SEMINAR REPORT

Statistical Forecast Models in Agriculture

By

POOJA A.

(2018-19-004)

Presented on 12/12/2019

Submitted in partial fulfillment of requirement of the course

STAT 591 Master's Seminar (0+1)

DEPARTMENT OF AGRICULTURAL STATISTICS

COLLEGE OF HORTICULTURE

KERALA AGRICULTURAL UNIVERSITY

VELLANIKKARA THRISSUR, KERALA- 680656

CERTIFICATE

This is to certify that the seminar report entitled **“Statistical Forecast Models in Agriculture”** has been solely prepared by **Pooja A. (2018-19-004)**, under my guidance and has not been copied from seminar reports of any seniors, juniors or fellow students.

Vellanikkara
25/01/2020

Dr. Laly John C.
Professor and Head
Dept. of Agricultural Statistics
College of Horticulture
KAU, Vellanikkara

DECLARATION

I, Pooja A. (2018-19-004) declare that the seminar entitled "**Statistical Forecast Models in Agriculture**" has been prepared by me, after going through various references cited at the end and has not been copied from any of my fellow students.

Vellanikkara
25/1/2020

Pooja A.
(2018-19-004)

CERTIFICATE

This is to certify that the seminar report entitled **"Statistical Forecast Models in Agriculture"** is a record of seminar presented by **Pooja A. (2018-19-004)** on 12th December, 2019 and is submitted for the partial requirement of the course STAT 591.

Dr. Anil Kuruvila
Professor
Department of Agricultural Economics
College of Horticulture, Vellanikkara

Dr. Reshmy Vijayaraghavan
Assistant Professor
Department of Plant Pathology
College of Horticulture, Vellanikkara

Dr. Sangeeta Kutty M.
Assistant Professor
Department of Vegetable Science
College of Horticulture, Vellanikkara

INDEX

LIST OF TABLES

LIST OF GRAPHS

1. Introduction

Forecasting is an important aid in effective and efficient planning. Reliable and timely forecasts provide important and useful input for proper informed planning in agriculture, which is full of uncertainties (Agrawal, 2007). It is needed to determine when an event will occur or a need arise, so that appropriate actions can be taken. Various statistical techniques are employed in forecasting agricultural systems with reasonable precisions well in advance for taking timely decisions. The widespread introduction of computers has led to readily available softwares, which inturn have made forecasting easy.

Objective advance estimates on crop production are necessary for making various policy decisions regarding stock, distribution and supply, pricing, etc. In the light of knowledge of yield forecasting, farmers can decide what they will undertake to do for the coming year to improve the characters based on which forecasting is done as well as how much land they will use, the acreage they will put into each of the various crops, etc.

2. Steps in forecasting

It includes data collection, fitting statistical model and prediction. One of the most difficult and time consuming part in forecasting is collection of valid and reliable data. Forecast can be no more accurate than the data on which it is based. Thus, the data collected should be reliable, accurate, timely and consistent. Similarly the statistical model fitted should be appropriate and judiciously selected for the data. This helps in making more realistic forecasts for future. The final part is to make the forecast.

3. Statistical forecast models

The following statistical forecast models are discussed.

Regression Models:-

- Multiple Linear Regression (MLR) models using plant characters
- Weather indices based MLR models

Time series models

- Exponential smoothing models
- Auto-Regressive Integrated Moving Average (ARIMA) models
- Artificial Neural Network (ANN) model

3.1 Regression Models

When we are concerned with the dependence of a random variable Y (dependent variable/response variable) on a quantity X (independent variable/ regressor/ predictor variable) which is a variable but not a random variable, the equation that relates Y to X is called regression equation. But often, a number of predictor variables affect the response variable. In such cases, the true relationship between Y and the predictor variables (X_1, $X_2,\dots,X_p$) is represented by a multiple regression model,

$$Y=\beta_0+ \beta_1X_1+ \beta_2X_2+\dots+ \beta_pX_p+ \varepsilon$$

where β_0 and β_i (i= 1,2,…,p) are parameters to be estimated and ε is random error.

