Sector-based pairs trading strategy with novel pair selection technique

Pranjala G. Kolapwar, *Member, IEEE*, Uday V. Kulkarni, and Jaishri M. Waghmare, *Member, IEEE*

***Abstract*—A pair trading strategy (PTS) is a balanced approach that involves simultaneous trading of two highly correlated stocks. This paper introduces the PTS-Return-based pair selection (PTS-R) strategy which is the modification of the traditional PTS. The PTS-R follows a similar framework to the traditional PTS, differing only in the criteria it employs for selecting stock pairs. Moreover, this paper proposes a novel trading strategy called Sector-Based Pairs Trading Strategy (SBPTS) along with its two variants, namely SBPTS-Correlation-based pair selection (SBPTS-C) and SBPTS-Return-based pair selection (SBPTS-R). The SBPTS focuses on the pairs of stocks within the same sector. It consists of three innovative phases: the classification of input stocks into the respective sectors, the identification of the best-performing sector, and the selection of stock pairs based on their returns. The goal is to identify the pairs with a strong historical correlation and the highest returns within the best-performing sector. These chosen pairs are then used for trading. The strategies are designed to enhance the efficacy of the pairs trading and are validated through experimentation on real-world stock data over a 10-year historical period from 2013 to 2023. The results demonstrate their effectiveness compared to the existing techniques for pair selection and trading strategy.**

***Impact Statement*—The introduction of the PTS-R and SBPTS strategies represents a significant advancement in pairs trading methodology, offering innovative approaches to pair selection within the stock market. By incorporating distinct criteria for identifying stock pairs, these strategies aim to optimize trading outcomes and improve overall efficacy. The utilization of historical correlation and return-based metrics, coupled with a focus on sector-based analysis, enhances the precision of pair selection and increases the potential for profitable trades. Through rigorous experimentation and validation over a substantial 10-year historical period, these strategies have demonstrated their superiority over existing techniques, showcasing their potential to revolutionize pairs trading practices. The impact of these strategies extends to traders, investors, and financial markets, providing valuable insights and tools for enhancing trading performance and decision-making processes.**

***Index Terms*—Pairs trading, sector-based pairs trading, cumulative annual returns, stock pair selection, correlation, and historical data analysis.**

I. Introduction

The traditional pairs trading is a popular and well-established trading strategy used by traders and investors to capitalize on the relative performance of two correlated stocks [1]. The strategy is built on the concept of mean reversion when the spread between the prices of the two stocks will revert to its historical average. The traders aim to profit from the temporary deviations in their relative prices [9] by identifying pairs of stocks with a historical correlation. However, various researchers [3], [4], [5] argue that the pairs trading may not yield substantial profits over extended periods. Despite this, pairs trading remains widely employed in the contemporary financial markets to generate profits [6]. A recent comprehensive examination of diverse pairs trading strategies is presented in a survey [1].

The methodology behind traditional pairs trading involves calculating the spread between the two stocks, often using statistical tools such as cointegration [13] or linear regression [7]. The z-score is commonly employed to quantify the deviation of the spread from its mean in terms of standard deviations. When the spread deviates significantly from its average, trading signals are generated, prompting traders to take long and short positions in the two stocks to profit from the expected convergence. The strategy involves taking a long position in the underperforming stock and a short position in the outperforming stock. The goal is to capture profits when the price ratio between the two stocks reverts to its mean, regardless of the direction of the overall market. The traditional PTS typically doesn't take sector-specific factors into account when selecting stock pairs. This can lead to pairs that are vulnerable to sector-wide trends or events. Moreover, the existing PTS typically identifies the pairs based on correlation or cointegration [20].

The objective of this research is to improve the traditional PTS by modifying the pair selection method and developing sector-based trading strategies to improve investment performance. The PTS and PTS-R, both share a common operational framework, the only distinction lies in the criteria employed for selecting stock pairs. The PTS relies on a distance measure for stock pair selection, whereas in the PTS-R, it is done by using cumulative annual returns of each stock. This paper also introduces a novel trading strategy, SBPTS along with its two variants, SBPTS-C and SBPTS-R. The adoption of a sector-based approach in trading strategies significantly mitigates several inherent limitations of traditional trading methods. These limitations include overlooking sector-specific dynamics, neglecting broader market trends, and many more. In the broader context of industries, market trends, and sector-specific factors, its goal is to enhance diversification, improve risk management, and increase the potential to capture opportunities.

The rationale behind the SBPTS is that the stocks within

Pranjala G. Kolapwar is with Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded(MH)-431606, Bharat (e-mail: pgkolapwar@sggs.ac.in).

Uday V. Kulkarni, Professor is with Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded(MH)-431606, Bharat (e-mail: uvkulkarni@sggs.ac.in).

Jaishri M. Waghmare, Associate Professor is with Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded(MH)-431606, Bharat (e-mail: jmwaghmare@sggs.ac.in).

the same sector tend to be influenced by similar economic and industry-specific factors. The strategy attempts to minimize the exposure to broader market movements and sector-specific risks by focusing on the pairs of stocks within a sector. This is useful during periods of heightened market volatility or when specific sectors experience unique events or news.

The paper is structured as follows. The section II provides an overview of the existing literature in the field. The section III presents the details of the PTS-R, the SBPTS, and its variants. The experimental setup, result analysis, and validation of the proposed approach are explained in the section IV. Section V summarizes the findings and discusses potential future directions.

II. Related work

This section deals with a brief review of the PTS along with the various pair selection methods that are used to construct the stock pairs. In recent years, there has been a growing interest in applying pairs trading to gain profit. This literature review aims to explore the key findings and highlight the main contributions related to pairs trading.

Gatev et al. [1] and Vidyamurthy [2] laid the groundwork for pairs trading. They introduced the concept of cointegration and mean-reversion as key components for selecting stock pairs. Recent research has incorporated machine learning (ML) techniques to enhance pairs trading strategies [15], [21]. Flori and Regoli [12] proposed a Long Short-Term Memory (LSTM) network to identify the pairs-trading opportunities. The LSTM networks are employed to reveal the potential pairs with the exploitable price divergence and the convergence.

Deep learning (DL) models have been explored by Krauss et al.[9]. They have proposed the statistical arbitrage on the S&P 500 using deep neural networks, gradient-boosted trees, and the random forest algorithm and presented review [11]. There has been a growing interest in applying reinforcement learning (RL) to financial trading problems [18]. The paper [10] proposed the reward-shaping method that can lead to more efficient and robust trading strategies by using improved RL.

Gupta and Chatterjee [14] proposed a methodology for selecting the stock pairs for the pairs trading while considering the lead-lag relationship between stock prices. This paper proposes a new distance measure known as dynamic cross-correlation type (DCCT) which incorporates the lead-lag relationship. Some recent works have extended pairs trading to market prediction. They also explore a pairs trading strategy in the foreign exchange market [21] by applying extreme value theory to enhance the risk management aspect of pairs trading.Naccarato et al. [34] proposed a pairs trading cointegration strategy, which identifies the prices and returns of each stock based on the cointegration relationship estimated using the Vector Error Correction Model (VECM). The new pairs trading model has been designed to incorporate a mean-reverting strategy [35]. This proposed strategy was applied to three pairs of real cointegrated stocks belonging to the European financial sector. F García et al. [36] developed the heuristics non-dominated sorting genetic algorithm II which analyses the trade-off between return, risk, and corporate social responsibility.

Wang et al. [16] applied deep learning techniques to predict stock prices and identify potential pairs for trading. In the study conducted by Miroslav Fil et al. [8], the effectiveness of pairs trading rules, considered the gold standard, is examined. The study emphasizes the influence of market factors on the strategy's optimal parametrization, suggesting the need for adjustments in response to modern market conditions.

