J. Electrical Systems 20-7s (2024): 3969-3974

Pagadala Srivyshnavi[1]
Dr. D. Aju[2*]

A Predictive Analytic Time Series Forecasting Model

Journal of Electrical Systems

ABSTRACT: Predictive Analytics is the process of using predictive modelling techniques to analyse the various types of data in agriculture, business management, engineering, weather forecasting, planning, meteorology etc. Most frequently using predictive modelling technique in Electrical load forecasting and Agricultural production forecasting is the time series forecasting technique. Particularly, in the electrical systems, the future electric load demand and peak load can be predicted by using several methods such as regression methods and time series methods; data science methods namely ML, DL, SVM, ANN, AI techniques etc. In the empirical study, a time series forecasting technique based on a selected ARIMA (1,1,2) model has been applied to an agricultural big data and obtained forecasts for Bengal gram average monthly prices of Andhra Pradesh state, India from Jan'2023 to Dec'2025. In the similar lines, the proposed time series forecasting modelling can be applied to forecast short term electric load demand by using the given previous bigdata.

Key Words: Time series forecasting technique, ARIMA (p,d,q) Model, Diagnostic tests, R-Language code.

1. Introduction

Data Analytics is the systematic computational analysis of data, which is an essential tool for all types of organizations. One of the modern forms of Data analytics is the predictive analytics which analyses the current and past data in predicting the future events. The various predictive analytic techniques can be broadly divided into four types namely; i. classification models, ii. clustering models, iii. time series forecasting models and iv. regression analytic models. Among four predictive analytic techniques, the time series forecasting, and regression based analytic techniques have a wide number of applications in analysing big data sets. Now-a-days, forecasting technique is an essential data analysis tool for research workers in almost all the fields of Science and technology. Several regression and time series forecasting techniques have been frequently applied in estimating demand for electric load in electrical systems.

Since, Bengal Gram (known as chana or chickpea or gram of grambango bean) accounts about 20% of the world pulses production, the time series data on Bengal gram monthly price data of Andhra Pradesh State in India during the period from January 1991 to November 2020 have been analysed and obtained forecasts of monthly wise Bengal gram prices in Andhra Pradesh state from 2023 to year 2025 by applying a predictive analytic in the present research work [1, 2].

2. Related Work

Vibhav Kumar and M.L Garg have presented trends and techniques with reference to predictive analytics[3]. Lee Chee Sun et.al., reviewed various applications of data science methods[4]. R.V.Mc Carthy et.al., have discussed about predictive analytics and their applications in a simple and systematic manner[5].

Under demand forecasting studies, Jamal Fattah et.al. have developed a forecasting approach by analysing a food company data[6]. Sanjay Tyagi et.al., have used different time series models including ARIMA model to obtain more accurate and effective predictions for sugarcane production in India [7]. Kumar Manoj and Madhu Anand have applied an univariate ARIMA model and obtained Sugar production forecasts[8]. T.W. Mapuwei et.al. have focused on forecasting tobacco yield in Zimbabwe. Box-Jenkin's methodology employed in building various types of ARIMA models; and R-Language code and certain performance metrics have been used in selecting the best ARIMA (1, 1, 0) model for forecasting [9]. In forecasting wheat production in India for upto 10 years, D.S. Bholanath et.al., have been analysed the time series data by using ARIMA models[10]. Medhat Rostum et.al., have used various stochastic forecasting techniques to forecast demand of electric load and compared them by measures of performances[11]. Cheng-Ming Lee and Chia-Nan Ko have predicted the short term electric load data by using ARIMA-model[12]. Grzegorz Dudek applied local linear model for short term load forecasting based upon time series patterns[13].

3. Methodology

Time series forecasting model, involves mainly five components namely (i) determining the objective (ii) specifying a forecasting model (iii) estimating and testing the forecasting model (iv) application of forecasting model and (v) evaluating and revising the forecasting model. In the present research work, univariate ARIMA (p,d,q) models have been applied to analyse the data by using predictive analytic time series forecasting technique.

