POVANA Update 2025

Abdullah Mamun

2025-03-21

## 1. Introduction

POVANA, a database of household income and expenditure surveys of countries available, has been created to cater to the needs of economic modeling and welfare analysis. With several past versions, starting from 2013, we have been able to manage data for 35 countries in the current form of POVANA. The database is key to providing household data to the MIRAGRODEP modeling framework. The database offers several features including household members’ demographic profile, consumption expenditures, farm and non-farm income, wage, remittance and transfer. In 2017 we have made several modifications to the structure of the database that include quantity of food consumption where available, GPS variables, and mapping of sectors that align with GTAP sectors etc. Since then, most countries introduced new surveys as latest as in 2022. This necessitates the updating of the POVANA database.

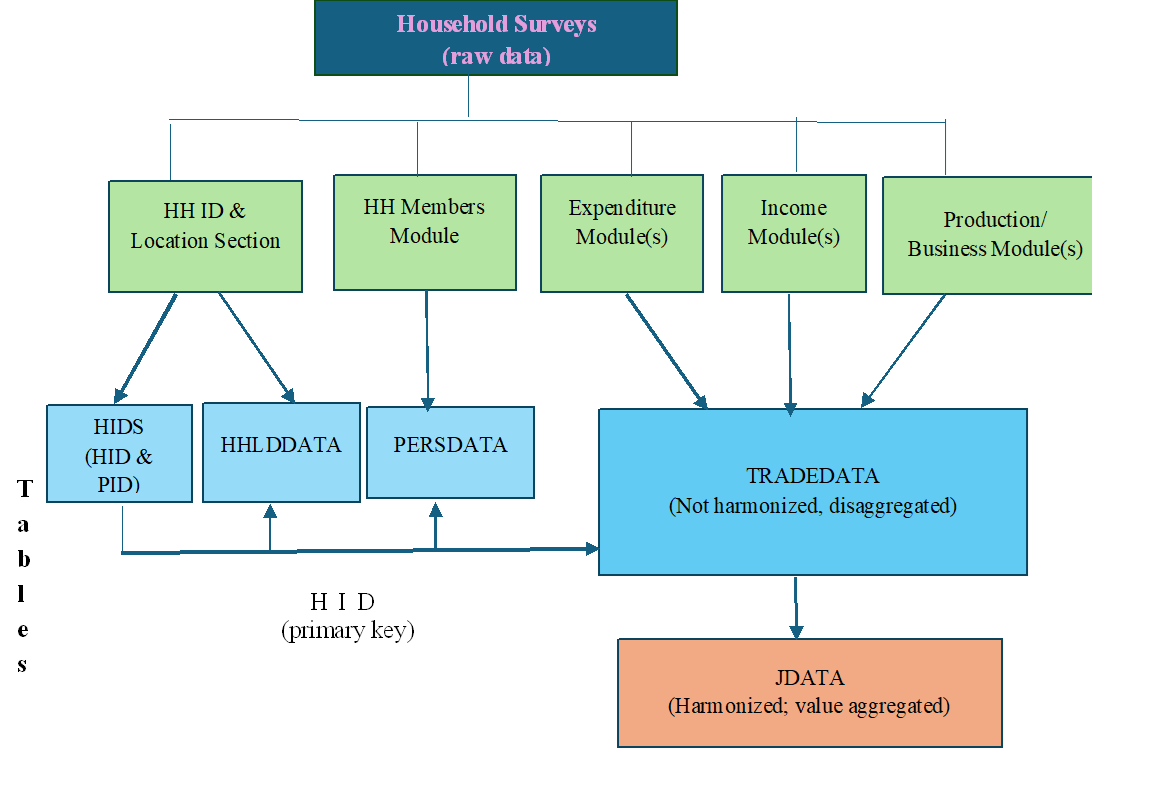
On the other hand, in the current form, POVANA database lacks the best practices of survey data processing, such as those recommended in Mancini and Vecchi (2022). One important recommendation from Mancini and Vecchi (2022) is that the household consumption aggregate, key welfare measure of household members, should be adjusted for price and spatial deflation to reflect the cost-of-living differences over time and across the national territories. The paper also discusses at length the criteria in choosing what to include and what not to in food and non-food consumption aggregates. Therefore, we think it is important that we adhere to these standard practices so that we feed consistent country data to MIRAGRODEP for any policy analysis.

In this note we have outlined a plan for an update of POVANA in 2025 and discussed the potential new features and changes in the structure of the database. Though the current form of POVANA has data for 35 countries, we have planned to make updates for 14 countries that are deemed important for global analysis given the resource constraints and time available.