The cointegration and correlation approach is generally considered superior for pairs trading due to its foundation in statistical theory and its ability to identify stable, mean-reverting relationships [11]. However, cointegration tests are built on the assumption of stationarity in data series, which isn't always met in real-world financial data. Financial time series often exhibit non-stationary behavior due to trends, seasonality, or other factors, posing challenges for the cointegration approach. In contrast, in certain market conditions or for specific investor goals, the accumulated or annual returns approach could be advantageous, particularly for its simplicity and alignment with performance-based strategies [36]. Therefore, the proposed return-based pair selection, PTS-R, addresses these considerations effectively. By directly using the return of each stock, PTS-R can be more relevant for investors focusing on profitability rather than statistical properties. Return-based methods offer more robustness in situations where data are non-stationary or cointegration assumptions are not fulfilled. Additionally, return-based selection can be more adaptable to different market conditions since it doesn't require stringent assumptions about the statistical properties of the data, such as stationarity.

The key motivation for the shift towards sector-based pairs trading lies in its ability to mitigate some of the limitations associated with traditional pairs trading. These limitations include the sensitivity of traditional pairs to broader market movements, the impact of macroeconomic factors, and the vulnerability of individual stocks to idiosyncratic events. So, to overcome the said limitations, SBPTS is proposed. This approach enhances the robustness of pairs trading strategies across diverse market conditions by reducing exposure to systematic risk.

III. Proposed model

This section describes PTS-R and the proposed novel trading strategy, SBPTS along with its two variants.

A. PTS-R

The PTS-R modifies the traditional PTS by utilizing annual cumulative returns for stock pair selection, replacing the conventional distance measure [14]. It is introduced to refine and improve pairs trading strategies, addressing limitations, and adapting to the evolving landscape of financial markets. In the subsequent sections, the fundamental steps of the PTS-R, outlined in Table I, will be thoroughly explained.

1) Selection of the stock pairs: The pairs trading involves selecting two correlated stock pairs by utilizing their relative performance. The input stock files used in the experimentation represent a selection of individual stocks from different sectors. These files represent the historical price data used

TABLE I
BASIC STEPS INVOLVED IN THE PTS-R

Steps	
1	Selection of the stock pairs
2	Calculation of the spread and the z-score
3	Defining the entry and exit threshold
4	Calculating cumulative returns of the strategy
5	Backtesting and risk management by profit/loss

for conducting the analysis, building the models, and making the investment decisions. Each stock file contains information such as the stock symbol, date, opening price, closing price, high and low prices, adjusted close price, trading volume, and other relevant data parameters. In the experimental approach, the focal parameter chosen for trading analysis is the adjusted closing price.

It is essential to perform the normalization and the pre-processing of the input stocks before proceeding with actual trading. Several methods are commonly employed for pre-processing, enhancing the quality and reliability of the data. Among these methods, Min-Max Scaling [26] is frequently utilized. In this technique, the data inputs are mapped into a predefined range $[0, 1]$ and $[-1, 1]$.

Let $\mathcal{F}$ be the set of M stock input files denoted as $\{f_m | m = 1, 2, 3, \ldots, M\}$ collected from YFinance API[1]. Each f_m is represented by a two-dimensional array of size $d \times p$, where d is the number of days and p is the number of input parameters. The historical price data for the file f_m at the i-th day and j-th input parameter is denoted as D_{ij}. Here, i varies for different days, while j remains constant and is equal to the adjusted closing price. The normalized and the preprocessed data are calculated as follows:

$$D'_{ij} = \frac{D_{ij} - \min(D_{ij})}{\max(D_{ij}) - \min(D_{ij})}, \quad \forall f_m \in \mathcal{F}, \forall i, \tag{1}$$

$$D''_{ij} = \log(D'_{ij}) - \log(D'_{ij}) \cdot (\text{shift} \cdot 1), \quad \forall f_m \in \mathcal{F}, \forall i, \tag{2}$$

where D'_{ij} is the normalized data and D''_{ij} is the preprocessed data obtained through the application of the Min-Max Scalar method and a logarithmic transformation with a shift, respectively.

The next step involves the identification of suitable stock pairs for trading. In traditional pairs trading [14], the stock pairs are selected based on statistical measures such as distance or correlation. These traditional methods may not adapt quickly to shifts in market dynamics, as the relationships between individual stocks can change over time. Hence, the return-based pair selection method is proposed. This method uses cumulative returns of individual stocks to find pairs for trading. To achieve this, the initial step involves calculating each stock's daily return, R. Subsequently, the cumulative return, R^c for each stock is computed, leading to the final step where the overall cumulative return, $R^{oc}_{f_m}$ for each stock file f_m is determined, where $m = 1, 2, 3, \ldots, M$. These calculations are performed as follows:

$$R_i = \frac{D''_{ij} - D''_{i-1,j}}{D''_{i-1,j}}, \quad \forall f_m \in \mathcal{F}, \forall i > 1, \tag{3}$$

[1] https://finance.yahoo.com.

where R_i represents the daily return for the i^{th} day, and $D_{i-1,j}$ is the adjusted closing price on the $(i-1)^{th}$ day. The condition $i > 1$ ensures that the cumulative return updates from the second day, using the initial value, R_1 as a reference point.

The calculation of R^c begins from day two onwards to accumulate the daily returns as:

$$R^c_i = (1 + R_i) \times R^c_{i-1}, \quad \forall f_m \in \mathcal{F}, \forall i > 1, \tag{4}$$

where R^c_{i-1} is the cumulative return on the $(i-1)^{th}$ day. The $R^{oc}_{f_m}$ is adjusted to incorporate the most recently computed value of R^c_i as:

$$R^{oc}_{f_m} = R^c_i, \quad \forall f_m \in \mathcal{F}, \forall i > 1. \tag{5}$$

After computing the overall cumulative returns for each stock file within the set $\mathcal{F}$, the process of stock pair selection begins. It involves identifying two stock files, f_p and f_q, that form a pair (f_p, f_q), where both belong to the set $\mathcal{F}$. This pair is chosen based on their overall cumulative return values, ensuring that f_p has the highest cumulative return among all files in $\mathcal{F}$, and f_q has the second-highest cumulative return.

The selection of this stock pair, driven by the maximization of the overall cumulative returns, forms the foundational step in the pair trading strategy, laying the groundwork for subsequent analysis and trading decisions.

2) Calculate the spread and z-score: After selecting the stock pair, PTS-R proceeds with the historical analysis. The historical analysis in pairs trading involves examining past price movements and statistical measures to identify patterns and potential trading opportunities by using the spread and z-score [14], [26]. This analysis helps traders to understand the behavior of the selected pair. It assists in selecting the best parameters for generating trading signals.

The spread, representing the relative performance of the two stock files, is computed either as the log price difference or the simple price difference. Subsequently, the z-score, a standardized measure of how the current spread deviates from its historical mean, is calculated using the mean and standard deviation of historical spread values. These metrics are used to generate buy and sell signals. When the spread deviates from its historical mean or exhibits certain patterns, traders may interpret this as an opportunity to make buy or sell decisions. These foundational calculations serve as key indicators in pairs trading, guiding decisions on when to enter or exit trades based on statistical measures. The spread is mathematically expressed as:

$$S_i = D''_{f_{p_{ij}}} - D''_{f_{q_{ij}}}, \quad \forall i, \tag{6}$$

where S_i is spread, and $D''_{f_{p_{ij}}}$ and $D''_{f_{q_{ij}}}$ are the adjusted closing prices of the stock files f_p and f_q at i^th trading day respectively.

