APPLICATION OF PANEL DATA ANALYSIS OF THE GROSS OUTPUT OF SOME PRINTING INDUSTRIES IN SOUTH-WEST NIGERIA

Asabe Ibrahim

Department of Mathematical Sciences, Adamawa State University, Mubi.

Abstract

This work focused on the econometric analysis of panel data using fixed effect (FE) and random effect (RE) model. The study aimed at estimating the FE model which examines group differences in intercepts which assume slopes and constant variances across groups; the RE model which estimates variance components for groups and error assuming the same intercept and slopes. It tested the individual effects related to fixed and random effects model; compared FE and RE using the Hausman test to determine whether FE or RE should be used. A pooled regression was also taken which shows that F values are not significant, indicating that the model fits the data. Significance of test was further measured using the t value(s) in which /t/>2 is considered significant. The overall regression value coefficient is also significant. Test of the linear hypothesis shows that the independent variables (labour capital and industrial cost) have a linear relationship with the dependent variable (gross output). Test of heteroscadasticity revealed a constant variable across and among the variables. FE regression results also reveal that all dummies are not different from zero and the test of fixed group effect shows that there is no fixed group effect in the model. In the RE, the real hypothesis of zero variance is not rejected and the Hausman test of specification is not rejected in favour of the fixed effect model.

Keywords: panel data, econometric analysis, printing industry, fixed effects and random effects.

Introduction

For a long time econometrics has been considered a highly specialized tool of research, yet its rapid growth and increasing use in economic planning and research require simplification and wider diffusion (Baltagi, 2001). Econometrics deals with the measurement of economic relationships. It may be considered as the integration of economics, mathematics and statistics for the purpose of providing numerical values for the parameters of economic relationships and the verification of economic theories (Koutsoyiannis, 2003).

An econometric study begins with a set of propositions about some aspects of the economy. The theory specifies a set of precise, deterministic relationships among variables. In line with this view, Frisch (1933) defines econometrics as the field of economics that concerns itself with the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by economic theory.

The Economist's *Dictionary of Economics on the other* defines econometrics as "the setting up of mathematical models describing mathematical models describing economic relationships (such as that the quantity demanded of a good is dependent positively on income and negatively on price), testing the validity of such hypotheses and estimating the parameters in order to obtain a measure of the strengths of the influences of the different independent variables."

Further, econometrics maybe defined generally as the "application of mathematics and statistical methods to the analysis of economic data. (Wikipedia, the Free Encyclopedia).

From the foregoing, therefore, econometrics is a tool used in mathematics, statistics and economics to measure relationships among economic variables in order to make some economic decision(s).

Many recent studies have made use of econometrics to analyze panel or longitudinal data sets. This may not be unconnected with the obvious advantages which accrue due to this approach. Two famous ones are the National Longitudinal Survey of Labour Market Experience (NLSY) and the Michigan Panel Study of Income Dynamics (PSID) (Green, 2003). Panel Data refers to cross-section repeatedly sampled overtime, but where the same economic agent has been followed throughout the period of sample. This has a data set containing series of observations per each n entity. In general, two principal segments of industry could be recognized in this study, newspaper and magazine publishing which is dominated by expatriates, governments and political parties; and commercial, and job printing. Panel data model is used in this work in line with the views of Damodar and Gujarati (2007) because of the following advantages.

- Since panel data related to individuals, firms, states and countries overtime, there is bound to be heterogeneity in these units. The techniques of panel data estimation can take such heterogeneity explicitly into account by allowing for individual-specific variables.
- By studying the repeated cross-section of observations, panel data are better suited to study the dynamics of change. Spells of unemployment, job turnover and labour mobility are better studied with panel data.
- Panel data enables us to study more complicated behavioural models.
- By making data available for several thousand units, panel data can minimize the bias that might result if we aggregate individuals or firms into broad aggregates.

Consequently, panel data can enrich empirical analysis in ways that may not be possible if we use only cross-section or time series data. One of the most frequent used panel data set is the panel study of income dynamics (PSID) developed by the Institute of Social Research at the University of Michigan (Johnson and John, 2003). Since 1968, researchers have collected information on more than 5000 families. Once a year family members are re-interviewed about their economic status and many socio economic and demographic characteristics. Arelano (2003) published a paper, which focused on two of the developments in panel data econometrics. Firstly, it provided a review of linear panel data models with pre-determined variables. Secondly, it discussed the implications of assuming that explanatory variables are pre-determined as opposed to the assumption that they are strictly exogenous in dynamic structural equations with unobserved heterogeneity.

