Original Research Article

Forecasting of Cropping Intensity of Kerala in India: a Multivariate Approach

ABSTRACT

A state like Kerala in Indian subcontinent, even though, blessed with the typical physiographical features contributing to its agricultural diversity, has an average cropping intensity of 132 percentages which shows the huge gap between its potential and actual production. A multivariate statistical approach known as Vector Auto Regressive (VAR) modeling was done in the present study to analyse the behavior of Cropping Intensity in reference to gross cropped area, net cropped area and annual rainfall of the state from 1965-66 to 2022-23 and to forecast the same. VAR(1) model was found to be better fitting for prediction as well as forecasting based on criteria like Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQ), Schwarz's Criterion (SC) and Akaike's Final Prediction Error criterion (FPE). The model suggested that Cropping Intensity is influenced by past Cropping Intensity, Gross Cropped Area, Net Cropped Area and Annual Rainfall.

Key words: *Agriculture, Forecasting, VAR(p), Cropping Intensity, Kerala*

INTRODUCTION

A state like Kerala in Indian subcontinent enjoys its agricultural boon from the presence of two monsoons in an year. Blessed with the typical physiographical features is an added golden feather to its agricultural diversity. Rice being the major staple crop of the state is cultivated in three different seasons such as Virippu, Mundakan and Puncha in most of the parts of Kerala. Many seasonal, annual and perennial crops are also cultivated extensively in the State. But, still the average cropping intensity of the State is 132 percentages which shows the huge gap between its potential and production.

A multivariate statistical approach was made in the present study to analyse the behavior of Cropping Intensity in reference to gross cropped area, net cropped area and annual rainfall of the state. Thus,the present study confines to a multivariate time series model called Vector Auto Regressive (VAR) modeling and forecasting of cropping intensity of Kerala in India.

A study conducted by Viswanathan (2014), using Kerala state statistics between 1995–96 and 2009–10, concluded that there was a decrease in the growth of area and production for most crops in the state along with their stagnant yield levels. Johnson (2018) noted the changes in land-use pattern and cropping pattern in the State of Kerala over 61 years from 1956–57 to 2016–17 and found that there was a clear shift in land use with a decline in net cropped area and a rise in area under non-agricultural use. Shijitha et.al (2020) concluded by observing that the regional variations in physiological characteristics influence the areal pattern of agricultural land use, crop diversification as well as the selection of crops, and results in variations in the cropping intensity in terms of productivity per unit of arable land. George and Sharma (2020) assessed crop diversification over a period of 30 years (1987-88 to 2016-17) along with the trends and growth in area of major crops of Kerala and revealed that the area under food crops in Kerala has been dwindling over the years. In the process, crop diversity suffered as farmers have shifted from diversified to specialized farming.

MATERIALS AND METHODS

The yearly data on four variables namely Gross Cropped Area in lakh hectares, Net Cropped Area in lakh hectares, Cropping Intensity in Percentages and Annual Rainfall in mm from 1965-66 to 2022-23 over a period of 58 years for the entire Kerala State were collected from Economic Review published by The Kerala State Planning Board (https://spb.kerala.gov.in) and the official website of The India Meteorological Department (https://mausam.imd.gov.in).

Gross Cropped Area refers to the land utilised for crop cultivation once or more than once in a year. This is also known as Total Cropped Area. Net Cropped Area is the actual land physically available for cultivation in a year. The ratio of the gross cropped area to net cropped area expressed in percentages gives the Cropping Intensity for the year. It is a measure of crop diversification and intensification. Since the presence of two monsoons greatly influence the agricultural production of the state annual precipitation can be considered to have an influence on the cropping intensity of the state.

