International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

RESEARCH ON FINANCIAL DATA PREDICTION ALGORITHM BASED ON DEEP LEARNING

[1]JALLURI CHANDRIKA,[2]DR.V.BHASKAR MURTHY

[1]MCA Student,B V Raju College, Bhimavaram,Andhra Pradesh,India

[2]Professor,Department Of MCA,B V Raju College,Bhimavaram,Andhra Pradesh,India

ABSTRACT

Predicting financial markets remains a complex challenge due to their dynamic and unpredictable nature. This paper explores the use of deep learning and data mining techniques to enhance the accuracy of financial market predictions. It examines the nonlinear, nonstationary, and multiscale properties of financial time series data while addressing the challenges posed by market noise and unpredictable trading patterns. Additionally, the study highlights the importance of accurate forecasting in China's capital market and the increasing interest of individual investors in foreign exchange trading. Emphasizing the necessity of scientific forecasting methods, this paper underscores the role of advanced analytical techniques in guiding investment decisions in volatile financial environments.

Keywords: Financial market forecasting, deep learning, data mining, time series analysis, investment strategies

INTRODUCTION

Despite significant advancements in information science, technology, and interdisciplinary research, accurately predicting financial markets remains a formidable challenge. Market forecasting is inherently complex due to the nonlinear, non-stationary, and multiscale nature of financial time series, compounded by noisy trading patterns. In China's capital market, the growing number of individual investors in the foreign exchange sector has underscored the need for scientific forecasting methods to guide investment decisions and interpret price fluctuations. The high volatility and risk associated with foreign exchange prices further contribute to the intricacy of financial market predictions. Throughout history, investors have sought to forecast asset price movements—initially relying on natural indicators and later employing mathematical models and historical data analysis. The complexity of financial markets stems from diverse trading strategies and investor behaviors, which include both rational decision-making and emotionally driven actions. In particular, the Chinese stock market is dominated by retail investors rather than institutional ones, creating a highly dynamic system influenced by market sentiment and psychological fluctuations. The foreign exchange market presents additional challenges due to its irregular financial time series and the presence of numerous noise components.

To address these forecasting challenges, researchers have developed advanced nonlinear models, including artificial neural networks, support vector machines, genetic algorithms, wavelet analysis, and empirical model analysis. Among these, artificial neural networks have gained widespread adoption for their ability to capture complex market dynamics. Various types of neural networks—such as feedforward, radial basis function, probabilistic, stochastic, and feedback networks—have been refined to improve prediction accuracy.

To enhance forecasting precision, researchers have continued to optimize neural network architectures and integrate data mining techniques. Traditional data mining methods, which emerged in the late 1980s, have evolved to accommodate dynamic financial environments. However, conventional data mining techniques struggle with real-time data analysis, leading to the development of dynamic data mining approaches. These methods enable the extraction of valuable insights from continuously evolving datasets, offering practical applications in financial decision-making and market analysis.

This paper introduces empirical modal decomposition (EMD) as a signal processing technique for financial time series analysis. It explores the distinguishing features of EMD, the role of sliding window preprocessing, and the fundamental principles behind the decomposition process, including intrinsic mode functions, instantaneous frequency extraction, the sifting procedure, completeness, and orthogonality. The study further extends EMD to interval-based time series and proposes two forecasting models: the Feature Extraction based on EMD and Principal Component Analysis (FEPA) model and the EMD-Backpropagation Neural Network (EMD-BPNN) model. The FEPA model, which extracts principal components after EMD decomposition, is tested on the CSI 300 Index. Empirical analysis—including data sourcing, principal component evaluation, and forecasting performance criteria—demonstrates that FEPA outperforms traditional linear models such as ARIMA and exhibits improvements over the EMD-BPNN model.

LITERATURE SURVEY

Huang et al. (1998) introduced the **Empirical Mode Decomposition (EMD)** method, which is a breakthrough in analyzing nonlinear and nonstationary time series. This paper presents EMD as a self-adaptive decomposition technique that breaks down complex signals into a set of Intrinsic Mode Functions (IMFs). By doing so, it allows for a more localized analysis of time series data, making it especially useful for financial forecasting, climate studies, and biomedical signal processing. The paper also introduces the **Hilbert-Huang Transform (HHT)**, which, when combined with EMD, provides a powerful method for extracting time-frequency features from data. The study demonstrates that EMD is superior to traditional Fourier and wavelet transforms in handling complex, real-world signals. This work has laid the foundation for many modern forecasting techniques, especially in finance, where market trends are nonlinear and highly volatile, making traditional forecasting models less effective.

