DUKE UNIVERSITY
MATH 585: INTRODUCTION TO ALGORITHMIC TRADING
FINAL REPORT
INSTRUCTOR: PROF. DAVID YE

Pairs Trading with Sparse Portfolios

Group 6
Lintai Hou
Jiazheng Sun
Yuemin Tang
Sijia Wang
David Zhao

Date: December 12, 2023

Abstract

This paper introduces an approach to pairs trading utilizing sparse portfolios, using different methods for feature selection and portfolio construction. Pairs trading, a popular market-neutral strategy, involves exploiting short-term misalignments between two correlated assets. Pairs trading involving two assets can naturally be extended to a portfolio of assets, but typically these portfolios are dense. Sparse portfolios enhance this strategy by automatically selecting a subset of the most relevant features, thereby improving interpretability, reducing dimensionality, and potentially enhancing trading performance by creating more meaningful statistical arbitrage opportunities.

In this paper, we will discuss 2 different methods of forming spare portfolios using machine learning and greedy search. We will discuss a new method of testing for cointegration when there are more than 2 time series, and how this new method gives the appropriate hedge ratios for the assets in our portfolio. We will discuss and compare several methods of determining trading signals using a stochastic approach, as well as risk management considerations for our trading strategy including using VaR (value at risk) and sentiment analysis. We will discuss how different parameters for our strategy are tuned, and the efficacy of our strategy in different trading periods.

Contents

List of Figures

List of Tables

1 Introduction

Pairs trading is considered widely as one of the most fundamental examples of statistical arbitrage. It is a market-neutral trading strategy that relies on the statistical relationship between two assets. The primary objective is to profit from the mean-reverting behavior of the spread between the two assets. The strategy assumes that assets deviate temporarily from their intrinsic values and that these deviations will eventually correct. To identify a mean-reverting (or co-integrated) pair, statistical tests need to be performed. While suitable pairs for pairs trading are usually highly correlated, correlation does not imply co-integration. Correlation is a measure of historical data, whereas co-integration is a measure of whether or not two non-stationary variables are stationary when combined in the future. A common statistical test used to measure cointegration is the Augmented Dicky-Fuller Test (ADF test). It is a statistical test with the null hypothesis that the time series has a unit root (meaning it is non-stationary), and an alternative hypothesis that that time series does not have a unit root, and is stationary/mean-reverting.

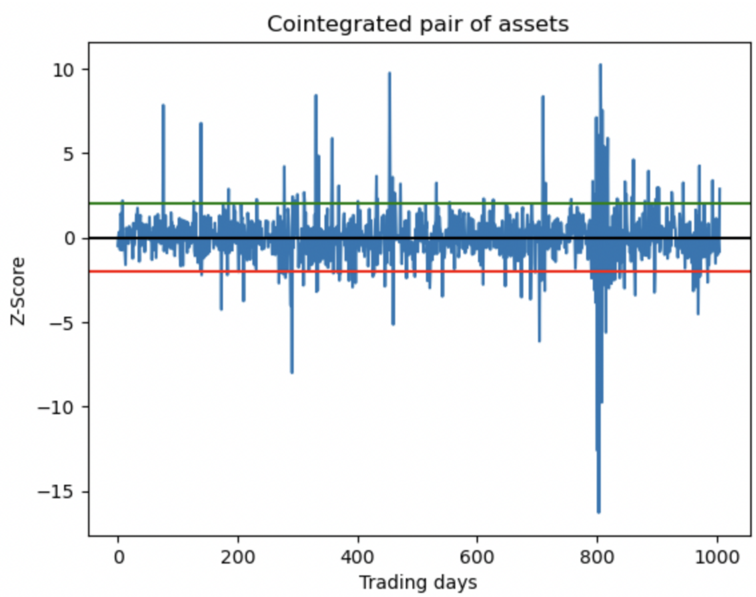


Figure 1: Mean Reverting Pair of Assets

As seen in the figure above, trading opportunities arise when the spread of the residuals of the assets deviate by a large amount. If the spread crosses the green or red band, we can long the undervalued asset, and short the overvalued asset because we expect the spread to return to 0.

A natural extension of pairs trading is to use the same mean-reverting nature to trade a portfolio of assets. While mean reversion is easy to identify in uni-variate time series, finding portfolios with $n > 3$ assets is a more challenging task. Traditional methods for identifying mean portfolios involve analyzing the entire time series of all the assets in the portfolio. These portfolios are called "dense". Dense portfolios are sub-optimal for a few reasons. Exploiting statistical arbitrage opportunities for dense portfolios involve considerable transaction costs. This also impacts the interpretability and the significance of the structural relationships that these portfolios highlights, and optimally mean reverting portfolios often behave like noise and don't form meaningful statistical arbitrage opportunities. Sparse portfolios are defined to be optimally mean-reverting portfolios with fewer (generally $n < 8$) assets that aim to improve transactions costs, interpretability. Sparse portfolios vary greater in price range, so market in-efficiencies that are found are more significant. In this paper, we will discuss the theories and implementations behind different ways to form these sparse portfolios. The general problem is as follows.

Suppose that S_{ti} is the value at time t of an asset S_i with $i = 1, ..., n$ and $t = 1, ..., m$. We form portfolios P_t of these assets with coefficients x_i with each asset following a Geometric Brownian Motion. If we assume that our portfolio follows an Ornstein-Uhlenbeck process (a Brownian Motion with a mean-reverting nature). Then the portfolio is given by

$$dP_t = \lambda(\overline{P} - P_t)dt + \sigma dZ_t, P_t = \sum_{i=1}^{n} x_i S_{ti}. \tag{1}$$

Where Z_t is standard Brownian Motion. We would like to maximize the mean reversion coefficient λ of P_t by adjusted weights x_i such that the number of positive (nonzero) coefficients is less than some k for $k > 0$. The following section will describe 2 different methods for sparse portfolio formation. The first uses a supervised machine learning method to estimate the relationship between variables and to make predictions. The second, a simply search on which assets to include in our portfolio. In checking for co-integration, the ADF test cannot handle more than 2 time series, so we will need to use other methods like the Johansen test.

There are some clear risks/downfall to consider in pairs trading. The first is that assets in our portfolios may suddenly stop being co-integrated, and we do not have a mean-reverting process anymore. We will discuss how we mitigate these risks later on in the paper using rigorous value-at-risk (VaR) calculations , maximum drawdown, and sentiment analysis. These measurements are used to help generate new entry/exit thresholds for our portfolios. The second downfall is that trades can only be entered when the spread deviates a certain amount. This means that we may potentially be waiting to enter trades, affecting our profitability. We will discuss how we implement other trading strategies with appropriate capital allocation to increase the frequency of our trading, and how we use multiple sparse portfolios to diversify our holdings, and to increase the frequency of trading. For this project, our universe of stocks is the S&P 500. It consists of the 500 leading publicly traded companies in the United States, where each stock in it is weighted by market capitalization. Our strategy can be easily extended to other financial products like commodities, futures, and currencies so long as the resulting portfolios are sparse and mean-reverting in nature.

The paper is organized as follows. In Section 2, we discuss the 2 different methods used to form sparse portfolios. The third section discusses using different trading signal based on a stochastic approach. The fourth section discusses our implementation of risk management and capital using a combination of different signals. Finally, we discuss our backtesting results with respect to return/Sharpe ratio/maximum drawdown for different testing periods, and potential improvements on our trading strategy.

