Financial Ratios as Predictors in Quantitative Models

Shuaitao Tan

QMSS5999

Introduction:

In recent years, the rise of quantitative finance has transformed the landscape of investing. It has grown crucial for both institutional and retail investors to incorporate data driven strategies into their investment methodologies. As technology advances and financial markets become more complex, traditional methods of fundamental analysis are often supplemented or even replaced by quantitative techniques.

Quantitative investing, which leverages mathematical models, statistical analysis, and large datasets to inform investment decisions, has gained significant prominence. This growing shift reflects a desire for more objective and systematic approaches to understanding market behavior that offer potential advantages in terms of precision and scalability. Understanding the foundations and principles of quantitative finance is therefore critical to situating this paper in the context of broader trends shaping modern investment strategies.

The objective of this thesis is to investigate the predictive power of traditional financial ratios, namely the Price-to-Earnings (P/E) ratio, in forecasting stock returns, with a specific focus on companies within the Standard and Poor's 500 (S&P 500) index. In other words, S&P 500 return will be the benchmark for this study. While traditional ratios like P/E have long served as cornerstones of fundamental analysis, the question arises: *To what extent can traditional financial ratios effectively predict stock returns, and does their predictive power vary across different market sectors?*

Furthermore, this thesis takes one step further down the road to explore whether integrating additional financial features, such as risk metrics, enhances the performance of models relying on these ratios.

Lastly, this study examines how advancements in machine learning techniques might improve the accuracy and efficiency of quantitative models compared to traditional methods. By addressing these questions, this research aims to evaluate the evolving role of financial ratios in modern quantitative finance and assess whether more sophisticated approaches can yield superior predictive outcomes.

This study provides practical insights to both individual and institutional investors. For non-specialists, this study also provides a reason to look into the quantitative finance field, and how quantitative modeling could be an accessible tool for the future.

Literature Review:

Background - Growth of Quantitative Market

The appeal of quantitative finance lies not only in its scalability and precision but also in its capacity to adapt to the complexities of modern financial markets, which are often too volatile and unpredictable for traditional fundamental analysis to adequately address. Citadel and Two Sigma, representatives of pioneering quant firms, rose to all time top 20 firms since 2017 have kept growing at an unprecedented pace. As for the market in general, measured by Russell 3000 index the three types of computer managed funds -index funds, ETFs, and quant funds- run around **35 percent** as of the year of 2019(Jansen, page 8). As these firms have grown in size and influence, they have also contributed to increased market liquidity and efficiency, further solidifying the importance of quantitative techniques. According to Stefan Jansen, there are "three trends have boosted the use of data in algorithm trading strategies and may further shift the investments industry from discretionary to quantitative styles: 1) the exponential increase in the availability of digital data 2) the increase in computing power and data storage capacity at a lower cost 3) The advances in statistical methods for analyzing complex datasets" (Jansen 2020[1], page 9). **The increased accessibility of information, coupled with advancements in data mining techniques, has made quantitative methodologies a more sophisticated and effective approach to navigating the complexities of volatile financial markets.** The growth of quantitative finance represents more than just a technological innovation; it reflects a paradigm shift in the way market participants approach investment decision-making.

---

[1] Jansen, S. (2020). *Machine learning for algorithmic trading: Predictive models to extract signals from the market and alternative data for systematic trading strategies* (2nd ed.). Packt Publishing.

Problems Related to Financial Ratios Efficacy

The rise of quantitative models poses challenges to traditional investing strategies based on fundamental analysis, prompting a reevaluation of established investment practices. It questions the predictive powers of the financial ratios and what possibilities there are to improve efficacy of these ratios.

The Price-to-Earnings (P/E) ratio is a widely used financial metric that compares a company's current share price to its earnings per share (EPS). It helps investors to evaluate companies on their profitability and potential scalability. The P/E ratio indicates how much investors are willing to pay per dollar of earnings. A high P/E ratio typically suggests that investors expect higher earnings growth in the future, while a low P/E ratio may indicate undervaluation or pessimistic market sentiment towards the company's future prospects.

Research has extensively explored the relationship between the P/E ratio and stock performance. Studies, such as those by Basu (1977)[2], have shown that stocks with lower P/E ratios often outperform their high P/E counterparts over long periods, a phenomenon known as the "low P/E effect." However, the effectiveness of the P/E ratio as a predictor of future stock returns is debated, as other factors—such as market conditions and risk metrics—can influence its accuracy (Gottwald, 2011[3]).

Variations of the P/E ratio, such as trailing P/E (based on historical earnings) and forward P/E (based on projected earnings), provide investors with different perspectives for assessing stock value, making the ratio a versatile and widely used tool in fundamental analysis. However, the P/E ratio's limitations become evident when used in isolation, as it fails to account for critical

---

[2] Basu, S. (1977). Investment performance of common stocks in relation to their price‐earnings ratios: A test of the efficient market hypothesis. The Journal of Finance, 32(3), 663-682. https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

[3] Radim Gottwald, "The Use of the P/E Ratio to Stock Valuation," *GRANT Journal* 0102 (2012): 21-23, https://www.grantjournal.com/issue/0102/PDF/0102gottwald.pdf.

factors such as growth potential, risk, or sector-specific characteristics. For instance, companies in growth-oriented sectors, such as technology, often exhibit higher P/E ratios compared to those in mature, capital-intensive industries like utilities or manufacturing. This raises an important question: Does the predictive power of the P/E ratio—and similar traditional financial metrics—vary across different sectors of the market? Understanding these sectoral differences is essential for evaluating the true effectiveness of financial ratios in modern investing, as their reliability may depend heavily on the market context in which they are applied.

