International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2017; 2(5): 14-21

www.mathsjournal.com
Received: 04-07-2017
Accepted: 05-08-2017

M Saranyadevi
Assistant Professor (GUEST), Department of Statistics, Government Arts College (Autonomous), Kumbakonam, Tamilnadu, India

Dr. A Kachi Mohideen
Assistant Professor, Department of Statistics, Periyar EVR College Autonomous, Trichy, Tamilnadu, India

Correspondence
M Saranyadevi
Assistant Professor (GUEST), Department of Statistics, Government Arts College (Autonomous), Kumbakonam, Tamilnadu, India

A stochastic modeling for paddy production in Tamilnadu

M Saranyadevi and Dr. A Kachi Mohideen

Abstract
This paper investigates predictive performance of time series analysis method for paddy production trends in Tamilnadu. In the present study had been carried out on the basis of time-series production data of paddy crop pertaining to the period 1960 to 2015 had been collected and forecasting the production for 2016 to 2020 using ARIMA (Autor Regressive Integrated Moving Average), Simple exponential smoothing, Brown exponential smoothing and Damped exponential smoothing models.

Keywords: Paddy production, ARIMA, simple exponential smoothing, brown exponential smoothing, damped exponential smoothing models

Introduction
Agriculture plays a vital role in India's economy. 54.6% of the population is engaged in agriculture and allied activities (census 2011).Tamil Nadu has historically been an agricultural state, while its advances in other fields launched the state into competition with other areas. Agriculture is heavily dependent on the river water and monsoon rains. Tamil Nadu is also the leading producer of kambu, corn, rye, groundnut, oil seeds and sugarcane in India. At present, Tamil Nadu is India's second biggest producer of rice.

Agriculture continues to be the most predominant sector of the State economy, as 70 percent of the population is engaged in Agriculture and allied activities for their livelihood. Tamil Nadu has all along been one of the states with a creditable performance in agricultural production with the farmers relatively more responsive and receptive to changing technologies and market forces. Heavy consumption of this food crop has become an important issue the yield of rice of our

Country. Considering these facts in mind it is very important to have timely forecast of this crop in Order to plan the strategies to increase the yield.

To make the best forecast of paddy production in India, appropriate time series forecasting Methods are discussed by Makridakis and Hibbon (1979), Pankratz (1983), Granger and Newbold (1986), Makridakis *et al*. (1998).One of the most successful time series forecasting techniques is Box-Jenkin type ARIMA model is given by Box and Jenkins (1970) which is the most commonly used linear statistical model. Accordingly the present study was aims to find appropriate model for forecasting paddy production in tamilnadu based on different smoothing techniques and ARIMA time series modelling and to select more appropriate forecasting model based on model selection criteria.

Materials
In the present study had been carried out on the basis of time-series production data of paddy Crop pertaining to the period 1960 to 2015 had been collected in Tamil Nadu from the official Website (http://www.indianmundi.com) and forecasting the production for 2016 to 2020 using ARIMA (Auto Regressive Integrated Moving Average), Brown exponential smoothing and Damped exponential smoothing models. SPSS package was used to estimate the model Parameters and all other related statistics.

Methods

After ensuring the presence of trend in the data, smoothing of the data is the next requirement for time series analysis. For smoothing the common techniques discussed by Gardner (1985) are simple Exponential smoothing (SES), Holt's linear exponential smoothing (LES), Additive Holt's model with a damping factor, brown's exponential smoothing. All these smoothing techniques have been employed to paddy production data set. Among these smoothing techniques, best one has been identified based on model selection criteria. ARIMA model also was employed and compared with the smoothing techniques. The computational procedures are discussed in detail in sequence below.

Simple exponential smoothing.

The simplest of the exponentially smoothing methods is naturally called "simple exponential smoothing" (SES). (In some books, it is called "single exponential smoothing".) This method is suitable for forecasting data with no trend or seasonal pattern. Using the naïve method, all forecasts for the future are equal to the last observed value of the series,

$$\hat{Y}_{T+h/T} = Y_T$$

for h=1,2,…. Hence, the naïve method assumes that the most current observation is the only important one and all previous observations provide no information for the future. This can be thought of as a weighted average where all the weight is given to the last observation.

Using the average method, all future forecasts are equal to a simple average of the observed data,

$$\hat{Y}_{T+h/T} = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

for h=1,2,….

Hence, the average method assumes that all observations are of equal importance and they are given equal weight when generating forecasts.