2 Background I: Sparse Mean Reversion Portfolios

2.1 Testing for Cointegration

As mentioned previously, one of the biggest drawbacks of the ADF test is that it is only able to be used to check colinearity for 2 separate time series. For our portfolios created with $n > 3$ portfolios, we need another method. The Johansen test allows us to check whether three or more time series are cointegrated. If we have cointegration, we will form a stationary series by taking a linear combination of the underlying assets in the portfolios.

In the test, we form a general vector autoregressive model where each value is vector valued and matrices are used as coefficients.

$$x_t = \mu + A_1 x_{t-1} + A_2 x_{t-2} + ... + A_p x_{t-p} + w_t, \tag{2}$$

where μ is the vector valued mean of the series, and A_is are the coefficient matrices for each lag in our portfolio, and w_t is standard brownian motion with mean 0.

We form a VECM (vector error correction model) by differentiating the series given above

$$\Delta x_t = \mu + A x_{t-1} + \Gamma_1 \Delta x_{t-1} + ... + \Gamma_p \Delta x_{t-p} + w_t, \tag{3}$$

where Δ is the Laplacian, A is the coefficient matrix for the first lag, and Γ_i are matrices for the following lags. When there is no cointegration, A = 0. If A is nonzero, an eigenvalue decomposition is carried out on A. The rank r of matrix A takes values from $0, 1, ...n-1$, where n is the number of assets in our portfolio. A rank of 0 means that there is no cointegration, whereas a rank $r >= 1$ means that there is cointegration with r assets in our portfolio. If we do have cointegration, we can look at the eigenvectors resulting from our eigenvalue decomposition, and the eigenvector corresponding to the largest eigenvalue gives the coefficients for each portfolio to form a stationary time series by taking a linear combination of them.

$$MeanRevertingPortfolio = v_0 * X_1 + v_1 * X_2 + ... + v_n * X_n, \tag{4}$$

where v is our eigenvector corresponding to the largest eigenvalue of matrix A, and X_i is each stock in our portfolio.

This is a key difference from the ADF test since the ADF test does not give these coefficients. However, we can make our Johansen test more rigorous by using the ADF test on the univariate time series from the equation above to test for cointegration as well. This procedure outlined will be used to test for cointegration on sparse portfolios formed with LASSO regression and greedy search.

2.2 LASSO Regression

LASSO (least absolute shrinkage and selection operator) regression is a supervised machine learning algorithm. The primary purpose of LASSO regression is to prevent overfitting by encouraging sparsity in the model. In this algorithm, data values are shrunk towards a central point. It performs an L1 regularization that adds a penalty equal to the absolute value of the magnitude of the coefficients. Because we are using the L1 norm for regularization rather than the L2 (euclidean) norm, some coefficients can go to zero, resulting in less assets selected. The goal of the algorithm is to minimize

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{5}$$

where y_i is the predicted value, and x_i are the actual values. The tuning parameter λ controls the strength of the L1 penalty. The greater the value of λ, the more sparse the model. LASSO regression is a well suited machine learning algorithm because we have a large number of features, and it encourages sparsity.

Using the time period from the training set (January 1st, 2017 - January 1st, 2021), we implemented a graphical LASSO (from which we can visualize an undirected graph with spare connections for higher interpretability) on daily lognormalized closing prices on stocks in the S&P 500.

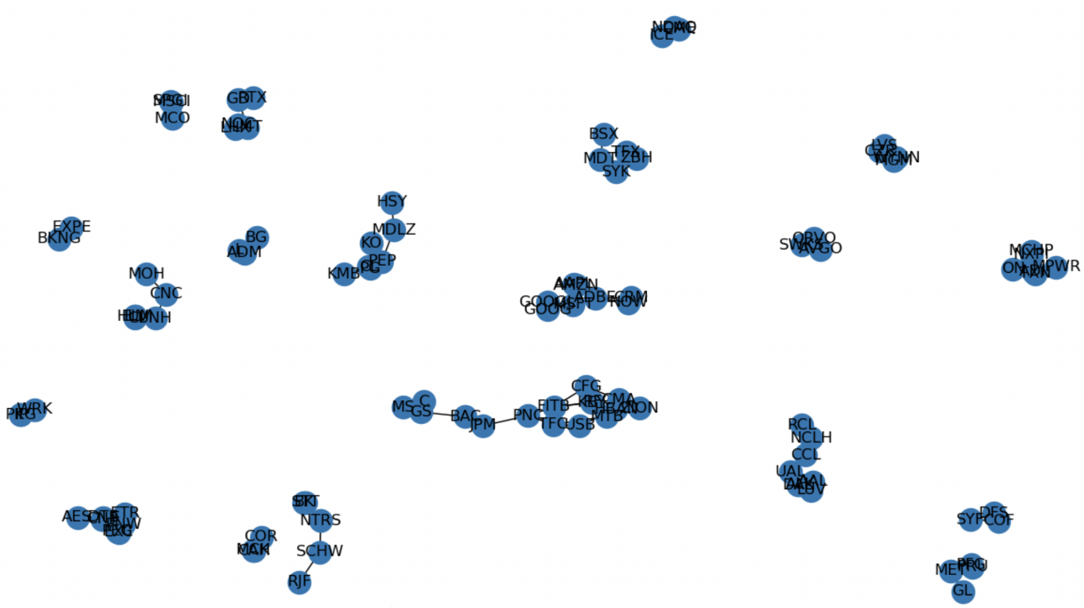
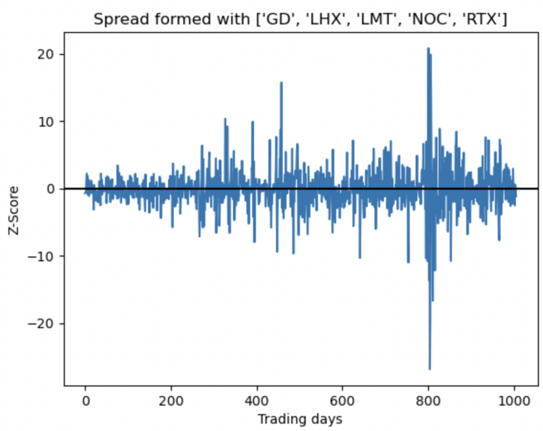
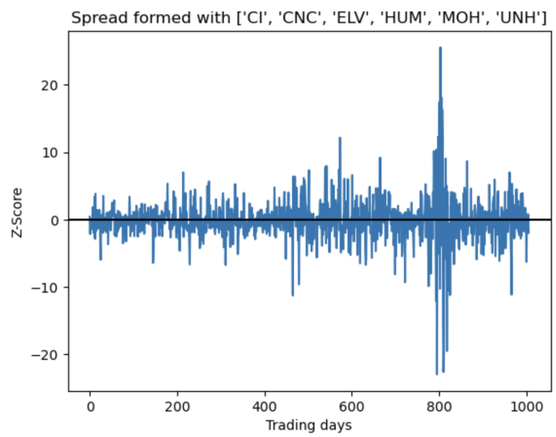


Figure 2: Sparse Portfolios Formed with LASSO Regression

Using the portfolios generated, we use the Johansen test as referenced in the previous section to test for cointegration. If we have cointegration, we have a suitable sparse portfolio to trade, with the appropriate hedge ratios (the eigenvector corresponding to the maximium eigenvalue) for each portfolio given from the Johansen test.



(a) Sparse Portfolio 1



(b) Sparse Portfolio 2

Figure 3: Sparse Portfolio Examples

2.3 Greedy Search

As suggested by [6], we can perform greedy search to obtain the approximate solution to the problem.

