Adaptive Pairs Trading with Bayesian MCMC Sampling

MTH 585 FINAL PROJECT

Amy Chang, Netra Mittal, Renjie Wang, Sayali Pingle

Contents

1 Introduction

This study elevates pairs trading by applying Bayesian methods and Probabilistic programming, enabling dynamic stock pair selection and co-integration tracking, thus quickly adapting to market changes. It contrasts with static frequentist methods which lack such responsiveness. Incorporating real-time risk tools like dynamic VaR and drawdown measures, along with Bayesian Gaussian Random Walks, it finely tunes to shifts in stock correlations for improved prediction reliability. The approach, building on Dias (2021)[1], not only refines trading strategy but also integrates direct uncertainty handling into decisions. This paper underlines our novel, adaptable, and risk-informed contributions, aiming to redefine trading practices amid market uncertainties.

1.1 Literature Review

In developing a refined approach to pairs trading, this research draws from a robust body of work that examines the integration of Bayesian methods into financial strategies. Traditional models in pairs trading, often based on the frequentist perspective, have been critiqued for their static nature, which typically results in less adaptability to market fluctuations (Gatev et al., 2006)[2].

Integrating insights from Jarrow (2010)[3], this study advances pairs trading models by emphasizing the Bayesian framework's response to model error and parameter uncertainty. Jarrow's analysis reveals the limitations of conventional models that overlook such errors, resulting in potential underestimation of risk. Our approach aligns with Jarrow's advocacy for a methodology that produces more reliable forecasts without relying on large samples—a significant move beyond the frequentist models that have traditionally shaped financial strategy development.

This shift is pivotal for real-time market adaptability and reflects a broader industry trend towards incorporating probabilistic programming to better manage complex, fluctuating financial systems. The enhanced approach to pairs trading in this research is informed by the intricacies identified in the work of Roberts and Rosenthal (2009)[4]. Their exploration into the Random Walk Metropolis (RWM) algorithm reveals significant constraints when applied to target distributions with i.i.d. product forms, highlighting a discrepancy between theoretical optimal scaling results and their practical applications.

In our methodology, we employ Bayesian Gaussian Random Walks, which, while recognized for their intricate handling of stochastic financial series, must be examined beyond the often-cited i.i.d. product structures to truly grasp the complexity of market dynamics and cointegration shifts—crucial for the robustness of convergence trading strategies.

1.2 Bayesian Model

The Bayesian Pairs Trading model operates under the premise that the relationship between pairs of stocks, quantified through the β coefficient, evolves over-time in a auto-stochastic manner, rather than remaining static. The Gaussian Random Walk element of the model allows for the sequential updating of beta, making it dependent on its last known value but with a random fluctuation that accounts for market volatility and other unforeseen changes.

In practical terms, this means that each new estimate of beta is influenced by the previous estimate, but with a random error that ensures flexibility and adaptability to new market information. This is a key distinction from traditional static models, where beta remains fixed over time and fails to capture dynamic market behaviors. The Bayesian framework enhances this model by providing a probabilistic backbone. Bayesian inference is utilized to update the beliefs about the beta parameter after observing new data, effectively learning from each new piece of information and refining the model's predictions continuously.

This probabilistic learning approach is crucial in trading environments characterized by uncertainty and frequent change. It allows traders to compute and adjust the probability distributions of future market states based on all historically observed data, rather than just relying on point estimates or static models. As a result, traders can make more informed decisions that take into account the likelihood of various outcomes, thereby managing risk more effectively.

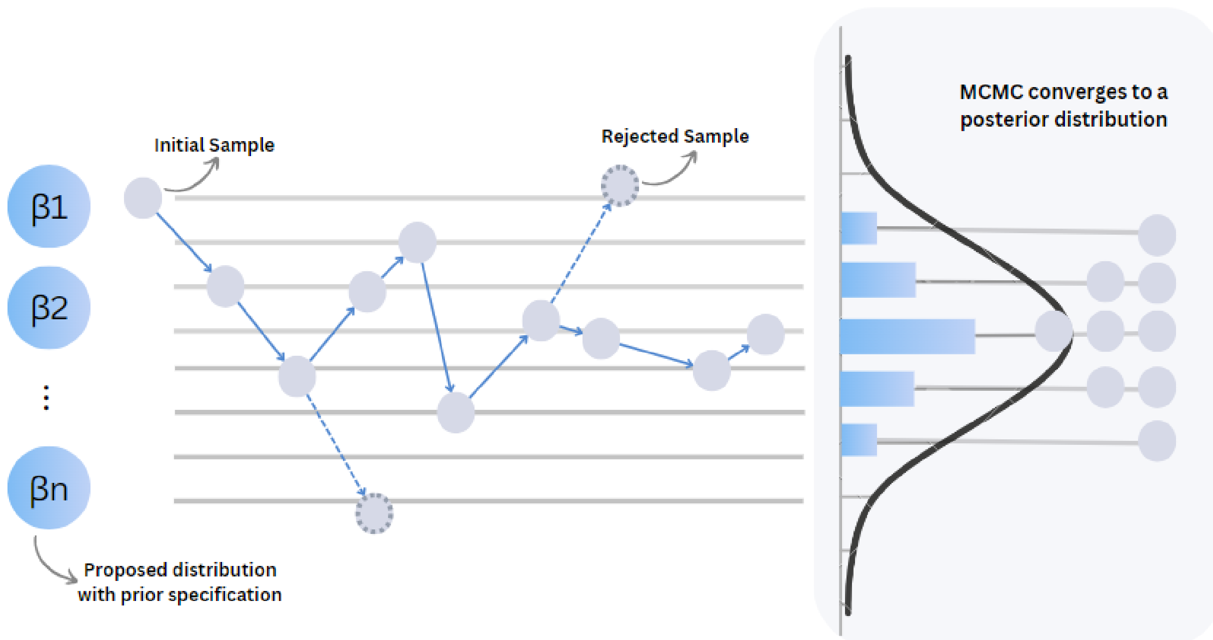

Figure 1: MCMC convergence process

In Figure 1, we showcase the methodology underlying our Bayesian model for dynamic pairs trading. The Markov Chain Monte Carlo (MCMC) convergence process is a statistical method used to approximate the posterior distribution of model parameters. It starts with an initial sample and iteratively proposes new parameter values. Some proposals are accepted and others are rejected based on a probability criterion that guides the algorithm toward areas of higher likelihood. Over many iterations, this process creates a chain of samples that, given enough time and under proper conditions, will represent the true posterior distribution of the parameters. This allows us to make probabilistic inferences about the parameters based on the observed data.

We used the PyMC Python library to execute our Bayesian MCMC model. The model starts by establishing priors that reflect our initial assumptions:

- We select an exponential prior for the volatility in beta's progression, implying a belief in the rare occurrence of large shifts:
$$\sigma_\beta \sim \text{Exponential}(50)$$

- Beta is modeled as a Gaussian Random Walk, reflecting our expectation that changes in beta over time are continuous yet subject to random variation:
$$\beta \sim \text{GaussianRandomWalk}(\sigma = \sigma_\beta)$$

- The relationship between the two stocks is captured through a regression model, indicating our prior belief in a proportional relationship:
$$\text{stock2} = \beta \cdot \text{stock1}$$

- A half-normal distribution is chosen for the observation noise's standard deviation, indicating a zero mean with a concentration of values around a small standard deviation, as large deviations are less likely:
$$\sigma_L \sim \text{HalfNormal}(0.1)$$

- The likelihood (L) of observing our data is assumed to be normally distributed, based on the central limit theorem and the prevalence of normal distributions in empirical financial data:
$$\text{L} \sim \text{Normal}(\mu = \text{stock2}, \sigma = \sigma_L)$$

- The posterior distribution is simulated using the MCMC sampling where 500 samples are considered to generate the posterior and a burn of another 500 samples. The sampling uses 4 CPU cores for parallel processing. Following is the PyMC command:
$$\text{trace(t)} = \text{pm.sample(500, tune=500, cores=4, model=model)}$$

As the market evolves, new data is incorporated, allowing for the real-time recalibration of beta. This leads to an updated posterior distribution, providing a broad range of beta estimates to guide trading decisions. This feedback loop demonstrates the adaptability of our strategy, optimized for a dynamic financial landscape.

2 Adaptive Pairs Trading

Adaptive pairs trading focuses on tracking a wide range of financial metrics, trading data, and company-specific indicators to pinpoint which variables are most significant in predicting stock performance. This data helps refine the selection process when focusing on stocks to trade. By meeting specific data-driven criteria, the portfolio can achieve a higher diversification level and the selected stocks can discover unique profit opportunities.

2.1 Dynamic Stock Selection

Dynamic stock selection highlights the adaptability of stock selection per trading period. Over the years company profitability and growth changes, so the portfolio should reflect these changes. By continuously updating the stocks based on market criteria, financial metrics, and sentiment analysis, survivorship bias can be avoided. Survivorship bias is selecting only the "survivors", or outperforming individuals, without looking more broadly at the whole dataset.[5] While the algorithm selects the best of the best, it is flexible enough to realize the best can change each day. Consequently, the algorithm can focus on trading company shares that have grown sufficiently large enough to be considered stable investments.

2.2 Universe Selection

The Universe Selection model creates Universe objects, which select the assets for an algorithm. As the universe changes, the algorithm is notified through an event handler. With this event handler, other parts of the algorithm can track the current universe constituents without breaking the separation of concerns design principle.

