Comparative Analysis Of Least Square Regression And Fixed Effect Panel Data Regression Using Road Traffic Accident In Nigeria

J.A. Kupolusi, R.A Adeleke, O. Akinyemi, B. Oguntuase

ABSTRACT: In this research work, attempt was made to critically analyze the effect of Federal Road Safety Corps (FRSC) to various categories of road traffic accident in Nigeria for a certain period of time over all the states of federation including Federal capital territory. This was done by using panel data regression model. The conventional OLS estimator applied to panel data has over time led to inconsistent estimate of the regression parameters due to lack of adequately handling individual specific effect of the parameters. A better and preferable estimation method was exploited in this analysis to obtain a more reliable result that can be used for prediction of likely future occurrence. Among all the estimation methods considered, only the fixed effect panel data regression method with heteroscedasticity variance-covariance tools gives a consistent estimate of the regression parameters.

KEY WORDS: FRSC, Heteroscedasticity Variance-Covariance, Outliers, Ordinary Least Square Regression, Road Traffic Accident, R^2, t-test, Fixed Effect and Random Effect

INTRODUCTION

For many years, the most challenging task in statistics has been the effort to device methods for making causal inferences from non experimental data. Also a difficult problem is how to statistically control the variables that cannot be observed. For expermentalists, the solution to that problem is easy. Random assignment to treatment groups makes those groups approximately equal on all characteristics of the subjects, whether those characteristics are observable or unobservable. But in non experimental research, the classic way to control for confounding variables is to measure them and put them in some kind of regression model. In this research work, we describe a class of regression method, called Fixed Effect Panel data regression model that ensureheteroscedasticity robust standard error estimates. We say an estimator is robust if it is not seriously affected by changes in the assumption on which it isbased. There are two basic data requirements for using fixed effect methods. Firstly, the dependent variable must be measured for each individual on at least two occasions. Those measurements must be directly comparable, that is they must have the same meaning and metric. Secondly, the predictor variables (regressors) of interest must change in value across those two occasions for some substantial portion of the sample. Fixed effect methods are pretty much useless for estimating the effects of variables that do not change overtime, like race and sex of course, some statisticians argue that it makes no sense to talk about causal effects of such variables anywhere (Sobel, 2000). The increase availability of data observed on cross-sections of units (like household, firms states, countries etc) and over time has given rise to a number of estimation approaches exploiting this double dimensionality to cope with some of the typical problems associated with economic data, first of all that of unobserved heterogeneity. Time wise observation of data from different observational units has long been common in other field of statistics (where they are often termed longitudinal data). In the panel data field as well as in others, the econometric approach is nevertheless peculiar with respect to experimental contexts, as it is emphasizing model specification, testing and tackling a number of issues arising from the particular statistical problems associated with economic data. In analyzing regression models, there are other types of data that are available generally for empirical analysis, namely; time series data, cross sectional data and panel data. In time series data, we observe the value of one or more variables over time like gross domestic product GDP for several years or quarters. In cross sectional data, values of one or more variables are collected for several sample units, or entities, at the same point in time like mass failure of students in 30 selected secondary schools for a given year. In panel data the same cross-sectional unit is surveyed over time. Panel data set are repeated observations on the same cross-section, typically of individuals or firms or state in microeconomics applications, observed for several time periods. Other terms used for such data include longitudinal data and repeated measures. A major advantage of panel data over others is increase precision in estimation. This is the result of an increase in the number of observations owing to combining or pooling several time periods of data for each individual. However, for a valid statistical inference one needs to control for likely correlation of regression model errors overtime for a given individual. In particular, the usual formula for ordinary least square (OLS) standard errors in a pooled OLS regression typically overstates the precision gains, leading to underestimated standard error and t-statistics that can be greatly inflated. A second attraction of panel data is the possibility of consistent estimation of the fixed effect models, which allows for unobserved individual heterogeneity that may be correlated with regressors. Such unobserved heterogeneity leads to omitted variable bias that could in principle be correlated by instrumental variables method using only a single cross section, but in practice it can be difficult to obtain a valid instrument. Data from a panel offers an alternative way to proceed if the unobserved individual-specific effects are assumed to be additive and time invariate. Most disciplines in applied statistics other than micro-econometric treat any unobserved individual heterogeneity as being distributed independently of the regressors, then the effects are called random. Compared to fixed effect model this stronger assumption has the advantage of permitting consistent estimator of all parameters, including coefficient of time-invariateregressors. However, random effects and pooled estimators are inconsistent if the true model is one with fixed effects. A third attraction of panel data is the possibility of learning more about the dynamics of individual behavior than is possible from a single cross-section. Thus a cross-section may yield an accident rate of 30% but we need panel data to determine whether the same 30% are in accident each year.

