# Time Series Analysis on Monthly Beer Production in Australia

Nuolai Xu*

University of California, Santa Barbara, Santa Barbara, United States

*Corresponding author email: nuolaixu@ucsb.edu

**Abstract.** Despite Australia's high enthusiasm for beer consumption, domestic beer production falls short of meeting the demand, creating business opportunities for potential investors. In this research, the method Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used to analyze a historical beer production data in Australia from January 1956 to December 1994 in each month. The optimal SARIMA model is identified through a rigorous process involving time series conversion, seasonal component removal, and data differencing to address non-stationarity. Residual analysis confirms the model's effectiveness. By dividing the dataset into training and test sets, the SARIMA model forecasts the next 24 periods, providing valuable insights for potential investors seeking optimal entry points into Australia's beer market. Results reveal a notable upward trajectory in monthly beer production, highlighting untapped growth potential within the domestic beer industry. The study's findings not only contribute to forecasting future production trends but also offer strategic guidance for investors, shedding light on sustained growth prospects in the Australian beer market.

**Keywords:** Time series; beer production; prediction; SARIMA.

## 1. Introduction

The history of human brewing can be traced back thousands of years. After a long period of development, alcohol beverage has also differentiated into many different types, among which beer is the most popular potable liquid in the world, its ranking only below tea and water. [1]. People are very willing to consume this type of alcohol, which contains relatively less alcohol, in various scenarios. In 2009, beer enthusiasts around the world consumed a total of 177.27 million kiloliters of beer [2]. Of this, Australia is an important component of this vast consumer market, consuming 1869 thousand kiloliters of beer, and each Australian consumed 87.9 liters of beer, which ranked seventh on the 2009 Global Per Capita Beer Consumption Ranking [3]. Such high beer consumption not only brings a unique beer culture to Australia but also brings business opportunities. Although the domestic population has a high enthusiasm for beer consumption, the production level of beer in Australia is not sufficient to meet the supply. In 2009, Australia only produced 1775 thousand kiloliters of beer, ranking 22nd among the beer-producing countries that year [4]. In this situation, investing in the beer industry in Australia may be a good idea, so reliable data analysis is required. For example, predicting the future beer production in Australia to assess investment risks. In this field, the SARIMA model for time series is a good choice.

A time series can also be called a sequence of data points essentially. It arranged in chronological order and usually in a constant time interval (such as 1 second, 3 minutes, and also in hours and days, etc). In that case, time series is similar to the discrete-time data so can be analyzed and processed in same methods [5]. A time series is a collection of time observations. Through relevant research, people can obtain future trends and potential, and use these results to provide a basis for different applications, such as production planning, control, optimization, etc. [6]. Many models have been proposed to solve the problem, such as the ARIMA (Auto Regressive Integrated Moving Average) model and also the SARIMA (Seasonal Auto Regressive Integrated Moving Average) model which is developed on this former basis [7-9]. Compared to ARIMA, the difference between them is that SARIMA has an extra set of autoregressive components and moving average components. These two additional components allow SARIMA models differencing data by non-seasonal differencing and by frequency seasonally. Thus, it will be much easier to know which parameters are the best [10]. Seasonality refers to repetitive patterns with a fixed frequency in data, such as daily, bi-weekly, and every four months. The research protocol involves utilizing the SARIMA (p, d, q) x (P, D, Q) model, where these letter parameters are determined based on data characteristics and model diagnostics. After outlining the optimal SARIMA model identified and the subsequent steps in data processing and analysis, Australia's beer production in the coming period will be predictable. Through intuitive images, the answer will be provided: Will the identified trend in Australian beer production, characterized by sustained growth, indicate untapped potential within the industry? The implications extend to providing prospective investors with valuable insights into the enduring growth prospects inherent in the local beer market.

To conclude, this study aims to utilize the programming language R to build a robust SARIMA model, offering a nuanced understanding of past beer production trends in Australia and, crucially, predicting future fluctuations. By combining data analysis with predictive modeling, this research contributes not only to understanding the dynamics of the Australian beer industry but also to inform strategic decisions for investors seeking opportunities in this thriving market.

## 2. Methods

### 2.1. Data Source

The data set for this paper is collected from the Kaggle website, which was updated by Shenbaga Kumar in 2018 for 476 observations of two variables.

### 2.2. Variable Selection

The dataset contains 476 observations of two variables, providing information on the monthly beer production in Australia in a specific month, measured in some unit of quantity (e.g., liters or barrels, which the original data does not explain). There are two variables, Month and Monthly beer production. The first one indicates the time and the second one indicates how many beers is produced in the specific month. covering the time period from January 1956 to August 1995, spanning nearly 39 years. The data is recorded on a monthly basis, with 12 observations per year. In order to follow the frequency, we only selected the data from January 1956 to December 1994 to build the time series.

