Data Integration and Poverty Mapping: The Experience of Sri Lanka

(Dr. D Dilshanie Deepawansa)

The ambition of the global development agenda in which the Sustainable Development Goal 2030 mainly focus on data-driven evidence-based policymaking to achieve the concept of no one behind the principle. It shades with quality, timely and reliable data with disaggregates statistics. Although many developing countries lack over half of the SDG indicators[1]. The developing countries provide statistics typically from household surveys, establishment surveys and censuses. Many surveys data provide reliable estimates at national, regional, province, district or other highly aggregate levels. Typically, the survey samples are inadequate to provide reliable estimates down to lower administrative levels. Hence, policymakers find it extremely challenging to arrive at effective targeting of socioeconomic programs. Therefore, it is important to explore alternative data access to complement the survey data.

An increase in the survey sample, data integration, and use of big data are the three options to the solution that can be used for providing statistical estimates for lower administrative levels. However, the above options also consist of several challenges and issues. For example, increasing the survey sample is not practical and statistically sound to provide precious estimates. When increasing the sample size, the organizations require additional physical resources and financial assistance. But many national statistical offices (NSOs) and statistical organizations do not possess those facilities. Further, when sample size increases it is more prone to increase non-sampling errors.

To provide reliable estimates to lower administrative levels, many developing countries have adopted small area estimation methods using survey data together with census or administrative data. However, this may present some technical issues even after combing conventional data sources such as surveys, censuses and administrative data. Therefore, it is important to use non-conventional data sources such as big data-particularly geospatial data (shadow area, car counts, road density, farmland type, roof material, and vegetation index) and mobile phone data. In addition, data integration is another efficient solution to provide new statistics to fill the data gaps by combining data from different data sources. Nevertheless, data integration is a challenging area.

Data integration is the combination of technical and business processes that are used to combine data from different sources into a meaningful single information system. Data Integration minimizes data duplication and reduces the respondent burden and increases cost efficiency. It contributes to increasing the quality of dimensions of official statistics of relevance, accuracy, reliability, timeliness, punctuality and accessibility (ESCAP, 2019). Data integration is one of the well-recognized methods used as the solution for providing statistical estimates for lower administrative levels, but it consists of several challenges and issues, such as; limited technical capacity, access to data and metadata quality and public acceptance are some of the main limitations when developing national statistical systems.

Many countries have attempted to compile poverty mapping using Small Area Estimation (SAE) technique. Sri Lanka does not have experience in data integration. However, DCS Sri Lanka has executed a small area estimation in 2005 and 2015 using household income and expenditure survey data and census of population and housing data in collaboration with the World Bank (Bedi, Coudouel, & Simler, 2007; DCS

[1] Asia and the Pacific SDG Progress Report 2020

& WB, 2015). Sri Lanka gained a remarkable decline in poverty headcount at the national level in the last two decades, but there is considerable spatial heterogeneity across districts. Sri Lanka usually provides poverty statistics based on Household Income and Expenditure Survey down to the district level. For the first time in its history, the Department of Census and Statistics (DCS), Sri Lanka implemented a poverty mapping exercise in 2005 applying a small area estimation technique and providing poverty estimates at the Divisional Secretariats level (DSS) combining the HIES 2002 data and Census of Population and Housing in 2001. The next SAE exercise was executed in 2015 using 2012/13 HIES data and Census of Population and Housing 2012 data. However, conventional SAE techniques have limitations which lead to an increase the model errors and reduce the precision of the estimates when there is a time gap between survey and census years. Further, usually, Census is conducted once in ten years, hence there is an issue when updating granular data timely.

