Household Poverty-Risk Analysis and Prediction Using Bayesian Ordinal Probit Models

Richard Puurbalanta[1,*], Atinuke O. Adebanji[2]

[1]University for Development Studies, Faculty of Mathematical Sciences, Department of Statistics, Navrongo Campus, Ghana
[2]Kwame Nkrumah University of Science and Technology, Faculty of Physical Sciences, College of Science, Department of Mathematics, Kumasi, Ghana

Abstract   Though the rate of poverty in Ghana has consistently declined over the years, some parts of the country still record substantially high figures [1], and this is a major concern for stake holders. Previous research to identify causal factors has commonly used the binary logit or probit models. These models, however, mask the effect of important intermediate information during the binary transformation of the response variable. This has the potential to misestimate the probability of poverty. In this study, the ordered probit model was used, thus creating a framework that includes the ordinal nature of poverty severity. The model was based on the round 6 dataset of the Ghana Living Standards Survey. Our findings show that poor and extremely poor were negatively affected by rural location, illiteracy, and Savannah ecological zone. Policies to eradicate poverty must therefore aim at optimizing these significant variables contributions to welfare conditions in the country.

Keywords   Household Poverty Severity, Predictors, Ordinal Probit Model, MLE, Bayesian estimation

1. Introduction

Poverty is the inability to retain a minimal standard of living measured in terms of basic consumption needs or some income required for satisfying them [2]. It is a global development issue [3], and its magnitude and burden is evident, even to the non-observant, especially in developing countries. Thus, the need for research designed to identify the determinants of poverty and to monitor the impact of programs aimed at poverty reduction is imperative.

Predictor variables commonly used in poverty studies include socio-economic and demographic attributes of the household [4], which in this study include age, sex, household size, rural/urban location, employment status, and educational level of household head. An additional covariate, ecological zone, was added in this study to estimate the effect of differences in agro-climatic conditions on level of wellbeing. Table 1 gives details of relevant variables extracted from the GLSS dataset for this study.

Previous research to investigate the link between poverty and potential predictors in regression analysis is abundant in the literatures, albeit with varying degrees of sophistication and often with mixed conclusions. For instance, [5] used the logistic regression model to identify household-level determinants of poverty in Albania, and concluded that the probability of being poor is influenced mainly by education and employment status of household head, the household composition and geographic divisions. Dudek and Lisicka [6] used the binary logit model but included interaction terms to describe poverty in Poland. They found that the probability of being poor is influenced by size of the household, living in a rural area, and working in manual positions. Akin to these findings, a recent study in Ghana by [7] using logistic regression to estimate the probability of being in poverty based on living standards survey (LSS) data, reported that large households, and households headed by illiterates negatively affect poverty levels in the country. An important and a notable conclusion reached in their study identified households in the savannah ecological zone of Ghana to be almost four times poorer than those living in the coastal and forest zones.

Most studies, including this one, have used household income or expenditure to measure poverty. A common approach has been to dichotomize the dependent variable into poor or non-poor, and use the binary logit or probit models (being the natural choices) to estimate the probability of a household being in poverty. These binary models, however, usually mask the effect of important intermediate information during the binary transformation of the response variable. This has the potential for incomplete and inaccurate conclusions.

What we do differently in this study is to adopt a framework for poverty analysis that allows for the response variable to assume all of its key response categories; "non-poor", "poor", and "extremely poor" [1]. We then use

* Corresponding author:
pubarichard@yahoo.com (Richard Puurbalanta)

the proportional odds ordered probit (OP) model [8], thus creating a framework that includes the ordinal nature of poverty severity. This approach allows thorough investigations to determine whether or not distances between the poverty categories differ, and the probability of slipping into the next category given that an individual is already in a lower-risk category, and vice versa.

The few previous studies that have used multi-classification methods to model poverty have largely ignored the ordinal nature of the variable. For example, in separate studies, [9] and [10] used a multinomial econometric approach in their bit to quantify the relationship between poverty and socio-economic and demographic variables. The intention to include the multi-categorical nature of poverty was explicit. However, the natural ordering of the variable was ignored. This can easily bias results of the study.

The ordered probit (OP) model is an extension of the ordinary binary probit model. It is generally used when categories of the dependent variable exceeds two, and follow a natural order. Ordered models are generally derived within a framework of unobservable underlying latent processes generating observable ordered outcomes. Although it is generally not possible to observe the latent variable, we make the assumption that some factors jointly determine its behaviour. Thus, the latent variable is assumed to be a random function. In this setting, the goal is to understand the influence of the latency on the probability of observing a particular outcome based on individual-related factors.

Proportional odds ordered probit models describe the relationship between covariates and the probability to be in a higher ordered category [1]. Intercepts for each comparison is different, but share the same set of estimated beta coefficients. The intercepts show the effect when $X = 0$. Unlike the partial proportional odds (PPO) model [11] and the generalized ordered probit (GOP) model [12], proportional odds ordered probit models are less complicated and simple to interpret.

