Prediction Model of Physical Goods Sales based on Time Series Analysis

Huizhen Ge[1, a], Lina Fang[1]

[1]School of Business Administration of Liaoning University of Science and Technology, Anshan City, Liaoning Province 114000, China
[*] Corresponding author: Ge Huizhen (Email: 2579602854@qq.com)

Abstract: With the advent of the era of big data, most online historical consumption data are collected and simply stored in the cloud, but they are not receiving enough attention. This paper takes "online sales of physical goods" as the research object, uses "time series analysis related theory and sales prediction theory" to study and analyze "data from January 2020 to December 2021", and finally uses EViews software, using ARMA model for modeling and prediction. Establishing a mathematical model to effectively analyze and predict the "online physical goods sales ability", to a certain extent, has a practical guiding significance for e-commerce enterprises to make the optimal business decisions.

Keywords: Physical goods, Time series analysis, ARMA model, Online consumption prediction.

1. Introduction

According to the 49th Statistical Report on the Development of The Internet in China, as of December 2021, the number of online shopping users in China had reached 842 million. As a typical representative of the new business form of the digital economy, online retail still maintained a fast growth rate. In 2021, online retail sales of physical goods accounted for 24.5 percent of the total retail sales of consumer goods[1]. Therefore, how to effectively develop and use the existing network consumer consumption data will play a key role in the operation and decision of many e-commerce enterprises in China.

Demand analysis and prediction of network consumers, as the links that are paying more and more attention to by e-commerce enterprises, play an increasingly important role in providing a good decision-making basis for e-commerce enterprises. This article to "physical goods online sales" as the research object, using "time series analysis related theory and sales forecast theory", the "January 2020-December 2021 data" research analysis and establish a model, so as to get the forecast results, to help electricity enterprises achieve accurate production, reduce inventory costs and other costs, and obtain greater profit space.

2. Research Status

On the demand analysis of network consumers, scholars have explored from different aspects. Zuo Wenming, Chen Shao et al. (2019) analyzed the complex behavioral decision-making process and influencing factors of online consumers in B2C activities based on prospect theory and fuzzy multi-attribute decision-making[2]. Dong Yan, Time et al. (2020), based on consumer behavior and marketing theory, analyze the influence mechanism of online marketing activities on online consumer purchasing behavior from a psychological perspective[3]. Kansra P, Oberoi S (2022) used a questionnaire to analyze online purchases among young people[4]。

With the growth of online shopping users, it is of great significance to use the consumption data left on the Internet to predict the purchase demand of consumers for enterprises to achieve precision marketing. Zhang Jingxuan, Zhang Weiwei (2021), predicts the consumer repurchase behavior based on the neural network[5]. Li Weiqing, Chi Maomao et al. (2021) to predict the preferences of online consumers from four perceptual dimensions[6]. Dong J, Huang T, et al. (2022) analyze historical user data in promotions based on the BERT-MLP prediction model to predict repeated purchases of consumers after the promotion[7]. Rajendrakumar R (2022) obtained the data by simple random sampling and ensured the model fitting using SEM[8]。

The analysis and prediction of time-series models are applied in various scenarios. Wang Xue (2019) applied the time-series prediction model to high-level discipline inquiry[9]. Yan Jinghua et al. (2020) applied the time series prediction to theft crimes to adjust the community patrol strategy[10]. Wang Zhigang et al. (2020) proposed a prediction model based on the plan review technology (PERT), and created a combined prediction model combined with the time series analysis model, which has good results in predicting the sales volume of supermarket chains[11].

Ke Miao et al. (2021) summarized the commodity sales forecast into multivariable time series forecast problems, and used the historical sales data of an e-commerce online store to build LSTM and network model under the TensorFlow framework. The research results have important guiding significance for e-commerce enterprises to improve marketing decisions and reasonable inventory management[12].

3. Theoretical Basis

Based on the above literature review, this paper adopts the combination of sales prediction theory and time-series model to analyze the historical data of online sales of physical goods, so as to predict the purchase behavior of consumers thereafter. Sales forecast refers to making an estimate of the quantity and amount of a commodity sold at a certain time in the future. Time series refers to the series obtained by arranging the values of an index at different times in the order of time.

