EGTRIB Journal
JOURNAL OF
THE EGYPTIAN SOCIETY OF TRIBOLOGY
VOLUME 21, No. 4, October 2024, pp. 82 - 92 **ISSN 2090 - 5882**

(Received July 28. 2024, Accepted in final form August 24. 2024) **jest.journals.ekb.eg**

A COMPARATIVE STUDY OF LUBRICANTS DEMAND FORECASTING

Islam M. Hammam, Amin K. El-Kharbotly and Yomna M. Sadek

Design and Production Engineering Department, Ain Shams University, Cairo, EGYPT.

ABSTRACT
Lubricants play a crucial role in various industries such as automotive, manufacturing, and energy, where accurate demand forecasting is essential for maintaining efficient supply chains, reducing costs, and ensuring timely product availability. This study evaluates the performance of traditional time series forecasting models on forecasting of lubricants demand, with a specific focus on demand exhibiting a linear increasing trend as a prevalent pattern in many situations. The models tested include ARMA, ARIMA, SARIMA, and Triple Exponential Smoothing (TES), which are widely used for forecasting in scenarios with linear and seasonal patterns. Two datasets were selected based on their linear trend characteristics, representative of the steady and consistent growth in demand for lubricants across different industries. The datasets were split into training and test sets, with model parameters optimized to minimize the Akaike Information Criterion (AIC).

Performance was measured using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results showed that SARIMA consistently outperformed the other models, with TES, ARIMA, and ARMA following in effectiveness. The study highlights the significance of accurate lubricants demand forecasting in improving supply chain efficiency. Furthermore, the presence of a linear increasing trend in demand data underscores the importance of selecting appropriate models that can effectively capture and project these trends, which are vital for informed decision-making in supply chain management of lubricant industries.

KEYWORDS
ARIMA, Exponential Smoothing, Increasing Linear Pattern, Seasonality, Time Series.

INTRODUCTION
Accurate demand forecasting plays a vital role in ensuring that lubricants businesses can meet customer demand efficiently while minimizing costs associated with overstocking or stockouts, leading to effective supply chain management. Good demand forecasting leads to optimized inventory levels, reduced lead times, improved production planning, and ultimately enhanced customer satisfaction. The ability to predict future lubricants

demand with precision allows companies to stay competitive in today's fast-paced markets, where consumer preferences and purchasing patterns can shift rapidly, [1]. Among the various patterns observed in lubricants demand data, a linear increasing trend is particularly significant. This trend is commonly seen in lubricants experiencing steady growth due to factors like market expansion, seasonal effects, or long-term consumer adoption. Accurately forecasting such trends is crucial for lubricants industries ranging from retail to manufacturing, where consistent demand growth can inform strategic decisions on inventory management, production scaling, and distribution logistics, [2].

Given the importance of capturing linear trends in demand data, several time series forecasting models have been developed to address this need. These models include traditional approaches like Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Triple Exponential Smoothing (TES). Each of these models offers unique strengths in dealing with specific aspects of time series data, such as trend and seasonality, making them valuable tools for demand forecasting in various applications. Although those traditional methods are not the only forecasting methods, machine learning techniques are starting to invade the field well, improving the forecasting accuracy and increasing the ability of dealing with complex demand. However, machine learning techniques also have limitations, [3]. Therefore, the focus of this study will be the traditional approach.

