Vector Autoregressive and Vector Error Correction Modelling of Impacts of Health and Education Sectors on Nigerian Economy

I. Ahmed[1]; T. D. Okpe[2]; M. O. Adenomon[3]

[1,3]Department of Statistics,
Nasarawa State University Keffi, Nigeria.

[2]Department of Mathematics and Statistics,
Federal Polytechnic, Nasarawa, Nigeria.
E-mail: ibrahimloko@gmail.com[1]; thankgodokpe@gmail.com[2]

***Abstract* — The aim of this research is to determine the relationship among GDP, Education and Health Sector of the economy by using Vector Autoregressive (VAR) and Vector Error Correction Model (VECM) models. Attempt was also made determine the behavior of GDP if there is a shock in the contribution of education and health to the Nigerian economy. The VECM and VAR models are employed on data series that are not stationary and have cointegration (long term) relationship. The VECM model can also be used to examine the movement in one variable to give a response regarding the shock produce by another variable through the graph of Impulse Response Function (IRF). Based on the data on GDP, Health and Education Sector in Nigeria over the periods from 1981 to 2018, we determined that the best VECM model is the (VECM(1)). Based on the graph of the IRF we have established that the response of GDP towards the shocks in the contribution of the education and health sector of the economy fluctuate and temporary over time. the proportion of shock towards the changed in GDP provides a negative response while proportion of shock in the change in GDP did not have a high contribution (effect) upon the education and health sector. It has become pertinent that the government should increase the rate of infrastructural development and funding of these sectors and also design effective policy implementation in order to increase the quality of services provided by these sectors.**

Keywords - *Gross Domestic Product (GDP), Health Sector, Education Sector, VAR, VECM.*

I. Introduction

The main problem faced by governments is allocating scarce resources across competing activities and sectors. The choice between alternative investments such as investment in education versus investment in physical infrastructure depends on society's objectives, which are represented by governmental decisions, and on the analysis between costs of the investment versus the future benefit to be derived from that investment. Since, economists see education as an investment, therefore, it is important to estimate its contribution to economic growth and/or its rate of return. Education represents both consumption and investment items in an economy. Education is valued for its immediate as well as its future benefits. This means that the distribution of educational investment affects future income distribution, thus, equity plays an important role in educational investment decisions. Different societies give different weight between the objectives of efficiency and equity in defining an educational investment. In general, centrally planned economies placed a higher weight on equity grounds in defining their educational policy investment than capitalist economies [1]. Health services are provided by the private and public sectors. From private sector, there are non-governmental organization, private for-profit providers, community-based organization and religious and traditional care givers.

The health status in Nigeria is ranked low among other developing country in the same category. Life expectancy is put at 52 years in 2011(according to World Bank) and crude death rate, in that same year as 14%. It is estimated that 124 out of 1000 new births do not survive beyond age 5. Only 39.56% of male and 42.25% of female survive up to the age of 65 years.

There are close to 3 million adults (ages 15-49) living with HIV. While the estimated HIV/AIDS prevalence rate is 3.7 [7]. Nigeria has large stock of health workers that is comparable to that of Egypt and South Africa. Beside the contribution of education on national economic growth, it also plays significant role in reducing income inequality, research done by [2] concluded that educational achievement as well as human capital development would positively reduce income inequality. In general, there is a consensus among the researchers that education influenced economic growth by reducing poverty incidence, social imbalances as well as income equality. Moreover, it gives a positive impact to the poor and needy to improve their live.

Investigation was done on the causal direction and long run relationship between government health expenditure, poverty and health status, in Nigeria. The Granger causality test and Vector Error Correction Model (VECM) was employed by [2] in establishing a strong causal bi-directional relationship running between life expectancy and poverty in Nigeria. Their study also reports the existence of a long-run relationship between poverty and health status. However, they found a non- significant long run relationship between health status and government health expenditure. They conclude that policies that would improve health status should be such as would promote adult literacy level, reduce the poverty and income disparity since, increasing budgetary allocation to funding health sector alone without reducing poverty level, would not be sufficient to improve the health status of the country..

