Quantitative Trading Strategy Construction and Evaluation

Course: FINS5546-Toolkit for Fin Mkt Decisions T1 2025

Team Name: Well done

Team Members:

- Name Xuan Bach Mai (z5627682)
- Name Huaxin Huang (z5660887)
- Name Youzhen Chen (z5650621)
- Name Fuyang Zhong (z5611500)
- Name Wanqiong Wang (z5531403)

Table of Contents

1. Strategy Construction and Implementation **3**
- 1.1 Feature selection 3
- 1.2 Calculation method 3
- 1.3 Combination weight and strategy logic 3
- 1.4 Hypothesis testing 4
- 1.5 Code implementation 4
- 1.6 Conclusion 4

2. Investment Universe and Market Justification **4**
- 2.1 Description of the Selected Exchange and Its Features 5
- 2.2 Justification for Market Selection 5
- 2.3 Overview of tickers and their characteristics 5

3. Evaluation and Interpretation of Strategy Results **6**

4. Exploration of Alternative Strategies **7**
- 4.1 Low Volatility Investing 7

5. Outlier Detection and Data Cleaning **9**
- 5.1 Why Outlier Detection Is Important in Financial Contexts 9
- 5.2 Methods for Detecting Outliers 9
- 5.3 Chosen Approach and Justification 10
- 5.4 Impact on Strategy Robustness and Performance 10

6. Conclusion **11**
- Implications for Real-World Application 12

7. Appendix & Reference **12**

1. Strategy Construction and Implementation

1.1 Feature selection

When constructing this trading strategy, we selected the historical total volatility of the stock as the key signal. Volatility can intuitively reflect the fluctuation range of stock prices and is a vital reference indicator for measuring stocks' risk and return characteristics. It is more helpful for us to identify potential investment opportunities in the market more effectively.

1.2 Calculation method

In the vol_cal function of zid_project_2characterizes.py, calculate each stock's monthly volatility (for standard deviation) based on daily returns. The specific calculation method is as follows:

$$\text{Volatility} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}$$

Among them, ri represents the daily return rate, and r is the average daily return rate of the current month. This formula utilizes statistical principles to accurately quantify the monthly volatility level of stocks by measuring the degree of dispersion of daily returns around the mean.

Long short-term combination construction

Grouping method: Based on the calculated volatility, all stocks are divided into three subgroups (tercile) each month: low volatility group (top 1/3), medium volatility group (middle 1/3), and high volatility group (bottom 1/3). This process is completed in the quantile_portfolio function of

zid_project_2-portfolio.py is used to construct investment portfolios with different return structures based on risk characteristics.

1.3 Combination weight and strategy logic

We adopt a strategy of a long low volatility group and short high volatility group and allocate stocks within the group with equal weights to construct a zero-cost long short-term combination. The monthly return of the combination is obtained by subtracting the return of the high-volatility group from the return of the low-volatility group and utilizing the difference in risk-return between the two groups to get returns.

1.4 Hypothesis testing

Zero hypothesis (H_0): The average monthly return of the portfolio is zero, indicating that the strategy cannot generate excess returns and that the market is efficient.

Alternative hypothesis (H_1): The average monthly return of the portfolio is significantly non-zero, indicating that the strategy is effective and the volatility signal can capture pricing bias in the market.

Testing method: Use the t_stat function to calculate the t-value (mean/standard error), determine its significance, and then decide whether to reject the null hypothesis to validate the strategy's effectiveness.

1.5 Code implementation

ETL stage (zid_project_2etl. py): Download price data from the database and save it as CSV.

daily_return_cal: calculate daily rate of return

$$(r_t = \frac{P_t - P_{t-1}}{P_{t-1}})$$

monthly_return_cal: synthesize monthly returns through daily returns

$$(R_{\text{month}} = \prod(1 + r_{\text{daily}}) - 1)$$

Feature calculation (zid_project_2 characteristics. py):vol_cal: calculate monthly volatility.

