**Introduction to Graph Alpha**

Graph Alpha is a quantitative trading strategy predicated on the hypothesis that financial assets (tickers) exhibiting strong correlations over a defined historical period (e.g., one year) will continue to demonstrate stronger correlations over a subsequent shorter period (e.g., one month). This strategy leverages the power of graph theory and statistical correlations to identify and exploit these relationships, generating potential alpha for investors.

**Concept and Rationale**

Graph Theory and Financial Markets:

Graph theory provides a robust framework for analyzing relationships among financial assets. In the context of Graph Alpha, each stock or financial instrument is represented as a node, and the correlation between any two stocks is represented as an edge connecting these nodes. The strength of the correlation determines the weight of the edge. By constructing such a graph, we can visualize and analyze the intricate web of relationships within a portfolio or the broader market.

Correlation Persistence:

The fundamental assumption of Graph Alpha is that correlations among certain stocks are persistent over time. Stocks that are strongly correlated over a longer historical period (e.g., one year) are likely to remain correlated over a shorter horizon (e.g., one month). This persistence can be attributed to underlying economic linkages, sectoral ties, or other market dynamics that do not change rapidly.

**Implementation of Graph Alpha**

Data Collection and Preprocessing:

1. Collect historical price data for a set of stocks over the past year.
2. Calculate daily returns or other relevant metrics (e.g., log returns, volatility-adjusted returns).Those metrics can be expanded to other metrics such as NDR factor, total turnover ratio and so on.
3. Constructing the Correlation Matrix: Compute the correlation matrix for the selected stocks based on their daily returns over the one-year period. The correlation matrix serves as the adjacency matrix for the graph, where each element represents the correlation between two stocks.

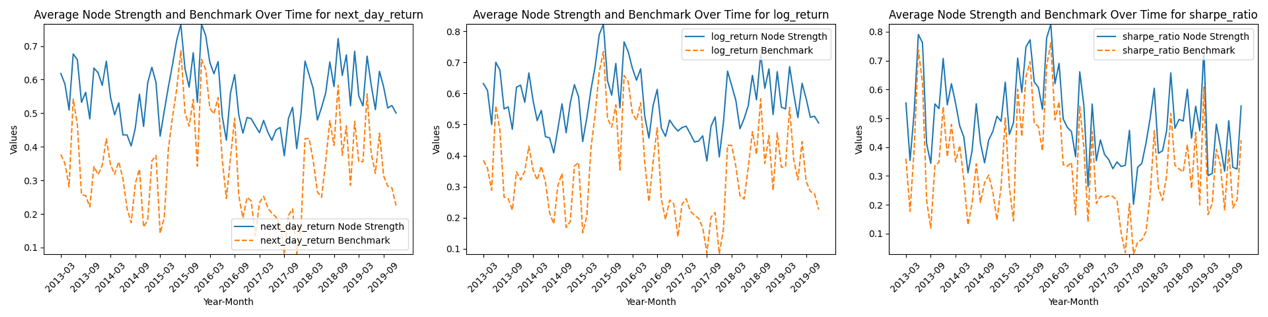
Graph Construction:

1. Create a graph where each node represents a stock, and each edge represents the correlation between two stocks.
2. Assign weights to the edges based on the strength of the correlations.

Identifying Strong Correlations:

1. Focus on the top 1% (or another percentile threshold) of correlations to identify the most strongly correlated pairs. These pairs are expected to exhibit strong correlations in the subsequent one-month period.

**Return Correlation Graph**



graph 1 NDR node strength vs benchmark (absolute return, log return, risk adjusted return)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Absolute return | Log return | Risk adjusted return |
| Benchmark | 0.3406 | 0.3500 | 0.3290 |
| Benchmark volatility | 0.1357 | 0.1411 | 0.168 |
| Average node strength | 0.5560 | 0.5685 | 0.4902 |
| Node strength volatility | 0.0918 | 0.0931 | 0.1400 |
| Excess Sharpe ratio | 3.5466 | 3.6223 | 1.5430 |

The excess Sharpe ratio is calculated as the difference between ratio of node strength and node volatility and ratio of benchmark and benchmark volatility

In the context of analyzing alpha strategies based on different return metrics—absolute return, log return, and risk-adjusted return (Sharpe ratio)—distinct patterns and insights emerge regarding the correlation and volatility of these returns over time.

