ISSN 2709-4170 (Print)
Transactions in Mathematical and Computational Sciences
Vol. 1(1)(2021) pp. 75–84

MODELING AND FORECASTING OF GROSS DOMESTIC PRODUCT PERCENTAGE SHARE OF EDUCATION SECTOR: A STATISTICAL STUDY IN PAKISTAN

MUHAMMAD WASIM AMIR[1,*], AYESHA BIBI[1], NOUREEN AKHTAR[2], ZEESHAN RAZA[3]
[1]Calaza System of Education, Miani, Sargodha, Pakistan
[2]Department of Statistics, University of Sargodha, Sargodha, Pakistan
[3]Govt. Elementary School No.1 Lalian, Chaniot, Pakistan

Abstract. Education is a very important medium for obtaining skills and knowledge. The world is continually changing so it is significant to teach and bring up those peoples who can understand and solve modern social problems. Forecasting techniques are necessary to find out the future trends that are meet the upcoming problems. For the desired purpose, different time series models are applied and selected the best one for better-forecasted values of the gross domestic product (GDP) percentage share on education. The best model is determined based on the lowest value of the Akaike information criterion (AIC) and Bayesian information criterion. Moreover, the quality and predictability power of the selected models are measured based on the minimum value of the mean error, root mean square error, mean percentage error, and mean percentage absolute error. The AIC criteria showed that the best time series model is the autoregressive integrated moving average (ARIMA) (2, 1, 1) to forecast the (GDP) percentage share on the education of Pakistan. Furthermore, the assumptions of the fitted model such as independence, no heteroscedasticity, and normality are also evaluated. The selected model showed that the (GDP) percentage share on education is decreasing slowly, which is not good for the education sector and development of the country.

Keywords: gross domestic product (GDP), education, time series modeling, ARIMA .

1. INTRODUCTION

Education plays a dynamic role in shaping successful people. It provides us the opportunity to become a productive member of a civilized society by attaining all the essential skills such as rules, regulation, basic norms, and the values of society. We learn how to

*Correspondence: Email of Corresponding author
Received: 20.01.2021 Accepted: 01.03.2021 Published: 04.03.2021

become an integrated personality and maintain the perpetuation of our culture. Furthermore, high-quality education permits us to lead a successful life, knowledge, intelligence and brings positive changes in our life. The importance of education also highlights the Holy Quran and Hadith. In the Holy Quran, the almighty Allah says "Are those who know equal to those who do not know? Only they will remember (who are) people of understanding". The Prophet (PBUH) says "when a man dies, his acts come to an end, but three, recurring charity, or knowledge (by which people benefit), or a pious son, who prays for him (the deceased)". Education has a multidimensional effect on society and plays a significant role in the development of countries. More, it is an instrument providing equal opportunities and higher income levels amongst the entire population. The literacy rate of the developed country also high from underdeveloped countries because they spend more on the education sector than in other countries. In the world, Pakistan has the 2nd highlight number of out-of-school children (OOSC) with an estimated 44% of the total population age 5-16 who are not enrolled in school. At the age of 5-9 years, 5 million children are not attending schools, and the OOSC number doubles after completing primary education. In provinces Sindh and Balochistan, gender inequality, geography, and socio-economic status are significant. 78% of girls are out of school in Balochistan and 52% of the poorest children along with 58% of girls are not enrolled in school (UNICEF, [1]). Pakistan ranks 50th about the percentage share of GDP on education in the world with a 2.9% share of the total GDP. The most contributed countries of percentage share in the education sector are Costa Rice, Belize, Bhutan, Oman, Moldova, Guyana with 7.40%, 7.38%, 7.03%, 6.85%,6.68%, and 6.34% respectively (The Global Economy, [2]). It is observed that Pakistan has a low GDP contribution as compared to other countries. The reasons for the low GDP share on education include political issues, border issues, improper policy, low literacy rate, etc. Over time, the population of Pakistan is increasing along with their needs, such as the number of schools and universities, well-educated staff, and developed research labs. While income consumption in the education sector is not enough to achieve all these needs. Every year a lot of students have not completed their education due to income problems. So it is necessary to plan in increasing the GDP% share on the education of Pakistan to fulfill the nation's necessities of education. Forecasting techniques are necessary to find out the trends for future planning of the country. Different studies highlighting the importance of forecasting are presented in the literature. Haider and Hanif [3] forecasted the inflation rate of Pakistan by using an artificial neural network model. Zakai [4] used the Box-Jenkins methodology to forecast the GDP of Pakistan. Itoh *et al*, [5] used a Bayesian network to

