Pamukkale Üniversitesi

Sosyal Bilimler Enstitüsü Dergisi

Pamukkale University Journal of Social Sciences Institute

ISSN 1308-2922 E-ISSN 2147-6985

Article Info/Makale Bilgisi
√Received/Geliş:01.12.2023 **√Accepted/Kabul:**02.04.2024
DOİ:10.30794/pausbed.1398790
Research Article/Araştırma Makalesi

Akusta, A. (2024). "Exploring CEEMDAN Decomposition for Improved Financial Market Forecasting: A Case Study on Dow Jones Index", *Pamukkale University Journal of Social Sciences Institute*, Issue 62, Denizli, pp. 19-35.

EXPLORING CEEMDAN DECOMPOSITION FOR IMPROVED FINANCIAL MARKET FORECASTING: A CASE STUDY ON DOW JONES INDEX

Ahmet AKUSTA*

Abstract

This study presents an innovative financial time series analysis approach by integrating Complete Ensemble Empirical Mode Decomposition (CEEMDAN) with the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model. The primary contribution of the research lies in significantly enhancing the predictive accuracy and understanding of the dynamics governing major stock indices. CEEMDAN adaptively decomposes complex financial time series into intrinsic mode functions (IMFs), a technique that has yet to be extensively utilized in this domain. IMFs are combined with ARIMAX's predictive proficiency, which accounts for the influence of historical trends and external factors. Our study showcases an R^2 of 0,93, aligning with some of the high-performing models in the literature. However, the unique strength of our model lies in its lag-free predicting of the DJIA, effectively mirroring its volatility and major movements with high fidelity, making it highly practical for financial applications.

Keywords: *Financial time series decomposition, ARIMAX modeling, Financial market forecasting, CEEMDAN.*

DOW JONES ENDEKSİNİN İLERİ ZAMAN SERİSİ ANALİZİ: CEEMDAN AYRIŞTIRMASI KULLANILARAK YAPILAN KAPSAMLI BİR ÇALIŞMA

Öz

Bu çalışma, Tam Topluluk Ampirik Mod Ayrıştırması (CEEMDAN) ile Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) modelini entegre ederek yenilikçi bir finansal zaman serisi analizi yaklaşımı sunmaktadır. Araştırmanın birincil katkısı, büyük hisse senedi endekslerini yöneten dinamiklerin tahmin doğruluğunu ve anlaşılmasını arttırmaktır. Daha önce bu alanda kapsamlı bir şekilde kullanılmayan CEEMDAN, karmaşık finansal zaman serilerini uyarlamalı olarak içsel mod fonksiyonlarına (IMF'ler) ayrıştırmak için yenilikçi bir şekilde uygulanmıştır. CEEMDAN'ın karmaşık finansal zaman serilerini uyarlanabilir bir şekilde IMF'lere ayrıştırma yeteneği, ARIMAX'ın tarihsel eğilimlerin ve dış faktörlerin etkisini hesaba katan tahmin yeterliliği ile birleştirilmiştir. Metodoloji, çeşitli büyük ABD hisse senedi endekslerini dışsal değişkenler olarak içeren kapsamlı Dow Jones Endüstriyel Ortalama (DJIA) analizi ile doğrulanmıştır. Çalışmamız, literatürdeki yüksek performanslı modellerle uyumlu olarak 0,93'lük bir R^2 skoru sunmaktadır. Bununla birlikte, modelimizin benzersiz gücü, DJIA'nın gecikmesiz tahmininde yatmaktadır. Endeksin volatilitesini ve önemli hareketlerini yüksek doğrulukla yansıtarak finansal uygulamalar için son derece pratik hale getirmektedir.

Anahtar Kelimeler: *Finansal zaman serisi ayrıştırması, ARIMAX modelleme, Finansal piyasa tahmini, CEEMDAN.*

*Lecturer Dr., Konya Technical Univeristy, Rectorate, KONYA.
e-mail:ahmetakusta@hotmail.com, (https://orcid.org/0000-0002-5160-3210)

1. INTRODUCTION

The evolution from traditional statistical methods to machine learning (ML) and hybrid ML techniques marks a significant advancement in financial forecasting. This shift, driven by the increasing complexity and volume of financial data, necessitates more sophisticated analysis techniques. Makridakis et al., (2018) underscore the potential of ML methods to surpass traditional approaches by streamlining the labor-intensive process of model selection and reducing forecasting errors. This enhancement in the accuracy and efficiency of financial forecasts lays the groundwork for exploring specific machine learning and deep learning techniques.

Building on this foundation, applying Convolutional Neural Networks (CONV1D) with Long Short-Term Memory (LSTM) networks and signal processing have shown promising results in predicting financial market movements. Studies by Wang et al., (2021) and Altan et al., (2019) demonstrate the effectiveness of these techniques in capturing the complex patterns within financial time series data, further emphasizing the potential of ML in financial forecasting. These advancements suggest a paradigm shift towards adopting more nuanced analytical frameworks to understand market dynamics better.

In particular, the superiority of LSTM models in financial markets has been highlighted by Alzheev and Kochkarov, (2020) and Gao, (2021) who note their ability to outperform traditional ARIMA models significantly. Alzheev and Kochkarov's research, focusing on the stock prices of Russian companies, and Gao et al.'s discussion on Seq2Seq models introduce a layer of complexity by incorporating attention mechanisms. These developments underline a broader trend of leveraging advanced computational techniques to refine predictive accuracy in financial forecasting.

However, Alp et al., (2021) present a contrasting viewpoint by affirming the applicability and effectiveness of ARIMA in specific contexts, such as predicting BIST price indices. This serves as a reminder of the continued relevance of traditional statistical methods, highlighting that deep learning models may not necessarily offer a clear advantage in specific scenarios. This nuanced perspective encourages a balanced consideration of traditional and modern financial forecasting approaches.

Further expanding the discussion, Erden, (2023) explores an array of deep learning models, including LSTM, GRU, and RNN, showcasing the robustness of RNN in predicting financial time series with high accuracy levels. This comprehensive approach emphasizes the importance of exploring various models to identify the most effective techniques for different financial forecasting applications.

Exploring machine learning techniques extends beyond financial markets into areas like energy consumption forecasting. The studies by Saranj and Zolfaghari, (2022) and Cai et al., (2019) delve into innovative applications of hybrid models, such as the combination of AWT-LSTM with ARIMAX-GARCH and the use of a gated 24-h CNN model. These studies showcase the potential of leveraging machine learning for forecasting in sectors with distinct challenges, such as fluctuating energy demands in commercial buildings and sports venues.

Introducing Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) into the mix marks a significant innovation in time series analysis. The versatility and effectiveness of CEEMDAN, as proven across various domains, including solar radiation prediction and agricultural forecasting, signify a leap forward in our ability to analyze and predict market dynamics with greater precision. The integration of CEEMDAN with deep learning techniques, such as LSTM, and its combination with methods like Wavelet Decomposition and N-BEATS illustrate the developing potential of hybrid models in financial forecasting.

