Original Research Article

Power transformer demand forecast with Box Jenkins ARIMA model

Özlem Kuvat[1*], Ege Adalı[2]

[1] Business Management, Faculty of Economics and Administrative Sciences, Balıkesir University, Balıkesir, Turkey
[2] Business Administration Graduate Student, Institute of Social Sciences, Balıkesir University, Balıkesir, Turkey

ARTICLE INFO

ABSTRACT

Demand forecasting is based on the principle of trying to forecast the demand for the outputs of enterprises in the field of manufacturing or service for the next periods. It requires the estimation of various future scenarios, if necessary, taking measures and taking steps, and during the application phase, the technique that is most suitable for the characteristics of the examined data set is selected and used. As a result of a healthy analysis carried out in this way, detailed plans and strict measures can be taken for the unknown, negative scenarios of the future.

This study analyzes the characteristics of a series of power transformers of a company operating in the electromechanical industry in the past years, and as a result of this analysis, the Box Jenkins Autoregressive Integrated Moving Average method (ARIMA), which best fits the results, is expected to occur for power transformers in the future. It was made to estimate the amount of demand.

Within the scope of this study, firstly, the most suitable model was tried to be determined by taking into consideration the past 132 months data of PTS. It was decided that the best choice among the alternative models was the ARMA (4,4) x (0,1) $_{12}$ model. The model was found to be stable and it was decided that the root mean square error (RMSE), mean absolute percentage error (MAPE) and Theil inequality coefficient values determined in the performance measurements were appropriate.

Keywords: ARIMA; Box Jenkins; Transformer demand forecast

1. Introduction

Demand forecasting is based on the principle of trying to forecast the demand for the outputs of businesses in the field of manufacturing or service for the next periods. In particular, it has gained a wide place in the manufacturing sector. Technological, commercial, economic, legal, administrative, political, cultural, social etc. in the global market. conditions are constantly changing; For this reason, in order to adapt to these changes in the fastest way, an effective demand forecast should be made and this should be implemented effectively. Demand forecasting the uncertainties and risks that the future carries as the focus of opinion, determining the situations that may cause problems, taking the necessary precautions for them, but if the problem cannot be prevented despite the measures taken, the problem is eliminated with the corrective actions determined by foreseeing these and similar negative scenarios and at least in the planning phase. it is aimed to reduce the negative impact compared to an unplanned process.

The main purpose of this study is; to analyze the demand series of the product group defined as power transformers of an enterprise operating in the electromechanical industry, to try to predict the amount of demand for power transformers using the Box Jenkins ARIMA method and to measure the performance of the study. The data used in the study consisted of the sales (monthly) of the power transformers sold between 2009-2019 and was provided by the implementing company.

In this study, it is aimed that the demand forecasting study for transformers, which are included in almost all electronic devices and have an important place in the sector, will be supportive of the planning activities for the enterprise.

2. Material and Methods

Box Jenkins ARIMA algorithm is a widely used method among linear time series prediction methods. It makes a strong assumption that future data values are linearly linked to current and past data values. In this way, it gives high accuracy results in estimation of stationary time series [1]. In the literature, it has been applied in different fields such as automotive export revenue [2], cotton price analysis [3], electricity price prediction [4], the remaining useful life of aircraft engines [5], near-term regional temperature and precipitation [6], forecast air temperature [7], gold price forecasting [8].

An ARMA model is labeled as an ARMA model (p, q), where in: p is the number of autoregressive terms; and q is the number of moving averages. In the equation, y_t representation observation values, $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ values error terms; $\emptyset_1, \emptyset_2, \ldots \emptyset_p$ and $\varphi_1, \varphi_2, \ldots \varphi_q$ values are the coefficients of the model. The model is formed as given in Equation 1. [2,9]:

$$y_t = c + \emptyset_1 y_{t-1} + \ldots + \emptyset_p y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \ldots - \varphi_q \varepsilon_{t-q} \qquad (1)$$

If the series is not stationary, the model is represented by ARIMA (p, d, q). Here d is the order of integration (number of differences), p is order of autoregressive process and q is order of moving average process. The seasonal ARIMA model is also classified as the ARIMA (p, d, q) x (P, D, Q)$_s$ model. Where P denotes the term SAR, Q denotes the term SMA for S seasonal period. D denotes the seasonal difference of order with period S [2, 9].

