International Multidisciplinary Research Journal 2012, 2(1):10-15
ISSN: 2231-6302
Available Online: http://irjs.info/

INTERNATIONAL MULTIDISCIPLINARY RESEARCH JOURNAL

Forecasting milk production in Tamilnadu

T. Jai Sankar[1] and R. Prabakaran[2]

[1]Assistant Professor, Department of Statistics, Bharathidasan University, Tiruchirappalli – 620 024, Tamilnadu, India.
[2]Vice-Chancellor, Tamilnadu Veterinary and Animal Sciences University, Chennai – 600 051, Tamilnadu, India.

Abstract
Tamilnadu, with a daily milk production of 145.88 lakh litres, is one of the leading states in milk production in India. And the Tamilnadu milk cooperatives play a major role in the development dairy within the state. This study aims at forecasting milk production in Tamilnadu, based on data on milk production during the years from 1978 to 2008. The study considered Autoregressive (AR), Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) processes to select the appropriate stochastic model for forecasting milk production in Tamilnadu. Based on ARIMA (p, d, q) and its components ACF, PACF, Normalized BIC, Box-Ljung Q statistics and residuals estimated, ARIMA (1, 1, 0) was selected. Based on the chosen model, it could be predicted that the milk production would increase to 7.15 million tons in 2015 from 5.96 million tons in 2008 in Tamilnadu.

Keywords: Milk production, Normalized BIC, forecasting, ARIMA.

INTRODUCTION

On global basis India is able to produce milk at very competitive prices by virtue of utilizing crop residues for rearing the animals. India is predominantly an agrarian society where animal husbandry forms the backbone of national economy. Dairying provides millions of small marginal farmers and landless labours means for their subsistence. Milch animals are reared mainly through the utilization of crop residues; thus milk production is essentially a subsidiary activity to agriculture. The planners recognized dairying, because of the potential impact it can make, as an instrument to bring about socio-economic transformations in the rural sector. Cows account was about 44% of the milk produced in India, buffalo 53%, and goat, sheep and camel etc. account for the balance 3%. Thus, dairying in India is a sharp contrast to the pattern of milk production in the advanced countries of the world, where specialized dairy farming is in vogue. Tamilnadu State's share in total milk production at the All India level was 5.38%.

Tamilnadu, with a daily milk production of 145.88 lakh litres, is one of the leading states in milk production in India. And the Tamilnadu milk cooperatives play a major role in the development dairy within the state. The milk cooperatives of Tamilnadu, with the help from the government and National Dairy Development Board (NDDB), have played substantial roles in taking the state to the current position. Besides, NDDB also undertakes methodical approach and appropriate strategy for the upliftment of Tamilnadu milk cooperatives. The dairy development programmes in Tamilnadu have been implemented through a wide network of co-operatives, which follow the three-tier structure where primary milk producers' co-operative societies remain at the base level. In the district level, there is a union of producers' co-operative societies, whereas, Federation of District Co-operative Milk Producers' Union remain at the top (state) level of the structure.

In Tamilnadu, Milk Producers' Cooperative Societies function at the village level, where milk producers get enrolled as members. The members get animal health cover for their animals; breed improvement is also carried out. District unions collect the milk produced at the village societies. In this background, this study was conducted to forecast the future milk production in the State, so as to help the policy planners to formulate needed strategies for achieving and sustaining the targets in the sector.

MATERIAL AND METHODS

As the aim of the study was to forecast milk production, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1970), was frequently used for discovering the pattern and predicting the future values of the time series data. Akaike (1970) discussed the stationary time series by an AR(p), where p is finite and bounded by the same integer. Moving Average (MA) models were used by Slutzky (1973). Hannan and Quinn (1979) suggested obtaining the order of a time series model by minimizing the errors for pure AR models, and Hannan (1980) for ARMA models. A second order determination method could be considered as a variance of Schwarz's Bayesian Criterion (SBC) which gives a consistent estimate of the order of an ARMA model. Hosking (1981) introduced a family of models, called fractionally differenced autoregressive integrated moving average models, by generalizing the 'd' fraction in ARIMA (p, d, q) model.

Stochastic time-series ARIMA models were widely used in time series data having the characteristics (Alan Pankratz, 1983) of parsimonious, stationary, invertible, significant estimated coefficients and statistically independent and normally distributed residuals. When a time series is non-stationary, it can often be made stationary by taking first differences of the series i.e., creating a new time series of successive differences (Y_t-Y_{t-1}). If first differences do not convert the series to stationary form, then first differences can be created. This is called second-order differencing. A distinction is made between a

Received: Sept 11, 2011; Revised: Dec 18, 2011; Accepted: Dec 26, 2011.

