Jung, Nak Hyun; Oh, Tae Yeon; Kim, Kang Hee

Article

AI-based pairs trading strategies: A novel approach to stock selection

Global Business & Finance Review (GBFR)

Provided in Cooperation with:
People & Global Business Association (P&GBA), Seoul

Suggested Citation: Jung, Nak Hyun; Oh, Tae Yeon; Kim, Kang Hee (2024) : AI-based pairs trading strategies: A novel approach to stock selection, Global Business & Finance Review (GBFR), ISSN 2384-1648, People & Global Business Association (P&GBA), Seoul, Vol. 29, Iss. 7, pp. 1-15, https://doi.org/10.17549/gbfr.2024.29.7.1

This Version is available at:
https://hdl.handle.net/10419/306024

GLOBAL BUSINESS & FINANCE REVIEW, Volume. 29 Issue. 7 (AUGUST 2024), 1-15
pISSN 1088-6931 / eISSN 2384-1648 | Https://doi.org/10.17549/gbfr.2024.29.7.1
ⓒ 2024 People and Global Business Association

GLOBAL BUSINESS & FINANCE REVIEW

www.gbfrjournal.org
for financial sustainability and people-centered global business

AI-based Pairs Trading Strategies: A Novel Approach to Stock Selection

Jung Nak Hyun[a], Oh Tae Yeon[b], Kim Kang Hee[c†]

[a]*Seoul Business School, aSSIST University, Seoul, Korea*
[b]*Seoul AI School, aSSIST University, Seoul, Korea*
[c]*LG Chem Life Science Division 6σ Master Black Belt, Seoul, Korea*

A B S T R A C T

Purpose: This study aims to explore the optimization of Stock Pairs Trading Strategies' performance using AI techniques, with a focus on accurately evaluating stock similarities and selecting the most suitable pairs.
Design/methodology/approach: A variety of AI models, including Autoencoders (AE), Vector Embeddings (VE), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), are utilized to assess the similarity between stocks, which is crucial in the stock pairs selection process for implementing the Pairs Trading strategy.
Findings: The implementation of the Pairs Trading strategy with stock pairs selected through AI models showed higher profitability than conventional methods. Strategies utilizing LSTM models demonstrated the highest performance, achieving an approximate cumulative return of 51.25353%. This indicates that AI models are capable of accurately assessing similarities and establishing effective trading strategies.
Research limitations/implications: The study highlights the potential of AI-based stock pair selection methods to enhance Pairs Trading Strategies' performance. This approach surpasses traditional statistical methods by better reflecting the stock market's complexity and dynamism, potentially offering investors more stable and higher returns.
Originality/value: The research contributes to the field by demonstrating the effectiveness of AI models in the stock pair selection process, suggesting a novel approach to enhancing Pairs Trading Strategies that could provide valuable insights for investors seeking more sophisticated investment strategies in the financial markets.

Keywords: Pairs Trading, Similarity Analysis, Autoencoder, Vector Embeddings, LSTM

I. Introduction

In financial markets, the challenge of identifying optimal investment strategies persists. Traditional strategies rely on historical data and static models, which fail to account for the dynamic nature of the financial environment. However, recent advancements in artificial intelligence (AI) have provided new tools for more precise decision-making. AI has revolutionized big data processing, offering deeper insights into financial patterns and market behaviors (Huang, 2017). Pairs Trading exploits the relative price movements between two correlated stocks (Gatev et al., 2006; Kim et al., 2016). Traditionally, it has used statistical methods to identify and capitalize on price differences, assuming these differences will

Received: May. 15, 2024; Revised: May. 31, 2024; Accepted: Jun. 9, 2024
† Corresponding author: Kim Kang Hee
E-mail: mbbkim@lgchem.com

revert to a mean over time (Krauss, 2017).

Existing similarity analysis methods in Pairs Trading include the Distance approach (Gatev et al., 2006; Do and Faff, 2010), Cointegration approach (Vidyamurthy, 2004; Rad et al., 2015), Time-series approach (Elliott et al., 2005; Cummins and Bucca, 2012), Stochastic control approach (Jurek and Yang, 2007; Liu and Timmermann, 2013), and Copula approach (Krauss and Stübinger 2015). Additionally, AI-based methods like Machine Learning (ML) (Huck, 2009) and Principal Component Analysis (PCA) (Avellaneda and Lee, 2010) have been explored.

These traditional methods have limitations in adequately responding to market volatility and complexity. To address these limitations, we have introduced a novel approach integrating AI techniques such as Autoencoders (AE), Vector Embeddings (VE), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) to refine stock pair selection. This research is distinguished by its utilization of AI to enhance real-time data analysis and advanced pattern recognition, which are often overlooked in conventional pairs trading methodologies. By examining AI's capacity to improve stock pair selection and trading decisions, this study offers significant insights into AI's potential to drive innovative developments in financial trading strategies.

II. Background and Related work

A. Pairs Trading

Pairs Trading is a statistical financial strategy that generates profits by exploiting price differences between two related assets (Kim et al., 2016). The core of this strategy is to trade on price differences between two assets when these differences occur due to temporary market anomalies (Gatev et al., 2006). The main assumption is that the price difference between two assets will regress to the mean over time (Cho et al., 2022). Therefore, it is important to select correlated asset pairs and ensure that the spreads between them fluctuate periodically and tend to revert to the mean (Caldeira and Moura, 2013). If spreads are fixed, there is no profit opportunity, but pairs with volatility and with spreads that regress to the mean provide investment opportunities. This regression property of spreads is called Cointegration (Kim et al., 2016). Thus, when the price of one asset rises and the price of another falls, one sells the asset that has risen and buys the asset that has fallen, waiting for the price differential to narrow again before taking profits (Kim et al., 2016). This strategy is based on the expectation that price differentials will revert to the mean over time, unaffected by market direction (SHIN, 2021).

Pairs Trading involves the following steps. First, two stocks are selected for Pairs Trading (Cho et al., 2022; Kim and Lee, 2014). Factors considered at this time include the company's industry, market capitalization, and market equity ratio, favoring stocks with similar characteristics to each other. Second, the correlation between the two selected stocks is analyzed to see how closely they are statistically related (Cho et al., 2022). Typically, the correlation coefficient is used to measure this. Third, based on the correlation analysis, if one stock is expected to rise when the other stock rises, then one stock should be purchased and the other sold (Cho et al., 2022). These decisions are adjusted according to the strength of the correlation. Additionally, pairs trading strategies incorporate short selling and margin trading. If a stock price is expected to decline, short selling can be executed by borrowing and selling the stock in advance. Subsequently, if the stock price does indeed fall, the stock can be repurchased at a lower price and returned, thus realizing a profit (Mendee and Jun m, 2021). Buying and selling occurs when correlations are temporarily weak or volatile. Such timing is detected using statistical methods to determine when to enter and escape the market efficiently.

