JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ARIMA Modeling for Leguminosae Production in Tamil Nadu

[1]Dr. T. Jai Sankar, [2]Ms. S. Sivasankari, [3]Ms. J. Lizy Perpetua

[1]Assistant Professor, [2]Research Scholar, [3]Research Scholar
Department of Statistics
Bharathidasan University, Tiruchirappalli – 620 023, India.

Abstract

This paper analyzes and determines the ARIMA modeling for *Leguminosae* production (Minor Pulses) in Tamil Nadu during the years from 1966 to 2017. The minor pulses are grown in limited areas in dry and hot conditions of semi-arid regions of tropics in South Asia, Africa and Mediterranean regions also play a great role in supplementing the diet of the resource poor farmers and consumers. These crops are usually grown in marginal lands with low inputs. Per capita availability of pulses in Tamil Nadu is lower than the recommended quantity. This study considers autoregressive (AR), moving average (MA) and ARIMA processes to select the appropriate ARIMA model for *Leguminosae* production in Tamil Nadu. ARIMA (p, d, q) and its components autocorrelation function (ACF), partial autocorrelation function (PACF), root mean square error (RMSE), mean absolute percentage error (MAPE), normalized BIC and ARIMA (2,1,0). Based on the selected model, minor pulses production in Tamil Nadu is projected to increase from 41.482 thousand tonnes in 2017 to 55.714 thousand tonnes in 2027.

Index Terms: ARIMA, BIC,PACF, Forecasting, MAPE, Minor Pulses, RMSE.

I. Introduction

Leguminosae are cultivated on a small scale by economically poor farming communities in the developing and less developed countries for their personal food, and only rarely grains are sold in the local market in the case of surplus stock. Drought resistant and requiring little fertilizer and care, these crops are under-utilized or neglected, although they are reasonable sources of protein and can increase food security in rural areas. In addition, minor pulses (Horse gram, moth bean, rice bean shown in Figure 1) are used to meet the nutritional requirement of a large section of the human population. In developing and less developed countries where people suffer from protein deficiency, the per capita consumption of pulses is decreasing because of the increasing human population, fluctuating market prices, and low crop productivity. To reverse this trend, increased research on pulses production and management is needed to improve food security and nutrition.

Figure 1. Varieties of Minor Pulses

Explicit management practices and economic threshold levels are only developed for major pests, which regularly cause economic losses and are not available for minor pulses. Here, we discuss the importance of minor pulses. Insects generally do not damage other minor pulses or plants may escape attack due to a short growth period and inherent resistance to pest attack. In general, minor pulses are cultivated in two cropping systems, as intercrop in the long duration main crops and as second crop on residual soil water after the harvest of long season crops in different geographic locations. For example, green gram, black gram, horse gram, and moth bean in the Indian subcontinent; fava bean/broad bean in West Africa; lentil in South West Asia; and rice bean and grass pea in Asia and East Africa. In this Figure 2 are explain the benefits of Legumes in human health. We consolidate scattered information on pest management in minor pulses crops and discuss how practices can be improved to enhance production of minor pulses and enhance their food quality.