We often want something between these two extremes. For example it may be sensible to attach larger weights to more recent observations than to observations from the distant past. This is exactly the concept behind simple exponential smoothing. Forecasts are calculated using weighted averages where the weights decrease exponentially as observations come from further in the past --- the smallest weights are associated with the oldest observations:

$$\hat{Y}_{T+h/T} = \alpha Y_T + \alpha(1-\alpha)Y_{T-1} + \alpha(1-\alpha)^2 Y_{T-2} + \cdots$$

Where $0 \leq \alpha \leq 1$ is the smoothing parameter. The one-step-ahead forecast for time T+1 is a weighted average of all the observations in the series $y_1, \ldots, Yt$. The rate at which the weights decrease is controlled by the parameter α.

Let an observed time series be $y_1, y_2, \ldots y_n$. Formally, the simple exponential smoothing equation takes the form of

$$\hat{Y}_{T+1} = \alpha Y_T + \alpha(1-\alpha)\hat{Y}_T$$

Where Y_T is the actual, known series value for time period T,$\hat{Y}_T$ is the forecast value of the variable Y for time period T,Y_T is the forecast value for time period T +1 and α is thesmoothing constant is given by Brown *et al*.(1961). The forecast $\hat{Y}_{T+1}$is based on weighting the most recent observation$Y_T$with a weight α and weighting the most recent forecast $Y_T$with a weight of 1−α.

Holt's exponential smoothing

Holt (2004) extended simple exponential smoothing to allow forecasting of data with a trend. This method involves a forecast equation and two smoothing equation

$$\hat{Y}_{T+h/T} = l_T + hb_T$$
$$l_T = \alpha Y_T + \alpha(1-\alpha)(l_{T-1} + b_{T-1})$$
$$b_T = \beta^*((l_T - l_{T-1}) + (1-\beta^*)b_{T-1}$$

where t,bt denotes an estimate of the level of the series at time t, bt denotes an estimate of the trend (slope) of the series at time tt, αα is the smoothing parameter for the level, $0 \leq \alpha \leq 1$ and β^* is the smoothing parameter for the trend, $0 \leq \beta^* \leq 1$.

Brown's exponential smoothing

Brown's exponential smoothing is given by Brown (1961). The smoothing equations are

$$l_T = \alpha Y_T + (1-\alpha)l_{T-1}$$
$$T_t = \alpha(l_t - l_{t-1}) + (1-\alpha)T_{t-1}$$

The h step ahead prediction equation is

$$\hat{Y}_{T+h} = l_T + ((h-1) + 1/\alpha)T_t.$$

Damped exponential smoothing

The smoothing equation is

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + \varphi T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)\varphi T_{t-1}$$

The h-step-ahead prediction equation is

$$\hat{Y}_{t+h} = L_t + \sum_{i=1}^{k} \varphi^i T_t$$

This is, you forecast y h-steps ahead by taking the last available estimated level state and multiplying the last available trend (slope), T_t, with ϕ^i= dampening factor.

ARIMA Model

A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally spaced intervals with a consistency in the activity and the method of measurement.

Moving Average Process

Moving average models were first considered by Slutsky (1927) and Wold (1938). The Moving Average Series can be written as

$$Y_t = e_t - \theta_1 e_{t-1}\, e_t - \theta_2 e_{t-2}\, e_t - \theta_3 e_{t-3} - \cdots - e_t - \theta_q e_{t-q} \quad (1)$$

We call such a series a moving average of order q and abbreviate the name to MA (q). Where, Y_t is the original series and e_t is the series of errors

Auto-Regressive Process

Yule (1926) carried out the original work on autoregressive processes. Autoregressive processes are as their name suggests regressions on themselves. Specifically, a p^{th}- order autoregressive process $\{Yt\}$ satisfies the equation

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \cdots + \Phi_p Y_{t-p} + \quad (2)$$

The current value of the series Y_t is a linear combination of the p most recent past values of itself plus an "innovation" term e_i that incorporates everything new in the series at time t that is not explained by the past values. Thus, for every$t$, we assume thate_i is independent of $Y_{t-1}, Y_{t-2}, Y_{t-3} \cdots Y_{t-q}$.

