

Forecasting Real GDP Rate through Econometric Models: An Empirical Study from Vietnam

Dr. Nguyen Thi Viet Nga¹

Abstract

Gross Domestic Product is an essential indicator of economic activity, and is usually used by decision makers to plan economic policy. This paper aims at modeling and forecasting the real Gross Domestic Product rate in Vietnam. For this purpose, using the Box- Jenkins methodology with the data covering the period between 1985 and 2018. The optimal model is ARIMA(0,1,2) model. Thanks to this model, the author forecasts the values of real Gross Domestic Product rate for some incoming years of 2019, 2020 and 2021. Statistical results show that the real Gross Domestic Product rate of Vietnam is steadily improving.

JEL Classification numbers: C53, E27.

Keywords: Real Gross Domestic Product rate, ARIMA Model, Box-Jenkins methodology, forecasting, Vietnam.

1. Introduction

Gross Domestic Product (GDP) of a country is the money value of all final goods and services produced by all the enterprises within the borders of a country in a year. It represents the aggregate statistic of all economic activity. The performance of economy can be measured with the help of GDP. According to Eurostat (1996) there are three ways in which the GDP of a country can be measured.

- a) GDP is the sum of gross value added of the various institutional sectors or the various industries plus taxes and less subsidies on products (which are not allocated to sectors and industries) - production approach,
- b) GDP is the sum of final uses of goods and services by resident institutional units (actual final consumption and gross capital formation), plus exports and minus imports of goods and services - expenditure approach,
- c) GDP is the sum of uses in the total economy generation of income account (compensation of employees, taxes on production and imports less subsidies, gross operating surplus and mixed income of the total economy) - income approach.(see The European System of Accounts ESA 1995, Eurostat, 1996).

Forecasting future economic outcomes is a vital component of the decision-making process in

¹ PhD, Department of Economics, Academy of Finance, 58 Le Van Hien, Bac Tu Liem, Hanoi, VIETNAM. E-mail: nguyenthivietnga@hvtc.edu.vn

central banks for all countries. Monetary policy decisions affect the economy with a delay, so, monetary policy authorities must be forward looking, i.e. must know what is likely to happen in the future. Gross Domestic Product is one of the most important indicators of national economic activities for countries. Scientific prediction of the indicator has important theoretical and practical significance on the development of economic development goals. For the forecasting of time series we use models that are based on a methodology that was first developed in Box and Jenkins (1976), known as ARIMA (Auto-Regressive-Integrated-Moving-Average) methodology. This approach was based on the World representation theorem, which states that every stationary time series has an infinite moving average (MA) representation, which actually means that its evolution can be expressed as a function of its past developments (Jovanovic and Petrovska, 2010). The rest of the paper is organized as follows: Section 2 describes literature review while in Section 3, theoretical background is given. In Section 4, the empirical results are presented. Section 5 is the forecasting and finally, conclusions are provided in Section 6.

2. Literature Review

Box and Jenkins (1976) methodology has been used extensively by many researchers in order to highlight the future rates of GDP. Wei et. al. (2010) used data from Shaanxi GDP for 1952-2007 to forecast country's GDP for the following 6 years. Applying the ARIMA(1,2,1) model, the author found that GDP of Shaanxi presented an impressive increasing trend. Maity and Chatterjee (2012) examined the forecasting of GDP growth rate for India using ARIMA(1,2,2) model and a time period of 60 years. The results of their study showed that predicted values follow an increasing trend for the following years. Zhang Haonan (2013) used three models ARIMA, VAR, AR(1) to examine the forecasting of per capita GDP for five regions of Sweden for the years 1993 - 2009. The results of the study showed all three models can be used for forecasting in the short run. However, the autoregressive first order model is the best for forecasting the per capita GDP of five regions of Sweden. Shahini and Haderi (2013) tested GDP forecasting for Albania using quarterly data from the first quarter of 2003 until the second quarter of 2013. For the forecasting, they used two model types ARIMA and VAR. Their results showed that the group of VAR model gave better results on GDP's forecasting rather than ARIMA model. Zakai (2014) investigated forecasting of Gross Domestic Product (GDP) for Pakistan using quarterly data from 1953 until 2012. Choosing a ARIMA(1,1,0) model, the author found out the size of the increase for Pakistan's GDP for the years 2013- 2025. Chaido (2015) also used the ARIMA(1, 1, 1) model for data from 1980-2013 to forecast the real GDP growth rate of Greece in 2015, 2016 and 2017.

3. Theoretical Background

The Box-Jenkins ARMA model is a combination of the AR (Autoregressive) and MA (Moving Average) models as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p P_{t-p} - \alpha_1 u_{t-1} - \alpha_2 u_{t-2} - \dots - \alpha_q u_{t-q} + u_t \quad (3.1)$$

The Box-Jenkins methodology consists of the following phases:

- Establishment of the stationarity of time series. The autocorrelation function (ACF) as well as Augmented Dickey- Fuller test (ADF) (1979) and Phillips-Perron (1988) test (PP) are used for stationarity testing of time-series.
- Model Identification of the model ARMA(p,q). To determine the order of ARMA(p,q), we use the sample of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series. These two plots are suggesting the model we should build. The parameter p of autoregressive operator is determined by the partial autocorrelation coefficient and the parameter q of the moving average operator is specified by the autocorrelation coefficient. In fact we use the limits $\pm \frac{2}{\sqrt{n}}$ for the non-significance of the two functions, so we will have a number ARMA models (a, b), where $0 \leq a \leq p$, $0 \leq b \leq q$. For the optimum model we are using the criteria of Akaike (AIC) and Schwartz (SIC).
- Model Estimation. The involvement of the white noise terms in an ARIMA model entails a nonlinear iterative process in the estimation of the parameters. maximum likelihood estimation is generally the preferred technique.
- Diagnostic checking of the model. With diagnostic checking we investigate whether the estimated model is acceptable and statistically significant, i.e. if it fits well to the data. Box and Jenkins for the adequacy of estimated ARIMA model suggested checking the randomness of the residuals, i.e. whether the residuals from the estimated ARIMA model is white noise, and are not serially correlated.
- Forecasting. One of the main reasons of the analysis of time series models is forecasting. The accuracy of the forecasts depends on the forecasting error. Moreover, a number of statistical measures are employed for this aim, such as root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and the inequality coefficient of Theil (U).

Then the forecast value one period ahead conditional on all information up to time, t, given at time t + k, as:

$$y_{t+k} = (\beta_1 + 2) y_{t+k-1} - (1 + 2\beta_1) y_{t+k-2} + \beta_1 y_{t+k-3} + \alpha_1 \varepsilon_{t+k-1} + \varepsilon_t \quad (3.2)$$

4. Empirical Results

The variable used in the analysis is the GDP growth (annual %) that span from 1985 to 2018. The source of data is the World Bank. The ARIMA approach is an iterative four-stage process of stationary, identification, estimation and testing.

4.1 Testing for Stationarity

Figures 1 and 2 represent the correlogram of the real GDP rate series with a pattern of up to the 16 lags in level and for first differences.

Figure 1: Correlogram of Real GDP Rate Series (Level)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.682	0.682	17.246	0.000
		2	0.333	-0.247	21.490	0.000
		3	0.172	0.113	22.652	0.000
		4	-0.004	-0.236	22.653	0.000
		5	-0.151	-0.044	23.619	0.000
		6	-0.323	-0.311	28.175	0.000
		7	-0.336	0.162	33.291	0.000
		8	-0.257	-0.134	36.403	0.000
		9	-0.273	-0.102	40.054	0.000
		10	-0.178	0.115	41.671	0.000
		11	0.002	0.075	41.672	0.000
		12	0.081	-0.096	42.036	0.000
		13	0.105	0.009	42.678	0.000
		14	0.030	-0.204	42.735	0.000
		15	-0.052	-0.096	42.908	0.000
		16	-0.143	-0.227	44.294	0.000

From the above figure we can conclude that the coefficients of autocorrelation (ACF) starts with a high value and declines slowly, indicating that the series is non-stationary. Also the Q-statistic of Ljung-Box (1978) at the 16th lag has a probability value of 0.000 which is smaller than 0.05, so we cannot reject the null hypothesis that the real GDP rate series is non-stationary. Thus, the series must be configured in first differences.

