Journal of Agriculture, Biology and Applied Statistics
Vol. 3, No. 2, 2024, pp. 57-68

URL : www.arfjournals.com
https://DOI:10.47509/JABAS.2024.v03i02.02

Forecasting Wheat Crop Area, Production & Productivity in India Using Arimax Model

S. R. Krishna Priya[1] and N. Kausalya[2]
[1]*Assistant Professor, Department of Statistics, PSG College of Arts & Science, Coimbatore*
[2]*Assistant Professor, Department of Statistics, Kristu Jayanti College, Bengaluru*

To cite this article

S.R. Krishna Priya & N. Kausalya (2024). Forecasting Wheat Crop Area, Production & Productivity in India using Arimax Model. Vol. 3, No. 2, pp. 57-68. *https://DOI:10.47509/JABAS.2024.v03i02.02*

Abstract: The Time Series approach has been used in this study to assess the Area, Production, and Productivity statistics of wheat for India. Annual data on Wheat Area, Production & Productivity for a period of 68 years (from 1951 to 2018) has been utilized for the study. The results have been predicted using the Box-Jenkins ARIMA approach and ARIMAX. The appropriate models have been selected based on the low Akaike Information Criteria (AIC) and error measures like Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE) have been used to assess the validity of the models. The results reveal that ARIMAX model with Temperature (1,1,1) & (0,1,1) is most suitable for Wheat Area & Production. For Wheat Productivity data, ARIMAX with Rainfall & Temperature (1,1,1) is found to be most appropriate.

Keywords: ARIMA, ARIMAX, MAE, MSE, MAPE, AIC

1. INTRODUCTION

Wheat is an important crop for food as well as a valuable commodity in international trade. Throughout history, the cultivation and commerce of this plant have shaped economies, communities, and cultures. Many parts of the world's culinary heritage are rich in traditions and recipes made using wheat. A sizable section of India's population relies primarily on wheat. For millions of people, mainly in northern and central India, it is their main source of sustenance. India's Agricultural economy is greatly influenced by wheat farming. In India, wheat is grown in a variety of agro climatic zones, from the dry regions of Rajasthan and Uttar Pradesh to the fertile plains of Punjab and Haryana. By diversifying, wheat agriculture is become more resilient to regional hazards. Wheat is essential to maintaining rural livelihoods, food security, and economic stability in India.

Article History Received : 19 March 2024; Revised : 17 April 2024; Accepted : 27 April 2024; Published : 30 September 2024

Previous attempts were made for forecasting purpose using ARIMA and ARIMAX models, by (Alkali *et al.*, 2019, Ray & Bhattacharya. 2020, Tamuke *et al.* 2018, Anggraeni *et al.* 2015, Andrews *et al.*, 2013, Peter & Silvia. 2012). Various efforts have been made to forecast the production and prices using ARIMAX models such as (Alharbi & Csala. 2022, Neong *et al.*, 2022, Ugoh *et al.*, 2021, Gopinath *et al.*, 2019, Kongcharoen & Kruangpradit. 2013, Wongssanao & Chaovanpoonphol. 2012). ARIMAX models have been used to forecast the currency (Ahmae. 2015, Nasiru *et al.*, 2013). ARIMAX Models have been used to forecast the diseases and industry areas (Aji *et al.*, 2021, Chi *et al.*, 2019, Shilpa & Sheshadri. 2019, Wangdi *et al.*, 2010, Fan *et al.*, 2009). For predicting weather variables ARIMAX model plays an important role (Wanishsakpone & Oswusu. 2020, Zhao *et al.*, 2015).

2. OBJECTIVES OF THE STUDY

The main objective of the present study is to forecast the Wheat crop Area, Production & Productivity in India using ARIMAX model.