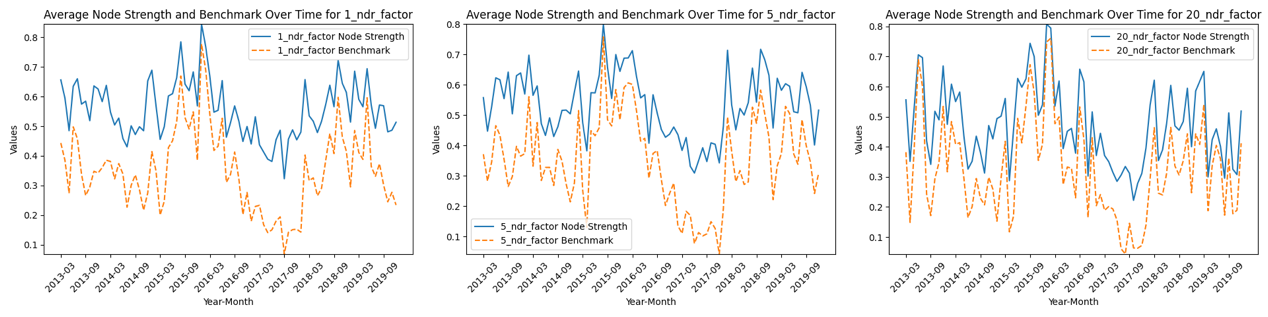
Absolute return focuses on the raw performance of the tickers without adjusting for volatility or logarithmic transformations. In the analysis, the absolute return metric shows a trend where, as the days shifted increase, the performance of excess returns diminishes. This is accompanied by an increase in the benchmark correlation and a decrease in node strength. These trends align with the understanding that future returns are more closely tied to recent financial data rather than data from further in the past.

Log return, calculated as the natural logarithm of the ratio of successive prices, provides a more normalized view of returns and is often preferred for its properties in continuous compounding. Among the three metrics analyzed, log return exhibits the greatest average node strength and the smallest volatility in node strength. This suggests that log return maintains stronger correlations over time and exhibits more stable relationships between the most correlated tickers.

The Sharpe ratio measures returns adjusted for risk, providing insights into the efficiency of the returns considering the associated volatility. When comparing the risk-adjusted return, we observe that the excess Sharpe ratio—calculated as the difference between the ratio of node strength to node volatility and the ratio of benchmark-to-benchmark volatility—decreases as the days shifted increase. This indicates an overall shrinking excess return performance, compounded by the decrease in node strength and the increase in volatility.

The analysis underscores that log return is a superior metric for evaluating the strength and stability of correlations between tickers. It provides a more normalized and consistent measure of returns, leading to higher average node strength, lower volatility in node strength, and the highest excess Sharpe ratio. This makes log return an effective metric for forecasting future returns based on historical correlations, especially when considering the diminishing predictive power of correlations as the prediction window extends.

**NDR Correlation Graph**

****

graph 2 NDR node strength vs benchmark (1dr,5dr, 20dr)

**Benchmark correlation and node strength comparison:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1dr | 5dr | 20dr |
| Benchmark | 0.3499 | 0.3511 | 0.3250 |
| Benchmark volatility | 0.1362 | 0.1447 | 0.1609 |
| Average node strength | 0.5566 | 0.5365 | 0.4742 |
| Node strength volatility | 0.0961 | 0.1065 | 0.1339 |
| Excess Sharpe ratio | 3.223 | 2.611 | 1.499 |

In the context of a graph-based alpha strategy, the underlying assumption is that tickers most closely related to a particular ticker over the past year will exhibit stronger correlations with that ticker within the next month. Analyzing this strategy reveals notable trends as the days shifted increase. Specifically, the performance of excess returns diminishes, characterized by an increase in the benchmark correlation and a decrease in node strength. These observations support the understanding that future returns are more closely linked to recent financial data than to data from further in the past.