Let I_k be the index set of the portfolio, i.e., I_k is the support set of vector x:

$$I_k = \{i \in [1, n] : x_i \neq 0\} \tag{6}$$

We can recursively compute I_k such that it is subject to the carnality constraint $|I_k| < k$. At step 1, we compute

$$I_1 = \text{argmax}_{i \in [1,n]} A_{ii}/B_{ii}. \tag{7}$$

Given an index set I_k at step k, finding approximate solution vector x as well as the extremal eigenvalue can be solved as a generalized eigenvalue problem of size k:

$$x_k = \text{argmax}_{x \in \mathbf{R}^n : x_{I_k^c} = 0} \frac{x^T A x}{x^T B x} \tag{8}$$

$$pred_k = \max_{x \in \mathbf{R}^n : x_{I_k^c} = 0} \frac{x^T A x}{x^T B x}. \tag{9}$$

Then, at step $k+1$, we add one additional index i_{k+1} that produces the highest predictability increase to the existing index set I_k:

$$I_{k+1} = I_k \cup i_{k+1} \tag{10}$$

where

$$i_{k+1} = \text{argmax}_{i \in I_k^c} \ \text{max}_{x \in \mathbf{R}^n : x_{J_i} = 0} \frac{x^T A x}{x^T B x} \tag{11}$$

$$J_i = I_k^c \backslash \{i\}. \tag{12}$$

At step n, we would have found approximate solution for all cardinalities lower than n. At each step, we need to solve $(n-k)$ generalized eigenvalue problems each costing $O(k^2)$, therefore the total complexity is $O(n^4)$.

This greedy search algorithm is computationally efficient compares to the NP-hard problem of exhaustive search, however, for the solution found from this algorithm to be optimal, the optimal solution needs to have support set at each carnality k that satisfies $I_k \subset I_{k+1}$. This suggests that the solution quality found using this method depends on the assumption about the optimal solution sets. For example, if the optimal mean reverting portfolio of cardinality 2 doesn't include the stock that make up the optimal mean reverting portfolio of cardinality 1, then the greedy search solution would not find that.

We can remedy this problem by running the recursive process forward or backward. I.e., on a forward step, we add an index to the index set just as before, but on a backward step, we delete an index that would result in the least predictability reduction:

$$I_{k+1} = I_k \backslash \{i_{k+1}\} \tag{13}$$

where

$$i_{k+1} = \text{argmax}_{i \in I_k^c} \ \text{max}_{x \in \mathbf{R}^n : x_{J_i} = 0} \frac{x^T A x}{x^T B x} \tag{14}$$

$$J_i = I_k^c \cup \{i\}. \tag{15}$$

For a cardinality target of n, we could specify that the recursive program runs forward or backward each with probability 0.5, except at the edge cases of $k = 1$ and $k = n$, it turns back with probability 1. We record the top portfolios for each cardinality found during this process.

We ran the greedy search algorithm with forward and backward steps on the S&P500 stocks in the time period of 2017-2021. We observed that the selected portfolio is mean-reverting in-sample, but performs poorly in out of sample periods. We suspect that the large amount of freedom given by the combination of the 500 stocks result in overfitting.

3 Background II: Trading Signals based on Stochastic Approach

3.1 Price-level Signals

We first look at the simplest case, in which we only have two stocks. For such a stock pair $\{P, Q\}$, we denote their prices at time t by P_t and Q_t. We perform linear regression of the log prices $\log(P_t)$ against $\log(Q_t)$ in the following form:

$$\log(P_t) = \beta \log(Q_t) + \alpha + \epsilon_t. \tag{16}$$

In pairs trading, we require that the residual ϵ_t is mean-reverting within a reasonable period of time. We can further quantify the normalized residual by the z-score defined as:

$$zscore_t = \frac{\epsilon_t - \mu}{\sigma}, \tag{17}$$

where μ is the mean value of the residual, and σ is the standard deviation.

From the discussions above we can see, z-score describes the spread of two (weighted) stocks at the price level, we can write the residual explicitly as:

$$\epsilon_t = \beta \log(Q_t) - \log(P_t) - \alpha. \tag{18}$$

We can use the z-score as a trading signal to enter or exit positions with the portfolio weights

$$\begin{aligned} \mathrm{wt}_P &= \frac{1}{1+\beta}, \\ \mathrm{wt}_Q &= \frac{\beta}{1+\beta}. \end{aligned} \tag{19}$$

And the PnL will be

$$\mathrm{PnL}(t, T) \sim \xi \mathrm{wt}_P (\epsilon_T - \epsilon_t) C, \tag{20}$$

where C is the capital, $\xi = 1$ if long the spread, $\xi = -1$ if short the spread.

From the z-score, we can form a trading rule. For example, if we set the enter threshold to be 2, then we short the spread at $zscore > 2$, long the spread at $zscore < -2$, and exit positions at $zscore \sim 0$.

While the z-score can provide us an excellent trading signal, there are some other inspiring research within the stochastic approach. In the research paper [3], Elliott *et al.* outlined an approach to pairs trading which explicitly models the mean reverting behavior of the spread in a continuous time setting. The observed spread y_t is defined as the difference between the two stock prices. It is assumed that the observed spread is driven mainly by a state process x_k plus some measurement error, ω_k, i.e.:

$$y_k = x_k + D\omega_k, \tag{21}$$

where x_k denotes the value of variable at time $t_k = j\tau$, for $k = 0, 1, 2, ...$ and $\omega_k \sim$ IID$N(0,1)$ with $D > 0$ is a constant measure of errors. The state variable x_k is assumed to follow a mean reverting process:

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\epsilon_{k+1}, \tag{22}$$

where $\sigma > 0$, $b > 0$, $a > 0$, and $\{\epsilon_k\}$ is IID $N(0,1)$ and independent of $\{\omega_k\}$.

The process mean reverts to $\mu = a/b$ with "strength" b and $x_k \sim N(\mu_k, \sigma_k)$, where

$$\mu_k \rightarrow \frac{a}{b}, \quad \text{at} \quad k \rightarrow \infty, \tag{23}$$

and

$$\sigma_k^2 \rightarrow \frac{\sigma^2\tau}{1-(1-b\tau)^2}, \quad \text{at} \quad k \rightarrow \infty. \tag{24}$$

We can rewrite the equation in the form

$$x_k = A + Bx_{k-1} + C\epsilon_k, \tag{25}$$

with $A = a\tau \geq 0$, $0 < B = 1 - b\tau < 1$, and $C = \sigma\sqrt{\tau}$. The discrete process can be approximated by a continuous process, i.e. $x_k \sim X_t$ where $\{X_t | t \geq 0\}$ satisfies the stochastic differential equation:

$$dX_t = \rho(\mu - X_t)dt + \sigma dB_t, \tag{26}$$

where $\rho = b$, $\mu = a/b$ and $\{B_t | t \geq 0\}$ is a standard Brownian motion. By assuming x_k follows that process, Elliott *et al.* considered the observed process y_k which is mainly driven by an AR(1) process in a discrete context.

Using the Ornstein-Uhlenbeck process as an approximation, the first passage time result for X_t will be

$$T = \inf\{t \geq 0, X_t = \mu | X_0 = \mu + c\sigma/\sqrt{2\rho}\} = \hat{t}\rho, \tag{27}$$

where

$$\hat{t} = 0.5\ln[1 + 0.5(\sqrt{(c^2-3)^2 + 4C^2} + C^2 - 3)], \tag{28}$$

and T means the time needed for the process X_t to reach μ for the first time given $t = 0$, $X_0 = \mu + c\sigma/\sqrt{2\rho}$, where c is a constant. Coefficients A, B, C and D in the equations are estimated using the state space model and the Kalman filter.