3. REASONS TO CHOOSE ARIMAX MODEL

ARIMAX (Auto Regressive Integrated Moving Average with eXogenous variables) models expands on the classic ARIMA model by adding exogenous parameters that exists, outside variables that may have an impact on the variable are attempting to forecast. Because of its adaptability, ARIMAX models can handle a wide range of data types, including time series data with trends, seasonality, and other patterns. ARIMAX models allow you to control for the confounding variables by including them in the model, thus providing more accurate estimates of the relationship between the predictor and the response variable. The validity and reliability of ARIMAX models are enhanced by diagnostic techniques for model evaluation including residual analysis and model selection criteria like AIC (Akaike Information Criterion).

4. MATERIALS & METHODS

4.1. Source of Data

The information is based on annual data for Wheat crop Area, Production and Productivity and the Weather Variables such as Rainfall & Temperature in India between the years 1951 to 2018 for the period of 68 years have been collected from the indiastat website.

4.2. ARIMA Model

In an ARIMA *(p, d, q)* model, the AR, I, and MA components are combined to form,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots \phi_p Y_{t-p} + \theta_{1\epsilon_{t-1}} + \theta_{2\epsilon_{t-2}} + \ldots + \theta_{q\epsilon_{t-1}} \qquad (1)$$

Y_t is the differenced series

$\phi_1, \phi_2, \ldots \phi_p$ are the autoregressive coefficients

θ_1, θ_2,... θ_q are the moving average coefficients.

4.3. ARIMAX Model

The ARIMAX model is an extension of the ARIMA model that incorporates one or more exogenous variables *(X)* that are thought to have an impact on the dependent variable *(Y)*.

The ARIMAX model is expressed mathematically as,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_{1\epsilon_{t-1}} + \theta_{2\epsilon_{t-2}} + \ldots + \theta_{q\epsilon_{t-q}} + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_k X_{t-k} + \epsilon_t \quad (2)$$

Where Y_t is the dependent variable
X_{t-1}, X_{t-2} …, X_{t-k} are exogenous variables at time
β_1, β_2, ..., β_k are coefficients associated with the exogenous variables
c is a constant term
ϕ_1, ϕ_2, ... ϕ_p are the autoregressive coefficients
θ_1, θ_2, ... θ_q are the moving average coefficients
ϵ_t is error term at time t.

Methods like maximum likelihood estimation (MLE) and least squares estimation are used to estimate the parameters of the ARIMAX model. Forecasts for future time points can be produced using the ARIMAX model, which combines the exogenous, moving average, and autoregressive components, once the model parameters have been calculated.

4.4. Goodness of fit

The performance of the ARIMAX model can be authorise using the error metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). It is important to perform model diagnostics, such as goodness of fit tests and residual analysis, to make sure the model accurately reflects the underlying patterns in the data.

Mean Absolute Error (MAE)

A statistic called Mean Absolute Error (MAE) is used to calculate the average absolute differences between a set of data's actual and anticipated values. It is frequently employed to assess regression model's performance.

The following is the formula for MAE,

$$MAE = \frac{1}{n} \Sigma_{i=1}^{n} \mid y_i - \hat{y}_i \mid \quad (3)$$

Where, n is the number of data points.

y_t is the actual value of the i-th data point.

$\hat{y}_i$ is the predicted value of the *i*-th data point.

Mean Squared Error (MSE)

A popular metric for calculating the average squared discrepancies between actual and anticipated values in a set of data is the mean squared error, or MSE. Regression analysis frequently uses it to evaluate a predictive model's accuracy.

The following is the formula for MSE,

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Where, *n* is the number of data points.

y_t is the actual value of the *i*-th data point.

$\hat{y}_i$ is the predicted value of the *i*-th data point.

Mean Absolute Percentage Error (MAPE)

An indicator of a forecasting or prediction model's accuracy is the Mean Absolute Percentage Error, or MAPE. The average percentage difference between the expected and actual values is computed.

The following is the formula for MAPE,

$$MAPE = \frac{1}{n}\Sigma_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \tag{5}$$

Where, *n* is the number of data points.

y_i is the actual value of the *i*-th data point.

$\hat{y}_i$ is the predicted value of the *i*-th data point.