3. Forecasting for Power Transformer

Forecasting study was carried out in an enterprise operating in the electromechanical industry. It was carried out to analyze the characteristics of a series of sales values of the power transformers produced in the past years and to estimate the expected amount of demand for power transformers for the company in the coming period as a result of this analysis. Transformers are machines that can change the voltage and current value of electrical energy according to the need, and have an important place in the Electrical Machinery and Equipment Sector. In addition, transformers are now available in almost every electronic device and are of importance for devices that make our lives easier in our daily lives.

Power transformers are transformers in high power class. These are generally used in electricity transmission and distribution networks. Transmission and distribution networks ensure that the production and consumption processes of electrical energy occur uninterruptedly and safely. In addition, widespread uses of power transformers can be found in production facilities and industrial centers to meet electrical requirements.

In this paper, firstly, descriptive statistics related to past sales data obtained for the purpose of demand forecasting were calculated, and then forecasting study was performed with Box Jenkins methods. Box Jenkins methods, which have a more complex structure than other estimation methods, allow the time series to be analyzed and evaluated in terms of stationarity and seasonality. In this paper, 132 observation values regarding the power transformer sales amount in 2009: 01 and 2019: 12 periods were taken into consideration. The time path graph is PTS (Power Transformer Sales) shown in Figure 1.
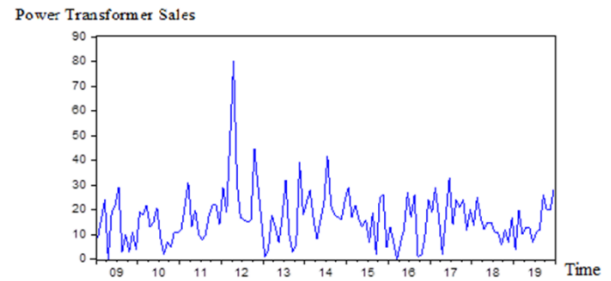

Fig. 1. Time path graph for power transformer sales

The graphic of the series gives the impression that it is stationary, but outlier values attract attention in some periods. Descriptive statistics data for the series are presented in the Table 1.

Table 1. Descriptive statistical data of the series

Parameters	Values
Average	16.71970
Median	16.00000
Maximum	80.00000
Minimum	0.000000
Standard deviation	10.76204

When descriptive statistics are analyzed, it is seen that the average value for the power transformer sales amount is 16.71 and the median value is 16. According to the results of Skewness, Kurtosis and Jarque Bera Probability value, it does not show normal distribution feature. The chart of the series is shown in Figure 2.

After evaluating the descriptive statistics, firstly the stationarity of the series was examined in order to decide on the analyzes. In this paper, Augmented Dickey-Fuller (ADF), Philips-Perron (PP) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) tests were applied for unit root test of power transformer sales series [10, 11, 12]. The test results are shown in Table 2.
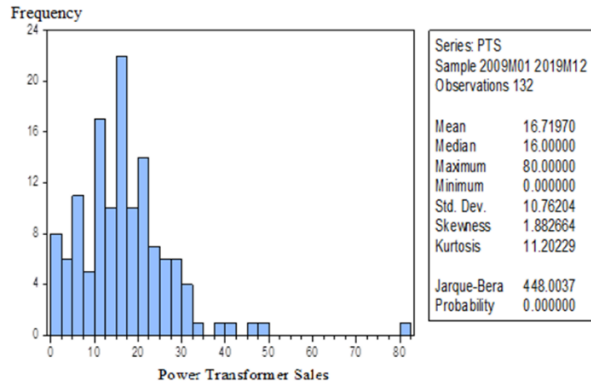
Fig. 2. Graph of the descriptive statistical data of the series

Table 2. Unit root stationary tests for Level (trending test results)

	ADF	PP	KPSS
Statistics	-7.924200	-7.875100	[b] 0.137009
[a] **Probability**	[a] 0.0000	[a]0.0000	-------
Critical value 1%	-3.480818	-3.480818	0.739000
Critical value 5%	-2.883579	-2.883579	0.463000
Critical value 10%	-2.578601	-2.578601	0.347000

[a]Probability: If the probability value for ADF and PP is greater than 0.05, the H0 hypothesis (H0: Series is not stationary) is accepted, series is not stationary. In case the probability is less than 0.05, H0 hypothesis is rejected. [b]KPSS $_{LM}$ statistics: H0 hypothesis (H0: Serial is stationary) is accepted when the LM statistic value calculated for KPSS is less than the critical value of 1%.

