Forecasting GDP Movements in Nepal Using Autoregressive Integrated Moving Average (ARIMA) Modelling Process

Surya Bahadur Rana[1]

Lecturer

Butwal Multiple Campus, Tribhuvan University, Nepal

Abstract

This study attempts to test the ARIMA model and forecast annual time series of GDP in Nepal from mid-July, 1960 to mid-July, 2018. The annual time series on GDP used in this study consists of total 59 observations. Out of them, three years' data from mid-July 2016 to mid-July 2018 have been used for in-sample forecasting and evaluation. The study uses univariate Box-Jenkins ARIMA modelling process to identify the best fitted model that describes the sample data set. The study examines a number of ARIMA family models and recommends ARIMA (0,1,2) as the most appropriate model that best describes the annual GDP series of the sampled period. The ARIMA (0, 1, 2) model incorporates zero lag order for autoregression, integrated with 2 lag order for moving average model using first difference operator. The ARIMA model forecasts documented in this study are not significantly different from actual because the actual annual GDP series observed in forecast period fall within 95 per cent confidence interval of estimates. Hence, ARIMA (0,1,2) model can best capture the GDP movement in Nepal for the sample period.

1. Introduction and Study Objectives

Modelling and forecasting of gross domestic product (GDP) have captured considerable interest of scholars at both academic and policy level in the context of several economies around the globe. The evolution of academic and policy interest in this area has been basically geared by the fact that GDP is considered as an important index of national economic development. Besides, GDP also

---

[1] Dr. Rana, a PhD from the Faculty of Management, Tribhuvan University, has a number of textbooks and more than two dozen of research papers to his credit. He can be reached at surryabrana@gmail.com.

helps judging the operating status of macro economy as a whole (Ning et al, 2010). The GDP measures national income and output for a given country. GDP is measured as total expenditure incurred for final goods and services produced in the country and include the expenditure incurred for consumption, investment, government expenditure and net exports. Alternatively, it can also be measured as value added by goods and services produced in an economy.

Over the last several decades, many theoretical and empirical attempts have witnessed the growing academic interest on GDP growth and its determinants. Lucas (1988) explains that increasing concern toward GDP growth is being accelerated by the need to achieve higher rate of economic growth in both developed and developing nations. The theoretical and empirical attempts toward this direction seem to have directed to explore two issues: why GDP growth rates across countries are different and what factors do cause such difference? The difference in GDP growth rate is highly reflected in the difference in living standard, and quality of life of the people living across countries. Some countries have lower level of investment, lower productivity, slower transformation in people's standard of living, and hence lower level of GDP growth (Ravallion & Chen, 1997).

The economy of the Federal Democratic Republic of Nepal is the one among the smallest economies in the world with a GDP of only $28.81 billion in 2018. Nepal has been the slower growing nation for the past 59 years (since 1960 until 2018) with an average annual growth rate in real GDP about 3.8 per cent (The World Bank, 2019). The constant unsatisfied needs and wide disparity of income and wealth among Nepalese people are basic shortcomings of Nepalese socioeconomic context. These factors have been the root causes for internal conflict over the years. However, in recent years, Nepal has made notable socio-economic progress. Literacy rates have increased, poverty rates have declined, gender disparities have narrowed, and social inclusion has improved (Asian Development Bank, 2013). Nepal now is striving toward achieving the sustainable development goals and graduating from the status of least-developed country. The support from international community in democratic transition of Nepal, to some extent, is expected to facilitate the country for making progress toward achievement of the Millennium Development Goals. In an attempt toward this direction, growth in GDP remains as the main target variable for Government of Nepal (GoN) for setting up effective and efficient strategies and policies for economic development. In this backdrop, it is necessary to provide a precise forecast of GDP in order to develop meaningful vision of the future trend of Nepalese economy. Framing appropriate strategies and policies for economic development with proper allocation of funds towards priority sectors requires a good estimate of GDP for some period ahead. It is only possible by using appropriate time series model for forecasting.

A time series represents equal spaced data points, in terms of periods, on a given variable which follows a chronological order. An examination of time series data is conducted with the expectations to discover historical patterns in the data set that may facilitate in preparing efficient forecast for the coming period. An economic time series exhibits two essential properties: non-stationary and

time-varying volatility. Non-stationary implies that a time series variable has no definite tendency to return to a linear trend. Time-varying volatility means that the volatility in economic time series varies over time. It should also be noted that non-stationarity and time-varying volatility in time series data can occur in combination as well. For this reason, no single best forecasting model exists. Thus, one of the basic issues in economic time series forecasting is that of trying to test the appropriateness forecasting model to the pattern of the available data.

Hence, this study focuses on modelling and quantifying the GDP movement in Nepal by using appropriate time series forecasting model. Several time series forecasting methods have been developed over the past decades. One of the most widely used in practice is the Box-Jenkins (1970) method. This method incorporates both moving average and autoregressive models. As Dobre and Alexandru (2008) argue, although autoregressive and moving average models were independently used before the invention of Box-Jenkins method, the basic contribution of the Box-Jenkins method rests on offering systematic methodology for identifying and estimating models that can incorporate both autoregressive and moving average approaches. This makes Box-Jenkins methodology, also known as Autoregressive Integrated Moving Average (ARIMA) model, a powerful class of models. Pankratz (1983) argues that most ARIMA models place significant emphasis on the recent past rather than the distant past and hence are appropriate for short-term forecasting. The underlying significance of ARIMA model is that it facilitates forecasting of the future trend of economic time series variables, which the policy makers can utilize for formulating policies based on the previous knowledge of such variables. In essence, the ARIMA model captures the information relating to the possible movement in future time series from the past data of the time series itself. In line with this argument, this paper attempts to use Box-Jenkins approach to test several ARIMA family models to decide on the best fitted model and forecast GDP movements using the best fitted ARIMA model in the context of Nepal using a data set of annual GDP- consisting of 59 observations from mid-July, 1960 to mid-July, 2018. Hence, the main purpose of this paper is to identify the appropriate time series ARIMA modelling of sampled time series data of annual GDP. Besides, the study also attempts to evaluate in-sample forecast of the time series variable on the fitted model.

The rest of this paper is organised thus: second section deals with literature review; third section describes data and methodology associated with Box-Jenkins process; fourth section discusses on the study results associated with model identification, parameter estimation, diagnostic checking and forecasting; and finally, fifth section presents conclusions of the study.

