B.Comp. Dissertation

Equity Portfolio Construction using Quarterly Institutional Holdings Data

By

Oh Chunguan Darius

Department of Information Systems and Analytics

School of Computing

National University of Singapore

2022/2023

B.Comp. Dissertation

Equity Portfolio Construction using Quarterly Institutional Holdings Data

By

Oh Chunguan Darius

Department of Information Systems and Analytics

School of Computing

National University of Singapore

2022/2023

Project No: H196350

Advisor: Assoc Prof Tan Wee Kek

Deliverables:

Report: 1 Volume

Abstract

Institutional holdings data released every quarter is publicly accessible and provides insights into what large funds are investing in. This paper aims to investigate the effectiveness of using this data to develop a systematic investment strategy via machine learning methods.

Our research found that following the holdings of smaller funds with historically low volatility and a good track record produced the best returns. Using the features generated from these holdings, we developed a XGBoost model to predict future stock returns, achieving a Pearson Information Coefficient of 10.5% and a Spearman Information Coefficient of 6.5% on withheld data. Our portfolio constructed from the model forecasts achieved a return-risk ratio of 1.00 when evaluated on withheld data, far outperforming the S&P 500 Index with a return-risk ratio of 0.67.

To summarize, we have developed an alpha-seeking trading strategy using easily accessible and cost-free data sources and it significantly outperforms a buy-and-hold strategy. However, we do recognize the limitations of cost-free data sources, and cost-benefit analysis can be applied in future research to determine whether it is beneficial to utilize other data sources. Future extensions to the research can also include exploring other objective functions for the stock prediction model, as well as a more complex covariance forecast model for portfolio optimization.

Subject Descriptors:

G.1.6 Optimization

I.2.6 Learning

I.7.0 General Document and Text Processing

Keywords:

Data Science & Business Analytics, Machine Learning, Financial Analytics

Acknowledgement

Firstly, I would like to express my gratitude to my mentor, Dr. Tan Wee Kek, for his invaluable guidance and support throughout the entire process of writing this research paper. His expertise and insights have been vital in shaping the direction of this project and improving its overall quality.

I would also like to extend my heartfelt thanks to my friends who took their time to share their feedback and suggestions regarding my research. It has helped to refine my ideas and expand my understanding of the subject matter.

Without their unwavering support, this research paper would not have been possible. I am deeply grateful for their contributions to this research.

Table of Contents

1 Introduction

In recent years, the use of institutional holdings data has become increasingly popular in developing trading strategies due to its increasing availability and quality. Large institutional investment managers, such as hedge funds and mutual funds, managing more than US$100 million worth of assets are required to disclose their equity holdings each quarter (Hayes, 2022). The report, known as the Securities and Exchange Commission's (SEC) Form 13F, is publicly available and relatively clean. The filing provides a comprehensive snapshot of their portfolio holdings, including their positions in individual stocks, bonds, and other securities. These data provide insights into what the top funds in the world are investing in, and how their investments change over time.

The accessibility of this information has made it possible for retail investors to also utilize this data in their own trading strategies. This can be observed by the rise of online platforms such as WhaleWisdom.com which provides quick tools to track and analyze fund filings in real time. This paper aims to investigate the effectiveness of using the SEC 13F institutional holding data to develop a systematic investment strategy via Machine Learning. The strategy would output a quarterly-rebalancing portfolio of stocks, which we will analyze the performance of the strategy as compared to past research as well as major market indices.

1.1 Research Problem and Motivation

Top asset management funds invest vast amounts of resources in order to develop alpha-generating strategies for their investors. However, retail investors do not have the same access to the same level of resources. Various key financial data sources, such as Bloomberg and Refinitiv, costs tens of thousands of dollars annually. On top of that, asset management funds also spend huge amounts recruiting the best talents globally, to devise their trading strategies. To level the playing field, retail investors are able to attempt to use 13F filings to pursue 'copycat' investing strategies, by replicating the moves made by large funds. This process is commonly known as "Alpha Cloning" (DiPietro, 2019). Retail investors can potentially benefit from the expertise and resources of these large funds without incurring the same level of costs.

To overcome the challenges required to manually analyze the holdings data of every single institution, machine-learning models can be used to extract the meaningful features in the data for stock-picking, and the output of the models can then be used to construct a stock portfolio which systematically rebalances itself every new quarter. On top of that, machine-learning models can also identify more advanced trading strategies rather than a simple "copycat" strategy. For example, we could potentially be able to identify overcrowding situations, where many large investors follow the same strategy or purchase the same stocks. While this may represent consensus in a stock performing well, it also amplifies market swings and could potentially exhibit a mean-reversion behaviour. Hence, it may be more beneficial for the portfolio to avoid stocks that are too crowded (Miori and Cucuringu, 2022).

This paper aims to study data that is readily accessible for retail investors, and whether such data can be used to capture some of the alpha that is generated via larger institutional funds. Our main goal is to level the playing field for small retail investors which do not have access to much financial resources, and do not meet the liquidity requirements for putting their money in larger institutional funds. While this method has its limitations, it does provide a viable option for retail investors looking to gain a competitive advantage in their investment strategies.

1.2 Research Objectives

First and foremost, the paper aims to propose a set of rules to identify funds where the 13F holdings data filed would accurately reflect the actual holdings of the fund. This is to address a key limitation of the 13F data, where asset management funds are only required to file the 13F form for their quarterly holdings up to 45 days after the last day of the calendar quarter (Hayes, 2022). This implies that the actual holdings of the fund may differ substantially from the holdings reported in the 13F form, particularly for funds that adopt a mid-to-high frequency trading strategy, due to the trades made during the time lag. Our initial hypothesis is that by classifying funds according to their characteristics and trading strategies, we can narrow down our data to focus on funds that adopt a low turnover, long-horizon trading strategy. These funds invest in stocks that they expect to have a good performance over a long period of time, and their holdings should not change very significantly in a short period of time as compared to funds which trade higher frequency. By filtering out funds that do not meet the criteria set, we can improve the accuracy of our analysis of the filings data and provide a more reliable basis for constructing our investment portfolios.

Secondly, the paper aims to identify useful features that can be extracted from the quarterly institutional holdings data to be used as leading indicators for individual stock performance. According to Angelini, Iqbal and Jivrag (2019), these features can be categorized into two broad categories: Conviction and Consensus. Conviction refers to the strength of belief of how the stock will perform, and this is depicted by funds concentrating a large percentage of their portfolio on a small handful of stocks. Consensus refers to the number of funds that have a similar belief of the stock's performance, and this is depicted by many funds holding the same stocks. Our initial hypothesis is that a combination of both Conviction and Consensus in a stock should be a good predictive indicator for future positive returns. However, as discovered in the study by Miori et al. (2022), too high a level of Consensus might be indicative of overcrowding behaviour, where we expect a reversion of price instead.

Lastly, with the future expected returns of individual stocks, the paper aims to identify the most optimal asset allocation methodology for portfolio construction. We will explore methods such as Markowitz Mean-Variance Portfolio Optimisation (Markowitz, 1952) and look at various portfolio performance metrics such as the portfolio's return-risk ratio and tail loss risk. We will also evaluate our portfolio against the benchmark S&P 500 Index, where we should expect that our portfolio should outperform the S&P500 Index which simply weighs allocation to the top 500 U.S. stocks via their market capitalisation.

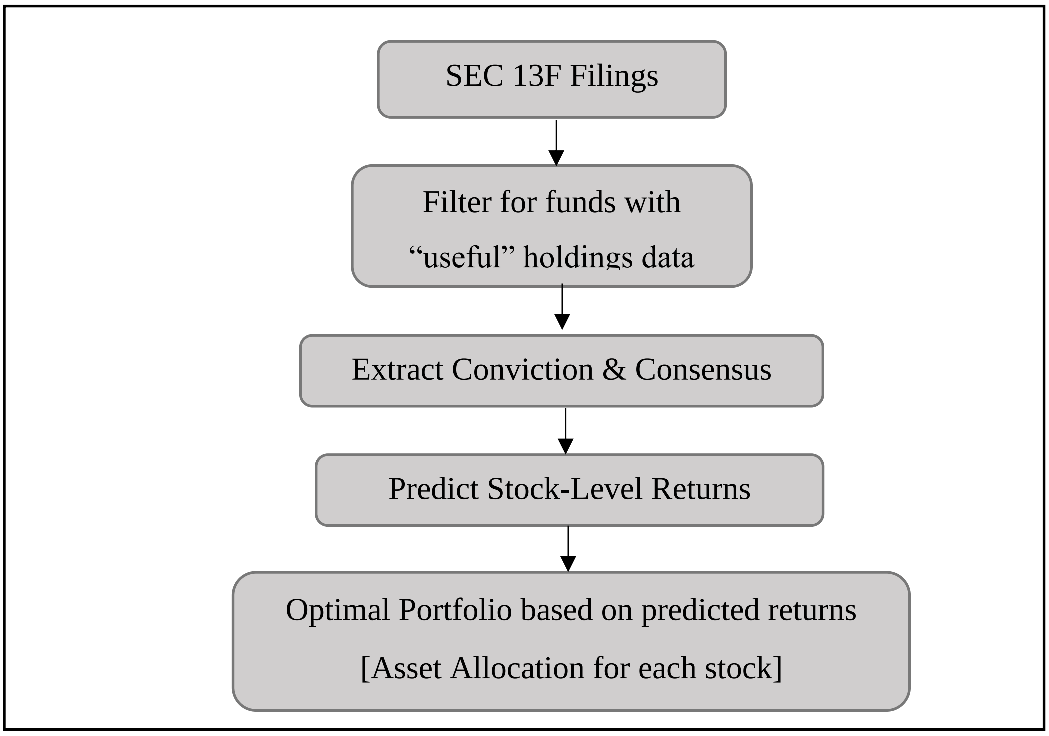


Figure 1: Flowchart of research deliverables

2 Literature Review

There have been several research papers in the past that have shown success in extracting features from 13F Filings to construct a portfolio which outperforms benchmark indices.

2.1 Best Ideas

Cohen et al. (2010) aimed to extract out the "Best Ideas" of funds via their 13F Filings. This is done by comparing the weights of the stocks in the fund's portfolio against other benchmark portfolios such as a market-capitalisation weighted portfolio. A greater difference in the weights of the stocks, also known as "portfolio tilt", would be indicative that the manager has a stronger view that the stock would perform well in the future. Hence, a fund "Best Idea" would be the stock which has the largest difference from a benchmark portfolio. A portfolio is then constructed by equally weighing each best idea.

The authors found that these "Best Ideas" generated statistically significant risk-adjusted returns over time and outperformed other positions in a fund's portfolio. Extending this to our research, it might not be sufficient to only look at the weights of the stocks in each fund's portfolio, as these weights would traditionally be correlated with features of the stocks such as market-capitalisation. Instead, we should consider extracting the portfolio tilt of the stocks in each fund's portfolio instead.

The authors also found that the performance of "Best Ideas" have a relationship with fund characteristics such as the size and historical performance of a fund. This is in line with our initial hypothesis that fund characteristics would determine the accuracy and reliability of the filing data in reflecting the actual holdings of the fund. Our research will further explore this relationship and propose filters which only retain the useful filing data.

One major limitation of the study is the assumption that trades made by the strategy are executed at the end of quarter. This is unrealistic as funds have up to 45 days after the end of quarter to file their holdings, and many funds typically file their holdings only on the deadline to hide their trading strategies. We would not have access to the filing data during the end of quarter to make the trades, and hence the strategy proposed by the study is unable to be productionized. Our study should ensure a realistic scenario where trades are executed only after the filing deadline where the data is available.

Additionally, it is important to note that the study chose to focus on the single best idea of each fund. While the study was able to provide some insights as to the usefulness of filings data in developing trading strategies, it clearly limits the full potential of the filings data in developing trading strategies as potentially valuable information could be extracted from the remaining holdings of each fund.

2.2 Systematic 13F Hedge Fund Alpha

Angelini et al. (2019) took a different approach and looked at extracting out the conviction and consensus of the funds' ideas. Conviction is measured by ranking stocks by position size in each manager's portfolio, while Consensus is measured by ranking stocks by the number of managers holding them. Two portfolios are constructed by taking equally weighted long positions in all stocks appearing in the top quintile of conviction and consensus. Additionally, a third portfolio is constructed, overlaying consensus with a conviction threshold. This is done by ranking stocks by the number of managers that has at least 7.5% of their portfolio invested in the stock and taking equally weighted long positions in all stocks appearing in the top quintile.

The study explored the hypothesis that fund characteristics do have an impact on the useful of the filings data. More specifically, the study looked at holdings of "Fundamental Equity" hedge funds, which are hedge funds that have a longer-term view on equity and compared against those outside of this category. The funds are categorized by a third-party paid data provider, Novus, which systematically identifies fund categories via a proprietary technique. The study concluded that following the holdings of "Fundamental Equity" hedge funds generated statistically significant positive returns, while following the holdings of non-"Fundamental Equity" hedge funds generated statistically insignificant negative returns. While we are unable to obtain the same set of data for use due to the costs involved, the paper agrees with our initial hypothesis that it is essential to identify funds with longer-term view on their investments as these funds would provide useful filings data which closely reflect their actual holdings.

The study found that a combination of Conviction and Consensus for stock picking led to the best portfolio performance, which also outperformed benchmark indices. In the feature engineering aspect of our research, we must ensure that we extract features that can capture both the conviction of each manager, and the consensus across all managers.

2.3 Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning

Fleiss, Cui, Stoikov and DiPietro (2020) made use of machine learning methods, such as Logistic Regression and XGBoost, to extract features from the 13F Filings as leading indicators for future stock performance.

In this study, stocks were first categorized by their industry sector. Funds were then divided into sub-funds, where each sub-fund would contain the holdings of only a single category. By classifying and dividing funds in such a manner, more granular fund features can be extracted from the data. The sub-funds were then further characterized and filtered by their size, historical performance, volatility, age, inflow and turnover. While the study did not specify the reason for this methodology, one potential advantage of splitting the funds could be to handle funds which deploy different trading strategies for different industries. However, as the authors did not provide any evidence that this has led to improvements in portfolio or model performance, our study will aim to validate the effectiveness of this methodology.

Using features which reflect conviction and consensus of a stock, such as the mean portfolio weights of funds holding a stock, the study trained binary classification models to predict the direction of a stock's future returns. A portfolio is then constructed by filtering for stocks with a predicted positive direction, taking the top 200 stocks ranked by average volume traded in the previous quarter, and equally weighing them.

The study found that the XGBoost model had the highest classification accuracy, and the portfolio constructed from the output of the XGBoost model significantly outperformed the S&P 500 in terms of annualized returns and risk-adjusted returns during backtesting. While it is more challenging, our study will explore the suitability of the model to predict both direction and magnitude of a stock's future returns, which could have the potential to achieve a more precise prediction of future returns.

We have also identified three limitations in this study. Firstly, the model was trained on only four years of data from 2013 to 2017. This is insufficient especially when the data has a quarterly frequency. A larger time horizon should be included to test the robustness of the model and recession periods should also be included for stress-testing.

Secondly, in the computation of fund-level features, such as the fund's returns and volatility, the author's methodology fell victim to lookahead bias. The author used the entire training period for the computation, rather than a lookback period. In our study, we will use an expanding lookback window in the computation to prevent lookahead bias.

Lastly, the model could only predict the direction of stock returns and not its magnitude. This implies that the model tells us what stocks to buy but is unable to solve the problem of how much of each stock to buy. While this would simplify the returns prediction model, it limits the potential portfolio construction methods that can be explored from the model output.

2.4 Constructing Equity Portfolios with Deep Reinforcement Learning

A year later, Fleiss et al. (2021) looked into using Deep Reinforcement Learning for the same purpose. The study employed a similar feature engineering methodology as that of the prior study. Funds were further decomposed into sub-funds and filtered via their sub-fund characteristics. Features reflecting conviction and consensus of the stocks were then extracted from the holdings data, and these were fed as inputs to the Reinforcement Learning model.

Two popular deep reinforcement learning models, A2C and PPO, were used in the study. The strategy acts as the agent, which takes buy and sell decisions for the portfolio. The features extracted from the holdings data as well as the existing portfolio weights represent the state of the model. The strategy performs an action each time step, rebalancing the portfolio and changing the state of the model. The reward is the positive returns gained by the portfolio as compared to the previous time step.

The study found that while using deep reinforcement learning models could generate alpha and outperform benchmark indices, it required a large number of time-steps in the training process to be able to generate actionable insights. More specifically, the strategy only worked well with a daily frequency sampling rate for the data and had significantly poorer performance with a quarterly frequency sampling rate. As our study aims to construct a quarterly-rebalancing portfolio which is easy to manage for retail investors and would not incur significant trading costs, we note that Deep Reinforcement Learning would not be a suitable model to use as it would require a daily-rebalancing portfolio to perform well.

2.5 Summary

Past Research Study	**Summary**	**Pros**	**Cons**
Cohen et al. (2010) Best Ideas	Analyzed performance of most overweighted stocks in each fund’s portfolio (Conviction).	Proposed portfolio tilt feature rather than portfolio weight. Displayed evidence of usefulness of 13F Data. Discovered relationship between fund characteristics and usefulness of 13F Data.	Unrealistic implementation of strategy. Omitted large portion of filings data.

