Impact of Household and Demographic Characteristics on Poverty in Bangladesh: A Logistic Regression Analysis

Nusrat Farah
Eastern Illinois University

Follow this and additional works at: http://thekeep.eiu.edu/lib_awards_2015_docs
Part of the Income Distribution Commons, and the Regional Economics Commons

Recommended Citation
http://thekeep.eiu.edu/lib_awards_2015_docs/3

This Article is brought to you for free and open access by the 2015 Awards for Excellence in Student Research and Creative Activity at The Keep. It has been accepted for inclusion in 2015 Awards for Excellence in Student Research and Creative Activity - Documents by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.
Impact of Household and Demographic Characteristics on Poverty in Bangladesh: A Logistic Regression Analysis

Nusrat Farah
Eastern Illinois University
Abstract

Poverty is multi-dimensional in nature and depends on interactions of various socio-economic factors. Several demographic and health factors can shape up the economic status of a household, and theory suggests that the ability of a household to earn a given level of income can depend on the characteristics internal to the household. While most studies done on poverty measurements rely on the income, expenditure and consumption data, this study has used data from Bangladesh Demographic and Health Surveys (DHS). The main objective of this paper is to identify the factors that have relative effect on poverty of the household. The principal component analysis was used to create an asset index which gave the Social Economic Status (SES) of each household. The variables were tested as a univariate model to see the effect on SES. Finally, a logistic regression was estimated based on this data with the SES (that is poor and non-poor) as the dependent variable and a set of demographic variables as the explanatory variables. The results presented in this paper suggest that the DHS data can be used to determine the correlates of poverty. The results also suggest that demographic and household data can describe poverty. The probability of a household being poor depends on the ownership of assets and other household data. A closer look was then taken to identify whether the results were driven by rural or urban property. The further analyses suggest that rural poverty is significantly affected by the demographic and household characteristics. The results were mostly driven by rural poverty rather than urban poverty. The paper has some policy values, as it suggests that rural people more suffer from the plight of poverty than do the urban people. Hence, the government needs to focus on improving the standard of living in rural areas of Bangladesh.

Keywords: Principal components analysis, Logistic regression, Demographic and Household Characteristics
Introduction

Poverty being multi-dimensional in nature is the outcome of various interactive socioeconomic factors. The incidence of poverty is purposive by multiple interactive factors operating at both micro and macro levels. 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.

Sahn and Stifel (2003) argued that 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 expenditure. In the developing countries, expenditure is preferred over income due to the difficulties involved in the measuring income. 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). Poverty can also be measured using demographic and household data.

The objective of this paper is to develop a poverty line using demographic and household data of Bangladesh. An asset index has been developed using Principal Component Analysis (PCA) following Achia, Wangombe and Khadioli (2010) from asset ownership variables in the Bangladesh Demographic and Health Survey (2011). A poverty line then applied to identify the poor and non-poor households. Finally, a logistic regression analysis has been done to identify the key determinants of poverty in Bangladesh.

Literature Review

The use of demographic and health survey data to the measure of poverty is not unique. Filmer and Pritchett (2001) used Demographic and Healthy Survey data to show that the relationship between wealth and enrollment in school can be estimated without income or expenditure data,
by using household asset variables. PCA provided acceptable and reliable weights for an index of asset to serve as a measure for wealth. In the four countries examined- India, Indonesia, Nepal and Pakistan - this approach produced reasonable results.


Booysen (2002) used demographic and healthy surveys to measure differences in socioeconomic status of South Africa households. The asset index used represented a comparable indicator of poverty in South Africa.

In their study, Geda et al. (2001) used binomial and ordered logit analysis and identified the following variables as the key determinants of poverty: size of household, places of residence (urban or rural), level of schooling and engagement in agricultural activity, both in rural and urban areas.

The study on the determinant of poverty was done by Oyugi et al (2000). In their study they used Probit Model to analysis the Welfare Monitoring survey (1994) data. The predictors (household characteristics) used in the study included holding area, livestock unit, the proportion of household members able to read and write, household size, sector of economic activity (agriculture, manufacturing/industrial sector or wholesale/retail trade), source of water for household use, and off-farm employment. The result showed that almost all the variables used were important determinants of poverty.

Rodriguez and Smith (1994) used a logistic regression model to estimate the effect of different economic and demographic variables on the probability of a household being in poverty in Costa
Rica. The source of the data was from National Household-Income (1986). Their results showed that poverty was higher for the household whose heads had a lower level of education.

