J. Ind. Soc. Agril. Statist.
61(2), 2007 : 80-83

Statistical Models for Forecasting Milk Production in India

Satya Pal, Ramasubramanian V. and S.C. Mehta
Indian Agricultural Statistics Research Institute, New Delhi
(Received : March, 2007)

ABSTRACT

In this paper, an attempt has been made to forecast milk production using statistical time-series modeling techniques - Double Exponential Smoothing and Auto-Regressive Integrated Moving Average (ARIMA). On validation of the forecasts from these models, ARIMA model performed better than the other one.

Keywords : Exponential smoothing, ARIMA model.

1. INTRODUCTION

Livestock farming has emerged as a commercial enterprise in India. Its growth depends on input feed and output price. Further, forecasting of milk production enables planners and policy makers to estimate the supply requirement of milk in future and formulate appropriate strategies to meet the growing demand. In this paper, an attempt has been made to forecast milk production using statistical time-series modeling techniques - Double Exponential Smoothing and Auto-Regressive Integrated Moving Average (ARIMA).

From the Annual Report "Development of Animal Husbandry, Dairying and Fisheries, Ministry of Agriculture, Govt. of India, New Delhi (2005 – 06)", the data on milk production in India for the period 1980-81 to 1999-2000 was utilized for model fitting and data for subsequent periods from 2000-01 to 2004-05 (Table 1) was used for validation. The analysis was carried out by using SPSS package.

2. METHODOLOGY

I. Selection of Appropriate Smoothing Techniques

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), Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES) and Adaptive Response Rate Simple Exponential Smoothing (ARRSES) which are described below:

(i) Simple exponential smoothing (SES)

For the time series $Y_1, Y_2, \ldots, Y_t$ forecast for the next value Y_{t+1} say F_{t+1}, is based on the weights α and $(1-\alpha)$ to the most recent observation Y_t and the recent forecast F_t respectively, where α is a smoothing constant. The form of the model is :

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

The choice of α has considerable impact on the forecast. The optimum value of α corresponding to minimum Mean Square Error (MSE) is then identified.

(ii) Double exponential smoothing (Holt's)

The form of the model is

$$L_t = \alpha Y_t + (1-\alpha)\ (L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$
$$F_{t+m} = L_t + b_t m$$

where L_t is level of the series at time t

b_t is slope of the series at time t

α and β (= 0.1, 0.2, …, 0.9) are the smoothing and trend parameters.

The pair of values of parameters α and β which gives minimum MSE are taken.

(iii) Triple exponential smoothing (Winter's)

This method is recommended when seasonality exists in the time series data. It is based on three smoothing equations – one for the level, one for trend, and one for seasonality. It is similar to Holt's method but with one additional equation to take care of seasonality. There are two different Winter's methods depending on modeling of seasonality - in an additive or multiplicative way.

(iv) Adaptive response rate simple exponential smoothing method

It has an advantage over above SES method and is used when smoothing constant α changes with the data-pattern. It is useful when quite a large number of items are to be forecasted.

II. Auto-Regressive Integrated Moving Average (ARIMA model)

(i) Model identification

At the outset, the stationarity of the series is examined. In case the data is found to be non-stationary, stationarity is achieved by differencing technique. For instance, the differencing of first order is

$$Z_t = Y_t - Y_{t-1}$$

The next step in the identification process is to find the initial values for the orders of non-seasonal parameters p and q, which are obtained by looking for significant correlations in the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. For this, sample ACFs and PACFs are plotted and compared with their theoretical counterparts available in the literature (for instance, see Pankratz 1984). For identifying the order of AR component, a common practice is to see for significant spikes in the first few lags of the PACF graph and for, MA component, that of ACF graph.

(ii) Estimation

The parameters are estimated by modified least squares technique appropriate to time series data.

(iii) Diagnostic checking

For the adequacy of the model, the residuals are examined from the fitted model and alternative models are considered, if necessary. If the first identified model appears to be inadequate then other ARIMA models are tried until a satisfactory model fits to the data.

The ARIMA model is given by (taking Z_t as the already 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).

Different models are obtained for various combinations of AR and MA individually and collectively (Makridakis *et al.* 1998). The best model is obtained on the basis of minimum value of Akaike Information Criteria (AIC) given by

$$AIC = -2 \log L + 2m$$

where m = p + q and L is the likelihood function.

