# Performance Assessment of GDP Forecasting Models for Cambodia's Economy

Research Paper

In partial fulfillment of the degree

Master of Public Policy, International Program

Submitted by

Yuthyea Than

51-198213

Submitted to

Graduate School of Public Policy

The University of Tokyo

Academic Advisor

Associate Professor Taisuke Nakata

June 2021

Tokyo, Japan

**Table of Contents**

## List of Tables and Figures

### Tables

**Figures**

## Performance Assessment of GDP Forecasting Models for Cambodia's Economy


### Abstract

Gross domestic product is a key indicator gauging national economic performance as well as state of domestic macroeconomy. Being aware of its future paths is critical in planning budgeting, managing financial position and developing economic and social development goals and strategies. By utilizing annual time series data of real gross domestic product, real export and import of goods and services of Cambodia from 1993 to 2020, three simple models such as autoregressive model, autoregressive integrated moving average model and vector autoregressive model are applied to model and forecast real gross domestic product of Cambodia aiming to find the best performing forecasting model. The forecast is produced by following five steps consisting of stationary tests, model identification, model estimation, model diagnostics and model forecasting. The results show that ARIMA(1,1,2) model performs the best in forecasting economic activities of Cambodia with the relatively lowest forecast accuracy metrics. Going forward, with adverse impacts from Covid-19 in 2020, Cambodia's economy still encounters a negative development in 2021, then realizes positive consistent growth from 2022 to 2025.

## 1. Introduction

Gross domestic product (GDP) is defined by the IMF[1] as the monetary value of final goods and services produced within borders of a country in a given period of time either a quarter or a year. It consists of not only transactions of goods and services in the market but also some nonmarket productions, for instance, education and defense services provided by the government.

Typically, GDP is regarded as a key indicator gauging national economic performance as well as state of domestic macroeconomy as a whole. Fluctuations of GDP which is known as business cycles affect economic agents differently in terms of income, employment, price level and interest rate. Due to a great importance of such an indicator, being forward-looking of its possible future paths and variations is crucial for policy makers and central bankers whose objective is to keep the economy on the desired path by adjusting fiscal, monetary and exchange rate policies. The effects of these policies, however, are often lagged because it takes a considerable amount of time for such policies to influence spending behaviors of the population (Mankiw, 2016). As such, successful implementations of stabilizing policies require the ability to reliably and accurately forecast future economic conditions, particularly GDP. In addition, reliable and accurate forecast of GDP can serve as a practical basis for planning state and local budgeting, managing financial position and developing economic and social development goals and strategies (Stock, 2001).

There are two main methods that can be applied to do forecasting either using univariate time series or multivariate economic models. The former is a simple single variable-based model with less data requirement, while the latter is relatively sophisticated and based on economic theories as well as assumptions with large data requirement. In data scarce environment, considering a trade-off between a model complexity and precision, simple univariate or VAR model is more likely to outperform large multivariate models (Robertson & Tallman, 1999).

In this paper, given a short available time series data, three simple time series models are applied to forecast real GDP of Cambodia's economy: Autoregressive (AR) model, Autoregressive Integrated Moving Average (ARIMA) model and Vector Autoregressive (VAR) model. The purpose of the paper is to test which of the three autoregressive model performs the best in forecasting the real GDP.

The rest of the paper is organized into four sections. Section 2 outlines existing empirical literature on GDP forecasting. Section 3 highlights methodologies of AR, ARIMA and VAR model analysis and forecasting as well as forecast accuracy metrics. Section 4 presents empirical results. The last section summarizes the study and provides conclusions.

[1] The International Monetary Fund

## 2. Review of Literature

Zhang (2013) conducts a performance comparison between AR, ARIMA and VAR model to forecast regional real GDP per capita of 25 counties in Sweden. The author applies annual data from 1993 to 2009 into the analysis; around 70% of the original data which is between 1993 and 2004 is utilized to fit the model, and that of the last five years is used to evaluate forecast ability of each model. For each model, the author follows five steps including stationarity test, model identification, model estimation, model diagnostics and model forecasting. Then, to assess the forecast performace, percentage errors and mean absolute percentage errors are calculated and compared across each model (Diebold & Mariano, 1995). The empirical result shows that, due to a small sample size, the simpler AR(1) model performs the best in forecasting the real GDP per capita for a period of 5 years with 1.71% mean absolute percentage error, followed by ARIMA(0,1,3) and VAR(1) and (2).

Agrawal (2018) studies various specifications of ARIMA model in an attempt to find out the best model describing the evolution of India's GDP and then to develop forecasts. The author utilizes quarterly real GDP data between second quarter of 1996 and second quarter of 2017 from Reserve Bank of India and Central Statistics Office. In terms of methodology, the author follows similar procedures as Zhang (2013). Based on the empirical analysis, there are two ARIMA specifications being considered such as ARIMA(1,0,1) and (1,0,2). However, between these two specifications, it is inconclusive which specification represents the data better. Subsequently, the author proceeds to dynamic, structural and fixed time forecasting by applying AR(1) and MA(2) model and studying their residuals. As a result, the paper concludes there is no significant differences between AR(1) and MA(2) model and forecasts from both models converge in the long run.

Wei et al. (2010) applies time series model, particularly ARIMA model to forecast the GDP of the Shaanxi province in China. Data of GDP, obtained from 2008 Shaanxi Statistical Yearbook, from 1952 to 2007 is used to set up the ARIMA model, and that between 2002 and 2007 is used to evaluate forecast accuracy. The authors, Zhang (2013) and Agrawal (2018) share the similar steps to construct the model. Following these procedures, the authors successfully construct ARIMA(1,2,1) model with white noise residual sequence. In addition, the paper concludes that ARIMA(1,2,1) model is acceptable for forecasting purposes as the relative error is within 5% range.

Robertson and Tallman (1999) illustrates some detailed steps in designing and constructing a VAR forecasting model to produce real time forecast in an attempt to improve the application of VAR model used to do forecasting in business contexts and in policy institutions. The authors mentions that forecasting with VAR model is conducted to see dynamic correlation patterns between observed data, then the observed data can be used to forecast future values. In the paper, they employ VAR model to forecast real GDP growth, inflation and unemployment of the US economy. Monthly data from January 1959 to December 1985 is utilized to fit the model, while the remaining of data which is until december 1997 is kept for testing purposes. The authors also highlights technical difficulties needed to be dealt with to implement practical applications as well as approaches to overcome those issues. In addition, the authors provide discussions regarding ways to improve forecast accuracy.

Andersson (2007) studies the performance of AR, VAR and random walk models to forecast real GDP growth of Swedish by employing forward looking surveys as explantory variables. The surveys include confidence of consumers and businesses, which are based on perspectives on the current and future economic conditions. The author utilizes quarterly data of real GDP of Swedish economy from first quarter of 1993 to fourth quarter of 2006. To assess forecast performace, mean errors, mean absolute errors and root mean square errors are calculated and compared. According to empirical analysis, the author concludes that VAR model exhibits superior performance compared to AR and random walk models. In addition, VAR(1) performs best for one quarter ahead forecast; VAR(2) performs best for four quarters ahead forecast; VAR(3) performs best for eight and twelve quarters ahead forecast.

## 3. Methodology

This section divides into three main parts. The first part discusses statistical background, stationarity tests, model identification, model estimation, model diagnostics and model forecasting of AR and ARIMA model together given their similar characteristics. The second part details those of VAR model with an addition of Granger causality test. The last part provides details to evaluate forecast accuracy.

### 3.1. AR and ARIMA Model

Application of univariate time series model forecasting is used widely in an economic and a financial field. Particularly, ARIMA model, popularized by Box and Jenkins (1976), has been increasingly utilizing to do forecasting due to its practical use for nonstationary time series data.

#### 3.1.1. Statistical Background

AR model is based on a concept that a current value of a time series can be written as a linear combination of its past values with a random error. An AR model of lag p, abbreviated as AR(p), can be expressed as follows:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t \quad ; \quad t = 1,2,\dots,T$$

Where $c$ is a constant term; $\phi_1, \phi_2, \dots, \phi_p$ are model parameters to be estimated and $\varepsilon_t$ is the error term which is a white noise process assumed to be a sequence of independently and identically distributed random variables with $E(\varepsilon_t) = 0\ and\ Var(\varepsilon_t) = \sigma^2; \varepsilon_t \sim iid\ N(0, \sigma^2)$.

ARIMA model is a combination of AR and MA model[2], called ARMA model, with an integrated process[3] denoted by I. ARMA model overcomes a drawback of AR model which ignores correlated unobserved noise structure in a time series which is sometimes informative in explaining the behavior of the series. An ARMA model of lag p and q, abbreviated as ARMA(p,q), can be expressed as follows:

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + v_1 \varepsilon_{t-1} + \cdots + v_q \varepsilon_{t-q}$$

$$= c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} v_j \varepsilon_{t-j} + \varepsilon_t \qquad ; \qquad t = 1,2,\dots,T$$

Where $c$ is a constant term; $\phi_1, \phi_2, \dots, \phi_p$ and $v_1, v_2, \dots, v_q$ are AR and MA model parameters to be estimated, respectively.

**3.1.2. Stationarity Tests**

Time series forecast exploits the past time series to forecast the future development. This requires the future to be similar to the past in a sense that correlations or distributions of the series in the future follow those of the past. In other words, a time series needs to have a stable mean and variance over time so that past relationships might be able to guide to the future. Technically, this is called stationarity.

The issue of stationarity can be dealt with by, first, using a line graph to get an overall idea on a time series. If the series exhibits an obvious trend, for instance, an upward or a downward pattern, the series is nonstationary. Alternatively, unit root tests can also be applied directly to determine the stationarity of the series. If the series is nonstationary, taking first or higher order differencing might make the series become stationary. Typically, differencing the series repeatedly can reduce the non-stationarity nature. However, it does not imply that the higher order differencing always better. According to Harvey (1989), differencing procedure is performed to extract information and to process data with the expense of information loss each time the procedure is used.

