Poverty Mapping

Innovative Approaches to Creating Poverty Maps with New Data Sources

Virginia Ziulu
Jessica Meckler
Gonzalo Hernández Licona
Jozef Vaessen

IEG Independent Evaluation Group | WORLD BANK GROUP World Bank • IFC • MIGA

IEG Methods and Evaluation Capacity Development Working Paper Series

1818 H Street NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

ATTRIBUTION
Please cite the report as: Virginia Ziulu, Jessica Meckler, Gonzalo Hernández Licona, and Jozef Vaessen. 2022. *Poverty Mapping: Innovative Approaches to Creating Poverty Maps with New Data Sources.* IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group. Washington, DC: World Bank.

MANAGING EDITORS
Jos Vaessen
Ariya Hagh

EDITING AND PRODUCTION
Amanda O'Brien

GRAPHIC DESIGN
Luísa Ulhoa
Rafaela Sarinho

Poverty Mapping

Innovative Approaches to Creating Poverty Maps with New Data Sources

Virginia Ziulu, Jessica Meckler, Gonzalo Hernández Licona, Jozef Vaessen

Independent Evaluation Group

August 2022

CONTENTS

AUTHORS

Virginia Ziulu[1]

Jessica Meckler[2]

Gonzalo Hernández Licona[2]

Jozef Vaessen[1, 2]

Corresponding Author

Jozef Vaessen: jvaessen@worldbank.org

Author Affiliations

[1] World Bank Independent Evaluation Group

[2] The Global Evaluation Initiative

ABSTRACT

Geographically disaggregated poverty data are vital for better understanding development issues and ensuring development efforts are directed to the places where they are most needed. Poverty has traditionally been measured by data on consumption, income, or assets. However, recent advances in computing power and the emergence of new methods has made it increasingly feasible to produce reliable, cost-effective, and timely poverty maps by extracting features from novel data sources such as satellite imagery, call detail records, and internet connectivity indicators.

This paper explores the methodological implications of using both traditional and novel data sources to generate poverty maps. Specifically, it examines the applications of (i) survey and census data; (ii) Global System for Mobile Communications, smartphone, and Wi-Fi indicators; (iii) call detail records; (iv) daytime and nighttime remote sensing imagery; and (v) the Survey of Well-being via Instant and Frequent Tracking for poverty mapping. Each section provides a brief overview of the data requirements, methodology, and applicability considerations of the data source under consideration. In addition, the paper discusses the usefulness and limitations of each approach in the field of evaluation, providing concrete examples of poverty maps created from each of the listed data sources.

ABBREVIATIONS

CDR	call detail record
CNN	convolutional neural network
DHS	Demographic and Health Surveys
ELL	Elbers, Lanjouw, and Lanjouw method
G	generation (for internet connectivity; 2G, 3G, and 4G)
LSMS	Living Standards Measurement Study
OLS	ordinary least square
SWIFT	Survey of Well-being via Instant and Frequent Tracking

ACKNOWLEDGMENTS

This paper provides an overview of five applications that use both traditional and novel data sources to generate granular representations of the spatial distribution of poverty. Poverty maps are an increasingly useful tool for evaluation, harnessing new sources of data to improve assessment of the relevance and effectiveness of development interventions.

This paper was produced as part of the Methodological Paper Series sponsored by the Independent Evaluation Group's Methods Advisory Function. The authors are grateful for the feedback provided by the editor and staff of the Paper Series: Ariya Hagh, Michael Harrup, Maurya West Meiers, and Jos Vaessen. A special thanks to Amanda O'Brien, Rafaela Sarinho, Alexander Hery, and Luísa Ulhoa for their support in editing, production, and graphic design.

INTRODUCTION

Poverty maps provide a granular representation of the spatial distribution of poverty within a country or geographical area. They are increasingly useful tools for evaluative analysis, enabling a more refined assessment of the relevance and effectiveness of development interventions by leveraging new data sources that can serve as informative proxies for subnational poverty levels. Traditionally, poverty maps have relied on survey or census data to derive poverty estimates. The emergence of nontraditional data sources such as satellite imagery, call detail records (CDRs), smartphone metadata, and Wi-Fi connectivity opens new possibilities for achieving more timely and accurate poverty estimates. The high level of disaggregation provided by these sources allows for the visualization of poverty estimates at the household or village level, generating a more nuanced estimation of poverty than many conventional approaches do.[1] Furthermore, comparing poverty estimates at different points in time permits examination of temporal changes in poverty at very high levels of spatial disaggregation.

Poverty maps enable various stakeholders (government officials, program managers, the media, and others) to deepen their understanding of poverty and its determinants, allowing them to target development policies and programs in a more informed manner. They are also particularly helpful in the context of evaluation, allowing evaluators to examine the effects of interventions on the incidence and magnitude of poverty, including changes over time.

As noted above, poverty maps have traditionally relied on survey or census data to derive poverty estimates. This was only feasible where recent and accurate data existed. Data can be outdated or missing, and large-scale data collection efforts can be time-consuming and expensive. For instance, a 2015 World Bank study on the availability of traditional poverty data concluded that there was no meaningful way of monitoring poverty using conventional sources (such as census or survey data) for over a third of the world's low- or middle-income countries (Serajuddin et al. 2015). This study revealed that among the 155 countries for which the World Bank monitors poverty data, 29 had no poverty data points and 28 had only one poverty data point during 2002–11.

There is therefore an urgent need for cost-efficient rapid tools that can develop up-to-date poverty estimates. The emergence of nontraditional data sources, recent advances in data science and artificial intelligence, and increased computational capacity open new possibilities for using more indirect proxies of poverty to derive accurate and timely poverty maps. Specifically, poverty maps could play a critical role in assessing the relevance of targeting and evaluating program effectiveness by using poverty proxies for spatial counterfactual analysis.

This paper aims to provide guidance on methods to create poverty maps based on different data sources. The methods explored can be categorized in two main groups:

1. Methods based on either or both household surveys and census data on assets, consumption, expenditures, and access to services. This category includes methods (for example, small-area estimation and some of its variants) that use data sources such as large-scale surveys. Such surveys include the Living Standards Measurement Study (LSMS) and national household surveys. These approaches primarily apply multivariate statistical techniques to derive poverty estimates.

2. Methods based on more indirect proxies for poverty estimation, such as remote sensing data, CDRs, and the Global System for Mobile Communications or smartphone subscriptions. These approaches mostly rely on the application of various machine learning techniques, remote sensing, and geospatial analysis for poverty estimation.

Five data sources and corresponding methods are considered for their potential usefulness for poverty estimation and mapping. Each section includes a brief overview of the data requirements, methodology, applicability considerations, limitations, examples, and helpful references. The purpose of this paper is to show evaluators and other stakeholders how to leverage different poverty proxies to estimate poverty rates in the context of evaluation.[2] Through greater knowledge and use of nontraditional data sources, more temporally and spatially disaggregated estimations of poverty can be produced in a timely and cost-efficient manner. These estimates provide a critical complement to traditional statistics, filling some existing data gaps and improving the understanding of coverage or outreach of policy interventions targeted toward poor and vulnerable groups, and the poverty alleviation effects of these interventions.

Notes

1 For example, poverty maps can identify small pockets of poverty within wealthier areas, information that would otherwise be masked by national poverty averages. Although geolocated survey data offer similar benefits, surveys are more costly to implement and have lower coverage (in time and space) than the aforementioned proxies.

2 Poverty maps also have a broader relevance in the development community at large. They can be used to enhance key policy and programmatic aspects, such as targeting and coordinating strategies at local levels. Timely poverty estimates also aid real-time decision-making during crises, such as pandemics and natural disasters. Although the broader applicability of poverty maps contributes to their overall usefulness and relevance for development programs and policies, this publication focuses specifically on creating and using poverty maps in the context of evaluation.

1

SURVEY AND CENSUS DATA

Using Survey and Census Data

Data Availability

Limited by availability of existing data; requires both census and household survey data

Correlation with Poverty Attributes

Income or consumption per household

Cost

Large-scale surveys and census exercises can be a significant portion of an evaluation budget. This method is therefore only possible if existing data are available

Expertise Needed

Multivariate statistics

This section discusses the methodological implications of using traditional data sources such as survey and census data for generating poverty maps.