When selecting assets, certain thresholds such as market cap, P/E ratio, or Debt/Equity ratio can be set. To limit consideration of companies with lower volatility and select those with stable gains, we focused on companies with a share price greater than $10 and average trading volume greater than fifty-thousand. To increase the chances of profitability, the universe selector was narrowed down to choose the top twenty companies with the highest market capitalization. Market Capitalization is important because it showcases a company's value by extrapolating based on its market value. Market capitalization is measured by multiplying the share price by the number of available shares. The final criteria used when refining the universe was the stock industry. Often in bullish markets, technology stocks do favorably and even in bearish markets the software companies are still a win according to John Freeman, vice president of equity research at the Center for Financial Research and Analysis (CFRA).[6]

2.3 Cointegration

Cointegration occurs when two or more non-stationary time series have a stable long-run relationship despite exhibiting some degree of short-term differences. This relationship is essential to the success of pairs trading[7]. To determine if such a relationship exists between stocks, cointegration tests must be performed to identify scenarios where two or more non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. These tests can also identify the degree of sensitivity of two variables to the same average price over a specified period of time.

Potentially Cointegrated Stocks

Figure 2: Two stocks, IBM and ABC, display their long-term price relationship which indicates potential for cointegration between them.

Figure 2 displays the long-term price relationship between two potentially cointegrated stocks, IBM and ABC.[1] It should be noted that ABC announced the completion of its name and stock ticker change from AmerisourceBergen Corp. (NYSE: ABC) to Cencora, Inc. (NYSE: COR) on August 30, 2023 to market the latest stage of its evolution as a leader in global pharmaceutical services. [8] From the graph, the stocks are see to have a large divergence in price in 2018 and 2019, but then they converge to a similar price range in 2020 and 2021. This short term difference and long-term similarity makes this pair a good candidate for potential cointegration.

To test for cointegration, there are numerous statistical tests one could perform to determine the relationship between two stocks. One test, the Johansen Test, was a top contender for many reasons such as its avoidance of choosing a dependent variable during calculation and its avoidance of issues stemming from errors carried from one step to the next. However, it proved difficult to implement as the test could only check for cointegration between twelve time series at time. Instead the Augmented Dickey-Fuller Test was implemented.

Previously, a study was done to test the most effective measurement of a relationship between two stocks. They calculated the distance between two stocks, the stationarity between two stocks, and the cointegration between two stocks. Figure 3 displays the results of this study by showing the cumulative excess returns (log scale) including transaction costs pairs trading strategies with a 2-standard deviation opening trigger versus equity premium. [7] In the graph, "stat" represents stocks traded according to stationarity, "dist" represents stocks traded according to their distance calculaation, and "coint" represents stocks traded according to their cointegration value.

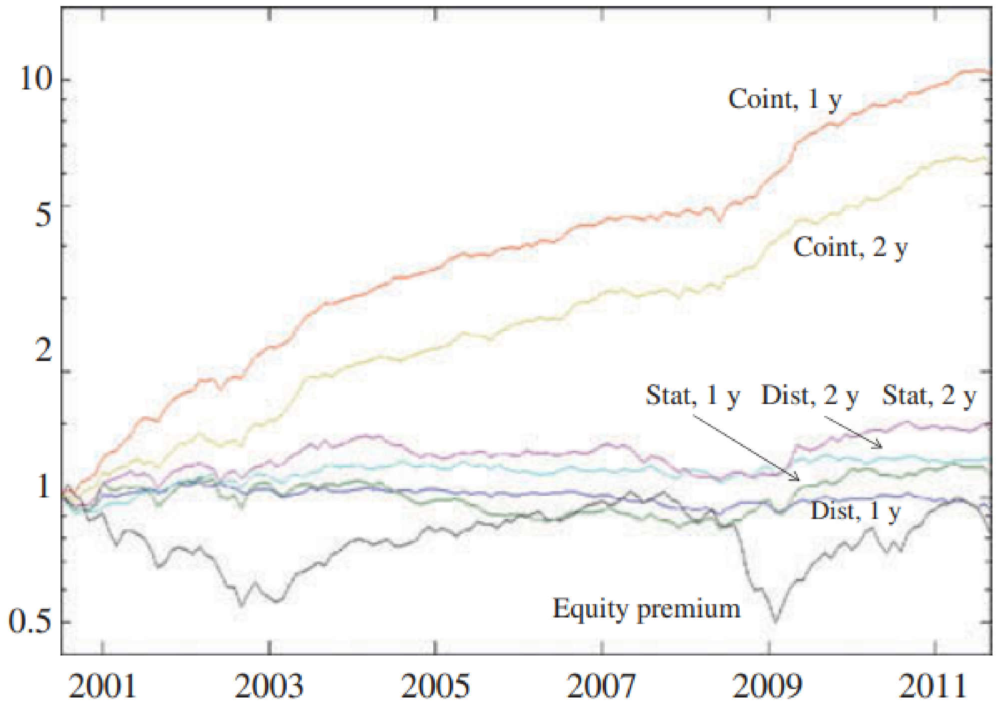

Figure 3: Cumulative excess returns (log scale) including transaction costs pairs trading strategies with a 2-standard deviation opening trigger versus equity premium.

The fundamental idea is to find two stocks whose prices are cointegrated, meaning they tend to move together over the long run, but may diverge in the short-term. Thus, cointegration is an accurate way to determine the optimal pair of stocks for trading. By buying one stock and simultaneously selling the other, a trader can profit from the price difference if and when the stocks converge back to their long-term relationship.

2.3.1 Augmented Dickey-Fuller

The Augmented Dickey-Fuller test is another test to determine the cointegration relationship between two variables. It derived from the Dickey-Fuller test which is a unit root test that determines if a time series is non-stationary.[9] A unit root is said to exist in a time series if $\alpha = 1$ following

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

where,

$$Y_t = \text{value of time series at time } t$$

and

$$X_e = \text{an exogenous variable}$$

The Dickey-Fuller test expanded on the unit root test while maintaining the same null hypothesis that alpha = 1 in the model equation

$$y_t = c + \beta t + \alpha y_{t-1} + \phi \Delta Y_{t-1} + e_t$$

where,

$$y_{t-1} = \text{lag 1 of time series}$$

$$\delta Y_{t-1} = \text{first difference of the time series at time } (t-1)$$

The Augmented Dickey-Fuller test expands the Dicker-Fuller equation to include a high order regressive process.

$$y_{t-1} = c + \beta t + \alpha y_{t-1} + \phi_1 \Delta Y_{t-1} \tag{1}$$
$$+ \phi_2 \Delta Y_{t-2} + \ldots + \phi_p \Delta Y_{t-p} + e_t \tag{2}$$

With the same null hypothesis, the Augmented Dickey-Fuller test statistic and p-value are calculated. If the p-value of is less than 5%, we can reject the null hypothesis that there is a unit root. This means that the time series is NOT non-stationary. Another way to interpret this is that the stock price probability distribution does not change when it is shifted in time, a characteristic known as stationary. Thus, by the p value, a pair of stocks are suitable for pairs trading. When comparing ADF statistic values amongst other pairs, the smallest value means the pair will have stronger cointegrating relationship. [10]

3 Algorithm

The model generates beta values for each time period (subjective to the trading frequency: here, daily resolution) obtained from the simulated posterior as depicted in the graph below.

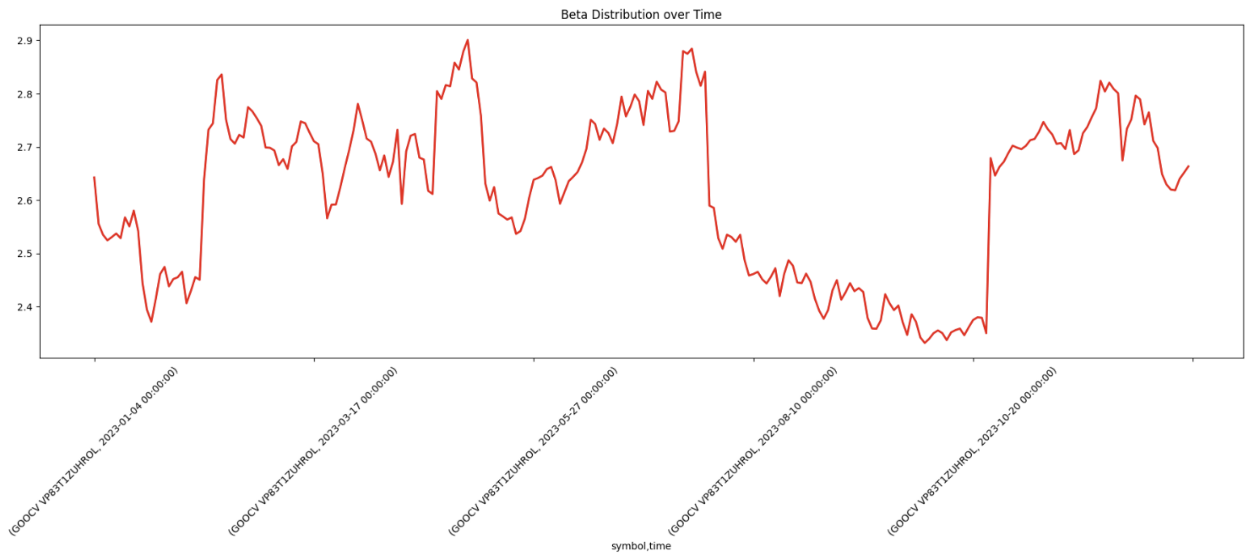

Figure 4: Beta distribution from 01/01/2023 to 01/01/2024 for GOOG and MSFT

1. High Beta: A high beta (e.g., above 1.5) suggests that the second stock (MSFT) is more volatile compared to the first stock (GOOG). This might indicate that the second stock has greater risk, but it might also offer higher potential returns.
2. Low Beta: A beta close to or below 1 suggests lower relative volatility, indicating more stability but possibly lower returns.
3. Rising Beta: Indicates the second stock is becoming more volatile relative to the first. This might be due to market conditions, specific events, or changing correlations.
4. Decreasing Beta: Indicates that the second stock is becoming less volatile relative to the first, suggesting increased stability or changes in underlying dynamics.