Figure 2: Correlogram of Real GDP Rate Series (First Differences)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.012	-0.012	0.0051	0.943
		2	-0.296	-0.296	3.2731	0.195
		3	0.111	0.113	3.7498	0.290
		4	0.099	0.012	4.1367	0.388
		5	-0.089	-0.028	4.4656	0.484
		6	-0.238	-0.241	6.8888	0.331
		7	-0.014	-0.063	6.8979	0.440
		8	0.140	0.024	7.8003	0.453
		9	-0.138	-0.122	8.7115	0.464
		10	-0.111	-0.059	9.3338	0.501
		11	0.156	0.049	10.615	0.476
		12	0.030	-0.054	10.666	0.558
		13	0.019	0.096	10.687	0.637
		14	-0.057	-0.083	10.883	0.695
		15	0.147	0.159	12.271	0.658
		16	-0.090	-0.226	12.824	0.686

From the figure 2 we can conclude that the Q-statistic of Ljung-Box at the 16th lag has a probability value larger than 0.05, so we cannot reject the null hypothesis that the real GDP rate series is stationary. The results of Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test on real GDP rate series are representing on Table 1.

Table 1: ADF and Phillip-Perron's Test

	Chuỗi tỷ lệ tăng trưởng GDP		Chuỗi sai phân của chuỗi tỷ lệ tăng trưởng GDP	
	C	C, T	C	C, T
ADF	-2,627521 (3)	-2,541152 (3)	-5,625594 (0) ***	-5,664354 (0) ***
PP	-2,553543 [3]	-2,433494 [3]	-5,625357 [1] ***	-5,664222 [1] ***

Note:

1. Lag length in () and Newey-West value using Bartlett kernel in []
2. Asterisks (***) denote statistically significant at 1% significance levels.

The results in Table 1 indicate that real GDP rate is stationary in first differences. Therefore, for our model ARIMA (p,d,q), we will have the value $d = 1$.

4.2 Identification of the Model

We can use the correlogram of figure 1 to determine the model ARMA (p,q), i.e. the values of parameters p and q. As already mentioned above, an AR(p) model has a PACF that truncates at lag p and an MA(q) has an ACF that truncates at lag q. In practice $\pm \frac{2}{\sqrt{n}}$ are the nonsignificance limits for both functions. We shall explore the range of models ARMA(a,b), $0 \leq a \leq p$, $0 \leq b \leq q$ for an optimum one. To do this we shall use the automatic model determination criteria AIC and SIC. The limits for both functions (ACF, PACF) are $\pm \frac{2}{\sqrt{34}} \approx \pm 0,343$. From figure 1, the ACF cuts off at lag 2 ($q = 2$) and the PACF at lag 1 ($p = 1$). Exploring the range of models $\{\text{ARMA}(p,q): 0 \leq a \leq 1, 0 \leq b \leq 2\}$ for the optimal on the basis of AIC and SIC. Thereafter we create Table 2 with the values of p and q as follows:

Table 2: Comparison of Models within the Range of Exploration Using AIC and SIC

p	q	AIC	SIC
0	1	3,21	3,31
0	2	2,90	3,00
1	0	3,22	3,31
1	1	3,11	3,25
1	2	2,92	3,06

The results from table 2 indicate that according to the criteria of Akaike (AIC), and Schwartz (SIC) the model ARMA is formulated to ARMA(0,2). As the model is stationary on first differences, i.e. ($d = 1$) our ARIMA model will be ARIMA (0,1,2).

4.3 Estimation of the Model

Thereafter we can proceed to estimating the above model. The following table 3 presents the results of this model.

Table 3: Estimation Model ARIMA (0,1,2)

Dependent Variable: D(GDP_RATE)					
Method: Least Squares					
Sample (adjusted): 1986 2018					
Included observations: 33 after adjustments					
Convergence achieved after 12 iterations					
MA Backcast: 1984 1985					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	-0.007689	0.036841	-0.208712	0.8360	
MA(2)	-0.953165	0.034221	-27.85325	0.0000	
R-squared	0.262889	Mean dependent var		0.099089	
Adjusted R-squared	0.239111	S.D. dependent var		1.154087	
S.E. of regression	1.006698	Akaike info criterion		2.909921	
Sum squared resid	31.41669	Schwarz criterion		3.000618	
Log likelihood	-46.01369	Hannan-Quinn criter.		2.940438	
F-statistic	11.05606	Durbin-Watson stat		1.970160	
Prob(F-statistic)	0.002281				
Inverted MA Roots	.98	-.98			

