OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Pair Trading Analysis of the Semiconductor Stock Market

Alex Taeyeon Shin
shin44649@gmail.com

ABSTRACT

This study investigates the strength of the co-relationships between major semiconductor manufacturers and large technology companies through pair analysis and predicts their relevant future performance using machine learning. In recent years, unpredictable global events—including the COVID-19 pandemic and international trade conflicts—have initiated significant fluctuations in the semiconductor industry's stability. The rapid emergence of generative artificial intelligence (AI) has increased demand for semiconductors, further disrupting the semiconductor supply chain and accelerating structural economic transformation.

We hypothesize that this economic instability is reflected in the performance relationships among semiconductor-related firms over a long-term period. To examine this hypothesis, we select ten major global technology companies highly relevant to the semiconductor industry, find a strong relationship, and analyze their future performance from 2019 to 2024. Four econometric pair analysis methods are applied to identify optimal pair combinations, and future pair performance is subsequently estimated using two machine-learning models.

The results indicate that, among the methods considered, only the Johansen cointegration method successfully captures long-term equilibrium relationships between the companies under study at the 95% confidence level. The decision tree and random forest models demonstrate the ability to predict the next-day movement, achieving an average accuracy exceeding 50% based on the forecasting z-score spread trends from the optimized company pairs. These findings suggest that conducting long-term pair analysis is feasible for the semiconductor industry. Although short-term movement projections may yield strong performance outcomes, they may not fully capture broader structural effects, including prolonged global disruptions.

INTRODUCTION

Semiconductors are indispensable materials in modern human life. These solid-state materials are widely used in most of the current electrical devices because of their unique intermediate electrical conductivity—positioned between that of metallic conductors and insulating materials—which makes them essential for manufacturing integrated circuits and other electronic components [1]. Devices such as smartphones, computers, electric vehicles, and household appliances all rely on semiconductors for processing power, connectivity, and interface control. Consequently, the semiconductor industry is a key driver of advanced technological innovation, enhancing productivity and promoting the quality of human life [2]. As a core pillar of the digital generation, the semiconductor market is one of the most strategically important sectors in the global economy, and especially, its importance is further heightened by the rapid emergence of AI (artificial intelligence).

In recent years, however, the semiconductor industry has experienced substantial instability. The COVID-19 pandemic severely disrupted global production and logistics systems, while rapid advancements in intellectual property development and intensified technical competition have further accelerated structural changes in the market [3]. These fluctuations are closely linked to geopolitical interventions and national strategic interests, resulting in chip shortages, supply chain fragmentations, and significant pricing volatility [4,5]. Especially, the US–China trade conflict has undermined the stability of the global semiconductor supply chain through export controls, sanctions, and policy initiatives aimed at reshoring or localizing chip manufacturing [5].

Such market volatility has prompted renewed interest in assessing the relationship between semiconductor manufacturers and large technology companies that consume chips [6]. Although rapid structural changes in the semiconductor market are difficult to evaluate within a limited time frame, examining the inter-firm relationships provides valuable insights into industry dynamics and future chip-cycle expectations. In this study, we hypothesize that a close, measurable relationship exists between chipmakers and upstream technology firms, and that this relationship manifests as pairing movements in their stock market performance over the long term. To test this hypothesis, we selected ten representative companies spanning both semiconductor manufacturing and technology consumption segments and analyzed their correlations and long-term equilibrium relationships using pair trade analysis. Pair trading is widely used in quantitative finance as a market-neutral strategy that identifies relative underperformance or outperformance by examining price spreads [7].

Selecting a diverse set of companies is critical to ensuring analytical robustness. A heterogeneous portfolio that captures different dimensions of market behavior enhances the reliability and interpretability of the results. The companies included in this study were selected based on an annual semiconductor industry report that ranks firms according to market performance and influence [8]. We also assume that, while chip markets may experience more short-term uncertainty during a crisis, longer-term stock price movements tend to stabilize as markets adjust and supply and demand gradually balance out. As time passes, structural resilience and adaptive capacity may weaken extreme interactions between firms. This assumption is consistent with conventional macroeconomic cycle theory, which suggests that markets eventually revert toward equilibrium after periods of disruption [9]. Accordingly, here we analyze stock price data from 2019-1-2 to 2024-12-30, which includes the COVID-19 pandemic period and the initial phase of global trade conflicts.

