FORECASTING THE DEMAND FOR PRODUCTION OF VEHICLES USING ARIMA MODEL AT FACTORY No. 3, THANG LONG METAL COMPANY

Nguyen Thi Thanh Binh[1], Pham Thi Minh Hue[1,*]

DOI: http://doi.org/10.57001/huih5804.2024.151

ABSTRACT

The research conducted statistical data analysis to forecast the demand for vehicle claw production for Factory No. 3 of Thang Long Metal Company. By applying the ARIMA time series forecasting model, the team found that the ARIMA model (1,1,3) was most consistent with the data from January 2011 to November 2022. The forecast results of the model have a fairly small error compared to reality. Thanks to this accurate forecast, businesses can make more efficient production plans and avoid unnecessary risks.

Keywords: *Forecast, production, time series, ARIMA.*

[1]Hanoi University of Industry, Vietnam
[*]Email: phamthiminhhue@haui.edu.vn
Received: 20/8/2023
Revised: 12/10/2023
Accepted: 25/5/2024

1. INTRODUCTION

The forecast of production demand at many company in Vietnam has not been paid attention to so far. This makes businesses face many ricks, such as a lack of supply of raw materials, production not meeting consumption demand, ect. This is the problem that Thang Long Metal Company encounters, typically with the motorcycle claw product line. The more accurate the results of the forecast, the more feasible production planning is. There have been many studies using ARIMA models around the world and in the country, such as:

According to Robert, et al., the ARIMA model is well suited for linear relationships between current and past data [1]. Furthermore, Brockwell, et al. suggest that the ARIMA model will forecast more accurately when the data are detailed according to each month of the year [2]. In [3], we compared the forecast of energy consumption in Shandong, China Using the ARIMA model, the GM model and the ARIMA-GM model for 2016 - 2020 will grow at an average annual rate of 3.9% by 2020. The incidence of malaria in Afghanistan [4] was predicted using self-regression integrated mobility average models (ARIMA) to build a predictive tool for malaria surveillance. For short-term projections, malaria incidence can be predicted based on the number of cases in the previous 4 months and the previous 12 months; For long-term prediction, malaria incidence can be predicted using rates from 1 and 12 months ago.

In Vietnam, there have also been a number of studies applying the Arima model, such as: analyzing the effectiveness of three forecasting models used to forecast the daily number of emergency patients at Cu Chi General Hospital, Ho Chi Minh City, Vietnam [5].

2. METHODOLOGY

2.1. Subject of study

The ARIMA model is first used to identify the features like cyclicity and trend of the time series data and to estimate the model parameters. The parameters are then adjusted by the steepest descent algorithm in the adaptive filtering method to reduce the prediction error.In this paper, the Research Team presents the process of building the Arima forecasting problem, applying the Arima time series forecasting model to forecast vehicle production in the next 12 months of 2022 and 2023. The research data is collected from the Factory No. 3 vehicle production statistics for 2011 - 2022, as shown in Table 1.

In this section, samples from the 2011 – 2022, Factory No. 3 are examined in detail to demonstrate how the selected methods are applied and explained. Table 1 shows the monthly change in demand from 2011 to 2022. The chart shows that the output of vehicles at Factory No. 3 tends to increase over time as shown in Fig. 1.

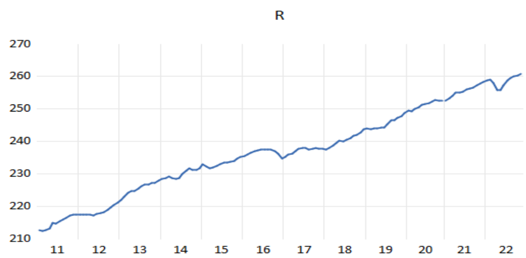


Fig. 1. Monthly changes in vehicle demand from 2011 - 2022

Table 1. Motorcycle claw production No. 3 in 2011-2022 (Unit: 1000 pcs)

Month \ Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	212.71	217.28	221.90	228.33	232.94	235.55	235.34	237.83	243.72	248.82	252.55	258.82
2	212.50	217.35	223.05	228.81	232.28	236.03	235.98	237.51	244.03	249.48	253.18	259.05
3	212.71	217.40	224.09	229.19	231.80	236.47	236.22	237.99	243.72	249.41	254.10	257.95
4	213.02	217.29	224.81	228.71	231.89	236.92	237.00	238.84	244.06	249.96	254.94	255.90
5	214.79	217.20	224.81	228.52	232.45	237.23	237.66	239.44	243.93	250.64	255.17	255.77
6	214.73	217.61	225.40	228.59	232.90	237.50	238.03	240.14	244.18	251.18	255.40	257.21
7	215.45	217.92	226.11	229.92	237.46	237.46	238.03	240.11	244.33	251.48	256.09	258.72
8	215.86	218.28	226.60	231.02	233.54	237.48	237.50	240.60	245.30	251.91	256.29	259.68
9	216.51	219.04	226.75	231.64	233.67	237.43	237.73	241.07	246.45	252.26	256.59	260.21
10	217.23	219.59	227.17	231.25	234.10	236.98	238.02	241.64	246.57	252.78	257.23	260.33
11	217.35	220.47	227.22	231.22	234.72	236.25	237.76	241.99	247.33	252.66	257.82	260.82
12	217.49	221.19	227.84	231.68	235.29	234.75	237.83	242.71	247.85	252.65	258.44	

2.2. Research methodology

ARIMA model (p, d, q): since the Box-Jenkins model only describes stationary or differentiated series, the ARIMA model (p, d, q) represents nonstationary, differentiated data series (d indicates the degree of difference). The Box-Jenkins method includes the following steps:

Step 1: Consider the stationarity of the series of observations.

The ARIMA model is only applicable to stationary series. A stationary process is a stochastic process represented by the sample mean and the variance of the error being constant over time. In fact, most economic data series (base series) are non-stationary. This means that those time series have a time-varying sample mean and variance. To get stationary data, use the difference between the data, where Z_t is the original data.

Differential degrees:

The 0th difference is I (0), which is the original data Z_t.

The first difference is I (1): $w_t = Z_t - Z_{t-1}$

The difference of order d is denoted by I(d): $w_t = Z_t - Z_{t-d}$

The ARMA model (p,q) applied to I (d) is called the ARIMA model (p,d,q).

The stationarity of a time series can be detected based on the graph of the time series and the Augmented Dickey-Fuller test, denoted ADF.

Step 2: Identification the Model

Identifying the ARIMA model (p, d, q) is to find the appropriate values of p, q, and q, where p is the order of autoregression, d is the order of difference of the surveyed time series, and q is the moving average order. The determination of σ and q will depend on the SPACF (Sample Autocorrelation Function) and SPACF (Sample Partial Autocorrelation Function) graphs. A sample autocorrelation function (SACF) is a function or graph of a sample's correlation at lags k = 1, 2,... The sample partial autocorrelation function (SPACF) is a list or graph of a sample's partial correlation values at lags k = 1, 2,... These graphs are simply the points of the SACF and the SPACF plotted against the lags.

Choose the value of p if the SPACF graph has a high value at the lags 1, 2,... p and then suddenly decreases and the SACF function fades out. Similarly, choose the value of q if the SACF graph has a high value at latency 1, 2,… q and decreases sharply after q, and the SPAC form fades out.

Step 3: Estimate the parameters of the ARIMA model (p, d, and q).

The parameters of the ARIMA model will be estimated by the least squares method (Ordinary Least Squares - OLS).

Step 4: Check model diagnostics.

After selecting a particular ARIMA model and estimating the parameters, find out if the selected model fits the data at an acceptable level because it is possible that another ARIMA model will also fit the data. The method of testing the selected model is to see if the error term ε_t of this model has pure randomness (white noise) or not; if yes, the model is accepted. If not, you have to start over from the beginning. The tests that can be used are the BP test (Box-Priere) or the Ljung-Box test with the Q statistic.

Step 5: Forecast by ARIMA model

Based on the most suitable ARIMA model, perform point prediction and confidence interval prediction.

3. RESULTS AND DISCUSSION

The data used is Factory No. 3 Vehicle Output for the period from January 2011 to November 2022 (unit: thousand products), this data was collected by the research team from the data of Factory No. 3 of Thang Long Metal Company (Fig. 1).

3.1. Model recognition

The forecast model of Factory No. 3 for the period from January 2011 to November 2022 is identified as shown in Table 2.

Table 2. Model recognition

Null Hypothesis: R has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.323145	0.4184
Test critical values:	1% level	-4.025426	
	5% level	-3.442474	
	10% level	-3.145882	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(R)
Method: Least Squares
Date: 05/02/23 Time: 02:15
Sample (adjusted): 2011M04 2022M11
Included observations: 139 after adjustments

Since P-Value > 0.05, we cannot reject the null hypothesis (H0: R has a unit root).

