Original Research Article

Statistical Modeling for Analysis of Growth and TrendPattern of WheatProduction in Selected States of India

ABSTRACT

In the present paper, the time series analysis of wheat production in some selected states of India has been carried out by fitting statistical models, viz. linear, exponential and cubic models. The secondary time series data on the production of wheat have been utilized for the analysis. The trend values have been evaluated by fitting the concerned models, and the validity of the models has been tested by using the Chi-square test statistic. Moreover, the coefficient of determination(R^2), root mean square error (RMSE), and relative mean absolute percentage error (RMAPE) have been computed to reveal the suitability of the concerned models for exploring the trend patterns of wheat production in the concerned states of India.

Comment [u1]: What are the main assumptions to be considered in the various location to determine the area as a potential for wheat production? For instance increase temperature, rainfall, drought etc. On which one were your model based. It is clear you did you various model?

Keywords: Time series, Linear model, Exponential model, Cubic model,Chi-square test,Coefficient of determination.

1. INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most significant cereal crops, and the second most important staple food crop in India after rice. It belongs to Gramineae(or Poaceae) family, and is a rich source of carbohydrate, protein, multi-vitamins and other vital nutrients.Wheat is mainly consumed in processed form, for instance, bread, biscuits, cookies, noodles,porridge, pudding, pasta,vermicelli, and so on. In India,mostly three species of wheat are cultivated i.e., *Triticum aestivum* (common wheat or bread wheat), *Triticum durum* (macroni or durum wheat) and *Triticumdicoccum* (emmer wheat).

In India, the largest area under coverage of wheat species is *Triticumaestivum* (common wheat or bread wheat), followed by *Triticum durum* (macaroni or durum wheat) and *Triticumdicoccum*(emmer wheat). The wheat species *Triticumaestivum is*commonly grown in all six agro-climatic zones of the country viz., Northern Hill Zone (NHZ), North West Plain Zone (NWPZ), North East Plain Zone (NEPZ), Central Zone (CZ), Peninsular Zone (PZ) and Southern Hill Zone (SHZ) (Joshi *et al.*[1]).

India is the second largest producer of wheat after China. In India, the leading state in the production of wheatwas Uttar Pradesh (35.50 million tonnes) during the year 2020-21, followed by Madhya Pradesh (17.62 million tonnes), Punjab (17.14 million tonnes), Haryana (12.36 million tonnes), Rajasthan (11.04 million tonnes), and Bihar (6.34 million tonnes). In India, the overall production of wheat was 109.52 million tonnes, and the wheat yield was 3.464 tonnes per hectare (Source: Directorate of Economics & Statistics, DAC&FW, Govt. of India [2]).

The agriculture sector is deemed to be a significant sector as it contributes towards the economic development of a nation. In recent years, there is significant rise in foodgrain production due to noteworthy developments in agriculture sector and evolution of high-yielding variety of seeds. The time series analysis of agricultural crops is of utmost importance for exploring the long-term trend pattern of the crops, and policy formulation regarding inventory management, price fixation, and transportation of the crops. Considering this fact, the trend analysis of crops has been dealt by various authors.Boken [3] applied time series analysis to forecast yield of spring wheatfor Saskatchewan, Canada by using well-known statistical models (viz. linear trend, quadratic trend, simple exponential smoothing, double exponential smoothing, simple moving averaging, and double moving averaging). Arunachalam and Balakrishnan [4] investigated the trends in area, production and productivity of wheat in India by utilizing non-linear as well as non-parametric regression models. Michel and Makowski [5] presented eight statistical models for analyzing wheat yield time series and predicted wheat yield at the national and regional scales, on utilizing data obtained through the Food and Agriculture Organization of the United Nations and the French Ministry of Agriculture.Dasyam*et al.*[6] modeled and forecasted the production of wheat in India by using parametric regression, exponential smoothing and Auto Regressive Integrated Moving Average (ARIMA) models. Ray *et al.*[7] proposed a hybrid model by combining Autoregressive Integrated Moving Average (ARIMA) and Wavelet Neural Network (WNN).Polisetty and Paidipati[8] examined the change point and trend analysis of wheat production in India using non-parametric methods viz. Pettitt's, Standard normal homogeneity (SNH) and Buishand's range tests. Yonar et al. [9] modeled and forecasted the production of wheat in South Asian region countries, viz. Afghanistan, Bangladesh, Bhutan, China, India, Nepal, and Pakistan, on utilizing ARIMA and Holt's linear trend models. Rao and Naidu [10] applied various non-linear models to forecast the area, production and productivity of wheat crop in India. Madhukar *et al*.[11] analyzed the temperature and precipitation trends and their impact on wheat yield across 29 Indian states using statistical methods.Some other noteworthy contributions towards time series analysis of crops have been made by Rajarathinam*et al.*[12], Tripathi *et al.*[13], Joshi *et al.* [14], Kumar and Menon [15], Paudel et al. [16], and Rana and Kumar [17].