3.1 Signal Generation

1. Initial Signal: The function calculates an initial trading signal based on the 'fixed beta' value taken from the MCMC simulated beta array at the position defined by the smoothing window. This signal should mean revert to 0 if β remains relatively stationary. The signal is defined as the difference between the product of 'fixed β' and $stock_1$ (GOOG), and $stock_2$ (MSFT).

$$\text{signal} = \text{fixed}_\beta \times \text{stock}_1 - \text{stock}_2$$

2. Smoothing: The signal is then smoothed using a rolling mean with the smoothing window to reduce noise and potential over-fitting to short-term fluctuations. This smoothing window is determined based on the volatility of the market and volatility of the stocks. If the volatility is high (consider

a bearish market) then a smaller smoothing window (3 days) is chosen unlike a 15 day window in case of low volatility.

3. Derivative: The derivative (difference) of the smoothed signal is calculated to detect changes in trend.

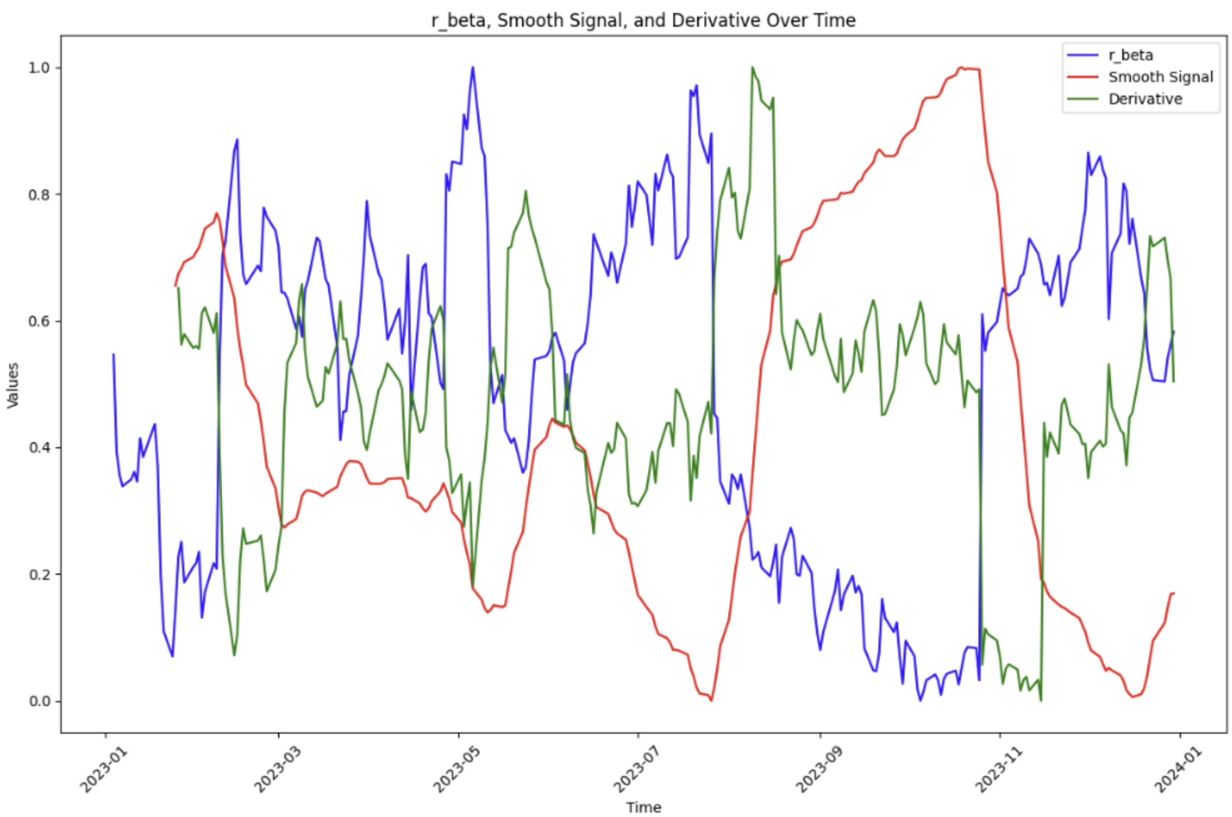

Figure 5: Signal from 01/01/2023 to 01/01/2024 for GOOG and MSFT with 15 day smoothing window

3.2 Strategy

The algorithm delineates three distinct trading scenarios contingent upon the dynamics of the "Smoothed signal" relative to its initial value and the zero line:

1. **Non-Trading Scenario:** If the "Smoothed signal" is above zero and moving downward, we short the portfolio, unlike going long when the signal is negative but the derivative is positive.

2. **Long Trade:** If the "Smoothed signal" crosses below its starting value, any existing long positions are closed as it may indicate a divergence from the mean. However, if the "Smoothed signal" then rises through zero, a new long position is initiated as it suggests a return to the mean.

3. **Short Trade:** If the "Smoothed signal" crosses above its starting value, any existing short positions are closed as it may indicate a divergence from the mean. However, if the "Smoothed signal" then falls through zero, a new short position is initiated as it suggests a return to the mean from the opposite direction.

3.3 Leverage Calculation

- long position

$$\text{leverage}_{\text{stock}_1} = \frac{\text{fixed\_beta}}{|\text{fixed\_beta}| + 1}$$

- short position

$$\text{leverage}_{\text{stock}_1} = -\frac{\text{fixed\_beta}}{|\text{fixed\_beta}| + 1}$$

This formula balances the leverage based on the strength of fixed β. A higher absolute value of beta leads to a higher fraction of the portfolio being dedicated to that position, albeit moderated by the denominator which adds 1 to avoid division by zero and reduce extreme leverage.

The leverage for stock 2 is similarly calculated as the inverse portion of the total leverage, ensuring the total adds up to 1 (or near it) but split according to the strength and direction indicated by fixed β.

The dynamically calculated leverage allows the strategy to adjust its exposure based on the confidence in the trade's direction and strength. This dynamic adjustment helps manage risk and optimize returns.

4 Risk Management

Risk management is a discipline for living with the possibility that future events may cause adverse affects.[11] For each industry, there is a different standard of what constitutes risk and there are different regulations for those risks depending on the particular environment. The management of risk at financial institutions encompasses many tasks such as managing market risk, credit risk, and operational risk.

Market risk is the risk of change in the value of a financial position or portfolio due to a change in the value of its components. Under market risk lies model risk and liquidity risk. Model risk is the risk associated with using a misspecified model for measuring risk. Liquidity risk is the risk associated with the marketability of an investment specifically, how quickly an investment can be bought or sold to prevent or minimize loss. Some specific tasks such as ensuring portfolios are diversified and optimized according to risk-return considerations help mitigate these risks. Financial institutions must also manage risk associated with not receiving promised repayments on outstanding investments which is called credit risk. Additionally, these institutions must manage risks related to inadequate or failed internal processes, people and systems, and external events which is operational risk.

To monitor market risk and liquidity risk, the trading strategy measures many specific values that help mitigate financial risks. The top risk measurements observed during this strategy are value at risk, expected shortfall, maintenance margin, drawdown, beta, stop loss, and slippage.

4.1 Value at Risk

Value at risk, or VaR, is a widely used risk metric in financial institutions. VaR can be thought of as the maximum loss that is not exceeded with a given high probability. Consider a portfolio of risky assets and a fixed time horizon Δt, and denote by $F_L(l) = P(L \leq l)$ the distribution function of the corresponding loss distribution. Given some confidence level $\alpha \in (0, 1)$, the VaR of a portfolio with loss L at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1 - \alpha$. Represented formally,

$$VaR_\alpha = VaR_\alpha(L) = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}$$

In probabilistic terms, VaR is a quantile of the loss distribution. In the trading strategy, an $\alpha = .95$ was used to maintain a portfolio loss less than the expected loss for the period specified.

4.2 Expected Shortfall

Expected shortfall, or ES, is closely related to value at risk. Consider the same portfolio from value at risk with the same distribution function $F_L(l) = P(L \leq l)$ corresponding to the loss distribution. For a loss L with $E(|L|) < \infty$ and distribution function F_L, the ES at confidence level $\alpha \in (0, 1)$ is defined as

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_L)\, du$$

where $q_u(F_L) = F_L^{\leftarrow}(u)$ is the quantile function of F_L. The condition $E(|L|) < \infty$ ensures the above integral is well defined. To further explain the relationship between expected shortfall and value at risk, they follow

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(L)\, du$$

Instead of establishing a specific confidence interval α, VaR is averaged over all levels $u \geq \alpha$ to look further into the tail of the loss distribution. ES_α depends on the distribution of L and $ES_\alpha \geq VaR_\alpha$. Thus, expected shortfall can be loosely interpreted as the expected loss that is incurred in the event that value at risk is exceeded. Overall, the main idea behind expected shortfall is that it oversees the amount of loss in the worst of cases.

4.3 Maintenance Margin

Initial margin is the amount of money that must be deposited at the time an investor enters a trading contract. The margin account is used to reflect the investor's gain or loss. At the end of the day, daily settlement, or marking to market, occurs and the funds are deposited or withdrawn from the margin account accordingly. To ensure that the balance in the margin account never becomes negative, a maintenance margin is established. Maintenance margin is typically slightly lower than the initial margin. Maintaining a good margin balance is critical with respect to risk because if at any time the margin account falls below the maintenance margin, the broker can make margin call.[12] A margin call is when the broker makes a request for more funds to be added to the margin account to meet the required maintenance balance. If the investor cannot add more funds, the broker can forcefully liquidate trading positions to recoup those funds. By maintaining a plentiful margin, the trading positions remain intact and the portfolio avoids any forceful liquidation of its assets.