Angelini et al. (2019) Systematic 13F Hedge Fund Alpha	Analyzed performance of stocks which are overweighted in many funds' portfolio (Conviction and Consensus)	Showed that filings data for funds with long-term view are most useful. Discovered that features combining conviction and consensus of funds led to best performance.	Fund categorization data obtained is proprietary and costly.
Fleiss et al. (2020) Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning	Introduced machine learning to learn patterns in 13F data features for predicting direction of future returns.	Introduced idea of sub-fund level analysis. Showed usefulness of Machine Learning methods in predicting stock returns.	No evidence to support claim that sub-fund level analysis is superior. Ignored prediction of magnitude of returns. Small time periods used for training and testing. Lookahead bias in fund feature computation.
Fleiss et al. (2021) Constructing Equity Portfolios with Deep Reinforcement Learning	Introduced deep reinforcement learning to learn patterns in 13F data features for portfolio construction.	Showed usefulness of Deep Learning methods in predicting stock returns.	Small time periods used for training and testing. Reinforcement Learning requires large amount of training data and frequent portfolio rebalancing

We observe that several of the past research studies fell victim to similar limitations.

A small time period being used for training and testing would lead to a model that is overfitted and not adaptable to changes in the market conditions in future periods. This is especially so because the financial markets experience cycles of varying economic regimes and stress periods. As such, it is essential to collect a sufficient large sample of data for training and testing in order to train a model that is generalizable across different market conditions.

In the context of researching a systematic trading model, it is crucial to detect any potential biases that is introduced in its implementation. These biases, which are often challenging to uncover, have a tendency to distort the results of backtesting and lead to misleadingly positive results. Consequently, such models underperform when put to actual practice. A primary objective of our research is to identify biases that have been introduced in our methodology and come up with ways to minimize or eliminate them.

In all four past research studies, the authors used naïve approaches for portfolio construction, such as equally weighing stocks which are predicted to do well or weighing stocks solely by their expected returns. There is potential to use more advanced methods for optimizing a target objective function or target portfolio metric. In our study, we will implement Markowitz's Mean-Variance Optimization (Markowitz, 1952) to maximize our portfolio's return-risk ratio. This method of constructing a portfolio would consider not only the expected returns of the stock, but also its expected volatility, allowing us to construct a portfolio with better risk-adjusted returns.

3 Data Methodology

3.1 Data Collection

3.1.1 SEC 13F Filings

There are several external data sources that provide easy access to 13F filings data, these data sources require paid subscription and do not go far back in history. This is a potential reason as to why the studies done by Fleiss et al. (2020, 2021) had limited training time frame, as the data sources that they had relied on could only provide 13F filings from the year of 2013.

To ensure that the data we collect would not be limited by external data sources, we have decided to manually scrape 13F filings data directly from the Electronic Data Gathering, Analysis and Retrieval (EDGAR) database provided by the SEC. This allows to obtain institutional holdings data from as far back as the early 2000.

The following key information are extracted from the 13F Filings:

1. Company Name – Name of fund
2. Central Index Key (CIK) – Unique identifier for fund
3. Date of filing – Date that 13F Filing is submitted
4. Date of reported holdings – Quarter end date of reported quarter
5. Name of Issuer – Company Name of the Stock Holding
6. CUSIP – Unique identifier for financial instruments
7. Value – Total value of holding in USD
8. Shares Amount – Total number of shares held

9. Put or Call – Value is set if holding is a put or call option, NA if it is a stock

However, one of the issues that we encountered during the data collection phase was the non-standardisation of the SEC 13F Filing Format before mid-2013. There was no formal specification for the format of SEC 13 Filings until mid-2013 where the SEC introduced formal guidelines defining the XML technical specification (Murphy, 2013). As such, companies submitted filings which had different form structures, making it difficult to implement a common scraper for all filings. Upon manual inspection of the different formats and accounting for several different cases in our data extraction code, we were able to scrape data from before mid-2013.

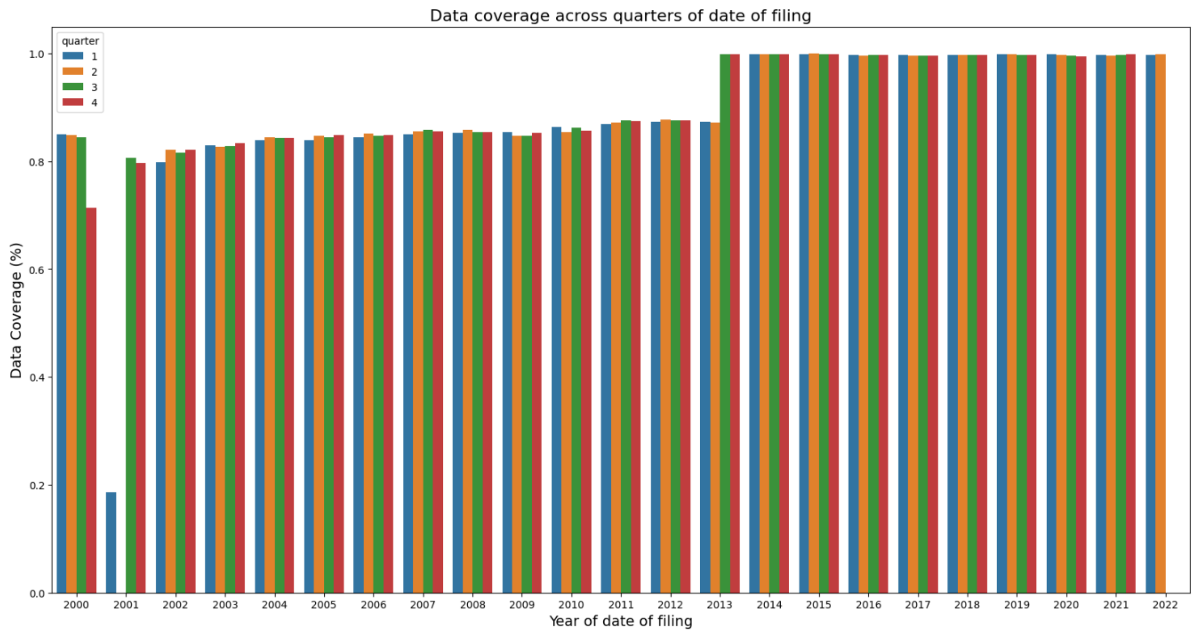


Figure 2: Data Coverage across quarters of date of filing

As shown in Figure 2, we were able to scrape approximately 80% of all 13F Filings from 2002 to mid-2013. Before 2002, our implemented scraper was unable to accurately extract the data from the filings as the formats used varied significantly. From mid-2013 onwards where the standardized formatting of 13F filings were introduced, we were able to scrape close to 100% of filings made in this period, barring a small number of edge cases.

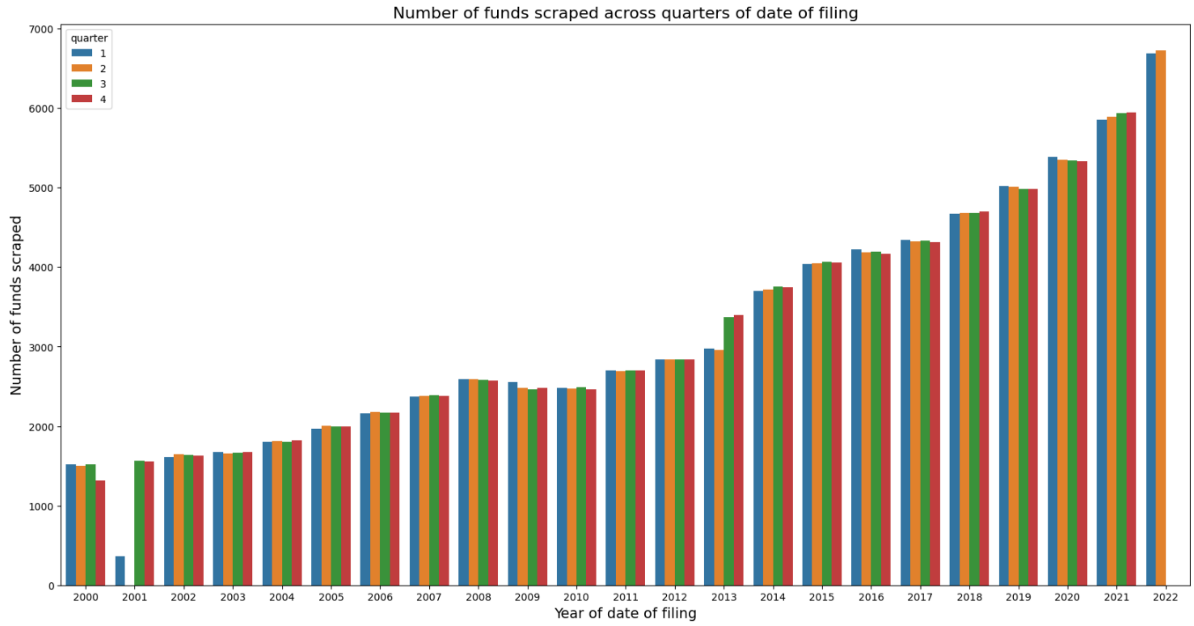


Figure 3: Absolute number of funds scraped across quarters of date of filing

From Figure 3, we can see that over the years, there has been a consistent increasing trend with regards to the number of large institutional investment managers required to make SEC 13F filings.

3.1.2 CUSIP-Ticker Information

Within the 13F Filings data, the unique identification of financial instruments are based on their CUSIP number. However, it is missing crucial information required for our research, notably the ticker symbol of the instrument, which is necessary for accessing price information as well as trade in the market.

We have, however, been able to locate an external API that offers a CUSIP-ticker mapping service, along with comprehensive information of each ticker. This service is provided by sec-api.io, and we were able to access the data through the free API calls that are available.

Using the CUSIP of a financial instrument, we were able to extract the following information:

1. Ticker Symbol – Unique identifier used to trade instrument in the market
2. Type of Security – Whether the instrument is a single stock or bond or an exchange-traded product (basket of securities)
3. Delisted – Whether the instrument has been delisted from the market
4. Sector of Security – The industry that the issuer of the security belongs to

3.1.3 Historical Price of Stocks

For our research, we have utilized the yfinance open-source python package to extract financial market data from Yahoo Finance. By specifying the ticker symbols and the date range of interest, we were able to retrieve the historical price information of the ticker symbols for the entirety of the specified period. This includes the daily open, high, low, close, and adjusted close price for each ticker symbol.

Although Yahoo Finance is a straightforward and cost-free data source for obtaining the historical price information, there is one key limitation to its usefulness. Yahoo Finance does not provide historical prices for delisted tickers. Consequently, our stock universe will not be able to include these delisted tickers. This may introduce survivorship bias to our trading strategy, as we may be excluding tickers that are delisted due to reasons such as bankruptcy or acquisition. As noted in the study by Gilbert and Strugnell (2019), ignoring delisted companies leads to survivorship bias, and as a result, an upward bias when measuring the performance of our proposed trading strategy.

3.2 Data Cleaning

To ensure that our implementation of the usage of 13F filings adheres to realistic conditions, our strategy is designed to rebalance only during each filing deadline, without access to information contained in filings submitted after the deadline. Hence, we start by removing filings that were submitted after their respective filing deadlines, 45 days after the end of the reporting quarter. This accounts for approximately 5.6% of all filings scraped. Although the SEC has the ability to impose penalties on funds who do not submit their filings on time, such penalties are typically only applied in cases where the failure to report is deemed intentional (Christoffersen, Danesh and Musto, 2015). Hence, although it is not a rare occurrence for missed deadlines, there have been relatively few instances historically where late filers have been penalized.

We filter holdings of funds to only include securities listed in the major equity exchanges in the US, namely NYSE, NASDAQ, NYSEMKT, NYSEARCA, OTC, BATS, INDEX. Additionally, we exclude non-equity holdings such as stock warrants, exchange-traded products, and put or call options. This allows us to focus on solely comparing the holdings of US single-stock equities.

We also exclude filings with formatting errors, such as the date of reported holding not being equal to the quarter end date of the reported quarter as well as the number of shares of the holding reported as zero.

Lastly, we aggregate all holdings by their unique combination of CIK and ticker symbol, allowing us to have a unique set of stock holdings information for each fund. This disregards additional information of the holdings specified in the 13F Filing such as the voting authority for the security as well as the fund's investment discretion and investment manager. We consider these information to be non-essential, and therefore unnecessary for differentiating between fund holdings.

4 Stock Returns Prediction

4.1 Fund Filter

According to the studies done by Cohen et al. (2010) and Angelini et al. (2019), the usefulness of 13F filings differ for funds with different characteristics. Specifically, funds which utilize high-frequency quantitative strategies would exhibit high frequency turnover due to significant alpha decay. Hence, the holdings data for such funds would not be useful as they do not represent a positive view of longer-term performance.

We wish to develop a series of filters based on the characteristics of the funds, to ensure that we only consider information from funds that hold a long-term view of their positions. In particular, we aim to focus on funds with low turnover and a positive track record of performance. This would be a strong indication of funds that adopt low-frequency strategies and have the ability to generate strong alpha.

The study by Fleiss et al. (2020) defined fund level features which we believe can accurately capture the different characteristics of the funds. Despite this, it has been observed that the computation process introduced lookahead bias. In light of this, we propose a novel computation method for the features that employs an expanding lookback window. As a result, the features of each fund changes quarterly, and the list of filtered funds will not remain constant from one quarter to another.

We define and extract the following fund level features:

Fund Feature	**Description**	**Definition**
Age	Age of fund in quarters	Number of quarters between latest filing date and first filing date
Historical Performance	Annualized return of fund	Geometric average of fund's annual returns
Historical Volatility	Annualized volatility of fund	Standard deviation of fund's annual returns
Size	Value of holdings reported in USD	Sum of USD value of holdings
Inflow	Inflow of new capital into fund	Change in fund size between each quarter that is not explained by the return of the fund
Turnover	Rate of which holdings are changed each quarter	Minimum of value of stocks bought or sold divided by size of fund

It should be noted that the fund level features calculated in our study represent our best possible approximation of the true features of each fund. To accurately compute the features of the fund, such as its historical performance and volatility, we require more granular data such as the daily trade information of the fund. However, such information is not publicly available due to its proprietary nature. As a result, we must rely on an assumption that the funds rebalance their positions only at the end of each quarter and maintain these positions until the end of the next quarter. This assumption is consistent with the one made by Fleiss et al. (2020).

In order to establish the fund level filters, we conduct exploratory data analysis to investigate the relationship between the fund features and their future performance. Specifically, we wish to identify characteristics of funds that would lead to excess future returns if their holdings were replicated, as this is indicative of funds which have an accurate long-term view of their holdings. To evaluate the future performance of funds based on their individual features, we rank them according to their features in comparison to other funds within the same quarter. Then, we group them into quintile groups and compute the average returns for the following quarter as an indicator of their future performance.

Correlation between fund features

Figure 4: Correlation heatmap between fund features

We identify a strong negative correlation between a fund's historical performance and its historical volatility. This implies that highly volatile funds tend to have a worse track record of performance as compared to less volatile funds.

Age

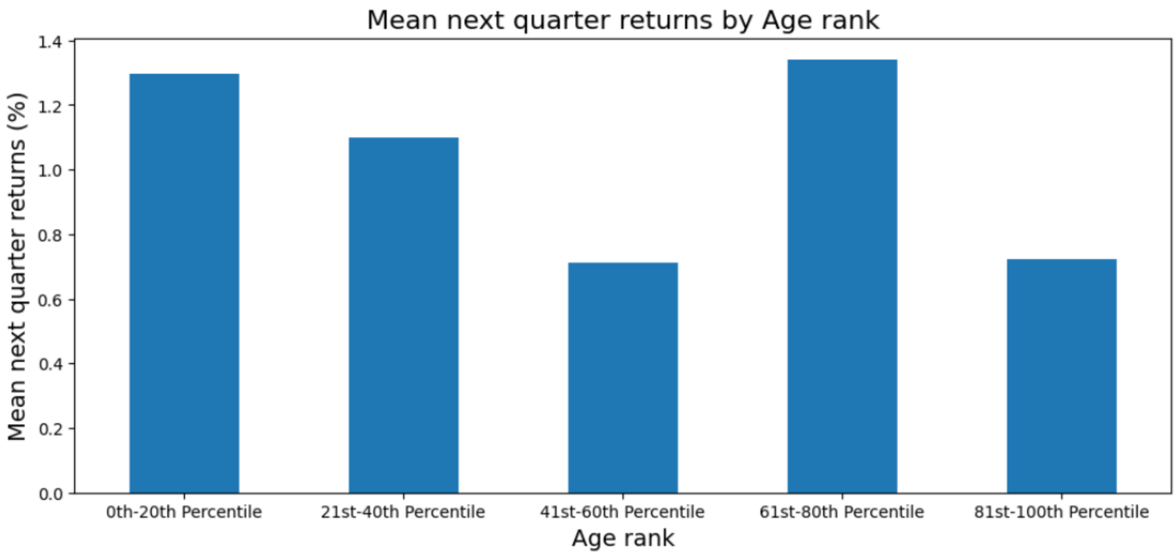


Figure 5: Relationship between fund's age and future performance

We do not observe a clear directional relationship between a fund's age and its future performance. However, according to Jones (2006), we should expect younger hedge funds to outperform older hedge funds. While this does not align our findings, the difference can be attributed to our computation of fund age using the first 13F filing date. As funds are only required to file their holdings when they manage more than $100USD AUM, the funds in our analysis are likely to have already existed before the first filing date and belonging to the "older" age group defined in the previous study.

As the fund's age is uncorrelated with the usefulness of its filings, we will not propose a filter for funds based on their age.