According to the table, the PTS series ADF, PP and KPSS unit root test result is stable at the level when compared with the critical values at 1%, 5% and 10% significance levels. Therefore, the series is I (0) for all three tests. If the series constitutes a stationary series, the parameter d is set to 0 and it can be shown as an integrated I(0). It becomes the ARMA series instead of the serial ARIMA. Since the PTS series is arranged with monthly data, seasonal parameters were also used in the trials to examine the seasonal inclusion status and it was worked with 12 seasonal periods. Model selection was made according to Akaike Information Criteria (AIC). AIC values for alternatives can be seen in Figure 3.
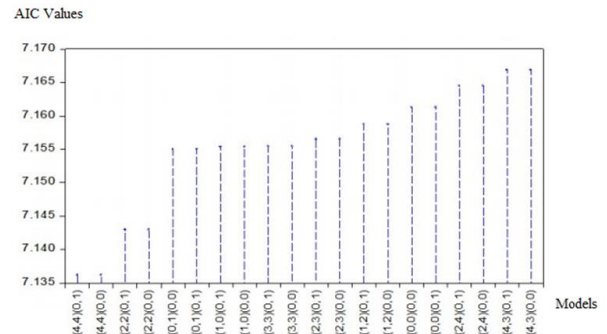

Fig. 3. AIC values of models

According to the lowest AIC criteria, it was decided that the best choice among the alternative models was the ARMA (4,4) x SMA $(0,1)_{12}$ model, which has 7.13627 AIC. The dummy variable is also used because the series contains outliers. Model estimates results can also be seen in Table 3.

Table 3. Model estimates results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
(c)	15.59907	0.891039	17.50660	0.0000
Dummy	38.39092	2.953495	12.99847	0.0000
AR(1)	-0.404994	0.086186	-4.699091	0.0000
AR(2)	-0.457775	0.090221	-5.073921	0.0000
AR(3)	-0.433998	0.096969	-4.475630	0.0000
AR(4)	-0.831789	0.120255	-6.916895	0.0000
MA(1)	0.541879	77.72067	0.006972	0.9944
MA(2)	0.596306	105.4550	0.005655	0.9955
MA(3)	0.541877	115.5032	0.004691	0.9963
MA(4)	0.999995	353.5997	0.002828	0.9977
SMA(12)	-0.021217	0.124388	-0.170571	0.8648
SIGMASQ	61.80003	1817.519	0.034002	0.9729

Model Statistics			
R-squared	0.462347	Mean dependent var	1.671970
Adj R-squared	0.413063	S.D. dependent var	1.076204
S.E. of regression	8.245	Akaike info criterion	7.196584
Sum squared resid	8.157604	Schwarz criterion	7.458657
Log likelihood	-4.629746	Hannan-Quinn criter.	7.303079
F-statistic	9.381131	Durbin-Watson stat .	1.978905
Prob(F-statistic)	0.000000		

In the model obtained, the coefficients are significant since the probability values of the constant (c), dummy and AR coefficients are less than 0.05. R-squared R^2 value for the model was 0.462347 and Adjusted R-squared R^2 value was obtained as 0.413063. While the F-statistics for the model was 9.381131, the probability value was obtained as 0.00. The residual values of the model were analyzed with the Jarque-Bera Normality test. Since the Jarque-Bera probability value (0.311> 0.05), the distribution of the residues is normal. Test results are given in Figure 4.
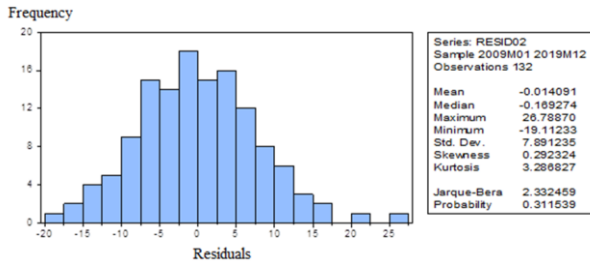