Autoregressive Integrated Moving Average (ARMIA) model

The Box and Jenkins (1970) procedure is the milestone of the modern approach to time series analysis. Given an observed time series, the aim of the Box and Jenkins procedure is to build an ARIMA model. In particular, passing by opportune preliminary transformations of the data, the procedure focuses on Stationary processes. In this study, it is tried to fit the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model. This model is the generalized model of the non-stationary ARMA model denoted by ARMA (p, q) can be written as

$$\Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \cdots \cdots - \theta_q e_{t-q} \quad (3)$$

Where, $Y_t$is the original series, for every$t$, we assume that e_t is independent of $Y_{t-1} + Y_{t-2} + Y_{t-3} + \cdots + Y_{t-p}$.A time series $\{Y_t\}$ is said to follow an integrated autoregressive moving average (ARIMA) model if the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. If $\{W_t\}$ follows an ARMA (p, q) model, we say that $\{Y_t\}$is an ARIMA (p, p, q) process. Fortunately, for practical purposes, we can usually take d = 1 or at most 2.

Consider then an ARIMA (p, 1, q) process. With $W_t = Y_t - Y_{t-1}$ we have

$$W_t = \Phi_1 W_{t-1} + \Phi_2 W_{t-2} + \cdots + \Phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \cdots \cdots \theta_q e_{t-q} \quad (4)$$

Box and Jenkins procedures

1. **Preliminary analysis:** create conditions such that the data at hand can be considered as the realization of a stationary stochastic process.
2. **Identification:** specify the orders p, d, q of the ARIMA model so that it is clear the number of parameters to estimate. Recognizing the behavior of empirical autocorrelation functions plays an extremely important role.
3. **Estimate:** efficient, consistent, sufficient estimate of the parameters of the ARIMA model (maximum likelihood estimator).
4. **Diagnostics:** check if the model is a good one using tests on the parameters and residuals of the model. Note that also when the model is rejected, still this is a very useful step to obtain information to improve the model.
5. **Usage of the model:** if the model passes the diagnostics step, then it can be used to interpret a phenomenon, forecast.

Jarque-Bera Test

We can check the normality assumption using Jarque-Bera (1978) test, which is a goodness of fit measure of departure from normality, based on the sample kurtosis (k) and Skewness(s). The test statistics Jarque-Bera (JB) is defined as

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right) \sim \chi^2_{(2)}$$

Where n is the number of observations and k is the number of estimated parameters. The statistic JB has an asymptotic chi-square distribution with 2 degrees of freedom, and can be used to test the hypothesis of Skewness being zero and excess kurtosis being zero, since sample from a normal distribution have expected Skewness of Zero and expected excess kurtosis of zero.

Ljung-Box test

Ljung-Box Test can be used to check autocorrelation among the residuals. If a model fit well, the residuals should not be correlated and the correlation should be small. In this case the null hypothesis is

$H0: \rho_1(e) = \rho_2(e) = \cdots \ldots = \rho_k(e) = 0$ is tested with the Box-Ljung statistic $Q* = N(N+1) \sum_{i=1}^{k} (N-k) \rho_k^2(e)$

Where, N is the no of observation used to estimate the model. This statistic Q* approximately follows the chi-square distribution with $(k-q)$ df, where q is the no of parameter should be estimated in the model. If Q* is large (Significantly large from zero), it is said that the residuals autocorrelation are as a set are significantly different from zero and random shocks of estimated model are probably auto-correlated. So one should then consider reformulating the model.

Model Performance Measures

The following measures have been used to study the different model performance.

Root Mean Square Error (RMSE) = $\left[\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}\right]^{\frac{1}{2}}$

Mean Absolute Error (MAE) =$\frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_i|}{n}$

Mean Absolute Percentage Error (MAPE) =$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100$

Where Y_i is the original paddy production in different years and $\hat{Y}_i$is the forecasted paddy production in the corresponding years and n is the number of years used as forecasting period.

NBIC(p,q)=In v*$(p,q) + (p+q)\left[\frac{In(n)}{n}\right]$;Where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

Results and discussion

Exponential Smoothing Linear model: Time plot (Fig. 1) of milled rice production data revealed that there is increased trend in the data period 1960-2015.