II. RESEACH METHODOLOGY

STATIONARY

Many data analyses of time series are based on the assumption that the time series is stationary. The process being stationary indicate that the mean, variance and autocorrelation functions are essentially constant and do not depend on time that is the first two moments are time invariant. By transformation we define a new series Z'_t as follows:

$$Z_t = \frac{Z_t^{\lambda} - 1}{\lambda}$$

where λ is a real number. Nnote that Z_t must not be negative. The successive changes in the series for all t, as follows:

$$Z_t = Z_t - Z_{t-1}$$

If the resulting series does not yet have a constant overall mean, we then compute the first differences of the first differences for all t. That is, denote the first differences of z_t as w_t*. Thus, the first differences of the w_t* series are;

$$w_t = w_t - w_{t-1} = (Z_t - Z_{t-1}) - (Z_t - Z_{t-2})$$

The resulting series is called the second differences of z_t.

COINTEGRATION

(i) Trace test

H_0 : There exist at most r eigen values which are positive.

H_1 : There exist more than r eigen values which are positive.

$$Tr(r) = -T \sum_{i=r+1}^{k} In(1 - \hat{\lambda}_i)$$

(ii) Test λ_{Max} whether there are $r \; or \; r + 1 \; vectors \; cointegration$

$$\lambda_{Max}(r, r+1) = -T \; In(1 - \lambda_1)$$

where:

$\hat{\lambda}_i$: The estimation of Eigen values

T : Number of observations.

k : Number of endogenous variables.

VECTOR AUTO REGRESSIVE (VAR)

Suppose that we measure three difference time series variables, say $y_{t,1}$, $y_{t,2}$, and $y_{t,3}$ VAR model for order 1, VAR(1) is as follows:

$$y_{t,1} = c_1 + \phi_1 \, y_{t-1,1} + \phi_2 \, y_{t-1,2} + \phi_3 \, y_{t-1,3} + \varepsilon_{t,1}$$
$$y_{t,2} = c_2 + \phi_2 \, y_{t-1,1} + \phi_2 \, y_{t-1,2} + \phi_3 \, y_{t-1,3} + \varepsilon_{t,2}$$
$$y_{t,3} = c_3 + \phi_3 \, y_{t-1,1} + \phi_2 \, y_{t-1,2} + \phi_3 \, y_{t-1,3} + \varepsilon_{t,3}$$

In general, model VAR(p) for m difference time series variable scan be defined as follows:

$$y_{t,i} = c_1 + \sum_{j=1}^{p} \phi_j y_{t-j,i}$$

$$y_t = c + \sum_{j=1}^{p} \phi_j y_{t-1} + \varepsilon$$

Where:

y_t : the element vector of $\mathbf{y}$ at time $\mathbf{t}$

ϕ_i : Matrix order n ×n which the elements are the coefficient of the vector y_{t-1} , for i=1,2,....p.

p: The length of lag

c: Vector intercept

ε_t: Random vector of shock.

VECTOR ERROR CORRECTION MODEL (VECM)

The VECM(p) with the cointegration rank r ≤ k is as follows:

$$\Delta_{yt} = c + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-1} + \varepsilon_i$$

Δ: Operator differencing, where $\Delta y_t = y_t - y_{t-1} \, y_{t-1}$

y_{t-1} Vector variable endogenous with the 1-st lag.

ε_t: Vector residual.

c: Vector intercept.

Π: Matrix coefficient of cointegration $(\Pi = \alpha\beta'$

α: vector adjustment, matrix with order (k × r) and

β: vector cointegration (long-run parameter) matrix (k × r)

Γ_i:: Matrix with order k × k of coefficient Endogenous of the i-th variable.

Testing for Normality

Jarque-Bera (JB) test

$$JB = \frac{n}{6}\left(S^2 + \frac{(k-3)^2}{4}\right)$$

Where

$$n: Number\ of\ sample$$

$$S: Expected\ skewness$$

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right]^{3/2}}$$

$$K: Expected\ exccess\ kurtosis$$

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2\right]^{2}}$$

Jarque-Bera (JB) (which is used in testing for normality for residuals) determined that the calculation used is as follows:

$$JB = \frac{n-k}{6}\left(S^2 + \frac{(k-3)^2}{4}\right)$$

Where k: Number of independent variables.