Estimation of categorical ordered models can be challenging because non-linear optimization methods are involved in the process. Moreover, the regression-based ordered probit model has a unique feature that includes the thresholds as part of unknown parameters to be estimated. This further complicates the estimation process: because of possible unequal spacing between categories, ordinary least squares (OLS) regression, which, theoretically, requires equal interval widths to function properly, is inappropriate. However, $J \to \mp\infty$ OLS regression becomes legitimate. Additionally, thorough procedures to fest predictability of the model is absent in OLS regression. This is important because the primary aim is to use the developed models for prediction of the unobserved.

For these reasons, we employ the Bayesian paradigm for parameter estimation and inference instead of the traditional ordinary least squares (OLS) approach. Bayesian estimation in particular is attractive largely because the latent variable is naturally incorporated as an additional parameter via data augmentation. With advancement in computing power, and the advent of Markov Chain Monte Carlo (MCMC) [13], implementation of the Bayesian framework is easy. A key advantage of Bayesian inference is to fully account for parameter uncertainty associated with the predicted values [14]. Besides, Bayesian inference is robust, and does not need to satisfy any regularity conditions to function properly.

The primary objective is to attempt to provide a full description of the character of poverty in a simple and straightforward language, and predict the probability of an unobserved household being in a particular severity category given some covariate characteristics of households in the study area. This way, even non-technical end-users such as government and social-support groups can understand how patterns of economic, social and demographic changes affect different classes of the poor in the country, helping to avoid blanket approach to combating poverty. We base our inference on the latest Ghana living standards survey (GLSS) dataset collected in 2012.

Table 1. Study Variables

Predictor Variables	
Variables	**Meaning**
Age	Householder age in years
Household size	Number of household members
Sex	1 if male, 0 female
Rural/Urban	1 if rural, 0 urban
Ecological zone	1 if coastal, 2 forest, 3 savannah, 4 Accra
Education level	1 if no education, 2 basic education (JSS or below), 3 SHS/Vocational (or equivalent certificate), 3 Tertiary
Employment status	1 self-employed, 2 unemployed, 3 wage employment
Response variable	
Poverty status	1 if Non-poor, 2 Poor, 3 Extremely poor

2. Measurement of Poverty

This work, in line with the Ghana Statistical Service (GSS), and following international common practice, adopts monetary measures of poverty. This approach measures poverty in relation to the amount of money necessary to meet some basic consumption needs such as food, clothing, and shelter [15].

Estimation of monetary measures of poverty requires us to choose between disposable income and total consumption expenditure as the indicator of wealth, the latter being the preferred choice in most developing countries, and in this study [1].

Moreover, to measure monetary poverty, we need to set minimum standards of the poverty indicator to separate the various categories of the poor from the non-poor. These are called poverty lines [15, 1]. The three categories of poverty lines proposed by [15] are based on family daily expenditure per adult equivalent. Households who spend at least $1.25 a

day are non-poor, whiles those who spend less than $1.25 are poor. The extremely poor are those with expenditure less than $1 a day. In Ghana, these thresholds are set at GHC 3.60 per day- non-poor, less than GHC3.60 –poor, and GHC2.17- extremely poor [1].

3. The Ordinal Probit Model

The ordinal probit (OP) model derives from the multinomial distribution, albeit with ranked categories, thus, its likelihood function is multinomially distributed. The multinomial distribution is an extension of the binomial distribution where, now, the number of parameters being modelled exceeds one. The density function for the multinomial distribution is:

$$Pro(P|Y) = \prod_{i=1}^{n} \prod_{j=1}^{J} p_i^{y_{ij}} \tag{1}$$

To derive the cumulative ordinal probit (OP) generalized linear model (GLM) from the multinomial distribution, let the responses $j = 1, \dots, J$ be arranged in order of magnitude, and α_j the corresponding thresholds associated with the ordering. Further let Y_i^* be a Gaussian random variable assumed to be latent [8, 16], and assigning values to α_j according to a regression function:

$$Y_i^* = X_i \beta + \varepsilon_i, \tag{2}$$

where X is $n \times p$ design matrix, β is a $p \times 1$ unknown vector of regression coefficients, and ε is the $n \times 1$ vector of independently and identically distributed $(i.i.d)$

measurement errors: $\varepsilon_i \sim N(0,1)$. Though the values of y_i^* cannot be directly observed, the rule that assigns y_i^* to α_j is that if y_i^* exceeds a given threshold, then, for example, a household falls in the jth category of poverty. This culminates in cumulative multiple binary outcomes:

$$Y_i = \begin{cases} j, \alpha_{j-1} < y_i^* \le \alpha_j \\ 0, \text{otherwise} , \end{cases} \tag{3}$$

where $\alpha_j \in \mathbb{R}$, and $\alpha_1 < \alpha_2 \dots < \alpha_J$.

Cleary, Y_i^* in our application, refers to the Gaussian expenditure line, and is asymptotic of the ordinal variable Y_i when $J \to \mp\infty$.

