Iran J Econ Stud, 2026, 15(2), 479-517

Shiraz University

Iranian Journal of Economic Studies

Journal homepage: ijes.shirazu.ac.ir

Does Signal Decomposition Improve Stock Market Forecasting? Uncovering High- and Low-Frequency Information with a Hybrid EMD–Random Forest Model

Mohamad Ahadzadeh[a], Fatemeh Ahadzadeh[b], Ruhollah Moradi[*c]

a. Department of Management and Economics, Science and Research Branch, Islamic Azad University Tehran, Iran.
b. Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran.
c. Department of Economics and Accounting, Imam Ali Officers' University, Tehran, Iran.

Highlights

- **The Proposed EMD–RF Hybrid Achieves the Lowest Observed Forecasting Errors Across All Evaluated Equity Indices.**
- **High- and Low-Frequency IMFs Provide Complementary Predictive Information.**
- **EMD–RF Forecasting Gains Vary in Statistical Significance Across Markets and Benchmarks.**
- **EMD-Based Decomposition Enhances RF Forecasting Performance Across Diverse Equity Markets.**

Article History

Received: 05 September 2026
Revised: 23 October 2026
Accepted: 30 October 2026
Published: 06 November 2026

JEL Classification

C22
C45
C53
G17

Keyword

Empirical Mode Decomposition (EMD)
Random Forest
High-frequency IMFs
Intrinsic Mode Functions (IMFs)

Abstract

Accurate equity index forecasting requires models capable of capturing temporal dependencies across multiple scales. This study develops a hybrid empirical mode decomposition–Random Forest (EMD–RF) framework and evaluates its performance against standalone Random Forest (RF), support vector machine (SVM), and XGBoost models across the CAC 40, S&P 500, SSE Composite, and Tehran Stock Exchange (TSE). Forecasting performance is assessed using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE), complemented by Diebold–Mariano (DM) tests and component-level ablation analysis. EMD–RF achieves the lowest observed values for all three error metrics across all four indices, demonstrating consistent improvements over the benchmark models. However, statistical significance varies across markets: improvements are significant against all benchmarks for the CAC 40 and TSE, significant only against XGBoost for the S&P 500, and statistically inconclusive for the SSE Composite. Ablation results further show that high- and low-frequency intrinsic mode functions (IMFs) provide complementary predictive information across all four markets, although their relative contributions vary by index. The residual component also contributes additional predictive information to differing degrees. These findings demonstrate that EMD-based decomposition enhances RF forecasting performance by preserving informative temporal structures, while highlighting the market-dependent contributions of frequency-specific components and the importance of distinguishing observed forecast improvements from statistically established gains.

* financialmanagement421@gmail.com
Doi: https://doi.org/10.22099/ijes.2026.57068.2149

Copyright © 2026 The Authors. Published by Shiraz University. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license. Non-commercial uses of the work are permitted, provided the original work is properly cited.

1. Introduction

In recent years, researchers have sought solutions capable of overcoming the limitations of traditional models and enhancing the accuracy of volatility forecasting. Volatility, in fact, is recognized as one of the primary indicators of risk and uncertainty in financial markets, reflecting the magnitude of price fluctuations over different time horizons. Financial market volatility plays an important role in investment decisions, option pricing, and financial market regulation (Poon & Granger, 2003). A precise understanding of volatility structure and its forecasting not only facilitates effective risk management and the design of efficient trading strategies but also contributes to analyzing investor behavior and the dynamics of international markets. Given the critical importance of volatility in financial markets, numerous studies have been conducted in this area (Khoojine 2023; 2024; Kang 2025; Karanasos et al. 2022; Letteri 2023; Case and Clements 2025; Mansilla-Lopez et al. 2025; Lee et al. 2002).

One of the most successful approaches in this field involves combining signal decomposition methods with machine learning algorithms. Decomposition techniques, such as Empirical Mode Decomposition (EMD), possess the capability to extract meaningful components from complex time series, allowing short-term and long-term volatility to be modeled separately. On the other hand, machine learning algorithms, particularly Random Forest, are capable of capturing nonlinear relationships and complex interactions among variables, providing accurate forecasts without requiring stringent assumptions about data distributions.

This study proposes a hybrid framework based on Empirical Mode Decomposition (EMD) and Random Forest, capable of accurately identifying and forecasting hidden patterns and temporal complexities in international stock market indices. By integrating the signal decomposition strength of EMD in extracting meaningful components from volatile time series with the machine learning capability to model nonlinear relationships, this innovative framework enables precise and robust short-term forecasts. Given the growing applications of machine learning models, numerous studies have explored their effectiveness in capturing financial market volatility (Diaz et al. 2024; Kumar et al. 2025; Han et al. 2024; Filipović and Khalilzadeh 2021).

The primary objective of this study is to investigate whether Empirical Mode Decomposition (EMD) enhances stock market index forecasting by decomposing original time series into informative intrinsic mode functions (IMFs) and a residual component, and to examine how the predictive contributions of high and low frequency components vary across markets. To this end, the study develops a hybrid EMD–Random Forest (EMD–RF) framework that integrates adaptive signal decomposition with nonlinear machine learning to capture the temporal structures embedded in financial time series. The study further evaluates whether EMD–RF reduces forecasting errors relative to standalone machine learning benchmarks across the examined equity indices and investigates whether the statistical significance of its predictive performance gains varies across markets

and benchmark models. By combining component-level ablation analysis with comparative forecasting evaluation and formal tests of predictive accuracy, this research aims to clarify the role of multiscale decomposition in equity index forecasting and establish a systematic basis for assessing the predictive contributions of distinct temporal components and the cross-market performance of hybrid forecasting models.

This study addresses two primary research questions concerning the role of Empirical Mode Decomposition (EMD) in stock market index forecasting and the predictive performance of the proposed hybrid EMD–Random Forest (EMD–RF) framework. Specifically, it examines how EMD-based decomposition into informative intrinsic mode functions (IMFs) and a residual component contributes to forecasting accuracy, how the predictive contributions of high and low frequency components vary across markets, and whether EMD–RF reduces forecasting errors relative to standalone machine learning benchmarks, with particular attention to differences in the statistical significance of its performance gains across markets and benchmark models. The research questions are presented below.

RQ1: To what extent does Empirical Mode Decomposition (EMD) improve stock market index forecasting by decomposing the original time series into informative intrinsic mode functions (IMFs) and a residual component, and how do the predictive contributions of high and low frequency components vary across markets?

RQ2: To what extent does the proposed hybrid EMD–Random Forest (EMD–RF) model reduce forecasting errors relative to standalone machine learning benchmarks across the examined stock market indices, and how do the statistical significance of its performance gains vary across markets and benchmark models?

This paper is organized as follows. Section 1 introduces the research problem and motivates the use of a hybrid EMD–Random Forest framework for equity index forecasting. Section 2 reviews the literature on financial forecasting, signal decomposition, and machine learning. Section 3 describes the data, methodological framework, and evaluation procedures. Section 4 presents the empirical results, including benchmark comparisons, component-level ablation analyses, and Diebold–Mariano tests. Section 5 discusses the findings, evaluates the hypotheses, and outlines the study's contributions and implications.

2. Literature Review

The literature on volatility forecasting with high frequency IMFs has increasingly focused on hybrid approaches that integrate advanced decomposition techniques with machine learning models. This section reviews key contributions in the field, highlighting how empirical mode decomposition (EMD) and ensemble learning methods have been employed to capture complex, non-linear patterns in financial time series. By examining these developments, the review

provides the foundation for the proposed framework and situates the study within the broader context of predictive modeling in international stock markets.

2.1. Machine Learning Models and Their Impact on Volatility Prediction

Machine learning models have increasingly become central to the field of volatility prediction due to their remarkable capacity to model complex, non-linear relationships and adapt to dynamic market environments. Unlike traditional econometric approaches, which often rely on strict assumptions and linear structures, machine learning algorithms can process vast amounts of high-frequency financial data, uncover latent patterns, and respond to regime shifts in real time. Techniques such as random forests, support vector machines, and gradient boosting have demonstrated superior predictive performance across various stock indices, capturing subtle interactions among market variables that conventional models may overlook. Furthermore, these models facilitate the integration of multiple data sources, including technical indicators, macroeconomic factors, and decomposed components of price series, enhancing both the granularity and robustness of volatility forecasts. The growing body of research underscores that machine learning-based frameworks not only improve short-term prediction accuracy but also provide valuable insights into the underlying drivers of market volatility, offering a transformative perspective for risk management and strategic investment decisions. Yan and Li (2024), in their study *"Machine Learning-Based Analysis of Volatility Quantitative Investment Strategies for American Financial Stocks,"* examined volatility-based quantitative investment strategies for twelve US financial stocks. They combined three time series models with six machine learning models, K-nearest neighbors, AdaBoost, CatBoost, LightGBM, XGBoost, and Random Forest, to create eighteen fusion models that accurately predicted volatility. Integrating these predictions with quantitative investing principles, their strategy achieved improved returns, with potential average annualized returns of 5 to 10%. Zhang et al. (2023), in their study *"Volatility Forecasting with Machine Learning and Intraday Commonality,"* applied machine learning models to forecast intraday realized volatility (RV) by exploiting commonality in intraday volatility across pooled stock data and incorporating a proxy for market volatility. They found that neural networks outperformed linear regressions and tree-based models due to their ability to capture complex latent interactions among variables. The results were robust when applied to stocks not included in the training set, providing evidence for a universal volatility mechanism. Additionally, they proposed a new method for forecasting one-day-ahead RVs using past intraday RVs as predictors, highlighting time-of-day effects that enhance forecasting accuracy. The methodology yielded superior out-of-sample forecasts compared to traditional approaches relying only on past daily RVs. Lubis and Samsudin (2025), in their study *"Using the Random Forest Method in Predicting Stock Price Movements,"* applied the Random Forest machine learning algorithm to forecast stock price movements. Using historical data from companies listed on the Indonesia Stock

Exchange (IDX), their model achieved 98% accuracy, an R-squared of 0.94, and a Mean Absolute Percentage Error (MAPE) of 0.40%. The study identified key factors affecting stock prices, including Previous, High, Low, Volume, and Change. By utilizing a large dataset of 104 stock codes and integrating interactive visualization through Streamlit, the research provides a robust predictive tool that outperforms traditional methods and aids investors in making more informed decisions.

2.2. Foundational EMD and Its Applications in Financial Markets

Empirical Mode Decomposition (EMD), introduced by Huang et al. (1998), provides an adaptive, data-driven approach to analyzing nonlinear and nonstationary time series without requiring predefined basis functions. By decomposing a signal into intrinsic mode functions (IMFs) and a residual component, EMD facilitates the representation of temporal dynamics across distinct characteristic scales. This property is particularly relevant to financial time series, in which short-term fluctuations and longer-term movements may exhibit different statistical characteristics and predictive relevance. Subsequent applications of EMD to financial forecasting have explored its capacity to characterize complex market dynamics and extract information that may be obscured in undecomposed observations. Lin et al. (2012) proposed a hybrid EMD–LSSVR model for foreign exchange rate forecasting to address the nonlinear and nonstationary characteristics of financial time series. The model decomposes exchange rate data into intrinsic mode functions (IMFs) and a residual component, forecasts each component individually using LSSVR, and aggregates the forecasts to obtain the final prediction. Empirical results show that EMD–LSSVR outperforms EMD–ARIMA, LSSVR, and ARIMA models without time-series decomposition. Nevertheless, the predictive value of individual components cannot be assumed to be uniform across markets or forecasting settings. Distinguishing the contributions of high-frequency IMFs, low-frequency IMFs, and the residual is therefore essential to understanding whether decomposition improves equity index forecasting and how its benefits depend on market-specific dynamics. Building on this literature, the present study evaluates these component-specific contributions through ablation analysis across four equity indices, examining how high-frequency IMFs, low-frequency IMFs, and the residual affect forecasting accuracy within the proposed EMD–RF framework.

2.3. EMD–Machine Learning Hybrid Models for Financial Forecasting

The integration of EMD with machine learning algorithms has emerged as a promising approach to financial forecasting, combining adaptive signal decomposition with the capacity of nonlinear learning models to capture complex predictive relationships. Within these hybrid frameworks, decomposition separates the original series into components that can be modeled individually or incorporated into a unified predictive architecture, potentially improving

forecasting accuracy relative to models trained on undecomposed data. Random Forest (RF), Support Vector Machines (SVM), and gradient-boosting methods offer complementary modeling capabilities for evaluating the predictive benefits of such integration. Cheng and Wei (2014) proposed a hybrid time-series Support Vector Regression (SVR) model based on Empirical Mode Decomposition (EMD) to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX). The proposed model was compared with autoregressive (AR) and SVR models to evaluate forecasting performance. Experimental results showed that the proposed model outperformed the benchmark models in terms of root mean squared error (RMSE), with greater forecasting accuracy during the more fluctuating period of 2000–2001. However, lower forecasting errors alone do not establish whether observed improvements are statistically significant, nor do aggregate performance measures reveal which decomposed components contribute most to predictive accuracy. A rigorous assessment of EMD-based hybrid models therefore requires comparisons with standalone machine learning benchmarks, component-level ablation analyses, and formal tests of predictive accuracy. Evaluating these dimensions across multiple equity indices can clarify both the generalizability of hybrid forecasting approaches and the extent to which their predictive advantages depend on market characteristics and benchmark selection.