With a concern of an over-differencing issue, two different unit root tests are applied which are Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. ADF test, developed by Dickey and Fuller (1979), tests the null hypothesis that there is a presence of a unit root in a time series, while the alternative hypothesis states that the series is stationary. Computed ADF statistics is usually a negative number and the more negative or less the statistics is, the higher chance of rejecting the null hypothesis

---

[2] Moving Average (MA) model expresses a current value of a time series as a function of its past errors with a random error. An MA model of lag q, abbreviated as MA(q), can be expressed as the following:

$$Y_t = \varepsilon_t + v_1 \varepsilon_{t-1} + v_2 \varepsilon_{t-2} + \cdots + v_q \varepsilon_{t-q} = \varepsilon_t + \sum_{j=1}^{q} v_j \varepsilon_{t-j} \quad ; \quad t = 1,2,\dots,T$$

[3] Integrated process is the use of differencing to transform nonstationary time series to stationary process. In ARIMA(p,d,q) model, d refers to the number of times data got differencing to become stationary.

at a certain level of confidence. KPSS test, developed by Kwiatkowski et al. (1992), can be said to be complimentary to ADF test. Contrary to ADF test, KPSS test tests the null hypothesis that a time series is stationary, while the alternative hypothesis states that there is a presence of a unit root in the series. Computed KPSS statistics is usually a positive number and the less positive or less the statistics is, the higher chance of failing to reject the null hypothesis at a certain level of confidence. In this paper, a time series is concluded as a stationary process once either ADF or KPSS test indicates so at a 5% significant level.

**3.1.3. Model Identification**

After conducting necessary data transformations leading to a stationary process, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to tentatively determine a lag length of AR and ARMA model. Then, information criterion is applied to verify the selection.

ACF measures correlations between $Y_t$ and $Y_{t-q}$, while PACF, controlling for intermediate lags in between, measures partial correlations between $Y_t$ and $Y_{t-p}$. Ghysels and Marcellino (2018) presents a method to determine a lag length of AR, MA or ARMA model using ACF and PACF plots. If ACF plot displays an exponentially declining trend and PACF plot presents some peaks, this suggests a pure AR(p) model, with p is the number of peaks (statistically different from zero) in the PACF plot. If PACF plot displays an exponentially declining trend and ACF plot presents some peaks, this suggests a pure MA(q) model, with q is the number of peaks (statistically different from zero) in the ACF plot. If both ACF and PACF plots exhibit an exponentially declining trend, this suggests the mix model, specifically ARMA(p,q).

Then, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are employed to verify the tentatively selected lag length of AR and ARIMA model. AIC, developed by Akaike (1974), is a statistical technique to select a model by comparing the fit of a different models to observed time series. A relatively better model is the one with a relatively lower AIC value. Similarly, BIC, developed by Schwarz (1978), is also another criterion for model selection by measuring a trade-off between a model fit and model complexity. A relatively better model is the one with a relatively lower BIC value. AIC and BIC can be computed using the following formulas:

$$AIC(p) = -2ln(L) + 2(p + q)$$

$$BIC(p) = -2ln(L) + ln(T)(p + q)$$

Where $L$ is a likelihood of the series with a certain model; $p$ and $q$ is a lag length of AR and MA model, respectively and $T$ is a number of observations in the stationary process.

### 3.1.4. Model Estimation

With the determined lag length p for AR model as well as the determined integrated process d, lag length p and q for ARIMA model, each model is to be estimated using maximum likelihood estimation method in Python 3.

### 3.1.5. Model Diagnostics

After the chosen model is estimated, diagnostics test is conducted to assess whether the fitted model is valid or not[4]. In order to evaluate a validity of the fitted model, a diagnostics test called Ljung-Box Q test is applied on residuals of the fitted model. Ljung-Box Q test, developed by Ljung and Box (1978), tests the null hypothesis that residuals autocorrelation is jointly zero, while the alternative hypothesis states that residuals are serially correlated. Ljung-Box Q test is computed as follows:

$$Q_{LB} = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k}$$

Where $T$ is a number of observations in a stationary process; $\hat{\rho}_k$ is autocorrelations at lag $k$ and $h$ is a number of lags being tested. Under the null hypothesis, $Q_{LB}$ statistics follows Chi-squared distribution ($\chi^2$). If $Q_{LB}$ statistics is higher than $\chi^2$critical value, the null hypothesis is rejected, and if $Q_{LB}$ statistics is less than or equal $\chi^2$critical value, the null hypothesis cannot be rejected. If the null hypothesis is rejected, this means the fitted model is invalid, which requires model reidentification by either decreasing or increasing lags length until the null hypothesis of Ljung-Box Q test cannot be rejected at a 5% significant level.

### 3.1.6. Model Forecasting

Following confirming that the fitted model possesses white noise residuals, the model can be used to do forecasting. Suggested by Stock and Watson (2020), iterated forecast which is one-period or one-step ahead forecast is applied. The central idea is that, by exploiting observed time series through period *T*, the model can produce one period forecast *T+1*. Then, by treating the forecasted value at period *T+1* as observed data combining with observed time series until period *T*, the model produces forecast for period *T+2*. This process keeps repeating or iterating until forecast is made for the desired forecast periods.

## 3.2. VAR Model

VAR model, proposed by Sims (1980) in his seminal work, has been extensively used to conduct multivariate time series analysis and forecasting. By extending univariate autoregressive model allowing more than one variable in the equation, VAR model is able to study a joint dynamic behavior of multiple variables and produces forecast.

---

[4] A valid fitted model is a model whose residuals are white noise with zero mean and no serial correlations.

### 3.2.1. Statistical Background

VAR model consists of a system of $k$ time series regressions, each of which has an intercept and $p$ lags of each of the $k$ time series variables. In other words, VAR model treats all variables symmetrically in a structural way. Each variable in the model possesses its own equation describing its relationship based on its own lags and lags of other variables. A VAR model with lag p, abbreviated as VAR(p), can be expressed as follows:

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t = c + \sum_{i=1}^{p} \Pi_i Y_{t-i} + \varepsilon_t \quad ; \quad t = 1,2,\dots,T$$

Where $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ is a $(n \times 1)$ vector of time series variables; $c$ is a $(n \times 1)$ vector of constants; $\Pi_i$ is a $(n \times n)$ vector of coefficient matrices and $\varepsilon_i$ is a $(n \times 1)$ vector of an unobservable zero mean and white noise process.

In this paper, two key factors are taken into consideration regarding a choice of variables to use in VAR model. First, the chosen variables have to reflect the ongoing structure of the economy as a whole. Second, a time series of the chosen variables must be available and in good quality.

### 3.2.2. Stationarity Tests

Initially, a line graph is used to investigate a time series of each variable. In case of an obvious upward or a downward trend, first order differencing is taken to eliminate nonstationary nature of the series. Then, unit root tests such as ADF and KPSS tests are applied on each transformed series to verify again. Among all variables in the model, if a time series is nonstationary, the differencing is applied on all other series. This process is repeated until each series is stationary.

### 3.2.3. Model Identification

Following conducting necessary data transformations leading to the stationary process for every series, AIC and BIC are used to select a lag length of VAR model. Formulas used to compute AIC and BIC are as follows:

$$AIC(p) = -2ln(L) + 2p$$

$$BIC(p) = -2ln(L) + pln(T)$$

Where $L$ is a likelihood of the series with a certain model; $p$ is a number of parameters in the model and $T$ is a number of observations in the stationary process.

#### 3.2.4. Model Estimation

With the selected lag length using AIC and BIC, VAR model is to be estimated using ordinary least square estimation method in Python 3.

#### 3.2.5. Granger Causality Test

Unlike AR and ARIMA model, VAR model exploits each of a time series to do forecasting. Hence, before proceeding to next steps, it is necessary to make sure whether the chosen explanatory variables are useful for forecasting by applying Granger causality test on the estimated coefficients. Granger causality test, developed by Granger (1969), test the null hypothesis that past values of explanatory variables does not jointly Granger cause an explained variable, while the alternative hypothesis is contrary to the null hypothesis. If the null hypothesis cannot be rejected at 5% significant level, other explantory variables need to be considered and replaced.

#### 3.2.6. Model Diagnostics

Given the fact that there is a system of $k$ equations in VAR model, there are $k$ regressions to be estimated with $k$ residuals to be evaluated. Similar to Ljung-Box Q test, Portmanteau test is applied to assess joint $k$ residuals from the fitted model. Portmanteau test tests the null hypothesis that autocorrelation of residuals from $k$ regressions up to certain lag lengths is jointly zero, while the alternative hypothesis states that residuals from $k$ regressions are serially correlated. If the null hypothesis is rejected, this means the fitted model is invalid, which requires model reidentification until the null hypothesis of Portmanteau test cannot be rejected at a 5% significant level.

#### 3.2.7. Model Forecasting

Similar to AR and ARIMA model, iterated forecast is used to obtain forecasted values for VAR model.

### 3.3. Forecast Accuracy Comparison

Following obtaining forecasted values from AR, ARIMA and VAR model, the forecasted values are compared to the actual ones and errors are computed. To objectively assess performance of the three models, three different metrics are computed and compared against one another. These metrics namely mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) can be computed as follows:

$$MAE = mean(|e_t|)$$

$$RMSE = \sqrt{mean(e_t)^2}$$

$$MAPE = mean\left(\left|\frac{e_t}{y_t}\right|\right)$$

The best performing model is the model with the lowest value of all metrics.

## 4. Empirical Results

This section divides into five main parts. The first part highlights sources and treatment of data. The second, third and fourth part discusses results of AR, ARIMA and VAR model. The last part compares forecast accuracy of each model and produces five-year ahead forecast with the most accurate model.

### 4.1. Data

In this paper, every time series is obtained from national accounts of National Institute of Statistics of Cambodia. For AR and ARIMA model, real GDP (RGDP) data is used. In addition, real export (REXP) and real import (RIMP) of goods and services data are utilized for VAR model. REXP and RIMP variables are chosen as explanatory variables based on criteria specified in 3.2.1[5]. These three annual time series are available from 1993 to 2020. The real value of the series is calculated using price level of year 2000 as the base year.

Data between 1993 and 2015, which is about 80 percent of the sample size, is used to train the model and the rest from 2016 to 2020, which is around 20 percent, is utilized to evaluate forecast accuracy of each model. **Table 1** shows summary statistics of each time series. Each series is measured in billion riels which is the local currency of Cambodia.

