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 0.975 and these factors are *ptb* (Price / Book), *debt_capital* (Total Debt / Capital), *bm* (Book / Market), dltt_be (Long-term Debt/Book Equity), and opmbd (Operating Profit Margin Before Depreciation). An evaluation on the patterns of the financial ratios across general industries and individual companies was done and showed that sales and earning related ratios were highly correlated to stock price and for stocks that were less correlated to their sales/earnings ratio some had cash balance/total liability ratio as their most important financial ratio. Other findings from this assessment showed that debt/asset ratio was important for financial services sector stocks, and price to booking ratio was the most common among the top five financial ratios for each stock. 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 Wilkins (2022), 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 (Ak et al., 2013), 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 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:

- How do general market predictors compare to predictors for individual stocks?
- 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 Regression model and Vector support machines (Gururaj et al., 2019). 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 0.975 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. Following this, we evaluated the top 5 financial ratios for each of the Dow 30 stocks and did an in-depth analysis on the patterns of these financial ratios for individual stocks as well as across industries. A comparison was made to the work done by Subramanian & Prabhu, 2014.

Data

1. **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

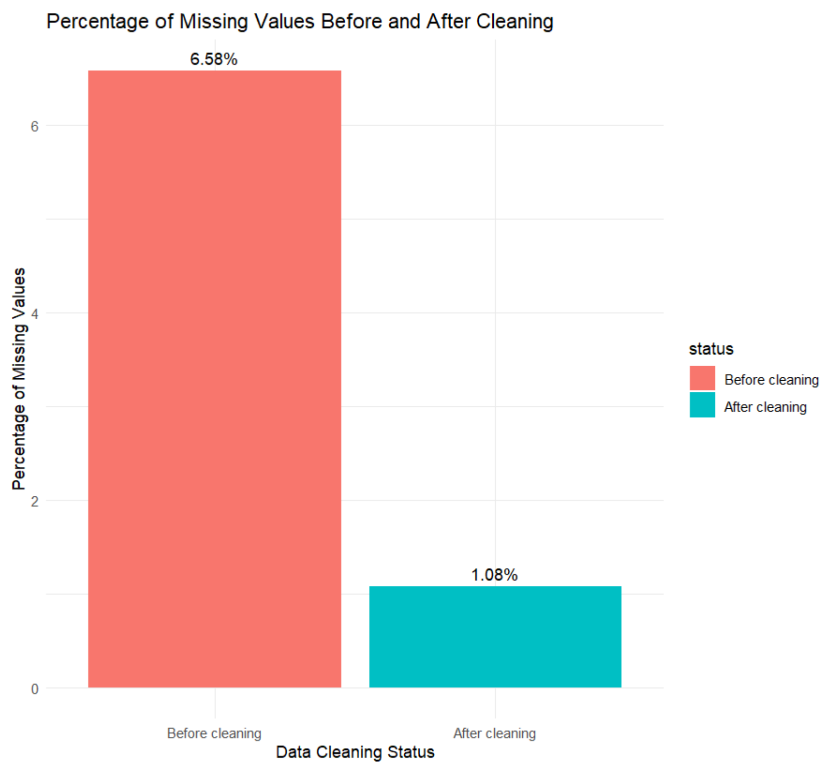
5 rows × 78 columns

    - 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.