Machine Learning Techniques and Model Building

Quantitative finance models are developed through feature engineering and model training, with different firms tailoring trading strategies and stock selection models based on factors such as size and risk preference, adapting these models to various market scenarios. In “Research on Quantitative Investment Strategies Based on Deep Learning”, Yujie Fang investigates the effectiveness of deep learning models in quantitative trading with a specific focus on ETF options. Fang’s study, centered on the 50 ETF options in the highly complex options market, highlights the advantages of deep learning over traditional quantitative investment strategies, stating that “the experimental results show that the quantitative investment strategy based on deep learning has higher returns than the traditional quantitative investment strategy, the yield curve is more stable, and the anti-fall performance is better” (Fang). Moreover, the study emphasizes the broad applicability of quantitative methodologies across diverse financial contexts. While Fang’s research focuses on the Random Forest, Long Short-Term Memory Network model, other scholars like Muhhamad Ali had taken different routes in studying the effectiveness of artificial neural networks and support vector machines (other machine learning methods) in predicting the closing pricing movement of the four stock exchanges. In their study[4], Ali et al. (2021) examine the predictive capabilities of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) in forecasting the directional movement of stock indexes, including the KSE-100, KOSPI, Nikkei 225, and SZSE composite. Using ten years of daily closing data (2011–2020) from Yahoo Finance and fifteen technical indicators as input features, the authors found that ANN consistently achieved higher accuracy and F-scores than SVM models across all indexes. Despite limitations such as excluding macroeconomic factors like exchange rates and interest rates, the study underscores the utility of ANN and SVM in minimizing investment risk by predicting market trends effectively. The authors suggest future exploration of ensemble methods, particularly LSTM networks with diverse input features and timeframes, to enhance prediction accuracy across various economic conditions.

These findings underscore the potential of machine learning techniques to revolutionize portfolio strategies, enabling models that can outperform traditional benchmarks such as the S&P 500. By leveraging advanced methods like deep learning and neural networks, researchers and practitioners have demonstrated improvements in return stability, predictive accuracy, and downside risk management. However, this raises two critical questions: *Does the increasing complexity of machine learning models inherently lead to better performance, or is there a point of diminishing returns?* Moreover, as these models excel in producing superior outcomes, *to what extent can such strategies consistently outperform benchmarks like the S&P 500 in diverse market conditions?* Exploring these questions is essential for understanding the trade-offs between model sophistication and practical applicability in quantitative finance.

[4] Ali, M., Khan, D. M., Aamir, M., Ali, A., & Ahmad, Z. (2021). Predicting the direction movement of financial time series using artificial neural network and support vector machine. *Complexity*, *2021*, 2906463. https://doi.org/10.1155/2021/2906463

Data and Methodology:

Market Growth

This thesis begins by examining the growth of the quantitative market through descriptive statistical analysis. As noted by the Corporate Finance Institute, "The total value of AUM is a measure of the size of a financial institution and a key performance indicator of success, as a larger AUM generally translates into larger revenue in the form of management fees[5]" (CFI, Assets Under Management). Consequently, Assets Under Management (AUM) is selected as the primary quantitative variable to demonstrate the increasing prominence of quantitative finance in the financial industry. Since the market is growing, a positive trend is expected. Renowned firms such as Citadel, Two Sigma, and Renaissance Technologies, widely regarded as leaders in quantitative finance, are chosen as representative examples for this analysis.

The data used in this study is sourced from AUM 13F[6], a platform that provides comprehensive information on AUM and other financial metrics of firms globally, ensuring the reliability and relevance of the data used for this research. Since the market is growing, a positive trend is expected. This marks not only the growth of the firm, but also hints the expanding need of investors.

---

[5] "Assets Under Management (AUM): A Measure of Success." *Corporate Finance Institute*.
[6] AUM 13F. (n.d.). *Firm holdings and assets under management*. Retrieved [insert date of access], from https://aum13f.com/

Hypothesis Testing

Data Overview:

In this study, the stocks related information are retrieved using the yahoo finance package through python. The stock selection is based on the S&P 500 index, providing a comprehensive list of large-cap U.S. equities. The range of stock data are different for the individual hypothesis because the goal of different hypothesis varies: in building the quantitative model for this thesis, 8 years of stock data are gathered and then cleaned to ensure the model is trained on enough data, but only 1 year of stock data are used for testing correlation between ratios and examining performance across different sectors.

Hypothesis 1: The inclusion of a greater number of financial features, such as financial ratios, enhances the predictive accuracy of quantitative models in financial analysis.

Independent Variables: The financial features included in the model, such as Price-to-Earnings (P/E), Price-to-Book (P/B),volatility, and other financial ratios.

