Application of Multilevel Logistic Model to Identify Correlates of Poverty in Ethiopia

Teshome Kebede Deressa

Department of Statistics, Haramaya University, P O Box 138, Dire Dawa, Ethiopia
E-mail: tashe.zgreat@gmail.com

Abstract: Implementation of multilevel model is becoming a common analytic technique over a wide range of disciplines including social and economic sciences. In this paper, an attempt has been made to assess the application of multilevel logistic model for the purpose of identifying the effect of household characteristics on poverty status in Ethiopia using household income, consumption and expenditure (HICE) survey data of 2011. Households are classified as either poor or non-poor based on the absolute poverty line set at yearly per capita consumption of Birr 3781. Accordingly, the random intercept only model indicates the existence of differences in poverty status among households across regions. The result of random intercept and fixed slope model show that the rates of poverty for households residing in Afar, Somali, SNNP, Benishangul-Gumuz and Gambela regions were higher than the average of all regions, while the rates for households residing in Harari and Addis Ababa regions were low compared to the average of all regions. The random coefficient model showed that the random effects of place of residence vary across regions in explaining poverty status. Further, this model was more appropriate to explain the regional variation than a model with fixed coefficients or empty model with random effects. Thus, researchers should take the advantage of multilevel models to identify correlates of poverty when the data structure is hierarchical like HICE survey.

Keywords: Correlates; Household; Multilevel; Poverty; Region

1. Introduction

The pursuit of a more efficient allocation of relatively scarce resources has led public decision makers in developing countries to a global reconsideration of public expenditure priorities. In this context, the analysis of poverty has always aroused the interest of researchers, public authorities and international organizations. In all economies of the contemporary world, serious objectives and priorities of public decision makers are to fight poverty, to improve the conditions of life for people and to reduce the gap between the social strata.

Poverty has a series of contested definitions and complex arguments that overlap and at times contradict each other. It is differently seen as a big or small phenomenon, as a growing or a declining issue, as an individual or a social problem, as a country or a regional problem and as urban or a rural problem (Chaudhry, 2003). This implies that the depth and dimension of poverty vary according to the country's situation. It is multi-dimensional and has to be looked at through a variety of indicators. Different indicators showed varied levels of poverty status for Ethiopia. For example, the life expectancy at birth in Ethiopia is approximately 54 years, which is substantially lower than the average of 77 and 67 years recorded for countries with high and medium human development indices, respectively (DIFD, 2008). The adult illiteracy rate was around 60 percent which is significantly higher than the average for Sub-Saharan Africa (SSA) and other developing countries. According to human development report in 2009, 38.7% of the total population was below absolute poverty line.

There were 676 maternal deaths for every 100,000 live births in the country (CSA and ICF, 2012). Information obtained from Ethiopian Demographic and Health Survey (EDHS) revealed that under-five mortality decreased from 166 deaths per 1,000 live births in the 2000 to 88 in 2011, while infant mortality decreased from 97 deaths per 1,000 live births in the 2000 to 59 in 2011. On the other hand, even though neonatal mortality rate decreased from 49 deaths per 1,000 live births in 2000 to 39 deaths per 1,000 live births in 2005, it has since then remained stable at 37 deaths per 1,000, as reported in the 2011 EDHS. In 2002, the proportion of population with access to safe and clean water was only 22% and it increased to 54% in 2011 (CSA and ICF, 2012).). The majority of households, 82%, used non-improved toilet facilities (91% in rural areas and 54% in urban areas) which in turn affect the health of the community. Moreover, there were 0.03 physicians per 1,000 people in Ethiopia (World Bank, 2005). Although the average annual growth rate in GDP was 9.5% in 2012/2013, accesses to health services were inadequate for the majority of the population, particularly in rural areas. Besides, women literacy rate has increased from 29% in 2005 to 38% in 2011 (CSA and ICF, 2012).).

Despite the above anomalies, Ethiopia is on the right track to achieve the Millennium Development Goal (MDG) № 1: *Halving Poverty by 2015*. In order to reduce poverty and achieve maximum benefit for the poor, the government of Ethiopia has formulated different poverty reduction strategies including the Growth and Transformation Plan (GTP), which is under implementation to attain rapid and broad-based economic

ISSN 1992-0407

growth (MoFED, 2012). However, to achieve the above objectives, it is necessary to have adequate information on the nature and determinants of poverty.

Most studies applied different statistical methods to investigate the correlates of household poverty using a combined data set, i.e. rural and urban settings, and typically at national level without considering the effects at regional and local indicators. However, some authors have pointed out the potential bias associated with this practice. For instance, some variables like asset ownership and other characteristics may exhibit different relationships with wealth at different levels (Vyas and Kumaranayake, 2006; O'Donnell and Van Doorslaer, 2008; Woldehanna, 2008). There are also cases where some variables are relevant in rural settings and not in urban settings and vice versa.