Fig. 4. Residual values analysis

Inverse roots of the AR polynomial are examined to control the stability of the model. The fact that the inverted roots are inside the circle or the modulus values are less than 1 indicate stability [13]. Looking at the polynomial given in Figure 5, it is seen that all the opposite roots are located
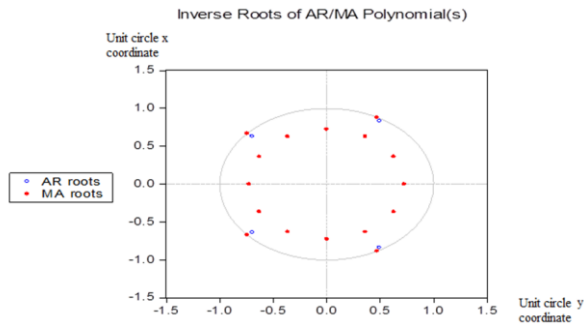

Fig. 5. Inverse roots inspection

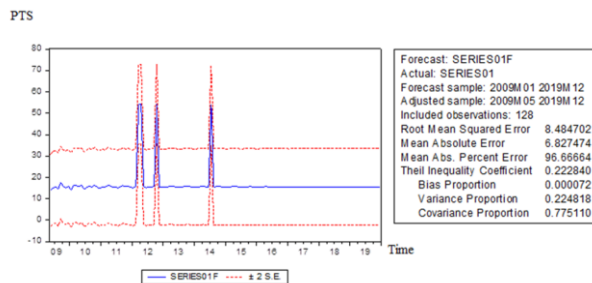Values obtained for the performance criteria are given in Figure 6.

Fig. 6. Performance criteria

The criteria taken into consideration during the estimation phase are:

- RMSE and MAPE values are expected to be small. These values are calculated as in the Equations (2) and (3) given below. Here, the real values of y_t show the estimated values of $\hat{y}_t$ and the number of T estimates [14]:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_t - \hat{y}_t)^2} \qquad (2)$$

$$MAPE = \frac{1}{T}\sum \left|\frac{y_t-\hat{y}_t}{y_t}\right| * 100 \qquad (3)$$

- Theil's inequality coefficient is expected to take a value close to zero.

- Bias proportion points to a systematic error. Its close to zero indicates that the results are reliable.

- Variance proportion value expresses how much the variability in the series can be predicted. The value of the series is regarded as unpredictable by the percentage value obtained here; the value for this reason is expected to be small.

- Covariance proportion value shows the non-systematic error. Achieving a larger covariance proportion compared to other coefficients indicates that the error is not systematic; therefore, this value is desired to be large [15].

According to the results here, the model is correctly determined and there is no systematic error. The non-systematic error percentage is 0.77 and 77% of the variability in the series could be predicted. As given in the graphic in Figure 6, it has been determined that the estimated last period values forecasted sales of 15.5990 units. It will be appropriate to consider this value for the next 3-6 months planning period. For the next periods, the estimation study can be repeated.

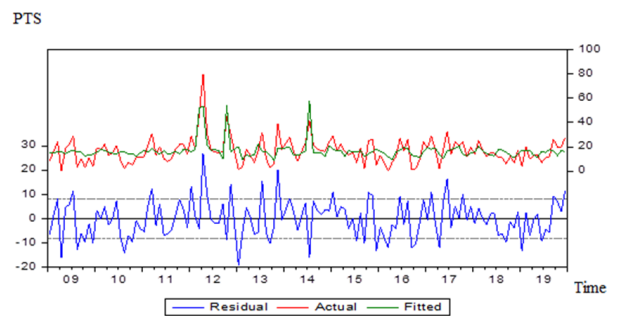Residual, actual and fitted values obtained for model are given in Figure 7.

Fig. 7. Comparison chart

According to the graph's appearance, it can be said that the prediction series follows the original series. Because of using dummy variable, outlier values could also be estimated According to the graphic obtained, the overlap status of the series can be monitored.

