Forecasting of milk production in Tamil Nadu: an application of Arima model

C.Balan*[1], M.Thirunavukkarasu[2], G. Senthil Kumar[3] and M. Boopathy Raja[4]
Department of Animal Husbandry Statistics and Computer Applications
Faculty of Basic Sciences, Madras Veterinary College Campus
Tamil Nadu Veterinary and Animal Sciences University
Chennai – 600 007, Tamil Nadu, India

ABSTRACT

A study was made to forecast the milk production in Tamil Nadu, using the milk production data of the State from 1978-79 to 2018-19 and the Auto Regressive Integrated Moving Average (ARIMA) model for estimating future milk production. The autoregressive (p) and moving average (q) parameters were identified based on the significant spikes in the correlogram plots of Partial Auto Correlation Function (PACF) and Auto Correlation Function (ACF) of time series data. The adequacy of the fitted model was verified by the test of significance of residuals using Box-Ljung statistic. The results indicated that ARIMA (0, 1, 0) model was found to be the good model, based on the minimum values of selection criteria, viz., Akaike Information criteria (AIC) and Bayesian Information Criteria (BIC). The results also indicated the non-significance of Box-Ljung statistic and that the residual was normally distributed. Based on the model, the predicted figures of milk production for the next five years will be viz., 2019-20, 2020-21, 2021-22, 2022-23 and 2023-24 are 8529, 8696, 8863, 9030 and 9197 thousand tons in the State, respectively.

Key Words: Milk Production, Tamil Nadu, Forecasting, ARIMA model

INTRODUCTION

Dairy subsector plays a vital role in sustaining rural livelihood in India. Although the per capita availability of milk in Tamil Nadu is currently (2018-19) 322 g per day, greater than the Indian Council of Medical Research (ICMR) recommendation of per capita milk consumption (300 g/ day), there is a continuous rise in the demand for milk and milk products, due to the increasing per capita income, changing food consumption pattern and rapid urbanization. Hence, predicting milk production in the future is important for the scientists, planners and administrators to frame suitable policy plans, so as to achieve the supply requirement of milk for the State. Hence, this study was attempted to forecast the milk production in Tamil Nadu for the next five years (2019-20 through 2023-24).

MATERIALS AND METHODS

Data

To achieve the objective of the study, secondary data on annual milk production

Corresponding Author Email: drbala005@gmail.com
[1]Assistant Professor
[2] Professor and Head
[3] Assistant Professor, Office of the Registrar, TANUVAS, Chennai - 600051
[4]Assistant Professor, Dept. of Veterinary and Animal Husbandry Extension Education, Veterinary College and Research Institute, Tirunelveli - 627351

in Tamil Nadu relating to the period of four decades (from 1978-79 to 1999-2000) were collected from various reports of Integrated Sample Survey, Directorate of Animal Husbandry and Veterinary Services (Govt. of Tamil Nadu) and National Dairy Development Board (from 2001-02 to 2018-19).

Analytical tools

Annual Compound Growth Rate (ACGR) Analysis

The ACGRs of milk production in Tamil Nadu were estimated by as detailed below:

$$\text{Log } Y_t = \text{Log } Y_0 + t \text{ Log } (1+G)$$

Where, G=Annual Compound Growth Rate, Y_o=Production in base year, Y_t = Production in t^{th} year, t=Time in series (1978-79, 1979-80, ..., 2018-19).

Hence, G= (Antilog (1+r) -1) × 100, r = Regression coefficient.

Auto Regressive Integrated Moving Average (ARIMA) model

Pal *et al.* (2007) and Deshmukh and Paramasivam (2016) had earlier found that ARIMA (1, 1, 1) model was the best model for forecasting milk production in India. Hence, in the present study, forecasting of milk production was done using Auto Regressive Integrated Moving Average (ARIMA p, d, q) model. Box and Jenkins (1976) suggested ARIMA (p, d, q) model for forecasting using a specific time series dataset. Where 'p' is the number of autoregressive terms, 'd' is the number of non-seasonal differences needed for stationarity and 'q' is the number of lagged forecast errors in the prediction equation. The lags of the stationary series in the forecasting equation are called as Auto Regressive (AR) terms and lags of the forecast errors are called as Moving Average (MA) terms. The time series which need to be differenced to be made stationary is said to be an integrated version of a stationary series.