**Table 1: Summary Statistics of RGDP, REXP and RIMP**

| | Min | Mean | Max | Std |
|---|---|---|---|---|
| RGDP | 8,521.5 | 27276.5 | 56,578.1 | 15320.5 |
| REXP | 1,322.7 | 24554.0 | 69240.5 | 21263.9 |
| RIMP | 2,722.4 | 28405.2 | 78543.9 | 23836.3 |

Based on the summary table, each time series is of high and positive value, so natural logarithm transformation of every series is taken in order to reduce heteroskedasticity issues.

### 4.2. AR model

#### 4.2.1. Stationarity Tests

The line graph of natural logarithm of RGDP shown in **Figure 1** clearly exhibits an upward trend which is nonstationary. In this sense, first order differencing is needed to transform the series.

The first order differencing of natural logarithm of RGDP, $\Delta ln(RGDP)$[6], is taken. The line graph of $\Delta ln(RGDP)$ and results of unit root tests of $\Delta ln(RGDP)$ are presented in **Figure 2** and **Table 2**, respectively.

---

[5] Cambodia is a small open economy with the size of total international trade both goods and services in real term around 250 percent of RGDP in 2020.

[6] $\Delta ln(RGDP) = ln(RGDP_t) - ln(RGDP_{t-1})$, which is a growth rate approximation of RGDP.

From the result of ADF test, test statistics (-2.909) is less than critical value at 10% significant level (-2.714) but higher than that at 5% significant level (-3.154). This means that the null hypothesis of unit root time series cannot be rejected at 5% with ADF test. However, based on the result of KPSS test, given that test statistics (0.108) is even less than critical values at 10% significant level (0.347), the null hypothesis of a stationary series cannot be rejected even at 10% level with KPSS test. Hence, $\Delta ln(RGDP)$ is stationary at 5% significant level.

**Figure 1: Line Graph of Natural Logarithm of RGDP**

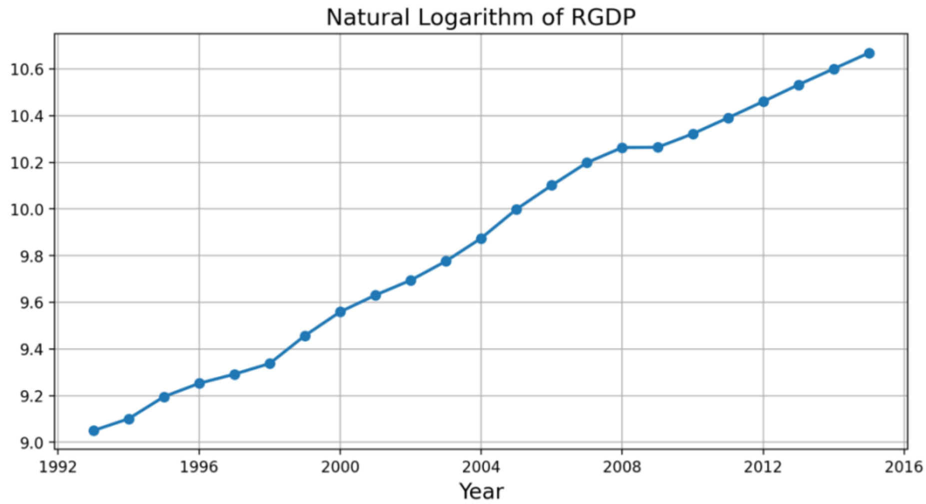



**Figure 2: Line Graph of $\Delta ln(RGDP)$**

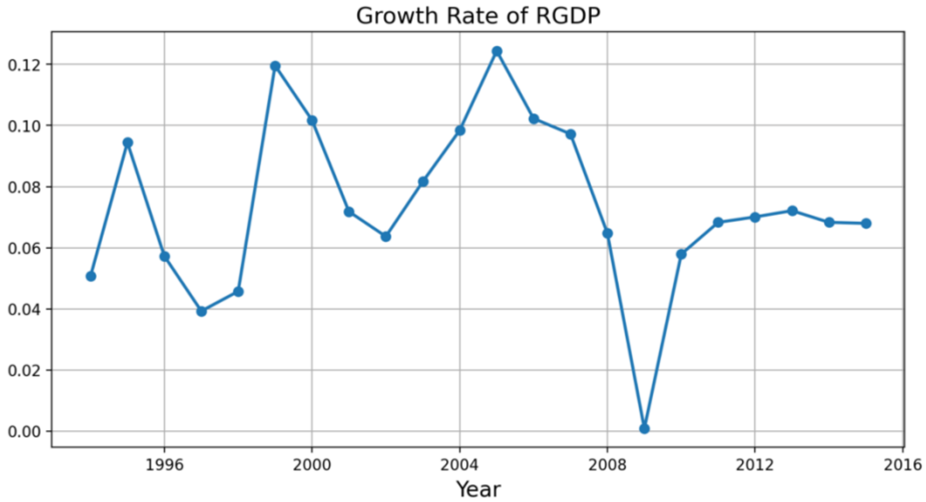


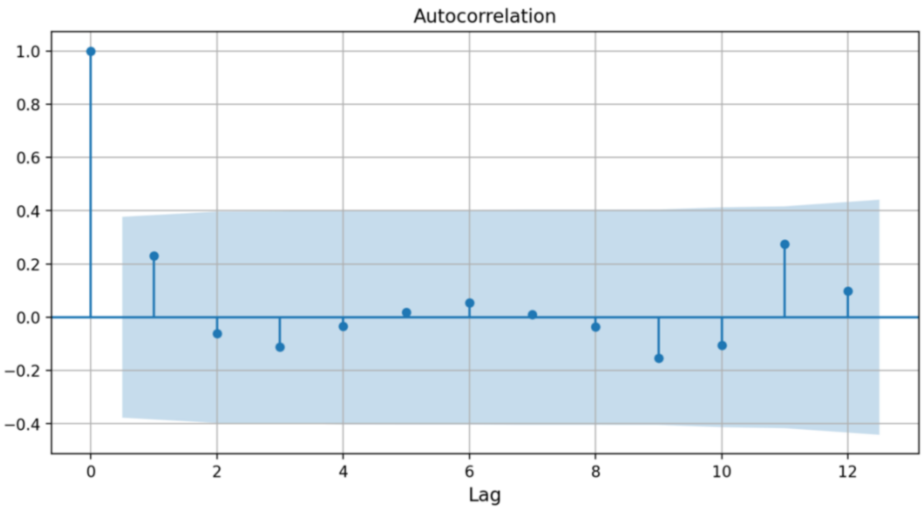

**Table 2: ADF and KPSS Unit Root Tests of $\Delta ln(RGDP)$**

| | | ADF | KPSS |
|---|---|---|---|
| Test Statistics | | -2.909 | 0.108 |
| Critical Value | 1% | -4.137 | 0.739 |
| | 5% | -3.154 | 0.463 |
| | 10% | -2.714 | 0.347 |

### 4.2.2. Model Identification

Following Ghysels and Marcellino (2018), ACF and PACF plot are used to determine lag lengths of AR, MA and ARMA model tentatively. Based on **Figure 3**, there are two noticeable peaks in both ACF and PACF plot at first and eleventh lag. However, due to a short series data, only the first peak is considered. From the first peak onward, coefficients start going to zero. This process suggests either AR(1), MA(1), or ARMA(1,1) model. In this section, only AR model is examined in detail. With the tentative one lag at hand, AIC and BIC up to 5 lags are computed and compared against one another to select the most preferred AR model. According to **Table 3**, AIC value (-85.295) and BIC value (-83.113) are the lowest with one lag. This means that, among all AR models, AR(1) is the most preferred model.

**Figure 3: Autocorrelation and Partial Autocorrelation Plot of $\Delta ln(RGDP)$**



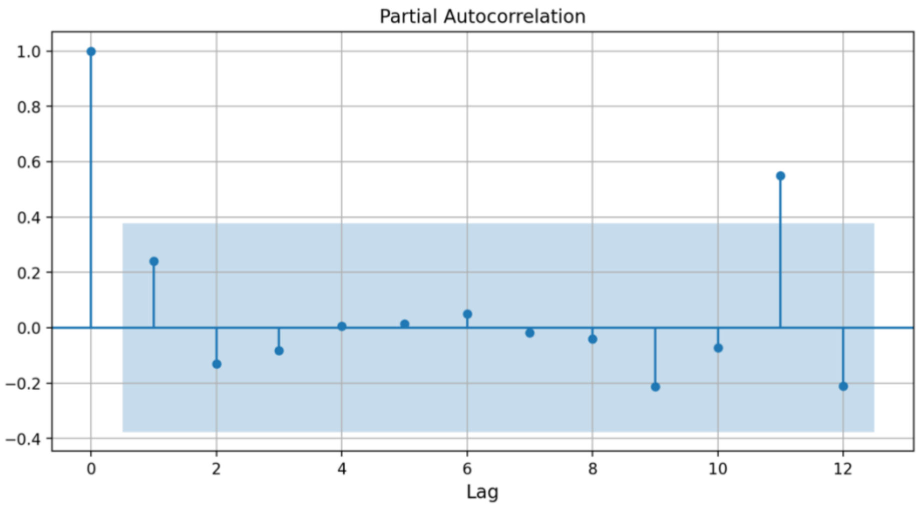



**Table 3: Information Criteria of AR Model with Different Lags**

| | | AIC | BIC |
|---|---|---|---|
| Lag | **1** | **-85.295** | **-83.113** |
| | 2 | -83.329 | -80.056 |
| | 3 | -83.431 | -79.067 |
| | 4 | -83.338 | -77.883 |
| | 5 | -81.896 | -75.349 |

### 4.2.3. Model Estimation

**Table 4** presents the estimates of the preferred AR(1) model. The coefficient of AR(1) model is statistically significant at any meaningful levels.

