Effects on the United States’ Unemployment Rate Change

Nathan Rosel, Emily Nguyen, Mavis Chi, Evelyn Kern

Professor Xiahua (Anny) Wei, Ph.D

BBUS 382: Introduction to Econometrics - Spring 2018

**Abstract**

This research focuses on the unemployment rate across the United States in the span of 30 years, starting from 1987 up until 2016. Our group researched variables that affected the unemployment rate during this period of time. To analyze the data, we performed multivariate regression analysis to explain the models we created. Additionally, we performed correlation testing, t-test, and F-test to validate the significance of our variables we utilized during this research. For our first model, we concluded that “Time” and the “Significant Event”, represented by the 2008 financial crisis both had significant effects on the unemployment rate in the United States. For our second model, it was determined that the change in annual personal expenditures, the change in annual average single household income, and the change in gross domestic products were significant.

Table of Contents:

Abstract……………………………………………………………………………………………….……....1

Introduction…………………………………………………………………………………………....……3

Purpose…………………………………………………………………………………………….….……....4

Research:

Data Selection and Description…………………………………….……………….…….5

Regression Model………………………………………………….………….……...5

Variables………………………………………………………………….……….…..5-6

Descriptive Statistics……………………………………………………………………..……….....…..7

Data Discussion and Model Specification…………………………………………………….....8

OLS Assumption…………………………………………………………………………..…………….8-9

Data Analysis……………………………………………………………………………...……………9-15

Conclusion…………………………………………………………………….………...……………….....15

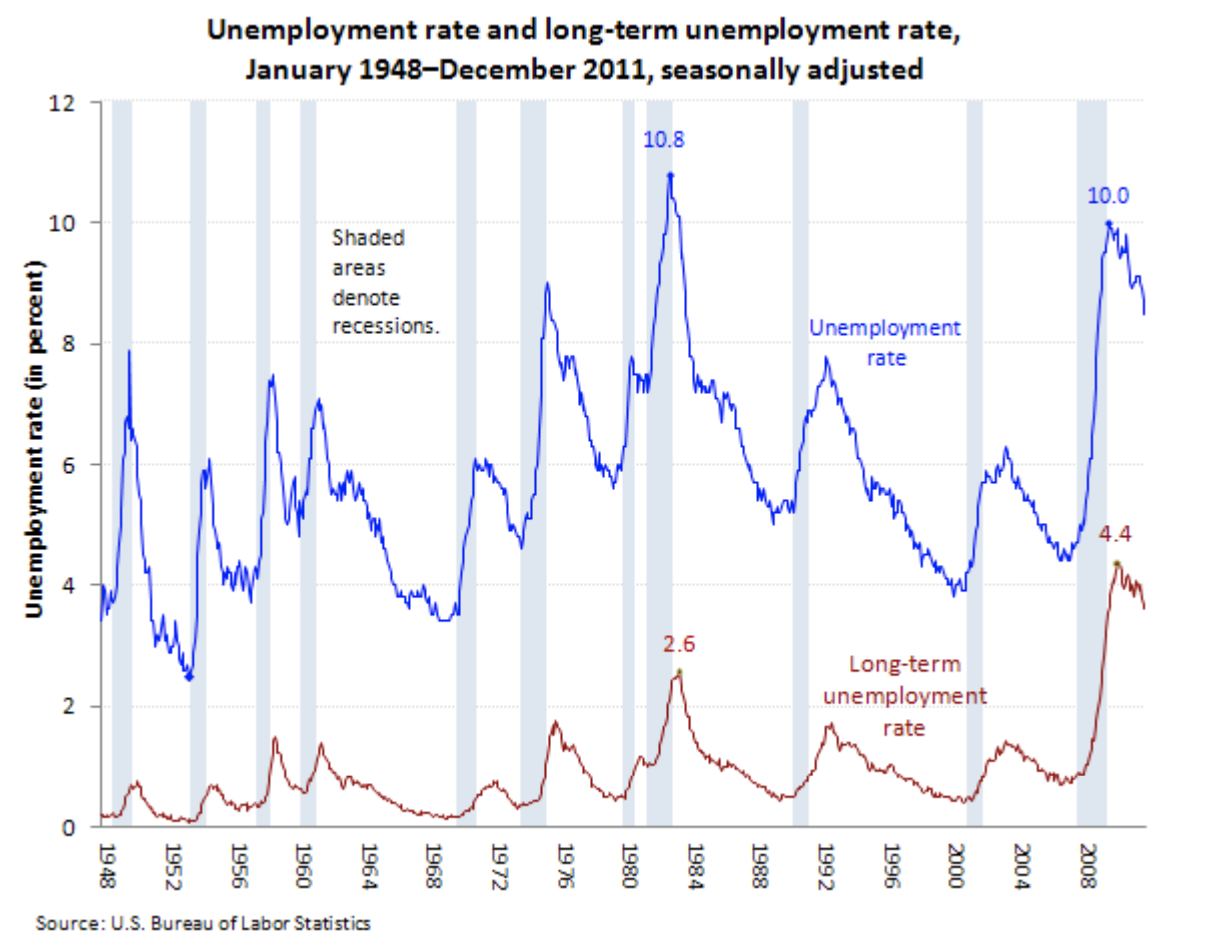
Reference………………………………………………………………….………………………………...16