We start our pair analysis by applying four relationship models to identify optimally paired companies that exhibit synchronized performance patterns, as defined by correlation and cointegration. Next, we employ machine learning algorithms to forecast how these pairs may move in the future based on performance spreads. With this combined approach, our study aims to assess the feasibility of long-term performance analysis in the semiconductor market and to better understand volatility in the semiconductor industry, particularly when the market is influenced by trade fluctuations and macroeconomic shocks.

METHODS

In this study, we employed a two-fold approach: pair analysis to investigate the interrelationship between the performance of semiconductor manufacturers and major technology companies that consume chips, and future performance prediction to forecast the next-day movement of the optimal pairing group. Figure 1 illustrates the overall workflow timeframe of the analytical process used in this study. The primary analysis consists of pair analysis and future movement prediction. Blue boxes represent the main analytical stages, whereas white boxes denote subordinate processes. The arrows indicate the direction of the sequential procedure, implying that each step is irreversible. All procedures were carried out in parallel for each variable to obtain a meaningful outcome through integrated evaluation. The analysis was conducted using Python scripts in a free, cloud-based Jupyter Notebook environment. The scripts used in this study are publicly available at the GitHub repository [10].

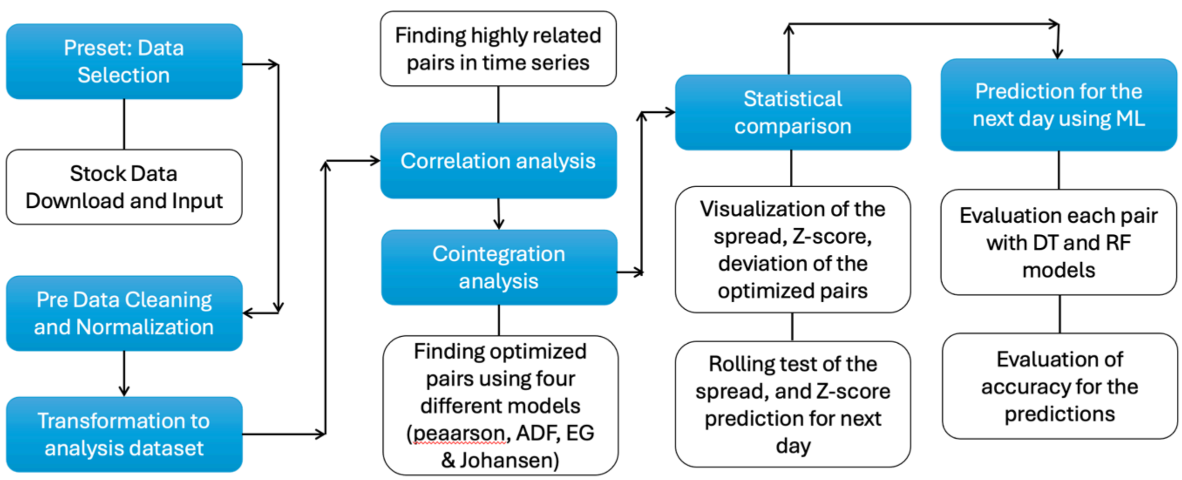

Figure 1: Flowchart of the overall analytical framework applied in this study. The primary analysis consists of two components: pair analysis and future movement prediction. Blue boxes represent the main analytical stage, while white boxes denote subordinate processes. Arrows indicate the sequential procedure of the analyses. Each process is irreversible, and all sub-procedures are conducted in parallel for each variable.