Our objective is to predict the probability of a household falling in or below the jth category given the observed covariates $x = (x_1, \dots, x_n)^T$. This probability is determined by the values of the latent variable Y_i^*, and is given by

$$p(y_i = j|x_i) = p(\alpha_{j-1} < y_i^* \le \alpha_j) \tag{4}$$

Since Y_i^* is Gaussian, and ε_i is assumed to be normally distributed, the outcome is a probit model, implying that the probability of falling in or below the jth category is:

$$p(y_i = j|x_i) = p(\alpha_{j-1} - X_i^T\beta < \varepsilon_i \le \alpha_j - X_i^T\beta)$$
$$= \Phi(\alpha_j - X_i^T\beta) - \Phi(\alpha_{j-1} - X_i^T\beta) \tag{5}$$

where $\Phi(.)$ is the cumulative distribution function (cdf) for the standard normal $\varepsilon_i \sim N(0,1)$:

$$F(\varepsilon_i) = \left(\frac{1}{2\pi}\right)^{-\frac{n}{2}} \exp(-\frac{1}{2}\varepsilon_i^2).$$

Thus, the likelihood function for the parameters is

$$L(\beta, \alpha|y = j) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left[\Phi(\alpha_j - X_i^T\beta) - \Phi(\alpha_{j-1} - X_i^T\beta)\right] \tag{6}$$

3.1. Distributional Assumptions and Normalizations of the OP Model

Several normalizations are often needed to identify the ordered model parameters. First, to preserve the positive signs of all the probabilities, it is required that $\alpha_j > \alpha_{j-1}$.

Second, if the support for the categories is to be the entire real line, then $\alpha_{j-1} = -\infty$ and $\alpha_j = +\infty$.

Third, following standard practice, we assume the continuous latent variable Y_i^* is standard normal [8, 16].

Finally, we assume, following [8], and [16], that x_i contains a constant term. Then it is required for $\alpha_0 = 0$. If this restriction is not imposed, adding a constant to α_j and another constant to the vector of beta coefficients results in identifiability problems (the model may appear to contain two constant terms or intercepts).

3.2. Parameter Estimation and Inference

3.2.1. Maximum Likelihood Estimation (MLE)

In our poverty severity application, given that our response is multinomial ordered with n_i observations and y_j categories, where $n_1 + \cdots + n_N = n$ and $j = 1, \dots, J$, we use the likelihood function for the OP model derived from (1):

$$L(\beta, \alpha|x, y) = \prod_{i=1}^{n} \sum_{j=1}^{J} \left[\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)\right]^{I(y_i=j)}, \tag{7}$$

the log-likelihood being

$$\log L(\beta, \alpha|x, y) = \prod_{i=1}^{n} \sum_{j=1}^{J} I(y_i = j)\log\left[\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)\right]$$
$$= \prod_{i=1}^{n} \sum_{j=1}^{J} I(y_i = j)\log\left[\Phi(\alpha_{y_i}) - \Phi(\alpha_{y_i-1})\right]$$
$$= \prod_{i=1}^{n} \sum_{j=1}^{J} I(y_i = j)\log P_{ij} \tag{8}$$

where I is the identity matrix.

The ML process chooses that estimators, $\hat{\beta}$ and $\hat{\alpha}$ of the set of unknown parameters, β, α which maximize the data [17, 18, 16].

Like many models, the curve of $L(\beta, \alpha | x, y)$ is non-linear [16] and $\hat{\beta}$ and $\hat{\alpha}$ is the point at which

$$\frac{\partial logL\,(P_{ij})}{\partial(\beta, \alpha_j)} = 0, j = 1, ..., J - 1. \tag{9}$$

Equation (9) is maximized to obtain the MLE of β, α_j subject to $\alpha_0 = -\infty, \alpha_1 = 0, \alpha_j = \infty$. This equation cannot easily be solved analytically. So we adopt the Newton-Raphson iterative algorithm [8].

A measure of how well the model fits is important, and is determined by the significance of the overall model fit statistic. Fit indices include the likelihood ratio test, which is approximately chi-square $\chi^2_{[P+J-2]} = 2[logL_{Model} - logL_{No\ Model}]$ for large n.

3.2.2. Bayesian Estimation via Gibbs Sampling

Following the Bayesian criteria, we set priors for the parameters, and build the posterior as:

$$f(\beta, \alpha | y_i^*, y, x) = \frac{\prod_{i=1}^{n}[\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)]p(\beta, \alpha)}{\int_{\Theta} \prod_{i=1}^{n}[\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)]d(\beta, \alpha)} \tag{10}$$

where $\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)$ is the data likelihood, $p(\beta, \alpha)$ is the density of the prior, and $\int_{\Theta} \prod_{i=1}^{n}[\Phi(\alpha_j - X\beta) - \Phi(\alpha_{j-1} - X\beta)]d(\beta, \alpha)$ is the integrated likelihood with Θ being the parameter space.

The $P \times 1$ matrix of beta coefficients β arises from the mean structure of the multivariate latent multivariate Gaussian variable Y^*, so we assign a multivariate prior $\beta \sim N_p(h, H)$, where h is the mean and H is the covariance matrix. If h is small H is large, this becomes non-informative prior.

The thresholds α_j, where $\alpha_1 < \alpha_2 ... < \alpha_J$ and $\alpha_0 = -\infty, \alpha_1 = 0, \alpha_j = \infty$ for $j = 1, ..., J$, are jointly distributed with Y^*. We therefor impose a conjugate multivariate normal distribution on α: $\alpha \sim N_{J-2}(t, T)$, where t is $J \times 1$ vector and T is diagonal matrix. For $t \to 0$ and $T \to \infty$, this is also a non-informative prior.