### 2.3. Research Protocol

This paper uses SARIMA model with parameters (p, d, q) x (P, D, Q). The SARIMA is developed from the ARIMA model that incorporates seasonal components, and it captures the autoregressive, differencing, and moving average components, accompanied by seasonal components denoted by capital letters (P, D, Q). The letter parameters are selected based on the data characteristics and model diagnostics. SARIMA models are effective for capturing linear and non-linear relationships in the data and handling seasonal patterns. The model is fitted to historical data and used to generate forecasts for future time points.

## 3. Results and Discussion

The optimal model identified for the dataset is ARIMA(0,1,3)(0,0,2)[12]. Upon importing the data from the CSV file, the resulting table, as depicted below, is created (table 1).

**Table 1.** Part of the original data set

| | Month | Monthly.beer.production |
|---|---|---|
| 1 | 1956-01 | 93.2 |
| 2 | 1956-02 | 96.0 |
| 3 | 1956-03 | 95.2 |
| 4 | 1956-04 | 77.1 |
| 5 | 1956-05 | 70.9 |
| 6 | 1956-06 | 64.8 |
| 7 | 1956-07 | 70.1 |
| 8 | 1956-08 | 77.3 |
| 9 | 1956-09 | 79.5 |
| 10 | 1956-10 | 100.6 |
| 11 | 1956-11 | 100.7 |

### 3.1. Original Time Series Visualization

Subsequently, the data is converted into a time series using the 'ts' function. As previously mentioned, the dataset encompasses monthly beer production records spanning from January 1956 to August 1995. The monthly production counts prompted us to set the frequency to 12. However, due to incomplete data for the year 1956, information from January 1995 onwards was excluded from the time series, resulting in the creation of another time series named 'beer_ts.' A visual representation of 'beer_ts' was then generated (Figure 1).

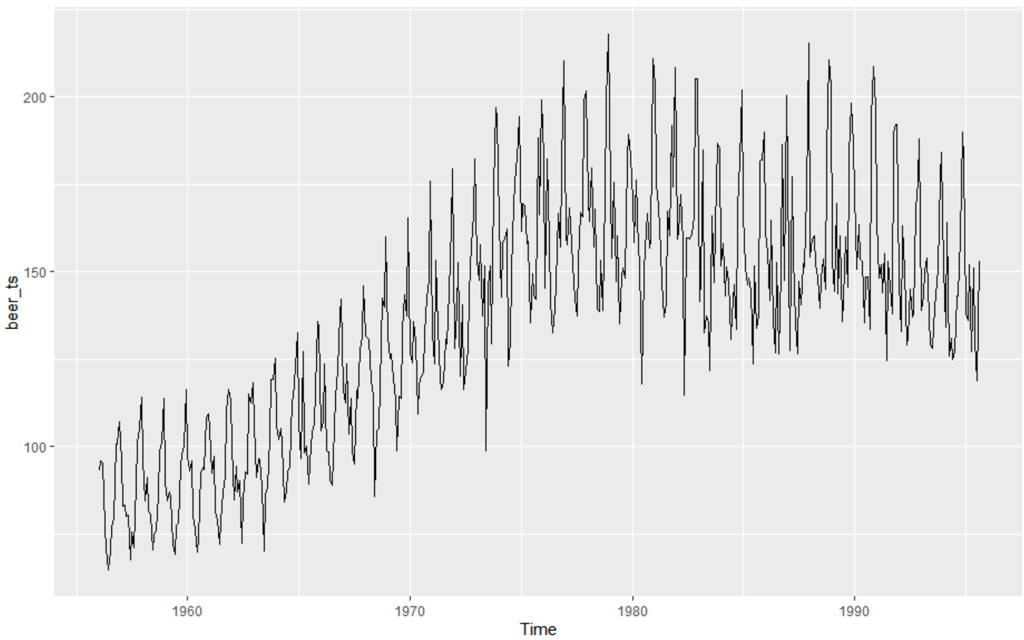


**Fig. 1** Original time series figure

Then we use the stl function in R to estimate the seasonal component of the data and remove it from beer_ts. The new data is called beer_ts_adj (Figure 2).