*Corresponding Author

T. Jai Sankar
Assistant Professor, Department of Statistics, Bharathidasan University,
Tiruchirappalli – 620 024, Tamilnadu, India.

Tel: +91-9444469629
Email: tjaisankar@gmail.com

second-order differences (Y_t-Y_{t-2}).

While Mendelssohn (1981) used Box-Jenkins models to forecast fishery dynamics, Prajneshu and Venugopalan (1996) discussed various statistical modeling techniques viz., polynomial, ARIMA time series methodology and nonlinear mechanistic growth modeling approach for describing marine, inland as well as total fish production in India during the period 1950-51 to 1994-95. Tsitsika *et al.* (2007) also used univariate and multivariate ARIMA models to model and forecast the monthly pelagic production of fish species in the Mediterranean Sea during 1990-2005.

The time series when differenced follows both AR and MA models and is known as autoregressive integrated moving averages (ARIMA) model. Hence, ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated using the Statistical Package for Social Sciences (SPSS) package and to fit the ARIMA models.

Autoregressive process of order (p) is,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \,;$$

Moving Average process of order (q) is,

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t \,;$$

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

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y_t is milk 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 ϕs and θs are coefficients to be estimated.

Trend fitting

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. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC) [as suggested by Schwartz, 1978] were used. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right]^{1/2} ; \quad MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right|$$

BIC(p,q) = ln v*(p,q) + (p+q) [ln (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.

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

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

In this study, the data on milk production in Tamilnadu were collected from the Tamilnadu Cooperative Milk Producers Federation, Government of Tamilnadu for the period from 1978 to 2008 and were used to fit the ARIMA model to predict the future production.

RESULTS AND DISCUSSION

Model identification

ARIMA model was designed after assessing that transforming the variable under forecasting was a stationary series. The stationary series was the set of values that varied over time around a constant mean and constant variance. The most common method to check the stationarity is to explain the data through graph and hence is done in Figure 1.

Figure 1 reveals that the data used were non-stationary. Again, non-stationarity in mean was corrected through first differencing of the data. The newly constructed variable Y_t could now be examined for stationarity. Since, Y_t was stationary in mean, the next step was to identify the values of p and q. For this, the autocorrelation and partial autocorrelation coefficients (ACF and PACF) of various orders of Y_t were computed and presented in Table 1 and Figure 2.

