

Maria Grazia Pittau (Sapienza University of Rome); Roberto Zelli (Sapienza University of Rome); Saida Ismailakhunova (World Bank)

Assessing Individual Poverty Status Using Repeated Cross-Sectional Surveys

When available, longitudinal or panel data are used to study poverty dynamics to evaluate how the poor react differently to macro-economic shocks, or generally to changes of the contextual environment where they live. Longitudinal data also contribute to the empirical literature on vulnerability to poverty, where vulnerability is defined as the (in)ability of individuals to protect themselves to fall into poverty when an adverse event occurs (Lopez-Calva and Ortiz-Juarez, 2014).

The importance of longitudinal data is widely accepted, however the availability of reliable panel data spanning over a sufficient time period is very rare in developing countries and cross-sectional surveys, evenly or unevenly repeated over time, are the only data available.

In absence of panel data, individual or family poverty is measured assessing the current poverty status ignoring dynamics over time. Recent studies tried to combined micro and macro determinants of poverty in order to simultaneously assess the impact of both individual and macroeconomic (regional) shocks on poverty (see, among others, Chen and Wang, 2015). However, even when family-level and regional-level factors are considered as potential determinants of poverty, results from such a static poverty analysis can be misleading since the influence of individual and structural characteristics on poverty might not be necessarily constant over time.

To overcome the shortcomings of traditional static poverty assessment, we propose a methodology that takes into account the dynamic dimension of poverty-- at both individual(family) and regional levels--when only when only a set of independent cross-sectional surveys are available. Differently from longitudinal data, repeated cross-sectional data consist of independent observations drawn from the same context (e.g. the same region) at many different time-points, and can therefore be treated as clustered within regions and time. Although we cannot fill the data gap that exists in many developing countries, we tried to develop an estimation method able to evaluate the dynamic effect of micro and macro predictors on individual poverty risk. We cannot follow the same individuals over time but we are able to capture social or economic changes since there there is a spatial-temporal dynamics in this collection of data that manifests itself (Cagnone et al., 2017). Therefore, despite the limitations of purely cross-sectional data, our approach estimates poverty dynamics based on repeated cross-sectional surveys resolving the problem of missing lengthy panel data.

In our analysis we distinguish between cross-sectional and longitudinal associations between macro-economic variables and poverty, taking into account the clustering of each observation (individual or family) within region&years. The proposed methodology relies on dynamic multilevel models that treat individual poverty status as a function of individuals' characteristics and circumstances, in interaction with time-varying and time-constant features of their economic contexts. For example being low educated in a region with high level of unemployment rate might have a different effect on poverty risk of being low educated in an area characterized by low level of unemployment and this effect might vary over time.

Our dynamic multilevel model explicitly accounts for the hierarchical nature of our data and for their different levels of variation: individuals (families), region-years, regions and time. Our individual-region-year data are clustered within region-year, region-years data are clustered within regions and also within years, with the potential for predictors at all four levels.

The complexity of these dynamic hierarchical models has prevented their use so far because of their well-known problems of convergence. Casting them in the Bayesian statistical framework offers a reasonable solution and recent developments of simulation techniques such as Markov chain Monte Carlo (MCMC) facilitates fitting these models with the aim of exploring real world complexities of data.

To specify the dynamic poverty model, we use the library `{\tt rstanarm}` for R which uses `{\tt lme4}` syntax to facilitate building complex hierarchical generalized linear models (Carpenter et al. 2017; Stan Development Team 2017). The function `{\tt stan_glmer}` allows to fit a fully Bayesian model using Hamiltonian Monte Carlo sampling and easily access the samples. The hierarchical covariance matrix for each set of hierarchical intercepts and slopes is decomposed into a correlation matrix and a pair of variances.

We empirically illustrate our approach for Kyrgyz Republic using independent cross-sectional Kyrgyz Integrated households budget and labor force surveys (KIHS) available over the period 2013--2017. The KIHS is designed to measure consumption-based poverty in the country and its regions and to analyze the socioeconomic dimensions of people's living standards. Each year the KIHS covers around 5,000 households and around 20,000 individuals. The sample design is a two-stage random sampling, stratified into 16 strata, representing urban and rural dimensions of the seven country regions (oblasts) (Batken, Jalal-Abad, Issyk-Kul, Naryn, Osh, Talas and Chui), the city of Bishkek and Osh city.

To anticipate our main findings, some significant associations between poverty status and individual/macro predictors hold both cross-sectionally and longitudinally. Particularly, at micro level education plays an important role in predicting poverty and its influence increases over time remaining important across regions. At regional level, unemployment rate matters substantially for poverty and this association holds cross-sectionally but not longitudinally. That is, there is a significant effect of enduring differences in oblast's level of unemployment, but the longitudinal variation in the level of unemployment (measured as deviation from the regional mean) over the period is not associated with variation in poverty risk. Moreover, unemployment rate affects individual poverty differently according to where people live: rural or urban areas. Per capita gross regional product is not associated with poverty neither cross-sectionally nor longitudinally. Instead is per capita GDP at country level that manifests a trickle-down effect over time.