Analysis and Prediction of Online Beer Sales Based on SARIMA Model

Shiyuan Wang[1, *]

[1]Shanghai Lixin University of Accounting and Finance, Shanghai, China

*Corresponding author: 161105104@stu.cuz.edu.cn

Abstract. With the boom of e-commerce in China, online shopping has become the mainstream way of shopping in Chinese. To explore the impact of online shopping on beer sales, this paper uses a time series SARIMA model to analyze online beer sales data from January 2020 to September 2022 obtained from Internet platforms and predicts online beer sales from October 2022 to September 2023. This paper first introduces the current research on beer sales in China, and then briefly analyzes the current situation of the beer industry. Thirdly, based on the real data of beer online sales on the Internet platform, SARIMA model is used to forecast the sales volume of next year. The result shows that beer online sales are expected to show an upward trend, with the industry being the most competitive in June 2023, and a small sales peak both in November 2022 and January 2023 due to the e-commerce carnival. Therefore, beer online sales are significantly affected by seasonality and platform promotions.

Keywords: Time Series; Beer Online Sales; SARIMA Model.

1. Introduction

According to the economic index data of the beer industry released by the China Alcoholic Drinks Association, the production of beer enterprises above designated size in China in 2021 was 35.62 million kiloliters, a year-on-year increase of 5.60%; sales revenue in year 2021, which was RMB 158.48 billion, increased 7.91% compared to year 2020; the profit registered a 38.41% growth to reach RMB 18.68 billion. China's beer market is highly concentrated, which has always been divided by five major groups, namely China Resources Beer, Tsingtao Beer, Yanjing Beer, AB InBev and Carlsberg. In 2020, the market share of these five major groups was up to 92%, but the regional division is obvious. For example, China Resources Beer sales are concentrated in Northeast China, Sichuan, Guizhou, Anhui and Jiangsu. Tsingtao Beer has obvious advantages in Shandong Province and the provinces along the Yellow River. While AB InBev has a high market share in coastal and South China, and the main market of Carlsberg is in the southwest region. In addition, there are some local brands with a high local market share, such as Venus Beer and Pearl River Beer, but they are not well known nationwide.

Currently, the sales channels of China's beer industry are divided into offline channels and online channels, of which offline channels mainly include Key Account channels and traditional channels such as stores and restaurants. Online channels are mainly e-commerce. Although the offline channel is still the main sales channel of the beer industry, from the perspective of the proportion of channels over the years, the share of offline channels is constantly decreasing, replaced by online channels dominated by e-commerce. In 2011, beer accounted for only 0.3% of sales in e-commerce channels, and in recent years, with the rapid development of e-commerce, the growth rate of online channels in the beer industry from 2016 to 2020 almost reached more than 20%. During the epidemic in 2020, online beer sales increased by 21.98%. In the future, online channels may become the beer sales channel with the greatest growth potential in the next stage.

Therefore, by analyzing the various factors affecting beer sales, this paper establishes an appropriate forecast model to pre-evaluate the online sales of beer in the future, so as to provide a judgment basis for whether beer companies should increase marketing investment online and help some beer companies improve the adjustment of sales channels.

2. Literature Review

At present, the research on beer sales at home and abroad mainly focuses on sales channels and marketing strategies. Tian Yongjie used the basic theory of channel conflict management to put forward implementation suggestions for the channel conflict management of Blue Horse Beer by combining short-term channel conflict and long-term channel conflict [1]. Wang Yunsheng evaluated the promotional effect of Le Cordon Bleu in the Heilongjiang market and gave suggestions on sales strategies via qualitative and quantitative analysis [2]. Wang Yijun conducted an actual analysis based on the current situation of QD beer online marketing and demonstrated the improvement strategy by combining marketing theory and actual business data analysis [3]. Lang Lijun conducted research on the marketing status of Yanjing Beer in the Xinjiang province, analyzed its macro and micro-environment, and combined it with relevant marketing mix theories to formulate marketing strategies for it in this market [4].