For instance, S. Makridakis et al., [4], aimed to evaluate the forecasting accuracy of various extrapolative time series methods through a structured competition. The objective is to identify which methods perform best under different conditions and to understand the factors influencing forecasting accuracy. The study concluded that users could improve forecasting accuracy by selectively choosing methods based on the type of data and forecasting horizon. It emphasized that there is no universally best method; rather, the effectiveness of a method can varies depending on the context. Everette S. Gardner, Jr., [5], aimed to review and evaluate the methodologies and practices associated with exponential smoothing in forecasting by providing guidelines for the application of exponential smoothing techniques. The paper concludes that the Holt-Winters method is generally favored for seasonal data, while Brown's linear trend model is noted for its theoretical advantages. J.W. Taylor, [6], aimed to explore and evaluate a new forecasting method that incorporates damped multiplicative trends into exponential smoothing techniques. The objective is to assess the effectiveness of this method compared to traditional approaches, particularly the Holt and Holt-Winters methods, which typically use additive trends. The findings indicate that the damped Pegels method performs competitively, often slightly outperforming the damped Holt method, particularly in series with strong trends. L. Elneel, M. S. Zitouni, H. Mukhtar, and H. Al-Ahmad, [7], investigated the forecasting of global mean sea level (GMSL) changes by analyzing the interrelationships among various climatic factors, including CO_2, CH_4, ocean heat, and temperature. The primary objective is to understand how these factors influence GMSL and to evaluate the effectiveness of different time series analysis models in predicting sea level changes. They concluded that while CO_2 and temperature are critical indicators of GMSL changes, the direct influence of GMSL on these variables is not statistically significant. The ARIMA and Prophet models demonstrate good performance in

forecasting, with the Prophet model showing higher confidence intervals for long-term predictions.

This paper focuses on evaluating the performance of these models on datasets characterized by a linear increasing trend. By comparing their effectiveness, we aim to provide insights into which models are best suited for lubricants demand forecasting in scenarios where demand consistently rises over time. The following sections will explore the methodology used in this study and present the results of our analysis, offering guidance on model selection for practitioners dealing with similar demand patterns.

METHODOLOGY

This section states the main steps of obtaining a forecast using various traditional models: Autoregression Moving Average (ARMA), Autoregression Integrated Moving Average (ARIMA), Seasonal Autoregression Integrated Moving Average (SARIMA), and Triple Exponential Smoothing (TES). The steps are targeting the best parameters combination of the candidate models to obtain the best result avoiding any possibility of overfitting.

The Auto Regression Moving Average Model

The ARMA model is a statistical method used to forecast future values of a time series variable based on its past values [8]. ARMA is a combination of two models: Auto-regressive (AR) model and Moving Average (MA) model. The AR model explains how the present value of a time series variable depends on its past values, while the MA model explains how the present value depends on the past errors. The AR function uses a linear combination of past values weighted by coefficients, while the MA function incorporates past forecast errors to refine predictions. The steps of obtaining a forecast using the ARMA model are as follows:

1. Data Preparation

Organizing and formatting the data sets for the time series variable to ensure your data is in a time series format, with equally spaced intervals. Check for missing values, outliers, and seasonality in the data. This step is necessary in all other models so it will be excluded from further sections for brevity.

2. Model Specification

Determine the orders of the ARMA model (p, q) by looking at the autocorrelation function (ACF) for (q) and partial autocorrelation function (PACF) for (p). The number of lagged values of predictors and number of lagged forecast errors that are necessary for AR model and MA model respectively can be obtained by observing (ACF) and (PACF).

3. Model Estimation

Estimate the coefficients of the ARMA model using a method such as maximum likelihood estimation or least squares estimation. The ARMA model is typically expressed as shown in (1):

$$Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

4. Model Selection

Among all possible number of lagged values that will be considered for AR model, the best model is selected based on AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). (AIC) and (BIC) are statistical measures commonly used for model selection and comparison in the field of statistics and machine learning. Both criteria provide a way to evaluate the goodness of fit of different models and help in selecting the most appropriate model among a set of competing models. AIC was developed by Hirotugu Akaike and is based on information theory. It measures the quality of a model by considering both its goodness of fit and its complexity. The AIC is defined as in (2). BIC, also known as Schwarz criterion, is similar to AIC but places a higher penalty on model complexity. BIC is derived from a Bayesian framework and considers the sample size of the data. The BIC is defined as (3). The penalty term in BIC is more severe than in AIC due to the inclusion of the logarithm of the sample size. Similar to AIC, lower BIC values indicate a better model fit, but BIC tends to favor simpler models more strongly than AIC.