2. Literature Review

Macroeconomic forecasting relies on the use of econometric models to provide an estimate of future progress expected to be undertaken in the economy. Forecasting future national productivity, such as GDP, is of obvious interest in empirical

macroeconomic analysis. Forecast of next year's GDP with some degree of precision can help in formulation of macroeconomic policies and strategies. Although, it is not often possible to generate a very accurate prediction of GDP, one can at least forecast, for example, the sign of next GDP. Time series forecasting is highly utilised in predicting economic and business trends including the movements in GDP and stock market, among others. The purpose of literature review in this paper is to discuss on some of the empirical evidences associated with application of ARIMA model in modelling and forecasting economic time series.

Many studies have been conducted in past to observe the applicability of Box-Jenkins models in time series modelling and forecasting associated with business and economy. One such earlier study was conducted by Klugh and Markham (1985) with a view to compare Box-Jenkins time series forecasts to preliminary milk price estimates. By using the forecasts for monthly time series data from 1979 through 1983, the study observed that Box-Jenkins time series models outperformed the preliminary milk price estimation procedure in five States of USA. Marcek's (1998) study on economic time series forecasting was associated with stock price prediction based on 163 observed data from January 2, 1997 to December 31, 1997. The study established ARIMA (2,1,1) model as the best predictor of future stock prices. Similarly, Amiri's (1999) study on the Support Vector Machine (SVM) approach for Box-Jenkins models recognised the ARIMA model as the benchmark model in evaluating the performance of new method. By using 200 observations of time series data those used by Brockwell and Davis (1991), this study focused on the SVM approach in the modelling of AR(2), MA(1) and ARIMA(2,1,1) process. Shabri (2001) examined the performance of the Box-Jenkins method in comparison to neural networks in forecasting time series. This study was based on seasonal time series forecasting using monthly data from International Airline Passengers for 1949-1960 periods, monthly Hotel Room average for 1973-1986, and Water Demand in Utopia for 1963-1982 periods. The study found Box-Jenkins methods being most compatible in modelling and forecasting the time series data used.

In a later period, Al-Shiab (2006) examined the univariate ARIMA forecasting model, using the Amman Stock Exchange general daily index between January 1, 2004 and August 10, 2008, with out-of-sample undertaken on the following seven days. By employing different diagnostic tests to find the best model describing the data, the selected model predicted that the stock price index could grow by 0.195 per cent point for seven days. Similarly, a study conducted by Dobre and Alexandru (2008) on modelling unemployment rate using Box-Jenkins procedure based on the monthly time series data in Romania during the period 1998-2007 observed ARIMA(2,1,2) as the most adequate model for the unemployment rate and provided the forecasts value of unemployment rate for January and February 2008.

In many of the recent studies, ARIMA model has been also used to model and forecast the movement in GDP. For example, in the context of India, Maity and Chattarjee (2012) observed significant ARMA terms in one period across the GDP

series over the period 1951-2011. The study established ARIMA(1,2,3) model as the best fitted model. The results on forecasting based on this model suggested an upward trend but growth rates showing opposite trend for future periods. Similarly, in the context of developed country, particularly Sweden, Zhang (2013) observed first order ARMA as the most appropriate model for modelling and forecasting GDP. On the other hand, Shahini and Haderi's (2013) study on GDP forecasting for Albania using quarterly data from the first quarter of 2003 until the second quarter of 2013 observed no better performance of ARIMA. This study particularly compared the forecasting accuracy of two model groups, namely ARIMA and VAR. The results of this study established the VAR group models better as compared to the ARIMA group models in forecasting GDP. In the context of Pakistan, Zakia (2014) concluded ARIMA (1,1,0) being the best fitted model for modelling and forecasting of quarterly GDP of Pakistan. In the context of Jordan, Touama (2014) applied the statistical analysis for prediction of the Jordanian GDP series over the period 2003-2013 by using ARIMA process. The study results indicated the ARIMA (0, 1, 2) model as adequate with greater forecasting accuracy to forecast the Jordanian GDP. Similarly, Dritsaki (2015) used Greece GDP data from 1980 to 2013 to fit ARIMA (1,1,1) model forecasting values for three years in future. The study indicated an upward trend in GDP forecast.

There are some studies during more recent period, which have further validated the estimation power of ARIMA models. For example, in an attempt to establish the forecasting model for Kenya's GDP over the period 1960-2012, Wabomba et al. (2016) followed the Box-Jenkins method to model GDP. The study established ARIMA (2,2,2) model as the best for modelling the Kenyan GDP. The results of in-sample forecast of this study showed predicted values within the range of 5 per cent and confirmed the adequacy and efficiency of the forecasting effect of the model in modelling annual growth of the Kenyan GDP. Besides, Abonazel and Abd-Elftah (2019) used Box-Jenkins approach to build the appropriate ARIMA model for Egyptian GDP data over the period 1965-2016. The study documented ARIMA(1,2,1) model as the appropriate model for Egyptian GDP and finally used the fitted ARIMA model to forecast the GDP of Egypt for the next ten years. Similarly, Miah et al. (2019) used ARIMA model building approach to forecast the GDP of Bangladesh for the period 1960-2017. The study found ARIMA(1,2,1) model as the best fitted one. Using model selection criteria and checking model adequacy, the study established the ARIMA(1,2,1) model being in suitable shape and concluded that forecast values of GDP in Bangladesh are expected to improve steadily over the next thirteen years.

The review of above cited literatures shows mixed evidences on the appropriateness of different ARIMA family models in modelling GDP growth. The difference in results obtained is basically attributed to the difference in the pattern of GDP movement across the countries. In the context of Nepal, the economy has been revealing below average growth in many of the past years. The performances of agriculture, industry, and services sectors were less than satisfactory causing a lower economic growth

rate. In recent years, the demand for product has been increasing due to growth in size of economy. However, the industry sector of Nepal is still not getting desired momentum in the line with demand side of the economy. The lack of effective capital allocation may be the basic responsible factor for lower growth over the past years. So, it is necessary to improve effectively the allocation and absorption of capital expenditure to upgrade from the low-investment, low-growth trade-off. The evidence on the pattern of GDP movements concerning small economy of Nepal is expected to contribute toward macroeconomic policy formulation and implementation. Therefore, in the light of these issues and based on experiences of past studies, this study focuses on testing and identifying an appropriate ARIMA model using univariate Box-Jenkins methodology for annual GDP data set from mid-July, 1960 to mid-July, 2018.

3. Research Methods

This study uses univariate Box-Jenkins method of modelling and forecasting non-seasonal time series data of annual GDP series in Nepal. The data set used in this study includes annual GDP series from mid-July, 1960 to mid-July, 2018 consisting of 59 observations for modelling, in-sample forecasting and evaluation. The annual time series data set on GDP used in this study were derived from official website of World Bank, World Development Indicator. The methodological procedures applied for modelling and forecasting GDP is the Box-Jenkins approach. This approach to modelling ARIMA process was described in an influential book entitled 'Time Series Analysis, Forecasting and Control' by statisticians Box and Jenkins (1970). Box-Jenkins modelling involves identifying an appropriate ARIMA process, fitting it to the data, and then using the fitted model for forecasting. The ARIMA process is a very rich class of possible models and this model usually makes it possible to find a process, which provides an adequate description to the data (Hyndman, 2002). It is a self-projecting time series forecasting method. It aims at finding an appropriate model so that residuals are minimised and they exhibit no pattern.