4.4 Drawdown

Drawdown is the peak-to-bottom loss over a given period of time in the performance curve of an investment. It is also important to note drawdown duration, the length of time between new equity highs, to prevent extreme loss in the portfolio. A good strategy will aim to limit peak-to-trough decline during an investment and will strategically trade should the length of time increase. For the trading strategy it was designed to have a drawdown less than 10% and a drawdown duration of less than four months.

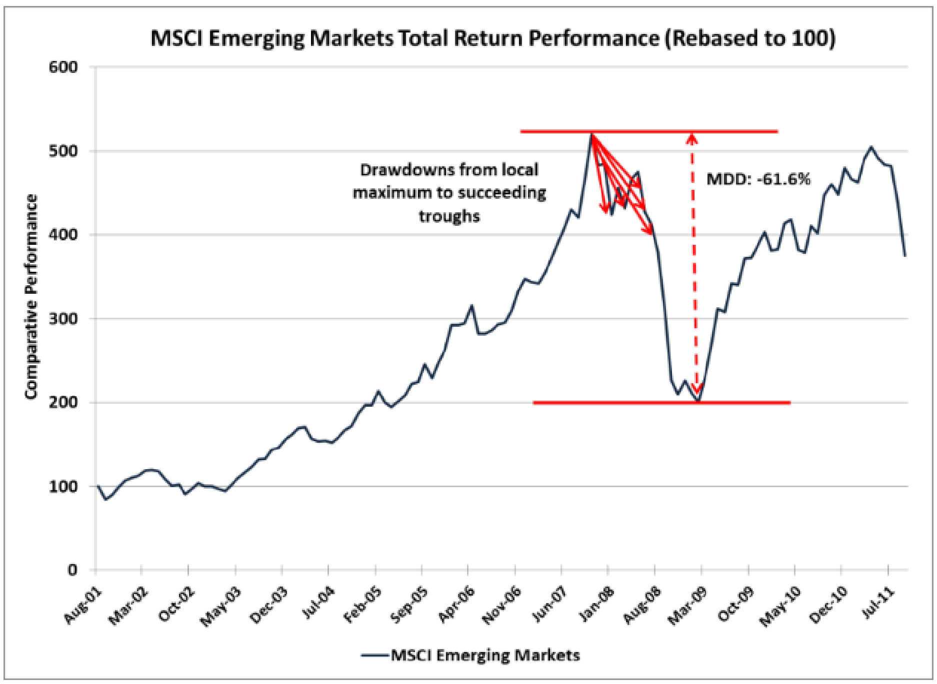

Figure 6: Drawdowns and Maximum Drawdown (MDD) for the MSCI Emerging Markets Index based on monthly total return data from September 2001 to September 2011.

In Figure 6, maximum drawdown (MDD) and local drawdown are displayed for the MSCI Emerging Markets Index for monthly returns from September 2001 to September 2011.[13] While there maybe be multiple drawdowns from a "local", or specific time horizon for which the investment is analyzed, there is only one maximum drawdown.In this figure maximum drawdown is -61.6% while the local drawdowns are significantly less and can be estimated to -30% or less for each red arrow representing each local drawdown.

4.5 Beta

Beta, or β is a parameter that measures systematic risk. It is derived from the capital asset pricing model (CAPM) which is a model that is used to relate the expected return from an asset to the risk of the return. [12] The risk in the return comprises of systematic risk, risk related to the return from the market as a whole and cannot be diversified away, and nonsystematic risk, risk that is unique to the asset and can be diversified away by choosing a large portfolio of different assets. CAPM follows

$$\text{Expected return on asset} = R_F + \beta(R_M - R_F)$$

where R_M is the return on the portfolio, R_F is the return on a risk-free investment, and β is the measure of the sensitivity of the stock's returns compared to the returns of the market. It is estimated from historical data as the slope obtained when the excess return on the asset over the risk-free rate is regressed against the excess return on he market over the risk-free rate. When $\beta = 0$, an asset's returns are not sensitive to returns from the market. If $\beta = 1$, the expected return on the asset is equal to the return on the market. Beta can even be used to measure a portfolio's volatility, or tendency of its value to change quickly and unpredictably.The portfolio's beta times the volatility of a well-diversified market index gives the portfolio's volatility. A good trading strategy hopes to have a $\beta > 1$, however, the trade off is taking on a higher amount of risk. By monitoring β, the portfolio's sensitivity and volatility to the market can be adjusted accordingly.

4.6 Stop Loss

There are many risk management methods regarding stop loss. One method, the stop-loss order, is used to exit or liquidate positions when the market goes against the portfolio's position. This order becomes a limit order when price reaches a certain specified level. This specified level is known as the stop-loss price because it is the price at which the investor would close their positions to "stop" the loss.

Some investors refer to the threshold at which to liquidate a long position should the stock begin to fall as a "stop-loss" rule. This threshold helps protect realized profits.[14] By implementing a "stop-loss" rule on the overall portfolio position, the trading strategy was able to limit how much loss was incurred over time.

4.7 Slippage

Slippage is the difference between model price and actual filled price due to timing differences and market data qualities. Generally, slippage is bad for algorithmic trading strategies as it causes deviation from the intended strategy. However, if caught at the right time, it could generate positive profit.

To calculate slippage and market impact (SMI), it follows

$$\text{SMI} = \frac{\sigma}{\sqrt{2\pi}}\sqrt{T}$$

where σ is the daily stock volatility and T is the time between when the algorithm receives the last trade price and the completion of the order execution.

To manage slippage, there are many considerations to take into account. One is to avoid trading during volatile time. Higher market volatility generally means higher slippage. Another method is to use a limit order including a stop-loss limit order, instead of a market order when possible. During market orders, there tends to be larger slippage due to the speed of the market. While limit orders, despite having uncertain fill times, tend to have smaller slippage.Additionally, higher liquidity in the market means lower slippage. High liquidity means that there are a lot of active market participants to accommodate the other side of the trades.[15]

5 Backtesting

After the initial construction of our quantitative strategy, we conducted a rigorous backtesting process to validate its effectiveness and robustness under different market conditions. The backtesting was structured as follows:

1. **In-Sample Period (Training):** We utilized a designated in-sample (IS) period to tune the parameters of our strategy. This period was selected to represent typical market conditions, enabling us to optimize the strategy's parameters based on historical data.

2. **Out-of-Sample Testing:** Following the in-sample optimization, we tested the strategy across three distinct out-of-sample (OOS) periods. These periods were chosen to evaluate the strategy's performance in scenarios that were not part of the training dataset, thereby providing an unbiased assessment of its predictive power and its ability to generalize across time.

3. **Stress Testing:** To assess the strategy's resilience under extreme market conditions, we conducted stress tests during two significant historical events:

   - **COVID-19 Pandemic Outbreak:** This stress test evaluated the strategy's performance during the market volatility induced by the COVID-19 pandemic, focusing on its capability to handle sudden and severe market downturns.
   - **2008 Financial Crisis:** A second stress test was performed to simulate the strategy's behavior during the 2008 financial crisis, providing insights into its effectiveness during periods of extreme financial stress and market disruptions.

This comprehensive backtesting framework allows us to ensure that our strategy is, not only theoretically sound, but also practical and robust across different market conditions and stress scenarios. We compare our strategy with the benchmark of S&P 500 and simple ADF pairs trading within technology sector.

5.1 In-Sample Backtest: 1/1/2017 - 1/1/2021

We conducted a detailed evaluation of our strategy's performance during the in-sample backtest period from January 1, 2017, to January 1, 2021. This timeframe was critical for fine-tuning our model's parameters and served as a robustness test against a variety of market conditions including periods of rapid growth, market corrections, and extended sideways movements.(Link to Backtest)

Table 1: In-Sample Backtest Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	0.657	0.614	-0.397
Return	66.20%	78.61%	-41.73%
Drawdown	25.40%	33.70%	62.10%

Our strategy achieved a Sharpe ratio of 0.657, which not only outperformed the S&P 500's Sharpe ratio of 0.614 but also significantly surpassed the ADF pairs trading within the technology sector, which had a negative Sharpe ratio of -0.397. This indicates that our model was exceptionally effective in dynamically adapting to different market conditions and capitalizing on short-term inefficiencies while maintaining a controlled level of risk.

Although the total return of 66.20% was slightly lower than the S&P 500's 78.61%, it far exceeded the ADF pairs trading strategy's performance, which suffered a substantial loss of 41.73%. This reflects our conservative approach during periods of high volatility and market uncertainty and underscores our strategy's capability to avoid large drawdowns, which were notably severe in the ADF pairs trading strategy at 62.10% compared to our strategy's maximum drawdown of 25.40%.

Figure 7 visually displays the equity curve of our strategy throughout the in-sample backtest period. The graph highlights the consistent growth of the portfolio and showcases how our strategy managed risk during periods of market stress. Key inflection points on the graph correspond to external economic events, illustrating our strategy's proactive adjustments to maintain stability and protect against downside risks. These visuals provide a clear indication of our strategy's superior performance and robustness compared to both the broad market and more specialized pairs trading strategies within the technology sector.

5.2 Out-of-Sample Backtest A: 1/1/2022 - 11/1/2022

The Out-of-Sample Backtest A was conducted from January 1, 2022, to November 1, 2022, a period marked by significant economic events, including varying market recoveries and instabilities due to ongoing global disruptions. This phase was instrumental in testing the adaptability and resilience of our strategy under real-world, unpredictable market conditions. Below, we detail the performance metrics of our strategy compared to the benchmarks. (Link to Backtest)

Our strategy achieved a Sharpe ratio of 1.148, significantly superior to the benchmark SPY's -0.749 and the ADF pairs trading's -0.654. This stark contrast highlights our strategy's robust risk management and ability to optimize returns despite significant market fluctuations. The positive Sharpe ratio not only

Figure 7: In-Sample Backtest Performance: 1/1/2017 - 1/1/2021

reflects effective asset selection but also timely rebalancing and risk diversification, which were crucial during the volatile trading periods encountered in 2022.

