Forecasting cattle and buffalo population in India – A time series analysis

Arya S Nair, M Thirunavukkarasu, A Serma Saravana Pandian, G Senthilkumar and C Balan

Abstract: Cattle and buffalo are known to be the symbol of wealth in rural India from time immemorial. To ensure the attainable socio-economic benefits from cattle and buffalo reaching the rural poor through effective strategies, a more complete extrapolation of cattle and buffalo population in the country for the future using forecasting tools was attempted in this study. Data on cattle and buffalo population from 1950-51 to 2016-17 were collected from various reports of BAHS and FAO. Various time series forecasting models were employed to identify the growth patterns and to predict the future trends in bovine population. The forecasting models used were compared to identify the best fit model. From the results of the study, it could be discerned that the Damped Trend Exponential Smoothing was found to be the best fit model for cattle population and forecasted value indicated that the cattle population in the country would be almost stagnant in the next three decades. The cattle population forecast showed a slightly decreasing trend from 2010-11 with 194.184 million to 187.661, 188.177, 188.191, and 188.192 millions in 2020-21, 2030-31, 2040-41 and 2050-51, respectively. However, in case of buffaloes, the Brown Exponential Smoothing model was found to be the best fit model and the buffalo population was predicted to increase from 2000-01 and the predicted populations were 116.663, 127.787, 138.910 and 148.921 million in 2020-21, 2030-31, 2040-41 and 2050-51 respectively.

Keywords: Cattle and buffalo population, Forecasting, Time series analysis

Introduction

Cattle and buffalo, providing food products like milk, meat and hide, have been generating productive employment and valuable income (Govt. of India, 2015-16 and Govt. of India, 2019), to the vast majority of rural households, majority of whom are small and marginal farmers, landless labourers and women. Total milk production in the country was 17 million tonnes in the year 1950-51 and since 1970s, milk production continued to rise, taking the country as the largest producer of milk in the world now by producing 13.1 per cent of world’s milk, with milk production reaching 176.35 million tonnes in 2017-18 and achieving self-sufficiency in milk production (Govt. of India, 2018-19). This impressive growth could be attributed to the concerted efforts of large number of small dairy farmers, milk cooperatives, and planners who made possible crossbreeding of local low producing bovines with exotic germplasm and high producing buffaloes.

To be able to effectively plan strategies for optimizing milk production and for augmenting rural livelihood in the country, prediction of future bovine population using forecasting tools is required. Hence, this study attempted to untangle the future of Indian bovine population, considering the past and present trends.

Materials and Methods

Data on bovine population from 1950-51 to 2016-17 were collected from the reports of FAO Statistics (www.fao.org) and the reports of Basic Animal Husbandry Statistics (of different years), Dept. of Animal Husbandry and Dairying, Ministry of Agriculture and Farmers’ Welfare, Govt. of India, both of various years. Different forecasting models were employed to predict the future bovine population. The results of forecasting models fitted were also compared to identify the best fit model. Among various forecasting models, Auto Regressive Integrated Moving Average [ARIMA] - p, d, q and Exponential Smoothing [ES] models were fitted.

Various combinations of ARIMA models viz., ARIMA (1,1,1), (1,1,0), (0,1,1), (0,1,0), (0,1,2), (1,1,2), (2,1,0), (2,1,1), (2,1,2), (1,2,1),...
(0,2,1), (1,2,0), (0,2,0), (0,2,2), (1,2,2), (2,2,0), (2,2,1), and (2,2,2) and ES models viz., Simple ES, Holt ES, Brown ES and Damped trend ES were tried. Choudhury and James (2014) used simple exponential smoothing, double exponential smoothing, damped-trend linear exponential smoothing and ARMA models for predicting crop yields.

