Project 1

Alfred Zhu

The Relationship Between Interest Rates, Inflation Rates and Productivity Rates

Alfred Zhu azz82

Introduction

As an economics major, I have been taught that lower interest rates are beneficial for businesses as they can afford to borrow more money. As a result of lower interest rates, they can fund more projects, invest in better equipment, and have higher productivity. I want to investigate this by comparing the monthly real output of all workers in the manufacturing sector against both the federal funds rate, the interest rate charged by banks in the federal funds market, and the discount rate, the interest rate charged by the FED to banks. I was also taught that interest rates were used to combat inflation as higher interest rates can be used to combat rapid inflation rates. I also want to see that for myself. For the inflation rate, I will use a dataset containing the consumer price index (CPI) for each month.

For this project, I would like to investigate these questions:

What is the relationship between interest rates and inflation rates?

What is the relationship between interest rates and productivity?

The datasets are all downloaded from FRED

The Federal Funds rate, measured in percent, is sourced from the Federal Reserve

The Discount rate from the US, measured in percent, is sourced from the International Monetary Fund

The Consumer Price Index in the US is sourced from the Organization for Economic Co-operation and Development

The Real Output for all Workers in Manufacturing Sector, measured in percent change at annual rate, is sourced from the Bureau of Labor Statistics

For the datasets that I am using each row represents a month.

```r
# Imports libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(lubridate)
```

```
## Loading required package: timechange
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(readr)

# Imports datasets
fedfunds <- read_csv("/Users/quaza/Documents/SDS322E/Project1Data/FEDFUNDS.csv")# 823 Rows
```

```
## Rows: 823 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (1): FEDFUNDS
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
discountrate <- read_csv("/Users/quaza/Documents/SDS322E/Project1Data/INTDSRUSM193N.csv")# 860 Rows
```

```
## Rows: 860 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (1): INTDSRUSM193N
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
cpi <- read_csv("/Users/quaza/Documents/SDS322E/Project1Data/USACPIALLMINMEI.csv")# 756 Rows
```

```
## Rows: 756 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (1): USACPIALLMINMEI
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
productivity <- read_csv("/Users/quaza/Documents/SDS322E/Project1Data/PRS30006042.csv")# 143 Rows
```

```
## Rows: 143 Columns: 2
## -- Column specification --------------------------------------------------------
```

```
## Delimiter: ","
## dbl  (1): PRS30006042
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Joining/Merging

With all the datasets imported and already tidy, I can join them all together using the inner_join function. Because all the dates are on the first of every month and are all formatted in the same way (YYYY-MM-DD), I can easily join them by the 'DATE' variable without modifying anything.

```r
#Joining/Merging

mydata <- productivity|># 1987/4/1 - 2022/10/1 (143 Rows)
  inner_join(fedfunds, by = 'DATE')|># 1954/7/1 - 2023/1/1 (823 Rows)
  inner_join(discountrate, by = 'DATE')|>#  1950/1/1 - 2021/8/1 (860 Rows)
  inner_join(cpi, by = 'DATE')|># 1960/1/1 - 2022/12/1 (756 Rows)
  mutate(date = ymd(DATE))|>
  # Changes the variable names so that I can actually remember them
  rename(productivity = PRS30006042, fedfunds = FEDFUNDS, discountrate = INTDSRUSM193N)
# 138 Rows after joining

head(mydata)
```

```
## # A tibble: 6 x 6
##   DATE       productivity fedfunds discountrate USACPIALLMINMEI date
##   <date>            <dbl>    <dbl>        <dbl>           <dbl> <date>
## 1 1987-04-01          6.4     6.37          5.5            47.5 1987-04-01
## 2 1987-07-01          6.5     6.58          5.5            48.0 1987-07-01
## 3 1987-10-01          9.8     7.29          6              48.6 1987-10-01
## 4 1988-01-01          0.6     6.83          6              48.8 1988-01-01
## 5 1988-04-01          3.3     6.87          6              49.4 1988-04-01
## 6 1988-07-01          0.6     7.75          6              50.0 1988-07-01
```

With our starting dataset of productivity, we have 143 rows to work with. Joining the dataset with the federal funds dataset did not result in any loss of rows for the productivity dataset as the time frame of the federal funds datset is larger than the productivity dataset. When joining with the discount rate there is a loss of 5 rows. This mainly is because the discount rate dataset does not extend 2021 so no observations in 2022 are listed in the dataset. Because the productivity dataset only records the months for January, April, July, and October, there are not as many rows lost due to the one year time difference between the productivity and discount rate datasets. Because the CPI dataset has a larger time frame that encapsulates the productivity dataset, there is no loss of rows. After all the joins, we are left with 138 rows to work with.

