Journal of Statistics *ISSN-P: 1684-8403*
Volume 27, 2023, pp. 96-102 *ISSN-E: 2957-8981*

Forecasting the Production of Centrifugal Sugar in Pakistan using ARIMA Modelling

Muhammad Kashif Saleem[*1], *Parwarasha Nazir*[2], *Fahim Nazeer*[3]

Abstract

Throughout history, people have been intrigued by what lies ahead. The purpose of this study is to develop an ARIMA model that can predict Pakistan's annual sugar production between the years 1974 and 2021, as well as provide forecasts for upcoming years. The ARIMA (1, 1, 1) model was found to be the best fit based on the minimum value of the Bayesian Information Criterion (BIC). According to the 95% confidence interval sugar forecast for Pakistan from 2021 to 2030, the anticipated sugar production in 2030 is 6881 million tons. Additionally, this study uncovered a rising trend in sugar production in Pakistan.

Keywords

Autoregressive integrated moving average (ARIMA), Forecasting, Centrifugal Sugar, Pakistan.

1. Introduction

Sugar that is separated from its liquid through a centrifugal machine is called centrifugal sugar. Pakistan has 81 sugar mills. In Pakistan, the private sector controls a large portion of the sugar industry. Of the 81 sugar mills, 40 (or 49%) are in Punjab, 32 (or 40%) are in Sindh, and 9 (or 11%) are in NWFP. (ICMA, 2021). After textiles, sugar is the second-largest agro-based sector in Pakistan. Pakistan is a significant cane-producing nation, ranking fifth in terms of cane area and fifteenth in terms of sugar production in the world. Its contributions to GDP and agriculture's value-added are 3.4 percent and 0.7 percent, respectively. 4.2% of manufacturing is made up of sugar. More than 1.5 million people are employed by the sugar sector, including managers, technologists, engineers, financial specialists, as well as skilled, semiskilled, and unskilled labourers (Nadeem *et al*. 2019). The demand-supply imbalance has resulted in higher sugar prices. The market's supply is insufficient to satisfy the demand currently. In 2009, more than a million hectares of sugar cane were planted, and this provided the raw materials for 81 sugar mills in Pakistan.

To begin with, there are several factors that impact the production of sugarcane in Pakistan. When compared to other nations, Pakistan's sugarcane yield per hectare is 20 to 30% lower on average. Moreover, the sucrose recovery rate of the types of sugarcane grown in Pakistan is 15-20% lower than those of other major sugarcane-producing countries.

* Corresponding author
[1] Government Associate College (B), Shahkot, Nanakana Sahib, Pakistan.
Email: ranakashifpu@gmail.com
[2] Government Graduate College for Women, Jaranwala, Faisalabad, Pakistan.
[3] Government Graduate College, Mandi Bahauddin, Pakistan.

Additionally, sugarcane requires a higher volume of water per hectare and takes up the land for a longer period of time each year in comparison to other crops. However, due to price support and input subsidies, sugarcane farming is more profitable for farmers than other competing crops, particularly in terms of water usage. Finally, despite no noticeable increase in the amount of sugarcane produced per hectare, the average price paid to sugarcane growers in Pakistan is significantly higher than the global price.

The Autoregressive integrated moving average (ARIMA) model is recommended because of its well-known ability to handle time series data. In contrast to other forecasting models, ARIMA can effectively handle data sets with unclear structural linkages or underlying economic models, if it is considered that past values of the series and prior error terms include information that may be used to make forecasts. The advantage of the ARIMA forecasting model is that all the data for the target variable need simply be time series data. Additionally, the ARIMA model consistently provides robust forecasts that exceed those from more complex structural models; as a result, it is preferred in practise. The ARIMA (p, d, q) model was introduced by Box and Jenkins in 1976 as a method for locating, estimating, and validating models for a particular time series dataset. The Autoregressive Integrated Moving Average (ARIMA) models are a group of models that Box and Jenkins (1976) propose. It seems to be applicable in a wide range of circumstances. However, picking the right ARIMA model might not be simple. According to a lot of literature, creating an accurate ARIMA model is an art that necessitates sound judgement and extensive experience. In the beginning, George Box and G. William Jenkins studied ARIMA models in-depth, and since 1968, their names have been used interchangeably with the general ARIMA process utilised in time series analysis, forecasting, and control (Ho and Xie, 1998; Wei, 2006).