ROAD TRAFFIC ACCIDENT IN NIGERIA

A Road Traffic Accident (RTA) is when a road vehicle collides with another vehicle, pedestrian, animal or geographical or architectural obstacle. The RTAs can result in injury, properties damage, fatal, minor and death. RTA results in the deaths of 1.2m people worldwide each year and injuries about 4time this number (WHO, 2004). In this study, a road traffic accident is defined as accident which took place on the road between two or more objects, one of which must be any kind of a moving vehicle (Jha et al, 2004). Road Traffic Accidents (RTAs) are increasing with rapid pace and presently these are some of the leading causes of death in Nigeria. The morbidity and mortality burden in developing countries like Nigeria is rising due to a combination of factors, including rapid motorization, poor road and traffic infrastructure as well as the behavior of road user (Nantulya and Reich, 2002). This contrasts with technologically advanced countries where the indices are reducing (Oskam et al, 1994; O'Neil and Mohan, 2002). Nigeria, a heavily motorized country with poor road conditions and transportation systems has a high rate of Road Traffic Accidents (RTAs) and the tendency is on increase. The recognition of RTA as a crisis in Nigeria inspired the establishment of the Federal Road Safety Commission (FRSC). The FRSC was established by the government of the Federal Republic of Nigeria vide Decree 45 of 1988 as amended by Decree 35 of 1992, with effect from 18th February 1988. The commission was charged with responsibilities for, among others, policymaking, organization and administration of road safety in Nigeria. Much attention has not been put into variation in the number of RTAs across states in Nigeria over a period of time. This study attempts to view and investigate various variations in the number of RTAs across all the states in Nigeria with respect to distribution of FRSC over a period of five years from 2002 to 2006. This is achieved by observing the relationship between national licensing scheme and various categories of RTAs (like killed, injured, fatal, serious and minor) in Nigeria to obtain a consistent estimate of individual effect in order to reduce variation across states and over years(i.e. bias adjustment for Heteroscedasticity Robust (HR)standard error estimates). The National Road Traffic Regulations, harmonized and standardized by the Federal Road Safety Commission, contains the guiding road traffic rules and regulations for a good road safety culture in Nigeria. Overtime the Federal Road Safety Commission (FRSC) has put in place various schemes to reduce road traffic accidents in Nigeria. These schemes include: - revision of the Highway Code, national FRSC and public education The Nigerian Highway Code was revised in 1989 to meet local and international specifications of road traffic management and crash control. The result is a culture-related guide for driver education. The well-illustrated Revised Highway Code was translated to the three major Nigerian languages: Hausa, Igbo, and Yoruba as well as Arabic. The French version is in the pipeline. The National Licensing Scheme represents a landmark in the achievements of the FRSC. The success of the scheme, introduced in 1989, has continued to provide a veritable avenue for ensuring a good road safety culture among drivers. The scheme is made up of: - National Driver's License, National Vehicle License, National Vehicle Identification Scheme, National Driver's Testing and vehicle Examination ,National Road Traffic Regulations ,Vehicle Identification Tag, Road Worthiness Validity Tag and Proof of Ownership CertificateIn its determination to restore the integrity of the Nigerian drivers' license and improve the capacity of the National FRSC, the Federal Road Safety Corps, FRSC, has set aside 1 October and 1 December 2010 to launch a new drivers' license and vehicle number plates respectively.

METHODOLOGY

This research work focuses at obtaining consistent estimate of fixed effect models for panel data regression applied to data of road traffic accidents whichoccurred in Nigeria between 2002 and 2006in the 36 states of the federation and FCT. Here, a secondary data set is used. This is obtained from published bulletin of the Bureau of Statistics .The states of the federation serve as cross-section units and the involved years as time periods.

SPECIFICATION OF MODEL

PANEL DATA MODELS

A general linear model for panel data permits the intercept and slope coefficients to vary over both individual and time, with

$$Y_{it} = \alpha_{it} + X'_{it}\beta_{it} + \mu_{it} \quad i= 1\ldots\ldots\ldots\ldots\ldots N, \quad t= 1\ldots\ldots\ldots\ldots T; \qquad \ldots\ldots\ldots\ldots\ldots(1)$$

Where Y_{it} is a scalar dependent variable, x_{it} is a kx1 vector of independent variable, μ_{it} is a scalar disturbance term, I indexes individual in a cross section, and t indexes time. This model is too general and is not estimable as there are more parameters to be estimated than observations. Further restrictions need to be placed on the extent to which α_{it} and β_{it} vary with i and t, and on the behavior of the error μ_{it} .

FIXED EFFECTS MODEL

The fixed effects model specifies

$$Y_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it}$$

for $i = 1\ldots.N, t = 1\ldots T$ ……….(1)

Where the individual-specific effects $\alpha_1, \alpha_2, \ldots\ldots\ldots\ldots\ldots\alpha_N$ measure unobserved heterogeneity that is possibly correlated with the regressors, X_{it} and β are k x1 vectors, and to start with the errors ε_{it} are iid $(0,\sigma^2)$.

The major challenge for estimation is the increase in N individual-specific effects as N becomes large. For this research we are interested in the slope parameter β. The N parameters α_1, α_2, …………α_N are nuisance parameters or incidental parameters that are not of intrinsic interest. There are several ways to consistently estimate β for linear model despite the presence of these nuisance parameters. Some of which are;

- **(a)** OLS in the within model
- **(b)** Direct OLS estimation of the model with indicator variables for each of the N fixed effects.
- **(c)** GLS in the within model
- **(d)** ML estimation conditional on the individual means $\bar{Y}_i$ i=1,………N and
- **(e)** OLS in the first-differences model

The fixed effect model used for this analysis is thus expressed as

$Y_{it} = \alpha_i + \beta' X_{ijt} + \mu_{it}$ where j= 1……5, i=1……37, t=1…..5

Where X_{ijk} is a k x j vector of regressors and β' = (β_1, β_2, β_3, β_4,β_5) and ………..(2)

$$X_{it} = \begin{pmatrix} killed \\ injured \\ fatal \\ serious \\ \min or \end{pmatrix}_{it}$$

i.e $(FRSC)_{it} = \alpha_i + \beta_1 killed_{it} + \beta_2 injured_{it} + \beta_3 fatal_{it} + \beta_4 serious_{it} + \beta_5 minor_{it} + \mu_{it}$ ……(3)

Here, we represent Y =FRSC, X1 = killed, X_2= injured, X_3 = fatal, X_4= serious, X_5= minor and $\mu_{it}$is error term in matrix form.

