Data Science Final Project

Daniel Glass

May 2024

1 Introduction

For this project, I have decided to use financial and stock market data to gain a better understanding of potential investments. I have not worked with financial or stock market data before, and I thought this would be a neat way to employ new skills while gaining additional insights that can be applied to my personal investing. The stock market can be a great tool for wealth generation and is one of the most accessible methods outside of traditional employment. However, information about the stock market is often sensationalized or obscured, making researching investments a difficult chore. With a few of the tools and skills learned in this class, I believe I can cut through much of the noise and find investments that better fit my interests, such as stocks that pay dividends and have low risk, with the potential to grow.

2 Literature Review

One reason I am interested in dividend stocks is that they appear to be more stable (Siddikee, 2018). This is because a consistent return in the form of a dividend incentivizes investors to keep the stock for a longer period, as opposed to other trading practices where investors prefer to move their investments around to meet an ever-changing market. Also, investors who are interested in dividend stocks may be following a similar idea of seeking long-term income generation. A paper from The Journal of Private Equity uses statistical analysis to study the impacts of a proposed daily dividend on stock volatility (Siddikee, 2018). The findings indicate that implementing a daily dividend policy did reduce stock return volatility for certain stocks, reinforcing the idea that frequent and consistent dividends can reduce volatility. In most cases, firms pay out dividends on a quarterly or monthly basis, so some of this effect is likely reduced, but perhaps still less volatile than firms that do not offer any dividend payments.

To drive this idea a little further, we can look to another paper from The Journal of Financial and Quantitative Analysis that looks closely at the relationship between a company's stock price and the dividends it pays out over time (Kim & Park, 2013). The paper shows that as fewer companies followed the traditional path of regularly paying out dividends, the tools used for predicting future stock prices based on those dividends weren't working as well. This further suggests that consistent dividend payments help lessen the volatility of a stock.

Dividend stocks attract retail investors more than other types of stocks according to a paper in The Journal of Finance, particularly older and low-income investors (Graham & Kumar, 2006). The paper also indicates that investors who focus on dividend yields may also exhibit longer holding periods. Older investors also tend to increase their position in a stock after positive dividend changes.

A paper published in the Journal of Finance brings up an interesting point. That the decision to pay dividends is not solely based on firm characteristics or traditional financial theories but also significantly influenced by market demand and investor sentiment towards dividend-paying stocks (Baker & Wurgler, 2004). The authors construct several stock price-based measures of investor demand for dividend payers to test this theory, and find that firms are more likely to initiate dividends when the market demand is high. Similarly, firms are more likely to omit dividends when the demand is low. The baby boomer generation entering retirement, or as the generation begins to pass away, could have a noticeable impact on a firm's decision to offer dividends. The investment strategy put forth in my paper may need to be reevaluated as the market potentially changes.

Another study from the Journal of Financial and Quantitative Analysis indicates that over the past several decades, there has been a decline in the propensity of firms to pay dividends (Banerjee et al., 2007). This trend is associated with improvements in stock market liquidity. As the market has become more liquid, with lower trading costs and higher trading activity, firms have become less inclined to distribute cash dividends. With the introduction of trading apps, the market could become even more liquid as there are fewer barriers to trading and education around the stock market. If this continues to affect a firm's propensity to pay dividends, this will further affect the degree at which this stock selecting method needs to be reevaluated.

3 Data

For this project, our primary focus lies in generating income through investments, thus we will primarily explore stocks that offer dividends. To identify these potential investments, we will commence with a sizable dataset and endeavor to streamline it. The initial data is sourced from stockanalysis.com.

The dataset includes information such as Symbol, Dividend Yield, and Payout Ratio, Dividend Ratio, Dividend Growth, Average 5 Year Growth, and Number of Year of Year Dividend Growth. Initially, we incorporate all stock tickers from the NASDAQ and NYSE, totaling around 5,600 stock tickers as of March 2024. Handling 5,600 stocks manually is quite cumbersome, and even with online tools, navigating through such a vast amount of data can be challenging. Moreover, the raw data from stockanalysis.com may be deceptive. For instance, a stock may be reported as paying dividends quarterly, but it has only commenced doing so in the past year, with no dividend payments prior to that. Consider an example where a stock has paid dividends only four times in the last decade. We will address these discrepancies in our base data as part of our project methodology. Our objective at this stage is to conduct data cleaning and filtering to refine the list to a more manageable size for further analysis.

