AGRICULTURAL PRODUCTION PATTERNS IN TAMIL NADU: INSIGHTS FROM VECTOR AUTOREGRESSIVE ANALYSIS USING PYTHON PROGRAMMING

R. Kamalanathan[1] A. Sheik Abdullah[1*] and A. Jawahar Farook[2]

•

[1] PG & Research Department of Statistics, Salem Sowdeswari College, Salem-10, India.
[2] Deputy Director of Statistics, Department of Economics and Statistics,
Tamilnadu Government, India.
[1]rknstat1@gmail.com, [1*]sheik.stat@gmail.com, [2]farookstat@gmail.com
Correspondence Email: sheik.stat@gmail.com

Abstract

Understanding agricultural production patterns is crucial for enhancing productivity and ensuring food security. This study explores the dynamics of agricultural production in Tamil Nadu using the Vector Autoregressive (VAR) model to capture the interdependence among various crop yields and rainfall over time. Employing Python programming for data analysis and modeling, the study leverages historical time-series data to identify trends, forecast production, and analyze the impact of external shocks on agricultural outputs. The research incorporates preprocessing techniques to ensure stationarity, optimal lag selection using Akaike's Information Criterion(AIC) and Bayesian Information Criterion(BIC), and diagnostic checks for model accuracy and stability. The findings provide insights into the temporal relationships among various crops and rainfall. Additionally, Impulse Response Functions(IRF) and variance decomposition analyses offer a deeper understanding of how shocks to one variable propagate through the system. The study demonstrates the utility of Python-based VAR models in agricultural forecasting and decision-making, offering policymakers and stakeholders a robust tool to improve resource allocation and agricultural planning in Tamil Nadu. This work highlights the potential of data-driven approaches to address challenges in the agricultural sector effectively.

Keywords: Time Series Analysis, Stationary, Impulse Response Functions, Augmented Dickey-Fuller Test, Granger Causality Test, Vector Autoregressive.

I. Introduction

Agriculture plays a pivotal role in Tamil Nadu's economy, serving as a cornerstone for livelihood, food security, and economic growth. The state's diverse agro-climatic conditions enable the cultivation of various crops, making it a vital contributor to India's agricultural output. Farook and Kannan[8], investigate the influence of climate change on the yields of Kharif and Rabi rice crops, focusing on the impact of maximum temperature, minimum temperature, and rainfall using a VAR model that incorporates Granger causality, IRF, and variance decomposition. Their results reveal significant temperature effects on both types of rice, with rainfall negatively affecting Kharif yields and both maximum temperature and rainfall reducing Rabi yields.

This paper focuses on forecasting the production of major crops, including Paddy, Cumbu, Cholam, Ragi and Maize. These crops not only fulfill the dietary needs of Tamil Nadu's population but also play a pivotal role in the state's agricultural economy and resilience. Their significance lies in their ability to support food security, adapt to diverse climatic conditions, and provide livelihood opportunities to farmers across the state. Promoting sustainable cultivation practices and value addition for these crops can further enhance their role in Tamil Nadu's agricultural development. However, the agricultural sector faces several challenges, including erratic rainfall patterns, fluctuating market dynamics, and the increasing need for efficient resource allocation. Understanding the complex interdependencies among these factors is critical for informed decision-making and sustainable agricultural development.

Modern analytical methods provide powerful tools to examine the interdependencies and dynamics within agricultural systems. VAR analysis stands out as a versatile econometric model for understanding the relationships between multiple time-series variables. The VAR model allows researchers to identify causal relationships, forecast trends, and evaluate the impact of external shocks, making it a powerful tool for analyzing agricultural production patterns. This study applies the VAR model to explore the interdependencies of agricultural production variables in Tamil Nadu, such as crop yields and rainfall. Leveraging the computational power and versatility of Python programming, the research provides a comprehensive analysis of historical data to uncover trends and derive actionable insights. The objective of this research is to forecast agricultural production, understand the influence of key factors, and provide a data-driven foundation for policymaking and resource optimization. By integrating modern computational tools with advanced econometric techniques, this study contributes to enhancing agricultural planning and addressing the challenges of the sector in Tamil Nadu.

Granger[4], introduced testable definitions of causality and feedback using simple two-variable models. He developed a causality test to determine the directional relationships between variables, providing a systematic approach to identify causal links. Additionally, he addressed the critical issue of instantaneous causality, further enriching the understanding of temporal relationships in time series data.

