Research Inventy: International Journal of Engineering And Science
Vol.16, Issue 1 (January 2026), PP 01-04
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Prediction of Shanghai Port Container Throughput Based on Multiple Linear Regression

Yu Lu
The Affiliated High School of Hangzhou Normal University
Corresponding Author: Yu Lu

Abstract: *This study focuses on Shanghai Port, utilizing multi-dimensional data from 2005 to 2024. It examines the impact of various factors on the port's container throughput from four perspectives: infrastructure, economic structure, port efficiency, and port development potential. A multiple linear regression model was constructed to predict Shanghai Port's container throughput. The results indicate that Gross Domestic Product (GDP) and the employed population significantly influence container throughput. After refinement, the model can accurately forecast future container throughput at Shanghai Port, providing a scientific basis for various development plans and decision-making processes.*
Keywords: *Prediction of Container Throughput in Shanghai Port, Influencing Factors, Multiple Linear Regression, Stepwise Regression.*

Date of Submission: 01-01-2026 Date of acceptance: 11-01-2026

I. INTRODUCTION

Ports, functioning as critical hubs in the global supply chain and barometers of regional economic development, have their operational efficiency and handling capacity directly impacting a nation's economic competitiveness. As a top-tier port that has ranked first globally in container throughput for over a decade, Shanghai Port serves as a vital gateway for China's foreign trade and a strategic resource for safeguarding the implementation of national major strategies and enhancing the country's global influence and competitiveness. Accurate prediction of its container throughput is therefore of paramount importance for port operations, shipping logistics planning, and policy formulation. Given Shanghai Port's unique status and the sufficiency of relevant data, this paper selects Shanghai Port as the research subject.

Zheng Wei [1] analyzed the foreign trade situation enabling Shanghai Port to maintain its position as the "top container port" from the perspective of new shipping routes, but the analysis lacked in-depth exploration based on mathematical models. Zhou Yuqin, Liu Xiaojia, and Liu Bowei[2] analyzed the competitiveness of Chinese coastal ports under the "Belt and Road" initiative, but their research did not extend into forecasting. Tang Simin, Lan Peizhen, and Zhu Jingjun[3] developed a multiple linear regression model to predict the port throughput of Xiamen Port; however, abrupt changes in variables led to deviations in their prediction results.

Against this backdrop, this study aims to construct a prediction model for Shanghai Port's container throughput based on multiple linear regression. Through model training and data screening, the objective is to establish a robust and reliable forecasting model. This research seeks not only to provide quantitative data support for the future development planning of Shanghai Port but also to assist relevant stakeholders in making more scientific decisions within a complex and volatile environment.

II. METHODOLOGY

2.1 **Selection of Research Variables**

Drawing upon four key dimensions—infrastructure, economic structure, port efficiency, and port development potential[2]—this study selects the container throughput of Shanghai Port Y(in 10,000 TEUs) as the explained variable. Eight influencing factors are chosen as the explanatory variables: Gross Regional Product x_1 (in 100 million yuan), Employed Population x_2 (in 10,000 persons), Total Cargo Throughput x_3 (in 10,000 tons), Number of Berths x_4 (units), General Public Budget Revenue x_5 (in 100 million yuan), Total Length of Operational Quays x_6 (meters), Share of Tertiary Industry Value-Added in GDPx_7(%), and Total Retail Sales of Consumer Goods (in 100 million yuan).

2.2 **Data Sources**

The data utilized in this study were sourced from the National Bureau of Statistics, the Shanghai Municipal Statistics Bureau, the China Statistical Abstract, and the Ministry of Transport. Through a process of screening, cross-referencing, and integration of these four sources, a consolidated dataset was compiled, encompassing relevant influencing factors for the container throughput of Shanghai Port from 2005 to 2024.

Table1 Shanghai Port Container Throughput: Influencing Factor Indicators, 2005–2025.

