IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Research on Financial Data Prediction Algorithm Based on Deep Learning

Siddhant Jadhav, Hritik Mali, Ashish Bhoria
Sinhgad Institute of Technology, Lonavala, India

Abstract: *Financial market forecasting remains a challenging research area due to the complex nature of financial markets and the inherent difficulty in predicting market trends. This paper explores the application of deep learning and data mining techniques to improve the accuracy of financial market predictions. It addresses the nonlinear, nonstationary, and multiscale characteristics of financial time series data, as well as the challenges posed by noisy trading components. The paper also discusses the significance of forecasting in the context of China's capital market and the growing interest of individual investors in foreign exchange trading. It emphasizes the need for scientific forecasting methods and theories to guide investment decisions in this volatile market.*

Keywords: Financial market forecasting

I. INTRODUCTION

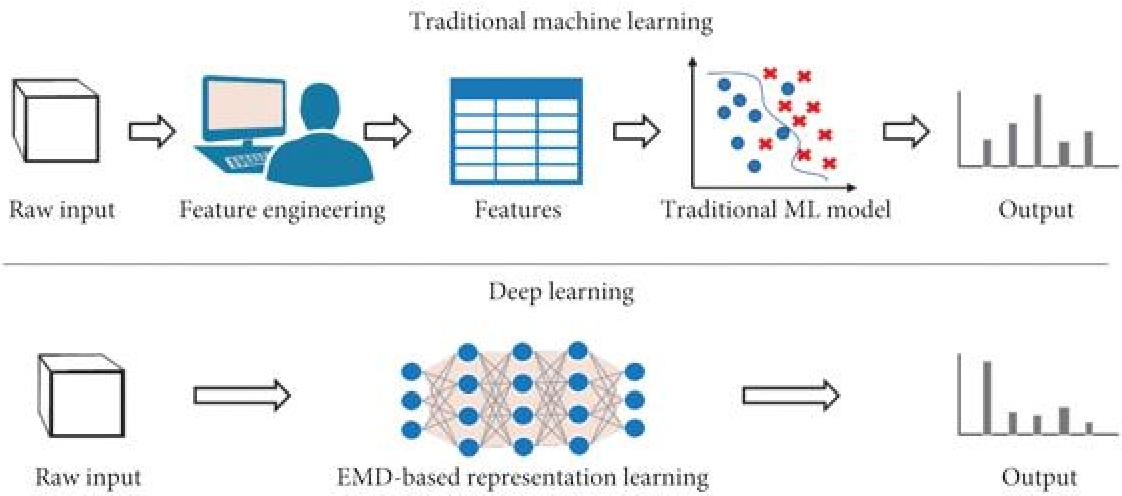
Despite significant advancements in information science, technology, and interdisciplinary research, accurately predicting the current and future state of financial markets remains a formidable challenge. Financial market forecasting is complex due to the nonlinearity, non-stationarity, and multiscale nature of financial time series, as well as the presence of noisy trading components. In China's capital market, the increasing number of individual investors participating in the foreign exchange market has emphasized the need for scientific forecasting methods to guide investments and explain price movements. The unpredictability and high-risk nature of foreign exchange prices make this market particularly intricate. Researchers worldwide have been attracted to studying financial market forecasting due to the potential economic benefits it holds.Throughout history, people have attempted to forecast changes in asset prices, initially relying on natural signs and later employing mathematical models and historical data analysis. The complexity of financial markets arises from the diverse investment styles and trading strategies of market participants, including both rational and irrational behaviours. The Chinese stock market, in particular, is characterized by a large number of retail investors and a lack of institutional investors, contributing to a super complex dynamic system influenced by emotional fluctuations. The foreign exchange market presents further challenges with its intricate interval financial time series and numerous noisy components.To tackle financial market forecasting, researchers have made significant progress in developing nonlinear mathematical models, such as artificial neural networks, support vector machines, genetic algorithms, wavelet analysis, and empirical model analysis. Artificial neural networks, in particular, have been extensively used to accurately simulate and represent the complexities of financial markets. Various types of neural networks, including feedforward neural networks, linear neural networks, radial basis function neural networks, probabilistic neural networks, stochastic neural networks, and feedback neural networks, have been developed and refined to improve forecasting accuracy.To achieve more precise financial market forecasting, researchers have invested considerable effort in advancing the algorithms and architectures of neural networks. They have also explored innovative approaches by integrating data mining techniques and sampling methods. Data mining, as a multidisciplinary field that emerged in the late 1980s, has evolved to handle dynamic data sources and environments. Traditional data mining approaches fall short in analyzing real-time data, which has driven the importance of dynamic data mining. The goal of dynamic data mining is to extract valuable knowledge from diverse and timely data, going beyond fixed datasets. It has practical applications in various business fields, providing decision support and improving data-driven operations.This paper introduces the application of empirical modal decomposition signal processing in financial time series analysis. It highlights the distinguishing features of empirical modal analysis and describes the pre-processing steps involving sliding windows. The principle of empirical mode decomposition (EMD) is then explained,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

