**Big Data Essay**

**Relationship between the Price-Earnings Ratio, Price-To-Book Ratio, Dividend Yield, Earnings per share, Payout ratio and Stock Price of Nifty 500 Index – Evidence from Indian Stock Market**

Table of Contents

[1. Introduction 3](#_Toc34972011)

[2. Literature Review 4](#_Toc34972012)

[2.1. Introduction 4](#_Toc34972013)

[2.2. Correlation Analysis 4](#_Toc34972014)

[2.3. Ordinary Least Squares (OLS) Regression 5](#_Toc34972015)

[2.4. Bayesian Linear Regression 6](#_Toc34972016)

[3. Data 7](#_Toc34972017)

[4. Methodology 9](#_Toc34972018)

[4.1. Regression Model 10](#_Toc34972019)

[4.2. Hypothesis 11](#_Toc34972020)

[5. Empirical Results 12](#_Toc34972021)

[5.1. Summary Statistics 12](#_Toc34972022)

[5.2. Changes Over Time 13](#_Toc34972023)

[5.3. Correlation Analysis 15](#_Toc34972024)

[5.4. Ordinary Least-Squares (OLS) regression 17](#_Toc34972025)

[5.5. Bayesian Linear Regression analysis 20](#_Toc34972026)

[5.6. OLS Regression Result 22](#_Toc34972027)

[5.7. Summary 24](#_Toc34972028)

[6. Conclusion 26](#_Toc34972029)

[Reference 27](#_Toc34972030)

[Appendices 29](#_Toc34972031)

[Appendix A : Python Code 29](#_Toc34972032)

[Appendix B : Graphs 33](#_Toc34972033)

[Appendix C: Tables 38](#_Toc34972034)

# 1. Introduction

The connection between the equity price and different financial indices obtains financial traders’ attention by attempting to predict the future financial representation in the equity market. Miscellaneous academics endeavoured to determine forecasting variables in the equity market and they detected meaningful indicators such as price-to-book ratio, stock price volatility and trading volume in comparison to other ratios and it has turned into a controversial topic. The Indian market has been flourishing substantially, with the NIFTY 500 as the predominant single financial index in India, covering the top 500 largest Indian company stocks. The aim is to assess the correlation between NIFTY 500 Index Price and 5 ratios under NIFTY 500: Price-Earnings ratio (P/E), Price-to-book ratio (PB), Dividend Yield (DY), Earnings per share, and Payout ratio, which applied to analyze the foreseeability of the NIFTY Index price in India equity market to measure whether these ratios could provide insights into the future direction of equity prices or not. In investors perspectives, they could use these rations to generate optimized and reasonable investment portfolios and strategies. The report will comprise a literature review section, data and empirical analysis result, conclusion, references and appendix to offer a conceptional overview of the topic.

# 2. Literature Review

# 2.1. Introduction

This area of the research provides a brief analysis of the methodologies as well as comparing each individual method with academic paper to demonstrate the usage.

## 2.2. Correlation Analysis

Correlation Analysis is popular within statistics when gauging the relation between two certain variables. As it assesses the strength of correlation between the variables, correlation analysis can be helpful. The correlation analysis can show positive and negative relationships. When the relation turns out positive and one variable increases, the other correlated variable will go up as well (James, 2018). A score from +0.5 to +1 indicates a strong relation. However, when a correlation is negative, the two related variables will move into reverse ways. This can be presented with a score between -0.5 and -1.

**Macroeconomic and banking system indicators**

In this academic paper, Larionova and Varlamovab used the correlation analysis to identify the relationship between the dynamics of the institutional environment and banking sector. The data from 2003 to 2012. According to their hypothesis, the relation between economic indicators and the banking indicators should be strong. The research showed that the elements of the institutional environment do influence the banking system (Larionova & Varlamovab, 2014). It showed that the banking system indicators relate to the households’ income and expenses.

## 

## 2.3. Ordinary Least Squares (OLS) Regression

OLS are often used as a quantitative measurement. As a linear least-squares type, it is particularly powerful in guessing the unfamiliar parameters in a linear regression model. These parameters are determined of a group explanatory variable according to the principle of least squares, which is reducing the difference of total squares between perceived dependent variables and the estimated ones in the linear regression (Frost, 2018).

**Impact of Factors Affecting the Underground Economy in Pakistan and Turkey**

The research aims to find out why underdeveloped countries might have greater shadow economies. For the purpose of the paper, secondary data have been used from 2000-2013. By applying the OLS model, the author scrutinised the impact of tax revenue, unemployment rate, inflation, internet users, Index of Economic Freedom, population and GDP growth rates of Pakistan and Turkey. The final results demonstrate the necessity for policies, which have the ability, to limit the measure of the underground economy (Suhail, 2017).

