PROSPECTS OF LIVESTOCK AND DAIRY PRODUCTION IN INDIA UNDER TIME SERIES FRAMEWORK

R.K.Paul, Wasi Alam and A.K.Paul

Indian Agricultural Statistics Research Institute, New Delhi -110012.

ABSTRACT

In spite of continuous decline in contribution of agriculture in total gross domestic product (GDP), share of livestock sector especially dairy production in total GDP has shown a continuous rise trend over the last 30 years. Autoregressive integrated moving average (ARIMA) approach has been applied for modeling and forecasting of milk production of India. Autocorrelation (AC) and partial autocorrelation (PAC) functions were estimated, which led to the identification and construction of ARIMA models, suitable in explaining the time series and forecasting the future production. A significant increasing linear trend in the total milk production in India has been found. To this end, evaluation of forecasting is carried out with mean absolute prediction error (MAPE), relative mean absolute prediction error (RMAPE) and root mean square error (RMSE). The best identified model for the data under consideration was used for out-of-sample forecasting up to the year 2015.

Keywords: ARIMA, Forecasting, Milk production, Stationarity, Trend

INTRODUCTION

Livestock sector plays a critical role in the welfare of India's rural population. Livestock provides stability to family income especially in the arid and semi-arid regions of the country. This sector is emerging as an important growth leverage of the Indian economy. As a component of agricultural sector, its share in gross domestic product has been rising gradually, while that of crop sector has been on the decline. In recent years, livestock output has grown at a rate of about five percent a year, higher than the growth in agricultural sector. Livestock are the best insurance against the vagaries of nature due to drought, famine and other natural calamities. Major part of the livestock population is concentrated in the marginal and small size of holdings. Livestock plays an important and vital role in providing nutritive food to families both in rural and urban areas. Hence, from the equity and livelihood perspective it is considered as an important component in poverty alleviation programmes. Agriculture was always considered to be the backbone of our society. This is why the monsoons have always been watched so closely. After all no rain means bad agriculture which translates to poor economic data. Therefore it would be interesting to note that the contribution of agriculture to total GDP has been declining over the years. It formed nearly 19% of total GDP way back in 2004-2005 and as per the latest data the contribution has declined to just 14% in 2012. The main reason behind this is the healthier growth in other sectors that contribute to GDP. Despite the capital invested in agriculture, it has grown by a mere 3.3% annually in the 11th 5-year plan. Though this is better than the 2.4% growth seen in the previous 5-year plan but is nowhere near the growth rates seen in other sectors.

India is the largest producer of milk. Livestock provides a flow of' essential food products, draught power, manure, employment, income, and export earnings. With an annual production of 74 million tonnes in 1998-99, initiation of Operation Flood in early seventies provided a stimulus to milk production, which has never looked back. The growth is on account of both improvements in productivity and shift in priorities towards buffalo and crossbred cattle. Yet, the productivity is low compared to the potential and world average. Meat output grew tremendously since eighties. This has resulted exclusively due to increase in number of animals slaughtered. Productivity of most meat producing species is low and shows no sign of growth. In short, excluding milk, other products are lagging behind on the productivity front.

Time-series analysis of milk production has been an important tool for livestock management and decision making as it reveals hidden trends and seasonality patterns. In livestock management, as well as in other cases, forecasting strategies are based on the development of either descriptive or explanatory models. In addition to the forecasting character, the multivariate descriptive models have the advantage that by “stepwise modelling”—namely by adding stepwise predictors and comparing the quality of fit, certain inferences concerning the importance of the predictors can be made. Descriptive models used to predict and analyze time series data attempt to decompose the dependent variable into four main components. Simple time trends, periodic fluctuations, predictors’ effect and the error component. A common realisation of this approach is the development of the multivariate ARIMA models (Box et al, 2007). Prajneshu and venugopalan (1997) applied this model as well as other parametric statistical modeling techniques, like polynomial function fitting, and nonlinear mechanistic growth modeling for describing trends in marine fish production data of the country. Singh et al (2007) applied statistical models for forecasting milk production in India. One advantage of the ARIMA approach is that it is able to provide a good understanding of the system. This model has been dominating time series analysis for several decades. In the present work, ARIMA model was used for modelling and forecasting of milk production of India and trend of production over the last three decades has been studied.

