**Data & Methods – Andrew Vo**

**I, Conceptual/Theoretical Model:**

Thesis: *Low interest rates for long period of time would make recessions more severe (GDP drops at higher rate during recessions)*

Motivation: After looking into the US recessions and interests rate history, I see a trend that when there is a recession, the interest rates would drop. I believe this is a solution from the government for the recession because sudden low interest rates would boost investments and businesses during a hard time like recession. However, if the interest rate is kept low for a long time, US would be stuck with a lot of debt. Also, what’s more important is that when there is a new recession, if the interest rate is already low, then the government cannot have a sudden interest rate drop to gain the same businesses boost. With both of these reasons combined, I believe *long term low interest rate* *would make recessions more severe.* Specifically, I would focus on analyzing the drop rate of GDP during recessions being affected by the drop rate of the interest rate.

**II. Data:**

1. Data Description
2. Data Source: I have combined the data sources from my MetaDataGuide. From the link below, the final combined dataframe that would be used for model is *df\_model*
3. Data process and combination: <https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Data_Process/DataGroup.ipynb>
4. This can be Time Series data since it’s a dataset of US economy Monthly from 1976 to 2010. This can also be a Pooled Cross Section data – when I examine the recessions periods, I can compare the data before and during the recession to see the impact of the interest rates on GDP recessions
5. Unit of analysis: GDP – dependent variable, 568 observations
6. Variable Description Table

 

1. Summary Statistics
2. Table of Summary Statistics





1. Value that is most difficult to believe

I think that the variables are decently reasonable. There are still some interesting points:

* GDP almost has the same trend as population
* Long-term Interest rate is also almost quite similar trend to federal\_funds\_rates
* Some variables seem to indicate recessions quite well (big drop near 2007 – Great Recession) such as employed\_rate and avg\_HPI

1. Report any dummy variables

* No Dummy variables but I may include one called *treatment* that turns on (1) when the interest drop is during the recession and turn off (0) otherwise. Also another dummy variable called *recession* that turns on (1) whenever the timeline is during a recession and (0) otherwise.

1. Any quantitative variables that have little variation – change little, does it affect your analysis?

* For this part, I have included boxplot codes at the bottom of the link <https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Data_Process/DataGroup.ipynb> to check for distribution
* CPI\_Inflation seem to have Small Variation and many outliers. This is expected due to simultaneous nature of consumer price index inflation and the outliers seem to packed to each other so there can be some usable trend there.
* Budget\_on\_education seem to have Small variation as well and some outliers are very high. I should eliminate these outliers for better model prediction

1. Are any of your variables very highly correlated? If so, why? How will this high degree of correlation impact your ability to estimate causal effects?

* I’ve already eliminated some variables that seem to be highly correlated before having the final dataframe such as *laborforce, education/GDP ratio*. I’m hesitating to include the data from *month\_InvestorFlow* file because the time range is quite narrow and start at 2007 – very close to Great Recession and I can’t capture any previous recessions but I can try when I want to purely focus on the Great Recession period.
* For the df\_model data I’ve combined, I believe there can be *some correlation* between the *income limits*: lowest, second, third, fourth and top 5 percent because they seem to have similar trends. But I also want to see how the income variety can affect the GDP so I’ll temporary keep it and test in model

1. Limitations: I believe there would be limitations in terms of limited observations. Since the data are mainly monthly and yearly, multiple data is being repeated for yearly variables. Also, for each year I would only have 12 observations because of the monthly data so for the Pooled Cross Section data purpose, I may need multiple years in the model instead of just 2. I would also run the Time Series model since I would have 568 observations to see the overall effects of variables on GDP through time. I also haven’t included one of my data file *month\_investorFlow* due to limited observations.

**III. Model Specification**

1. Dependent variable: GDP (gdp variable from Data Appendix) – Yearly US GDP
2. log(gdp)=β0+δ0d2+β1treatment+δ1d2∗treatment+β2CPI+β3CPI\_Inflation+β4avgHPI+β5long\_term\_interest+ ... +β15top\_5\_percent
3. All variables from Variable description table will be included with their exact name.

2 new dummy variables: *d2* – during recession period and *treatment* – interest rate drop from previous interest rate

1. Some main coefficients:

|  |  |
| --- | --- |
| Variable | Coefficient |
| D2 | Negative |
| Treatment | Negative |
| CPI | Positive |
| avgHPI | Positive |
| Employed\_rate | Positive |