Utilizing the calculated spread S_i, the z-score Z at i^{th} trading day, Z_i is determined by:

$$Z_i = \frac{S_i - \mu}{\sigma}, \quad \forall i, \tag{7}$$

where μ is the mean and σ is the standard deviation of the spread over the historical period.

The spread and z-score are the important metrics for understanding, analyzing, and optimizing the pairs trading strategy. They enable traders to make informed decisions, identify potential trading opportunities, and manage risks effectively in dynamic market conditions.

3) Define the entry and exit threshold: Following the computation of the spread and z-score between the chosen stocks, the subsequent pivotal task is to define entry and exit points. In the PTS, traders monitor the spread and z-score to identify entry and exit points [14]. It suggests that the stocks are likely to revert to their mean relationship when the spread deviates significantly from its historical behavior and the z-score exceeds a predefined threshold (e.g., $+1$ or -1). The traders might execute a pairs trade by going long on the relatively undervalued stock and short on the relatively overvalued stock, expecting the spread to revert to its historical average. The traders can strategically define entry and exit points that align with the dynamics of the selected stock pair, enhancing the precision and effectiveness of the pair's trading strategy.

4) Calculate cumulative returns of the strategy: After establishing well-defined entry and exit points, the subsequent critical step involves assessing the performance of the trading strategy. Evaluating the performance of the trading strategy encompasses various metrics, and the selection of these metrics depends on the particular objectives and preferences of traders or investors. The common performance metrics are total return, annualized return, compound annual growth rate, maximum drawdown, cumulative return, and average annual return. The cumulative returns are often favored for assessing strategy performance because they encapsulate the total impact of investment decisions over time [25], [27]. This metric considers both positive and negative movements, providing a comprehensive picture that is valuable for investors looking at the long-term effectiveness of their strategies.

To calculate the strategy's cumulative annual return, it first needs to calculate the trade's position. The positions in trading, P_i represent the quantity of a financial instrument (e.g., stocks, bonds, currencies) that a trader holds at a specific point at the i^{th} trading day. So, P_i is the trade position, and it indicates how much of the financial instrument is held at a particular trade. The P_i can be a positive value, indicating a buy position, or a negative value, indicating a sell position, or a zero value, indicating no position. It is obtained by:

$$P_i = \begin{cases} -1, & \text{if } Z_{i-1} > \theta, \\ 1, & \text{if } Z_{i-1} < -\theta, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where, Z_{i-1} is the z-score at the $(i-1)^{th}$ trading day, and θ is the predefined z-score threshold for entering a position. The use of Z_{i-1} allows the trading strategy to incorporate information from the last day to make decisions for the current day. The value of P_1 is set to 0 as no trade is made on the first day of the trading strategy based on the provided conditions. The conditions for determining P_1 are based on Z_0.

The daily return for the strategy based on the positions and changes in the spread is given by:

$$R_i^s = P_{i-1} \cdot (S_i - S_{i-1}), \quad \forall i > 1, \tag{9}$$

where R_i^s represents the strategy return on the i^{th} trading day, P_{i-1} is the position on the $(i-1)^{th}$ trading day, and S_i and S_{i-1} are the spread values on the i^{th} and $(i-1)^{th}$ trading days, respectively.

The resulting value represents the profit or loss generated by the strategy on each trading day. The annual cumulative return for the year y, denoted as R_y^c, is obtained by summing the daily returns across all trades within that year.

$$R_y^c = \sum_{i=1}^{N_y} R_i^s, \quad \text{for } y = 1, 2, \ldots, Y, \forall i > 1, \tag{10}$$

where N_y is the total number of trading days in a year y and Y is the total number of years. The overall cumulative return of the strategy denoted as R^{oc}, is computed by summing the cumulative returns across all individual years as:

$$R^{oc} = \sum_{y=1}^{Y} R_y^c, \quad \text{for } y = 1, 2, \ldots, Y. \tag{11}$$

By applying these formulas to each trade, traders can evaluate the individual trade performances and the cumulative return over multiple trades. This quantitative analysis serves as a key metric for assessing the overall effectiveness and profitability of the pairs trading strategy.

5) Backtesting and risk management by profit/loss: Now, that a framework has been established for calculating returns and cumulative returns in the PTS-R, the next crucial steps involve backtesting and implementing risk management measures based on profit and loss. The backtesting is a critical process in trading and investment strategy development. It involves evaluating a trading strategy using historical market data to assess how it would have performed in the past. The strategy performance is backtested by various parameters like initial capital, final portfolio value (training), final portfolio value (test), profit/Loss (training), and profit/Loss (test). The profit/loss [14] metric serves as an evaluation parameter for the strategy's performance, providing insights into the overall financial gain or loss experienced during the specified trading period and is calculated as:

$$Profit/Loss = Final\ Capital - Initial\ Capital. \tag{12}$$

These insights aid in refining the strategy and guide traders in making well-informed decisions during the live execution of the trading strategy by using Alpaca API, QuantConnect, AlgoTest, etc.

B. SBPTS

The methodology employed to develop and validate the SBPTS is explained in this section. It encompasses data collection, sector classification, pair selection, trading strategy implementation, and performance evaluation. This approach seeks to enhance the trading performance by considering the sector-related factors when forming the pairs. The innovation continues with the evolution of the two variants, SBPTS-C and SBPTS-R which represent the improvements over the traditional methods.

The basic outline of the overall methodology of the SBPTS is given in Table II.

TABLE II
OUTLINE OF THE BASIC METHODOLOGY OF THE SBPTS

Steps	
1	Classify the normalized stocks into the respective sector
2	Identify the best-performing sector
3	Select the stock pairs from the best-performing sector
4	Execute the trading strategy outlined in I across steps 2 to 5

1) Classify the normalized stocks into their respective sector: This is an important step in organizing and analyzing the data in the SBPTS. Initially, the sector information is retrieved for each stock file from the Alpha Vantage API, and this data is used to classify the stock files into the respective sectors by applying a classification algorithm. Various ML and DL algorithms can be used for the classification. It's important to consider the specific characteristics of the dataset, the available computational resources, and the desired level of interpretability while selecting an ML or DL algorithm. The ML algorithm, Support Vector Machine (SVM) with Radial Basis Kernel is well-suited for time series data (stock data) classification due to its ability to handle non-linear patterns, high-dimensional spaces, and small sample sizes while being robust to overfitting [28].

The classification process starts with an initialization step, establishing an empty set $\mathcal{S}$ of C sectors denoted as $\{s_c|c = 1, 2, 3, \ldots, C\}$ that will be used to store the sectors classified during subsequent steps. The goal is to classify stocks from a given set of stock files $\mathcal{F}$ containing data file as D''_{ij} into their respective sectors. The process unfolds through several key steps. Initially, sector information for each stock file f_m is retrieved using Alpha Vantage API. Subsequently, an SVM model is employed to train on D''_{ij} and classify stocks into their respective sectors. Following the classification, the identified sectors are assigned to the corresponding stocks within the set F. This SVM model is ready for use in classifying historical stock data into their respective sectors based on the learned relationships during the training phase.

In conclusion, the final output is the comprehensive set S, which consolidates all the classified sectors obtained through the SVM-based classification of the provided stock files of the set F. This procedure facilitates the organization and categorization of stocks into distinct sectors for further analysis and interpretation.

