

ON THE IDENTIFICATION OF CORE DETERMINANTS OF POVERTY: A LOGISTIC REGRESSION APPROACH

[*, 1]Osowole, O.I., [2]Ugbechie, Rita, [3]Uba, Ezenwanyi
1. Department of Statistics, University of Ibadan, Nigeria
2. Department of Mathematics, College of Education, Agbor, Delta State, Nigeria
3. National Root Crops Research Institute, Umahia, Abia State, Nigeria
[*]Email of the corresponding author: dosowole@yahoo.com

ABSTRACT
This study carried out a logistic regression modelling of poverty status of households in Nigeria to identify possible determinants of poverty using the 2003/2004 National Living Standard Survey (NLSS) data. This approach differs from classical regression methods that place inappropriate restrictions on the residuals of the model. The results of the logistic regression showed that household size and educational group for highest level attained by the household head were the most significant determinants of poverty. Others include sex of the household head, age in years of the household head, father's education level, father's work, mother's work, and occupation group of the household head. It strongly recommends that moderate household size and acquiring of formal education be encouraged to reduce poverty prevalence.
Keywords: poverty incidence, multidimensional concept, lack of representation and freedom, household's consumption expenditure, classical regression methods, logistic regression model

1.0 Introduction
The issue of poverty in many developing countries is a very crucial one going by its intensity, incidence and severity. The situation in Nigeria presents a paradox, because despite the fact that the nation is rich in natural resources, the people are poor. World Bank (1996) referred to this situation as poverty in the midst of plenty. In 1992, for instance, 34.7 million Nigerians (one-third of the population) were reported to be poor, while 13.9 million people were extremely poor (World Bank, 1996). The incidence of
poverty increased from 28.1 percent in 1980 to 46.3 percent in 1985. The poverty problem grew so worse in the 1990s that in 1996, about 65.6 percent of the population was poor, while the rural areas account for 69.3 percent (FOS, 1999). Recent data showed that in 2004, 54.4 percent of Nigerians were poor (FRN, 2006). Also, more than 70 percent of the people are poor, living on less than $1 a day. Similarly, Nigeria's Human Development Index (HDI) of 0.448 ranks 159th among 177 nations in 2006, portraying Nigeria as one of the poorest countries in the world (UNDP, 2006, IMF, 2005).

Poverty is a multi-dimensional concept. It is hunger, lack of shelter, being sick and not being able to see a doctor, not having access to school, not knowing how to read, not having a job, fear for the future, living one day at a time, losing a child to illness brought about by unclean water, powerlessness, lack of representation and freedom. Well-being can be termed as coming out of poverty. It may be defined as ability to function in the society in order to achieve certain functioning of beings and doings (Ahmad, 1995).

The measurement and analysis of poverty have traditionally relied on reported income or consumption and expenditure as the preferred indicators of poverty and living standards. Income is generally the measure of choice in developed countries while the preferred metric in developing countries is an aggregate of a household's consumption expenditures, Sahn and Stifel (2003). The choice of expenditures over income is influenced by the difficulties involved in the

measuring income in the developing countries. Similarly with the expenditure data the limitation is the extensive data collection which is time-consuming and costly as stated by Vyas and Kumaranayake (2006).

This study differs from the use of classical linear regression methods which are inappropriate when the dependent variable, y, takes on values of zero and one only. Further more when classical methods are used for binary dependent variable models, the implied model of the conditional mean places inappropriate restrictions on the residuals of the model. Also the fitted value of y from these linear regressions is not restricted to lie between zero and one. Therefore the logistic regression model is adopted because it has a specification which is designed to handle the specific requirements of binary dependent variable models.

