Using Big Data and Small Area Estimation to transform poverty mapping

opml.co.uk/insights/using-big-data-and-small-area-estimation-transform-poverty-mapping

Social protection programmes and appropriate policy are impactful drivers of poverty reduction, but they need up-to-date, comprehensive and accurate data in order to effectively tackle the causes of poverty. In this article we'll discuss advanced statistical techniques using Big Data to generate granular poverty estimates at a much lower cost and more frequently than ever before, enabling more dynamic conversations to help end extreme poverty.

Limitations of current poverty estimation methods

Current methods and analysis techniques often used in poverty reduction policies rely on expensive surveys and infrequent census data. This means that policy discussions can often lack dynamism, and might be at risk of sliding off the political agenda between census years.

Also, current poverty estimation and analysis often lack granularity, with a focus frequently at an 'urban vs rural' level rather than comparing districts or even villages. This can result in public resources being inefficiently deployed rather than being tailored for the specific area that needs it most.

Thanks to new estimation techniques, harnessing Big Data, our 'Small Area Estimation' method allows policymakers to generate more frequent, granular poverty estimates. This provides country-wide statistics at a fraction of the cost of a census, lowering costs for governments, NGOs and think tanks. It can also help determine how current policies impact poverty rates and guide future policy.

What is small area estimation?

Small Area Estimation is a statistical modelling technique. It allows a user to estimate data for small sub-populations that are not necessarily included in a household survey. It does this by using common covariates (a variable related to the dependent variable being studied) from a different dataset.

Small Area Estimation is not a new technique. Our guide illustrates the use of the Fay-Herriot model, which was first published in 1979. However, in the past, it has always used a combination of sample surveys and population censuses, resulting in researchers having to wait 10 years for the next census to generate the succeeding poverty map.

Now, using our innovative estimation technique, we can use granular geospatial data as an alternative to population censuses, enabling us to:

- Generate poverty estimates in and out of sample areas;
- See the spatial distribution of poverty across the country; and
- Generate poverty estimates much more frequently.

This new development has been made possible by technological improvements such as high-speed computers, as well as key innovations in modelling tools and data integration which allow analysis between geospatial and household data.

An 8-step procedure

Our Data & Evidence to End Extreme Poverty programme (DEEP) has produced an SAE guide. This provides an 8-step procedure which outlines the whole process of Small Area Estimation using geospatial data, from sourcing the geospatial data online to producing poverty maps like that of Bangladesh in Figure 1.

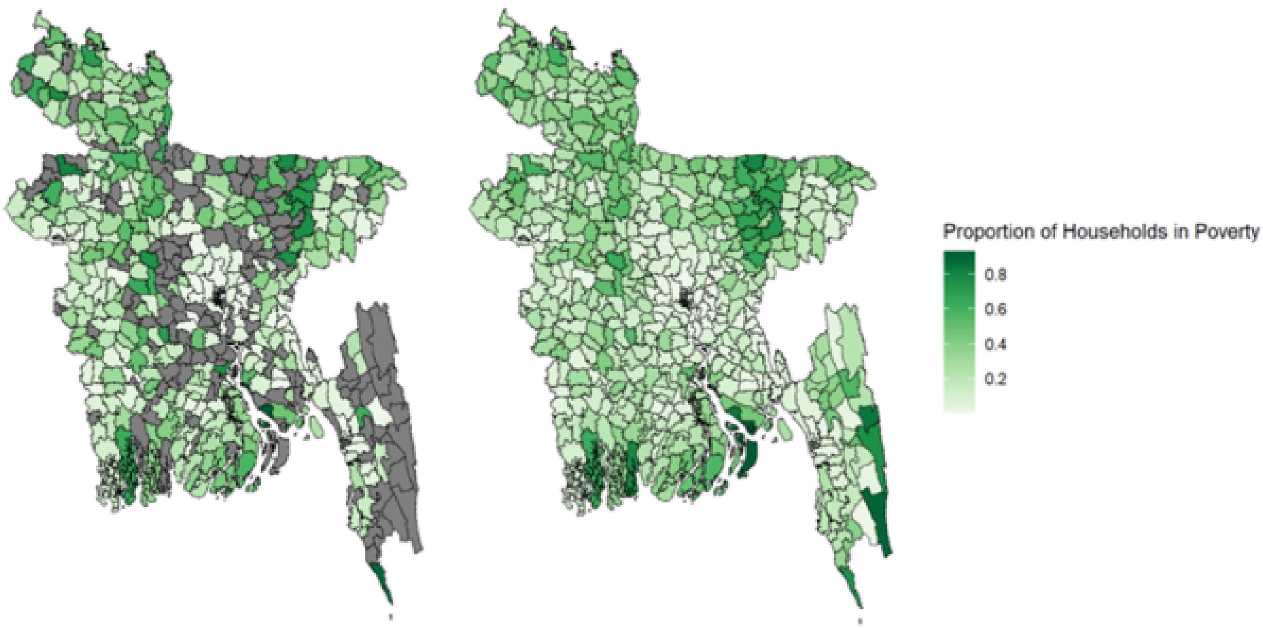
The first step is to gather the geospatial data and harmonize the data into a single rasterised dataset. We can then merge the geospatial and household datasets and fit the Fay-Herriot model, carrying out the necessary diagnostic checks. After running the model, the results are visualised on a country-wide map.