Pairs Trading serves as a defensive strategy against market volatility (Kim et al., 2016). In the stock market, correlations exist between stocks, and these correlations can be used to predict the movement of one stock over another (Kim and Lee, 2014). Therefore, the goal of Pairs Trading is to create profits by focusing on the relative price differences between specific stocks rather than the overall direction of the market (Kim et al., 2016).

B. Similarity Analysis

Similarity analysis is crucial in Pairs Trading, as accurately assessing and selecting similarities among stocks directly impacts strategy performance. Various methods and techniques can be used for this analysis. Mathematical, and statistical methods are commonly used to analyze the relationship between stocks. Distance Approach involves calculating the distance between price series of stocks to identify pairs with similar price movements (Gatev et al., 2006; Do and Faff, 2010). Cointegration analysis is used to analyze correlations in time series data and to understand long-term correlations (Caldeira and Moura, 2013; Krauss, 2017; Vidyamurthy, 2004; Rad et al., 2015). Time-Series Approach uses time-series models to analyze the price movements of stock pairs and generate trading signals (Elliott et al., 2005; Cummins and Bucca, 2012). Stochastic control theory is used to determine optimal trading strategies and portfolio holdings (Jurek and Yang., 2007 Liu and Timmermann, 2013) The copula method models the dependency structure between the price movements of stock pairs (Krauss and Stübinger, 2015). Machine learning and AI techniques are used to analyze stock similarity, enabling more accurate data-driven pattern recognition and similarity assessment (Huck, 2009; Avellaneda and Lee, 2010). For example, AI techniques such as Autoencoders (AE), Vector Embeddings (VE), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) can evaluate stock similarity (Seo et al., 2021; SEO and YUN, 2019; Kim et al., 2021; Lee and Kang, 2023). AE can compress data into a low-dimensional space and evaluate similarity by computing angles between vectors of this low dimension (Bank et al., 2023). VE maps data from a high-dimensional to a low-dimensional vector space and can thus compute the similarity of the transformed vectors. Cosine Similarity uses the cosine value of the angle between such vectors to express the similarity of two vectors as a value between 0 (not at all similar) and 1 (completely similar) (Yang, 2019). RNN, LSTM, and GRU, can be used to learn patterns in stock price time series data. In this process, RNN, LSTM, and GRU store data changes and patterns over time, which allows them to extract low-dimensional characteristics of the data (JUNG et al., 2023). The similarity between the low-dimensional characteristic vectors thus obtained can be computed using cosine similarity to evaluate the similarity of patterns between different stocks. Finally, the data can be analyzed through financial indicators and factors. When evaluating the similarity between stocks, it is important to consider the characteristics of the stocks, the industry sector, and the financial statements. Indicators commonly used for comparison and analysis among stocks include Price-to-Earnings Ratio (PER), asset size, and market capitalization. Stock similarity analysis can be performed by utilizing a synthesis of these various methods and techniques. This helps to determine the relevance and similarity between stocks and to select suitable pairs of stocks for Pairs Trading. Accurate analysis and selection of similarity among stocks plays an important role in improving the performance of the strategy.

Seo et al. (2021) studied a similarity-based anomaly detection model called Long Short-Term Memory Variational Autoencoder (LSTM-VAE), which utilizes Autoencoders (AE). In the study by SEO and YUN (2019), AE was employed, while Kim et al. (2021) developed a predictive model that leverages AE to explain regional real estate bubble conditions.

The study by Lee and Kang (2023) used an artificial neural network-based Vector Embedding method to classify movement modes from unlabeled people's GPS movement trajectory data; in the study by Yang et al. (2019), Word2Vec, a word embedding model, was used to convert them into vectors and compare their performance on Euclidean similarity, Cosine similarity, and Augmented Jaccard similarity using real data. In the study by Park and Kim (2019), documents were vectorized using only the core words, studied a multi-vector document embedding methodology in which a single document is represented by a set of multiple vectors by decomposing the various topics the document contains.

C. Factors Influencing Returns in Pair Selection

In Pairs Trading, pair selection is one of the most important factors that directly affect the rate of return. Pair selection can be divided into two methods. The first method is to search for pairs by investigating all possible combinations among the selected stocks (Krauss, 2017; Caldeira and Moura, 2013). And it is a method of searching for pairs through combinations, arranged in meaningful groups by sector (Do and Faff, 2010; Dunis et al., 2010). To summarize the factors that pair selection has on rates of return, first, irrational trading resulting from market inefficiencies can cause price divergence among similar stocks, which can be exploited to generate revenue (Do and Faff, 2010). However, as competition among competitors to take advantage of such opportunities intensifies, profitability may decline. Second, the profitability of Pairs Trading Strategies is highly influenced by market conditions, especially during periods of market instability, such as the financial crisis, which showed high profitability (Kim et al., 2016). Third, trading costs are an important factor in the profitability of Pairs Trading Strategies. The larger the trading costs, the lower the net profit of the strategy (Kim et al., 2016; Gatev et al., 2006). Trading costs significantly impact various financial aspects, including returns and liquidity (Park, 2018). Finally, selecting pairs within the same industry group may exhibit higher profitability relative to other groups (Kim and Lee, 2014). This may be due to economic factors that are commonly influenced by the characteristics of that industry. These factors play an important role in the success of a Pairs Trading Strategy and are important factors to consider when developing and implementing a strategy.

D. Research on Pairs Trading

Pairs Trading Strategies select highly correlated pairs of stocks and use these to take advantage of price differences (Kim and Lee, 2014). The greater the correlation, the greater the likelihood that a price change in one stock will affect another, and thus the more successful the trading strategy is likely to be. Pairs Trading assumes that price differences between stocks tend to regress around a constant mean (Caldeira and Moura, 2013). Therefore, it is important to analyze whether the selected pairs of stocks have mean-reversion characteristics. If such characteristics exist, price differentials can be reduced, and profits can be generated. Trading performance should also be considered when selecting a pair of stocks. Trading performance is important because Pairs Trading is very quick (Do and Faff, 2012). In other words, whether the selected stock pair is easy to trade, and liquid may affect profitability.

