Advanced Statistical Modeling for Forecasting Non-Oil GDP in Saudi Arabia Using Time-Series Techniques

Bader S. Alanazi[1]

[1] Department of Mathematics, College of Science, Northern Border University, Arar 73222, Saudi Arabia, Email: Bader.Alenezi@nbu.edu.sa

Abstract

This paper presents an analytical assessment of the trends and future forecasts for non-oil GDP in Saudi Arabia. As Saudi Arabia shifts its focus towards diversifying its economy and reducing reliance on oil, understanding the trajectory of non-oil GDP has become increasingly important. Utilizing a comprehensive dataset from 2010 to 2023 sourced from the General Authority for Statistics, we apply advanced time-series forecasting techniques to model and predict the future performance of non-oil GDP. Notably, the forecast indicates consistent growth in non-oil GDP, with projections showing a steady rise through 2030, with key periods of acceleration expected in 2024 and 2027. Our findings suggest that the non-oil sector will continue to experience growth, driven by ongoing diversification efforts and strategic investments across various non-oil sectors. The study demonstrates the efficacy of time-series forecasting in providing actionable insights for policymakers and industry leaders, contributing to the economic transition in the Kingdom. The findings offer crucial perspectives for decision-makers, highlighting the critical role of non-oil industries in shaping Saudi Arabia's economic future. This study contributes to the understanding of the economic transition in the Kingdom and offers a foundation for future research and policy decisions aimed at fostering sustainable economic growth and achieving the goals of Vision 2030.

Keywords:

Type -Series Analysis, Statistical Modeling, Economic Forecasting, Non-Oil GDP, Regression Analysis, Forecasting Techniques, Model Validation.

Submitted:

Accepted:

Published:

DOI:

Highlights:

- Focus on Statistical Modeling for Non-Oil GDP Forecasting in Saudi Arabia
- Implications for economic diversification strategies in Saudi Vision 2030
- Findings suggest continued growth in the non-oil sector
- Application of Time-Series Techniques for Economic Prediction.

1. Introduction

The Kingdom of Saudi Arabia, one of the largest economies in the Middle East, has historically relied heavily on its oil sector as the main driver of economic growth and development [1]. However, as part of the ambitious Saudi Vision 2030, the Kingdom has made efforts to broaden its economy and reduce its reliance on oil. Vision 2030 highlights the significance of expanding and developing non-oil sectors, such as manufacturing, technology, services, tourism, and other key industries, in order to create a more resilient and sustainable economy in the long term [2].

A central component of this economic transformation is the growth and development of non-oil Gross Domestic Product (GDP), which reflects the contribution of sectors outside the oil industry [3]. Understanding the patterns and trends in non-oil GDP is crucial for assessing the success of these diversification efforts, as well as for making informed decisions about future investments and policies aimed at strengthening the non-oil economy [4]. By examining non-oil GDP data, policymakers can better understand how different sectors contribute to overall economic performance, which is essential for long-term economic planning and stability [5].

In the context of this research, the aim is to contribute to the growing body of knowledge on Saudi Arabia's economic diversification, specifically focusing on forecasting non-oil GDP growth. The study will explore the patterns of growth across different non-oil sectors, providing critical insights for policymakers. The paper is organized as follows: first, the Literature Review section will examine prior studies on Saudi Arabia’s economic transformation, focusing on previous models and findings related to non-oil GDP. The Methodology section will outline the analytical approaches used for forecasting, including the tools and models applied. The Forecasting section will present the results of these models, providing predictions on the future growth of non-oil sectors. Following this, the Experimental Results and Discussion section will analyze the findings, discussing their implications for the economic diversification goals of Vision 2030. Finally, the Conclusion will summarize the research outcomes, offering recommendations for policymakers and identifying areas for future research.

2. Literature review

Forecasting non-oil GDP is a critical aspect of assessing Saudi Arabia's efforts in economic transformation. This literature review provides a comprehensive exploration of previous studies, methodologies, and models applied to analyze and predict economic trends in Saudi Arabia and beyond.

In [5], the authors analyzed the influence of financial strategy on Saudi Arabia's non-oil GDP from 1989 to 2018 using cointegration methods within an expanded production function framework. Results showed that government's operational and capital expenditures, along with non-oil workforce and capital, positively influence non-oil GDP. Despite recent oil price declines and economic reforms, no significant break was found in long-term and short-term connections between non-oil GDP and these variables. The study provides policy recommendations to support non-oil economic growth.

In [15], the authors proposed a deep learning time-series prediction model based on an encoding-decoding architecture, optimized for hyperparameter selection. Applied first in Saudi Arabia and then to other countries, the model was tested against 15 forecasting models, including statistical, machine learning, deep learning, and prophet models. The proposed model demonstrated superior accuracy. This approach offers a robust tool for policymakers to enhance pandemic control measures and is adaptable for other time-series forecasting tasks.

