PAK Publishing Group
Growing Knowledge for Future

Proceedings Book of ICETSR, 2014, Malaysia
Handbook on the Emerging Trends in Scientific Research
ISBN: 978-969-9347-16-0

A Spatial Econometrics Analysis of Educational Distribution and Regional Income Disparities in Nigeria

Habibu Mohammed Umar
School of Economics, Finance and Banking, College of Business, University Utara Malaysia
Russayani Ismail
School of Economics, Finance and Banking, College of Business, University Utara Malaysia
Lim Hock Eam
School of Economics, Finance and Banking, College of Business, University Utara Malaysia

Abstract

Understanding the determinants of regional economic disparities is very important for designing an effective policy framework to address regional inequalities and their disruptive potentials. The role of educational attainment in the regional income determination has been well documented in the economic literature. So far, little attention was given to the importance of educational distribution in regional income determination. To contribute in this respect, this paper developed a regional production function model that incorporates education inequality as a determinant of regional income disparity in Nigeria. Using micro data from the recent Living Standard Measurement Survey (LSMS) on Nigeria, we calculate inequality of education using Theil index. Using a cross-sectional spatial econometric approach, we found evidence that equitable distribution of education has significant and positive effects on regional income level. Interestingly, the measure of the education distribution out-performs that of educational attainment in the model. The results also confirm the role of education inequality in accounting for regional income differences in Nigeria. We conclude that investing for equitable distribution of education will be very effective policy strategy, both for improving regional economic performance and regional economic convergence in Nigeria.

1. Introduction

Within country regional economic disparities rose as a critical issue of regional development policies, as researchers and policy makers alike noticed the increasing level of inequality in the presence of drastic national economic growth. Regions within a national economy do not usually have equal potentials for economic growth and development. Where the disparity in the potentials is large, regional disparities tend to persist and may even increase if not mitigated by necessary policy interventions. Thus, governments' intervention, at both national and regional level is required to minimize the disparity and its disruptive potentials. While variations in regional economic performance are recognized in developed and developing countries alike, the problems of regional inequality are most acute in the developing countries. This is because, in most developing countries, geographical divides reflect ethnic or racial dichotomy which are prone to rising intense political tension. Similarly, the disequilibrating effects of regional economic disparity tend to be greater during the early stages of development, thus retarding the realization of other national development objectives.

There exists regional disparity in Nigeria, particularly with respect to socio-economic indicators; such as per capita income, unemployment rate, educational attainment, intra-regional inequality (see, for example, UNDP report of 2012 on Nigeria; Umar, Russayani & Abdul-Hakim, 2013). The economic indicators of Poverty, unemployment and inequality in Nigeria have a strong regional concentration, showing significant levels of regional disparity in the country. The recent Nigeria living standards measurement survey of 2013 by the country's Bureau of statistics (NBS) shows variations in average income levels of the households as well as per capita GDP level across regions in the country with the northern region having the lowest average household income and GDP per capita levels. The survey also reveals that relative poverty was most apparent in the northern part of the country, with the north-west and north-east of the country having poverty rates as high as 77.7% and 76.3% respectively, compared to the south-west with only 59.1% and the trend continuous with almost all socio-economic indicators. This scenario, if not checked, is capable of planting distrust and hopelessness in the minds of people in the disadvantaged regions of the country which can consequently be counterproductive to the nation's economic growth. In their research, Wilkinson and Pickett (2011) illustrated the very strong positive relationship between income disparities and social problems in 29 OECD countries and concluded that the expanded income gap could cause social conflicts in any social setups.

