



Panel Data Regression Modeling of Insurance Penetration on the Nigeria Economic Growth.

Chibo, Onyesom

Department of Statistics,
Captain Elechi Amadi Polytechnic, Rumuola, Port Harcourt, Nigeria
E-mail:chiboonyesom5@gmail.com.

ABSTRACT

This research present estimation of panel data regression models with individual effects. We discuss estimation techniques for both fixed and random effects panel data regression models. We derive experimental design and correlation estimators, and show their consistency. We discovered that the coefficient of determination $R^2 = 0.2687$, shows a fairly positive relationship, while the adjusted R (\bar{R}) 0.7574, show a stronger positive relationship .noting that insurance sector contributes to the growth of the country's economic growth. $t = 3.2079$, another indication of stronger relationship.

Keywords: Panel data, endogenous variables, experimental design, insurance, gross domestic product.

1.0 INTRODUCTION

Panel data or Longitudinal data consist of repeated observations from repetition of surveys with the same set of cross-sectional units for the same information over time .Insurance companies in Nigeria generate a data bank on Insurance Premium collected by its agents (Insurance Brokers) over time .The profitability of the Insurance company is dependent on the tangibility, liquidity, risk, etc. Insurance business is a risk transfer mechanism where the insured is put in the same pecuniary position that he was before the occurrence of the insured event.

Panel data are helpful in constructing and testing more complicated behavioral hypotheses. Panel data contain information on inter-temporal dynamics and may allow controlling the effects of unobserved variables in estimating a model, and in carrying out dynamic analysis. In panel data, if there are N cross section units and T time periods, then total number of observations will be NT. Therefore, the panel degrees of freedom are more providing, more variability than the cross sectional data or time series data.

1.1 Benefits of Panel data. Hsiao (2003) posited the benefits from using panel data to include; controlling for individual heterogeneity. Panel data suggests that individual firms, states, are heterogeneous. Time series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results (Moulton, 1986, 1987). Baltiga and Levin (1986) consider panel data estimation of cigarette demand across 46 American States. Consumption is modeled as a function of lagged consumption, price and income. Panel data give more informative data, more variability, less collinearity among the variables, more degree of freedom, and more efficiency. Time series studies are plagued with multicollinearity; for example, in the case of the cigarette demand, there is high collinearity between price and income in the aggregate time series for the USA. Panel data are better able to study the dynamics of adjustment.

1.2 Statement of the problem

The continuous decline in the growth of Nigeria economy is a great concern to the nation and the international community. A cursory look at the Nigeria Gross Domestic Product (GDP) from independence (1961) till date is of great concern to economic observers and the citizens. This continuous

decline and its negative impact on the prices of goods and services, and great impact on the inflationary trends of the country. The purpose of this research is therefore to peruse the conditions and examine the several sectors of the economy, especially the insurance industry, that contributes to the general growth of the Nigeria economy.

1.3 Objectives of the study

Insurance companies play a pivotal role in the economic development of any nation. Insurance is a risk transfer mechanism, whereby the insured is put in the pecuniary position he was before the occurrence of the insured event by the payment of a consideration called the premium into the pool of people facing similar situation, adequate for the degree of the risk being covered. However, the broad objective of this study is to appraise the way insurance fund (profitability) impacts on the Nigeria economic growth. Intend Specifically, the study intend to:

- (1) To establish the extent to which insurance companies contribute to the Nigeria economic growth.
- (2) To determine the various components of insurance growth that contribute to the general growth of the company.

2.0 Literature Review

Panel data econometrics is a continuously developing field. The increasing availability of data observed on cross –sections of units (like households, firms, countries etc),and over time has given rise to a number of estimation approaches exploiting this double dimensionally to cope with some of the typical problems associated with economic data.