**Introduction**

Unemployment rates have been volatile throughout the years we have studied. We have reason to believe the failures in our economy have been damaging to the employment rate. The start of the war in the Middle East accompanied by the spikes in oil prices contributed to the recession in the 1990’s. Unemployment rates skyrocketed to a whopping 7.8%, while private employments decreased to 1.9%. Results from failures in financial communities such as banks and business, created a downfall in the employment rates.

As the years went on, another recession sprung up from the Y2K (Year 2000 Problem). The computer and software industry began to boom in the early 2000’s, causing companies to flourish while stock prices started to increase. Towards the end of year, computer purchases started to decline and as a result, businesses started to go bankrupt. The federal reserve did not pay much attention to these events and continued to increase their credit rates. Consequently, stock prices worsened and the economy crashed. Unemployment rates rose to 5.7% which was above the natural unemployment rate. This major event set the stage for the next one to come -- the Great Recession.

The Great Recession in 2008 was caused by banks loaning to borrowers with poor credit scores. Most of these borrowers defaulted on their mortgages, resulting in a multitude of banks drowning in their debt. Due to this, the economy went downhill, as unemployment gradually rose to 9.5%. By October 2009 it increased an extra 0.5%, ending at a 10%. As we looked at these recessions, it was obvious that they impacted the rates considerably. We wanted to take a look into depth about what other factors would tie into the unemployment rate in the US for the past 30 years.



***Table 1-1: Unemployment Rate and Long-Term Unemployment Rate - United States***

**Purpose**

Since the unemployment rate change has been very volatile throughout the years, being affected by several financial and economic events, and had a significant effect on the lives of many citizens, our group wanted to research the other variables that may have affected the unemployment rate change as well. Our original assumption was the unemployment rate per year will increase as GDP decreases, inflation rate increases, personal expenditures increase, and when household income decreases.