However, there are less research in beer demand and sales forecasting, which mainly focusing on the time series forecasting method. Qiao Shuhong and Sun Debao used a forecast-weighted average model calculated by the seasonal index method based on moving averages and chain indexes to forecast sales in the beer industry [5]. Chen Yuke applied a time series model to predict beer sales in the next year based on the beer sales data of a city for seven consecutive years [6]. Yang Junqi obtained historical data by POS machine and other tools to analyze the market seasonal law of beer and predict the data mining algorithm of the future sales market [7]. Lv Xiaoguo made a prediction on the sales volume of beer in the sixth year based on the five-year monthly sales data of a certain brand of beer in Nantong, using trend prediction analysis and autoregressive prediction analysis methods, and compared the two forecasts, believing that the prediction analysis was more effective [8]. Based on the characteristics of beer sales, using the time series and PERT model, Hu Yanjie forecasted the sales data for the first four months of 2010 by analyzing the sales data from 2008 to 2009 [9]. Mao Hui combined the James-Stein compression method with the Dlahart group seasonal forecasting method and applied it to the sales forecast of a large beer company in Wuhan [10].

3. Fundamental Analysis of Beer Industry

3.1 Analysis of Production and Sales

According to Euromonitor, as shown in the figure below, the average growth rate per annum of global beer sales from 2007 to 2020 was only 0.65%, and beer sales in most countries have basically stopped growing, of which China's average annual growth rate of 1.51% has been at a high level, but still showing a downward trend. Affected by the epidemic in 2020, China's beer sales fell significantly, with sales of 42.78 million kiloliters, a year-on-year decrease of 7%.
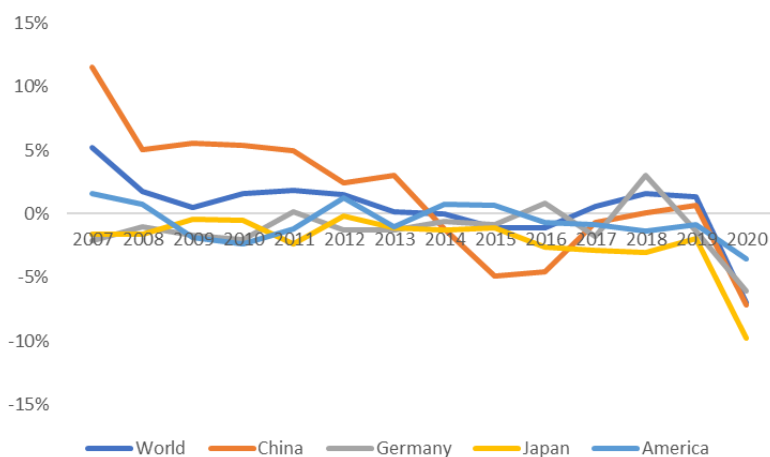

Fig. 1 Average Growth Rate of Sales in Different Countries from 2007 to 2020

China is a big beer producer and consumer. According to data from the National Bureau of Statistics, from 2013 to 2020, the production of beer in China had continued to decline. In 2020, it completed 34.11 million kiloliters, a year-on-year decrease of 7.0%. While beer production rebounded slightly in 2021, data showed that it reached 35.624 million kiloliters in 2021, an increase of 5.6% compared to 2020.

From the perspective of monthly sales, as shown in the figure 2, it can be seen that the sales of beer have obvious seasonal change laws, with parabolic distribution with weather temperature, with the largest sales volume in the hot summer and the lowest sales in the cold winter.
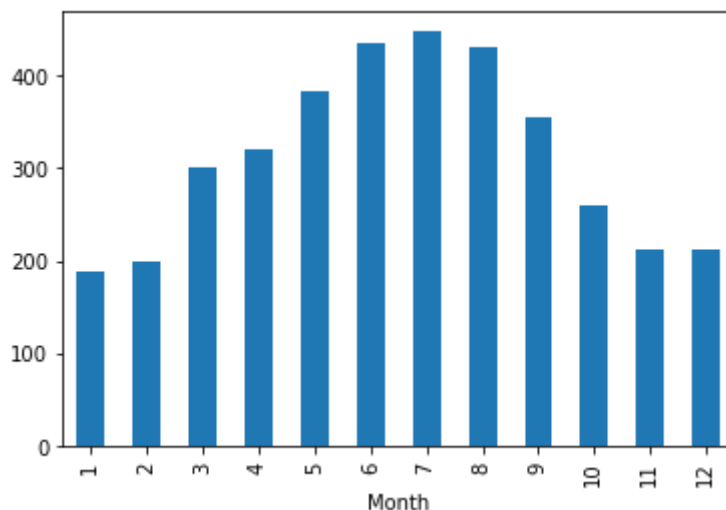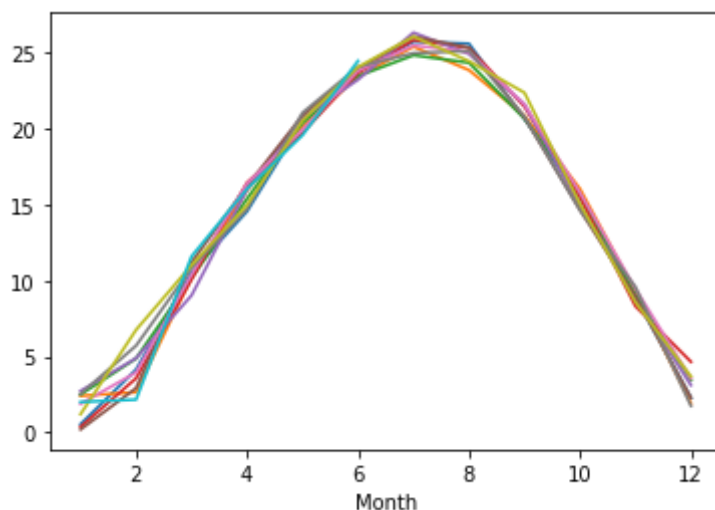