In regional economics literature, a number of explanations have been put forward to explain the observed variation in regional economic performance. The commonly identified forces typically involve the interplay of geographic, historical, socio-economic, endowment of resources and institutional factors which can bring or constrain development in the lagging regions depending on their availability or otherwise (see Kitson, Martin & Tyler, 2004; Trendle & Pears, 2004; Takahashi, 2007; Rodriguez-Pose & Tselios, 2010). Similarly, some explanations were offered from the political economic view-point that regional differences in economic outcomes can stem from long-standing power imbalances between the advantaged and the disadvantaged regions, allied to institutional weaknesses and racial disadvantages (Mancini, 2009). However, the differing economic performance and socio-economic divide between the north and the south in Nigeria may not be unconnected with the differences in the level of educational attainment and possibly its distribution which can be attributed to the differing historical experiences of the two main regions (i.e. northern region and southern region) especially with respect to education. Mustapha, (2005) pointed out "the misguided colonial education policy in northern Nigeria and different levels of ethnic receptivity to western education, produced a huge development gap between the southern and northern regions of Nigeria". Notwithstanding, the interplay of environmental factors peculiar to the regions and other historical factors, the persistence and often the widening regional development gap in Nigeria also raises serious concern about the effectiveness of post-independence national policies in producing an inclusive society.

The paper is aimed at shedding more light on the sub-national level of educational inequality and its possible effects on regional economic performance and disparity in Nigeria. The focal point is to examine whether differences in regional educational attainment and its distribution determine the regional income disparity in Nigeria. To achieve that goal, the paper rises and addresses the following questions: Does educational distribution matter for regional economic performance? How much regional disparity is due to differences in the regional distribution of education in Nigeria? In this paper, regional education is measured by average and inequality levels, so as to discriminate between endowment (Attainment) and distribution (Inequality) in education as inputs in the model and also compares the magnitude and significance of their coefficients. The findings contribute to the empirical literature on regional economics and economics of education by providing evidence on the role of educational attainment and its distribution in regional income determination and also showing which of these factors prevails in shaping regional income disparity within a country.

The paper is organized as follows: section 2 discusses the theoretical and empirical literature on the effect of education on regional economic outcomes. Section 3 presents the data and describes the econometric approach used. Section 4 presents the regression results. Section 5 concludes.

2. Theoretical and Empirical Considerations

The economic literature had, for a long time, emphasized the importance of human capital as a key driver of economic performance (Lucas, 1988 and Barro, 1991). However, imbedded in human capital stock of a nation lies the level of educational attainment of the nation. This implies that, education is a leading component of human capital and it is well argued that the level of education drives economic performance because it increases the ability to create new technologies or adapt and implement the existing technology. It is further argued that education promotes greater political participation among the populace and also minimizes crime rate (Becker, 2009). Focusing on regional impact of education, Polasek, Schwarzbauer & Sellner, (2010) promote a hypothesis saying policies that could increase education could be very good growth strategies for sub-national economies.

However, regions tend to exhibit differences in economic performance as some regions outperform others in terms of output, income, economic growth rate, and general wellbeing of the populace. According to Nijkamp & Abreu, (2009), regional economic disparity is normally determined by a multiplicity of factors such as natural resource endowments, quantity and quality of labour, capital availability and access, productive and overhead investments, entrepreneurial culture and attitude, physical infrastructures, technological infrastructure and progress, openness, and effective public support systems. Alone this line, some studies specifically investigated factors that account for differences of economic performance across regions. For example, Takahashi, (2007) Investigated the sources of regional income disparity in Vietnam focusing mainly on the role of human capital and land endowments. The author classified the country into two regions; namely, Red River Delta located in the north (the RRD) and Mekong Delta located in the south (the MKD). The findings suggest that difference in returns to education and assets (land) rather than the difference in asset holding or stock of human capital (educational attainment) are the leading factors to account for economic disparity across the regions. Similarly, Ledyaeva & Linden, (2008) applied the modified Barro and Sala-i-Martin's framework to the question of unequal regional development. Their focus is on the determinants of regional per capita income growth in Russia for a period of ten years (1996-2005). They utilized both panel and cross sectional data. The results suggested that initial level of region's economic development; domestic investment and exports are the most important drivers of regional economic performance in Russia.

