European Journal of Statistics and Probability
Vol.8, No.3, pp, 35-38, November 2020
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)

# APPLICATION OF LEAST SQUARE DUMMY VARIABLE (LSDV) IN ESTIMATION OF COMPENSATION OF EMPLOYEE

**[1]Okoroafor, I. B.; [1]Okechukwu, C. N.; [1]Anuonye, R. C.; and [1]Uka, C. O.**
*[1]Department of Statistics, Federal College of Agriculture, Ishiagu, Ebonyi State.*
Corresponding author: ok.iykeben@gmail.com

**ABSTRACT**: *This research was conducted to estimate compensation of employee using least square dummy variable (LSDV) regression model. The data used in this work were secondary data sourced from National Bureau of Statistics (NBS) from 1981 to 2006. The variables considered were compensation of employee as the dependent variable, fixed capital, price of goods, tax and surplus as the independent variables. The data were analyzed using (STATA 13). The results obtained revealed that F-value of 3874.05 was statistically high suggesting the overall model was good fitted. The $R^2$ -value 0.9989 was also high which indicated that 99.89% of the total variation was accounted for by the independent variables included in model while the remaining 0.11% unexplained was accounted for by the white noise. Again, all the differential intercept coefficients have negative signs. Also, several differential slope coefficients have negative signs which implied that they were negatively related to compensation. Again, the result revealed that compensation is not statistically significantly related to fixed capital, price, tax and surplus. However, none of the differential slope coefficients is statistically significant. Of all the three differential intercept coefficients only $\alpha_4$ was statistically significant. Since none of the differential slope coefficients was statistically significant, it concluded that the differential slope coefficients are not different from the slope coefficient of the base/comparison group (power sector.*

**KEYWORD**: compensation, dummy variable, panel data, fixed effect and employee

## INTRODUCTION

Panel data, also called longitudinal or cross sectional time series data are data where multiple cases are observed at two or more time periods. There are two kinds of information in panel data; the cross sectional information reflects in the differences between subjects and time series or within-subject information reflects in the change within subjects over time (Okoroafor *et al* 2012). Panel data regression techniques allow one to take advantage of these different types of information. Panel data regression model developed for situations where the error term $\varepsilon_{it}$ is assumed to vary non-stochastically over i and t, making it a fixed effect model(Nwabueze *et al* 2012). In panel data analysis, the term fixed effects estimator, also known as the within estimator is used to refer to an estimator for the coefficients in the regression model. Fixed effects models have been applied to social and economic problems see (Ahmed and Sobhi 2009), Baltagi (2008), Treisman (2000) and (Hsiao and Kamil 1997).

## THEORETICAL FRAMEWORK

Fixed effect (least square) model was considered in this paper. Statistically, fixed effect model represents the observed quantities in terms of explanatory variables where the quantities are treated

as non-random. It varies non-stochastically over i and t. If we assume fixed effects, we impose time independent effects for each entity that are possibly correlated with regressors (Gujarati 2006, Gujarati 1996) and (Gujarati and Porter 2009).
The general fixed effect model is given as:

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \quad (1)$$

Where $\varepsilon_{it} = \varepsilon_i + v_{it}$ and $\varepsilon_i$ is are individual unit specific time-invariant effects. The subscript i on the intercept term suggest that the intercepts of the units may be different which may be due to special features of each of the units (four sectors).

## ANALYTICAL FRAMEWORK

The Least Squares Dummy variable model was used to analyzed the data. A dummy variable is a binary variable that is coded to either 1 or 0 commonly used to examine group and time effects in regression analysis. Least square dummy variable is a way to take into account the individuality of the sectors (Okoroafor 2012). This is done by letting the intercept vary for each sector but still assume that the slope coefficients are constant across the sectors or time periods (Hsiao 2003). Here, least square dummy variable regression model was applied to examine compensation of employee function when all the coefficients vary across individuals (the sectors and time periods). Consider:

$$C_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_1 F_{1it} + \beta_2 T_{2it} + \beta_3 P_{3it} + \beta_4 S_{4it} + \varepsilon_{it} \quad (2)$$