Fig. 2 Monthly Sales

Fig. 3 Average Temperature

3.2 Analysis of Area

China's beer production and sales have obvious regional differences. As matters stand, the beer production is mainly concentrated in North China, East China, Guangdong and Sichuan and the places centered on Shandong, while beer production in the western region is relatively small. Domestic beer production shows regional characteristics, and the relatively developed Bohai Rim, Yangtze River Delta, Pearl River Delta Economic Circle and their surrounding areas have relatively high beer production. Domestic beer production is mainly concentrated in the eastern coastal region.

According to data from the National Bureau of Statistics, from 2013 to 2020, the top five provinces in China's beer production were Shandong, Guangdong, Zhejiang, Henan and Liaoning. Among them, since 2000, Shandong and Guangdong have maintained the first and second positions in the beer

production of various provinces in the country. In 2020, Shandong ranked first in the country with a beer output of 4.58 million kiloliters, accounting for 13.4% of the country. Guangdong ranked second with 3.575 million kiloliters of beer production, accounting for 10.5%. From the changes in beer production in various provinces in the past two years, beer production in most provinces and cities has declined, and only a few provinces and cities have increased their output. Among them, the overall beer production of Zhejiang, Jiangsu, Anhui and Fujian in East China has increased in the past two years. Beer production in Henan, Liaoning, Heilongjiang, Hubei and other places has declined significantly.

The production data disclosed by the National Bureau of Statistics is basically consistent with the overall trend of sales. Beer companies tend to arrange production plans on the basis of sales volume. Thus, the industry as a whole is basically in a state of production and sales balance.

3.3 Analysis of Channel

According to data from Euromonitor, offline channels accounted for about 88.9% in 2020, still the mainstream channel for beer sales. Further subdivided, offline channels can be divided into non-existing channels, which are supermarket retail channels, including KA channels, traditional channels, etc., and existing drinking channels such as catering and entertainment, both of which account for about 50. High-end and ultra-high-end beer is mainly sold in catering and entertainment through ready-to-drink beer channels. However, in recent years, with the rapid growth of e-commerce, the proportion of online channels has gradually increased. In accordance with the statistics from Euromonitor, the compound annual growth rate of beer online sales from 2015 to 2020 was close to 27%, far faster than the growth rate of offline channels. E-commerce has become the third largest sales channel after offline independent stores and supermarkets, and the proportion continues to grow. It is possible for online sales to strongly promote the nationalization process of some beer companies, and even affect the industry competition pattern.

4. SARIMA Model's Prediction of Online Beer Sales

Although there are many ways to predict sales, from the basis of the forecast and the description of the market by the forecast results, the market forecasting methods are divided into two categories, which are qualitative forecasting methods and quantitative forecasting methods. Qualitative prediction methods mainly include derivation method, transduction analogy method and correlation factor method etc. Quantitative forecasting methods include causal analysis methods with market factors as independent variables, such as regression model forecasting, and time series analysis with time independent variables. However, after the above analysis, it is found that beer sales have strong seasonality, so it is more accurate to forecast with the time series method. Based on the actual sales volume of the market by month, this model makes month-by-month forecasts of future market trends.