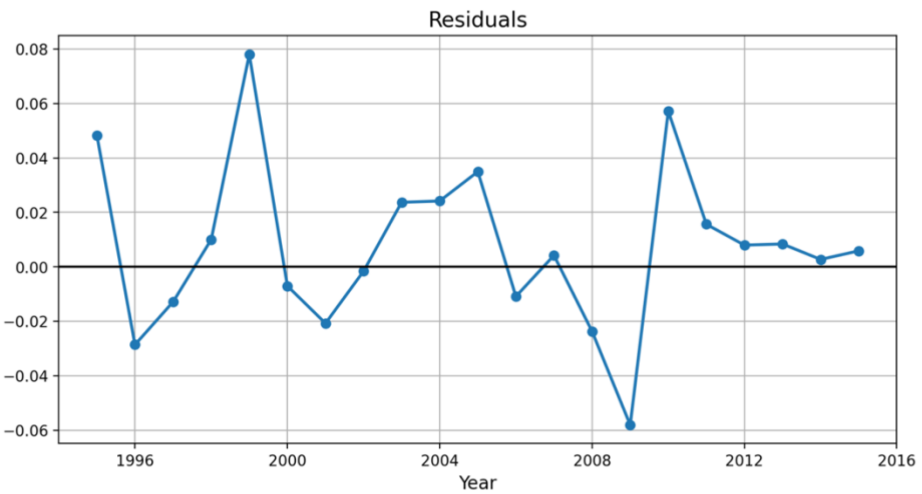
**Table 4: Regression Results of AR(1) Model**

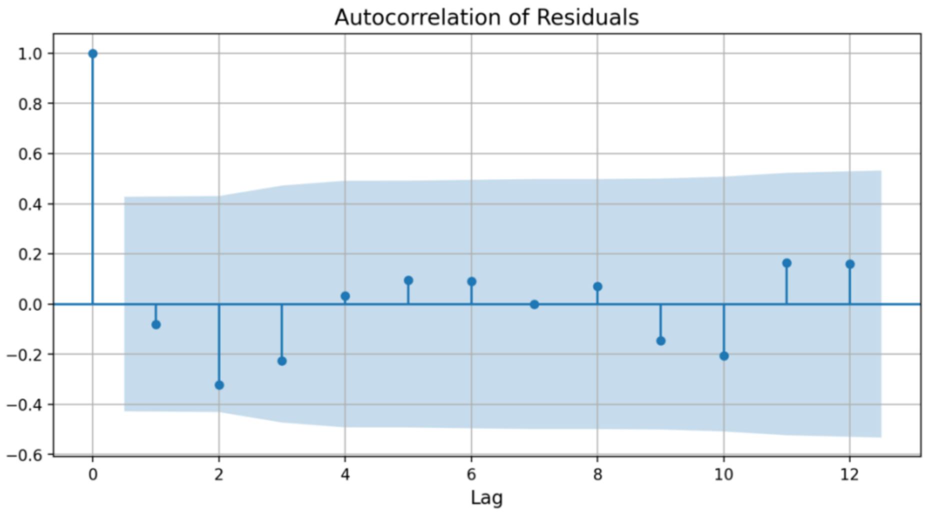
| Number of observations: 22 | | | | | | AIC = -85.295<br>BIC = -83.113 |
|---|---|---|---|---|---|---|
| | Coefficients | Standard error | t-statistics | p-value | [0.025 | 0.975] |
| AR.L1 | 0.9093 | 0.108 | 8.415 | 0.000 | 0.698 | 1.121 |
| $Sigma^2$ | 0.0009 | 0.000 | 3.629 | 0.000 | 0.000 | 0.001 |

### 4.2.4. Model Diagnostics

As shown in **Figure 4**, residuals of the fitted AR(1) model are around zero and autocorrelation of residuals are not significantly different from zero given that they are within 95% confidence interval.

**Figure 4: Residuals and Autocorrelation of Residuals Plot of AR(1) Model**






In addition, results of Ljung-Box Q test applied on residuals of the fitted AR(1) model in **Table 5** show that the p-value is always higher than 10% significant level for any lags up to twelve lags. In this sense, the null hypothesis of jointly zero residuals autocorrelation or white noise residuals cannot be rejected at 5% level, which means AR(1) model is a valid model for forecasting.

**Table 5: Ljung-Box Q Test Results of AR(1) Model**

| | | Ljung-Box Q Statistics | p-value |
|---|---|---|---|
| Lag | 1 | 0.157 | 0.691 |
| | 2 | 2.780 | 0.249 |
| | 3 | 4.142 | 0.246 |
| | 4 | 4.172 | 0.383 |
| | 5 | 4.444 | 0.487 |
| | 6 | 4.705 | 0.582 |
| | 7 | 4.705 | 0.695 |
| | 8 | 4.889 | 0.769 |
| | 9 | 5.736 | 0.765 |
| | 10 | 7.619 | 0.665 |
| | 11 | 8.910 | 0.630 |
| | 12 | 10.267 | 0.592 |

### 4.2.5. Model Forecasting

Applying the fitted AR(1) model with iterated forecast approach, forecasted values of RGDP can be obtained and presented in **Figure 5** and **Table 6**. Regarding in-sample forecast, AR(1) model fails to forecast indirect effects of Global Financial Crisis on Cambodia's economy in 2009. In terms of out-sample forecast, compared to actual real economic activities, AR(1) model forecasts relatively lower activities and the gap keeps widening as forecast horizons increase except in 2020 due to Covid-19 shock.

**Figure 5: Graphical Comparison of AR(1) Model Forecasting to Actual RGDP**

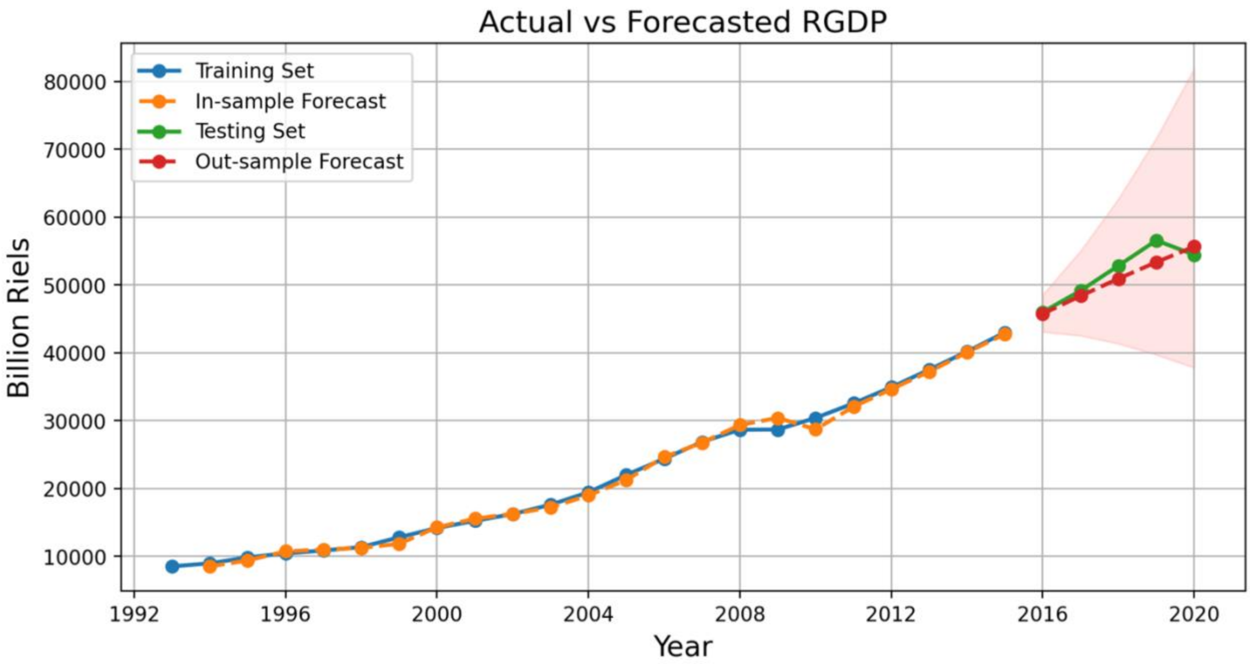



**Table 6: Forecasted Results of AR(1) Model**

| Year | Actual RGDP | Forecasted RGDP | Forecasted Growth Rate | Error |
|---|---|---|---|---|
| 2016 | 45,999.748 | 45,752.492 | 6.38% | -0.54% |
| 2017 | 49,176.889 | 48,398.579 | 5.78% | -1.58% |
| 2018 | 52,849.994 | 50,937.362 | 5.25% | -3.62% |
| 2019 | 56,578.089 | 53,361.375 | 4.76% | -5.69% |
| 2020 | 54,454.336 | 55,665.595 | 4.32% | 2.22% |

### 4.3. ARIMA model

#### 4.3.1. Stationarity Tests

As discussed in **4.2.1,** $\Delta ln(RGDP)$ is a stationary process. Hence, the order of integration is one for ARIMA model.

#### 4.3.2. Model Identification

Again, ACF and PACF plots suggest either MA(1) or ARMA(1,1) model. However, according to **Table 7**, AIC value (-86.380) and BIC value (-82.016) are the lowest with one lag for AR term and two lags for MA term. Hence, among all ARIMA models, ARIMA(1,1,2) is the most preferred model.

**Table 7: Information Criteria of ARIMA Model with Different Specifications**

| ARIMA Model | AIC | BIC |
|---|---|---|
| (0,1,1) | -65.901 | -63.719 |
| (1,1,1) | -83.717 | -80.444 |
| (2,1,1) | -81.347 | -76.983 |
| (0,1,2) | -71.536 | -68.263 |
| **(1,1,2)** | **-86.380** | **-82.016** |
| (2,1,2) | -83.512 | -78.056 |
| (3,1,2) | -82.433 | -75.887 |
| (0,1,3) | -75.001 | -70.637 |
| (1,1,3) | -84.569 | -79.114 |
| (2,1,3) | -82.911 | -76.365 |
| (3,1,3) | -81.008 | -73.371 |
| (4,1,3) | -78.676 | -69.948 |

### 4.3.3. Model Estimation

**Table 8** presents the estimates of the preferred ARIMA(1,1,2) model. The coefficient of AR(1) term is statistically significant at any meaningful levels, while that of MA(1) and MA(2) term are not statistically significant different from zero.