Furthermore, our strategy realized a return of 21.19%, contrasting sharply with SPY's decline of -18.37% and ADF pairs trading's -40.37%. This performance is attributed to our algorithm's capacity to identify and leverage upward trends within the technology sector, which outpaced the general market downturn experienced by more traditional sectors during this period.

The drawdown for our strategy was limited to 10.90%, substantially lower than SPY's 24.50% and the ADF pairs trading's 49.00%. This resilience is largely due to our strategic hedging practices and the implementation of advanced stop-loss techniques, which effectively minimized potential losses during sudden market drops.

Figure 8 illustrates the equity curve for our strategy throughout the testing period. The graph not only

Table 2: Out-of-Sample Backtest A Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	1.148	-0.749	-0.654
Return	21.19%	-18.37%	-40.37%
Drawdown	10.90%	24.50%	49.00%

demonstrates consistent growth but also shows how the strategy effectively navigated through periods of heightened market volatility. The dips and recoveries mapped in the curve correspond to external economic impacts, showcasing our strategy's proactive adjustments which capitalized on recovery phases promptly and efficiently.

This performance visualization and analysis underscore the sophisticated nature of our trading algorithms, which are designed to adapt dynamically to changing market conditions, safeguarding investments while capitalizing on growth opportunities.

5.3 Out-of-Sample Backtest B: 1/1/2016 - 1/1/2017

During the period from January 1, 2016, to January 1, 2017, we conducted our Out-of-Sample Backtest B to evaluate the performance of our strategy under market conditions distinct from the in-sample testing phase. This timeframe was characterized by significant volatility within the technology sector, which heavily influenced the dynamics of our pairs trading strategy, focused predominantly on this sector. (Link to Backtest)

Table 3: Out-of-Sample Backtest B Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	-0.292	0.83	-0.61
Return	-5.26%	13.60%	-15.33%
Drawdown	11.90%	8.90%	25.30%

The strategy's Sharpe ratio of -0.292 significantly underperformed compared to the S&P 500 benchmark (0.83) but did outperform the Benchmark Pairs ADF's -0.61. This illustrates the challenges posed by the heightened volatility within the technology sector during this period, which included regulatory uncertainties affecting major tech companies, shifting consumer preferences, and rapid technological advancements that disrupted traditional business models.

The return of -5.26% starkly contrasts with the S&P 500's positive return of 13.60%. This underperformance highlights the limitations of our static cointegration models, which struggled to adapt to the rapidly evolving conditions within the technology sector, particularly those driven by unexpected innovations and regulatory changes.

Moreover, the strategy's drawdown of 11.90%, although better than the Benchmark Pairs ADF's

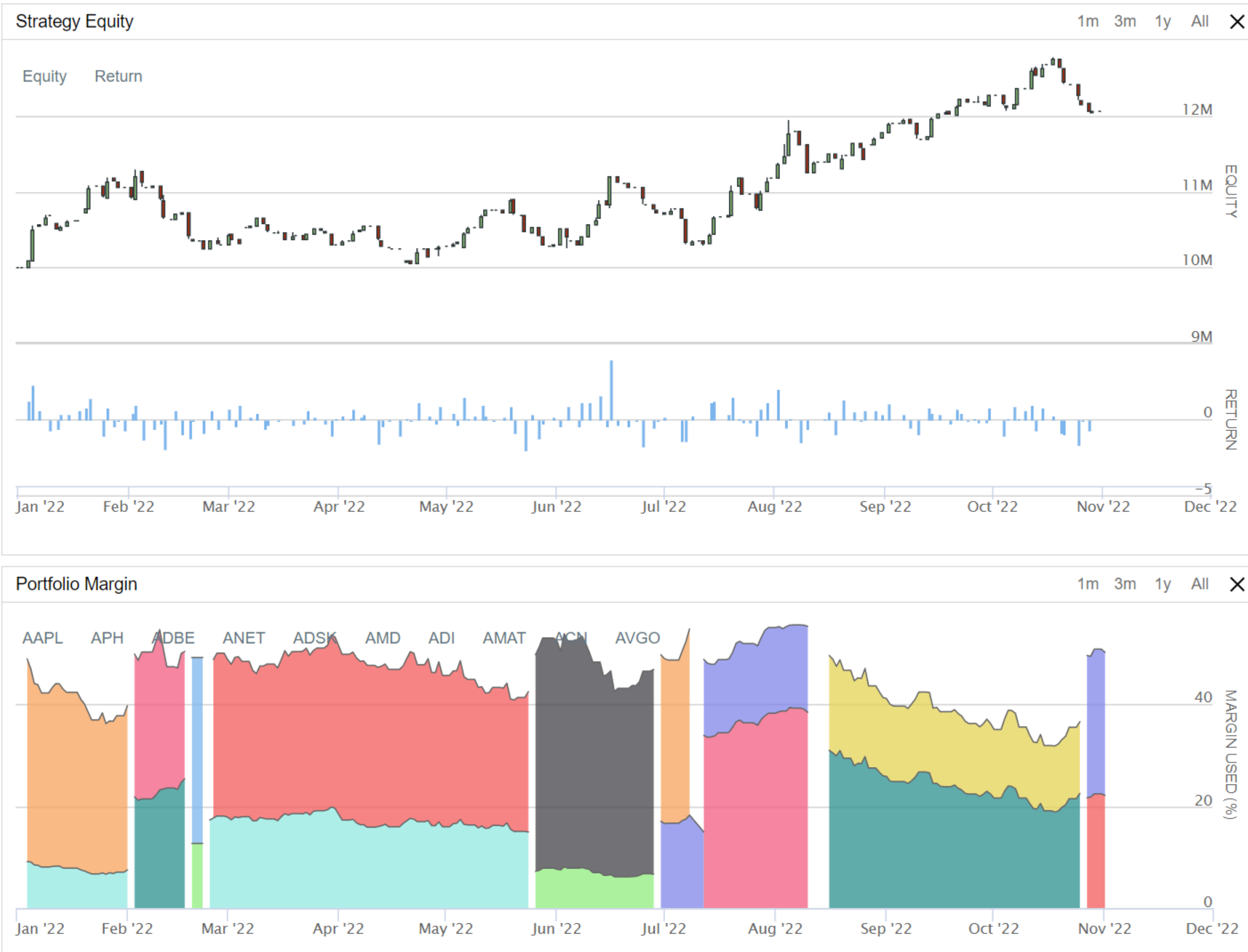

Figure 8: Out-of-Sample Backtest A Performance: 1/1/2022 - 11/1/2022

25.30%, was higher than the S&P 500's 8.90%, indicating a need for improved risk management protocols capable of responding more effectively to swift and unpredictable market changes.

Figure 9 visually depicts the equity curve of our strategy, highlighting periods of underperformance relative to the benchmarks. This graphical representation not only illustrates the fluctuations but also underscores the strategy's sensitivity to the rapid shifts occurring within the technology sector.

The insights gained from this backtest underscore the necessity for more dynamically responsive trading models that can swiftly adapt to sudden market changes. Future enhancements will focus on developing more agile and flexible risk management strategies that can better accommodate the volatility typical of the technology sector, aiming to minimize drawdowns and improve overall return profiles.

Figure 9: Out-of-Sample Backtest B Performance: 1/1/2016 - 1/1/2017

5.4 Out-of-Sample Backtest C: 10/1/2023 - 4/1/2024

The Out-of-Sample Backtest C, covering the period from October 1, 2023, to April 1, 2024, presented unique challenges that resulted in the strategy underperforming relative to the benchmark. This phase allowed us to rigorously test the adaptability and resilience of our strategy under real-world, unpredictable market conditions, which proved particularly volatile. (Link to Backtest)

The strategy's Sharpe ratio of -0.42 significantly lagged behind the S&P 500 benchmark, which achieved a robust 3.08. This discrepancy underscores the challenges our strategy faced during this testing period. Similarly, the negative return of -2.09% contrasts starkly with the benchmark's substantial gain of 23.54%. Furthermore, our strategy's drawdown of 18.60% was substantially higher than the benchmark's 5.80%, indicating less efficient management of peak-to-trough declines during the backtest period.

Table 4: Out-of-Sample Backtest C Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	-0.42	3.08	-1.296
Return	-2.09%	23.54%	-18.40%
Drawdown	18.60%	5.80%	24.20%

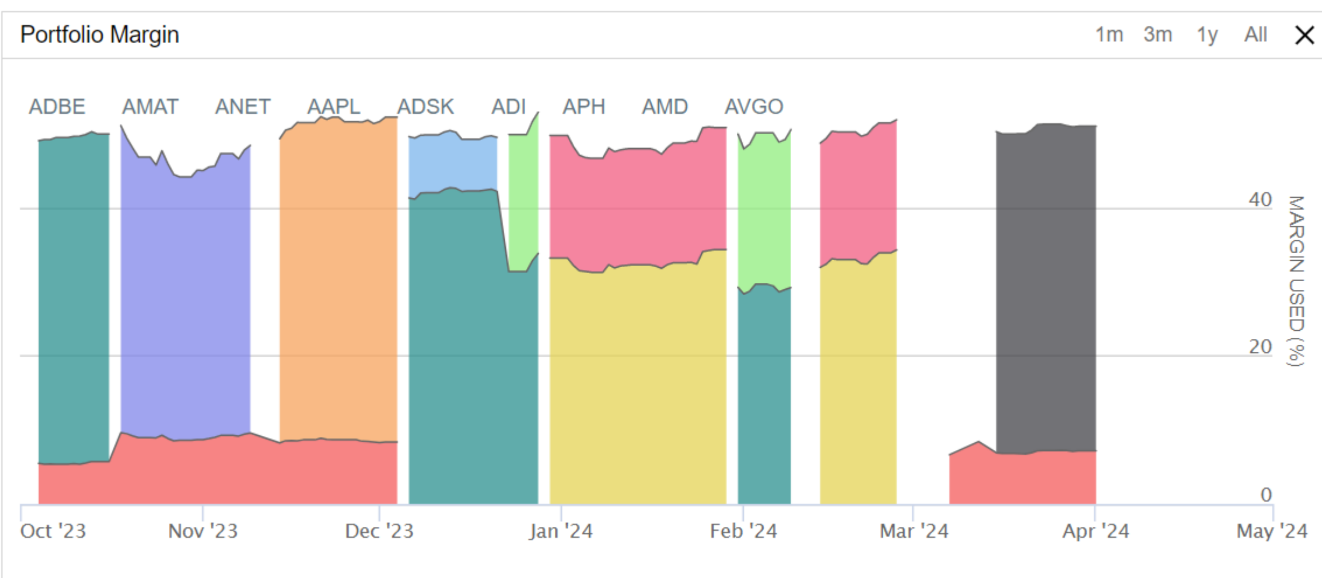

Figure 10: Out-of-Sample Backtest C Performance: 10/1/2023 - 4/1/2024

This underperformance can be attributed to several unanticipated events within the technology sector, a primary focus of our pairs trading strategy. Increased volatility and significant pair divergence disrupted the typical cointegration relationships our strategy relies on, leading to adverse trading outcomes.