TERMINOLOGIES USED

Killed: accidents with death recorded at the spot

Fatal: accidents with major injuries with the possibility of recorded death later.

Serious: accidents that cause major injuries. Without death.

Minor: accidents that cause minor injuries with fewer or nobody injured and little or no vehicle damage.

Injured: these are accidents that cause many different injuries to virtually any part of the body, depending on the circumstance of the crash and the severity of the impact.
Equation (3) above is expressed as

$$Y_{it} = \alpha_i + \beta' X_{ijt} + \mu_{it} \quad \text{……………(4)}$$

In order to avoid the problem of outliers, there is need for linear transformation of the variables. That is

$$\log Y_{it} = \alpha_i + \beta' \log X_{ijt} + \mu_{it} \quad \text{…(5)}$$

Let$\log_e Y_{it} = Y^*_{it}$ and$\log_e X_{it} = X^*_{it}$ now (5) becomes

$Y^*_{it} = \alpha_i + \beta' X^*_{it} + \mu_{it}$ i= 1……..37, t=1…5 …….…….(6)

and β=1……5 where $\propto_1 = \propto_2 = \propto_3 = \text{………} = \propto_{37} = 0$

ASYMPTOTIC PROPERTIES OF A GOOD ESTIMATOR

(1) **Asymptotic unbiasedness;** an estimator $\hat{\theta}$ is said to be an asymptotically unbiased estimator of θ if $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$

(ii) **Consistency** ; $\hat{\theta}$ is said to be a consistent estimator of θ if it approaches the true value θ as the sample size gets larger. A sufficient condition for consistency is that mean square error of$\hat{\theta}$, MSE ($\hat{\theta}$) tends to zero as n increases indefinitely.

(iii) **Asymptotic Efficiency;** Let $\hat{\theta}$ be an estimator of θ. The variance of the asymptotic distribution of$\hat{\theta}$ IfV$(\widehat{\theta})$ is consistent and its asymptotic variance is smaller than the asymptotic variance of all other consistent estimators of θ, $\hat{\theta}$ is called asymptotic efficient.

(iv) **Asymptotic Normality;** An estimator $\hat{\theta}$ is said to be asymptotically normally distributed if its sampling distribution tends to the normal distribution as the sample size n increases indefinitely.

FIXED EFFECTS ESTIMATOR

This is obtained by subtracting the time-averaged model

$\bar{Y}^*_i = \alpha_i + \bar{X}'^*_i \beta + \bar{\mu}_i$ from the original model. Then

$$Y^*_{it} - \bar{Y}^*_i = \left(X^*_{it} - \bar{X}^*_i\right)' \beta + \left(\mu_{it} - \bar{\mu}_i\right) \quad \text{…………(7)}$$

So fixed effect α_i is eliminated, along with time-invariant regressors since

$X^*_{it} - \bar{X}^*_i = 0$ if $X^*_{it} = X^*_i$ for all t.

From the equation above, let

$$\left(Y^*{}_{it} - \bar{Y}_i^*\right) = \ddot{Y}_{it},$$

$$\left(X_{it}^* - \bar{X}_i^*\right) = \ddot{X}_{it} \text{ and}$$

$$\mu_{it} - \bar{\mu}_i = \ddot{\mu}_{it}$$

Therefore the new equation is thus:

$$\ddot{Y}_{it} = \ddot{X}'_{it}\beta + \ddot{\mu}_{it} \quad \text{.............(8)}$$

By using OLS estimation, equation (8) yields the within estimator or fixed effects estimator $\hat{\beta}_W$ i.e from (8)

$$\ddot{\mu}_{it} = \ddot{Y}_{it} - \ddot{X}'_{it}\beta \quad ;$$

$$\hat{\ddot{Y}}_{it} = \ddot{X}'_{it}\hat{\beta}$$

$$\ddot{\mu}'_{it}\ddot{\mu}_{it} = \left(\ddot{Y}_{it} - \ddot{X}'_{it}\beta\right)'\left(\ddot{Y}_{it} - \ddot{X}'_{it}\beta\right)$$

$$= \ddot{Y}'_{it}\ddot{Y}_{it} - \ddot{Y}_{it}{}'\ddot{X}_{it}\beta - \ddot{X}_{it}\beta'\ddot{Y}_{it} + \ddot{X}'_{it}\beta'\ddot{X}_{it}\beta$$

$$= \ddot{Y}'_{it}\ddot{Y}_{it} - 2\beta'\ddot{X}'_{it}\ddot{Y}_{it} + \beta'\ddot{X}'_{it}\ddot{X}_{it}\beta$$

$$\frac{\partial y(\ddot{\mu}'_{it}\ddot{\mu}_{it})}{\partial\hat{\beta}} = 0 - 2\ddot{X}'_{it}\ddot{Y}_{it} + 2\ddot{X}'_{it}\ddot{X}_{it}\hat{\beta}$$

$$\approx 2\ddot{X}'_{it}\ddot{Y}_{it} = 2\ddot{X}'_{it}\ddot{X}_{it}\hat{\beta}$$

$$\hat{\beta}_W = \left(\ddot{X}'_{it}\ddot{X}_{it}\right)^{-1}\ddot{X}'_{it}\ddot{Y}_{it} \quad \text{.......(8a)}$$

$$\hat{\beta}_W = \sum_{i=1}^{37}\sum_{t=1}^{10}\left(\ddot{X}'_{it}\ddot{X}_{it}\right)^{-1}\ddot{X}'_{it}\ddot{Y}_{it}$$