4 Methods

Filtering for Dividend Information

We start by running a regression with the 3-year trend of each stock as the dependent variable, and Dividend Yield, Payout Ratio, Dividend Growth, Average 5-Year Growth, and Number of Years of Year-over-Year Dividend Growth as the independent variables. The regression model is as follows:

$$\begin{aligned}\text{Change 3Y} = \beta_0 &+ \beta_1 \times \text{Div. Yield} \\ &+ \beta_2 \times \text{Payout Ratio} + \beta_3 \times \text{Div. Growth} \\ &+ \beta_4 \times \text{Div. CAGR 5Y} + \beta_5 \times \text{Years} + \epsilon\end{aligned}$$

Our first area of focus is Dividend Growth and Average 5-Year Growth. Since both of these variables are positively correlated with positive stock growth with statistical significance, filtering any stocks with negative dividend growth will help us find stocks that will be great passive earners in the future, while also providing growth of the underlying investment. Next, we look at the dividend yield (Div. Yield) column. The dividend yield is how much the stock pays in dividends each year, as a percentage of the stock price. Meaning the higher the percentage, the more ”bang for your buck.”

$$\text{Dividend Yield} = \left(\frac{\text{Annual Dividends Per Share}}{\text{Stock Price}}\right) \times 100$$

We bring this into our R environment as a dataframe and remove all rows that have ”NA” or ”0” in the Div. Yield column. If this column does not contain any information, or has a value of 0, that indicates that no dividends are paid for that stock. By doing this, we already reduce the number of stocks that I am interested in from 5,600 to around 2,000.

Looking for Longevity

While some stocks have a high yield, in some cases, it could be too high. To get an idea of this, we look at the column Dividend Payout Ratio. This is the percentage of the firm’s profits that are paid out as dividends.

$$\text{Payout Ratio} = \left(\frac{\text{Dividends Per Share}}{\text{Earnings Per Share}} \right) \times 100$$

A ratio of 100 percent means that all of the profits are paid out as dividends, and anything over 100 percent likely means the firm will need to accrue debt to make dividend payments. This seems unsustainable and risky. Since I am looking for long-term income generation, I will next remove all tickers that have a payout ratio of over 100 percent. This brings us down to 1,725 tickers left. We need to reduce this further to make a more in-depth analysis of each stock practical.

Historical Consistency

Next, I want to look at the historical consistency of the dividend payments. The reasoning here is some dividend payments may be high, but are not paid consistently. The original data pulled from stockanalysis.com shows only the most recent year. As an example, the ticker EC has a very high dividend yield at around

30 percent, but only had consistent quarterly dividends in the past year. Before that, this firm has just one or two dividend payments per year.

To get an idea of the firm's consistency with dividend payments, we use the GetDividends function call in the quantmod R package. This package pulls information related to stocks from Yahoo Finance. In this case, we are pulling specifically dividend information. In this project, I pull dividend information for the last 4 years. If the company has had consistent dividends for the last 4 years, I have greater confidence in the company to continue the trend. This could easily be expanded to look further back. I have GetDividends pull dividend information starting January first, 2020, with the end date set to the local machine's system date. The dividend information is pulled for all the tickers in our data frame that have not yet been filtered out. Then we count how many entries/payments have been made and record that number as a new column. Most firms make dividend payments on a quarterly basis, meaning that there should be at least 16 payments made in the last 4 years. If the firm makes dividend payments on a monthly basis, we would expect to see at least 48 dividend payments. Since I am looking for income generation via dividends, I will begin to select my near-final stocks by taking the top 25 tickers with the highest dividend yield and another set of 25 with the highest number of dividend payments, meaning payments are made more frequently. The reason I choose 25 is that the free version of the API from AlphaVantage.com will let users run 25 queries each day, which will come into play in the next section.

Financial History

I would like to look at a firm's profits at this point. We want to look at firms that have reported profits for most years or quarters. We will be using the API from AlphaVantage.com. We will look at the EPS history for each ticker and count how many times a firm reported negative profits, break-even profits, or positive profits. We can do this by looking at Earnings Per Share (EPS). Earnings Per Share is the portion of profits that is allocated to each individual stock.

$$\text{Earnings Per Share} = \frac{\text{Net Income}}{\text{Shares Outstanding}}$$

With a free account with AlphaVantage.com, we can run an R code to gather historical EPS data for my two groups of near-final stock tickers. We do this by counting the total number of financial reporting periods and the number of those in which the firm reported negative profits. Then we divide the count of negative profits by the number of reporting periods. This will give us a ratio of how often that firm has reported negative profits.

$$\text{Negative Profits Ratio} = \frac{\text{Number of Negative Reports}}{\text{Total Number of Reports}}$$

I want to exclude any tickers with a ratio of 0.20 or higher, meaning that the firm has reported negative profits 20 percent or more of the time.

