Volume : 3 | Issue : 1 | January 2014 • ISSN No 2277 - 8179

Research Paper

Stochastic Modeling for Forecasting of India's Milled Rice Production

Statistics

KEYWORDS : Brown, Damped exponential smoothing; ARIMA model

M.Thirunavukkarasu	Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, INDIA
A.Rajarathinam	Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, INDIA

ABSTRACT

Time series analyses, were used for modeling and forecasting future milled rice production in India. Initially, each series was recorded yearly from 1960 to 2013. These data include trend variations which allow the use of ARIMA (AutorRegressive Integrated Moving Average), Brown exponential smoothing and Damped exponential smoothing models for predictions of future behavioral patterns. All model forecasting accuracy degree of statistical validity.

1. Introduction

Time -series forecasting is a technique that helps to predict what will occur in the future if trends do not change. A time series, also called historical or chronological series, is a sequence of observed values of a variable referring to different moments or periods of time, which are generally regular. Univariate time -series analysis consists of making use of these data to elaborate a model that describes the behavior of this variable acceptably for the past and, based on this model, allows making satisfactory forecasts for the future.

2. Materials and Methods

The main objective of this research was to model and forecast the Milled Rice production of India. The data used for the present study have been taken from United States Department of Agricultural for the periods 1960-2013. As our data are univariate time series, so Univariate time series Models have been tried to fit on the current data.

3. METHODOLOGY

3.1 Exponential Smoothing forecasting methodology

After ensuring the presence of trend in the data, smoothing of the data is the next requirement for time series analysis. a linear trend into account represent an improvement on simple exponential smoothing, it does not deal with more complex type of trends. Exponential smoothing is a method, conceived of by Robert Macaulay in 1931 and developed by Robert G. Brown during world war II , for extrapolative forecasting from series data. The more sophisticated exponential smoothing methods seek to isolate trends or seasonality from irregular variation. Where the patterns are found the more advanced methods identify and model these patterns. The model can then incorporate those patterns into the forecast. When used for forecasting, exponential smoothing uses weighted averages of the past data. The both smoothing constant weights can range between 0 to 1.0.

3.1.1 Brown's Exponential smoothing

The smoothing equations are

$$L_t = \alpha Y_t + (1-\alpha) L_{t-1} \quad (1)$$

$$T_t = \alpha (L_t - L_{t-1}) + (1-\alpha) T_{t-1} \quad (2)$$

The h-step-ahead prediction equation is

$$\hat{Y}_{t+h} = L_t + (\ h-1)+1/\alpha) T_t \text{, h= 1,2,...} \quad (3)$$

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, by $(\ h-1)+1/\alpha)$

3.1.2 Damped-Trend Exponential smoothing

The smoothing equation is

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + \varphi T_{t-1}) \quad (4)$$

$$T_t = \gamma (L_t - L_{t-1}) + (1-\gamma) \varphi T_{t-1} \quad (5)$$

The h-step-ahead prediction equation is

$$\hat{Y}_{t+h} = L_t + \sum_{i=1}^{k} \varphi^i T_t \quad (6)$$

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.

3.2 Stochastic Box-Jenkins forecasting methodology

This model popularly known as Box - Jenkins (1970) forecasting model is a regression based model in which forecasted values are obtained by regressing past values of the variable itself and current and past values of error term with different lag length. The ARIMA model combines autoregressive and moving average process with integration of order d, which should be stationary. The application of the model requires four steps, such as identification, estimation, diagnostic checking and forecasting. As the model is very powerful and useful with stationary data, the stationary of the series can be tested with correlogram or unit root tests. After verifying stationary properties, it is essential to find out highest order of autoregressive process 'p' through partial autocorrelations function and highest order of moving average process 'q' through autocorrelation function. The same can be verified with alternative model selection criteria such as Akaike Information Criteria, Schwarz Bayesian Criteria, Adjusted $R^{2,}$ and Final Prediction Error and so on. Once the order of p and q selected the specified regression model is estimated with maximum likelihood estimation procedure. The next step is to go for diagnostic checking whether the estimated results are correct or not. It's obtained through above mentioned model selection criteria's. Also it is verified with ACF and PACF obtained from the residual of the specified ARIMA model. If the residual is free from all classical assumption of the regression model and stationary then the model is correct. Additionally, diagnostic checking has been performed with χ^2 and Ljung-Box Q statistics. Lastly obtain the forecasted values by estimating the regression model with different forecasting period ahead recursively.