5. RESULTS & DISCUSSION

5.1. ARIMA Model

The data set of Wheat Area, Productivity, and Production is not stationary. Apply the Unit Root test to determine if the data is stationary or non-stationary. The Augmented Dickey Fuller Test (ADF) is used in unit root testing. Except for lags, every variable has unit roots. ACF and PACF were also tested before and after differencing, to determine the stationary mode. Although the interest rate test statistic is around the critical value, the unit root null hypothesis cannot be ruled out. All of the variables difference logs are taken in order to render the variables stationary. Rerunning the ADF test shows that all variables, including Wheat Area, Production, and Productivity, are now stationary in all three sets of data.

The lowest AIC score, 2.600, is reported in the results. Out of all the options that the ARIMA models created, ARIMA (0,1,0) represents the best outcome, based on the ACF plot, PACF plot and number of differencing. Using the same methodology, the forecasted Production and Productivity of Wheat crop were determined. The results show that the Production and Productivity had the lowest AIC scores, of 12.378 and 5.341 respectively. These values indicate the best result that the ARIMA model could have produced. ARIMA (0,1,1) and ARIMA (0,1,1) are determined to be the most suitable models, based on the number of differencing, ACF and PACF plots.

Table 1: Parameter Estimates for ARIMA model

		Wheat Area		
Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMA (0,1,0)	0.295	0.106	2.789	0.006
		Wheat Production		
Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMA (0,1,1)	1.392	0.268	5.195	0.000
MA (1)	-0.374	0.132	-2.827	0.006
σ^2 value	11.156	1.700	6.561	0.000
		Wheat Productivity		
Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMA (0,1,1)	39.286	8.231	4.772	0.000
MA (1)	-0.460	0.158	-2.900	0.005
σ^2 value	12678.13	1724.27	7.352	0.000

Table 2: Error Measures for ARIMA model

Criteria	*AIC*	*MAE*	*MSE*	*MAPE*
Area	2.600	0.591	0.340	1.870
Production	5.341	2.785	7.759	2.561
Productivity	12.378	78.572	6173.559	2.211

4.2. ARIMAX model fitting for Wheat Area, Production & Productivity

Finding an appropriate ARIMA model for the endogenous variable is the first stage in creating an ARIMAX model. Before modelling, the exogenous variable must be stationary in order to apply the ARIMAX model. Using the best weather contributor's temperature and rainfall, taking into account the non-stationary behaviour of the series under study, ARIMAX models were attempted to increase forecast performance. Proceed to ARIMAX model, after stationary mode has been confirmed.

The ARIMAX model has been used to represent the Weather Variables for a variety of combinations, including Wheat Area, Productivity, and Production with Rainfall. Rainfall has an improper impact on Wheat Area, Production & Productivity when it comes to forecasting.

When it comes to the Weather Variable, Temperature, the same process applies. For further forecasting, only the Wheat Area and Wheat Production, in addition to the weather variable temperature, have the appropriate effect. The Wheat Productivity does not has the influence with the weather variable temperature for forecasting.

For Area with Temperature, the ARIMAX approach has 1 and 1 lags. This indicates that there is 1 AR lag and 1 MA lag. It provides minimum AIC score of 2.571. Out of all the options that were generated for the Wheat Area, ARIMAX (1,1,1) represents the best outcome.

The ARIMAX technique has 0 and 1 lags for Wheat Production with Temperature. This shows that there is a single MA lag and no AR lag. According to the data, it offers an AIC score of 5.364. ARIMAX (0,1,1) is the best result out of all the possibilities that were created for the Wheat Production.

The parameter estimates and error measures for the combination of Wheat Area with Temperature and Wheat Production with Temperature are given in the table 2 & 3.