Plots for a Final Cut

”Now that we have narrowed down the 5600 or so stocks to just a few handfuls, it might be useful to make a plot. For both groups of tickers (top 25 in frequency and top 25 in growth), I want to plot the dividend growth on the x-axis and the dividend frequency on the y-axis. Since both of these factors are valuable for my objective, the farther away from the origin the ticker is, the more valuable it is to me. Let's start by plotting the set of high dividend frequency stocks. This is shown in Figure 1. In Figure 2, we will show the scatter plot for high growth stocks. The scatter plot helps us determine how we want to further refine our search. After viewing the plots, choose your own cutoff point to eliminate the points closest to the origin.

From here, we can combine the data frames into one final data frame that we will use to make our selection. After combining the data, we need to remove duplicates in the Symbol column. I was left with 18 stock tickers to research. This is a small enough group that individual research will not be troublesome.

One option from here is to use GetQuotes from the quantmod R package and ggplot to visualize the pricing trends for each of these 18 stocks. However, at this point, it may be easier to just find each ticker on Google or Yahoo Finance."

5 Findings

Our findings are 24 stock tickers that meet our criteria for healthy dividend Growth, strong historical dividend frequency, and positive historical profits. From here we can individually research each stock ticker, read news, and view the stock price trends.

6 conclusion

In summary, this project aimed to analyze financial data to identify potential dividend-paying stocks for investment. Despite limited prior experience, the study applied various filtering methods to narrow down a dataset of 5,600 stock tickers to 24 promising candidates with healthy dividend yields, consistent frequency, and positive historical profits. These findings will guide further research and informed decision-making in building an investment portfolio aligned with long-term financial goals.

References

Siddikee, M. N. (2018). Do Daily Dividends Reduce Stock Return Volatility and Value-at-Risk? *The Journal of Private Equity, 21*(4), 75-86. Euromoney Institutional Investor PLC. Retrieved from https://www.jstor.org/stable/10.2307/26497446

Kim, C.-J., & Park, C. (2013). Disappearing Dividends: Implications for the Dividend—Price Ratio and Return Predictability. *Journal of Money, Credit and Banking, 45*(5), 933-952. Wiley. Retrieved from https://www.jstor.org/stable/23463563

Graham, J. R., & Kumar, A. (2006). Do Dividend Clienteles Exist? Evidence on Dividend Preferences of Retail Investors. *The Journal of Finance, 61*(3), 1305-1336. Retrieved from https://www.jstor.org/stable/3699324

Baker, M., & Wurgler, J. (2004). A Catering Theory of Dividends. *The Journal of Finance, 59*(3), 1125-1165. Wiley for the American Finance Association. Retrieved from https://www.jstor.org/stable/3694732

Banerjee, S., Gatchev, V. A., & Spindt, P. A. (2007). Stock Market Liquidity and Firm Dividend Policy. *The Journal of Financial and Quantitative Analysis, 42*(2), 369-397. Cambridge University Press on behalf of the University of Washington School of Business Administration. Retrieved from https://www.jstor.org/stable/27647301

Linear Regression Results

.

	Estimate	Std. Error	t value	Pr($> \|t\|$)
(Intercept)	40.8235	5.0215	8.1300	1.26×10^{-15}
Div. Yield	-6.8227	0.9817	-6.9500	6.57×10^{-12}
Payout Ratio	0.0011	0.0021	0.5340	0.5933
Div. Growth	0.2419	0.0468	5.1640	2.91×10^{-7}
Div. CAGR 5Y	0.4851	0.1791	2.7080	0.0069
Years	-0.5631	0.1831	-3.0750	0.0022

Residual standard error: 77.66 on 1000 degrees of freedom

Multiple R-squared: 0.08041, Adjusted R-squared: 0.07581

F-statistic: 17.49 on 5 and 1000 DF, p-value: 2.2×10^{-16}

Figure 1) .

