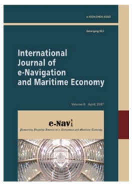
e-Navi

Available online at http://www.e-navi.kr/
e-Navigation Journal



Original article

A study on forecasting model of container cargo throughput of Vietnam's seaport

Tan Vinh Nguyen[a,*], Hoang Phuong Nguyen[a]

[a] Academy of Politics Region II, Ho Chi Minh City, Vietnam. Corresponding Author: vinhnt@hcma2.edu.vn

Abstract

Seaports play a huge role in Vietnam's economy, being a border gate for economic and cultural exchanges with outsiders, especially the role of goods circulation. Container cargoes are one of the types of goods with large and increasing volume through Vietnam seaports. However, the heterogeneity between the seaport and the connected infrastructure greatly affects the capacity and efficiency of port investment. This is also one of the main causes leading to a shortage of goods, excess ports in some port areas. The root cause is that the planning has not kept up with the growth of the amount of goods arriving at the port, because the issue of forecasting the volume of goods through the port is not accurate. Therefore, it is necessary to develop models of forecasting container cargo through the ports with general, scientific, and high accuracy to serve the strategy, planning and development of seaport system; the work of planning and investment in the development of seaports, shipping fleets and other auxiliary transport infrastructure works. The purpose of this study is to build suitable forecasting models with high accuracy and reliability on the total volume of container cargo throughput of the Vietnamese seaport system. Based on the methods of a statistical survey, synthesis, regression analysis, and correlation to evaluate the influence of factors on container cargo volume through Vietnam's seaports in the period of 2004-2019. By incorporating more economic factors into the regression model, the paper focuses on forecasting container cargo through the Vietnamese seaport systems, going into cargo-based forecasting in tons and TEUs. The results of this study contribute to complete the rationale for forecasting, especially forecasts related to the shipping industry and the forecast for container cargo throughput of the seaport. Finally, selecting models for forecasting container cargo volume throughput of seaports by Vietnamese conditions.

Keywords: container cargo throughput, seaports, forecasting model, economic factors, shipping industry.

https://doi.org/10.52820/j.enavi.2020.14.075

1. Introduction

The physical and technical infrastructure of the transport industry is the infrastructure of society. The transportation industry is like a vascular system, which circulates blood throughout the body to exchange substances of the economy. Transport creates a premise and environment for all sectors and economic sectors in all areas of development. Transport facilitates improving people's knowledge, promote the strengths of each region, expand exchanges, contribute to attracting external investment (Tolley & Turton, 2014).

Vietnam's seaport system is part of the transport infrastructure. It is not only to meet the requirements of loading, unloading, storage, transshipment of cargoes and passengers going to and from ports arising from domestic socio-economic development needs but also acts as a driving force for the process of development and international economic integration of coastal regions, regions and localities and the whole country (Vietnam's Transport Sector, 2017). Moreover, it is the basis for reaching out to the sea, developing the maritime economy and maritime services to become the leading spearhead in the maritime economy. Finally, it also contributes greatly to the consolidation of national defense and security, and the maintenance of national sovereignty over seas and islands (X. P. Nguyen & Pham, 2019).

Statistics show that over 90% of Vietnam's exports and imports are through the Vietnamese seaport system. In particular, the proportion of containers in total cargo through the years has changed markedly. If in 2000 it accounted for the lowest proportion of 13.15% of all products, then by 2015 it, along with dry goods, rose to the top position compared to other types of goods (X. P. Nguyen & Pham, 2019). The proportion of containers has a stable growth rate and the highest growth rate compared to other ports in Asia. Besides, the rate of containerization in Vietnam is relatively high, from 2004 to now it has remained above 30% (Pham, 2019). From that, it can be seen that the seaport plays a huge role in Vietnam's economy, being a border gate for economic and cultural exchanges with the outside, especially the role of goods circulation. Container cargoes are one of the most important goods of all import and export goods through the Vietnam seaport system (Athirah, Musa, & Keng, 2019).

However, at present, the Vietnam port system faces several difficulties that need to be addressed. Although the quality and technology level is significantly improved especially in port infrastructure, there still exists asynchronous between the seaport and connected infrastructure. More specifically, it is the asynchronous scale, especially on the implementation process between investment projects to build technical infrastructure systems connecting to the port (Zainal & Jeevan, 2019). It can be mentioned that both the access channel and the logistics hub greatly affect the operational capacity and investment efficiency of the port. This is also one of the main causes leading to a shortage of goods, excess ports in some port areas. That is the lack of synchronization between the port and the transport network in some typical port groups. The underlying reason is that the planning has not kept up with the growth of cargo volume at the port and this is because the forecast of cargo throughput is not accurate (Dragan, Keshavarzsaleh, Intihar, Popović, & Kramberger, 2020).

Building a model to accurately forecast total cargo volume in general and container cargo in particular through seaports in the long-term, medium-term, and short-term is very meaningful (H. P. Nguyen, 2019). It not only helps the strategy and planning of the port system development scientifically, accurately, avoid overcrowding, lack of goods, overloaded ports, asynchronous transport systems connected to seaports, causing congestion for loading or unloading goods in or out of seaports, but also help businesses operating in seaports, import-export business and logistics can build business strategies effectively, close to reality (Du, Wang, Yang, & Niu, 2019).

