

Research on Optimization Model of Stock Index Forecasting Considering Cross-Market Linkage Factors

Hui Liu^{1*}

¹ School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, China

Email Address

750476337@qq.com (Hui Liu)

*Correspondence: 750476337@qq.com

Received: 14 June 2024; Accepted: 12 July 2024; Published: 24 July 2024

Abstract:

With the development of financial integration, early warning signals generated by a single market will be rapidly transmitted to other markets through the correlation network, resulting in price linkages. Firstly, the combination of mutual information theory and cointegration theory proves that the price of Bitcoin and the IXIC are strongly correlated with each other and that they have a stable equilibrium relationship in the long run. In order to validate the predictive effect of Bitcoin transaction features on stock indices, a Bayesian optimization-based Long Short-Term Memory network is introduced, and Bitcoin transaction features are incorporated into the feature set as inputs to the model for the prediction of stock indices. The experimental results show that the model with the addition of hybrid features has a significant improvement in both prediction error and goodness-of-fit, confirming the usefulness of Bitcoin trading features in predicting stock indices.

Keywords:

Stock Index Forecast, Cross-Market Linkages, LSTM, Bayesian Optimization

1. Introduction

Stock market forecasting has always been one of the biggest challenges for financial markets because of the complex, noisy, chaotic, dynamic, volatile and non-parametric nature of its time series, where several micro- and macro-economic attributes and characteristics affect price formation, increasing the complexity of the forecasting task. Satoshi Nakamoto developed the first cryptocurrency, Bitcoin, in 2008 [1]. Its emergence represents an innovation in digital currencies, reflecting the current trend of globalization of world politics and economics, which continues to impact the traditional financial industry [2]. The contemporary financial industry is constantly evolving and various financial activities are integrated with each other, and their individual risks can be quickly transferred to other relevant markets in the event of a crisis in one market, through the irrational herd effect of investors, asset-liability correlations and market expectations.

As cryptocurrencies come into the public eye, scholars have begun to focus on the relationship between this new investment product and the traditional stock market. The dynamic efficiency in Bitcoin price fluctuations is consistent with the adaptive market hypothesis (AME), where the presence of emotions and the generation of events change market efficiency [3]. Corbet et al. found cryptocurrencies to be relatively isolated from financial and economic assets by combining time and frequency domain analysis [4]. Chinese scholars Luo Mei and Zhou Weihua et al. argue that there is a linkage between Bitcoin and stock prices, and that stock returns can be predicted based on the level of Bitcoin returns during periods of high volatility in Bitcoin prices. The market risk that exists in Bitcoin can be transmitted to the stock market causing stock price fluctuations [5,6].

Machine learning models are widely used for stock prediction in existing research. Sim et al. verified the applicability of convolutional neural networks for image recognition to stock market prediction [7]. LSTM networks have been admired by a wide range of researchers because of their excellent performance in processing financial time series data [8,9]. Meanwhile, the factors affecting the stock market are also of interest to researchers, and many different external data sources are used for stock market forecasting, including market data, technical indicators, Wikipedia traffic, market news, social media comments, etc [10,11].

This paper focuses on the cross-market price linkage between Bitcoin and the stock market, and selects the trading data of Bitcoin and the IXIC from February 2017 to August 2023, and finds that there is a strong correlation between the trading characteristics of Bitcoin and the IXIC and a significant cointegration relationship between the two sample time series through the theory of mutual information and cointegration test. This paper introduces the LSTM network model, fuses the Bitcoin trading features with the basic stock features and technical features to form a hybrid feature set as the input of the model, and predicts the next day's closing price of the stock index, and sets up a control test for the focused Bitcoin trading features to verify the improvement of the model's prediction performance by the Bitcoin trading features. In addition, this paper improves the original LSTM network and proposes a Bayesian optimization-based LSTM (BO-LSTM) prediction model, which improves the tuning efficiency of the model and significantly improves the prediction performance of the model.