II. Review of Literature

This section reviews and discusses several foundational and relevant papers that provide context and support for this article. Granger and Newbold[5], critically examines the pervasive issue of spurious regression in econometrics, particularly in time series analysis, and its implications for applied research. The authors highlight the frequent reporting of regression equations with high R^2 values but alarmingly low Durbin-Watson statistics, indicating significant autocorrelation in residuals. Despite warnings in econometric textbooks, such errors persist in respected literature, leading to inefficiencies, suboptimal forecasts, and invalid significance tests of coefficients. The findings of Dickey and Fuller[2] hold considerable importance for hypothesis testing and parameter estimation in autoregressive models, particularly in identifying whether a time series is stationary or nonstationary. By offering methods to test the unit root hypothesis ($H_0 : p = 1$), the paper contributes to the broader econometric literature on time series analysis. It underscores the importance of understanding the behavior of estimators in near-nonstationary environments, providing a foundation for more robust inference in such contexts.

Runkle[12], critically evaluates the utility of unrestricted VARs in understanding the interrelationships among key macroeconomic variables such as interest rates, money, prices, and output. The evidence highlights significant limitations in drawing strong conclusions using this approach. Granger[6] explored the intricate connection between causality, statistical methods, and their practical implications, emphasizing the importance of thoughtful evaluation in both theoretical and applied settings. Johansen[9], explores the statistical framework for analyzing nonstationary VAR processes that are integrated of order 1, focusing on cointegration properties. The authors derive the maximum likelihood estimator for the cointegration space and propose a likelihood ratio test to evaluate the dimensionality of the cointegration vectors. Additionally, they develop tests for linear hypotheses about the cointegration vectors.

Barkley et al.,[1] provide an in-depth examination of the institutional framework of the Saudi economy using a VAR model. The study reveals that external variables, such as world inflation rates and Saudi oil policies, play a crucial role in shaping the country's economic outcomes. Waggoner and Zha[15], develops Bayesian methods to compute the exact finite-sample distribution of conditional forecasts in VAR models, addressing parameter uncertainty and expanding their applicability. The study highlights the practical utility of these methods in assessing monetary policy impacts and analyzing scenarios with specific economic conditions. This advancement enhances the reliability of VAR-based macroeconomic forecasts for policy and practical applications.

Stock and Watson[14], analyze the role of VARs in macro-econometrics, emphasizing their effectiveness in data description and forecasting. While VARs excel in capturing dynamic relationships among time series, the study highlights their limitations in structural inference and policy analysis due to the "identification problem," which requires economic theory or institutional insights to resolve. Zivot[16] offers a comprehensive overview of the VAR model, emphasizing its versatility in analyzing multivariate time series, especially in economic and financial contexts. Also primarily focused on VAR for stationary time series, while also previewing its extension to nonstationary series with cointegration. Sasikumar and Sheik[13], provided text highlights the critical role of financial market volatility in shaping investment decisions and regulatory policies. It effectively underscores the significance of time series modeling, particularly in the context of forecasting stock market dynamics. The discussion about exploring the interconnections between variables such as the dollar rate, crude oil, and fuel prices using a VAR model is particularly compelling.

Hamzah et al.,[7] present a thorough application of the VAR model to investigate the export dynamics of Indonesia's major agricultural commodities—coffee beans, cacao beans, and tobacco—over a ten-year period. The study effectively demonstrates the suitability of the VAR model for multivariate time series analysis, particularly for capturing the intricate dynamic relationships between endogenous and exogenous variables. By evaluating VAR models with varying lag structures (VAR(1) to VAR(5)), the researchers employ a rigorous model selection process based on well-established statistical criteria, including AIC, Corrected AIC, Schwarz Bayesian Criterion (SBC), and Hannan-Quinn Information Criterion (HQIC). The selection of the VAR(2) model as the best fit is robustly supported by these criteria, significantly enhancing the reliability and credibility of the findings.

III. Data Source and Basic Statistics

The data for this study were obtained from the official website of the Department of Economics and Statistics, Government of Tamil Nadu, India (https://www.tn.gov.in/crop/stat.html), covering the period from 1990-91 to 2022-23. This source provides reliable and comprehensive statistical information, ensuring the validity and accuracy of the analysis conducted in this research. Table 3.1 presents the basic statistical summary of the dataset used in this study. This summary provides an initial understanding of the data distribution and variability, which is essential for further analysis. Paddy and Ragi are more stable compared to Maize and Cholam, which exhibit higher fluctuations in yield. Cholam's skewness and kurtosis suggest potential outliers or irregular growth conditions.