The researches related to forecasting cargo throughput of ports, as well as the number of containers handled by seaports around the world, are also quite diverse. Farhan et al. forecast cargo throughput in Hong Kong port in their study (Farhan & Ong, 2018). They have forecasted the amount of cargo passing through the port according to the type of cargo, with the forecasting method based on the regression model between the number of goods going through the port and the economic factors affecting depending on the type of goods that the influence of different factors. However, the volume of goods through the port not only depends on GDP, total import and export turnover but also depends on many other factors such as total industrial value, the total value of agriculture, forestry, and fishery, total investment, population, ... (Gao, Luo, & Zou, 2016). Intihar et al. have researched forecasting cargo throughput using the SARIMA model. In this study, they focus on short-term forecasts of total quarterly cargo throughput for Hong Kong and Kaoshung ports using the SARIMA model (Intihar, Kramberger, & Dragan, 2017). Mathilde Jansen has focused on predictive research for three types of ports: transshipment ports or export and import ports (Port in Salalah, Oman; Port in Aukland, New Zeland); Greenfield port (new port) or Brownfield port (existing port) represented by Filyos port in Turkey and port of Rotterdam; ports in a developed and developing economy (Dares Salaam Port in Tanzania and Port of Copenhagen Malmo) (Intihar, Kramberger, & Dragan, 2015). For each type of port, the forecasting method is different, but all must be based on the influence of input factors such as GDP, import, export, population, etc. However, this study only listed and showed the appropriate forecasting method, not showing the forecast results, and the data collected over 10 years (Mohamed Ismael & Vandyck, 2016).

Currently, in Vietnam, there have been many studies on forecasting cargo throughput of ports, including forecasting container cargoes to serve the strategy and planning for the development of Vietnam's seaport system and port group to 2020, orientation to 2030 and port investment projects of the Ministry of Transport. However, these forecasts are only macro, the accuracy is not high, the forecasting method still has many disadvantages and time for making old predictions.

Therefore, it is necessary to have models of forecasting container cargo through the ports of general nature, science, and high accuracy to serve the strategy of planning the development of the seaport system. At the same time, it also meets the requirements of the investment and development planning for seaports, shipping fleets and another auxiliary transport infrastructure works. The paper focuses on building appropriate forecasting models with high accuracy and reliability on the total container cargo volume through the Vietnamese seaport system. Methodically, this study uses statistical survey methodology to collect secondary data on cargo throughput, as well as data on influencing factors. Moreover, the author combines regression and correlation analysis methods to study the relationship of factors affecting container cargo throughput of Vietnamese seaports to develop and select appropriate forecasting models for 2020 and 2030. Regarding research tools, the use of Eviews software in predictive calculations has resulted in a fast, convenient, and highly reliable calculation.

2. Theoretical foundations and methodology

2.1. Correlation between important indicators in socio-economic development and cargo throughput of seaports

To make a forecast of container throughput, the forecaster's approach is to establish a correlation between economic variables and cargo throughput. Can be noticed, the volume of freight as well as the volume through the seaport, if not counting the transportation network formed, the development of vehicles, equipment, and transportation at each stage, depend mainly on the development of the economy. At the macro level, the scale and quality of development of the economy are expressed through the indicators of production, investment, external economic activities, and consumption (Ismael & Vandyck, 2015).

The development scale of the country is reflected in the Gross domestic product (GDP). With the same size of GDP, different structure of the industry, transport demand is different. Therefore, where permissible, it is possible to replace GDP with GDP of industries to reflect the production structure. Regarding the volume of container cargo through the seaport, it is considered to be influenced by the following production factors: i) GDP is the gross domestic product or gross domestic product, is the value of the volume of goods and services produced, which greatly affects the volume of goods through the seaport; ii) Total industrial value is an equally important economic factor affecting container throughput of ports (Quandt, Beinke, Ait-Alla, & Freitag, 2017); iii) The total value of agriculture, forestry and fishery together with the value of industrial production have positively impacted the volume of container cargo through the seaport (H. P. Nguyen, 2019); iv) The more the total investment capital (DT), the more the infrastructure is developed, the more favorable conditions for goods loading and unloading, thereby increasing the volume of goods through seaports (Azwa, Rosni, Saiful, & Saadon, 2019).

The total import-export turnover or total import value M and total export turnover X are also factors that have a direct impact on the volume of transport, especially foreign sea transport in general, as well as the volume of goods through seaports in particular. Total import and export turnover is also an important factor that directly affects the volume of goods through seaports, the greater the import and export turnover, the greater the volume of goods through the seaport (Salleh, Alias, Jeevan, Hanafiah, & Ngah, 2019).

The following criteria can be selected, which are specific to the consumer factor: population (DS), final consumption fund (C). A large number of products directly serve the consumption of the population. The level of product consumption depends on the consumption fund (C) or per capita income of the population (expressed in GDP / DS).

2.2. The basis for conducting forecasting

The forecast study for the volume of cargo through the port in general and the amount of container cargo through the seaport, in particular, is conducted based on the following data sources: (1) Correlation between important socio-economic factors and cargo throughput of seaports; (2) Current situation and socio-economic development strategy in the planning period; (3) The general plan for the development of transport in general and the overall plan for the development of the Vietnamese port system has been approved by the Ministry of Transport or the Prime Minister; (4) International and regional contexts affect cargo throughput of seaports; (5) Demand for transit and transshipment transportation of regional countries for Vietnam; (6) Master planning of specialties: electricity, coal, cement, fertilizer, steel, ...); (7) Planning, feasibility and pre-feasibility studies for ports have been previously studied.

2.3. Forecasting method

In this section, the author focuses on the forecasting methods applied to forecast the stock in general and containerized cargo throughput of seaports, in particular, has been used in Vietnam and some countries in the world. In general, these methods are also based on the above general theoretical and predictive methods.

It can be seen that the methods of forecasting cargo throughput of seaports in particular, also apply the methods of forecasting demand for sea transport in general. But now, in Vietnam, in forecasting cargo through the seaport often combines two methods: the socio-economic scenario method and the extrapolation method through the models (Azwa et al., 2019).

