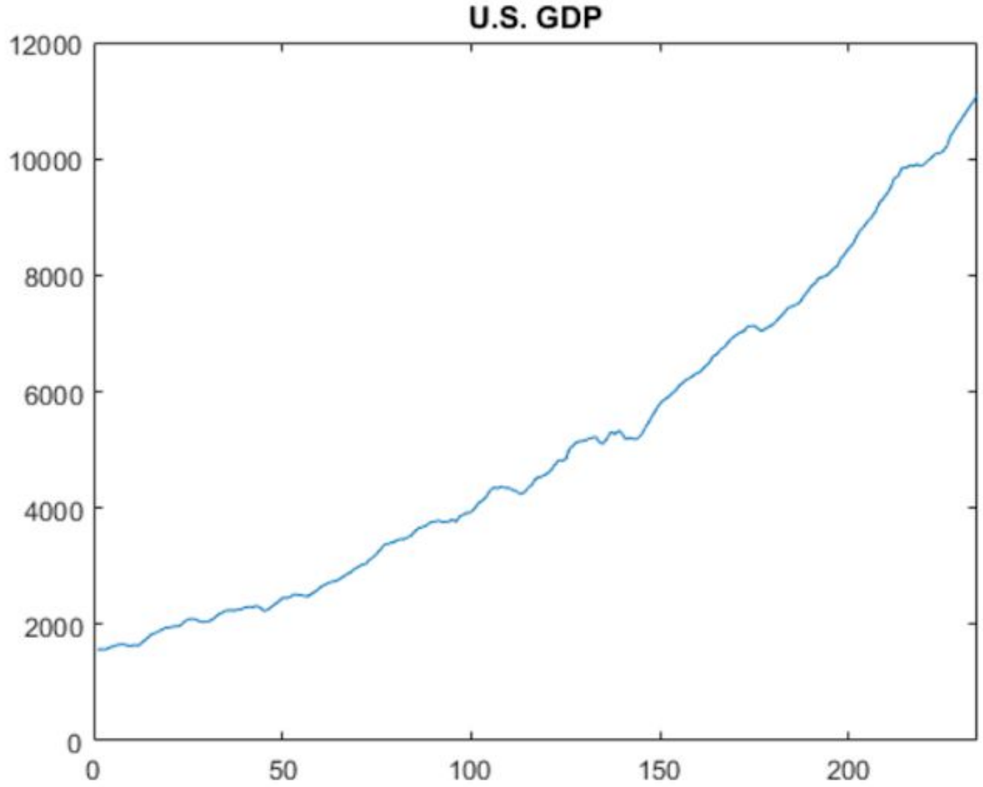
ARIMA Model

Concepts in Time Series

What is Stationary data?

“Stationary” means that the statistical structure of the series is independent of time.

```
figure
plot(Y)
xlim([0,N])
title('U.S. GDP')
```



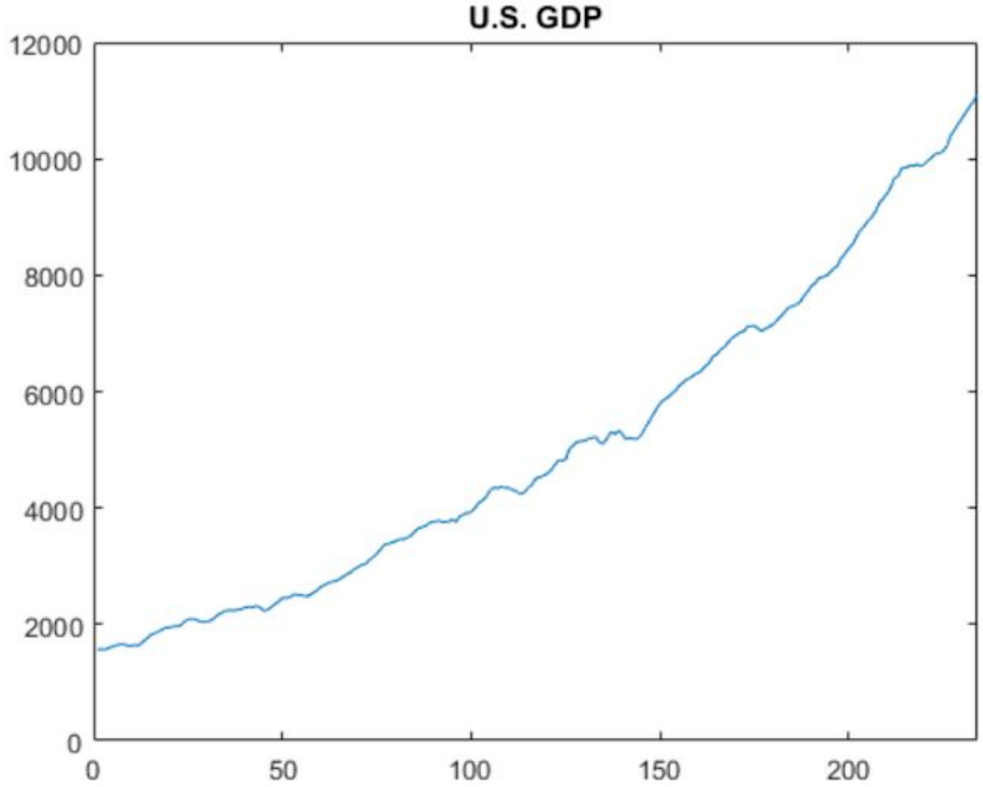
Source: https://www.mathworks.com/help/econ/nonseasonal-differencing.html

Concepts in Time Series

How do we know data is Stationary?