As literacy increases, so also does the demand for printed materials. Far back in 1965, there were four nationally-distributed newspapers in Nigeria and more than fifteen additional newspapers aimed at regional or local markets. The sudden rise in population of students requires books and other printed materials. Commercial ventures require handbills; business cards and host of printing forms. A growing market existed for magazines with a local flavour and for literature of general interest.

This study emphasized on firm based form of data, in order to establish relationship among some economic factors (labour, capital, industrial cost, gross output), make some estimation as well as test some useful hypotheses. These economic factors contribute to production in printing industries. In seeking to understand the functions of industries such as printing industries, a theoretical model needs to be built. All models are mental simplification of reality. A model builder seeks to capture the fundamental features of the system being studied.

Industry is a product of industrialization, a process that developed alongside scientific and technological advancement. Studies have shown that many qualitative factors can influence the gross output of printing firms. These factors include; capital, labour industrial cost and value added which form the focus of the study.

Statement of the problem

This study sought to construct an economic model that relates some economic factors(labor, capital, industrial cost and gross output) of 31 printing firms, test relationships between these factors, as well as examine how the factors contributes to production.

There are several advantages of using panel data, some of which are;

1) Panel data enables us to study more complicated behavioral models.
2) The techniques of panel data estimation can take care of heterogeneity explicitly by allowing for individual-specific variables.
3) Panel data gives more information on a data, more variability; less co-linearity among variables, more degree of freedom and more efficiency by combining time series and cross sections of observation.
4) Panel data are better able to study the dynamics of adjustment, spells of un-employment, job turn over studied with panels.

Model Specification and Econometric Models

Econometric models are mathematical expression of economic theory. A useful economic model emerges from a theoretical economic preposition which can be cast in mathematical form.

The Components of econometric model are classified into three; viz, the variables, the parameters and the disturbance terms. In general, variables are entities that can take many forms in a model. They are classified into two categories; endogenous and exogenous variables. Parameters are constants in a model. Their values may be known or unknown and can be estimated using sample data. Disturbance terms are terms included in econometric models to represent factors that are not adequately accounted for. They represent whatever; gap may approximate the behavior given by econometric models.

Properties of Econometric Models

Good econometric model possess the following properties:

Simplicity: a model must be sufficiently simple in order to allow for analytical operations to be performed on it.

Relevance: model ability to relate to an important economic problem, that is to say an econometric model is worth constructing when its estimation provides a useful answer to important and relevant econometric quest.

Explanatory: model should be able to explain as adequately as possible the real economic behavior to which it relates.

Theoretical plausibility: model should be plausible in the context of well-established economic theory.

Method

This research work used the Cobb-Douglas production function in order to study the relationship between output and input. The Cobb-Douglas production function in its stochastic form may be expressed as:

$$y = \beta_0 X_{1it}^{\beta_1} X_{2it}^{\beta_2} X_{3it}^{\beta_3} \ell^{ui}$$

y = Gross output (in naira) (is the dependent variable we are trying to predict).

X_1 = Labour (number of employers)

X_2 = Capital (in naira)

X_3 = Independent cost (in naira)

X_1, X_2, X_3: are the coefficients or multipliers that describe the size of the effect the independent variables are having on the dependent variable y.

Ά = Constant term (is the value y is predicted to have when all the independent variables are equal to zero)

U = Error term (it represents those factors that are not adequately accounted for in the model).

e = base of natural logarithm. This shows clearly that the relationship between gross output and the three inputs (i.e. Labour, Capital and Industrial Cost) is nonlinear. If we then log transform the model in order to linearize it, we obtain:

$Lny_i = Ln\beta_o + \beta_1 LnX_{1i} + \beta_2 LnX_{2i} + \beta_3 LnX_{3i} + U_i$

$Lny_i = \alpha + \beta_1 LnX_{1i} + \beta_2 LnX_{2i} + B_3 LnX_{3i} + U_i$

Functional forms of one-way panel data model

Fixed group effect model:

$y_{it} = (\alpha + \mu_i) + X_{it}\beta + V_{it}$

Where $V_{it} \approx IID\ (0, \sigma^2)$

Random group effect model:

$y_{it} = \alpha + X_{it} + (\mu_i + V_{it})$

Where $V_{it} \approx IID\ (0, \sigma^2_v)$

The dummy variable is part of the intercept in the fixed effect model and a part of error in the random effect model.

