Journal of Modern Mathematics and Statistics 4 (2): 53-57, 2010
ISSN: 1994-5388


# Stochastic Modeling for Cattle Production Forecasting

[1]T. Jai Sankar, [1]R. Prabakaran, [2]K. Senthamarai Kannan and [2]S. Suresh
[1]Madras Veterinary College, University of Tamilnadu Veterinary and Animal Sciences, Chennai 600-007, India
[2]Department of Statistics, University of Manonmaniam Sundaranar, Tirunelveli 627-012, India

**Abstract:** This study proposes a technique using Autoregressive Integrated Moving Average (ARIMA) Model for cattle production. Stochastic modeling and forecasting plays a vital role in many fields such as agricultural production, animal husbandry economics, stock prices prediction, etc. ARIMA Model was introduced by Box and Jenkins. Hosking has introduced a family of models called fractionally differenced autoregressive integrated moving average models by generalizing the d fraction in ARIMA (p, d, q) models. Mandal was using ARIMA Model for analyzing sugarcane production. This study analysis the design of ARIMA process to select the appropriate model for cattle production in Tamilnadu. These results are verified on the basis of various diagnostic checking and error analysis which is used to forecast the future values. Also, results are shown by graphically and numerically.

**Key words:** Cattle production, AIC, BIC, ARIMA, forecasting, India

## INTRODUCTION

India is an agricultural country with about 70% of its population dependent on income from agriculture. Cattle and buffaloes are maintained for milk production, motive power for various farm operations, village transport, irrigation and production of manure. The animals are generally maintained for agricultural byproducts and crop residues. The small income farmers and diary developers are well based on the cattle production. But the cattle production is very low for the past 25 years.

In fact livestock and human are dependent on each other. Cattle were raised mainly to get the male calves which were used for agriculture fields and dung for enriching the soil. Higher the number of the cattle maintained meant the higher the availability of the bullock /draught power and the farm yard manure, due to which the productivity and the production is higher.

## MATERIALS AND METHODS

In this study, the source of data for cattle production in Tamilnadu is collected from the Department of Animal Husbandry and Veterinary Services, Government of Tamilnadu for the period 1970-2008. ARIMA model was introduced by Box and Jenkins (1970) and is used for discovering the pattern and predict the future values of the time series data. Akaike (1970) discussed with the stationary time series by an AR (p), p is finite and bounded by the same integer. The Moving Average (MA) models were first used by Slutzky (1973). Hannan and Quinn (1979) for pure AR models and Hannan (1980) for ARMA models, suggest obtaining the order of a time series model by minimizing the errors. Prajnesh and Venugopalan (1996) have 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 of the country during the period 1950-51 to 1994-95. Model parameters were estimated using the Statistical Package for Social Sciences (SPSS) package to fit the ARIMA models. ARIMA process for any variable involves four steps: Identification, estimation, diagnostic checking and forecasting. Each of these four steps is explained for cattle production.

**ARIMA process:** The time series when differenced follows an AR and MA model is known as autoregressive integrated moving averages (ARIMA) model. Autoregressive process of order (p) is:

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

**Corresponding Author:** T. Jai Sankar, Madras Veterinary College, University of Tamilnadu Veterinary and Animal Sciences, Chennai 600-007, India

Moving average process of order (q) is:

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

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

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

Where $\varepsilon_t$'s are independently and normally distributed with zero mean and constant variance $\sigma^2$ for t = 1, 2,...n. The different models can be obtained for various combinations of autoregressive and moving average. The best model is obtained with the following diagnostics low Akaike's Information Criteria (AIC) which is defined by:

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

Where m = p+q+P+Q and L is the likelihood function. Since -2 logL is approximately equal to n (1+log2Π) + $n\log\sigma^2$, where $\sigma^2$ is the mean square error. Also AIC can be written as:

$$AIC = (n (1+\log 2\Pi) + n \log \sigma^2 + 2m)$$

and Schwartz Bayesian Criteria (SBC) is defined by:

$$SBC = \frac{\log \sigma^2 + (m \log n)}{n}$$

To check the adequacy for the residuals using Q statistic. A modified Q statistic is the Box-Ljung Q statistic is defined by:

$$Q = \frac{n (n+2)\sum rk^2}{(n\text{-}k)}$$

Where:

$r_k$ = The residual autocorrelation at lag k
n = The number of residuals

The Q statistic is compared to critical value from Chi square distribution. If the p-value associated with Q statistic is small ($p<\alpha$), the model is consider in adequate. Forecasting the future periods using the parameters for the tentative model has been selected.