Bogale *et al.* (2005) investigated the determinants of rural poverty in Ethiopia. They used logit model to identify determinants of poverty using one-year rural household survey data collected in three rounds in three districts of Ethiopia. The results indicated that entitlement failures resulted in lack of household resource endowments to crucial assets such as land, human capital and oxen.

Sepahvand (2009), using data from the 1997 Ethiopian Rural Household Survey (ERHS), identified determinants of rural poverty using the Foster-Greer-Thorbecke model (Foster *et al.*, 1984). He found that the incidence of rural poverty is high for villages that had less potential for agriculture. Moreover, the study also indicated that age of the household head and size of farmland are directly related to poverty status of households. Furthermore, households headed with less educated member were more vulnerable to incidence of poverty.

Mamo (1997) used multivariate analysis to analyze the determinants of standards of living in Addis Ababa using the first round Ethiopian Urban Household Survey (EUHS) conducted in 1994. He estimated a multinomial logit model to assess the likelihood of being poor using socioeconomic and demographic variables. The author found that education, access to credit, employment status, gender, marital status and food shortage were significant determinants of poverty status.

Generally, most previous studies on the correlates of poverty applied different models using nationally aggregated or to some extent urban/rural disaggregated data. However, given the diverse agro ecological and social setup of the country, application of aggregated data models to assess the status and intensity of poverty has little implication to design and implement sound policies and strategies. This implies the need for applying multilevel models that consider the effect of various regional and local level covariates. The main objective of this study was, therefore, to assess the application of multilevel models in identifying correlates of household poverty in Ethiopia.

2. Source of Data and Methodology

2.1. Source of Data

The 2011 Household Income, Consumption and Expenditure (HICE) survey for Ethiopia was used in this study. The data were collected to provide basic information on the standard of living of households, individuals and the society as a whole in Ethiopia. The survey that covered both rural and urban areas of the country was conducted by the Central Statistical Agency (CSA) in 2010/2011. For the purpose of representative sample selection, the country was divided into three broad categories, i.e., rural, major urban centers and other urban areas. Based on this division, two stage (for rural & major urban) and three stage (for other urban) stratified sampling technique were adopted to select a representative sample. After cleaning the data based on relevant variables, this study used information obtained from a total of 27,833 households.

2.2. Definition of Variables and Working Hypotheses

2.2.1. Dependent Variable

The dependent variable, poverty status of households, is measured based on per capita consumption of households. A household is considered to be poor if its total consumption per capita is below the official poverty line; that is ETB[1] 3781 per year (MoFED, 2012). The variable is, therefore, considered as binary which takes a value of 1 if the household is poor and 0 otherwise.

2.2.2. Independent Variables

Sex: It is widely believed that the gender of the household head significantly influences household poverty, and more specifically households headed by women are poorer than those headed by male**.** For example, Geda *et al* (2005) found that the households headed by males reduce the probability of being poor. This might be expected to be of particular importance in Ethiopia.

Age: Age of a household head is measured in complete years and is treated as a continuous variable. Households, whose heads is in higher age groups significantly lower the possibility of remaining poor (Khalid *et al.*, 2005; Meng *et al.,* 2007; Qureshi and Arif, 2001).

Family Size: It is number of household members. It is hypothesized that the larger the household size, the higher the level of poverty incidence, and vice versa (Meng *et al.*, 2007).

Dependency Ratio: It is the ratio of the number of family members not in the labor force (young or old) to those in labor force within household. One might expect a high dependency ratio will be associated with greater poverty (Minot and Boulch, 2005).

Employment Status: In order to take into consideration the different employment characteristics of the household

[1] ETB is the monetary unit in Ethiopia.

head, employment is distinguished as a categorical data comprised of the formal sector, informal sector and self-employed. Datt and Jolliffe (1997) found a positive relationship for sectors of employment (being self-employed and employed in formal sector) with per capita consumption. Similar result might be expected in the case of Ethiopia.

Educational Level: Educational attainment of the head of the household also significantly reduces the probability of remaining in the poor group. High educational attainment may imply a greater set of employment opportunities and specifically in the rural context, a better awareness of the full potential of new agricultural technologies and associated agricultural practices (Khalid *et al.*, 2005).

Location of Household: In order to know the importance of place of residence in the poverty status of the household, location dummy (rural/urban) was included.

Landholding: It is a dummy variable which takes a value of 1 if the household owns agricultural land and 0 otherwise. It is hypothesized that ownership of agricultural land has positive effect on pulling a household out of poverty trap.

Region: Ethiopia has nine regions namely: Tigray, Afar, Amhara, Oromiya, Somali, Benishangul-Gumuz, SNNP, Gambela and Harari and two Administrative Cities (Addis Ababa and Dire Dawa). Hence, to compare poverty across administrative regions, the dummies of region were included in the model.