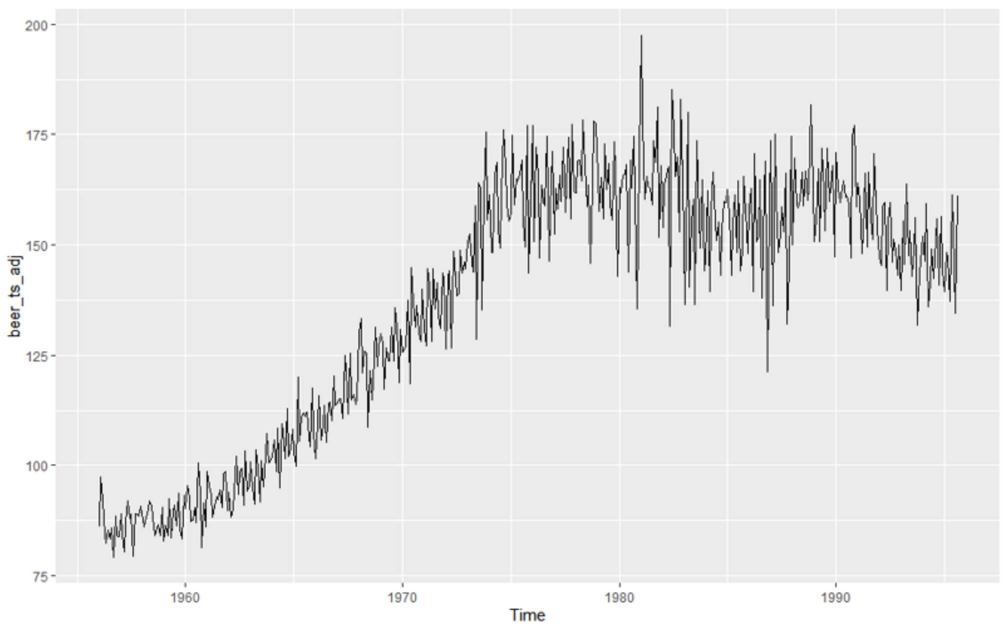


**Fig. 2** Adjusted time series figure

### 3.2. Stationarity Assessment

Then the stationarity of the dataset needs to be tested. Stationarity refers to the stability of properties in statistics, including mean, variance, and autocorrelation structure over time, etc. Stationarity is pivotal in time series analysis and it is a fundamental requirement to establish the effectiveness of the analysis. In a stationary time series, statistical properties remain constant, enhancing the reliability of the models. Non-stationary time series, which may exhibit trends or seasonality, can introduce disturbances that compromise the accuracy of predictions. Stationarity provides the foundation for building effective models, making accurate predictions, and performing robust statistical inferences.

Thus, the 'stl' function was employed to estimate and eliminate the seasonal component from 'beer_ts,' resulting in a modified dataset named 'beer_ts_adj.' The next step involved an testing of the Autocorrelation Function (ACF) as well as the Partial Autocorrelation Function (PACF) for 'beer_ts_adj.' The ACF is response for measuring the correlation between the lagged values of a time series and itself at different time intervals (Figure 3). By examining how correlated a series is with its past values, the ACF helps identify patterns, seasonality, and potential autocorrelations. The PACF, on the other hand, assesses the correlation between a certain lag and the series, removing the influence of shorter lags (Figure 4). It provides insights into the direct relationship between observations at different time points, helping to identify the true underlying dependencies in the data.