Vector Auto Regressive (VAR) Model: VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables. As an example suppose that we measure three different time series variables, denoted by $Y_{t,1}, Y_{t,2}$ and $Y_{t,3}$. The vector autoregressive model of order 1, denoted as VAR(1), is as follows:

$$Y_{t,1} = \alpha_1 + \emptyset_{11}Y_{t-1,1} + \emptyset_{12}Y_{t-1,2} + \emptyset_{13}Y_{t-1,3} + w_{t,1}$$

$$Y_{t,2} = \alpha_2 + \emptyset_{21}Y_{t-1,1} + \emptyset_{22}Y_{t-1,2} + \emptyset_{23}Y_{t-1,3} + w_{t,2}$$

$$Y_{t,3} = \alpha_3 + \emptyset_{31}Y_{t-1,1} + \emptyset_{32}Y_{t-1,2} + \emptyset_{33}Y_{t-1,3} + w_{t,3}$$

Each variable is a linear function of the lag 1 values for all variables in the set. In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.

Autocorrelation Function: Autocorrelation Functions (ACF) and Partial Autocorrelation Function (PACF) generally produce plots that are very important in finding the values p, q and r for Autoregressive (AR) and

Moving Average (MA) models.Autocorrelation measures the degree of similarity between a given time series and the lagged version of that time series over successive time periods. Autocorrelation guides the determination of order of ARIMA and MA models by providing insights into the number of lag terms to include.

ADF test: Augmented Dickey Fuller test (ADF Test) is a common statistical test used to test whether a given Time series is stationary or not. The ADF test belongs to a category of tests called 'Unit Root Test'. A unit root is said to exist in a time series of the value of alpha = 1 in the below equation.

$$Y_t = \alpha Y_{t-1} + \beta X_e + \varepsilon$$

where, $Y_t$is the value of the time series at time *'t'* and $X_e$is an exogenous variable. The presence of a unit root means the time series is non-stationary. Besides, the number of unit roots contained in the series corresponds to the number of differencing operations required to make the series stationary.

The Augmented Dickey-Fuller test evolved based on the above equation and is one of the most common form of Unit Root test. The ADF test expands the Dickey-Fuller test equation to include high order regressive process in the model.

$$\mathrm{Y_t} - \mathrm{c} + \beta \mathrm{t} + \alpha \mathrm{Y_{t-1}} + \emptyset_1 \Delta \mathrm{Y_{t-1}} + \emptyset_2 \Delta \mathrm{Y_{t-2}} + \cdots + \emptyset_{\mathrm{p}} \Delta \mathrm{Y_{t-p}} + \mathrm{e_t}$$

Since the null hypothesis assumes the presence of unit root, that is α=1, the p-value obtained should be less than the significance level (say 0.05) in order to reject the null hypothesis.

Decision Criteria for lags: The number of lags to be considered as optimum to fit a VAR model can be obtained based on the following four Information Criteria.

(i) Akaike Information Criterion (AIC): The AIC is a measure of the relative goodness of fit of a model. AIC does not provide a test of a model in the sense of testing a null hypothesis; i.e., AIC can tell nothing about how well a model fits the data in an absolute sense. In the general case, the AIC is :

$$AIC = 2k - \ln(L)$$

where k is the number of parameters in the statistical model, and L is the maximised value of the likelihood for the estimated model. AIC with a correction for finite sample sizes is given by:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

(ii) Hannan-Quinn Information Criterion (HQ):where n denotes the sample size. Thus, AICc is AIC with a greater penalty for extra parameters.

In statistics, the Hannan–Quinn information criterion (HQC) is a criterion for model selection. It is an alternative to Akaike information criterion (AIC) and Bayesian information criterion (BIC). It is given as:

$$HQC = -2L_{max} + 2k \ln(ln(n))$$

where L_{max} is the log-likelihood, *k* is the number of parameters and *n* is the number of observations.

(iii) Schwarz's Criterion (SC): In statistics, The Schwarz Information Criterion (SC) is a criterion for model selection among a finite set of models. It is based, in part, on likelihood function, and it is closely related to AIC. The formula for SC is:

$$-2\ln p(x|k) \approx SC = -2\ln L + k\ln(n)$$

Under the assumption that the model errors or disturbances are independent and identically distributed according to a normal distribution and that the boundary condition that the derivative of the log likelihood with respect to the true variance is zero,

$$SC = n\ln(\sigma^2) + k\ln(n)$$

where σ² is the error variance and is defined as:

$$\sigma^2 = 1/n\sum_{i=1}^{n}(x_i - \bar{x})^2$$

(iv) Akaike's Final Prediction Error criterion (FPE): It provides a measure of model quality by simulating the situation where the model is tested on a different data set. After computing several different models, comparison can be done using this criterion. According to Akaike's theory, the most accurate model has the smallest FPE. Akaike's Final Prediction Error (FPE) is defined by the following equation:

$$FPE = \det\left(\frac{1}{N}\sum_{1}^{N} e\left(t, \widehat{\theta_N}\right)\left(e\left(t, \widehat{\theta_N}\right)\right)^T\right)\left(\frac{1 + \frac{d}{N}}{1 - \frac{d}{N}}\right)$$

where *N* is the number of values in the estimation set, *e*(*t*) is a *ny*-by-1 vector of prediction errors, θ_N represents the estimated parameters and *d* is the number of estimated parameters.

Estimation: Despite their seeming complexities, VAR models are quite easy to estimate. The equation can be estimated using ordinary least squares with the assumptions of stationarity, absence of multicollinearity and conditional mean of zero for error term. The estimates will be consistent and can be evaluated using traditional t-statistics and p-values.

Test for heteroscedasticity: The Lagrange Multiplier (LM) test is one of the principal tools to detect ARCH and GARCH effects in data analysis. It is a test of no conditional heteroskedasticity against an ARCH model. The test is easy to compute from an auxiliary regression involving the squared least squares (LS) residuals. The LM statistic is asymptotically distributed as χ^2 under the null hypothesis.

Test for Normality: In statistics, the Jarque–Bera test is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution. The test is named after Carlos Jarque and Anil K. Bera. The test statistic is always nonnegative. If it is far from zero, it signals the data do not have a normal distribution. The test statistic JB is defined as

$$JB = \frac{n}{6}(s^2 + \frac{1}{4}(k-3)^2$$

where n is the number of observations (or degrees of freedom in general); S is the sample skewness, K is the sample kurtosis

$$s = \frac{\mu_3}{\sigma^3} \qquad k = \frac{\mu_4}{\sigma^4}$$

Where μ_3 and μ_4 are the estimates of third and fourth central moments, respectively, $\bar{y}$ is the sample mean, and σ^2 is the estimate of the second central moment, the variance.

If the data comes from a normal distribution, the JB statistic asymptotically has a chi-squared distribution with two degrees of freedom, so the statistic can be used to test the hypothesis that the data are from a normal distribution. The null hypothesis is a joint hypothesis of the skewness being zero and the excess kurtosis being zero. Samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0 (which is the same as a kurtosis of 3).

Test for structural break: Many models assume that the relationship between variables stays constant across the entire period. However, there are cases where changes in factors outside of the model cause changes in the underlying relationship between the variables in the model. Structural break models capture exactly these cases by incorporating sudden, permanent changes in the parameters of models. The CUSUM test for instability is appropriate for testing for parameter instability in the intercept term. It is best described as a test for instability of the variance of post-regression residuals. The CUSUM test statistic is computed from the one-step-ahead residuals of the recursive least squares model. It is based on the intuition that if β changes from one period to the next then the one-step-ahead forecast will not be accurate and the forecast error will be greater than zero.

Granger Causality: Granger-causality statistics test whether one variable is statistically significant when predicting another variable. The Granger-causality statistics are F-statistics that test if the coefficients of all lags of a variable are jointly equal to zero in the equation for another variable. As the p-value of the F-statistic decreases, evidence that a variable is relevant for predict another variable increases.