Where $D_2 = 1$ if the observation belong to telecommunication sector and 0 otherwise; $D_3 = 1$ if the observation belong to transportation sector and 0 otherwise; $D_4 = 1$ if the observation belong to education sector and 0 otherwise. $\alpha_1$ represents the intercept of power sector while $\alpha_2$, $\alpha_3$ and $\alpha_4$ are the differential intercept coefficients, telling us how much the intercepts of telecommunication, transportation and education differ from the intercept of power, which is our comparison sector. When the intercepts and slope coefficients are assumed to be different for all the sectors, our least square dummy variable model in equation (2) can be extended to take care of this situation. Multiplying each of the sector dummies by each of the sector sample units, equation (2) becomes:

$$C_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_2 F_{2it} + \beta_3 T_{3it} + \beta_4 P_{4it} + \beta_5 S_{5it} + \rho_1(D_{2i}F_{2it}) + \rho_2(D_{2i}T_{3it}) + \rho_3(D_{2i}P_{4it}) + \rho_4(D_{2i}S_{5it}) + \rho_5(D_{3i}F_{2it}) + \rho_6(D_{3i}T_{3it}) + \rho_7(D_{3i}P_{4it}) + \rho_8(D_{3i}S_{5it}) + \rho_9(D_{4i}F_{2it}) + \rho_{10}(D_{4i}T_{3it}) + \rho_{11}(D_{4i}P_{4it}) + \rho_{12}(D_{4i}S_{5it}) + \varepsilon_{it} \quad (3)$$

Where the $\rho$'s are the differential slope coefficients while $\alpha_2$, $\alpha_3$ and $\alpha_4$ the differential intercepts. If one or more of the $\rho$ coefficients are statistically significant, then one or more of the slope coefficients are different from the base group.

## PRESENTATION OF RESULTS

| Source | SS | df | MS | | | |
|---|---|---|---|---|---|---|
| | | | | Number of obs | = | 104 |
| Model | 6.95E+10 | 19 | 3.00E+00 | F( 19, 84) | = | 3874.05 |
| Residual | 7.93E+07 | 84 | 3.40E+10 | Prob > F | = | 0.0000 |
| Total | 6.96E+10 | 103 | 3.72E+10 | R-squared | = | 0.9989 |
| | | | | Adj R-squared | = | 0.9986 |
| | | | | Root MSE | = | 971.52 |

| C | Coefficient. | Std. Err. | t | P>t | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| F | -.9999979 | 4.303101 | -0.23 | 0.817 | -9.557185 | 7.55719 |
| T | 1.000013 | 1.137227 | 0.88 | 0.382 | -1.261487 | 3.261513 |
| P | .9999485 | 4.271207 | 0.23 | 0.815 | -7.493815 | 9.493712 |
| S | .0000234 | 2.152762 | 0.00 | 1.000 | -4.28098 | 4.281027 |
| $D_2$ | -.0037172 | 585.4803 | -0.00 | 1.000 | -1164.296 | 1164.288 |
| $D_3$ | -.0053244 | 864.7985 | -0.00 | 1.000 | -1719.752 | 1719.741 |
| $D_4$ | -2503.564 | 641.5814 | -3.90 | 0.000 | -3779.419 | -1227.709 |
| $\rho_1$ | -.0000144 | 4.551924 | -0.00 | 1.000 | -9.052015 | 9.051986 |
| $\rho_2$ | -2.00002 | 2.743221 | -0.73 | 0.468 | -7.455216 | 3.455177 |
| $\rho_3$ | .0000593 | 4.565379 | 0.00 | 1.000 | -9.078697 | 9.078816 |
| $\rho_4$ | -.0000225 | 3.115443 | -0.00 | 1.000 | -6.195422 | 6.195377 |
| $\rho_5$ | 3.92E-07 | 4.34935 | 0.00 | 1.000 | -8.649159 | 8.64916 |
| $\rho_6$ | -2.000013 | 1.147246 | -1.74 | 0.085 | -4.281437 | .2814108 |
| $\rho_7$ | .0000519 | 4.273799 | 0.00 | 1.000 | -8.498866 | 8.49897 |
| $\rho_8$ | -.0000323 | 4.229612 | -0.00 | 1.000 | -8.411081 | 8.411016 |
| $\rho_9$ | 1.71516 | 4.313469 | 0.40 | 0.692 | -6.862645 | 10.29297 |
| $\rho_{10}$ | -1.802769 | 1.1379 | -1.58 | 0.117 | -4.065608 | .4600688 |
| $\rho_{11}$ | -.1976739 | 4.271386 | -0.05 | 0.963 | -8.691793 | 8.296446 |
| $\rho_{12}$ | -.8605286 | 14.8424 | -0.06 | 0.954 | -30.37627 | 28.65521 |
| _cons | .0028512 | 458.8676 | 0.00 | 1.000 | -912.5057 | 912.5114 |