The unknown values y_i^* are treated as nuisance parameters to be estimated along with β, α. Conditional on β, α, x, and y_i, y_i^* is a truncated normally distributed variable with mean $X\beta$ and variance 1. That is, the latent variable has a multivariate normal prior distribution with density:

$$p(y^* | \alpha, \beta, y, x) = N(v, V) \tag{11}$$

where v is the mean, and V is the variance. This is a doubly truncated normal distribution (truncated to the left by α_{j-1} and to the right by α_j).

Posterior estimation is done by setting up the Gibbs sampler [19], which requires us to derive full conditionals for all parameters:

$$p(\beta | y^*, y, \alpha) = p(y^* | \beta)p(\beta)$$

$$= \prod_{i=1}^{n_i} exp\left\{-\frac{1}{2}(y^* - X\beta)\Sigma^{-1}(y^* - X\beta)\right\} \times exp\left\{-\frac{1}{2}(\beta - h)^T H^{-1}(\beta - h)\right\} \tag{12}$$

$$p(\alpha | y^*, y, \beta) = p(y | \alpha)p(\alpha) \tag{13}$$

where $p(y | \alpha)$ is an indicator variable. The likelihood function of α_j is thus

$$L(\alpha_j | y_i) = \prod_{i=1}^{n} \sum_{j=1}^{J} \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \le \alpha_j) \times p(\alpha) \tag{14}$$

where $\mathbf{1}$ is $J \times 1$ vector of ones, $\alpha_1 < \alpha_2 ... < \alpha_{J-1}$ and $\alpha_0 = -\infty, \alpha_1 = 0, \alpha_j = \infty$ for $j = 1, ..., J - 1$. Thus, we obtain the full conditional posterior of α_j conditioned on $\alpha_{j \neq -j}$:

$$p(\alpha_j | \alpha_{j \neq -j}, y, \beta, \sigma^2, \phi) = \prod_{i=1}^{n}\left[\sum_{j=1}^{J} \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \le \alpha_j)\right] \times N_{J-1}(t, T) \times \mathbf{1}(\alpha_1 < \alpha_2 ... < \alpha_{J-1}).$$

$$= \prod_{i=1}^{n}\begin{Bmatrix} \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \le \alpha_j) + \\ \mathbf{1}(y_i = j + 1) \times \mathbf{1}(\alpha_j < y^* \le \alpha_{j+1}) \end{Bmatrix} \times exp\left\{-\frac{1}{2T}(\alpha_j - t)^2\right\} \times \mathbf{1}(\alpha_1 < \alpha_2 ... < \alpha_{J-1}) \tag{15}$$

This is a truncated normal distribution. We allow $t \to 0$ whiles $T \to \infty$ to impose vague prior on α_j. Thus,

$$p(\alpha_j | \alpha_{j \neq -j}, y, \beta, \sigma^2, \phi) = \prod_{i=1}^{n} \mathbf{1}(\alpha_j^{inf} < \alpha_j < \alpha_j^{sup}) \times exp\left\{-\frac{1}{2H}(\alpha_j - t)^2\right\} \tag{16}$$

MCMC techniques then make draws from the uniform distribution in the interval:

$$\left.\begin{array}{l} \alpha_j^{inf} = max\{max(y_i^* | y_i = j), \alpha_{j-1}\} \\ \alpha_j^{sup} = min\{min(y_i^* | y_i = j + 1), \alpha_{j+1}\} \end{array}\right\} \equiv [max(y: y \in j - 1), min(y: y \in j)]$$

$$p(y_i^*|y, \alpha, \beta) = p(y|y_i^*, \alpha, \beta)p(y_i^*)$$
$$= \prod_{i=1}^{n} \sum_{j=1}^{J} \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \le \alpha_j) \times exp\left\{-\frac{1}{2}(y_i^* - v)^T V^{-1}(y_i^* - v)\right\} \qquad (17)$$

The Gibbs algorithm to generate posterior samples is extended in the following steps:

i.   Choose initial values for all parameters including thresholds.
ii.   Sample the latent variable Y_i^* from truncated normal distributions on the truncation interval $[\alpha_{j-1}, \alpha_j]$.
iii.   Sample thresholds $\alpha_j, j = 1, \dots, J$ from uniform distributions on the interval
$[max\{max(y_i^*|y_i = j), \alpha_{j-1}\}, min\{min(y_i^*|y_i = j + 1), \alpha_{j+1}\}]$.
iv.   Sample the regression coefficients from $c = (X^T X)^{-1} X^T y^*$.
v.   Repeat the steps until enough samples are drawn for inference.

3.3. Posterior Predictions and Model Evaluation

Predictions of the unobserved Y_{i+1} are made using the posterior predictive density:

$$f(Y_{i+1}|y, \beta, \alpha) = \int_\Theta f(Y_{i+1}|y, \beta, \alpha)f(\beta, \alpha|y)d(\beta, \alpha), \qquad (18)$$

where Θ is the parameter space. The posterior predictive density is the distribution of possible future observations that could arise from the current model.