forecasting the future academic level. Perara [6] applied a Box-Jenkins type model to forecast foreign direct investment in SAARC countries. Nie *et al.* [7] forecasted the college student's future career choice base on big data. Wasomba *et al.* [8] applied an autoregressive integrated moving average model to forecast the GDP of Kenya and disclose that the ARIMA (2, 2, 2) model is best for future forecasting of Kenyan GDP. Uwimana *et al.* [9] forecasted the 20 Africa's largest economies and disclosed that the average speed of the economy of Africa will be 5.52% in the future. Miah *et al.* [10] forecasted the GDP of Bangladesh and concluded that the ARIMA (1, 2, 1) is an adequate model for future forecasting. Abonazel and Abd-Elftah [11] used the Box-Jenkins methodology to forecast the GDP of Egypt. For furthermore studies about time series forecasting see (Guo *et al.* [12], Green and Armstrong [13], Moe *et al.* [14], Praveen *et al.* [15], Akpensuen *et al.* [16]). From the literature, we have found that no study still has been done to forecasts the GDP% share on the education sector of Pakistan. So the objective of this study is to forecast the GDP% share on the education sector of Pakistan based on an appropriate time series model (TSM) that fulfills the required model assumption, which may improve the GDP% share on the education sector of Pakistan that meet the requirement of Pakistan nation. The rest of the study is following as: Section 2 contains the material and method of the time series model. Section 3 provides the results and discussion. The concluding remarks and recommendations are presented in section 4.

2. Material and Method

For model building and forecasting, secondary data of percentage share of Gross Domestic Product (GDP) on Education have been used in this article from the period 1971-2017. There are a variety of TSMs available in the literature to forecasting the time series data but this article is considered Box-Jenkin's (1976) methodology. In Box-Jenkins methodology the Autoregressive Integrated Moving Average (ARIMA) is a more accumulatively class of model for a time series forecasting. The ARIMA model is symbolized by ARIMA (p,d,q), where "p" is the order of autoregressive identified by autocorrelation function (ACF), "d" denoted the order of difference for the time series to become stationary, and "q" is the order of moving average identified by partial autocorrelation function (PACF). For further details about these types of time series models and their parameter estimation check (Amin *et al.* [17]; Chatfield, [18]; Tsay, [19]).

2.1. **Model Specification Criterion.**

The key issue in time series forecasting is to identify the best model. The TSM is identified based on some information criteria's which include

Schwarz Bayesian information criteria (SBIC) and Akaike's information criteria (AIC). Akaike's (1973) presented AIC criteria for model specification and mathematically defined as:

$$AIC = -2\log(maximumlikelihood) + 2k \tag{1}$$

where k = p+q+1 (if model includes intercept) otherwise $k = p+q$. The TSM model which displays the minimum value of AIC is preferred over other fitted TSM models (Tsay, [19]). The other model identification criterion is SBIC and is computed as:

$$SBIC = -2log(maximumlikelihood) + 2k\log(n) \tag{2}$$

It also concerns the minimum value for a model to be a good fit for the data (Tsay, [19]).

3. Time Series Model Diagnostics

The best-fitted model requires some assumptions for better estimates of the forecasted values such as independence, normality, no autocorrelation of residuals. The assumption of residual normality is examined by utilizing a Periodogram and normal probability plot (Amin *et al.* [17]). The run test is employed for checking the autocorrelation in the residuals as recommended by (Gujarati [20]). ACF, PACF, and Box- Pierce tests are also used to detect the autocorrelation in the data (Amin *et al.* 2014).

4. Results and Discussion

In this research, various time series models are fitted on the GDP% share of the education time series data of Pakistan. The purpose of fitting multiple TSMs on this data is to gain reliable forecasts on the premises of statistical measures. For the desired purpose, different TSMs are fitted and the results of these models are reported in Table 1. We also accumulate the results of the model selection and validity criteria's in Table 1. Based on model selection criteria, we find model "L" i.e. ARIMA (2, 1, 1) has minimum AIC value, and use this model for future forecasting of government expenditure on education. Table 1, also summarizes the results of the different run tests on the residuals to confirm the superiority of the selected model related to the independence assumption of the model. It is noticed that the selected model passes all the tests and adequate for the data.

4.1. **Time Series Models Comparison.** (A) Random walk, (B) Random walk with drift = 0.0240476, (C) Constant mean = 2.28251, (D) Linear trend = 2.06492 + 0.00906615t, (E) Exponential trend = exp(0.712889 + 0.00353414 t), (F) S-curve trend = exp(0.840908 +

$-0.457503/t$), (G) Simple moving average of 2 terms, (H) Simple exponential smoothing(ES) with $\alpha = 0.9999$.