The most common ARIMA model includes three parameters: p, d, and q. The parameter p is the order of autoregressive parameters, d is the order of differencing parameters, and q is the order of moving average parameters. In order to model a time series data with this approach, at first the series must be stationary. Statistically, the series is confirmed to be stationary if the n values seem to fluctuate with constant variation around a constant mean. In other words, a stationary time series exhibits mean reversions with finite and time-invariant variance and also produces a theoretical correlogram that decreases with an increase in lag length. The fundamental belief of this assumption is that mean (μ_x) and variance (σ^2_x) do not depend on time. This assumption of ARIMA model is applicable to the AR process (Box and Jenkins, 1970). If time series data set is non-stationary, differencing process is used to make it stationary. It involves calculating successive changes in the values of the data series. For example, the first order differences of the time series values

x_1, x_2, x_3, . . .,x_n are written as:

$$z_t = x_t - x_{t-1} \qquad \qquad . . . (1)$$

In Equation (1), $t = 2,....n$. If the first order differences of the original time series values are also non-stationary, then second order differences are used to produce stationary time series values. The second order differences of the time series values, x_1, x_2, x_3, . . .,x_n, are written as:

$$z_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) \qquad . . . (2)$$

In Equation (2), $t = 3, 4, ...,n$. The Box-Jenkins models are based on the idea that a time series can be usefully regarded as generated from a series of uncorrelated independent shocks (ε_t). The sequence of ε_t is called a white noise if each value of ε_t are normally and independently distributed having no pattern. It has a zero mean, $E(\varepsilon_t) = 0$, and a constant variance ($\sigma^2_{\varepsilon t}$) as such that error variance is lower than the variance of z_t. The autoregressive component of ARIMA model uses immediate past values of the series to predict future values by estimating an observation as a weighted sum of previous observations. An AR model uses 'p' number of lags in its regression, and hence is called an autoregressive model of order 'p', denoted as AR(p). The general form of AR(p) model is given by Equation (3).

$$z_t = \mu + \phi_1 z_{t-1} + \phi_2 z_{t-2} +...+ \phi_p z_{t-p} + \phi_t \quad . . . (3)$$

In Equation (3), z_t is a stationary difference time series, ε_t is a hypothetical white noise, and ϕ_1, ϕ_2, ..., ϕ_p are AR model parameters to be estimated.

The moving average (MA) component of ARIMA model attempts to model an underlying time series by averaging it for a sufficient period of time. The 'invertibility' is the underlying assumption of MA process in ARIMA modelling. It refers to the linear stationary process behaving as infinite representation of AR process. Invertibility is essential as such that time series can be represented by a finite order MA process or convergent autoregressive process. In essence, the idea behind this method is to eliminate the influence of shorter-term fluctuations. The general form of the MA model of order q is denoted as MA(q), and the model is represented as:

$$z_t = \mu \, \phi \, \theta_1 \theta_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad . . . (4)$$

In Equation (4), z_t is a difference time series, ε_t is a hypothetical white noise, and θ_1, θ_2,...,θ_q are MA model parameters to be estimated.

The ARIMA model incorporates both autoregressive and moving average components similar to ARMA. However, one distinction between ARMA and ARIMA is the integration component in the ARIMA model, which is necessary for the stationarity. In reality, most economic variables are non-stationary. Hence, they have to be taken through the transformation process called differencing before they become stationary. Therefore, ARIMA model informs the reader that the time series in question has gone through an integration process before being

used for analysis. The ARIMA model presents a time series of a variable as a linear function of past values of the variable and random error terms. ARIMA (p,d,q) is combination of AR, MA and difference terms as follows:

$$z_t = \mu + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} - \phi_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t. \qquad \ldots (5)$$

In Equation (5), z_t is a difference time series, ε_t is a hypothetical white noise, and ϕ_1, ϕ_2, ..., ϕ_p; ϕ_1, ϕ_2,...,ϕ_q are autoregressive and moving average model parameters to be estimated.

Under Box-Jenkins Process, a careful observation of the behaviour of autocorrelation function (ACF) and partial autocorrelation function (PACF) is necessary to determine the tentative ARIMA model. Box and Jenkins (1970) suggest that the number of lags should be no more than n/4 autocorrelations. The autocorrelation coefficient is the measure of direction and degree of linear association between the stationary time series random variables z_t and z_{t+k} separated by constant time lag 'k'. The sample autocorrelation coefficient at lag 'k' (r_k) is used to estimate the theoretical autocorrelation function. The sample autocorrelation coefficient is given by:

$$r_k = \frac{\Sigma(z_t - \hat{z})(z_{t+k} - \hat{z})}{\Sigma(z_t - \hat{z})^2} \qquad \ldots (6)$$

In Equation (6), z_t is the data from stationary time series, z_{t+k} is the data from k time period ahead of t, and $\hat{z}$ is the mean of the stationary time series.

The estimated PACF and ACF are used - in choosing one or more ARIMA models that fit the available data. The PACF measures correlation between time series observations that are k time periods apart after controlling for correlations at intermediate lags (that is, lags less than k). In other words, it measures the strength of the relationship between observations in a series controlling for the effect of intervening time periods. Therefore, the idea of using PACF is to measure how lag variables are related. The PACF is given by:

$$\hat{\phi}_{kk} = \frac{\sum_{j=1}^{k-1} \phi_{k-1,\,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,\,j}\, r_j} \qquad (if\ k = 2,\ 3,\ \ldots) \quad \ldots (7)$$

Where,

$$\phi_{kj} = \phi_{k-1,\,j} - \phi_{kk}\, \phi_{k-1,\,k-j}\,; \quad (j = 1,\ 2,\ 3,\ \ldots,\ k-1)$$

The ACF and PACF are used to find a stationary time series. If the sample ACF and PACF of the difference series either cuts off fairly quickly or dies down fairly quickly, the time series values are considered stationary. In this case, model is not strictly

AR or MA but an ARIMA model. Once a stationary time series has been selected, a tentative model can be identified by examining the behaviour of the ACF and PACF. In the mixed model, both the ACF and PACF die down exponentially.

