Three Centuries of Economic Data

A tiny data science project exploring Monetary Policy and the three Pillars of Economics

I. Introduction

This data science project will analyse three centuries of economic data to examine two aspects: the first will be an in-depth analysis of the Bank of England's monetary policy and its impact on inflation and unemployment within the UK. The other facet of this of this report will be exploring GDP growth, its link to unemployment, and how it fares when calculated on a per-capita basis. At a time where interest rates are at unprecedented low levels and some questioning the role of the Bank of England in the aftermath of the Brexit referendum result, the effectiveness of its monetary policy and its political independence, this discussion is as pertinent as ever.

Central Banks around the world are responsible for setting national monetary policies, in which interest rates and money supply is adjusted in order to either stimulate or restrain the economy. The effect of monetary policy on inflation is a well-documented economic phenomenon. It is widely understood that central banks adjust base interest rates to manage inflation levels. Decreasing interest rates spurs economic activity by promoting lending and discouraging saving, thus increasing inflation as a result; while increasing interest rates decreases inflation [1]. Meanwhile an empirical model, named after A. W. Phillips [2], known as Phillip's Curve represents an inverse relationship between unemployment and inflation. The model suggests the level of unemployment has a direct effect on inflation, claiming that increasing unemployment decreases inflation and vice versa. Monetary policy aims to flatten the Phillip's curve - where inflation is at a minimum whilst still containing unemployment rates - an idealistic visualisation of this curve is shown in figure. On the other side of the same coin, the relationship between GDP change and unemployment change is also well documented. Okun's law [3] states that a 2% drop in unemployment is conducive of a 1% drop in GDP change.

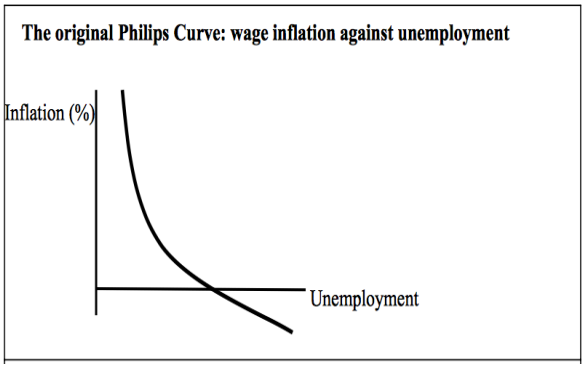

Figure 1: the original Phillips Curve showing theoretical relationship of inflation against unemployment [1]

This project will thus leverage UK's economic data to examine Monetary policy and the intricate balance between these three pillars of the economy: GDP growth, inflation and unemployment. To conduct this project, the data is first collected, an analysis strategy devised, the data is prepared for analysis, which is then followed by the actual analysis and its findings. This report will replicate this structure.

II. Data Collection

The data required for this project were pooled from 3 different sources. As such, it was present in various formats and covered differing date ranges – offering the first challenge in this data science project. The main data source, which spurred this study, comprised inflation and unemployment rates in the UK between the years 1871-2009[2]. The GDP and interest rates data were then drawn from the Bank of England's 'three centuries of economic data', covering the years 1831-2009[3]. Finally, the population statistics in the UK over the years 1971-2009 were collected from the World Bank[4].

The data merged to create two data sets: one with the year as the index and inflation, interest rates, GDP growth and unemployment as the columns, each represented as a percentage and to one decimal point for uniformity. Once the data was compiled, the disparity in date ranges was visible where there were any missing values across the row. To treat this, the years with missing values were removed, leaving the date range in which the all data collected intersect. Thus the date range that remained available in the first data set of this project covered 1871-2009. The data was

1. http://www2.econ.iastate.edu/classes/econ102/bishnu/short_note.pdf
2. https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/13848
3. http://www.bankofengland.co.uk/research/Pages/onebank/threecenturies.aspx
4. http://data.worldbank.org/indicator/SP.POP.TOTL?locations=GB

inherently temporal and discrete in nature. Each column presented the corresponding economic measurement for that year, as a percentage and is thus a quantitative ratio data type. Where:

- Interest Rate: the base interest rate set by the Bank of England that determines the cost of borrowing. One arm of the monetary policy used for inflation-targeting and is set as an independent variable in this project
- Inflation: the rate at which the general price of goods and services rises. Central banks attempt to control inflation and prevent deflation, with inflation targets set at 2%.
- Unemployment: percentage of labour force that is unemployed but actively seeking employment. Optimal rate of unemployment lies somewhere between 4-5%.
- GDP growth: year-on-year change in GDP, measured as a percentage.

