Canadian Journal of Business and Information Studies, 5(3), 63-69, 2023

Publisher homepage: www.universepg.com, ISSN: 2663-7820 (Online) & 2663-7812 (Print)

https://doi.org/10.34104/cjbis.023.063069

Canadian Journal of Business and Information Studies

Journal homepage: http://www.universepg.com/journal/cjbis

Forecasting the GDP in the United States by Using ARIMA Model

Maysoon A. Sultan*

Department of Mathematics, Faculty of Science, Hafr- Albatin University, Saudi Arabia.

*Correspondence: maisona@uhb.edu.sa (Dr. Maysoon A. Sultan, Assistant Professor, Dept. of Mathematics, Faculty of Science, Hafr- Albatin University, Saudi Arabia).

ABSTRACT

This paper aims to give an overall view of the Gross Domestic Product (GDP) in the United States and determine the optimal model to predict the growth of GDP by using Autoregressive Integrated Moving Average Model (ARIMA). The ARIMA model was performed for 93 years from 1929 to 2022 of Gross Domestic Product, Billions of Dollars, Annually from Federal Reserve Economic Data (https://fred.stlouisfed.org). The researcher conclude that the estimated model of the first order difference for the logarithm of GDP (DLGDP) series is ARIMA (1,1,1) with coefficients: C = 0.057064, AR (1) = 0.489046 & MA (1) = 0.265583 where S.E. of regression equals 0.051529, R-squared value is about 0.412974, Durbin-Waston statistic (1.961008) and the probability of F-statistic equals (0.000000), which gives the forecast value 0.10436 of LGDP in 2022, while the actual value equals 0.8818 with very low relative error 1.617%, therefore, the forecast value is close to the actual value and indicates that the ARIMA (1,1,1) model has a good fitting effect.

Keywords: Forecasting, GDP, Autoregressive integrated moving average (ARIMA), and Model.

INTRODUCTION:

GDP refers to Gross Domestic Product and is a criterion measure of the value added. It is created through the production of goods and services in certain country over a certain period of time, In this way, it also measures the yields and earnings which was gained from the production, as well as measuring the total amount spent on the final goods and services. The Gross Domestic Product (GDP) is the sum of all the final goods and services produced in the country within a specified period of time (usually one year).

The overall monetary value of these goods and services is taken together, which gives us the GDP. Various methods are used for computing GDP, such as value added method, expenditure methods. GDP is one of most important economic indicators that reflects the nature of the economic activities, and is a tool of evaluating the economic performance at the same time it helps as well in predicting some indicators such as inflation and unemployment (Islam and Alam, 2019; R. Carter Hill, 2015).

Although, GDP is single most significant indicator that seizes the economic activity. However, it lacks to supply a suitable measure of people's financial living and well-being making it less appropriate. Hence, using alternative indicators maybe more adequate. This indicator is based on symbolic GDP (known as GDP at current prices or GDP in value) and is available in various measures such as, US dollars and US dollars per capita. It is not an excellent option for the comparisons because progressing in developments is not only caused are real evolution but also requires variation in prices and PPPs (S. Dutta, 2022). The best way to the comprehend a country's economy is by looking at its Gross Domestic Product (GDP), it measures the

country's total output, this includes everything produced by the public and all the companies in the country, it helps to follow economic fluctuations, the development of policies on the population and determining economic policies (S. Dutta, 2022).

MATERIALS AND METHODS:

The data of Gross Domestic Product, Billions of Dollars, Annual, (GDP) were obtained from Federal Reserve Economic Data (https://fred.stlouisfed.org). The MRIMA model was performed during 93 years from 1929 to 2022 by using Stationary test (Unit Root of Augmented Dickey-Fuller) which was performed on the GDP series, also autocorrelation and partial autocorrelation function graphs was performed to determine the laying of difference and the appropriate transformation should be used to be converted to stationary series.

The researcher will deter-mine the appropriate model of ARIMA (p, d, q), by selecting the model that have a larger significant coefficient and highest R-squared value with smallest values of Akai Info Criterion, Schwarz Criterion and SIGMASQ (G.E.P. Box, 2015; Gujarati, 2009; H.H. Fan, 2009).