Subsequently, it is important to assess whether the model built fits the data sufficiently for prediction within and, possibly, beyond the observed data. That is, we determine a checking function for both predicted data and actual observations to assess whether the predictive model adequately reproduces at least some key features of the actual observations to conclude the model fits well. In standard ML analyses, a repertoire of fit measures includes R^2 and likelihood-ratio χ^2 statistics. In Bayesian analyses, predictive model selection is checked within the framework of Bayesian decision theory [20], where prediction accuracy is based on the Bayesian expected loss (the expected value of the loss function given the data) [20]. The loss function $L(Y_i, Y_{i+1})$, represented here by the squared prediction error loss, assigns zero loss to correct predictions [21]. For the multi-categorical application, the Bayesian expected loss is expressed as

$$E[L(Y_{i+1}, Y_i)] = \frac{1}{n}\sum_{i=1}^{n} E\left\{\left((Y_{i+1} - Y_i)\right)^T (Y_{i+1} - Y_i)|y\right\}$$
$$= \frac{1}{n}\sum_{i=1}^{n} E\left[\left\{(Y_{i+1} - E(Y_{i+1}|y)) - (Y_{i+1} - E(Y_{i+1}|y))\right\}^2|y\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E\left\{(Y_{i+1} - E(Y_{i+1}|y))^2|y\right\} + \frac{1}{n}\sum_{i=1}^{n}(Y_{i+1} - E(Y_{i+1}|y))^2. \qquad (19)$$

The optimal Bayes predictor (the category with the largest probability) is found at
$Y_{i+1} = E(Y_{i+1}|y)$ where $E[L(Y_{i+1}, Y_i)]$ is at its minimum.

Alternative procedures to measure fit include Bayesian p-values. To calculate the p-value, we first define a test statistic $T(y)$ (high values of which show poor fit) to be a function of the data, and $T(y_{sim})$ to be the same function but applied to simulated or training data. We then compute the Bayesian p-value as:

$$p - value = p[T(y_{sim}) \ge T(y)]. \qquad (20)$$

Equation (20) represents the proportion of simulated future datasets whose functional values $T(y_{sim})$ exceed that of the function $T(y)$ applied to the original data. That is, the probability that a future observation would exceed the observed data, given the model. So, an extreme p-value implies poor model fit [22]. We jointly apply the two fit measures (Bayesian expected loss and Bayesian p-value) to determine the model's ability to forecast future values.

4. Data Description

The study used secondary data from the Ghana Living Standard round 6 Survey. The Ghana Living Standard Survey (GLSS) focuses on the household as the socioeconomic unit, but collects information on individuals within the household and on the communities in which the households are identified. Data are gathered on wide ranging issues including demographic characteristics, education, economic activities, and health.

For the GLSS round 6, a total of 16772 households were sampled across Ghana. Cumulative proportional odds probit models are constructed to categorize households using a set of cut-point values (non-poor, poor, and extremely poor) based on the micro data of the GLSS.

5. Preliminary Analysis

In this section, we examine the relationship between poverty and its correlates using the ordered probit model. We subject the model to relevant cross-validation tests by randomly splitting our dataset into two parts: one part (5000) as hold-out set and the other (11772) as validation set. Our sample size of 16772 is relatively large to permit a hold-out dataset. We use comparisons of the two datasets to investigate the model's ability to estimate parameters and accurately predict future observations under similar conditions. This step is important [23], especially when dealing with non-Gaussian models.

First, using the hold-out set, we conducted a graphical analysis to test conformity of the model with the proportional odds assumptions [17]; the assumption that independent variables' effect on the cumulative odds does not change from one cumulative odds to the next. Results obtained show a fairly affirmative picture (Figure 1). Some deviations were observed in the effects of location and ecological zone, though.

Next, to examine the usefulness of the model for prediction, we fit a one-covariate model to the hold-out dataset. The model was fit using the procedure detailed in this study and implemented in [24]. We ran the MCMC chain for 10000 iterations. Posterior summaries for prediction and estimation were based on the last 5000 iterations of the chain.

Summaries of results for Bayesian expected loss (BEL) (0.21), and Bayesian p-value (0.364) show that the model appears very effective in predicting unobserved values, the low BEL values indicating barely noticeable mis-prediction losses.

Using 95% posterior credible intervals to summarize the range of possible values (Table 2), it is obvious that the model was successful at capturing the data-generating process. This means that the model provides correct conclusions about the association between the response and the predictor.
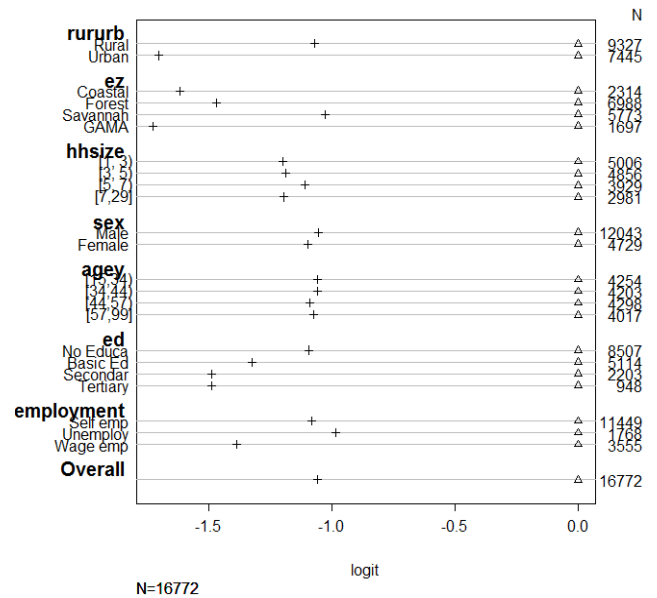

