Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

A Study of Trans-LSTM Based Arbitrage Strategy for Science and Technology Boards

Na Hu[1], Mingming Qu[2,a,*]

[1]College of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou, Gansu, 730101, China
[2]Yantai Aizhi Intelligent Technology Co., Yantai, Shandong, 264000, China
[a]vincent_aismart@163.com
[*]Corresponding author

Abstract: *In the present research, we develop a fusion model that is based on the positive combination of LSTM and Transformer models and is aimed at creating a deep learning-based statistical arbitrage strategy. The model was subjected to comparative analysis with both the LSTM model and the statistical cointegration approach, testing it through the Science and Technology Board (STB) dataset during both bull and bear market conditions. A complete analysis of the results shows that the new model with fusion trans_LSTM demonstrates better efficiency, enabling higher returns, and confirms that the overall stock market trends were confirmed during the backtesting of bull and bear market trainings. This quality demonstrates that the models' performance can be judged by the fact that the trans_LSTM fusion model shows effectiveness and reliability in stock market arbitrage. Thereby, it can be seen that this investigation contributes to solutions for further improvement of the role of deep learning in the financial market sphere.*

Keywords: *Arbitrage strategies; Deep learning; trans_LSTM fusion models*

1. Introduction

As China's financial market is progressing extensively, the exponential growth of financial data is not a luxury for an investor but a challenge posed by data heterogeneity, noise accumulation, and spurious correlations. Depending on the power of machine learning to solve investment problems in two spheres, the use of machine learning algorithms for the development of quantitative investment strategies is a salient issue across academic institutions and investment circles[1]. This background has a good platform for deep learning, a kind of machine learning for the purpose of developing advanced statistical arbitrage. The research becomes data-driven by sourcing trading data from the KSC market for the span of June 1, 2020, and June 1, 2023. The screening and normalization preprocessing operations are applied to the flow data. Subsequently, a trans_LSTM fusion model is developed to predict the spread data and establish a statistical arbitrage strategy. Following this, the strategy outcomes with traditional statistical methods are compared. Finally, relevant conclusions and recommendation are drawn.

2. Design of Trans_LSTM Fusion Models

The trans_LSTM fusion model represents a hybrid architecture, integrating two distinct neural network structures: the Transformer and the LSTM (long short-term memory). This innovative product suite links the attention mechanism of Transformer with the ability of LSTM to retain data for long periods, capturing several dependency relations within the sequence.

In addition, it gets a sense of relationships between positions by utilizing the self-attention mechanism, which allows the transformer to learn from different parts of the input sequence, ignoring the rule of sequential step-by-step computation. Typically, it utilizes multiple attention heads, with every head assigned to discover different sequence contingencies. This way, the model not only has a better grasp of the input sequence but also has a wider horizon for what follows next. The further we go in attention layers, the more a full-connected feed-forward neural network comes to aid with non-linear relationship recognition. To the contrary, the transformer is defined by the fact that it cannot automatically use the positional data that is present in the input sequence. For the mechanism to pivot, this is achieved through the incorporation of positional encoding. It nevertheless cannot give a satisfactory answer to that question because of the trade-off between the model information and positional information when generalizing to a more complex scenario like the situation in which the relative positions of objects are likely to be smaller after some transformation in the self-attention module.

On the other hand, there is LSTM, a more sophisticated development of RNN that was built to remedy the problem of vanishing gradients. Cell state and gating mechanisms, such as forget gates, input gates, and output gates, are the strengths of LSTM in the way they can easily manage long-term dependencies in the learning experience[2]. The best architecture of the model allows it to efficiently store short-term dependencies for sequences; hence, it is most appropriate for accelerating the time-series data. The forget gate selects the amount of the previous state that is going to be forgotten, the input gate defines the extent of the new content to be entered, and the output gate tolerates the remaining, which is used to deliver the results. The transposition of the layered architecture enables LSTM to discover these various sequences, from local short-term fluctuations to long-term trends, and delivers the velocity that overcomes the sharp rise and fall of gradient values.