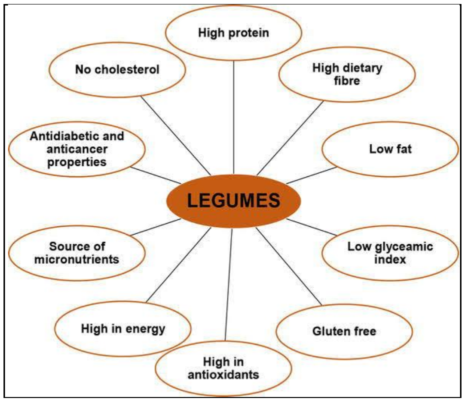

Figure 2. Benefits of Legumes

II. Material and Methods

The aim of the study was to design and develop ARIMA modelling for *Leguminosae* (Minor pulses) productions in Tamil Nadu, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1976), was frequently applied for discovering the pattern and predicting the future values of the time series data. Box and Pierce (1970) measured the distribution of residual autocorrelations in ARIMA. Akaike (1970) found the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were applied by Slutzky (1973). Jai Sankar et al. (2010) also applied ARIMA (1,1,0) Model for stochastic modeling for cattle production and forecast the yearly production of cattle in Tamil Nadu during the period of 1970 to 2010. Jai Sankar (2011) used a Stochastic model approach to model and forecast fish product export in Tamil Nadu during the period of 1969 to 2008. Jai Sankar et al (2011) applied ARIMA (1,1,0) model to Stochastic for Bovine Production Forecasting in Tamil Nadu during the years from 1970 to 2008. Jai Sankar and Prabakaran (2012) applied ARIMA (1,1,0) model to model and forecast milk production in Tamil Nadu during the period of 1978 to 2008. Jai Sankar (2014) also used ARIMA (0,1,1) model for design of a Stochastic Modeling for Egg Production Forecasting in Tamil Nadu during the years from 1996 to 2008. Jai Sankar and Pushpa (2019) also applied the ARIMA (2,1,0) model for Design and Development of Time Series Analysis for *Saccharum officinarum* Production in India during the years from 1950 to 2017. Jai Sankar and Pushpa (2022) applied ARIMA (1,1,0) model for *Solanum tuberosum* production in India during the years from 1950 to 2018. Jai Sankar and Pushpa (2022) used ARIMA (0,1,1) model for Bajra *Pennisetum glaucum* production in India during the years from 1951 to 2018. Jai Sankar and Pushpa (2022) used ARIMA (0,1,1) model for *Musa paradisiaca Linn* production in India during the years from 1961 to 2019.

Description of the model

The time series when differenced follows both AR and MA models and is known as ARIMA model. Hence, ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

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

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

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, Y_t is *Leguminosae* production, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1, 2,, n; d is the fraction difference while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting: The Box-Ljung Q statistics was used to transform the non-stationary data into stationarity data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, RMSE, MAPE, and BIC as suggested by Schwarz Gideon (1978) were used as below:

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right]^{1/2} ; \quad MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Y_i - \hat{Y}_i)}{Y_i}\right| \quad \text{and} \quad BIC(p,q) = \ln v^*(p,q) + (p+q)\,[\ln(n)/n]$$

where, p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

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

where, n is the number of residuals and rk is the residuals autocorrelation at lag k.

III. Results and Discussion

In this study, data were collected from the International Crops Research Institute for the Semi-Arid Tropics of the *Leguminosae* (minor pulses) production, Government of India for the period 1966 to 2017 (Table 1) and applied to ARIMA. A model for predicting and forecasting future production of *Leguminosae*.

Table 1: Actual *Leguminosae* Production (1000 tons) in Tamil Nadu.

Year	Production	*Year*	Production	*Year*	Production	*Year*	Production
1966	7.041	*1979*	10.598	*1992*	22.895	*2005*	12.718
1967	6.253	*1980*	13.917	*1993*	26.833	*2006*	22.114
1968	8.289	*1981*	10.020	*1994*	21.360	*2007*	13.273
1969	7.189	*1982*	12.028	*1995*	13.779	*2008*	12.199
1970	-1.000	*1983*	12.915	*1996*	14.907	*2009*	14.963
1971	12.673	*1984*	14.799	*1997*	17.047	*2010*	18.051
1972	12.548	*1985*	15.985	*1998*	20.850	*2011*	0.000
1973	12.623	*1986*	17.678	*1999*	20.209	*2012*	14.578
1974	6.544	*1987*	16.469	*2000*	21.997	*2013*	45.863
1975	6.723	*1988*	13.106	*2001*	18.732	*2014*	63.559
1976	10.702	*1989*	20.038	*2002*	14.354	*2015*	48.483
1977	12.901	*1990*	23.748	*2003*	14.103	*2016*	21.088
1978	13.805	*1991*	24.218	*2004*	15.294	*2017*	41.482

