

Survey-to-Survey Imputation of Poverty: Methodology, Limitations, and Replication Code

This repository contains the replication code for the paper "The why, the how, and the when to impute: a practitioners' guide to survey-to-survey imputation of poverty" by Paul Corral, Andres Ham, Peter Lanjouw, Leonardo Lucchetti, and Henry Stemmler (February 18, 2025).

Overview

Survey-to-survey (S2S) imputation has become an essential tool for poverty measurement when traditional household consumption or income surveys are unavailable. This repository provides code to replicate the model-based simulations and analyses presented in our paper, which examines the methodological challenges and potential pitfalls of S2S imputation.

Files to run:

1. \2.scripts\0.Main_master_file.do

See "Path_for_reproduction.xlsx" to see where all figures and tables are created.

Key Features

- Simulation of populations with controlled properties for methodological testing
- Implementations of multiple imputation techniques
- Tools for analyzing biases from sampling issues and sample corrections
- Simulations for examining parameter changes over time
- Analysis of omitted variable bias in poverty predictions

Requirements

The code mainly uses Stata for analysis. We recommend:

- Stata (version 17 or higher)
- Required user-written Stata packages:
 - `wentropy` (for entropy-based reweighting)
 - `hetmireg` (for heteroskedastic MI regression)
 - `lassopmm` (for lasso with predictive mean matching)
 - `sp_groupfunction` (for aggregation of indicators)
 - `groupfunction` (for aggregation of indicators)

Data Access

All data utilized is simulated and created in the underlying scripts.

Citation

If you use this code in your work, please cite:

Corral, P., Ham, A., Lanjouw, P., Lucchetti, L., Stemmler, H. (2025). Stress-Testin Survey-to-Survey Imputation: Understanding When Poverty Predictions Can Fail. World Bank Policy Research Working Paper (11192).

Main Findings

Our simulations demonstrate:

1. Standard bias-correction techniques may be insufficient when source and target surveys differ fundamentally
2. S2S methods tend to replicate the welfare distribution of the source survey, making them unreliable for measuring changes in inequality and poverty
3. Omitted variable bias affects poverty predictions, particularly when imputing across time periods with significant economic changes

Contact

For questions about the code, please contact Paul Corral: pcorralrodas@worldbank.org