Assumptions made are:-

i. Regression model is linear in parameters
ii. Absence of multi-collinearity among $X_1, X_2, \dots, X_p$
iii. Errors are assumed to be independently and identically distributed normal random variables with mean zero and constant variance σ^2

i.e., **$\varepsilon \sim IIN(0, \sigma^2)$**

3.1.1 Model Fit Parameters

1. Significance of the F-value in the ANOVA

We can assess the overall model fit using the significance of the F-value and adjusted R^2.

The test statistic follows an F-distribution with k and (n – k – 1) degrees of freedom.

$$\mathbf{F} = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{R^2}{1-R^2}\left(\frac{n-k-1}{k}\right)$$

If test statistic of F is significant then one / more regression coefficients are significant then we proceed with regression analysis. When we interpret the model fit, the F-test is the most critical, as it determines if the overall model is significant. If the model is insignificant, we do not interpret the model further. If the model is significant, we proceed by interpreting individual variables.

2. **Adjusted $\mathbf{R^2}$**

R^2 is a non-decreasing function i.e., as the number of regressors increases, $\mathbf{R^2}$ almost invariably increases. So to compare two $\mathbf{R^2}$ terms, one must take into account the no. of X variables. The $\mathbf{R^2}$ thus defined is known as adjusted $\mathbf{R^2}$. As the number of X variables increases, adjusted $\mathbf{R^2}$ increases less than unadjusted $\mathbf{R^2}$. Adjusted $\mathbf{R^2}$ may be negative but $\mathbf{R^2}$ is necessarily non negative.

$$\textbf{Ad. } R^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right)$$

The adjusted R^2 is a measure of how much the model explains while controlling for model complexity.

3.1.2 Forecast accuracy measure

Mean Absolute Percentage Error (MAPE):-

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - F_t|}{Y_t}\times 100$$

Where, Y_t is the yield at time t, and F_t is the forecasted yield

3.1.3 Multiple Linear Regression (MLR) models using plant characters

This approach uses data on plant characters like crop stand, number of tillers, root length, leaf area, number of flowers, *etc*., as such for forecasting the crop yield. Some of these characters are direct components of yield while some may indirectly affect the crop yield. Thus, choice of plant characters as explanatory variables is done based on the knowledge concerning their influence on crop yield. These models are developed in a simple manner by utilizing data at one point of time only during the crop growth period (Jain *et al.* 1985; Agrawal and Jain, 1996). Multiple linear regression model used is,

$$Y=\beta_0+ \beta_1X_1+ \beta_2X_2+\ldots+ \beta_pX_p+ \varepsilon$$

where Y is yield, X_i (i=1,.2,…,p) are plant characters, β_0 and β_i are constants to be estimated and ε is the error term.

3.1.3.1 Pre-harvest forecast models for prediction of sugarcane yield

Singh and Bapat (1988) developed pre harvest model based on yield attributes using MLR, to forecast sugarcane yield about four months before harvesting. The study was carried out at Kolhapur during 1977-78 to 1979-80. About 168 fields were selected for the study and out of different attributes selected, only three characters contributed significantly to yield. The model used was:

$$\mathbf{Y= \beta_0 + \beta_1 \mathit{X_1} + \beta 2\mathit{X_2}+ \beta_3 \mathit{X_3} + \varepsilon}$$

where Y is sugarcane yield, X_1 is number of canes/plot, X_2 is height of cane, X_3 is girth of cane and ε is error.

Forecast models were fitted using data during 1977-79 in different crop growth periods *i.e.,* 8^{th}, 9th and 10^{th} month after planting. The regression equations fitted are given in the Table.1.