The second dataset was created by combining the real GDP value per year, measured in GDP million, and the population, covering the years 1971-2009.

The core objective of this study is really an exploration of the UK's economy by leveraging the data of the past 100 years. A limited number of features are chosen in order to allow more in-depth exploration between them. To align with the phenomenon mentioned in Section I, first the relationship between base interest rates and inflation is tested using regression analysis with inflation as the dependant variable, followed by a similar analysis of unemployment and inflation rates in order to observe if a Phillip's curve is apparent. Okun's law will also be tested using this data, by comparing the GDP growth against the change in unemployment rate each year. Besides the regression analysis, clustering the data will be attempted as well as producing a network graph based on the information inferred from the preceding analysis. Then finally, the GDP per capita is calculated for each year and compared to the actual GDP change to observe if an interesting pattern emerges.

The data therefore needs to be prepared in order to enable for this analysis to be conducted, this is explained in section III. Section V will present the results of this analysis and discuss some of the findings that were uncovered in the process, before section VI will comprise a review of the analysis along with some of the limitations identified and potential further work.

III. Data Preparation

This section explains how the data was prepared prior to conducting the analysis. First, the data is cleaned to ensure uniformity in all the variables that would be tested against each other and outliers were identified in case they affected the results. By using basic statistics, some initial exploratory data analysis is also carried out to characterise the data. Finally, the technique used that allowed easy comparison of GDP against the GDP per capita in the latter parts of this report is briefly described.

Abnormalities in economies occur every so often during history. Times where, often due to political or financial disruptions, economic data is askew from expected levels and thus would likely affect the explorations in this project. Box plots were the method used to quickly observe if any of these outliers existed, and some of the results were examined both with and without outliers. Since inflation is the main dependant variable in this study, outliers were examined based on the inflation rate and box-plots were the method used to identify outliers. The box plot in figure 2.a that clearly illustrated that outliers were present within the inflation rate data. The inflation rates that lay beyond the 10% minimum and maximum quantiles were then deemed outliers and the observation year that coincided with this outlier were removed from the source data. A box-plot was then reproduced to visualise if the outliers were cured, and figure 2.b shows that this was indeed the case.

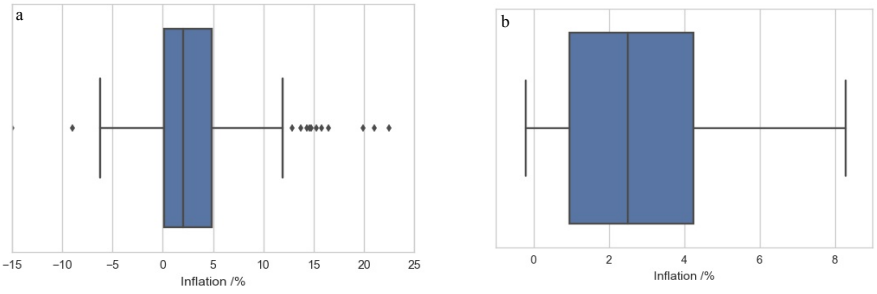

Figure 2: Inflation rate occurrences represented as a box plot, with a. including all the instances and b. following the removal of above 90% and below 10% interval

A statistical summary of the main data set is displayed in table 1 following the exclusion of the observed years with an outlier. This statistical summary was then modeled using density distribution graphs of the inflation, unemployment

and interest rates and shown in figure 3 – which indicates that all three variables experienced a few irregular observations to the right of the mean even after outliers were removed. The similarity between unemployment and inflation density distribution is visible, with both following a slight-left skew and similar spread of observations. The base interest meanwhile experiences limited spread from the mean in comparison, with most observations occurring around the mean and that is evident in its lower standard deviation.

The key observation from table 1 and its illustration in figure 3 is that abnormalities still existed in the data despite the treatment of the outliers. This is evident for example in the interest rates statistics, where the maximum observation was 15% although the mean was 5.1% and a standard deviation of 2.3%.