Table 1. Model validity and selection criteria's for a percentage share of GDP on education

Model	AIC	HQC	SBIC	RMSE	MAE	MAPE	ME	MPE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	-2.57	-2.57	-2.57	0.28	0.21	10.56	0.02	-0.16	OK	OK	***	OK	**
(B)	-2.51	-2.50	-2.47	0.28	0.21	10.61	0.00	-1.28	OK	OK	***	OK	**
(C)	-1.43	-1.42	-1.40	0.48	0.37	20.41	0.00	-6.76	**	***	***	OK	*
(D)	-1.44	-1.41	-1.36	0.47	0.33	18.72	0.00	-6.56	*	***	***	OK	**
(E)	-1.42	-1.39	-1.34	0.47	0.35	18.87	0.06	-3.81	**	***	***	OK	**
(F)	-1.49	-1.46	-1.41	0.46	0.33	18.12	0.06	-3.66	*	***	***	OK	***
(G)	-1.94	-1.93	-1.90	0.37	0.26	14.02	0.04	-1.00	***	**	***	OK	**
(H)	-2.53	-2.51	-2.49	0.28	0.20	10.33	0.02	-0.16	OK	OK	***	OK	**
(I)	-2.52	-2.50	-2.48	0.28	0.22	10.70	0.00	1.57	OK	OK	OK	OK	OK
(J)	-2.41	-2.38	-2.33	0.29	0.21	10.85	-0.05	-3.62	OK	OK	***	OK	**
(K)	-2.18	-2.17	-2.14	0.33	0.25	12.51	0.00	2.16	OK	OK	OK	OK	*
(L)	-2.94	-2.89	-2.82	0.22	0.17	8.02	0.00	-0.73	OK	OK	OK	OK	OK
(M)	-2.84	-2.81	-2.76	0.23	0.19	8.87	0.01	0.78	OK	OK	OK	OK	OK
(N)	-2.78	-2.73	-2.66	0.23	0.18	8.81	0.01	0.74	OK	*	OK	OK	OK
(O)	-2.77	-2.76	-2.73	0.24	0.20	9.51	0.02	0.76	OK	OK	OK	OK	*
(P)	-2.75	-2.69	-2.59	0.23	0.18	8.30	0.01	0.47	OK	OK	OK	OK	OK

(I) Brown's linear ES with $\alpha = 0.9341$, (J) Holt's linear ES with $\alpha = 0.9999$ and $\beta = 0.0228$, (K) Brown's quadratic ES with alpha = 0.7052, (L) ARIMA(2,1,1), (M) ARIMA(0,1,2), (N) ARIMA(1,1,2), (O) ARIMA(1,1,0), (P) ARIMA(2,1,2).

Table 2. ARIMA Model Co-efficient Summary

Parameter	Estimate	Stnd. Error	t-cal	P-value
AR(1)	1.35845	0.118758	11.4388	0.000000
AR(2)	-0.603184	0.118319	-5.09793	0.000008
MA(1)	0.968382	0.0196459	49.2919	0.000000

We have fitted the numerous time series models on the GDP% share on education. On the evidence of Table 1 and Table 2, our best-fitted model fond to be ARIMA (2, 1, 1) for GDP% share on education is given by

$$\hat{y}_t = 1.35845\hat{y}_{t-1} - 0.0603184\hat{y}_{t-2} + 0.968382\hat{e}_{t-1} \tag{3}$$

where $\hat{y}_t$ is the GDP% share on education for time t years,
$\hat{y}_{t-1}$ is the forecasted GDP% share on education at lag one year,
$\hat{y}_{t-2}$ is the forecasted GDP% share on education at lag two year,
$\hat{e}_{t-1}$ is residual at lag one year.

4.2. **Testing selected model assumptions.**

Each model is fitted on certain assumptions. A model that fulfills its assumptions can be considered as the most accurate and reliable model for forecasting a particular time series variable. In the current study, the fitted model requires assumptions of normality, no heteroscedastic, and no autocorrelation of residuals.

