Neutrosophic Sets and Systems, Vol. 90, 2025
University of New Mexico

Evaluating the Role of Investor Sentiment Dynamics in Tadawul Stock Market Using Hybrid Wang-Mendel Fuzzy Logic: A Bridge to Neutrosophic Uncertainty

Abdullah Alawajee[1], Mohd Tahir Ismail[1*], S. Al Wadi[2] and Jamil J. Jaber[3,4*]

[1] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia. aaalanazi103@gmail.com; m.tahir@usm.my

[2]Department of Finance, Faculty of Business, The University of Jordan, Aqaba 77110, Jordan. s.alwadi@ju.edu.jo;

[3]Department of Mathematical Sciences, Faculty of Science and Technology, The National University of Malaysia-Qatar, Lusail, Qatar, j.jaber@ju.edu.jo, jameljaber2011@hotmail.com

[4]Department of Finance, Faculty of Business, The University of Jordan, Aqaba 77110, Jordan. j.jaber@ju.edu.jo, jameljaber2011@hotmail.com

***** Correspondence: j.jaber@ju.edu.jo, m.tahir@usm.my

Abstract: This study enhances the precision of predicting daily Saudi stock market closing prices and returns by integrating the Maximum Overlapping Discrete Wavelet Transform (MODWT) spectral model with the Wang-Mendel (WM) method, the latter known for generating interpretable fuzzy rules and its financial applications like the Wang transform for risk pricing. WM forms the foundation for Neutrosophic Logic, which expands it by adding indeterminacy (I) and falsity (F) to truth (T) for advanced uncertainty modeling. Utilizing five wavelet basis functions and daily Tadawul closing prices (August 2017 - September 2022), inputs for Model (1) predicting returns (SMR) were selected as Real Estate Investment Trusts return(REIT), Real Estate Management & Development returns (REMD), and Brent oil returns (ROIL), while Model (2) predicting closing prices (NSM) used Real Estate Management & Development closing prices (NREMD) and Brent oil closing prices (NOIL), following variable selection via multiple regression, multicollinearity (tolerance, VIF), fixed/random effects, unit root, and Granger causality tests, with min-max normalization applied. Results show MODWT-d4 with ARIMA(1,0,0) (non-zero mean) for Model (1) and MODWT-d4 with ARIMA(5,1,0) for Model (2) outperformed other wavelets on the 80% training data based on lower error; furthermore, the MODWT-C6-WM hybrid model surpassed traditional models in forecasting the remaining 20% for returns (Model 1), while the MODWT-La8-WM hybrid excelled for price forecasting (Model 2).

Keywords: Wavelet; Wang-Mendel; Neutrosophic ; Stock Market

1. Introduction

The Saudi Stock Exchange (Tadawul) has experienced remarkable expansion in recent years. In 2020, it ranked among the world's top 10 exchanges by market capitalization, securing ninth place globally among 67 financial markets and third place among emerging markets (WFE 2020). This prominent market features high liquidity and is regulated by the Capital Market Authority (CMA). Established in 1988, the Tadawul All Share Index (TASI) serves as its primary benchmark [1].

Financial analysts predict Tadawul will continue its growth trajectory, driven by increasing foreign investment. External investors are drawn to the market despite its inherent volatility and potential risks, largely due to the prospect of significant capital gains. While recent political and global economic uncertainty has heightened market fluctuations—raising investor concerns—this environment also presents opportunities for higher returns. Although Tadawul remains comparatively volatile, the potential rewards may outweigh the risks for investors willing to take calculated positions. Foreign investors maintain a positive outlook on the Saudi economy and Tadawul's future. This optimism stems from the Kingdom's diversified financial assets, expanding investment opportunities in sectors like telecommunications and banking, a stable regulatory framework, and the attractiveness of non-dollar-denominated assets amid a weaker US dollar. Overall, Saudi Arabia's influential economic position, combined with its regulatory stability, underpins Tadawul's appeal [2,3].

The stock markets of the Gulf Cooperation Council (GCC) countries are significantly influenced by key factors including geopolitical risk, crude oil price fluctuations, and the COVID-19 pandemic. Research has consistently explored these relationships. [4], examining stock return predictability from 2007 to 2019, identified geopolitical risk and crude oil returns as crucial determinants. Their findings indicate that heightened geopolitical risk reduces stock returns, while rising oil prices enhance returns. The distinct impact of the COVID-19 pandemic on GCC stock markets was specifically addressed by [5]. Furthermore, [6] analyzed the oil-stock market nexus from 2004 to 2019, employing Rademacher oil price decomposition and quantile regression. Their results confirm a significant short- and long-term relationship, demonstrating that oil prices exert a substantial influence on GCC stock returns. Expanding the analysis to sectoral impacts, [7] investigated the effects of oil price volatility and geopolitical risk on GCC stock sectors. Their study revealed significant impacts on both the energy and financial sectors, with oil volatility exhibiting a particularly pronounced effect during periods of high market returns.

Wavelet Transform (WT) is a mathematical technique widely employed in signal processing and time-frequency decomposition to break down non-stationary signals into constituent frequency components [8,9]. When applied to stock price data, WT decomposes it across multiple scales and frequencies, enabling a granular analysis of volatility patterns. This method effectively captures complex volatility structures and enhances forecasting of future market trends. By providing a more precise representation of market behavior, WT empowers investors and analysts to enhance investment decision-making. The utilization of WT in finance, particularly for stock market analysis, has garnered significant attention. A key application area is volatility analysis – measuring the

magnitude of price fluctuations over time – which is critical for investors and analysts to assess potential risks and returns [10].

The prediction of stock movements in Saudi Arabia's Tadawul exchange employs diverse models with varying inputs and sample sizes, where advanced techniques like machine learning and time series analysis show significant promise. Notable approaches include: a hybrid Long Short-Term Memory (LSTM) model integrated with Discrete Wavelet Transform (DWT) and ARIMA, achieving 97.54% accuracy using historical closing prices [11]; ARIMA applied to Tadawul All Share Index (TASI) data, though outperformed by LSTM[12]; a hybrid MODWT-Adaptive Network-based Fuzzy Inference System (ANFIS) incorporating oil prices and repo rates [13]; GARCH-family models (GARCH(1,1), EGARCH(1,1), GRJ-GARCH(1,1)) for volatility estimation [14]; and a Dynamic Evolving Neural Fuzzy Inference System (DENFIS) combined with MODWT using lagged prices and oil data [8]. Despite demonstrating effectiveness, inherent market volatility necessitates ongoing model refinement to enhance predictive accuracy.

The Wang-Mendel (WM) methodology, developed for generating interpretable and accurate fuzzy rule-based systems in complex data environments, has evolved significantly since its inception, finding notable application in finance, particularly through the widely used Wang transform for pricing financial instruments under risk [15]. In this study, we present a novel approach building on this foundation, predicting stock market closing prices and returns of the Tadawul All Share Index (TASI) using a hybrid methodology that combines maximal overlap discrete wavelet transform (MODWT) functions with the WM model. Our proposed approach leverages the WM methodology's ability to handle uncertainty and provide interpretable results for financial forecasting by first using MODWT to decompose stock market data into multiple scales and frequency bands, which are then fed as input to the WM model. While the WM methodology shows promise, its effectiveness can be limited by data complexity and variable selection [16] . Our approach is an innovative and powerful tool that can provide investors and financial analysts with reliable predictions of stock market volatility.

Neutrosophic uncertainty extends traditional fuzzy logic by explicitly incorporating indeterminacy (I) as a fundamental component alongside truth (T) and falsity (F), offering a more nuanced framework for modeling real-world uncertainty. Introduced by [17], neutrosophic logic generalizes several predecessor theories: (1) fuzzy sets [18], which represent partial membership but lack explicit indeterminacy; (2) intuitionistic fuzzy sets [19], which track membership and non-membership yet still conflate uncertainty with conflict; and (3) paraconsistent logic, which tolerates contradictions but does not quantify ignorance [20]. Unlike these models, neutrosophic logic operates on three independent dimensions $(T, I, F \in]^{-}0, 1^{+}[)$, enabling the distinct representation of ambiguity (e.g., incomplete data), conflict (e.g., contradictory evidence), and classical uncertainty (e.g., probabilistic likelihood) [17]. For instance, in financial markets, investor sentiment might exhibit high truth (T = 0.8, strong bullish indicators), moderate falsity (F = 0.3, some bearish signals), and notable indeterminacy (I = 0.4, reflecting unresolved geopolitical factors)—a scenario poorly captured by fuzzy or intuitionistic methods.

