Jute price forecasting in Murshidabad market of west Bengal using ARIMA technique

Ravi Ranjan Kumar, Akhilesh Kumar Gupta, Chinmayee Patra and Mamata

Abstract
The present study is an attempt to forecast the prices of jute in the markets of Murshidabad district of West Bengal. The time series data on monthly price of jute required for the study was collected from AGMARKNET website for the period January, 2002 to December, 2019. The seasonal ARIMA model was used for the modelling of price using the Box-Jenkins technique and best model was selected on the basis of lowest RMSE, AIC and MAPE values and the best identified model was ARIMA (1,1,1)(1,0,0)\[12\]. The fitted values for the in-sample period and predicted values for out-sample period were closer to real time price values. The best identified model was used for predicting the future prices of 12 months (January, 2020 to December, 2020). The forecasted price rises from the month of March, 2020 and stabilizes thereafter. The analysis was done in “R” statistical software.

Keywords: ARIMA, AIC, MAPE, forecasting, jute, time series

Introduction
Jute, known as the golden fiber, is an important traditional cash crop in India. In fact, jute is the second most important natural fiber in terms of global consumption after cotton. One of the important characteristics of jute is that it is free from health hazards and environmental pollution. Jute is also versatile, durable, reusable, cheap, and superior to synthetic fibers. Jute is regarded as the best natural substitute for nylon and polypropylene. In agriculture as well as industry, jute directly and indirectly supports employment, commerce and other economic activities. In trade and industry however, in term used as raw jute. Raw jute plays an important role in the country’s economy, particularly in the eastern and north eastern state. West Bengal is the major jute growing state sharing about three fourth of the country’s production (Mondal et al., 2014)\[13\]. Although jute has been recognized as a solution to produce eco-friendly products for the future, the production of jute and jute products are actually declining globally. This is mainly due to the availability of plastic substitutes, which is likely to continue in the future. India is the single largest producer of jute goods in the world, contributing about 60% of the total production (Kundu D.K.,2016)\[12\]. In India, Its grown on 0.686 m.ha. area, and production of 9.6 million bales (of 180kgs each). West Bengal is the most leading state of jute production in the country, followed by Bihar and Assam. In West Bengal, it is grown on 0.515 m.ha. area, and production of 7.511 million bales (of 180kgs each), (India, 2018-19). About 90% of the jute farmers in India belong to the marginal and small categories of which almost 65% are with 1 hectare and below and 25% with 2 ha and below land holdings (Kundu D.K.,2016)\[12\].

The present study is aimed to forecast the wholesale prices of jute for the markets of Murshidabad district of West Bengal. As the price of jute keep changing from time to time, it creates risks to producers, traders and consumers involved in production, marketing and consumption of jute. Thus, it is important to forecast the jute prices. Price forecasts are critical to market participants who make production and marketing decisions, and to policy makers who administer commodity programs and assess the market impacts of domestic or international events (Sharma, 2015)\[15\].

Literature Review

Materials and Methods

Data Collection

The time series data on monthly average price of jute in the markets of Murshidabad district of West Bengal for the period of 18 years from January, 2002 to December, 2019 were collected from AGMARKNET website for present study. Murshidabad district includes 5 markets viz., Beldanga, Jangipur, Jiaganj, Kasimbazar, Lalbagh. Monthly prices of these markets were averaged to use as monthly price for overall Murshidabad District and the same was for modelling and forecasting.