The time series model originated from the autoregressive (autoregressive AR) model proposed by Yule in 1927; the British scientist Walker then expanded it to a general model. Slutzky invented the moving average (moving average MA)

model; based on this model, Wold proposed the famous autoregressive moving average (auto regressive moving averageARMA) model in 1938[13-15]. Box's "Prediction and Control", published in collaboration with Jenkins (1970), perfected these models and introduced the ARIMA model in detail, from which the time series theory developed rapidly[16,17]。

Statistics (website: https: / / data.stats.gov. The cn /) "official website gets data for" online sales of physical goods ". This paper takes "online sales of physical goods" as the research object, takes "cumulative online retail sales from January 2020 to December 2021 and cumulative online retail sales of physical goods" as the sample data (Table 1), and the data is the time series data. Note: The physical goods here stand against the virtual goods on the network, that is, the sum total of the morphological and real goods sold on each platform on the network.

4. Data Processing of Online Sales of Physical Goods

4.1. The Source of The Data

After multiple access, from the " National Bureau of

Table 1. Initial data of online sales of physical goods

Database: monthly data. Time: Last 24 Months				
Time.	Cumulative value of online retail sales (billion yuan)	Cumulative growth in online retail sales (%)	Cumulative value of online retail sales of physical goods (billion yuan)	Cumulative growth in online retail sales of physical goods (%)
12/2021	130883. 5	14. 1	108042. 4	12
November 2021	118749. 2	15. 4	98056. 1	13. 2
October 2021	103765. 4	17. 4	84979. 3	14. 6
September 2021	91870. 9	18. 5	75041. 8	15. 2
August 2021	81226. 7	19. 7	66340	15. 9
July 2021	71108. 1	21. 9	58130	17. 6
June 2021	61132. 5	23. 2	50262. 7	18. 7
May 2021	48238. 9	24. 7	39376. 9	19. 9
April 2021	37638. 1	27. 6	30773. 5	23. 1
March 2021	28092. 7	29. 9	23067. 3	25. 8
February 2021	17586. 8	32. 5	14412. 2	30. 6
January 2021				
December 2020	117601. 3	10. 9	97590. 3	14. 8
November 2020	105374. 1	11. 5	87792	15. 7
October 2020	91275. 3	10. 9	75619. 3	16
September 2020	80064. 7	9. 7	66476. 9	15. 3
August 2020	70325. 9	9. 5	58650. 7	15. 8
July 2020	60785. 1	9	51017. 6	15. 7
June 2020	51500. 9	7. 3	43480. 7	14. 3
May 2020	40175. 8	4. 5	33738. 8	11. 5
April 2020	30698	1. 7	25751	8. 6
March 2020	22168. 8	−0. 8	18535. 7	5. 9
February 2020	13712	−3	11233. 2	3
January 2020				

4.2. Data Processing of Online Sales of Physical Goods

4.2.1. Outliers Handling Method

The causes of outliers are: weather effects, human error, and unscientific means of collection.

Elimination method: if a value violates the Raida criterion, that is, if the experimental data x is subject to a normal score $P（|x − μ| > 3σ） ≤ 0.003μσ$Then, if the data and data are the mean and standard deviation of the experimental data respectively, the data for or should be removed as outliers.$x > μ + 3σ$$x < μ − 3σ$

To replace the outliers with the data close to the real value, this is the substitution method. In general, the discrete-type data Mass or median substitution are available, while continuous data can be replaced with means.

4.2.2. Outlier Handling Method Used in This Article

The data in this paper are collected from real data, so

outliers may also exist and require outliers processing. This article removes outliers. Given that the outliers data in this paper are not large, using this method.

In this paper for the real data, online physical goods sales into the original time series for research problems, to establish the monthly cycle time series model, on the basis of the cumulative value of online sales and physical goods online sales cumulative value two dimensions for comparative analysis, through modeling prediction, on the line of physical goods demand forecast. Raw data are shown in Tables 2.

After the observation that the data for January 2020 and January 2022 were blank states, and the data for other time periods were normal, the data from January 2020 and January 2022 / 2022 were excluded. The time series effect after handling the outliers was better improved. The processed data are shown in Table 2.and the processed time series plots are shown in Figures1.2.

