Analysis factors determining U.S. Stock return rate*

Yunshu Zhang

April 19, 2024

This paper utilizes data from the World Bank to examine the relationship between U.S. Stock return rate and key indicators. By using linear regression models, it revealed a strong correlation between U.S. Stock return rate and factors such as the price-to-earnings ratio, also known as the PE ratio, is the ratio between a company's stock price and its earnings per share (EPS) and industry category. These findings are significant as they underscore the relationship between monetary policy and broader economic conditions. These findings can provide valuable guidance to investors in understanding the factors influencing Stock return rate in the U.S. and make effective investment strategies.

Introduction

The U.S. stock market is known as one of the largest and most influential financial markets globally, attracting investors from all over the world. With its vast array of listed companies spanning various industries and sectors, it offers a wealth of investment opportunities to individuals and institutions alike. Investors flock to this market in search of both short-term gains and long-term wealth accumulation. There are some common investments such as stock, bond, real estate, and forex. At the same time, many international investors closely monitor the performance of the U.S. stock market to guide their investment decisions.

In this paper, we will investigate the factors influencing the return rates of the U.S. stock through a linear regression analysis. The estimand of our analysis will be on understanding how various factors impact the return rate of the U.S. Stock. We interest in factors such as industry category and PE ratio. For example, how two different industry's return rate change from the 2010 to 2024. By analyzing these variables, we found that different industry have different return rate.

---

*Code and data are available at:https://github.com/Yunshu921/repro_microplastics_effect_on_fish.git

Unlike some literature focus on comparing different mutual funds which includes: money market funds, bond or fixed income funds, stock or equity funds, or hybrid funds (Wikipedia contributors 2024a). This paper focus on explore how diverse factors, specifically, industry category and PE ratio influence the return rate of stock.

There are five chapters in this paper. This first section is introduction. Section 2 is data-related, emphasizing the strengths and limitations of the dataset. Methodologies adopted from the data set and key terminologies are also discussed. Then subsection of section 2 outlines the analysis results, primarily examining the relationship between return rate of stock and factors. In Section 3, we employ a linear regression model to further analyze these trends. Section 4 is dedicated to visualizing the model's outcomes through graphical representation and summary the result from previous analysis. Finally, the concluding section explores the factors influencing the observed trends, analysis of the bias and ethical concerns, and weakness and next steps.

Data

Data Description and Measurement

The data used in this paper is gained from the World Bank (Wikipedia contributors 2024b). In the World Bank Open Data, we can choose data category we interested in and then download the specific dataset. Also, there are various finical resources such as Yahoo! finance which is is a media company owned by Yahoo! Networks. It provides financial news, data and commentary, including stock quotes, press releases, financial reports and original content. It also offers a number of online tools for personal financial management. It publishes original stories written by its team of journalists, in addition to publishing paid partner content from other sites. The site is ranked #20 on SimilarWeb's list of the largest news and media sites (Wikipedia contributors 2024c). We can downloaded data from the website: https://finance.yahoo.com/quote/ACGL/history?frequency=1wk. In this website, we can downloaded historical data about ACGL stock monthly from 2010 to 2024. We choose the World Bank since Databank provide the database has advanced features for selecting and slicing data sets, and executing custom queries, and data downloads. The raw dataset contains 228 different kinds of U.S. stock, each stock has 171 historical data from 2010 to 2024. And there are 17 variables. They are:
Qualitative measurements: Id means the name of stock company and different industries.
Quantitative measurements:
PE means the ratio between a company's stock price and its earnings per share.
RE means the a measure of the change in the price of a stock over a specific period of time, typically expressed as a percentage. Stock return reflects the return earned by investors who hold the stock during the given time frame.
ROE (Return on Equity): This ratio measures a company's profitability relative to its shareholder equity. It shows how much profit a company generates with the money shareholders

have invested. A higher ROE indicates more efficient use of equity capital. It is a internal tool for company managers to make efficient decisions.

OPTOEBT (Operating Income Before Extraordinary Items): This metric represents a company's earnings from its core business operations before accounting for extraordinary items such as restructuring charges, litigation settlements and natural disasters. It gives insight into how well a company is generating profits from its primary activities, excluding one-time events listed above.

OCFTOOPERATEINCOME (Operating Cash Flow to Operating Income Ratio): This ratio indicates how efficiently a company converts its operating income into cash flow. It shows the proportion of operating income that's converted into cash, which is important for assessing a company's liquidity and ability to meet short-term obligations.

