Nicolai Suppa (Centre for Demographic Studies, Bellaterra, Barcelona, Spain), Sabina Alkire (OPHI, University of Oxford), Usha Kanagaratnam (OPHI, University of Oxford), Ricardo Nogales (OPHI, University of Oxford)

# Assessing Normative Decisions in Poverty Measurement

Normative decisions are an inescapable part of any poverty measurement exercise. Reaching a normative decision is, however, a non-trivial exercise, as rather diverse aspects have to be balanced, including conceptual aspect, peoples' values, various forms of technical expertise, and data availability. These aspects then have to be fuelled into a public debate, while simultaneously accounting for the actual purpose of the measure (e.g., monitoring national poverty). The present paper seeks to examine normative implications and other policy-relevant consequences of such decisions more rigorously. Assessing the sensitivity and robustness of particular results is more than a compulsory exercise. In fact, knowledge to which extent results are sensitive to decisions can actually facilitate the decision process itself, as it may be easier to agree on a reasonable range rather than a specific value for a particular parameter, like the poverty cutoff. Moreover, it is actually vital for the decision process to understand the different normative implications in the first place. Finally, a profound knowledge of policy-relevant empirical implications, may help to better appreciate and characterise alternative design options (e.g. union vs non-union approach or unequal weighting schemes).

The present paper focuses on one particular implementation of a multidimensional poverty measure, as specific normative decisions and assumptions vary to some degree with the approach adopted. We note, however, that the principle flow of analyses can be applied to other poverty measures as well (including monetary poverty measures).

In the measurement of multidimensional poverty two types of normative decisions received particular attention. First, the cross-dimensional *poverty cutoff*, introduced by Alkire and Foster (2011) offers an intermediate solution between the polar cases of union and intersection (Atkinson, 2003). For example, several scholars recently advocated a union approach (e.g., Pattanaik and Xu, 2018, Datt 2018, Dotter and Klasen2014), where a single deprivation suffices for being considered poor. Second, assigning a particular *weighting scheme* on a normative basis may cause uneasiness among scholars and reluctance among policy makers. Indeed, one may expect different weighting schemes to give different results. *Selecting indicators*, however, received less attention and is, somewhat surprisingly, less disputed.

A large share of robustness analyses seeks to ascertain robustness for the entire domain of a particular parameter. Unlike this line of research, we follow the Atkinson report (2016:171) and distinguish local and global robustness. Therefore, we also illustrate how to obtain a reasonable or relevant parameter ranges. Finally, we pay special attention to the simple headcount ratio, which would at least be part of any alternative poverty measure proposal.

The present paper assesses the sensitivity parametric choices with respect to several outcomes, which may reveal critical normative implications or are particular pertinent to policy makers, or both:

1. the *headcount ratio* of a poverty measure must range in a certain interval to be useful and informative for public policy. Moreover, overly strong changes in the headcount ratio, triggered by seemingly minor parameter modifications, may undermine credibility and agreeability of the measure.
2. *subnational rankings* are important for the credibility of a measure as well (it should be broadly consistent with the common sense). Often subnational rankings are essential for allocating funds, too. Therefore, overly responsiveness to parameter changes, may invite battles for funding.
3. Different measures may identify different people as poor, despite similar headcount ratios. Therefore, we also examine *poverty sets*, and explore to which extent alternative measures identify the same people as poor. It may turn out that entirely different sub-populations would be considered poor. Normative implications like these, are essential for the decision maker.
4. Finally, a generalisation of the previous analysis is to examine the *correspondence of the underlying welfare variable* for alternative parametrizations compared against our preferred one. In our case this is the weighted deprivation count. This analysis reveals how poor households are under alternative parameter choices. Complete reversals, for instance, would be normatively problematic.

The empirical analyses use data from latest round of the global MPI, an internationally recognised poverty measure, which has several advantages for the purpose of this paper: (i) it offers comparable cross-country data, (ii) it allows us to adopt methods which account for sampling errors (i.e. inferential exercises), and (iii) it allows us to return to the original household level data and to perform micro-level analyses, like comparing poverty sets. It is important to note, that the purpose of the global MPI is to allow cross-country comparisons. Therefore, boundary values like 0 or 1 for the simple headcount ratio play a sensible role in this particular context. The primary interest of the present paper is however in applications of the global MPI to countries in which it could serve as a starting point for a useful national poverty measure as well. The reason is that we expect in-depth analyses for countries with headcount ratios close to the boundary values either impracticable or informative.