encompassing the EMD decomposition process, instantaneous frequency, intrinsic mode functions, sifting process, completeness, and orthogonality. The paper further introduces the EMD decomposition process for interval-type time series. The EMD algorithm serves as the foundation for constructing the FEPA (Feature Extraction based on EMD and Principal Component Analysis) model and the reference EMD-BPNN (Backpropagation Neural Network) model. The extraction of principal components after EMD decomposition is a crucial step in the FEPA model. Several forecasting models are built for the CSI 300 Index, and their empirical analysis, including data sources, principal component analysis, forecasting performance criteria, and discussions, are presented. The FEPA model exhibits superior forecasting performance compared to linear models like ARIMA and demonstrates improvement over the EMD-BPNN model to a certain extent.

II. ARCHITECTURE FOR DEEP LEARNING IN FINANCIAL MARKET FORECASTING

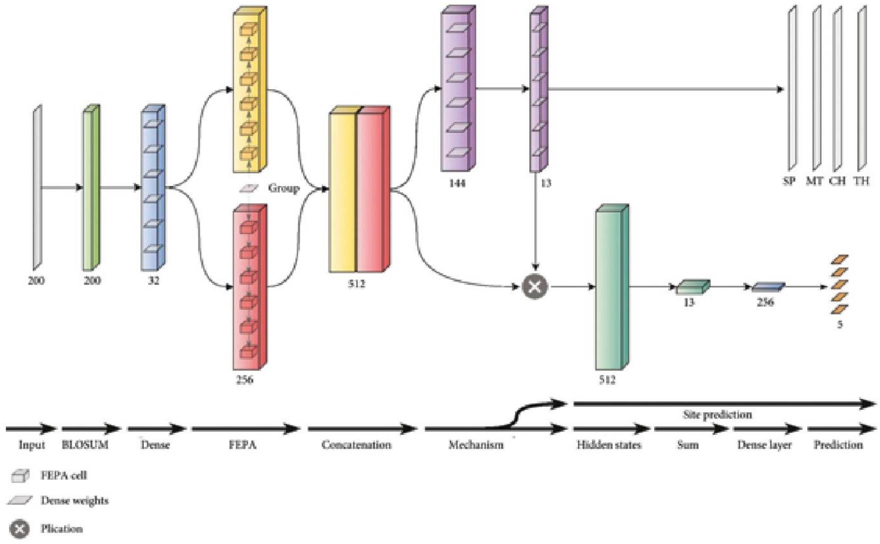
The forecasting of foreign exchange market prices is challenging due to their stochastic nature and the presence of nonlinear, nonstationary, and multiscale characteristics. Traditional forecasting methods struggle to capture the complexity and variability of these fluctuations, leading to reduced accuracy. To address this, a novel model called FEPA (Combining Empirical Modal Decomposition, Principal Component Analysis, and Neural Network) is proposed in this study. The model integrates the empirical modal decomposition algorithm, principal component analysis, and neural network to and predict price fluctuations in the foreign exchange market.



