IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Garden Pea (Pisum sativum) production: Forecasting model for Himachal Pradesh province in India

Bharath JPV, Ashu Chandel, RK Gupta, RS Chandel, Shivani Thakur, Dejen A., Pawan

Department of Basic Sciences, College of Forestry, Dr YS Parmar UHF, Nauni, Solan (173230)

Abstract

The present investigation was carried out on the time series data from 1996-2016 (21 years including 2016). The secondary data were collected from Directorate of Agriculture, 2018, Shimla, Himachal Pradesh, India. The Regression analysis and Autoregressive Models were fitted. However, based on Adjusted-R square ($\overline{R^2}$), Root Mean Square Error (RMSE), Theil's U statistic and F-chow statistic test the Quadratic model in regression analysis was found best fit for the estimation of production of garden pea. Whereas, on the basis of Akaike's Information Criterion (AIC), RMSE, Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) and Ljung-Box coefficient were selected ARIMA (2, 2, 0) as the best fit model for estimation of garden pea.

Keywords: Time Series, Regression Analysis, Autoregressive Models, Adjusted-R Square ($\overline{R^2}$), Root Mean Square Error (RMSE), Theil's U Statistic, F-Chow Statistic, Quadratic Model, Garden Pea (*Pisum Sativum*).

Introduction:

Vegetables are part of plants consumed as food by human beings. Vegetables are high in vitamins, minerals, dietary fiber and are cheap source of proteins. The women and children in developing countries are worst affected by the deficiency of vitamins and minerals (Samantaray *et al.*, 2016). Consumption of vegetables helps in reducing the incidence of cancer, stroke, cardiovascular disease and other chronic ailments. World Health Organization recommends daily consumption of 400 grams of vegetables and fruits daily.

The diverse agro-climatic conditions of India are conducive to cultivate a large array of vegetable crops. In Himachal Pradesh 87.31 thousand hectare of area is under vegetable cultivation with production of 1755.43 thousand tonnes (FAO, 2017). The garden pea (*Pisum sativum*) belongs to family Fabaceae of order Fabales. It is originated from Mediterranean region. These contain minerals such as magnesium, potassium and calcium. Pea is an important vegetable grown throughout the world. In India, it is grown mainly as a winter vegetable in the plains of North India and as a summer vegetable in the hills of other states. *Pisum sativum* the white flowered horticultural or garden pea also known as sweet pea. Garden pea is highly nutritive and rich source of digestible protein, carbohydrates and vitamins. It is used as a fresh vegetable or in soup, canned, processed or dehydrated

(Franklin, 2000). The area under cultivation in India is 554 thousand hectare, with production of 5524 thousand tons. In Himachal Pradesh 24.37 thousand hectares is under Peas cultivation with production of 294.96 thousand tons (NHB, 2018).

Vegetables are perishable in nature. Hence, forecasting the yield has great significance. Forecasting is the process of making predictions about the future by analysing the trend of time series data. This trend in production and productivity is valuable to the farmers, processing industries, banks and government. It is extremely useful in formulation of policies regarding the stock, distribution and supply of the produce to different areas in the country and outside the country. The forecasts with reasonable precisions before harvests help in timely decisions. Hence, the research paper was carried out to predict the production of garden pea in Himachal Pradesh.

Material and methods:

The study belongs to the state of Himachal Pradesh in India. Data were subjected to various regression models *viz.*, linear, quadratic, cubic and compound models. Based on the significance of autocorrelation of different lags, production of garden pea was predicted using autoregressive models. Time was taken as an independent variable for forecasting production. The model having significant $\bar{R}^2$, significant t-statistic for regression coefficient, lowest RMSE, lowest Theil's U statistic and non-significant F value in chow test was selected for forecasting of production of garden pea.

Model fitting

Regression Analysis:

It is a statistical technique for investigating and modeling the relationship between variables. It can be used to predict the value of the response variable in terms of the regressor variable. Regression analysis was performed with linear, quadratic, cubic and compound models.

