

Stochastic Time Series Analysis for Milk Production in Tamil Nadu

P. Pushpa and T. Jai Sankar

Abstract

"This study presents a stochastic time series analysis for milk production in Tamil Nadu from 1978 to 2022. The increasing demand for milk and dairy products necessitates a thorough understanding of production trends and patterns in the region. Through rigorous data collection and preprocessing, this study identifies non-seasonal components influencing milk production. ARIMA (0,1,2) model is selected on the basis of its optimal fit, as determined by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model captures the underlying stochastic processes governing production fluctuations and provides reliable forecasts for future milk yield. The result indicates a significant upward trend in milk production, with periodic fluctuations that align with climatic and economic factors. The findings highlight the importance of strategic planning in dairy production to meet growing demands and ensure sustainable growth in Tamil Nadu's dairy sector. This research contributes valuable insights for policymakers and stakeholders aiming to enhance milk production efficiency and food security in the region."

Keywords: ARIMA, Forecasting, Box-Ljung, BIC, Milk Production.

Introduction

"Milk production in Tamil Nadu plays a pivotal role in the state's economy and cultural landscape. As one of India's leading milk-producing states, Tamil Nadu contributes significantly to the national dairy output. The region's favorable climate, combined with a rich tradition of livestock rearing, has facilitated the establishment of a robust dairy sector. Over the years, milk production in Tamil Nadu has evolved from traditional practices to a more organized and commercially viable industry. The state's commitment to improving dairy farming practices, enhancing breeding techniques, and implementing innovative technologies has led to increased milk yields and better quality products. Additionally, the establishment of cooperative societies, such as the Tamil Nadu Cooperative Milk Producers' Federation (Aavin) has empowered local farmers and improved their livelihoods. The demand for milk and dairy products has surged due to

population growth, urbanization, and rising health consciousness among consumers. This trend is further fueled by the growing middle class, which has led to increased consumption of dairy products like cheese, yogurt, and ice cream. Consequently, understanding the dynamics of milk production in Tamil Nadu is essential for meeting future demands and ensuring food security. Through a stochastic time series analysis, this study aims to uncover the patterns and factors influencing milk production, providing valuable insights for policymakers and stakeholders in the industry. Milk is rich in several essential nutrients that contribute to overall health. It is an excellent source of calcium, which is vital for maintaining strong bones and teeth. The proteins in milk support muscle growth and repair, while its healthy fats provide energy and aid in the absorption of fat-soluble vitamins (Figure 1)."

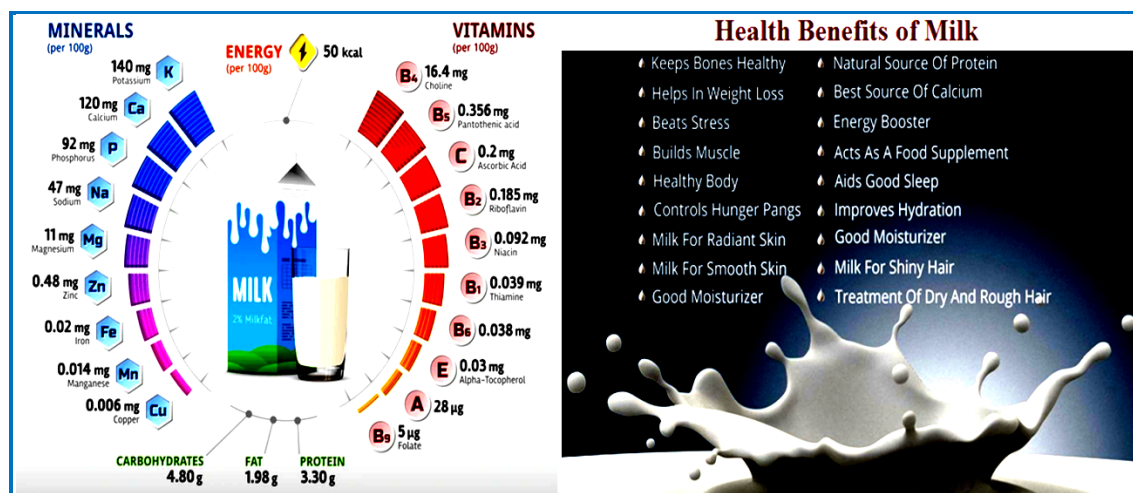


Figure 1 - Minerals, Vitamins and Health Benefits of Milk.