2.4. Stock Index Trends and Market Volatility

Figure 1 illustrates the trends of the CAC 40, S&P 500, SSE Composite, and TSE stock indices over the study period. These trends indicate the overall direction of the market and changes in average prices, but they alone do not provide precise information on the magnitude of volatility. However, analyzing trends can offer important clues about potential fluctuations; for instance, periods of strong upward or downward movement are often accompanied by higher volatility, and sudden shifts in the indices can highlight critical points of changing market risk. Therefore, observing and examining index trends plays a key role in identifying and anticipating dynamic changes in market volatility.

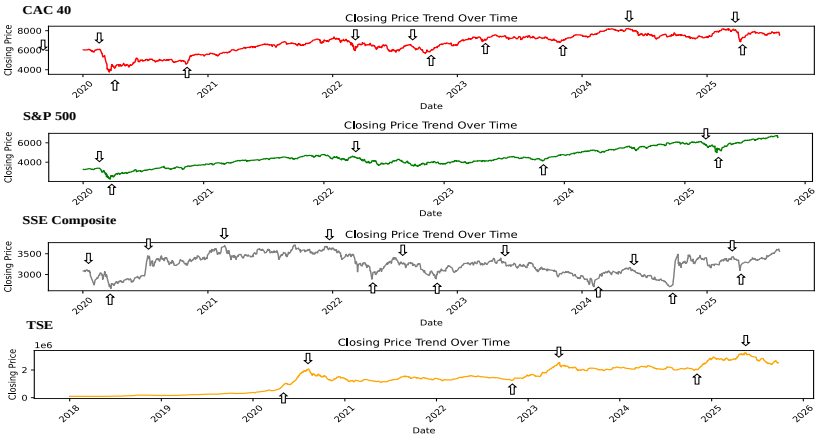


Figure 1. Trends of Market Indices

Sources: *Authors' own elaboration*

2.5. Random Forest (RF) in Volatility Forecasting

Random Forest (RF) has emerged as a highly effective machine learning technique for volatility forecasting due to its ensemble-based structure and robustness against overfitting. By aggregating the predictions of multiple decision trees, RF captures complex, non-linear interactions among financial variables that traditional linear models often fail to identify. Its inherent ability to handle high-dimensional data and noisy inputs makes it particularly suitable for modeling high-frequency market fluctuations and capturing sudden shifts in volatility. Applications of RF in financial markets have demonstrated superior predictive accuracy across diverse stock indices, enabling more precise estimation of risk and improved portfolio management strategies. Furthermore, the model's interpretability, through feature importance measures, provides valuable insights into the key drivers of volatility, offering both predictive power and explanatory value in understanding market dynamics. Lamine and Brijlal (2024), in their study "*Forecasting Stock Market Realized Volatility Using Random Forest and Artificial Neural Network in South Africa*," stated that conventional time series models such as GARCH are the preferred models in the literature. However, they adopted two novel non-linear machine learning algorithms, Random Forest (RF) and Artificial Neural Network (ANN), to predict realized volatility for the JSE Basic Material (JBIND) and Financial (JFIN) indices over five years. Their results showed that the Random Forest model outperformed the ANN for both indices, while the impact of COVID-19 on model performance was found to be ambiguous. Campisi et al. (2024), in their study *"A Comparison of Machine Learning Methods for Predicting the Direction of the US Stock Market on the Basis of Volatility Indices,"* examined the use of volatility indices to predict the US stock market direction. Using data from January 2011 to July 2022, they

applied various machine learning methods and evaluated performance with accuracy, ROC-AUC, and F-measure. Results showed that machine learning models outperformed classical linear regression, with random forests and bagging achieving the highest predictive accuracy.

2.6. Research Hypotheses

The primary objective of this study is to introduce and validate an innovative hybrid approach that combines the adaptive signal processing capabilities of Empirical Mode Decomposition (EMD) with the robust predictive power of the Random Forest (RF) machine learning algorithm for high-frequency volatility forecasting. To guide this investigation, the following two research hypotheses, conceptually illustrated in Figure 2, are proposed:

H1: EMD-based decomposition improves stock market index forecasting by separating the original time series into informative intrinsic mode functions (IMFs) and a residual component, with the predictive contributions of high-frequency and low-frequency components varying across markets.

H2: The proposed hybrid EMD–Random Forest (EMD–RF) model achieves lower forecasting errors than standalone machine learning benchmarks across the examined stock market indices, while the statistical significance of its performance gains varies across markets and benchmark models.

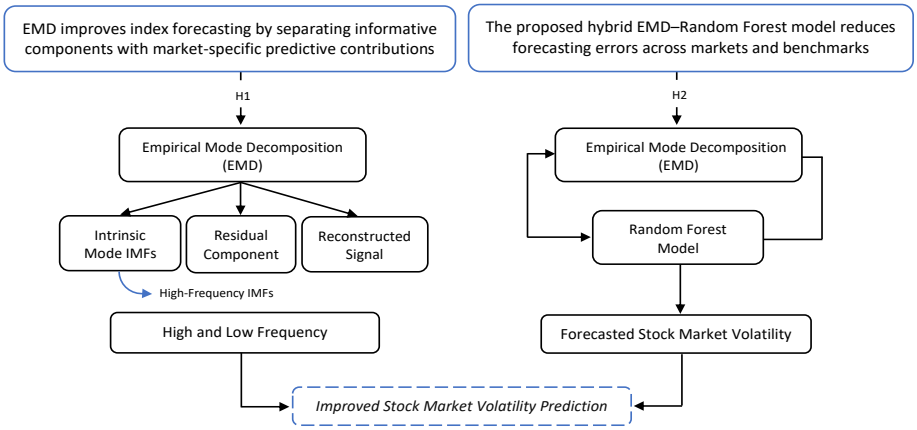


Figure 2. Research Hypotheses in EMD-RF Volatility Forecasting

Sources: *Authors' own elaboration*

3. Methodology

3.1. Research Design

This study employs a quantitative, comparative design to evaluate a hybrid empirical mode decomposition–Random Forest (EMD–RF) framework for one-step-ahead forecasting of daily log returns for four international equity indices: the CAC 40, S&P 500, SSE Composite, and TSE. The target variable is the daily

log return, calculated as the natural logarithm of the ratio of an index's closing price on a given trading day to its closing price on the preceding trading day. Thus, the framework forecasts the next trading day's log return, not the closing index level or financial market volatility. This distinction is essential because intrinsic mode functions (IMFs) represent return-series components at different temporal scales and are not direct measures of volatility. EMD decomposes the daily log-return series into IMFs and a residual component, which are classified into high- and low-frequency groups using estimated mean periods and dominant frequencies. A Random Forest (RF) regressor is trained separately on lagged observations of each IMF and the residual to generate one-step-ahead component forecasts, which are aggregated into the final predicted log return. Performance is compared with standalone RF, support vector regression (SVR), and extreme gradient boosting (XGBoost), using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE). Diebold–Mariano (DM) tests assess the statistical significance of predictive-accuracy differences. A systematic ablation analysis compares the complete EMD–RF framework with standalone RF applied to undecomposed returns and variants excluding high-frequency IMFs, low-frequency IMFs, or the residual; it also examines the effect of reducing lagged inputs. This design distinguishes overall forecasting performance from the incremental contributions of individual components. Evaluation preserves chronological ordering and aims to prevent future observations from entering predictive-feature construction. All analyses are conducted in Python using scientific computing, signal decomposition, and machine learning libraries.

3.2. Population and Sample

The empirical sample comprises four benchmark equity indices representing distinct national financial markets: the CAC 40 (France), the S&P 500 (United States), the SSE Composite (China), and the Tehran Stock Exchange (TSE) index (Iran). These indices were selected to facilitate a cross-market assessment of the EMD–RF framework across heterogeneous financial environments rather than to represent all international equity markets statistically. As reported in Table 1, the dataset contains 1,432 daily observations for the CAC 40, 1,452 for the S&P 500, and 1,353 for the SSE Composite, covering 2020–2025, alongside 1,842 daily observations for the TSE covering 2018–2025. The differing sample periods reflect market-specific data coverage and should be considered when interpreting cross-market comparisons. Data for the CAC 40, S&P 500, and SSE Composite were obtained from Yahoo Finance, whereas TSE data were sourced from *Rahavard* 365. The inclusion of these four indices enables the analysis to assess whether the observed forecasting gains and the predictive contributions of different frequency components are consistent across markets with different institutional and economic characteristics. The empirical findings indicate that EMD–RF yields lower observed forecasting errors for all four indices, while DM test results reveal variation in the statistical significance of these gains across

markets and benchmark models. The sample therefore provides a comparative basis for examining the cross-market performance of the proposed framework, although the findings should be interpreted within the limits of the selected indices, observation periods, and data sources.

Table 1. Overview of Selected Stock Market Indices

Index	Country	Frequency	Observations	Period
CAC 40	France	Daily	1432	2020–2025
S&P 500	United States	Daily	1452	2020–2025
SSE Composite	China	Daily	1353	2020–2025
TSE	Iran	Daily	1842	2018–2025

Source: *Source: Authors' compilation*

3.3. Empirical Mode Decomposition (EMD)

Empirical mode decomposition (EMD) is employed to represent financial time series through a data-adaptive decomposition of their oscillatory behavior into intrinsic mode functions (IMFs) and a residual component. The procedure iteratively identifies local extrema, constructs upper and lower envelopes using spline interpolation, and subtracts their mean from the current signal to obtain candidate IMFs. Each candidate is subjected to the standard IMF criteria, which require the numbers of zero crossings and extrema to differ by no more than one and the local mean defined by the envelopes to be approximately zero. Validated components are extracted sequentially until the prescribed stopping criterion is met, leaving a residual that captures the remaining slower-moving variation. In the proposed EMD–Random Forest (RF) framework, the resulting components provide the basis for examining and forecasting financial dynamics across different temporal scales. High- and low-frequency IMFs are analyzed separately to assess their respective predictive contributions, while the residual is retained to evaluate the additional information associated with the lower-frequency structure of the series. This component-based design is supported by the empirical findings: excluding either frequency group increases forecasting errors across all four equity indices, although the magnitude of the deterioration varies by market and evaluation metric. The contribution of the residual also differs across indices and is comparatively modest for the CAC 40. These results motivate treating EMD not merely as a preprocessing operation but as an analytical framework for investigating the complementary and market-dependent predictive information contained in different components of financial time series.

3.4. High- and Low-Frequency IMF Analysis

Figure 3 presents the EMD procedure implemented in this study to decompose financial time series into oscillatory components operating at different temporal scales. The process begins with the financial signal, represented by stock index closing prices, followed by the calculation of logarithmic returns, which express relative price changes in additive form. The resulting series is then subjected to the iterative EMD sifting procedure, in which local extrema are

identified, upper and lower envelopes are constructed, and their mean is subtracted from the current signal to isolate an oscillatory component. Each candidate component is evaluated against the IMF sifting criteria before being retained as a valid intrinsic mode function (IMF). Once extracted, the IMF is stored and removed from the current signal, leaving a residual that serves as the input for subsequent extraction. This procedure continues until the stopping criteria are satisfied, producing an ordered set of IMFs and a final residual component. As illustrated in Figure 3, this sequential decomposition provides a structured representation of the original financial series without imposing a predetermined set of basis functions, allowing its oscillatory behavior and slower-moving variation to be examined separately.

Following decomposition, the extracted IMFs and residual are characterized and examined to distinguish their frequency content and assess their relevance to the underlying financial dynamics. The analysis incorporates descriptive statistics, including the mean, standard deviation, variance, and energy, alongside measures of dominant frequency based on the fast Fourier transform (FFT), persistence through the Hurst exponent, complexity through sample entropy, and correlation with the original series. These complementary diagnostics provide a basis for differentiating relatively high-frequency components, which represent faster fluctuations, from lower-frequency components and the residual, which capture more slowly evolving variation. The frequency-based classification is further supported by visual inspection of the decomposed components and the assessment of their individual and cumulative energy contributions. As indicated in Figure 3, this stage establishes the component representation used in the subsequent EMD–Random Forest forecasting framework. By preserving the distinct temporal characteristics of the decomposed series, the procedure enables the predictive contribution of different frequency groups and the residual to be investigated systematically, providing a methodological basis for the component-level analysis without presupposing that their relative importance is identical across equity markets.