An asset-index approach to the measuring of poverty is one alternative to income or consumption and expenditure. This approach, although lacking data on income, consumption and expenditure, collects information on ownership of a range of durable assets, which include: car/track; refrigerator; television; radio; bicycle; telephone and solar power; housing characteristics, which include material of dwelling floor; roof and toilet facilities; and access to basic services, which includes electricity supply and sources of drinking water. This approach was used by Achia, Wangombe and Khadioli (2010) to determine the explanatory variables for poverty using demographic and health survey data of Kenya.

Previous studies on poverty in Bangladesh relied on the expenditure and consumption data and used the poverty line computed from the Bangladesh Integrated Household Survey (BIHS) data using the cost of basic needs method or food energy intake method. While literature on poverty measurement is by now relatively advanced and abundant, there are very few studies dealing with finding the determinant or causes of poverty.

Few papers studied the impact of household characteristics on poverty in the case of Bangladesh, while most of these papers focused on few household and demographic characteristics. Also, a limited number of papers can be attributed to the impact of household and demographic characteristics on poverty in general.

Rahman (2013) examined how household characteristics affect the probability of a household being poor via labor market interactions. The central premise was that the ability of a household to exploit available income-earning opportunities is shaped by characteristics inherent to the household. He concluded that poverty is high in households with young household heads, low-
level education of the household heads, female heads, disability of household members, larger size, a predominance of female members, excessive dependency burden or a high proportion of female workers. However, no study was undertaken to see the impact of demographic characteristics on poverty particularly referring Bangladesh.

Methodology

The Data

The data used to analyze the poverty is taken from the 2011 Demographic and Health Surveys (DHS) for Bangladesh. The survey covered both rural and urban populations. The survey collected information relating to demographic and detailed information on asset ownership, access to public services and housing characteristics. A household was defined as a person or a group of people related or unrelated to each other who live together in the same dwelling unit and share a common source of food.

The survey is based on a two-stage stratified sample of households. In the first stage, 600 Enumeration Areas (EAs) were selected with probability proportional to the EA size, with 207 clusters in urban areas and 393 in rural areas. A complete household listing operation was then carried out in all the selected EAs to provide a sampling frame for the second-stage selection of households.

In the second stage of sampling, a systematic sample of 30 households on average was selected per EA to provide statistically reliable estimates of key demographic and health variables for the country as a whole, for urban and rural areas separately, and for each of the seven divisions. With this design, the survey selected 18,000 residential households to conduct the survey. Finally, 17,142 households were survey and the data was collected, summarized and presented in the household and demographic survey report.
Computation of an Asset Index using Principal Components Analysis

Principal Component Analysis (PCA) has been used to create an asset index based on data from the Bangladesh Demographic Household Survey (2011). The Bangladesh Demographic Household Survey (2011) included information regarding the ownership of durable goods, housing characteristic, access to services, along with basic demographic information concerning household size and composition.

Using PCA, the household variables were recoded into dichotomous variables, distinguishing between a household that owns the particular asset with the household that does not own that particular asset, or for which a particular statement about access to services is true with one that does not own the asset or for which the statement is not true. Hence, all variables take on a value of zero or one. The only variable that is included in the PCA as a continuous variable is the number of household members sharing a room for sleeping purposes.

The PCA is a multivariate statistical technique used to reduce the number of variables without losing too much information in the process. The PCA technique achieves this by creating a fewer number of variables, which explains most of the variation in the original variables.

The new variables that are created are linear combinations of the original variables. The first new variables will account for as much of the variation in the original data as possible.

Given \( p \) variables \( X_1, ..., X_p \) measured in \( n \) households, the \( p \) principal components \( Z_1, ..., Z_p \) are uncorrelated linear combinations of the original variable, \( X_1, ..., X_p \) given as:

\[
Z_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p
\]

\[
Z_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p
\]

\[...
\]

\[
Z_p = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p
\]
This system of equations can be expressed as $z = Ax$, where $z = (Z_1, ..., Z_p)$, $x = (X_1, ..., X_p)$ and $A$ is the matrix of coefficients.

The coefficient of the first principal component, $a_{11}, ..., a_{1p}$, are chosen in such a way that the variance of $Z_1$ is maximized subject to the constraint that $a_{11}^2 + ... + a_{1p}^2 = 1$. The variance of this component is equal to $\lambda_1$, the largest eigenvalue of $A$.

The second principal component is completely uncorrelated with the first component and has variance equal to $\lambda_2$, the largest eigenvalue of $A$. This component explains additional but less variation in the original variable than the first component subject to the same constraint. Further, principal components (up to the maximum of $p$) are defined in a similar way.