## 

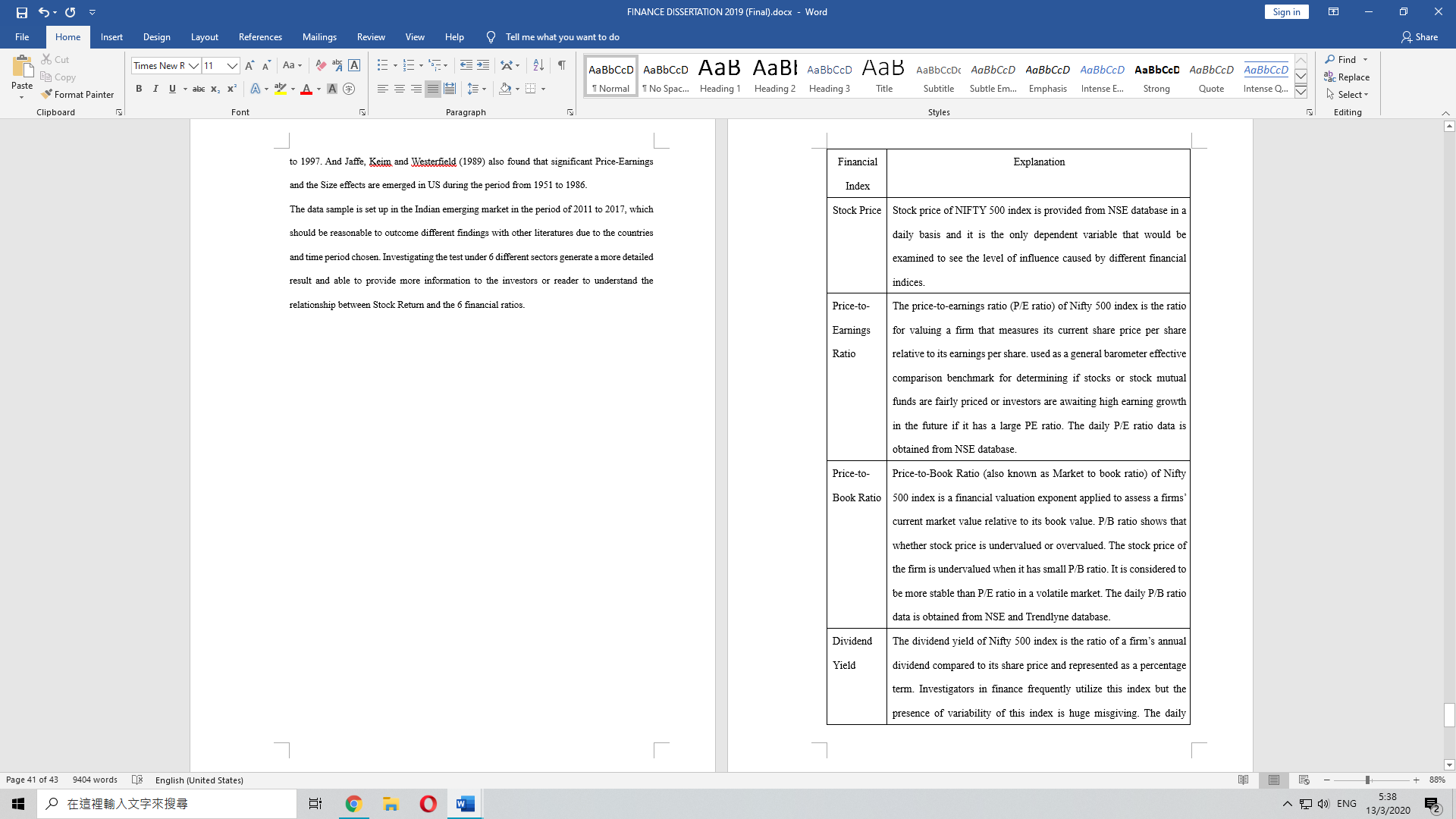
## 2.4. Bayesian Linear Regression

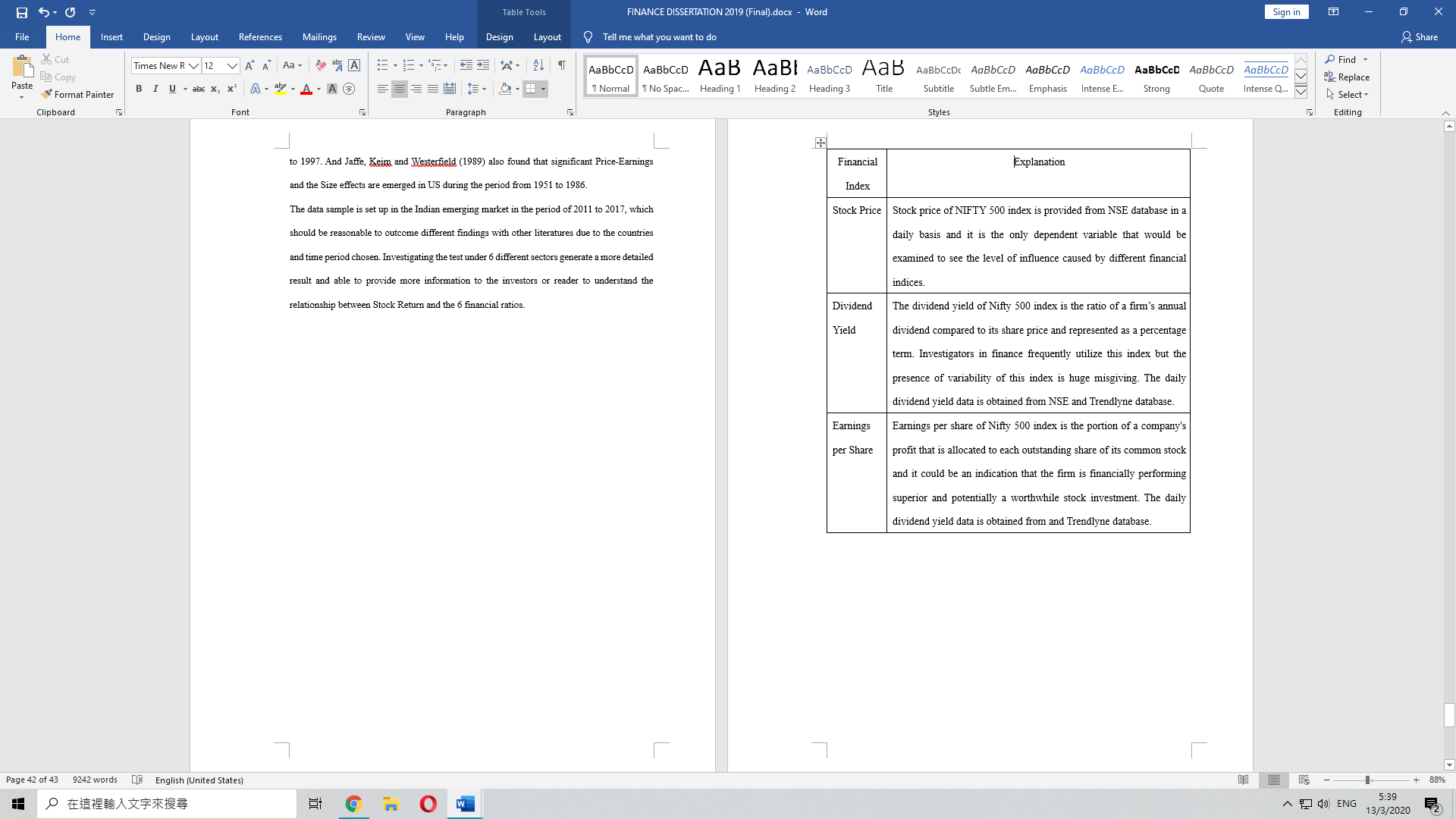
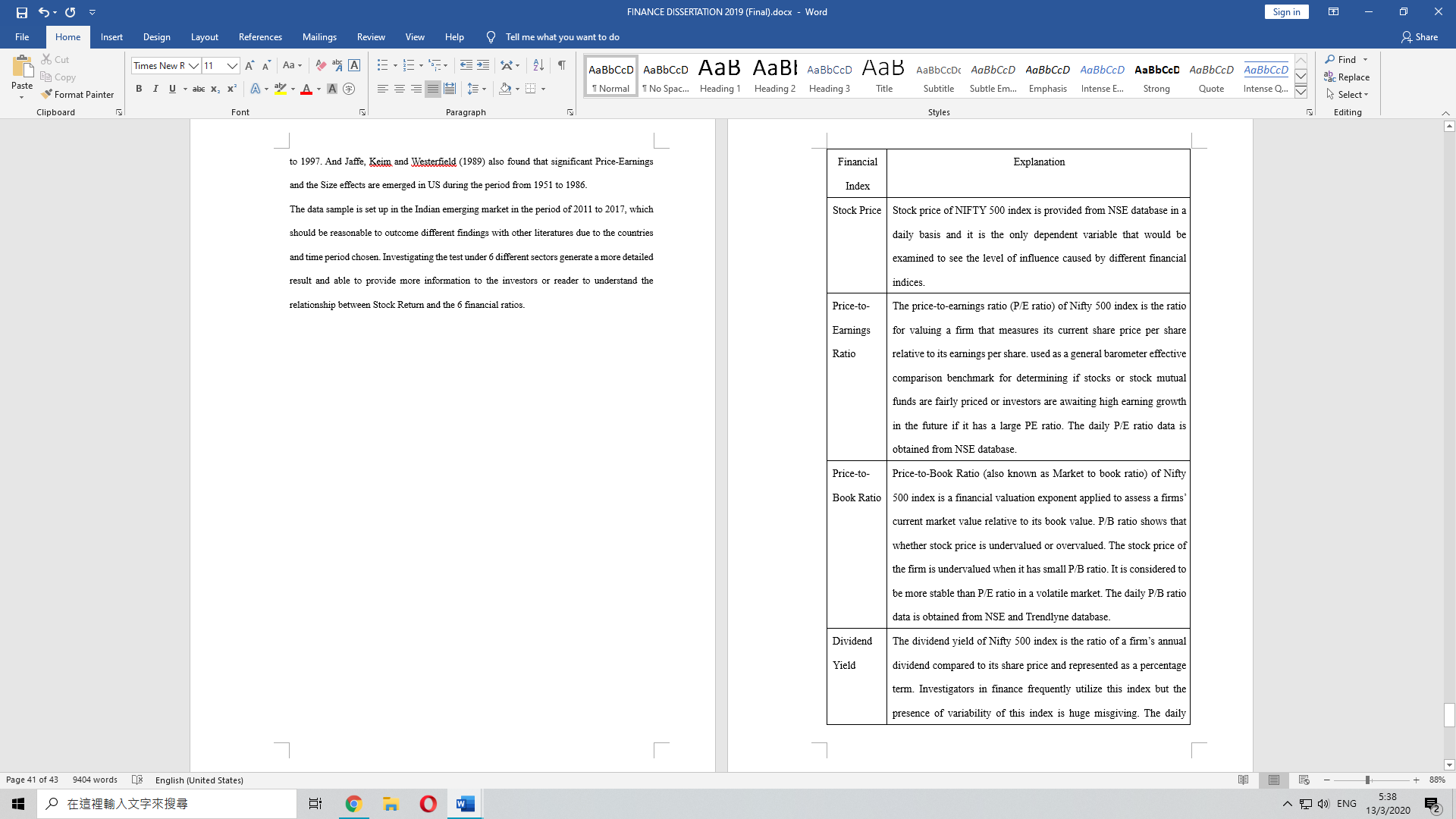
The characteristic of Bayesian models is that it calculates or measures the likelihood of an event before collecting the data. This is known as prior probability. This enables the model to adjust itself continuously due to posterior distribution. This is the process post the prior probability whereby information is being processed and updated. Posterior distribution allows the Bayesian model to create the response, which differs from other models. Therefore, Bayesian regression is not to discover the optimal value. The more data points the posterior distribution contains, the narrower the spread will be and vice versa (Verma, 2017).

**Forecasting Using a Large Number of Predictors**

The academic paper leveraged Bayesian regression for predicting focusing on time series. In the paper, the Bayesian regression is researched under a double-exponential prior. The aim was to analyse the forecasting performance using the Bayesian regression method. This was then compared with the principal component regression. The conclusion of the paper showed evidence that the Bayesian regression method provides an alternative to the principal components. The forecasted outcome of both models is strongly correlated and contains comparable mean square forecast errors (De Mol, Giannone & Reichlin, 2006).

# 3. Data

From NSE (National Stock Exchange of India Ltd.) database and Trendlyne Database, I have collected daily historical data from January 2010 to December 2019 of NIFTY 500 index stock price, which is the weighted average of 500 Indian firm stocks in different industries and is one of the two principal stock indices applied in the Indian equity market. Besides, I have gathered daily historical data from January 2010 to December 2019 with different ratios of NIFTY 500 index such as Price-Earnings Ratio, Price-To-Book Ratio, Dividend Yield and Earnings per share. After converting the data into one excel file, I started to set up my time series from 4 January 2010 to 31 December 2019 and convert the Payout Ratio by using Dividend Yield divided by Earnings Yield, which gathers 2473 observations per ratio. The detail of the financial index is provided below:



Giannetti (2007) found the market timing ability did exist based on the independent variable PE ratio for the 500 Index and Kurach and Słoński (2015) also measured the influence between stock performance with PE ratio and Dividend Yield with the sample of approximately 600 observations for a decade from 2001 they discovered the important transiency in present annual earnings and verify the appearance of transiency. Other researchers have been using similar financial indices to test their hypotheses and I complement more independent variables in this essay.

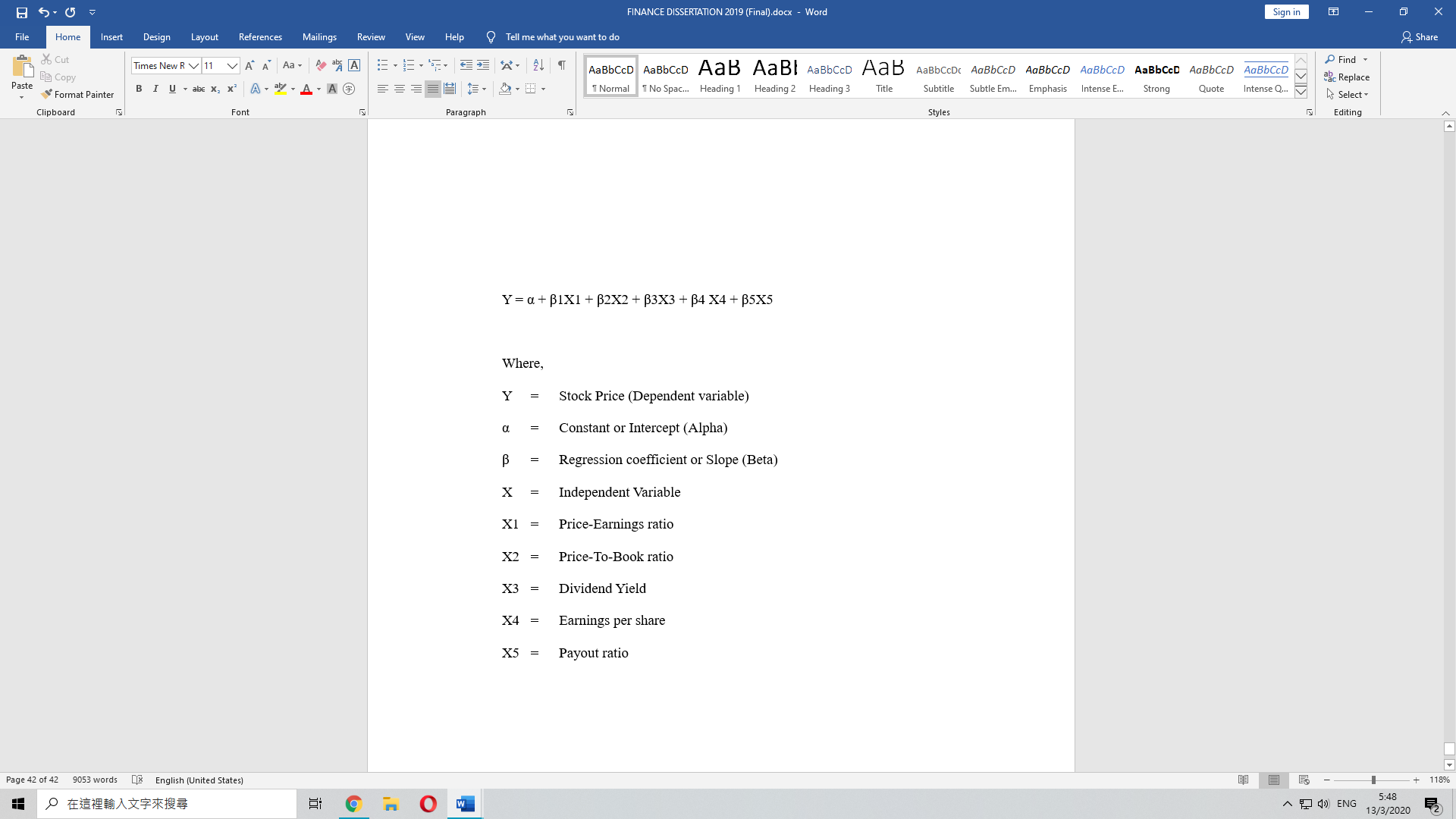
# 4. Methodology

The essay scrutinizes the correlation between Stock Price and 5 independent variables based on NIFTY 500 index, which are Price-Earnings Ratio, Price-To-Book Ratio, Dividend Yield and Earnings per share and Payout ratio. As mentioned above, I collected all of these variables data from databases in a daily basis in the latest ten years from 4 January 2010 to 31 December 2019 and NIFTY 500 index is the weighted average of 500 Indian firms. Appendix A indicates how to analyse data with different ratios collected from the database through different statistical models such as Correlation Analysis and Linear Regression Model by using Python.

