CEEW-Columbia Survey: Overview of Sampling Strategy

Johannes Urpelainen 2014-10-10

States

- Bihar
- Jharkhand
- Madhya Pradesh
- Odisha
- Uttar Pradesh
- West Bengal

Sampling Strategy for Villages

- In total, we survey 714 villages in 51 districts.
- We begin by randomly sampling one district within each administrative division. Since West Bengal has large administrative divisions, two districts are sampled within each administrative division in West Bengal.
- Within the administrative divisions, a district is chosen for the sampling with a probability proportional to population size.
- Within each district, we select 7 large villages (from a sample containing 50% of district population) and 7 small villages (from a sample containing 50% district population).

State	Divisions	Districts	Villages	Responses
Bihar	9	9	126	1,512
Jharkhand	5	5	70	840
Madhya Pradesh	10	10	140	1,680
Odisha	3	3	42	504
Uttar Pradesh	18	18	252	3,024
West Bengal	3	6	84	1,008
Total	**48**	**51**	**714**	**8,568**

Table 1: Proposed Sampling Scheme

Representativeness of Survey Villages

- Within each district, the villages are representative of the district.
- Within each administrative division, districts are sampled proportional to their sizes. However, since we only select one or two districts within each administrative division, conclusions about the administrative division are shaped by the random draw of the district.
- At the state level, the same is true – the random draw of a small number of districts shapes the

outcome. It is important to consider the context of each specific district in drawing conclusions. To verify the representativeness of the survey, it is useful to compare the population characteristics to Census 2011 numbers.

- Across the entire study area, the sample is broadly representative since we draw 51 districts. However, survey weights must be constructed to account for variation in the population sizes of administrative divisions and states. Columbia researchers will create survey weights and share them with CEEW and Shakti Foundation.

Sampling within Revenue Villages

- Our initial plan was to rely on Census 2011 household lists. Unfortunately, these lists are not reliable in most of the study villages.
- As an alternative, we came up with an innovative sampling strategy that ensures that our sampling within each revenue village is representative. This method is itself notable, as it improves on standard survey practice in the energy field.
- The procedure is as follows:
    - Survey team enters the revenue village and consults *panchayat* officials or the village elder to secure an up-to-date list of all habitations and their household numbers.
    - The 12 survey households are allocated to the habitations in proportion to household numbers.
    - Within each habitation, the survey team selects a random spot based on available maps and sketches of the habitation. From this random spot, the survey team randomly selects a direction. Every 10th household in that direction is chosen for the survey.