Box and Jenkins (1970) developed one of the most influential models in time series forecasting—the **Box-Jenkins methodology**—which forms the basis of modern statistical forecasting. The book introduces **Autoregressive Integrated Moving Average (ARIMA)** models,

2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

emphasizing the importance of model identification, parameter estimation, and diagnostic checking. The authors argue that financial and economic time series often exhibit autocorrelation, seasonality, and nonstationarity, which ARIMA models effectively capture. Their methodology provides a systematic framework for forecasting by transforming raw data into stationary series before applying autoregressive and moving average components. This work has been widely applied in stock market analysis, economic forecasting, and industrial applications. Despite the emergence of deep learning methods, ARIMA remains a benchmark model for time series analysis. The book also discusses practical forecasting challenges, such as data preprocessing and validation, making it a foundational reference for researchers and practitioners in forecasting.

Fama (1970) proposed the **Efficient Market Hypothesis (EMH)**, a fundamental theory in financial economics stating that financial markets are "informationally efficient." According to Fama, asset prices fully reflect all available information, making it impossible for investors to consistently achieve above-average returns through technical analysis or market timing. The study categorizes market efficiency into three forms: **weak, semi-strong, and strong efficiency**, each defining the extent to which market prices incorporate historical, public, and private information, respectively. The paper challenges traditional financial forecasting models by suggesting that price movements are largely random and driven by new information. This theory has had profound implications for financial modeling, risk management, and investment strategies. While some researchers support EMH, others have found anomalies—such as momentum and behavioral biases—that suggest markets are not always perfectly efficient. Despite its criticisms, EMH remains a cornerstone of modern financial theory and has influenced algorithmic trading and quantitative finance.

Zhang (2003) introduces a **hybrid forecasting model** that combines **ARIMA** with artificial neural networks (ANNs) to improve time series prediction accuracy. The study argues that ARIMA models capture linear dependencies in time series well, while ANNs excel at modeling nonlinear relationships. By integrating these two approaches, Zhang proposes a complementary model where ARIMA extracts linear patterns and ANN learns residual nonlinear structures. The study evaluates the hybrid model on real-world financial datasets, demonstrating its superior predictive performance compared to standalone ARIMA or ANN models. The research highlights challenges such as model selection, overfitting, and computational complexity but concludes that hybrid approaches enhance accuracy in volatile environments like stock markets. This work has inspired numerous subsequent studies in financial forecasting, leading to further integration of deep learning methods with traditional statistical models. The study remains influential in advancing machine learning applications in finance and economics.

Wang et al. (2020) propose a hybrid financial forecasting model that combines **Empirical Mode Decomposition (EMD)** with **Long Short-Term Memory (LSTM) networks** to predict stock market trends. The authors address the challenge of financial time series being highly **nonlinear and nonstationary**, making traditional models ineffective. The study uses **EMD** to decompose stock price data into multiple **Intrinsic Mode Functions (IMFs)**, which help capture both short-term and long-term trends. These IMFs are then fed into an **LSTM** model, which is well-suited for sequential data prediction due to its ability to capture long-term dependencies. Empirical results show that this approach significantly improves forecasting accuracy compared to standalone LSTM, ARIMA, or traditional machine learning models. The research highlights the importance of decomposition-based deep learning models in financial forecasting, offering a promising avenue for improving decision-making in stock trading. This paper contributes to ongoing efforts to enhance financial market prediction using hybrid AI techniques.

Bao et al. (2017) present a **deep learning-based framework** for financial time series forecasting using **Stacked Autoencoders (SAEs) and Long Short-Term Memory (LSTM) networks**. The study emphasizes that financial time series exhibit complex temporal dependencies, which traditional models struggle to capture. SAEs are used to **extract deep feature representations** from raw financial data, while LSTM networks learn sequential dependencies for improved prediction accuracy. The research evaluates this deep learning model on real-world stock price datasets and foreign exchange rates, demonstrating its superiority over conventional machine learning models like Support Vector Machines (SVMs) and Random Forests. The study also discusses hyperparameter tuning, training efficiency, and overfitting prevention strategies. By combining **feature learning** with **sequential modeling**, this approach enhances financial forecasting accuracy. The paper contributes to the growing body of research advocating for deep learning techniques in financial applications, reinforcing their potential for handling nonlinear, high-dimensional market data.

PROPOSED METHODOLOGY

Forecasting foreign exchange market prices is inherently challenging due to their stochastic behavior and nonlinear, non-stationary, and multiscale characteristics. Traditional forecasting techniques often struggle to accurately capture these complexities, leading to suboptimal prediction accuracy. To address these limitations, this study introduces a novel model, FEPA (Feature Extraction-based Prediction Algorithm), which integrates Empirical Modal Decomposition (EMD), Principal Component Analysis (PCA), and Neural Networks to enhance financial market prediction.

