Journal of Agriculture, Biology and Applied Statistics
Vol. 1, No. 1, 2022, pp. 25-37

URL : www.arfjournals.com
https://doi.org/10.46791/jabas.2022.v01i01.02

A Comparative Study between (ARIMA - ETS) Models to Forecast Wheat Production and its Importance's in Nutritional Security

Shweta Shrivastri[1], Khder Mohammed Alakkari[2], Priyanka Lal[3], Aynur Yonar[4*], and Shikha Yadav[5]

[1]*CSIR- National institute of science communication and policy research (NIScPR), New Delhi*
[2]*Department of Statistics and Programming, University of Tishreen, Lattakia, Syria*
[3]*Department of Agricultural Economics & Extension, Lovely Professional University, Punjab, India*
[4]*Selçuk University, Faculty of Science, Deparment of Statistics, Konya, Turkey*
[5]*Research Scholar, Department of Geography, Delhi School of Economics, University of Delhi, India*
[*]*Corresponding author email: aynursahin@selcuk.edu.tr*

To cite this article

Shweta Shrivastri, Khder Mohammed Alakkari, Priyanka Lal, Aynur Yonar & Shikha Yadav. (2022). A Comparative Study between (ARIMA-ETS) Models to Forecast Wheat Production and its Importance's in Nutritional Security. *Journal of Agriculture, Biology and Applied Statistics*. Vol. 1, No. 1, pp. 25-37.

Abstract: Wheat, which has been one of the most important food sources for humans for centuries, has a special place in the economy of India and its states due to its substantiality in food security, trade, and industry. Thus, the forecasting of wheat production is of great importance because it allows them to cope with food problems they may encounter in the future. This study aims to set up models and forecasts for wheat production in India and its five states: Uttar Pradesh (UP) - Punjab (PB) - Madhya Pradesh (MP) - Haryana - Rajasthan (RJ) by using annual time series data from 1956 to 2020. The ARIMA and ETS models were conducted and the best-fitted models to forecast wheat production were selected with respect to performance indicators of Akaike information criterion (AIC) and root mean square error (RMSE) for each state and India. The most appropriate models were determined as ARIMA (1,1,0) for UP, ARIMA (0,1,1) for PB, ETS (M, M, N) for MP, ETS (M, MD, N) for Haryana, ETS (M, N, N) for JR, and ETS (A, A, N) for India for forecasting wheat production from 2021 to 2030.

Keywords: ARIMA; ETS; Time series;Production; Forecasting.

Introduction

Due to the advent of the Green Revolution during the 1960's, India's agriculture has transformed from a food deficit to a food surplus. This led to the alleviation of poverty and hunger in the country (Raeboline *et al.*, 2019). But the recent figures of India's Hunger

Article History Received : 15 January 2022; Revised : 09 February 2022; Accepted : 19 March 2022; Published : 30 June 2022

Index speak of a completely different picture of food grain production in the country. The Global Hunger Index 2021 was an impudent reminder that in the future, the situation of scarcity is next door and many people around us are deprived of adequate food to survive (https://www.downtoearth.org.in/news/food/india-is-reeling-under-hunger-will-governmentintervention-during-covid-19-help--79753). In the past, the government's policy focused on increasing production, which later changed to increasing productivity, and now the approach is towards sustainability. India is undisputedly the second-highest producer of major food grains, viz., rice, wheat, etc. The point of concern here is whether we can sustain the population of the whole country, which is increasing at a growth rate of around 1%, also accounting for the post-harvest losses. Thus, in order to address the demand-supply gap and food security in the country where the production is resource-intensive, cereal-centric, and regionally biased (https://www.fao.org/india/fao-in-india/india-at-a-glance/en/), the trend needs to be studied. Therefore, the present study on forecasting wheat production in India was taken as a modest step to cover the gap.

For forecasting wheat production in the country, the top five wheat-producing states were selected for the study. The analysis was based on the data collected from Uttar Pradesh, Punjab, Madhya Pradesh, Haryana, and Rajasthan during the study period of 1956 to 2020. For the purpose of forecasting, two types of models were used in the study, i.e., ARIMA and the Exponential Smoothening Method (ETS). The most popular method of forecasting is the Box-Jenkins model, which uses a time series autoregressive integrated moving average (ARIMA) model for forecasting (Box and Jenkins, 1976). Previous studies on forecasting of wheat production have depicted an increase in production (Prabakaran *et al.* (2013); Dasyam *et al.* (2015)) growing at an average growth rate of 4% per year (BholaNath *et al.* 2019). But after the jolt received due to the COVID situation, the trend of production forecasts needs to be re-studied.

