# Time series forecasting for tobacco product sales employing SARIMA, ETS, and TBATS models

URAZ COREKCI, MEHTAP DURSUN
Industrial Engineering Department
Galatasaray University
Istanbul, TÜRKIYE

*Abstract:* — In this study, performance of three different time series are examined for predicting the monthly sales of tobacco products in the FMCG industry. Key forecast accuracy metrics, such as MAE, MSE, MASE, and sMAPE, are used to compare and assess each model using actual anonymized sales data. While ETS offers reliable performance with a more straightforward structure, the SARIMA model successfully captures regular seasonal patterns. Although it can occasionally result in overfitting, TBATS shows great versatility when modeling multiple or non-integer seasonalities. Python libraries like statsmodels and tbats are used to implement all models in the Google Colab environment. The results leads us in the selection of suitable forecasting techniques for stock management and operational planning in highly regulated and seasonal industries such as tobacco.

*Keywords:* —Time Series Forecasting, SARIMA, ETS, TBATS, Tobacco Sales, FMCG Sector, Forecast Accuracy Metrics

Received: April 17, 2025. Revised: May 21, 2025. Accepted: June 23, 2025. Published: September 11, 2025.

## 1. Introduction

Successful forecasting has become a central concern for firms within the fast-moving consumer goods (FMCG) industry. Firms are required to rely on reliable forecasting models to predict future demand and facilitate operational efficiency because of shifting demand patterns especially in seasonal sales, intricate supply chains, and seasonally influenced consumption patterns. This is especially true for the tobacco industry, where there may be seasonal fluctuations, regulatory trends, holidays and changes in consumer trends that all affect sales.

This study aims to test and compare the forecasting performance of three well-known time series forecasting models in predicting tobacco monthly sales: TBATS (Trigonometric, Box-Cox, ARMA errors, Trend and Seasonal components), ETS (Error-Trend-Seasonality), and SARIMA (Seasonal Autoregressive Integrated Moving Average). They are commonly used models in practice for demand forecasting in various types of industries and are well known in literature for their capacity to capture various characteristics of various types of data [1].

Real anonymized sales data of a tobacco products organization are used for analysis purposes so that the research considers actual real-life problems encountered in the industry. TBATS is formulated to handle more complex patterns, i.e., multi and non-integer seasonality, prevalent in high-frequency or irregular data series, but SARIMA is specifically tuned to fit to time series with periodic and easily identifiable seasonal patterns. ETS provides a flexible model that adjusts to the fluctuations and changes in trend and seasonality. Models are run using Python with couple of libraries like statsmodels and TBATS. After the performance is completed, it is tested using popular measures of accuracy like MAE (Mean Absolute Error), MSE (Mean Squared Error), MASE (Mean Absolute Scaled Error), and sMAPE (symmetric Mean Absolute Percentage Error). Preparation operations like Box-Cox transformation, and differencing are also utilized in order to enhance model performance.

This study aims to determine the best-performing approach and also aims to be practically useful for demand planning and strategic market decision-making by testing these models in the framework of the tobacco industry. Finally, the results assures the opportunity to future research on hybrid models and automation in predictive analytics and add to the expanding significance of research into time series forecasting [2].

## 2. Methodology

Using a quantitative modeling approach, this study compares and assesses the forecasting performance of three sophisticated time series models in the context of predicting sales for fast-moving consumer goods (FMCG): TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components), ETS (Error-Trend-Seasonality), and SARIMA (Seasonal Autoregressive Integrated Moving Average).

1. Data Collection and Preprocessing

An FMCG company that operates in the Turkish market provided the dataset's anonymized historical sales data. The data shows distinct seasonal patterns, trends, and erratic fluctuations over two years at a weekly frequency. The following procedures are utilized to prepare the data for forecasting:

- Box-Cox Transformation: Applied to stabilize variance and improve model fit.
- Differencing: Conducted as necessary to achieve stationarity for SARIMA.

2. Model Implementation

a. SARIMA: The SARIMA model is chosen since it can easily identify seasonal and trend patterns in the given data. The Augmented Dickey-Fuller (ADF) test is used in order to evaluate if the data is stationarity and find the necessary degree of differencing prior to model identification [3]. In order to attain the best forecasting performance, model orders (p, d, q)(P, D, Q)[s] are then determined by analyzing autocorrelation (ACF) and partial autocorrelation (PACF) plots, these are all implemented in a Google Colab environment [4].

b. ETS: The ETS model is selected for its performance in non-linear seasonal data and is naturally capable of handling exponential smoothing. Its components are error, trend, and seasonality which are implemented using the statsmodels library [1].

c. TBATS: This model is selected due to its versatility in simulating seasonal patterns using the Python TBATS package. During the initial training, normalizing parameters and the Box-Cox transformation are optimized [5].