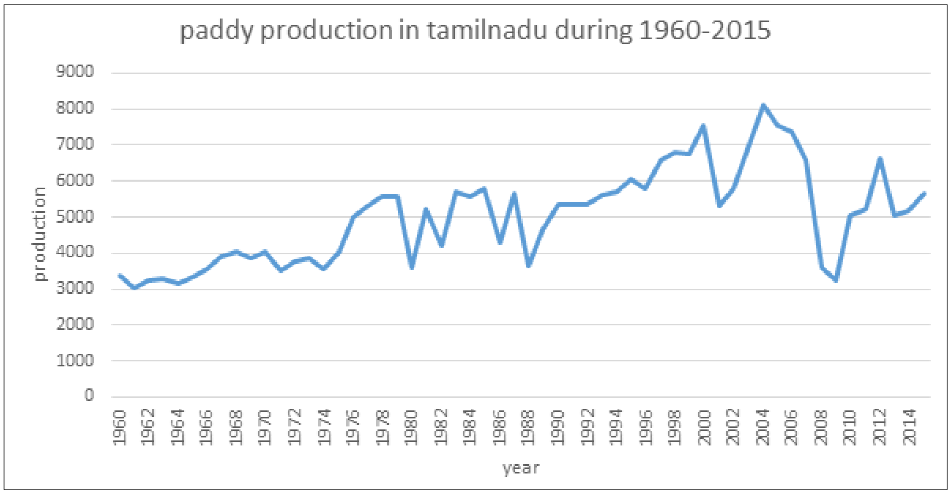

Fig 1: Time plot of paddy production in tamilnadu.

For smoothing of the data, Brown's exponential smoothing technique was found to be most appropriate. Various combinations of Level and Trend based on range between 0.1 to 0.9 with increments of 0.1 were tried and mean absolute percentage error (7.787) was least for α= 0.214 (Fig. 2 and Table 4).

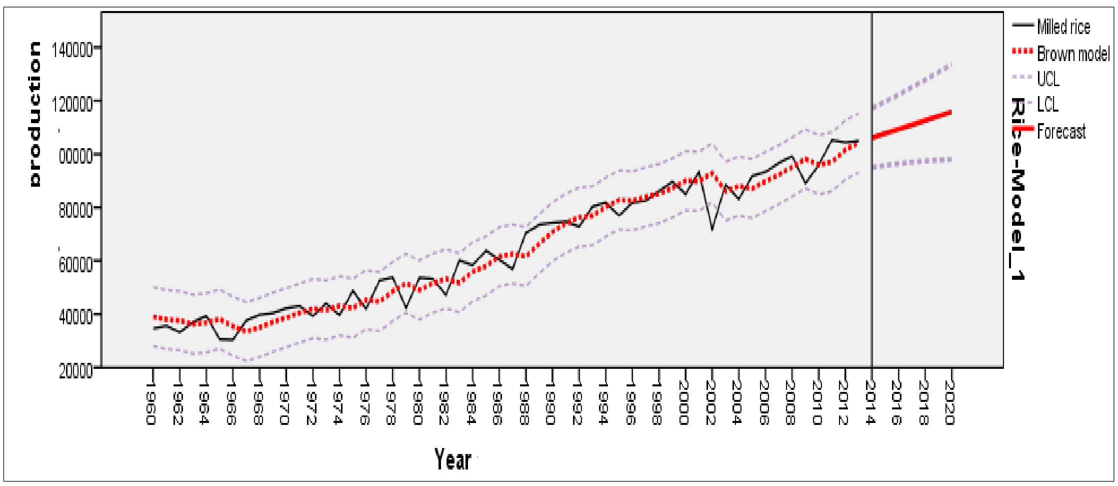

Fig 2: Forecasting of Brown smoothing Exponential

Damped exponential smoothing technique was found to be most appropriate. Various combinations of Level and Trend based on range between 0.1 to 0.9 with increments of 0.1 were tried and mean absolute percentage error(6.896) was least for α= 0.214,γ=0.00006688 and φ = 0.899 is shown in figure.3 and table 4.

