Sensitivity Analyses in Poverty Measurement: The Case of the Global Multidimensional Poverty Index

Sabina Alkire
University of Oxford

Ricardo Nogales Carvajal
Universidad Privada Boliviana

2019

IARIW-World Bank

Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World" Washington, DC, November 7-8, 2019

Sensitivity Analyses in Poverty Measurement: The Case of the Global Multidimensional Poverty Index

Sabina Alkire

Usha Kanagaratnam

Ricardo Nogales

Nicolai Suppa

Paper Prepared for the IARIW-World Bank Conference

Washington, DC, November 7-8, 2019

Sensitivity analyses in poverty measurement. The case of the global multidimensional poverty index.

Sabina Alkire, Usha Kanagaratnam, Ricardo Nogales, Nicolai Suppa*

October 2019

Early draft; references incomplete; comments welcome: nsuppa@ced.uab.es

Abstract

This paper provides an extensive sensitivity analyses of the global multidimensional poverty index (MPI), which is a counting-based measure of acute poverty covering over 100 developing countries. Empirically, the paper probes the sensitivity of poverty measures and comparisons to modifications in key parameters. Outcomes studied include the adjusted headcount and headcount ratios and their subnational rankings, as well as the exact set of people who are identified as poor. The parameters that are adjusted include the poverty cutoff, weights or deprivation values, and indicators. Multidimensional poverty measures are generated using 10 alternative poverty cutoffs, 231 alternative weighting schemes, and six alternative indicator selections, in addition to the global MPI baseline specifications. Comparisons across 1226 subnational regions for 98 countries are assessed using the percent of pairwise comparisons for an alternative parameter that are robust in comparisons with the global MPI baseline. Assessments of the fit between poverty sets in relation to the global MPI poverty set use the Jaccard coefficient. Overall, the outcomes show little sensitivity when parameters are changed within plausible ranges, but there are a number of general findings of potential interest that emerge. Finally, the present paper also suggests 'second-order' sensitivity analyses to deepen the understanding of the underlying methods by varying poverty cutoffs and indicators simultaneously. The union-based measures are less stable than the base-line measure.

Keywords: poverty measurement, sensitivity analysis, multidimensional poverty, global poverty

JEL Codes: I32, C43

*This manuscript benefited from comments made by James Foster, Natalie Quinn, and participants of the HDCA 2019 in London. Sabina Alkire (sabina.alkire@qeh.ox.ac.uk), Usha Kanagaratnam (usha.kanagaratnam@qeh.ox.ac.uk) and Ricardo Nogales (ricardo.nogales@qeh.ox.ac.uk) Oxford Poverty and Human Development Initiative (OPHI), University of Oxford, UK; Nicolai Suppa Center for Demographic Studies, Autonomous University of Barcelona, Spain and OPHI, University of Oxford, UK (nsuppa@ced.uab.es). Nicolai Suppa gratefully acknowledges funding by the Spanish Ministry of Science, Innovation and Universities Juan de la Cierva Research Grant Programs (IJCI-2017-33950), the European Research Council (ERC-2014-StG-637768, EQUALIZE project); and the CERCA Programme, Generalitat de Catalunya, and the other authors the support from DFID. All errors remain our own.

1 Introduction

Poverty measurement inescapably entails value judgments. Considerable attention has been paid to the selection of a poverty measurement methodology, hence to clarifying which of various desirable properties candidate measures fulfil.[1] Yet even after the important question of a methodology is settled, no less momentous value judgements remain. Monetary poverty measures, for instance, require equivalence scales, price indices and other decisions taken in creating the welfare aggregate, potentially purchasing power parities, and a poverty line. Multidimensional measures typically require the definition of indicators, weights or deprivation values, and a poverty cutoff.

These parametric value judgements, however, frequently attract the lion share of criticism. In case of the World Bank's dollar-a-day, critics questioned, for instance, the estimation procedure for the international poverty line (e.g., Deaton, 2010, Klasen *et al.*, 2016, Kakwani and Son, 2016) or the purchasing power parities (e.g., Deaton and Dupriez, 2011, Ackland *et al.*, 2013). In the case of the 2010 global MPI, critics were particular worried about the chosen weights (Ravallion, 2011, 2012), the poverty cutoff (Duclos and Tiberti, 2016, Datt, 2018, Pattanaik and Xu, 2018), and the indicator selection (Dotter and Klasen, 2014).

Different strategies can be adopted to assess the extent to which comparisons that result from a chosen poverty measure would differ if alternative parameters had been used. If the results are highly sensitive to the parameters used, then the requirement for compelling justification of the chosen parameters is essential; the requirement may be less exacting if the poverty results are quite consistent for a range of plausible parameters. The most general strategy for both unidimensional and multidimensional poverty is dominance analysis.[2] However empirically dominance results may or may not emerge and thus potentially leaving policy makers without counsel, and even if first dominance is found in two periods, it is not possible to ascribe a cardinal distance to the ranking. And in the case of multidimensional poverty, the sample size will rarely support more than three indicators.

The more commonly applied strategies are sensitivity analyses, which explore the effect of alternative parameter values (e.g., the poverty cutoff) on selected outcome variables or relationships (e.g., subnational rankings in terms of simple headcount ratios). Sensitivity analyses are common in poverty measurement and have been carried out for various parameter-outcome combinations.[3]

This paper provides an extensive sensitivity analysis for the global MPI. The global MPI is a cross-country comparable multidimensional poverty measure using the method of Alkire and Foster (2011), that has been calculated since 2010 and is currently available for over 100 countries (Alkire and Santos, 2014, Alkire *et al.*, 2019b). Previous research directly employed the global MPI to compare countries and subnational re-

[1] See, for instance, Sen (1976), Foster *et al.* (1984) among others.

[2] Dominance techniques for monetary poverty lines and equivalence scales are discussed in Atkinson (1987), Foster and Shorrocks (1988), Atkinson (1992), Davidson and Duclos (2000). Adaptations to measures of multidimensional poverty can be found in Duclos *et al.* (2006), Chakravarty and D'Ambrosio (2006), Lasso de la Vega (2010), Alkire and Foster (2011), Yalonetzky (2014).

[3] Sensitivity to alternative equivalence scales, for instance, has been explored in Buhmann *et al.* (1988), Coulter *et al.* (1992), Banks and Johnson (1994), Jenkins and Cowell (1994), Burkhauser *et al.* (1996), De Vos and Zaidi (1997), Duclos and Mercader-Prats (1999) among many others.

gions (Alkire and Seth, 2015, Alkire *et al.*, 2017, Jindra and Vaz, 2019). Moreover, the global MPI also serves as an international benchmark for country-level studies employing similar or different specifications (Vijaya *et al.*, 2014, Espinoza-Delgado and Klasen, 2018, Ogutu and Qaim, 2019, Datt, 2018, World Bank, 2017, 2018). The present paper assesses the sensitivity of the global MPI in terms of simple levels of the headcount ratio,[4] its implied subnational rankings, and poverty sets with respect to 10 alternative poverty cutoffs (including union and intersection), 231 weighting schemes (including 21 'plausible' ones), and six alternative indicator selections (where one living standard indicator at a time is removed).

This paper complements previous research on sensitivity analyses in poverty measurement in several important ways. First,[5] since sensitivity analyses in poverty measurement probe the empirical consequences of normatively relevant information they fundamentally differ from sensitivity analyses in other fields, like econometrics. Therefore, the present paper provides a conceptual integration into a social choice framework, which guides the interpretation of empirical evidence. The paper also observes that sensitivity analyses in poverty measurement are an integral component of the initial process fixing the parameters: the availability of this evidence actually allows those involved in measurement design to take this normatively critical information into account.

Second, the paper includes a focus on poverty sets. Previous research has assessed sensitivity in terms of simple headcount ratios, average gaps, adjusted headcount ratios, their respective subnational rankings, the absolute number of the poor, and time trends (Ravallion and Bidani, 1994, Ravallion and Sen, 1996, Coulter *et al.*, 1992, Chen and Ravallion, 2010, Alkire and Santos, 2014, Santos and Villatoro, 2018, Kakwani and Son, 2016). Analyses of poverty sets have been used to compare, for instance, income poverty and material deprivation measures (e.g., Nolan and Whelan, 2011, Whelan *et al.*, 2004), income and consumption expenditure poverty measures (e.g., Meyer and Sullivan, 2012), or alternative methods to identify households living below the poverty line in India (Alkire and Seth, 2013), but they have not yet been systematically incorporated in the literature on sensitivity analysis to parameters choices. In assessing the extent to which households are consistently identified as poor under four alternative weighting schemes, Santos and Villatoro (2018) recently introduced poverty sets into robustness analyses. The present paper, however, adopts a more systematic approach and, moreover, seeks to identify thresholds where poverty sets begin to diverge. Since changes in poverty sets are not necessarily detected by changes in headcount ratios their distinct analysis seems imperative. Moreover, robust poverty sets may contribute to both credibility and social acceptance of a given poverty measure.

Third, the parameters ranges are more general. Previous research on multidimensional poverty tends to explore sensitivity of few particularly relevant alternative parameters (Alkire and Santos, 2014, Espinoza-Delgado and Klasen, 2018, Santos and Villatoro, 2018), while Pasha (2017) explored certain statistical weights. Assessing the robustness of comparisons to a range of 'plausible parameters' was suggested by Sen (1999, p. 78) cf. World Bank (2017, p. 171). The analysis of plausible parameter

[4]The final paper will include adjusted headcount ratios; this draft focuses on the headcount ratios.
[5]This section will be developed in the subsequent version of this paper.

ranges offers a sensible middle-ground between analysing a few selected parameter values and the entire domain, in particular because expecting ‘global insensitivity’ is, in fact, theoretically unsubstantiated. This paper covers the plausible range of parameters more intensively, and also explores the general range of parameters in order to build a more general understanding of robustness empirically.