Conclusively, the success of LSTM in the field of sequence modeling demonstrates that it is a key component in the various fields of time-series data analysis, natural language processing, and so many other domains. This can be generally said to be an essential precursor for the successful temporal element adaptation of deep learning tasks with time elements. Time-dependent tasks are very often disturbed because of both short-term and long-term temporalities. This idea is related to situations where, in the future, the data will, without doubt, be mostly based on the data that was collected during the past. Conversely, short-term dependence occurs when future data is primarily affected by more recent historical data. To address the Transformer's limitations in capturing intricate positional relationships, the input data is initially encoded using sine and cosine functions. These encoded features are then subjected to parallel feature extraction in the self-attention layer, followed by the utilization of LSTM for the preliminary extraction of time-series features. Subsequently, a fully-connected layer conducts a linear transformation on the features extracted by both LSTM and the self-attention layer, culminating in the output of the spread prediction sequence.

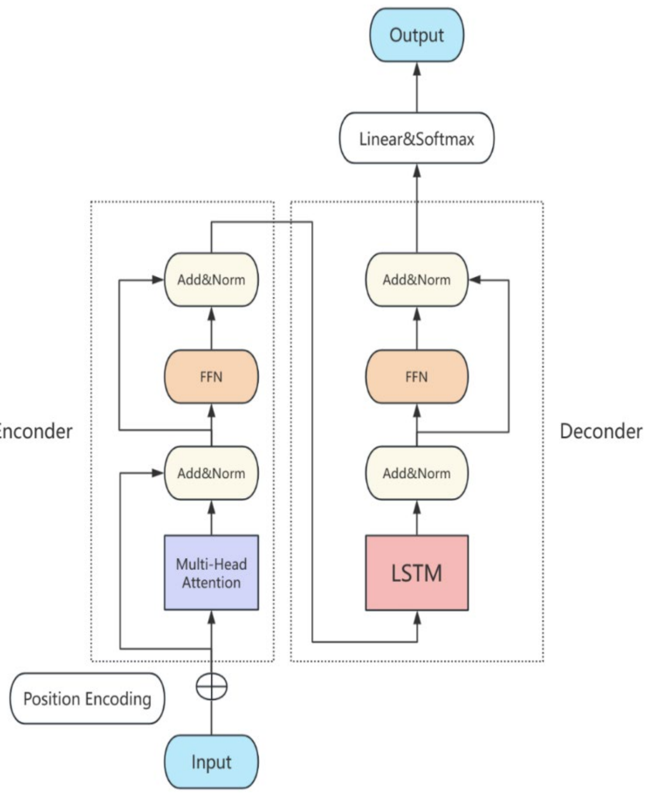

Figure 1: Trans_LSTM Fusion Model Structure

As depicted in Fig. 1, this study introduces the trans_LSTM fusion model, incorporating the Transformer's self-attention and multi-head attention mechanisms for the initial processing of input sequences. The model comprises encoding and decoding blocks, where each encoding layer includes

Multi-Head Attention (MHA), Fully Connected Networks (FFN), and Regularization Layers (Add & Norm). Each decoding layer, in turn, features two Multi-Head Attention layers for assimilating global information. Subsequently, an LSTM module is integrated into the decoder, enabling the transmission of this global information to the LSTM layer for enhanced processing and the capture of longer-term dependencies. This synergistic approach significantly boosts the model's capacity to effectively discern both global and local dependencies within sequence data.

The process is delineated as follows:

1) Input spread data $\boldsymbol{X_m}$ is segmented to emphasize local features within the time dimension, thereby generating a between-sequence matrix $\boldsymbol{X_n}$. This matrix is then processed through the Multi-Head Attention mechanism to derive local feature representations $\boldsymbol{I_m}$.

$$\boldsymbol{I_m} = MultiHeadAttention\left(\boldsymbol{X_n}\right) \tag{1}$$

2) Feature weights for different time series $\boldsymbol{S_m}$ are calculated independently, which then facilitate the computation of local self-attention scores $\boldsymbol{H}^{(m)}$.

$$\boldsymbol{S_m} = Softmax(ReLU\left(\left(\boldsymbol{W_q I_m}\right)^T \boldsymbol{W_k I_m}\right)) \tag{2}$$

$$\boldsymbol{H}^{(\mu)} = \sum_{m=1}^{n} \boldsymbol{S_m W_v I_m} \tag{3}$$

Where, $\boldsymbol{W_q}$, $\boldsymbol{W_k}$, $\boldsymbol{W_v}$ represents the weight matrix.