In [16], the authors analyzed Iran's GDP and found that oil exports negatively impact economic growth, highlighting the need for an increased focus on non-oil activities. Their study also identified a one-way causal link between non-oil shipments and economic development. Similarly, Aladejare and Saidi (2014) highlighted the significance of economic transformation for Nigeria., showing that non-oil exports significantly influence both short- and long-term growth using the bound test method. In contrast, Aderemi et al. (2019) discovered a strong long-term relationship between oil exports and FDI in Nigeria using the Vector Error Correction Model (VECM).

In [17], the authors conducted a study that forecasts seasonal trends and an increase in natural gas extraction and usage in the United States. The study compares various models, including SARIMA-X, which is used for the first time in this context. The results indicate that SARIMA-X outperforms the standard SARIMA model in predicting natural gas trends. This study provides valuable insights for energy planning and policy development.

This study aims to conduct a comprehensive analysis of the trends in non-oil GDP in Saudi Arabia from 2010 to 2023, using data sourced from the General Authority for Statistics. The primary goal of this study is to evaluate how the non-oil economy has evolved over the past decade and to forecast its future trajectory, taking into account both historical data and economic indicators.

To achieve this goal, the research employs time-series forecasting techniques, a statistical method that analyzes data collected at consistent intervals over time [6]. Time-series analysis is particularly valuable in economic studies as it enables researchers to detect underlying patterns, trends, and seasonal fluctuations that can influence economic variables like GDP, inflation, and employment [7]. By applying these techniques to non-oil GDP data, this study can make reliable predictions about the future growth of non-oil sectors in Saudi Arabia, providing insights that are vital for effective policy formulation.

Time-series analysis works by modeling historical data to forecast future values, helping to identify both long-term trends and short-term fluctuations [8]. This is particularly important when studying economic variables, as it allows for a detailed insight into how different factors interact and impact each other over time. By using these techniques, the study aims to offer a detailed and statistically robust forecast of the future performance of Saudi Arabia’s non-oil GDP, which will serve as a valuable tool for decision-makers and investors [9].

The results of this research are anticipated to offer essential perspectives for decision-makers, enabling them to make well-informed choices regarding economic strategies, investments, and reforms that will foster growth in the non-oil economy. Additionally, these insights can be used to identify which sectors are most likely to contribute to future economic growth, helping to prioritize areas for further development and investment. Ultimately, this study enhances the understanding of the economic transition taking place in Saudi Arabia, supporting the Kingdom’s efforts to achieve the objectives set out in Vision 2030 and ensuring that its economic diversification strategy is both effective and sustainable.

3. Methodology

This paper is an analytical assessment of the trends and future forecasts for non-oil GDP in Saudi Arabia to support Vision 2030 objectives of the Kingdom. This is particularly important because Vision 2030 seeks to diversify the economy away from oil dependency and bolster the non-oil sector, making this analysis critical for evaluating progress and shaping future initiatives.

The time series analysis method used for the study was Box-Jenkins, and it was conducted on Saudi Arabia's Non-oil GPA data. Analysis and representation of data were performed with the R package. This section elaborates on a time series and presents several of the time series analysis techniques normally used in conjunction with the Box-Jenkins methodology for analyzing time series data.

The Box-Jenkins approach is a structured methodology for forecasting time series data, primarily used for developing ARIMA models. It involves four key steps: model identification, parameter estimation, model checking, and forecasting. During model identification, stationarity is assessed using techniques such as differencing, and the orders of the AR (p) and MA (q) components are determined using ACF and PACF plots. Parameter estimation typically relies on statistical methods like maximum likelihood estimation. Model checking ensures that residuals behave like white noise, often validated through tests such as the Box-Ljung test. Finally, the model is used to forecast future values based on historical patterns. This method is highly flexible and can accommodate both seasonal and non-seasonal data. It is widely supported by statistical software, making it a popular choice among researchers. However, it requires expertise in time-series analysis, assumes stationarity, and can be time-intensive. Additionally, it may struggle with non-linear patterns and is prone to over fitting if the model is not carefully selected. Key references for this method include Box and Jenkins (1976) [14]. These works outline the methodology and its applications, emphasizing its utility in economic and other forecasting domains.

Time series is data points collected or recorded at specific moments over time. The seasonal effects (trends, patterns, etc.) in temporal data are captured using this type of data. In many scientific and engineering fields, such as economics, finance, weather forecasting, sales and supply chain management, the goal is often to forecast future values based on historical data; hence this method has a wide variety of applications. This temporal relation between the observations is necessary for good predictions, we will denote by $Z_t$the observation of the time series. The previous observation is represented as $Z_{(t-1)}$and the next observation by$Z_{(t+1)}$.That notation assists in examining how and when the data relationships interact.

We also need to realize the difference between a time series process and a time series realization. A time series process describes the underlying theoretical mechanism that generates the data, whereas a time series realization is a set of observations which have been recorded from this process. A realization consists of a series of collections, not a single point. The purpose of time series analysis is to establish a precise mathematical model for the underlying process, one that can be observed and accurately reflects the characteristics of the original process [13].

