

## Sales Forecast with YSA, ARIMA and ARIMAX Methods: An Application in the White Goods Sector

**Melih YÜCESAN**

Munzur University

Faculty of Engineering

62000 Tunceli, Turkey

[orcid.org/0000-0001-6148-4959](https://orcid.org/0000-0001-6148-4959)

[melihyucesan@munzur.edu.tr](mailto:melihyucesan@munzur.edu.tr)

### 1. Introduction

Reliable forecasting of sales will increase the efficiency of the business. Increased productivity plays an important role in increasing the quality of business strategy and reducing costs. In order to increase the competitiveness of the enterprise, the manager must make the right decision based on the existing information. Forecasting plays a vital role in the planning and operation of the retail industry. Sales forecast; marketing, production, etc. (Shih and Chung, 2008).

White goods sector, which is included in some sources of electrical household appliances and in some of the durable consumer goods, has a wide range of products. These products are mainly; refrigerator, washing machine, dishwasher, oven, cooker and broom and toaster, robot, fruit press, blender and mixer, freezer, air conditioner, water heater (T.C. ministry of science, industry and technology, 2012). The white goods sector, which is one of the important sectors in Turkey, is among the sectors that are affected by the changing and lively competition conditions with the recent technological developments and globalization (Alptekin, 2010, p. 18).

Despite the fact that sales forecasting studies have been implemented in many areas, sales forecasting for the white goods sector has not been done yet. This study aims to fill the gap in the literature. Apart from the white goods sector, sales forecasting studies have been implemented in many areas like food, retail, clothing, automotive.

### 2. Material Method

In this study, the sales amount of a district dealer selling refrigerators, washing machines, dishwashers, televisions and small home appliances was used between 2014-2017. The average monthly sales averages of refrigerators washing machines, dishwashers, television and small house appliances products were calculated as 19, 18, 11, 44, 42 respectively. For all years, an increase in the amount of sales is observed in summer. Dummy variables were used to explain the monthly effect of the sales amount.

In addition, dollar exchange rate, holiday days, producer price index (PPI), consumer confidence index (CCI) and housing sales in the region were used as input variables while making sales forecasts. PPI is an index that measures price changes over time in producer prices of goods and services manufactured in the country's economy. CCI current situation assessments of the consumer's personal financial situation and general economy with the monthly consumer tendency survey is a display aimed at measuring the expectations of the period and the near future spending and saving trends. ARIMA, ARIMAX, ANN methods were used to make sales forecasting studies. The explanations of the methods are presented below.

The time series model using the Box-Jenkins method was proposed by Box and Jenkins (1970). This approach is widely used in the literature because it is simple and gives good results. Often this method is referred to as ARIMA method. The ARIMA method is quite different from other methods because it does not contain explanatory variables that show the tendency of the prediction series. When an ARIMA model is being created, the current series must be stationary. If the series is not stationary then the difference of the series must be taken before using the method (Gahirwal, 2013). In this method, the future values of the depended variable are assumed that will be expressed as past values and past errors of dependent variable. The ARIMA method is expressed in Equation 1.

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \theta_q \varepsilon_{t-q} \quad (1)$$

p and q take integer values and represent the level of the model. The value of  $\varepsilon_t$  represents the error. We assumed zero mean and constant variance for errors. ARIMA model consists of AR, MA and I parts. I indicates at which level the depended variable is stationary. For  $q = 0$ , the model becomes an AR model. For  $p = 0$ , the model becomes the MA model. The main motivation in creating the ARIMA model is to determine the appropriate model level (p, d, q). For example ARIMA (1,2,3); It shows that the AR part is at the first level, the second level of the series is stationary, and the MA part is at the third level. Calculated ARIMA results'  $R^2$  values of refrigerators, washing and dishwashing machines, TV, small household appliances models were calculated as 0.5588, 0.4814, 0.4084, 0.4932 and 0.2748 respectively.

In recent years, the ARIMA method has been used frequently, since it is simple and gives good results. However, the ARIMA method has some disadvantages. One of them is insufficient to explain the real problems. Because only one variable is used in this method. It is possible to use more than one argument in the ARIMAX method (Fan et al., 2009; Jalalkamali et al., 2015). Instead of estimating the dependent variable  $Y_t$  with only the past data, it is intended to estimate  $Y_t$  with both the past values and the explanatory variables  $X_t$ . Thus, more accurate results are obtained compared to the ARIMA method (Cool et al., 2009; Neter et al., 1996). The  $R^2$  values of refrigerators, washing and dishwashing machines, TV, small household appliances models were calculated as 0.9152, 0.8957, 0.8099, 0.7285, and 0.6859 respectively.

ANNs are one of the most powerful computer modeling techniques based on statistical methods. The use of ANN is quite common in defining models when the physical model is inadequate and identifying difficulties (Ezugwu et al., 2005). ANNs are machine learning algorithms aimed at solving computational processes in specific

areas using a large number of interconnected processing elements (Yucesan et al., 2017). The Mean Squared Error (MSE) is one of the main parameters used in the performance measurement of ANN models. The calculation of MSE is shown in Equation 2.

$$OKH = \frac{1}{N} \sum (f_i - y_i)^2 \quad (2)$$

MATLAB 2016, a commercial software, was used to create ANN models. The input and output variables were normalized using the min-max normalization method before the program was run. Later, ANN properties and structure were determined. *Purelin* transfer functions are applied for output layers and *tansig* transfer functions applied for hidden layers. Model; training, verification, test data. In this study, the data about the 70:15:15 ratios were taken at random from the main verb. Levenberg-Marquardt training algorithm is used in the model. This algorithm gives very good results in nonlinear data (Yu and Wilamowski, 2011).

The accuracy of the ANN model is directly related to the number of neurons in the hidden layer. It was aimed to increase the number of neurons from 1 to 30 neurons and to find the model with the highest R value. The training, verification, test and calculated  $R^2$  values for all products were calculated as 0.9557, 0.72415, 0.67688 and 0.86904, respectively.

### 3. Result

The white goods sector, which is one of the important sectors in Turkey, is among the sectors that are affected by the technological developments and globalization that have been experienced in recent years and the changing competitive conditions. The first methods used in literature for sales forecasting are statistical methods such as trend analysis and exponential correction. However, these methods are insufficient to determine the trend of sales forecasts. Using time series forecasting models to make sales forecasts is both easy and relatively good.

In this study, ANN method used which produces very good results when many factors are effective on sales. ARIMA method used which gives very good results despite its simplicity without using explanatory variables and also ARIMAX method used which is very successful in determining linear relations. Sales forecast model for the white goods sector was proposed with mentioned methods. 46 months of sales have been used for dishwashers, refrigerators, small appliances and televisions. The exchange rate, holiday days, consumer confidence index, housing sales in the region and producer price index are used as explanatory variable. As far as we know, there is no sales forecasting work for the white goods sector in the literature yet.

It can be said that the most accurate estimates are obtained by using the ANN method when mean squared error (MSE) compared which is the performance criterion. Although the ANN method gives good estimation results at an acceptable level, it is still thought that this method should be improved. Increasing the explanatory variables is considered to have a positive effect on the results obtained from ARIMAX and YSA methods. ARIMAX models are insufficient to solve complex nonlinear problems. On

the other hand, ANN is insufficient to explain linear relationships. For these reasons, it is considered that, if we cannot determine the linear and nonlinear parts of the problems as in the sales forecast problems, the future use of ARIMAX and ANN methods as hybrids may increase the accuracy of estimation.