

# FE-511-B: Final Project

Sean Trinh

November 30, 2019

## 1 Problem Description

**Introduction.** The state of modern economies is measured in gross domestic product (GDP), or the total value of final goods produced and services provided in a country during one year. In other words, GDP is an aggregate calculation of a country's consumer spending, business investment, government spending, and net exports in a given time period. An important distinction to note is that nominal GDP is evaluated at current market prices, which means that its value is affected by inflation or deflation. Real GDP, on the other hand, is not.

In the United States, there are two major stock indexes: the S&P 500 and the Dow Jones Industrial Average (DJIA). The S&P 500 is an index that includes 500 companies, spanning all sorts of industries, and is based on each individual company's market capitalization, meaning companies with higher market capitalization will have a bigger impact on the calculation of the index than companies with lower market capitalization. The DJIA on the other hand includes thirty companies, spanning some of the major industries, and is based on each individual company's stock price.

Economists and investors alike are interested in finding ways to better track and predict the state of the US economy. For example, economists have done a lot of research into the yield curves of the US fixed income market. Since the 1970s, very time the yield curves in the United States have inverted, meaning yields on short term bonds have risen above the yields on long term bonds, the US has entered a recession shortly after.

This project will seek to explore the validity of using either the S&P 500 or the DJIA or both in tracking the state, health, and growth of the US economy. As the S&P 500 and DJIA have much more frequent and regular reportings than US GDP, which is reported on a quarterly basis, it may be beneficial to economists and investors to keep track of the more frequent reportings of both indexes in order to predict whether they should be optimistic, or bullish, on the economy or pessimistic, or bearish, on the economy. In other words, does positive growth in the S&P 500 and/or DJIA indexes necessarily predict a positive growth in US GDP, and vice-versa? If there exists a positive correlation between the indexes and US GDP, do the indexes overstate or understate how well the economy is doing?

**Thesis.** The S&P 500 and DJIA keep track of what each index deems as the top 500 and top 30 companies in the United States, respectively. Because of this, each index may be prone to overstating how well the overall US economy is doing, because neither is keeping track of smaller companies that may be performing poorly compared to the tracked, top companies. However, since

these companies in theory make up a large and significant portion of the economy, it should still be the case that positive growth in these indexes should predict positive growth in US GDP.

Because of this, positive growth in the S&P 500 and DJIA indexes predict a positive growth in US GDP, and vice-versa, but are prone to overstating how well the economy is doing.

## 2 Data and Methodology

**Data.** The data that will be used to conduct the regressions and analyses includes quarterly data on the nominal US GDP, quarterly S&P 500 data, and quarterly DJIA data. Although S&P 500 and DJIA data is reported more frequently than quarterly, quarterly data must be used in order to do proper regressions on nominal US GDP, which is only reported quarterly. Also, since the data for nominal US GDP only goes back to the 1940s, all sets of data will be scoped to start at the earliest quarterly reporting of nominal US GDP data and end at the most recent quarterly reporting.

The Excel spreadsheet used to conduct this analysis contains all of the data described and can be found here [https://github.com/seantrinh/fe\\_511\\_final\\_project/blob/master/fe\\_511\\_data.xlsx](https://github.com/seantrinh/fe_511_final_project/blob/master/fe_511_data.xlsx).

**Methodology.** Three regressions to test the hypothesis will be done, all of which will be done using the programming language, R. First, a linear regression will be done to determine the relationship between the S&P 500 and US Nominal GDP. Standard regression results will be reported and analyzed, and a regression equation like the following will be produced:

$$USNominalGDP_t = \alpha * S\&P500_t + \epsilon$$

A second linear regression will be done to determine the relationship between the DJIA and US Nominal GDP. Standard regression results will be reported and analyzed, and a regression equation like the following will be produced:

$$USNominalGDP_t = \alpha * DJIA_t + \epsilon$$

Finally, a multivariate linear regression will be done to determine the relationship between the S&P 500 and DJIA and US Nominal GDP. Standard regression results will be reported and analyzed, and a regression equation like the following will be produced:

$$USNominalGDP_t = \alpha * S\&P500_t + \beta * DJIA_t + \epsilon$$

Using the regression results and equations produced from these three regressions, an analysis will be conducted, and a conclusion on the hypothesis will be made.

Common studies on indicators of GDP growth will lag US GDP in response to a variable. In other words, regressions sometimes try to see how a change in variable affects GDP a number of quarters later. An example regression would look like the following:

$$USNominalGDP_{t+n} = \alpha * S\&P500_t + \beta * DJIA_t + \epsilon$$

In the above equations, t represents the variables at a given time. In only the above equation does n represent a number of quarters after the given time, hence lagging. Lagging will not be

done with these regressions, because the idea of this study and hypothesis is to see if either or both of the indexes are able to simulatenously track the direction of GDP so that economists and investors can estimate or predict the quarter's GDP before it is reported. In other words, the idea of this study is not to see how changes in either index affect GDP in future quarters. In theory, GDP is strongly correlated to both indexes and mostly made up of the value produced by and performances of companies in those indexes in a specific, given quarter, not future or previous quarters.