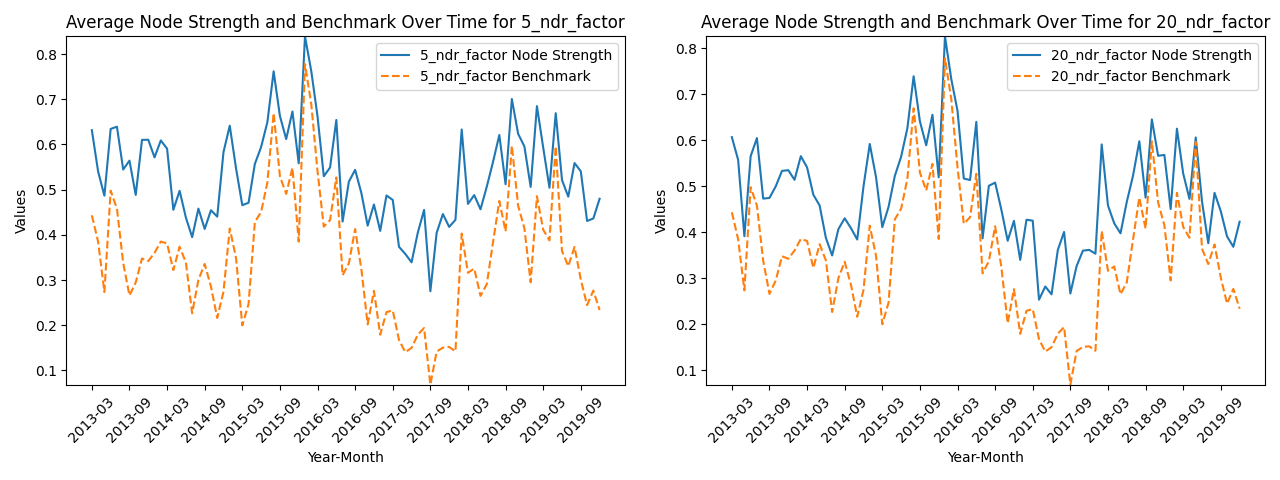
As the days shifted increase, the financial markets display a recency effect, where recent data more accurately reflects current market conditions compared to older data. This diminishing relevance of past correlations over time can be attributed to several factors. Financial markets are dynamic and continuously influenced by new economic indicators, geopolitical events, and company-specific developments. The constant influx of new information alters historical correlations, reducing their predictive power for future returns.

Moreover, extending the prediction window increases market volatility and noise, which can obscure signals derived from historical correlations. This makes it more challenging to identify meaningful patterns that can be exploited for excess returns. Additionally, many financial metrics exhibit mean-reverting behavior over time, meaning that extreme values observed in the past tend to revert to the mean, diminishing the strength of previously observed correlations.

The analysis also reveals that the benchmark correlation, representing the average correlation among all tickers, tends to rise with increased days shifted. This suggests a general market movement where overall market conditions become more synchronized, reducing the differentiation in excess returns. Conversely, the node strength, representing the average correlation of the most correlated tickers with a particular ticker, declines over time. This decline underscores that the strongest relationships observed in the past year lose their predictive power over time, emphasizing the importance of recent data for forecasting future returns.

Additionally, I introduced another metric called the excess Sharpe ratio. The excess Sharpe ratio is calculated as the difference between the ratio of node strength to node volatility and the ratio of benchmark volatility. Analysis of this metric shows that as the days shifted increase, the Sharpe ratio decreases, indicating an overall shrinking excess return performance in addition to the decrease in node strength. Furthermore, increasing the shift also raises volatility, which brings more uncertainty and risk. This comprehensive evaluation highlights the diminishing effectiveness of historical correlations for predicting future returns as the prediction window extends.