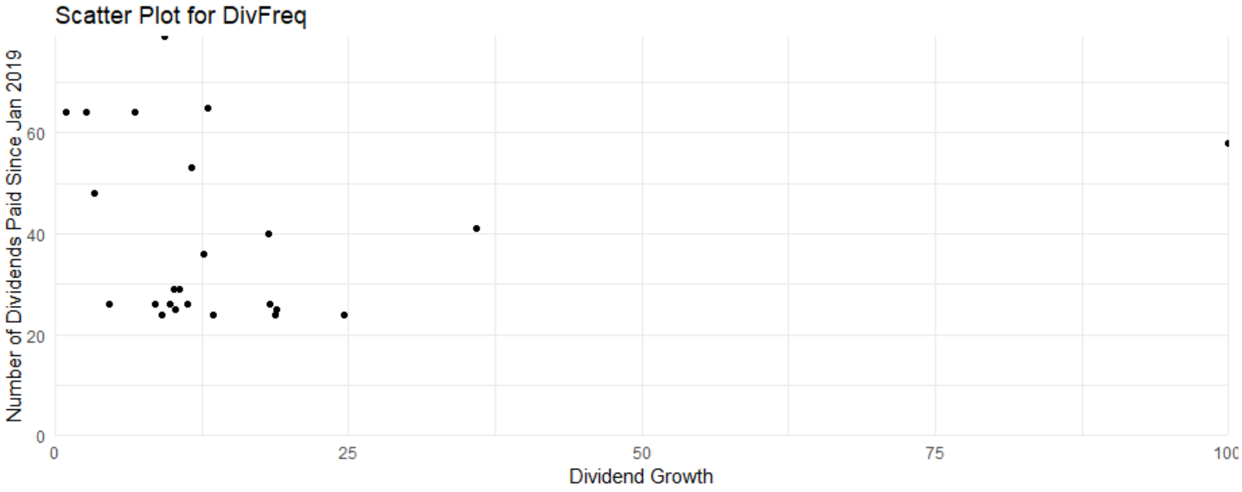
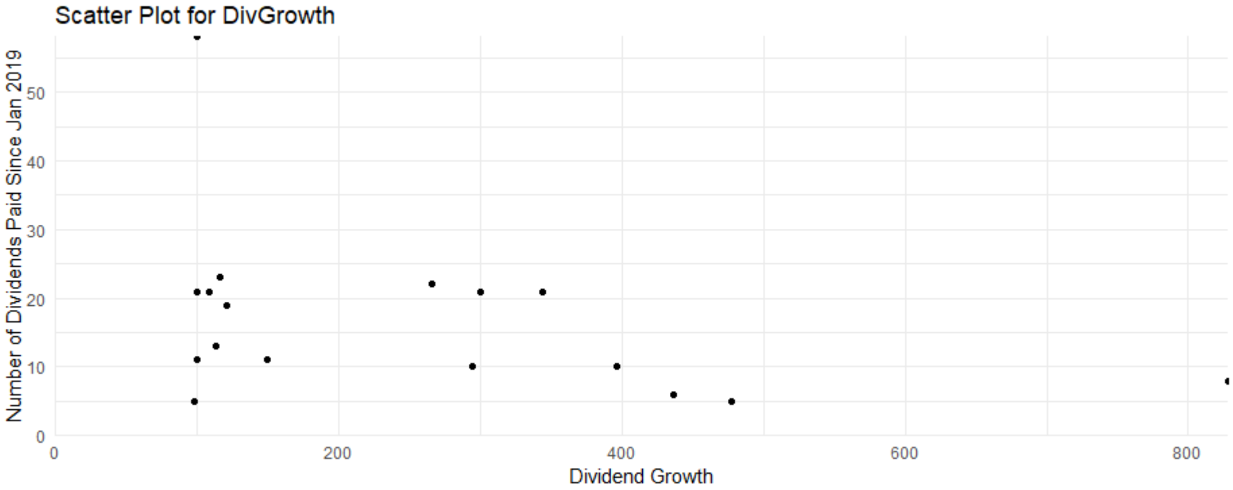

Scatter Plot for DivFreq
Number of Dividends Paid Since Jan 2019
0
20
40
60
0
25
50
75
100
Dividend Growth

Figure 2) .


Scatter Plot for DivGrowth
Number of Dividends Paid Since Jan 2019
0
10
20
30
40
50
0
200
400
600
800
Dividend Growth