Table 3.1: *Basic Statistics*

Basic Statistics	Rainfall (in mm)	Paddy (in Tonnes)	Cholam (in Tonnes)	Cumbu (in Tonnes)	Ragi (in Tonnes)	Maize (in Tonnes)
Maximum	1401.1	8141300	868940	296270	362343	2989945
Minimum	598.1	3222776	153856	56505	114429	43820
Mean	973.47	6253034.87	363440.03	155494.72	230889.75	1043979.06
Median	985.8	6610607	345820	146132	227476	759112
Skewness	0.1087	-0.75	1.25	0.48	0.1350	0.69
Kurtosis	-0.6067	-0.19	2.75	-0.65	-1.0332	-1.20
CV	0.19867	0.21404	0.40525	0.40354	0.29949	1.0408
SD	190.4448	1317966.677	145036.983	61790.6097	68093.8147	1070031.06

IV. Methodology

The VAR model is a statistical model used to capture the linear interdependencies among multiple time series. The VAR model is an extension of the univariate autoregressive (AR) model to multivariate time series data, allowing each variable to be a function not only of its past values but also of the past values of all other variables in the system. The following flowchart is representing the procedure of VAR model.

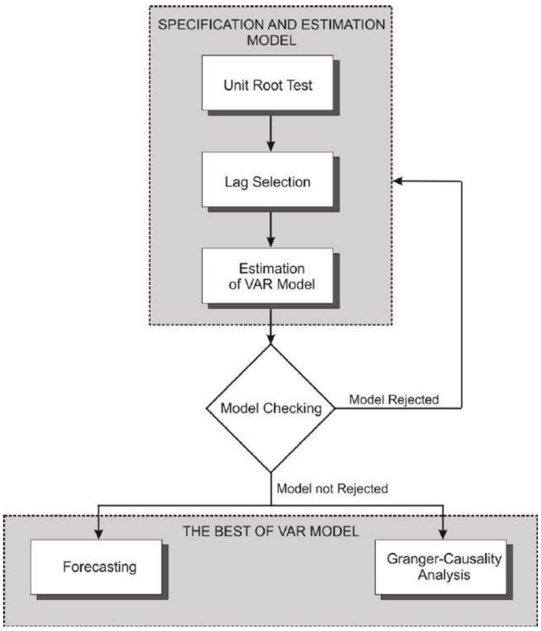

Figure 4.1: *Flow Chart of VAR Analysis Procedure*

4.1. Assumptions of VAR Model

The following are basic assumptions for performing VAR analysis

- *No Multicollinearity* - Variables should not be perfectly correlated.
- *Linearity* - The relationships between variables are assumed to be linear.
- *No Serial Correlation in Residuals* - Check for autocorrelation in residuals after fitting the model using tests like the Ljung-Boxtest.
- *Homoscedasticity* - The variance of residuals should be constant over time.

4.2. General Form of VAR(p) Model

For a system of *k* variables, the VAR(*P*) model is written as

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \in_t \tag{1}$$

where $Y_t = \begin{bmatrix} y_{1t} & y_{2t} & \cdots & y_{kt} \end{bmatrix}'$ vector of k variables at time *t*, A_i is $k \times k$ coefficient matrices for lag *i*, p is number of lags and $\in_t = \begin{bmatrix} \in_{1t} & \in_{2t} & \cdots & \in_{kt} \end{bmatrix}'$ Vector of error terms (innovations), assumed to be white noise, i.e., $\in_t \sim N(0, \Sigma)$.

4.3. VAR(1) Model with Two Variables

Consider a VAR(1) model(first order VAR) with two variables

$$Y_{1t} = \alpha_1 + \phi_{11} Y_{1(t-1)} + \phi_{12} Y_{2(t-1)} + \in_{1t} \tag{2}$$

$$Y_{2t} = \alpha_2 + \phi_{21} Y_{1(t-1)} + \phi_{22} Y_{2(t-1)} + \in_{2t} \tag{3}$$

where α_1 and α_2 are the intercept terms, $\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$ are the coefficient for lagged values and $\in_{1t}, \in_{2t}$ are the error terms for each equation. The above equation (2) and (3) can be expressed in matrix form as

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Y_{1(t-1)} \\ Y_{2(t-1)} \end{bmatrix} + \begin{bmatrix} \in_{1t} \\ \in_{2t} \end{bmatrix} \tag{4}$$