2.0 Data and Methods

The 2003/04 Nigeria Living Standard Survey (NLSS) data from the National Bureau of Statistics (formerly Federal Office of Statistics) were used in this study. The sample design was a two-stage stratified sampling. The first stage involved the selection of 120 Enumeration areas (EAs) in each of the 36 states and 60 EAs at the Federal Capital Territory (FCT). The second stage was the random selection of five housing units from each of the selected EAs. A total of 21,900 households were randomly interviewed across the country with 19,158 households having consistent information (NBS, 2005). For the purpose of this study, the secondary data was first stratified into rural and urban sectors. The second stage was the stratification of the rural area based on the six geo-political zones of Nigeria namely: South West, South East, South South, North Central, North East and North West. The next stage involved the selection of all the sampled rural households in each of the geo-political zones. The data set provides detailed records on household expenditure and household characteristics. Data were collected on the following key elements: demographic characteristics, educational skill and training, employment and time use, housing and housing conditions, social capital, agriculture, income, consumption expenditure and non-farm enterprise. Some of the variables captured in the survey included sector of the country, sex of the household head, age in years of the household head, marital status of the household head, religion of the

household head, father's educational level, father's work, mother's educational level, mother's work, household size, expenditure of own produce, household expenditure on food, occupation group the household head belongs, educational group for highest level attained by the household, literacy of the household head and educational age grouping.

2.0.1 Logistic Regression Model

Logistic regression is part of a category of statistical models called generalized linear models. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous such as presence/absence or success/failure. In instances where the independent variables are categorical, or a mix of continuous and categorical, the logistic regression is preferred. The logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related, or of equal variance within each group. The relationship between the predictor and response variables is not a linear function .In logistic regression, instead, the logistic regression function is used, which is the logit transformation of θ. The logistic regression model describes the relationship between a dichotomous response variable y, coded to take values one and zero for success and failure, and the k explanatory variables x_1, $x_2,\ldots,x_k$. The explanatory variable can be quantitative or indicator variables referring to the levels of categorical variables. Since y is a binary variable, it has a Bernoulli distribution with parameter P = P(y = 1). Suppose that $Y_1, \ldots, Y_n$ are independent Bernoulli variables, and let P_i denote the mean value i.e. $P_i = E(Y_i) = P(Y_i = 1)$. The mean value P_i can be expressed in terms of the explanatory variables $x_{i,1}, x_{i,2}, \ldots, x_{i,k}$ as

$$P_i = \frac{1}{1+\exp\left[-\beta_0 - \sum_{i=1}^{k} \beta_j x_{ij}\right]} \quad (1)$$

If we apply the logit-transformation to the equation above, we get a linear relationship between logit (P_i) and the explanatory variables, that is

$$\text{logit}\left(P_i\right) = \log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{i=1}^{k} \beta_j x_{i,j} \quad (2)$$

The equation is sometimes called the logistic form of the model. Note that logit is the logarithm of the odds of success for the given values $x_{i,1}, x_{i,2}, ..., x_{i,k}$ of the explanatory variables. The parameters of this model are estimated using the method of maximum likelihood. The first order conditions for this likelihood are nonlinear. Thus, obtaining parameter estimates requires an iterative solution. Each of the binary responses represents an event with the coding of y as a zero-one variable. This coding yields a number of advantages (Greene, 1997). In logistic regression, hypotheses on significance of explanatory variables cannot be tested in quite the same way as in linear regression. The response variables here are Bernoulli distributed and exact distribution is not known. There exists fairly good approximation to the distribution of the test statistics. These are the log likelihood ratio statistics referred to as the -2log Q, where Q is the likelihood statistic and the Wald statistic.

2.0.2 Hypothesis Testing in Logistic Regression

2.0.2.1 Wald Test

The Wald test follows immediately from the fact that the information matrix for generalized linear models is given by

$$I(\beta) = \frac{(X'WX)}{\varphi} \quad (3)$$

so the large sample distribution of the maximum likelihood estimator $\hat{\beta}$ is multivariate normal, that is

$$\hat{\beta} \sim N_p(\beta, (X'WX)^{-1}\varphi) \quad (4)$$

with mean β and variance-covariance matrix $(X'WX)^{-1}\varphi$.