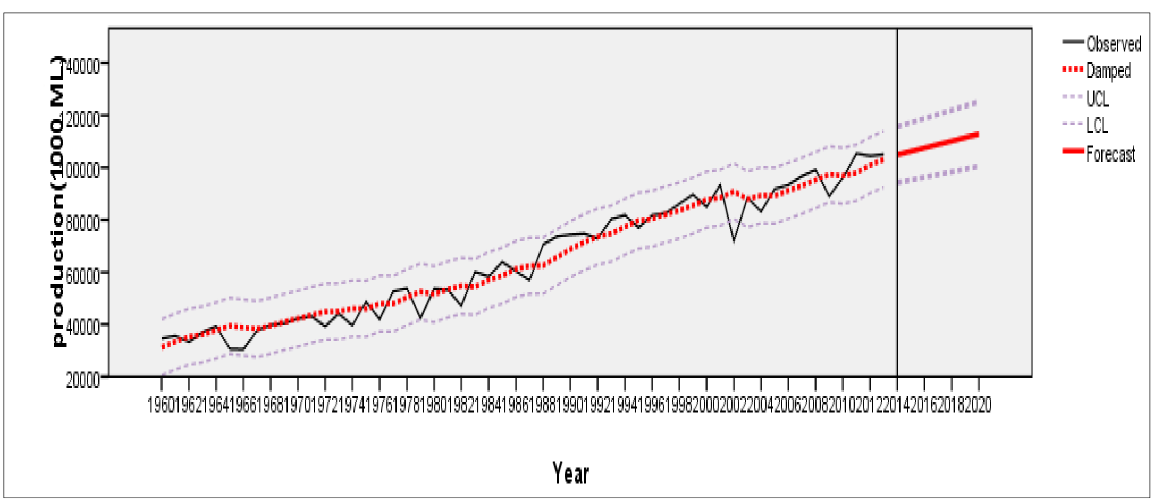

Fig 3: forecasting damped exponential smoothing

ARIMA Model identification

The milled rice production is stationary check of the series revealed that it was non-stationary. Merely by using the first differencing technique, it was made stationary and thus the values of d was 1. The graphs of the sample ACF and PACF were plotted in figure -4 and table -1.

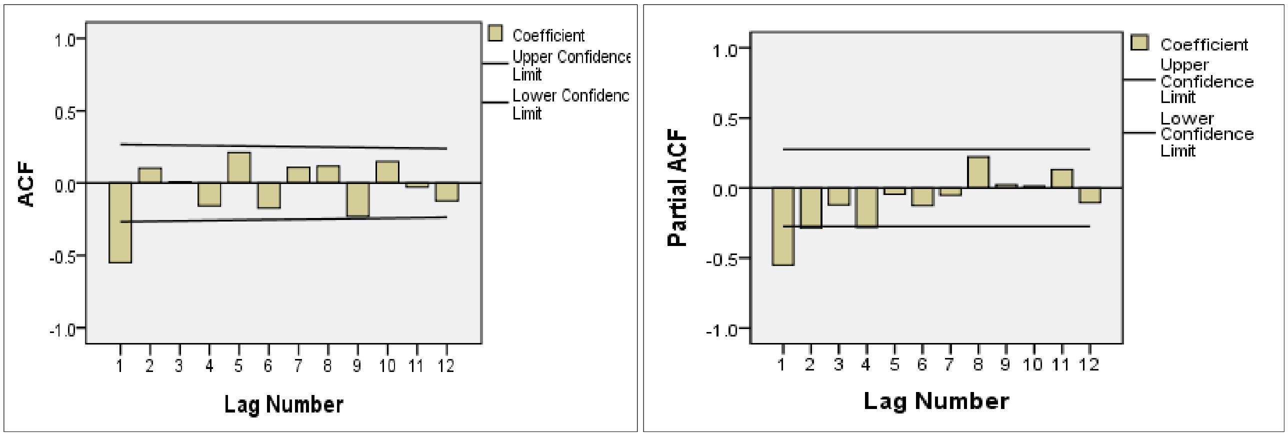

Fig 4: ACF and PACF of differenced data

Table 1: (ACF and PACF for paddy production)

Lag	Auto correlation		Box –Ljung statistics			Partial auto correlation	
	Value	df	Sig	Value	df	Value	Df
1	-0.548	.134	16.914	1	000	-0.545	.137
2	0.101	.132	17.511	2	000	-0.284	.137
3	0.006	.131	17.514	3	001	-0.130	.137
4	-0.154	.130	18.958	4	001	-0.284	.137
5	0.212	.128	21.680	5	001	-0.047	.137
6	-0.174	.127	23.560	6	001	-0.126	.137
7	0.108	.126	24.274	7	001	-0.054	.137
8	0.116	.124	25.176	8	001	0.223	.137
9	-0.231	.123	28.714	9	001	0.026	.137
10	0.146	.121	30.201	10	001	0.015	.137
11	-0.027	.120	30.202	11	001	0.131	.137
12	-0.123	.119	30.328	12	002	-0.105	.137

Tentatively selected twenty ARIMA models at different values of p,d and q were estimated. Comparison among family of different parametric combination of ARIMA (p,d,q) was done according to the minimum values of Normalize BIC,RMSE,MAPE,MAE and maximum value of R^2which are given in table 2.