Table 2. The data after the processing

Database: monthly data

Time: Last 24 Months

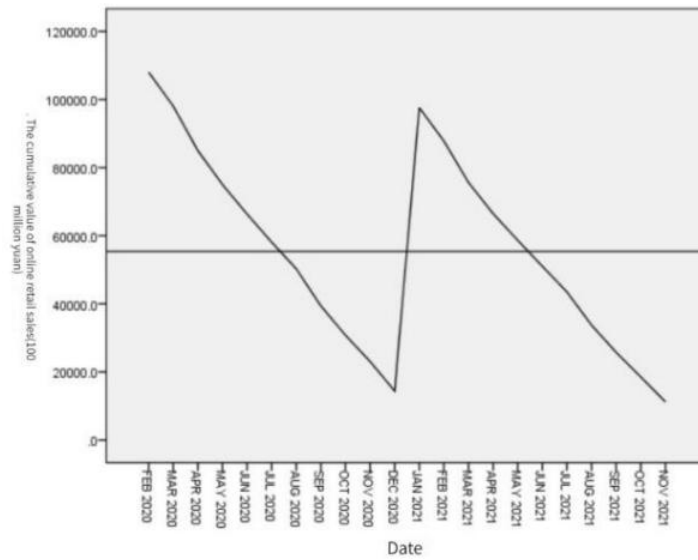| Time	Cumulative value of online retail sales (billion yuan)	Cumulative growth in online retail sales (%)	Cumulative value of online retail sales of physical goods (billion yuan)	Cumulative growth in online retail sales of physical goods (%)
December 2021	130883.5	14.1	108042.4	12
November 2021	118749.2	15.4	98056.1	13.2
October 2021	103765.4	17.4	84979.3	14.6
September 2021	91870.9	18.5	75041.8	15.2
August 2021	81226.7	19.7	66340	15.9
July 2021	71108.1	21.9	58130	17.6
June 2021	61132.5	23.2	50262.7	18.7
May 2021	48238.9	24.7	39376.9	19.9
April 2021	37638.1	27.6	30773.5	23.1
March 2021	28092.7	29.9	23067.3	25.8
February 2021	17586.8	32.5	14412.2	30.6
December 2020	117601.3	10.9	97590.3	14.8
November 2020	105374.1	11.5	87792	15.7
October 2020	91275.3	10.9	75619.3	16
September 2020	80064.7	9.7	66476.9	15.3
August 2020	70325.9	9.5	58650.7	15.8
July 2020	60785.1	9	51017.6	15.7
June 2020	51500.9	7.3	43480.7	14.3
May 2020	40175.8	4.5	33738.8	11.5
April 2020	30698	1.7	25751	8.6
March 2020	22168.8	−0.8	18535.7	5.9
February 2020	13712	−3	11233.2	3

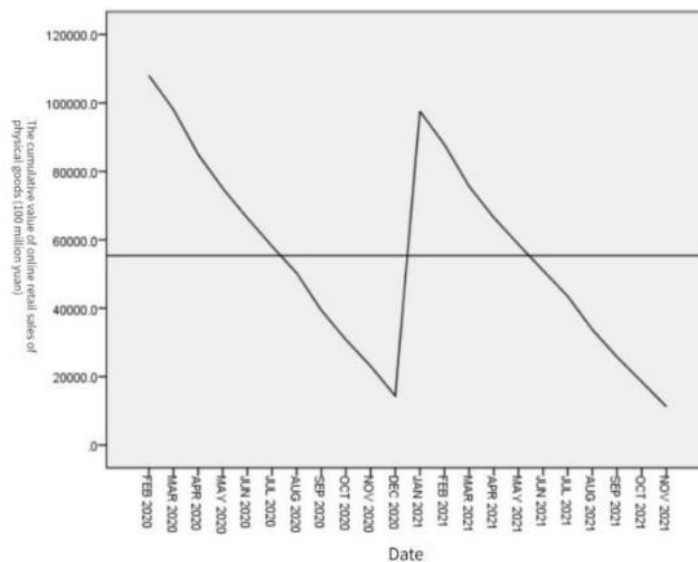**Figure 1.** Time chart of cumulative online retail sales

Figure 2. Time chart of cumulative online retail sales of physical

From Figures 1.2. the trend of the timing chart obtained after data processing changes significantly. The cumulative value of online retail sales shows the same pattern as the cumulative value of online retail sales of physical goods, and the data is in line with the actual situation.