The data used in this paper comes from the sales of the beer industry on e-commerce platforms. And the online sales data of beer companies from January 2020 to September 2022 are obtained from a big data platform crawler, which is shown in Table 1. In the following pages, based on historical data, the trend analysis method of time series is used to forecast the sales of beers with strong seasonality from October 2022 to March 2023.

Table 1. Online Beer Sales by month from 2020 to 2021

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2020	166	54	116	144	211	313	166	165	131	119	202	105
2021	181	91	113	156	213	328	159	171	173	137	156	128
Mean	173	72	115	150	212	321	163	168	152	128	179	117

In this paper, the SARIMAX function is used to adapt to the seasonal ARIMA model, and the time series model is established in monthly cycles. The modeling process is as follows:

(1) Draw a timing diagram to determine whether the series has obvious trends or periodicity and whether it contains seasonal components.

(2) Stationary analysis.

(3) Select the optimal parameter values of the SARIMA(p,d,q)(P,D,Q)S time series model.

(4) Validate the forecast sales volume for the coming year, from October 2022 to September 2023, and draw conclusions.

4.1 Analysis of Timing Diagram

The figure below selects the data from January 2020 to September 2022 for time series forecast analysis. It can find that the online sales of beer have a clear upward trend from January 2020 to September 2022, while there is a sharp decline after June 2022.
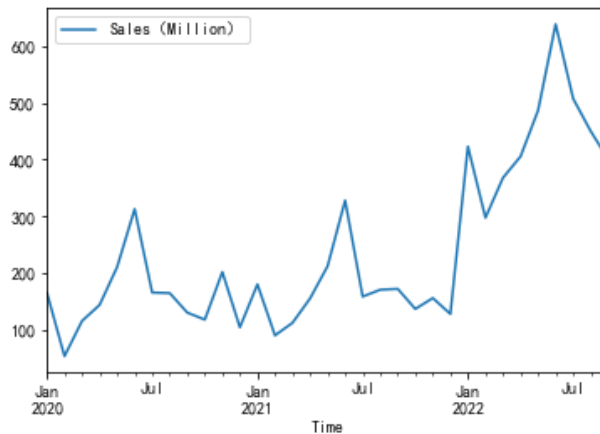

Fig. 4 Online Beer Sales from January 2020 to September 2022

4.2 Stationary Analysis

Through the observation of the ACF and PACF of the data, it is found that the P value of the ACF tailing and the unit root test of ADF is 0.38>0.05, which means that the series is not stable. In other words, the beer sales data has obvious seasonality, which is consistent with the analysis of 3.1 Analysis of Production and Sales.

Compared to the second-order difference of the original data, it is found that the p-value of the ADF unit root test is 0.04<0.05 so that the sequence become stable after running second-order difference.
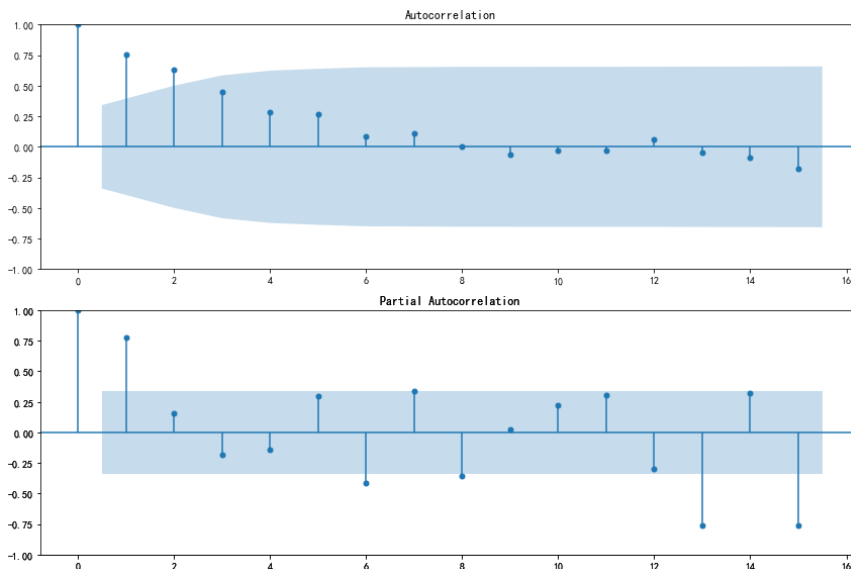