Moreover, Autoregressive models of first, second and third orders were also used to study the trend in the data. ARIMA models were fitted for forecasting the future values of time series data.

By considering time as an independent variable, various linear and nonlinear regression models were used for prediction of production of garden pea:

(i) Linear Model: $Y_t = a + bt + e_t$

(ii) Quadratic Model: $Y_t = a + bt + ct^2 + e_t$

(iii) Cubic Model: $Y_t = a + bt + ct^2 + dt^3 + e_t$

(iv) Compound Model: $Y_t = ab^t + e_t$

where,

Y_t is the time series values of dependent variable (Production), t is time period, a is intercept, e_t is error term and b, c and d are regression coefficients

Autocorrelation:

Autocorrelation is a special case of correlation. It is the relationship, not between two or more different variables, but between the successive values of the same variable.

$$\rho_{(Y_t, Y_{t-k})} = \frac{cov.(Y_t, Y_{t-k})}{\sqrt{var(Y_t)}\ \sqrt{var.(Y_{t-k})}}$$

where,

$\rho_{(Y_t, Y_{t-k})}$ =Autocorrelation between Y_t and Y_{t-k}

$$\text{cov.}\,(Y_t, Y_{t-k}) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y}_t)\,(Y_{t-k} - \overline{Y}_{t-k})$$

$$\text{var}\,(Y_t) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y}_t)^2$$

$$\text{var}\,(Y_{t-k}) = \frac{1}{n}\sum_{t=1}^{n}(Y_{t-k} - \overline{Y}_{t-k})^2$$

Autoregressive Integrated Moving Average (ARIMA) Models:

ARIMA are such class of models that explain a given time series data based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. An ARIMA model is characterized by 3 terms: p, d and q. Here, p is the order of autoregressive terms, q is order of a moving average terms and d is the number of differencing required to make the time series stationary (Nochai and Nochai, 2006).

An ARIMA model is one where the time series will be differenced at least once to make it stationary and we combine the autoregressive and the moving average terms. So the equation becomes

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_1 + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_q \epsilon_{t-q}$$

Hence, ARIMA Predicted as $\boldsymbol{Y_t}$ = Constant + Linear combination lags of Y (upto p lags) + Linear Combination of lagged forecast errors (upto q lags).

Autoregressive Models:

Sometimes when the regression analysis consists of time series data, the lagged values of the response variable are considered as the regressors (independent variables). Based on the order of lagged values, various autoregressive models can be written as following:-

(i) First order autoregressive model: $Y_t = \phi_1 Y_{t-1} + e_t$

(ii) Second order autoregressive model: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$

(iii) Third order autoregressive model:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$$

where,

ϕ_1, ϕ_2, ϕ_3 = coefficients of the autoregressive models

Stationarities of time series data:

First step is to make the stationary series because autoregressive terms mean a regression model that uses its own lags as predictors. Linear regression models work best when the predictors are not correlated and are independent of each other. The most common approach is to difference the time series. Depending on the complexity of the series, more than one differencing may be needed. The value of d is the minimum number of differencing needed to make the series stationary and if the series is already stationary, then d = 0.

Model adequacy

Model adequacy investigates the fit of regression model to the available data. It includes residual analysis, R^2, $\overline{R^2}$ etc. Standard error and t- statistic were computed to test the significance of parameters of the regression models used in the study. Adjusted coefficient of multiple determination $\overline{R}^2$, root mean square error and Thiel's inequality coefficient had been used to check the adequacy of the fitted models.