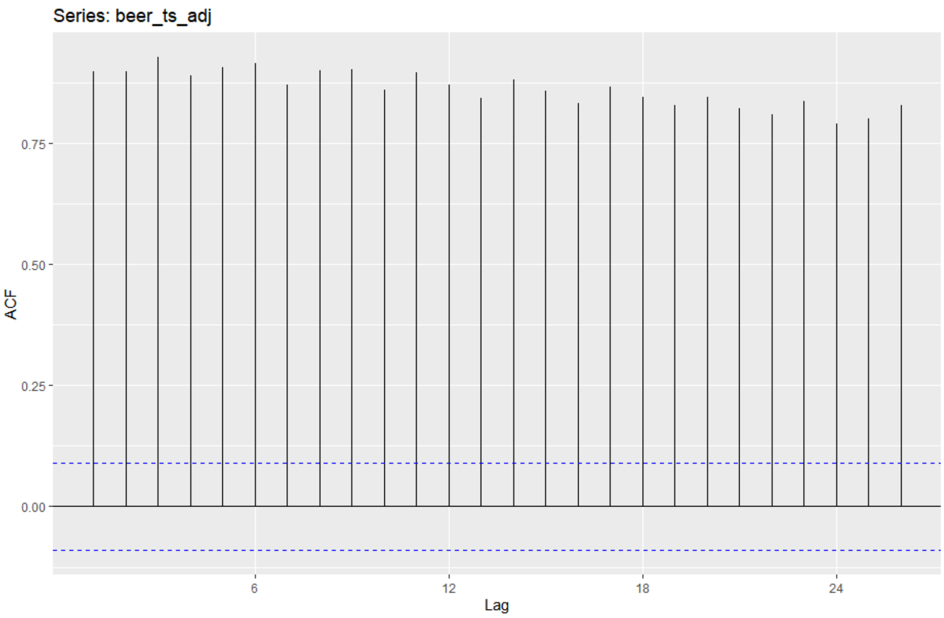


**Fig. 3** The auto-correlation function of beer_ts_adj

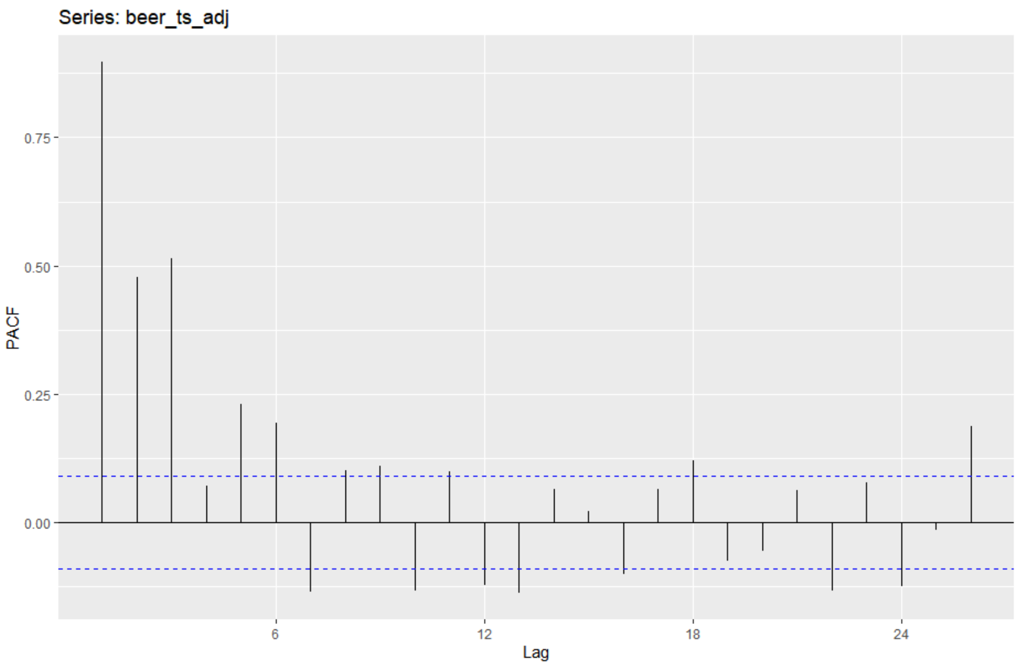


**Fig. 4** The partial auto-correlation function of beer_ts_adj

Unfortunately, the ACF and PACF plots did not exhibit desirable characteristics. A confirmatory analysis using the 'adf.test' and 'kpss.test' functions revealed that the time series was not stationary.

### 3.3. Differencing for Stationarity

To transform a non-stationary time series into a stationary one, you can employ differencing. Differencing involves computing the differences between consecutive observations. This helps remove trends or seasonality, making the series more stable. If a time series still exhibits non-stationarity after differencing, further differencing may be applied until stationarity is achieved. Consequently, differencing the data was performed, resulting in a new dataset named 'beer_diff,' the visual representation of which is presented in the plot (Figure 5, 6, 7).

