Journal of Agriculture, Biology and Applied Statistics
Vol. 3, No. 2, 2024, pp. 47-56

URL : www.arfjournals.com
https://DOI:10.47509/JABAS.2024.v03i02.01

Forecast of Milk and Eggs Production of India using ARIMA Model

S. R. Krishna Priya[1] and N. Naranammal[2]
Assistant Professor[1], Research Scholar[2]
Department of Statistics, PSG College of Arts & Science, Coimbatore -14

To cite this article

S.R. Krishna Priya & N. Naranammal (2024). Forecast of Milk and Eggs Production of India using ARIMA Model. Vol. 3, No. 2, pp. 47-56. *https://DOI:10.47509/JABAS.2024.v03i02.01*

Abstract: Milk and eggs are the most important sources of nutrition consumed daily in every household in many forms. There is an increase of milk and egg consumption due to the increase in population in India. Per capita consumption of milk and eggs has been drastically increasing over the past decades. Therefore, forecasting milk and egg production is necessary to make a strategic plan for the government. In this study, milk and egg production in India has been forecast using auto regressive integrated moving average (ARIMA) model. Milk and egg production data from year 1950-51 to 2019-20 has been used for the development of the model. Results revealed that ARIMA (2,2,0) and ARIMA (1,2,2) are the best models for forecasting milk and egg production respectively. The ACF and PACF plots of residuals do not have autocorrelation. It implies that the proposed model is the best model for forecasting. Result of the goodness of fit measure, mean absolute percentage error is satisfactory. The production of milk and eggs from years 2020-21 to 2024-25 has been forecasted using the best ARIMA model. From the forecasted value, it is clear that there will be an increasing demand for milk and eggs in the upcoming years.

Keywords: Milk, Egg, ARIMA, Forecasting, Stationarity, Poultry, Dairy, Modelling.

1. Introduction

Milk provides essential nutrients for strong bones and teeth, such as calcium and vitamin D. It is an important food consumed daily in different forms. Approximately 150 million households around the globe are engaged in milk production. Milk provides relatively quick returns for small-scale producers and is an important source of cash income. In recent decades, developing countries have increased their share of global dairy production. This growth is mostly the result of an increase in the number of producing animals rather than a rise in productivity per head. A recent study shows that India is the largest producer of milk in the world, accounting for nearly 10% of the world production. Various milk-based products that can be manufactured commercially in a rural industry, such as paneer, cheese, curd, ghee,

Article History Received : 17 March 2024; Revised : 11 April 2024; Accepted : 21 April 2024; Published : 30 September 2024

etc. India has the world's largest dairy herd with over 300 million bovines, producing over 187 million metric tons of milk. India is first among all countries in both production and consumption of milk. Most of the milk is domestically consumed, though a small fraction is also exported. Poultry eggs are an important source of high-quality proteins, minerals, and vitamins to balance the human diet. The per capita availability of eggs was 86 eggs per year in 2019–20. Poultry production in India has taken a quantum leap in the last four decades, emerging from the use of unscientific farming practices to commercial production systems with state-of-the-art technological interventions. Egg and milk production in India is crucial for ensuring nutritional security and providing affordable protein and essential nutrients to a vast population. Understanding the significant value of milk and eggs in everyday life needs the ability to forecast milk and egg production.

ARIMA is a time series model used widely for forecasting in many fields. Researchers have used ARIMA for forecasting cotton production, prices of tomatoes, production of jasmine flower, prices of onion, rice and wheat yield of SAARC nations, wholesale groundnut prices, wheat yield, prices of paddy and farm prices of ragi (Poyyamizhi and Mohideen, 2017; Amerander, 2018; Prakash and Muniyandi, 2014; Ashwini *et al.*, 2015; Sahu *et al.*, 2015; Bannor and Sharma, 2016; Gurung *et al.*, 2017; Darekar and Reddy, 2017; Jadhav *et al.*, 2017).

The present study is an attempt to forecast milk and eggs production of India using ARIMA model.

2. Materials and Methods

2.1. Data Description

In this study, annual data of milk and egg production of India from year 1950-51 to 2019-20 have been used for forecasting. The data is collected from indiastat.com. ARIMA model has been developed for modeling and forecasting.

2.2. Stationarity Test

A time series is said to be stationary if its underlying generating process is based on a constant mean and constant variance, with its autocorrelation function (ACF) essentially constant through time. Otherwise, it is called non-stationary. A common statistical test to check the stationarity of the data is the Augmented Dickey Fuller (ADF) test. The null hypothesis that a unit root exists in a time series sample is tested with an augmented Dickey–Fuller test (ADF). If the null hypothesis is rejected, it implies that the data is stationary.

