**Final Research Paper – Andrew Vo**

**I, Empirical Methodology**

1. **Conceptual Framework:**

Thesis: *Lowering interest rate during a recession would make the recession less severe (a sign of lowering interest rate during recession would slightly increase GDP during that period)*

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. I want to test the significant of this theory so I would run a Pooled Cross Section model to compare and signify the effect of interest rate drop before and during the recession period. I would also try to run an extra Time Series model to see the general effect of interest rate drop on GDP.

1. **Data**
2. Data Description
3. 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*
4. Data process and combination: <https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Data_Process/DataGroup.ipynb>
5. 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
6. Unit of analysis: GDP – dependent variable, 568 observations and 288 observations after omitting all empty data points
7. Variable Description Table

Already existed variables:



Variable

Unit

CPI

Number

CPI\_Inflation

Number

avg\_HPI

Number

long\_term\_interest

Percent per year

federal\_funds\_rates

Percent per month

budget\_on\_education

Millions of dollars

gdp

Millions of dollars

Manually created variables:



1. Summary Statistics
2. Table of Summary Statistics (After omitting unused and empty data)





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

* Dummy variables *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: After adjusting the dataset variables at the beginning of this link:

<https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Final_Paper/Pooled_Cross_Section.ipynb>

I can see that a lot of observations have been omitted and I can only focus on 1 recession period which is the Great Recession around 2007. Also, as I go along with my model there are signs of heteroscedasticity as well.

1. **Model Specification and Estimation Strategy**
2. Dependent variable: lgdp (log of gdp variable from Data Appendix) – Yearly US GDP

|  |
| --- |
| log(gdp)=β0+δ0recession+β1treatment+δ1recession∗treatment+β2lfederal\_funds\_rate |

+β3CPI\_Inflation+β4avgHPI+β5long\_term\_interest+ ... +β15top\_5\_percent

1. All variables from Variable description table will be included with their exact name.
2. Notes: I’m using the log of GDP as dependent variable in order to both reduce some heteroscedasticity error as well as focus on the rate of GDP change with log function. The model specified above is specifically for Pooled Cross Section model in which the coefficient for recession\*treatment would identify the effect of interest rate drop during the recession on the change of GDP during that recession period. Thus δ1 is expected to be Positive. The coefficient for recession dummy variable would be negative and for the treatment would be positive because generally the recession would lower the GDP and the sign of interest rate drop would boost the GDP. I have shown the general effect of interest rate on GDP from this link (Heteroskedasticity-Robust statistic included):

<https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Final_Paper/Normal_OLS_Regression.ipynb>

1. Some main coefficients:

|  |  |
| --- | --- |
| Variable | Coefficient |
| Recession | Negative |
| Treatment | Positive |
| Recession \* Treatment | Positive |
| Lfederal\_funds\_rate | Negative |
| avgHPI | Positive |
| Employed\_rate | Positive |

**II, Results**

**Notes:** For three OLS Pooled Cross Section results below, my goal is to estimate the effect of interest-rates drop on GDP during recession. For the last result I have included Heteroskedasticity Robust statistics as well. For more details, please go to link:

<https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Final_Paper/Pooled_Cross_Section.ipynb>

OLS Estimates of the Effect of Interest-rates Drop on GDP Depedent

Variable: Log of yearly GDP 1987-2010

==========================================================================

Dep. Variable: log(GDP) R-squared: **0.182**

Model: OLS Adj. R-squared: 0.173

Method: Least Squares F-statistic: 19.71

Date: Tue, 05 May 2020 Prob (F-statistic): 1.44e-11

Time: 10:44:52 Log-Likelihood: -64.411

No. Observations: 270 AIC: 136.8

Df Residuals: 266 BIC: 151.2

Df Model: 3

Covariance Type: nonrobust

==========================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------