- Plots
- Summary Statistics
- Statistical Tests

```
figure
plot(Y)
xlim([0,N])
title('U.S. GDP')
```



Source: https://www.mathworks.com/help/econ/nonseasonal-differencing.html

Concepts in Time Series

How to convert a Trendy data to “Stationary” data?

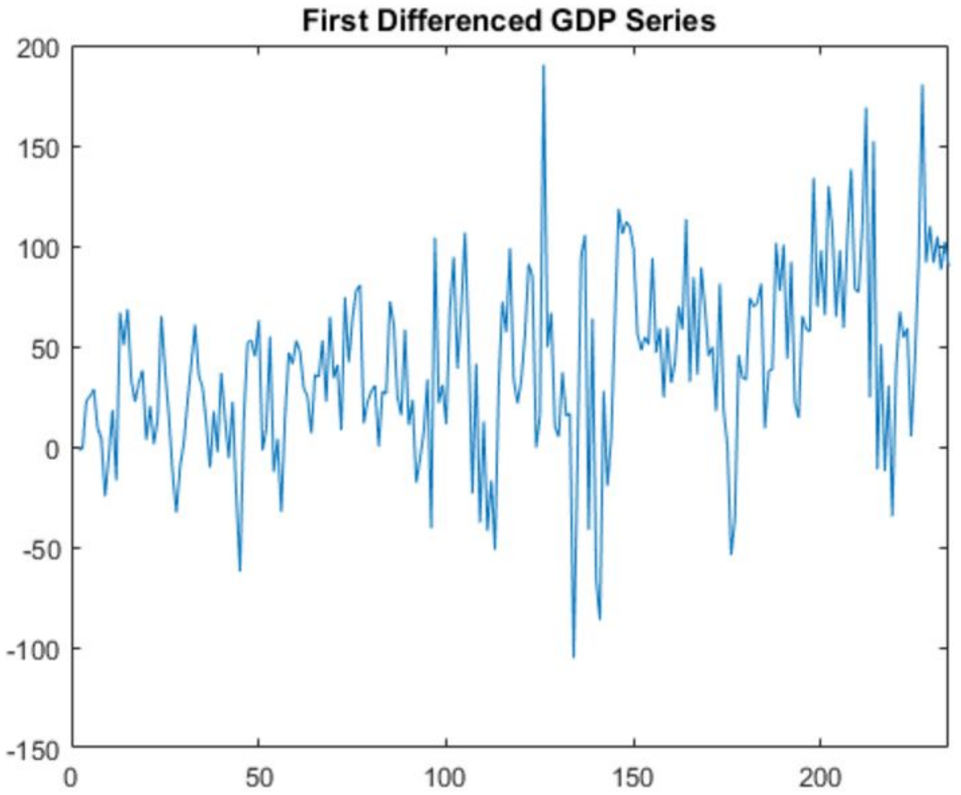
One way to Convert Trendy Data to Stationary Data is to Difference it or De-Mean it!

The time series has a clear upward trend.

Take a first difference of the series to remove the trend,

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}.$$

First create a differencing lag operator polynomial object, and then use it to filter the observed series.



Source: https://www.mathworks.com/help/econ/nonseasonal-differencing.html

Concepts in Time Series

How to convert a Trendy data to “Stationary” data?

After differencing it twice, this data now appears to be “stationary”.

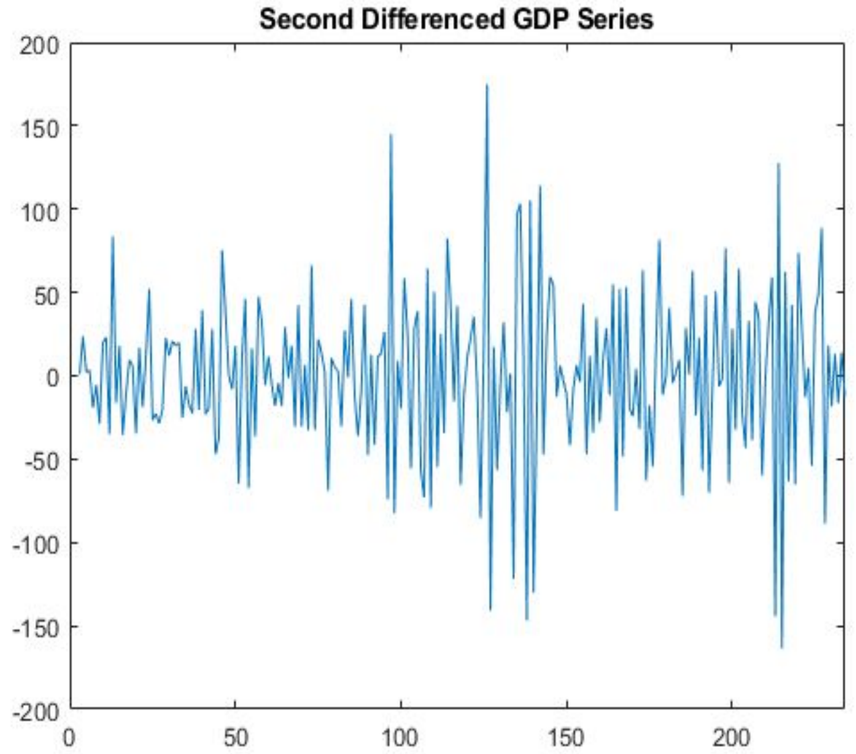
You can double check it with an Augmented Dickey-Fuller test*

The time series has a clear upward trend.

