Google Trends' Ability to Predict Unemployment and Inflation in the World & Testing the Phillips Curve

Audrey Webb, Victor Yoolim Jin, Arion Thai, Mehma Bhasin

Introduction

Google has become the most predominant search engine in much of the world. We decided to test whether or not, due to its popularity and frequency of use, it could measure certain countries' economic factors based on number of searches. One of the foundations of our research project was the use of "Google Trends", which measures the volume of searches throughout different time periods and regions of the world. Our first question is whether Google Trends data accurately reflects the unemployment and inflation rates in the world. We then decided to test an important historical relationship which is whether the short-run Phillips Curve(referred to as "Phillips Curve" below) is accurate about the inverse relationship between inflation and unemployment. The Phillips Curve states that decreased unemployment rate in an economy will correlate with higher rates of inflation.

Data Extraction and Cleanup:

Our world unemployment and inflation data came from data.worldbank.com. Data from this website is detailed and provided usably and efficiency. The Google search data came from google.com/trends, where the data provided is easily accessible and functional. We created a skeleton which created all of the necessary folders where we will be downloading our data into. The Skeleton also includes code to download a zip folder from the World Bank website, unzip and extracts the specific "csv" file we need from the unzipped file. However, for Google Trends, downloading the data directly from their website using R proved too unwieldy, so we downloaded the data manually. Though we were able to obtain the download links from Google Trends, using the download.file function within R to download the "csv" files resulted in corrupt "csv" files, perhaps due to Google Trends' requirement for a web browser when downloading files.

At first, our data focused on the top four countries that Google search "unemployment" the most. These countries are the United States, Canada, South Africa, and Ireland. Then looking at each of these countries' Google search data on the key word "inflation," we found it insufficient for Ireland. Due to the lack of inflation data for Ireland, we chose the sixth country that google searches "unemployment" the most–New Zealand. Nigeria, being the fifth highest country, was also an option, but, similarly to Ireland, we chose not to analyze it due to insufficient and unreliable data. Our unemployment and inflation data now focuses on these four countries - United States, Canada, South Africa, and New Zealand. We limited our data to years 2006 to 2013 as these years contained the most data and would be the most relevant.

We began our project by cleaning our World Bank data on Unemployment and Inflation. Then, we cleaned our Google Trends data. Our Google Trends data had separate comma-separated-value ("csv") files for each country we were interested in, and for each country's employment and inflation rates. In total, we had ten "csv" files with unorganized data to clean and analyze. Our World Bank data listed every country's inflation and unemployment rates from 1960s to 2013. This was more information than we were interested in, so we filtered the rows of the data frame by identifying the country names under the country.names column, and deleted all the columns that were unrelated to the topic. Lastly, we transposed the data frame to make it easier to analyze. The Google Trends data gave us weekly search frequencies of a specific keyword–unemployment and inflation. Due to the data being organized in weekly values, we chose to weight the weekly frequencies to daily values and averaged it out annually for years 2006 to 2013. The code below is the complete function for weighting weekly values to daily values and eventually to annual averages after cleaning Google Trend files.

```
rel_freq <- function(file, data = "data") { #x is the clean data_frame file
  x <- read.csv(file)
  y <- NULL
  for (i in 2006:2013) {
    a <- which(x$from_year == i & x$to_year == i)
    b <- NULL
    for (j in 1:length(a)) {
      b[j] <- x[a[j], data]
    }
    sum1 <- sum(b*7)
    c <- setdiff(which(x$from_year == i | x$to_year == i), a)
    d <- NULL
    for (k in 1:length(c)) {
      if(x[c[k], 'from_year'] == i) {
        d[k] <- (31 - x[c[k], 'from_day'] + 1) * x[c[k], data]
      } else if(x[c[k], 'to_year'] == i) {
        d[k] <- x[c[k], 'from_day'] * x[c[k], data]
      }
    }
    sum2 <- sum(d)
    e <- as.numeric(as.Date(paste(i, 12, 31, sep = '-')) -
                    as.Date(paste(i, 1, 1, sep = '-')) + 1)
    y[i - 2005] <- (sum1 + sum2)/e
  }
  names(y) <- 2006:2013
  return(y)
}
```