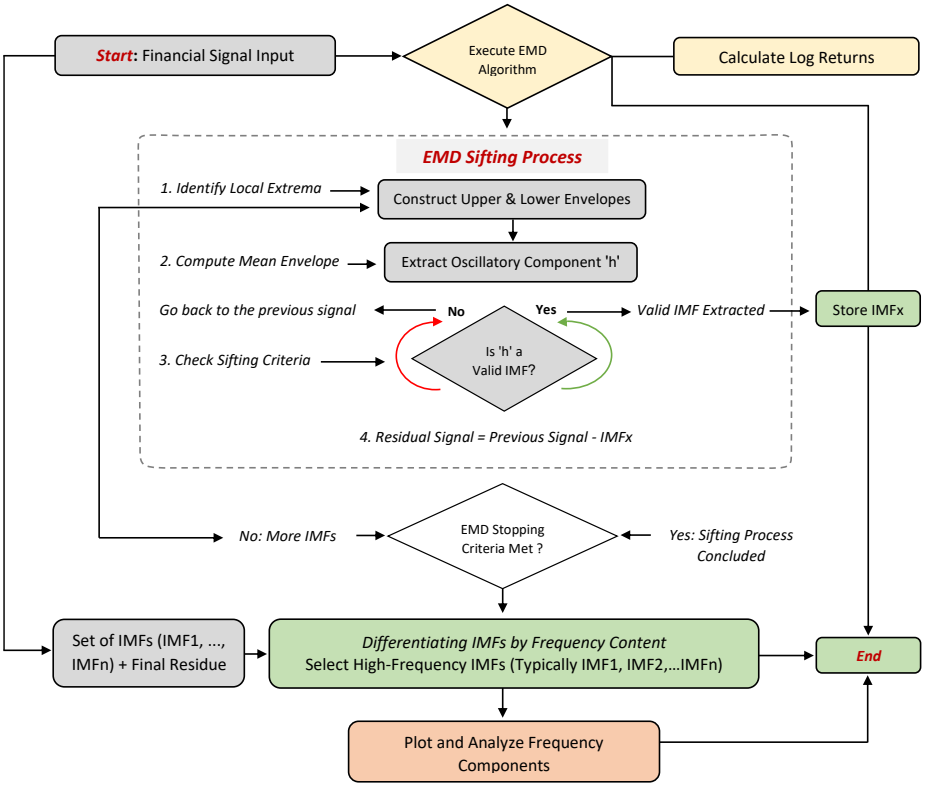


Figure 3. Flowchart for Extracting High-Frequency IMFs Using EMD
Sources: *Authors' own elaboration*

3.5. Random Forest Framework for IMF-Based Volatility Prediction

Figure 4 illustrates the proposed Random Forest (RF) forecasting procedure within the EMD-based hybrid framework, detailing the progression from data input and feature engineering to data preparation, model estimation, prediction aggregation, performance evaluation, and output visualization. Following EMD decomposition, the relevant intrinsic mode function (IMF) components and trend component are used to construct the inputs for forecasting their respective volatility dynamics. Feature engineering is performed to derive informative predictors, after which the time-series data are divided chronologically into training and testing sets, with the first 80% of observations allocated to model estimation and the subsequent 20% reserved for out-of-sample evaluation. This temporal split is performed without shuffling, thereby preserving the chronological order of observations and preventing future observations from entering the training set. The RF model generates an ensemble of decision trees through bootstrap sampling of the training data and random feature-subset selection, thereby introducing diversity among the individual trees. For regression, the tree-level forecasts are aggregated by averaging their predictions

to obtain the final estimates of the IMF and trend volatilities. Predictive performance is assessed using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE), which jointly characterize forecast accuracy from complementary perspectives. Finally, the aggregated forecasts are compared with the corresponding actual values over time to assess the model's ability to reproduce the observed fluctuations in the IMF and trend components. Although the chronological holdout provides an assessment of out-of-sample predictive performance, additional evaluation using walk-forward or rolling-origin validation would further strengthen the robustness assessment by examining forecast accuracy across multiple sequential forecast origins. Overall, this framework provides a systematic basis for evaluating the contribution of the RF forecasting stage within the broader EMD–RF architecture while distinguishing model estimation from out-of-sample predictive evaluation.

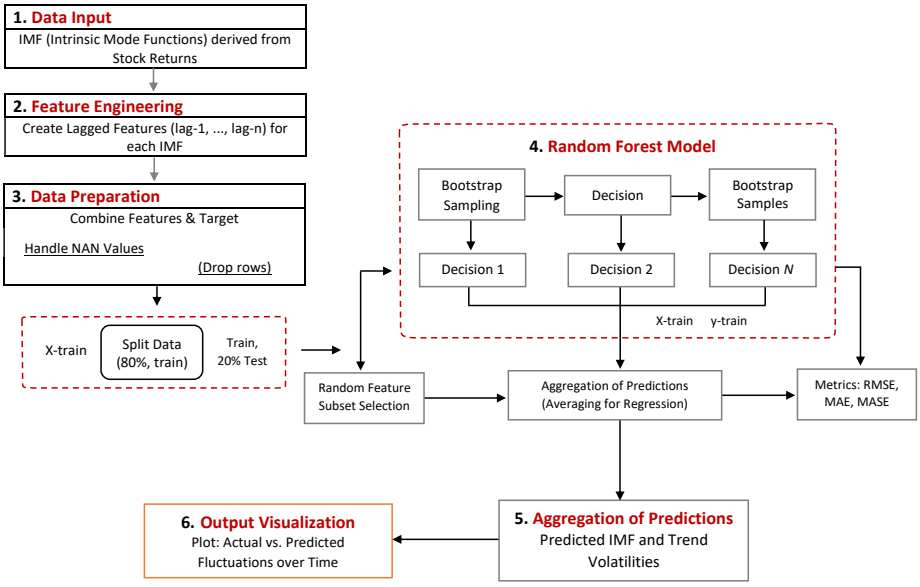


Figure 4. Random Forest Model Methodology for Stock Volatility Prediction Using IMF

Sources: *Authors' own elaboration*

3.6. Model Configuration and Validation

Table 2 summarizes the Random Forest (RF) hyperparameter search space and the experimental protocol used to ensure systematic model optimization and reproducibility. Eight model parameters were specified, with six subjected to optimization and the splitting criterion and random seed held fixed. Hyperparameter selection was conducted using Randomized Search CV with 50 iterations and five-fold time-series cross-validation, using root mean squared error (RMSE) as the optimization criterion. This design accommodates temporal

ordering and reduces the risk of information leakage associated with conventional random cross-validation. Lag selection was guided by the partial autocorrelation function (PACF), using a 95% significance level and a candidate range of 5–20 lags. The data were divided chronologically into training and test sets, comprising 80% and 20% of the observations, respectively. The fixed random seed of 42 supports computational reproducibility. Collectively, these settings establish a structured framework for model selection and evaluation. However, the reliability of the resulting out-of-sample estimates also depends on ensuring that lag selection and any preprocessing steps, including empirical mode decomposition (EMD), are performed without using information from the test period.

Table 2. Random Forest Hyperparameter Configuration and Experimental Settings

Panel A. Random Forest Hyperparameters	
Hyperparameter	Search Space / Value
N-Estimators	{100, 200, 300, 500}
Max-Depth	{5, 10, 15, 20, None}
Min-Samples-Split	{2, 5, 10}
Min-Samples-Leaf	{1, 2, 4}
Max-Features	{'sqrt', 'log2', 0.5, 1.0}
Bootstrap	{True, False}
Criterion	Squared-error (fixed)
Random-State	42 (fixed)
Panel B. Optimization and Experimental Settings	
Hyperparameter Optimization Method	Randomized Search CV
Cross-Validation Strategy	Time Series Split
Number of CV Splits	5
Number of Randomized Search Iterations	50
Optimization Metric	RMSE
Lag Selection Method	Partial Autocorrelation Function (PACF)
PACF Significance Level	95%
Candidate Lag Range	5–20
Training Set / Test Set	80% / 20%
Random Seed	42

Source: *Authors' own elaboration*

3.7. Hybrid EMD–RF Framework for High-Frequency Volatility Forecasting

Figure 5 illustrates the methodological workflow of the proposed hybrid empirical mode decomposition–Random Forest (EMD–RF) framework, from time-series input and signal decomposition to model training and out-of-sample evaluation. In the first stage, the input time series is decomposed using EMD into intrinsic mode functions (IMFs) and a residual component. The resulting output comprises a two-dimensional NumPy array containing the extracted components, the number of IMFs, and the residual, thereby providing a structured representation of the original series at multiple temporal scales. The workflow then identifies the high-frequency IMFs, with logarithmic returns calculated from the financial signal as part of the subsequent processing stage. A further EMD

application is used to identify the high-frequency components. This sequence reflects the framework's emphasis on isolating short-term fluctuations before incorporating them into the machine-learning pipeline. Importantly, the diagram distinguishes the general decomposition output from the subsequent high-frequency identification stage, making explicit the transition from signal representation to component selection.

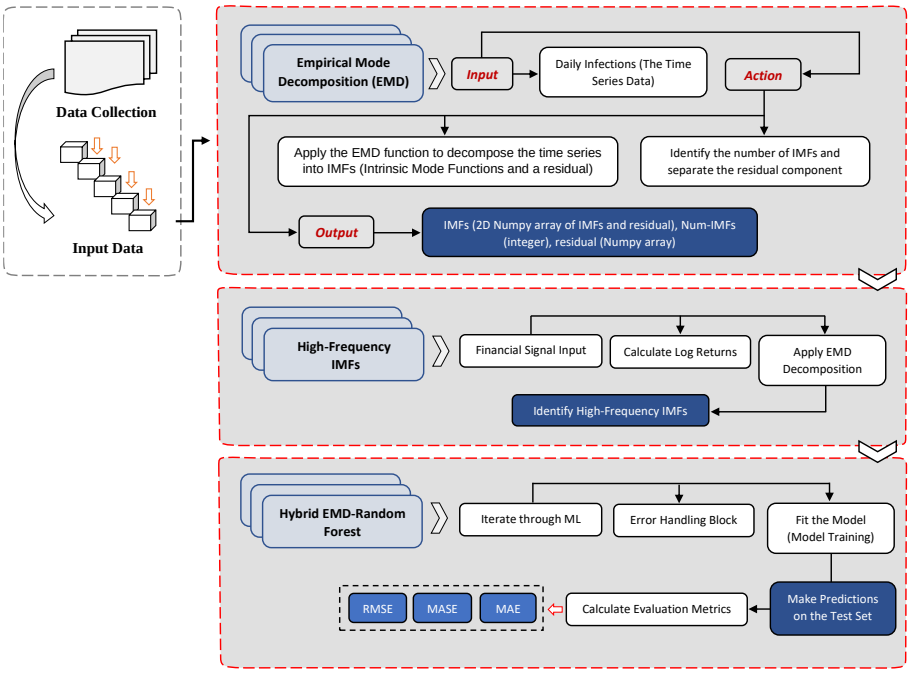


Figure 5. Hybrid EMD-Random Forest Approach in Research Process Flow
Sources: *Authors' own elaboration*

In the final stage, the hybrid EMD–RF workflow combines iterative model execution, error handling, and Random Forest training before generating predictions for the held-out test set. The resulting forecasts are evaluated using root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE). Together, these metrics assess forecast errors from complementary perspectives: RMSE penalizes larger errors more heavily, MAE measures their average absolute magnitude, and MASE expresses forecast accuracy relative to a scaled benchmark. The workflow therefore links frequency-oriented signal processing with nonlinear predictive modeling and a multidimensional evaluation procedure. Its methodological rationale is to investigate whether isolating high-frequency information can provide useful inputs for forecasting. However, the schematic alone does not establish the framework's empirical superiority; that conclusion depends on the comparative results and statistical tests reported elsewhere in the study.

3.8. Accuracy Metrics Analysis

In this study, the predictive accuracy of the hybrid EMD–Random Forest model is evaluated using three standard performance metrics: root mean square error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE).

3.8.1. RMSE (Root Mean Squared Error)

RMSE measures the square root of the mean squared differences between predicted and actual values. It is highly sensitive to large errors, making it particularly useful for detecting the model's ability to capture extreme fluctuations in financial volatility. Lower RMSE indicates closer alignment between predicted and observed market behaviors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

Where:

- y_i ctual observed value of the target variable at instance i
- $\hat{y}_i$ predicted value from the model at instance i
- n: total number of observations

3.8.2. MAE (Mean Absolute Error)

MAE represents the average magnitude of prediction errors, providing a balanced assessment of overall model performance. Unlike RMSE, it is less sensitive to outliers, offering a robust measure of how consistently the model captures typical market movements.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

Where:

- y_i ctual observed value of the target variable at instance i
- $\hat{y}_i$ predicted value from the model at instance i
- n: total number of observations

3.8.3. MASE (Mean Absolute Scaled Error)

MASE evaluates forecasting accuracy by scaling the model's mean absolute error against the mean absolute error of a naïve benchmark forecast calculated from the training data. This scale-free metric facilitates comparisons across forecasting models and time series with different magnitudes. For this study, MASE is calculated as follows:

$$MASE = \frac{\frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|}{\frac{1}{T-1}\sum_{j=2}^{T}\left|y_j^{\text{train}} - y_{j-1}^{\text{train}}\right|} \quad (3)$$

Where:

- y_t: Actual observed value of the target variable at test observation t.
- $\hat{y}_t$: Predicted value of the target variable at test observation t.
- n: Total number of observations in the test set.
- T: Total number of observations in the training set.
- y_j^{train}: Actual value of the target variable at training observation j.
- y_{j-1}^{train}: Actual value at the preceding training observation.