Testing for Stability

The stability system VAR can be from the inverse roots characteristics polynomial of AR. A VAR system is said to be stable (stationary) if all roots have a modulus of less than one and all are contained within the unit circle. The equation can be rewritten as:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

If this mechanism is started at certain time, for example at t=1, then we have :

$$y_1 = c + \phi_1 y_0 + \varepsilon_1;$$
$$y_2 = c + \phi_1 y_1 + \varepsilon_2;$$
$$= c + \phi_1 (c + \phi_1 y_0 + \varepsilon_1) + \varepsilon_1;$$
$$= (I_k + \phi_1)c + {\phi_1}^2 y_0 + \phi_1\varepsilon_1 + \varepsilon_2$$

$$y = (I_k + \phi_1 + \cdots + {\phi_1}^{t-1})c + {\phi_1}^t y_0 + \sum_{i=0}^{t-1} {\phi_1}^1 \varepsilon_{t-1}$$

$$y_t = (I_k + \phi_1 + \cdots + {\phi_1}^{t-1})c + {\phi_1}^t y_0 + \sum_{i=0}^{t-1} {\phi_1}^1 \varepsilon_{t-1}$$

therefore vector ($y_1, y_2, \ldots, y_t$) can be determine by vector ($y_0, \varepsilon_1, \ldots, \varepsilon_t$) and The joint distribution of is determined by joint distribution of ($y_1, y_2, \ldots, y_t$) is determined by joint distribution of ($y_0, \varepsilon_1, \ldots, \varepsilon_t$).

$$y_t = (I_k + \phi_1 + \cdots + {\phi_1}^{t-1})c + {\phi_1}^t y_0 + \sum_{i=0}^{t-1} {\phi_1}^1 \varepsilon_{t-1}$$

If all the eigen values of $\varnothing_1$ are less than 1 in absolute values, then the order of $\varnothing_1^i$, i=0,1,2...is summable. And the model y_t is stochastic process and defined as:

$$y_t = \mu \sum_{i=0}^{\propto} {\phi_1}^1 \varepsilon_{t-1}, t = \cdots - 1, \ldots$$

Formally Y_t can be said to stable if $\det(I_{Kp} - \phi z) = \det(I_K - - - \phi_1 z \ldots \phi_p z^p) \neq 0 \quad for\ z\| \leq 1$ This condition is called the stability condition.

Impulse Response Function (IRF)

The IRF is a method that can be used to determine the response of an endogenous variable toward a shock from the other variables. A Vector Autoregressive can be written as the form of Vector Moving Average (VMA). As an illustration, we used two variables in the form of matrix VAR as follows:

$$y_t = b_{10} + b_{12}z_t + \alpha_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t = b_{20} + b_{21}y_t + \alpha_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$$

In matrix notation it can be written as

$$\begin{pmatrix} 1 & b_{12} \\ b_{12} & 1 \end{pmatrix}\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} b_{10} \\ z_{20} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

Or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

Where

$$B = \begin{pmatrix} 1 & b_{12} \\ b_{12} & 1 \end{pmatrix}, X_t \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \Gamma_0 = \begin{pmatrix} b_{10} \\ z_{20} \end{pmatrix}, \Gamma_1 = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix}, and\ \varepsilon_t = \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

III. Results And Discussion

The first step of modeling time series is to check whether or not the time series data are stationary. To check the stationary of the data we can use time series plot, correlogram ACF and unit root test. Plot time series

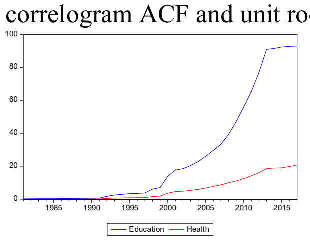

Figure 1. Plot time series data -Education and Health

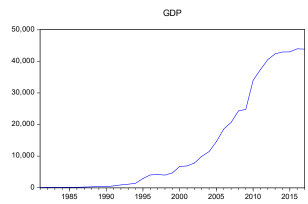

Figure 2. Plot time series data GDP

Table 1. Unit Root Test for variable Education

Exogenous: Constant

Lag Length: 1 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.047446	0.9475
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

Table 2. Unit Root Test for variable Health

Null Hypothesis: HEALTH has a unit root

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.459537	0.9988
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

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

Table 3. Unit Root Test for variable GDP

Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.994880	0.9998
Test critical values:	1% level	-3.626784	
	5% level	-2.945842	
	10% level	-2.611531	

At any lag, the three variables do not pass through the significance $\alpha = 0.05$, this means that the p-values of lag 0 to lag 5 are greater than 0.05. Thus, it is not sufficient evidence to reject Ho, so we conclude that the data are nonstationary.