$V_{it} \approx IID(0, \sigma^2_v)$ indicates that errors are independent identically distributed.

NATATIONS

y_i = Dependent variable (DV) mean of group i.

X_i = Means of independent variables (IVS) at times t

y.. = Overall means of the DV

X.. = Overall means of the IVS

n = The number of groups or firms

T = The number of times periods

N = nT = total number of observations

k = the number of regressions excluding dummy variables

K = (k + 1) (including the intercept)

Regression Analysis

Sometimes it is difficult to identify or measure all the explanatory variables relevant to a study, but it is imperative to think is hard to identify and measure as many relevant ones as possible. If one explanatory variable is involved, the concept is called a simple regression. When two or more explanatory variables are involved, we move from the concept of simple regression to multiple regressions in which case, there are more than two variables in the model to handle.

The two variable regression models often represent a gross over-simplification of real life situations in econometric studies. In this case, one independent variable would be inadequate in explaining or predicting changes in the dependent variable. The simplest form of the multiple regressions is the three variable regression models, which includes one dependent variable and two explanatory variables.

In the context of this work, we are using four variable regression models that consist of one dependent variable (Y) and three explanatory or independent variables (X_1, X_2, X_3).

This involves stacking time series and cross-sectional observation together and then running a common regression called the ordinary least square on them.

The classical linear regression model is based on several simplified assumptions.

Assumption

1. The regression model is linear in the parameters
2. The value of the regressors the $X^{'s}$, are fixed in repeated sampling.
3. For given $X^{'s}$, the mean value of the disturbance U_i is zero
4. The regression model is correctly specified
5. There is no exact linear relationship in the regressor
6. For given$^{'s}$, the variance of U_i is constant

Estimation of the regression coefficients

The multi-co linearity problem suggests rewriting the equation as

$y = Z_*\delta + Z_\mu \mu + \varepsilon$

as

Where $y = ZB + Z_\mu^*[\,^\alpha{}_\mu] +, \varepsilon$

$Z^*_\mu = [I_{NT}\ Z_\mu]$

Cleary the column spaces generated by Z_μ and by Z^*_N are the same. Introducing the projectors onto its orthogonal complement;

$P^*_\mu = Z_\mu (Z_\mu Z)^{-1} Z^1_\mu$, $M_\mu = I_{NT} - P_\mu$

Using the decomposition of multiple regressions provides an easy way for constructing the OLS estimator of β, namely.

$\beta\ as = (Z^1 M_\mu Z)^{-1} Z^1 M_{\mu y} y$

$V(\beta_{OLS} / Z, Z_\mu) = \sigma^2_\varepsilon (Z^1 M_\mu Z)^{-1}$

Fixed effect models (FEM)

There are several strategies for estimating fixed effect models. The least squares dummy variables model (LSDV) uses dummy variables, whereas the within effects does not. These strategies produced the identical slopes of non-dummy independent variables.

The between effects model also does not use dummies, but produces different parameter estimates.

LSDV's is widely used because it is relatively easy to estimate and interpret. The functional form of LSDV model is $y_i = i\alpha_i + X_i\beta + \varepsilon_i$. It however becomes problematic when there are many groups or subjects in the panel data. If T is fixed and $N \rightarrow \infty$, only coefficients of regressors are consistent. $\alpha + N_i$, are not consistent since the number of parameters increases as N increases (Baltagi 2001). This problem is called the incidental parameter problem. Under this circumstance LSDV is useless, calling for another strategy, the within effect model.

The within effect model does not used dummy variables, but uses deviation from group means. Thus, this model is the OLS of: *equation* without an intercept. The parameter estimates of regressors are identical to those of LSDV. Since no dummy is used, the within effect model has a larger standard errors. Hence R^2 of the within effect model is not correct because an intercept is suppressed.

The between group effect model, also called the group mean regression, uses the group means of the dependent and independent variables. The model uses aggregated data to test effects between groups (or individuals), the functional form of the between effect is given by $y_{it} = \alpha + \chi_{it} + \varepsilon_i$.