Box and Jenkin's methodology consists of the following three iterative steps which have been used to build ARIMA (p,d,q) models for forecasting: (i) model Identification (ii) selection of estimators and (iii) validation of the ARIMA (p,d,q) model. With usual notation, the following five types of unvariate ARIMA (p,d,q) models have been fitted to empirical time series data by using R-Language code:

---

[1]Research Scholar, Email Id: vyshu.sri@gmail.com
[2*]Associate Professor, Department of Computer Science & Engineering VIT, Vellore. India – 632014
Email Id: daju@vit.ac.in
Copyright © JES 2024 on-line : journal.esrgroups.org

(i) ARIMA (0, 0, 0) model : $Y_t = \alpha + e_t$ (3.1)

(ii) ARIMA (1, 0, 1) model : $y_t = \alpha + \phi_1 y_{t-1} + e_t - \theta_1 e_{t-1}$ (3.2)

(iii) ARIMA (1, 1, 1) model : $(1-\phi_1 B)\ (1-B) y_t = \alpha + (1-\theta_1 B)\ e_t$ (3.3)

(iv) ARIMA (1, 0, 0) model : $y_t = \alpha + \phi_1 y_{t-1} + e_t$ (3.4)

and (v) ARIMA (1, 1, 2) model : $(1-\phi_1 B)(1-B) y_t = \left(1-\theta_1 B - \theta_2 B^2\right) e_t$ (3.5)

Here, B is the Backward shift operator : $B(e_t) = e_{t-1}$

- **Selecting best ARIMA (p,d,q) model:**

Diagnostic Tests:

1. Shapiro and Wilk have developed a statistical test procedure to obtain normal sample.
W- Statistic for testing normality of the data is given by

$$W = \frac{\left[\sum_{i=1}^{n} a_i x_{(i)}\right]^2}{s^2} \quad (3.6)$$

Where, $S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$ (3.7)

and $\bar{x} = \left(\sum_{i=1}^{n} x_i \Big/ n \right)$ .Here, the coefficients a_i's are defined as $a_i = \frac{\left[\sum_{j=1}^{n} m_j v^{ij}\right]}{C}$ (3.8)

Here, m_j's are the expected values of the standard normal order statistics $x_{(i)}$'s respectively; V^{ij}'s denote the elements of Inverse covariance matrix $V^{-1} = \left(\left(V^{ij}\right)\right)^{-1}_{nxn}$; C is the vector norm which is given by $C = \|V^{-1}m\| = \left(m^{1}v^{-1}v^{-1}m\right)^{1/2}$ and m is a (n x 1) vector. If both mean vector 'm' and the covariance matrix V are known, then the coefficients a_i's can be directly determined. For the use of W-test statistic, various approximations to the coefficient a_i's are given in the literature from time to time by different researchers. Shapiro and Wilk suggested approximations to the coefficients a_i's and tabulated them for sample sizes from n = 2 to 50[14].

Remark: For smaller values of W- statistic, the null hypothesis of normality of sample data may be rejected.

2. Grubbs Test statistic for outliers detection is given by

$$G = \underset{i=1,2,....,n}{\text{Max}} \frac{|x_i - \bar{x}|}{s} \quad (3.9)$$

Where $\bar{x}$'s are sample arithmetic means; s is sample standard deviation; and X_i is data point to be tested. Later, this outlier is to be expunged [15].

3. In the non-stationary case, consider simple Auto Regressive (AR) process

$$\Delta y_t = (\rho - 1) y_{t-1} + e_t \quad \text{or} \quad \Delta y_t = \delta y_{t-1} + e_t \quad (3.10)$$

Here, Δ is the first difference operator and $\delta = \rho - 1$

Dickey and Fuller (DF) test statistic for testing $\delta = 0$ is given by $DF_{cal} = \frac{\hat{\delta}}{SE(\hat{\delta})}$ (3.11)

Where $\hat{\delta}$ is obtained by estimating the model (3.10).

Augmented DF statistic for testing the stationarity in the time series $(\gamma = 0)$ is given by $ADF_{Cal} = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$

Where, $\hat{\gamma}$ is obtained by estimating the model: $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + e_t$ (3.12)

and β is Coefficient on a time trend.

Also, SE $(\hat{\gamma})$ is the standard error of $\hat{\gamma}$.

The critical values for ADF test statistics are given by Dickey and Fuller. ADF_{cal} is a negative number. The larger negative reveals stronger non-stationarity in the time series data [16].