Autoregressive process of order (p) is,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \quad (7)$$

Moving Average process of order (q) is

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (8)$$

and the general form of ARIMA model of order (p,d,q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (9)$$

Volume : 3 | Issue : 1 | January 2014 • ISSN No 2277 - 8179

where Y_t is milled rice production, ε_t's are independently and normally distributed with zero mean and constant variance σ_2 for t = 1,2,..., n ; d is the fraction differenced while interpreting AR and MA and ϕ s and θ s are coefficients to be estimated.

The ARIMA model is given by (taking Z_t as the first differenced series, in our case d=1)

$$(Z_t - \mu) - \alpha_1 (Z_{t-1} - \mu) - \ldots - \alpha_p (Z_{t-p} - \mu) =$$

$e_t - \beta_1 e_{t-1} - \ldots - \beta_q e_{t-q}$ is called as ARIMA (p,1,q) of order (p,q).

4. Evaluation Criteria

The performances of different approaches have been evaluated on the Box-Ljung Q statistics was used to transform the non-stationary data in to stationarity data and to check the adequacy for the residuals and Bayesian Information Criterion (BIC) were used. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics which are given by

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right| \times 100 \qquad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum (Y_t - F_t)^2} \qquad (11)$$

Where Y_t is the original milled rice yield in different years and F_t is the forecasted rice yield in the corresponding years and n is the number of years used as forecasting period.

BIC(p,q) = In v* (p,q) + (p+q) [In(n) /n] ; 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.

$$Q = \frac{n(n+2)\sum_{i=1}^{k} k^{2}}{(n-k)} \qquad (11)$$

where n is the number of residuals and irk is the residuals autocorrelation at lag k.

5. Results And Discussion

Exponential Smoothing Linear model: Time plot (Fig. 1) of milled rice production data revealed that there is increasing trend in the data period 1960 to 2013

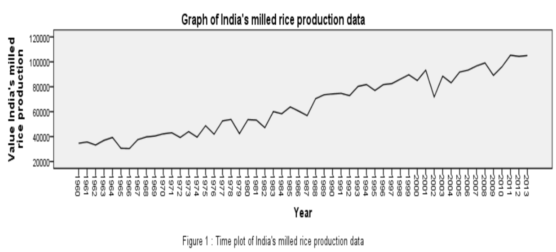

Fig. 1 Time plot of India's milled rice production

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 = 0.204 (Fig. 2 and Table 4).

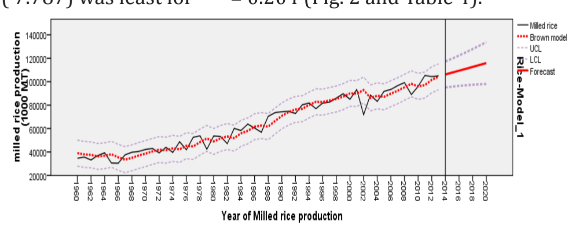

Fig. 2 Forecasting of Brown smoothing Exponential

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 = 0.204, γ =0.00006669 and φ =0.999 (Fig. 3 and Table 4).

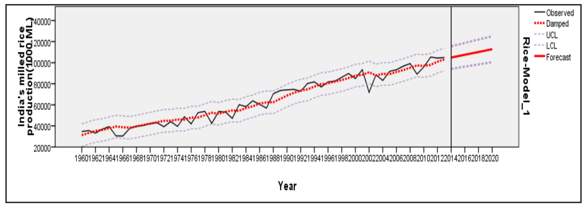

Fig. 3 Forecasting of Damped Exponential smoothing

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 sample ACFs and PACFs were plotted (Fig. 4) and Table 1.