2.3. Model Specification

2.3.1. Poverty Measures

For this research, Foster *et al.* (1984) P_α class of poverty measures were used to aggregate poverty and measure incidence, depth and severity of poverty. The general formula for the FGT class of poverty measures is:

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{q}\left[\frac{z-y_i}{z}\right]^\alpha ; \alpha \geq 0 \qquad (1)$$

where y_i is the ranked welfare indicator (per capita consumption) and z is poverty line. The parameter α is a measure of the sensitivity of the index to poverty and the poverty line . Larger values of α put higher weight on the poverty gaps of the poorest people. By setting $\alpha = 0$, the equation reduces to a headcount index. If $\alpha = 1$, the above equation becomes a poverty gap index, aggregating the proportionate poverty gap, which shows the shortfall of the poor's income from the poverty line, expressed as an average over the whole population. If $\alpha = 2$, the equation indicates the squared poverty gap index, which indicates severity of poverty.

2.3.2 Multilevel Logistic Models

The household data used for this analysis are nested within regions. To avoid bias in the parameter estimates and to estimate the impact of region level variables on the reported poverty status, a multilevel modeling was employed. The multilevel strategy deals with the problem of clustering which arises as a result of the hierarchical nature of the data, and estimates a random effect term which in this paper represents the extent to which poverty status varied across regions (Stephenson, 2009). Since the data were from 11 regions, to analyze such data, Goldstein (1991; 1995) developed the basic (two level) multilevel model for a binary response which is written as follows:

$$y_{ij} = \pi_{ij} + \varepsilon_{ij} \qquad (2)$$

where $\varepsilon_{ij} \sim iid\ N(0,\sigma_\varepsilon^2)$, takes the value 0 or 1 for each household i (0 = non-poor, 1 = poor) in region j, π_{ij} is the probability of being poor for household i in region j and ε_{ij} is a household-level error.

Random Intercept Only Model

The empty two-level model for a dichotomous outcome variable refers to a population of classes (level-two units, i.e. regions) and specifies the probability distribution for class-dependent probabilities without taking further explanatory variables into account. This model only contains random classes and random variation within regions. It can be expressed with logit link function as follows (Snijders and Bosker, 1999).

$$\text{logit}(\pi_{ij}) = \gamma_{00} + u_{0j} \qquad (3)$$

where $u_{0j} \sim iid\ N(0,\tau_{00})$, γ_{00} is the population average of the transformed probabilities and u_{0j} is the random deviation from this average for region j.

Random Intercept and Fixed Slope Model

In the random intercept logistic regression model, the intercept is the only random effect meaning that the regions differ with respect to the average value of the response variable. It represents the heterogeneity between regions in the overall response. The logistic random intercept model expresses the log odds, i.e. the logit of π_{ij} , as a sum of a linear function of the explanatory variables and a random region-dependent deviation u_{0j} . That is,

$$\text{logit}(\pi_{ij}) = \beta_{0j} + \sum_{h=1}^{k}\beta_h x_{hij} \qquad (4)$$

where the intercept term β_{0j} is assumed to vary randomly and is given by the sum of an average intercept γ_{00} and region-dependent deviations u_{0j} . That is,

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (5)$$

As a result

$$\text{logit}\,(\pi_{ij}) = \gamma_{00} + \sum_{h=1}^{k} \beta_{hj} x_{hij} + u_{0j} \qquad (6)$$

Note that $\gamma_{00} + \sum_{h=1}^{k} \beta_h x_{hij}$ is the fixed part of the model and u_{0j} is called the random part of the model.

The Random Coefficient Model

So far, we have allowed the probability of being poor to vary across regions, but we have assumed that the effects of the explanatory variables are the same for each region. We will now modify this assumption by allowing the difference between explanatory variables within a region to vary across regions. To allow for this effect, we will need to introduce a random coefficient for those explanatory variables. So a random coefficient model represents heterogeneity in relationship between the response and explanatory variables. As mentioned above, the response variable in this study, poverty status was binary. Therefore, the statistical models used in this analysis will be the two-level random coefficient multilevel regression model. The model with p household-level predictors and q region-level predictors can be expressed as:

$$\text{logit}\,(\pi_{ij}) = \beta_{0j} + \sum_{h=1}^{p} \beta_{hj} x_{hij} + \sum_{h=1}^{q} u_{hj} x_{hj} \qquad (7)$$

where

$\beta_{0j} = \gamma_{00} + u_{0j} \quad i = 1, 2, 3, ...n_j; \quad j = 1, 2, 3, .., J$

Now equation (7) can be rewritten as:

$$\text{logit}\,(\pi_{ij}) = \gamma_{00} + \sum_{h=1}^{p} \beta_{hj} x_{hij} + u_{0j} + \sum_{h=1}^{q} u_{hj} x_{hj} \qquad (8)$$

The first part of equation $\gamma_{00} + \sum_{h=1}^{p} \beta_{hj} x_{hij}$ is called the fixed part of the model. The second part $u_{0j} + \sum_{h=1}^{q} u_{hj} x_{hj}$ is called the random part.