<u>Historical Performance</u>

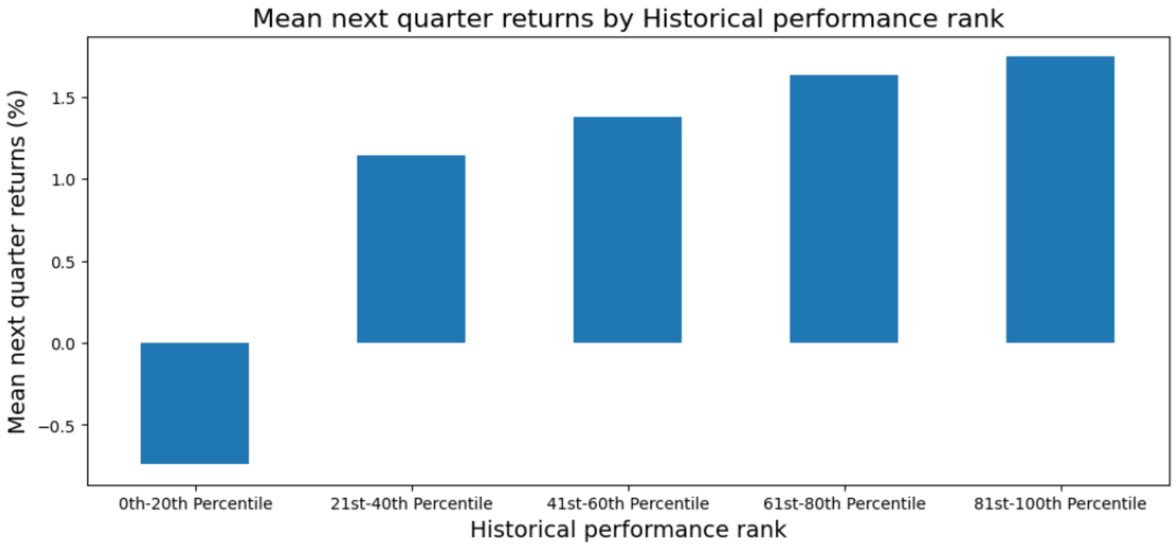


Figure 6: Relationship between fund's historical performance and future performance

We observe a clear positive relationship between a fund's historical performance and its future performance. This is intuitive as it tells us that funds with a history of strong performance are likely to continue performing well in the future. It is also expected because we are computing historical performance via the performance of the filed quarterly holdings. If tracking the holdings of a fund's past filings has yielded excess returns, it is probable that the fund has a strong long-term view of its holdings. This, in turn, should lead to stronger returns should we track the holdings of a fund's future filings.

We propose to filter away the bottom quintile of funds ranked by historical performance each quarter.

Historical Volatility

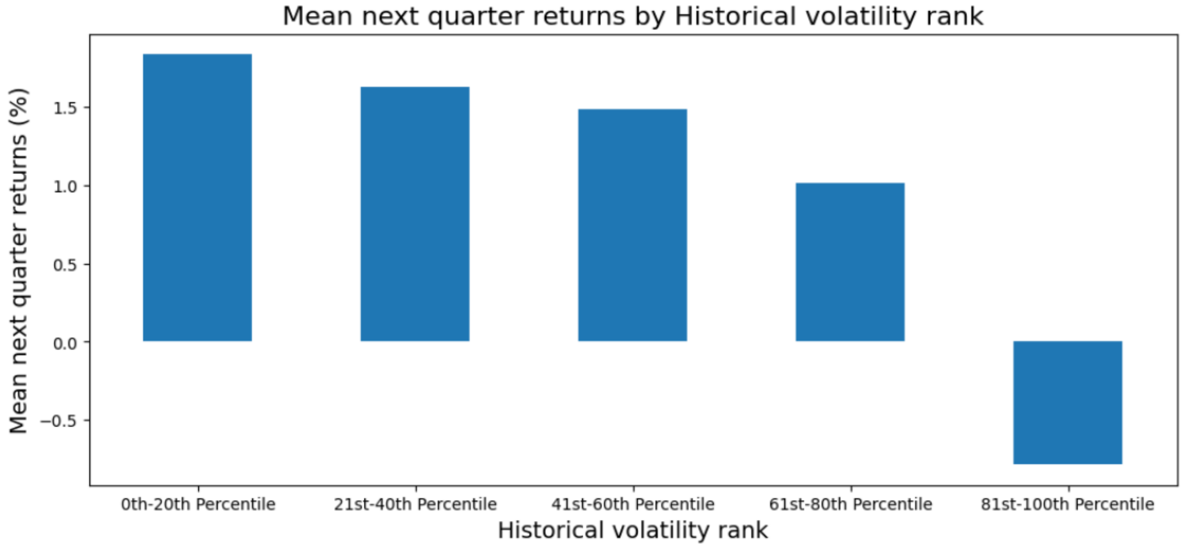


Figure 7: Relationship between fund's historical volatility and future performance

Our findings indicate a negative monotonic relationship between a fund's historical volatility and its future performance. Specifically, we observe that the funds belonging to the group of the highest historical volatility performed significantly worse in the future than those of the other groups. In fact, it was the only group to average a loss. This is also expected because we hypothesize that market makers and high-frequency trading firms make up this group. As such firms do not hold a long-term view of their holdings, replicating their quarterly holdings would likely lead to highly volatile performances caused by randomness. Based on these observations, we do not expect favourable performance from tracking the future holdings of such funds.

We propose to filter away the top quintile of funds ranked by historical volatility each quarter.

Size

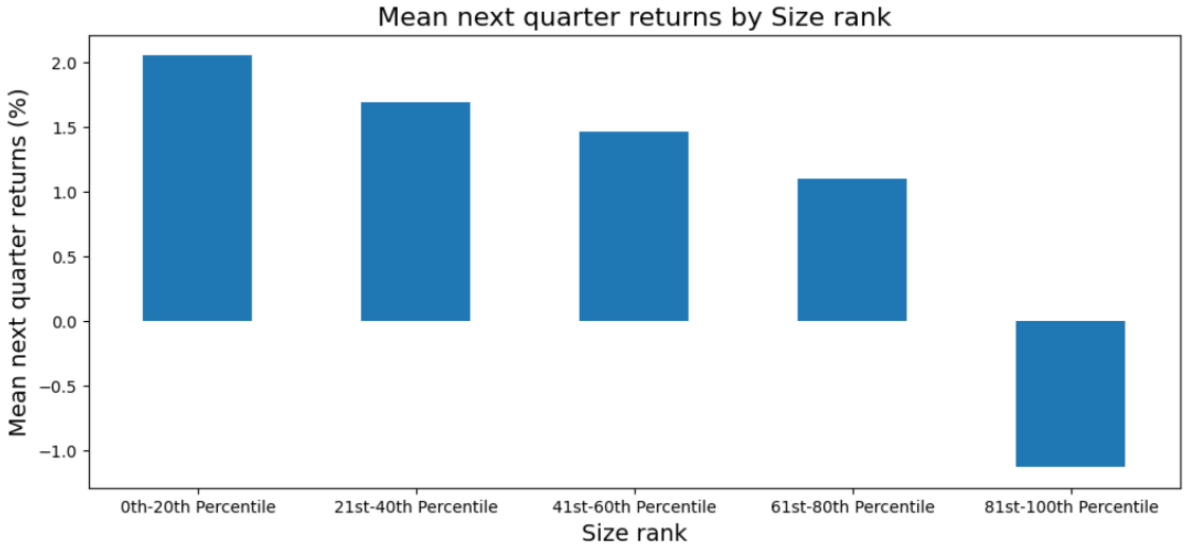


Figure 8: Relationship between fund's size and future performance

Our findings show that there is a negative monotonic relationship between the fund's size and its future performance. This agrees with Jones (2006), which also found that smaller funds consistently outperformed larger funds. This could be due to the ability of smaller funds to take advantage of minor market inefficiencies and opportunities that might not be feasible for larger funds to pursue. However, it is surprising that the largest funds would average such negative performance. Upon diving deeper into this specific group of funds, we find that the group contains brokers, market makers and funds which employ mid-to-high frequency quantitative strategies, which could explain the poor performance when tracking their holdings.

We propose to filter away the top quintile of funds ranked by size each quarter.

Inflow

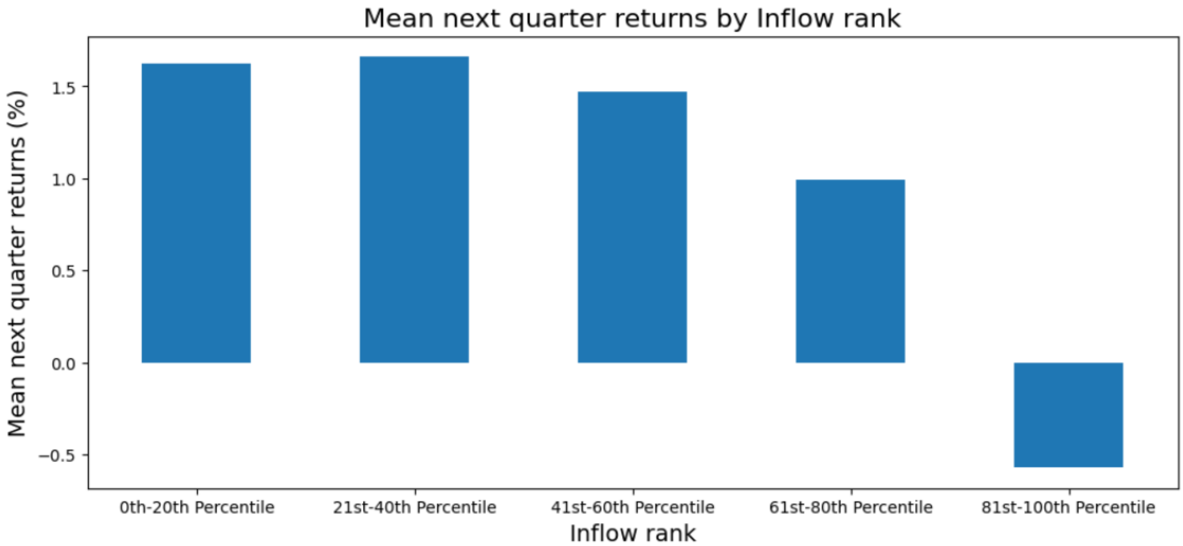


Figure 9: Relationship between fund's inflow and future performance

We observe a clear negative relationship between a fund's inflow and its future performance. This finding contradicts the common expectation that funds with larger inflow outperform those with smaller inflows, as inflow is a measure of investor sentiments of the fund (Chen, 2022). A possible explanation for this is the violation of the assumption in the computation of a fund's inflow, which assumes that a fund's reported holding remains constant until the end of the next quarter. Funds that trade in mid-to-high frequency would make more trades within the quarter, and potentially earn more profits within the quarter, leading to understated returns and overstated inflows. Upon further examination of this specific group of funds, we discovered that the group does indeed contain brokers, market makers and funds which employ mid-to-high frequency strategies.

We propose to filter away the top quintile of funds ranked by inflow each quarter.

Turnover

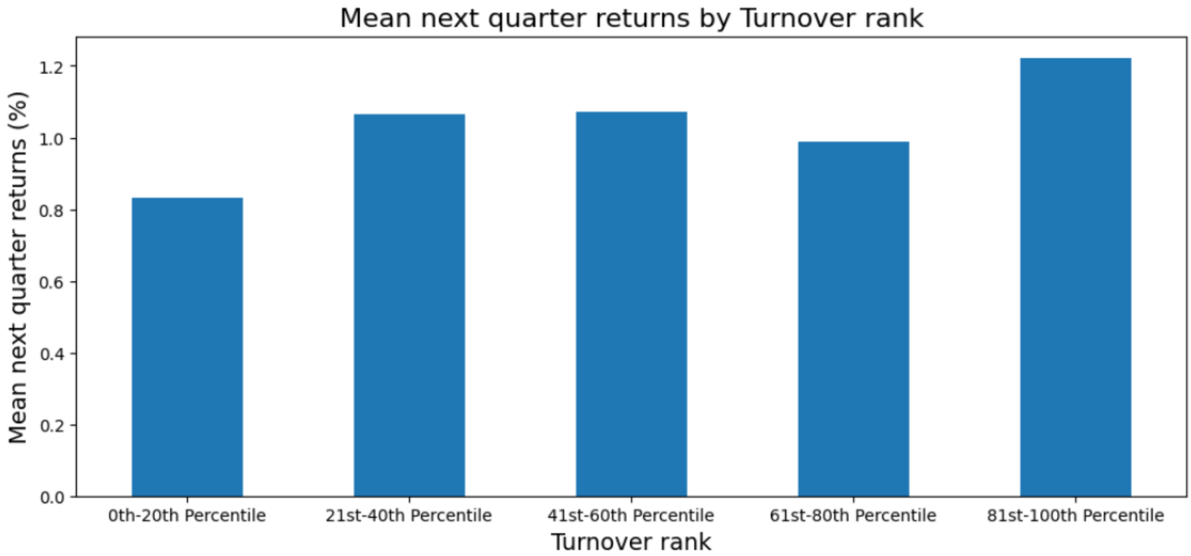


Figure 10: Relationship between fund's turnover and future performance

Our study indicates that there is no significant relationship between a fund's turnover and its future performance. This contradicts our initial hypothesis that market makers and high frequency trading firms should typically have high turnover, resulting in poor future performance. Upon further analysis of these firms, we came to a realization that these firms employ market-making and arbitrage strategies in the same basket of stocks every quarter. Hence, they are often holding a similar set of stocks at the end of each quarter, resulting in an understated turnover ratio when calculated via their quarterly filings.

Given that a fund's turnover ratio cannot differentiate between funds with a long-term or short-term view of their holdings, we will not propose a filter for funds based on their turnover.

Sample Analysis

We examine the fund level features of three different types of funds, a market maker firm, a mid-to-high frequency quantitative fund and an asset management firm.

date_holding	Age Quintile	Historical Performance Quintile	Historical Volatility Quintile	Size Quintile	Inflow Quintile	Turnover Quintile
2017-03-31	1	1	4	5	5	5
2017-06-30	1	1	5	5	5	5
2017-09-30	2	1	5	5	5	5
2018-03-31	2	1	5	5	5	5
2018-09-30	2	1	5	5	5	5

Figure 11: Market Maker -

Jane Street's quintile of features for each quarter from 2017-2018

As a market maker does not hold directional view of its holdings, we do not wish to use data from its holdings in our strategy. The filings of Jane Street are filtered away each quarter as it does not pass the filters for historical performance, volatility, size and inflow.

date_holding	Age Quintile	Historical Performance Quintile	Historical Volatility Quintile	Size Quintile	Inflow Quintile	Turnover Quintile
2017-03-31	4	1	5	5	4	5
2017-06-30	4	1	5	5	4	5
2017-09-30	4	1	5	5	4	5
2017-12-31	4	1	5	5	4	5
2018-03-31	4	1	5	5	4	5
2018-06-30	4	1	5	5	4	5
2018-09-30	4	1	5	5	4	5

Figure 12: Quant fund -

D.E. Shaw's quintile of features for each quarter from 2017-2018

As D.E. Shaw does not hold a long-term view of its holdings, we do not wish to use data from its holdings in our strategy. The filings of D.E. Shaw are filtered away each quarter as it does not pass the filters for historical performance, volatility and size.

date_holding	Age Quintile	Historical Performance Quintile	Historical Volatility Quintile	Size Quintile	Inflow Quintile	Turnover Quintile
2017-03-31	4	5	3	4	3	5
2017-06-30	4	4	3	4	3	5
2017-09-30	4	4	3	4	3	5
2017-12-31	4	4	3	4	3	5
2018-03-31	4	4	3	4	3	5
2018-06-30	4	4	3	4	3	5
2018-09-30	4	4	3	4	3	5

Figure 13: Asset Management Firm -

Bridgewater's quintile of features for each quarter from 2017-2018

As Bridgewater holds a long-term view of its holdings, we wish to include data from its holdings in our strategy. The filing of Bridgewater passes our proposed set of filters.