4.4. Granger Causality Test

The Granger causality test evaluates whether past values of one time-series variable (X_t) provide statistically significant information about another variable (Y_t) beyond the information contained in its own past values. This is done by comparing two regression models: a restricted model (without X_t) and an unrestricted model (including X_t). Restricted model(No X_t) is defined as

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \in_t \tag{5}$$

where Y_t is current value of the dependent variable, β_i is coefficients of lagged Y_t, p is number of lags and ϵ_t is error term (white noise). Then the Unrestricted Model (Including X_t) is defined as

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{p} \gamma_i X_{t-i} + \in_t \tag{6}$$

where X_t is lagged values of the independent variable and γ_i coefficients for the lagged X_t. The Granger causality test uses an F-test to compare the restricted and unrestricted models is

$$F = \frac{(RSS_R - RSS_U)/p}{RSS_U/(n-k)} \tag{7}$$

here RSS_R is Residual Sum of Squares from the restricted model, RSS_U is Residual Sum of Squares from the unrestricted model, p is number of lags tested, n number of observations and k is Total number of parameters in the unrestricted model. Decision rule of the above F – Statistic, reject H_0 if the p-value is less than the significance level (0.05) and the rejection indicates that X_t Granger-causes Y_t.

4.5. Impulse Response Functions

Lütkepohl[11], offers an overview of IRFs in VAR models, emphasizing their role in analyzing dynamic relationships and responses to shocks. IRF is a fundamental method in time-series econometrics, widely applied within the framework of VAR models. This analysis is designed to evaluate the impact of an unexpected change, referred to as a shock or impulse, in one variable on the other variables within a system. By tracing the effects of such a shock over subsequent time periods, IRF reveals how interconnected variables respond and adapt dynamically.

This approach is particularly useful for understanding the magnitude, direction, and duration of these interactions, offering a comprehensive view of the causal relationships and feedback mechanisms present in complex systems. The IRF measures the effect of a one-time shock in one variable ($\in_{jt}$) on all variables in the system (Y_{it}) over time. The IRF formula is

$$\psi_h = \frac{\partial Y_{t+h}}{\partial \in_t} \tag{8}$$

4.6. Augmented Dickey-Fuller (ADF) Test

Dickey and Fuller[3], examines the statistical properties of a time series model of the form $Y_t = \alpha + \beta Y_{t-1} + \in_t$, where Y_t is fixed and $\in_t$ are independent and normally distributed random variables with mean 0 and variance σ^2. The focus is on the likelihood ratio test for the joint hypothesis $(\alpha, \beta) = (0, 1)$, which corresponds to a random walk without drift.

ADF test is also known as unit root test. It is a statistical procedure used in time-series analysis to determine whether a time series is stationary or non-stationary. A stationary time series has a constant mean and variance over time, while a non-stationary time series exhibits trends, seasonality, or varying variance. If a time series has a unit root, it means that it is non-stationary and requires differencing to achieve stationarity before applying models like VAR or Auto Regressive Integrated Moving Average (ARIMA). It is sensitive to the choice of lag length (p), which can be selected using criteria like AIC or BIC.

V. Results and Discussions

5.1. ADF Test

In this section, the results of the analysis are presented systematically, focusing on the key findings derived from the data. Before applying the VAR model, it is essential to verify whether the selected data is stationary. The ADF test evaluates this by testing the null hypothesis that the series has a unit root (non-stationary). A low p-value (< 0.05) indicates rejecting the null hypothesis, suggesting that the series is stationary. From the table 5.1, p-values of Rainfall, Paddy, and Ragi have p-values < 0.05, indicating that these data series are stationary. The remaining data series are non-stationary. Accordingly, the crops Paddy and Ragi are suitable for a VAR model, while for the other crops, a Vector Error Correction Model (VECM) or ARIMA model can be applied.

Table 5.1: *Augmented Dickey-Fuller Test*

Crops	ADF Test Statistic	p - value
Rainfall	-4.568784031659	0.0001474082433563
Paddy	-3.966250739089	0.0015979164166240
Cholam	-1.967101070968	0.3011817556291899
Cumbu	-2.607373774359	0.0914828150744177
Ragi	-3.585822200935	0.0060379338458998
Maize	-1.018083747077	0.7466116628126567

The assessment of linearity is conducted using a linear regression model. The calculated R^2 value is 0.08842 for the relationship between Rainfall and Paddy, and 0.000315 for Rainfall and Ragi. Based on these results, we can proceed with applying a VAR model only for Paddy and Ragi.