Elliott *et al.* suggested a pairs trading strategy by firstly choose a value of $c > 0$, then we enter the positions when $y_k \geq \mu + c\sigma/\sqrt{2\rho}$ or $y_k \leq \mu - c\sigma/\sqrt{2\rho}$, and exit the positions at time T later where T is the first passage time defined above. However, they did not give clear explanation how to choose the optimal value of c which can be regarded as a threshold to open a pairs trading position.

That strategy has a similar enter rule but a quite different exit rule comparing to our strategy. According to the research paper, we should exit the position at a pre-calculated time T according to statistical calculations, no matter what the price is at that time. In the backtesting with real market data, that strategy has poor performance compared to the trading rules with the z-score, because we are heavily relying on the accuracy of the statistical results from historical data and is not observing the market when we are trading. The movement of the market is much more complicated than the model described, so although conceptually it provides a good explanation, it cannot perform well in the real market.

3.2 Return-level Signals

From the equation

$$\alpha + \epsilon_t = \beta \log(Q_t) - \log(P_t), \tag{29}$$

we can also write

$$\alpha + \epsilon_{t-1} = \beta \log(Q_{t-1}) - \log(P_{t-1}). \tag{30}$$

Since the difference of log price is the return, we denote the (daily) return of stock P and Q by PR_t and QR_t, and we have

$$\Delta\epsilon_t = \beta \log(QR_t) - \log(PR_t). \tag{31}$$

From the equation above we can see, the (weighted) spread of the log returns of two stocks is the change of the residual. Comparing with the equations of price-level signals, we can define an r-score to further quantify the normalized $\Delta\epsilon_t$:

$$rscore = \frac{\Delta\epsilon_t - \mu'}{\sigma' 4}. \tag{32}$$

Since the spread of the return of two stocks is the rate of changing of the residual, we can consider the r-score as the derivative to time t of the z-score. It can tell us how the z-score is moving. We will further discuss how to make use of the r-score in our trading rules in the next subsection.

There have been extensive studies on the topic of the return-level signals. In the research paper [4], Do *et al.* proposed a pairs trading strategy by modelling mispricing at the return level, as opposed to the more traditional price level. The model also incorporates a theoretical foundation for a stock pricing relationship in an attempt to remove *ad hoc* trading rules, which are prevalent in the paper [3].

The proposed model adopted more general modelling, i.e.:

$$x_k = A + Bx_{k-1} + C\epsilon_k, \tag{33}$$

$$y_k = x_k + \Gamma U_k + D\omega_k, \tag{34}$$

where x_k, for $k = 0, 1, ...$, is a state process. One main difference in that model is that here y_k is the observed spread defined as the difference of the asset returns, i.e.:

$$y_k = R_{1,k} - R_{2,k}, \tag{35}$$

where $R_{1,k}$ and $R_{2,k}$ are the returns of asset S_1 and S_2, respectively, at time k.

Another difference between that approach comparing to Elliott *et al.* is the existence of variable U_k where U_k is an exogenous input. If $\Gamma = 0$, the model here will be the same as the previous model.

The authors used the residual spread function $G(R_t^A, R_t^B, U_t)$ to quantify the state of disequilibrium, and they specified the residual spread function as:

$$G_t = G(P_{1,t}, P_{2,t}, U_t) = R_{1,t} - R_{2,t} - \Gamma r_{m,t}, \tag{36}$$

where $r_m = R_m - r_f$ denotes the excess of market return over risk free rate. We can estimate the parameters in the model using the state space model and the Kalman filter with historical data,.

In the research paper, the authors proposed a trading rule that we take a long-short position whenever the accumulated spread δ_k exceeds a certain threshold, and we exit at the first passage time.

Although that strategy sounds plausible, as we will see in the next subsection, it cannot perform well in real market. After all, we are directly trading the stock price, not the stock return, if we enter and exit positions only based on the return level signal, our strategy won't be stable for the long run.

3.3 Further Discussions

While the return-level signals sounds very appealing, it is actually not a promising signal for our trading rules. From the example in the figure we can see, when the price

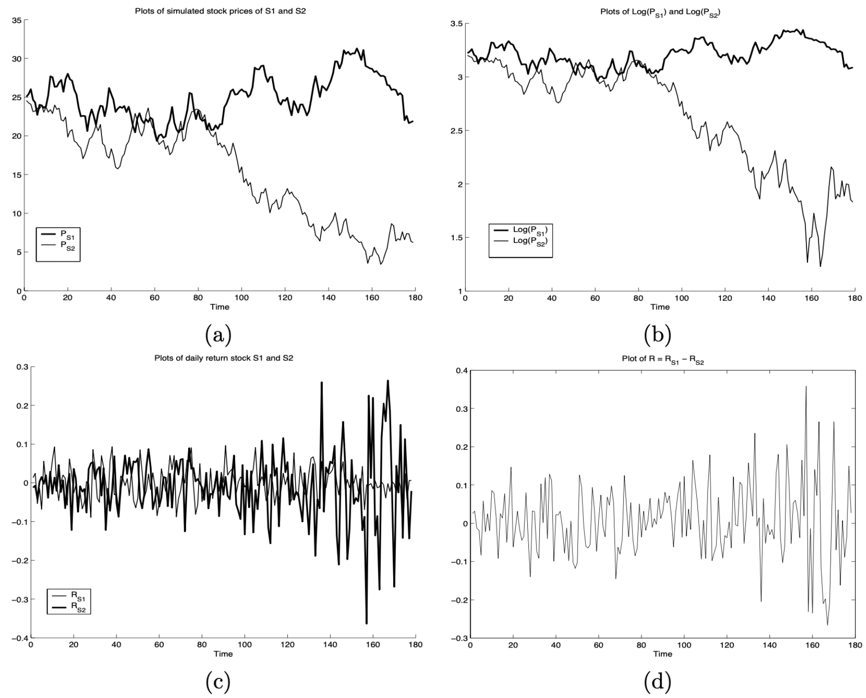


Figure 4: Examples of Price-level and Return-level signals (From [2])

of the two stocks diverge, we still have mean-reverting signals for the return level. From that observation, we should not use the return-level signal to decide whether the pair is mean-reverting, or when should we enter the positions.

However, since the r-score is the changing rate of the z-score, we can use it as a stop-loss signal. That is, we set an “early liquidate threshold” in addition to the “regular” liquidate threshold, if z-score is greater than the early liquidate threshold and r-score is positive and greater than a certain threshold, the pair is possibly broken and we should exit the positions.

To reduce the noise in the return signal, in the actual implementations, we use the average value of rolling P days ($P = 3$ in our programs). If we can see the z-score is quickly and consistently moving from zero for several days, we are more confident that the mean-reverting portfolio might be broken and we should exit the positions.

The return level signal can provide us the opportunity to exit earlier and avoid greater loss. We will discuss the stop-loss strategies in more detail in the next section.

4 Risk Management and Capital Allocation

4.1 Value At Risk

Value at Risk (VaR) serves as a critical metric in financial risk assessment, representing the potential minimum loss expected within a defined probability over a specified time under certain market conditions. Our strategy utilizes the historical simulation methodology to compute the daily VaR of the portfolio. We analyze the historical returns over a 20-day lookback period, arranging them in ascending order and finding the VaR at a chosen confidence level of 5%. The calculation is then adjusted by the size of the position.