Year	Y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2024	5150.63	5.4	2480.3	78958.92	1024	8728	84619	78.2	15872.3
2023	4915.83	4.72	2487.4	75277	1002	7992.46	77317	75.2	16385.8
2022	4730.3	4.86	2457	66832.16	1013	7487.22	76829	74.3	14918.2
2021	4703.33	4.71	2489.4	69826.76	1037	7606.9	76480	73.8	16818.3
2020	4350.34	4.16	2488.2	65104.69	1024	7047.52	75817	73.4	15189.2
2019	4330.26	3.8	2481	66351.32	1032	7383.44	75818	72.9	15469
2018	4201.02	3.6	2475	68392	1054	7120.85	75410	70.9	14874.8
2017	4023.31	3.29	2466	70542	1078	6556.66	72473	70.7	13699.52
2016	3713.31	2.99	2467	64482	1152	6090.83	74066	70.9	12588.21
2015	3653.7	2.69	2458	64906	1238	5212.54	75161	68.3	11605.7
2014	3528.53	2.53	2467	66953.94	1220	4464.84	75091	65.3	10592.68
2013	3361.68	2.32	2448	68273.41	1191	4025	74487	63.7	9693.15
2012	3252.94	2.13	2399	63740	1183	3737.62	74459	61	8833.2
2011	3173.93	2.01	2356	62432	1164	3103	72742	58.6	8052.21
2010	2906.9	1.79	2302.7	56320	1160	2743	72537	57.9	6901.39
2009	2500.2	1.57	2210.3	49467.47	1145	2537.58	72274	60	5786.83
2008	2800.6	1.45	2140.6	50808	1141	2420.05	68191	56.5	5053.35
2007	2615.2	1.29	2063.6	49227	174	1748.8	22194	55.1	4250.23
2006	2171.9	1.06	1964.1	47040	175	1581.8	22194	52.6	3681.66
2005	1808.4	0.92	1890.3	44317	183	1259.31	20069	52.1	3230.66

2.3 **Selection of Forecasting Methodology**

Numerous methods are currently available for forecasting port throughput[4]. Commonly used approaches include: 1) time series analysis[5], which predicts based on historical data trends and seasonal patterns; 2) statistical regression methods; 3) artificial neural networks; and 4) genetic algorithms, among others. The appropriate forecasting method should be selected according to the specific circumstances of the port in question, and predictions should then be made in conjunction with the actual data. Based on a statistical analysis of historical data, the multiple linear regression model can effectively screen for variables that have a significant impact on container throughput. Therefore, this study employs this model to construct a forecasting model for the container throughput of Shanghai Port.

III. Model Formulation and Solution

3.1 **Multiple Linear Regression Analysis and Forecasting Model**

Linear regression is a statistical analysis method that utilizes regression analysis in mathematical statistics to determine a quantitative relationship of interdependence between two or more variables[6]. When the regression analysis includes two or more independent variables and the relationship between the dependent variable and the independent variables is linear, it is referred to as multiple linear regression analysis. The form of the multiple linear regression model used in this paper is as follows:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

b_i—coefficient i=1,2,⋯,n; x_j—influencing factors j=1,2,⋯,n; Y—Annual Port Container Throughput; n—Number of Influencing Factors[7]

3.2 **Regression Analysis**

Table 2 Initial Regression Results of the Multiple Linear Regression Model

Variables	Unstandardized Coefficients		Standardized Coefficients	t	P	VIF
	B	Standard Error	Beta			
C	574.128	1612.838	-	0.356	0.729	-
x_1	604.475	152.034	0.882	3.976	0.002***	57.868
x_2	2.24	1.102	0.449	2.033	0.067*	57.381
x_3	0.001	0.011	0.008	0.068	0.947	15.635
x_4	0.709	0.914	0.26	0.776	0.454	132.22
x_5	0.377	0.192	0.971	1.967	0.075*	286.717
x_6	-0.018	0.019	-0.375	-0.948	0.363	184.316
x_7	-63.447	34.084	-0.553	-1.861	0.090*	103.786
x_8	-0.113	0.076	-0.556	-1.478	0.167	166.284
R²			0.991			
Adjusted R²			0.984			
F			F=145.619 P=0.000***			

The regression results are as follows.

$$Y = 574.128 + 604.475x_1 + 2.24x_2 + 0.001x_3 + 0.709x_4 + 0.377x_5 - 0.018x_6 - 63.447x_7 - 0.113x_8$$

IV. Testing and Modification of the Shanghai Port Container Throughput Model

4.1 **Test for Economic Significance**

Holding other variables constant, the regression results indicate that: For every 100-million-yuan increase in Gross Regional Product, container throughput increases by 6.04475 million TEUs; For every 10,000-person increase in the Employed Population, container throughput increases by 22,400 TEUs; For every 10,000-ton increase in Total Cargo Throughput, container throughput increases by 10 TEUs; For every additional Berth, container throughput increases by 7,090 TEUs; For every 100-million-yuan increase in the General Public Budget Revenue, container throughput increases by 3,770 TEUs; For every one-meter increase in the Total Length of Operational Quays, container throughput decreases by 180 TEUs; For every 1-percentage-point increase in the Share of Tertiary Industry Value-Added in GDP, container throughput decreases by 634,470 TEUs; For every 100-million-yuan increase in the Total Retail Sales of Consumer Goods, container throughput decreases by 1,130 TEUs. In summary, x_6, x_7, x_8 are inconsistent with established economic theory.