<u>Summary</u>

Proposed set of filters:

- Historical Performance > 20th Percentile Rank
- Historical Volatility < 80th Percentile Rank

- Size < 80th Percentile Rank
- Inflow < 80th Percentile Rank

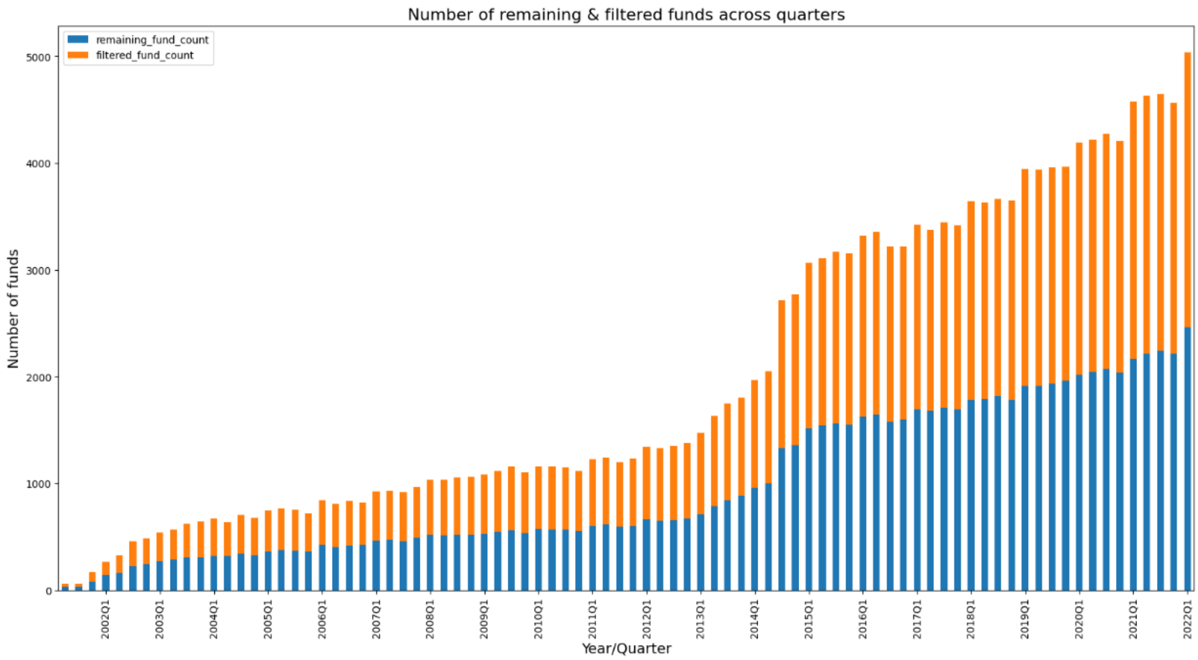


Figure 14: Number of remaining and filtered funds each quarter

By implementing our proposed list of filters, approximately 50% of funds are filtered every quarter. In contrast to the study by Angelini et al. (2019) where their definition of funds with a long-term view made up more than half of all funds, our set of filters appear to be more stringent. It is therefore probable that the funds that remained after filtering consist predominantly of those with a long-term view on stocks.

4.2 Stock Universe

For our study, our stock universe would consist only of constituents of the S&P 500 Index. The S&P 500 Index consists of the 500 largest market-capitalisation equities listed in the US market. However, if we use the latest constituents of the S&P 500 Index, our analysis will fall victim to survivorship bias. This is because over time, better performing stocks will be added to the index while poorer performing stocks will be dropped off. Hence, we would be limiting our portfolio to stocks that are likely to have performed well in the recent few years. Instead, we will build a portfolio consisting of all the unique constituents of S&P 500 Index from the year 2000 up to the year 2022, giving us a total of 960 unique tickers. This reduces survivorship bias as it consists of tickers that have been dropped off the index due to poor performance, as well as tickers that have been added to the index due to strong performance. Koker (2019) proposed the same approach to creating a survivorship bias-free data set of the S&P 500 Index.

4.3 Feature Engineering

After filtering away funds which do not adhere to our proposed set of filters, we will perform feature engineering using the holdings data of the remaining funds to obtain stock-level features for our stock universe which reflect the conviction and consensus of the fund managers.

The following features are extracted from the holdings data as well as historical price data:

1. x_1 - Percentage of funds in the industry sector holding the stock
2. x_2 - Mean portfolio weights of funds in the industry sector holding the stock
3. x_3 - Median portfolio weights of funds in the industry sector holding the stock
4. x_4 - 25th percentile portfolio weights of funds in the industry sector holding the stock
5. x_5 - 75th percentile portfolio weights of funds in the industry sector holding the stock
6. x_6 - Last 21 trading day historical return of stock
7. x_7 - Last 42 trading day historical return of stock
8. x_8 - Last 63 trading day historical return of stock

As the distribution of portfolio weights for each stock is often non-symmetrical and highly skewed in either direction, we believe that including features that account for such distributional properties would aid in the forecasting of stock performance. For example, a right skewed portfolio weights distribution would signify lower consensus, while a left skewed portfolio weights distribution would signify higher consensus.

In addition, we include features that incorporate recent price movements as stock prices are believed to exhibit momentum behaviour in very short periods and exhibit mean-reversion behaviour in longer periods of time (Poterba and Summers, 1988).

4.4 Model Methodology

4.4.1 Model Design

By employing the stock-level features, we wish to train a machine learning model capable of forecasting the future performance of each individual stock in our stock universe. Our hypothesis posits that there are hidden patterns in the features that can serve as leading indicators of a stock's future performance.

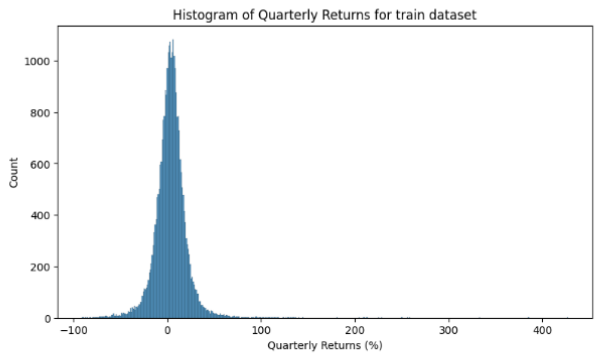
The study by Fleiss et al. (2020) approached the stock prediction task as a classification problem, where the model's objective was to predict the direction of the stock returns, disregarding its magnitude. According to Ou and Wang (2009), predicting the direction of stock returns is a more straightforward and accurate problem compared to predicting their actual returns. However, our research aims to delve into a more intricate problem involving the prediction both the direction and magnitude of the stock returns. This is because we believe that the magnitude of the stock returns is a crucial input for optimal portfolio construction.

Specifically, we will experiment with three different machine learning models: Linear Regression, XGBoost and LightGBM model. The predictor variables of the model would be the stock-level features that we have extracted from our data sources, while the response variable of the model would be the next quarter return of the stock.

Our training dataset will consist of data from the start of 2006 to the end of 2018 while our test dataset will consist of data from the start of 2019 to the second quarter of 2022. We do not include data before 2006 as there are a limited number of fund filings (<300), and data that is too old may not be as relevant for our model. Both our training dataset and test dataset contain stress periods, namely the 2008 Financial Crisis and the 2020 Covid Crash respectively, allowing us to train and evaluate the model on different economic regimes.

To evaluate the accuracy of our proposed prediction models, we will look at Root Mean Square Error (RMSE) which is a standard metric used in evaluating regression models, as well as Information Coefficient (IC), which is a widely accepted measure in the financial industry to quantify the predictive ability of a strategy (Zhang, Guo and Cao, 2020). Information Coefficient is defined as the correlation between the actual and predicted stock returns. However, correlation can be computed using either the Pearson or Spearman's definition. Pearson correlation evaluates the linear relationship between two variables based on its raw values, whereas Spearman correlation evaluates the monotonic relationship between two variables based on its ranked values. In our study, we will explore both definitions to evaluate our models.

4.4.2 Outlier Analysis



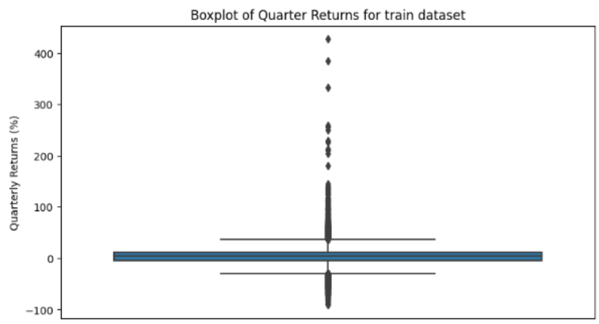


Figure 15: Histogram and Boxplot of Quarterly Returns for train dataset

We observe from our train dataset that the quarterly returns of stocks contain many outliers outside the interquartile range. This is expected as past research has also observed that while the distribution of stock returns follows a symmetric bell-shaped curve, the distribution has fat tails, where the probability of observing extreme stock movements is greater than a normal distribution (Toth and Jones, n.d.).

These outliers could potentially skew our model's predictions as it produces much larger residuals than other data points. A common approach to handling outliers in the stock returns distribution is winsorization, where extreme values of the distribution are limited to reduce the effects of these outliers. A study done by Dai and Chang (2021) found that by limiting the stock returns to its interquartile range, they were able to improve the out-of-sample predictive ability of their stock return forecast model. We adopt a similar approach where the upper threshold and lower threshold of stock returns is determined by the interquartile range of the stock returns distribution in the train dataset. The computation of the thresholds is done in the train dataset to prevent lookahead bias.

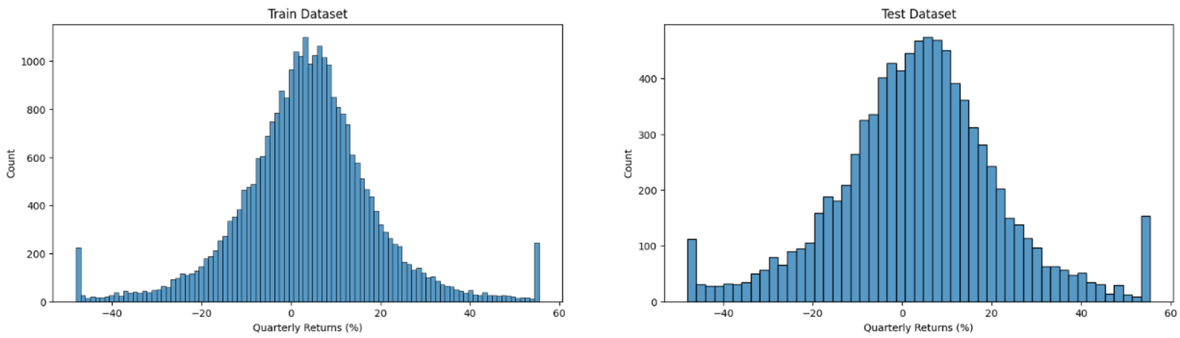


Figure 16: Histogram of Quarterly Returns for train and test dataset after transformation

4.4.3 Exploratory Data Analysis

We start by visualizing the distribution of the predictor variables and its relationship with the response variable via a scatterplot. The visualizations are documented in Appendix A. To summarize the findings of the exploratory analysis, we do not observe any linear relationships between the predictor variables and the response variable.

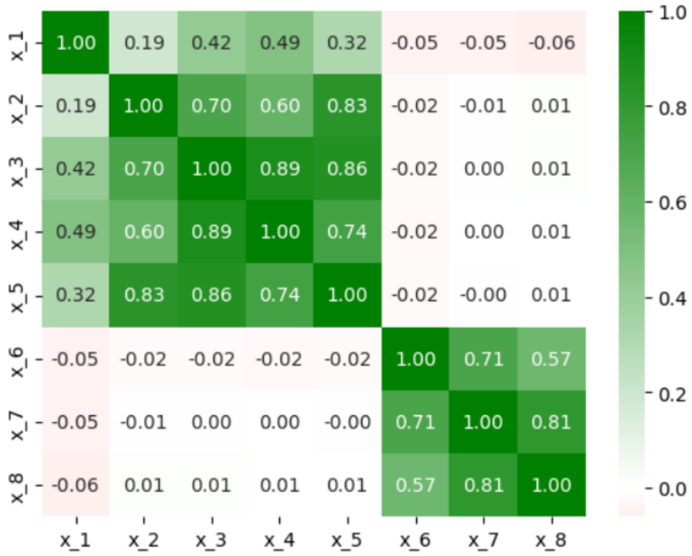


Figure 17: Correlation Heatmap between predictor variables

We find that the features describing the distribution of the portfolio weights, x_2 to x_5, are highly correlated with each other. Similarly, features describing the distribution of the recent stock price movements, x_6 to x_8, are highly correlated with each other. These two distinct sets of features are uncorrelated with each other.

4.4.4 Model Experimentation

4.4.4.1 Baseline Model

As a starting point to the prediction problem, we employ a naïve method of using price momentum to predict stock returns. We will use the previous quarter's stock return as a prediction for the subsequent quarter's stock return. This serves as a baseline model that sets the foundation for further iterative development. Our subsequent models which are more complex, are expected to outperform the results of this baseline model.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Baseline	0.2107	0.0632	0.0359	0.2659	-0.0180	-0.0118

4.4.4.2 Linear Regression

We develop a multiple linear regression model which examines the relationship between the response variable and multiple predictor variables. The model assumes a linear relationship between the variables and aims to find the best-fit line that can predict the response variable based on the predictor variables. It allows for the identification of the effect of each predictor variable while controlling for the other variables in the model.

The output of the model is summarized below.

Variable	**Coefficient**	**P-Value**	**Significant (Y/N)**
Constant	0.0415	0.000	Y
x_1	-0.0524	0.000	Y
x_2	-0.5309	0.002	Y
x_3	-0.1021	0.784	N
x_4	0.4559	0.251	N
x_5	0.2189	0.194	N
x_6	-0.0543	0.000	Y
x_7	-0.0993	0.000	Y
x_8	0.1456	0.000	Y
R^2: 0.011		$Adjusted\ R^2$: 0.011	

The model tells us that the mean and median portfolio weights of funds holding the stock, as well as the recent price movements of the stock, have a statistically significant linear relationship with future returns. Other variables in the model do not have a statistically significant relationship with future returns.

However, from our exploratory analysis, we found that the predictor variables do not have a linear relationship with future returns. This violates the assumption of the linear regression model which assumes that the predictor variables have a linear relationship with the response variable. Furthermore, we have identified high correlation values between the predictor variables which introduces multicollinearity. Multicollinearity generates high variance of the estimated coefficients. Hence, the results of the linear regression may be unreliable due to its inability to capture nonlinear relationships as well as the presence of multicollinearity.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Linear Regression	0.1530	0.1053	0.0830	0.1923	-0.1092	-0.0635

The linear regression model is able to achieve a lower RMSE score in both the training and test dataset as compared to the baseline model. However, the information coefficient of the test set predictions is negative and even lower than that of the baseline model. This suggests that the model is not generalizable and performs poorly in stock return prediction during the withheld time period.

4.4.4.3 XGBoost

XGBoost is a powerful and popular machine learning model designed for building efficient and accurate gradient boosting models. Gradient boosting is an ensemble learning technique that combines multiple weak learners, such as decision trees, into a single strong learner. XGBoost uses a gradient boosting framework to iteratively train many decision trees in sequence, each one trying to correct the errors of its predecessor.

We start by training a benchmark XGBoost model using the default set of parameters as defined in the functions of the XGBoost open-source python package. The objective function of the model aims to minimize squared loss.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Benchmark XGBoost	0.1218	0.6555	0.5145	0.1921	0.0920	0.0480

The benchmark XGBoost model performed exceptionally on the train set with an unrealistic information coefficient of more than 0.5, yet its performance dropped significantly on the test set to less than 0.1. It is highly probable that the model has overfitted during the training phase, resulting in a model that performs well on the data it is trained with, but generalizes poorly and produces subpar results when tested on withheld data.

To reduce overfitting of the model, we will tune the hyperparameters of the model to optimize its performance. These hyperparameters control the complexity of the model, allowing us to reduce the complexity of the model in order to improve its generalizability.

To tune these hyperparameters, one can use techniques like grid search, random search, or Bayesian optimization. Grid search involves trying all possible combinations of hyperparameters within a predefined range, while random search randomly selects combinations of hyperparameters to evaluate. Bayesian optimization is a sophisticated technique that uses a probabilistic model to predict the best set of hyperparameters to experiment next based on previous results. In our study, we will make use of Bayesian optimization which allows us to search a wide range of combinations of hyperparameters without the need for long training time. The objective of the Bayesian optimization process is to select the best hyperparameters to minimize the RMSE of the model when evaluated using unseen data.

During each step of the hyperparameter tuning process, the model will be evaluated via k-fold cross-validation. Our training data will be split into k subsets, and the model will be trained on k-1 of the subsets while using the last remaining subset for validation. This process is repeated multiple times, with each subset serving as the validation set exactly once, and the performance metrics of the validation set are averaged across all folds. Cross-validation is a crucial step in hyperparameter tuning because it allows us to access the model's performance on data that the model has not seen during training without looking into the withheld test dataset. This helps us to select hyperparameters that generalize well to new data and avoid overfitting.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Tuned XGBoost	0.1497	0.2609	0.1734	0.1864	0.1048	0.0648

After the tuning process, we observe that the tuned model performed worse than the untuned model across all three evaluation metrics for the train set, while the tuned model performed better than the untuned model across all three evaluation metrics for the test set. This implies that hyperparameter tuning has allowed us to reduce overfitting and improve our model's generalizability to unseen data.

4.4.4.4 LightGBM

Similar to XGBoost, LightGBM is a high-performance gradient boosting framework that uses a tree-based learning algorithm. However, the main difference between the two lies in the way that the decision trees are constructed. LightGBM uses a novel approach called "Histogram-based Gradient Boosting," which bins the continuous feature values into discrete bins and then constructs decision trees based on these bins. This allows for faster and more memory-efficient training.

We will first train a benchmark LightGBM model using the default set of hyperparameters defined in the functions of the LightGBM open-source python package. The objective function of the model aims to minimize squared loss. Following that, we use the same approach as we did in the development of the XGBoost model, where we make use of Bayesian Optimization and K-Fold Cross Validation techniques to tune the hyperparameters of the model.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Benchmark LightGBM	0.1392	0.4971	0.3646	0.1876	0.1019	0.0737
Tuned LightGBM	0.1460	0.3552	0.2659	0.1877	0.0647	0.0542

We encounter the same challenge with the benchmark LightGBM model, where the default set of hyperparameters causes overfitting of the model. This can be observed by the strong performance on the train set followed by a sharp drop in performance on the test set. After hyperparameter tuning, we were able to reduce overfitting as shown by the drop in the performance of the train set. However, we note that hyperparameter tuning did not improve the performance of the model on the test set, as the RMSE remained relatively constant while both IC values dropped.