2.3. Auto Correlation and Partial Auto Correlation Function

Autocorrelation among the elements in a time series is the correlation between the observations in the series lagged by time periods. The autocorrelation function (ACF) is defined by time lags k=1, 2, 3, the values of ACF along the y-axis. This is another graphical tool to examine stationary properties with respect to their mean.

The partial autocorrelation function (PACF) is similar to an ACF. It measures the correlation between the observations of a time series at different lags, but it removes the effects of the correlations at shorter lags. For stationary time series data, the spikes of ACF and PACF will be closer to zero or inside the confidence limits. However, if the spikes are large and outside the confidence limits at some point, they will be outside the confidence limits.

2.4. ARIMA Model

An autoregressive integrated moving average (ARIMA) is a statistical analysis model that uses time series data to understand the data set and predict future trends. ARIMA makes use of lagged moving averages to smooth time series data. It is a time series forecasting method that combines auto regression (AR), differencing (I), and moving average components (MA). If the data is stationary, ARIMA model is reduced to ARMA model. If the data is non stationary, it is differentiated until it becomes stationary. The ARIMA model consists of three parameters (p,d,q). p is the order of auto regression which is obtained from PACF, q is the order of moving average obtained from ACF and d is the order of differencing. The ARIMA model involves four stages, such as:

Identification: Identification consists of deciding the values of p,d,q. The patterns of spikes created corresponding to the values of PACF and ACF plots can be used to find the values of p and q. The value of d is calculated by the number of times the data is differentiated to become stationary.

Estimation of Parameters: After identifying the model, the parameters of AR and MA are estimated using least squares with a minimum error sum of squares.

Diagnostic Checking: After estimating the parameters, diagnostic checking of the model is done to verify its adequacy. Examining ACF and PACF of residuals determines whether the residuals have autocorrelation or not. If the residuals of the fitted model do not have autocorrelation, the model is adequate enough to forecast the time series data.

Forecasting: After finding an adequate model, the data is forecasted using the best ARIMA model.

2.5. Goodness of Fit Measures

Goodness of fit measures such as AIC and MAPE have been used to evaluate the models.

2.5.1. Akaike information criterion (AIC)

$$AIC = -2\log L + 2n \tag{1}$$

Where, L is likelihood function, n is number of hyperparameters.

2.5.2. Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{1}{n}\Sigma_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

Where,

n- number of observations yi - actual value and $\hat{y}_i$ - predicted value

3. Results and Discussion

3.1. Forecasting of Milk Production

ARIMA model is developed to forecast the milk production of India. The time series plot of milk production of India is presented in Figure 1.

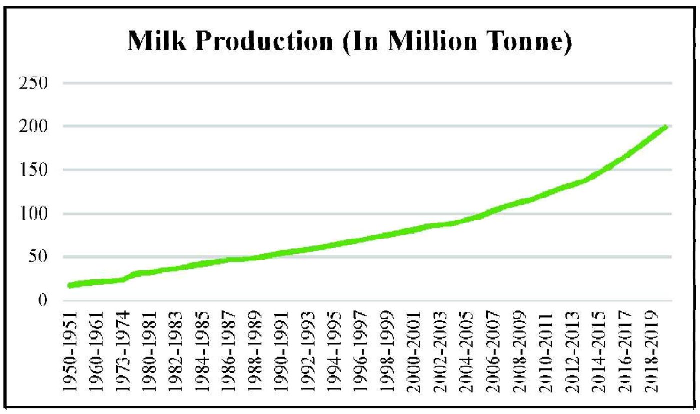


Figure 1: Time Series Plot of Milk Production

From figure 1, it is noticed that there is linear upward trend in the data. It implies that, every year there is an increase in milk production. The ACF and PACF plot of milk production is presented in Figure 2.