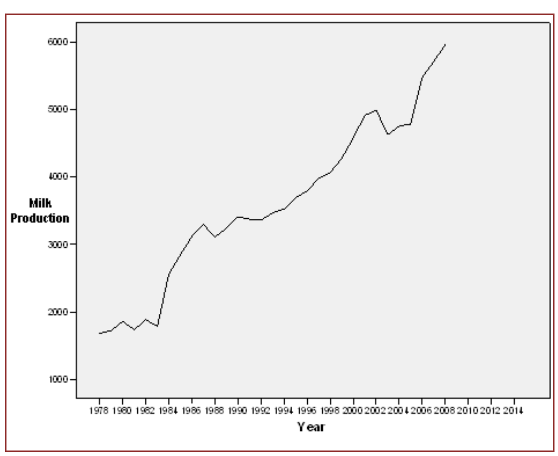

Fig 1. Time plot of milk production in Tamilnadu

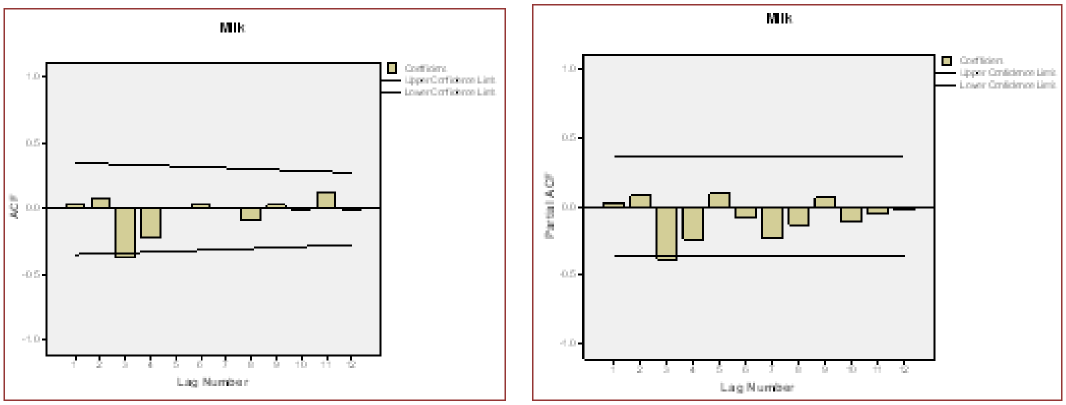

Fig 2. ACF and PACF of differenced data

Table 1. ACF and PACF of milk production

Lag	Auto Correlation		Box-Ljung Statistic			Partial Auto Correlation	
	Value	Df	Sig.	Value	Df	Value	Df
1	0.027	0.174	0.025	1	0.876	0.027	0.183
2	0.084	0.171	0.266	2	0.875	0.083	0.183
3	-0.376	0.168	5.302	3	0.151	-0.384	0.183
4	-0.222	0.165	7.127	4	0.129	-0.234	0.183
5	0.004	0.161	7.128	5	0.211	0.096	0.183
6	0.026	0.158	7.155	6	0.307	-0.085	0.183
7	-0.002	0.155	7.155	7	0.413	-0.232	0.183
8	-0.086	0.151	7.476	8	0.486	-0.131	0.183
9	0.027	0.148	7.509	9	0.584	0.067	0.183
10	-0.011	0.144	7.515	10	0.676	-0.108	0.183
11	0.124	0.141	8.293	11	0.687	-0.045	0.183
12	-0.013	0.137	8.302	12	0.761	-0.022	0.183

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 11.195.

Table 2. BIC values of ARIMA (p, d, q)

ARIMA (p, d, q)	BIC values
(0, 1, 0)	11.245
(0, 1, 1)	11.197
(0, 1, 2)	11.290
(1, 1, 0)	**11.195**
(1, 1, 1)	11.243
(1, 1, 2)	11.410
(2, 1, 0)	11.340
(2, 1, 1)	11.487
(2, 1, 2)	11.483

Model estimation

Model parameters were estimated using SPSS package and the results of estimation are presented in Tables 3 and 4. R^2 value was 0.96. Hence, the most suitable model for milk production was ARIMA (1, 1, 0), as this model had the lowest normalized BIC value, good R^2 and better model fit statics (RMSE and MAPE).

Table 3. Estimated ARIMA model of milk production

	Estimate	SE	t	Sig.
Constant	-5980.471	9723.478	-0.615	0.544
AR 1	0.018	0.193	0.091	0.928

Table 4. Estimated ARIMA model fit statistics

Fit Statistic	Mean
Stationary R-squared	0.015
R-squared	0.966
RMSE	227.534
MAPE	4.491
Normalized BIC	11.195

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 milk production in Tamilnadu.

Table 5. Residual of ACF and PACF of milk production

Lag	ACF		PACF	
	Mean	SE	Mean	SE
Lag 1	-0.001	0.183	-0.001	0.183
Lag 2	0.084	0.183	0.084	0.183
Lag 3	-0.360	0.184	-0.362	0.183
Lag 4	-0.204	0.206	-0.232	0.183
Lag 5	0.032	0.213	0.108	0.183
Lag 6	0.014	0.213	-0.086	0.183
Lag 7	-0.021	0.213	-0.240	0.183
Lag 8	-0.107	0.213	-0.135	0.183
Lag 9	0.013	0.215	0.047	0.183
Lag 10	-0.018	0.215	-0.133	0.183
Lag 11	0.120	0.215	-0.061	0.183
Lag 12	-0.021	0.217	-0.044	0.183

The ACF and PACF of the residuals are given in Figure 3, which also indicated the 'good fit' of the model. Hence, the fitted ARIMA model for the milk production data was:

$$Y_t = -5980.471 - 0.018Y_{t-1} + \varepsilon_t$$

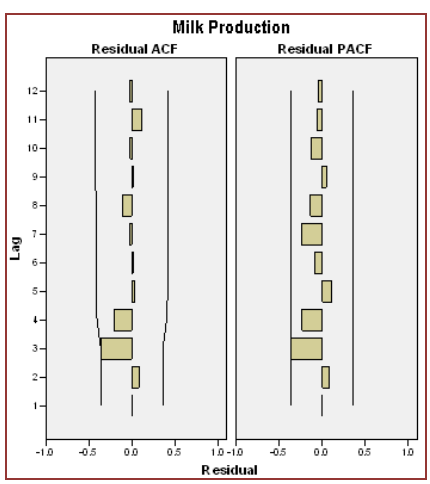

Fig 3. Residuals of ACF and PACF

Forecasting

Based on the model fitted, forecasted milk production (in million tons) for the year 2009 through 2015 respectively were 6081, 6219, 6400, 6584, 6770, 6959 and 7150 million tons (Table 6). To assess the forecasting ability of the fitted ARIMA model, the measures of the sample period forecasts' accuracy were also computed. This measure also indicated that the forecasting inaccuracy was low. Figure 4 shows the actual and forecasted value of milk production (with 95% confidence limit) in the State.