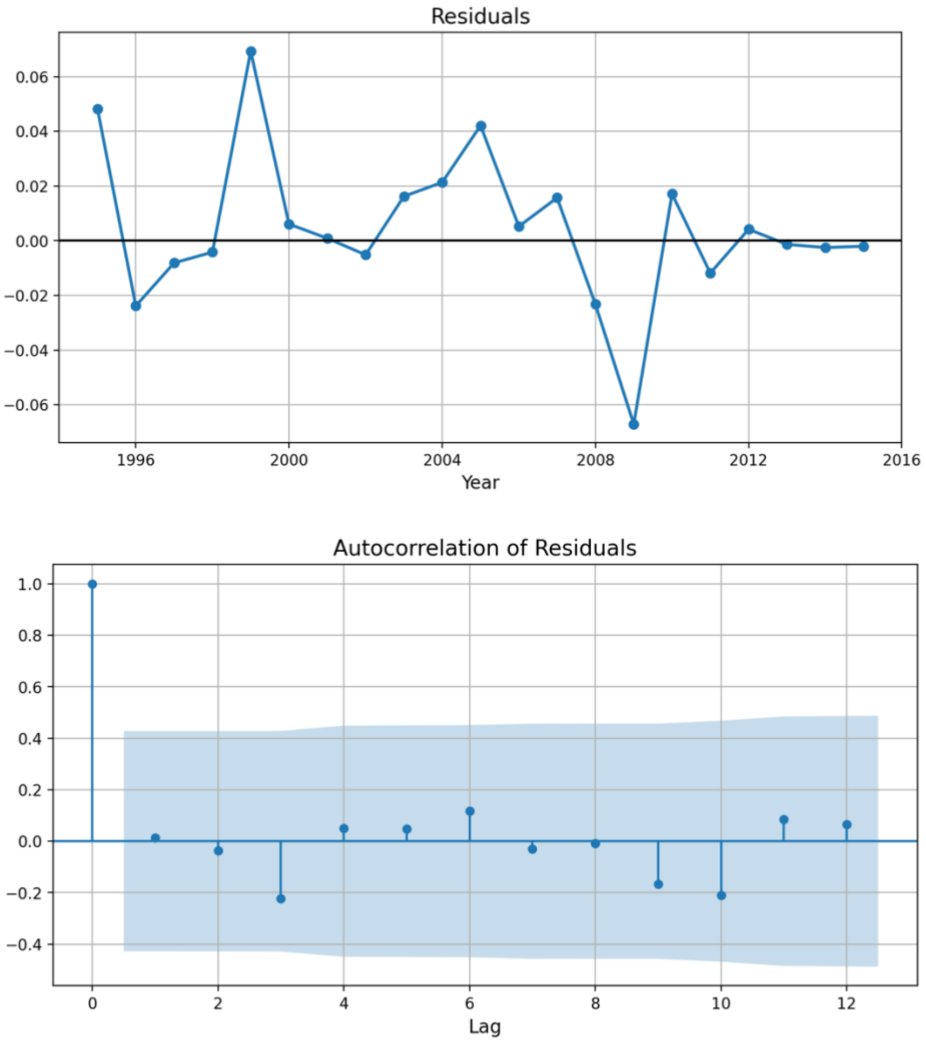
**Table 8: Regression Results of ARIMA(1,1,2) Model**

| Number of observations: 22 | | | | | | AIC = -86.380<br>BIC = -82.016 |
|---|---|---|---|---|---|---|
| | Coefficients | Standard error | t-statistics | p-value | [0.025 | 0.975] |
| AR.L1 | 0.9994 | 0.010 | 98.536 | 0.000 | 0.980 | 1.019 |
| MA.L1 | -0.4590 | 0.528 | -0.870 | 0.385 | -1.494 | 0.576 |
| MA.L2 | -0.4532 | 0.371 | -1.221 | 0.222 | -1.181 | 0.274 |
| Sigma$^2$ | 0.0007 | 0.000 | 1.823 | 0.068 | -0.00005 | 0.001 |

### 4.3.4. Model Diagnostics

Based on **Figure 6**, residuals of the fitted ARIMA(1,1,2) model are around zero and autocorrelation of residuals are not significantly different from zero given that they are within 95% confidence interval.

**Figure 6: Residuals and Autocorrelation of Residuals Plots of ARIMA(1,1,2) Model**




In addition, results of Ljung-Box Q test applied on residuals of the fitted ARIMA(1,1,2) model in **Table 9** show that the p-value is always higher than 10% significant level for any lags up to twelve lags. In this sense, the null hypothesis of jointly zero residuals autocorrelation or white noise residuals cannot be rejected at 5% level, which means ARIMA(1,1,2) model is a valid model for forecasting.

**Table 9: Ljung-Box Q Test Results of ARIMA(1,1,2) Model**

| | | Ljung-Box Q Statistics | p-value |
|---|---|---|---|
| Lag | 1 | 0.004 | 0.946 |
| | 2 | 0.040 | 0.980 |
| | 3 | 1.362 | 0.714 |
| | 4 | 1.436 | 0.837 |
| | 5 | 1.503 | 0.912 |
| | 6 | 1.950 | 0.924 |
| | 7 | 1.983 | 0.960 |

| | | |
|---|---|---|
| 8 | 1.985 | 0.981 |
| 9 | 3.104 | 0.959 |
| 10 | 5.020 | 0.889 |
| 11 | 5.377 | 0.911 |
| 12 | 5.608 | 0.934 |

#### 4.3.5. Model Forecasting

Applying the fitted ARIMA(1,1,2) model with iterated forecast approach, forecasted values of RGDP are obtained and presented in **Figure 7** and **Table 10**. Regarding in-sample forecast, ARIMA(1,1,2) model also fails to forecast indirect effects of Global Financial Crisis on Cambodia's economy in 2009. In terms of out-sample forecast, compared to actual real economic activities, ARIMA(1,1,2) model forecasts marginally higher activities until 2019 with less than 1% error. Due to Covid-19 shock, the error increases significantly in 2020.

**Figure 7: Graphical Comparison of ARIMA(1,1,2) Model Forecasting to Actual RGDP**

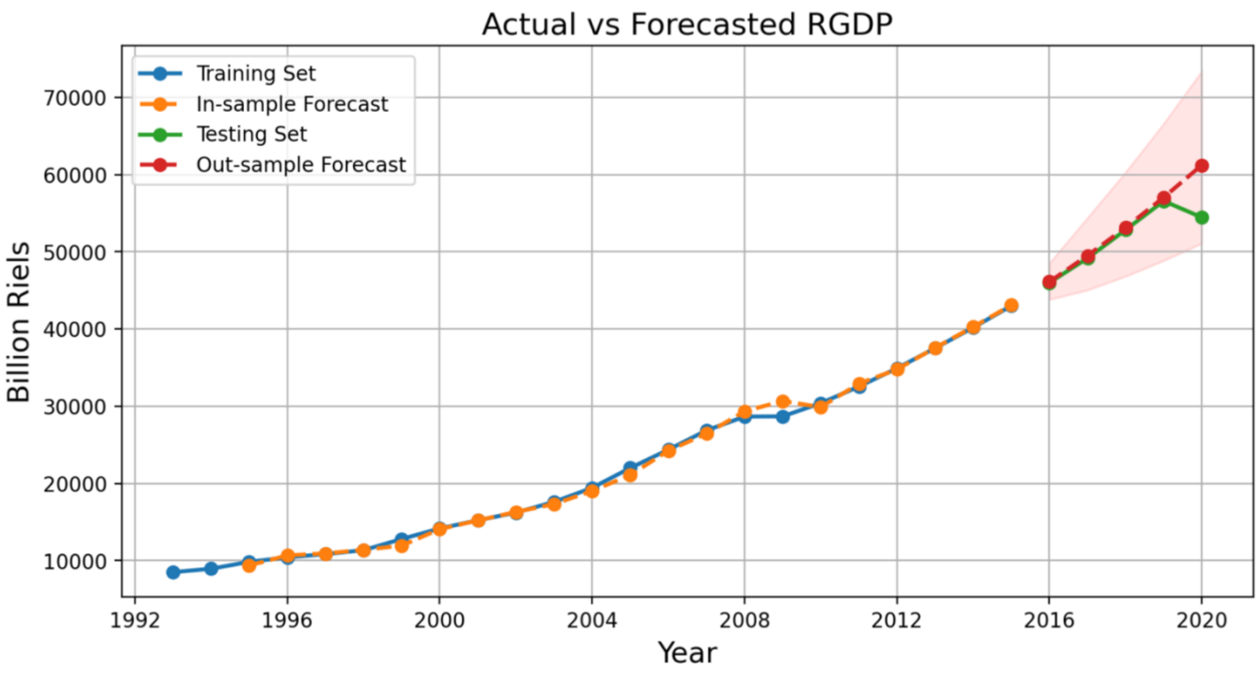



**Table 10: Forecasted Results of ARIMA(1,1,2) Model**

| Year | Actual RGDP | Forecasted RGDP | Forecasted Growth Rate | Error |
|---|---|---|---|---|
| 2016 | 45,999.748 | 46,128.242 | 7.25% | 0.28% |
| 2017 | 49,176.889 | 49,516.826 | 7.35% | 0.69% |
| 2018 | 52,849.994 | 53,152.219 | 7.34% | 0.57% |
| 2019 | 56,578.089 | 57,052.244 | 7.34% | 0.84% |
| 2020 | 54,454.336 | 61,235.996 | 7.33% | 12.45% |

### 4.4. VAR model

#### 4.4.1. Stationarity Tests

The line graph of natural logarithm of REXP and RIMP shown in **Figure 8** clearly exhibits an upward trend which is nonstationary. A first order differencing of natural logarithm of REXP and RIMP, $\Delta ln(REXP)$ and $\Delta ln(RIMP)$, are taken. The line graph of $\Delta ln(REXP)$ and $\Delta ln(RIMP)$ and results of unit root tests are presented in **Figure 9**, **Table 11** and **Table 12**, respectively.