MATERIALS AND METHODS

Data description

Data on growth rate of agriculture and allied sectors as well as livestock sector from the year 1994 to 2004 and time series data of India's milk production (in million tonnes) from 1979 to 2011 are taken from the report of the working group on animal husbandry and dairying for the eleventh five year plan (2007-20012) and partly from Department of Animal Husbandry, Dairying and Fisheries, Ministry of Agriculture (http://www.nddb.org/English/Statistics/Pages/Milk-Production.aspx). Milk production data in million tonnes from the year 1979 to 2007 has been used for model development and data from 2008 to 2011 have been used for model validation purpose. The SAS 9.3 statistical software package has been used for data analysis.

Autoregressive Integrated Moving Average (ARIMA) Model

A generalization of ARMA models which incorporates a wide class of non- stationary time-series is obtained by introducing the differencing into the model. ARIMA econometric modeling takes into account historical data and decomposes it into an autoregressive (AR) process where there is a memory of past events and integrated (I) process which accounts for stabilizing or making the data stationary, making it easier to forecast, and a moving average (MA) of forecast errors, such that the longer the historical data, the more accurate forecast will be as it learns from over time. The simplest example of a non-stationary process which reduces to a stationary one after differencing is Random Walk. A process $\{ y_t \}$ is said to follow an

Integrated ARMA model, denoted by ARIMA (*p*, *d*, *q*), if $\nabla^d y_t = (1-B)^d \varepsilon_t$ is ARMA (*p*, *q*). The model is written as

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad (1)$$

where $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p$

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$

$\varepsilon_t \sim WN(0, \sigma^2)$, *WN* indicating White Noise.

B is the backshift operator such that $By_t = y_{t-1}$,

The integration parameter *d* is a nonnegative integer. When *d* = 0, ARIMA (*p*, *d*, *q*) model reduces to ARMA (*p*, *q*) model.

The ARIMA methodology is carried out in three stages, viz. identification, estimation and diagnostic checking. Parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage and adequacy of tentatively selected model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration. An excellent discussion of various aspects of this approach is given in Box *et al*. (2007). Most of the standard software packages, like SAS, SPSS and EViews contain programs for fitting of ARIMA models.

Estimation of Parameters

Estimation of parameters for ARIMA model is generally done through Nonlinear least squares method. Fortunately, several software packages are available for fitting of ARIMA models. To this end, in this paper, SAS software package is used. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for ARIMA model are computed by:

$$AIC = T' \log(\sigma^2) + 2(p+q+1) \quad (2)$$

and

$$BIC = T' \log(\sigma^2) + (p+q+1)\log T' \qquad (3)$$

where T' denotes the number of observations used for estimation of parameters and σ^2 denotes the Mean square error.

RESULTS AND DISCUSSION

Trend estimation

Assuming presence of deterministic linear trend in the time series, following model is fitted:

$$Y_t = \mu + \delta t + \varepsilon_t,\ t = 1, 2, \ldots, T \qquad (4)$$

where Y_t is the total milk production in million tones in the t^{th} year, μ is the general mean, δ is the coefficient of trend ε_t's are uncorrelated with zero mean and constant variance σ_ε^2. Let

$$\hat{e}_t = Y_t - \hat{\mu} - \hat{\delta} t$$

The fitted trend equation is obtained as:

$$Y_t = 21.449 + 2.872\,t$$
$$(1.553) \quad (0.079)$$

where the values within brackets ( ) denote corresponding standard errors of estimates. The trend is found to be significant at 1% level of significance.

It is evident from Fig.1 that the growth rate (percent share in GDP on the basis of 1993-94 prices) in agriculture sector over the years has been steady and fluctuating significantly depending upon the monsoon and other climatic factors. Of late there has been deceleration of agricultural growth. On the contrary, livestock sector has shown a steady growth and thus providing stability to the overall family income. Fig.2 depicts % share of agriculture and livestock to total GDP and also share of livestock to agriculture.