Stacking observations over time period for a given individual and over the N individuals, the within estimator for β (β_W) in (8a) becomes normal OLS estimator

$$\hat{\beta}_W = \left(\ddot{X}'\ddot{X}\right)^{-1}\ddot{X}'\ddot{Y} \quad \text{..........(8b)}$$

CONSISTENCY OF THE FIXED ESTIMATOR

The fixed effect estimator of $\hat{\beta}_W$ is consistent if

$$prob.\lim(NT)^{-1}\sum_{i=1}^{37}\sum_{t=1}^{5}\left(X_{it}^* - \bar{X}_i^*\right)\left(\mu_{it} - \bar{\mu}_i\right) = 0$$

For all N→∞ or T →∞ and $E\left(\mu_{it} - \bar{\mu}_i \middle| X_{it}^* - \bar{X}_i^*\right) = 0$

DERIVATION OF THE VARIANCE OF THE FIXED ESTIMATOR

By using matrix algebra, consider equation (6) of i^{th} observation

$$Y^*{}_{it} = \alpha_i + \beta' X^*{}_{it} + \mu_{it}$$

Where X^*_{it} and β are K x 1 vectors. For the i^{th} individuals, stack all T observations, so

$$\begin{pmatrix} y^*_{i1} \\ . \\ . \\ . \\ y^*_{i5} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix}\alpha_i + \begin{pmatrix} x'^*_{i1} \\ . \\ . \\ . \\ x'^*_{i5} \end{pmatrix}\beta + \begin{pmatrix} \mu_{i1} \\ . \\ . \\ . \\ \mu_{i5} \end{pmatrix} \quad \text{i=1...N or}$$

$$Y^*{}_i = e\alpha_i + X^*{}_i\beta + \mu_i \quad \text{i=1...N} \quad \text{............(9)}$$

Where e = (1,1,1,1, ………………..1) is a Tx1 vector of ones, x^*_i is a TxK matrix, and y^*_i and μ_i are Tx1 vectors. To transform model (9) to the within model that subtracts the individual-specific mean, introduce the TxT matrix

$Q = I_T - T^{-1}ee'$ ………….(10)

Pre-multiplication of w by the matrix Q converts w to deviations from the mean, since

$QW_i = W_i - e\bar{w}'_i$ ………………..(11)

Where W_i is a TxM matrix with ith row w'_{it} and $\bar{w}_i = T^{-1}\sum_{t=1}^{T} w_{it}$ is a mx1 vector of averages.

Result (11) is obtained by using $e'W_i = T\bar{w}_i'$. Note also that QQ' =Q using e'e =T and Qe=0, so Q is idempotent.

Pre-multiplying the fixed effect model (9) for the ith individual by Q yields

$$QY_i^* = QX_i^*\beta + Q\mu_i$$ i=1………….37……… (12)

Using Qe= 0. This is the within model (7), since equivalently

Y_i^* - $e\bar{y}_i^{*\prime}$ =(X^*_i - $e\bar{x}_i^{*\prime}$)β + (μ_i - $e\bar{\mu}_i$)   Using (11)

Thus pre-multiplication by Q yields the within model. An OLS estimation of (12) yields $\hat{\beta}_w$ with variance matrix, assuming independence over i, equals to

$$V(\hat{\beta}_W)=\left[\sum_{i=1}^{37} X_i'^* Q'QX_i^*\right]^{-1}\sum_{i=1}^{37} X_i'^* Q'V(Q\mu_i|X_i^*)QX_i^*\left[\sum_{i=1}^{37} X_i'^* Q'QX_i^*\right]^{-1}$$
…..(13)

Beginning with the assumption that μ_{it} are iid[0,σ_μ^2], so that μ_i are iid [0,σ_μ^2I]. the Tx1 error $Q\mu_i$ is then independent over i, with mean zero and variance V($Q\mu_i$)=QV(μ_i)Q'

=σ^2_μQQ' =σ^2Q

Then from equation (13)

$$\sum_{i=1}^{37} X_i'^* Q'V(Q\mu_i|X_i^*)QX_i^* = \sum_{i=1}^{37} X_i'^* Q'\sigma_\mu^2 QQX_i^*$$

$$= \sigma_\mu^2 \sum_{i=1}^{37} X_i'^* Q'QX_i^*$$ …………(13a)

And $X_i'^* Q'QX_i^* = (QX_i^*)'(QX_i^*)$

= $\sum_{t=1}^{5}$(X_{it}^* - $\bar{X}_i^*$)(X_{it}^* - $\bar{X}_i^*$)' …………(13b)

Substituting (13a) and (13b) into (13)

$$V(\beta_W)=\left[\sum_{i=1}^{37}\sum_{t=1}^{5}(X_{it}^*-\bar{X}_i^*)(X_{it}^*-\bar{X}_i^*)'\right]^{-1}\times$$

$$\times\left[\sum_{i=1}^{37}\sum_{t=1}^{5}(X_{it}^*-\bar{X}_i^*)(X_{it}^*-\bar{X}_i^*)'\right]\sigma_\mu^2\left[\sum_{i=1}^{37}\sum_{t=1}^{5}(X_{it}^*-\bar{X}_i^*)(X_{it}^*-\bar{X}_i^*)'\right]^{-1}$$

Recall that $(X_{it}^* - \bar{X}_i^*) = \ddot{X}_{it}$ *and* $\sigma_\mu^2 = \hat{\ddot{\mu}}_{it}\hat{\ddot{\mu}}_{is}$

$$V(\hat{\beta}_W)=\left[\sum_{i=1}^{37}\sum_{t=1}^{5}(\ddot{X}_{it}\ddot{X}'_{it})\right]^{-1}\sum_{i=1}^{37}\sum_{t=1}^{5}\sum_{s=1}^{5}(\ddot{X}_{it}\ddot{X}')\hat{\ddot{\mu}}_{it}\hat{\ddot{\mu}}'_{is}\left[\sum_{i=1}^{37}\sum_{t=1}^{5}(\ddot{X}_{it}\ddot{X}'_{it})\right]^{-1}$$
..(14)

Equation (14) is the robust estimate of the asymptotic variance for short panel i.e panel data with little time effect.

