IPO Underpricing Data Project

Brandon Parmanand

Fall 2022

Professor Dane Blevins

Introduction

An Initial Public Offering (IPO) is when stock of a private corporation is listed publicly. An IPO can be a traditional IPO or a spin-off. Spin-offs are usually part of corporations restructuring themselves with regards to the new spinoff company having a better chance of being more profitable alone. There are also three different ways to issue an IPO. SPAC, Traditional, and direct listing. Traditional IPOs are the most standard which could take years, SPAC or Special Purpose Acquisition Company that is more streamlined as it does not have business operations. Lastly, there is the direct listing which has no underwriting, and the company shares are listed directly. IPO Underpricing is what occurs when the issued price is lower than the closing price on the IPO date. What are the variables that can help determine underpricing? Can I select variables that can strictly predict it? The end product of the project will be a multiple regression model to provide insights into what variables or circumstances can assist with predicting underpricing in IPO stocks.

Dataset

The dataset contained 1996 observation of 443 variables including Underpricing. Through some initial data exploration 26 variables were selected, including the IPO date, total shares offered, price, the IPO auditor, whether it was in the different exchanges, the business states, and various financial measurements such as EBITDA, net income, book value of assets, operating cashflow, and revenue. I also included variables such as the type of IPO, as SPAC IPOs are designed to reduce volatility, and whether the IPO is a traditional or spin-offs. Additionally, I included whether the company is an emerging company and the close price of the IPO. I included the latest current close price and shares outstanding. Looking at these variables can help understand

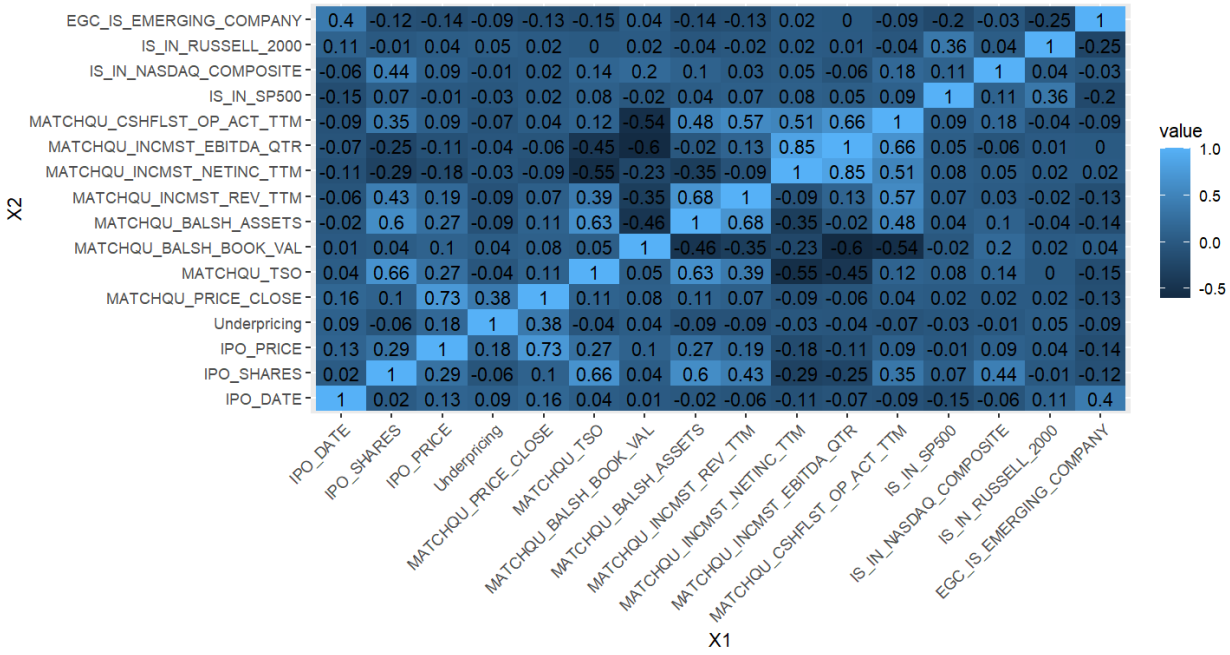
the look into the performance of the stock through time comparing the initial data to the current data.

