Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 4, No. 1, pp. 44-58, 2024
Research Article
Received: 03/10/2023 Accepted: 13/05/2024

Dual-Class Stocks: Can They Serve as Effective Predictors?

Veli SAFAK

Abstract
This paper investigates the three stocks of Kardemir Karabuk Iron Steel Industry Trade & Co. Inc. (Kardemir), the 24th largest industrial company in Turkey, listed on the Borsa Istanbul under tickers KRDMA, KRDMB, and KRDMD. Despite sharing identical attributes except for voting power, these stocks have displayed notable price divergences over an extended period from January 2001 to July 2023. Through an extensive analysis, this paper identifies and quantifies the price divergence patterns, revealing a compelling arbitrage opportunity through pair trading with a maximum potential gain of 361.21%. Employing wavelet coherence analysis, this study documents a strong coherence among the stock prices for the majority of the analysis period. Additionally, it demonstrates that the use of a sliding-window approach in selecting the training set significantly improves predictive performance based on 2,408 long short-term memory (LSTM) models. Notably, the fixed-horizon approach is found to lead to a statistically significant underestimation of future prices. Finally, the empirical findings emphasize that even when predictors exhibit strong coherence with the target variable, they may adversely impact predictive performance.
Keywords: Dual-class stock; long short-term memory; stock price prediction; wavelet analysis

1. Introduction

The establishment of Kardemir, Turkey's inaugural integrated iron and steel factory, dates back to September 10, 1939, when it was initiated by İsmet İnönü, who served as Prime Minister during that time. This significant step was part of the broader national industrialization efforts championed by the republic's founder, Mustafa Kemal Atatürk. The main activity subject of the company is the production and sale of all kinds of crude iron and steel products, coke, and coke by-products. It was listed in the Borsa Istanbul on Jun 1, 1998, with 3 stocks: group A (ticker: KRDMA), group B (ticker: KRDMB), and group D (ticker: KRDMD). The Group A shareholders have the right to elect 4 members to the Board of Directors, the Group B shareholders have the right to elect 2 members to the Board of Directors, and the Group D shareholders have the right to elect 1 member to the Board of Directors. Apart from this voting privilege, there are no other privileges.

This stock structure with different voting privileges is known as dual-class stock structure. There is significant cross-country evidence suggesting that investors pay a premium for stocks with voting privileges. The pioneering empirical investigation in this domain was conducted by Lease, McConnell, and Mikkelson [37], who demonstrated that higher vote shares in the United States are associated with a premium of approximately 5%. Horner [28] examined dual-class stocks in Switzerland and observed a voting premium of merely about 1%. Zingales [26] identified a substantial premium of roughly 80% in Italy. Smith and Amoako-Adu [3] detected a premium of around 19% in Canada during the period 1988-1992, which closely resembles the premium documented in Sweden by Rydqvist [22] at 15%. Additionally, Megginson [48] provided evidence of a premium of around 13% in the United Kingdom.

Voting premium is not the only interesting phenomenon about the dual-class structures. There is also evidence suggesting that prices of dual-class stock also tend to exhibit high co-integration [1]. Since 2014, the relative price ratio of GOOG (without voting power) and GOOGL (with voting power) ranged between 0.9459 and 1.05. Wu [19] used the co-integration between GOOG and GOOGL and designed a pair trading strategy. Pair trading constitutes a market-neutral tactic centered on the selection of stock pairs grounded in their relative prices or alternative indicators. The primary objective is to pinpoint pairs that exhibit a substantial level of correlation or cointegration, indicative of their tendency for synchronized price movements. This strategy finds prevalent usage among hedge funds and can be further refined through the assimilation of supplementary insights, such as volatility, anti-persistence, or qualitative information derived from financial reports. Diverse methodologies, spanning statistical assessments, machine learning algorithms, and genetic programming, can be employed to unearth lucrative pairs and formulate trading cues. The efficacy of pair trading extends across various asset categories and market conditions, with certain investigations intimating heightened effectiveness in periods of market decline [18], [25], [9], [11], and [5].

The academic interest in dual-class stock structure stems from the curiosity of understanding the stock prices. Forecasting stock prices is a classic problem laying in the intersection of finance, computer science, and economics. Various methods have been developed and used to forecast future stock prices. Fundamental analysis based

*Corresponding author

Veli SAFAK*; Carnegie Mellon University, Qatar; e-mail: vsafak@andrew.cmu.edu;

00000-0001-9302-1879

on companies' financial statements and technical analysis based on various indicators formulated as functions of past price action are cornerstones in quantitative finance.

In parallel to reductions in computation cost and increase in the volume of accessible data, researchers have developed more sophisticated methods to forecast price action in foreign exchange and stock markets. In recent years, deep learning algorithms have gained attraction among researchers. There are four major types of deep learning algorithms: convolutional neural network, deep neural network, recurrent neural network, and long short-term memory. While this paper exclusively employs LSTM models, it provides a brief overview of the other methods. Additionally, it presents a list of studies utilizing methods not emphasized in this paper for reference. For a more in-depth literature review, readers may refer to [47] and [54].

Convolutional neural networks (CNNs) are a class of deep neural networks specifically designed for processing and analyzing visual data, such as images and videos. CNNs are characterized by their ability to automatically learn hierarchical representations of features from input data. The key innovation of CNNs lies in the use of convolutional layers, which apply convolutional operations to input data. These operations involve small, learnable filters that scan the input in a systematic way, capturing local patterns and spatial relationships. This enables the network to recognize low-level features, such as edges and textures, and progressively build more abstract and complex representations through subsequent layers. Typically, CNN architectures consist of convolutional layers followed by pooling layers, which downsample the spatial dimensions of the data, reducing computational complexity. Fully connected layers are then employed to make predictions or classifications based on the learned features. The strength of CNNs lies in their ability to automatically extract relevant features from raw input data, making them highly effective in tasks such as image classification, object detection, and image segmentation. The hierarchical feature learning in CNNs mimics the human visual system, contributing to their success in various computer vision applications.

Numerous studies have explored CNNs in stock market prediction and explored various aspects, including model comparisons, graph theory integration, technical indicator application, multi-indicator feature selection, ensemble models, event-driven prediction, and unique architectural approaches. These studies contribute to advancing the utilization of deep learning techniques for enhanced stock market forecasting [10], [12], [13], [14], [16], [20], [29], [30], [34], [35], [40], [42], [44], [49], [51], and [53].

Deep neural networks (DNNs) represent a class of artificial neural networks characterized by their depth, involving multiple layers of interconnected nodes or neurons. These networks are designed to automatically learn hierarchical representations of features from input data, allowing them to capture intricate patterns and relationships. The architecture of DNNs typically consists of an input layer, one or more hidden layers, and an output layer. Each layer contains nodes that process information and pass it to subsequent layers, with weighted connections determining the strength of these interactions. The depth of DNNs facilitates the extraction of complex and abstract features from raw input, enabling them to effectively model intricate relationships in data. Training DNNs involves adjusting the weights of connections through backpropagation, where the network learns by minimizing the difference between predicted and actual outputs. This iterative learning process enhances the network's ability to generalize and make accurate predictions on new data. DNNs have demonstrated considerable success in various domains, including image and speech recognition, natural language processing, and reinforcement learning. Their capacity to automatically learn hierarchical representations contributes to their effectiveness in capturing intricate patterns and solving complex tasks, making them a prominent tool in machine learning research.