	Interest Rate /%	Unemployment /%	Inflation /%	GDP Growth /%
Mean	5.1	4.9	2.4	2.1
Standard Deviation	2.3	3.1	2.8	2.35
Minimum Value	0.5	0.4	-1.1	-5.3
Maximum	15.0	11.9	8.3	7.9

Table 1: Summary of statistics of first data set including the inflation, unemployment and interest rates

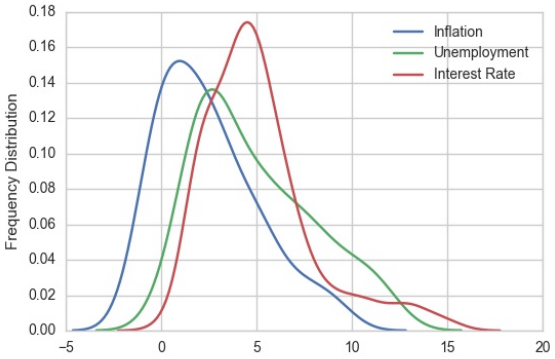

Figure 3: density distribution of inflation, unemployment and interest rates between 1871-2009

To calculate the GDP per capita for the latter parts of the analysis, GDP was first divided by the population for each year from 1971-2009. Then to allow easy comparison of GDP per capita against the actual GDP, the first observations in 1971 were normalised to a value of 100 for both entries to allow intuitive visualisation in the analysis section. Table 2 below shows a slice of the final data table with the normalised values of GDP.

Year	Real GDP /GBP million	Population	GDP per Capita Normalised /-	Real GDP Normalised /-
1971	507,493	55,928,000	100	100
…	…	…	…	…
2009	1,139,859	62,260,500	201.8	224.6

Table 2: Slice of the second data set, with GDP growth, Real GDP measured in GBP millions, Population and then GDP and GDP per capita normalised to the value of 100 in 1970 for both to allow easy comparison

IV. Analysis

Figure 4.a visualises the base interest rate, inflation and unemployment over time, covering the years 1871 – 2009 and including the outliers identified in section III. Figure 4.b shows the same data but excluding the outliers that were identified earlier. The most striking outliers in the inflation rate appear around the years 1915-1925, where the effects of the first world war likely resulted in stints of high inflation preceding stints of high deflation. The decision to model the relationship between the two excluding the outliers is easily justified by looking at this figure as those bouts of high inflation and deflation would have been impacted minimally by the central bank's interest rates at the time. In any case, a relationship between interest rates and inflation can be observed with ease. Particularly in the years following 1960, where fluctuations in base interest rates appear to be mirrored in the inflation rate. The relationship between the unemployment rate and inflation was carried out differently to that with the base interest rate; outliers were included as aberrations in inflation rates are usually met with similar shifts in unemployment rate, and excluding these outliers may obstruct the Phillips curve from being potentially uncovered. Once again, a relationship between unemployment and inflation is visible, particularly in the years following 1950 and the years preceding 1915, where shifts in the unemployment rate were met with similar shifts in inflation.

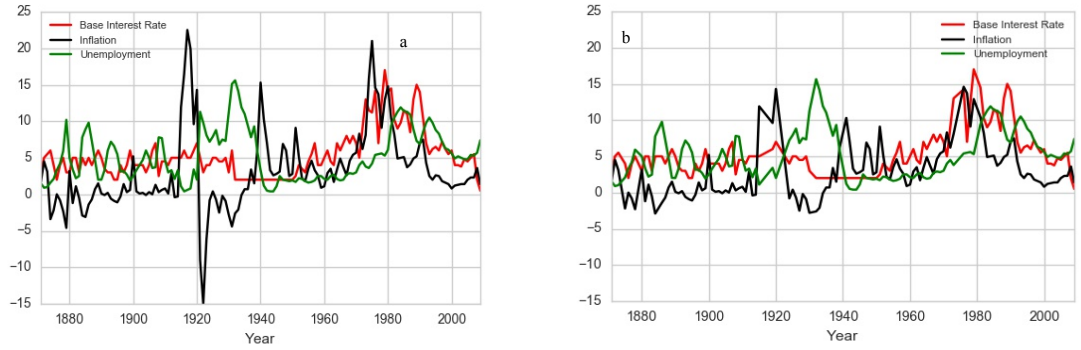

Figure 4: plot of inflation rates, interest rates and unemployment through the years 1871 - 2009: a. includes all the data points and b. excludes those deemed outliers in Section III.

To model the relationships between these variables, first a scatter plot was drawn between the base interest rate and the corresponding inflation rate both for the data set including the outliers, figure 5Figure 5.a, and excluding the outliers, figure 5.b. Similarly, the relationship between the unemployment rate and inflation is modelled using a scatter plot, with a second-order regression-line fit to the data to attempt to model the Phillip's curve in figure 1.