4.4.5 Final Model

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Baseline	0.2107	0.0632	0.0359	0.2659	-0.0180	-0.0118
Linear Regression	0.1530	0.1053	0.0830	0.1923	-0.1092	-0.0635
Tuned XGBoost	0.1497	0.2609	0.1734	0.1864	0.1048	0.0648
Tuned LightGBM	0.1460	0.3552	0.2659	0.1877	0.0647	0.0542

The XGBoost and LightGBM model outperformed the baseline model across the three performance metrics. However, the Linear Regression model has a lower RMSE yet higher information coefficient than the baseline model. The XGBoost model performs the best when evaluated using both RMSE and information coefficient, while the LightGBM model performs just slightly worse than the XGBoost model. These results are expected as we observed non-linear relationships between the predictor variables and the response variable. Linear Regression would be unable to capture such complex relationships, while LightGBM and XGBoost have proven itself to be capable of modelling complex relationships.

Depending on the specific use case of each stock prediction model, one might choose to optimize for a different performance metric. Regression methods in general are more commonly evaluated by RMSE. When applied to the case of stock returns prediction, optimizing for RMSE implies that we want our model to most accurately predict the future value of the stock. Maximizing information coefficient implies that we want our model to be able to maximize the positive relationship between the predicted values and the actual values.

For our research, we will focus on optimizing for Spearman Information Coefficient. Predicting the actual future returns of each stock is an extremely difficult task, and yet does not add much value in our portfolio construction. When determining the asset allocation in our portfolio, the relative order or ranking of the stocks is more important than its price or returns (Saha, Gao and Gerlach, 2021). Since the XGBoost model outperforms the other models for all the performance metrics that we considered, we will use the predictions of the XGBoost model when constructing our portfolio.

According to industry insights from Gleiser and McKenna, portfolio managers generally view strategies with greater than 0.05 IC as “good”, while strategies with greater than 0.1 IC are viewed as “very good”. As our model falls between the two categories, we can infer that our model is satisfactory in its predictive power. Furthermore, our model also outperformed the best performing stock prediction model explored by Zhang, Guo and Cao (2020) which generated an IC of 0.045, although there are differences in the date range and assets that is being considered.

4.4.6 Comparison with methodology in past literature

In the study by Fleiss et al. (2020), the author further divided funds into sub-funds based on the industry sector of the stock holding. We wish to test the effectiveness of this methodology. We adopt the same approach of first dividing funds into their respective sub-funds. In order to come up with a set of sub-fund filters, we conduct exploratory data analysis to investigate the relationship between the sub-fund features and their future performance. The findings are similar to that of the analysis of the fund features, as documented in Appendix B. As such, we propose the same set of filters to be used for sub-funds. The input to the feature engineering step would consist of sub-fund holdings data instead of fund holdings data. We then develop a new hyperparameter-tuned XGBoost model and document its results.

	Train Set			**Test Set**		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Tuned XGBoost – Fund	0.1497	0.2609	0.1734	0.1864	0.1048	0.0648
Tuned XGBoost – Sub-fund	0.1485	0.2904	0.1855	0.1873	0.0566	0.0457

The introduction of sub-fund level holdings for feature generation resulted in a decrease in model performance on the test set across all evaluation metrics. Consequently, we conclude that an additional step to divide funds into sub-funds is unnecessary and counterproductive, as it not only adds complexity but also fails to enhance the model's value.

5 Portfolio Allocation

5.1 Benchmark Portfolio

We will benchmark our constructed portfolios against two different baseline portfolios. The first baseline portfolio, termed SPY Portfolio, would allocate its entire portfolio weights to buying and holding SPY shares. SPY is one of the most popular exchange-traded funds that seek to track the S&P 500 index. This is akin to a simple buy-and-hold strategy of a major market indices, gaining profits due to market beta without seeking any market alpha.

Our second baseline portfolio, which we will term Stock-Universe portfolio, would be a portfolio that holds equal weights of all stocks in our stock universe. This serves to eliminate the possibility that the performance of the constructed portfolio is caused by the definition of our stock universe.

5.2 Alpha Portfolio

5.2.1 Portfolio Construction

We construct two portfolios using the output from our stock prediction model. We would expect these portfolios to outperform the baseline portfolio, as it serves to prove the predictive ability of our stock prediction model. The portfolio will rebalance the day after the filing deadline of each quarter using the features generated from the quarterly filings, and its allocation remains constant until the next rebalancing.

Our first alpha portfolio, termed Equally-Weighted Alpha portfolio, ranks the stocks by predicted returns and selects the top 200 stocks to construct a portfolio. The portfolio will equally weigh each of the 200 stocks.

The second alpha portfolio, termed Alpha-Weighted Alpha portfolio, similarly ranks the stocks by predicted returns and selects the top 200 stocks to construct a portfolio. However, instead of equally weighing each of the 200 stocks, the stocks are weighted proportional to the magnitude of their predicted returns.

5.2.2 Comparison against Benchmark

We evaluate our alpha portfolios against the benchmark portfolios based on the test date range, where our portfolios are constructed using predictions generated from our stock prediction model.

On top of considering the annualized return and volatility of the portfolio which are simple metrics to reflect the returns and risk of the portfolio, we introduce two additional portfolio metrics. Return-risk ratio is calculated by dividing annualized returns by annualized volatility. It represents the expected returns per unit risk, allowing us to compare the risk-return trade-off of different portfolios. Value-at-risk 5% is an alternative risk measure that quantifies the potential losses that one might lose with a 5% probability. The volatility of the portfolio, calculated as the standard deviation of portfolio returns, penalizes for upside moves as much as downside moves. However, value-at-risk simply captures downside risk and is the most commonly-used risk measure to do so.

	Test Date Range [2019-02-15 to 2022-08-15]			
	Annualized Returns	Annualized Volatility	Return-Risk Ratio	Value-At-Risk 5%
SPY portfolio	15.06%	22.50%	0.6693	-21.84%
Stock-Universe portfolio	18.06%	26.21%	0.6890	-24.93%
Equally-Weighted Alpha portfolio	24.63%	30.63%	0.8041	-25.60%
Alpha-Weighted Alpha portfolio	27.54%	31.95%	0.8619	-24.86%

Our Stock-Universe portfolio performs slightly better than the SPY portfolio, with a higher return-risk ratio. This tells us that our definition of stock universe has likely caused an upward bias in portfolio performance.

However, this is not a significant issue as our alpha portfolios significantly outperform both the Stock-Universe portfolio and SPY portfolio, displaying the value-add that our model prediction offers. Our alpha portfolios generated a return-risk ratio greater than 0.8 while both the benchmark portfolios generated a return-risk ratio of less than 0.7. This is due to the portfolios' ability to allocate to stocks that generate much higher returns. While the side effect of an alpha-seeking strategy is its increased volatility, our alpha portfolios were able to generate a larger increase in returns compared to its increase in risk, resulting in an increase in the portfolios' return-risk ratio. This illustrates the predictive ability of our stock prediction model to identify better performing stocks for our portfolio to allocate to.

When we compare both alpha portfolios, we see that our Alpha-Weighted Alpha portfolio outperforms our Equally-Weighted Alpha portfolio. Specifically, weighing the portfolio by the magnitude of predicted returns allowed us to further improve the returns as well as the return-risk ratio of the portfolio, albeit increasing the portfolio's volatility. However, when we look at the downside risk measure, we see that value-at-risk dropped. This implies that the upside movements caused by the larger portfolio returns led to greater volatility, although downside risk fell. Hence, we can conclude that our Alpha-Weighted Alpha portfolio is able to generate greater returns than our Equally-Weighted Alpha portfolio while taking on less downside risk.

The positive difference between the alpha portfolios also illustrates the effectiveness of having our stock prediction model predict not just the direction of returns, but also its magnitude.

5.2.3 Parameter Adjustments

In the construction of our Alpha-Weighted Alpha portfolio, the number of stocks for consideration was fixed at 200. However, we recognize the possibility of adjusting the risk appetite of the portfolio by varying the number of stocks in the portfolio. We will explore varying the number of assets in the portfolio, and its effects on the performance of the Alpha—Weighted Alpha portfolio.

	Test Date Range [2019-02-15 to 2022-08-15]			
	Annualized Returns	Annualized Volatility	Return-Risk Ratio	Value-At-Risk 5%
N=50	31.86%	37.41%	0.8515	-29.5%
N=100	29.69%	34.70%	0.8556	-27.22%
N=200	27.54%	31.95%	0.8619	-24.86%
N=500	22.59%	28.43%	0.7945	-24.04%

When we decrease the number of stocks to hold in our portfolio, we concentrate our holdings on a fewer number of stocks and reduce the diversification of our portfolio. Consequently, this results in an increase in annualized returns as we allocate greater weights to stocks that we predict to perform the best. However, because of the loss in diversification, the volatility and value-at-risk of the portfolio increases, indicating an increase in risk. The return-risk ratio remains largely similar, denoting that risk increases proportional to returns.

Conversely, when increasing the number of stocks to hold in our portfolio, we increase diversification of our portfolio and reduce our portfolio’s allocation to each individual stock. This results in a decrease in returns coupled with a decrease in risk. The return-risk ratio drops significantly, as the returns of the portfolio decreased substantially, while the risk of the portfolio only decreased slightly.

The optimal number of stocks to hold in our Alpha portfolio ultimately depends on the risk appetite of the individual investor. Investors who are willing on to take on more risk can choose to decrease the number of stocks in their portfolio and concentrate their holdings, while risk-adverse investors can increase the number of stocks to diversify their holdings.

5.3 Alpha-Sigma Portfolio

5.3.1 Portfolio Construction

While we have proven that the output of our stock prediction model allowed us to generate a portfolio which outperforms a buy-and-hold strategy, we note that more can be done as our alpha portfolios do not consider the risk of individual stocks as well as the relationship between pairs of stocks in its asset allocation process.

Markowitz (1952) proposed the mean-variance portfolio optimization approach and was awarded the Noble Prize in Economics in 1990. The model considers both the expected return and risk of the portfolio and aims to maximise the expected return for any given level of risk or minimize expected risk for any given level of return. Using the theory proposed by Markowitz, we will design a portfolio construction methodology which takes into account both the returns and risk of the individual assets. We term this portfolio as Alpha-Sigma portfolio.

We define our objective function as:

$$\max_w U = E(r_p) - \gamma\sigma_p^2$$

where,

$E(r_p)$ is the expected return of the portfolio.

σ_p^2 is the variance of the portfolio.

γ is the coefficient of risk aversion.

The expected return and variance of the portfolio is calculated as:

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i)$$

$$\sigma_p^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(r_i, r_j)$$

where,

w_i is the weight allocation to asset i.

r_i is the returns of asset i.

n is the number of assets in the portfolio.

$w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$ to indicate a long-only portfolio.

From the objective function formulation, we observe two key inputs to the portfolio construction process. These include the expected returns of each constituent asset in the portfolio, as well as the covariance matrix between them. The covariance matrix plays a crucial role in the process as it not only incorporates the volatility of each individual asset but also the correlation between assets, which is a measure of diversification that also impacts portfolio variance.

There exists extensive research on covariance forecasting, such as the study by Symitsi, Symeonidis, Kourtis and Markellos (2018) which has shown promising results by using complex models such as the Orthogonal Generalized Autoregressive Conditional Heteroskedasticity (O-GARCH) model and Vector Heterogeneous Autoregressive (VHAR) model. However, as the focus of our research lies in the stock returns prediction model, we adopt a naïve approach of forecasting covariance by using historical realized covariance as the forecast value.

Similar to the construction of our alpha portfolios, the portfolio will rebalance the day after the filing deadline of each quarter using the features generated from the quarterly filings, and its allocation remains constant until the next rebalancing. The expected returns of each individual asset, r_i, is the forecast output from our stock prediction model. Adopting the approach by Symitsi et al. (2018), the covariance matrix forecast is computed using a rolling window of 1000 observations of past realized asset returns.

The Alpha-Sigma portfolio considers the entire stock universe, and the weight allocated to each stock is determined by optimizing for the objective function specified above. We fix the coefficient of risk aversion, γ, to be 0.5 and we add a constraint for the maximum allocation to any individual asset to be 5% to prevent our portfolio from being overly-concentrated on a few assets.

5.3.2 Comparison against Alpha Portfolio and Benchmark

We evaluate our Alpha-Sigma portfolio based on the test date range and compare it against the Alpha-Weighted Alpha portfolio and the SPY portfolio.

Test Date Range [2019-02-15 to 2022-08-15]

	Annualized Returns	Annualized Volatility	Return-Risk Ratio	Value-At-Risk 5%
SPY portfolio	15.06%	22.50%	0.6693	-21.84%

Alpha-Weighted Alpha portfolio	27.54%	31.95%	0.8619	-24.86%
Alpha-Sigma portfolio	29.99%	29.82%	1.0058	-18.91%

As compared to our Alpha portfolio, our Alpha-Sigma portfolio achieved a higher annualized return with a lower annualized volatility. This results in a significant increase in the return-risk ratio. By considering portfolio variance in our portfolio construction process, we were able to improve the performance of our strategy yet reduce its risk at the same time.

Furthermore, the value-at-risk of our Alpha-Sigma portfolio is even lower than that of the SPY portfolio, indicating that the Alpha-Sigma portfolio has a lower downside risk than that of our benchmark market index. Yet, the Alpha-Sigma portfolio is able to achieve almost twice the returns performance than the index. This clearly shows the effectiveness of our systematic investment strategy as compared to a simple buy-and-hold strategy.

5.3.3 Parameter Adjustment

A key input to the portfolio optimization function of our Alpha-Sigma portfolio is the coefficient of risk aversion, γ. While our study has fixed γ to be 0.5, we wish to further explore varying the values of γ and observing its effects on the performance of the Alpha-Sigma portfolio.

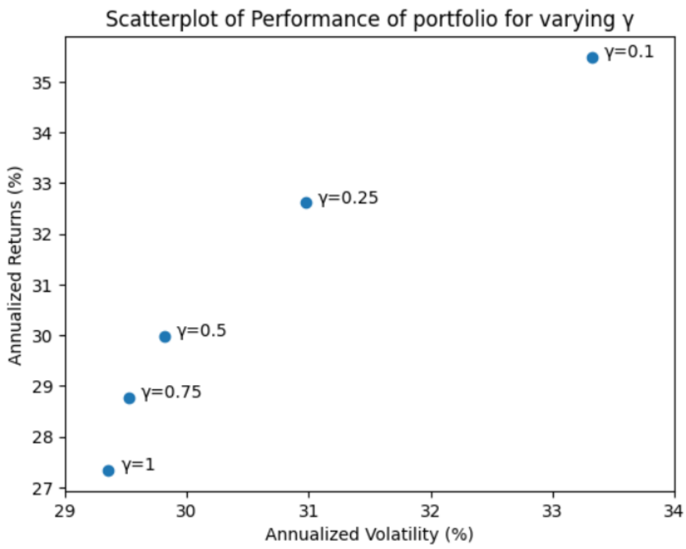


Figure 18: Scatterplot of Annualized Returns & Annualized Volatility for Alpha-Sigma portfolios of varying γ

As we increase γ, our objective function will place greater emphasis on reducing portfolio variance and less emphasis on increasing portfolio returns. Hence, the portfolio will see a decrease in annualized volatility with a corresponding decrease in annualized returns.

The optimal value for γ ultimately depends on the risk appetite of the individual investor. Investors willing on to take on more risk should select a low value of γ while risk-averse investors should select a high value of γ.

5.4 Final Portfolio

In the subsequent parts of our study, our discussions will focus solely on the implementation of the Alpha-Sigma portfolio with γ=0.5 as it has shown superior performance when compared to portfolios generated using other methodologies and reflects a balanced risk appetite.

5.4.1 Time Series Performance Analysis

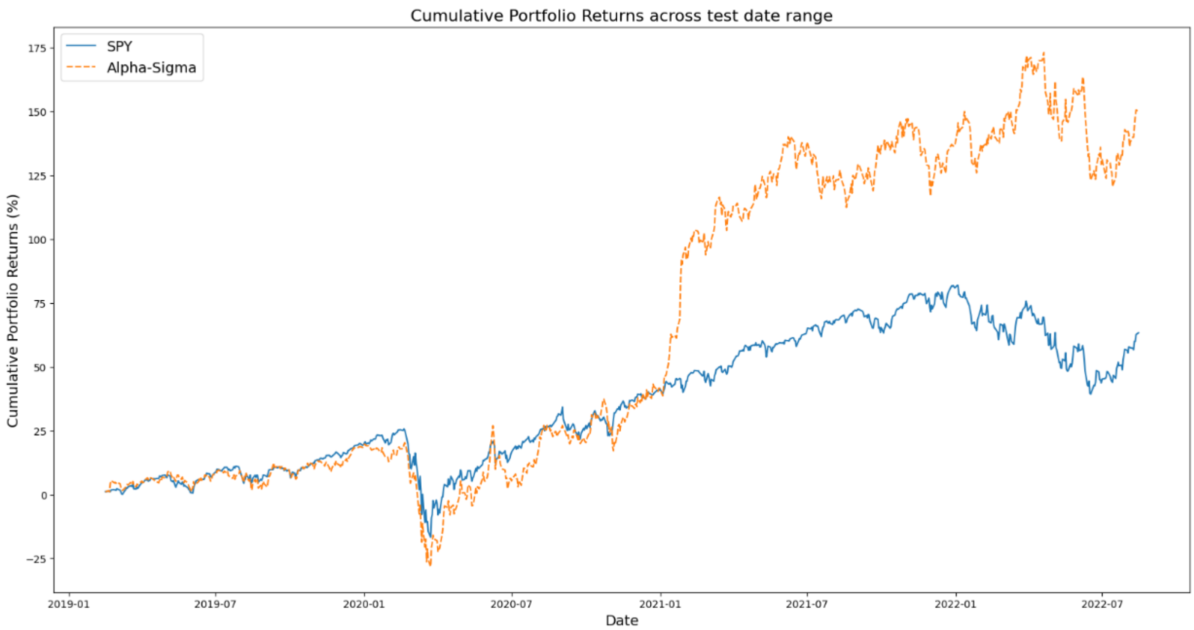


Figure 19: Cumulative Portfolio Returns Time Series across Test Date Range

We further categorize our test date range into three different economic regimes and evaluate the performance of the portfolios across the different regimes.