With the statistics models applied, summary statistics table, statistical analysis based on changes over time graphs, correlation Analysis table and graph, OLS and Bayesian Linear Regression comparison graphs would be generated and located in Section 5.

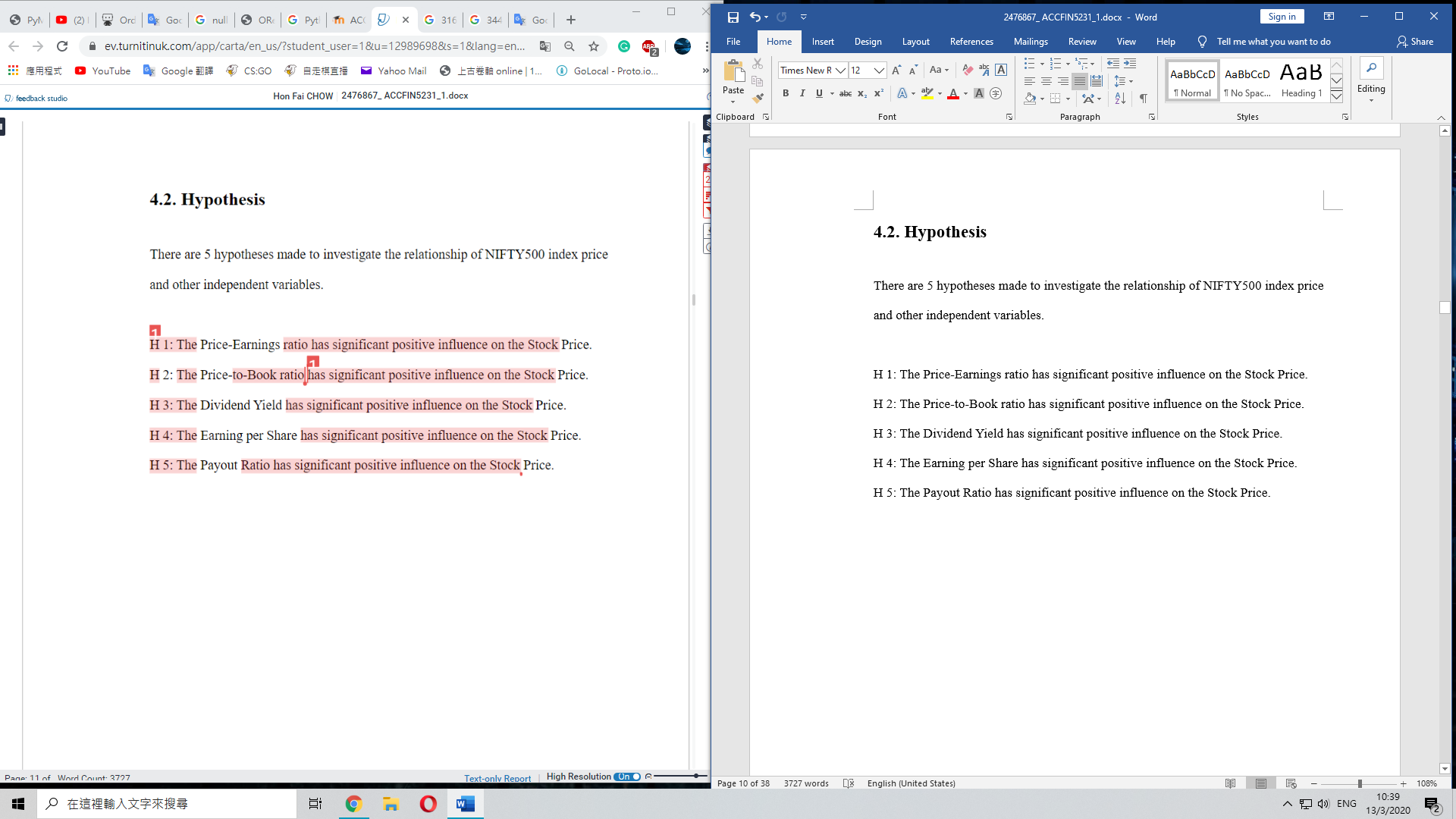
## 4.1. Regression Model

Data with a different independent variable (Financial Indices) that could be used to forecast the future stock price of Nifty 500 Index are gathered and applying regression line could form a simple mathematical formula, which is indicated following:



## 4.2. Hypothesis

There are 5 hypotheses made to investigate the relationship of NIFTY500 index price and other independent variables.



# 5. Empirical Results

Through using Python (Appendix A), the data sample presented in Section 3 and the methodology illustrated in Section 4 is applied to illustrate the correlation between Stock Price and other independent variables. In this section, Summary Statistics

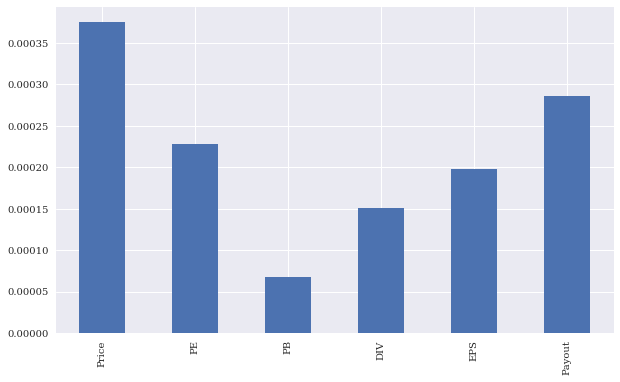
Statistical analysis (changes over time), Correlation Analysis, OLS regression, Bayesian Linear Regression analysis and summary would be provided.

## 5.1. Summary Statistics

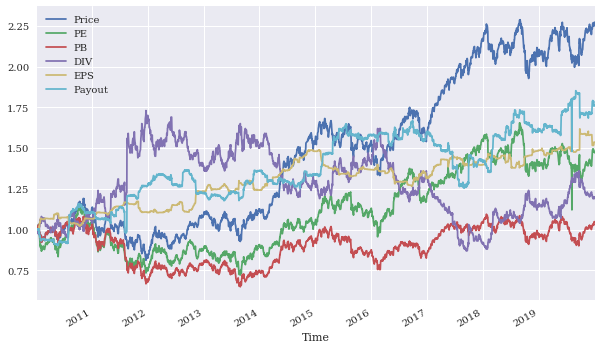
The observations of all variables are 2473, which should be enough to provide a representative data sample size for different Statistical analysis models to analyse the correlation between Price and other ratios. The average and standard deviation of all 5 independent variables seem to be quite low except the EPS with mean 270.69 and standard deviation 30.76. The mean and standard deviation of Price are 6411.96 and 1985.84, which are more immense compared to other variables.

## 5.2. Changes Over Time

Statistical analysis methods are often based on changes over time and not the absolute values themselves. There are different ways to calculate the changes in a time series over time such as percentage changes and logarithmic (log) return. From a statistics point of view, absolute changes are not optimal because they are dependent on the scale of the time series data itself. Therefore, percentage changes are usually preferred. Using the Python Code could derive the percentage change in different dependent variables to visualizes their mean values per column (see Figure 1)



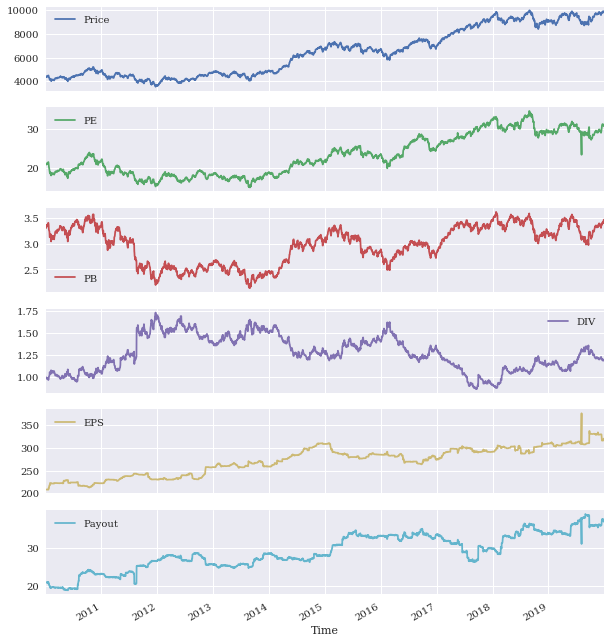
**Figure 1.** Mean values of percentage changes with all variables as bar plot

The other method that could be applied is log returns. In some cases, they could provide a different overview by easily applying data into the model and the financial context.Figure 2 indicates the Cumulative log returns for the single financial time series. (See Appendix 1.2 for the Python Code)

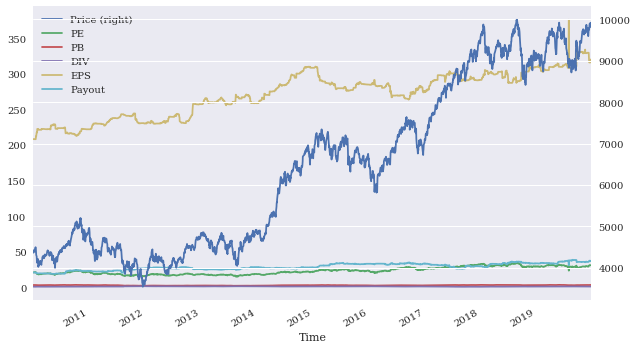
**Figure 2.** *Cumulative log returns over time (Daily)*

Rolling statistics are financial tradition to use since they are the basic tools for financial chartists and technical traders. Appendix B.1, B.2, B.3 and B.4 indicate plots the rolling correlation between Price and PE ratio, PB ratio, Dividend Yield and EPS ratio with a single financial time series. (see Appendix 1.5 for the Python Code)

## 5.3. Correlation Analysis

This section indicated how to come up with some supporting statistical evidence for the stylized fact that the Stock Price with PE ratio and PB ratio are highly positively correlated, the Stock Price and dividend yield are highly negatively correlated and the remaining variables and the Stock Price is not correlated. (See Figure 3, 4)

**Figure 3.** *Financial variables time series data as line plots (different subplots)*



**Figure 4.** *Financial variables time series data as line plots (same plots)*

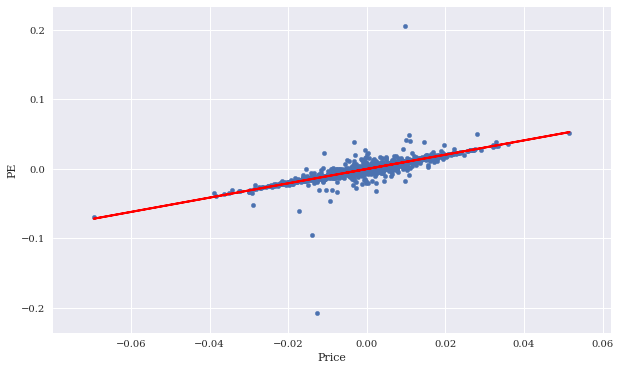
As mentioned above, statistical analysis in general relies on returns instead of absolute changes or even absolute values. Hence, calculating log returns before processing any further analysis should be executed. Appendix B.1 and B.2 indicate the variability of the log returns of the Nifty 500 and other variables over time and as a scatter matrix. Table 1 finds that there is a positive relationship between the Stock Price and P/E ratio and P/B ratio with correlation respectively 0.812527 and 0.957325. Furthermore, it demonstrates that PE ratio has positive correlations with PB ratio (0.776223) and Payout ratio (0.402194) and has negative correlations with Dividend Yield (-0.631299) and Earnings per share (-0.614523). At the meantime, Dividend Yield indicates high

negative correlations compared with all variables except Payout ratio (0.456143) and EPS (0.03032360).