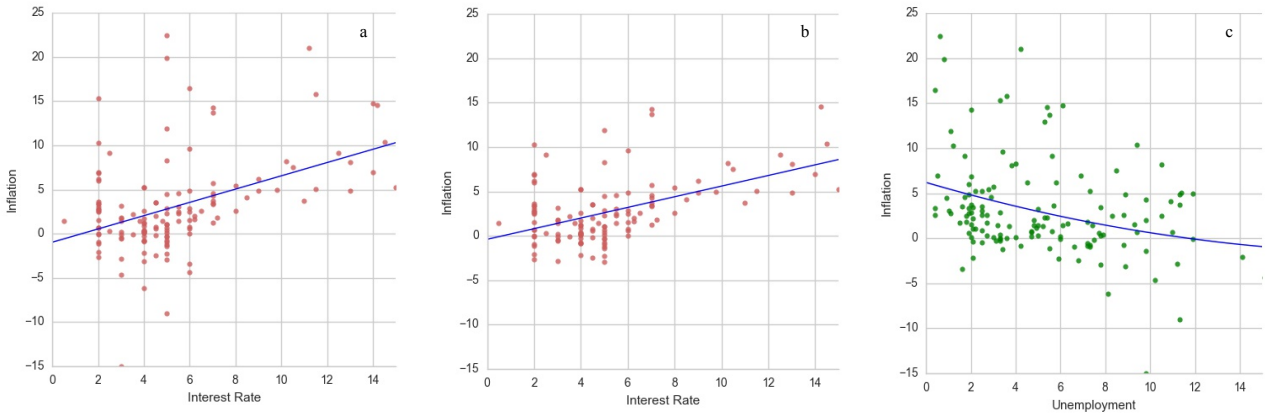

Figure 5: Regression lines fit to the scatter plots where a. includes represents interest rates and inflation relationship including all points, b. shows the same interest rate and inflation relationship but only excluding any outliers and c. shows a weak Phillips Curve on the data of unemployment and inflation

Spearman's coefficient for the data-set in figure 5.a is used to quantify the correlation between interest rates and inflation, as this function is more resilient to data that includes outlier, returning a value of 0.47. Pearson's coefficient is then computed for the data-set excluding the outliers, giving a value of 0.44. Therefore, using the regression analysis returned that there is a present, albeit weak, relationship between the base interest rate and inflation regardless if outliers were considered. A slight second-order regression line was fit on the scatter-plot in Figure 5.c that displayed the relationship between unemployment and inflation. Pearson's coefficient was again calculated as outliers were present, which gave a value of 0.32. A weak inverse relationship was thus identified, but perhaps more interestingly, that data followed a slight Phillips curve like that in literature, shown in figure 1. The coefficient of determination was highest, maybe understandably so, in the regression analysis omitting the outliers. However, all three returned a relatively modest coefficient of determination value between 0.5 and 0.7, meaning the regression lines adequately modelled the relationships between the variables.

Utilising this found relationships of interest rates and unemployment on inflation, a multi-regression analysis was conducted to observe if a relationship exists between all three. With inflation as the dependant variable, a 3D scatter plot was visualised and a multi-regression plane fit to the data, as shown in figure 6. This visualisation helps to articulate the relationship between all three variables, where observations with high interest rates and low unemployment rates correlate to high inflation rates. Conversely, low base interest rates and high unemployment rates correlate to low inflation rates. The coefficient of determination for this regression analysis was 0.36, which is expectedly lower than those discovered in the 2-dimensional analysis in figure 5. However, the effect between both interest rates and unemployment in unison on inflation is still profound. The curse of dimensionality is somewhat visible in this figure though, where the data points may be too sparse in order to draw this regression analysis as a conclusive finding of this relationship.