Against this backdrop of methodological advancements, the Dow Jones Industrial Average (DJIA) presents a unique challenge due to its volatile nature. The complexity of stock market data, characterized by non-linear and non-stationary features, often eludes traditional prediction models. Our study introduces an innovative approach to synthesizing CEEMDAN with ARIMAX models to address these challenges. This methodology, which leverages the strengths of CEEMDAN for time series decomposition and incorporates exogenous variables into the forecasting model, aims to enhance the predictability of financial time series, particularly the DJIA.

The practical implications of this research are profound. By translating complex models into actionable insights, we aim to bridge the gap between theoretical advancements and their application in financial markets. This study contributes to the academic literature on financial time series analysis and offers valuable strategies for market participants seeking to navigate the complexities of financial forecasting.

2. LITERATURE REVIEW

The advent of big data and advances in computational technology have provided traders and investors with many different ways of looking at the stock market. Compute power for just a couple hundred dollars a month. Good computing power is now available for just a couple hundred dollars a month. In recent years, the availability of intraday stocks, options, futures, and foreign exchange data has increased dramatically. Recently, the concept of "big data" has become popular. In short, big data is a large data set that cannot be managed using a traditional database, so a new data management method is necessary.

The literature reflects this evolution, with studies delving into various aspects of stock market prediction, ranging from clustering and unsupervised learning to advanced time-series analysis and integrating computational methods with traditional financial theories.

An introduction and background to clustering in general is given by (Aghabozorgi et al., 2015), showing the importance of clustering as a significant technique for categorizing large amounts of data, including big data and cloud computing. The authors mentioned that unsupervised approaches like clustering algorithms are vital in data deluge and are highly applied in time series data from various scientific domains. Gamboa (2017) gives an insight into unsupervised learning and its impact on data analysis. Gamboa has listed deep learning methods in unsupervised learning and has analyzed novel time series approaches. He concluded that deep learning methods with unsupervised feature learning significantly impact venues with high time dependencies.

There is a rich literature on time series and forecasting. Foreseeing the intrinsic worth of observational studies in connection with Randomized Controlled Trials (RCTs), Kontopantelis et al., (2015) extol the virtues of Interrupted Time Series analyses, in particular, where an RTC may be impractical. The potential of investigating longitudinal datasets in "real-world" data analysis is unlocked through this approach. Working in a similar vein, Wang and Wang, (2015) develop a stochastic time-effective function neural network combined with principal component analysis for dynamic forecasting in financial markets; the empirical analysis demonstrates the superiority of this model over traditional neural networks with economic time series forecasting. (Alaoui et al., (2015) employed wavelet transform to reveal the Islamic indices' co-movement dynamics that offer new investment opportunities for the investors, along with new views of market risk and correlations. Then, Sen and Chaudhuri, (2016) gave topics time series forecasting by analyzing and predicting stock market data using an effective decomposition and forecasting approach. Secondly, Siami-Namini et al., (2018) studied the superiority of traditional time series forecasting models, like ARIMA, with the help of deep learning techniques, particularly LSTM, and showed significantly that it outperforms. Lastly, Cao et al., (2019) recommend implementing the hybrid forecasting models consisting of the different variants of EMD with LSTM to improve the stock market price forecasting accuracy.

A noteworthy development noted in the literature is the application of advanced computational techniques with traditional financial models. Bao et al., (2017) propose a novel deep-learning framework that combines wavelet transforms and stacked autoencoders with extended short-term memory networks to forecast stock prices and achieve better predictive accuracy and profitability, andNti et al., (2020) conducted a systematic review of machine learning-based stock market prediction studies, whereby support vector machines and artificial neural network algorithms are the most commonly employed. Baek et al., (2020) employ the MS-AR (1) model to analyze the changes in the United States stock market volatility after the outbreak of COVID-19, using feature selection methods to identify effective financial indicators that reflect the increased industry risk and the sensitivity to the pandemic.

Ayala et al. (2021) propose a hybrid approach that uses technical indicators and machine learning (ML). This technique will help to predict the stock market. Many deep learning models were tested for the best trading decision-making. Using the Diebold-Yilmaz connectedness index, Bahloul and Khemakhem, (2021) analyzed the dynamics between commodity returns and Islamic market indices. This research will explain the market impact during COVID-19 shocks by comparing the Commodity indices and Islamic indices; commodities indices seemed to be the most crucial stock, followed by Islamic indices in terms of returns and volatilities. Althelaya et al., (2021) proposed a deep neural network model utilizing multiresolution analysis. Eventually, the model proposed in this research could track the S&P500 index outperforming the traditional techniques.

Future research trends and emerging directions in modeling stock markets and prediction can leverage several studies Shah et al., (2019) and Gandhmal and Kumar, (2019) to combine other methods with machine-learning techniques to enhance the accuracy of the predictions. Costola et al., (2023), Zhao et al., (2023), and Noh and Park, (2023) emphasize the inclusion of context-based factors such as News Sentiment, time-lag relationship, and climate change risks as part of stock market analysis in order to improve the prediction, while Dua and Tuteja, (2023) and (Kashyap, (2023) will bring out the importance of prediction models that have the flexibility to adapt to the ever-changing market conditions and structural reforms, especially during a financial crisis.

3. DATA ANALYSIS

Data analysis is a crucial process in extracting meaningful insights from data. It requires employing a variety of approaches and strategies to analyze and summarize data in order to find correlations, patterns, and trends.

3. 1. Description of the Dataset

The chosen dataset for this study provides a broad overview of the performance of the three leading stock indices in the US. The dataset taken from Yahoo Finance captures the period from 2018 to 2023, before and during the market fluctuations caused by the pandemic and the stages during the economic recovery.

The dataset consists of the closing values for the following stock market indices: the Dow Jones Industrial Average (DJIA), the NASDAQ Composite (IXIC), the NYSE Amex Composite Index (XAX), the S&P 500 (GSPC), the Russell 2000 (RUT), the NYSE Composite (NYA) and the CBOE Volatility Index (VIX). These indices were chosen as they encompass a range of market segments, from blue-chip corporates to technology-focused companies, small-cap firms, and volatility measures.

Table 1 summarizes the key variables and their respective statistical descriptors:

Table 1. Descriptive Statistics

Statistic	DJIA	IXIC	VIX	XAX	GSPC	RUT	NYA
Mean	29778,4	10796,5	20,8299	3070,62	3569,77	1769,6	14212,9
75%	33821,3	13314,3	24,16	3985,88	4181,17	1951,33	15731,2
25%	25985,2	7964,24	15,57	2471,89	2888,21	1557,24	12748,4
50% (Median)	29823,9	11085,2	19,11	2714,93	3647,49	1717,47	14097,3
Min	18591,9	6192,92	10,85	1317,22	2237,4	991,16	8777,38

3. 2. Initial Data Analysis

The seasonal decomposition of the DJIA presented here is a time series analysis that decomposes the original data into three distinct components: trend, seasonal, and residual components.