The intercept-only model does not explain any variance of the dependent variable. It only decomposes the variance into two independent components: σ_ε^2, which is the variance of the lowest level (household-level) errors ε_{ij}, and τ_{00}, which is the variance of the highest-level (region level) errors u_{0j}. Using this model, we can define the intraclass correlation ρ by the equation:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma_\varepsilon^2} \qquad (9)$$

The intraclass correlation indicates the proportion of the variance explained by the grouping structure in the population.

Multilevel Model Selection Criteria

The AIC and the BIC are two common measures for comparing maximum likelihood models. Given two models fit on the same data, the model with the smaller value of the information criterion is considered to be better (Akaike, 1974 and Schwarz, 1978). In this paper these two model selection criteria were used to suggest the best model.

3. Results and Discussion

3.1 Descriptive Statistics Results

As it can be seen from Table 1, out of the total household heads, 31.5% were female headed and the remaining 68.5% were male headed. With respect to poverty status, 77.5% female headed households were categorized under poor and the remaining 22.5% female headed households were belonging to non-poor category. The study also illustrates that 12.84% of urban and 52.95 of rural household heads were below the poverty line, and thus categorized as poor. The proportion of poor rural households were higher than urban.

In terms of education, about 94.2% of the household heads in Ethiopia were literate with different level of schooling, the largest part of the sample population being in primary school. Based on literacy status, non-poor household heads did much better than poor household heads. In each level of schooling, most of the poor households were tend to be lower in number as compared to non-poor households. Similar to the non-poor household heads, most of the poor household heads concentrate in primary school while the number of heads with school levels higher than secondary school was very small.

It can be viewed from Table 1 that the majority of household head respondents (73.10% and 26.90%) were self-employed for both non-poor and poor categories, respectively. Households who owned agricultural land comprises 19.95 % of the poor and 80.89% of the non-poor. Regionally, the distribution of poverty was highest in Afar region (47.9%) followed by Somali region (46.48%) and lowest in Addis Ababa City administration (14.27%).

Table 1. Descriptive results for categorical variables disaggregated by poverty status.

Variable	Categories	Non-poor Count	Non-poor Percent	Poor Count	Poor Percent	Total	χ² value
Sex	Male	13334	69.89	5744	30.11	19078	
	Female	6786	77.49	1971	22.51	8757	173.051*
Place of residence	Urban	15264	87.16	2249	12.84	17513	
	Rural	4856	47.05	5466	52.95	10322	5215.804*
Educational level	Illiterate	578	62.02	354	37.98	932	
	Primary School	5699	70.29	2409	29.71	8108	
	Secondary School	2820	81.53	639	18.47	3459	79.857*
	College & above	2904	82.90	599	17.10	3503	
Employment status	Self	11744	73.10	4321	26.90	16065	
	Formal	4721	70.11	2013	29.89	6734	29.187*
	Informal	543	77.02	162	22.98	705	
Landholding	Yes	10680	80.09	2655	19.91	13335	779.152*
	No	9438	65.10	5060	34.90	14498	
Region	Tigray	1597	69.74	693	30.26	2290	
	Afar	697	52.05	642	47.95	1339	
	Amhara	4071	80.42	991	19.58	5062	
	Oromiya	4556	79.23	1194	20.77	5750	
	Somali	920	53.52	799	46.48	1719	
	Benishangul-Gumuz	742	55.83	587	44.17	1329	111.02*
	SNNP	2767	70.50	1158	29.50	3925	
	Gambela	719	53.54	624	46.46	1343	
	Harari	470	70.25	199	29.75	669	
	Addis Ababa	3207	85.73	534	14.27	3741	
	Dire Dawa	373	56.01	293	43.99	666	

** = Significant at 0.05*

The chi-square test result presented in Table 1 was also used to test whether or not there is a significant association between poverty status of household and each predictor variables independently. These tests revealed that all predictor variables showed a significant ($P \leq 0.05$) association with poverty status.

According to the results computed from the data (Table 2), the average age of the poor household heads (44.92 year) was significantly ($P \leq 0.05$) greater than that of the non-poor (40.46 year). However, standard deviation for poor households was relatively higher showing relatively higher dispersion from the mean age. The mean family size of the sample respondent, was found to be 4.74 persons per household. The average family sizes of the sampled poor and non-poor households were 5.77 and 3.71 persons, respectively. It shows that the mean household size of the poor category was greater than the non-poor category. The average dependency ratio for the sample data was computed to be about 81, which indicates that every 100 persons of economically productive age group were responsible to take care of themselves as well as additional 81 persons (children and aged population) based on the survey. The mean dependency ratios for poor and non-poor households were estimated to be 127% and 63.4%, respectively. The survey has also indicated high variation in dependency ratio for poor households than the non-poor (0.994 vs 0.761).