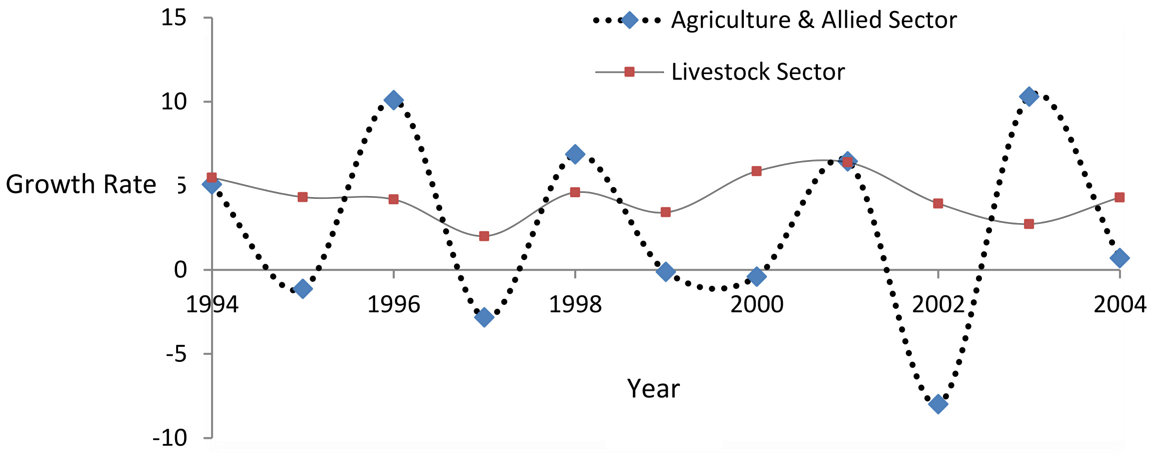

Fig. 1: Growth rate in agriculture sector and livestock sector over the year

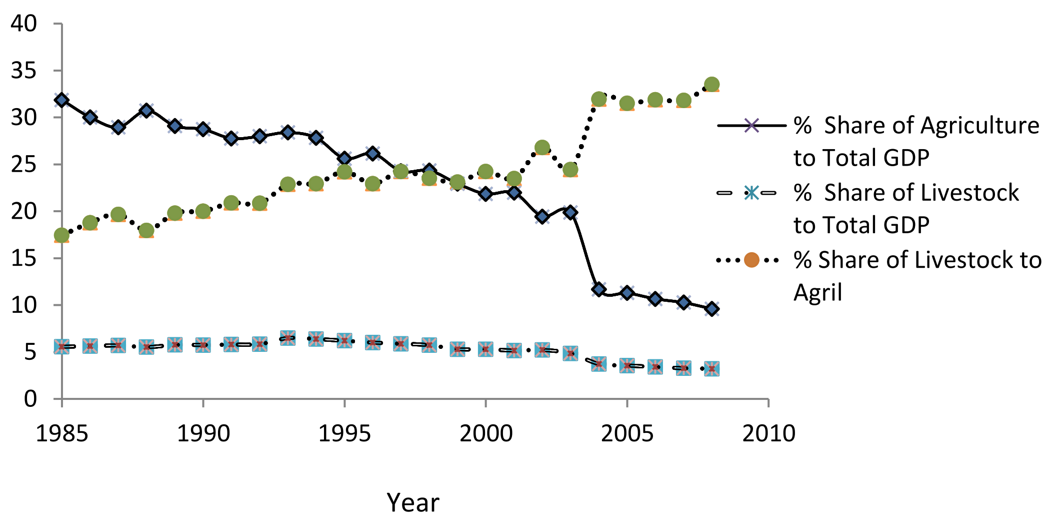

Fig. 2: % share of agriculture and livestock to total GDP and % share of livestock to agriculture

Fitting of ARIMA model

The detrended residual series $\hat{e}_t$ is found to be nonstationary. In order to attain stationarity, differencing is done. From the estimated autocorrelation function (acf), reported in Table 1, and figure 2a to 2d, it is found that it decays very slowly thereby requires to be differenced so that the resulting series depicts a pattern for a possible ARMA modeling. In order to select the order of the ARIMA model, unit root test proposed by Dickey and Fuller (1979) is applied for parameter ρ in the auxiliary regression

$$\Delta y_t = \rho\, y_{t-1} + \alpha_1\, \Delta y_{t-1} + \varepsilon_t$$

where $\Delta y_t = y_t - y_{t-1}$. The relevant null hypothesis is H_0: ρ = 0 and the alternative is H_1: ρ < 0. For the given data, the estimate of ρ is computed as -0.64 with calculated *t*-statistic as -4.62, which is less than the critical value of *t* at 5% level of significance, i.e. -1.95 (Franses, 1998, Page 82). Therefore, H_0 is not rejected at 5% level and so ρ = 0. Thus, there is presence of one unit root and so differencing is required. Usually, differencing is applied until the acf shows an interpretable pattern with only a few significant autocorrelations. On taking the first difference of the original series, it is seen that only a few acfs, reported in Table 1, are high making it easier to select the order of the model. On taking the second differencing of the original series it is seen that the sum of the autocorrelations of double differenced series is -0.492 which concludes that the series is over differenced (Franses, 1998).