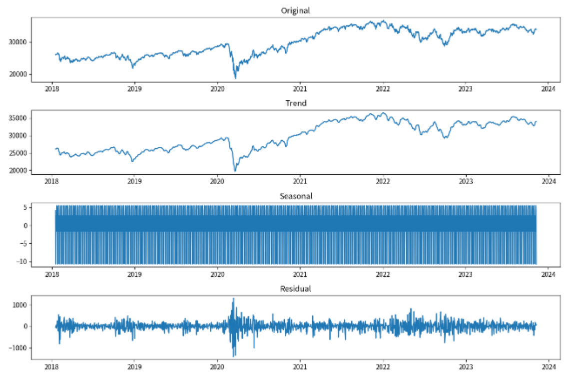

Figure 1. Seasonal Decomposition Results

In Figure 1, the first plot displays the DJIA's actual values from early 2018 to late 2023, highlighting notable fluctuations that hint at economic trends and market shifts. The second plot, showing the extracted trend, smooths out these fluctuations to reveal the Index's overarching movement, initially declining and then steadily rising, possibly reflecting market corrections and economic growth. The third plot focuses on the seasonal component, illustrating consistent, predictable patterns that suggest the Index's annual cyclical influences. Lastly, the residual plot captures the randomness left after removing trend and seasonal elements, marked by volatility indicating periods of market instability or unforeseen events.

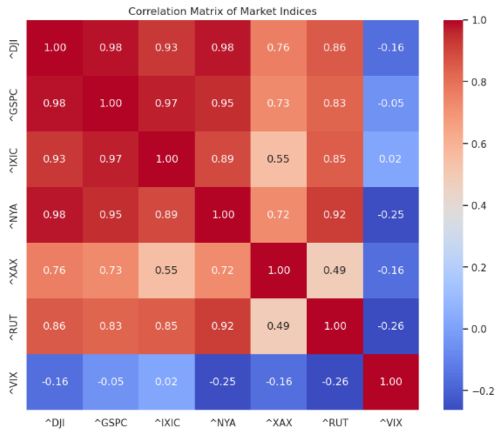

Figure 2. Correlation Matrix

The A correlation matrix of the leading U.S. stock indices is displayed in Figure 2, highlighting the relationships between them. The S&P 500, NASDAQ Composite, NYSE Composite, NYSE Amex Composite Index, and Russell 2000 all show favorable correlations with the DJIA. The indices typically move in the same direction as these positive correlations indicate.

The VIX (Volatility Index) has a negative correlation with all of the indices. The VIX has the opposite of the equation of a standard stock index. Developed by the Chicago Board Options Exchange (CBOE), the VIX is a gauge of market volatility based on implied options movements rather than historical data. It is recognized by many as the foremost indicator of volatility and investor sentiment (Shaikh and Padhi, 2015).

3. 3. Advanced-Data Analysis Using CEEMDAN

In this analysis of the price action of the DJIA, we utilize the novel CEEMDAN algorithm to decompose time series into constituent Intrinsic Mode Functions (IMFs), each established at different frequency ranges. Compared to EEMD, another popular algorithm for decomposing time series, CEEMDAN orders these IMFs from high to low frequency, with the original data being more simplified (Ban and Shen 2022).

Through this decomposition, we aim to gain a more profound and intricate comprehension of the market's intrinsic dynamics, paving the way for more informed and strategic forecasting endeavors.

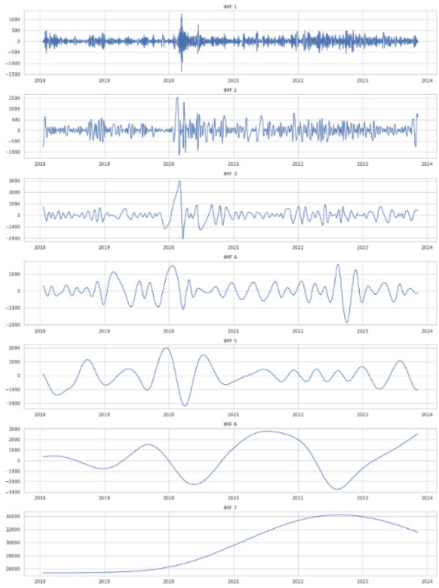

Figure 3. Ceemdan Decomposition Results

Figure 3 shows the first IMF, which captures the highest frequency oscillations in the DJIA data, representing noise or the Index's most rapid changes. This stems from the market's short-term volatility and may be attributable to daily trading activities, news events, or immediate market reactions to economic data releases.

The following IMFs (IMF2, IMF3) likely represent higher frequency market cycles or intermediate-term fluctuations. These may be associated with short-term investor sentiment cycles, trading strategies, or impacts of quarterly financial reporting.

IMFs 4 and 5 may pick up intermediate-term market cycles. These may reflect the impact of policy changes, business cycles, or seasonal tendencies in investment behavior.

IMF 6 exhibits even lower frequency components. These could represent long-term investment cycles or significant economic trends over five years. This IMF might show macroeconomic trends such as expansion or contraction phases.

The seventh IMF, displaying the lowest frequency, trends upwards consistently over the period. This trend component encapsulates the long-term movement in the DJIA, abstracting from cyclical and irregular behaviors. It could denote sustained growth or decline in the market, driven by long-standing economic fundamentals, structural changes in the economy, or persistent trends in corporate earnings growth.

3. 4. Lags

The analysis generated ten lagged instances of the DJIA's daily closing values. These lags, sequentially denoted from Lag 1 to Lag 10, represent the series of closing prices offset by one to ten trading days, respectively. Given their significance to the model, the original ten lags were narrowed to three.

4. METHODOLOGY

Python is the main programming language used in this study, and it is known for its adeptness in data processing. The flexibility and ease of use of Python make it ideal for the complex requirements of time series analysis. With this approach Python's libraries—Pandas(McKinney et al., 2010) and NumPy (Harris et al., 2020) for data handling, EMD-signal (Laszuk, 2017) for CEEMDAN decomposition, and Statsmodels (Seabold and Perktold,

2010) for ARIMAX modeling are used. They all streamline the workflow from data collection to sophisticated analysis.

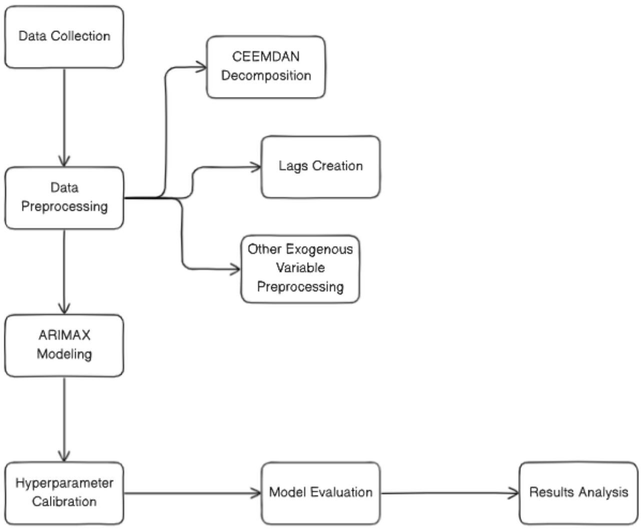

Figure 4. Methodology Workflow

Data pre-processing is the next step, as seen in Figure 4. During the data pre-processing step, the raw data is filtered and changed. It ensures the quality and integrity of the data for robust analysis and removes any discrepancies that could distort the analytical outcomes.

Once the data have been refined, the workflow branches into two parallel streams: CEEMDAN and Exogenous variables pre-processing. The CEEMDAN decomposes the time series into Intrinsic Mode Functions (IMF) to reveal the underlying frequency components of the multicomponent time series data. Meanwhile, the Exogenous Variables Pre-processing step ensures the exogenous variable preparation process and makes the data manipulation used in the predictive model.