Figure 1. Test of Proportional Odds Assumption

Table 2. ML and MCMC Estimates for the Ordinal Probit Model

Coefficient		Ordered Probit Model					
		ML Estimation			Bayesian Estimation		
		Estimate	95% C.I.		Posterior Mean	Credible Interval	
						2.5%	97.5%
Intercept		1.551	1.409	1.633	1.523	1.411	1.633
Age		0.001	0.000	0.003	0.001	-0.0003	0.003
Household size		0.122	0.113	0.130	0.122	0.114	0.129
Sex	Male	-0.016	-0.075	0.043	-0.016	-0.072	0.039
	Female	Reference	-	-	-	-	-
Location	Rural	0.654	0.595	0.714	0.655	0.598	0.712
	Urban	Reference	-	-	-	-	-
Ecological zone	Coastal	0.280	0.118	0.447	0.275	0.116	0.435
	Forest	0.280	0.128	0.438	0.001	-0.078	0.079
	Savannah	0.876	0.723	1.036	0.596	0.521	0.676
	Accra	Reference	-	-	-	-	-
Education	No education	1.087	0.887	1.300	1.094	0.898	1.291
	Basic	0.720	0.518	0.934	0.773	0.672	0.876
	Secondary	0.316	0.099	0.544	0.368	0.311	0.425
	Tertiary	Reference	-	-	-	-	-
Employment	Self	0.079	-0.004	0.163	0.084	-0.005	0.163
	unemployed	0.163	0.052	0.275	0.163	0.058	0.273
	Wage	Reference	-	-	-	-	-
α_1		1.991	0.919	2.494	1.997	1.091	2.497
α_2		3.381	3.772	4.282	3.381	3.770	4.283
α_3		Fixed	-	-	Fixed	-	-
Pseudo-R^2		0.293	-	-	0.401	-	-

6. Model Results

In this section, we fit the model on the remaining dataset. Inference compared MCMC simulations (via Gibbs sampling) with the commonly used ML estimation. Results of the two models were generally identical. However, the Bayesian pseudo-R^2 value of 0.401 shows that the Bayesian model is superior to the ML model (pseudo-R^2 value = 0.293). More importantly, the Bayesian expected loss ($BEL = 0.17$), and p-value ($P = 0.464$) show that the Bayesian model is suitable for parameter estimation and prediction. For the MLE, coefficients of the threshold are $\alpha_1 = 1.997$ and $\alpha_2 = 3.381$, and differ significantly from each other (Table 2), with α_2 (poor) being more likely regardless of the covariates. The threshold values for the Bayesian model are similarly distributed (Table 2), so the use of the ordered probit model that categorizes the poor according to severity levels is better than using a single model to describe poverty. The observed differences in the threshold parameters basically reflect the differential parameterizations between the two approaches.

Table 2 summarizes and compares results of ML estimates and the Bayesian inference. The results show that all but the coefficients for sex, age and self-employment were statistically significant in both models. Whiles forest ecological zone was statistically significant in the ML model, its contribution to the Bayesian model was negligible.

Many of the statistically significant covariates were generally strongly related with the response variable. The coefficient of the variable household size for example, is statistically significant in both models; its contribution to change in log-odds ratio being 1.130. The impact of a unit change in the variable no education on the log-odds of a change in economic status is similarly high ($OR = 2.965$ for the MLE and $OR = 2.986$ for the Bayesian model). The converse is true for households whose heads have higher education. Using Tertiary education as reference, and the ML model, the odds ratios are 2.965, 2.054, and 1.372 for no education, basic, and secondary education. Living in the savannah ecological zone or rural place negatively impacts the log-odds of poverty among households in Ghana by 2.401 and 1.923 respectively. We also observed strong association between poverty severity and location (Rural/Urban). The estimated model is given in Table 2.

Marginal effects of the covariates were estimated for the Bayesian model (Table 3). In ordered modelling, a unit increase in the independent variable changes the probability of falling in the jth alternative by the marginal effect in percentage terms. For this study, results of Table 3 show that every one unit increase in household size decreases the probability of falling in the non-poor category by 2.8%, but increases the probability of poor and extremely poor by 1.9% and 0.09% respectively, implying that a 1 unit increase in household size tends to deteriorate the economic conditions of households (from non-poor through poor to extremely poor).