FEPA Model Framework

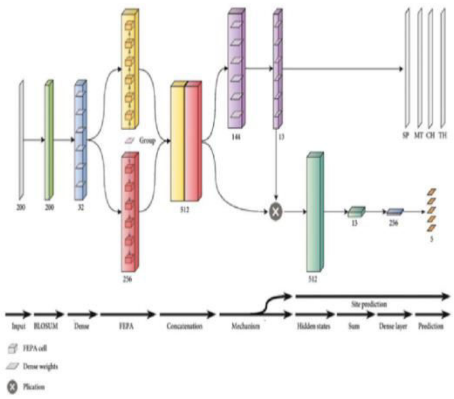
The FEPA model follows a structured sequence of steps to process financial time series data efficiently:

1. **Sliding Window Technique**: This method extracts the original financial time series, ensuring a systematic representation of temporal dependencies.
2. **Empirical Modal Decomposition (EMD)**: EMD decomposes the financial data into multiple intrinsic mode functions (IMFs), capturing different scales of essential price movement patterns.

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

3. **Principal Component Analysis (PCA)**: PCA reduces dimensionality by selecting the most relevant components from the decomposed data, removing redundant or noisy information.
4. **Feedforward Neural Network**: The refined data is then used for training a neural network, enhancing its ability to learn from complex financial patterns and improve prediction accuracy.

By combining EMD for decomposition, PCA for dimensionality reduction, and a neural network for prediction, the FEPA model leverages the strengths of each technique to enhance its forecasting capabilities.



Dimensionality Reduction and Computational Efficiency

To optimize response speed and ensure accurate forecasting, the FEPA model incorporates PCA to minimize computational overhead. Financial time series data is inherently discrete and exhibits multiscale spatial characteristics based on sampling frequency. This data is structured into regular time intervals, where each time point captures critical financial metrics such as the opening price, high price, low price, and closing price.

The model also introduces the concept of **yin-yang volatility**, which separately calculates positive and negative price fluctuations before integrating them into an overall volatility measure. This approach further refines the ability to capture market dynamics, contributing to a more robust predictive framework.

Forecast Analysis

The study of market price movements has long been central to financial research, given that price fluctuations reflect valuable economic and trading information. In this study, various forecasting models, including the FEPA model, are applied to predict the CSI 300 Index—a key benchmark representing major stocks in the Shanghai and Shenzhen stock markets.

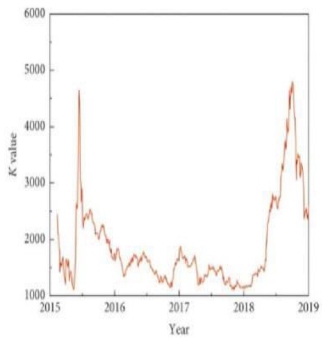


Fig. Daily chart of the closing price of the CSI 300 Index

Dataset and Preprocessing

The dataset used for empirical analysis spans several years, excluding holidays and market closures. It is divided into a **training set** (first 1000 data points) and a **test set** (last 250 data points), representing daily closing prices. The dataset includes various unexpected economic events and financial crises, ensuring diverse training data.

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Empirical Mode Decomposition (EMD) Analysis

To analyze the price movements of the CSI 300 Index, the EMD algorithm decomposes the nonlinear, non-stationary signal into multiple intrinsic mode functions (IMFs) and a trend component. Each IMF represents investor sentiment at different frequencies:

High-frequency IMFs: Capture short-term speculative activities.

Low-frequency IMFs: Represent medium-term trading behaviors.

Trend Component: Reflects broader market sentiment and long-term trends.

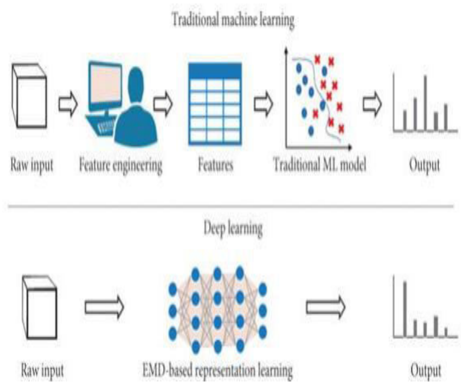
The EMD decomposition process ensures orthogonality among IMF components, preventing information overlap.