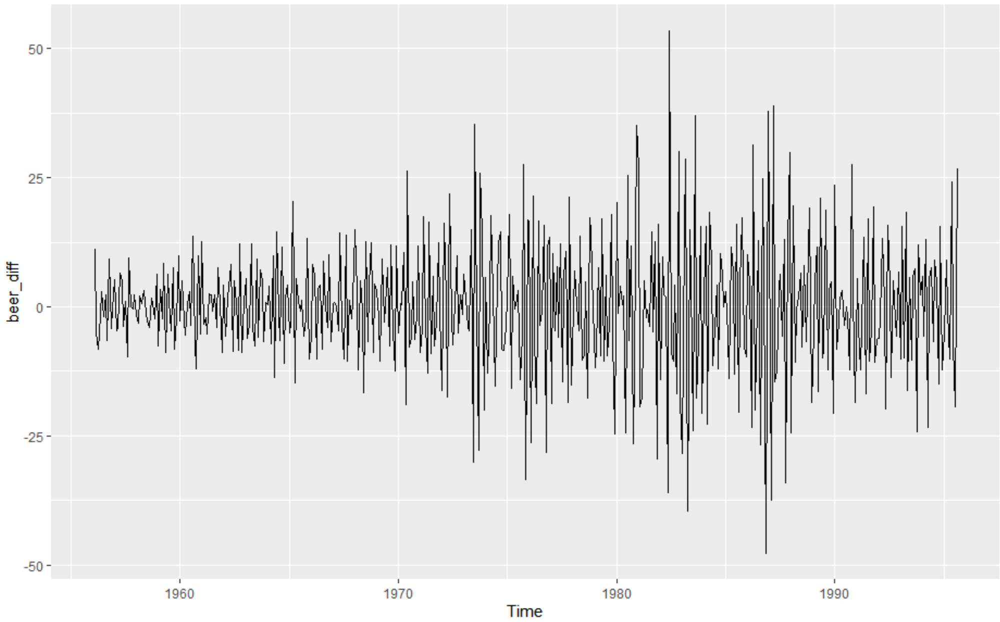


**Fig. 5** Differenced data set

Once again, the stationarity is tested, and now they look stationary.

