DEEP RESEARCH INSIGHT 01

How can new technology support better measurement of extreme poverty?

Paul Jasper

DATA & EVIDENCE
deep
TO END EXTREME POVERTY

June 2021

Available at: https://doi.org/10.55158/DEEPWPR1

Key points

- The data revolution – the use of new data sources and analytical techniques derived from new technology – can significantly influence efforts to measure and tackle extreme poverty in low- and middle-income countries (LMICs).
- While new data sources will not replace existing tried and tested approaches to poverty measurement, such as in-person survey and census data, the two can be used together to provide better evidence on poverty in LMICs.
- The review identified many studies combining traditional data with data from remote sensors (e.g. satellite imagery) and the data exhaust (e.g. Call Detail Records (CDRs)) for high-resolution poverty mapping and targeting. This approach is known as a 'data sandwich'. There are fewer studies on the use of other forms of new data to measure poverty in LMICs (e.g. citizen-reported data and online data).
- The recommendation is that DEEP's future research focus on three main areas. First, further exploring 'data sandwich' approaches, including how innovative analytical methods can make high-resolution poverty estimates more precise. Second, trying less explored data collection approaches – e.g. crowd-sourcing or mobile phone surveys – to measure poverty in LMICs. Third, investigating how an integrated data system for measuring poverty in LMICs could work.

Summary

Policies, strategies and programmes require access to high-quality and timely information to tackle extreme poverty and to monitor progress towards achieving the Sustainable Development Goals (SDGs). However, this information is not always available, is available at high cost, or only at the overall national level (i.e. not at lower levels, such as in regions or municipalities). Traditional approaches to poverty measurement rely on household surveys or census data, which are costly and therefore collected infrequently, leading to gaps in data. New sources of data – including remote sensing data, data from the data exhaust, online data, crowd-sourced data, and mobile phone survey data – and novel statistical techniques can start to fill this gap and to enhance the toolkit of approaches available for measuring and investigating extreme poverty.

This Research Insight summarises the findings of a longer review produced by the Data and Evidence to end Extreme Poverty (DEEP) research programme to explore how innovation in data collection, data processing, and data analysis might provide solutions to 'pinch points' in policymaking and policy management for poverty reduction. The focus is on exploring the different data sources that can be used to measure and investigate poverty, including their advantages and limitations. The focus is on data sources that can be used to help measure extreme poverty directly, as well as related attributes such as determinants, proxies, or correlates of poverty. The research insight discusses if these new technologies can enable better insights into the incidence, distribution, and severity of extreme poverty over space and time, and where the most promising avenues for adopting these methods or investing in further research lie.

Introduction

The World Development Report 2021 'Data for Better Lives' argues that, 'innovations resulting from the creative new uses of data could prove to be one of the most life-changing events of this era for everyone. (…) the transformations emerging from the data revolution could touch all aspects of societies and economies.'[1] The same holds for efforts to measure poverty and track progress towards SDG 1.1 (eradicating extreme poverty globally by 2030).

This Research Insight summaries a longer review of approaches that use data from new technologies to measure and research poverty in LMICs. It focuses on approaches that examine data sourced from remote sensing technology, the data exhaust, online activities, crowd-sourcing efforts, and remote surveys. It also briefly touches on statistical learning techniques that allow researchers to extract value from these data sources. This Insight provides an overview of how innovations in the use of new data are affecting the global effort to measure poverty, and assesses where future research in the context of DEEP could add the greatest value in delivering on the promises of the 'data revolution'.

The poverty data gap

The starting point for most approaches reviewed here is what we label the 'poverty data gap'. This data gap exists because, for various reasons, the traditional data ecosystem for measuring extreme poverty in LMICs, which uses in-person sample survey and census data, provides stakeholders with information that is either out of date, not available at the right geographical level, or too static. New data sources can help close this gap.

The importance of statistical learning

Extracting insights and value from new data sources to help close the poverty data gap relies on improvements in the analytical tools used by researchers and analysts – in particular 'statistical learning' -methods for rapidly analysing very large datasets.

In this context, statistical learning methods are mainly used for two purposes. First, to process unstructured data so they can be used for analysis. For example, statistical learning algorithms extract relevant information from large sets of satellite images, which it would take humans a very long time to go through. Second, to estimate or predict poverty indicators in a certain geography or time period for which an estimate was not previously available. For example, estimating poverty rates in regions where a survey has not been implemented – sometimes referred to as 'out-of-sample prediction'.

Traditional data: in-person surveys and censuses

Traditional data sources – chiefly in-person surveys and censuses – currently form the backbone of efforts to measure poverty across the world (See Box 1). However, they are expensive and challenging to implement at a high quality. Nationally representative surveys easily cost between 1 and 2 million US dollars. This means data collection does not happen very often, with censuses implemented roughly once every 10 years and sample surveys every three to five years in many LMICs. This contributes to the 'poverty data gap', with data often being out of date.

[1] https://openknowledge.worldbank.org/handle/10986/35218, p. 3.