A study by Gatev et al. (2006) analyzed the risk and profit characteristics of Pairs Trading Strategies using daily data for the U.S. stock market from 1962 to 2002; a study by Do and Faff (2010) investigated whether Pairs Trading Strategies are still effective. They noted that Pairs Trading remains useful under certain conditions and market environments, but that market changes and external factors can change the effectiveness of the strategy; Do and Faff (2012) analyzed whether the profitability of pairs trading strategies remains robust when transaction costs are considered. The study examined Pairs Trading Strategies under various market conditions and trading cost levels, and the results suggest that profitability can be maintained up to a certain level of trading costs. However, profitability tends to decline as costs increase, underscoring the importance of traders strictly controlling costs and choosing optimal entry and liquidation timings. The study by Yun and Kim (2011) utilized the Kalman Filter (KF) to calculate spreads between stocks in extracting buy and sell signals by exploiting high-frequency data. Caldeira and Moura (2013) studied a portfolio of pairs trading strategies based on Cointegration, a statistical arbitrage approach that identifies long-term relationships between stock prices. Using Brazilian stock market data from 2005 to 2012, they applied the Cointegration test to select stock pairs and evaluate the profitability of the strategy. The results show significant excess returns, demonstrating that Cointegration-based pairs trading is effective even during market crises. Kim and Lee (2014) empirically analyzed a pairs trading strategy using Bollinger Bands to account for industry-specific volatility in the Korean stock market. The study by Kim et al. (2014) studied Pairs Trading using the price ratio of the KOSPI 200 and S&P 500 index futures; the study by Kim et al. (2016) studied Pairs Trading Strategies for the foreign futures market and used orthogonal regression to select pairs. orthogonal regression analysis and ADF tests were used to select the covaried pairs.

The study by Fallahpour et al. (2016) studied the process of simultaneously learning and optimizing stock pair selection and trading rules using the Q-learning algorithm, a reinforcement learning model. The study results showed that reinforcement learning can be used to make real-time decisions based on market data, thereby improving the profitability of Pairs Trading Strategies. The study by Krauss (2017) studied statistical arbitrage strategies for Pairs Trading were examined in detail and analyzed various approaches, including distance, covariance, time series, and stochastic control. Highlighting the simplicity and profitability of Pairs Trading in various markets and asset classes, identifying similar stocks, and studying the optimization of trading signals, the study by Brim (2020) investigated how to optimize the profitability of Pairs Trading Strategies using deep learning. He presented how a deep reinforcement learning model can select optimal pairs based on market data and determine when to buy and sell to maximize profitability. The Sarmento and Horta (2020) study utilized OPTICS-based Principal Component Analysis (PCA) to search for similar stock pairs and utilized a novel prediction-based trading model using ARIMA, LSTM, and LSTM Encoder-Decoder. the SHIN (2021) studied a pairs trading strategy with stop loss levels applied to KOSPI under the universal assumption that stocks follow a geometric Brownian model. Cho et al. (2022) studied pairs of similar price flows through K-means, a type of grouping which is a type of grouping. The study by Keshavarz Haddad and Talebi (2023) analyzed the profitability of Pairs Trading Strategies. Using data from the Toronto Stock Exchange, they evaluated the effects of stock selection and trading strategies using covariance, distance, and copula methodologies. In particular, the study covered the period including before and after the COVID-19 pandemic to confirm the stability of the strategies even under market crisis conditions. The results showed that copula-based trading strategies exhibited the highest profitability.

E. Analysis and Limitations of Prior Research

The results of the analysis of the previous studies are presented below. Prior studies explored various aspects of Pairs Trading Strategies and presented the following key findings. First, the basic principles of Pairs Trading. Based on statistical theory and the characteristics of the market, this strategy uses relative price differences between related assets to generate profits. It seeks profits by trading on the price difference between two assets until the difference is recognized and adjusted as a market anomaly (Gatev et al., 2006). Second is the importance of issue similarity analysis (Caldeira and Moura, 2013). The accurate assessment and selection of similarity between stocks has a direct impact on the strategy's profitability and performance, and for this purpose, statistical methods, machine learning and AI techniques, and analytical methods utilizing financial indicators and factors are used.

Limitations of previous studies are as follows. First, is the reliance on historical data (Gatev et al., 2006). Most previous studies use models based on historical data, which limits their ability to accurately reflect current and future financial market conditions. As a result, models may have difficulty dealing with rapid market volatility and unpredictable events. Second is the static optimization approach. Many prior studies use static optimization based on data collected over a fixed period (Caldeira and Moura, 2013; Kim et al., 2014). Such approaches cannot adequately account for market changes over time, making it difficult to respond flexibly to real-time market changes. Third is the complexity of similarity assessment. The complexity and diversity of similarity assessment methodologies among stocks make accurate stock selection and similarity measurement difficult (Caldeira and Moura, 2013; Kim et al., 2016; Sarmento and Horta, 2020; Cho et al., 2022). Finally, the study is specific to trading strategies. Although the selection of stock pairs has a direct impact on earnings, recent research has focused on AI-based trading strategies (Fallahpour et al., 2016; SHIN, 2021; Brim, 2020).

To overcome the limitations of previous research, this study investigates the optimal pair selection model utilizing artificial intelligence technologies such as AE, VE, LSTM, GRU, and RNN. The study proposes a model that leverages these AI technologies to process dynamic data, recognize complex market patterns, and evaluate enhanced similarities, ultimately aiming to improve profitability.