Various studies have investigated the application of deep neural networks (DNNs) in stock market prediction. One study utilized a DNN model with novel input features and a plunge filtering technique, demonstrating notable profitability [52]. Another proposed a DNN model using the Boruta feature selection technique, outperforming some other machine learning models [32]. Additionally, a study employed boosted approaches in a DNN model to predict stock market crises, highlighting their relevance in price prediction [43]. Another research revealed the superiority of DNNs over shallow neural networks and representative machine learning models [27]. Lastly, a study on a deep factor model suggested a nonlinear relationship between stock returns and factors, outperforming linear models and other machine learning methods [21].

Recurrent neural networks (RNNs) constitute a category of artificial neural networks specifically designed to process sequential data by incorporating temporal dependencies. Unlike traditional feedforward neural networks, RNNs possess internal memory mechanisms, allowing them to retain information about previous inputs and use it to influence subsequent predictions. The architecture of RNNs includes recurrent connections that form loops, enabling information to persist within the network over time. This inherent memory capacity makes RNNs well-suited for tasks involving sequential patterns, such as natural language processing, time series prediction, and speech recognition. However, traditional RNNs suffer from challenges like the vanishing gradient problem, which hinders their ability to effectively capture long-range dependencies in sequential data. To address this limitation, variants like long short-term memory (LSTM) networks and gated recurrent unit (GRU) networks have been developed. These architectures incorporate specialized memory cells and gating mechanisms, allowing for improved information retention and flow through the network. The success of RNNs lies in their ability to model and comprehend sequential dependencies, making them valuable tools in diverse applications. Despite their effectiveness, ongoing research aims to further enhance their capabilities and address remaining challenges to advance the field of sequential data analysis.

Several studies have explored RNNs for financial prediction. One introduced a CRNN model combining convolutional neural network and recursive neural network and demonstrated its outperformance relative to LSTM and CNN models in forecasting Forex pairs' prices [24]. Another proposed a multi-task RNN model with Markov Random Fields (MRF), employing a multi-multilayer perceptron (MMLP) for feature extraction without reliance on technical indicators [6]. A study presented an RNN-Boost model incorporating technical indicators, sentiment features, and Latent Dirichlet allocation (LDA) features, demonstrating superior performance over a single-RNN model [46]. In a different approach, a Deep and Wide Neural Network (DWNN) model integrated CNN's convolution layer into the RNN's hidden state transfer process, achieving a 30% reduction in prediction mean squared error compared to a general RNN model [38]. Additionally, an Attention-based RNN (ARNN) with wavelet denoised input was proposed, combining autoregressive integrated moving average (ARIMA) and a RNN model output for enhanced forecasting [55].

Overall, this paper presents results of three primary analyses. It begins by examining the historical voting premium within Kardemir stocks. This analysis represents the first documentation of enduring discrepancies among Kardemir stocks. The identification of such consistent differences provides traders with the potential to strategically employ pair trading techniques in a profitable manner. The second analysis examines the coherence between their daily returns through wavelet coherence analysis. While Kardemir stocks generally exhibit strong coherence, there were prolonged instances where the coherence among them weakened. Finally, a comparative analysis of LSTM models with different specifications is provided. The variations in specifications arise from three sources: (i) the decision to use or not use dual stock prices as predictors, (ii) the number of lags employed as predictors, and (iii) the selection of the training set. This final analysis does not intend to compare LSTM models with various other machine learning models. Instead, it focuses on addressing two key questions:
Question 1: Can past dual-class stock prices effectively forecast future prices of each other?
Question 2: Does the use of longer price lags enhance forecast performance in LSTM models?

2. Methodology

2.1. Data and Variables

The data used in this study is based on daily high and low prices (in Turkish liras) for Kardemir stocks (tickers: KRDMA, KRDMB, KRDMD) between Jan 2001 and July 2023. The sole data source is Bloomberg. To capture the most likely intraday price, this study uses daily mid-prices. The daily mid-prices are calculated as follows:

$$HL = 0.5(high + low) \tag{1}$$

The daily premium of Group i over Group j is calculated as the percentage difference between daily mid-prices for Group i and daily mid-prices for Group j.

$$v_{ij} = HL_i/\mathrm{HL}_j - 1 \tag{2}$$

2.2. Wavelet Coherence Analysis

This paper uses the continuous wavelet transform (CWT) to quantify the magnitude, direction, and lead-lag effects between Kardemir stocks. This approach has a number of advantages. First, it uncovers the dynamic relationship between these stocks, allowing me to distinguish between periods at which prices are linked. Secondly, using the CWT, it is possible to identify changes in the direction of the relationship over time. Finally, the CWT provides insights about the relationship between these stocks at different time horizons simultaneously.

According to Torrence and Campo [7], the wavelet coefficients $W_{\varepsilon,\tau}$ associated with a time series $f(t)$ are calculated as:

$$W_{\varepsilon,\tau} = \sum_{t=1}^{n} f(t)\,\psi^*\left[\frac{t-\tau}{\varepsilon}\right] \tag{3}$$

where * represent the complex conjugate, $\varepsilon > 0$ is the scale associated with the wavelet and $\tau \in [-\alpha, \alpha]$ is the window location and $1/\varepsilon$ is the normalization factor. Here, the Morlet wavelet with wave number $\omega_0 = 6$ is used following Grinsted et al. [2]. More specifically, the Morlet wavelet is formulated as:

$$\psi(t) = \pi^{0.25} e^{i\omega_0 t} e^{\frac{-t^2}{2}}. \tag{4}$$

The cross-wavelet power spectrum is calculated as the product of two wavelet coefficients and represents the common variation between two time series over time and scale. It is formulated as:

$$W_{\varepsilon,\tau}(f,g) = W_{\varepsilon,\tau}(f) W_{\varepsilon,\tau}^{*}(g). \tag{5}$$

Like the correlation, the wavelet squared coherency is defined by normalizing the smoothed cross-wavelet power spectrum by the smoothed wavelet power spectrum associated with the individual time series:

$$\rho_{\varepsilon,\tau}^2 = \frac{\left| Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right|^2}{\left| Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f)\right)\right|^2 \left| Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(g)\right)\right|^2} \tag{6}$$

where Q is the smoothing operator. By construction, $\rho_{\varepsilon,\tau}^2$ takes values between 0 and 1. It implies no comovement when $\rho_{\varepsilon,\tau}^2 = 0$, and perfect comovement when $\rho_{\varepsilon,\tau}^2 = 1$. To identify statistically significant squared coherency regions, the study uses a Monte-Carlo method with 1,000 iterations.

To uncover lead-lag effects, the following wavelet multi-scale phase is used:

$$\theta_{\varepsilon,\tau}(f,g) = tan^{-1}\left(\frac{\mathcal{J}\left(Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right)}{\mathcal{R}\left(Q\left(\varepsilon^{-1} W_{\varepsilon,\tau}(f,g)\right)\right)}\right). \tag{7}$$

Here, $\mathcal{J}$ and $\mathcal{R}$ represent the imaginary and real components of the wavelet coefficients. Phase arrows are utilized within wavelet coherence plots to depict the direction of simultaneous movement and the effects of leading or lagging. Arrows pointing east (west) signify being in (out of) sync, while arrows pointing north (south) indicate that one time series leads (lags) the other. When the phase arrow points in a northeast (southeast) direction, it means that the two series are in sync, but the second one (or first one) leads the first one (or second one). Differing outcomes are conveyed by arrows facing northwest and southwest.