The paper is structured as follows: Section 2 offers a review of the related literature on predicting stock market volatility. Section 3 provides a detailed explanation of the methodology and mathematical approach. Section 4 describes the data used in the study, while Section 5 presents experimental results and compares them with existing models. Section 6 outlines the limitations of this study and suggests future research directions. The paper is concluded in Section 7.

2. Literature Review

Predicting the stock market is crucial for investors and financial analysts to optimize investment decisions and risk management strategies. The development of various models and techniques has significantly enhanced forecasting capabilities. One foundational approach is the Autoregressive Conditional Heteroskedasticity (ARCH) model introduced by [21], which estimates the conditional variance of returns using past squared residuals [22]. ARCH models effectively capture volatility clustering—the phenomenon where periods of high volatility tend to cluster together. [22] extended this framework with the Generalized ARCH (GARCH) model, incorporating both lagged conditional variances and lagged squared residuals to improve accuracy, particularly in modeling long-term volatility persistence. Demonstrating the continued evolution of such methods, [23] employed a Time-Varying Parameter Vector Autoregression (TVP-VAR) model, based on Diebold and Yilmaz's technique, to analyze daily data (2013–2020) and investigate volatility spillovers and dynamic linkages between crude oil, gold, and Chinese stock markets.

Artificial Neural Networks (ANNs) are a popular technique widely applied to stock price prediction. [24] demonstrated their potential early by developing a hybrid ANN-fuzzy logic system that achieved a notable 86.17% prediction accuracy; however, the timeframe referenced ("last 20 years") relative to their publication date requires clarification. More recently, the Random Forest (RF) model has gained significant attention. [25] utilized RF to investigate the predictive power of industry returns—constructed according to the Global Industry Classification Standard using data from 1972 to 2016—on overall stock market movements. This study found the RF model delivered superior accuracy and out-of-sample prediction performance compared to linear regression and neural networks. Advancing volatility forecasting specifically, [10] introduced a novel hybrid approach integrating the Exponential GARCH (EGARCH) model with a selected Maximal Overlap Discrete Wavelet Transform (MODWT) function, aiming to enhance stock market volatility prediction accuracy.

To enhance stock market volatility forecasting accuracy, [26] introduced the Trend Deterministic Data Preparation (TDDP) method, designed to remove trend components from data before input into machine learning (ML) models. Applying TDDP alongside algorithms including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and decision trees to 10 years of daily Bombay Stock Exchange (BSE) data (2001-2010), they found TDDP significantly improved model accuracy and prediction consistency, with ANNs demonstrating superior performance among individual algorithms. Building on this, [27] employed an ensemble approach combining ANNs, decision trees, SVMs, and K-Nearest Neighbors (KNN) on the same BSE dataset. This ensemble

outperformed all individual models, achieving the highest prediction accuracy; notably, ANNs and SVMs were the strongest base models, and their combination with others yielded further gains. Both Patel studies acknowledged that external factors (e.g., political events, market trend shifts) not incorporated into their models could impact accuracy. More recently, [28,29] proposed a hybrid model integrating the Fuzzy Inference Rules by Descent Method (FIR.DM), the heuristic gradient descent (FS.HGD) and hybrid neural fuzzy inference system (HyFIS) with MODWT functions for predicting Tadawul All Share Index (TASI) volatility, demonstrating its superiority over standard models like ARIMA and traditional models.

[8] developed a hybrid MODWT-Haar-DENFIS model to forecast volatility (LSCP) in the Saudi stock market (Tadawul), using 4,609 daily observations (2006-2024) capturing multiple crises. Their model decomposed volatility via MODWT with the Haar wavelet and fed the low-frequency components alongside key inputs—lagged closing price (LCP), logarithmic oil price (Loil), cost of living (LCL), and interbank rate (IB)—into a DENFIS system. This hybrid approach significantly outperformed benchmarks (ARIMA, standalone DENFIS, other MODWT-DENFIS models), achieving the lowest test errors (RMSE: 2.40, MAPE: 41.99%). Analysis identified LCP as having a negative impact and Loil a positive impact on volatility, while LCL and IB showed mixed effects; the study concluded wavelet-integrated neuro-fuzzy systems greatly enhance forecasting in emerging markets.

3. Methods and Mathematical Models

This section outlines the methodology of the hybrid forecasting model integrating Wang and Mendel's (WM) Fuzzy Logic System with Wavelet decomposition. The process begins with dataset normalization using the Min-Max model to standardize input variables, including REIT (Real Estate Investment Trusts), REMD (Real Estate Management and Development), and OIL prices. The TASI (Tadawul All Share Index) daily closing prices serve as the output variable.

To address localized variations such as transient events and anomalies, the dataset undergoes decomposition via the Maximal Overlap Discrete Wavelet Transform (MODWT). This splits the data into two components: detail coefficients (high-frequency features), which capture short-term fluctuations and main dataset characteristics, and approximate coefficients (low-frequency features), representing smoothed trends and serving as the output variable. In the subsequent stage, external economic variables (REIT, REMD, OIL) are incorporated as inputs into the WM Fuzzy Logic System, which handles uncertainties and nonlinearities in the data. The hybrid framework combines wavelet-processed features with fuzzy logic to enhance forecasting robustness. Finally, accuracy criteria tests are applied to evaluate model performance and select the optimal hybrid model. Figure 1 visually encapsulates this workflow, emphasizing the synergy between wavelet-based feature extraction, fuzzy logic processing, and empirical validation to achieve reliable financial forecasting.

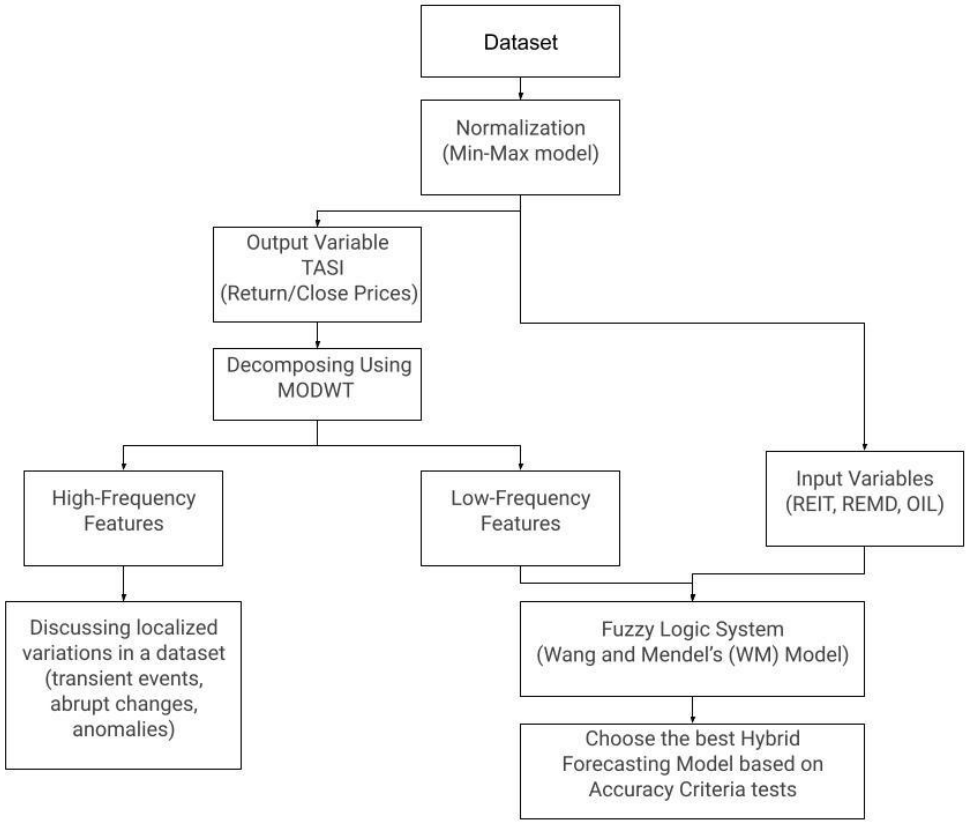


Figure 1. The flowchart of hybrid WM with Wavlet model

3.1. Wavelet transform

Time series analysis is a powerful tool for understanding trends, discovering hidden patterns, and predicting future events. In 1807, Joseph Fourier demonstrated that any 2π function could be expressed as a sum of sinusoidal components with appropriate coefficients. The Fourier transform is a well-known signal analysis method that decomposes a signal into its component sine and cosine functions as shown in (1),

$$f(t) = \frac{a_0}{2} + \sum_{n\in N} a_n \cos nt + \sum_{n\in N} b_n \sin nt, \tag{1}$$

where $a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\, dt$ and $b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\, dt$. The a_n and b_n are the sine and cosine Fourier coefficients, respectively with $n \in Z, f \in L_1^r[-\pi, \pi], (t \text{ or } r) \in R$. These sine and cosine coefficients were modified to reconstruct the signals known as deterministic [30].