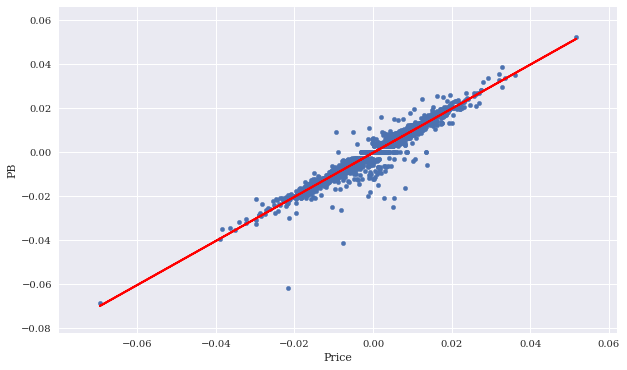
## 

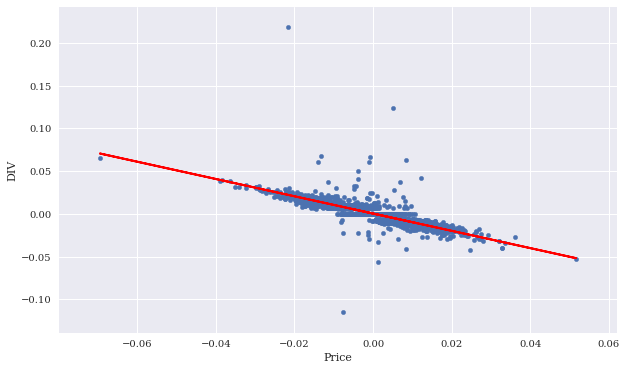
## 5.4. Ordinary Least-Squares (OLS) regression

OLS regression analysis graphs are to be implemented to testify the relationship between the Nifty 500 stock price and other independent variables with log returns as a scatter plot from Figure 5 to 9, which demonstrate the same correlations as Table 1

but in visualized diagrams.

**Figure 5.** *Log returns of the Nifty 500 price and PE ratio as a scatter matrix*

**Figure 6.** *Log returns of the Nifty 500 price and PB ratio as a scatter matrix*

Figure 5 reveals a positive correlation between the Stock Price and P/E ratio with a correlation of 0.812527 and Figure 6 reveals a positive correlation between the Stock Price and PB ratio with a correlation of 0.957325. Besides, Figure 7 reveals a negative correlation between the Stock Price and Dividend with a correlation of -0.775486.

**Figure 7.** *Log returns of the Nifty 500 price and Dividend Yield as a scatter matrix*

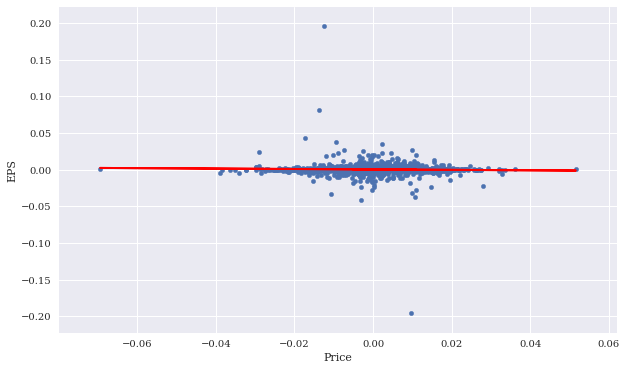
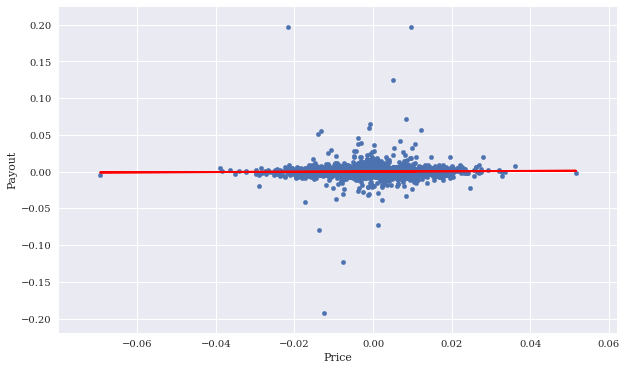
**Figure 8.** *Log returns of the Nifty 500 price and EPS as a scatter matrix*

Figure 8 and 9 reveals that there is no correlation between the Stock Price and Earnings per share and Payout ratio.

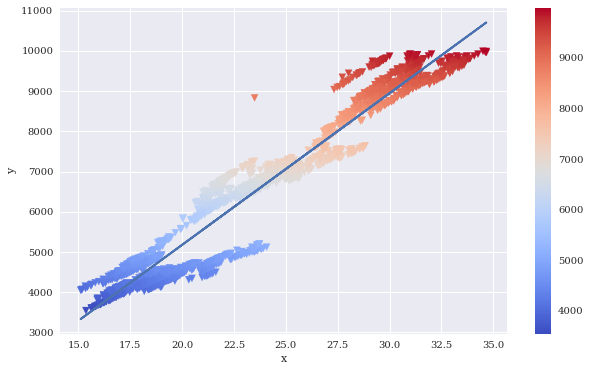


**Figure 9.** *Log returns of the Nifty 500 price and Payout ratio as a scatter matrix*

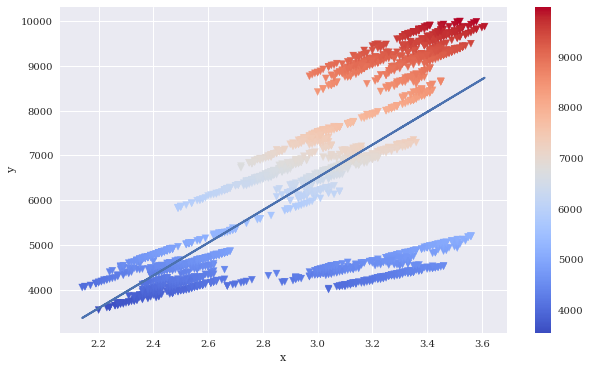
## 5.5. Bayesian Linear Regression analysis

Bayesian Linear Regression could provide regression lines based on the different estimates to generate the posterior distribution for the different parameters as

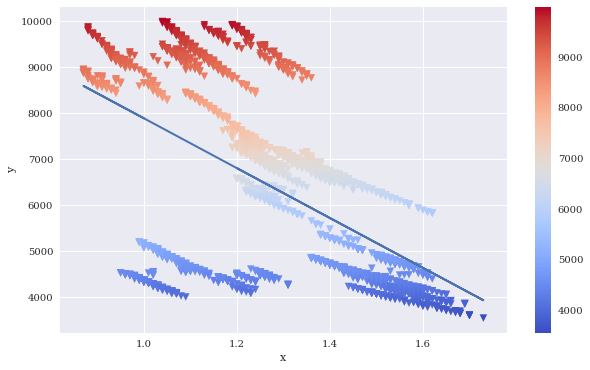
well as all single estimates per sample. However, due to the technological issues, the graphs are not provided as it took an extended period to run in Python since the sample size has 2473 observations and the python code need to run 4 chains and more 100 tunes to produce a representative posterior distribution. However, Figure 10 to 12 indicate a single estimate plotline with the fixed values for the two parameters of the regression line. Figure 10 and 11 display that there are positive correlations between stock price and PE ratio and PB ratio, whereas Figure 12 displays a negative correlation between stock price and dividend yield.



**Figure 10.** *Bayesian Regression correlation between the Nifty 500 price and PE ratio*



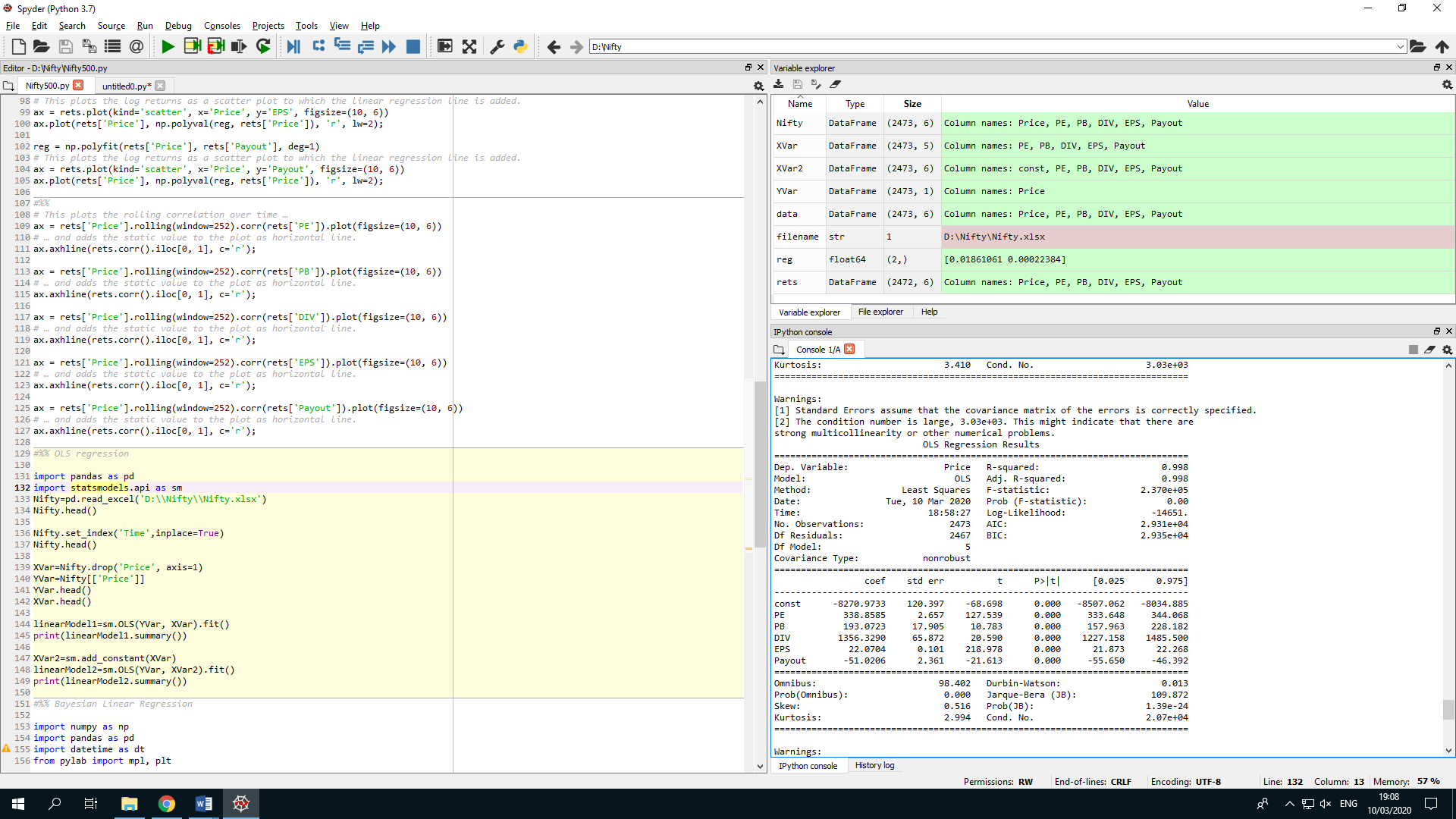
**Figure 11.** *Bayesian Regression correlation between the Nifty 500 price and PB ratio*