The FEPA model follows a sequential process: it applies the sliding window technique to extract the original financial time series, performs empirical modal decomposition to obtain different scales of essential functions, and then applies principal component analysis to reduce the dimensionality of the decomposed data and eliminate redundant information. The model combines decomposition reconstruction and integration, enhancing its learning ability for financial time series and improving prediction accuracy. The feedforward neural network is used for training and prediction within the FEPA model.The flowchart of the FEPA model illustrates its overall structure and steps. The EMD decomposition algorithm is employed to handle the nonlinear and stochastic characteristics of financial data, addressing the prediction challenge. Principal component analysis is then used for dimensionality reduction of the intrinsic mode function components, extracting the most informative data and reducing noise interference. The reduced-dimension data is fed into the neural network model for training and prediction, taking advantage of the neural network's adaptability and prediction performance. Thus, the FEPA model combines the strengths of both EMD decomposition and neural network modelling.

DOI: 10.48175/IJARSCT-10971

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023



The FEPA model also incorporates dimensionality reduction through principal component analysis to improve response speed, considering the need for fast and accurate financial forecasting. From a theoretical standpoint, the FEPA model is considered an ideal forecasting model.

In financial time series analysis, the raw time series data is usually discrete and exhibits multiscale spatial characteristics based on the data's sampling frequency. The data can be divided into regular time series with corresponding time intervals representing different time scales. For instance, hourly data can be subdivided into intervals of one hour, with each interval representing one data point comprising the opening price, high price, low price, and closing price. Additionally, the volume for each time interval within a selected time scale can be used for data mining. The concept of yin-yang volatility is introduced, which calculates positive and negative volatility separately and combines them to capture the overall volatility of the price time series.Overall, the FEPA model combines advanced techniques such as empirical modal decomposition, principal component analysis, and neural networks to tackle the challenges of forecasting foreign exchange market prices, offering improved accuracy and responsiveness in financial predictions.

III. FORECAST ANALYSIS

The study of market price movements has always been a popular topic in finance research as price movements reflect information. Researchers have been focusing on utilizing trading information to understand and predict financial markets. The fluctuations in foreign exchange prices, for example, have a significant impact on global economies, making it crucial to analyse and predict their behaviour. In this paper, various forecasting models, including the FEPA model, are used to predict the CSI 300 Index and compare their performance.

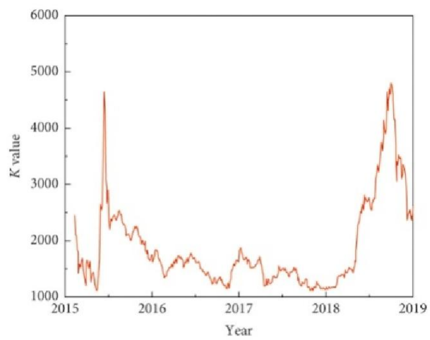


Fig. Daily chart of the closing price of the CSI 300 Index

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

The CSI 300 Index represents the major stocks in the Shanghai and Shenzhen stock markets, providing an overview of their overall performance. The dataset used for empirical analysis covers recent years, excluding holidays and other factors. It is divided into a training set and a test set, with the training set containing the first 1000 data points and the test set containing the last 250 data points representing the daily closing price of the index. This dataset encompasses various unexpected events and financial crises, providing ample training data for the model.

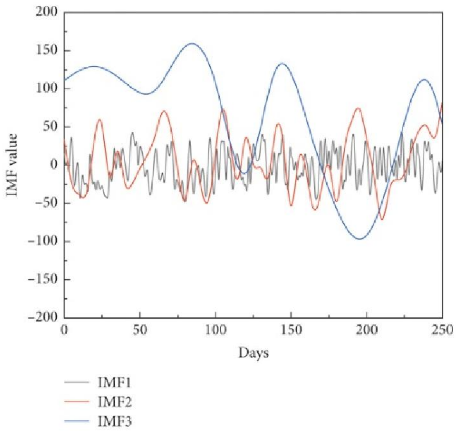


Fig. IMF component chart of the closing price of the CSI 300 Index

To analyze the price fluctuations of the CSI 300 Index, the Empirical Mode Decomposition (EMD) algorithm is applied to decompose the nonlinear and nonstationary signal into a series of Intrinsic Mode Functions (IMFs) and a trend term. Each IMF component reflects investor sentiment fluctuations at different frequencies, with high-frequency IMFs capturing short-term operations and low-frequency IMFs representing medium-term operations. The trend term reflects the overall trend of investor sentiment. The EMD decomposition process ensures that the individual IMF series are orthogonal to each other.