In this paper, an attempt is made to explore the trend of wheat production in some selected states of India. The analysis is done using well-known statistical models viz. linear, exponential and cubic models. The accuracy of the concerned models has been evaluated using coefficient of determination (R^2), root mean square error (RMSE) and relative mean absolute percentage error (RMAPE). Moreover, the validity of models has been examined using chi-square (χ^2) test of "Goodness of Fit".

2. DATA AND METHODOLOGY

2.1 Source of Data

The present paper deals with the analysis of secondary time series data on wheat production pertaining to the period (2011-2020) in some selected states of India. The time series data is obtained through the records ofDirectorate of Economics & Statistics, DAC&FW, Govt. of India.

2.2 Terminologies and Notations

In the present analysis, we have considered three wheat growing states of India, viz. Uttar Pradesh (S1), Haryana (S2), and Bihar (S3). In these states, we observe various trends of wheatproduction during the concerned period of study.

2.3 Fitting of Statistical Models to the Data

In order to analyze the growth and trend patterns of wheat production in the concerned states S1, S2 and S3, we compute the trend values by fitting linear, exponential and cubic models to the time series data on wheat production as follows:

(a) Linear Model:

$$y_t = a + bt \text{…………………………} (1)$$

where y_t denotes the time series value at time t.The values of constants 'a' and 'b' are obtained on using the principle of least squares by solving the following normal equations:

$$\sum y_t = na + b \sum t \text{……………………………} (2)$$

$$\sum t y_t = a \sum t + b \sum t^2 \text{………………………} (3)$$

where 'n' represents the number of observed values.

(b) Exponential Model:

$$y_t = ae^{bt} \text{………………………………} (4)$$

Taking natural log on both sides of above equation, we have

$$log_e y_t = log_e\, a + bt\, log_e\, e$$

i.e., $Y_t = A + bt$ ………………………. (5)

where $Y_t = log_e y_t$, $A = log_e\, a$, and $log_e\, e = 1$

The normal equations for estimating the values of 'A' and 'b' are as follows:

$$\sum Y_t = nA + b \sum t \text{……………………..…….} (6)$$

$$\sum t Y_t = A \sum t + b \sum t^2 \text{…………………....} (7)$$

Finally, the value of 'a' is obtained on using

$$a = antilog\,(A)$$

(c) Cubic Model:

$$y_t = a + bt + ct^2 + dt^3 \text{…………………………} (8)$$

The values of constants 'a', 'b','c' and 'd' are obtained on solving the following normal equations.

$$\sum y_t = na + b\sum t + c\sum t^2 + \mathrm{d}\ \sum t^3 \text{……................. (9)}$$

$$\sum ty_t = a\sum t + b\sum t^2 + c\sum t^3 + d\sum t^4 \text{……................. (10)}$$

$$\sum t^2y_t = a\sum t^2 + b\sum t^3 + c\sum t^4 + d\sum t^5 \text{……................. (11)}$$

$$\sum t^3y_t = a\sum t^3 + b\sum t^4 + c\sum t^5 + d\sum t^6 \text{……................. (12)}$$

3. DATA ANALYSIS AND RESULTS

The secondary time series data on wheat production in states S1, S2 and S3 of India is presented in Table 1. The trend values are obtained on fitting linear, exponential andcubic models to the data in the concerned states, and are depicted in Tables 2, 3 and 4, respectively. Moreover, the model equations for linear, exponential and cubic trends in the respective states are elaborated in Table 5.