```r
# Finds the dates that are lost
productivity |>
  anti_join(discountrate, by = 'DATE')|>
  head()
```

```
## # A tibble: 5 x 2
##   DATE       PRS30006042
```

```
##   <date>           <dbl>
## 1 2021-10-01         5.9
## 2 2022-01-01         3.7
## 3 2022-04-01         3.3
## 4 2022-07-01        -0.2
## 5 2022-10-01        -2.6
```

The months that are lost are the dates from 2021-10-1 to 2022-10-1. It is a bit of a shame as those months are the period near the end of the quarantine which could have interesting economic insights.

Wrangling

Let's now explore our variables and see some potential relationships before visualization.

The CPI variable by itself does not tell me much however it is essential for calculating the inflation rate which I can more easily interpret. With this dataset, I can calculate the inflation rate between every observation in the mydata dataframe.

```
# Converts the CPI variable to inflation rate and adds it to the mydata dataframe
# Inflation rate = (CPI2-CPI1)/CPI1 * 100

# Sets up the denominator by creating a vector and adding a previous CPI not in mydata.
# The CPI value is from the date 1987-1-1
# This date follows the date pattern in the productivity dataset which goes 1-4-7-10
# removes the last element in CPI as that won't be used in the calculation
denominator<- c(46.91647, mydata$USACPIALLMINMEI[1:137])
# Calculates the inflation rate between the months in mydata
inflation <- ((diff(c(46.91647, mydata$USACPIALLMINMEI)))/denominator)*100
# adds inflation rate to mydata
mydata <- mydata|>
  mutate(inflationrate = inflation)

# We now have all our data in a single nice dataframe!
head(mydata)
```

```
## # A tibble: 6 x 7
##   DATE       productivity fedfunds discountrate USACPIALLMI~1 date       infla~2
##   <date>            <dbl>    <dbl>        <dbl>         <dbl> <date>       <dbl>
## 1 1987-04-01          6.4     6.37          5.5          47.5 1987-04-01   1.35
## 2 1987-07-01          6.5     6.58          5.5          48.0 1987-07-01   0.976
## 3 1987-10-01          9.8     7.29          6            48.6 1987-10-01   1.32
## 4 1988-01-01          0.6     6.83          6            48.8 1988-01-01   0.347
## 5 1988-04-01          3.3     6.87          6            49.4 1988-04-01   1.21
## 6 1988-07-01          0.6     7.75          6            50.0 1988-07-01   1.20
## # ... with abbreviated variable names 1: USACPIALLMINMEI, 2: inflationrate
```

Now that we have an inflation rate variable, we can also create a categorical variable to describe whether the inflation rate is relatively low or high to make things more simple.

Officially, the FED sets the target inflation rate at around 2%, however we are looking at data by the month (or every 3 months) rather than the year which is what the FED looks at. As a result I am going to determine a subjective number that will determine whether the inflation rate is high or not.

I will check for the look at the center of inflation rates to have an idea of what values are high and what values are low.

```r
# Checks for the mean of the inflation rates
mean(mydata$inflationrate)
```

```
## [1] 0.6555716
```

```r
# The data seems to be centered at around 0.66% so we will set that as our subjective determining value

# Creates a categorical variable indicating whether or not the inflation rate is high or low
mydata <- mydata |>
  mutate(inflationcat = ifelse(inflationrate > 0.66, 'high', 'low'))
```

We can now check some summary statistics for our variables

```r
# Makes a table counting the number of high inflation
table(mydata$inflationcat)
```

```
##
## high  low
##   70   68
```

Overall, there are 70 months with high inflation and 68 months with low inflation. This makes sense as I created the variable based on the mean or center of the inflation rate variable

We can also look at the distribution of high and low inflation rates during a specific time period. Let's see the distribution of high and low inflation rates during the COVID-19 pandemic. We saw that there were record inflation numbers during the pandemic so I expect that there are more months with higher inflation.