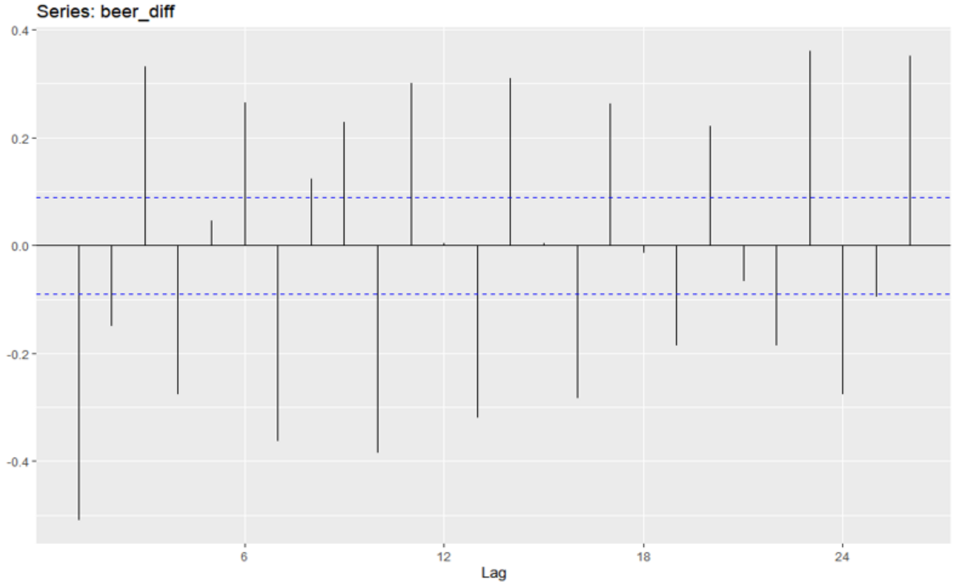


**Fig. 6** The autocorrelation function of differenced data set

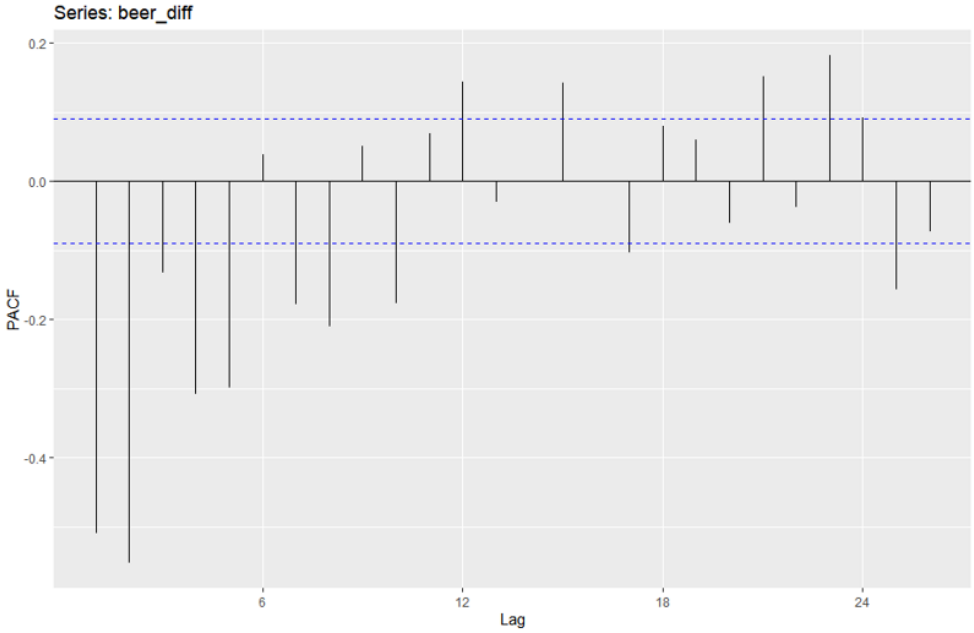


**Fig. 7** The partial autocorrelation function of differenced data set

### 3.4. Residual Analysis and Model Evaluation

In the subsequent analysis, an attempt was made to fit a Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Initially, an ARIMA model was constructed, followed by the development of the SARIMA model. The ACF and PACF plots of the SARIMA model's residuals were examined (Figure 8, 9, 10).