Table1: Time series data on wheat productionin selected states of India

Year	*Production (in million tonnes) for the states		
	S1	**S2**	**S3**
2011	30.29	12.68	4.79
2012	30.30	11.12	5.36
2013	30.25	11.80	5.08
2014	22.42	10.35	3.99
2015	25.43	11.35	4.74
2016	30.06	11.55	5.11
2017	31.88	11.16	5.74
2018	32.74	12.57	6.47
2019	33.82	11.88	5.58
2020	35.50	12.36	6.34

(*Source:Directorate of Economics & Statistics, DAC&FW,Govt. of India)

Table 2: Trend values for linear, exponential andcubicmodels in state S1

Year (t)	Production(y_t)	Trend Values		
		Linear Model(L_t)	Exponential Model (E_t)	Cubic model (C_t)
2011	30.29	27.08	29.91	31.80

2012	30.30	27.79	30.61	28.49
2013	30.25	28.50	31.32	26.84
2014	22.42	29.21	32.05	26.54
2015	25.43	29.91	32.80	27.27
2016	30.06	30.62	33.57	28.72
2017	31.88	31.33	34.35	30.59
2018	32.74	32.04	35.16	32.56
2019	33.82	32.75	35.98	34.32
2020	35.50	33.46	36.82	35.56

Table 3: Trend values for linear, exponential and cubicmodels in state S2

Year (t)	**Production (y_t)**	**Trend Values**		
		Linear Model (L_t)	**Exponential Model (E_t)**	**CubicModel(C_t)**
2011	12.68	11.44	10.90	12.54
2012	11.12	11.49	10.95	11.60
2013	11.80	11.55	11.00	11.11
2014	10.35	11.60	11.05	10.96
2015	11.35	11.66	11.11	11.07
2016	11.55	11.71	11.16	11.34
2017	11.16	11.76	11.21	11.69
2018	12.57	11.82	11.27	12.02
2019	11.88	11.87	11.32	12.23
2020	12.36	11.93	11.38	12.25

Table 4: Trend values for linear, exponential and cubicmodels in state S3

Year (t)	**Production (y_t)**	**Trend Values**		
		Linear Model (L_t)	**Exponential Model (E_t)**	**Cubic Model (C_t)**
2011	4.79	4.55	5.31	5.16
2012	5.36	4.72	5.47	4.78

2013	5.08	4.89	5.65	4.65
2014	3.99	5.06	5.83	4.71
2015	4.74	5.24	6.02	4.91
2016	5.11	5.41	6.21	5.21
2017	5.74	5.58	6.41	5.55
2018	6.47	5.75	6.61	5.87
2019	5.58	5.92	6.82	6.12
2020	6.34	6.09	7.04	6.25

Table 5: Model equations for linear, exponential and cubic trends in selected states of India

States	Linear Model	Exponential Model	Cubic Model
S1	$y_t = 29.914 + 0.709\,t$	$y_t = 2\text{E} - 19e^{0.0231\,t}$	$y_t = 27.265 + 1.144\,\text{t} + 0.362t^2 - 0.052t^3$
S2	$y_t = 11.655 + 0.054\,t$	$y_t = 7\text{E} - 4e^{0.0048\,t}$	$y_t = 11.068 + 0.207\,\text{t} + 0.083\,t^2 - 0.015t^3$
S3	$y_t = 5.235 + 0.170\,t$	$y_t = 2\text{E} - 27e^{0.0314\,t}$	$y_t = 4.913 + 0.260\,\text{t} + 0.046\,t^2 - 0.009t^3$

In Tables 2, 3 and 4, the term ‘y_t’ denotes the observed value of wheat production(in million tonnes)for the year ‘t’($t = 2011,\ 2012,\ \ldots\ ,2020$). Moreover, ‘$L_t$’denotes the linear trend value of wheat production for the year ‘t’. In a similar manner,‘E_t’denotes the exponential trend value of wheat production, and‘C_t’ denotes the cubic trend value of wheat production.

In order to illustrate the relative influence of linear, exponential and cubic trend values on the observed values of wheat production for the states S1, S2 and S3, the graphical plots are obtained and demonstrated in Figs. 1 to 9.