Then, using the best model, two kinds of forecasts were performed: sample period forecasts and post-sample period forecasts. The former was used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. Forecasting accuracy of different models was identified by using the measures of indices like Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Ahmad and Ahmad (2013) also compared the ARIMA model and Exponential Smoothing Method in making a prediction, by examining the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

**ARIMA model**

The data used in the study were non-stationary and non-seasonal. ARIMA model is a combination of an Auto Regressive (AR) process and a Moving Average (MA) process applied to a non-stationary data series. A combined model that contains p (AR) term and q (MA) term is called ARMA (p,q). If the object series is differenced ‘d’ times to achieve stationary, the model is classified as ARIMA (p, d, q) model as discussed by Box et al. (2007). The basic criteria for choosing the best fit of ARIMA (p, d, q) model are given in Table 1.

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

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

Where,

- \( Y = \) Value at \( t \)th year;
- \( \varepsilon_t 's = \) Error terms which are independently and normally distributed with mean zero and constant variance \( \sigma^2 \) for \( t = 1, 2, \ldots, n; \)
- \( \mu = \) Constant and
- \( \phi_i \) and \( \theta_j = \) Coefficients to be estimated.

**Simple Exponential Smoothing**

The Simple Exponential Smoothing (SES) model is a time series forecasting technique that can be defined using an additive model used to analyze data which have no trend and seasonal pattern. This is a method of estimation of forecasts of single weight or parameter. Greater weights are assigned to recent observation and smaller weights to distant observation (Sharpe et al. 2010). The model as given below:

\[ F_{t+1} = F_t + \alpha (y_t - F_t) \]

New forecast value at time \( t+1 = \) Old forecast at time \( t + \alpha (\text{Error in the last forecast}) \)

The smoothing constant \( \alpha \) value is selected based on error minimization approach (Talwar and Goyal, 2019).

**Brown’s Linear (Double) Exponential Smoothing model**

The double exponential smoothing model is used to model time series data which have trend, but not seasonality (Brown, 1963). Here, \( F' \) denotes a simple smoothed value and \( F'' \) denotes a double smoothed value:

\[ F'_t = \alpha Y_t + (1 - \alpha) F'_{t-1} \]

\[ F''_t = \alpha Y_t + (1 - \alpha) F''_{t-1} \]

\[ \alpha_t = F'_t + (F'_t - F''_t) = 2F''_t - F'_t \]

\( \alpha \) is estimated smoothed level at time \( t \)

\[ b_t = \frac{\alpha}{1 - \alpha} (F'_t - F''_t) \]

\( b_t \) shows the estimated trends at the end of time period \( t \), for \( m \) period ahead forecast is

\[ F_{t+m} = \alpha_t + m b_t \]

**Holt’s Linear (Double) Exponential Smoothing model**

Holt’s method can be implemented for the time series data demonstrating a trend (Hanke and Wichern, 2008). This method is appropriate for non stationary data and to make short term forecast. In this technique, level and trend components are smoothed separately using different parameters \( \alpha \) and \( \beta \). Holt’s double exponential smoothing method uses three equations one each for level, trend and forecast.

\[ L_t = \alpha Y_t + (1 - \alpha) (L_{t-1} + b_{t-1}) \]

\[ b_t = \beta (L_t - L_{t-1}) + (1 - \beta) b_{t-1} \]

\[ F_{t+m} = L_t + b_t m \]

where,

- \( L_t \) = Level of time series at period \( t \)
- \( b_t \) = trend (slope) estimate of time series at time period \( t \)
- \( F_{t+m} \) = forecast at \( m \) period ahead of time \( t \)

\( \alpha \) and \( \beta \) are smoothing constants for level and trend with their values lying between 0 and 1.
Damped Trend Exponential Smoothing method

The forecasts generated by Holt’s linear method display a constant trend (increasing or decreasing) indefinitely into the future. Since empirical evidence indicated that this method tended to over-forecast, especially for longer forecast horizons, Gardner and McKenzie (1985) introduced a parameter that dampens the trend to a flat line sometime into the future.