Year	Months after planting	Model	Ad.R^2
1977-78	8	$Y= -114.11 + 1.01X_1 + 69.21X_2$	0.73
	9	$Y= -73.44 + 0.87X_1 + 57.73X_2$	0.51
	10	$Y= -223.57 + 0.99X_1 + 62.14X_2$	0.58
1978-79	8	$Y= -218.12 + 0.63X_1 + 65.52X_2 +19.51X_3$	0.66
	9	$Y= -302.88 + 0.64X_1 + 63.94X_2 +15.86X_3$	0.70
	10	$Y= -278.80 + 0.54X_1 + 79.84X_2 +12.58X_3$	0.69

Table. 1. Forecast models for sugarcane yield

Regression coefficients of X_3 were not significant for year 1977-79. So it was not included in the model. Among different models tried, it was found that model for 8th month of planting of year 1977-78 was more appropriate due to high adjusted R^2 and predictability. Results in Table 2. show that forecasted yield was closer to actual yield.

Year of forecast	Actual yield at harvest	Using equation of year	Forecasted yield at 8 months
1978-79	74.64	1977-78	**72.35**
1979-80	73.80	1977-78	**75.14**
		1978-79	78.44

Table 2. Actual yield *vs.* forecasted yield of sugarcane

Thus at earliest, it's feasible to forecast yield of sugarcane based on these attributes about four months before harvesting. Usually, these models are developed by utilizing data at one point of time during the crop growth period (Agrawal and Jain, 1996).

3.1.4 Weather indices based MLR models

These models are used for forecasting crop yields, crop pest counts, etc. The predictor variables used are transformed weather indices rather than weather variables as such. The model is given by,

$$\mathbf{Y= a_0} + \sum_{i=1}^{p} \boldsymbol{a_i} \ \boldsymbol{Z_i} + \sum_{i\neq j}^{p} \boldsymbol{a_{ij}} \ \boldsymbol{Z_{ij}} + \boldsymbol{e}$$

where $Z_i = \sum_{w=1}^{m} r_{iw} \ X_{iw}$; $Z_{ij} = \sum_{w=1}^{m} r_{ijw} \ X_{jw} \ X_{jw}$,

with Y as the crop yield or pest count and *e* as the random error. r_{iw} / r_{ijw} is the correlation coefficient of Y with i^{th} / product of i^{th} and j^{th} weather variables in w^{th} week and *m* and *p* denote week of forecast and number of weather variables used, respectively. Stepwise regression analysis was used for selecting significant variables (Draper and Smith, 1981; Gomez and Gomez, 1984). Here Z_i 's and Z_{ij}'s are functions of the basic weather variables (rainfall, maximum temperature, minimum temperature, *etc*.). These models utilize data on weather parameters over weeks within years during the crop growth, response variable, if crop yield is available, over years and if pest count is available, over weeks for various years.

3.1.4.1 Weather indices based pre-harvest forecasting of wheat yield

Garde *et al*. (2012) reported that reliable yield forecast for wheat can be made about two and half months before harvest using weather indices based MLR models. They used yield data of wheat (q/ha) for 25 years (1982-83 to 2006-07) of Uttar Pradesh and weekly weather data for 25 years (collected from IMD). The weekly weather variables observed were maximum temperature, minimum temperature, rainfall, pan evaporation, relative humidity at 7.00 hours and 14.00 hours and sunshine hours. Models for different periods from 14 week coinciding with later elongation period to 20th week (early maturation period) were constructed. Based on adjusted R^2 value (Table 3.) and earliest appropriate time of forecast, it was found that, model for 15^{th} week was appropriate.

Pre-harvest week no.	**Model**	**Ad. R^2**
15	$Y= 12.944 + 0.161\ Z_6 + 0.103\ Z_{2,4} + 0.011\ Z_{3,7}$	0.78

Table 3. Forecast model for wheat

Also, model of 15^{th} week had lower error percentage value as in Table 4.

Year	**Actual yield**	**Forecast yield**	**MAPE (%)**
2006-07	24.22	24.94	2.98

Table 4. Actual yield *vs.* forecasted yield of wheat

Thus, it provides pre harvest forecast in 15^{th} week of cropping season of wheat about two and half months before harvesting.