In recent decades, WT has emerged as a powerful alternative to traditional time series analysis techniques such as Fourier analysis. WT is a mathematical tool that can effectively detect features in a signal while minimising the effects of noise by adjusting to fit data in both the time and frequency domains [8]. There are three types of WT: MODWT, discrete WT (DWT), and continuous WT (CWT), which share similar characteristics. The primary distinction between DWT and MODWT is the former can only be used with a limited number of observations, while the latter may be used with any quantity of data. WT is an extension of the Fourier transform[8,10]. It is divided into two types: the mother wavelet $\phi(t)$, which defines the high-frequency components or detailed coefficients, and the

father wavelet $\varphi(t)$, which represents the low-frequency components or smooth coefficients as indicated in (2) and (3), respectively [9] .

$$S_{j,k} = \int \phi_{j,k} f(t)dt, \tag{2}$$

$$d_{j,k} = \int \varphi_{j,k} f(t)dt, \tag{3}$$

where $S_{j,k}$ and $d_{j,k}$ denote smooth and detailed coefficients, respectively, J denotes the maximum scale sustainable by the number of data points and the two types of wavelets

$$\phi_{j,k} = 2^{\left(\frac{-j}{2}\right)} \phi\left(t - \frac{2^j k}{2^j}\right), \tag{4}$$

$$\varphi_{j,k} = 2^{\left(\frac{-j}{2}\right)} \varphi\left(t - \frac{2^j k}{2^j}\right). \tag{5}$$

Smooth coefficients capture the most important features of the original data, while detailed coefficients reveal the main variations in the data. The father wavelets and mother wavelets need to meet the requirements of $\int \phi(t)dt = 1$ and $\int \varphi(t)\, dt = 0$. The MODWT provides several useful transform functions, including Haar, Daubechies (d4), Coiflet (C6), least asymmetric (LA8), and best-localized (bl14)[29].

32. *Fuzzy Inference (Mamdani Model)*

The Mamdani Model (MM) and Neutrosophic Logic both address uncertainty but operate differently: MM uses fuzzy sets with single membership values ($\mu \in [0,1]$) for rule-based inference, ideal for control systems, while Neutrosophic Logic generalizes this approach with a triplet ⟨T, I, F⟩ (where T, I, F $\in]^{-}0, 1^{+}[$) to explicitly quantify truth, indeterminacy, and falsity independently, making it better suited for high-ambiguity domains like finance or healthcare. While MM relies on defuzzification for crisp outputs, Neutrosophic Logic preserves granular uncertainty through all stages of reasoning.

The Mamdani model (MM) is a widely used fuzzy logic-based inference system developed by Lotfi Zadeh in the 1970s. Due to its simplicity, the MM has found numerous applications in fields such as artificial intelligence, engineering, and medical diagnosis. [31] introduced the principles of fuzzy set theory and proposed a fuzzy inference system that employed linguistic fuzzy variables for both input and output. The fuzzy set theory provides tools for effectively representing linguistic concepts, variables, and rules, making it a natural model for representing human expert knowledge. In the same paper, a four-step procedure was proposed for computing fuzzy inference systems, which involved fuzzification, inference, aggregation, and defuzzification.

The MM of fuzzy logic was first introduced in studies by [32,33]. It is a linguistic variable-based inference system in which both the antecedent and consequent parts of the rules consist of linguistic variables, hence applicable to multi-input and single-output systems. For instance, we can define a fuzzy IF-THEN rule as follows for the input linguistic variables $Z_n = \{Z_1, \ldots, Z_n\}$ and $W_n = \{W_1, \ldots, W_n\}$, with output lingusitic variable Y_n = $\{Y_1, \ldots, Y_n\}$ IF Z_1 is W_1 and … and Z_n is W_n THEN Y is $\{Y_1, \ldots, Y_n\}$.

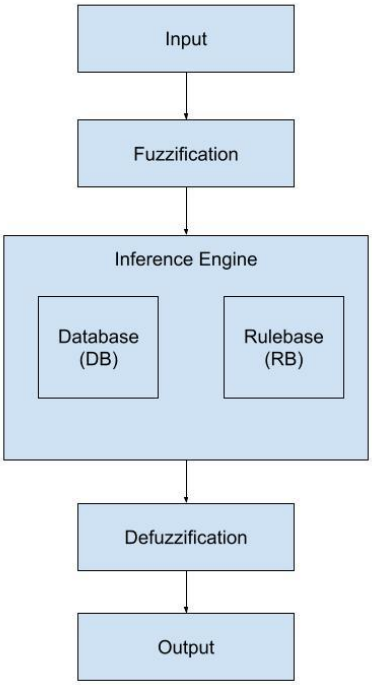


Figure 2. FRBS and Mamdani model.

The flow chart illustrates the integrated learning and prediction phases of a Fuzzy Rule-Based System (FRBS) using the Mamdani model (MM), where the learning phase (implicitly embedded in the Database [DB] and Rulebase [RB]) first processes training data through structure identification to define input/output variables, linguistic terms (e.g., "Low"/"Medium"/"High"), and initial rule structures, followed by parameter estimation to optimize membership function shapes and rule weights—executed simultaneously or sequentially—to build the knowledge base (DB storing fuzzy set definitions and membership parameters, RB housing fuzzy IF-THEN rules); this foundation enables the prediction phase, depicted in Figure 2's standard MM architecture, where new crisp Input undergoes Fuzzification (using DB to convert data to linguistic values), leverages the Knowledge Base (DB/RB) within the Inference Engine (applying RB's rules via Mamdani's MIN-MAX operations for reasoning and fuzzy output generation), and culminates in Defuzzification to transform aggregated fuzzy outputs into actionable crisp Predicted Values [34-36].

Learning approaches for FRBSs that utilize a space partitioning approach employ the segmentation of the variable domain. They subsequently leverage this partitioning to derive membership function parameters and utilize these spatial division methodologies to construct FRBSs. The Wang-Mendel methodology ("WM"), introduced by Wang and Mendel (1992) using the Mamdani framework, involves four learning stages: Step 1: Uniformly segment the input and output domains of provided numerical data into fuzzy regions constituting the database. Here, fuzzy regions denote intervals corresponding to linguistic descriptors. The span of these regions depends on the quantity of linguistic terms. Consider a temperature scale from 1 to 5. Defining linguistic terms "cold", "neutral", and "hot" with region breadth 2 yields intervals [1,3], [2,4], [3,5]. Triangular membership functions are constructed using vertex points: for [1,3], points a=1, b=2 (peak where membership = 1),

and c=3. Step 2: Formulate fuzzy IF-THEN rules encompassing training instances using Step 1's database. Compute membership degrees for all training values. For each data point and variable, assign the linguistic term achieving highest membership. Repeat across all training cases to establish comprehensive rules. Step 3: Assign a weight to each rule by multiplicatively combining antecedent and consequent membership degrees (product operator). Step 4: Establish the definitive rulebase by eliminating duplicate rules, retaining higher-weighted versions when conflicts occur. The resulting system follows the Mamdani inference framework.

3.3 Wang and Mendel's technique (WM)

Wang and Mendel's technique (WM), introduced in their seminal 1992 paper, is a foundational data-driven method for generating fuzzy rule-based systems (FRBS) directly from numerical data. Pioneering the automated extraction of fuzzy rules, WM addressed the critical bottleneck of manual knowledge acquisition, particularly for Mamdani-type FRBS where both antecedents and consequents are fuzzy sets, making it especially useful for regression tasks (function approximation). Its simplicity, noise tolerance, and ability to handle high-dimensional problems made it revolutionary. This spurred widespread applications in control systems (e.g., power grid stabilization), time-series prediction (e.g., financial forecasting), and pattern recognition. WM's significant legacy includes derivatives like WM+ (extended for regression) and its influence on hybrid neuro-fuzzy systems, cementing its enduring role in modern computational intelligence [37-40]

Below is a step-by-step explanation of the WM method, including system space partitioning, equations, and its application to regression.

Step 1: System Space Partitioning, involves dividing the input-output space into overlapping fuzzy regions defined by linguistic terms (e.g., "Low", "Medium", "High"). This process begins by normalizing the domain intervals of all input variables (x_1, x_2, …,xn) and the output variable (y) to a predefined range, typically [0, 1], to ensure consistent scaling. Subsequently, each normalized input and output variable is partitioned into these fuzzy regions by assigning appropriate membership functions (MFs) – such as triangular, trapezoidal, or Gaussian shapes – which mathematically define the degree of belonging for any value within each linguistic set and inherently create overlaps between adjacent regions.