The Box-Jenkins methodology for analyzing and modeling a time series involves following steps of model identification, parameter estimation and model validation:

Auto Regressive Process of order (p) is

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

Moving Average Process of order (q) is

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_p \varepsilon_{t-q} + \varepsilon_t$$

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

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

where, Y_t - milk production at t^{th} year,

ε_{t-} independently and normally distributed with mean zero and constant variance for t = 1, 2, 3, …, n

ϕ_p and θ_p - coefficients to be estimated

Trend and stationarity of data

The first step in the time series analysis is the trend test for assessing the trend present in the data by using the Mann-Kendall test. The next step in identifying the perfect model is to find out the stationery of data, which is assessed by Augmented

Dickey-Fuller (ADF) Test. If the test statistic is significant, we can conclude that the data set is stationary, otherwise we need to go in for differentiation to make it stationary. In case the data are found to be non-stationary, stationarity is achieved by differencing technique. For instance, the differencing first order is $Y_t - Y_{t-1}$. If the first differences do not convert the series to stationary form, then the second differences can be created. It is called as the second order differencing ($Y_t - Y_{t-2}$). Further, the seasonality of the data series was also checked by the Webel- Ollech (WO) test. If the data series has the linear trend, stationarity and non-seasonality, then the ARIMA model is considered to be the best model for future projection.

Parameter estimation and diagnostic checking

In order to identify the order of Auto Regressiveness for the value 'p' and the order of Moving Average for the value of q', correlograms of PACF and ACF, respectively were examined. According to Tripathi *et al.,* (2014), the parameters 'p' and 'q' were obtained by looking for significant spikes in autocorrelation and partial autocorrelation functions. The lowest value of AIC and BIC were used as selection criteria for identification of the best fit ARIMA model.

$AIC = -2\log L + 2m$, where, $m = p + q$ and L is the likelihood function (Akaike, 1974)

In addition, the lower values of MSE (Mean Square Error) RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) also indicated that identified model was the most appropriate model to forecast milk production.

$$MAPE = \frac{1}{n}\sum\left|\frac{Y_t - F_t}{Y_t}\right| X100$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2}$$

$$BIC = \ln v\,(p, q) + (p + q)\,[\ln (n)/n]$$

where, Y_t is original milk yield in different years (t),

F_t is the forecasted milk yield in the corresponding years (t),

p is order of autoregressive (AR),

q is order of moving average (MA),

v is the estimate of variance, and

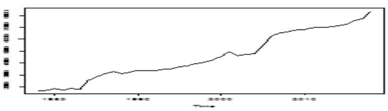
n is the number of observations.

Finally, the model verification is concerned with checking the residuals of the model using Box- Ljunk statistic. If the residuals are normally distributed, it can be concluded that the model is the best fit, otherwise unfit for forecasting. Forecasting for the next five years starting from 2019-20 to 2023-24 was done using the best fit ARIMA model. The 'R' software version 3.6.3 was used for time series data analysis for developing ARIMA models and forecasting of milk production in Tamil Nadu.

RESULTS AND DISCUSSION

The original series of milk production are represented in Figure 1, which implied that the production showed an increasing trend over the years. Various tests for the presence of trend, stationarity and seasonality of data series were performed and these results are presented in Table 1. The Mann Kendal test for trend analysis was performed to test the significance of the trend. The value of the test statistic showed a significant and positive trend (Table 1). Trend analysis also showed a considerable increase in all the years in milk production in Tamil Nadu from 1681 thousand tons in 1978-79 to 8362 thousand tons in 2018-19. In addition, the overall Annual Compound Growth Rate (ACGR) for the period starting from 1978-79 to 2018-19 for milk production was estimated as 3.97 per cent and presented in Table 4. The highest ACGR (8.48 per cent) was observed during the period 1978-79 to 1987-88. Whereas, the ACGR for the periods 1988-89 to 1997-98, 1998-99 to 2007-08, 2008-09 to 2018-19 were 2.38, 4.83 and 1.90 percentages respectively (Table 4).