**Analysis and trend fitting techniques:** For evaluating the AR, MA and ARIMA process adequacy, various reliability statistics like $R^2$, Stationary $R^2$, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Maximum Absolute Percentage Error (MaxAPE), Mean Absolute Error (MAE), Maximum Absolute Error (MaxAE) and Normalized BIC have been used. Lesser the various reliability statistics better will be the efficiency of the model in predicting the future cattle production. For calculating the Box-Ljung, Q statistics have also been used.

## RESULTS AND DISCUSSION

**Model identification:** ARIMA model is designed after assessing which varies the variable under forecasting as a stationary series. The stationary series is the set of values vary over period of time around a constant mean and constant variance. The stationarity is checked by graphical representation.

Figure 1 shows that the data is non-stationary. Non-stationarity in mean is corrected through first differencing of the data. For this purpose, the various autocorrelations up to 12 lags were computed and the same along with their significance which is tested by Box-Ljung test are shown in Table 1.

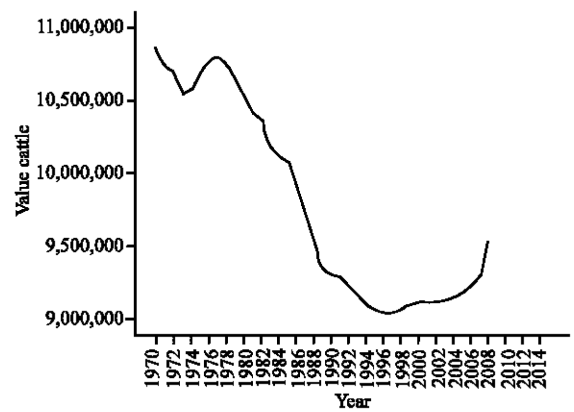


Fig. 1: Time plot of cattle production in tamilnadu

Table 1: ACF and PACF of cattle production

| Lag | Auto correlation value | Std. error[a] (Df) | Box-ljung statistic Sig.[b] | Value | Partial auto correlation Df | Value | Std. error (Df) |
|---|---|---|---|---|---|---|---|
| 1 | 0.560 | 0.156 | 12.877 | 1 | 0.000 | 0.560 | 0.162 |
| 2 | 0.356 | 0.154 | 18.235 | 2 | 0.000 | 0.062 | 0.162 |
| 3 | 0.218 | 0.152 | 20.299 | 3 | 0.000 | -0.006 | 0.162 |
| 4 | 0.090 | 0.150 | 20.664 | 4 | 0.000 | -0.065 | 0.162 |
| 5 | 0.162 | 0.147 | 21.871 | 5 | 0.001 | 0.184 | 0.162 |
| 6 | 0.169 | 0.145 | 23.235 | 6 | 0.001 | 0.042 | 0.162 |
| 7 | 0.158 | 0.143 | 24.464 | 7 | 0.001 | 0.016 | 0.162 |
| 8 | 0.136 | 0.140 | 25.407 | 8 | 0.001 | -0.004 | 0.162 |
| 9 | 0.073 | 0.138 | 25.686 | 9 | 0.002 | -0.019 | 0.162 |
| 10 | 0.023 | 0.136 | 25.716 | 10 | 0.004 | -0.037 | 0.162 |
| 11 | -0.156 | 0.133 | 27.086 | 11 | 0.004 | -0.253 | 0.162 |
| 12 | -0.251 | 0.131 | 30.767 | 12 | 0.002 | -0.139 | 0.162 |

[a]The underlying process assumed is independence (white noise), [b]Based on the asymptotic chi-square approximation