Root Mean Square Error (RMSE):

The RMSE represents the standard deviation of the differences between predicted values and actual values. This statistic is also known as fit standard error and standard error of regression. It is an estimate of the standard deviation of error component of the data.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n-k}}$$

where,

MSE = Mean Sum of Squares due to Error

Y_t= actual value of the response variable

$\hat{Y}_t$= estimated value of the fitted model

Coefficient of multiple determination:

R^2 is the percentage of total variation in the dependent variable that can be explained by the independent variables,

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

where,

RSS = Regression Sum of Squares

TSS = Total Sum of Squares

ESS = Error Sum of Square

Adjusted coefficient of multiple determination:

In case of more than one regressor variables, R^2 sometimes misleads results regarding the fitting of model. So, instead of using R^2 we use $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{\frac{ESS}{(n-k-1)}}{\frac{TSS}{(n-1)}}$$

$$= 1 - \frac{(1-R^2)(n-1)}{(n-k)}$$

where,

ESS = Error Sum of Square

N = number of observations

K = number of variable

Theil's inequality coefficient (U)

A systematic measure of the accuracy of the forecasts has been suggested by H. Theil (1924 – 2000). This measure is called inequality coefficient and can be defined by the expression

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}A_i^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}P_i^2}}$$

where,

A_i= actual value of the dependent variable

n = number of observations

P_i = predicted value of the dependent variable

Note: When $P_i = A_i$ then U = 0 that means the model attains perfect forecast.

If P_i= 0 then U= 1 and the predictive power of model is worse. When U > 1 then it is preferable to accept zero change extrapolation that means $Y_{t+1} = Y_t$

Mean absolute error (MAE):

Absolute error is the amount of error in the measurements. It is the difference between measured value and true value.

$$MAE = \frac{1}{n}\Sigma|x_i - \bar{x}|$$

where,

n = the number of errors

$|x_i - \bar{x}|$ = the absolute error

Mean Absolute Percentage Error (MAPE):

It is the average of the sum of all the percentage errors for a given data set taken without regard to sign.

$$MAPE = \frac{1}{n}\Sigma\left|\frac{e_t}{Y_t}\right| \times 100$$

where,

Y_t = actual observation for time period t

e_t = error at time t

MAPE= mean absolute percentage error

n = time periods

Akaike's Information Criterion (AIC)

AIC compares the quality of a set of statistical models to each other. It is calculated using formula

$$AIC = 2(log - likelihood) + 2K$$

where,

K = number of model parameters

Log-likelihood = measure of model fit

Model validation

A model that fits well to the data helps in final application, but we are not always sure that the best fit model is also a good predictor or not. Some new factors negligible at the time of modeling may affect the response variable significantly in the future. In such situations the forecasting is almost useless. Validation is performed to determine if the model will function well in real situations or not. In the validation step, we investigate the stability of regression coefficients. Data splitting is one of the techniques for validation of a model. It has its significance when collection of new data for validation purpose is not possible or data collection is very costly. In case of time series data, the time may be used as a basis for data splitting. Half splitting may be used if we have no basis for splitting.

Chow test

If the variation in the regression coefficients is significant, then the model is not valid for prediction. Chow test is one of the approaches to test the significance of change in the regression coefficients over the time by using statistic.

$$F = \frac{\frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right]}{k}}{\frac{(\sum e_1^2 + \sum e_2^2)}{(n_1+n_2-2k)}}$$

where,

$\sum e_p^2$ = Pooled error variance

$\sum e_1^2$ = First half sample error variance

$\sum e_2^2$ = Second half sample error variance

n_1 = Number of observations in first sample

n_2 = Number of observations in second sample

k = Number of regression coefficients in the model

If F ≥ F_{tab}. For (k, $n_1 + n_2 - 2k$) , degrees of freedom then the F value is significant and the model is not valid for prediction.

Ljung – Box Test

This is a way to test for the absence of serial autocorrelation up to specified lag k. The null hypothesis is that our model does not show lack of fit i.e. the model is just fine. Alternative hypothesis is that the model shows lack of fit.

$$Q(m) = n(n+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j}$$

where,

r_j = the accumulated sample autocorrelations

m = the time lag

$\chi^2_{1-\alpha,k}$= chi-square table value at significance level α and k degrees of freedom.

If $Q(m) > \chi^2_{1-\alpha,k}$ infers that the model shows lack of fit and model is not suitable for the estimation.

Results and Discussion

1. Regression analysis of garden pea production

The estimation of production of garden pea was tried using different models. The best model among them was selected based on the different statistical criterion.

