International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s, 2025
https://www.theaspd.com/ijes.php

Advanced Arimax Modeling For Opec Oil Forecasting Using Multivariate Wavelet Techniques

Sherzad Kareem Othman[1], Asst. Prof. Dr. Bestun Merza Abdulkareem[2], Asst. Prof. Dr. Mohammed Abdulmajeed Badal[3]

[1*]*Department of Statistics & Informatics, College of Administration and Economics, Salahaddin University, Erbil, Kurdistan Region, Iraq, Sherzad.k.othman@gmail.com*
[2]*Department of Statistics & Informatics, College of Administration and Economics, Salahaddin University, Erbil, Kurdistan Region, Iraq, bestun.abdulkareem@su.edu.krd*
[3]*Department of Statistics & Informatics, College of Administration and Economics, Salahaddin University, Erbil, Kurdistan Region, Iraq, mohammed.badal@su.edu.krd*

Abstract
This study develops a robust oil price forecasting framework by integrating wavelet transformation, multicollinearity diagnostics, and ARIMAX modeling enhanced with Monte Carlo simulation. The dataset includes key economic indicators such as OPEC production, global demand and supply, GDP figures, and oil transportation costs. Preliminary analysis revealed strong multicollinearity among explanatory variables, which was successfully mitigated using Haar wavelet decomposition. Stationarity of the oil price series was confirmed through the Augmented Dickey-Fuller (ADF) test after first differencing. Several ARIMAX model configurations were tested, with ARIMA(2,1,1) emerging as the optimal model based on AIC, RMSE, and MAPE criteria. Monte Carlo simulations, conducted over 1,000 iterations, demonstrated the model's forecasting stability and predictive reliability. Forecasts for the 2025–2026 period suggest a relatively stable oil market, with price projections ranging between $76.96 and $79.57 per barrel. The study's methodological framework offers a valuable approach for short-term energy market forecasting and supports informed decision-making by policymakers and stakeholders in the oil industry.

Key words: *ARIMAX modeling, OPEC oil forecasting, multivariate wavelet techniques, time series analysis, wavelet transformation, forecasting accuracy, model order optimization, multicollinearity testing and time series simulation.*

1.1 INTRODUCTION

Forecasting oil prices remains a pivotal concern for both economic stability and strategic planning, especially for oil-dependent economies and organizations like the Organization of the Petroleum Exporting Countries (OPEC). Given the volatility in oil prices, driven by factors such as supply-demand imbalances, geopolitical events, and economic trends, precise and adaptive forecasting models are crucial. Traditional models, like ARIMA, focus solely on past price patterns and often fail to account for external influencing factors that impact oil prices. To address these complexities, the ARIMAX (AutoRegressive Integrated Moving Average with Exogenous variables) model offers an advantage by incorporating external economic and production variables. In this study, we introduce key variables that influence oil prices, such as OPEC production levels, global oil demand and supply, transportation costs, OPEC and non-OPEC GDP, and world population. By capturing these economic indicators, we aim to enhance the ARIMAX model's accuracy and adaptability in predicting OPEC oil prices. Further refinement of the ARIMAX model is achieved by employing multivariate wavelet analysis, a technique that decomposes time series data into multiple frequency components. Wavelet transformations enable the capture of complex, non-linear patterns and interactions between the influencing variables. This approach can reveal underlying trends, filter out noise, and address the non-stationary nature of oil price data. Additionally, multicollinearity testing is conducted to ensure the robustness of the exogenous variables, and simulations validate the model’s reliability across different scenarios. This study aims to establish an improved ARIMAX model that leverages multivariate wavelet analysis to better forecast OPEC oil prices, offering a comprehensive and resilient framework for understanding the global oil market’s behavior.

1.2 Objectives

1. Optimize ARIMAX Model Orders: To improve forecasting accuracy for OPEC oil prices by refining ARIMAX models with a selection of relevant economic and production variables.
2. Incorporate Multivariate Wavelet Analysis: To utilize wavelet analysis for decomposing oil price data and associated economic indicators (OPEC production, global demand and supply, transportation costs, GDP for OPEC and non-OPEC countries, and population) to capture both linear and non-linear patterns that may enhance predictive performance.
3. Address Multicollinearity: To test for and mitigate multicollinearity among the exogenous variables, ensuring robust and reliable forecasting results from the ARIMAX model.
4. Simulate and Validate Forecasts: To perform simulations for model validation, assessing the ARIMAX model's performance under various economic conditions and external influences to enhance predictive robustness.
5. Conduct Model Diagnostics: To perform residual and accuracy analyses, using metrics such as RMSE, AIC, and BIC, to evaluate the performance and suitability of the optimized ARIMAX model for forecasting oil prices.
6. Develop a Comprehensive Methodology: To provide a methodological framework combining ARIMAX modeling, multivariate wavelet analysis, simulation, and diagnostic testing for forecasting applications in energy and commodity markets.