ARIMA modelling and exponential smoothening methods use the same factors for predicting, but there are some differences as well. The theoretical underpinning of differences shows that non-stationary data is found in the case of ETS models while some ARIMA models are stationary. Research in the past also indicated that if forecasting is the objective, then ETS proved to be a simpler model, as ARIMA models have the important and challenging task of selecting the correct order (Makridakis *et al.* 1982, Fildes *et al.* 1998). In this paper, we have tried to approach the forecasting of wheat production using both methods.This paper is organized into the following heads: Section 2 comprises data and methodology, followed by section 3, which gives the empirical results of the study, followed by section 4, comprising conclusions.

Data and Methology

In this research, we aim to predict the production of wheat in five Indian states (Uttar Pradesh (UP)-Punjab (PB)-Madhya Pradesh (MP)-Haryana-Rajasthan (RJ) and India). The study period runs from 1956 to 2020 on an annual basis. To forecast wheat production up

to the year 2030, we use two types of models (ARIMA and ETS) and compare their results. We will use data spanning from 1956-2015 for estimation and training by models, and data from 2016-2020 to validate the models. Before that, our methodology goes through several stages:

1. DATA Exploration

To visualize the data features (patterns, unusual observations, changes over time). We need to plot the data, and then translate that through descriptive statistics and normal distribution of the data using the following statistic:

$$Jarque - Bera = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \quad (1)$$

Where *n*: number of observations, *S*: Skewness, *K*: kurtosis.

2. DATA Stationary

Time series that have a trend and volatility are not stationary, will affect the value of the time series at different time. Thus, the series cannot predicted in the long run. One way to determine whether a time series is stationary or not is to use a unit root. The time series in augmented Dickey - Fuller test is described by the equation (Dickey and Fuller, 1981):

$$\Delta y_t = c + \alpha.t + \delta y_{t-1} + \beta_{p-1}\Delta y_{t-p+1} + \varepsilon_t \quad (2)$$

Wherec: constant, α: coefficient on a time trend, *p*: lag order of the autoregressive process. The ADF test is carried out under the null hypothesis δ = 0 (not stationary) against the alternative of δ < 0 (stationary). If the null hypothesis is not rejected, we perform the first difference to make the series stationary:

$$y'_t = y_t - y_{t-1} \quad (3)$$

3. Estimation Models

To forecast pulses production up to year 2027, we use three types of models:

3.1. ARIMA Model

ARIMA models are the most widely used statistical models for time series forecasting, this is done by describing the autocorrelation in the data (Box *et al.*, 2015). These models are divided into three parts, according to their nomenclature (AutoRegressive - Integrated - Moving Average) (*p*, *d*, *q*).

Autoregressive (*p*) refers to predicting a variable using a linear set of its preceding values, the model of order p can written as (Mishra *et al.*, 2021a):

$$y_t = c + \beta_P y_{t-p} + \varepsilon_t \quad (4)$$

Where β_P: parameters of model, P: lag order of the autoregressive process, ε_t: error term. Integrated (*d*) refers to the degree of stationary of a variable that is determined using ADF test.

Moving average (*q*): uses past forecast errors in in regression. The equation will be in the form:

$$y_t = c + \varepsilon_t + \beta_q \varepsilon_{t-q} \quad (5)$$

Where β_q: parameters of model, *q*: lag order of the moving average, ε_t: error term.

Whereas (d) is determined by ADF test, (*p*) and (*q*) are determined by the autocorrelation function *R*(*p*) and the partial autocorrelation function *R*(*p*), which are given according to the following ,Mishra *et al.*, 2021b):

$$\boldsymbol{\rho}(p) = \frac{Cov(y_t, \boldsymbol{y}_{t+p})}{\sigma^2} \quad (6)$$

$$(\boldsymbol{\rho}(p-1) \quad \boldsymbol{\rho}(p-2) \ldots . \quad \boldsymbol{\rho}(0)) \ (\beta_q) = R(p) \quad (7)$$

3.2. ETS Model

Whereas the ARIMA model describes autocorrelation in the data, the exponential smoothing model (ETS) is based on describing the trend in the data, which was suggested by Holt (1957) and Winters (1960).ETS models are a systematic development in which exponential smoothing models (ETS) are combined into a nonlinear dynamic model. Analysis of these models using state-space based likelihood calculations provides support for model selection and calculation of forecast standard errors (Hyndman *et al.*, 2002).

Interested in the model in three main component of time series: trend (T), seasonal (S), error (E). Reflects the trend term of the long-term movement of time series, and the error term is the unpredictable component of the time series. In our case, do not care about the seasonal term because the data annual. The components we need are combined in our model, in various additive and multiplicative combinations to produce y_t. We have additive model y_t = T+E or multiplicative model like y_t = T. E. where the individual components of the model are given as follows:

E [A,M]

T [N,A,M,AD,MD]

S [N,A,M]

Where N = none, A = additive, M = multiplicative, AD = additive dampened, and MD = multiplicative dampened (damping uses an additional parameter to reduce the impacts of the trend over time).the models that we are interested in estimating can be written (after selecting S [N]) in the following table:

Table 1: State Space Equations for each of the Models in the Holt's Nonlinear

Trend	Additive Error Models	Trend	Multiplicative Error Models
N	$y_t = l_{t-1} + \varepsilon_t$ $l_t = l_{t-1} + \alpha\varepsilon_t$	**N**	$y_t = l_{t-1}(1 + \varepsilon_t)$ $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$
A	$y_t = l_{t-1} + b_{t-1} + \varepsilon_t$ $l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	**M**	$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t$
AD	$y_t = l_{t-1} + \phi b_{t-q} + \beta\varepsilon_t$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	**MD**	$y_t = (l_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $l_t = (l_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\varepsilon_t$

Where parameters: α: smoothing factor for the level, β: smoothing factor for the trend, ϕ: damping coefficient. And initial states: *l*: initial level components, *b*: initial growth components, which is estimated as part of the optimization problem.

4. Performance Indicators

In order to choose the best prediction model of the same type, we using akaike information criterion (AIC), which is given as follow:

$$-2log\ L(\hat{\theta}) + 2k \tag{8}$$

Where$\hat{\theta}$ maximum value of the likelihood function. Whereas the lowest value for it gives us the best model.

To compare the prediction performance of the two models used, we first test the validity of the model by calculating root mean square error (RMSE) between the estimated data and actual data (Out and In of sample), which indicates the standard deviation of errors as its value depends on the values of the variables:

$$\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{9}$$

Where $\hat{y}_t$: the forecast value, y_t: the actual value, *n*: number of fitted observed. The last stage is to predict the wheat production for the states of the study sample until 2030, the model that hasthe least values of (RMSE) is the best.

Empiricial Study

To know the development and trends in wheat production for (Uttar Pradesh (UP)-Punjab (PB)-Madhya Pradesh (MP)-Haryana-Rajasthan (RJ) and India), we present the following figure:

Clearly, the visual shows us that wheat production has increased stationary during the studied period in all states and consequently in India. But it is also noticeable that there

was a significant decline in MP wheat production between 2000 and 2010. Descriptive statistics show the most important values of changes in data through the following table:

The table 2 shows us that the probability distribution for all is close to normal (except for MP). As the probability (Jarque-Bera) is larger than the level of significance (5%), The mean wheat production of India was 51355.52 thousand tons during the study period, and the largest production of Haryana was a mean of 5997.529 thousand tons. The probability distribution of variables indicates a kurtosis of type platykurticand no skewness. This indicates a stable development in production values in all states and India, as seen in figure 1. As we notice in the figure 1, the MP variable's values are not normal distribution, as we notice that kurtosis is of type mesokurtic, which indicates unstable volatility in the values of the MP. In order to find out the effect of these volatility and trends on the stationary variables, we use the ADF test, and we get the following results:

The table shows us that (Haryana - India) is stationary in level with linear trend as shown in figure 1. For the rest of the states, the high volatility during the studied period and power of trend made it a stationary series at the first difference. With the aim of forecasting the wheat production for all states and India up to 2030, we use (ARIMA- ETS) models; we use the data during the period (1956-2015) to estimate using models (training) and (2016 - 2020) to verify the validity of the model (testing). The following figure (2) shows auto and partial correlation function of the variables according to their degree of stationary, where the figure (2) indicates that auto and partial correlation function is absent after the first lag for (d(UP), d(PB), d(MP), d(RJ)), and gradually decreases for each (Haryana, India). The following table shows the results of estimating the ARIMA model for the all states and India. The best model for each variable is chosen based on auto correlation function (figure 2), stationary (table2), and akaike information criterion, as we get the following results:

The table shows us that (UP-PB-MP) have better out-of-sample prediction results as the value of (RMSE-Testing) is less than the value of (RMSE-Training) for these models, and the models of (Haryana-RJ-India) failed to predict lower values after 2015. The best model is ARIMA (0,1,1) for PB, which achieved the lowest values of AIC and RMSE among the selected models. We estimate the ETS model for all variables and get the following results:

The table shows that the best ETS model among the estimated models is for Haryana (M, MD, and N), which has the lowest values of AIC-RMSE in and out of the sample. The next step is to choose the best model for forecasting wheat production up to 2030. We rely on visualization to choose the best model for each variable that shows us the actual data (blue line) and the expected data using the ARIMA model (red line) and the ETS model (green line), in-sample (1956-2015) and out-of-sample (2016-2020). The best model is the one that makes predictions that are close to actual data:Figure 3 shows that the ARIMA model achieved a better fit than the ETS model for (UP-PB0), and the ETS model achieved a better fit for (MP-Haryana-RJ-India). This is in line with the results of the RMSE indicator

shown in table (4-5). Accordingly, we choose the appropriate model for each variable according to the following table:The next step is to compute wheat production forecast values for each variable to year 2030:The figure shows us that all states and India are expected to achieve an increase in wheat production, but at varying rates. The table shows us that (MP) is expected to achieve the largest growth in wheat production. The lowest expected growth rate is (PB), which gives rise to greater interest in increasing production.