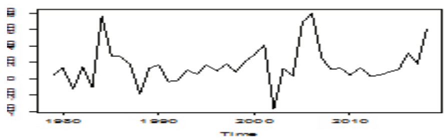
Figure 1: Line diagram of original milk production



The seasonality of the data series was checked by the WO test. The non-significant test statistic found in WO test confirmed the absence of seasonality (Table 1). Based on the results, it was confirmed that there was

Estimation of ARIMA model was performed after checking the stationarity of original data series. The most common method to check stationarity is Augmented Dicky Fuller (ADF) test. The ADF test revealed a non-significant test statistic of original data series and concluded that the data were non stationary. Hence, the first order differentiation was performed and made a stationarity of the data. The test statistic of ADF test revealed a significant value and confirmed that first order differenced data were stationary and the line diagram is depicted in Figure 2. Thus, the value of d was fixed as 1 in ARIMA (p, d, q) model.

Figure 2: Line diagram of diffrenced data series



a positive trend, stationarity in first order differentiation and non-seasonality of the data series and hence, it was concluded that the usage of ARIMA (p, d, q) model was the best model for projection of milk production in Tamil Nadu.

Table 1. Test of significance for the presence of trend, stationarity and seasonality of data

S. No.	Particulars	Name of the test	Test statistics	P value	Result
1	Test for Trend	Mann- Kendal test	8.92	<0.01	Presence of trend
2.	Test for Stationarity	Augmented Dicky Fuller test (ADF)for original data	-1.72	0.68	Data were non-stationary
		Augmented Dicky Fuller (ADF) test for 1 differenced data (d=1)	-3.85	0.03	Data were stationary
3.	Test for Seasonality	Webel-Ollech (WO) test	0.00	0.59	Absence of seasonality

The next step in ARIMA model was identification of the parameters of p and q. The p and q parameters were identified based on the significant spikes in the plots of PACF and ACF of the different time series. According to Figures 3 and 4, there were no significant spikes found and the value of p and q were fixed as 0, for both the parameters p and q.

Figure 3

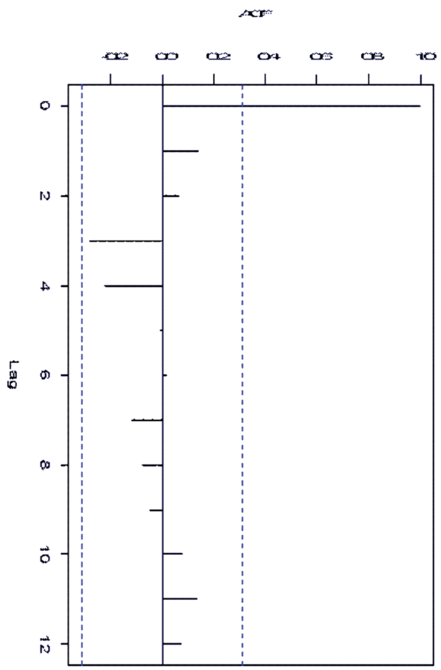
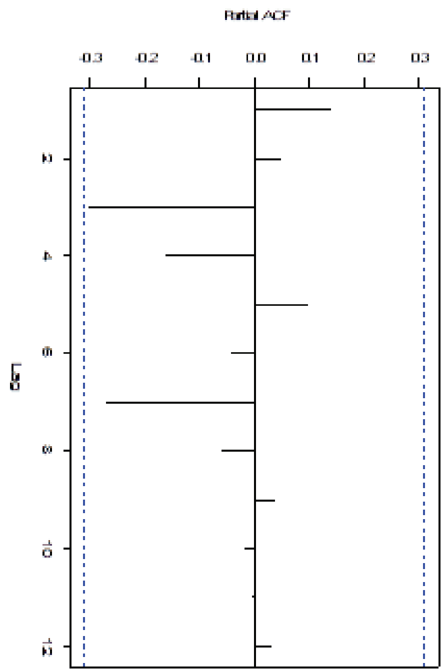