3. Data and Methods

To examine the role of educational distribution on regional economic performance in Nigeria, we used cross-sectional data extracted from the revised Nigeria Living Standards Measurement Survey (LSMS, 2013) provided by the World Bank. The survey was carried out in partnership with World Bank (WB), the Nigeria Bureau of Statistics (NBS) and other relevant agencies. The survey was meant to collect economic information on households, their characteristics, welfare as well as their agricultural activities. Using the LSMS data, we extracted the relevant information and computed the educational distribution (inequality) using Theil index. The value of this measure lies between 0 and 1; a value closer to one meaning higher level of inequality and vice versa. The index represents the least fraction of total education that must be redistributed to achieve perfect equality. However, going by what is available on the data set, we used per capita expenditure as a proxy of income as it was done in a number of studies including Takahashi, (2007) and Barrett and Reardon, (2001). Appendices 1&2 display some basic descriptive statistics of the main and control variables used in this study.

The main novelty of this study are the use of micro data to measure regional educational distribution and examine its role in regional income determination and disparity in Nigeria, and the consideration of the problem of spatial dependence known as spatial autocorrelation in economics literature. Spatial dependence can be expected in a data set with observations that are collected from different locations. The problem of spatial interdependence could be more severe in cross sectional studies that involve micro level data of households or firms, rather than studies that involve macro data. This is because, in studies that involve micro-level data, the data sets normally contain large numbers of observations which are characterised by spatial relationships that are better measured as a declining function of distances between agents (Bell & Bockstael, 2000). The problem arises when an observation on a variable in a particular location is strategically influenced by the same variable observations of the neighbouring areas (Anselin, 2001). For example, expenditure decisions of an area can be influenced by the expenditure decision of the neighbouring locations. In a study that covers large geographical space, detection of the spatial autocorrelation problem could be very important due to the potential econometric problems arising from it. Presence of spatial autocorrelation leads to the violation of some statistical assumptions used in the traditional analysis approach i.e. the assumptions of uncorrelated error terms and independent observations (see Anselin & Rey, 1991; LeSage and Pace, 2009 for details).

However, when applying spatial models, it is important to detect the nature and pattern of the spatial dependence so as to be able to know which appropriate model will suit the data. There are mainly two types of spatial dependence identified in the spatial econometrics literature; spatial lag and spatial error (Anselin, 1988). In the former, the dependent variable in a particular place is said to be affected by the independent variables in both the place and other places as well, while in the later the error terms across different spatial units are correlated generally due to omitted variables which are themselves spatially correlated. To address the problem of spatial dependence as proposed by Anselin (1988), two spatial models that contain one type of spatial interaction effect are normally put to use; spatial lag model and spatial error model. The first model contains a spatially lagged dependent variable as appeared in equation (1), while the second model incorporates a spatially autoregressive process in the error term as can be seen in equation (2).

The spatial lag model takes the form:

$$Y = \rho W y + X\beta + \varepsilon \qquad (1)$$

While the spatial error is defined as:

$$Y = X\beta + \mu \qquad (2)$$

$$\mu = \gamma W \mu + \varepsilon$$

Where Y is a vector of *N* observations on the dependent variable; W is an $N \times N$ spatial weights matrix; ρ *and* γ are spatial autoregressive parameters; X is an $N \times K$ matrix of observations on the exogenous variables, with associated $K \times 1$ regression coefficient vector β, ε is a vector of residuals, μ is an independently and normally distributed error term with a constant variance, and W is an $N \times N$ spatial weight matrix. According to Anselin (2002), models of this nature require specialised estimation techniques, like maximum likelihood, GMM or instrumental variables. In this study, we implemented maximum likelihood estimation method using GeoDa statistical software package.

