Spare Parts Demand Forecasting based on ARMA Model

REN Xi*[a], ZHANG Xiao-fei[b]
[a]Unit 92493, Huludao, Liaoning China 125000; [b]Unit 92635 Qingdao, Shandong China 266000
* Corresponding author: renxi126@163.com

ABSTRACT

Starting from the time series of factors, the level of analysis time, data types, and forecasting accuracy, based on the characteristics of the data sequence to be analyzed. ARMA model to predict sequence requirements must be stable, that factors in the time range of the study subjects must be subjected to the same requirements. If the given sequence is not stationary sequence, you must do on a given sequence of preprocess, smoothing it, then by ARMA model. Example is analyzed by Eview software, the validity of the model is verified.

Keywords: time series; spare parts; demand forecast; ARMA model;statistical method;data preprocess;order determination;adaptability test

1. INTRODUCTION

Time series forecasting method is a basic forecasting method with a long history and vitality. Its basic idea is to reveal the law of demand value changing with time through the historical data of time series, and extend this law to the future, so as to predict the future. Time series forecasting method has attracted extensive attention because it only needs to use past records, easy data collection, low cost, simple calculation and easy execution [1][2].

2. ANALYSIS OF INFLUENCING FACTORS OF TIME SERIES

2.1 Time level

Table 1.Comparison between traditional statistical method and gray forecasting method

Method	Minimum required number of data	Data type	Data interval
Simple exponential smoothing	5~10	Equidistant	Short interval
Holt's exponential smoothing	10~15	Same trend	Short or medium interval
Winter's exponential smoothing	5	Same trend and regularity	Short or medium interval
Regression analysis	10 or 20	Same trend and regularity	Short or medium interval
Causal regression	10	Various forms can be mixed with each other	Short, medium and long intervals
Time sequence compression	2 peaks	Same trend, regular and self-adjusting	Short or medium interval
Gray forecasting	4	Equal and unequal spacing	Short, medium and long intervals
ARMA	30	Equidistant	Short, medium and long interval

The time level affects the forecasting method in three aspects: first, select the appropriate forecasting method according to the future time span. Generally speaking, qualitative methods are generally used for long-term forecasting, while quantitative forecasting methods are mostly used for medium-term and short-term situations; The second is the historical data required by a forecasting method. Some forecasting methods need little historical data, while others must ensure a certain amount of historical data, otherwise the forecasting method cannot run; Finally, the effective forecasting length of one method. Some forecasting methods can predict one or two time cycles in the future, while others can predict multiple cycles. The basic demand conditions and application scope of the main forecasting methods are shown in Table 1[6].

Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), edited by Xianye Ben, Proc. of SPIE Vol. 12462, 124620G · doi: 10.1117/12.2661048

2.2 Data type

Some data series may contain both seasonal type and trend type, or consist of simple mean and random disturbance, or periodic type. Because different forecasting methods have different forecasting abilities for different data types, it is particularly necessary to predict different data types with appropriate forecasting methods.

2.3 Forecasting accuracy

The first-line indicator closely related to the forecasting method is the forecasting accuracy. For some cases, the forecasting accuracy of 10% is enough, while for some precision industries, the accuracy of 0.2% will be far lower than the use requirements.

Other influencing factors such as ease of use and availability of relevant software can be used as indicators to evaluate the advantages and disadvantages of a forecasting method. Nevertheless, the most important indicator is the forecasting accuracy mentioned above.

The forecasting problem based on time series is studied for two problems: firstly, the spare parts demand time series is preprocessed, and then the corresponding mathematical model is established for the processed time series, and the forecasting analysis is carried out.

3. TIME SERIES CHARACTERISTIC ANALYSIS

The assumption that the future "looks like the past" is a very important assumption in time series analysis, which is called stationarity. However, time series variables become unstable in various ways, two of which are particularly relevant to the regression analysis of time series data [7]:

1) The sequence may show continuous long-term movement, that is, the sequence has a trend;

2) The overall regression may be unstable over time, that is, there may be mutations in the overall regression.