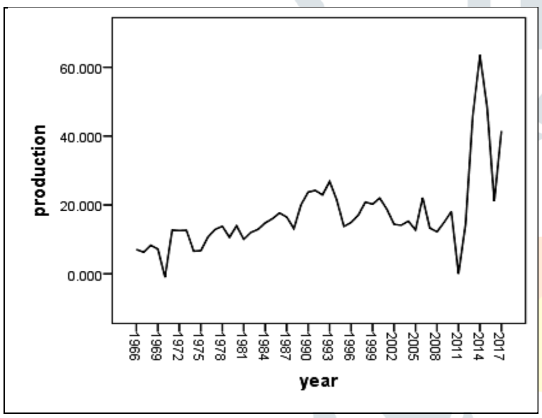

Figure 3. Time plot of *Leguminosae* production

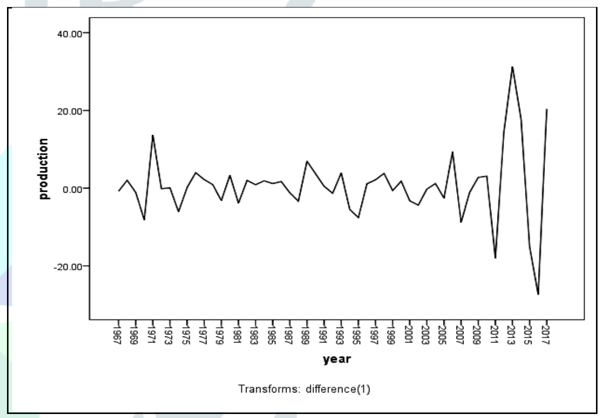

Figure 4. Time plot of *Leguminosae* production with d = 1

Figure 3 depicts that the data used were non-stationary. The time series plot of *Leguminosae* minor pulses production from 1966 to 2017 with d=1 was created as shown in Figure 4 and it is also non-stationary. The newly constructed variable Y_t could now be examined for stationarity. Since, Y_t was stationary in mean, the next step was to identify the values of p and q. For this, the ACF and PACF of various orders of Y_t were computed and presented in Table 2 and Figure 5.

Table 2: ACF and PACF of *Leguminosae* Production

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic			Partial Autocorrelation	Std. Error
			Value	df	Sig.b		
1	0.640	0.135	22.554	1	0.000	0.640	0.139
2	0.342	0.133	29.103	2	0.000	-0.115	0.139
3	0.201	0.132	31.422	3	0.000	0.054	0.139
4	0.155	0.131	32.820	4	0.000	0.048	0.139
5	0.037	0.129	32.901	5	0.000	-0.142	0.139
6	-0.012	0.128	32.909	6	0.000	0.042	0.139
7	0.084	0.127	33.349	7	0.000	0.175	0.139
8	0.081	0.125	33.770	8	0.000	-0.104	0.139
9	0.034	0.124	33.846	9	0.000	0.004	0.139
10	-0.033	0.122	33.920	10	0.000	-0.070	0.139
11	-0.020	0.121	33.949	11	0.000	0.017	0.139
12	-0.035	0.119	34.035	12	0.001	-0.018	0.139
13	0.028	0.118	34.092	13	0.001	0.150	0.139
14	0.049	0.116	34.268	14	0.002	-0.059	0.139
15	0.047	0.115	34.437	15	0.003	-0.005	0.139
16	0.033	0.113	34.524	16	0.005	0.003	0.139

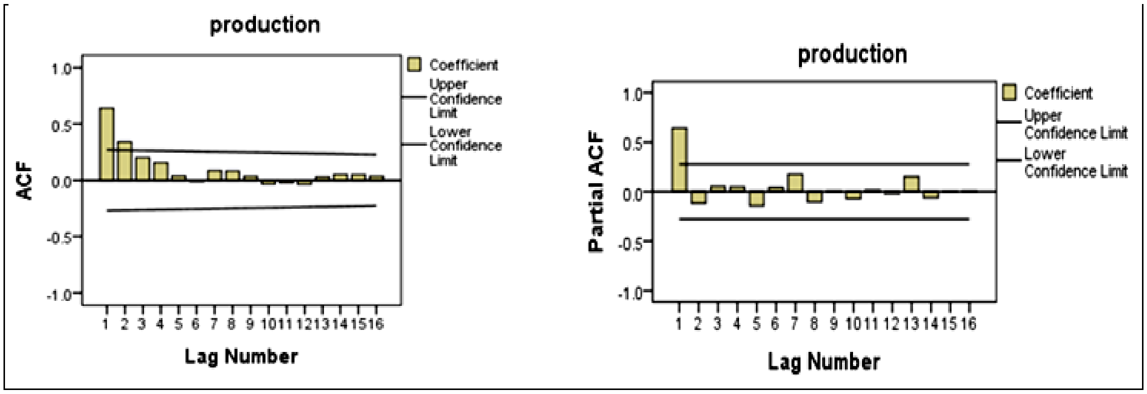

Figure 5. ACF and PACF of differenced data

The ARIMA models are discussed with values differenced twice (d=1) and the model which had the minimum normalized BIC was chosen. The various ARIMA models and the corresponding normalized BIC values are given in Table 3. The value of normalized BIC of the chosen ARIMA was 4.333.