The performances of different approaches have been evaluated on the basis of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) which are given by

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

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

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

3. RESULTS AND DISCUSSION

(i) Exponential smoothing model:

Table 1 shows the yearly milk production for the period 1980-81 to 2004-05. Time plot (Fig. A) of milk production data revealed that there is increasing trend in the data. For smoothing of the data, Holt's double exponential smoothing technique was found to be most appropriate. Various combinations of α and β both ranging between 0.1 to 0.9 with increments of 0.1 were tried and Mean Square Error (0.4219) was least for $\alpha = 0.9$ and $\beta = 0.2$ (Table 2). The fitted model is given by

$$L_t = 0.9Y_t + 0.1(L_{t-1} + b_{t-1})$$

$$b_t = 0.2(L_t - L_{t-1}) + 0.8b_{t-1}$$

$$F_{t+m} = L_t + b_t m$$

where m = 1, 2, ..., 5 and the initial values for level L_t and trend b_t are 30.37 and 2.46 respectively.

Table 1. Data on milk production in India for the period 1980-81 to 2004-05

Sl. No.	Year	Observed milk production (Million tonnes)
1	1980 – 81	31.6
2	1981 – 82	34.3
3	1982 – 83	35.8
4	1983 – 84	38.8
5	1984 – 85	41.5
6	1985 – 86	44.0
7	1986 – 87	46.1
8	1987 – 88	46.7
9	1988 – 89	48.4
10	1989 – 90	51.4
11	1990 – 91	53.9
12	1991 – 92	55.7
13	1992 – 93	58.0
14	1993 – 94	60.6
15	1994 – 95	64.0
16	1995 – 96	66.2
17	1996 – 97	69.1
18	1997 – 98	72.1
19	1998 – 99	75.4
20	1999 – 2000	78.3
21	2000 – 01	80.6
22	2001 – 02	84.4
23	2002 – 03	86.2
24	2003 – 04	88.1
25	2004 – 05	91.0

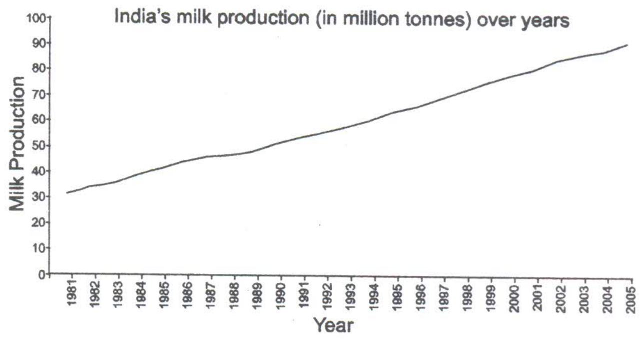

Fig. A

Table 2. Forecasts of milk production in India using double exponential smoothing and ARIMA (1, 1, 1) models

Sl. No.	Year	Observed Milk Produc-tion (Million Tonnes)	Forecast of Milk production	
			Double Exponential Model	ARIMA (1, 1, 1)
1	2000-01	80.6	81.06 (0.57)	80.80 (0.25)
2	2001-02	84.4	83.85 (0.65)	83.26 (1.35)
3	2002-03	86.2	86.64 (0.51)	85.73 (0.55)
4	2003-04	88.1	89.42 (1.50)	88.20 (0.11)
5	2004-05	91.0	92.21 (1.33)	90.66 (0.37)
	MAPE		0.9122	0.5262
	RMSE		0.8848	0.5807

Note : The figures in brackets are the % deviations of forecast from their observed ones

(ii) ARIMA model

The stationary check of the series revealed that it was non-stationary. Merely by using the first differencing technique, it was made stationary (Fig. B) and thus the value of d was 1. The graphs of sample ACFs and PACFs were plotted (Figs. C and D). On matching plots with the theoretical ones of various ARIMA processes, the PACF of AR(1) compared well with the sample PACF as spikes cut off after lag 1. Hence the order of AR component p was taken