These deviations from stationarity will endanger the forecasting and inference based on time series. Fortunately, however, we have statistical methods to find trends and mutations. Once these situations are found, we can adjust the setting form of the model.

The main methods of time series forecasting include exponential smoothing method, regression analysis method, regression method, ARMA method and Gray forecasting method. The research on the characteristics of time series by these methods mainly includes: stability detection, sequence analysis, and the selection of trend items. The following mainly studies the spare parts demand forecasting method based on ARMA model.

4. SPARE PARTS DEMAND FORECAST BASED ON ARMA MODEL

In the actual use of equipment, the obtained spare parts demand data is usually non-stationary. How to reasonably predict the future spare parts demand according to the non-stationary data sequence. ARMA model forecasting requires that the sequence must be stable, that is, the influencing factors of the research object must be basically the same within the research time range. If the given sequence is not a stationary sequence, the given sequence must be preprocessed to stabilize it, and then modeled with ARMA model. The modeling steps are shown in Figure 1.

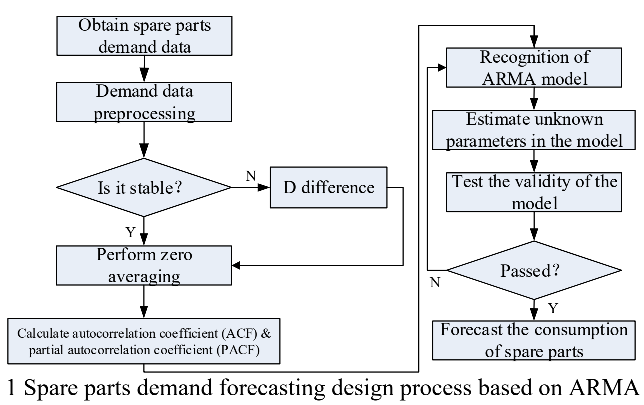


Figure 1 Spare parts demand forecasting design process based on ARMA model

Time series analysis is an effective modeling method, which requires that the observation data samples should not be less than 30, and the test time should be equal. However, in engineering practice, due to the constraints of various factors, there are often less than 30 observation data samples with different time intervals. Therefore, it is particularly important to study the time series data processing method of small samples. Some literates use SVM to model and predict small sample data, but how to reasonably define the model is very difficult[8].

The process of spare parts demand forecasting method is as follows: firstly, obtain the historical demand data sequence of equipment spare parts, preprocess the demand data sequence, test the stability of the sequence, and finally select the forecasting method to predict and analyze the future demand of spare parts.

4.1 Data preprocessing

Before establishing the time series model, the data must be preprocessed to eliminate those abnormal samples that do not comply with the statistical law, and the basic statistical characteristics of these samples must be tested to ensure the reliability and confidence of the time series model and meet certain accuracy requirements.

For the original data sequence of spare parts demand $X=\{x_1,x_2,\cdots,x_t,\cdots\},t=1,2,\cdots,n$, there are generally the following two characteristics:

1) There are few historical data of spare parts, and the amount of general data is less than 30, which belongs to a typical small sample problem;

2) There may be unequal intervals in the historical demand data of spare parts;

3) The data size of individual parameters is abnormal.

Therefore, the test data must be preprocessed before statistical analysis. Firstly, the values with obvious errors, that is, the possible outliers in the original data, are eliminated. At the same time, in order to ensure the integrity and continuity of data, the eliminated data are interpolated; Then, in order to ensure the equal interval of the test data, the original data is interpolated and supplemented again[9].

The stationarity of time series is an important prerequisite for time series modeling. The purpose of stationarity test is to determine whether there is a random trend or determine the trend in the sequence, otherwise there will be a pseudo regression problem[10]. Stationarity test mainly includes parameter test method, non-parametric test method and sequence diagram test method[11]. After the stationarity detection of the time series, according to the detection results, if it is a stationary series, it is directly predicted by ARMA model. If it is a non-stationary series, the series is processed by difference until it is stable, and then it is modeled by ARMA model and predicted by correlation analysis.