At the core of the analysis, we employ ARIMAX modeling, which incorporates both the intrinsic elements derived from CEEMDAN decomposition and the carefully prepared exogenous variables. This amalgamation is paramount because the DJIA is influenced by events worldwide, predicated on economic directives and respective market sentiments; this amalgamation includes a peak into the future while deciphering the causal effects of influential spikes (roughly), yielding the most accurate results.

Through this modeling, the results will be hyperparameter calibrated to get the best possible fit for the financial time series data. This calibration is essential to make the model increasingly correct in its predictions.

Model Validation is a critical juncture in the workflow and rigorously examines the model's performance against a set of metrics to verify its predictive accuracy and reliability. The model's predictive capacity is substantial compared to the basic ARIMA and LSTM models.

4. 1. ARIMAX Model Setup and Exogenous Variables

The methodological framework implemented in this study is predicated on the fusion of the CEEMDAN and the ARIMAX to predict the DJIA.

Integrating the ARIMAX model within financial analytics is instrumental for multivariate time series prediction, demonstrating its efficacy in various domains (Adekanmbi, 2017). ARIMA models are significant in prediction models since they have been the focus of time series as they depend on the series' past. ARIMAX models are almost identical to ARIMA models, but ARIMAX includes exogenous inputs, which helps us deal with endogeneity (Liu et al., 2020).

In the model, ARIMAX, the serialization in the (AR) model proposes that the value series in the current is a function in past values series. The (MA) aspect of the model proposes that the forecast error becomes a linear collection of the last errors. Integration (I) transforms are the purpose of differencing the data series to change its stationery series (Kotu and Deshpande, 2019). Since the ARIMA model combines these three aspects, it can fit past models efficiently and predict future points well.

Due to their robustness and efficiency in handling financial time series, the ARIMA and ARIMAX models have garnered attention for their application in financial markets, particularly for short-term stock price forecasting (Adebiyi et al., 2014). Moreover, the adaptability of ARIMA models has been evidenced in forecasting the dynamics of epidemiological data, such as cumulative COVID-19 cases, showcasing their capacity to model complex patterns across disparate data (Benvenuto et al., 2020).

The ARIMAX model's capacity to incorporate external predictors, lags, and IMFs created by CEEMDAN Analysis augments its forecasting accuracy. The model utilizes the following final dataset as exogenous variables:

Table 2. Variables

Target Variable	Dow Jones Industrial Average
Exogenous Variable	S&P 500 (^GSPC), NASDAQ Composite (^IXIC), New York Stock Exchange Composite (^NYA), AMEX Composite (^XAX), Russell 2000 (^RUT), and the CBOE Volatility Index (^VIX), Lags, IMFs

Table 2 lists all of the exogenous variables that are carefully incorporated into the ARIMAX framework in order to precisely estimate the trajectory of the DJIA in the future. The stock market's temporal dependencies and possible autoregressive features are acknowledged by including lagged variables. The DJIA can be dissected on multiple scales using the IMFs obtained by CEEMDAN, which can capture oscillatory patterns ranging from short-term noise to long-term trends. With the use of intricate relationships with associated market indices, this integrative method seeks to uncover hidden structures in the time series in order to create a reliable forecasting model.

4. 2. Methodological Choices: Why CEEMDAN and ARIMAX are Suitable for This Study

The methodological underpinning of this study is predicated on the integration of CEEMDAN and ARIMAX. CEEMDAN advances empirical mode decomposition (EMD) by solving its mode-mixing issues, offering a refined approach for signal analysis that is particularly effective for complex, non-linear data like financial time series (Yang et al., 2017).

Unlike traditional Fourier or wavelet analysis, EMD adaptively decomposes signals into intrinsic mode functions (IMFs) without relying on pre-established bases, ensuring an authentic data representation (Huang and Yao, 2014). Empirical Mode Decomposition (EMD) has proven to be an effective methodology for noise reduction in time series analysis, enabling the adaptive decomposition of signals into intrinsic mode functions (IMFs) and a residual component. This process allows for a detailed extraction of the signal's oscillatory behaviors, effectively minimizing noise within the dataset (Chui and He, 2024). The effectiveness of EMD, and by extension of CEEMDAN, in addressing noise-related issues underscores its significance in refining signal analysis and laying the groundwork for the development of more sophisticated decomposition techniques.

CEEMDAN builds upon ensemble EMD (EEMD), which introduces noise to alleviate mode-mixing by strategically adding noise at each decomposition stage, achieving more precise results without reconstruction errors, thus providing more apparent component features (Colominas et al., 2012; Singh et al., 2021).

This study employs CEEMDAN's decomposition capabilities with ARIMAX's predictive modeling to analyze the financial time series. The integration leverages the strengths of both methods—CEEMDAN's adaptive decomposition and ARIMAX's consideration of historical and external factors—to offer a tailored approach to time series forecasting (Li et al., 2018).

In addition to analyzing financial time series, CEEMDAN has already shown its versatility in many fields. It was used to forecast temperatures (Chen et al., (2022) and for hydrological studies (Zhang et al., 2021), but it was also used in combination with deep learning models such as LSTM to improve forecasting accuracy (Wu et al., 2022).

In hydrology, CEEMDAN helps extract meaningful information from complex hydroclimatic time series data, aiding in understanding and predicting hydrological patterns. It has also been employed in fault diagnosis, seismology, traffic flow prediction, and medicine, showcasing its adaptability and versatility (Zhai et al., 2020).

Furthermore, the CEEMDAN technique allows for further flexibility. It has been integrated with other techniques, such as Support Vector Regression (SVR) and variational mode decomposition (VMD), to create hybrid forecasting models for power load and runoff.

4. 3. Rationale Behind Selection of Hyperparameters

The selection of hyperparameters for the ARIMAX model was a meticulous process guided by several analytical techniques and statistical measures. Table 3 shows Stationarity Test Statistics.

Table 3. Stationarity Test Statistics

Metric	Original Series	First Difference
ADF Statistic	-1,635646	-11,89877
p-value	0,464537	5,64E-22
Critical value (1%)	-3,434884	-3,434884
Critical value (5%)	-2,863542	-2,863542
Critical value (10%)	-2,567836	-2,567836

In the pre-modeling phase, the stationarity analysis was conducted to inform the specification of the ARIMAX model, particularly in determining the order of differencing (d). The Augmented Dickey-Fuller (ADF) test was employed to assess the stationarity of the DJIA series.

The ADF test statistic for the original time series of the DJIA index was computed to be -1,681318. This value does not surpass the critical threshold at any conventional significance level, remaining within the bounds of the established critical values. Moreover, the p-value associated with the test is 0,46, considerably exceeding the alpha level of 0,05. This outcome does not allow for rejecting the null hypothesis, which posits the existence of a unit root within the series. As a result, it is inferred that the original time series is non-stationary, characterized by a time-dependent mean or variance.

Upon differencing the series once, the ADF test statistic markedly decreases to -11,97547, significantly lower than the critical values at the 1%, 5%, and 10% levels. The p-value associated with this test statistic is virtually zero, offering robust evidence against the null hypothesis of a unit root in the differenced series.