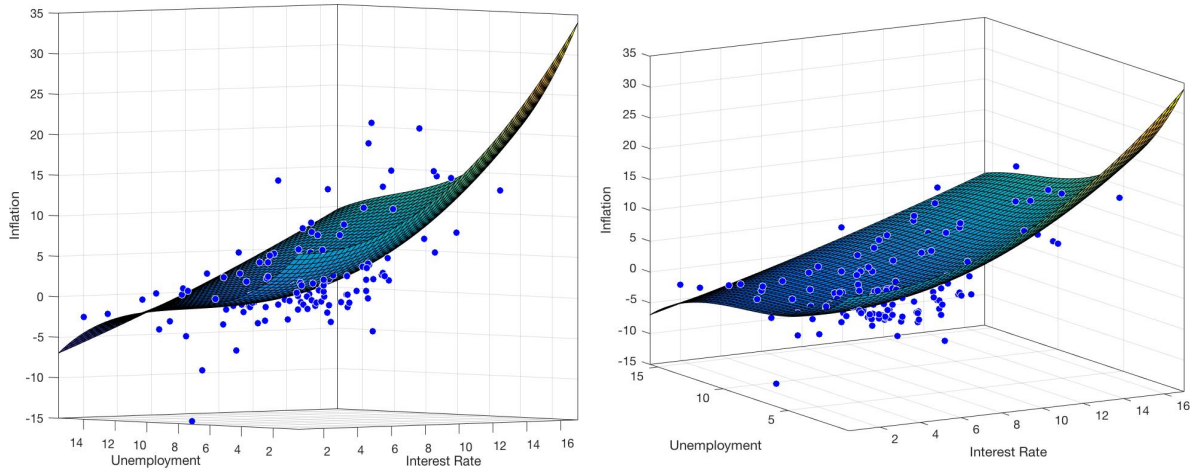

Figure 6: 3D multi-regression where interest rates and unemployment are the independent variables and inflation the dependant. Both visualisations are different angles of the same plot to better enable interpretation.

Following the regression analysis, a simple partitioning clustering task was performed in order to categorise the years in the data based on the similarities in the rates of inflation and unemployment, as these are two economic indicators that are usually reflective on the health of the economy. The results shown in table 3 indicate the yearly instances of each economic climate based on the data covering the years 1871-2009. The results are also visualised on figure 7 with the clustered clearly marked.

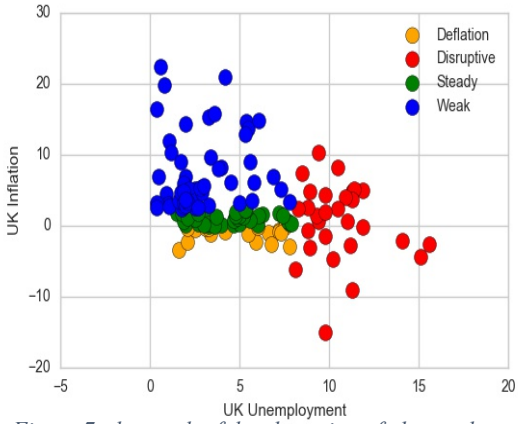

Figure 7: the result of the clustering of observed years based on the unemployment and inflation rates.

Cluster	Observations
Steady	38
Deflation	20
Weak	52
Disruptive	29

Table 3: number of observations in each cluster visualised in figure 7.

All instances in the 'deflation' cluster were in the years preceding 1929 and instances in the 'disruptive' cluster were predominantly in the years between 1981 – 1996, where high unemployment rates were rife. Interestingly, all instances following the political independence of the BoE in 1997 were in the 'steady' economic climate cluster, bar the financial crisis in 2008 in which inflation reached almost double the target rate at 3.6%. This observation may indeed be coincidental, but since the Bank's independence, inflation has been aligned to its target, with a mean of 1.6% and standard deviation of 0.4%. Unemployment has also been at near-optimal levels, with a mean of 5.7% and a standard deviation of 0.9%. Of course there are a multitude of other factors that affect inflation and unemployment, such as currency rates and stock-markets, but in times where the BoEs independence is being questioned[1] this kind of analysis offers a quick indication of the success of the Bank of England's monetary policy on inflation-targeting since their independence.

The change in GDP is then reviewed against the change in unemployment in a similar vein to the relationships earlier and the results are visualised in figure 8 below. With a Pearson's coefficient of -0.7, an inverse correlation is revealed where an increase in unemployment correlates to a decrease in GDP growth. The relationship also ties closely with literature and a coefficient of determination of 0.72, which was greater than those in figure 5. Taking into consideration the Pearson's coefficient and coefficient of determination meant that this was the most prominent relationship explored thus far.