Figure 4



Further, various combinations of ARIMA models were fitted and their AIC and BIC values are presented in Table 2. The model which had minimum AIC and BIC values were chosen as the best fit model. From Table 2, it was observed that the lowest BIC value 556.63 was found in ARIMA (0, 1, 0) model and hence, it was the best fit model for prediction of milk production. Additionally, the auto. arima ( ) function used for data analysis in 'R' software also showed that ARIMA (0, 1, 0) model was the fitted model. Pal *et al.* (2007) and Deshmukh and Paramasivam (2016) found that ARIMA (1,1,1) model was the best fit model for forecasting milk production in India, for the periods 1980-81 to 2004-05 and 1961 to 2012-13, respectively.

Table 2. Goodness of fit statistics of various ARIMA (p, d, q) models

ARIMA (p, d, q)	MSE	RMSE	MAPE	AIC	BIC
ARIMA (1, 0, 0)	9.74	220.39	4.50	567.81	574.67
ARIMA (1, 1, 0)	0.52	226.63	3.91	554.39	559.46
ARIMA (1,1,1)	0.57	226.58	3.90	556.37	563.13
ARIMA (0, 1, 0)	**0.04**	**229.14**	**4.10**	**553.25**	**556.63**
ARIMA (0, 0, 1)	3.78	265.86	5.80	582.71	589.57
ARIMA (1, 0, 1)	8.49	215.98	4.34	568.23	576.80
ARIMA (0, 1, 1)	0.36	227.05	3.93	554.53	559.60

The adequacy of the fit model was verified by the test of significance of residuals tested by Box-Ljung statistic and the estimated parameters are provided in Table 3. As the results indicated non significance of Box-Ljung statistic, it was concluded that the residual was normally distributed. This finding proved that the selected ARIMA (0, 1, 0) model was an appropriate model for forecasting milk production in Tamil Nadu. Based on the best fitted model ARIMA (0, 1, 0), actual and predicted milk production are presented in Table 4 and Figure. According to the figure, the trend lines for actual and predicted milk productions were closer to each other in Tamil Nadu. Hence, by using ARIMA (0, 1, 0) model, the next five years' milk production was forecasted with 95% confidence interval and which is presented in Table 5. As the results indicated, the milk prediction for the years 2019-20, 2020-21, 2021-22, 2022-23 and 2023-24 would be 8529, 8696, 9030 and 9197 thousand tons, respectively.

Table 3. Estimates of the ARIMA model fitted for forecasting milk production in Tamil Nadu

Parameters	Estimates	SE	Log likelihood	Box-Ljung test	
Constant	167.03	36.68	-274.62	16.04	P value= 0.59

Table 4. ACGR, Actual and Predicted Milk Production in Tamil Nadu (in '000' tons)

Year	Actual	Predicted	Residuals
1978-79	1681	1679	1.51
1979-80	1727	1848	-121.03
1980-81	1860	1894	-34.03
1981-82	1738	2027	-289.03
1982-83	1886	1905	-19.03
1983-84	1788	2053	-265.03
1984-85	2562	1955	606.98
1985-86	2846	2729	116.98
1986-87	3118	3013	104.98
1987-88	3295	3285	9.98
ACGR (1978-79 to 1987-88)=8.48%			
1988-89	3109	3462	-353.03
1989-90	3238	3276	-38.03
1990-91	3410	3405	4.98
1991-92	3375	3577	-202.03
1992-93	3357	3542	-185.03
1993-94	3468	3524	-56.03
1994-95	3524	3635	-111.03
1995-96	3694	3691	2.98
1996-97	3791	3861	-70.03
1997-98	3977	3958	18.98
ACGR (1988-89 to 1997-98)=2.38%			
1998-99	4061	4144	-83.03
1999-00	4273	4228	44.98
2000-01	4574	4440	133.98
2001-02	4988	4741	246.98
2002-03	4622	5155	-533.03
2003-04	4752	4789	-37.03
2004-05	4784	4919	-135.03
2005-06	5474	4951	522.98
2006-07	6277	5641	635.98
2007-08	6540	6444	95.98
ACGR (1998-99 to 2007-08)=4.83%			
2008-09	6651	6707	-56.03
2009-10	6787	6818	-31.03
2010-11	6831	6954	-123.03
2011-12	6968	6998	-30.03
2012-13	7005	7135	-130.03
2013-14	7049	7172	-123.03
2014-15	7132	7216	-84.03
2015-16	7244	7299	-55.03
2016-17	7556	7411	144.98
2017-18	7742	7723	18.98
2018-19	8362	7909	452.98
ACGR (2008-09 to 2018-19) =1.90% Overall ACGR(1978-79 to 2018-19)=3.97%			