The technical precondition for the use of spatial models is the availability of spatial weight matrix (W). The spatial weights matrix is an $N \times N$ nonnegative matrix that describes the proximity of every observation (spatial unit) with the rest of observations that are considered in the sample. A location (observation) appears as both row and column of the matrix as the nonzero matrix elements of W_{ij} that indicates the relation between the observations (row) i and (column) *j*. By convention, observations of the same location are excluded, since no spatial unit can be viewed as its own neighbour. Therefore, the diagonal elements of the matrix are all set to zero (W_{ii} = 0). There are different types of weights matrices that are used in spatial modelling, depending on the nature and phenomenon being studied. According to Bell and Bockstael, (2000), if the units of observation are households or firms the spatial relationship will be best captured by considering distance decay effects between points, because any one decision unit is small relative to the geographical sphere of influence. Going by the suggestion of Bell and Bockstael, (2000), we chose 'Inverse Distance Decay' type of spatial weight matrices. This type of matrix provides greater weighting to observations that are closer to each other than those that are further apart. In the matrix, the weighting between location *i* and *j*, with $W_{ij} = 1/d_{ij}$, for $d_{ij} < D_{max}$, and $W_{ij} = 0$ otherwise, where d_{ij} is the distance between the centroids of location *i* and *j* and D_{max} is a threshold distance.

Various testing procedures are used to capture spatial dependency in data, and the most widely used procedures in the literature for testing spatial autocorrelation are Moran's I statistics and Lagrange Multiplier (LM) tests. The later has an important advantage over the former as it allows testing to know which model best suits the data between spatial lag and spatial error specifications.

Therefore, we apply both the classic LM tests proposed by Anselin (1988) and the robust version of it proposed by Anselin, Bera, Florax & Yoon (1996). Both the two tests are based on the residuals of the OLS model and follow a chi-squared distribution with one degree of freedom.

3.1. Econometric Approach

In this study, regional income (y) is hypothesized to be the function of the following regional features; the demographic characteristics of the region (Demo) such as Age and household size, the regional level of Educational attainment (Edu), the regional educational inequality (EduIneq), the regional industry structure (Indstry) and the general economic condition of the region (Ecr), in this case proxied by the regional per capita GDP.

Hence, the model is specified as:

$$lnY_{ij} = \beta_0 + \beta_1 Demo_{ij} + \beta_4 Indstry_{ij} + \beta_5 Ecr_{ij} + \beta_4 Edu_{ij} + \beta_5 EduIneq_{ij} + \mu_{ij} \qquad (3)$$

The equation (3) is estimated by OLS (Ordinary Least Square) and Spatial Error Model. In addition, the Blinder-Oaxaca Decomposition analysis is performed to gain further insight on the impact of regional education inequality on income disparity.

4. Regression Results

The first analytical step involves the estimation of the regional income model using Ordinary Least Square (OLS) technique. The next step of the analysis involves checking for the presence of spatial autocorrelation in the residuals of the estimated OLS equation, and if the hypothesis of no spatial autocorrelation cannot be rejected, then, we proceed to test whether the spatial lag model or the spatial error model is more appropriate to describe the data. The results of the OLS diagnostic test are shown on table 1.

Table-1. Diagnostics for Spatial Dependence (Autocorrelation)

Test	Stat. value	P-value
Moran's I (error)	39.642	0.000
LM lag	97.392	0.000
Robust LM lag	4.810	0.028
LM error	972.015	0.000
Robust LM error	879.434	0.000

Five test statistics that are against spatial autocorrelation are reported in the diagnostic output as shown on table1.The first one is called Moran’s I statistics. As we mentioned earlier, the Moran’s I statistic has great power in detecting spatial dependence (spatial autocorrelation), but it is less helpful in suggesting which alternative specification should be used. To this end, the Lagrange Multiplier test statistics are more useful and suggestive, as they point to the better alternative model specification to be used. The next two test statistics on the table (LM-Lag and Robust LM-Lag) pertain to the spatial lag model as the alternative, if are found to be more significant against their alternative error tests. The next two (LM-Error and Robust LM-Error) refer to the spatial error model as the alternative when are found to be significant alternatively. The results on table one suggest that we can’t reject the hypothesis of spatial dependence due to the lower probability values associated with all the tests and it can be concluded that the residuals from the OLS estimation exhibit spatial autocorrelation. Here, both the Robust LM-Lag statistic and Robust LM-Error statistics are significant with p-values (0.028 and 0.000 respectively). This suggests that a spatial error specification better suits the data and should be estimated next (see Anselin, 2005; for more details on the model selection decision rule).