5.2. Lag Selection

Selecting the optimal lag length in a model is an important step because it determines how many past observations of the variables in the model are used to predict their future values. Incorrect lag selection can lead to misleading results.

Table 5.2: *Optimum Lag Selection*

Lag	AIC	BIC	HQIC
0	64.48	64.67	64.54
1	19.01	19.96	19.30
2	20.83	22.54	21.35
3	21.21	23.68	21.96
4	20.84	24.08	21.83
5	23.73	27.73	24.95

The table 5.2 presents the optimal lag order selection for a VAR model based on three selection criteria: AIC, BIC, and HQIC. The minimum AIC value is 19.01, the minimum BIC value is 19.96, and the minimum HQIC value is 19.30 at lag1. The optimal lag length is 1, as determined by all three criteria. This indicates that the relationships between the variables are best captured by considering only the most recent lag in the VAR model.

5.3. Impulse Response Function

In the IRF analysis of Rainfall to Paddy, a unit shock (unexpected change) in rainfall is used, and the graph illustrates its effect on paddy production over time. A one-unit shock in Rainfall results in a significant positive response in Paddy at the start, as indicated by the sharp rise in the impulse response curve. The shock in Rainfall has an immediate positive effect on Paddy production, indicating that rainfall is generally beneficial in the short term. The negative response observed in the subsequent periods may be due to the harmful effects of excessive rainfall, such as waterlogging, soil nutrient depletion, or flooding.

By Period 10, the effect of the shock in Rainfall on Ragi almost vanishes, indicating that the system stabilizes over time. In periods where the confidence intervals include zero, the impact of the shock may not be statistically significant. Initially, the impact is statistically significant (confidence intervals do not cross zero), but as the intervals narrow toward zero over time, the significance of the effect decreases.

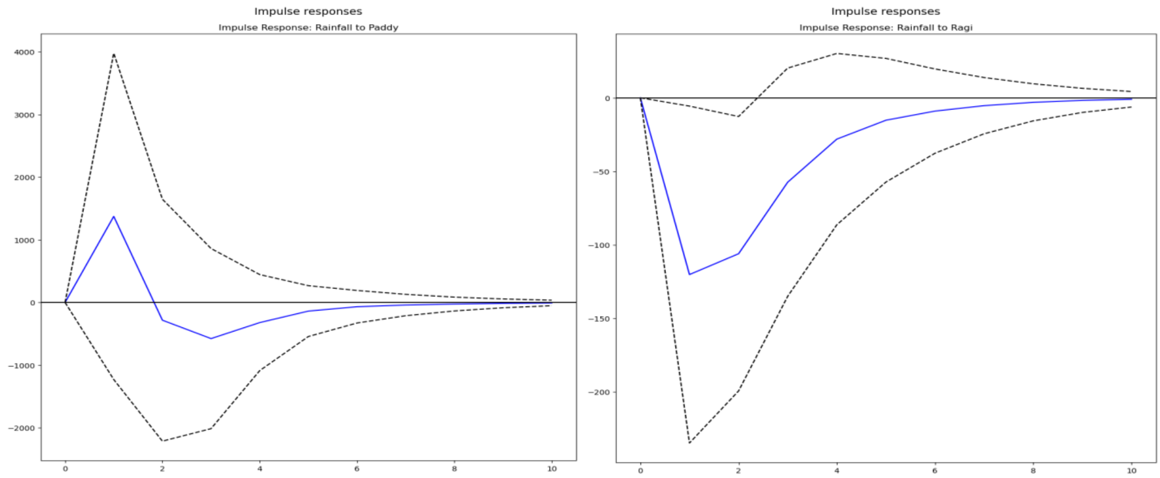

Figure 5.1: *Impulse Response Function Rainfall to Paddy and Ragi*

5.4. VAR Model

Lee[10], conducts an in-depth analysis of the causal relationships and dynamic interactions between asset returns, real activity, and inflation in the postwar United States using a multivariate VAR framework, providing valuable insights into the complex interplay between macroeconomic indicators and financial variables. The presented VAR model estimated using Ordinary Least Squares (OLS). It consists of two equations, each corresponding to one of the endogenous variables: Paddy and Ragi. A total of 33 observations were utilized for estimation, providing the basis for analyzing interdependencies among these variables. The log-likelihood value of -479.937 serves as a measure of the model's overall fit, with higher (less negative) values indicating better fit. To assess and compare model performance, information criteria such as the AIC, BIC, and HQIC are computed. These criteria penalize model complexity, and lower values indicate a better balance between fit and simplicity. Among them, BIC is often favoured due to its stricter penalty for model complexity, making it a preferred metric for model selection.