2.3.1. Extrapolation methods

The essence of extrapolation is to extend the rules that have been formed in the past to make predictions for the future. The basic assumption of this method is the preservation of the rhythm, relationships, and the rules of development of past predictors for the future. The information provided for the extrapolation method is data about the dynamics of the past forecasting object over a certain number of years, typically requires a historical period whose data is multiple times larger than the forecasting period. The extrapolation method has the advantage of being simple, however, the main drawback is that it is impossible to calculate the influence of the objective factors on the forecast results (Brezinski & Redivo-Zaglia, 1994).

The forecast for cargo throughput of ports by this method is based on the correlation between cargo throughput of seaports and some key socio-economic indicators.

Extrapolation method via linear function:

If the data series only depends on the time variable, it is most convenient to use the extrapolation method according to the trend function according to the theory of forecasting methods, this is the forecasting method by trend models. In addition to the types of trend functions mentioned above, there are also weighted trend functions. The weighted trend function is an extension of the above six types, from the view that: The closer the data is to the current, the more informative, so people give each node of the series a weight so that the latter has a higher weight than the previous one. Therefore, the weighted trend function gives greater accuracy, but the calculation is also more complicated (Pace & Shaw, 2000).

The nature of time variables involves many inevitable developmental factors. If the data series is highly correlated with the time variable, the prediction in this model is very convenient. The disadvantage of this function is that it accepts the hypothesis: The future is the law of a long past. That is only suitable for short-term forecasts. In this day and age, with the rapid development of science and technology, it is found that every 5-7 years new factors are affecting the object that changes its trend. That limits the applicability of the trend function.

Extrapolation method via regression model:

According to the theory of forecasting method, this is the method of forecasting by regression analysis. However, the model requires some fairly strict conditions, which must ensure the following basic principles: (1) The function form must reflect the true nature of the correlation between the demand for goods transportation of cargo through the seaport and the selected socio-economic criteria; (2) The selected socio-economic criteria must be strictly those that have the greatest impact on the demand for goods transportation through the seaport); (3) The function construction data series must be large enough; (4) The regression coefficients are not 0; (5) There is no multicollinearity between independent variables, if they are variable types; (6) If there is autocorrelation phenomenon, it is necessary to study and add independent variables into the model (Mangan, Kutz, Brunton, & Proctor, 2017).

Regression functions for forecasting in planning often take the form:

$$Y_i = a_0 + a_1.X_{1i} + a_2.X_{2i} + U_i \quad (1)$$

where:

a_0, a_1, a_2: regression coefficients;

Y_i: the volume of freight transport through the port in a year i;

X_{1i}: the value of import and export turnover in a year i;

X_{2i}: GDP value in a year i.

The regression model is widely used in countries, especially the old SEV block, because their data increases steadily, so the correlation coefficient is tight and the forecasting error is quite small. In Vietnam, the regression model has been used for many years in the forecast. However, in the forecast of cargo throughput of the port, the model only mentioned two main economic indicators, namely GDP and import-export turnover, not to mention other indicators. Especially with container cargo through the port is also influenced by many other socio-economic indicators.

2.3.2. The method of extrapolating trend with some other form

The method uses the trend function and self-regression:

Combining the trend function and regression results in better results, namely smaller forecast errors. The content of this method is as follows: Assuming the time series has a tendency function form of f(t), proceed to determine the function f(t) and calculate the autocorrelation coefficient, the prediction model has the following form:

$$Y_t = r.Y_{t-1} + (a + b.t) + \varepsilon_t \quad (2)$$

Method using trend and season functions:

Methods used for trend and seasonal functions include the Seasonal index method and harmonic interpretation method.

The season index method has the following forecasting model:

$$Y_t = X_t.\omega_t + \varepsilon_t \quad (3)$$

where:

X_t is a trend component determined from a series of data that has been averaged sliding with the number of steps equal to the number of seasons.

ω_t is the seasonal fluctuating component calculated by the formula: $\omega_t = Y_t/X_t$.

The method of explaining the harmonic has a prediction model as follows:

$$Y_t = X_t + \omega_t + \varepsilon_t \quad (4)$$

Transport demand is distinctly seasonal. Freight transport depends heavily on the weather. Therefore, the model combining the trend function with another model has many advantages. It makes forecasts more diverse, richer, and more accurate. But all are based on the core assumption that a trend exists so that the unconditional time series cannot be used. In fact, within a few years, the trend has not changed much, so the short-term forecast will give good results.

Extrapolation method is forecasted through the elastic model:

The essence of this method is to determine the correlation function of transport volume with the gross domestic product (GDP). Specifically considering the correlation between the growth rate of the transport volume and the GDP growth rate at a certain time (t_i):

$$E(t) = \frac{(y_t - y_{t-1})/y_{t-1}}{(x_t - x_{t-1})/x_{t-1}} = \frac{a_{vt}(\%)}{a_{GDP}(\%)} \quad (5)$$

where: y_t, y_{t-1} are the transport volume (the volume of goods through the port) in year t and year $_{t-1}$; x_t, x_{t-1} is the GDP values in year t and t-1; E (t) is the elastic coefficient.

After building the correlation function:

$$E(t) = F(y_t, x_t) \quad (6)$$

The value of elastic coefficient E (t) can be determined at any time in the future. Based on the local socio-economic development planning, the GDP growth rate of each period in the future can be determined. This will determine the growth rate of the volume of cargo transporting through the port at future times, plus the volume of cargo handled through the current port. From there, forecast the volume of goods required through the port in the future (Dickhoff, Charity, & Mahzoon, 2017). This method is commonly applied to forecast cargo throughput with advantages and disadvantages analyzed above. However, this forecasting method is based entirely on GDP, and the elasticity coefficient is often calculated incorrectly.