From February 2019 to February 2020, our alpha portfolio returned approximately 13% while SPY, the US equity benchmark, returned approximately 20%. During this time period, the U.S. stock market experienced low volatility and rose at a slow and steady pace.

During the Covid Crash which happened between February 2020 to April 2020, our alpha portfolio also performed worse than the SPY portfolio. We can categorize this event as a black swan event, as the entire financial market did not anticipate this crash. Hence, we would not expect our model to be able to anticipate it, as it merely follows the sentiments of larger institutional investors. Furthermore, the time lag in the filing also prevents our portfolio from reacting quickly enough to the negative news.

From April 2020 onwards, U.S. stocks rebounded and entered a bull market, as we see a recovery of the SPY prices past its previous peak. During this period, our alpha portfolio significantly outperformed the SPY portfolio. Within a few months, the alpha portfolio was able to recoup its losses and caught up to the performance of the SPY portfolio, eventually surpassing it by a large margin.

While we do not have a sufficiently long period of test data to confidently conclude the performance of our alpha portfolio in different economic regimes, our findings suggest that our portfolio could outperform the S&P 500 during periods of higher volatility but underperform during periods of low volatility and black swan events.

5.4.2 Past Literature Comparison

In addition to comparing our portfolio's performance against a major U.S. market index as benchmark, we wish to also assess it against models from previous literature which utilize 13F institutional holdings data as input.

	Date Range [2013-12-31 to 2017-12-31]		
	Annualized Returns	Annualized Volatility	Return-Risk Ratio
Alpha-Sigma portfolio	27.35%	15.47%	1.7676
Fleiss et al. (2020) Portfolio	15.0%	15.0%	0.987

When we compare our portfolio performance against the performance of the strategy proposed by Fleiss et al. (2020) for the years 2014 to 2017, we found that our portfolio achieved much better annualized returns while maintaining a similar level of annualized volatility. Thus, the return-risk ratio of our portfolio was much higher than that of the portfolio constructed by Fleiss et al. (2020).

However, as the test date range of the paper differed from ours, the data from years 2014 to 2017 belong to our training period and was used in the training of our stock prediction model. Hence, we note that the comparison may be biased and not entirely fair.

Date Range [2019-01-01 to 2019-12-31]

	Annualized Returns	Annualized Volatility	Return-Risk Ratio
Alpha-Sigma portfolio	38.06%	16.66%	2.2849
Fleiss et al. (2021) Portfolio	21.6%	-	1.82

Likewise, when we compare our portfolio performance against the performance of the strategy proposed by Fleiss et al. (2021) for the year 2019, our portfolio also outperformed theirs with a higher return-risk ratio. Since the year 2019 falls within our test period, this is a fair comparison as the data was not used for training our model.

6 Conclusion

6.1 Summary

Our initial overarching aim of this research is to develop a robust trading strategy which outperforms a buy-and-hold strategy yet does not require resources unavailable to the average retail investor.

We believe that we have largely achieved this as our proposed strategy has outperformed the major market index while using only data sources that are widely accessible and cost-free. Furthermore, the implementation of the strategy is simple and straightforward, as the model only requires quarterly updates for portfolio allocation, making it user-friendly. Retail investors can simply execute the model quarterly and allocate their portfolio accordingly, should they choose to rely solely on this trading strategy.

For more sophisticated retail investors with other trading strategies, the output of our trading strategy can serve as an input to their overall trading strategy, potentially via a weighted-average of multiple different trading strategies. This serves to further diversify their portfolio, as their trading strategies are not reliant on any single data source or investment hypothesis. Additionally, investors can adjust the weights of the different trading based on their discretionary views of which types of strategies are best suited to perform in the existing economic conditions. We would recommend this approach as our study concluded that in certain economic conditions, our model could not outperform the market beta, and it would be better off to follow a buy-and-hold strategy or explore other trading strategies.

We also took various steps to ensure the robustness of the proposed trading strategy. By developing our own proprietary method to collect and clean the SEC 13F Filings data, we were able to collect a larger dataset as compared to those offered by external data vendors. This allowed us to train and evaluate our model with a longer time horizon spanning across different economic regimes and stress events. Consequently, we were able to observe specific market conditions which our strategy performs well in, and market conditions which our strategy fails to beat the market beta.

We placed emphasis on identifying and eliminating potential biases in the development of our trading strategy. In past literature, we identified sources of lookahead bias and survivorship bias which could lead to an upward bias of backtest results. Having overstated backtest results for a trading strategy is detrimental as it would often lead to losses when used in real-world trading. Hence, we came up with solutions to reduce these biases so that our backtest results would be more accurate and reflective of real-world trading performance. Although we reduced upward bias of our backtest results, we still saw better results as compared to past literature due to improvements made in the data filtering process as well as portfolio construction process.

6.2 Limitations

6.2.1 Data

Due to our research being limited to cost-free data sources, we note that there are several drawbacks caused by this limitation. While we were able to reduce survivorship bias when defining our stock universe, we were not able to completely eliminate it. We recognize that there is still a potential source of survivorship bias in our trading strategy. This is a result of Yahoo Finance as our data source for historical stock price information. Free-to-use market price data sources, such as Yahoo Finance, do not provide access to data of delisted tickers. As tickers are often delisted for reasons linked to poor performance, such as bankruptcy and acquisition, filtering away these tickers due to lack of data is likely to cause an upward bias to our backtest results (Gilbert and Strugnell, 2019).

Cohen et al. (2010) introduced the concept of using portfolio tilt instead of portfolio weights as the feature to determine a fund’s level of conviction in a particular stock. While the potential value added by this change remains unproven, the hypothesis behind the feature is plausible.

However, to calculate the portfolio tilt on a quarterly basis, we would require historical market capitalisation data for each stock during the period. Vendors offering access to this information charge a monthly or annual subscription fee for their services. Therefore, we were not able to explore further on the portfolio tilt feature due to limitations with our data sources.

6.2.2 Trading Implementation

In our research, we did not consider potential trading costs involved with the purchase or sale of stocks in the portfolio. Retail investors engage the services of stock brokerage platforms such as Interactive Brokers and TDAmeritrade in order to trade. Some brokers charge a trading fee, often as a percentage of the value of the trade, while others offer commission-free trading. Hence, investors who implement our strategy using brokers that charge trading fees would likely see that the strategy's real world performance would be lower than that of the backtest.

6.2.3 Market Unpredictability

According to Toth and Jones (n.d.), extreme events occur more often than one would expect if stock returns followed a normal distribution. These events, commonly referred to as Black Swan events, are unexpected and difficult to predict under normal circumstances. An example of such an event would be the Covid Crash in February 2020, an event within our test period, which came as a surprise to the entire market. We concluded that our strategy would not perform well in the occurrence of such events as it relies on following the moves of the larger financial institutions. If these institutions were unable to anticipate such events, our strategy would similarly not be able to do so. While they are able to react swiftly to such events and make the necessary portfolio adjustments, our strategy is limited by the time lag of the filings and would not capture the changes promptly. This limits our strategy from performing well if it were fully systematized, as discretionary adjustments must be made during such unforeseeable circumstances.

6.3 Recommendation for Future Work

Recognizing the limitations of our current work, we note that future research on this topic could help address them. As time passes, there will be a larger amount of institutional holdings data available for training and testing, allowing for further improvement to the robustness of the strategy. Furthermore, we were unable to obtain historical prices of delisted tickers as well as historical market capitalisation data due to the costs involved, which limited the scope of our research. Future research could look into these other data sources and explore whether the addition of such data has any significant impact to the performance of the strategy.

For retail traders who do not have access to commission-free stock trading, our strategy would not be optimal as it does not consider trading costs in the portfolio construction process. This means that our strategy would rebalance the portfolio to achieve higher returns, even though the rebalancing may result in a trading cost exceeding the potential higher returns. Future research can extend the objective function of the portfolio optimization process to consider trading costs involved. This would result in the reduction of the portfolio turnover as trading costs increases.

We also identified two possible extensions to the research. Firstly, although information coefficient is used in the evaluation of our stock prediction model, it is not being considered during the training phase of the model. The objective function of our current models aim to solely minimize squared error, and the resultant model performance could vary when evaluated with multiple evaluation metrics. Becker, Fox and Fei (2008) also recognized this limitation when using "simple" objective functions in a stock prediction model. They proposed using multi-objective algorithms to simultaneously optimize for multiple performance metrics and found that they produced a more well-balanced model performance. We believe that by defining an objective function which optimizes for maximizing information coefficient while minimizing squared error, we can achieve a higher information coefficient while keeping a relatively low squared error.

As the focus of our study is on the stock prediction model, we spent little effort in implementing a complex and accurate methodology for covariance forecasting in the portfolio construction step. However, there is much room for improvements in this aspect, as there have been extensive research into the problem of covariance forecasting. Models such as Vector Heterogeneous Autoregressive (VHAR) model and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model have shown stellar results in predicting the covariance matrix of a basket of stocks (Symeonidis et at, 2018). By implementing complex models which offer a more accurate covariance prediction, this is likely to reduce the volatility of the trading strategy while maintaining its returns, further increasing the risk-adjusted returns of the strategy.

References

Hayes, A. (2022, February 15). SEC form 13F: What it is, filing requirements, and key issues. Investopedia. Retrieved from https://www.investopedia.com/terms/f/form-13f.asp.

DiPietro, D. M. (2019). Alpha Cloning: Using Quantitative Techniques and SEC 13f Data for Equity Portfolio Optimization and Generation. The Journal of Financial Data Science, Vol. 1, Issue 4, pp. 159–171. https://doi.org/10.3905/jfds.2019.1.008.

Miori, D. and Cucuringu, M. (2022, September 20). SEC FORM 13F-HR: Statistical Investigation of Trading Imbalances and Profitability Analysis.

U.S. Securities and Exchange Commission (2021, January 25). How to Read a 10-K/10-Q. Retrieved from https://www.sec.gov/oiea/investor-alerts-and-bulletins/how-read-10-k10-q.

Wikipedia (2022, August 24). Efficient Frontier. Retrieved from https://en.wikipedia.org/wiki/Efficient_frontier.

Murphy, E. M. (2013, May 14). Adoption of Updated EDGAR Filer Manual. Retrieved from https://www.federalregister.gov/documents/2013/05/21/2013-11921/adoption-of-updated-edgar-filer-manual.

Gilbert, E. and Strugnell, D. (2019, August). Does Survivorship Bias really matter? An Empirical Investigation into its Effects on the Mean Reversion of Share Returns on the JSE Securities Exchange (1984-2006).

Christoffersen, S., Danesh, E. and Musto D. (2015, August 15). Why Do Institutions Delay Reporting Their Shareholdings? Evidence from Form 13F.

Cohen, R., Polk, C. and Silli, B. (2010). Best Ideas.

Angelini, L., Iqbal, M. and Jivrag, F. (2019, October). Systematic 13F Hedge Fund Alpha.

Fleiss, A., Cui, H., Stoikov, S. and DiPietro, D. M. (2020, Winter). Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning. Journal of Financial Data Science, Vol. 2, Issue 1, pp. 45–60. https://doi.org/10.3905/jfds.2019.1.022.

Fleiss, A., Kumaar, A., Rida, A., Shin, J., Lai, X., Fang, V., Chen, J. and Li, A. (2021, November 8). Deep Reinforcement Learning & Feature Extraction For Constructing Alpha Generating Equity Portfolios. http://dx.doi.org/10.2139/ssrn.3958478.

Markowitz, H.M. (1952). Portfolio Selection. Journal of Finance, Vol. 7(1), pp. 77-79.

Jones, M. (2006, September 7). Examination of fund age and size and its impact on hedge fund performance.

Chen, J. (2022, July 20). Fund Flow: Definition, Example, and How To Interpret. Retrieved from https://www.investopedia.com/terms/f/fund-flow.asp.

Koker, T. (2019, May 12). Creating a Survivorship Bias-Free S&P 500 Dataset with Python.

Poterba, J. and Summers, L. (1988, March). Mean reversion in stock prices, evidence and implications. Journal of Financial Economics 22, pp. 28-59.

Ou, P. and Wang, H. (2009, December). Prediction of Stock Market Index Movement by Ten Data Mining Techniques. Journal of Modern Applied Science Vol. 3, No.12.

Zhang, F., Guo, R. and Cao, H. (2020, October). Information Coefficient as a Performance Measure of Stock Selection Models.

Toth, D. and Jones, B. (n.d.). Against the Norm: Modeling Daily Stock Returns with the Laplace Distribution.

Dai, Z. and Chang, X. (2021, June 21). Predicting Stock Return with Economic Constraint: Can Interquartile Range Truncate the Outliers?.

Saha, S., Gao, J. and Gerlach, R. (2021, June 21). Stock Ranking Prediction Using List-Wise Approach and Node Embedding Technique.

McKenna, D. and Gleiser, L. (2010, May). Converting Scores into Alphas.

Symitsi, E., Symeonidis, L., Kourtis, A. and Markellos, R. (2018, September 3). Covariance forecasting in equity markets. Journal of Banking and Finance 96, pp. 153-168.

Becker, Y., Fox, H. and Fei, P. (2008). An Empirical Study of Multi-Objective Algorithms for Stock Ranking.