The implications of this analysis are twofold: firstly, the original DJIA index exhibits non-stationarity, an inherent property of many financial time series that manifests as trends or random walks. Secondly, and more critically, stationarity is achieved after the first difference. This indicates that a single differencing (d=1) is adequate for stabilizing the mean, rendering the time series suitable for subsequent analysis within the ARIMAX framework.

In model selection for the ARIMAX time series analysis, we are guided by the principle of parsimony, also known as Occam’s Razor, a fundamental concept in various scientific disciplines. It emphasizes the importance of simplicity and minimizing the number of parameters in model design (Bröcker, 1998). The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, along with the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC), provide critical insights into the selection process.

In Figure 5, the ACF plot exhibits a gradual decline in the autocorrelation coefficients as the lags increase. This pattern suggests a slowly decaying correlation with past values that does not cut off abruptly. Such a decay is characteristic of a series where an MA component may be present. The ACF does not show a sharp cut-off, which would typically indicate the order of the MA component. However, the sustained significance across the lags hints at a more complex MA structure than a simple MA(1) model.

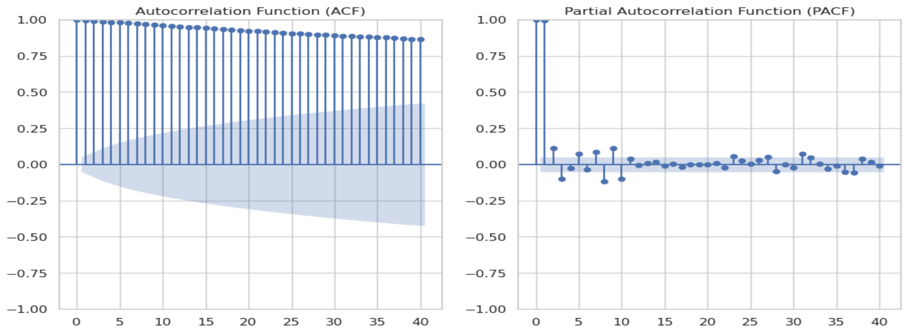

Figure 5. ACF and PACF Plots

In contrast, the PACF plot reveals a significant spike at lag 1, which sharply cuts off after that, a signature indicative of an AR(1) component. The PACF of the subsequent lags falls within the confidence bounds, implying that these lags do not provide additional significant explanatory power for the variation in the time series beyond what is already explained by the first lag. However, the second lag is marginally significant, suggesting that the AR component may extend beyond the first lag.

The significance of the second lag in the PACF plot should be considered. In the context of ARIMAX modeling, which incorporates seasonal patterns and exogenous variables, even slight indications of additional AR or MA components can be relevant, depending on their contribution to the model's predictive power and the robustness of the estimation.

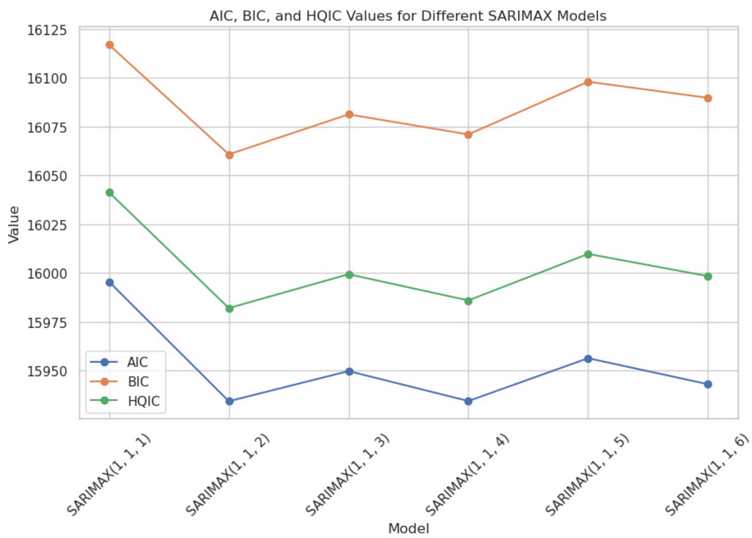

Figure 6. Information Criteria Comparison of Models

Considering the AIC and HQIC values in Figure 6, which are essential metrics for model comparison, we observe that the ARIMAX(1, 1, 2) model achieves the lowest AIC value of 15934.248, suggesting a superior fit to the data compared to other model specifications within the set range. Although the ARIMAX(1, 1, 4) model shows a similar AIC value, we prefer the ARIMAX(1, 1, 2) model for its simplicity, aligning with the parsimony principle.

The selection of the ARIMAX(1, 1, 2) model over the ARIMAX(1, 1, 4) is further substantiated by the trade-off between model complexity and the likelihood of overfitting. A model with fewer parameters (ARIMAX(1, 1, 2)) is less likely to overfit the data compared to a model with more parameters (ARIMAX(1, 1, 4)), especially when the information criteria values are similar.

Based on the observed ACF and PACF plots and the lowest AIC and competitive HQIC values, the ARIMAX(1, 1, 2) model is the most suitable forecasting model. This model strikes a compromise between complexity and explanatory power by effectively capturing the dynamics of the underlying time series data while retaining a simplified form.

4. 4. Validation Process

To make sure that our results are reliable and robust, we carefully validated the ARIMAX(1, 1, 2) model that we utilized for the advanced time series analysis of the DJIA.

Table 3. Performance Metrics of the Model

Metric	Value
MAE	270,556823
MSE	115343,3258
RMSE	339,622328
R^2	0,936476

The table illustrates that model performance metrics play a pivotal role in assessing the accuracy and predictive quality of the ARIMAX(1, 1, 2) model. Specifically, an R^2 value of 0,93 suggests that the ARIMAX(1,1,2) model accounts for about 93% of the variance in the data. This high R^2 value implies that the model perfectly fits the observed data.

Table 4. Cross-Validation Metrics of the Model

Metric	Value
CV AIC	9690,749577
CV BIC	9784,085054
CV HQIC	9726,834279

This table presents the cross-validation metrics, which are pivotal in the validation process. They include the Cross-Validation Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC). These cross-validation values are notably lower than their standard counterparts, indicating a more robust model performance in cross-validation scenarios. This suggests that the model fits the training data well and generalizes effectively to new data.

Table 5. Significance Metrics of the Model

	coef	std err	z	P>Z	[0,025	0,975]
ar,L1	0,5066	0,080	6,350	0,000	0,350	0,663
ma,L1	-0,6415	0,119	-5,397	0,000	-0,875	-0,409
ma,L2	-0,1794	0,077	-2,329	0,020	-0,330	-0,028

The coefficient analysis table is essential for understanding the influence of AR (AutoRegressive) and MA (Moving Average) components in the ARIMAX(1, 1, 2) model. This table demonstrates that the statistical significance of AR and MA components confirms their substantial impact on the model. The p-values and confidence intervals associated with these coefficients highlight their importance in the model's predictive capability. The AR and MA components are significant for the model, indicating their critical role in accurately forecasting the DJIA.

4. 5. Predictions

The ARIMAX(1,1,2) model's predictive performance on the DJIA closing values is depicted in Figure 7 through the actual versus predicted values graph for the specified period between 2022/09 and 2023/11. The graphed lines demonstrate the model's capacity to track the Index's movement over time.