Table 2: Normal distribution and descriptive statistics for wheat production in fivestatus and India during the period 1956-2020

Status	*Normality J-B (Prob)*	*Mean*	*Standard Deviation*	*Maximum*	*Minimum*	*Skewness*	*Kurtosis*
UP	0.078200	17032.70	9827.334	33815.50	2715.000	-0.035598	1.630002
PB	0.061214	10192.50	5513.716	18261.00	1302.000	-0.269109	1.668411
MP	0.000002	6092.628	4914.505	19607.14	1031.000	1.437727	4.148909
Haryana	0.060467	5997.529	4050.843	12780.00	611.0000	0.155150	1.594443
RJ	0.070145	4508.769	3021.169	10916.12	785.0000	0.534274	2.094218
India	0.133253	51355.52	30332.79	107860.5	7998.000	0.159797	1.822616

Table 3: ADF test result

Countries	*ADF (prob)*		*Integrated degree*
	Level I(0)	*First difference I(1)*	
UP	0.2806	0.0000	I(1) with constant
PB	0.2463	000	I(1)with constant
MP	0.9997	000	I(1) with trend
Haryana	0.0330	——	I(0) with trend
RJ	0.5189	000	I(1)with constant
India	0.0048	——	I(0) with trend

Note: UP- Uttar Pradesh, PB – Punjab, MP – Madhya Pradesh, HR- Haryana, RJ – Rajasthan,

Table 4: The results of estimating ARIMA models

Countries	*Model*	*Parameters ARIMA*			*AIC*	*Training RMSE*	*Testing RMSE*
		drift	*AR*	*MA*			
UP	(1,1,0)	381.1	-0.29	-	17.33	3570.44	1925
PB	(0.1.1)	253.3	-	-0.36	14.08	1492.1	745.5
MP	(0,1,4)	245.8	-	-0.28	17.01	4515.8	1631
Haryana	(1,0,8)	-	0.99	0.15	15.61	6204.3	11615
RJ	(1,1,0)	-	-0.40	-	15.84	3991.7	8956
India	(1,0,2)	-	0.99	0.29	19.48	4768	7886

Table 5: The results of estimating ETS models

Countries	*Model*	*Parameters ETS*			*Initial stats*		*AIC*	*Training*	*Testing*
		α	β	ϕ	*l*	*b*		*RMSE*	*RMSE*
UP	(A,A,N)	0.756	-	-	2577.8	-386.7	1117.4	1335.8	5329.3
PB	(M,A,N)	0.927	-	-	1069.3	218.4	1038.1	0.1921	1000.1
MP	(M,M,N)	0.489	-	-	1602	1.044	1058.1	0.1942	1682.8
Haryana	(M,MD,N)	0.879	-	0.976	593.87	1.109	971.9	0.1005	483.89
RJ	(M,N,N)	0.463	-	-	1101.1	1.041	1002.1	0.1601	1176.3
India	(A,A,N)	0.615	-	-	6867.47	1430.9	1227.7	3343.4	6752.3

Table 6: The best model for forecasting wheat production to 2030 for each status and India

Status	*UP*	*PB*	*MP*	*Haryana*	*RJ*	*INDIA*
Model	ARIMA(1.1.0)	ARIMA(0.1.1)	ETS(M,M,N)	ETS(M.MD.N)	ETS(M,N,N)	ETS(A,A,N)

Table 7: Forecasting points in wheat production (2021-2030) for all status and India.

	UP	*PB*	*MP*	*Haryana*	*RJ*	*INDIA*
2021	33586.94089	17778.59038	20465.20207	12261.15631	11442.2454	109414.7833
2022	34066.2593	18031.96869	21359.59001	12507.50107	11885.02736	110971.0988
2023	34545.5777	18285.347	22293.06527	12752.42778	12344.94372	112527.4143
2024	35024.8961	18538.72532	23267.33606	12995.82459	12822.65751	114083.7298
2025	35504.21451	18792.10363	24284.18528	13237.58488	13318.85746	115640.0452
2026	35983.53291	19045.48194	25345.47373	13477.60716	13834.25891	117196.3607
2027	36462.85132	19298.86026	26453.14352	13715.79514	14369.60492	118752.6762
2028	36942.16972	19552.23857	27609.22165	13952.05761	14925.66728	120308.9917
2029	37421.48812	19805.61688	28815.8237	14186.30843	15503.24765	121865.3071
2030	37900.80653	20058.9952	30075.15773	14418.46646	16103.17871	123421.6226
RATE %	12.84	12.82	46.9	17.59	40.73	12.8