Table 2. Continuous variables disaggregated by poverty status.

Variable	Poverty Status: Non-poor Mean	Non-poor St. Dev.	Poor Mean	Poor St. Dev.	z/t value
Age	40.46	14.54	44.92	15.94	-71.09*
Family size	3.71	2.200	5.770	2.158	-22.26*
Dependency ratio	0.634	0.761	1.270	0.994	-51.44*

** = Significant at 0.05*

Moreover, the z/t-test was used to know the mean variations between the poor and non-poor in terms of continuous explanatory variables. The analysis of z/t-test also showed that there was a significant ($P \leq 0.05$) statistical difference between poor and non-poor in terms of three variables.

3.2. Poverty Indices

The estimated poverty indices for Ethiopia using HICE survey (2010/2011) were presented in Figure 1. Based on total poverty line, absolute headcount index stood at about 28%, which indicates the percentage of the sampled population who was unable to meet the required minimum amount of 2, 200 kcal per person per day. In other words, this proportion of households could not attain the minimum amount of consumption (ETB 3781) to satisfy the minimum calorie requirement per adult equivalent per year.

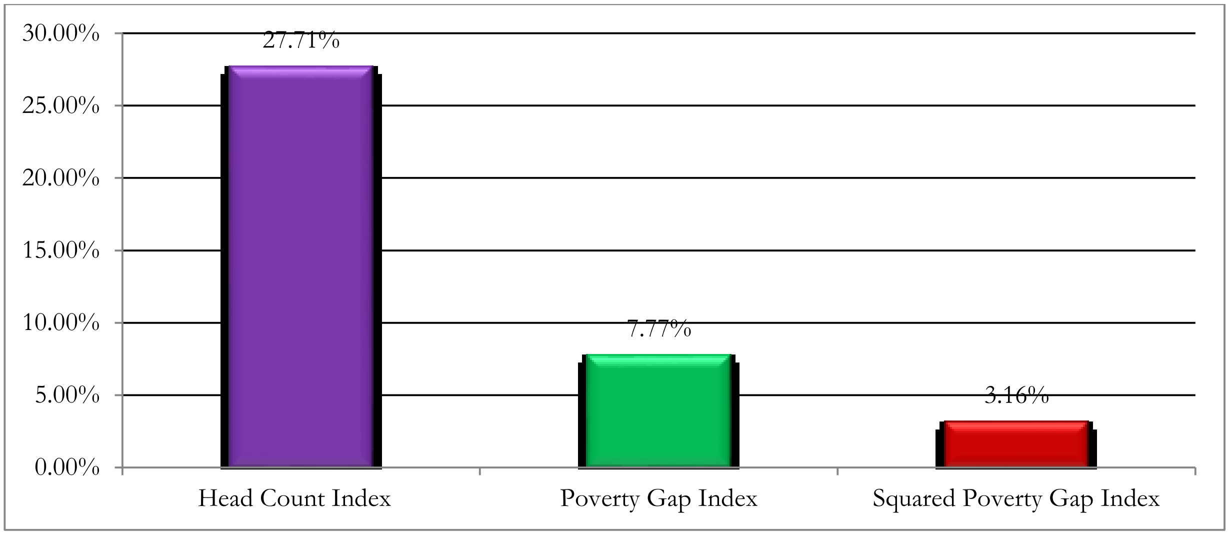

Figure 1: Absolute Poverty Indices in Ethiopia.

The poverty gap index $(\alpha = 1)$ which captures the average proportional shortfall (i.e., the difference between per capita consumption and total poverty line and then divided by the total poverty line) is 7.77%. This means average consumption needed to bring the poor above the poverty line or the minimum level of living is 7.77% of the poverty line. It indicates the percentage of consumption expenditure deficit the poor faces so as to uplift the poor from the poverty line. If one simply adds up the difference between the expenditure measure and poverty line for all those who were below, one would obtain the total cost required to eliminate poverty. Similarly, the squared poverty gap index $(\alpha = 2)$ in consumption expenditure, was 3.16%, which could indicate the severity of poverty by assigning more weight to the poor.

3.3. Intercept Only Model

Is there significant variation in poverty status at household and region level? To answer this question, the intercept-only model was estimated. Table 3 shows the parameter estimate of the average log odds (grand mean log odds) and the variance component. As one can see, the variance component (τ_{00}) is statistically significant suggesting that there was significant variance in poverty status of households at region level. This indicates the multilevel character of household data should not be ignored.

With the existence of hierarchically structured data, applying traditional regression models violates the assumption of independence of observations and increases Type-I error (Kreft & De Leew, 1998). Another way to examine clustered data is to compute intraclass correlation, which is a measure of the degree of dependence of households belonging to the same region. The intraclass correlation can also be interpreted as the fraction of total variability that is due to the region level (Kreft & DeLeew, 1998). When the logistic model is applied, the level-one residuals are assumed to follow the standard logistic distribution, which has a mean of 0 and a variance of $\pi^2/3 = 3.26$ (Snijders & Bosker, 1999). Using $\rho = \frac{\tau_{00}}{\tau_{00} + 3.26}$, the intraclass correlation coefficient is 0.146. Where τ_{00} is between region variance and $\sigma_{\varepsilon}^2 = 3.26$ is within region variance. Thus, about 14.6% of the total variance in poverty status of households is attributed to differences between regions in the country.

