IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Milk Production Forecasting In India Using ARIMA And VAR Time Series Models

[1]MONICA. S , [2]KAVITHA. S , [3]SARANYA. V

[1]Ph.D Research Scholar, [2]Assistant Professor, [3]Ph.D Research Scholar
[1]Department of Statistics
[1]Periyar University, Salem-11, India

Abstract: Milk production is an essential component of India's agricultural farming system; despite the area's capacity for milk and dairy products, there is always a high demand for milk and milk products among the general public. From 128 million tonnes in 2011 to 463 million tonnes in 2040–41, milk output has surged. For many years, India has maintained its top spot in the production of milk. In India, the dairy industry is expanding at a 10% annual pace. However, no long-term studies have been conducted in the area to anticipate the volume of milk production. As a result, the purpose of this study The study's goal is to determine the best forecasting technique for milk production in order to have an impact on both future production sustainability and public policy. Secondary data were utilised in the study and were gathered from NDDB (1991 to 2022) and FAOSTAT (1961 to 2022). Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR) models were utilised after the stationarity of the data had been verified using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). According to the findings, ARIMA was shown to be a better model for (1, 1, 1) appropriate when using the SPSS programme to forecast milk. 463 million tonnes of milk are anticipated to be produced by 2041.

Index terms: **ARIMA, Milk production, Vector Autoregression (VAR), Forecasting**

I. Introduction

Dairy products like butter, cheese, and milk powder are all hydrated with milk. People constantly have a high demand for milk and milk-related goods. Since milk is the only complete meal in nature, it plays a crucial role in everyone's daily lives as well as those of their homes. It offers every critical vitamin needed for bodily development and growth. It is the primary protein source. Most nations enhance their production systems and methods to boost milk output in order to meet the demands of the population (Haenlein and Wendorf 2006). Today, computerised daily monitoring of cow performance is possible thanks to a number of technological advancements. According to DAHD&F (2014), the milk and milk product segment of the Indian dairy sector is worth Rs. 3.6 lakh crores. Furthermore, it keeps expanding at a 10% yearly pace. Approximately 730 million tonnes of milk were produced by dairy farms worldwide in 2011 using 260 million dairy cows (Food perspective, 2012). In order to estimate expected performance and compare forecast and actual performance, statistical models are needed when using automated milk yield recording systems for early disease diagnosis. According to Jacobs and Siegford (2012), earlier studies on modelling milk production in cows have mostly concentrated on fitting linear or nonlinear deterministic models to daily, weekly, or monthly milk measurements from lactations using partial or whole lactation data sets. To establish appropriate development strategies for the region, researchers and dairy development agents must have a thorough understanding of the current conditions (Tassew and Seifu 2009). The global dairy market is worth $187 billion. According to www.fao.org, India is the world's greatest producer and consumer of milk, followed by the USA, China, Russia, and Brazil. China, Algeria, Indonesia, Brazil, and Russia account for 86% of global imports of milk and milk products, while New Zealand, the European Union, Argentina, Australia, and the Philippines are the top exporters. 97% of the milk and milk product exports worldwide are from these five nations (FICCI, 2012). There are 54 million tonnes of cow milk, 66 million tonnes of buffalo milk, and five million tonnes of goat milk. Therefore, milk production is an essential component of the agricultural farming system in India. However, despite the area's favourable climate and potential for milk and dairy products, India's milk output is relatively modest in comparison to some of the dairy industries in tropical nations. . Village dairy cooperatives have over 14 million farmers as members (www.nddb.org). India now has a dairy industry that is self-sufficient, but that situation could change soon. The population is expanding steadily, and the demand for milk and milk products is rising significantly. Therefore, it is crucial to have a thorough understanding of past sector behaviour in order to design appropriate development strategies that fit the area. This will help to assess the advantages and disadvantages of previously implemented strategies and also provide the necessary framework for establishing future growth targets. This study's objectives were to analyse daily milk production trends, fit the right model, and predict milk production in the future based on previously collected data.

II. Materials and methods

2.1 Data Collections: Time series analysis refers to techniques or procedures that dissect a series into manageable portions and enable the identification of patterns as well as the creation of estimations and forecasts (Kantz and Schreiber 2004). Secondary data was used for the investigation. It was gathered between 1991–1992 and 2040–2041 from the National Dairy Development Board and the Food and Agriculture Organisation. These time series models include MA, AR, ARIMA and VAR models serve the majority of the time series data is stationary of this study.