Linear, quadratic, cubic and compound models were presented in **Error! Reference source not found.** along with their statistical parameters and their significance. Adjusted R-square values range from 0.887 in linear to 0.969 in compound which suggests that all the models were best fit for the estimation of garden pea production.

Minimum root mean square error value was seen in quadratic model followed by compound, cubic and linear models. RMSE was 6236 in quadratic and 6318, 6446 and 11508 in compound, cubic and linear models respectively. F-chow test show that linear models was not suitable for estimation. Hence among the rest three models available quadratic model was best suited for the estimation of garden pea production. Quadratic models contain minimum value of RMSE and maximum values of $\bar{R}^2$ and lower Theil's U inequality coefficient.

Table 1.1. Statistical parameters curve estimation models for prediction of garden pea production.

Statistical Model	Regression Coefficients		Standard error	t - statistic	$\bar{R}^2$	RMSE	Theil's U	F (Chow test)
Linear	a	4905.081	5207.84	-0.942				
	b	5208.856	414.752	12.59*	0.887	11508	0.1768	10.233*
Quadratic	**a**	**8494.033**	**4784.57**	**3.865***				
	b	**-895.261**	**1001.76**	**-0.894**				
	c	**277.46**	**44.223**	**6.274***	**0.963**	**6236**	**0.0987**	**0.94**
Cubic	a	3014.531	6957.44	3.871*				
	b	1786.735	2675.28	0.668				
	c	-20.339	279.835	-0.073				
	d	9.024	8.353	1.080	0.963	6446	0.0955	1.38
Compound	a	3070.785	702.629	18.60*				
	b	1.113	0.005	233.5*	0.969	6318	0.0967	2.49

Hence quadratic model with equation $\hat{Y}_t = 18494.033 - 895.261(t) + 277.46(t^2)$ where t = given year – 1996 was used for the estimation of pea production in Himachal Pradesh.

Table 1.2. Trend in production of garden pea predicted using curve estimation model.

Year	Garden pea production	Predicted production of garden pea			
		Linear	**Quadratic**	**Cubic**	**Compound**
1996	13600	303.7749	14876.23	14789.95	14549.64
1997	14550	5512.631	16813.35	16578.84	16195.82
1998	20900	10721.49	18305.39	18435.34	18028.25
1999	23650	15930.34	19352.35	20413.59	20068.01
2000	23640	21139.2	20954.23	22567.75	22338.55
2001	24340	26348.05	23111.02	24951.96	24865.98
2002	24980	31556.91	25822.74	27620.36	27679.37
2003	30435	36765.77	29089.38	30627.1	30811.08
2004	29328	41974.62	32910.93	34026.33	34297.11
2005	34445	47183.48	37287.41	37872.18	38177.56
2006	53103	52392.33	42218.81	42218.81	42497.06
2007	52746	57601.19	47705.12	47120.35	47305.27
2008	53226	62810.05	53746.36	52630.96	52657.49
2009	54512	68018.9	60342.51	58804.79	58615.28
2010	56640	73227.76	67493.59	65695.96	65247.14
2011	60701	78436.61	75199.58	73358.64	72629.35
2012	91640	83645.47	83460.5	81846.97	80846.79
2013	101710	88854.32	92276.33	91215.08	89993.98
2014	100071	94063.18	101647.1	101517.1	100176.1
2015	117012	99272.04	111572.8	112807.3	111510.2
2016	119010	104480.9	122053.3	125139.6	124126.8

The estimated values of garden pea production from year 1996 to 2016 were presented in **Error! Reference source not found.** It was observed in table that the predicted values from quadratic model were nearer to actual values than any other models.

Autoregressive models for garden pea production:

Autocorrelation coefficients along with standard error were presented in **Error! Reference source not found.**. As it shows that there was significant correlation coefficient up to first four lags. Maximum autocorrelation was found in first lag with value of 0.834 and followed by subsequent lags with values of 0.652, 0.518, 0.352 in second, third and fourth lag respectively.