2.3. Long Short-Term Memory (LSTM)

While training a recurrent neural network, each iteration receives an update proportional to the partial derivative of the error function with respect to its current weight. When the gradient is vanishingly small, the training may slow and, in some cases, stops [39]. The long short-term memory technique [41] is developed as a potential solution for the vanishing gradient problem. The LSTM approach is widely used in predicting stock prices because of its capacity to recognize patterns and generate more accurate predictions compared to other methods [36], [17], [33], [23], and [31].

An LSTM unit consists of a cell, and within this cell, there are three gates that manage the movement of information and regulate the cell state. These gates include an input gate, an output gate, and a forget gate. The LSTM units are then interconnected, forming a chain where each individual cell acts as a memory module within the LSTM architecture. Figure 1 illustrates a standard LSTM architecture and Figure 2 shows a standard LSTM cell architecture.

In Figure 2, f_t, i_t, and o_t respectively represent the forget gate, input gate, and output gate. Also, X_t is the input, h_t is the output, C_t is the cell state, and $\hat{C}_t$ is the internal cell state. Based on the input, previous output, and previous cell state ($X_t, h_{t-1},$ and C_{t-1}); f_t, i_t, o_t, $\hat{C}_t$, C_t, and h_t are calculated as follow:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, X_t] + b_f\right) \tag{8}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{9}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \tag{10}$$
$$\hat{C}_t = tanh(W_C \cdot [h_{t-1}, X_t] + b_C) \tag{11}$$
$$C_t = i_t \cdot \hat{C}_t + f_t \cdot C_{t-1} \tag{12}$$
$$h_t = o_t \times tanh(C_t) \tag{13}$$

Here, σ represents the sigmoid function and $tanh$ represents the hyperbolic tangent function.

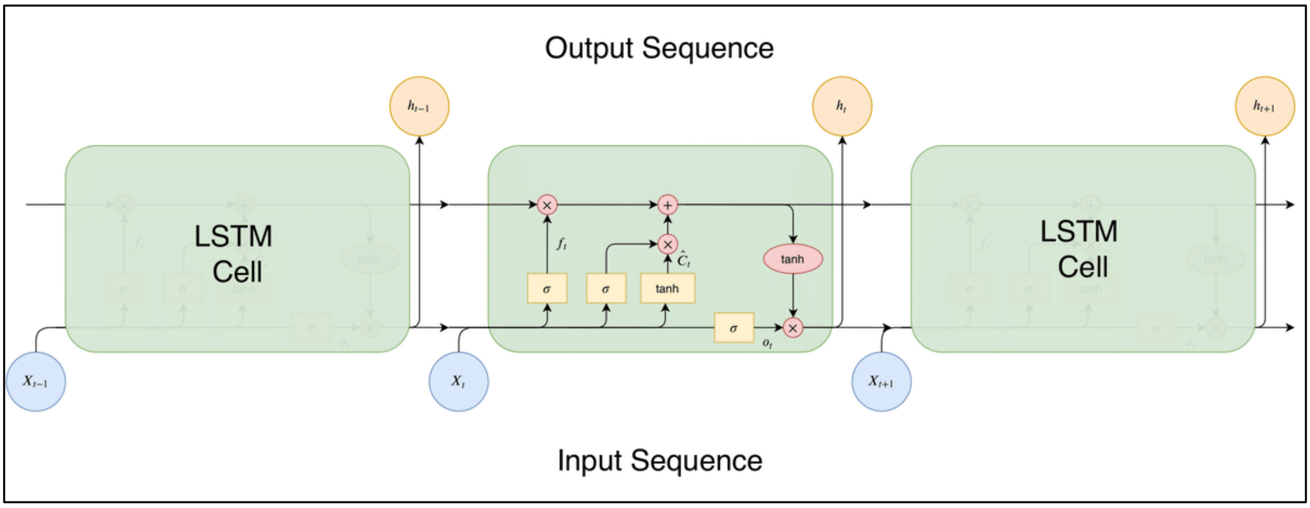

Figure 1: LSTM Architecture (source: [45])

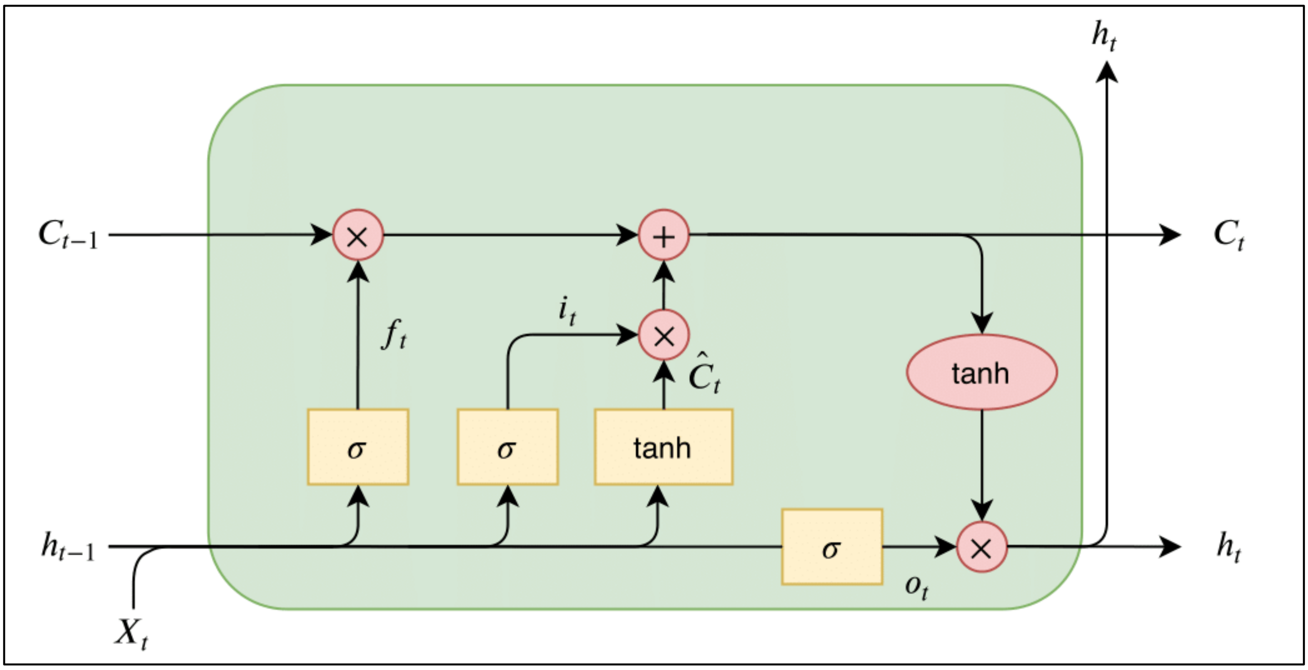

Figure 2: LSTM Cell Architecture (source: [45])

An LSTM model, operating as a black-box method, has the potential to exhibit overfitting issues, diminishing its effectiveness when applied to new, untrained data. To gauge the extent of overfitting, data is typically divided into two subsets in the realm of machine learning: the training set and the validation set. The training set is utilized for model development, while the validation set remains untouched during the training phase and serves to evaluate the model's predictive performance.

Traditionally, practitioners have favored training their models on large datasets with numerous observations, a practice grounded in the law of large numbers. However, this study posits that this conventional approach may not be well-suited for forecasting financial variables. In essence, it argues that a model trained on the daily prices of an asset spanning long periods (several lagged prices) may not perform as effectively as a model trained solely on the more recent daily prices.

