International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Pattern Recognition - Product Sales Analysis Using SARIMA Model in Time Series Forecasting

Ruchir Kulkarni[1], Milind Rane[2]

Bachelor of Technology, Department of Electronics Engineering, Vishwakarma Institute of Technology, Pune, India

Professor, Department of Electronics Engineering, Vishwakarma Institute of Technology, Pune, India

Abstract: ***Product sales analysis is the most important part for any working manufacturing industry. It not only provides an over sales record of their currently manufactured product, but also it helps them predict it performance in the near future. This performance can affect the company in a positive or negative way. For this effective analysis, I have used a SARIMAX model in time series forecasting, which I will be describing in this paper. This paper will explain why there is a need of such model instead of using a simpler form of regression to predict the sales values.***

Keywords: Product Sales, SARIMAX, time series, regression

1. Introduction

By using an ARIMA model, we are able to forecast a time series in series past values [1]. Time series analysis is a sequence of datapoints measured particular at successive time points. This particular method makes an attempt to understand, the underlying context of the data points, or to make forecasts, that is predictions about the future values of those particular data points. In this paper I have used an extension of the ARIMA model, Seasonal ARIMA, which is also called as the SARIMA model [3].

2. Dataset

The dataset used in the following product sales analysis consists of above ten lakh samples. The dataset ranges samples over a period many years and the samples range from daily to weekly basis. The dataset consists of multiple variables, which are the product number, the date of sale, the product category type, the warehouse from which it was manufactured and sold, the date at which a particular product was manufactured and sold and the overall units of the product ordered in that particular shipment. All these factors are taken into consideration to build a solid forecasting model to gauge the consumption of the sold products and their categories and the overall changing need of the product in the near future can be predicted using the time series forecasting in this given scenario. Considering the randomness of each of the variable, there is a dire need to establish a relationship between them to create a successful model to access the sale. Now due to this complex structure, simple regression techniques can be less to no effective in making a reliable model. Hence mapping all the variables against a time sequence can give us its usage value variance over a certain period and considering these past values, the future requirement can be predicted easily [6]. This type of model can be used to gauge the basic demand and supply model for a particular goods manufacturing company or be it a shipment-based company whose sole purpose is to transport a product from 'A' location to 'B' location.

3. Methodology

ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models. An ARIMA model is characterized by 3 terms: p, d, q where,

'p' is the order of the AR term.

'q' is the order of the MA term.

'd' is the number of differencing required to make the time series stationary.

If a time series, has seasonal patterns, then you need to add seasonal terms and it becomes SARIMA, short for 'Seasonal ARIMA' [2]. The term 'Auto Regressive' in ARIMA means it is a linear regression model that uses its own lags as predictors. Linear regression models, work best when the predictors are not correlated and are independent of each other.To make a series stationary the most common approach is to difference it. That is, subtract the previous value from the current value. Sometimes, depending on the complexity of the series, more than one differencing may be needed. The value of d, therefore, is the minimum number of differencing needed to make the series stationary. And if the timeseries is already stationary, then d = 0.

The SARIMA model built here is good.But for the sake of completeness, let's try and force an external predictor, also called, 'exogenous variable' into the model [3][4]. This model is called the SARIMAX model. The only requirement to use an exogenous variable is you need to know the value of the variable during the forecast period as well.

4. Implementation

Some distinguishable patterns appear in the results. Starting from the very first, we get the overall demand of the products category wise, ranging from 1 to 28.

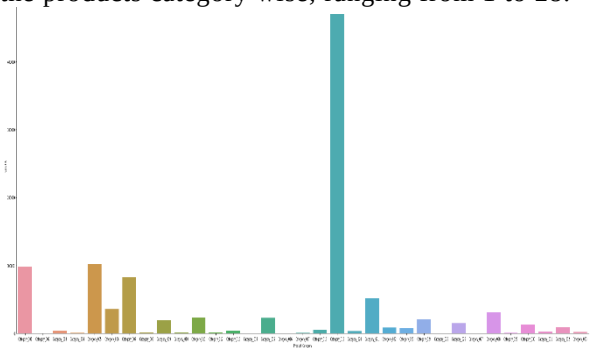

Figure 1: Category wise product count

Next, we need to figure out how much demand is coming from the customers for each of the warehouse that is delivering the products.

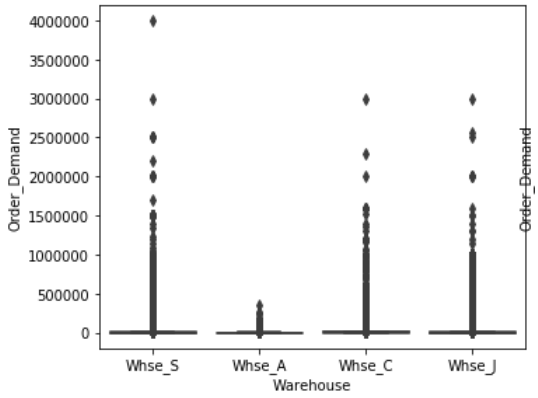

Figure 2: Demand Distribution

Now after acquiring the above information, we can now plot the information for the past values and study the overall trend by mapping those values.