III. Methodology

A. AI-based Pairs Trading Model (APTM)

APTM, the AI algorithm proposed in this study, is divided into Pair Selection and Trading Strategy stages. In the Pair Selection phase, each algorithm (AE, VE, RNN, LSTM, and GRU) is used to encode the stock data D into a low-dimensional characteristic space as shown in the following equation.

$$V = Encoding_{Algo}(D),$$

$$Algo \in \{AE(D), VE(D), RNN(D), LSTM(D), GRU(D)\}$$

The similarity matrix S is generated by calculating the cosine similarity between the encoded data as shown in the following equation.

$$S_{ij} = CosineSimilarity(V_i, V_j) = \frac{V_i \cdot V_j}{\| V_i \| \| V_j \|}$$

Find the i and j with the largest values from the similarity matrix S and extract the corresponding stocks.

$$S1, S2 = \arg max_{ij} S_{ij}$$

The Trading Strategy phase utilizes the equity ratio and the Z-score to determine buy and sell signals. The stock ratio is defined by the following equation, where S1 and S2 are the price of the two selected stocks.

$$Ratio = \frac{S1}{S2}$$

The Z-score is calculated as in the following equation, MA(S1) is the moving average of the first stock, MA(S2) is the moving average of the second stock, and SD(S1-S2) is the standard deviation of the price difference between the first and second stocks.

$$Z-score = \frac{MA(S1) - MA(S2)}{SD(S1 - S2)}$$

Adding the buy and sell conditions to the trading strategy can be represented as follows.

$$If\ Z < -1,\ then\ Buy\ S1 \text{ and } Sell\ S2 \times Ratio$$
$$If\ Z > 1,\ then\ Sell\ S1 \text{ and } Buy\ S2 \times Ratio$$
$$If\ -0.5 < Z < 0.5 \text{ and } profit\ is\ positive,\ then\ Exit\ S1,\ S2\ positions$$

Where Z is the Z-score, S1 and S2 are the two selected shares, and Ratio is the ratio of the two share prices. If the Z-score is below -1, buy S1 and sell S2; if the Z-score is above +1, sell S1 and buy S2. If the Z-score is between -0.5 and 0.5 and the profit is positive, the position is liquidated.

B. Data & Features

Price data for the equity stocks comprising the portfolio will be collected from the S&P 500. The collected data will be refined, and missing values and outliers will be processed. The data used in the study are shown in Table 1. and the period of study is from January 2, 2014, to December 29, 2023. Pair Selection is performed using each model (AE, VE, LSTM, GRU and RNN) using the APTM model based on the collected data.

C. Evaluation Metrics

The APTM was trained from January 2, 2014, to December 29, 2020, and back tested for each model from December 30, 2020, to December 29, 2023, using the models trained during the train period.

To select similar stock pairs, this study evaluates and compares the traditional statistical techniques of Merit and Artificial Function techniques of AE and VE, and deep learning techniques of RNN, LSTM, and GRU models using performance metrics (JUNG et al., 2023). The performance metrics for this study are as follows Cumulative Returns represents the overall performance of an investment strategy and is a cumulative representation of returns over the entire period (JUNG et al., 2023). Annual Return shows the return of an investment on an annualized basis and is used to evaluate the annual performance of an investment (JUNG et al., 2023). Annual Volatility represents the volatility of an investment on an annualized basis and is used to measure the riskiness of an investment (JUNG et al., 2023). The Sharpe Ratio represents the ratio of an investment's excess return to its risk and is a measure of how efficiently an investment has generated returns relative to its risk (Sharpe, 1998; Cheong et al., 2023). The higher the ratio, the better the investment strategy. The Sortino Ratio represents the ratio of an investment's excess return to its negative volatility (downside risk), similar to the Sharpe Ratio but specifically evaluating profitability relative to downside risk by considering only negative volatility (Sortino, 1991). The higher this index is, the better the strategy performs. Max Drawdown (MDD) is a metric that shows the likelihood of suffering the largest losses at the peak of an investment strategy and is used to determine the riskiness of a strategy (JUNG et al., 2023; Magdon-Ismail et al. 2004). Utilizing these indicators and evaluating the overall performance of each model provides a better understanding of the performance, risk, and efficiency of the investment strategy.

Table 1. Dataset

	Count (%)	Period
Total Dataset	2,516 (100 %)	2014.01.02 ~ 2023.12.29
Training Dataset	3,339 (70 %)	2014.01.02 ~ 2020.12.29
Test Dataset	835 (30 %)	2020.12.30 ~ 2023.12.29

IV. Research Findings

A. Descriptive Statistical Analysis

The data used in this study is for the Industrials sector, which is a component of the S&P 500 and covers the 78 stocks that make up this sector, and since data is not available for holidays, data from the immediately preceding business day was obtained and used. Data were collected for 2,516 days from January 2, 2014, to December 29, 2023, Of the 78 stocks in the Industrials sector, 69 stocks with the same number of stock price data for the study period were used to measure similarity.

A total of 9 stocks were selected from the research models, excluding duplicates. The stock pairs selected for each model are as follows: Axon Enterprise (Symbol AXON) and Trane Technologies (Symbol TT) for the Cointegration model, Expeditors International (Symbol EXPD) and Xylem Inc. (Symbol XYL) for the AE model, A. O. Smith (Symbol AOS) and Expeditors International (Symbol EXPD) for the VE model, Caterpillar Inc. (Symbol CAT) and Parker Hannifin (Symbol PH) for the RNN model, Rockwell Automation (Symbol ROK) and Union Pacific Corporation (Symbol UNP) for the LSTM model, and Caterpillar Inc. (Symbol CAT) and Parker Hannifin (Symbol PH) for the GRU model.

The date-specific rates of return for the selected stocks are shown in Figure 1. and the technical statistic data for the collected stocks are shown in Table 2.

B. Results

1. Cointegration Pair Selection

Most of the time (80%) during the study period was allocated to measuring the p-value of each

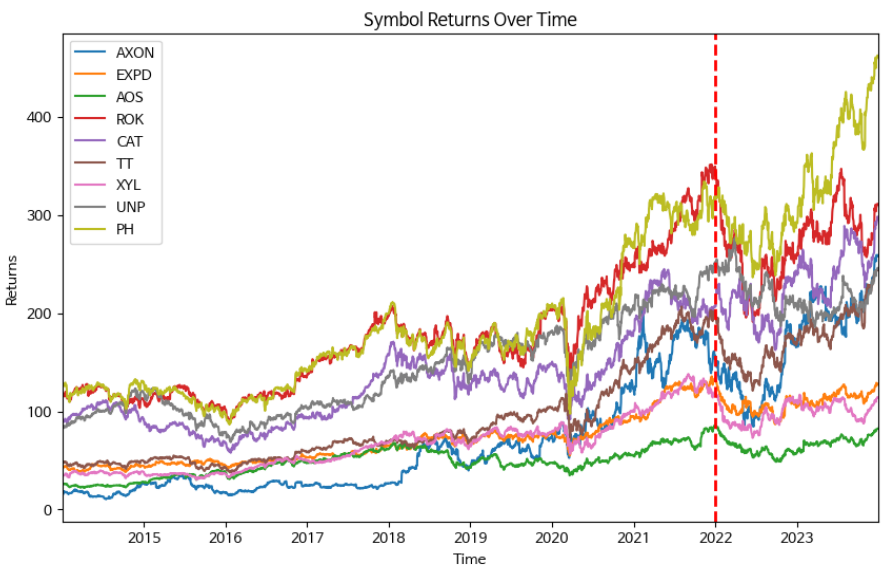

Figure 1. date-specific rates of return

Table 2. Descriptive statistics

Symbol	count	mean	std	min	25%	50%	75%	max
AXON	2516	78.88436	64.84886	10.50	24.24	59.00	125.785	259.08
EXPD	2516	75.47801	27.23007	38.21	49.9675	72.195	99.445	135.62
AOS	2516	51.08163	14.87522	22.34	40.2175	51.155	62.3325	85.85
ROK	2516	190.1461	68.25568	89.71	123.2275	176.58	249.8	351.35
CAT	2516	145.2579	57.65722	57.91	96.2625	134.38	194.8425	298.12
TT	2516	101.2274	54.36388	37.85881	52.51164	79.25911	150.955	243.95
XYL	2516	71.70726	27.71086	30.46	46.0725	72.9	91.8125	138.03
UNP	2516	154.659	52.44174	68.79	105.43	151.45	203.6825	276.69
PH	2516	201.5074	86.56979	86.51	124.7625	173.355	276.3775	462.25

competition pair using the Cointegration model, and the remaining 20% was set aside as a test period to perform additional p-value measurements for each competition pair. The results of the experiment are as follows. The Cointegration model similarity results are summarized in Table 3, and the buy and sell signals during the test period are visually represented in Figure 2. The results of the analysis show that the stock pair of Axon Enterprise and Trane Technologies exhibited the lowest p-value of 0.000317. This low p-value indicates a strong covariance relationship between the two stocks, which means that the price difference between the two stocks is likely to remain in constant balance over time. This study also introduced a method of using the p-value subtracted from 1(i.e., 1 minus p-value) as the similarity. The application of this method allows for a clearer identification of stock pairs with high similarity (i.e., those with low p-values) and allows for similarity comparisons with other models in the stock pair selection process. in the case of Axon Enterprise and Trane Technologies, the 1 minus p-value was subtracted from 0.999683, which indicates a very high degree of similarity.

2. AE Pair Selection

During 80% of the study period, the AE model was used for training, and the remaining 20% of the study period was used as a test interval to measure similarity for each pair of stocks. The experimental results are as follows. The similarity results of the AE model are summarized in Table 4, while the buy and sell signals during the test period are visually represented in Figure 3. The similarity results show that the Expeditors International and Xylem Inc. stock pair has the highest similarity of 0.999851.

Table 3. Cointegration similarity

No	Stock 1	Stock 2	p-value	Similarity
1	**Axon Enterprise**	**Trane Technologies**	**0.000317**	**0.999683**
2	Hubbell Incorporated	Quanta Services	0.000333	0.999667
3	Automated Data Processing	Waste Management	0.000373	0.999627
4	Eaton Corporation	United Rentals	0.000492	0.999508
5	Dover Corporation	Jacobs Solutions	0.000658	0.999342

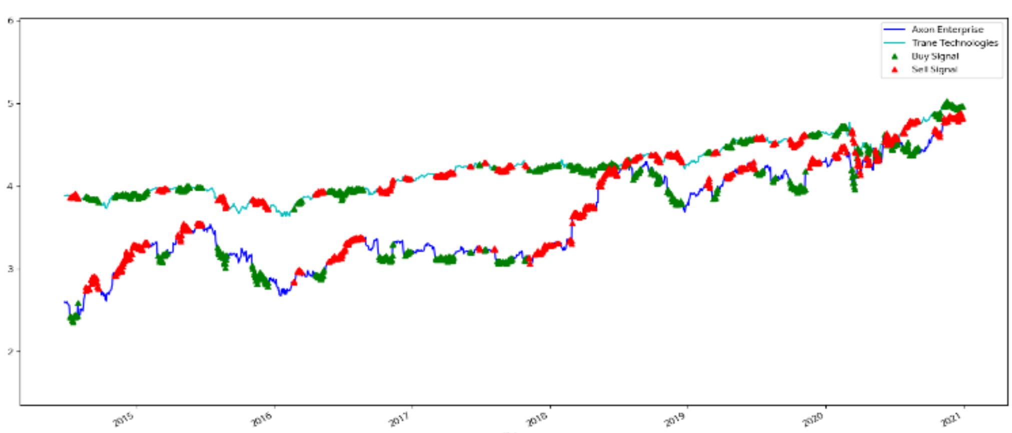

Figure 2. Cointegration buy & sell signal

Table 4. Autoencoder similarity

No	Stock 1	Stock 2	Similarity
1	**Expeditors International**	**Xylem Inc.**	**0.999851**
2	CSX	Rollins, Inc.	0.999479
3	Delta Air Lines	Southwest Airlines	0.999072
4	CSX	Fastenal	0.998912
5	American Airlines Group	Delta Air Lines	0.998669

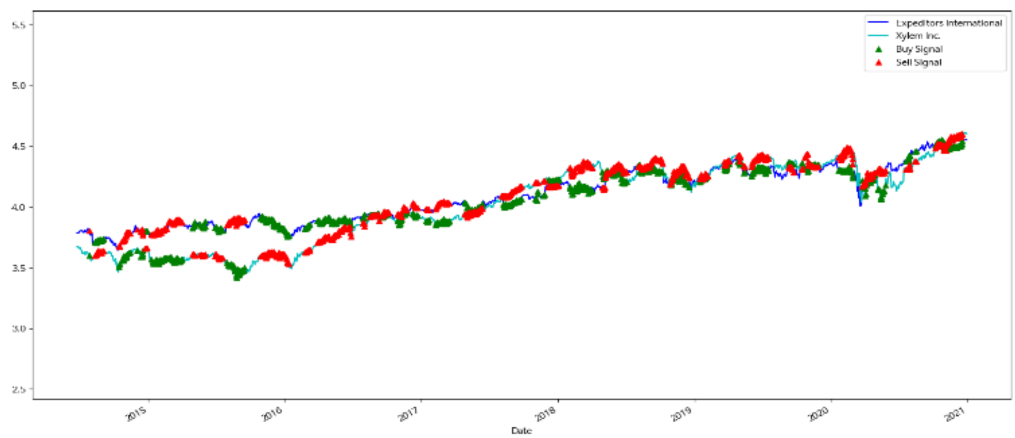

Figure 3. Autoencoder buy & sell signal

3. VE Pair Selection

During 80% of the study period, training was performed using the VE model, and the remaining 20% of the study period was used as a test interval to measure similarity for each pair of stocks. The experimental results are as follows. The similarity results of the VE model are summarized in Table 5, and the buy and sell signals during the test period are visually represented in Figure 4. The similarity results show that the stock pair of A.O.Smith and Expeditors International has the highest similarity of 0.99994.