The data were analyzed with Econometrics Views (EViews) Release 10.

RESULTS AND DISCUSSION:

The GDP Data During 1929 - 2022 is plotted in Fig. Below:

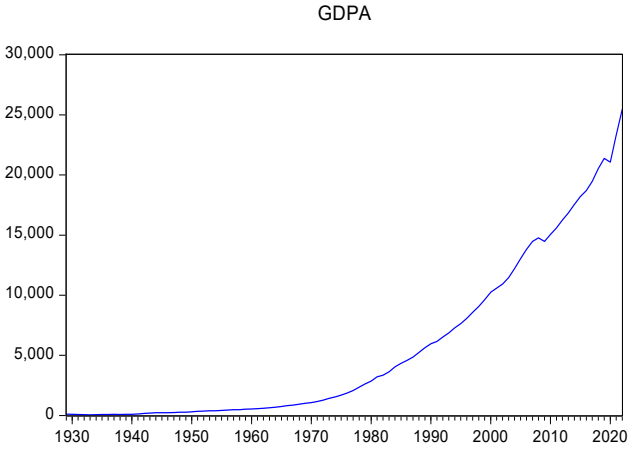


Fig. 1: The above figure shows that the GDP series has exponential shape.

Table 1: Augment Dickey-Fuller Unit Root Test on GDP.

Null Hypothesis: GDPA has a unit root				
Exogenous: Constant				
Lag Length: 2 (Automatic - based on SIC, maxlag=11)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			6.611147	1.0000
Test critical values:	1% level		-3.503879	
	5% level		-2.893589	
	10% level		-2.583931	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(GDPA)				
Method: Least Squares				
Date: 03/01/23 Time: 22:32				
Sample (adjusted): 1932 2022				
Included observations: 91 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDPA(-1)	0.056835	0.008597	6.611147	0.0000
D(GDPA(-1))	0.049686	0.121919	0.407537	0.6846
D(GDPA(-2))	-0.469731	0.155390	-3.022922	0.0033
C	62.87125	36.91042	1.703347	0.0921
R-squared	0.596336	Mean dependent var		278.9788
Adjusted R-squared	0.582417	S.D. dependent var		396.0853
S.E. of regression	255.9530	Akaike info criterion		13.97083
Sum squared resid	5699540.	Schwarz criterion		14.08119
Log likelihood	-631.6726	Hannan-Quinn criter.		14.01535
F-statistic	42.84197	Durbin-Watson stat		1.858602
Prob (F-statistic)	0.000000			

Table 1 shows that the Augment Dickey-Fuller statistic is 6.611147 with P value (1.0000) is not statistically significant value at 1%, 5%, 10% level, so; we cannot reject the null hypothesis; that GDPA has a unit root, and we conclude that the series of GDP is no stationary. As in **Fig. 1** the original series has exponential shape, so we should try to eliminate its nonstationary by using the logarithm of the GDP.

Table 2 The LGDP Data During 1929 - 2022 is plotted in **Fig. 2**. According to **Fig. 2** and **Table 2**, the results show that the Augment Dickey-Fuller statistic of LGDP is -0.907100 with P value (0.7820) is not statistically significant value at 1%, 5%, 10% level, so; we cannot reject the null hypothesis; that LGDP has a unit root, and we conclude that the series of LGDP is still non-stationary.

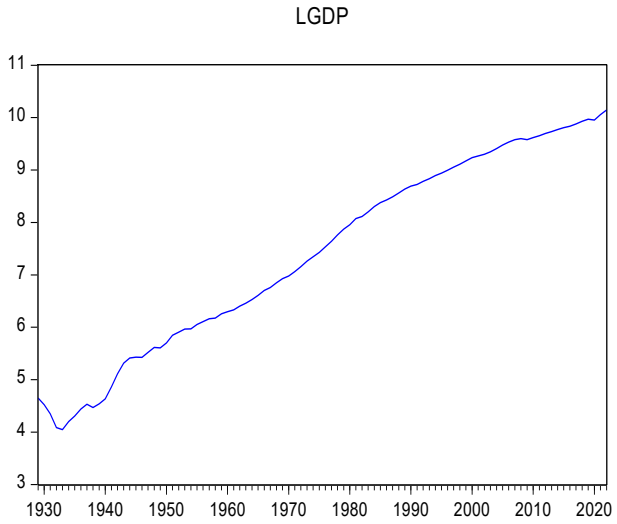


Fig. 2: The LGDP Data During 1929 - 2022 is plotted in figure.

