Estimating Poverty in India without Expenditure Data: A Survey-to-Survey Imputation Approach

David Newhouse & Pallavi Vyas

Motivation

- Latest poverty estimates based on consumption expenditure for India are from 2011-12.
- Understanding Indian poverty trends critical for both global poverty estimates and national policy debate.
- Usual (“Line-Up”) method used by the World Bank tends to overestimate poverty decline.

Data

- Consumption Expenditure: National Sample Survey (NSS) 4 Rounds of Data.
  - 61 (2004-05); 66 (2009-10); 68 (2011-12) and 72 (2014-15)
- Rainfall: Climate Hazards Group InfraRed Precipitation with Station (CHIRPS).
- The 2014-15 survey doesn’t collect data on household consumption expenditure.
- Contains variables in common with earlier rounds of NSS data (61, 66, 68) referred to as the source datasets.
- Wording and/or recall periods are similar.
- Sampling frame is common across all rounds of surveys.

Methodology

- Use Small Area Estimation methods (Ellbers, Lanjouw, and Lanjouw 2003) to predict welfare conditional on common variables in a 2014-15 “target” data set which does not contain household expenditure.

Step 1: Estimate relationship between per capita expenditure and explanatory variables using OLS in source surveys from 2004, 2009, and 2011. Explanatory variables chosen using the LASSO from a pool of candidate variables including:
  - Demographic: household size, age and gender of head, religion and caste.
  - Labor: Household principal industry, occupation and means of livelihood.
  - Expenses on Miscellaneous Services: Household services, recreation and transport.
  - District Explanatory variables: District means of household variables listed above.
  - All household and district variables are interacted with a linear time trend.
  - Add Rainfall Shocks: District’s deviation from mean historical rainfall and in square form.

Step 2: OLS regression:

$$\ln(y_{cht}) = X'\beta + u_{cht}$$

- Systematic component of poverty decline.
- **U**_{cht} is the disturbance term.
- **\eta**_{dcht} is the district cluster component. **\epsilon**_{cht} is the household component.

Step 3: Because of heteroskedasticity, in the error term and spatial correlation, re-estimate the equation Generalized Least Squares (GLS).

$$\ln(y_{cht}) = X'\beta_{GLS} + u_{cht}$$

- GLS weights are predicted variances of the error term from the OLS model.

Step 4: Predict welfare using Monte-Carlo simulations.

$$\tilde{y}_{cht} = X'\beta + \tilde{\eta}_{cht} + \tilde{\epsilon}_{cht}$$

- Simulate welfare in the target dataset by drawing **\eta**_{dcht}, **\epsilon**_{cht} and \beta 100 times.
- Assumptions re: parameters are the following:
  - **\beta** ~ N(0, \sigma_\beta^2)
  - **\epsilon**_{cht} ~ N(0, \sigma^2_{\epsilon})
  - **\eta**_{dcht} ~ N(0, \sigma^2_{\eta})

- Steps 1-4 are repeated separately for urban and rural households and then aggregated to obtain a national estimate.

Main Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>National Poverty rate</td>
<td>40.9</td>
<td>33.1</td>
<td>21.6</td>
<td>12.7</td>
</tr>
<tr>
<td>Urban Poverty rate</td>
<td>28.1</td>
<td>22.6</td>
<td>13.4</td>
<td>10.4</td>
</tr>
<tr>
<td>Rural Poverty rate</td>
<td>63.6</td>
<td>53.1</td>
<td>24.9</td>
<td>13.8</td>
</tr>
</tbody>
</table>
| Source: India National Sample Survey Office (NSSO) Surveys.

Model Selection

- **Three additional models were considered.**
  1. District Dummies x Time Trend: Each district level variable is interacted with a linear time trend.
  2. Expenditures at the Extensive Margin: Dummy variables for positive income/service expenditures.
  3. Constant Coefficient Model: Use only 2011 data as source data, no time trend interactions (most common method).

- **Four models tested by:**
  - **Projecting forward into 2011-12** based on 2004-05 and 2009-10 data. Compare actual poverty in 2011-12 with predictions of the four models.

Conclusions:

- The preferred model generates estimates of 10.4% in urban areas and 13.8% in rural areas, implying a national estimate of 12.7% for 2014-15. These estimates imply poverty reduction from 2011-12 at a rate that is approximately half as fast as the Line-Up method in urban areas.
- Combining a recent survey without consumption expenditure with multiple rounds of past expenditure surveys can generate accurate and informative poverty estimates.