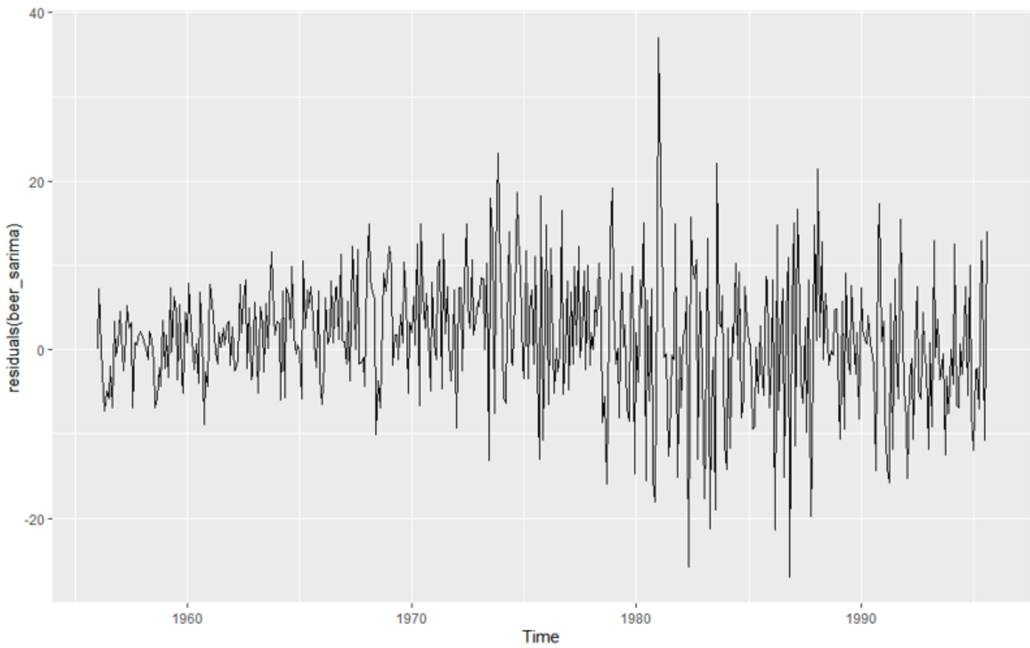


**Fig. 8** The residuals of model

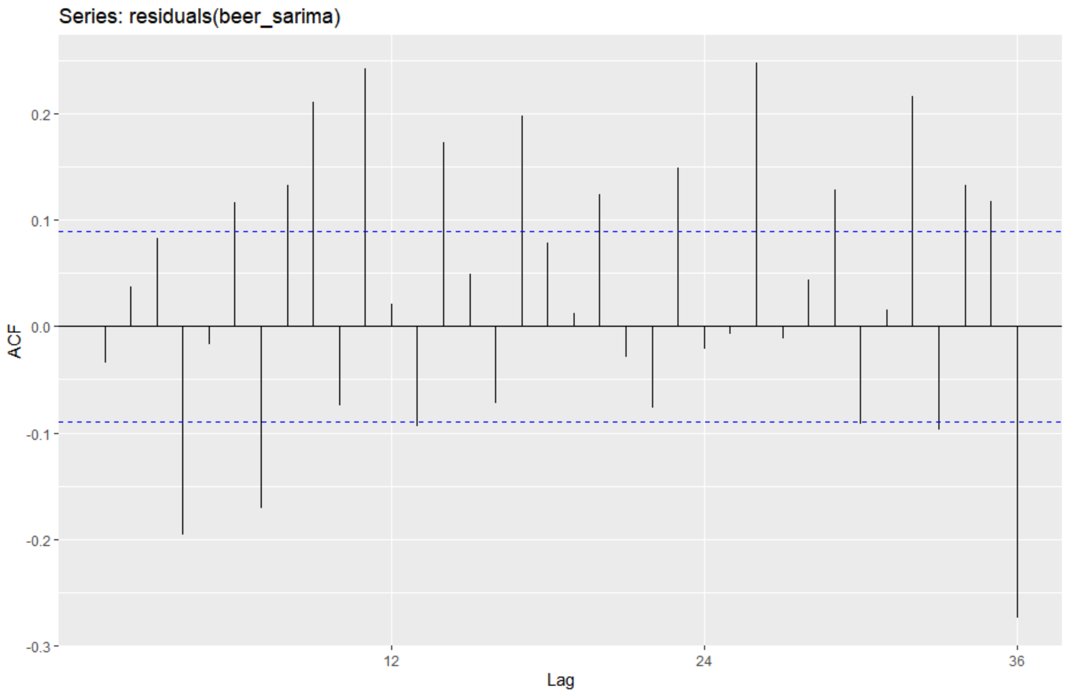


**Fig. 9** The auto-correlation function of the residuals

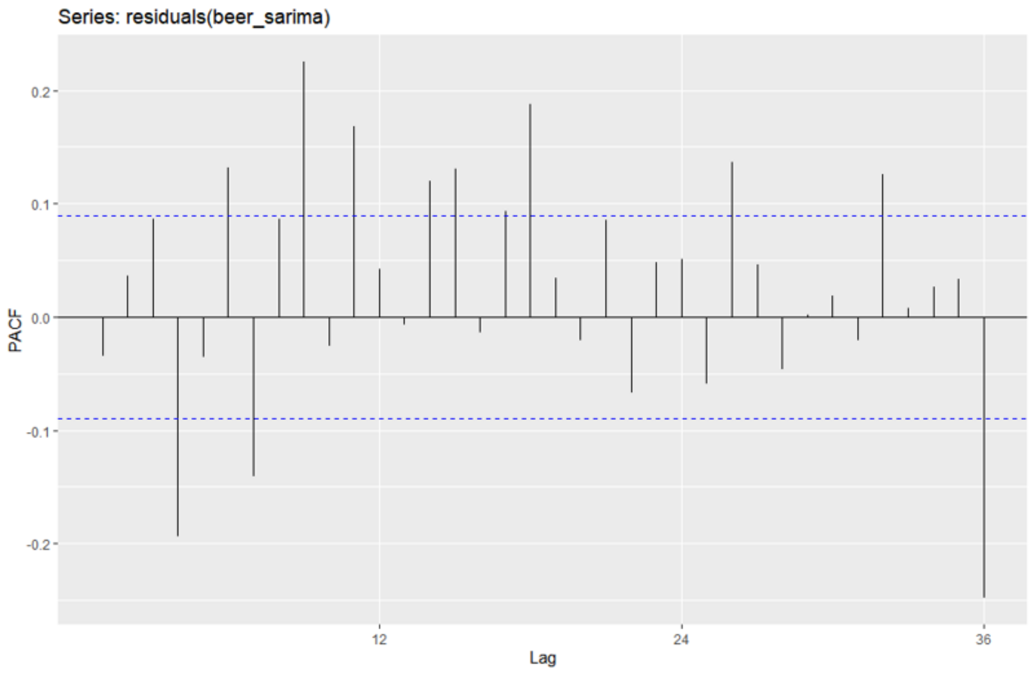


**Fig. 10** The partial autocorrelation function of the residuals

### 3.5. Forecasting Future Trends

The dataset was then divided into two parts to evaluate the model's behavior and usually be called as the training set and the test set. Once the effectiveness of the model is deemed qualified, the model was then utilized to forecast the upcoming 24 periods, representing a projection for the next 2 years. With this prediction, we can determine the growth cycle of local beer production and thus determine the optimal investment timing (Figure 11).