Dependent Variable: The predictive accuracy of the quantitative model, measured using metrics such as R-squared and Mean Absolute Error (MAE).

Testing Approach:

<u>Autocorrelation analysis</u> is a crucial first step in testing this hypothesis, as it identifies the degree of redundancy among financial features. Autocorrelation is problematic in referring to a lack of independence between variables. In other words, while there are many financial features available, if these ratios exhibit strong autocorrelation, they will provide overlapping information. This indicates that the model may be negatively impacted by adding redundant variables. It could potentially create a bias due to repetitive input into the model. By calculating

autocorrelation coefficients and evaluating their statistical significance, this step flags variables with high redundancy as less informative for improving predictive performance. By having the understanding between how the financial features relate to each other, it establishes a basis for the following step.

The second step involves systematically adding variables with relatively low correlations to the model to evaluate their contribution to predictive accuracy. From a broader perspective, since these variables present low correlations with each other, this thesis hypothesized that these independent variables are meaningful predictors. These financial variables are able to capture diverse aspects of financial performance. When each one of these variables is added to the model, the model shall present an increasing accuracy.

By sequentially integrating these features into the model and measuring changes in accuracy using metrics R-square, a trend line is able to reflect how the model's predictive power changes over time. On the Y axis, R squared represents the accuracy of the model, and on the X axis, it represents the number of financial features added to the model.

Together, these steps comprehensively test the hypothesis by differentiating the impact of redundant versus complementary variables, providing a robust evaluation of the relationship between financial features and model accuracy.

Hypothesis 2: The predictive power of traditional financial ratios, such as the Price-to-Earnings (P/E) and Price-to-Book (P/B) ratios, varies in accuracy when applied to different sectors of the market.

To test Hypothesis 2, from December 2023 to December 2024, a total of 370 days of s&P 500 data is prepared where an extra week's data is used as a buffer. By converting returns to percentages and removing outliers, the new dataframe contains necessary information to study P/E, P/B, and one year return.

Global Industry Classification System (GICS) is a system for categorizing the companies across the world. The data is then segmented according to GICS sectors to enable sector-specific analysis.

Then regression analysis is conducted based on the newly generated dataframe. In order to examine the predictive power of P/B and P/E, the testing process is divided into two stages. Where the first test is an overall market regression analysis to examine how the ratios perform in simple context. Next, the study is conducted sector by sector. However, in order to have enough data to study, a minimum of 10 companies is required to proceed. For each sector, besides the regression coefficients, p values, and R squared are also calculated for valuation of the result. In order to ensure robust results, the potential usage of adding constant terms is reserved.

A bar chart is created at the end of this study for a clear visualization of the sector difference in the predictive power of the ratios. It allows easy access for analysis and statistical summary.

Hypothesis 3: The application of machine learning techniques significantly enhances the performance and predictive accuracy of quantitative financial models compared to traditional methods.

This study employs an unsupervised machine learning approach to develop and test a quantitative trading strategy using S&P 500 stock data spanning from 2016 to 2024. The methodology framework integrates technical analysis with modern portfolio theory through a three-stage process: feature engineering, cluster analysis, and strategy development.

In the first phase of building this quantitative model, several financial features are not available directly through the yahoo finance package. “RSI is a momentum oscillator that is widely used in technical analysis of stocks and commodities to identify changes in momentum and price direction”[7]. In other words, RSI as a financial feature provides the model insight into the sentiment of the individual stock. At the same time, a high RSI over 70, is generally considered as overbought and a RSI lower than 30, often treated as oversold. RSI is not the only significant technical indicator, Garman Klass Volatility and Bollinger Bands also plays a role in accessing the volume of the stocks. Garman Klass volatility emphasizes the measurement of daily volume, whereas Bollinger Bands looks into how the characteristic of stock price and volume changes over time. Additionally, Moving Average Convergence Divergence (MACD) adds another dimension to the financial features pool: the typical MACD range of -1 to 1 refers to that approximately 68% of data in a normal distribution fall within one standard deviation of the mean. This range can be considered as the normal fluctuation around the mean. Besides the

[7] Investopedia. (n.d.). *Relative strength index (RSI).* Retrieved December 25, 2024, from https://www.investopedia.com/terms/r/rsi.asp

MACD, the dollar volume and Average True Range all provide insight into how the stock performs relative to the market.

Into the next phase of the model building, due to the large amount of data present, monthly data and monthly return is then calculated in order to perform the task with a higher efficiency. The Fama French Factor[8], a well known asset pricing model, is introduced to the monthly return data frame whereas valuable information like market risk premium, size factor, and value factor etc are now all taken into account. This helps to take the next step to use the K means algorithm to conduct a cluster analysis on the stocks. This analytical framework helps to determine the upward trend with the help of RSI criterion (>70).

From this cluster analysis, the final model uses an unsupervised machine learning model to create a trading strategy which is then used to compare to the S&P 500 Benchmark return. The graph will use cumulative return as the y axis because it better shows how the difference between the model and the benchmark differentiate over time.