4.3. Time-series Stability Judgment of The Data

4.3.1. Time Series Stability Judgment of The Cumulative Value of Online Retail Sales

This paper judges the stability of the time series by observing the correlation graph of the time series. The specific judgment method is as follows: drawing the autocorrelation and partial autocorrelation graph. If the correlation graph of the sequence shows an approximate linear decay, it shows that it is a non-stationary sequence, and if the correlation graph shows an exponential function or a sinusoidal function to decay, it is a stationary sequence.
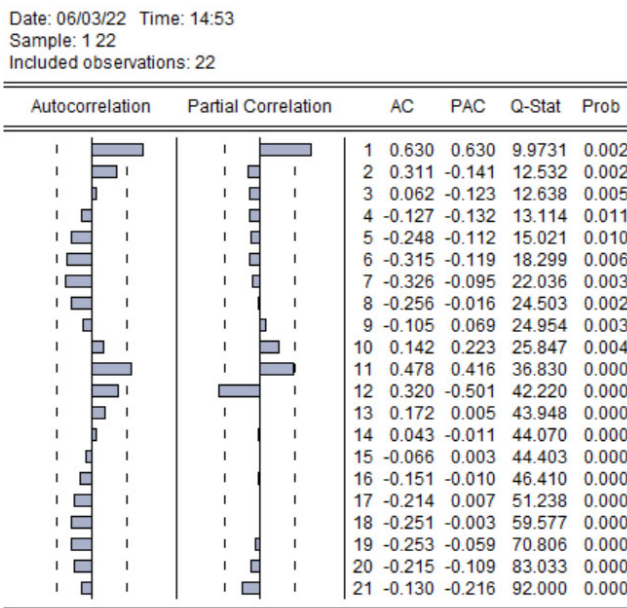

Figure 3. Autocorrelation and partial autocorrelation figure of the cumulative value of online retail sales

As can be seen from Figure 3. the time series of the cumulative value of online retail sales is stable, so there is no need to use the difference method to stabilize the sequence.

4.3.2. Time series stability judgment of the cumulative value of online retail sales of physical goods

The time series stability judgment method of the cumulative value of online retail sales of physical goods is the same as above. Results are shown in Figures 4.
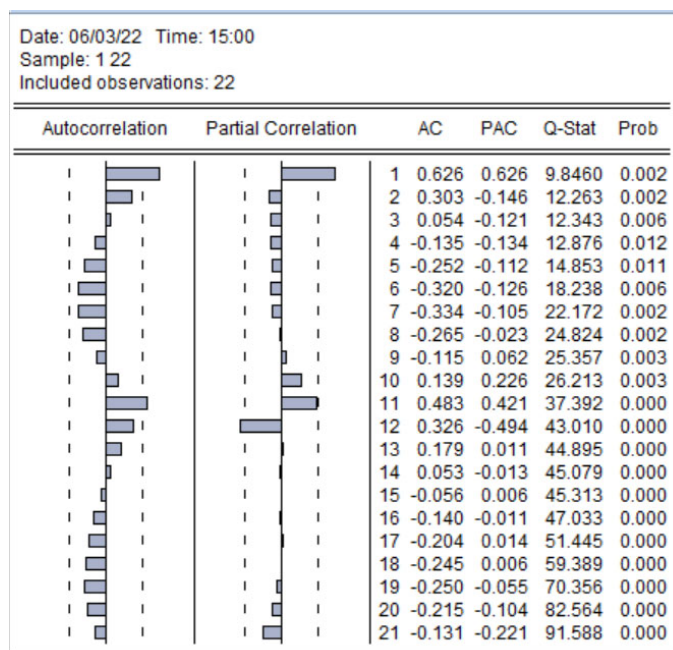

Figure 4. Autocorrelation and partial correlation diagram of the cumulative value of online retail sales of physical goods

As can be seen from Figure 4.the time series of the cumulative value of online retail sales is stable, so there is no need to use the difference method to stabilize the sequence.