Each principal component is uncorrelated with all the others and the squares of its coefficients sum to one. The principal component analysis involves finding the eigenvalues and eigenvectors of the correlation matrix.

**Logistic Regression Model**

To identify key determinants of poverty, a dichotomous variable indicating whether the household is poor or not was calculated. That variable was named as Social Economic Status (SES). That is, SES is equal to 1 if the household is poor and 0 otherwise.

On the basis of Pearson's Chi-square statistic, it was determined whether the predictor variables—age of household head, size of household, educational level of the household head, type of residence (rural or urban), ethnicity, religion, sex ratio, dependency ratio, child-woman ratio and proportions of female members in the household were associated with the poverty index, SES. This helped to identify the important predictor variables associated with social status before setting up the final logit model. The main focus is to identify the determinants of poverty using the logit model from the available data.
A logistic regression model was used, given by

\[
\text{Logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{12} X_{12}
\]

where \( X_1, \ldots, X_{10} \) were the predictor variables—type of residence (rural or urban), educational level of the household head, region, size of household, age of household head, sex of household head, land ownership, sex ratio, dependency ratio, child-woman ratio, proportions of female members in the household, household condition (households with able-bodied persons and persons with disabilities), respectively; and \( p \) denoted the probability that the household was poor.

The forward selection, backward elimination and stepwise (logistic) regression methods determined automatically which variables to add or drop from the model. The conditional options use a computationally faster version of the likelihood ratio test.

**Results**

**The poverty index**

Table 1 show all variables used in the construction of the asset index and the result of the PCAs. Because of standardization, all principal components will have a mean close to 0. The standard deviation is also given for each of the components, and these will be the square root of the eigenvalue.

Interpretation of the principal components is based on finding which variables are most strongly correlated with each component, i.e., which of these numbers are large in magnitude, the farthest from zero in either a positive or negative direction. The correlations between the principal components and the original variables are very important, as it will help to create the asset index, which will be needed to determine the poverty line. The first principal components have been used, as there is zero correlation between the components.
The results of PCA indicate that the first principal component explains 18.3% of the variation in the original variables and each subsequent component explains a decreasing proportion of variance.

The scree plot in Figure 1 shows the proportion of variance explained by each principal component and indicates that the first four components would sufficiently explain the original variables.
In the construction of the social economic index, only the factor score (that's the eigenvectors) of the first principal component are used.

**Cross Tabulations:**

This section presents social economic status cross-tabulated by characteristics of the household, like education, household size, religion, region, age of household head, household owns land, sex ratio, child-woman ratio, dependency ratio and proportion of female members in household.

The asset index derived from the BDHS data was employed to calculate an estimate of the headcount poverty index for Bangladesh. The asset index at the 40th percentile is employed as the poverty line.

**Table II:** Values of Pearson's $\chi^2$ statistics on cross-classifying demographic characteristics with SES

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>$\chi^2$ value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of place of residence</td>
<td>4564.6259</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Highest educational level</td>
<td>3469.5520</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Religion</td>
<td>11.2236</td>
<td>0.0008</td>
</tr>
<tr>
<td>Number of household members</td>
<td>196.766</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Age of household head</td>
<td>241.4687</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Sex of household head</td>
<td>38.5561</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Region</td>
<td>915.3960</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Household owns land</td>
<td>1061.4747</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Sex ratio</td>
<td>18.8026</td>
<td>0.0003</td>
</tr>
<tr>
<td>Child-woman ratio</td>
<td>97.3628</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Dependency ratio</td>
<td>52.225</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Proportions of female members in household</td>
<td>1.1177</td>
<td>0.2904</td>
</tr>
</tbody>
</table>
The results indicate that there is association between SES and the following predictor variables: type of place of residence; religion; number of household members; highest education level; age of household head; sex of household head; region; land ownership; child-woman ratio; dependency ratio; and household condition.

According to the residence, the result shows that most cases of the very poor are in rural areas. The distribution of households by education and social economic status results show that the groups with the highest poverty cases have no education while those with the higher education have lower cases of poverty. The distribution of households by religion and the SES result show that religion is not significantly associated with poverty.

The household size was classified using three household size categories (1-5, 5-15 and 15+). The result shows that poverty is highest for a household with 9 or more members and lower for households of smaller sizes. As per the age of the household head, the results show that poverty increases with the age of the household head. As per the sex of the household head, it can be seen that most household heads are male. Households with a female head are poorer than the households with a male household head.