Statistical parameters of first three fitted autoregressive models were presented in the **Error! Reference source not found.**. It was clearly visible that all the three models have at least one significant coefficient. Table also includes standard error, t – statistic, $\overline{R}^2$, root mean square error, Theil's U inequality constant and F chow test values. $\overline{R}^2$ value ranges from 0.91 to 0.913. Highest $\overline{R}^2$ was found in second autoregressive model followed by first and third.

Table Error! No text of specified style in document.**.1. Autocorrelation of different lags for pea production.**

Lag	Autocorrelation	Standard Error
1	0.834	0.203 *
2	0.652	0.198 *
3	0.518	0.193 *
4	0.352	0.188 *
5	0.206	0.182
6	0.141	0.176
7	0.061	0.170
8	-0.115	0.164

Table Error! No text of specified style in document.**.2.** Statistical parameters of first, second and third order autoregressive model for estimation of garden pea production.

Autoregressive Models	Coefficients		SE	t - statistic	$\overline{R}^2$	RMSE	Theil's U	F(Chow test)
First order	Φ_1	0.993	0.041	24.038 *	0.910	9794.89	0.154	0.9961
Second order	Φ_1	**1.027**	**0.227**	**4.527***	**0.913**	**9612.39**	**0.1513**	**0.983**
	Φ_2	**-0.031**	**0.237**	**-0.132**				
Third order	Φ_1	0.971	0.283	3.431*	0.904	10159.5	0.1587	0.385
	Φ_2	-0.091	0.442	-0.206				
	Φ_3	0.111	0.336	0.331				

According to the $\overline{R}^2$ value all the models were found good fit. Since there were no significant values of F-chow test no models need to be rejected. So, the selection of models was based on the root mean square error and Theil's U inequality coefficient. In second order Root mean square error was minimum with values 9612.39 followed by first and third order models with values 9794.89 and 10159.5 respectively. Theil,s U statistic was minimum in second order model followed by first order and third order autoregressive models. Hence second order autoregressive model was chosen over first and third order models for the prediction of garden pea production.

The equation $\widehat{Y}_t = 1.027(Y_{t-1}) - 0.031(Y_{t-2})$ was used for the prediction of garden pea production in Himachal Pradesh. The predicted and actual values were given in **Error! Reference source not found.**

Table Error! No text of specified style in document.**.3.** Trend in the production of garden pea using autoregressive models.

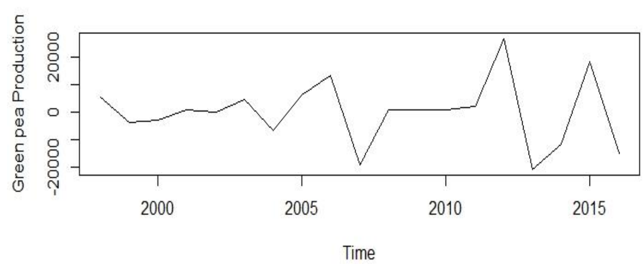
Year	Garden Pea production	Predicted garden pea production		
		First AR	Second AR	Third AR
1996	13600	-	-	-
1997	14550	13504.62	13545.1	13455.3
1998	20900	14447.95	14519.0	14381.7
1999	23650	20753.42	21010.9	20473.0
2000	23640	23484.13	23637.2	22670.3
2001	24340	23474.2	23541.2	23115.3
2002	24980	24169.29	24260.5	24100.8
2003	30435	24804.8	24895.9	24657.3
2004	29328	30221.54	30478.3	29971.8
2005	34445	29122.31	29171.4	28472.2
2006	53103	34203.42	34461.1	34145.2
2007	52746	52730.56	53463.4	51667.7
2008	53226	52376.06	52515.3	50192.2
2009	54512	52852.7	53019.4	52761.3
2010	56640	54129.68	54325.1	53926.3
2011	60701	56242.75	56470.5	55928.2
2012	91640	60275.27	60574.9	59819.3
2013	101710	90997.28	92222.8	89717.7
2014	100071	100996.7	101600.	97129.3
2015	117012	99369.15	99603.3	98056.2
2016	119010	116191.3	117053	115767

Error! Reference source not found. contains the predicted values of production of garden pea from yeas 1996 to 2016. In this table it was clearly visible that the predicted values from second order autoregressive model were nearer to actual values than any other order.