2.2 Tools: The exponential smoothing approach, single equation regression models, simultaneous equation regression models, autoregressive integrated moving average models (ARIMA), and vector autoregression are the five main forecasting methods that are now accessible. For precise and reliable forecasting, ARIMA and VAR are the better options. These techniques are often employed (Chaudhari and Tingre, 2013; Pal et al., 2007).

2.3 ARIMA: Time series that have been distinguished by AR and MA models are referred to as autoregressive integrated moving averages. In an ARIMA (p, d, q) time series, p stands for the quantity of autoregressive terms (AR), d for the number of difference iterations the series requires to reach stationary behaviour (I), and q for the quantity of moving average terms (MA). It is often referred to as the "box Jenkins methodology". Sankar and Prabakaran (2012) used the autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) approaches to anticipate milk output in Tamil Nadu. ARIMA was utilised by Chaudhari and Tingre (2013) to anticipate milk output. Hossain and Hassan (2013) used cubic and linear models to anticipate milk, meat, and egg output in Bangladesh.

Auto Regressive Process of order (p) is,

$$y_t = \mu + \phi_1 y_{t-1} + \phi_1 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

Moving Average Process of order (q) is,

$$y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

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

$$y_t = \mu + \phi_1 y_{t-1} + \phi_1 y_{t-2} + \cdots + \phi_p y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

Where, Y_t is milk production, are independently and normally distributed with zero mean and constant variance for t = 1, 2,….., n; and $\phi_p$and θ_p are also estimated

There are five steps to follow

(i) MODEL DESCRIPTION : The autocorrelation function (ACF) and partial autocorrelation function (PACF) are initially used to determine if the data are stationary. Finding the starting values for the orders of the non-seasonal parameters p and q, which are found by searching for significant correlations in the ACF and PACF plots, is the next stage in the identification procedure.

(ii) ESTIMATION : In general, this computation is performed using the least squares approach, however occasionally we must use nonlinear (in parameter) estimate techniques. Software tools like SPSS and GRETL were employed for the study since they are readily available and simple to use.

(iii) DIAGNOSTIC CHECKING : The residuals from the fitted model are assessed for model adequacy, and alternative models are taken into consideration if appropriate. Other ARIMA models are explored until a good model fits the data if the initial found model seems to be insufficient. The optimal model is determined based on the minimal value of Akaike Information Criteria (AIC) provided by Makridakis et al. (1998). Different models are derived for various combinations of AR and MA separately and collectively.

$$AIC = -2 \log L + 2m \quad (4)$$

Where m = p+q and L be the likelihood function.

(iv) FORECASTING : We make a five-year prediction from 2024 to 2040 because if we project too far into the future, forecasting errors quickly rise VAR model is a multivariate time series data application of the univariate autoregression methodology. A multi-equation system called a VAR model treats all of its variables as endogenous. Every dependent variable is represented by a single equation. All dependent variables in the system are represented by lagged values on the right-hand side of each equation; contemporaneous variables are absent of VAR(p) model.

$$y_t = \propto + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t$$

(5)

ε_t = Zero mean (white noise) ; B_i = (I = 1,2,….,p) (n×n) coefficient matrices; ∝ = (n×1) vector of intercepts

