WORLD BANK GROUP

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 (Elbers, 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's 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 its square term.

- **Step 2**: OLS regression:

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

$$u_{cht} = \eta_{ct} + \epsilon_{cht}$$

  - U_{cht} is the disturbance term.
  - η_{ct} is the district cluster component. ϵ_{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.

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

- Simulate welfare in the target dataset by drawing η_{ct}, ϵ_{cht}, and β 100 times.
- Assumptions re: parameters are the following:

$$\tilde{\beta} \sim N(\hat{\beta}_{GLS}, Var(\hat{\beta}_{GLS}))$$

$$\tilde{\eta}_{ct} \sim N(0, \hat{\sigma}_\eta^2)$$

$$\hat{\sigma}_\eta^2 \sim Gamma(\bar{\sigma}_\eta^2, Var(\hat{\sigma}_\eta^2))$$
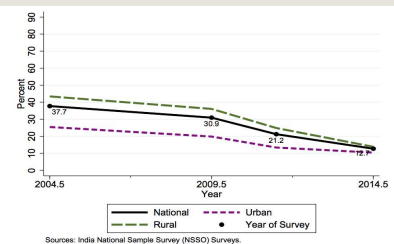
$$\tilde{\epsilon}_{cht} \sim N(0, \hat{\sigma}_{\epsilon_{cht}}^2)$$

- Use simulated welfare vectors to calculate national poverty estimates and standard errors for 2014-15.
- Steps 1-4 are repeated separately for urban and rural households and then aggregated to obtain a national estimate.

Main Results

	2004-05	2009-10	2011-12	2014-15
National				
Poverty rate	38.9	31.7	21.6	12.7
Standard Error	0.4	0.4	0.4	0.7
Urban				
Poverty rate	25.4	19.8	13.4	10.4
Standard Error	0.6	0.5	0.4	0.7
Rural				
Poverty rate	43.4	36.1	24.8	13.8
Standard Error	0.4	0.5	0.5	0.8

Sources: India National Sample Survey Office (NSSO) Surveys and staff estimates.

Sources: India National Sample Survey (NSSO) Surveys.

Model Selection

- **Three additional models were considered**.
1. District Dummies*Time Trend: Each district level variable is interacted with a linear time trend.
2. Expenditures at the Extensive Margin: Dummy variables for positive misc. 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.
  - **Reverse Projection into 2004-05** based on 2009-10 and 2011-12 data. Compare actual poverty in 2004-05 with predictions of the four models.
- All three additional models perform worse overall than the base model in these projection tests.

Predicted 2014-15 Poverty Rates ($1.90 Per Day)

	Model 1	Model 2	Model 3	Model 4
National	12.7	18.8	17.4	15.4
Urban	10.4	13.1	15.4	10.0
Rural	13.8	21.6	18.4	18.0

Sources: India National Sample Survey Office (NSSO) Surveys.

Comparison of Actual Poverty in 2011-12 with Forward Projection

	Actual	Model 1	Model 2	Model 3	Model 4
National	21.1	21.1	24.3	19.7	25.1
Urban	13.4	16.5	18.7	15.9	23.9
Rural	24.8	23.2	27.0	21.6	25.7

Comparison of Actual Poverty in 2004-05 with Reverse Projection

	Actual	Model 1	Model 2	Model 3	Model 4
National	37.5	40.7	48.9	64.1	28.5
Urban	25.4	30.4	37.0	48.4	19.5
Rural	43.4	45.8	54.7	71.7	32.9

Sources: India National Sample Survey Office (NSSO) Surveys.
Model 1: Final Model
Model 2: District dummies*Time Trend
Model 3: Expenditures at the Extensive Margin
Model 4: Constant Coefficient Model

Preferred Model vs. Typical Line-Up Method Predictions for 2014-15

	Model	Line-up
National	12.7	11.0
Urban	10.4	7.1
Rural	13.8	12.9

Source: India National Sample Surveys (NSSO) Surveys.

Other Robustness Checks

1. **Elasticity and Semi-Elasticity of predictions compared with typical line-up methods:** Results imply an elasticity of poverty of -2.8 which is consistent with past values based on actual survey data.
2. **Implied State Level Results:** Predicted poverty reduction is greater in states with higher rates of GDP growth.
3. **Predicted 2014-15 Poverty rates at $3.20 per day and $5.50 per day**
   $3.20 line: 49.4% rural, 33.4% urban, and 44.2% nationally. Decline of 16% from 2011-12.
   At $5.50 line: 83% rural, 65% urban and 77% nationally. Decline of 12% from 2011-12.

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.