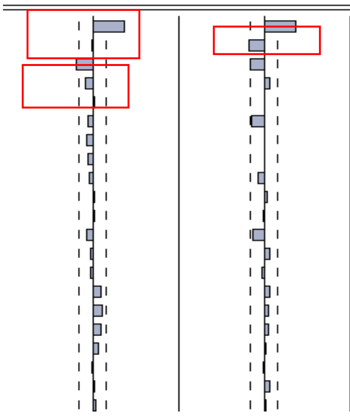
Determine the order of p, d, and q.

To determine the order of the moving average and autoregressive components, focus on the correlation plot of "R" in the differences first. The team that built the correlation graph in the first spreads confirmed that "R" was constant in the first spreads. The purpose of this step is to find all possible models for estimation. ARIMA model identification is to determine the order of AR(p) and MA(q) according to BIC by eq.

To determine the order of the autoregressive component ("p"), the Partial Autocorrelation column (PACF) or Partial Correlations column must be observed. In the column, observe a confidence band on either side. Out-of-range values suggest possible ordering of the autoregressive component. Looking at the correlation graph, the first lag is the highly significant AR (1) component, then lags 2 and 3 are consistent and can be tested. For the purpose of this study, only one AR (1) component is considered.

Table 3. ACF, PACF Autocorrelation column

Date: 05/02/23 Time: 02:18
Sample (adjusted): 2011M02 2022M11
Included observations: 141 after adjustments



	AC	PAC	Q-Stat	Prob
1	0.391	0.391	21.988	0.000
2	-0.005	-0.186	21.992	0.000
3	-0.202	-0.156	27.980	0.000
4	-0.084	0.080	29.028	0.000
5	0.025	0.012	29.123	0.000
6	-0.050	-0.142	29.492	0.000
7	-0.074	0.000	30.314	0.000
8	-0.051	0.004	30.703	0.000
9	-0.042	-0.079	30.968	0.000
10	0.023	0.049	31.051	0.001
11	0.017	-0.008	31.094	0.001
12	-0.076	-0.134	31.992	0.001
13	-0.029	0.072	32.124	0.002
14	-0.022	-0.032	32.198	0.004
15	0.095	0.069	33.642	0.004
16	0.114	0.060	35.751	0.003
17	0.096	0.050	37.245	0.003
18	0.066	0.030	37.964	0.004
19	-0.008	-0.012	37.976	0.006
20	0.021	0.069	38.052	0.009
21	0.043	0.028	38.356	0.012

After the AR order was identified, to identify the order of the moving average component ("q"), the team observed the autocorrelation function (ACF). The ACF suggests that there are two orders of MA. We can see that lags 1 and 3 exceed the confidence band. Therefore, there may be two moving average components: MA (1) and MA (3). Therefore, there may be two moving average components, MA (1) and MA (3), shown in the ACF column in Table 3.

3.2. Estimation and verification with ARIMA model

When determining the components of the ARIMA model, the Group decided to choose a suitable estimation model, which is one of two models: ARIMA (1, 1, 1) and ARIMA (1,1,3)

- In the Jenkins box method, phase 2:
- Estimate the models we identified in phase 1.
- Choose a model based on the importance of coefficient estimates.
- Based on model criteria such as Schwartz, Akaike, and Hannan-Quinn

The model with the smallest value in the model's criteria and the most significant coefficient will be the best fit of the ARIMA (1, 1, 1) and ARIMA (1, 1, 3) models, as shown in Tables 4 and 5.

Table 4. ARIMA Model Results (1,1,1)

Dependent Variable: D(R)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/02/23 Time: 12:15
Sample: 2011M02 2022M11
Included observations: 141
Convergence achieved after 31 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.339917	0.065593	5.182192	0.0000
AR(1)	0.202304	0.176788	1.144331	0.2545
MA(1)	0.239765	0.183300	1.308045	0.1930
SIGMASQ	0.220651	0.019375	11.38817	0.0000

R-squared	0.169815	Mean dependent var	0.341220
Adjusted R-squared	0.151635	S.D. dependent var	0.517381
S.E. of regression	0.476543	Akaike info criterion	1.384828
Sum squared resid	31.11174	Schwarz criterion	1.468481
Log likelihood	-93.63039	Hannan-Quinn criter.	1.418822
F-statistic	9.341128	Durbin-Watson stat	1.971434
Prob(F-statistic)	0.000012		