Table 2

Model ARIMA (p,d,q)	Values of model selection criteria				
	R^2	RMSE	MAPE	MAE	NBIC
ARIMA(0,1,0)	0.92	6832.441	8.514	5018.619	17.734
ARIMA(0,1,2)	0.95	5308.000	6.418	3600.844	17.345
ARIMA(0,1,3)	0.94	5365.265	6.809	3912.826	17.358
ARIMA(1,1,0)	0.94	5418.628	6.647	3914.068	17.258
ARIMA(1,1,1)	0.93	5377.368	6.258	4230.258	17.366
ARIMA(1,1,2)	0.94	5364.873	6.325	3978.258	17.258
ARIMA(1,1,3)	0.94	5389.235	6.258	3974.258	17.369
ARIMA(2,1,0)	0.93	5471.365	6.214	4358.369	17.369
ARIMA(2,1,1)	0.93	5354.214	7.258	4258.369	17.258
ARIMA(2,1,2)	0.93	5347.214	6.247	4563.698	17.582
ARIMA(2,1,3)	0.93	5389.256	6.987	4583.258	17.014
ARIMA(3,1,0)	0.94	5321.214	6.369	3984.258	17.025
ARIMA(3,1,1)	0.93	5421.256	6.258	3958.214	17.025
ARIMA(3,1,2)	0.92	5496.214	6.147	3987.214	17.025
ARIMA(3,1,3)	0.91	5421.639	6.741	3947.214	17.045
ARIMA(4,1,0)	0.92	5490.214	6.852	3825.147	17.026
ARIMA(4,1,1)	0.945	5497.325	6.963	3921.258	17.023
ARIMA(4,1,2)	0.942	5436.215	6.874	3987.214	17.147
ARIMA(4,1,3)	0.948	5438.254	6.987	3978.214	17.369
ARIMA(4,1,4)	0.935	5436.147	6.897	3910.410	17.361

From the above table it is clear that among the different ARIMA model the ARIMA (0,1,1) has the higher value of R^2 with minimum values of Normalize BIC,RAPE,MAE in comparison to that of the other model and the tentative model for the paddy production data set was ARIMA (0,1,1). By using the SPSS software package the model parameters were estimated and presented in table 3. It is clear from the table 3 that all the parameter estimates were found to be highly significant. Since all the model selection criterian measures were found to be minimum and the coefficient determination was 0.95 which means that 95% of variation in the data series was explained by the ARIMA (0, 1, 1) model.

Table 3: Estimates of ARIMA (0, 1, 1) for paddy production

coefficients	Estimates	S.E	t-value	p-value
Constant	1351.489	191.246	7.067	.000
MA1	0.756	0.095	7.732	.000

Finally the fitted ARIMA (0,0,1) model was found to be most appropriate one among ARIMA time series model employed to the paddy production. The trend values in paddy production in the figure-5