Table 1 Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the milk production data

Lag	Actual series		First differenced series	
Lag	ACF	PACF	ACF	PACF
1	.897	.897	.619	.619
2	.797	-.045	.441	.094
3	.701	-.030	.383	.127
4	.606	-.052	.339	.067
5	.515	-.038	.328	.092
6	.430	-.027	.207	-.119
7	.353	-.020	.122	-.054
8	.281	-.029	.008	-.155
9	.213	-.040	.021	.067
10	.143	-.061	.051	.050
11	.074	-.055	.017	.001
12	.011	-.036	.034	.072
13	-.051	-.058	-.015	-.049
14	-.110	-.048	-.111	-.179
15	-.164	-.042	-.044	.098
16	-.212	-.039	-.096	-.153

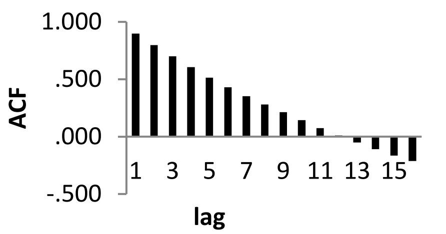

Fig 2a ACF of original series

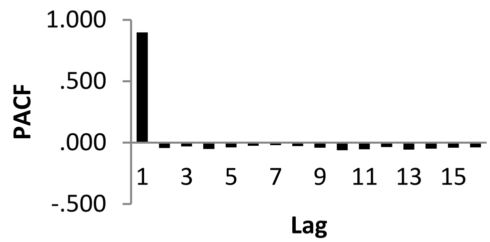

Fig 2b PACF of original series

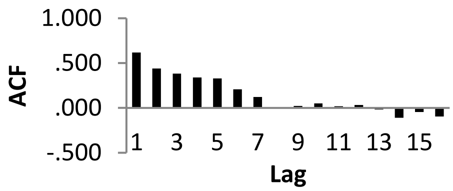

Fig 2c ACF of 1st differenced series

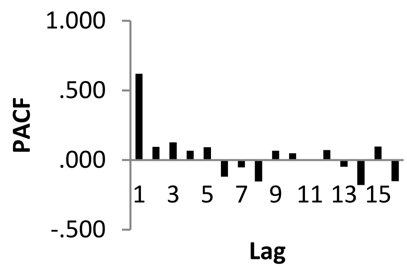

Fig 2d PACF of 1st differenced series

Therefore after differencing once, the resultant series becomes stationary. On examining its autocorrelation functions (acf) and partial autocorrelation functions (pacf) and on the basis of minimum AIC and BIC values, the ARIMA(*1*,*1*,*0*) model is selected. Parameter estimates along with corresponding standard errors of fitted ARIMA(*1*,*1*,*0*) model are reported in Table 2.

Table 2. Parameter estimates along with standard errors (S.E.)

Parameters	Estimate	S.E.	*t*-value	Significance
Constant	*2.817*	*0.458*	*6.156*	*<0.001*
AR1	*0.612*	*0.178*	*3.433*	*0.002*

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 16 lags were computed and the same along with their significance tested by Box-Ljung statistic are provided in Table 3. 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 India. The ACF and PACF of the residuals are given in Figure 4, which also indicated the 'good fit' of the model.