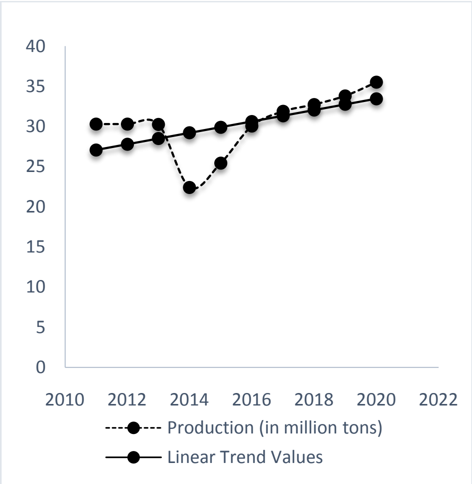


Fig.1: Trend values for Linear Model in state S1

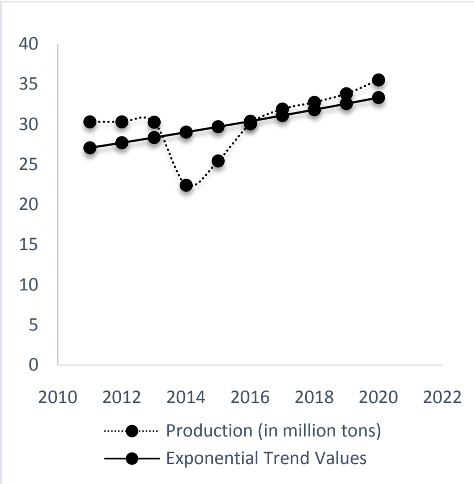


Fig.2: Trend values for Exponential Model in state S1

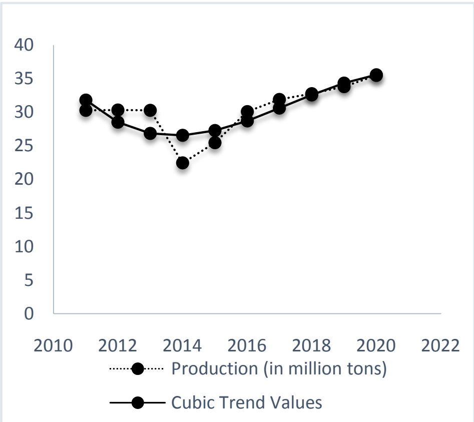


Fig.3: Trend values for Cubic Model in state S1

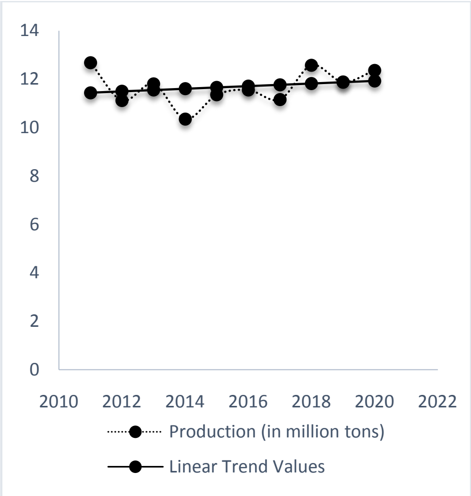


Fig. 4: Trend values for Linear Model in state S2

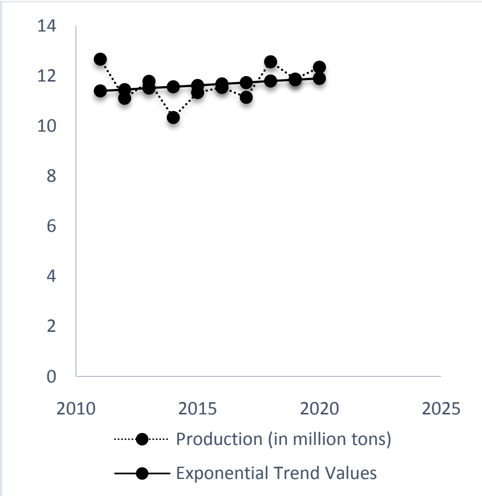


Fig. 5: Trend values for Exponential Model in state S2

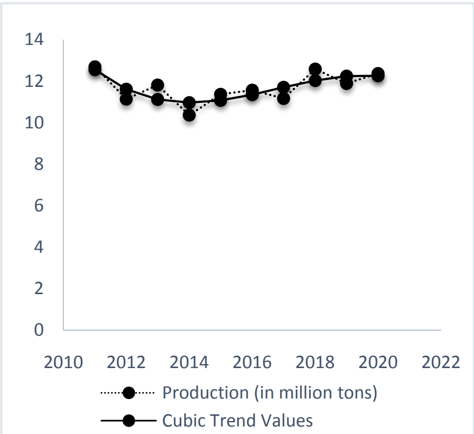
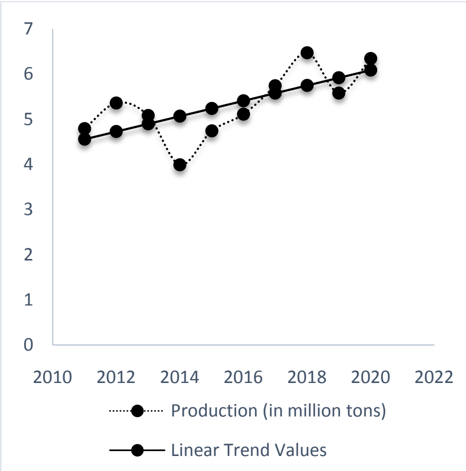


Fig. 6: Trend values for Cubic Model in state S2



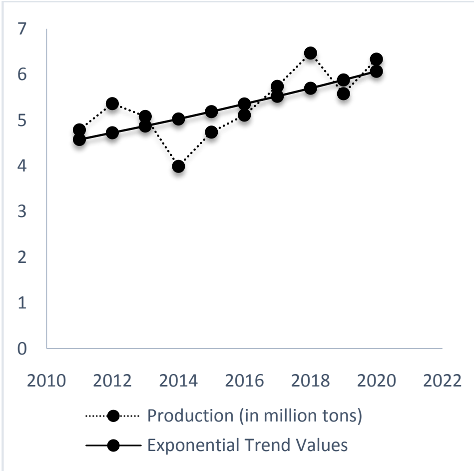


Fig.7: Trend values for Linear Model in state S3

Fig.8: Trend values for Exponential Model in state S3

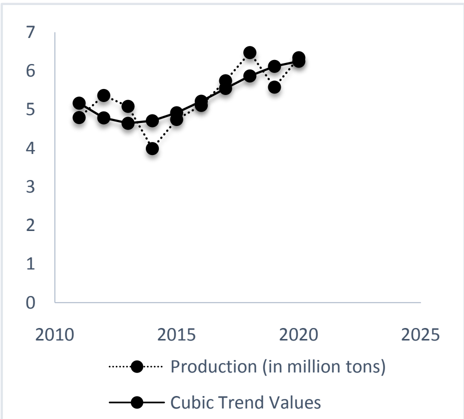


Fig.9: Trend values for Cubic Model in state S3

In order to test the suitability of various fitted models, we have computed the coefficient of determination (R^2), Root Mean Square Error (RMSE) and Relative Mean Absolute Percentage Error (RMAPE) for the selected states, by using the following formulae:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

and

$$RMAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100$$