Nowadays, poverty mapping through data integration and artificial intelligence are popular methods. Satellite imagery is a potentially useful source complementary to conventional data sources such as surveys, censuses and administrative data to predict the incidence of poverty and poverty mapping. However, images that are naturally unstructured, and noisy and analysis are prone to be difficult statistically. It requires a computer version technique to map the spatial distribution of poverty. The Philippines and Thailand have executed poverty mapping exercises in collaboration with the Asian Development Bank using data from publicly accessible satellite images integrated with household income, and expenditure and censuses using the technique by Jeal et al. (2016) as a case study to examine the feasibility of using satellite imagery to enhance the granularity of poverty statistics compiled using conventional methods. The prediction was generally aligned with published poverty estimates by the government (Hofer, et al., 2020).

The big data concept is another source of data in use to compile granular poverty statistics for data-driven decision making to reduce poverty. The big data is collected from various sources. Some examples are; social networks web text documents, large-scale e-commerce, sensor network, medical records etc. Big data exhibits several characteristics of multiple Vs: volume, velocity, variety, variability, veracity, validity, vulnerability, and volatility (ESCAP U. , Asia-Pacific guidelines to data integration for official statistics, 2020). In addition, there are characteristics such as most of these data captured and owned by the private sector and data integration differs from the traditional method in many dimensions volume, velocity, variety and veracity (Dong & Srivastava, 2013) However, analysis of big data has challengers in capturing data, data storage, data analysis, information privacy, visualization and sharing and transferring.[2] When considering the use of big data integration into official statistics National Statistical Agencies should be addressed the legislation, privacy, financial, management, methodological and technical issues and challenges introduced by The United Nations, Economic Commission for Europe.[3] The World Bank has done a feasibility study of a bottom-up method combining household survey data with contemporaneous satellite imagery to track frequent changes in local population density (Engstrom, Newhouse, & Soundararajan, 2019) .

Conclusion

[2] https://en.wikipedia.org/wiki/Big_data

[3] United Nations, Economic Commission for Europe, "In-depth review of big data", Conference of European Statisticians, sixty session plenary session, Paris, 9−11 April 2014 (ECE/CES/2014/7). Available at https://unece.org/DAM/stats/documents/ece/ces/2014/7-Indepth_review_of_big_data.pdf.

The granularity of poverty statistics is very important for evidence-based policy decisions to achieve the SDG to achieve the concept of "no one behind the principle". Many developing countries have not compiled over half of the SDG indicators due to the lack of data. Many countries typically produce statistics for policy decisions for the development of their countries based on surveys and censuses. But those are insufficient to provide reliable statistics for more granular levels. Therefore, it is important to go beyond the traditional type of data. Currently, there is a trend to use data integrations methods and big data for compilations of poverty mapping for evidence-based policy decisions. However, there are limitations to these methods. Despite this, it is important to examine the use of these approaches meaningfully to improve the quality dimensions of official statistics to use evidence-based policy decisions for the countries' developmental objectives.

References

Bedi, T., Coudouel, A., & Simler, K. (. (2007). *More than a pretty picture: using poverty maps to design better policies and interventions.* World Bank Publications.

DCS, & WB. (2015). *The Spatial Distribution of Poverty in Sri Lanka.* Department of Census and Statistics and World Bank.

Dong, X. L., & Srivastava, D. (2013). Big data integration. *IEEE 29th international conference on data engineering (ICDE)*, 1245-1248.

Engstrom, R., Newhouse, D. L., & Soundararajan, V. (2019). Estimating small area population density using survey data and satellite imagery: An application to Sri Lanka. *World Bank Policy Research Working Paper*, 8776.

ESCAP. (2019, July). Integrated Statistics: A journey worthwhile. *Stat Brief*.

ESCAP, U. (2020). *Asia-Pacific guidelines to data integration for official statistics.* Economic and Social Commission for Asia.

ESCAP, U. (2020). *Asia-Pacific guidelines to data integration for official statistics.* ADB.

Hofer, M., Sako, T., Martinez Jr, A., Addawe, M., Bulan, J., Durante, R. L., & Martillan, M. (2020). *Applying Artificial Intelligence on Satellite Imagery to Compile Granular Poverty Statistics.* Asian Development Bank Economics.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790-794.

.