```r
# Filters for dates during or after 2020 and checks for the distribution of high and low interest rates
mydata2020 <- mydata |>
  filter(date >= ymd('2020-01-01'))

table(mydata2020$inflationcat)
```

```
##
## high  low
##    3    4
```

There are more months with low inflation. There still seems to be a roughly even spread of low and high inflation rates during this period however this may not tell the whole story as it does not show a sense of time.

```r
mydata2020|>
  # We only need to look for the date and inflation variables
  select(date, inflationrate, inflationcat)|>
  arrange(date)|>
  head(7)
```

```
## # A tibble: 7 x 3
##   date       inflationrate inflationcat
##   <date>             <dbl> <chr>
## 1 2020-01-01         0.243 low
```

```
## 2 2020-04-01        -0.613 low
## 3 2020-07-01         1.06  high
## 4 2020-10-01         0.497 low
## 5 2021-01-01         0.459 low
## 6 2021-04-01         2.09  high
## 7 2021-07-01         2.23  high
```

Although our scope is limited, the inflation rate does seem to increase during the later part of the dataset as the interest rate from January to July has increased. The actual percentage of the two latter months is also greater at 2% while the other months are either a bit greater than or lower than 1%.

Let's check on the relationship of interest rates and inflation rate. Are interest rates different when the inflation rate is high or low?

```
# Checks to see if the mean discount rate is different based on inflation rate
mydata |>
  group_by(inflationcat)|>
  summarize(MeanDiscountRate = mean(discountrate))|>
  head()
```

```
## # A tibble: 2 x 2
##   inflationcat MeanDiscountRate
##   <chr>                   <dbl>
## 1 high                     3.79
## 2 low                      2.63
```

```
# Checks to see if the mean federal funds rate is different based on inflation rate
mydata |>
  group_by(inflationcat)|>
  summarize(MeanFedFundRate = mean(fedfunds))|>
  head()
```

```
## # A tibble: 2 x 2
##   inflationcat MeanFedFundRate
##   <chr>                  <dbl>
## 1 high                    3.98
## 2 low                     2.36
```

There does seem to be a difference in interest rates between months that have high and low inflation rates. For the mean discount rate, there is a bit over 1% difference between discount rates during months with high inflation rates and months with low inflation rates with months with high inflation rates also having higher discount rates. The same can be said with the federal funds rate and inflation rate except that the difference between the mean federal funds rates during low and high inflation rates is greater than the difference between the mean discount rates.

Let's also look at the relationship between interest rates and productivity.

```
# Let's first see at a glance if the months with the highest productivity also have low interest rates
mydata |>
  # We are right now focusing only on interest rates and productivity
  select(date, productivity, fedfunds, discountrate)|>
  arrange(desc(productivity))|># sorts by the highest productivity
  head()
```

```
## # A tibble: 6 x 4
##   date       productivity fedfunds discountrate
##   <date>            <dbl>    <dbl>        <dbl>
## 1 2020-07-01         53.1     0.09         0.25
## 2 2010-04-01         11.5     0.2          0.75
## 3 2009-07-01         10.6     0.16         0.5
## 4 1987-10-01          9.8     7.29         6
## 5 1997-10-01          9.5     5.5          5
## 6 1994-10-01          9.4     4.76         4
```

For the productivity variable, the three highest percent change at annual rate all have interest rates lower than 1 percent. This does seem to point towards lower interest rates correlating with higher productivity, but the next highest positive percent changes in productivity all have interest rates greater than 4% which is relatively high

Let's make sure our assumption of high interest rates is correct by checking for the mean interest rates for both the Federal Funds market and the Discount window.