3.9. Statistical Significance and Ablation Analysis

To assess the statistical reliability of the forecasting results and identify the contributions of individual model components, this study employs the Diebold–Mariano (DM) test and a systematic ablation analysis. The DM test evaluates whether the differences in predictive accuracy between the proposed EMD–RF framework and the benchmark models are statistically significant. The test compares the models' forecast errors under a specified loss function, thereby distinguishing statistically meaningful differences in predictive performance from differences that may arise from sampling variability. The analysis is conducted using forecasts evaluated over the same test period, with the loss function and test specification applied consistently across model comparisons. The interpretation of the results accounts for the dependence structure of forecast errors, as appropriate to the forecasting horizon and evaluation design.

The ablation analysis examines the incremental contribution of the decomposition procedure and its constituent components to the overall forecasting performance of the EMD–RF framework. The complete model is compared with a standalone Random Forest model trained on the undecomposed daily log-return series and with alternative specifications that selectively exclude the high-frequency IMFs, low-frequency IMFs, or residual component. The analysis also evaluates the effect of reducing the number of lagged inputs. All configurations are assessed using the same forecasting target, test period, and performance metrics to support consistent comparisons. Differences in predictive accuracy across these configurations provide evidence regarding the relative contribution of each component to the framework's performance. Together, the DM test and ablation analysis serve complementary purposes: the former assesses the statistical significance of differences between competing forecasting models, whereas the latter investigates the contribution of specific architectural components to predictive performance. This combined evaluation provides a more comprehensive basis for interpreting the empirical results beyond comparisons of error metrics alone.

4. Results

4.1. Descriptive Statistics of Stock Market Indices

Table 3 provides a detailed descriptive statistics overview for the dataset used in this study, offering essential insights into the distribution and volatility characteristics of four major international stock market indices: CAC 40, S&P 500, SSE Composite, and TSE. Key statistical measures such as Kurtosis, Skew, Max, Min, Percentiles (25%, 50%, 75%), Standard Deviation (Std), Mean, and Count are presented for each index, shedding light on the market dynamics over the study period. The Kurtosis values reflect the distribution's tail behavior, with all indices exhibiting negative kurtosis, suggesting relatively flat distributions with fewer extreme outliers. The Skew values reveal the asymmetry of each index's return distribution, with CAC 40 showing a negative skew, indicating more frequent lower returns, while S&P 500 and TSE demonstrate a more balanced or slightly positive skew. The Max and Min values reflect the range of the indices' returns, with the CAC 40 reaching the highest maximum return of 8239.990, followed by S&P 500 and SSE Composite, while TSE experienced the lowest returns with a maximum of 3,252.21.

The Standard Deviation values indicate the level of volatility for each index, with TSE exhibiting the highest volatility (867,383.6) compared to the other indices, suggesting greater market fluctuations. In contrast, the SSE Composite has a much lower standard deviation (228.99465), reflecting less volatility. The Mean values show that CAC 40 had the highest average return (6690.126), followed by S&P 500 and SSE Composite, while TSE had the lowest mean return (1,351.46), signifying a less stable market with significant volatility. The Percentiles (25%, 50%, 75%) provide further insight into the data distribution, with CAC 40 and S&P 500 showing relatively higher returns compared to SSE Composite and TSE. These detailed statistics form the foundation for understanding market dynamics and are crucial for accurately forecasting volatility using the hybrid EMD-Random Forest model proposed in this study. All values are thoroughly presented in the table, ensuring a comprehensive understanding of the financial time series data across the four indices.

Table 3. Descriptive Statistics Overview of the Dataset

Index	Count	Mean	Std	Min	25%	50%	75%	Max	Skew	Kurtosis
CAC 40	1432	6690.12	988.183	3754.84	6061.64	6891.12	7428.39	8239.99	-0.68295	-0.30852
S&P 500	1452	4499.08	950.0096	2237.39	3895.58	4358.18	5202.68	6753.72	0.36389	-0.51273
SSE Composite	1353	3240.05	228.9946	2660.16	3067.76	3262.80	3404.03	3715.37	-0.21068	-0.74260
TSE	1842	1,351,4	867,383	92,849	326,469	1,414,2	2,069,6	3,252,2	-0.01166	-1.0452

Source: *Authors'own calculations*

4.2. Annual Market Performance and Volatility Analysis

The tables from 4 to 7 provide a detailed overview of yearly market metrics and volatility analysis for four major international stock market indices: CAC 40, S&P 500, SSE Composite, and TSE. These tables present key statistics such as the number of trading days (NOTD), mean values, minimum and maximum

values, standard deviation (SD), kurtosis, and skewness for each index over multiple years. The data captures the market's behavior, fluctuation patterns, and overall volatility across different periods, which is crucial for understanding both short-term and long-term dynamics. The kurtosis and skewness values, in particular, offer insights into the distribution shape and asymmetry of the returns, providing a deeper understanding of the market's risk characteristics.

Table 4. Yearly Statistics of Trading Days and Market Metrics for CAC 40 Index

Year	NOTD	Mean	Minimum	Maximum	SD	Kurtosis	Skewness
2020	257	5077.86	3754.8400	6111.2402	543.1583	-0.5315	0.23221
2021	258	6427.13	5399.2099	7181.1098	452.6653	-0.6811	-0.4854
2022	257	6434.02	5676.8701	7376.3701	366.3683	-0.3747	0.28262
2023	255	7237.24	6594.5698	7596.9101	190.9424	0.1411	-0.4612
2024	256	7654.35	7130.0400	8239.9902	298.9356	-1.0223	0.42794
2025	149	7775.12	6863.0200	8206.5595	263.6973	0.8769	-0.8182

Source: *Authors' own calculations*

Table 5. Yearly Statistics of Trading Days and Market Metrics for S&P 500 Index

Year	NOTD	Mean	Minimum	Maximum	SD	Kurtosis	Skewness
2020	253	3217.855	2237.399	3756.070	319.226	0.161685	-0.69824
2021	252	4273.385	3700.649	4793.060	287.443	-1.00921	-0.19137
2022	251	4098.514	3577.030	4796.560	291.830	-0.82760	0.369589
2023	250	4283.729	3808.100	4783.350	230.258	-0.87385	0.068361
2024	252	5428.235	4688.680	6090.270	368.070	-0.98150	0.028142
2025	194	6052.013	4982.770	6753.720	389.071	-0.39977	-0.26171

Source: *Authors' own calculations*

Table 6. Yearly Statistics of Trading Days and Market Metrics for SSE Composite Index

Year	NOTD	Mean	Minimum	Maximum	SD	Kurtosis	Skewness
2020	243	3128.419	2660.169	3473.0690	232.80620	-1.4923	-0.2124
2021	243	3539.656	3357.739	3715.3701	74.794839	-0.4158	-0.2034
2022	242	3225.556	2886.429	3632.3300	160.03255	-0.3382	0.43559
2023	242	3175.740	2898.879	3395.0000	116.27905	-0.6395	-0.5262
2024	242	3058.579	2702.185	3489.7749	198.76243	-0.7928	0.3876
2025	141	3362.876	3096.575	3615.7170	103.0328	0.1787	0.3463

Source: *Authors' own calculations*

Table 7. Yearly Statistics of Trading Days and Market Metrics for TSE Index

Year	NOTD	Mean	Minimum	Maximum	SD	Kurtosis	Skewness
2018	241	126,675.30	92,849.93	195,480.1	33,945.6	-1.2606	0.59836
2019	240	245,711.4	156,154.9	377,012.6	62,429.8	-1.1858	0.11151
2020	242	1,170,372	353,807.7	2,078,546	511,101	-1.2877	-0.1897
2021	238	1,307,661	1,095,697	1,575,761	118,005	-0.9729	0.2992
2022	240	1,415,638	1,207,926	1,647,314	96,757	-0.998	-0.0084
2023	239	2,030,954	1,524,795	2,535,236	226,569	-0.0482	-0.60342
2024	239	2,182,562	1,970,901	2,848,900	176,953	4.0607	2.03498
2025	163	2,813,884	2,395,694	3,252,219	212,662	-0.7136	0.21094

Source: *Authors' own calculations*

4.3. Multi-Scale Stock Index Decomposition

Figure 6 presents the Empirical Mode Decomposition (EMD) of the CAC 40 Index, providing a multiscale characterization of the temporal dynamics of the French equity market. This adaptive, data-driven method decomposes the original nonstationary signal into seven intrinsic mode functions (IMFs) and a residual component, each capturing fluctuations at a distinct temporal scale. IMFs 1–3 represent high-frequency fluctuations associated with short-lived market movements, transient trading activity, and abrupt responses to incoming information. These components capture rapid price adjustments and short-term deviations in market behavior, providing a basis for examining high-frequency herding and potential overreactions to information shocks. IMFs 4–6 capture intermediate-frequency fluctuations, reflecting medium-horizon market adjustments associated with changes in investor sentiment, evolving expectations, and portfolio reallocation. IMF 7 represents lower-frequency, persistent oscillations that reflect broader market dynamics associated with macroeconomic conditions, policy developments, and sustained changes in investor confidence. The residual component captures the underlying long-run trend of the CAC 40 Index after the oscillatory components have been extracted. Taken together, this decomposition separates market fluctuations across distinct temporal horizons and provides a multiscale framework for investigating the temporal structure of herding behavior in the French equity market.

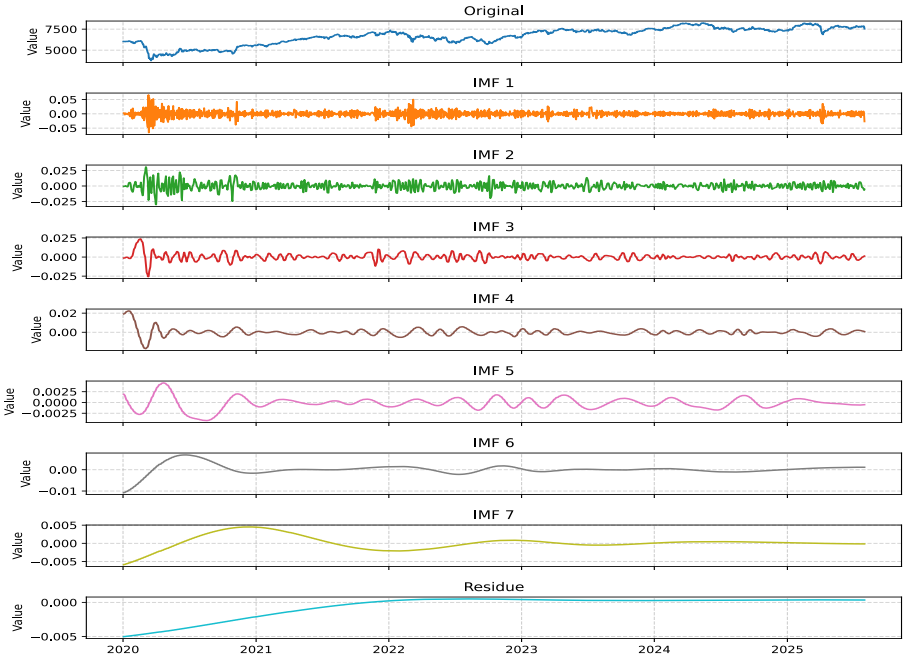


Figure 6. Time-Frequency Analysis of the CAC 40 Index using Empirical Mode Decomposition

Sources: *Authors' own calculations*

Figure 7 presents the Empirical Mode Decomposition (EMD) of the S&P 500 Index, providing a multiscale characterization of the temporal dynamics of the U.S. equity market. This adaptive, data-driven method decomposes the original nonstationary signal into seven intrinsic mode functions (IMFs) and a residual component, each capturing fluctuations at a distinct temporal scale. IMFs 1–3 represent high-frequency fluctuations associated with short-lived market movements, rapid changes in investor sentiment, and immediate price responses to incoming information and market shocks. These components characterize short-horizon market dynamics and provide a basis for examining high-frequency herding and transient collective reactions. IMFs 4–6 capture intermediate-frequency fluctuations, reflecting medium-horizon market adjustments associated with sectoral rotations, portfolio rebalancing, and shifts in collective expectations. IMF 7 represents lower-frequency, persistent oscillations that reflect broader market dynamics associated with macroeconomic conditions, monetary and fiscal policy developments, and sustained changes in investor confidence. The residual component captures the underlying long-run trend of the S&P 500 Index after the oscillatory components have been extracted. Collectively, this decomposition distinguishes market fluctuations across temporal horizons and establishes a multiscale analytical framework for investigating the evolution of volatility and herding behavior in the U.S. equity market.