where y_t denotes the observed value of wheat production (Y), and$\bar{y}$ is the mean value of the variable Y.Also,$\hat{y}_t$ is the trend value of the variable Y, which is obtained on fitting the respective statistical model (such as linear or exponential or cubic model, as the case may be) to the variable Y.

The values of R^2, RMSE and RMAPE for the concerned statesare obtainedon fitting linear, exponential and cubic models, and presented in Table 6.

Table 6: Model evaluation for wheat production in selected states of India

States	Models	R^2	RMSE	RMAPE
S1	Linear	0.31	3.04	8.72
	Exponential	0.33	3.00	8.79
	Cubic	0.69	2.03	5.85
S2	Linear	0.05	0.67	4.64
	Exponential	0.05	0.67	4.62
	Cubic	0.59	0.44	3.45
S3	Linear	0.47	0.52	8.77
	Exponential	0.49	0.51	8.60
	Cubic	0.62	0.44	7.46

From Table 6, the following results are obtained:

(i) In each of the three states S1, S2 and S3, the values of R^2 are more for the cubic model as compared to the linear and exponential models. Moreover, the values of R^2 are nearly the same for both linear and exponential models in each state.

(ii) In each states S1, S2 and S3, we observe that$R^2 > 0.5$for the cubic model, whereas $R^2 < 0.5$ for the linear and exponential models. Hence, among all the three models, cubic model is the best fitted model.