Incorporating VaR into our risk management strategy, we establish a stop-loss criterion. We will liquidate portfolio positions when the risk threshold determined by VaR is exceeded. The VaR is compared with the unrealized gains or losses of each portfolio position. When the unrealized loss is greater than the VaR, it indicates that the current loss has breached our risk tolerance threshold. In such circumstances, positions are liquidated as a measure to mitigate further losses.

4.2 Z-Score Stop Loss

The Z-Score is a critical metric in mean-reversion pairs trading, serving as a measure of the deviation of a stock price from its historical mean. In our trading strategy, a Z-Score threshold of 1.9 is used as an entry point for trades. This threshold is predicated on the assumption that it signifies a substantial deviation from the mean, representing a potential convergence to the mean in the future.

However, an excessively high Z-Score is indicating a potential disruption in the mean-reverting relationship of the stock pair. This could be attributed to shifts in market dynamics or fundamental changes within the companies constituting the pair. In our trading model, a Z-Score exceeding 3 triggers the liquidation of the concerned portfolio. Beyond this point, the mean-reverting nature of the pair is likely broken, necessitating an exit from the position to mitigate risks.

4.3 News Sentiment Stop Loss

Many previous works [9, 10] indicate that there is a causal relation between financial news sentiments and the stock price. In a mean-reverting strategy, where traders expect prices to revert to a historical average, sudden changes in sentiment may signal an upcoming trend or deviation from the mean. In such cases, we want to use sentiment analysis to detect early warnings of potential shifts in market sentiment and use it as an early stop loss signal.

In the following sections, we will first introduce the dataset we use and the news sentiment signals we extracted from the dataset. Then we explain the sentiment analysis models we consider and choose to use for practical purposes. With the model chosen, we further explain how we use it to acquire a sensible and more accurate polarity score for each news article. Based on the extracted sentiment signals, we propose two news impact metrics and explain how we use them as early stop loss signals. We also provide our preliminary investigation on the relation between stock price and news signals, and insights on how to further improve and utilize the sentiment signals in future works.

4.3.1 Dataset

The Tiingo News Feed dataset, developed by Tiingo, monitors financial news releases across 120 different news providers related to US Equities. The dataset encompasses 10,000 US Equities, starts in January 2014, and is delivered on a second frequency. It provides us with titles and concise abstract descriptions of articles related to a certain stock of a given time frame. Since the dataset starts in 2014, we only integrate sentiment analysis for backtesting periods after 2014. Given the news data, we create the following time series for each company:

1. Total volume of news V_d^{tot}: the total number of articles given a day d
2. Negative news volume V_d^-: the total number of negative articles given a day d
3. Positive news volume V_d^+: the total number of positive articles given a day d
4. Neutral news volume V_d^0: the total number of neutral articles given a day d
5. Sentiment polarity score: the difference between the number of positive and negative articles as a fraction of non-neutral articles $P_d = \frac{V_d^+ - V_d^-}{V_d^+ + V_d^-}$
6. News volume change: the difference between the total number of articles given a day V_d^{tot} and the average number of articles of the previous n days (in practice, we preset $n = 5$) $\frac{1}{n}\sum_n V_{d_i}^{tot}$ as a fraction of the sum of them $Q_d = \frac{V_d^{tot} - \frac{1}{n}\sum_n V_{d_i}^{tot}}{V_d^{tot} + \frac{1}{n}\sum_n V_{d_i}^{tot}}$.

4.3.2 Sentiment Analysis Tools

Valence Aware Dictionary and Sentiment Reasoner (VADER) is a lexical database and rule-based sentiment analysis tool that is specifically attuned to the sentiments expressed in social media. It assigns a polarity score to text data, indicating the positivity, neutrality, or negativity of the expressed sentiment. It incorporates a pre-built sentiment lexicon that considers both the valence and the context of words to provide a more nuanced sentiment analysis. Moreover, VADER is particularly effective for analyzing short texts, making it suitable for short tweets and headlines.

While VADER considers context to some extent, it may not fully capture the complexity of nuanced language or sarcasm present in financial news. Financial news often includes domain-specific language and jargon. VADER's lexicon may not be optimized for such specialized terminology, potentially leading to less accurate sentiment analysis.

In comparison, FinBert is trained on financial news corpus to understand the nuances of financial language. However, the BERT-based model requires more than 400MB of memory and takes a few hundred milliseconds for inference on CPU instances. This makes it unsuitable for research on quantconnect platform and real-time predictions in production environments.

Overall, since VADER offers a quick and accessible solution for sentiment analysis, we choose to use it for research purposes and fast inference during trading.

4.3.3 Polarity Score

To get better results, instead of analyzing headlines, we work with the article descriptions, which will provide more explanation of the articles' sentiment to compensate for VADER's disadvantages in financial domain. Specifically, we use VADER to rate individual sentences within the article description rather than the entire text, then calculate the average compound score C for all sentences. the overall sentiment S_i of a given article i's description paragraph will be

$$S_i = \begin{cases} Positive, & \text{if } C > upperbound \\ Negative, & \text{if } C < lowerbound \\ Neutral, & \text{otherwise,} \end{cases} \tag{37}$$

where $C = \frac{1}{n}\sum_{j=1}^{n} c_j$, *upperbound* and *lowerbound* are the preset thresholds. In practice, we find 0 to be a good preset value for both *upperbound* and *lowerbound*. c_j denotes the compound score of the j-th sentence, it is between -1 (indicating strong negative sentiment) and 1 (indicating strong positive sentiment).

4.3.4 News Impact Scores for Each Portfolio

Here we propose two metrics using the time series we extracted from the Tiingo news dataset. Given a mean-reverting portfolio idx, which contains m stocks, and the normalized weights of the portfolio $\{w_1, \cdots, w_m\}$, the two news impact metrics are:

the sentiment impact metric (SI)

$$SI_{idx} = \frac{|w_i \cdot w_j| \cdot |P_d^i - P_d^j|}{\sum_{i \in [1,m], j \neq i} |w_i \cdot w_j|} \tag{38}$$

the volume impact metric (VI)

$$VI_{idx} = \frac{|w_i \cdot w_j| \cdot |Q_d^i - Q_d^j|}{\sum_{i \in [1,m], j \neq i} |w_i \cdot w_j|} \tag{39}$$

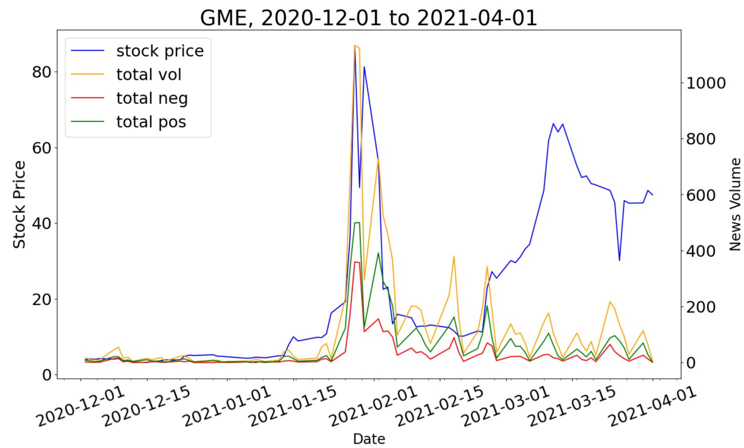
The metrics above are based on the assumption that what influences mean-reverting behavior the most is the high relative news impact amongst the stocks in the portfolio. Take the sentiment impact metric for an example, if the sentiment polarity scores between each pair of the stocks in the portfolio are all similar, then the prices of stocks will have the same moving direction. In contrast, if the sentiment polarity score of one stock is drastically different than the others, it will cause the price of that stock to increase or drop faster than others, leading to mean-reverting no more. The same assumptions are made with the news volume changes among the stocks in each portfolio.