Step 2: Generate Fuzzy Rules from Data Pairs

In the first stage, for each input-output data pair $(x_1^{(k)}, x_2^{(k)}, \ldots x_n^{(k)}, y^{(k)})$, compute the membership degree of each variable to its ifuzzy sets. For instance, find the fuzzy set A_{ij} with the highest membership for $x_i^{(k)}$ as following:

$$\mu_{A_{ij}}(x_i^{(k)}) = \max_j \mu_{A_{ij}}(x_i^{(k)}) \quad (6)$$

Repeat for the output $y^{(k)}$, identifing its best matching fuzzy set B_j.

In the second stage, construct a fuzzy rule, involve map the input fuzzy sets to the output fuzzy set

$$\text{IF } x_1 \text{ is } A_{1j} \text{ AND } x_2 \text{ is } A_{2j} \text{ AND}.\ldots x_n \text{ is } A_{nj}, \text{THEN y is } B_j.$$

where A_{1j}, A_{2j},…, A_{nj}, B_j are the regions with max membership for $(x_1^{(k)}, x_2^{(k)}, \ldots x_n^{(k)}, y^{(k)})$.

Step 3: Assign a Degree to Each Rule. In this stage, the rule strength is assigned to each fuzzy rule by calculating the product of the membership degrees of the input values in their respective antecedent fuzzy sets, representing the combined degree to which the rule's conditions are satisfied by the current inputs.

$$\text{Strength}^{(k)} = \prod_{i=1}^{n} \mu_{A_{ij}}(x_i^{(k)}) \cdot \mu_{B_j}(y^{(k)}) \tag{7}$$

In this stage, repeat for the output $y^{(k)}$, identifying its best matching fuzzy set B_j. If multiple rules have the same antecedents but different consequents, retain the rule with the highest strength.

Step 4: Fuzzy Inference (Mamdani Model)

For a new input $x^* = (x_1^*, x_2^*, \ldots x_n^*)$, compute membership degrees to each fuzzy set.

In the first stage, compute the firing strength α_j for each rule as following:

$$\alpha_j = \min(\mu_{A_{ij}}(x_1^*), \mu_{A_{ij}}(x_2^*), \ldots, \mu_{A_{ij}}(x_n^*))$$

Or

$$\alpha_j = \prod_{i=1}^{n} \mu_{A_{ij}}(x_i^*) \tag{8}$$

In the second satge, apply the firing strength to the consequent fuzzy set B_j :

$$\mu_{B'_j}(y) = \min(\alpha_j, \mu_{B_j}(y)) \quad \text{(clipping)}$$
$$\mu_{B'_j}(y) = \alpha_j, \mu_{B_j}(y) \quad \text{(scaling)} \tag{9}$$

Then combine all clipped/scaled consequents using the max operator:

$$\mu_{agg}(y) = \max_j \mu_{B'_j}(y) \tag{10}$$

Step 5: Defuzzification: Convert the aggregated fuzzy output to a crisp value y^*:

- Centroid Method:

$$y^* = \frac{\int y.\,\mu_{agg}(y)\,dy}{\int \mu_{agg}(y)\,dy} \tag{11}$$

-Weighted Average (simplified): If singletons are used for consequents:

$$y^* = \frac{\sum_{j=1}^{M} \alpha_j.\, c_j}{\sum_{j=1}^{M} \alpha_j} \tag{12}$$

where c_j is the centroid of B_j.

For regression, the FRBS approximates a function f:Rn→R. The WM method constructs rules that map inputs to outputs using fuzzy logic, effectively capturing non-linear relationships. The Wang-Mendel method operates in two sequential phases: during training, the WM algorithm generates fuzzy rules directly from numerical training data and stores them in the rule base; subsequently, in the testing phase, the system computes the output y^* for any test input x^* using Mamdani inference, which aggregates fired rules through fuzzy operations (fuzzification, rule evaluation, aggregation, and defuzzification) to produce crisp outputs.

3.4 Error criteria test

We evaluate the accuracy of our method using mean absolute percentage error (MAPE), mean error (ME), MAE, and root mean squared error (RMSE). The MAPE, also known as the mean absolute percentage deviation, is a statistical measure of prediction accuracy expressed as a percentage. It is defined as

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right|, \tag{13}$$

where X_t is the actual value, F_t is the forecasted value and n is the sample size. The absolute value in (13) is summed for every forecasted point in time and divided by the number of fitted points. MAE and ME are respectively defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|X_t - F_t|, \tag{14}$$

$$\text{ME} = \frac{1}{n}\sum_{t=1}^{n}(X_t - F_t), \tag{15}$$

L

E

t (t = 1,..., n) denote the number of input variables. The mean square error (MSE) is defined as

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - F_t)^2 \tag{16}$$

The RMSE, also known as the root mean squared deviation, is a frequently used measure of the estimators' differences. It measures the average error produced by the model in predicting the outcome of an observation. It is defined as

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{17}$$

4. Data Description

This study utilizes a dataset of 1,410 daily closing prices for the Tadawul All Share Index (TASI), sourced from the Saudi Exchange (Tadawul) and spanning 1 August 2017 to 1 September 2022, to analyze Saudi equity market dynamics across distinct economic phases—capturing pre-pandemic stability under oil-driven conditions (Aug 2017–Feb 2020), the acute COVID-19 crisis with its historic 23% single-day crash (Mar 2020), extreme volatility, and the post-pandemic recovery (2021–Sep 2022) driven by oil rebounds and Vision 2030 diversification—enabling assessment of black-swan resilience, sectoral vulnerabilities, and data-driven investment strategies in a transformative era.

This study's statistical summary (Table 1) encompasses 1,410 daily observations of Tadawul market indices and oil prices: the TASI ranged from a COVID-19 pandemic low of 5,959.69 to a recovery peak of 13,820.35, averaging 8,808.20 amid high volatility (Std. Dev. 1,837.41); the REIT (real estate trusts) varied between 3,501.77 and 6,824.32 (mean 4,379.84, Std. Dev. 686.68); the REMD (real estate development) fluctuated from 2,294.42 to 5,039.92 (mean 3,522.42, Std. Dev. 656.43); while Brent crude oil exhibited extreme volatility (Std. Dev. 18.85), plunging to $19.33 during the 2020 demand collapse and surging to $127.98 in the 2022 supply crisis, averaging $65.93 and underscoring its pivotal role in the Saudi market's dynamics.

Table 1. A Statistical summary of TASI, REIT, REMD and Oil closing prices in Tadawul .

	N	Mean	Std. Dev.	Minimum	Maximum
TASI	1410	8808.197	1837.413	5959.690	13820.350
REIT	1410	4379.837	686.676	3501.770	6824.320
REMD	1410	3522.420	656.427	2294.420	5039.920
OIL	1410	65.927	18.851	19.330	127.980

5. Empirical Results and Discussion

A hybrid model integrating Wang and Mendel's technique with MODWT was proposed to forecast Tadawul stock market closing prices from 2017 to 2022, utilizing wavelets like Haar, d4, LA8, C6, and bl14 within the MODWT framework. This method involves three core steps: decomposition,

which breaks down the price signal into wavelet coefficients across different scales and positions; filtering, which extracts relevant information from these coefficients using filter banks; and reconstruction, which synthesizes the filtered coefficients back into an approximation of the original signal. MODWT categorizes the data into detail series (capturing high-frequency fluctuations) and approximation series (capturing the trend), specifically designed to handle the large fluctuations typical in financial data, and the model's performance was assessed using an accuracy indicator.

The proposed hybrid MODWT-WM models represent a novel method for predicting Saudi stock market returns (SMR) and closing prices (NSM), utilizing Tadawul stock data and oil prices from Investing.com. The methodology involves four key steps: firstly, collecting the financial time series data; secondly, normalizing both input and output variables (SMR, NSM) using Min-Max scaling; thirdly, decomposing the SMR and NSM data using five distinct MODWT functions (Haar, d4, LA8, C6, bl14) to separate them into low-fluctuating approximation coefficients and high-fluctuating detail coefficients, with the approximation coefficients serving as the primary outputs for forecasting; and finally, applying a robust modeling approach (WM) to establish the relationship between the input financial variables and these SMR/NSM approximation coefficients for each MODWT function, resulting in the hybrid MODWT-WM models designed for accurate and reliable forecasting.

A comparative study then evaluates the best-performing MODWT-WM hybrid models against alternative MODWT functions and the traditional WM model, using an 80/20 data split where 80% of the original Saudi stock market and oil price data trains the models and selects the most effective one, while the remaining 20% tests its performance. This proposed hybrid model offers an innovative approach by combining MODWT decomposition (using approximation coefficients of SMR and NSM), relevant input financial variables, and hybrid modeling, resulting in significantly more accurate and reliable forecasts than traditional methods. The study conclusively highlights the superiority of the MODWT-WM models over both the alternative functions and the standalone WM model, demonstrating their strong potential for broader financial forecasting applications.