4. Conclusion

The electricity we use almost every moment of our daily life goes through many stages until it reaches us. The majority of

electricity carried by power lines is used in workplaces and homes. Electricity is also an important sign of modernization and is of great importance in our lives as we cannot fully realize it. The health, traffic, transportation, education, sports, commerce, technology, communication, security, water, energy, manufacturing, press and broadcast sectors are the only ones that come to mind first of electricity-dependent fields.

Currently, many of us are not aware of how much electricity covers most of our lives. Unfortunately, we cannot use the electrical energy that contributed so much to our lives from the moment it was produced. The voltage of electrical energy, which can be produced in different plants by using different sources such as wind or solar energy, is firstly raised in the substations. Then, it is necessary to travel long distances and reach cities efficiently. Therefore, electrical energy is sent from the substations to electricity transmission lines. These transmission lines are often called high voltage lines. After the electricity reaches the substations in the neighborhood, the voltage of the electricity is lowered. The electricity, whose voltage is lowered, is directed to the distribution lines in the neighborhoods before moving to the houses. Here, the inter-polar voltage of electricity is reduced; delivered to our homes.

Of the machines called transformers; Electricity is really important in this long and difficult process that takes place until it reaches our homes. In addition to all of these, transformers are no longer found in almost every electronic device, but perhaps fill a gap, whose importance can be understood only in its absence.

Within the scope of this paper, firstly, Akaike Information Criteria has been applied to determine the most suitable model considering the past 132 months data of PTS. As a result, it was decided that the best choice among alternative models was the ARMA $(4,4) \times (0,1)_{12}$ model. The model was found to be stable and it was decided that the RMSE, MAPE and Theil inequality coefficient values determined in performance measurements were appropriate.

These results show that Box Jenkins methods are suitable to be used in transformer production demand estimation studies. With this study, it is seen that the use of quantitative estimation methods can be used as a supportive tool in this sector planning studies. Similar approaches can be used for different sectors and can provide useful results.

ORCID

Özlem Kuvat   0000-0001-9864-7512

Ege Adalı   0000-0002-5715-1040

References

[1] Büyükşahin, Ü. and Ertekin, Ş. 2020. A feature-based hybrid ARIMA-ANN model for univariate time series forecasting. *Journal of the Faculty of Engineering & Architecture of Gazi University*, *35*(1), 468-478.

[2] Karakaş E. 2019. Forecasting Automotive Export Revenue of Turkey using ARIMA Model. *Journal of Yasar University,* 14/55, 318-328.

[3] Kaya L., Doğan, Z. and Binici, T. 2015. A comparative investigation of alternative estimation methods in non-stationary time series: analysis of cotton price. *Atatürk University Journal of Graduate School of Social Sciences* 19(2), 401-421.

[4] Contreras, J., Espinola, R., Nogales, F. J. and Conejo, A. J. 2003. ARIMA models to predict next-day electricity prices. *IEEE Transactions On Power Systems*, 18(3), 1014-1020.

[5] Ordóñez, C., Lasheras, F. S., Roca-Pardiñas, J. and de Cos Juez, F. J. 2019. A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines. *Journal of Computational and Applied Mathematics*, 346, 184-191.

[6] Lai, Y. and Dzombak, D. A. 2020. Use of the autoregressive ıntegrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting*, *35*(3), 959-976.

[7] Wang, H., Huang, J., Zhou, H., Zhao, L. and Yuan, Y. 2019. An integrated variational mode decomposition and ARIMA model to forecast air temperature. *Sustainability*, 11(15), 4018.

[8] Guha, B. and Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science,* 4(2), 117-121.

[9] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H. and Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1847979018808673.

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

[11] Phillips, P. C. and Perron, P. 1988. Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346.

[12] Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), 159-178.

[13] Nyoni, T. 2019. Modeling and forecasting inflation in Tanzania using ARIMA models. MPRA Paper No. 92458, posted 03 Mar 2019 19:07 UTC, 12p.

[14] Şahan M. and Okur Y. 2016. Estimation of the solar radiation using some meteorological data for the mediterranean region with the artificial neural network. *SDU Journal of Science (E-Journal),* 11(1), 61-71

[15] Bozkurt, H. Y. Zaman Serileri Analizi, 2. Baskı, *Ekin Kitapevi,* Bursa, Türkiye, 2013.