2) Identify the best performing sector: Investors and traders often consider various factors to make informed decisions when selecting a sector for investment or trading. There are various sector selection measures used to identify the performance of the sector[29]–[33]. These metrics collectively aid in discerning the sector that exhibits optimal performance, guiding strategic decisions, and enhancing overall portfolio management. The common are explained below.

The average sector return (ASR) is a measure that expresses the stock's increase in value over a designated period. The identification of the sector that has generated the highest returns over time is obtained by comparing the annual returns of different sectors [17], [27].

The ASR of each sector is obtained by calculating the overall cumulative returns of the individual stock, $R^{oc}_{f_m}$, calculated in the (5) for all the files present in each sector s_c. Moreover, the sector return, R_{s_c} is the average return across all individual files within the sector and is calculated as:

$$R_{s_c} = \frac{1}{N_{ms}} \sum_{m=1}^{N_{ms}} (R^{oc}_{f_m}), \quad f_m \in s_c, s_c \in S, \tag{13}$$

where s_c is the sector from the set S and N_{ms} is the total number of stock files in the sector s_c. The higher annual returns indicate stronger performance and may attract more investor interest.

The Sharpe ratio (SR) is a measure used to evaluate the risk-adjusted return of the sector or portfolio [23]. This ratio helps investors to understand the return of an investment relative to its risk. It is calculated for each sector s_c, denoted by ψ_{s_c} as:

$$\psi_{s_c} = \frac{R_{s_c} - R_f}{\sigma_{s_c}}, \quad s_c \in S, \tag{14}$$

where R_f is the is the risk-free rate. The risk-free rate refers to the hypothetical return on an investment with no risk of financial loss, typically represented by government bonds or similar low-risk securities. If $\psi_{s_c} > 0$, it means that the returns earned by an investment exceed the risk-free rate, and generate positive returns above what could have been earned with no risk. When $\psi_{s_c} < 0$, the returns earned by an investment do not adequately compensate for the level of risk taken and the investment is not performing well enough to justify the associated risk. It is a widely used metric in finance for comparing the performance of different sectors and strategies [17], [26].

The beta value represented as β, is a measure of a stock or sector's sensitivity to market movements [28]. A $\beta > 1$ indicates that the sector is more volatile than the overall market, while $\beta < 1$ indicates lower volatility. The β of sector s_c with respect to the market β_{s_c} is defined by:

$$\beta_{s_c} = \frac{Cov(R_{s_c}, R_{mk})}{Var(R_{mk})}, \quad s_c \in S, \tag{15}$$

where R_{mk} is the market return, $Cov(R_{s_c}, R_{mk})$ is the covariance between the returns of sector s_c and the market return, and $Var(R_{mk})$ is the variance of market returns. The R_{mk} indicates how much an investor would have gained or lost by investing in the overall stock market over a specific period. It serves as a benchmark for evaluating the performance of individual stocks, sectors, or portfolios relative to the broader market.

The price-to-earnings (P/E) ratio is a commonly used evaluation metric in finance that measures the current market price of a company's stock relative to its earnings per share (EPS) [33]. It provides insights into how much investors are willing to pay for each unit of earnings generated by the company[2].

[2]https://groww.in/p/pe-ratio.

The earnings per share for stock f_m in sector s_c is ε_{f_m} and calculated as:

$$\varepsilon_{f_m} = \frac{R_{s_s}}{N_{f_m}}, \quad f_m \in s_c, s_c \in S, \tag{16}$$

where R_{s_c} represents the average returns of the sector s_c and N_{f_m} represents the total number of outstanding shares for stock f_m. The P/E ratio for an individual stock f_m, ρ_{f_m} in sector s_c is:

$$\rho_{f_m} = \frac{P_{f_m}}{\varepsilon f_m}, \quad f_m \in s_c, s_c \in S, \tag{17}$$

where P_{f_m} represents the market price of the individual stock f_m. It is determined by the supply and demand dynamics in the stock market and can fluctuate throughout the trading day based on various factors such as investor sentiment, company performance, economic conditions, and market trends. The average price-to-earnings ratio for the sector s_c is ρ_{s_c} given by:

$$\rho_{s_s} = \frac{1}{N_{ms}} \sum_{m=1}^{N_{ms}} \rho_{f_m}, \quad f_m \in s_c, s_c \in S, \tag{18}$$

where N_{ms} represents the total number of stocks in sector s_c. The metrics discussed, ASR, SR, β value, and P/E ratios serve as key indicators for evaluating and identifying the best-performing sectors. The identification of the best-performing sector can be crucial for investors and traders looking to allocate their capital effectively. To accomplish this goal, this paper proposes a method that introduces a systematic and data-driven strategy called the Total Score Weighting Method (TSWM). It is an approach used to select the best-performing sector by assigning weights to key performance metrics and calculating a total score for each sector. To implement TSWM effectively, it is essential to establish a hierarchy among the metrics by assigning priorities. The primary goal of the investor is to generate high profits with minimum risk. The priority assigned to these metrics is based on their significance in evaluating the performance and risk associated with a sector. Let $P_{R_{s_c}}$, $P_{\psi_{s_c}}$, $P_{\beta_{s_c}}$ and $P_{\rho_{s_c}}$ represent the priorities assigned to ASR, SR, β value, and P/E ratio respectively, for sector s_c.

The process of assigning priority values involves subjective judgment, expert opinion, or predefined criteria. In this approach, the priority values for each metric are determined through predefined criteria based on the condition $P_{R_{s_c}} > P_{\psi_{s_c}} \wedge P_{\beta_{s_c}} > P_{\rho_{s_c}}$.

The fundamental steps in the TSWM involve assigning weights to key metrics based on their priority hierarchy, normalizing the weights to ensure they sum up to 1, calculating the total score for each sector, and then identifying the sector with the highest total score as the best-performing one. This work is illustrated in the algorithm 1. The given rules provide a structured approach to assigning weights to the set of sectors S based on their performance in different metrics, with varying priority levels.

Let W be a weight matrix of size 4×1, where each element corresponds to the weight assigned to a specific metric denoted as $w_{R_{s_c}}$, $w_{\psi_{s_c}}$, $w_{\beta_{s_c}}$, and $w_{\rho_{s_c}}$. Then, the weight assignment rules can be expressed as follows:

$$P_{total} = \sum_{x \in X} P_x, \quad \text{for } X = \{R_{s_c}, \psi_{s_c}, \beta_{s_c}, \rho_{s_c}\} \tag{19}$$

$$W_x = \frac{P_x}{P_{total}}, \quad \text{for } X = \{R_{s_c}, \psi_{s_c}, \beta_{s_c}, \rho_{s_c}\} \tag{20}$$

The normalization should ensure that the sum of weights, W_x, for all metrics in set X equals 1. This is commonly represented by the equation $\sum_{x \in X} W_x = 1$. This ensures that higher-priority metrics receive greater weights, contributing more significantly to the overall evaluation. By employing the defined rules, following algorithm systematically computes the total score for each sector, facilitating the selection of the best-performing sector having highest total score value.

Algorithm 1 Identify the best-performing sector

```
Input: Set of sectors S = {s_c | c = 1, 2, 3, ..., C}
Output: best_sector
TS_max = −∞
for each sector s_c in S do
    Read R_{s_c}, ψ_{s_c}, β_{s_c}, ρ_{s_c}
    Calculate Total Score TS_{s_c} for sector s_c:
    TS_{s_c} = R_{s_c} · w_{R_{s_c}} + ψ_{s_c} · w_{s_c} + β_{s_c} · w_β + ρ_{s_c} · w_{ρ_{s_c}}
    Identify best-performing sector:
    if TS_{s_c} > TS_max then
        TS_max = TS_{s_c}
          best_sector = s_c
    end if
end for
Print the best performing sector: best_sector
```

3) Select the stock pairs from the best performing sector: This step involves a dual consideration of the correlation method and return-based criteria. Various correlation and distance-based measures can be employed to assess the relationships between stocks like Pearson correlation [3], Sum of Squares Difference (SSD) [14], Dynamic Time Warping (DTW) [22], Cross-Correlation Type Measure (CCT) [21], and Dynamic Cross-Correlation Type Measure (DCCT) [14].