These two metrics reflect the drastic changes in the market sentiment, which could help us detect the deviation from mean-reverting in the near future. In our strategy, besides the z-score stop loss threshold 3, we set an early liquidate threshold for z-score. We will initiate an early liquidation when the sentiment metric scores exceed a preset threshold, indicating much higher news impact on one or a subset of stocks, while z-score is above its early liquidate threshold. In Section 7, we provide results for using each of the two metrics in our news sentiment stop loss.

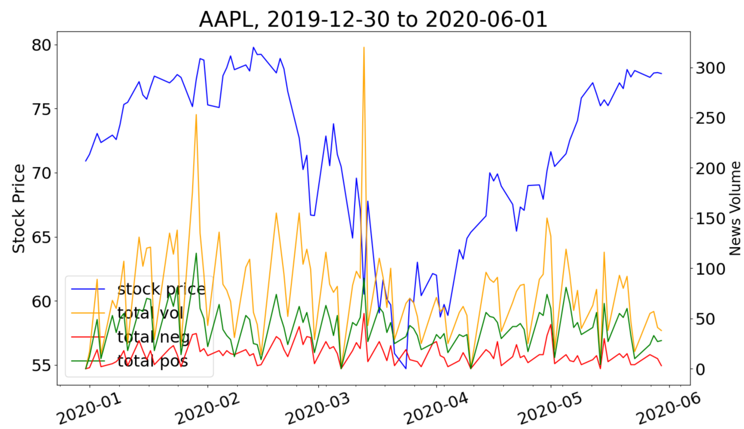
4.3.5 Investigation on Stock Price and News Sentiment

In Figure 5, we show the preliminary results that confirm our assumption of the relation between stock price and news sentiment. We picked two extreme examples to showcase our findings. Figure 5(a) shows the stock price and daily new volume of Gamestop during the short squeeze in 2021, while Figure 5(b) shows those of Apple during the start of COVID. We observe that the drastic high news increases are often during or before the equally drastic changes in stock prices.

Moreover, we perform the Granger causality test [11] to check if the daily news volume has a causality relation with the stock price for all the stocks in the mean-reverting portfolio we use. We find that most of the stocks' prices have a casual relation with the stock price, with different lag periods.



(a) GameStop During Short Squeeze in 2021



(b) AAPL During COVID 2020

Figure 5: Investigation of Relation between Stock Price and News Volume

4.3.6 Future Works

Firstly, we choose VADER as our sentiment analyzing tool because of its quick inference speed during backtesting. Since we are only trading several portfolios with daily trading frequency, we only need to do sentiment analysis once per day. This means that during live trading, the computational cost for calculating sentiments is not unbearably high.

Consequently, we could potentially get more accurate sentiment results with FinBert or other pretrained Large Language Models.

Furthermore, currently, the news impact metrics are only used as stop loss signals for mean-reverting strategy. However, sentiment analysis, which contributes to a more comprehensive analysis of the overall market trend, can also help traders understand the prevailing market sentiment for momentum trading since it relies on identifying and capitalizing on existing trends. We can use the sentiment impact signals to identify the stocks with the strongest sentiment trends, and allocate more capital to them to maximize profit.

Finally, we are trading before the market closes right now, so the news sentiment scores are calculated based on the current day's news articles. However, it is likely that the news sentiment information is already reflected in the stock price by the market closing time. By analyzing the news sentiment/news volume change throughout the day, we could catch profit early on.

4.4 Return-level Spread Stop Loss

As discussed in Section 3, we have the following equation for the return level of two stocks:

$$\Delta\epsilon_t = \beta \log(QR_t) - \log(PR_t). \tag{40}$$

According to that equation, in addition to the z-score, we can define an r-score to further quantify the normalized residual of the spread at the return level:

$$rscore = \frac{\Delta\epsilon_t - \mu}{\sigma}. \tag{41}$$

While the return level spread cannot be used to justify the mean-reversion of the price level spread, it can help us form an "early" stop loss strategy. Since the r-score is basically the changing rate of the z-score, we can set up an "early liquidate threshold" in addition to the regular liquidate threshold, if z-score is greater than the early liquidate threshold and r-score is positive and greater than a certain threshold, we think the pair is possibly broken and the price level spread will not come back zero in a reasonable time period, we should exit the positions at that time.

To reduce the noise of the return level signal, we calculate the average value of rolling P days (we set $P = 3$ in actual programs). If the return signal is consistently pointing at the direction that pulls the z-score away from zero, we are confident that the mean-reverting portfolio might be broken.

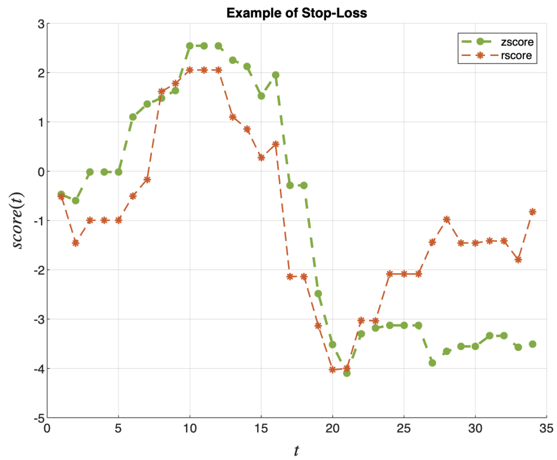


Figure 6: Example of Stop loss with the Return Level Signal

The return-level spread signal can be combined with the sentiment score signal to decide when to liquidate early. If we can detect the broken of the mean-reverting portfolio at the early liquidate threshold, we can avoid the loss at the “regular” liquidate threshold.

From the figure of z-score and r-score we can see, if we can make use of the r-score signal, we are able to exit the positions when z-score reaches ~ -2.5 because the r-score is also very negative at that time. If we only use the liquidate threshold ~ -3.5, we will exit the positions at z-score ~ -4, which will cause greater loss.

5 Combining Momentum and Mean-reverting

5.1 Momentum Trading Background

Momentum trading is an investment strategy that focuses on buying stocks demonstrating an upward price trend while selling from those exhibiting a downward trend. This approach is based on the hypothesis that stocks following a consistent trend are likely to maintain in the future. A key belief of momentum strategy is that the market is not always efficient in quickly incorporating available information into asset valuation, as shown in previous works [7], confirming the continuous applicability of momentum trading strategies within U.S. asset markets.

5.2 Incorporating Momentum and Mean Reversion Trading

Our backtests over different time periods suggest that sole reliance on the mean-reverting strategy may result in times when our capital is uninvested for any stocks. This may be due to our strict position exit conditions. Similar findings are also presented in work [8], wherein the author identifies the existence of both short-term momentum trends and long-term mean reversion patterns in the foreign exchange market. The author also mentioned the parallels between the dynamics of the foreign exchange markets and stock markets, suggesting the potential of using similar strategies across these domains. Consequently, to optimize capital utilization, we have integrated a momentum trading approach during these short-term uninvested periods. This strategy utilizes traditional momentum indicators, such as the Relative Strength Index (RSI) and Bollinger Bands (BB), aiming to make a profit on short-term market movements.

