Academic Journal of Engineering and Technology Science
ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

Predicting Container Throughput Based on Combined Model of Principle of Least Squares and Validity

Zheyu Zhang

Shanghai Maritime University, Shanghai 200000, China
zhangzheyu@stu.shmtu.edu.cn

ABSTRACT. *With exponential smoothing model and linear regression model, two linear regression models are established based on the principles of least squares and validity respectively, and compared with the prediction results of single prediction model. The results show that the combination of the principle based on validity principle proposed in this paper. The prediction result of the model is more reasonable and has certain practicality. Through this method, the container throughput of Shenzhen Port in 2020 was forecasted.*

KEYWORDS: *Container throughput, Combined model, Effectiveness*

1. Introduction

Shenzhen Port is not only one of the earliest ports established by China's international liner companies for regular liner services. The international liner services of far and near seas cover the major ports in the world's twelve major shipping areas, forming a huge network of routes. One of the widest range of ports. Shenzhen Port has been the second largest container port in China for 14 years, taking the Pearl River Delta, China's largest export source of goods, as its hinterland. The total amount of foreign trade imports and exports has ranked first in China's large and medium-sized cities for 13 consecutive years, with a large total volume, strong carrying capacity, The characteristics of the wide radiation range, the trend and potential of its port development and container throughput have received special attention from the world shipping industry. The turbulent world economic situation in recent years has resulted in the instability of trade volume, which directly affects the market of international and regional container hub ports. The industrial structure transfer and upgrading that China is facing will also have an impact on the port market. Nature and the strategic status and role of container transportation in Shenzhen Port. It is of great significance to port and economic development to conduct in-depth research on its throughput generation factors and rules, and to use a more scientific and reasonable method to make guiding predictions [1] [2]. This

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

paper provides a method that can be used for reference in predicting the container throughput of Shenzhen Port.

2. Methodology

2.1 Exponential smoothing

Exponential smoothing is actually a special weighted moving average method. Its characteristics are: First, the exponential smoothing method further strengthens the effect of recent observations on the predicted value during the observation period, and the weights given to the observations at different times are different, thereby increasing the weight of the recent observations, so that Forecasts can quickly reflect actual market changes. The weights are reduced in equal series. The first term of this series is the smoothing constant a, and the common ratio is 1-a. Second, the exponential smoothing method has scalability for the weights given to the observations. You can take different values of a to change the rate of change of the weights.

Quadratic exponential smoothing is re-smoothing of the first exponential smoothing. It applies to time series with a linear trend. Its prediction formula is:

$$S_t^{(2)} = a \times S_t^{(1)} + (1-a) \times S_{t-1}^{(2)} \tag{1}$$

Where $S_t^{(2)}$ and $S_{t-1}^{(2)}$ are the quadratic exponential smoothing of t period and t-1 period respectively; a is the smoothing coefficient.

2.2 Linear regression model

The regression analysis method is to apply the inertia principle and related principles of the development of things to seek and establish a correlation model between the independent variable that affects the change in container throughput and the dependent variable of container throughput. Then predict the future value of container throughput based on the correlation model and the future value of the independent variable. The container throughput of Shenzhen Port has a good correlation with the gross domestic product (GDP). GDP is selected as the independent variable in the correlation analysis, and container throughput is used as the dependent variable. The corresponding regression model is established as follows:

$$X_{(t)} = a \times Y_{(t)} + b \tag{2}$$

Where t is the years; X (t) is the container throughput; Y (t) represents GDP.

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

2.3 Combined model based on the principle of least squares

Due to the complexity of the system, in many cases, simply using a specific model for inversion is often one-sided, and different models are appropriately combined in some way, and the information provided by various models is comprehensively used. Improve the limitations of a single model, improve inversion accuracy, and reduce errors [3]. Exponential smoothing Because the quadratic exponential smoothing model is a single time series model, it does not reflect the influence of economic and other influencing factors on throughput changes and has certain limitations, and the linear regression model can just make up for the above shortcomings. Combining the two to build a new model can not only make full use of the advantages of less data modeling of the quadratic exponential smoothing model, but also make up for the lack of linear terms and can not reflect the defects related to factors, thereby improving the prediction accuracy.