Conclusion

In this study, the ARIMA and ETS models were employed to model and forecast the wheat production in India and its five states. The wheat production data during the period 1956–2015 was used for building models, *i.e.*, training, and the data during the period 2016–2020 was utilized to verify the validity of the models, i.e., testing. The best models were determined as ARIMA (1,1,0) for UP, ARIMA (0,1,1) for PB, ETS (M, M, N) for MP, ETS (M, MD, N) for Haryana, ETS (M, N, N) for JR, and ETS (A, A, N) for India. Using these models, wheat production forecasts from 2021 to 2030 were calculated for each state and India. The forecasts demonstrate that the biggest increase in wheat production is expected to be in MP, with 46.9%. It is followed by RJ with 40.73%, UP with 12.84%, PB with 12.82%, India with 12.8%, and Haryana with 17.59%, respectively. By evaluating the estimated wheat production rates obtained in this study and population growth together, economists can make predictions about whether these wheat production amounts will be sufficient for the future. Thus, if a

problem such as insufficient or excessive wheat production is detected, early measures can be taken.For future work, other time series models or machine learning techniques can be employed for modeling and forecasting the wheat production of the states and India.

Competing interests: The authors declare that they have no conflict of interest.

Funding: No funding received for this research.

References

Box, G. E. P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed). Hoboken, New Jersey: John Wiley & Sons.

Bhola N, Dhakre DS, Bhattacharya D (2019). Forecasting wheat production in India: an ARIMA modelling approach. *J Pharmacogn Phytochem* 8(1):2158–2165

Dasyam R, Pal S, Rao VS, Bhattacharyya B (2015). Time series modelling for trend analysis and forecasting wheat production in India. *Int J Agric Environ Biotechnol* 8(2): 303–308.

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

Fildes, Robert, Michèle Hibon, Spyros Makridakis, and Nigel Meade. (1998). "Generalising about univariate forecasting methods: further empirical evidence." *International Journal of Forecasting* 14 (3): 339–58. doi:10.1016/S0169-2070(98)00009-0.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.

Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.

Prabakaran K, Sivapragasam C, Jeevapriya C, Narmatha A (2013) Forecasting cultivated areas and production of wheat in India usingARIMA model. *Golden Res Thoughts* 3(3):23–32

Makridakis, Spyros, A P Andersen, R Carbone, Robert Fildes, Michèle Hibon, R Lewandowski, J Newton, Emanuel Parzen, and Robert L Winkler. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting* 1 (2): 111–53. doi:10.1002/for.3980010202.

Mishra, P., Al Khatib, .M.G., Sardar, I. *et al.*,(2021a), Modeling and Forecasting of Sugarcane Production in India", *Sugar Tech*, **23**, 1317–1324. doi: 10.1007/s12355-021-01004-3.

Mishra P., Yonar. A., Yonar. H., Kumari. B., Abotaleb, M., Das, S.S., Patil, S.G.,(2021b).State of the art in total pulse production in major states of India using ARIMA techniques. Current Research and Food Science, **4**:800–806. doi. :10.1016/j.crfs.2021.10.009.

Raeboline, A., Eliazer, L., Ravichandran, K., & Antony, U. (2019). *The impact of the Green Revolution on indigenous crops of India*. *9*, 1–10.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342.