The ARIMA model building process involves four steps: identification, estimation, diagnostic checking, and forecasting. At the identification stage ACF and PACF are used to choose one or more ARIMA models that seem appropriate. In the next step, the parameters of specific Box-Jenkins model for a given series are estimated. This stage confirms whether the selected AR and/or MA parameters have the lowest sum of the squared residuals, residuals are not correlated to one another, and the model is parsimonious or not over-specified. In an ARIMA model, parsimony is the key issue. A parsimonious model gives better forecast than an over-parameterised model. In the third step, model diagnostic checking is conducted. Model diagnostic for Box-Jenkins models is similar to model validation for non-linear least square fitting. In other words, the error terms, ε_t, is assumed to follow the assumptions for a stationary unvaried process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance. The best fitted Box-Jenkins model can satisfy these assumptions of the residual distribution. If these assumptions are not satisfied, it is necessary to search the most appropriate model. The analysis of the residuals can provide some clues toward more appropriate model. The residual analysis is based on random residuals test statistics known as Ljung Box test (Q). According to this test, if Q $\leq$ chi-square, the ACF patterns are white noise and the model is the appropriate. On the other hand, if Q > chi-square, the ACF patterns are not white noise and the model-building cycle has to be repeated. If the model is valid after diagnostic checking, the model can be used for forecasting. The optimum forecast is one that has the minimum error. Other methodological issues associated with ARIMA modelling and forecasting are discussed in the ensuing section.
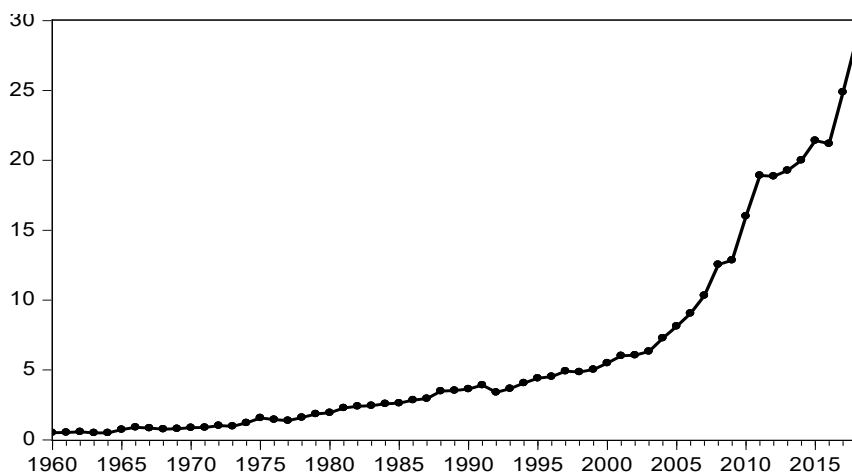

Figure 1. *Annual Level GDP Series, 1960-2018*

4. Data Analysis and Discussions

For the purpose of empirical development and verification of appropriate ARIMA model, this study considers the annual GDP series from mid-July, 1960 to mid-July, 2018 consisting of total 59 observations. The empirical development and analysis of the appropriate ARIMA model has been conducted through the steps mentioned below as suggested by Box and Jenkins (1970).

4.1 Identification for Stationarity

The first step in developing the Box-Jenkins model is determining if the series is stationary. A unit root test determines whether a time series variable is non-stationary using an autoregressive model. The test statistic proposed in this study is the Augmented Dickey-Fuller (ADF) Test. This test uses the existence of a unit root as the null hypothesis. In this study, the correlogram representing ACF and PACF of the level series of the annual GDP non-seasonal time series data for the sampled period has been plotted and the ADF unit root test has been performed. Before testing for unit root, it is necessary for visualise the pattern of raw data over the samples period. Figure 1 shows the plot of annual level GDP series for the period 1960-2018. An observation of Figure 1 reveals that there is a clear pattern of upward movement in GDP series over the period.

Figure 2 shows the correlogram consisting of ACF and PACF for annual level GDP level series for 24 lags (only 10 lags reported here). As revealed in Figure 2, the ACF declines very slowly up to 24 lags and they are outside the 95 per cent confidence interval. Similarly, the PACF drops immediately after the first lag. It is an indicative of a non-stationary series. So, correlogram in Figure 2 gives an initial indication for non-stationary level GDP series, which has been further confirmed by ADF unit root test results reported in Table 1.

Autocorrelation	Partial Correlation	Lag	AC	PAC	Q-Stat	Prob.
		1	0.896	0.896	49.835	0.000
		2	0.812	0.045	91.461	0.000
		3	0.750	0.075	127.61	0.000
		4	0.677	-0.072	157.59	0.000
		5	0.607	-0.021	182.17	0.000
		6	0.539	-0.043	201.89	0.000
		7	0.467	-0.056	217.00	0.000
		8	0.383	-0.116	227.36	0.000
		9	0.315	0.009	234.52	0.000
		10	0.266	0.048	239.73	0.000

Figure 2. *Correlogram of annual Level GDP Series, 1960-2018*

The ADF unit root test shows that annual level GDP series for the period 1960-2018 is non-stationary as the p- value of test statistic is greater than 5 per cent level. So, the result suggests no sufficient evidence to reject the null hypothesis of unit root.

Table 1

The ADF Unit Root Test Results on Annual Level GDP Series, 1960-2018

			t-Statistic	*Prob.**
Augmented Dickey-Fuller test statistic			3.008146	1.0000
Test critical values:	1% level		-3.568308	
	5% level		-2.921175	
	10% level		-2.598551	
*MacKinnon (1996) one-sided p-values.				

Source.Author's computation using level GDP series in Eviews 10 version

A log transformation of level GDP series were also conducted taking natural logarithm of GDP to see whether the natural logarithm of GDP series (LnGDP) would ensure the stationarity. The ACF and PACF of LnGDP series showed the similar pattern as to that of level GDP series and the result of ADF unit root test again revealed no sufficient evidence to reject the null of unit root in LnGDP series. Hence, in the next step, the autocorrelation and partial autocorrelation functions using first difference operator of LnGDP series have been examined to confirm the stationarity in first difference log GDP series, D(LnGDP). The results are reported in Figure 3 showing correlogram of ACF and PACF of D(LnGDP) series. An observation of Figure 3 shows that most of the ACF and PACF are within 95 per cent confidence inter vals up to specified 24 lags (only reported up to 10 lags here). So, first difference log GDP series is said to be stationary.