Take a first difference of the series to remove the trend,

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}.$$

First create a differencing lag operator polynomial object, and then use it to filter the observed series.



Source: https://www.mathworks.com/help/econ/nonseasonal-differencing.html

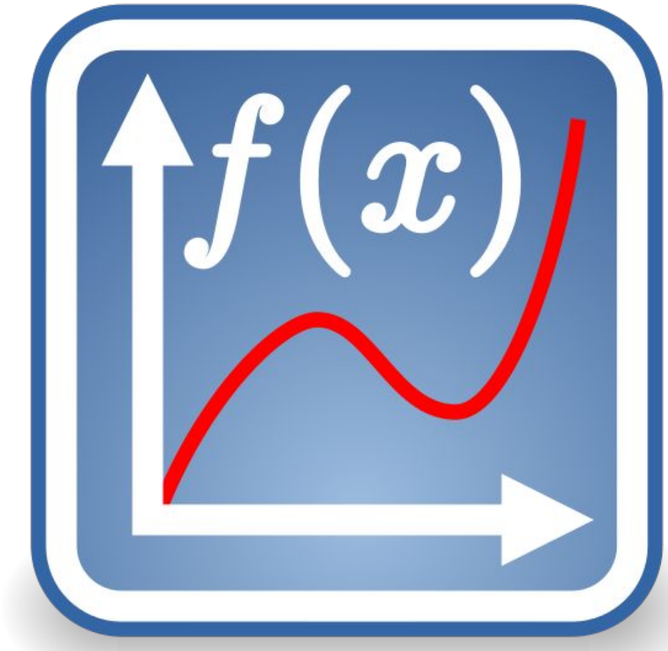
Link;
http://stats.stackexchange.com/questions/44647/which-dickey-fuller-test-should-i-apply-to-a-time-series-with-an-underlying-mode
https://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test

Concepts in Time Series

Why is Stationarity important in Time Series Modeling?

Stationarity allows preserving model stability, i.e. a model whose parameters and structure are stable over time.

Stationarity matters because it provides a framework in which averaging (used in AR and MA processes) can be properly used to describe the time series behavior.

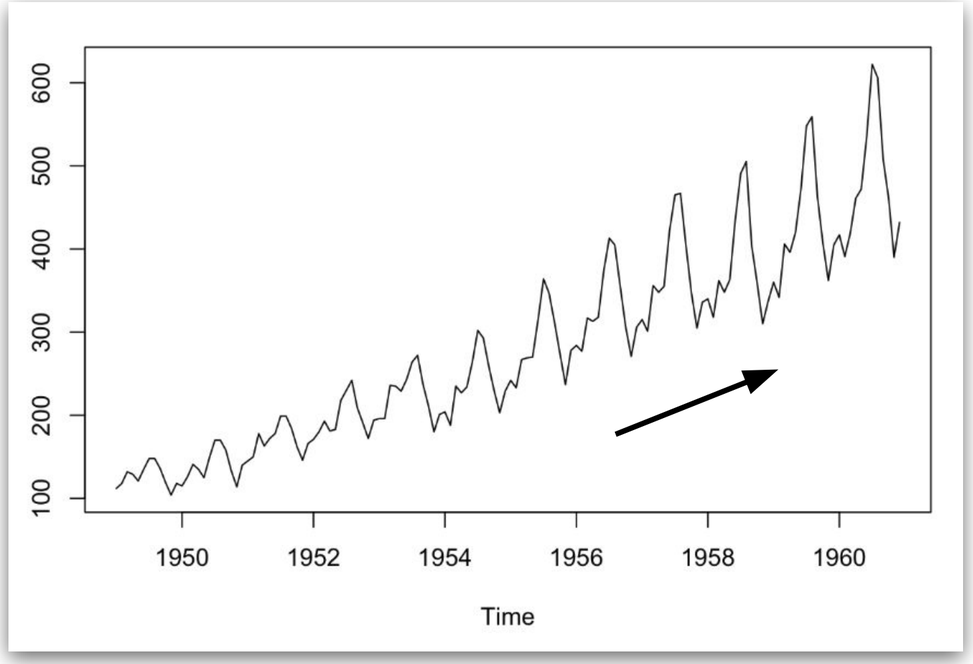


Concepts in Time Series

What are the other properties of time series data?

Most time series contain one or more of the following components:

- Trend
- Seasonal
- Cyclical
- Irregular (or Residual)

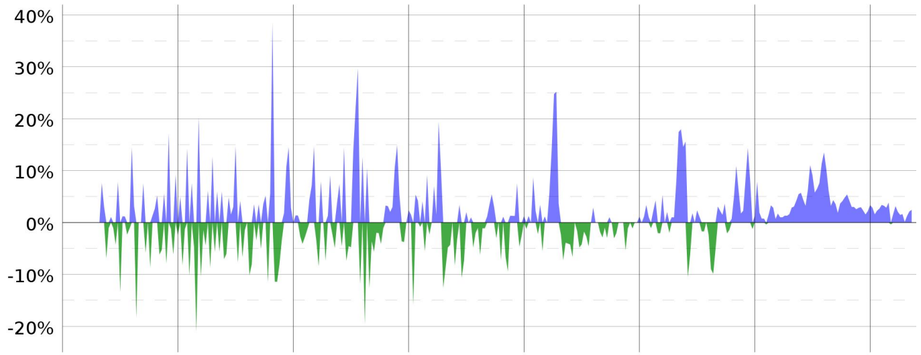


Concepts in Time Series

How do we change stock prices to be more stationary?

Stock prices are typically trending (up or down). But in any case, they have changing mean over time.

Hence we must difference stock prices to get daily, monthly or annual returns to make them “stationary”



Agenda

Concepts in Time Series

Time series terminology

AR - Auto regressive

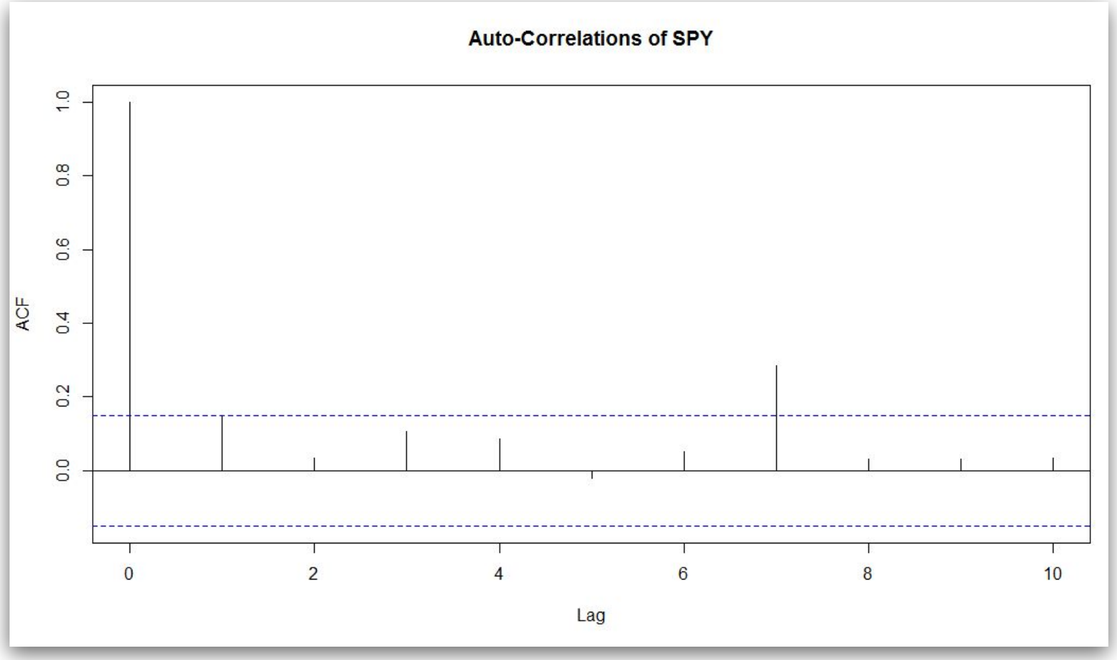
MA- Moving Average

Putting it all together

ARIMA Model

Time Series Terminology: Auto Correlation (AR)

A correlation of a variable with itself at different time periods in the past is known as “Autocorrelation.



How can we identify what value of lag we should use?

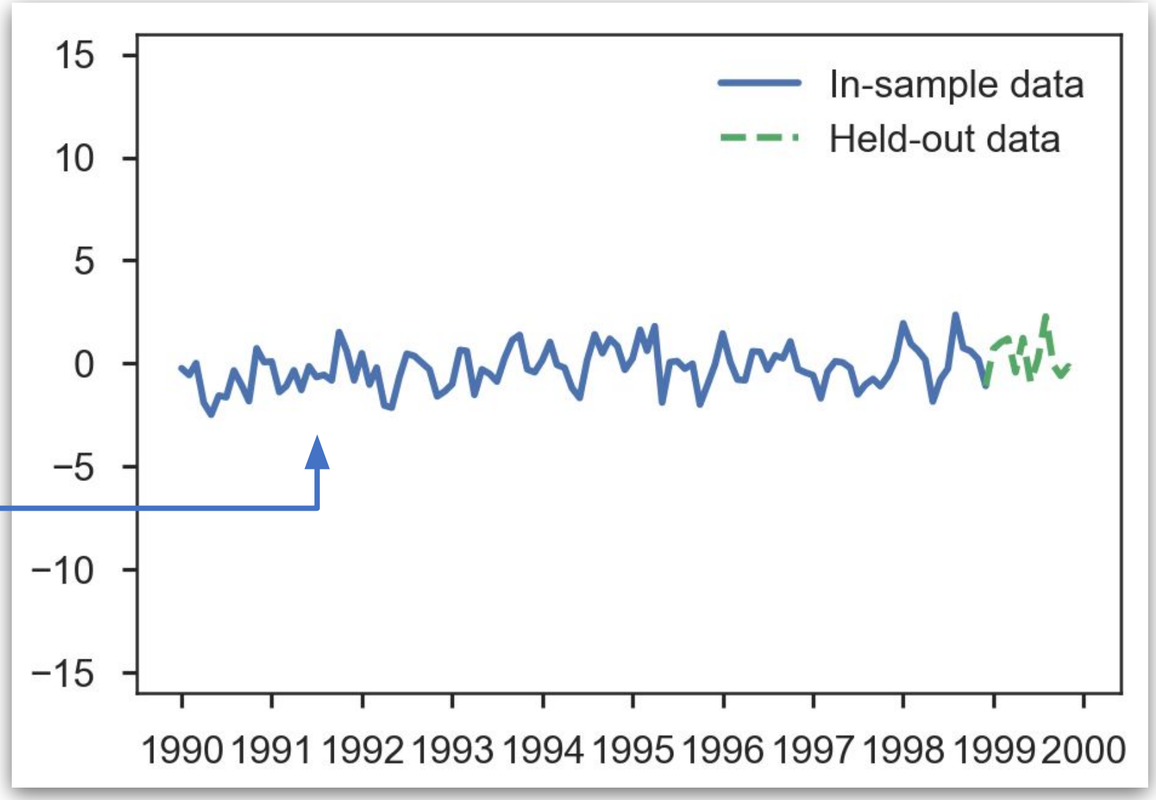
Time Series Terminology: Auto Regression (AR)

An AR process is where autoregression occurs. Our goal is to find the “correct” time lag that best captures the “order” of such an AR process. This is not a one-step procedure but is an Iterative process.

