**Business Understanding**

In June 2017, Australia took the world record for the longest run of uninterrupted growth in the developed world (Bagshaw & Massola 2017). 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 (measured quarterly by the ABS), 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 examined with the data:

* **What are the main factors that contribute to GDP in Australia? (Main)**
  + **Does unemployment rate have an effect on GDP?**
  + **Does increase in GDP lead to increase in ASX50 stock prices?**
  + **How does interest rate changes affect GDP?**
  + **If given the right variables can GDP be accurately predicted?**

**Research Approach**

There are three ways that a nation’s GDP is traditionally measured, and conceptually they all should deliver the same estimate (Australian Bureau of Statistics 2012). They are:

* The income approach measures income generated by the economy.
* The production approach calculates the sum of the value of goods and services produced by each industry minus the those used in production.
* The expenditure approach measures final expenditures on goods and services.

As it did not matter which method we choose as long as we were consistent in our logic, we choose the expenditure approach to form our starting basis. This decision was not to discount the importance of variables with the other approaches, but due to the expenditure approach for calculating GDP variables having the most readily available information.

In doing so we took a progressive approach to model building and examining factors that contributed to GDP.

* Model 1 examined the variables of the Expenditure Approach (GDP = C + I + G + NX).
* Model 2 examined and used indicators for the components of the Expenditure Approach e.g. using CPI and Sales as indicators instead of Consumption.
* Model 3 took 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(C) + Investment(I) + Government spending(G) + Net Exports(NX)). A simple linear regression model was used to examine the variables.

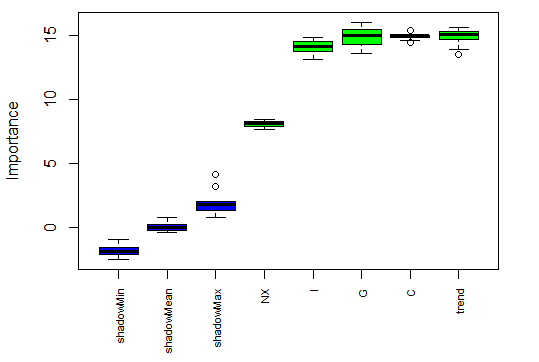
The hold out method (Wikipedia 2018) was used to split the available data into a training set and testing set. As this was a time series, the training and test split was not done randomly, but the training set corresponded to first 70% of the available data and the testing set corresponded to remainder 30% of the available data for each combination. This was to ensure the model would not be built using training set data it would not have had access to at the time of prediction (using test set data). Given this was a time series of annual data, a trend variable was also added.

*Table 1: Model 1 variables*

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

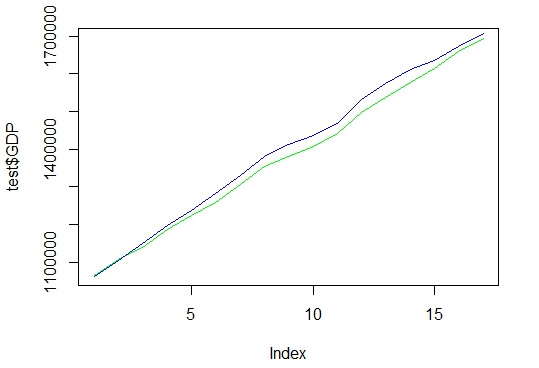
What the simple linear regression model found was that whilst the Consumption variable had the highest average (three times the size of Investment), Investment was found to be the variable that impacted the model the most (based on both P-value and varImp score).

Similarly with Net Exports even though it had a negative average, and was much smaller than government spending average, it was found to have a greater impact on the linear regression model than government spending (based on both P-value and varImp score).



Alternatively, the Boruta package was also applied to the data and it showed a different result. The weighting Boruta package gave to net exports in part reflected the size of this amount when compared to the other variables. It was interesting however, that despite Consumption doubling or tripling the size of Government Spending and Investment respectively, the Boruta package gave similar weightings to all three variables for their importance to GDP.

The linear regression model generated by the training data set was then applied to the test data.



The model fit (as measured by R-squared) was 99.9, whilst accuracy of predictions (as measured by RMSE) was 23384.34. Whilst the simple linear regression model assisted us in better understanding the variables we were dealing with and how they related to GDP, the overly high and almost perfect R-squared highlighted the questionability of using a simple linear regression model to predict GDP with these variables.

# Reference List:

##### Bagshaw, E. & Massola, J. 2017, *GDP: Australia grabs record for longest time without a recession*, SMH, viewed 15 May 2018, < https://www.smh.com.au/business/the-economy/gdp-australia-grabs-record-for-longest-time-without-a-recession-20170607-gwm0o2.html >.

Wikipedia. 2018, *Training, test, and validation sets*, San Francisco, viewed 10 May 2018, < https://en.wikipedia.org/wiki/Training,\_test,\_and\_validation\_sets >.

Australian Bureau of Statistics 2012, *Defining and measuring GDP*, viewed 28 April 2018, <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1301.0~2012~Main%20Features~Defining%20and%20measuring%20GDP~221>.

Investopedia 2018, *What is GDP and why is it so important to economists and investors?*, viewed 28 April 2018, < https://www.investopedia.com/ask/answers/199.asp>.