The choice of measure depends on the characteristics of the data and the trader's preferences. It's important to note that no single measure is universally better, and the choice may also depend on the trading strategy being employed. The Pearson Correlation [14] is frequently used as a measure for pair selection. It measures the strength and direction of the linear association between the variables and ranges between -1 and 1. A value of 1 indicates a perfect positive linear correlation, -1 indicates a perfect negative linear correlation, and 0 indicates no linear correlation. The Pearson correlation coefficient, γ_{f_p,f_q} between the adjusted closing prices of stock files f_p and f_q from the set F over all trading days, d, is calculated as:

$$\gamma_{f_p,f_q} = \frac{\sum_{i=1}^{d}(D''_{f_{p_{ij,i}}} - \overline{D''_{f_{p_{ij}}}})(D''_{f_{q_{ij,i}}} - \overline{D''_{f_{q_{ij}}}})}{\sqrt{\sum_{i=1}^{d}(D''_{f_{p_{ij,i}}} - \overline{D''_{f_{p_{ij}}}})^2}\sqrt{\sum_{i=1}^{d}(D''_{f_{q_{ij,i}}} - \overline{D''_{f_{q_{ij}}}})^2}}, \tag{21}$$

where $D''_{f_{p_{ij,i}}}$ and $D''_{f_{q_{ij,i}}}$ represents the adjusted closing prices of stock file f_p and f_q at the i^{th} trading day respectively. $\overline{D''_{f_{p_{ij}}}}$ and $\overline{D''_{f_{q_{ij}}}}$ represents the mean adjusted closing prices of stock file f_p and f_q for all trading days, d respectively. The Pearson correlation coefficient is computed for all stock files within set F to determine the pair exhibiting the highest correlation.

Though these methods have been widely used and can be effective, they may not always capture the dynamic nature of market conditions. The relationships between individual stocks can change, and the traditional methods may not adapt quickly to shifts in market dynamics. Hence, the return-based pair selection method explained in the section III-A1 is proposed which takes a broader perspective by focusing on sector-level performance.

C. Variants of the SBPTS

This sub-section deals with the two variants of the SBPTS which differ in the pair selection criteria. The pair selection in SBPTS-C is done by correlation measure whereas in SBPTS-R, it is done by considering returns of individual stock.

IV. Experimentation and Result Analysis

This section deals with the various results obtained by the proposed strategy on historical data and discusses the insights. All the experiments were conducted by applying the algorithms and formulas outlined in section III. The experimentation involved implementing a PTS, PTS-R, SBPTS-C, and SBPTS-R using the 30 input stock data files randomly collected from Sensex30 and NASDAQ each from Yfinance API and AlphaVantage API. For the experimentation, a dataset covering a period of 10 years, from January 1, 2013, to January 1, 2023, was employed. In the forthcoming sub-sections, we delve into the experimental details of various strategy variants, along with the basic pairs trading strategy.

TABLE III
An Overview of the Key Components of the PTS and SBPTS

Steps	
1	Pair formation by using Pearson correlation measure and cumulative returns as per the selected strategy (correlation threshold coefficient = 0.8)
2	Define entry threshold = 1.0 and exit threshold = 0.5 while calculating
3	Calculate the short and long positions based on z-score value (-1 for short, sell signal and 1 for long, buy signal)
4	Calculate cumulative returns of the PTS, PTS-R, SBPTS-C, and SBPTS-R returns
5	If the z-score crosses the exit threshold, the corresponding positions are changed to zero, closing the market position

A. PTS

The basic PTS is implemented in this paper which involves selecting pairs of stocks by using Pearson correlation. Establishing a threshold coefficient is essential to employ the Pearson correlation. Generally, its value is close to +1 (this value is 0.8 i.e. 80%). As per the threshold used in the experimentation, 182 stock pairs for Sensex30 and 161 for NASDAQ datasets have been identified as having a value greater than 0.8. Table IV showcases the upper 10 rows with the highest correlation coefficient values among the identified stock pairs. Utilizing the findings presented in Table IV over the Sensex30 dataset, the selection process has identified a stock pair as (AXISBANK.NS, ASIANPAINT.NS). Similarly for the NASDAQ dataset, the (AAPL, JNJ) stock pair is selected as a potential candidate for trading. After identifying the stock pairs, a pairs trading strategy is implemented as outlined in Table III.

B. PTS-R

In this modified strategy, the cumulative returns of all the stocks by using daily returns which represent the percentage change in the stock's price from one day to the next are calculated. The pairs of stocks will be identified as having the highest overall cumulative returns over the historical period. Fig. IV-B and Fig. IV-B show the PTS-R analysis for Sensex30 and NASDAQ datasets respectively. In the provided graph

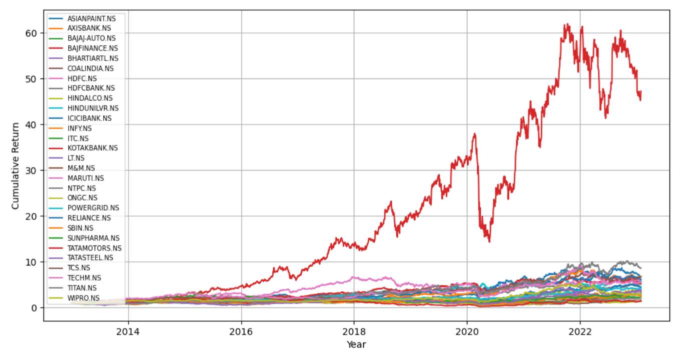

Fig. 1. PTS with Return-Based Pair Selection: Cumulative Returns of Sensex30 Dataset Stocks Over Historical Period

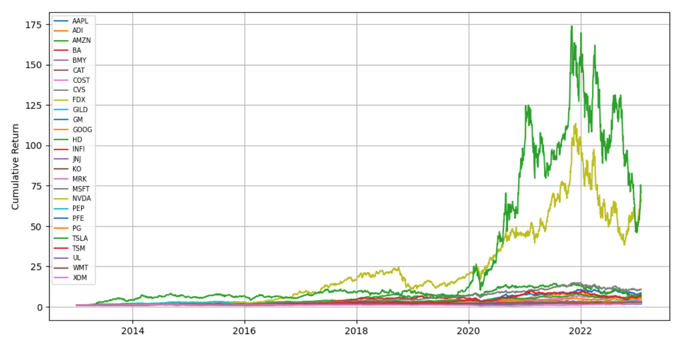

Fig. 2. PTS with Return-Based Pair Selection: Cumulative Returns of NASDAQ Dataset Stocks Over Historical Period

(Fig. IV-B), the stock pair (BAJFINANCE.NS, TITAN.NS) from the Sensex30 dataset exhibits the highest cumulative returns, leading them to be selected as the pair for PTS-R. Similarly, as depicted in the graph (Fig. IV-B), the stock pair (NVDA, TSLA) is identified as the pair for PTS-R within the NASDAQ dataset. Further, the trading strategy is applied and the final profit/loss will be calculated as mentioned in Table III.