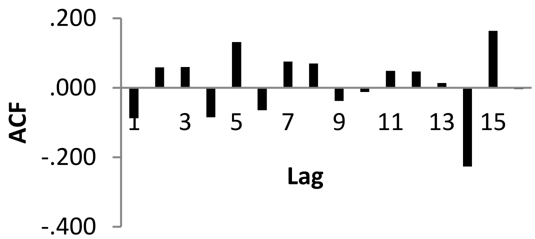

Fig 4a ACF

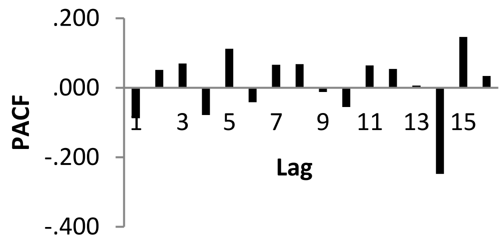

Fig 4b PACF

Fig. 4: ACF and PACF of residual series

Table 3 ACF and PACF of the residuals of fitted ARIMA model

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	P-value
1	-.087	.179	.238	1	.626
2	.059	.176	.350	2	.839
3	.060	.173	.471	3	.925
4	-.085	.169	.724	4	.948
5	.131	.165	1.354	5	.929
6	-.065	.162	1.515	6	.959
7	.076	.158	1.744	7	.973
8	.070	.154	1.952	8	.982
9	-.038	.150	2.014	9	.991
10	-.012	.146	2.021	10	.996
11	.049	.142	2.138	11	.998
12	.047	.138	2.254	12	.999
13	.014	.134	2.265	13	1.000
14	-.227	.129	5.345	14	.980
15	.164	.124	7.088	15	.955
16	-.004	.120	7.089	16	.972

Validation

One-step ahead forecasts of milk production along with their corresponding upper confidence interval and lower confidence interval for the year, 2008 to 2011 in respect of above fitted model are reported in Table 4.

Table 4 Validation of the model for forecasting milk production in India (in million tonnes)

Years	Actual	Forecasts by ARIMA(*1,1,0*)	Lower Confidence Limit	Upper Confidence Limit
2008	112.2	112.24	110.25	114.23
2009	116.4	115.99	112.21	119.76
2010	121.8	119.37	113.90	124.84
2011	127.3	122.54	115.51	129.57

For measuring the accuracy in fitted time series model, Mean absolute error (MAE), Mean absolute percentage error (MAPE) and Relative mean absolute prediction error (RMAPE) are computed by using the formulae given in eqs. 5, 6 and 7. The MAPE, RMAPE and RMSE values for fitted ARIMA(1,1,0) model are respectively computed as 1.9, 1.5% and 2.68.

$$\text{MAPE} = 1/4\sum_{i=1}^{4}\left|y_{t+i} - \hat{y}_{t+i}\right| \tag{5}$$

$$\text{RMAPE} = 1/4\sum_{i=1}^{4}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\}\times 100 \tag{6}$$

$$\text{RMSE} = \sqrt{1/4\sum_{i=1}^{4}\left(y_{t+i} - \hat{y}_{t+i}\right)^2} \tag{7}$$

Forecasting

The best model i.e. ARIMA (1,1,0) model as given in eq. 8, was used for forecasting of total milk production in India for the period 2012-2015 and the same along with the forecast error variance is reported in table 5.

$$\Delta Y_t = \alpha_0 + \alpha_1 \Delta Y_{t-1} + \varepsilon_t \tag{8}$$

i.e. $$Y_t = \alpha_0 + (1+\alpha_1) Y_{t-1} - \alpha_1 Y_{t-2} + \varepsilon_t \quad (9)$$

where $\Delta Y_t = Y_t - Y_{t-1}$

So the forecast for the year 2012 to 2015 can be computed by eq. 9.

Table 5 Forecasts of milk production (in million tonnes)

Years	Forecast (Million Tonnes)	Lower Confidence Limit	Upper Confidence Limit
2012	132.13	130.17	134.10
2013	136.48	132.57	140.40
2014	140.48	134.58	146.39
2015	144.23	136.39	152.07

CONCLUSION

It has been found that there is a significantly increasing trend in the total milk production in India. ARIMA (1,1,0) model quite satisfactorily captured the variation present in the data set. The model demonstrated a good performance in terms of explained variability and predicting power. The findings of the present study provided direct support for the potential use of accurate forecasts in decision making and livestock management in India.

REFERENCES

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2007). *Time-Series Analysis: Forecasting and Control*. 3rd edition. Pearson education, India.

Dickey D A and Fuller W A. 1979. Distribution of the estimators for the autoregressive time series with a unit root. *J. Amer. Stat. Assoc.* **74**: 427-31.

Franses, P. H. (1998). *Time series models for business and economic forecasting*. Cambridge university press, U. K.

Prajneshu and Venugopalan, R. (1997). Statistical modeling for describing fish production in India. *Ind. J. Anim. Sci*. **67**, 452-456.

Singh S., Ramasubramanian,V. and Mehta, S.C. (2007). Statistical Models for forecasting milk production of India. *J. Ind. Soc. Agril. Statist.* **61**, 2007: 80-83.