3.1.4.2 Rice yield forecasting using weather indices based MLR model

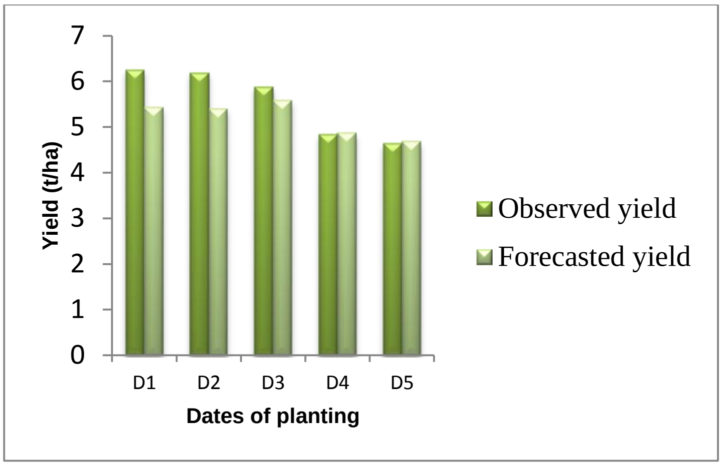
A similar study was conducted at KAU by Ravindran (2018), on forecasting rice yield using weather indices based MLR model. The field experiments were conducted at Agricultural Research Station, Mannuthy. The yield data for 25 dates of planting (May, 2013 to November, 2017) were collected for variety Kanchana (105-110 days). The observations were made on weekly weather variables like maximum temperature, minimum temperature, forenoon relative humidity, afternoon relative humidity, forenoon vapour pressure deficit, bright sunshine hours, rainfall and pan evaporation. The model used was

$$\mathbf{Y = 12.143 + 1.596\ Z_8}$$

where, Y is rice yield

and $Z_8 = \sum_{w=1}^{7} r_{8,w} . X_{8,w}$ (X_8 – Pan evaporation)

The proximity of observed and forecasted yield can be compared in the Graph 1. This shows the accuracy of the model.



Graph.1 Yield forecast of rice variety Kanchana

3.1.5 Logistic regression models

Logistic regression is the statistical forecasting model preferred for analysis of qualitative responses. The model is of the form: P (Y=1) = $\frac{1}{1+\exp(-Z)}$

where Z= $\sum \beta_i X_i$; a function of weather variable X_i.

The forecast model is: **$Y= b_0 + b_1p_1$**

The logistic regression is used for obtaining probabilities of occurrence of the different categories (Misra *et al*, 2004; Agrawal *et al*, 2004).

3.1.5.1 Logistic regression in wheat yield forecasting

Kumar *et al.* (2016) concluded that logistic regression model with "two group classification" can be recommended for forecasting wheat yield by the 11th week after sowing. Time series data (1971-72 to 2009-10) on yield of wheat of and weekly weather data for 39 years of Kanpur was collected. They fitted linear trend equation for yield and estimated yield using the equation. The crop years grouped into two as shown in Table 5.

Good year	Actual yield – Estimated yield > 1
Bad year	Actual yield – Estimated yield < 1

Table 5. Classification of crop years

Probability of good year, p_1 was then found and was used as regressor. Models were fitted as in Table 6. using **$Y = b_0 + b_1 p_1 + b_2 T$**

Year(T) was used as regressor in order to incorporate technological advancement over the years in yield.

SMW	Forecast model	Adj R^2	MAPE(%)
52	**$Y = -1120.18 + 3.93p_1 + 0.57\ T$**	**0.903**	**6.11**
1	$Y = -1118.89 + 3.88p_1 + 0.57\ T$	0.898	6.34
2	$Y = -1118.66 + 3.85p_1 + 0.57\ T$	0.894	6.51
3	$Y = -1118.71 + 3.85p_1 + 0.57\ T$	0.892	6.61

Table 6. Wheat yield forecast models using logistic regression

52nd standard meteorological week has best model fit as compared to others as it explained about 90.30% variability in the yield. Table 7. indicates precision of forecast values as they lie close to the actual values.