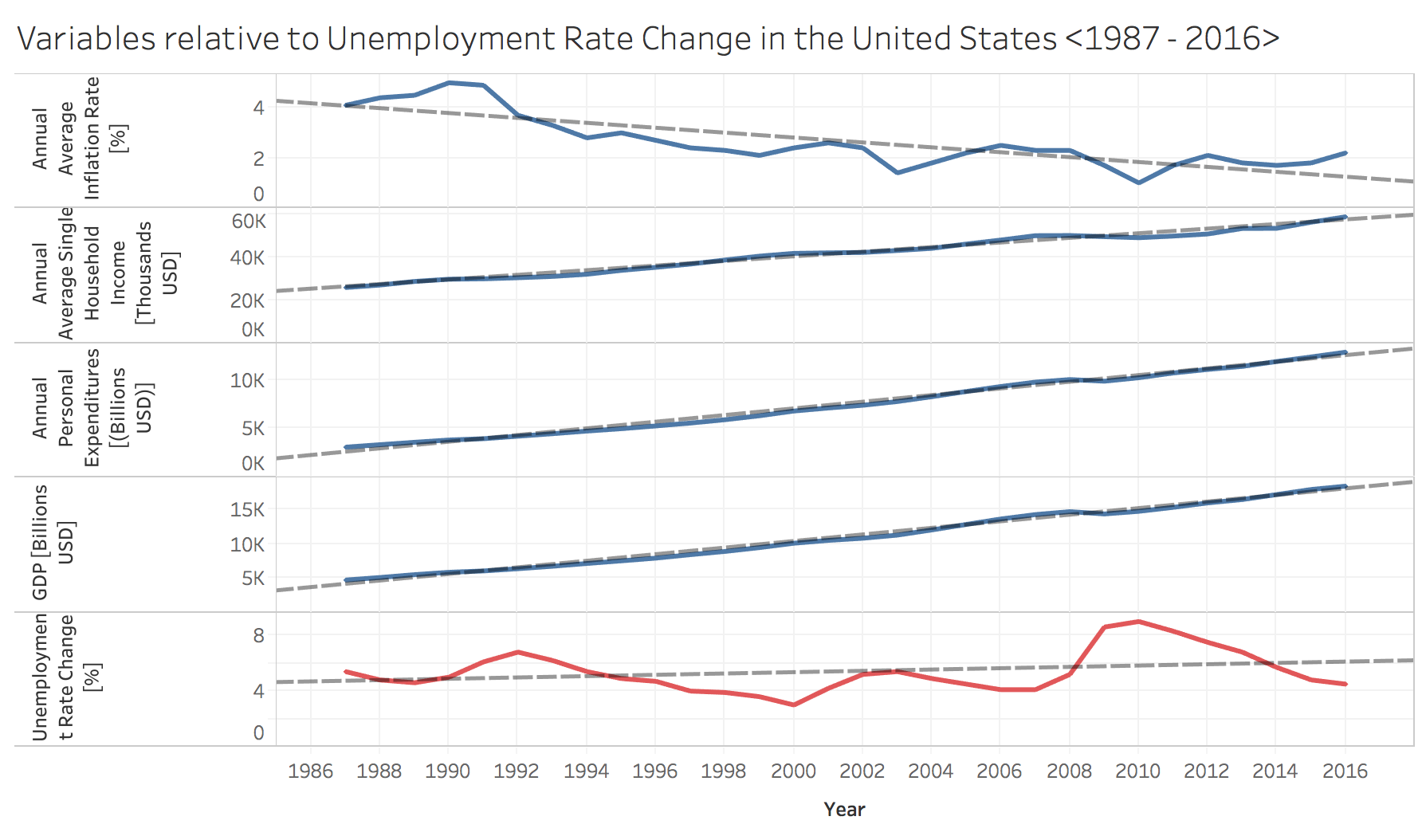
**Data Selection and Description**

The focus on the unemployment rate was primarily centered around whether our explanatory variables time (**T**), annual GDP (**GDP**), annual average inflation rate (**InflR**), annual average single household income (**Inc**), annual person expenditure (**PerExp**), and/or significant event, based on the 2008 subprime mortgage backed securities crisis, (**SigEvent**) had an affected on the unemployment rate, in the US, in the past 30 years. Our sample size consisted of the years between 1987 and up until 2016. The model that we have includes all of the following listed above, as well as, an error term () to accomodate for possible omitted variables. Below you will see the general form of the regression along with the descriptions of the explanatory variables:

***(*UnEmply = β0 + β1T+ β2GDP + β3InflR + β4Inc + β5PerExp + β6SigEvent + ui*)***

|  |  |  |
| --- | --- | --- |
|  | **Variable** | **Description** |
| 1 | UnEmply | The change in the United State’s unemployment rate per year (in%) |
| 2 | T | The change in time per year |
| 3 | GDP | United States’ Gross Domestic Products (in billions USD) |
| 4 | InfR | Annual average inflation rate per year (in %) |
| 5 | Inc | Annual average single household income per year (in thousands USD) |
| 6 | PerExp | Annual personal expenditures per year (in billions USD) |
| 7 | SigEvent | Indicator = 1, if year is greater than 2008 and 0, otherwise |