DEBTTOASSETS (Debt to Assets Ratio): This ratio measures the proportion of a company's assets financed by debt. It indicates the company's risk exposure. The formula is: Total Debt / Total Assets. A higher ratio suggests a higher degree of financial risk.

CURRENT (Current Ratio): This ratio evaluates a company's ability to pay its short-term liabilities with its short-term assets. It's calculated by dividing current assets by current liabilities. A ratio above 1 indicates that the company has more current assets than liabilities, implying good liquidity.

APTURN (Accounts Payable Turnover): This ratio measures how efficiently a company manages its accounts payable by comparing purchases to average accounts payable. It indicates how quickly a company pays off its suppliers. A higher turnover ratio suggests efficient management of accounts payable.

YOYEPS_BASIC (Year-over-Year Growth Rate of Basic Earnings per Share): This variable shows the percentage change in a company's earnings per share (EPS) from one year to the next. It reflects the company's growth or decline in profitability on a per-share basis over time.

DUPONT_ROE (DuPont ROE): The DuPont analysis breaks down ROE into three components: profitability, efficiency, and financial leverage. It helps investors understand what is driving changes in ROE over time and compare companies within the same industry.

There is a limitation in this dataset. Firstly, we get all prices of different stocks after stock price adjustment. Thus there are some inconsistency between our dataset and online data in Yahoo! Finance. For instance, let's consider the case of Arch Capital Group Ltd. (ACGL) stock, which changed from 104 to 33.730 during the period from April 29, 2011, to May 31, 2011. On May 12, 2011, ACGL underwent a 3:1 stock split. This means that for every share an investor held prior to the split, they received three shares after the split. Simultaneously, the price was adjusted proportionally to maintain the same total market value of the investor's holdings. Therefore, if the stock was trading at 90 per share before the split, it might now trade at 30 per share after the split. This adjustment explains why the price of the stock decreased from 104 to 33.730.

For the measurement used in this dataset, the World Bank follows a strict manner to collect data from authoritative sources. They will do data validation, standardization and dissemination, and then they aggregates and analyzes data to recognize patterns and offer guidance for policy decisions.

Table 1: A summary table of cleaned data

X	date	industry	pe
1	2010/2/26	Financials	4.599015
2	2010/3/31	Financials	4.740131
3	2010/4/30	Financials	4.698480
4	2010/5/31	Financials	4.518467
5	2010/6/30	Financials	4.578697
6	2010/7/30	Financials	4.809782

Table 1 shows first ten row of selected data from cleaned dataset, which includes 6 variables and 38988 observations. The variables are date (in month), industry, pe that is PE ratio (in percentage), ret that is the return rate (in percentage), marketvalue and optoebt (in dollars).

In this paper, we use R (R Core Team 2024), reshape2 (Wickham 2007), dplyr (Wickham et al. 2023), stargazer (Hlavac 2022), lmtest (Zeileis and Hothorn 2002) and ggplot2 (Wickham 2016).

Data Visualization

Whether the industry is an important factor of return rate

Now, we wonder if the category of industry influence the return rate.
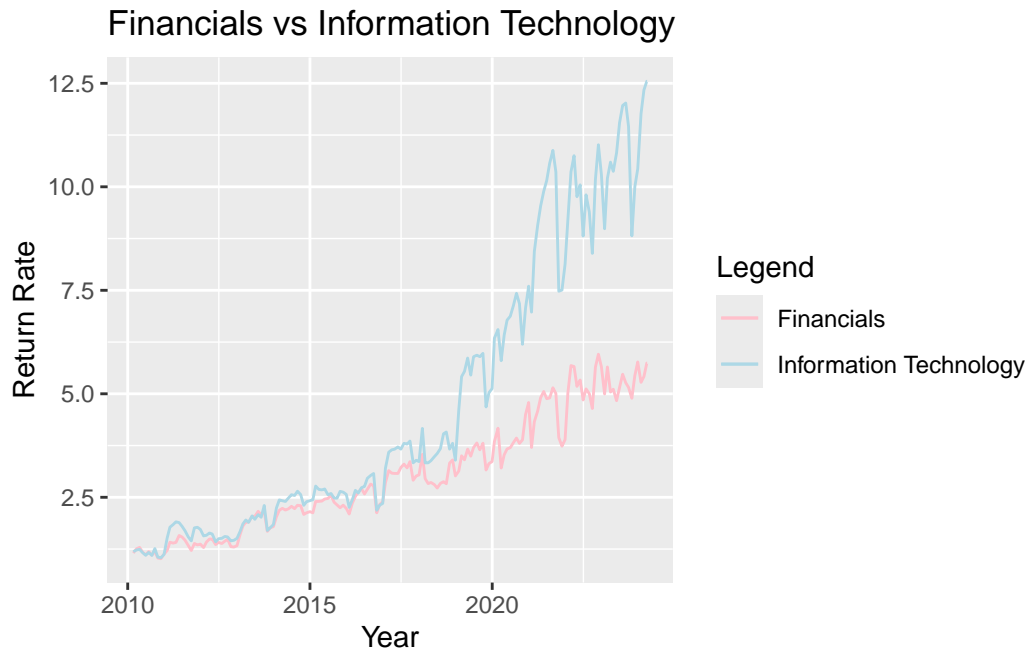
Figure 1: Overall increase in return rate from 2010 to 2024 for Financials and Information Technology