Mehmood *et al*. (2018) used the ARIMA model to depict their study on sugarcane production predictions in Pakistan. Ali *et al*. (2015) described the study's attempts to anticipate production and yield of Pakistan's two primary cash crops, cotton and sugarcane, using the ARIMA models of forecasting. The ARIMA model was used by Vishwajith *et al*. (2016) to predict the productivity of sugar and sugarcane. Hossain *et al*. (2015) categorised the study whose primary goal is to find the Auto-Regressive Integrated Moving Average (ARIMA) model that could be utilised to predict the production of sugarcane in Bangladesh. For the forecasting and production of maize in India, Sharma *et al*. (2018) put together the ARIMA model. Forecasting is a crucial technique for determining the acreage, output, and productivity of any crop in the near future. Tripathi *et al*., (2014). Used historical data from 1950–51 to 2008–09, univariate autoregressive integrated moving average (ARIMA) models were used to anticipate the rice area, production, and productivity of Odisha. These predictions were then compared to predictions for all of India. Many other crops were forested using ARIMA Model.

The primary objective of this research is to identify an appropriate time series model for Pakistan's sugar production data and utilize it to analyse and predict the sugar yield from 2022 to 2030. The findings of this study are intended to aid in decision-making related to future sugar production in Pakistan.

2. Data and methodology

Data related to sugar production in Pakistan is acquired from Directorate of Crop Reporting Service, Pakistan. Data is considered for the spans 1974 to 2021. The 48 years' worth of annual data regarding Pakistan's production of sugar is statistically analysed. In the literature, a variety of models have been used to predict time series data; however, this study uses ARIMA modelling to predict Pakistan's sugar yield. It is the most all-encompassing type of stochastic model for studying time series data. The ARIMA models incorporate moving average (MA), differencing (or integrated), and autoregressive (AR) elements. If a model exclusively includes autoregressive terms, it is referred to as an AR model. An ARIMA model is used when a non-stationary series is brought into stationary behaviour via differencing. 'p' stands for the order of the autoregressive process, 'q' stands for the order of the moving average process, and 'd' specifies the order of differencing the series to make it stationary. This is the general version of ARIMA.

The general form of the AR process for order p is represented by the symbol AR (p) as follows:

$$X_t = \Theta_1 x_{t-1} + \Theta_2 x_{t-2} + \text{--------} + \Theta_q x_{t-q} + z_t \quad (1)$$

The following is the general form of the MA process of order q:

$$X_t = Zt + \Theta_1 z_{t-1} + \Theta_2 z_{t-2} + \text{-----} + \Theta_q z_{t-q} \quad (2)$$

In general, Z_t is regarded as an ARMA (p, q) model if it is stationary for time t since it contains both the AR and MA features. Here is the ARMA (p, q) model:

$$X_t = \Theta_1 x_{t-1} + \Theta_2 x_{t-2} + \text{-------} + \Theta_q x_{t-q} + z_t + \Theta_1 z_{t-1} + \Theta_2 z_{t-2} + \text{--------} \Theta_\mu z_{t-\mu} \quad (3)$$

Model identification, model estimate, diagnostics, and forecasting are the four main steps in the ARIMA process as shown in Figure 1. First, make sure our variable of interest is stationary. Additionally, low Akaike information criteria, Bayesian information criteria MSE, MAPE, and other metrics are used to identify the model's best suitable coefficients. In the following stage, forecasting, the model's performance is finally validated by looking at the residuals using an ACF plot of residuals and a Ljung Box test.

Initially, data is presented in the form of line chart in Figure 2 to visualize the stationarity of data. Cleary, we see that there exists increasing trend in the sugar production data. This also indicates that the sugar production data is not stationary. To check the stationarity in the data, we make a plot at first differences, as shown in Figure 3. We observe that the sugar production data becomes stationary after taking first difference. This also helps to determine one component of the ARIMA model that we are going to use in the following to search for a suitable ARIMA model. To search for other suitable parameters of ARIMA model, that is, AR and MA terms, we use the information of ACF and PACF. Note that PACF is used to choose the appropriate AR term and ACF is used to choose the MA term. Both ACF and PACF are shown in Figure 4.