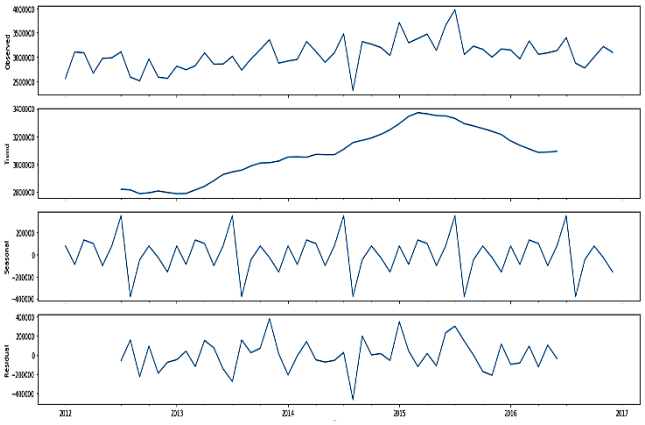

Figure 3: Trend Analysis

The above graphs show the overall observed changes over the period of four years.

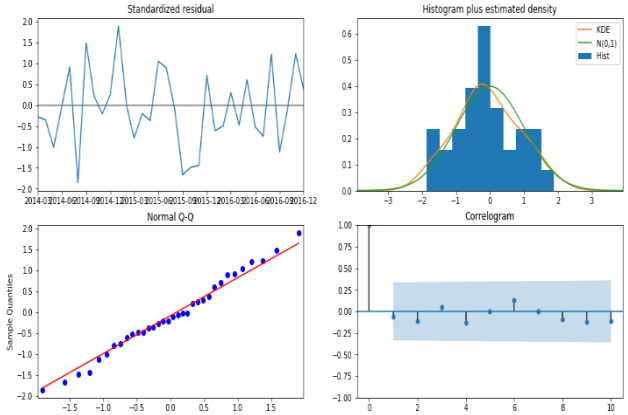

Figure 4: Standard Deviation and Histogram for estimated density

Above presented is the standard deviation observed, which gives us the data response and fitted response.

Now the following graphs presented is the output of all the analysis conducted till now. The forecasting of the demand changes that can be observed in the upcoming years and how will the overall sales go through a transition can be observed and studied from the presented output graphs using the SARIMA model in time series analysis.

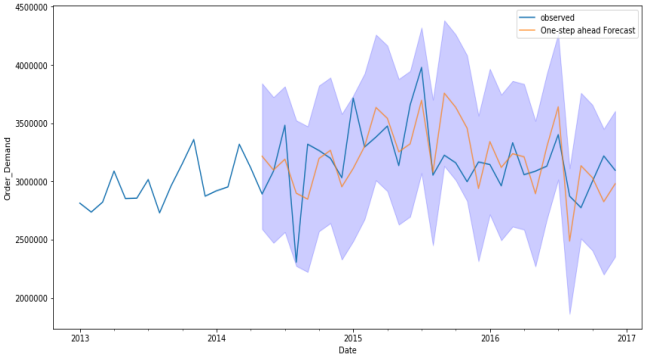

Figure 5: Observed trend vs One step ahead forecasting.

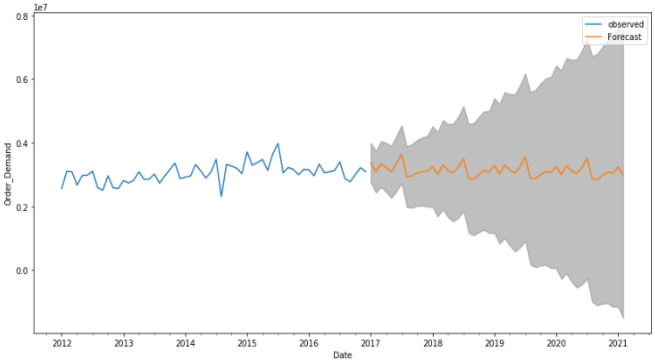

Figure 6: Observed trend vs forecast for the next 5 years.

5. Conclusion

By using the time series forecasting, I been able to construct an ARIMA model to gauge the overall tendencies in the product sales domain by analyzing the sales from 2012 to 2016, and step by step developing a model to predict the behavior of the market in the upcoming five years that is 2017 to 2021. These kinds of studies not only will help the companies to gauge the strength of their product but also will be able predict when if the product will fail or succeed

in generating a revenue and thus can withdraw it before causing a huge loss. Higher accuracy results can be obtained by getting more precise samples and lowering the time intervals in the algorithm for better fortification.

References

[1] Chen, S., et al. "The time series forecasting: from the aspect of network." arXiv preprint arXiv:1403.1713 (2014).
[2] Devi, B. Uma, D. Sundar, and P. Alli. "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50."
[3] Box, George EP, and George C. Tiao. "Intervention analysis with applications to economic and environmental problems." Journal of the American Statistical Association 70.349 (1975): 70-79.
[4] L---Stern Group Ly Pham, Time Series Analysis with ARIMA – ARCH/GARCH model in R
[5] Anderson, David Raymond. Model based inference in the life sciences: a primer on evidence. New York: Springer, 2008.
[6] Chatfield, Chris. The analysis of time series: an introduction. CRC press, 2013.