The hypothesis of the fixed effect model, along with the parameter space, may be written as follow:

$$y = Z\sigma + Z_v\mu + \varepsilon$$
$$\varepsilon \sim (0, \sigma^2_\varepsilon I_{NT}),\ \varepsilon \perp (Z, _{Zv}).$$

Individual effects

Consider the equation $y = Z\beta + Z_\mu\mu + \varepsilon$ where μ's are the ordinates at the origin of each individual hyper-plane and eventually measure the "individual effects". In this case, the μ's are estimated as

$\mu = (Z'_\mu m_z z_\mu)^{-1} z'_\mu m_z y$

$\mu = y_i - Z^1_i\,\beta_{OLS}\ 1 \leq i \leq \mu$

$M_z = I_{NT} - ZZ^+,\ Z_i = 1\sum Z_{it}: K\chi 1$

Now, α is an "average" ordinate at the origin and the μ'_I s are differences between that "average" ordinate at the origin and the ordinates at the origin of each individual hyper planes. Thus α represent an "average effect" and the μ'_i measure eventually the "individual effects" in deviation form.

The coefficients α and μ_i are accordingly estimated as follows:

$$\hat{\alpha} = \bar{y} - \bar{z}\hat{\beta}_{OLS}$$

$$\hat{\mu}_i = \bar{y}_i - \hat{\alpha} - z'_i\hat{\beta}_{OLS}$$

$$= (\bar{y}_i - \bar{y}) - (\bar{z}_i - \bar{z})'\hat{\beta}_{OLS} \quad 1 \leq i \leq N$$

$$where \quad \bar{z} = \tfrac{1}{NT}\sum_{1 \leq i \leq N}\sum z_{itit} : kx1$$

$$\mu_N = 0$$

Random effects models (REM)

The one-way random effect model is given by

$y_{it} = \alpha + \beta_{\chi it} + \mu_i + V_{it}$

$W_{it} = \mu_i + v_{it}$ where $\mu_i \sim IID(0,\sigma^2_\mu)$ and $v_{it} \sim IID(0,\sigma^2_v)$

The μ_i are assumed independent of V_{it} and X_{it}, which are also independent of each other for all i and t.

A random effect model is estimated by generalized least square (GLS) when the variance structure is known and by feasible generalized least squares (FGLS) when the variance is unknown. Compared to FEM, REM relatively difficult to estimate.

Generalized least squares (GLS)

When Ω is known (given), GLS based on the true variance components is BLUE and all the FGLS estimators considered are asymptotically efficient as either n or T ~ ∞ (Baltagi 2001).

The Ω matrix is given as below:

$$\Omega = \begin{bmatrix} \delta_\mu^2 + \delta_v^2 & \delta_\mu^2 --- \delta_\mu^2 \\ \delta_\mu^2 & \delta_\mu^2 + \delta_v^2 --- \delta_\mu^2 \\ --- & --- \quad --- \\ \delta_\mu^2 & \delta_\mu^2 \quad --- \delta_\mu^2 \end{bmatrix}$$

In GLS, one just needs to compute θ using the Ω matrix.

$\theta = 1 - \sqrt{\dfrac{\delta_v^2}{T\delta_\mu^2 + \delta_v^2}}$. Then transform variable as

$$y_{it}^* = y_{it} - \theta \bar{y}_{i.}$$

$$X_{it}^{\&*} = X_{it} - \theta \bar{X}_{i.\ for\ all\ X_k}$$

$$\alpha^* = 1 - \theta$$

Finally, run OLS with the transformed variables:

Since Ω is often unknown, FGLS is more frequently used rather than GLS.

Feasible generalized least squares (FGLS)

If Ω is unknown, then θ has to be estimated first using

$\hat{\sigma}_\mu^2$ *and* $\hat{\sigma}_v^2$

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_\mu^2 + \hat{\sigma}_v^2}} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_{btw}^2}} \quad where$$

$$\hat{\sigma}_v^2 = \frac{SSE_{within}}{nT - n - k} = \frac{e'e_{within}}{nT - n - k} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{t}(v_{it} - \bar{v}_{i.})^2}{nT - n - k}$$

Where v_{it} are the residuals of the LSDV. The $\hat{\sigma}_v^2$ comes from the effect model (group mean regression).

