Q-STEP 'HOW TO' GUIDES:

EXAMPLE OF A TIME SERIES

Creator: Patrick English

This guide takes you through an example of time series data and time series analysis. All the data was downloaded directly from the World Bank. Terminology is used from the main page and users should refer to this for definitions.

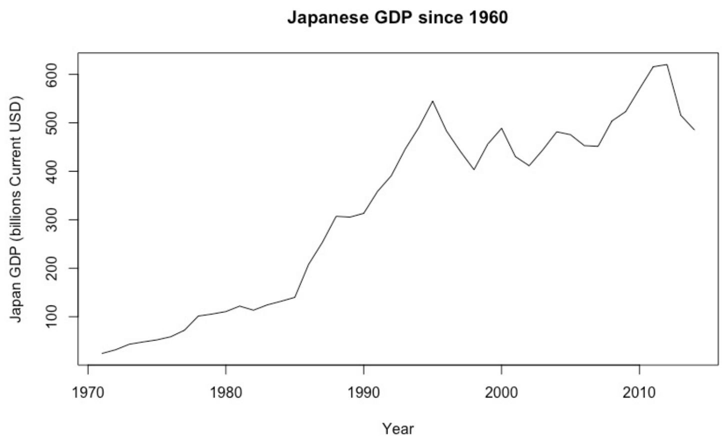
Example 1 - GDP in Japan

Suppose we want to know the extent to which we can predict Japanese GDP over time using the following three variables:

Population: Total population (Millions of people)
School Enrolment: Gross ratio of total enrolment into primary education
Carbon Emissions: CO_2 emissions (metric tonnes per capita)

To do this, we need to fit a time series regression model to the data and examine the extent to which these three factors can predict changing GDP over time. It is the temporal nature of time series analysis makes the process unique, introducing all kinds of different considerations and potential pitfalls which researchers must walk through.

All time series analysis should start with a detailed descriptive section on the nature of the dependent variable. It is important for readers to see the time series data which is going to be worked with, so that they can get a sense for themselves of the structure and any potential trends or unit root processes in the data.



The figure plotted above represents Japan's reported GDP (in billions of current US Dollars) from 1970 through until 2017. We will now run a time series analysis on the data to attempt to explain variation in Japanese GDP using other macro-indicators about the country.

Unit Root Tests

The first thing we need to establish before running a time series analysis is whether our not our dependent variable contains trend (or some other stochastic process that we need to account for). We will need to establish stationarity in our modelling approach if the variable is indeed found to be non-stationary over time.

The easiest way to test for stationarity is to use an augmented Dickey Fuller (ADF) test. ADF tests establish whether or not a 'unit root' (a non-stationary, stochastic process in a time series data sample) is present in our series. The null hypothesis of the test supposes that there is a unit root underlying our data (and that the data is non-stationary), while a rejection of the null hypothesis in an ADF would suggest that the data did not contain a unit root. The ADF test is able to account for different 'trend' and 'lag' components which enable better analysis of the unit root hypothesis.

In the case of the Japanese GDP data, an ADF test clearly indicates that the data is not stationary. We can see from the R output on the right that the null hypothesis (that the series is has a unit root) fails to be rejected in all cases.

This therefore requires us to use time series methods to account for the stochastic process when modelling the series. If we do not, then we will run a heavy risk of pulling out a spurious result.

Note: though the (augmented) Dickie Fuller test is a test for unit roots, such tests are actually not reliably able to differentiate between a unit root and other types of non-stationary processes - particularly autoregressive terms. Furthermore, they are not reliable in correctly rejecting the null hypothesis when working with short time-series data. See Pickup (2014).

```
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag   ADF p.value
[1,]   0 1.481   0.961
[2,]   1 0.647   0.823
[3,]   2 1.441   0.958
[4,]   3 0.830   0.875
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -0.702   0.791
[2,]   1 -1.088   0.657
[3,]   2 -0.738   0.778
[4,]   3 -1.052   0.669
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -2.19   0.483
[2,]   1 -3.14   0.129
[3,]   2 -2.19   0.485
[4,]   3 -3.08   0.149
```

A popular route to establishing stationarity in a time series is to **difference** the variables - in effect create a measure of 'annual change', where the differentiated annual value of the dependent variable Y (dYt) is worked out as the value of Y at time t subtracted by the value of Y at time t-1.

However, since a unit root test (such as a Dickey-Fuller) can suggest a unit root when the autocorrelation is in fact caused by autoregressive (AR) and/or moving average (MA) components, it is best practice to first model the data using techniques to account for AR/MA terms first. If this approach is unsuccessful, then we can thus conclude that the series is integrated and move to including integration points in the model specification (thus differencing the time series).