**Business Understanding**

In June 2017, Australia took the world record for the longest run of uninterrupted growth in the developed world **(SMH)**. It has now been 107 quarters since Australia had a technical recession, defined as two consecutive quarters of negative economic growth.

As one of the most widely used economic indicators, Gross Domestic Product (GDP) is used to gauge the health of a country’s economy (Investopedia 2018). GDP represents the total dollar value of all goods and services a country produced over a specific time period, often referred to as the size of the country’s economy.

Given the importance of having a healthy economy to the wellbeing of a country’s citizens, our team considered it would be important to try and understand what are the factors driving GDP and whether it was possible to predict future GDP of Australia using other information.

Our team viewed choosing this topic as a learning opportunity, to better understand how we as individuals and as a community can contribute to the economy. As a team of data analysts without any formal qualifications in economics, will try to decode the economic jargon and provide insights on the important factors that influence a country’s economy.

**Research questions**

As there are two types of GDP that economists use to measure a country’s economy, our regression model will disregard real GDP (economic output adjusted for the effects of inflation) and solely focus on predicting nominal GDP (a country’s economic output without an inflation adjustment). The research questions that we try to answer with the data:

* **Can GDP be accurately predicted given the historical economic and non-economic factors?**
* **Which** **Economic and Non-Economic factors are most influential to nominal GDP?**
* **Does unemployment rate have an effect on GDP?**

**Research Approach**

There were three common methods used in the measurement of GDP

* Income Approach
* Production Approach
* Expenditure Approach

As it did not matter which method we choose as long as we were consistent in our logic, we choose the expenditure method. It had the most readily available information for calculating GDP based on the formula:

* Model 1 looks at the Expenditure Approach (GDP = C + I + G + X)
* Model 2 looks at the indicators for the components of the Expenditure Approach e.g. using CPI and Sales as indicators instead of Consumption
* Model 3 takes into consideration other available data sources, not limited to the model e.g. population of Australia

**Model 1**

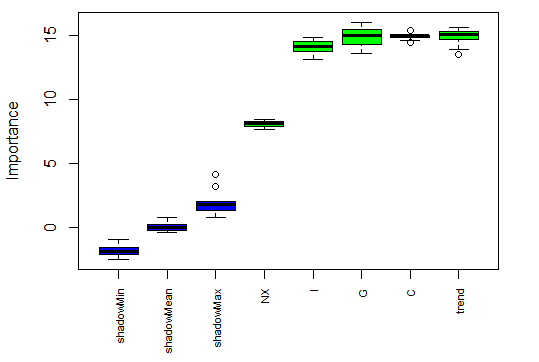
To start, Model 1 investigated the importance of the variables used in the expenditure formulae (GDP = Consumption + Investment + Government spending + Net Exports). Given this was a time series of annual data, a trend variable was added.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Average | Pr(>|t|) | varImp score |
| Consumption (C) | 455752 | 8.30e-07 | 5.97 |
| Investment (I) | 143344 | 8.98e-08 | 6.71 |
| Government Spending (G) | 190278 | 0.045887 | 2.07 |
| Net Exports (NX) | -1589 | 0.000576 | 3.79 |

Consumption was the largest amount in linear regression model, Investment was found to be the highest affected variable

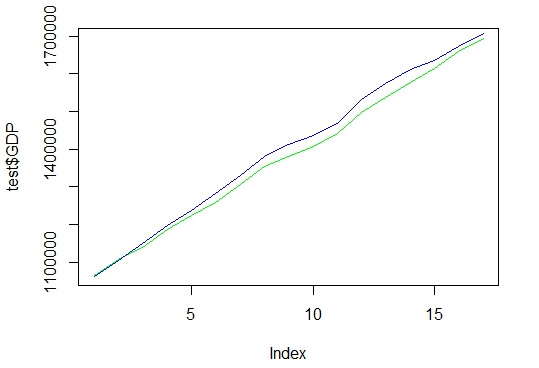
Similarly with Net Exports even though it was a negative average, much smaller than Government spending, it was deemed more important to calulating GDP

Var Imp score showed a similar ranking of variable importance



Boruta package showed a different result, I gave similar weightings to the first three and less important weighting to net exports.

The data set was split into train and test and predictions generated were as follows.



If these four are know they are good predictors