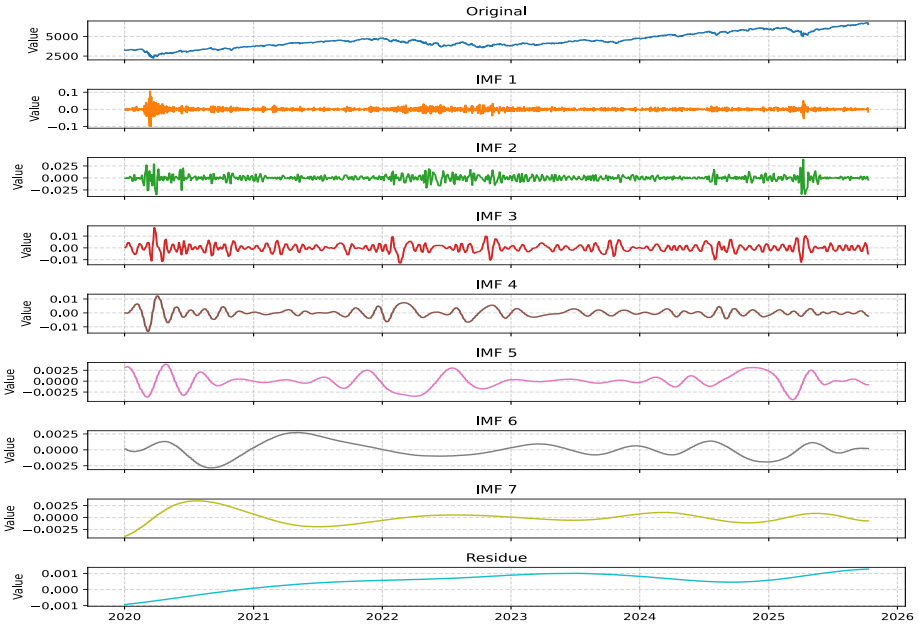


Figure 7. Time-Frequency Analysis of the S&P 500 Index using Empirical Mode Decomposition

Sources: *Authors' own calculations*

Figure 8 illustrates the Empirical Mode Decomposition (EMD) of the SSE Composite Index, providing a multiscale characterization of the nonlinear temporal dynamics of the Chinese equity market. This adaptive, data-driven method decomposes the original nonstationary signal into seven intrinsic mode functions (IMFs) and a residual component, each capturing fluctuations at a distinct temporal scale. IMFs 1–3 represent high-frequency fluctuations associated with short-lived market movements, speculative trading activity, and rapid responses to incoming information. In the Chinese equity market, these fluctuations provide a basis for examining transient herding behavior in response to policy announcements, liquidity shocks, and global financial disturbances. IMFs 4–6 capture intermediate-frequency fluctuations, reflecting medium-horizon market adjustments associated with sectoral reallocation, changes in liquidity conditions, and shifts in investor sentiment. IMF 7 captures lower-frequency, persistent oscillations associated with broader market dynamics, including macroeconomic conditions, structural reforms, and sustained changes in investor confidence. The residual component isolates the underlying long-run trend of the SSE Composite Index after the oscillatory components have been extracted, revealing the market's underlying trajectory independently of shorter-term fluctuations. Collectively, this multiscale decomposition distinguishes fluctuations across temporal horizons and provides a framework for examining the evolution of market volatility and herding behavior in the Chinese equity market, from transient speculative activity to persistent market-wide movements.

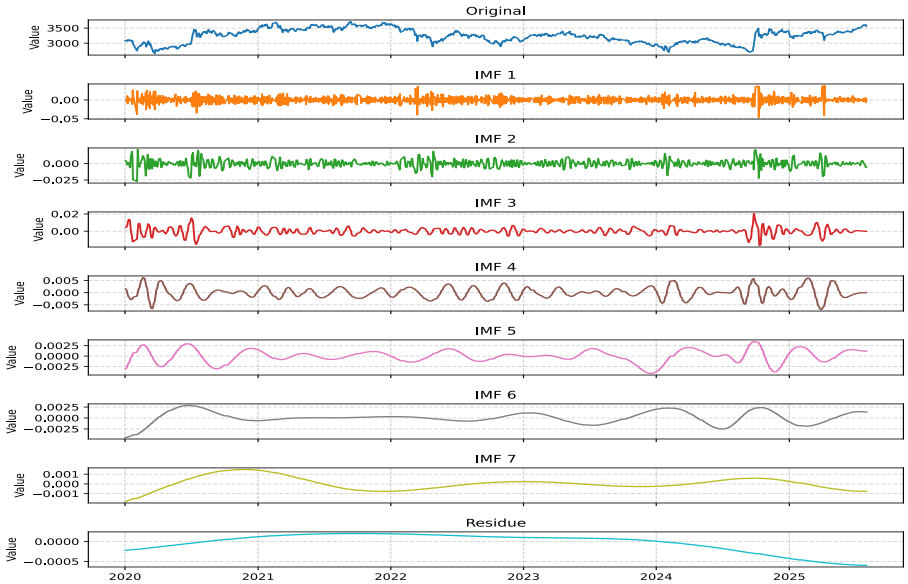


Figure 8. Time-Frequency Analysis of the SSE Composite Index using Empirical Mode Decomposition

Sources: *Authors' own calculations*

Figure 9 presents the Empirical Mode Decomposition (EMD) of the Tehran Stock Exchange (TSE) Index, providing a multiscale characterization of the nonlinear temporal dynamics of the Iranian equity market. This adaptive, data-driven method decomposes the original nonstationary series into seven intrinsic mode functions (IMFs) and a residual component, each capturing fluctuations at a distinct temporal scale. IMFs 1–3 capture high-frequency fluctuations associated with short-lived market movements, speculative trading activity, and rapid price responses to changing market conditions. These components characterize short-horizon market dynamics and provide a basis for examining transient herding behavior in response to abrupt policy changes and macroeconomic uncertainty. IMFs 4–6 capture intermediate-frequency fluctuations, reflecting medium-horizon market adjustments associated with shifts in investor sentiment, capital flow reallocations, and responses to domestic and global economic developments. IMF 7 represents medium-term oscillatory dynamics, capturing more persistent market cycles associated with structural economic changes, monetary policy developments, and sustained shifts in investor confidence. The residual component isolates the underlying long-run trajectory of the TSE Index after the oscillatory components have been extracted, distinguishing the market's broader trend from its cyclical and higher-frequency fluctuations. Collectively, this decomposition distinguishes short-term volatility, intermediate-frequency adjustments, medium-term cycles, and the underlying long-run trend, providing a structured framework for investigating the temporal dynamics of market volatility and herding behavior in the Iranian equity market.

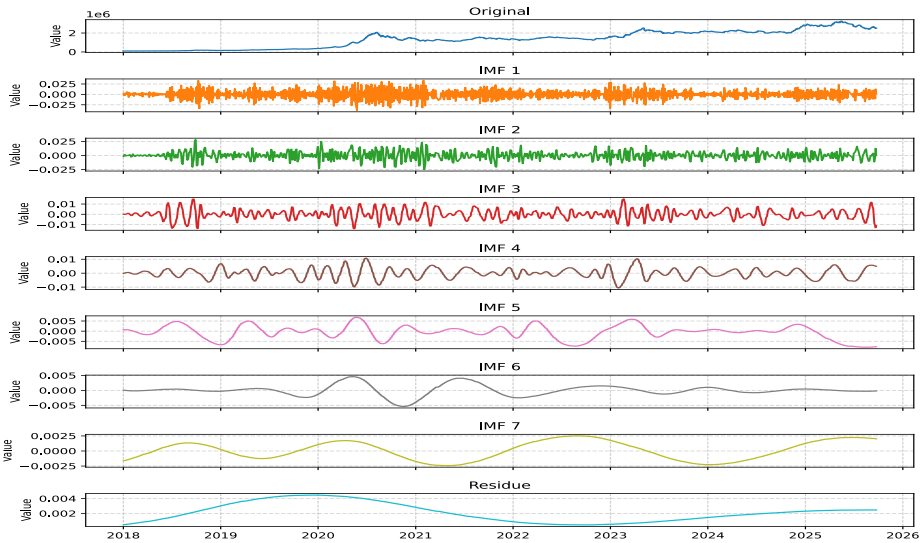


Figure 9. Time-Frequency Analysis of the Tehran Stock Exchange (TSE) Index using Empirical Mode Decomposition

Sources: *Authors' own calculations*

4.4. EMD-Based High- and Low-Frequency IMF Identification

Figure 10 presents the empirical mode decomposition (EMD) of the CAC 40 Index, separating the original time series into high-frequency components (IMF1–IMF3) and lower-frequency components (IMF4–IMF7), with the residual representing the underlying trend. Table 8 quantifies the statistical and spectral characteristics of the extracted IMFs, providing an empirical basis for the proposed frequency-specific forecasting framework. The high-frequency components account for 86.94% of cumulative energy, with IMF1 alone contributing 60.64%. Despite their substantial energy contribution, these components exhibit near-zero correlations with the original series (0.013, 0.010, and 0.004), indicating that their individual oscillatory patterns have limited linear association with the overall index. Their relatively high sample entropy, declining from 1.290 for IMF1 to 0.540 for IMF3, also indicates greater signal complexity in the high-frequency modes. Moreover, their dominant frequencies decrease from 0.419 to 0.031, confirming the progressive reduction in oscillation frequency across these components. Collectively, these findings support treating IMF1–IMF3 as the high-frequency inputs for volatility forecasting, while avoiding the assumption that low correlation alone establishes that these components are purely noise.

Table 8. Intrinsic Mode Function (IMF) Statistics and Signal Characteristics for CAC 40 Index

IMF	Mean	Std	Energy	Dominant Frequency	Hurst Exponent	Sample Entropy	Correlation Original	Energy Percentage	Cumulative Energy
1	0.000115	0.010656	0.162507	0.419287	-0.001056	1.290037	0.013363	60.6353	60.6353
2	0.000024	0.005584	0.044625	0.141859	-0.044030	0.794978	0.010050	16.6507	77.2860
3	0.000262	0.004245	0.025880	0.031447	0.388559	0.540236	0.003890	9.6564	86.9424
4	0.000088	0.003762	0.020268	0.013976	0.841979	0.308511	-0.036366	7.5625	94.5049
5	-0.000179	0.001330	0.002577	0.009085	0.956197	0.185718	0.026588	0.9615	95.4664
6	0.000219	0.002236	0.007220	0.004892	0.993333	0.033856	-0.226462	2.6939	97.3061
7	0.000183	0.001847	0.004930	0.002096	0.997351	0.019518	-0.199108	1.8397	98.1603

Source: *Authors' own calculations*

The lower-frequency components exhibit distinct statistical properties. Sample entropy declines consistently from IMF4 (0.309) to IMF7 (0.020), while the Hurst exponent rises from 0.842 to approximately 0.997, indicating increasingly regular and persistent dynamics as frequency decreases. IMF5–IMF7, in particular, combine low sample entropy with high Hurst exponents (0.956–0.997), consistent with strongly persistent behavior. Their dominant frequencies are also lower than those of IMF1–IMF3, reinforcing the frequency-based distinction. Nevertheless, IMF4's Hurst exponent of 0.842 indicates substantial persistence despite its classification as an intermediate-frequency component. These results support the proposed hybrid framework's separation of complex high-frequency fluctuations from more persistent lower-frequency dynamics, providing a statistical rationale for applying Random Forest to the high-frequency components without implying predictive gains before they are empirically demonstrated.

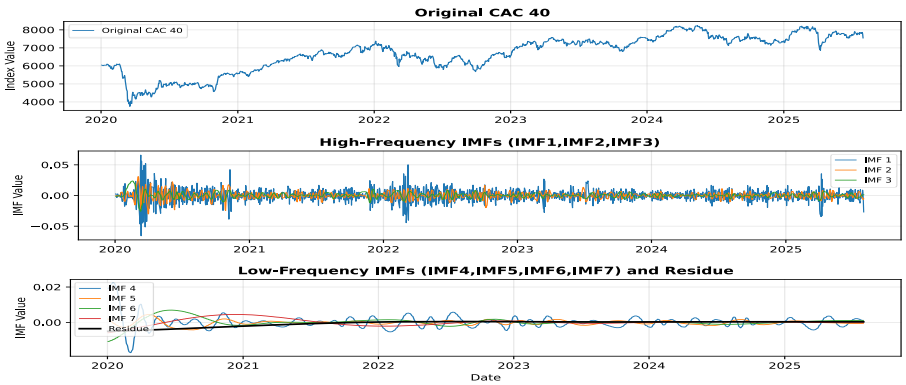


Figure 10. High- and Low-Frequency IMFs Analysis of the for CAC 40 Index
Sources: *Authors' own calculations*

The EMD results in Figure 11 and Table 9 reveal a pronounced separation between high- and low-frequency dynamics in the S&P 500 Index. High-frequency IMFs 1–3 account for 92.87% of cumulative energy, with IMF 1 alone contributing 67.56%, yet exhibit near-zero correlations with the original index (0.032, 0.007, and 0.004, respectively). Their elevated sample entropy (1.147–0.593) and low Hurst exponents (0.013 to 0.232) indicate greater signal complexity and limited persistence. IMF 4 raises cumulative energy to 97.08% but exhibits a Hurst exponent of 0.813, distinguishing it from the less persistent high-frequency components and underscoring the importance of the specified IMF classification. By contrast, IMFs 5–7 display progressively lower sample entropy (0.150–0.039) and higher Hurst exponents (0.958–0.995), indicating increasingly regular and persistent dynamics. The residual component accounts for 0.89% of total energy and has a negative correlation with the original index (−0.255), distinguishing its long-run trajectory from the oscillatory modes. These results support a frequency-specific modeling strategy that separates complex short-term fluctuations from more persistent lower-frequency dynamics.