Fig. 5 ACF and PACF of Original Data

4.3 Parameter Selection

When selecting parameters for a seasonal ARIMA time series model, this paper uses Grid Search to iteratively explore different combinations of parameters. For each combination of parameters, it fit a new seasonal ARIMA model using the SARIMAX function of the statsmodels module and evaluate its overall quality. When evaluating and comparing statistical models equipped with different parameters, rank each parameter based on its suitability or ability to accurately predict future data points. It uses the Akaike Information Standard (AIC) value, which is returned by using the ARIMA model installed using statsmodels. AIC measures how well a model fits into the data, taking into account for the overall complexity of the model. In the case of a large number of features, models that fit the data will be given a larger AIC score than models that use fewer features to achieve the same fitness. Therefore, the parameters corresponding to the model that produce the lowest AIC value are the most suitable parameters for this model.

Using the SARIMAX function, the output results show that SARIMA(0, 1, 1) (1, 1, 1)12 yields the lowest AIC value of 64.41.

4.4 Validation and Prediction

Now it have obtained a model of the time series, which can use for forecasting. Comparing the forecasted values first with the actual values of the time series will help understand the accuracy of the forecast. Get the sales forecast values and associated confidence intervals for March 2002 to September 2022 by using get_prediction function and conf_int function, as shown in the following table.

Table 2. Actual Sales Values and Associated Confidence Intervals

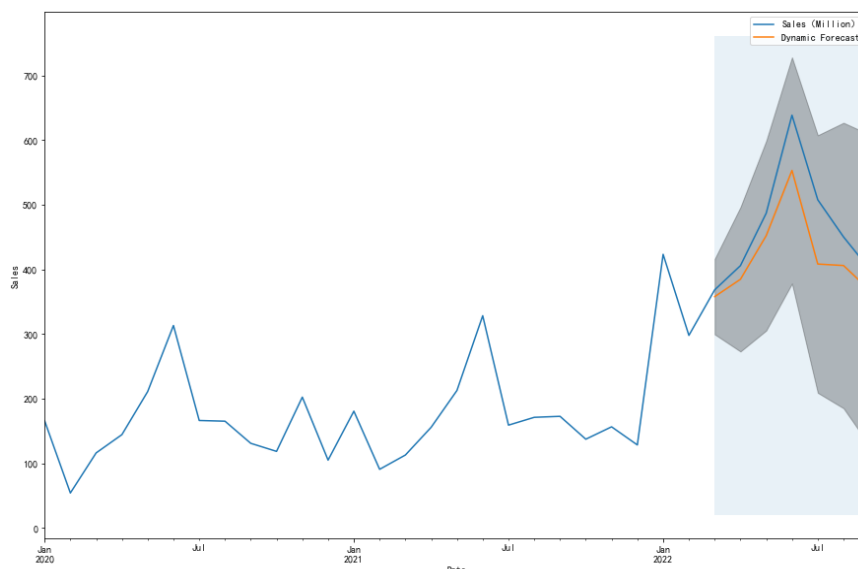| Date	Lower Sales （Million）	Upper Sales （Million）	Actual Sales （Million）
2022-03	299.48	416.10	368.26
2022-04	347.06	457.01	405.70
2022-05	422.70	529.17	486.81
2022-06	544.18	648.49	638.33
2022-07	477.17	580.03	507.56
2022-08	436.33	538.14	450.13
2022-09	329.51	430.51	400.75

Fig. 6 Actual Sales and Forecast Sales from January 2020 to September 2022

Then to calculate the mean predicted value, the equation is

$$\text{mean predicted value} = (\text{Lower Sales} + \text{Upper Sales}) \div 2 \tag{1}$$

Using dynamic forecasts to plot the graph, it can see that the trend between the actual and forecasted values is basically the same: sales have been rising month by month since March, peaking in June, and gradually decreasing in July.