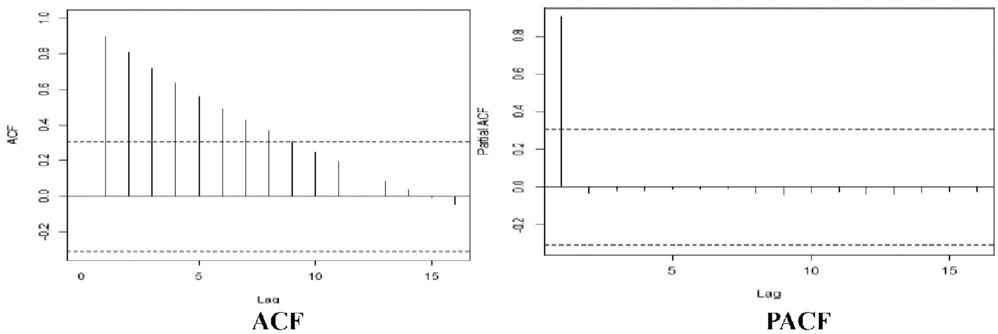


Figure 2: ACF and PACF plot of Milk Production

From the ACF and PACF plot, it is clear that the data is non-stationary. Additionally, ADF test is used to check the stationarity and the result is presented in Table 1.

Table 1: Results of ADF Test of Milk Production

Dickey-Fuller Statistic	*Lag Order*	*p-value*
2.9761	3	0.99

The p value of ADF test is greater than 0.05. So, null hypothesis given the data is non stationary is accepted. It implies that the milk production data is non-stationary. The ARIMA model can be identified using ACF and PACF plot. But trying different combination of ARIMA model and choosing the best model from lower AIC values is the best method to determine the ARIMA model. The AIC values of different combination of ARIMA model is given in Table 2.

Table 2: AIC Values of Different ARIMA Models

Model	*AIC*
ARIMA (0,2,0)	107.8363
ARIMA (0,2,1)	106.4257
ARIMA (0,2,2)	106.6538
ARIMA (1,2,0)	108.8098
ARIMA (1,2,1)	107.9061
ARIMA (2,2,0)	105.5229
ARIMA (2,2,1)	107.4839
ARIMA (3,2,0)	107.5063
ARIMA (4,2,0)	109.0141
ARIMA (5,2,0)	110.5513

ARIMA (2,2,0) has the lowest AIC value of 105.52 among the different models. So, the chosen best model among different combination is ARIMA (2,2,0). Next step is to diagnostic checking of the chosen model. The ACF, PACF and time series plot of residuals of ARIMA (2,2,0) model is presented in Figure 3.

The spikes of lags of ACF and PACF plots are within the confidence limit. It implies that there is no autocorrelation of the residuals. From this diagnostic checking, it is concluded that ARIMA (2,2,0) is the best model for forecasting milk production. Using ARIMA (2,2,0) model, milk production for year 2020-21 to 2024-25 has been forecasted and values are presented in Table 3.

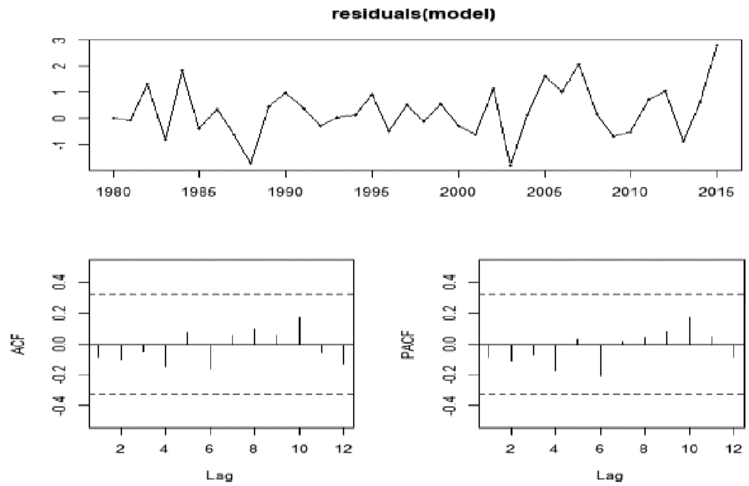


Figure 3: Time Series, ACF and PACF Plots of Residuals of ARIMA (2,2,0) Model

Table 3: Forecast Value of Milk Production

Year	*Forecast Value (Million Tonne)*
2020-21	209.0539
2021-22	219.8541
2022-23	230.6521
2023-24	241.4204
2024-24	252.1915
MAPE	5.24

Goodness of fit measure mean absolute percentage error is calculated and the result is satisfactory.

3.2. Forecasting of Egg Production:

ARIMA model is developed to forecast the egg production of India. The time series plot of milk production of India is presented in Figure 4.