***Table 1-2: Variables and Descriptions***

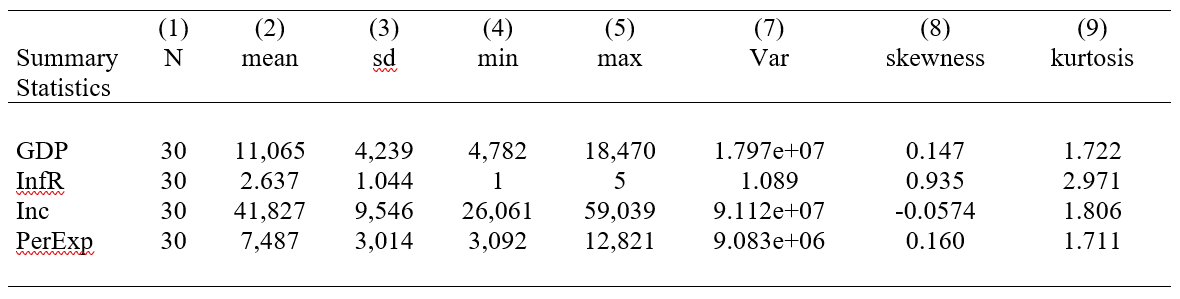
****

***Table 1-3: Variables Relative to Unemployment Rate Change in the United States[[1]](#footnote-1)***

The graph above utilizes the data we have researched all compiled into a graphical visualization. Prior to performing a regression analysis, we can see the overall trend of our regressors on the dependent variable, which is the Unemployment Rate Change in the United States. As shown in the graph above, the Unemployment Rate Change in the United States makes a gradual increase over the period of 30 years. In comparison with the four regressors during this period of time, the United States’ GDP, Annual Personal Expenditures, and Annual Average Single Household Income increase. However, the Annual Average Inflation Rate decreases over the period of 30 years from 1987 to 2016 which proves the Phillips Curve theory. It was a concept that was developed by A. W. Phillips that states that the inflation rate and the unemployment rate has an inverse relationship. When there is economic growth, there comes more inflation which leads to more jobs and unemployment and vice versa.

The method behind including these specified variables are due to their likelihood of interrelationship with the dependant variable, unemployment rate. This study’s intention is to arrive at an explanation for the original research question, what factors have affected the US unemployment rate for the past 30 years? In the last 30 years, we saw multiple spikes in the US unemployment rate that may, or may not have contributed from numerous factors. Our main objective is to identify which factors, when coming into contact with our regression model affects this rate.

**Descriptive Statistics**

****

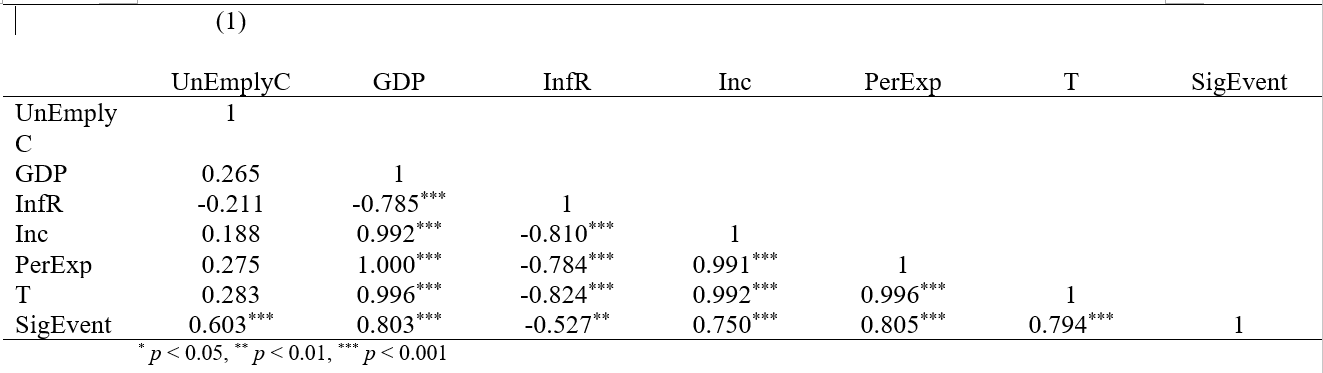
***Table 1-4: Descriptive Statistics***

Shown in the table above, are the variables for GDP, Annual Average Inflation Rate, Annual Single Household Income, and Annual Personal Expenditures summarized. We can note how all of the variables are not perfectly skewed or have a perfect kurtosis. The average household income has a skewness of -0.0574 which indicates that it is almost symmetrical unlike the United States inflation rate with a skewness of .935 which is almost highly skewed to the left. The inflation rates kurtosis (2.971) indicates that it is almost a normal distribution because it is close to 3, but the other variables are platykurtic which indicated it is less than 3.