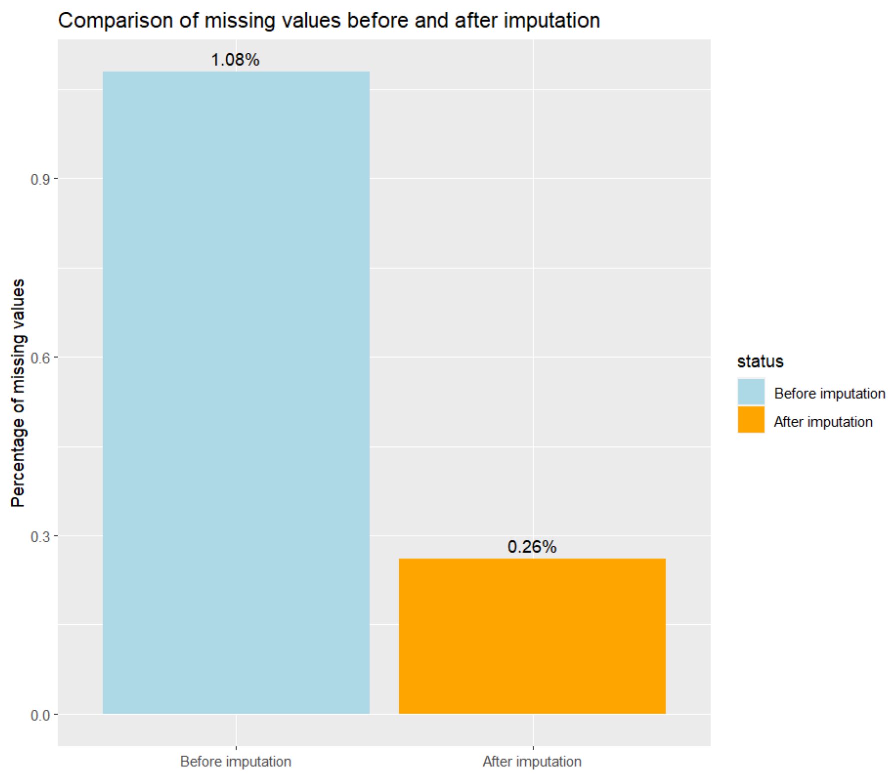
	symbol	date	open	high	low	close	volume	adjusted
1	AAPL	2010-01-04	7.622500	7.660714	7.585000	7.643214	493729600	6.505279
2	AAPL	2010-01-05	7.664286	7.699643	7.616071	7.656429	601904800	6.516529
3	AAPL	2010-01-06	7.[illegible]5429	7.686786	7.526786	7.534643	552160000	6.412873
4	AAPL	2010-01-07	7.562500	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

2. **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.

After taking these steps, we believe we have effectively improved the quality of our dataset. For example, below is a chart of the percentage of missing values in the dataset before and after we performed the data cleaning in Step 2. As the chart shows, removing financial ratios with >500 NAs is necessary and beneficial because it significantly reduces the percentage of missing values from 6.58% to 1.08%, indicating that the missing values were concentrated in a few specific columns. This reduction in missing values can improve the quality of our data and the accuracy of any subsequent model training.



Another example shown below indicates the effectiveness of the imputation missing values in Step 3. Before the imputation, there was a significant percentage of missing values (1.08%) in the original dataset, which could potentially affect the accuracy of any analysis. However, after performing the imputation, the percentage of missing values reduced to 0.26%, indicating the effectiveness of the imputation method and the necessity of performing it.



Modeling

In this section, we describe the process of selecting the final model for our analysis. We started by exploring different modeling techniques and evaluating their performance using appropriate metrics. We then used a systematic approach to select the best-performing model based on the R-squared value. We first normalize the price column as the stock price has a minimum value of $8.39, and maximum value of $555.15. Normalization will help us to have a better understanding of the model performance.

1. Model experiments

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.

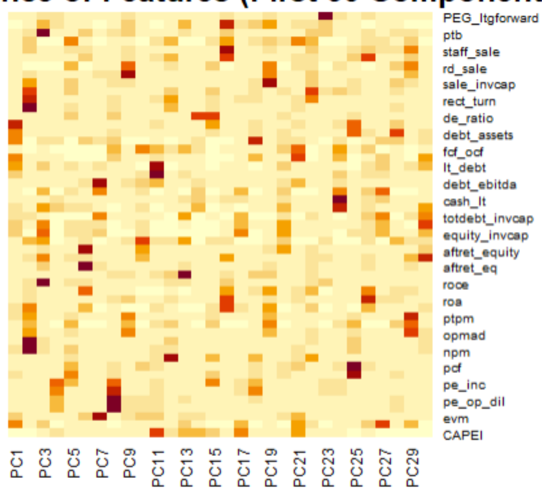
Lasso Regression: Cross validation was used to search for the best lambda which was used to create the lasso regression model. The fitted model was used to predict the original dataset and generated an R-squared value of 0.534.