**Figure 12.** *Bayesian Regression correlation between the Nifty 500 price and Dividend Yield*

## 5.6. OLS Regression Result

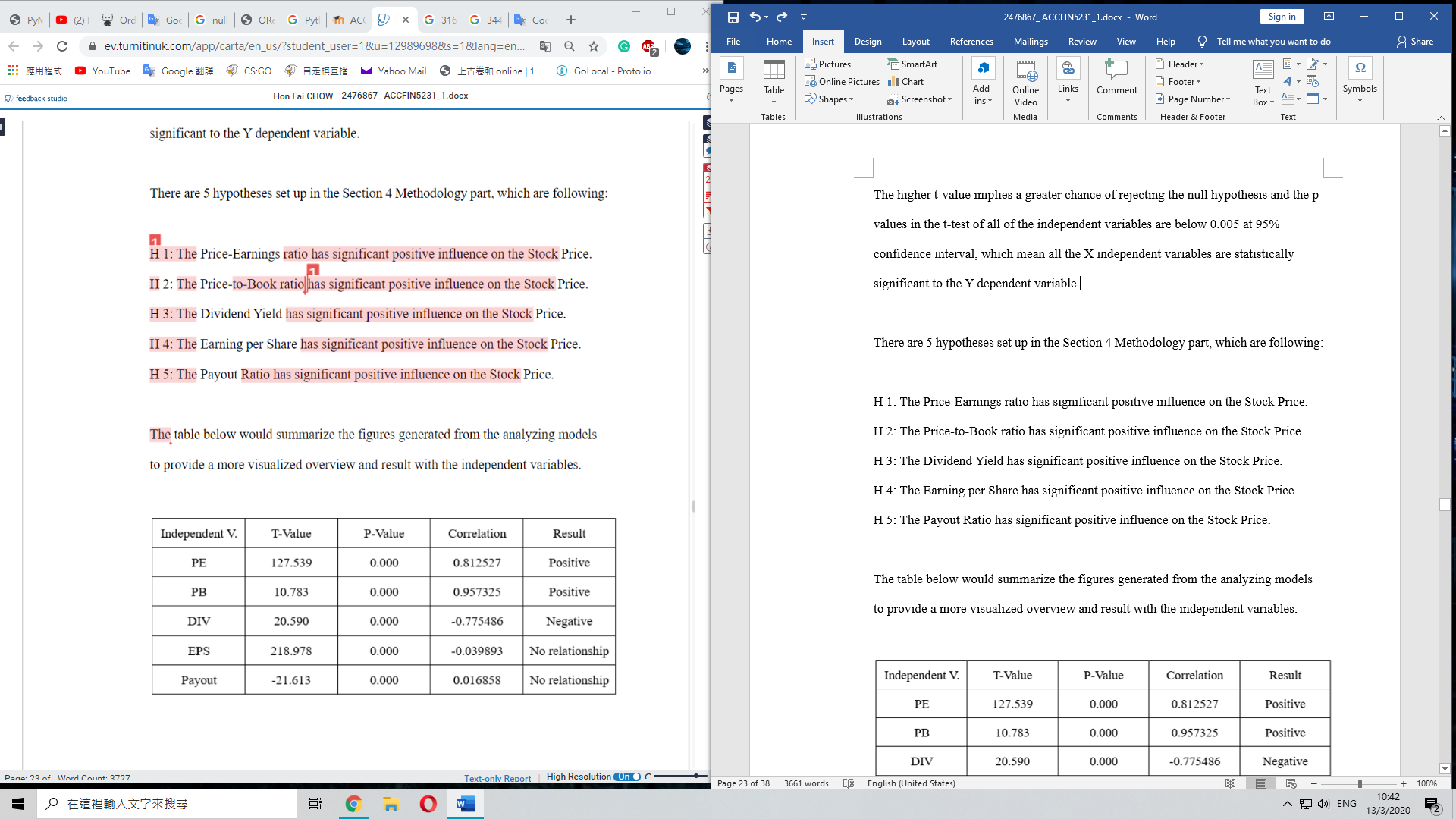
The OLS regression result measure the R-squared, the Y-intercepts (coef) and the corresponding t-value and p-value of t statistic, which demonstrates in Table 3 with scrutinizing the correlation and the statically significance level among Stock Price the 5 independent variables from January 2010 to December 2019

**Table.3** represent the OLS Regression Results among Price and other 5 independent variables from 2010 to 2019.

In statistics, the coefficient of determination that is denoted by R-squared is the ratio of variance in the explanatory variable that is able to estimate from the independent variable. The R-squared is 0.998, which is close to 1 and means the linear regression model fit well. The large overall F-statistic (2.37e+05) and the Prob value of the F-statistic less than 0.05 indicate that there is a good amount of linear relationship between the dependent variable (Price) and the independent variables (PE ratio, PB ratio, Dividend Yield, EPS and Payout ratio).

The higher t-value implies a greater chance of rejecting the null hypothesis and the p-values in the t-test of all of the independent variables are below 0.005 at 95% confidence interval, which means all the X independent variables are statistically significant to the Y dependent variable.

There are 5 hypotheses set up in the Section 4 Methodology part, which are following:



The table below would summarize the figures generated from the analyzing models to provide a more visualized overview and result with the independent variables.



## 5.7. Summary



The results fail to reject the H1 hypothesis since the P value of the P/E ratio is under 0.05 and the t-statistics value of it is 127.539, which implies strong statistically significant positive correlation on the price with a correlation of 0.813. This signify stocks or stock mutual funds are fairly priced or the investors could utilize the P/E ratio to predict the price. Mahmood and Waheed (2014) possessed the similar findings that P/E ratio is strongly relevant with the stock performance with the sample size more than 100 companies from 1998 and 2009, and Guler Aras and Mustafa Kemal Yilmaz (2008) also stated that P/E ratio could be a good predictive variable to forecast the future price.



The results fail to reject the H2 hypothesis since the P value of the P/B ratio is under 0.05 and the t-statistics value of it is 10.783, which implies strong statistically significant positive correlation on the price with a correlation of 0.957. The results fail to reject the H2 hypothesis, which mean stock price is overvalued or the investors could utilize the PB ratio to predict the price. Jeffrey Pontiff and Lawrence D. Schall (1997) and Guler Aras and Mustafa Kemal Yilmaz (2008) both had the same finding that price to book ratio is a good indicator to forecast the market price return



The results reject the H3 hypothesis since the correlation between the price and dividend yield is 0.957, which show negative correlation between parameters. However, it is strong statistically significant as the P value of the it is under 0.05 and the t-statistics value of it is 20.590, Hence, this display strong statistically significant negative correlated with the price so investors could utilize the Dividend Yield to forecast the price but the presence of variability of this index is huge misgiving. Mahmood and Waheed (2014) stated that Dividend Yield has negative impact with the stock performance with the sample size more than 100 companies from 1998 and 2009, while Kee Hong Bae, Jeong-Bon Kim(1998) and WU, J., HU, Y. and LEE, C. (2011) also stated that the Dividend Yield is a pointer to forecast the market price.



The results reject the H4 hypothesis because it has no relationship with the stock price with a correlation of -0.04, even though the EPS is statistically significant with a P value under 0.05 and a t-statistics value of 218.978. This means that the investors should not use the EPS to predict the price since there is no relationship between stock price and EPS.



The results reject the H5 hypothesis because it has no relationship with the stock price with a correlation of 0.0, even though the EPS is statistically significant with a P value under 0.05 and a t-statistics value of -21.613. This means that the investors should not use the payout ratio to predict the price since no relation between stock price and payout ratio.

# 6. Conclusion

The essay researches the connection and the ability to predict equity price in Nifty 500 index using as Price-Earnings Ratio, Price-To-Book Ratio, Dividend Yield, Earnings per share and Payout ratio as predictive variables. The data sample includes 2473 observations with daily historical data from January 2010 to December 2019 with different ratios in NIFTY 500 index.

Correlation analysis, OLS and Bayesian regression models are applied to assess how it might affect the aforementioned variables on firm stock return. The outcome display that Price-Earnings and Price-to-Book ratio contain statistically significant positive correlation with the Stock Price and Dividend Yield has statistically significant negative correlation on the stock price.

Reference :

Aras, G., & Yilmaz, M. K. (2008). Price-earnings ratio, dividend yield, and market-to-book ratio to predict return on stock market: Evidence from the emerging markets. Journal of Global Business and Technology, 4(1), 18-30.

Bae, K. H., & Kim, J. B. (1998). The usefulness of earnings versus book value for predicting stock returns and cross corporate ownership in Japan. Japan and the World Economy, 10(4), 467-485.

De Mol, C., Giannone, D., & Reichlin, L. (2006). Forecasting using a large number of predictors is bayesian regression a valid alternative to principal components? (pp. 3-28). Retrieved from <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp700.pdf?ffd9f4561990919f315127a2eee70375>

Frost, J. (2018). Ordinary least squares - Statistics By Jim. Retrieved 7 March 2020, from <https://statisticsbyjim.com/glossary/ordinary-least-squares/>

Giannetti, A. (2007). The short term predictive ability of earnings-price ratios: The recent evidence (1994–2003). *The Quarterly Review of Economics and Finance*, *47*(1), 26-39.

James, E. (2018). What is Correlation Analysis? Definition and Exploration. Retrieved 6 March 2020, from <https://blog.flexmr.net/correlation-analysis-definition-exploration>

Larionova, N., & Varlamovab, J. (2014). Correlation analysis of macroeconomic and banking system indicators (pp. 2-7). Retrieved from <https://core.ac.uk/download/pdf/82645401.pdf>

Mahmood, H., & Waheed, A. (2014). Relationship Between Price/Earning, Dividend Yield, Size and Stock Returns of Listed Pakistani Firms. Journal of Basic and AppliedScientific Research, 4(10), 83-90.

