Understanding the Effects of Financial Ratios on Stock Performance

(Team #42) Xi Lu, Guchuan Gao, Ligeng Peng, Daniel Tong & Isabelle Victoria Zelazny

Abstract

Understanding how to make successful investments is important for both investment businesses and individuals, as effective allocation of capital is extremely important for their success and financial health. This study improves the understanding of the relationship between financial ratios and stock prices by using data from the Dow 30 stocks. Random Forest was used to determine the top five most important ratios for predicting a stock's performance. Through factor analysis, we have identified five key factors that drive stock performance with a strong R2 of 0.964 and these factors are ptb(Price/Book), bm(Book/Market), debt_capital(Total Debt/Capital), equity_invcap(Common Equity/Invested Capital), and totdebt_invcap(Total Debt/Invested Capital). By implementing factor analysis to predict future stock performance, we can make more informed investment decisions and potentially generate higher returns on our investments.

Introduction

A company's stock price fluctuates every day by market forces, but it is also strongly influenced by its own financial fundamentals. Financial ratios provide investors with opportunities to evaluate a company's true performance and are often crucial fundamentals used in stock analysis. According to Glenn Wilkins, "6 Basic Financial Ratios and What They Reveal" article in Investopedia 2022[1], the following six financial ratios are typically expected to influence stock prices: working capital ratio, the quick ratio, earnings per share (EPS), price-earnings (P/E), debt-to-equity, and return on equity (ROE). By understanding which financial ratios best predict the returns of a stock, investors can make better informed, data driven decisions. The results of this analysis can help businesses seeking investors tailor company financial ratios to align with ones that predict better performance [2], while also helping both companies and individuals with their own investment decisions.

Methodology

The key objective of this project is to analyze financial ratios' correlations with stock prices among Dow 30 stocks and use factor driven analysis to understand how financial ratios drive stock performance in different industries. We approached this topic by putting forward a primary research question, which of the financial ratios (dependent variable) best predicts the returns (independent variables) of a stock from the Dow 30? To further enhance the robustness of our study, we also are attempting to answer the following supporting research questions; which combinations (interaction terms) of financial ratios best predicts the return of a stock from the Dow 30, which combinations of financial ratios suggest poor performer/good performer for a stock in the next 3 months, and do different industries have different financial ratios as their predictors?

This study uses R for data analysis, visualization, and model predication. The researchers first cleaned the data by following a four-step approach covered in the Data section. Next, potential multicollinearity between financial ratios was evaluated and then PCA and Lasso regression were considered for dimensionality reduction. Dimensionality reduction ensured that key financial ratios that have major impacts on the stock price can be identified and included as inputs into the modelling. Lasso regression was selected as the method for dimensionality reduction and explanations for the exclusion of PCA are discussed in the Modelling section.

We reference the existing models to predict stock prices with Linear model and Vector support model [4]. Our approach (Random Forest model) is novel for several reasons. First, Random Forest is a machine learning algorithm that is particularly well-suited to handling complex data with multiple input variables, such as financial ratios. Second, the high R squared value of 0.964 suggests our model can explain a significant amount of the variation in stock prices based on the input variables. Compared to the existing linear regression model with an R squared value of 0.73 and support vector regression model with an R squared value of 0.93, our model can explain more of the variation in stock prices. This suggests that our model is more accurate and reliable to predict stock prices than the existing models. The Future Work sections includes future includes future methodologies. We also consider creating multi linear regression models to establish correlations between financial ratios and stock prices [3] for our next step to further evaluate if our model is the best model for this problem.