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. We started by splitting the data into 2/3 training and 1/3 testing sets. We then trained the model on the training set and evaluated its performance on the testing set using R-squared.

2. Model performances

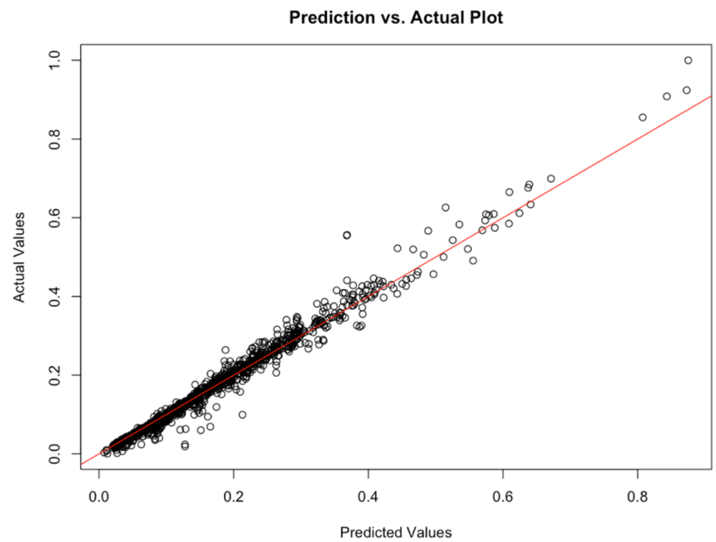
For the purposes of identifying feature importance, PCA did not yield satisfactory results, as the amount of variance explained was split quite evenly among most of the principal components, with vastly different combinations of important features for each component (heatmap shown below). The heatmap 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 generally not a reliable way of determining feature importance, and the principal components it creates are more often used as a new feature space entirely. We will move on to lasso regression as our next potential technique.

Significance of Features (First 30 Components)

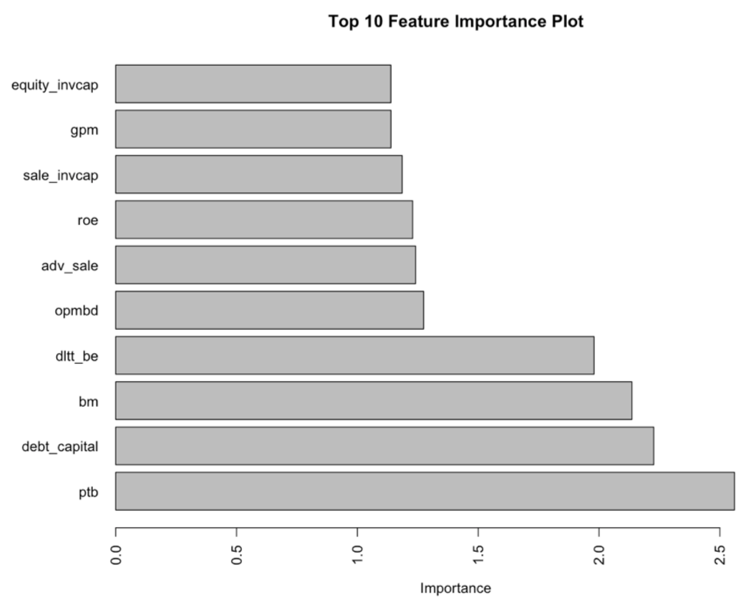


Lasso regression's R-squared value indicates only 53.4% of the stock price variability can be explained by a total of 69 financial ratios. By examining some of the largest coefficients of the features, accruals/average assets, Total debt/Total Assets are found to have the largest positive impact on stock prices while Advertising Expense/Sale, cash flow margin and capital ratio are found to have the largest negative impact on stock prices. We move on to Random Forest, so that our identified important ratios may be more accurate.