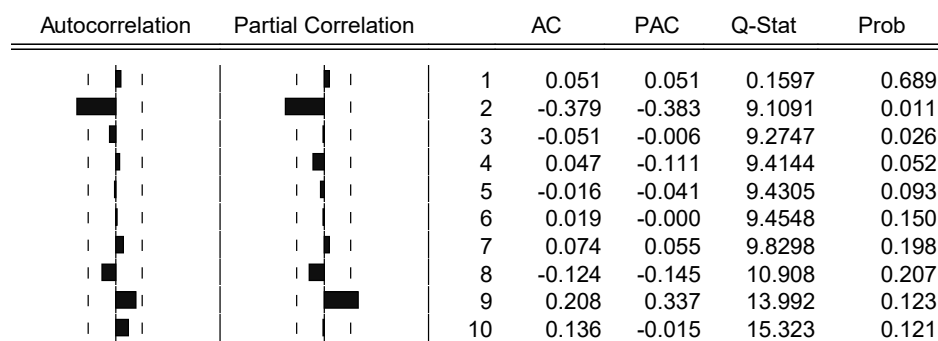| Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.051	0.051	0.1597	0.689
		2	-0.379	-0.383	9.1091	0.011
		3	-0.051	-0.006	9.2747	0.026
		4	0.047	-0.111	9.4144	0.052
		5	-0.016	-0.041	9.4305	0.093
		6	0.019	-0.000	9.4548	0.150
		7	0.074	0.055	9.8298	0.198
		8	-0.124	-0.145	10.908	0.207
		9	0.208	0.337	13.992	0.123
		10	0.136	-0.015	15.323	0.121

Figure 3. *The Correlogram for First Difference of Log GDP Series, 1960-2018*

Table 2 shows the results of ADF unit root test for stationary on first difference series of log GDP. The appropriate lag length was selected based on the Akaike Information Criterion (AIC). The test statistic employed to ensure data stationary is again the Augmented Dickey-Fuller (ADF) test. The reported values are test statistic with corresponding p-value. The results in Table 2 reveal the fact that there is sufficient evidence to reject the null hypothesis of unit root at 1 per cent

level as p-value of the test statistic is less than 1 per cent. Hence, the results of this workout confirm at 1 per cent level that the first difference of log GDP series is stationary. Therefore, the tentative order of differencing is identified as d = 1.

Table 2

The ADF Unit Root Test Results on First Difference of Log GDP Series, 1960-2018

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.483295	0.0000
Test critical values:	1% level	-3.552666	
	5% level	-2.914517	
	10% level	-2.595033	

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

Source. Author's computation using first difference of log GDP series in Eviews 10 version

The tentative orders of autoregressive and moving average processes have been determined by examining the ACFs and PACFs of the first difference of log GDP series in Figure 3 computed for 24 lags. The ACF of D(LnGDP) series in Figure 3 shows that autocorrelation is significant at lag order 2 and the ACF shows good exponential decay and a damped sine-wave pattern after lag 2. The ACF at lag 2 shows larger and significant negative spike. This implies that the autocorrelations of the successive pairs of observations in time period 2 is not within sampling error of zero. All other autocorrelations thereafter are within 95 per cent confidence limit. Hence, the tentative order of moving average process can be 2 (that is, q = 2).

The Figure 3 also shows the partial autocorrelation functions (PACFs) for the first difference of log GDP series. The partial autocorrelation for lag 2 is -0.383 and it is significantly different from zero. Also, there is a negative significant spike at lag 2 and other partial autocorrelations are within 95 per cent confidence limit. So, the tentative order of autoregressive process also can be 2 (that is, p = 2). In Figure 3, both ACF and PACF have almost the same pattern. So, this is an ARIMA process. The pattern is typical to autoregressive and moving average processes of lowest order 0 to highest order 2, that is, tentative AR(p) = 2 and MA(q) = 2. Thus, the study identified the tentative order of autoregressive parameter (p), order of differencing parameter (d), and order of moving average parameter (q). These are tentative orders of autoregressive and moving average components because the judgment about order of these parameters is made very subjectively in Box -Jenkins process. Therefore, multiple ARIMA (p, d, q) models are tested for a range of values of p and q to establish a more accurate model.

4.2 Estimation of the model and diagnostic checking

In this study, estimation of the best ARIMA model is based on volatility, standard error of regression, adjusted R- square, Akaike Information Criterion (AIC) and

Schwarz Information Criterion (SIC). An ARIMA model is said to be the best if it has the least volatility measured by Sigma-square (SIGMASQ). Lowest value of sigma-square makes predictive power of the estimated model the most efficient. Similarly, the best model must have higher adjusted R-square (Adj. R2), and lower values of standard error of regression (SER), AIC and SIC. The study estimated more models of 0, 1 and 2 orders in autoregressive and moving average process in order to determine the right specification based on Akaike Information Criterion (AIC). The correlogram of D(LnGDP) series suggests the necessity of introduction of both analysed variable lags and the lags of the error in the process of estimation. The results of estimated models of ARIMA family are reported in Table 3. This shows choices of ARIMA models based on SIGMASQ, Adj. R2, SER, AIC and SIC.

Table 3

Test Results of ARIMA(p, d, q) Model Fitting

Model	SIGMASQ	Adj. R2	SER	AIC	SIC
ARIMA (0,1,1)	0.0092	-0.0254	0.0983	-1.7501	-1.6435
ARIMA (1,1,0)	0.0092	-0.0336	0.0987	-1.7427	-1.6361
ARIMA (1,1,1)	0.0078	0.1058	0.0918	-1.8296	-1.6875
ARIMA (2,1,0)	0.0079	0.1171	0.0912	-1.8951	-1.7885
ARIMA (0,1,2)	0.0078	0.1322	0.0905	-1.9107	-1.8041
ARIMA (2,1,1)	0.0079	0.1049	0.0919	-1.8651	-1.7230
ARIMA (1,1,2)	0.0077	0.1206	0.0911	-1.8809	-1.7388
ARIMA (2,1,2)	0.0078	0.1163	0.0913	-1.8766	-1.7345

Source. Author's computation running various ARIMA family models in Eviews 10 version

Table 3 indicates that the best model among all fitted models is ARIMA (0,1,2), which has the highest adjusted R-square and the lowest values of SER, AIC and SC. At this stage, it is necessary to emphasise that lower values of AIC and SIC are not only the necessary conditions for the best model. These criteria are simply used to establish first a model with lower AIC and SIC, which later has to be passed through parameter significant test and residual diagnostic test.

Table 4 exhibits the estimate of parameter along with corresponding standard errors, t-statistics, and p-values for the ARIMA(0,1,2) Model. Also reported are the statistics of fit consisting of adjusted R-square, standard error of regression, F-statistic, AIC and SC. (*) sign indicates that results are significant at 1 per cent level. It can be seen from Table 4 that the coefficient of MA(2) is highly significant at 1 per cent level. The model also represents a goodness of fit to the sampled data. It implies that the ARIMA(0,1,2) model comfortably explains the data and is suitable for forecasting.