III. RESULT AND DISCUSSION

3.1 Results of descriptive study variables

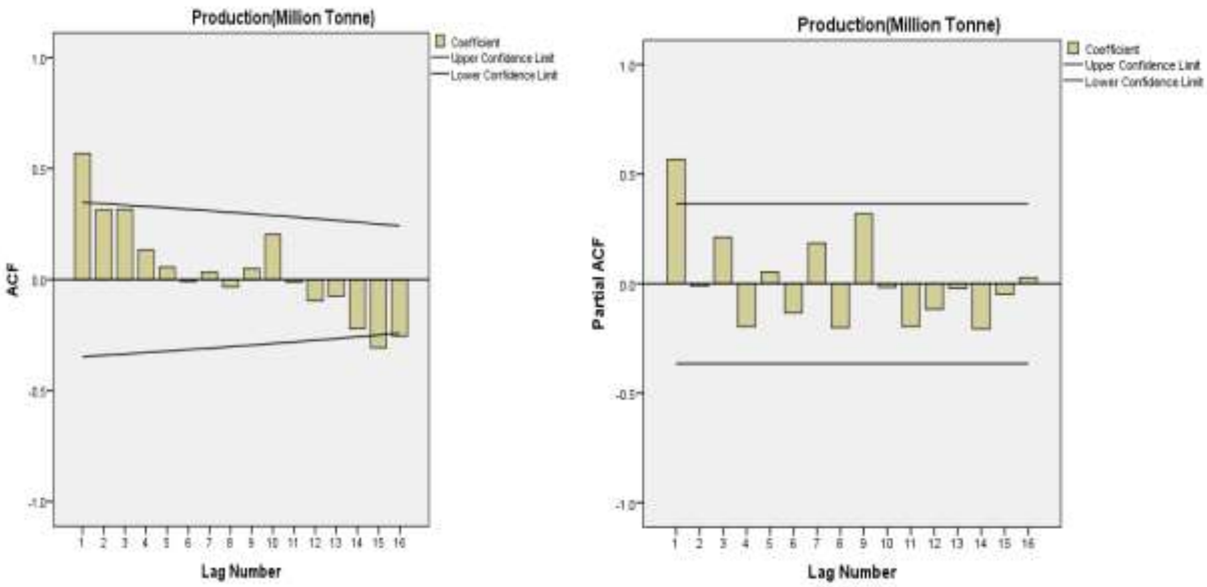
ACF and PACF are used to examine the stationarity of the data. The fact that the ACF and PACF values in Fig. 1 fall between the range of -0.5 and 0.5 shows that the data is stable. Based on the requirements of minimal MAPE and BIC values, all ARIMA models were assessed for accuracy of fit in SPSS software. All ARIMA models are included in Table 1 along with their MAPE and BIC values.

Table 3.1.1: Gretl and Spss of models

Model	Gretl		SPSS	
	AIC	SBC	SBC	MAPE
ARIMA (1,0,0)	32.031	34.899	1.568	1.780
ARIMA (1,1,0)	32.966	35.053	0.972	1.021
ARIMA (1,1,1)	49.053	46.484	**0.628**	**0.762**
ARIMA (0,1,1)	48.234	45.875	0.870	0 .781
ARIMA (0,0,1)	60.969	62.799	1.75	1.878
ARIMA (1,0,1)	NA	NA	-0.621	0.933
ARIMA (0,1,0)	NA	NA	-0.137	0.854

ARIMA (1, 1, 1) is taken into consideration for further study since it has the lowest MAPE (0.6) and BIC (0.1) values. As shown in Table 2, the additional diagnostic metrics R square (0.99) and RMSE (0.9) show that the model is well-fit and suitable for forecasting. The findings of the earlier research by Sankar and Prabakaran (2012) and Chaudhari and Tingre (2013) are different from those of our current study since their investigation revealed that ARIMA (1, 1, 0) was the most appropriate model. The findings of this investigation are consistent with studies by Pal, et al. (2007), which found that ARIMA (1,1,1) is the best model.

FIGURE 1: ACF and PACF values



GRETL is yet another field of study software. Additionally, in this instance, Akaike and Schwarz criteria were assessed on all ARIMA models. The minimal values of the required criterion, the Akaike and Schwarz criteria, were discovered to be ARIMA (1, 0, 0), and this led to additional analysis. Table 4 provides the projection from 2023 to 2041 and Fig. 3 shows it graphically.

Table 3.1.2 : Model fit of mean and statistics

Model Fit			
Fit Statistics	**Mean**	**Fit Statistics**	**Mean**
Stationary R-squared	**0.147**	**MaxAPE**	**2.698**
R-squared	**0.999**	**MAE**	**2.083**
MaxAE	**3.679**	**MAPE**	**0.736**
Normalized BIC	**0.189**	**RMSE**	**0.952**

The GRETL programme was used to perform the Vector Autoregression approach. The results reveal that the AIC is 3.1, the BIC is 3.2, the HQC is 3.1, and the R square is 0.99. This suggests that the model is well-fitted for further investigation. Table 5 and Figure 4 illustrate the forecast of milk production using the VAR technique from 2023 to 2041.

IV. Findings and Conclusion

The forecasting results vary greatly depending on the software programme employed. When using SPSS software, the ARIMA (1, 1, 1) model is best suited for milk production forecasting.