According to the region, the result shows that Chittagong has the lowest number of the very poor while Rangpur has the highest number of very poor households. But the number of non-poor households is high in the case of Dhaka. For the land ownership, it is seen that the non-poor household owns land more than the poor households. Most of the poor households do not have any land ownership. The result shows that households having persons with disability are more likely to live in poverty than households of able-bodied persons. This is indicative of relatively high poverty among households having persons with a disability.
The sex ratio is not significant in describing poverty for the households. The dependency ratio also shows that the households with a high dependency ratio are poorer than the household with a low dependency ratio. The proportion of female members also is not describing any influence on poverty. The poverty estimates are not directly comparable, given that different poverty lines, equivalence scale, time and data set are employed in estimating the headcount poverty index. The results are comparable with the Household Budget Survey of Bangladesh (2011).

**Logistic Regression Analysis**

The final model that was fit to the data was given by:

\[
\text{Logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10}
\]

Where,

- \(X_1\) = residence
- \(X_2\) = education of household head
- \(X_3\) = region
- \(X_4\) = number of household members
- \(X_5\) = age of household head
- \(X_6\) = sex of household head
- \(X_7\) = land ownership
- \(X_8\) = child-woman ratio
- \(X_9\) = dependency ratio
- \(X_{10}\) = Household condition

This logistic regression equation was arrived at using a forward stepwise selection method. Using the data, the logistic regression has been estimated to determine the factors affecting poverty in Bangladesh.
The results of the logistic regression are summarized in Table 3.

**Table III: Effects for the best fitted logistic regression**

<table>
<thead>
<tr>
<th>Logistic Regression for SES</th>
<th>Summary Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Null Deviance</td>
</tr>
<tr>
<td></td>
<td>Model Deviance</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
</tr>
<tr>
<td></td>
<td>p-Value</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Coefficients</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>Wald Value</th>
<th>p-Value</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Exp(Coef)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.98</td>
<td>0.08</td>
<td>-11.93</td>
<td>&lt; 0.0001</td>
<td>-1.14</td>
<td>-0.82</td>
<td>0.38</td>
</tr>
<tr>
<td>Education</td>
<td>-0.20</td>
<td>0.01</td>
<td>-18.48</td>
<td>&lt; 0.0001</td>
<td>-0.22</td>
<td>-0.17</td>
<td>0.82</td>
</tr>
<tr>
<td>No. of hh members</td>
<td>0.19</td>
<td>0.01</td>
<td>19.14</td>
<td>&lt; 0.0001</td>
<td>0.17</td>
<td>0.21</td>
<td>1.21</td>
</tr>
<tr>
<td>Age of household head</td>
<td>0.39</td>
<td>0.06</td>
<td>6.59</td>
<td>&lt; 0.0001</td>
<td>0.27</td>
<td>0.50</td>
<td>1.47</td>
</tr>
<tr>
<td>Owns land</td>
<td>1.14</td>
<td>0.04</td>
<td>30.66</td>
<td>&lt; 0.0001</td>
<td>1.07</td>
<td>1.22</td>
<td>3.14</td>
</tr>
<tr>
<td>Sex of household head</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.9223</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Region</td>
<td>-0.04</td>
<td>0.01</td>
<td>-3.94</td>
<td>&lt; 0.0001</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.96</td>
</tr>
<tr>
<td>Type of residence</td>
<td>2.01</td>
<td>0.04</td>
<td>46.09</td>
<td>&lt; 0.0001</td>
<td>1.92</td>
<td>2.09</td>
<td>7.43</td>
</tr>
<tr>
<td>Dependency ratio</td>
<td>0.12</td>
<td>0.03</td>
<td>3.49</td>
<td>0.0005</td>
<td>0.05</td>
<td>0.18</td>
<td>1.13</td>
</tr>
<tr>
<td>Child woman ratio</td>
<td>-0.35</td>
<td>0.03</td>
<td>-10.79</td>
<td>&lt; 0.0001</td>
<td>-0.41</td>
<td>-0.28</td>
<td>0.71</td>
</tr>
<tr>
<td>Household Condition</td>
<td>0.38</td>
<td>0.03</td>
<td>7.03</td>
<td>&lt; 0.0001</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Discussion**

The results show that all the variables significantly explain household socio-economic status except for the sex of household head. Theoretically, we can also explain this fact, as most of the household heads are male so poor households definitely have male heads. This variable is not explanatory for the social economic status.