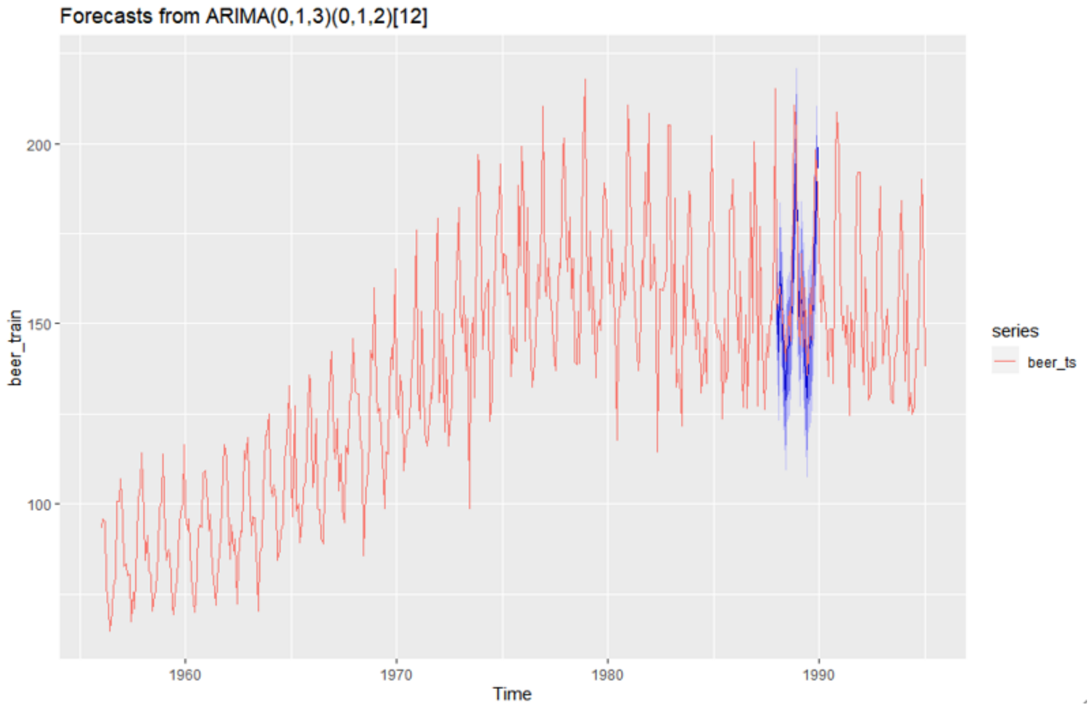


**Fig. 11** The forecast of future production

## 4. Conclusion

In the pursuit of understanding Australia's beer production dynamics, we leveraged the monthly beer production dataset spanning from 1956 to 1995, employing the powerful Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The discernment of SARIMA as the optimal choice was rooted in its exceptional performance in capturing and forecasting complex temporal patterns within the data.

The prognostications illuminate a compelling narrative. Beyond the recurrent seasonal oscillations, our analysis unveils a sustained upward trajectory in Australia's monthly beer production, a trend notably distinct from the preceding years leading up to 1995. This observation hints at untapped growth potential within the domestic beer industry, laying the groundwork for future expansion.

The implications of this discernment extend far beyond the immediate scope of our investigation. For prospective investors, these findings offer invaluable insights into the enduring growth prospects intrinsic to the local beer industry. The identified trend not only informs strategic decisions regarding market entry but also underscores the resilience and promising future of Australia's beer production landscape. As we delve deeper into the uncharted territory of this industry, our study not only contributes to forecasting future production trends but also provides a roadmap for stakeholders seeking to capitalize on the sustained growth within the Australian beer market.

## References

[1] Nelson M. The Barbarian's Beverage: A history of beer in ancient Europe. Routledge Taylor & Francis Group, 2005.

[2] Kirin Holdings. Global Beer Consumption by Country in 2009. Kirin Institute of Food and Lifestyle Report, 2010.

[3] Box G E P, et al. Time Series Analysis: Forecasting and Control. Hoboken, NJ, USA:Wiley, 2015.

[4] Han Z, et al. A Review of Deep Learning Models for Time Series Prediction. in IEEE Sensors Journal, 2021, 21: 7833-7848.

[5] Warren Liao, Chang P C. Impacts of forecast inventory policy and lead time on supply chain inventory-A numerical study. International Journal of Production Economics, 2010, 128: 527-537.

[6] Liljana Ferbar Tratar. The modeling method study of fault fuzzy forecast system. Computer Measurement and Control, 2002, 10: 23-25.

[7] Liu C, Hoi S C H, Zhao P. Online ARIMA algorithms for time series prediction. 13th AAAI Conf. Artif. Intell, 2016, 1-7.

[8] Brendan Artley. Time Series Forecasting with ARIMA , SARIMA and SARIMAX-A deep-dive on the gold standard of time series forecasting. Working paper, 2022.

[9] Yue Ming. Prediction and Development Strategy of Liquor Market Based on Random Forest and Rule Integration Method. Tianjin University, 2023.

[10] Lv Xiaoguo. Sales Forecast Analysis of a Brand of Beer in Nantong. Occupational Time and Space, 2007, 3(2): 2.