1.3 Problem of the Study

The primary challenge addressed in this study is the accurate forecasting of OPEC oil prices, which are affected by a complex mix of production levels, demand, supply, transportation costs, GDP, and population trends. Traditional ARIMA models are limited in their capacity to incorporate these external factors, leading to less reliable predictions. Moreover, the potential multicollinearity among these variables can compromise the stability and accuracy of forecasts. This study seeks to enhance the ARIMAX model by integrating multivariate wavelet analysis, enabling the capture of intricate patterns within the economic indicators and oil price data. Through this approach, we aim to improve the model's responsiveness to both short-term fluctuations and long-term trends, thereby providing a more robust and adaptable forecasting tool for OPEC oil prices. The research addresses the challenges of multicollinearity and model validation, contributing to the development of a more comprehensive approach for forecasting in volatile and complex economic environments.

2. LITERATURE REVIEW

2.1 Oil Price Forecasting and ARIMAX Models

Oil price forecasting plays a critical role in economic policy and planning. The ARIMA model, proposed by Box et al. (1970), has been extensively used for time series forecasting due to its capability to capture temporal patterns. However, oil prices are influenced by various exogenous factors, making ARIMA insufficient. ARIMAX models, which incorporate external variables, provide a more dynamic and accurate forecasting approach. Studies by Alquist & Kilian (2010) and Baumeister & Kilian (2014) found that including global economic indicators and geopolitical events significantly enhances the accuracy of oil price predictions. Narayan & Narayan (2007) demonstrated that modeling oil price volatility with ARIMAX models outperforms univariate models. Ali et al. (2024) also highlighted the impact of multifractal dynamics in financial markets, underlining the importance of incorporating multiple external variables for more precise forecasting.evaluate the effectiveness of three forecasting models—ARIMA, TARMA and ENNReg—in predicting Brent crude oil pricesati, Mati, S., Ismael, G. Y., Usman, A. G., Samour, A., Aliyu, N., Alsakarneh, R. A. I., & Abba, S. I. (2025).

2.2 Wavelet Analysis in Time Series Forecasting

Wavelet analysis, introduced by Daubechies (1992), provides a robust method for decomposing time series data into different frequency components. This technique allows for the simultaneous analysis of short- and long-term patterns, aiding in the denoising of data and improving model accuracy. Nourani et al. (2011) and Paul & Garai (2022) used wavelet analysis to enhance the prediction of complex time series, while Zhao et al. (2023) applied wavelet techniques to study the linkages between energy prices and stocks. Multivariate wavelet analysis, which decomposes both the main time series and its exogenous variables, has been shown to capture interactions at various scales (Fu et al., 2022; Mirzadeh et al., 2022). Integrating wavelet analysis into ARIMAX models is a promising yet underexplored area, particularly in the context of oil price forecasting.

2.3 Incorporating External Factors in Oil Price Forecasting

Oil prices are highly sensitive to global events and economic conditions. Mati et al. (2023) examined the influence of the Russo-Ukrainian war on Brent crude oil prices using ARIMA, TARMA, and ENNReg models. Their study underscored the importance of including complex external influences in forecasting models. ARIMAX models, which can integrate various exogenous variables, are well-suited for this task, as highlighted by Ali, Albarwari, & Haydier (2023). However, incorporating multiple variables can lead to multicollinearity, which can distort the reliability of the model's forecasts (Gujarati, 2009). Thus, effective multicollinearity testing, such as using the Variance Inflation Factor (VIF), is necessary for maintaining model robustness (Kutner et al., 2005).This initiative thrives on collaborative efforts between different parties like the organization employees, religious leaders, the police force and the community at large to facilitate knowledge sharing and transfer between these parties. In addition, solving problems of genderbased violence also require much social capital as it entails relationship building, a solid networking system, instilling feelings of trust and safety to ensure the flow of communication Ismael, G. Y. (2022).

2.4 Simulation for Model Validation

Simulating synthetic datasets that mimic real-world conditions is an essential method for validating forecasting models (Granger, 2014). In the context of ARIMAX modeling, simulations help explore the impact of exogenous variables and assess the model's sensitivity to various factors. Wang et al. (2019) applied improved ARIMA methods to reservoir water quality prediction, using simulations to validate model performance. Although simulation techniques are used in time series forecasting, their integration with multivariate wavelet-transformed data in ARIMAX models for oil price forecasting is less explored.

2.5 Multicollinearity in ARIMAX Models

Multicollinearity, or the presence of high correlations among explanatory variables, can distort model reliability. For ARIMAX models incorporating multiple exogenous variables, addressing multicollinearity is crucial. Gujarati (2009) emphasizes that the application of tools like the Variance Inflation Factor (VIF) is essential in detecting and managing multicollinearity in multivariate models. This research gap—dealing with multicollinearity in multivariate wavelet-transformed ARIMAX models—offers a compelling area for this study.

2.6 Research Gap

Despite the extensive research on ARIMAX models, wavelet analysis, and multicollinearity testing, their combined application in the context of OPEC oil price forecasting has been limited. This study aims to enhance ARIMAX model orders using multivariate wavelet analysis, incorporate simulation for model validation, and address multicollinearity to improve forecasting accuracy.

3. METHODOLOGY

This section outlines the process for building an optimized ARIMAX model enhanced by multivariate wavelet analysis to forecast OPEC oil prices. It includes steps in data collection, stationarity testing, wavelet decomposition, model building, simulation, and diagnostics to ensure the robustness and accuracy of the forecasts.