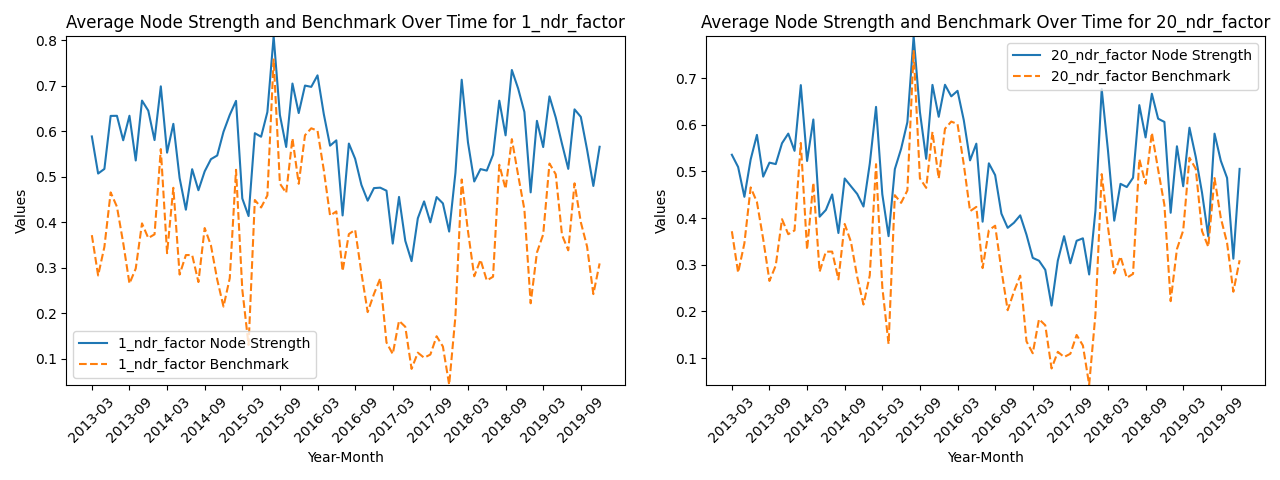
**Cross-metrics analysis:**



graph 3 5dr,20dr for data period 1dr for graph period NDR node strength vs benchmark

**Benchmark correlation and node strength comparison:**

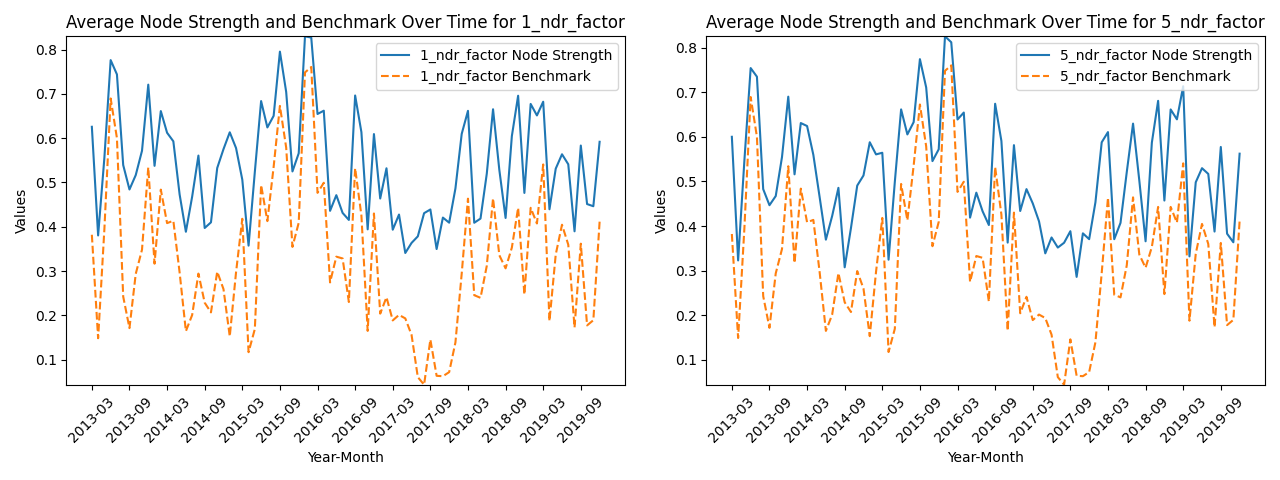
|  |  |  |
| --- | --- | --- |
|  | 5dr | 20dr |
| Benchmark | 0.3499 | 0.3499 |
| Benchmark volatility | 0.1362 | 0.1609 |
| Average node strength | 0.5314 | 0.4854 |
| Node strength volatility | 0.1362 | 0.1125 |
| Excess Sharpe ratio | 1.3325 | 2.1400 |



graph 4 1dr,20dr for data period 5dr for graph period NDR node strength vs benchmark