SMW	Year	Observed yield	Forecasted yield
52	2007-08	30.08	33.31
	2008-09	33.56	33.88
	2009-10	32.31	34.45

Table 7. Observed and forecasted yield of wheat using logistic regression

3.2 Time series Models

Time series (TS) data refer to an ordered sequence of values of a variable at equally spaced time intervals. In this approach, it is assumed that some aspects of the past pattern will continue to remain in the future. Time series models are usually applicable in short term forecasting. Different types of data patterns should be considered, so that the models can be utilized appropriately (Makridakis *et al.*). There are four types of time series components, *viz.*

i. Horizontal- when data values fluctuate around a constant value
ii. Trend- when there is long term increase or decrease in the data
iii. Seasonal- when series is influenced by seasonal factor/recurs on regular periods
iv. Cyclical- when the data exhibit rises and falls that are not of a fixed period

3.2.1 Exponential Smoothing Models

The idea of exponential smoothing is to smooth the original series the way the moving average does and to use the smoothed series in forecasting future values of the variable of interest. In exponential smoothing, however, we want to allow the more recent values of the series to have greater influence on the forecast of future values than the more distant observations.

Simple Exponential Smoothing (SES) Models: Forecast F_{t+1} is based on weighing the most recent observation Y_t with α and weighing recent forecast using 1-α.

$$\mathbf{F_{t+1} = \alpha\, Y_t + (1\text{-}\alpha)F_t}$$

where, F_t - recent forecast and **α** - smoothing constant, $0 < \alpha < 1$

Double (Holts) Exponential Smoothing (DES) Models: It allows forecasting data with trends.

$$\mathbf{F_{t+m} = L_t + b_t m}$$

where, $L_t = \alpha\, Y_t + (1\text{-}\alpha)\,(L_{t-1} + b_{t-1})$ and $b_t = \beta\,(L_t - L_{t-1}) + (1 - \beta)\, b_{t-1}$, $0 < \beta < 1$

Holt-Winter's Multiplicative Seasonal (HWMS) Models: It is used if seasonality exists in time series data.

$$\mathbf{F_{t+m} = (L_t + b_t m)S_{t-s+m}}$$

where, $L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1})$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \gamma\frac{Y_t}{L_t} + (1-\gamma)S_{t-s}\,, \quad 0 < Ƴ < 1$$

3.2.1.1 Exponential smoothing model to forecast coconut oil price

Indraji (2016), tried to forecast price of coconut oil price using data of monthly average price of coconut oil for Kochi market from January 1990 to December 2015. Holt-Winters' Multiplicative Seasonal (HWMS) model was identified from among the different exponential smoothing models like SES, DES, HWAS and HWMS models for the coconut oil price at Kochi market. The parameters were estimated as followed in Table 8.

Parameter	**Α**	**β**	**Ƴ**
Estimate	1.00	0.28	0.00

Table 8. Estimated parameters

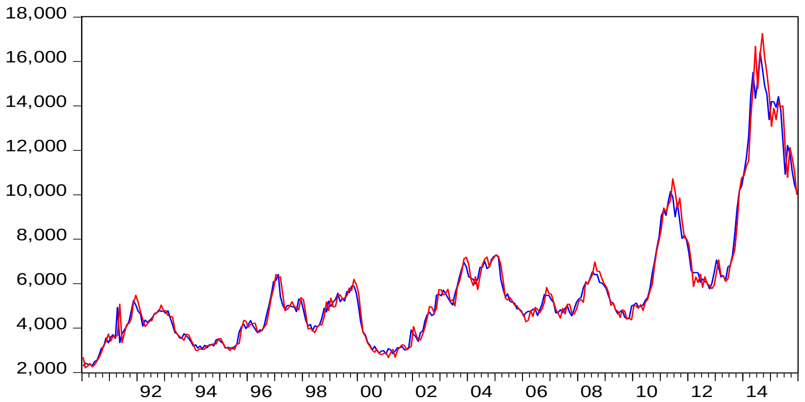
Model used was, $L_t = \frac{Y_t}{S_{t-12}}$

$$b_t = 0.28(L_t - L_{t-1}) + 0.72b_{t-1}$$

$$S_t = S_{t-12}$$

$$F_{t+m} = (L_t + b_t m)S_{t-12+m}$$

The fit of the HWMS model for coconut oil price at Kochi market can be seen from Graph 2.