Table 2: Augment Dickey-Fuller Unit Root Test on LGDP.

Null Hypothesis: LGDP has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic - based on SIC, maxlag=11)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-0.907100	0.7820
Test critical values:	1% level		-3.503049	
	5% level		-2.893230	
	10% level		-2.583740	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LGDP)				
Method: Least Squares				
Date: 03/02/23 Time: 00:23				
Sample (adjusted): 1931 2022				
Included observations: 92 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP(-1)	-0.002564	0.002827	-0.907100	0.3668
D(LGDP(-1))	0.601413	0.078590	7.652515	0.0000
C	0.044738	0.022013	2.032317	0.0451
R-squared	0.399758	Mean dependent var		0.061103
Adjusted R-squared	0.386269	S.D. dependent var		0.063580
S.E. of regression	0.049809	Akaike info criterion		-3.129175
Sum squared resid	0.220804	Schwarz criterion		-3.046943
Log likelihood	146.9420	Hannan-Quinn criter.		-3.095985
F-statistic	29.63672	Durbin-Watson stat		1.835399
Prob (F-statistic)	0.000000			

Further, the first order difference is performed and the D (LGDP) series is obtained as in the following table:

Table 3: Augment Dickey-Fuller Unit Root Test on D (LGDP).

Null Hypothesis: D (LGDP) has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=11)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-5.084361	0.0000

Test critical values:	1% level		-3.503049	
	5% level		-2.893230	
	10% level		-2.583740	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D (LGDP,2)				
Method: Least Squares				
Date: 03/02/23 Time: 00:15				
Sample (adjusted): 1931 2022				
Included observations: 92 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LGDP(-1))	-0.399174	0.078510	-5.084361	0.0000
C	0.025791	0.006943	3.714665	0.0004
R-squared	0.223138	Mean dependent var		0.002330
Adjusted R-squared	0.214506	S.D. dependent var		0.056145
S.E. of regression	0.049760	Akaike info criterion		-3.141711
Sum squared resid	0.222845	Schwarz criterion		-3.086890
Log likelihood	146.5187	Hannan-Quinn criter.		-3.119585
F-statistic	25.85073	Durbin-Watson stat		1.823067
Prob (F-statistic)	0.000002			

The Augment Dickey-Fuller statistic of D (LGDP) is (-5.084361) with P value (0.0000) is statistically significant value at 1%, 5%, 10% level, so; we reject the null hypothesis; that D (LGDP) has a unit root, and we conclude that the series of D (LGDP) is stationary. The autocorrelation and the partial correlation function graphs of D (LGDP) series are plotted in the figure. In the below table the autocorrelation of the D (LGDP) series is significantly non zero when the lag order is q=1 or q=2, as it is basically in confidence band when the lag order is greater than 2. The same goes as well for partial auto-correlation where we take p=1 or p=2, hence the final order with 0, 1, 2 in autoregressive moving average pre-estimation is performed on the sample series.

Table 4: Correlogram of D (LGDP).

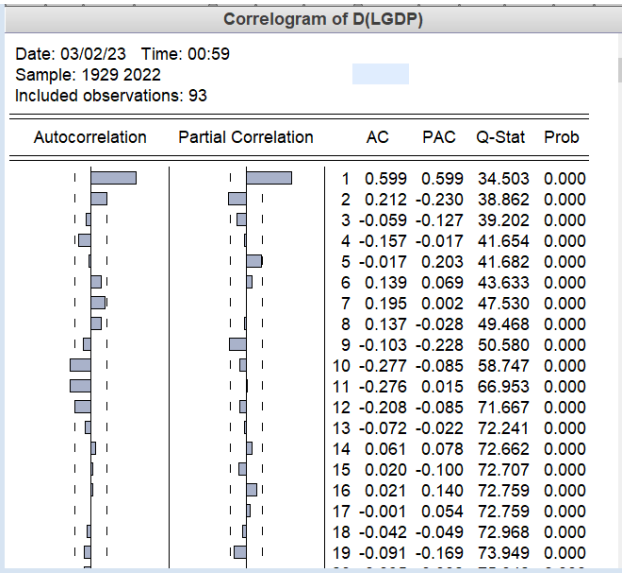


Table 5: The Results of Estimated Model of ARMA (p, q).