$$\hat{\sigma}_\mu^2 = \hat{\sigma}\,between - \frac{\hat{\sigma v},}{T}\ where$$

$$\hat{\sigma}^2\,between = \frac{SSEbetween}{n - k}$$

Transforming the variables using $\hat{\theta}$ and then run OLS:

Testing group effects

The null hypothesis is that all dummy parameters except one are zero:

$$Ho: \quad \mu_1 = \mu_2 = \ldots \mu_{n-1} = 0$$

This hypothesis is tested by the test, which is based on loss of goodness of fit.

$$F = \frac{(e'e_{efficiency} - e'e_{Robust})/(n-1)}{(e'e_{Robust})/(nT - n - k)}$$

$$= \frac{(R^2{}_{Robust} - R^2{}_{Efficient})/(n-1)}{(1 - R^2{}_{Robust})/(nT - n - k)} \approx F(n-1,\ NT - n - k)$$

If the null hypothesis is rejected, one may conclude that the fixed group effect model is better than the pooled OLS model.

Fixed time effect and two-way fixed effect model.

For fixed time effects model, one need to change n and T, i and t in the formula.

$$Model: \quad y_{it} = \sigma + T_i + \beta' X_{it} + \varepsilon_{it}$$

$$Withingeffectmodel: \quad (y_{it} - \bar{y}_{*t}) = \beta'(\chi_{it} - \bar{\chi}_{*t}) + (\varepsilon_{it} - \bar{\varepsilon}_{*t})$$

$$Dummy\ coefficient: d_t^* = \bar{y}_{*t} - \beta' \bar{\chi}_{*t}$$

$$Standard\ errors: \ Se^*_k = Se_k \sqrt{\frac{df_{error}^{within}}{df_{error}^{LSDV}}}$$

$$= Se_k \sqrt{\frac{nT - k}{nT - n - T - k + 1}}$$

$$H_0 : \mu_1 = \mu_2 = ... = \mu_{n-1} = 0 \text{ and } t_1 = t_2 = ... = t_{t-1} = 0$$

$$F - test,\ F = \frac{(e^1 e_{Efficient} - e' e_{Robust})/(n + T - 2}{(e^1 e_{Robust})/(nT - n - k + 1)} \approx F[(n + T - 1), (nT - T - k + 1$$

Testing random effects (LM Test)

The null hypothesis is that cross-sectional variances components are zero: H_0: $\delta^2 = 0$ Breusch a and pagan (1980) developed the lagrange multiplier (LM) test given by

$$LM_\mu = \frac{nT}{2(T-1)}\left[\frac{e' DDe}{e' e} - 2\right]^2$$

$$= \frac{nT}{2(T-1)}\left[\frac{T^2 \bar{e}' \bar{e}}{e'_i e} - 1\right]^2 \approx \chi^2(1)$$

Where $\bar{e}$ the nxl vector of the group is specific means of pooled regression residuals, and $e'e$ is the SSE of the pooled OLS regression. Baltagi (2001) also present some LM test

$$LM_\mu = \frac{nT}{2(T-1)}\left[\frac{\sum(\sum e_{it})^2}{\sum\sum e^2_{it}} - 1\right]$$

$$= \frac{nT}{2(T-1)}\left[\frac{\sum(T \bar{e}_{i.})^2}{\sum\sum e^2_{it}} - 1\right]^2 \approx \chi^2_{(1)}$$

Poolability Tests

In order to run the pool ability test, one need to run group by group OLS regressions and/or time by time OLS regressions. This test is performed to known if slopes are the same across groups or overtime. The null hypothesis of the pool ability test across group is

$$H_0 : \quad \beta_{ik} = \beta_k .$$

The null hypothesis of the pool ability test over time is

$$H_0 : \quad \beta_{ik} = \beta_k .$$

It uses the F-statistic

$$F_{Obs} = \frac{(e'e - \sum e'_i e_i)/(n-1)k}{\sum e'_i e_i / n(T-k)} \approx F[(n-1)k, n(T-k)]$$

Where $e'e$ is the SSE of the pooled OLS and $e'_i e_i$ is the SSE of the OLS regression for group i.

If the null hypothesis is rejected, then the panel data are not poolable. Under this circumstance, you may go to the random coefficient model or hierarchical regression model.