In response, we plan to enhance our strategy by integrating more robust risk management measures and leveraging advanced analytical tools to better anticipate and adapt to such market anomalies. These strategic enhancements, along with an increased investment in computational resources, aim to refine our strategy's performance in future testing phases, ensuring it can withstand and capitalize on the dynamics of volatile market conditions.

5.5 Stress Testing A: 3/1/2020 - 4/1/2020

During March 2020, as global markets faced unprecedented volatility triggered by the COVID-19 pandemic, we conducted Stress Testing A to assess the resilience of our trading strategy under extreme market conditions. The results, detailed below, provide insights into how effectively the strategy managed the crisis compared to the benchmarks. (Link to Backtest)

Table 5: Stress Testing A Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	-1.796	-1.065	-1.345
Return	-10.54%	-19.99%	-24.75%
Drawdown	10.90%	28.30%	30.00%

Despite the overall negative Sharpe ratios, which reflect losses during this turbulent period, our strategy exhibited greater resilience (-1.796) compared to the SPY benchmark (-1.065) and the ADF pairs trading benchmark (-1.345). This resilience demonstrates the effectiveness of our strategy's risk management under rapidly deteriorating market conditions.

Our strategy's return of -10.54%, though negative, significantly outperformed the SPY benchmark's -19.99% and was less severe than the ADF pairs trading benchmark's -24.75%. This comparative outperformance underscores our strategy's robust mechanisms for limiting losses during severe market downturns, where traditional and specialized investment approaches faced more substantial declines.

The strategy's drawdown was contained to 10.90%, considerably lower than SPY's 28.30% and the ADF benchmark's 30.00%. This highlights our strategy's capability to protect investments by minimizing potential losses during the peak impacts of the crisis.

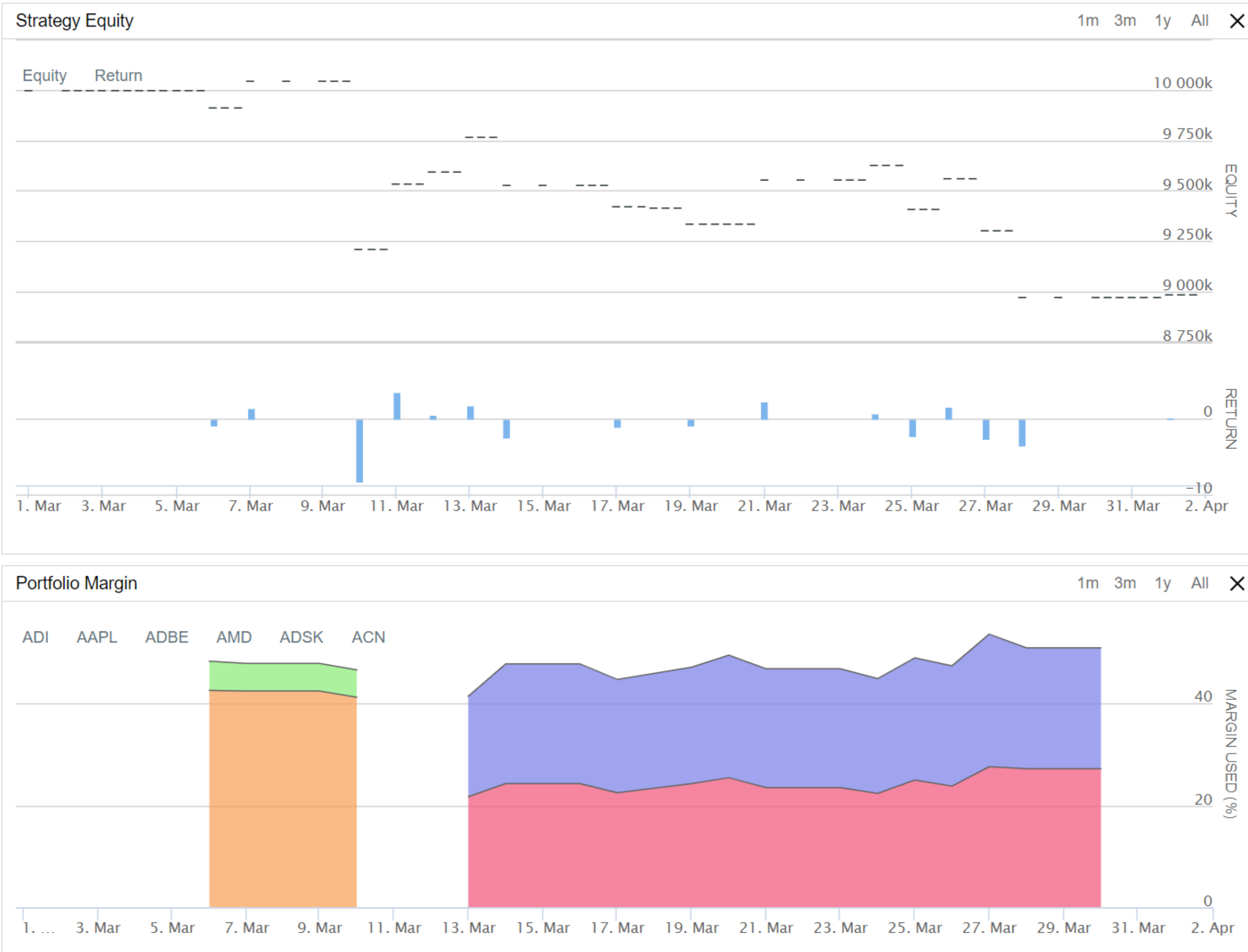

Figure 11: Stress Testing A Performance: 3/1/2020 - 4/1/2020

Figure 11 illustrates the equity trajectory of our strategy throughout the crisis. The graph clearly shows how our strategy managed to mitigate risks and preserve capital more effectively, especially during the most intense periods of market stress.

The outcomes from this stress test underline the importance of adaptive risk management techniques that can dynamically respond to sudden and severe market shifts. Our strategy's performance during March 2020 is indicative of a successful application of such techniques, which helped to moderate losses and stabilize the portfolio as market conditions rapidly changed.

5.6 Stress Testing B: 9/1/2008 - 1/1/2009

The second stress test, Stress Testing B, was conducted to evaluate the performance of our trading strategy during the 2008 financial crisis' period characterized by extreme market turmoil and significant economic downturns. (Link to Backtest)

Table 6: Stress Testing B Performance Comparison

Metric	Strategy	Benchmark SPY	Benchmark Pairs ADF
Sharpe Ratio	1.097	-0.89	-0.927
Return	9.91%	-28.27%	-46.07%
Drawdown	15.40%	40.10%	58.40%

Our strategy achieved a Sharpe ratio of 1.097, significantly outperforming both the S&P 500 benchmark and the ADF pairs trading benchmark during one of the most challenging periods in recent financial history. This performance underscores our strategy's superior risk management capabilities and its proficiency in capitalizing on volatility within the technology sector.

With a positive return of 9.91

Additionally, our strategy's maximum drawdown of 15.40% was considerably lower than the benchmarks', which recorded 40.10% and 58.40% respectively. This demonstrates the resilience of our trading approach during extreme market conditions, effectively protecting the investment from the depths of the market's falls.