Our Random Forest model achieved an R-squared value of 0.975, which 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. We will use these results to proceed with our analysis.



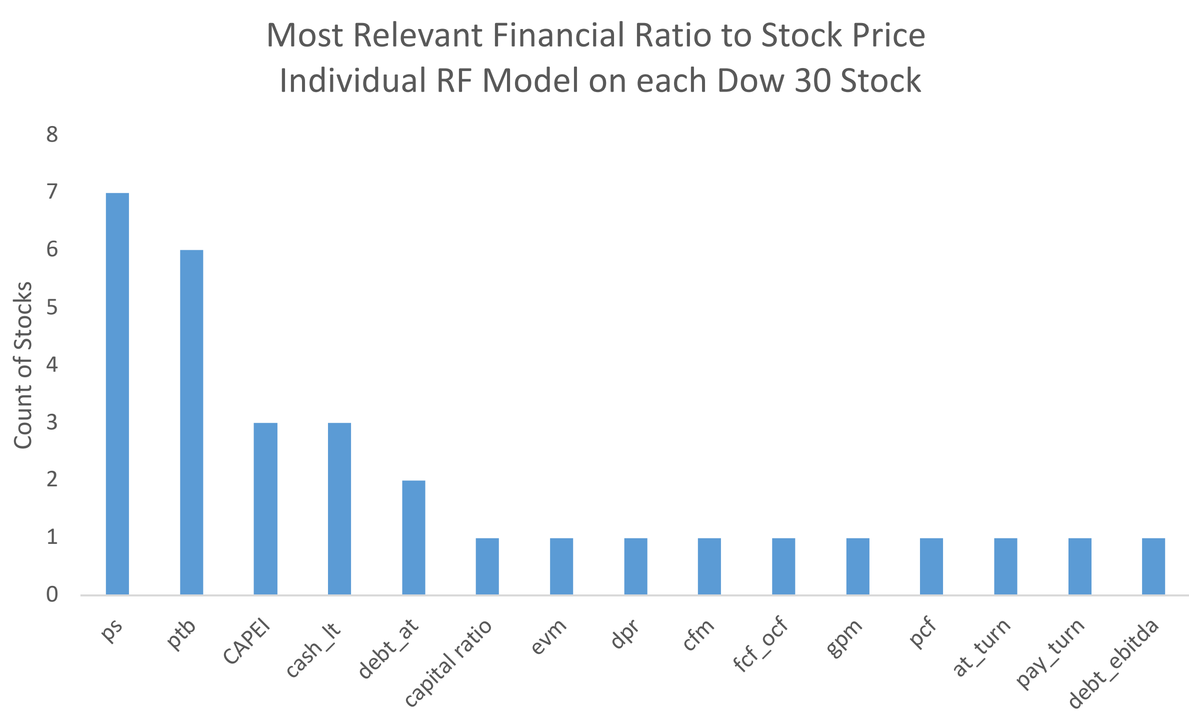
Discussion

Among the top 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 versus 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 small sample size. Our evaluation of the top 5 financial ratios for each of the Dow 30 stocks as well as for industry specific groupings of stock shows that patterns exit among stocks as a whole, industry specific, and also vary for individual stocks and this is discussed further in the next section. 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.

Experiment & Evaluation

Now that we've identified the top 5 financial ratios for the general market, we want to see what the results may look like for each individual stock. We split the data by stock and train individual random forest models for each subset, identifying new sets of top 5 ratios for each stock. The chart below summarizes how many times each ratio was found to be the most relevant ratio for a specific stock.



The following table displays all top 5 ratios found for each stock.