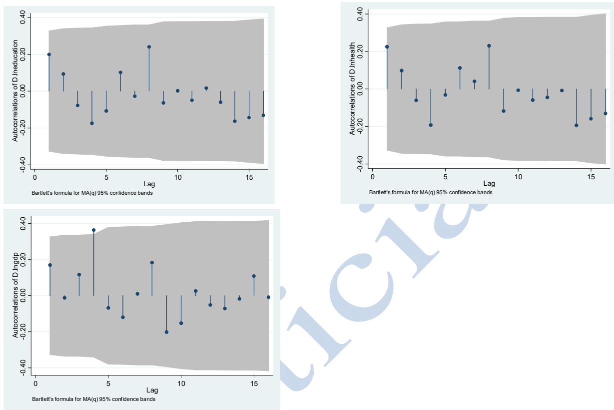

Figure 7. ACF variable Education, Health, GDP, Box-Cox(1) after first differencing.

Figures 7 shows that from lag 1 to lag 2 and up to lag 16 decreases tend to zero. Thus, we can conclude that based on correlogram ACF, the three variables data after the differencing are stationary.

Table 4. Unit root test variable Education Box-Cox(1) after first differencing.

Null Hypothesis: D(LNEDUCATION) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on AIC, maxlag=2)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.617968	0.0007
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

Table 5. Unit root test variable Health Box-Cox(1) after first differencing.

Null Hypothesis: D(LNHEALTH) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic - based on AIC, maxlag=2)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.662031	0.0006
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

Table 6. Unit root test variable GDP Box-Cox(1) after first differencing

Null Hypothesis: D(LNGDP) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic - based on AIC, maxlag=2)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.775406	0.0005
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

Unit Root Test for Stationarity show that the three variables after the first differencing passes through the significant level α = 0.05'. This means that the p-value at lag 0 to lag 5 is less than 0.05. The Ho is then rejected and we conclude that the data are stationary.

Cointegration Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.530455	38.41569	29.79707	0.0040
At most 1	0.256281	12.71201	15.49471	0.1258
At most 2	0.074842	2.644898	3.841466	0.1039

Model Estimation

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-74.59776	NA	4.401786	4.319872	4.364310	4.335212
1	12.12300	163.5306*	0.032837*	-0.578457*	-0.489580*	-0.547777*
2	12.44212	0.583523	0.034150	-0.539549	-0.406234	-0.493529

Table 9. Endogenous variables: LNEDUCATION

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-74.28377	NA	4.323512	4.301929	4.346368	4.317270
1	8.887339	156.8369*	0.039506*	-0.393562*	-0.304685*	-0.362882*
2	9.596660	1.297045	0.040179	-0.376952	-0.243636	-0.330932

Table 10. Endogenous variables: LNHEALTH

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-73.23681	NA	4.072438	4.242103	4.286542	4.257443
1	9.204584	155.4609*	0.038796*	-0.411691*	-0.322813*	-0.381010*
2	9.869205	1.215307	0.039558	-0.392526	-0.259210	-0.346505

The table above indicate that the lag optimal is at lag 1, hence, the VECM(p) model which is used is VECM (1).

Standard errors in ( ) & t-statistics in [ ]

	Lngdp	Lneducation	Lnhealth
LNGDP(-1)	0.969232	0.292977	0.287380
	(0.10642)	(0.10441)	(0.10414)
	[9.10753]	[2.80596]	[2.75950]
LNEDUCATION(-1)	-0.681975	0.433146	-0.502856
	(0.44298)	(0.43462)	(0.43350)
	[-1.53951]	[0.99660]	[-1.16000]
LNHEALTH(-1)	0.710305	0.261125	1.199319
	(0.46991)	(0.46104)	(0.45985)
	[1.51158]	[0.56638]	[2.60808]
C	1.352160	-1.290069	-1.339054
	(1.02308)	(1.00377)	(1.00118)
	[1.32165]	[-1.28522]	[-1.33748]
R-squared	0.993780	0.993792	0.993456
Adj. R-squared	0.993197	0.993210	0.992842
Sum sq. resids	0.986670	0.949778	0.944869