2. Model Construction

The construction process of LSTM stock index prediction model based on Bayesian optimization is as follows:

- (1) Obtain historical data of stock indices and Bitcoin trading to generate technical characteristics of stock indices using Talib library;
- (2) Data pre-processing. Including missing values, duplicate values and outliers, the processed data are converted into array form for standardization. Sliding window processing is performed on the array, rolling 5 days as a group to predict the stock closing price of the next day with the stock characteristic data of the past 5 days;
- (3) The preprocessed data are divided into training set, validation set and test set respectively in chronological order according to the ratio of 6:2:2;

(4) Depending on the purpose of the experiment, a control group is set up by different combinations of features, chosen to input technical features of stock indices and Bitcoin trading features;

(5) Construct the LSTM network. The MSE function is used for the loss function, the adam optimizer is used for the optimizer, the relu is chosen for the activation function, and a dropout layer is added to the network to alleviate the overfitting problem of the model;

(6) Hyperparameter optimization. The Bayesian optimization algorithm is used to update the model hyperparameters iteratively to determine the combination of hyperparameters that minimizes the target value and obtain the final model;

(7) Model Comparison. Compare the differences between the actual and predicted values of different control groups to determine the predictive performance of Bitcoin trading characteristics on stock indices;

3. Empirical Analysis

This section focuses on the effectiveness of Bitcoin trading features on the prediction of IXIC, adding Bitcoin trading features to the model input dimension data, and comparatively analyzing the role of Bitcoin trading features on the improvement of the accuracy of the prediction results through controlled experiments.

3.1. Data Acquisition

In this paper, data related to Bitcoin and the IXIC from February 2017 to August 2023 are selected as a research sample. Bitcoin trading data was obtained from the CoinMarketCap website, which is one of the most recognizable cryptocurrency ticker websites and the main source of cryptocurrency ticker data. The IXIC was obtained from the Investing website. In addition, in order to make the sample data more adequate and to provide support for predicting short-term stock fluctuations, this paper generates several stock technical indicators by means of a third-party library, Talib, and manual calculations, which is a very powerful third-party library for calculating indicators for technical analysis, including several categories of popular indicators, and has been widely used by researchers for technical analysis of financial market data.

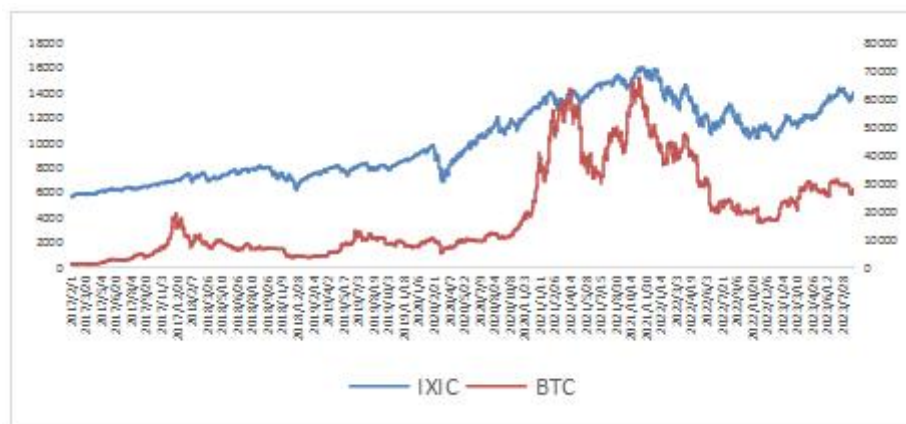


Figure 1. BTC and IXIC Price Trend.

Table 1. LSTM input dimension features.

Features	Specific eigenvectors
Basic	Open, Close, High, Low, Volume
Bitcoin	Open, Close, High, Low, Volume, MarketCap
Technical	MA, Momentum, Volatility, Volume and Liquidity, BIAS, BRAR, CR
Hybrid	Basic, Bitcoin Trading and Technical features

3.2. Data Preprocessing

Unlike the traditional stock market, the Bitcoin market is not closed, so the Bitcoin market still generates transaction data during the stock market closure, this paper takes the stock market transaction date as the primary, and takes the Bitcoin transaction data of the corresponding transaction date, so as to ensure that the data of the two are one-to-one correspondence. In order to solve the non-comparability of the data indicator variables due to the differences in nature, magnitude, order of magnitude and other attributes, this paper adopts the maximum-minimum standardization method to convert the preprocessed data indicators into dimensionless standardized values. Finally, the preprocessed feature sequences are divided into training set, validation set and test set in the ratio of 6:2:2 and chronological order.

3.3. Correlation test

In this paper, we use MIC (Maximum Information Coefficient) to measure the degree of association between the indicator data. MIC belongs to the largest information-based nonparametric explorations, which are not limited to linear functional relationships between variables, but also cover nonlinear functional relationships, especially for high-dimensional data [12].

Table 2. Maximum Information Relevance Ranking.

