Introduction to Time Series Analysis using R

By: Alice Roberts

Introduction

Data that depends on time; Y=F(t).

Time Series

Here are 3 different definitions of a time series :

1. A time series is a series of data points indexed (or listed or graphed) in time order
2. A time series is a sequence of observations collected at some time intervals.
3. Any metric that is measured over regular time intervals makes a Time series.

Why do we care about Time Series?

It's Everywhere

Data obtained from observations collected sequentially over time are extremely common.

In the biological sciences, we observe the electrical activity of a heart at millisecond intervals.

The list of areas in which TS are studied is virtually endless

It's Informative

The analytical study of a Time Series is important so as to forecast regarding the fluctuation of the data in the future, on the basis of the trend studied from the data.

So, Time series analysis may be regarded as a decision making factor of any concern for their future plan and estimate.

It helps us understand past behaviour.

Stationarity

Key Concept of Time Series Analysis

2

Stationarity?

Stationarity is a critical assumption in time series models, and it implies homogeneity in the series; series behaves in a similar way regardless of time.

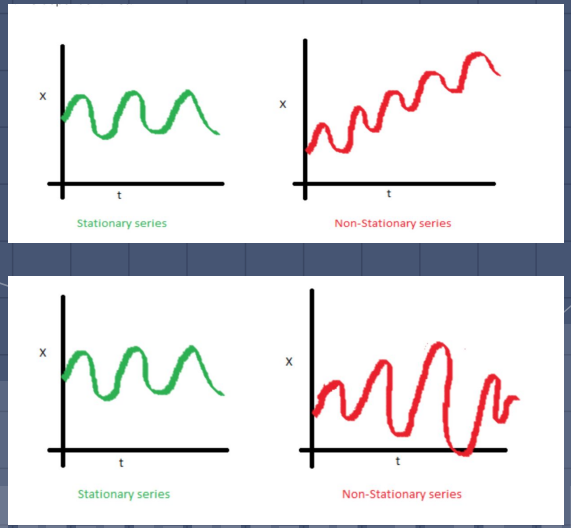
Properties :

The mean of the series should not be a function of time rather should be a constant.

The variance of the series should not a be a function of time.

Testing for Stationarity:

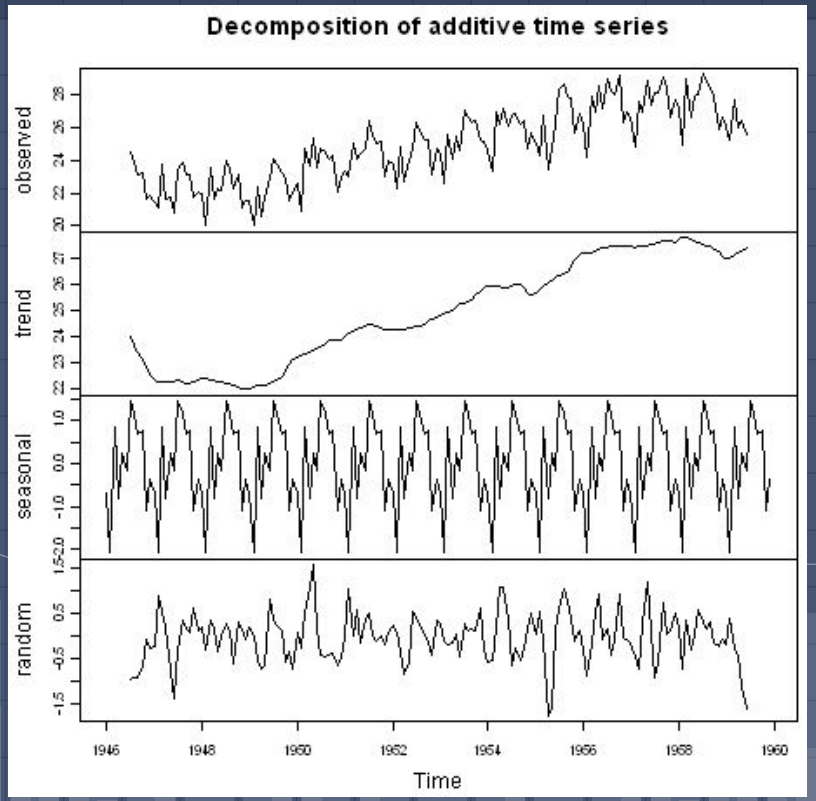
R has a test called the “Augmented Dicky Fuller test”. If the p-value obtained from ADF is less than 0.05 then we consider the TS stationary.