Table 9. Intrinsic Mode Function (IMF) Statistics and Signal Characteristics for S&P 500 Index

IMF	Mean	Std	Energy	Dominant Frequency	Hurst Exponent	Sample Entropy	Correlation Original	Energy Percentage	Cumulative Energy
1	-0.000139	0.011312	0.185704	0.443143	0.012573	1.147259	0.032402	67.5593	67.5593
2	-0.000081	0.006080	0.053643	0.137147	-0.044937	0.729279	0.007359	19.5154	87.0747
3	0.000037	0.003312	0.015919	0.059959	0.232081	0.593177	0.004216	5.7913	92.8659
4	0.000093	0.002824	0.011586	0.025500	0.813429	0.382413	-0.013741	4.2151	97.0810
5	-0.000013	0.001565	0.003556	0.006892	0.958470	0.150317	0.082948	1.2937	98.3747
6	0.000029	0.001177	0.002012	0.002757	0.991572	0.072247	-0.033602	0.7318	99.1066
7	0.000045	0.001300	0.002456	0.002068	0.994628	0.039215	-0.254862	0.8934	99.1066

Source: *Authors'own calculations*

The spectral characteristics further reinforce this distinction: dominant frequency declines from 0.443 for IMF1 to 0.002 for IMF7, indicating a

progressive shift toward slower oscillations. The combination of declining dominant frequency and sample entropy with increasing Hurst exponents across the later IMFs is consistent with increasingly smooth and persistent dynamics. Accordingly, IMF1–IMF3 provide the designated inputs for high-frequency volatility forecasting, whereas IMF4–IMF7 represent intermediate- to lower-frequency fluctuations.

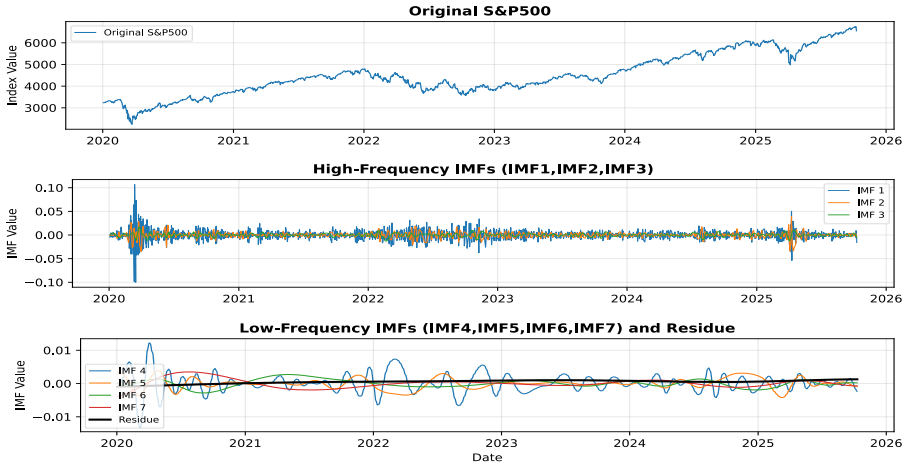


Figure 11. High- and Low-Frequency IMFs Analysis of the for S&P 500 Index
Sources: *Authors' own calculations*

Based on the decomposition presented in Figure 12 and the corresponding statistical summary in Table 10, the empirical mode decomposition (EMD) of the SSE Composite Index separates the series into components with distinct frequency and statistical characteristics, providing an empirical basis for the proposed modeling framework. The high-frequency IMFs (IMF1–IMF3) account for 92.59% of cumulative energy, with IMF1 alone contributing 59.87%, highlighting the substantial energy concentrated in the fastest oscillatory modes. Despite this contribution, their correlations with the original index are close to zero (0.0646, 0.0232, and 0.0082, respectively), indicating weak linear associations with the overall series. Their sample entropy values decline from 1.3983 for IMF1 to 0.5309 for IMF3, indicating decreasing complexity across the high-frequency components. The Hurst exponents of IMF1 and IMF2 are slightly negative (−0.0038 and −0.0660), while IMF3 has a value of 0.1976, suggesting limited persistence in these modes. IMF4 marks a transition toward more persistent behavior, with a Hurst exponent of 0.8076, and therefore should not be classified as a high-frequency component under the specified grouping.

Table 10. Intrinsic Mode Function (IMF) Statistics and Signal Characteristics for SSE Composite Index

IMF	Mean	Std	Energy	Dominant Frequency	Hurst Exponent	Sample Entropy	Correlation Original	Energy Percentage	Cumulative Energy
1	0.0001	0.0086	0.1004	0.4763	-0.0038	1.3983	0.0646	59.8680	59.8680
2	-0.0001	0.0052	0.0360	0.0991	-0.0660	0.8294	0.0232	21.4518	81.3198
3	-0.0001	0.0037	0.0189	0.0459	0.1976	0.5309	0.0082	11.2685	92.5883
4	-0.0000	0.0022	0.0066	0.0214	0.8076	0.4262	-0.0142	3.9167	96.5050
5	-0.0001	0.0015	0.0030	0.0089	0.9517	0.1917	0.0708	1.7718	98.2768
6	0.0001	0.0013	0.0023	0.0037	0.9943	0.0719	-0.0352	1.3957	98.6043
7	0.00001	0.0006	0.0005	0.0022	0.9955	0.0321	0.0835	0.3275	99.6725

Source: *Authors'own calculations*

The lower-frequency IMFs (IMF5–IMF7) exhibit substantially higher Hurst exponents (0.9517–0.9955) and lower sample entropy (0.1917–0.0321), indicating increasingly persistent and regular dynamics. Their dominant frequencies also decline from 0.0089 to 0.0022, consistent with progressively slower oscillations. These results support distinguishing the complex, rapidly oscillating high-frequency components from the more persistent lower-frequency modes. Accordingly, IMF1–IMF3 are the designated inputs for high-frequency volatility forecasting, while IMF4–IMF7 represent the transition toward intermediate- and lower-frequency dynamics. This decomposition provides a statistical rationale for the proposed modeling strategy, although its predictive value must be established through out-of-sample evaluation.

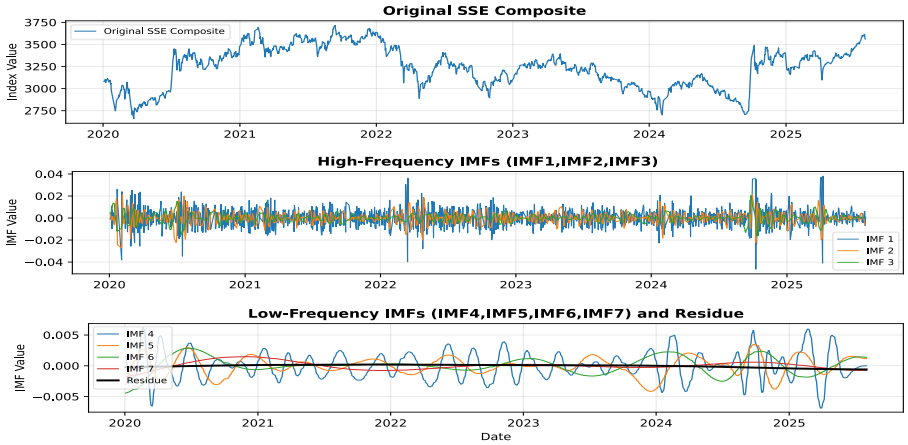


Figure 12. High- and Low-Frequency IMFs Analysis of the for SSE Composite Index
Sources: *Authors' own calculations*

Based on the decomposition of the TSE Index depicted in Figure 13 and the statistical evidence reported in Table 11, the empirical mode decomposition (EMD) separates the series into components with distinct frequency characteristics, providing a statistical basis for the proposed forecasting

framework. The high-frequency IMFs (IMF1–IMF3) account for 84.55% of cumulative energy, with IMF1 alone contributing 49.86%. Despite their substantial energy contribution, these components exhibit weak correlations with the original index (−0.0103, −0.0169, and 0.0041, respectively), indicating limited linear association with the overall series. Their sample entropy values decline from 1.3304 for IMF1 to 0.5347 for IMF3, reflecting decreasing complexity across the high-frequency modes. The Hurst exponents of IMF1 and IMF2 are slightly negative (−0.0273 and −0.0128), whereas IMF3 has a higher value of 0.4574, indicating a transition toward greater persistence. IMF4, with a Hurst exponent of 0.8909, exhibits substantially stronger persistence and therefore belongs to the intermediate-frequency group rather than the designated high-frequency category.

Table 11. Intrinsic Mode Function (IMF) Statistics and Signal Characteristics for TSE Index

IMF	Mean	Std	Energy	Dominant Frequency	Hurst Exponent	Sample Entropy	Correlation Original	Energy Percentage	Cumulative Energy
1	-0.0001	0.0098	0.1763	0.2357	-0.0273	1.3304	-0.0103	49.8635	49.8635
2	0.0001	0.0067	0.0822	0.0831	-0.0128	0.7145	-0.0169	23.2631	73.1266
3	-0.0001	0.0047	0.0404	0.0331	0.4574	0.5347	0.0041	11.4264	84.553
4	0.0001	0.0036	0.0238	0.012	0.8909	0.3882	0.01	6.737	91.29
5	-0.0005	0.0033	0.0208	0.0043	0.9836	0.1083	-0.179	5.8869	97.1769
6	0.00001	0.0018	0.0058	0.0033	0.9981	0.0463	-0.0379	1.6359	98.8128
7	0.0001	0.0015	0.0042	0.0016	1.0001	0.0281	0.0774	1.1872	99.9811

Source: *Authors'own calculations*

The lower-frequency components (IMF5–IMF7) exhibit Hurst exponents ranging from 0.9836 to 1.0001, indicating highly persistent dynamics, alongside low sample entropy values that suggest comparatively regular behavior. Their dominant frequencies also decline progressively, reaching 0.0016 for IMF7, consistent with increasingly slow oscillations. These characteristics distinguish the lower-frequency modes from the more complex high-frequency components and support a frequency-specific modeling strategy. Accordingly, IMF1–IMF3 are the designated inputs for high-frequency volatility forecasting, while IMF4–IMF7 capture increasingly persistent intermediate- to lower-frequency dynamics. The decomposition provides a statistical rationale for this separation, although its practical forecasting value must be established through out-of-sample performance evaluation.

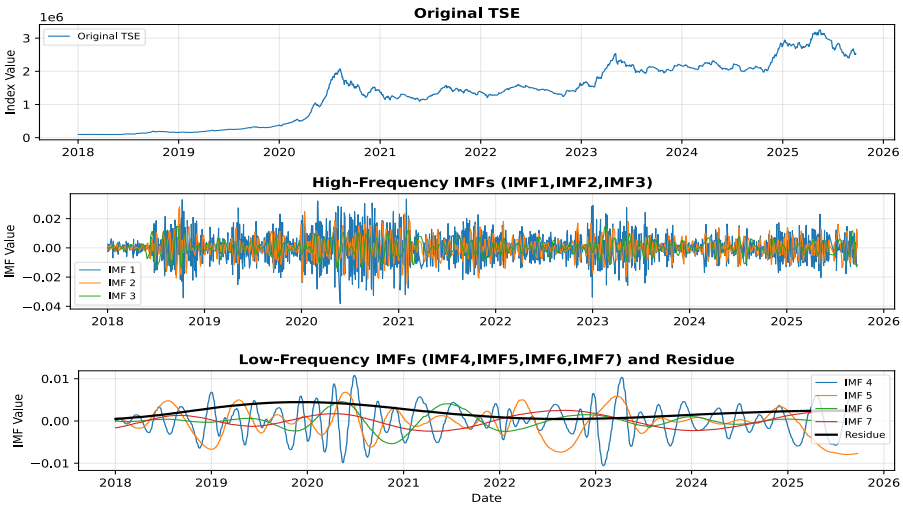


Figure 13. High- and Low-Frequency IMFs Analysis of the TSE Index

Sources: *Authors' own calculations*

4.5. Hybrid EMD–RF Forecasting Superiority Across Stock Indices

Table 12 provides strong empirical evidence of the forecasting effectiveness of the proposed EMD–RF model across four major equity markets, with the model achieving the lowest RMSE, MAE, and MASE in all four markets. For the CAC 40, EMD–RF records an RMSE of 0.009569, an MAE of 0.006910, and a MASE of 0.536268, outperforming all three benchmark models across every evaluation metric. The Diebold–Mariano (DM) test results further indicate that these improvements are statistically significant relative to SVM ((DM=2.9567), (p=0.0031)), XGBoost ((DM=4.2882), (p<0.0001), subject to verification of the unrounded p-value), and RF ((DM=2.4073), (p=0.0161)). For the S&P 500, EMD–RF again achieves the lowest errors, with an RMSE of 0.011124, an MAE of 0.007435, and a MASE of 0.552866. Although the numerical results consistently favor the proposed model, the DM tests establish statistically significant superiority only over XGBoost ((DM=2.2000), (p=0.0278)); the differences relative to SVM ((p=0.5204)) and RF ((p=0.4258)) are not statistically significant at the 5% level. A similar pattern emerges for the SSE Composite, where EMD–RF attains the lowest RMSE (0.011460), MAE (0.007254), and MASE (0.686035), demonstrating consistent numerical gains across all three error measures. However, none of the corresponding DM tests is statistically significant at conventional levels, with p-values of 0.3712 for SVM, 0.2070 for XGBoost, and 0.3449 for RF. The strongest overall performance is observed in the TSE market, where EMD–RF achieves an RMSE of 0.008252, an MAE of 0.006582, and a MASE of 0.560585, substantially outperforming all benchmark models across every metric. The corresponding DM statistics are 6.7071 against SVM, 6.9999 against XGBoost, and 6.6006 against RF, with all reported p-values displayed as 0.0000, indicating strong statistical evidence against equal predictive

accuracy; the exact p-values should be reported using an appropriate threshold rather than as zero. Taken together, these findings establish EMD–RF as the best-performing model in terms of point-forecast error across all four markets, while the DM results provide particularly strong evidence of statistically significant gains in the CAC 40 and TSE markets. Although the improvements over SVM and RF in the S&P 500 and the gains against all benchmarks in the SSE Composite are not statistically significant at the 5% level, the consistently lower error measures suggest that the proposed framework offers a promising and broadly applicable approach to equity-index forecasting. The cross-market evidence is consistent with the potential value of combining empirical mode decomposition with random forests, while also highlighting the importance of accounting for market-specific differences when assessing statistical forecasting gains.