"Milk production plays a crucial role in providing essential nutrition, contributing to food security and the global economy. As a rich source of calcium, proteins, vitamins, and minerals, milk supports health and development, making it a staple in diets worldwide. Economically, dairy farming provides income and employment, particularly in rural areas, supporting millions of farmers. It also contributes to agricultural sustainability by utilizing crop residues as animal feed and generating organic fertilizers. This makes milk production important not only for individual health but also for broader economic and agricultural systems.

Material and Methods

The current behavior of the variable under investigation is described by the autoregressive integrated moving average (ARIMA) model in terms of linear

relationships with its historical values. All that is needed for this extrapolation method is historical time series data for the variable being studied. The main purpose of ARIMA models is to forecast the associated variable. ARIMA residual autocorrelations were measured by Box and Pierce (1970) [1]. Slutzky (1973) applied Moving Average (MA) models [2]. As described by Akaike (1983), the stationary time series is defined as being bounded by the same integer [3]. A model based on ARIMA (1,1,0) was developed by Jai Sankar and Prabakaran (2012) for forecasting milk production in Tamil Nadu for the period from 1978 to 2008 [4]. According to Jai Sankar and Pushpa (2019), *Saccharum officinarum* production in India was validated into the period 1950-2017 and ARIMA (2,1,0) model was applied up to 2022 [5]. Jai Sankar and Pushpa (2020) considered ARIMA (0,1,1) model for stochastic forecasting analysis for *Arachis hypogaea* production in India during the years 1950-2017 [6]. Forecasting of milk production in India with ARIMA and VAR time series models from 1991 to 2012, Sagar Surendra Deshmukh and Paramasivam (2016) developed an ARIMA (1,1,1) model, which forecasts through up to 2017 years [7]. Yashavanth Basavapatna Subbanna et al. (2021) found ARIMA (1,2,1) model for forecasting of buffalo milk production in India through time series approach [8]. Dipriya R. Lyngkholi et al. (2022) considered trend analysis of milk production in India during the period of 1989 to 2019 [9]. Vishnu Shanka et al. (2023) applied ARIMA (1,1,2) modeling and forecasting of milk production in the Western Zone of Tamil Nadu. [10]. Nilay Adhikari et al. (2024) considered ARIMA (2,1,0) model for growth, instability and forecasting of milk production in Assam through ARIMA applications during the years from 1995-96 to 2020-21 [11]. The ARIMA (1,1,1) model was used by Monica et al. (2024) for milk production forecasting in India using ARIMA and VAR time series models for the period from 1961 to 2022 [12]. Jai Sankar and Pushpa (2023) calculated ARIMA (0,1,2) model for implementation of stochastic time series forecasting ARIMA model for *Hordeum vulgare* production in India during the years from 1960 to 2020 [13]. Statistically independent and normally distributed residuals were significant features of stochastic time-series ARIMA models (Alan Pankratz, 1983) which were widely used to analyze time series data [14].

"In this study, a four-step ARIMA model was used, consisting of identification, estimation, diagnostic checking, and forecasting. Model parameters were considered to fit the ARIMA models.

AR process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$;

MA process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

ARIMA process of order (p, d, q) is,

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

where Y_t - Milk production, ε_t 's - independently and normally distributed with zero mean and constant variance σ^2 for $t = 1, 2, \dots, n$; d - the fraction differenced while interpreting AR and MA, and ϕ 's and θ 's - coefficients to be valued.