Table 4

Estimation Results of ARIMA(0,1,2) Model [Dependent Variable: D(LnGDP)]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0692*	0.0070	9.8369	0.0000
MA(2)	-0.4273*	0.1039	-4.112	0.0001
SIGMASQ	0.0078*	0.0013	6.1333	0.0000
R-squared	0.1627	Mean dependent var		0.0696
Adjusted R-squared	0.1322	S.D. dependent var		0.0971
S.E. of regression	0.0905	Akaike info criterion		-1.9107
Sum squared resid.	0.4500	Schwarz criterion		-1.8041
Log likelihood	58.4105	Hannan-Quinn criter.		-1.8692
F-statistic	5.3428	Durbin-Watson stat		1.8586
Prob(F-statistic)	0.0076			
Inverted MA Roots	.65			

Source. Author's computation running ARIMA (0, 1, 2) model in EViews 10 version

In selecting a good model, the need for parsimonious model has been considered. A parsimonious model produces better results than an over-parameterised model. With many variables, the goodness of fit of the model improves with an increasing R-square. But such model is penalized with a reducing adjusted R-square, which may tilt towards zero or even become negative if there are too many irrelevant variables added to the model. Similarly, too many variables consume degrees of freedom more. As a result, those variables contribute little to the significance of the dependent variable.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.068	0.068	0.2852	
		2	-0.015	-0.020	0.2999	0.584
		3	-0.010	-0.007	0.3060	0.858
		4	0.042	0.043	0.4187	0.936
		5	0.039	0.033	0.5162	0.972
		6	-0.008	-0.011	0.5202	0.991
		7	0.155	0.160	2.1658	0.904
		8	-0.126	-0.154	3.2640	0.860
		9	0.167	0.204	5.2341	0.732
		10	0.022	-0.022	5.2708	0.810
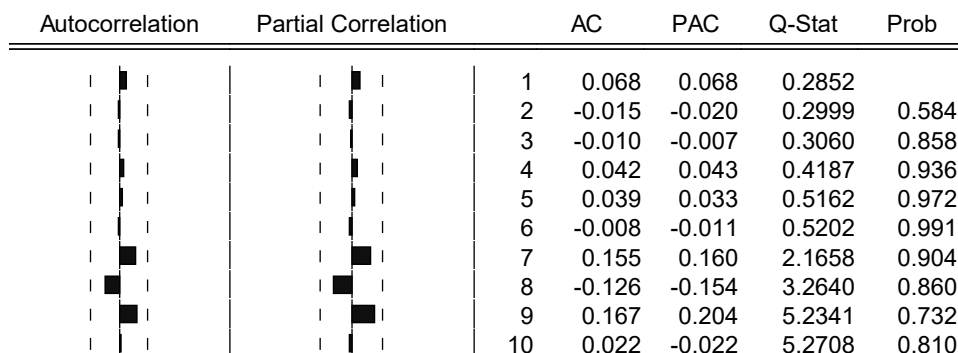
Figure 4. *Correlogram of Residuals for ARIMA (0,1,2) Model*

Before forecasting with the final model, it is necessary to perform various diagnostic tests in order to validate the goodness of fit of the model. A very strict assumption, in choosing the most accurate model, is that

residuals are assumed to have white noise, and uncorrelated. Thus, the adequacy of the Box-Jenkins model is evaluated by examining the residuals. For a randomly distributed residuals, the autocorrelations and partial autocorrelations are statistically zero. If they are not, the fitted model is not correct. The autocorrelations and partial autocorrelations of the residuals up to fixed lags are calculated to test for autocorrelation. For this purpose, the model residual autocorrelations for 24 lags have been examined in this study using the Ljung- Box Q statistic tests for serial autocorrelation. Figure 4 and 5 present correlogram of the residual and correlogram of squared residual, respectively with autocorrelations and partial autocorrelations of the of ARIMA(0,1,2) model. Also reported are the corresponding p-values.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.011	0.011	0.0071	0.933
		2	0.185	0.185	2.1290	0.345
		3	-0.026	-0.030	2.1705	0.538
		4	-0.119	-0.158	3.0777	0.545
		5	-0.111	-0.104	3.8900	0.565
		6	-0.008	0.050	3.8944	0.691
		7	-0.079	-0.044	4.3187	0.742
		8	0.054	0.025	4.5193	0.807
		9	0.070	0.071	4.8625	0.846
		10	0.082	0.062	5.3480	0.867

Figure 5. *Correlogram of Residuals Squared for ARIMA (0,1,2) Model*

As reported in Figure 4 and 5, the correlogram of the residuals and correlogram of squared residuals are both flat. It indicates that all information has been captured by the model. As it can be observed, p-values are all above 5 per cent level of significance. Employing Ljung-Box Q statistic, the result indicates that model residual autocorrelations and partial autocorrelations of all lags are not significantly different from zero. This implies that all the spikes are within the 95 per cent confidence limit, residuals are random and the model is a good fit to data. The best fitted ARIMA (0, 1, 2) model includes zero lag order of autoregression, first difference operator, and 2 lag order of moving average components. Finally, the forecast will be based on the ARIMA(0,1,2) model.

Figure 6 presents inverse roots of AR and MA characteristic polynomials for the stability of the selected ARIMA(0,1,2) model. The inverse roots of AR/MA polynomials show whether the estimated model is stable (stationary) or not. If all roots are less than one and fall inside the unit circle, the model is said to be stable. In otherwise case, the impulse response standard errors are not considered valid. An observation of Figure 6 demonstrates that the selected ARIMA model is stable as inverse roots of the characteristic polynomials are inside the unit circle meaning that they are less than unit.

ARIMA(0,1,2) is used to fit the D(LnGDP) series, and the result is shown in Figure 7. The figure shows actual values by solid line, fitted values by upper dotted line and residual of the model by lower dotted line. The actual, fitted and residual graph shows the actual and fitted values of the dependent variable and the residuals from the regression. The actual values are the sum of fitted values and residuals. The residuals are shown with a bound of plus minus one estimated standard error. A visual observation of Figure 7 shows that there is a good fit of actual and fitted values of dependent variable because maximum values of residuals are within the bound of plus minus one estimated standard error.
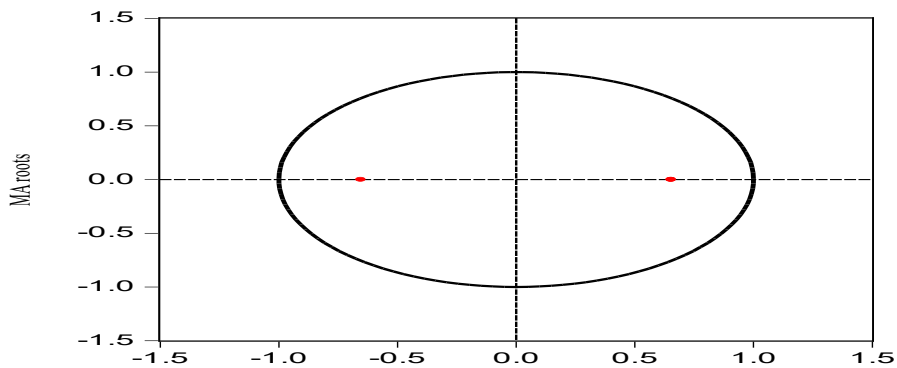