Appendix A – EDA Visualizations

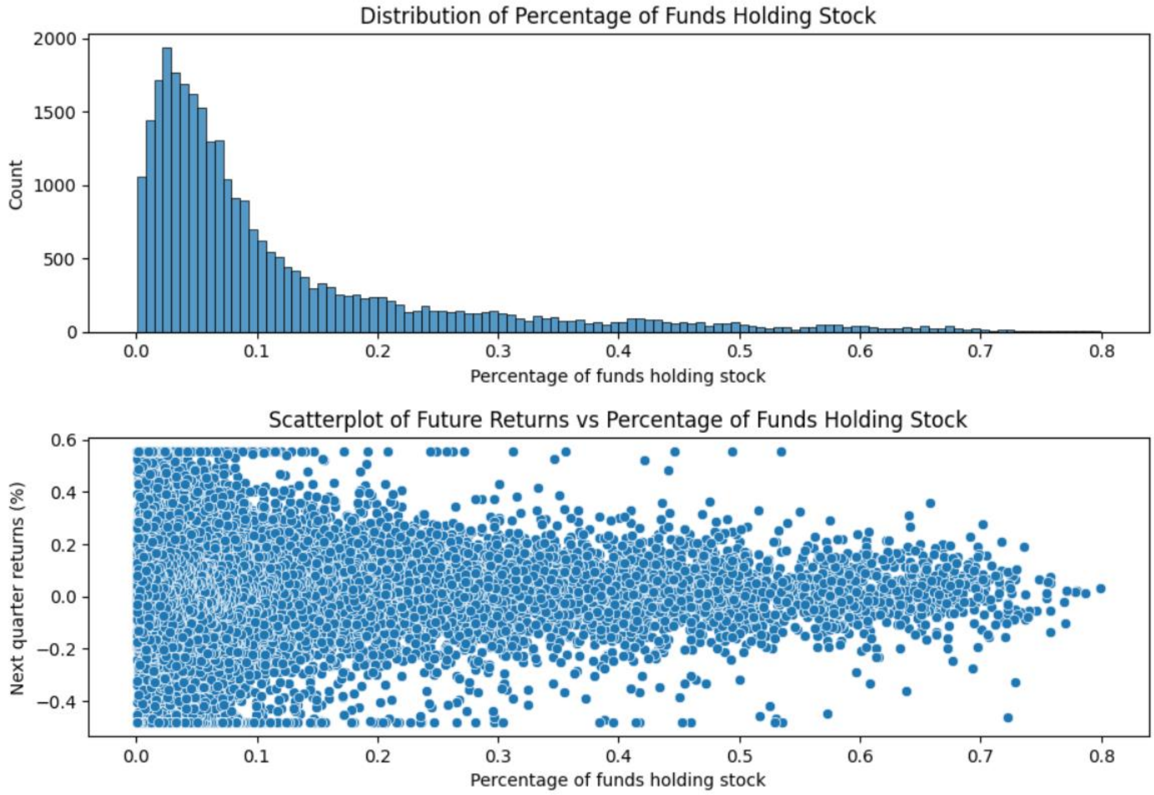


Figure 20: EDA of feature x_1

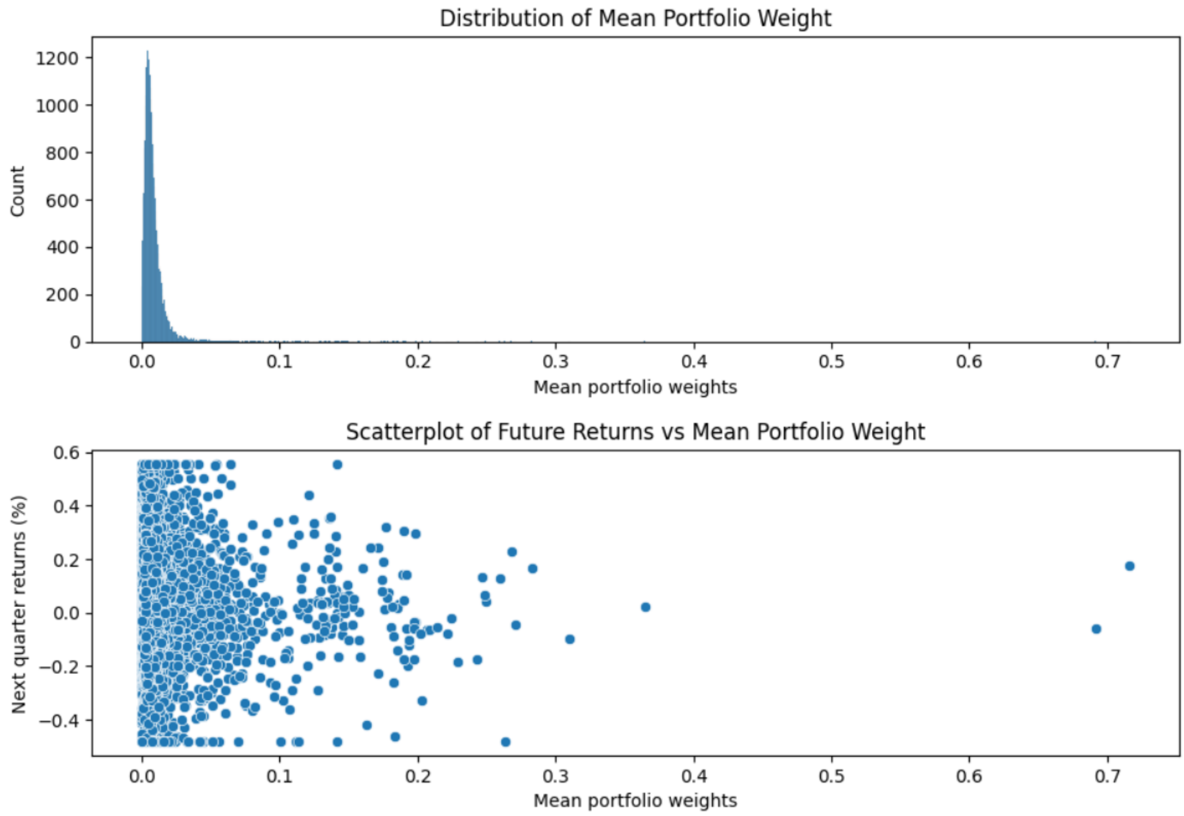


Figure 21: EDA of feature x_2

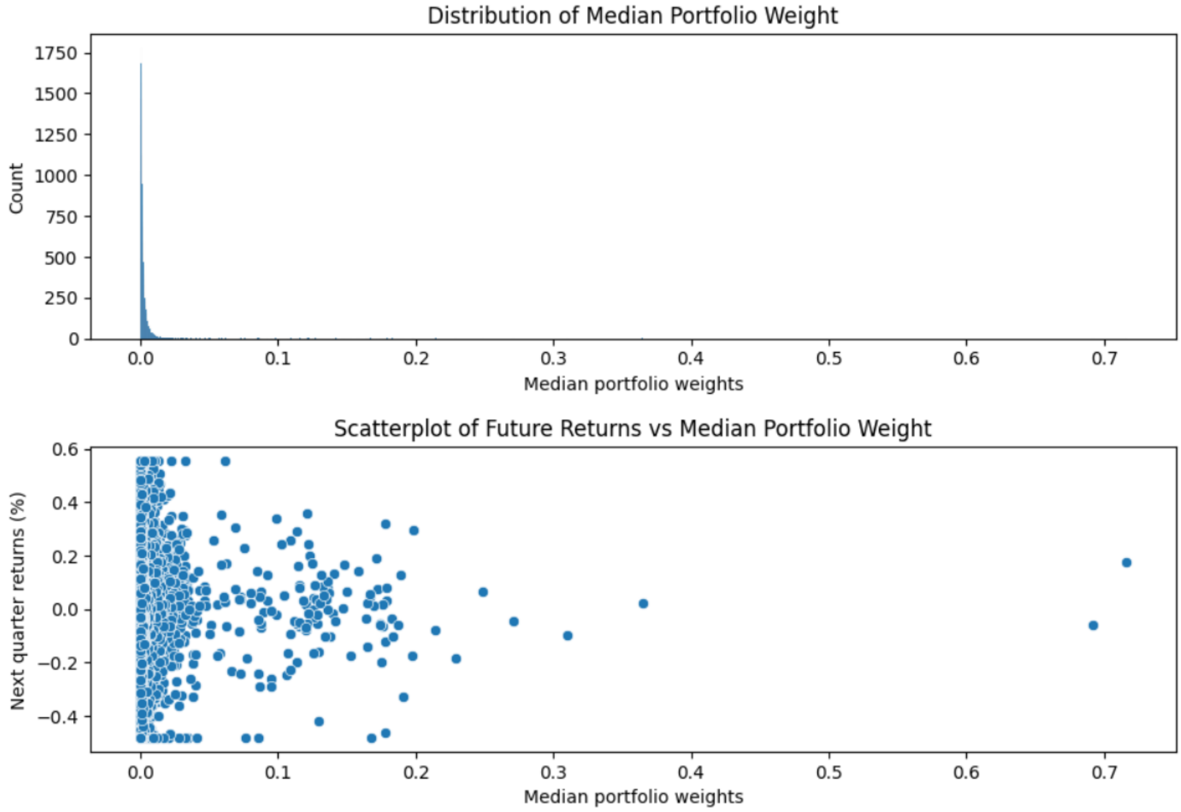


Figure 22: EDA of feature x_3

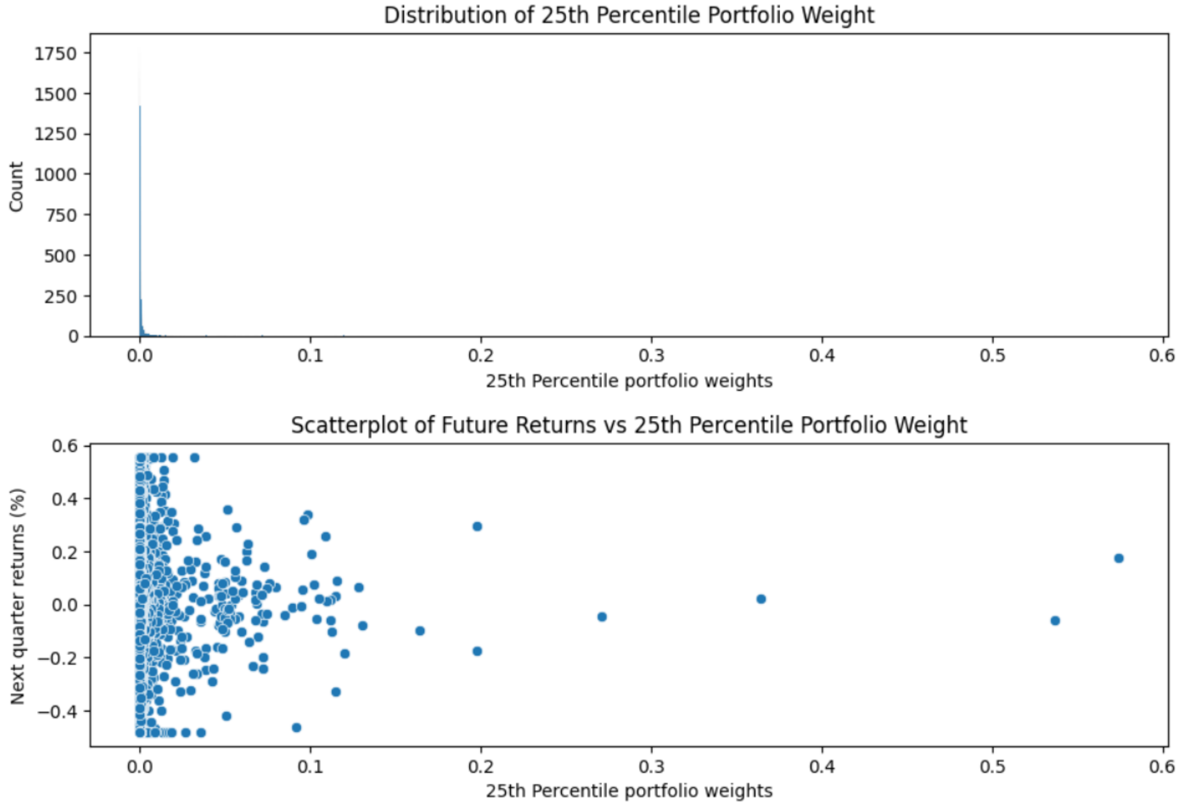


Figure 23: EDA of feature x_4

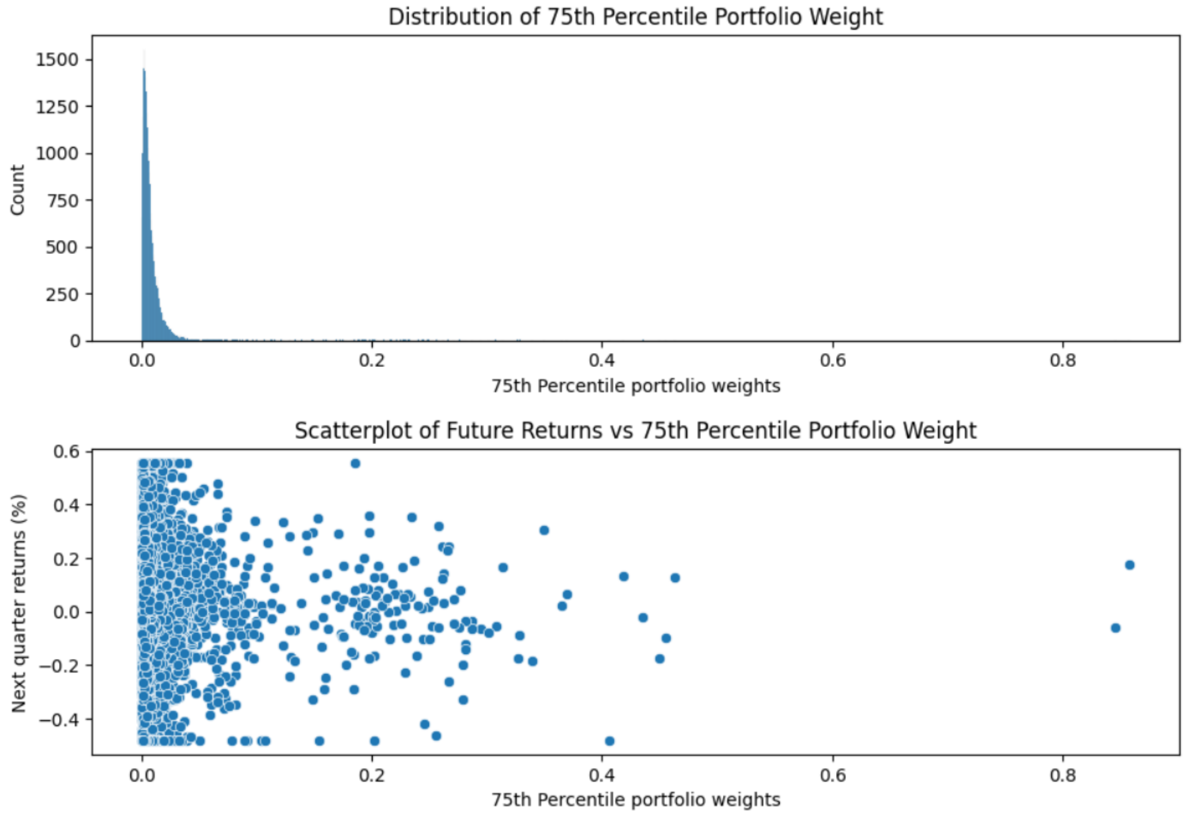


Figure 24: EDA of feature x_5

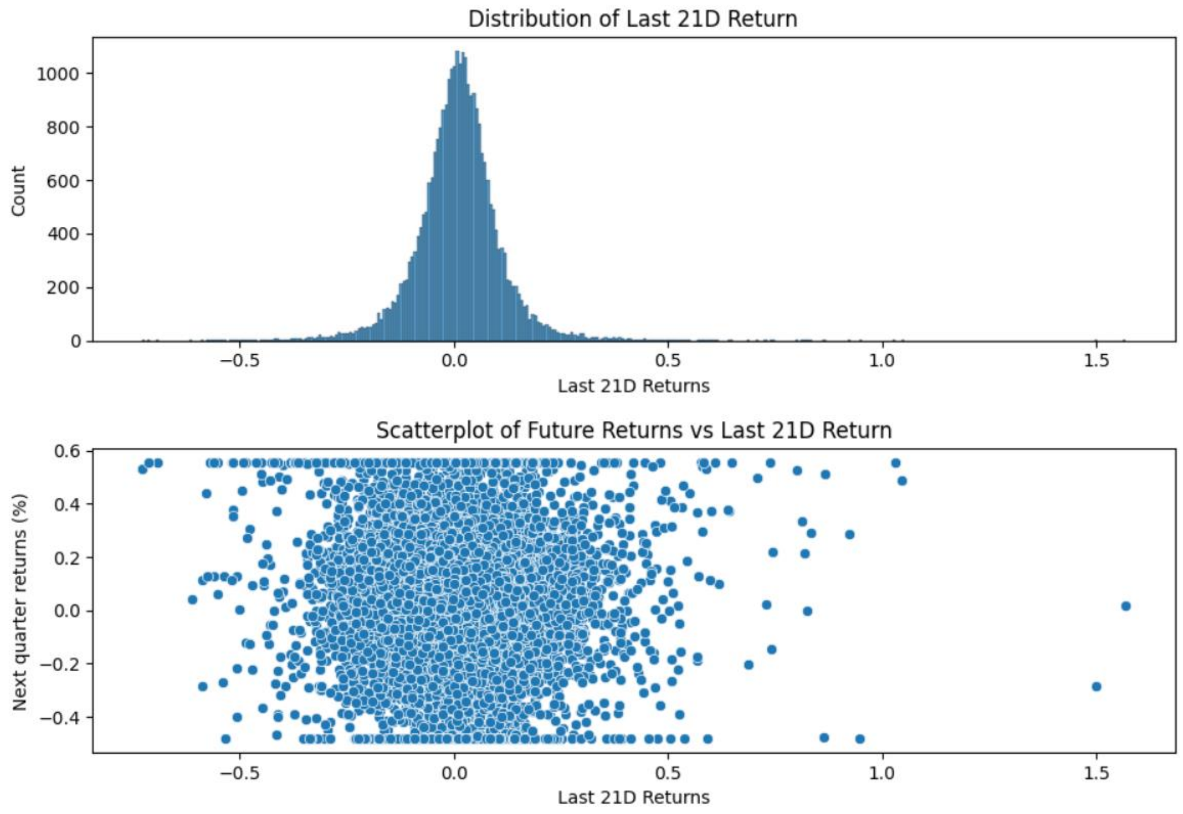


Figure 25: EDA of feature x_6

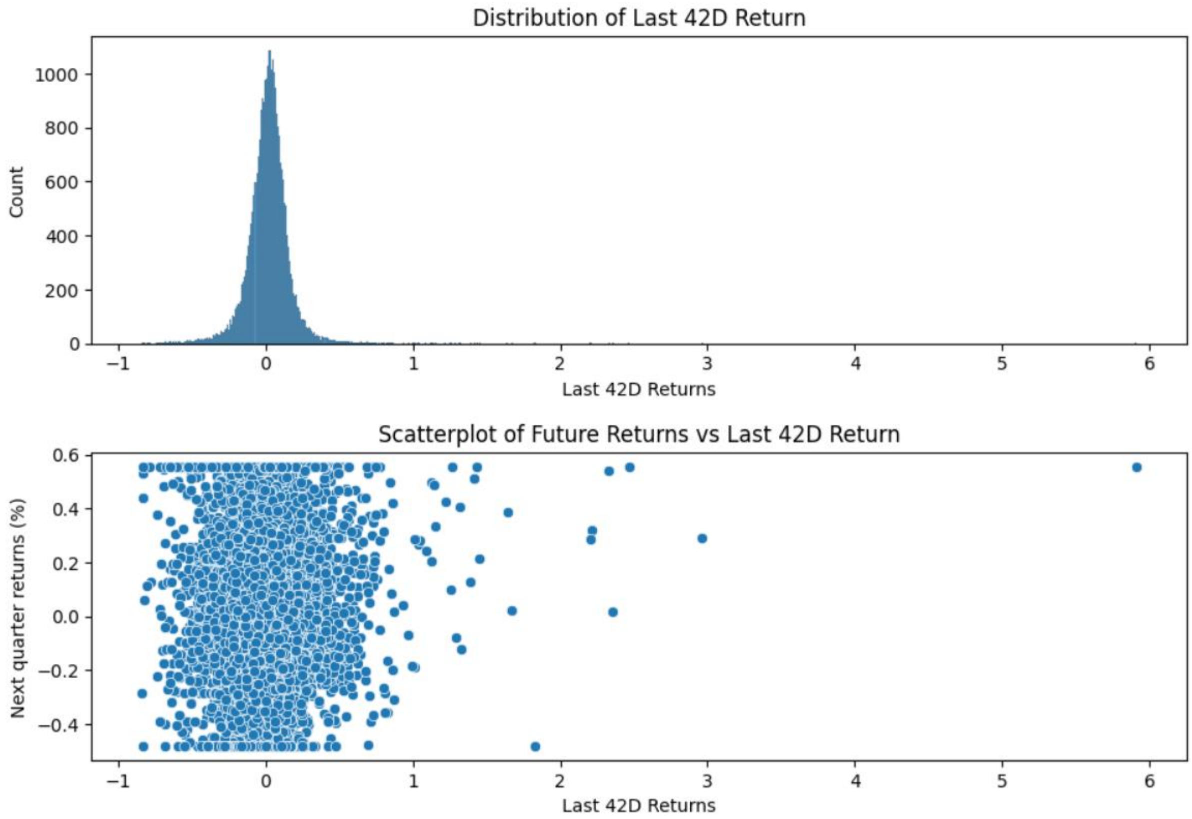


Figure 26: EDA of feature x_7

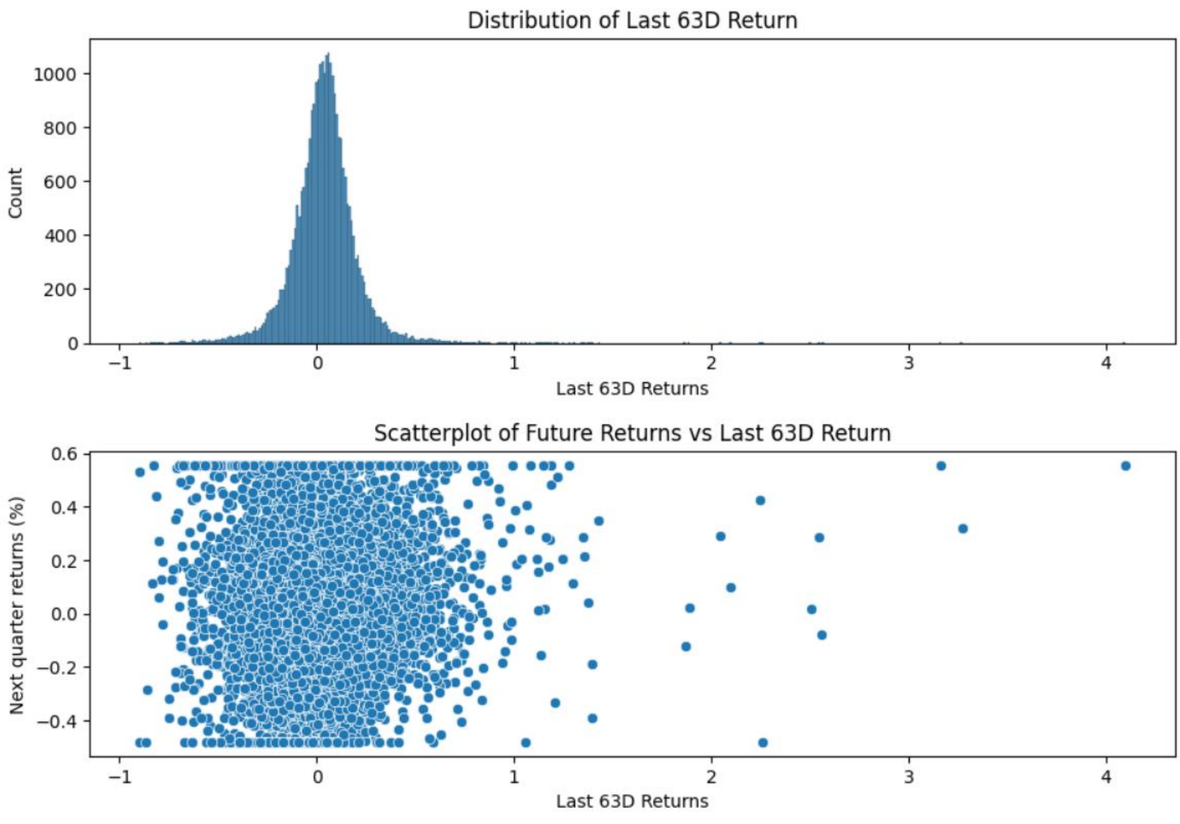


Figure 27: EDA of feature x_8

Appendix B – Sub-fund Feature Analysis

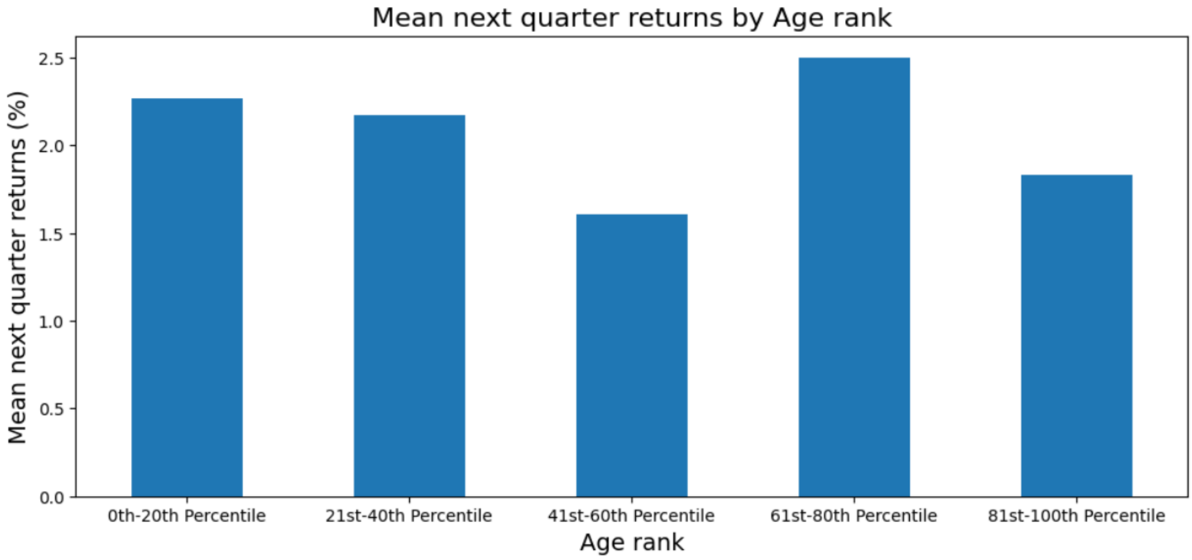