4. RNN, LSTM, and GRU Pair Selection

The RNN, LSTM, and GRU models were used to train for 80% of the study period, while the remaining 20% was used as a test interval to measure similarity for each pair of stocks. The experimental results are as follows: the similarity results for RNN, LSTM, and GRU are organized in Table 6, and the buy and sell signal for the test period can be seen in Figures 5, 6, and 7. The RNN model shows the highest similarity of 0.999614 for the Caterpillar Inc. and the Parker Hannifin stock pairs, the LSTM model shows the highest similarity of 0.999763 for the Rockwell Automation and the Union Pacific Corporation stock pairs, and the GRU model shows the highest similarity of 0.999735 for the Caterpillar Inc. and the Parker Hannifin stock pairs.

C. Evaluation

This study utilized various AI models to evaluate the performance of equity Pairs Trading Strategies. The models included in this study are Cointegration, AE, VE, RNN, LSTM, and GRU. Each model formulated the same trading strategy based on the

Table 5. Vector embeddings similarity

No	Stock 1	Stock 2	Similarity
1	**A. O. Smith**	**Expeditors International**	**0.999940**
2	Fastenal	Rollins, Inc.	0.999927
3	Jacobs Solutions	Republic Services	0.999892
4	A. O. Smith	Paccar	0.999871
5	Paychex	Robert Half	0.999838

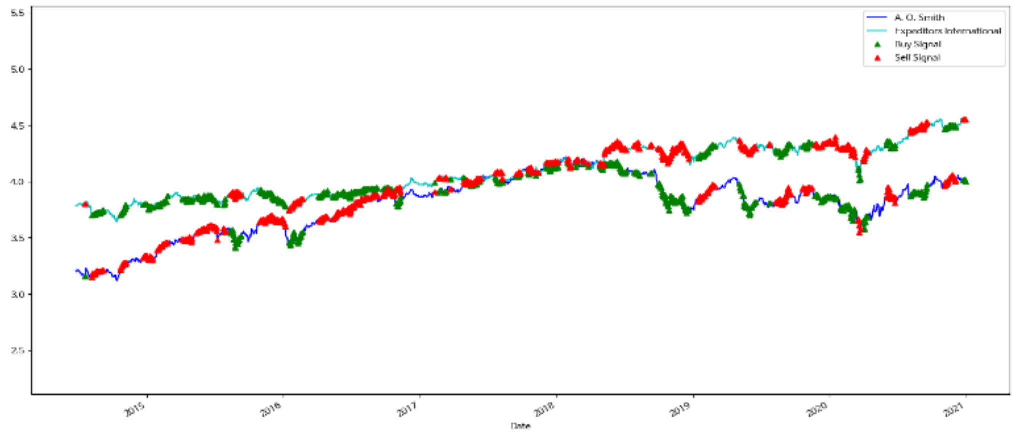

Figure 4. Vector embeddings buy & sell signal

Table 6. RNN, LSTM, and GRU similarity

MODEL	No	Stock 1	Stock 2	Similarity
RNN	1	**Caterpillar Inc.**	**Parker Hannifin**	**0.999614**
	2	Automated Data Processing	IDEX Corporation	0.999563
	3	Rockwell Automation	Union Pacific Corporation	0.999525
	4	Nordson Corporation	Waste Management	0.999518
	5	Automated Data Processing	Waste Management	0.99938
LSTM	1	**Rockwell Automation**	**Union Pacific Corporation**	**0.999763**
	2	Caterpillar Inc.	Parker Hannifin	0.999691
	3	Lockheed Martin	Northrop Grumman	0.999663
	4	Automated Data Processing	IDEX Corporation	0.99963
	5	Ametek	Dover Corporation	0.999607
GRU	1	**Caterpillar Inc.**	**Parker Hannifin**	**0.999735**
	2	Automated Data Processing	IDEX Corporation	0.999653
	3	Rockwell Automation	Union Pacific Corporation	0.999626
	4	Nordson Corporation	Waste Management	0.999526
	5	Lockheed Martin	Northrop Grumman	0.999477