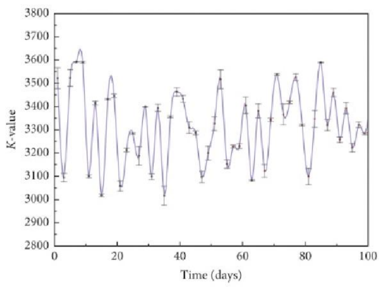


Fig. CSI 300 daily K-line chart.

Before conducting the principal component analysis, factor analysis is performed to assess the suitability of applying the principal component analysis. The Kaiser-Meyer-Olkin (KMO) value, a measure of sampling adequacy, is found to be greater than 0.5, satisfying the conditions for conducting principal component analysis. Principal component analysis is then used to reduce the dimensionality of the IMF series by mapping them into low-dimensional mutually uncorrelated principal components. The eigenvalues of the CSI 300 Index IMF series after the principal component analysis demonstrate that the cumulative contribution of the first four eigenvalues exceeds 85%.For training data input to the neural network, the first 22 principal components of the CSI 300 Index are extracted. The performance of the FEPA model, along with other forecasting models, is tested and compared. The empirical analysis focuses on the prediction effect of the interval EMD decomposition model, which replaces the forward-rolling EMD decomposition algorithm used in the FEPA model. The interval EMD model is tested for short-term trend prediction of the closing price, highest price, and lowest price of the CSI 300 Index.The results show that the interval EMD model outperforms the FEPA model in terms of prediction accuracy. The interval EMD decomposition method yields smaller prediction error values for the three time series: closing price, highest price, and lowest price. Additionally, the interval EMD decomposition method demonstrates a slight improvement in the hit rate when predicting the closing price of the index.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Compared to the FEPA model, the interval EMD decomposition method achieves better forecasting performance for the CSI 300 Index.

IV. DEEP LEARNING-BASED ANALYSIS OF EMPIRICAL FOREX RESULTS' PREDICTION

The exchange rate is a crucial factor in a country's economic system, impacting both domestic and foreign economies. Fluctuations in the exchange rate reflect changes in the international purchasing power of the local currency. The foreign exchange market is influenced by various factors, such as the securities market, commodity market, crude oil futures market, and import/export trade volume of each country.

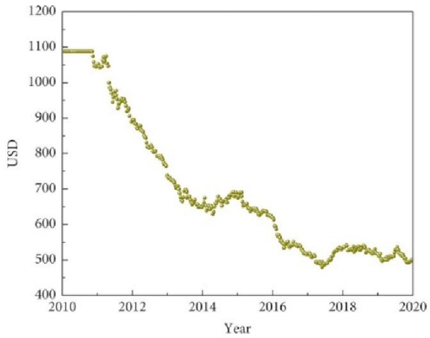
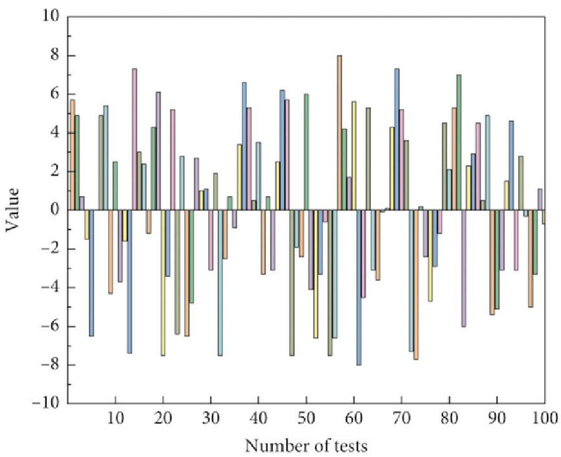


Fig. Daily Chart

Exchange rate fluctuations often exceed acceptable levels, causing significant impacts on the economy and the lives of residents. China adopted a floating exchange rate system in 2005, and the addition of the RMB to the IMF's Special Drawing Rights currency basket in 2017 paved the way for the internationalization of the RMB.

