FORECAST CONTAINER THROUGHPUT VOLUMES OF SEA PORT CLUSTER NUMBER FIVE BY ARIMA MODEL

Ha Minh Hieu
haminhhieu06@gmail.com
Faculty of Commerce, University of Finance and Marketing, HCMC, Vietnam
Nguyen Duc Bang
nguyenducbang@gmail.com
Faculty of Economics and Law, University of Finance and Marketing, HCMC, Vietnam

Abstract

This study applies ARIMA model in forecasting the number of containers through seaport cluster No. 5 according to Decision No. 1745 / QD-BGTVT dated August 3, 2011. Data were collected from the Vietnam Maritime Administration, the Vietnam Port Association (VPA) for the period 1995-2020 and processed using the software "R". The results show that ARIMA model (0.2.1) is suitable for forecasting container throughput of seaport cluster No. 5 in the period of 2021-2025 with an average forecast of 6.77% corresponding 946,716 TEUs. With this result, policymakers and businesses will help to formulate policies, plans as well as to plan properly and reasonably.

Keywords: *Forecast, TEU, ARIMA, Container, Throughput.*

1. Introduction

According to the planning of development of Vietnam's seaport system up to 2020, orientation to 2030 and the detailed plan approved in Decision No. 1745 / QD-BGTVT dated August 3, 2011, the seaport cluster number 5 is the inter-port system between Ho Chi Minh City (Ho Chi Minh City), Dong Nai and Ba Ria - Vung Tau, including the entire system of ports in Ho Chi Minh City on the Saigon River, Nha Be River. and the deep-water port Cai Mep - Thi Vai in Ba Ria - Vung Tau. This port cluster system plays an important role not only for transportation but also for the development of the southern key economic region, which contributes more than 40% of the country's GDP and attracts more than 50% FDI of Việt Nam[1]. Therefore, forecasting container throughput volumes seaport cluster No. 5 is a very important activity for the government, seaport operators, FDI investors, and import-export enterprises in the region in establishing planning, planning policy, operation management. The more accurate the forecast results, the more feasible the policy making as well as the plan are. The port container throughput projections are often used directly in container port operations and often use a TEU (Twenty Equipvalent Unit) of one TEU equivalent to one 20 feet container. This forecast also relates to the purchase of additional machinery, equipment, supplies, and worker placement and placement. Moreover, unlike

[1] http://www.baodongnai.com.vn/kinhte/201905/vung-kinh-te-trong-diem-phia-nam-vai-tro-dan-dat-trong-phat-trien-2944651/

other manufacturing industries, the container terminal industry cannot increase the capacity and size of the container terminal and yard to meet changing needs but must rely on strategies such as inventory management. warehouse, outsourcing, overtime. Therefore, forecasts are necessary for control and planning of the container port system as well as for the port operator in decision-making and planning. In this study, the authors use the univariate forecasting method, in which the future value time series is assumed based only on past values to predict the future container throughput of ports. hybrid. Historical data is analyzed to identify models and build future forecasting models. And ARIMA model is the model selected on the "R" treatment tool to forecast container throughput of port cluster No. 5 in the next 5 years. Research data were collected from the Vietnam Maritime Administration and the Vietnam Port Association (VPA) for the period 1995-2020.

Table 1: Statistics of TEUs throughput of port cluster No. 5 for the period 1995-2020

Unit: TEU

1995	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**
80,101	18,377	526,446	413,663	686,900	649,068	684,578
2002	2003	2004	2005	2006	2007	2008
380,492	1,037,443	1,308,840	1,525,316	1,850,052	3,036,453	3,388,037
2009	2010	2011	2012	2013	2014	2015
3,388,037	3,003,803	4,250,974	4,491,350	5,886,428	6,953,892	7,689,640
2016	2017	2018	2019	2020		
7,876,100	8,594,634	9,537,225	11,273,997	12,221,725		