The literature on statistical models, for panel data has experienced enormous growth over the last decades with several recent textbooks focusing on that subfield of econometrics. For example, Arellano (2005), Hsiao (2003), Baltagi (2005), Woodridge (2010)

The relevant literature on the relationship between Insurance and economic growth is reviewed as follows: Akinlo (2012) examined the effects of insurance on economic growth in Nigeria from 1986 to 2010. The structure, growth of insurance sub-sectors, and the direction between insurance causality and economic growth in Nigeria were addressed in the study. An error correction model analysis and cointegration technique was adopted .Alhassan and Fiador (2014) examined long –run causal relationship between insurance penetration and economic growth for Ghana during the period 1990 to 2010, using autoregressive distributed lag (ARDL) bounds approach to cointegration by Pesaran, Shin and Smith (1996, 2001).The study found a long run positive relationship between insurance penetration and economic growth.

Akinlo and Apaninsile (2014) examined the relationship between insurance and economic growth in sub Saharan Africa over the period 1986 – 2015.Pooled OLS, fixed effect model and generalized method of moment panel model were employed in the estimation. The estimation of the dynamic panel data result show that insurance has positive and significance impact on economic growth in the sub-Sahara Africa.

Olayungbo (2015) investigated the asymmetric non-linear relationship between from insurance and economic growth in Nigeria from 1976-2010.The conclusion is that asymmetric effect is present in Nigeria’s insurance market. Haise and Sumegi (2008) applied cross country panel data analysis from 29 European countries in the period from 1992-2005.The insurance variables is the premium income and total net investment of insurance companies. An estimation of ordinary least squares on unbalanced panel was done for the EU-15 Countries.

Omoke (2012) makes use of insurance density (premium per capita) as a measure for insurance market activity and real GDP for economic growth in Nigeria between 1970 -2008.The Johanson co integration and vector error correction approach were used to estimate the relationship among the variables. According to Pesaran, M H.(2015),the linear dynamic panel model include lag dependent variables as covariates along with the unobserved effects ,fixed or random, and exogenous regressors. The presence of lagged dependent variables allows for the modeling of a partial adjustment mechanism. Kugler and Ofoghi (2005) evaluate both long-run relationship and Granger-causality between insurance market size and economic growth for the United Kingdom using net written premium for each insurance market in the UK for the period 1966-2003.Using Johansen cointegration test the authors find a long –run

relationship between development in insurance market size and economic growth for all insurance components.

Essi, et al ,(2018) evaluated the effect of health expenditure per capita and the percentage of Gross Domestic Product (GDP) on infant mortality rate (IMR) on five countries using three model methods namely; Pooled Ordinary Least Squares (POLS),Fixed Effect Model (FEM) and Random Effect Model (REM).The three European countries showed that health expenditure and the percentage of GDP on health expenditure has no significant effect on IMR, while in the two African countries, these factors were highly significant, and had a negative correlation with their IMR. Im, et al (2003) developed a more flexible and computationally simple unit root for panel using the likelihood method. They noted that the IPS is not as restrictive as the Levin, et al unit root test. They noted that IPS test allows for a heterogeneous coefficient .The IPS test is a set of ADF test. O’Connel et al proposed a feasible GLS estimator for ρ where an assumed non-zero **form for the correlation** between the disturbances is employed. To overcome the limitation that the correlation matrix must be specified, Bai and Ng 2004 proposed an approach based on separating the data into a common factor component that is highly correlated across the series.

3.0 METHODOLOGY

Regression analysis of Panel data is a data structure which is panel data. Generally, parameter estimation in the regression analysis with cross – section data is done by estimating the least squares method called Ordinary Least Square (OLS) Regression Method Data Panel will give the result of estimation which is Best Linear Unbiased Estimation (BLUE). The data for the study was sourced from the electronic site of the Nigeria Stock Exchange.