TABLE IV
PTS-Correlation calculation for each stock from January 2013 to January 2023

Sensex30 Dataset			NASDAQ Dataset		
Stock 1	Stock 2	Correlation value	Stock 1	Stock 2	Correlation value
AXISBANK.NS	**ASIANPAINT.NS**	**0.9897**	**AAPL**	**JNJ**	**0.9831**
SBIN.NS	LT.NS	0.9826	GOOG	AMZN	0.9826
BAJFINANCE.NS	COALINDIA.NS	0.9785	COST	MRK	0.9823
LT.NS	BAJAJ-AUTO.NS	0.9769	FDX	JNJ	0.9808
SBIN.NS	COALINDIA.NS	0.9760	AMZN	JNJ	0.9802
ITC.NS	BAJAJ-AUTO.NS	0.9750	AAPL	INFI	0.9794
LT.NS	COALINDIA.NS	0.9747	INFI	JNJ	0.9780
SBIN.NS	BAJFINANCE.NS	0.9730	GOOG	JNJ	0.9773
TITAN.NS	COALINDIA.NS	0.9706	INFI	MRK	0.9731
SBIN.NS	BAJAJ-AUTO.NS	0.9678	FDX	PFE	0.9729

C. SBPTS

The classification of the input stock files into their respective sectors is an important step in organizing and analyzing the data in the SBPTS. This paper uses an SVM algorithm to classify the normalized stock data into respective sectors. By training an SVM model on historical data and relevant features, it can learn patterns and relationships that enable it to classify stock pairs into their respective sectors. Among the 30 input stock data files, 24 (80%) stock files are allocated for training, leaving 6 (20%) designated for testing. This enables to classification of the data into their respective sectors accurately as shown in Table V.

TABLE V
Sector Classification

Dataset	**Sector**	**Number of stocks**
Secsex30	Communication services	4
	Energy	2
	Information Technology	4
	Pharmaceuticals	1
	Industrial	1
	Automobiles	7
	Materials	7
	Utilities	2
	Financial	2
NASDAQ	Technology	8
	Manufacturing	6
	Trade and services	5
	Healthcare	6
	Consumer goods	5

After classifying the sock data into respective sectors, the best-performing sector is identified by using AR, SR, β value, and P/E ratio. The results for selecting the best-performing sector are shown in Table VI. When applying the algorithm outlined in the section III-B1 to the Sensex30 dataset, the Materials sector emerges with the highest total score. As a result, it is chosen as the best-performing sector. Similarly, for the NASDAQ dataset, the Technology sector achieves the highest total score and is selected as the best-performing sector. The TSWM offers a systematic way to select the best-performing sector, enhancing the decision-making process for investors looking to achieve their financial goals. After identifying the best-performing sector, apply the strategy outlined in Table III. Referring to table VII and focusing on SBPTS correlation analysis (SBPTS-C), the chosen stock pairs encompass SHREECEM.NS and GRASIM.NS from the Sensex30 dataset, as well as AAPL and CRM from the NASDAQ dataset. These all pairs exhibit the highest correlation values and hence selected as the better candidates for trading.

D. SBPTS-C

In this proposed variant of SBPTS, the Pearson correlation method is used to select the stock pairs from the best-performing sector shown in the Table VII. The stock pairs (SHREECEM.NS, GRASIM.NS) from the Sensex30 dataset and (AAPL, CRM) from the NASDAQ dataset are selected for trading. These all pairs exhibit the highest correlation values and hence are selected as the better candidates for trading.

E. SBPTS-R

In this approach, the stock pairs having the highest cumulative returns from the best performing over the historical period are selected. Fig. IV-E and Fig. IV-E showcase the pairs of stocks (JSWSTEEL.NS, SHREECEM.NS) for Sensex30 and (NVDA, MSFT) for NASDAQ are identified for trading.

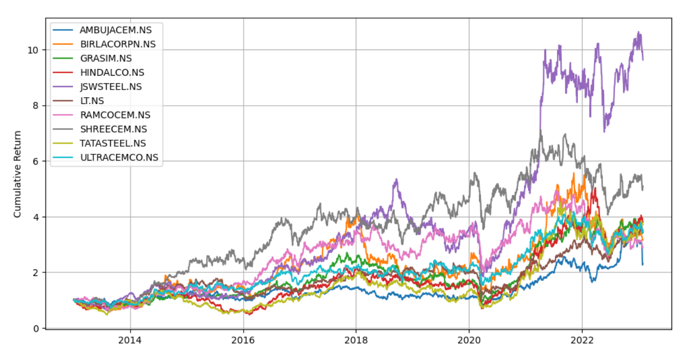

Fig. 3. SBPTS with Return-Based Pair Selection: Cumulative Returns of Sensex30 Dataset Stocks Over Historical Period

Four pairs are identified from PTS, PTS-R, SBPTS-C and SBPTS-R shown in the Table VIII. Subsequently, the trading strategy is implemented by using these pairs, and the resulting profit/loss is computed. The comparison between PTS and SBPTS is outlined using the cumulative returns achieved from

TABLE VI
PERFORMANCE OF SECTORS FOR DIFFERENT MEASURE

Sector	Annual Returns	Sharp Ratio	Beta Value	P/E Ratio	Total Score
SENSEX30 Dataset					
Communication services	5.2278	-3.2483	0.2989	1.0006	3.2790
Energy	5.6630	-2.7627	0.3977	1.0009	3.2989
Information Technology	2.9495	-2.5682	0.3332	1.0005	2.7150
Healthcare	2.3792	-2.6072	0.3235	1.0004	2.0959
Industrials	3.0278	-2.7750	0.4130	1.0000	1.6658
Automobiles	0.8986	-1.8658	0.2269	1.0009	0.2606
Utilities	2.7450	-2.2102	0.3225	1.0015	1.8588
Materials	5.7997	-2.8008	0.2823	1.0001	**4.2813**
Financials	1.7240	-2.2856	0.3213	1.0000	0.7597
NASDAQ Dataset					
Technology	0.4001	0.7333	0.5065	1.0118	**0.81036**
Manufacturing	0.3821	0.5903	1.2304	1.0008	0.71714
Trade and Services	0.1108	0.7205	0.8250	1.0016	0.55374
Healthcare	0.0029	0.5745	0.5745	1.0024	0.31744
Consumer Goods	-0.1064	0.5423	0.5423	1.0032	0.37500

TABLE VII
SBPTS-CORRELATION CALCULATION FOR EACH STOCK FROM JANUARY 2013 TO JANUARY 2023

Materials Sector from Sensex30 Dataset			Technology Sector from NASDAQ Dataset		
Stock1	Stock2	Correlation value	Stock1	Stock2	Correlation value
GRASIM.NS	**SHREECEM.NS**	**0.9785**	**AAPL**	**CRM**	**0.9896**
AMBUJACEM.NS	BIRLACORPN.NS	0.9701	GOOG	NVDA	0.9821
ULTRACEMCO.NS	GRASIM.NS	0.9512	ADI	CRM	0.9808
ULTRACEMCO.NS	SHREECEM.NS	0.9500	TSM	CRM	0.9800
ULTRACEMCO.NS	LT.NS	0.9567	ADI	CSCO	0.9758
SHREECEM.NS	LT.NS	0.9434	TSM	ADI	0.9702
HINDALCO.NS	RAMCOCEM.NS	0.9400	AAPL	ADI	0.9612
TATASTEEL.NS	ULTRACEMCO.NS	0.9356	CRM	CSCO	0.9601
BIRLACORPN.NS	LT.NS	0.9297	ADI	ADBE	0.9600
TATASTEEL.NS	LT.NS	0.9220	CRM	ADBE	0.9567