The smoothing equations are,

\[ L_t = \alpha Y_t + (1-\alpha) (L_{t-1} + \varphi T_{t-1}) \]

\[ T_t = \gamma (L_t - L_{t-1}) + (1-\gamma) \varphi T_{t-1} \]

The m-step-ahead prediction equation is

\[ \hat{y}_{t+m} = L_t + \sum_{i=1}^{m} \varphi^i T_i \]

This is the forecast \( y \), m-steps ahead by taking the last available estimated level state and multiplying the last available trend (slope) \( T_i \), with \( \delta_i \) = damping factor.

Results and Discussion

Among various models fitted, the model with the lowest normalized Bayesian Information Criterion (BIC) value and better model fit statistics like higher \( R^2 \) and the lowest Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) was selected as the best fit model.

Forecasting cattle population

The criteria adopted for model selection for forecasting of cattle population are given in Table 2, like the values of \( R^2 \), RMSE, MAPE, MAE and Normalized BIC. The BIC value directly compares the information loss of models. Hence, a lower BIC value suggests a better model or best fit model. It can be discerned that the Damped Trend Exponential Smoothing was found to be the best fit model, since its BIC value (28.209) was the lowest among all the models, along with reasonably lower values of RMSE (1247937.145), MAE (539261.451) and MAPE (0.443) and the highest \( R^2 \) value (0.984).

After this model selection, the model parameters were estimated and the results of the estimates are given in Table 3. Based on the best fit Damped Trend Exponential Smoothing model, forecasting of cattle population was carried out at two stages viz., sample period forecasts for the period from 2000-01 to 2016-17 (to develop confidence in model) and post sample period forecasts for 2020-21, 2030-31, 2040-41 and 2050-51 and the results are displayed in Table 4.

As could be seen, the Indian cattle population forecast shows that the cattle population had been slightly decreasing from 194.184 million in 2010-11 to 187.661, 188.177, 188.191, and 188.192 million numbers in 2020-21, 2030-31, 2040-41 and 2050-51, respectively. The results clearly indicate that the cattle population in the country would be almost stagnant in the next three decades. This underlines the fact that concerted programmes are required to be framed and implemented to ensure higher productivity to continue to satisfy the increasing demand for milk.

Forecasting buffalo population

The criteria adopted for model selection for forecasting of buffalo population are given in Table 5, like the values of \( R^2 \), RMSE, MAPE, MAE and Normalized BIC. From the table, it can be found out that the Brown Exponential Smoothing model was found to be the best fit model, since the BIC value was found to be the lowest (26.606), with the high \( R^2 \) value of 0.996. RMSE (577912.272), MAPE (0.445) and MAE (336752.723) were also lower in this model. The model parameter was estimated and the results of the estimate are given in Table 6.

Based on the best fit Brown Exponential Smoothing model, forecasting of buffalo population was carried out at two stages

### Table 1 Criteria for choosing the appropriate forecasting model

| Selection Criterion | Notation |
|---------------------|----------|
| Bayesian Information Criterion | \( n \log(MSE) + K \log n \) |
| Coefficient of Determination | \( 1 - \frac{\text{Error sum of square}}{\text{Total sum of squares}} \) |
| Root Mean Square Error | \( \sqrt{\frac{1}{n-k} \sum \tilde{e}_t^2} \) |
| Mean Absolute Error | \( \frac{1}{n} \sum_{t=1}^{n} |\tilde{e}_t| \) |
| Mean Absolute Percent Error | \( \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\tilde{e}_t}{y_t} \right) \times 100 \) |

Where \( k = \) Number of parameters in the statistical model; \( n = \) Sample size; \( y_t = \) Observed value; and \( \tilde{e}_t = \) Difference between the observed and estimated values.
viz., sample period forecasts for the period from 2000-01 to 2016-17 (to develop confidence in model) and post sample period forecasts for 2020-21, 2030-31, 2040-41 and 2050-51. Results are displayed in Table 7. From the forecasted values, it is evident that buffalo population is increasing linearly. The actual value of buffalo population during 2000-01 was found to be 93.831 million against the predicted value of 93.762 million.