3) Feature vectors produced by the Transformer model are concatenated to form the feature vector representation $\boldsymbol{A}$.

4) This feature vector representation $\boldsymbol{A}$ is input into the LSTM model for further extraction of time-series information.

$$\boldsymbol{Y_m^t} = LSTM\left(\boldsymbol{A}, \boldsymbol{W_m}, \boldsymbol{Y_m^{t\text{-}1}}, \boldsymbol{\theta_m}\right) \tag{4}$$

Where, $\boldsymbol{Y_m^t}$ denotes the hidden state at time t; LSTM(·) is the LSTM model's function for extracting time-series features, $\boldsymbol{\theta_m}$ is the hyperparameters of the LSTM model, and $\boldsymbol{W_m}$ is the parameter matrix learned during the LSTM model's training.

3. Empirical Analysis of Statistical Arbitrage Strategy Based on Trans_LSTM

3.1 Data Collection and Cleansing

The dataset utilized in this study comprises trading data from the Science and Technology Innovation Board (STB) over three natural years, from June 1, 2020, to June 1, 2023. It includes 104 stocks, spanning 784 trading days, and encompassing a total of 81,536 data points. The data was sourced from the Cathay Pacific database. A statistical description of the market data is provided in Table 1.

Table 1: Description of the Market's Statistics

	Market Raw Data	Market Spread Data
Max	1495.58	1479.11
Minimum	2.34	-1489.80
Variance	11960.83	16936.97
Standard Deviation	109.37	130.14
Median	49.113	0.89

Prior to model input, the data undergoes a cleaning process. Most missing data instances correspond to periods of stock delisting or listings of less than one year. Their removal is deemed to have no significant impact on the model's learning capability. Consequently, for handling missing values and outliers, this paper adopts a strategy of direct deletion.

3.2 Data preprocessing

The deep learning model deployed in this study directly engages with the spread data. Consequently, the initial step involves constructing the spread data of stock series as follows:

$$Spread{=}Close\ price[symboll]{-}Close\ price[symbol2] \tag{5}$$

Where, $Close\ price$ represents the closing price of the day, and $symbol$ corresponds to the stock in question.

3.2.1 Data Standardization

Standardization not only enhances model performance but also significantly supports data interpretability and visualization. This, in turn, renders the application of machine learning in the financial domain more robust and effective.

$$spread = \frac{spread - \mu}{\sigma} \tag{6}$$

Where, the left-hand side $spread$ denotes the standardized data points, while the right-hand side $spread$ represents the original data points. μ and σ are the mean and variance of the data, respectively.

3.2.2 Data Conversion and Model Parameterization

1) Data Conversion

Considering the LSTM model operates within the realm of supervised learning, it necessitates the preprocessing and conversion of time series data into sample data and corresponding labeled data prior to training. The objective herein is nonlinear fitting of the samples, categorizing the problem as one of regression. An adaptive methodology was employed for segmenting the sample and labeled data. Specifically, data sequences from the $t\text{-}m$ to the t minute were designated as sample data, while the sequence data at the t+1 point were treated as the corresponding labeled data. This component gives advantages through the usage of historical data of t types length for training later, followed by the projection of t-1 data points. The main gist of this approach involves correlating present data sets to their previous values in order to forecast future datasets and thereby enhance the model's practicality in discovering nonlinear associations in sequence data. This methodology thus presents an effective and flexible approach to special cases of supervised learning through the homogenous subscription and addressing of sample and labeled data in a precise manner.