Table 3: Estimated ARIMA Model Fit Statistics

Performances of different ARIMA (p,q,d) models of *Leguminosae* yield in Tamil Nadu								
ARIMA(p,q,d)	**Stationary R^2**	**R^2**	**RMSE**	**MAPE**	**MaxAPE**	**MAE**	**MaxAE**	**Normalized BIC**
0,1,0	0.000	0.380	8.938	44.643	886.431	5.492	30.610	4.458
0,1,1	0.004	0.382	9.012	44.904	863.880	5.550	28.256	4.551
0,1,2	0.232	0.524	7.994	43.579	851.677	5.160	27.127	4.389
1,1,0	0.000	0.380	9.027	44.675	883.524	5.504	30.317	4.555
1,1,1	0.058	0.416	8.855	44.019	856.977	5.447	25.074	4.593
1,1,2	0.276	0.551	7.842	42.693	833.558	5.051	22.529	4.427
2,1,0	**0.274**	**0.550**	**7.774**	**46.184**	**797.517**	**5.375**	**21.348**	**4.333**
2,1,1	0.314	0.575	7.635	44.884	854.696	5.167	21.086	4.374
2,1,2	0.337	0.589	7.585	45.054	851.844	5.125	19.966	4.438
3,1,0	0.341	0.591	7.486	44.094	893.696	4.997	21.425	4.334
3,1,1	0.355	0.600	7.482	44.575	891.763	5.026	21.346	4.410
3,1,2	0.369	0.608	7.486	44.115	875.630	4.955	21.052	4.489

Model parameters were estimated and reported in Table 3 and Table 4. The model verification is concerned with checking the residuals of the model to improve on the chosen ARIMA (p,d,q). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders, up to 24 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 5. This proves that the selected ARIMA model is an appropriate model.

Table 4: Estimated ARIMA Model of *Leguminosae* Production

		Estimate	**SE**	**t**	**Sig.**
Constant		0.776	0.751	1.034	0.307
Auto regressive	**Lag 1**	0.126	0.132	0.950	0.347
	Lag 2	-0.599	0.149	-4.005	0.000
Difference		1			

Table 5: Residual of ACF and PACF of *Leguminosae* Production

Lag	**ACF**	**Std. Error**	**PACF**	**Std. Error**	***Lag***	**ACF**	**Std. Error**	**PACF**	**Std. Error**
1	-0.149	0.140	-0.149	0.140	*13*	0.024	0.163	-0.096	0.140
2	0.080	0.143	0.059	0.140	*14*	0.055	0.163	0.062	0.140
3	-0.251	0.144	-0.237	0.140	*15*	0.021	0.163	0.053	0.140
4	0.079	0.152	0.011	0.140	*16*	-0.018	0.163	-0.064	0.140
5	0.055	0.153	0.100	0.140	*17*	-0.036	0.163	0.013	0.140
6	-0.052	0.154	-0.105	0.140	*18*	-0.119	0.163	-0.136	0.140
7	0.058	0.154	0.060	0.140	*19*	-0.066	0.165	-0.111	0.140
8	-0.035	0.154	0.25	0.140	*20*	-0.011	0.165	-0.056	0.140
9	0.118	0.154	0.069	0.140	*21*	0.040	0.165	-0.015	0.140
10	-0.207	0.156	-0.170	0.140	*22*	0.038	0.166	-0.003	0.140
11	0.018	0.162	-0.034	0.140	*23*	-0.023	0.166	-0.036	0.140
12	-0.091	0.162	-0.046	0.140	*24*	0.052	0.166	0.088	0.140

The ACF and PACF of the residuals are given in Figure 6, which also indicated the 'good fit' of the model. Hence, the fitted ARIMA model for the *Leguminosae* production data was