3. Building and selecting a forecasting model of containerized cargo through Vietnamese seaports

3.1. Data

The cargo throughput of Vietnam's seaport system from 2004-2019 (in the direction of cargo) is shown in Table 1.

Table 1: Cargo throughput of Vietnam's seaport system from 2004 to 2019 (Tons)

Year	Export	Import	Inland	Quantity
2004	28996532	23085672	21200345	82423517
2005	35898771	25429020	20101336	91115974
2006	34506936	35008781	22718636	103129467
2007	37871274	39854243	25858321	114108213
2008	47070686	41297572	28966202	127771588
2009	50549745	45923152	29955658	139161413
2010	57581434	49057029	33120927	154497732
2011	62494375	58567996	42940624	181116296
2012	63726431	72364262	42810885	196579572
2013	101402208	69399935	60226900	251218000
2014	74671210	79483042	75500668	259144580
2015	80445578	74430809	87941009	286579768
2016	90344822	83359170	79783547	294549952
2017	104770998	84603672	97629713	328795476
2018	115979683	93731891	118956980	373027220
2019	109952704	122728457	138804902	427816389

From the data table, it is noted that the volume of goods through the Vietnam seaport system in the 16 years from 2004 - 2019 increased sharply with an average growth rate of 11.39%, especially increasing rapidly in 2013 (up 27.8% compared to 2012). In general, exports through ports are higher than imports through ports except for 2006, 2007, 2012, and 2014. If the volume of imports through the port tends to increase steadily over the years, the volume of exports through the port fluctuates more and more but in general, the trend is increasing.

According to the statistics of the Maritime Administration, we have a Table 2 of statistics on container cargo throughput of the Vietnam seaport system in the period 2004– 2019.

Table 2: Container cargo throughput of Vietnam's seaport system from 2004 to 2019 (Tons)

Year	Export	Import	Inland	Quantity
2004	4536800	4728052	1573856	10838708
2005	5754300	5006321	4064551	14825172
2006	6614126	7122966	2512405	16249497
2007	8308103	8455361	3567459	20330923
2008	9986997	10452703	5039532	25479232
2009	11253782	12724515	5623492	29601789
2010	13733319	15512191	7948367	37193877
2011	16695774	21002167	11588391	49286332
2012	19098551	24690903	11670564	55460018
2013	19390749	27521828	15390763	62303340
2014	24153255	32411096	16372590	72936941
2015	26079793	34421287	18929000	79430080
2016	32379263	36972236	21716724	91068223
2017	35578174	41039777	24495388	101113339
2018	40994537	47104621	30295550	118394708
2019	44119419	54857491	27371427	126348337

In general, container cargo throughput increased steadily with an average growth rate of 20.19% and especially the fastest increase in 2011 (an increase of nearly 32% compared to 2010).

Table 3 shows the container cargo throughput of Vietnam seaports in the cargo direction for the period 2004-2019 in TEU units.

In general, the volume of containerized cargo through the port in the direction of import and export is almost the same and tends to increase steadily (in tons, the volume of containers going through the port is slightly higher than that of export). Domestic container cargo through the port also increased but increased slowly and always lower than the direction of export and import.

Table 3: Container cargo throughput of Vietnam's seaport system from 2004 to 2019 (TEU)

Year	Export	Import	Inland	Quantity
2004	496954	513020	136875	1146849
2005	477946	501026	366615	1345587
2006	747555	729971	240059	1717585
2007	883253	874557	285425	2043235
2008	1055586	1034224	343024	2432834
2009	1266160	1217479	436126	2919765
2010	1475048	1428496	507124	3410668
2011	1837189	1878405	773571	4489165
2012	2046790	2105408	871114	5023312
2013	2189192	2248051	1102004	5539247
2014	2686993	2675655	1158356	6521004
2015	2872179	2951088	1386913	7210180
2016	3357548	3162052	1496396	8015996
2017	3488921	3515206	1621899	8626026
2018	4016859	4167195	2214186	10398240
2019	4484230	4686872	2356095	11527197

Table 4: Statistics of Vietnam's import and export turnover from 2004-2019 (Million USD)

Year	Export	Import	Total turnover
2004	14483	15637	30120
2005	15029	16218	31247
2006	16706	19745	36451
2007	20149	25256	45405
2008	26485	31969	58454
2009	32447	36761	69208
2010	39826	44891	84717
2011	48561	62765	111326
2012	62685	80714	143399
2013	57096	69949	127045
2014	72237	84389	156626
2015	96906	106750	203656
2016	114529	113780	228309
2017	132032	132032	264064
2018	150186	148048	298234
2019	162400	165600	328000

Table 5: Statistical data of industrial value, agro-forestry and fishery value, investment capital value of Vietnam from 2004-2019 (Billion VND)

Year	Industrial value	Agro-forestry and fishery value	Investment capital value
2004	198326.1	112112	115109
2005	227342.4	114990	129460
2006	261092.4	122150	147993
2007	305080.4	127651	166814
2008	355624.10	132888	189319
2009	416612.80	137112	213931
2010	486637.10	142711	243306
2011	568140.60	147765	309117
2012	646353.00	158108	333226
2013	701183.80	162593	371302
2014	811181.70	169436	400183
2015	884999.20	177603	371172
2016	962592.80	182317	391732
2017	1051311.10	186381	420341
2018	1126375.00	193650	461551
2019	1232141.60	198685	516937

3.2. Using extrapolation through a linear function

The model set is:

MH1 is the first-order function model:

$$Y_t = \beta_1 + \beta_2.t + U_t \quad (7)$$

where: Y_t is the symbol of container cargo throughput at the port in year t; t is chronological order.