Analytical Method

Box-Jenkins (ARIMA) Model: The Box-Jenkins models (1976), are especially suited to short term forecasting because most ARIMA models place greater emphasis on the recent past rather than the distant past. The ARIMA model analyses and forecasts equally spaced univariate time series data. In this study, the analysis performed by ARIMA is divided into four stages viz., identification, estimation, diagnostics and forecasting. R programming software was used for time series analysis and developing ARIMA models and forecasting jute prices. The ARIMA(p,d,q) model can be represented by the following general forecasting equation:

\[ Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{d} \theta_j \varepsilon_{t-j} + \varepsilon_t \]

where, \( Y_t \) is soybean prices, \( \mu \) is the mean of series, the \( \phi_1, \ldots, \phi_p \) are the parameters of the AR model, the \( \theta_1, \ldots, \theta_q \) are the parameters of the MA model and the \( \varepsilon_{t-j}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q} \) are the noise error terms. The value of \( p \) is called the order of AR model while the value of \( q \) is called the order of the MA model (Gupta et al., 2018) [10].

Since seasonal time series data is taken for this study, ARIMA model can be extended easily to handle seasonal aspects denoted as ARIMA(p,d,q) (P,D,Q), where the small letter parentheses part (p,d,q) indicates the non-seasonal part of model while the capital letter part (P,D,Q) indicates the seasonal part of models being the number of periods per season (Barathi 2011) [11]. The general seasonal autoregressive integrated moving average (SARIMA) model written as follows:

\[ \Phi_P(B^s)\phi_p(B) \nabla^D \nabla^d Y_t = \theta_Q(B)\theta_q(B^s)\varepsilon_t \]

where,

- \( \Phi_P(B^s) = (1 - \phi_1B^s - \ldots - \phi_PB^{sp}) \) is the seasonal AR operator of order \( P \);
- \( \phi_p(B) = (1 - \phi_1B - \ldots - \phi_PB^p) \) is the regular AR operator of order \( p \);
- \( \nabla^D = (1 - B^s)^D \) represents the seasonal differences and \( \nabla^d = (1 - B)^d \) the regular differences;
- \( \theta_Q(B^s) = (1 - \theta_1B^s - \ldots - \theta_QB^{sq}) \) is the seasonal moving average operator of order \( Q \);
- \( \theta_q(B) = (1 - \theta_1B - \ldots - \theta_QB^q) \) is the regular moving average operator of order \( q \);
- \( \varepsilon_t \) is a white noise process;

Identification Stage

Model identification involves defining the orders (p, d, and q) of the AR and MA components of the time series model. The first stage of ARIMA modelling is to identify that the variable, which is about to be forecasted, is a stationary time series or not. Stationary means the values of variable over time fluctuates around a constant mean and variance. The ARIMA model cannot be made until we make the series stationary. First, we have to take the difference of the time series ‘d’ times to obtain a stationary series to obtain an ARIMA (p, d, q) model, where ‘d’ is the order of differencing.

There should a Caution is to be taken in differencing the time series as over differencing will lead to increase the standard deviation, rather than the reduction. The best idea is to start the differencing with the lowest order (first order, d = 1) and then test the unit root problems for the data.

Before moving further, we will have to test the differenced time series for stationarity (unit root problem) by Augmented Dickey-Fuller test (ADF). After that, best fit ARIMA models were developed using the data from January 2002 to December 2019 and used to forecast the prices during harvesting season.

Criterion of Model Selection: The best model was selected based on the following criterions.

I) Low Akaike Information Criteria (AIC): AIC is estimated by

\[ AIC = -2 \ln(L) + 2K \]

Where, \( K \)= Number of Parameters and \( L \)= Maximized log likelihood

II) Mean Absolute Percent Error (MAPE): It is a measure of percentage error for the model and it can be easily to understood for model accuracy.

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \]

Where, \( y_i \) = \( i \)th actual value and \( \hat{y}_i \) = \( i \)th predicated value

III) R.M.S.E (Root Mean Square Error): Root Mean Square Error is defined as the differences between the value of population and sample predicted by a model and actually observed values.

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p}} \]

Results and Discussion

Model identification

The first step in developing a ARIMA model is to determine if the monthly jute price series are stationary. For this, we...
used the Augmented Dickey-Fuller test (ADF), which was performed to determine if the series is stationary or not. The test confirmed that the data was nonstationary for without difference. In this case differencing of lag 1 gave significant result, so with differencing of lag 1 (d = 1) is stationary in respect to mean and variance. Thus, there is no need of further differencing the time series and then the adopted difference order is d = 1 for the ARIMA (p, d, q) model. The test statistic and its p-value is presented in Table 1.