$$AIC = -2\log L + 2k \tag{2}$$

$$BIC = -2\, log\, L + k\, log\, n \tag{3}$$

5. Forecasting and Model Evaluation

Use the estimated ARMA model to forecast future values of the time series variable. The forecasted values are obtained by recursively applying the ARMA model to the past observed values of the time series variable. Evaluate the performance of the ARMA model by comparing the forecasted values with the actual values using metrics such as mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) shown in (4), (5), and (6) in order to compare it with the other proposed models. This step is necessary in all other models so it will be excluded from further sections for brevity.

$$MSE = \frac{1}{T} \sum_{t=1}^{t} (x_t - y_t)^2 \tag{4}$$

$$MAE = \frac{1}{T} \sum_{t=1}^{t} |x_t - y_t| \tag{5}$$

$$RMSE = \left(\frac{1}{T} \sum_{t=1}^{t} (x_t - y_t)^2\right)^{1/2} \tag{6}$$

The Auto Regression Integrated Moving Average Model

The ARIMA model is a statistical method used to predict future values based on past observations [8]. It is a combination of three components: autoregression (AR), differencing (I), and moving average (MA). The AR component involves using past values of the time series to predict future values. The MA component involves using past forecast errors to predict future values. The I component involves differencing the time series to make it stationary, which means that its statistical properties do not change over time. The steps of obtaining a forecast using the ARIMA model are as follows:

1. Model Specification

Determine the order of the ARIMA model parameters by observing the (ACF) and (PACF) plots for p and q. Test the stationary and white noise by the augmented Dickey–Fuller (ADF) and then select the best order of differencing or (d) for the time series based on ADF test.

2. Model Estimation
Estimate the coefficients of the ARIMA model using a method such as maximum likelihood estimation or least squares estimation. The differenced series at time t of the ARIMA model is typically expressed as in (7). Once the parameters p, d and q are selected, the model can be fit to the data using methods such as maximum likelihood estimation.

$$\Delta^d Y_t = C + \varphi_1 \Delta^d Y_{t-1} + \cdots + \varphi_p \Delta^d Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{7}$$

3. Model Selection
Selecting the best parameters value for p and q using (AIC) and (BIC) criteria.

The Seasonal Auto Regression Integrated Moving Average Model
The SARIMA model is a statistical model extending the ARIMA model to capture both seasonal patterns and non-stationarity in data [8]. The model is denoted as SARIMA (p, d, q) (P, D, Q, s), where p, d, and q are the non-seasonal orders, P, D, and Q are the seasonal orders, and s is the length of the seasonal cycle. The steps of obtaining a forecast using the SARIMA model are as follows:

1. Model Specification
For the selection of the parameters (p, d, q) (P, D, Q, s), we begin by identifying the seasonal period (s). This is done by examining the data plot and the autocorrelation function (ACF) plot to look for repeating patterns. The non-seasonal parameters (p, d, q) can be determined similarly to the ARIMA model. For seasonal parameters, s is determined by detecting the seasonal pattern repeated in ACF and PACF. D is obtained after applying seasonal differencing, checking if the seasonal pattern has been adequately removed by using the ADF test or visual inspection of the ACF. Significant spikes at the seasonal lags ACF and PACF Plots are indicators for Q and P respectively.

2. Model Estimation
Estimate the coefficients of the SARIMA model using a method such as maximum likelihood estimation or least squares estimation .

