INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 337-341

e-ISSN : 2583-1062

Impact Factor : 5.725

TIME SERIES PREDICTION OF GINGER PRODUCTION USING ARIMAX AND FEED-FORWARD NEURAL NETWORK REGRESSION

Divyashree[1], Pavana Kumar S. T[2]

[1]Master Scholar, Dept. of Mathematics, Davanagere University, Karnataka

[2]College of Community Science, Central Agricultural University, Tura-794005

Corresponding Email: pvnkmr625@gmail.com

DOI: https://www.doi.org/10.58257/IJPREMS31705

ABSTRACT

The data series tested for stationarity using an augmented dickey-fuller test and data was found non-stationary, and it was analyzed using Autoregressive Integrated Moving Average with explanatory variables model with first differencing, and Feed-Forward Neural Network regression. The ARIMAX model outperformed the feed-forward regression model with leading indicators. The Autoregressive Integrated Moving Average with explanatory variables ARIMAX (1, 1, 1) model predicted the production of Ginger with less root mean square and mean absolute percentage error than the NNAR (1, 2) model, which considers the average of 1000 networks with linear output. The residuals of the fitted models were subjected to the Box-Pierce test and it was found that the residuals from the ARIMAX model were independent while the residuals of NNAR (1, 2) showed dependency over lag. Therefore, the Autoregressive Integrated Moving Average with explanatory variables (ARIMAX) model is the better prediction model over NNAR (1, 2) in terms of performance indicators and residual insignificance.

Keywords: Ginger Production, ARIMAX, Fee-Forward Regression, Prediction, Indicators

1. INTRODUCTION

Ginger (Zingiber officinale) is a herbaceous perennial plant of the family Zingiberaceae, used as a spice, food, flavouring agent, and medicine. Its generic name Zingiber is derived from the Greek zingiberis, which comes from the Sanskrit name of the spice, singabera. The Latin name, Zingiber, means "shaped like a horn" and refers to the roots, which resemble a deer's antlers. Long cultivated by the ancient Chinese and Hindus, Ginger was one of the first oriental spices known in Europe. Throughout the early centuries, Ginger was thought to have medicinal powers. It was often used by pregnant women for morning sickness. The spice has a slightly biting taste and is used, usually dried and ground, to flavour bread, sauces, curry dishes, confections, pickles, and Ginger ale. Its fresh rhizome is used in cooking. It is the most alkaline-promoting food. It has a pH of 5.6 to 5.9, like that of figs, fennel, leeks, parsnips, and romaine lettuce. The alkalinity of foods depends on many variables, including growing conditions and processing.

India is the major producer of ginger having production of 655000 MT of ginger from 132000 ha area under its cultivation (National Horticulture Board, 2014). The productivity of ginger in India is more (3,417 kg/ha) than the average productivity (2,546 kg/ha) in the world. The USA is having the highest (51,925 kg/ha) productivity of ginger in the world. The leading Ginger growing Indian states are Assam, Maharashtra, west Bengal, Gujarat, Kerala, Meghalaya, Mizoram, Karnataka, and Arunachal Pradesh. (Annon.2020). Ginger is a tropical species native to Southeast Asia, belonging to the family Zingiberaceae. The English term ‘ginger’ originated from the Sanskrit word Sringavera. Botanically known as Zingiber Officinale, it is the most popular hot spice in the world. It is marketed in different forms such as raw ginger, dry ginger, bleached dry ginger, ginger powder, ginger oil, ginger oleoresin, gingerale, ginger candy, ginger beer, brined ginger, ginger wine, ginger squash, ginger flakes, etc.

Ginger has a long and well-documented history of both culinary and medicinal use throughout world history, especially in Chinese, Indian, and Japanese medicinal care. In Indian Ayurvedic medicine, ginger is used as an anti-inflammatory (Bode and Dong. 2011). ANN was first developed in the 1940s and the development has experienced a revival with Hopfield’s effort in iterative auto-associable neural networks. ANN has had wide applications in many spheres of life. According to Maier and Dandy, in recent years, Artificial Neural Networks (ANNs) have become extremely popular for prediction and forecasting techniques in a numeral of areas, including water resources, environmental science, power generation, and as well as medicine and finance.

The prediction of response variables using explanatory variables provides better accuracy. The climatic parameters would also be included to understand the functional relationship between the variables. The trend of the response variables and their interrelationship with the explanatory variables also can be studied. The changepoint (Pavan Kumar et al., 2021; Pavan Kumar et al., 2023) in the production of ginger can also be studied for valuable insights in future works to understand the crop in a Broadway and take decisions. Being a high-value crop, Ginger needs intensive management practices starting from rhizome selection to harvest. Important agronomic practices included in ginger cultivation are the selection of good quality disease-free rhizomes, proper rhizome treatment, proper land preparation, and raised beds to provide adequate drainage, planting, weed management along with mulching, irrigation, and nutrient management. The cost of cultivation of ginger is very high. Hence, farmers think twice before going for the cultivation of ginger. The price of ginger is highly volatile in nature it fluctuates based on the quantity of products available in the market and export demand. In this situation, farmers need a forecast of the production of Ginger to decide whether to opt for ginger or not. Hence, in this paper an attempt was made to predict the production of ginger by using explanatory variables area and productivity, using ARIMAX and NNAR models.