[8] French, K. (n.d.). *Fama/French 5 factors (2x3).* Dartmouth College. Retrieved December 25, 2024, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library/f-f_5_factors_2x3.html

Results:

Growth of Importance

Since 2000, firms like Two Sigma, Citadel, and Renaissance Technologies have demonstrated exceptional growth in the financial markets, solidifying their positions as leaders in quantitative finance. The sustained increase in assets under management (AUM) among these firms reveals a strong and consistent upward trend, reflecting both the stability of their business models and the high level of confidence they command from clients and investors. This trajectory underscores the growing indispensability of quantitative finance within the broader financial services industry. The success of these firms highlights not only their ability to leverage advanced quantitative models and technologies but also the shifting paradigm in investment management, where data-driven strategies are increasingly favored over traditional methods. This finding reinforces the critical role quantitative finance plays in shaping the future of the industry and signals its continued expansion as a key pillar of financial innovation.

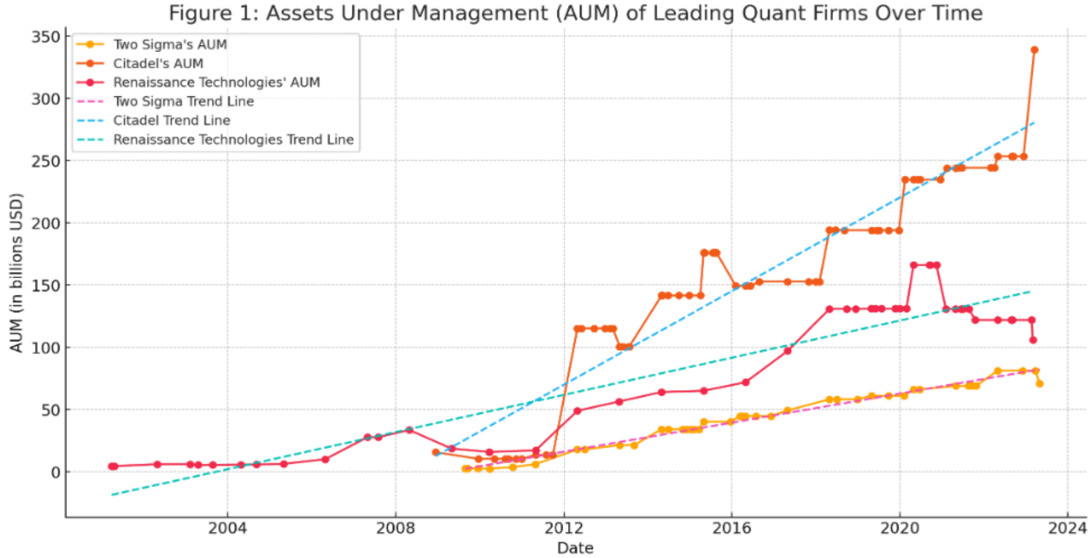


Figure 1: Assets Under Management (AUM) of Leading Quant Firms Over Time

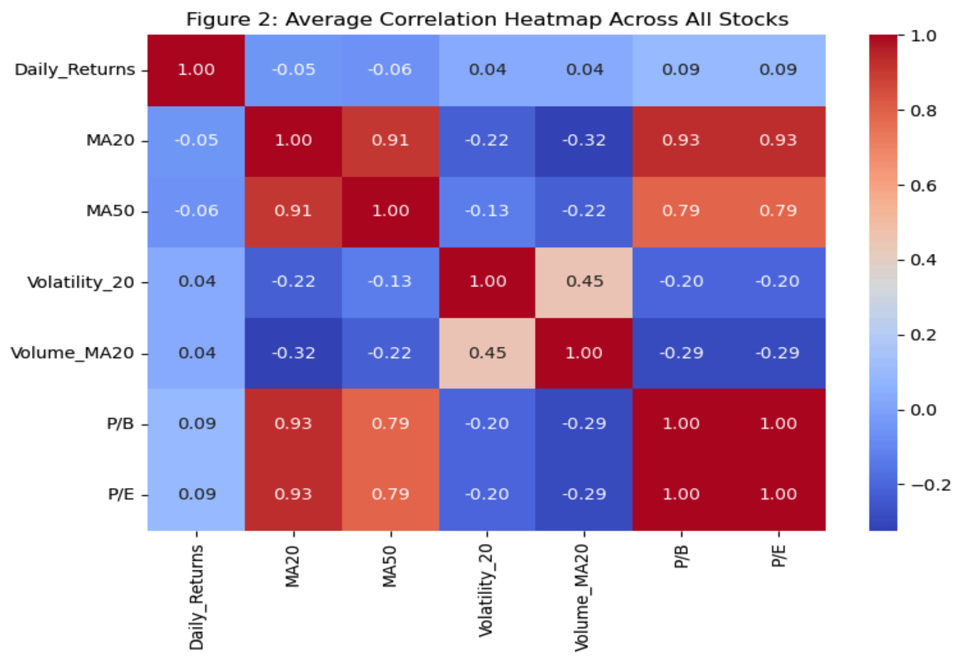
Hypothesis 1: The inclusion of a greater number of financial features, such as financial ratios, enhances the predictive accuracy of quantitative models in financial analysis.