Table 3. Marginal Effects for the Bayesian Ordinal Probit Model

Poverty Status		Marginal effects for		
		Non-poor	Poor	Extremely poor
Age		0.000	0.000	0.000
Household size		-0.028	0.019	0.009
Sex	Female	-0.008	0.005	0.003
Location	Urban	0.149	-0.101	-0.049
Ecological zone	Forest	-0.002	0.001	0.001
	Savannah	-0.154	0.101	0.053
	Accra	0.072	-0.050	-0.022
Education	Basic	0.076	-0.052	-0.024
	Secondary	0.129	-0.090	-0.039
	Tertiary	0.147	-0.104	-0.043
Employment	unemployed	-0.017	0.011	0.005
	Wage	0.021	-0.014	-0.007

Moreover, urban location increases the probability of non-poor by 14.9%, but is associated with being 10.1% and 4.9% more likely to be in the poor or extremely poor categories respectively. Similarly, having a higher level of education significantly decreases the expected levels of poverty on the log-odds scale, tertiary education contributing the largest in this respect.

Based on results of Table 2, the Bayesian ordered probit (BOP) model equations for prediction are:

$$log\left[\frac{p(non-poor)}{p(poor) + p(extremely\ poor)}\right] = 1.997 + 0.112_{hh} + 0.655_{loc} + 0.596\ _{sav} + 0.275_{Co}$$

$$+1.094_{NoEd} + 0.773_{basic} + 0.368_{sec} + 0.084_{self} + 0.163_{unem} \qquad (21)$$

$$log\left[\frac{p(poor)}{p(extremely\ poor)}\right] = 3.381 + 0.112_{hh} + 0.655_{loc} + 0.596\ _{sav} + 0.275_{Co}$$

$$+1.094_{NoEd} + 0.773_{basic} + 0.368_{sec} + 0.084_{self} + 0.163_{unem} \qquad (22)$$

where hh = Household size, loc = location (rural/urban), sav = savannah ecological zone, Co = coastal ecological zone, $NoEd$ =No education, $basic$= Basic level education, sec = Secondary school certificate, and $self$ = self-employed, $unem$ = unemployed.

Predictions using (17) show that for a Ghanaian, the probability of being non-poor, poor or extremely poor, respectively, are 75.93%, 14.15% and 9.92%.

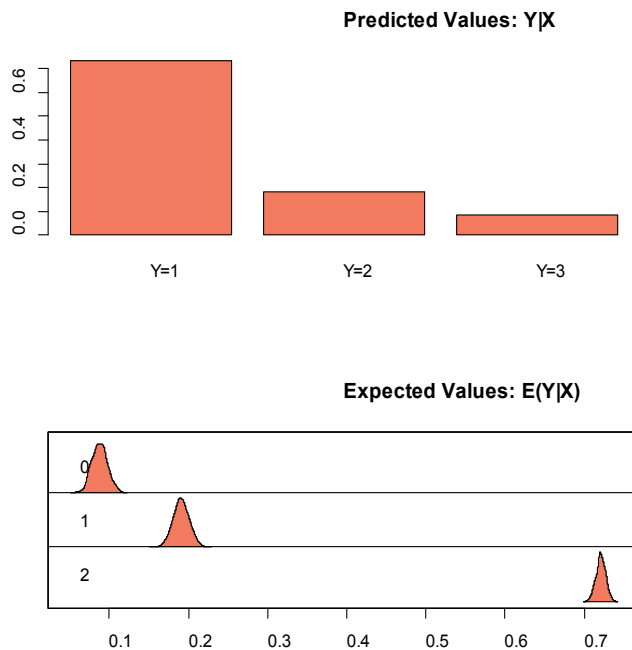Figure 2 presents results to compare predicted and expected probabilities of falling into poverty.

Figure 2. Predicted and Expected Probabilities of Falling in a Poverty Category

7. Discussion

The study employed the proportional odds cumulative probit model that ordered the population into three distinct categories (non-poor, poor, and extremely poor). The model was built to forecast the risk of household poverty in Ghana. Inference compared MCMC simulations (via Gibbs sampling) with the commonly used ML estimation. A key difference between the Gibbs sampler and that for the ML is the necessity to sample latent data from doubly truncated normal distributions. Though results were generally identical, the Bayesian pseudo-R^2 value of 0.401 shows that the Bayesian model is superior to the ML model with a pseudo-R^2 value of 0.310.

The estimated marginal effects clearly highlighted the strong association between poverty and its correlates. Similar findings were made by [25] who used both a bivariate and an ordered probit model to estimate the effects of different socio-economic and demographic variables on the probability of a household being in poverty in Malaysia. In their analyses, the authors found that the probability of being in poverty was higher for rural families. [26] also

applied the binary logit model to assess the causal relationship between poverty and some household-specific economic and demographic explanatory variables in Nigeria. They similarly found that being located in a rural community and having a large household size significantly increases the probability of falling below the poverty line. Many other studies have shown poverty to be predominantly rural and affected by large household size [5, 6].

Furthermore, our estimation results show significant variations in the relationship of the response variable with the different levels of categorical covariates. We see in Table 2, for example, that effect of the coefficients for education are inversely related to the probability of poverty, reducing the likelihood of dropping from non-poor to poor or extremely poor as education advances. This scenario was also found with the ecological zone variable. The high log-odds ratio value of the savannah ecological zone variable corroborate the work of [7] who reported that households in the savannah ecological zone of Ghana were almost four times poorer than those living in the coastal and forest zones in 2005/06. This could be attributed to the fact that the seaports and heavy industrial plants (along with the recently discovered crude oil) are found predominantly across the middle and coastal belts of the country, where both the rich agricultural lands and tropical rain forest coincide.