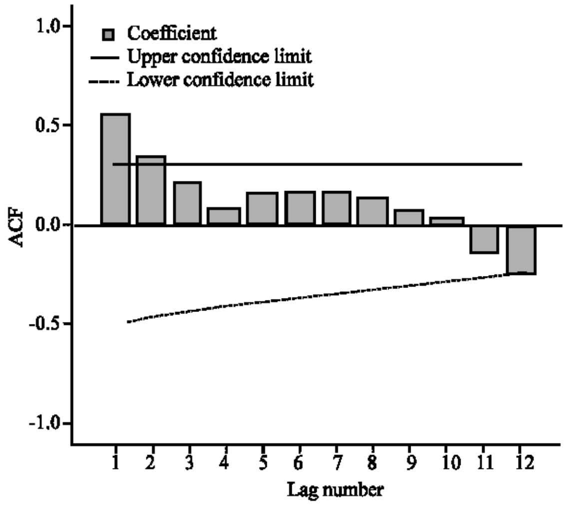


Fig. 2: ACF of differenced data

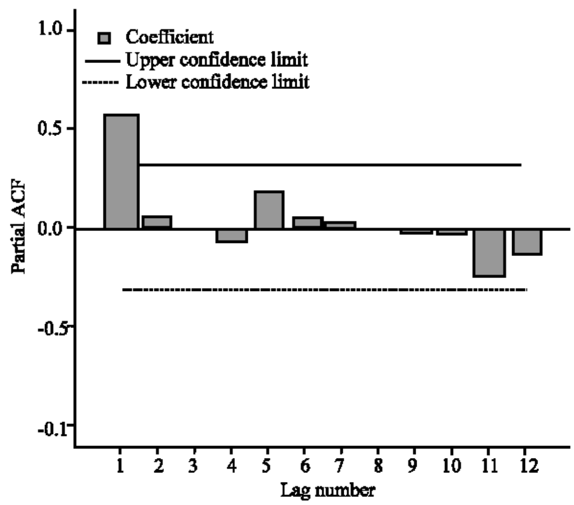


Fig. 3: PACF of differenced data

The graphs of ACF and PACF are shown in Fig. 2 and 3. The tentative ARIMA models are described with differenced once and model is chosen which has minimum normalized BIC (Bayesian Information Criterion).

The models and corresponding normalized BIC values are shown in Table 2. The value of normalized AIC is 959.78 and $R^2$ value is 99%. So the most suitable model for Cattle Production is ARIMA (1, 1, 0) as this model has the lowest AIC value.

**Model estimation:** Model parameters were estimated using SPSS package. Results of estimation are shown in Table 3 and 4.

**Diagnostic checking:** Based on the estimation, the autocorrelations and partial autocorrelations of the residuals of various orders are analysed. For this purpose, the various autocorrelations up to 12 lags were computed and the same along with their significance are shown in Table 5. As the results show, none of these autocorrelations is significantly different from zero at a

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

| ARIMA (p, d, q) | AIC values | BIC values |
|---|---|---|
| (0, 1, 2) | 964.19 | 972.38 |
| (1, 1, 0) | 959.78 | 966.30 |
| (2, 1, 0) | 961.74 | 964.65 |

Table 3: Estimated ARIMA model of cattle production

| ARIMA | Estimate | Std error | T | P-sig. |
|---|---|---|---|---|
| Constant | -10247810.729 | 5409130.414 | -1.895 | 0.066 |
| AR 1 | 0.659 | 0.147 | 4.476 | 0.000 |

Table 4: Estimated ARIMA model fit statistics

| Fit statistic | Mean |
|---|---|
| Stationary $r^2$ | 0.482 |
| $R^2$ | 0.990 |
| RMSE | 68725.053 |
| MAPE | 0.483 |
| MaxAPE | 1.711 |
| MAE | 47774.553 |
| MaxAE | 164316.026 |
| Normalized BIC | 22.563 |

Table 5 : Residual of ACF and PACF of cattle production

| | ACF | | PACF | |
|---|---|---|---|---|
| Lag | Mean | SE | Mean | SE |
| Lag 1 | 0.028 | 0.162 | 0.028 | 0.162 |
| Lag 2 | 0.009 | 0.162 | 0.008 | 0.162 |
| Lag 3 | 0.077 | 0.162 | 0.076 | 0.162 |
| Lag 4 | -0.178 | 0.163 | -0.183 | 0.162 |
| Lag 5 | 0.076 | 0.168 | 0.091 | 0.162 |
| Lag 6 | 0.024 | 0.168 | 0.013 | 0.162 |
| Lag 7 | -0.019 | 0.169 | 0.007 | 0.162 |
| Lag 8 | 0.029 | 0.169 | -0.018 | 0.162 |
| Lag 9 | -0.044 | 0.169 | -0.019 | 0.162 |
| Lag 10 | 0.143 | 0.169 | 0.155 | 0.162 |
| Lag 11 | -0.101 | 0.170 | -0.134 | 0.162 |
| Lag 12 | -0.272 | 0.173 | -0.273 | 0.162 |