Source: Vietnam Maritime Administration; VPA, 2020

2. Literature Review and Research Model

There are many models used to forecast container throughput of ports, the most common being a regression model to identify and measure causality of variables such as the research of T.C. Guo (1993) or H.D.Liu. et al. (1996). It is necessary to consider the estimated error caused by the factors of predictive uncertainty, so some studies find it is necessary to adjust the forecast using alternative techniques. As a study by J.G. Gooijer et al. (1989) used diverse time series models to define the navigational steel traffic flow at the port of Antwerp. Meanwhile W.H.K. Lam (2004) proposes a No-Ron (NN) network model to forecast cargo traffic through Hong Kong ports. And, C.C.Chou (2004) used classified integral representation method, a new method to calculate the opacity to predict the total volume of import and export containers for Taiwan ports and found that the forecast results. better than seasonal forecast (SARIMA). In addition to individual models, some authors have used combined models to predict container throughput through ports such as studies by Shih-Huang Chen et al. (2010) using genetic programming. (GP), analytical method (X-11) and moving average integrated with automatic seasonal regression (SARIMA) to forecast container throughput through major Taiwan ports or such as Gang Xie et al. (2012) used the least squares-based approaches to support vector regression (LS SVR) model to forecast the container throughput at the proposed ports and the results show the resulting. The proposed combination can achieve better predictive performance than the individual approaches. According to Mingfei Niu et al (2017), this group of authors built a synthetic model to combine many models such as VMD-ARIMA-HGWO-SVR for prediction for the purpose of improving stability and accuracy. of container throughput through Chinese ports. In summary, the above forecasting models have their own advantages and disadvantages and depend on statistical data. Meanwhile, according to Robert et al., (1979), the ARIMA model is very suitable for linear relationships between present data and past data. Moreover, Brockwell et al., (2001) also suggested that the ARIMA model will forecast more accurately when the data are detailed in the year.

3. Methods

Box & Jenkins (1970) first introduced the Autoregressive Integrated Moving Average (ARIMA) model in time series analysis, known as the Box-Jenkins method. The ARIMA model is combined by 3 main components: AR (Self-Regression Component), I (Stationary Calculation of Time Series) and MA (Moving Average Component). According to Gujarati (2006) as well as R. Carter Hill et al., (2011) to use the ARIMA model in time series prediction, it must go through 4 steps.

Step 1: Identify the model

To use ARIMA model (p, d, q) in prediction, it is necessary to identify three components p, d and q of the model. The component d of the model is identified through the time series stationary test. If the time series stops at order 0 we denote I (d = 0), if the first difference of the stop sequence we denote I (d = 1), if the second difference of the stop series we denote I (d = 2). To test chain stop using the modified Dickey – Fuller unit root test (ADF) and Phillips-Perron test (D. Dickey and W. Fuller, 1979): $\Delta Yt = \beta 0 + \beta 1t + \pi Yt{-}1 + \sum_{j=1}^{t} \Delta Yt - 1 + \varepsilon$. Test hypothesis H0 with $\pi = 0$, H1 with $\pi < 0$ and use student statistics (symbol t). After the stationary test, we will determine the order of the self-regression process (AR) and the moving average process (MA) through the autocorrelation chart (ACF) and partial correlation chart (PACF). According to G. Box and Jenkin (1970), the p-order regression procedure, symbol AR (p) is: $(Yt - \delta) = \alpha 1(Yt{-}1 - \delta) + \alpha 2(Yt{-}2 - \delta) + \cdots + \alpha p\,(Yt{-}p - \delta) + ut$.

Where: Yt is the time series, δ is the expectation of the series Yt, ut is the white noise. According to D. N. Gujarati and D. C. Porter (2009), the q-order moving average procedure, denoted MA (q) is: $Yt = \mu + \beta 0ut + \beta 1ut{-}1 + \beta 2ut{-}2 + \cdots + \beta qut{-}q$. Combining we have ARIMA model (p, q) as follows:

$$Yt = \Theta + \alpha 1Yt{-}1 + \alpha 2Yt{-}2 + \cdots + \alpha pYt{-}p + \beta 0ut + \beta 1ut{-}1 + \beta 2ut{-}2 + \cdots + \beta qut{-}q.$$

Identification of the ARIMA model (p, d, q) is to find suitable values of p, d, q, where d is the degree of difference of the time series to be investigated, p is the order of self-regression and q is the order moving average. Determining p and q will depend on the graphs PACF = f (t) and ACF = f (t).

Step 2: Estimating the parameters and selecting the model.

The model parameters will be estimated using the "R" tool. The process of model selection is the process of experimenting and comparing adjusted criteria R^2, AIC and Schwarz until we choose the best model for prediction.