Fourth, robustness to the inclusion or omission of indicators is assessed. Previous robustness analyses of multidimensional poverty measures routinely consider the sensitivity to the poverty cutoff, alternative weighting schemes and alternative indicator definitions (Alkire and Santos, 2014, Alkire and Seth, 2015, Santos and Villatoro, 2018, Espinoza-Delgado and Klasen, 2018), whereas the inclusion or omission of indicators tends to be overlooked. Santos and Villatoro (2018) take selected indicator omissions into account, and assess the robustness to indicator omissions together at the same time with other modifications of the specification such as weights. While understanding the effects of multiple parameter adjustments is also of value, in this paper, we choose to change exactly one parameter at a time in order to have a clear documentation of how each change affects the measured outcome. We find, for instance, that dropping a single indicator may result in changes of simple headcount ratios or subnational rankings, which are quantitatively similar to reasonable changes of the poverty cutoff or the weighting scheme. The present paper, therefore, argues that the indicator inclusion or omission should be more routinely subjected to sensitivity analyses.

Fifth, the robustness of subnational rankings is assessed. Previous assessments have focused on national rankings and not on subnational groups such as age, region, or rural-urban area. This paper illustrates a methodology for considering disaggregated units, which could then be applied to other subnational groups.

Finally, the majority of previous sensitivity analyses in poverty measurement, varies a single parameter at a time (Dhongde and Minoiu, 2013, p. 17) while only some studies assess the robustness across several parameter variations simultaneously (Santos and Villatoro, 2018, p. 71–72). The present paper, however, introduces a ‘second-order’ sensitivity analysis, which assesses the extent to which a certain parameter sensitivity depends on further, so far fixed, parametric choices. Our results suggest, for instance, that a union cutoff may entail a higher sensitivity of both national headcount ratios and implied subnational rankings, to modified indicator selections, compared against our non-union reference cutoff. Second-order sensitivity analyses may provide instructive insights into methodological questions, and thus offer additional guidance for devising better poverty measures.

The remainder of the paper is structured as follows: section 2 provides the conceptual background on value judgments and sensitivity analyses in poverty measurement. Section 3 introduces the underlying data and methodology, whereas section 4 and 5 contain the first and second-order results. Finally, section 6 offers some concluding remarks

2 Conceptual Considerations

This section will be expanded subsequently.

Value judgments as described above are inescapably part of any exercise in poverty measurement. A multidimensional poverty measure must satisfy various actors in the society: the policy actors must reflect that the measure illuminates deprivations that policy action can redress; taxpayers must be satisfied that the deprivations are those that must be confronted in a fair society; poor persons and communities must judge that the poverty measure reflects important elements of their experience of poverty. Understood in this way, the task of specifying a poverty measure can therefore be reframed as translating social value judgments into parametric choices. In a plural society, a range of legitimate and reasonable views on social value judgements is to be expected: between groups and indeed between the same person who may change his or her view. For that reason, poverty measures that are used for public policy should seek to elicit the range of plausible parameter values, and offer for policy use poverty measures that are robust to this range. The present paper illustrates how such an exercise might unfold.

3 Data and Methods

This paper undertakes a very data intensive and computationally demanding exercise. This section introduces the components of that exercise. The datasets used for the global MPI represent 5.7 billion people in 101 countries whose 'poverty sets' are used in some analyses. Other exercises focus on 98 countries which can be disaggregated into 1226 subnational regions. These are introduced in the section 3.1. Next, we introduce the global MPI in section 3.2 which forms the baseline parameterization against which all alternatives are assessed. Note that robustness tests are done for the associated multidimensional Headcount Ratio (in this paper), and also that a later exercise omits one living standard indicator at a time. Section 3.3 articulates the three parameters that are varied and the general and plausible ranges that are implemented in the results section. Section 3.4 defines the first of the two core sensitivity methodologies implemented, namely the proportion of robust pairwise comparisons, which is implemented on subnational regions. Section 3.5 defines the second methodology, the Jaccard coefficient, which is used to evaluate the similarity between sets of the poor who identified using an alternative parameter vs the baseline MPI. Each of the core methodologies is then implemented and interpreted with a set of parameter changes in the results section.

3.1 Data

The data used for the estimations in this paper were prepared strictly following the methodology outlined in (Alkire *et al.*, 2019b). In 2019 the global MPI was calculated for 101 countries and 5.7 billion people (2016 population figures). The underlying data is drawn from DHS, MICS, PAPFAM, and some national surveys. All surveys were carried between 2007 and 2018, with 75 surveys covering 5.1 billion people dating between 2013–2018. The global MPI is also routinely estimated for subnational units to the extent the data permits. Effectively, 1226 subnational regions can be meaningfully analysed.[6] The number of available subnational regions varies significantly with countries, from 2 to more than 30. Only for three countries the underlying survey does not permit any subnational analysis at all (Armenia, Bosnia and Herzegovina, Saint Lucia). The specific details of each dataset preparation are found in Alkire *et al.* (2018, 2019b); the papers also detail policies related to missing values, missing indicators, sample drop and so on are also specified. For example if a dataset is missing an indicator, the other indicators in that dimensions are re-weighted such that their weights sum to one-third.

3.2 The global MPI

The global MPI employs the dual cutoff-counting approach suggested by Alkire and Foster (2011), and it is implemented using 10 indicators categorized into three conceptual dimensions, with equal nested weights and a poverty cutoff of one-third. The indicators, deprivation values or weights, and dimensional categories are provided in table 1; the poverty cutoff is one-third. For details on the global MPI methodology please see Alkire *et al.* (2019b) and the references cited therein. It suffices for this paper version to remind the reader that each household is identified as deprived or non-deprived in each indicator; their weighted deprivations are summed into a deprivation score, and they are identified as poor if their deprivation score is equal to or great than one-third. Considering sampling weights, the persons who are identified as poor constitute the 'poverty set' and the percentage of the population who are identified as poor constitute the 'simple headcount ratio' for any country or subnational unit.

This paper provides certain results which are too extensive to depict graphically for all countries, hence are shown only for a set of countries. Table 2 provides an overview of H and MPI from the 2019 round of estimation for the set of selected countries, which vary by region and poverty level.[7] Since subnational data is heavily used, the number of available regions per country are shown as well.

3.3 Parameter variations

We explore the sensitivity of the global MPI with respect to a) the poverty cutoff; b) the weighting structures; and c) the inclusion or omission of indicators. For the poverty cutoff, we implement 10 distinct values, including union ($k = 1\%$) and intersection ($k = 100\%$). Following Alkire *et al.* (2019a), we consider the range of plausible values to be $k = 20\%$ to $k = 50\%$. For the deprivation values or indicator weights, we implement 231 alternative weighting schemes, which include dropping one and even two dimensions (via assigning a weight of zero). More specifically, we consider for

[6] The global MPI table on subnational regions (Table 5) only disaggregates 1,119 subnational regions because it does not disaggregate low-poverty countries; this paper, in contrast, disaggregates all possible countries, which total 98 countries including low-poverty countries.

[7] Selected countries may be changed in subsequent versions of this paper. Suggestions are welcome.

Table 1: The global MPI

Dimension	Indicator	Deprived if ...	Weight
Health	Nutrition	Any person under 70 years of age for whom there is nutritional information is *undernourished*.	$\frac{1}{6}$
	Child mortality	A child under 18 years of age has died in the household in the five-year period preceding the survey.	$\frac{1}{6}$
Education	Years of schooling	No household member aged 10 years or older has completed *six years* of schooling.	$\frac{1}{6}$
	School attendance	Any school-aged child is not attending school up to the age at which he/she would complete *class 8*.	$\frac{1}{6}$
Living Standards	Cooking fuel	A household cooks with dung, agricultural crop, shrubs, wood, charcoal or coal.	$\frac{1}{18}$
	Sanitation	The household's *sanitation facility* is *not improved* (according to SDG guidelines) or it is *improved but shared* with other households.	$\frac{1}{18}$
	Drinking water	The household does *not have access* to *improved drinking water* (according to SDG guidelines) or safe drinking water is at least a *30-minute walk* (roundtrip) from home.	$\frac{1}{18}$
	Electricity	The household has *no electricity*.	$\frac{1}{18}$
	Housing	The household has *inadequate housing*: the floor is of natural materials or the roof or walls are of natural or rudimentary materials.	$\frac{1}{18}$
	Assets	The household does *not own more than one* of these *assets*: radio, TV, telephone, computer, animal cart, bicycle, motorbike, or refrigerator, and does not own a car or truck.	$\frac{1}{18}$

Notes: More details on the indicators can be found in Alkire *et al.* (2019b).

Table 2: Selected results of the global MPI

Country name	ISO code	survey	year	MPI	H	# regions
Thailand	THA	MICS	2015-2016	0.003	0.79	5
Algeria	DZA	MICS	2012-2013	0.008	2.10	7
India	IND	DHS	2015-2016	0.123	27.91	36
Ghana	GHA	DHS	2014	0.138	30.07	10
Bangladesh	BGD	DHS	2014	0.198	41.70	7
Senegal	SEN	DHS	2017	0.288	53.17	14
Mozambique	MOZ	DHS	2011	0.411	72.45	11
Niger	NER	DHS	2012	0.590	90.47	8

each dimension all weighting schemes assigning values from 0–100% in increments of 5 percentage points. Following Alkire and Santos (2014), we consider the range of plausible weights to be between 25–50% for each dimension. Thus we consider 21 out of our 231 weighting structures to be 'plausible'.