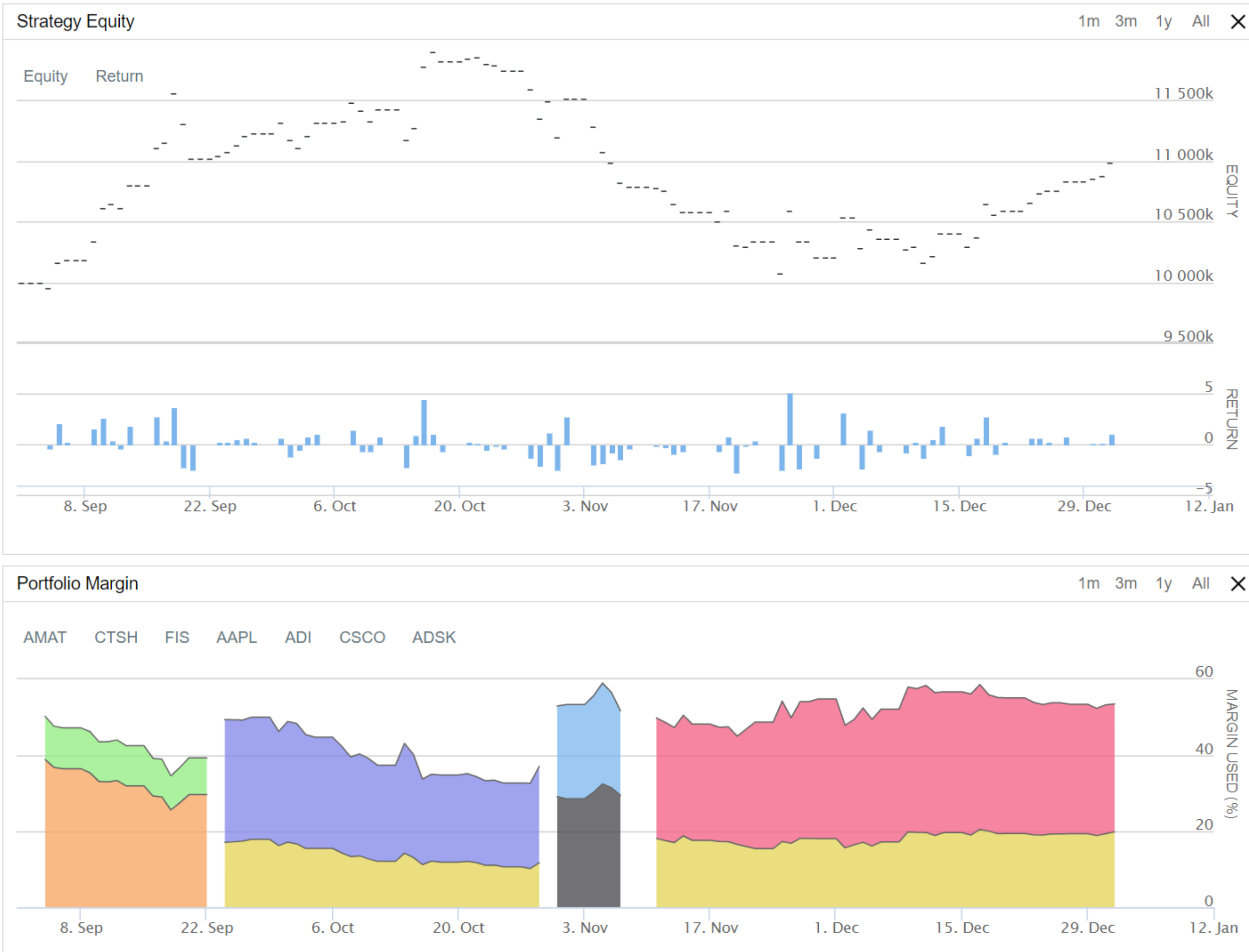

Figure 12: Stress Testing B Performance: 9/1/2008 - 1/1/2009

The accompanying figure 12 visually illustrates the equity curve of our strategy throughout the financial crisis. The graph not only shows the strategy's stability but also its capability to seize opportunities during recovery phases, effectively managing risks and exploiting market inefficiencies even under severe stress.

The successful outcomes of Stress Testing B validate the importance of adaptable, dynamic trading strategies that can respond swiftly to extraordinary market conditions. The impressive performance of our pairs trading within the technology sector during the 2008 crisis serves as a testament to our nuanced understanding of sector-specific dynamics and our ability to leverage discrepancies and convergences between closely related stocks.

5.7 Live Trading: 4/23/2024 - 4/29/2024

We initiated our live trading phase with a starting capital of $10,000,000 on April 23, 2024. By 10:34 AM on April 29, 2024, our strategy demonstrated its real-world applicability by achieving a return of 0.48%. This growth resulted in an increase in our holdings to $10,047,768.02, reinforcing the effectiveness of our strategy under live market conditions.

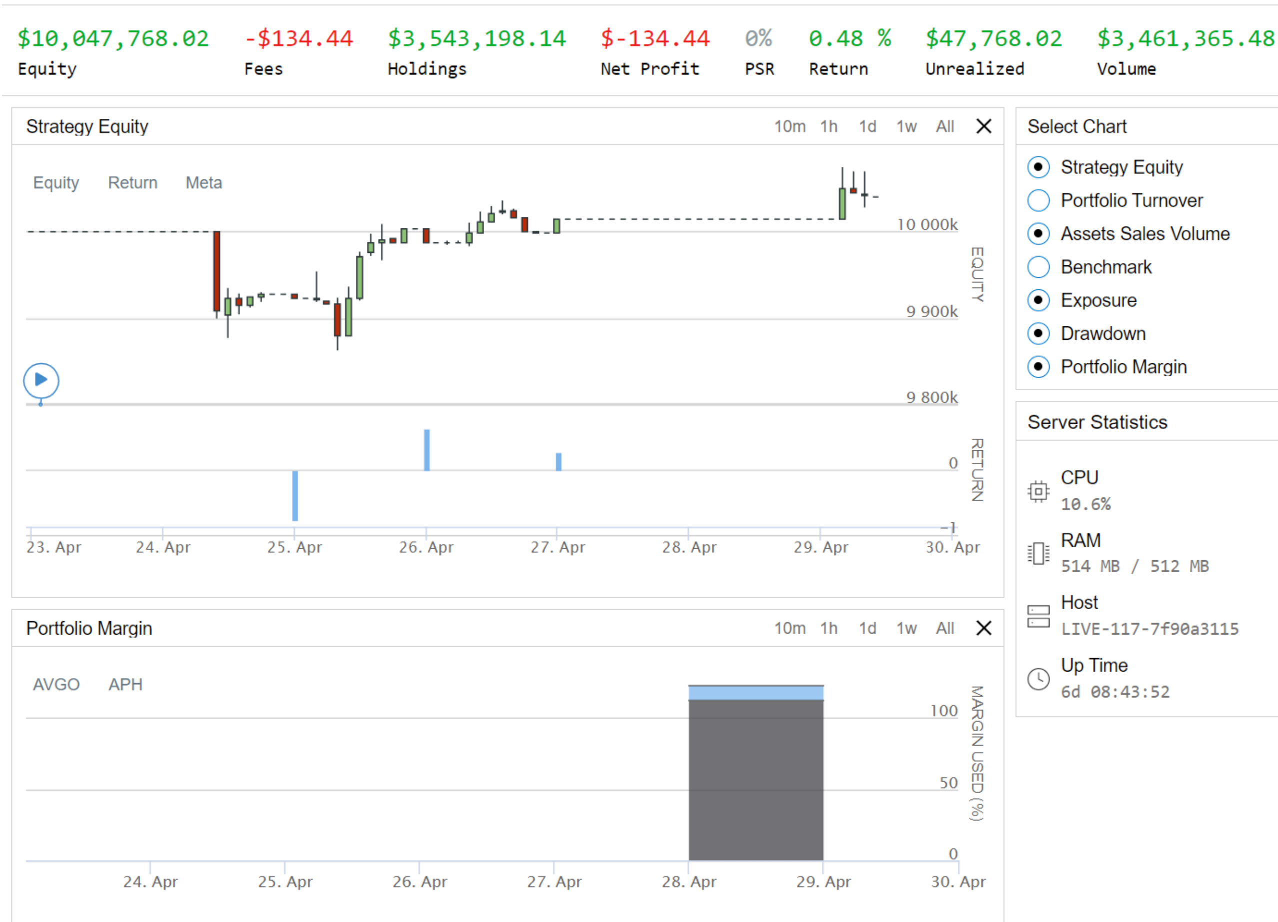

Figure 13: Live Trading Performance from 4/23/2024 to 4/29/2024.

This brief period of live trading serves as a preliminary validation of our approach, suggesting that the strategies developed during back-testing can be successfully transitioned into operational trading. We continue to monitor and optimize our strategy to adapt to ongoing market conditions, aiming to maximize returns while managing risks efficiently.

6 Future Improvements

The trading strategy developed has a sound foundation. However, should the project be continued in the future there are a few components that would be worthwhile to improve upon. The optimization of sentiment analysis, the implementation of dynamic parameter tuning, and the incorporation of diversification through multiple trading pairs are the top concerns for future development. Additionally, increasing compute power to handle the calculations required for the strategy would be crucial for future success.

6.1 Sentiment Analysis

For future improvements, a trading strategy could incorporate sentiment analysis. Sentiment analysis is the process of analyzing large volumes of text to determine whether it expresses a positive, a negative, or a neutral sentiment.[16] With the growing trend of social media and the prevalence of the internet, public opinion is widely accessible about nearly any topic. By learning about the public's thoughts and opinions about companies and the overall economy, market sentiment can be understood.

To learn about market and corporate sentiment, the algorithm could access news from Tiingo, a financial data platform, and processes it using the Natural Language Toolkit (NLTK) Python library. The NLTK library analyzes texts based on a set of predefined rules and heuristics to determine its sentiment. This sentiment analysis approach is called rule-based sentiment analysis because it uses groups of words, also known as lexicons, to classify other words as either having a positive, or good feeling, or a negative, or bad feeling. The library can scan for the lexicons, and then it totals up the sentiment score based on the volume of words and the sentiment score of each category as determined by the preexisting rules. The other scores a word could receive is neutral or compound. Thus, for the NLTK sentiment intensity analyzer, four polarity scores, or grades, can be received: positive, negative, neutral, or compound. Once the algorithm determines the average polarity score for a stock, the ones with a higher polarity score are selected as this means the stock has a positive sentiment.A positive sentiment might indicate growth or success in profits for that company in the future.

This biggest hurdle for implementing sentiment analysis is processing time and computer processing power.When a backtest runs on an algorithm, one must be mindful how often sentiment analysis will be called. Since the designed trading strategy incorporated an adaptive pairs trading strategy, sentiment analysis needed to be run for every new stock during the pairs transition.