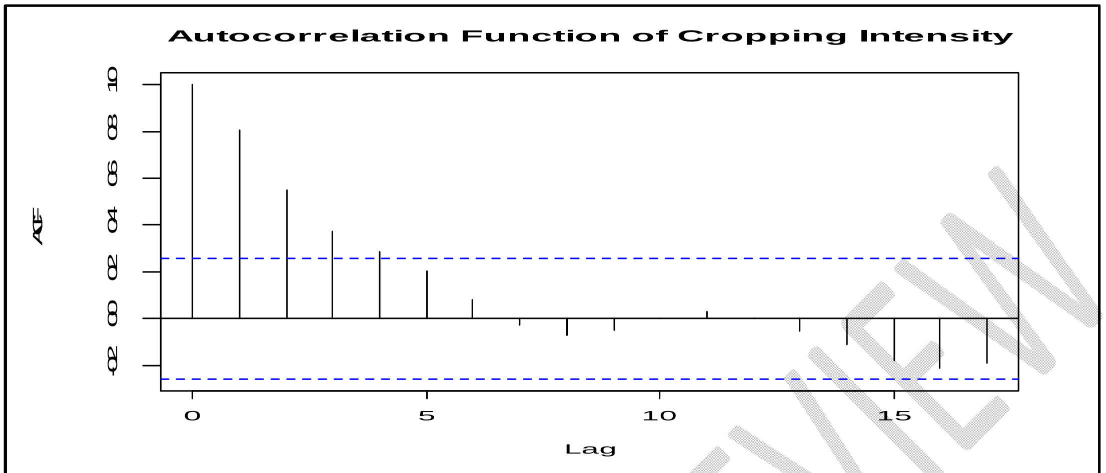
Forecast Error Variance Decomposition: Forecast error decomposition separates the forecast error variance into proportions attributed to each variable in the model.Intuitively, this measure helps us judge how much of an impact one variable has on another variable in the VAR model and how intertwined our variables' dynamics are.

RESULTS AND DISCUSSION

The four time series were considered as a single vector with column headings as Cropping Intensity, Gross Cropped Area, Net Cropped Area and Annual Rainfall. Then Vector Auto Regressive model with p lags was applied to the data.

To consider the degree of persistence in the data, autocorrelation functions were made use of. The plot of autocorrelation functions of the croppingintensity is given in Figure 1.

Figure 1 Plot of autocorrelation functions of the time series



As the persistence in Cropping Intensity is relatively high unit root test was done using Augmented Dickey Fuller Test. In case of Cropping Intensity, p values of estimates and F value were less than 0.05 and the null hypothesis of unit root was rejected. Summary of Augmented Dickey Fuller Test is given in Table 1.

Table 1 Summary of Augmented Dickey Fuller Test

Sl. No.	**Data**	**Coefficients**		**R^2**	**Adjuster R^2**	**F statistic**
1	Cropping Intensity	Intercept	36.6747*	0.2764	0.2346	6.619*
		z.lag.1	-0.2686*			
		tt	-0.0384*			
		z.diff.lag	0.3085*			

* Significant at 5% level

After creating a multivariate object for four time series, information criteria such as Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQ), Schwarz's Criterion (SC) and Akaike's Final Prediction Error criterion (FPE) were used to decide upon the number of lags to be included. All of the information criteria suggested that the use of 1 lag would be appropriate, which would imply to set p=1 when estimating the model. Vector Auto Regressive Estimation with 1 lag (p=1) was done using Cropping Intensity, Gross Cropped Area, Net Cropped Area and Annual Rainfall as endogenous variables and the obtained roots of characteristics polynomials were 1, 0.6899, 0.6899 and 0.1397.

The estimation results suggested that the system is stable (the characteristic roots were interpreted as eigenvalues in this case). There would appear to be many insignificant variables in this model, where Cropping Intensity is influenced by past Cropping Intensity, Gross Cropped Area, Net Cropped Area and Annual Rainfall. Summary of estimation results is given in Table 2.

Table 2. Summary of Estimation using VAR (1) Model

Cropping Intensity				
Model	ci = -0.2104 (ci.l1) + 4.194(gca.l1) -4.260(nca.l1) -0.0061(rf.l1) + 0.0134			
R^2	**Adjusted R^2**	**F Statistic**	**MAPE (%)**	**RMSE**

