Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.27373
Page: 807-818

Comparative Performance of Statistical and LSTM Based Arbitrage in the Indonesian Stock Market

[1] **Yunita**
School of Computing, Telkom University, Bandung, 40257, Indonesia

[2] **Indwiarti**
School of Computing, Telkom University, Bandung, 40257, Indonesia

[3] **Deni Saepudin**
School of Computing, Telkom University, Bandung, 40257, Indonesia

Article Info

Article history:

Accepted, 26 December 2025

Keywords:

Cointegration;
Deep arbitrage;
Indonesian stocks;
LSTM;
Statistical arbitrage.

ABSTRACT

This study compares the performance of traditional statistical arbitrage and Long Short-Term Memory (LSTM)-based deep arbitrage strategies in generating returns and risk-adjusted performance in the Indonesian stock market. A quantitative approach is employed using long-only trading simulations on daily closing prices of blue-chip financial sector stocks listed on the Indonesia Stock Exchange from April 2015 to April 2025. Stock pairs are selected based on correlation and cointegration criteria, while spread volatility is modeled using a GARCH (1,1) framework. To ensure a genuine out-of-sample evaluation, the sample is divided into an in-sample period from April 2015 to August 2021 for model training and parameter optimization, and an out-of-sample period from September 2021 to April 2025 for performance assessment. Strategy performance is evaluated using portfolio return and Sharpe ratio. The empirical results show that both strategies are feasible in the Indonesian market; however, the LSTM-based deep arbitrage strategy significantly outperforms the traditional statistical arbitrage approach, achieving a higher out-of-sample portfolio return (735% versus 482%) and a superior Sharpe ratio (1.67 versus 0.69). These findings indicate that deep learning-based arbitrage can provide substantial improvements in both return and risk-adjusted performance under long-only trading constraints in an emerging market context.

Corresponding Author:

Yunita,
School of Computing,
Telkom University, Bandung, Indonesia
Email: yyunitaa3@gmail.com

1. INTRODUCTION

The Indonesian stock market has experienced substantial growth in recent years, both in terms of investor participation and trading activity. As of September 25, 2024, the number of registered investors reached 6,001,573 Single Investor Identifications (SIDs), with approximately 744,000 new investors added within the same year, of which nearly 79% are under the age of 40 [1]. Beyond signaling increasing market participation, this growth has important implications for quantitative trading strategies. Higher investor participation is typically associated with improved liquidity, tighter bid–ask spreads, and lower effective transaction costs, all of which are critical conditions for the feasibility of arbitrage-based trading strategies, particularly in emerging markets.

Within this context, statistical arbitrage has been widely studied as a quantitative trading framework that exploits mean-reversion behavior between pairs of assets exhibiting strong comovement. Since its development in the 1980s, statistical arbitrage has relied on applied mathematical tools such as correlation analysis, cointegration testing, and stochastic modeling of price spreads to identify temporary mispricing opportunities [2]. In this study, arbitrage is formulated as a relative-value trading strategy, where investment decisions are based on the relative mispricing between two economically related assets rather than absolute price movements.

Despite its mathematical elegance, traditional cointegration-based statistical arbitrage faces several practical and modeling limitations. Fixed trading thresholds may fail to adapt to changing volatility regimes, while unadjusted spread dynamics can lead to delayed or noisy trading signals. These issues reduce strategy robustness, especially under time-varying market volatility [3]. From an applied mathematics perspective, these limitations motivate the need for volatility-aware modeling and systematic threshold optimization to improve signal stability and trading performance[4].

Recent advances in applied machine learning have introduced deep arbitrage strategies based on Long Short-Term Memory (LSTM) networks, which are designed to model nonlinear temporal dependencies in financial time series. Prior studies report that LSTM-based approaches can outperform traditional statistical arbitrage in terms of portfolio return and risk-adjusted performance [5]However, most existing work focuses on developed markets, relies on high-frequency data, and assumes unrestricted long–short trading mechanisms. Such settings limit the direct applicability of these methods to regulated emerging markets, where arbitrage strategies must operate under long-only trading constraints while still preserving a relative-value interpretation of spread dynamics.

In the Indonesian stock market, empirical evidence comparing traditional statistical arbitrage and deep arbitrage under daily data frequency and long-only trading constraints remains limited[6]. This gap is particularly important from an applied mathematics standpoint, as it raises questions about how classical econometric tools (cointegration and volatility modeling) can be systematically integrated with modern deep learning techniques to implement relative-value arbitrage when short selling is not permitted [7]. Therefore, this study aims to empirically compare cointegration-based statistical arbitrage and LSTM-based deep arbitrage strategies by evaluating portfolio return and Sharpe ratio using Indonesian financial sector stocks within a relative-value, long-only trading framework.

From a modeling perspective, this study explicitly formulates an applied mathematical pipeline consisting of: (i) correlation and cointegration analysis to identify equilibrium relationships between stock pairs; (ii) construction of residual-based price spreads as a measure of relative mispricing; (iii) volatility scaling of spreads using a GARCH(1,1) model; (iv) optimization of trading thresholds via grid-search techniques; (v) signal generation through both rule-based statistical arbitrage and LSTM-based classification; and (vi) out-of-sample performance evaluation using portfolio return and Sharpe ratio [8]. By systematically justifying each modeling stage, this research contributes to the applied mathematics literature by demonstrating how classical stochastic models and modern deep learning can be integrated to enhance relative-value arbitrage performance under long-only constraints in an emerging market setting[9].