3. Model Selection
Selecting the best parameters value using (AIC) and (BIC) criteria.

The Triple Exponential Smoothening Model
Triple Exponential Smoothing, also known as Holt-Winters' Exponential Smoothing [9], is an extension of the exponential smoothing method that allows you to model time series data with both trend and seasonality. It's one of the most effective forecasting methods for data that exhibits both of these characteristics. The steps of obtaining a forecast using the TES model are as follows:

1. Model Initialization

- **Level Initialization: The initial level Lt in (8) is often set to the average of the first season data.**
- **Trend Initialization: The initial trend Tt can be set as the average change in level over the first season, obtained from (9)**
- **Seasonal Initialization: The initial seasonal indices St can be estimated by (10) as the ratio of the observed value to the level for each point in the first cycle.**
- **Forecast: Assuming additive demand patterns, the forecast of the data set can be obtained using (11):**

$$L_t = \alpha \cdot \frac{Y_t}{s_{t-m}} + (1-\alpha).(L_{t-1} + T_{t-1}) \quad (8)$$

$$T_t = \beta \cdot (L_t - L_{t-1}) + (1-\beta).T_{t-1} \quad (9)$$

$$S_t = \gamma \cdot \frac{Y_t}{L_t} + (1-\gamma).s_{t-m} \quad (10)$$

$$F_{t+k} = (L_t + k.T_t).s_{t+k-m} \quad (11)$$

2. Model Fitting

The parameters α (the smoothing parameter for the level), β (the smoothing parameter for the trend), and γ (the smoothing parameter for seasonality) need to be optimized. This can be done using fit () function in python which automatically selects the parameters to minimize the loss function. The seasonal period m is also selected based on the domain knowledge of the data.

3. Forecasting

Once the model is fitted, you can use the final values of Lt, Tt, and St to forecast future values.

RESULTS

The candidate models in this study are tested on two data sets obtained from [10] and [11]. These datasets were selected based on their linear trend patterns and were then visually inspected using seasonal decomposition to identify their trend and seasonality patterns, as illustrated in Figure 1 for Dataset 1 as an example. The datasets were divided into training and test sets, with the training set used to fit the models and the test set reserved for model evaluation. Figure 2 displays Data set 1 split into training and test sets with an 80 % to 20 % ratio, which aligns with the common practice of data splitting. The results of the candidate models for each data set are presented in the following sections.

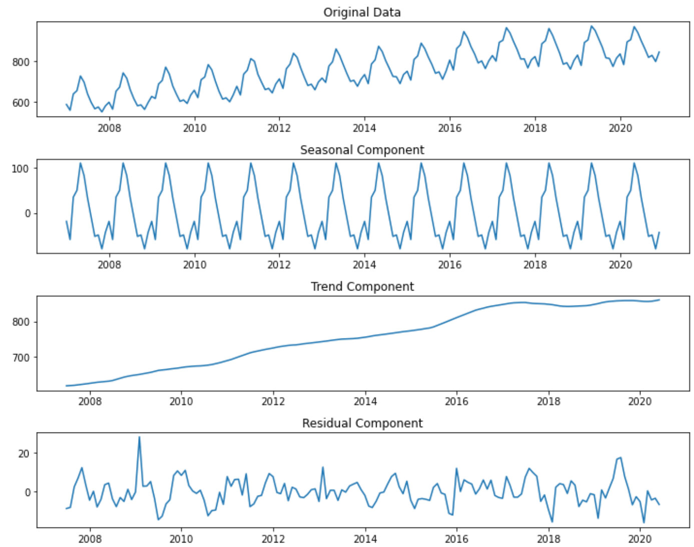


Fig. 1 Seasonal decomposition of data set 1.