Table 4.1***:*** *Forecasting*

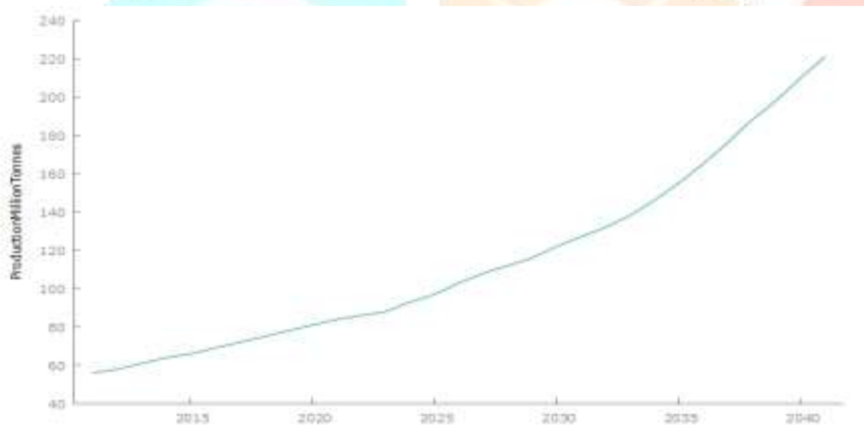
Year	***Production***	***Prediction***	***LCL Production***	***UCL Production***	***N residual***
2011	128				
2012	132	132.80	132	135	1
2013	138	137.97	136	139	-1
2014	146	145.88	140	146	0
2015	155	154.73	145	152	0
2016	165	165.04	150	159	0
2017	176	176.49	155	162	1
2018	188	188.07	161	168	-2
2019	198	197.83	167	175	-1
2020	210	209.600	173	179	0
2021	221	221.002	179	182	0
2022		233.652	186	188	1
2023		246.081	193	195	
2024		257.321	201	205	
2025		270.010	207	215	
2026		281.022	214	219	
2027		293.351	222	225	
2028		304.538	230	232	
2029		317.908	238	239	
2030		328.850	246	249	
2031		341.401	254	259	
2032		353.136	262	264	
2033		365.343	271	276	
2034		378.189	280	285	

2035		390.981	289	291	
2036		402.010	298	308	
2037		414.431	307	312	
2038		426.962	316	318	
2039		437.550	325	329	
2040		449.932	334	342	
2041		460.293	344	349	

If GRETL is utilised, ARIMA (1, 0, 0) is the best model to employ.

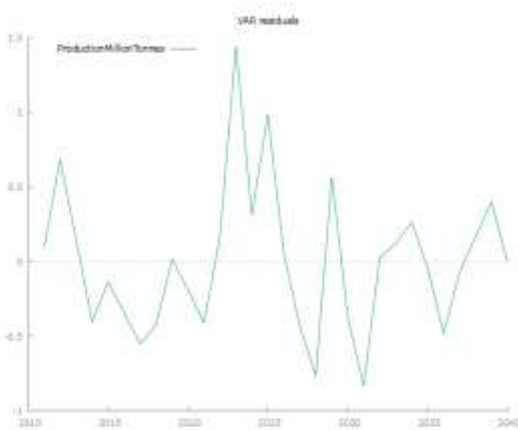
When compared to the other two softwares, SPSS forecast appears to be more accurate. Forecast numbers for the previous year are closer to actual milk production figures.

FIGURE 2: Forecasting ARIMA model using SPSS



The milk output for the year 2041 is anticipated to be 460 million tonnes using ARIMA (1,1,1) and SPSS, 462 million tonnes using ARIMA (1,0,0) and 463million tonnes using the VAR model.

FIGURE 3: Forecasting VAR for milk production using Gretl FIGURE 4: Forecasting ARIMA for milk Production using Gretl