**Data Discussion and Model Specification**

As shown in ***Table 1-2***, we have included two binary variables to account for time series and the 2008 financial crisis ***(SigEvent)*** that may have affected the unemployment rate change during the period of 30 years. Our decision to include these two variables.

For our model, we will be focusing on using a 95% confidence interval to determine the significance of our data.



***Table 1-5: Correlation Matrix***

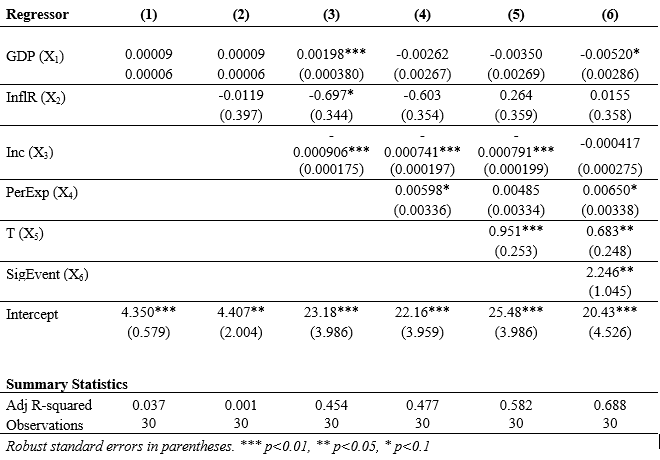
The gross domestic product, average household income, personal expenditures, time in years all had positive weak correlations with the unemployment rate change. The inflation rate had a negative weak correlation. The significant events variable had a moderately positive correlation. Paired independent variables in our model are average household income; personal expenditures and time in years; Significant Events. Gross Domestic Products and Personal expenditures had a perfect correlation of 1.00 which proves perfect multicollinearity. Other variables in the correlation table had significantly high correlations but were not exactly 1.

**OLS Assumptions**

1. The dependent variable y can be calculated as a linear function of a specific set of independent variables plus an error term. ✓
2. Our sample consists of n-paired observations that are drawn randomly from the population. The number of observations is greater than the number of parameters to be estimated ( n > k ). The independent variables are nonstochastic whose values are fixed ✓
3. The mean of the error terms has an expected value of 0 given values for the independent variables ✓
4. There is no perfect multicollinearity
   1. In our correlation table we correlated our independent variables against the dependent variable. Significant Events and T, and personal expenditures and average household income should be correlated against each other to be tested for perfect multicollinearity
5. The error terms all have the same variance and are not correlated with each other
   1. Our yearly data of unemployment rates suffered a bit from autocorrelation because the next year depended on unemployment from the year before. This caused error terms in different observations are correlated with each other

**Data Analysis**

1. **Original regression Model**

****

***Table 1-7: Original Regression Model***

We created several regression models to determine the relationship between the regressors and dependent variables. The model is accounted for heteroskedasticity in the errors of the regression. As shown in the graph above, the sixth model proved to be the most significant out of all of the other regression models. We decided to use the adjusted R-squared to measure our models because of our relatively small sample size, although the sample is considered to have a “normal distribution”, by having at least 30 samples. The sixth model has an adjusted R-squared of 0.688, meaning that about 68.8% of the change in unemployment rate is explained by the regressors we have chosen in our model.