Graph 2. Actual *vs.* forecasted price of coconut oil

The actual and forecasted values are also in close agreement in the Table 9.

Month	Actual price (Rs./quintal)	Forecast price
Jan-16	9296 (9000-9600)	9736
Feb-16	9093 (8950-9300)	9233
Mar-16	8448 (8200-9000)	8318
MAPE (%)		4.43

Table 9. Actual yield *vs.* forecasted yield of coconut oil

3.2.2 Box –Jenkins ARIMA Models

ARIMA stands for Autoregressive Integrated Moving Average. It was popularized by George Box and Gwilym Jenkins in 1973. Auto-regression means variable to be forecasted are related to past values of itself at varying time lags. Moving average means value of time series at time t is influenced by a current error term and weighted error terms in the past.

General ARIMA (p ,d ,q) model:-

$$\left(1-\sum_{i=1}^{p}\phi_i B^i\right)(1-B)^d Y_t = \left(1-\sum_{j=1}^{q}\theta_j B^j\right)\varepsilon_t$$

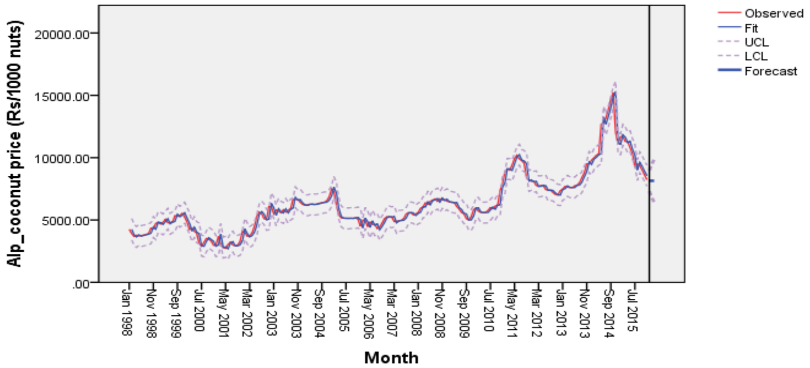
where, Y_t is the value of time series at time t, d - Order of differencing, p - Order of Auto regression, q - order of Moving Average, ϕ_i (i=1,2,…,p) - Auto Regressive parameters and θ_j (j = 1,2,…,q) - Moving Average parameters.

3.2.2.1 ARIMA model for coconut price

Indraji (2016) demonstrated forecasting the price of coconut of Alappuzha market using ARIMA model. Monthly average price of coconut for Alappuzha market from January, 1998 to December, 2015 (216 months) was collected from Coconut Development Board. The model used was ARIMA (0,1,1).

$Y_t - Y_{t-1} = e_t - 0.066Be_{t-1}$ or $(1-B)\,Y_t = (1-0.066B)e_t$

where, Y_t is the price of coconut and B is backshift operator.



Graph 3. Actual and forecasted coconut price

Graph 3. shows the accuracy of actual and predicted coconut price. Their exact values are given in the Table 10.

Month	**Actual price Rs/quintal**	**Forecast price**
16-Jan	8000	8144 (7306-8983)
16-Feb	8000	8144 (6805-9484)
16-Mar	7967	8144 (6445-9844)
MAPE (%)		4.25

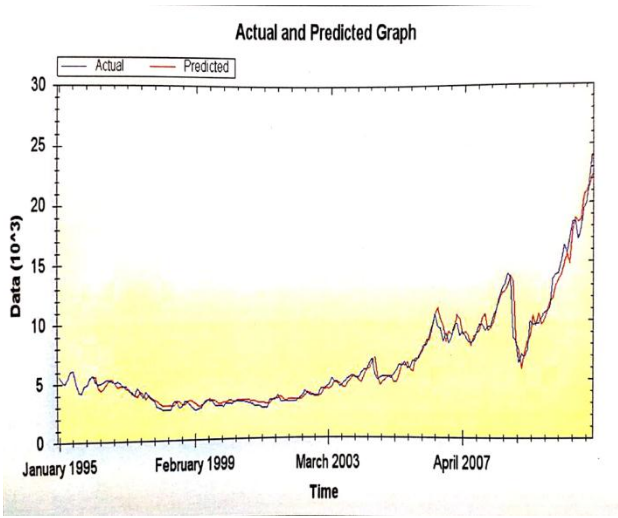
Table 10. Actual and forecasted values of coconut price