2) Model Parameter Setting

In the current work, this model is used, which is named the TRANS-LSTM Fusion Model. The two are layered proprietary LSTM neural network model hidden layers, and between successive layers is a dropout layer to mitigate overfitting risk. Therefore, a connecting layer is then added and runs through as the output layer. In this study, the model’s hyperparameters of the epochs (epochs = 200), look-back time step (look_back = 10), learning rate (learning_rate = 0.001), the number of hidden layer neurons (the first being 50, while the second is 10), as well as the dropout ratio for the dropout layer, are specified.

3.3 Arbitrage Strategy

Ten thousand yuan of first capital is assumed, which, under the condition of not taking into account commissions, spelling duties, and the and the daily limit of transactions, implies two assets that are unpaid: A and B. The minimum trading unit is set at 100 shares. For model initialization, 200 days of data are input, employing a stochastic trading strategy thereafter, with daily training sessions. To ascertain a long-term relationship between the two stocks, the average spread between them is typically utilized as a criterion for pair trading. Namely, trading actions are initiated when the stock spread diverges from this benchmark. In empirical analyses, for accurate evaluation of the stocks' high and low prices, it is often necessary to standardize the spread data, enabling a more rigorous analysis and

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

prediction. The standardization formula is:

$$M\_spread_t = spread_t - mean\left(spread_t\right) = y_t - \beta x_t - \alpha \tag{7}$$

If the sequence of spreads of two stocks M_spread_t, after undergoing decentralized processing, conforms to a normal distribution, with μ representing the mean and σ denoting the standard deviation of this distribution, the specific trading rule model is established as follows:

1) Opening a Position:

If at time $M\_spread_t > \mu + k\sigma(k > 0)$, Stock A is overvalued, execute a purchase of Stock B shares in the ratio of α: 1.

Conversely, if Stock A is undervalued at time $M\_spread_t < \mu - k\sigma(k > 0)$, buy Stock A shares at the same α:1ratio.

2) Closing the Position:

Positions are closed in the reverse direction of the opening operation under the condition that when M_spread_t, Stock A returns to the vicinity of the mean value μ.

3) Stop Loss:

When $M\_spread_t > \mu + \Gamma\sigma(\Gamma > k > 0)$, Stock A is deemed overvalued. Then close the position

Similarly, when $M\_spread_t < \mu - \Gamma\sigma(\Gamma > k > 0)$, if Stock B is considered overvalued, proceed with the forced closure of positions.

To mitigate the issue of having the majority of spreads fall within a certain range—thus triggering fewer signals—and to prevent the majority from falling outside this range, resulting in excessive trading signals and elevated transaction costs, the model sets the threshold to initiate the base signal $\mu \pm 1\sigma$, establish the close position signal at $\mu \pm 0.2\sigma$, and determine the stop-loss signal at $\mu \pm 2\sigma$.

3.4 Empirical Analysis

An extensive examination of the trading data from the Science and Technology Boards between June 1, 2020, and June 1, 2023, yielded trading data encompassing 784 trading days and 104 stocks. Using Matlab 2016, the spreads of 38,329 stock pair groups were calculated. These spread data were then input into the constructed trans_LSTM model. To explore various data sets and parameters, one pair of stocks was ultimately selected for further analysis and discussion. For comparative analysis, this study selects the LSTM model and the cointegration model, a prevalent model in statistical methods, against the trans_LSTM fusion model.

3.4.1 Bull Market

During the cointegration test, it was observed that the stock prices displayed unstable trends when compared to their original series. When the data underwent logarithmic transformation followed by first-order differencing, the differenced series achieved smoothness, successfully passing the Augmented Dickey-Fuller (ADF) test, thereby indicating first-order monointegration. Subsequent cointegration testing, employing the Engle-Granger two-step method, involved conducting a smoothness test on the residual series. All tested residuals met the criteria for smoothness; however, the outcomes from the error correction model did not exhibit adequate fitting, suggesting that the series, based on these results, should not be considered for cointegration-based statistical arbitrage. This conclusion limits the selection pool for investors, potentially elevating investment risk[3]. It is important to highlight that this inference is drawn from a detailed examination of both smoothness and cointegration within the stock price series. The results of the cointegration test are presented in Table 2.