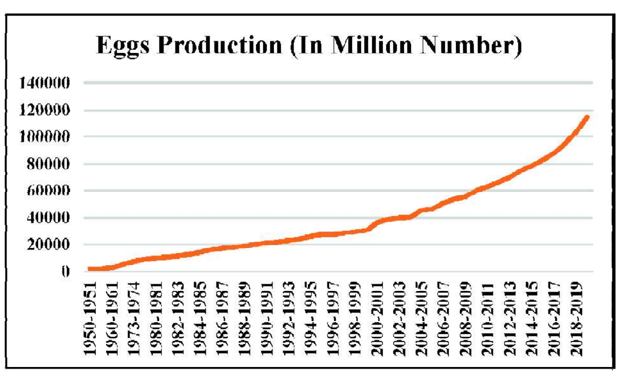
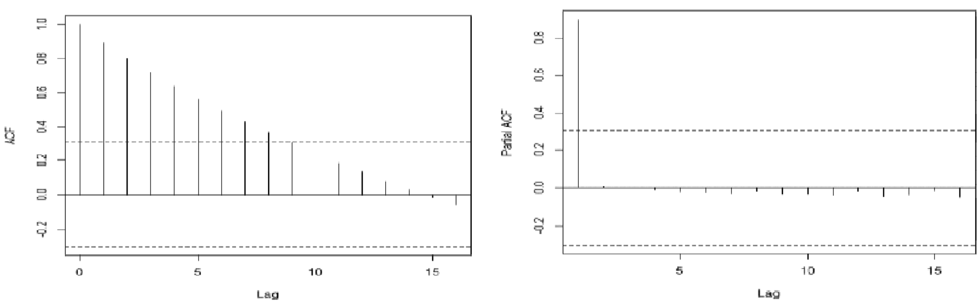


Figure 4: Time Series Plot of Egg Production

There is a linear upward trend in egg production of India. The ACF and PACF plot of egg production is presented in Figure 5.



ACF PACF

Figure 5: ACF and PACF Plots of Egg Production

From the ACF and PACF plot, it is clear that the data is non-stationary. Additionally, ADF test is used to check the stationarity and the result is presented in Table 4.

Table 4: Results of ADF Test of Egg Production

Dickey-Fuller Statistic	*Lag Order*	*p-value*
3.399	3	0.99

The p value of ADF test is greater than 0.05. So, null hypothesis given the data is non stationary is accepted. It implies that the milk production data is non-stationary. The best ARIMA model is be identified using ACF and PACF plot. But trying different combination of ARIMA model and choosing the best model from lower AIC values is the best method to determine the best ARIMA model. The AIC values of different combination of ARIMA model is given in Table 5.

Table 5: Different Combination of ARIMA Model for Egg Production

Model	*AIC*
ARIMA (0,2,0)	608.8729
ARIMA (0,2,1)	590.6239
ARIMA (0,2,2)	590.9002
ARIMA (0,2,3)	590.3969
ARIMA (0,2,4)	591.0953
ARIMA (0,2,5)	593.0833
ARIMA (1,2,0)	598.2952
ARIMA (1,2,1)	591.472
ARIMA (1,2,2)	589.1609
ARIMA (1,2,3)	591.1579

Model	*AIC*
ARIMA (1,2,4)	592.773
ARIMA (2,2,0)	596.2898
ARIMA (2,2,1)	593.0768
ARIMA (2,2,2)	591.1578
ARIMA (3,2,0)	594.488
ARIMA (3,2,1)	594.1014
ARIMA (3,2,2)	592.6401
ARIMA (4,2,0)	595.3856
ARIMA (4,2,1)	595.9158
ARIMA (5,2,0)	594.9318

Among different combination ARIMA (1,2,2) has the lowest AIC value. Therefore, ARIMA (1,2,2) is the best model among other models. After identifying the best model, diagnostic checking has been done to the chosen model. The ACF, PACF and time series plot of residuals of ARIMA (1,2,2) model is presented in Figure 6.