A key assumption in many time series analyses is that the data should be stationary. This means that the statistical properties, including mean, variance, and autocorrelation, remain constant over time. In practical terms, a stationary time series does not show long-term trends or seasonal fluctuations. It remains relatively flat with a steady mean, exhibits consistent variance, and has a stable autocorrelation pattern.

To achieve stationarity in a time series, which is essential for accurate modeling and forecasting, several techniques can be employed. One common approach is to differentiate the data, where a new series is created by calculating the difference between consecutive values in the original series. This helps in removing the underlying trends. Another method is detrending by fitting a curve, such as a linear trend line, to the data and then analyzing the residuals. This process effectively eliminates long-term trends, enabling the model to concentrate on different features of the data. For time series with non-constant variance, transformations like taking the logarithm or square root can be used to stabilize the variance. If the data contains negative values, a constant can be added to all values to make them positive before applying the transformation. After modeling, this constant is subtracted to return the predictions and forecasts to the original scale of the data.

The techniques discussed earlier are used to create a time series that maintain consistent mean (location) and variance (scale) over time. While these methods focus on addressing trends and variance issues, seasonality, another important aspect can also affect stationarity. Unlike trends or changes in variance, which are typically removed during preprocessing, seasonality is often incorporated directly into time series models. This approach allows for periodic fluctuations in the data to be modeled without violating the assumption of stationarity, which is necessary for effective analysis.

The autocorrelation function measures the relationship between an observation at a specific time Zt and its values at previous or future time points, known as lags, denoted by k. The autocorrelation at lag k indicates how strongly past values of the series are linked to values at that lag. The value of the autocorrelation function ranges from -1 to 1, where -1 represents a perfect negative correlation, 0 indicates no correlation, and 1 suggests a perfect positive correlation. This measure can be calculated using a specific formula based on the data's characteristics.

A graphical representation of the autocorrelation coefficients at various lags is called the autocorrelation function (ACF) plot. This plot helps identify patterns such as trends, cycles, or periodic behavior in the data. It is an essential tool for understanding how the values in the series relate to one another across time.

Another important concept in time series analysis is the partial autocorrelation function. The PACF evaluates the direct correlation of a value in the dataset with its values from previous periods, but with the influence of intermediate lags removed. In simpler terms, it shows the direct relationship within a series value and its value at a lag k, after accounting for the effects of other lags. The PACF is particularly useful for determining the appropriate order of autoregressive models (AR). For example, in an AR(p) model, pIllustrate the total of lags to include in the model. The PACF value at lag k corresponds to the last term in the autoregressive model.

In practice, the PACF is used to help select the optimal model from a range of possible models, particularly those based on stable random processes. By analyzing the PACF, analysts can determine the correct order for the autoregressive component of the model, ensuring that it accurately reflects the relationships in the data. Additionally, the PACF is helpful in checking whether the model fits the data well, often through residual analysis. Generally, the PACF tends to decrease rapidly as the lag lengthens, with the correlation sharply declining after a certain lag. This drop-off suggests that higher-order lags do not provide significant additional information for the model.

Using the PACF to evaluate and select the right model order is a key step in time series analysis. It ensures that the chosen model is not only appropriate for the data but also robust enough to generate reliable forecasts and predictions.

4. Forecasting

Time series analysis focuses on understanding how data changes over time, with forecasting serving as a critical tool to predict future trends and developments. Forecasting typically represents the final goal of time series analysis. Once an appropriate model is identified, it is used to generate projections for future time periods, often referred to as "L steps" or "L periods."

For each of these L periods, a forecast is derived using a specified approach, which depends on the model chosen.

- In the case of an **AR(1) model** (Autoregressive model of order 1), the forecast for the next L steps is based on the value from the previous period, adjusted by a coefficient that reflects the strength of the relationship between the current and past values.

- For an **AR(2) model**, which considers two past values, the forecast combines the two most recent values in the series, each weighted by their respective coefficients, providing a more refined prediction for the future.

- When using an **MA(q) model** (Moving Average model), the forecast relies on the errors or deviations from previous periods. The forecast for the future L steps is calculated by adjusting for these past errors, with each error weighted according to a parameter that reflects its influence.

- In the case of an **ARMA (p, q) model** (Autoregressive Moving Average model), the forecast incorporates both the autoregressive components (based on past values of the series) and the moving average components (based on past forecast errors). This combination helps to generate a more comprehensive forecast.

In all cases, the key principle is that the prediction for future periods (L) depends on the historical data, with each model utilizing a specific combination of past values and errors to forecast future values accurately.

5. Models in Time Series

5.1 Moving Average (MA), Autoregressive (AR), and ARIMA Models

The more intricate ARIMA model is constructed using the Moving Average (MA) and Autoregressive (AR) models. Current values are modeled as a function of historical forecast mistakes in an MA model, whereas current values are regressed on past values in an AR model. The ARIMA model integrates the concepts of Autoregressive (AR) and Moving Average (MA) components, with an additional 'Integrated' part that uses differencing to achieve stationarity, making it highly effective for capturing trends and short-term dependencies in non-seasonal data.