For the pair analysis, we applied four models: Pearson correlation, the Augmented Dickey-Fuller (ADF) test, the Engle-Granger two-step method, and the Johansen cointegration test. The Pearson method is used to evaluate correlations between variables, whereas the other three methods are used to identify cointegration among variables. In stock market analysis, correlation and cointegration are conceptually related but they have distinct technical meanings and uses [11]. Correlation measures the extent to which two variables move in the same direction, while cointegration evaluates whether two or more variables maintain a stable long-term relationship. All these analytical methods rely on the p-value from statistical hypothesis testing to assess whether the test assumptions are satisfied [12]. Typically, a significance level of 0.05 is used as the threshold for hypothesis testing. Here, the rank (*r*) represents the number of correlation or cointegration vectors under the null hypothesis (H_r). In the Pearson correlation test, the p-value is derived from the correlation coefficient (r) using its sampling distribution. If $p \leq 0.05$, the correlation between the variables is determined to be statistically pronounced; otherwise, it is not [13].

The Augmented Dickey-Fuller (ADF) test determines if a time series is stationary ($p \leq 0.05$) or non-stationary ($p \geq 0.05$). A stationary series is one in which the statistical properties of the spread between variables remain constant over time, indicating a required condition for confirming cointegration. The Engle-Granger two-step method is used to determine whether two non-stationary variables are cointegrated ($p \leq 0.05$), showing the presence of a long-term equilibrium relationship; however, this method is limited to detecting a single cointegrating relationship. The Johansen method extends this approach by providing a statistical framework for identifying long-term equilibrium relationships among multiple non-stationary variables. More details about these methods are available in the literature [13]. In this study, a “time series” refers to a sequence of observations recorded continuously over a specific period.

For future movement prediction, we employed two supervised learning algorithms: a decision tree (DT) and a random forest (RF). Both models make predictions through a hierarchical sequence of decision rules [14]. A decision tree is a single tree-like structure, making it easy to visualize and interpret the decision-making process and outcomes. On the other hand, a random forest is an ensemble learning method that combines multiple decision tree structures to improve predictive performance. This approach can minimize the risk of overfitting, which is common in single decision tree models. As a result, random forests can generally demonstrate higher and more robust predictive accuracy. However, the method requires greater computational resources and may be slow due to the complexity of aggregating multiple decision processes.

RESULTS

Pair Analysis

The stock trading dataset for the selected companies was obtained from Yahoo Finance, which provides open, high, low, close, and adjusted close prices [15]. The exchange rate between the Korean won and the US dollar was also added. The data was imported into a Python script and cleaned during preprocessing by removing entries corresponding to non-trading days (e.g., weekends and holidays) and periods of unusual activity caused by extreme volatility or regulatory events. Subsequently, the stock prices of SK and SE were converted from Korean won to US dollars to enable direct comparison with the other companies. In this study, the conversion process is referred to as "normalization". A new data frame containing only the adjusted closing price for each company was then created and further transformed into a structure suitable for statistical analysis. Using the "statsmodels" Python module [18], the Pearson Correlation, the Augmented Dickey-Fuller (ADF) test, the Engle-Granger two-step method, and Johansen Cointegration tests were applied. The results of these analyses were visualized using heat maps to facilitate comparison among the company pairs.

Figure 2 shows the heatmap of correlation values for all company pairs, along with a table of the selected correlations (> 0.6) and their corresponding p-values. Overall, the correlation coefficients range between -1 and 1, indicating varying degrees of association among the companies. Five pairs exhibit relatively high correlation. For example, the AMD-NVDA pair shows a correlation coefficient of 0.72, indicating a strong positive relationship. However, the corresponding p-values are extremely small, so it is very unlikely that the observation of such a correlation by chance is negligibly low. Therefore, this relationship can be considered as "statistically rock-solid", meaning that the result is statistically unlikely to happen in random [13].

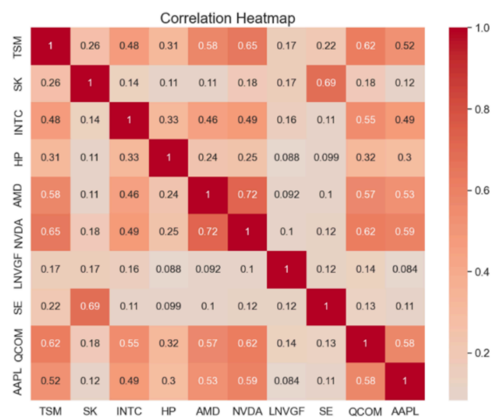

Figure 2: The heatmap of the Pearson Correlation test with p-values for all entries (10 selected companies). The scale bar is for a 0-to-1 scale, and the number inside each block indicates the p-value for the examined pairs.