4.2 ARMA model order determination

Parameter estimation and model order determination are important contents of establishing spare parts consumption forecasting model, which affect each other. Based on the above model identification, the unknown parameters of ARMA (P, q) model, namely autoregressive coefficient, moving average coefficient and white noise variance, are estimated by using sample moment estimation method, least square estimation method or maximum likelihood estimation method and got $\hat{\varphi}_1, \hat{\varphi}_2, \cdots \hat{\varphi}_p, \hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_q, \hat{\sigma}^2$. The following table shows the discrimination principles of ARMA (p, q) model[12].

Table 2 Discriminant principles of ARMA (p, q) model

Model	AR（p）	MA（q）	ARMA（p,q）
auto-correlation function	Tailing, exponential decay or oscillation	Finite length, q-step truncation	Tailing, exponential decay or oscillation
Partial auto-correlation function	Finite length, p-step truncation	Tailing, exponential decay or oscillation	Tailing, exponential decay or oscillation

Box Jenkins proposed to use the truncation of sample auto-correlation function $\hat{\rho}_k$ or sample auto-correlation function $\hat{\gamma}_k$ and sample partial auto-correlation function $\hat{\varphi}_{kk}$ to judge MA model and AR model respectively, and determine the order by asymptotic normality and asymptotic χ^2 test [13].

Order determination, parameter estimation and test are combined step by step. For the above preliminary determination $MA(q_0)$ and $AR(p_0)$ by truncation, the corresponding order shall also be determined through white noise test. The order of ARMA model is mainly determined by F-test, FPE, AIC and BIC order determination criteria. For specific methods, refer to literature [11].

4.3 Model adaptability test

After the identification, order determination and parameter estimation of the model, the remaining problem is to judge whether the model is appropriate to describe the time series, that is, the adaptability test of the model. The adaptability of the model refers to the degree to which a time series model explains the dynamics of the system (that is the correlation of data series). The suitable model of a time series $\{a_t\}$ should completely or basically explain the dynamics of the system (that is the correlation of data series), so the residual sequence $\{a_t\}$ in the model should be a white noise sequence. In essence, the adaptability test of the model is to test whether the sequence is white noise sequence. The most important one is the independence test of sequence $\{a_t\}$.

The progress of computer technology has greatly promoted the development of time series analysis. At present, many statistical software can be used for time series analysis. Commonly used software mainly includes S-PLUS, MATLAB, Gauss, TSP, Eviews and SAS software[14]. This paper mainly studies the characteristic form of time series through Eviews analysis software.

Eview is the abbreviation of economic views. It is an econometric software package launched by GMS company[15]. The original intention of the software is to analyze the quantitative law of socio-economic relations and economic activities by using econometric methods and techniques. The core of econometric research is to design models, collect data, estimate models, test models, use models to predict, solve models and use models. Eview is an indispensable tool to complete the above tasks.

5. CASE ANALYSIS

In the actual use of equipment, the obtained spare parts demand data is usually non-stationary. How to reasonably predict the future spare parts demand according to the non-stationary data sequence. ARMA model forecasting requires that the sequence must be stable, that is, the influencing factors of the research object must be basically the same within the research time range. If the given sequence is not a stationary sequence, the given sequence must be preprocessed to stabilize it, and then modeled with ARMA model. The modeling steps are shown in Figure 1.

Now, the statistical sequence of spare parts used for an equipment from the first quarter of 2006 to the first quarter of 2014 is as follows (the unit of spare parts demand sequence is calculated quarterly (0 means lack of statistical data)):43,42,47,45,46,43,0,43,47,48,46,42,49,48,46,42,47,45,49,47,48,41,0,46,48,43,49,42,48,23,41,44,46。

In order to verify the forecasting effectiveness of the model, the first 32 data in the series are selected for modeling and analysis. It can be seen from the above statistical data that the spare parts demand sequence of the system $\{x_t\}$ is a stationary random demand sequence. The number of samples $\{x_t\}$ is 33, greater than 30, does not belong to small sample series, so statistical methods can be used to analyze the series.