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

$$\mathbf{Y_t = 0.776 + 0.126Y_{t-1} - 0.599Y_{t-2} + \varepsilon_t}$$

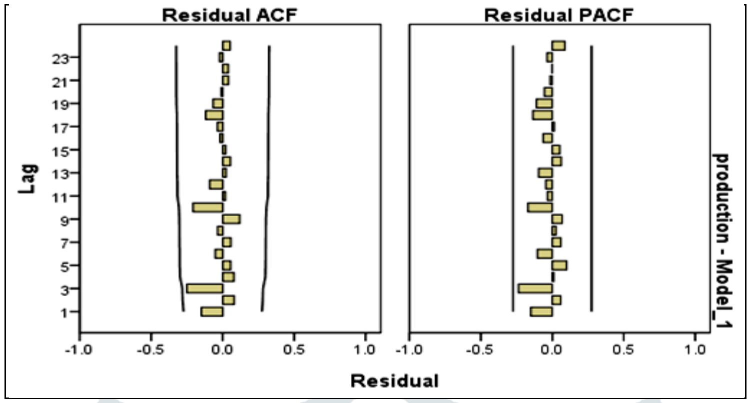

Figure 6. Residuals of ACF and PACF

Forecasting value of *Leguminosae* production (1000 tons) from the year 2018 to 2027 given in Table 6. To assess the forecasting ability of the fitted ARIMA (p,d,q) model, important measures of the sample period forecasts' accuracy were computed. Figure 7 shows the actual and forecasted value of *Leguminosae* production (1000 tons) data with 95% confidence limits.

Table 6: Forecast of *Leguminosae* Production (1000 tons)

Year	Forecast	UCL	LCL
2018	61.586	77.212	45.960
2019	53.048	76.576	29.520
2020	41.084	66.094	16.073
2021	45.835	71.593	20.076
2022	54.736	82.873	26.599
2023	54.154	85.302	23.007
2024	49.896	82.774	17.017
2025	50.852	84.927	16.777
2026	54.665	90.306	19.023
2027	55.714	93.190	18.239

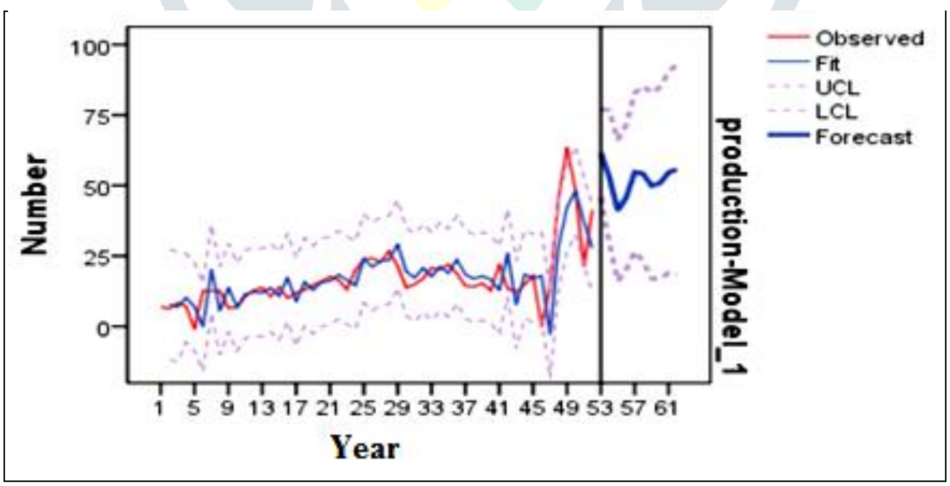

Figure 7. Actual and Estimate of *Leguminosae* Production (1000 tons)

IV. Conclusion

The results showed that the production would not remain stable throughout the year. The most appropriate ARIMA model for *Leguminosae* (minor pulses) production forecasting of data was found to be ARIMA (2,1,0). From the temporal data, it can be found that forecasted Production would increase from 41.482 thousand tonnes in 2017 to 55.714 thousand tonnes in 2027 in Tamil Nadu. Using time series data from 1966 to 2017 on *Leguminosae* production, these results will be useful for the government to take necessary steps to increase *Leguminosae* production in Tamil Nadu in future.

References

[1] Akaike, H. (1970). Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics*. **22**: 203-270.