5.1. Selecting variables

5.1.1 Multicollinearity test

Table 2 shows The correlation between input and output variables for model 1 and model 2. In model (1), the SMR exhibits a strong positive correlation (0.690 ≥ 50%) with REMD return, indicating REMD movements significantly influence Tadawul's overall market performance; it shows a weak positive correlation (0.335 < 50%) with REIT returns, suggesting a limited relationship; and demonstrates an extremely weak positive correlation (0.086 < 50%) with Brent Oil price returns (ROIL), implying negligible linear influence from oil prices. Critically, all correlations between independent variables are weak (below the 50% threshold): REIT and REMD show a weak positive correlation (0.318 < 50%), while both REIT-ROIL (0.020 < 50%) and REMD-ROIL (0.026 < 50%) exhibit near-zero correlations, confirming no significant multicollinearity exists between these predictors as all inter-variable correlations fall substantially below the 50% benchmark.

In model (2), NSM exhibits a strong positive correlation (0.690 ≥ 50%) with the Real Estate Management and Development Index (REMD) closing prices, indicating that Tadawul's overall market performance moves significantly in tandem with the real estate sector; it shows a weak positive correlation (0.335 < 50%) with Real Estate Investment Trust (REIT) closing prices, suggesting a limited relationship; and demonstrates an extremely weak positive correlation (0.086 < 50%) with normalized Brent Oil prices (ROIL), revealing negligible linear influence from oil markets. Crucially, correlations between independent variables are all below the 50% threshold: REMD and ROIL show a near-zero correlation (0.026 < 50%), while REIT and ROIL (0.020 < 50%) and REIT and REMD (0.318 < 50%) similarly display weak linkages, confirming no significant multicollinearity exists between predictors as all inter-variable correlations fall substantially below the 50% benchmark for strength.

Collinearity diagnostics confirm the absence of significant multicollinearity in both models, with all tolerance values substantially exceeding the critical threshold of 0.1 and all VIF values falling well below the conservative limit of 10. In Model (1), the independent variables (REIT, REMD, ROIL) show robust metrics (REIT: Tolerance=0.899/VIF=1.113; REMD: Tolerance=0.899/VIF=1.113; ROIL: Tolerance=0.999/VIF=1.001), while Model (2) exhibits near-ideal results for its normalized predictors (NREMD and NOIL: Tolerance=0.999/VIF=1.001 for both). These metrics—where tolerance measures unique variable variance and VIF quantifies coefficient variance inflation—collectively indicate minimal shared variance between predictors across both models. Crucially, Model (1) demonstrates no concerning inter-variable correlations among its three inputs, while Model (2) confirms virtually independent predictors. Since collinearity compromises regression stability, these results validate both models' reliability for both Tadawul return/closing price (SMR/NSM) analysis, ensuring unbiased coefficient estimates and robust predictive power without collinearity-induced instability.

Table 2. The correlation between input and output variables for model 1 and model 2

Models		SMR	REIT	REMD	ROIL	Collinearity statistics	
	SMR	1.000	0.335	0.690	0.086	Tolerance	VIF
Model (1)	REIT		1.000	0.318	0.020	0.899	1.113
	REMD			1.000	0.026	0.899	1.113
	ROIL				1.000	0.999	1.001
		NSM	NREMD	NOIL		Tolerance	VIF
Model (2)	NSM	1.000	-0.283	0.790			
	NREMD		1.000	-0.023		0.999	1.001
	NOIL			1.000		0.999	1.001

5.1.2 Multiregression models

In model (1) for table 3, The OLS regression model for SMR demonstrates strong overall significance ($F(3,1406) = 459.9$, $p < 0.01$) and explains approximately 49.5% of SMR variance (R^2 =

0.4953, Adj. $R^2 = 0.4942$). The intercept (constant) of 0.11925 ($t = 7.083$, $p < 0.01$) indicates the baseline SMR when all independent variables are zero. Among predictors, REMD show the strongest positive influence ($B = 0.58861$, $SE = 0.01817$, $t = 32.393$, $p < 0.01$), where a 1% increase in REMD corresponds to a 0.59% rise in SMR. The REIT also significantly predict SMR ($B = 0.15377$, $SE = 0.02401$, $t = 6.404$, $p < 0.01$), with a 1% REIT increase associated with a 0.15% SMR gain. The ROIL exhibit a smaller but still significant positive effect ($B = 0.07476$, $SE = 0.02128$, $t = 3.512$, $p < 0.01$), where a 1% oil price return increase relates to a 0.07% SMR increase. All coefficients are statistically significant at the 1% level, confirming each variable's distinct contribution to modeling Tadawul market returns.

The fixed effect regression model demonstrates excellent explanatory power for SMR, showing highly significant overall fit ($F(3,1401) = 460.253$, $p < 0.01$) and explaining approximately 49.6% of SMR variance ($R^2 = 0.49636$, Adj. $R^2 = 0.49349$). All three independent variables exhibit statistically significant positive relationships with SMR at the 1% level: REIT show a coefficient of 0.15501 ($SE = 0.02403$, $t = 6.45$), indicating a 1% increase in REIT returns corresponds to a 0.155% rise in SMR; REMD demonstrate the strongest effect with a coefficient of 0.58935 ($SE = 0.01819$, $t = 32.41$), where a 1% REMD increase predicts a 0.589% SMR gain; while ROIL have the smallest but still significant influence ($B = 0.07461$, $SE = 0.0213$, $t = 3.50$), suggesting a 1% oil price return increase associates with a 0.075% SMR increase. These results confirm REMD as the dominant driver of Tadawul market returns, with REIT and ROIL providing additional significant explanatory power in the fixed effects specification.

The random effects model demonstrates strong explanatory power for SMR, with a highly significant overall fit ($\chi^2(3) = 1379.77$, $p < 0.01$) and explaining approximately 49.5% of SMR variance ($R^2 = 0.49529$, Adj. $R^2 = 0.49421$). The intercept (constant) of 0.11925 ($SE = 0.01684$, $t = 7.083$, $p < 0.01$) indicates the baseline SMR when all predictors are zero. All independent variables show statistically significant positive effects at the 1% level: REIT exhibit a coefficient of 0.15377 ($SE = 0.02401$, $t = 6.404$), meaning a 1% increase in REIT returns corresponds to a 0.154% rise in SMR; REMD display the strongest influence with a coefficient of 0.58861 ($SE = 0.01817$, $t = 32.393$), where a 1% REMD increase predicts a 0.589% SMR gain; while ROIL have a modest but significant effect ($B = 0.07476$, $SE = 0.02129$, $t = 3.512$), indicating a 1% oil price return increase associates with a 0.075% SMR increase. These results confirm REMD as the dominant driver of market returns in the random effects specification.

In model (2) for table 4, the OLS model demonstrates exceptionally strong explanatory power for NSM, with a highly significant overall fit ($F(2,1407) = 1595$, $p < 0.01$) and explaining 69.4% of NSM variance ($R^2 = 0.6939$, Adj. $R^2 = 0.6935$). The intercept (constant) is 0.02600 ($SE = 0.01100$, $t = 2.278$, $p < 0.05$), indicating a small but significant baseline closing price when predictors are zero. Critically, NOIL exhibit a powerful positive relationship ($B = 1.05600$, $SE = 0.02000$, $t = 53.112$, $p < 0.01$), where a 1-unit increase in normalized oil prices corresponds to a 1.056-unit surge in NSM. Conversely, the NREMD shows a significant negative effect ($B = -0.26000$, $SE = 0.01400$, $t = -18.010$, $p < 0.01$), implying a 1-unit increase in normalized real estate prices associates with a 0.260-unit decline in Tadawul's market closing prices. These opposing forces highlight oil prices as the dominant driver of NSM, while real estate sector performance exerts a substantial but inverse influence on overall market valuations.