Random Forest Top 5 Financial Ratios* for DOW 30 Stocks**

Stock Ticker	AAPL	AMGN	AXP	BA	C	CAT	CRM
Top 5 ratios	evm	dpr	debt_at	ps	ps	ps	cfm
	equity_invcap	bm	sale_invcap	roce	ptb	ptb	de_ratio
	roe	opmbd	dltt_be	opmbd	dpr	cash_lt	debt_assets
	pcf	adv_sale	ptb	opmad	rect_turn	rd_sale	sale_equity
	ptb	ps	debt_invcap	cash_debt	sale_equity	CAPEI	accrual

Stock Ticker	CSCO	CVX	DIS	DOW	GS	HD	HON
Top 5 ratios	cash_lt	fcf_ocf	gpm	cash_lt	debt_at	debt_ebita	pay_turn
	bm	ptb	evm	pcf	CAPEI	debt_at	at_turn
	pay_turn	opmbd	pay_turn	pe_inc	staff_Sale	roe	ptb
	ptb	CAPEI	ps	pe_exi	cfm	efftax	ps
	de_ratio	npm	rect_turn	ptb	lt_debt	capital_ratio	sale_invcap

Stock Ticker	IBM	INTC	JNJ	JPM	KO	MCD	MMM
Top 5 ratios	pcf	cash_lt	ptb	ptb	CAPEI	ps	ptb
	ps	adv_Sale	lt_debt	bm	ps	opmad	bm
	ptb	debt_assets	pe_op_di	ps	fcf_ocf	CAPEI	ps
	opmad	de_ratio	pe_op_basic	debt_at	ptb	debt_ebitda	rd_sale
	bm	Gprof	CAPEI	efftax	pe_op_dil	opmbd	evm

Stock Ticker	MRK	MSFT	NKE	PG	TRV	UNH	V
Top 5 ratios	pcf	cash_lt	ptb	ptb	CAPEI	ps	ptb
	ps	adv_Sale	lt_debt	bm	ps	opmad	bm
	ptb	debt_assets	pe_op_di	ps	fcf_ocf	CAPEI	ps
	opmad	de_ratio	pe_op_basic	debt_at	ptb	debt_ebitda	rd_sale
	bm	Gprof	CAPEI	efftax	pe_op_dil	opmbd	evm

Stock Ticker	VZ	WBA	WMT				
Top 5 ratios	ps	ptb	CAPEI				
	pcf	CAPEI	ps				
	ptb	ps	de_ratio				
	bm	sale_equity	debt_capital				
	rect_turn	debt_capital	ptb				

*Reference appendix for detailed description of each financial ratio

**A total of 32 stocks have been listed in DOW 30 at one point from the analysis period from 2010 – 2022

An in-depth analysis is then conducted to understand the patterns of these financial ratios across the general industries as well as individual companies, based upon Random Forest model output table above.

- Sales and earnings-related ratios such as ps(prices/sales) and CAPEI (price/earnings) are highly correlated to stocks prices. This aligns with the consensus from many investment websites, which typically list PE or PS ratio as key indicators of a stock's performance. 10 companies in DOW30 have either ps or CAPEI as their most influential financial ratio while 24 of them have these 2 ratios in their top 5 rankings.
- Among companies whose stock prices are less correlated to their sales/ earnings ratios, 3 of them have cash balance/total liability ratio(cash_lit) as their most important financial ratio. These companies are Cisco System Inc, Dow Inc and Intel Corp. Since this ratio is calculated by adding a company's total cash reserve and dividing that by its total current liability, it measures a company's liquidity and shows a company's ability to cover its short-term obligations using only cash and cash equivalents. For investors who are interested in buying these 3 companies, it is important to pay attention to their liquidity ratios as they have proved to be highly influential in driving their stock prices for the past 20 years.
- Debt/Asset ratio(debt_at) is a particularly important stock price indicator for companies in the financial services sector. 4 of these companies in DOW30, American Express Co, Goldman Sachs Group Inc, JPMorgan Chase & Co, and Visa Inc have total debt/ total asset ratio in their top 5 rankings. It is also worth noting that debt_at ratio does not typically serve as an important ratio for many of the other DOW 30 stocks.
- Out of the original top 5 ratios found for the general market, Price to booking ratio(ptb) was by far the most common among the top 5s for each stock, appearing in 61.29% of the stocks. Book/market ratio (bm) was next at 29.03%, while the remaining 3 (debt_capital, dltt_be, opmbd) were much rarer. Out of those 5 factors, it seems that ptb alone was significantly present among the individual top 5s. Price to booking ratio(ptb), which measures the market' s valuation of a accompany relative to its book values, tends to be a prominent factor in technology stocks including Apple Inc, Microsoft Corp and IBM. This could speak to the fact that many of these technology companies experienced various growth phases in the past 2 decades and their stock prices tend to deviate away from their companies' fundamental values more often due to fluctuation of investor sentiment, and could explain its frequency among many of the top 5s.
- Some companies have unique financial ratios in their top rankings. This includes gross profit margin for Walt Disney Co, asset turnover ratio for Travelers Companies Inc, payables turnover for Honeywell International Inc and capital ratio for Nike Inc. This certainly serves as a reminder to investors that each company's circumstance may lead to its stock price being impacted by different financial ratios.