This study makes three primary contributions. First, it provides one of the first systematic comparisons between traditional statistical arbitrage and LSTM-based deep arbitrage in the Indonesian stock market under a relative-value, long-only trading assumption[10]. Second, it extends the arbitrage literature by focusing on daily data and long-only trading rules, reflecting realistic regulatory and market conditions. Third, it highlights the role of volatility-adjusted spread modeling and threshold optimization as key applied mathematical components for improving arbitrage robustness and risk-adjusted performance.

2. RESEARCH METHODE

The data were obtained from AN, a paid web-based financial data provider that supplies historical equity data for the Indonesian Stock Exchange (IDX). The dataset consists of daily closing prices of seven blue-chip financial sector stocks listed on the Indonesian Stock Exchange, covering a ten-year period from April 6, 2015 to April 6, 2025, with a total of 2,418 trading-day observations. All prices are denominated in Indonesian Rupiah (IDR) and follow the official IDX trading calendar (Jakarta time, UTC+7), where non-trading days correspond to weekends and official market holidays[2].

The raw dataset contains daily time, open, high, low, close, and trading volume information. However, this study utilizes only the time and closing price data, as the arbitrage strategies are based on price spread dynamics and cointegration analysis, which primarily rely on closing prices. This study uses adjusted closing prices, which account for corporate actions such as stock splits and dividends to ensure price continuity and avoid artificial structural breaks in the time series[11]. Days with missing observations due to market closures were not forward-filled; instead, the chronological trading sequence defined by the IDX calendar was preserved.

Stocks were selected based on sector classification and continuous daily trading activity to ensure data availability and liquidity. Only stocks with uninterrupted trading histories during the sample period were included. Consequently, survivorship bias may be present due to the exclusion of delisted or inactive stocks; however, this approach ensures data consistency and realistic arbitrage execution. The ten-year observation period spans multiple market regimes, including pre-pandemic, pandemic, and post-pandemic phases of the Indonesian stock market, which enhances the robustness and generalizability of the empirical results.

To ensure a genuine out-of-sample evaluation, the dataset is divided chronologically into a training (in-sample) period and a testing (out-of-sample) period. The in-sample period spans from April 6, 2015 to August 31, 2021 and is used for model training and parameter optimization, including threshold selection and LSTM training[2], [12]In the deep arbitrage model, the input features consist of historical spread values derived from paired stock prices, while the target variables are defined as trading signals indicating buy and sell actions based on predefined threshold rules. The out-of-sample period covers September 1, 2021 to April 6, 2025 and is reserved exclusively for performance evaluation, where the trained model receives only historical spread information to generate buy and sell signals. All trading signals are executed using a close-to-close convention, where buy and sell decisions are based on information available at the market close and transactions are assumed to occur at the corresponding closing prices. This design ensures consistency with long-only trading constraints and portfolio return calculations [13].

The arbitrage strategy involving two paired stock data has two requirements for the data pair, it must be correlated and cointegrated. The historical correlation of a stock pair is said to be suitable if the correlation value was $(r) \geq 0.8$. The correlation test can use Formula 1. The correlation test uses the daily closing stock prices. The correlation value is denoted by r. Calculating the correlation value requires several variables, including X_i dan Y_i as the closing stock price, with i, $\bar{X}$ and $\bar{Y}$ as the average closing stock price and n as the amount of data. The cointegration test can use the stationarity test with the Augmented Dickey-Fuller (ADF) test as in the Formula 2, with a p-value < 0.05. A significance level of 0.05 means that there is a 5% risk of rejecting H0, where a value of 5% is considered small for stationarity and cointegration calculation [14]. The cointegration test uses the log value of the closing stock price, denoted by ln. This caused by how it can stabilize variance and reduce the effect of data scale [15]. Thus, it can be analyzed more accurately. The ADF test requires the variables α and γ as regression parameters, ε_t as the error term, $\ln(P_t)$ and $\ln(P_{t-1})$ as the log closing stock prices at time t.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=0}^{n}(X_i - \bar{X})^2 \sum_{i=0}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

$$\Delta \ln(P_t) = \alpha + \gamma \ln(P_{t-1}) + \varepsilon_t \quad (2)$$

The ADF test has two hypotheses, hypothesis 0 and hypothesis 1. Hypothesis 0 states that "The time series data has a unit root and is non-stationary." On the contrary, hypothesis 1 states that "The time series data does not have a unit root and is stationary." A pair of stocks is said to be cointegrated if each stock is non-stationary (h0 accepted), but their linear combination produces a stationary residual (h0 rejected) [16]. After ensuring that the stock pair meets these two conditions, two models will be created using statistical arbitrage and deep arbitrage.