2. METHODOLOGY

Data - The data on Ginger Area, Production, and Productivity was collected for 48 years (1970-2017) in India. The techniques like ARIMAX and NNAR models were employed to predict the production of ginger using area and productivity and to evaluate the model based on the accuracy criteria.

ARIMAX model

When an ARIMA model includes other time series as input variables, the model is sometimes referred to as an ARIMAX model i.e., in addition to past values of the response series and past errors, the response series is modeled using the current and past values of the input series. An ARIMAX form of the model is presented as follows: Based on the ARIMA model, the ARIMAX model can take the impact of covariates into account by adding the covariate to the right hand of the ARIMA model equation. The equation of the ARIMAX model is presented as follows:

$$\Phi(B)\Delta^d y_t = \mu + \Theta(B)x_t + \Theta(B)\varepsilon_t$$

Initially, it is required to test the stability of the response series. If the stationary condition is not satisfied, the non-stationary can be removed by an initial differencing step. Calculate the statistics describing the characteristics of the response series, for example, ACF and PACF, to determine the parameters p and q. Estimate the unknown parameters of the model and test the significance as well as the residual series. Conduct the same procedure to the input series as the response series. Estimate the cross-correlation coefficient between the response series and the input series to determine the configuration of the ARIMAX model. Establish diagnostic analysis to verify that the model corresponds to the characteristics of the data. The ARIMAX model concept requires testing of the stationarity of exogenous variables before modeling. The transformed variable is added to the ARIMA model in the second step, in which the lag length r is also estimated. A nonlinear least squares estimation procedure is employed to estimate the parameters of the ARIMAX model (Bierens, 1987). Fortunately, the ARIMAX model can be fitted to data by using a software package, like SAS, MATLAB, EViews, and R. In the present investigation, SAS, Version 9.3 is used for data analysis. The commonly used measures for validation are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Standard Error. In this study, validation of the forecasts was done by computing MAPE (Mean Absolute Percentage Error) for the hold-out data as it is a scale-independent measure. Akaike Information Criteria (AIC) / Schwartz Bayesian Criteria (SBC). The residuals of fitted models were examined for adequacy of the fitted model. The final forecasts were validated by using a lower MAPE value.

NNAR model

Neural Network Autoregression (NNAR) is a model that uses lagged values in the time series data as input for prediction. NNAR (p, k) denotes that the hidden layer has k nodes and p-lagged inputs. It is a deep learning model that simulates the human brain's neural network (Forecasting, 2022). Unlike the earlier two models, which assume that the time series data involves only a linear relationship, the NNAR model can study non-linear relationships between the predictor and predictand. The accuracy measures are obtained based on the above-discussed models and in-sample forecast. The accuracy measures used in this study are Root Mean Square Logarithmic Error (RMSLE), Mean Absolute Square Error (MASE), and Mean Absolute Percentage Error (MAPE). Next, the best model is chosen based on these accuracy metrics and model assumptions. The extent of land planted with rice in India during the next five years is forecast using the most accurate model.

3. RESULTS AND DISCUSSION

The results of the stationarity test are presented in Table 1. Results revealed that the data shows nonstationary with unit root ($p=0.99$) according to adf test. Hence the data was transformed to stationary by taking differencing. To predict the production of ginger, area, and productivity were used as explanatory variables. Model ARIMAX (1,1,1) for prediction of production with area and productivity was found to be a good fit with better accuracy measures. Model parameters ar (1), ma (1), area, and productivity were found significant (Table. 2) at $p<0.001$ level.

Table 1: ADF Test for Stationary

```
Augmented Dickey-Fuller Test

data:  data_1
Dickey-Fuller = 1.0018, Lag order = 3, p-value = 0.99
alternative hypothesis: stationary
```

Table 2: Coefficient tests

```
z test of coefficients:

              Estimate Std. Error z value  Pr(>|z|)
ar1            0.89749    0.22177  4.0469 5.191e-05 ***
ma1           -0.81400    0.25088 -3.2445  0.001176 **
Area           3.99916    0.36608 10.9242 < 2.2e-16 ***
Productivity 95.41859   10.05843  9.4864 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model exhibited a better accuracy measures RMSE (27.62), MAE (14.75), and MAPE (7.36) (Table 3). The coeffi cients of the model were associated with less standard error i.e., ar(1) 0.22, ma (1) 0.25, area (0.36) and productivity (1 0.05). Moving average MA (1) coefficient was found negative. The residuals tested with Box-Pierce test and found no n-significant with significance level p=0.9858 (Table 4). The Akaike information criteria exhibited 456.45 and has neg ative log likelihood -223.22for the ARIMAX model.

Table 3: Summary of Model Fit

```
Coefficients:
         ar1      ma1    Area  Productivity
      0.8975  -0.8140  3.9992       95.4186
s.e.  0.2218   0.2509  0.3661       10.0584

sigma^2 estimated as 779.6:  log likelihood = -223.22,  aic = 456.45

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE
Training set 1.191107 27.62965 14.75557 -2.653598 7.366737 0.3479213
                    ACF1
Training set -0.01483451
```

Table 4: Box-Pierce test for Residuals Autocorrelation

```
Box-Pierce test
X-squared = 0.14595, df = 3, p-value = 0.9858
```

The results from the table 5 & 6 reveals that, the NNAR (1,2) model was fitted to the data with 1000 epochs and decay 0.5. The model 3-2-1 was fitted with 11 weights with linear output units. The model exhibits accuracy criteria such as RMSE (52.06), MAE (27.20), MAPE (9.17). The accuracy criteria of the model NNAR (1,2) was found inferior to that of ARIMAX model. The residuals also fail to explain the independence as per Box-Pierce test (p<0.001), found highly significant (Table 7).

Table 5: Model Summary

```
Model:  NNAR(1,2)
Average of 1000 networks, each of which is
a 3-2-1 network with 11 weights
options were - linear output units  decay=0.5
sigma^2 estimated as 2710
```

Table 6: Model Accuracy of NNAR (1, 2)

```
                   ME     RMSE      MAE       MPE     MAPE      MASE
Training set 1.139739 52.06213 27.20571 -5.406668 9.177011 0.6414832
                  ACF1
Training set 0.5454846
```

Table 7: Box-Pierce test for Residual Autocorrelation-

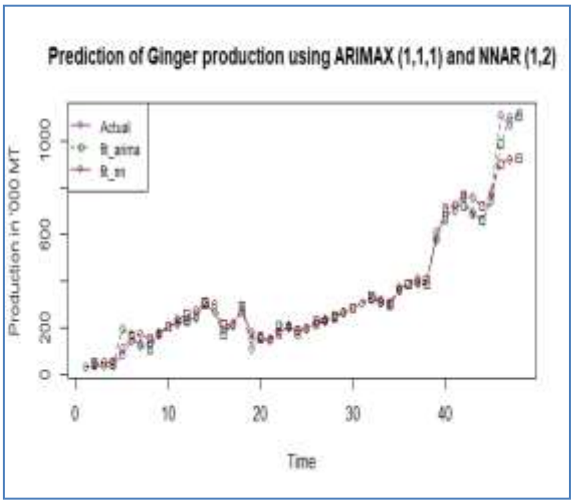
```
Box-Pierce test