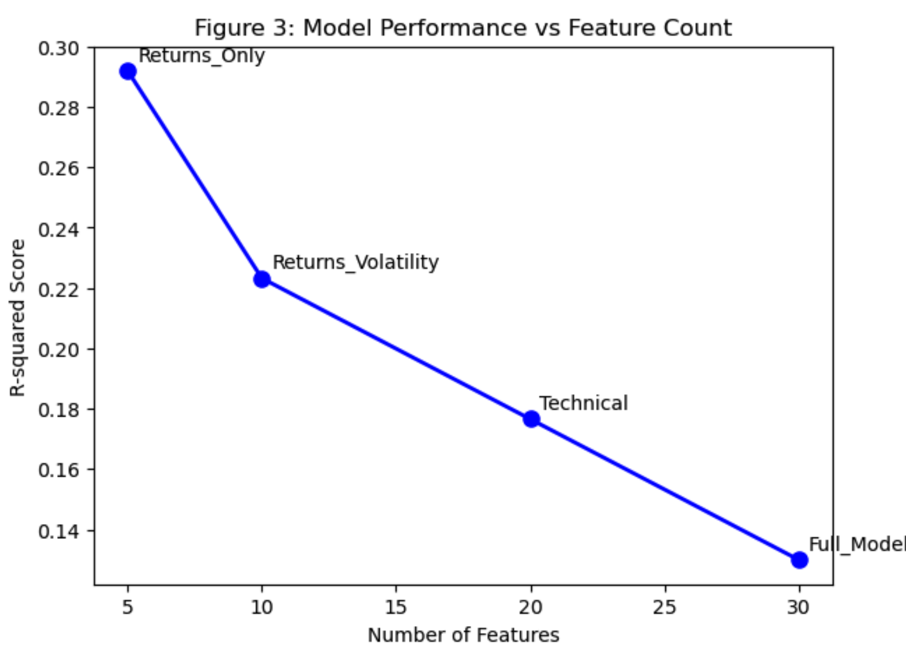
The testing of hypothesis 1 begins with an autocorrelation study on how financial features relate to each other. In this specific context, five stocks are randomly picked from sp500, and then daily returns, P/B, P/E, and volume these ratios are used to conduct the autocorrelation analysis. The correlation is valued across the stocks chosen. From the heatmap, Based on the correlation heatmap analysis, several significant relationships emerge among the studied variables. The most notable finding is the strong positive correlation ($r = 0.93$) between MA20 and both P/B and P/E ratios, suggesting a robust relationship between short-term moving averages and valuation metrics. Similarly, MA50 exhibits a moderately strong positive correlation ($r = 0.79$) with these valuation measures, albeit slightly weaker than MA20. Interestingly, daily returns show minimal correlation with other variables, with correlation coefficients ranging from -0.06 to 0.09, indicating relative independence in daily price movements. The analysis also reveals a moderate positive correlation ($r = 0.45$) between volatility and volume moving averages, suggesting some relationship between price variability and trading activity. Furthermore, the perfect correlation ($r = 1.00$) between P/B and P/E ratios indicates these valuation metrics move in complete synchronization. The negative correlations between volume moving averages and both MA20 (-0.32) and MA50 (-0.22) suggest an inverse relationship between trading volume and price trends. These findings provide valuable insights into the interrelationships among various market indicators and valuation metrics, which could be particularly relevant for investment decision-making and market analysis.

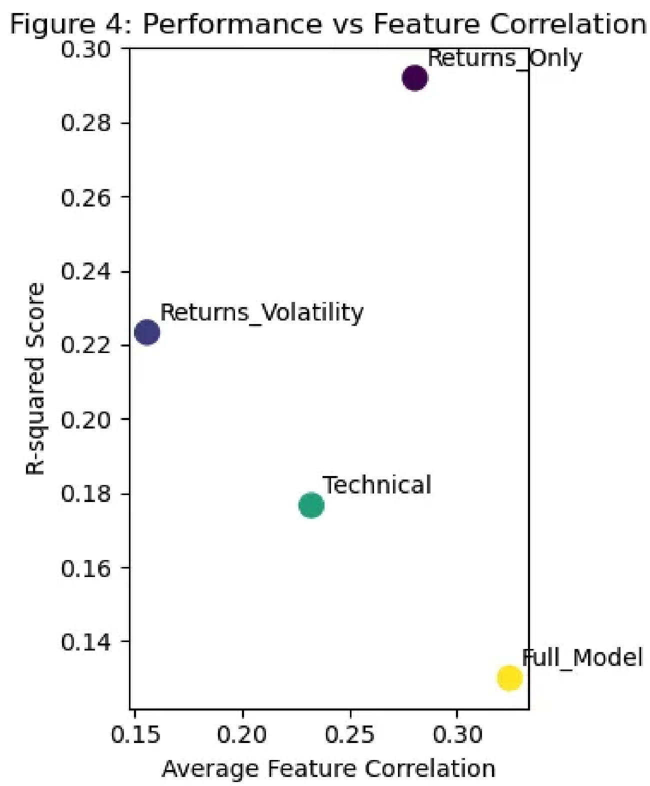
Since some of the variables are highly correlated to each other, it indicates that inclusion of more variables does not necessarily improve the model. This leads to the next step of hypothesis 1 testing: testing model improvement based on the number of features. In this study, a linear machine learning model is used, where the independent variables are the financial features, including returns, technical indicators, volatility, and other metrics. The dependent variable is model performance, which is assessed by adding features step by step into the model. The full model is defined as the model that includes the maximum number of financial features.