Creating a statistical arbitrage model begins by taking the spread value between stocks at each point in time using Formula 3. Calculating the spread requires two regression parameters, namely α and β. In addition, the closing log prices of the two stocks are denoted by $\ln(P_{A,t})$ dan $\ln(P_{B,t})$.

$$spread_t = \ln(P_{A,t}) - (\alpha + \beta \ \ln(P_{B,t})) \quad (3)$$

Then, volatility and mspread values are needed at each point in time to develop a trading strategy. Volatility can be calculated using the GARCH(1,1) method with Formula 4. GARCH(1,1) is used because this method has proven to be more effective than other volatility calculation methods [17]. The volatility at each time is denoted by σ_t. Calculating volatility requires GARCH(1,1) parameters such as ω, α, β, and the spread value at time t by $spread_{t-1}$.

$$\sigma_t = \sqrt{\omega + \alpha spread^2{}_{t-1} + \beta \sigma^2{}_{t-1}} \quad (4)$$

Calculating mspread(t) can be done using Formula 5. The variables needed to calculate mspread(t) are spread(t) and the spread average generated in Formula 3. In addition, trading costs denotated by c are required. The value of c is set at 0.5%, as used on investment platforms in Indonesia[18].

$$mspread_t = spread_t - \overline{spread} - c \quad (5)$$

Statistical arbitrage trading strategies include the timing of buying stocks, selling stocks, and stop-losses. The timing of buying stock A can use the conditions in Formula 6, and the timing of buying stock B can use the conditions in Formula 7. The timing of selling stocks can use the conditions in Formula 8. The stop-loss timing for stock A can use the conditions in Formula 9, and the stop-loss timing for stock B is in Formula 10. Stop-loss is a rule that stipulates that an investment position will be sold automatically when the price reaches a predetermined maximum loss level These conditions are applied in the hope of making a profit in the period between buying and selling the stock.

In Formulas 6 to 10, there are threshold values with variables λ_1, λ_2, and λ_3. Maximizing the threshold value can be done using the gridsearch method. Gridsearch compares the portfolio return values obtained from all combinations of λ_1, λ_2, and λ_3. The value of λ_2 is in the range of 0-10, λ_1 is in the range of λ_2-30, and λ_3 is in the range of λ_1-50. The value of λ_1 will definitely be greater than λ_2, because λ_1 represents a signal to buy stocks as the spread value is considered to be increasing. The value of λ_2 represents a normal spread. This becomes a signal to sell stocks at a profit, because the spread has returned to normal. The value of λ_3 represents a stop-loss signal because the spread is widening, which is considered a loss when selling stocks.

$$mspread_t > (\lambda_1 * \sigma_t) \,\&\, mspread_t < (\lambda_3 * \sigma_t) \quad (6)$$

$$mspread_t < (-\lambda_1 * \sigma_t) \,\&\, mspread_t > (-\lambda_3 * \sigma_t) \quad (7)$$

$$-\lambda_2 * \sigma_t < mspread_t < \lambda_2 * \sigma_t \quad (8)$$

$$mspread_t > \lambda_3 * \sigma_t \quad (9)$$

$$mspread_t < -\lambda_3 * \sigma_t \quad (10)$$

Definition of Symbols
t : time index
mspreadt : mispricing spread at time t
σt: volatility of the mispricing spread at time t
λ1: lower threshold parameter for trading signal activation
λ2: threshold parameter defining the no-trading (neutral) region
λ3: upper threshold parameter indicating extreme mispricing
rtr: portfolio return at time t
Vt : portfolio value at time t

Simulation visualization requires a z-score to make it easier to read. Calculating the z-score can be done by dividing the mspread by volatility, as shown in Formula 11.

$$z_t = \frac{mspread_t}{\sigma_t} \quad (11)$$

The deep arbitrage strategy utilizes the previously constructed spread and z-score series derived from paired stock prices. Prior to model training, each observation is labeled based on its corresponding z-score following the decision rules defined in Formula 12. Label 1 represents a buy signal for stock A, label 2 represents a buy signal for stock B, label 3 represents a sell signal, label 4 represents a stop-loss signal, and label 0 represents a hold or no-action signal.

The input to the LSTM model is formulated as rolling sequences of historical spread values using a fixed lookback window of 20 trading days. Specifically, at each time step ttt, the input sequence is defined as [spreadt−20, ..., spreadt−1] [spread_{t-20}, \ldots, spread_{t-1}] [spreadt−20, ..., spreadt−1], while the target output corresponds to the trading signal label at time ttt. This sequence-based representation enables the LSTM to capture temporal dependencies and dynamic mean-reversion behavior in the spread process.

The LSTM model adopts a stacked architecture consisting of two LSTM layers, each with 50 hidden units, followed by a fully connected dense layer with a softmax activation function for multi-class classification[19]. Dropout regularization with a rate of 0.2 is applied between LSTM layers to mitigate overfitting. Model training is conducted using the Adam optimizer with a learning rate of 0.001, and categorical cross-entropy is employed as the loss function.

Training is performed using a batch size of 32 over a maximum of 100 epochs. Early stopping is implemented with a patience of 10 epochs based on validation loss to prevent overfitting. To address class imbalance arising from the predominance of hold (label 0) observations, class weighting is applied during model training to penalize misclassification of minority trading signals.

