A Brief Introduction to Demand Forecasting using ARIMA models

Nadia Tarannum J[1], Sri Vidya M S[2]

[1]*Student, Dept. of Computer Science & Engineering, Rashtreeya Vidyalaya College of Engineering, Karnataka, India*
[2]*Assistant Professor, Dept. of Computer Science & Engineering, Rashtreeya Vidyalaya College of Engineering, Karnataka, India*

---***---

Abstract - *Demand forecasting, a field of predictive analysis attempts to understand and estimate the amount of commodities a customer will need in the foreseeable future. It involves the use of data, particularly historical data for sales along with statistical estimates obtained from the test markets. Inventory management, production, planning or deciding to enter a new market is some of the major area of uses for demand forecasting. Proper demand forecasting helps a business by providing valuable information about their potential in the current market helping the management make better informed decisions for the company. In this paper, we will briefly discuss and understand the features of a given time series data used by these models while focusing on the working of ARIMA (Auto Regressive Integrated Moving Average) which is among the many designed models exercised to forecast the future demand of different commodities along with a literature review of the same. We also discuss the mathematical equation and explanation of each of the terms of the ARIMA model alongside discussing the two types of models which are Seasonal and Non Seasonal ARIMA models. Towards the end, we discuss on the final steps on how to select from the above two models with the data we have to work with at a given point in time.*

Key Words: future, demand, forecasting, ARIMA, time series, model

1. INTRODUCTION

Predicting the future with absolute certainty cannot be done by anyone. But in this time and generation it is best advised to be prepared for the future or at least have an estimate of it. The concept of forecasting makes predictions by using historical data to make informed estimates, which helps in determining the direction in which the future trends are going to be. Similarly, demand forecasting when done accurately can be very essential to a firm as it would allow them to produce just the right quantity at the right time thus saving on raw materials, labour, building costs, etc. Efficient and effective planning can be achieved with the help of forecasting. Industrially advanced countries where the demand conditions are so uncertain that the supply conditions are most popularly known to use demand forecasting in contrast to developing countries where supply is often the limiting factor. Better accuracy in forecasting can be achieved by using both quantitative and qualitative methods Forecasts are frequently expressed in language that are applicable to capability management, production, and SC planning. The goal is to comprehend the difference between the plan and actual data, and not just be as accurate as possible. This will make it easier to improve future forecasts and estimate how well foretelling approaches operate. Additionally, forecasts are made for the near future to help in the long term. Shorter forecasts, on the whole, are rather accurate. There are a number of tools that can be used to forecast demand:

- By analysing time series data,
- Analysis through regression.
- Estimates given by experts

For example, time series analysis, such as exponential smoothing, uses retail data from the past to forecast future demand.

Regression analysis is carried out to determine the degree of dependence between two variables.

A chemical analysis method that is based on historical demand and expert judgments on future development could be used to create an expert estimate.

This document is template. We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace (copy-paste) the content with your own material. Number the reference items consecutively in square brackets (e.g. [1]). However the authors name can be used along with the reference number in the running text. The order of reference in the running text should match with the list of references at the end of the paper.

2. Literature Survey

Zhang [1] provides a hybrid forecasting model which over here is used for nonlinear statistics in his article. This model combines ARIMA with genetic programming (GP), allowing it to outperform both ANN and ARIMA forecasting models. He decided to give it a try since ARIMA is frequently used to obtain a highly accurate linear forecasting model, although it cannot anticipate nonlinear statistics accurately. Similarly, for nonlinear statistics, artificial neural networks (ANN) are frequently used to develop more accurate forecasting models than ARIMA, although the meaning of the hidden layers of ANN is difficult to decipher, and it also does not produce a mathematical equation. Towards the end, to demonstrate the usefulness of the suggested forecasting model, certain real data sets were used, which proved to be successful.

In this study, Khalid Yunus et al [2] introduces a modified auto regressive integrated moving average (ARIMA) modelling methodology that may capture time correlation and hence the potential distribution of determined wind-tempo time collection from the data available. The strategy includes frequency decomposition (dividing the wind speed into high-frequency (HF) and low-frequency (LF) components), shifting, and restricting, as well as differencing and electrical transformation, which are all used in a trending ARIMA modelling system.

The paper by Fattah [3] is a major contribution to predicting demand in smaller companies using the statistic approach. They used the ARIMA method to model this problem supported by four performance criterions that they had set which were standard error, maximum likelihood, Akaike criterion and also the Schwarz Bayesian criterion. The ARIMA (1, 0, 1) model was selected and validated with the assistance of another historical demand with similar conditions. The results achieved with this model proved that it may be better to forecast demand for the given problem providing managers reliable guidelines in making decisions. They further planned to create an ARIMA-RBF which is radial basis function combination to attain higher accuracy.