Tests for subsets of β are based on the corresponding marginal normal distributions.

2.0.2.2 Likelihood Ratio Tests and The Deviance

We will show how the likelihood ratio criterion for comparing any two nested models, say $w_1 \subset w_2$, can be constructed in terms of a statistic called the deviance and an unknown scale parameter φ.

Consider first comparing a model of interest w with a saturated model Ω that provides a separate parameter for each observation. Let $\hat{\mu}_i$ denote the fitted values under w and let $\hat{\theta}_i$ denote the corresponding estimates of the canonical parameters. Similarly, let $\bar{\mu}_0 = y_i$ and $\bar{\theta}_i$ denote the corresponding estimates under Ω.

The likelihood ratio criterion to compare these two models in the exponential family has the form

$$-2\log\lambda = 2\sum_{i=1}^{n} \frac{y_i(\bar{\theta}_i - \hat{\theta}_i) - b(\bar{\theta}_i) + b(\theta_i)}{a_i(\varphi)} \quad (5)$$

Assume that $a_i(\varphi) = \frac{\varphi}{p_i}$ for known prior weights p_i. Then we can write the likelihood-ratio criterion as follows:

$$-2\log\lambda = \frac{D(y,\hat{\mu})}{\varphi} \quad (6)$$

The numerator of this expression does not depend on unknown parameters and is called the

deviance:

$$D(y,\hat{\mu})=2\sum_{i=1}^{n} p_i y_i[(\bar{\theta}_i-\text{涂}\theta_i)-b(\bar{\theta}_i)+b(\theta_i)] \quad (7)$$

The likelihood ratio criterion −2 log L is the deviance divided by the scale parameter φ, and is called the scaled deviance.

For the logistic regression model used in this study, per capita household expenditure was selected as the dependent variable. This was coded into poor (1) and non-poor (0). Households with per capita expenditure less than the poverty line, z were deemed poor and those with per capital expenditure greater than the poverty line were regarded non-poor households. The poverty line was defined as 2/3 of mean per capita household expenditure.

3.0 Results and Discussion

The poverty line was obtained as N23, 734 per month. The STATA software version 10 was employed for the logistic modelling of poverty determinants. The independent variables selected were the household-based determinants of poverty as hypothesized in literature (Datt, 1998; Al-Saleh, 2000; Ajakaiye and Adeyeye, 2002). They were Sex (sex of the household head), Ageyrs (age in years of the household head), Fatheduc (father's education level), Fathwrk (father's work), Motheduc (mother's education level), Mothwrk (mother's work), Hhsize (household size), Occgrp (occupation group of the household head), Edgrp (educational group for highest level attained by the household head) and Lit (literacy of the household head). The logistic regression modelling (Table 1) showed that Sex, Ageyrs, Fatheduc, Fathwrk, Mothwrk, Hhsize, Occgrp, and Edgrp were significant while, Motheduc and Lit were insignificant at 5% level of significance.

Also, Fatheduc, Fathwrk, Hhsize, and Occgrp had positive effect on poverty status while, Sex, Ageyrs, Mothwrk and Edgrp showed negative effect on household poverty status. However both Hhsize and Edgrp exerted most influence on household poverty status. The implication of this is that poverty status of a given household increases with increase in the size of the household. Also, the poverty status of the household is worse when the household head has no education than when he has one form of education or the other.