Auto Regressive (AR) process

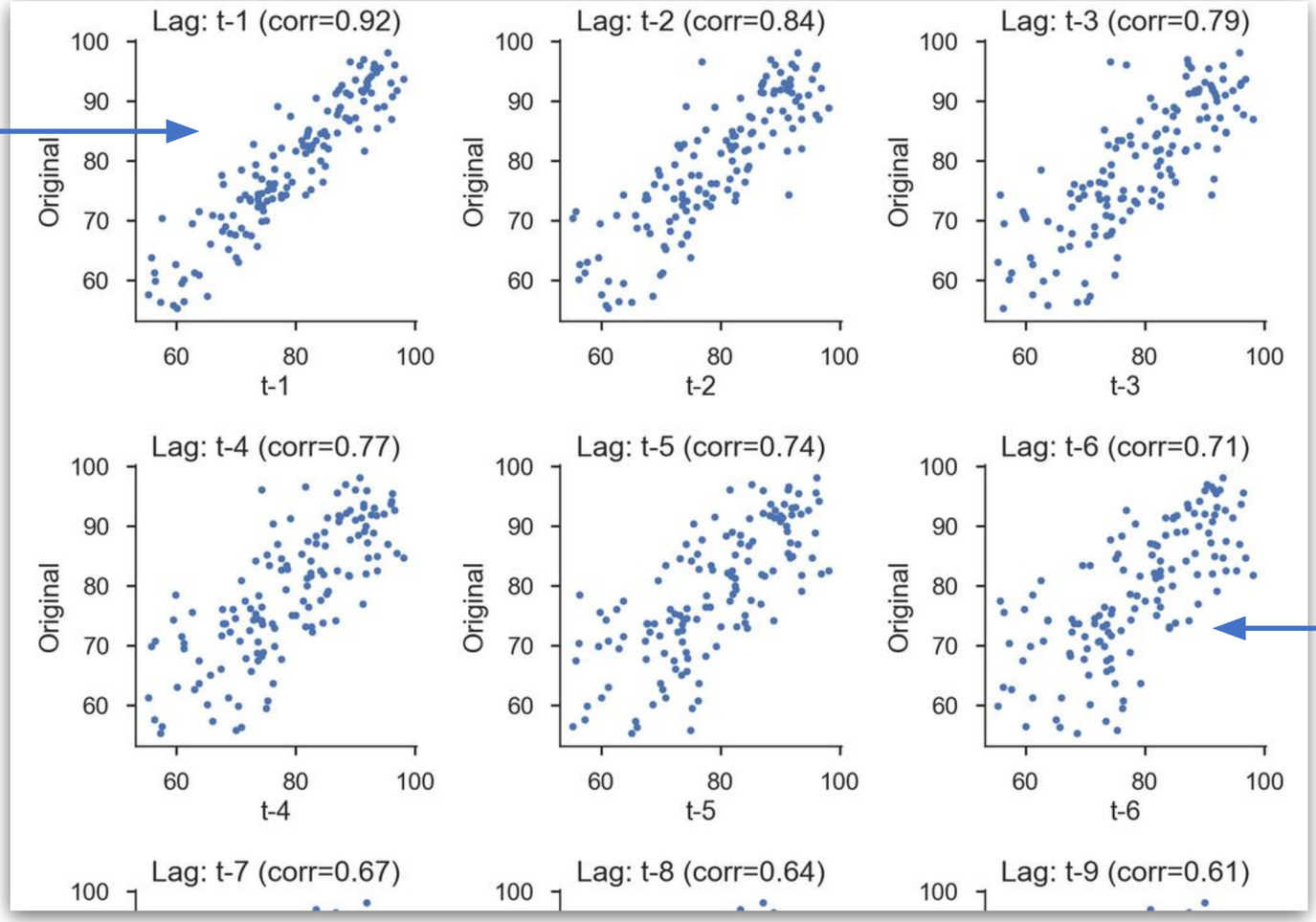
An AR process is where autoregression occurs. Our goal is to find the “correct” time lag that best captures the “order” of such an AR process. This is not a one-step procedure but is an Iterative process.

Here is how an AR process with a time period lag of 1 looks like.