Yi Shan Lee [4] in their paper wanted to beat the restriction of SVM model which does not take into account the time correlation information between distinct data points in statistics, reducing the educational competency of SVM in real-world use. They proposed the Forecasting model called Taylor Expansion as an alternative to the SVM and developed a completely unique hybrid methodology by combining autoregressive integrated moving average and Taylor Expansion Forecasting in order to exploit the great forecasting capacity of financial statistic data with noise. On numerous commodity future prices, both theoretical evidence and empirical results were achieved, demonstrating that the suggested hybrid model enhances predicting accuracy by a significant margin.

Theresa Hoang Diem Ngo et al [5] aimed to replicate a time sequence event as an occupation of its history values in order to help analysts detect the outline with the assumption that the pattern motivation will continue in the future. The autocorrelation function (ACF) and partial autocorrelation function (PACF) have identical behaviors to the other autocorrelation functions, which can be used to identify a suitable statistic replica using the Box-Jenkins style. The research also explores the meaning of the structure estimate, verifies the diagnostics, and validates the forecasts in addition to model recognition.

In their paper Babu [6] & Reddy [7] wanted to show that for forecasting statistic data arising from multiple applications, a mixture of linear and nonlinear models is used to provide a more accurate prediction model than a personal linear or nonlinear model. The models explored during this paper are a replacement for the hybrid ARIMA-ANN model for forecasting, given the time series data. Though prevailing models gave forecasts with better accuracy in comparison to individual models, there was still space for further improvements within the correctness of the character of the given statistic taken under consideration. The suggested hybrid model was tested against individual models and a few existing hybrid models using a simulated data set and experimental data sets. The findings of all of those distinct models demonstrated that the suggested hybrid model of this research has greater prediction accuracy for both one-step-ahead and multistep-ahead forecasts.

Pham [8] wanted to develop a completely unique hybrid forecasting model that included autoregressive integrated moving average ARIMA, artificial neural networks ANNs, and k-means clustering as part of their research. Only these models and k-means clustering are commonly used to create hybrid forecasting models at various levels of complexity in their respective areas. By integrating the weights generated by the discount mean square forecast error approach, the ultimate forecasting value was determined. The suggested model was tested on a few different data sets to see how well it could forecast in three different ways: MSE, MAE, and MAPE. The achieved results proposed that the designed model gave the most nominal performance in MSE, MAE, and MAPE.

Khairalla [9] presented an innovative hybrid method to mix machine learning model (SVM) with statistic model which supported autoregressive moving average (ARIMA). This paper geared towards solving the constraints of this ARIMA model in financial statistic forecasting by joining it with SVM model. Proposed method capitalized on the distinctive power of ARIMA and SVM models in nonlinear and linear modelling. Through the assistance of the outcomes from this model, it had been clearly indicated that the combined model was an efficient attempt to increase forecasting accuracy as compared to when the models were utilized individually.

In this paper by Kardokos [10] , they attempt to compare four realistic methods for forecasting the generation of electricity using grid-connected Photovoltaic (PV) plants, i.e. the Seasonal Autoregressive Integrated Moving Average (SARIMA) modelling, modified SARIMA modelling, as a result of a previous modification of the SARIMA model, SARIMAX modelling (SARIMA modelling with an exogenous factor), and ANN-based modelling. At the conclusion of this comparison, interesting conclusions were gained on the necessity and benefits of using exogenous components in a time series model. To calculate the predicting errors of the SARIMA and SARIMAX models, intra-day prediction updates were implemented. The comparison of these results highlighted variations in accuracy between the two models used.

3. ARIMA Model

Box Jenkins models also called ARIMA, which is the acronym for Auto Regressive Integrated Moving Average uses time series data to predict solutions to the problem at hand. There are three terms which we need to get familiar with. The first two are auto regression and moving average models the combination of which gives us the ARIMA models and finally the time series data it uses. A time series can be defined as data in sequence which is observed over a period of time. The AR component of ARIMA demonstrates that the changing variable of attention is regressed on its own personal lagged (i.e., prior) values. The MA element demonstrates that the regression error is essentially a linear mixture of error components whose values occurred simultaneously and several times in the past. I component shows that information value were changed with the distinction among their values and the preceding values (and this differencing manner can also additionally be carried out multiple times). Each of these aspects has the primary purpose of making the model as accurate as possible. With the use of a single variable described from historical data, ARIMA models can represent both stationary and non-stationary time series data and produce accurate forecasts. This model is in distinct when compared to other as it does not assume any recurring events in the data it is provided with for the forecast but rather aims at describing the autocorrelations in the data. Since it might possibly include autoregressive terms, moving average terms and differencing (a method used for converting non-stationary time series into stationary one) operations we can have various abbreviations