4.2 **Goodness-of-Fit Test**

The R^2-value represents the goodness of fit, with values closer to 1 indicating a better fit. The results show an R^2-value of 0.991 and an Adjusted R^2-value of 0.984, demonstrating that the equation exhibits a high goodness of fit.

4.3 **Issues and Modifications**

4.3.1 **Multicollinearity Test**

The Variance Inflation Factor (VIF) quantifies the extent to which the variance of a regression coefficient estimate is inflated due to multicollinearity[8]. A VIF value greater than 10 indicates the presence of multicollinearity issues. The model's Mean VIF of 125.525875 indicates severe multicollinearity within the model.

4.3.2 **Stepwise Regression**

After applying the stepwise regression procedure, six variables—Total Cargo Throughput x_3 (10,000 tons), Number of Berths x_4 (units), General Public Budget Revenue x_5 (100 million yuan), Total Length of Operational Quays x_6 (meters), Share of Tertiary Industry Value-Added in GDP x_7 (%), and Total Retail Sales of Consumer Goods x_8 (100 million yuan)—were eliminated. The Gross Regional Product x_1 (100 million yuan) and Employed Population x_2 (10,000 persons) were retained. The regression results based on the final set of variables are as follows.

Table3 Stepwise Regression Results

	Unstandardized Coefficients		Standardized Coefficients	t	P	VIF
	B	Standard Error	Beta			
C	-723.887	485.028	0	-1.492	0.154	-
x_1	547.113	32.35	0.798	16.912	0.000***	2.568
x_2	1.171	0.235	0.235	4.973	0.000***	2.568
R²	0.985					
Adjusted R²	0.984					
F	F=567.879，P=0.000***					

The resulting model is as follows.

$$Y = -723.887 + 543.117x_1 + 1.171x_2$$

In the adjusted model, the R^2-value is 0.985 and the Adjusted R^2-value is 0.984, indicating a high goodness of fit. The p-values for all independent variables are less than 0.05, demonstrating that the model is statistically significant at the 5% confidence level.

Following the stepwise regression, multicollinearity was tested again. The resulting VIF value of 2.5685 indicates that multicollinearity has been effectively mitigated in the model.

V. Summary

The forecasting of Shanghai Port's container throughput is of significant importance for its construction and development. This study analyzed various factors influencing the container throughput of Shanghai Port, established an initial model, and subsequently refined it by selecting the key influencing factors. The resulting multiple linear regression forecasting model is relatively robust and accurate for this specific port, and can provide valuable insights for port infrastructure planning and trade development strategy formulation in Shanghai. However, the analysis of the multiple linear regression model's applicability for short-term forecasting has not been thoroughly explored in this study and warrants further investigation in future research.

REFERENCES

[1]. Zheng, W., 2025. Viewing new foreign trade markets through changes in shipping routes.Ningbo Economy (Finance & Economics Perspective) (09), 38–39.DOI: CNKI:SUN:NBJJ.0.2025-09-011.

[2]. Zhou, Y.Q., Liu, X.J., Liu, B.W., 2025. Analysis of the competitiveness of Chinese coastal ports under the Belt and Road Initiative. Journal of Shanghai Maritime University 46, 58–67.DOI: 10.13340/j.jsmu.202402290023.

[3]. Tang, S.M., Lan, P.Z., Zhu, J.J., 2018. A typical factor prediction model for port throughput.Journal of Shanghai Maritime University 39, 41–44, 102.DOI: 10.13340/j.jsmu.2018.02.008.

[4]. Ge, X.F., 2024. Prediction of container throughput at Zhoushan Port based on grey prediction model.Pearl River Water Transport (11), 25–27.DOI: 10.14125/j.cnki.zjsy.2024.11.014.

[5]. Liang, X.Z., Zhao, X., Yang, M.G., et al., 2024. Ensemble forecasting of port container throughput based on secondary decomposition and model selection strategy.Management Review 36, 52–64.DOI: 10.14120/j.cnki.cn11-5057/f.2024.08.019.

[6]. Zhou, X.J., Jing, Z.Y., 2013. An empirical analysis of logistics demand forecasting in Hebei Province based on multiple linear regression model.Logistics Technology (09), 270–272.

[7]. Huang, B.J., Lin, J.S., Zheng, X.Y., et al., 2013. Prediction of passenger throughput in civil transport airports based on multiple linear regression analysis.Mathematics in Practice and Theory 43, 172–178.DOI: CNKI:SUN:SSJS.0.2013-04-026.

[8]. Lin, J.H., Ren, X.Y., 2025. Research on influencing factors of agricultural product logistics demand in Shandong Province: based on analysis of multiple linear regression model.Logistics Technology 48, 14–17.DOI: 10.13714/j.cnki.1002-3100.2025.12.004.