Table 3: Parameter Estimates for ARIMAX model with Temperature

Wheat Area with Temperature				
Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMAX (1,1,1)	4.0854	1.8179	2.2472	0.0282
Temperature	-0.1272	0.0613	-2.0733	0.0423
AR (1)	0.7703	0.1407	5.4718	0.0000
MA (1)	-0.9999	1596.598	-0.0006	0.9995
σ^2 value	0.637658	34.0958	0.0187	00.9851
Wheat Production with Temperature				
Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMAX (0,1,1)	-5.1837	10.40219	-0.4938	0.6200
Temperature	0.2211	0.3458	0.06395	0.5248
MA (1)	-0.3916	0.1330	-2.9428	0.0045
σ^2 value	11.0782	1.7756	6.2388	0.0000

Table 4: Error Measures for ARIMAX model with Temperature

Criteria	*AIC*	*MAE*	*MSE*	*MAPE*
Area with Temperature	2.571	0.202	0.401	0.665
Production with Temperature	5.364	0.199	0.044	0.177

Table 5: ARIMAX model forecasted values for ten years from 2019 to 2028 for Wheat Area & Production with Temperature

Years	*ARIMAX (1,1,1,) for Wheat Area with Temperature*	*ARIMAX (0,1,1) for Wheat Production with Temperature*
2019	29.729	100.980
2020	29.876	102.708
2021	30.008	104.449
2022	30.114	106.222
2023	30.217	107.990
2024	30.328	109.739
2025	30.431	111.497
2026	30.519	113.277
2027	30.600	115.064
2028	30.678	116.856