Table 1: Results for Logistic Regression Model

Variable	Coefficient	Standard Error	Z	P> |Z|	Remark
Constant	-0.64103	0.16456	-3.9	0.000	S
Sex	-0.25712	0.04766	-5.39	0.000	S
Ageyrs	-0.01028	0.00116	-8.83	0.000	S
Fatheduc	0.04308	0.00578	7.45	0.000	S
Fathwrk	0.01141	0.00163	6.99	0.000	S
Motheduc	0.00865	0.00509	1.7	0.090	NS
Mothwrk	-0.00722	0.00194	-3.73	0.000	S
Hhsize	0.29496	0.00705	41.82	0.000	S
Occgrp	0.05594	0.00846	6.61	0.000	S
Edgrp	-0.24548	0.00995	-24.66	0.000	S
Lit	0.08366	0.04542	1.84	0.065	NS

S= significant at α = 5%, NS= insignificant at α = 5%

AIC = 1.20283; Deviance = 23021.71378; Scale Deviance = 1.20237

4.0 Conclusion and Recommendation

Binary dependent variable logistic model has been used in this study as an alternative to classical regression methods which place inappropriate restrictions on the residuals of the model. The estimated logistic regression showed that both household size and the educational group for highest level attained by the household head were the most significant determinants of poverty in Nigeria. The findings of this study agree with previous studies (Lanjouw and Ravallion, 1994; Al-Saleh, 2000; Ajakaiye and Adeyeye, 2002; and Biewen and Jenkins, 2002). Hence there is the need to encourage individuals and families to have moderate household sizes so that their children would be adequately catered for. Also the need for formal education should not be ignored. People should be adequately sensitized about the various benefits of acquiring formal education as a basic step toward the reduction of poverty prevalence in Nigeria.

REFERENCES

1. Ahmad M. (1995). "Poverty in Pakistan: Concept, Measurement, Nature, Incidence and Review of Strategies to Alleviate Poverty" Paper presented in the Seminar of Pakistan Institute

of Development Economics, Islamabad.
2. Ajakaiye, D.O. and Adeyeye, V.A.2002. Concepts, Measurements and Causes of Poverty. *CBN Economic and Financial Review. Prod. 39:4.*
3. Al-Saleh, J.2000. Poverty Assessment in Palestine: Preliminary Report. *Inter-Stat. 22: 83-106.*
4. Biewen, M. and Jenkins, S.P. 2002. Accounting for poverty differences between the United States, Great Britain and Germany. ISER Working Paper No 2002-14

5. Datt, G. 1998. Computational Tools for Poverty Measurement and Analysis. FCND Discussion Paper No. 50. Food and Consumption and Nutrition Division, International Policy Research Institute 2033k Street, N.W. Washington, D.C., 20006 USA

6. Federal Office of Statistic (FOS) (1999), Poverty and Agricultural sector in Nigeria FOS, Abuja, Nigeria.
7. Federal Republic of Nigeria (FRN) (2006). Poverty Profile for Nigeria. National Bureau of Statistics (NBS) FRN.

8. Foster, J. Greer, J and Thorbecke, E. (1984) 'A class of Decomposable Poverty Measures', *Econometrica*, 52(3): 761-776.
9. Green, W. H. (1997). "*Econometric Analysis*." Prentice Hall International, Inc., USA: 450-53
10. IMF (2005). Nigeria: Poverty Reduction Strategy Paper— National Economic Empowerment and Development Strategy. IMF Country Report No. 05/433.
11. Lanjouw, P and Ravallion, M 1994. Poverty and Household Size. Policy Research Working Paper 1332. The World Bank: Washington, DC.
12. Sahn, D.and Stifel, D (2003). Exploring Alternative measure of welfare in Absence of Expenditure Data. *Review of income and wealth*, **49**: 463-489
13. United Nations Development Program (UNDP) (2006). Beyond scarcity: Power, poverty and the global water crisis. Human Development Report 2006.
14. Vyas, S and Kumaranayake, L (2006). Constructing social-economic status indices, How to use PCA. *Health Policy and Planning*, **21**, 6, 459-468
15. World Bank (1996) "Poverty in the Midst of Plenty: The challenge of growth with inclusion in Nigeria" A World Bank Poverty Assessment, May 31, World Bank, Washington, D.C.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar