**Sketch of Empirical Illustration**

The ultimate goal of the empirical illustration is to compare ranking households on the basis of potentials versus other forms of ranking (i.e., means test, proxy means test, and community ranking). To do so, we need to observe or generate each of these ranking schemes for the same households. Note that ranking on potentials and proxy means requires auxiliary data to estimate the model used to generate the rankings for the households of interest. This is often done by using multiple datasets: one (often nationally-representative) survey to estimate the model and then another more specific to the project areas to generate the scores or rankings. To keep things simple, we can “simulate” this procedure using one dataset split into a training and test sample.

The structure of the Alatas et al. data is presented in the table below. The data includes information on 640 “subvillages” from which 5,756 households were sampled. There is a fair amount of missing information in the data, so we are left with 5,040 complete cases. The subvillages are from three different provinces in Indonesia (North Sumatra, South Sulawesi, and Central Java) and were randomly assigned to three different targeting schemes (community-based targeting, hybrid, and proxy means testing). The table below presents the distribution of subvillages across these nine groups with the number of observations in each group in parentheses.

|  |  |  |  |
| --- | --- | --- | --- |
| Province | Community | Hybrid | Proxy Means Test |
| North Sumatra | 55 (429) | 56 (441) | 51 (386) |
| South Sulawesi | 61 (476) | 64 (505) | 64 (502) |
| Central Java | 98 (777) | 97 (783) | 94 (741) |

We, of course, only observe the community-based rankings for those subvillages that conducted the ranking exercise and there is no way to replicate this exercise for the subvillages that used hybrid or proxy-means targeting. (We may be able to retrieve the community rankings for hybrid subvillages, but I have not yet been able to locate that information.) For our training dataset, I thus propose using the data from the hybrid and proxy-means subvillages to estimate both our model and a more traditional proxy-means model (i.e., ordinary linear regression or random forest). We would then use the models to generate rankings on the basis of potentials and proxy-means scores for the subvillages that used community-based targeting. That is, the community-based targeting subvillages would be the test sample. This strategy would allow us to compare the three different ranking schemes within the subvillages that used community-based targeting. Technically, we also observe consumption for each household in the community-targeting subvillages, so we could also generate a fourth ranking akin to a full means test.

As a side note, Alatas et al. note that the provinces are fundamentally different. As a result, it might be beneficial to estimate unique models for each province. This may depend on how many observations we require for the BART procedure.