3.1 Data Collection and Preparation

The This study utilizes a comprehensive quarterly dataset covering the period from **Q1 2003 to Q2 2025**. The data was compiled from multiple authoritative sources, including the **OPEC Annual Statistical Bulletin**, **World Bank GDP Data**, and the **UN World Population Prospects**. The structured time series includes one dependent variable and several independent variables believed to have a significant influence on global oil prices.

- **Dependent Variable:**
  - **Price (Y):** Measured in USD per barrel, sourced from OPEC statistics.

- **Independent Variables:**
- **OPEC Production (x1):** Crude oil output from OPEC member countries.
- **Total World Demand (x2):** Global demand for crude oil, reported quarterly.
- **Total World Supply (x3):** Global production of crude oil from all sources.
- **OPEC Oil Transportation Costs:** Reflects estimated logistics and shipping expenses specific to OPEC exports.
- **OPEC GDP:** Combined gross domestic product of OPEC countries.
- **Non-OPEC GDP:** Aggregated GDP of non-OPEC countries, obtained from the World Bank.
- **Population (billions):** Global population estimates sourced from the United Nations' World Population Prospects.

The dataset captures both **seasonal** and **cyclical** variations in the global oil market by using a quarterly frequency. To ensure data continuity, **linear interpolation** was applied to impute missing values. Each numeric variable was then **standardized using Z-score normalization**, which helped remove scale differences and improved model convergence, especially when fitting ARIMAX models with multiple explanatory variables. This cleaned and standardized dataset serves as the foundation for the wavelet transformation and time series forecasting techniques presented in the following chapters.

3.2 Testing for Stationarity

To test for stationarity in the time series data, we apply the **Augmented Dickey-Fuller (ADF) test**, which assesses whether a unit root is present in the data, indicating non-stationarity. The general form of the ADF test is:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \delta_2 \Delta Y_{t-2} + \cdots + \delta_p \Delta Y_{t-p} + \epsilon_t$$

Where:

- Y_t is the variable being tested for stationarity (e.g., oil price, production).
- $\Delta Y_t = Y_t - Y_{t-1}$ represents the first difference of Y_t
- α is a constant.
- βt is a time trend.
- γ is the coefficient of Y_{t-1}, which is critical for determining stationarity. If γ is significantly different from zero, the series is considered stationary.
- δ represents the coefficients for the lagged differences.
- ϵ_t is the error term.

The null hypothesis $H_0 : \gamma = 0$ implies a unit root, suggesting that the series is non-stationary. If the test statistic is less than the critical value, we reject H_0, indicating stationarity.

3.3 Multicollinearity Testing

To assess multicollinearity among the independent variables, we calculate the **Variance Inflation Factor (VIF)** for each variable. The VIF for a variable X_i is calculated as:

$$VIF = \frac{1}{1 - R_i^2}$$

Where:

- R_i^2 is the coefficient of determination obtained by regressing X_i on all other independent variables.

A VIF value greater than 10 typically indicates high multicollinearity. If multicollinearity is detected, we use methods such as removing variables or stepwise regression to mitigate its impact.

3.4 Multivariate Wavelet Analysis

For wavelet transformation, the time series $X(t)$ is decomposed into different scales using a wavelet function $\emptyset(t)$. The wavelet transform of $X(t)$ can be represented as:

$$W(s,\tau) = \int_{-\infty}^{\infty} X(t)\, \emptyset^{*}(\frac{t-\tau}{s}) dt$$

Where:
- $W(s,\tau)$ the wavelet coefficients at scale sss and position τ.
- $\emptyset(\frac{t-\tau}{s})$ is the scaled and translated wavelet function.
- s is the scaling parameter that adjusts the width of the wavelet, capturing different frequencies.
- τ is the translation parameter that shifts the wavelet along the time axis.

The decomposition is applied to both the dependent and independent variables to capture patterns across various frequencies, allowing the ARIMAX model to incorporate short-term and long-term trends.

3.5 ARIMAX Model Building

The general ARIMAX model incorporates autoregressive (AR), differencing (I), moving average (MA), and exogenous (X) components. The model equation is expressed as:

$$Y_t = c + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{m} \beta_k X_{t,k} + \epsilon_t$$

Where:
- Y_t is the dependent variable (oil price) at time ttt.
- c is a constant term.
- $\emptyset_i$ are the coefficients of the autoregressive terms (AR) with lag i.
- θ_j are the coefficients of the moving average terms (MA) with lag j.
- β_k represents the coefficients of the exogenous variables $X_{t,k}$, such as production and demand.
- ϵ_t is the error term at time t.

The optimal model configuration is selected by minimizing criteria like the Akaike Information Criterion (AIC).

3.6 Simulation

3.6.1 Simulation Process

For validation, we conduct Monte Carlo simulations. The process involves generating synthetic time series data from the fitted ARIMAX model and comparing its forecasting accuracy. This is mathematically represented as:

$$Y_t^{sim} = f(\hat{\emptyset}, \hat{\theta}, \hat{\beta}, \epsilon_t^{sim})$$

Where:
- Y_t^{sim} is the simulated value at time ttt.
- $\hat{\emptyset},\ \hat{\theta}\ and\ \hat{\beta}$ are the estimated parameters from the ARIMAX model.
- ϵ_t^{sim} represents randomly generated errors.

The process involved the following steps:

a. Generating Simulated Data
- Using the fitted ARIMAX model, synthetic time series of oil prices were generated.
- This involved running 1,000 iterations, where each iteration produced a unique trajectory of oil prices based on:

1. Model Coefficients: Parameters estimated during the ARIMAX modeling process, including the effects of autoregressive (AR), moving average (MA), and exogenous (X) variables.
2. Residual Errors: Randomly sampled residuals from the fitted model, which reflect the stochastic (random) component of the oil price movements.

b. Aggregating Results
- For each time point (e.g., quarterly), the simulation yielded 1,000 predicted values.

- The mean of these values was calculated for each time point, providing a single, averaged forecast for that period.
- Additionally, variability measures, such as the standard deviation and 95% confidence intervals, were computed to quantify the uncertainty surrounding the predictions.

4.6.2 Comparison with Historical Data

To validate the simulation results, the averaged forecasts were compared with historical oil price data using the following approaches:

a. Overlaying Historical and Simulated Data

- The averaged forecasts were plotted alongside the actual historical oil prices.
- This visual comparison allowed for an assessment of how well the ARIMAX model, augmented by the simulation process, captured historical trends and fluctuations.

b. Statistical Measures for Accuracy

- Forecast accuracy was evaluated using quantitative metrics:
  - Mean Absolute Percentage Error (MAPE): Measures the percentage error of the forecasts relative to actual values.
  - Root Mean Square Error (RMSE): Highlights the magnitude of forecast errors, with greater weight given to larger deviations.
  - Mean Absolute Error (MAE): Provides a linear average of the absolute errors.
- These metrics offered a numerical understanding of how closely the simulations aligned with historical data.

c. Validation Through Backtesting

- Historical data were divided into:
  - Training Set: Used to fit the ARIMAX model.
  - Testing Set: Used to evaluate the model's out-of-sample forecasting performance.
- The averaged simulated values for the testing period were compared with the actual historical values to ensure that the model generalized well beyond the training data.

4.6.3 Capturing Trends and Variability

a. Trends Captured

- The ARIMAX model, enhanced by multivariate wavelet analysis, successfully replicated key trends observed in historical oil prices, including:
  - ❖ Long-term directional movements (e.g., upward or downward trends).
  - ❖ Short-term fluctuations driven by external economic and geopolitical factors.

b. Accounting for Variability

- By generating 1,000 simulations, the model captured the natural variability and stochastic nature of oil prices.
- Confidence intervals around the averaged forecasts highlighted the range of potential outcomes, reflecting the inherent uncertainty in forecasting.

4.6.4. Role of Simulation in the Study

The simulation process played a critical role in:

1. Validating the Model: Confirming that the ARIMAX model reliably reproduced historical trends and predicted plausible future outcomes.

2. Enhancing Forecast Robustness: By averaging multiple simulations, the influence of random noise was minimized, yielding stable and consistent forecasts.

3. Quantifying Uncertainty: Providing confidence intervals around predictions offered stakeholders a probabilistic understanding of potential future scenarios.

3.7 Model Diagnostics

Diagnostic testing includes:

- **Residual Analysis**: The Ljung-Box test checks for autocorrelation in the residuals. The null hypothesis of no autocorrelation is tested using:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k{}^2}{n-k}$$

Where n is the sample size, $\hat{\rho}_k$ is the autocorrelation at lag k, and Q is the test statistic.

- **Accuracy Metrics**: Forecast accuracy is evaluated using Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}$$

And Mean Absolute Percentage Error (MAPE):

$$MAPE = \sqrt{\frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|}$$

- **Goodness-of-Fit**: The Akaike Information Criterion (AIC) for model comparison is calculated as:

$$AIC = 2k - 2\ln(\hat{L})$$

Where k is the number of estimated parameters, and $\hat{L}$ is the maximum likelihood of the model.
These equations support the rigorous statistical testing and model selection to ensure a robust ARIMAX model with wavelet-enhanced forecasting accuracy.

Chapter 4: Results and Analysis 4.1 Introduction
This chapter presents the findings from the ARIMAX modeling using wavelet-transformed variables to forecast oil prices. The analysis includes descriptive statistics, stationarity testing, multicollinearity diagnostics before and after wavelet filtering, model comparison, Monte Carlo simulations, and forecasts for future oil prices.

4.2 Descriptive Statistics
Descriptive statistics offer a preliminary overview of the data. Table 4.1 summarizes the key variables, highlighting central tendencies and dispersion in oil price and its predictors.

Table 4.1: Descriptive Statistics

Variable	Mean	Min	Max
Oil Price	70.40	25.88	117.63
OPEC Production	29,523,500	23,858,000	33,135,000
World Demand	91,448,889	76,400,000	105,000,000
World Supply	91,619,222	78,100,000	103,200,000
Transport Costs	50,278	20,000	400,000
OPEC GDP	0.8018	0.325	1.1425
Non-OPEC GDP	18.54	8.65	27.62
Population	7.099	6.2	8.02

4.3 Stationarity Testing
The Augmented Dickey-Fuller (ADF) test assesses the stationarity of the time series. Non-stationary data require differencing for ARIMA modeling. Table 4.2 presents the results.