One potential reason for this discrepancy lies in the fact that a long-range training set encompasses both downward and upward market trends. Such training data may not yield accurate predictions when applied to data sampled during a trend in a single direction. Consequently, utilizing more recent data points as the training set may lead to superior predictive performance.

In this paper, the use of rolling training sets is proposed for predicting the next observation. In this scenario, the 5,300th, 5,301st, 5,302nd, 5,303rd, and 5,304th observations serve as the training set to forecast the 5,305th observation when the training window is set to 5. This approach ensures that every observation is predicted based on the most recent price action, rather than relying on price action from hundreds of days ago.

To test the effectiveness of this new approach, 2 training set rules are used:

Approach 1 (Fixed horizon): The entire dataset is divided into two mutually exclusive and collectively exhaustive sets, namely a training set and a test set. The first 5,282 observations are used as the training set and the remaining 300 observations are used as the test set. In this case, all observations among the last 300 observations are used for forecasting based on a single model developed by using the first 5,282 observations.

Approach 2 (Sliding window): For every observation among the last 300 observations, the prior 5, 10, 20, and 50 observations are used as the training set. In this approach, a model is trained to forecast the next observation for each training window.

In total, 2,408 LSTM models with configurations above are trained to forecast the last 300 observations in the sample. As a preprocessing step, a transformation on the daily mid-prices by subtracting 100 from each value and then scaling the results by a factor of 1/100 is performed, resulting in the formula x/100-1. This scaling operation has the effect of confining all observations within the range of -1 to 1, consistent with the range of tanh function used in LSTM models. It's important to note that this scaling choice is based on the assumption that the daily mid-prices will consistently remain below 100 Turkish Lira (TRY). The selection of this threshold is based on the observation that all data points within the fixed-range training set are significantly lower than the chosen threshold. Consequently, the study has intentionally refrained from constraining the model to only produce forecasts that surpass the maximum value observed in the training set.

Finally, a deliberate decision was made to avoid using the conventional min-max scaling method. This choice was driven by the understanding that min-max scaling assumes prior knowledge of the range of daily mid-prices in the test set. However, in this context, the range of these mid-prices is considered unknown since they are the very values we aim to forecast.

3. Results

3.1. Historical Premiums

Figure 3 shows that KRDMA was traded at a premium relative to KRDMB for 1,171 days out of a total of 5,582 days. In 2020, the premium of KRDMA over KRDMB was the strongest when KRDMA predominantly traded at a premium for most of that year. Conversely, KRDMD consistently saw substantial discounts relative to both KRDMA and KRDMB. For 4,360 days, KRDMA was traded at a premium. Similarly, KRDMB was traded at a premium status for 4,318 days. Figure 1 also demonstrates a substantial decrease in premiums paid for KRDMA and KRDMB over KRDMD, starting from 2018. A reversion occurred starting in 2021 for KRDMD discounts. Since 2021, both KRDMA and KRDMB have consistently been traded at a discount relative to KRDMD.

Summary statistics of these premiums are given in Table 1. It clearly shows that substantial premiums paid for KRDMA and KRDMB over KRDMD between 2001 and 2023. At their heights these premiums reached 235.87% and 361.21% respectively for KRMDA and KRDMB.

Table 1. Summary Statistics of Premiums

	KRDMA over KRDMB	*KRDMA over KRDMD*	*KRDMB over KRDMD*
Minimum	-59.70%	-33.08%	-33.21%
1st Quartile	-19.10%	5.99%	12.98%
Median	-7.71%	35.47%	57.82%
Mean	-10.30%	49.44%	75.73%
3rd Quartile	0.00%	73.51%	128.13%
Maximum	57.06%	235.87%	361.21%

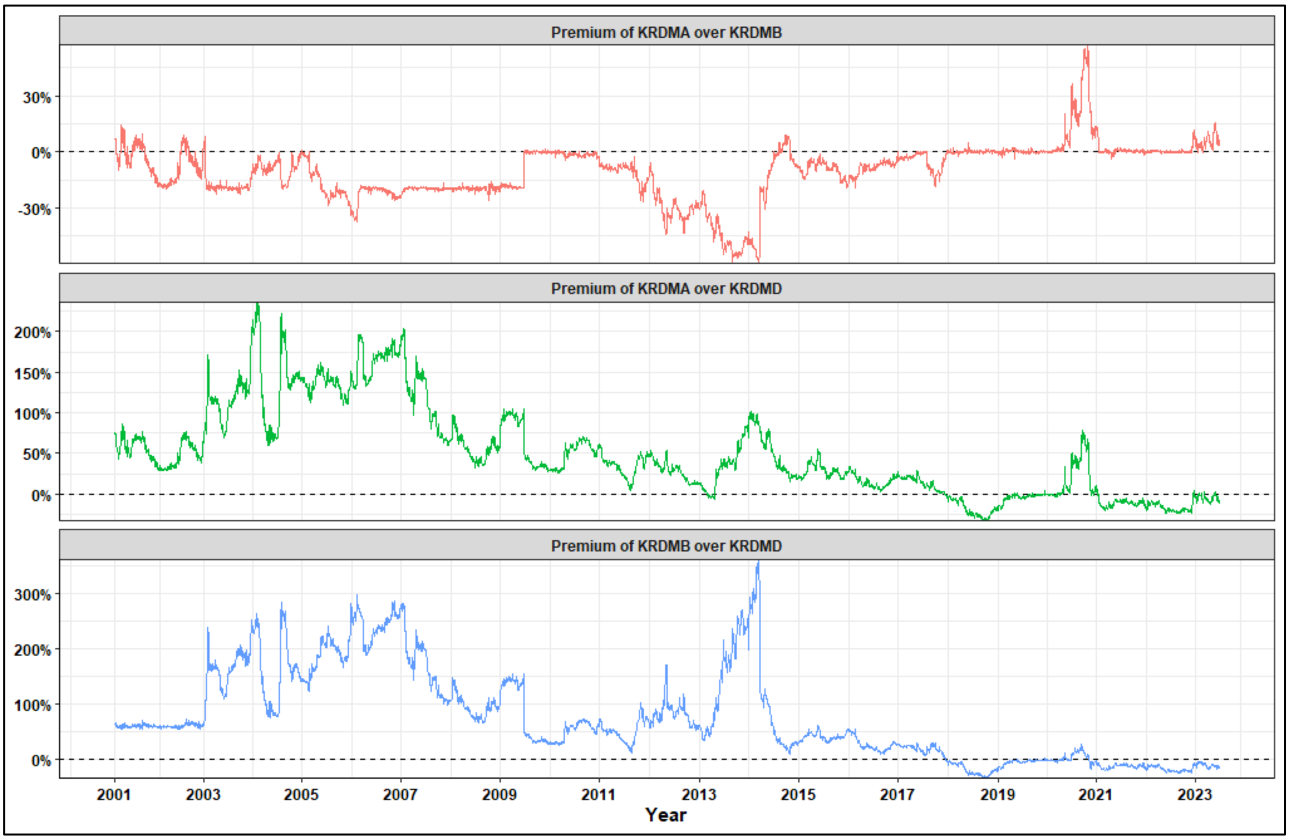

Figure 3: Historical Premiums

3.2. Wavelet Coherence Analysis

This section presents the dynamic relationship between daily returns (calculated as percentage change in daily mid-prices) of Kardemir stocks by using the wavelet coherence technique explained above.