This article uses the definition of the best combined model for the same independent variable, and now it is generalized so that it can be used to select combinations of models with independent variables. Assume $S_0(X_t)$ is the actual value($t=1,2,...,n$). $S_i(X_{it})$ is the prediction value of the i-th method of $S_0(X)$ ($i=1,2,...,n$). Then $S(X)=\sum_{i=1}^{m} k_i S_i(X_i)$ is called a combined model of m prediction methods, where k_i is the weight of the i-th prediction in the combined model. If $K=(k_1,k_2,...,k_m)^T$ can be found to minimize the sum of the squared prediction errors of the combined model, the combined model is called the best combined prediction model in the sense of least squares estimation. Using the Lagrange multiplier method, we get

$$K=\frac{E^{-1}R^T}{RE^{-1}R^T} \tag{3}$$

$$E=\begin{bmatrix} \sum_{t=1}^{n} e_1^2(X_{1t}) & \sum_{t=1}^{n} e_1(X_{11})e_2(X_{2t}) & ... & \sum_{t=1}^{n} e_1(X_{1t})e_m(X_{mt}) \\ \sum_{t=1}^{n} e_2(X_{2t})e_1(X_{1t}) & \sum_{t=1}^{n} e_2^2(X_{2t}) & ... & \sum_{t=1}^{n} e_2(X_{2t})e_m(X_{mt}) \\ ... & ... & & ... \\ \sum_{t=1}^{n} e_m(X_{mt})e_1(X_{1t}) & \sum_{t=1}^{n} e_m(X_{mt})e_2(X_{2t}) & ... & \sum_{t=1}^{n} e_m^2(X_{mt}) \end{bmatrix} \tag{4}$$

Where $e_i(X_{it})=S_0(X_i)-S_i(X_{it})$ is the prediction error of the i-th method, $R=(1,1,1,...,1)$. So the combined model is

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

$$S(X) = S_1(X_1) + S_2(X_2) \tag{5}$$

Where X_1 is the forecast year; $S_1(X_1)$ is the predicted value of the container throughput of the exponential smoothing model; X_2 is GDP value of the forecast year; $S_2(X_2)$ is the linear regression forecast value.

2.4 Combined model based on validity principle

The literature proposes a method for determining the combination weighting coefficient based on the principle of validity [4].

$$A_{it} = 1 - \left| \frac{y_t - y_{it}^{'}}{y_t} \right| \tag{6}$$

$$E = \frac{1}{N} \sum_{t=1}^{N} A \tag{7}$$

$$S = \frac{1}{N} \left[\sum_{t=1}^{N} (A_t - E) \right] \tag{8}$$

$$f_i = \frac{S_i}{\sum_{j=1}^{2} S_j} (i = 1, 2) \tag{9}$$

Among them, the total number of data to be predicted is N, y_t is the actual value of container throughput, and $y_{1t}^{'}$, $y_{2t}^{'}$, A_{1t} and A_{2t} are the predicted values and precision sequences using the smooth exponential model and linear regression model, respectively. E are the mean of the series; f_1 and f_2 are the weighting coefficients of these two models. $y_t^{'}$ is the predicted value of the combined model as follow.

$$y_t^{'} = f_1 y_{1t}^{'} + f_2 y_{2t}^{'} (t = 1, 2, ..., N) \tag{10}$$

3. Example study

This article selects Shenzhen's container throughput data for a total of 10 years from 2008 to 2017 as an example, as shown in Table 1 below. (Data source: China Statistics Bureau and Port Yearbook)