The geospatial data DEEP uses include accessibility data from Malaria Atlas, demographic maps from Meta, VIIRS nightlights data, and topography data from WorldPop. This provides over 150 geospatial covariates which give us sufficient explanatory power for out-of-sample prediction.

Comparing methods to estimate poverty in Bangladesh

In Figure 1, there are two maps showing poverty in Bangladesh. The variable illustrated reflects the likelihood that a household in the upazila (a small administrative region) is among the poorest 20% in Bangladesh.

Figure 1.



The graphic on the left shows data we used from the DHS (Demographic and Health Survey). The grey areas indicate areas not covered by this data. We found that 157 out of 522 upazilas were not included – more than 30% of the regions.

For the graphic on the right, we used our Small Area Estimation model to estimate poverty. This provides poverty estimates for all regions, which may be especially significant because we believe that areas not covered by the initial data, such as the Chittagong Hill tracts and the Sundarbans, may be disproportionately disadvantaged.

We intend to repeat this process on DEEP's other focus countries to learn if we can identify any common characteristics. Furthermore, we are producing time-series poverty distribution maps in Bangladesh to observe which areas are responding best to current policies and which areas might be lagging behind.

The benefits of Small Area Estimation

By using Big Data, this innovative Small Area Estimation method adds another key layer of analysis to a normal household survey. We can robustly extrapolate results to generate poverty estimates in small regions across the whole country.

It could influence and improve policy evaluation and social protection programs. Frequently updated, granular poverty estimates can reveal which areas respond most positively to different policy measures, enabling governments to make services more bespoke to the area in need.

Furthermore, the public availability of geospatial data ensures that this method is available to a wider audience of agencies. This should add valuable momentum to the poverty reduction conversation.

The method also has academic benefits, since it:

- Closes data gaps by producing policy estimates for previously hard-to-survey areas;
- Integrates data by combining geospatial and household data;
- Improves the precision of estimates compared to previous methods.

Research from UC Berkeley has found using geospatial data in Small-Area Estimation to be very precise. When compared to ground-truth data in Togo, the Small Area Estimation explained 84% of the variation in wealth (Chi et al., 2022).

A real-world purpose

Leveraging data and technology presents a powerful opportunity to drive meaningful change and accelerate progress towards ending extreme poverty. It provides a cost-effective and accurate way to generate high-frequency, granular poverty estimates, enabling policymakers and organisations to make data-driven decisions and tailor their activities to the areas that need them most.

If you're interested in learning more about this innovative approach or would like to explore potential collaborations, we encourage you to download the full DEEP guide on Small Area Estimation. Alternatively, you can reach out to the author, Nick Lindsay, at Nick.Lindsay@opml.co.uk or another member of the Data Innovation team for further information.

About the author:

Nick Lindsay is an Assistant Consultant in the Data Innovation team, specialising in quantitative methods and machine learning. He has technical experience using advanced statistical techniques to leverage national datasets and employing mixed-methods approaches.