From figure 3, we observe a declining trend in model performance as more variables are added, leading to a rejection of the hypothesis. The model's performance appears to deteriorate with the inclusion of additional features, suggesting that a simpler model with fewer carefully selected variables might be more effective. This finding aligns with the principle of parsimony in statistical modeling and indicates that multicollinearity among financial features may be introducing noise rather than enhancing predictive power.

Building upon the previous findings, Figure 4 provides a visual representation of the relationship between model performance (R-squared score) and average feature correlation. The analysis demonstrates that the Returns_Only model, utilizing just return-based features, achieves the highest R-squared score of approximately 0.29, despite having a relatively high average feature correlation of 0.30. In contrast, the Full_Model, which incorporates all available financial features, shows the lowest performance with an R-squared score of about 0.13, accompanied by the highest feature correlation. The Returns_Volatility combination yields an intermediate performance with an R-squared score of 0.22, while the Technical indicators model shows slightly lower performance at 0.18. This inverse relationship between model complexity and performance further supports the rejection of hypothesis 1, suggesting that simpler models focused on specific feature sets may be more effective for financial prediction than more complex models incorporating numerous potentially redundant variables.







Hypothesis 2: The predictive power of traditional financial ratios, such as the Price-to-Earnings (P/E) and Price-to-Book (P/B) ratios, varies in accuracy when applied to different sectors of the market.

The regression analysis reveals significant variations in the predictive power of traditional financial ratios across different market sectors, providing strong support for Hypothesis 2. The Energy and Utilities sectors demonstrate the strongest predictive relationship, with notably high R-squared values of 0.7254 and 0.5374 respectively, indicating that P/E and P/B ratios explain a substantial portion of return variation in these sectors.

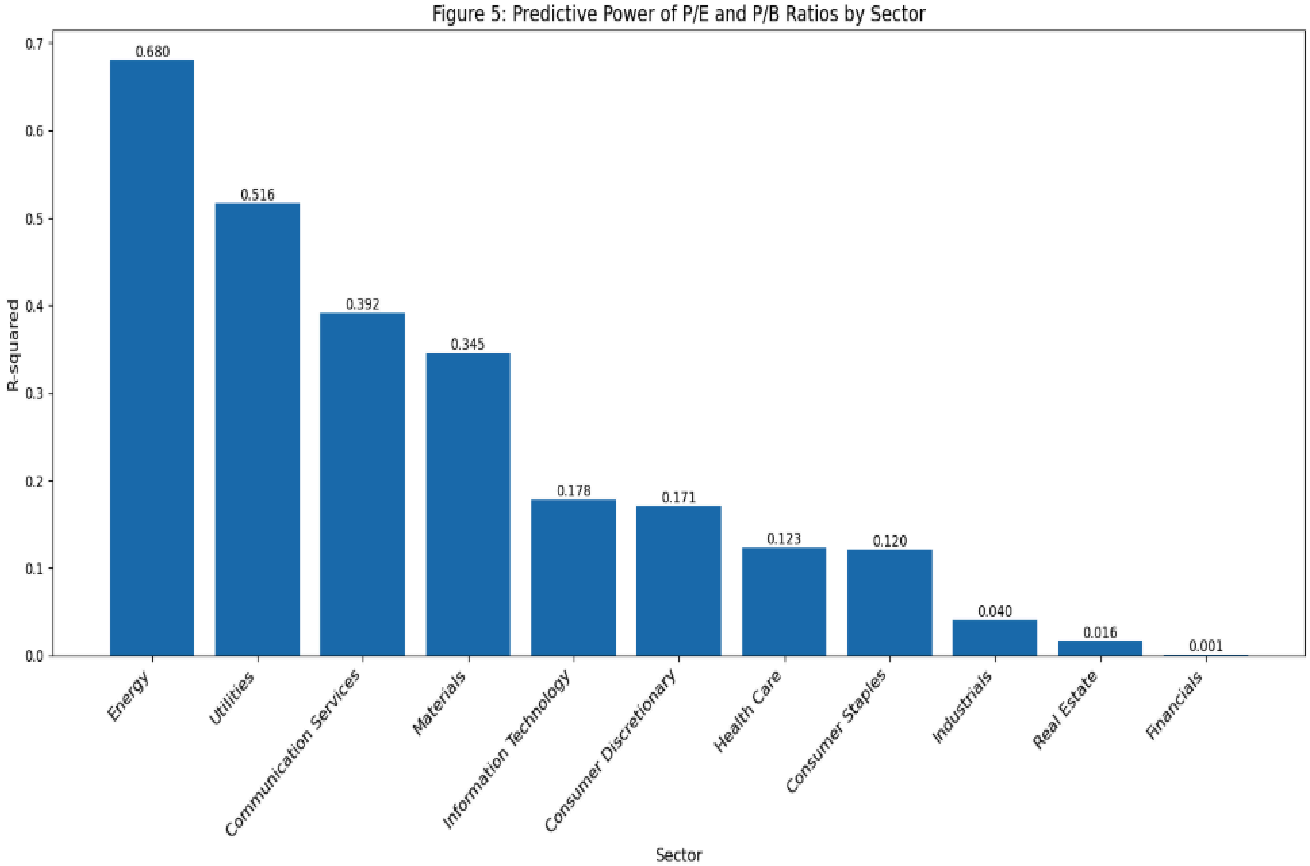
In contrast, traditional financial ratios show remarkably weak predictive power in the Financials and Real Estate sectors, with R-squared values of just 0.0022 and 0.0168, suggesting these metrics may not be suitable for valuation in these particular sectors. One potential reason is that macroeconomic factors contribute more to the growth of Financial and Real Estate industries, while the P/B and P/E are not the best financial indicators in capturing market sentiment. At the same time, regulatory policy has a strong influence over the overall performance. In other words, companies with large physical assets may not be fully predicted in terms of model using financial ratios, ratios related to operating conditions of the companies could be the alternative solution.