In terms of inclusion, this paper implements six alternative specifications in which one living standard indicator at a time is removed and the other five indicators re-weighted.[8] This choice allows an analyses for all countries and can be thought of as a lower bound for indicator related effects. Importantly, most of the analyses explore the effect when only a single parameter is modified. This limitation is relaxed in section 5, where we assess the sensitivity of results with respect to variation in a third parameter–a second-order sensitivity analyses so to say.

3.4 The proportion of robust pairwise comparisons

To assess subnational rankings in poverty, we follow Alkire and Santos (2014) (cf. Alkire *et al.*, 2015) in drawing on pairwise comparisons, which allows us to account for sampling errors. While other studies have focused mainly on pairwise comparisons across countries, the principal ingredient for this analysis are pairwise comparisons between two subnational regions. Using statistical tests (which account for a potentially non-zero covariance), we can conclude for any pairwise comparison of two subgroups whether poverty is higher in one region than in the other, or whether poverty in both regions is not significantly different. Thus, given a country with $g = 1, \ldots, G$ subnational regions, we obtain a value for each pairwise comparison between two subnational regions g and h in terms of a particular poverty measure P

$$O_P^\theta(g,h) = \begin{cases} 1 & \text{if } P_g^\theta > P_h^\theta \\ 0 & \text{if } P_g^\theta = P_h^\theta \\ -1 & \text{if } P_g^\theta < P_h^\theta \end{cases} \tag{1}$$

where θ summarises the underlying parameter choices used to specify P. Alkire and Santos (2014) seek to assess the robustness of a pairwise ordering under several parametrizations simultaneously. This paper, instead, considers a pairwise comparisons for two regions to be robust, if the same ordering for a given poverty measure is observed under the reference parametrization and *one* alternative parametrization, i.e. θ and θ', respectively. Finally, we count pairwise comparisons robust across reference and alternative parametrization and define the share of robust pairwise comparisons R_P, more succinctly, as the number of robust pairwise comparison relative to the total possible pairwise comparisons:

$$R_P = \frac{\sum_{1 \le g < h \le 1}^{G} \mathbb{I}\left[O_P^\theta(g,h) = O_P^{\theta'}(g,h)\right]}{0.5G(G-1)} \tag{2}$$

[8]Some countries lack already one health indicator, so dropping the other one would result in discarding an entire dimension and create difficulties in interpretation.

where $\mathbb{I}(\cdot)$ is the indicator function taking the value of 1 if a poverty ordering between regions g and h is identical under both parametrizations, and 0 if not. The advantage of the present approach is that with incremental parameter changes thresholds where pairwise comparisons begin to change can be precisely identified. The normalisation with the maximum possible pairwise comparisons allows cross-country comparability and provides a useful intuition. More specifically, as the $R_P \in [0,1]$ and its value is 1, if the baseline is compared with itself, one can read the R_P as the percentage to which the original ordering is retained under the alternative parametrization. While the investigation of pairwise comparisons across subnational regions for a given country has an arguable and immediate intuition and relevance, it must be noted that its interpretation is complex for several reasons. First, the number of subnational regions in a country, and their relative population shares, vary, and this variety itself affects the analysis and its policy salience. For example, the change of a single pairwise comparisons results in stronger reductions of R_P for countries with fewer regions. For instance, the R_P will fall by 0.1 for a country with 5 regions, and by 0.02 or 0.0013 for a country with 10 or 20 regions. Therefore, sharper decreases are to be expected for countries with fewer subnational regions. Second the population shares of regions may vary greatly. It may be that one country has five regions with 20% of the population each, whereas in one country 50% of the population is in one region, 35% in the second region, and 5% in each of the last two regions. Population shares affect many things: it may be that the standard errors of the regions differ; it may also be that the policy salience of certain switches vary.

Third, a meaningful interpretation requires context. A very robust subnational ranking may, for instance, emerge from grave subnational disparities, which are more likely to be preserved under alternative parametrizations. Alternatively, high robustness may also be observed, simply because all regions are completely poor or non-poor. A prerequisite for a more informative robustness analyses is thus that the underlying measure according to which the regions are ranked possess sufficient discriminatory power (i.e. its value are not to close to 0 or 1).

3.5 The Jaccard-coefficient

To assess the extent to which the poverty sets between the baseline measure and the measure using an alternative parameter vary, we will use the Jaccard-coefficient (Jaccard 1901; cf Sneath 1957). Recall that the poverty set, denoted as $\mathbb{P}$, of a poverty measure is the set of individuals who are identified as poor, because their deprivation score is greater than or equal to the poverty cutoff.

Naturally, the size of $\mathbb{P}$ (i.e. H) may change upon some parameter variation in θ. $\mathbb{P}$ may, however, change even if H remains the same. That is, two measures may each identify 20% of the population as poor, but the same people may not be identified as poor in both cases. This is quite important, as we shall see. Within the frame of sensitivity analyses, we seek to quantify to which extent poverty sets change upon modest parameter variations compared to our baseline specification. There are many measures of similarity that could be used (Alkire *et al.*, 2015, ch. 10). The Jaccard-coefficient offers one instructive way to quantify the similarity of two sets:

$$J = \frac{|\mathbb{P}(\theta) \cap \mathbb{P}(\theta')|}{|\mathbb{P}(\theta) \cup \mathbb{P}(\theta')|} \tag{3}$$

Intuitively, J is the number of persons identified as poor according to both poverty measures, relative to the number of persons identified as poor by at least one of the poverty measures. The Jaccard-coefficient equals 1 only if both poverty sets are exactly the same. However, J declines if the headcount ratio for an alternative parametrizations would change, even while the jointly identified poor units remain exactly the same. Moreover, J also declines for a decreasing overlap, while both headcount ratios remain equal. The subsequent analyses will explore to which extent alternative parametrization of poverty measures (in particular alternative weighting schemes and indicator selections) result in poverty sets similar to the reference choice. Note, that for the poverty cutoff k this is a trivial exercise, which may explain why poverty sets have been neglected in empirical analyses sofar. Specifically, J can be inferred from the simple change in H: for $k_1, \leq k_2$, $J(k_1, k_2) = \frac{H(k_1)}{H(k_2)}$. Finally, similar to the poverty measure itself, J loses its practicability once almost everybody or nobody is identified as poor in a particular society. In our empirical application the estimation of this proportion takes survey weights into account.

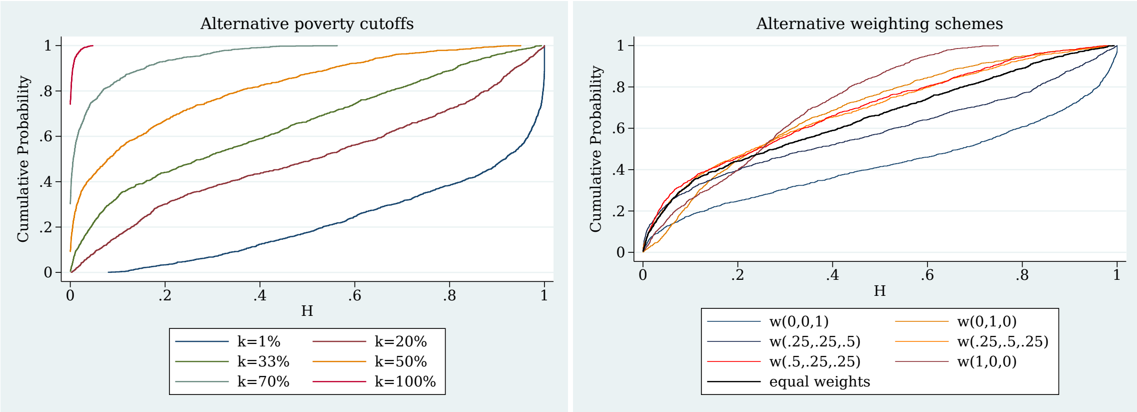
4 Results

4.1 Simple and adjusted headcount ratios

A convenient tool to initially explore the effects of parameter modifications is the cumulative distribution function (cdf): it describes the distribution in sufficient detail such that potential issues like excessive bounded values can be detected. Excessive values of 0 or 1 are undesirable as a poverty measure essentially becomes impracticable: it loses its discriminatory power and progress or failure in fighting poverty can no longer be adequately monitored.Moreover, cdfs introduce anonymity, i.e. country and region names are omitted, thereby, allowing an efficient inspection of several parametrizations simultaneously.

Figure 1 shows cdfs of all subnational regions in all countries for both six alternative poverty cutoffs and seven weighting schemes and highlights that $H\left(k = \frac{1}{3}\right)$ more or less exhausts the value range, what is sensible for a cross-country poverty measure. Moreover, $H\left(k = \frac{1}{5}\right)$ and $H\left(k = \frac{1}{2}\right)$ result in higher (lower) headcount ratios, without manifesting in excessive boundary values of 0 and 1. A union cutoff ($k = 1\%$) however implies a headcount ratio of 100% for 11% of the subnational regions, more than 90% for more than 50% of the regions, and more than 50% for 80% of the regions. Conversely, an intersection identification results in zero poverty for 75% of the regions. Thus, given the current indicator set, both union and intersection cutoffs render the simple headcount ratio at least partially impracticable. For alternative weighting schemes, figure 1 shows that assigning higher weights to the living standard dimensions, implies a downward shift of the cdf and thus, by tendency, higher H. Note, however, that even very unequal extreme weighting schemes, which effectively

Figure 1: Cumulative distribution functions the simple headcount ratio.