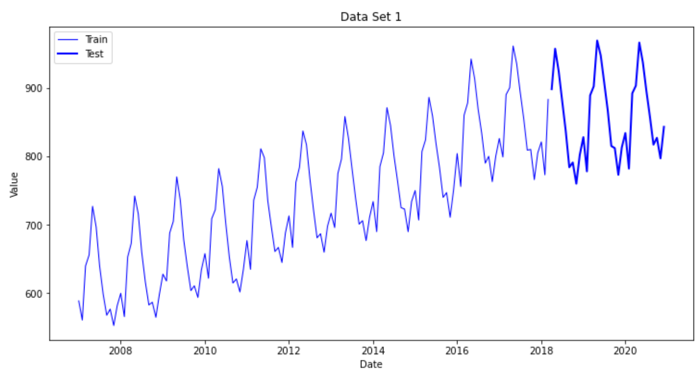
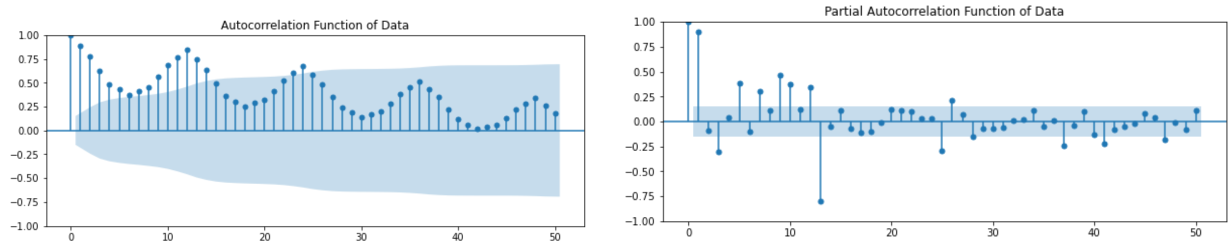


Fig. 2 Splitting of data set 1 to train and test sets.

Results for Data Set 1

After analyzing Dataset 1 using seasonal decomposition graphs, it is clear that the data exhibits a linear increasing tr end and a seasonal pattern of 6 or 12. The ACF and PACF plots shown in Figures 3. a and 3. b were generated to detect significant lag values. Based on this analysis, Table 1 presents the parameters of each candidate model along with the various performance measures used for model evaluation, where the selected parameters are based on minimizing the AIC. Table 1 also includes the TES parameters obtained after fitting the model and its performance compared to other models. Figure 4 shows the plot of forecasts of all models against the test set. These results conclude that the SARIMA has the best performance for data set 1 among all candidate models. TES comes next and then ARIMA. With the ARMA has the worst performance.



(a) **(b)**

Fig. 3 Analysis of data set 1 using different plots, where: (a) ACF of data set 1; (b) PACF of data set 1.

Table 1. Models parameters and performance for data set 1.

Model	Model Parameters	MSE	RMSE	MAE
ARMA	(1,2) intercept	14871.1739	121.9474	104.0142
ARIMA	(1,1,4) intercept	4835.331	69.5365	60.0942
SARIMA	(2,1,1) (0,1,1) [12]	186.7501	13.6657	11.6480
TES	$\boldsymbol{\alpha}$ = 0. 695, $\boldsymbol{\beta}$ = 5.961, $\boldsymbol{\gamma}$ = 1.219	511.8	22.623	19.3039

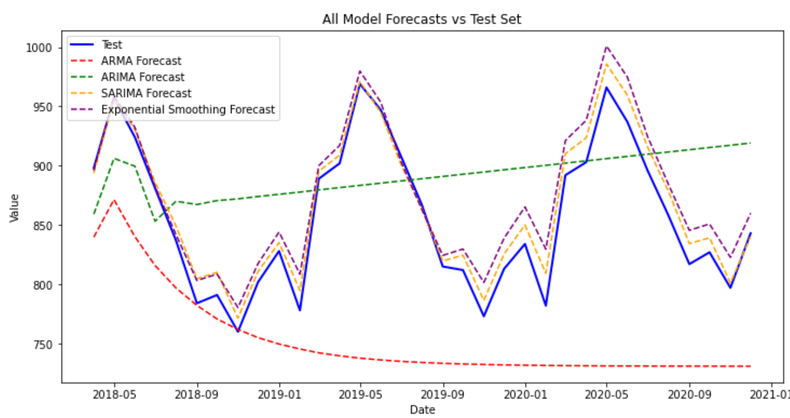
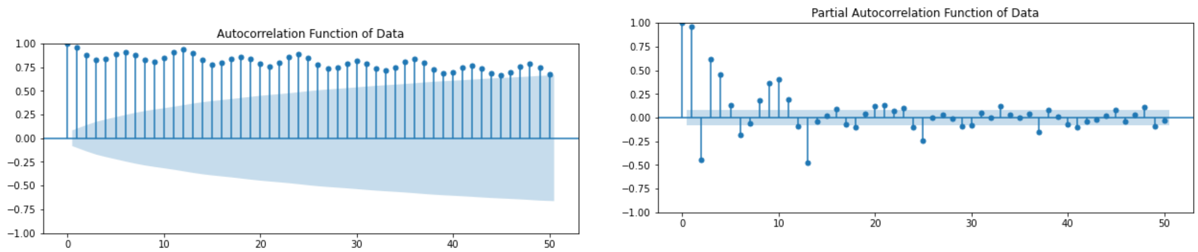