5. Empirical Analysis

This paper is for the acquisition value of online physical goods sales data, which is taken as the observed value in the time series analysis model, where the time index is measured in months. And through modeling prediction and analysis, compare the benefit and effect of modeling. This part is a time-series prediction model built by the EViews software. Time-series analysis of the monthly cycle selected data from February 2020 to December 2021 to forecast online physical goods sales in November 2020.

5.1. Forecast and Modeling Process of Online Sales of Physical Goods

The modeling process of online physical goods sales forecast is shown in Figure 5.
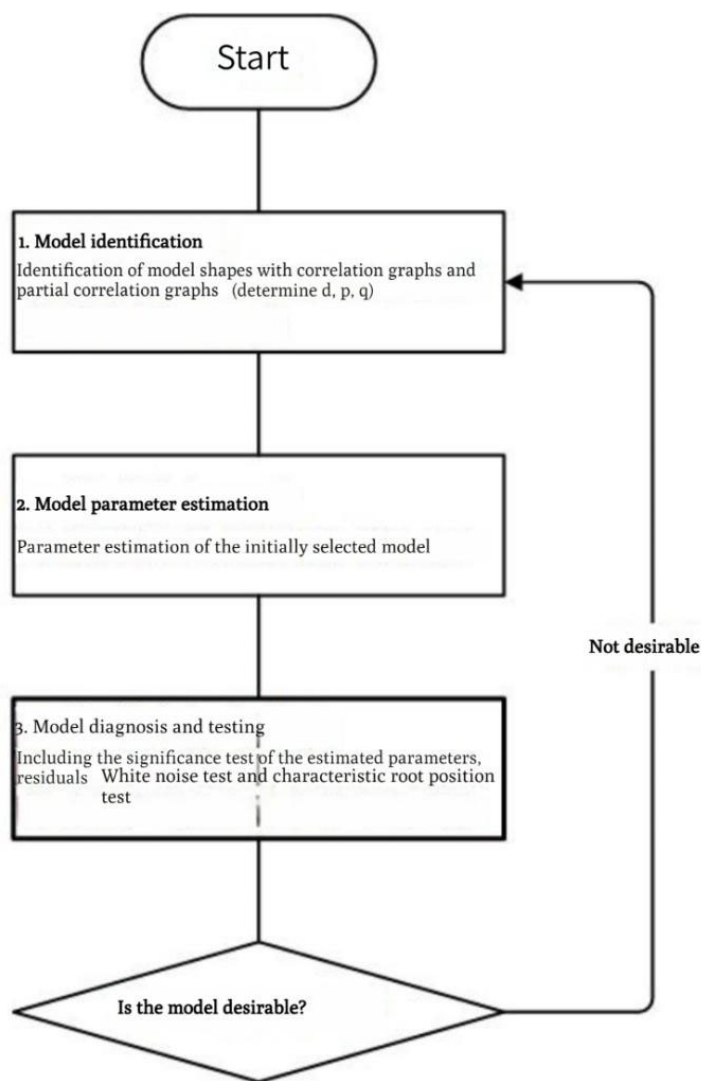