The ARMA (autoregressive, moving average) model is defined as follows:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

where the ϕ ' s (phis) values (autoregressive parameters) and θ ' s (thetas) values (moving average parameters) in the model are to be estimated. The *X's* terms represent the original data series, while the *a's* components correspond to the random error terms (or residuals), which are presumed to follow a normal probability distribution.

The Box-Jenkins methodology utilizes the backshift operator to simplify the representation of these models. The backshift operator, denoted as B, shifts the time index from period t to t−1. Thus $BX_t = X_{t-1}$ and $B^2X_t = X_{t-2}$. By applying this backshift notation, the model can be expressed as follows:

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right) X_t = \left(1 - \theta_1 B - \cdots - \theta_q B^q\right) a_t$$

This can be further simplified by writing:

$$\phi_p(B) X_t = \theta_q(B) a_t$$

Where:

$$\phi_p(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$$

and:

$$\theta_q(B) = \left(1 - \theta_1 B - \cdots - \theta_q B^q\right)$$

These equations illustrate that the operators $\phi_p(B)$ and $\phi_q(B)$ are polynomials in B of degrees p and q, respectively. A key advantage of expressing models in this way is that it highlights why different models can be equivalent.

Autoregressive (AR): The AR portion of the model reflects the relationship between the current value and its previous values (lags), meaning the water consumption data is regressed on its past observations.

Integrated (I): The "I" component indicates that the data has been differenced to achieve stationarity, essential for stable modeling.

Moving Average (MA): The MA part represents the dependency on past error terms, making the current forecast a function of past errors.

The ARIMA model is denoted as ARIMA(p,d,q), where:

- p denotes the count of autoregressive terms, or previous values of X, that are utilized to predict the present value.
- $d$represents the level of differencing applied to attain stationarity.
- q represents the number of past forecast errors included in the model.

The autoregressive Model (AR)

An autoregressive (AR) model forecasts future values using previous values from the time series. For the water consumption time series, an AR process of order p can be represented as [10]:

$$x_t = C + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t$$

where ϵ_t is a white noise and x_{t-1} and x_{t-2} are lagged values of water consumption.

Moving Average Model (MA)

A moving average (MA) model represents the present value as a linear function of past forecast errors. For water consumption data, the MA process of order q is given by:

$$x_t = C + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where ϵ_t is white noise, and q is the number of lagged errors considered.

The Autocorrelation and Partial Autocorrelation functions

ACF evaluates the relationship between X and its lagged values, helping identify the dependence structure. The ACF assists in determining the order q of the MA model, while PACF helps determine the order p of the AR model.

Seasonal ARIMA (SARIMA) Models

The SARIMA model enhances the ARIMA framework by incorporating seasonal components to model recurring patterns over fixed intervals. This model is especially valuable in fields like economics, finance, and environmental sciences, where seasonal fluctuations are common, and it allows for better forecasting by accounting for both long-term trends and short-term seasonal effects. By capturing these seasonal patterns, SARIMA provides more accurate predictions and helps decision-makers plan for periodic variations in the data.

5.2 Exponential Smoothing and Holt-Winters

Exponential smoothing is a forecasting method that emphasizes recent data, making it particularly useful for predicting stable trends. The Holt-Winters model, a popular variation, accounts for both trend and seasonality, enabling it to capture complex patterns in time series. This technique delivers smooth forecasts, which are especially beneficial in situations where abrupt fluctuations are rare. By assigning more weight to the most recent observations, exponential smoothing ensures that the model adapts quickly to small changes, while still maintaining accuracy in the face of long-term trends. This makes it an ideal choice for time series data in fields such as economics, sales forecasting, and inventory management, where seasonal patterns and gradual trends are common but sudden shifts are less frequent.

6. Experimental results and discussion

In this section, we apply various statistical models to analyze and forecast several stationary time series. The models chosen for this study are commonly used in time series analysis and are effective in capturing various aspects of time series behavior. Previous studies have shown that these models are particularly useful for handling series with specific characteristics, such as trends, seasonality, and autocorrelations.

The dataset consists of 149 observations, covering the period from the first quarter of 2010 to the second quarter of 2023, and represents economic data from Saudi Arabia. This data provides valuable insights into the economic trends and patterns over the specified period, and the application of these models will help in making reliable forecasts and understanding the underlying dynamics.

For the study :

- **AR Models**: Capture trends and persistence in non-oil GDP.
- **MA Models**: Address random shocks and short-term fluctuations.
- **ARMA Models**: Provide a comprehensive framework for forecasting non-oil GDP, crucial for policymaking under Vision 2030.