Trend Fitting: The Box-Ljung Q statistics was used to convert the non-stationary data into stationarity data and also to validate the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, a range of reliability statistics like R squared, Stationary R squared, RMSE, MAPE and BIC were applied. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right]^{1/2} \text{ and } MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

$$BIC(p, q) = \ln v^*(p, q) + (p + q) \left[\frac{\ln(n)}{n} \right]$$

where p and q - order of AR and MA processes; n - number of observations; and v^* - approximate of white noise variance σ^2 .

$$Q = \frac{n(n+2) \sum_{i=1}^k rk^2}{(n-k)}$$

where n - number of residuals and rk - residuals autocorrelation at lag k.

In this analysis, the data on milk production in Tamil Nadu were collected from the Directorate of Animal Husbandry and Veterinary Services, Government of Tamil Nadu for the period from 1978 to 2022 (Table 1) and were applied to fit the ARIMA model to predict the future production".

Table 1 - Actual Milk Production ('000 metric tonnes) in Tamil Nadu

Year	Production	Year	Production	Year	Production
1978	1681	1993	3524	2008	6651
1979	1727	1994	3694	2009	6787
1980	1860	1995	3791	2010	6831
1981	1738	1996	3977	2011	6968
1982	1886	1997	4061	2012	7005
1983	1788	1998	4273	2013	7049
1984	2562	1999	4574	2014	7132
1985	2846	2000	4907	2015	7244

1986	3118	2001	4988	2016	7556
1987	3109	2002	4622	2017	7742
1988	3238	2003	4752	2018	8362
1989	3410	2004	4784	2019	8759
1990	3375	2005	5474	2020	9790
1991	3357	2006	6276	2021	10107
1992	3468	2007	6540	2022	10317

Results and Discussion

"In this analysis, to fit an ARIMA model, the process for any variable involves four steps: identification, estimation, diagnostic and forecasting. ARIMA (p,d,q) is steady to make certain stationarity through reading the graph or time plot of the given data. Figure 3 suggests that the data is non-stationary.

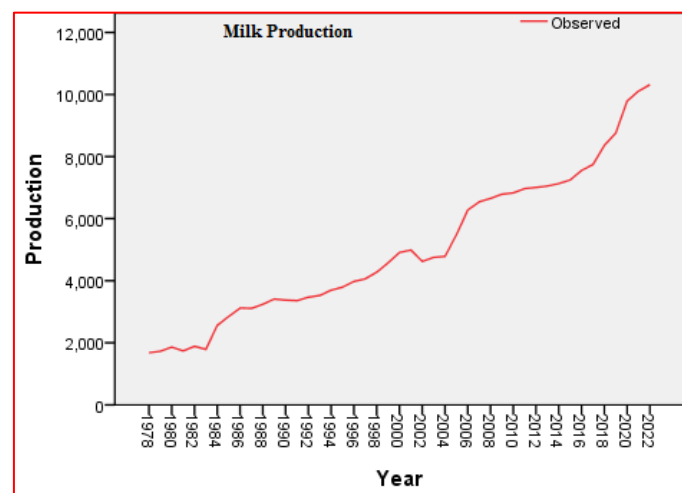


Figure 2 - Time plot for milk production

The autocorrelation and partial autocorrelation coefficients of various orders of Y_t are calculated (Table 2). The graphs of ACF and PACF are produced (Figure3). The models and corresponding BIC values are specified (Table 3). The value of normalized BIC is 11.260 and R squared value is 0.991 in the most appropriate model for milk production is ARIMA (0,1,2) as this model has the lowest BIC value.