Figure 5. Modeling process of online sales forecast of physical goods

1.model recognition. Combined with the previous stationarity test, the time series is stationary, so the value of d (the number of times the sequence needs a difference) is 0, meaning that the sequence selects the ARMA model. Then p and q are identified by the autocorrelation maps and the partial autocorrelation maps.
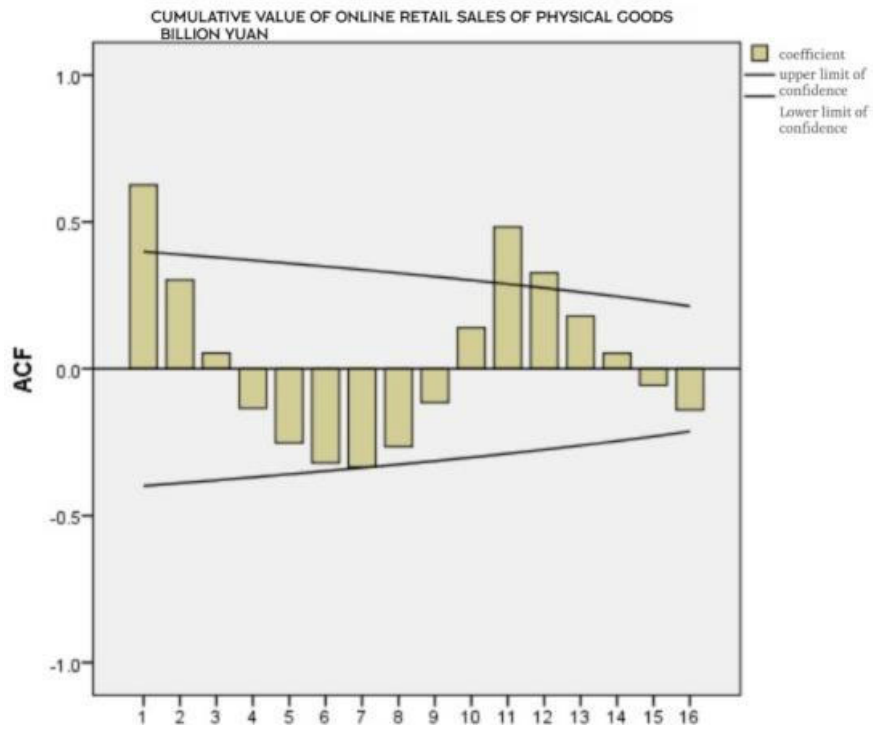
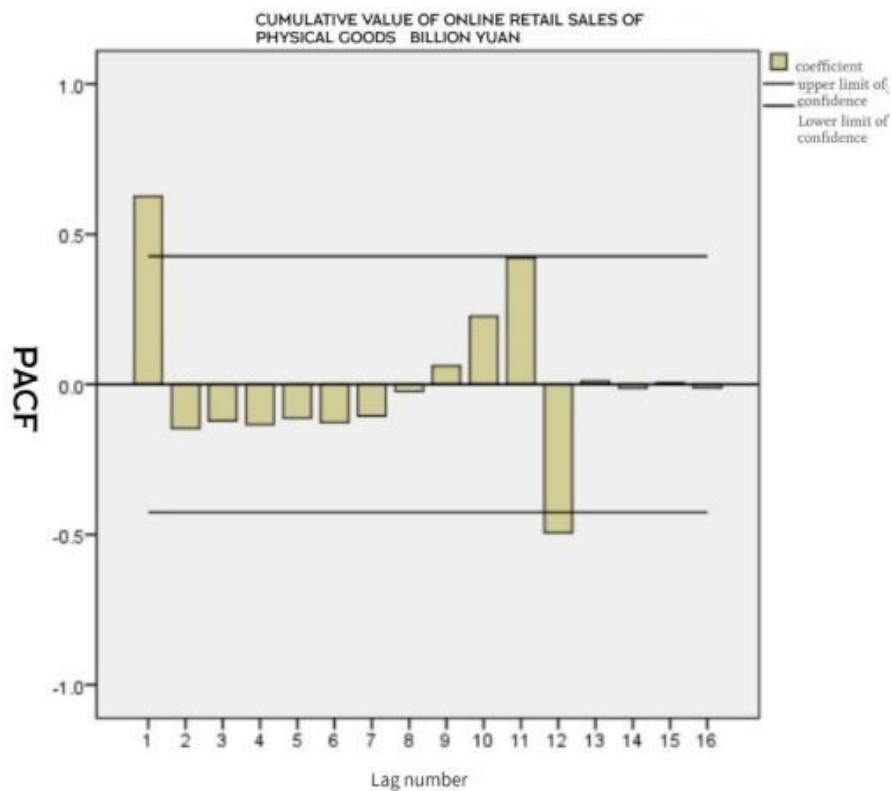**Figure 6.** Autocorrelation diagram

Figure 7. partial autocorrelation diagram

2.By observing acf (autocorrelation) and pacf (partial autocorrelation), acf shows obvious tail, see Figure 6. while pacf drops rapidly at order 2, see Figure 7.which means tail cut at order 2. The dataset was then paired and AR (2) model was used for the data set.