Table 3. Intercept Only Model.

Fixed Effect	Coefficient	S.E.
Intercept (γ_{00})	-0.122*	0.007
Random Effect	Var. Comp.	S.E.
Region-level (τ_{00})	0.556*	0.126

** = Significant at 0.05*

3.4 The Random Intercept and Fixed Slope Model

The results of the intercept only model indicate very clearly that there was significant variation in poverty status of household at two levels of analysis. Now we return to the question of whether the model specified in equation (6) can account for this variance. Table 4 below gives the parameter estimates of the fixed effects and the variance component of this multilevel model. From the model estimates $\gamma_{00} = 0.361$ is the expected log-odds of poverty status for an average household. Introducing eight level-1 variables decreased the intraclass correlation to 0.043.

Table 4. The Random Intercept and Fixed Slope Model.

Fixed Effects	Estimate	S.e.	OR
Intercept (γ_{00})	0.361*	0.008	.
Dependency ratio	0.450*	0.034	1.568
Family size	0.376*	0.014	1.456
Age	-0.035*	0.002	0.965
Female	0.471*	0.063	1.601
Rural	0.239*	0.065	1.269
Landholding	-0.990*	0.056	0.372
Primary school complete	-0.588**	0.270	0.555
Secondary school complete	-0.372*	0.094	0.689
College & above	-0.554*	0.070	0.575
Employed in formal sector	-0.295*	0.061	0.744
Self-employee	-0.702*	0.170	0.496
Random Effects	Random Component	S.E.	
Region level (τ_{00})	0.145*	0.004	

** = Significant at 0.01; ** = Significant at 0.05; S.e = Standard error; OR = Odds ratio*

The variance of random effect of the intercept and fixed slope model (0.145) decreased compared to random effect of the intercept only model (0.556). The reduction of the random effects of the intercept variance is due to the inclusion of fixed explanatory variables. That is, taking into account the fixed independent variables can provide extra predictive value on poverty status in each region. The random intercept $\tau_{00} = 0.145$ (P $\leq$ 0.01) indicates that poverty status differs from region to region in terms of measured covariates (Table 4). This implies that there is still unexplained variation on poverty status across regions.

The fixed part of the model presented in Table 4 reveals that covariates family size, dependency ratio, landholding of household, age, sex (female), type of place of residence (rural), educational attainment and employment status of household head were statistically significant at 5% level of significance. This implies that all variables were correlated with the probability of being poor at household level (level-1) in Ethiopia. For example, the coefficient of family size (0.376) indicates that log of being poor increase by an average of about 0.456 for each increase in household member fixing other covariates. Looking at the results of multilevel logistic regression estimated above (Table 4), the sign for sex (female) of the head was positive and statistically significant at 5% level of significance. The average odd of being poor for female headed households were 1.602 times that for male headed households. Education was grouped into four categories ranging from illiterates to those who have attended higher education (college and above). The odds of being poor with education level of elementary school, secondary school and college and above was found to be 0.555, 0.689 and 0.575 times that of the illiterates (reference category), respectively.

The random intercept and fixed slope multilevel logistic regression model also helps to compare poverty status across regions. The results of this model in Table 5 show the estimated regional random effects of intercept in eleven regions of Ethiopia. Among these regions, the random effect of intercept for poverty status in Afar, Somali, SNNP, Benishangul-Gumuz, Gambela, Harari and Addis Ababa were statistically significant at 5% level of significance. The estimated random regional effects revealed the average poverty status in a particular region.

Table 5. Estimated random effects of intercepts for each region in explaining poverty status.

Effect	Subject	Estimate	S.e.	Z value	P-value	OR
Intercept	Region Tigray	0.306	0.256	1.190	0.2320	1.36
Intercept	Region Afar	0.885	0.212	4.170	0.000*	2.42
Intercept	Region Amhara	-0.204	0.129	-1.580	0.1130	0.82
Intercept	Region Oromiya	-0.007	0.071	0.011	0.9180	0.99
Intercept	Region Somali	0.288	0.067	4.280	0.000*	1.33
Intercept	Region Ben-Gumuz	0.210	0.100	2.040	0.041*	1.23
Intercept	Region SNNP	0.544	0.286	1.990	0.028*	1.72
Intercept	Region Gambela	0.225	0.116	1.935	0.027*	1.25
Intercept	Region Harari	-0.202	0.044	21.59	0.001*	0.82
Intercept	Region Addis Ababa	-0.041	0.003	162.24	0.000*	0.96
Intercept	Region Dire Dawa	-0.047	0.038	1.559	0.2110	0.95

*** = Significant at 0.05; S.e = Standard error; OR = Odds ratio*

The result also depicts that the average poverty status in Afar and SNNP was very high compared to the average of all regions. On the contrary, Harari and Addis Ababa regions had better performance in the average reduction of household poverty compared to the average of all regions as the odds ratio were less than one (Table 5).