The code does the following:
1. First chooses a year from 2006 to 2013 using a 'for loop'.
2. It recognizes the data into two types : one that has the same 'from_year' and 'to_year' (from_year being the year that the week starts from, and to_year being the year the week ends.)
3. For the weeks that have the same 'from_year' and 'to_year', it multiplies the data values by 7 (because we first want daily values, not weekly values. We assume that the weekly value is representative of the value of every day within that week). It then puts all the data into the vector("b")
4. For the weeks that have different 'from_year' and 'to_year', it seeks out one of two cases.
4a. If we're trying to find the annual average of 2005, for example, and the year of interest is in the 'from_year': for example, the google data shows the weekly average from 2005-12-30 to 2006-01-05. Then we know that the last day of the year is the 31st, so we subtract the day in which the week starts(30) from the last day(31), and add 1. Lastly, we multiply that by the weekly data value and store it into a vector("d").
4b. If the year of interest is in the 'to_year': for example, the google data shows the weekly average from 2004-12-29 to 2005-01-04 . We know the first day of the year is 1, and so we subtract the first day(1) from the day of the week(4) and add it by 1. Lastly, we multiply that by the weekly data value and store it into a vector("d").
5. Then after getting the two variables, it takes the sum of the entire vectors to find the year total for that year, and divides it by the number of days within that year.
6. The function then uses the 'for' loop to repeat the process from year 2006 to 2013.

Data Analysis:

In our analysis, we first compared the World Bank unemployment and inflation data to each other. The results show that there is an inverse correlation between inflation and unemployment. We represented this information using data visualization tools. As you can see in Figure 1, the four graphs represent each country we are analyzing and shows the unemployment data on the x-axis and the inflation data on the y-axis. Each point on the graph represents a year's unemployment and inflation rate for each country, and the red line in each plot is the line of best fit for each country (otherwise known as the regression line). Since the slope of this line is negative, it suggests an inverse correlation. From our analysis, you can see that all four countries suggest a negative correlation between inflation and unemployment, thereby confirming that the data we used from the World Bank supports the Phillips curve theory.

Figure 1

Next, we wanted to see how accurately Google Trends data is at predicting the actual unemployment and inflation for each country (which we showed with our World Bank data). We created four graphs (one for each country) comparing the World Bank's actual unemployment and actual inflation data with Google Trends data for the search words "unemployment" and "inflation" for each country, respectively. In our Google Trends and World Bank unemployment comparison graphs we discovered that for each country in our sample there is a very similar overlapping between actual unemployment rates and Google Trends search rates for "unemployment". This phenomenon is astoundingly accurate for the United States graph, probably due to Google's popularity as a search engine here. In Figure 2, the four graphs show years on the x-axis, unemployment data from the World Bank on the left y-axis, and unemployment data from Google Trends on the right y-axis. They are plotted together in one graph in order for easier data visualization in showing the similarity between the World Bank and Google Trends data. Overall, there is great similarity between the two data sets in the graphs, New Zealand's graph shows data dissociation from years 2010 to 2013.