IXIC	BTC	MIC
Close	MarketCap	0.9239
	Low	0.8877
	Close	0.8849
	Open	0.8747
	High	0.8736
	Volume	0.6478

From the Table 2, it can be seen that the closing price of the IXIC has a strong informational correlation with the trading characteristics of Bitcoin to a significant degree, with MIC coefficients higher than 0.6 in all cases, including the MIC coefficient of Bitcoin market capitalization and the closing price of the stock index, which is as high as 0.9239.

In order to more rigorously argue the degree of association between the Bitcoin price and the IXIC, this paper applies the theory of cointegration to test the cointegration relationship between the IXIC and the Bitcoin price. The cointegration test is used to determine whether a linear combination of a set of non-stationary series has a stable equilibrium relationship. The original series is found to be non-stationary through the unit root test, so the sample series is taken as a first-order difference, and the sample series after the first-order difference is stationary and satisfies the first-order monotonicity. Next, the Engle-Granger (EG) two-step cointegration test is used to test the sample time series for cointegration. The experiment shows that the residual series obtained by least squares regression on the two sample time series are

smooth, so this paper concludes that the relationship between the IXIC and the price of Bitcoin is smooth and there is a long-term stable equilibrium relationship.

3.4. Analysis of Projected Results

In order to verify the impact of Bitcoin transaction features on the prediction performance, this paper sets up control experiments based on different combinations of features, and verifies the enhancement effect of Bitcoin transaction features on the accuracy of the prediction results by comparing the prediction results of different control groups, and the control groups are set up as shown in the Table 3.

Table 3. Control groups setup.

Control Groups	Features Combination
Group 1	Basic
Group 2	Basic + Bitcoin
Group 3	Basic + Technical
Group 4	Hybrid

The performance of a deep learning model depends largely on the selection of hyperparameters, which are the adjustment buttons that control the structure, function and efficiency of the model. The hyperparameter optimization method chosen in this paper is Bayesian optimization, which first determines the hyperparameters affecting the performance of the LSTM network and their corresponding ranges, defines the hyperparameter search space, randomly initializes the hyperparameter combinations in the search space and constructs the network model to be trained in this way, uses the data from the training set for training the model, and inputs the data from the validation set into the trained model for prediction. Calculate the mean square error (MSE) of the true and predicted values as the target value for optimization, and construct a TPE agent model consisting of hyperparameters and target value. The hyperparameter combinations are iteratively sampled in the search space using the acquisition function, and the agent model is updated based on the sampling results until the target optimized value converges. The determined LSTM network model that minimizes the objective value is the optimal hyperparameter combination, and the determined LSTM network model is the final model. (Table 4)

Table 4. Hyper-parameter space.

Hyper-parameter	Space
LSTM Layer Units	[4, 128]
Dense Layer Units	[4, 18]
Batch Size	[8, 128]
Dropout Rate	(0.05, 0.1, 0.2, 0.3)
Initial Learning Rate	(0.001, 0.0005, 0.0001)

In this paper, 100 groups of hyperparameter combinations are extracted from the hyperparameter space for training and optimization respectively, and the optimal hyperparameter combinations obtained from each control group after Bayesian optimization are shown in Table 5:

Table 5. Optimal hyper-parameter combinations.

Control Groups	LSTM Layer Units	Dense Layer Units	Batch Size	Dropout Rate	Initial Learning Rate
Group 1	65	29	16	0.1	0.001
Group 2	128	119	10	0.05	0.001
Group 3	115	22	20	0.1	0.0005
Group 4	110	17	22	0.05	0.0005

The state variables in the test set are input into the final model for prediction. For the prediction of the closing price trend of the stock index on the test set, the predicted closing price derived from each control group is compared with the real closing price, and the closing price prediction chart is shown in Figure 2, in which the control group II and control group IV are the prediction results after adding Bitcoin trading features. It is obvious from the chart that the model fits better on the test set, and the predicted closing price can overlap with the real closing price better, which indicates that the final model has excellent performance for the prediction of stock closing price in terms of both prediction accuracy and generalizability.