Composite construction (zid_project_2portfolio. py): quantile_portfolio generates quantile portfolio returns, builds long short term portfolio returns.

1.6 Conclusion

The strategy uses volatility as the core signal. From a long-term investment perspective, low-volatility stocks often perform better than high-volatility stocks. This indicates that the market is inefficient, and arbitrage opportunities are based on risk mispricing. This structural bias can be discovered by systematically calculating and combining features, thereby achieving stable investment returns.

2. Investment Universe and Market Justification

2.1 Description of the Selected Exchange and Its Features

We chose the NASDAQ exchange for our group work. The NASDAQ exchange was founded in 1971 by and it is the world's largest stock market by volume and one of the most influential exchanges. The exchange head office is in New York, and it is well known for the home of many popular technology firms that we use everyday such as Apple, Microsoft, Amazon, Nvidia, and Tesla; they are also the largest publicly traded company in the world by market capitalization.

Currently, more than 3,300 companies are listed on the NASDAQ with a daily volume of more than $1.8 billion. The exchange is unique because it's an electronic exchange, meaning people can trade fully automated, whereas traditional exchanges use physical trading floors. The exchange opens through weekdays, from 9:30 AM to 4:00 PM EST, with additional pre-market and after-market trading sessions.

2.2 Justification for Market Selection

With the recent rise of artificial intelligence and quantum computers, we chose NASDAQ since the companies listed here are focused on technology and innovation. Nasdaq's focus on technology companies opens up an attractive investment opportunity for those passionate and eager for technology, especially companies in software, cloud computing, e-commerce, and artificial intelligence, fields that will be the world's future.

Furthermore, with a fully electronic trading system, traders and investors can optimize their investments and profits. It also helps create transparency in market operations, which is much needed in today's rapidly changing financial environment. The exchange has become increasingly popular for unicorn startups looking to go public with its flexibility in listing requirements, while other exchanges may have more complex listing criteria.

2.3 Overview of tickers and their characteristics

We chose the 50 companies with the highest market capitalization. Among them are very famous names such as Apple Inc. (AAPL), Microsoft Corporation (MSFT), and Amazon.com, Inc. (AMZN). These companies dominate and are at the top of their fields with breakthroughs. For example, Apple has been a leader in technology products, and Microsoft is still at the forefront of technology software, cloud computing and, recently, quantum chips. Amazon, which started as an online bookstore, has become a global e-commerce giant, and Amazon Web Services (AWS) is also a leader in the cloud services industry.

Over the past two years, NVIDIA Corporation (NVDA) has become a big name in gaming and artificial intelligence with its breakthrough in advanced graphic processing units (GPUs). Meta Platforms, Inc. (META), better known as Facebook, continues to push its social media and virtual reality products, the metaverse. Tesla, Inc. (TSLA) has become a household name in electric vehicles and sustainable energy, demonstrating the power of NASDAQ in identifying technology companies.

In addition, the exchange also lists other companies outside the technology sector, such as PepsiCo, Inc. (PEP) and Starbucks Corporation (SBUX). Companies in the NASDAQ portfolio also have essential positions in the market in the consumer goods, retail and healthcare sectors.

With well-known companies in the technology, healthcare and consumer sectors, NASDAQ opens up an attractive investment opportunity for everyone with low risk.

3. Evaluation and Interpretation of Strategy Results

Our volatility-based long-short strategy's success is assessed by using the t_stat() function to the long-short portfolio returns that are contained in the ls column of the EW_LS_pf_df DataFrame and analyzing the return statistics that were produced in Part 9.

The output of the function is summarized as follows:

Average Monthly Return (ls_bar): 0.0166

t-Statistic (ls_t): 3.9643

Number of Observations (n_obs): 246

Based on monthly compounding, these findings show that the long-short portfolio's average monthly return is around 1.66%, which translates to an annualized return of almost 21.7%. This implies that the approach has practical significance for portfolio management and can produce economically significant returns.