(iii) In each states, the values of RMSE are least for cubic model as compared to the linear and exponential models. Furthermore, the values of RMSE are nearly the same for both linear and exponential models.

(iv) In each states, the values of RMAPE are least for cubic model as compared to the linear and exponential models. Also, the values of RMAPE are approximately the same for both linear and exponential models.

3.1 Formulation of Hypotheses

We test the following null hypotheses:

H_{0L}: Linear model fits the given data on wheat production.

H_{0E}: Exponential model fits the given data on wheat production.

H_{0C}: Cubic model fits the given data on wheat production.

against the following respective alternative hypotheses:

H_{1L}: Linear model does not fit the given data on wheat production.

H_{1E}: Exponential model does not fit the given data on wheat production.

H_{1C}: Cubic model does not fit the given data on wheat production.

The above mentioned hypotheses for model fitting on wheat production are testedusing chi-square test statistic,in theconcernedstates S1, S2 and S3 of India.

3.2 Hypotheses Testing and Validation

The chi-square valueshave been computed for the linear, exponential and cubic models (i.e., χ_L^2 , χ_E^2 and χ_C^2) in the concerned states of India, and the findings are depicted in Table 7. The chi-square values,on fitting the concerned models, have been obtained using the following formulae:

$$\chi_L^2 = \sum_{t=1}^{n} \frac{(y_t - L_t)^2}{L_t} = \sum_{t=1}^{10} \frac{(y_t - L_t)^2}{L_t},$$

$$\chi_E^2 = \sum_{t=1}^{n} \frac{(y_t - E_t)^2}{E_t} = \sum_{t=1}^{10} \frac{(y_t - E_t)^2}{E_t},$$

$$\chi_C^2 = \sum_{t=1}^{n} \frac{(y_t - C_t)^2}{C_t} = \sum_{t=1}^{10} \frac{(y_t - C_t)^2}{C_t},$$

where the terms 'y_t', 'L_t', 'E_t' and 'C_t' have been utilized from the Tables 2, 3 and 4, for the concerned states S1, S2 and S3 of India.

Table 7: Values of chi-square statistic on fitting linear, exponential and cubicmodels

States	Chi-square values		
	Linear Model (χ_L^2)	**Exponential Model (χ_E^2)**	**CubicModel (χ_C^2)**
S1	3.1593	3.1048	1.5069
S2	0.3921	0.3925	0.1704
S3	0.5216	0.5087	0.3733

The tabulated values of chi-square (χ^2) at 1% and 5% levels of significance with 9 degrees of freedom are given, respectively, by

$$\chi^2_{0.01,9} = 21.67 \text{ and } \chi^2_{0.05,9} = 16.92$$

From Table 7, the following results are obtained:

(i) $\chi^2_{L(S_i)} < \chi^2_{0.01,9}$ and $\chi^2_{L(S_i)} < \chi^2_{0.05,9}$ $(i = 1,2,3)$

(ii) $\chi^2_{E(S_i)} < \chi^2_{0.01,9}$ and $\chi^2_{E(S_i)} < \chi^2_{0.05,9}$ $(i = 1,2,3)$

(iii) $\chi^2_{C(S_i)} < \chi^2_{0.01,9}$ and $\chi^2_{C(S_i)} < \chi^2_{0.05,9}$ $(i = 1,2,3)$

Hence, on the basis of above results, the null hypotheses H_{0L}, H_{0E} and H_{0C} are acceptedat 1% and 5% levels of significance. So, we conclude that the linear, exponential and cubic models fit the given time series data on wheat production for the concerned states S1, S2 and S3 of India.

4. DISCUSSION AND CONCLUSION

The present paper deals with time series analysis of wheat production in some selected states of India. The secondary time series data on wheat productionpertaining to the period (2011-2020) have been utilized for the analysis. The growth and trend patterns of wheat production have been examined by fitting well-known statistical models, viz. linear model, exponential model and cubic model to the concerned time series data for selected states of India.

It has been observed from the empirical results of section 3 that thecubic model is more precise and suitable, as compared to the linear and exponential models, for exploring the trends of wheat production in the concerned states S1 (Uttar Pradesh), S2 (Haryana)and S3 (Bihar) of India.The growth patterns of wheat production in the states S1 (Uttar Pradesh)and S3 (Bihar) is slightly increasing. Moreover, in the state S2 (Haryana), we observe a constant growth pattern of wheat production.