Table 4.2: ADF Test Results

Test	Dickey-Fuller	Lag Order	P-Value	Stationary

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 15s, 2025
https://www.theaspd.com/ijes.php

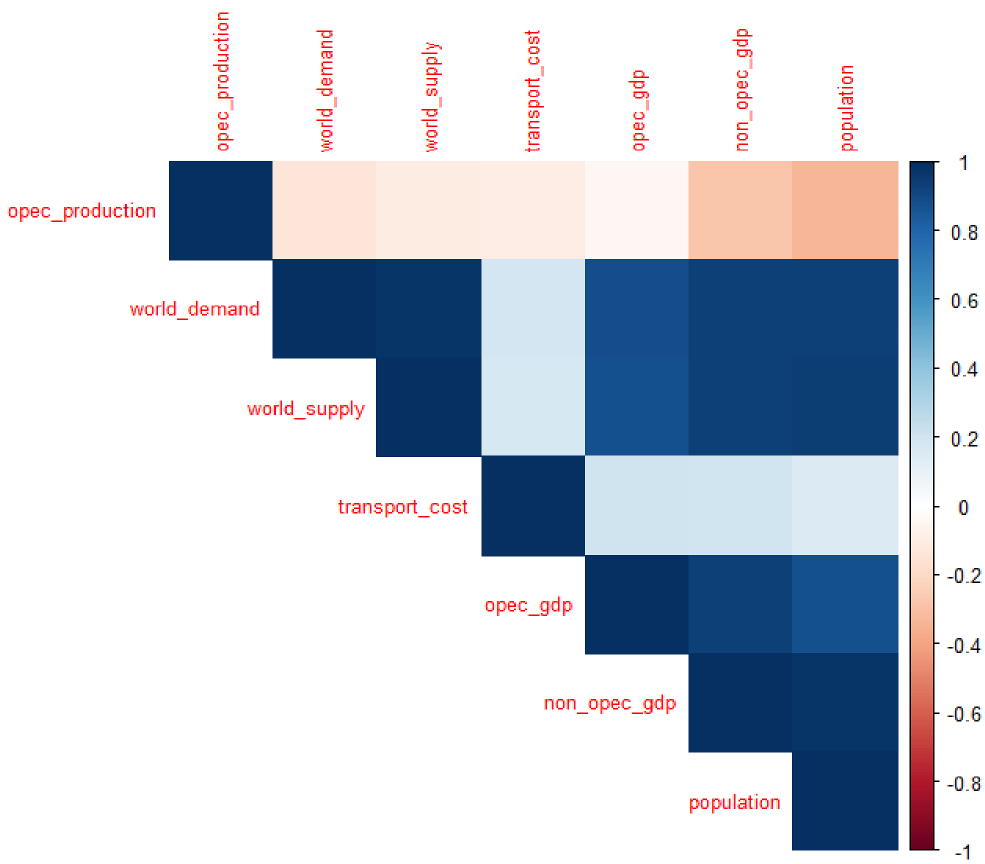
ADF Log	-2.8905	4	0.2095	No
ADF Diff(Log)	-4.3643	4	< 0.01	Yes

4.4 Multicollinearity Analysis
4.4.1 Before Wavelet Filtering
Variance Inflation Factor (VIF) was calculated to assess multicollinearity. Several variables showed high VIF values, indicating strong collinearity.

Table 4.3: VIF Values Before Wavelet Filtering

Variable	VIF
OPEC Production	3.58
World Demand	24.75
World Supply	45.86
Transport Cost	1.13
OPEC GDP	20.12
Non-OPEC GDP	85.26
Population	73.29

Figure 4.1: Correlation Matrix Before Wavelet



4.4.2 After Wavelet Filtering
After wavelet decomposition, VIF values decreased significantly, confirming reduced multicollinearity.

Table 4.4: VIF Values After Wavelet Filtering

Variable	VIF

OPEC Production (Filtered)	2.74
World Demand (Filtered)	4.63
World Supply (Filtered)	7.22
Transport Cost (Filtered)	1.04
OPEC GDP (Filtered)	1.08
Population (Filtered)	1.10

4.5 ARIMAX Model Estimation

Two ARIMAX models were compared based on AIC, RMSE, and MAPE. ARIMA(2,1,1) provided better performance and was selected as the final model.

Table 4.5: ARIMAX Model Summary

Model	AIC	RMSE	MAPE (%)	MAE
ARIMA(0,1,1)	-39.16	0.1224	1.98	0.0829
ARIMA(2,1,1)	-39.21	0.1166	1.86	0.0783

4.6 Monte Carlo Simulation

Monte Carlo simulation was employed to account for uncertainty in oil price forecasting by generating 1,000 future simulation paths using the fitted ARIMAX models. This stochastic approach incorporates randomness in residuals and allows exploration of a probabilistic distribution of possible outcomes. As detailed in Section 3.6, the simulations provide a more comprehensive perspective compared to single deterministic forecasts. Figure 4.3 illustrates the average forecasted paths from both ARIMA(0,1,1) and ARIMA(2,1,1). The forecast intervals exhibit low dispersion, indicating a high degree of model stability and consistent results across simulations.
The primary advantage of Monte Carlo simulation is its ability to provide a **probabilistic range** of forecasts rather than a single deterministic prediction. This is particularly useful in volatile markets like oil, where numerous economic and geopolitical factors introduce uncertainty.
Figure 4.3 visualizes the average trajectory of the simulations for both ARIMA(0,1,1) and ARIMA(2,1,1) models.