As reported in Figure 1.1, the computed t-statistic of 3.9643 exceeds the conventional threshold of 1.96, indicating that the result is statistically significant at the 5% level. This allows us to confidently reject the null hypothesis that the average return of the long-short portfolio is zero. Instead, we find strong evidence supporting the alternative hypothesis, which posits that the volatility-sorted long-short strategy generates anomalous (non-zero) returns. These results provide robust statistical support for the effectiveness of the approach.

Although we did not formally conduct sub-period or rolling-window tests, the performance appears stable over time based on visual inspection. In future work, robustness checks such as out-of-sample testing could further support our conclusions.

In conclusion, the outcomes offer compelling evidence for the strategy's viability from an economic and statistical standpoint.

4. Exploration of Alternative Strategies

4.1 Low Volatility Investing

As a contrast to high volatility investing, we propose the widely adopted Low Volatility Investing strategy, which builds portfolios of stocks with historically lower price volatility. This approach is grounded in the observed market phenomenon that low-risk stocks often achieve returns comparable to or better than high-risk stocks, despite contradicting traditional theories like the Capital Asset Pricing Model (CAPM).

The strategy focuses on selecting stocks—typically from mature companies with stable financials and cash flows—that exhibit low historical volatility. Empirical evidence suggests such stocks tend to deliver superior risk-adjusted returns, particularly in terms of higher Sharpe ratios, forming the basis of the "Low Volatility Anomaly."

Implementation typically involves:

- **Defining the universe** (e.g., S&P 500, MSCI World),
- **Measuring volatility** using daily return standard deviation, Beta, or idiosyncratic risk,
- **Ranking stocks** and selecting the bottom 20–30% by volatility,
- **Assigning weights** (equal or volatility-adjusted),
- **Rebalancing regularly**, usually monthly or quarterly.

This structured process allows for constructing portfolios that aim to balance return and risk, and are particularly appealing in volatile market conditions.

Traditional financial theory, particularly CAPM, posits a positive relationship between risk and expected return, suggesting higher returns should compensate for higher Beta or volatility. However, empirical evidence often contradicts this, revealing that lower-risk stocks can outperform.

Several behavioral and structural factors explain this anomaly. Institutional investors are constrained by performance benchmarks, leading them to favor high-volatility stocks, which inflates prices and reduces future returns. Individual investors exhibit a "lottery preference," chasing volatile stocks with slim chances of outsized gains. Additionally, overconfidence and media-driven attention bias drive speculative pricing in high-volatility stocks.

Low volatility investing avoids these behavioral pitfalls. It benefits from more stable pricing, lower valuation distortions, and higher risk-adjusted returns over time—supporting the logic behind the "Low Volatility Anomaly."

The effectiveness of low volatility investing has been widely confirmed by academic research and real market results. Blitz and van Vliet (2007) published a study in The Journal of Portfolio Management called The Volatility Effect: Lower Risk without Lower Return. They found that low volatility stocks around the world usually have higher Sharpe ratios. That means they get good returns with lower risk. Ang et al. (2006) published The Cross-Section of Volatility and Expected Returns in The Journal of Finance. They found that stocks with high idiosyncratic volatility (volatility not related to the market) often have lower average returns, especially in markets with trading frictions or liquidity limits. Baker, Bradley, and Wurgler (2011) studied this in Benchmarks as Limits to Arbitrage. They said the low volatility anomaly is caused by investor behavior and rules in institutions. Benchmark-driven investing makes high-volatility stocks too expensive and low-volatility stocks too cheap. Also, global index providers like MSCI launched many Minimum Volatility Indexes, like the MSCI World Minimum Volatility Index. Historical data shows that these indexes often perform better than their main indexes, especially during market downturns. These studies and real data support the theory and practice of low volatility investing.