In order to test the "Goodness of Fit" of the linear, exponential andcubic models for the states S1, S2 and S3, the chi-square test statistic values (i.e., χ_L^2, χ_E^2 and χ_C^2) have been computed for the respective states. These values are then compared with the tabulated values of chi-square at 1% and 5% levels of significance. It has been observed that all the considered models fit the given time series data on wheat production for the concerned states.

The present study could be enhanced further by considering the scenario of wheat production in the other states of India. Moreover, on considering the benefits and usefulness of wheat, the potential farmers could be encouraged for its cultivation.

References

1. Joshi, A., Kumar, A. and Kashyap, S. (2020). Genetic analysis of yield and yield contributingtraits in bread wheat. *International Journal of Agriculture, Environment and Biotechnology*, 13(2): 119-128.

2. Anonymous (2020). Agricultural Statistics at a Glance. *Directorate of*

Economics&Statistics, DAC&FW, Government of India.

3. Boken, V.K. (2000). Forecasting spring wheat yield using time series analysis: A case study for the Canadian Prairies. *Agronomy Journal*, 92(6): 1047-1053.
4. Arunachalam, R. and Balakrishnan, V. (2012). Statistical modeling for wheat (*Triticum aestivum*) crop production. *International journal of statistics and applications,*2(4): 40-46.
5. Michel, L. and Makowski, D. (2013). Comparison of statistical models for analyzing wheat yield time series. *PLoS One,* 8(10): 1-11.
6. Dasyam, R., Pal, S., Rao, V.S. and Bhattacharyya, B. (2015). Time series modeling for trend analysis and forecasting wheat production of India. International Journal of Agriculture, Environment and Biotechnology, 8(2): 303-308.
7. Ray, M., Rai, A., Ramasubramanian, V. and Singh, K.N. (2016). ARIMA-WNN hybrid model for forecasting wheat yield time-series data. *Journalof the Indian Society of Agricultural Statistics*, 70(1):63-70.
8. Polisetty, K. and Paidipati, K.K. (2020). Statistical assessment of trend analysis on production of wheat crop over India. *Sarhad Journal of Agriculture*, 36(1): 178-184.
9. Yonar, A., Yonar, H., Mishra, P., Kumari, B., Abotaleb, M. and Badr, A. (2021). Modeling and forecasting of wheat of South Asian region countries and role in food security.*Advances in Computational Intelligence*, 1: 1-8.
10. Rao, S.G. and Naidu, G.M. (2021). Statistical modeling on area, production and productivity of wheat in India. *International Journal of Agricultural & Statistical Sciences*, 17(2): 539-543.
11. Madhukar, A., Kumar, V. and Dashora, K. (2022). Temperature and precipitation are adversely affecting wheat yield in India. *Journal of Water and Climate Change*, 13(4): 1631-1656.
12. Rajarathinam, A., Parmar, R.S. and Vaishnav, P.R. (2010). Estimating models for area, production and productivity trends of Tobacco (*Nicotiana tabacum*) crop for

Anand region of Gujarat state, India. *Journal of Applied Sciences*, 10(20): 2419-2425.

13. Tripathi, R., Nayak, A.K., Raja, R., Shahid, M., Kumar, A., Mohanty, S., Panda, B.B., Lal, B. and Gautam, P. (2014). Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models. *Advances in Agriculture*, 2014: 1-9.
14. Joshi, P., Gautam, P. and Wagle, P. (2021). Growth and instability analysis of major crops in Nepal.*Journal of Agriculture and Food Research*, 6: 1-6.
15. Kumar, M. and Menon, S.V. (2022). Statistical modeling and trend analysis of jackfruit production in the districts of Kerala in India. *International Journal of Agriculture, Environment and Biotechnology*, 15(03): 745-752.
16. Paudel, A., Basnet, K.B., Paudel, A., Gurung, B. and Poudel, U., (2022). Trend analysis ofarea, production, productivity, and supply of potato in Sindhulidistrict and Nepal: A comparative study. *Malaysian Journal of Sustainable Agriculture*, 6(1): 29-37.
17. Rana, S.K. and Kumar, M. (2022). Growth rate and instability analysis of sugarcane in selected states of India. *International Journal of Agriculture, Environment and Biotechnology*, 15(04): 837-843.