**Benchmark correlation and node strength comparison:**

|  |  |  |
| --- | --- | --- |
|  | 1dr | 5dr |
| Benchmark | 0.3252 | 0.3252 |
| Benchmark volatility | 0.1609 | 0.1609 |
| Average node strength | 0.5401 | 0.5154 |
| Node strength volatility | 0.1206 | 0.1289 |
| Excess Sharpe ratio | 2.4573 | 1.9960 |



graph 5 1dr,5dr for data period 20dr for graph period NDR node strength vs benchmark

**Benchmark correlation and node strength comparison:**

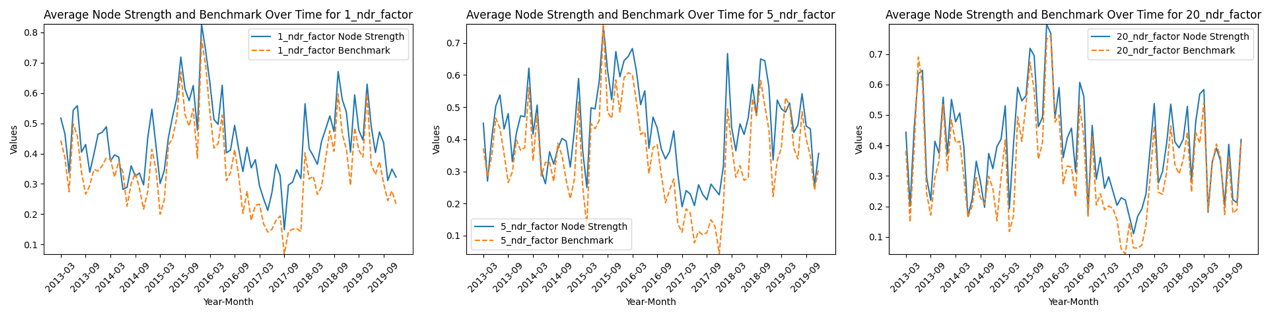
|  |  |  |
| --- | --- | --- |
|  | 1dr | 20dr |
| Benchmark | 0.3511 | 0.3511 |
| Benchmark volatility | 0.1447 | 0.1447 |
| Average node strength | 0.5585 | 0.4933 |
| Node strength volatility | 0.1005 | 0.1163 |
| Excess Sharpe ratio | 3.1308 | 1.815 |

The process do improve the performance of 20dr, while decrease the performance of 1dr and 5dr.

The cross NDR analysis highlights significant insights into the effectiveness of different data periods and graph periods on the performance of graph alpha. Specifically, using a 20-day rolling period (20DR) as the graph period tends to decrease both the node strength and the excess Sharpe ratio of graph alpha when the data period is set to 1-day or 5-day rolling periods (1DR or 5DR). This suggests that a longer graph period may dilute the immediate impact and predictive power of the most recent data, thus weakening the strength of the correlations and diminishing the returns adjusted for risk.

Conversely, when shorter rolling periods such as 1DR or 5DR are used as the graph period, the performance of the alpha significantly improves compared to using 20DR for both the data period and the graph period. This indicates that shorter rolling periods capture more relevant and recent financial data, leading to stronger and more predictive correlations. Consequently, the alpha derived from these correlations exhibits better performance, evidenced by higher node strength and excess Sharpe ratio.

**Excluding same industry analysis:**

****

graph 6 node strength after excluding same industry(1dr,5dr,20dr)

**1，5，20 9图**

The operation of excluding tickers from the same industry from the list of most correlated stocks significantly reduces the node strength. This indicates that stocks within the same industry tend to exhibit higher correlations with each other compared to those across different industries. This intra-industry correlation is likely due to shared economic factors, industry-specific news, and similar market conditions affecting companies within the same sector similarly.