Definition

Household surveys are useful for collecting reliable information on the demographic and socioeconomic characteristics of a population of interest. Data are usually collected from a sample of randomly selected households, and statistical inference is used to estimate population parameters. The LSMS is the World Bank's flagship household survey program. It focuses on strengthening household survey systems in client countries and improving the quality of microdata to better inform development policies. The LSMS encompasses a multitopic survey that is customized to local contexts and uses best practices in survey methodology. The data collection and data management processes are coordinated by national statistical offices with the support of the World Bank LSMS team.

The Demographic and Health Surveys (DHS) are nationally representative household surveys that provide data for a wide range of indicators, such as population, health, and nutrition. Standard DHS surveys have large sample sizes (usually between 5,000 and 30,000 households) and are typically conducted every five years to allow comparisons over time. The DHS project is funded by the United States Agency for International Development. Household surveys, including the LSMS, DHS, and national surveys, can be used in combination with household-level census data to generate poverty maps. They provide detailed information about key variables contributing to poverty, with census data providing broader geographic coverage of household information.

Data Sources

For this method, survey and census data are combined to estimate poverty rates. A typical household survey collects data on a range of dimensions related to household and individual well-being, including but not limited to demographics, education, health, fertility, migration, labor, housing, savings and credit, income, and consumption. Surveys can be customized to a specific country context. Data can be collected via face-to-face methods, remotely, or both (significant innovations in remote data collection were prompted by the coronavirus [COVID-19] pandemic). Country-level implementation of household surveys relies on instruments such as household, community, price, and facility questionnaires.

Variables on consumption and income can be used to determine the economic status (and poverty levels) of households across a country. These variables can be complemented by other data sources, such as the socioeconomic modules included in household surveys, to derive a multidimensional poverty estimate. The concept of multidimensional poverty includes not only the insufficiency of economic resources but also the lack of basic rights (such as access to food, health, housing, social security, and education [CONEVAL 2017]).

Methods

The primary method for estimating geolocated poverty incidence rates using survey and census data is small-area estimation. Small-area estimation refers to a family of statistical imputation techniques that combines census and survey data to derive poverty estimations disaggregated to small geographical units (such as cities, towns, villages, or census divisions).

The creation of poverty maps for small areas is complex. Household survey data typically include income or consumption variables (which are required to derive poverty estimates) but are not representative at lower levels of disaggregation because of insufficient sample sizes. But the opposite is true for census data, which are representative at lower levels but do not usually contain sufficient information on consumption or income. To overcome these data limitations, the small-area estimation method aims to link income or consumption variables from survey data to other variables available in the census, so that they can be applied to census-level units of observation.

Using the small-area estimation method requires both national census data and household survey data (such as DHS, LSMS, or national income-expenditure surveys) for the country of interest. The survey and census data used to apply this method must share a common set of variables (associated with poverty levels). Furthermore, the data sources must be close in time (the accepted gap is three to five years). This is important to ensure that the characteristics of the populations have not significantly changed between the two surveys, since the method relies on the assumption that the estimated model of consumption or income from the survey is applicable to the census-level observations.

The small-area estimation method usually consists of two steps: (i) calibration of a statistical model based on survey data, and (ii) application to the comprehensive census data. In the first step, multiple linear regression analysis is used to estimate a model of household income or consumption based on survey data, restricting the explanatory variables in the model to the subset available in both the survey and the census. In the second step, the estimated model parameters are applied to the census data. The output from these two steps is an estimate of income or consumption for every household in the census. These estimates are then aggregated at the desired geographical level (for example, municipalities, districts, or villages).

Various techniques can be used to conduct small-area estimations. The most commonly employed methods include the ELL method (named after researchers Elbers, Lanjouw, and Lanjouw, [2003]), Empirical Bayes Prediction, Hierarchical Bayes, and Best Linear Unbiased Prediction. The World Bank has been a pioneer in the development of the small-area estimation method for creating poverty maps. Different variations of this technique have been applied to countries such as Albania, Bolivia, Bulgaria, Cambodia, China, Ecuador, Indonesia, Mexico, Morocco, Thailand, and Vietnam (Bedi, Coudouel, and Simler 2007).

Applicability Considerations

In the context of evaluation, the use of household survey data for poverty mapping is fairly limited by the substantial data requirements. Costs for this method vary. If household survey data are not available, data collection costs are likely to be prohibitive for individual evaluations. For example, the LSMS is estimated to cost approximately US$1.7 million per survey per country (commensurate with similar survey efforts by other organizations; see, for example, SDSN TReNDS [2018]). However, piggybacking on existing data from LSMS and national household surveys is quite feasible. LSMS data are publicly available via a database of completed surveys conducted in 38 countries from 1980 to the present.[1] Data access is characterized as (i) direct data access; (ii) public-use data files; or (iii) data available from an external repository. Data sets categorized as direct data access can be downloaded immediately. Data sets categorized as public-use data files can be accessed after registering with the World Bank Microdata Library and applying for access. This application requires a description of the intended use of the data. For data sets categorized as available from an external repository, the World Bank Microdata Library provides links to partners' websites. In addition, an access policy is outlined for each data set in the study description; this policy includes the name of a contact individual, access conditions, and citation requirements.

In some countries, national survey data are available and can be used for evaluative purposes. Similarly, the DHS program has been running for over 30 years and has produced over 320 surveys in 90 countries. DHS surveys can be directly accessed from the website of the United States Agency for International Development but viewing and downloading DHS microdata requires registration as a DHS data user.[2] DHS data set access is granted only for legitimate research purposes.

Such analysis can be completed using standard software packages for statistical analysis such as R, Stata, or SPSS. The World Bank has also released a publicly available Stata package, which can be used to conduct small-area estimation.[3] The visualization of the poverty estimates in maps might also require access to geospatial software such as QGIS (open source) or ArcGIS. The application of such estimation techniques requires knowledge of multivariate statistics and data manipulation and processing skills

Examples

Example 1: Poverty Maps Using Household Surveys in Brazil

Elbers, Lanjouw, and Leite (2008) validated the application of the ELL method based on a poverty map of Minas Gerais, a state in southeastern Brazil. This exercise was motivated by the fact that the 2000 Brazil census included additional income information as part of the census data collection procedure: (i) a single question on income of the household head was added to the traditional questionnaire collected from all households, and (ii) a more detailed questionnaire on income was fielded to 12.5 percent of households. These additional data provided an opportunity to compare the predicted poverty estimates produced by the ELL method with the actual household income figures obtained from the census data. For computational ease, the analysis focused on the state of Minas Gerais only.

After examining the estimates in nearly 1,000 municipalities, the researchers concluded that the poverty estimates produced by the ELL method were closely aligned to the actual observed poverty rates in those municipalities. Furthermore, the authors found that confidence intervals for those estimates were moderate.

Example 2: Poverty Maps Using National Household Surveys in Bolivia

As described by Arias and Robles in “The Geography of Monetary Poverty in Bolivia: The Lessons of Poverty Maps,” the World Bank, in conjunction with the Social and Economic Policy Analysis Unit and the National Institute of Statistics developed a poverty map of Bolivia using the ELL method (Arias and Robles 2007).

The main data sources for this exercise were the National Population and Housing Census of 2001 and household surveys that were conducted through the Program for the Improvement of Household Surveys and the Measurement of Living Conditions and carried out by the National Institute of Statistics in 1999, 2000, and 2001. Data from these sources were combined to obtain a larger sample that could be disaggregated according to the main administrative regions (departments) and areas in Bolivia. The method linked household consumption expenditure with variables measured in the household surveys and the census to impute the missing expenditure data.

Example 3: Multidimensional Poverty Maps Using National Household Surveys in Mexico

CONEVAL (2017) developed a poverty map of Mexico, disaggregated at the municipality level, using the small-area estimation method. A novel element in this case was the use of a multidimensional poverty measure. This estimation was based on a combination of data on economic well-being (income) and social rights (such as access to food, health, education, social security, or dignified housing). Income data were obtained from the Intercensal Survey, and information for the multidimensional measurement of poverty was extracted from the Socioeconomic Conditions Module of the National Survey of Household Income and Expenditure. The study, within the multidimensional approach of measuring poverty, also produced granular estimates on food insecurity and lack of access to social security.