However, the introduction of the trans_LSTM fusion model offers a novel perspective. By analyzing the spread series across all stock pairs, this model uncovers arbitrage opportunities previously unidentified through cointegration theory, evaluating the predictability and accuracy of spread data. This illustrates that, for instance, with traditional cointegration tests, they are not capable

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

of fully tracking the status of latent arbitrage potential in the market. In comparison, the structural array knowledge of this new union provides investors with a greater variety of investments and a more accurate prediction of arbitrage opportunities, which is expected to greatly influence the forthcoming investments.

Table 2: Cointegration Test

	688123-688202
ADF test	Non-stationary - Non-stationary
The smoothness of first-order difference series	Smooth-Smooth
Residual Smoothness	Smooth
Error correction model mse	6630.9121
Error correction model mae	74.2715
Error correction model r2	10.8164

The next step involves running forecasts for stocks, and then arbitrage strategies are enacted in both the bull and bear markets upon reading the predictions from the correlation models. The phrases "bull market" and "bear market" are not merely random lingo in finance but a crucial attribute, structuring the moves of the stock and financial markets' mood and direction. A bull market is depicted by an upward movement, which is marked by investors' hopeful attitude about near-future economic conditions, and the prices of investments advance along with other assets. During this time of positive economic performance, stock markets are based on good statistics, and investors are confident in the corporate earnings, which enable the stock markets. In a bull market, there is a tendency to think that it's safe to go to riskier investment strategies with high rates of return. This is a common reaction in a bull market. Addressing the contrarian view, it must be mentioned that a bear market refers to a negative trend where investors exhibit pessimism concerning the economic future. This is an environment where stock prices and other assets decline in value. Rather than being separate events, the bear market and recession may happen concurrently, coupled with economic indicators such as inflationary pressures, falling corporate earnings, and decreased confidence among investors. Apparently, conservative plans and potentially some hedging are popular actions in such periods when investors try to reduce market risks. When the wind is to buy and sell, it is the phases of bull and bear that appear in the stock market. Investors invest by purchasing and selling stock under various market conditions. On the other hand, they are instructed to gain insight on market behavior and price fluctuations of assets. They are crucial, indeed, for formulating investment strategies and for taking manageable risks. Thus, this work chooses to use up-and-down market conditions as a sampling model because it provides a convenient method for arbitrage strategy trading research. Data study and observation of the stock market lead to the conclusion that March 1, 2021–June 1, 2021 must be given a bull market period, and a bear market backtesting period must be June 1, 2022–September 1, 2022.

In the spread forecast chart, the red curve represents the spread forecast by the trans_LSTM fusion model, the green curve by the LSTM model, and the black curve depicts the actual spread. Meanwhile, in the net value curve forecast chart, the red curve illustrates the net value trajectory of the trans_LSTM fusion model, with the blue curve representing that of the LSTM model. The analysis involves Juchen (688123) and Meidixi (688202) stock pairs.

Figure 2: 688123-688202 Forecast and Actual Value

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

Figure 2 presents the comparison curve between the forecasted and actual spreads of the stock pairs Juchen (688123) and Meidixi (688202). The predictions from the trans_LSTM fusion model more closely align with the actual values compared to those from the LSTM model.

Table 3: 688123-688202 Forecast and Actual Value

	MAE	MSE	RMSE	R^2
trans_LSTM	0.0520	0.0042	0.0658	0.9367
LTSM	0.0907	0.0114	0.1069	0.8298

Table 3 delineates the comparison of performance metrics between the forecasted and actual values for the stock pairs Juchen (688123) and Meidixi (688202). From the perspective of Mean Absolute Error (MAE), the trans_LSTM fusion model, with an MAE of 0.0520, significantly outperforms the LSTM model's MAE of 0.0907. Similarly, regarding the Mean Squared Error (MSE), the trans_LSTM fusion model's MSE of 0.0042 is notably lower than the LSTM model's MSE of 0.0114. The Root Mean Squared Error (RMSE) for the trans_LSTM fusion model is 0.0658, substantially smaller than the LSTM model's RMSE of 0.1069. Additionally, from the R^2 metric perspective, the trans_LSTM fusion model, with an R^2 of 0.9367, significantly exceeds the LSTM model's R^2 of 0.8298. These results collectively indicate that the trans_LSTM fusion model surpasses the LSTM model in predicting the spread for the stock pairs of Juchen (688123) and Meidixi (688202).