Notes: Underlying data from subnational level. Left panel only varies poverty cutoffs, underlying weights follow equal-nested structure. Right panel only varies weights, while underling poverty cutoff is one third. For instance, $w(.5,.25,.25)$ assigns a weight of 50% to health, and 25% to each schooling and living standards, while indicators inside a dimension are equally weighted.

only retain a single dimension, do not induce excessive boundary values.

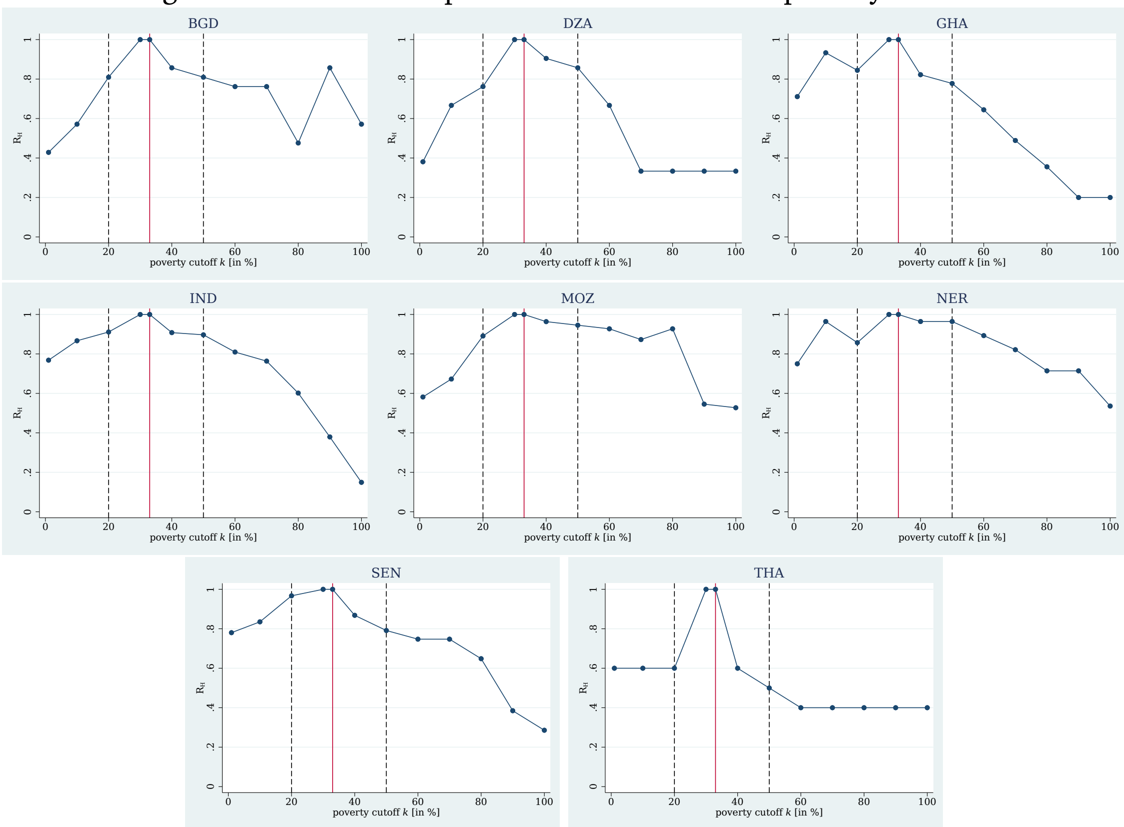
4.2 Subnational rankings

Alternative poverty cutoffs. To which extent do subnational orderings in terms of headcount ratios change upon varying the poverty cutoff? Figure 2 shows percent of robust pairwise comparisons in comparison with the global MPI baseline specifications, for eight countries all k cutoffs ranging from 0–100%, in terms of the headcount ratio. What we see is that patterns of robustness vary across country, with generally higher robustness in the poorest countries, and generally a hump-shaped pattern with the lowest robustness at intersection. Thus, subnational poverty orderings tend to increasingly change with more distant poverty cutoffs. Moreover, for $k \in \left[\frac{1}{5}, \frac{1}{2}\right]$ we observe $R_H > 0.7$ for most countries.[9] Additionally, no country is found to be entirely insensitive to changes in the poverty cutoff and nine countries which do display pretty insensitive responses all have 6 or less subnational regions (except Jordan). Finally, under union identification 75% of the countries exhibit a share of robust pairwise comparisons with the reference specification of at most 75% and often in fact much less.

Alternative weighting schemes. How strong do subnational orderings change if alternative weighting schemes are adopted? Figure 3 illustrates the share of robust pairwise comparisons, where each point within the simplex unambiguously describes the underlying weighting scheme. The tiny triangle indicates the 'plausible' ranges of weights, and the black dot indicates the global MPI baseline weights. Two observations are salient: first several countries exhibit mostly very high values of .95 or more for the weighting schemes within the contracted simplex (DZA, IND, MOZ, SEN), while other countries display still high values of .85 or more (BGD, GHA, NER). A few countries

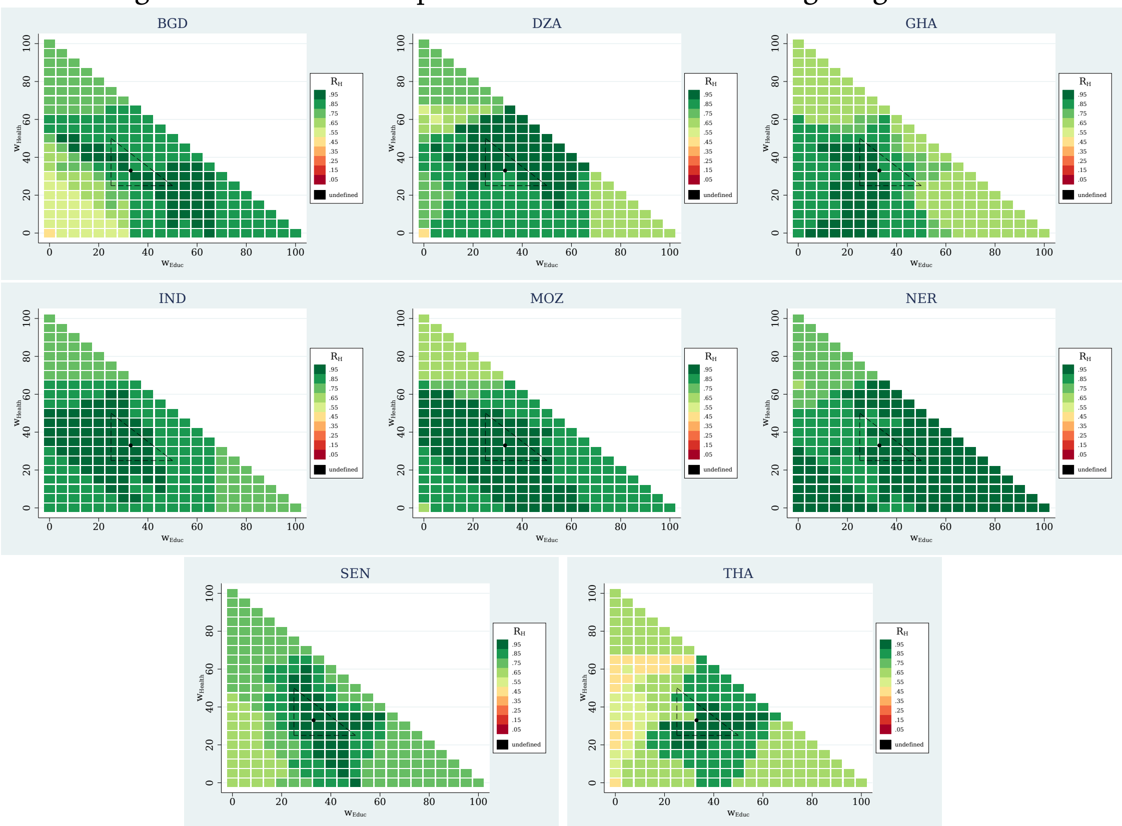
[9] Smaller values emerge in some cases, in particular for countries with very few subnational regions.

Figure 2: Pairwise comparisons with alternative poverty cutoffs.



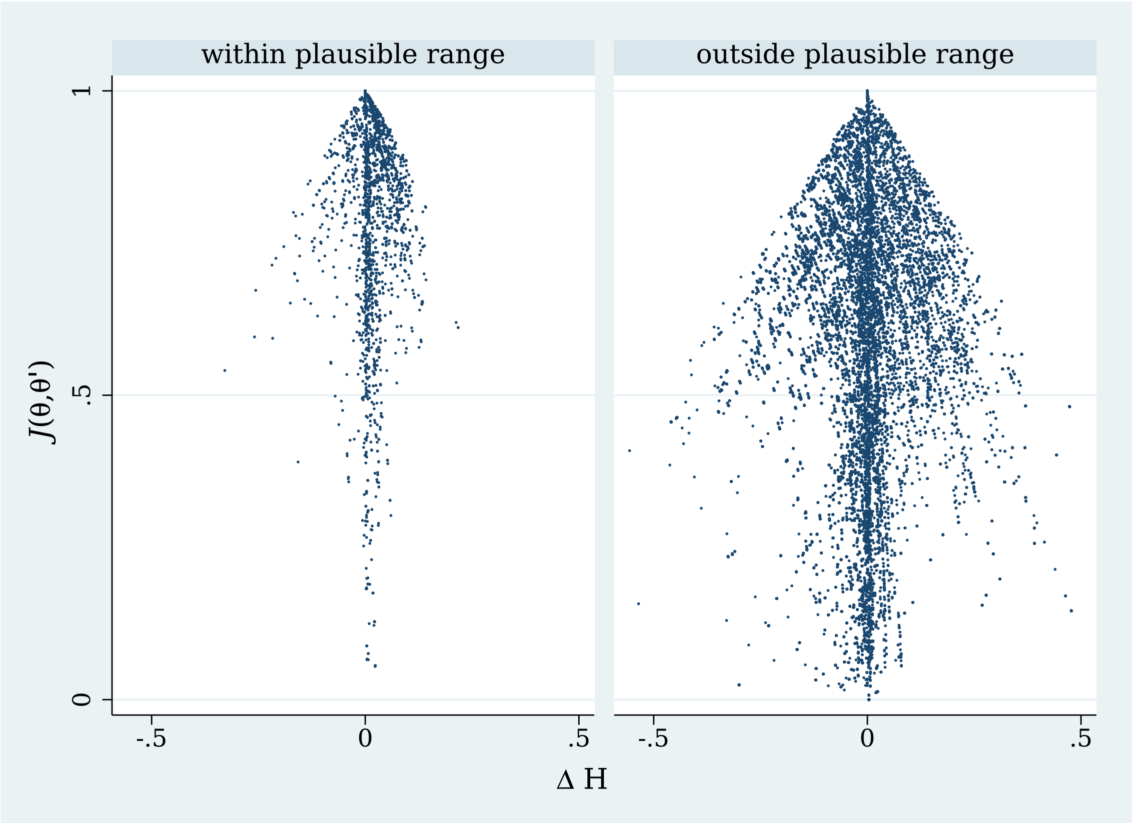
Notes: Pairwise comparisons are performed for subnational regions in terms of the simple headcount ratio.