In a similar manner, one can test for the pool ability of tests over time. The hypothesis is $H_0 : \quad \beta_{tk} = \beta_k$. The F-test is

$$F_{Obs} = \frac{(e'e - \sum e'_t e_t)/(T-1)k}{\sum e'_t e_t / T(n-k)} \approx F[(T-1)k, T(n-k)]$$

Where $e'_t e_t$ is SSE of the OLS regression at time t.

Results and discussion

Some of the results are presented in a general manner using tables

The data used for this work is a secondary data, which covers records on the gross output, labour, capital and the industrial cost of some 31 printing firms in Nigeria for the periods of 6 years.

Results

Table 1: Descriptive Analysis

VARIABLES	STATISTICS						
	Mean	s.e	Median	Mode	s.d	Var	c.l
Gross output	89.937	15.580	16.592	3	212.489	45151.561	30.738
Capital	14.446	1.912	3.713	1.475	26.077	680.033	3.772
Labour	112.226	14.312	39	14	195.189	38098.89	28.236
Industrial cost	32.064	5.656	7.102	7.102	77.137	5950.097	11.15846

Key: s.e = standard error
s.d = standard deviation
var = variance
c.l = confidence limit

Table 1 gives the summary of the entire data.

Table2: Results obtained on the Test of overall regression (pooled OLS)

Source	SS	Df	MS	F-ratio
Model	429.631	3	143.210	644.42
Residual	40.446	182	.222	
Total	470.077	185	2.541	

$R^2 = 0.9140$

From the above results, the computed value of F(3,182) = 1.46 is significant indicating that the model fits the data well. The fitted model is given as;

$$\hat{Y} = 0.1205 + 0.2612X_1 + 0.1358X_2 + 0.7719X_3$$

Se = (0.1602) (0.0586) (0.0412) (0.0489)

t = (0.75) (4.46) (3.29) (15.79)

The coefficients are: $\alpha = .1205$
$\beta_1 = .2612$
$\beta_2 = .1358$
$\beta_3 = .7719$

The coefficient on the constant term, "α" is obviously significant. This is the intercept for the regression line. It is the default predicted value of dependent variable when all of the other variable equal zero.

Significance is typically measured by t-statistic. From the result, t is statistically significant in all the depended variables (with t>2).

The model fits the data well with p<.0000, which indicates that the coefficients are significant at 99.99 +% level, and R-squared = .9140.

Testing the linear hypothesis:

After running the overall regression above, we can proceed with tests of linear hypothesis on the covariates.

To test;

1. Ho: 1nl = 0
   F (1, 182) = 19.86 Prob>F = 0.0000
   The result means that labour (in logs) are significantly different from zero at 0.000%
2. Ho: 1nk = 0
   F (1, 182) = 10.85
   Prob>F = 0.0012
   This result also means that capital (in logs) is significant different from zero at 1.2%
3. Ho: 1nc = 0
   F (1, 182) = 249.25
   Prob>F = 0.0000
   This means that industrial cost (in logs) is significantly different from zero at 0.00%

To test the null hypothesis that all covariates are zero,

4. Ho: 1nl = 0
   1nk = 0
   1nc = 0
   F(3, 182) = 644.42
   Prob> F = 0.0000

This means that we cannot accept the hypothesis of significance at 0.00%, i.e. some of them are significantly different than zero at this level.

Test for Heteroscedasticity:

The result obtained in testing constant variance using Breusch-Pagan/cooked-Weisberg test for heteroscedaticity based on the fitted value of Y is given below:

$\chi^2_1 = 0.28$

Prob>χ^2 = 0.5940

This indicates an insignificant result and hence the null hypothesis of constant variance is not rejected. We may conclude that there is a constant variance across the variables.

Estimation of fixed effects regressions:

Table 3: Within Effect Regression (without intercept)

Source	SS	Df	MS	F-ratio
Model	1829326.83	3	609775.61	362.70
Residual	307660.94	183	1681.207	
Total	2136987.76	186	11489.18	

$R^2 = 0.8560$ F (3, 183) = 1.46

Coefficients obtained are: $\beta_1 = 0.2494$

$\beta_2 = -0.2362$

$\beta_3 = 2.0139$

The above coefficients are called the "within" estimators because it relies on variations within individuals rather than between individuals.

The model is fitted thus

$\hat{Y} = 0.2494X_1 - 0.0236X_2 + 2.0139X_3$

se = (0.5529) (0.31324) (0.0730)

t = (4.51) (-0.08) (27.58)

This output tests the null hypothesis that all dummy parameter are zero.