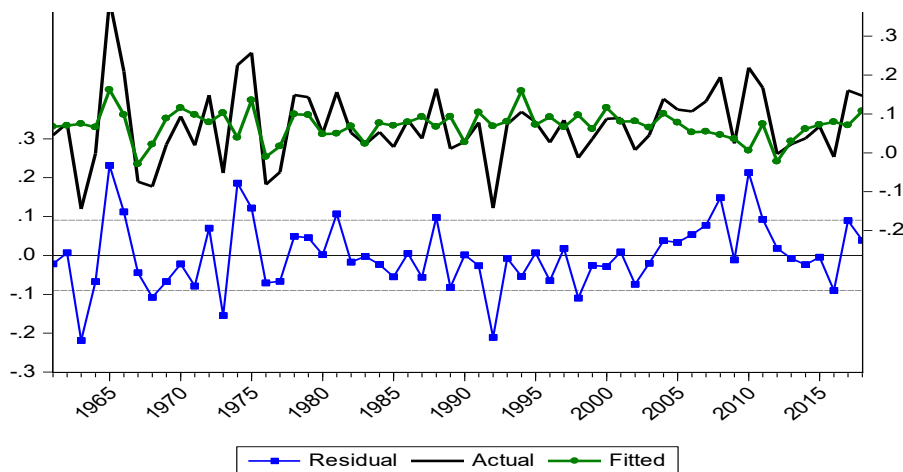Figure 6. *Inverse Roots of AR/MA Polynomials*

Figure 7. *Actual, Fitted, and Residual Graph of First Difference LnGDP Series, 1960-201*

4.3 Forecasting

Before providing an in-sample forecast of GDP, it is necessary to check whether there is any structural break in the model. For this purpose, Chow break point test has been conducted. Chow breakpoint test fits the equation separately for
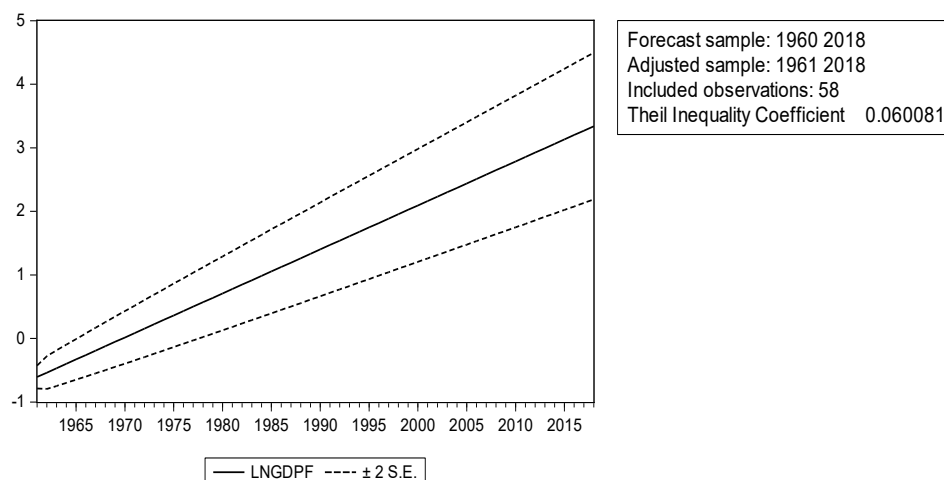
each subsample to see whether or not there are significant differences in the estimated equation. If there is a significant difference, it shows a structural break in the relationship. To carry out Chow test, it is necessary to partition the data into at least two sub-samples selecting a breakpoint. It tests the null hypothesis that there is no break at specified breakpoint. For the purpose of this study, the year 2003 has been selected as the breakpoint because there are some major events (such as cease fire between rebels and government in January and pulling out of peace talks with government in August ending seven-month truce, and resurgence of violence and frequent clashes between students/activists and police during the following months) took place during this in year in the history of Nepal. The results of Chow test are revealed in Table 5.

Table 5

Results of Chow Breakpoint Test for Structural Break: 2003

F-statistic	2.1026	Prob. F(3,52)	0.1111
Log likelihood ratio	7.0438	Prob. Chi-Square(3)	0.0705
Wald Statistic	6.9356	Prob. Chi-Square(3)	0.0740

As shown in Table 5, all test statistics (F-statistic, log likelihood ratio and Wald statistic) are not significant at 5 per cent level. It supports the hypothesis that there is no sufficient evidence to reject the null of no structural break. Hence, it can be concluded that there is no structure break in the selected ARIMA(0,1,2) model.

Forecast sample: 1960 2018
Adjusted sample: 1961 2018
Included observations: 58
Theil Inequality Coefficient   0.060081

LNGDPF   ---- ± 2 S.E.

Figure 8. *In-Sample Forecast of GDP*

The essence of fitting the ARIMA model is to forecast future values of the series. Therefore, after the validation of ARIMA(0,1,2) model through diagnostic checking procedure, the model has been used in this study to provide an in-sample forecast for the annual GDP series, 1960 – 2018. It is appropriate to choose the predicted value ahead for forecasting because ARIMA model is adequate for the forecasting. The Figure 8 shows in-sample forecast of GDP for the period, 1960-2018. The

forecasted value of LnGDP fall within plus-minus two estimated standard error. The Theil Inequality Coefficient is 0.06, which is closer to zero. It implies that forecast value of time series is closer to the observed values. So forecast is better.

Finally, the predicted values of annual GDP series for the year 2016 through 2018 for in-sample forecasting are reported in Table 6 along with 95 per cent confidence interval of estimates.

Table 6

Actual and Predicted Values of Annual GDP using ARIMA (0,1,2) Model for 2016-2018 Periods

Year	Actual	Predict	95% Confidence Interval	
			Lower	Upper
2016	21.19	23.26	19.33	27.75
2017	24.88	22.70	18.87	27.08
2018	28.81	28.07	23.34	33.49

Source. Author's computation running forecast command in Eviews 10 version

The actual data of 3 years in-sample forecasting period are from 2016 to 2018. The ARIMA model forecasts documented in this study are not significantly different from actual because all actual annual GDP series observed in 3 years fall within 95 per cent confidence interval of estimates. Hence, ARIMA(0,1,2) model can best capture the GDP movement in Nepal for the sample period. The result of this study confirms to Touama (2014), who indicates that ARIMA (0, 1, 2) model is adequate to forecast the Jordanian GDP with greater forecasting accuracy.