MH2 is a quadratic function model:

$$Y_t = \beta_1 + \beta_2.t + \beta_3.t^2 + U_t \quad (8)$$

MH3 is a log-linear function model:

$$Ln(Y_t) = \beta_1 + \beta_2.t + U_t \quad (9)$$

Conducting regression of container cargo throughput Vietnam seaport in tons, by time variable to consider which regression function is consistent with the data collected. With the help of Eviews software, the results shown in Table 6.

According to the data in Table 6, considering the criteria of the 3 models, it is found that the second-order function is the most suitable (the indicators measuring the forecast accuracy are the smallest).

Do the same for containerized cargo through Vietnamese ports (in TEU units, according to cargo dimensions shown in Table 7). In summary, looking at the summary of the trend, it is noted that the regression model of containerized cargo throughput over time in the period of 2004 - 2019 is secondary.

Table 6: Summarizing the regression function of containerized cargo through Vietnam seaport in tons, by time variable from 2004-2019

Regression function	MAE	MAPE	RMSE
$\hat{Y}_t = -25984803 + 4851694.t$	12100570	190.6485	14229170
$\hat{Y}_t = 9272980 + 2883369.t + 301348.6.t^2$	2187976	41.28144	2640346
$\widehat{Ln(Y_t)} = 14.40987 + 0.181959.t$	5698168	15.99821	11368010

Table 7: Summarizing the regression function of containerized cargo through Vietnam seaport in TEU, by time variable from 2004-2019

Unit	Regression function	RMSE
	The total amount of goods	
T	$\hat{Y}_t = 9272980 - 2883369t + 301348.6t^2$	2640346
TEU	$\hat{Y}_t = 9272980 - 2883369t + 301348.6t^2$	234026.9
	The amount of goods in the export	
T	$\hat{Y}_t = 3633364 - 932645.1t + 100216.2t^2$	1228914
TEU	$\hat{Y}_t = 279982.6 - 73517t + 9502.844t^2$	86849.26
	The amount of goods in the import	
T	$\hat{Y}_t = 4130473 - 1173811t + 124369.3t^2$	1458418
TEU	$\hat{Y}_t = 340562.9 - 86612.73t + 10078.17t^2$	107078.9
	The amount of goods in the domestic	
T	$\hat{Y}_t = -590975.9t + 71469.94t^2$	1249221
TEU	$\hat{Y}_t = 501607.8 - 112050.4t + 7350.28t^2$	88497.2

3.3. Using extrapolation method by the regression model

As mentioned in section 2, it can be seen that the increase and decrease of cargo throughput of ports in general, and the volume of containers through ports, in particular, are closely related to the level of economic development. In other words, it is influenced by macroeconomic factors, including 5 basic economic factors, namely: GDP, total import-export turnover, total industrial value, the total value of agriculture, forestry and fisheries, and total investment.

In addition, there are many other factors affecting the volume of goods through seaports such as: *(1) Import price:* The higher the import price, the lower the volume of goods through the seaport; *(2) Transaction value:* Port service demand is the result of import and export demand; *(3) Population:* The bigger the population, the greater the demand leads to an increase in the volume of goods through the seaport; *(4) Exchange rate:* Unfavorable exchange rate will not promote trade and make the volume of goods through seaports decrease; *(5) Interest rate:* The interest rate determines the value of GDP and GDP is an important influence signal for the loading and unloading port because it determines the volume of exports and imports, these are also variables in determining GDP function; *(6) Inflation:* The inflation rate also influences the determination of GDP; *(7) Transactions between a country's largest trading partner and the country itself:* This factor is determined from the variable "transaction value". The transaction between the two largest customers can be used as a proxy for the transaction of the whole region.

These factors are expressed in quantitative terms. Besides, there are also macroeconomic parameters, such as macroeconomic events, which may affect cargo throughput but do not have historical data. In this study, the authors focus on studying the influence of five economic factors on container throughput of the Vietnamese seaport system.

3.3.1. Single regression model

MH1 $$Y_t = \beta_1 + \beta_2.X_t + U_t \quad (10)$$

where: X_i is the symbol of influencing factors in year t, Y_t is the symbol of containerized cargo throughput of the seaport in year t.

MH2 $$LnY_t = \beta_1 + \beta_2.LnX_t + U_t \quad (11)$$

Conducting regression of total container cargo throughput of the Vietnamese seaport system in tons, with the variables as affecting factors, with the support of Eviews software we have the results shown in Table 8.

The results in Table 8 show that, among the factors affecting the total container cargo throughput of the Vietnamese seaport system in tons, the import-export turnover is the most influential with the largest determination coefficient R^2 = 0.99151, then the industrial value. However, all models have R^2> d, so it is suspected that there is a fake regression phenomenon, except for the regression model of goods volume by import-export turnover. Therefore, only MH 1.2 and MH 2.2 are suitable. Compare MH 1.2 and MH 2.2 found that MH 1.2 has a smaller RMSE, so choose MH 1.2. Moreover, the evaluation results in Table 9 show that the model does not exist defects so choosing MH 1.2 is the most suitable model.