4. Empirical Results

In the present study, under empirical investigation, ARIMA (1,1,2) has been selected and obtained the forecast values for Bengal gram monthly process from Jan' 2023 to Dec' 2025. Empirical results have been presented below:

Table 1: Descriptive Statistics for Bengal Gram Monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020

Bengal gram(Price)	
Mean	2407.39
Standard Error	70.70
Median	1924
Mode	1040
Standard Deviation	1339.64
Kurtosis	-0.60
Skewness	0.72
Range	5024.66
Minimum	625.34
Maximum	5650
Count	359

The results obtained from descriptive statistics reveal that the mean and standard deviation of Bengal Gram Monthly price data during the study period are respectively given by Rs. 2407.39 and Rs. 1339.64 per quintal.

Table 2: Diagnostic Tests for Bengal Gram Monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020

Diagnostic Test	**Calculated Value of Test Statistic**	**Probability**
W	0.9081	0.00*
G	2.42	1.00

***: Highly significant at 0.01 level**

It has been observed from the Shapiro and Wilk test, the time series data needs normal transformation of data. The corresponding test statistic value is 0.9081 which is highly significant at 1% level.

Grubbs test statistic value is given by 2.42 which is not significant at 5% level. It reveals that no outliers exist in the transformed data.

Table 3: ADF Test for Stationarity in Bengal Gram Monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020

	Calculated value of ADF Test Statistic	**Probability**
No Difference	-3.273	0.07581
First Difference	-8.1798	0.01*

***: Significant at 0.01 level**

The ADF test statistic value for no difference (d=0) is given by -3.273 which is not significant at 5% level. At first differencing, the calculated value of ADF test statistic is given by -8.1798 which is highly significant at 1% level. Thus, the value of d is taken as '1' in the selected ARIMA model.

Table 4: Performance Evaluation Metrics Values of Various ARIMA (p, d, q) Models for Bengal Gram monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020.

Model	R^2 Value	$\overline{R}^2$ Value	RMSE	MAPE	Normalized BIC
ARIMA(0,0,0)	0.983	0.9830	178.860	7.072	10.816
ARIMA(1,0,1)	0.976	0.9759	205.727	5.628	10.670
ARIMA(1,1,1)	0.993	0.9929	116.305	3.919	9.9230
ARIMA(1,0,0)	0.986	0.9860	160.465	4.885	10.418
ARIMA(1,1,2)	0.993	0.9929	116.237	3.888	9.9050

From the results obtained from the performance evaluation metric analysis, the ARIMA (1,1,2) model has been selected as an appropriate model for forecasting Bengal gram monthly price data. For the largest $\overline{R}^2$ value; smallest RMSE, MAPE and normalized BIC are respectively given by 0.9929; 116.237, 3.888 and 9.9050.

The necessary graphic representations for trend line; ACF and PACF; residuals from ARIMA (1,1,2) with drift and forecasts from ARIMA (1,1,2) with Drift have been respectively represented in the following four figures:

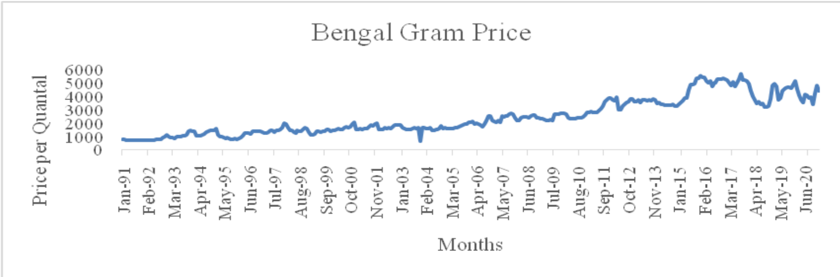


Fig 1: Trend line for Bengal Gram Monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020.