Auto Correlation with lags

Notice how tightly packed it is to y values one time period ago



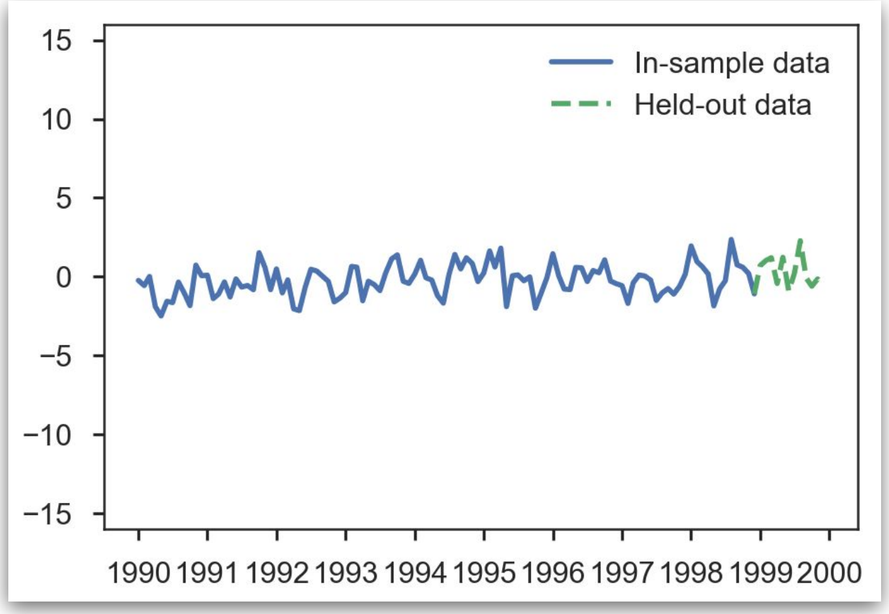
Notice how dispersed it is to y values 6 time periods ago

Lag 1

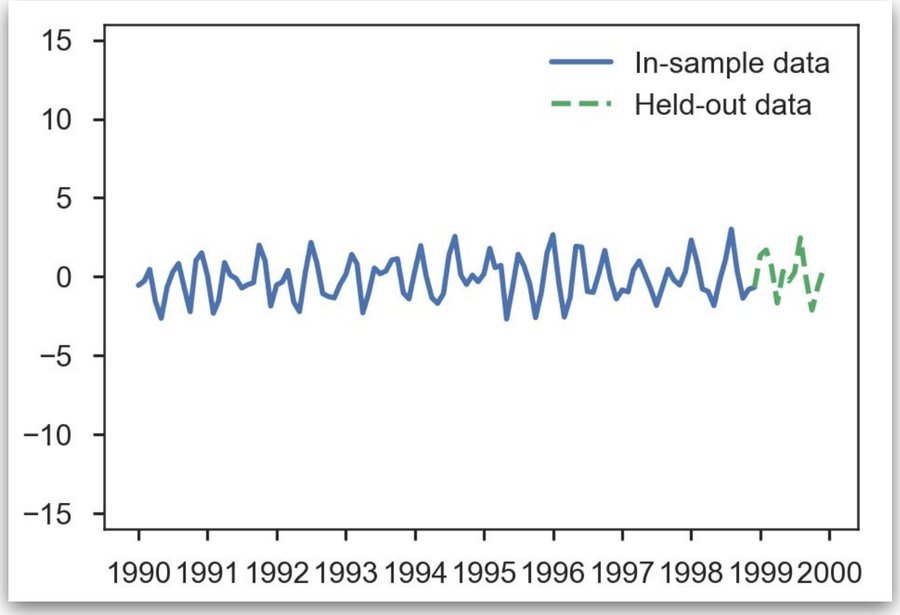
Google Cloud

Understanding Lags

Here is how an AR process with a time period lag of 1 looks like.



Here is how an AR process with a time period lag of 2 looks like.



Agenda

Concepts in Time Series

Time series terminology

AR - Auto regressive

MA- Moving Average

Putting it all together

ARIMA Model

Understanding an MA process

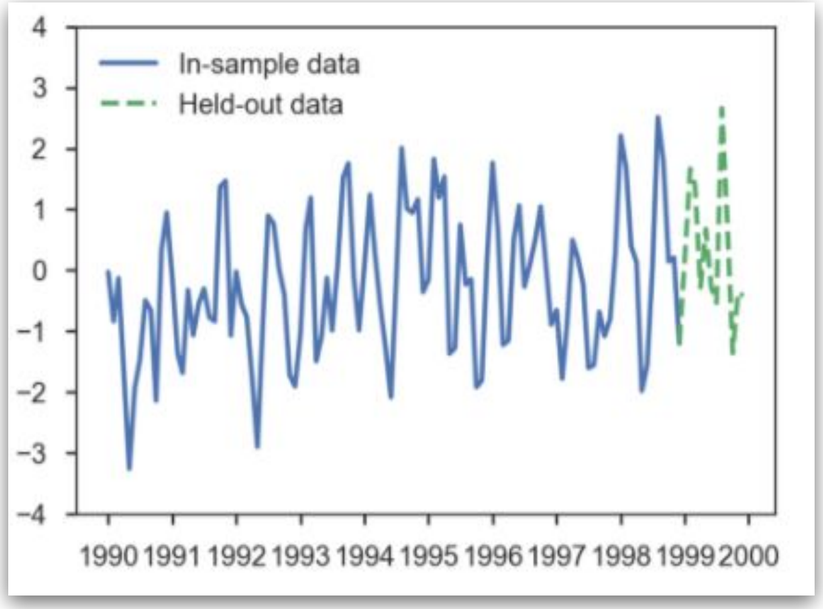
Moving Average (MA) models do not take the previous y values as inputs, but rather take the previous error terms. How do we get error terms if we don't predict?

We will attempt to predict the next y value based on a moving average (MA) - that's how we get error terms: how wrong our moving average was compared to the actual value. But there can be more than one error term since there can be many moving averages...