Nonetheless the empirical analysis relies on point estimates for all 105 countries and more than 1200 sub-national regions. The global MPI is based on the Alkire-Foster method and comprises ten indicators organised in three dimensions (health, education, and living standards). Our preferred specification is characterised by an equal-nested weighting scheme and a poverty cutoff of k=33%, meaning an household is considered poor if it is deprived in 33% or more of the (weighted) possible deprivations. Our analyses cover 10 different choices for the poverty

cutoff (including the union and intersection approaches), 231 different weighting schemes ranging from equal-nested weights to retaining a single dimension only. We also explore six alternative indicator selections, where compared to our preferred specification one living standard indicator is dropped at a time.

We argue that 20-45% is as reasonable range of alternative poverty cutoffs, which is motivated by the deprivation value of being deprived in two living standard indicators (about 11%). Moreover, we consider weighting schemes, where a single dimension receives a 50% weight (while the other two receive 25% each), as an upper bound of a reasonable weighting schemes, since assigning an even higher weight to a single dimension, would require substantial argumentation.

**Selected preliminary results**

- In terms of alternative indicator selections we observe that the *headcount ratio* of the global MPI can drop up to 10%-points if a single living standard indicator is removed, irrespective of whether the poverty cutoff is chosen as k=1% (union) or k=33%. On the other hand, we observe that drops of 10-30%-points (or even more than 50%-points in one case) only occur for k=1%. Interestingly, we also observe in several instances, where the union approach implies very high poverty rates (i.e. 90% or more), that the headcount still may drop with up to 10-%points when removing a single indicator. This indicates that even when poverty seems to be literally pervasive, this assessment may not be backed by other simultaneously experienced deprivations.
- We find sub-national *rank robustness* for the headcount ratio to be relatively high for relevant alternative choices of the poverty cutoff and the weighting schemes. More specifically, we find for most countries that around 80% of possible pair-wise comparisons among the subantional units, which were significant under the original parametrization, are also significant at the alternative parametrizations. Our results also suggest that rank-robustness sometimes occurs beyond the range of reasonable poverty cutoffs and weighting schemes. In general, however, we observe lower subnational rank-robustness once we approach the boundary values of the parameter space (i.e. k close to union or intersection and weighting schemes assigning extraordinary high weights to a single dimension).
- In our analyses to which extent parameter choices *diverge poverty sets*, we find for most countries a relatively stable identification of the poor for the relevant parameter spaces. More specifically, when comparing pairwise our preferred specification with alternative parametrizations, we find for many countries that 80% or more of those who are poor according to one of the measures are actually poor according to both. This share is found to fall gradually with more distant, i.e. unequal weighting schemes (for any given poverty cutoff). For the special cases of intersection and union we observe as expected poverty sets to coincide for alternative weighting schemes, as long as we only allow strictly positive

weights. However, once we allow for zero-weights, i.e. dropping one or more entire dimensions, poverty sets tend to diverge sharply as well.

- Finally, we find that different relevant weighting schemes, tend to assign households similar deprivation counts. Even when an entire dimension is dropped, it rarely happens that somebody considered considered very poor according to one scheme is suddenly almost non-poor according to the other scheme. This stability at least partly results from the fact that still 6–8 indicators are remaining according to which the deprivation is count is constructed, i.e. simultaneously experienced deprivation still matters. Indeed, this observation highlights the weights' function as so-called 'deprivation values'. Only for weighting schemes which assign almost exclusively all weight to a single dimension, we sometimes find househholds to be assessed systematically different from the equal weighting scheme.
- Across all analyses we observe that the more in-depth analyses like rank-robustness and analysis of poverty sets deliver more reliable results for countries with a poverty rate larger than 1.5% and more informative (or non-trivial) results for countries not having a too high headcount ratio (say, less 80%). This is not surprising, since the analysis of very small proportions of the population, say 1%, is hampered by statistical problems and for certain parameter combinations there are no poor people in the in the first place. However, for headcount ratios typically produced by national measures in practice our additional analyses deliver instructive insights.

**Preliminary conclusions**

- We conclude that in many instances, the empirical implications of the global MPI are locally insensitive to moderate changes in weighting schemes and the poverty cutoff. At the same time our results document, that the global MPI is not globally robust to parametric decisions, which essentially reflects that effectively fundamentally different measures can actually be designed within the same framework. Likewise, it highlights that decisions should be made well-informed in the first place. Therefore, it is vital for the decision process to account for these normative implications in the first place.
- Given the relatively low burden in computation, we recommend to perform analyses along the suggested lines on a routinely basis. Normative implications then can be identified early on, and results on local insensitivity results may facilitate reaching an agreement on some of the parameters.
- A potential caveat of the headcount ratio under a union-cutoff is its potentially strong sensitivity to the indicator selection. We find that this is not only a theoretical possibility, but does indeed occur. Therefore, we suggest to qualify the supposedly advantage of union and intersection approaches of being independent of weighting schemes, as this claim only holds for strictly positive weights. In practice, this may become problematic, for instance, if

upcoming indicator updates cause pronounced changes in the headcount, which then may undermine credibility and support of the entire measure.