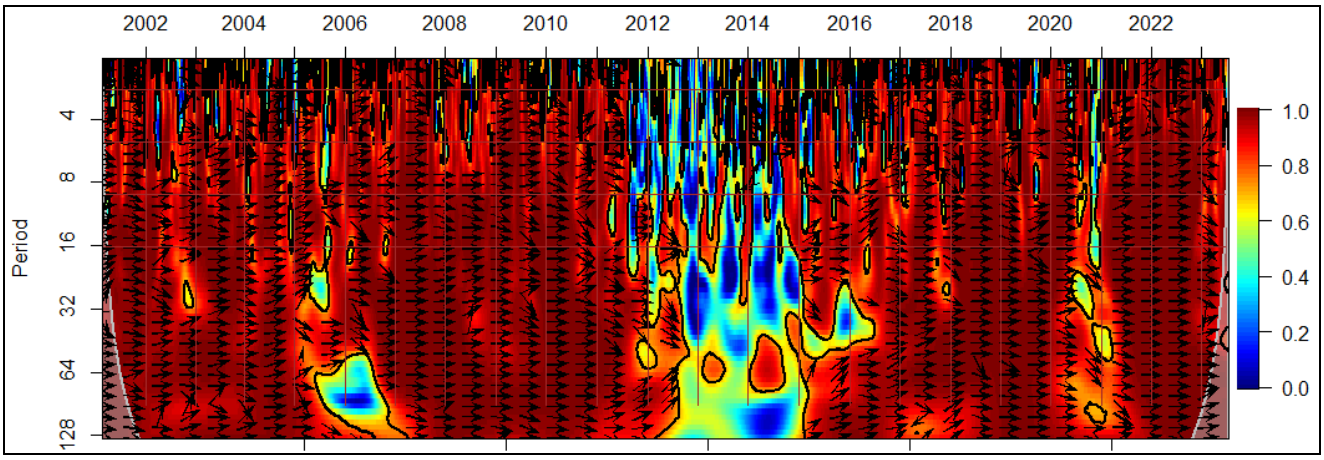

Figure 4: Wavelet Coherence between KRDMA and KRDMB Daily Returns

As depicted in Figure 4, the coherence between the daily returns of KRDMA and KRDMB remained consistently strong throughout the analyzed period, with only a few exceptions. The first notable deviation occurred in 2006, spanning periods 64 to 128, during which there was a clear lack of coherence between the daily return patterns of the two stocks. The second significant divergence surfaced in 2012 and persisted for over two years. During this period, the daily returns of KRDMA and KRDMB exhibited noticeable discrepancies. It's worth highlighting that over this time frame, the premium of KRDMA over KRDMB reached its lowest point, showing a substantial decline of -59.64%. Importantly, the figure also emphasizes the absence of a substantial cause-and-effect relationship between these two series. Instead, their temporal progression displayed synchronized movements, without any prominent identifiable temporal precedence or lag.

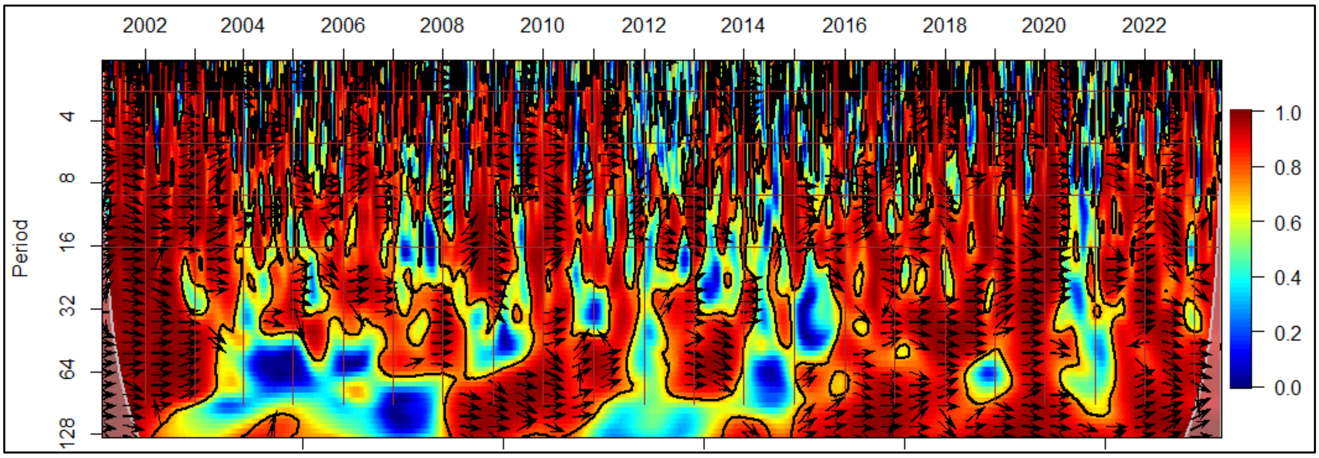

Figure 5: Wavelet Coherence between KRDMA and KRDMD Daily Returns

Figure 5 illustrates the results of the wavelet coherence analysis applied to the relationship between KRDMA and KRDMD. Across most of the time span, from January 2001 to July 2023, the daily returns of these two stocks displayed significant synchronization. It's worth noting that during this period, the coherence between the two stocks was notably strong and consistent, contributing to their aligned behavior.

Valuable insights can be derived from the coherence patterns at various time scales. Specifically, between 2003 and 2008, as well as intermittently in the first quarter of 2012 and throughout 2014, the coherence associated with longer cycles, particularly those spanning from 64 to 128 days, was relatively weak when compared to the coherence observed within shorter cycles, ranging from 8 to 16 days. This observation underscores temporal variations in the degree of synchronization across different scales, highlighting periods of heightened and diminished shared behavior.

Moreover, a clear absence of coherence becomes evident in the latter part of 2020 across various time cycles. During this specific period, the synchronization between the two stocks was notably absent. What's particularly noteworthy is that this timeframe coincided with a substantial premium of over 75% attributed to KRDMA over KRDMD. This convergence of factors highlights the potential interplay between coherence patterns and premium fluctuations, implying intricate dynamics at play in the relationship between these stock returns during this period.