- When the model involves moving average terms only it can be called an MA model and similarly an AR model when it involves only autoregressive terms
- The abbreviation ARMA can be used when there is no differencing involved

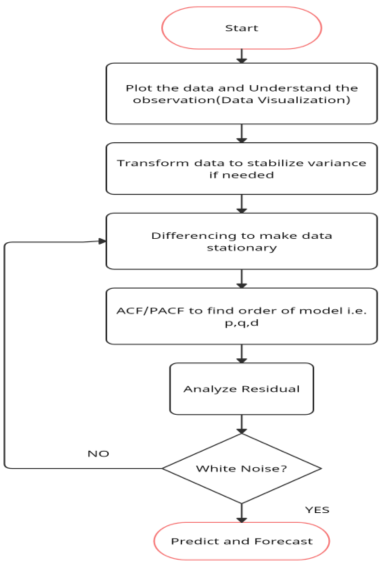

Fig. 1.Basic algorithm of the ARIMA model

Fig. 1 explains the basic algorithm followed by the ARIMA model for prediction which follows the steps as shown in the flow chart above.

3.1 Mathematical Description

The ARMA (p, q) model is given by Eq. 1 for a time series data indicated by X_t, where t is an integer index and X_t are real numbers.

$$Xt - \propto_1 X_{t-1} - \cdots - \propto_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \quad (1)$$

This is equivalent to Eq. 2

$$\left(1 - \sum_{i=1}^{p'} \propto_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \quad (2)$$

Lag operator is denoted by L and $\propto_i$ is a parameter of the autoregressive section of the model, θ_i the parameters of moving average section and ε_t, the term that denotes error. Independent variables are sampled from a normal distribution with zero mean, and error terms are assumed to be identically distributed.

We must now presume that the polynomial $\left(1 - \sum_{i=1}^{p'} \propto_i L^i\right)$ has d as a unit root of multiplicity. Then it can be re- written as Eq.3

$$\left(1 - \sum_{i=1}^{p'} \propto_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \varphi_i L^i\right) (1 - L)^d \quad (3)$$

The ARIMA (*p*, *d*, *q*) process demonstrates the factorization property with $p=p'-d$, and is given by Eq. 4

$$\left(1 - \sum_{i=1}^{p'} \varphi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \quad (4)$$

And hence can be thought as a particular case of an ARMA (*p*+*d*, *q*) containing autoregressive polynomial with *d* unit roots. It can also be simplified as follows in Eq. 5

$$\left(1 - \sum_{i=1}^{p'} \varphi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \quad (5)$$

3.2 Types of ARIMA Models

ARIMA models can be broadly classified into two types:

- Non seasonal ARIMA Models
- Seasonal ARIMA Models

3.2.1 Non seasonal ARIMA model

In this approach, there are three variables to consider in ARIMA (p, d, q) model

p = specifies the periods to lag for e.g. this aids in the adjustment of the line that is being fitted to forecast the current series (If P= 3, the three prior periods of the time series in the autoregressive component of the calculation are used.).

d = Using differencing, we must convert a time series into a series with no trend or seasonality (stationary series). The number of differencing transformations required by the time series to become stationary (where the mean and variance are constant throughout time) is denoted by the letter d. The difference between the current and prior time periods is used as the first differencing value. If the values do not rotate around a constant mean and variance, we proceed to the second differencing with the first differencing's values until a stationary series is obtained

q = The lag of the error component, which is a part of the time series that is not explained by trend or seasonality, is represented by this variable.

3.2.2 Seasonal ARIMA (SARIMA) models:

Time series used when exhibits seasonality, seasonal model comes into the picture. It is similar to the ARIMA models, along with a few other parameters to keep the seasonality in account. SARIMA is depicted as

ARIMA (p, d, q) (P, D, Q) m, where

p — the number of lags

d — differencing degree

q — moving average terms present

m — this is the number of stages in each season

(P, D, Q)— is the (p, d, q) for the seasonal part of the data.

3.3 Autocorrelation function plot (ACF):

The term "autocorrelation" refers to the degree to which a time series is associated with its previous values. The ACF plot is used to examine the relationship between the various points, including the lag unit. The x-axis shows the correlation coefficient, while the y-axis shows the number of delays.

Normally, an ARIMA model either uses the AR or MA expression. It is only on rare occasions that these terms are used together. The ACF plot aids in determining which of these phrases should be used to a given time series. When there is a positive autocorrelation at lag 1, the AR model is utilized; when there is a negative autocorrelation at lag 1, the MA model is utilized. Then a Partial Autocorrelation Function (PACF) is plotted.