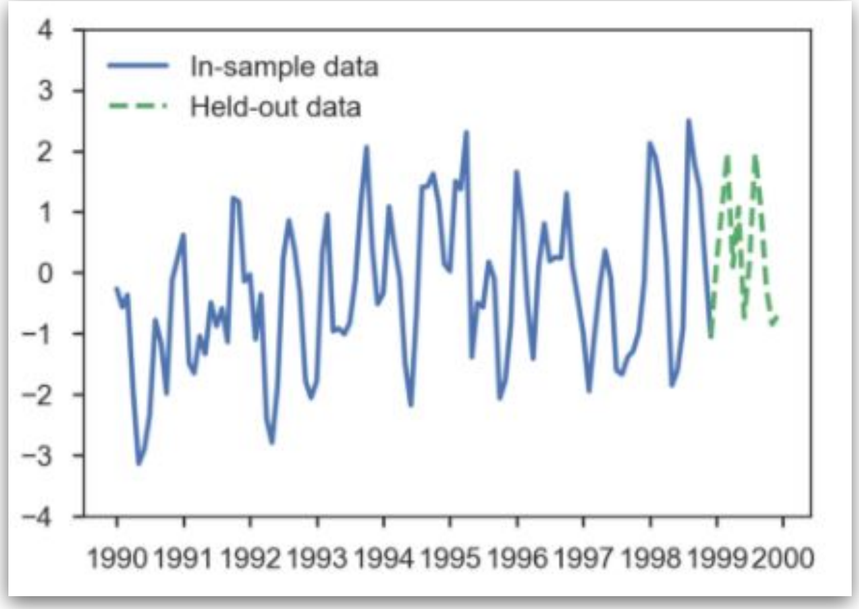
How can we identify what number of Error terms we should use?

Finding number of error terms to use

The model is iteratively built by assuming some value, q, for the number of previous error terms we must input into our forecast to calculate our prediction for future values.



Here is how an MA process with one error term looks like.



Here is how an MA process with 2 error terms looks like.

Agenda

Concepts in Time Series

Time series terminology

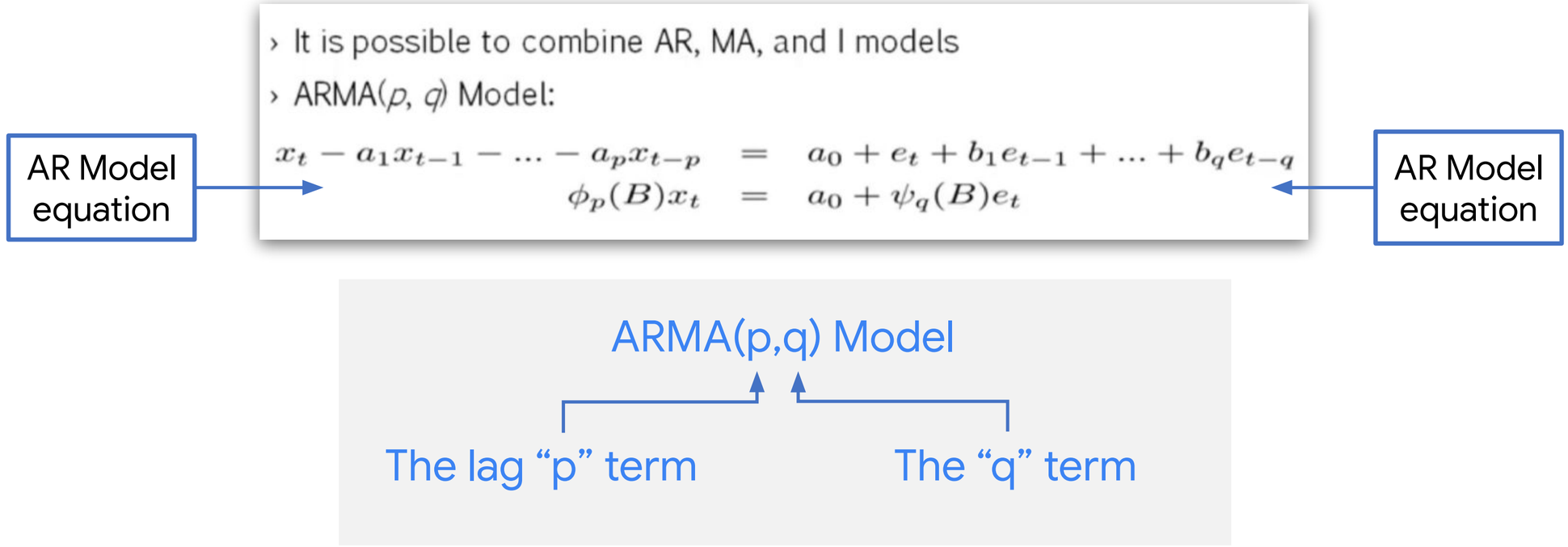
AR - Auto regressive

MA- Moving Average

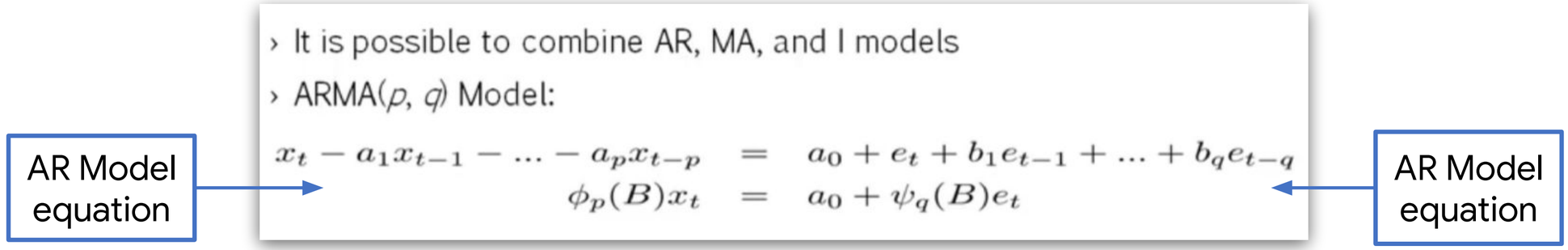
Putting it all together

ARIMA Model

How does a combined model look like?