Table 2 - ACF and PACF of Milk Production

Lag	AC	Std. Error (white noise)	Box-Ljung Statistic	PAC	Std. Error	Lag	AC	Std. Error (white noise)	Box-Ljung Statistic	PAC	Std. Error
	Value	Df	Sig. (Chi-Square Approx.)	Value	Df		Value	Df	Sig. (Chi-Square Approx.)	Value	Df
1	0.914	0.144	40.156	0.914	0.149	17	-0.042	0.115	209.164	-0.037	0.149
2	0.823	0.143	73.476	-0.075	0.149	18	-0.085	0.113	209.723	0.028	0.149
3	0.732	0.141	100.442	-0.051	0.149	19	-0.111	0.111	210.718	0.039	0.149
4	0.657	0.139	122.694	0.047	0.149	20	-0.138	0.109	212.331	-0.060	0.149

5	0.588	0.138	140.975	-0.013	0.149	21	-0.164	0.107	214.713	-0.027	0.149
6	0.527	0.136	156.016	-0.002	0.149	22	-0.201	0.104	218.417	-0.089	0.149
7	0.475	0.134	168.592	0.026	0.149	23	-0.236	0.102	223.759	-0.035	0.149
8	0.431	0.132	179.192	0.005	0.149	24	-0.266	0.100	230.866	-0.003	0.149
9	0.388	0.130	188.059	-0.014	0.149	25	-0.297	0.097	240.171	-0.052	0.149
10	0.344	0.129	195.212	-0.036	0.149	26	-0.323	0.095	251.812	-0.020	0.149
11	0.298	0.127	200.721	-0.036	0.149	27	-0.350	0.092	266.216	-0.048	0.149
12	0.249	0.125	204.691	-0.041	0.149	28	-0.366	0.090	282.837	0.011	0.149
13	0.197	0.123	207.247	-0.055	0.149	29	-0.369	0.087	300.860	0.026	0.149
14	0.139	0.121	208.569	-0.070	0.149	30	-0.367	0.084	319.854	-0.008	0.149
15	0.079	0.119	209.013	-0.061	0.149	31	-0.363	0.081	339.798	-0.023	0.149
16	0.016	0.117	209.033	-0.076	0.149	32	-0.357	0.078	360.568	-0.012	0.149

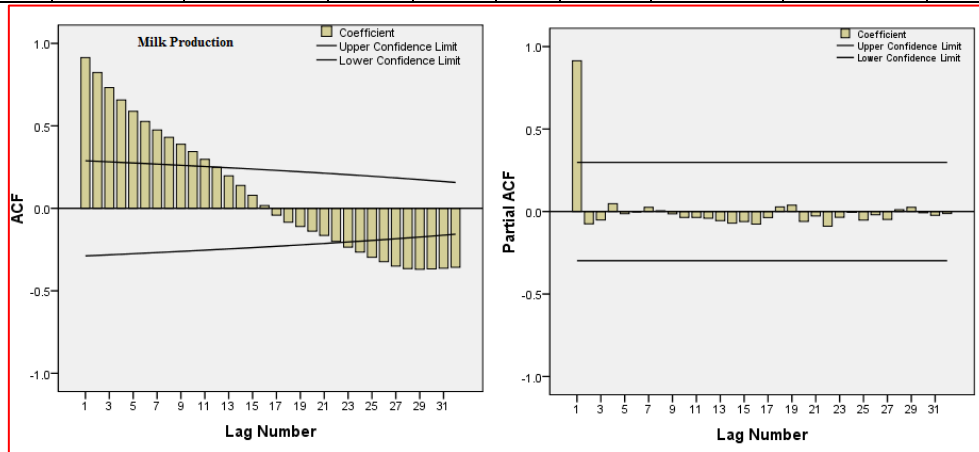


Figure 3 - ACF and PACF of differenced data

Model Estimation: Model parameters were found and accounted (Table 4 and Table 5). The model verification is concerned with examining the residuals of the model to progress on the chosen ARIMA (p,d,q). This is done through validating the autocorrelations and partial autocorrelations of the residuals of various orders, up to 32 lags were considered and the same along with their significance which is checked by Box-Ljung test are given (Table 6). This proves that the chosen ARIMA model is a suitable model.