Subsequently, arbitrage trading based on the forecasts of these two models will be executed in accordance with the arbitrage strategy.

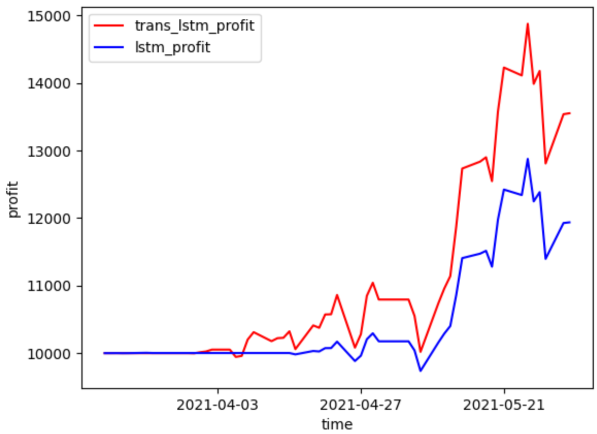

Figure 3: 688123-688202 Net Return Comparison

Figure 3 illustrates the net return comparison curve for the arbitrage strategy involving the stock pairs Juchen (688123) and Meidixi (688202), depicting an overall upward trend. The trans_LSTM fusion model presented extra-ordinary performance during the initial test. In spite of the kink during the ascending period of trend that looks like a temporary fluctuation or deviation due to the different spreading facts, the trend line still rose. This is a more positive sign that the strategy is capable of adjusting to the evolving market environment and has the agility to capitalize on current bullish stock market trends.

The following analysis will examine the trading processes and results with respect to the previous studied models based on the trans_LSTM fusion model and the LSTM model method, while taking into account the internal invariant investment strategy.

Table 4: 688123-688202 Backtesting Trading Results

	trans_LSTM	LSTM
Backtest Time	2021.3.1-2021.6.1	
Initial Margin	10000	
Number of trades	48	40
Total Profit	13551.2090	11936.5170
Maximum Retracement	0.1390	0.1149
Annualized Return	2.7057	1.1447

From the trading backtesting results presented in Table 4, the maximum drawdown experienced using the trans_LSTM fusion model was 0.1390, compared to 0.1149 with the LSTM model. The trans_LSTM fusion model executed 48 trades, while the LSTM model executed 40 trades. The

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

annualized return generated by the trans_LSTM fusion model was 2.7057, significantly surpassing the LSTM model's return of 1.1447. Considering maximum drawdown and annualized returns, the trading strategy employing the trans_LSTM fusion model yielded higher returns than the LSTM model, aligning with the prevailing bull market conditions.

A comparative analysis of forecasted versus actual spread values reveals that the trans_LSTM fusion model's curve more closely follows the actual spread trend of the stock pairs, demonstrating closer alignment with real values. Evaluation metrics such as MAE, MSE, RMSE, and R^2, along with the trading backtesting outcomes—specifically annualized returns and maximum drawdown rates—underscore the superior performance and effectiveness of the trans_LSTM fusion model relative to the LSTM model, in both bull and bear market scenarios.

3.4.2 Bear Market

In the context of the bear market, stock pairs Zhuoyue New Energy (688196) and Ottaway (688516) were selected for comparative analysis. The results of the cointegration test are displayed in Table 5.