## DISCUSSION OF RESULTS

The result of the table above revealed that F-value of 3874.05 was statistically high suggesting the overall model was good fitted. The $R^2$ -value 0.9989 was also high which indicated that 99.89% of the total variation was accounted for by the independent variables included in model while the remaining 0.11% unexplained was accounted for by the white noise. Again, all the differential intercept coefficients have negative signs. Also, several differential slope coefficients have negative signs which implied that they were negatively related to compensation. Again, the result revealed that compensation is not statistically significantly related to fixed capital, price, tax and surplus. However, none of the differential slope coefficients is statistically significant. Of all the three differential intercept coefficients only $\alpha_4$ was statistically significant. The slope coefficient of fixed capital is –

0.9999979 for power sector; for telecommunication sector it is –1.0000123(–0.9999979 – 0.0000144), that of transportation sector is –3.0000179(–0.9999979 –2.00002) and – 0.9999386(– 0.9999979 +0.0000594) for education sector. Finally, since none of the differential slope coefficients was statistically significant, it means that none of the slope coefficients is different from the slope coefficient of the base/comparison group (power sector).

## CONCLUSION

Obviously**,** least square dummy variable regression model (LSDV) allows for heterogeneity among sectors by allowing each sector to have its own different intercept value. The difference may be due to some special features of each sectors such as managerial style, policies and programs or the type of market each sector is serving.

References

Ahmed, H. Y., and Sobhi, M. R. (2007), A Comparative study for Estimation Parameters in Panel Data Model, Journal Econometrics pp 1 - 15.

Baltagi, B. H. (2008), Econometrics Analysis of Panel Data 4th ed. John Wiley and sons Ltd.

Ayadi, J. O., (2002), "Foreign Direct Investment and Economic Growth in Nigeria: A seemingly unrelated model". African Review of Money, Finance and Banking, Supplementary Issue of Savings and Development. Milan Italy pp5 – 25.

Gujarati, D. N. (2006), "Essentials of Econometrics, 3rd ed. Mcgraw-hill, New York.

Gujarati, D. N., and Porter, D. C. (2009), "Basic Econometrics, 5th ed. Mcgraw-hill, New York.

Hsiao, C., (2003), Analysis of Panel Data 2nd ed. Cambridge University Press.

Hsiao, C. and Kamil A. T., (1997), A Panel Data Analysis of Liquidity Constraints and Firm Investment. Journal of America Statistics Association, 92, 455 – 464.

National Bureau of Statistics, (2007), National Accounts of Federal Republic of Nigeria, 27 – 52.

Nwabueze, J. C., Okoroafor, I. B. and Okoroafor-Iyke, F. E., (2012) Estimation of compensation of Employee Function: Panel Data Analysis of National Accounts of Nigeria, Indian Journal of Education, Information Management Vol. 1(4) 152 – 156.

Okoroafor, I. B., Nwabueze, J. C. and Okoroafor-Iyke, F. E., (2012), Panel Data Analysis of National Accounts of Nigeria: Random Effect Model Approach, Indian Journal of Education, Information Management Vol. 1(4) 170 – 173.

Treisman, D., (2000), The Causes of Corruption: Across-national Study Journal of Public Economics, 76, 399 – 457.