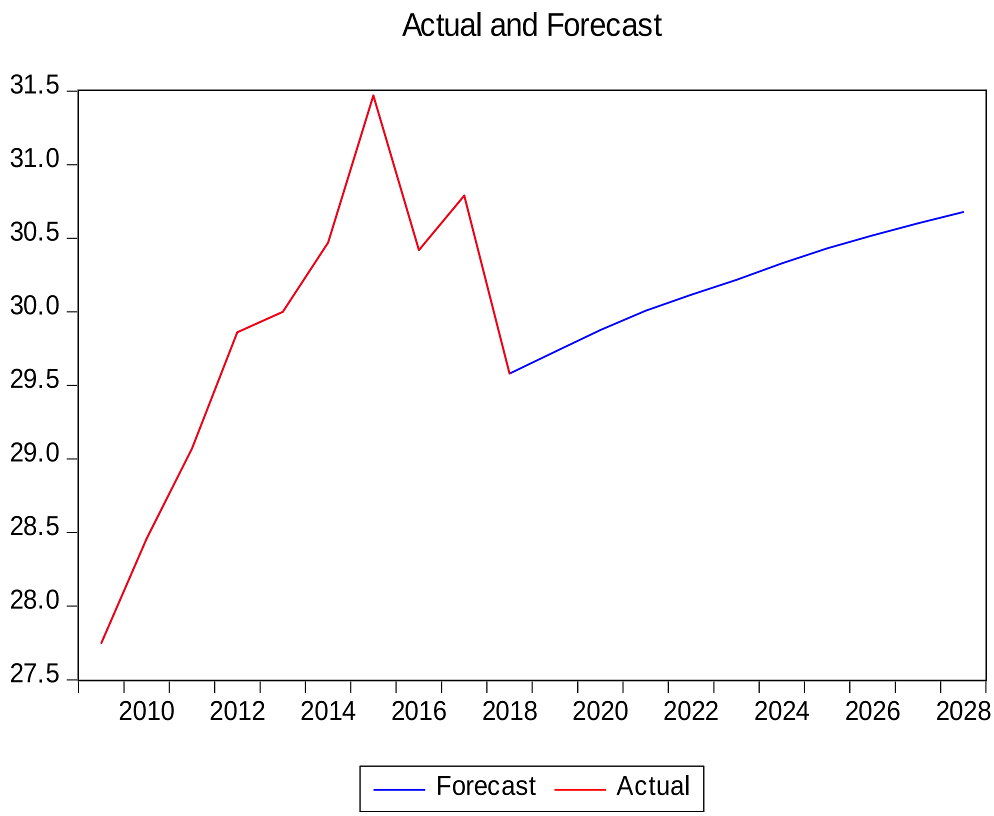

Figure 1: Actual and Forecast graph of Wheat crop Area with Temperature

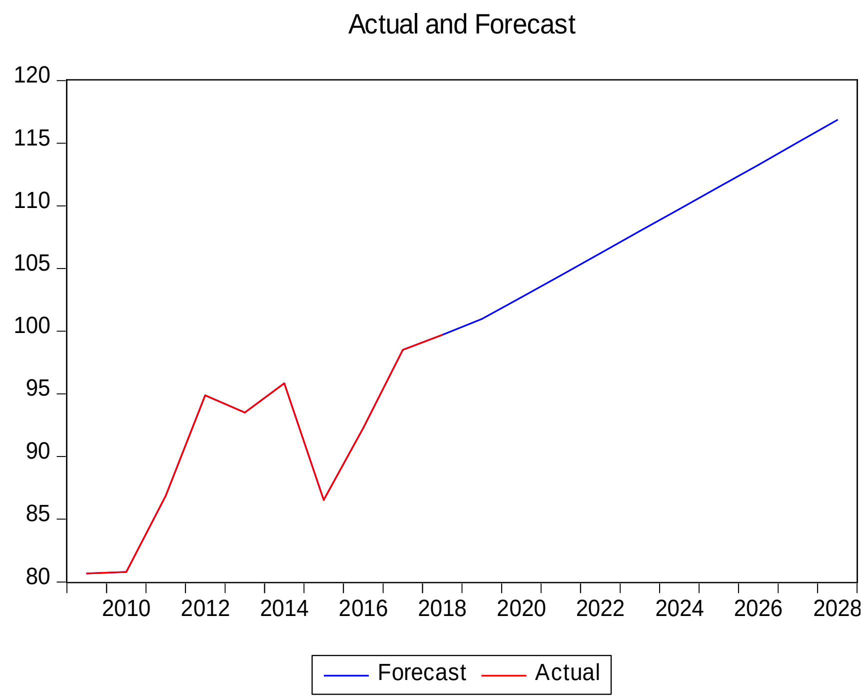

Figure 2: Actual and Forecast graph of Wheat crop Production with Temperature

By considering both the Weather Variables, Rainfall & Temperature, into the consideration, only the Wheat crop Productivity has the good impact for the forecasting. The ARIMAX method has lags of 1 and 1 for Productivity with Rainfall & Temperature. As a result, there is only one MA lag and one AR lag, hence the fitted model is ARIMAX (1,1,1). The findings indicate that it offers an AIC of 12.421. This value indicates the best result among all the choices that were produced for the Wheat Productivity. The parameter estimates and error measures are shown below in the table 5.

Table 6: Parameter Estimates for ARIMAX model with Rainfall & Temperature

Variable	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
ARIMAX (1,1,1)	511.4678	413.2341	1.2377	0.2206
Rainfall	-0.1741	0.1469	-1.1849	0.2406
Temperature	-9.0389	9.6803	0.9337	0.3541
AR (1)	0.3571	0.2856	1.2503	0.2160
MA (1)	-0.7476	0.1841	-4.0602	0.0001
σ^2 value	12082.04	2115.981	5.7098	0.0000

Table 7: Error Measures for ARIMAX model with Rainfall & Temperature

Criteria	*AIC*	*MAE*	*MSE*	*MAPE*
Productivity with Rainfall &Temperature	12.421	47.603	2290.176	1.381

Table 8: ARIMAX forecasted values for ten years from 2019 to 2028 for Wheat Productivity with Rainfall & Temperature

Years	*ARIMAX (1,1,1) for Wheat Productivity with Rainfall & Temperature*
2019	3322.593
2020	3326.871
2021	3346.350
2022	3368.326
2023	3392.666
2024	3418.608
2025	3444.491
2026	3469.548
2027	3494.333
2028	3518.948