3.5. The Random Coefficient Model

So far, we have allowed the probability of being poor to vary across regions, but we have assumed that the effects of the explanatory variables are the same for each region. We will now modify this assumption by allowing the difference between urban and rural areas within a region to vary across regions. To allow for this effect, we will need to introduce a random coefficient for place of residence as one can see in Table 6 below. One can also test the significance of the added parameters, τ_{55} (variance in the slopes of place of residence) and τ_{05} (covariance between region and place of residence), using a Wald test. The test statistic is 39.136, which is approximately Chi-Square distributed with 2 degree of freedom ($P \leq 0.001$).

Table 6. The Random Coefficient Model.

Fixed Effects	Estimate	S.E.
Intercept (γ_{00})	0.361*	0.008
Intercept (γ_{50})	0.240*	0.065
Dependency ratio	0.459*	0.035
Family size	0.369*	0.014
Age	-0.035*	0.002
Female	0.487*	0.063
Landholding	-0.972*	0.057
Primary school complete	-0.476*	0.091
Secondary school complete	-0.611*	0.104
College & above	-0.516*	0.110
Employed in formal sector	-0.302*	0.063
Self-employee	-0.708*	0.174
Random Effects	**Var. Comp.**	**S.E.**
τ_{00} (var. in intercept)	0.0718	0.0046
τ_{55} (var. in area slopes)	0.4731	0 .1135
$\tau_{05} = \tau_{50}$ (covariance)	-0.125	0.0220

** = Significant at 0.05*

At the 5% level of significance, both parameters are non-zero (γ_{00} and γ_{50}), which implies that the effect of place of residence does indeed vary across regions. On average (after adjusting for the other explanatory variables), the log odd of being poor was 0.240 higher for rural areas than for urban areas. Depending on the value of u_{5j}, the difference in a given region can be larger or smaller than 0.240. That means, the average effect of rural is $\gamma_{50} = 0.240$, but the effect for region j is $\beta_{5j} = \gamma_{50} + u_{5j} = 0.240 + u_{5j}$, where

$$\begin{bmatrix} u_{0j} \\ u_{5j} \end{bmatrix} \sim N(0, \Omega_u): \ \Omega_u = \begin{bmatrix} \tau_{00} & \tau_{05} \\ \tau_{50} & \tau_{55} \end{bmatrix} = \begin{bmatrix} 0.0718 & -0.125 \\ -0.125 & 0.4731 \end{bmatrix}.$$

For the model specified in Table 6, the residual variance between regions is a function of rural.

$$\begin{aligned} \text{var}(u_{0j} + \text{rural } u_{5j}) &= \text{var}(u_{0j}) + 2 \text{ rural cov}(u_{0j}, u_{5j}) + \text{rural}^2 \text{var}(u_{5j}) \\ &= \tau_{00} + 2 \text{ rural}\tau_{05} + \text{rural}^2 \tau_{55} \end{aligned}$$

Because rural is a (0, 1) variable, $\text{rural}^2 = \text{rural}$. Thus, the above equation becomes:

$$\text{var}(u_{0j} + \text{rural } u_{5j}) = \tau_{00} + (2\tau_{05} + \tau_{55})\text{rural}$$

The equation can be more simplified for rural and urban separately. For rural areas (rural = 1).

$$\begin{aligned} \text{var}(u_{0j} + \text{rural } u_{5j}) &= \tau_{00} + 2\tau_{05} + \tau_{55} \\ &= 0.0718 + 2(-0.125) + 0.4731 = 0.2949 \end{aligned}$$

For urban areas (rural = 0), which leads to $\text{var}(u_{0j} + \text{rural } u_{5j}) = \tau_{00} = 0.0718$.

Hence, there was greater region level variation in the probability of being poor in rural areas than in urban areas in the country.

Besides, Table 7 shows the cases for the empty model, random intercept and fixed slope model and random coefficient model. Each model had its own AIC and BIC.

Table 7. Model Selection Criteria.

Model	AIC	BIC
Empty Model	20819.56	20836.03
Fixed Slope Model	6083.665	6181.622
Random Coefficient Model	5844.545	5912.723

The AIC and BIC reported were measures of model misfit; when we add explanatory variables to the model, AIC and BIC are expected to go down. After examining each model, the random coefficient model had the lowest value of both criterion (AIC and BIC) since lower values of these statistics indicate a better fitting model by adjusting for the number of explanatory variables. This indicates that the random coefficient model was found to give a better fit as compared to the empty and random intercept and fixed slope model to predict poverty status of household in Ethiopia.