Data

Datasets: Two datasets were used in this study (the first few rows of each dataset are shown below). First, the independent variables which were obtained from the Wharton Research Data Services (https://wrds-www.wharton.upenn.edu/) and includes the Dow 30 stocks' 75 Financial ratios by month from Jan 2010 to Dec 2022.

	gvkey	permno	adate	qdate	public_date	CAPEI	bm	evm	pe_op_basic	pe_op_dil	...	adv_sale	staff_sale	accrual	ptb	PEG_trailing
0	12141	10107	20090630.0	20090930	20100131	17.445	0.181	9.091	17.835	17.949	...	0.024	0.0	-0.111	5.975	2.039
1	12141	10107	20090630.0	20091231	20100228	17.365	0.166	9.138	15.332	15.582	...	0.024	0.0	-0.063	5.630	0.910
2	12141	10107	20090630.0	20091231	20100331	17.722	0.166	9.138	15.662	15.917	...	0.024	0.0	-0.063	5.746	0.930
3	12141	10107	20090630.0	20091231	20100430	18.480	0.166	9.138	16.329	16.595	...	0.024	0.0	-0.063	5.992	0.970
4	12141	10107	20090630.0	20100331	20100531	15.296	0.181	9.939	13.163	13.368	...	0.024	0.0	-0.064	4.859	1.119

5 rows × 78 columns

Second, the dependent variable which were obtained from Yahoo Finance via R tidyquant and includes the Dow 30 stock prices by month from Jan 2010 to Dec 2022.

	symbol	date	open	high	low	close	volume	adjusted
1	AAPL	2010-01-04	7.622500	7.660714	7.585000	7.643214	493729600	6.505279
2	AAPL	2010-01-05	7.664286	7.699643	7.616071	7.656429	601904800	6.516529
3	AAPL	2010-01-06	7.[illegible]5429	7.686786	7.526786	7.534643	552160000	6.412873
4	AAPL	2010-01-07	7.562500	7.571429	7.466071	7.520714	477131200	6.401017
5	AAPL	2010-01-08	7.510714	7.571429	7.466429	7.570714	447610800	6.443573
6	AAPL	2010-01-11	7.600000	7.607143	7.444643	7.503929	462229600	6.386732
7	AAPL	2010-01-12	7.471071	7.491786	7.372143	7.418571	594459600	6.314082
8	AAPL	2010-01-13	7.423929	7.533214	7.289286	7.523214	605892000	6.403145
9	AAPL	2010-01-14	7.503929	7.516429	7.465000	7.479643	432894000	6.366061
10	AAPL	2010-01-15	7.533214	7.557143	7.352500	7.354643	594067600	6.259671

Data Cleaning: A four step approach was used to clean the data. Step 1: Remove rows with N/A close prices as it is required as a dependent variable for the regression model. Step 2: Remove Financial ratios with >500NAs, which translates to ~15% of each financial ratio's 3255 total observations (30 stocks over ~109 months). Removing these ratios is deemed more beneficial than imputing them, as it may introduce additional bias in the analysis. Step 3: Imputing remaining missing financial ratios with each stock's closest adjacent financial ratio. The imputation method is stated as each stock's closest adjacent financial ratios. Step 4: 4 financial ratios are missing for the entire duration for JPM and TRV. They are removed from the model.

Modelling

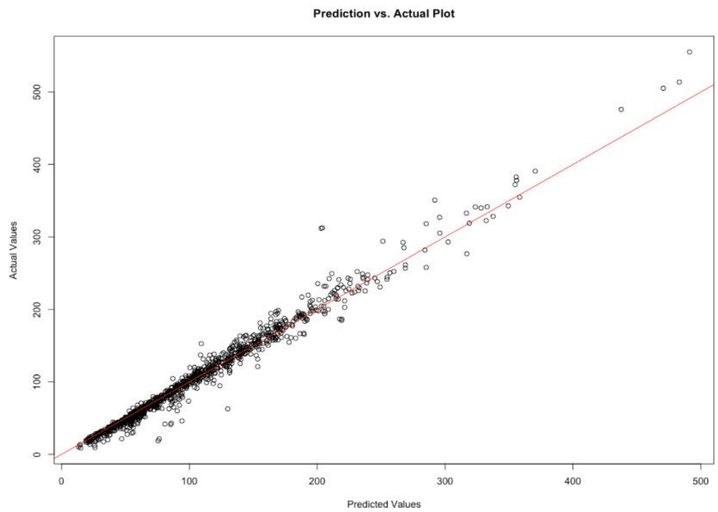
Lasso Regression: Cross validation was used to search for the best lambda which was used to create the lasso regression model. The fitted model was used to predict the original dataset and generated an R-squared value of 0.534.