The form panel data regression equation is similar to ordinary least squares ‘i e

$$y_{it} = \alpha + \beta'X_{it} + \varepsilon_{it} \dots (3.1)$$

3.1The Random Effect Model: The linear model we shall employ in the analysis is ;

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \{ I = 1,2,\dots,a,j = 1,2 \dots n \}, \dots (3.2)$$

where both the treatment effects τ_i and ε_{ij} are random variables. We will assume that the treatment effects τ_i are normally and independently distributed $N(0, \sigma_\tau^2)$ random variables, and that the errors are $N(N(0, \delta^2))$, random variables and that the τ_i and ε_{ij} are independent .Because τ_i is independent of ε_{ij} , the variance of any observation is ;

$$V(y_{ij}) = \sigma_\tau^2 + \delta^2$$

The variance σ_τ^2 and δ^2 are the variance components and the model (3.1) is the components of variance or random effect model. The observations in the random effects model are normally distributed because they are linear combinations of the two normally and independently distributed random variables $\tau_i + \varepsilon_{ij}$. However, unlike the fixed effects case in which all the observations y_{ij} are independent, in the random model the observations y_{ij} are only independent if they come from different factor levels.

We can express the covariance structure of the observations in the single –factor random effects model through the covariance matrix of the observations.

y = [y_{1,2}, y₁₂, y₂₁, y₂₂, y₃₁, y₃₂.] . and a 4 x 4 covariance matrix of the observations is;

$$\text{Cov.}(y) = \begin{matrix} & \delta_{\tau}^2 + \delta^2 & \delta^2 & 0 & 0 \\ & \delta_{\tau}^2 & \delta_{\tau}^2 + \delta^2 & 0 & 0 \\ & 0 & 0 & \delta_{\tau}^2 + \delta^2 & 0 \\ & 0 & 0 & 0 & \delta_{\tau}^2 + \delta^2 \end{matrix}$$

The main diagonals of this matrix are the variances of each individual observation and every off-diagonal elements in the covariance of a pair of observations..

3.2 The Analysis of variance for the Random Effect Model.

The basic ANOVA sum of squares identity: $SS_{\text{TOTAL}} = SS_{\text{TREATMENT}} + SS_{\text{ERROR}}$ is still valid. Testing hypotheses about individual treatment effects is very meaningful because they were selected randomly So the test hypotheses about the variance component δ_{τ}^2 ;

$H_0 : \delta_{\tau}^2 = 0$ vs $H_1 : \delta_{\tau}^2 > 0$. If $\delta_{\tau}^2 = 0$, all treatments are identical ,but if $\delta_{\tau}^2 < 0$,variability exists between treatments. But SS_E / δ^2 is distributed as chi-square with $N - a$ degree of freedom, and under the null hypotheses $SS_{\text{TREATMENT}} / \delta^2$ is distributed as chy-square with $a-1$ degree of freedom.

Both random variables are independent .Thus, under the null hypothesis, $\delta_{\tau}^2 = 0$, the ratio

$$F_0 = [SS_{\text{TREATMENT}}/a - 1] / [SS_{\text{ERROR}}/N-a] \\ = MS_{\text{TREATMENT}} / MS_{\text{ERROR}} \dots (3.2)$$

is distributed as F with $a-1$ and $N-1$ degree of freedom.

But the expected mean square:

$$E (MS_{\text{TREATMENT}}) = \frac{1}{a-1} -1 E (SS_{\text{TRETMENT}}) = 1/a-1 E \{ \sum_{i=1}^a y_{i.1}^2 /n - \frac{y_{..}^2}{N} \}. \\ = \frac{1}{a-1} E \{ \frac{1}{n} \sum_{i=1}^a (\sum_{j=1}^n \mu + \tau_i + \epsilon_{ij})^2 - \frac{1}{N} (\sum_{i=1}^a \sum_{j=1}^n \mu + \tau_i + \epsilon_{ij})^2 \}. \dots (3.5)$$

Under H_0 , both the numerator and denominator in (3.5) are unbiased estimator of δ^2 , whereas, under H_1 the expected value of the numerator is greater than the expected value of the denominator.