Example 4: Poverty Maps Using Living Standards Measurement Study Survey Data in Nicaragua

Sobrado and Rocha used data from a 2005 LSMS in Nicaragua to create a poverty map of the country (World Bank 2008). The 2005 Census of Nicaragua and the 2005 LSMS were used as data sources, and the authors included only data from questions that were either the same or similar in both the sources of information. The authors then compared the 2005 poverty map with one created in 1995 to identify changes in the distribution of poverty. Through this exercise, the authors found a decrease in the incidence of poverty and in the poverty gap index for almost all regions of Nicaragua. The authors recommended that policy makers in Nicaragua use the 2005 poverty map as a targeting tool (in addition to other tools) because the map showed both the distribution of poverty and how the distribution had shifted since 1995.

Example 5: Poverty Maps Using Household Survey Data in Ecuador, Madagascar, and South Africa

Gabriel Demombynes et al. (2002) created poverty maps for Ecuador, Madagascar, and South Africa by combining survey and census data. Although the three countries differ significantly in geography, stage of development, and so on, the researchers found that the poverty estimates generated from this exercise were plausible (that is, the estimates generated from the census data matched well with estimates calculated directly from the survey data) and sufficiently precise (that is, at a lower level of disaggregation than was possible through the household survey data alone).

For the Ecuador map, the researchers used data from a 1990 census conducted by the National Statistical Institute of Ecuador and a 1994 household survey based on the LSMS. For the Madagascar map, they used data from a 1993 census conducted by the National Institute of Statistics, a 1993–94 household survey conducted by the Ongoing Household Survey, and data on spatial and environmental outcomes at the *fivondrona* (communes) level. For the South Africa map, they used data from the 1995 October Household Survey, an Income and Expenditure Survey conducted at approximately the same time, and a 1996 population census.

The researchers examined the extent to which the poverty estimates from the census matched the poverty estimates from the household surveys (at the level represented in the survey). The poverty estimates for Ecuador were relatively close to the results of the census, with all but two regions within 95 percent confidence intervals. The estimates for Madagascar were also relatively close, except for one or two strata that were not well explained by the first-stage regression (for example, the adjusted R^2 for the rural Antsiranana stratum was 0.292, the lowest of any of the models explored). The estimates for South Africa were also deemed satisfactorily close.

Based on these results, the researchers found that across the three countries, the poverty estimates at the census level aligned overall with the household survey estimates, with the standard errors at the stratum level being consistently lower than those derived solely from the household survey data.

The researchers also explored how far the census-based poverty estimates can be disaggregated, using the household survey sampling errors to benchmark acceptable levels of precision. For all three countries, they could generate poverty estimates at the third administrative level with similar levels of precision to the household survey data (at the representative stratum level of the survey). This exercise demonstrated how this method can provide useful information about the incidence of poverty levels across regions.

Notes

1 See the Living Standards Measurement Study database at https://microdata.worldbank.org/index.php/catalog/lsms.

2 See the Demographic and Health Surveys data sets: https://dhsprogram.com/data/available-datasets.cfm.

3 The Stata package is available at https://github.com/pcorralrodas/SAE-Stata-Package.

2

SURVEY OF WELL-BEING VIA INSTANT AND FREQUENT TRACKING DATA

Using Survey of Well-being via Instant and Frequent Tracking Data

Data Availability

Limited

Correlation with Poverty Attributes

Income, expenditures, or both per household

Cost

Relatively less expensive than traditional data collection exercises

Expertise Needed

Analytic skills (post-survey completion)

This section discusses the methodological implications of using Survey of Well-being via Instant and Frequent Tracking (SWIFT) surveys as a data source for estimating poverty. Building on insights from the previous section, the applicability of SWIFT data for generating poverty maps is discussed in this section.

Definition

The World Bank Group's SWIFT is a rapid assessment tool that estimates household income or expenditures to measure household poverty. SWIFT does not collect direct income or consumption data; instead, it collects poverty correlates such as household size, ownership of assets, or education levels, and then converts them to poverty statistics using estimation models. These poverty correlates are collected through a customized questionnaire consisting of 10–15 questions, which typically takes approximately five minutes to administer (hence, "swift").

SWIFT survey data are collected through computer-assisted personal interviews, enabling data to be collected using tablets or smartphones and uploaded to a data cloud, making them accessible in real time. Analysts can then download the data and convert them into poverty and distributional statistics.

Data Sources

To derive the survey questionnaires, the SWIFT team develops a model based on household survey data. To ensure optimal results, there should be at least two rounds of highly comparable household survey data (such as the LSMS). These data sets should be no more than five years apart, and at least one of them should be no more than three years old.

Given that these data requirements are not always satisfied, some of the requirements can be relaxed in some cases. First, if the latest survey was carried out within the previous two years or is in progress, the SWIFT team can produce models using only the latest survey data, assuming that consumption patterns did not change significantly since the data were collected. Second, if the latest survey is older than five years, but there is a survey in progress, the SWIFT team can create a questionnaire to include variables that are likely to be in models that will be developed from the new survey.

Based on the model trained on the available survey data, the SWIFT team creates a questionnaire and collects data on poverty correlates, such as household size, ownership of assets, education levels, employment status, and so on, to estimate household income or expenditures. There are several versions of the SWIFT survey, including the classic SWIFT 1.0, which is described here; the SWIFT Plus, which can be used in locations experiencing economic shocks; the SWIFT-COVID19, which is specific to the COVID-19 situation; and the SWIFT 2.0, which can be used when there are no reliable or recent data for the location of interest. SWIFT surveys can also be included in any household survey to incorporate a poverty lens.

Since the SWIFT program launched in 2014, more than 100 SWIFT surveys have been or are being conducted in over 50 countries. However, unlike LSMS data, which are available online in the World Bank Microdata Library, SWIFT survey data are not accessible for public use at this time.

Methods

SWIFT relies on the availability of both consumption and nonconsumption data collected through a national household survey. The SWIFT survey model is derived by imputing consumption data based on the consumption data available in the household survey and collecting specific nonconsumption data through a custom questionnaire.

To derive a stable model, a cross-validation exercise is first conducted. The relevant household survey data are split randomly into 10 subsamples (or folds). Nine of these folds are used for training the model, and the remaining fold is used for testing. A model is estimated from nine folds by running a stepwise ordinary least square (OLS) regression, and the performance of the model is evaluated in the remaining fold. Because the remaining fold was not included when the model was trained, no performance indicators in the remaining fold are subject to the problem of overfitting.

This cross-validation exercise is intended to determine the optimal threshold of the p-value for the OLS regression equation. After the selection of the optimal p-value, OLS is applied to the full sample of data to estimate a model. Once the model for estimating household consumption is complete, the next step is to develop the questionnaire to collect nonconsumption data. It is critical that at this stage researchers consider the survey sampling design, as this highly influences the sampling precision of the survey. Finally, poverty rates are estimated using the multiple imputation method. The accuracy of SWIFT estimations relies on strong underlying models, which in turn rely on the quality and accuracy of the underlying large data sets used when designing the SWIFT survey.

Applicability Considerations

The use and accessibility of SWIFT surveys for poverty estimates is somewhat limited, given the data requirements. For a SWIFT survey, national-level data must be available to identify the poverty correlates the survey will measure. Recent, high-quality data may not be available for all countries.

Further, the SWIFT method is designed to provide poverty estimates for a specific geographic area or target group, such as national poverty estimates or poverty estimates for participants of a particular program. Since these poverty estimates cannot be disaggregated below the target level of the model, SWIFT survey estimates cannot (at this time) be used to develop poverty maps. However, given the method's light touch and relatively low cost, SWIFT poverty estimates could be useful for assessing poverty in the context of evaluations. Large-scale surveys and censuses are elaborate exercises that require significant resources; the SWIFT method offers an efficient way to obtain poverty estimates in certain contexts.

However, because SWIFT aims to produce models specific to areas and populations in which projects are being implemented, the method is well suited to measure the impact of specific interventions on the income levels of target beneficiaries. Furthermore, given its relatively low cost, SWIFT could be implemented at times to better understand changes in poverty rates. Both of these features make the data well suited to the generation of granular poverty maps in various geographic contexts.