The buffalo population was predicted to be increasing from 2000-01 and the predicted populations were 116.663, 127.787, 138.910 and 148.921 million in 2020-21, 2030-31, 2040-41 and 2050-51 respectively. The increasing trend of buffalo population in the past, present and future could be attributed to the higher fat content of its milk which fetches higher price. This inherent characteristic of buffalo must have adequately persuaded the farmers to rear more and more buffaloes over time.

While Prasad et al. (2004) found the average annual growth rates of cattle and buffaloes in India as 0.67 and 1.63 per cent, respectively during the period 1951-92, Prabu et al. (2012) found a positive growth of cattle during 1997 and 2003 census periods also in the country. Borah and Halim (2014) too found a positive average annual growth rate of cattle (1.83 per cent) in India.

Table 2 Criteria for model selection for cattle population

| Model       | R²  | RMSE          | MAPE | MAE  | BIC   |
|-------------|-----|---------------|------|------|-------|
| ARIMA (1,1,1) | 0.982 | 1279634.152   | 0.449 | 848577.423 | 28.343 |
| ARIMA (1,1,0) | 0.982 | 1356200.215   | 0.529 | 1003762.456 | 28.459 |
| ARIMA (0,1,1) | 0.982 | 1356200.231   | 0.529 | 1003762.478 | 28.459 |
| ARIMA (0,1,0) | 0.968 | 1689753.124   | 0.721 | 1372901.548 | 28.826 |
| ARIMA (0,1,2) | 0.981 | 1333415.456   | 0.508 | 964140.245  | 28.498 |
| ARIMA (1,1,2) | 0.982 | 1286037.456   | 0.452 | 855883.968  | 28.426 |
| ARIMA (2,1,1) | 0.983 | 1258719.012   | 0.464 | 876411.743  | 28.456 |
| ARIMA (2,1,2) | 0.984 | 1271794.014   | 0.457 | 862154.756  | 28.549 |
| ARIMA (0,2,1) | 0.979 | 1367974.365   | 0.347 | 654400.431  | 28.479 |
| ARIMA (1,2,0) | 0.979 | 1368205.348   | 0.346 | 652181.654  | 28.480 |
| ARIMA (1,2,1) | 0.979 | 1356819.489   | 0.339 | 639172.698  | 28.389 |
| ARIMA (1,2,2) | 0.979 | 1376650.564   | 0.372 | 703860.124  | 28.566 |
| ARIMA (2,2,0) | 0.980 | 1382529.154   | 0.374 | 707267.423  | 28.648 |
| ARIMA (2,2,1) | 0.980 | 1380997.324   | 0.379 | 715601.123  | 28.646 |
| ARIMA (2,2,2) | 0.982 | 1325554.781   | 0.427 | 808053.465  | 28.638 |
| Simple ES    | 0.966 | 1751810.123   | 0.756 | 1444439.146 | 28.824 |
| Holt ES      | 0.981 | 1331488.154   | 0.340 | 641808.135  | 28.347 |
| Brown ES     | 0.981 | 1318859.215   | 0.335 | 631730.131  | 28.256 |
| Damped trend ES | 0.984 | 1247937.145 | 0.443 | 539261.451 | 28.209 |

Table 3. Estimates of the best fit Damped Trend Exponential Smoothing model for cattle population

| Model         | Parameters          | Estimate | SE      | t       | Sig. |
|---------------|---------------------|----------|---------|---------|------|
| Damped trend ES | Alpha (Level)       | 1.000    | 0.292   | 3.420   | 0.001|
|                | Gamma (Trend)       | 1.000    | 0.946   | 1.050   | 0.295|
|                | Phi (Trend damping factor) | 0.701 | 0.199   | 3.510   | 0.001|

Table 4 Forecasts of cattle population (in million)

| Year   | Actual | Predicted | LCL(95%) | UCL(95%) | Residual |
|--------|--------|-----------|----------|----------|----------|
| 2000-01| 191.924| 192.591   | 190.088  | 195.095  | -0.667   |
| 2010-11| 194.184| 194.673   | 192.170  | 197.176  | -0.488   |
| 2020-21| -      | 187.661   | 177.929  | 197.394  | -        |
| 2030-31| -      | 188.177   | 161.331  | 215.023  | -        |
| 2040-41| -      | 188.192   | 150.548  | 225.836  | -        |
| 2050-51| -      | 188.192   | 142.963  | 233.422  | -        |
Indian J Dairy Sci 73(3): 268-273