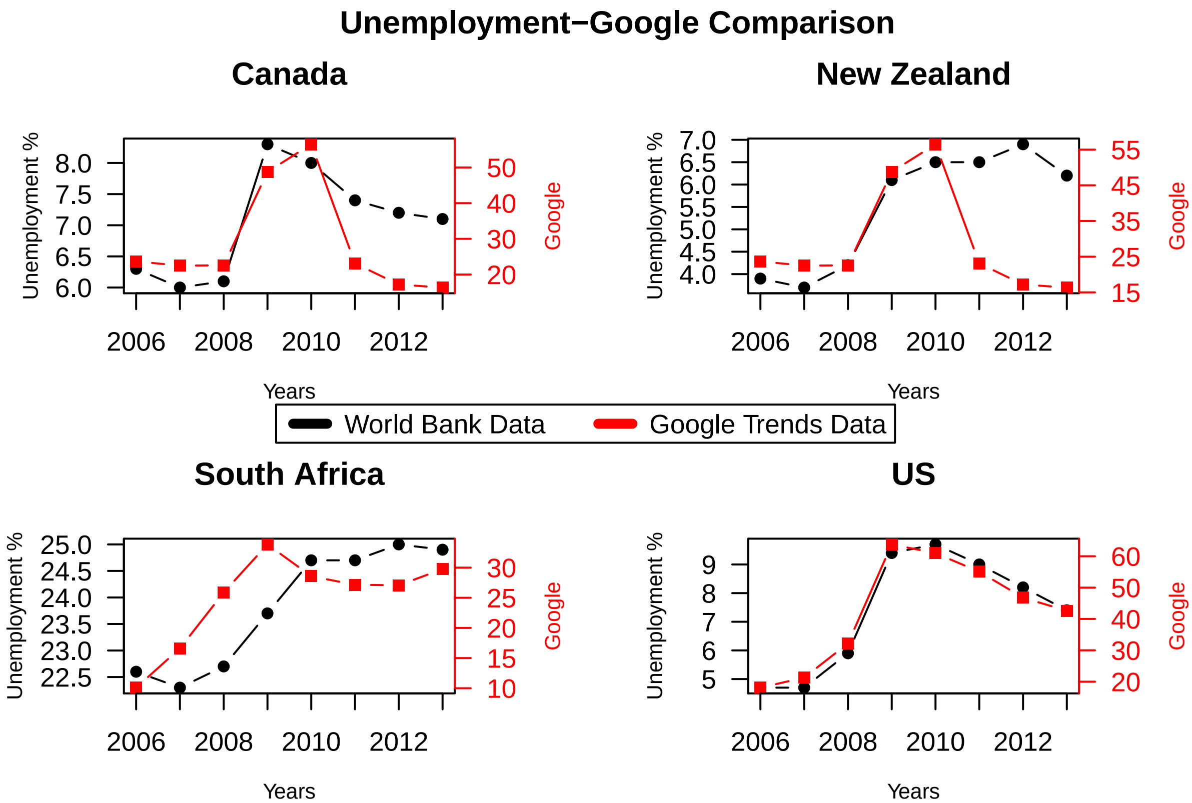

Figure 2

However, for our Google Trends and World Bank inflation comparison graphs we found that Google Trends data did not predict the inflation rate as accurately as it did with unemployment, just by analyzing google trends search for “inflation”. In Figure 3, these four graphs individually show the years on the x-axis, inflation data from the World Bank on the left y-axis, and inflation data from Google Trends on the right y-axis. They are plotted together in one graph in order for easier data visualization in comparing the World Bank and Google Trends data, which is not as similar for inflation as it is with unemployment. Around 2008, when inflation was relatively high for all four countries, the two plots correlate and the red (Google Trends) and black (World Bank) plots somewhat overlap. However, in the years to follow, both plots dissociate from each other.