5.3 Entering Signals

To determine the appropriate momentum direction for each stock, we compute the autocorrelation metric over a lookback period of 50 days. Autocorrelation is the correlation of a time series with its own past values and serves as a measure of how the historical values of a variable are related to its current value. A higher absolute value of autocorrelation suggests stronger correlations, indicating that a stock is more suitable for momentum trading strategies. The autocorrelation is calculated using the following formula:

$$\text{Autocorrelation}(\tau) = \frac{\sum_{t=1}^{N-\tau}(X_t - \bar{X})(X_{t+\tau} - \bar{X})}{\sum_{t=1}^{N}(X_t - \bar{X})^2} \tag{42}$$

In our methodology, stocks exhibiting an autocorrelation value greater than a positive threshold indicate a positive correlation between their current price and the prices of the past 50 days, and we employ a trend-following strategy in such cases. Conversely, for stocks with autocorrelation values below a negative threshold, we use a trend-reversal strategy. Stocks that do not meet either threshold are considered currently unsuitable for momentum trading. This approach of applying momentum strategies can potentially enhance the efficiency of our trading strategy.

5.4 Trading Signals

In our strategy, we employed two indicators—Relative Strength Index (RSI) and Bollinger Bands (BB)—to determine the entry points for long or short positions in stocks.

RSI is a momentum oscillator to measure the speed and change of price movements. Calculated over a specific lookback period, it helps identify the overbought or oversold conditions of an asset. An RSI value exceeding 70 typically signals overbought, indicating strong positive price momentum, while a value below 30 suggests an oversold condition, reflecting negative price momentum. For trend-following stocks in our strategy, we adopt a long position when the RSI is above 70 and a short position when it is below 30.

Bollinger Bands offers a statistical perspective on stock prices and volatility over time. It consists of a middle band that represents a moving average, accompanied by upper and lower bands determined by adding or subtracting two standard deviations from the middle band. The distance of stock prices to these bands serves as an indicator of the market's overbought or oversold status. In our methodology, the upper and lower bands are constructed using 2 standard deviations from the middle band. A stock price falling below the lower BB band triggers a long position, while a price exceeding the upper band triggers a short position.

Regarding the exit strategy for momentum positions, we liquidate long positions when the current stock price surpasses the upper BB band or when the RSI declines below the lower band. These conditions often indicate a nearing peak in upward momentum, suggesting a potential downward trend reversal shortly. In contrast, short positions are closed when the stock price drops below the lower BB band or when the RSI rises above the upper threshold. This framework serves as an effective management of risk exposure.

6 Overview of Our Strategy

6.1 Trading Process Flowchart

Figure 7 is a flowchart of our trading process, with each step discussed in previous sections.

6.2 Other Trading Considerations

In addition to the outlined trading process, our strategy incorporates several key elements to enhance realism and effectiveness.

6.2.1 Trading Cost and Slippage

To make our backtest more realistic and reliable, we incorporate a brokerage model of Interactive Brokers Brokerage. This model simulates the trading conditions and costs associated with a brokerage firm. This closely resembles the actual trading environment with transaction costs, margin requirements, and other trading constraints.

Furthermore, we employ the Volume Share Slippage Model to estimate the costs due to price movement between the initiation and execution of a trade. The model considers the volume of our order relative to the total trading volume of the stock. Using this approach to model market impact provides a more realistic assessment of trading costs and potential profits or losses.

6.2.2 Margin and Buying Power Constraints

The margin model employed in our algorithm is critical, particularly for strategies like pairs trading which rely on little capital for short positions while utilizing this cash for long positions. We adopt a margin model with a leverage multiplier of 2.5, meaning that we can leverage our capital by at most 2.5 times. his serves as a constraint for shares of each stock we can hold in our portfolio.

Initially, we encountered issues with insufficient buying power. To resolve this, we employed a check against available cash before setting target holdings for each stock. If the cash is inadequate for the intended trade amount, we will reduce the order size to the current available amount of cash accordingly. By incorporating these checks and adjustments, our algorithm dynamically adapts to the current account cash balance and avoids placing orders that cannot be fulfilled due to insufficient funds.

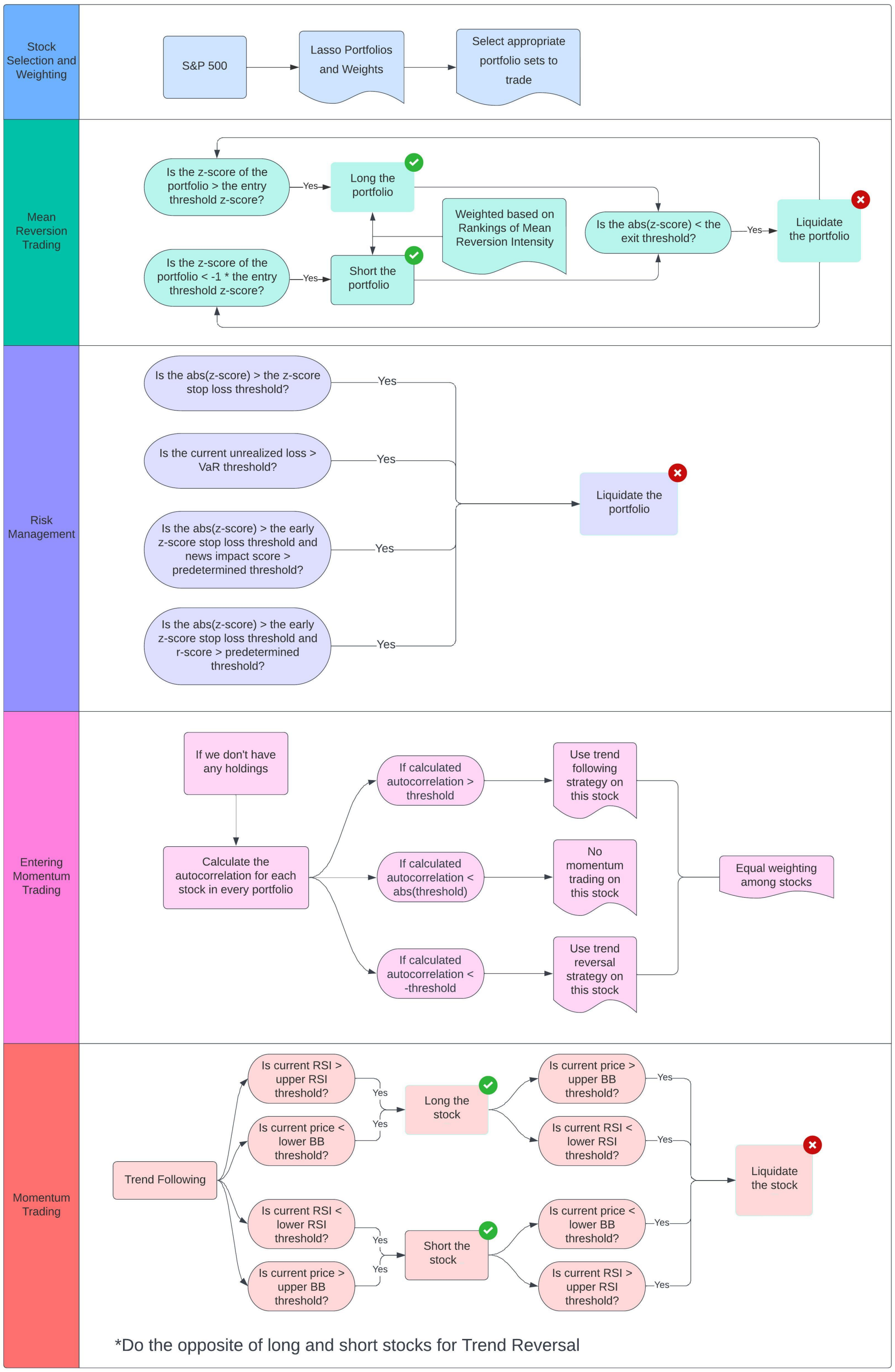


Figure 7: Overview of the trading process of our strategy

6.2.3 Trading Time

To further reduce insufficient buying power error and margin, we tried to use scheduled events to trade at specific times every day. Prices at market close are more indicative of a stock's true value since they incorporate all the information and activities of the entire trading day. Therefore, Instead of using the default OnData function which trades at the market opening, we create a scheduled event to trade 10 minutes before the market closes. By scheduling trades shortly before the market closes, we can use these more stable prices to make decisions, potentially leading to more informed and effective trading choices.