We can take a further look to identify which variables explain poverty in urban and rural areas. This will help to determine whether the factors affecting poverty in rural and urban areas are the same. If not, then which factors are more explanatory in rural compared to urban?

The results show an interesting pattern. Almost all of the variables are explanatory for rural poverty except the regions, meaning that poverty in rural areas does not significantly depend on which part of the country they live. This is due to the vicious circle of poverty. Those who are poor are not migrating to other regions to earn more and are remaining poor, and those who are migrating to earn better income are not able to come out of poverty.
The households having able-bodied persons are less poor than the households having persons with disabilities. It is obvious because in rural areas of Bangladesh sometimes able-bodied persons remain unemployed. So, households with disabled persons suffer more from poverty.

The variables sex ratio and proportions of females in households are not significant, which supports the previous conclusion from the chi-square test. The dependency ratio does not also explain poverty in rural areas. This may be the reason most rural people work in the informal agricultural sector and do not have any retiring age. So, in rural areas the dependency ratio is not very high and does not explain poverty.

For urban poverty, none of the variables are significant except land ownership, region and household condition. Owning land is explanatory, as this can be a factor for pulling out a
household from being poor to non-poor. Land can be used for agricultural purposes or for resale. So, owning land is a very important indicator of being a non-poor household. Also, region, which was not significant in rural poverty, is significant in urban poverty, as in urban areas most households depend on the urban industrial sector for their living. Regions with industries have fewer poor households compared to regions with agricultural economy. This is because of job availability and high wages in the industrial sectors compared to the agricultural sector. So, region plays an important role in the case of urban poverty.

It is important to note that households having persons with disabilities suffer the plight of poverty in urban Bangladesh because of lack of opportunities of education and jobs. The logistic regression of rural and urban poverty suggests that most of the results in the initial logistic regression is driven by rural poverty.

Based on these findings, it can be concluded that rural poverty is high in households with young household heads, low-level education of the household heads, female heads, disability of household members, larger household size, excessive dependency burden and a high proportion of female workers.

The implications of the above findings for policy development are monumental, especially for a developing country like Bangladesh. The poverty of female-headed households or households with disability emphasizes the essentiality of bringing these households beneath the umbrella of social protection.

In the long run government will have to come up with measures that can absorb these households in remunerative employment, which is adequate to ensure a simple living. Introducing compulsory education for all, nurturing the idea of small family, promoting healthy habits and
family planning and ensuring civil rights of disabled persons can improve the poverty situation in Bangladesh.

**Conclusion**

Poverty measured on the basis of asset ownership is increasingly being used in research but there are some limitations on their use. According to Filmer and Pritchett (2001), the asset-based measures are more reflective of the long-run household wealth, failing to capture short-run wealth to the household. Therefore, if we are interested to see an effect that is associated with current resources to the household, then an index based on asset may not be the best measure. Also Falkingham and Namazie (2002) pointed out that the ownership does not capture the quality of the asset (Owning a poor-quality asset will not help to live a better standard of living). But asset-based measures do not account for this fact. Also, some variables may have a different relationship with the asset index across sub-groups; for example, ownership of farm land and live-stocks may be more reflective of wealth in rural areas than urban areas.

The multivariate analysis shows that an increase in educational level has an impact on the probability of a household being non-poor. Achieving higher education can lead a household from being poor to non-poor. The regression result shows that a rural family has a high probability of being poor. The rural/urban variable is statistically significant and this variable can be the reason of a household being poor.

The logistic regression of rural and urban poverty suggests that most of the variation was explained by rural poverty. The other demographic factors that affect the probability of a household being poor are the age of household head, religion, number of household members, sex of household head, land ownership, child-woman ratio and household condition. The sex of household head and dependency ratio when tested as a univariate model was statistically
significant with the social economic status, but it's not significant when included in the multivariate analysis.

The rural urban variability was significant when they were accounted for separately. This can be attributed to the fact that the assets included in the asset index are by their nature urban rather than rural, and therefore are biased against rural areas. Also, formal housing, water and sanitation facilities are urban services, which support the conclusion of having significant results in rural areas. As the country develops, the rural population will get access to these services, which will help the household to be non-poor. This paper shows the key determinants of poverty on the basis of demographic and household data.

This paper focuses on the point that rural areas suffer due to poverty compared to urban areas. The rural standard of living is far less than the urban standard of living in Bangladesh. Hence, the government should come up with policies that can ensure reduction of the plight of the poor people in rural Bangladesh.
Reference