Step 3: Test the model

To ensure the model is suitable, the error of the model must be white noise, you can use the ACF correlation chart or Breusch-Godfrey test to check the autocorrelation of the error. For variable variance, either the White or ARCH test can be used. Besides, to evaluate the reliability of the predictive model, the study uses MAPE index (Mean Absolute Percent Error). According to C. Lewis (1983), the MAPE greater than or equal to 50%, the forecast is incorrect, 20% - 50% is valid, 10% -20% is a good forecast, and less than 10% is a perfect forecast. The MAPE index is defined as follows:

$$\text{MAPE} = \frac{100}{n} x \sum_{i=2}^{n} \left|\frac{xt-\mu t}{xt}\right|$$

In which "xt" is the true value and "μt" is the forecast value at time t, n is the total number of predictions.

Step 4: Forecast

After testing the error of the forecasting models, if appropriate, it will be used in the forecasting.

4. Results

4.1. Building ARIMA model

To build the ARIMA model, we must first test the solitude of the chain of container throughput through the port. Dickey-Fuller (ADF) and Phillips-Perron (PP) test results show that / t-stat / are both less than the critical value. Therefore, it is concluded that the chain of container throughput through the port does not stop. Shown in figure 1

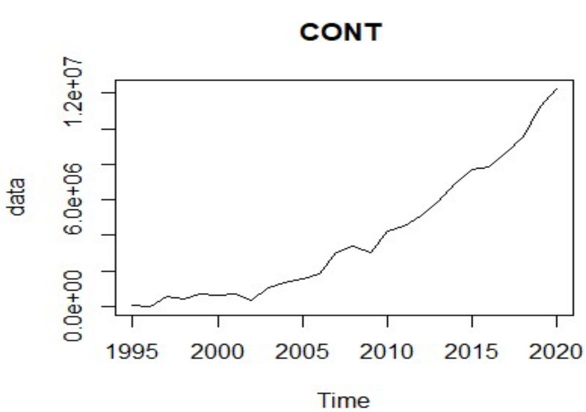

Figure 1: Zero difference graph

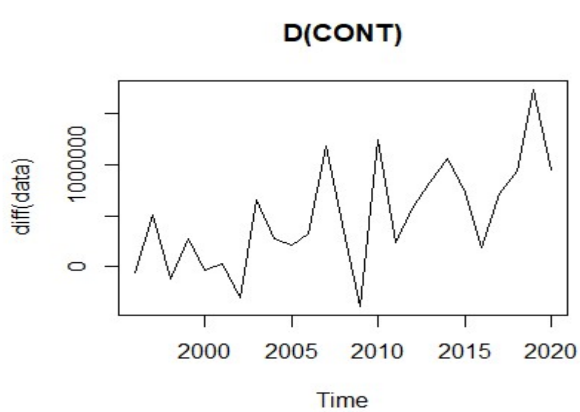

Figure 2: Graph of difference of order 1

Since the container throughput chain through the port does not stop at the zero difference, the first difference must be taken to test the stopping and the result of the chain still does not stop shown in Figure 2. Continuing to take the second difference, the production chain. The volume of containers through the stopped port is shown in Figure 3. Results of the ADF and PP tests are shown in Table 2 after the chain has stopped. And the stationary test results show Table 2 with graph ACF in figure 4 and PACF in figure 5, showing the chain stopped and significant P <0.05, which means that it is reliable to predict.