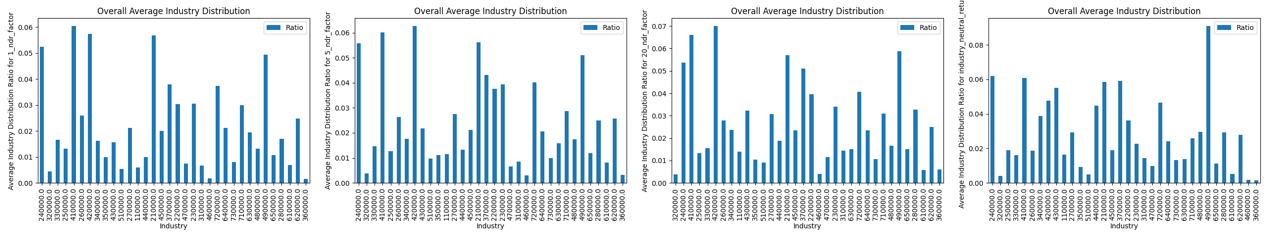
After implementing this exclusion, the node strength—which measures the average correlation of a ticker with its most correlated tickers—often struggles to outperform the benchmark. In some instances, the benchmark even exceeds the node strength. This phenomenon underscores the critical role of industry effects in driving stock correlations. By removing these industry effects, the remaining correlations are typically weaker, highlighting the less pronounced relationship between stocks from different sectors. Consequently, this adjustment demonstrates the challenge of identifying strong, cross-industry correlations that can consistently outperform the broader market benchmark.

**Average industry exposure:**

To investigate the industry exposure of the nodes in the graph, we conduct the following research:

1. Count the number of tickers from top correlated list that are from the same industry as the original ticker, and calculate the average ratio of that number with the total number of top correlated tickers
2. Calculate the spread of industries within the most related list

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NDR | 1dr | 5dr | 20dr | Indu\_neu return |
| Average ratio | 0.5435 | 0.5175 | 0.4439 | 0.2322 |



Graph 7 industry exposure(1dr, 5dr,20dr，indu\_neutral)

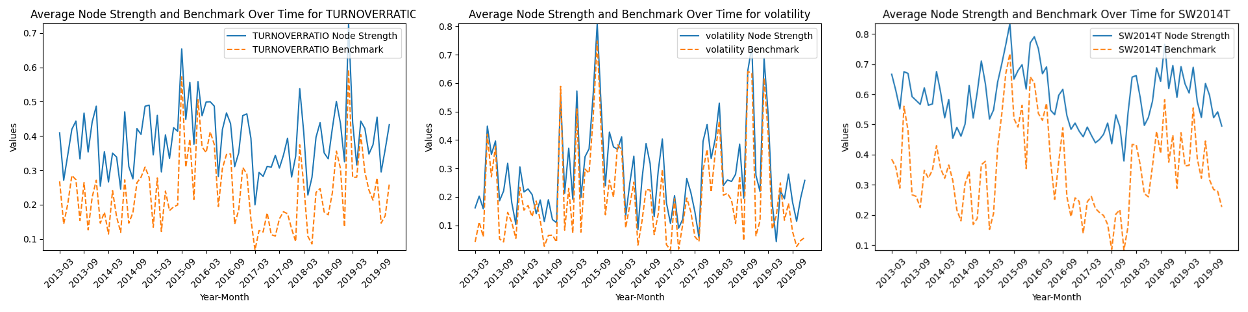
The results of the analysis reveal significant insights into the behavior of tickers within the same industry and their impact on graph alpha. The ratio indicates that nearly half of the most correlated tickers come from the same industry. This finding reinforces the idea that tickers within the same industry tend to behave similarly, showcasing a strong intra-industry correlation.

Additionally, the second graph, which illustrates the industry exposure, suggests that the graph alpha is unevenly exposed to different industries. This uneven exposure could potentially introduce unwanted risks from specific industries. Such industry-specific risks can arise due to various factors like regulatory changes, economic conditions, and sector-specific events, which might not be diversified away if the portfolio is heavily concentrated in a few industries.

In summary, while the high intra-industry correlation provides a strong basis for predicting future returns within the same industry, the uneven industry exposure highlights the need for a balanced approach to mitigate industry-specific risks. These findings emphasize the importance of considering both intra-industry correlations and overall industry exposure when constructing a robust graph-based alpha strategy.