$$Paddy_t = \alpha_1 + \phi_{11} Year_{t-1} + \phi_{12} Rain\_fall_{t-1} + \phi_{13} Paddy_{t-1} + \in_{Paddy} \tag{9}$$

$$Ragi_t = \alpha_2 + \phi_{21} Year_{t-1} + \phi_{22} Rain\_fall_{t-1} + \phi_{23} Ragi_{t-1} + \in_{Ragi} \tag{10}$$

In equation (9), α_1 represent a Constant term or baseline value for Paddy that is independent of other variables, $\phi_{11} Year_{t-1}$ is the influence of year from the previous period on Paddy, $\phi_{12} Rain\_fall_{t-1}$ is the influence of the previous period's value of Rainfall on Paddy, $\phi_{13} Paddy_{t-1}$ captures the effect of Paddy's own previous value on its current state, reflecting consistency or persistence in its behavior over time. $\in_{Paddy}$ is error term, in other words random shocks or unexplained variation in Paddy. A similar interpretation applies to equation (10).

Table 5.3: *Actual and Forecasted Values*

	Actual Values			Forecasted Values		
Year	Rainfall (in mm)	Paddy (in Tonnes)	Ragi (in Tonnes)	Rainfall (in mm)	Paddy (in Tonnes)	Ragi (in Tonnes)
1990-91	714.6	5782440	316240	714.60	5782440.00	316240.00

1991-92	898.9	6596260	310610	899.76	6289285.80	301198.87
1992-93	862	6805720	291000	919.44	6536957.02	268547.41
1993-94	1171.9	6749810	330970	906.59	6389536.28	261099.76
1994-95	933.8	7558710	285020	982.96	7036997.82	245495.86
1995-96	750.6	5290030	221060	904.29	6492306.99	243275.31
1996-97	1121.2	5805300	190530	935.48	5804547.51	251970.63
1997-98	1133.8	6893730	217940	1008.06	6142549.85	185342.88
1998-99	1080.4	8141300	240610	979.00	6365657.99	189505.47
99-2000	896.8	7532100	245940	929.53	6479790.50	197842.16
2000-01	785.3	7366320	259490	908.25	6254337.95	229478.04
2001-02	795.2	6583630	235310	889.79	6187846.23	252689.53
2002-03	731	3577108	140169	919.15	6039911.11	245827.78
2003-04	1034.6	3222776	176381	1001.40	5301287.55	229395.41
2004-05	1078.9	5061622	154085	1085.72	5907828.88	215194.89
2005-06	1304.1	5209433	132172	1040.43	5910618.67	181648.21
2006-07	859.7	6610607	148148	1090.67	6093582.41	141017.99
2007-08	1164.8	5039954	175944	945.14	5658135.72	192788.63
2008-09	1023.1	5183385	169944	1068.17	6173918.51	184698.28
2009-10	937.8	5665258	160939	1032.94	5964952.35	198239.93
2010-11	1165.1	5792415	171096	1000.19	5825139.33	200128.79
2011-12	937.1	7458657	224862	1051.55	6196705.45	176887.79
2012-13	743.1	4050334	138011	948.11	6276321.94	220050.68
2013-14	790.6	7115195	362343	1012.75	5400115.14	227811.75
2014-15	987.9	7949437	349628	929.35	6886061.78	317211.53
2015-16	1118	7374681	282054	951.42	7110411.10	278908.07
2016-17	598.1	3554113	114429	1002.30	6876184.23	231659.58
2017-18	1017.2	6638450	321332	1003.97	5084997.01	239505.93
2018-19	698.9	6131550	255975	1006.66	6960801.78	273190.79
2019-20	985.8	7265161	274474	950.78	6149976.56	282109.08
2020-21	1232.8	6881725	288627	984.29	6684976.18	247023.14
2021-22	1401.1	7906373	227476	1056.34	7087419.11	228083.97
2022 -23	1170.6	7556567	206553	1065.69	6998552.23	165293.84
2023-24	-	-	-	1025.00	6567790.00	186125.00
2024-25	-	-	-	1025.00	6229389.00	202381.00
2025-26	-	-	-	1038.00	6319154.00	214837.00
2026-27	-	-	-	1040.00	6419746.00	219749.00
2027-28	-	-	-	1040.00	6463104.00	221745.00
2028-29	-	-	-	1041.00	6483788.00	223006.00
2029-30	-	-	-	1043.00	6500922.00	223927.00
2030-31	-	-	-	1045.00	6517247.00	224571.00
2031-32	-	-	-	1048.00	6532388.00	225030.00
2032-33	-	-	-	1050.00	6546546.00	225385.00