Figure 28: Relationship between sub-fund's age and future performance

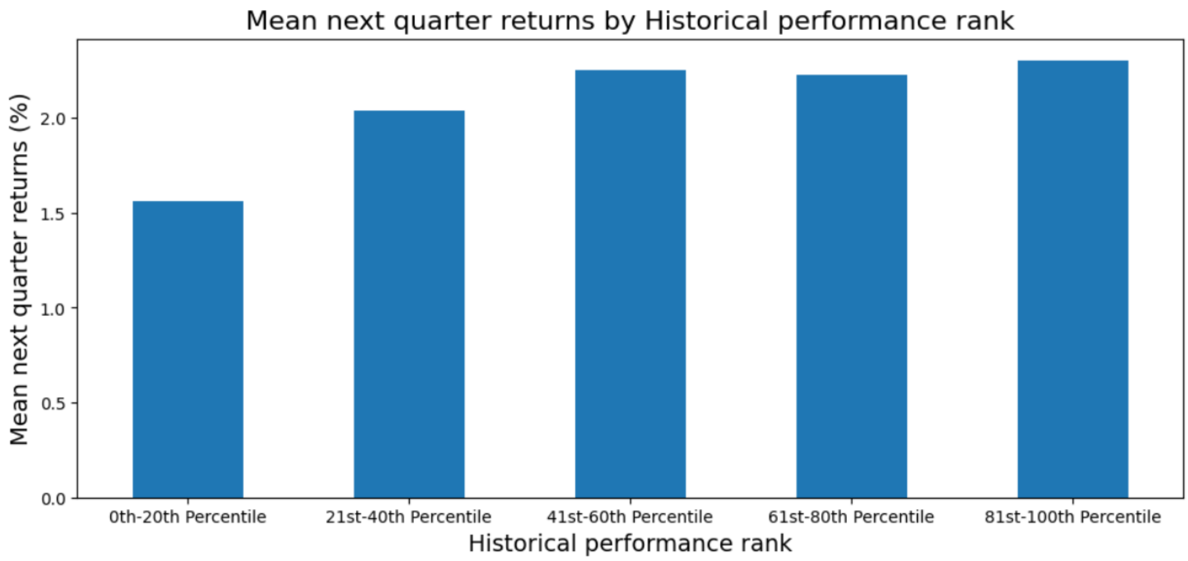


Figure 29: Relationship between sub-fund's historical performance and future performance

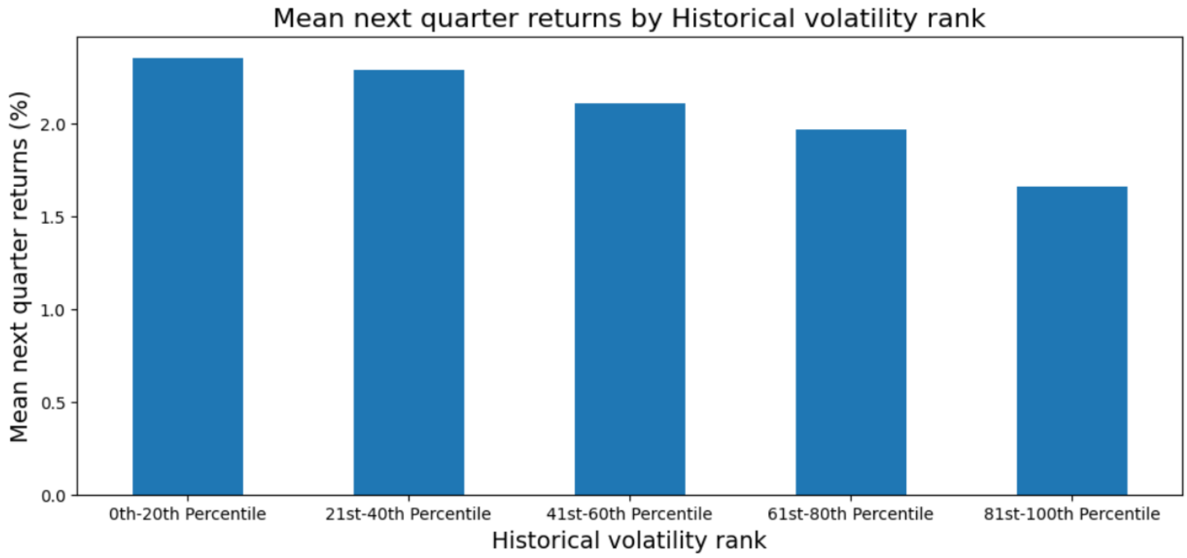


Figure 30: Relationship between sub-fund's historical volatility and future performance

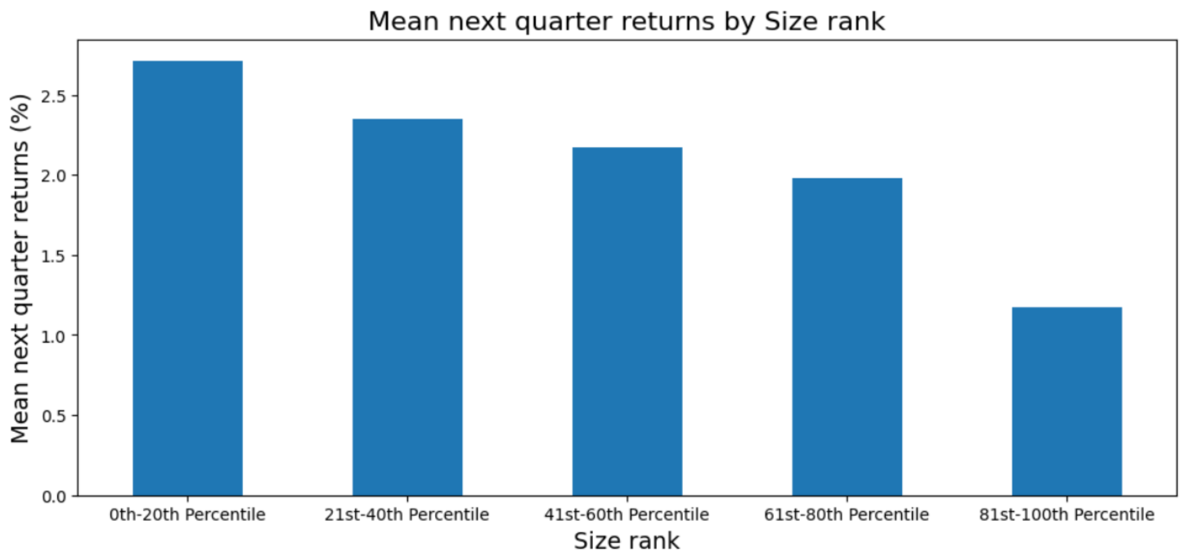


Figure 31: Relationship between sub-fund's size and future performance

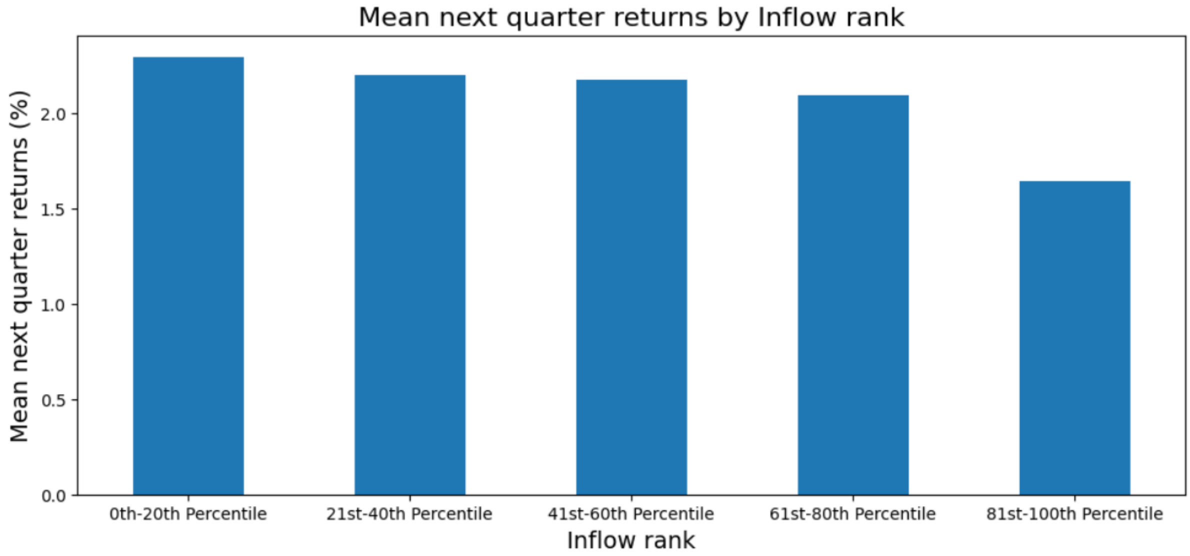


Figure 32: Relationship between sub-fund’s inflow and future performance

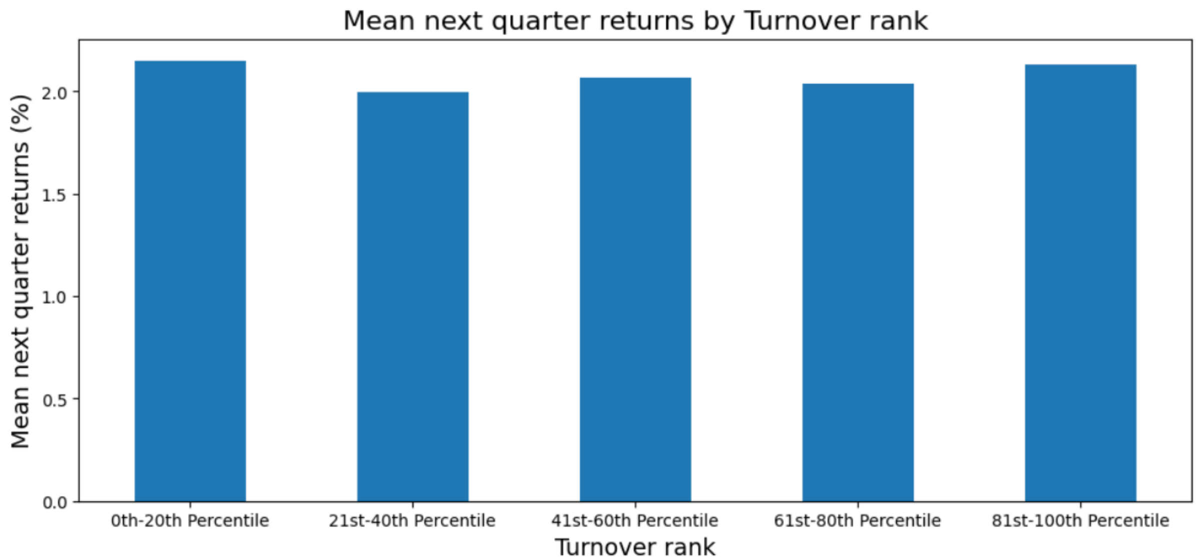


Figure 33: Relationship between sub-fund’s turnover and future performance