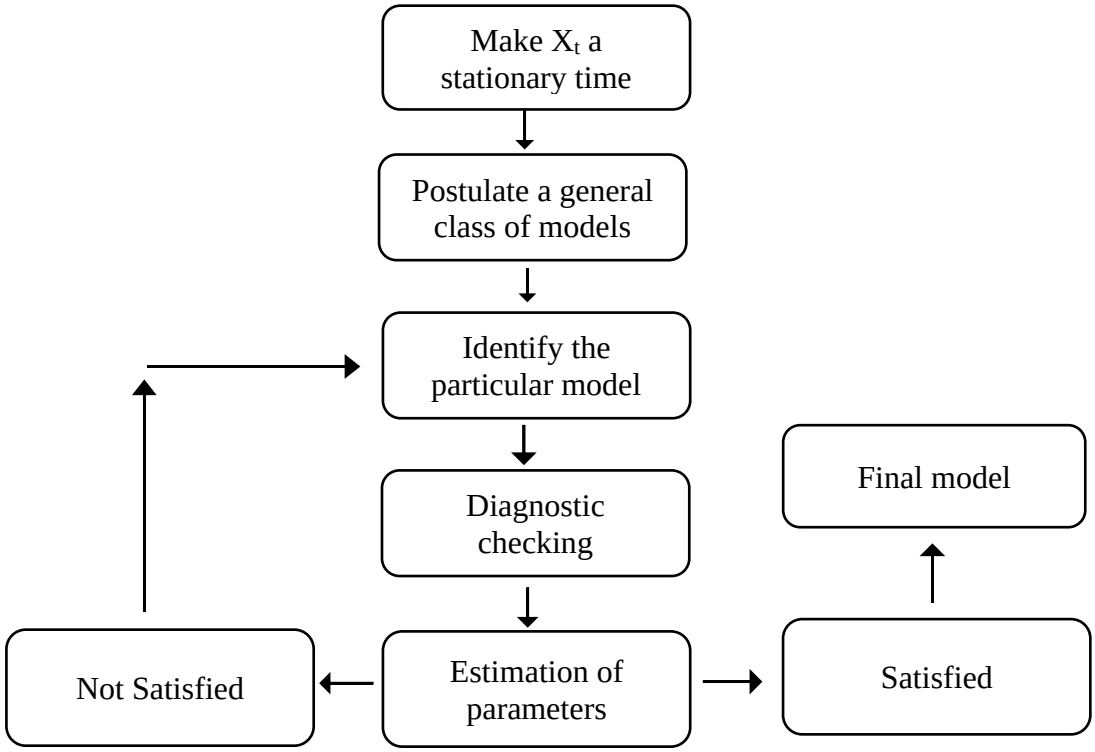

Figure 1: Flow chart of Box and Jenkins methodology.

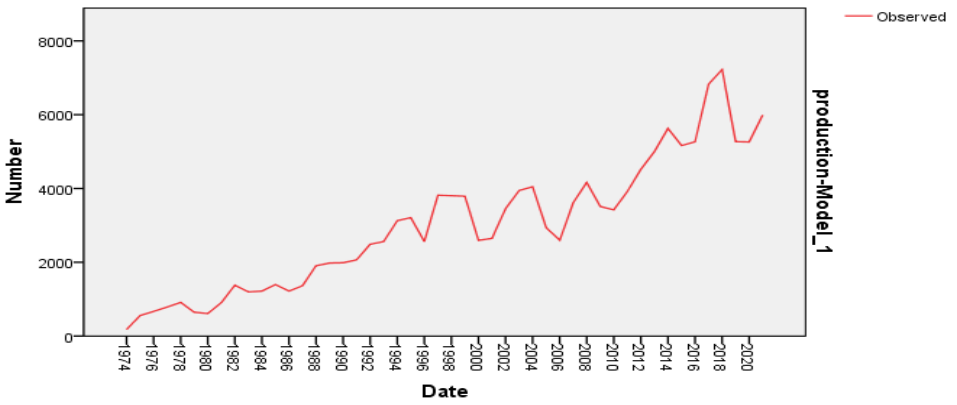

Figure 2: The visualization in the form of line chart of sugar production data.