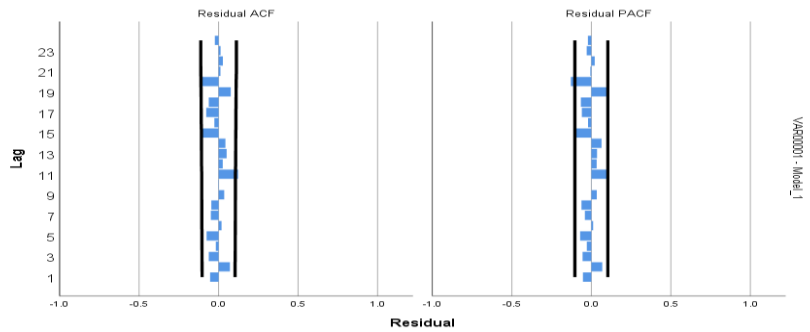
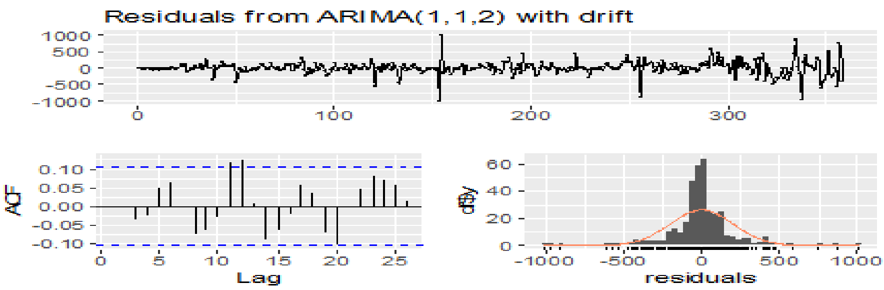


Fig 2: ACF and PACF of Bengal Gram Monthly Price Data (Rabi Harvest)Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020.



.Fig 3: Residual Analysis for Bengal Gram Monthly Price Data (Rabi Harvest) of Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020

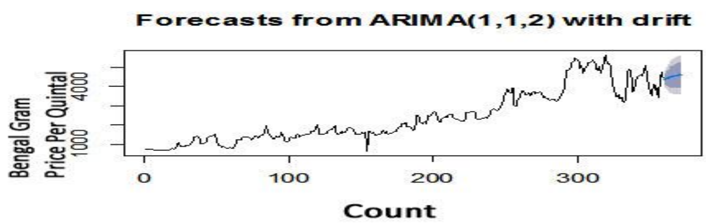


Fig4: Forecast analysis for Bengal Gram Monthly Price Data (Rabi Harvest)

Andhra Pradesh state in India during period from Jan' 1991 to Nov' 2020

The cautious inspection of ACF and PACF graphs of residuals upto 24 lags exhibit that volatility is present in the data. The predicted values of Bengal gram monthly average prices of Andhra Pradesh State in India for 36 months from Jan' 2023 to Dec' 2025 obtained by using the estimated selected ARIMA (1, 1, 2) model have been presented along with their corresponding 95% confidence limits in the Table -5.

Table -5: Predicted Bengal Gram Monthly Average Prices (Rs. Per quintal) of Andhra Pradesh State in India along with 95% confidence limits from Jan'2023 to Dec' 2025

S No	Year	Predicted	LCL-95%	UCL-95%
1	Jan-23	5175	3887	6595
2	Feb-23	5186	3893	6612
3	Mar-23	5198	3900	6630
4	Apr-23	5210	3908	6647
5	May-23	5223	3916	6664
6	Jun-23	5235	3924	6681

7	Jul-23	5248	3933	6698
8	Aug-23	5261	3943	6716
9	Sep-23	5275	3953	6733
10	Oct-23	5288	3963	6751
11	Nov-23	5302	3973	6768
12	Dec-23	5316	3984	6786
13	Jan-24	5331	3995	6803
14	Feb-24	5345	4007	6821
15	Mar-24	5360	4018	6839
16	Apr-24	5374	4030	6857
17	May-24	5389	4042	6875
18	Jun-24	5404	4054	6893
19	Jul-24	5419	4067	6911
20	Aug-24	5435	4079	6930
21	Sep-24	5450	4092	6948
22	Oct-24	5466	4105	6966
23	Nov-24	5481	4118	6985
24	Dec-24	5497	4131	7003
25	Jan-25	5513	4144	7022
26	Feb-25	5529	4157	7041
27	Mar-25	5545	4171	7059
28	Apr-25	5561	4184	7078
29	May-25	5577	4198	7097
30	Jun-25	5594	4212	7116
31	Jul-25	5610	4226	7135
32	Aug-25	5627	4240	7154
33	Sep-25	5643	4254	7173
34	Oct-25	5660	4268	7192
35	Nov-25	5676	4282	7212
36	Dec-25	5693	4296	7231

5. Conclusion

In the present research paper, five types of ARIMA model have been specified and estimated by using Bengal gram monthly average prices data of Andhra Pradesh State in India during the period from Jan' 1991 to Nov' 2020. R-Language code along with certain performance evaluation metrics have been applied and selected ARIMA model for the purpose of forecasting.