Table 8: Regression models of total container cargo throughput of the Vietnamese port system according to the influencing factor

MH	Regression function	R^2	d	RMSE
1.1	$\hat{Y}_t = -41331673 + 210.8142GDP_t$	0.95487	0.12697	8165376
1.2	$\hat{Y}_t = 392.8091XNK_t$	0.99151	1.38721	3542734
1.3	$\hat{Y}_t = -8337762 + 104.2017CN_t$	0.99110	0.45497	3625329
1.4	$\hat{Y}_t = -1.04E + 08 + 1055.413NN_t$	0.91400	0.26421	11272140
1.5	$\hat{Y}_t = -14364462 + 247.2862DT_t$	0.94731	0.24665	8823292
2.1	$Ln(\hat{Y}_t) = -17.033 + 2.657Ln(GDP_t)$	0.98654	0.74416	6150373
2.2	$Ln(\hat{Y}_t) = -5.682 + 1.03Ln(XNK_t)$	0.98823	1.37403	5480111
2.3	$Ln(\hat{Y}_t) = 1.330468Ln(CN_t)$	0.97959	0.54088	1 754314
2.4	$Ln(\hat{Y}_t) = -41.916 + 4.982Ln(NN_t)$	0.97265	0.87505	8919737
2.5	$Ln(\hat{Y}_t) = -1.6193 + 1.539Ln(DT_t)$	0.98350	1.07182	5962614

where: R^2 is the determining factor, used to assess the suitability of the model, it explains the influencing factor explaining what percentage of the change of containerized cargo through the seaport; d: Durbin - Watson statistics used to detect defects of the model.

Table 9: Evaluate the regression model of total containerized cargo throughput of Vietnam's seaport system by unit tons, by import-export turnover

Defects	Statistical testing	Conclusion
The phenomenon of fake regression model	d = 1.38721> R^2 = 0.99151	No defects
The expected random error is not zero due to the wrong type of function or lack of variables	p = 0.1321 (Ramsey)	No defects
Variance of errors varies	p = 0.0635 (White)	No defects
Self correlation	p = 0.3942 (BG – first-order)	No defects
Random errors do not follow standard rules	p = 0.9421 (JB)	No defects

Proceeding the same way with containerized cargo through the Vietnamese seaport (in TEU units, by cargo dimension) with regression results in Table 10. In Table 10, the economic factors affecting to run a single regression model are GDP_t - Vietnam's gross domestic product year t; XNK_t - Total import and export turnover of Vietnam in year t; XK_t - Vietnam's export turnover in year t; NK_t - Vietnam's import turnover in year t; CN_t - total industrial value of Vietnam in year t; NN_t - the total value of agriculture, forestry and fishery of Vietnam in year t; DT_t - total investment capital of Vietnam in the t year.

From the regression model of container volume through the Vietnam seaport system with economic factors. It can be seen that the volume of container cargo through the port depends on the most on the total industrial value and the total import and export turnover.

3.3.2. Regression multiple regression model

Because production depends on many factors, it is necessary to combine these factors to build a multi-factor model, ie a variable that depends on many independent variables. From the built models, it is necessary to select the best model from MH1 and MH2 with Eq. 12 and Eq. 13.

Conducting a regression of total containerized cargo through the Vietnamese seaport system in tons according to influencing factors, with the support of Eviews software that has resulted in Eq. 14 and Eq. 15.

MH1 $$Y_t = \beta_1 + \beta_2.GDP_t + \beta_3.XNK_t + \beta_4.CN_t + \beta_5.NN_t + \beta_6.DT_t + U_t \quad (12)$$

MH2 $$LnY_t = \beta_1 + \beta_2.Ln(GDP_t) + \beta_3.Ln(XNK_t) + \beta_4.Ln(CN_t) + \beta_5.Ln(NN_t) + \beta_6.Ln(DT_t) + U_t \quad (13)$$

MH1 $$\widehat{Y_t} = -866359.5 - 84.41GDP_t + 210.06XNK_t + 56.74CN_t + 100.83NN_t + 58.33DT_t + e_t \quad (14)$$

p = (0.8517) (0.029) (0.0001) (0.0168) (0.2519) (0.0333)

MH2 $$Ln\widehat{Y_t} = -19.75 + 1.43Ln(GDP_t) + 0.24Ln(XNK_t) - 0.88Ln(CN_t) + 1.33Ln(NN_t) + 0.93Ln(DT_t) + e_t \quad (15)$$

p = (0.044) (0.3737) (0.3215) (0.1482) (0.0525) (0.0000)

Realizing that in MH1 and MH2, there exist some regression coefficients that are not statistically significant and the sign of the estimates is wrong. Specifically, in MH1, the regression coefficient associated with agricultural, forestry, and fishery value variables are not statistically significant and the sign of the coefficient associated with the GDP variable is negative. In MH2, the regression coefficients associated with the GDP variable, import-export turnover, and industrial value are not significant and the sign of the coefficient associated with the industrial value is also negative. This is because, in the multiple regression model, multi-collinearity has occurred that the independent variables have a linear relationship with each other, namely, economic factors themselves are linearly related to each other. To overcome this phenomenon, we have to remove variables from the model. With the support of Eviews software, we can choose the model in which the regression coefficients are statistically significant and have the smallest RMSE:

$$\widehat{Y_t} = -3269576 + 295.5914XNK_t + 65.61218DT_t + e_t$$

With p = (0,0013) (0,0000) (0,0000); R^2 = 0.997462 and RMSE = 1936280. Moreover, the results of the disability assessment of the above model in Table 10 show that the model does not exist defects, so this model is chosen.

Proceeding the same way with containerized cargo through the Vietnamese seaport (in TEU units, according to cargo dimensions) with results shown in Table 11. From the data of Table 11, it can be seen that the factor of Vietnam's total investment capital is important and indispensable in the regression models of container cargo throughput of seaports, with other economic factors. In the previous forecasts, this factor was not mentioned but only focused on GDP and import and export turnover. For the Vietnamese seaport system: The most appropriate model is a regression model of cargo throughput of seaports according to the total import, export, and total investment capital of Vietnam.