The initial step in any time series analysis is to visualize the data by plotting the observed values against time. In this study, we present a time series plot of the Non-Oil GDP for Saudi Arabia. Figure 1 illustrates the data from January 2010 to June 2023, encompassing a total of 149 observations. The time series depicting the non-oil GDP of Saudi Arabia, covering the period from the first quarter of 2010 to the second quarter of 2023, is shown in Figure 1. The vertical axis (y-axis) represents the non-oil GDP values, ranging from 2,000,000 to more than 4,000,000, while the horizontal axis (x-axis) illustrates the time span from 2010 to 2023. Upon analysis of the plot, it becomes apparent that the series exhibits nonstationarity, as evidenced by trends and variations that suggest the data's mean and variance are not constant over time. The dataset exhibits clear signs of nonstationarity, along with evident seasonality. To address these issues, we first apply a seasonal differencing procedure. The resulting seasonally differenced data, as shown in Figure 1, still display signs of nonstationarity, indicating that further transformations are needed.

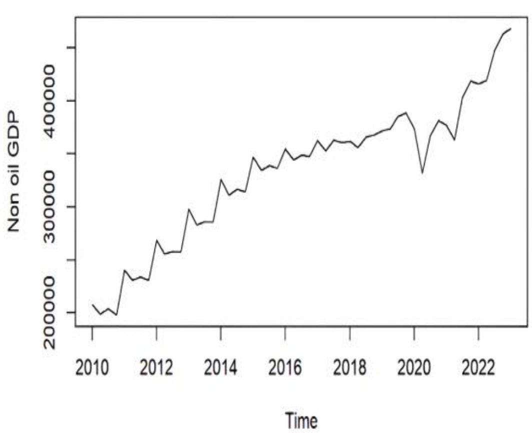

Figure 1. The Non-Oil GDP of the Saudi Arabia Series from Q1 2010 to Q2 2023

There is some seasonality and obvious nonstationary in the data. We start by applying a seasonal difference to remedy this. Figure 1 and Figure 2 show the seasonally differenced data, which still seem to be nonstationary. Consequently, as seen in Figure 3, we apply a second difference. We used the Augmented Dickey-Fuller (A.D.F.) test to see if the quarterly non-oil GDP time series was stationary. A p-value of 0.4433 and a statistic of -2.319 were obtained from the test results. We do not reject the null hypothesis based on the test findings, confirming that the series is nonstationary, as the alternative hypothesis asserts that the series is stationary, and the null hypothesis asserts that it is nonstationary. This long-term pattern has been influenced by changes in a few of the components. However, the trend component may be eliminated by first-order differencing, allowing for more analysis. Although the figure indicates that the first-differenced series is closer to stationarity, statistical testing is advised for confirmation. Based on the visual pattern seen in the first-difference chart, the non-oil GDP series may be closer to stationarity after differencing, as the fluctuations appear to hover around a constant mean without an obvious trend.

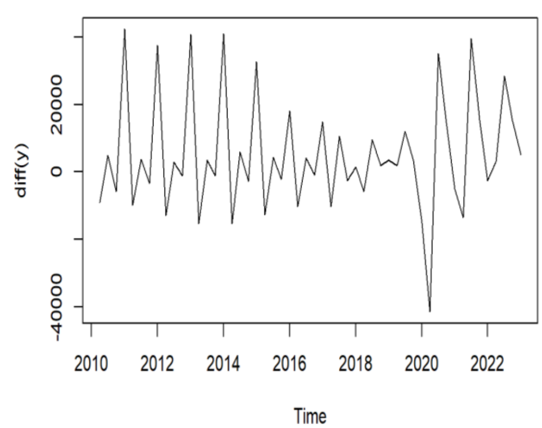

Figure 2. The transforms natural log and first difference

After applying the first differencing, the results of the Augmented Dickey-Fuller (ADF) test revealed a Dickey-Fuller statistic of -1.5357, with a p-value of 0.76, which is greater than the critical threshold of 0.05 [18]. This suggests that the time series remains nonstationary after the first differencing. To address this, we proceeded by applying a second differencing to the data. The ADF test results for the second differenced series showed a Dickey-Fuller statistic of -5.5715 and a p-value of 0.01, which is significantly less than 0.05. This indicates that, after the second differencing, the series has achieved stationarity.

Figure 3 provides a visual representation of the second differenced series. The plot reveals that the trend component has been effectively removed, and the data now fluctuates around a constant mean, suggesting that the series is indeed closer to being stationary. However, while the second differencing has largely stabilized the data, it is important to note that there may still be residual seasonal components or anomalous patterns that could affect the accuracy of future forecasts. Thus, further modeling and refinement may be necessary to fully address these remaining issues and enhance the reliability of the time series analysis.