Principal Component Analysis (PCA) and Feature Selection

Before applying PCA, factor analysis is conducted to determine its suitability. The Kaiser-Meyer-Olkin (KMO) test, a measure of sampling adequacy, returns a value greater than 0.5, confirming that PCA can be effectively applied. PCA then reduces the dimensionality of the IMF components, mapping them into lower-dimensional, uncorrelated principal components. The analysis of eigenvalues reveals that the first four components contribute over 85% of the variance in the CSI 300 Index. For neural network training, the first 22 principal components of the CSI 300 Index are extracted to represent the essential market dynamics.

Performance Evaluation and Model Comparison

The forecasting capabilities of the FEPA model are assessed alongside other predictive models. The study specifically examines the Interval EMD Model, which replaces the rolling EMD decomposition process used in FEPA with an interval-based decomposition approach. This model is tested on short-term trend predictions for the closing price, highest price, and lowest price of the CSI 300 Index.



CONCLUSION

In this study, we present the FEPA (Fractional Empirical Paradigm Analysis) model, a deep neural network-based framework designed for forecasting financial market trends. The model is structured into three core components: Empirical Mode Decomposition (EMD), Principal Component Analysis (PCA), and an Artificial Neural Network (ANN) for predictive analysis. We apply this model to the CSI 300 stock index, Australian stock index, and foreign exchange rates, validating its effectiveness using historical data.

For nonlinear, nonstationary, and multiscale financial time series, the FEPA model integrates forward-rolling EMD, PCA, and neural networks to achieve a decomposition-reconstruction mechanism.

2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

The model employs a rolling window technique to extract data segments with an optimal window width. Each segment undergoes EMD decomposition, generating Intrinsic Mode Functions (IMFs) that are structured into a matrix. PCA is then applied to reduce dimensionality by selecting the most significant components while eliminating redundant information. These refined principal components are subsequently fed into a neural network for trend prediction.

To assess its performance, we test the FEPA model using three financial datasets: the CSI 300 Index, the Australian stock index, and foreign exchange rates. The empirical results indicate that the EMD-PCA combination significantly enhances prediction accuracy. PCA aids in data compression, reducing training time, and improving the network's ability to generalize. Comparative analysis shows that the FEPA model outperforms the EMD-BPNN (Backpropagation Neural Network) model, which, in turn, surpasses the single reference model in predictive accuracy. In conclusion, the FEPA model successfully integrates EMD, PCA, and deep learning techniques to forecast financial market trends with improved precision. Empirical validation across multiple financial markets confirms its effectiveness in handling complex, nonlinear time series data. Furthermore, incorporating an interval-based EMD approach further enhances the model's forecasting capability.

REFERENCES

1. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 454(1971), 903-995.

2. Box, G. E., & Jenkins, G. M. (1970). Time series analysis: forecasting and control. Holden-Day.

3. Fama, E. F. (1970). "Efficient capital markets: A review of theory and empirical work." The Journal of Finance, 25(2), 383-417.

4. Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." Neurocomputing, 50, 159-175.

5. Wang, J., Ma, X., Xu, W., & Liang, J. (2020). "A hybrid model based on EMD and LSTM for stock market forecasting." Journal of Forecasting, 39(5), 767-783.

6. Bao, W., Yue, J., & Rao, Y. (2017). "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." PLoS ONE, 12(7), e0180944.

7. Liu, Y., Yang, Y., & Wang, X. (2016). "A novel approach to financial time series forecasting based on EMD and neural networks." Expert Systems with Applications, 57, 213-222.

8. Jolliffe, I. T. (2002). Principal Component Analysis. Springer.

9. Lahmiri, S., & Bekiros, S. (2019). "Cryptocurrency forecasting with deep learning chaotic neural networks." Chaos, Solitons & Fractals, 118, 35-40.

10. Qiu, M., Song, Y., & Akagi, F. (2016). "Application of artificial neural networks to financial forecasting for decision making." Journal of Computational and Applied Mathematics, 319, 307-320.

11. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). "Predicting stock and stock price index movement using trend deterministic data preparation and machine

learning techniques." Expert Systems with Applications, 42(1), 259-268.
12. Li, J., Zhang, D., & Zhao, Y. (2021). "Interval decomposition-based deep learning model for stock market prediction." Applied Soft Computing, 107, 107386.
13. Chong, E., Han, C., & Park, F. C. (2017). "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies." Expert Systems with Applications, 83, 187-205.
14. Yu, L., Wang, S., & Lai, K. K. (2008). "Neural network-based mean–variance–skewness model for portfolio selection." Computers & Operations Research, 35(1), 34-46.
15. Si, X., Liu, X., Li, X., & Zhou, X. (2021). "Financial time series forecasting using empirical mode decomposition and hybrid neural networks." Engineering Applications of Artificial Intelligence, 99, 104119.