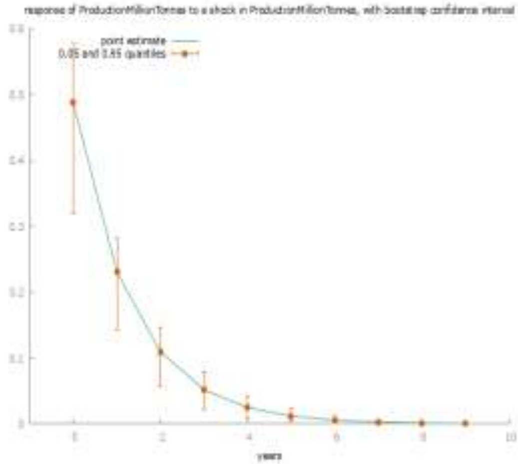


Table 4.2: milk forecasting by ARIMA using GRETL

Observation	Prediction	Standard error	95% (Confidence interval)
2023	246.081	0.4879	(242.9057, 249.9118)
2024	257.321	0.629	(253.3118, 261.3179)
2025	270.010	0.751a	(266.8471, 275.8532)
2026	281.022	0.801	(278.3983, 285.4044)
2027	293.351	0.864	(289.1324, 297.1385)
2028	304.538	0.998	(301.3397, 308.3458)
2029	317.908	1.428	(314.5470, 321.5531)
2030	328.850	1.876	(323.4263, 331.4324)
2031	341.401	1.952	(339.9775, 348.9836)
2032	353.136	2.134	(349.1848, 357.1909)
2033	365.343	2.465	(359.4080, 371.4142)
2034	378.189	2.875	(373.8301, 382.8362)
2035	390.981	2.933	(386.5482, 395.5543)
2036	402.010	3.105	(396.6859, 408.6920)
2037	414.431	3.196	(411.006, 419.013)
2038	426.962	3.354	(422.945, 431.952)
2039	437.550	3.496	(433.427, 441.433)
2040	449.932	3.854	(443.764, 452.770)
2041	463.293	3.912	(459.428, 469.434)

Table 4.3: milk forecasting by VAR

Observation	Prediction	Standard error	95% (Confidence interval)
2023	245.395	0.491	(243.0157, 289.7128)
2024	257.154	0.692	(253.2541, 261.1479)
2025	270.128	0.781	(266.4232, 276.0532)
2026	280.567	0.832	(277.2548, 286.0432)
2027	294.215	0.931	(289.4508, 296.3519)
2028	303.850	1.312	(302.7186, 309.5432)
2029	316.724	1.496	(315.0538, 322.5186)
2030	329.025	2.125	(322.7835, 331.4324)
2031	340.512	2.165	(340.4826, 348.6350)
2032	353.035	2.351	(348.3265, 357.2876)
2033	365.498	2.498	(359.0586, 371.3591)
2034	377.864	3.125	(374.3671, 382.2863)
2035	391.348	3.217	(387.0512, 396.3672)
2036	402.110	3.451	(397.8653, 407.9126)
2037	413.821	3.632	(411.0106, 418.0328)
2038	426.639	3.715	(423.5892, 432.2386)
2039	438.031	3.805	(433.1087, 441.3583)
2040	448.912	3.952	(442.614, 452.9106)
2041	462.965	4.020	(459.0496, 469.3124)

According to statistics, milk production is expanding at a CAGR of 3.61%, while milk consumption is growing at a CAGR of 6.28% (USDA GAIN report- Dairy and Indian Products Annual, 2013). Thus, there is a requirement for a quick action strategy to accelerate the production of milk to India. However, GOI is putting several programmes into place. comparable to the Intensive Dairy Development Programme, strengthening Infrastructure for the Production of Clean and High-Quality Milk.

REFERENCES

[1] Indian J. Dairy Sci. 66: 72-78. DAHD&F. (2014). Annual Report 2014-15, Department of Animal Husbandry and Fisheries,
Ministry of Agriculture, GOI, India

[2] Box G, Jenkins E (2015) Time series analysis: forecasting and control. Wiley, Hoboken

[3] Haenlein Y, Wendorf W (2006) Sheep milk. Handbook of milk of non-bovine mammals. Blackwell, Oxford, pp 137–194

[4] Makridakis S., Wheelwright S.C., Hyndman R. J. (1998). Forecasting - Methods and Applications. John Wiley and Sons: New York.

[5] Chaudhari D. J. and Tingre A. S. (2013). Forecasting of milk production in India: An application of ARIMA model.

[6] Hossain J. and Hassan F. (2013). Forecasting of Milk, Meat and Egg Production in Bangladesh. Research Journal of Animal,
Veterinary and Fishery Sciences 1: 7–13.

[7] Hipel KW, Mcleod AL, Lennox WC (1977) Advances in box-jenkins modeling: model construction. Water Resources 13:567–575

[8] Pal S., Ramasubramanian V., and Mehata S.C. (2007). Statistical Models for Forecasting Milk Production in India. J. Ind. Soc.
Agril. Statist. 61: 80–83.

[9] Beck N, Katz JN (2011) modeling dynamics in time-series–crosssection political economy data. Ann Rev Polit Sci 14:331–352

[10] Sankar T. J. and Prabakaran R. (2012). Forecasting milk production in Tamil Nadu. International Multidisciplinary Research
Journal. 2: 10–15.

[11] http://en.wikipedia.org/wiki/Milk#cite_ref-2 as retrieved on 22/05/2014

[12] http://www.fao.org/agriculture/dairy-gateway/milk-production/en/#.VqdWj_l97IV

[13] http://www.dairyco.org.uk/market-information/farming-data/cow-numbers/world-cow-numbers/ retrieved 23 Mar 2014

[14] http://www.nddb.org/English/Genesis/Pages/Genesis.aspx as retrieved on 22/05/2014