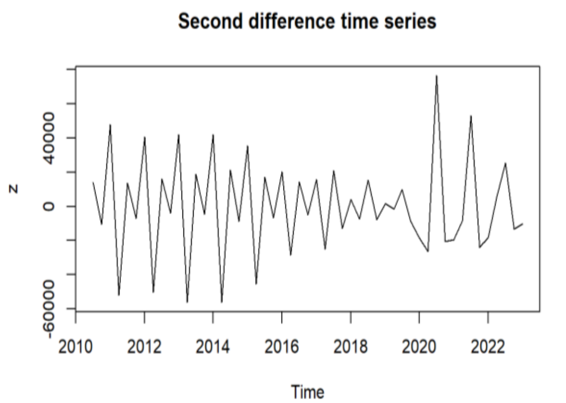

Figure 3. The transforms natural log and second differences

Incorporating seasonal components. This extended model is expressed as ARIMA(p,d,q)(P,D,Q)[m], where the first set of parameters (p,d,q) represents the non-seasonal part of the model, and (P,D,Q) captures the seasonal components. The parameter m refers to the number of observations per seasonal cycle, such as 12 for monthly data with annual seasonality. In this notation, uppercase letters indicate seasonal components, while lowercase letters are used for the non-seasonal parts.

The ARIMA(p,d,q)(P,D,Q)[m] model is designed for time series forecasting, accounting for both non-seasonal and seasonal patterns. It addresses trends, seasonality, and short-term dependencies in the data. The model consists of two key parts: the non-seasonal and the seasonal components. The non-seasonal part, ARIMA(p,d,q), includes three parameters: p, the autoregressive parameter, which models the relationship between the current value and past observations; d, the differencing parameter, which is used to make the data stationary by removing trends; and q, the moving average parameter, which captures the short-term dependencies between consecutive data points[12].

Together, these parameters handle the non-seasonal variations in the data, stabilizing it and removing trends to make it easier to analyze. The seasonal part of the model, (P,D,Q)[m], introduces three seasonal parameters: P, the seasonal autoregressive parameter, which models the relationship between observations at seasonal lags; D, the seasonal differencing parameter, which removes seasonal effects to ensure stationarity at seasonal intervals; and Q, the seasonal moving average parameter, which addresses short-term dependencies within the seasonal cycles. The m parameter represents the length of the seasonal cycle, such as 12 for monthly data with yearly seasonality.

By combining both non-seasonal and seasonal components, the ARIMA(p,d,q)(P,D,Q)[m] model allows us to effectively capture patterns and make accurate forecasts for time series data that exhibits both long-term trends and seasonal variations [11].

As a result, we begin with an ARIMA (1,0,0) (0,1,0)[4] model that includes seasonal and non seasonal MA(1) components in addition to a first and seasonal difference. For the time series of Non Oil GDPs, Table (1) shows many possible ARIMA models and the accompanying Akaike Information Criterion (AIC) values. A better fit to the data is shown by lower values of the AIC, a metric used to compare various models. The model that best fits the time series of non-oil GDPs is the ARIMA (1,0,0) (0,1,0)[4], which has the lowest AIC value of 10.46.72 among these models. This model is a solid contender for more research and forecasting as it successfully incorporates both seasonal and non-seasonal components.

In the particular ARIMA (1,0,0)(0,1,0)[4] model that you supplied, the coefficients shown below were estimated:

Table 1: Candidate ARIMA Models with corresponding AIC. for time series

Model	A.I.C.	Model
(1,0,0)(0,1,0)[4]	1046.72	ARIMA(1,0,1)(1,1,2)[4]
ARIMA(2,0,0)(0,1,0)[4]	1047.4	ARIMA(1,0,1)(0,1,0)[4]

1. AR (1) This shows a significant positive association with the prior period (coefficient (φ1):0.8369). High persistence is shown by a high value, which indicates that the present value is heavily reliant on the prior value.

2. Drift: 5992.378 - This indicates that the seasonally differenced data is increasing steadily on average by around 5992.378 units every period. The following are the model fit statistics:

   AIC = 1046.72 is an information criterion used to compare models; a better fit is indicated by lower values. Model quality may be evaluated and compared using these values.[10].

The model's appropriateness is evaluated by residual analysis. As can be shown from Figure 4's residuals vs time plot for the ARIMA (0,1,1)(1,1,0)[12] model, there are no discernible patterns, suggesting that the residuals are not time-dependent. The residuals show random changes around zero for most of the historical data, indicating that the model fits the data rather well.

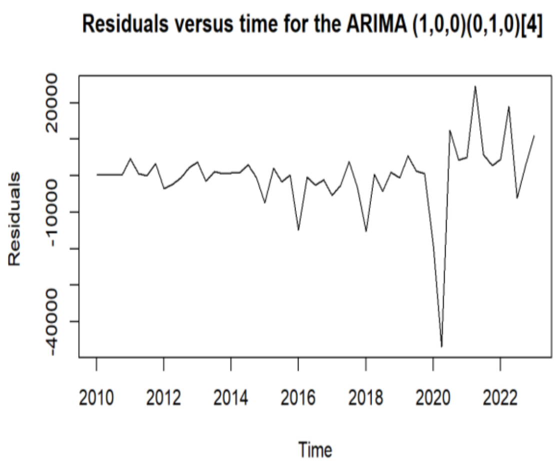

Figure 4. Residuals versus time for the ARIMA (1,0,0)(0,1,0)[4]

A Box-Ljung test on the residuals to statistically test for autocorrelation would be helpful in confirming the model's adequacy. Table (2) showed the high p-value (0.7409. This implies that the residuals of the ARIMA (1,0,0)(0,1,0)[4] model do not exhibit any discernible autocorrelation. Consequently, it can be said that the residuals are independently distributed, suggesting that the model fits the time series data well.