Materials and Communication Services sectors exhibit moderate predictive relationships, with R-squared values of 0.3678 and 0.3551, while the remaining sectors show relatively weak correlations. The analysis also reveals varying significance of individual ratios across sectors, with the P/B ratio showing high significance in Energy ($p=0.0017$) and Utilities ($p=0.0002$) sectors, while the P/E ratio demonstrates stronger significance in sectors such as Materials ($p=0.0055$). The varying dynamics of the underlying industries highlight that the significance of financial ratios is volatile. For example, for the energy industry, companies tend to have large tangible assets, making P/B ratio more relevant. However, for the materials industry, P/E is more dominant in its application on the economic conditions of the company.

These findings highlight the importance of considering sector-specific characteristics when applying traditional valuation metrics, rather than using a one-size-fits-all approach across the market. This sector-dependent variation in predictive power suggests that investors and analysts should adjust their valuation approaches based on the specific sector they are analyzing.

Sector	R-squared	P/E Coefficient (p-value)	P/B Coefficient (p-value)	Sample Size
Energy	0.7254	0.9208 (0.0560)	4.3451 (0.0017)**	21
Utilities	0.5374	1.7730 (0.0476)*	8.7682 (0.0002)**	28
Materials	0.3678	0.6823 (0.0055)**	1.4479 (0.1410)	25
Communication Services	0.3551	0.1253 (0.8174)	4.5898 (0.0306)*	17
Information Technology	0.1776	0.2074 (0.0111)*	0.5019 (0.0523)	55
Consumer Discretionary	0.1703	0.7035 (0.0194)*	0.0008 (0.9987)	37
Health Care	0.1642	0.1594 (0.1480)	0.5960 (0.0216)*	49
Consumer Staples	0.1136	-0.0739 (0.7280)	1.3703 (0.0705)	31
Industrials	0.0265	0.0800 (0.4910)	0.6704 (0.2790)	66
Real Estate	0.0168	0.0614 (0.5325)	-0.0344 (0.9589)	27
Financials	0.0022	0.0699 (0.7166)	-0.0566 (0.8577)	65

Figure 5: Predictive Power of P/E and P/B Ratios by Sector



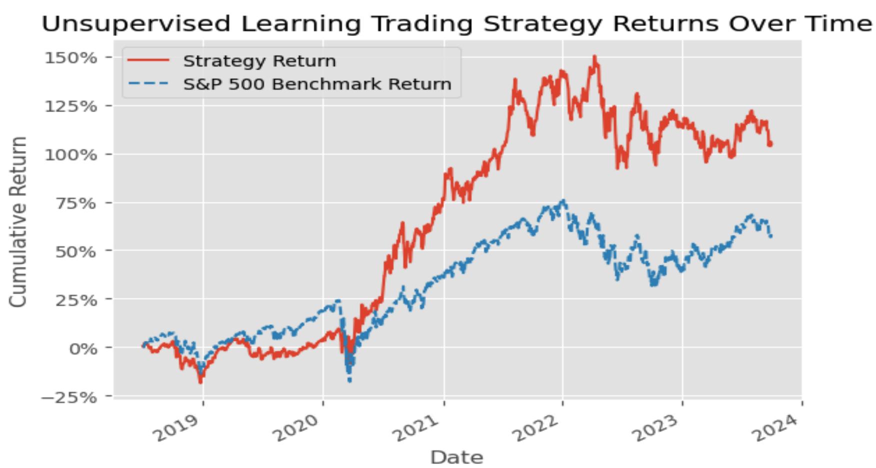
Hypothesis 3: The application of machine learning techniques significantly enhances the performance and predictive accuracy of quantitative financial models compared to traditional methods.

The empirical results demonstrate that the unsupervised learning trading strategy significantly outperformed the S&P 500 benchmark over the period from 2019 to early 2024. The strategy, which employs K-means clustering analysis of financial ratios, technical indicators and Fama-French factors, generated a cumulative return of approximately 110% compared to the benchmark's return of around 65%, representing an outperformance of 45 percentage points over the five-year period.