Table 5: Cointegration Test

	688196-688516
ADF test	Stationary - Non-stationary
The smoothness of first-order difference series	Smooth-Smooth
Residual Smoothness	Smooth
Error correction model mse	566.6223
Error correction model mae	20.1518
Error correction model r2	2.1551

The findings lead to the exclusion of the aforementioned stock pairs when considering statistical arbitrage using cointegration-based theory, resulting in a narrowed field of options for investors and an increased investment risk. This conclusion stems from a comprehensive examination of the smoothness and cointegration of stock price series.

For the stock pairs Zhuoyue New Energy (688196) and Ottaway (688516), Figure 4 illustrates the comparison curves between the forecasted and actual spreads. The curves indicate that predictions from the trans_LSTM fusion model are more closely aligned with the actual values than those from the LSTM model.

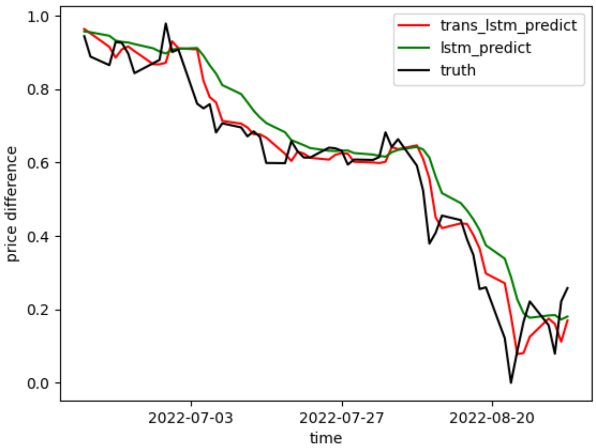

Figure 4: 688196-688516 Forecast and Actual Value

Table 6: 688196-688516 Forecast and Actual Value

	MAE	MSE	RMSE	R^2
trans_LSTM	0.0474	0.0043	0.0656	0.9335
LTSM	0.0704	0.0088	0.0941	0.8631

Table 6 illustrates the comparative analysis of performance metrics between the predicted and actual values for the stock pairs Zhuoyue New Energy (688196) and Ottaway (688516). The trans_LSTM fusion model exhibits a significantly lower Mean Absolute Error (MAE) of 0.0474 compared to the LSTM model's 0.0704. In terms of Mean Squared Error (MSE), the trans_LSTM fusion model's MSE is 0.0043, substantially lower than the LSTM's 0.0088. From the Root Mean Squared Error (RMSE) perspective, the trans_LSTM fusion model scores 0.0656, markedly below the

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

LSTM's 0.0941. Furthermore, the trans_LSTM fusion model achieves an R^2 value of 0.9335, significantly higher than the LSTM model's 0.8631. These numbers show the crossover model is better than the LSTM model in correctly forecasting the movement of a given pair of Zhuoyue New Energy (688196) and Ottaway (688516) stock share prices.

Correspondingly, arbitrage strategies will represent the culmination of the indication obtained from the modeling techniques.

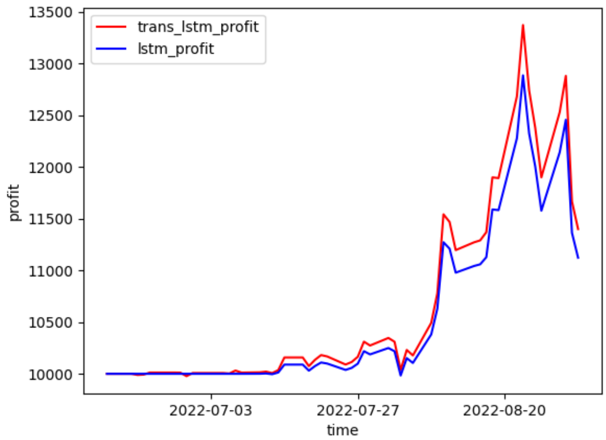

Figure 5: 688196-688516 Net Return Comparison

The net return curve on the chart is presented in Fig. 5 for the arbitrage trade that was performed between the stock pairs Zhuoyue New Energy (688196) and Ottaway (688516). There is a graph that heavily swings and fluctuates, demonstrating the final stock dividend returns that do not meet the suggested level.