Fig. 4 All model forecasts vs test set of data set 1.

Results for Data Set 2

The analysis of this data through seasonal decomposition indicates a linear trend and seasonal factor of 12. After the proper selection of the parameters of the candidate models from ACF and PACF shown in Figures 5.a and 5.b, the error measures of each model shown in Table 2 indicate again that the SARIMA has the best performance.



(a) **(b)**

Fig. 5 Analysis of data set 2 using different plots, where: (a) ACF of data set 2; (b) PACF of data set 2.

Table 2 Models parameters and performance for data set 2.

Model	Model Parameters	MSE	RMSE	MAE
ARMA	(5,5)	90.7493	9.5262	7.5507
ARIMA	(2,1,1) intercept	187.8405	13.7055	11.5160
SARIMA	(0,1,4) (0,1,1) [12]	14.9952	3.8724	3.3210
TES	$\alpha = 0.392$, $\beta = 4.705$, $\gamma = 0.356$	65.1872	8.0739	6.9027

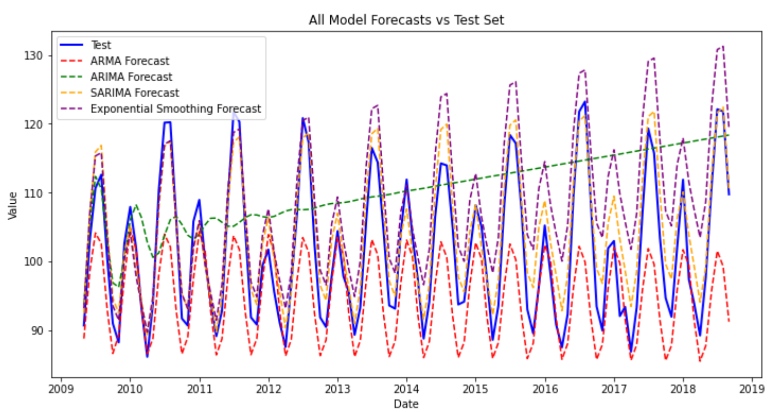


Fig. 6 All model forecasts vs test set of data set 2.

DISCUSSION

The analysis of the results reveals that the SARIMA model consistently outperforms other traditional time series models in handling data with both a linear increasing trend and seasonal components. SARIMA's superior performance is attributed to its capacity to model both non-seasonal and seasonal variations simultaneously. This makes it highly suitable for lubricants demand forecasting in situations where seasonality, alongside steady growth, plays a critical role. Its effectiveness in Dataset 1, where these patterns were prominent, underscores its reliability for such use cases. The Triple Exponential Smoothing (TES) model also performed well, particularly in datasets with clear seasonal fluctuations. TES's ability to account for both trend and seasonality ensures it remains a robust choice for forecasting in situations where demand patterns are primarily additive. However, its slightly lower performance compared to SARIMA suggests that TES may not capture more complex seasonality as effectively.

On the contrary, the ARIMA model, although useful for datasets with a strong trend, struggled to perform well in Dataset 2, which exhibited clear seasonality. Its limited ability to model seasonal components resulted in poorer accuracy compared to SARIMA and TES. This highlights ARIMA's limitations in scenarios where seasonality is integral to the data. However, ARIMA remains a valid choice for datasets where trends dominate but seasonality is minimal or absent. The ARMA model, which performed the weakest overall, failed to capture the increasing trend in Dataset 1 and was only somewhat effective in Dataset 2, where the linear trend was less pronounced. This reinforces the importance of carefully selecting forecasting models based on the specific characteristics of the dataset. ARMA's simplicity is its strength in stable, non-trending datasets, but it lacks the sophistication required for complex demand forecasting with both trend and seasonal components.