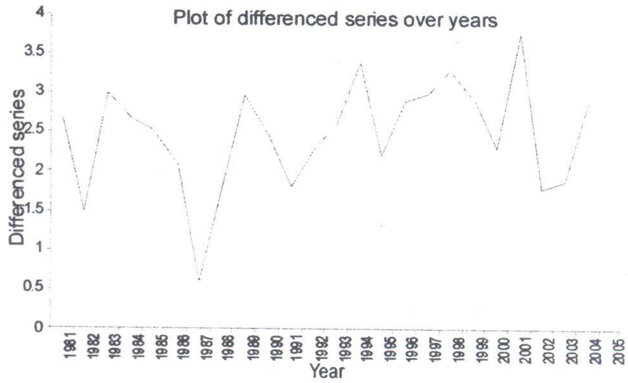

Fig. B

as 1. Also, in order that the proposed model adequately represent the data and at the same time have lesser number of parameters an MA component of order 1 was also added to the model. In addition, using SPSS package for different values of p and q (0, 1 or 2), various ARIMA models were fitted and the

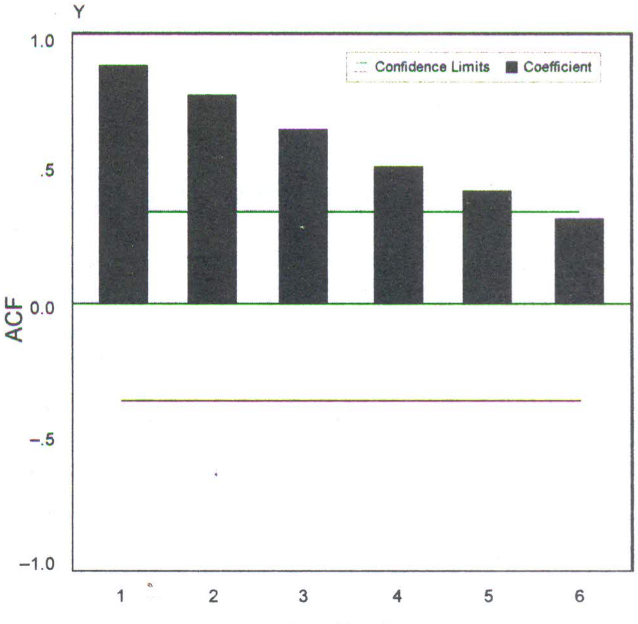

Fig. C

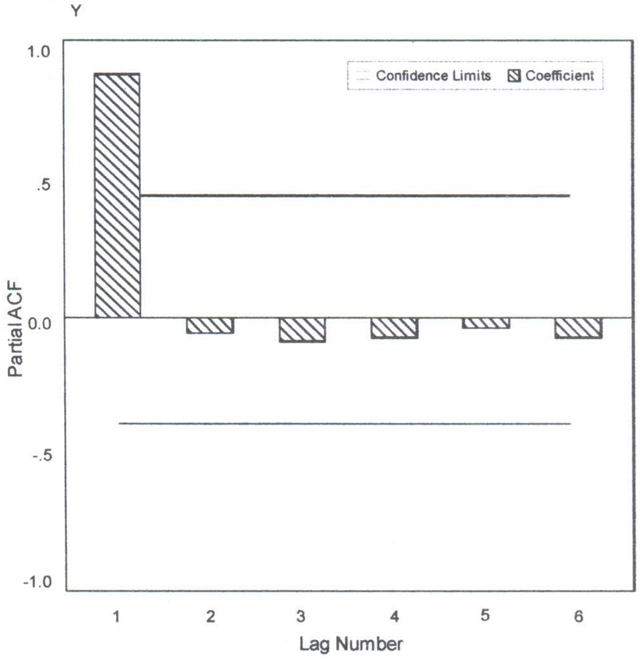

Fig. D

appropriate model was chosen corresponding to minimum value of the selection criterion i.e. Akaike Information Criteria (AIC). In this way, ARIMA (1, 1, 1) model was found to be the best model (Table 2). The fitted model is given by

$$Y_t = 2.47 - 0.45\varepsilon_t - 0.18Y_{t-1}$$

For this model the MSE came out to be 0.4219 which is less than that of fitted Exponential model. Performance evaluation measures viz. MAPE and RMSE were computed for the forecasted milk production for the years 2001 to 2005 (Table 2). Comparison of the results revealed that among the models fitted, ARIMA(1, 1, 1) model came out to be performing better when the forecasts were validated.

REFERENCES

Gardner, E.S. (1985). Exponential smoothing – The state of the art. *J. Forecasting,* **4**, 1-28.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998). *Forecasting – Methods and Applications.* John Wiley and Sons, New York.

Pankratz, A. (1983). *Forecasting with Univariate Box - Jenkins Models: Concepts and Cases.* John Wiley and Sons, New York.