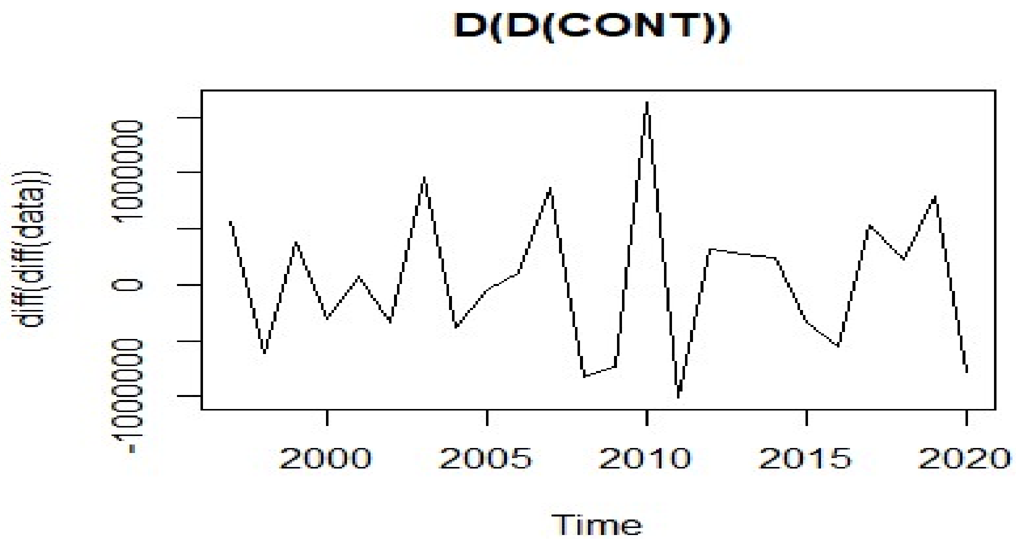

Figure 3: Graph of difference of order 2

Table 2: Stability TEST results

Variale	Value ADF	P-value
CONT	0.077666	>0.99
D(CONT)	-3.9485	0.02511
D(D(CONT))	-5.4249	<0.01

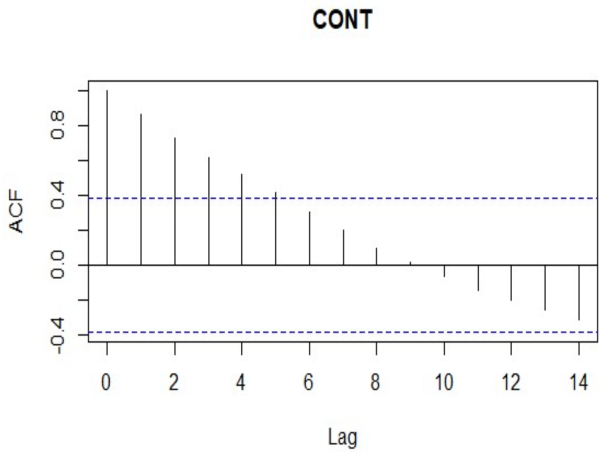

Figure 4: ACF chart

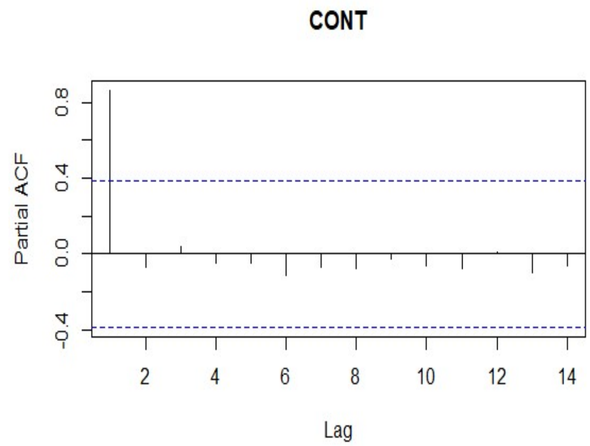

Figure 5: PACF chart

4.2. Estimation of model selection parameters

To find the most suitable prediction model, it is necessary to use empirical methods by comparing adjusted R2, AIC and Schwarz with the smallest AIC. The comparison results show that ARIMA model (0,2,1) is the most suitable model for statistical data set. The results of the models are shown in Table 3.

Table 3. AIC results of some models

Model	AIC
ARIMA (0,2,2)	702.2987
ARIMA (0,2,0)	712.3595
ARIMA (1,2,0)	707.0652
ARIMA (0,2,1)	701.254
ARIMA (1,2,1)	702.7549

4.3. Test the model

To know if the ARIMA model (0,2,1) violates the assumptions of the regression model, it is necessary to do some more tests. The White test shows that the model has no variance and variance. The Breusch-Godfrey test showed that the error had no similar correlation, the Jacque-Bera test showed that the random error had a normal distribution. Moreover, the calculation and comparison of the MAPE index of the models showed that ARIMA (0,2,1) had higher accuracy in terms of both reference data and test data shown in the table 4 and the model error are shown in Table 5

Table 4. Regression results of ARIMA model (0,2,1)

	Estimate	S. E	t.value	p.value	Lag
MA1	-0.754	0.111	-6.81	3.85e-07	1

n = 26; 'sigma' = 483936.9; AIC = 701.254; SBC = 702.5121

Table 5. Model error

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
136105.7	464951.5	338816.6	0.7716496	20.78972	0.6075287	-0.2493524