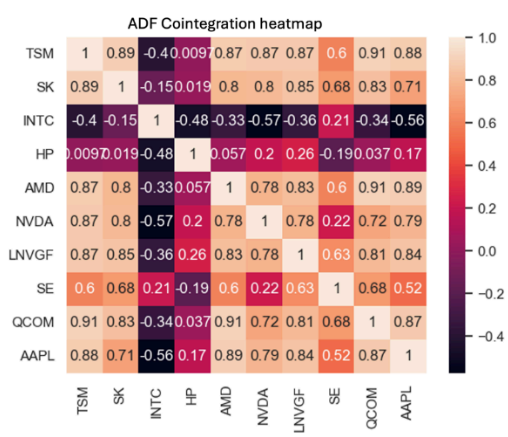

Figure 3: The heatmap of the ADF (Augmented Dickey-Fuller) Cointegration test with p-values for all entries (10 selected companies). The scale bar is for a 0-to-1 scale, and the number inside each block indicates the p-value for the examined pairs.

Figure 3 presents a heatmap of p-values from the ADF test applied to individual time series to assess stationarity. All p-values are greater than 0.05, suggesting that the null hypothesis (H_0)—that the series contains a unit root—cannot be rejected. This result implies that the time series are non-stationary. Consequently, the likelihood of cointegration across the entire set of series is low, as statistical properties such as the mean, variance, and covariance change over time.

A similar outcome is observed in the cointegration heatmap produced using the Engle-Granger two-step method, as shown in Figure 4. All p-values exceed 0.05, indicating no statistically significant evidence of cointegration between the examined pairs. Furthermore, no pairs show even marginal evidence of cointegration ($p < 0.10$). This suggests that, although some pairs are correlated, the time series do not exhibit stable long-term equilibrium relationships. The observed correlations are therefore unlikely to reflect long-term equilibrium relationships and may instead arise from short-term movement or random variation.

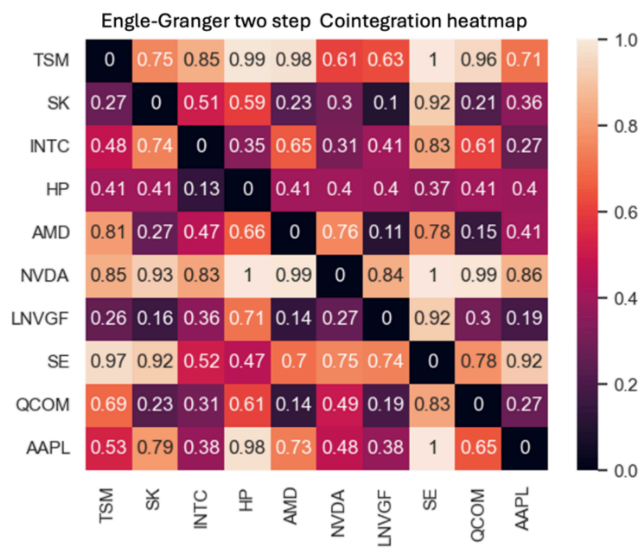

Figure 4: The heatmap of the Engle-Granger two-step Cointegration test with p-values for all entries (10 selected companies). The scale bar is for a 0-to-1 scale, and the number inside each block indicates the p-value for the examined pairs.

Unlike the previous methods, the Johansen cointegration test provides meaningful results. First, the critical values under the null hypothesis (H_0), which assumes no cointegration relationship ($r = 0$), are examined at the 95% confidence level. These critical values serve as a threshold for determining whether non-stationary time series are statistically related through a long-run equilibrium relationship [13]. Then, the trace statistic is calculated to test the null hypothesis for $r > 0$. This statistic is compared with the corresponding critical value to determine the number of cointegration relationships (r) among the variables [16]. If the trace statistic is higher than the critical values, the null hypothesis is rejected, suggesting the presence of at least one cointegrating relationship [16]. Accordingly, the pairs with trace statistics greater than > the 95% critical values (15.4943) have at least one integration relationship, which means the spread between the variables is stationary or mean-reverting. This suggests that these pairs may maintain a long-run equilibrium relationship in the trading. Table 1 lists the identified pair and their corresponding trace statistic values.