1. https://www.ft.com/content/2616611e-a665-11e6-8b69-02899e8bd9d

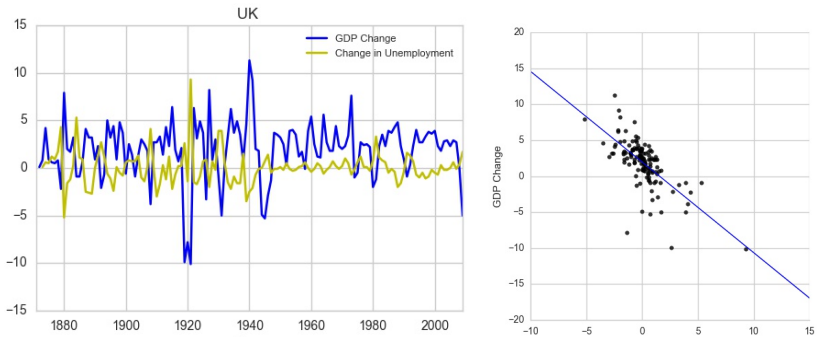

Figure 8:plot of GDP change per year and Unemployment change per year (left) along with a regression analysis of this relationship (right)

Based on the intuition inferred from this data, and some prior knowledge, a simple network graph of the results is produced to visualise the relationship between Monetary policy and the three economic pillars. The results of this network graph are shown in figure 9. This illustrates that altering the interest rates, which is controlled by the Bank of England, directly affects inflation rates, which is also affected by the rate of unemployment. Although this relationship is reciprocated as increasing unemployment decreases inflation rates, hence the undirected edge, and also decreases GDP growth. Although it should be made clear that other factors come into play between these four variables and the relationship between them is more complex than that displayed. This is merely the conclusion drawn from this data science project.

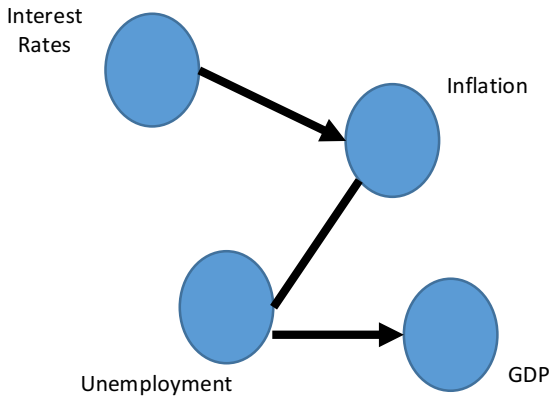

Figure 9: graphical network representation based on the relationships inferred in this project, with Interest Rates, Inflation, Unemployment and GDP as the nodes.

Looking at GDP growth, inflation and unemployment rates gives a perhaps deceivably rosy view of the UK economy over the past 3 decades or so. Since approximately 1980, GDP has grown at an average of 3%, inflation has been contained near its target, with an average of 1.6% and a standard deviation of 0.6%, and unemployment has been steadily declining, as can be observed on figure 4. However, upon closer inspection, the deceit that looking at these statistics alone can be is uncovered by simply using the GDP per capita measurement. This arguably gives a more accurate representation of the economic growth of a country. To break this down, GDP is a measurement of the productivity and output of an economy and is adjusted for inflation. This is the measurement used to primarily assess economic growth by comparing the change on GDP value year-on-year. However, GDP naturally does not directly account for change in population, whereby in some cases GDP growth may in fact not be a true representation of the economic growth as population grew at a greater rate as to negate this growth in GDP.

Figure 10 shows the result of this analysis in a side-by-side bar chart, using the normalised values of GDP and GDP per capita shown in table 2 earlier, where the orange bar indicates the real GDP value per year and the blue bar shows the GDP value per capita. This visual makes it pretty clear that by and large GDP has grown fairly consistently over the past few decades, shown by the orange bar. However, the GDP per capita has grown comparatively slower from the turn of the millennium, shown by the blue bar, and the disparity between the two measurements increases with each year.