From **Table 11**, the ADF test statistics of $\Delta ln(RExp)$ (-6.128) is less than critical value even at 1% significant level (-3.788). This means that the null hypothesis of unit root time series can be rejected at 1% level. From **Table 12**, the ADF test statistics of $\Delta ln(REIMP)$ (-2.245) is more than critical value at 10% significant level (-2.714). This means that the null hypothesis of unit root time series cannot be rejected at 10% level with ADF test. However, the KPSS test statistics of $\Delta ln(RIMP)$ (0.138) are less than critical values even at 10% significant level (0.347), which means the null hypothesis of a stationary time series cannot be rejected at 10% level with KPSS test. Hence, $\Delta ln(REXP)$ and $\Delta ln(RIMP)$ are stationary at a 5% significant level.

**Figure 8: Line Graph of Natural Logarithm of REXP and RIMP**

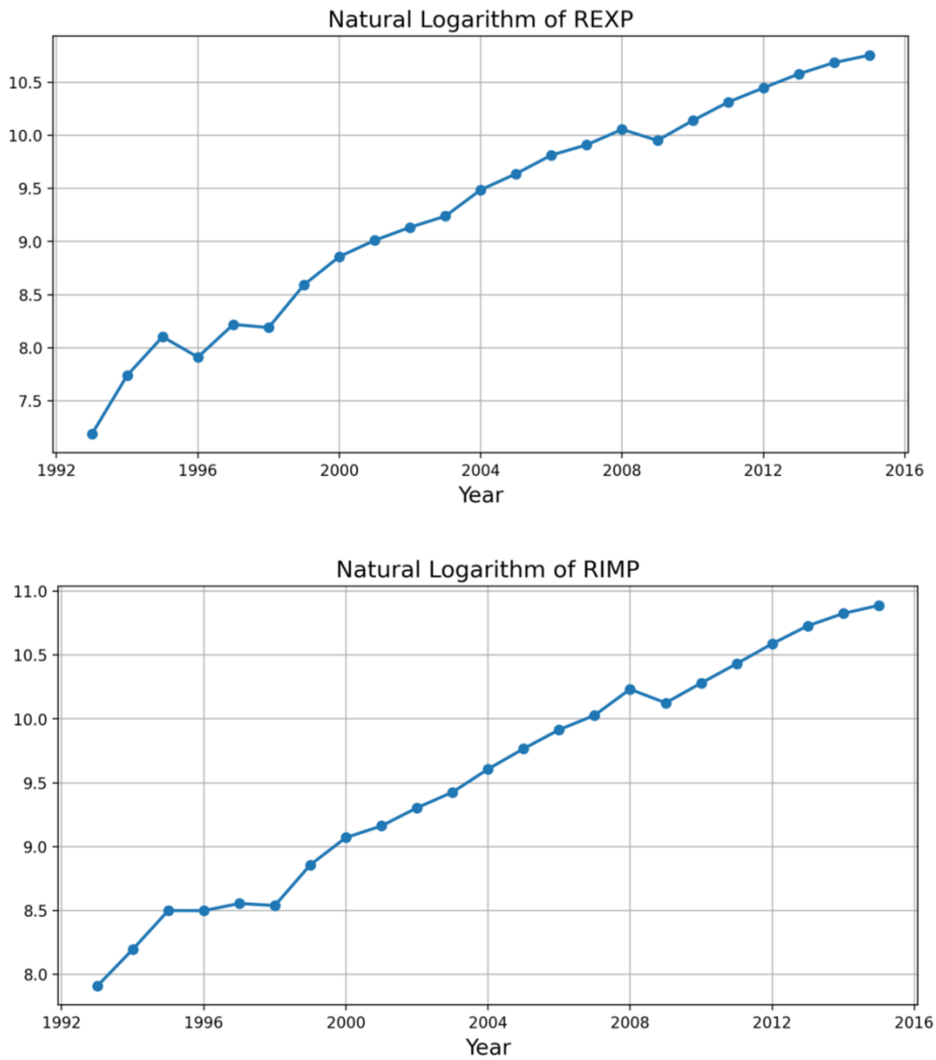
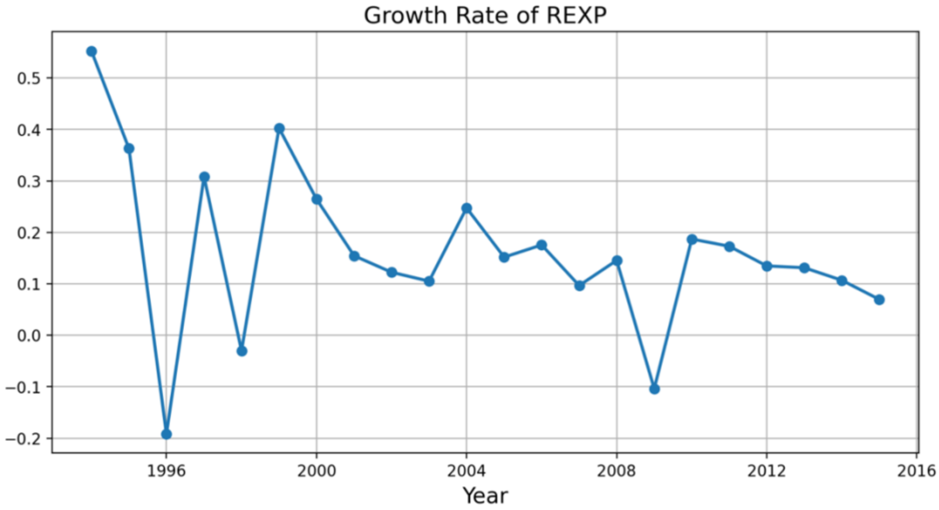


**Figure 9: Line Graph of $\Delta ln(REXP)$ and $\Delta ln(RIMP)$**



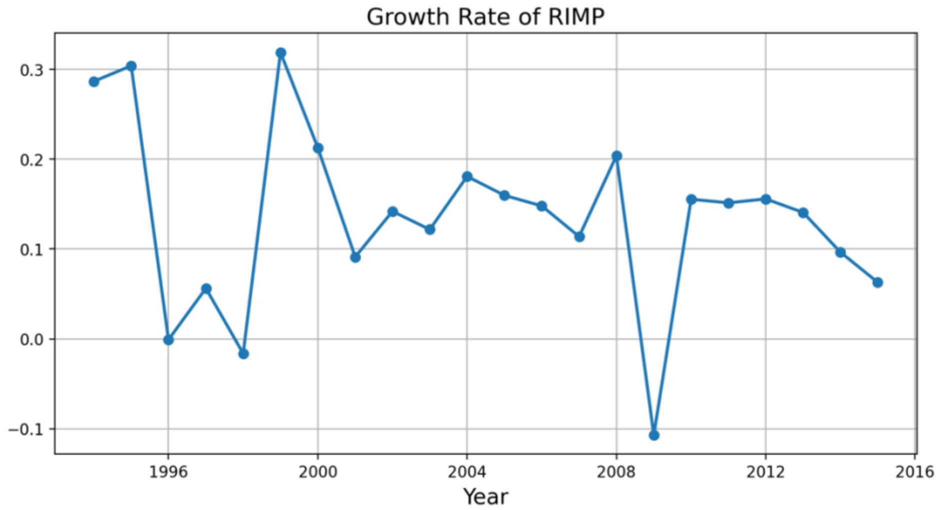




**Table 11: ADF and KPSS Unit Root Tests of $\Delta ln(REXP)$**

| | | ADF | KPSS |
|---|---|---|---|
| Test Statistics | | -6.128 | 0.359 |
| Critical Value | 1% | -3.788 | 0.739 |
| | 5% | -3.013 | 0.463 |
| | 10% | -2.646 | 0.347 |

**Table 12: ADF and KPSS Unit Root Tests of $\Delta ln(RIMP)$**

| | | ADF | KPSS |
|---|---|---|---|
| Test Statistics | | -2.245 | 0.138 |
| Critical Value | 1% | -4.137 | 0.739 |
| | 5% | -3.154 | 0.463 |
| | 10% | -2.714 | 0.347 |

### 4.4.2. Model Identification

Due to a small number of stationary observations of $\Delta ln(RGDP)$, $\Delta ln(REXP)$ and $\Delta ln(RIMP)$, only up to four lags are allowed in VAR model. According to **Table 13**, AIC value (-21.55) and BIC value (-19.62) are the lowest with four lags. Hence, among all VAR models, VAR(4) is the most preferred model.

**Table 13: Information Criteria of VAR Model with Different Lags**

| | | AIC | BIC |
|---|---|---|---|
| Lag | 1 | -17.70 | -17.10 |
| | 2 | -18.16 | -17.11 |
| | 3 | -19.55 | -18.06 |
| | **4** | **-21.55** | **-19.62** |

### 4.4.3. Model Estimation

**Table 14** presents the estimates of the preferred VAR(4) model. In RGDP and REXP equation, most of the estimated coefficients are statistically significant at 5% significant level, while only a few of those are statistically significant at 5% level in RIMP equation.