Figure 1 shows the overall trend of return rate of Financials and Information Technology from 2010 to 2024. The overall trend of this figure shows a increase in both industry over years. From 2010 to 2024, for Information Technology, the return rate of this industry increased from 5% in 2020 to 12.5% 2024 approximately. Compared to financials, whose return rate increased from 3.75% in 2020 to 5.625% 2024 approximately. Although the return rate fluctuates along with time, specially for Information Technology, it has a huge drop from 11% to 7.5% approximately around 2021. The overall tendency for return rate is rising and this finding can give the investors some confidence that they can trust the future of stock market.

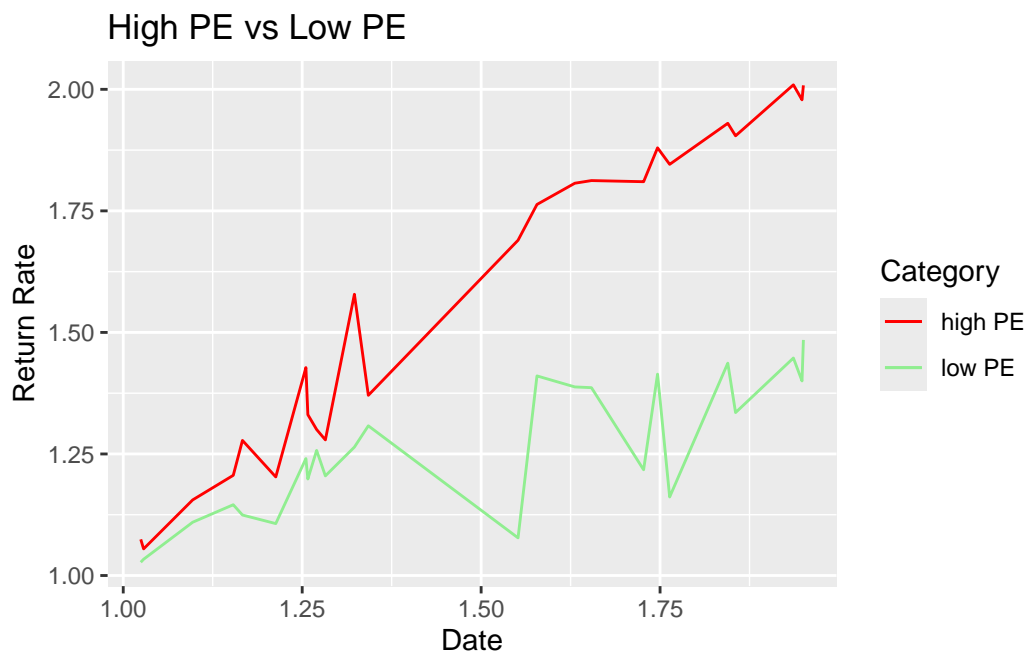**Whether the PE is an important factor of return**

Figure 2: Overall increase in return rate from 2010 to 2024 during the change of PE

In Figure 2, the x axis can be transferred to date of time that is from 2010 to 2024. It shows the overall trend of return rate of high PE stock and low PE stock from 2010 to 2024. The overall trend of this figure shows a increase in both industry over years. From 2010 to 2024, for low PE stock, the return rate increased from 1% in 2010 to 1.5% 2024 approximately. Compared to high PE stock, whose return rate increased from 1% in 2020 to 2% 2024 approximately. For this two conditions, even though there is a difference in overall rise in return rate, the scope of changing is not enormous. In this way, investors could consider low PE stock for some situation. In addition, we can find that high PE stock which in red line always has higher return rate than low PE stock that in green line. At the same time, one observation is fluctuations happened in low PE stock more frequently than high PE stock. This indicates the importance of maintaining the high PE, since high PE implies stability to some extent.