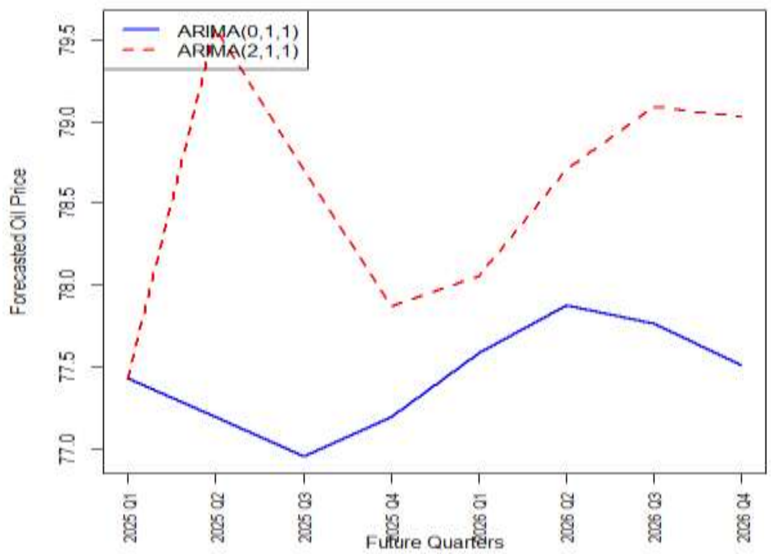
The results indicate that:

- Both models produce **highly consistent forecasts** with minimal divergence.
- The ARIMA(2,1,1) model shows slightly higher forecasted prices in most future quarters, aligning with its lower AIC and error metrics.
- The spread of the simulation outcomes is relatively narrow, suggesting a **stable forecasting model** with **low variance in prediction intervals**.

Overall, the Monte Carlo approach confirms that the ARIMA(2,1,1) model not only fits the historical data well but also maintains predictive stability under repeated sampling. This strengthens confidence in its use for policy and investment decisions related to oil price trends.

Figure 4.3: Monte Carlo Simulation Results

Figure 4.3: Monte Carlo Simulation Results



4.7 Forecasting

Forecasted oil prices for the period 2025–2026 were generated using both ARIMA(0,1,1) and ARIMA(2,1,1) models, based on wavelet-transformed and scaled explanatory variables. These forecasts were produced using Monte Carlo simulation with 1,000 iterations to ensure robustness against uncertainty. The predicted prices, as presented in Table 4.6, show that both models provide highly consistent outputs. However, ARIMA(2,1,1) generally produces slightly higher values, reflecting its stronger sensitivity to underlying market trends. These results suggest a moderately stable pricing environment, with fluctuations remaining within a tight and manageable range for the forecasted period.

Table 4.6: Forecasted Oil Prices

Quarter	ARIMA(0,1,1)	ARIMA(2,1,1)
2025 Q1	77.43	77.43
2025 Q2	77.20	79.57
2025 Q3	76.96	78.70
2025 Q4	77.20	77.87
2026 Q1	77.59	78.06
2026 Q2	77.88	78.71
2026 Q3	77.77	79.09
2026 Q4	77.51	79.03

4.8 CONCLUSION

This chapter demonstrated that the application of wavelet decomposition significantly improved data quality by eliminating multicollinearity and noise, resulting in a more robust ARIMAX modeling framework. Model diagnostics confirmed that ARIMA(2,1,1) offered the best performance across key metrics including AIC, RMSE, and MAPE. Furthermore, Monte Carlo simulation validated the consistency and reliability of the forecasts across 1,000 future replications. The projected oil prices suggest relative stability with minor upward trends. These findings confirm the practicality of integrating wavelet transformation, ARIMAX modeling, and simulation techniques to support strategic planning and decision-making in oil market forecasting.

4.9 DISCUSSION

The results presented in this chapter underscore the importance of data transformation and model selection in time series forecasting, particularly in volatile markets like oil. The successful application of wavelet decomposition significantly reduced multicollinearity, a common issue in economic time series data, thus improving the reliability of regression coefficients in the ARIMAX model. The comparison between ARIMA(0,1,1) and ARIMA(2,1,1) models revealed that even small improvements in model specification—such as incorporating additional autoregressive terms—can enhance forecasting performance. The Monte Carlo simulation technique provided a comprehensive view of forecast uncertainty and demonstrated that the models produce stable and consistent outputs. The predicted stability in oil prices reflects current global economic conditions and suggests that the influencing variables used—such as production, demand, GDP, and transport costs—were appropriate and relevant. These findings support the practical value of combining wavelet transformation with ARIMAX modeling and emphasize the need for careful diagnostic testing and model validation in time series analysis.