0.7762	0.7589	45.08*	2.3622	3.6001

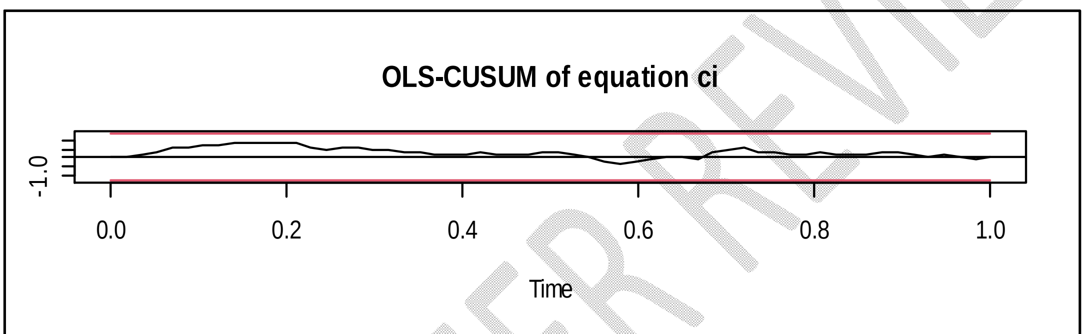
gca: Gross Cropped Area, nca: Net Cropped Area, ci: Cropping Intensity, rf: Annual Rainfall, l1: lag of 1

To test for heteroscedasticity in the residuals a multivariate ARCH Lagrange-Multiplier test was performed where a chi squared value of 450 was obtained with p value greater than 0.05 and this indicated that there is absence of heteroscedasticity in the data.

To consider the distribution of residuals normality test was applied where a chi squared value of 6.0795 was obtained with p value greater than 0.05. In case of skewness and kurtosis, the chi squared values were 3.7513 and 2.3282 respectively with p values greater than 0.05. This indicated that the residuals are fairly distributed normally.

To test for the structural break in the residuals a Cumulative Sum (CUSUM) test was applied. It was found that there is no break in the respective confidence intervals. Plot of CUSUM test is given in Figure 2.

Figure 2. Plot of CUSUM test



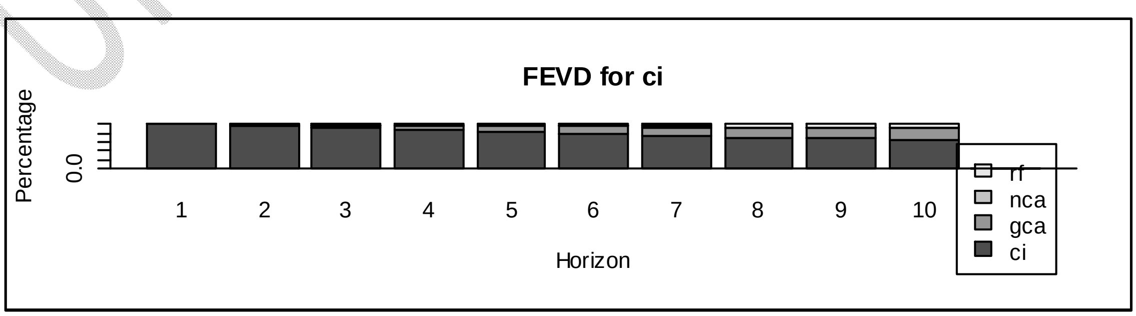
Test for Granger causality revealed that Cropping Intensity does not Granger-cause the other three variables. There is instantaneous causality in case of Cropping Intensity w.r.t other three variables. Summary of test for Granger Causality is given in Table 3 below.

Table 3. Summary of Granger Causality Test

Sl. No.	**Variable**	**Granger Causality (F Statistic)**	**Instantaneous Causality (χ^2 Statistic)**
1	Cropping Intensity	1.0047	28.478*

On setting the steps ahead to ten, Forecast Error Variance decompositions were generated. The results suggested that Cropping Intensity was largely determined by Gross Cropped Area shocks. Plot of forecast error variance decompositions is given in Figure 3.