Inverted AR Roots	.20
Inverted MA Roots	-.24

Table 5. ARIMA Model Results (1,1,3)

Dependent Variable: D(R)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/02/23 Time: 12:17
Sample: 2011M02 2022M11
Included observations: 141
Convergence achieved after 29 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.341056	0.052579	6.486608	0.0000
AR(1)	0.412897	0.055106	7.492770	0.0000
MA(3)	-0.299014	0.089653	-3.335220	0.0011
SIGMASQ	0.209469	0.020779	10.08059	0.0000

R-squared	0.211886	Mean dependent var	0.341220
Adjusted R-squared	0.194628	S.D. dependent var	0.517381
S.E. of regression	0.464311	Akaike info criterion	1.334450
Sum squared resid	29.53507	Schwarz criterion	1.418103
Log likelihood	-90.07872	Hannan-Quinn criter.	1.368444
F-statistic	12.27760	Durbin-Watson stat	1.940763
Prob(F-statistic)	0.000000		

Inverted AR Roots	.41		
Inverted MA Roots	.67	-.33+.58i	-.33-.58i

The identified models were checked for fit based on the test parameters: coefficient of determination R^2, Bayesian Information Criterion (BIC), mean squared error RMSE, and rating index. estimate the accuracy of the MAPE (Mean

Absolute Percent Error) forecast model. It is found that ARIMA (1, 1, 3) is the model that satisfies the most usage criteria, but after building the above models, the following results are obtained: In the ARIMA model (1, 1, 3), there are estimates of the parameter with statistical significance (with a p-value < 0.05) and the minimum BIC criterion.

Therefore, the ARIMA model (1, 1, 3) is the model used for the subsequent estimation. To check the autocorrelation, display the correlation histogram of the ARIMA model (1, 1, 3) and see the Ljung Box Q statistic, where the null hypothesis is "residual is white noise" as shown in Table 6.

Table 6. Comparison table of criteria between 2 models ARIMA (1,1,1) and ARIMA (1,1,3)

Criteria	Model		Best Model
	A) ARIMA(1,1,1)	B) ARIMA(1,1,3)	
AR p-value	0.2545	0.0000	B
MA p-value	0.1930	0.0011	B
SIGMASQ	0.220651	0.209469	B
Log likelihood	-93.63039	-90.07872	B
Akaike	1.384828	1.334450	B
Schwarz	1.468481	1.418103	B
Hannan-Quinn	1.418822	1.368444	B

As can be seen in the figure below, the p-values of the Q-statistic are all greater than 0.05, which confirms that the residuals are white noise. The final step is to confirm whether the inverse AR/MA roots are inside the unit circle.

Thus, the error of the ARIMA model (1, 1, 3) is a stationary series with a normal distribution. This error is white noise as shown in Table 7.

Table 7. Check for residuals with white noise

Date: 05/02/23 Time: 12:52
Sample (adjusted): 2011M02 2022M11
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.024	0.024	0.0812	
		2	-0.064	-0.064	0.6716	
		3	0.025	0.028	0.7644	0.382
		4	-0.038	-0.044	0.9810	0.612
		5	0.088	0.095	2.1383	0.544
		6	-0.063	-0.076	2.7329	0.603
		7	-0.049	-0.030	3.0999	0.685
		8	0.021	0.007	3.1684	0.787
		9	-0.095	-0.091	4.5328	0.717
		10	0.043	0.042	4.8235	0.776
		11	0.046	0.038	5.1461	0.821
		12	-0.101	-0.092	6.7518	0.749
		13	0.040	0.037	7.0089	0.798
		14	-0.042	-0.043	7.2955	0.837
		15	0.077	0.079	8.2532	0.827
		16	0.073	0.044	9.1160	0.824
		17	0.026	0.068	9.2244	0.865
		18	0.075	0.049	10.137	0.859
		19	-0.006	0.008	10.144	0.897
		20	-0.019	-0.015	10.202	0.925
		21	0.014	0.009	10.232	0.947

3.3. Diagnostics and Forecast

In the next step, having determined the most appropriate demand model for the forecast, we must make the forecast. To do this and thus predict trends and develop forecasts, the study used ARIMA forecasting. Presenting the results of the sales forecasts obtained by the team by applying the ARIMA model (1, 1, 3) for the next 12 months from December 2022 to November 2023, as shown in Table 8 and Fig. 2.