Table 5.3 provides an actual and forecasted of yearly values of Rainfall, Paddy and Ragi over the period from 1990-91 to 2032-33. The actual and estimated rainfall values over the time period are depicted in Figure 5.2(a). The estimated rainfall values are relatively more stable compared to the actual values, indicating that the forecasting model has smoothed out the extreme variations seen in the actual data. From the overall trend, it appears that both actual and forecasted rainfall have shown similar trajectories in recent years, with both stabilizing around the 1100 mm mark. The consistency of the forecasted values over time suggests that the forecasting method might be using a simplified approach that doesn't capture short-term weather fluctuations.

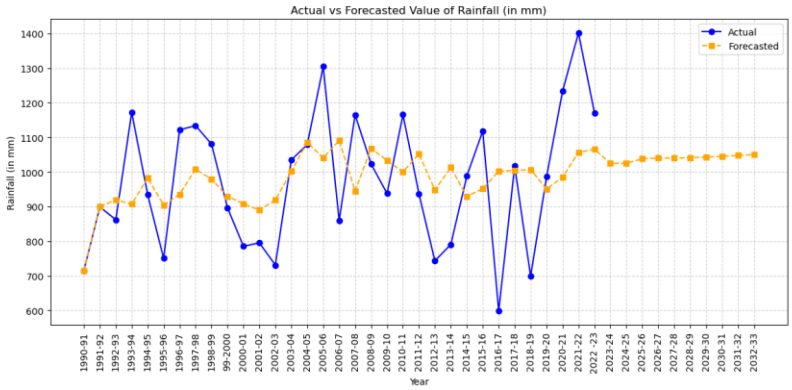

Figure 5.2(a): *Actual vs Forecasted Value of Rainfall (in mm) over the Years*

Figure 5.2(b) illustrates the historical variations in paddy production alongside the actual and forecasted values. From around 2020 onwards, the forecast remains steady, showing no major variation. This stability might suggest an assumption of consistent external conditions affecting paddy production. The closeness of the blue and red lines in many years suggests that the forecasting model performs well in capturing production trends.

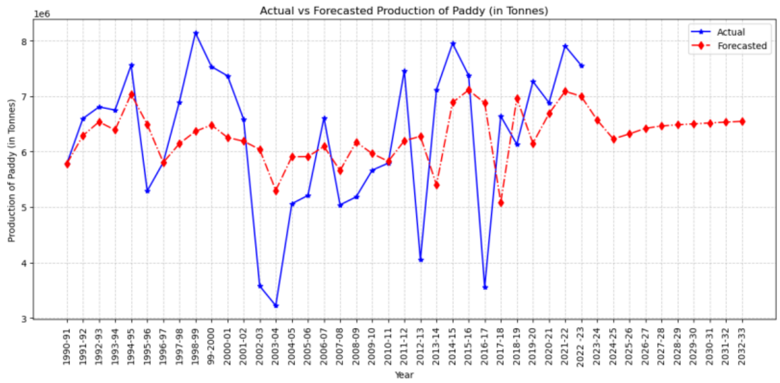

Figure 5.2(b): *Actual vs Forecasted Paddy Production (in Tonnes) over the Years*

Figure 5.3(c) depicts the historical trends in Ragi production along with the actual and forecasted values. The actual production of Ragi exhibits significant fluctuations over the years, with periods of both steep increases and decreases. The forecasted production generally smooths out the variability, following the overall trend of actual production but without sharp deviations. The forecasted production stabilizes in the later years, indicating an assumption of steadier production levels. This could be based on historical trends or constraints in the forecasting method.