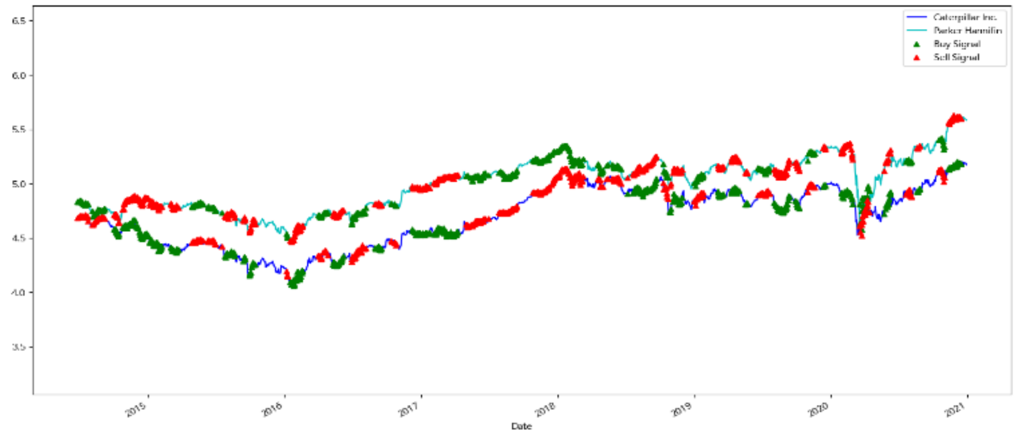

Figure 5. RNN buy & sell signal

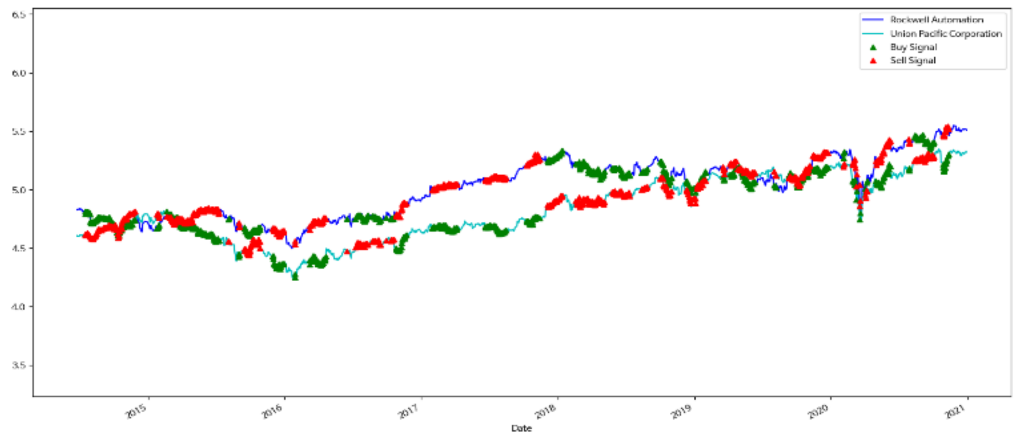

Figure 6. LSTM buy & sell signal

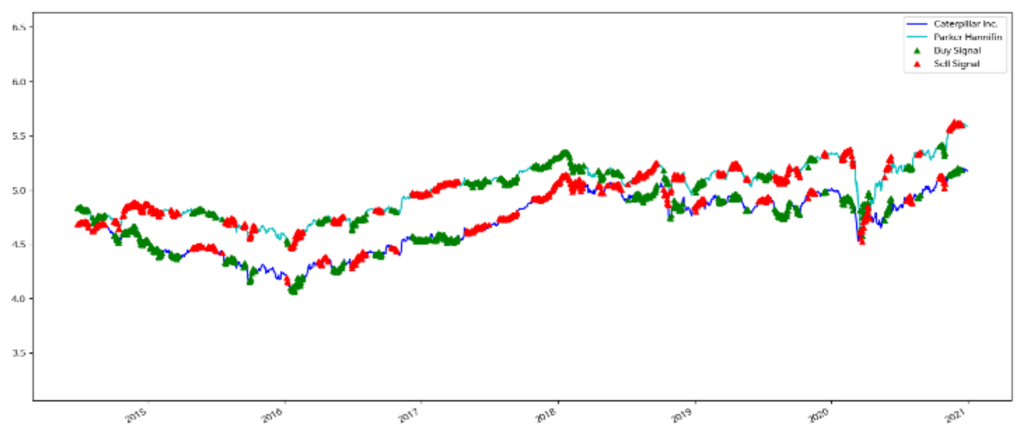

Figure 7. GRU buy & sell signal

similarity between the stock pairs (Stock 1 and Stock 2). The test set was used to measure each model's cumulative return, annual return, annual volatility and risk-adjusted return on the Sharpe Ratio and Sortino Ratio and maximum loss rate. Study results include.

The experimental results, indicate that the LSTM models achieved the highest performance with a rate of return of 51.25353% the VE model also performed well, recording a rate of return of about 9.206619%. According to Sharpe (1994), the Sharpe Ratio measures the risk-adjusted return of an investment, indicating how well the LSTM models performed in maximizing returns while managing overall volatility. Similarly, Sortino (1994) emphasizes that the Sortino Ratio assesses the return of an investment relative to the downside risk, demonstrating that the LSTM models effectively mitigated downside risk while achieving high returns. In particular, the Sortino Ratio of these two models was much higher than those of the other models, indicating superior compensation for downside risk. Additionally, all models recorded relatively low Maximum Drawdown (MDD) rates, which is consistent with research by Magdon-Ismail et al. (2004), highlighting that lower MDD rates are indicative of models that can generate stable returns with minimal significant losses.

V. Conclusions

This study evaluated the performance of Stock Pairs Trading Strategies utilizing various deep learning models and investigated the effectiveness of similarity-based trading strategies between stock pairs. The experimental results demonstrated that the LSTM models performed the best, with each achieving a rate of return of approximately 51.25353%.

This study differs from previous studies in several ways. First, rather than relying solely on historical data for pair selection, this study utilized a variety of AI models that can accommodate the variability of real-time data and complex market conditions. This overcomes the limitations of existing statistical or static models and is differentiated by its ability to respond more flexibly to market uncertainty. Secondly, this study presents a methodology for pair selection through similarity measures, which allows for a more granular understanding of the dynamic relationships in the stock market. In contrast to existing studies, which have primarily employed pair selection methodologies based on fixed statistical attributes and economic assumptions, this study recognizes more complex market patterns through the learning capabilities of AI models and reflects this in trading strategies.

This research contributes to the academic literature by demonstrating how AI technologies can be effectively integrated into financial trading strategies to improve their performance. It establishes a comprehensive framework for utilizing AI in analyzing stock market data, setting a solid foundation for subsequent studies in financial analytics. Employing a range of AI models to evaluate similarities between stock pairs, this study introduces an innovative methodological approach, potentially enriching future research in finance and AI. In practice, this study offers actionable insights for financial practitioners. It demonstrates that AI-driven pairs trading strategies can significantly outperform traditional methods, especially in volatile markets. The use of AI enables traders and financial analysts to identify profitable trading opportunities more accurately and efficiently. This, in turn, has the potential to increase returns while managing risks effectively.

The limitations of this study are as follows. Firstly, the study was limited to a specific deep learning model and stock pair, and thus lacks validation for various market conditions and other asset classes. Secondly, there is a general limitation of deep learning models in that the internal mechanisms of the models and their interpretation for the market are relatively difficult to comprehend.

Future research should investigate the generalizability of our model by applying it to diverse markets, asset classes, and market conditions. Additionally, research is needed to develop methodologies to enhance the interpretability of the model and to fortify the stability and risk management capabilities of the model. Furthermore, additional research to continuously refine optimal pair selection methods and trading strategies is crucial.

This study presents a novel approach to applying AI models to Pairs Trading Strategies. The objective is to enhance the performance and market responsiveness of trading systems.

References

Avellaneda, M., & Lee, J.-H. (2010) Statistical arbitrage in the US equities market. *Quantitative Finance, 10*(7), 761-782.