(p, q)	R-squared	F-statistic	Prob (F-statistc)	Akaike info criterion	Schwarz criterion	Durbin-Watson stat	SIGMASQ
(0,1)	0.357125	24.99804	0.000000	-2.976803	-2.895107	1.647405	0.002783
(1,0)	0.393897	29.24485	0.000000	-3.034934	-2.953238	1.744855	0.002624
(1,1)	**0.412974**	**20.87056**	**0.000000**	**-3.044942**	**-2.936013**	**1.961008**	**0.002541**
(2,0)	0.058324	2.787136	0.066922	-2.598606	-2.516909	0.879439	0.004076
(0,2)	0.100916	5.050953	0.008337	-2.642383	-2.560686	0.989980	0.003892
(1,2)	0.401556	19.90631	0.000000	-3.025747	-2.916818	1.846568	0.002590
(2,1)	0.408514	20.48950	0.000000	-3.037393	-2.928464	1.874683	0.002560
(2,2)	0.113967	3.815924	0.012689	-2.635430	-2.526501	0.952259	0.003835

The above table shows the results of ARMA (p, q) model for different parameters. To select the optimal model, we should compare the results of the significant for parameters: R-squared, Akai Info. Criterion, Schwarz Criterion and SIGMASQ. According to the above table, we found that the models ARMA

(2, 2), ARMA (2, 0), ARMA (0, 2) have the lowest R-squared. Once finding it, select the model that has a larger significant coefficient and the highest R-squared value with the smallest values of Akai Info. Criterion, Schwarz Criterion and SIGMASQ. The ARMA (2, 2) model didn't pass the parameter significance test, in addition, it has a low R-squared value. Models: ARMA (2, 0), ARMA (0, 2) have the lowest R-squared values. Hence, we should compare between the models that have the highest value of R-squared: ARMA (1, 1), ARMA (2, 1). Therefore, we conclude that ARMA (1, 1) model is the best which satisfied the parameter significance test with the highest R-squared value and the lowest values of the Akai Info. Criterion, Schwarz Criterion and SIGMA-SQ.

Table 6: The Estimated Results of ARIMA (1, 1, 1) Model.

Dependent Variable: D (LGDP)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 03/02/23 Time: 21:42				
Sample: 1930 2022				
Included observations: 93				
Convergence achieved after 62 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.057064	0.013867	4.115191	0.0001
AR(1)	0.489046	0.078050	6.265838	0.0000
MA(1)	0.265583	0.102370	2.594358	0.0111
SIGMASQ	0.002541	0.000273	9.303313	0.0000
R-squared	0.412974	Mean dependent var		0.059089
Adjusted R-squared	0.393187	S.D. dependent var		0.066149
S.E. of regression	0.051529	Akaike info criterion		-3.044942
Sum squared resid	0.236314	Schwarz criterion		-2.936013
Log likelihood	145.5898	Hannan-Quinn criter.		-3.000960
F-statistic	20.87056	Durbin-Watson stat		1.961008
Prob (F-statistic)	0.000000			
Inverted AR Roots	.49			
Inverted MA Roots	-.27			

So, the estimated model of the DLGDP series ARIMA (1, 1, 1) is:

$LGDP = 0.057064 + 0.265583x_t + 0.489046x_{t-1}$
DLGDP=0.057064+0.489046AR (1) + 0.265583MA (1), with S.E. of regression equals 0.051529

The R-squared value is about 0.412974 which is statistically significant value. Durbin-Waston statistic (1.961008) is found to be 2, so there is no first-order autocorrelation either positive or negative. Also it's more than R-squared, which means that this model is not spurious. The probability of F-statistic equals 0.000000 which is statistically significant at level 5% meaning that the explanatory variables are jointly significant to D (LGDP).