3. Evaluation Metrics

Model performance was compared using the following metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Symmetric Mean Absolute Percentage Error (sMAPE)
- Mean Absolute Scaled Error (MASE)

Each model was used to forecast over the selected period, and the metrics above were applied to assess predictive accuracy.

4. Tooling and Environment

All modeling, preparations, and evaluation procedures were implemented in Python using a Google Colab environment. Libraries used include pandas, numpy, statsmodels, tbats, sklearn, and matplotlib for visualization in some cases.

## 3. Case Study

Anonymized monthly sales data from a large tobacco company is used in this study for the purposes of supporting demand planning and inventory optimization in a fast-moving consumer goods (FMCG) industry, the forecasting accuracy of three time series models—SARIMA, ETS, and TBATS—were compared.

Key characteristics of the dataset:

- Period: 36 months of historical sales
- Frequency: Monthly
- Seasonality: Yearly cycles observed
- Data pre-processing:
    - Box-Cox transformation applied
    - Differencing performed for stationarity

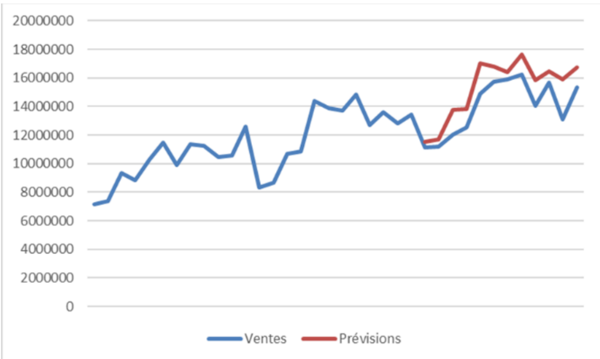
Statsmodels and tbats libraries in Google Colab was used to build and test the models in Python. This real-life study shows how useful time series forecasting models can be for making strategic decisions in industries with strict rules and changing demand patterns. The table below shows a summary of the inputted dataset.

Table I. Sales Data for 2022–2023

| Date | Sales | Date | Sales |
|---|---|---|---|
| 31.01.2022 | 7124558 | 31.01.2023 | 8321357 |
| 28.02.2022 | 7389722 | 28.02.2023 | 8682800 |
| 31.03.2022 | 9314775 | 31.03.2023 | 10660718 |
| 30.04.2022 | 8841353 | 30.04.2023 | 10847752 |
| 31.05.2022 | 10284916 | 31.05.2023 | 14401043 |
| 30.06.2022 | 11457984 | 30.06.2023 | 13867112 |
| 31.07.2022 | 9926079 | 31.07.2023 | 13691731 |
| 31.08.2022 | 11352148 | 31.08.2023 | 14844190 |
| 30.09.2022 | 11220371 | 30.09.2023 | 12693086 |
| 31.10.2022 | 10479377 | 31.10.2023 | 13579008 |
| 30.11.2022 | 10598419 | 30.11.2023 | 12831733 |
| 31.12.2022 | 12580839 | 31.12.2023 | 13447342 |

The dataset was well-fitted by the SARIMA model, which particularly captured the consistent seasonality patterns found in the sales data. SARIMA accomplished stable performance metrics, especially for datasets with almost regular periodic fluctuations, as the table below indicates.

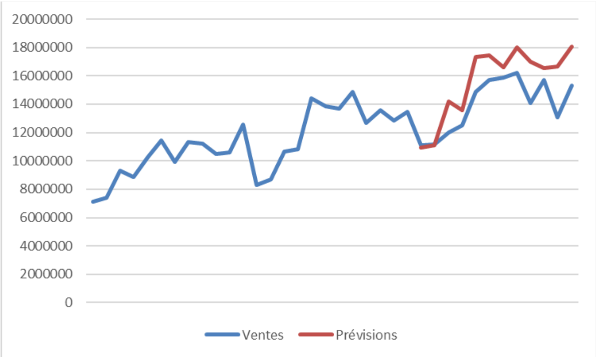
Figure I. Graphical representation of observed sales and forecasts for SARIMA




According to the findings, SARIMA works well with time series that have consistent and significant seasonal components.

Strong adaptability to intricate seasonal patterns and erratic trends was shown by the TBATS model. The performance metrics listed below demonstrate its capacity to manage multiple or non-integer seasonality.