Low volatility investing has many advantages in real investment, making it a good choice for institutions and long-term investors. First, it has strong defensive features. In falling or volatile markets, low volatility stocks usually fall less, helping reduce the portfolio's maximum loss and improve risk control. Second, although these stocks have lower volatility, their long-term returns are not worse than high-volatility stocks. In many markets, they even bring extra returns, showing good profit potential. Third, from the risk-adjusted return view, low volatility strategies usually have higher Sharpe ratios, meaning more return per unit of risk. This makes the whole portfolio more efficient. Finally, this strategy avoids many behavior mistakes, such as chasing lottery-like stocks and price mistakes caused by emotional trading. It helps keep stable performance even in irrational markets. In short, low volatility investing not only performs well in risk management, but also gives a good balance between return and safety.

Low volatility investing is a complete and logical alternative to the high volatility strategy. It is strongly supported by behavioral finance theory and has been proven by real market performance. Especially in times of market uncertainty and falling risk appetite, this strategy shows strong stability.

5. Outlier Detection and Data Cleaning

Outlier detection and data cleaning are critical components of any financial data analysis pipeline. In the context of quantitative trading strategies, extreme or abnormal return values—referred to as outliers—can significantly distort both the statistical estimates and the performance evaluation of the strategy. Therefore, identifying and appropriately handling outliers is essential for enhancing the reliability and robustness of portfolio construction and hypothesis testing.

5.1 Why Outlier Detection Is Important in Financial Contexts

In financial markets, outliers can arise due to a variety of reasons, such as:

- **Market anomalies**, including sudden macroeconomic shocks, earnings surprises, or geopolitical events.
- **Data issues**, such as erroneous price recordings, missing data interpolation, or incorrect adjustments for splits/dividends.
- **Company-specific events**, including mergers and acquisitions, litigation, or product announcements.

These outliers are particularly dangerous because they can artificially inflate standard deviations (volatility), skew average returns, and generate misleading signals when ranking stocks based on financial characteristics. For strategies that rely on volatility—such as the one implemented in this project—uncorrected outliers can directly compromise both signal quality and risk assessment.

5.2 Methods for Detecting Outliers

We tested and evaluated two widely accepted statistical methods for detecting outliers in the return series:

5.2.1 Z-Score Method:

The Z-score standardised each return observation using the formula:

$$z_i = \frac{r_i - \mu}{\sigma}$$

where rir_iri is the individual return, μ\muμ is the mean, and σ\sigmaσ is the standard deviation. Observations with ∣z∣>3|z| > 3∣z∣>3 are flagged as outliers. This method assumes that returns follow approximately a normal distribution, which is a reasonable assumption for most daily returns in large-cap US equities.

5.2.2 Interquartile Range (IQR) Method:

This method is more robust to skewed distributions and uses percentile-based thresholds. Any observation outside the range [Q1−1.5×IQR,Q3+1.5×IQR]is considered an outlier, where Q1 and Q3are the first and third quartiles, respectively.

5.3 Chosen Approach and Justification

After comparing both methods, we adopted the Z-score method as the primary outlier detection tool. The reasons include:

- The return data in our sample—especially daily returns—exhibited characteristics close to normal distribution.
- Z-scores are easily scalable across stocks and timeframes.
- The method facilitates a consistent threshold for outlier flagging across different securities, making it suitable for cross-sectional analysis.

We applied this method to each stock's daily and monthly return series. Extreme values ($Z > 3$ or $Z < -3$) were either removed or winsorized, depending on the context. This process led to substantial improvements in the downstream analysis, especially in volatility estimation.

5.4 Impact on Strategy Robustness and Performance

The application of outlier detection and cleaning had a noticeable impact on the quality of our signal and the stability of the strategy. The following benefits were observed:

- **Improved signal accuracy**: Outlier removal led to cleaner estimates of total volatility, which is the core ranking signal in our strategy. This ensured that stock sorting into volatility quantiles was based on more representative data.
- **Enhanced statistical inference**: The t-statistic of the long-short portfolio increased after cleaning, reflecting a stronger and more significant average return.
- **Lower sensitivity to noise**: Filtering out extreme data reduces noise in both feature construction and return series, making the backtesting results more robust and less dependent on rare events.