3.3 The Fixed Effect Model:

The single effect factor analysis of variance for the fixed effect model. We express:

$$y_i = \sum_{j=1}^n y_{ij} , \bar{y}_i = y_i/n , I = 1,2 \dots n \\ y_{..} = \sum_{i=1}^a \sum_{j=1}^n y_{ij} , \bar{y} = \frac{y_{..}}{N}$$

Where N = 'a.n' is the total number of observations. We are interested in testing the equality of the a treatment means; that is $E(y_{ij}) = \mu + \tau_i = \mu_i$, $i = 1, 2, \dots, a$. The appropriate hypotheses are :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_n \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_j \quad \text{for at least one pair } i, j.$$

In the fixed effect model we break the ith treatment mean μ_i into two components such that: $\mu_i = \mu + \tau_i$. We usually think of μ as an overall mean, so that: $\sum_{i=1}^a \mu_i / a = \mu$.

This definition implies that $\sum_{i=1}^a \tau_i = 0$., That is ,the treatment effect or factor effect can be thought of as deviations from the overall mean. Consequently, we can construct the hypotheses in terms of τ_i . as:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_n = 0 \quad \text{vs} \quad H_1 : \tau_i \neq 0 \quad \text{for at least one } i$$

3.4 Decomposition of the total sum of squares.

We decompose the total sum of squares, the total corrected sum of squares as:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

is used as a measure of overall variability of the data..

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}).$$

But the cross product term is equal to zero i e, $2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) = 0$, because,

$$\sum_{i=1}^a (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^a (y_{i.} - n \bar{y}_{i.})^2 = \sum_{i=1}^a (y_{i.} - n (y_{i.} / n))^2 = 0$$

$$\text{Therefore, we have } \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 .$$

$$\text{That implies: } SS_{TOTAL} = SS_{TREATMENT} + SS_{ERROR}.$$

Where $SS_{TREATMENT}$ is called the Sum of Squares due to treatment (ie between treatment) and SS_{ERROR} is called Sum of Squares due to error (ie within treatment)..There are 'an' = N total Observations; thus SS_T has N-1 degree of freedom. There are ' a ' levels of the factor (and a treatment means) so, $SS_{TREATMENT}$ has a-1 degree of freedom .Finally, there are ' n' replicates within any treatment providing n-I degree of freedom with which to estimate the experimental error. Because there are a treatment, we have a(n-1) = 'an' =N – a degrees of freedom for error.

The model: $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ {i=1,2,.. .a j =1,2,.. .n.

Describes two different situations with respect to treatment effects If we estimate the model parameters (μ, τ_i, δ^2) , this is called the fixed effect model. The τ_i are random variables, and knowledge about the

particular ones investigated is relatively useless. Instead we test hypotheses about the variability of the τ_i and try to estimate this variability. This is called the random effect model.

4.0 Test of hypotheses

We wish to test the hypotheses on the effect of time period (years) on the economic growth of the insurance companies vis-a vis the country, that is :

Ho(1) : Period has no effect on the Profitability, Tangibility, and Liquidity of the insurance company vis a vis the country.

Ho (2): Profitability, Tangibility and Liquidity have no relationship with the growth of the insurance companies, .vis a vis the country

$\alpha = 0.05$ (Test statistic)

Critical Region: Reject Ho , if F-cal is greater than F 0.05 (89,18) for period. Here F- cal = 1.0236 while $F_{0.05,(9,19)} = 2.96$.

Since the test hypothesis support H1, we have reason not to accept Ho (1)., and we conclude that the various time periods have effect on the economic growth of the companies..

Similarly, Ho (2): Profitability, Liquidity and Tangibility have no relationship with the growth of insurance company, vis a vis the country.

$\alpha = 0.05$

Critical Region. Reject Ho (2),if F.cal is greater than $F_{(0.05, 2,18)}$.

Since the test hypothesis support H1(2) ,we have reasons not to accept Ho(2),and conclude that the variables ;Profitability, Tangibility and Liquidity have effect on the growth of insurance companies.