The Bank Group's SWIFT team provides support to teams interested in using SWIFT surveys in their studies.[1] This support may enhance opportunities to use this method. Given their limited scope, SWIFT surveys cost less than US$100,000 per country to implement and are substantially cheaper than longer survey exercises. SWIFT survey results can be interpreted by applying standard analytical techniques.

Example: Estimating Poverty Rates in Uganda Using SWIFT

Heitmann and Buri (2019) used results from a SWIFT survey in conjunction with CDR data to estimate poverty rates in Uganda as part of a larger study on using satellite imagery to estimate poverty at neighborhood levels. The survey focused on Northern Uganda, covering 9,037 households in the Karamoja, Mid-North, West Nile, and Adjumani administrative areas. The location of each household surveyed was geolocated using GPS. The researchers aimed to identify correspondence between CDR and household survey data by matching phone numbers across the two data sources to explore additional methods to predict poverty through CDR data.

The survey responses did not overlap with the CDR data well because of the randomized design of the survey. Of the 9,037 households surveyed, only 222 were also present in the CDR data. The researchers therefore did not have sufficient observations to draw meaningful prediction models from this exercise. They instead used household information aggregated by cell-tower catchment area to estimate poverty rates. Even so, these models had an extremely low explanatory power, with an R^2 of 0.01.

The researchers concluded that in such cases, research teams should consider conducting a light-touch baseline survey to understand the general market share of cell phone usage and then design a survey that over-samples in a statically controllable manner to achieve sufficient overlap between survey and CDR data sets.

Notes

1 See the Survey of Well-being via Instant and Frequent Tracking Team web page at https://worldbankgroup.sharepoint.com/sites/Poverty/Pages/SWIFT-06202018-141205.aspx (user ID and passcode required).

3

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS, SMARTPHONE, AND WI-FI CONNECTIVITY DATA

Using Global System for Mobile Communications, Smartphone, and Wi-Fi Connectivity Indicators

Data Availability

Several data sources are publicly available (such as Facebook's advertising data). Other data sources are proprietary

Correlation with Poverty Attributes

Household wealth

Cost

Low to high, depending on the data and techniques used

Expertise Needed

Statistical analysis or machine learning (depending on the data and method to be used)

This section discusses the methodological implications of using Global System for Mobile Communications, smartphone, and Wi-Fi connectivity indicators as a main data source for generating poverty maps.

Definition

The set of data sources examined here comprises indicators on connectivity (for example, internet speed and network coverage) and technology use (for example, the prevalence of high-end smartphones or certain mobile phone operating systems); this section explores their usefulness in creating poverty maps in different locations. The use of these indicators assumes that connectivity data provide strong predictive information on income levels and can be used to predict the socioeconomic situation in that location. For example, an area with fewer smartphones and lower Wi-Fi connectivity would suggest lower wealth levels relative to an area with a higher prevalence of high-end smartphone use and fourth generation (4G) internet connectivity.

Data Sources

The variables that are particularly useful for this type of analysis include network access (2G, 3G, or 4G networks, Wi-Fi connectivity, and so on), the mobile operating systems used (Android, iOS, Windows), and the brands of smartphone used (Apple, Samsung, Motorola, and so on). Some of this information is publicly available but at different levels of geographic and temporal disaggregation, depending on the country of interest.[1] Additionally, technology companies such as Facebook tend to possess more granular data, which can be extremely useful for this type of analysis. Some of these data are publicly available for research purposes (for example, network coverage maps from Facebook), but data that might identify users remain proprietary and confidential. Such proprietary data may be accessible, however, through an agreement with the owner of the data and a clear statement on its intended use.

Methods

Methods used to generate poverty maps vary greatly depending on the type of data used.

The simplest method, typically applied in the case of models relying only on connectivity data, is ridge regression. Ridge regression is an extension of OLS linear modeling, which is particularly useful for multivariate regression problems where the explanatory variables are suspected to be highly correlated (exhibiting multicollinearity; Hastie, Tibshirani, and Friedman 2009). This method aims to avoid overfitting a model when there are many predictors.

Other approaches apply more complex models, such as convolutional neural networks (CNNs) and transfer learning, to a combination of connectivity data and satellite imagery to derive micro-estimates of wealth (Chi et al. 2021). CNNs are deep-learning algorithms that assign weights to various features of an image. Transfer learning is a machine-learning method in which a model developed and trained for a task is reused as the starting point for a model on a different task. Transfer learning approaches have the advantage of reducing the time and computing resources needed to train a new model.

Applicability Considerations

There is some potential for the use of connectivity data in the context of evaluation. Publicly available data (such as Facebook's advertising data) can be used to create poverty maps through relatively simple techniques (such as regression analysis) and that do not require specialized software or additional computing resources. Research on the use of connectivity data, however, is still incipient and fairly limited, and therefore the limitations of these models are not yet fully understood. Preliminary research suggests that this modeling approach performs better in urban areas and is highly dependent on penetration rates.[2]

Another method uses Facebook's poverty estimates—derived using the method described under Example 2 below—which are publicly available in a tabular format for 135 countries at a very granular spatial resolution. But these estimates are only available for 2021. The method could be replicated for other years; however, this would require access to Facebook's proprietary data, and substantial expertise in machine learning and appropriate computing resources to run image-based models.[3]

Examples

Example 1: Poverty Maps Using Facebook's Publicly Available Advertising Data

Fatehkia, Coles et al. (2020) developed a model using Facebook advertising data to estimate household wealth in India and the Philippines. The authors used the Facebook Marketing Application Programming Interface to query the number of Facebook users matching certain criteria to obtain insights into the spatial distribution of users by device type (for example, iOS, Windows, Android), access to connectivity (for example, 2G, 3G, 4G, Wi-Fi), and use of high-end devices (the latest releases of Apple iPhones and Samsung Galaxy phones).

The Facebook Marketing Application Programming Interface only provides an estimate of the monthly active users matching the specified criteria at different levels of geolocation (the most disaggregated level is the city level). In addition to Facebook penetration data, the fraction of users in each location with access to these different features was computed using a ridge regression approach, with the assumption that these insights provide signals on the underlying distribution of poverty. For comparison, the authors also collected nighttime and daytime satellite data, which were processed using CNNs to extract relevant features. The estimates obtained by this model were validated against a wealth index, which was constructed using principal component analysis based on data from the DHS.

In the case of the Philippines, the authors concluded that a model featuring Facebook data performed roughly similarly to a model based only on satellite data (with slightly better performance in urban areas). This conclusion is important because models based on Facebook's public data are considerably simpler to implement than models using satellite data. In India, however, where Facebook penetration is lower, satellite data performed better.

Example 2: Poverty Maps Using Facebook's Proprietary Connectivity Data

Chi et al. (2021) developed the first micro-estimates of wealth that cover the populated surface of all 135 low- and middle-income countries at 2.4 kilometer resolution. The estimates were generated by applying machine-learning algorithms to vast and heterogeneous data from satellites, mobile phone networks, topographic maps, and aggregated and anonymized connectivity data from Facebook. Data sources included road density, land cover, elevation, slope, precipitation, population, nighttime lights, satellite imagery, and specific features derived from Facebook's proprietary data.

The authors found the resulting estimates of wealth to be quite accurate. Depending on the method used to evaluate performance, the model explained 56–70 percent of the actual variation in household-level wealth in low- and middle-income countries. In particular, information on mobile connectivity was highly predictive of subregional wealth, with 5 of the 10 most important features in the model related to connectivity.

This approach was further validated on "ground truth" measurements of wealth from DHS and local or regional surveys, where available. This validation was conducted using spatial markers in the survey data to link each village to the various data sources used in the study.[4] Considering the impact of the COVID-19 pandemic on the launch of new development interventions and the importance of detailed wealth estimates for better targeting, Facebook has provided free access to these estimates for public use.[5]

Notes

1 See the International Telecommunication Union's Statistics page at https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx; see Facebook's Marketing Application Programming Interface web page at https://developers.facebook.com/docs/marketing-apis/; see Meta's Data for Good webpage at https://dataforgood.facebook.com/dfg/tools.