Table 10: Evaluation of multiple regression models of containerized cargo throughput of Vietnam's seaport system in tons

Defects	Statistical testing	Conclusion
The phenomenon of fake regression model	d = 2.3428> R^2 = 0.9976	No defects
The expected random error is not zero due to the wrong type of function or lack of variables	p = 0.1318 (Ramsey)	No defects
Variance of errors varies	p = 0.076 (White)	No defects
Self correlation	p = 0.3495 (BG – first-order)	No defects
Random errors do not follow standard rules	p = 0.4006 (JB)	No defects

Table 11: Summary table of regression function of multiple container cargo throughput of Vietnam seaport system from 2004-2019 according to TEU

Unit	Regression function	RMSE
	The total amount of goods	
T	$\hat{Y}_t = -3269576 + 3295.59XNK_t + 65.61DT_t$	1936280
TEU	$\hat{Y}_t = -254617.9 + 24.57XNK_t + 6.74DT_t$	166529.6
	The amount of goods in the export	
T	$\hat{Y}_t = 202.3067XK_t + 21.7869DT_t$	565416.5
TEU	$\hat{Y}_t = -94619.91 + 4.2206XNK_t + 2.5067CN_t$	63182.73
	The amount of goods in the import	
T	$Ln(\hat{Y}_t) = -20.461 + 2.3446Ln(NN_t) + 0.7383Ln(DT_t)$	758394.3
TEU	$\hat{Y}_t = 10.1942XNK_t + 2.3494DT_t$	71444.98
	The amount of goods in the domestic	
T	$\hat{Y}_t = -7592190 + 27.553GDP_t + 53.058XNK_t$	1291545
TEU	$\hat{Y}_t = -155294.3 + 5.7502XNK_t + 1.0453DT_t$	94722.38

3.4. Select a model of forecasting containerized cargo throughput of Vietnam ports by year

To make the best choice of models, we rely on the forecast accuracy of RMSE. And the selected models are presented in Table 12.

Recognizing the multiple regression model is the model with the smallest RMSE of the three models. So the model to choose forecasts is:

$$\hat{Y}_t = -3269576 + 3295.59XNK_t + 65.61DT_t$$

Similarly, the combined results of forecast models for containerized cargo through Vietnam seaports are shown in Table 13.

Table 12: Select the forecasting model of total container cargo throughput of Vietnam's seaport system in tons

Sample regression function	RMSE
$\hat{Y}_t = 9272980 - 2883369.t + 301348.6t^2$	2640346
$\hat{Y}_t = 392.8091XNK_t$	3542734
$\hat{Y}_t = -3269576 + 3295.59XNK_t + 65.61DT_t$	1936280

Table 13: Summary of the forecasting model of containerized cargo throughput of Vietnamese ports

Unit	Regression function	R^2	MAPE
	The total amount of goods		
T	$\hat{Y}_t = -3269576 + 3295.59XNK_t + 65.61DT_t$	0.998	0.013
	p = (0.0013) (0.0000) (0.0000)		
TEU	$\hat{Y}_t = -254617.9 + 24.57XNK_t + 6.74DT_t$	0.998	0.008
	p = (0.0031) (0.0000) (0.0000)		
	The amount of goods in the export		
T	$\hat{Y}_t = 202.3067XK_t + 21.7869DT_t$	0.998	0.012
	p = (0.0000) (0.0000)		
TEU	$\hat{Y}_t = -94619.91 + 4.2206XNK_t + 2.5067CN_t$	0.998	0.008
	p = (0.0046) (0.0000) (0.0000)		
	The amount of goods in the import		
T	$Ln(\hat{Y}_t) = -20.461 + 2.3446Ln(NN_t) + 0.7383Ln(DT_t)$	0.996	0.005
	p = (0.0000) (0.0000) (0.0000)		
TEU	$\hat{Y}_t = 10.1942XNK_t + 2.3494DT_t$	0.997	0.01
	p = (0.0000) (0.0000)		
	The amount of goods in the domestic		
T	$\hat{Y}_t = -590975.9t + 71469.94t^2$	0.983	0.069
	p = (0.0000) (0.0000)		
TEU	$\hat{Y}_t = 501607.8 - 112050.4t + 7350.28t^2$	0.984	0.067
	p = (0.0228) (0.0000) (0.0000)		

From Table 13, the following conclusions can be made: (1) Regression coefficients associated with statistically significant p-values are close to zero); (2) The models selected are suitable not to violate any defects; (3) The determination coefficients of the models are very high, most of which are over 99% of economic factors causing explanations over 99% of the change of containerized cargo through ports; (4) Especially for evaluating the predictive quality of a model usually based on MAPE: MAPE <0.01, the model is very good; MAPE <0.05, the model is good; MAPE <0.1, the model satisfies; MAPE ≥ 0.1, the model is not satisfied. As can be seen, there are no models with MAPE> 0.1 so the models are satisfied. Most of the forecasting models are of good quality, except the forecasting models for containerized cargo through inland ports with MAPE> 0.05 but still satisfied; (5) Although the amount of containerized cargo through the seaport is influenced by the factors as analyzed above when applying the multiple regression model, the factors interact with each other affecting the model so when making the best model selection, it is impossible to include all the influencing factors into the model; 6) From the construction models, we realize that it is not applicable as before: a forecasting model for goods types. Here when building a forecasting model for containerized cargo through the seaport, the model is completely different when the port area is different, when the cargo dimensions are different, even when the units are different.

5. Conclusions

In the process of national economic development, marine economic development is an indispensable and extremely important part. With great sea potential, paying attention to investment and developing the marine economy is an essential task. With the trend of faster shipping development, the urgent problem is that planning must keep up with that development. To do so, the forecasting work for sea transport must be accurate and reliable, including the forecast of the total cargo volume through the seaport (due to the increasing trend of containerization).