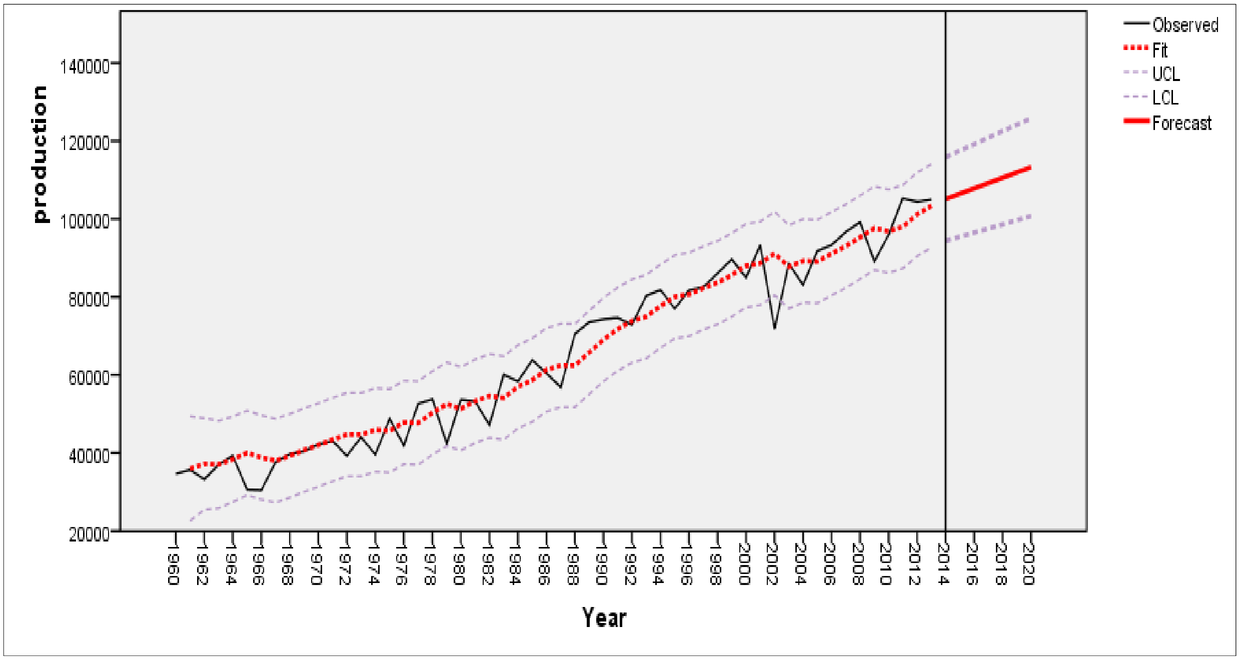

Fig 5: trend values for paddy production based on ARIMA model

Trends in Paddy Production Based On Different Smoothing Techniques

Table 4: Estimates of parameter an their testing

Fitted statistic	SES (α=0.55)	Holt's α =0.205 β=0.00003	Brown's α =0.204	Damped α =0.234 β=0.00099 φ=0.989	ARIMA (0,1,1)
R^2	0.92	0.95	0.93	0.93	0.95
RMSE	6056.365	5283.815	5521.746	5330.277	5308.000
MAPE	8.283	6.938	7.795	6.897	6.792
MAE	4984.68	3875.958	4390.487	3894.645	3885.821
Normalized BIC	17.498	17.589	17.369	17.384	17.345

From the results presented in table 4 it is clear that among the different smoothing techniques, the Holt's winter smoothing technique was found to be better one in comparison to other smoothing techniques. Since the model selection criteria values were found to be less than that of other models. Since the data set contains the linear trend the exponential smoothing techniques was not found suitable. As all the ACF and PACF values were found to be within the confidence interval the residuals due to the Holt's linear exponential smoothing possessed a white noise is given in figure- 6.

The trend in paddy production based on Holt's linear exponential smoothing in the figure-7. Increased trends are observed. The forecast of paddy production based on holt's and ARIMA (0,0,1) models are given in the table 5.