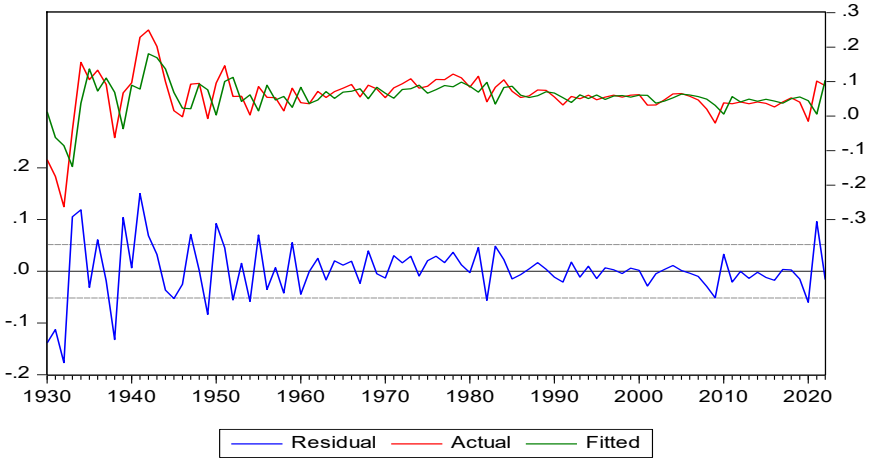
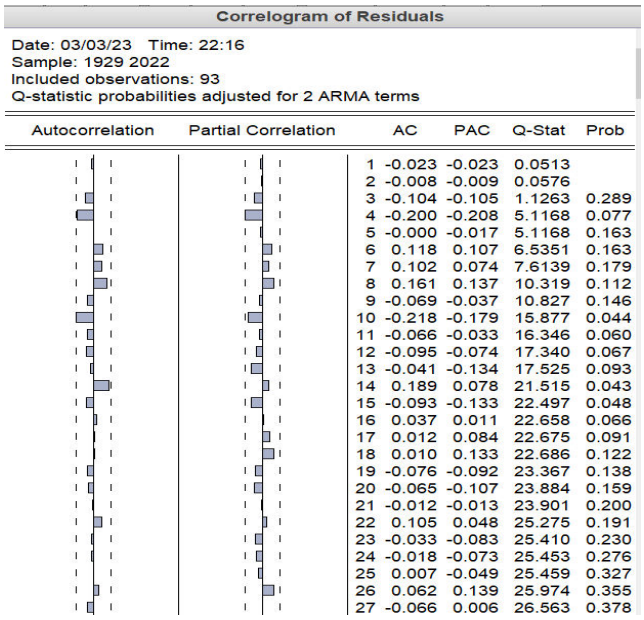


Fig. 3: The actual & fitted series are passing closely and the D (LGDP).

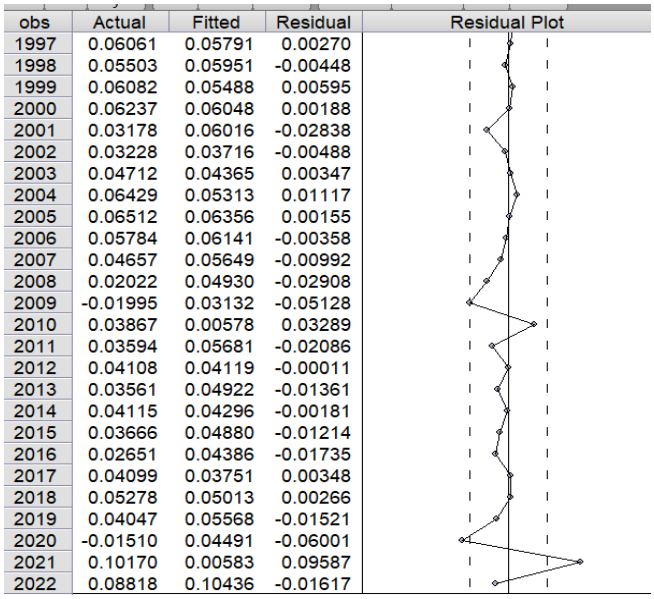
As shown in **Fig. 3**, the actual & fitted series are passing closely and the D (LGDP) has been fore-casted and is passing throw 50% confidence interval; so the forecasting of D (LGDP) is significant and the ability of forecasting the model is satisfactory.