ASSUMPTIONS OF FIXED EFFECT STANDARD ERROR

(X_{i1}, ……….,X_{iT},…………….μ_{it}) are iid over i=1,…n

E(μ_{it}/X_{i1},………………….X_{iT},)=0 (strict exogeneity)

$Q\bar{x}\bar{x} \equiv E\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'$ is nonsingular

(no perfect multicollinearity)

E($\mu_{it}\mu_{is}$| X_{i1}, …….X_{iT}) =0 for t≠s conditionally serially uncorrelated errors).

TESTING FOR FIXED EFFECT MODEL

(i) When the true model is fixed effect as considered above, OLS yields biased and inconsistent estimates of the regression parameters. This is an omission variable bias due to the fact that OLS deletes the individual dummies when in fact they are relevant. A joint significance of individual specific is tested i.e H_o ; α_1 =α_2 =,…………,=α_{36}=0 performing an F-test Here

$$F_{obs} = \frac{\left((RRSS - URSS)/(N-1)\right)}{URSS/(NT-N-K)} \approx F_{N-1,N(T-1)-K}$$
…….(22)

This is a Chow test with the restricted residual sums of squares (RRSS) being that of OLS on the pooled model and the unrestricted residual sum of squares,(URSS) being that of the Least Square Dummy Variable (LSDV) or fixed effect regression. For a large N, the within transformation is performed and use that residual sum of squares as the URSS. **Note:** within effect regression package divides the residual sum of squares by NT-K instead of NT-N-K from the LSDV regression, there is need to adjust the variances obtained from the within regression (7) by multiplying the variance- covariance matrix by (NT-K)/[NT-N-K]. The significance of this F_{Obs}-test shows that fixed effect estimator is most appropriate.

We can also use

$$F_{obs} = \frac{(R_{UR}^2 - R_R^2)/(N-1)}{(1-R_{UR}^2)/(NT-N-K)} \approx F_{N,N(T-1)-K}$$
………(23)

Where R^2_{UR} = Unrestricted R^2, R^2_R = Restricted R^2 of OLS

TESTING THE OVERALL SIGNIFICANCE OF β

Assuming that ε_{it} are normally distributed and the null hypothesis is $H_O : \beta_2 = \beta_3 = ... = \beta_6 = 0$ against

$$H_1 : \beta_2 \neq \beta_3 \neq ... \neq \beta_6 \neq 0$$

Then the F-statistic follows that

$$F_{cal} = \frac{R^2 / (K-1)}{(1-R^2)/(NT-N-K)} \approx F_{tab\alpha,(K-1,NT-N-K)}$$

Reject H_0 if $F_{cal.} > F_{tab}$ i.e the coefficients are significantly different

TESTING FOR INDIVIDUAL FIXED EFFECT COEFFICIENT

The t-statistic test for individual coefficient is given thus;

t-statistic $$|t| = \frac{\beta_i}{S.E(\beta_i)} \approx t_{\alpha/2(n-1)d.f}$$

if $|t| > t_{\alpha/2}$ then the coefficient is significant.

RESULTS

TEST FOR FIXED EFFECT MODEL

When the true model is fixed effect as considered, OLS yield biased and inconsistent estimates of the regression parameters. This is an omission variable bias due to the fact that OLS delete the individual dummies when in fact they are relevant. A joint significance of individual specific effects is tested i.e Hypothesis: H_o; $\alpha_1 = \alpha_2 =, ..., _{36} = 0$ performing an F-test against

H_1; $\alpha_1 \neq \alpha_2 \neq, ..., \neq \alpha_{36} \neq 0$

Test Statistic

$$F_{obs} = \frac{(R^2_{UR} - R^2_R)/(N-1)}{(1-R^2_{UR})/(NT-N-K)} \approx F_{N,N(T-1)-K}$$

Fobs. = <u>(0.806083 – 0.251262)/(37-1)</u>

(1-0.806083)/(185-37-5)

Fobs = <u>0.0154117</u>

0.0013561

= 11.3647 ;F(N-1, N[T-1]-K) = F(36,143) =1

Decision; reject Ho if Fobs >F(N-1,N[T-1]-k) Conclusion; fixed effect model is most appropriate for this analysis that is the intercepts are not random since the calculated value is greater than the tabulated value

TABLE OF RESULTS

Table I POOLED LEAST SQUARE REGRESION(OLS)

variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	15045.79	34.566	-11.050	-38.78	25.38	15.953	0.25	0.247	17218.6 2	-12398.67	74.096	0.000	1.7985
Std.Err.	846.732	7.800	2.950	5.332	3.077	4.147							
T-stat	17.769	4.431	-3.745	-7.274	8.249	3.8460							
Prob.	0.0000	0.0000	0.0002	0.0000	0.000	0.0001							

$$\hat{y}_{it} = 15045.79 + 34.56\hat{X}_{1it} - 11.050\hat{X}_{2it} - 38.78\hat{X}_{3it} + 25.38\hat{X}_{4it} + 15.95\hat{X}_{5it}$$