These findings underscore the critical need for lubricant businesses and practitioners to carefully assess the underlying patterns in their data when selecting a forecasting model. Models like SARIMA, which can simultaneously address both trend and seasonality, provide the best overall performance for datasets exhibiting such characteristics. Furthermore, even slight improvements in forecast accuracy, such as those achieved by SARIMA, can have substantial benefits for lubricant businesses, including more efficient inventory management, reduced costs, and improved customer satisfaction.

CONCLUSIONS AND FUTURE WORK

This study has demonstrated the importance of selecting appropriate time series forecasting models for lubricants demand forecasting where linear increasing trend and seasonality existed. SARIMA emerged as the most effective model, capable of accurately capturing both trend and seasonal patterns, making it a highly versatile option for demand forecasting in a variety of industries. TES also performed well, proving valuable in cases where seasonality was dominant, though it was less flexible in handling more complex patterns. ARIMA, while useful for purely trending data, showed limitations in datasets with significant seasonal components. ARMA, with its simplicity, was the least effective in this study, reinforcing the need for more sophisticated models in demand forecasting.

The results of this study highlight that even small improvements in lubricants demand forecasting precision can lead to significant gains in supply chain management. Accurate demand forecasting can drive better decision-making, optimize inventory levels, and reduce operational costs. For future research, incorporating more advanced models, such as ARFIMA or structural time series models, and exploring the impact of external variables, such as economic conditions, could further improve the accuracy and applicability of demand forecasting in various real-world contexts. For the future work of this research, other factors affecting the lubricants demand can be considered in further study with models that can handle different factors on demand.

REFERENCES

1. Chopra S. and Meindl P., "Supply chain management: strategy, planning, and operation", Pearson, (2016).

2. W. S. C. Sail & Makridakis H. R. J., "Forecasting: Methods and Applications (3rd ed.)", (1998).
3. Makridakis S., Spiliotis E., and Assimakopoulos V., "The M4 Competition: 100,000 time series and 61 forecasting methods," Int J Forecast, vol. 36, no. 1, pp. 54–74, Jan. (2020), doi: 10.1016/j.ijforecast.2019.04.014.
4. Makridakis S., Hibon M., and Moser C., "Accuracy of Forecasting: An Empirical Investigation," J R Stat Soc Ser A, vol. 142, no. 2, p. 97, 1979, doi: 10.2307/2345077.
5. Gardner E. S., "Exponential Smoothing: The State of the Art," (1985).
6. Taylor J. W., "Exponential smoothing with a damped multiplicative trend," Int J Forecast, vol. 19, no. 4, pp. 715–725, Oct. (2003), doi: 10.1016/S0169-2070(03)00003-7.
7. Elneel L., Zitouni M. S., Mukhtar H., and Al-Ahmad H., "Examining sea levels forecasting using autoregressive and prophet models," Sci Rep, vol. 14, no. 1, Dec. (2024), doi: 10.1038/s41598-024-65184-0.
8. Box G. E. P., Jenkins G. M., and Reinsel G. C., "Time series analysis : forecasting and control", John Wiley, (2008).
9. "6.4.3.5 Triple Exponential Smoothing", Accessed: Aug. 26, (2024). [Online]. Available: https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc435.htm
10. "GitHub - awesomedata/awesome-public-datasets: A topic-centric list of HQ open datasets", Accessed: Jul. 29, (2024). [Online]. Available: https://github.com/ awesomedata /awesome-public-datasets
11. "Find Open Datasets and Machine Learning Projects | Kaggle", Accessed: Jul. 30, (2024). [Online]. Available: https://www.kaggle.com/datasets