data:  residuals(fit_nn)
X-squared = 13.985, df = 1, p-value = 0.0001843
```

e-ISSN : 2583-1062

Impact Factor : 5.725

Vol. 03, Issue 06, June 2023, pp : 337-341



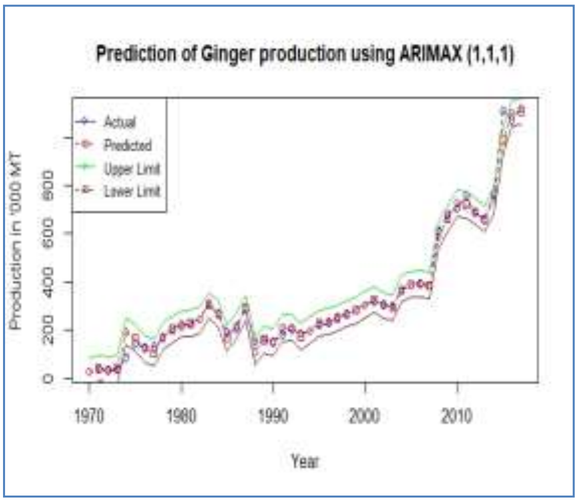


Fig 1: Prediction of ginger production using ARIMAX (1,1,1) & NNAR (1,2) model

Fig 2: Interval Prediction of ginger production using ARIMAX (1,1,1)

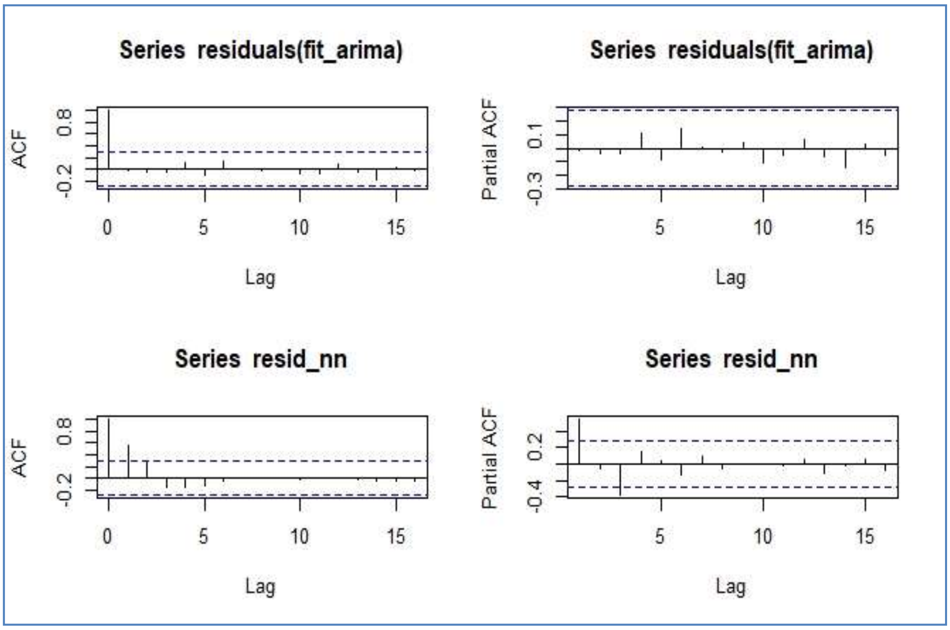


Fig 3: ACF and PACF plots for the residuals of ARIMAX (1,1,1) and NNAR(1,2)

The plot (Fig 1) shows that, the observed values of production of ginger were closely predicted by the ARIMAX model than the NNAR (1,2) model. The lower and upper limits were predicted with significance level $p<0.05$ for the ARIMAX model (Fig 2). The ACF and PACF plots for ARIMAX model and NNAR models showed that, the plots were within the limit showing normal and independent for ARIMAX model whereas for NNAR model doesn't show normal and independent (Fig 3). Therefore, Autoregressive Integrated Moving Average with explanatory variables (ARIMAX) model considered to be the better prediction model over NNAR (1, 2) in terms of performance indicators and residuals insignificance. (Ahmar et al., 2023).

4. CONCLUSION

This study was conducted to build prediction model for the production of ginger in India by utilising 48-year time series data with the explanatory variables. The predictions are based on the data collected about the variable from 1970 to 2017. The data are first plotted to illustrate the prediction performances of ARIMAX and NNAR model. Then the Augmented Dickey–Fuller test is applied to check for non-stationarity of the data. Hence, differencing is carried out. The data becomes stationary after the first differencing. After testing for accuracy, ARIMAX (1,1,1) is the best method to forecast the target variable compared to NNAR (1,2) model.

5. REFERENCES

[1] Ahmar, A. S Singh, P. K Ruliana, R Pandey, A K Gupta, S. Comparison of ARIMA, Sutte ARIMA, and Holt Winters, and NNAR Models to Predict Food Grain inIndia.Forecasting2023,5,138–152. https://doi.org/10.3390/ forecast5010006

Vol. 03, Issue 06, June 2023, pp : 337-341

e-ISSN : 2583-1062

Impact Factor : 5.725

[2] Bierens H J 1987. ARMAX model specification testing, with an application to unemployment in the Netherlands. Journal of Econometrics, 35: 161-190.

[3] Bode AM, Dong Z. (2011) The Amazing and Mighty Ginger. In: Benzie IFF, Wachtel- Galor S, editors. Herbal Medicine: Biomolecular and Clinical Aspects. 2nd edition. Boca Raton (FL): CRC Press/Taylor & Francis; 2011. Chapter 7. Available from: https://www.ncbi.nlm.nih.gov/books/NBK92775/

[4] Forecasting: Principles and practice (2nd ed). 11.3 Neural network models. (n.d.). Retrieved October 13, 2022, from https:// otexts.com/ fpp2/ nnetar. html

[5] National Horticulture Board (2014). Area, production statistics, retrieved from http://nhb.gov.in/area%20_production.html

[6] Pavankumar, S.T., Lahiri, B. and Alvarado, R. (2021) Estimation of trends in Covid- 19 Infections and deaths across WHO regions and India: Multiple change point analysis approach. Spat. Stat. 49: 100538. https://doi.org/10.1016/j.spasta.2021.100538

[7] S.T. Pavan Kumar, Biswajit Lahiri, M.M. Nageswararao, Rafael Alvarado, Silkame N. Sangma, (2023) Trend analysis and changepoint detection of monthly, seasonal and annual climatic parameters in the Garo Hills of Northeast India, Ecological Informatics, Volume 75, 2023, 102104, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2023.102104.