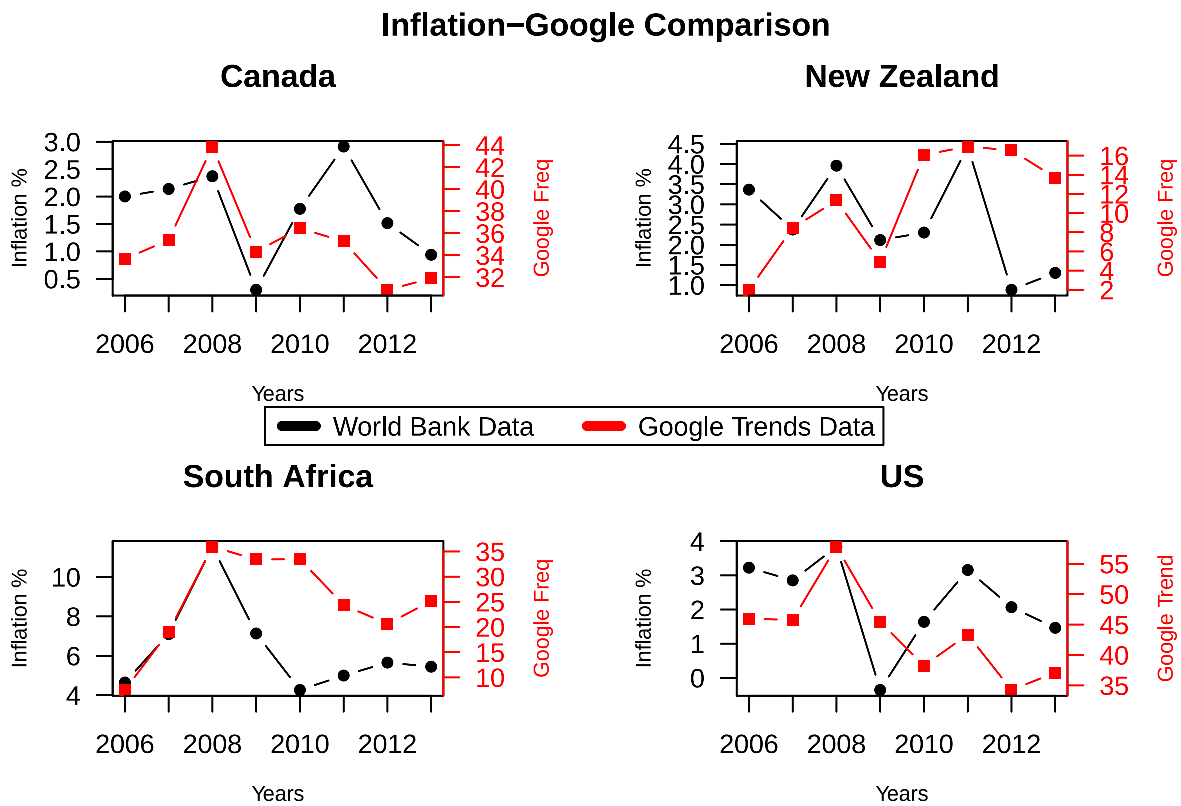

Figure 3

Looking at the graphs individually, we notice that the correlation between Google Trends and World Bank data for the United States does not exist from years 2009 to 2010 and from 2012 to 2013. The same disassociation occurs with Canada from years 2010 to 2011 and 2012 to 2013, South Africa from years 2010 to 2013 (where the correlations were all opposites), and New Zealand from years 2006 to 2007 and 2012 to 2013.

We believe that despite the lack of similarity between the Google Trends and World Bank inflation data, there is a phenomenal connection between these two data sets. We believe the connection extends from years of high unemployment and layoffs causing immediate attention throughout a country therefore leading citizens to go home, go on the popular search engine, Google, and search for jobs. Although inflation does affect the current lives of citizens, they are more likely interested in the consequences of inflation than in inflation itself, hence the low rate of "inflation" keyword Google searches. We supposed people are more likely to search for prices of goods that have increased, rather than just the term "inflation." A tactic we could have used to find more information on inflation is search phrases in instances where people would actually search for relating to inflation. It is not guaranteed that this would provide more Google Trends inflation data for us, but it is another tactic that can be used.

Next, we wanted to compare the inflation and unemployment data from Google Trends, as shown in Figure 4.

Figure 4

By doing such a comparison, we can see whether there is any correlation between the search frequencies of economic key terms related to unemployment and inflation. As shown in the graph, the correlation between unemployment and inflation for Canada, New Zealand, and South Africa are positive, with an especially strong positive correlation in New Zealand and South Africa. This does not reflect the Phillips Curve, and we believe this is due to Google now just becoming a popular search engine in those countries. This is reflected in Appendix C and D, which shows the greatly increasing search frequencies. The United States, being a country that has been using Google for quite a while, is the only country that has a negatively correlated search frequency between inflation and unemployment. Despite the actual unemployment and inflation rates showing a negative correlation, this shows that at least for the three countries, except the United States, the Google Trend data is greatly misleading when showing the relationship between inflation and unemployment. This is not to say that Google search frequencies are not affected by the two economic factors, however, as we have seen a similar trend of searches of keywords related to inflation and unemployment as the actual economy changed.