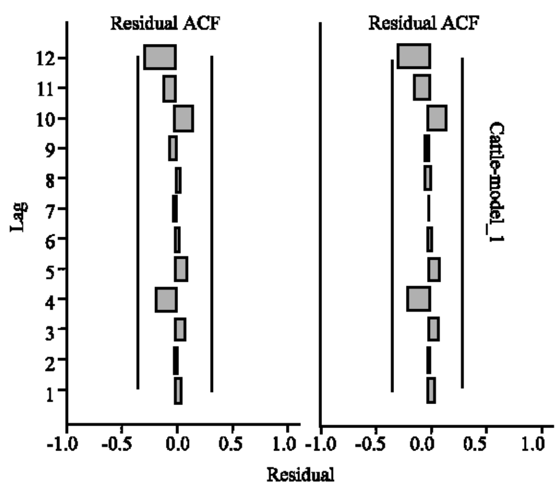


Fig. 4: Residuals of ACF and PACF

Table 6 : Forecast of cattle production

| Years | Actual production | Predicted | LCL | UCL | Residual |
|---|---|---|---|---|---|
| 1970 | 10859345 | - | - | - | - |
| 1971 | 10731248 | 10737991 | 10552490 | 10923493 | -67430 |
| 1972 | 10695387 | 10610588 | 10471077 | 10750099 | 847990 |
| 1973 | 10541932 | 10637269 | 10497758 | 10776780 | -95337 |
| 1974 | 10572378 | 10408062 | 10268551 | 10547573 | 164316 |
| 1975 | 10692345 | 10561464 | 10421954 | 10700975 | 130881 |
| 1976 | 10764573 | 10742184 | 10602673 | 10881695 | 223890 |
| 1977 | 10801119 | 10784700 | 10645189 | 10924211 | 164190 |
| 1978 | 10736542 | 10799481 | 10659970 | 10938991 | -62939 |
| 1979 | 10649721 | 10670007 | 10530496 | 10809518 | -20286 |
| 1980 | 10532345 | 10570277 | 10430766 | 10709788 | -37932 |
| 1981 | 10424556 | 10434515 | 10295004 | 10574026 | -99590 |
| 1982 | 10365500 | 10334796 | 10195285 | 10474307 | 307040 |
| 1983 | 10218734 | 10309610 | 10170099 | 10449121 | -90876 |
| 1984 | 10123950 | 10106788 | 99672770 | 10246299 | 171620 |
| 1985 | 10093567 | 10048016 | 99085050 | 10187527 | 455510 |
| 1986 | 99611340 | 10061830 | 99223190 | 10201341 | -100696 |
| 1987 | 97456720 | 98638890 | 97243790 | 10003400 | -118217 |
| 1988 | 95800230 | 95954570 | 94559460 | 97349670 | -154340 |
| 1989 | 93531410 | 94643900 | 93248790 | 96039010 | -111249 |
| 1990 | 93144420 | 91989020 | 90593910 | 93384130 | 115540 |
| 1991 | 92978670 | 92859820 | 91464710 | 94254930 | 118850 |
| 1992 | 92374510 | 92857400 | 91462290 | 94252510 | -48289 |
| 1993 | 91675830 | 91981810 | 90586700 | 93376920 | -30598 |
| 1994 | 90961210 | 91238350 | 89843240 | 92633460 | -27714 |
| 1995 | 90743890 | 90530740 | 89135630 | 91925850 | 213150 |
| 1996 | 90521740 | 90658690 | 89263580 | 92053800 | -13695 |
| 1997 | 90465420 | 90450870 | 89055760 | 91845980 | 145500 |
| 1998 | 90783450 | 90521360 | 89126250 | 91916470 | 262090 |
| 1999 | 90992340 | 91103630 | 89708530 | 92498740 | -11129 |
| 2000 | 91159280 | 91258110 | 89863000 | 92653220 | -98830 |
| 2001 | 91197650 | 91414920 | 90019810 | 92810030 | -21727 |
| 2002 | 91263360 | 91386060 | 89990960 | 92781170 | -12270 |
| 2003 | 91356790 | 91487310 | 90092200 | 92882420 | -13052 |
| 2004 | 91410430 | 91616520 | 90221420 | 93011630 | -20609 |
| 2005 | 91783210 | 91661460 | 90266350 | 93056560 | 121750 |
| 2006 | 91993450 | 92262090 | 90866980 | 93657200 | -26864 |
| 2007 | 92843710 | 92382720 | 90987610 | 93777830 | 460990 |
| 2008 | 95302680 | 93672320 | 92277210 | 95067430 | 163036 |
| 2009 | - 97209070 | 95813960 | 98604180 | - | |
| 2010 | - 98768770 | 96066240 | 10147130 | - | |
| 2011 | - 10011751 | 96138360 | 10409665 | - | |
| 2012 | - 10134471 | 96162420 | 10652700 | - | |
| 2013 | - 10250933 | 96208870 | 10880979 | - | |
| 2014 | - 10365021 | 96314750 | 11098568 | - | |
| 2015 | - 10479298 | 9649874 | 11308721 | - | |