3.3 Artificial Neural Network (ANN) model

It is based on artificial intelligence which is the computational ability of a digital computer combined with the desirable functions of the human brain. ANN does not require any prior assumptions of model form. It is capable of identifying underlying pattern of time series. It captures non-linearity in time series data.

3.3.1 ANN model to forecast rubber price

Varghese (2011) used ANN model to forecast the price of natural rubber in India using month-wise modal prices of natural rubber grade RSS-4 in the domestic market of Kottayam from January 1995 to February 2011 (194 months). The results are presented in Graph 4. and Table 11.



Graph 4. Actual and forecasted graph of rubber

Month	Forecasted price (Rs./kg)	Actual Price (Rs./kg)
March	225	227
April	219	240
May	223	222

Table 11. Actual price *vs.* forecasted price of rubber

4. Conclusion

Reliable and timely forecasts provide important and useful input for proper planning in agriculture which is full of uncertainties. Statistical forecast models will help to provide objective crop yield as well as price forecasts with reasonable precision for taking timely decision.

5. Discussion

Q.1 What is meant by the term ‘smoothing’ in time series?

Smoothing is a technique applied to time series to remove the variation between time steps to help us better see patterns or trends in time series.

Q.2 Do these models have any uses other than yield and price forecasting?

Yes, they do. Logistic models can be used to forecast and forewarn pest and disease incidence. Also, models have been developed to forecast weather and climate.

Q.3 Explain what is meant by multi-collinearity in brief.

Multi-collinearity means there is no exact linear relationship between the explanatory variables or none of the regressors can be written as exact linear combinations of the remaining regressors in the models.

Q.4 Which of these models is considered to be the best for forecasting?

It depends on the data available. All these models are selected on the basis of comparative studies. The model with least MAPE values would be considered to be apt for a particular set of data.

6. References

Agrawal, R. and Jain, R. C. 1996. Forecast of sugarcane yield using eye estimate along with plant characters. *Biom. J*. 38(5): 731-739.

Agrawal, R. and Mehta, S. C. 2007. Weather based forecasting of crop yields, pests and diseases – IASRI models. *J.Indian Soc.Agric. Statist.* 61(2): 255-263.

Agrawal, R., Mehta, S. C., Bhar, L. M., and Kumar, A. 2004. Development of weather based forewarning system for crop pests and diseases

Draper, N.R. and Smith, H. 1981. *Applied RegressionAnalysis*, (2nd Ed.). John Wiley and sons, New York, 709p

Garde, Y. A., Singh, S., Mishra, G. C., and Singh, T. 2012. Weather based pre-harvest forecasting of wheat at Ghazipur (UP). *Indian J. Agric. Sci*. 8(2): 325-328.

Gomez, K. A. and Gomez, A. A. 1984. *Statistical Procedures in Agricultural Research*. (2nd Ed.). Wiley paperback, New York, 680 p.

Indraji, K. N. 2016. Price forecast models for coconut and coconut oil. M.Sc. (Ag. Stat.) thesis, Kerala Agricultural University, Thrissur, 123p.

Jain, R. C., Jha, M. P., and Agrawal, R. 1985. Use of growth indices in yield forecast. *Biom. J*. 27(4): 435-439.

Kumar, A., Kumari, V., and Agrawal, R. 2016. Use of ordinal logistic regression in crop yield forecasting. *Mausam 67*(4): 913-918.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. 1998. *Forecasting: Methods and Applications*. John Wiley, New York, 923p.

Misra, A. K., Prakash, O., and Ramasubramanian, V. 2004. Forewarning powdery mildew caused by *Oidium mangiferae* in mango (*Mangifera indica*) using logistic regression models. *Indian J. Agric. Sci*. 74(2): 84-87.