Figure 3: Pairwise comparisons with alternative weighting schemes.



Notes: Pairwise comparisons are performed for subnational regions in terms of the simple headcount ratio. Dot in the centre indicates equal weighting structure, small black contracted simplex indicates plausible weighting schemes.

Figure 4: Jaccard coefficient and change in headcount ratio for alternative weights



Notes: Analysis of national level data for alternative weighting schemes; plausible weights assign a value between 25–50% to every dimensions (i.e. health, education, and living standards).

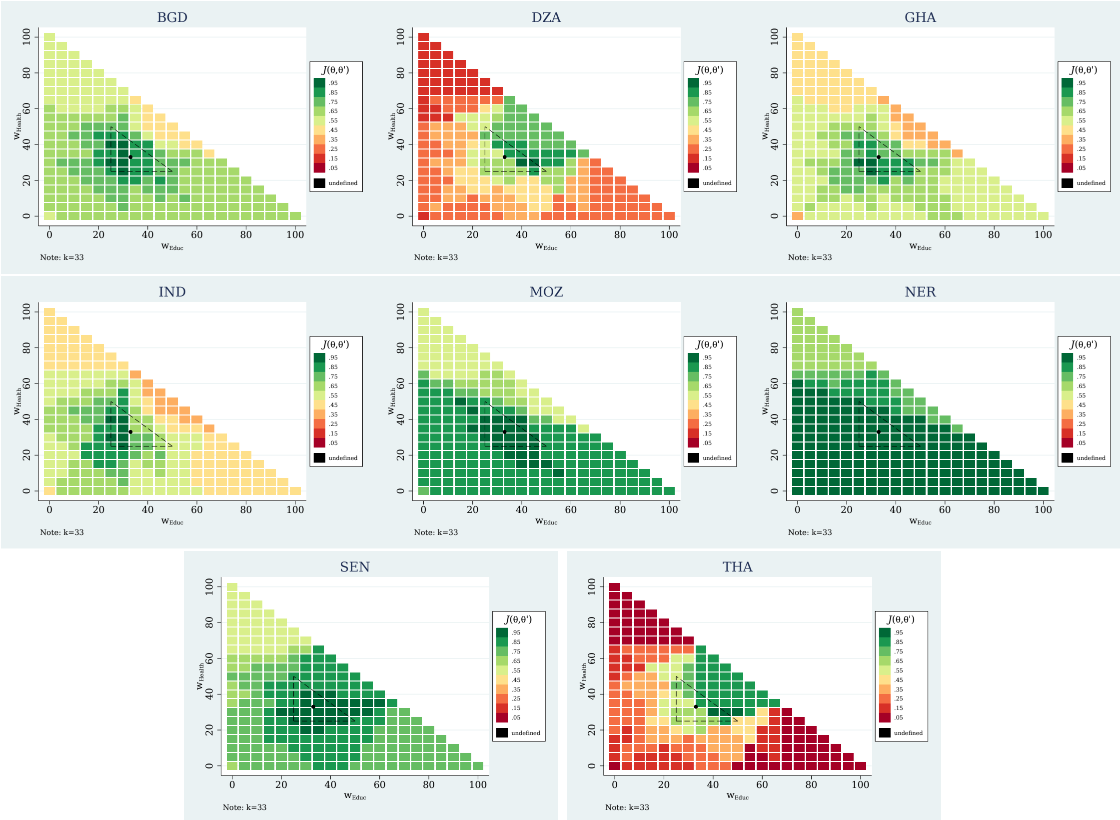
(e.g., THA) contain lower values of around 0.65. Second, the evidence, in general, suggests high values within simplex (and sometime beyond), but significant decreases can occur in particular for more unequal weighting schemes.[10]

4.3 Poverty sets

Alternative weighting schemes. A change in the overlap of poverty sets can occur without a change in the Headcount ratio. Even for equal headcount ratios—the number of persons identified as poor is identical—the set of persons identified as poor can be quite different. So to what extent do we identify decreases in overlap that are not indicated by changes in the H. Figure 4 plots the Jaccard-coefficient for against the associated change in the headcount ratio for alternative weighting schemes. Each dot represents a comparison for one country with an alternative weighting scheme against our baseline specification. On the left are dots 21 weighting schemes × 101 countries, whereas on the right are 210 weighting schemes × 101 countries. The dots on the 'zero' axis indicate that the headcount ratio for the alternative weight is the same as the baseline weight. But if the Jaccard is not 1, then different people are being identified as poor. Inspection reveals that it is indeed common to observe the

[10]For some countries we find under rather unequal weighting structures a share of robust pairwise comparisons of around 35%.

Figure 5: Comparison of poverty sets for alternative weighting schemes.



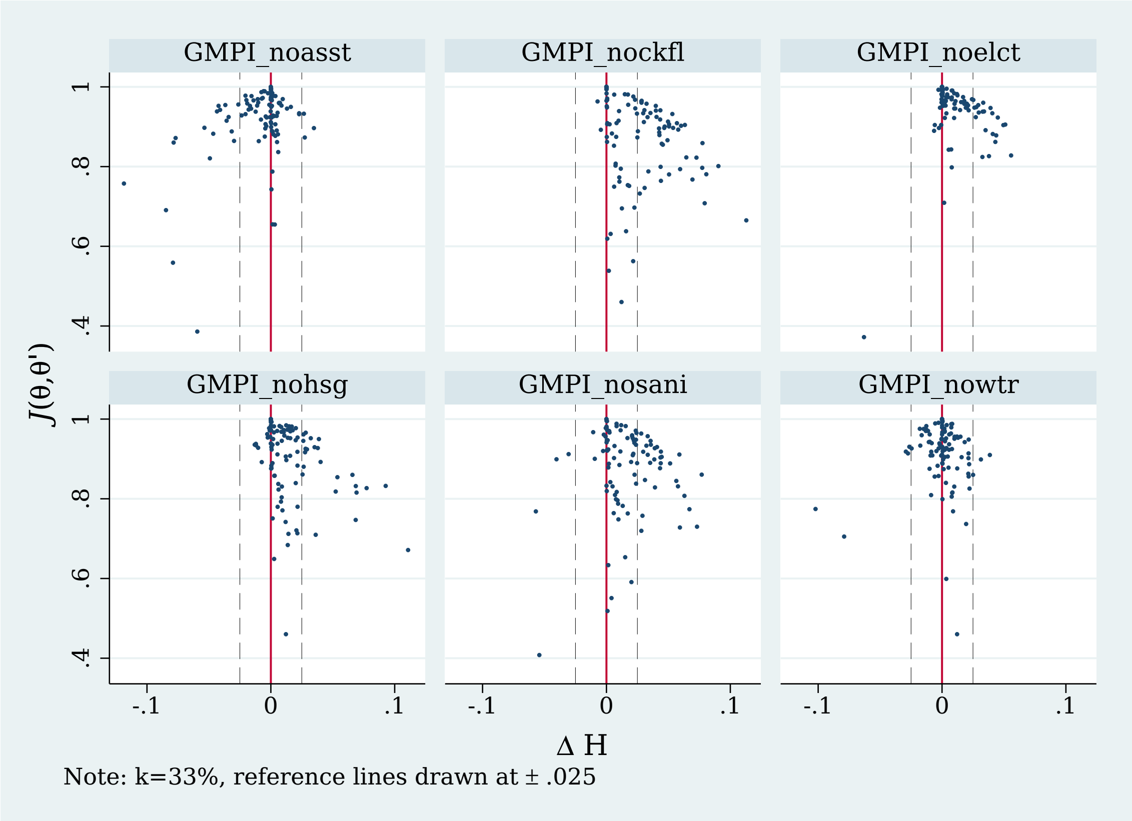
Notes: The point in centre of the graph indicates the equal weighting scheme, the dashed contracted simplex covers all dimensional weighting schemes, which assign a weight between 25–50% to every dimension, which we call a plausible weighting scheme; J is undefined if *both* poverty sets are empty.

Jaccard-coefficient J to drop significantly without being driven by notable changes in the headcount ratio. Importantly, this finding also holds for the more interesting subset plausible weighting schemes, indicating that a dedicated poverty set analysis is advisable.