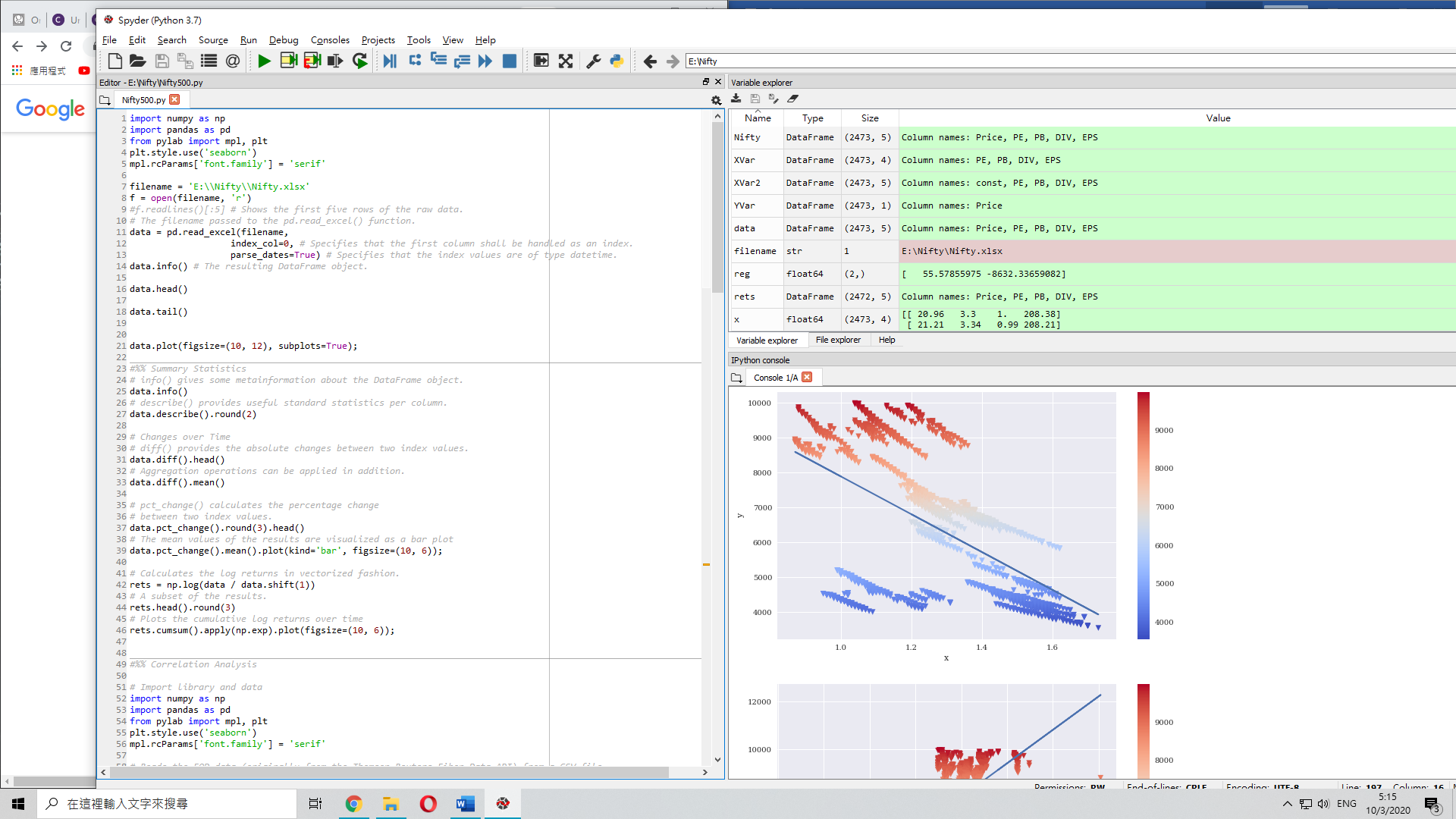
Pontiff, J., & Schall, L. D. (1998). Book-to-market ratios as predictors of market returns. Journal of Financial Economics, 49(2), 141-160.

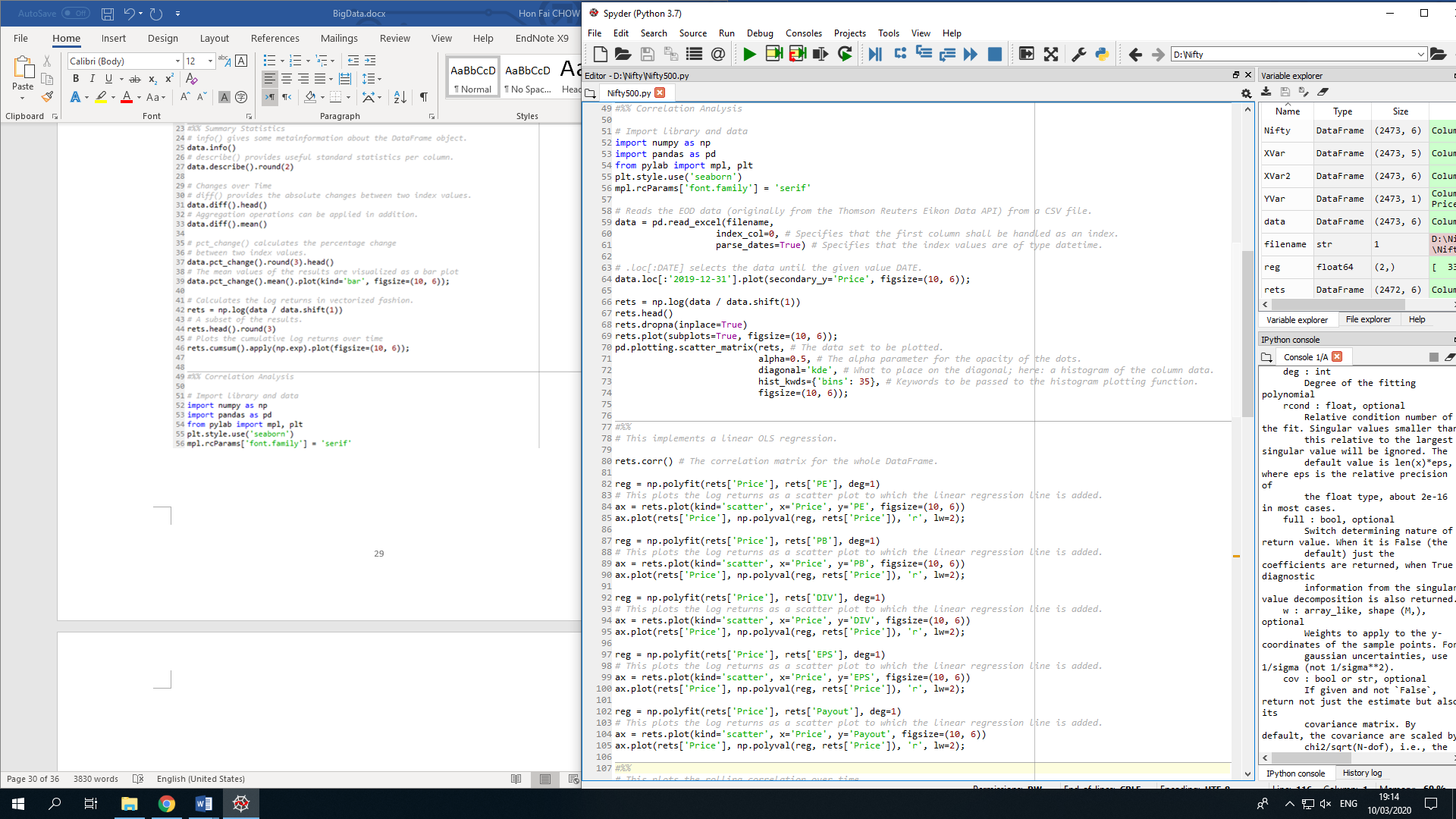
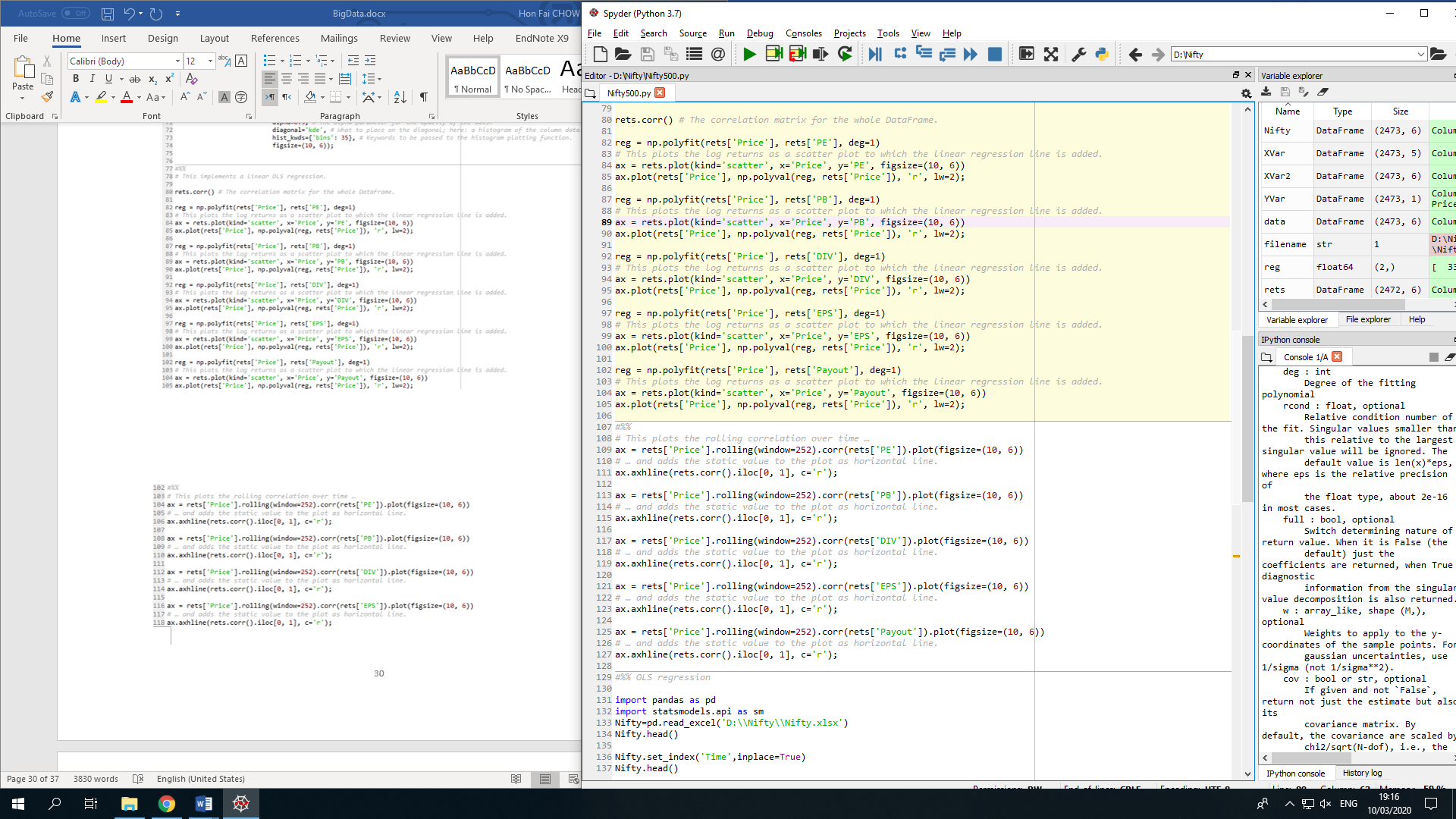
Suhail, G. (2017). Using Ordinary Least Squares to Measure the Impact of the Factors Affecting Underground Economy: A Comparison between Pakistan and Turkey (pp. 2-6). Retrieved from http://jbepnet.com/journals/Vol\_4\_No\_3\_September\_2017/7.pdf