Table 8. Forecast results of demand from 12-2022 to 12-2023

2022M12	260.6983
2023M01	261.0588
2023M02	261.4079
2023M03	261.7523
2023M04	262.0947
2023M05	262.4364
2023M06	262.7776
2023M07	263.1188
2023M08	263.4599
2023M09	263.8010
2023M10	264.1420
2023M11	264.4831

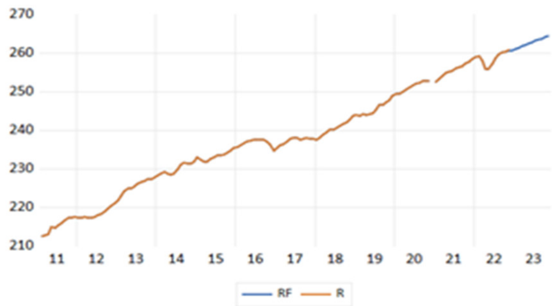


Fig. 2. Forecast diagram of demand from 12-2022 to 12-2023

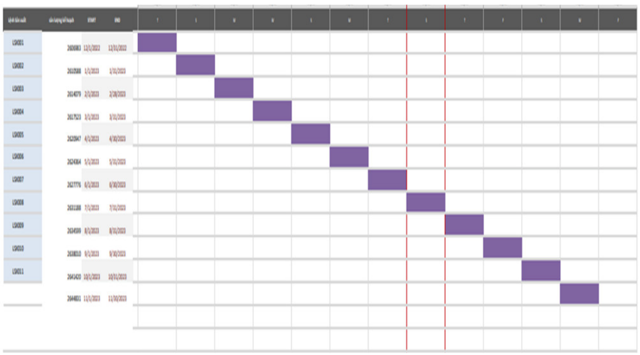
From there, we researched and developed a planned Gantt chart. Production plan from 12-2022 to 12-2023, as shown in Table 9. With the plan in Table 9, you can switch to the Gantt chart as follows:

- Each task is represented in the diagram by a paragraph with a job timetable.

- The beginning and end points of each task correspond to the earliest start and end times of each job.

- Glove jobs are marked with large dots at the ends of the line segment. Gloveless jobs have a time reserve of white space, as shown in Table 9.

Table 9. Gantt diagram showing production plan 12-2022 to 12-2023



4. CONCLUSIONS

Having researched, developed, identified, and selected the ARIMA model Forecasting demand for production of vehicles using ARIMA model at Factory No. 3, Thang Long Metal Company applied to the demand forecast of production demand for enterprises in the period from the next 12 months on the basis of selecting ARIMA models (1, 1, 3) for the most suitable results.

The forecast results show that the production level of ARIMA vehicles at factory No. 3, Thang Long Metal Company, tends to increase slowly in the future (up 2.3% compared to the previous year). The forecast also shows that to meet growth needs, businesses must take measures to promote and invest in technology to achieve the desired sales.

The study also uses gantt charts to help businesses have a new direction in planning production based on ARIMA forecasts for the next 12 months. With gantt charts, it also helps businesses track the progress and implementation of the project from the beginning to the end.

REFERENCES

[1]. Anwar Mohammad, Lewnard Joseph, Parikh Sunil, Pitzer Virginia, "Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence," *Malaria Journal*, 15, 2016. DOI: 10.1186/s12936-016-1602-1.

[2]. Box G., Jenkins G., *Time Series Analysis-Forecasting and Control*. San Francisco: Holden Day, 553p, 1970.

[3]. Li Shuyu, Li Rongrong, "Comparison of forecasting energy consumption in Shandong, China Using the ARIMA model, GM model, and ARIMA-GM model," *Sustainability (Switzerland)*, 9, 2017. DOI: 10.3390/su9071181.

[4]. Cancino Susan, Cancino Giovanni, Cancino Ricketts Daniel, "A Box Jenkins ARIMA model for modeling and forecasting blackberry production in Colombia," *Económicas CUC*, vol. 44, no. 1, 69-82, 2022. https://doi.org/10.17981/kinh.44.1.2023.

[5]. Duong T. A., Nguyen T. P., "Forecasting the daily number of emergency patients in hospitals: A case study," *HUFLIT Journal of Science*, 7(2), 13, 2022. https://hjs.huflit.edu.vn/index.php/hjs/article/view/119