Pair	Trace Statistic (r>0)	95% CV (r>0)
SK-LNVGF	21.143496	15.4943
AMD-QCOM	18.734221	15.4943
AMD-NVDA	17.732616	15.4943
LNVGF-QCOM	16.869254	15.4943
AMD-LNVGF	16.621693	15.4943

Table 1: The optimally combined pairs with the trace statistics and 95% critical values, identified from the Johansen Cointegration test among the 10 selected companies.

Spread Analysis

As a preparation stage of future movement prediction analysis, the spread between the two variables in the optimized pairs identified through the Johansen cointegration test is evaluated using a z-score. The z-score measures the standardized deviation of the current spread from its historical mean over a specified time period. This metric is subsequently used to estimate the relative movement of the pair for the next trading day. In general, the z-score indicates the level of deviation of the current value from the historical average of the spread. Values of 0, ±1, ±2 represent: (1) neutral performance, where the spread is close to the historical mean; (2) moderate deviation, indicating potential upward or downward momentum; and (3) statistically significant deviation, suggesting an exceptional movement away from the equilibrium level. Positive (+) and negative (-) values indicate upward and downward movement, respectively. The time-series z-scores for the optimized pairs are presented in Figure 5. In addition, we explored multiple approaches, including ordinary least squares (OLS) and rolling mean and standard deviation calculations of the spread, to support the prediction analysis. These approaches can capture broader deviation from the long-run equilibrium relationship. However, in our study, for several pairs, the resulting spreads were excessively wide, making them less suitable for precise prediction. The corresponding results are provided in the shared scripts [10]

Prediction of future movement

Based on the calculated z-scores, we predicted the future performance of the optimized pairs using two machine learning algorithms: decision tree (DT) and random forest (RF), targeting the next-day trends. This approach aims to evaluate the extent to which the relationships between each pair remain effective over the long term. For qualitative evaluation, “accuracy”—defined as the proportion of correctly classified samples—was used as the primary metric. In addition, the F1 score was employed together with "precision" and "recall" to assess the performance of the model applied. The F1-score is often considered a more reliable evaluation metric for classification performance because it provides a balanced measure of precision and recall. It is defined as the harmonic (balanced) mean of precision and recall. The following formulas describe the evaluation metrics used in this study [17].

$$Accuracy = Number\ of\ correct\ predictions / Total\ number\ of\ predictions$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

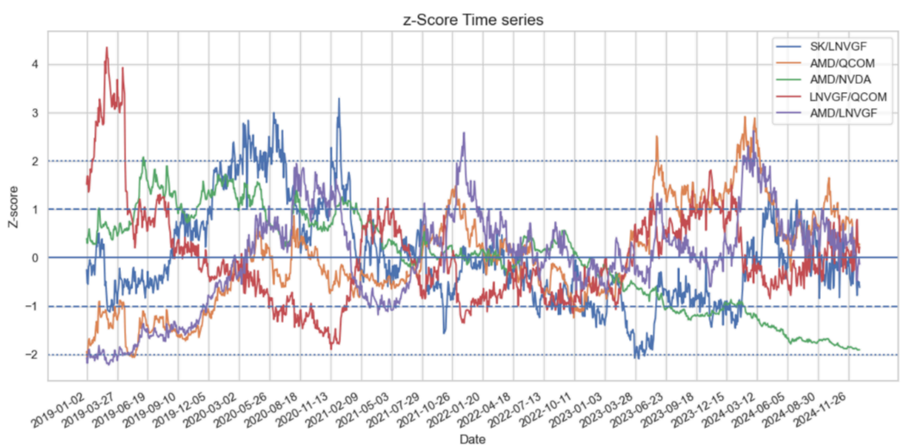

Figure 5. The time-series z-score spread values of the optimized paired companies, identified from the Johansen Cointegration test.