Table 3 - BIC values of ARIMA(p,d,q)

ARIMA (p,d,q)	BIC Values
0,1,1	11.262
0,1,2	11.260
1,1,0	11.262
1,1,1	11.358
1,1,2	11.352
2,1,0	11.354
2,1,1	11.426

2,1,2	11.439
3,1,0	11.291
3,1,1	11.282
3,1,2	11.378

Table 4 - Estimated MA Model of Milk Production

	Estimate	SE	t	Sig.
Constant	-9854.811	9641.377	-1.022	0.313
MA1	-0.368	0.151	-2.431	0.020
MA2	-0.440	0.156	-2.815	0.008

The ACF and PACF of the residuals are specified (Figure 4) and also indicated 'good fit' for the selected ARIMA model of the milk production data is

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

$$Y_t = -9854.811 + 0.368\varepsilon_{t-1} + 0.440\varepsilon_{t-2} + \varepsilon_t$$

Table 5 - Estimated ARIMA Model Fit Statistics

ARIMA (p,d,q)	Stationary R ²	R ²	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC
0,1,1	0.125	0.990	245.167	3.832	27.748	164.600	713.897	11.262
0,1,2	0.219	0.991	234.623	3.725	26.335	162.952	674.693	11.260
1,1,0	0.125	0.990	245.167	3.832	27.748	164.600	713.897	11.262
1,1,1	0.138	0.990	246.462	3.806	28.022	164.299	717.933	11.358
1,1,2	0.234	0.991	235.331	3.590	25.402	159.329	650.791	11.352
2,1,0	0.142	0.990	245.863	3.771	27.738	164.039	710.659	11.354
2,1,1	0.175	0.990	244.194	3.665	26.322	161.378	674.375	11.426
2,1,2	0.252	0.991	235.461	3.536	25.451	158.116	652.052	11.439
3,1,0	0.279	0.992	228.207	3.677	23.796	158.608	623.382	11.291
3,1,1	0.361	0.993	217.649	3.374	24.488	139.642	627.394	11.282
3,1,2	0.371	0.993	218.796	3.395	24.262	141.552	621.600	11.378

Table 6 - Residual of ACF and PACF of Milk Production

Lag	ACF		PACF		Lag	ACF		PACF	
	Mean	SE	Mean	SE		Mean	SE	Mean	SE
1	-0.065	0.151	-0.065	0.151	17	-0.027	0.172	0.007	0.151
2	-0.076	0.151	-0.080	0.151	18	-0.288	0.172	-0.283	0.151
3	-0.205	0.152	-0.217	0.151	19	0.010	0.183	-0.159	0.151
4	-0.089	0.158	-0.136	0.151	20	0.084	0.183	-0.059	0.151
5	-0.011	0.160	-0.076	0.151	21	0.255	0.183	0.147	0.151
6	-0.024	0.160	-0.110	0.151	22	0.094	0.191	0.131	0.151
7	-0.039	0.160	-0.126	0.151	23	-0.132	0.192	-0.052	0.151
8	-0.066	0.160	-0.152	0.151	24	-0.031	0.194	0.076	0.151
9	-0.045	0.160	-0.158	0.151	25	-0.011	0.195	0.024	0.151
10	0.002	0.161	-0.139	0.151	26	-0.012	0.195	-0.104	0.151
11	0.070	0.161	-0.079	0.151	27	-0.025	0.195	-0.020	0.151
12	0.048	0.161	-0.082	0.151	28	-0.037	0.195	-0.030	0.151
13	-0.061	0.162	-0.184	0.151	29	-0.038	0.195	-0.096	0.151
14	0.165	0.162	0.079	0.151	30	-0.070	0.195	-0.045	0.151
15	0.141	0.166	0.147	0.151	31	0.028	0.196	-0.012	0.151
16	-0.152	0.169	-0.158	0.151	32	0.070	0.196	0.049	0.151

The forecasted value of milk production (quantity in '000 metric tonnes) for the years 2023 through 2030 respectively is given by 10429.48, 10617.91, 10929.80,

11246.71, 11568.64, 11895.60, 12227.57 and 12564.56 in Table 7. We calculated significant measures of the forecasts' accuracy for the sample period in order to evaluate the fit of an ARIMA (p,d,q) model. This measure shows that the forecasting inaccuracy is low. Figure 5 indicates that the actual and forecasted value of milk production data with 95% confidence limits".