For the data shown above, the Pauta criterion is applied to calculate the mean value $\bar{x} = 43.8571$, $\sigma = 6.6548$, and the residuals corresponding to the data in the second quarter of 2013 can be $|V_{30}| = 23.8571$, as $|V_{30}| = 23.8571 > 3\sigma = 19.9644$, according to the Pauta criterion,outlier x_{30} should be eliminated. In order to maintain the integrity of the data, it is necessary to supplement the eliminated outliers. According to the Lagrange interpolation principle, select x_{28} , x_{29} , x_{31} and x_{32} four points for interpolation to obtain x_{30} = 43.75. Considering that the actual demand for spare parts cannot be non integer, the final value is 44.

In addition, the time series analysis requires equal intervals. As can be seen from the above table, the statistics of spare parts in the third quarter of 2007 and the second quarter of 2010 are missing in the series $\{x_t\}$, mainly because the detachment performed ocean training tasks in the third quarter of 2007 and the second quarter of 2010 respectively, resulting in the lack of equipment support and maintenance records. After the Lagrange interpolation polynomial is used to interpolate and supplement it, x_7 =44.75. Considering that the actual demand for spare parts cannot be non integer, the value x_7 is 45, x_{14} is obtained as 46 by the same method, that is, the latest correction sequence $\{x_t\}$ is:43,42,47,45,46,43,45,43,47,48,46,42,49,48,46,42,47,45,49,47,48,41,46,46,48,43,49,42,48,44,41,44。

For the spare parts demand series $\{x_t\}$, it can be seen from the figure above that the series is stable. According to the above figure, AR (1), MA (1), ARMA (1,1) and other models can be used for continuous fitting and comparison. Due to the space relationship, only the correlation analysis diagram of ARMA (1,1) is given here.

The relevant information of ARMA (1,1) can be obtained through software analysis, as shown in Figure 3. For model comparison and model selection, it is often not to simply look at a certain index, but to comprehensively consider all aspects of the situation, make a comprehensive judgment and select an optimal model. The following factors are generally considered:

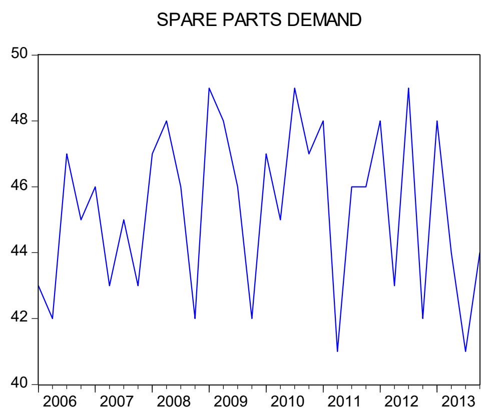


Figure 2 Spare parts demand sequence

Sample (adjusted): 2006Q2 2013Q4
Included observations: 31 after adjustments
Convergence achieved after 12 iterations
MA Backcast: 2006Q1

	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.999552	0.001337	747.5144	0.0000
MA(1)	-0.965930	0.059778	-16.15850	0.0000

R-squared	-0.050287	Mean dependent var	45.38710
Adjusted R-squared	-0.086503	S.D. dependent var	2.525568
S.E. of regression	2.632538	Akaike info criterion	4.836115
Sum squared resid	200.9774	Schwarz criterion	4.928630
Log likelihood	-72.95978	Hannan-Quinn criter.	4.866272
Durbin-Watson stat	2.379613		

Inverted AR Roots	1.00
Inverted MA Roots	.97

Figure 3 Spare parts demand sequence related parameter indicators

1) R-squared and adjusted R-squared. The larger the two indicators, the better.