Prior to model prediction, the entire dataset was split into training and testing set in an 80% to 20% ratio using the train_test_split function provided by scikit-learn [18]. Specifically, 80% of the z-score data were used for model training, while the remaining 20% were used for model testing. The prediction analysis was conducted using the scikit-learn Python library [18]. The results are summarized in Table 2 and 3 for the decision tree (DT) and random forest (RF) models, respectively. The evaluation metrics are reported as a 1/100 rate (percentage). For each optimally combined pair identified from the Johansen cointegration test, the next movement is represented as a binary variable (1 or 0), and the corresponding precision, recall, F1-score, and accuracy are calculated. If the F1-score for class 1 is higher than that for class 0, the next-day prediction is interpreted as an upward movement, along with the corresponding accuracy value.

Conversely, if the F1-score for class 0 is higher than that for class 1, the next-day prediction is interpreted as downward movement. For example, for the SK/LNVGF pair using DT model, the next-day movement is predicted to be upward with an accuracy of 52.36% because the F1-score for class 1 exceeds that for class 0, (e.g., 0.66 > 0.19: namely 66% > 19%). In contrast, when using the RF model for the same pair, the predicted movement is downward with an accuracy of 49.61% because the F1-score for class 0 is higher than for class 1 (e.g., 0.61 > 0.29, likely 61% > 29%). These results suggest that, for the SK/LNVGF pair, the DT model predicts upward movement more frequently than the RF model.

Interestingly, the overall predictions exhibit different trends between the DT and RF models. The DT model provides predictions based on how a single tree splits the data, while the RF model builds multiple decision trees, each independently splitting the data on its own, and aggregates their predictions. By averaging the results from many trees, the RF model reduces overfitting caused by variability or randomness in individual trees. As a result, the RF model is generally considered more stable and robust than the DT model. Despite this intrinsic difference, the DT model in this study achieves a slightly higher average prediction accuracy (51.50%) than the RF model (48.98%). Feature importance was also evaluated for each pair in the individual prediction analysis, measuring each feature's contribution to the model's predictions during testing. However, as this ranking does not make any significant effect on the overall prediction accuracy, the corresponding results are not presented here, but shown in the open shared link [10].

Pair	Up (1) or Down (0)	Precision	Recall	f1-score	Accuracy	Next Day prediction
SK/LNVGF	0	0.67	0.11	0.19	0.5236	Up
	1	0.51	0.94	0.66		
AMD/QCOM	0	0.53	0.2	0.29	0.4803	Up
	1	0.47	0.8	0.59		
AMD/NVDA	0	0.59	0.81	0.68	0.5630	Down
	1	0.46	0.22	0.3		
LNVGF/QCOM	0	0.58	0.25	0.35	0.4961	Up
	1	0.47	0.79	0.59		
AMD/LNVGF	0	0.5	0.85	0.63	0.5118	Down
	1	0.56	0.19	0.28		

Table 2: The prediction parameters obtained using the Decision Tree model. Values are presented on 1/100 scale (e.g. 1 = 100%, 0 = 0%).

Pair	Up (1) or Down (0)	Precision	Recall	f1-score	Accuracy	Next Day prediction
SK/LNVGF	0	0.5	0.78	0.61	0.4961	Down
	1	0.48	0.21	0.29		
AMD/QCOM	0	0.52	0.3	0.38	0.4803	Up
	1	0.46	0.68	0.55		
AMD/NVDA	0	0.75	0.14	0.24	0.4764	Up
	1	0.44	0.93	0.6		
LNVGF/QCOM	0	0.54	0.27	0.36	0.4803	Up
	1	0.46	0.73	0.56		
AMD/LNVGF	0	0.51	0.78	0.61	0.5157	Down
	1	0.55	0.26	0.36		

Table 3: The prediction parameters obtained using the Random Forest model. Values are presented on 1/100 scale (e.g. 1= 100%, 0 = 0%).