Table 7: 688196-688516 Backtesting Trading Results

	trans_LSTM	LSTM
Backtest Time	2023.3.1-2023.6.1	
Initial Margin	10000	
Number of trades	47	38
Total Profit	11399.6630	11122.5840
Maximum Retracement	0.1475	0.1368
Annualized Return	0.7763	0.5946

After performing the trading backtest in Table 7, the trans_LSTM fused model showcases the largest drawdown of 0.1475, while the LSTM model has a slightly lower value of 0.1368 for the stock alternatives. By 47 trades, the trans_LSTM fusion model facilitated trade in contrast to the LSTM model's 38 neutral trades. Amongst the findings, the annualized return resulting from the fusion of the trans_LSTM model is 0.7763, compared with the LSTM model return of 0.5946. In this case, the LSTM trading strategy fusion model exhibited the ultimate performance with the highest annual return and maximum drawdown values as compared to the LSTM model.

A comparison between the predicted values of the trans_LSTM model and the actual spreads shows that the curve of the model takes closer alignment to the real spread trends of the stock pairs and brings the values more in accordance with real figures. The MSE, MAE, R^2, and RMSE measurements, in conjunction with the mentioned performance metrics like annualized returns and maximum drawdown reported from the backtesting evaluation metrics of the proposed TRANS-LSTM fusion model in this study, show improvement in the effectiveness and performance of the model compared to the conventional LSTM model. Moreover, the raw results displayed match the run-backtesting analysis executed within different market conditions in both bull and bear markets. Investors who invest in the bull market gain a boost in return rates; correspondingly, investors in the bear market draw low-scale returns when the stock market surges.

4. Conclusions

The paper included the use of the trans_LSTM fusion model in an attempt to elucidate the spreads

ISSN 2616-5775 Vol. 7, Issue 4: 62-71, DOI: 10.25236/AJCIS.2024.070409

of the Science and Technology Board data; it was revealed that the model was characterized with a higher accuracy than the LSTM model itself when the datasets were compared to the actual spread values in the analysis. This model suggests the trend of the spread, and it closely corresponds to the spread values exactly as our predictions, showcasing similarities in their curve. Furthermore, the deployed trans_LSTM fusion approach to arbitrage reveals completely invisible arbitrage opportunities that traditional statistical methods and the cointegration theory as applied in numerous research studies cannot disclose. By applying the classical results of the LSTM versions against our trans_LSTM model, we can not only improve the performance of our model in both volatility and stability market conditions but also establish the position of the trans_LSTM model as an even more effective model. This means that the hypothesis to apply the deep learning model to statistics for statistical arbitrage has both academic substance and practical authenticity. We have developed a unified framework that will help to alleviate the statistical arbitrage challenges we face while giving out new research directions and ways of updating our deep learning methodologies for financial industry improvement. What's more, this trans_LSTM fusion model, which we propose, not only performs well in spreading prediction but also proves to be robust and general even across different scenarios, emphasizing the model's capability of adapting to various financial data sets and diverse market dynamics. This model is a very useful forecasting tool for traversing across spread movements in financial markets, discovering thereby the statistical arbitrage opportunities with a precisely located foundation. In the end, we apply the more sophisticated Transformer LSTM fusion model within the deep learning approach to the financial domain, which has demonstrated clearly improved performance in identifying and predicting spreads relative to other approaches. We hope that this fresh perspective on deep learning technologies in finance will be adopted not only by academics but also by practical research initiatives and yield useful results for addressing the statistical arbitrage problem in the market.

References

[1] Elliott R J, Hoekd V D, Malcolm W P. Pairs Trading [J]. Quantitative Finance, 2005, 3(5):271-276.
[2] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions [J]. European Journal of Operational Research, 2017, 270(2).
[3] Zhijing He, Sudan Jin. Realization of pair trading strategy in steel industry based on cointegration theory [J]. Investment and Entrepreneurship, 2020, 31 (23): 14-16.