As we include more variables to alleviate the omitted variable bias (OVB), the table shows by including more variables, the more significant our regression model becomes. An analysis on the effect of the regressors on the dependent variable is as follows:

* *As national GDP increases by $1 USD per year, then the unemployment percent will be reduced by 0.00520%*
* *As the annual inflation rate increases by 1% per year, then the unemployment change will increase by 0.0155%*
* *For every dollar (USD) increase in the annual average single household income, unemployment rate will decrease by 0.000417%*
* *For every dollar (USD) increase in the annual average personal expenditures, unemployment will increase by 0.00650%*
* *As time passes year-by-year, the unemployment rate will increase by 0.683%*
* *As a result of the 2008 subprime mortgage backed securities financial crisis, the unemployment rate will increase by 2.246% each year*

**Outcome**

Overall, the first five regressions all had small t-values which did not explain the dependent variable. When we created a new variable in STATA called SigEvent (for significant events in the years that are greater than 2008) and T (Time), the t-value became significant at a 1.96 t-stat. Our SigEvent was 2.15 and the T t-value was 2.75. To validate the significance of these two variables, we performed a joint hypothesis-test with a significance level of 95%.

*H0: SigEvent = 0 and T = 0*

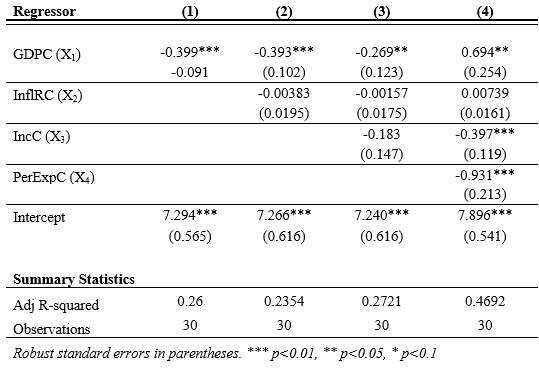
*HA: SigEvent ≠ 0 or T ≠ 0*

After calculating the F-statistic, it had a value of 5.56. Under the 95% confidence level, it has a p-value of 0.0107 < 0.05. Therefore, we reject the null hypothesis and that both variables are statistically significant from zero, which meant both time and 2008 financial crisis affect the unemployment rate change. We believe that the 2008 subprime mortgage crisis affected unemployment rates significantly in that time because from 2007 to 2008, the rate went from 4.4% to 10%. Additionally, as time passed year-by-year, there also could be various unpredicted events that may significantly affect the unemployment rate in the United States.

Surprisingly, the annual average inflation rate and annual personal expenditures did not have a significant effect on the unemployment rate change - as the annual personal expenditures had a t-value of -1.5164 while the annual average inflation rate had a t-value of 0.0433. One would assume that these variables would affect the unemployment rate by increasing it. If individuals spend more per year, then they are more likely to become unemployed. If the inflation rate increases per year, then it would increase the price of goods and services, making goods and services more expensive, thus increasing the likelihood of an increase in unemployment. Under the 95% confidence interval, annual personal expenditures was not statistically significant in our model. However, under the 90% confidence interval, the annual personal expenditures would be a significant variable to the regression.

**b) Second regression model**

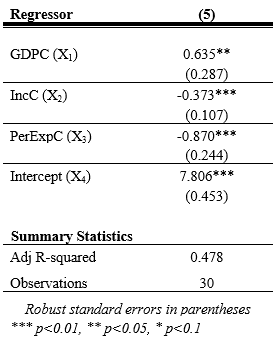
In the original regression model, we found that the only two significant variables were time and significant events. However, the other variables did have a slight effect on the change in unemployment rate. To better understand the relationship between the GDP, inflation rate, annual single household income and annual personal expenditure, we performed a second multiple regression with a change in the unit for these variables. In this regression, we changed all the units of the regressors into the rate of change (%); omitted the time and significant event as they yielded the same result as the original model. The result from this new model provided different outcomes. The change in GDP, annual single household income, annual personal expenditure have become more significant, but the change in the inflation rate still remained insignificant. This is shown in the table below.



***Table 1-8: Second Regression Model***

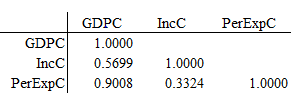
The first three regressions indicated that a positive change in GDP, inflation rate and annual single household income resulted in a decrease in unemployment rate. However, the first three models had less than 30% chance that the change in unemployment rate could be explained by the change of the regressors in the first three models. In the last model, there was a higher chance (46.92%) that the change in the regressors could explain the change in the unemployment rate. Nevertheless, the result from the last regression contradicted with a small part in our research theory, which was, “As unemployment increases, there is a drop in the GDP and inflation rate”. To better understand the meaning of this, we ran two more tests.