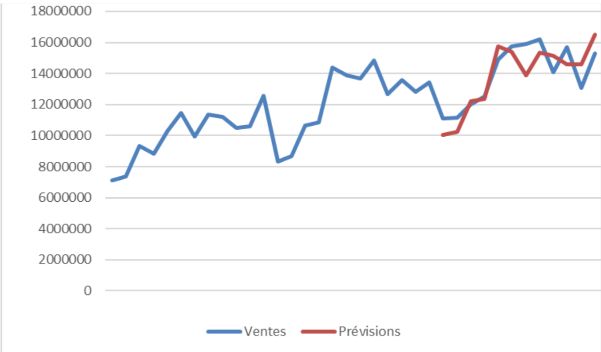
Figure II. Graphical representation of observed sales and forecasts For TBATS




When it came to capturing more complex temporal dynamics, TBATS performed marginally better than other models, as the table shows.

Overall, the ETS model's forecasts were the most accurate and reliable, especially because of its smoothing methods. The table illustrates how ETS successfully managed the seasonal and trend components.

Figure III. Graphical representation of observed sales and forecasts For ETS




ETS achieved the best forecasting performance, making it a reliable choice for short-term sales prediction.

## 4. Conclusıons

This study looked at three well-known time series forecasting models—SARIMA, ETS, and TBATS—and compared them all in depth. It used monthly sales data from the tobacco industry from 2022 to 2023. Each model had its own strengths and weaknesses based on the data patterns, especially when it came to seasonality, trend behavior, and noise.

We used MSE, MAE, MASE, and sMAPE to test how well the models could predict the future. The table below shows the results.

Table II. Model performance comparison based on evaluation metrics

| | SARIMA | TBATS | ETS |
|---|---|---|---|
| MSE | 2247216646170 | 3994781737501 | 1135061949756 |
| MAE | 1332390 | 1684355 | 937577 |
| MASE | 1.07 | 1.35 | 0.75 |
| sMAPE | 9.06% | 10.97% | 6.76% |

Overall, the ETS model proved to be the most dependable, offering consistent, precise forecasts with a more straightforward structure. Particularly appropriate for datasets with regular patterns and no major differences, its consideration of Error, Trend, and Seasonality components made modeling simple and straightforward.

However, for datasets with regular seasonal structures, SARIMA performed well. In situations where the seasonal effect was evident and recurring, SARIMA was able to generate accurate forecasts by utilizing autoregressive/moving average components and differencing to capture both seasonal and non-seasonal dynamics. Nevertheless, it made the modeling process more difficult because it required careful hyperparameter tuning and stationarity checks using methods like the ADF test.

The TBATS model performed better in scenarios with multiple or non-integer seasonality, such as overlapping monthly and annual cycles, despite requiring more computation. Its use of trigonometric terms and the Box-Cox transformation allowed for greater flexibility in handling complex time series. However, its risk of overfitting and longer training times may limit its applicability in some real-time business scenarios.

All things considered, this comparison showed that no single model is always better than another and that the features of the data, the forecasting horizon, and the organization's operational requirements should all be taken into consideration when selecting a model. Hybrid approaches or model ensembles may provide even better performance in the future for the tobacco industry, where seasonal peaks and regulatory impacts are critical factors.

Strategic decision-making can be aided by the knowledge gathered from this study, especially in the areas of demand forecasting, inventory planning, and sales optimization. Businesses can increase forecast accuracy, cut waste, and become more responsive to market fluctuations by utilizing the right time series model.

This study provides a number of new questions for investigating. In order to improve forecasting accuracy, one promising approach is to incorporate different variables into the models, such as economic indicators, public health initiatives, holidays, public events and fiscal policies. Furthermore, investigating hybrid modeling strategies that blend machine learning algorithms like SARIMA-LSTM with statistical techniques may improve handling of patterns in sales data. The accuracy of the models could be tested by extending the current methodology to other FMCG categories. Future research might also concentrate on using a Python-based pipeline which feeds from the daily data to automate the forecasting process, allowing for real-time application and improving operational effectiveness. Finally, more accurate forecasting results would result from better data processing.

*Acknowledgment*

This research is financially supported by Galatasaray University Research Fund FBA-2025-1297.

## References

[1]. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts. Retrieved from: https://otexts.com/fpp2/

[2]. Perone, G. (2021). Comparison of ARIMA, ETS, NNAR, TBATS and Hybrid Models to Forecast the Second Wave of COVID-19 Hospitalizations in Italy. The European Journal of Health Economics, 23(6), 917–935. Retrieved from: https://link.springer.com/article/10.1007/s10198-021-01347-4

[3]. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431.

[4]. Statistics by Jim. (2021). Autocorrelation and Partial Autocorrelation in Time Series Data. Retrieved from: https://statisticsbyjim.com/time-series/autocorrelation-partial-autocorrelation/

[5]. Skorupa, G. (2019). Forecasting time series with multiple seasonalities using TBATS in Python. Intive Developers. Retrieved from: https://medium.com/intive-developers/forecasting-time-series-with-multiple-seasonalities-using-tbats-in-python-398a00ac0e8a