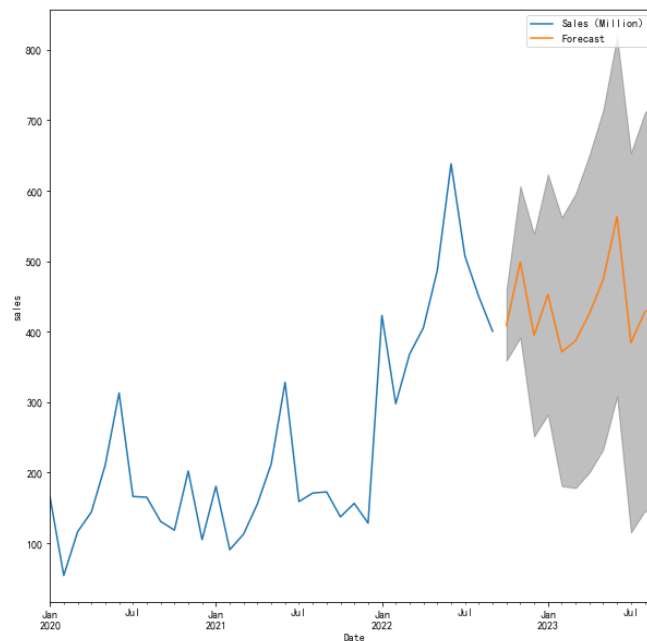Finally, used the seasonal ARIMA time series model to forecast sales from October 2022 to September 2023.

Fig. 7 Actual Sales and Forecast Sales from January 2020 to September 2023

It can be seen from the figure that from the perspective of the whole year, the sales volume of beer online channels has generally shown an increasing trend; On a monthly basis, online sales are expected to peak in June and reach small peaks in November and January.

5. Conclusion

In the context of the explosive development of big data technology and online shopping, the acquisition of big data has become more convenient, and online demand forecasting in the beer industry has become easier. By constructing a time series model, this paper predicts the change trend of beer online sales in the following year and provides theoretical support for the layout of beer enterprises in online channels.

This paper first introduces the current research on beer sales in China, and then briefly analyzes the present situation of the beer industry. Thirdly, it uses a SARIMA model to predict the sales in the next year based on the real data of online beer sales on the Internet platform.

The result shows that the method in this paper has practical significance for the study of beer online sales prediction. It suggests that beer online sales are expected to show an upward trend, with the industry being the most competitive in June 2023, and a small sales peak both in November 2022 and January 2023 due to the e-commerce carnival. Inferred from this, e-commerce carnival also is a key node for beer sales growth. In conclusion, online beer sales are significantly affected by seasonality and platform promotions.

Due to the limited data sources, the scope of this paper is not comprehensive enough, and choosing other time dimensions for modeling and prediction may obtain more accurate results, which needs

further research. The online sales data of beer in this article will be affected by consumer preferences and the epidemic, but because the time series forecasting method does not consider external factors for the time being, there is a defect of forecast error. And when it encounters large changes in the outside world, there will be a large deviation. In the future, the research on beer online sales data can increase the analysis of consumer preferences and epidemic dimensions, so that the problem research can be gradually improved.

References

[1] Tian Yongjie. Research on the Improvement of the Management of Marketing Channel Conflicts in Landmark Beer. Xi'an: Xi'an University of Technology, 2007.

[2] Wang Yunsheng. Heilongjiang Blue Ribbon Beer Differentiated Marketing Strategy. Wuhan: Huazhong Agricultural University, 2013.

[3] Wang Yijun. The Research of Networking Marketing Strategy of QD Brewery. Qingdao: Ocean University of China, 2015.

[4] Lan Lijun. Marketing Strategy Research of Yanjing Beer in Xinjiang. Shihezi: Shihezi University, 2017.

[5] Qiao Shuhong, Sun Debao. Analysis and Research on Sales Forecasts in the Beer Industry. Beer Science and Technology, 2000, (11):54-58.

[6] Chen Yuke. Trend Analysis and Forecast of Product Sales. Journal of Western Chongqing University (Nature Science Edition), 2003, (02):59-61.

[7] Yang Junqi, Xing Zhanlei, Li Qing, Xu Xiaoyan, Liu Xiaoli. Study on Seasonal Regular, Trend Value and Data Digging Method about Beer Market.Liquor-making Science & Technology, 2004, (06):104-107.

[8] Lv Xiaoguo. Forecast Analysis of a Certain Brand of Beer in Nantong. Career Horizon, 2007, (03):78-79.

[9] Hu Yanjie, Peng Lin, Xue Wenwen, Sun Yong. Research on Beer Sales Forecasting Method Based on Time Series and PERT. China Market, 2011, (15):91-94.

[10] Mao Hui. Study on Seasonal Forecasting Models and Applications for Short Time Series. Wuhan: Wuhan Polytechnic University, 2015.