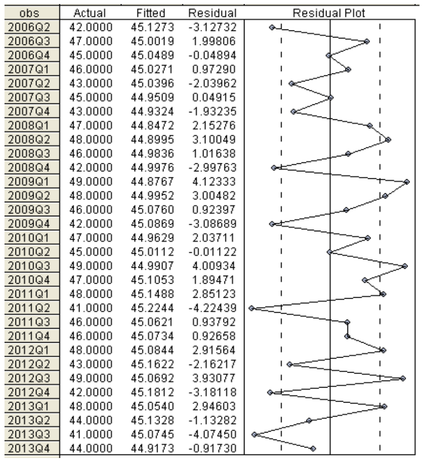
2) Whether the estimation of each parameter in the model fitting is significantly non-zero. Generally speaking, the prob term in the above figure is significantly less than 0.05.

3) AIC guidelines. Akaike info criterion and Schwarz criterion in the above figure. Generally speaking, the smaller the two indicators, the better.

4) DW value depends on whether the value is close to 2. That is, the Durbin Watson stat term in the figure above represents whether there is a first-order auto-correlation in the residual sequence after model fitting. When the DW value is close to 2, it means that there is no auto-correlation in the residual sequence; When the DW value is close to 0, there is a strong auto-correlation in the residual sequence; When the DW value is close to 4, it means that there is no auto-correlation in the residual sequence.

5) The number of parameters to be estimated. Generally speaking, the fewer parameters need to be estimated, the better. Less estimation can reduce the deviation in the estimation process as much as possible. According to the above judgment criteria, combined with the above figure, the overall fitting effect of ARMA (1,1) model is ideal (the dotted line in Figure 4 is the 95% upper and lower confidence interval of the predicted value). Therefore, the model is finally selected to fit the stationary series. The form of the final model is as follows:

$$X_t = 0.9996X_{t-1} + \varepsilon_t - 0.9659\varepsilon_{t-1}, \varepsilon_t \sim WN(0,\sigma^2) \quad (1)$$



obs	Actual	Fitted	Residual	Residual Plot
2006Q2	42.0000	45.1273	-3.12732	
2006Q3	47.0000	45.0019	1.99806	
2006Q4	45.0000	45.0489	-0.04894	
2007Q1	46.0000	45.0271	0.97290	
2007Q2	43.0000	45.0396	-2.03962	
2007Q3	45.0000	44.9509	0.04915	
2007Q4	43.0000	44.9324	-1.93235	
2008Q1	47.0000	44.8472	2.15276	
2008Q2	48.0000	44.8995	3.10049	
2008Q3	46.0000	44.9836	1.01638	
2008Q4	42.0000	44.9976	-2.99763	
2009Q1	49.0000	44.8767	4.12333	
2009Q2	48.0000	44.9952	3.00482	
2009Q3	46.0000	45.0760	0.92397	
2009Q4	42.0000	45.0869	-3.08689	
2010Q1	47.0000	44.9629	2.03711	
2010Q2	45.0000	45.0112	-0.01122	
2010Q3	49.0000	44.9907	4.00934	
2010Q4	47.0000	45.1053	1.89471	
2011Q1	48.0000	45.1488	2.85123	
2011Q2	41.0000	45.2244	-4.22439	
2011Q3	46.0000	45.0621	0.93792	
2011Q4	46.0000	45.0734	0.92658	
2012Q1	48.0000	45.0844	2.91564	
2012Q2	43.0000	45.1622	-2.16217	
2012Q3	49.0000	45.0692	3.93077	
2012Q4	42.0000	45.1812	-3.18118	
2013Q1	48.0000	45.0540	2.94603	
2013Q2	44.0000	45.1328	-1.13282	
2013Q3	41.0000	45.0745	-4.07450	
2013Q4	44.0000	44.9173	-0.91730	