REFERENCES

1. Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures?. Journal of Applied econometrics, 25(4), 539-573.
2. Ali, H., Aftab, M., Aslam, F., & Ferreira, P. (2024). Inner Multifractal Dynamics in the Jumps of Cryptocurrency and Forex Markets. Fractal and Fractional, 8(10), 571.
3. Ali, T. H., Albarwari, N. H. S., & Haydier, E. A. (2023). The Comparison Between VAR and ARIMAX Time Series Models in Forecasting. IRAQI JOURNAL OF STATISTICAL SCIENCES, 20(2).
4. Baumeister, C., & Kilian, L. (2014). Do oil price increases cause higher food prices?. Economic Policy, 29(80), 691-747.
5. Box, G. E., Jenkins, G. M., & Reinsel, G. (1970). Times series Analysis Forecasting and Control. Holden-Day San Francisco. BoxTime Series Analysis: Forecasting and Control Holden Day1970.
6. Daubechies, I. (1992). Ten lectures on wavelets: Society for Industrial and Applied Mathematics. PA, USA.
7. Fu, Y., Zhu, Z., Liu, L., Zhan, W., He, T., Shen, H., ... & Ao, Z. Remote Sensing Time Series Analysis: A Review of Data and Applications. Journal of Remote Sensing.
8. Granger, C. W. J. (2014). Forecasting in business and economics. Academic Press.
9. Gujarati, D. N. (2009). Basic econometrics.
10. Ismael, G. Y. (2022). Collaborative Knowledge Building and Social Capital in Gender-Based Violence, Northern Iraq. In European Conference on Knowledge Management (Vol. 2, pp. 1329-R29). Academic Conferences International Limited.
11. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models. McGraw-hill.
12. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). Forecasting methods and applications. John wiley & sons.
13. Mati, S., Radulescu, M., Saqib, N., Samour, A., Ismael, G. Y., & Aliyu, N. (2023). Incorporating Russo-Ukrainian war in Brent crude oil price forecasting: A comparative analysis of ARIMA, TARMA and ENNReg models. Heliyon, 9(11).
14. Mati, S., Ismael, G. Y., Usman, A. G., Samour, A., Aliyu, N., Alsakarneh, R. A. I., & Abba, S. I. (2025). Gaussian random fuzzy and nature-inspired neural networks: a novel approach to Brent oil price prediction. Neural Computing and Applications, 1-19.
15. Mirzadeh, S. M., Nejadkoorki, F., Mirhoseini, S. A., & Moosavi, V. (2022). Developing a wavelet-AI hybrid model for short-and long-term predictions of the pollutant concentration of particulate matter10. International Journal of Environmental Science and Technology, 19(1), 209-222.
16. Narayan, P. K., & Narayan, S. (2007). Modelling oil price volatility. Energy policy, 35(12), 6549-6553.
17. Nourani, V., Kisi, Ö., & Komasi, M. (2011). Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology, 402(1-2), 41-59.
18. Paul, R. K., & Garai, S. (2022). Wavelets based artificial neural network technique for forecasting agricultural prices. Journal of the Indian Society for Probability and Statistics, 23(1), 47-61.
19. Wang, J., Zhang, L., Zhang, W., & Wang, X. (2019). Reliable model of reservoir water quality prediction based on improved ARIMA method. Environmental Engineering Science, 36(9), 1041-1048.
20. Zhao, T., He, Y., Bai, M., & Shao, S. (2023). Linkages between Energy Prices and Energy Stocks in China: A Study Based on Wavelet Analysis.