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

Table 1 Shenzhen Port's container throughput and GDP indicators over the years

years	GDP of Shenzhen (trillion yuan)	Shenzhen Port Throughput (10,000 TEU)
2008	7806.54	2141.62
2009	8201	1825.01
2010	9581.5	2250.97
2011	11505.53	2257.09
2012	12950.1	2294.13
2013	14500.23	2327.84
2014	16001.82	2403.74
2015	17502.86	2420.46
2016	19492.6	2397.93
2017	22438.39	2520.87

3.1 Exponential smoothing model prediction

The data is exponentially smoothed. And the damping coefficient is tested one by one from 0.1 to 0.1, and the relatively small error of 0.3 is selected. The results are shown in Figure 1 and 2. The above chart shows that the accuracy of the exponential smoothing is higher.

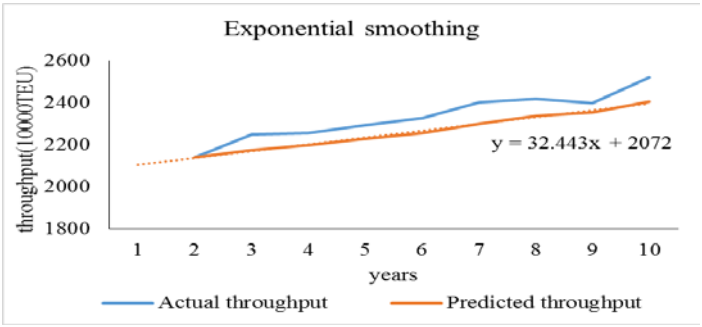

Figure. 1 Result of the exponential smoothing

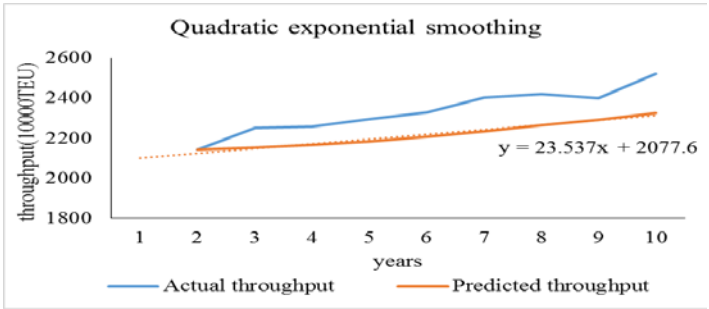

Figure. 2 Result of the quadratic exponential smoothing

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

3.2 Linear regression model prediction

Because in the following we need to use data from 16 and 17 years for verification, and because of the 2008 subprime mortgage crisis, the annual throughput in 2009 declined. Therefore, when using a univariate linear regression model, the data for these three years need to be removed.Use the remaining seven years of data to make predictions. The model with GDP as independent variable and throughput as dependent variable is as follows. The displayed R-squared value is 0.9494. So a high degree of fit could be used for prediction.

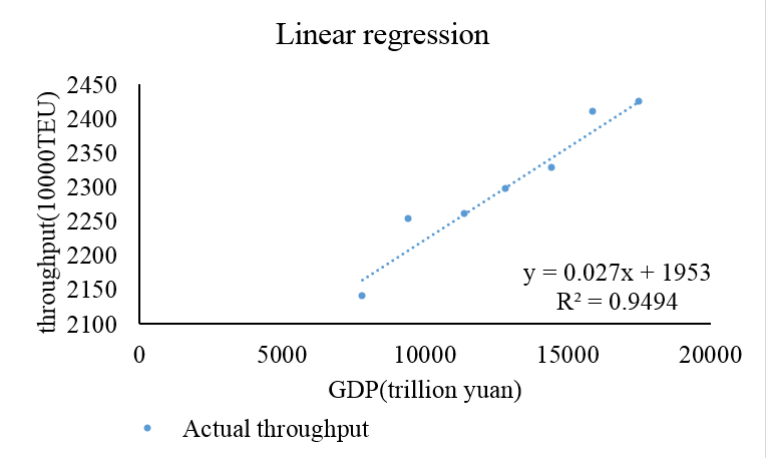

Figure. 3 Result of the linear regression

3.3 Prediction of linear regression model based on least squares principle

Compare the predicted value obtained from the exponential smoothing model and the linear regression model with the actual value. Use the formulas (3), (4), and (5) to obtain the combined model: $y_t = 3.428X_1(t) - 2.428X_2(t)$ In the formula, $X_1(t)$ and $X_2(t)$ respectively represent the predicted value of the linear regression model and the exponential smoothing model.