Table 3. The OLS, fixed effect , and random effect of model (1)

Tests	Variables	B	Std. error	t-test	F/Chisq-test	Adj. R-square	R-square
OLS	(Constant)	0.11925	0.01684	7.08300***	F:459.9*** on 3 and 1406 DF	0.4942	0.4953
	REIT	0.15377	0.02401	6.40400***			
	REMD	0.58861	0.01817	32.39300***			
	ROIL	0.07476	0.02128	3.51200***			
Fixed effect	REIT	0.15501	0.02403	6.44980 ***	F:460.253*** on 3 and 1401 DF	0.49349	0.49636
	REMD	0.58935	0.01819	32.40810 ***			
	ROIL	0.07461	0.02130	3.50350 ***			
Random effect	(Constant)	0.11925	0.01684	7.08300***	Chisq: 1379.77*** on 3 DF	0.49421	0.49529
	REIT	0.15377	0.02401	6.40420***			
	REMD	0.58861	0.01817	32.39250***			
	ROIL	0.07476	0.02129	3.51230***			

*Signif. codes: '***' 0.01 '**' 0.05 '*' 0.1*

The fixed effects model demonstrates strong explanatory power for NSM, with a highly significant overall fit ($F(2,1402) = 721.392$, $p < 0.01$) and explaining approximately 50.7% of NSM variance ($R^2 = 0.50717$, Adj. $R^2 = 0.50471$). Both normalized independent variables show statistically significant positive relationships at the 1% level: a one-unit increase in the NREMD corresponds to a 0.32381-unit rise in NSM (SE = 0.015, t = 21.583, $p < 0.01$), while a one-unit increase in NOIL predicts a stronger 0.56036-unit increase in NSM (SE = 0.02108, t = 26.581, $p < 0.01$). These results highlight oil prices as the dominant positive driver of market closing prices, with real estate sector performance also contributing significantly but with approximately half the magnitude of oil's influence. The absence of a reported constant term suggests the fixed effects specification effectively accounts for entity-specific intercepts in modeling Tadawul's closing price dynamics.

The random effects regression demonstrates strong explanatory power for NSM, with a highly significant overall fit ($\chi^2(2) = 1429$, $p < 0.01$) and explaining approximately 50.5% of NSM variance ($R^2 = 0.50508$, Adj. $R^2 = 0.50438$). While the constant term is statistically insignificant (B = -0.00752, SE = 0.04184, t = -0.18), both normalized predictors show highly significant positive relationships at the 1% level: a one-unit increase in NREMD corresponds to a 0.31837-unit rise in NSM (SE = 0.01507, t = 21.13), and a one-unit increase in NOIL predicts a stronger 0.56713-unit increase in NSM (SE = 0.02115, t = 26.81). These results confirm NOIL as the dominant driver of market closing prices with nearly 80% greater impact than NREMD (0.567 vs 0.318 units), reinforcing oil prices' critical role in Tadawul's valuation dynamics within the random effects framework. Random effect and fixed effect models use years as panal data.

Table 4. The OLS, fixed effect , and random effect of model (2)

Tests	Variables	B	Std. error	t-test	F/Chisq-test	Adj. R-square	R-square
OLS	(Constant)	0.02600	0.01100	2.27800**	1595*** on 2 and 1407 DF	0.6935	0.6939
	NREMD	-0.26000	0.01400	-18.01000 ***			
	NOIL	1.05600	0.02000	53.11200 ***			
Fixed effect	NREMD	0.32381	0.01500	21.58300***	721.392 on 2 and 1402 DF	0.50471	0.50717
	NOIL	0.56036	0.02108	26.58100 ***			
Random effect	(Constant)	-0.00752	0.04184	-0.17970	Chisq: 1429 on 2 DF	0.50438	0.50508
	NREMD	0.31837	0.01507	21.13090***			
	NOIL	0.56713	0.02115	26.81230***			

Signif. codes: '***' 0.01 '**' 0.05 '*' 0.1

5.1.3 Stationary and Causality Tests

For the Dickey-Fuller (ADF) test, the null hypothesis (H0) states that the time series has a unit root (non-stationary); at a 5% significance level, if the p-value ≤ 0.05, reject H0 (concluding the series is stationary), otherwise fail to reject H0 (indicating non-stationarity). For the Granger Causality test, the null hypothesis (H0) asserts that X does not Granger-cause Y; at 5% significance, reject H0 (concluding X Granger-causes Y) if the p-value ≤ 0.05, otherwise fail to reject H0 (finding no Granger causality from X to Y). Both tests use p-value ≤ 0.05 as the criterion for rejecting the null hypothesis[41].

In table 5, for the Dickey-Fuller (ADF) tests, all series (REIT, REMD, ROIL) overwhelmingly reject the null hypothesis of a unit root (non-stationarity) at the 1% level, as evidenced by their highly significant p-values (< 2e-16) and large negative t-statistics (-25.870, -24.654, -26.486); thus, all are stationary in model (1). For the Granger Causality tests, the null hypothesis that each variable (REIT, REMD, ROIL) does not cause SMR is rejected: REIT→SMR (p=0.0035<5%), REMD→SMR (p=0.0359<5%), and ROIL→SMR (p=7.954e-09<1%) all show statistically significant p-values below 0.05, confirming REIT, REMD, and ROIL each Granger-cause SMR in model (1).

In model (2), for the ADF tests, NREMD fails to reject the null hypothesis of non-stationarity (t-stat = -0.385, p-value = 0.701 > 5%), indicating NREMD is non-stationary, while NOIL rejects the null hypothesis (t-stat = -2.253, p-value = 0.0244 ≤ 5%), confirming NOIL is stationary. For Granger causality, NREMD does not Granger-cause NSM (p-value = 0.3316 > 5%), but NOIL significantly Granger-causes NSM (p-value = 0.0002 ≤ 1%), demonstrating predictive causality from NOIL to NSM.

Table 5. Dickey Fueler Test (ADF)-Unit Root test and Granger Causality Tests

Models	Null Hupothesis	Dickey Fueler Test (ADF)- Unit Root test at lag1		Granger Causality Tests	
		t-stat.	P-value	F-statistcs	P-value
Model (1)	REIT=>SMR	-25.870	<2e-16***	5.6561	0.0035**
	REMD=>SMR	-24.654	<2e-16***	3.3299	0.0359**
	ROIL=>SMR	-26.486	<2e-16***	18.7740	7.954e-09***
Model (2)	NREMD=>NSM	-0.385	0.7010	1.1042	0.3316
	NOIL=>NSM	-2.253	0.0244**	8.5007	0.0002***

Signif. codes: '***' 0.01 '**' 0.05 '*' 0.1

5.3 Results of WM for model

Table 6 presents the wavelet transform (WT) function results for the output variable (SMR/NSM) using an 80% training dataset, comparing the performance of various WT functions paired with ARIMA models. Model accuracy was evaluated using four error metrics: ME, RMSE, MAE, and MASE. The Haar wavelet paired with ARIMA(5,0,0) exhibits minimal bias (ME: 0.0000243) and moderate predictive accuracy (RMSE: 0.0715, MAE: 0.0466), but its MASE (0.2814). The d4 wavelet with ARIMA(1,0,0) achieves the lowest ME (0.0000212) and lowest MAE (0.0459), with competitive RMSE (0.0705) and exceptional MASE (0.0174). The LA8 wavelet with ARIMA(1,0,0) shows slight bias (ME: 0.0000334) and mid-range errors (RMSE: 0.0707, MAE: 0.0471, MASE: 0.1911). The bl14 wavelet with ARIMA(4,0,4) is the weakest performer, with the highest ME (0.0001622), worst RMSE (0.0784), worst MAE (0.0556), and poorest MASE (0.2424). The C6 wavelet with ARIMA(1,0,0) delivers the best RMSE (0.0700) and second-best MAE (0.0468), but its MASE (0.1561) is significantly higher than the d4 variant. Optimal Selection, the d4 wavelet with ARIMA(1,0,0) is superior, combining the lowest MAE, near-best RMSE, and unmatched MASE (indicating near-perfect scalability), while maintaining minimal bias and model simplicity in model (1).

In model (2), The Haar wavelet with ARIMA(5,1,0) achieves the strongest performance: minimal bias (ME: 0.000109), lowest errors across RMSE (0.0139), MAE (0.0074), and MASE (0.0703). The d4 wavelet with ARIMA(5,1,0) shows similar low bias (ME: 0.000069) but higher errors (RMSE: 0.0153, MAE: 0.0079, MASE: 0.1566). The LA8 wavelet with ARIMA(0,1,3) with drift and C6 wavelet with ARIMA(0,1,2) with drift exhibit moderate bias (ME: 0.000067/0.000427) but significantly elevated errors (RMSE: 0.0309/0.0318, MAE: 0.0151/0.0158, MASE: 0.7326/0.7337). The bl14 wavelet with ARIMA(1,1,0) with drift, despite near-zero bias (ME: 0.000002), yields the weakest results: highest RMSE (0.0435), MAE (0.0222), and MASE (0.8539).