3.4 Differencing

Seasonal differencing is a type of differencing that takes into account the seasons and the variances between the current value and the prior season's value. ACF decays slowly in a purely seasonal AR model, while PACF reaches zero. ACF shuts out at zero in a purely seasonal MA model, and vice versa. When seasonal auto-correlation is positive, AR models are employed, and when seasonal auto-correlation is negative, MA models are employed.

4. Final steps on deciding the model

The following steps can be followed to decide on the model that needs to ARIMA model that needs to be built:

•Stationarity Check: If the time series contains a seasonality or trend component, it must be stationary before ARIMA may be used to forecast it.

•Difference Check: As previously stated, if the time series is not stationary, it must be stationarized using differencing.

•Validation sample filtering: This step is used to verify the model's accuracy. To do this, a train test validation split might be employed.

•Choose your AR and MA terms: It must be decided whether to include an AR term(s), MA term(s), or both using the ACF and PACF.

•Create the model: The model must be created, and the forecasting periods must be set to N.

•Test the created model by comparing the predicted values to the actual values in the validation sample.

5. CONCLUSIONS

ARIMA is a popular model in statistics and econometrics for calculating the probability of events occurring over time. Through this paper, we get a brief introduction to what demand forecasting is and how ARIMA model is used to predict the future demand using time series data. A literature survey on the same gives the developments in this field over the past years and the improvements that have been seen by the different proposed models. Seasonal and Non- seasonal ARIMA models have also been discussed along with the steps to identify when a model should be chosen. In addition to the autoregressive, differencing, and average terms for each season, a seasonal model must account for the number of occurrences in each season. ARIMA models can be created and implemented using a variety of software tools, such as Python and others.

ACKNOWLEDGEMENT

I would like to express my gratitude towards Mrs. Sri Vidya M S for guiding me and providing me with valuable feedback throughout my journey along with Dr. Ramakanth Kumar P, Head of the Department, Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering for all his guidance and support.

REFERENCES

[1] Guisheng Zhang, Xindong Zhang, Hongyinping Feng, "Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion", The Journal of Knowledge Engineering Volume 33Issue 5October 2016 pp 501–516

[2] Yi-Shian Lee, Lee-Ing Tong, "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming", Based Systems Volume 24Issue 1February, 2011 pp 66–72

[3] Mergani A. Khairalla, Xu Ning, "Financial Time Series Forecasting Using Hybridized Support Vector Machines and ARIMA Models, Proceedings of the 2017 International Conference on Wireless Communications, Networking and Applications October 2017 Pages 94–98

[4] C. Narendra Babu , B. Eswara Reddy, "A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data", Applied Soft Computing Volume 23October, 2014 pp 27–38

[5] Van-Nam Huynh, Warut Pannakkong, Van-Hai Pham, "A Novel Hybridization of ARIMA, ANN, and K-Means for Time Series Forecasting", International Journal of Knowledge and Systems Science Volume 8Issue 4October 2017 pp 30–53
[6] S. I. Vagropoulos, G. Chouliaras, E. G. Kardakos, C. K. Simoglou, A. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting", Engineering 2016 IEEE International Energy Conference (ENERGYCON),2016
[7] Jamal Fattah1, Latifa Ezzine1, Zineb Aman2, Haj El Moussami2, and Abdeslam Lachhab. "Forecasting of demand using ARIMA model", International Journal of Engineering Business Management, October 2018
[8] W. Jacobs, A. M. Souza and R. R. Zanini – "Combination of Box-Jenkins and MLP/RNA Models for Forecasting Combining Forecasting of Box-Jenkins", IEEE LATIN AMERICA TRANSACTIONS, 14(4), 2016.
[9] Takaomi HIRATA, Takashi KUREMOTO, Masanao OBAYASHI, Shingo MABU, "Time Series Prediction Using DBN and ARIMA", International Conference on Computer Application Technologies, IEEE, 2015.
[10] Kalid Yunus, Torbjorn Thiringer, and Peiyuan Chen, "ARIMA-Based Frequency-Decomposed Modeling of Wind Speed Time Series", IEEE Transaction on Power Systems, 31(4), 2546-2556, 2016.
[11] E. Guresen, G. Kayakutlu, T. U. Daim, "Using Artificial Neural Network Models in Stock Market Prediction", Expert Systems with Applications, 38, 10389-10397, 2011.
[12] T. M. Choi, Y. Yu, K. F. Au, "A Hybrid SARIMA Wavelet Transform Method for Sales Forecasting", Decision Support Systems, 51, 130-140, 2011.
[13] Of Output Growth in a Seven-Country Dataset", Journal of Forecasting, 23, 405-430, 2004.