Intercept 15.9988 0.026 618.226 0.000 15.948 16.050

treatment -0.0935 0.039 -2.370 0.019 -0.171 -0.016

recession 0.4935 0.141 3.505 0.001 0.216 0.771

treatment **0.1002** 0.166 **0.604** 0.546 -0.227 0.427

during-recession

Breusch Pagan test for Hetereoskedasticity

'p-value': 9.14209174788533e-06, 'f-value': 9.48340659733261

White test for Hetereoskedasticity

'p-value': 2.1729749135484476e-06, 'f-value': 14.273149528315761

**Description**: The table above is the Pooled Cross Section regression result from regressing the log of yearly GDP on 3 variables: *treatment, recession* and *treatment-during-recession* (the product: *treatment \* recession*). The coefficient for *treatment-during-recession* will indicate the effect of interest-rates drop during recession of GDP change. Here, the coefficient is 0.1 which indicates that if there is interest-rates drop during recession, the GDP is predicted to increase 10%, holding all other variables constant. However, since the t-statistic for such coefficient is only 0.604 and the Adjusted R squared is only 0.182 which means this prediction is insignificant. This can be due to limited number of variables (only 3). Also, there is clearly Heteroskedasticity error since the p-value for 2 heteroskedasticity tests are extremely low.

OLS Estimates of the Effect of Interest-rates Drop on GDP Depedent

Variable: Log of yearly GDP 1987-2010

(All Variables included)

==========================================================================

Dep. Variable: log(GDP) R-squared: 0.999

Model: OLS Adj. R-squared **0.999**

Method: Least Squares F-statistic: 2.611e+04

Date: Tue, 05 May 2020 Prob (F-statistic): 0.00

Time: 10:48:44 Log-Likelihood: 917.53

No. Observations: 270 AIC: -1799.

Df Residuals: 252 BIC: -1734.

Df Model: 17

Covariance Type: HC3

==========================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------

Intercept 15.2896 0.021 718.714 0.000 15.248 15.331

treatment -0.0058 0.001 -4.669 0.000 -0.008 -0.003

recession -0.0371 0.006 -6.378 0.000 -0.049 -0.026

treatment **0.0140** 0.005 **2.814** 0.005 0.004 0.024

-during-recession

CPI\_Inflation 0.0111 0.006 1.940 0.054 -0.000 0.022

avg\_HPI -0.0304 0.012 -2.446 0.015 -0.055 -0.006

long\_term -0.0306 0.008 -3.992 0.000 -0.046 -0.015

\_interest

budget\_on\_

education 0.0434 0.008 5.233 0.000 0.027 0.060

population 0.8046 0.034 23.376 0.000 0.737 0.872

employed\_percent 0.2267 0.027 8.474 0.000 0.174 0.279

unemployed\_percent 0.0467 0.020 2.309 0.022 0.007 0.086

lowest -0.0470 0.013 -3.487 0.001 -0.074 -0.020

second 0.0042 0.016 0.258 0.796 -0.028 0.037

third 0.0200 0.020 1.003 0.317 -0.019 0.059

fourth -0.0831 0.029 -2.852 0.005 -0.140 -0.026

top\_5\_percent 0.1006 0.025 4.012 0.000 0.051 0.150

log(CPI) 0.3471 0.027 12.663 0.000 0.293 0.401

log(federal\_funds -0.0048 0.010 -0.481 0.631 -0.025 0.015

\_rates)

Breusch Pagan test for Hetereoskedasticity

'p-value': 2.675257327527334e-08, 'f-value': 5.1339545777744755

White test for Hetereoskedasticity

'p-value': 0.018436627582135184, 'f-value': 4.069421663493036

**Description**: The table above is the Pooled Cross Section regression result from regressing the log of yearly GDP on 3 variables *treatment, recession*, *treatment-during-recession* and all other variables from the data. The coefficient for *treatment-during-recession* in this case is 0.014 (much smaller than previous result but still positive) which indicates that if there is interest-rates drop during recession, the GDP is predicted to increase 1.4%, holding all other variables constant. In this case, the t-statistic for such coefficient is 2.814 and the Adjusted R squared is very high (0.999) which means this prediction very statistically significant. However, there is still clear signs of heteroskedasticity error. There may be no heteroscedasticity error under 1% significant level for White test, but it seems that the data still clearly has heteroscedasticity

OLS Estimates of the Effect of Interest-rates Drop on GDP Depedent

Variable: Log of yearly GDP 1987-2010

(All Variables included and Hetero Robust Statistics)

==========================================================================

Dep. Variable: log(GDP) R-squared: 0.999

Model: OLS Adj. R-squared **0.999**

Method: Least Squares F-statistic: 2.611e+04

Date: Tue, 05 May 2020 Prob (F-statistic): 0.00

Time: 10:48:44 Log-Likelihood: 917.53

No. Observations: 270 AIC: -1799.