Conclusion

Unlike some of the most popular financial ratios found on some investment websites such as price earnings ratio (PE) and return on equity (ROE), our random forest model output based on the Dow 30 stocks in the past 20 years indicates a variety of financial ratios that have impacted these stocks' prices. The outcome of our analysis is not intended to challenge any existing consensuses. Instead, this proves our hypothesis that each stock has its own unique circumstances that may lead to its stock prices being correlated to a different set of financial ratios. As a smart investor, it is important to conduct the fundamental analysis of each individual stock rather than following an overly broad guideline. The top 5 detailed financial rankings for each of the Dow 30 stocks provide a useful tool to guide investors in evaluating their stock performance down the road. At the same time, some of the common patterns that are observed among financial services and technology sectors could serve as a reference for investment into similar firms in the same industry.

References

Wilkins, G. (2022, May 4). *6 Basic Financial Ratios and What They Reveal*. Investopedia. Retrieved April 1, 2023, from 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).

Ak, B. K., Dechow, P. M., Sun, Y., & Wang, A. Y. (2013). The use of financial ratio models to help investors predict and interpret significant corporate events. *Australian journal of management*, 38(3), 553-598.

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.

Gururaj, V., Shriya, V. R., & Ashwini, K. (2019). Stock Market Prediction using Linear Regression and Support Vector Machines. *Int J Appl Eng Res*, 14(8), 1931-1934.