Figure 5 provides more detail for selected countries on how the overlap in poverty sets changes if a slightly modified weighting schemes is adopted.[11] Several observations stand out: First, most countries have values of 0.65+, and often in fact more 0.85, for weighting schemes within the 50–25–25 ranges. Around 10% of country-weighting scheme combinations have J of 0.55 or less for plausible weighting schemes (distributed over 46 countries). Second, high poverty countries (e.g., SEN, MOZ, NER) tend to have overlap values of 95% within plausible (and, e.g., NER also for very unequal) weighting schemes. Essentially, this finding merely mirrors the impracticability of headcount ratios equalling 1: if everybody is poor anyway, weights do not matter any more. In low poverty countries, like DZA or THA, more unequal weighting schemes tend to result in dramatic reductions of J to 0.05–0.15 implying that actually almost entirely different households are identified as poor. Finally, that declines in the overlap do not necessarily indicate a problem, but simply call for complementary

[11] More results will be added in a subsequent version of the paper. Additionally, an online appendix will be provided.

Figure 6: Jaccard coefficient for alternative indicator selections.



Notes: Jaccard coefficient calculated for national level for 101 countries; each sub-graph contains one dot per country; remaining living standard indicators are re-weighted to sum-up to one third.

evidence. In India, for instance, plausible weighting schemes allocating a weight of 45% to living standards result in J-values of 0.70. In fact, the higher weight for living standards increases H as previously just non-poor, but in particular in living standards deprived households now become poor (while few of the previously people became non-poor). The normatively critical question policy makers have to face is whether these in particular in living standard-deprived households should be considered poor or not. Thus, the previous analysis shows that this decision has relevant implications.

Alternative indicator selections. Alternative indicator selections, which result in substantially different headcount ratios, could be expected to reduce the overlap in poverty sets. Figure 6 presents that the Jaccard coefficient for instances in which one living standard indicator is dropped (and the other living standard indicators are re-weighted). However we find that in many cases, the headcount ratio changes can be minor, but the set of poor often vary quite a lot. Indicator drops often result in Jaccard values of 0.6–0.8. These changes potentially could be compared with other parametric variations. Dropping in an indicator results in changes of Jaccard that are of a magnitude, which in the case of alternative weighting schemes are usually only achieved for more unequal weighting schemes. So a single indicator drop can shift poverty sets as much as a major change in indicator weights. While the discussions of robustness have been driven by concerns regarding weights, it seems that indicator decisions should perhaps receive more attention. Finally, more frequent reductions in the overlap of poverty sets seem to emerge from removing cooking fuel, housing, and sanitation.

5 Second-order sensitivity analyses

This section explores to which extent the previous sensitivity results depend upon so far fixed parameters like the poverty cutoff k. More specifically, this section first examines the role of the poverty cutoff in sensitivity analyses of indicator selections, in particular with respect to levels in H and subnational rankings.[12] Moreover, this section also illuminates the role of the poverty cutoff in the sensitivity analysis of poverty sets to alternative weighting schemes.

5.1 The poverty cutoff and the sensitivity the indicator selections.

Figure 7 (a), plots the absolute change with respect to the original headcount ratio for all 101 countries if one living standard indicator is dropped. In the right hand panel, the other five living standard indicators are re-weighted to one-third. The blue dots refer to a situation in which the original headcount ratio uses the union cutoff, and the red dots refer to the global MPI cutoff of one-third.

Two observations are salient. First the headcount ratio can drop up to 10%-points if a single living standard indicator is removed, irrespective of the poverty cutoff, if the original headcount ratio exceeds 10%. Second, however, drops of 10–30 percentage points or higher are only observed under a union identifications. Importantly, this finding is not driven by weight-adjustments inherent to indicator removals under non-union cutoffs.[13]

How do subnational orderings within each country that can be disaggregated (98 countries) respond to changes in the indicator set under different poverty cutoffs? Figure 7 (b) plots the percent of pairwise comparisons between subnational regions of a country that are robust against the original headcount ratio of that country, when a particular living standard indicator (the color of the dot), is removed. The figure shows that 80% or more of the subnational ordering at baseline are commonly retained, if a single living standard indicator is dropped, independent of the poverty cutoff. Additionally, however, larger reductions in pairwise comparisons are more commonly observed under union identification. Finally, lower values of pairwise comparisons R_H can originate from different indicator drops, but are mostly associated with removing drinking water, cooking fuel, or housing (and occasionally sanitation).

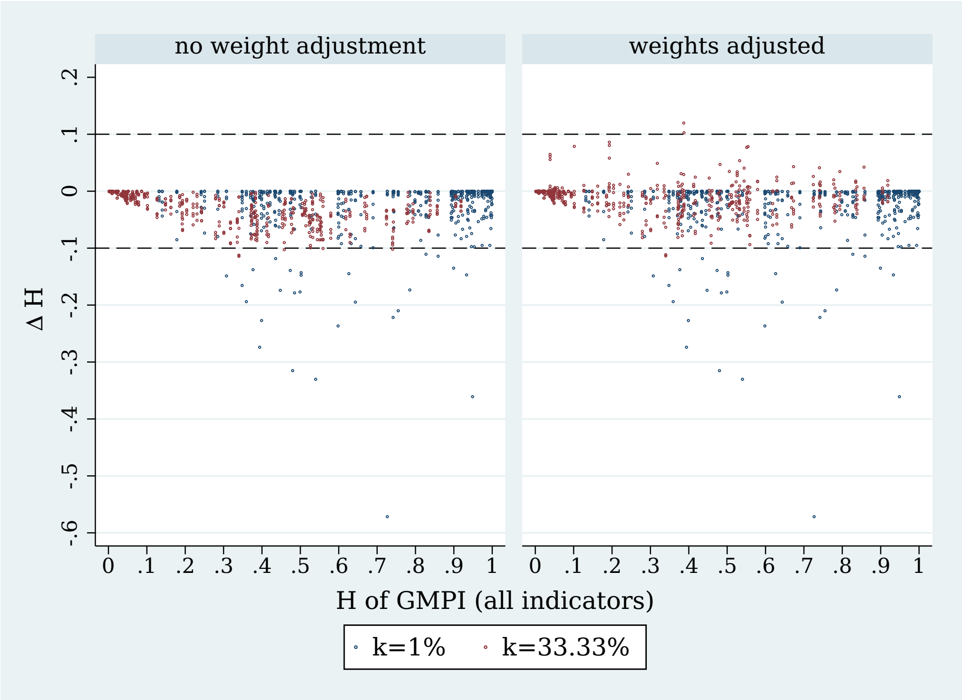
Since the previous results indicate that removing a single living standard indicator may affect both headcount ratios and the subnational ordering, a natural follow-up

[12]The analysis of poverty sets in this setting is somewhat more demanding as different poverty cutoffs usually imply quite different incidences, which are however key to rationalise findings of the suggested poverty sets comparisons.

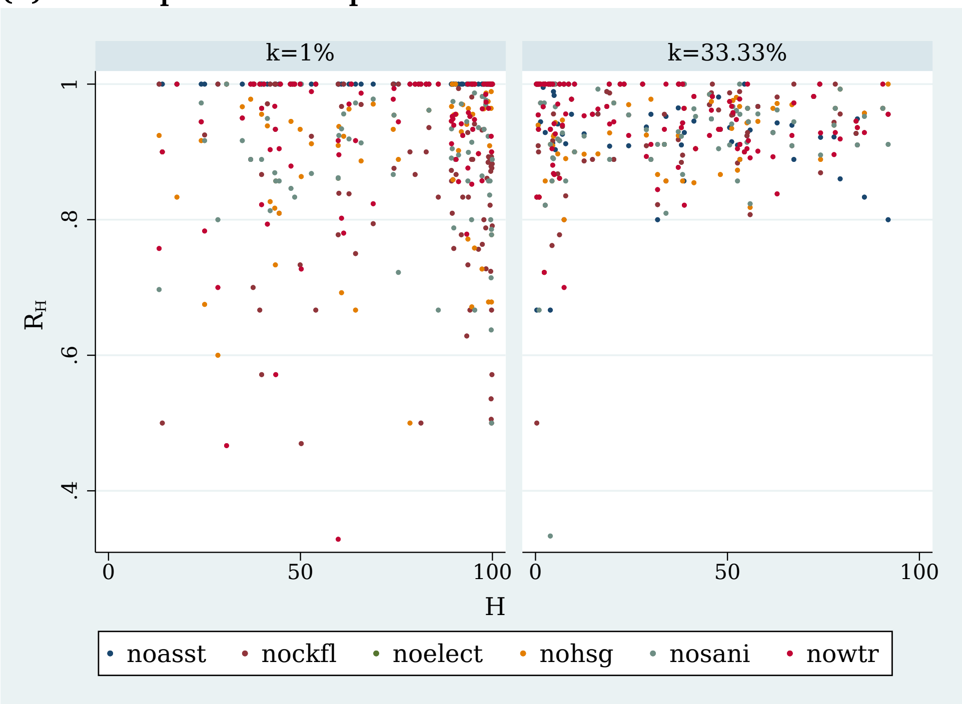
[13]Removing a single indicator within an equal-nested weighting scheme implies the remaining indicators of that dimension to receive a slightly higher weight.