2 The fraction of users of each product (such as smartphones or Wi-Fi) varies greatly across countries and tends to be higher in urban areas. Penetration rates are typically computed as the ratio of users to the estimated population of the area of interest. If the penetration rate is low, the data might not be representative of the entire population. More important, the association with poverty levels is significantly weaker in such cases.

3 The following geographically disaggregated data from Facebook require a license or are restricted: number of cells towers, number of Wi-Fi access points, number of mobile devices, number of Android devices, and number of iOS devices.

4 Indeed, based on the strength of these results, the government of Nigeria is using these estimates as the basis for social protection programs. Likewise, the government of Togo is using these estimates to target mobile money transfers to hundreds of thousands of the country's poorest mobile subscribers.

5 The estimates can be found and downloaded from the Humanitarian Data Exchange website: https://data.humdata.org/dataset/relative-wealth-index.

4

CALL DETAIL RECORD DATA

Using Call Detail Record Data

Data Availability

Proprietary data. For some countries (for example, Senegal) there might be publicly available anonymized data

Correlation with Poverty Attributes

Asset-based wealth and consumption-based wealth per household

Cost

Low

Expertise Needed

Data processing, geospatial analysis, machine learning

This section discusses the methodological implications of using CDRs as a data source for generating poverty maps.

Definition

CDRs obtained from mobile network operators provide highly granular real-time data that can be used to assess socioeconomic behavior, including consumption, mobility, and social patterns. CDRs have been successfully used to predict poverty in some countries with (i) models that attempt to predict welfare based on call activity only, and (ii) combined models that use telephone data and remote sensing covariates.

Data Sources

CDR data include encrypted user ID, location area code, cell ID, time stamp, and event ID. Location area code and cell ID jointly determine the geographical location (coordinates) of the cell tower. The event ID records the type of the transaction: call in, call out, text messaging, and web browsing. In addition, researchers can typically infer the type of phone in use (including the brand), the vendor, the model, and the system. This information can be used as a proxy for the user's disposable income. For the purpose of poverty mapping, CDRs can be used in conjunction with other poverty estimates or satellite imagery (daytime, night-time, or both).

Methods

Raw CDR data are typically noisy and require preprocessing before they can be analyzed. Once cleaned, CDRs can be used to create geographical segments by constructing Voronoi polygons or grids at the desired resolution level.[1] Estimating poverty rates from CDR features is an example of a supervised machine-learning problem, one in which input data are used to predict known outputs—in this case, CDR features and poverty rates. Once the model is built, it can be applied to new input data for which the corresponding output is unknown. In this case, the unknown output would be either different geographies or different points in time. Supervised machine-learning problems are either classification based, in which the output variable is one of a discrete set of classes (for example, poor or not poor), or regression based, in which the output variable is a continuous real number expressed as a decimal, ratio, or percentage.

Applicability Considerations

The use of CDRs for poverty mapping in the context of evaluation is fairly limited. Although these data can be obtained at zero or low cost, CDRs are largely proprietary and can be accessed only through an agreement with mobile network operators. Furthermore, to ensure the representativeness of data, agreements are needed with multiple mobile network operators with coverage in the area of interest. Some countries, however, have made anonymized CDR data sets publicly available (for example, Senegal); in these instances, there may be greater opportunities for using CDRs for poverty mapping.

Some specific expertise is needed to derive poverty maps from CDRs, including experience in advanced data cleaning and manipulation and in geospatial analysis. No advanced computing resources are likely to be needed. All parts of the analysis can be completed with a combination of Excel, Python, or R (open source), and geospatial software such as QGIS (open source).

Examples

Example 1: Guatemala Poverty Map

The World Bank conducted a CDR analysis focused on five administrative departments in the southwest region of Guatemala, using mobile phone data to predict observed poverty rates and generate poverty maps. The study used encrypted CDR data for August 2013, aggregated at the municipality level. To test the validity of the CDR analysis, the findings were compared with World Bank poverty estimates based on Guatemala's National Living Conditions Survey for 2006 and 2011 and the 2002 Population and Housing Census.

The findings from the study indicate that CDR-based research methods may replicate poverty estimates obtained from traditional forms of data collection at a fraction of the cost. Although the poverty estimates produced by CDR analysis did not perfectly match those generated by surveys and censuses, the results show that more comprehensive data could greatly enhance their predictive power. CDR analysis has especially promising applications in low-income countries where limited fiscal and budgetary resources complicate the task of survey data collection.

Example 2: Rwanda Poverty Map

Blumenstock, Cadamuro, and On (2015) constructed a poverty map of Rwanda using an anonymized database containing records of billions of interactions on Rwanda's largest mobile phone network. These data were complemented with follow-up phone surveys of a geographically stratified random sample of 856 individual subscribers, which included questions regarding asset ownership, housing characteristics, and several other basic welfare indicators. Given the geographic information contained in the CDR data, the authors were able to map each data point to small divisions created using Voronoi polygons. The data were analyzed using a supervised-learning algorithm to generate wealth predictions at a very fine degree of spatial granularity. Out-of-sample predictions were generated for the characteristics of the remaining 1.5 million Rwandan mobile phone users who did not participate in the survey. By comparing the model's results with other sources of data, the study showed that CDR data were predictive of individual-level asset-based wealth in Rwanda.

In 2018, Blumenstock expanded this analysis by applying a simplified version of the previous model to Afghanistan. The objective was to demonstrate the accuracy of a model that could be replicated and generalized to different countries. The study relied on several rounds of interviews with 1,234 Afghan citizens. As in the case of Rwanda, each respondent's survey responses were matched to their corresponding CDRs. The simplified model was applied to both Rwanda and Afghanistan with results similar to those observed in the original model. The author further investigated whether a model trained with data from one country (in this case, Rwanda) could be used to predict the wealth of a different country (Afghanistan). However, the results were only slightly better than those that would be obtained by random guesses, indicating the need to retrain the model with country-specific data.

Notes

1 Voronoi polygons partition a plane into regions proximate to items or objects within a defined set.

5

DAYTIME AND NIGHTTIME REMOTE SENSING IMAGERY DATA

Using Remote Sensing Imagery Data

Data Availability

Low- or medium-resolution satellite data are publicly available. The highest-resolution data must be purchased from specialized vendors

Correlation with Poverty Attributes

Household expenditure or household wealth

Cost

Additional computing resources (for example, a graphics processing unit) might be needed

Expertise Needed

Remote sensing, machine learning

This section discusses the methodological implications of using remote sensing imagery as a data source for generating poverty maps.

Definition

Remote sensing is the process of observing, monitoring, and detecting the physical characteristics of an area from a distance, using sensors located in aerial platforms (such as satellites, aircrafts, and drones). For generating poverty maps, the most commonly used type of remote sensing data are satellite data, including both daytime and nighttime data. Both types of satellite images have a raster data format, and therefore consist of a numerical gridded representation, where each cell of the grid is associated with a specific geographical location.

A considerable advantage of this type of data is its large temporal and geographic reach. Daily satellite images are available from reliable sources for the whole world from approximately the 1980s to the present. Furthermore, satellite data can be disaggregated into small areas—such as cities or villages—depending on the image resolution. This results in consistent and comparable data suitable for conducting meaningful comparisons across time and geographies.

Data Sources

Low- or medium-resolution daylight satellite imagery data are available from government programs such as Landsat (National Aeronautics and Space Administration / United States Geological Survey) and Sentinel (European Space Agency) from 1972 onward. Although these data are collected daily, images with high cloud coverage are typically discarded. Higher-resolution satellite data are available for purchase from specialized vendors, but publicly available satellite data are typically sufficient to generate poverty maps.

Most common sources of nighttime satellite data include the Defense Meteorological Satellite Program, available for 1993–2017, and the Visible Infrared Imaging Radiometer Suite, available for 2012 to the present. These data sets provide an estimate of radiance, defined as a stable measure of brightness as seen from space, filtered to remove extraneous features such as biomass burning and aurora. Nighttime light data are available in different series, depending on the sensor that was used to capture the image. Although data from different series can be combined to perform time series analysis, this requires prior calibration of the data.