3.Model parameter estimation. The online sales data of physical goods is defined as variable X, and the parameters of the obtained model AR (2) were estimated using EViews software, resulting in Figure 8.

```
Dependent Variable: X
Method: Least Squares
Date: 06/08/22   Time: 19:58
Sample (adjusted): 3 22
Included observations: 20 after adjustments
Convergence achieved after 5 iterations
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	43742.89	14761.15	2.963380	0.0087
AR(1)	0.747190	0.243499	3.068559	0.0070
AR(2)	-0.096047	0.233919	-0.410600	0.6865

R-squared	0.475221	Mean dependent var	50613.50
Adjusted R-squared	0.413482	S.D. dependent var	25889.68
S.E. of regression	19827.47	Akaike info criterion	22.76501
Sum squared resid	6.68E+09	Schwarz criterion	22.91436
Log likelihood	-224.6501	Hannan-Quinn criter.	22.79416
F-statistic	7.697295	Durbin-Watson stat	1.996562
Prob(F-statistic)	0.004167		

Inverted AR Roots	.58	.16

Figure 8. Parameter estimation results

According to the picture, the model is expressed as follows:

$$x_t = 43742.89 + 0.747190 x_{t-1} - 0.096047 x_{t-2} + \mu_t, \quad \mu_t \sim N(0, \sigma 2) \quad (1)$$

According to the test of the parameters, it is found that the AR (1) model is more applicable, so the AR (1) model is finally selected.

5.2. 5.2 Forecast and analysis of online physical goods sales

Forecast the sales data for February 2022 based on the data from January 2020-December 2021, and finally compare it with the real data, to verify the accuracy of the forecast. The predicted results are shown in Table 4-1 and Figure 9.

Table 3. Shows the prediction results

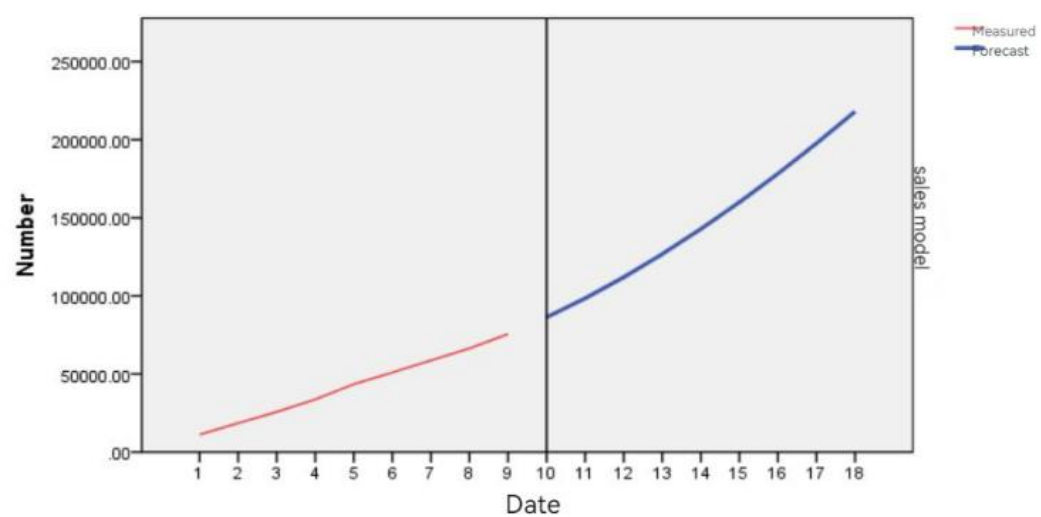| model		10	11	12
Sales volume-Model _1	calculate	86299.39	98458.34	111993.24
	UCL	93205.02	114169.39	138039.95
	LCL	79620.65	83745.48	88294.90

Figure 9. Graph the prediction results

After the prediction, the forecast data for February 2022 is obtained, and the true value is 87792, and the predicted value is 86299.39. Thus, the deviation between the forecast data and real data is very small, so the electricity enterprise can adopt the method of prediction for the company subsequent management and business decisions, develop network marketing strategy, to help electric business enterprises to achieve accurate production, reduce the inventory cost caused by product backlog and other costs, and obtain greater profit space.

6. Summary

By analyzing and combining the prediction model of online physical commodity sales, this paper obtains the

following conclusions, hoping to provide technical support for e-commerce enterprises to develop marketing strategies.

The deviation between forecast data and real data is very small, so the electricity enterprise can adopt the method of prediction for the subsequent management and business decisions, develop network marketing strategy, to help electric business enterprises to achieve accurate production, reduce the product backlog of inventory cost and other costs, and obtain greater profit space.

Therefore, e-commerce enterprises can adopt appropriate prediction methods to develop more accurate marketing strategies when their own capabilities allow.