CONCLUSION

In this study, we aimed to identify highly correlated pairs of semiconductor-related companies using pair trade analysis and to predict the next-day performance using machine learning models. We believed that the close relationship between chipmakers and top stream technology firms could be quantified through financial market analysis, as such relationships may be reflected in their stock market performance. We also expected that if a pair of companies are highly interconnected, their relationship may strengthen during periods of economic instability.

By applying various pair analysis models to ten major semiconductor-related companies, we observed no strong Pearson correlation from January 2, 2019, to December 31, 2024—a period encompassing the COVID-19 pandemic and recent geopolitical trade conflicts. This suggests that the companies did not consistently exhibit synchronized movements in their asset performance. A similar pattern was observed in the cointegration tests that attempt to identify tradable pairs with long-term equilibrium relationships. However, only the Johansen cointegration test revealed statistically significant cointegration relationships among several stock price series and identified optimized pairs among the selected stock price series of the studied companies. The observed optimal pairs are considered statistically robust because the Johansen method evaluates multiple variables simultaneously. These identified cointegrating relationships

can be interpreted as certain company pairs maintain long-term equilibrium dynamics, potentially influenced by broad market fluctuations and macroeconomic factors.
Recent structural changes in the semiconductor market over a relatively short period have made it difficult to evaluate with simple analytical models. Thus, as we hypothesized, examining inter-firm relationships provides valuable insights into industry dynamics and expectations for future semiconductor cycles. Since measuring market volatility is still highly challenging, particularly in the rapidly evolving semiconductor industry, we estimated the future performance of optimized company pairs based on their z-scores, which quantify deviations from the historical mean. In this way, the spread in the stock performance between paired companies can be characterized in advance.

Pair trading analysis has been widely accepted as a strategic tool for evaluating volatile sectors like semiconductors. One of its key advantages is the ability to mitigate market risk by evaluating the relevant performance of two related entities, for example, a chip manufacturer and a major tech consumer in this study. Our analysis relies heavily on the relationship between two companies across these two sectors, examining their vertical interdependence. Although no strong structural relationship was observed between the selected companies, this study seeks to quantify the resilience of the semiconductor-related trade dynamics by introducing a new method based on Johansen cointegration, extending beyond simple price tracking. This attempt represents a novel contribution related to traditional pair trading strategies, which typically rely solely on either correlation or cointegration metrics.

Using two supervised learning algorithms, decision tree (DT) and random forest (RF), we predicted the next-day movements of the optimized cointegrated pairs. The prediction accuracy exceeded 50%, which falls within a practically acceptable range for financial forecasting (approximately ~50-60%). Despite differences in analytical methodology, the DT and RF models produced slightly similar results, with the RF model yielding marginally lower prediction accuracy in this study. This result indicates that predictive performance may be improved by incorporating additional models or alternative feature sets. We also examined spread measures derived from rolling means and rolling standard deviations, in addition to z-scores. However, these features varied substantially across company pairs, reducing their suitability for consistent prediction. Our future studies would focus more on advanced ensemble forecasting methods, such as Extreme Gradient Boosting (XGBoost), Gradient Boosting, or Support Vector Machine (SVM), to improve predictive performance. Although our technical approach, which applies machine learning algorithms to the Z-spread between the two co-integrated companies, demonstrates a reasonable level of accuracy, several technical challenges remain before it can achieve practical market performance, especially given the inherent difficulty of trading market prediction. The key issues to be addressed include: 1. Look-ahead Bias in Z-score calculation: Z-score values depend on the choice of a rolling window (20 or 60 days), which can introduce bias if not handled carefully. 2. Limitation of raw Z-spread: The Z-spread does not fully capture dynamic features such as the rate of changes (e.g., velocity), which may carry important predictive information, 3. Definition of prediction framework (classification or regression): The prediction must be defined for the next-day spread (Regression) or the direction (Classification). These challenges are not addressed in this present study but will be dealt with in the future study.

In summary, our analysis demonstrates that evaluating the relationship between semiconductor manufacturers and chip-consuming technology companies is feasible even under intensified economic and geopolitical uncertainty. The results provide a useful framework for understanding how the market performance of chip-consuming firms may influence the long-term dynamics of semiconductor manufacturers.