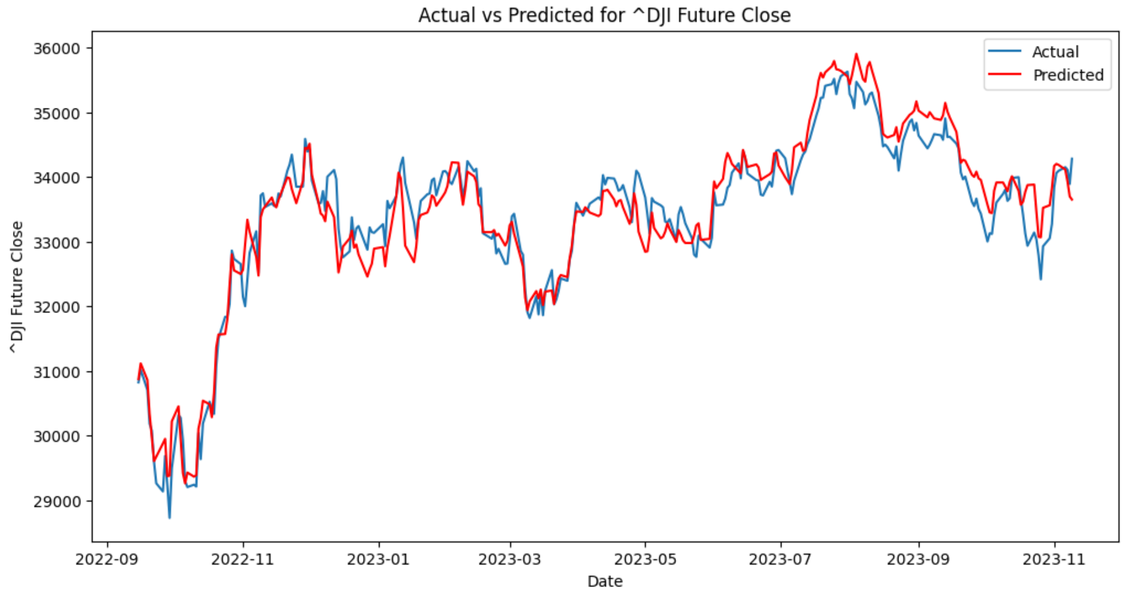

Figure 7. Actual vs. Predicted DJIA Values

The visual representation exhibits the model's ability to replicate the intrinsic fluctuations of the DJIA. The actual and predicted values' congruity indicates a strong model alignment with the time series volatility.

The model's fidelity is particularly notable in its response to significant market movements. It accurately delineates the pronounced decline in late 2022 and emulates the market's subsequent recovery. The peaks occurring around mid-2023 are also well-represented, indicating the model's responsiveness to pivotal market trends.

The accuracy of the ARIMAX(1,1,2) model is quantitatively expressed through an R^2 score of 0.93 for the DJIA future close predictions. This high R^2 value signifies that the model's predictions account for a substantial portion of the variability in the actual closing prices, underscoring its predictive precision.

5. COMPARING ACCURACY WITH OTHER MODELS

The accuracy of the three models was compared to assess the predictive power of the ARIMAX model. As shown in the table, these models included ARIMA, LSTM, and ARIMAX. The evaluation of model performance was based on four key metrics:

Table 6. Performance Comparison of Across Different Models

Metric	ARIMA	LSTM	ARIMAX
MAE	2.562,92	422.176,61	270,56
MSE	7.421.517,65	183.051,57	115.343,33
RMSE	2.724,25	427,85	339,62
R^2	-3,080	0,899	0,936

Based on the results in the table, ARIMAX outperforms both LSTM and ARIMA in terms of MAE, MSE, RMSE, and R^2, indicating its superiority in forecasting financial time series data in this context. LSTM also demonstrates strong performance, while ARIMA lags in all evaluated metrics.

Figure 8 shows ARIMA, LSTM, and ARIMAX model predictions, respectively. While a fundamental approach, the ARIMA model shows limitations in its predictive capability, as evidenced by its negative R^2 value. This suggests that the model does not adequately capture the complexities inherent in the data.

The R^2 scores demonstrate explanatory power the models (Hong and Rhee, 2022). The LSTM model achieves an R^2 score of 0.899, while the ARIMAX model scores 0.936. These results indicate their ability to explain a significant portion of the variance in the data. Conversely, the ARIMA model's negative R^2 score of -3.080 suggests it performs worse than a basic constant model. It's worth noting that R^2 is not a symmetric function, as highlighted by scikit-learn documentation, and may result in negative values, indicating poor model performance (www.scikit-learn.org, 2024). This underscores the inadequacy of the ARIMA model, ~~that~~ which means that the regression performed poorly~~.~~ (Chicco et al., 2021) ~~)~~ in capturing underlying data patterns compared to the more advanced LSTM and ARIMAX models.

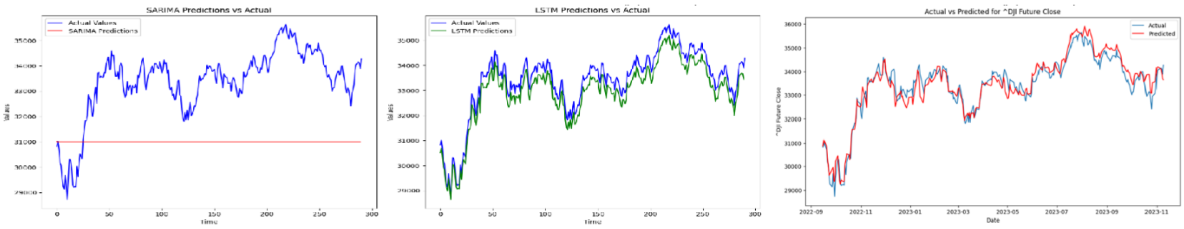

Figure 8. Comparison of Prediction Across Different Models

On the other hand, the model ARIMAX enhances all the contrastive metrics against all the different models by having the lowest MAE, RMSE, and the highest R^2 value with 0,936. This model's predictions are highly accurate compared to the other models. This is likely because the model ARIMAX can also consider external factors (exogenous variables).

6. DISCUSSION

The study's claim of enhanced predictive accuracy of the univariate ARIMA model can be advanced by including exogenous variables in the ARIMAX model. Including external variables undoubtedly broadens the analytical nature of the model since more variables that may impact the financial market are covered. However, their inclusion into the model adds complexity, possibly leading to overfitting. Does including many exogenous variables enhance the model's forecast ability, or does it lead to a complex model that cannot be used in practice?

At the same time, using CEEMDAN to generate the Intrinsic Mode Functions (IMFs) is an innovative aspect of this research. It offers a more in-depth look into the underlying dynamics of the DJIA. The attention paid to the selection and validation of hyperparameters in the model is also notable. However, it is worth considering the principle of simplicity. This principle dictates that we always prefer the simplest models.

The lack of overfitting in the relatively simple ARIMA (1,1,2) model, the stationarity of the data after differencing (Integration parameter I=1), and the significance of the AR and MA coefficients indicate a good and robust model. Moreover, the cross-validated information criteria and the R^2 score of 0,93 confirm that the model can generalize well. Therefore, the most crucial aspect of this study is the balance between keeping the model simple enough to be helpful in marketing analytics and yet being able to understand the fine details of the financial market.