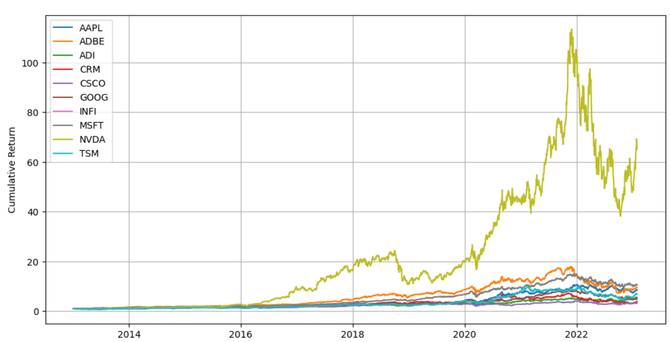

Fig. 4. SBPTS with Return-Based Pair Selection: Cumulative Returns of NASDAQ Dataset Stocks Over Historical Period

each approach shown in Fig. IV-E. Comparing the overall returns of all four strategies, SBPTS-R stands out as the strategy with the highest cumulative return, suggesting that within the NASDAQ dataset, the SBPTS has performed exceptionally well. While PTS-R also shows favorable returns, SBPTS-C demonstrates substantial growth, indicating the effectiveness of the sector-based approach in capturing potential profit opportunities. Explicitly comparing SBPTS-R and SBPTS-C within the same sector-based framework would offer a more comprehensive understanding of their relative performance and validate the effectiveness of the chosen strategy. The comparison underscores the advantages of the sector-based approach over traditional methods, providing a clearer validation of the proposed strategy based on accumulated returns and correlation within a consistent sectoral context as shown in Fig. IV-E.

TABLE VIII
SELECTED PAIRS FOR FOUR STRATEGIES

Strategy	**Sensex30**	**NASDAQ**
PTS	(AXISBANK.NS, ASIANPAINT.NS)	(AAPL, JNJ)
PTS-R	(BAJFINANCE.NS, TITAN.NS)	(NVDA, TSLA)
SBPTS-C	(GRASIM.NS, SHREECEM.NS)	(AAPL, CRM)
SBPTS-R	(JSWSTEEL.NS, SHREECEM.NS)	(NVDA, MSFT)

F. Validation and robustness

The validation and robustness analysis of the SBPTS involves a comprehensive evaluation of the strategy's effectiveness and adaptability under various market conditions. This study aims to verify the performance of SBPTS and ascertain its reliability across different scenarios. This paper leverages the Python backtrader library and integrates it with Alpaca's

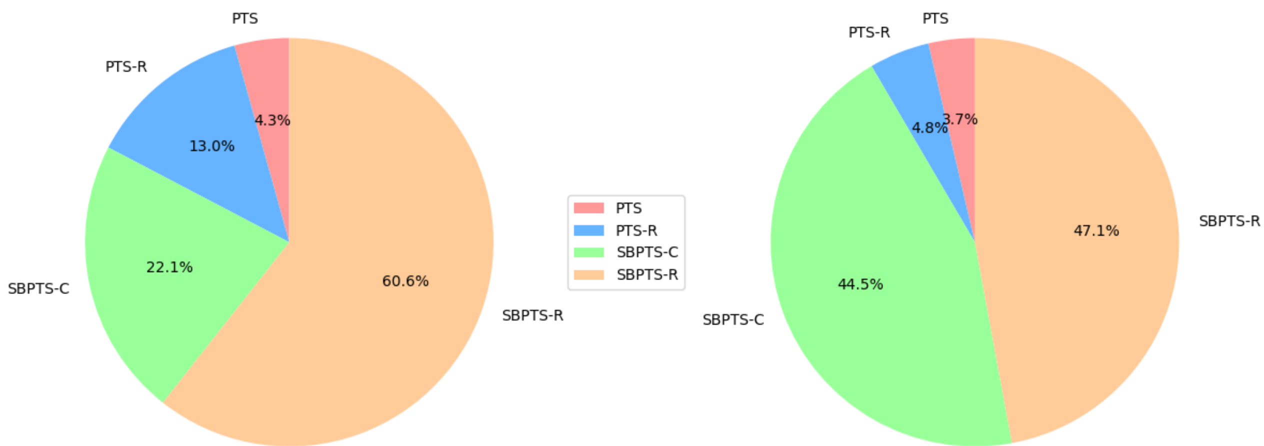

Fig. 5. Comparison of the overall cumulative returns achieved by PTS, PTS-R, SBPTS-C, and SBPTS-R for Sensex30 and NASDAQ datasets, respectively, over the historical period of 10 years.

paper trading API to conduct thorough backtesting. It is based on predefined parameters, including a window size, an entry threshold, an exit threshold, and an initial capital. The out-of-sample testing is performed to assess how well the proposed strategy performs on historical data not used in the strategy development process. It provides a more reliable estimate of the strategy's potential effectiveness in live markets and helps avoid the pitfalls of over-optimization based on training data. The overall performance of the trading strategy on the selected stock pairs (mentioned in Table VIII) is promising. It successfully navigated both the training and test phases, consistently generating positive returns by considering the initial capital of Rs. 100,000. The analysis examined the impact of transaction costs, specifically a slippage rate of 0.1% per trade, on the returns of the pairs trading strategy. With transaction costs included, returns decreased slightly from 1054.90% to 1051.0%. This 3.89% reduction highlights the significant effect transaction costs have on trading strategy profitability. It stresses the importance of accounting for transaction costs to assess a strategy's performance accurately. These findings emphasize the need for traders to consider transaction costs to evaluate and optimize the strategies for real-world implementation effectively.

V. Conclusions

The proposed SBPTS with a return-based pair selection technique demonstrates promising potential in exploiting price relationships within the same sector to achieve favorable trading outcomes. The SBPTS strategy offers several advantages. It leverages sector-specific information to enhance the accuracy of pair selection, increasing the potential for profitable trades. By concentrating on specific sectors, SBPTS aligns with the concept that stocks within the same sector tend to exhibit similar market behaviors, which can lead to more effective trading opportunities. SBPTS, by capitalizing on sector-specific correlations and price relationships, has shown the ability to generate substantial returns, making it a promising approach for pairs trading in different market environments. After conducting backtesting on the proposed SBPTS approach, the results indicate this strategy's superior effectiveness and robustness. The future scope lies in enhancing the effectiveness and applicability of sector-based pairs trading strategies in diverse market environments. This can be achieved by addressing limitations such as missed stock opportunities, lack of diversification, and sector-specific risk. By incorporating a hybrid strategy that combines sector-based and cross-sector approaches, future sector-based strategies can be significantly improved. The cross-sector approach involves identifying synergies and opportunities across different sectors, leveraging strengths from one sector to complement weaknesses in another. This comprehensive approach allows for capturing missed opportunities among eligible stocks, promoting diversification, and mitigating sector-specific risks.

References

[1] Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. "Pairs Trading: Performance of a relative-value arbitrage rule." The Review of Financial Studies 19.3 (2006): 797-827.

[2] Vidyamurthy, Ganapathy. "Pairs Trading: quantitative methods and analysis." Vol. 217. John Wiley & Sons, 2004.

[3] Sarmento, Simão Moraes, and Nuno Horta. "Enhancing a pairs trading strategy with the application of machine learning." Expert Systems with Applications 158 (2020): 113490.

[4] Figueira, Miguel, and Nuno Horta. "Machine Learning-Based Pairs Trading Strategy with Multivariate." Available at SSRN 4295303.