Df Residuals: 252 BIC: -1734.

Df Model: 17

Covariance Type: nonrobust

==========================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------

Intercept 15.2896 0.025 616.078 0.000 15.241 15.338

treatment -0.0058 0.001 -4.266 0.000 -0.008 -0.003

recession -0.0371 0.006 -6.645 0.000 -0.048 -0.026

treatment during **0.0140** 0.004 **3.633** 0.000 0.006 0.022

recession

CPI\_Inflation 0.0111 0.006 1.791 0.074 -0.001 0.023

avg\_HPI -0.0304 0.013 -2.414 0.016 -0.055 -0.006

long\_term\_

interest -0.0306 0.007 -4.322 0.000 -0.044 -0.017

budget\_on\_

education 0.0434 0.008 5.785 0.000 0.029 0.058

population 0.8046 0.043 18.729 0.000 0.720 0.889

employed\_percent 0.2267 0.028 8.143 0.000 0.172 0.281

unemployed\_

percent 0.0467 0.024 1.976 0.049 0.000 0.093

lowest -0.0470 0.012 -3.964 0.000 -0.070 -0.024

second 0.0042 0.017 0.254 0.799 -0.029 0.037

third 0.0200 0.020 0.991 0.323 -0.020 0.060

fourth -0.0831 0.029 -2.819 0.005 -0.141 -0.025

top\_5\_percent 0.1006 0.027 3.768 0.000 0.048 0.153

lCPI 0.3471 0.030 11.480 0.000 0.288 0.407

lfederal\_funds\_

rates -0.0048 0.012 -0.408 0.683 -0.028 0.018

==========================================================================

Warnings:

[1] Standard Errors are heteroscedasticity robust (HC3)

**Description**: The table above is the Pooled Cross Section regression result from regressing the log of yearly GDP on 3 variables *treatment, recession*, *treatment-during-recession* and all other variables from the data under **heteroscedasticity-robust standard errors**. The coefficient for *treatment-during-recession* in this case is still 0.014. Moreover, the t-statistic for such coefficient has increased to 3.633 which means this prediction is even more statistically significant with heteroskedasticity robust t statistics.

**III, Discussion and Conclusion**

**Conclude:** From the above Pooled Cross Section models, I can conclude that it’s statistically significant under heteroscedasticity robust statistics for GDP to increase during Recession thanks to a decrease of interest rates. In particular, if there is a sign of interests rates decrease during recession, holding all other variables constant GDP can increase 1.4%.

**Discussion:**

I have chosen the topic of my project to be predicting the effect of interest-rates drop on GDP changes. My initial thoughts were a bit different, I want to predict more in specific of how much interest-rates has to drop for an effect on GDP increase instead of just having a signal of interest-rates decrease during recession of no matter what amount. However, by seeing the positive coefficient with statistically significant, I can still conclude that the recession can be less extreme when there is interest-rate drop.

I want to make this prediction because by March of 2020, president Trump has declared the interest-rates drop from 2.25% to nearly 0% during this Corona Virus pandemic (technically a recession). I believe that my model can conclusion can be applicable for policy changes during economy hardships like Trump’s decision and it would be further expressed if US GDP can benefit from this decision.

I’m just a bit concern that the effect of such interest-rate drop of recession GDP is quite big (1.4%) and the R Squared is quite too high (0.999) which may indicate some variable bias in my data even though I constantly have tests for Hetereoskadasticity. Also, there can be some time lag effect of interest rate drop on GDP because generally the effect is not always sudden. However I haven’t figured out how to include time lag effect in the python model and this can be further included in future.

Besides this model for policy change on interest-rates during recession, I also contruct normal OLS regression for GDP all variables to gasp the general effects of interest-rates as well as other variables on GDP overall. My models can be further improved if it can specifically predict the amount of interest-rate changed needed for the GDP to benefit.

**Links to my models:**

Pooled Cross Section Model: <https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Final_Paper/Pooled_Cross_Section.ipynb>

* Normal OLS Regression: <https://gitlab.com/hiep.vo/econ342_RecessionUs/-/blob/master/Final_Paper/Normal_OLS_Regression.ipynb>