Table 7: Correlogram of Residuals.



The autocorrelation and the partial autocorrelation function graphs of residual series in the above figure, show that the residual is white noise which indicates that the model is valid. Firstly, we do the forecast inside the sample to check the power of the model in forecasting (Hossain *et al*., 2020).

Table 8: Forecast inside the Sample.



The above graph shows that the forecast value of LGDP in 2022 is 0.10436 while the actual value is equal to 0.8818 with a very low relative error 1. 617%, so the forecasted value is close to the actual value which indicates that the model has a good fitting effect. Secondly, by using Box-Jenkies for forecasting GDP during the upcoming five years from 2023 to 2027, the results are shown in the table below:

Table 9: Forecast outside the Sample:

Year	Forecasting of LGDP values	Forecasting of GDP values
2023	9.600642	14774.26
2024	9.657706	15641.86
2025	9.714771	16560.42
2026	9.771835	17532.91
2027	9.828900	18562.52

CONCLUSION:

Autoregressive Integrated Moving Average Model ARIMA (1, 1, 1) is acceptable to the predictive purpose of forecasting the Gross domestic product (GDP):

$$LGDP = 0.057064 + 0.265583x_t + 0.489046x_{t-1}$$

With S.E. of regression equals 0.051529, R-squared value is about 0.412974, Durbin-Waston statistic (1.961008) and the probability of F-statistic equals (0.000000).

ACKNOWLEDGEMENT:

With due respect and obeisance, I would like to express my utmost gratitude, indebtedness and appreciation to my husband Dr. Husham Hamadto for his immaterial support.

CONFLICTS OF INTEREST:

All authors declare no conflict of interest with the contents of this research work.

REFERENCES:

1) G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung. (2015). Time Series Analyses: Forecasting and Control. 5th ed. *Wiley, New York.* https://www.wiley.com/en-us/Time+Series+Analysis:+Forecasting+and+Control,+5th+Edition-p-9781118675021
2) Gujarati, D. N. and Porter, D.C. (2009). Basic Econometrics, Fifth Edition, *McGraw-Hill Companies. United State Military Academy, West Point and University of Southern California*, USA.
3) Hossain MF, Nandi DC, and Uddin KMK. (2020). Prediction of banking sectors financial data of Dhaka stock exchange using Autoregressive Integrated Moving Average (ARI-

MA) approach, *Int. J. Mat. Math. Sci*., **2**(4), 64-70. https://doi.org/10.34104/ijmms.020.064070

4) H.H. Fan, L.Y. Zhang. E. (2009).Views Statistical Analysis and Application. *China Machine Press, Beijing*, https://www.researchgate.net/publication/329812

5) Islam MT, and Alam MJ. (2019). The relationship between informal economy and GDP growth: a study on South-Asian developing countries. *Can. J. Bus. Inf. Stud*., **1**(5), 01-09. https://doi.org/10.34104/cjbis.019.0109

6) R. Carter Hill, William E. Griffiths and Guay C. Lim (2011). Principle of Econometrics, Fourth Edition. *John Wiley & Sons, Inc, USA*. http://zalamsyah.staff.unja.ac.id/wp-content/-uploads/sites/286/2019/11/7-Principles-of-Econometrics-4th-Ed.-R.Carter-Hill-et.al_.-1.pdf

7) S. Dutta, (2022). "Introductory- Economics". *New Age International (P) Limited Publishers, New Delhi*. http://150.129.131.246:8090/jspui/bitstream/123456789/28/8/introductory%20economics.pdf

Citation: Sultan MA. (2023). Forecasting the GDP in the United States by Using ARIMA Model, *Can. J. Bus. Inf. Stud.*, **5**(3), 63-69. https://doi.org/10.34104/cjbis.023.063069