3.4 Prediction of combined model based on validity principle

Compare the predicted value obtained from the exponential smoothing model and the linear regression model with the actual value. Use the formulas (6), (7), (8), (9)and (10) to obtain the combined model: $y_t = 0.5049X_1(t) + 0.4951X_2(t)$ In

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

the formula, $X_1(t)$ and $X_2(t)$ respectively represent the predicted value of the linear regression model and the exponential smoothing model.

4. Results

In order to evaluate the prediction effect of the model, this paper uses Shenzhen 2016 and 20017 container throughput data to verify the model. After searching the statistics through the data, it can be obtained that the container throughput of Shenzhen in 2016 was 23.9793 million TEUs, and the GDP was 1949.26 billion yuan. In 2017, the container throughput was 25.207 million TEUs and the GDP was 2.243839 billion yuan. The above four models are used to predict the container throughput in 2016 and 2017, and the results are shown in Table 2.

Table 2 Error comparison of four prediction models

Model	Prediction of 2016 (10000TEU)	Error(%)	Prediction of 2017 (10000TEU)	Error(%)
Exponential smoothing	2355.43	1.8	2405.06	4.5
Linear regression	2479.3	3.4	2558.84	1.5
Based on Least squares combined model	2780.01	15.9	2932.22	16.32
Based on validity combined model	2418	0.84	2482.7	1.5

From Table 2, each model can be evaluated as follows: First, the prediction effect of the exponential smoothing model is worse than the linear regression model from a single model, indicating that the linear regression model can better reflect the current container throughput development trend of Shenzhen Port. Second, the accuracy of the exponential smoothing linear regression model established by the validity is better than that of the single exponential smoothing model and linear regression model, which should be used as the optimal model for Shenzhen container throughput prediction. Third, the combined model based on the principle of least squares seems to have a large error [5]. It may be a problem with the data itself, but it is not suitable for throughput prediction.

5. Conclusion

In this paper, a combination of quadratic exponential smoothing model and linear regression model, the validity model principle, has achieved good prediction results. Therefore, a combined model based on the validity principle is used to predict the throughput of Shenzhen Port in 2020 by searching the Internet. It is expected that the GDP of Shenzhen Port in 2020 may reach 2868 billion yuan, and the predicted throughput of Shenzhen Port in 2020 is 29.51 million TEU. As there is not much data collected in this paper, the model should continue to verify the

ISSN 2616-5667 Vol.3, Issue 2: 9-16, DOI: 10.25236/AJETS.2020.030202

accuracy of the combined model by predicting the container throughput of other cities and ports.

References

[1] N. Chen, M.Q. Zhu, Z.W. Yu: Port throughput prediction based on logarithmic exponential smoothing, Journal of Wuhan University of Technology, vol. 9 (2005), 77-79.
[2] Y. Liu, Y.M. Chen: Application of Grey System Model to Throughput Forecasting of Inland Ports, Waterway Engineering, vol. 4 (2006), 31 -34.
[3] X.C. Ma, X. Wang, H.F. Li: Prediction of port container throughput based on combined model, Traffic standardization, vol. 1 (2009), 74-77.
[4] Y.J. Liu, Q. Xiao: Application of Combined Model in Forecasting Port Container Throughput, Journal of Dalian Maritime University, vol. 3 (2006), 26-28.
[5] W. Chen, S.J. Zhang, C.H. Hou, et al. Discussion on the calculation method of regression coefficient of quadratic exponential smoothing prediction model, Statistics and Decision, vol 19 (2016), 11-12.