```
# Checks for the mean interest rates
mean(mydata$discountrate)
```

```
## [1] 3.222246
```

```
mean(mydata$fedfunds)
```

```
## [1] 3.18529
```

The average interest rate is a bit over 3% so our assumption that interest rates at and over 4% is relatively high is correct.

Lets also check for the months with the lowest productivity

```
mydata |>
  # We are focusing only on interest rates and productivity
  select(date, productivity, fedfunds, discountrate)|>
  arrange(productivity)|># sorts by the lowest productivity
  head()
```

```
## # A tibble: 6 x 4
##   date       productivity fedfunds discountrate
##   <date>            <dbl>    <dbl>        <dbl>
## 1 2020-04-01        -43.1     0.05         0.25
## 2 2009-01-01        -22.9     0.15         0.5
## 3 2008-10-01        -21       0.97         1.25
## 4 2008-07-01        -13.1     2.01         2.25
## 5 2009-04-01         -8.6     0.15         0.5
## 6 1991-01-01         -8.5     6.91         6.5
```

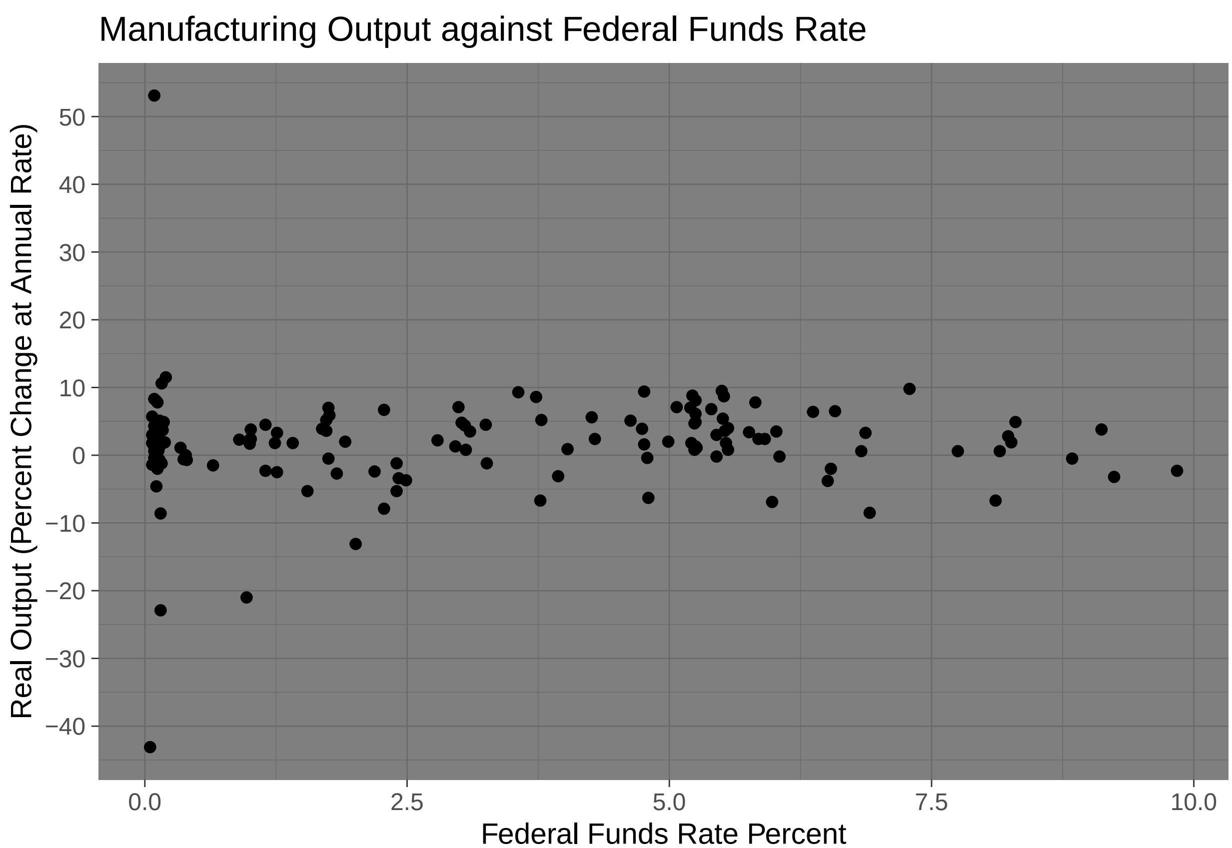
The months with the lowest percent change at annual rate also have relatively low interest rates.

Visualizing

Visualization 1 (2 variables)

Let's Visualize the relationship between productivity and interest rates.