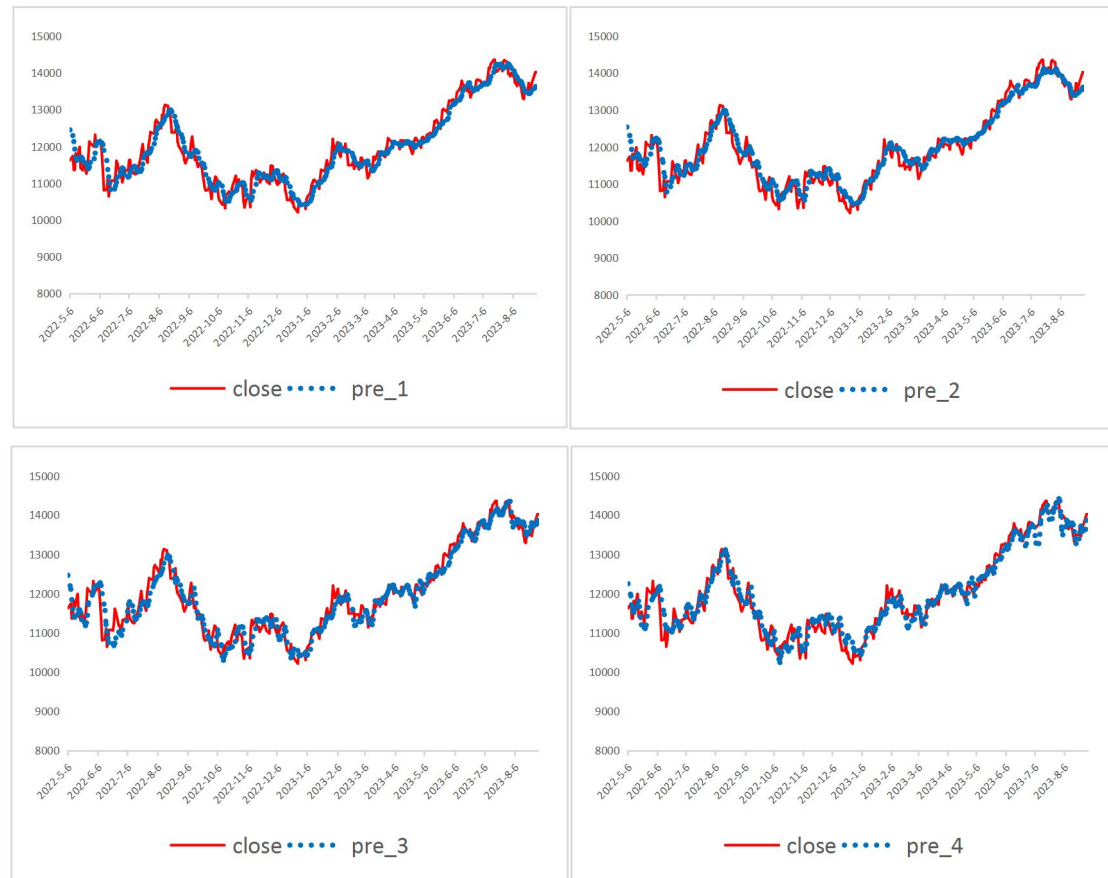


Figure 2. Control groups test set prediction trend.

In order to more specifically analyze the prediction performance of each control group model on the test set data, this paper introduces commonly used regression model evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R^2 (Coefficient of Determination, also known as Goodness of Fit) (Table 6).

RMSE, MAE, MAPE indicators reflect the prediction accuracy of the model, and R^2 reflects the degree of model fitting. Comparison of the indicators between control group 1 and control group 2, control group 3 and control group 4 reveals that the prediction accuracy and the degree of fit of the model are significantly improved after the addition of the Bitcoin trading features, in which the RMSE decreases by 3.35% and 6.67%, the R^2 improves by 0.5% and 0.89%, and the MAE and the MAPE also decrease, which indicates that the addition of Bitcoin trading features has a significant improvement on the prediction performance of the model, and the Bitcoin trading features can significantly improve the model prediction performance. This indicates

that the addition of Bitcoin trading features has a significant effect on the model's prediction performance, and Bitcoin trading features can be used to predict the price movement of the stock market. By overall comparison, Control Group IV has the best prediction performance, compared to the baseline model Control Group I. After adding mixed features, RMSE, MAE, and MAPE decreased by 11.52%, 13.02%, and 12.57%, respectively, and R^2 improved by 1.65%. It indicates that adding different feature dimensions is better than using basic features alone for prediction, the reason is that the network can obtain more useful information by extracting features of different dimensions for training, and this information can help to improve the prediction accuracy of the model effectively.

Table 6. Comparison of test set regression metrics for control groups.

Control Groups	RMSE	MAE	MAPE	R^2
Group 1	0.0269	0.0215	3.6787	0.9292
Group 2	0.0260	0.0206	3.5585	0.9338
Group 3	0.0255	0.0199	3.4141	0.9362
Group 4	0.0238	0.0187	3.2164	0.9445

In order to further prove the effectiveness of the LSTM model based on Bayesian optimization constructed in this paper in stock price prediction, RNN, GRU and BiLSTM models are also introduced as comparison models, and the prediction results of the four models are shown in Table 7. From the table, it can be seen that the RMSE, MAE, and MAPE indicators of the BO-LSTM model proposed in this paper are significantly lower than those of the other baseline models, and its coefficient of determination is the highest among the four models, and the BO-LSTM model is better than the baseline model both in terms of prediction accuracy and the degree of fitting, which indicates that the prediction effect of the model constructed in this paper is better.