Table 6. Forecast of milk production (in million tons) in Tamilnadu

Year	Actual	Predicted	LCL	UCL
1978	1681	--	--	--
1979	1727	1785	1264	2307
1980	1860	1817	1318	2316
1981	1738	1957	1481	2433
1982	1886	1809	1342	2277
1983	1788	1925	1459	2391
1984	2562	1904	1440	2367
1985	2846	2786	2323	3250
1986	3118	3280	2818	3743
1987	3295	3227	2765	3690
1988	3109	3364	2902	3827
1989	3238	3203	2741	3666
1990	3410	3262	2799	3724
1991	3375	3597	3134	4059
1992	3357	3522	3059	3984
1993	3468	3352	2890	3815
1994	3524	3563	3101	4026
1995	3694	3709	3247	4172
1996	3791	3818	3355	4280
1997	3977	3925	3462	4387
1998	4061	4128	3666	4591
1999	4273	4220	3758	4683
2000	4574	4411	3948	4873
2001	4909	4797	4334	5259
2002	4988	5171	4709	5634
2003	4622	5154	4692	5617
2004	4753	4568	4105	5030
2005	4784	4725	4263	5188
2006	5474	5052	4590	5515
2007	5705	5780	5317	6242
2008	5961	6052	5589	6514
2009	--	6081	5618	6543
2010	--	6219	5477	6961
2011	--	6400	5316	7484
2012	--	6584	5243	7926
2013	--	6770	5213	8327
2014	--	6959	5213	8705
2015	--	7150	5234	9067

CONCLUSION

The most appropriate ARIMA model for milk production forecasting was found to be ARIMA (1, 1, 0). From the forecast available from the fitted ARIMA model, it can be found that forecasted production would increase to 7.15 million tons in 2015 from 5.96 million tons in 2008. That is, using time series data from 1978 to 2008 on milk production, this study provides evidence on future milk production in the State, which can be considered for future policy making and formulating strategies for augmenting and sustaining milk production in the State.

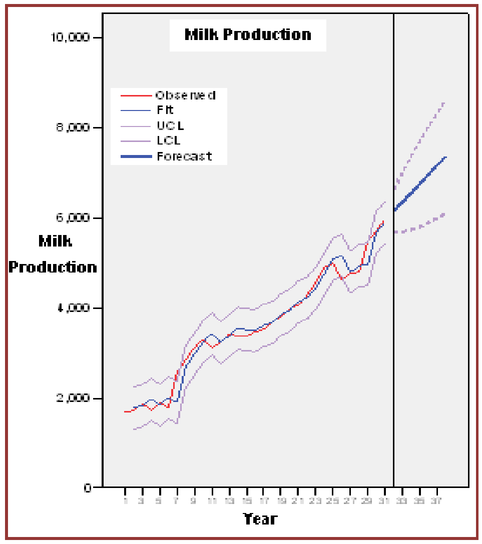

Fig 4. Actual and estimate of milk production

REFERENCES

[1] Akaike H. 1970. Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics* 22: 203-270.

[2] Alan Pankratz. 1983. Forecasting with univariate Box–Jenkins models - concepts and cases. John Wiley, New York, Page 81.

[3] Box G E P and Jenkins J M. 1970. Time Series Analysis - Forecasting and Control. Holden-Day Inc., San Francisco.

[4] Hannan E J and Quinn B G. 1979. The determination of the order of an autoregression. *Journal of Royal Statistical Society* B(41): 190-195.

[5] Hannan E J. 1980. The estimation of the order of an ARMA process. *Annuals of Statistics* 8: 1071-1081.

[6] Hosking J R M. 1981. Fractional differencing. *Biometrika* 68(1): 165-176.

[7] Mendelssohn R. 1981. Using Box-Jenkins models to forecast fishery dynamics: identification, estimation and checking. *Fishery Bulletin* 78(4): 887-896.

[8] Prajneshu and Venugopalan R. 1996. Trend analysis in all India marine products export using statistical modeling techniques. *Indian Journal of Fisheries* 43(2): 107-113.

[9] Slutzky E. 1973. The summation of random causes as the source of cyclic processes. *Econometrica* 5:105-146.

[10] Tsitsika E V, Maravelias C D and Haralabous J. 2007. Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models. *Fisheries Science* 73: 979-988.