Figure 7: Sensitivity to indicator selection by poverty cutoff

(a) Incidence

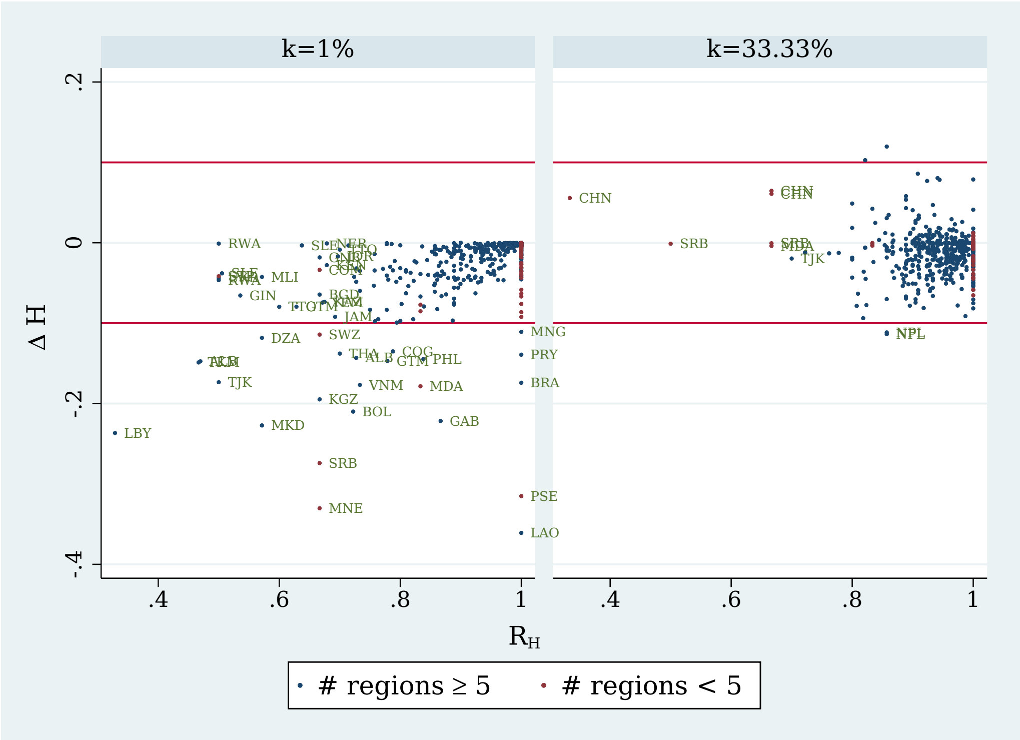


(b) Robust pairwise comparisons



Notes: Each observation represent a particular country-indicator set combination; six alternative indicator selections are covered. Panel (a) is based on 101 countries, panel (b) on 98. Headcount ratio on x-axis refers to specification without indicator drop and to $k = 1\%$ and $k = 33\%$, respectively.

Figure 8: Incidence and pairwise comparisons for indicator selections by poverty cutoff



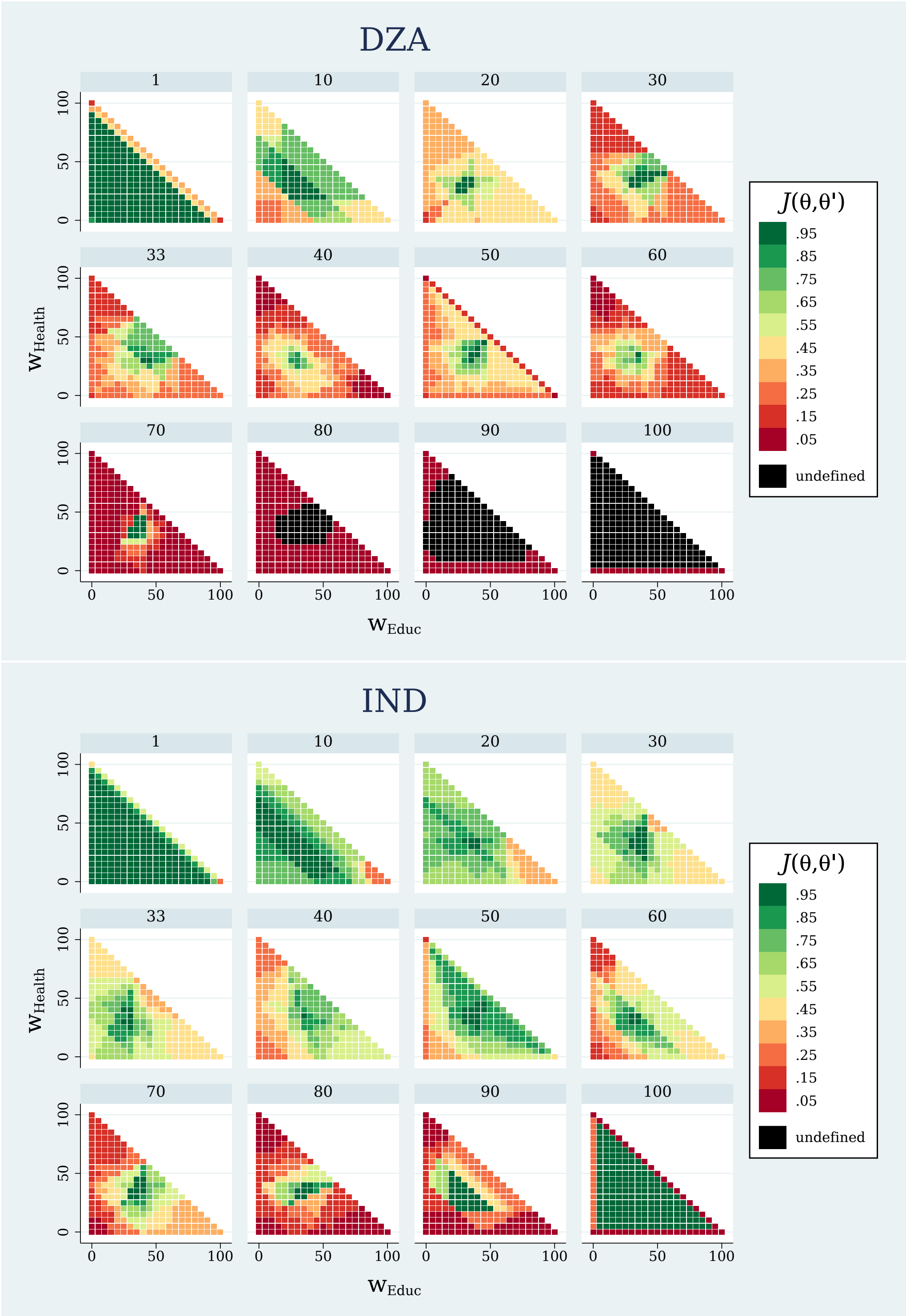
Notes: Analysis for 98 countries; each panel contains 98 × 6 dots, where each dot represents a particular country-indicator set combination, ΔH refers to national difference in the headcount ratio between with and without drop of single living standard indicator under respective k.

question is actually the phenomenon is observed just from different perspectives. Figure 8 also considers 98 countries, and drops one living standard indicator at a time (so each country has six dots), re-weights the indicators, and computes the alternative H. It then subtracts the alternative headcount ratio (considering one indicator dropped) from the original (baseline and union) headcount ratio. It then plots ΔH against R_H and distinguishes the countries into two colours, according to the number of subnational regions they have. The figure shows that in principal, drops in pairwise comparisons and headcount ratios can coincide. More importantly, however, changes in subnational rankings may well occur without larger changes in incidence, and vice versa. Therefore, separate analyses are advisable. Additionally, the figure also reveals that countries with less than 5 regions actually either have nearly 100% of robustness or have quite high account for the drops in pairwise comparisons, and this is a function of having fewer possible pairwise comparisons.

5.2 Poverty cutoff in sensitivity analyses of poverty sets to weights

Section 4 documents that plausible alternative weighting schemes poverty, by and large, tend to identify the same people as poor. Does this result depend on the choice of our reference cutoff $k = \frac{1}{3}$? How would the overlap in poverty sets change with alternative weighting schemes if a different poverty cutoff had been chosen in the first place? Note that for answering this question the poverty cutoff of both measures is

Figure 9: Poverty set analysis for alternative weighting schemes by poverty cutoff



Notes: Denoted poverty cutoff applies to both reference and alternative specification. Each square represents a particular weighting scheme. J is undefined if *both* poverty sets are empty.

changed the same way. Figure 9 shows the Jaccard-coefficient for the 231 different weighting schemes by poverty cutoffs ranging from union to intersection for two countries. First note that for union and intersection identification, figure 9 reflects the theoretically well-known result that weights are irrelevant—unless assigning zero weights is allowed. Effectively dropping one or more entire dimensions, then frequently results in a substantial decrease of J to 0.15 and less. Second, the area around the plausible triangle of equal weights with relative high overlap in poverty sets tends to decrease with k, implying that fewer weighting schemes leave the overlap in poverty sets essentially unaffected. Third, the magnitude of the decline in J tends to increase with k, indicating a decreased robustness to alternative weighting schemes.

6 Concluding Remarks

First, in measurement exercises robustness of outcomes to local parametric variation is desirable. Non-robustness does, however, not automatically disqualify a measure, but essentially represents normatively relevant information. Therefore, sensitivity analyses are best viewed as an integral part of the process determining the parameters as they allow better informed decisions in the first place.

Second, this paper documents the GMPI to be rather insensitive to local parameter variations in most instances of application. More specifically, subnational rankings in terms of simple headcount ratios are, for instance, found to be quite similar with the ranking under the preferred poverty cutoff ($k = \frac{1}{3}$). A similar pattern is observed for plausible alternative weighting schemes compared against equal weights. However, this paper, also provides evidence that more substantial changes in the poverty cutoff or the weighting schemes do have relevant implications, e.g., for subnational rankings.