Thirunavukkarasu and Rajarathinam (2014), while forecasting milled rice production in India using data from 1960-61 to 2013-14, found that the most appropriate models were the ARIMA, Brown’s model and Damped model. Chaudhari and Tingre (2015), while forecasting egg production in India, based on data from 1979-80 to 2010-11, found that the model which had minimum normalized BIC value (i.e., ARIMA (0,1,0)) was found to be the best model for predicting Indian egg production. Celik (2016) used Holt, Brown and Damped Trend exponential smoothing methods for forecasting production of cereals in Turkey based on the data from 1965 to 2015. Celik and Sengul (2016), while predicting the number of poultry in Turkey from 2016 to 2025 found that Damped exponential smoothing method was the best model for predicting the number of turkeys.

Conclusions
Cattle population would be almost stagnant in the next three decades. This underlines the fact that concerted programmes are required to be framed and implemented to ensure higher productivity to continue to satisfy the increasing demand for milk. However, the buffalo population would increase linearly in the future, due to the much preferred higher fat content of its milk which fetches higher price, adequately persuading the farmers to rear more and more buffaloes. Provision of adequate quality inputs like feed, fodder and health cover needs to be continuously provided to further augment the productivity among cows and to satisfy the increasing demand from the increasing number of buffaloes, if milch bovines are to be continuously exploited to ensure food security and improved rural livelihood in the country.

Table 5 Criteria for model selection for buffalo population

| Model            | $R^2$ | RMSE     | MAPE  | MAE    | BIC    |
|------------------|-------|----------|-------|--------|--------|
| ARIMA (1,1,1)    | 0.999 | 567161.153 | 0.501 | 381150.716 | 26.788 |
| ARIMA (1,1,0)    | 0.999 | 569149.589 | 0.520 | 393204.556 | 26.722 |
| ARIMA (0,1,1)    | 0.999 | 577715.345 | 0.539 | 410116.112 | 26.752 |
| ARIMA (0,1,0)    | 0.999 | 563598.589 | 0.596 | 401268.973 | 26.789 |
| ARIMA (0,1,2)    | 0.999 | 565862.894 | 0.512 | 425689.178 | 26.987 |
| ARIMA (1,1,2)    | 0.999 | 602589.241 | 0.536 | 394685.156 | 26.897 |
| ARIMA (2,1,0)    | 0.999 | 586947.235 | 0.548 | 375984.679 | 26.698 |
| ARIMA (2,1,1)    | 0.999 | 614859.245 | 0.591 | 389125.279 | 26.746 |
| ARIMA (2,1,2)    | 0.999 | 598674.259 | 0.459 | 412689.265 | 26.823 |
| ARIMA (1,2,1)    | 0.999 | 547785.295 | 0.437 | 338227.316 | 26.723 |
| ARIMA (1,2,0)    | 0.999 | 625475.653 | 0.436 | 323676.971 | 26.914 |
| ARIMA (0,2,1)    | 0.999 | 545105.745 | 0.439 | 342153.624 | 26.639 |
| ARIMA (0,2,0)    | 0.999 | 598647.258 | 0.612 | 396587.415 | 26.789 |
| ARIMA (0,2,2)    | 0.999 | 602146.359 | 0.658 | 412689.345 | 26.989 |
| ARIMA (1,2,2)    | 0.999 | 576298.345 | 0.589 | 402356.246 | 26.874 |
| ARIMA (2,2,0)    | 0.999 | 563897.156 | 0.754 | 396589.125 | 27.215 |
| ARIMA (2,2,1)    | 0.999 | 587569.217 | 0.753 | 435689.456 | 27.198 |
| ARIMA (2,2,2)    | 0.999 | 589246.256 | 0.659 | 482678.159 | 28.234 |
| Simple ES        | 0.996 | 1265498.454 | 1.459 | 1140561.574 | 28.174 |
| Holt ES          | 0.999 | 571353.153 | 0.466 | 357039.178 | 26.655 |
| Brown ES         | 0.996 | 577912.272 | 0.445 | 336752.723 | 26.606 |
| Damped Trend ES  | 0.999 | 576353.100 | 0.492 | 376768.763 | 26.745 |