Source of variation	Sum of squares	Degree of freedom	Mean Squares	F – Ratio
Time Period (year)	53474.94	9	5941.66	1.0236
Profitability,Liquidity,Tangibility	230886.22	2	115443.11	19.8878
Error	104485.14	18	5804.73	
Total	388846.30	29		

Source; computation by the author, (March, 2023)

The Regression Model; $R^2 = \frac{\text{Regression Sum of Squares}}{\text{Total Sum of Squares}} = \frac{SSR}{SST}$.

$$= \frac{104485.14}{388846.30} = 0.2687.$$

$$R = \sqrt{\text{Regression S. S}} = \sqrt{0.2687} = 0.5184.$$

This implies that about 51.84 per cent of the variation in the tail length is accounted for by the variation of Profitability, Liquidity and Tangibility in the economic growth of insurance companies. Coefficient of determination is the amount of variability in one of the variables accounted for by the variability in the other variables.

4.1 Test of existence of relationship between the variables (profitability, liquidity, and tangibility).

Ho: $\rho = 0$.(there is no correlation between the variables)

Ha : $\rho \neq 0$.(there is correlation between the variables).

$$t = \frac{r - \rho}{\sqrt{\frac{1-r^2}{n-2}}}, \text{ where } t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.5184}{.1616} = 3.2079$$

But $t_{0.05, 28} = 2.0484$. Since $t = 3.2079 > t_{0.05, 28} = 2.0484$, we do not accept Ho., There is correlation among the endogenous variables. We can see that the calculated t falls in the area rejection, therefore, Ho is rejected at the 0.05 level. This indicates enough evidence to say that there is correlation between the endogenous variables.

5.0 CONCLUSION AND RECOMMENDATIONS

R Square shows the magnitude of the influence or ability of the endogenous variables simultaneously in describing the response variables. Since the value is more than 0.5, then the ability of the predictor variables is strong in explaining the response variables. The adjusted R square is the magnitude of the influence or ability of predictor simultaneously in explaining the response variables by observing the standard error. The F statistic is the value of F test which is a simultaneous test of panel regression. This F value indicates the significance level of influence of predictor variables to response variables. Based on these tests support and the inferences derived, this study support government legislation to easy the bottlenecks encountered by investors in the establishment of insurance companies. Furthermore, since there is strong relationship between the endogenous variables and the exogenous variable, investment in insurance companies should support government monetary and fiscal policies. Researchers should expand this subject matter to include other variables of economic growth, and other methods of analysis to add to the literature of insurance and economic growth.

REFERENCES

- Akinlo, T., & Apanisile, O. (2014) Relationship between Insurance and Economic Growth in Sub-Saharan Africa: A panel data analysis. *Modern Economy*, 5, 120- 127.
- Arellano, M., and Bover, O., (1990) Another look at the Instrumental-variable Estimation for Error Components Models. Discussion Paper 7. London School of Economics .Center for Economic performance
- Bai, J. and Ng, S. (2004) A Panic Attack on Unit Roots and Co-integration. *Econometrica* 72, 1127-77
- Breitung, J. and S. Das., (2005) Testing for Unit Roots in panels with a factor structure. *Econometric Theory*; 24, 88-108
- Breitung, J. and Das, S. (2005). Panel Unit Root Tests Under cross-sectional Dependence. *Statistica. Neerlandica* 59, 41 *Review of Economics and statistics* 68: 4-33
- Das, P. (2019). *Econometrics in Theory and Practice. Analysis of Cross Section, Time Series and Panel Data with Stata 15.1* .Springer Nature Singapore Pte Ltd 2019.
- Essi, I.D., and Ogbonda, U.E., (2018). Panel Modeling of health expenditure per capita and percentage of GDP on infant mortality rate. *IIARD International Journal of Applied Science and Mathematical theory* 4(2) 35-48.

- Evereart,G.,(2013) Orthogonal to backward mean transformation for dynamic panel data models. *Econometrics Journal* 16,179-221.
- Haiss, P.,& Sumegi,K.(2008).The Relationship between Insurance and Economic growth in Europe. A theoretical and empirical analysis. *Emperica*, 35, 405-431.*The International Association of the study of Insurance Economics*, 35,(1018 – 5895)
- Montgomery,D.C.(2013) Design and Analysis