7 Backtesting Results

We trade with five portfolios, which are ['MCD', 'YUM'], ['FAST', 'GWW'], ['GD', 'LHX', 'LMT', 'NOC', 'RTX'], ['CI', 'CNC', 'ELV', 'HUM', 'MOH', 'UNH'],['CDNS', 'INTU', 'SNPS']. In Table 2, we show backtesting results for trading these portfolios with equal capital allocation (CA), and with weighted capital allocation (CA) based on the strength of each portfolio's mean reversion (given from the Johansen test). We also perform an ablation study on using different stop losses, including VaR stop loss, Return-level Spread (RS) stop loss, and News Impact stop loss (based on SI and VI metrics), under the equal capital allocation setting. Using multiple stop loss conditions in general give us a lower drawdown rate, and higher return during "hard" periods, such as 2022/1/1-2022/11/1 and the stress testing period, because we avoid big losses. However, using more stop losses also means being conservative during "easy" periods where we can potentially make more profit.

Table 1: Backtesting Periods

	Time Range
In Sample (IS in short)	(2017/1/1-2021/1/1)
Out of Sample A (OOS A in short)	(2016/1/1 - 2017/1/1)
Out of Sample B (OOS B in short)	(2022/1/1-2022/11/1)
Out of Sample C (OOS C in short)	(2023/3/10-2023/10/10)
Stress Testing (ST in short)	(2020/3/1-2020/3/31)

Table 2: Backtesting Results

	Equal CA									Weighted CA		
	VaR Only			VaR + RS + SI			VaR + RS + VI			VaR + RS + VI		
	SR↑	Return↑	DD↓	SR↑	Return↑	DD↓	SR↑	Return↑	DD↓	SR↑	Return↑	DD↓
IS	1.241	144.8%	24.2%	1.163	124.00%	22.7%	1.287	143.25%	22.7%	1.435	**154.95%**	19.3%
OOS A	1.649	21.60%	4.8%	1.276	17.09%	6.0%	1.339	17.06%	4.5%	1.686	**22.88%**	4.9%
OOS B	0.278	4.90%	7.8%	0.755	10.16%	8.5%	0.428	6.64%	7.8%	0.915	**13.20%**	5.8%
OOS C	1.357	**18.46%**	6.0%	0.897	14.62%	8.0%	0.682	10.08%	3.7%	1.144	14.27%	4.1%
ST	-1.062	-10.77%	21.8%	-0.943	**-9.22%**	21.8%	-1.062	-10.77%	21.8%	-1.271	-13.54%	23.10%

Moreover, our trading strategy shows outperformance relative to the S&P 500 benchmark, from which we derived our stock selection. Our strong performance can be seen when

Table 3: Live Trading Results

	Weighted CA VaR + SRS + VI		
	SR↑	Return↑	DD↓
2023/12/4 - 2023/12/11	2.119	-1.409%	2.6%

Table 4: Benchmark Results

	SPY		
	SR↑	Return↑	DD↓
IS	0.618	79.02%	33.7%
OOS A	0.836	13.9%	9.7%
OOS B	-0.75	-18.31%	26.2%
OOS C	1.175	12.49%	7.9%
ST	-0.797	-13.13%	29.3%

examining risk-adjusted returns, as indicated by the Sharpe Ratio. During both in-sample and all out-of-sample periods, our strategy maintained a Sharpe Ratio consistently around a value of 1, indicating strong performance under bullish or normal market conditions. During the OOS period B where the SPY exhibited negative returns, our strategy not only achieved positive returns but also maintained notable risk-adjusted returns of 0.915.

In terms of risk management, our strategy displayed a smaller value of drawdowns for all backtesting periods compared to the benchmark. This outcome shows the effectiveness of our comprehensive risk management approach, which includes both traditional value-at-risk measures and sentiment analysis. This combination provides good results in mitigating downside risks.

The stress test period with increased market volatility due to Covid-19, presents some challenges in generating high returns for our strategy. Still, our strategy's drawdown (23.1%) was less than that of SPY (29.3%), indicating our strategy's ability to manage downside risks. In the future, we plan to explore the integration of more dynamic hedging and capital allocation to further improve our strategy's robustness against volatile market environments.

8 Future Work

In the future, we plan to implement several enhancements to our strategy.

One of the developments is applying refactoring of portfolio weights from Lasso Regressions, and performing more backtests on trading cost for determining the trade-off between rebalance frequency and cost efficiency.

We also aim to explore other capital allocation methods, such as risk parity and return permanence. Such dynamic cash allocation among selected portfolios can potentially enhance our overall risk-adjusted returns.

Another possible improvement is the extension of our sentiment analysis stop loss from our pairs trading strategy to our momentum trading segment as well. Using it as an early stop-loss threshold for exiting positions can further minimize downside risk.

Lastly, we will optimize our algorithm's hyperparameters using grid search methods. This will enable us to have more consistent performance across different time periods and market conditions, generating a more robust and adaptable trading algorithm.

9 Quantconnect Code

QC code for backtesting period 2016/1/1-2017/1/1

References

[1] Krauss, C., *Statistical Arbitrage Pairs Trading Strategies: Review and Outlook*, Journal of Economic Surveys Vol. 31, No. 2, pp. 513–545 (2017).

[2] Puspaningrum, H., *Pairs trading using cointegration approach*, PhD thesis, University of Wollongong (2012).

[3] Elliott, R.J., Van Der Hoek, J. and Malcolm, W.P. *Pairs trading*, Quantitative Finance 5(3): 271–276 (2005)

[4] Do, B., Faff, R. and Hamza, K., *A new approach to modeling and estimation for pairs trading*, Proceedings of 2006 Financial Management Association European Conference (2006)

[5] Galenko, A., Popova, E. and Popova, I., *Trading in the presence of cointegration*, The Journal of Alternative Investments 15(1): 85–97 (2012)

[6] d'Aspremont, A., *Identifying small mean-reverting portfolios*, Quantitative Finance, 11(3):351–364 (2011)

[7] Serban, A. F., *Combining mean reversion and momentum trading strategies in foreign exchange markets*, Journal of Banking & Finance, 34(11), 2720–2727 (2010)

[8] Asness, C. S., Moskowitz, T. J., & Pedersen, L. H., *Value and momentum everywhere*, SSRN Electronic Journal, (2009)

[9] Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grcar, and Igor Mozetic. *The effects of twitter sentiment on stock price returns*, PloS one, 10(9):e0138441 (2015)

[10] Wesley S. Chan. *Stock price reaction to news and no-news: drift and reversal after headlines*, Journal of Financial Economics, 70(2):223–260 (2003)

[11] Cees Diks and Valentyn Panchenko. *A new statistic and practical guidelines for non-parametric granger causality testing*, Journal of Economic Dynamics and Control, Computing in economics and finance 30(9):1647–1669 (2006)