Compositions of an original time series

We decompose the original time series, into several independent components:

- **Trend–** Overall direction of the series (upwards, downwards, etc)
- **Seasonality–** monthly or quarterly patterns
- **Cycle -** recurrent variations in TS (business cycles)
- **Irregular remainder –** random noise left after extraction of all the components



What if our time series is NOT stationary?

We can make it stationary by differencing.

- Differencing a time series means to subtract each data point in the time series from its successor.

  Note:

- Sometimes we take a log transformation of the data stationary on the variance.

No Differencing (d=0)	$Y_t^{'} = Y_t$
1st Differencing (d=1)	$Y_t^{'} = Y_t - Y_{t-1}$
2nd Differencing (d=2)	$Y_t^{'} = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2})$

ARIMA Forecasting

Auto-Regressive Integrated Moving Average

3

AR(p)

Auto Regressive (AR)

-Regressions on themselves

-When a value from a time series is regressed on previous values from that same time series.

-extracts the influence of the previous periods' values on the current period.

-AR(P)

Formula:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

I (d)

Integrated (I)

- Subtract time series with its lagged series to extract trends from the data.

-Differencing is one of the most commonly used mechanisms for extraction of trends.

-I(d)

Formula:

No Differencing (d=0)	$Y_t^{'} = Y_t$
1st Differencing (d=1)	$Y_t^{'} = Y_t - Y_{t-1}$
2nd Differencing (d=2)	$Y_t^{'} = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2})$

MA (q)

Moving Average (MA)

-extract the influence of the previous period's error terms on the current period's error

-Idea with moving average is to remove all the zigzag motion from the time series to produce a steady trend. We do this through averaging adjacent values of a time period.

-MA(q)

Formula:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

ARIMA (p,d,q) - Questions:

When we analyse our time series, we have 2 main questions:

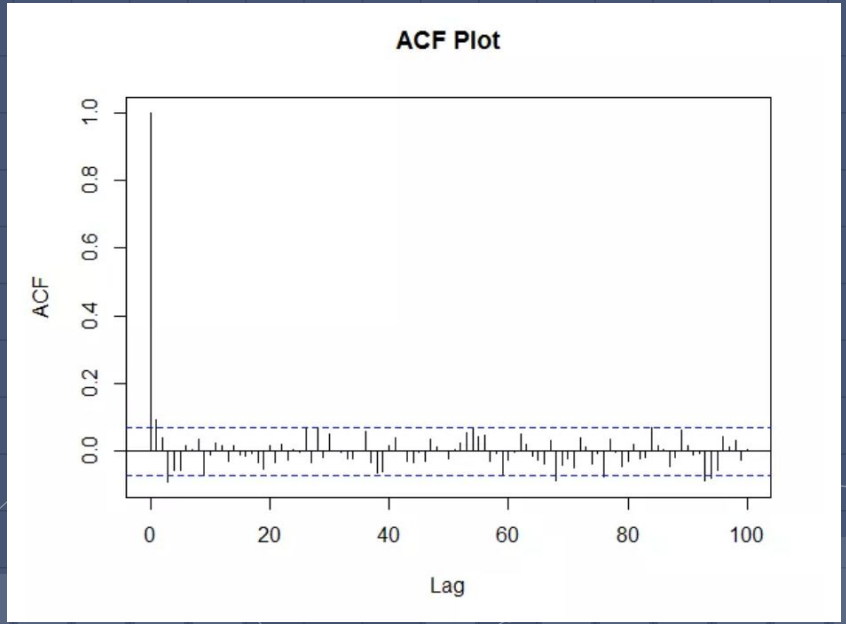
1. Is it an AR and/or MA process?