### 3 Regression, Code, and Results

#### S&P 500 and US Nominal GDP

**Code.** The following R code was used to load the quarterly data for the S&P 500 and US Nominal GDP and run a linear regression on the data to produce a regression equation and summary. The correlation coefficient between the two variables was also produced.

```
1 library("readxl")
2 my_data <- as.data.frame(read_excel("Documents/Classes/FE/fe511/fe_511_final_
  project/fe_511_data.xlsx"))
3 us_nominal_gdp <- my_data[, "us_nominal_gdp"]
4 sp_500 <- my_data[, "sp_500"]
5 regression_sp_500 <- lm(us_nominal_gdp ~ sp_500)
6 summary(regression_sp_500) #Summary of the regression
7 cor(us_nominal_gdp, sp_500) #Correlation coefficient between the two variables
```

**Regression.** Figure 1 outlines the results of the linear regression on US Nominal GDP and the S&P 500. Additionally, the correlation coefficient between the two variables was reported to be **0.9582932**.

The regression equation, based on the results, is as follows:

$$USNominalGDP_t = 8.3843 * S\&P500_t + 1340.9416$$

#### DJIA and US Nominal GDP

**Code.** The following R code was used to load the quarterly data for the DJIA and US Nominal GDP and run a linear regression on the data to produce a regression equation and summary. The correlation coefficient between the two variables was also produced.

```
1 library("readxl")
2 my_data <- as.data.frame(read_excel("Documents/Classes/FE/fe511/fe_511_final_
  project/fe_511_data.xlsx"))
3 us_nominal_gdp <- my_data[, "us_nominal_gdp"]
4 djia <- my_data[, "djia"]
5 regression_djia <- lm(us_nominal_gdp ~ djia)
6 summary(regression_djia) #Summary of the regression
7 cor(us_nominal_gdp, djia) #Correlation coefficient between the two variables
```

```

Call:
lm(formula = us_nominal_gdp ~ sp_500)

Residuals:
    Min       1Q   Median       3Q      Max
-5023.0 -1223.7  -362.9   1338.6   6364.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1340.9416    134.3904   9.978  <2e-16 ***
sp_500        8.3843     0.1471  57.004  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1786 on 289 degrees of freedom
Multiple R-squared:  0.9183, Adjusted R-squared:  0.918
F-statistic: 3249 on 1 and 289 DF, p-value: < 2.2e-16

```

Figure 1: Regression Results - S&P 500.

**Regression.** Figure 2 outlines the results of the linear regression on US Nominal GDP and the DJIA. Additionally, the correlation coefficient between the two variables was reported to be **0.9603199**.

The regression equation, based on the results, is as follows:

$$USNominalGDP_t = 0.9468 * DJIA_t + 1397$$

## DJIA, S&P 500, and US Nominal GDP

**Code.** The following R code was used to load the quarterly data for the S&P 500, DJIA, and US Nominal GDP and run a multivariate linear regression on the data to produce a regression equation and summary.

```

1 library("readxl")
2 my_data <- as.data.frame(read_excel("Documents/Classes/FE/fe511/fe_511-final-
   project/fe_511_data.xlsx"))
3 us_nominal_gdp <- my_data[, "us_nominal_gdp"]
4 sp_500 <- my_data[, "sp_500"]
5 djia <- my_data[, "djia"]
6 regression_both <- lm(us_nominal_gdp ~ sp_500 + djia)
7 summary(regression_both) #Summary of the regression

```

**Multivariate Regression.** Figure 3 outlines the results of the linear regression on US Nominal GDP and the S&P 500 and DJIA.

The regression equation, based on the results, is as follows:

```

Call:
lm(formula = us_nominal_gdp ~ djia)

Residuals:
    Min       1Q   Median       3Q      Max
-5696.9 -1315.4  -380.3   1411.3   5794.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.397e+03  1.305e+02   10.71  <2e-16 ***
djia         9.468e-01  1.617e-02   58.53  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1743 on 289 degrees of freedom
Multiple R-squared:  0.9222, Adjusted R-squared:  0.9219
F-statistic: 3426 on 1 and 289 DF, p-value: < 2.2e-16

```

Figure 2: Regression Results - DJIA.

$$USNominalGDP_{t+n} = 0.6461 * S\&P500_t + 0.8742 * DJIA_t + 1390.5869$$

## 4 Discussion

**Observations.** Test. Test. Put some observations here.

**Conclusion.** Test. Test. Either reject the hypothesis or fail to reject the hypothesis here.

**Further Steps.** Test. Test. Put down some thoughts as to what can be done next to further test the hypothesis/main idea.

```

Call:
lm(formula = us_nominal_gdp ~ sp_500 + djia)

Residuals:
    Min       1Q   Median       3Q      Max
-5652.1 -1309.4  -373.9   1425.8   5837.1

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1390.5869    132.0020   10.535 < 2e-16 ***
sp_500        0.6461     2.0370    0.317 0.751354
djia         0.8742     0.2295    3.808 0.000171 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1746 on 288 degrees of freedom
Multiple R-squared:  0.9222, Adjusted R-squared:  0.9217
F-statistic: 1708 on 2 and 288 DF,  p-value: < 2.2e-16

```

Figure 3: Multivariate Regression Results - S&P 500 and DJIA.