Methods

There are essentially three types of models that use machine-learning techniques to derive poverty proxies using remote sensing (for example, satellite image) data. The simplest isa feature-based prediction model that uses quantifiable geospatial features (such as the number of buildings in a region, the length of roads, or points of interest). This type of model typically relies on random forest regression to generate a prediction.[1]

The second type of model is an image-based prediction model that extracts geospatial features directly from remote sensing imagery. These models rely on deep-learning algorithms such as CNNs, which are particularly suited to working with images.[2] The third type of model combines the two approaches described above: geospatial features are combined with satellite imagery to estimate a poverty-related proxy.

Applicability Considerations

There is strong potential for the use of satellite imagery to derive poverty maps in the context of evaluation. Both daytime and nighttime satellite imagery data are freely available from reliable sources at a low or medium resolution, which is sufficient to create poverty maps. However, a higher level of expertise is needed to develop poverty maps using satellite imagery. This includes experience in manipulating Earth observation data, machine learning (specifically deep learning and CNNs), and programming.

Additional computing resources (such as access to graphics processing unit clusters) might be needed to use this approach, especially when using CNNs. Access to graphics processing unit clusters must be purchased, with plans typically priced per hour of use.

All parts of the analysis can be completed with a combination of Excel, Python (open source), and geospatial software such as QGIS (open source). Different research centers have also made publicly available the Python code used to create their poverty maps, which would be an excellent starting point for developing new maps.[3]

Examples

Example 1: Poverty Maps Using Geospatial Features

Tingzon et al. (2019) developed a poverty map of the Philippines using geospatial features extracted from OpenStreetMap (a crowdsourced online mapping platform). The features used in the model included roads, buildings, and points of interest (such as parks, schools, hospitals, and cinemas). These features were extracted within a 5 kilometer radius for rural areas and a 2 kilometer radius for urban areas.

A random forest regression model was trained on these features, both separately and jointly, to predict socioeconomic well-being. The authors found that using roads, buildings, or points of interest alone already explained 49–55 percent of the variance, with roads being the best predictor (R^2 = 0.55). Training a model on all three types of OpenStreetMap features resulted in a slightly higher R^2 value (0.59). Furthermore, the authors tested the performance of the model by adding nighttime lights data. This resulted in an increased R^2 (0.63) for wealth prediction.

Example 2: Poverty Maps Using Satellite Imagery

Yeh et al. (2020) trained deep-learning models to predict survey-based estimates of asset wealth across approximately 20,000 African villages using temporally and spatially matched multispectral daytime satellite imagery and nighttime lights data (30 meter per pixel Landsat and < 1 kilometer per pixel nighttime lights imagery). The authors trained a CNN—ResNet 18 architecture—to predict village- and year-specific measures of wealth. The main objective was for the machine-learning model to identify those features present in daytime and nighttime satellite imagery that are predictive of asset wealth.

The study found that a deep-learning model trained on this type of imagery data can explain approximately 70 percent of the spatial variation in asset wealth across Africa and up to 50 percent of the changes in wealth over time when aggregating the village-level data to the district level. Notably, CNNs trained only on nighttime lights or only on multispectral daytime imagery performed similarly to each other and almost as well as the combined model, suggesting that these two inputs contain similar information, at least for predicting spatial variation in wealth in Africa. This model also outperformed simpler models based only on geospatial features.[4]

But deep-learning models tend to be less interpretable than other machine-learning approaches. Deep neural network constructs typically combine multiple hidden layers and neurons, resulting in millions of features. From the perspective of human interpretability, it is very difficult to track the complex interactions that occur among the features underpinning the model's output. Although there have been advances in the academic literature toward more interpretable models (through fields such as interpretable artificial intelligence), deep-learning models continue to be considered a "black box."

Example 3: Poverty Maps Using a Combination of Geospatial Features and Satellite Imagery

Puttanapong et al. (2020) developed a model to predict the spatial distribution of poverty in Thailand based on the integration of multiple data sources, including geospatial features and features extracted from satellite imagery. Specifically, the study used the following sources: land surface temperature, normalized difference vegetation index, intensity of lights, geocoded data on built-up and non–built-up areas, geocoded human settlement data, land cover maps, and crowdsourced data from OpenStreetMap (road count, road length, points of interest, and built-up areas).

The study applied several computational techniques to examine the relationship between geospatial features and the proportion of people living below the poverty line using conventional methods of estimating poverty levels. The methods applied in the study included generalized least squares, neural networks, random forest, and support vector regression. Results suggested that intensity of night lights and other variables that approximate population density are highly associated with the proportion of an area's population who are living in poverty. The random forest technique yielded the highest level of prediction accuracy among the methods considered in this study, perhaps because of its ability to fit complex association structures even with small- and medium-size data sets.

Notes

1 Random forest is a supervised machine-learning algorithm that combines predictions from multiple decision trees using an ensemble approach.

2 Deep learning encompasses supervised machine-learning algorithms, which mimic the structure of the human brain by using multilayered neural network architectures.

3 See, for example, the data available at the following webpages: https://github.com/sustain-lab-group/africa_poverty, https://github.com/nealjean/predicting-poverty, https://github.com/jmather625/predicting-poverty-replication.

4 A similar approach has been developed by Stanford University's Sustainability and Artificial Intelligence Lab, which has been tested across several African countries (Jean et al. 2016). Other studies have also built on this approach, such as Babenko et al. (2017), Tingzon et al. (2019) and Heitmann and Buri (2019).

CONCLUSIONS AND SUGGESTIONS

This paper explores both traditional and novel methods that can be used to derive geographically disaggregated poverty estimates and poverty maps. Granular and up-to-date poverty data are critical for responding to questions regarding the relevance and effectiveness of policy interventions in the context of evaluation. The approaches outlined above provide a brief overview of some of the data and methodological alternatives that can be used to generate cost-efficient poverty maps.

Many of these methodological options can be derived using publicly available data and existing resources at a relatively low cost. This is an important consideration given that traditional in-the-field poverty data collection for a country or an area of interest is expensive and time-consuming.

Finally, this paper also introduces some readily available resources in the form of data sets that can be directly used to plot detailed poverty maps (such as the wealth index jointly developed by Facebook and University of California, Berkeley) or code repositories that provide all the implementation details that are essential to replicate some of the methods described herein.

BIBLIOGRAPHY

Aiken, Emily L., Guadalupe Bedoya, Aidan Coville, and Joshua E. Blumenstock. 2020. "Targeting Development Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan." Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies. 2020.

Alkire, Sabina. 2014. "Towards Frequent and Accurate Poverty Data." *OPHI Research in Progress* 43b, Oxford University.

Amin, Samia, Jishnu Das, and Markus Goldstein, eds. 2008. *Are You Being Served? New Tools for Measuring Service Delivery*. Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/6921/424820PUB0ISBN1LIC0disclosed0Feb131.pdf?sequence=1&isAllowed=y.

Arias, Omar, and Marcos Robles. 2007. "The Geography of Monetary Poverty in Bolivia: The Lessons of Poverty Maps." In *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, edited by Bedi, Tara, Aline Coudouel, and Kenneth Simler, 67–89. Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/6800/414470PAPER0Pr101Official0Use0Only1.pdf?sequence=1&isAllowed=y.

Asian Development Bank. 2001. *Handbook for Integrating Poverty Impact Assessments in the Economic Analysis of Projects*. Manila: Asian Development Bank https://www.adb.org/sites/default/files/institutional-document/31336/integrating-poverty-impact-assessment.pdf.

Ayush, Kumar, Burak Uzkent, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. 2020. "Efficient Poverty Mapping Using Deep Reinforcement Learning." arXiv preprint arXiv:2006.04224. https://arxiv.org/abs/2006.04224.

Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz. 2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico." *arXiv preprint arXiv:1711.06323*. https://arxiv.org/pdf/1711.06323.pdf.

Bedi, Tara, Aline Coudouel, and Kenneth Simler, eds. 2007. *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/6800/414470PAPER0Pr101Official0Use0Only1.pdf?sequence=1&isAllowed=y.

Blumenstock, Joshua E. 2018. "Estimating Economic Characteristics with Phone Data." *AEA Papers and Proceedings* 108: 72–76. https://par.nsf.gov/servlets/purl/10062319.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–1076. https://doi.org/10.1126/science.aac4420.

Chambers, Ray, and Nikos Tzavidis. 2006. "M-Quantile Models for Small Area Estimation." *Biometrika* 93 (2): 255–268. https://www.jstor.org/stable/20441279?refreqid=excelsior%3A1efdc510f7c748519df160fb2e365063.

Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. 2021. "Micro-Estimates of Wealth for all Low- and Middle-Income Countries." *arXiv preprint arXiv:2104.07761*. https://arxiv.org/abs/2104.07761.

CONEVAL (Consejo Nacional de la Política de Desarrollo Social). 2017. *Metodología para la Medición de la Pobreza en los Municipios de México, 2015*. Mexico City: CONEVAL. https://www.coneval.org.mx/Medicion/Documents/Pobreza_municipal/Metodologia_municipal_2015.pdf.

Corral, Paul, Kristen Himelein, Kevin McGee, and Isabel Molina. 2021. "A Map of the Poor or a Poor Map?" Policy Research Working Paper 9620, World Bank, Washington, DC. (). https://openknowledge.worldbank.org/bitstream/handle/10986/35442/A-Map-of-the-Poor-or-a-Poor-Map.pdf?sequence=1&isAllowed=y.

DANE (Departamento Administrativo Nacional de Estadística). 2020. "El DANE presenta por primera vez la medida de la pobreza multidimensional de las cabeceras municipales a nivel de manzana." DANE, Bogota. https://www.dane.gov.co/index.php/en/actualidad-dane/5186-el-dane-presenta-por-primera-vez-la-medida-de-la-pobreza-multidimensional-de-las-cabeceras-municipales-a-nivel-de-manzana.

Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. 2017. "Data Gaps, Data Incomparability, and Data Imputation: A Review of Poverty Measurement Methods for Data-Scarce Environments." Policy Research Working Paper 8282, World Bank, Washington, DC. https://documents1.worldbank.org/curated/en/551171513690220305/pdf/Data-gaps-data-incomparability-and-data-imputation-a-review-of-poverty-measurement-methods-for-data-scarce-environments.pdf.

Decuyper, Adeline, Alex Rutherford, Amit Wadhwa, Jean-Martin Bauer, Gautier Krings, Thoralf Gutierrez, Vincent D. Blondel, and Miguel A. Luengo-Oroz. 2014. "Estimating Food Consumption and Poverty Indices with Mobile Phone Data." *arXiv preprint arXiv:1412.2595*. https://arxiv.org/abs/1412.2595.

Demombynes, Gabriel, Chris Elbers, Jenny Lanjouw, Peter Lanjouw, Johan Mistiaen, and Berk Özler. 2002. "Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries." Discussion Paper 2002/39, World Institute for Development Economics, Helsinki. https://www.econstor.eu/handle/10419/52910.

Egedy, Tamás, and Bence Ságvári. 2021. "Urban Geographical Patterns of the Relationship between Mobile Communication, Social Networks and Economic Development—The Case of Hungary." *Hungarian Geographical Bulletin* 70 (2): 129–148 ().DOI:10.15201/hungeobull.70.2.3.

Elbers, Chris, Tomoki Fujii, Peter Lanjouw, Berk Özler, and Wesley Yin. 2007. "Poverty Alleviation through Geographic Targeting: How Much Does Disaggregation Help?" *Journal of Development Economics* 83 (1): 198–213. https://doi.org/10.1016/j.jdeveco.2006.02.001.

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–364. https://doi.org/10.1111/1468-0262.00399.

Elbers, Chris, Peter Lanjouw, and Phillippe George Leite. 2008. "Brazil within Brazil: Testing the Poverty Map Methodology in Minas Gerais." Policy Research Working Paper 4513, World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/6575/wps4513.pdf?sequence=1&isAllowed=y.

Elvidge, Christopher D., Paul C. Sutton, Tilottama Ghosh, Benjamin T. Tuttle, Kimberly E. Baugh, Budhendra Bhaduri, and Edward Bright. 2009. "A Global Poverty Map Derived from Satellite Data." *Computers & Geosciences* 35 (8): 1652–1660. https://doi.org/10.1016/j.cageo.2009.01.009.

Engstrom, Ryan, Jonathan Samuel Hersh, and David Locke Newhouse. 2017. "Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being." Policy Research Working Paper 8284, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/29075.

Fatehkia, Masoomali, Benjamin Coles, Ferda Ofli, and Ingmar Weber. 2020. "The Relative Value of Facebook Advertising Data for Poverty Mapping." In *Proceedings of the International AAAI Conference on Web and Social Media* 14: 934–938. https://ojs.aaai.org/index.php/ICWSM/article/view/7361.

Fatehkia, Masoomali, Isabelle Tingzon, Ardie Orden, Stephanie Sy, Vedran Sekara, Manuel Garcia-Herranz, and Ingmar Weber. 2020. "Mapping Socioeconomic Indicators Using Social Media Advertising Data." *EPJ Data Science* 9 (1): 22. https://doi.org/10.1140/epjds/s13688-020-00235-w.

Ghosh, Malay, and John N. K. Rao. 1994. "Small Area Estimation: An Appraisal." *Statistical Science* 9 (1): 55–76. DOI: 10.1214/ss/1177010647.

Gram-Hansen, Bradley J., Patrick Helber, Indhu Varatharajan, Faiza Azam, Alejandro Coca-Castro, Veronika Kopackova, and Piotr Bilinski. 2019. "Mapping Informal Settlements in Developing Countries Using Machine Learning and Low Resolution Multi-Spectral Data." In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 361–368. https://arxiv.org/pdf/1901.00861.pdf.

Grosh, Margaret E., Paul Glewwe. 1995. "A Guide to Living Standards Measurement Study Surveys and Their Data Sets." Living Standards Measurement Study Working Paper 120, World Bank, Washington, DC. https://documents1.worldbank.org/curated/en/270551468764720584/pdf/multi-page.pdf.

Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Head, Andrew, Mélanie Manguin, Nhat Tran, and Joshua E. Blumenstock. 2017. "Can Human Development Be Measured with Satellite Imagery?" In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, article 8. https://dl.acm.org/doi/10.1145/3136560.3136576.

Heitmann, Soren, and Sinja Buri. 2019. Poverty Estimation with Satellite Imagery at Neighborhood Levels: Results and Lessons for Financial Inclusion from Ghana and Uganda. Washington, DC: World Bank. https://www.ifc.org/wps/wcm/connect/2cae89ee-dea3-4a7e-ba79-77c9011cbd0f/IFC_2019_Poverty+Estimation+with+Satellite+Imagery+at+Neighborhood+Levels.pdf?MOD=AJPERES&CVID=mHZhcxB.

Henninger, Norbert. 1998. *Mapping and Geographic Analysis of Poverty and Human Welfare—Review and Assessment*. Washington, DC: World Resources Institute. http://pdf.wri.org/poverty_and_human_welfare.pdf.

Henninger, Norbert, and Mathilde Snel. 2002. *Where Are the Poor? Experiences with the Development and Use of Poverty Maps*. Washington, DC and Arendal, Norway: World Resources Institute and UNEP/GRID-Arendal. https://files.wri.org/d8/s3fs-public/pdf/wherepoor.pdf

Hentschel, Jesko, Jean Olson Lanjouw, Peter Lanjouw, and Javier Poggi. 1998. "Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador." Policy Research Working Paper 1928, World Bank, Washington, DC.

Hentschel, Jesko, and Peter Lanjouw. 1998. "Using Disaggregated Poverty Maps to Plan Sectoral Investments." *PREM Notes 5*, World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/11544/multi_page.pdf?sequence=1&isAllowed=y.

Hernandez, Marco, Lingzi Hong, Vanessa Frias-Martinez, Andrew Whitby, and Enrique Frias-Martinez. 2017. "Estimating Poverty Using Cell Phone Data: Evidence from Guatemala." Policy Research Working Paper 7969, World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/26136/WPS7969.pdf?sequence=1&isAllowed=y.

Hersh, Jonathan, Ryan Engstrom, and Michael Mann. 2021. "Open Data for Algorithms: Mapping Poverty in Belize using Open Satellite Derived Features and Machine Learning." I*nformation Technology for Development* 27 (2): 263–292. https://doi.org/10.1080/02681102.2020.1811945.

Hyman, G. 2013. "Poverty Maps from Unsatisfied Basic Needs Indicators in Latin America." International Crisis Group. https://www.ciesin.columbia.edu/repository/povmap/methods/NBI.pdf.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–794. https://doi.org/10.1126/science.aaf7894.