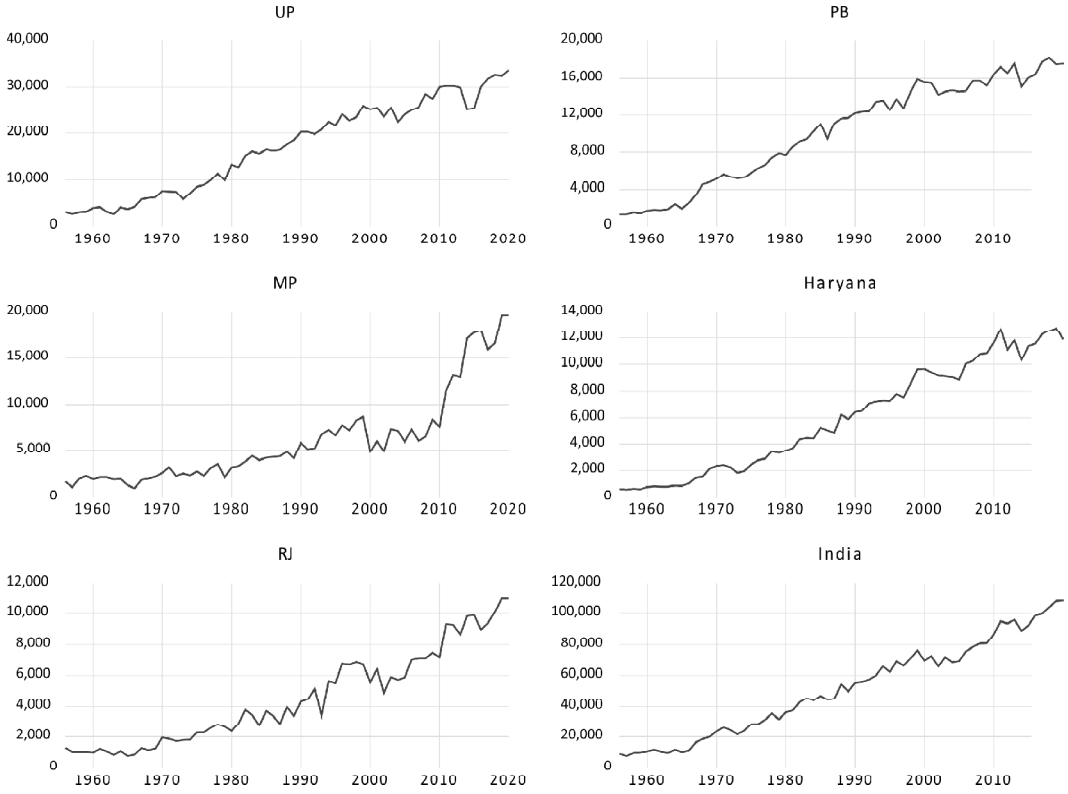
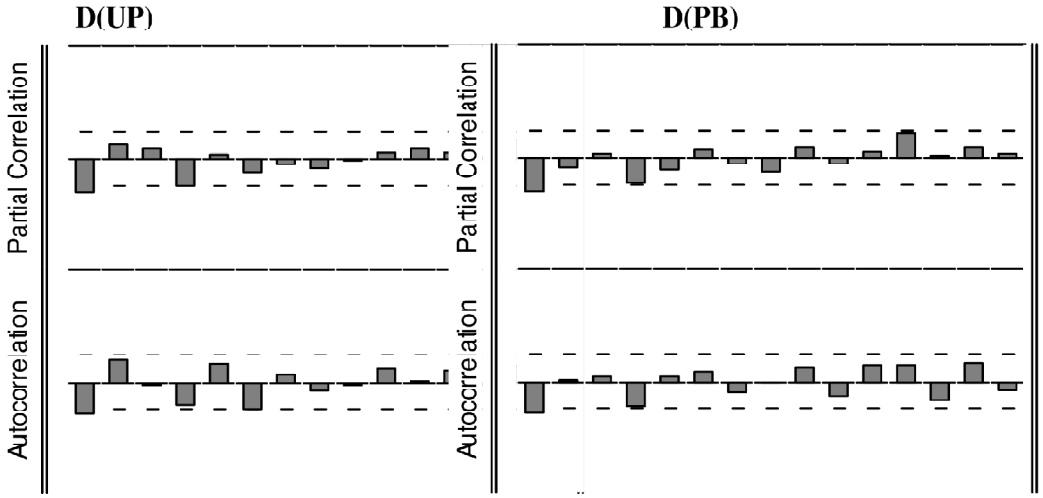


Figure 1: The development of wheat production in (Uttar Pradesh (UP) –Punjab (PB) –Madhya Pradesh (MP) –Haryana–Rajasthan (RJ) and India) during the period 1956-2020