Model

After conducting exploratory analysis on the dataset, we've identified a positive correlation between stock return rates and industry and PE rate. Both variables exhibit an upward trend,

Table 2: Correlation matrix of four variables: return rate, market value, Operating Income
Before Extraordinary Items and PE rate

```
            ret marketvalue optoebt pe
ret           1          NA      NA NA
marketvalue  NA           1      NA NA
optoebt      NA          NA       1 NA
pe           NA          NA      NA  1
```

suggesting a potential linear relationship. Firstly, we want to check if there are correlations among variables we interested.

In Table 2, we can make sure that there is no correlation among four variables: return rate, market value, operating income before extraordinary items and PE rate.

Next we show several linear regression models:

Model 1

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

where:

- Y_1 is the return rate.
- $X_{1,2,3}$ are PE rate = 0.0587, marketvalue = 0.180, and optoebt = -0.00106
- β_0 is the intercept or constant term, which represents the expected value of return rate when all X variables equal to zero.
- β_1 is the slope coefficient or the estimated change in return rate for a one-unit increase in the PE rate.
- β_2 is the slope coefficient or the estimated change in return rate for a one-unit increase in the marketvalue.
- β_3 is the slope coefficient or the estimated change in return rate for a one-unit increase in the optoebt.
- ϵ is the error term or the deviation of the actual value of return rate from the predicted value based on the regression equation.

Model 2

$$Y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where:

- Y_2 is the return rate.
- $X_{1,2}$ are marketvalueand and optoebt.
- β_0 is the intercept or constant term, which represents the expected value of return rate when all X variables equal to zero.
- β_1 is the slope coefficient or the estimated change in return rate for a one-unit increase in the marketvalue.
- β_2 is the slope coefficient or the estimated change in return rate for a one-unit increase in the optoebt.
- ϵ is the error term or the deviation of the actual value of return rate from the predicted value based on the regression equation.

Model 3

$$Y_3 = \beta_0 + \beta_1 X_1 + \varepsilon$$

where:

- Y_3 is the return rate.
- X are marketvalue for stock.
- β_1 is the slope coefficient or the estimated change in return rate for a one-unit increase in the marketvalue.
- β_0 is the intercept or constant term, which represents the expected value of return rate when the marketvalue equal to zero.
- ϵ is the error term or the deviation of the actual value of return rate from the predicted value based on the regression equation.

Model 4

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

where:

- Y_1 is the return rate.
- $X_{1,2,3,4}$ are PE rate, marketvalue, optoebt and the fourth variable gets 1 or 0. Specifically, if the industry is Information Technology we get 1, otherwise we get 0.
- β_0 is the intercept or constant term, which represents the expected value of return rate when all X variables equal to zero.
- β_1 is the slope coefficient or the estimated change in return rate for a one-unit increase in the PE rate.
- β_2 is the slope coefficient or the estimated change in return rate for a one-unit increase in the marketvalue.
- β_3 is the slope coefficient or the estimated change in return rate for a one-unit increase in the optoebt.

- β_4 is the slope coefficient or the estimated change in return rate for a one-unit increase in the IT_industry.
- ϵ is the error term or the deviation of the actual value of return rate from the predicted value based on the regression equation.

Result

Model result

Table 3: Summary of model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0173780	0.0114446	1.5184482	0.1289406
pe	0.0587663	0.0104016	5.6497506	0.0000000
marketvalue	0.1801678	0.0113930	15.8139110	0.0000000
optoebt	-0.0010600	0.0114229	-0.0927935	0.9260699

Table 4: Summary of model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0166398	0.0114092	1.4584517	0.1447548
marketvalue	0.1808316	0.0113791	15.8915101	0.0000000
optoebt	-0.0006328	0.0114171	-0.0554262	0.9558003

Table 5: Summary of model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0000000	0.0050568	0.00000	1
marketvalue	0.1623643	0.0050569	32.10746	0

Table 6: Summary of model 4

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.7140173	0.0624484	27.4469303	0.0000000
pe	0.0004761	0.0000829	5.7440870	0.0000000
marketvalue	0.0000000	0.0000000	16.2374757	0.0000000
optoebt	-0.0000067	0.0001830	-0.0364113	0.9709553