4. Conclusion

Estimates and policy recommendations based on a model without considering the nature and structure of data can be seriously biased. This study attempted to identify the most reliable model among the three multilevel logistic models (the random intercept only, the random intercept and fixed slope and the random coefficient) to analyze poverty when the data contain hierarchical structures. Using data from Ethiopia, the random coefficient model was found to be more appropriate than others to predict poverty status at households' level. This suggests that researchers should consider the nature of hierarchically structured datasets when modeling poverty at household level. Moreover, multivariate analyses techniques which considers the hierarchical nature of the data can be routinely incorporated to obtain better estimates of parameters for policy inputs at various level. Failure to account the hierarchical nature of the data may lead to biased estimate and wrong conclusions.

5. Acknowledgment

The feedback of Mr. Gutu Adugna (Lecturer of Statistics at Arba Minch University) deserves special appreciation for taking his incredible time in editing and reviewing the article. I am also indebted to the Central Statistical Agency (CSA) for providing the necessary materials and making the data ready for analysis.

6. References

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Bogale, A., Hagedorn, K. and Korf, B. 2005. Determinants of poverty in rural Ethiopia. *Quarterly Journal of International Agriculture* 44 (2): 101-120.

Chaudhry, I. 2003. An Empirical Analysis of the Determinants of Poverty in Pakistan: A Case Study of Bahawalpur District with Special Reference to Cholistan, PhD Thesis, Pakistan.

CSA (Central Statistical Agency) and ICF International. 2012. *Ethiopia Demographic and Health Survey 2011*. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Agency and ICF International.

Datt, G. and Jolliffe, D. 1999. Determinants of Poverty in Egypt 1997. Discussion Paper No. 75, Food Consumption and Nutrition Division, IFPRI Washington DC.

DIFD (Department for International Development) 2008. World Development Indicators (www.dfid.gov.uk.) (Accessed on Jun 20012)

Foster, J., Greer, J. and Thorbecke, E. 1984. A Class of Decomposable Poverty Measures, *Econometrica* 52: 761-766.

Geda, A., Jong, N.de, Kimenyi, M. S. and Mwabu, G. 2005. Determinants of Poverty in Kenya: A Household Level Analysis. Working Paper Series 44 Department of Economics, University of Connecticut.

Goldstein, H. 1995.*Multilevel Statistical Models*. London: Edward Arnold, New York, Halsted.

Goldstein, H. 1991. Nonlinear multilevel models with an application to discrete response data. *Biometrika* 78: 45-51.

Khalid, U., Shahnaz, L. and Bibi, H. 2005. Determinants of poverty in Pakistan: A multinomial logit approach. *The Lahore Journal of Economics* 10: 65-81.

Kreft, I. and De Leeuw, J. 1998. *Introducing Multilevel Modeling*. Sage Publications, UK.

Mamo, A. 1997. Modeling poverty and its determinants in Addis Ababa: A focus on multinomial logit selection. *Ethiopian Journal of Economics* 5(2): 1-36.

Meng, X., Gregory, R. 2007. Urban poverty in China and its contributing factors (1986-2000). *Review of Income and Wealth* 53: 167-189.

MoFED (Ministry of Finance and Economic Development). 2012. Ethiopia's Progress towards Eradicating Poverty: An Interim Report on Poverty Analysis Study (2010/2011), Addis Ababa, Ethiopia.

Minot, N., Baulch, B. 2005. Poverty Mapping with Aggregate Census Data: What is the Loss in Precision? *Review of Development Economics* 9: 5-24.

O'Donnell, O., Van Doorslaer, E., Wagstaff, A. & Lindelow, M. 2008. Analyzing Health Equity Using Household Survey Data - A Guide To Techniques And Their Implementation. World Bank, Washington DC.

Qureshi, K. S & Arif, G.M. 2001. Profile of Poverty in Pakistan (1998-1999). MIMAP Technical Paper Series No. 5. Pakistan Institute of Development Economics, Islamabad, Pakistan.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics* 6: 461–464.

Sepahvand, M. 2009. The Analysis of Rural Poverty in Ethiopia: Regarding the three Measurements of Poverty. MA Thesis, Uppsala University, Sweden.

Snijders, T.A.B. and Bosker, R.J. 1999. *Multilevel Analysis*. Thousand Oaks, CA, Sage.

Stephenson, R. 2009. Community Influences on Young People's Sexual Behavior in 3African Countries. *American Journal of Public Health* 99: 102-109.

Vyas, S. and Kumaranayake, L. 2006. Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning* 21: 459-468.

Woldehanna, T. 2008. Correlates of poverty in urban and rural Ethiopia. *Ethiopian Journal of Agricultural Economics* 7: 49-81.

World Bank. 2005. *World Development Report: Attacking Poverty*, Oxford University Press, New York, USA.