Figure 3 Plot of forecast error variance decompositions

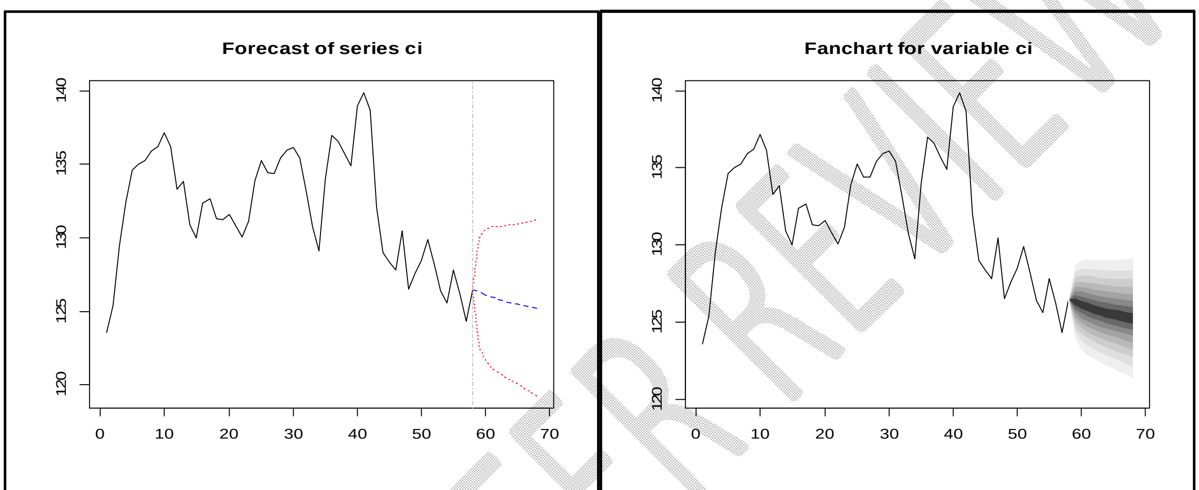


Forecasting was done for ten years ahead using the VAR (1) model identified and they are presented in Table 4 below. Plot and Fanchart for the forecasts are given in Figure 4

Table 4. Forecasted Values using VAR (1) Model

Year	2023-24	2024-25	2025-26	2026-27	2027-28	2028-29	2029-30	2030-31	2031-32	2032-33
Cropping Intensity	126.32	126.13	125.96	125.82	127.70	125.59	125.49	125.41	125.32	125.23

Figure 4 Plot and Fanchart for the forecasts using VAR (1) Model



CONCLUSION

Crop Intensification is the best strategy in mitigating hunger and poverty in an economy like Kerala which is found to have a shift from producer state to consumer state. Trends in cropping intensity of the state showed a decline along with decline in gross cropped area and net cropped area. The present study attempted to model the cropping intensity of Kerala State in India using data from 1965-66 to 2022-23 through a multivariate time series model called Vector Auto Regressive model with single lag (VAR (1)) and forecasting of cropping intensity was done for ten years ahead. The estimation results suggested that the system is stable and there would appear to be many insignificant variables in the model, where Cropping Intensity is influenced by past Cropping Intensity, Gross Cropped Area, Net Cropped Area and Annual Rainfall

REFERENCES

- Andrews, S (2013) Dynamics of cropping pattern shifts in Kerala: sources and determinants. Agric. Situat. India.70(6):15-28.
- George, G. P., & Sharma, H. O (2020) Determinants of crop diversification in Kerala- a temporal analysis. Journal of Tropical Agriculture, 58(1). Retrieved from https://jtropag.kau.in/index.php/ojs2/article/view/882

- Johnson, D (2018) Cropping Pattern Changes in Kerala, 1956–57 to 2016–17. Review of Agrarian Studies, 8(1):65-99
- Joseph, B. and Joseph, K.J (2005) Commercial agriculture in Kerala after the WTO. S. Asia Econ. J. 6(37):37-57.
- Saleena, E.C (2017) Performance appraisal of Indian agricultural system with special reference to agricultural scenario in the state of Kerala. J. Bus. Manag.19(10):5-11.
- Shijitha, M., Francis, K. M., & Sabu, P. J (2020) Physiological Characteristics of Land and Crop Diversification in Kerala. International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 4, page no.1647-1650, April-2020, Available :http://www.jetir.org/papers/JETIR2004422.pdf
- Viswanathan, P. K (2014) The Rationalisation of Agriculture in Kerala: Implications for the Natural Environment, Agro-ecosystems, and Livelihoods. Agrarian South: Journal of Political Economy, 3(1)