Table 6 Estimate of the best fit Brown Exponential Smoothing model for buffalo population

| Model          | Parameter | Estimate | SE   | t      | Sig.    |
|----------------|-----------|----------|------|--------|---------|
| Brown ES       | Alpha     | 0.671    | 0.065| 10.31  | 0.000   |

Table 7 Forecasts of buffalo population (in million)

| Year  | Actual | Predicted | LCL(95%) | UCL(95%) | Residual |
|-------|--------|-----------|----------|----------|----------|
| 2000-01 | 93.831 | 93.762    | 92.605   | 94.921   | 0.068    |
| 2010-11 | 107.375 | 107.757   | 106.600  | 108.916  | -0.382   |
| 2020-21 - | 116.663 | 112.815   | 120.513  | -        |
| 2030-31 - | 127.787 | 109.455   | 146.120  | -        |
| 2040-41 - | 138.910 | 100.185   | 177.636  | -        |
| 2050-51 - | 148.921 | 87.957    | 209.886  | -        |
References

Ahmad, WKAW, S Ahmad (2013) ARIMA model and exponential smoothing method: A comparison. AIP Conf Proc 1522.
Borah M, Halim RA (2014) Dynamics and performance of livestock and poultry sector in India: A temporal analysis. J Acad Indus Res 3: 1-9
Box GEP, Jenkins GM, Reinsel GC (2007) Time – series analysis: Forecasting and control. (3). Pearson education, India
Brown RG (1963) Smoothing, forecasting and prediction of discrete time series. Englewood Cliffs NJ: Prentice-Hall
Choudhury A, James J (2014) Crop yield prediction using time series models. J Econ Educ 15: 53-68
Celik S (2016) Forecasting Production of some cereal in Turkey by time series analysis. Int J Inf Res Rev 3: 2887-2897
Celik S, Sengul T (2016) Forecasting numbers of poultry in Turkey using Exponential Smoothing techniques. Int J Sci Res 5: 2277-8179
Chaudhari DJ, Tingre AS (2015) Forecasting eggs production in India. Indian J Anim Res 49: 367-372
Gardner ES, E McKenzie (1985) Forecasting trends in time series. Manag Sci 31: 1237-1246
Government of India (2015-16), NSS 72nd Round Survey (June 2015 – June 2016) on Employment and Unemployment
Government of India (2018-19), Department Animal Husbandry and Dairying, Annual Report, P.1.
Government of India (2019), Basic Animal Husbandry Statistics, 2019, P.99
Hanke JE, Wichern DW (2008) Business Forecasting. 8th Ed. Pearson Education International; Harlow, Essex
Prabhu M, Kumar GS, Pandian ASS, Selvakumar KN, Varathan BJ (2012) Dynamics of livestock population - India vis-à-vis Tamil Nadu. Tamil Nadu J Vet Sci 8: 266-270
Prasad S, Singh R, Lal K, Mishra SN (2004) Growth of livestock in India. Pashudhan Anusandhan 4: 47-53.
Sharpe R, Vaux RD, Velleman PF (2010) Business Statistics, 2nd Edition, Addison Vesley - Pearson Education; Boston
Talwar A, Goyal CK 2019. A comparative study of various exponential smoothing models for forecasting coriander price in Indian commodity market. Int Bull Manag Econ 10: 143-155
Thirunavukkarasu M, A Rajarathinam (2014) Stochastic Modelling for forecasting of India’s milled rice production. Int Sci Res 3: 2277-8179