S.E. equation	0.175595	0.172280	0.171835
F-statistic	1704.351	1707.671	1619.285
Log likelihood	13.66310	14.34904	14.44232
Akaike AIC	-0.536839	-0.574946	-0.580129
Schwarz SC	-0.360892	-0.399000	-0.404182
Mean dependent	8.188930	1.902563	0.512802
S.D. dependent	2.128982	2.090822	2.031069

Determinant resid covariance (dof adj.)	3.60E-07
Determinant resid covariance	2.53E-07
Log likelihood	120.2033
Akaike information criterion	-6.011294
Schwarz criterion	-5.483454

Impulse Response Function

The changes in GDP can be attributed to varying factors, namely the level of export, agricultural production, oil export and other goods manufactured here in Nigeria.

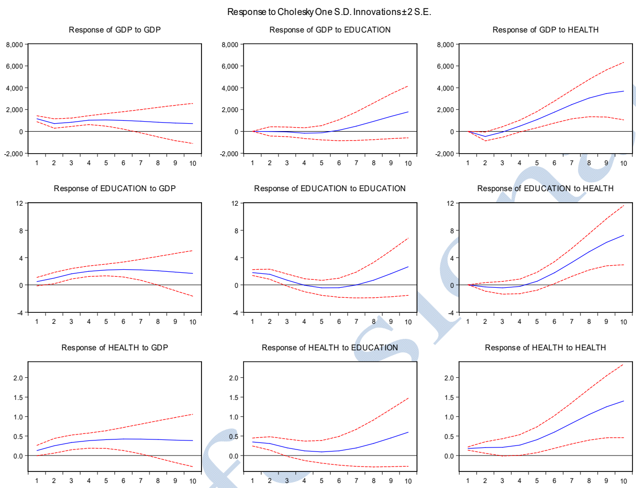

Figure 8. Graph of response of Health and Education Sector values toward the change in Gross Domestic Product (GDP

IV. Conclusion

Based on the discussion and results detailed above, the data on the Nigerian Gross Domestic Product (GDP), Contribution of the health sector to GDP (health) and the contribution of the education sector (education) can be modeled by using Vector Error Correction Model (1), VECM (1). By using this model, it was found that the Nigerian GDP, Education Sector and Health Sector have a co-integration relationship at rank = 1.

By using Impulse Response Function (IRF) it was found that when GDP changes the contribution of the health and education sector provides a positive response. On the other hand, the proportion of shock towards the changed in GDP provides a negative response. Thus, the proportion of shock in the change in GDP did not have a high contribution (effect) upon the education and health sector. On the other hand, the proportion of shock towards the changes in the contribution of the education and health sector of the economy has a high contribution (effect) on GDP.

In conclusion, Government should increase the rate of infrastructural development and funding of these sectors. Also government should design effective policy implementation in order to increase the quality of services provided by these sectors. Most specifically, to reduce the unnecessary delay in our education system thereby enabling the system to produce the required manpower for sustainable economic growth in Nigeria.

Acknowledgment

The authors are grateful to anonymous reviewers for their valuable comments on the original draft of this manuscript

References

[1] Adeniyi, O. and Abiodun, N. (2011). Health Expenditure and Nigerian Economic Growth. European Journal of Economics, Finance and Administrative Sciences.

[2] Bhargava, A., Jamison, D., Lau, L., & Murray, C. (2001). Modelling the effects of health on economic growth. Journal of Health Economics, 20(3), 423–440.

[3] Fuente, Angel de la (2006). "Education and Economic Growth: A Quick review of the evidence and some Policy Guidelines". A Paper Presented "Globalization and Challenges for Europe and Finland" organised by the Secretariat of the Economic Council.

[4] Johansen S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica. 1991 Nov; 59(6):1551–80.

[5] Montgomery D, Jennings C, Kulahci M. (2008). Introduction to time series analysis and forecasting. New York: John Wiley and Sons Interscience Publication; 2008.
[6] Sekar P. (2010). Diagnostic checking of time series models. Indian Journal of Science and Technology. 2010 Sep; 3(9):1026–31
[7] World Health Organization. (2018). Statistical information system. World Health Organization, WHO: Geneva. www.who.int/whosis/en.

.

.