<table>
<thead>
<tr>
<th>Without difference</th>
<th>Dickey-Fuller statistic = -3.3176</th>
<th>p-value =0.0697</th>
</tr>
</thead>
<tbody>
<tr>
<td>With difference</td>
<td>Dickey-Fuller statistic = -5.2579</td>
<td>p-value =0.01</td>
</tr>
</tbody>
</table>

This test allows to go further in the steps for ARIMA model development which are to find out the appropriate values of (p,d,q) (P,D,Q). It was done by observing Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) values. The Auto Correlation Function helps in choosing the appropriate values for ordering of moving average terms (MA) and Partial Auto-Correlation Function for those autoregressive terms (AR). The ACF graph showing significant spike at lag 1 but there is no significant spikes at seasonal lags (lag 12, 24) which gives us the non-seasonal and seasonal MA orders i.e. q=1 and Q=0. Similarly, PACF graph showing significant spike at lag 1 but no significant spikes at seasonal lags (lag 12, 24) which gives us the non-seasonal and seasonal AR orders i.e. p=1 and P=1. These are tentative orders and we tried some other models similar to these models.

![ACF of differenced data at lag 1](image1)

**Fig 2:** Autocorrelation Function (ACF) of first differenced series by lag

![PACF of differenced data at lag 1](image2)

**Fig 3:** Partial Autocorrelations Function of first differenced time series by lag
The lowest values of RMSE, MAPE, and AIC, the model ARIMA (1,1,1) and ARIMA (1,1,1)(1,0,0)[12] were best fitted model for Murshidabad district. The results of these coefficients are given in Table 2. ARIMA model was estimated after transforming the variables under study into stationary series through computation of either seasonal or non-seasonal or both, order of differencing. A careful examination of ACF and PACF up to 24 lags revealed the presence of seasonality in the data.

<table>
<thead>
<tr>
<th>S.N</th>
<th>MODELS</th>
<th>RMSE</th>
<th>MAPE</th>
<th>AIC</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ARIMA (1,1,1)</td>
<td>10.45032</td>
<td>0.346278</td>
<td>2664.61</td>
<td>6.34127</td>
</tr>
<tr>
<td>2.</td>
<td>ARIMA (1,1,1)(1,0,0)[12]</td>
<td>10.12219</td>
<td>0.341314</td>
<td>2666.27</td>
<td>5.827544</td>
</tr>
</tbody>
</table>

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARIMA (Darekar and Reddy, 2017) [7]. After model validation, MAPE value found to be 6.34127 and 5.827544 for ARIMA (1,1,1) and ARIMA (1,1,1)(1,0,0)[12] respectively. The model ARIMA (1,1,1)(1,0,0)[12] have lowest MAPE over ARIMA (1,1,1). Hence we conclude that model ARIMA (1,1,1)(1,0,0)[12] was best fitted for forecasting jute price in market of Murshidabad district of west Bengal. The results show that autocorrelations of residuals were not significantly different from zero at any reasonable level. This proved that the selected ARIMA model was an appropriate model for forecasting soybean price in market of Murshidabad district. The result of Box Ljung Q statistics also confirms the same. The results of diagnostics a represented in fig.3 and fig.4.