To ensure a realistic and bias-free evaluation, a rolling-window training framework with an expanding window scheme is employed. The model is initially trained using the in-sample period. Subsequently, at each re-training step, the training window is expanded to include all available historical observations up to the current time point, and the trained model is used to generate trading signals exclusively for future observations. This rolling-window protocol ensures that the LSTM model adapts to evolving market conditions while strictly preserving the out-of-sample nature of performance evaluation.

$$signal(z_t) = \begin{cases} 4, & |z_t| > \lambda 3 \\ 3, & |z_t| < \lambda 2 \\ 2, & \lambda_1 < z_t < \lambda_3 \\ 1, & -\lambda_3 < z_t < -\lambda_1 \\ 0, & others \end{cases} \tag{12}$$

The generated labels serve as the target outputs of the model, while the spread time series spread(t)spread(t)spread(t) is used as the input feature. Rather than applying a percentage-based data split, this study adopts a chronological date-based partitioning to ensure a realistic time-series evaluation and to prevent data leakage. The dataset is divided into an in-sample period used exclusively for model training and parameter optimization, and an out-of-sample period reserved solely for performance evaluation.

All model training, including LSTM parameter learning and threshold optimization, is conducted using data from April 6, 2015 to August 31, 2021. The out-of-sample period spans from September 1, 2021 to April 6, 2025, during which the trained models generate trading signals using only historical information available up to each decision point. This chronological split ensures that no future information is used during model training, thereby eliminating look-ahead bias.

Input sequences are normalized using statistics computed exclusively from the in-sample period. These normalization parameters are then fixed and applied to the out-of-sample data, ensuring that information from the test period does not influence model training. The processed input sequences are subsequently fed into the LSTM model to generate buy, sell, and stop-loss signals for paired stocks[20].

Both statistical arbitrage and deep arbitrage models generate portfolio return and Sharpe ratio values, which are used as the primary performance evaluation metrics. Portfolio return is calculated using Formula 15, while the Sharpe ratio is computed using Formula 19. Trading simulations are visualized using time-series plots and summarized in transaction tables. The trading simulation plots include markers indicating buy and sell executions for each stock, as well as threshold boundaries corresponding to entry, exit, and stop-loss conditions.

Each row of the trading transaction table will be calculated for portfolio return per transaction. Then, the equity per transaction will be calculated, resulting the portfolio return. The portfolio return per transaction is calculated using Formula 13 in decimal form, by subtracting the closing stock log price at the time of purchase from the closing stock log price at the time of sale. Equity per transaction can be calculated using Formula 14. Equity is denotated by V and t denotates the time. Portfolio return can be calculated using Formula 15, which requires V_T as the final equity value and V_0 as the initial equity value.

$$\mathrm{Return_t} = \ln(P_{sell,t}) - \ln(P_{buy,t}) \tag{13}$$

$$\mathrm{V_t} = \mathrm{V_{t-1}} * (1 + \mathrm{Return_t}) \tag{14}$$

$$\mathrm{Return_{portfolio}} = \left(\frac{V_T - V_0}{V_0}\right) * 100\% \tag{15}$$

The academic literature does not prescribe a fixed benchmark for interpreting the Sharpe ratio. However, in practical investment analysis, a Sharpe ratio greater than 1.0 is often viewed as suggesting good risk-adjusted performance [21]. This rule of thumb should be interpreted with caution because the Sharpe ratio is sensitive to the time horizon, volatility characteristics, and non-normal return distributions.

The risk parameter in the Sharpe ratio calculation is set to zero because for daily data, its value is very small compared to the market return. Before calculating the Sharpe ratio using Formula 18, several variables are needed, such as the average daily portfolio return, or denoted by $\overline{r_t}$. Also, the standard deviation of the portfolio return, or σ_r. The daily portfolio return can be calculated using Formula 16, with the position in A held, with a range of 0 to 1, denoted by ${p_t}^A$. The value of ${p_t}^B$ has the same meaning as ${p_t}^A$, but uses stock B. The value of σ_r can be calculated using Formula 17.

$$r_t = {p_t}^A * (\ln(P_{A,t}) - \ln(P_{A,t-1})) + {p_t}^B * \Delta \ln(P_{B,t}) \tag{16}$$

$$\sigma_r = \sqrt{\frac{1}{N-1}\sum_{t=0}^{N}(r_t - \bar{r})^2} \tag{17}$$

$$\text{Sharpe} = \frac{\bar{r_t}}{\sigma_r} * \sqrt{252} \tag{18}$$

3. RESULT AND ANALYSIS

Before the data pairs are processed, pre-processing is necessary to facilitate data processing because the data is clean and in the required format. First, stock closing data must be in numerical form and time data must be in year-month-date format. Next, all data entered must be in the same date order as the other data. Next, a data plot was generated to observe the price movements of all imported stock data. The data plot in this study is presented in Figure 1.

Figure 1. Daily closing price dynamics of selected Indonesian financial sector stocks (2015–2025)

In Figure 2, the highest correlation value is 0.92. This correlation level is observed in both stock pairs: BBCA–BBRI and CIMB–BMRI. A correlation value of 0.92 has exceeded the minimum threshold of 0.8 commonly required for pairs trading in statistical arbitrage strategies. Therefore, the next step is to examine their cointegration using the ADF test. While both BBCA–BBRI and CIMB–BMRI satisfy the statistical prerequisites for pairs trading, the performance outcomes differ across pairs [18], [22]. The BBCA–BBRI pair consistently delivers higher portfolio returns and Sharpe ratios under both strategies, suggesting more stable mean-reversion dynamics and superior spread behavior. In contrast, the CIMB–BMRI pair exhibits weaker trading performance despite comparable correlation levels, indicating that correlation and cointegration alone are not sufficient determinants of arbitrage profitability[22].