We also observe that being male reduced the risk of falling into poor or extremely poor 0.984. This finding is consistent with the work of [26] in Nigeria, who concluded that households with female heads were more likely to be poorer. This variable's effect is, however, not statistically significant.

8. Conclusions

The results we obtained show that out of the seven observed covariates, age, sex, as well as one level of ecological zone (forest) and self-employed do not significantly affect the distribution of household poverty in Ghana.

The most powerful predictors of extreme poverty include rural location, Savannah ecological zone and household heads with little or no education. We therefore entreat pro-education and pro-rural development agencies such as the Savannah Accelerated Development Authority (SADA) to intensify efforts aimed at combating poverty. We also recommend more diverse research into the area because the ability to correctly predict who is at risk should be the first step, and an integral part of efforts to combat poverty.

To improve parameter estimation, since the data was collected at different geographical locations, further research into the area using spatial analysis tools, instead of the standard regression methods used in this study, is recommended.

REFERENCES

[1] Ghana Statistical Service (2014). Poverty Profile in Ghana. http://www.statsghana.gov.gh/docfiles/glss6/GLSS6_Poverty%20Profile%20in%20Ghana.pdf.

[2] Townsend, P. (2014). International Analysis Poverty. Routledge.

[3] United Nations Development Programme (2010). What Will It Take to Achieve the Millennium Development Goals?– An International Assessment.

[4] United Nations Development Programme (2014). Summary: Human Development Report 2014; Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience

[5] Tomori, M., Elena, Zyka, E., & Bici, R. (2014). Identifying Household Level Determinants of Poverty in Albania Using Logistic Regression Model (SSRN Scholarly Paper No. ID 2457441). Rochester, NY: Social Science Research Network.

[6] Dudek, H., & Lisicka, I. (2013). Determinants of poverty–binary logit model with interaction terms approach. Ekonometria, (3 (41), 65–77.

[7] Ennin C.C., Nyarko P.K., Agyeman A., Mettle F.O., Nortey E.N.N. (2011). Trend Analysis of Determinants of Poverty in Ghana: Logit Approach. Research Journal of Mathematics and Statistics 3(1): 20-27, 2011, ISSN: 2040-7505.

[8] Agresti, A. (2007). An Introduction to Categorical Data Analysis. Wiley.

[9] Umer Khalid, Lubna Shahnaz and Hajira Bibi (2005). Determinants of Poverty in Pakistan: A Multinomial Logit Approach. The Lahore Journal of Economics. 10 : 1 (Summer 2005) pp. 65-81

[10] Samir B-E. Maliki, Abderrezak Benhabib, Abdelnacer Boutedja (2012). Quantification of the Relationship Poverty Education in Algeria: A Multinomial Econometric Approach. Topics in Middle Eastern and North African Economies, electronic journal, Volume 14, Middle East Economic Association and Loyola University Chicago, September, 2012, http://www.luc.edu/orgs/meea/.

[11] Fullerton A. S., 2009. A conceptual framework for ordered logistic regression models. Sociological methods and research 38 (2), 306-47.

[12] Terza, J. V., 1985. Ordinal probit: a generalization. Communications in Statistics-Theory and Methods 14 (1), 1-11.

[13] Gelfand, A.E. and A.F.M. Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities." Journal of the American Statistical Association.

[14] Garthwaite, P., Kadane, J. and O'Hagan, A. (2004) Elicitation Journal of the American Statistical Association.

[15] Ravallion, Martin; Chen, Shaohua; Sangraula, Prem (May 2008). Dollar a Day Revisited (Report). Washington DC: The World Bank.

[16] William H. Greene, David A. Hensher (2009). Modeling Ordered Choices.

[17] McCullagh, P., 1980. Regression models for ordinal data. Journal of the Royal Statistical Society, 42 (2), 109-142.

[18] Lance A. Waller, Carol A. Gotway (2004): Applied Spatial Statistics for Public Health Data. John Wiley & Sons, Inc., Hoboken, New Jersey. Pp 327.

[19] Stuart Geman and Donald Geman, (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE.

[20] Geisser, S. and Eddy, W. (1979) A predictive approach to model selection.

[21] Berger, J.., 1985. Statistical Decision Theory and Bayesian Analysis. Springer Verlag, New York, New York, USA.

[22] Rubin, D.B. 1984. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician." Annals of Statistics 12:1151–1172.

[23] Lewis, S. and Raftery, A. (1997) Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. Journal of the American Statistical Association, 92, 648–655.

[24] R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL http://www.R-project.org.

[25] Saidatulakmal and Madiha Riaz (2012). Demographic Analysis of Poverty, Rural-Urban Nexus. Research on Humanities and Social Sciences. Vol.2, No.6, 2012.

[26] Osowole, O.I., Ugbechie, Rita, Uba, Ezenwanyi (2012). On The Identification of Core Determinants of Poverty: A Logistic Regression Approach. Mathematical Theory and Modeling. Vol.2, No.10, 2012.