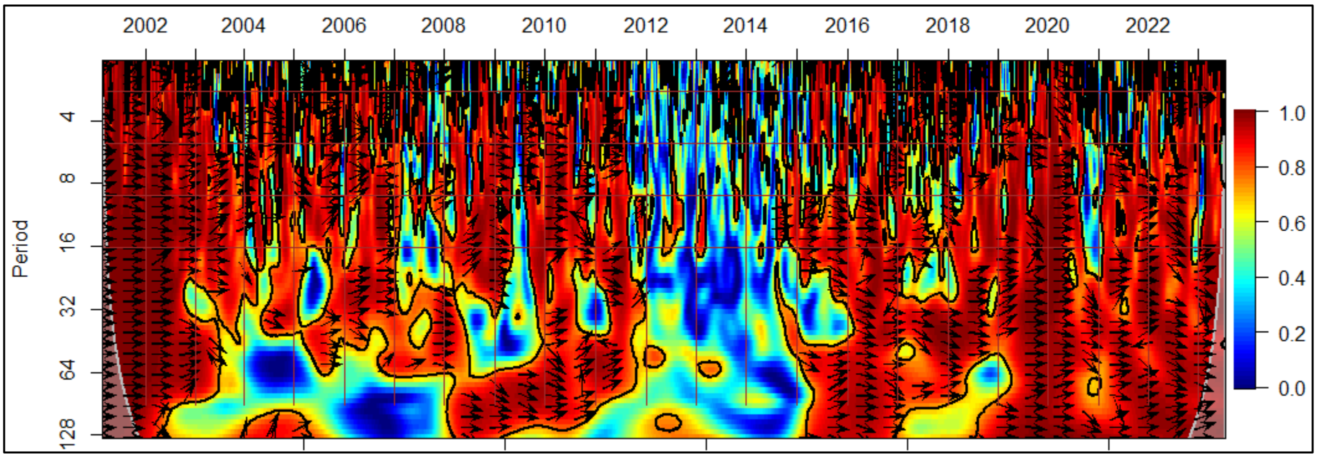

Figure 6: Wavelet Coherence between KRDMB and KRDMD Daily Returns

The final wavelet coherence plot, depicted as Figure 6, reveals the weakest coherence observed among Kardemir stock returns, specifically between KRDMB and KRDMD. Notably, the coherence between the daily returns of KRDMB and KRDMD was particularly weak, primarily spanning the years from 2012 to 2015. This period coincided with a time when the premium paid for KRDMB over KRDMD reached its peak, skyrocketing to an unprecedented level of 361.21%.

Across the broader time frame spanning from January 2001 to 2012, a robust coherence pattern was evident among the daily returns of the three Kardemir stocks across various time cycles. This robust coherence paradigm underwent a transformation, transitioning to a less robust coherence configuration from 2012 to 2015, only to reemerge in 2016. Another episode of coherence weakness emerged towards the latter part of 2020, encompassing all Kardemir stocks within shorter cycles. Typically, the highest coherence was observed between KRDMA and KRDMB.

Furthermore, the wavelet plots emphasize the absence of a clearly discernible leading or lagging relationship between various Kardemir stocks. Notably, instances of significant divergence in the daily mid-prices of Kardemir stocks coincided with the absence of coherence, particularly within longer cycles. This observation suggests a potential complex interplay between coherence dynamics and price disparities in the long-term context**.**

3.3. Long Short-Term Memory (LSTM) Models

This section presents the predictive performance results obtained from a total of 2,408 LSTM models. These models were trained with the specific goal of forecasting the daily mid-prices of the last 300 observations. To evaluate the effectiveness of dual-class stocks as predictors for each other, two sets of models are employed.

In the first set of models, the prediction models did not incorporate the historical price movements of dual stocks as lagged variables when forecasting future mid-prices. In contrast, the second set of models was designed to include lagged dual-class stock mid-prices as factors for predicting future mid-prices. For these lagged variables, two options are considered: using 4 lags and 9 lags of daily mid-prices. All these models were trained following one of the training-set rules outlined above. In summary, there are three significant distinctions among these specifications:

Criterion 1: Inclusion of dual-class stock prices as predictors (Possible values: Yes or No).

Criterion 2: The choice of the number of lags of previous stock prices as predictors (Possible values: 4 or 9).

Criterion 3: Selection of the training set (Possible values: The first 5,282 observations, the most recent 5 observations, the most recent 10 observations, the most recent 20 observations, or the most recent 50 observations).

To evaluate and compare their predictive performance, three key metrics are employed: root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Lower metric values indicate better predictive performance. These statistics for all models are reported in Tables 3-5. In Table 2, the results of t-tests based on the null hypothesis suggesting that the mean prediction error is zero are presented. The table displays mean prediction errors and p-values (in parentheses) associated with these t-tests.

Table 2 reveals that using a fixed-range training set, as opposed to a sliding window, consistently leads to underprediction, except for KRDMB with 4 lags and dual-class stock prices. Supporting this finding, Tables 3-5 show that the model employing a 5-day training window outperforms others across all predictive performance metrics. This finding aligns with the thesis against extended training periods in financial data forecasting and is consistent with prior research on investor and managerial myopia [15], [4], [8], and [50].

It's worth highlighting that the validation set, comprising the last 300 data points, coincided with a period marked by strong coherence between the daily returns of KRDMA, KRDMB, and KRDMD. Despite this robust coherence, the inclusion of dual stocks as predictors did not yield any noticeable additional insights into future stock prices. In fact, there is often a modest decline in accuracy as illustrated in Tables 3-5.

4. Conclusions

In this study, a case of dual-class stock structure on Borsa Istanbul was examined. This case is marked by a distinctive characteristic: prolonged disparities among three stocks, namely KRDMA, KRDMB, and KRDMD. These disparities reached staggering heights, with differentials soaring as high as 361.21%. To put this into perspective, consider that the most significant divergence observed between GOOG and GOOGL on NASDAQ has been approximately 5% since 2005. This stark contrast highlights that arbitrage opportunities between KRDMA, KRDMB, and KRDMD may present greater profit potential.

Moreover, this study also sheds light on the fact that even when there exists a strong coherence between dual-class stock prices, these prices may not necessarily serve as reliable predictors for future price movements of each other. Lastly, the study offers substantial empirical evidence supporting the practice of favoring shorter training periods over extended ones, contrary to the common practice of employing larger training sets with numerous observations and lags.

Overall, these findings underscore the importance of exercising caution when choosing training sets and predictors while training LSTM models. The inclusion of additional predictors, even those strongly coherent with the target variable, and the extension of training sets to encompass past values in financial time series have the potential to diminish predictive performance, ultimately resulting in poor forecasting.

Table 2. T-test Results

	Lag = 4		Lag = 9	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.020 (0.639)	0.034 (0.474)	0.025 (0.595)	0.034 (0.466)
KRDMB	0.021 (0.603)	0.030 (0.487)	0.019 (0.658)	0.015 (0.736)
KRDMD	0.026 (0.573)	0.025 (0.603)	0.030 (0.548)	0.035 (0.484)
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.038 (0.467)	0.031 (0.585)	-0.003 (0.963)	0.451 (0.049)
KRDMB	0.052 (0.294)	0.027 (0.594)	0.017 (0.764)	-0.001 (0.991)
KRDMD	0.048 (0.377)	0.033 (0.584)	0.031 (0.629)	0.008 (0.904)
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.097 (0.149)	0.139 (0.044)	0.009 (0.907)	0.064 (0.451)
KRDMB	0.087 (0.199)	0.041 (0.535)	0.121 (0.118)	0.094 (0.241)
KRDMD	0.137 (0.075)	0.050 (0.544)	0.097 (0.247)	-0.015 (0.857)
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	0.038 (0.655)	0.117 (0.165)	0.035 (0.710)	0.171 (0.114)
KRDMB	0.123 (0.130)	0.134 (0.105)	0.167 (0.073)	-0.022 (0.815)
KRDMD	0.062 (0.461)	0.215 (0.025)	0.239 (0.014)	0.257 (0.024)
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
KRDMA	-0.122 (0.009)	-0.126 (0.010)	-0.635 (<0.001)	-0.166 (0.005)
KRDMB	-0.167 (<0.001)	0.074 (0.099)	-0.399 (<0.001)	-0.343 (<0.001)
KRDMD	-0.210 (<0.001)	-0.328 (<0.001)	-0.742 (<0.001)	-0.378 (<0.001)