**Other Metrics Correlation Graph**

graph 8 turnover ratio and volatility graph node strength

**Benchmark correlation and node strength comparison: **

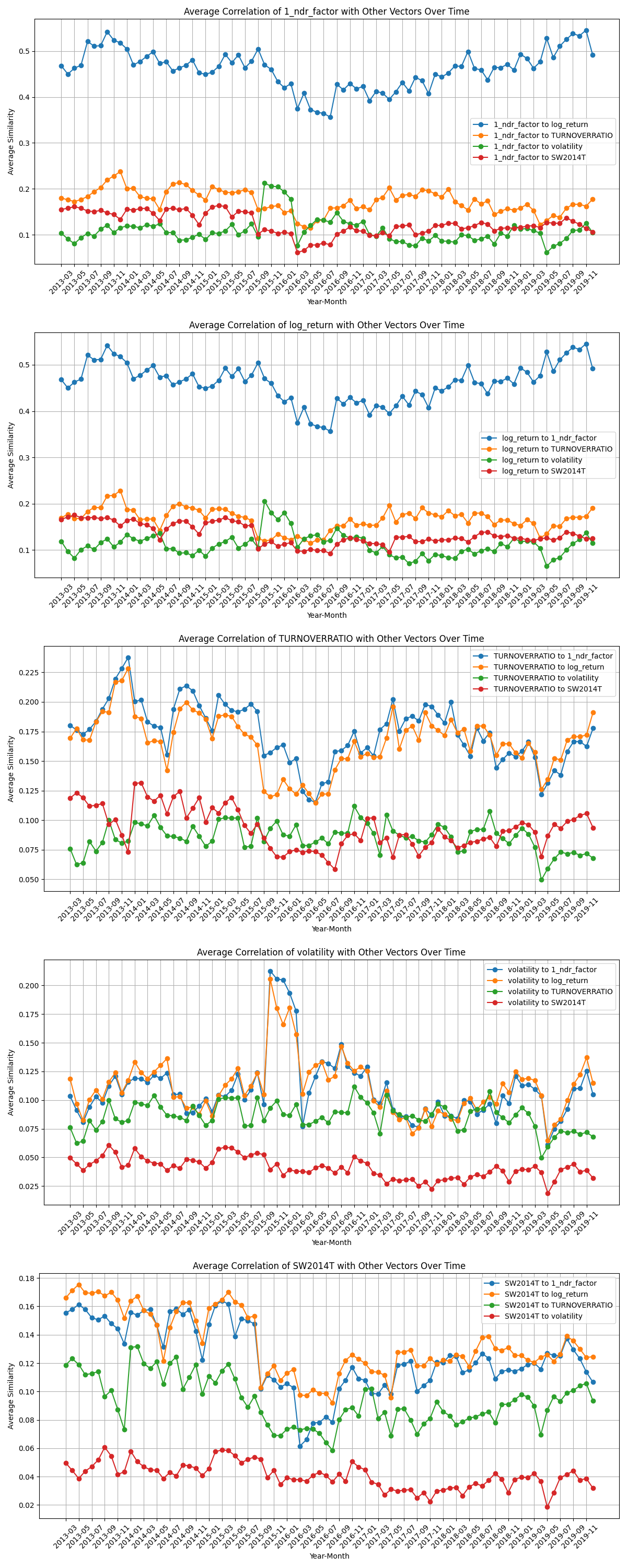
|  |  |  |  |
| --- | --- | --- | --- |
|  | Turnover ratio | Volatility | Industry(by SW2014T) |
| Benchmark | 0.2355 | 0.2025 | 0.3500 |
| Benchmark volatility | 0.1052 | 0.1680 | 0.1411 |
| Average node strength | 0.3916 | 0.2862 | 0.5895 |
| Node strength volatility | 0.0931 | 0.1589 | 0.094 |
| Excess Sharpe ratio | 1.9676 | 0.5923 | 3.7712 |

In the context of turnover ratio and volatility, the analysis reveals that both the benchmark and node volatility metrics exhibit patterns similar to those observed with returns. However, a notable distinction arises with the node strength for turnover ratio and volatility, which is significantly lower compared to returns. This indicates that while turnover ratio and volatility are important metrics, their correlations with other tickers are weaker.

Moreover, the excess Sharpe ratio for turnover ratio and volatility also tends to be lower. This suggests a lack of strong correlation between these metrics and price changes. The impact of turnover ratio and volatility on price movement is likely more complex and less explicit compared to direct return metrics. Turnover ratio may be influenced by various trading behaviors and market liquidity factors, while volatility can be affected by broader market conditions and investor sentiment, adding layers of complexity to their relationship with price changes.