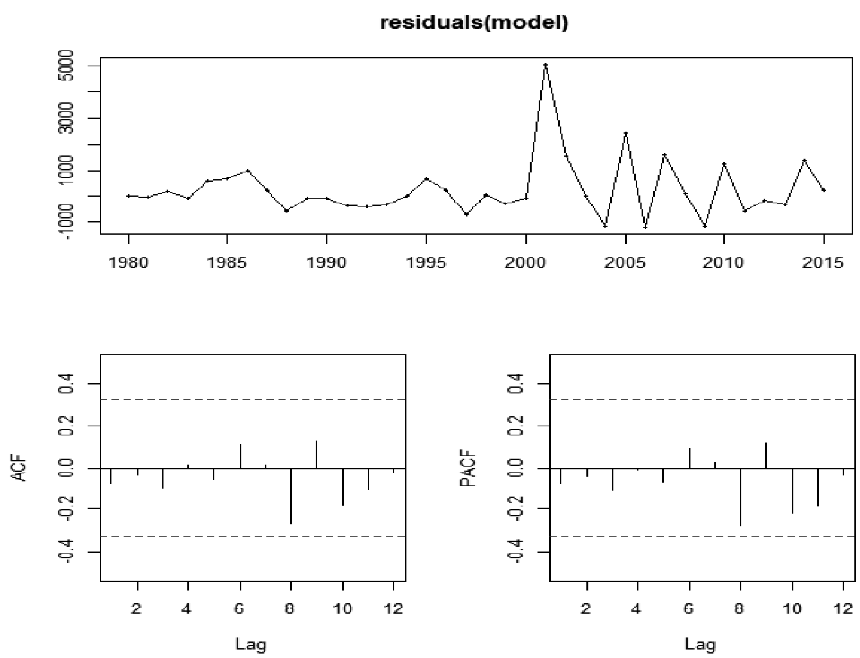


Figure 6: Time Series, ACF and PACF Plots of Residuals of ARIMA (1,2,2) Model

From Figure 6 it is clear that, the residuals of ARIMA (1,2,2) model does not have autocorrelation. So, it is diagnosed that, ARIMA (1,2,2) is the best model for forecasting egg production. Using the best model, egg production for year 2020-21 to 2024-25 have been forecasted and results are presented in Table 6.

Table 6: Forecast Values of Egg Production

Year	*Forecast Value (in Million Numbers)*
2020-21	123421.2
2021-22	134206.8
2022-23	146618.0
2023-24	160541.1
2024-25	175870.5
MAPE	6.74

4. Conclusion

Egg and milk production are vital in India for improving nutritional security and providing livelihoods to millions of farmers. In this study, the ARIMA model is developed to forecast milk and egg production in India. The time series data for milk and egg production is non-stationary. So, ARIMA is used for the development of models. After fitting different combinations of (p,d,q) values, ARIMA (2,2,0) and ARIMA (1,2,2) are the best models for forecasting milk and egg production respectively. From the forecast values, there is an increase in milk and egg production. It is concluded that there will be increasing demand for milk and eggs as the population increases. Forecasting egg and milk production is crucial for ensuring food security and stabilizing market prices. It aids policymakers and farmers in planning and resource allocation, reducing the risk of supply-demand imbalances.

References

Amarender, A. (2018). Price Forecasting of Tomatoes. *International Journal of Vegetable Science*. 25(2):176-184.

Ashwini, S. D., Pokharkar, V.G., Gavali, A. V., and Yadav, D.B. (2015). Forecasting the prices of onion in Lasalgaon and Pimpalgaon market of Western Maharashtra. *International Journal of Tropical Agriculture*. 33(4):3563-3568.

Bannor, R.K. and Sharma, M. (2016). Modelling and forecasting wholesale groundnut prices in Bikaner district of Rajasthan for marketing intelligence. *Applied Research Journal*. 1(4):1-20.

Dareker, A. and Reddy, A.A. (2017). Forecasting of common paddy prices in India. *Journal of Rice Research*. 10(1):71-75.

Gurung, B., Panwar, S., Singh, K.N., Banerjee, R., Gurung, S.R., and Rathore, A. (2017). Wheat yield forecast using detrended yield over a sub-humid climatic environment in five districts of Uttar Pradesh, India. *Indian Journal of Agricultural Sciences*. 87(1):87- 91.

Jadhav, V., Reddy, B. V. C., and Gaddi, G. M. (2017). Application of Arima Model for Forecasting Agricultural Prices. *Journal Agriculture Science Technology*. 19(5): 981-992.

Poyyamozhi, S. and Mohideen, A. K. (2017). Forecasting of Cotton Production in India Using ARIMA Model. *Asia Pacific Journal of Research*. 18(1): 70-74.

Prakash, K. and Muniyandi, B. (2014). Application of ARIMA Model for Forecasting Production of Jasmine Flower in Madurai District of Tamil Nadu, India. *American International Journal of Research in Humanities, Arts and Social Sciences*. 6(3): 279-285.

Sahu, P.K., Mishra, P., Dhekale, B.S., Vishwajith, K.P. and Padmanaban, K. (2015). Modelling and Forecasting of area, production, yield and total seeds of rice and wheat in SAARC counties and the world towards food security. *American Journal of Applied Mathematics and Statistics*. 3(1): 34-48.