The strategy's performance trajectory exhibits several notable characteristics. First, during the market downturn in early 2020 associated with the COVID-19 pandemic, the strategy demonstrated a comparable drawdown to the benchmark, suggesting similar initial vulnerability to systematic market shocks. However, the recovery phase revealed superior performance characteristics, with the strategy achieving a steeper recovery trajectory from mid-2020 onwards. Particularly striking was the period between 2021 and 2022, where the strategy's cumulative returns diverged significantly from the benchmark, reaching peak outperformance with returns approaching 150%.

Of particular interest is the strategy's behavior during various market regimes. The clustering methodology, utilizing an RSI threshold above 70 to identify upward trends, appeared most effective during the post-2020 recovery period. The integration of Fama-French factors with technical indicators seemingly enhanced the model's ability to capture both momentum effects and fundamental market premiums. However, it's worth noting that the strategy's outperformance showed some moderation in 2023-2024, while still maintaining a substantial cumulative performance advantage over the benchmark.

These findings support the initial hypothesis that machine learning techniques can enhance the performance of quantitative financial models. The strategy's consistent outperformance, particularly during periods of market recovery and momentum, suggests that the unsupervised learning approach effectively captured profitable market patterns that may not be readily apparent through traditional analysis methods.



Conclusions:

From the result, this thesis identifies a strong growing trend of quantitative firms size which are proven by the growth of leading quant firms like Citadel. It also points out that there is an expanding need from retail investors for access to quantitative investment. By testing the three hypotheses the thesis provided, it can be concluded that adding features do not always improve the performance of the quantitative models, on the contrary, adding repetitive features that contain high correlation with each other leads to deterioration effect on the predictive power of the model. It is essential to determine the financial features that are less correlated with each other but meaningful enough to capture diverse aspects of the market.

What's more, while P/B and P/E remain powerful ratios to be used in making financial judgement. Their predictive power varies vastly across industries. Therefore, when it comes to making investment decisions, sector analysis can be quite helpful in stock selection. It also builds an important signal for model building in the future.

At last, the unsupervised machine learning trading model this thesis has provided a good measurement of how quantitative modeling could elevate the investment performance of investors. The cumulative return greatly exceeded the benchmark return. However, while the model itself has limitations to be discussed in the later chapter, it provides valuable insight on how machine learning techniques could be an indispensable tool for future investors in terms of capturing patterns, signaling trends, and making predictions.

Limitations:

While this thesis presents the quantitative model with a better performance in the cumulative return of the trading strategy, it does not account for any of the surprise events in the past, especially market shocks. For example, there would have been no possibility to foresee any geopolitical risks involved in investing. Natural disaster or war introduces significant volatility that models struggle to predict. This limitation means that the model may overestimate potential returns by not considering the impact of sudden, unpredictable events that can lead to significant market volatility. As a result, investors relying solely on this model might face unexpected losses during periods of market disruption. Therefore, incorporating mechanisms to account for such shocks could enhance the model's robustness and reliability.

Another limitation of the thesis is the hypothesized distinctive boundary between human discretionary approach and quantitative modeling approach in trading strategies. While in this thesis, quantitative modeling seems to be the better option, in the real world, investors take a mixed approach in their investment strategies. A mixed approach allows investors to combine the intuitive insights and experience of human decision-making with the precision and data-driven analysis of quantitative models. A mixed approach potentially offers more adaptivity and balance as investors are able to respond to market anomalies more effectively.

The last limitation related to the model is it does not consider the transaction cost and the frequency of its trade. At the same time, the stock trade does not reflect any over the counter (OTC) opportunity. The cumulative return may be significantly lower than what the model has now.

Reference

Ali, M., Khan, D. M., Aamir, M., Ali, A., & Ahmad, Z. (2021). Predicting the direction movement of financial time series using artificial neural network and support vector machine. *Complexity, 2021,* Article 2906463. https://doi.org/10.1155/2021/2906463

AUM 13F. (n.d.). Firm holdings and assets under management. Retrieved December 10, 2024, from https://aum13f.com

Basu, S. (1977). Investment performance of common stocks in relation to their price‐earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance, 32*(3), 663–682. https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

Corporate Finance Institute. (n.d.). Assets under management (AUM): A measure of success. *Corporate Finance Institute.* Retrieved December 10, 2024, from https://corporatefinanceinstitute.com

French, K. (n.d.). *Fama/French 5 factors (2x3).* Dartmouth College. Retrieved December 25, 2024, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library/f-f_5_factors_2x3.html

Gottwald, R. (2012). The use of the P/E ratio to stock valuation. *GRANT Journal, 0102,* 21–23. https://www.grantjournal.com/issue/0102/PDF/0102gottwald.pdf

Jansen, S. (2020). *Machine learning for algorithmic trading: Predictive models to extract signals from the market and alternative data for systematic trading strategies* (2nd ed.). Packt Publishing.

Investopedia. (n.d.). *Relative strength index (RSI).* Retrieved December 25, 2024, from https://www.investopedia.com/terms/r/rsi.asp