Table 3. KRDMA Predictive Performance Results

	Lag = 4		Lag = 9	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7433	0.8106	0.8006	0.8080
MAE	0.5077	0.5466	0.5578	0.5575
MAPE	3.387	3.6617	3.7375	3.7516
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.9027	0.9698	1.0020	1.1226
MAE	0.6230	0.6745	0.7044	0.7667
MAPE	4.2066	4.5580	4.8132	5.2047
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.1606	1.1915	1.4265	1.4671
MAE	0.8411	0.7863	0.9940	1.0660
MAPE	5.7301	5.3693	6.8049	7.3317
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4709	1.4571	1.6515	1.8731
MAE	1.1134	1.0881	1.2367	1.3784
MAPE	8.0633	7.7143	8.9525	9.8199
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8173	0.8510	1.1949	1.0517
MAE	0.5873	0.6132	0.9246	0.7664
MAPE	3.9752	4.1735	6.3747	5.2103

Table 4. KRDMB Predictive Performance Results

	Lag = 4		**Lag = 9**	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.6834	0.7567	0.7440	0.7757
MAE	0.4601	0.4995	0.5047	0.5251
MAPE	3.2067	3.4693	3.5233	3.6607
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8587	0.8917	1.0084	1.0299
MAE	0.5756	0.6030	0.6862	0.7005
MAPE	4.0155	4.2513	4.8465	4.9389
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.1757	1.1435	1.3400	1.3932
MAE	0.7796	0.7850	0.8755	0.9492
MAPE	5.4603	5.5653	6.2216	6.6879
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4014	1.4328	1.6128	1.6676
MAE	1.0322	1.0424	1.2265	1.2750
MAPE	7.6067	7.6053	9.0592	9.5743
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7612	0.7824	0.9884	0.9943
MAE	0.5407	0.5490	0.7306	0.7220
MAPE	3.8301	3.9241	5.2094	5.0590

Table 5. KRDMD Predictive Performance Results

	Lag = 4		Lag = 9	
Training Window = 5	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.7910	0.8321	0.8596	0.8564
MAE	0.5608	0.5811	0.6050	0.6117
MAPE	3.4491	3.5723	3.7123	3.7537
Training Window = 10	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.9408	1.0293	1.1151	1.1721
MAE	0.6584	0.7155	0.7678	0.8209
MAPE	4.0577	4.4043	4.7642	5.0716
Training Window = 20	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.3288	1.4243	1.4513	1.4485
MAE	0.9194	1.0142	1.0385	1.0465
MAPE	5.6409	6.3443	6.4796	6.5548
Training Window = 50	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	1.4489	1.6598	1.6948	1.9712
MAE	1.1061	1.2782	1.2880	1.5104
MAPE	7.1237	8.1918	8.1252	9.7193
Fixed Horizon	*Dual-Stock = No*	*Dual-Stock = Yes*	*Dual-Stock = No*	*Dual-Stock = Yes*
RMSE	0.8730	0.9280	1.2142	1.1088
MAE	0.6322	0.6894	0.9765	0.8100
MAPE	3.9094	4.2723	6.3353	5.0450

Declaration of Interest

The author declares that there is no conflict of interest.

Acknowledgements

The author would like to express sincere gratitude to Ayşegül Alaybeyoğlu, the editor, Muhammet Mustafa Bahşi, the section editor, and the two anonymous referees for their valuable suggestions, feedback, and professionalism.

References

[1] A. Bylund, "What's the difference between Alphabet's stock tickers, GOOG and GOOGL?," The Motley Fool, https://www.fool.com/investing/2022/07/27/whats-the-difference-between-goog-and-googl/ (accessed Aug. 16, 2023).

[2] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to Geophysical Time Series," Nonlinear Processes in Geophysics, vol. 11, no. 5/6, pp. 561–566, 2004. doi:10.5194/npg-11-561-2004

[3] B. F. Smith and B. Amoako-Adu, "Relative prices of dual class shares," The Journal of Financial and Quantitative Analysis, vol. 30, no. 2, p. 223, 1995. doi:10.2307/2331118

[4] C. D. Rio and R. Santamaria, "Stock characteristics, investor type, and market myopia," Journal of Behavioral Finance, vol. 17, no. 2, pp. 183–199, 2016. doi:10.1080/15427560.2016.1170682

[5] C. Erten, N. Chotai, and D. Kazakov, "Pair trading with an ontology of SEC Financial Reports," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020. doi:10.1109/ssci47803.2020.9308384

[6] C. Li, D. Songand D. Tao, "Multi-task Recurrent Neural Networks and Higher-order Markov Random Fields for Stock Price Movement Prediction: Multi-task RNN and Higer-order MRFs for Stock Price Classification", ACM, Jul. 2019. doi: 10.1145/3292500.3330983.

[7] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," Bulletin of the American Meteorological Society, vol. 79, no. 1, pp. 61–78, 1998. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

[8] C. W. Holden and L. L. Lundstrum, "Costly trade, managerial myopia, and long-term investment," Journal of Empirical Finance, vol. 16, no. 1, pp. 126–135, 2009. doi:10.1016/j.jempfin.2008.05.001

[9] C.-H. Chen, W.-H. Lai, and T.-P. Hong, "An effective correlation-based pair trading strategy using genetic algorithms," Computational Collective Intelligence, pp. 255–263, 2021. doi:10.1007/978-3-030-88081-1_19

[10] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," Expert Systems with Applications, vol. 129, pp. 273–285, 2019. doi:10.1016/j.eswa.2019.03.029

[11] E. Tokat and A. C. Hayrullahoğlu, "Pairs trading: Is it applicable to exchange-traded funds?," Borsa Istanbul Review, vol. 22, no. 4, pp. 743–751, 2022. doi:10.1016/j.bir.2021.08.001

[12] H. Maqsood et al., "A local and global event sentiment based efficient stock exchange forecasting using Deep Learning," International Journal of Information Management, vol. 50, pp. 432–451, 2020. doi:10.1016/j.ijinfomgt.2019.07.011

[13] H. S. Sim, H. I. Kim, and J. J. Ahn, "Is deep learning for image recognition applicable to stock market prediction?," Complexity, vol. 2019, pp. 1–10, 2019. doi:10.1155/2019/4324878

[14] H. Yang, Y. Zhu, and Q. Huang, "A multi-indicator feature selection for CNN-Driven Stock Index Prediction," Neural Information Processing, pp. 35–46, 2018. doi:10.1007/978-3-030-04221-9_4

[15] I. Nyman, "Stock market speculation and managerial myopia," Review of Financial Economics, vol. 14, no. 1, pp. 61–79, 2005. doi:10.1016/j.rfe.2004.06.002

[16] J. Eapen, D. Bein, and A. Verma, "Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019. doi:10.1109/ccwc.2019.8666592

[17] J. M.-T. Wu et al., "A graphic CNN-LSTM model for stock price prediction," Artificial Intelligence and Soft Computing, pp. 258–268, 2021. doi:10.1007/978-3-030-87986-0_23

[18] J. P. Ramos-Requena, M. N. López-García, M. A. Sánchez-Granero, and J. E. Trinidad-Segovia, "A cooperative dynamic approach to pairs trading," Complexity, vol. 2021, pp. 1–8, 2021. doi:10.1155/2021/7152846

[19] J. Wu, A pairs trading strategy for GOOG/GOOGL using machine learning, https://cs229.stanford.edu/proj2015/028_report.pdf (accessed Aug. 16, 2023).