Table 6: Summary of model 4

	Estimate	Std. Error	t value	Pr(>|t|)
IT_industry	1.6452908	0.1591668	10.3368949	0.0000000

Table 3 shows $\beta_0 = 0.0174$, $\beta_1 = 0.0588$, $\beta_2 = 0.180$, and $\beta_3 = $ -0.0011

Table 4 shows $\beta_0 = 0.0166$, $\beta_1 = 0.181$, and $\beta_2 = $ -0.0006

Table 5 shows $\beta_0 = 0.0000$, and $\beta_1 = 0.162$

Table 6 shows $\beta_0 = 1.7140$, $\beta_1 = 0.0005$, $\beta_2 = 0.0000$, and $\beta_3 = $ -0.0000067, and $\beta_4 = $ 1.6452908

Summary all the results from four tables, we can conclude that PE rate and category industry: Information Technology have positive linear correlation with return rate.
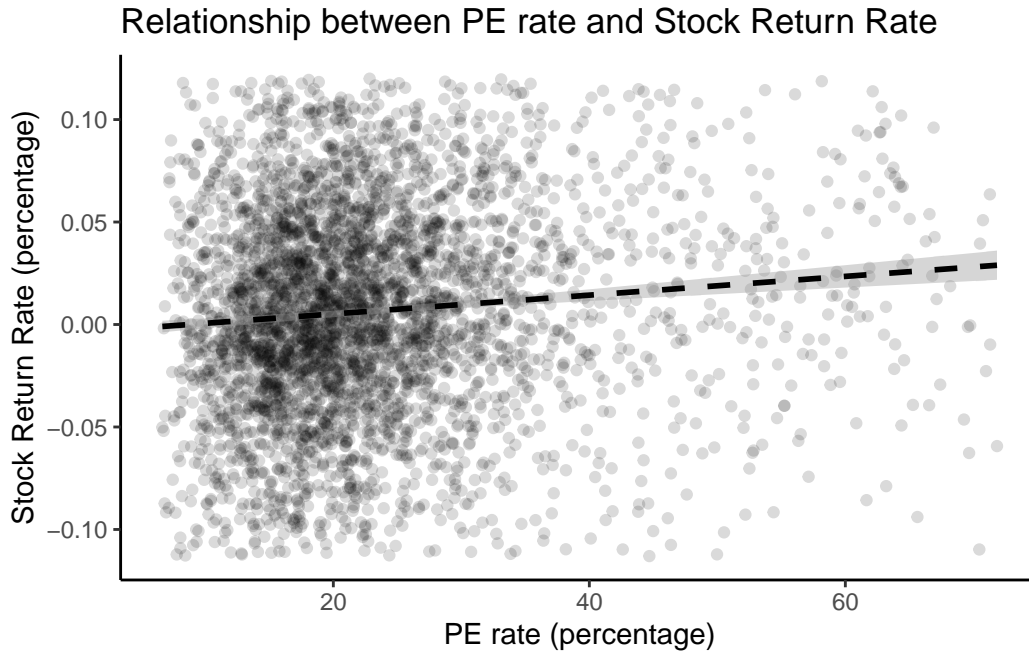

Figure 3: The model that shows the relationship between PE rate and Stock Return Rate

Figure 3 shows a fitted graph of the linear regression model illustrating the relationship between the Price-to-Earnings (PE) rate and Stock Return Rate. The plot shows the trend of stock return rate concerning changes in the PE rate. A positive relationship would suggest that as the PE ratio increases, so does the stock return rate. The regression line should confirm this trend, indicating a statistically significant positive slope.

Discussion

Evolution of Information Technology

Economic factors can influence stock return rate

Bias and ethical concerns

Weaknesses and next steps

Reference

Hlavac, Marek. 2022. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Social Policy Institute. https://CRAN.R-project.org/package=stargazer.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wikipedia contributors. 2024a. "Mutual Fund — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=Mutual_fund&oldid=1215343960.

———. 2024b. "World Bank — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=World_Bank&oldid=1219531902.

———. 2024c. "Yahoo! Finance — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=Yahoo!_Finance&oldid=1218474500.

Zeileis, Achim, and Torsten Hothorn. 2002. "Diagnostic Checking in Regression Relationships." *R News* 2 (3): 7–10. https://CRAN.R-project.org/doc/Rnews/.