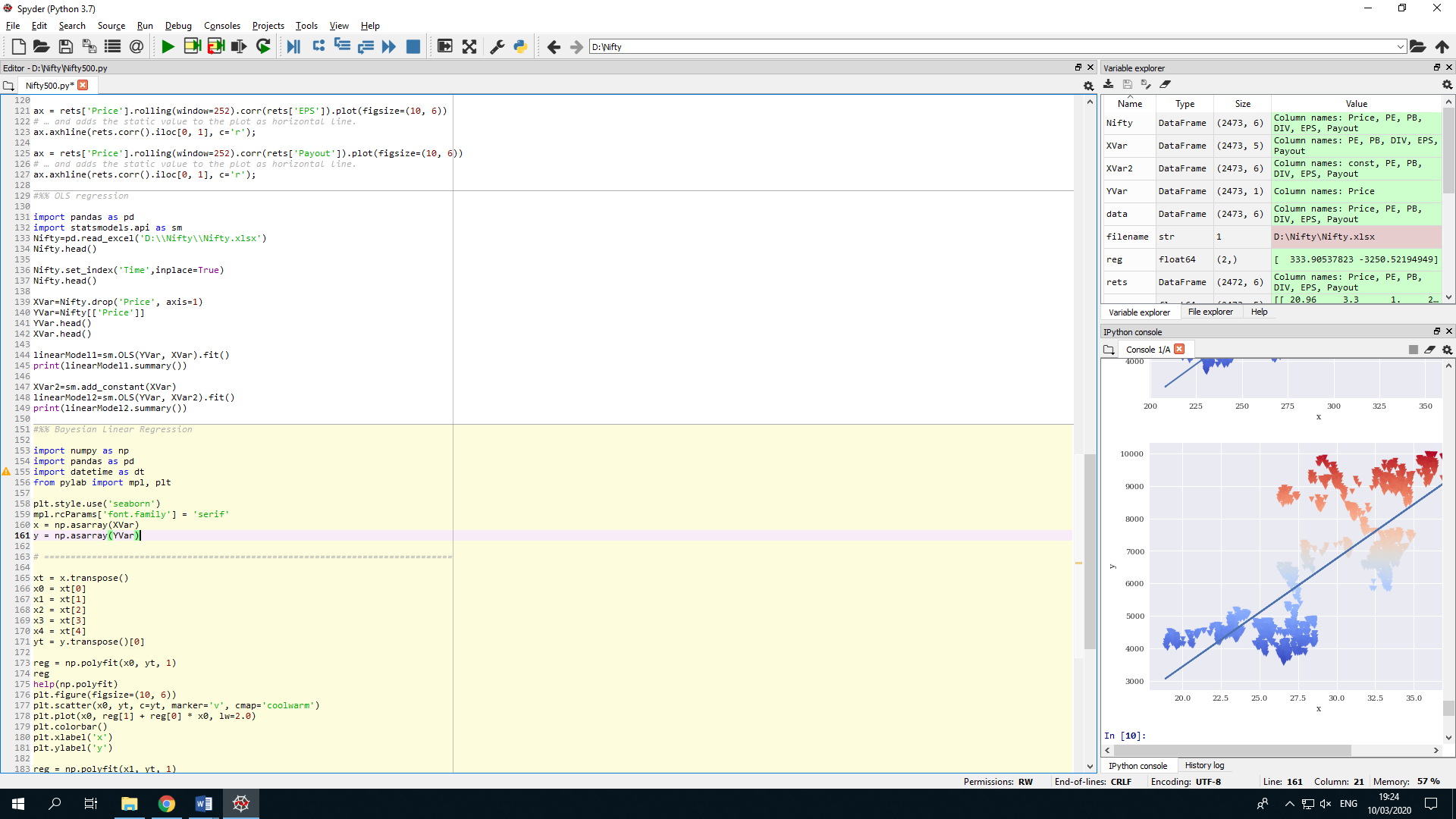
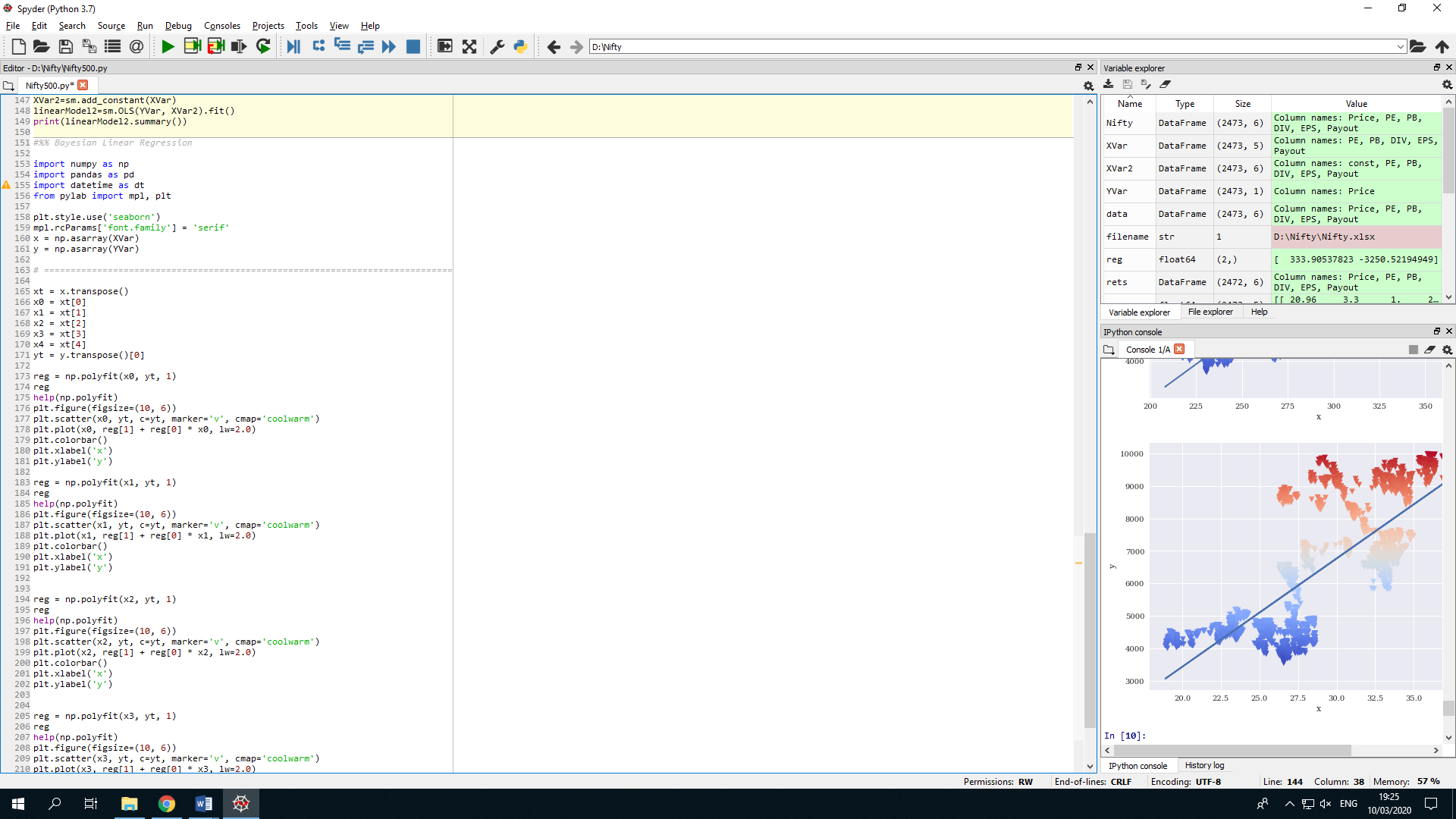
Verma, J. (2017). Linear Regression: Frequentist and Bayesian. Retrieved 9 March 2020, from https://medium.com/markovian-labs/linear-regression-frequentist-and-bayesian-447f97c8d330

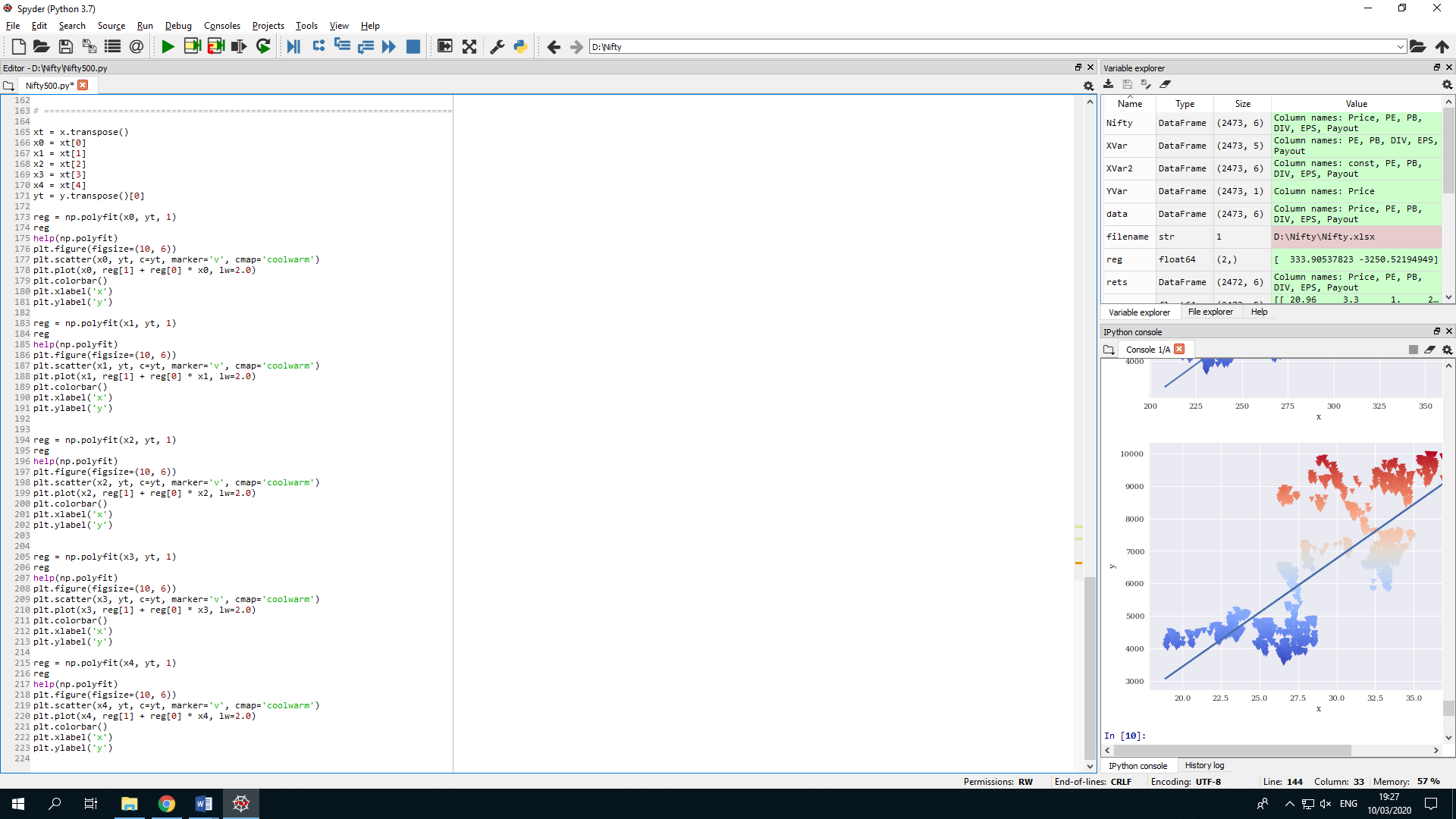
Wu, J. L., Hu, Y. H., & Lee, C. (2011). Can Dividend Yields Out‐Predict Uk Stock Returns Without Short Rates?. The Manchester School, 79(6), 1179-1196.