Table 6. Comparative analysis of WT functions for time series forecasting (SMR/NSM) on 80% dataset

Models	WT function	ARIMA (p,d,q)	ME	RMSE	MAE	MASE
Model (1) SMR	Haar	ARIMA(5,0,0) with non-zero mean	0.0000243	0.0714972	0.0465899	0.28139942
	d4	ARIMA(1,0,0) with non-zero mean	0.0000212	0.0705075	0.0458996	0.01742848
	LA8	ARIMA(1,0,0) with non-zero mean	0.0000334	0.0706689	0.0471436	0.19109392
	bl14	ARIMA(4,0,4) with non-zero mean	0.0001622	0.0784477	0.0555830	0.24244136
	C6	ARIMA(1,0,0) with non-zero mean	0.0000332	0.0700261	0.0468008	0.15605790
Model (2) NSM	Haar	ARIMA(5,1,0)	0.00010941	0.01391046	0.00742997	0.07030864
	d4	ARIMA(5,1,0)	0.00006860	0.01531472	0.00792509	0.15660277
	LA8	ARIMA(0,1,3) with drift	0.00006674	0.03087276	0.01510803	0.73263673
	bl14	ARIMA(1,1,0) with drift	0.00000218	0.04350793	0.02217769	0.85389322
	C6	ARIMA(0,1,2) with drift	0.00042735	0.03178395	0.01580071	0.73373673

In table 7, The standalone WM model performs worst, showing the highest ME (0.4889), highest RMSE (0.5144), and highest MAE (0.4889). Among MODWT variants, MODWT-LA8-WM achieves the lowest RMSE (0.5071) and lowest MAE (0.4777), while MODWT-Haar-WM yields the lowest ME (0.4775). Other models cluster closely in performance: MODWT-d4-WM (RMSE: 0.5082), MODWT-bl14-WM (RMSE: 0.5072), and MODWT-C6-WM (RMSE: 0.5072). The WM+ARIMA direct hybrid underperforms all MODWT methods (RMSE: 0.5123). MODWT-LA8-WM is optimal, delivering the lowest overall prediction error (RMSE: 0.5071 and MPE/MAPE:193.39188).

Table 7. Comparative Analysis of WT functions with WM model on 20% dataset.

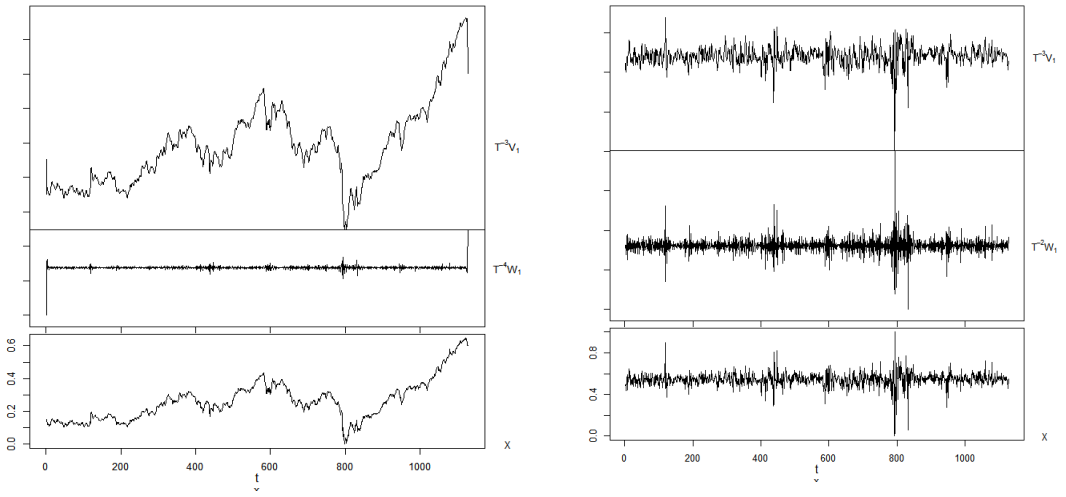
Models	WT-functions	ARIMA (p,d,q)	ME	RMSE	MAE	MPE	MAPE
Model (1)	WM		0.017484448	0.05595585	0.04232359	3.1415715	7.693322
	MODWT-Haar-WM	ARIMA(5,0,0) with non-zero mean	0.04703219	0.08347616	0.07013046	9.8435865	14.20015
	MODWT-d4-WM	ARIMA(1,0,3) with non-zero mean	0.045624002	0.08243784	0.06872849	9.4343836	13.790122

	MODWT-LA8-WM	ARIMA(1,0,0) with non-zero mean	0.045934525	0.07563944	0.06198297	9.1679558	12.49258
	MODWT-bl14-WM	ARIMA(1,0,0) with non-zero mean	0.037286744	0.06841368	0.0562723	7.3111347	11.143294
	MODWT-C6-WM	ARIMA(1,0,0) with non-zero mean	0.004739437	0.05704839	0.04196709	0.9043283	7.762031
	WM+ARIMA direct	ARIMA(1,0,1) with non-zero mean	0.027161288	0.06007143	0.04555499	4.8171896	8.042239
Model (2)	WM		0.4888634	0.5143932	0.4888634	201.42239	201.42239
	MODWT-Haar-WM	ARIMA(5,1,0)	0.477505	0.5076291	0.477505	194.24748	194.24748
	MODWT-d4-WM	ARIMA(5,1,0)	0.478008	0.5081521	0.478008	194.94025	194.94025
	MODWT-LA8-WM	ARIMA(0,1,4) with drift	0.4777372	0.5071005	0.4777372	193.39188	193.39188
	MODWT-bl14-WM	ARIMA(1,1,0) with drift	0.4779889	0.5072267	0.4779889	193.49367	193.49367
	MODWT-C6-WM	ARIMA(1,1,0) with drift	0.47779	0.5071562	0.47779	193.46531	193.46531
	WM+ARIMA direct	ARIMA(1,1,1)	0.4821028	0.512321	0.4821028	200.59867	200.59867

In Figure 3, In model (1), the visualization integrates MODWT (C6) decomposition and ARIMA modeling: the top-left plot displays T^3V_1, representing the approximation coefficients (smooth component) at decomposition level 3, which captures medium-term trends through low-frequency oscillations along the time axis (0–1000). Adjacent to it, the top-right plot shows T^2W_1, depicting detail coefficients (wavelet component) at level 2, where sharp peaks indicate high-frequency fluctuations and transient events. Dominating the center section, T^2V1 plots squared approximation coefficients at the coarsest level (C6), quantifying trend energy (variance) with y-values (0.0–0.8)—peaks signal high-volatility phases (requiring ARIMA adjustment), while valleys reflect stable periods.

In model (2), the visualization integrates MODWT (LA8) decomposition with ARIMA(0,1,4) with drift modeling: the T^3V_1 plot displays level-3 approximation coefficients, capturing the medium-term trend with amplitude fluctuations (0.0–0.6) over time indices 0–1000—where valleys (~0.0) reflect stable phases and peaks (~0.6) signal trend shifts necessitating differencing—while the T^4W_1 plot shows level-4 detail coefficients, exposing persistent volatility clusters (e.g., spikes at *t*=400/800) that disrupt long-term dynamics; the ARIMA(0,1,4) with drift processes the T^3V_1-reconstructed series, where first-order differencing (I=1) stabilizes non-stationary trends in T^3V_1, four

moving-average terms (MA(4)) model short-term disturbances linked to T^4W_1's volatility spikes, and the drift (μ) quantifies the persistent linear trend observed in T^3V_1's baseline, synthesizing multi-scale decomposition for robust forecasting.



Model(2)-MODWT(La8) Model(1)-MODWT (C6)

Figure 3. MODWT of La8 and C6

6. Limitations and Future Work

This study has several limitations warranting further investigation. First, the restricted input variables for predicting returns (SMR: REIT/REMD/ROIL returns) and closing prices (NSM: REMD/NOIL prices) may overlook other influential factors; future research should incorporate additional variables (such as macroeconomic variables, or sector-specific metrics) to enhance predictive power. Second, analysis was confined to Saudi Arabia's Tadawul market; expanding validation to international exchanges (e.g., U.S., European, Asian, and MENA markets) would test model generalizability. Third, the 2017–2022 daily data timeframe limits exposure to diverse market regimes; extending the study period to include multiple economic cycles would strengthen robustness. Future work should also explore hybridizing the MODWT-WM framework with deep learning architectures and real-time implementation for trading systems.