Discussion: the output above is the result of ordinary pool regression which shows that all the variables are significant but the coefficient of determination is very low (R^2=25.12%). This implies that 25.12% of the total variation in the dependent variable is explained by the set of independent variables. Therefore the model does not best fit the regression variables; there is need for better model to fit the variables. Hence the model of the output above is expressed as

GENERALIZED LEAST SQUARE METHODS

Considering the result above, by using weighted least square method, we are able to test for heteroscedasticity of the error terms. The outcome is shown in the table 2 below. It is obvious that this result gives a better estimate of panel data regression. {White cross-sectional standard errors and covariance test for the presence of heteroscedasticity in time series data and white period standard errors and covariance test for that of cross-section data.}

Table 2 PANEL LEAST SQUARE REGRESSION

variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	15045.79	34.566	-11.050	-38.78	25.383	15.953	0.25	0.230	17457.44	-2066.445	12.013	0.0000	1.1519
Std.Err.	2102.831	19.371	7.327	13.242	7.641	10.301							
T-stat	7.155	1.784	-1.508	-2.929	3.321	1.548							
Prob.	0.000	0.0761	0.133	0.003	0.001	0.123							

Discussion: both the ordinary pooled method and ordinary panel data method give the same result only that in ordinary panel data method, some of the hidden variables are well spelt out. The variables minor and injured are significant in pooled regression method but not significant in panel data method. The variables are “minor and injured”. Hence the model of the output above is thus expressed as

$$\hat{y}_{it} = 15045.79 + 34.566\hat{X}_{1it} - 11.050\hat{X}_{2it} - 38.78\hat{X}_{3it} + 25.38\hat{X}_{4it} + 15.95\hat{X}_{5it}$$

Table 3PANEL LEAST SQUARE REGRESSION WITH WHITE DIAGONAL STANDARD ERRORS AND COVARIANCE FOR CROSS SECTION AND PERIOD FIXED

Variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	19587.93	68.169	-25.714	-54.82	22.145	17.031	0.52	0.363	15874.96	-2025.47	3.335	0.0000	1.7431 8
Std.Err.	3478.82	33.602	10.454	22.65	13.491	12.031							
T-stat	5.630	2.028	-2.459	-2.419	1.641	1.415							
Prob.	0.000	0.044	0.015	0.016	0.103	0.159							

Discussion: Here both the cross section and time series variables are kept constant. This result shows that not all the variables are significant and there is an improvement in the value of coefficient of determination which is now 51.92%; yet this result did not show us among all other tests that it will best fit the regression model. There is need to test for other parameter of interest in the model. The model of this result is thus expressed

$$\hat{y}_{it} = 19587.93 + 68.169\hat{X}_{1it} - 25.714\hat{X}_{2it} - 54.820\hat{X}_{3it} + 22.145\hat{X}_{4it} + 17.03\hat{X}_{5it}$$

Table 4 PANEL LEAST SQUARE REGRESSION WITH ONLY CROSS SECTION FIXED

Variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	17513.51	68.571	-22.256	-51.61	22.192	15.952	0.49	0.35	15962.88	-2029.118	3.486	0.0000	1.773
Std.Err.	3799.170	24.872	9.633	14.693	9.450	10.636							
T-stat	4.609	2.756	-2.310	-3.512	2.348	1.499							
Prob.	0.0000	0.0066	0.0223	0.0006	0.0202	0.1359							

Discussion: considering the result above, fixing only the cross section variables gives a better estimate of the regression model, only that the adjusted R-square is small. Hence the model of the output above is thus expressed as

$$\hat{y}_{it} = 17513.51 + 68.57\hat{X}_{1it} - 22.25\hat{X}_{2it} - 51.61\hat{X}_{3it} + 22.25\hat{X}_{4it} + 15.95\hat{X}_{5it}$$

Table 5 PANEL GENERALISED LEAST SQUARE REGRESSION (PERIOD WEIGHTS) WITH WHITE CROSS SECTION STANDARD ERRORS AND COVARIANCEEFFECT WITH PERIOD FIXED AND CROSS SECTION CONSTANT

variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	14321.33	7.013	-7.697	-19.34	39.070	-0.839	0.41	0.377	17846.57	-2039.66	13.421	0.0000	1.114
Std.Err.	2006.337	21.740	9.742	18.548	13.477	21.414							
T-stat	7.138	0.322	-0.790	-1.042	2.898	-0.039							
Prob.	0.0000	0.7474	0.4305	0.2985	0.0042	0.2985							

Discussion: since the R-square residual is still very low for the weighted statistics (R^2=40.83%) there is need for further re-estimation of the model. Hence the model of the output above is thus expressed

Table 6 PANEL GENERALISED LEAST SQUARE REGRESSION (CROSS SECTION WEIGHTS) WITH WHITE PERIOD STANDARD ERRORS AND COVARIANCEEFFECT: CROSS SECTION FIXED AND PERIOD CONSTANT

Variable	Constant	X_1	X_2	X_3	X_4	X_5	R^2	$\bar{R}^2$	StdEr.	Loglikel	F-stat	ProbF	D.W
Coef.	16553.37	-11.740	2.179	-14.64	5.134	13.111	0.80	0.744	17180.32	-1856.601	14.109	0.0000	2.165
Std.Err.	1138.553	6.075	2.192	4.307	4.586	2.120							
T-stat	14.538	-1.932	0.993	-3.399	1.119	6.182							
Prob.	0.0000	0.0553	0.3219	0.0009	0.2648	0.0000							