Box 1: Living Standard Measurement Surveys

Living Standard Measurement Surveys (LSMS) are the leading example of international poverty measurement efforts. In these surveys representative samples of households are asked questions about their consumption and expenditure, to compute daily consumption levels, which are then compared to a poverty line to identify households living in extreme poverty. Other poverty-related indicators (e.g. household asset ownership) are also estimated. Estimates of extreme poverty are produced at the geographical level at which the sample is representative – which is usually the national level or one administrative level below that. Combining these data with census data can help estimate the absolute numbers of people in extreme poverty. Researchers can also combine census data with survey data to produce higher resolution poverty estimates (e.g. at lower administrative levels).

Despite their advantages, new data sources do not render survey and census data obsolete, for two reasons. First, because these traditional data serve as the 'ground truth' against which statistical learning algorithms are trained. This means that these survey data provide the 'true picture' of poverty in a certain time period or geography that is used to build models that then try to predict poverty outside of the sample. Second, given the wealth of information collected in them, there is currently no substitute for the depth of insights that household surveys can provide on different forms of poverty, its manifestations, and its determinants.

Table 1: New sources of data included in the review

Data source	Definition
Remote sensing	Data collected from satellites, planes, and other kinds of earth observation sensors.
Data exhaust	Data that are inadvertently produced as a by-product of people's interactions with digital services in their daily lives (e.g. when they use the internet or their mobile phones).
Online information	Data that individuals share online, such as online articles, social media posts, and Wikipedia.
Citizen-reported data	We define citizen-reported data – sometimes referred to as crowd-sourced data – as data deliberately submitted by citizens via their mobile phones or the internet.
Remote surveys	These are online or mobile phone surveys.

Remote sensing

Remote sensing refers to data – mostly stored in the form of images – that are collected from satellites, planes, and other kinds of earth observation sensors. Over recent decades the coverage, availability, and frequency of these images has increased significantly, and their costs have dropped considerably. All of this has increased interest in exploring their use for monitoring the SDGs, some of which can be directly observed from space (e.g. target 2.4 on land cover under productive and sustainable agriculture).

While using these data to monitor SDG target 1.1 on extreme poverty is more difficult (we can't observe household consumption from space), attempts are being made to combine remote sensing information with surveys in order to predict poverty, and to close the poverty data gap. Features that researchers have explored in this area are night-time lights (human light sources that can be seen from space, revealing areas of human activity/habitation), vegetation indices (vegetation cover that can be seen from space), geographic conditions, climate, and characteristics of dwellings (e.g. roof types). To identify features that are relevant to predicting poverty, recent approaches have employed 'transfer learning' analyses, where algorithms are left to their own devices to extract such features. Innovations in this area are happening fast, and methodological advances in statistical learning are feeding into them. Overall, given their low cost, high spatial resolution, high coverage, and the frequency with which they are updated, data from remote sensors offer considerable potential for our efforts to measure and address extreme poverty.

Data exhaust

The data exhaust refers to sources of data that are inadvertently produced as a by-product of people's interactions with digital services in their daily lives (e.g. when they use the internet or their mobile phones). Mobile phone ownership and internet penetration have increased significantly over the past decade, including in LMICs, with approximately 5.2 billion mobile phone subscribers worldwide by the end of 2019 – roughly 67% of the global population. With this has come vast amounts of data. Researchers are experimenting with using these data to estimate or 'predict' poverty estimates for areas, time periods, and at geographical levels that it was not previously possible to address. Again, statistical learning methods are a part of this process. A key set of data for these approaches is Call Detail Records (CDRs), which contain meta-data on mobile phone activity (e.g. location, movement, number of calls). Other approaches experiment with data from mobile money, airtime, and credit transactions to map poverty or to identify potential beneficiaries for social assistance.

There are two important benefits of the data exhaust. First, the data exhaust from mobile phones allows individual-level predictions of phone-owner wealth or poverty. Second, in comparison remote sensing at least, the data exhaust can provide information that is more relevant to the problem of extreme poverty measurement, since the data are more directly related to poverty and wealth (e.g. mobile money use).

However, there are three main drawbacks. First, data in the data exhaust are mostly owned by private companies, which means that accessing these data can be difficult for policymakers and researchers. Second, mobile phone ownership and internet access is not universal, which makes data exhaust coverage non-universal, as well as having the potential to bias results. Third, there are ethical concerns with using these data, relating to issues of privacy and citizen surveillance.

Online information

Online information refers to data that individuals share online, such as online articles, social media posts, and Wikipedia entries. In contrast to the data exhaust, these data are not a by-product of online interactions but are explicitly created by users (e.g. writing a Wikipedia article). Despite the trend of advancing internet penetration, our review identified only a limited number of applications of online information for measuring poverty. Some of these employ geo-located online data in the same way as CDR or satellite imagery data (e.g. to produce high-resolution maps of poverty). Others focus on specific components of the poverty measurement challenge: for example, automatically analysing news and social media data to track food price fluctuations, which can feed into poverty estimates.