Data Exploration

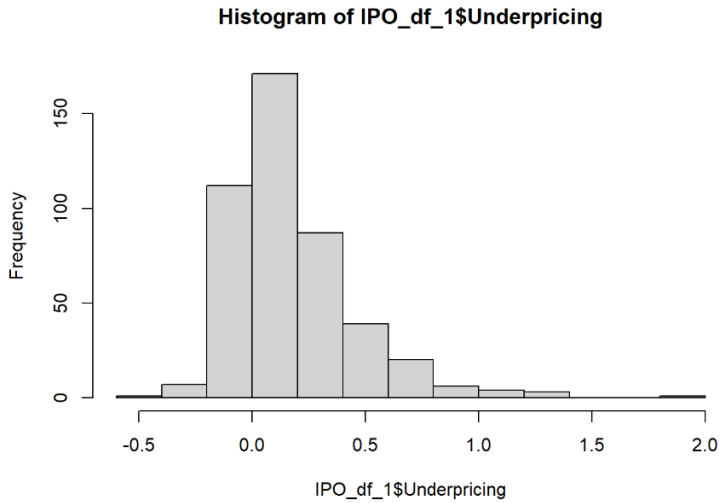
For in depth data exploration, I wanted to look at some summary statistics, data skewness, correlation, and any potential issues I would face while fitting my linear regression model. With the initial data exploration, specifically looking into brief summary statistics, I dropped the end of the quarter date (MATCHFY DATE QTR) and extraordinary income (PRIORQU INCMST EXTRAORDINARY ITEMS TTM) variables due to it having an excessive amount not applicable instances. Additionally, I made the decision to omit the current price (curr_price_close) and current shares outstanding (curr_tso) due to choosing to focus on a singular point in time rather than looking at underpricing over time. At this point I also removed variable indicating whether the IPO was in the DIJA 30 since there was no useful data in the dataset for it. I additionally added in State Region which is the regional part of the US where the corporation is located and also business state country to perhaps look at what regions can impact underpricing.

To create my correlation matrix, I removed the categorical variables and only included numerical variables. There mostly was not a lot of correlation between the independent variables. There was some notable correlation including negative correlation between Book Value and Assets, EBITDA, and Cash flow from operating income. EBITDA and Net Income are highly correlated with a correlation coefficient of 0.85. Since it is significantly greater than 0.7, I elected to remove EBITDA from the model as it had the most amount of not applicable values in the data. IPO shares and TSO for the end of the quarter along with IPO price and quarter price have a significant positive correlation but that is sensical as they are related. Additionally, I found that

all the financial metrics such as revenue, net income, cash flow, etc. have higher correlations with each other but not significantly higher than 0.7.



I also looked at the skewness of the variables with an emphasis of looking into Underpricing. It seems to be slightly skewed when looking at the histogram but still has a bell shape. However, book value and assets don't seem to be skewed from histograms.



Some issues I encountered were wanting to use whether the IPO issuance was a traditional IPO or spin-off IPO in order to determine the effects on underpricing. About 1% of the data was spin-off instances so therefore it would be difficult to conclude if spin-offs had a significantly different effect on underpricing than traditional. Additionally, less than 1% of the data were SPAC or direct listing IPO types so it would also be difficult to conclude whether the IPO type influenced underpricing.

Taking all the data exploration into consideration, some additional dimension reduction was needed. I removed the total shares outstanding for the quarter as it duplicates the IPO shares but had significantly more NAs in addition to removing the balance sheet book values as it had a significant number of NAs. Once all the variables were removed, I removed all instances of NA values and am left with 874 observations of 19 variables in the data frame.

Model

I ran the linear regression model with the 18 variables on Underpricing. This led to R-squared value of 0.4676 where 46.76% of underpricing variability is determined by the included variables. To generate a simplified but statistically significant model, I ran OLS stepwise models. Once running both the OLS Forward stepwise and backward stepwise models, they both ended with a 0.4657 R-squared. This determined that the simplified models are statistically better due to fewer variables needed but for the final model I chose to go with the forward stepwise model with 11 variables as it still contained key variables that are important theoretically to derive at underpricing.

Formula

Underpricing ~ MATCHQU_PRICE_CLOSE + IPO_PRICE + IPO_DESC + AUD_NAME_IPO_DATE + IS_IN_RUSSELL_2000 + BUS_STATE + IPO_SHARES + IS_IN_SP500 + IPO_TYPE + EGC_IS_EMERGING_COMPANY + IPO_DATE

Results

I expected that spin-offs have a greater chance of being underpriced than IPOs due to the status of the parent company which can drive investors toward it. In the regression model, that seems to be the case with a p-value of zero concluding that there is a correlation between the type of IPO. However, more spin-off data would be needed to accurately gauge the effect. Additionally the price of the IPO also has a strong correlation with the underpricing. Whether or not the IPO was in the Russel 2000 was also a good indicator of the underpricing.

Conclusion

With an R^2 score of 0.46, I do not think the variables chosen can efficiently predict underpricing. If I were to continue exploring this project, I would look into breaking up the states into regions, breaking out the dates into months or years to see what effect that would have.