2. What order of AR/MA process do we need to use?

Autocorrelation Function

What is ACF?

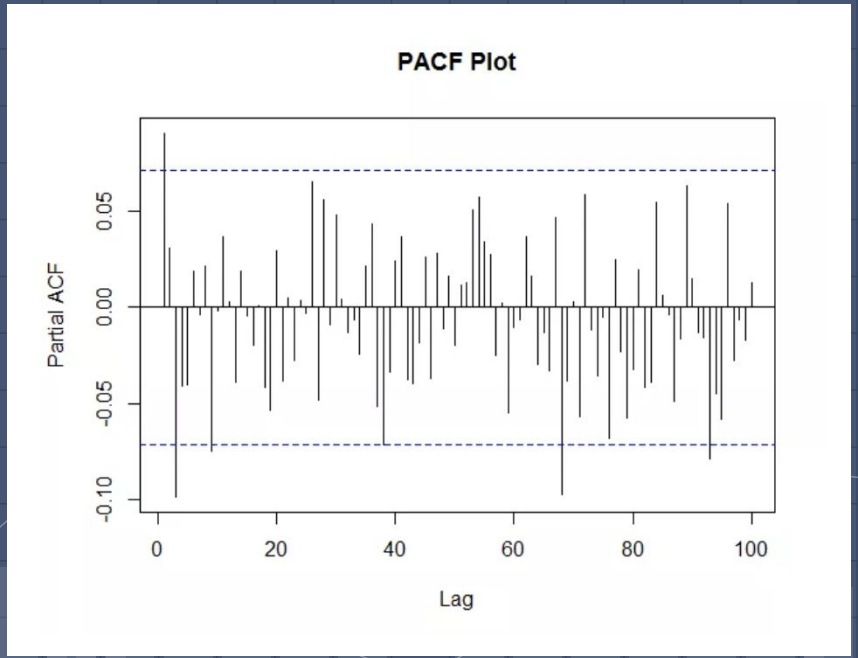
It is the correlation of a Time series with lags of itself.



Partial Autocorrelation Function

What is PACF?

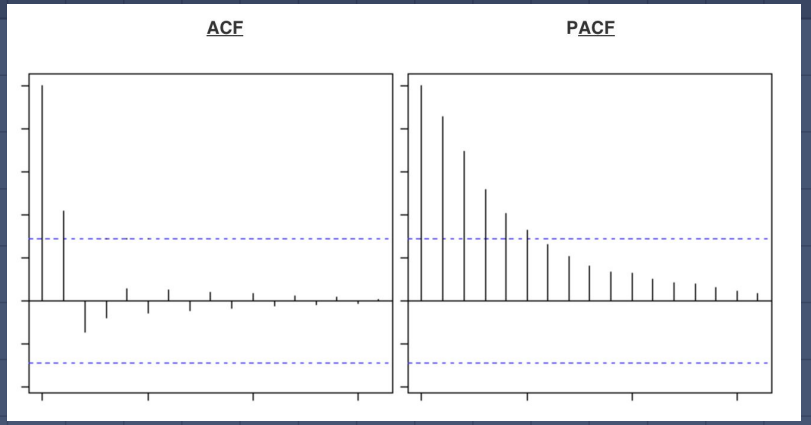
It is the correlation of time series with a lag of itself, with the linear dependence of all the lags between them removed.



Identification of AR/MA Models and its Order

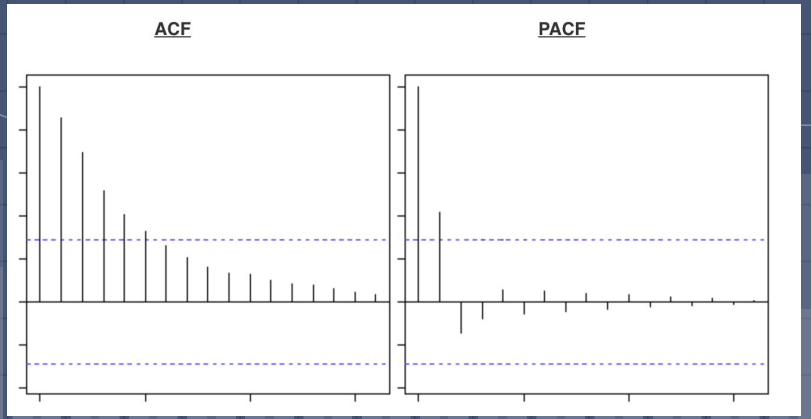
Identification of q order of MA Model

For MA Models, the PACF will dampen exponentially and the ACF plot will be used to identify the order of the MA process. If we have one significant spike at lag 1 on the ACT, then we have an MA model of order 1