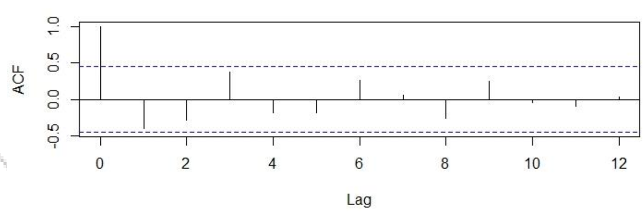
ARIMA for garden pea production

Time series plot presented in **Error! Reference source not found.** looks stationary after double differencing. Plot contains few outliers. Plot was presented. The ACF and PACF plots of the differenced time series were presented in **Error! Reference source not found.** and **Error! Reference source not found.** respectively. The ACF plot show no significant spike which suggests that there were no moving average terms. In the PACF plots there was a significant spike at lag 2 which suggest there might be a autoregressive term.

Plot 1. Time series plot of differenced garden pea production.



Plot 2. **Autocorrelation of garden pea production at different lags**



ARIMA models of order (0, 2, 0), (1, 2, 0) and (2, 2, 0) were fitted to the time series. Models along with their AIC, RMSE, MAE, MAPE and Ljung-Box coefficient **were** presented in **Error! Reference source not found..Error! Reference source not found.**

There was a significant autocorrelation in Ljung-Box coefficient which suggests that ARIMA (0, 2, 0) was not suitable for the prediction of garden pea production. Among the ARIMA models presented ARIMA (2, 2, 0) had minimum AIC. RMSE, MAE and MAPE were also minimum in ARIMA (2, 2, 0). So, ARIMA (2, 2, 0) was best suited for prediction of garden pea production in Himachal Pradesh.

Plot 3. Partial autocorrelation of garden pea production.

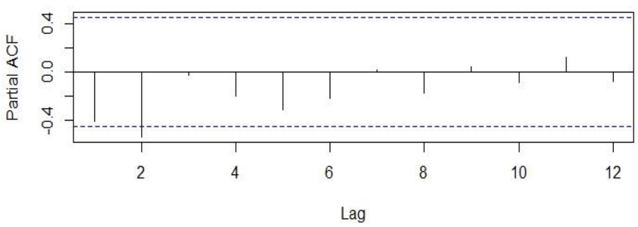


Table 3.1. Diagnostic tools for model selection criterion of garden pea production.

Model	AIC	RMSE	MAE	MAPE	Ljung-Box coefficient
ARIMA (0 , 2, 0)	411.647	11061.5	7624.10	12.931	8.103*
ARIMA (1, 2, 0)	410.124	10030.6	6848.10	11.726	5.351
ARIMA **(2, 2, 0)**	405.849	8351.86	5483.73	10.348	1.698

There was a significant autocorrelation in Ljung-Box coefficient which suggests that ARIMA (0, 2, 0) was not suitable for the prediction of garden pea production. Among the ARIMA models presented ARIMA (2, 2, 0) had minimum AIC. RMSE, MAE and MAPE were also minimum in ARIMA (2, 2, 0). So, ARIMA (2, 2, 0) was best suited for prediction of garden pea production in Himachal Pradesh.

The autoregressive and moving average coefficients of ARIMA (2, 2, 0) were presented in the **Error! Reference source not found.**

Table Error! No text of specified style in document.**.** Coefficients of ARIMA (2, 2, 0) model for garden pea production.