reasonable level. This proves that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals are shown in Fig. 4. It also indicates good fit of the model. So the fitted ARIMA model for the cattle production data is:

$$Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$$

$$Y_t = -10247810.729 + 0.659 Y_{t-1} + \varepsilon_t$$

**Forecasting:** Forecasted value of cattle production (Quantity in numbers) for the year 2009 through 2015 are shown in Table 6. To assess the forecasting ability of the fitted ARIMA model, important measures of the sample period forecasts' accuracy were computed.

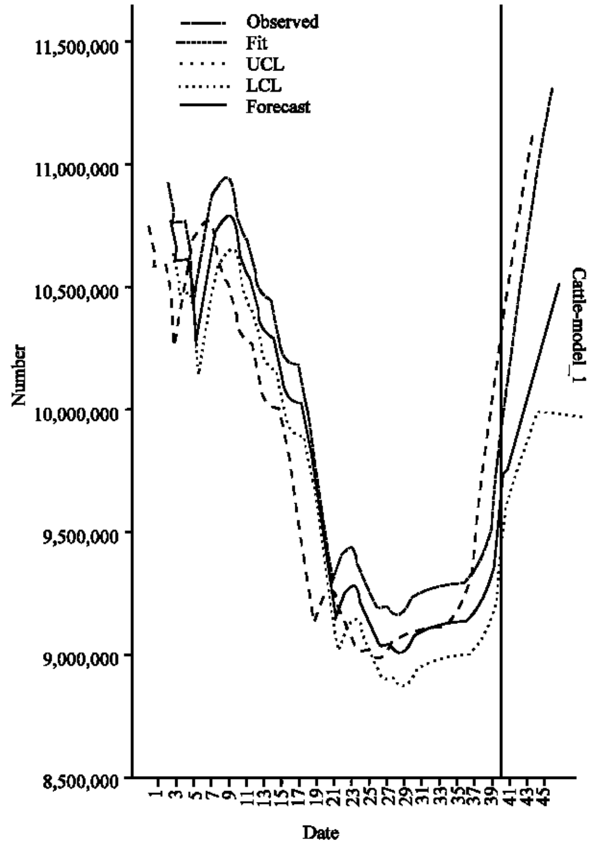


Fig. 5: Actual and estimate of cattle production

This measure indicates that the forecasting inaccuracy is low. Figure 5 shows that the actual and forecasted value of cattle production data with 95% confidence limits.

The constructed model designed for cattle production is found to be ARIMA (1, 1, 0). Based on the numerical calculations and graphical representations, it can be found that forecasted production for the year 2009 is <2010 but in subsequent years the production increases. The validity of the forecasted values can be verified for the period from 1970-2008 regarding cattle production. This study provides evidence on complete cattle production data.

## CONCLUSION

The estimated results indicate that there is an increase in the cattle production which will improve the economy of the state. This provides evidence in favour of Box-Jenkins methodology as it applies to cattle production and future efficiency.

## REFERENCES

Akaike, H., 1970. Statistical predictor identification. Ann. Inst. Stat. Math., 22: 203-217.

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

Hannan, E.J. and B.G. Quinn, 1979. The determination of the order of an autoregression. J. R. Stat. Soc. Ser. B, 41: 190-195.

Hannan, E.J., 1980. The estimation of the order of an ARMA process. Annu. Statistics, 8: 1071-1081.

Prajnesh and R. Venugopalan, 1996. Trend analysis in all India marine products export using statistical modeling techniques. Indian J. Fish., 43: 107-113.

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