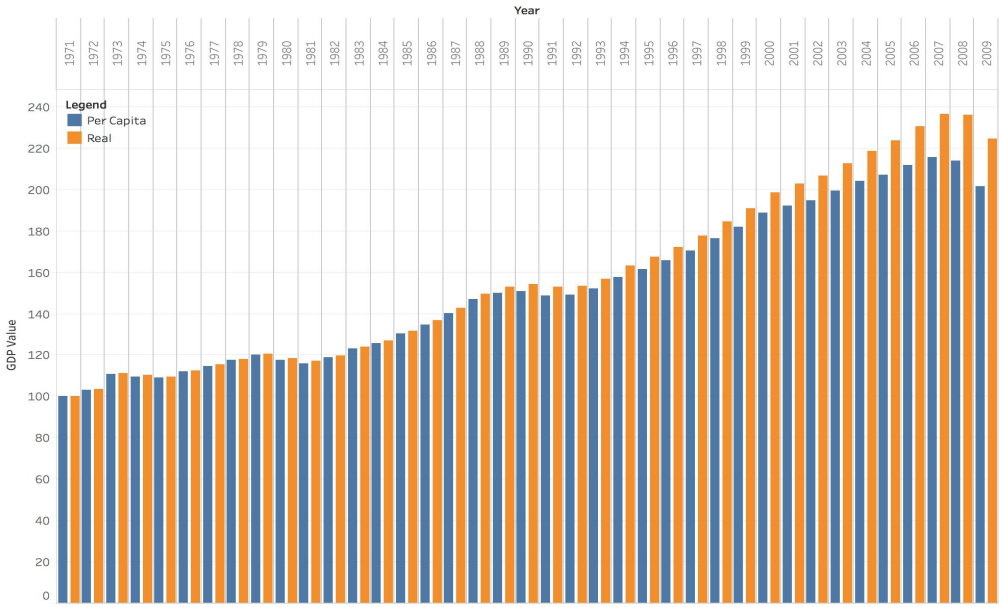

Figure 10: : Real GDP and GDP per Capita represented in side-by-side bars, where 1971 figure is normalised to value of 100

V. Review

In review, a relationship was indeed uncovered between inflation rates and both interest rates and unemployment. Although this is not a ground-breaking discovery, it is beneficial in that it confirms that the UK economy has conformed to literature in how interest rates, inflation, unemployment and GDP are correlated to each other. The 3D multi-regression plot offered an intuitive way to comprehend the trend between three of these variables, where high unemployment and interest rates coincided with high inflation. This multi-regression analysis surprisingly uncovered that the inflation is correlated just as much with both interest rates and unemployment in unison, as separately.

Clustering the economic data points based on the rates of inflation and unemployment revealed perhaps the most striking discovery in this project. Since the Bank's independence in 1997, inflation rates have maintained consistently, within a small tolerance, around their target of 2% and unemployment levels have also been contained within their ideal levels. This was visible in the 'Steady' economic climate cluster in figure 7. This may be coincidental and a plethora of other factors affect inflation and unemployment levels beyond just the interest rates, but it does extract a compelling discovery that perhaps political independence of the central bank has allowed it to better set an effective monetary policy free from political interference. The visualisation of the real GDP and GDP per capita in figure 10 was a more idealistic way to assess the economic growth of the UK for the past few decades. It showed, that although GDP has grown consistently, the growth in population has dampened this increase in GDP. This lag in growth since the turn of the millennium can go some way to explaining that although we have experienced consistent growth, living standards have not been reflective of this growth and the frustrations that has brewed within society that may have possibly led to the Brexit vote is more understandable.

A caveat that must be reiterated is that indeed correlation does not mean causation. For instance, as mentioned in Section V, there are a plethora of other factors that affect these economic measurements that were outside the scope of this project to explore. So in terms of further study, there are many aspects that may be interesting to incorporate within the analysis in this project. For instance, testing the relationship between these economic pillars and other factors such as currency exchange rate, government bond yields, oil prices, FTSE 100 index, et cetera, could offer a more holistic and comprehensive view of economic trends. Following this, the network diagram shown in figure 9 could benefit from being updated and expanded to include any new found relationships. In addition, the data examined in this project could be expanded by analysing quarterly data as opposed to yearly.

Tools

All of the data analysis was conducted through **Python**. From loading and combining the data sets to the clustering task. Most of the plots were produced using the Seaborn extension in Python, Scipy was used to calculate the regression coefficients and coefficients of determination. The 3D multi-regression was visialised using **Matlab** and the side-by-side bar chart using **Tableau**.

References

[1] Bernanke, B.S. and Mishkin, F.S., 1997. *Inflation targeting: a new framework for monetary policy?* (No. w5893). National bureau of economic research.

[2] Phillips, A.W., 1958. The Relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–19571. *economica*, *25*(100), pp.283-299.

[3] Lee, J., 2000. The robustness of Okun's law: Evidence from OECD countries. *Journal of macroeconomics*, *22*(2), pp.331-356.