[2] Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day, California, USA. https://link.springer.com/chapter/10.1057/9781137291264_6.

[3] Box, G.E.P. and Pierce, D.A. (1970). Distribution of Residual Autocorrelations in ARIMA Models. *J. American Stat. Assoc.*, **65**: 1509-1526. https://www.jstor.org/stable/ 2284333?seq=1.

[4] Jai Sankar, T., Prabakaran, R., Senthamarai Kannan, K. and Suresh, S. (2010). Stochastic Modeling for Cattle Production Forecasting. *International Journal of Modern Mathematics and Statistics.* **4(2)**: 53-57. https://medwelljournals.com/fulltext/?doi=jmmstat.2010.53.57

[5] Jai Sankar, T. (2011). Forecasting Fish Product Export in Tamilnadu - A Stochastic Model Approach. *International Journal of Recent Research in Science and Technology*. **3(7)**: 104-108. https://updatepublishing.com/journal/index.php/rrst/article/view/749/734.

[6] Jai Sankar, T., Senthamarai Kannan, K., Suresh, S. and Prabakaran, R. (2011). Stochastic Modeling for Bovine Production Forecasting. *International Journal of Agricultural and Statistical Sciences*. **7(1)**: 141-148. https://www.researchgate.net/publication/289812991_Stochastic_modelling_for_bovine_production_forecasting.

[7] Jai Sankar, T. and Prabakaran, R. (2012). Forecasting Milk Production in Tamilnadu. *International Multidisciplinary Research Journal*. **2(1)**: 10-15. https://www.academia.edu/62866649/Forecasting_milk_production_in_Tamilnadu?bulk.

[8] Jai Sankar, T. (2014). Design of a Stochastic Forecasting Model for Egg Production. *International Journal of Innovative Science, Engineering & Technology*. **1(6)**: 319-325. https://www.ijiset.com/v1s6/IJISET_V1_I6_52.pdf.

[9] Jai Sankar, T. and Vijayalakshmi, C. (2017). Implementation of ARIMA Model for Ghee Production in Tamilnadu. *International Journal of Pure and Applied Mathematics.* **113(6)**: 56-64. https://acadpubl.eu/jsi/2017-113-pp/articles/6/7.pdf.

[10] Jai Sankar, T., Pushpa, P. and Vijayalakshmi, C. (2017). Stochastic Forecasting Analysis for Rice Production in India. *Far East Journal of Theoretical Statistics.* **53(2)**: 63-75. http://www.pphmj.com/abstract/10739.htm.

[11] Jai Sankar, T., Pushpa, P. and Vijayalakshmi, C. (2017). Stochastic Time Series Analysis for Wheat (*Triticum aestivum*) Production in India. *International Journal of Pure and Applied Mathematics.* **113(7)**: 111-119. https://acadpubl.eu/jsi/2017-113-pp/articles/7/14.pdf.

[12] Jai Sankar, T. and Pushpa, P. (2019). Design and Development of Time Series Analysis for *Saccharum officinarum* Production in India. *A Journal of Composition Theory.* **12(9)**: 203-211. http://www.jctjournal.com/gallery/29-sep2019.pdf.

[13] Jai Sankar, T. and Pushpa, P. (2022).Design and Development of Stochastic Modelling for *Solanum tuberosum* Production in India. *Journal of Algebraic Statistics.* **13(2)**: 3353-3361. https://www.publishoa.com/index.php/journal/article/view/1067.

[14] Jai Sankar, T. and Pushpa, P. (2022). Design and Development of ARIMA Model for Bajra (*Pennisetum glaucum*) Production in India. *Journal of Algebraic Statistics.* **13(2)**: 3512-3523. https://publishoa.com/index.php/journal/article/view/1089.

[15] Jai Sankar, T. and Pushpa, P. (2022). Design and Development of Stochastic Modelling for *Musa paradisiaca Linn* Production in India. *Journal of Algebraic Statistics.* **13(2)**: 3591-3599. https://publishoa.com/index.php/journal/article/view/1103.

[16] Slutzky, E. (1973). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**: 105-146. https://www.scribd.com/document/388512380.

[17] Schwarz Gideon (1978). Estimating the Dimension of a Model. *Ann. Statist.*, **6(2)**: 461-464. https://doi.org/10.1214/aos/1176344136.