```
# Plots the relationship between productivity and interest rate (federal funds)
mydata|>
  ggplot(aes(x = fedfunds, y = productivity))+
  labs(x = 'Federal Funds Rate Percent',
       title = 'Manufacturing Output against Federal Funds Rate') +
  scale_y_continuous(name = 'Real Output (Percent Change at Annual Rate)', breaks = seq(-50, 50, 10)) +
  geom_point()+
  theme_dark()
```



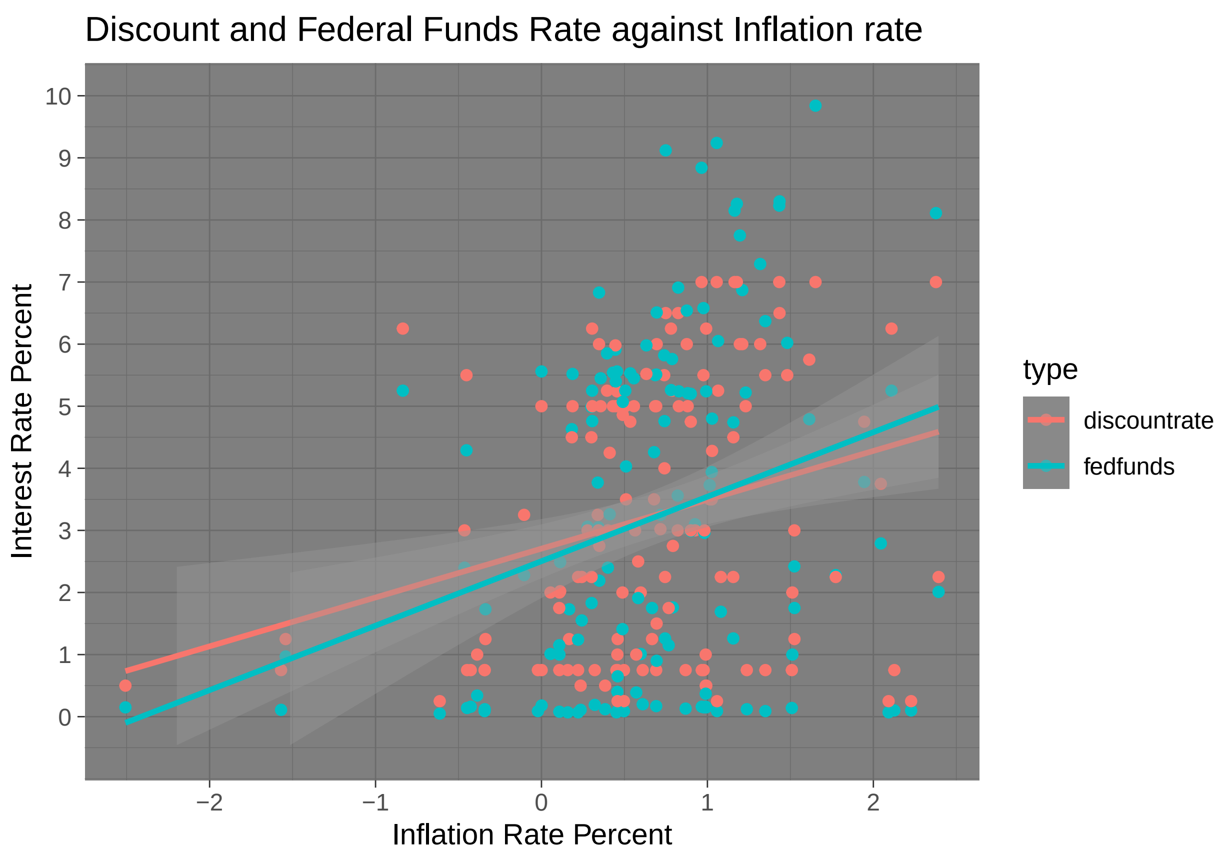
There does not seem to be a clear relationship between the interest rate and productivity. As interest rate increases, the real output in percent change at annual rate stay between -10% and 10%. There are however 2 distinct outliers where the absolute value of the percent change at annual rate is greater than 40%. Those points both happened during the year of 2020 so the cause of those high absolute values may be because of the supply shocks that happened during that year and not because of interest rates. There are also other confounding variables that may be the cause of those outliers. Overall, there does not seem to be a clear relationship between interest rates and productivity.

Visualization 2 (3 Variables)

Let's now visualize the relationship between interest rates and inflation.