7.CONCLUSION

This study, focusing on the DJIA, presents a comprehensive ARIMAX model incorporating major US stock indices, temporal lags, and outputs from CEEMDAN decomposition as exogenous variables. This research exhibits an R^2 score compared to some existing models in the literature. Its precision in capturing major market movements without lag is a crucial aspect of real-world financial forecasting.

Integrating CEEMDAN decomposition with the ARIMAX model is an essential methodological advancement that provides qualitative insights into market behavior. Using IMFs from the decomposition, this approach improves the model's predictive capability, allowing it to adapt to the complex dynamics of financial markets.

The literature reveals a range of R^2 scores in financial forecasting models. For instance, (Assous et al., 2020) achieved an R^2 of 0,993 with an ARIMA model, and (Bhagat et al., 2022) reported an R^2 of 0,91 using an ARIMAX model for crude oil price forecasting. While (Nurita, 2022) reported a lower adjusted R^2 value to a low of 35.1%. When ARIMAX models were used in different settings, the R^2 scores varied slightly; for example, (Yucesan et al., 2018) had 68-72%.

Our model has an R^2 of 0.93, which aligns with some of the highest-performing models in the literature. However, its efficacy at tracking the DJIA in real-time sets our model apart from the competition. This means that we can better track DJIA volatility and large swings in the DJIA compared to other models.

A forecasting model's significance in financial forecasting is not solely based on its R2 score. Its accuracy in capturing real-time market fluctuations and practical applicability are also determining factors. The model's sensitivity to significant market swings is confirmed by its exact tracking of the peaks in mid-2023, its precise recognition of the downturn in late 2022, and the recovery phase that followed.

This field has room for more investigation and advancement. Future research could extend this methodology to additional financial indices or macroeconomic indicators. Additionally, the model could be improved by incorporating sentiment analysis from news articles or social media, enhancing its ability to predict extreme market events or anomalies.

Furthermore, investigating alternative AI and machine learning approaches, such as ensemble methods or neural networks, could lead to new developments in financial forecasting. Developing even more complex and precise forecasting tools may result from combining these methods with conventional econometric models.

REFERENCES

Adebiyi, A. A., Adediran, A., & Ayo, C. K. (2014). *Stock Price Prediction Using the ARIMA Model*. https://doi.org/10.1109/uksim.2014.67

Adekanmbi. (2017). *ARIMA and ARIMAX Stochastic Models for Fertility in Nigeria*. https://api.semanticscholar.org/CorpusID:158086854

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/J.IS.2015.04.007

Alaoui, A. O., Dewandaru, G., Azhar Rosly, S., & Masih, M. (2015). Linkages and co-movement between international stock market returns: Case of Dow Jones Islamic Dubai Financial Market index. *Journal of International Financial Markets, Institutions and Money*, *36*, 53–70. https://doi.org/10.1016/J.INTFIN.2014.12.004

ALP, S., YİĞİT, Ö. E., & ÖZ, E. (2021). PREDICTION OF BIST PRICE INDICES: A COMPARATIVE STUDY BETWEEN TRADITIONAL AND DEEP LEARNING METHODS. *Sigma Journal of Engineering and Natural Sciences*, *38*(4), 1693–1704. https://dergipark.org.tr/en/pub/sigma/issue/65287/1004946

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons & Fractals*, *126*, 325–336. https://doi.org/10.1016/J.CHAOS.2019.07.011

Althelaya, K. A., Mohammed, S. A., & El-Alfy, E. S. M. (2021). Combining deep learning and multiresolution analysis for stock market forecasting. *IEEE Access*, *9*, 13099–13111. https://doi.org/10.1109/ACCESS.2021.3051872

Alzheev, A. V., & Kochkarov, R. A. (2020). Comparative analysis of ARIMA and LSTM predictive models: Evidence from Russian stocks. *Finance: Theory and Practice*, *24*(1), 14–23. https://doi.org/10.26794/2587-5671-2020-24-1-14-23

Arakelyan, A., Nersisyan, L., Hakobyan, A., Arakelyan, A., Nersisyan, L., & Hakobyan, A. (2016). Application of MATLAB in -Omics and Systems Biology. *Applications from Engineering with MATLAB Concepts*. https://doi.org/10.5772/62847

Assous, H. F., Al-Rousan, N., AL-Najjar, D., & Al-Najjar, H. (2020). Can International Market Indices Estimate TASI's Movements? The ARIMA Model. *Journal of Open Innovation Technology Market and Complexity*. https://doi.org/10.3390/joitmc6020027

Baek, S., Mohanty, S. K., & Glambosky, M. (2020). COVID-19 and stock market volatility: An industry level analysis. *Finance Research Letters*, *37*, 101748. https://doi.org/10.1016/J.FRL.2020.101748

Bahloul, S., & Khemakhem, I. (2021). Dynamic return and volatility connectedness between commodities and Islamic stock market indices. *Resources Policy*, *71*, 101993. https://doi.org/10.1016/J.RESOURPOL.2021.101993

Ban, W., & Shen, L. (2022). PM2.5 Prediction Based on the CEEMDAN Algorithm and a Machine Learning Hybrid Model. *Sustainability 2022, Vol. 14, Page 16128*, *14*(23), 16128. https://doi.org/10.3390/SU142316128

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. https://doi.org/10.1371/JOURNAL.PONE.0180944

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA Model on the COVID-2019 Epidemic Dataset. *Data in Brief*. https://doi.org/10.1016/j.dib.2020.105340

Bhagat, V., Sharma, M., & Saxena, A. (2022). Modelling the nexus of macro-economic variables with WTI Crude Oil Price: A Machine Learning Approach. *2022 IEEE Region 10 Symposium, TENSYMP 2022*. https://doi.org/10.1109/TENSYMP54529.2022.9864544

Borsa İstanbul Hissesi Örneği Caner ERDEN, B. (2023). Derin Öğrenme ve ARIMA Yöntemlerinin Tahmin Performanslarının Kıyaslanması: Bir Borsa İstanbul Hissesi Örneği. *Yönetim ve Ekonomi Dergisi*, *30*(3), 419–438. https://doi.org/10.18657/YONVEEK.1208807

Bröcker, J. (1998). Operational Spatial Computable General Equilibrium Modeling. *The Annals of Regional Science*. https://doi.org/10.1007/s001680050079

Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, *236*, 1078–1088. https://doi.org/10.1016/J.APENERGY.2018.12.042

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, *519*, 127–139. https://doi.org/10.1016/J.PHYSA.2018.11.061

Chen, L., Liu, X., Zeng, C., He, X., Chen, F., & Zhu, B. (2022). Temperature Prediction of Seasonal Frozen Subgrades Based on CEEMDAN-LSTM Hybrid Model. *Sensors*. https://doi.org/10.3390/s22155742

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623/SUPP-1

Colominas, M. A., Schlotthauer, G., Torres, M. E., & Flandrin, P. (2012). Noise-Assisted Emd Methods in Action. *Advances in Adaptive Data Analysis*. https://doi.org/10.1142/s1793536912500252