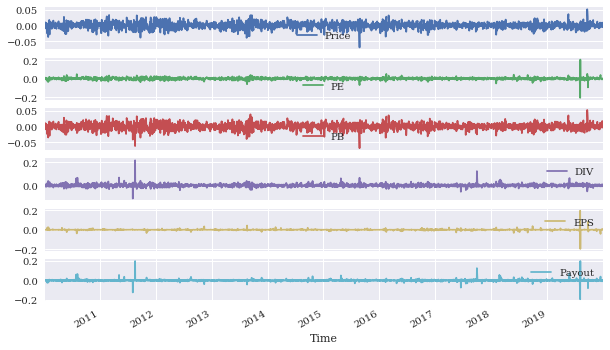
Appendices :

Appendix A : Python Code

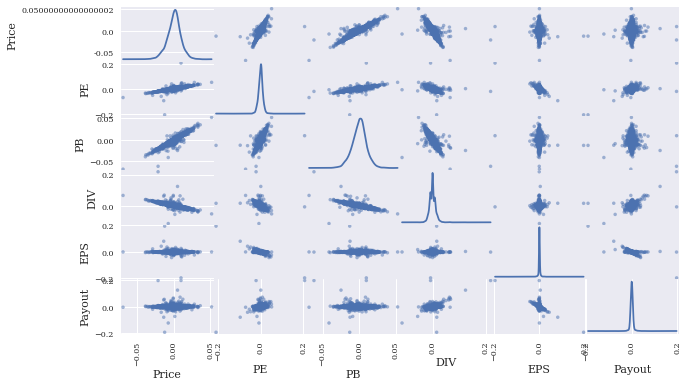




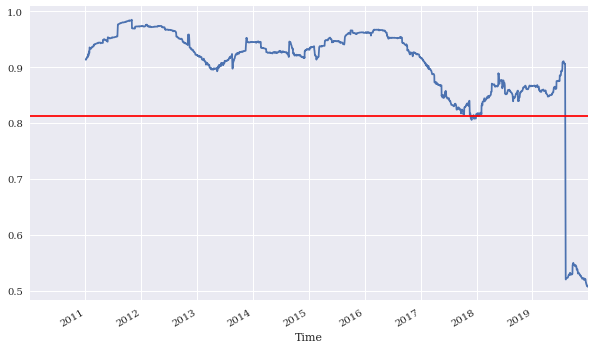


 Appendix B : Graphs

**Figure B1.** Log returns of the log return of the Nifty 500 and other variables over time



**Figure B2.** Log returns of the Nifty 500 and other variables as a scatter matrix

**Figure B3.** Correlation between Price and PE ratio over time (static and rolling)



**Figure B4.** Correlation between Price and PB ratio over time (static and rolling)



**Figure B5.** Correlation between Price and Dividend Yield over time (static and rolling)



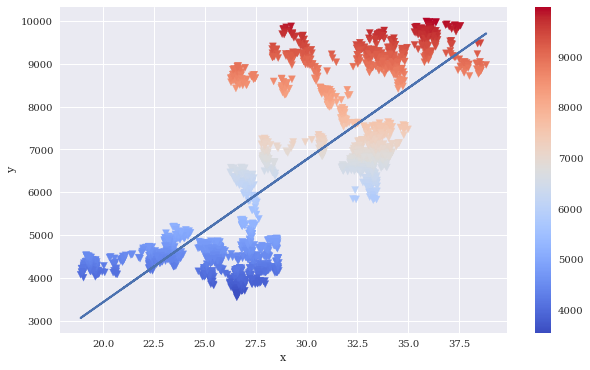
**Figure B6.** Correlation between Price and EPS ratio over time (static and rolling)



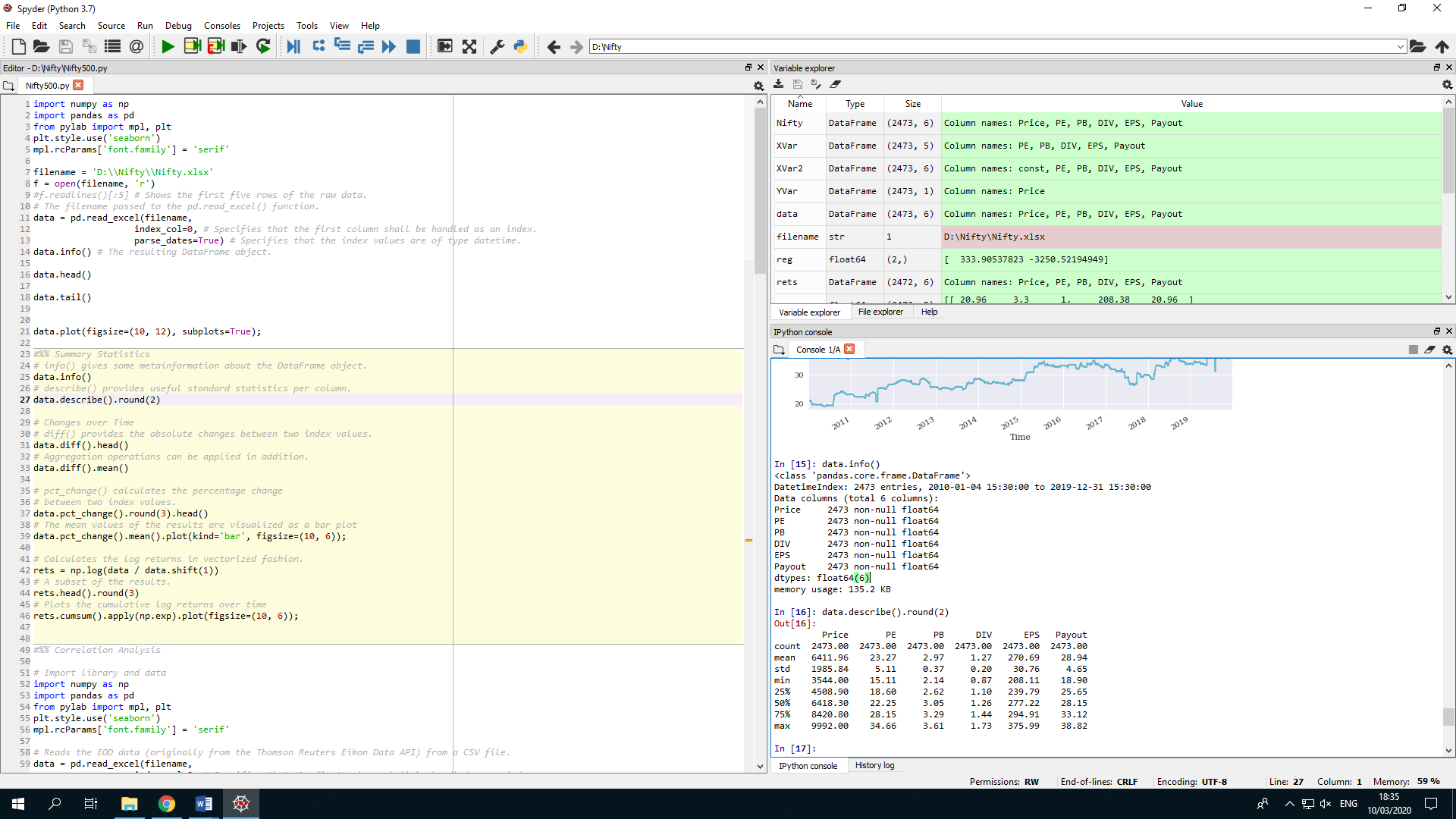
**Figure B7.** Correlation between Price and Payout ratio over time (static and rolling)



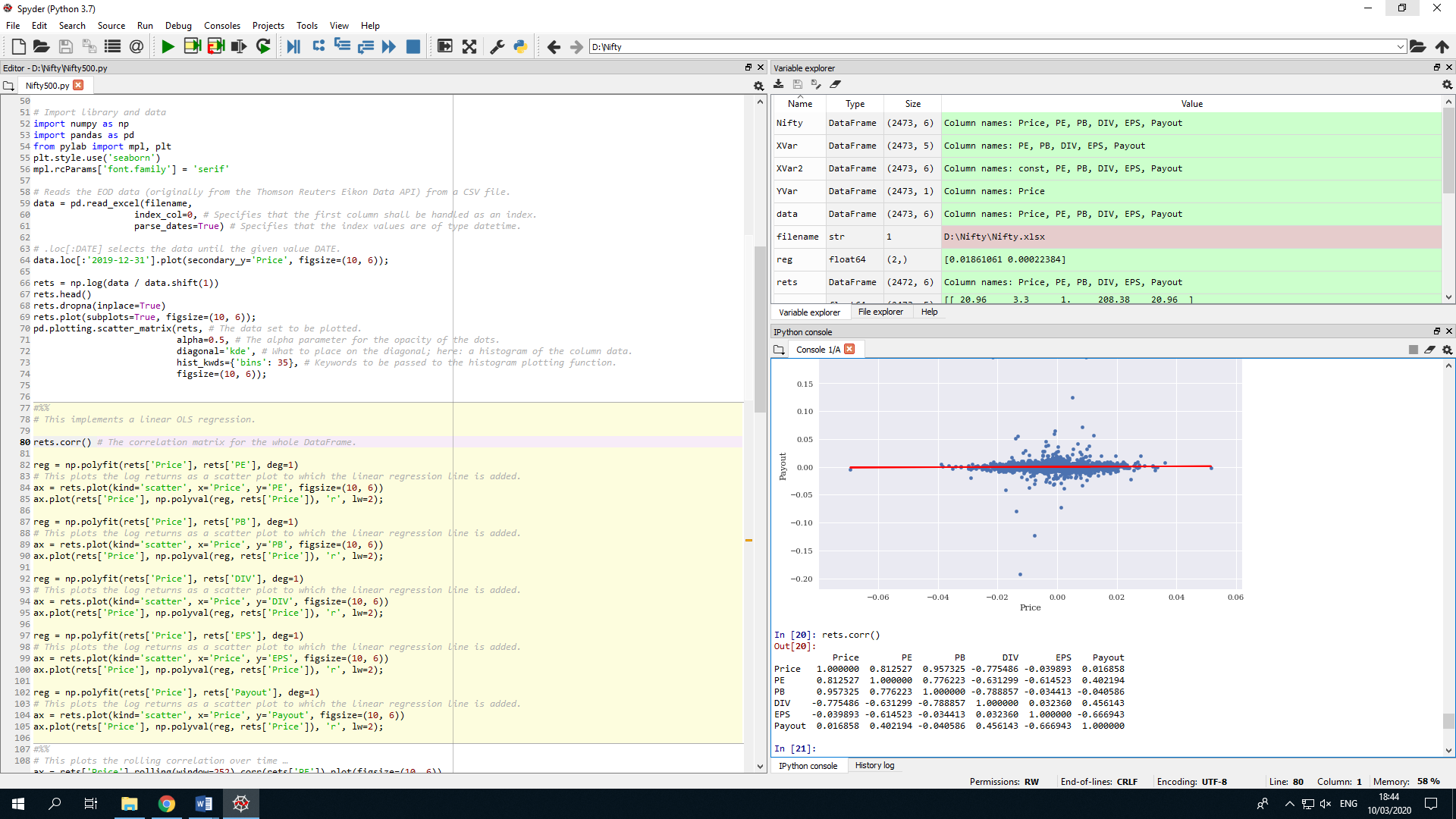
**Figure B8.** *Bayesian Regression correlation between the Nifty 500 price and EPS ratio*



**Figure B9.** *Bayesian Regression correlation between the Nifty 500 price and Payout ratio*

Appendix C : Tables

**Table C1.** Summary Statistics of all the variables from 2010 to 2019



**Table C2.** Log return correlation of all variables from 2010 to 2019