```r
# Plots the interest rate against inflation rate
mydata|>
  pivot_longer(cols = c(fedfunds,discountrate), names_to = "type", values_to = "interest_rate")|>
  ggplot(aes(x = inflationrate, y = interest_rate))+
  labs(x = 'Inflation Rate Percent', title = 'Discount and Federal Funds Rate against Inflation rate')+
  scale_y_continuous(name = 'Interest Rate Percent', breaks = seq(0,10,1), limits = c(-0.5,10))+
  geom_point(aes(color = type)) +
  geom_smooth(aes(color = type), method = 'lm')+
  theme_dark()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

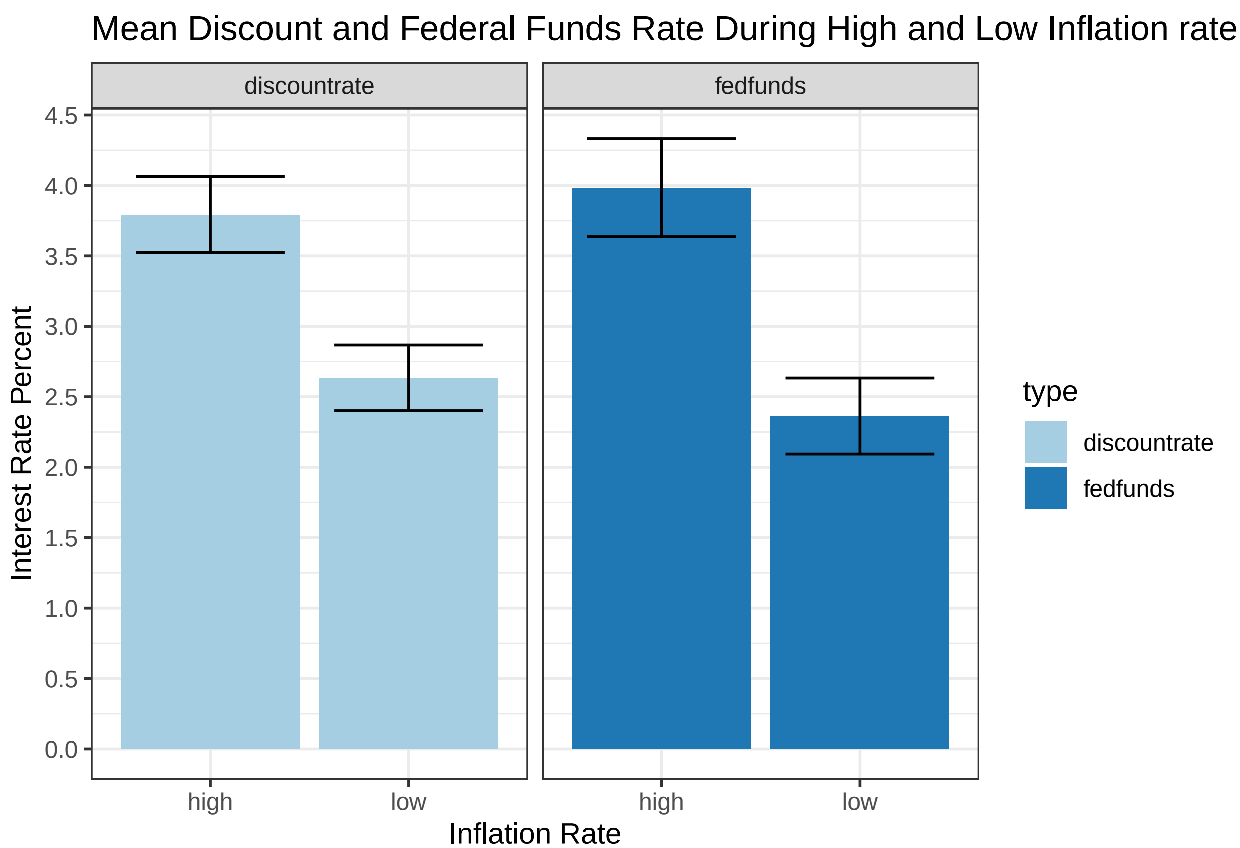


There does seem to be a positive relationship between interest rates for both the discount window and the federal funds market and the inflation rate. There is however a lot of noise between inflation rates at 0% to 1.5% as most of the points are concentrated at that interval. This makes things harder to interpret. There also isn't as clear of a relationship between interest rates and inflation rate at that interval. A reason for this is could be because of the time variable. Inflation rates may climb up before the Federal Reserve adjusts their interest rate and then it falls. What the plot does show is that when inflation rates are low, interest rates also are low. This may be because the Federal Reserve is trying to inject liquidity into the market during an economic slowdown and thus is lowering the interest rate.

Visualization 3 (3 Variables)

We will now plot the mean interest rates during low and high inflation

```r
# Creates a barplot of the mean interest rates for both the discount rate and the federal funds rate
# For both high inflation rates and low interest rates
mydata|>
  pivot_longer(cols = c(fedfunds,discountrate), names_to = "type", values_to = "interest_rate")|>
  ggplot(aes(x = inflationcat, y = interest_rate))+
  labs(x = 'Inflation Rate',
       title = 'Mean Discount and Federal Funds Rate During High and Low Inflation rate')+
  scale_y_continuous(name = 'Interest Rate Percent', breaks = seq(0,5,0.5))+
  geom_bar(aes(fill = type), stat = 'summary', fun = 'mean')+
  geom_errorbar(stat = "summary", fun.data = "mean_se", width = 0.75) +
  facet_wrap(vars(type))+
  scale_fill_brewer("type", palette = "Paired")+
  theme_bw()
```