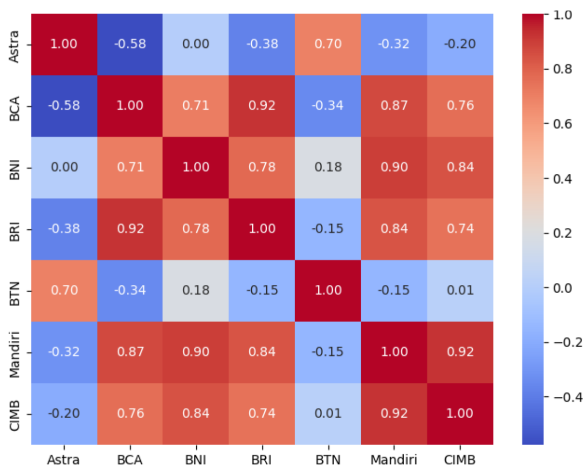

Figure 2. Stock pair correlation heatmap (darker colors indicate stronger positive correlations)

Based on Table 1, the p-value is obtained by testing stationarity using the ADF test. The BBCA and BBRI pair generate p-value with the ADF test using residual data. Meanwhile, each stock in the pair generates p-value with the ADF test using closing price data. Table 1 shows that both the BBCA–BBRI and CIMB–BMRI pairs are stationary. Considering the price movement patterns shown in the data plot in Figure 1, BCA–BRI exhibits a more similar behavior compared to CIMB–BMRI. Thus, this study will proceed by implementing the statistical arbitrage and deep arbitrage methods using the BBCA–BBRI stock pair.

Table 1. Correlation and Cointegration

Stock	Correlation	P-value	Explanation
BBCA-BBRI	0.92	0.004	
BBCA	-	0.71	
BBRI	-	0.45	Meeting the correlation and cointegrated requirements
CIMB-Mandiri	0.92	0.002	
CIMB		0.58	
Mandiri		0.65	

The process begins by calculating the spread at each time $t$for the BBCA–BBRI stock pair using Formula 3. Then, the volatility at each time $t$is calculated using the GARCH (1,1) method as shown in Formula 4. After that, the modified spread (mspread) at each time $t$is computed using Formula 5. Finally, the z-score at each time $t$is obtained using Formula 11.

After deriving the mspread and volatility variables, the optimal threshold values are selected through a grid search, following the procedure described in the Research Method chapter. This study produces the optimal threshold values of $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, and $\lambda_3 = 3.0$. By applying these thresholds in the statistical arbitrage and deep arbitrage models, the portfolio return results are obtained as presented in Table 2.

To assess whether the observed differences in portfolio performance are statistically meaningful, formal statistical inference was conducted. Mean return differences between the statistical arbitrage and deep arbitrage strategies were evaluated using two-sided t-tests. To account for potential serial correlation and heteroskedasticity in daily stock returns, Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) standard errors were applied. Statistical significance was evaluated using p-values, and 95% confidence intervals were reported.

Table 2. Portofolio Return and Sharpe Ratio Results of Two Methods

Method	Portfolio Return	Sharpe Ratio
Statistical Arbitrage	482%	0.69
Deep Arbitrage (LSTM)	735%	1.67

Figure 3 presents a visualization of trading simulations using the statistical arbitrage method. In this method, there are 61 stock buy-sell transactions, divided into 30 BBCA stock buy-sell transactions and 31 BBRI stock buy-sell transactions. In addition, Figure 4 presents a visualization of trading simulations using the deep arbitrage (LSTM) method. The deep arbitrage method with the LSTM model resulted in 37 stock buy-sell transactions, divided into 18 BBCA stock buy-sell transactions and 19 BBRI stock buy-sell transactions.

Figure 3. Trading simulation of the statistical arbitrage strategy using Z-score spread dynamics

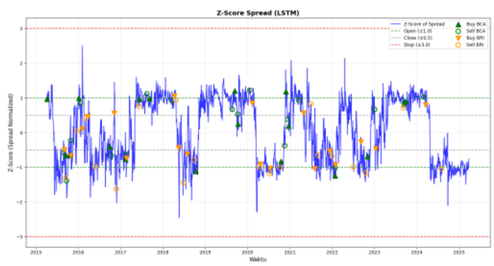

Figure 4. Trading simulation of the LSTM-based deep arbitrage strategy using spread dynamics.

A trading transaction table needs to be created to clarify the timing of buying and selling stocks. Therefore, Table 3 and Table 4 are presented to provide a clear overview of trade execution timing. Table 3 reports the trading transactions generated by the statistical arbitrage strategy, while Table 4 presents the trading transactions generated by the LSTM-based deep arbitrage strategy. For brevity and readability, only the first five and the last five trading transactions are shown for each strategy, while the full transaction records are omitted.