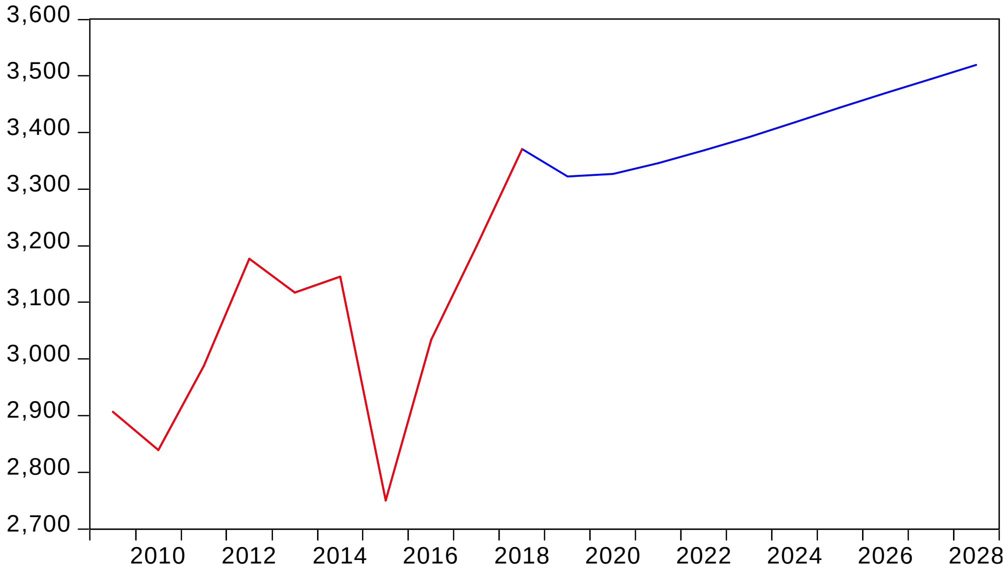

Figure 3: Actual and Forecast graph of Wheat Productivity with Rainfall & Temperature

4.3. Residual Values of ARIMAX Model for Wheat Area, Production & Productivity

Using ARIMAX model, the forecasted values have been calculated for Area, Production and Productivity of Wheat crop in India. The Error Values for Area, Production with Temperature and Productivity with Rainfall & Temperature of Wheat crop in India have been calculated.

Table 8: Residual Values for ARIMAX model

Variables	*Criteria*	*ARIMAX Model*
Area (Temperature)	MAE	0.2022
	MSE	0.4015
	MAPE	0.6655
Production (Temperature)	MAE	0.1997
	MSE	0.0444
	MAPE	0.1773
Productivity (Rainfall & Temperature)	MAE	47.6034
	MSE	2290.176
	MAPE	1.381

The table 8, represents the error values (MAE, MSE and MAPE) for ARIMAX. From comparing two models for Area, Production and Productivity of annual Wheat crop in India, ARIMAX model having the least error values. Hence, it is considered, ARIMAX model is the most suitable model for forecasting the Wheat Area, Production & Productivity in India.

5. CONCLUSION

The ARIMAX model is found to have the lowest error values when comparing with ARIMA model using error metrics like Mean Absolute Deviation (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE). For the present dataset, the ARIMAX model performs better. The ARIMAX (1,1,1) model is found to be the suitable model for Wheat Area with Temperature after evaluating these two models for Wheat Area. The ARIMAX (0,1,1) model is shown to have the best fit for Wheat Production with Temperature. The ARIMAX (1,1,1,) model is shown to have the best fit for Wheat Productivity when Temperature and Rainfall are taken into consideration.

REFERENCES

Ahmae, I. S., Setiawan, S., Suhartono, S., & Masun, N. H. (2015, December). Forecasting of monthly inflow and outflow currency using time series regression and ARIMAX: The Idul Fitri effect. In *AIP Conference Proceedings* (Vol. 1691, No. 1). AIP Publishing.