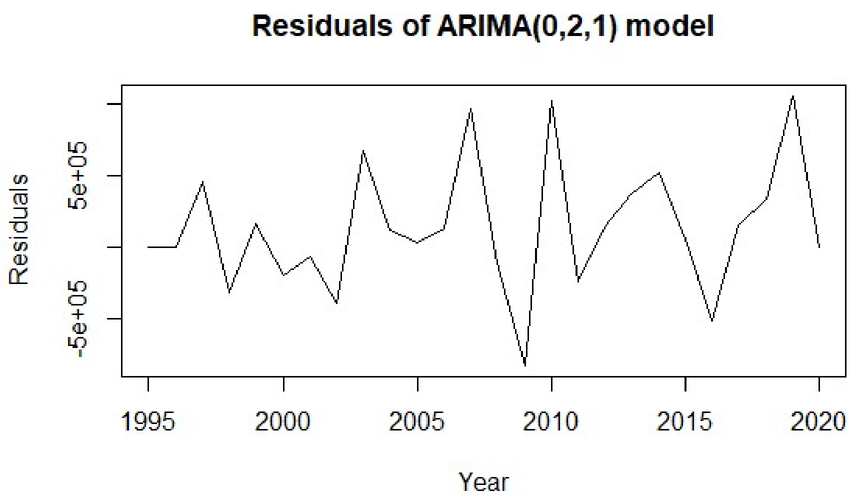

Figure 6: Steady test chart of residuals

Continuing to test the residue of the residue in the ARIMA model (0,2,1), we see that H0, the residual of the model is not stationary and the remainder of the model has no correlation phenomenon. It can be concluded that the statistics are sufficiently reliable for the forecast. Results are shown in figure 6; Table 6; figure 7 and table 7

Table 6. Results of the stationary residue test of the model

Type	Lag	ADF	P-value
no drift no trend	0	-5.63	<=0.01
	1	-3.93	<=0.01
	2	-2.62	0.0112
with drift no trend	0	-6.19	<=0.01
	1	-4.83	<=0.01
	2	-3.59	0.0158
with drift and trend	0	-6.36	<=0.01
	1	-5.13	<=0.01
	2	-4.20	0.0163

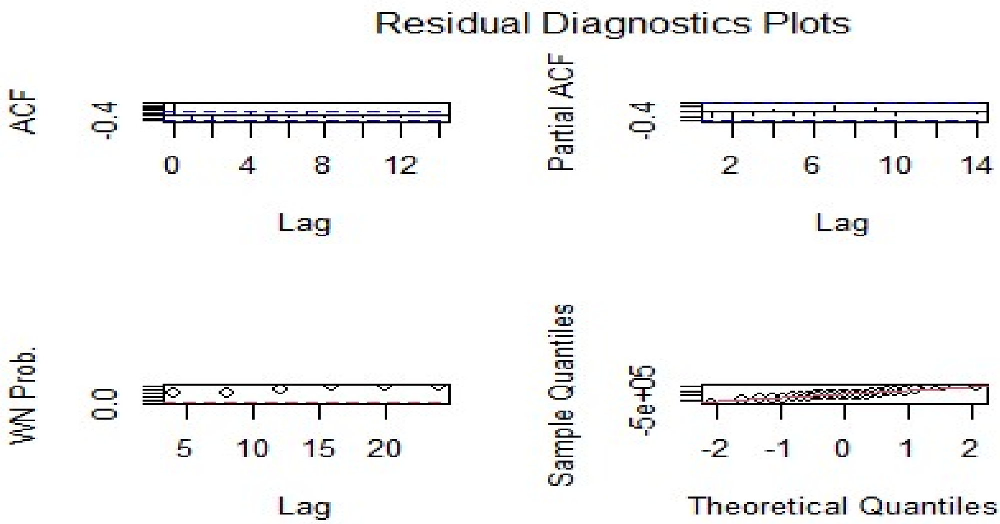

Figure 7: ACF and PACF diagram of remainder

Table 7. Test of similar correlation

Lag	**LB**	**p.value**
4	3.17	0.530
8	6.99	0.538
12	7.98	0.787
16	8.28	0.940
20	8.78	0.985
24	9.00	0.998

Continuing to test the ARCH effect to test the variance of variance, the test results show that H0: The variance of the model is unchanged, which means that the data is statistically significant. The ARCH effect test is shown in Figure 8 and Table 8 shows the results of testing the variance of error