**Table 14: Regression Results of VAR(4) Model**

| Number of observations: 18 | | | | AIC = -21.55<br>BIC = -19.62 |
|---|---|---|---|---|
| Results for RGDP equation | | | | |
| | Coefficients | Standard error | t-statistics | p-value |
| constant | 0.107 | 0.035 | 3.059 | 0.002 |
| L1.RGDP | 1.766 | 0.361 | 4.890 | 0.000 |
| L1.REXP | -0.187 | 0.161 | -1.161 | 0.246 |
| L1.RIMP | -0.599 | 0.152 | -3.930 | 0.000 |
| L2.RGDP | 0.562 | 0.579 | 0.972 | 0.331 |
| L2.REXP | 0.194 | 0.083 | 2.335 | 0.020 |
| L2.RIMP | -0.462 | 0.215 | -2.141 | 0.032 |
| L3.RGDP | -2.197 | 0.740 | -2.968 | 0.003 |
| L3.REXP | 0.313 | 0.130 | 2.405 | 0.016 |
| L3.RIMP | 0.019 | 0.150 | 0.130 | 0.897 |
| L4.RGDP | 0.855 | 0.378 | 2.258 | 0.024 |
| L4.REXP | 0.223 | 0.108 | 2.063 | 0.039 |
| L4.RIMP | -0.390 | 0.182 | -2.137 | 0.033 |

Results for REXP equation

| | Coefficients | Standard error | t-statistics | p-value |
|---|---|---|---|---|
| constant | 0.320 | 0.064 | 4.938 | 0.000 |
| L1.RGDP | 4.536 | 0.667 | 6.797 | 0.000 |
| L1.REXP | -0.232 | 0.298 | -0.778 | 0.436 |
| L1.RIMP | -1.786 | 0.281 | -6.337 | 0.000 |
| L2.RGDP | -1.975 | 1.069 | -1.846 | 0.065 |
| L2.REXP | 0.854 | 0.153 | 5.566 | 0.000 |
| L2.RIMP | -1.178 | 0.398 | -2.957 | 0.003 |
| L3.RGDP | -1.959 | 1.367 | -1.433 | 0.152 |
| L3.REXP | 1.018 | 0.241 | 4.223 | 0.000 |
| L3.RIMP | -1.108 | 0.278 | -3.987 | 0.000 |
| L4.RGDP | 1.742 | 0.699 | 2.490 | 0.013 |
| L4.REXP | 0.620 | 0.200 | 3.097 | 0.002 |
| L4.RIMP | -1.23 | 0.337 | -3.661 | 0.000 |

Results for RIMP equation

| | Coefficients | Standard error | t-statistics | p-value |
|---|---|---|---|---|
| constant | 0.251 | 0.111 | 2.250 | 0.024 |
| L1.RGDP | 3.735 | 1.146 | 3.257 | 0.001 |
| L1.REXP | -0.147 | 0.512 | -0.288 | 0.773 |
| L1.RIMP | -1.443 | 0.484 | -2.980 | 0.003 |
| L2.RGDP | -1.569 | 1.838 | -0.854 | 0.393 |
| L2.REXP | 0.640 | 0.263 | 2.429 | 0.015 |
| L2.RIMP | -0.887 | 0.684 | -1.296 | 0.195 |
| L3.RGDP | -0.954 | 2.349 | -0.406 | 0.685 |
| L3.REXP | 0.697 | 0.414 | 1.684 | 0.092 |
| L3.RIMP | -0.789 | 0.477 | -1.653 | 0.098 |
| L4.RGDP | 0.809 | 1.202 | 0.673 | 0.501 |
| L4.REXP | 0.307 | 0.344 | 0.893 | 0.372 |
| L4.RIMP | -0.678 | 0.579 | -1.171 | 0.242 |

#### 4.4.4. Granger Causality Test

According to **Table 15**, the test statistics of Granger causality test (3.418) is higher than critical value at 5% significant level (2.641). This means the null hypothesis that joint four-period past values of REXP and RIMP do not Granger cause RGDP can be rejected at 5% level. Hence, REXP and RIMP explanatory variables with four lags are able to jointly forecast RGDP.

**Table 15: Granger Causality Test Results of VAR(4)**

| | | |
|---|---|---|
| Test Statistics | | 3.418 |
| Critical Value | 1% | 4.004 |
| | 5% | 2.641 |
| | 10% | 2.119 |

#### 4.4.5. Model Diagnostics

**Table 16** shows results of Portmanteau test on joint twelve lags residuals of VAR(4) model. The test statistics (83.00) is less than critical value even at 10% significant level (87.74). This means the null hypothesis that autocorrelation of residuals up to twelve lags is jointly zero cannot be rejected at 5% level. Hence, VAR(4) model is a valid model for forecasting.

**Table 16: Portmanteau Test Results of VAR(4) Model**

| | | |
|---|---|---|
| Test Statistics | | 83.00 |
| Critical Value | 1% | 102.8 |
| | 5% | 92.81 |
| | 10% | 87.74 |

#### 4.4.6. Model Forecasting

Applying the fitted VAR(4) model with iterated forecast approach, forecasted values of RGDP are obtained and presented in **Figure 10** and **Table 17**. Regarding in-sample forecast, similar to AR(1) and ARIMA(1,1,2), VAR(4) model cannot forecast indirect effects of Global Financial Crisis on Cambodia’s economy in 2009. In terms of out-sample forecast, compared to actual real economic activities, VAR(4) model forecasts relatively higher activities for all forecast horizons until 2019 and the error gets worse in 2020 due to Covid-19 shock.

**Figure 10: Graphical Comparison of VAR(4) Model Forecasting to Actual RGDP**

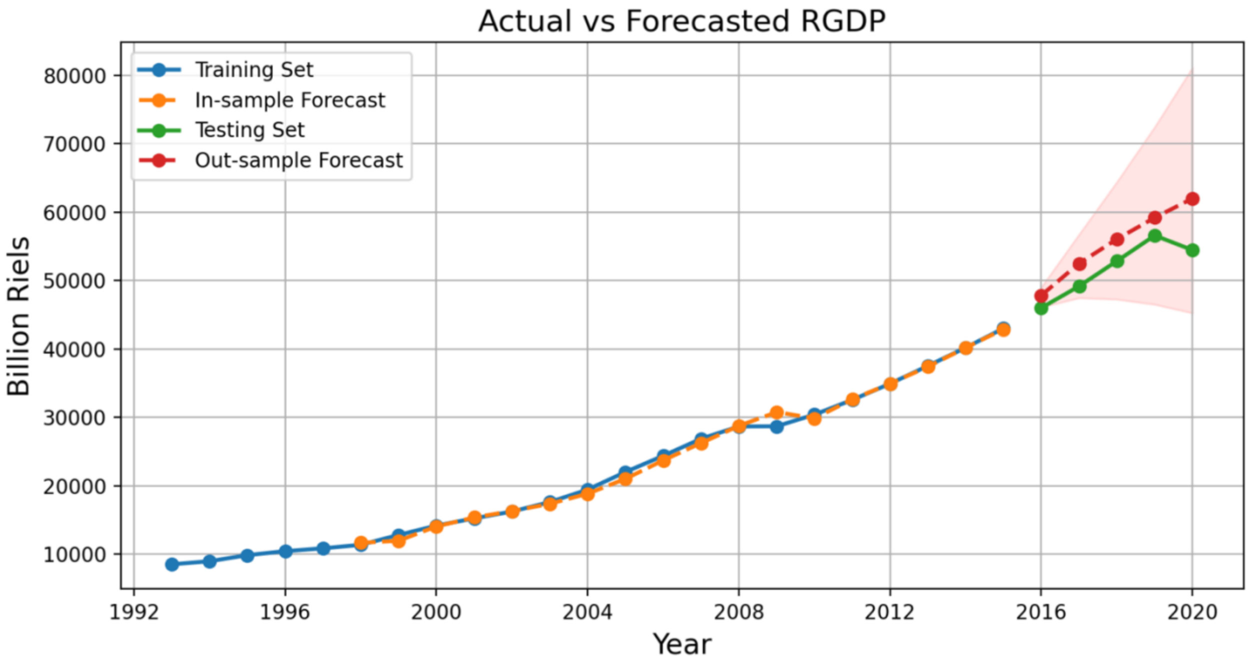


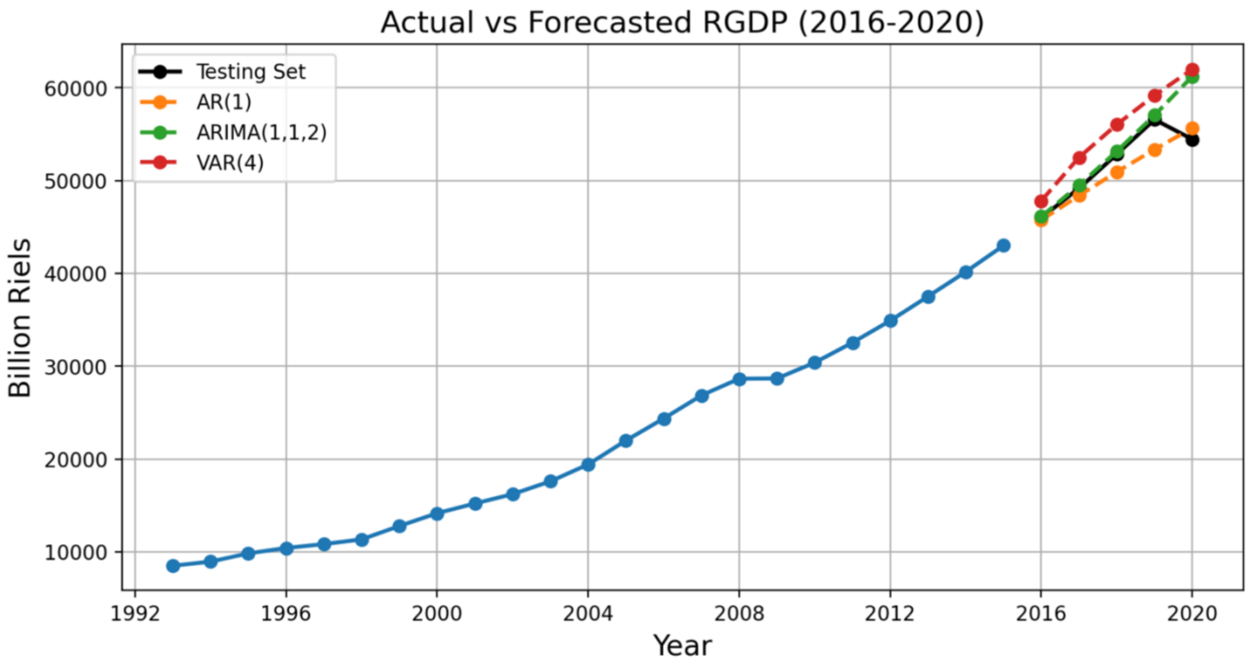

**Table 17: Forecasted Results of VAR(4) Model**

| Year | Actual RGDP | Forecasted RGDP | Forecasted Growth Rate | Error |
|---|---|---|---|---|
| 2016 | 45,999.748 | 47,829.556 | 11.21% | 3.98% |
| 2017 | 49,176.889 | 52,508.679 | 9.78% | 6.78% |
| 2018 | 52,849.994 | 56,050.104 | 6.74% | 6.06% |
| 2019 | 56,578.089 | 59,183.711 | 5.59% | 4.61% |
| 2020 | 54,454.336 | 61,998.046 | 4.76% | 13.48% |

### 4.5. Forecast Accuracy Comparison

**Figure 11** provides a visual comparison of actual RGDP against forecasted values from AR(1), ARIMA(1,1,2) and VAR(4) between 2016 and 2020. Overall, all models forecast an upward trend of real economic activities. This is not surprising due to the consistent development of Cambodia's economy, reflecting in positive growth rate in the training dataset. Compared to actual real economic activities, ARIMA(1,1,2) model produces quite accurate forecasts from 2016 to 2019. All models fail to forecast the downturn in 2020 which is caused by Covid-19 shock.