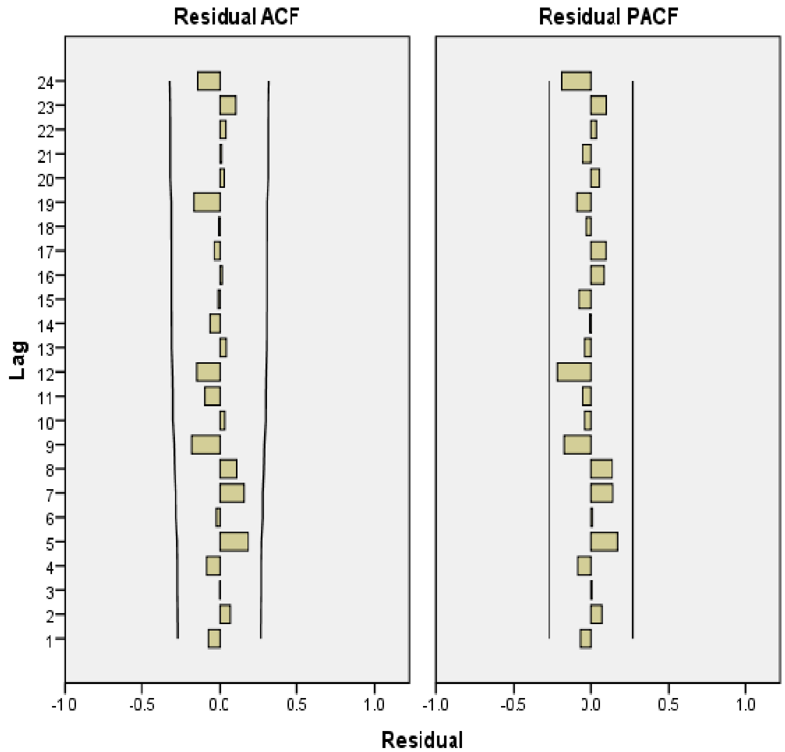

Fig 6: residual plots of ACF AND PACF of holt's trend method

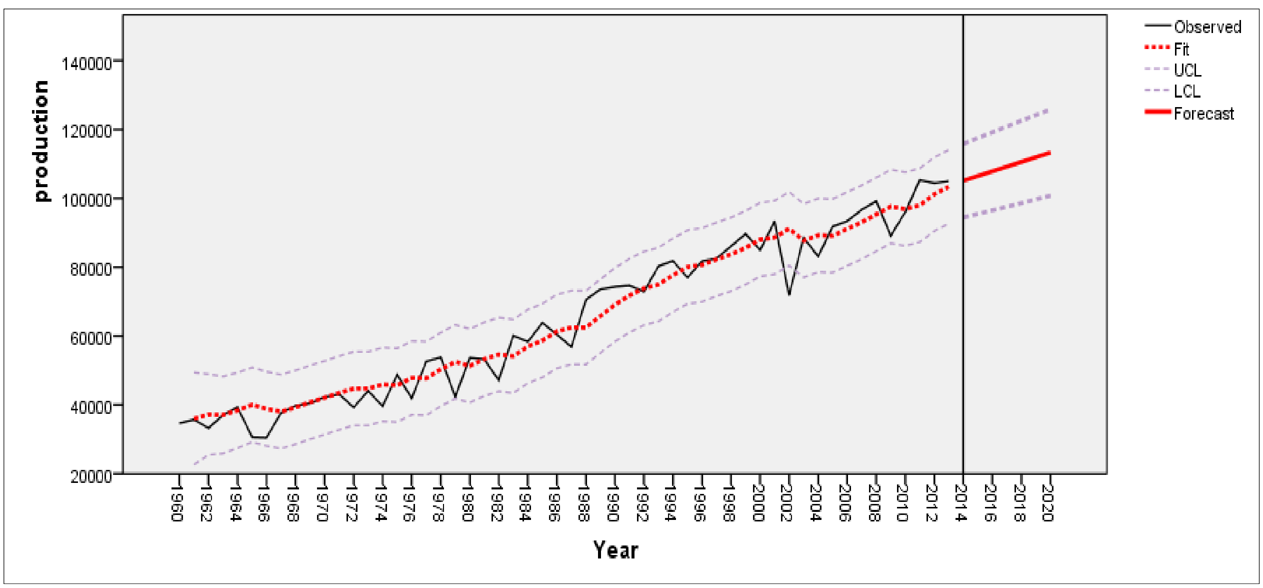

Fig 7: trend values for paddy production based on Holt's methods

Table 5

Year	production	
	Holt's method	ARIMA(0,0,1)
2016	10519	10258
2017	11253	10369
2018	12358	11258
2019	12458	12369
2020	12469	12565

Conclusion

In this paper, we developed three model for Paddy production in Tamilnadu, were found to be ARIMA (0,1,1), Brown's (α= 0.205), and Damped (α= 0.234,γ = 0.0099,ϕ = 0.989). Performance evaluation measures viz. We can see that the RMSE and MAPE (in Table 4) for each model very small. Among the different exponential smoothing method Holt's winter smoothing was selected based on model selection criteria. Among the times series model ARIMA (0,1,1) was found to be an appropriate model. These two models were found to be more appropriate model to forecast the paddy production data set. However, Holts winter model may be preferred which is a non-parametric model; based on the original time series paddy production data and the stationary condition need not be verified, whereas the ARIMA time series models are based on stationary data.

References

1. Box GEP, Jenkins GM. Time Series Analysis, Forecasting and Control. San Francisco, Holden- Day, California, USA, 1970.
2. Box GEP, GM Jenkins. Time Series Analysis, Forecasting and Control. San Francisco, Holden-Day, California, USA, 1976.
3. Brown RG, RF Meyer, DAD Esopo. The fundamental theorem of exponential smoothing, Operations Research. 1961; 9(5):673-685.
4. Gardner ES. Exponential smoothing-The state of the art, J Forecasting. 1985, 1-28.
5. Holt CC. Forecasting seasonal and trends by exponentially weighted moving averages, International Journal of Forecasting. 2004; 20(1):5-10.
6. Makridakis S, Wheelwright SC, Hyundman RJ. Forecasting: methods and applications, J Wiley and Sons: New York, 1998.
7. Makridakis S, Hibbon M. Accuracy of forecasting an empirical investigation, J Roy. Statist. Soc. A. 1979; 41(2):97-145.
8. Pankratz A. Forecasting with univariate box Jenkins models concept and cases, J.Wiley and sons, New York, 1983.
9. Wold HOA. A study of the analysis of stationary time series, Uppasala: Almqvist and Wiksells, 1983.
10. Yule GU. Why do we sometimes get nonsense-correlations between time series, Journal of Royal Statistical Society. 1926; 89(1):1-63.