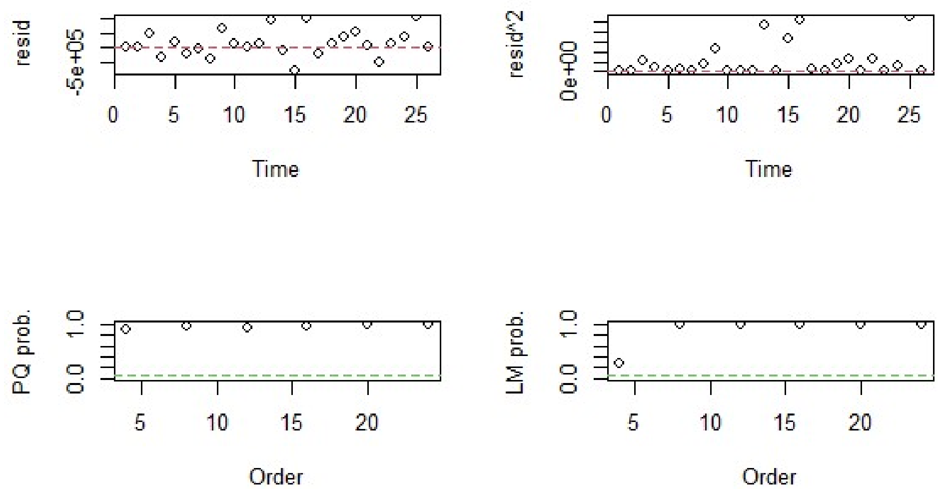

Figure 8. Testing the ARCH effect

Table 8. Results of testing variance of variation

order	**PQ**	**p.value**	**LM**	**p.value**
4	1.049	0.902	3.789	0.285
8	2.763	0.948	0.482	1.000
12	5.695	0.931	-0.001	1
16	7.098	0.971	0.148	1
20	7.167	0.996	0.245	1
24	7.352	1.000	0.385	1

4.4. Forecast

From the test results, we have the forecast results of the container throughput through the port cluster No. 5 to 2025 are shown in Table 9 and Figure 9.

Table 9. Forecast of container throughput through port cluster No. 5 from 2021-2025

Unit: TEU

	Point Forecast	Low 95%	High 95%
2021	13,168,441	12,199,542	14,137,340
2022	14,115,157	12,566,832	15,663,482
2023	15,061,873	12,942,984	17,180,762
2024	16,008,589	13,301,200	18,715,978
2025	16,955,305	13,633,835	20,276,775

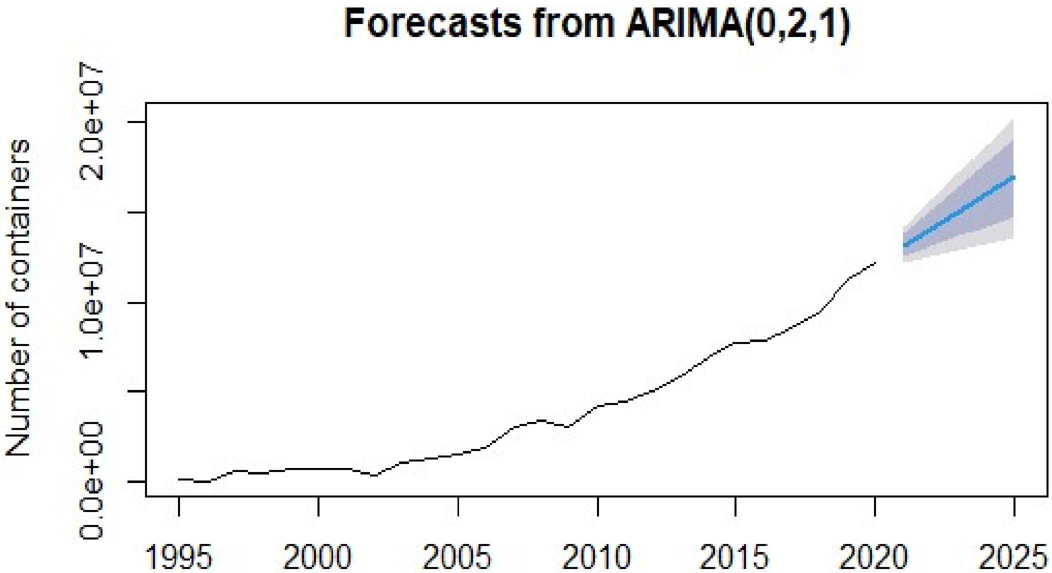

Figure 9: Forecast chart of container throughput through port cluster No. 5 in the period of 2021-2025

The forecast results in Table 9 show that the container throughput through the port cluster No. 5 is forecast to increase on average annually in the period 2021 - 2025 of 6.77% or 946,716 TEUs. In which, the lowest increase volume with reliability p> 0.05% is 2.22%, equivalent to 367,290 TEUs and the volume of containers through the port cluster No. 5 with reliability P <0.05% is 10.69%, corresponding to the level of output of 1,526,142 TEUs.