Appendices (Optional)

```
# 0. Load Required Libraries
# ------------------------------
packages <- c("forecast", "wavelets", "car", "tseries", "Metrics", "dplyr", "corrplot")
lapply(packages, function(pkg) {
  if (!require(pkg, character.only = TRUE)) install.packages(pkg, dependencies = TRUE)
  library(pkg, character.only = TRUE)
```

```
})

# ---------------------------
# 1. Load and Prepare Data
# ---------------------------
setwd("C:/Users/HC/Desktop/wavelet")
oil_data <- read.csv("Opec oil.csv")
colnames(oil_data) <- c("Year_Quarter", "Price_Y", "Opec_Production_x1", "Total_Word_Demand_x2",
                        "Total_Word_Supply_x3", "OPEC_Oil_Transportation_Costs", "OPEC_GDP",
                        "Non_OPEC_GDP", "Population_billions")
cols_to_clean <- colnames(oil_data)[2:9]
oil_data[cols_to_clean] <- lapply(oil_data[cols_to_clean], function(x) as.numeric(gsub(",", "", x)))
oil_data <- na.omit(oil_data)

cat("Table 4.1: Descriptive Statistics\\n")
print(summary(oil_data))

# ---------------------------
# 2. ADF Test
# ---------------------------
cat("Table 4.2: ADF Test Results\\n")
adf_log <- adf.test(log(oil_data$Price_Y), alternative = "stationary")
adf_diff <- adf.test(diff(log(oil_data$Price_Y)), alternative = "stationary")
print(adf_log)
print(adf_diff)

# ---------------------------
# 3. Time Series and Correlation Matrix
# ---------------------------
oil_price <- ts(log(oil_data$Price_Y), frequency = 4)
opec_production <- ts(oil_data$Opec_Production_x1, frequency = 4)
world_demand <- ts(oil_data$Total_Word_Demand_x2, frequency = 4)
world_supply <- ts(oil_data$Total_Word_Supply_x3, frequency = 4)
transport_cost <- ts(oil_data$OPEC_Oil_Transportation_Costs, frequency = 4)
opec_gdp <- ts(oil_data$OPEC_GDP, frequency = 4)
non_opec_gdp <- ts(oil_data$Non_OPEC_GDP, frequency = 4)
population <- ts(oil_data$Population_billions, frequency = 4)

cor_matrix <- cor(data.frame(opec_production, world_demand, world_supply, transport_cost,
                             opec_gdp, non_opec_gdp, population), use = "complete.obs")
cat("Figure 4.1: Correlation Matrix\\n")
corrplot(cor_matrix, method = "color", type = "upper", tl.cex = 0.8)

# ---------------------------
# 4. VIF Before Wavelet Filtering
# ---------------------------
raw_df <- data.frame(opec_production, world_demand, world_supply, transport_cost,
                     opec_gdp, non_opec_gdp, population)
cat("Table 4.3: VIF Values Before Wavelet Filtering\\n")
print(vif(lm(oil_price ~ ., data = raw_df)))
```

```
# ------------------------------
# 4.4.2 After Wavelet Filtering
# ------------------------------
wavelet_vars <- list(
  opec_production = dwt(opec_production, filter = "haar")@W$W1,
  world_demand = dwt(world_demand, filter = "haar")@W$W1,
  world_supply = dwt(world_supply, filter = "haar")@W$W1,
  transport_cost = dwt(transport_cost, filter = "haar")@W$W1,
  opec_gdp = dwt(opec_gdp, filter = "haar")@W$W1,
  non_opec_gdp = dwt(non_opec_gdp, filter = "haar")@W$W1,
  population = dwt(population, filter = "haar")@W$W1
)
min_len <- min(sapply(wavelet_vars, length))
wavelet_df <- as.data.frame(lapply(wavelet_vars, function(x) head(x, min_len)))
oil_price_trunc <- head(oil_price, min_len)

remove_aliased <- function(df, y) {
  fit <- lm(y ~ ., data = df)
  valid_vars <- names(coef(fit))[!is.na(coef(fit))]
  df[, valid_vars[valid_vars != "(Intercept)"), drop = FALSE]
}
wavelet_df_clean <- remove_aliased(wavelet_df, oil_price_trunc)
wavelet_df_scaled <- scale(wavelet_df_clean)

cat("Table 4.4: VIF Values After Wavelet Filtering\\n")
print(vif(lm(oil_price_trunc ~ ., data = as.data.frame(wavelet_df_scaled))))

# ------------------------------
# 5. Fit ARIMA(0,1,1) and ARIMA(2,1,1)
# ------------------------------
model_011 <- Arima(oil_price_trunc, order = c(0,1,1), xreg = wavelet_df_scaled)
model_211 <- Arima(oil_price_trunc, order = c(2,1,1), xreg = wavelet_df_scaled)

cat("Table 4.5: ARIMAX Model Summary\\n")
print(summary(model_011))
print(summary(model_211))

# ------------------------------
# 6. Monte Carlo Forecasts
# ------------------------------
future_exog <- matrix(rep(tail(wavelet_df_scaled, 1), 8), ncol = ncol(wavelet_df_scaled), byrow = TRUE)
colnames(future_exog) <- colnames(wavelet_df_scaled)

set.seed(123)
sim_log_011 <- replicate(1000, simulate(model_011, nsim = 8, xreg = future_exog))
sim_log_211 <- replicate(1000, simulate(model_211, nsim = 8, xreg = future_exog))

forecast_011 <- exp(apply(sim_log_011, 1, mean))
forecast_211 <- exp(apply(sim_log_211, 1, mean))

last_actual <- tail(oil_data$Price_Y, 1)
```

```
forecast_011_adj <- forecast_011 * (last_actual / forecast_011[1])
forecast_211_adj <- forecast_211 * (last_actual / forecast_211[1])

forecast_quarters <- paste0(rep(2025:2026, each = 4), " Q", rep(1:4, 2))
forecast_table <- data.frame(
  Quarter = forecast_quarters,
  ARIMA_011 = round(forecast_011_adj, 2),
  ARIMA_211 = round(forecast_211_adj, 2)
)

cat("Table 4.6: Forecasted Oil Prices\\n")
print(forecast_table)

# ----------------------------
# 7. Forecast Plot – Figure 4.3: Monte Carlo Simulation Results
# ----------------------------
plot(forecast_011_adj, type = "l", col = "blue", lwd = 2,
     ylim = range(c(forecast_011_adj, forecast_211_adj)),
     main = "Figure 4.3: Monte Carlo Simulation Results",
     ylab = "Forecasted Oil Price", xlab = "Future Quarters",
     xaxt = "n")
axis(1, at = 1:8, labels = forecast_quarters, las = 2, cex.axis = 0.8)
lines(forecast_211_adj, col = "red", lwd = 2, lty = 2)
legend("topleft", legend = c("ARIMA(0,1,1)", "ARIMA(2,1,1)"),
       col = c("blue", "red"), lwd = 2, lty = 1:2)
```