Bank, D., Koenigstein, N., & Giryes, R. (2023). Autoencoders. In *Machine learning for data science handbook: Data mining and knowledge discovery handbook* (pp. 353-374). Springer.

Brim, A. (2020). Deep reinforcement learning pairs trading with a double deep Q-network. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0222-0227). IEEE.

Caldeira, J., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Available at SSRN 2196391.

Cheong, M., Kim, Y., & Kim, K. (2023). Management of the Pension Fund in Korea: Sharpe Ratio as a Measurement. *Global Business and Finance Review, 28*(6), 95-111. doi:10.17549/gbfr.2023.28.6.95

Cho, P., Lee, M., & Song, J. W. (2022). Clustering-driven pairs trading Portfolio Investment in Korean Stock Market. *Journal of Korean Society of Industrial and Systems Engineering, 45*(3), 123-130.

Cummins, M., & Bucca, A. (2012) Quantitative spread trading on crude oil and refined products markets. *Quantitative Finance, 12*(12). 1857-1875.

Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal, 66*(4), 83-95.

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research, 35*(2), 261-287.

Dunis, C. L., Giorgioni, G., Laws, J., & Rudy, J. (2010). *Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities*. Working Paper, Liverpool Business School.

Elliott, R.J., Van Der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance, 5*(3), 271-276.

Fallahpour, S., Hakimian, H., Taheri, K., & Ramezanifar, E. (2016). Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach. *Soft Computing, 20*, 5051-5066.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies 19(3):797-827.

Huang, S. (2017). A Big Data Analysis System for Financial Trading. *Global Business and Finance Review, 22*(3), 32-44.

Huck, N. (2009) Pairs selection and outranking: an application to the S&P 100 index. *European Journal of Operational Research, 196*(2), 819-825.

Jung, N. H., Oh, T., & Kim, K. H. (2023). A Study on DRL-based Efficient Asset Allocation Model for Economic Cycle-based Portfolio Optimization. *Journal of Korean Society for Quality Management, 51*(4), 573-588.

Jurek, J. W., & Yang, H. (2007) *Dynamic portfolio selection in arbitrage*. Working paper, Harvard University.

Keshavarz Haddad, G., & Talebi, H. (2023). The profitability of pairs trading strategy in stock markets: Evidence from Toronto stock exchange. *International Journal of Finance & Economics, 28*(1), 193-207.

Kim, B., Choi, H. S., & Kim, S. W. (2016). A Study on Pairs Trading Performance in Global Futures Markets. *Korean Management Science Review, 33*(4), 1-15.

Kim, J., & Lee, S. (2014a). The Study on Pairs Trading Strategies by Industry for Stable Profits. *Journal of Industrial Innovation, 30*(4), 61-80.

Kim, J., Kim, J., & Lee, S. (2014b). Developing Pairs Trading Rules for Arbitrage Investment Strategy based on the Price Ratios of Stock Index Futures. *Journal of Korean Society of Industrial and Systems Engineering, 37*(4), 202-211.

Kim, Y., Park, J., & OH, H. Y. (2021). Analysis of Outliers In Real Estate Prices Using Autoencoder. *Journal of the Korea Institute of Information and Communication Engineering, 25*(12), 1739-1748.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys, 31*(2), 513-545.

Krauss, C., & Stübinger, J. (2015). *Nonlinear dependence modeling with bivariate copulas: statistical arbitrage pairs trading on the S&P100*. IWQW Discussion Paper Series, University of Erlangen-Nürnberg, Nürnberg.

Lee, J., & Kang, Y. (2023). Classification of Travel Mode in Smartphone Trajectories using Vector Embedding. *Journal of Korean Society for Geospatial Information Science, 31*(3), 3-12.

Liu, J., & Timmermann, A. (2013) Optimal convergence trade strategies. *Review of Financial Studies, 26*(4), 1048-1086.

Magdon-Ismail, M., Atiya, A. F., Pratap, A. & Abu-Mostafa, Y. S. (2004). On the maximum drawdown of a Brownian motion. *Journal of Applied Probability, 41*(1), 147-161.

Mendee, E., & Jun, S. (2021). Do Short Selling and Margin Trading Affect Price Randomness? *Global Business and Finance Review, 26*(3), 1-13.

Park, J., & Kim, N. (2019). Multi-Vector Document Embedding Using Semantic Decomposition of Complex Documents. *Journal of Intelligence and Information Systems, 25*(3), 19-41.

Park, K. (2018). The Comparison of Trading Costs between Two Markets: KOSPI and KOSDAQ. *Global Business and Finance Review, 23*(4), 1-22.

Rad, H., Low, R. K. Y., & Faff, R. W. (2015) *The profitability of pairs trading strategies: distance, cointegration, and copula methods*. Working paper, University of Queensland.

Sarmento, S. M., & Horta, N. (2020). Enhancing a pairs trading strategy with the application of machine learning. *Expert Systems with Applications, 158*, 113490.

Seo, J., Park, J., Yoo, J. W., & Park, H. (2021). Anomaly Detection System in Mechanical Facility Equipment: Using Long Short-Term Memory Variational Autoencoder. *Journal of Korean Society for Quality Management, 49*(4), 581-594.

Seo, M., & Yun, W. (2019). Condition Monitoring and Diagnosis of a Hot Strip Roughing Mill Using an Autoencoder. *Journal of Korean Society for Quality Management, 47*(1), 75-86.

Sharpe, W. F. (1998). *The sharpe ratio. Streetwise-the Best of the Journal of Portfolio Management, 3*, 169-85.

SHIN, D. (2021). Optimal Pairs-Trading Application and Performance with Cutting Loss for KOSPI Stocks under the Geometric Brown Motion. *Journal of The Korean Data Analysis Society, 23*(3), 1169-1183.

Sortino, F. A., & Van Der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management Summer 1991, 17*(4), 27-31.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. Hoboken, NJ: John Wiley & Sons.

Yang, Y., Lee, B., Kim, J., & Lee, K. Y. (2019). Development of An Automatic Classification System for Game Reviews Based on Word Embedding and Vector Similarity. *The Journal of Society for e-Business Studies, 24*(2), 1-14.

Yun. J., & Kim, K. (2011). Performance Analysis of Pairs Trading Strategy Utilizing High Frequency Data: Evidence from the Korean Stock Market. *Asian Review of Financial Research, 24*(4), 1153-1172.