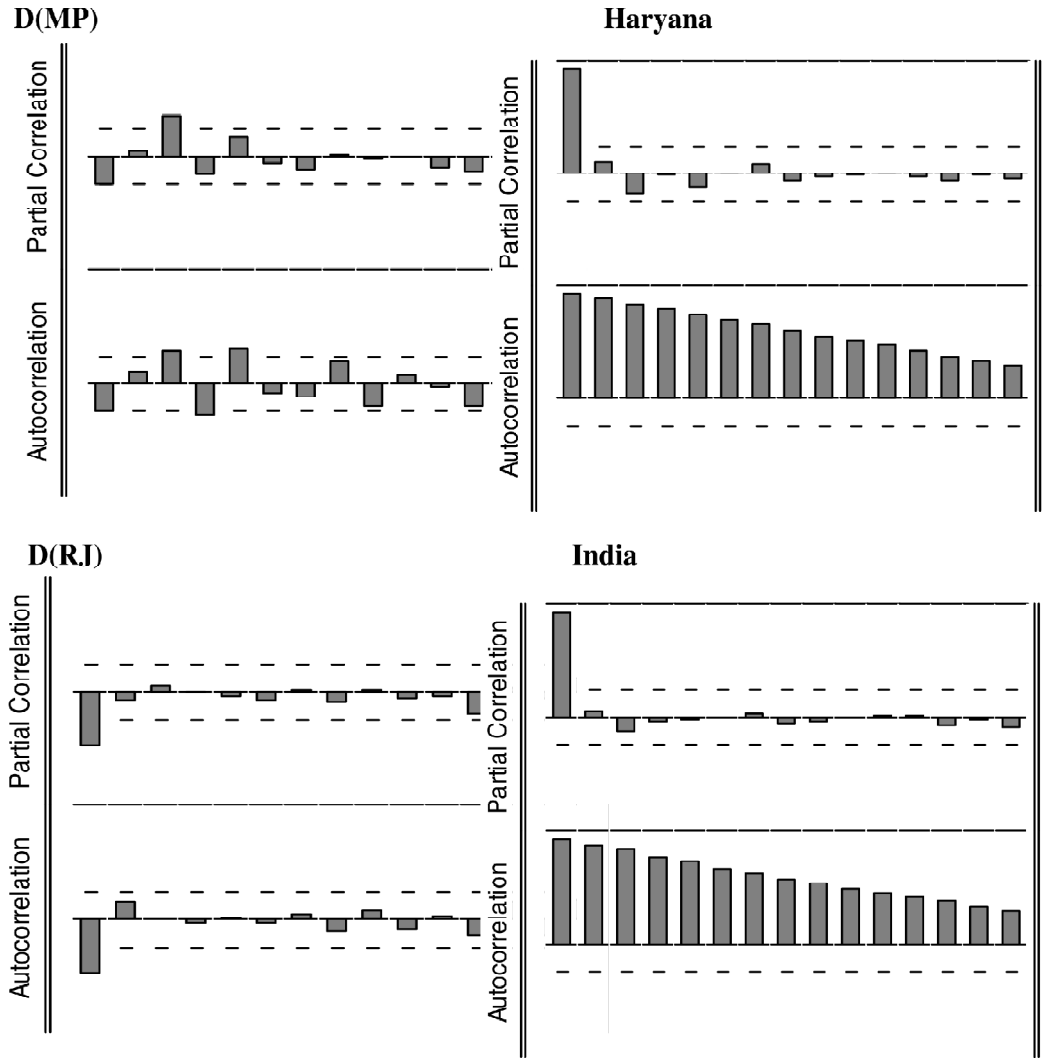


Figure 2: Auto and Partial Correlation function for all variables

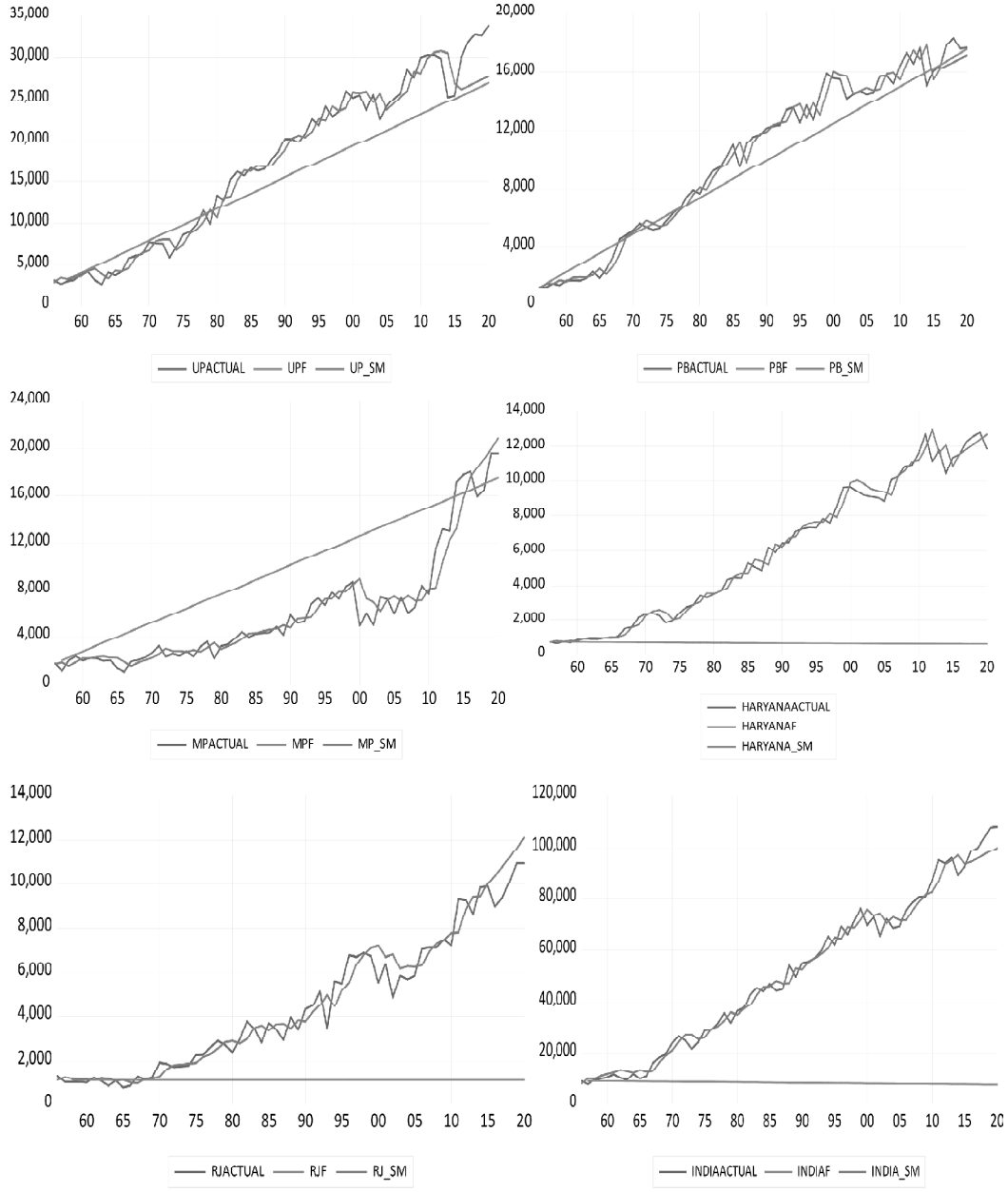


Figure 3: Actual and Forecast