REFERENCES

1. Singh, M., Sargent Jr., J.F. and Sutter, K.M. (2023) *Semiconductors and the semiconductor industry*. Congressional Research Service (CRS) Report R47508. Available at: https://crsreports.congress.gov (Accessed: 29 April 2026).
2. Yeboah, L., Oppong, P., Abdul Malik, A., Acheampong, P., Morgan, J., Addo, R. and Henyo, B.W. (2024) 'Exploring innovations, sustainability and future opportunities in semiconductor technologies', *International Journal of Advanced Nano Computation Analysis*, 3(2). Available at: https://doi.org/10.20944/preprints202409.1307.v3.
3. Haramboure, A., Lalanne, G., Schwellnus, C. and Guilhoto, J. (2023) *Vulnerabilities in the semiconductor supply chain*. OECD Science, Technology and Industry Working Papers. Available at: https://doi.org/10.1787/6bed616f-en.
4. Ivanova, M. and Ivanov, D. (2024) 'Responding to the ripple effect from systemic disruptions: empirical evidence from the semiconductor shortage during COVID-19', *Modern Supply Chain Research and Applications*, 6(4), pp. 354–375. Available at: https://doi.org/10.1108/MSCRA-03-2024-0011.
5. Alexander, P. (2025) 'The price of semiconductors and the macroeconomy: a real business cycle approach', *Economics*, 23 June.
6. Bibi, J. (2025) 'The US–China semiconductor war: how it impacts South Korea's tech industry', *International Journal of Social Science Bulletin*, 3(6), 25 June.
7. Sarmento, S.M. and Horta, N. (2020) 'Enhancing a pairs trading strategy with the application of machine learning', *Expert Systems with Applications*, 158, p. 113490. Available at: https://doi.org/10.1016/j.eswa.2020.113490.
8. Companies Market Cap (no date) 'Largest semiconductor companies by market cap'. Available at: https://companiesmarketcap.com/semiconductors/largest-semiconductor-companies-by-market-cap/ (Accessed: 29 April 2026).
9. Kydland, F.E. and Prescott, E.C. (1982) 'Time to build and aggregate fluctuations', *Econometrica*, 50(6), pp. 1345–1370. Available at: https://doi.org/10.2307/1913386.
10. shin44649-commits (no date) *Stock Analysis1*. Available at: https://github.com/shin44649-commits/Stoct-Analysis1.git (Accessed: 29 April 2026).
11. Wang, T. and Xu, G. (2024) 'The correlation between different country stock market indices', *Highlights in Business, Economics and Management*, 46, pp. 283–286. Available at: https://doi.org/10.54097/n1hfy470.
12. Ha, G.G., Lee, J.W. and Nobi, A. (2015) 'Threshold network of a financial market using the p-value of correlation coefficients', *Journal of the Korean Physical Society*, 66, pp. 1802–1808. Available at: https://doi.org/10.3938/jkps.66.1802.

13. Gujarati, D.N. and Porter, D.C. (2009) *Basic econometrics*. New York: McGraw-Hill Irwin.
14. Kinasih, A.N.S., Handayani, A., Ardiansah, J.T. and Damanhuri, N.S. (2024) 'Comparative analysis of decision tree and random forest classifiers for structured data classification in machine learning', *Science in Information Technology Letters*, 5(2), pp. 13–24. Available at: https://doi.org/10.31763/sitech.v5i2.1746.
15. Yahoo Finance (no date) *Yahoo Finance*. Available at: https://finance.yahoo.com (Accessed: 29 April 2026).
16. Dwyer, G.P. (2015) *The Johansen test for cointegration*. White Paper, April.
17. Yacouby, R. and Axman, D. (2020) 'Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models', in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*. Available at: https://aclanthology.org/2020.eval4nlp-1.9.
18. Statsmodels (no date) *Statsmodels documentation*. Available at: https://www.statsmodels.org/stable/index.html (Accessed: 29 April 2026).

ABOUT THE AUTHOR

Alex Tayeon Shin is a junior at Saint Paul Daechi Academy, interested in business with plans to pursue a major in business administration. He hopes to build his career as a business leader in the field of international commerce.