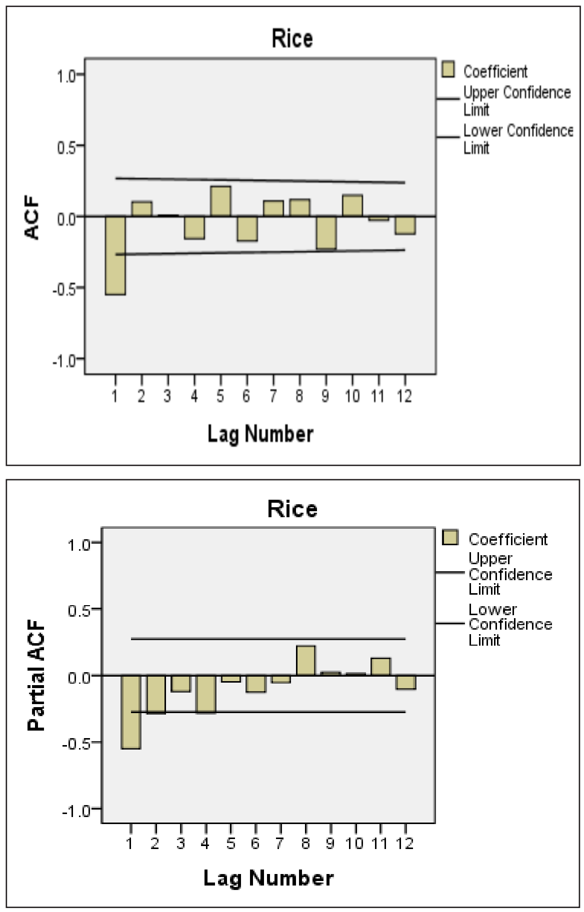

Fig 4. ACF and PACF of differenced data

The tentative ARIMA models are discussed with values differenced once (d=1) and the model which had the minimum normalized BIC was chosen. The various ARIMA models and the corresponding normalized BIC values are given in Table 2. The value of normalized BIC of the chosen ARIMA was 17.371. R^2 value was 0.947. Hence, the most suitable model for milled rice production was ARIMA (0,1,1), as this model had the lowest normalized BIC value and better model fit statics (RMSE and MAPE).

Table 1. ACF and PACF of milled rice production

Lag	Autocorrelation		Box-Ljung Statistics			Partial autocorrelation	
	value	Df	Sig	Value	Df	Value	Df
1	-.549	.134	16.911	1	.000	-.549	.137
2	.102	.132	17.510	2	.000	-.285	.137
3	.007	.131	17.514	3	.001	-.120	.137
4	-.156	.130	18.967	4	.001	-.283	.137
5	.212	.128	21.690	5	.001	-.047	.137
6	-.173	.127	23.541	6	.001	-.126	.137
7	.108	.126	24.279	7	.001	-.053	.137
8	.117	.124	25.173	8	.001	.221	.137
9	-.231	.123	28.712	9	.001	.023	.137
10	.148	.121	30.202	10	.001	.014	.137
11	-.027	.120	30.254	11	.001	.130	.137
12	-.123	.119	31.328	12	.002	-.104	.137

Table 2.

ARIMA (p,d,q)	BIC values
(0,1,0)	17.371
(0,1,1)	17.316
(0,1,2)	17.403
(1,1,0)	17.466
(1,1,1)	17.402
(1,1,2)	17.494
(2,1,0)	17.475
(2,1,1)	17.495
(2,1,2)	17.592

ARIMA Model estimation: Model parameters were estimated using SPSS package and the results of estimation are presented in Table 3 and 4.

Table 3. Estimated ARIMA model of milk production

	Estimate	SE	T	Sig.
Constant	1362.075	194.296	7.010	.000
MA1	0.751	0.098	7.630	.000

Table 4. Estimates of Parameters and their Testing

Fit Statistic	ARIMA(0,1,1)	Brown's α = (0.204)	Damped α = 0.229, γ = 0.00007, ϕ = 0.999
R-squared	0.947	0.943	0.949
RMSE	5340.346	5520.359	5337.483
MAPE	6.786	7.787	6.896
Normalized BIC	17.316	17.306	17.387

ARIMA Model Diagnostic checking : The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARIMA, which has been done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, various autocorrelations up to 12 lags were computed and the same along with their significance tested by Box-Ljung statistic are provided in Table 5. As the results indicate, none of these autocorrelations was significantly different from zero at any reasonable level. This proved that the selected ARIMA model was an appropriate model for forecasting milled rice production in India.

Table 5. Residual of ACF and PACF of milled rice production

Lag	ACF		PACF	
	Mean	SE	Mean	SE
Lag 1	-0.071	0.137	-0.071	0.137
Lag 4	-0.083	0.139	-0.088	0.137
Lag 6	-0.022	0.144	0.012	0.137
Lag 8	0.112	0.147	0.133	0.137
Lag 10	0.029	0.153	-0.038	0.137
Lag 12	-0.143	0.154	-0.215	0.137
Lag 14	-0.061	0.157	-0.005	0.137
Lag 16	0.015	0.157	0.083	0.137
Lag 18	-0.008	0.158	-0.028	0.137
Lag 20	0.027	0.161	0.055	0.137
Lag 22	0.036	0.161	0.034	0.137
Lag 24	-0.138	0.163	-0.187	0.137