Model	coefficients		Standard error	t-statistic
ARIMA (2, 2, 0)	AR1	-0.6077	0.1909	3.183*
	AR2	-0.5472	0.1909	2.866*

Table Error! No text of specified style in document. Predicted values of garden pea production using ARIMA (2, 2, 0)

Year	Garden pea production	Predicted values	**Year**	Garden pea production	Predicted values
1996	13600	**13593.92**	**2007**	52746	**60126.05**
1997	14550	**14566.12**	**2008**	53226	**56535.83**
1998	20900	**16743.41**	**2009**	54512	**63601.81**
1999	23650	**24887.85**	**2010**	56640	**54850.18**
2000	23640	**25633.12**	**2011**	60701	**57815.27**
2001	24340	**27277.17**	**2012**	91640	**63126.53**
2002	24980	**26118.70**	**2013**	101710	**105186.59**
2003	30435	**25267.97**	**2014**	100071	**109755.93**
2004	29328	**32996.58**	**2015**	117012	**116966.92**
2005	34445	**29574.33**	**2016**	119010	**129068.11**
2006	53103	**39370.00**			

Fit was done using ARIMA (2, 2, 0). Predicted values and actual values of the production under garden pea from year 1996 to 2016 were presented in **Error! Reference source not found.**

Summary

For the production of garden pea models *viz.*, linear, quadratic, cubic and compound were fitted. Linear models showed lack of stability and quadratic model was selected because of the high $\bar{R}^2$ and lowest RMSE values. So, the equation $\hat{Y}_t = 18494.033 - 895.261(t) + 277.46(t^2)$ is used to estimate the garden pea production. Where, $\widehat{Y}_t$ = values of independent variable (production) for year t and t = given year – 1996. Significant autocorrelation up to four lags were shown but significant coefficients were found up to third order only. All the models were found well fitted with significant coefficient. But second order was chosen over first and third order because of highest $\overline{R}^2$ and lowest RMSE values. Based on the significant spikes in ACF and PACF plots of stationery time series ARIMA models of order (0, 2, 0), (1, 2, 0) and (2, 2, 0) were fitted. Based on AIC, RMSE and MAPE, ARIMA (2, 2, 0) was found best fit. ARIMA (0, 2, 0) showed significant Ljung-Box coefficient not suitable for estimation. Based on the RMSE, MAE and MAPE values of all the models ARIMA (2, 2, 0) was found best fit. Hence ARIMA (2, 2, 0) is selected as best for the forecasting of garden pea production in Himachal Pradesh.

References:

Directorate of Agriculture. 2018. Directorate of Agriculture, Shimla - 5 (HP).

FAO. 2017. *Food and Agriculture Organization.* FAOSTAT. http:// www. fao. org/ foastat / en / #data/QC.

Fatoki O, UA Mbata, GA Olulude and O Abass. 2010. An application of ARIMA model to Nigeria gross domestic production. *International Journal of Statistics and Systems* **5**: 310-320.

Gupta D, Sahu PK and Banerjee R. 2009. Forecasting jute production in major contributing countries in the world. *Journal of Natural Fibers* **6**: 127-137.

Suleman N and Sarpong S. 2012. Forecasting milled rice production in Ghana using box-jenkins approach. *International Journal of Agricultural Management and Development* **2**: 79-84.

Suresh KK and Priya SK. 2011. Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Technology* **13**: 23-26.

Tripathi R, Nayak AK, Raja R, Shahid M, Kumar A, Mohanty S, Panda BB, Lal B and Gautam P. 2014. Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models. *Advances in Agriculture* **3**: 54-58.

Vishawajith KP, Sahu PK, Dhekale BS and Mishra P. 2016. Modelling and forecasting sugarcane and sugar production in India. *Indian Journal of Economics and Development* **12**: 71-80.

Wali VB and Lokesh DBH. 2017. Forecasting of area and production of cotton in India and Karnataka using ARIMA model. *Indian Journal of Economics and Development* **13**: 723-728.

Kumar A. 2015. Statistical Models for Forecasting the Production of Important Fruit and

Vegetable Crop in Himachal Pradesh. M.Sc. Thesis. Department of Basic Sciences, College of

Forestry, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Nauni, Solan. 53p.