PCA: Features were projected into a new dimension space, creating principal components that are orthogonal to one another. The components were ranked based on their explained variance in the data and analyzed by their specific feature construction.

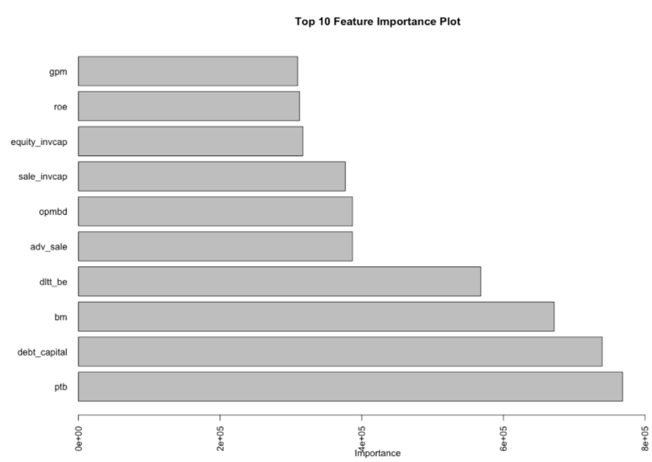
Random Forest Model: Many decision trees were used, and their outputs were combined. Random Forest can capture non-linear relationships between the input variables and the target variable. This model also provided feature importance, which can help identify which input variables are most predictive of the target variable

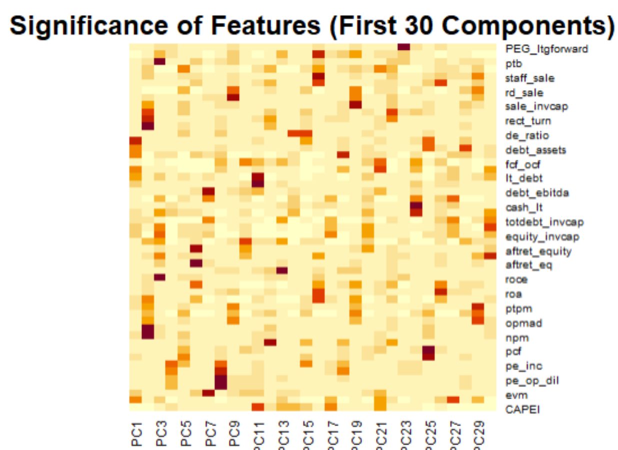
Preliminary Results

An R-squared value of 0.964 is a very promising result and suggests that the Random Forest model has strong predictive power for the given data. See below for Random Forest Prediction vs Actual Plot:



There is an importance function in R's Random Forest package which was used to extract variable importance measures for the model. Variable importance is a measure of how useful a particular predictor variable is in predicting the response variable. Below are the top 10 most important financial ratios from the importance function.





For the purposes of feature selection, PCA did not yield satisfactory results, as the amount of variance explained was split quite evenly among most of the principal components, with hugely different combinations of features for each component (heatmap shown above). The heatmap visualizes the significance of each financial ratio relative to the first 30 principal components normalized by the percentage of variance each component explains such that we can compare between components. It is clear from the visualization that it is exceedingly difficult to identify which features overall are contributing the most to the variance in the data. This was to be expected, as PCA is not considered a good method of feature selection, and the principal components it creates are more often used as a new feature space entirely. Lasso regression's R-squared value indicates only 50% of the stock price variability can be explained by a total of 69 financial ratios. By examining some of the largest coefficients of the features, accruals /average assets, Total debt/Total Assets are found to have the largest positive impact on stock prices while Advertising Expense/Sale, cash flow margin and capital ratio are found to have the largest negative impact on stock prices.