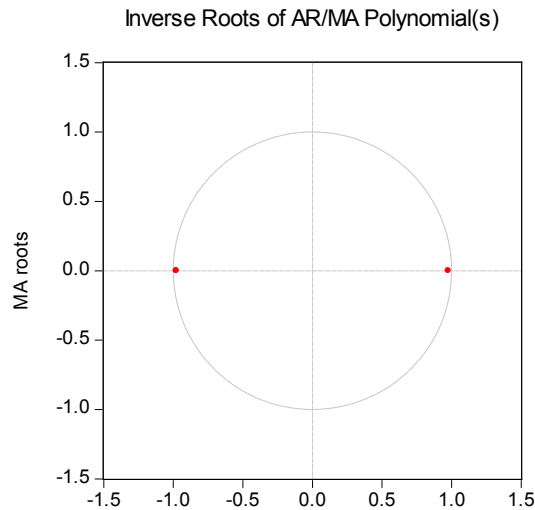
The results in table 3 indicate that both coefficients are statistically significant at 1% level of significance. The non-linear techniques used by Eviews involved an iterative process that is converged after 12 iterations. The roots are ± 0.98 , both inside the unit circle indicating stationarity and invertibility respectively. The chosen model as summarized in Table 3 is ARIMA(0,1,2) and is given by

$$D(\text{GDP_RATE}_t) = -0.953165 \varepsilon_{t-2} + e_t$$

t-stat.	(-27.85325)
prob.	[0.000]
s.e	{0.034221}

On the following diagram the inverse roots of AR and MA characteristic polynomials for the stability of ARIMA model are presented.

Figure 3: Inverse Roots of AR and MA



From diagram 3 we can see that the ARIMA model is stable since the corresponding inverse roots of the characteristic polynomials are in the unit circle.

4.4 Diagnostic Checking of the Model

Diagnostic checking of the model, help us to check if the estimated model is acceptable and statistical significant that means that the residuals are not auto correlated and follow normal distribution. For checking autocorrelation we use Q statistic of Ljung-Box (1978) and normality test using Jarque-Bera (JB) test (1980). The figures below represents the tests of the autocorrelation and normality of the residuals of the model ARIMA(0,1,2).

Figure 4: Histogram of the residuals of model ARIMA (0,1,2)

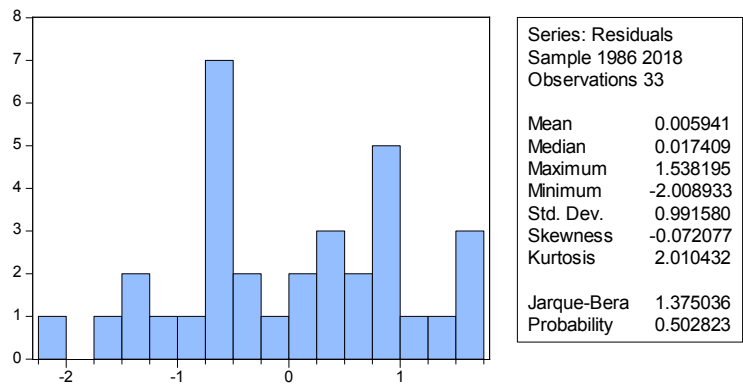
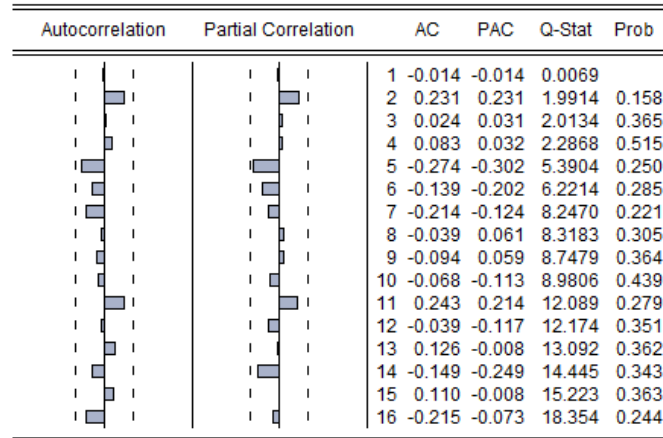


Figure 5: Correlogram Residuals of Model ARIMA (0,1,2)