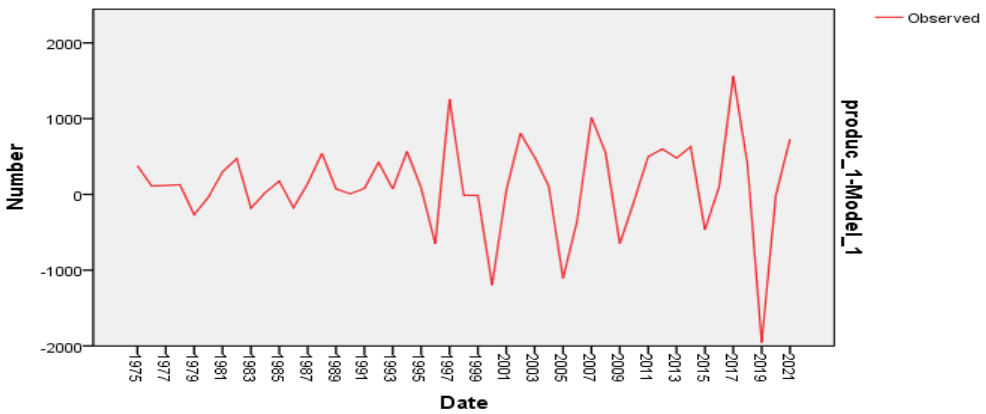

Figure 3: Visualization of sugar production data at first difference.

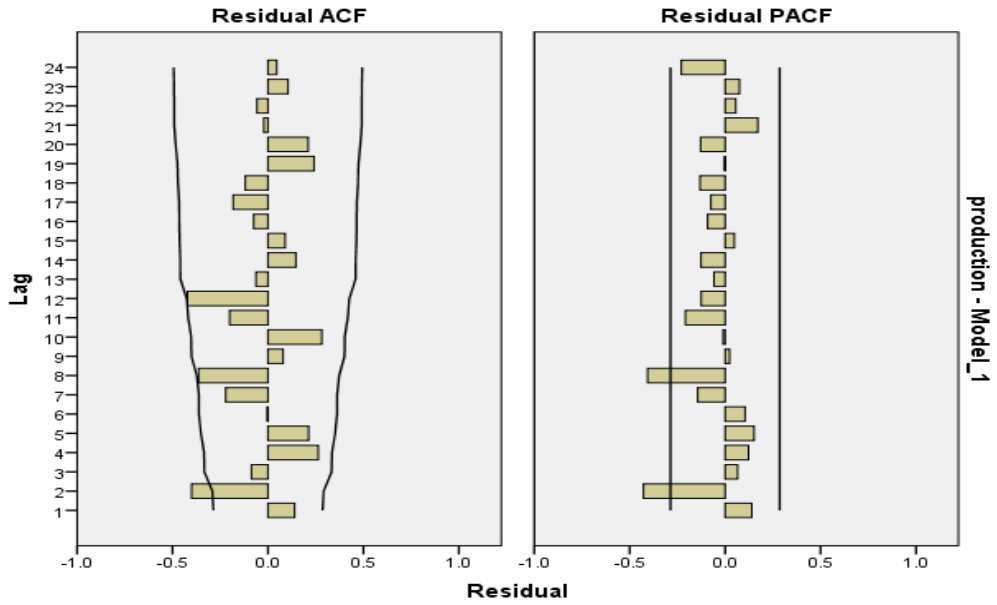

Figure 4: ACF and PACF of sugar production data.

Table 1: Fitted ARIMA models and BIC.

Models	BIC
ARIMA (1,1,1)	12.800
ARIMA (0,1,1)	12.912
ARIMA (3,2,1)	12.871
ARIMA (1,1,0)	12.961
ARIMA (2,2,1)	12.853
ARIMA (2,2,0)	13.329

From these plots, we choose six candidate models that are shown in Table 1. AIC and BIC are two most frequent used criteria to choose the best fitted model. In this case, we have considered the BIC criterion for this purpose since Shibata (1976) demonstrated that the order of auto regression tends to overestimate by AIC. The BIC information along with different fitted models is shown in Table 1. Clearly, we see that the model ARIMA (1, 1, 1) achieve the minimum value of BIC. As a result, this model is considered to be the best fitted model in the class of ARIMA models to model and forecast the sugar production of Pakistan.

One of the goals of a fitted model is to forecast the future values once its parameters have been identified. In our case, we have taken the forecast for sugar production for next ten years using fitted model ARIMA (1, 1, 1). The forecasted values and their 95% confidence intervals are shown in Table 2 and Figure 6.

Figure 6 shows the actual production of sugar in Pakistan during the period of 1974-2021 and the forecast for the period of 2022-2030, along with 95% confidence band. We can see that the forecast follows the trend of the data. The sugar production will be reached at 6881 tons.

Table 2: Forecasted values of sugar production of Pakistan using ARIMA (1, 1, 1).

