Appendix H
Poverty Data Availability in Micro data Catalog and Cost Estimation of Statistical System Improvement

Micro data Catalog

The International Household Survey Network (IHSN) provides a public catalog with listings for 4,224 surveys and censuses conducted in low- and middle-income countries, along with metadata, survey questionnaires, manuals, and reports, if available. Roughly 1,200 are identified as containing income, consumption, or other poverty-related data. Data cannot be obtained directly from the IHSN catalog, but a third of the micro data sets are available from external data repositories.

Through the Central Micro data catalog\(^1\) the World Bank provides a public listing of 1,762 surveys of various types along with their metadata and related documentation including 60 LSMS surveys. Micro data can be downloaded directly from the source, while others require approval before data are made available.

A larger micro data catalog is maintained for internal use by World Bank staff. It contains more than 5,000 surveys, of which somewhat less than a thousand contain household income or expenditure data. This catalog includes the harmonized datasets used for much of the World Bank’s poverty work. However restrictive covenants placed on many of underlying surveys by the owners or producers as a condition of their transfer to the World Bank limit their use even within the World Bank. Because the World Bank lacks a uniform policy on data acquisition, the terms of use for data sets acquired from countries or other producers are often ad hoc and poorly documented.

There are also issues related to sharing data inside the Bank. Internal sector silos hinder data harmonization and dissemination. Within the Bank, data appear to have not been adequately shared among staff working on poverty. The lack of a consistent flow of poverty data among the Bank networks and regions was cited as a key internal constraint to obtaining data beyond the external constraints posed by countries. In the Staff Survey, Focus Groups, and the case studies, Bank staff frequently noted that data needed to be shared better across the regions and the networks. Although about half (49 percent) of the internal survey respondents believed that Bank staff “almost always” or “frequently” shared poverty data,
slightly less than half (43 percent) maintained that this occurred only “sometimes,” occasionally,” or “hardly ever.”

The creation of the Global Poverty Working Group, regional data platforms, and the internal Micro data catalog containing harmonized datasets from different sources are all efforts to address some of these problems. However work on the harmonized databases, important for conducting cross-country studies and monitoring progress over time, remains fragmented, undertaken by different units at different times and without sufficient resources committed to ensure continuity. Bank-sponsored surveys are typically planned by local Bank teams with national counterparts. Although the teams may receive advice from the LSMS unit or other experienced staff, decisions about sampling, content of the questionnaire, and processing of the data may be made without adequate consideration of comparability or adherence to recommended standards. Likewise the terms under which data can be used and disseminated are often idiosyncratic, failing to take into account the Bank’s and general public interest in open access. Surveys obtained by World Bank staff for use in their work, may or may not be made available to other units or included in the Micro data Catalog. And because of the lack of common standards, considerable effort must be expended on ex post harmonization of data sets, which may still fail to resolve inconsistencies.

The cost of improving statistical system

The World Bank and other partners have attempted to estimate the shortfall in funding needed to upgrade statistical systems in developing countries. In 2004 the consensus estimate included in the Marrakech Action Plan for Statistics (MAPS) was $140-$160 million per year in additional resources for low-income and lower-middle-income countries. The MAPS cost estimates and the work program helped donors coordinate their support and encouraged partnerships for statistical capacity building.

Although MAPS achieved its objectives and stimulated additional investments in statistics, it did not address all of the shortcomings in national statistical systems. The World Health Organization (WHO) has recently estimated the cost of scaling up investment in global civil registration and vital statistics (CRVS) for 73 countries to be on the order of $3.82 billion over 10 years. Taking into account domestic contributions and recurrent expenditures, they conclude that an additional $1.99 billion is required over a 10-year period, or an average of $199 million per year; $40 million more than the Marrakech estimate per annum, for just one important statistical tool.
Morten Jerven at Simon Fraser University (2014) has estimated that using survey techniques for collecting data for the eight MDGs along with a population census would cost $1.08 billion per annum, assuming that the majority of MDG development data are survey-based and that poverty analysis requires annual collection of survey data. Demombynes and Sandefur (2014) have refined Jerven’s estimate by identifying the funding gap taking into account preexisting spending on household surveys. Focusing on countries below $2,000 per capita GDP in PPP dollars yields a total cost to international donors of closing all remaining survey gaps of less than $300 million per annum, which they point out is a fairly small share of global aid budgets.

The differences between these estimates demonstrate, as much as anything, the lack of consensus on the statistical tools needed to monitor a comprehensive development agenda. There are efforts underway with involvement from the World Bank to estimate the full cost and incremental investment needed to monitor the post-2015 agenda using broadly agreed assumptions and estimation methods. These estimates will be presented to the Financing for Development Conference scheduled for July 2015.

2 Some of the efforts to construct harmonized data sets and analytical tools include CLSP and I2D2 (DECRG), SEDLAC and LABLAC (LAC region), ECAPOV (ECS region), and SHIP (Africa region).