In the research process of this paper, due to the limitation of data access channels and its own limited research ability, the research content of this paper is not very perfect, and the selection of specific platforms and categories for modeling and prediction will get better results. After that, the research and analysis of the online physical commodity sales data can increase the analysis of the platform and category dimensions, such as studying the influencing factors of JD's Xiaomi mobile phone demand, so as to make the problem research more specific and accurate.

Acknowledgment

References

[1] China Internet Information Center. The 49th Statistical Report on Internet Development in China [R]. Beijing: China The Internet Information Center, 2022.

[2] Dong Yan, time, when the sweet rain. The Influence of Online Marketing on Online Consumer Purchase Behavior [J]. Economic problem exploration Cable, 2020 (10): 45-55.

[3] Zuo Civilization, Chen Shaojie, Wang Xu, Chen Huaqiong. A Multi-attribute Behavior Decision Model of Network Consumers Based on Foreground Theory[J]. Journal of Management Engineering, 2019,33 (03): 125-135.

[4] Kansra Pt,Oberoi S .Determinants of Online Purchase Intention Among Young Consumers in Punjab: A Cross-Sectional Study[J].International Journal of Social Ecology and Sustainable Development (IJSESD), 2022, 13.

[5] Zhang Jingxuan, Zhang Weiwei. Prediction of e-purchase behavior based on neural network [J]. Applied mathematics into Exhibition, 2021,10 (10): 7.

[6] Li Weiqing, Chi Maomao, Wang Weijun. Online Consumer Preference Prediction Study Based on Perceived Value [J].management Newspaper, 2021,18 (06): 912-918.

[7] Dong J,Huang T,Min L,et al.Prediction of Online Consumers'Repeat Purchase Behavior via BERT-MLP Model[J]. Electronic Research and Application, 2022 (006-003).

[8] Rajendrakumar R .The prediction of online shopping consumers' attitude based on certain statements of consumer and marketing factors with reference to post graduate students in Coimbatore City[J].International Journal of Business Information Systems, 2022, 39.

[9] Wang Xue. High-level subject prediction studies based on time-series models [J]. Intelligence Journal, 2019,38 (6).

[10] Yan Jinghua, Hou Miao Miao. Time series prediction of theft crime based on LSTM Network [J]. Data points Analysis and Knowledge Discovery, 2020,4 (11): 84-91.

[11] Zhigang Wang. Research on Deep prediction prediction prediction based on Chain chains [J]. Modern Computer, 2020 (34).

[12] Ke Miao, Huang Huaguo, et al. Based on LSTM Neural Network [J]. Fujian Normal University Journal of Fuqing Branch School, 2020 (05): 25-33.

[13] Huang Hongmei. Time-series analysis (M) was applied. Tsinghua University Press, 2016.

[14] Nie Shuyuan. The Early Development of the Time-Series Analysis (Dj.Northwestern University, 2012.

[15] Chris Chatfield.The Analysis of Time Series MTaylor and Francis,2013.

[16] Box G E P,Jenkins G M.Time Series Analysis,Forecasting,and Control[M].San Francisco,California: Holden Day1976.

[17] Box G EP,Pierce D ADistribution of Autocorrelations in Autoregressive Moving Average Time Series Models[J].Journal of the American Statistical Association, 197065(332):1509-1526.