## 2. Current POVANA structure and plan for improvement

Household survey data is processed to provide a set of tables that will describe households’ transactions in a compatible way with the MIRAGRODEP framework, as represented by Figure 1. It describes how household income and expenditure survey data of the countries of research interest is processed and populated into various tables. Our goal is to get a consistent and reliable POVANA database that can be fed into MIRAGRODEP or other household modeling works.

Figure 1: Data processing steps in POVANA



In the above flow chart, POVANA data as processed in R, contains several tables, each with specific purpose or uses. All tables are connected with unique Household ID (primary key in SQL language). Value data are stored either in TRADEDATA (not harmonized, disaggregated) or JDATA (harmonized, aggregated by commodity group and household). Keeping the disaggregated data gives us the flexibility of appropriate aggregation depending on the study objectives.

*Constructing consumption aggregates* Mancini and Vecchi (2022) describe four criteria for constructing National Consumption Aggregates (NCA) . These include: (i) comprehensiveness, (2) relevance, (3) typical consumption, and (4) valuation. POVANA update 2025 aims to maintain these criteria when processing household survey data. Where possible, we plan to include all consumption, including all monetary expenditures on goods and services consumed plus the monetary value of all consumption from income in kind, such as food produced on the family farm, and the value of owner-occupied housing evaluated at market prices. Mancini and Vecchi (2022) provide an excellent list of items for consideration in table 4.1 on page 36, suggesting what to include and what not to include. This will be a guiding principle in the update.

*Adjusting for price variation* As argued in Mancini and Vecchi (2022), we plan to implement price adjustment in consumption and income aggregates so that we get real consumption and income value instead of nominal values. This adjustment will reflect cost-of-living differences over time and across countries. It is possible that we have appropriate price deflators such as consumption price index or true cost-of-living (TCLI) index.

*Robust income data* Processing income data, particularly those from farm and nonfarm activities are important. In this regard, where available we plan to process all sources of income including wage, farm and nonfarm income, remittance, social security benefit, other transfer income.

*Household strata* In this update we also plan to introduce some additional features in the structure, particularly in household background. Currently, we process household members’ education, occupation and industry in the background. To assess the impact of global trade and environmental policies or price shocks on poverty, food security and nutrition across different strata or sub-populations within a country, we plan to construct these sub-populations, such as agricultural self-employed , non-agricultural self-employed, rural wage labor stratum , urban wage labor stratum etc.

*Mapping of sectors and commodities* After standardizing the item labels, we need to map commodities or services to code that keep standard nomenclatures available in different international classification systems. In this case we opted to follow GTAP classification of codes or sectors in line with the need for MIRAGRODEP. While we have kept three-lettered codes for non-food goods and services same as in GTAP, we have customized foods and crop items so that we get disaggregated level of information on production and consumption expenditures by households.

*Sensitivity analysis and reproducibility* Our goal is to provide consistent and reliable POVANA data for any household modeling and impact or distributional analysis. To achieve this, we will carry out several sensitivity analyses at aggregate level and perform robustness check. After addressing all issues related to outliers and missing data imputation, the sensitivity analysis will be done to check if poverty incidence and income aggregates conform with official data. In this regard, we can check for cumulative distribution function of per capita income or head count difference curves, as suggested in Mancini and Vecchi (2022).

All data will be processed in R. Special attention will be given to automate data processing as much as possible, particularly for commodity mapping, imputation, and sensitivity analysis. Disaggregated data will allow us to prepare data in other formats such as those being in used in RIAPA model. R codes will be prepared and documented well so that results can be reproduced efficiently and with greater accuracy.

## 3. Update plan: list of priority countries

The following table lists the countries for which household survey data can be obtained. The last column identifies the countries which will be updated with the current resources in 2025. A total of 14 countries (indicated by ‘Yes’) are proposed here. Selection of countries is done considering geographical coverage, country size and income groups.

Table 1: List of countries



## 4. Final output

Final output of this update will be a compilation of processed data files in both gdx (compatible with GAMS) and rds (compatible in R) format. Individual as well as one compiled file, identified by country and year of survey, will be produced.

## Reference

Mancini, C., and Vecchi, G. (2022). On the construction of a consumption aggregate for inequality and poverty analysis. World Bank Group. Washington DC.

Hertel, T., Verma, M., Ivanic, M., Magalhaes, Ludena, C., and Rios, Ana R. (2015). GTAP-POV: A framework for assessing the national poverty impacts of global economic and environmental change. GTAP Technical Paper No. 31, GTAP.