In this study, the USD-RMB and EUR-RMB exchange rates are selected for analysis due to the significance of the US dollar and the euro in the SDR basket of currencies, as well as the EU and the US being China's major trading partners. The daily chart of the USD-RMB exchange rate shows skewness and kurtosis values indicating a deviation from a normal distribution. Similarly, the EUR-RMB exchange rate does not follow a normal distribution.

The study employs a FEPA model, which undergoes training with a training set and a validation set. The model's performance is continuously adjusted using hyperparameters to achieve good results in the validation set. After confirming the hyperparameters, the model is trained again to minimize the global error between predicted and actual data. Finally, the test set is used for prediction and comparison.



The FEPA model demonstrates strong generalization ability and effectively predicts series volatility in financial price time series. The model's performance is compared with other models such as ARIMA, GARCH, WD-BPNN, LPP-BPNN, and EMD-BPNN using indicators like MDA, MAPE, RMSE, and DS.The study also examines the impact of the time factor on forex deep learning prediction results. The establishment of the FEPA-EMD model requires an effective FEPA model. Parameters are selected, and the model's applicability is verified through rolling predictions

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

based on true values. The FEPA-EMD model fits the trend of change, but deviations exist, with some predicted data being higher or lower than the true data.To address the non-stationarity of time series data, the first-order difference is often used to eliminate smoothness and extract relevant information. The differenced data exhibit fluctuations around 0, indicating a stable mean and the elimination of trends.In summary, the study utilizes the FEPA model to predict exchange rate fluctuations and demonstrates its effectiveness in capturing series volatility. The model's performance is evaluated and compared with other models, and the impact of the time factor on prediction results is analyzed.

V. CONCLUSION

In this paper, we introduce a novel FEPA (Fractional Empirical Paradigm Analysis) model based on deep neural networks for forecasting financial market trends. The model consists of three main parts: empirical modal decomposition, principal component analysis, and an artificial neural network for prediction. We apply this model to the CSI 300 stock index, Australian stock index, and foreign exchange rates, and validate its effectiveness using historical data.For nonlinear, nonstationary, and multiscale financial time series, the FEPA model combines the forward-rolling EMD (Empirical Mode Decomposition) method, principal component analysis, and neural networks to achieve a decomposition-reconstruction integration. The authors employ a rolling window approach with an appropriate window width to extract data segments. Each extracted segment undergoes EMD decomposition, and the resulting IMF (Intrinsic Mode Function) components are arranged in a matrix. Principal component analysis is then used to reduce the dimensionality by extracting the principal components that capture the essential information. These principal components are fed into the neural network for prediction.The CSI 300 Index, Australian stock index, and foreign exchange rates are chosen as test data to evaluate the model's performance. The empirical results demonstrate that the EMD decomposition algorithm, in combination with principal component analysis, effectively improves prediction performance. Principal component analysis helps compress redundant data, reduce training time, and enhance prediction accuracy. The FEPA model outperforms the EMD-BPNN (Backpropagation Neural Network) model, which, in turn, performs better than the single reference model.In summary, the FEPA model introduced in this paper leverages empirical modal decomposition, principal component analysis, and neural networks to forecast financial market trends. By combining these techniques, the model achieves improved prediction accuracy for nonlinear, nonstationary, and multiscale financial time series. The empirical results validate the effectiveness of the FEPA model across different markets, and the model's performance is further enhanced by incorporating an interval EMD model.

REFERENCES

[1] Yu Haishu, Cai Jihua, Xia Hong. Application of Arima Model in Stock Price Forecast [J]. Economist, 2015 (11): 156-157.

[2] Ariyo A A,AdewumiAO,AyoCK.Stock price prediction using the ARIMA model[C]//2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. 26-28 March2014, Cambridge, UK.IEEE, 2014:106-112.

[3] Shi Jia, Liu Wei, Feng Zhichao, et al. Analysis and Forecast of Stock Market Price Law based on ARIMA Model [J]. Statistics and applications, 2020, 9 (1): 101-114.

[4] Patel J,ShahS,ThakkarP,etal.Predicting stock market index using fusion of machine learning techniques[J].Expert Systems with Applications,2015,42(4):2162-2172.