It was also observed that when using SW2014T, a tertiary industry classification metric, to determine a node's neighbors, the resulting node strength was the highest among all metrics evaluated. This observation aligns with the expectation that securities within the same industry tend to exhibit higher correlations with one another. Consequently, utilizing industry-specific classification significantly enhances the predictive power of the model, indicating that industry-related information is crucial for accurately identifying and leveraging the relationships between securities.

Correlation analysis:



The analysis of neighbor similarity for various distinct parameters over time has revealed significant insights into their relationships and implications for portfolio diversification. The neighbor similarity is evaluated by calculating the ratio of times a particular ticker's nearest neighbor is consistently selected based on a specific parameter to the total number of neighbor selections.

It has been observed that the 1-day return (1DR) or the overall neighbor daily return (NDR) exhibits a high correlation with log returns over time. This strong correlation is attributed to the fact that NDR is calculated in a manner similar to return, inherently reflecting similar movements in market prices.

In contrast, other parameter pairs exhibit low correlations, typically below 0.25. This indicates that these parameters maintain relative independence from each other. Specifically, the similarity of SW2014T (tertiary industry classification) with graphs constructed using return-based parameters, such as log return and NDR, is notably low. This reduced similarity can likely be explained by the more granular and detailed nature of the SW2014T classification, which does not align closely with return-based metrics.

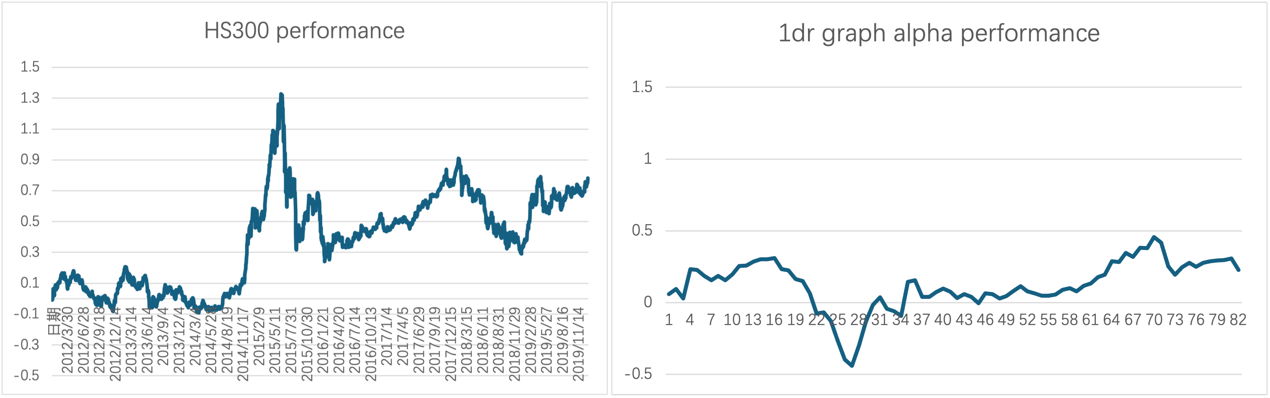
The findings from this analysis suggest that the diverse set of parameters used in the study exhibit good independence from each other. This independence is beneficial for portfolio construction as it enhances diversification, thereby effectively reducing systematic risk. By incorporating parameters with low correlations, investors can create portfolios that are more resilient to market fluctuations and systemic shocks.

Overall, the study highlights the importance of using a varied set of parameters in financial analysis to achieve robust and diversified portfolios. The independence of these parameters helps mitigate risks and improves the stability and performance of investment strategies over time.

**Alpha performance analysis:**

**之后在全市场计算，剔除industry bias**

**因子在不同板块的表现可能**



In comparing the performance of the 1-day return (1DR) with the HS300 index, it becomes evident that the raw 1DR performs worse than the benchmark. This conclusion is drawn from two key observations: higher volatility and relatively lower absolute return. The raw 1DR demonstrates significantly greater volatility, indicating that its daily returns fluctuate more dramatically compared to those of the HS300 index. This increased variability suggests a higher risk profile, making 1DR less stable and predictable for investors. Additionally, despite its higher risk, the 1DR shows a lower absolute return than the HS300, meaning that, on average, it generates smaller gains.

1. 多空的pnl
2. 市值的暴露
3. 对于SW2014T的因子，当前用的是log return 计算alpha,有市值暴露，是否用size neutral return 来计算