Table 2 displays the estimated coefficients of the OLS and the spatial error model. The second, third and fourth columns of the table contain the estimated results of the OLS model, while the last three columns provide the coefficients and z-values of the maximum likelihood version of the spatial error model.

Table-2. OLS and Spatial Error Model Estimates

	OLS Model			Spatial Error model		
	Coefficient	t value	P value	Coefficient	Z value	P value
(Intercept)	5.0690***	117.4	0.0000	4.809***	144.130	0.000
Gender	0.0023	0.15	0.8828	0.006	0.451	0.652
Age	0.0005	1.26	0.2078	0.000	1.040	0.298
Inddmy	-0.0537***	-3.58	0.0003	-0.049***	-4.353	0.000
Sector	0.0859***	6.86	0.0000	0.107***	9.779	0.000
HHsize	-0.0403***	-13.81	0.0000	-0.026***	-14.923	0.000
Schooling	0.0234***	17.08	0.0000	0.019***	21.194	0.000
Theil	-0.2633***	-5.361	0.0000	-0.109*	-1.656	0.090
Gdppc	6.4e-06***	2.83	0.0046	1.2e-05***	6.839	0.000
Rgd	-0.236***	-3.55	0.0004	-0.235***	-3.35	0.005
Lambda	n.a.	n.a.	n.a.	0.767***	23.908	0.000
N	**4979**			**4979**		
R^2	0.296			0.369		
Prob>F	0.000			n.a.		

Note: ***, ** and * represent 1%, 5% and 10% level of significance respectively

The OLS model and the spatial error version of the model, as shown on table 2, can explain about 30-37% of the cross regional variation in average income level in Nigeria. The results indicate two variables as insignificant in both models. The regional average age of the households (Age) and the proportion of women as house head (Gender) were not significant in either the OLS or the spatial error model. Interestingly, all the remaining variables are found to be significant with the expected signs in both models. The proportion of the working population that is in agriculture (Inddmy), household size (Hhsize), the measures of educational inequality (Theil) and Rgd (regional dummy variable) have a significant negative sign on their coefficients, meaning that all are associated with low regional incomes. The remaining variables that comprise educational attainment (schooling), the proportion of the working population living in urban areas (sector) and GDP per capita (Gdppc) are all significant with positive signs on coefficients, meaning that higher values of these variables are associated with high regional incomes.

However, the relative sizes of the coefficients are very important. As shown on table 2, the measure of educational inequality out-performed that of educational attainment in both models. Although, it has lower significance level, but its coefficient is higher in absolute terms (i.e. -0.109 against 0.019) than that of educational attainment suggesting that one unit (i.e one percentage point in this case) increase of the measure of educational inequality is associated with about 11% decrease in regional income. By contrast, regional income level increases by only 1.9 % for every additional year of schooling. Thus, distribution of education is more important in regional income determination, than a skewed educational attainment for the few portion of the populace.

To know how much of regional income disparity is explained by the differences in the levels of regional educational inequality, we used decomposition method of analysis popularized by Blinder and Oaxaca (see Jann, 2008 for details). In the analysis, we include regional educational attainment and regional household size as additional predictors. As it can be seen from Appendix 3, the decomposition output reports the predicted mean of each region and their difference in the first panel. In the sample, mean of the log per capita income is 4.88 for the southern region and 4.71 for the northern region, showing an income gap of 0.17. The three variables used explained up to 20% of the regional income gap in Nigeria. Therefore, adjusting these factors in the northern region to the levels obtainable in the southern region can ameliorate the regional income gap.

The second panel of the decomposition output shows the contribution of each variable to the regional income gap. Educational inequality accounts for about 8.2% of the explained regional income gap in Nigeria. However, differences in regional educational attainment and the regional average household size account for about 6.4% and 6.0% of the regional income disparity respectively. This is consistent with the findings of the regression analysis that shows higher explanatory power of educational distribution than educational attainment in accounting for regional income disparity in Nigeria.