Table 3. Statistical Arbitrage Trading Transactions

No	Buy date (year-month-date)	Buy details	Sell date (year-month-date)	Sell details	Return (%)
1	2015-04-08	Buy BBCA	2015-05-06	Sell BBCA	-10.4%
2	2015-05-12	Buy BBCA	2015-06-18	Sell BBCA	-0.72%
3	2015-06-24	Buy BBRI	2015-07-03	Sell BBRI	5.07%
4	2015-09-07	Buy BBRI	2015-08-04	Sell BBRI	0%
5	2015-11-02	Buy BBRI	2015-08-28	Sell BBRI	2.44%
...	...	...	...	...	...
57	2023-04-27	Buy BBCA	2023-03-27	Sell BBCA	3.88%
58	2023-10-25	Buy BBCA	2023-04-18	Sell BBCA	1.38%
59	2024-04-26	Buy BBCA	2024-04-24	Sell BBCA	8.74%
60	2024-09-04	Buy BBRI	2024-08-26	Sell BBRI	11.11%
61	2024-09-25	Buy BBRI	2025-03-27	Sell BBRI	-23.94%

Table 3 reports the trading transactions generated by the statistical arbitrage strategy over the full evaluation period. Each transaction record includes the corresponding buy and sell dates, the traded asset, and the realized transaction-level return. The results indicate that the statistical arbitrage strategy generates a mix of profitable and loss-making trades, reflecting the inherent sensitivity of threshold-based mean-reversion signals to changing market conditions. Several transactions exhibit negative returns, particularly during periods of heightened volatility, suggesting that fixed entry and exit thresholds may lead to suboptimal trade timing when market dynamics deviate from stable mean-reversion behavior.

Table 4. Deep Arbitrage Trading Transactions

No	Buy date (year-month-date)	Buy details	Sell date (year-month-date)	Sell details	Return(%)
1	2015-04-06	Buy BBCA	2015-08-28	Sell BBCA	0.26%
2	2015-09-02	Buy BBRI	2015-09-14	Sell BBRI	12.59%
3	2015-09-17	Buy BBCA	2015-09-21	Sell BBCA	2.94%
4	2015-10-21	Buy BBCA	2015-10-28	Sell BBCA	-2.07%
5	2015-10-30	Buy BBRI	2015-12-22	Sell BBRI	15.22%
...	...	...	...	...	...
33	2022-11-3	Buy BBCA	2023-01-02	Sell BBCA	0.15%
34	2023-01-12	Buy BBRI	2023-09-08	Sell BBRI	13.13%
35	2023-09-12	Buy BBCA	2023-09-18	Sell BBCA	-0.28%
36	2023-09-25	Buy BBCA	2024-03-01	Sell BBCA	10.86%
37	2024-03-18	Buy BBRI	2024-08-02	Sell BBRI	10.58%

Table 4 presents the trading transactions generated by the LSTM-based deep arbitrage strategy over the same evaluation period. Similar to Table 3, each transaction reports the buy and sell dates, traded asset, and realized transaction-level return. Compared to the statistical arbitrage strategy, the deep arbitrage approach executes fewer trading transactions while exhibiting a lower proportion of loss-making trades. This pattern suggests that the LSTM model applies more selective entry and exit decisions, potentially filtering out weaker arbitrage signals and reducing exposure to unfavorable market movements[19], [20].

Tables 3 and 4 provide a comparative overview of transaction-level performance between the statistical arbitrage and LSTM-based deep arbitrage strategies[23], [24]. While the statistical arbitrage approach generates a larger number of trading transactions, it also records a higher incidence of loss-making trades. In contrast, the deep arbitrage strategy executes fewer but more selective trades, resulting in a lower proportion of negative returns at the transaction level.

This contrast indicates a fundamental difference in trading behavior between the two approaches. The statistical arbitrage strategy reacts more frequently to deviations in the spread, which increases trading opportunities but also exposes the portfolio to higher noise and short-term market fluctuations[5], [10]. Conversely, the LSTM-based strategy appears to prioritize signal quality over trading frequency, leading to fewer entries but improved trade-level outcomes. These findings suggest that differences in overall portfolio performance are driven not only by return magnitude, but also by trade selectivity and timing efficiency.

Discussion

The discussion extends beyond restating numerical results by interpreting the underlying mechanisms that drive the observed performance differences between statistical arbitrage and LSTM-based deep arbitrage strategies.