5. Conclusion and Implications

This study attempted to establish appropriate ARIMA model to describe non-seasonal time series data using Box-Jenkins methodology. For the purpose, time series data on annual GDP from the period mid-July, 1960 to mid-July, 2018 were examined as the sample time series data. The time series data set included a total of 59 annual observations. Initially the level series was found non-stationary. The unit root test of level series data suggested for the first differencing of log GDP to produce the stationary series. So that first order differencing of log GDP was applied to make the series stationary. The ADF unit root test on the first difference of log GDP series was conducted again to confirm the stationarity of data. The ADF test suggested the stationarity of first difference of log GDP series. Going through the process of identification and estimation, several ARIMA family models were estimated. Diagnostic checking on all estimated models was conducted based on SIGMASQ, Adj. R2, SER, AIC and SIC for selecting the most efficient model. The results of diagnostic check established the ARIMA(0,1,2) model as the most efficient model. The estimated parameters of ARIMA (0, 1, 2) were significant at 1 per cent level. The selected model was also used to provide

the in-sample forecasts for each of the year from the 1961 to 2018. The proposed model performed best for in-sample forecasting. These results are consistent to the findings of Touama (2014) in Jordanian context, who documented the adequacy of ARIMA (0, 1, 2) model in forecasting Jordanian GDP with greater accuracy. Overall, the Box-Jenkins methodology of time series forecasting used in this study is considered to have the robust forecast of the annual GDP series for the given period as all observed annual GDP series during the forecast period are within the 95 per cent confidence interval of estimate.

One basic limitation of this study is the length of GDP data points used. Although ARIMA modelling process can be applied if the length of data points exceeds 50 observations, it is still not adequate. Nepal has not yet adopted the practice of announcing GDP in quarterly basis and therefore study was confined to the use of annual GDP series limiting the number of annual observations to 59. This study examined the movement in aggregate GDP in Nepal to test the efficient ARIMA model. In Nepal, however, there are different sectoral components of GDP such as industrial, agriculture and service sector GDP. The movement in these components of GDP may follow different patterns, and hence unique ARIMA models may be efficient to capture the movement in each of these components of GDP. Therefore, future studies are suggested to test the efficient forecasting models for sectoral GDP. Moreover, this study was simply confined to the selection of best performing ARIMA model in capturing GDP movements in Nepal. There are several other econometric forecasting models such as neural networks, vector autoregression models and so on. Therefore, future studies may be directed towards a comparison of these several forecasting models.

References

Abonazel, M. R., & Abd-Elftah, A. I. (2019). Forecasting Egyptian GDP using ARIMA models. *Reports on Economics and Finance, 5*(1), 35-47.

Al-Shiab, M. (2006). The predictability of the Amman stock exchange using the univariate ARIMA model. *Journal of Economic and Administrative Sciences*, *22*(2), 1-18.

Amiri, S., Rosen, D.V., & Zwanzig, S. (2009). The SVM approach for Box-Jenkins model. *REVAST, 7*(1), 23-36.

Asian Development Bank. (2013). *Nepal fact sheet*. Philippines, Manila: Author.

Box, G.E.P., & Jenkins, G.M. (1970). *Time series analysis, forecasting and control*. San Francisco: Holden Day.

Brockwell, P.J., & Davis, R.A. (1991). *Time series: Theory and methods*. New York: Springer.

Dobre, I., & Alexandru, A.A. (2008). Modeling unemployment rate using Box-Jenkins procedure. *Journal of Applied Quantitative Methods, 3*(2), 156-166.

Dristaki, C. (2015). Forecasting real GDP rate through econometric models: An empirical study from Greece. *Journal of International Business and Economics, 3*(1), 13-19.

Hyndman, R. (2002). Box-Jenkins modelling. In H. Daellenbach, & R. Flood, (Eds.), *The informed student guide to management science*. London: Thomson.

Klugh, B.F., & Markham, J. (1985). *A comparison of Box-Jenkins time series forecasts to preliminary milk price estimates* (SRS staff report number, 85-14). United States Department of Agriculture,.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*(1), 3-42.

Maity, B., & Chatterjee, B. (2012). Forecasting GDP growth rates of India: An empirical study. *International Journal of Economics and Management Sciences, 1*(9), 52-58.

Marcek, M. (1998). *Economic time series forecasting: Box-Jenkins methodology and Ssgnal processing approach* (Working paper). Opava: The Silesian University, Institute of computer science.

Miah, M. M., Tabassum, M, & Rana, M. S. (2019). Modeling and forecasting of GDP in Bangladesh: An ARIMA approach. *Journal of Mechanics of Continua and Mathematical Science, 14*(3), 150-166. doi: 10.26782/jmcms.2019.06.00012

Ning, W., Kuan-Jiang, B., & Zhi-Fa, Y. (2010). Analysis and forecast of Shaanxi GDP based on the ARIMA model. *Asian Agricultural Research, 2*(1), 34-41.

Pankratz, A. (1983). *Forecasting with univariate Box-Jenkins models*. New York: John Wiley & Sons.

Ravallion, M., & Chen, S. (1997). What can new survey data tell us about recent changes in distribution and poverty. *World Bank Economic Review, 11*(2), 357-382.

Shabri, A. (2001). Comparison of time series forecasting methods using Neural networks and Box-Jenkins model. *Mathematika, 17*(1), 25-32.

Shahini, L., & Haderi, S. (2013). Short term Albanian GDP forecast: One quarter to one year ahead. *European Scientific Journal, 9*(34), 198-208.

The World Bank. (2019, *December*). *Nepal development update*. Washington DC: Author.

Touama, H. Y. (2014). Application of the statistical analysis for prediction of the Jordanian GDP by using ARIMA time series and Holt's linear trend models for the period 2003 – 2013. *Mathematical Theory and Modelling, 4*(14), 19-26.

Wabomba, M. S., Mutwiri, M. P., & Fredrick, M. (2016). Modeling and forecasting Kenyan GDP using ARIMA models. *Science Journal of Applied Mathematics and Statistics, 4*(2), 64-73. doi: 10.11648/j.sjams.20160402.18

Zakia, M. (2014). A time series modeling on GDP of Pakistan. *Journal of Contemporary Issues in Business Research, 3*(4), 200-210.

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

Conflict of interest

The author has no conflict of interest in the research work.