[5] Chang, Victor, et al. "Pairs trading on different portfolios based on machine learning."" Expert Systems 38.3 (2021): e12649.

[6] Lu, Jing-You, et al. "Structural break-aware pairs trading strategy using deep reinforcement learning." The Journal of Supercomputing 78.3 (2022): 3843-3882.

[7] Brim, Andrew. "Deep reinforcement learning pairs trading." (2019).

[8] Fil, Miroslav. “Gold Standard Pairs Trading Rules: Are They Valid?.” arXiv preprint arXiv:2010.01157 (2020).
[9] Krauss, Christopher, Xuan Anh Do, and Nicolas Huck. “Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500.” European Journal of Operational Research 259.2 (2017): 689-702.
[10] Wang, Cheng, Patrik Sandås, and Peter Beling. “Improving pairs trading strategies via reinforcement learning.” 2021 International Conference on Applied Artificial Intelligence (ICAPAI). IEEE, 2021.
[11] Krauss, Christopher. “Statistical arbitrage pairs trading strategies: Review and outlook.” Journal of Economic Surveys 31.2 (2017): 513-545.
[12] Flori, Andrea, and Daniele Regoli. “Revealing pairs-trading opportunities with long short-term memory networks.” European Journal of Operational Research 295.2 (2021): 772-791.
[13] Blázquez, Mario Carrasco, and Camilo Prado Román. “Pairs trading techniques: An empirical contrast.” European Research on Management and Business Economics 24.3 (2018): 160-167.
[14] Gupta, Kartikay, and Niladri Chatterjee. “Selecting stock pairs for pairs trading while incorporating lead–lag relationship.” Physica A: Statistical Mechanics and its Applications 551 (2020): 124103.
[15] Nabipour, Mojtaba, et al. “Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis.” IEEE Access 8 (2020): 150199-150212.
[16] Wang, Jai-Jen, Jin-Ping Lee, and Yang Zhao. “Pair-trading profitability and short-selling restriction: Evidence from the Taiwan stock market.” International Review of Economics and Finance 55 (2018): 173-184.
[17] Han, Chulwoo, Zhaodong He, and Alenson Jun Wei Toh. “Pairs trading via unsupervised learning.” European Journal of Operational Research 307.2 (2023): 929-947.
[18] Varghese, Nelson Vithayathil, and Qusay H. Mahmoud. “A Hybrid Multi-Task Learning Approach for Optimizing Deep Reinforcement Learning Agents.” IEEE Access 9 (2021): 44681-44703.
[19] Sarmento, Simão Moraes, and Nuno Horta. “Enhancing a pairs trading strategy with the application of machine learning.” Expert Systems with Applications 158 (2020): 113490.
[20] Yan, Tingjin, and Hoi Ying Wong. “Equilibrium pairs trading under delayed cointegration.” Automatica 144 (2022): 110498.
[21] Gupta, Kartikay, and Niladri Chatterjee. “Financial time series clustering.” Information and Communication Technology for Intelligent Systems (ICTIS 2017)-Volume 2 2. Springer International Publishing, 2018.
[22] Berndt, Donald J., and James Clifford. “Using dynamic time warping to find patterns in time series.” Proceedings of the 3rd international conference on knowledge discovery and data mining. 1994.
[23] Kim, Taewook, and Ha Young Kim. “Optimizing the pairs-trading strategy using deep reinforcement learning with trading and stop-loss boundaries.” Complexity 2019 (2019): 1-20.
[24] Hoffmann, Marc, Mathieu Rosenbaum, and Nakahiro Yoshida. “Estimation of the lead-lag parameter from non-synchronous data.” (2013): 426-461.
[25] Rad, Hossein, Rand Kwong Yew Low, and Robert Faff. “The profitability of pairs trading strategies: distance, cointegration, and copula methods.” Quantitative Finance 16.10 (2016): 1541-1558.
[26] Fil, Miroslav, and Ladislav Kristoufek. “Pairs trading in cryptocurrency markets.” IEEE Access 8 (2020): 172644-172651.
[27] Chang, Victor, et al. “Pairs trading on different portfolios based on machine learning.” Expert Systems 38.3 (2021): e12649.
[28] Usmani, Mehak, et al. “Stock market prediction using machine learning techniques.” 2016 3rd international conference on computer and information sciences (ICCOINS). IEEE, 2016.
[29] Violita, Cynthia Eka, and S. Soeharto. “Stock liquidity and stock return.”Jurnal Bisnis dan Manajemen 3.2 (2019): 111-122.
[30] https://www.scribd.com/document/345894135/Risk-and-Return.
[31] Zivot, Eric, and Jiahui Wang. Modeling financial time series with S-PLUS. Vol. 2. New York: Springer, 2006.
[32] Chan, Ngai Hang. Time series: Applications to finance with R and S-Plus. Vol. 837. John Wiley & Sons, 2011.
[33] R. Sari, “Analysis of the Effect of Earnings per share, Price earning ratio and Price to book value on the stock prices of state-owned enterprises”, GRFM, vol. 1, no. 1, pp. 25 - 32, Mar. 2021, https://doi.org/10.52970/grfm.v1i1.117.
[34] Naccarato, Alessia, Andrea Pierini, and Giovanna Ferraro. “Markowitz portfolio optimization through pairs trading cointegrated strategy in long-term investment.” Annals of Operations Research 299.1 (2021): 81-99.
[35] Yang, Jen-Wei, et al. “Pairs trading: The performance of a stochastic spread model with regime switching-evidence from the S&P 500.” International Review of Economics & Finance 43 (2016): 139-150.
[36] García García, Fernando, et al. “What is the cost of maximizing ESG performance in the portfolio selection strategy? The case of The Dow Jones Index average stocks.” Entrepreneurship and Sustainability Issues 9.4 (2022): 178-192.

Pranjala G. Kolapwar received a B.E. in computer science and engineering from Gadgebaba University, Amravati, Bharat, and an M.Tech. in Computer Science and Engineering (CSE) from Swami Ramanand Teerth Marathwada University (SRTMU), Nanded, Bahrat and pursuing the Ph.D. from the same university. She is working as an assistant professor in the Department of CSE, SGGS Institute of Engineering & Technology, Nanded. She has 18+ years of teaching experience and has published research papers in national and international journals and conferences. Her primary research area includes Machine learning, deep learning, and stock management.

Uday V. Kulkarni received a doctoral degree from SRTMU, Nanded in 2002. He has a total of 30 years of teaching experience and is currently working as a Professor in the Department of CSE at SGGSIE&T, Nanded. His research interests include Fuzzy Neural Networks and Pattern Classification. He is a recipient of a national-level gold medal in the Computer Engineering Division for his research paper Fuzzy Hypersphere Neural Network Classifier published in the Journal of the Institution of Engineers in 2004 and the best paper award for the research paper presented in an international conference held at Imperial College London, U.K., 2014. He has published more than fifty research papers in the field of Neural Networks, Fuzzy Logic, and Hybrid Computing Systems in reputed conferences and journals.

Jaishri M. Waghmare received a B.E. degree in Computer Science and Engineering (CSE) from Swami Ramanand Teerth Marathwada University (SRTMU), Nanded, and the M.Tech. degree in CSE from the Indian Institute of Technology (IIT) Mumbai, Bharat, and a Ph.D. degree in CSE from SRTMU, Nanded, in 2000, 2009, and 2018, respectively. She is working as an associate professor in the Department of CSE, SGGS Institute of Engineering & Technology, Nanded. She has 23+ years of teaching experience and has published many research papers in national and international journals and conferences. Her areas of interest are Intelligent systems, neural networks, and fuzzy logic.