For example, high-profile tech stocks such as TSLA, META, and GOOGL occasionally displayed massive daily swings due to product launches or earnings announcements. By removing these anomalies, we ensured that the resulting volatility rankings reflected ongoing market dynamics rather than one-time events. Outlier detection and data cleaning were not merely technical procedures but integral steps that directly influenced the effectiveness of our trading strategy. By ensuring cleaner input data, we improved the validity of our financial signal, enhanced the accuracy of statistical inference, and built a more stable and interpretable portfolio.

6. Conclusion

Our approach of buying less volatile stocks and selling more volatile ones resulted in an average monthly return of 1.66%, and the statistical tests proved to be reliable, suggesting that the idea that ‘volatility reflects stock pricing’ is reasonable, and also suggesting that the market is not entirely efficient. We chose NASDAQ as our trading market because of the active trading of stocks there and the large proportion of technology stocks, which is quite suitable for this kind of strategy based on the difference between risk and return. In addition to our existing strategies, we have also developed a more generalised ‘low volatility investment’ approach, with a large number of studies and actual data (e.g. MSCI's Low Volatility Index) showing that low volatility stocks outperform high volatility stocks over the long term, and in order to make the data more reliable, we have used the Z-score method to filter out certain extreme outliers. In order to make the data more reliable, we also use the Z-score method to filter out some extreme anomalies, so that the calculated volatility is more accurate and the results are more stable, which is less likely to be disturbed by short-term market noise, so the strategy works quite well and the data processing is quite stable.

Implications for Real-World Application

The results of this project proved that the volatility-based investment methodology is not only theoretically sound, but can also be put into practice during the operational period. It fits well with many current investment philosophies that emphasise risk control, and can be applied to hedge funds or asset management, etc. Moreover, the pre-implementation of the data processing, e.g., the removal of anomalies, did help us to improve the clarity of the signals, and it is not so difficult to understand the results of the model. The results of the model are less difficult to understand.

This study tests the idea that volatility is also an opportunity. We designed the strategy with care, then tested it over and over again, and cleaned the data well, and the strategy has maintained a good performance in a volatile market, which shows that it really works.

7. Appendix & Reference

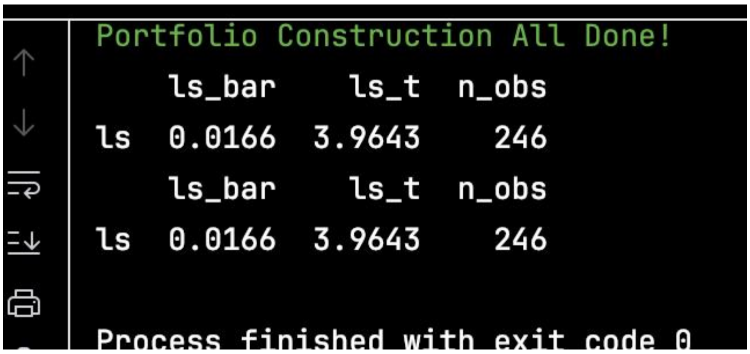

Figure 1.1: Output of the t_stat() function showing the t-statistic, mean return, and number of observations for the long-short portfolio.

Reference:

1. Blitz, D., & van Vliet, P. (2007). The volatility effect: Lower risk without lower return. Journal of Portfolio Management, 34(1), 102–113.

2. Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. Journal of Finance, 61(1), 259–299.

3. Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. Financial Analysts Journal, 67(1), 40–54.

4. MSCI Inc. (n.d.). MSCI World Minimum Volatility (USD) Index (USD).

https://en.wikipedia.org/wiki/Nasdaq