The forecasts of Bengal gram monthly average prices (Rs. per quintal) of Andhra Pradesh along with 95% confidence limits for the Period from Jan'2023to Dec'2025 have been obtained by using the selected best predictive analytic ARIMA (1, 1, 2) model. Farmers would be benefited by taking appropriate decisions based on our forecasts. Bengal gram farmers can be encouraged to continue in producing the Bengal gram if they get stabilized prices. Forecasts will be helpful to farmers to make future Plans either for storing or selling their production.
In the similar lines the proposed procedure can be extended to find forecasts for the demand of electric load in the electrical system.

References

[1] A Agmarknet, Bengal Gram-Chickpea(Chana), Chana Price hire, 230 Oct, 2021, https://agmarknet.gov.in>be......pdf.

[2] Directorate of Economics and Statistics, A.P. Bengal Gram Price – Rabi Harvest (Rs. per quintal) from Jan' 1991 to Nov' 2020 (359 data points).

[3] Vaibhav Kumar and M.L. Garg, Predictive Analytics: A Review of Trends and Techniques, International Journal of computer Applications, 182(1)(2018), 31-37.

[4] Lee Chee Sun, Cheang, Peek Yeng Sharon, Moslehpour, Massond, Predictive Analytics in Business Analytics: Decision Tree, Advances in Decision Sciences, 26(1)(2022), 1-29.

[5] R.V. Mc Carthy, M.M. Mc Carthy, W. Caccucci, L. Halahi, Applying Predictive Analytics, Springer, 2022.

[6] Jamal Fattah, Latifa Ezzine, Zineb Aman and Haj EI Moussami, Abdeslam Lachhab Forecasting of Demand using ARIMA Model, International Journal of Engineering Business Management, 10 (2018), 1-9.

[7] Sanjay Tyagi, Shalini Chandra and Gargi Tyagi. Statistical Modelling and Forecasting Annual Sugarcane Production in India: Using Various Time Series Models, Annals of Applied Biology, 2023. DOI :10.1111/aab.12825 (accessed on 2023-01-24), https://www.x-mol.com

[8] Kumar Manoj and Madhu Anand, An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India, Studies in Business and Economics, 9 (2014), 81-94.

[9] T.W.Mapuwei, J.N.Dava, M.Kachaka and B.Kusotera, An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe, American Journal of Modelling and Optimization, 9 (2022), 15-22, doi:10, 12691/ajmo-9-1-3.

[10] D.S.Bhola Nath, Dhakre and Debasis Bhatta Charya. Forecasting Wheat Production in India: An ARIMA Modelling Approach, Journal of Pharmacognosy and Phytochemistry, 8 (2019), 2158-2165.

[11] Medhat Rostum, Amr Zamel, Hassan Moustafa, Ibrahim Ziedam, "Electrical Load Forecasting: A Methodological overview", International Lurnal of Engineering and Technology, 9(5) (2020) 842.

[12] Cheng – Ming Lee and Chia – Nan Ko, "Short term Lord : Forecasting using Lifting Scheme and ARIMA Models, Expert systems with Application 38(5) (2011) 5902-5911.

[13] Grazegorg Duck, "Pattern based Local Linear Regression Model for short term and Forecasting", Electric Power Systems, 130(2016) 139-147.

[14] S.S. Shapiro, and M.B.Wilk, An Analysis of Variance Test for Normality (Complete Samples) Biometrika, 1965, 52 (3-4), 591-611.

[15] Grubbs, F. Procedures for Detecting Outlying Observations in Samples, Technometrics, 11 (1969), 1-21.

[16] D.A.Dickey, and W.A.Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of American Statistical Association, 71 (1979), 427-434.