As with the data exhaust, online information comes with an inherent bias towards wealthier populations. In addition, given its intentional nature, it is biased towards populations with a sufficiently high level of education for active interaction with the internet. And in terms of population coverage it is less relevant than the data exhaust, as mobile phone ownership is much more prevalent than internet access in LMICs. An important feature that differentiates online information from other data sources is that it often comes in an unstructured text format, which means researchers may need to employ statistical learning techniques to pre-process, engage with, and make us of it.

Citizen-reported data

Citizen-reported data – sometimes referred to as crowd-sourced data – are data that are purposefully submitted by citizens via their mobile phones or the internet. In contrast to surveys, including remote surveys, this means that respondents or contributors self-select into sharing data. Depending on the technology available, a wide variety of different data types can be shared: images (including tagged with text), voice recordings, videos, text, GPS coordinates, and other mobile phone sensor data.

The crowd-sourcing of data has been used frequently in environmental sciences, health, geographic mapping, and the monitoring and evaluation of social programmes, but our review found very few applications of these data for measuring poverty. One example is crowd-sourcing price data in order to construct price indices. Another is using crowd-sourced geographic data from the OpenStreetMap (OSM) platform and combining them with satellite data in order to improve poverty mapping.

Proponents of citizen-reported data point to the potentially empowering nature of contributing data, i.e. the potential of raising citizen agency in the process of data production and use. It is clear, also, that the variety of data types that can be collected with this approach holds great promise for new ways of trying to estimate poverty. However, this approach can also face issues of poor data quality and bias, given the unstructured way in which these data are collected.

Remote surveys

Remote surveys are surveys carried out in LMICs using mobile phones. In contrast to citizen-reported data, these surveys employ a sampling approach and do not rely on self-reporting. In contrast to the traditional face-to-face approach to poverty measurement, data are collected remotely via either phone calls or SMS messages. Over the last decade the interest in this data collection approach has significantly increased for three main reasons. First, costs per interview are significantly lower than in face-to-face surveys. Second, mobile phone ownership in LMICs has increased significantly, which means that larger shares of the population can be reached in this way. Third, the COVID-19 pandemic has prevented many face-to-face interviews going ahead.

In remote surveying, whether via a call or via SMS, interviews must be short. This means this approach can't be used to deliver comprehensive consumption or expenditure modules to measure poverty. Researchers therefore have to investigate alternative indicators and proxies related to poverty that can be measured using this approach. In our review we found few uses of remote surveys to measure poverty in LMICs, which indicates that further research might be needed in this area. On the other hand, there has been a lot of activity in this area since the beginning of the COVID-19 pandemic. For example, in the US, the census bureau has been running an online 'household pulse survey', an online survey to assess the effects of COVID-19 on household socio-economic well-being. The landscape around remote surveys is changing rapidly.

Possible areas of DEEP research

Our review revealed that there is a very active and rapidly increasing literature exploring the benefits and costs of using innovative data approaches to measure poverty in LMICs. There are still many open questions relating to how to extract the best value out of this new data ecosystem, while dealing with relevant issues around (for example) data quality, ethical concerns, and biases. We suggest that the following three main areas could be of particular interest for further investigation by DEEP:

- First, our review revealed that there is still comparatively little research on the potential uses of online data, citizen-reported data, and remote surveys for measuring poverty in LMICs. Exploring methods to deal with data that are not text-based or questionnaire-based (e.g. images from smartphones) is a particularly exciting – but experimental – frontier.
- Second, we also find that many approaches integrate data from different 'new sources' in order to improve poverty estimates. Further research would be useful on how to best do this integration across data sources, in particular taking into account advanced statistical learning methods and focusing on extreme poverty as the target of estimation.
- Third, we suggest investigating how such estimation and data collection approaches could be integrated into national statistics ecosystems in a useful way. This implies tackling questions related to ethical concerns, data quality, trust in highly modelled results, and actual demand for new poverty measures. It could also include testing prototypes and capacity building in specific national statistics offices, who are often the main national institution responsible for producing, analysing, and publishing national statistics on poverty.

About DEEP

Our mission is to build evidence, insights, and solutions that help end extreme poverty globally.

We aim to contribute to new global and national data and evidence that governments, decision makers, citizens and researchers can use to improve people's lives and support the world's poorest people in their efforts to escape extreme poverty.

We are a consortium of the Universities of Cornell, Copenhagen, and Southampton led by Oxford Policy Management, in partnership with the World Bank's Development Data Group and funded by the UK Foreign, Commonwealth & Development Office.

Suggested Citation: Jasper, P. (2021) How can new technology support better measurement of extreme poverty? *DEEP Research Insight 01*. Data and Evidence to end Extreme Poverty Research Programme: Oxford.

Based on the full working paper:

Jasper, P.; Luna Hernandez, A; Scott, M. and Tzavidis, N. (2021) How can new technology support better measurement of extreme poverty? *DEEP Working Paper 02*. Data and Evidence to end Extreme Poverty Research Programme: Oxford.

For more information email **contact@povertyevidence.org**

@povertyevidence

@povertyevidence

Or visit **povertyevidence.org**