Third, poverty sets are an important outcome for sensitivity analyses, in addition to the subnational rankings or time trends, among others. For the global MPI we find the poverty sets to be largely congruent for plausible alternative weighting schemes. We do, however, also document for the majority of countries sharp decreases of this congruence once more unequal weighting schemes are applied, in particular when one or more dimensions are effectively dropped. Importantly, our results also indicate that in many instances poverty sets do change significantly, without being driven by simple changes in the incidence. Therefore, a distinct poverty set analysis is an advisable sensitivity check.

Fourth, the present paper finds that removing a single living standard indicator can in fact have a similar influence in terms of subnational rankings or poverty sets as local changes of the poverty cutoff or the weighting schemes. Importantly, since living standard indicators receive a lower weight than other indicators, these results are best conceived as a lower bound. Since indicator construction and selection is evidently critical for multidimensional poverty measures it should, therefore, receive more academic attention.

Finally, the second-order sensitivity analysis highlights that design decisions in crafting a poverty measure are highly interdependent. Among other things, the results indicate that a union cutoff may entail a higher sensitivity of both the national simple headcount ratio and the implied subnational rankings, to slightly modified indicator selections.

References

Ackland, R., Dowrick, S., and Freyens, B., 2013. Measuring global poverty: Why PPP methods matter, *Review of Economics and Statistics*, 95 (3), 813–824.

Alkire, S., Ballon, P., Foster, J., Roche, J.M., Santos, M.E., and Seth, S., 2015. *Multidimensional Poverty Measurement and Analysis: A Counting Approach*, Oxford: Oxford University Press.

Alkire, S. and Foster, J., 2011. Counting and multidimensional poverty measurement, *Journal of Public Economics*, 95 (7-8), 476–487.

Alkire, S., Kanagaratnam, U., Nogales, R., and Suppa, N., 2019a. The first revision of the global MPI: Empirical insights and robustness, mimeo, Oxford Poverty and Human Development Initiative (OPHI), Oxford University.

Alkire, S., Kanagaratnam, U., and Suppa, N., 2018. The global multidimensional poverty index (MPI): 2018 revision, OPHI MPI Methodological Notes 46, Oxford Poverty and Human Development Initiative, University of Oxford.

Alkire, S., Kanagaratnam, U., and Suppa, N., 2019b. The global multidimensional poverty index (MPI): 2019, OPHI MPI Methodological Notes 47, Oxford Poverty and Human Development Initiative, University of Oxford.

Alkire, S., Roche, J.M., and Vaz, A., 2017. Changes over time in multidimensional poverty: Methodology and results for 34 countries, *World Development*, 94, 232–249.

Alkire, S. and Santos, M.E., 2014. Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index, *World Development*, 59, 251–274.

Alkire, S. and Seth, S., 2013. A comparison of methods identifying BPL households, *Economic and Political Weekly*, 48 (2), 49–57.

Alkire, S. and Seth, S., 2015. Multidimensional poverty reduction in India between 1999 and 2006: Where and how?, *World Development*, 72, 93–108.

Atkinson, A.B., 1987. On the Measurement of Poverty, *Econometrica*, 55 (4), 749–64.

Atkinson, A.B., 1992. Measuring poverty and differences in family composition, *Economica*, 59 (233), 1.

Banks, J. and Johnson, P., 1994. Equivalence scale relativities revisited, *The Economic Journal*, 104 (425), 883–890.

Buhmann, B., Rainwater, L., Schmaus, G., and Smeeding, T.M., 1988. Equivalence scales, well-being, inequality, and poverty: Sensitivity estimates across ten countries using the luxembourg income study (LIS) database, *Review of Income and Wealth*, 34 (2), 115–142.

Burkhauser, R.V., Smeeding, T.M., and Merz, J., 1996. Relative inequality and poverty in Germany and the United States using alternative equivalence scales, *Review of Income and Wealth*, 42 (4), 381–400.

Chakravarty, S.R. and D'Ambrosio, C., 2006. The Measurement Of Social Exclusion, *Review of Income and Wealth*, 52 (3), 377–398.

Chen, S. and Ravallion, M., 2010. The developing world is poorer than we thought, but no less successful in the fight against poverty, *The Quarterly Journal of Economics*, 125 (4), 1577–1625.

Coulter, F.A.E., Cowell, F.A., and Jenkins, S.P., 1992. Scale relativities and the extent of inequality and poverty, *Economic Journal*, 102 (414), 1067–1082.

Datt, G., 2018. Distribution-sensitive multidimensional poverty measures, *The World Bank Economic Review*.

Davidson, R. and Duclos, J.Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality, *Econometrica*, 68 (6), 1435–1464.

De Vos, K. and Zaidi, M.A., 1997. Equivalence scale sensitivity of poverty statistics for the member states of the european community, *Review of Income and Wealth*, 43 (3), 319–333.

Deaton, A., 2010. Price indexes, inequality, and the measurement of world poverty, *American Economic Review*, 100 (1), 5–34.

Deaton, A. and Dupriez, O., 2011. Purchasing power parity exchange rates for the global poor, *American Economic Journal: Applied Economics*, 3 (2), 137–166.

Dhongde, S. and Minoiu, C., 2013. Global poverty estimates: A sensitivity analysis, *World Development*, 44, 1–13.

Dotter, C. and Klasen, S., 2014. The multidimensional poverty index: Achievements, conceptual and empirical issues, Occasional paper, UNDP Human Development Report Office, New York.

Duclos, J.Y. and Mercader-Prats, M., 1999. Household needs and poverty: With application to Spain and the U.K., *Review of Income and Wealth*, 45 (1), 77–98.

Duclos, J.Y., Sahn, D.E., and Younger, S.D., 2006. Robust multidimensional poverty comparisons, *Economic Journal*, 116 (514), 943–968.

Duclos, J.Y. and Tiberti, L., 2016. Multidimensional poverty indices: A critical assessment, *in:* M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*, Oxford University Press, 1602, chap. 23.

Espinoza-Delgado, J. and Klasen, S., 2018. Gender and multidimensional poverty in nicaragua: An individual based approach, *World Development*, 110, 466–491.

Foster, J., Greer, J., and Thorbecke, E., 1984. A class of decomposable poverty measures, *Econometrica*, 52 (3), 761–66.

Foster, J.E. and Shorrocks, A.F., 1988. Poverty orderings and welfare dominance, *Social Choice and Welfare*, 5 (2-3), 179–198.

Jenkins, S.P. and Cowell, F.A., 1994. Parametric equivalence scales and scale relativities, *The Economic Journal*, 104 (425), 891–900.

Jindra, C. and Vaz, A., 2019. Good governance and multidimensional poverty: A comparative analysis of 71 countries, *Governance*.

Kakwani, N. and Son, H.H., 2016. Global poverty estimates based on 2011 purchasing power parity: where should the new poverty line be drawn?, *The Journal of Economic Inequality*, 14 (2), 173–184.

Klasen, S., Krivobokova, T., Greb, F., Lahoti, R., Pasaribu, S.H., and Wiesenfarth, M., 2016. International income poverty measurement: which way now?, *The Journal of Economic Inequality*, 14 (2), 199–225.

Lasso de la Vega, C., 2010. Counting poverty orderings and deprivation curves, *in:* J. Bishop, ed., *Studies in Applied Welfare Analysis: Papers from the Third ECINEQ Meeting*, Bingley: Emerald Group Publishing Limited, *Research on Economic Inequality*, vol. 18, chap. 7, 153–172.

Meyer, B.D. and Sullivan, J.X., 2012. Identifying the disadvantaged: Official poverty, consumption poverty, and the new supplemental poverty measure, *Journal of Economic Perspectives*, 26 (3), 111–36.

Nolan, B. and Whelan, C.T., 2011. *Poverty and Deprivation*, Oxford University Press.

Ogutu, S.O. and Qaim, M., 2019. Commercialization of the small farm sector and multidimensional poverty, *World Development*, 114, 281–293.

Pasha, A., 2017. Regional perspectives on the multidimensional poverty index, *World Development*, 94, 268–285.

Pattanaik, P.K. and Xu, Y., 2018. On measuring multidimensional deprivation, *Journal of Economic Literature*, 56 (2), 657–672.

Ravallion, M., 2011. On multidimensional indices of poverty, *Journal of Economic Inequality*, 9 (2), 235–248.

Ravallion, M., 2012. Mashup indices of development, *The World Bank Research Observer*, 27 (1), 1–32.

Ravallion, M. and Bidani, B., 1994. How robust is a poverty profile?, *The World Bank Economic Review*, 8 (1), 75–102.

Ravallion, M. and Sen, B., 1996. When method matters: Monitoring poverty in bangladesh, *Economic Development and Cultural Change*, 44 (4), 761–792.

Santos, M.E. and Villatoro, P., 2018. A multidimensional poverty index for latin america, *Review of Income and Wealth*, 64 (1), 52–82.

Sen, A.K., 1976. Poverty: An ordinal approach to measurement, *Econometrica*, 44, 219–231.

Sen, A.K., 1999. *Development as Freedom*, Oxford: Oxford University Press.

Vijaya, R.M., Lahoti, R., and Swaminathan, H., 2014. Moving from the household to the individual: Multidimensional poverty analysis, *World Development*, 59 (C), 70–81.

Whelan, C.T., Layte, R., and Maître, B., 2004. Understanding the mismatch between income poverty and deprivation: A dynamic comparative analysis, *European Sociological Review*, 20 (4), 287–302.

World Bank, 2017. *Monitoring Global Poverty*, Herndon: World Bank Publications.

World Bank, 2018. *Poverty and Shared Prosperity 2018: Completing the Poverty Puzzle*, World Bank Publications.

Yalonetzky, G., 2014. Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables, *Social Choice and Welfare*, 43 (4), 773–807.