7. Conclusions

This study significantly advances the forecasting of Saudi stock market returns (SMR) and closing prices (NSM) through a novel hybrid MODWT-WM framework. Rigorous diagnostics confirmed effectively no multicollinearity among predictors for both models, ensuring robust variable selection. Regression analyses revealed distinct drivers: Model (1) (SMR) showed strong positive effects from real estate sector returns (REMD, REIT) and Brent oil returns (ROIL), with

REMD exhibiting the dominant influence. Model (2) (NSM) demonstrated a powerful positive relationship with oil prices (NOIL) and a significant negative effect from real estate prices (NREMD) in OLS, though fixed/random effects models aligned on positive contributions from both. The hybrid approach outperformed traditional methods, with wavelet-fuzzy integration (MODWT-C6-WM for returns, MODWT-La8-WM for prices) achieving superior predictive accuracy. These results underscore the efficacy of combining spectral decomposition with interpretable fuzzy rules to navigate Tadawul's non-linear dynamics, offering stakeholders actionable insights for strategic decision-making amid volatility. Future work should expand variable scope and integrate deep learning architectures. Critically, while the Wang-Mendel method relies on classical fuzzy logic for rule extraction, its inability to disentangle indeterminacy from randomness (e.g., in Tadawul stock volatility analysis) underscores the potential for neutrosophic enhancements—a direction will explore in future research sections.

Refrences

1. Al Rahahleh, N.; Kao, R. Forecasting volatility: Evidence from the Saudi stock market. *Journal of Risk Financial Management* **2018**, *11*, 84.
2. Alsalloum, A.S. The Impact of the Financial Reforms Influenced by Religious Financial Principles on the Stock Market in the Kingdom of Saudi Arabia. Victoria University, 2023.
3. Otto, H.-P. *Essays on the Saudi Stock Exchange*; Tectum-ein Verlag in der Nomos Verlagsgesellschaft mbH & Co. KG: 2022.
4. Alqahtani, A.; Bouri, E.; Vo, X.V. Predictability of GCC stock returns: The role of geopolitical risk and crude oil returns. *Economic Analysis Policy* **2020**, *68*, 239-249.
5. Alkhatib, K.; Almahmood, M.; Elayan, O.; Abualigah, L. Regional analytics and forecasting for most affected stock markets: The case of GCC stock markets during COVID-19 pandemic. *International Journal of System Assurance Engineering and Management* **2022**, 1-11.
6. Ziadat, S.A.; McMillan, D.G. Oil-stock nexus: The role of oil shocks for GCC markets. *Studies in Economics and Finance* **2022**, *39*, 801-818.
7. Bouri, E.; Hammoud, R.; Abou Kassm, C. The effect of oil implied volatility and geopolitical risk on GCC stock sectors under various market conditions. *Energy Economics* **2023**, *120*, 106617.
8. Hamadneh, N.N.; Jaber, J.J.; Sathasivam, S. Estimating Volatility of Saudi Stock Market Using Hybrid Dynamic Evolving Neural Fuzzy Inference System Models. *Journal of Risk Financial Management* **2024**, *17*, 377.
9. Al Wadi, S.; Al Singlawi, O.; Jaber, J.J.; Saleh, M.H.; Shehadeh, A.A. Enhancing Predictive Accuracy through the Analysis of Banking Time Series: A Case Study from the Amman Stock Exchange. *Journal of Risk Financial Management* **2024**, *17*, 98.
10. Alshammari, T.T.; Ismail, M.T.; Hamadneh, N.N.; Al Wadi, S.; Jaber, J.J.; Alshammari, N.; Saleh, M.H. Forecasting Stock Volatility Using Wavelet-based Exponential Generalized Autoregressive Conditional Heteroscedasticity Methods. *Intelligent Automation Soft Computing* **2023**, *35*, 2589-2601.
11. Jarrah, M. Long Short-Term Memory and Discrete Wavelet Transform based Univariate Stock Market Prediction Model. *Journal of Information Organizational Sciences* **2024**, *48*, 263-277.

12. Alasiri, R.A.; Qahmash, A. Analysis and Forecasting of Saudi Stock Market Using Time Series Algorithms. In Proceedings of the 2023 3rd International Conference on Computing and Information Technology (ICCIT), 2023; pp. 340-347.
13. Alenezy, A.H.; Ismail, M.T.; Wadi, S.A.; Tahir, M.; Hamadneh, N.N.; Jaber, J.J.; Khan, W.A. Forecasting stock market volatility using hybrid of adaptive network of fuzzy inference system and wavelet functions. *Journal of Mathematics* **2021**, *2021*, 9954341.
14. Mhmoud, A.; Dawalbait, F. Estimating and forecasting stock market volatility using GARCH Models: Empirical Evidence from Saudi Arabia. *International journal of engineering research technology* **2015**, *4*, 464-471.
15. Kreinovich, V.; Nguyen, H.T.; Sriboonchitta, S. A new justification of Wang transform operator in financial risk analysis. *International Journal of Intelligent Technologies and Applied Statistics* **2009**, *35*.
16. Zhang, L.; Gao, X.-W.; Li, S.-M.; Zhao, J.-P.; Wang, J.-S. Constrained fuzzy predictive control based on Wang-Mendel model. *Control Decision* **2010**, *25*, 1384-1388.
17. Smarandache, F.; Abdel-Basset, M.; Vazquez, M. Neutrosophic Sets and Systems, Vol. 86, 2025. *Neutrosophic Sets Systems* **2025**, *86*, 61.
18. Zadeh., L.A. Fuzzy sets. . *Information and control* **1965**, *8*, 338–353.
19. Atanassov, K. Intuitionistic fuzzy sets. fuzzy sets and systems 20 (1), 87-96. **1986**, 80034-80033.
20. Smarandache, F. A unifying field in Logics: Neutrosophic Logic. In *Philosophy*; American Research Press: 1999; pp. 1-141.
21. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society* **1982**, 987-1007.
22. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* **1986**, *31*, 307-327.
23. Dai, Z.; Zhu, H. Dynamic risk spillover among crude oil, economic policy uncertainty and Chinese financial sectors. *International Review of Economics Finance* **2023**, *83*, 421-450.
24. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting* **1998**, *14*, 35-62.
25. Ciner, C. Do industry returns predict the stock market? A reprise using the random forest. *The Quarterly Review of Economics Finance* **2019**, *72*, 152-158.
26. Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications* **2015a**, *42*, 259-268.
27. Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting stock market index using fusion of machine learning techniques. *Expert systems with applications* **2015b**, *42*, 2162-2172.
28. Alenezy, A.H.; Ismail, M.T.; Jaber, J.J.; Wadi, S.A.; Alkhawaldeh, R.S. Hybrid fuzzy inference rules of descent method and wavelet function for volatility forecasting. *Plos one* **2022**, *17*, e0278835.
29. Alenezy, A.H.; Ismail, M.T.; Wadi, S.A.; Jaber, J.J. Predicting stock market volatility using modwt with hyfis and fs. hgd models. *Risks* **2023**, *11*, 121.
30. McBratney, A.B.; Minasny, B.; Stockmann, U. *Pedometrics*; Springer: 2018.

31. Zadeh, L.A. Calculus of fuzzy restrictions. In *Fuzzy sets and their applications to cognitive and decision processes*; Elsevier: 1975; pp. 1-39.
32. Mamdani, E.H.; Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies* **1975**, *7*, 1-13.
33. Mamdani, E.H. Application of fuzzy algorithms for control of simple dynamic plant. In Proceedings of the Proceedings of the institution of electrical engineers, 1974; pp. 1585-1588.
34. Riza, L.S.; Bergmeir, C.; Herrera, F.; Benítez, J.M. frbs: Fuzzy rule-based systems for classification and regression in R. *Journal of statistical software* **2015**, *65*, 1-30.
35. Sugeno, M.; Kang, G. Structure identification of fuzzy model. *Fuzzy sets systems* **1988**, *28*, 15-33.
36. Sugeno, M.; Yasukawa, T. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on fuzzy systems* **1993**, *1*, 7.
37. Wang, W.; Shao, J.; Jumahong, H. Fuzzy inference-based LSTM for long-term time series prediction. *Scientific Reports* **2023**, *13*, 20359.
38. Wang, L.-X.; Mendel, J.M. Generating fuzzy rules by learning from examples. *IEEE Transactions on systems, man, cybernetics* **1992**, *22*, 1414-1427.
39. Ishibuchi, H.; Nozaki, K.; Tanaka, H.; Hosaka, Y.; Matsuda, M. Empirical study on learning in fuzzy systems by rice taste analysis. *Fuzzy sets systems* **1994**, *64*, 129-144.
40. Casal-Guisande, M.; Cerqueiro-Pequeño, J.; Bouza-Rodríguez, J.-B.; Comesaña-Campos, A. Integration of the Wang & Mendel algorithm into the application of fuzzy expert systems to intelligent clinical decision support systems. *Mathematics* **2023**, *11*, 2469.
41. Saleh, M.H.; Alkhawaldeh, R.S.; Jaber, J.J. A predictive modeling for health expenditure using neural networks strategies. *Journal of Open Innovation: Technology, Market,*

Complexity **2023**, *9*, 100132.

Received: Jan. 15, 2025. Accepted: July 5, 2025