[20] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," 2016 7th International Conference on Cloud Computing and Big Data (CCBD), 2016. doi:10.1109/ccbd.2016.027

[21] K. Nakagawa, T. Uchida, and T. Aoshima, "Deep factor model," ECML PKDD 2018 Workshops, pp. 37–50, 2019. doi:10.1007/978-3-030-13463-1_3

[22] K. Rydqvist, "Dual-class shares: A Review," Oxford Review of Economic Policy, vol. 8, no. 3, pp. 45–57, 1992. doi:10.1093/oxrep/8.3.45

[23] Ko, Ching-Ru, and Hsien-Tsung Chang. "LSTM-Based Sentiment Analysis for Stock Price Forecast." PeerJ Computer Science, vol. 7, 11 Mar. 2021, p. e408, https://doi.org/10.7717/peerj-cs.408.

[24] L. Ni et al., "Forecasting of forex time series data based on Deep Learning," Procedia Computer Science, vol. 147, pp. 647–652, 2019. doi:10.1016/j.procs.2019.01.189

[25] L. Zhang, "Pair trading with machine learning strategy in China Stock Market," 2021 2nd International Conference on Artificial Intelligence and Information Systems, 2021. doi:10.1145/3469213.3471353

[26] L. Zingales, "The value of the voting right: A study of the Milan stock exchange experience," Review of Financial Studies, vol. 7, no. 1, pp. 125–148, 1994. doi:10.1093/rfs/7.1.125

[27] M. Abe and H. Nakayama, "Deep learning for forecasting stock returns in the cross-section," Advances in Knowledge Discovery and Data Mining, pp. 273–284, 2018. doi:10.1007/978-3-319-93034-3_22

[28] M. R. Horner, “The value of the corporate voting right,” Journal of Banking and Finance, vol. 12, no. 1, pp. 69–83, 1988. doi:10.1016/0378-4266(88)90051-9
[29] M. U. Gudelek, S. A. Boluk, and A. M. Ozbayoglu, “A deep learning based stock trading model with 2-D CNN trend detection,” 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017. doi:10.1109/ssci.2017.8285188
[30] M. Wen, P. Li, L. Zhang, and Y. Chen, “Stock market trend prediction using high-order information of Time Series,” IEEE Access, vol. 7, pp. 28299–28308, 2019. doi:10.1109/access.2019.2901842
[31] N. Foysal Ahamed, and M. Mahmudul Hasan. “Predicting Stock Price from Historical Data using LSTM Technique.” Journal of Artificial Intelligence and Data Science 3.1: 36-49.
[32] N. Naik and B. R. Mohan, “Stock price movements classification using machine and deep learning techniques-the case study of Indian Stock Market,” Engineering Applications of Neural Networks, pp. 445–452, 2019. doi:10.1007/978-3-030-20257-6_38
[33] Niu, Hongli, et al. “A Hybrid Stock Price Index Forecasting Model Based on Variational Mode Decomposition and LSTM Network.” Applied Intelligence, vol. 50, no. 12, 17 July 2020, pp. 4296–4309, https://doi.org/10.1007/s10489-020-01814-0.
[34] P. Oncharoen and P. Vateekul, “Deep learning using risk-reward function for stock market prediction,” Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018. doi:10.1145/3297156.3297173
[35] P. Patil, C.-S. M. Wu, K. Potika, and M. Orang, “Stock market prediction using ensemble of graph theory, machine learning and Deep Learning Models,” Proceedings of the 3rd International Conference on Software Engineering and Information Management, 2020. doi:10.1145/3378936.3378972
[36] P. Srivastava and P. K. Mishra, “Stock market prediction using RNN LSTM,” 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021. doi:10.1109/gcat52182.2021.9587540
[37] R. C. Lease, J. J. McConnell, and W. H. Mikkelson, “The market value of control in publicly-traded corporations,” Journal of Financial Economics, vol. 11, no. 1–4, pp. 439–471, 1983. doi:10.1016/0304-405x(83)90019-3
[38] S. Basodi, C. Ji, H. Zhang, and Y. Pan, “Gradient amplification: An efficient way to train deep neural networks,” Big Data Mining and Analytics, vol. 3, no. 3, pp. 196–207, 2020. doi:10.26599/bdma.2020.9020004
[39] R. Zhang, Z. Yuan, and X. Shao, “A new combined CNN-RNN model for sector stock price analysis,” 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018. doi:10.1109/compsac.2018.10292
[40] S. Cai, X. Feng, Z. Deng, Z. Ming, and Z. Shan, “Financial News quantization and Stock Market Forecast Research based on CNN and LSTM,” Lecture Notes in Computer Science, pp. 366–375, 2018. doi:10.1007/978-3-030-05755-8_36
[41] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735
[42] S. Liu, C. Zhang, and J. Ma, “CNN-LSTM neural network model for quantitative strategy analysis in stock markets,” Neural Information Processing, pp. 198–206, 2017. doi:10.1007/978-3-319-70096-0_21
[43] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, “Forecasting stock market crisis events using deep and Statistical Machine Learning Techniques,” Expert Systems with Applications, vol. 112, pp. 353–371, 2018. doi:10.1016/j.eswa.2018.06.032
[44] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, “Stock price prediction using LSTM, RNN and CNN-sliding window model,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017. doi:10.1109/icacci.2017.8126078
[45] Thorir Mar Ingolfsson, “Insights into LSTM architecture,” Thorir Mar Ingolfsson, https://thorirmar.com/post/insight_into_lstm/ (accessed Aug. 16, 2023).
[46] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, “Leveraging Social Media News to predict stock index movement using RNN-Boost,” Data & Knowledge Engineering, vol. 118, pp. 14–24, 2018. doi:10.1016/j.datak.2018.08.003
[47] W. Jiang, “Applications of deep learning in stock market prediction: Recent progress,” Expert Systems with Applications, vol. 184, p. 115537, 2021. doi:10.1016/j.eswa.2021.115537
[48] W. L. Megginson, “Restricted voting stock, acquisition premiums, and the market value of corporate control,” The Financial Review, vol. 25, no. 2, pp. 175–198, 1990. doi:10.1111/j.1540-6288.1990.tb00791.x
[49] X. Ding, Y. Zhang, T. Liuand J. Duan, “Deep learning for event-driven stock prediction”, AAAI Press, Jul. 2015.
[50] X. Sheng, S. Guo, and X. Chang, “Managerial myopia and firm productivity: Evidence from China,” Finance Research Letters, vol. 49, p. 103083, 2022. doi:10.1016/j.frl.2022.103083
[51] Y. Liu, Q. Zeng, H. Yang, and A. Carrio, “Stock price movement prediction from financial news with deep learning and knowledge graph embedding,” Knowledge Management and Acquisition for Intelligent Systems, pp. 102–113, 2018. doi:10.1007/978-3-319-97289-3_8
[52] Y. Song, J. W. Lee, and J. Lee, “A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction,” Applied Intelligence, vol. 49, no. 3, pp. 897–911, 2018. doi:10.1007/s10489-018-1308-x
[53] Y. Zhao and M. Khushi, “Wavelet denoised-resnet CNN and LIGHTGBM method to predict forex rate of Change,” 2020 International Conference on Data Mining Workshops (ICDMW), 2020. doi:10.1109/icdmw51313.2020.00060
[54] Z. Hu, Y. Zhao, and M. Khushi, “A survey of Forex and Stock Price Prediction using Deep learning,” Applied System Innovation, vol. 4, no. 1, p. 9, 2021. doi:10.3390/asi4010009
[55] Z. Zeng and M. Khushi, “Wavelet denoising and attention-based RNN- Arima model to predict forex price,” 2020 International Joint Conference on Neural Networks (IJCNN), 2020. doi:10.1109/ijcnn48605.2020.9206832