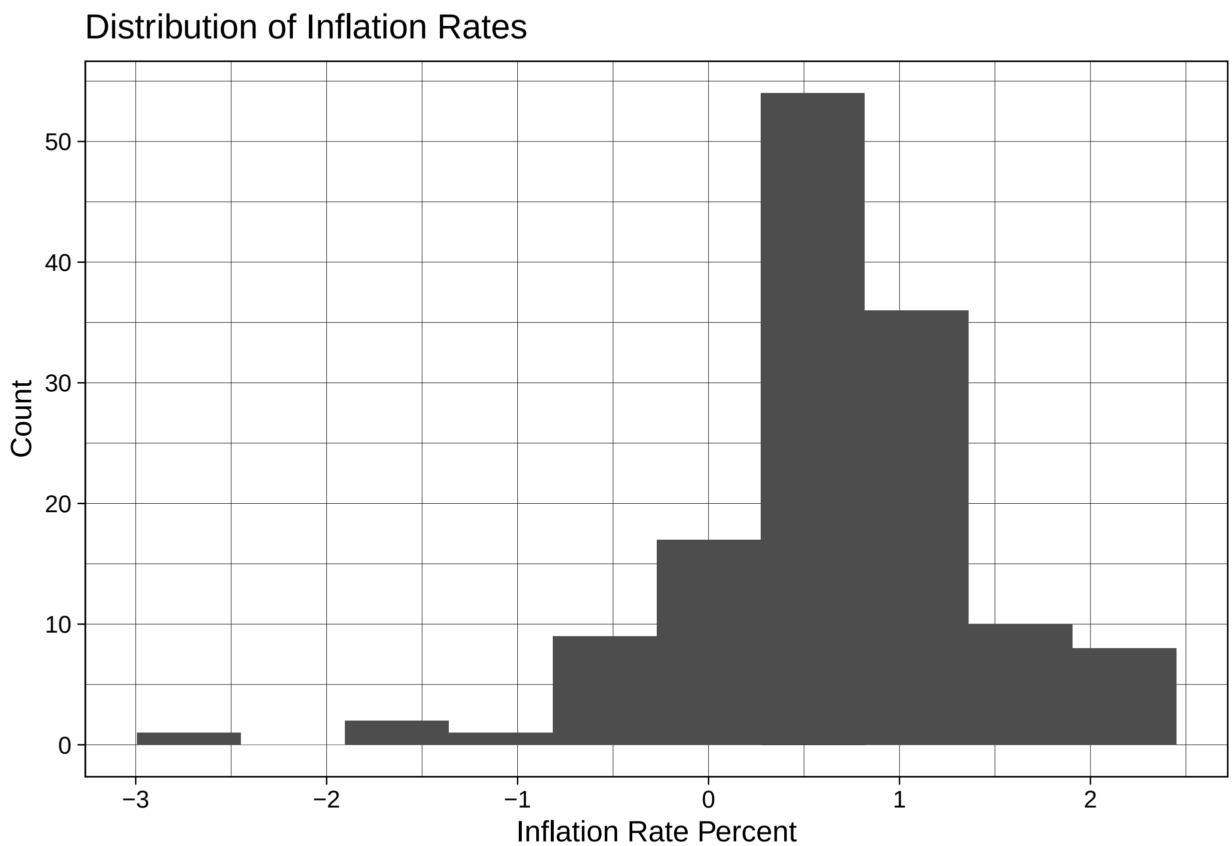


We have previously looked at the mean interest rates grouped by high and low inflation rates and this again shows that the mean interest rates do seem to be different based on whether the inflation rate is high or low. I am also more confident making this observation as the error bars for the mean interest rates between high and low inflation rates do not overlap each other. The federal funds rate do seem to have higher and lower interest rates when compared to the discount rate. This plot points to that high inflation rates results in on average higher interest rates and low inflation rates results in on average lower interest rates.

Visualization 4 (1 Variable)

Lastly, I want to look at the distribution of inflation rates

```r
# Plots a histogram of inflation rates
mydata|>
  ggplot(aes(x = inflationrate))+
  labs(x = 'Inflation Rate Percent',
       title = 'Distribution of Inflation Rates')+
  scale_y_continuous(name = 'Count', breaks = seq(0,60,10))+
  geom_histogram(fill = 'grey30', bins = 10)+
  theme_linedraw()
```



The distribution of inflation rates are roughly symmetrical, however there is a left skew. Most of the inflation rates are centered between 0.5 to 1.5 percent which tells us that the inflation rate every three months should fall between those intervals. The distribution also tells us that although we do see some deflation (inflation rate $< 0\%$), it is not very common and positive inflation rates outweigh them.

Discussion

There does not seem to be a clear relationship between interest rates and productivity. In visualization 1, the distribution of the real output percents remain to be roughly the same as interest rates increase. Furthermore, both the lowest and highest real output percent are at when the interest rate is low. The federal funds rate was at 0.09% when the productivity was at its highest and 0.05% when the productivity is at its lowest. There seems to be some other variable that could affect the real output in percent change at annual rate.

For the relationship between interest rates and inflation rates, there does seem to be a positive relationship between them as can be seen with visualizations 2 and 3. In visualization 2, the linear regression line has a positive slope, and the mean interest rates from visualization 3 are different from each other when grouped by low and high inflation rate. The interest rate when there is high inflation is around the upper 3% while the interest rate when there is low inflation is around the mid 2%. The error bars also do not overlap each other in visualization 3 which makes me confident with the assertion that the mean interest rate is different based on inflation rate, with high inflation resulting in higher interest while low inflation resulting in lower interest. This also seem to show that the Federal Reserve does follow traditional macroeconomics where when inflation is low, interest rates should also be low to help boost the economy and increase liquidity into the market.

Overall the project is somewhat challenging as I have to come up with how I should analyze each variable, and what plots I should make with these variables. It is also hard for me to come up with what exactly I should analyze in the first place as I first thought about analyzing animal data before settling onto economic data. I learned that I should not over think things and just find datasets about topics that I am interested in. This also makes the actual process of data analysis more fun as I could surprise myself with my own findings based on my prior knowledge of the topic.

I would like to thank Professor Brandl for teaching me these macroeconomic concepts.