**Figure 11: Line Graph of Actual RGDP and Forecasted Values from Different Models**




Forecast accuracy metrics namely MAE, RMSE and MAPE are computed in order to evaluate the performance of each model. For objective assessment, only forecasted values between 2016 and 2019 are used to compute the metrics. According to **Table 18**, ARIMA(1,1,2) model produces the lowest MAE (311.20), RMSE (334.75) and MAPE (0.60). Hence, ARIMA(1,1,2) model is the best performing model among these three models. Therefore, ARIMA(1,1,2) model can be applied for practical forecasting analysis with a high precision.

**Table 18: Forecast Accuracy Metrics of Different Models**

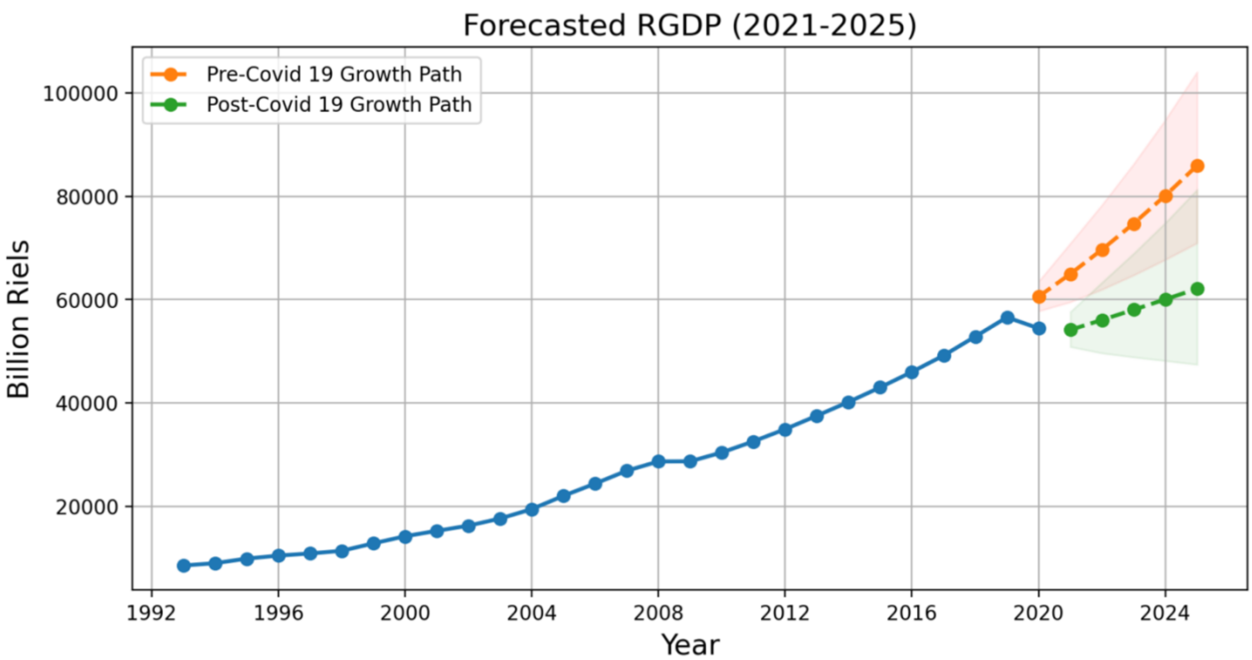
| | AR(1) | **ARIMA(1,1,2)** | VAR(4) |
|---|---|---|---|
| MAE | 1,538.73 | **311.20** | 2,741.83 |
| RMSE | 1,915.22 | **334.75** | 2,805.31 |
| MAPE | 2.86 | **0.60** | 5.35 |

By applying the best performing model with the series from 1993 to 2020, baseline future economic development for the next five years can be forecasted. Based on **Table 19** and **Figure 12**, Cambodia's economy is still overcoming the adverse impacts of Covid-19 shock in 2021 with a negative growth rate (-0.57%), then recovering from the shock in 2022 with a positive growth rate about 3.48% per year until 2025. However, the growth path of Cambodia's economy after Covid-19 crisis does not seem to be converging to that before Covid-19 shock in the near term without major government's interventions aiming to boost the economy.

**Table 19: Forecast of ARIMA(1,1,2) Model**

| Year | Forecasted RGDP | Forecasted Growth Rate |
|---|---|---|
| 2021 | 54,146.243 | -0.57% |
| 2022 | 56,071.815 | 3.56% |
| 2023 | 58,039.910 | 3.51% |
| 2024 | 60,050.575 | 3.46% |
| 2025 | 62,103.830 | 3.42% |

**Figure 12: Line Graph of Forecasted RGDP from ARIMA(1,1,2)**




**5. Summary and Conclusion**

The paper aims to test which of the three simple time series models namely AR, ARIMA and VAR model performs the best in forecasting the real GDP of Cambodia's economy given the limited available time series data. Annual observations of RGDP, REXP and RIMP from 1993 to 2020 are used for the empirical analysis. The series is split into two sets which are training set and testing set. The former consisting of about 80 percent of the sample size (between 1993 and 2015) is used to fit the model, while the latter consisting of around 20 percent (from 2016 to 2020) is used to evaluate forecast accuracy.

By applying a graphical method and unit roots test, the first order differencing of RGDP, $\Delta ln(RGDP)$ is concluded as a stationary process. Then, in addition to inspecting plot of autocorrelation and partial autocorrelation, information criteria confirms AR(1) and ARIMA(1,1,2) as the most preferred models given the training dataset. Following estimating the preferred models by maximum likelihood estimation, residuals of the fitted models are evaluated using Ljung-Box Q test which then determines residuals of both models are white noise. With iterated forecast approach, AR(1) and ARIMA(1,1,2) forecast an average growth rate of 5.30% and 7.32% for RGDP between 2016 and 2020, respectively. With respect to VAR model, REXP and RIMP variables are chosen as explanatory variables. In a similar process, $\Delta ln(REXP)$ and $\Delta ln(RIMP)$ are stationary and VAR(4) is chosen as the most preferred model using information criterion. Applying ordinary least square estimation, estimated coefficients of VAR(4) are obtained and then utilized to conduct Granger causality test, while residuals of the fitted model are used to run Portmanteau test. Both tests conclude that VAR(4) is a valid forecasting model. With iterated forecast approach, VAR(4) forecasts an average growth rate of 7.62% for RGDP from 2016 to 2020.

Given the consistent positive development over the past few decades, AR(1), ARIMA(1,1,2) and VAR(4) forecast increased economic activities for all forecast horizons between 2016 and 2020. However, except the unexpected Covid-19 shock in 2020, AR(1) and VAR(4) models tend to relatively under-forecast and over-forecast economic activities, respectively, while ARIMA(1,1,2) model produces a slightly above real economic development. In terms of forecast accuracy metrics, ARIMA(1,1,2) model produces the most accurate forecast with the lowest value of MAE (311.20), RMSE (334.75) and MAPE (0.60). For future development, ARIMA(1,1,2) model expects a negative growth rate of -0.57% for RGDP in 2021, then forecasts a positive economic development with an average growth rate of 3.48% from 2022 to 2025. However, unless there are significant policy packages aiming to boost domestic economy, the post Covid-19 growth path is more likely to diverge from the one before Covid-19 shock.

In conclusion, in data scarce environment like Cambodia, a simple ARIMA model can be used to produce short- and medium-term forecasts to get a preview of economic development path in the future. However, the forecasts from the model are baseline forecasts which assumes that there are no policy changes or major internal or external shocks. In this regard, with baseline forecasts at hands, policy makers and central bankers are able to come up with development plans and policies in a timely manner so that their desired development paths could be achieved. Last but not least, even as the most precise model among those studied in this paper, ARIMA(1,1,2) is unable to detect the emerging shock to the economy. This limitation requires further studies into more sophisticated models such as ARIMA model with an exogenous variable (ARIMAX) or VAR model with an exogenous variable (VARX).

**References**


Agrawal, V. (2018). *GDP Modelling and Forecasting using ARIMA: An Empirical Study from India.* Budapest: Central European University .

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 716-723.

Andersson, J. (2007). *Forecasting Swedish GDP Growth.* Sweden: Lund University.

Box, E.P., G., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control, Revised Edition.* University of Michigan: Holden-Day.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 427-431.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 253-263.

Ghysels, E., & Marcellino, M. (2018). *Applied Economic Forecasting using Time Series Methods.* New York, the United States of America: Oxford University Press.

Granger, C. W. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 424-438.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter.* New York, the United States of America: Cambridge University Press.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* , 159-178.

Ljung, G. M., & Box, G. E. (1978). On a Measure of Lack of Fit in Time Series Models. *Oxford University Press*, 297-303.

Mankiw, N. G. (2016). *Macroeconomics.* New York, The United States of America: Worth Publishers.

Robertson, J. C., & Tallman, E. W. (1999). *Vector Autoregressions: Forecasting and Reality*. The United States of America: Federal Reserve Bank of Atlanta.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 461-464.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 1-48.

Stock, J. H. (2001). Time Series: Economic Forecasting. *International Encyclopedia of the Social & Behavioral Sciences*, 15721.

Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics.* United Kingdom: Pearson Education .

Wei, N., Bian, K.-j., & Zhi-fa, Y. (2010). Analysis and Forecast of Shaanxi GDP Based on the ARIMA Model. *Asian Agricultural Research*, 34-36.

Zhang, H. (2013). *Modeling and forecasting regional GDP in Sweden using autoregressive models.* Sweden: Hogskolan Dalarna University.