5. Conclusions

This paper provided an analysis on the determinants of regional income level and disparity in Nigeria, with emphasis on the role of educational distribution. We used school attainment levels of individuals in the data set and calculated education inequality measures using Theil index and Gini coefficient. Additionally, relevant factors available on the data set were considered in the model to see whether the potential relationship can be affected by the introduction of the additional variables. These include variables pertaining to the demographic profile, educational attainment and inequality, regional industry structure and general economic conditions of the regions. Using spatial econometric technique, the role of space is considered in the model and we were able to take care of the spatial autocorrelation problem as it may have led to biased or inefficient estimators.

Overall, our findings have shown that both educational attainment and distribution matter for regional income. Higher educational attainment is associated with higher income level and this is in line with the general belief that educational achievement has a positive relationship with measures of economic performance. Contrarily, Educational inequality negates regional income level. The negative effect of education inequality on income is robust to the inclusion of regional dummies in the model, the inclusion of educational attainment and the use of spatial error model to control for spatial autocorrelation problems. The findings of this paper show vividly that the association between inequality in education and regional income level is stronger than that of between regional income and educational attainment. Similarly, educational inequality is found to have accounted for most of the explained regional disparity in Nigeria. Finally, the results suggest that government policies carried out to promote regional economic performance and narrowing regional disparities should not only take into account the level but also the distribution of educational attainment, making access to formal education at different stages to a wider section of the population possible.

References

Anselin, L., (1988). Spatial econometrics: Methods and models. Springer, 4.

Anselin, L., (1995). Local indicators of spatial association-LISA. Geographical analysis, 27(2): 93-115.

Anselin, L., (2001). Spatial econometrics. A companion to theoretical econometrics, 310330.

Anselin, L., (2002). Under the hood issues in the specification and interpretation of spatial regression models. Agricultural economics, 27(3): 247-267.

Anselin, L. (2005). Exploring spatial data with GeoDaTM: A workbook. *Urbana*.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J., (1996). Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics, 26(1): 77-104.

Barrett, C. B., Reardon, T., & Webb, P., (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: Concepts, Dynamics, and Policy Implications. Food Policy, 26(4): 315-331

Barro, R. J., (1991). Economic growth in a cross section of countries. The Quarterly Journal of Economics, 106(2): 407-443.

Becker, G. S., (2009). Human capital: A theoretical and empirical analysis, with special reference to education. University of Chicago Press.

Bell, K. P., & Bockstael, N. E., (2000). Applying the generalized-moments estimation approach to spatial problems involving micro-level data. Review of Economics and Statistics, 82(1): 72-82.

Castelló?Climent, A., & Domenech, R., (2008). Human capital inequality, life expectancy and economic growth. The Economic Journal, 118(528): 653-677.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. Spatial Economic Analysis, 5(1): 9-28.

Fleisher, B., Li, H., & Zhao, M. Q., (2010). Human capital, economic growth, and regional inequality in China. Journal of Development Economics, 92(2): 215-231.

Jann, B., (2008). A Stata implementation of the Blinder-Oaxaca decomposition. Stata Journal, 8(4): 453-479.

Kitson, M., Martin, R., & Tyler, P., (2004). Regional competitiveness: An elusive yet key concept?. Regional Studies, 38(9): 991-999.

Ledyaeva, S., & Linden, M., (2008). Determinants of economic growth: Empirical evidence from Russian regions. European Journal of Comparative Economics, 5(1): 87-105.

Lucas, Jr, R. E., (1988). On the mechanics of economic development. Journal of Monetary Economics, 22(1): 3-42.

Mancini, L., (2009). Comparative trends in ethno-regional inequalities in Ghana and Nigeria: Evidence from demographic and health surveys. Oxford: CRISE, Documento de trabajo, 72: 18.