Year	Forecast	Lower limit	Upper limit
2022	6021	4940	7102
2023	6094	4864	7324
2024	6189	4916	7462
2025	6296	5009	7583
2026	6409	5117	7701
2027	6525	5231	7819
2028	6643	5348	7937
2029	6762	5467	8057
2030	6881	5586	8176

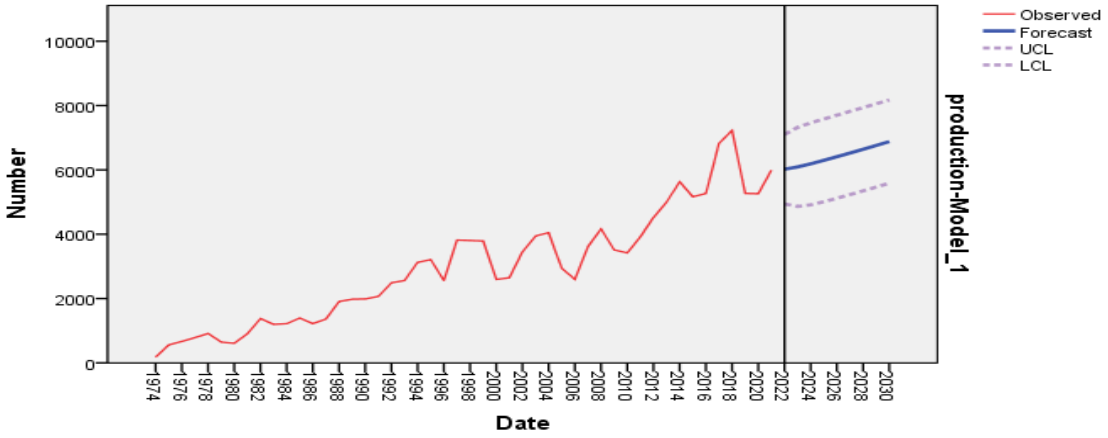

Figure 6: Visualization of forecasted values of sugar production of Pakistan.

3. Conclusion

Pakistan's economy, much like other developing countries, relies heavily on the agriculture sector, with crops playing a significant role. Among these crops, sugarcane dominates the cash crops. To forecast the sugar production for the period of 2022-2030, this study utilized the minimum criterion of BIC and developed an ARIMA (1, 1, 1) model, which was found to be the best fit. The study concluded that the forecast aligns with the historical trend and can be used for future policies regarding sugar production.

Furthermore, this research aimed to develop an ARIMA model that can predict Pakistan's annual sugar production from 1974 to 2021, as well as provide forecasts for upcoming years. The study's 95% confidence interval sugar forecast for Pakistan from 2021 to 2030 predicted a sugar production of 6881 million tons in 2030. In addition to the forecast, the study also discovered a rising trend in sugar production in Pakistan. Therefore, the findings of this study can be used to aid decision-makers in Pakistan's sugar industry.

References

1. Ali, S., Badar, N., and Fatima, H. (2015). Forecasting production and yield of sugar cane and cotton crops of Pakistan for 2013-2030. *Sarhad Journal of Agriculture*, 31(1), 1-10.
2. Ho, S. L. and Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers and industrial engineering*, 35(1-2), 213-216.
3. Hossain, M. M., and Abdulla, F. (2015). Forecasting the sugarcane production in Bangladesh by ARIMA model. *Journal of Statistics Applications and Probability*, 4(2), 297.
4. Mehmood, Q., Sial, M. H., Riaz, M., and Shaheen, N. (2019). Forecasting the Production of Sugarcane in Pakistan for the Year 2018-2030, using Box-Jenkin's Methodology. *JAPS, Journal of Animal and Plant Sciences*, 29(5), 1396-1401.
5. Nadeem, M., Hussain, A., and Ur Rehman, H. (2019). Forecasting of sugar production in Pakistan using time series analysis. *Journal of Applied Research on Industrial Engineering*, 6(4), *292-304.* doi: 10.22105/jarie.2019.193814.1129
6. Sharma, P. K., Dwivedi, S., Ali, L., and Arora, R. K. (2018). Forecasting maize production in India using ARIMA model. *Agro Economist-An International Journal*, 5(1), 1-6.
7. Tripathi, R., Nayak, A. K., Raja, R., Shahid, M., Kumar, A., Mohanty, S., and Gautam, P. (2014). Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models. *Advances in Agriculture*, 2014.
8. Vishawajith, K. P., Sahu, P. K., Dhekale, B. S., and Mishra, P. (2016). Modelling and forecasting sugarcane and sugar production in India. *Indian Journal of Economics and Development*, 12(1), 71-79.
9. Wei, W. W. (2006). *Time Series Analysis Univariate and Multivariate Methods.* New York : Greg Tobin.