Lee, Kamwoo, and Jeanine Braithwaite. 2020. "High-Resolution Poverty Maps in Sub-Saharan Africa." *arXiv preprint arXiv:2009.00544*. https://arxiv.org/ftp/arxiv/papers/2009/2009.00544.pdf.

Lee, Myeong, Rachael Dottle, Carlos Espino, Imam Subkhan, Ariel Rokem, and Afra Mashhadi. 2017. "A Tool for Estimating and Visualizing Poverty Maps." *Myeong Lee* (website). https://myeonglee.com/publications/tool-estimating-and-visualizing-poverty-maps#:~:text=%22Poverty%20maps%22%20are%20designed%20to,different%20dimensions%20of%20poverty%20determinants.

Leidig, Mathias, and Richard M. Teeuw. 2015. "Quantifying and Mapping Global Data Poverty." *PLOS ONE* 10 (11): e0142076. https://doi.org/10.1371/journal.pone.0145591.

Malgioglio, Silvia, Nobuo Yoshida, and Johannes Hoogeveen. 2019. "New Monitoring Methods and Tools Make Development More Effective." *Data Blog*, October 29, 2019. https://blogs.worldbank.org/opendata/new-monitoring-methods-and-tools-make-development-more-effective.

Molina, Isabel, and J. N. K. Rao. 2010. "Small Area Estimation of Poverty Indicators." *Canadian Journal of Statistics* 38 (3): 369–385. https://doi.org/10.1002/cjs.10051.

Nguyen, Minh, Paul Andres Corral Rodas, João Pedro Azevedo, and Qinghua Zhao. 2018. "SAE: A Stata Package for Unit Level Small Area Estimation." Policy Research Working Paper 8630, World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/30650/WPS8630.pdf?sequence=1&isAllowed=y.

Piaggesi, Simone, Laetitia Gauvin, Michele Tizzoni, Ciro Cattuto, Natalia Adler, Stefaan Verhulst, Andrew Young, Rhiannan Price, Leo Ferres, and André Panisson. 2019. "Predicting City Poverty Using Satellite Imagery." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 90–96. https://openaccess.thecvf.com/content_CVPRW_2019/papers/cv4gc/Piaggesi_Predicting_City_Poverty_Using_Satellite_Imagery_CVPRW_2019_paper.pdf.

Puttanapong, Nattapong, Arturo M. Martinez Jr., Mildred Addawe, Joseph Bulan, Ron Lester Durante, and Marymell Martillan. 2020. "Predicting Poverty Using Geospatial Data in Thailand." ADB Economics Working Paper Series 630, Asian Development Bank, Manila. https://www.adb.org/sites/default/files/publication/666711/ewp-630-predicting-poverty-geospatial-data-thailand.pdf.

Rodas, Paul Corral, Isabel Molina, and Minh Nguyen. 2021. "Pull Your Small Area Estimates up by the Bootstraps." *Journal of Statistical Computation and Simulation* 91 (16): 3304–3357. https://doi.org/10.1080/00949655.2021.1926460.

Šćepanović, Sanja, Igor Mishkovski, Pan Hui, Jukka K. Nurminen, and Antti Ylä-Jääski. 2015. "Mobile Phone Call Data as a Regional Socio-Economic Proxy Indicator." *PLOS ONE* 10 (4): e0124160. https://doi.org/10.1371/journal.pone.0124160.

SDSN (Sustainable Development Solutions Network) TReNDS (Thematic Research Network on Data and Statistics). 2018. *Household Surveys Shape Policy Investments*. Global Partnership for Sustainable Development Data. https://www.data4sdgs.org/sites/default/files/2018-11/LSMS%20Case%20Study.pdf.

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. 2015. "Data Deprivation: Another Deprivation to End." Policy Research

Working Paper 7252, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/21867.

Smith-Clarke, Chris, and Licia Capra. 2016. "Beyond the Baseline: Establishing the Value in Mobile Phone Based Poverty Estimates." In *Proceedings of the 25th International Conference On World Wide Web*, 425–434. https://doi.org/10.1145/2872427.2883076.

Smith-Clarke, Christopher, Afra Mashhadi, and Licia Capra. 2014. "Poverty on the Cheap: Estimating Poverty Maps Using Aggregated Mobile Communication Networks." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 511–520. https://doi.org/10.1145/2556288.2557358.

Sparks, Corey, and Joey Campbell. 2014. "An Application of Bayesian Methods to Small Area Poverty Rate Estimates." *Population Research and Policy Review* 33 (3): 455–477. http://dx.doi.org/10.1007/s11113-013-9303-8.

Statistics South Africa. 2020. "Mapping Vulnerability to Covid-19." https://www.statssa.gov.za/?p=13875.

Steele, Jessica E., Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, et al. 2017. "Mapping Poverty Using Mobile Phone and Satellite Data." *Journal of The Royal Society Interface* 14 (127): 20160690. https://doi.org/10.1098/rsif.2016.0690.

Tarozzi, Alessandro, and Angus Deaton. 2007. "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas." *The Review of Economics and Statistics* 91 (4): 773–792. http://dx.doi.org/10.2139/ssrn.997829.

Tingzon, Isabelle, Ardie Orden, Stephanie Sy, Vedran Sekara, Ingmar Weber, Masoomali Fatehkia, Manuel Garcia-Herranz, and Dohyung Kim. 2019. "Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-Sourced Geospatial Information." In *AI for Social Good ICML 2019 Workshop*. https://pdfs.semanticscholar.org/9d96/bbd1bab6f66015096336052bd86662e14c6d.pdf.

United States Agency for International Development. 2022. "Demographic and Health Survey (DHS)." https://dhsprogram.com/Methodology/Survey-Types/DHS.cfm.

World Bank. 2008. *Nicaragua Poverty Assessment*. Washington, DC: World Bank. https://documents1.worldbank.org/curated/en/235491468297893170/pdf/397360ESW0vol110gray0cover01PUBLIC1.pdf.

World Bank. 2020. "Tools for Poverty and Equity Rapid Data Collection." World Bank Poverty Global Practice Virtual Summer University, Washington, DC, July 7, 2020.

World Bank. 2022. "About LSMS." https://www.worldbank.org/en/programs/lsms/overview

World Bank. 2022. "SWIFT—Rapid Poverty Assessment Tool." World Bank, Washington, DC.

World Bank. n.d. *Survey of Well-being via Instant and Frequent Tracking (SWIFT): Estimating Consumption for Household Poverty Measurement*. Washington, DC: World Bank. https://www.ifc.org/wps/wcm/connect/64f11adb-ab01-4207-93cd-dd2c-c51af16c/SWIFT-booklet-05.pdf?MOD=AJPERES&CVID=m9Or9Ia.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature Communications* 11 (1): article 2583. https://www.nature.com/articles/s41467-020-16185-w.

Yoshida, Nobuo. 2017. "SWIFT: Survey of Well-being via Instant and Frequent Tracking." https://slideplayer.com/slide/12436138/.

Yoshida, Nobuo, X. Chen, S. Takamatsu, K. Yoshimura, S. Malgioglio, S. Shivakumaran, K. Zhang, and D. Aron. 2021. "The Concept and Empirical Evidence of SWIFT Methodology (Draft)." World Bank, Washington, DC.

Yoshida, Nobuo, R. Munoz, A. Skinner, C. Kyung-eun Lee, M. Brataj, W. Durbin, D. Sharma, and C. Wieser. 2015. *Swift Data Collection Guidelines Version 2*. Washington, DC: World Bank. https://documents1.worldbank.org/curated/en/591711545170814297/pdf/97499-WP-P149557-OUO-9-Box391480B-ACS.pdf.

Zhao, Xizhi, Bailang Yu, Yan Liu, Zuoqi Chen, Qiaoxuan Li, Congxiao Wang, and Jianping Wu. 2019. "Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh." *Remote Sensing* 11 (4): 375. https://doi.org/10.3390/rs11040375.

IEG
Independent Evaluation Group
WORLD BANK GROUP
World Bank • IFC • MIGA

The World Bank
1818 H Street NW
Washington, DC 20433

The World Bank
1818 H Street NW
Washington, DC 20433