How does a combined model look like?



ARMA(p,q) Model

The lag “p” term signifies how many prior time periods that each observation is highly correlated to, Increasing “p” (longer lag) would increase the dependency on previous values further.

The “q” term signifies how many prior time periods we’re considering for observing sudden trend changes.

Agenda

Concepts in Time Series

Time series terminology

AR - Auto regressive

MA- Moving Average

Putting it all together

ARIMA Model

ARIMA Model

Let’s go one step further and try to add another term to the ARMA model - it’s called ARIMA.

What does the “i” stand for?

“I” stands for Integrated: we are combining AR and MA techniques into a single integrated model: ARIMA.

“I” helps to “stationarize” the data.

ARIMA Model

ARIMA (p,d,q) Model

In an ARIMA model, there will be three parameters that will be needed.

› ARMA(*p*, *q*) Model:

$$x_t - a_1 x_{t-1} - \ldots - a_p x_{t-p} = a_0 + e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q}$$
$$\phi_p(B) x_t = a_0 + \psi_q(B) e_t$$

The additional “d” term tells ARMA that we are now predicting the DIFFERENCE between one prior period and the new period, rather than predicting the new period’s value itself.

d=1 may cause stationarity for a model
d=2 may capture exponential movements in our series

What would a typical ARIMA Model look like?

What does the model mean? → ARIMA(1,0,2)

We can add many more variations to the above including the following:

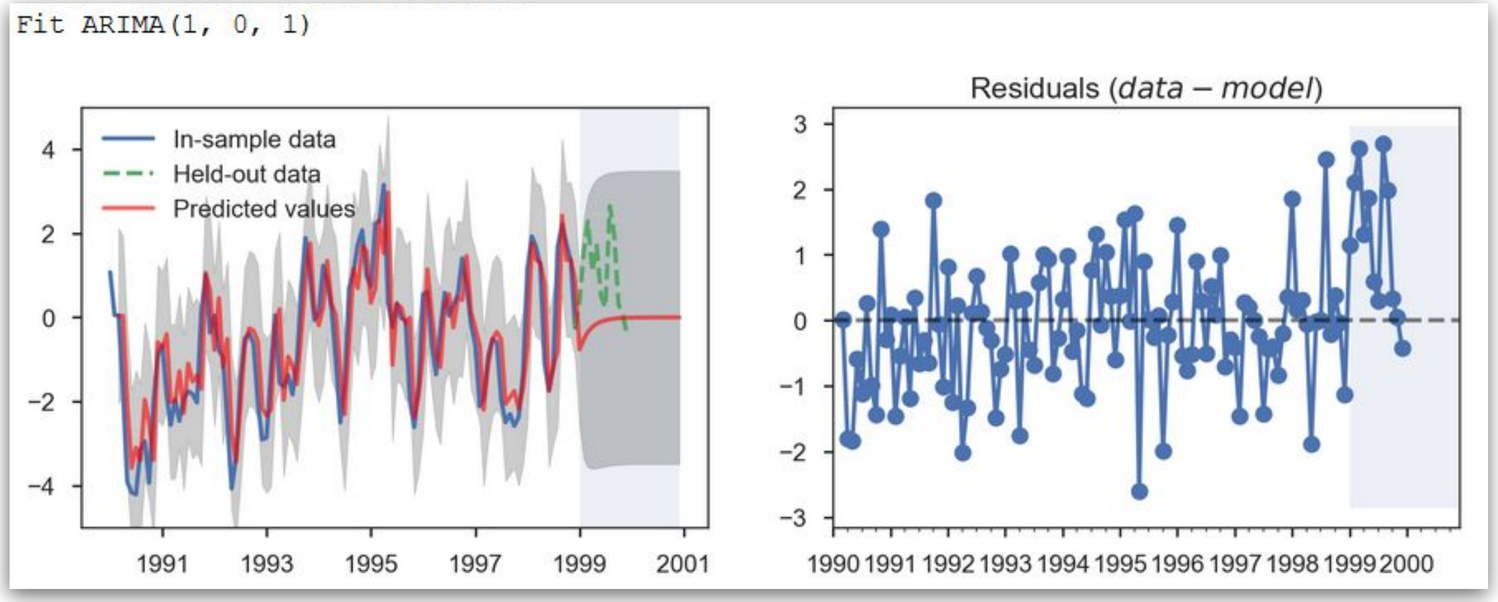
- Seasonality (SARIMA)
- ARFIMA
- SARIMAX,
- Add your own variation here

(1,0,2) implies that each term in the model is correlated with one output prior (p=1) and two prior error terms may be helpful in predicting the next y value (q=2). The zero in the middle implies that the data is already stationary (d=0).

Evaluating our ARIMA Model

In addition, there are a few key tactics we can explore:

- Plotting our residuals
  - If we do not observe a pattern in our residual error terms, we can stop iterating
- Ljung-Box Test
  - We can mathematically test the above assumption using the Ljung-Box test



* Source: https://wudan0707.wordpress.com/time-series-analysis-assignment-4/