Mustapha, A. R., (2006). Ethnic structure, inequality and governance of the public sector in Nigeria: United Nations Research Institute for Social Development.

Nijkamp, P., & Abreu, M. A., (2009). Regional development theory: Vrije Universiteit, Faculty of Economics and Business Administration.

Perugini, C., & Martino, G., (2008). Income inequality within European regions: Determinants and effects on growth. Review of Income and Wealth, 54(3): 373-406.

Polasek, W., Schwarzbauer, W., & Sellner, R., (2010). Human capital and regional growth in Switzerland (No. 250). Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS).

Rodríguez-Pose, A., & Tselios, V., (2010). Inequalities in income and education and regional economic growth in Western Europe. The Annals of Regional Science, 44(2): 349-375.

Takahashi, K., (2007). Sources of regional income disparity in rural Vietnam: Oaxaca-Blinder Decomposition.

Trendle, B., & Pears, W., (2004). The role of education in regional income determination-a cross sectional study of small areas in Queensland. Australasian Journal of Regional Studies, The, 10(3): 383

Umar, H., Ismail, R., & Abdul-Hakim, R., (2014). Regional inequality of educational attainment in Nigeria. British Journal of Economics, Management & Trade, 4(3): 420-430.

Wilkinson, R. and Pickett, K., (2011). The spirit level. Bloomsbury Press. Available from
http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/.

Appendix

Appendix-1. Description of the key continuous variables

Variable	Description	Unit	Mean	Std. Deviation
lnPEX	Log of household per capita expenditure	'000	4.79	0.37
hhsize	Total number of household members	people	6.00	3.00
Age	Age of household head	years	49.5	15.0
Edu	Educational attainment from 0-21 (illiterate=0,….Doctorate=21)		5.87	5.73
Totexp	Total household expenditure per annum	'000	417782	408469
Theil index	Measure of educational inequality that takes a value between 0 &1 (0=perfect equality; 1=perfect inequality		0.27	0.16
GDP	States Gross Domestic product	Million	8123.9	6143.7
GDPPC	States Gross Domestic product per capita	'000	2382.8	3320.1
N=4979				

Note: Expenditure and GDP are measured in local currency (Naira).

Appendix-2. Description of some discrete variables

Variable	Description	Frequency		Percentage
Sector	Household living in Rural or Urban areas (rural=0; urban=1)	0	3,365	68.0
		1	1,614	32.0
Industry	Whether a household head is working in the agricultural sector or not (Agriculture=1; otherwise=0)	0	1,112	22.0
		1	3,867	78.0
Gender	Whether a household head is male or female (male=1; female=0)	0	740	15.0
		1	4,239	85.0
Number of observations	4979			

Appendix 3

Blinder-Oaxaca decomposition Number of obs = 4979
Model = linear
Group 1: rgd1 = 0 N of obs 1 = 2493
Group 2: rgd1 = 1 N of obs 2 = 2486

ln_income	Coef. (Robust)	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Overall						
group_1	4.882681	.007128	685.00	0.000	4.868711	4.896652
group_2	4.712083	.0074949	628.71	0.000	4.697393	4.726773
Difference	.1705981	.0103432	16.49	0.000	.1503259	.1908704
Explained	.2061769	.013659	15.09	0.000	.1794056	.2329481
Unexplained	-.0355788	.0151712	-2.35	0.019	-.0653138	-.0058437
Explained						
hhsize	.0595504	.0041815	14.24	0.000	.0513548	.067746
Edu	.0642645	.0043449	14.79	0.000	.0557487	.0727804
theil	.0823619	.0129911	6.34	0.000	.0568999	.107824
Unexplained						
hhsize	-.1102372	.0177589	-6.21	0.000	-.1450441	-.0754303
Edu	.0332144	.0097056	3.42	0.001	.0141918	.0522371
theil	-.1952181	.0283819	-6.88	0.000	-.2508457	-.1395906
_cons	.2366622	.0392835	6.02	0.000	.1596679	.3136565