First, we noticed that no matter how many regressions we ran, the inflation rate or change in the inflation rate remained insignificant. We eliminated this variable in the next multiple regression. The result is shown in the table below.



***Table 1-9: Regression without Inflation Rate***

Similar to the fourth model, this model had almost a 50% chance that the unemployment rate could be explained by the regressors. However, the change in GDP still contradicted with our research theory, then we realized that there must have been a high correlation between these three regressors. So, we tested this theory and we were correct as it is shown in the correlation table below.



***Table 1-10: Second Correlation Matrix***

Next, we had another theory about the relationship between the change in annual single household income and the change in annual personal expenditure. We assumed that these two regressors yielded the same result for the change in unemployment rate that translated to a 1% increase in the change in annual income and the change in annual personal expenditure would result in the same percentage decrease in the change of unemployment rate. To test this theory, we performed a joint hypothesis testing as follows:

*Ho: IncC = 0 AND PerExpC = 0*

*Ha: IncC ≠ 0 OR PerExpC ≠ 0*

With a p-value (0.0021) less than a 5% level of significance, we concluded that the change in the annual single household income and the change personal expenditure do not yield the same result for the change in the unemployment rate.

*\*\*\*Note: We use the regression in model (5) for this hypothesis testing.*

In this second model, we concluded that when we manipulated the data by transforming them into the same unit, it gave us different outcomes. The significance level of this model is similar to the original model as they had the same range for the adjusted R-squared. The second model provided us with more insight information about this topic. We learn that there are many ways we can analyze the data and how to interpret the result from the findings. Moreover, to completely understand a set of data, it required constant testing using different techniques and each finding births a new theory.

**Conclusion**

In conclusion, our research found that both of our models interpreted the effect on unemployment rate in the United States differently. This was based on the variables we utilized for each model and whether they were significant or not. Overall, our group was able to collect data on the unemployment rate in the United States, followed by data on our regressors. One of our challenges during this research project was that not all of our data were updated and consistent. We did not have data for 2017 for some of our variables, so we had to find a common ending point to which we should end our time horizon. Another major challenge we faced is that our sample size was too small. Our data only consisted of 30 samples, which is still considered as a normal distribution, however it was still considerably small and did not fit the Central Limit Theorem[[2]](#footnote-2). This is because there is limited data to which we can access as we go back in years. If we were to do this research project again, we would make sure that we have a sizable sample size so that our data would be more normally distributed and there would be less errors in our data. This research project allowed us not only to apply the concepts we learned in class, but also to perform an in-depth statistical and economical analysis on the unemployment rate in the United States.

**References**

Amadeo, K. (n.d.). The Great Recession of 2008: What Happened, and When? Retrieved May 2, 2018, from https://www.thebalance.com/the-great-recession-of-2008-explanation-with-dates-4056832

Bureau of Labor Statistics Data. (n.d.). Retrieved May 15, 2018, from https://data.bls.gov/timeseries/CUUR0000SA0L1E?output\_view=pct\_12mths

History.com Staff. (2017). Great Recession Timeline. Retrieved May 2, 2018, from https://www.history.com/topics/great-recession-timeline

Karlgaard, R. (2012, July 16). The Not So Great Recession... Retrieved May 2, 2018, from https://www.forbes.com/forbes/2010/0329/opinions-rich-karlgaard-great-recession-digital-rules.html#5630066e50b7

U.S. Office of Management and Budget, Gross Domestic Product [FYGDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FYGDP, June 5, 2018.

U.S. Bureau of the Census, Median Household Income in the United States [MEHOINUSA646N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MEHOINUSA646N, June 5, 2018.

U.S. Bureau of Labor Statistics, Unemployment Rate: 20 years and over [LNS14000024], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS14000024, June 5, 2018.

1. *Graph was created through Tableau using the data we acquired during research*  [↑](#footnote-ref-1)
2. *Central Limit Theorem (CLT) states: when the sample size approaches infinity, the sample mean will be normally distributed* [↑](#footnote-ref-2)