1. Trading Frequency and Return Efficiency
   Although the LSTM-based deep arbitrage strategy generates fewer trades compared to the traditional statistical arbitrage approach, it achieves substantially higher cumulative portfolio returns. This result suggests that the LSTM model does not merely increase trading activity, but rather improves trade selectivity[3]. By learning temporal dependencies and nonlinear patterns in the spread dynamics, the LSTM model is able to filter out low-quality mean-reversion signals and concentrate capital allocation on high-confidence opportunities. In contrast, the threshold-based statistical arbitrage strategy reacts mechanically to spread deviations, which may result in more frequent but less profitable trades, particularly during periods of elevated market noise.
2. Performance Across Volatility Regimes
   The two strategies also exhibit distinct performance characteristics across volatility regimes. Statistical arbitrage performs relatively well during stable, low-volatility periods when mean-reversion assumptions hold consistently. However, its performance deteriorates during high-volatility or regime-shift periods, where fixed thresholds may trigger premature entries or delayed exits. In contrast, the LSTM-based strategy demonstrates greater adaptability during volatile market conditions, as it dynamically adjusts to evolving spread behavior learned from historical sequences[25]. This adaptive capability enables the LSTM model to avoid unfavorable trades during extreme volatility, contributing to its higher risk-adjusted performance.
3. Sensitivity Analysis and Robustness
   To assess robustness, sensitivity analyses were conducted using alternative sample periods and varying threshold parameter configurations. Specifically, the evaluation window was shifted across subperiods representing pre-pandemic, pandemic, and post-pandemic market conditions, and threshold parameters were adjusted within reasonable ranges around their baseline values. Across these variations, the relative performance ranking between the two strategies remained unchanged, with deep arbitrage consistently outperforming statistical arbitrage[26]. While absolute portfolio return levels varied across configurations, the qualitative conclusions were stable, indicating that the results are not driven by a particular time window or parameter choice.
4. Interpretation of Sharpe Ratios
   While a Sharpe ratio above one is often cited as a practical benchmark for favorable risk-adjusted performance, its interpretation depends critically on the return distribution and evaluation horizon. In this study, Sharpe ratios are computed using daily returns over a multi-year out-of-sample period, which mitigates small-sample distortions[7]. The substantially higher Sharpe ratio achieved by the LSTM-based strategy therefore reflects not only higher average returns, but also improved volatility management. Conversely, the lower Sharpe ratio observed for the statistical arbitrage strategy indicates that its higher return volatility offsets its profitability, particularly during unstable market phases.
5. Implications and Limitations
   Overall, the findings indicate that deep arbitrage methods based on LSTM models offer superior efficiency in identifying profitable arbitrage opportunities under long-only constraints in emerging markets such as Indonesia. However, this performance advantage comes at the cost of increased model complexity and reduced interpretability. Moreover, the reliance on daily data limits the model's responsiveness to intraday dynamics, suggesting potential extensions using higher-frequency data. These considerations highlight the trade-off between model adaptability and simplicity, which should be carefully weighed in practical implementations.

4. CONCLUSION

This study empirically compared traditional cointegration-based statistical arbitrage and LSTM-based deep arbitrage strategies in the Indonesian stock market under realistic long-only trading constraints. Using daily closing price data from April 6, 2015 to April 6, 2025, the analysis was conducted within a strictly chronological evaluation framework to ensure a genuine out-of-sample assessment. All model estimation, parameter optimization, and LSTM training were performed exclusively on the in-sample period (April 6, 2015–August 31, 2021), while portfolio performance was evaluated solely on the out-of-sample period (September 1, 2021–April 6, 2025). This design eliminates look-ahead bias and ensures that all reported results reflect forward-looking trading performance. The empirical results demonstrate that both statistical arbitrage and LSTM-based deep arbitrage strategies are applicable in the Indonesian stock market context. However, the deep arbitrage strategy consistently outperforms the traditional approach in terms of cumulative portfolio return and risk-adjusted performance, as measured by the Sharpe ratio. Although the LSTM-based strategy executes fewer trades, it achieves higher profitability by exhibiting greater trade selectivity and improved timing efficiency. These findings indicate that superior performance is driven not by increased trading frequency, but by the ability to filter noisy signals and adapt to evolving market conditions.

From an applied mathematics perspective, this study highlights the benefits of integrating classical econometric techniques such as cointegration testing, volatility modeling via GARCH, and threshold-based trading rules with modern deep learning methods. The volatility-adjusted spread construction and rolling-window LSTM framework allow the deep arbitrage model to adapt across different market regimes, resulting in more stable out-of-sample performance. Sensitivity analyses further confirm that the relative performance advantage of deep arbitrage is robust across alternative sample periods and parameter configurations. Despite these contributions, several limitations should be acknowledged. The analysis focuses on a single highly correlated and cointegrated stock pair, which constrains the generalizability of the results. In addition, the use of daily data limits the model's ability to capture intraday dynamics. Future research may extend this framework by evaluating multiple stock pairs simultaneously, incorporating transaction cost heterogeneity, or applying the proposed methodology to higher-frequency data. Overall, this study provides evidence that LSTM-based deep arbitrage offers a more efficient and robust approach to exploiting mean-reversion opportunities in emerging markets under realistic trading constraints. By clearly defining a clean evaluation design and combining applied mathematical modeling with deep learning, this research contributes to the growing literature on quantitative arbitrage strategies in developing financial markets.

5. REFERENCES

[1] Y.-Y. Chen, W.-L. Chen, and S.-H. Huang, "Developing arbitrage strategy in high-frequency pairs trading with filterbank cnn algorithm," in *Proceedings - 2018 IEEE International Conference on Agents, ICA 2018*, 2018, pp. 113–116. doi: 10.1109/AGENTS.2018.8459920.

[2] Y. Saleh, M. Pradana, and B. R. Kartawinata, "Comparative Forecasting of Indonesian Blue-Chip Stocks Using ARIMA, ANN and LSTM Models," in *2025 IEEE 16th Control and System Graduate Research Colloquium, ICSGRC 2025 - Conference Proceedings*, 2025, pp. 239–244. doi: 10.1109/ICSGRC65918.2025.11159813.

[3] Y. Zhao, S. Xu, and J. Ossowski, "Deep Learning Meets Statistical Arbitrage: An Application of Long Short-Term Memory Networks to Algorithmic Trading," *J. Financ. Data Sci.*, vol. 4, no. 4, pp. 133–150, 2022, doi: 10.3905/jfds.2022.1.103.