Identification of p order of AR Model

For AR Models, the ACF will dampen exponentially and the PACF will be used to identify the order(p) of the AR model. If we have one significant spike at lag 1 on the PACF, then we have an AR model of the order 1



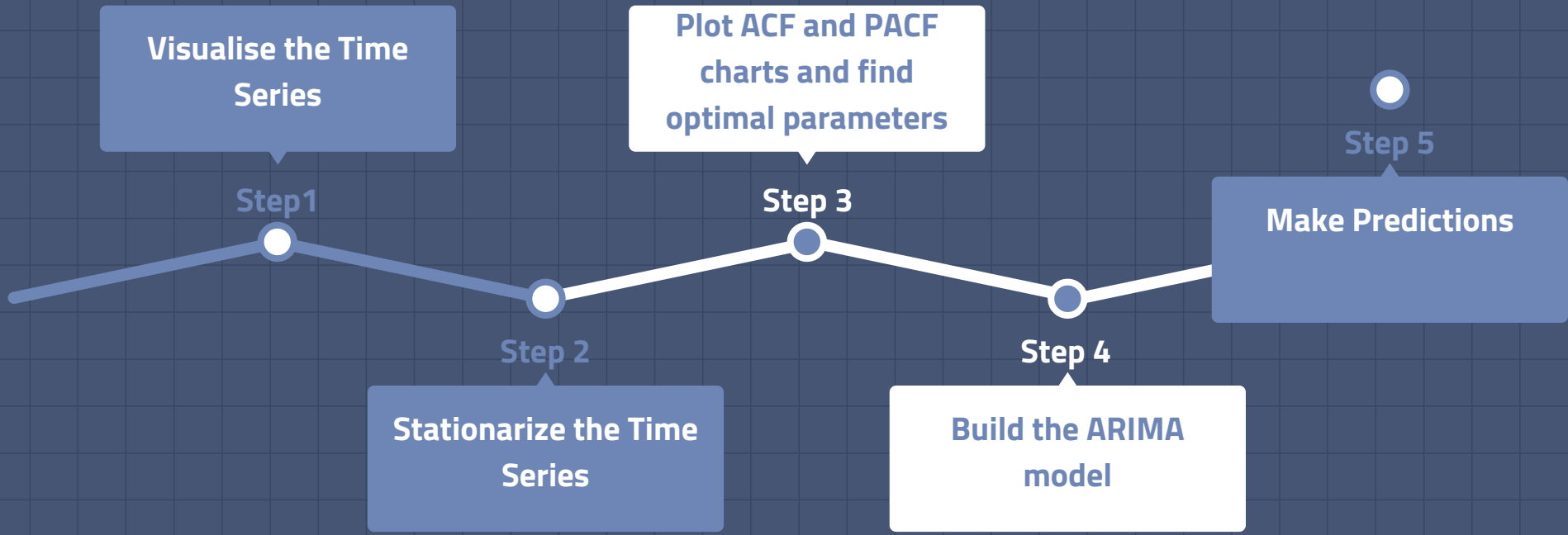
Building ARIMA Model

Now that we have identified the model and the order associated with it, we can build our ARIMA model.

Once we have built our ARIMA model we can make predictions on our time series.

We will move to our R code shortly to show how this is done.

Recap : Steps for TSA



Stock Forecasting

https://github.com/aliceroberts10/WorkShop-TSA

"Prediction is very difficult,
especially if it's about the future."
-Niels Bohr

CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/
- https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html
- https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

THANKS!

Any questions?

You can find me at:

- Github: Aliceroberts10
- Roberts.alice04@gmail.com