Discussion

Among the 5 key financial ratios identified by the random forest model, price/book and book and book/market ratio reflect the attractiveness and overall stock market sentiment towards as stock. Meanwhile, the remaining 3 ratios are more related to the stock fundamentals. For example, the proportion of debt a company uses to finance its assets relative to the equity may indicate if a company is highly leveraged and its risk of insolvency. These two different types of financial ratios are often related to the two common techniques used in stock analysis, fundamental vs technical analysis. The former is more useful for long-term investments while the latter is more useful for short-term trading and market timing. Further studies can be conducted to understand how these two types of financial ratios impact stock prices in the short and long term. However, this will fall outside of our project scope.

Overall, the 5 financial ratios identified through our analysis are different than those identified through the research findings from Investopedia. However, our analysis is based on the Dow 30 stocks and serves only as a good indicator of a very small sample size. This certainly also serves as a good reminder that each company may have different financial ratios that may influence their stock performance. As a stock investor, it is important to analyze each stock instead of following general investment advice. As such, an area of focus in our next step will be understanding how companies in different industries may have unique financial ratios that are critical to their stock performance.

Future Work (Esther)

Task	Task Details	Assigned To	Progress	Start	End
Topic & Proposal	Topic Research and Selection	Everyone	100%	2-16-23	3-02-23
	Dataset Research and Selection	Everyone	100%	2-16-23	3-02-23
	Proposal	Everyone	100%	3-02-23	3-12-23
	Proposal Presentation	Everyone	100%	3-12-23	3-26-23
Data	Data Collection, Cleaning and Prep.	Everyone	100%	3-12-23	3-26-23
	Data Exploration/ Variables Selection	Everyone	100%	3-12-23	3-26-23
	Analysis/Modeling	Everyone	100%	3-19-23	4-02-23
	Progress Report	Everyone	100%	3-26-23	4-02-23
Evaluation & Visualization	Experiment and Evaluation	Everyone	30%	4-02-23	4-09-23
	Final Report	Everyone	0%	4-09-23	4-16-23
	Final Presentation	Everyone	0%	4-09-23	4-19-23

We will review project milestones such as data collection, cleaning, variable selection, modeling, visualization, and completion of the progress report as part of our "midterm exam." As for now, we are ahead of schedule, and have a promising model. We are looking forward to adding some additional models in the next step. Some considerations include create clusters among Dow 30 stocks to understand patterns , and use back-testing to assess the viability of a trading strategy or pricing model by applying CUSUM analysis on historical financial ratios combined with regression models to see how stock price would have played out retrospectively using historical time series data

As part of our final deliverables, our "final exam" will focus on finishing the final report, creating a solid model to categorize financial ratios into 5 factors and determine which factor plays the most significant role in stock performance. The project activities and progress are tracked using a Gantt chart and Azure DevOps. We also have weekly meetings to show the progress to the team members.

References

[1] Glenn Wilkins. 6 Basic Financial Ratios and What They Reveal. Investopedia. 2022. https://www.investopedia.com/financial-edge/0910/6-basic-financial-ratios-and-what-they-tell-you.aspx#:~:text=There%20are%20six%20basic%20ratios,return%20on%20equity%20(ROE).

[2] B Korcan Ak, Patricia M Dechow, Yuan Sun, and Annika Yu Wang. (2013) The use of financial ratio models to help investors predict and interpret significant corporate events. Sage Journals. Vol 38. Issue 3

[3] Subramanian, K. & Prabhu, M. K. (2014). Predicting Stock Prices Using Financial Ratios: A Multiple Linear Regression Analysis. Journal of Finance and Accounting, 2(7), 383-387.

[4] Vaishnavi Gururaj, Shriya V R and Dr. Ashwini K (2019). Stock Market Prediction using Linear Regression and Support Vector Machines. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 8, pp. 1931-1934