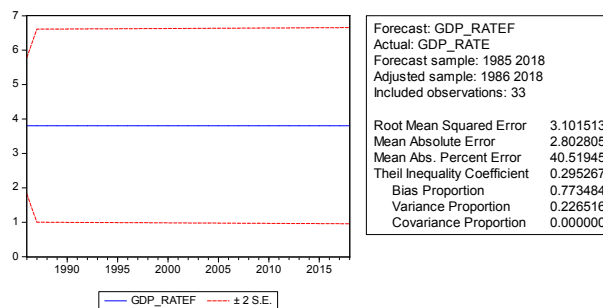


The results of figure 4 indicate that the residuals of ARIMA(0,1,2) model follow normal distribution. Moreover, the results of figure 5 indicate that the Q statistic of Ljung-Box for all the 16 lags has values greater than 0.05 thus the null hypothesis cannot be rejected i.e. there is no autocorrelation for the examined residuals of the series.

5. Forecasting

In figure 6 we represent the criteria for the evaluation of the forecasts of the model ARIMA(0,1,2)

Figure 6: Forecast Accuracy Test on the Model ARIMA (0,1,2)



The results in figure 6 indicate that the inequality coefficient of Theil has a high value $U = 0.295$ which means that our model does have a good forecasting ability. Table 6 below summarizes the forecasting results of the real GDP rate over the period 2019 to 2021.

Table 6: The Real GDP Rate Forecasts

Year	2019	2020	2021
The Real GDP Rate Forecasts	6,41%	6,57%	6,6%

6. Conclusion

In this paper, using Box - Jenkins technique, we are trying to forecast the real GDP rate in Vietnam for the next three years with an ARIMA model. After checking for the stationarity of the data series, we find the appropriate ARIMA (p, d, q) process. The corresponding correlogram helped in choosing the appropriate p and q for the data series. An ARIMA(0,1,2) model was created through the data used and estimating this model we found that the real GDP rate for the years 2019, 2020 and 2021 is forecast to be 6.41%, 6.57% and 6.6% respectively. Results of the study will be helpful for the policy makers to formulate effective policies for attracting foreign direct investment. Furthermore, the findings of the study will also help the managerial business executives for implementing the new project ideas or taking decisions concerned with the expansion of the existing business.

References

- [1] Box, G. E. P. & Jenkins G. M. "Time series analysis. Forecasting and control", Holden-Day, San Francisco, 1976.
- [2] Chaido, D., "Forecasting Real GDP Rate through Econometric Models: An Empirical Study from Greece", *Journal of International Business and Economics*, 3(1), 2015, pp. 13-19.
- [3] Dickey, D. A. & Fuller W. A., "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, 74(366), 1979, pp. 427-431.
- [4] The European System of Accounts ESA, 1995, <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/CA-15-96-001>.
- [5] Eurostat yearbook 1996: A statistical view Europe 1985-1995, <http://aei.pitt.edu/72805/>
- [6] Jarque. C. M., & Bera A. K., "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters* 6(3), 1980, pp. 255-259.
- [7] Jovanovic, B. & Petrovska M. "Forecasting Macedonian GDP: Evaluation of different models for short-term forecasting", Working Paper, National Bank of the Republic of Macedonia, 2010.
- [8] Ljung, G. M., & Box G. E. P. "On a measure of a lack of fit in time series models", *Biometrika*, 65(2), 1978, pp. 97-303.
- [9] Maity, B., & Chatterjee B. "Forecasting GDP growth rates of India: An empirical study", *International Journal of Economics and Management Sciences*, 1(9), 2012, pp. 52-58.
- [10] Phillips, P. C. B. & Perron P., "Testing for a unit root in time series regression", *Biometrika*, 75(2), 1988, pp. 335-346.
- [11] Shahini, L. & Haderi S. "Short term Albanian GDP forecast: One quarter to one year ahead. *European Scientific Journal*", 9(34), 2013, pp. 198-208.
- [12] Wei Ning, Bian Kuan-jiang & Yuan Zhi-fa, "Analysis and forecast of Shaanxi GDP based on

the ARIMA model”, Asian Agricultural Research, 2(1), 2010, pp. 34-41.

[13] Zakai, M., “A time series modeling on GDP of Pakistan”, Journal of Contemporary Issues in Business Research, 3(4), 2014, pp. 200-210.

[14] Zhang, H., “Modeling and forecasting regional GDP in Sweden using autoregressive models”, Working Paper, Hogskolan Dalarna University, Sweden, 2013.