**Fig 3:** Residuals analysis of model ARIMA (1,1,1)(1,0,0)[12].

**Forecasting using selected model ARIMA (1,1,1)(1,0,0)[12]:**

The above selected model ARIMA (1,1,1)(1,0,0)[12], which we are fitting to our time series data. The results of forecast of prices of jute are shown in the Table 3. The forecasts indicate that there are narrow variations in between the actual and forecasted values of prices of jute in the selected state and the forecasted values of prices showed an increasing trend in the future months. The result suggests that the price of jute will range from 3788.46 to 3789.67 from January, 2020 to December, 2020. The result shows the price increases from the month of March, 2020 and stabilizes thereafter. The graphical representation of sample period forecasts and post sample period forecasts are presented in which shows that actual values of prices and fitted values from selected model were closer.
Table 3: Forecasts of jute prices in Murshidabad district

<table>
<thead>
<tr>
<th>Month</th>
<th>Year</th>
<th>Forecasted prices</th>
<th>Low 80%</th>
<th>High 80%</th>
<th>Low 95%</th>
<th>High 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>2020</td>
<td>3788.468</td>
<td>2759.562</td>
<td>4817.375</td>
<td>2214.892</td>
<td>5362.045</td>
</tr>
<tr>
<td>Feb</td>
<td>2020</td>
<td>3788.948</td>
<td>2716.377</td>
<td>4861.519</td>
<td>2148.593</td>
<td>5429.304</td>
</tr>
<tr>
<td>Mar</td>
<td>2020</td>
<td>3789.105</td>
<td>2674.414</td>
<td>4903.796</td>
<td>2084.332</td>
<td>5493.877</td>
</tr>
<tr>
<td>Apr</td>
<td>2020</td>
<td>3789.055</td>
<td>2633.736</td>
<td>4944.374</td>
<td>2022.147</td>
<td>5555.964</td>
</tr>
<tr>
<td>May</td>
<td>2020</td>
<td>3789.089</td>
<td>2594.51</td>
<td>4983.667</td>
<td>1962.139</td>
<td>5516.038</td>
</tr>
<tr>
<td>Jun</td>
<td>2020</td>
<td>3789.059</td>
<td>2556.469</td>
<td>5021.65</td>
<td>1903.975</td>
<td>5574.144</td>
</tr>
<tr>
<td>Jul</td>
<td>2020</td>
<td>3789.179</td>
<td>2519.714</td>
<td>5058.645</td>
<td>1847.699</td>
<td>5730.659</td>
</tr>
<tr>
<td>Aug</td>
<td>2020</td>
<td>3789.094</td>
<td>2483.794</td>
<td>5094.393</td>
<td>1792.81</td>
<td>5755.377</td>
</tr>
<tr>
<td>Sep</td>
<td>2020</td>
<td>3789.302</td>
<td>2449.127</td>
<td>5129.478</td>
<td>1739.68</td>
<td>5838.924</td>
</tr>
<tr>
<td>Oct</td>
<td>2020</td>
<td>3789.401</td>
<td>2415.234</td>
<td>5163.568</td>
<td>1687.794</td>
<td>5891.009</td>
</tr>
<tr>
<td>Nov</td>
<td>2020</td>
<td>3789.566</td>
<td>2382.228</td>
<td>5196.904</td>
<td>1637.228</td>
<td>5941.903</td>
</tr>
<tr>
<td>Dec</td>
<td>2020</td>
<td>3789.675</td>
<td>2349.93</td>
<td>5229.419</td>
<td>1587.755</td>
<td>5991.374</td>
</tr>
</tbody>
</table>

Fig 5: Fitted and forecasted values of jute prices in Murshidabad district.

Conclusion
The paper forecasted jute prices for the year 2020 by using historical monthly prices. The study analysed the time series data of prices of jute through ARIMA modelling and concluded that ARIMA (1,1,1) (1,0,0) [12] is best identified model for forecasting of jute prices in Murshidabad market. The study also concluded that there is rise in prices from the month of January, 2020. The fitted values by selected ARIMA model were closer to the real time price values. The ARIMA model forecasted prices revealed an increase in the prices of jute in the future years and also demand for the crop. Hence, increase in the area of production of jute and their sale in the suitable markets can be planned suitably.

References
12. Kundu DK. Potentials of Jute in Sustainable Family Farming for Livelihood Security in India and Recent...