Table 12. Forecasting Performance Comparison and Diebold–Mariano Test Results

Panel A: CAC 40 Forecasting Performance

Model	RMSE	MAE	MASE	DM statistic	DM p-value
EMD-RF (Proposed)	0.009569	0.006910	0.536268	—	—
Support Vector Machine	0.010256	0.007673	0.586850	2.9567	0.0031
XGBoost	0.011515	0.008406	0.656044	4.2882	p<0.0001
Random Forest (RF)	0.010840	0.007874	0.593313	2.4073	0.0161

Panel B: S&P 500 Forecasting Performance

Model	RMSE	MAE	MASE	DM statistic	DM p-value
EMD-RF (Proposed)	0.011124	0.007435	0.552866	—	—
Support Vector Machine	0.011425	0.007524	0.556892	0.6427	0.5204
XGBoost	0.012703	0.007767	0.570441	2.2000	0.0278
Random Forest (RF)	0.011541	0.007597	0.559575	0.7964	0.4258

Panel C: SSE Composite Forecasting Performance

Model	RMSE	MAE	MASE	DM statistic	DM p-value
EMD-RF (Proposed)	0.011460	0.007254	0.686035	—	—
Support Vector Machine	0.012007	0.007335	0.689490	0.8942	0.3712
XGBoost	0.013549	0.007492	0.698663	1.2618	0.2070
Random Forest (RF)	0.012204	0.007389	0.693275	0.9446	0.3449

Panel D: TSE Forecasting Performance

Model	RMSE	MAE	MASE	DM statistic	DM p-value
EMD-RF (Proposed)	0.008252	0.006582	0.560585	—	—
Support Vector Machine	0.011277	0.008843	0.757845	6.7071	p<0.0001
XGBoost	0.011784	0.009209	0.789213	6.9999	p<0.0001
Random Forest (RF)	0.011035	0.008676	0.743576	6.6006	p<0.0001

Sources: *Authors' own calculations*

4.6. Ablation Analysis of the Hybrid EMD–RF Framework Across Stock Indices

Table 13 presents an ablation analysis designed to quantify the contributions of empirical mode decomposition (EMD), frequency-specific intrinsic mode functions (IMFs), the residual trend component, and the selected component structure to the forecasting performance of the proposed EMD–RF framework across four major equity markets. For the CAC 40, removing EMD and applying random forests directly

to the raw series increases RMSE, MAE, and MASE by 8.099%, 10.434%, and 10.366%, respectively, highlighting the predictive value of decomposition-based preprocessing. Excluding either high-frequency or low-frequency IMFs also deteriorates performance across all three metrics, with the removal of high-frequency IMFs producing larger increases in forecast errors. By contrast, excluding the residual trend results in only marginal changes, increasing RMSE, MAE, and MASE by 0.052%, 0.260%, and 0.022%, respectively. This pattern suggests that the residual component contributes relatively little incremental predictive information for the CAC 40 under the evaluated configuration, whereas the frequency-specific IMFs play a more substantive role. For the S&P 500, the full EMD–RF model achieves the lowest values across all three error metrics. Removing EMD increases RMSE by 11.219%, MAE by 3.309%, and MASE by 3.301%, while excluding high-frequency IMFs raises the corresponding errors by 6.446%, 0.753%, and 1.936%. Removing low-frequency IMFs or the residual trend likewise worsens all three metrics. These consistent increases indicate that both frequency-specific information and the residual component contribute to predictive accuracy in this market, although their relative effects differ across evaluation metrics. The SSE Composite exhibits a similarly informative but more heterogeneous pattern. Removing EMD increases RMSE, MAE, and MASE by 10.454%, 3.984%, and 1.903%, respectively, while excluding low-frequency IMFs produces particularly pronounced deterioration in MAE (8.712%) and MASE (5.950%). Excluding high-frequency IMFs has comparatively modest effects, increasing RMSE by 2.871%, MAE by 0.469%, and MASE by 1.520%. Removing the residual trend also increases all three error measures, although the changes are smaller than those associated with removing EMD or the low-frequency IMFs for most metrics. Taken together, the CAC 40, S&P 500, and SSE Composite results demonstrate that the full model provides the most favorable overall configuration among the variants examined, while the magnitude of each component's contribution varies across markets and loss measures.

Table 13. Ablation Study of the EMD-RF Forecasting Framework

Panel A: CAC 40 Forecasting Performance

Model Variant	RMSE	MAE	MASE	Δ RMSE (%)	Δ MAE (%)	Δ MASE (%)
Full EMD-RF	0.009569	0.006910	0.536268	0.000	0.000	0.000
Raw RF (No EMD)	0.010344	0.007631	0.591856	8.099	10.434	10.366
w/o High-Freq IMFs	0.009998	0.007278	0.564450	4.483	5.326	5.255
w/o Low-Freq IMFs	0.009855	0.007139	0.553729	2.989	3.314	3.256
w/o Residue (Trend)	0.009574	0.006928	0.536387	0.052	0.260	0.022

Panel B: S&P 500 Forecasting Performance

Model Variant	RMSE	MAE	MASE	Δ RMSE (%)	Δ MAE (%)	Δ MASE (%)
Full EMD-RF	0.011124	0.007435	0.552866	0.000	0.000	0.000
Raw RF (No EMD)	0.012372	0.007681	0.571117	11.219	3.309	3.301
w/o High-Freq IMFs	0.011841	0.007491	0.563567	6.446	0.753	1.936
w/o Low-Freq IMFs	0.011723	0.007527	0.559699	5.385	1.237	1.236
w/o Residue (Trend)	0.011537	0.007554	0.561690	3.713	1.601	1.596

Panel C: SSE Composite Forecasting Performance						
Model Variant	RMSE	MAE	MASE	Δ RMSE (%)	Δ MAE (%)	Δ MASE (%)
Full EMD-RF	0.011460	0.007254	0.686035	0.000	0.000	0.000
Raw RF (No EMD)	0.012658	0.007543	0.699089	10.454	3.984	1.903
w/o High-Freq IMFs	0.011789	0.007288	0.696463	2.871	0.469	1.520
w/o Low-Freq IMFs	0.011849	0.007886	0.726852	3.394	8.712	5.950
w/o Residue (Trend)	0.011594	0.007533	0.694124	1.169	3.846	1.179
Panel D: TSE Forecasting Performance						
Model Variant	RMSE	MAE	MASE	Δ RMSE (%)	Δ MAE (%)	Δ MASE (%)
Full EMD-RF	0.008252	0.006582	0.560585	0.000	0.000	0.000
Raw RF (No EMD)	0.011136	0.008752	0.745387	34.949	32.969	32.966
w/o High-Freq IMFs	0.010390	0.008125	0.691952	25.909	23.443	23.434
w/o Low-Freq IMFs	0.009326	0.007272	0.619309	13.015	10.483	10.475
w/o Residue (Trend)	0.008316	0.006706	0.571150	0.776	1.884	1.885

Sources: *Authors' own calculations*

The TSE provides the strongest evidence of the predictive value of the proposed framework. Removing EMD increases RMSE, MAE, and MASE by 34.949%, 32.969%, and 32.966%, respectively, representing the largest deterioration observed across the four markets. Excluding high-frequency IMFs also substantially increases forecast errors, by 25.909% in RMSE, 23.443% in MAE, and 23.434% in MASE. Removing low-frequency IMFs produces further increases of 13.015%, 10.483%, and 10.475%, respectively, indicating that both short-term fluctuations and lower-frequency dynamics contribute materially to predictive accuracy in this market. Excluding the residual trend leads to smaller but consistent increases of 0.776% in RMSE, 1.884% in MAE, and 1.885% in MASE. Overall, the ablation results show that the complete EMD–RF framework achieves the lowest forecast errors across all three evaluation metrics in each of the four markets, supporting the combined value of decomposition-based preprocessing and the inclusion of the extracted signal components. The magnitude of the improvements is particularly pronounced for the TSE, while the results for the CAC 40, S&P 500, and SSE Composite reveal meaningful differences in the relative contributions of high-frequency, low-frequency, and residual information. These cross-market patterns suggest that EMD-based preprocessing offers a consistent empirical advantage within the experimental design, while the extent to which individual components contribute to forecasting accuracy depends on the characteristics of the market under study. Accordingly, the findings support the effectiveness of the proposed hybrid framework and motivate further investigation of market-specific component selection and model configuration. The forecasting results for the CAC 40 index are illustrated in Figure 14, which presents the predicted values and facilitates a visual assessment of the model's forecasting performance.

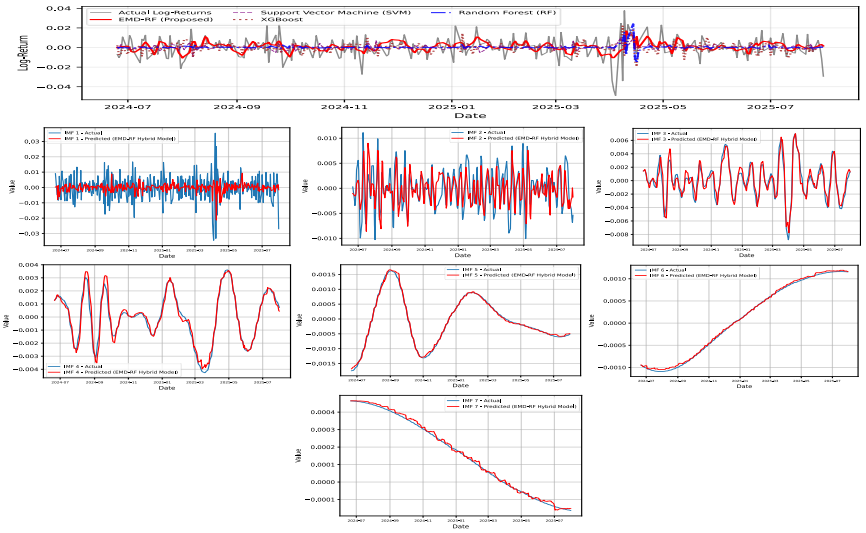


Figure 14. High- and Low-Frequency IMF Forecasting of CAC 40 Index Using Hybrid EMD-RF Model

Sources: *Authors' own calculations*

The forecasting results for the S&P 500 index, as described above, are illustrated in Figure 15, showing the predicted values.

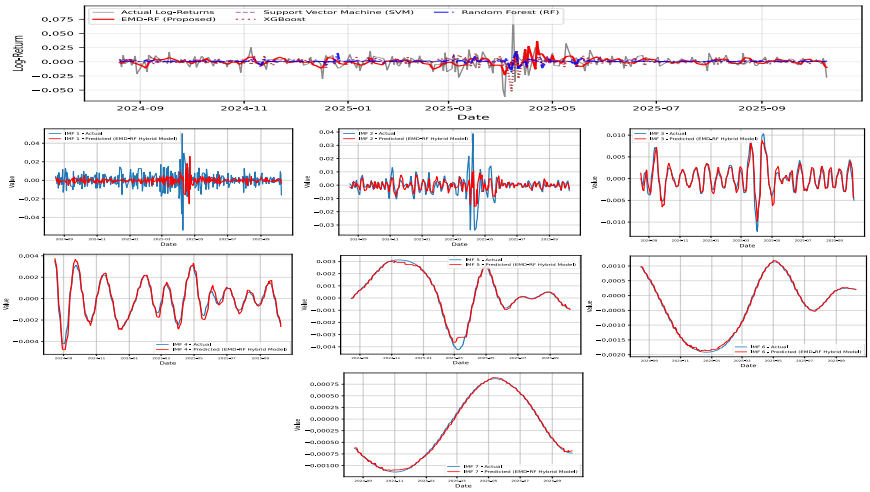


Figure 15. High- and Low-Frequency IMF Forecasting of S&P 500 Index Using Hybrid EMD-RF Model

Sources: *Authors' own calculations*

The forecasting results for the SSE Composite index, as described above, are illustrated in Figure 16, showing the predicted values.