From the result, X_1 and X_3 are statistically significant since t>2, while X_2 is statistically insignificant with t < 2.

Within effect Regression (with intercept) group variable: firm.

R^2: Within = 0.7346 F(3, 152)

Between = 0.9622 corr(u_Xb) = 0.5363

Overall = 0.9132

The group variable here is the firm and the there are 31 groups and 6 observations per group.

The coefficients are positively related and are fitted as:

$\hat{Y} = 0.3402 + 0.2858X_1 + 0.0832X_2 + 0.6676X_3$

$\sigma_u = 0.3522$

$\sigma_e = 0.4048$

$\rho = 0.4309$ (this is the fraction of variance due to u_i)

This tests the hypothesis that all dummy parameters are zero.

Ho: all dummies is zero

From the above result, F(30, 152) = 3.16 Prob>F = 0.0000

The null hypothesis is therefore not rejected.

Within effect regression (grouped variables; years):

In this case, we consider the within regression when the group variable is 'years'. There are 6 groups and 31 observations per group.

R^2: within = 0.9159 F(3, 177) = 642.65

Between = 0.9388

Prob>F = 0.0000

Overall = 0.91388

corr(u_i, Xb) = -0.1276

The coefficients are also positively related and fitted as;

$\hat{Y} = 0.1678 + 0.2345X_1 + 0.1413X_2 + 0.7935X_3$

σ_u = 0.0921
σ_e = 0.4706
ρ = 0.0369 (fraction of variance due to u_i)
Ho: All dummy variables are zeros
F(5, 177) = 1.12 Prob>F=0.3510

Comparing the above two results (i) the case where 'firms' is the group variable and (ii) where 'years' is the group variable), we examine the following:

1. There is a positive correlation between u_i and Xb when the group variable is firms, while a negative correlation or no correlation when the group variable is years.
2. R^2within is better of when years is the group variable or when the number of groups is less (i.e. 6) in this case.
3. In both cases, the null hypothesis of zero dummy variables is not rejected.
4. In both cases, all the coefficients are statistically and thus have larger degree of freedom, smaller MSE, and smaller standard errors of parameters.

Conclusion

Significance can be typically measured by t-statistics, typically, a t-statistic above 2 or below -2 is considered significant at the 95% level.

The degree of freedom (d.f) is number of observations minus number of variables
The p-value is a matter of convenience for us. This tells us at what level our coefficients are significant. If it is significant at the 95% level, then we have p < 0.05. if it is significant at the 99% level, then p < 0.01. Conclusively, from the above findings, it is recommended that in any application with panel data, fixed effect (FE) model is more appropriate.

Reference:

Okoronka, U.A. (2004). Model-Based Instructional Strategies as Determinants of Students' Learning Outcomes in Secondary School Physics. Unpublished PhD Thesis University of Ibadan.

Baltagi, B. H (2001) Econometric Analysis of Panel Data; John Wesley &Sons, New York.

Chikwendu .C.U. The structure of Nigeria industries and the utilization of scientific and technological manpower. The Nigerian Journal of Economics and Social Studies.

Cincera, M. (2002). Econometrics of panel Data (2009) http://homepages.ulb.ac.be/~mcincera/cours/panel.html

Damodar, N. and Gujarati, S. (2007). Basic Econometrics; Tata McGraw-Hill, Delhi

Duncan, A. (LID25) Cross Section & panel Data Econometrics.

Greene, H. (2000). *Econometric Analysis* by Dorling Kindersley (India)

Gujarati, D. (1978). Basic Econometrics, London; Mc Grill Hill

Hsiao, C. (1986). Analysis of Panel Data; Cambridge University press, Cambridge

Ibrahim, A. (2009). Econometric Analysis of the Gross Output of Some Printing Industries: An application of Panel Data Analysis. Unpublished M.Sc. project University of Ibadan.

Jeeshim and Kuscc 625 (2005) Linear Regression Models for Panel Data using STATA

Johnson J. Econometric Methods, New York, Mc Grill Hill.

Koutsoyiannis, A. (2003). Theory of Econometrics; Palgrave, New York

Omotosho, M.Y. (2000). Econometrics "A Practical Approach" Yosode Book Publishers, Ibadan

Robert, Y. (2003). A premier for Panel Data Analysis