Table 7. Comparison of model regressors.

Models	RMSE	MAE	MAPE	R^2
BO-LSTM	0.0238	0.0187	3.2164	0.9445
RNN	0.0268	0.0211	3.4789	0.9295
GRU	0.0267	0.0214	3.6509	0.9303
Bi-LSTM	0.0244	0.0197	3.3098	0.9417

4. Conclusions

Focusing on the price linkage between Bitcoin and the stock market, this paper finds informative correlations and cointegrations between Bitcoin prices and the IXIC time series, confirming that there is indeed a price linkage between the two. The prediction accuracy of the model is significantly improved after adding Bitcoin trading features, Bitcoin trading features are effective for stock price prediction, and Bitcoin prices contain useful information that can predict stock indices, which is of some significance to investors and market regulatory organizations, and also provides new ideas and empirical foundations for subsequent related research. In addition, the hyperparameter tuning problem has been a challenge in deep learning research, this paper improves the original LSTM model and constructs an LSTM prediction model based on Bayesian optimization, which improves the efficiency of the hyperparameter optimization compared with the traditional method, and is able to find the optimal hyperparameter combinations faster and more accurately. By comparing with the

baseline model, the model proposed in this paper has better prediction performance, which verifies the effectiveness of the model.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This work was supported by the Graduate Research Innovation Fund of Anhui University of Finance and Economics, grant number ACYC2022134.

Acknowledgments

The author would like to acknowledge Graduate Student Research Innovation Fund of Anhui University of Finance and Economics for funding this research.

References

- [1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008. Available online: <https://bitcoin.org/bitcoin.pdf> (accessed on 12 July 2024).
- [2] Jia, L.P. Bitcoin's Theory, Practice, and Impact. *Studies of International Finance*, 2013, 12, 14-25, DOI: CNKI:SUN:GJJR.0.2013-12-002.
- [3] Gradojevic N, Kukolj D, Adcock R, et al. Forecasting Bitcoin with technical analysis: A not-so-random forest? *International Journal of Forecasting*, 2023, 39(1), 1-17. DOI: <https://doi.org/10.1016/j.ijforecast.2021.08.001>.
- [4] Corbet S, Meegan A, Larkin C, et al. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics letters*, 2018, 165, 28-34, DOI: 10.2139/ssrn.3070288.
- [5] Luo M, Gu P, Jiang Y L. Bitcoin's Return Correlation with the Dow Jones Indexes. *China Accounting Review*, 2020, 18(01), 1-16.
- [6] Zhou, W.H.; Li, Y.L.; Tan, J. A study on the correlation between cryptocurrency and stock market risk. *China Soft Science*, 2021, S1, 116-126, DOI: 10.3969/j.issn.1002-9753.2021.z1.014.
- [7] Sim, HS.; Kim, H.I.; Ahn, J.J. Is deep learning for image recognition applicable to stock market prediction? *Complexity*, 2019, 2019(1), 4324878. DOI: <https://doi.org/10.1155/2019/4324878>.
- [8] Hu Y W. Stock Prediction Based on Optimized LSTM Model. *Computer Science*, 2021, 48(S1), 151-157. DOI: 10.11896/jsjcx.200400011.
- [9] Lin Y, Yan Y, Xu J, et al. Forecasting stock index price using the CEEMDAN-LSTM model. *The North American Journal of Economics and Finance*, 2021, 57, 101421. DOI: <https://doi.org/10.1016/j.najef.2021.101421>.
- [10] Xu, X.C.; Tian, K. A New Approach to Stock Index Prediction Based on Financial Text Sentiment Analysis. *The Journal of Quantitative & Technical Economics*, 2021, 38(12), 124-145, DOI:10.13653/j.cnki.jqte.2021.12.009.

- [11]Liu, J.; Lu, Z.; Du, W. Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. 2019. CorpusID:102352388. Available online: <https://api.semanticscholar.org/> (accessed on 12 July 2024).
- [12]Reshef, D.N.; Reshef, Y.A.; Finucane, H.K. et al. Detecting novel associations in large data sets. *Science*, 2011, 334(6062), 1518-1524. DOI: 10.1126/science.1205438.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)