[4] R. A. Andhika Viadinugroho and D. Rosadi, "Long Short-Term Memory Neural Network Model for Time Series Forecasting: Case Study of Forecasting IHSG during Covid-19 Outbreak," in *Journal of Physics: Conference Series*, 2021, vol. 1863, no. 1. doi: 10.1088/1742-6596/1863/1/012016.

[5] M. Diqi and I. W. Ordiyasa, "Enhancing Stock Price Prediction in the Indonesian Market: A Concave LSTM Approach with Runrelu," *J. Autom. Mob. Robot. Intell. Syst.*, vol. 18, no. 3, pp. 69–77, 2024, doi: 10.14313/jamris/3-2024/24.

[6] A. Flori and D. Regoli, "Revealing Pairs-trading opportunities with long short-term memory networks," *Eur. J. Oper. Res.*, vol. 295, no. 2, pp. 772–791, 2021, doi: 10.1016/j.ejor.2021.03.009.

[7] B. Zhan, S. Zhang, H. S. Du, and X. Yang, "Exploring Statistical Arbitrage Opportunities Using Machine Learning Strategy," *Comput. Econ.*, vol. 60, no. 3, pp. 861–882, 2022, doi: 10.1007/s10614-021-10169-8.

[8] V. Chang, X. Man, Q. Xu, and C.-H. Hsu, "Pairs trading on different portfolios based on machine learning," *Expert Syst.*, vol. 38, no. 3, 2021, doi: 10.1111/exsy.12649.

[9] D. Yoshikawa, "A new investment method for statistical arbitrage using eigendecomposition to generate stationarity," in *Proceedings - 2023 14th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2023*, 2023, pp. 663–666. doi: 10.1109/IIAI-AAI59060.2023.00130.

[10] K. Priya, P. Balaraman, and C. Sundar, "Machine Learning vs. Traditional Models: A Comparative Analysis of LSTM and ARIMA for Stock Prediction With Management Education Applications," in *Integrating AI and Machine Learning into Business and Management Education*, SRM Institute of Science and Technology, India: IGI Global, 2025, pp. 215–246. doi: 10.4018/979-8-3373-2150-9.ch009.

[11] H. Rad, R. K. Y. Low, and R. Faff, "The profitability of pairs trading strategies: distance, cointegration and copula methods," *Quant. Financ.*, vol. 16, no. 10, pp. 1541–1558, 2016, doi: 10.1080/14697688.2016.1164337.

[12] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020, doi: 10.1007/s10462-020-09838-1.

[13] B. Do and R. Faff, "Does simple pairs trading still work?," *Financ. Anal. J.*, vol. 66, no. 4, pp. 83–95, 2010, doi: 10.2469/faj.v66.n4.1.

[14] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: Performance of a relative-value arbitrage rule," *Rev. Financ. Stud.*, vol. 19, no. 3, pp. 797–827, 2006, doi: 10.1093/rfs/hhj020.

[15] R. Marti and A. Reimat, "Bitcoin and global stock portfolio: Dynamics across short and long-term horizons," 2025, doi: https://doi.org/10.1016/j.frl.2025.108918.

[16] H. Margaretha, P. M. Tauhalomoan, P. Widjaja, and F. V Ferdinand, "Predictive Analysis of Stock Prices Using ARIMA and Hybrid ARIMA-LSTM Models in the Indonesian Market," in *22nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2025*, 2025. doi: 10.1109/ECTI-CON64996.2025.11101223.

[17] H. Yirong and L. Yi, "Forecasting conditional volatility based on hybrid GARCH-type models with long memory, regime switching, leverage effect and heavy-tail: Further evidence from equity market," *North Am. J. Econ. Financ.*, vol. 72, 2024, doi: 102148.

[18] S. S. Wibisono, K. Leviana, T. S. Muliadiredja, H. Margaretha, F. V Ferdinand, and J. Utomo, "Stock Price Prediction on LQ45's Banking Sector Using Long-Short-Term Memory and Convolutional Neural Network," in *Proceedings - 2024 2nd International Conference on Technology Innovation and Its Applications, ICTIIA 2024*, 2024. doi: 10.1109/ICTIIA61827.2024.10761865.

[19] A. M. Hudzaifa, V. V Yang, and D. Y. Faidah, "THE IMPACT OF THE PRESIDENTIAL ELECTION ON IDX COMPOSITE PREDICTIONS USING LONG SHORT TERM MEMORY," *Barekeng*, vol. 18, no. 4, pp. 2397–2412, 2024, doi: 10.30598/barekengvol18iss4pp2397-2412.

[20] A. Hidayat and A. Sihabuddin, "Evaluating the impact of multiple leading indicator in forecasting next day stock price with lstm," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 2, pp. 293–303, 2021, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101374902&partnerID=40&md5=5b62bae8d6f776af84d6097217794969d

[21] Marverick, "What the Sharpe Ratio Means for Investors," 2025.

[22] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Syst. Appl.*, vol. 184, 2021, doi: 10.1016/j.eswa.2021.115537.
[23] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.
[24] A. A. Alalwan, N. P. Rana, Y. K. Dwivedi, and R. Algharabat, "Social media in marketing: A review and analysis of the existing literature," *Telemat. Informatics*, vol. 34, no. 7, pp. 1177–1190, 2017, doi: 10.1016/j.tele.2017.05.008.
[25] Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah, "Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 131–138, 2021, doi: 10.22146/jnteti.v10i2.1417.
[26] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.