Costola, M., Hinz, O., Nofer, M., & Pelizzon, L. (2023). Machine learning sentiment analysis, COVID-19 news and stock market reactions. *Research in International Business and Finance*, *64*, 101881. https://doi.org/10.1016/J.RIBAF.2023.101881

Dua, P., & Tuteja, D. (2023). Inter-linkages between asian and U.S. stock market returns: A multivariate garch analysis. *Macroeconometric Methods: Applications to the Indian Economy*, 339–376. https://doi.org/10.1007/978-981-19-7592-9_12/COVER

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, 100190. https://doi.org/10.1016/J.COSREV.2019.08.001

Gao, Z. (2021). Stock Price Prediction with ARIMA and Deep Learning Models. *2021 IEEE 6th International Conference on Big Data Analytics, ICBDA 2021*, 61–68. https://doi.org/10.1109/ICBDA51983.2021.9403037

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/S41586-020-2649-2

Hong, J., & Rhee, J. K. (2022). Genomic Effect of DNA Methylation on Gene Expression in Colorectal Cancer. *Biology*, *11*(10), 1388. https://doi.org/10.3390/BIOLOGY11101388/S1

Huang, B. L., & Yao, Y. (2014). Batch-to-batch Steady State Identification via Online Ensemble Empirical Mode Decomposition and Statistical Test. *Computer Aided Chemical Engineering*, *33*, 787–792. https://doi.org/10.1016/B978-0-444-63456-6.50132-0

Kashyap, S. (2023). Review on volatility and return analysis including emerging developments: evidence from stock market empirics. *Journal of Modelling in Management*, *18*(3), 756–816. https://doi.org/10.1108/JM2-10-2021-0249/FULL/PDF

Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I., & Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ*, *350*. https://doi.org/10.1136/BMJ.H2750

Kotu, V., & Deshpande, B. (2019). Time Series Forecasting. *Data Science*, 395–445. https://doi.org/10.1016/B978-0-12-814761-0.00012-5

Laszuk, D. (2017). Python implementation of Empirical Mode Decomposition algorithm. *GitHub Repository*. https://doi.org/10.5281/zenodo.5459184

Li, Y., Li, Y., Chen, X., Yu, J., Yang, H., & Wang, L. (2018). A New Underwater Acoustic Signal Denoising Technique Based on CEEMDAN, Mutual Information, Permutation Entropy, and Wavelet Threshold Denoising. *Entropy*. https://doi.org/10.3390/e20080563

Liu, J., Sun, T., Luo, Y., Yang, S., Cao, Y., & Zhai, J. (2020). An Echo State Network Architecture Based on Quantum Logic Gate and Its Optimization. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2019.09.002

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward. *Plos One*. https://doi.org/10.1371/journal.pone.0194889

McKinney, W., & others. (2010). Data structures for statistical computing in python. *Statsmodels: Econometric and Statistical Modeling with Python*, *445*, 51–56.

Noh, J. H., & Park, H. (2023). Greenhouse gas emissions and stock market volatility: an empirical analysis of OECD countries. *International Journal of Climate Change Strategies and Management*, *15*(1), 58–80. https://doi.org/10.1108/IJCCSM-10-2021-0124/FULL/PDF

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review 2019 53:4*, *53*(4), 3007–3057. https://doi.org/10.1007/S10462-019-09754-Z

Nurita, D. (2022). Dispotition Effect and Momentum. *Jurnal Manajerial*. https://doi.org/10.30587/jurnalmanajerial.v9i02.3918

Saranj, A., & Zolfaghari, M. (2022). The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models. *Energy Reports*, *8*, 7657–7679. https://doi.org/10.1016/J.EGYR.2022.06.007

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Statsmodels: Econometric and Statistical Modeling with Python*.

Sen, J., & Chaudhuri, T. D. (2016). *A Framework for Predictive Analysis of Stock Market Indices : A Study of the Indian Auto Sector*. https://arxiv.org/abs/1604.04044v1

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies 2019, Vol. 7, Page 26*, *7*(2), 26. https://doi.org/10.3390/IJFS7020026

Shaikh, I., & Padhi, P. (2015). The implied volatility index: Is 'investor fear gauge' or 'forward-looking'? *Borsa Istanbul Review*, *15*(1), 44–52. https://doi.org/10.1016/J.BIR.2014.10.001

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Singh, S. A., Singh, S. A., Devi, N. D., & Majumder, S. (2021). A study on sleep stage classification based on a single-channel EEG signal. *Electronic Devices, Circuits, and Systems for Biomedical Applications: Challenges and Intelligent Approach*, 135–152. https://doi.org/10.1016/B978-0-323-85172-5.00016-2

Scikit-learn 1.4.1 documentation. from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2score.html#sklearn-metrics-r2-score

Wang, J., & Wang, J. (2015). Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. *Neurocomputing*, *156*, 68–78. https://doi.org/10.1016/J.NEUCOM.2014.12.084

Wang, Q., Kang, K., Zhihan, Z., & Cao, D. (2021). Application of LSTM and CONV1D LSTM Network in Stock Forecasting Model. *Artificial Intelligence Advances*. https://doi.org/10.30564/aia.v3i1.2790

Wu, X., Sun, C., & Hao, X. (2022). Stock Closing Price Interval Prediction Based on CEEMDAN-WTD-Bilstm-Transformer Model. *BCP Business & Management*. https://doi.org/10.54691/bcpbm.v20i.898

Yang, X. D., Luo, M., Tao, L., & Song, G. (2017). ECG Signal De-Noising and Baseline Wander Correction Based on CEEMDAN and Wavelet Threshold. *Sensors*. https://doi.org/10.3390/s17122754

Yucesan, M., Gul, M., & Celik, E. (2018). Performance comparison between ARIMAX, ANN and ARIMAX-ANN hybridization in sales forecasting for furniture industry. *Drvna Industrija*, *69*(4), 357–370.

Zhai, Y., Yang, X., Peng, Y., Wang, X., & Bai, K. (2020). Multiscale Entropy Feature Extraction Method of Running Power Equipment Sound. *Entropy*. https://doi.org/10.3390/e22060685

Zhang, J., Jin, Y., Sun, B., Han, Y., & Yang, H. (2021). Study on the Improvement of the Application of Complete Ensemble Empirical Mode Decomposition With Adaptive Noise in Hydrology Based on RBFNN Data Extension Technology. *Computer Modeling in Engineering & Sciences*. https://doi.org/10.32604/cmes.2021.012686

Zhao, C., Hu, P., Liu, X., Lan, X., & Zhang, H. (2023). Stock Market Analysis Using Time Series Relational Models for Stock Price Prediction. *Mathematics 2023, Vol. 11, Page 1130*, *11*(5), 1130. https://doi.org/10.3390/MATH11051130

Beyan ve Açıklamalar (Disclosure Statements)

1. Bu çalışmanın yazarları, araştırma ve yayın etiği ilkelerine uyduklarını kabul etmektedirler (The authors of this article confirm that their work complies with the principles of research and publication ethics).

2. Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir (No potential conflict of interest was reported by the authors).

3. Bu çalışma, intihal tarama programı kullanılarak intihal taramasından geçirilmiştir (This article was screened for potential plagiarism using a plagiarism screening program).