Table 3. Tests for independence and autocorrelation

Test	Test Statistic	Pr-value	Expected Number of Runs
Runs overhead and underneath the median	1.64019	0.100966	24
Runs up and down	0.773	0.439495	30
Box-Pierce Test	13.45	0.337111	33

It is observed that the fitted model residuals are uncorrelated and one does not depend on the other (cf. Table 3). Three tests are applied to identify whether they show irregular patterns are not. From Table 3, it is noticed that the first test indicates the sequence of runs overhead and underneath the median. Results disclosed that runs are equal to 30 with its expected value is 24. Since the Pr-value for this test is greater than the level of margin 0.05. So, the residual of the selected model is random. The second test is related to the runs down and up. The probable number of runs for such a test is equal to 33 with the expected number of runs is equal to 30.33. From the provided evidence, the Pr-value for runs down and up median test is equivalent or greater than to the level of significance i.e. 0.05, so we conclude that the residuals are random. The Box-Pierce test is built on the sum of squares of the initial 15 autocorrelation coefficients. Since the Pr-value for this test is greater than a 5% margin of error, so the selected model residuals are uncorrelated. The assumption of normality is also justified by the normal probability plot as shown in Figure 1. From Figure 1, it is noticed that the residuals of ARIMA (2, 1, 1) are normally distributed.

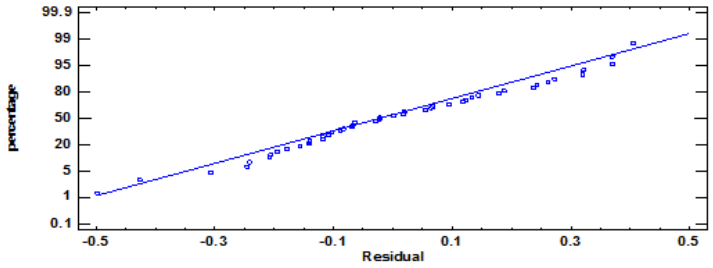

FIGURE 1. Residuals normal probability plot of government expenditure on education (%GDP)

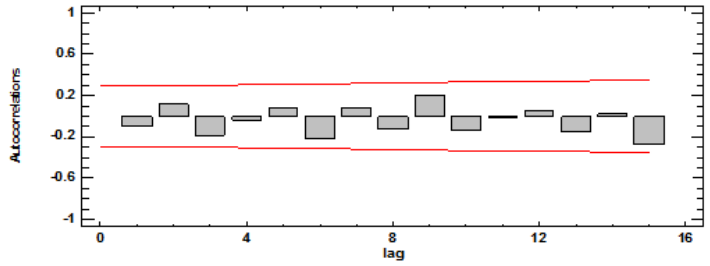

FIGURE 2. Residuals autocorrelation plot of government expenditure on education (%GDP)

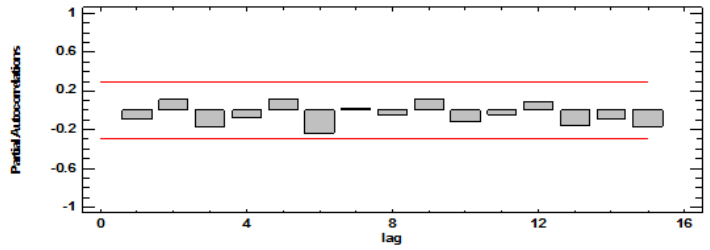

FIGURE 3. PACF plot of government expenditure on education(%GDP)

From Figure 2 and Figure 3, we noticed that the residuals of the fitted model are also uncorrelated. These results indicated that ARIMA (2, 1, 1) is the most suitable as well as a more accurate time series model for forecasting the government expenditure on education

as this model meets all its required assumptions. Forecasted values of government expenditure on education based on ARIMA (2, 1, 1) are shown in Table 4 and Figure 4. From Table 4, we observed that the government expenditure on education of Pakistan would become 2.3943% of total GDP in 2030. These forecasted values are based on statistical models, so the best-fitted model provides more accurate forecasting under the assumption that the environmental conditions remain the same.

Table 4: GDP share on education forecasts for model ARIMA (2, 1, 1) in percentage

Year	2020	2022	2025	2030
Forecasts	2.62205	2.36794	2.2905	2.3943
lower	1.66462	1.2861	1.20307	1.3020
upper	3.5794	3.4496	3.37792	3.4865

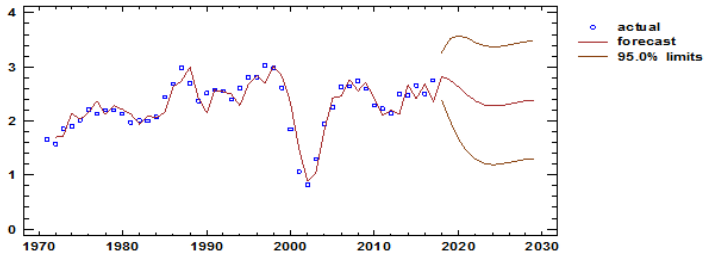

FIGURE 4. Forecasted plot of GDP% share on education for model ARIMA (2, 1, 1)