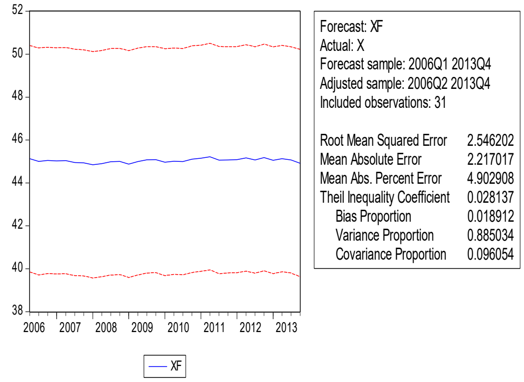


Figure 4 Spare parts demand forecast sequence   Figure 5 Deviation between predicted value and actual value

According to the above analysis results, the overall accuracy of multi sample spare parts demand forecasting based on ARMA model meets the error requirements. With the increase of sample data, the forecasting results will be more accurate. This must be paid attention to when applying ARMA model to predict the demand for spare parts.

According to the forecasting model, the demand for spare parts in the first quarter of 2014 is predicted, and the predicted value is 45.3249, rounded to 46. It is consistent with the actual demand value of spare parts, which verifies the effectiveness of the model.

6. CONCLUSION

The forecasting error of ARMA model is small, and the forecasting effect meets the needs of the actual situation. It can better solve the problem that the demand for ship equipment spare parts can not be accurately predicted, and has good engineering application value. It is more convenient and intuitive to study the historical demand sequence through relevant statistical software. This method also has important reference significance for the life forecasting of other complex systems and has broad application prospects.

REFERENCES

[1] Wang Naichao, Kang Rui. Research on spare parts demand generation, propagation and analysis algorithm [J]. Journal of Aeronautics, 2008.29 (5): 1163-1167

[2] Yu Guohua, Huang Houkuan. Selection method of time series model [J]. Journal of Guangxi Normal University, 2003,21 (1): 191-194

[3] Shou Zhaoyong, Yang Yuanyuan. Modeling method and process of time series problems [J]. Mathematical theory and application, 2012.32 (1): 112-120

[4] Hunter J,Mcintoshn. Knowledge-based eventdetection in complex time series data[J].Artificial Intelligence in Medicine,1999,16:271-280.

[5] Qu Li, Zhang Qun. Overview of spare parts inventory management [J]. Laboratory research and discussion, 2006,25 (7): 875-880

[6] Li Ping, he Xianzhong. Exploration and analysis of RBS and its engineering of the US Navy [C]. Song Tailiang. Proceedings of the Third Symposium on equipment reliability, maintainability and supportability. Beijing: technical foundation management center of the general equipment department, 2005: 347-354

[7] Wang Fang. Research on segment traffic volume forecasting based on support vector machine [D]. Master's thesis of Nanjing University of Aeronautics and Astronautics, 2007:57-72

[8] Wang Lin, Zeng Yurong. Determination model of demand function of infrequent spare parts under small sample data [J]. Journal of Huazhong University of science and Technology (NATURAL SCIENCE EDITION), 2004,32 (9): 96-99

[9] Chen Zhuo. Research on equipment defect forecasting based on time series [D]. Master's thesis of Liaoning University of engineering and technology, 2005: 9-15
[10] Zhang Hengxi, Guo Jilian, Zhu Jiayuan. Small sample multivariate data analysis method and its application [M]. Xi'an: Northwest University of Technology Press, 2002:1-3
[11] Wang Weixin. Applied time series analysis [M]. Nanning: Guangxi Normal University Press, 1999:24-56
[12] Xue Ziyun, Yang Jiangtian, Zhu Hengjun. Overview of mechanical fault forecasting models [J]. Mechanical strength, 2006, 28 (s): 60-65
[13] Carlo G.Subset. ARMA model indentification using genetic algorithms[J].Journal of Time Series Analysis,2000,21(5):559-570 .
[14] Zhang Xiaodong. Application of Eviews in Econometrics [M]. Beijing: Machinery Industry Press, 2006:19-24