Appendix – Financial Ratio Reference

Variable Name	Type	Description
PERMNO	double	PERMNO (PERMNO)
GVKEY	string	Global Company Key (GVKEY)
CUSIP	string	CUSIP IDENTIFIER - HISTORICAL (CUSIP)
TICKER	string	EXCHANGE TICKER SYMBOL - HISTORICAL (TICKER)
PEG_1yrforward	double	Forward P/E to 1-year Growth (PEG) ratio (PEG_1yrforward)
CAPEI	double	Shillers Cyclically Adjusted P/E Ratio (CAPEI)
bm	double	Book/Market (bm)
PEG_ltgforward	double	Forward P/E to Long-term Growth (PEG) ratio (PEG_ltgforward)
evm	double	Enterprise Value Multiple (evm)
pe_op_basic	double	Price/Operating Earnings (Basic, Excl. EI) (pe_op_basic)
pe_op_dil	double	Price/Operating Earnings (Diluted, Excl. EI) (pe_op_dil)
pe_exi	double	P/E (Diluted, Excl. EI) (pe_exi)
pe_inc	double	P/E (Diluted, Incl. EI) (pe_inc)
ps	double	Price/Sales (ps)
pcf	double	Price/Cash flow (pcf)
dpr	double	Dividend Payout Ratio (dpr)
ptb	double	Price/Book (ptb)
PEG_trailing	double	Trailing P/E to Growth (PEG) ratio (PEG_trailing)
divyield	double	Dividend Yield (divyield)
efftax	double	Effective Tax Rate (efftax)
GProf	double	Gross Profit/Total Assets (GProf)
aftret_eq	double	After-tax Return on Average Common Equity (aftret_eq)
aftret_equity	double	After-tax Return on Total Stockholders Equity (aftret_equity)
aftret_invcapx	double	After-tax Return on Invested Capital (aftret_invcapx)
gpm	double	Gross Profit Margin (gpm)
npm	double	Net Profit Margin (npm)
opmad	double	Operating Profit Margin After Depreciation (opmad)
opmbd	double	Operating Profit Margin Before Depreciation (opmbd)
pretret_earnat	double	Pre-tax Return on Total Earning Assets (pretret_earnat)
pretret_noa	double	Pre-tax return on Net Operating Assets (pretret_noa)
ptpm	double	Pre-tax Profit Margin (ptpm)
roa	double	Return on Assets (roa)
roce	double	Return on Capital Employed (roce)
roe	double	Return on Equity (roe)
capital_ratio	double	Capitalization Ratio (capital_ratio)
equity_invcap	double	Common Equity/Invested Capital (equity_invcap)
debt_invcap	double	Long-term Debt/Invested Capital (debt_invcap)
totdebt_invcap	double	Total Debt/Invested Capital (totdebt_invcap)
invt_act	double	Inventory/Current Assets (invt_act)
rect_act	double	Receivables/Current Assets (rect_act)
fcf_ocf	double	Free Cash Flow/Operating Cash Flow (fcf_ocf)
ocf_lct	double	Operating CF/Current Liabilities (ocf_lct)
cash_debt	double	Cash Flow/Total Debt (cash_debt)
cash_lt	double	Cash Balance/Total Liabilities (cash_lt)
cfm	double	Cash Flow Margin (cfm)
short_debt	double	Short-Term Debt/Total Debt (short_debt)
profit_lct	double	Profit Before Depreciation/Current Liabilities (profit_lct)
curr_debt	double	Current Liabilities/Total Liabilities (curr_debt)
debt_ebitda	double	Total Debt/EBITDA (debt_ebitda)
dltt_be	double	Long-term Debt/Book Equity (dltt_be)
int_debt	double	Interest/Average Long-term Debt (int_debt)

Variable Name	Type	Description
int_totdebt	double	Interest/Average Total Debt (int_totdebt)
lt_debt	double	Long-term Debt/Total Liabilities (lt_debt)
lt_ppent	double	Total Liabilities/Total Tangible Assets (lt_ppent)
de_ratio	double	Total Debt/Equity (de_ratio)
debt_assets	double	Total Debt/Total Assets (debt_assets)
debt_at	double	Total Debt/Total Assets (debt_at)
debt_capital	double	Total Debt/Capital (debt_capital)
intcov	double	After-tax Interest Coverage (intcov)
intcov_ratio	double	Interest Coverage Ratio (intcov_ratio)
cash_conversion	double	Cash Conversion Cycle (Days) (cash_conversion)
cash_ratio	double	Cash Ratio (cash_ratio)
curr_ratio	double	Current Ratio (curr_ratio)
quick_ratio	double	Quick Ratio (Acid Test) (quick_ratio)
at_turn	double	Asset Turnover (at_turn)
inv_turn	double	Inventory Turnover (inv_turn)
pay_turn	double	Payables Turnover (pay_turn)
rect_turn	double	Receivables Turnover (rect_turn)
sale_equity	double	Sales/Stockholders Equity (sale_equity)
sale_invcap	double	Sales/Invested Capital (sale_invcap)
sale_nwc	double	Sales/Working Capital (sale_nwc)
Accrual	double	Accruals/Average Assets (Accrual)
RD_SALE	double	Research and Development/Sales (RD_SALE)
adv_sale	double	Avertising Expenses/Sales (adv_sale)
staff_sale	double	Labor Expenses/Sales (staff_sale)