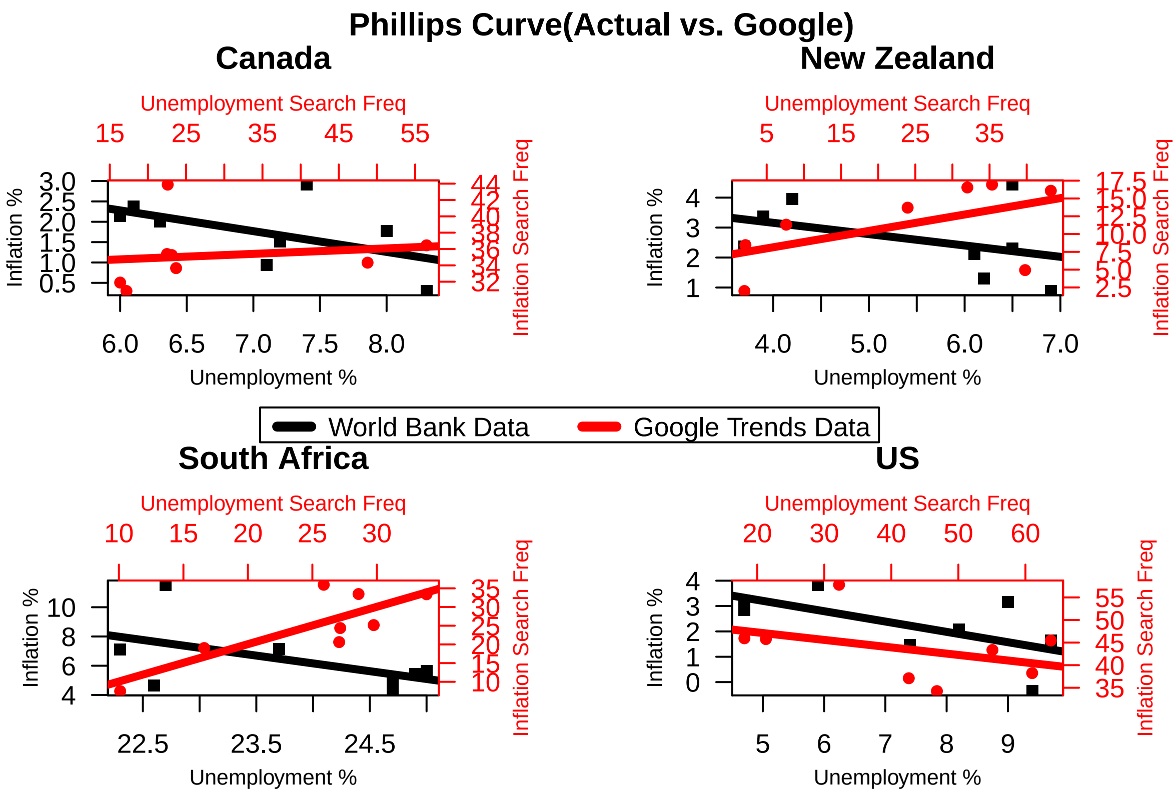

Figure 5
Finally, we created a side-by-side comparison between both the actual Phillips curve we observed using the World Bank and the Google Trends' Phillips curve. The graphs illustrates inflation as the y-axis and unemployment as the x-axis. From here, we see that the World Bank unemployment and inflation rates successfully reflects the Phillips Curve, but the Google Trends data does not.

Conclusion:

We believe that Google Trends data does have the potential to accurately predict economic activities in the world. However, as of now, Google Trends does not have enough elongated data in countries outside of the United States to accurately reflect the unemployment and inflation rates in the world, making Google Trends data unreliable for countries outside of the United States. Our theory behind this is because of Google's infancy and the recent growing popularity of Google (making its way outside of the United States) during the past decade, we don't have enough data to accurately predict economic factors. As for our second question on whether the Phillips Curve is accurate about the inverse relationship between inflation and unemployment, we believe that this is true. Though this historical theory is not supported by the Google Trends data (possibly due to our theory), it is supported by the actual world data obtained from the World Bank. Overall, Google Trends currently does not accurately reflect economic factors for most parts of the world, and the Phillips Curve theory is supported by world economic data.

Appendix

A. World Bank Averages(Unemployment)

Countries	2006	2007	2008	2009	2010	2011	2012	2013
Canada	6.30	6.00	6.10	8.30	8.00	7.40	7.20	7.10
New Zealand	3.90	3.70	4.20	6.10	6.50	6.50	6.90	6.20

Countries	2006	2007	2008	2009	2010	2011	2012	2013
South Africa	22.6	22.3	22.7	23.7	24.7	25.7	25.0	24.9
United States	4.70	4.70	5.90	9.40	9.70	9.00	8.20	7.40

B. World Bank Averages(Inflation)

Countries	2006	2007	2008	2009	2010	2011	2012	2013
Canada	2.00	2.14	2.37	0.30	1.78	2.91	1.51	0.94
New Zealand	3.37	2.38	3.96	2.12	2.30	4.43	0.88	1.30
South Africa	4.64	7.09	11.5	7.13	4.26	5.00	5.65	5.45
United States	3.23	2.85	3.84	-.35	1.64	3.17	2.07	1.46

C. Google Trend Averages(Unemployment)

Countries	2006	2007	2008	2009	2010	2011	2012	2013
Canada	23.7	22.5	22.6	48.8	56.5	23.1	17.2	16.4
New Zealand	2.00	2.10	7.60	39.8	43.2	35.3	32.0	24.0
South Africa	10.1	16.6	25.9	33.9	28.6	27.2	27.1	29.8
United States	18.1	21.3	32.2	63.8	61.0	55.1	46.8	42.6

D. Google Trend Averages(Inflation)

Countries	2006	2007	2008	2009	2010	2011	2012	2013
Canada	33.7	35.4	43.9	34.3	36.5	35.3	30.9	31.9
New Zealand	2.00	8.60	11.6	5.00	16.6	17.4	17.0	13.9
South Africa	7.50	19.0	35.9	33.4	33.5	24.4	20.6	25.2
United States	45.9	45.8	57.8	45.5	38.2	43.4	34.3	37.1

E. Inflation vs. Unemployment Correlation values

Category	Canada	New Zealand	South Africa	United States
World Bank	-0.5317	-0.4040	-0.5206	-0.6252
Google	0.1465	0.5434	0.8157	-0.3964