Table 7 - Forecast of Milk Production

Year	Predicted	LCL	UCL
2023	10429.48	9955.92	10903.03
2024	10617.91	9815.55	11420.27
2025	10929.80	9756.37	12103.23
2026	11246.71	9794.09	12699.33
2027	11568.64	9882.44	13254.84
2028	11895.60	10004.45	13786.74
2029	12227.57	10151.61	14303.53
2030	12564.56	10318.95	14810.17

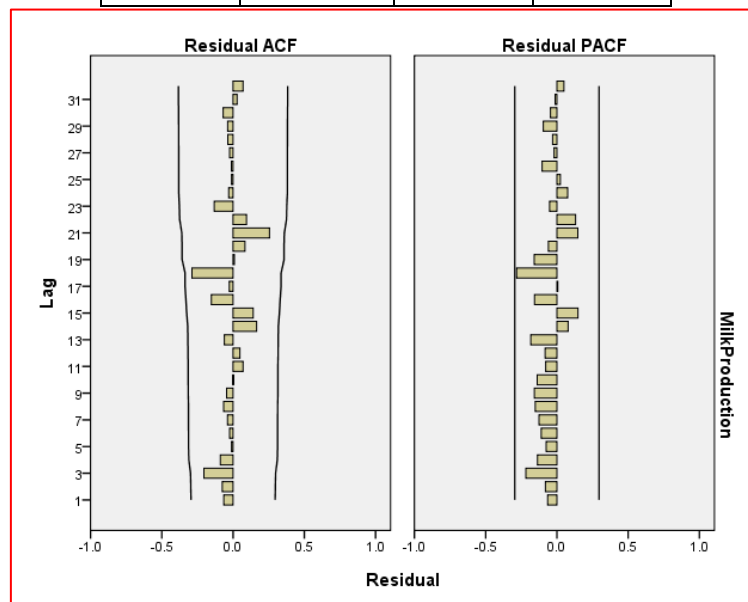


Figure 4 - Residuals of ACF and PACF

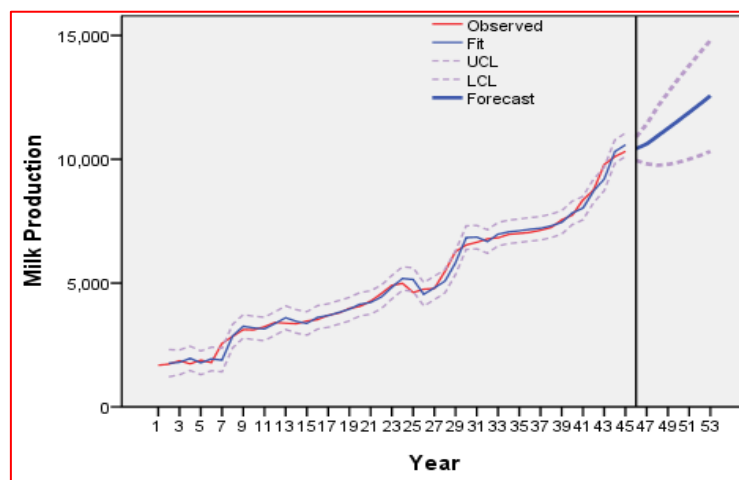


Figure 5 - Actual and Estimate of Milk Production

Conclusion

"The most suitable ARIMA model for forecasting milk production was determined to be ARIMA (0,1,2). Based on time series data from 1978 to 2022, the forecast indicates that milk production in India will rise from 10,317 metric tonnes in 2022 to 12,565 metric tonnes by 2030. This significant projected growth highlights the steady increase in milk production over the years and underscores the importance of the dairy sector in the country's economy. The results of this study not only provide valuable insights into future production trends but also serve as a critical tool for policymakers. By utilizing this forecast, decision-makers can develop targeted strategies to address the growing demand for milk, enhance infrastructure, and support farmers to boost production. Furthermore, the study offers a foundation for creating long-term plans aimed at strengthening India's dairy industry, ensuring sustainability, and securing the livelihoods of millions involved in dairy farming across the country. These forecasts can play a key role in shaping future policies to meet both domestic consumption needs and export potential".