Discussion: among all the estimation methods considered, only the estimator with weighted least square best fit the panel data regression model of about R^2= 80% and the necessary heteroscedasticity that are robust to covariance have been reduced drastically. This result is thus considered for the analysis of road traffic accident as considered in the research. The model of the estimate is thus expressed as $\hat{y}_{it} = 16553.37 - 11.74\hat{X}_{1it} + 2.17\hat{X}_{2it} - 14.64\hat{X}_{3it} + 5.134\hat{X}_{4it} + 13.11\hat{X}_{5it}$

TESTING THE OVERALL SIGNIFICANCE OF β

Assuming that ε_{it} are normally distributed and the null hypothesis is

H_O : $\beta_2 = \beta_3$ …………………= β_6 = 0 against

H_1 : $\beta_2 \neq \beta_3$ ……………………= $\beta_6 \neq 0$

Then the F-statistic is

$$F_{cal} = \frac{R^2/(K-1)}{(1-R^2)/(NT-N-K)} \sim F_{tab\ \propto,\ (K-1,\ NT-N-K)}$$

Test Statistic $$F_{cal} = \frac{0.806083/5}{1-0.806083/142} \approx F_{tab\ 0.05,(5,\ 142)}$$

$$F_{cal} = \frac{0.1612166}{0.0013656}$$

= 118.05551 $F_{tab\ 0.005(5,142)}$ =1

Conclusion: since $F_{cal} > F_{tab}$that is the coefficients are significantly not all equal to zero. This result does not mean that all the regression coefficients are significant. There is need to test for individual coefficients separately to ascertain the variable of interest in the model.

TESING FOR INDIVIDUAL FIXED EFFECT COEFFICIENT

The t-statistic test for individual coefficient is given as;

t-statistic /t/ = $\frac{\beta_i}{S.E(\beta_i)} \sim t_{\alpha/2(n-1)}$d.f

Test for killed as variable:

t-statistic /t/ = -11.74054/6.075199

=1.932536

$t_{\propto/2(n-1)} = t_{0.025,26} = 2.03$

If the null hypothesis Ho is true, the probability of obtaining as much as 1.9324 or greater (in absolute value) is only 0.0553

Test for injured as variable:

t-statistic /t/= 2.179590/2.192900

= 0.993931

$t_{\propto/2(n-1)} = t_{0.025,26} = 2.03$

Test for minor as variable;

t-statistic /t/=13.11164/2.120911

=6.182080

$t_{\propto/2(n-1)} = t_{0.025,26} = 2.03$

Test for serious as variable:

t-statistic /t/ = 5.134714/4.586666

=1.119487

$t_{\propto/2(n-1)} = t_{0.025,26} = 2.03$

Test for fatal as variable:

t-statistics /t/= -14.64450/4.307495

= 0.0009

$t_{\propto/2(n-1)} = t_{0.025,26} = 2.03$

INTERPRETATION OF RESULTS

Only variables minor and fatal are significantly while other variables like killed, injured and serious are not significant. That is only the case of road accident with minor and fatal accident has significant impact on the uniform license scheme.

SUMMARY

Among all the analyses considered, only the analysis of fixed effect variables in either cross-sectional data or time series give desired result.Themodel for the analysis is thus expressed as

$$\hat{y}_{it} = 16553.37 + 2.17 Injured_{it} - 11.74 killed_{it} + 13.11 Minor_{it} - 14.64 fatal_{it} + 5.1347 Serious_{it}$$

From the result, the coefficient of determination is about 80% that is about 80% of total variation was explained by the model.

CONCLUSION

Based on the analysis carried out, we arrived at the following conclusion. It was shown in table I that using pooled least square method to analyze this type of data will lead to inconsistence of the regression parameters. Pooled OLS method shows that all the regression parameters are significant. But coefficient of determination that explains variations between two or more linear related variables was very low. That is 25% of total variation in FRSC was explained by various categories of road traffic accident in Nigeria. This method does not fit the model well. Also considering the panel least square method in table 2, 3 and 4 with both effect fixed and constant do not fit the model well. Table 5 and 6 gave consistent estimate of the regression parameters. Time period has no significant effect on the number of FRSC distributed to all the states of the federation over the period of five years since all the regression parameters are almost not significant in table 5.

that is FRSC distributed to all the states has not really made a significant impact in the reduction of road traffic accident in the country. Therefore there is need to increase the number of FRSC staff in order to reduce road traffic accident in Nigeria. Table 6 shows that using fixed effect panel data with fixed cross-section variable and time period hold constant gives a consistent estimate of the regression parameters. Here this method is best to estimate this kind of dataset. In the result, 80.17% of total variation in FRSC is explained by different categories of road traffic accident in Nigeria. Based on this result, categories of road traffic accident like injured and serious have no significant effect on the number of FRSC distributed to all the states. Other categories like killed, minor and fatal show that there is a significant relationship between the numbers distributed. Then a unit increase in FRSC distributed will cause approximately 12% decrease in the number of people that will be killed by road traffic accident, 15% decrease in the number of fatal accident and 13% increase in the number of minor accident will occur in Nigeria.