5. CONCLUSION

Education is the basic need of any country all over the world. Over time the requirements of schools, universities, teachers, and research centers are increases. It is necessary to plan in increasing the GDP% share on the education sector of Pakistan to fulfill the nation's necessities of education. So, forecasting is the fundamental tool to alarm about the nation's needs in advance. In this study, we fitted the time series model on the GDP% share on the education sector of Pakistan on the premises of historical data i.e. 1971-2017. For the selection of the best model, the AIC and SBC model selection criteria are considered. On the model selection criteria, we find the finest model for forecasting GDP% share in the education sector is the ARIMA (2, 1, 1). On the evidence of this model, we have originated that the GDP% share in the education sector would become 2.3943% of total

GDP in 2030. The study presents an insight to national policymakers to increase the budget of the education sector. The study also recommends the government of Pakistan to build new universities, schools, colleges and highlight the importance of education.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewers for their helpful, valuable comments and suggestions in the improvement of this manuscript.

AUTHOR'S CONTRIBUTIONS

All authors equally contributed to this work. All authors read and approved the final manuscript.

REFERENCES

[1] United Nations International Children's Emergency Fund (2019). UNICEF, Pakistan, Education. https://www.unicef.org/pakistan/education.

[2] The Global Economy (2017). Business and economic data. Pakistan: Education spending, percent of GDP. https://www.theglobaleconomy.com/Pakistan/Education spending/.

[3] Haider, A., & Hanif, M. N. (2009). Inflation forecasting in Pakistan using artificial neural networks. *Pakistan economic and social review*, **47**(1), 123-138.

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

[5] Itoh, Y., Itoh, H., & Funahashi, K (2015). Forecasting future students' academic level and analyzing students' feature using schooling logs. *In 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, **2015**, 288-291.

[6] Perera, P. (2015). Modeling and Forecasting Foreign Direct Investment (FDI) into SAARC for the Period of 2013-2037 with ARIMA. *International Journal of Business and Social Science*, **69**(2), 264-272.

[7] Nie, M., Yang, L., Ding, B., Xia, H., Xu, H., & Lian, D. (2016). Forecasting career choice for college students based on campus big data. *In Asia-Pacific Web Conference*, 359-370, Springer, Cham.

[8] Wabomba, M. S., Mutwiri, M. P., & Fredrick, M. (2016). Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models. *Science Journal of Applied Mathematics and Statistics*, **4**(2), 64-73.

[9] Uwimana, A., Xiuchun, B., & Shuguang, Z. (2018). Modeling and forecasting Africa's GDP with time series models. *International Journal of Scientific and Research Publications*, **8**(4), 41-46.

[10] Miah, M. M., Tabassum, M., & Rana, M. S. (2019). Modelling and Forecasting of GDP in Bangladesh: An ARIMA Approach. *Journal of Mechanics of Continua and Mathematical Science*, **14**(3), 150-166.

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

[12] Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L (2015). Predicting students performance in educational data mining. *In 2015 International Symposium on Educational Technology (ISET)*, **2015**, 125-128.

[13] Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, **68**(8), 1678-1685.

[14] Moe, Z. H., San, T., Tin, H. M., Hlaing, N. Y., & Tin, M. M (2018). Evaluation for teacher's ability and forecasting student's career based on big data. *In International Conference on Big Data Analysis and Deep Learning Applications*, **2018**, 20-27. Springer, Singapore

[15] Praveen, B., Talukdar, S., Mahato, S., Mondal, J., Sharma, P., Islam, A. R. M. T., and Rahman, A. (2020). Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. *Scientific reports*, **10**(1), 1-21.

[16] Akpensuen, S. H., Edeghagba, E. E., Alhaji, A. G., & Joel, S. (2019). Time Series ARIMA Model for Predicting Nigeria Net Foreign Direct Investment (FDI). *World Scientific News*, **128**(2), 348-362.

[17] Amin, M., Amanullah, M., & Akbar, A. (2014). Time series modeling for forecasting wheat production of Pakistan. *The Journal of Animal and Plant Sciences*, **24**(5), 1444-1451.

[18] Chatfield, C. (1995). *The Analysis of Time Series an Introduction*. 5th Ed. Chapman and Hall.

[19] Tsay, R. S. (2002). *Analysis of Financial Time Series: Financial Econometrics.* John Wiley and Sons Inc.

[20] Gujarati, D. N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.