Acknowledgements

"This research was supported by St. Peter's Institute of Higher Education and Research, India under the seed money for research project. The authors thank the reviewer and editor for their valuable suggestions for improving this research article".

References

- [1] G.E.P. Box and D.A. Pierce, *Distribution of Residual Autocorrelations in ARIMA Models*, J. American Stat. Assoc., **65**, 1509-1526, (1970).
- [2] E. Slutsky, *The summation of random causes as the source of cyclic processes*, Econometrica, **5**, 105-146, (1973).
- [3] H. Akaike, *Statistical Predictor Identification*, Annals of Institute of Statistical Mathematics, **22**, 203-270, (1983).
- [4] T. Jai Sankar and R. Prabakaran, *Forecasting milk production in Tamil Nadu*, International Multidisciplinary Research Journal, **2**(1), 10-15, (2012).
- [5] T. Jai Sankar and P. Pushpa, *Design and development of time series analysis for Saccharum officinarum production in India*, A Journal of Composition Theory, **12**(9), 203-211, (2019).

- [6] T. Jai Sankar and P. Pushpa, *Stochastic forecasting analysis for peanut (Arachis hypogaea) production in India*, Journal of Critical Reviews, 7(12), 2394-5125, (2020).
- [7] Sagar Surendra Deshmukh and Paramasivam, *Forecasting of milk production in India with ARIMA and VAR time series models*. Asian Journal of Dairy and Food Research, 35(1), 17-22, (2016).
- [8] Yashavanth Basavapatna Subbanna, Sanjiv Kumar and Sharath Kumar Maddur Puttaraju, *Model to forecasting buffalo milk production in India: Time Series Approach*, Buffalo Bulletin, 40(2), 235-243, (2021).
- [9] Dipriya R. Lyngkhai, S. Basanta Singh, Ram Singh and Hehlangki Tyngkan, *Trend analysis of milk production in India*. Asian Journal of Dairy and Food Research, 41(2), 183-187, (2022).
- [10] S.Vishnu Shanka, R. Ajaykumar, S. Ananthakrishnan, A. Aravinth Kumar, K. Hari Shankar, T. Sakthiselvi and C. Navinkumar, *Modeling and forecasting of milk production in the Western Zone of Tamil Nadu*. Asian Journal of Dairy and Food Research, 1(1), 1-6, (2023).
- [11] Nilay Adhikari, Ram Singh, Th. Onchoila Maring and P. Harigovind, *Model for growth, stability and forecasting of milk production in Assam: ARIMA application*, International Journal of Agriculture Extension and Social Development, 7(6), 39-45, (2024).
- [12] S. Monica, S. Kavitha and V. Saranya, *Milk production forecasting in India using ARIMA and VAR time series models*, Jilin Daxue Xuebao (Gongxueban) / Journal of Jilin University (Engineering and Technology Edition), 43(9), 99-107, (2024).
- [13] T. Jai Sankar and P. Pushpa, *Implementation of stochastic time series forecasting ARIMA model for Hordeum vulgare production in India*. International Journal of Agricultural and Statistical Sciences, 19(1), 133-139, (2023).
- [14] Alan Pankratz, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, John Wiley & Sons, New York, (1983).

Author information

P. Pushpa, Department of Mathematics, St. Peter's Institute of Higher Education and Research, Avadi, Chennai, Tamil Nadu, India.

E-Mail: pushpastats@gmail.com

T. Jai Sankar, Department of Statistics, Bharathidasan University, Tiruchirappalli,
Tamil Nadu, India.

E-Mail: tjaisankar@gmail.com

Corresponding author: P. Pushpa