6.2 Dynamic Parameter Tuning

Significant enhancements to our stress testing approach have been achieved by adjusting the `self.lookbackUniverse` parameter from 90 days to 120 days. This modification has led to substantial improvements in the performance metrics during stress scenarios, as detailed below.

6.2.1 COVID-19 Pandemic Scenario (March 2020):

With the extension of the `self.lookbackUniverse` to 120 days, our strategy demonstrated a robust response to the market volatility induced by the COVID-19 pandemic: (Link to Backtest)

- Sharpe ratio improved from -1.796 to 4.146,
- Return increased from -10.54% to 9.50%,
- Drawdown decreased from 10.90% to 5.80%.

Table 7: Comparison of Stress Testing A Performance Before and After Parameter Adjustment

Metric	Before Adjustment	After Adjustment
Sharpe Ratio	-1.796	4.146
Return	-10.54%	9.50%
Drawdown	10.90%	5.80%

6.2.2 2008 Financial Crisis Scenario (September 2008 - January 2009):

The adjustments during the 2008 financial crisis period resulted in even more pronounced improvements: (Link to Backtest)

- Sharpe ratio soared from 1.097 to 4.655,
- Return escalated from 9.91% to 43.96%,
- Drawdown shrank from 15.40% to 8.30%.

Table 8: Comparison of Stress Testing B Performance Before and After Parameter Adjustment

Metric	Before Adjustment	After Adjustment
Sharpe Ratio	1.097	4.655
Return	9.91%	43.96%
Drawdown	15.40%	8.30%

The dynamic tuning of the self.lookbackUniverse parameter underscores the importance of adaptable parameters, affirming their pivotal role in enhancing the resilience and efficacy of our trading strategy under conditions of extreme market duress. This strategic adjustment has underscored our model's capability to better account for a wider array of historical market data, enabling a more precise and steadfast evaluation of market trends and risk factors. Subsequently, these enhancements have been instrumental in the substantial risk management and performance improvements realized during the significant market stresses evaluated.

6.3 Diversification Through Multiple Pairs

Our initial results, based on a single trading pair due to computational limits, indicate promising potential for risk diversification. However, to better hedge market exposure, we implemented our strategy across three different pairs for a shorter period. This diversification is expected to mitigate unsystematic risk more effectively.

Table 9 presents the performance metrics for the strategy implemented on multiple pairs form 9/1/2008 to 1/1/2009:

Table 9: Performance Metrics: Single Pair vs. Multiple Pairs

Metric	Single Pair	Multiple Pairs
Sharpe Ratio	1.097	1.882
Return	9.91%	13.60%
Drawdown	15.40%	8.00%

As indicated by the Sharpe ratio increase from 1.097 to 1.882, and a return boost from 9.91% to 13.60%, investing in multiple pairs has shown to reduce risk while enhancing returns. Moreover, the strategy's drawdown improved from 15.40% to 8.00%, demonstrating its effectiveness in protecting the portfolio against deep losses.

Our roadmap for scaling this strategy includes circumventing computational constraints and expanding our investment across an even wider array of pairs. This will enable a more robust risk distribution and is anticipated to drive higher risk-adjusted returns for our portfolio.

Future iterations of our strategy, with enhanced computational capabilities, will allow for the application of this diversified approach over longer time frames and across varied market conditions.

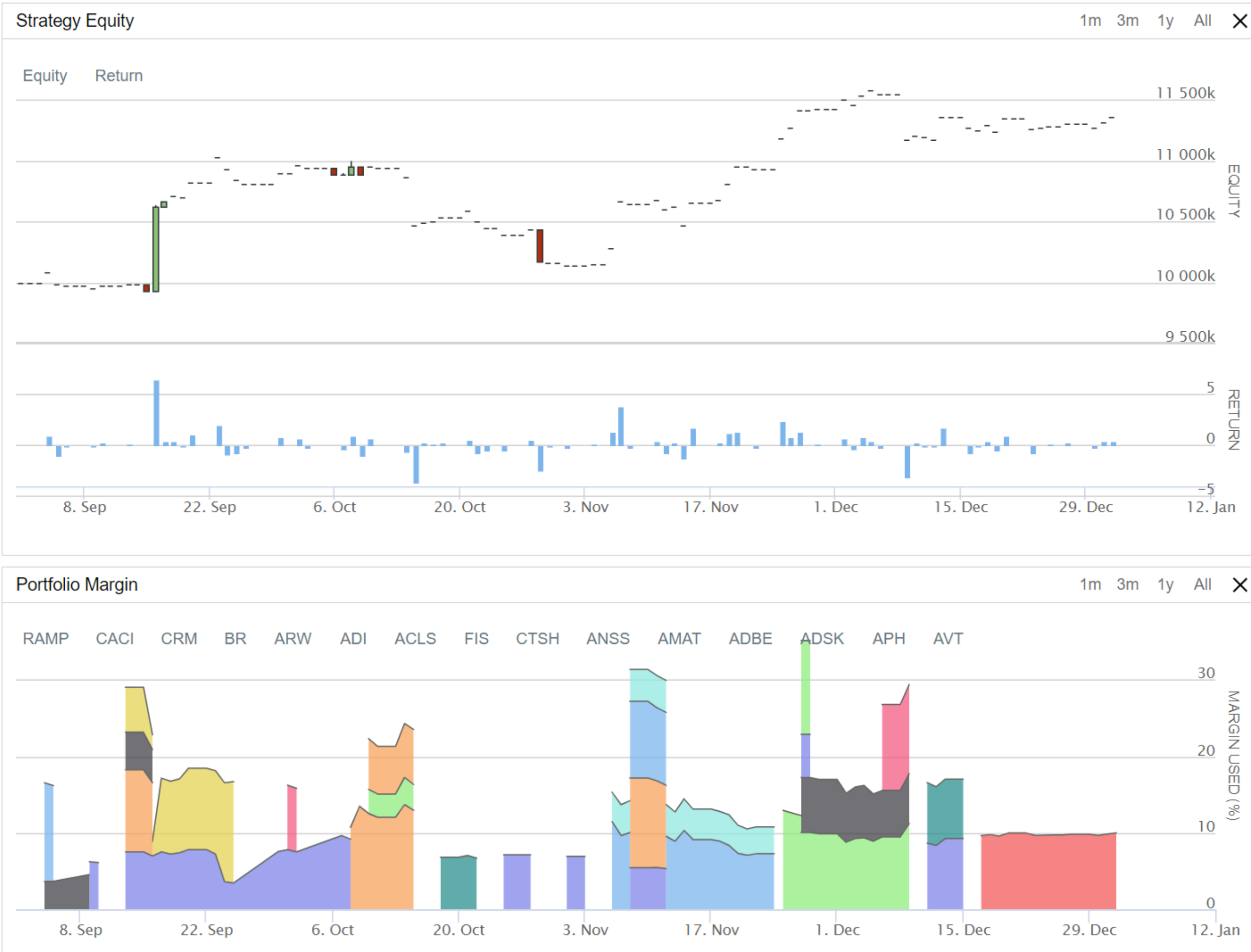

Figure 14: Comparative Performance of Single vs. Multiple Pairs Strategy

References

[1] "A bayesian approach to stock trading." http://hdl.handle.net/20.500.12648/8623, 2021.

[2] "Pairs trading: Performance of a relative-value arbitrage rule." http://stat.wharton.upenn.edu/~steele/Courses/434/434Context/PairsTrading/PairsTradingGGR.pdf, 2006.

[3] "Bayesian analysis of contingent claim model error." https://www.sciencedirect.com/science/article/abs/pii/S0304407699000202?via%3Dihub, 2000.

[4] "Optimal scaling of random-walk metropolis algorithms on general target distributions." https://www.sciencedirect.com/science/article/pii/S0304414919306982#b56, 2020.

[5] B. Miller, "How 'survivorship bias' can cause you to make mistakes." https://www.bbc.com/worklife/article/20200827-how-survivorship-bias-can-cause-you-to-make-mistakes.

[6] I. Ferre, "Tech stocks: What wall street experts recommend in a bear market." https://finance.yahoo.com/news/tech-stocks-what-wall-street-experts-recommend-in-a-bear-market-200929903.html.

[7] N. Huck and K. Afawubo, "Pairs trading and selection methods: Is cointegration superior?," *Applied Economics Vol. 47 Issue 6, p599-613.*, 2015.

[8] B. Wire, "Amerisourcebergen becomes cencora, in alignment with the companyâs growing global footprint and central role in pharmaceutical access and care," *Yahoo Finance*, 2023.

[9] S. Prabhakaran, "Augmented dickey fuller test (adf test) - must read guide." https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test.

[10] QuantStart, "Johansen test for cointegrating time series analysis in r." https://sakai.duke.edu/access/content/group/2a01ffab-810c-4caa-875b-b9bf50039564/Lecture8.5/Johansen%20Test%20for%20Cointegrating%20Time%20Series%20Analysis%20in%20R%20_%20QuantStart.pdf.

[11] A. F. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management*. Princeton University Press, 2015.

[12] J. C. Hull, *Options, Futures, and Other Derivatives*. Pearson Education, Inc., 2015.

[13] P. Ramani, "Sculpting investment portfolios: Maximum drawdown and optimal portfolio strategy." https://blogs.cfainstitute.org/insideinvesting/2013/02/12/sculpting-investment-portfolios-maximum-drawdown-and-optimal-portfolio-strategy/.

[14] Z. Kakushadze and J. A. Serur, "151 trading strategies," 2018.

[15] I. Analyst, "What is slippage and how do you avoid it in trading?." https://www.ig.com/en/trading-strategies/what-is-slippage-and-how-do-you-avoid-it-in-trading--190319#:~:text=Trading%20in%20markets%20with%20low%20volatility%20and%20high,to%20accommodate%20the%20other%20side%20of%20your%20trades.

[16] IBM, "What is sentiment analysis?." https://www.ibm.com/topics/sentiment-analysis.