Singh, B. H. and Bapat, S. R. 1988. Pre-harvest forecast models for prediction of sugarcane yield. *Indian J. Agric. Sci*. 58(6): 465-469.

Varghese, R. 2011. Price behavior of natural rubber in India. M.Sc. (Ag.) thesis, Kerala Agricultural University, Thrissur, 109p.

KERALA AGRICULTURAL UNIVERSITY

COLLEGE OF HORTICULTURE, VELLANIKKARA

Department of Agricultural Statistics

STAT 591: Master's Seminar

Name	: Pooja A.	**Venue**: Seminar Hall
Admission No	: 2018-19-004	**Date** : 12-12-2019
Major Advisor	: Dr. Laly John C.	**Time** : 9.15 am

Statistical Forecast Models in Agriculture

Abstract

Forecasting is an integral part of agriculture for providing valuable information regarding storage, sales, export, import, *etc*. The use of statistical models in forecasting crop yield and prices of agricultural commodities holds great significance. Predicting the yield of a crop well in advance before harvesting will help farmers as well as planners to take timely decisions. Price forecast aids farmers to take decision to sell or stock their produce and the government to formulate suitable price policies.

For reliable forecast, valid and consistent data are essential. The data collected through systematic investigation need to be accurate, relevant and timely. An appropriate statistical model is then judiciously selected and fitted for the data. Finally, future values are predicted using past and present observations. Although no statistical model can help to make exact forecasts, even approximate values can help in formulating future plans.

Statistical forecast models in agriculture are broadly classified into regression and time series models. Multiple Linear Regression (MLR) technique has been extensively used in developing models for crop yield forecasting. Singh and Bapat (1988) developed pre-harvest model based on yield attributes using MLR, to forecast sugarcane yield about four months before harvesting. Usually, these models are developed by utilizing data at one point of time during the crop growth period (Agrawal and Jain, 1996). Garde *et al*. (2012) reported that reliable yield forecast for wheat can be made about two and half months before harvest using weather indices based MLR models.

Logistic regression models are widely used in situations where the response variable is qualitative. Logistic regression is used for obtaining probabilities of different categories.

Kumar *et al.* (2016) concluded that logistic regression model with "two group classification" can be recommended for forecasting wheat yield by the 11th week after sowing.

Time series forecast models use previously observed values to predict future values. Indraji (2016) demonstrated the possibility of exploiting the time series models like Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing models in forecasting the price of coconut and coconut oil respectively, for different markets of Kerala.

Artificial Neural Network (ANN) model has been accepted as an attractive tool for capturing non-linearity in time series data for forecasting purposes. Varghese (2011) used ANN model to forecast the price of natural rubber in India.

A sound database along with reliable forecasting is very much essential in agriculture. At present, a national research wing in forecasting is operating at Indian Agricultural Statistics Research Institute, New Delhi. Establishment of a similar forecasting division at state level can play a crucial role in this process.

References

Agrawal, R. and Jain, R. C. 1996. Forecast of sugarcane yield using eye estimate along with plant characters. *Biom. J*. 38(5): 731-739.

Garde, Y. A., Singh, S., Mishra, G. C., and Singh, T. 2012. Weather based pre-harvest forecasting of wheat at Ghazipur (UP). *Indian J. Agric. Sci*. 8(2): 325-328.

Indraji, K. N. 2016. Price forecast models for coconut and coconut oil. M.Sc. (Ag. Stat.) thesis, Kerala Agricultural University, Thrissur, 123p.

Kumar, A., Kumari, V., and Agrawal, R. 2016. Use of ordinal logistic regression in crop yield forecasting. *Mausam 67*(4): 913-918.

Singh, B. H. and Bapat, S. R. 1988. Pre-harvest forecast models for prediction of sugarcane yield. *Indian J. Agric. Sci*. 58(6): 465-469.

Varghese, R. 2011. Price behavior of natural rubber in India. M.Sc. (Ag.) thesis, Kerala Agricultural University, Thrissur, 109p.

26