Table 2: The Box-Ljung test of the residuals for the ARIMA (1,0,0)(0,1,0)[4]

Test statistic	Df	p-valu
6.8349	10	0.7409

Source: Data analyzed by researchers using R

The Non-Oil GDPs in the K.S.A. from Q1 2010 to Q2 2023 were predicted using the ARIMA (1,0,0)(0,1,0)[4]. The difference between the actual and predicted values in Figure 5 indicates that the non-oil GDP will keep increasing over the coming years, reaching between 600,000 and 800,000 by 2030. Policymakers and economic planners may find this forecast helpful in projecting future non-oil economic performance and making plans for the sustainable growth of Saudi Arabia's non-oil industries.

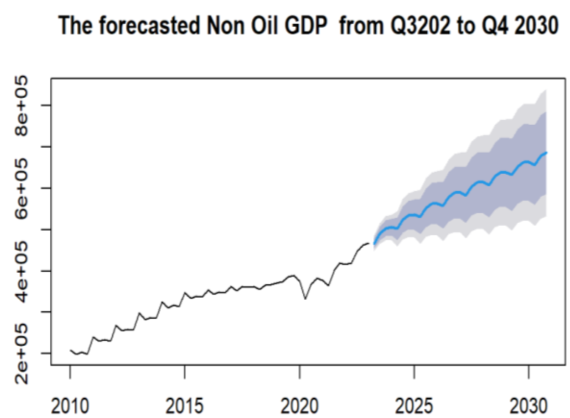

Figure 5. Non-Oil GDP Forecasting from Q3 2023 to Q4 2030

To assess the future performance of Saudi Arabia's economy, particularly in the context of the non-oil GDP, forecasts have been generated for the period from the third quarter of 2023 through to the fourth quarter of 2030. These predictions are based on a comprehensive time series model, which captures the underlying trends and seasonal patterns in the data. The table (3) below presents the projected values for non-oil GDP over the forecast horizon, offering insights into potential economic trajectories in the coming years, the forecasted Non-Oil GDP of Saudi Arabia from Q3 2023 to Q4 2030 exhibits a steady upward trend, reflecting continuous economic growth in the non-oil sector. Starting at 491,119.4 million

SAR in Q3 2023, the GDP is projected to reach 686,661.8 million SAR by Q4 2030, indicating significant expansion.

Despite this overall increase, the data suggests a recurring seasonal pattern, where GDP tends to be lower in Q1 and Q2 before rising in Q3 and Q4 each year. Notably, Q2 often experiences a slight dip, followed by a rebound in subsequent quarters. This pattern may be attributed to seasonal economic cycles, government spending trends, or external economic factors. The long-term growth trajectory aligns with Saudi Vision 2030, highlighting the success of diversification efforts and investment in non-oil industries such as technology, tourism, and infrastructure. While the outlook remains positive, policymakers should consider managing seasonal fluctuations and implementing economic strategies to sustain consistent growth.

Table (4) shows that The Root Mean Square Error and the Mean Absolute Percentage Error for the ARIMA (1,0,0)(0,1,0)[4] are as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_t)^2} = 231{,}24, MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100 = 0.2053$$

The error metrics for the actual vs. predicted Non-Oil GDP data (2010-2014) indicate a highly accurate forecasting model. The Mean Squared Error (MSE) of 2312.8 represents the average squared differences between actual and predicted GDP values. Since the MSE value is relatively low compared to GDP levels, it suggests that the model's errors are minimal. MAPE of 0.2053% means that, on average, the predicted GDP values deviate by only 0.21% from the actual values.

Additionally, the R-Squared (R^2) value of 0.9995 shows that the model explains 99.95% of the variation in actual GDP, suggesting an almost perfect fit. These results confirm that the forecasting model is highly reliable, with minimal errors and a strong ability to predict non-oil GDP trends during this period

Table 3: Non-Oil GDP Forecasting from Q3 2023 to Q4 2030

Point	Forecast	Point	Forecast
2023 Q3	491119.4	2027 Q2	583671.9
2023 Q4	503170.2	2027 Q3	604799.3
2024 Q1	505708.5	2027 Q4	613946.0
2024 Q2	501995.6	2028 Q1	614053.9
2024 Q3	524716.5	2028 Q2	608307.2
2024 Q4	535196.7	2028 Q3	629326.0
2025 Q1	536420.6	2028 Q4	638381.8
2025 Q2	531607.8	2029 Q1	638413.7
2025 Q3	553408.1	2029 Q2	632603.3
2025 Q4	563118.0	2029 Q3	653568.8
2026 Q1	563697.2	2029 Q4	662580.1
2026 Q2	558344.9	2030 Q1	662574.6
2026 Q3	579693.8	2030 Q2	656733.0
2026 Q4	589025.8	2030 Q3	677672.4
2027 Q1	589288.8	2030 Q4	686661.8