ACGR - Annual Compound Growth Rate estimated for actual milk production in Tamil Nadu

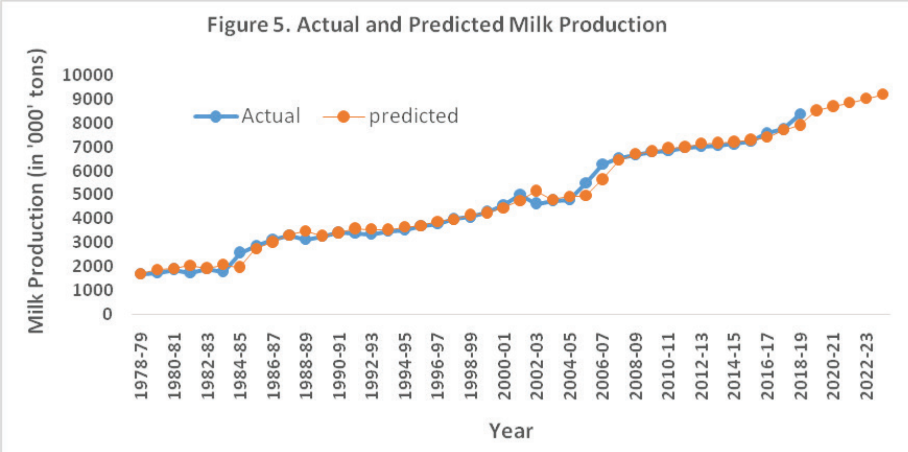


Table 5. Forecasted milk production in Tamil Nadu (in '000' tons)

Year	Forecasted Milk production	95 % LCL	95 % UCL
2019-20	8529	8069	8989
2020-21	8696	8045	9347
2021-22	8863	8066	9661
2022-23	9030	8109	9951
2023-24	9197	8167	10227

CONCLUSION

Based on the goodness of fit statistics and selection criteria that the ARIMA (0, 1, 0) model was found to be the most appropriate model for forecasting milk production in Tamil Nadu. Using this model for forecasting, it could be concluded that milk production in Tamil Nadu would increase from 8529 thousand tons in 2019-20 to 9197 thousand tons in 2023-24. These future projections of milk production in the State would be helpful for policy making, formulation of various schemes for improving milk production, strengthening the infrastructure for the potential export of dairy products and improving rural livelihood in the State.

REFERENCES

Akaike,H. (1974). A new look at the statistical model identication. IEEE Transactions on Automatic Control AC-19, 716-723.

Box, G.E.P and Jenkins, G.M. (1976). Time series analysis: Forecasting and control, Holden-Day, San Francisco. pp. 575.

Deshmukh, S.S and Paramasivam, R. (2016). Forecasting of milk production in India with ARIMA and VAR time series models. ***Asian Journal of Diary and Food Research,*** **35(1):**17-22.

Pal S., Ramasubramanian, V and Mehata, S.C. (2007). Statistical Models for Forecasting Milk Production in India.

Journal of the Indian Society of Agricultural Statistics, 61: 80–83.

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,UsingAutoregressive Integrated Moving Average Models. Advances in Agriculture, Hindawi Publishing Corporation. 2014:1-9.