5. Discussion and Conclusion

Thus, the application of forecast results helps policy makers to deal with the increase in container throughput through the port cluster No. 5 in the future. The government should have solutions to develop infrastructure, connect logistics networks to avoid traffic congestion, which is inherently traffic and the gateway to ports located in this port cluster area 5 have been and are congested as well. serious degradation. For businesses doing business in port operations, they should have strategies to develop and expand ports as well as build and link more satellite systems or have plans to purchase handling equipment to improve their competitiveness. competition as well as efficient port operation. As for import-export production and business enterprises or FDI enterprises, there are plans to choose a cluster of import and export ports to avoid congestion, overload at this port cluster, which will incur a lot of logistics costs. business results.

Research and application of ARIMA model to forecast container throughput of port cluster No. 5 with the aim of finding the best forecasting model for forecasting container throughput of ports in Vietnam. Research results show that, ARIMA model (0,2,1) shows good forecast of container throughput through port and ARIMA model application has been done in a number of studies around the world. However, this study has not compared the forecast results with some other models, these limitations will be addressed in the next study.

5. References

1. C.C. Chou (2004). A study on forecasting the container volume of international ports in Taiwan area. Ph.D. Dissertation, Department of Shipping and Transportation Management, National Taiwan Ocean University.
2. C. Lewis (1983). Industrial and business forecasting methods. *Journal of Forecasting*, vol. 2, pp. 194-196.
3. D. N. Gujarati and D. C. Porter (2009). Basic Econometrics, 5 ed., vol. 5, Canada: Mc GrawHill, pp. 777-784.
4. D. Dickey and W. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, vol. 74, pp. 427-431.
5. G. Box and Jenkin (1970). Time Series Analysis, Forecasting and Control, 4 ed., San Francisco: Holden-Day, 1970, pp. 234-239.
6. Gang Xie, Shouyang Wang, Yingxue Zhao, Kin Keung Lai (2012). Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study. *Applied Soft Computing*, Volume 13, Issue 5, May 2013, Pages 2232-2241
7. H.D. Liu, H.H. Chang (1996). The study on the model for forecasting the port volume. Chinese Technology 17, 89–97.
8. J.G.de Gooijer, A. Klein (1989). Forecasting the Antwerp maritime steel traffic flow: A case study. *Journal of Forecasting* 8, 381–398.
9. Mingfei Niu, Yueyong, Shaolong Sun, Yu Liu (2017). A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. Applied Mathematical Modelling, Volume 57, May 2018, Pages 163-178
10. P. J. Brockwell and R. A. Davis (2001), Introduction to Time Series and Forecasting, 2rd ed., New York: Springer Link, pp. 180-196.
11. R. B. MILLER and J. C. HICKMAN (1973). Time series analysis and forecasting. *Transaction of society of actuaries*, vol. 25, no. 1, pp. 267-329.
12. R. Hill, W. E. Griffiths and G. C. Lim (2011). Principles of Econometrics, 4 ed., New Jersey: John Wiley & Sons, Inc., pp. 512-517
13. Shih-Huang Chen & Jun-Nan Chen (2010). Forecasting container throughputs at ports using genetic programming. *Expert Systems with Applications*, volume 37, Issue 3, 15 March 2010, Pages 2054-2058.

14. T.C. Guo (1993). The analysis for thec argo volume and capacity of KeelungPort. The study on the integration development and deepen for the ports in Taiwan area-The development plan for KeelungPort. The Institute of Transportation, Ministry of Transportation and Communications.

15. W.H.K. Lam, P.L.P. Ng, W. Seabrooke, E.C.M. Hui (2004). Forecasts and reliability analysis of port cargo throughput in HongKong. Journal of Urban Planning and Development 130(3), 133–144.