Aji, B. S., & Rohmawati, A. A. (2021, August). Forecasting number of COVID-19 cases in Indonesia with ARIMA and ARIMAX models. In *2021 9th international conference on information and communication technology (ICoICT)* (pp. 71-75). IEEE.

Alharbi, F. R., & Csala, D. (2022). A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. *Inventions*, *7*(4), 94.

Alkali, M. A. (2019). Assessing the forecasting performance of ARIMA and ARIMAX models of residential prices in Abuja Nigeria. *Asia Proceedings of Social Sciences, 4*(1), 4-6.

Andrews, B. H., Dean, M. D., Swain, R., & Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. *Society of Actuaries*, 1-54.

Anggraeni, W., Vinarti, R. A., & Kurniawati, Y. D. (2015). Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study. *Procedia Computer Science*, *72*, 630-637.

Chiu, L. Y., Rustia, D. J. A., Lu, C. Y., & Lin, T. T. (2019). Modelling and forecasting of greenhouse whitefly incidence using time-series and ARIMAX analysis. *IFAC-PapersOnLine*, *52*(30), 196-201.

Fan, J., Shan, R., Cao, X., & Li, P. (2009). The analysis to tertiary-industry with ARIMAX model. *Journal of Mathematics Research*, *1*(2), 156.

Gopinath, M., & Kavithamani, M. (2019). Production of Sugarcane Forecasting using ARIMAXModel. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*, 12.

Kongcharoen, C., & Kruangpradit, T. (2013, June). Autoregressive integrated moving average with explanatory variable (ARIMAX) model for Thailand export. In *33rd International Symposium on Forecasting, South Korea* (pp. 1-8).

Nasiru, S., Luguterah, A., & Anzagra, L. (2013). The efficacy of ARIMAX and SARIMA models in predicting monthly currency in circulation in Ghana. *Mathematical Theory and Modeling*, *3*(5), 73-81.

Neog, B., Gogoi, B., & Patowary, A. N. (2022). Development of Hybrid Time Series Models for Forecasting Autumn Rice Using ARIMAX-ANN and ARIMAX-SVM. *Ann. For. Res*, *65*(1), 9119-9133.

Peter, Ï., & Silvia, P. (2012, September). ARIMA vs. ARIMAX–which approach is better to analyze and forecast macroeconomic time series. In *Proceedings of 30th international conference mathematical methods in economics* (Vol. 2, pp. 136-140).

Ray, S., & Bhattacharyya, B. (2020). Statistical modeling and forecasting of ARIMA and ARIMAX models for food grains production and net availability of India. *Journal of Experimental Biology and Agricultural Sciences*, *8*(3), 296-309.

Shilpa, G. N., & Sheshadri, G. S. (2019). ARIMAX model for short-term electrical load forecasting. *Int. J. Recent Technol. Eng.(IJRTE)*, *8*(4).

Tamuke, E., Jackson, E. A., & Sillah, A. (2018). Forecasting inflation in Sierra Leone using ARIMA and ARIMAX: A comparative evaluation. Model building and analysis team. *Theoretical and Practical Research in the Economic Fields*, *9*(1), 63-74.

Ugoh, C. I., Uzuke, C. A., & Ugoh, D. O. (2021). Application of ARIMAX Model on Forecasting Nigeria's GDP. *American Journal of Theoretical and Applied Statistics*, *10*(5), 216.

Wanishsakpong, W., & Owusu, B. E. (2020). Optimal time series model for forecasting monthly temperature in the southwestern region of Thailand. *Modeling Earth Systems and Environment*, *6*, 525-532.

Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010). Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. *Malaria Journal*, *9*(1), 1-9.

Wongsanao, T., & Chaovanapoonphol, Y. (2012). Maize price forecasting in northern region of Thailand using ARIMAX model. *Journal of Agricultural Research and Extension*, *29*(2), 58-71.

Zhao, E., Zhao, J., Liu, L., Su, Z., & An, N. (2015). Hybrid wind speed prediction based on a self-adaptive ARIMAX model with an exogenous WRF simulation. *Energies*, *9*(1), 7.