Values Using (ARIMA – ETS) Models for all variables Out and IN samples

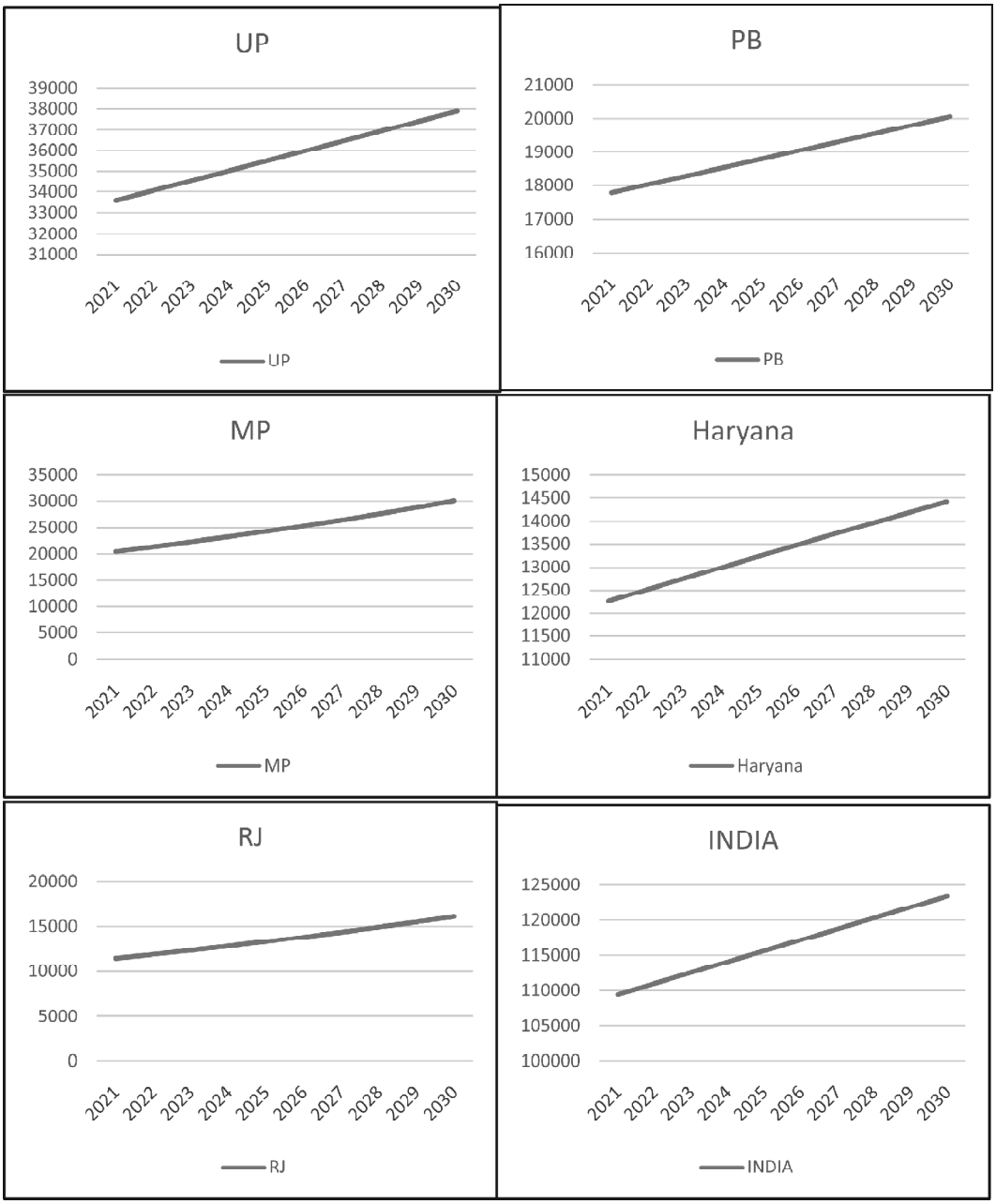


Figure 4: Forecasting in wheat production to 2030 for all states and India