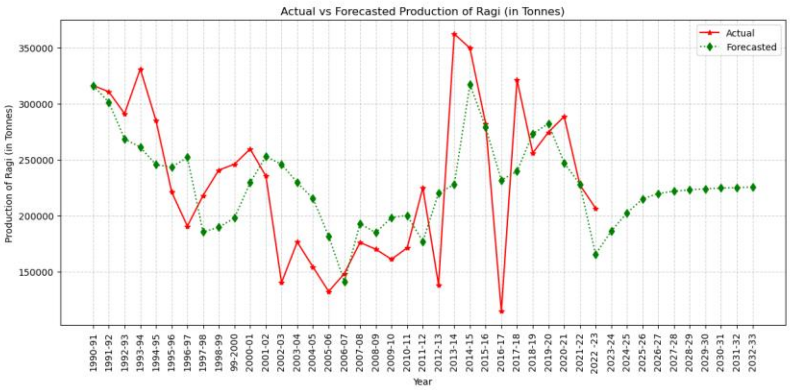

Figure 5.2(c): *Actual vs Forecasted Ragi Production (in Tonnes) over the Years*

VI. Conclusion

Rainfall shocks can have both positive and negative effects on Paddy production. Measures to mitigate short-term disruptions, such as drainage systems or adaptive farming techniques, may help stabilize yields. Over time, the negative impact subsides and stabilizes near zero, suggesting that the system recovers after the shock in Rainfall, and the long-term relationship between Rainfall and Ragi is relatively neutral. The forecast predicts a slow and steady improvement in Rainfall, which is expected to positively influence the production of both Paddy and Ragi. In Rainfall, the accuracy of the forecast can be assessed by comparing the actual data points with the forecasted ones. In some periods, the forecasted values align well with the actual values, while in others, there are noticeable discrepancies. The forecasting model provides a general sense of the production trend for Paddy but struggles with accurately predicting extreme changes. The forecast stabilizes future production, which might oversimplify the dynamics of agricultural production influenced by factors like climate, policy, and market conditions. Refining the model with additional variables or enhancing the methodology could improve accuracy, especially for capturing sharp variations in production. The forecasting model captures the general trend of Ragi production but struggles with periods of high volatility. Future adjustments to the model, such as incorporating additional variables or using more robust techniques, could improve accuracy, particularly during periods of extreme variability. Ragi demonstrates more resilience and consistent growth, whereas Paddy shows variability and slower recovery, signalling the need for specific attention to enhance its yield.

References

[1] Barkley Rosser, J. Jr, Richard, and Sheehan, G. (1995). A Vector Autoregressive Model of the Saudi Arabian Economy, *Journal of Economics and Business*, 47: 79 – 90.

[2] Dickey, D. A, and Fuller, W, A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366): 427-431.

[3] Dickey, D. A, and Fuller, W, A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49(4): 1057-1072.

[4] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3): 424–438.

[5] Granger, C. W. J. and Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2(2): 111–120.

[6] Granger, C. W. J. (1988). Some Recent Development in a Concept o f Causality. *Journal of Econometrics*, 39(1-2): 199-211.

[7] Hamzah, L. M, Nabilah, S. U, Russel, E, Usman, M, Virginia, E. and Wamiliana, (2020), Dynamic Modelling and Forecasting of Data Export of Agricultural Commodity by Vector

Autoregressive Model, *Journal of Southwest Jiaotong University*, 55(3): 1-10.

[8] Jawahar Farook, A. and Senthamarai Kannan, K. (2014). Climate Change Impact on Rice Yield in India – Vector Autoregression Approach, *Sri Lankan Journal of Applied Statistics*, 16(3): 161-178.

[9] Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12(2-3): 231–254.

[10] Lee. B. S. (1992). Causal Relations Among Stock Returns, Interest Rates, Real Activity, and Inflation. *The Journal of Finance*, 47(4): 1591-1603.

[11] Lütkepohl, H. (2008). Impulse Response Function. *The New Palgrave Dictionary of Economics*, 1–5.

[12] Runkle, D. E. (1987). Vector Autoregressions and Reality. *Journal of Business & Economic Statistics*, 5(4): 437-442.

[13] Sasikumar, R. and Sheik Abdullah, A. (2017). Vector Autoregressive Approach for Impact of Oil India Stock Price on Fuel Price in India. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 3(1-2): 41-47.

[14] Stock, J. H. and Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4): 101–115.

[15] Waggoner, D. F. and Zha T. (1999). Conditional Forecasts in Dynamic Multivariate Models. *The Review of Economics and Statistics*, 81(4): 639-651.

[16] Zivot, E. and Wang, J. Modeling Financial Time Series with S-Plus®, Springer Science+Business Media New York, 2003.