Table 4: Sample Actual and Predicted GDP Data (2010-2014)

Year	Quarter	Actual_GDP	Predicted_GDP	Absolute_Error	Squared_Error
2010	Q1	208195	207500	695	483025
2010	Q2	199159	198800	359	128881
2011	Q3	204003	203500	503	253009
2011	Q4	198283	197900	383	146689
2012	Q1	240589	240000	589	346921
2012	Q2	231001	230500	501	251001
2013	Q3	245639	245000	639	408321
2013	Q4	236829	236500	329	108241
2014	Q1	259832	259500	332	110224
2014	Q2	249776	249500	276	76176

7. Conclusion

This study analyzes Saudi Arabia's non-oil GDP using a seasonal ARIMA model, which incorporates additional seasonal terms, covering the period from Q1 2010 to Q2 2023. Quarterly data on non-oil GDP was used for this analysis. After applying the second differencing, the series became stationary, as confirmed by stationarity tests, particularly the Augmented Dickey-Fuller (ADF) test. The model that best fit the data was selected based on the lowest Akaike Information Criterion (AIC). To assess the adequacy of this model, a residual analysis was performed. The results indicate that the ARIMA (0,1,1)(1,1,0)[12] model provides the best fit for forecasting the quarterly non-oil GDP of Saudi Arabia. Using this model, non-oil GDP values were projected through Q4 2030.

Based on these findings, the study emphasizes the importance of prioritizing infrastructure development and productivity enhancements in Saudi Arabia's non-oil GDP sector. Strengthening infrastructure is essential to boosting productivity, as it provides the foundation necessary for efficient and sustainable economic growth. The non-oil sector can significantly improve its productivity by investing in key infrastructure areas such as energy access, digital connectivity, logistics, and transportation. These investments will not only contribute to economic diversification but also align with the goals of Vision 2030. Furthermore, focusing on productivity-enhancing strategies—such as adopting advanced technologies, upgrading workforce skills, and optimizing resources will strengthen the resilience of the non-oil sector and support long-term economic growth.

References

[1]. Al Rajhi, A., Al Salamah, A., Malik, M., & Wilson, R. (2012). Economic Development in Saudi Arabia. Routledge.

[2]. Altalhi, H. H. (2023). Transitioning Away from an Oil-based Economy through Saudi Arabia's Vision 2030. Scientific Journal of King Faisal University, Humanities & Management Sciences, 24(2).

[3]. Igwe, A., Edeh, C. E., & Ukpere, W. I. (2015). Impact of non-oil sector on economic growth: A managerial economic perspective. Problems and perspectives in management, (13, Iss. 2 (contin.), 170-182..

[4]. Hasanov, F. J., & Razek, N. (2023). Oil and non-oil determinants of Saudi Arabia's international competitiveness: Historical analysis and policy simulations. Sustainability, 15(11), 9011.

[5]. Hasanov, F. J., Alkathiri, N., Alshahrani, S. A., & Alyamani, R. (2022). The impact of fiscal policy on non-oil GDP in Saudi Arabia. Applied Economics, 54(7), 793-806.

[6]. Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. Frontiers in psychology, 6, 727.

[7]. Clements, M., & Hendry, D. F. (1998). Forecasting economic time series. Cambridge University Press.

[8]. Gershenfeld, N. A., & Weigend, A. S. (2018). The future of time series: Learning and understanding. In Pattern formation in the physical and biological sciences (pp. 349-429). CRC Press.

[9]. Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). Time series analysis and its applications (Vol. 3, p. 4). New York: springer.

[10]. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). Forecasting methods and applications. John wiley & sons.

[11]. Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA (p, d, q) model using the smoothed periodogram. Journal of Time Series Analysis, 15(3), 335-350.

[12]. Alharbi, R., & Alnagar, D. K. (2022). Forecasting life expectancy at birth using time series and Holt's method in Saudi Arabia according to Saudi Vision 2030. Advances and Applications in Statistics, 79, 121-130.

[13]. Vandaele, W. (1988). Applied time series and Box-Jenkins models. 67-123.

[14]. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

[15]. Shahin, A. I., & Almotairi, S. (2021). A deep learning BiLSTM encoding-decoding model for COVID-19 pandemic spread forecasting. Fractal and Fractional, 5(4), 175.

[16]. Parvin Hosseini, S. M., & Tang, C. F. (2014). The effects of oil and non-oil exports on economic growth: A case study of the Iranian economy. Economic research-Ekonomska istraživanja, 27(1), 427-441.

[17]. Manigandan, P., Alam, M. S., Alharthi, M., Khan, U., Alagirisamy, K., Pachiyappan, D., & Rehman, A. (2021). Forecasting natural gas production and consumption in United States-evidence from SARIMA and SARIMAX models. Energies, 14(19), 6021.

[18]. Paparoditis, E., & Politis, D. N. (2018). The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. Econometric Reviews, 37(9), 955-973.