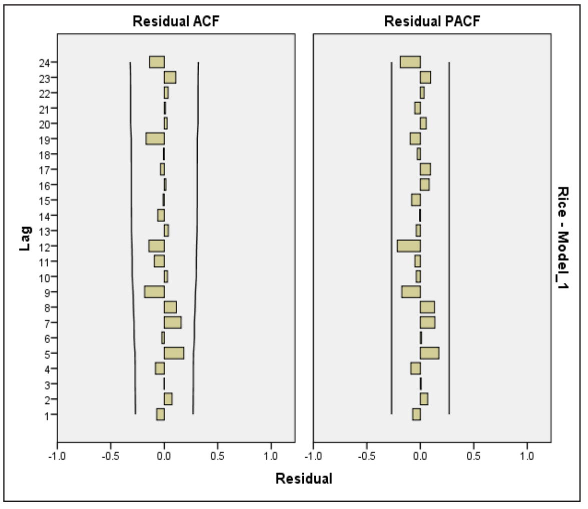

Fig 5. Residuals of ACF and PACF

The ACF and PACF of the residuals are given in Figure 5 which also indicated the 'good fit' of the model. Hence, the fitted ARIMA model for the milled rice production data was:

$$Y_t = 1362.075 - 0.751\varepsilon_{t-1} + \varepsilon_t$$

Forecasting. Based on the model fitted, forecasted rice product (Million Tones) for the year 2014 to 2020

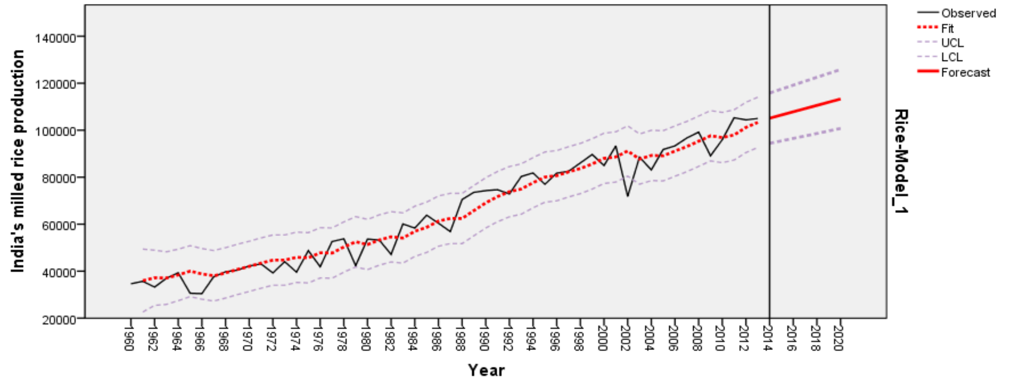

Fig 6. Forecasting of ARIMA model

Volume : 3 | Issue : 1 | January 2014 • ISSN No 2277 - 8179

Table 6. Forecasting of India's milled rice production (1960 - 2013)

Year	Rice	ARIMA	Brown's	Damped
2001	93340	88629	89837	88467
2002	71820	91166	92921	90923
2003	88530	87702	86117	87889
2004	83130	89271	88025	89374
2005	91790	89101	87053	89282
2006	93350	91134	89805	91192
2007	96690	93049	92269	93022
2008	99180	95319	95237	95196
2009	89090	97644	98194	97441
2010	95980	96873	95994	96862
2011	105310	98012	97122	97992
2012	104400	101194	101595	100998
2013	105000	103356	104213	103107
2014	-	105128	106125	104869
2015	-	106490	107748	106197
2016	-	107852	109372	107525
2017	-	109215	110995	108851
2018	-	110577	112619	110176
2019	-	111939	114242	111501
2020	-	113301	115866	112824

The most appropriate stochastic model for milled rice production forecasting was found to be ARIMA , Brown's and Damped model. From the forecast available from the fitted ARIMA, Brown's and Damped model, it can be found that forecasted rice production would increase to 113301(MT), 115866(MT) , 112824(MT) in 2020 from 105000(MT) in 2013. This study provides evidence on future rice product in the country, which can be considered for future policy making and formulating strategies for augmenting and sustaining rice product in India.

6. Conclusion

In this paper, we developed three model for India's milled rice production, were found to be ARIMA (0,1,1), Brown's (α = 0.204), and Damped (α = 0.229, γ = 0.00007, φ = 0.999). Performance evaluation measures viz. We can see that the RMSE and MAPE (in Table 4) for each model very small. All model R^2 values are 95%, so results revealed that among the models much better fitted

Acknowledgements

The second author expresses his gratitude to the UGC for providing fellowship to carry out this work under scheme of BSR-Fellowship.

REFERENCE

1. Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998). Forecasting - Methods and Applications. John Wiley and Sons, New York. | 2. Gardner, E.S. (1985). Exponential smoothing - The state of the art. J. Forecasting, 4, 1-28. | 3. Pankratz, A. (1983). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley and Sons, New York. | 4. Javed, M.S., 1996. Determination of Comparative economics of some planting patterns of sugarcane. Pakistan J. Agric. Sci., 33:50-3. | 5. Box, G.E.P., and G.M. Jenkins, 1976. Time Series Analysis and Control, Revised Ed, Holden Day. |