RECOMMENDATION

In view of the analysis carried out in this research, to maintain safety of life of motorist and commuter, we recommend that;

- **(i)** Government should recruit more people to the scheme so that drivers can be checkmated on our road and make sure that motorist obey road sign rules.
- **(ii)** The staff of the scheme must be well educated on when to and not to stop vehicle on high speed
- **(iii)** Our motorist must be enlightened and familiar to road signs in the country
- **(iv)** Government need to construct good road and maintain the existing once to allow motorist to travel near and far in the country.
- **(v)** Government should decentralize the staff of the scheme to both major and minor roads so that minor accident across the country will be drastically reduced.
- **(vi)** Government should enforce law that will bind the motorist on over speeding
- **(vii)** Government should disallow under age driving on highways

REFERENCES

[1] Amemiya T (1971). The Estimation of the Variances in a Variance- Components Model."InternationalEconomic Review, 12, 1{13}

[2] Anderson TW, Hsiao C (1981).Estimation of Dynamic Models with Error Components."Journal of the American Statistical Association, 76, 598{606}

[3] Ahn, S.G., Y.H. Lee and P. Schmidt (2001): "GMM Estimation of Linear Panel DataModels with Time-varying Individual Effects," Journal of Econometrics, 102, 219-255.

[4] Arellano M (1987). Computing Robust Standard Errors for Within Group Estimators."Oxford Bulletin ofEconomics and Statistics, 49, 431{434}

[5] Arrelano, M.(2003). Panel Data Econometrics, Oxford: Oxford University Press.

[6] Baltagi BH (2001). Econometric Analysis of Panel Data (3rd edition) John Wiley & Sons, Hoboken, NJ

[7] Bertrand, Marianne; Duflo, Esther and Mullainathan, Sendhil. "How Much Should We TrustDifferences-in-Differences Estimates?" Quarterly Journal of Economics, 2004, 119(1), pp. 249-75.

[8] Brillinger, D. (1981). Time Series Data Analysis and Theory .San Francisco: Holden-Day.

[9] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," Review ofEconomic Studies, 47, 225-238.

[10] Damodar N. Gujarati (2007). Basic Econometrics 4th Edition Greene. Model for Panel data 2008

[11] Greene, W.H. 2002. Econometric Analysis, 5th ed. Upper Saddle River, NJ: Prentice-Hall.

[12] Hansen, C. (2007). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data with T is Large," manuscript, Graduate School of Business, University of Chicago; forthcoming, Journal of Econometrics

[13] Kezdi, G., 2002. Robust standard error estimation in fixed-effects panel models. Mimeo

[14] Kezdi, Gabor, 2004, "Robust Standard Error Estimation in Fixed-Effects Panel Models",forthcoming. Hungarian Statistical Review.

[15] Kiefer, N.M. (1980): "A Time Series-Cross Section Model Fixed Effects with anIntertemporal Factor Structure," Unpublished manuscript, Department of Economics, Cornell University

[16] Laird NM, Ware JH (1982).Random-Effects Models for Longitudinal Data." Biometrics,38, 963 {974}

[17] MacKinnon JG, White H (1985). Some Heteroscedasticity-Consistent Covariance MatrixEstimators with Improved Finite Sample Properties "Journal of Econometrics, 29,305{325}

[18] Mohan D. An analysis of road traffic fatalities in Delhi India. Accid Anal Prev1985,17:33-45

[19] Mundlak Y (1978). On the Pooling of Time Series and Cross Section Data " Econometrica,46(1), 69{85}

[20] MunnellAH (1990). Why Has Productivity Growth Declined? Productivity and PublicInvestment." New England Economic Review, pp 3{22}

[21] NantulyaV.M and M.R Reich, 2002. The neglected epidermic; Road traffic injuries in developing countries BMJ; 324: 1139-1141

[22] Nerlove M (1971). Further Evidence on the Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections." Econometrica, 39, 359{382}

[23] Nickell, S. (1981): “Biases in Dynamic Models with Fixed Effects,” Econometrica, 49,1417-1426.

[24] O’Neill, B and D Mohan, 2002, reducing motor vehicle crash dealth and injuries in newly motorizing countries, BMJ; 324:1142-1145

[25] Oskam,J.J: kingma and H.J. Klasen, 1994, The Groningen trauma study. Injury Pattern in a Dutch trauma centre. Eur. J. Emergency Med, 1:167-172

[26] Pesaran MH (2004). General Diagnostic Tests for Cross Section Dependence in Panels Technical Report1229, CESifo Working Paper Series.

[27] Pinheiro J, Bates D, DebRoy S, Sarkar D (2007). nlme: Linear and Nonlinear Mixed EffectsModels. R package version 3.1-86,

[28] Pinheiro JC, Bates D (2000). Mixed-Effects Models in S and S-PLUS. Springer-Verlag, NewYork.R Development Core Team (2008). R: A Language and Environment for Statistical Computing.R Foundation for Statistical Computing,

[29] Soderlund N, Zwi AB. Traffic related mortality in industrialized and less developed countries.Bull World Health Organ 1995;73:175-82

[30] Sobel,M.E. (2000), “Causal Inference in the Social Sciences,” Journal of the American Statistical Association, 95,647–651

[31] Stock J, Watson M (2006). Heteroscedasticity-Robust Standard Errors for Fixed EffectsPanel Data Regression." Technical Report 0323, NBER Working Paper Series.

[32] WHO, 2004 world Report on Road Traffic injury prevention: Summary. World health organization, Geneva, Switzerland.

[33] White H (1984).A Heterosedasticity-Consistent Covariance Matrix and a Direct Test forHeteroscedasticity. Econometrica, 48, 817{838}

[34] Windmeijer F (2005). \A Finite Sample Correction for the Variance of Linear Efficient Two-Steps GMM Estimators. Journal of Econometrics, 126, 25{51}

[35] Wooldridge JM (2002). Econometric Analysis of Cross-Section and Panel Data.