The study assessed economic factors affecting container throughput of ports. Building forecasting models by extrapolation method through linear functions, single regression models, and multiple regression models. On that basis, the study has selected a model of forecasting containerized cargo through the Vietnamese seaport in the future by year with high reliability, meeting MAPE criteria (the indicator measuring the accuracy of the forecast). As a result, the study has conducted a detailed forecast in cargo direction, in two units (Tons and TEU) of container cargo through Vietnam seaports. This can be said to be an important basis to help managers in the work of adjusting the planning, development strategy of the seaport system, as well as the infrastructure connected to the Vietnamese seaport system.

6. Acknowledgements

Acknowledgements and Reference heading should be left justified, bold, with the first letter capitalized but have no numbers. Text below continues as normal.

References

Athirah, N., Musa, B., & Keng, L. (2019). Challenge of Malaysia Port toward Shipping Revolution. *International Journal of E-Navigation of Maritime Economy*, *12*, 55–63.

Azwa, N., Rosni, M., Saiful, M., & Saadon, I. (2019). Seaport quality and seaport competitiveness : An innovative assimilation in. *International Journal of E-Navigation and Maritime Economy*, *12*, 1–10.

Brezinski, C., & Redivo-Zaglia, M. (1994). Extrapolation methods. *Applied Numerical Mathematics*. https://doi.org/10.1016/0168-9274(94)00015-8

Dickhoff, W. H., Charity, R. J., & Mahzoon, M. H. (2017). Novel applications of the dispersive optical model. *Journal of Physics G: Nuclear and Particle Physics*. https://doi.org/10.1088/1361-6471/44/3/033001

Dragan, D., Keshavarzsaleh, A., Intihar, M., Popović, V., & Kramberger, T. (2020). Throughput forecasting of different types of cargo in the adriatic seaport Koper. *Maritime Policy & Management*, 1–27.

Du, P., Wang, J., Yang, W., & Niu, T. (2019). Container throughput forecasting using a novel hybrid learning method with error correction strategy. *Knowledge-Based Systems*, *182*, 104853.

Farhan, J., & Ong, G. P. (2018). Forecasting seasonal container throughput at international ports using SARIMA models. *Maritime Economics & Logistics*, *20*(1), 131–148.

Gao, Y., Luo, M., & Zou, G. (2016). Forecasting with model selection or model averaging: a case study for monthly container port throughput. *Transportmetrica A: Transport Science*, *12*(4), 366–384.

Intihar, M., Kramberger, T., & Dragan, D. (2015). The relationship between the economic indicators and the accuracy of container throughput forecasting. In *IAME 2015 Conference Kuala Lumpur, Malaysia*.

Intihar, M., Kramberger, T., & Dragan, D. (2017). Container throughput forecasting using dynamic factor analysis and ARIMAX model. *Promet-Traffic&Transportation*, *29*(5), 529–542.

Ismael, H. M., & Vandyck, G. K. (2015). Forecasting Container Throughput at the Doraleh Port in Djibouti through Time Series Analysis. In *Applied Mechanics, Mechatronics And Intelligent Systems-Proceedings Of The 2015 International Conference (Ammis2015)* (p. 341). World Scientific.

Mangan, N. M., Kutz, J. N., Brunton, S. L., & Proctor, J. L. (2017). Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rspa.2017.0009

Mohamed Ismael, H., & Vandyck, G. K. (2016). Forecasting container throughput at the doraleh port in djibouti through time series analysis. In *Proceedings of the 2015 International Conference on Applied Mechanics, Mechatronics and Intelligent Systems (AMMIS2015)* (pp. 341–350). World Scientific.

Nguyen, H. P. (2019). Blockchain - an indispensable development trend of logistics industry in Vietnam : Current situation and recommended solutions. *International Journal of E-Navigation of Maritime Economy*, *13*, 14–22.

Nguyen, X. P., & Pham, V. T. (2019). The orientation for the development strategy of seaport system in Ho Chi Minh city by key measures to enhance the efficiency of port system management. *International Journal of E-Navigation of Maritime Economy*, *11*, 24–32.

Pace, C. N., & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins: Structure, Function, and Genetics*. https://doi.org/10.1002/1097-0134(2000)41:4+<1::aid-prot10>3.3.co;2-u

Pham, V. T. (2019). The expanding tendency of logistics major in the 4 . 0 industrial revolution : A case study in Vietnam. *International Journal of E-Navigation of Maritime Economy*, *11*, 1–13.

Quandt, M., Beinke, T., Ait-Alla, A., & Freitag, M. (2017). Simulation Based Investigation of the Impact of Information Sharing on the Offshore Wind Farm Installation Process. *International Journal of E-Navigation of Maritime Economy*, *7*, 042–054. https://doi.org/10.1155/2017/8301316

Salleh, N. H. M., Alias, N. A., Jeevan, J., Hanafiah, R. M., & Ngah, A. H. (2019). A Perspective of Malaysian Marine Training Providers and Shipowners on Communication Issues Onboard Merchant Vessels. *International Journal of E-Navigation of Maritime Economy*, *11*, 33–43.

Tolley, R., & Turton, B. J. (2014). *Transport systems, policy and planning: a geographical approach*. Routledge.

Vietnam's Transport Sector. (2017). Vietnam's Transport Sector. *Vietnam National Transport Strategy Study (VITRANSS)*.

Zainal, Z., & Jeevan, J. (2019). The Evolution of Seaport Competitiveness in Malaysia Seaport System. *International Journal of E-Navigation of Maritime Economy*, *12*, 100–110.

Received **21 May 2020**

Revised **24 June 2020**

Accepted **27 June 2020**