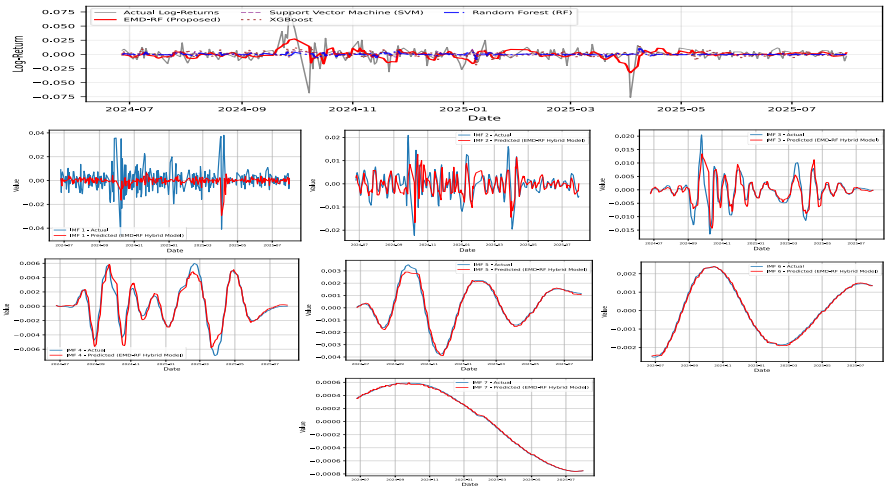


Figure 16. High- and Low-Frequency IMF Forecasting of SSE Composite Index Using Hybrid EMD-RF Model

Sources: *Authors' own calculations*

The forecasting results for the TSE index, as described above, are illustrated in Figure 17, showing the predicted values.

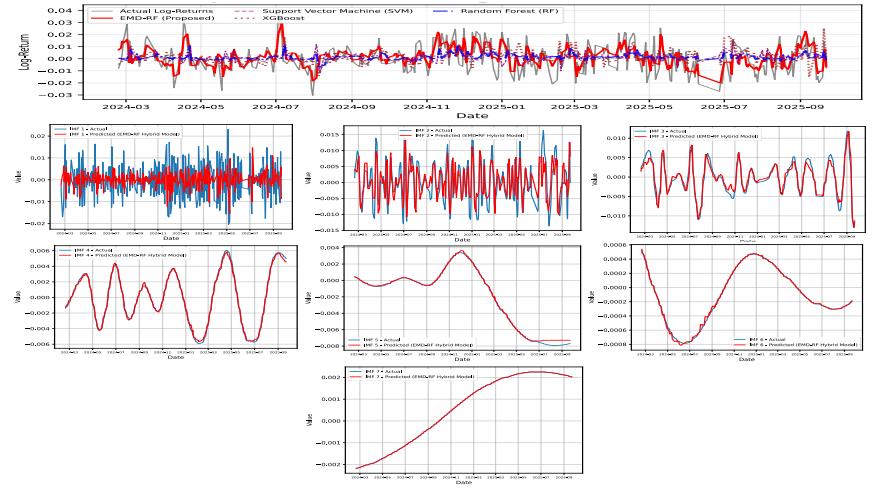


Figure 17. High- and Low-Frequency IMF Forecasting of TSE Index Using Hybrid EMD-RF Model

Sources: *Authors' own calculations*

5. Discussion and Conclusion

The findings establish empirical mode decomposition (EMD) as a valuable component of the proposed hybrid forecasting framework, supporting the premise that equity index prediction benefits from representing financial time series across

multiple temporal scales. The ablation analysis indicates that the contribution of EMD extends beyond preprocessing: decomposing the original series enables the Random Forest (RF) model to exploit predictive information more effectively than modeling the undecomposed series. This improvement is consistent across all four markets and all three evaluation metrics, highlighting the robustness of the decomposition-based representation. Equally important, excluding either high-frequency or low-frequency intrinsic mode functions (IMFs) increases forecasting errors in every market, demonstrating that both frequency groups contribute complementary information to the prediction process. Their relative contributions, however, vary with market characteristics and evaluation criteria. For the SSE Composite, removing low-frequency IMFs produces particularly pronounced deterioration in MAE and MASE, underscoring the importance of lower-frequency information for accurate point prediction in this market. For the TSE, excluding high-frequency IMFs leads to substantial increases across all three error metrics, highlighting the predictive relevance of higher-frequency fluctuations. The residual component also contributes to the full framework, although its incremental effect varies across indices and is comparatively modest for the CAC 40. Taken together, these findings support H1: EMD-based decomposition improves forecasting accuracy by organizing the original series into informative components, while the predictive contributions of high-frequency IMFs, low-frequency IMFs, and the residual depend on the market under consideration. This component-level evidence advances a multiscale interpretation of equity index forecasting, in which predictive performance emerges from retaining complementary temporal information rather than relying on a single frequency range.

The comparative evaluation further supports H2 by showing that EMD–RF achieves the lowest RMSE, MAE, and MASE among the examined models for each of the four equity indices. Its consistent ranking relative to standalone Support Vector Machine (SVM), XGBoost, and RF models indicates that the proposed framework's advantage is not confined to one benchmark or one error criterion. The Diebold–Mariano (DM) tests add an important inferential dimension to this comparison. For the CAC 40 and TSE, EMD–RF's predictive-accuracy gains are statistically significant against all three benchmarks. For the S&P 500, the evidence establishes a statistically significant advantage over XGBoost, while the pairwise differences relative to SVM and RF are not statistically significant at conventional levels. For the SSE Composite, EMD–RF records the lowest errors, although the DM tests do not establish statistically significant pairwise differences against the evaluated benchmarks. These results offer a coherent interpretation of the model's performance: the direction of the observed error differences consistently favors EMD–RF, whereas the statistical strength of that advantage varies across markets and competing algorithms. The particularly clear gains for the TSE and CAC 40 demonstrate the value of combining decomposition with nonlinear learning in these settings, while the results for the S&P 500 and SSE Composite show why point-forecast rankings

and formal tests of predictive accuracy should be considered jointly. Accordingly, H2 is supported by the uniformly lower observed forecasting errors of EMD–RF, with the DM evidence specifying where its comparative advantage is most strongly established.

The principal contribution of this study is to connect model-level forecasting performance with a systematic examination of the temporal information that underpins that performance. Benchmark comparisons can identify which model produces the lowest errors, but component-level ablation provides additional insight into why a decomposition-based architecture may be effective and which elements of the decomposed signal matter most. The present findings show that EMD contributes value beyond the standalone RF baseline and that both high- and low-frequency IMFs improve the performance of the complete hybrid model across all four indices. At the same time, the variation in component contributions demonstrates that the predictive value of temporal information is market-dependent. This distinction is central to the interpretation of hybrid financial forecasting: decomposition is not merely a technical transformation of the input series, but a means of representing different forms of temporal variation in a way that supports nonlinear prediction. From a modeling perspective, the results motivate further investigation of market-adaptive forecasting architectures that preserve multiple frequency bands while allowing their relative contributions to reflect the characteristics of individual indices. From an applied perspective, the consistently lower statistical errors establish EMD–RF as a strong candidate for equity index point forecasting among the methods evaluated here. Any extension from improved forecast accuracy to trading profitability, portfolio performance, or risk-management benefits would, however, require evaluation with measures specifically designed for those objectives. This distinction preserves the central implication of the findings while separating evidence of predictive accuracy from claims about downstream economic outcomes.

In conclusion, the empirical evidence supports both hypotheses and identifies EMD–RF as the best-performing forecasting framework among the evaluated alternatives across the CAC 40, S&P 500, SSE Composite, and TSE. The results demonstrate that EMD-based decomposition improves the predictive performance of RF and that the high-frequency and low-frequency IMFs contain complementary information whose relative importance varies across markets. The residual component provides additional predictive information to differing degrees, further illustrating the value of examining the decomposed series at the component level. The DM tests complement the error-based comparisons by showing that the statistical significance of EMD–RF's advantage is strongest for the CAC 40 and TSE, selective for the S&P 500, and not statistically established for the SSE Composite. The overall evidence therefore supports a clear conclusion: integrating EMD with Random Forests yields consistently lower observed forecasting errors across the four examined indices, while the inferential strength of the gains remains market- and benchmark-dependent. By combining cross-market model evaluation with frequency-specific ablation, this study

provides a more detailed understanding of how signal decomposition and nonlinear learning work together in equity index forecasting. Its central contribution is to demonstrate not only the forecasting value of the hybrid architecture, but also the complementary and market-dependent roles of the temporal components from which its predictions are constructed.

Table 14. Abbreviation

CAC 40	Cotation Assistée en Continu 40 (French Stock Market Index)
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Function
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
RF	Random Forest
RMSE	Root Mean Square Error
SD	Standard Deviation
S&P 500	Standard & Poor's 500 Index
SSE Composite	Shanghai Stock Exchange Composite Index
TSE	Tehran Stock Exchange

Source: *Authors' own elaboration*

Author Contributions

Author Contributions: Mohammad Ahadzadeh (MA) contributed to conceptualization, methodology, formal analysis, data curation, and writing of the original draft. Fatemeh Ahadzadeh (FA) contributed to interpretation of results and writing and editing. Ruhollah Moradi (RM) contributed to supervision, validation, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

No external funding was received.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data were collected from the Rahavard 365 database and are available from the corresponding author upon reasonable request.

Acknowledgements

We acknowledge that artificial intelligence tools (ChatGPT) were used solely for English language editing in this article.

References

Campisi, G., Muzzioli, S., & De Baets, B. (2024). A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices. *International Journal of Forecasting*, 40(3), 869-880. https://doi.org/10.1016/j.ijforecast.2023.07.002.

Case, Justin and Clements, Adam, Stock Market Volatility and Business News (June 29, 2025). *The University of Auckland Business School Research Paper Forthcoming*, Available at SSRN: https://ssrn.com/abstract=5207116 or http://dx.doi.org/10.2139/ssrn.5207116.

Cheng, C. H., & Wei, L. Y. (2014). A novel time-series model based on empirical mode decomposition for forecasting TAIEX. *Economic Modelling*, 36, 136-141. Doi: 10.1016/j.econmod.2013.09.033.

Díaz, J. D., Hansen, E., & Cabrera, G. (2024). Machine-learning stock market volatility: Predictability, drivers, and economic value. *International Review of Financial Analysis*, 94, 103286. https://doi.org/10.1016/j.irfa.2024.103286.

Filipović, D., & Khalilzadeh, A. (2021). Machine learning for predicting stock return volatility. *Swiss Finance Institute Research Paper*, (21-95). http://dx.doi.org/10.2139/ssrn.3995529.

Han, B., Liu, A., Chen, J., & Knottenbelt, W. (2025). Can machine learning models better volatility forecasting? A combined method. *The European Journal of Finance*, 1-22. https://doi.org/10.1080/1351847X.2025.2553053.

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: *Mathematical, Physical and Engineering Sciences*, 903-995. https://doi.org/10.1098/rspa.1998.0193.

Kang, W., de Gracia, F. P., & Ratti, R. A. (2025). Stock market volatility and commodity prices. *Macroeconomic Dynamics*, 29, e45. https://doi.org/10.1017/S1365100524000324.

Karanasos, M., Yfanti, S., & Hunter, J. (2022). Emerging stock market volatility and economic fundamentals: The importance of US uncertainty spillovers, financial and health crises. *Annals of operations research*, 313(2), 1077-1116. https://doi.org/10.1007/s10479-021-04042-y.

Khoojine, A. S., Feng, Z., Shadabfar, M., & Khoojine, N. S. (2023). Analyzing volatility patterns in the Chinese stock market using partial mutual information-based distances. *The European Physical Journal B*, 96(12), 165. https://doi.org/10.1140/epjb/s10051-023-00628-6.

Kumar, S., Rao, A., & Dhochak, M. (2025). Hybrid ML models for volatility prediction in financial risk management. *International Review of Economics & Finance*, 98, 103915. https://doi.org/10.1016/j.iref.2025.103915.

Lamine, D., & Brijlal, P. (2024). Forecasting stock market realized volatility using random forest and artificial neural network in South Africa. International *Journal of Economics and Financial Issues*, 14(2), 5. https://doi.org/10.32479/ijefi.15431.

Lee, W. Y., Jiang, C. X., & Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of banking & Finance*, 26(12), 2277-2299. https://doi.org/10.1016/S0378-4266(01)00202-3.

Letteri, I. (2023). Volts: A volatility-based trading system to forecast stock markets trend using statistics and machine learning. arXiv preprint arXiv:2307.13422. https://doi.org/10.48550/arXiv.2307.13422.

Lin, C. S., Chiu, S. H., & Lin, T. Y. (2012). Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*, 29(6), 2583-2590. Doi: 10.1016/j.econmod.2012.07.018.

Lubis, M. A., & Samsudin, S. (2025). Using the random forest method in predicting stock price movements. *Journal of Dinda*. Inf. Technol. 5(1), 28-35. https://doi.org/10.20895/dinda.v5i1.1765.

Mansilla-Lopez, J., Mauricio, D., & Narváez, A. (2025). Factors, Forecasts, and Simulations of Volatility in the Stock Market Using Machine Learning. *Journal of Risk and Financial Management*, 18(5), 227. https://doi.org/10.3390/jrfm18050227.

Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2), 478-539. DOI: 10.1257/002205103765762743.

Yan, K., & Li, Y. (2024). Machine learning-based analysis of volatility quantitative investment strategies for American financial stocks. *Quantitative Finance and Economics* , 8, 364-386. doi: 10.3934/QFE.2024014.

Zhang, C., Zhang, Y., Cucuringu, M., & Qian, Z. (2024). Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*, 22(2), 492-530. https://doi.org/10.48550/arXiv.2202.08962.