



Testing the Robustness of the Revised Multidimensional Poverty Index for Arab Countries



Shared Prosperity **Dignified Life**





Shared Prosperity **Dignified Life**



VISION

ESCWA, an innovative catalyst for a stable, just and flourishing Arab region

MISSION

Committed to the 2030 Agenda, ESCWA's passionate team produces innovative knowledge, fosters regional consensus and delivers transformational policy advice. Together, we work for a sustainable future for all.



Economic and Social Commission for Western Asia

Testing the Robustness of the Revised Multidimensional Poverty Index for Arab Countries



United Nations
Beirut

© 2021 United Nations
All rights reserved worldwide

Photocopies and reproductions of excerpts are allowed with proper credits.

All queries on rights and licenses, including subsidiary rights, should be addressed to the United Nations Economic and Social Commission for Western Asia (ESCWA), e-mail: publications-escwa@un.org

Author: Noha Omar and Sama El Hage Sleiman.

The findings, interpretations and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the United Nations or its officials or Member States.

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Links contained in this publication are provided for the convenience of the reader and are correct at the time of issue. The United Nations takes no responsibility for the continued accuracy of that information or for the content of any external website.

References have, wherever possible, been verified.

Mention of commercial names and products does not imply the endorsement of the United Nations.

References to dollars (\$) are to United States dollars, unless otherwise stated.

Symbols of United Nations documents are composed of capital letters combined with figures. Mention of such a symbol indicates a reference to a United Nations document.

United Nations publication issued by ESCWA, United Nations House, Riad El Solh Square,
P.O. Box: 11-8575, Beirut, Lebanon.

Website: www.unescwa.org

Contents

	<i>Page</i>
Introduction	1
I. Sources of Uncertainty	2
A. Aggregation methods	2
B. Weights for MPI	4
C. Poverty cut-off	5
II. External Robustness Process – How to Determine Robustness	6
References	13

Introduction

The construction of a multidimensional poverty index (MPI) involves crucial decisions: the choice and definition of indicators, deprivation cut-offs, weights and poverty cut-offs. Determining the values of these parameters affects the identification of the poor, the contributions of dimensions and indicators to MPI, and the ranking of countries. These decisions involve normative choices (Alkire and Jahan, 2018), but also empirical and statistical considerations. It is critical to use uncertainty analysis and sensitivity analysis to set the parameter values in constructing MPI (Saisana and others, 2005). The present section sets out the various tools for examining the robustness properties of the revised Arab MPI and its partial indices of the headcount ratio and the poverty intensity, as well as the sensitivity of countries' ordering when assumptions change. An ordering of entities (countries or population groups) is defined as robust when the order is unchanged as one of the parameter values changes. For regional policy relevance, the Arab MPI must be robust and consistent with various parameter specifications.

Various methods can be used to evaluate different kinds of ordering robustness. They can be categorized into two complementary strands: internal validity and external robustness checks. External robustness considers the effect of changes in parameters, such as weights, poverty cut-off, and aggregation on country rankings. Internal validity assesses the choice of indicators, the deprivation cut-offs within indicators, potential redundancy of information, and consistency and coherence of correlations among indicators within and between dimensions. It guides the choice of indicators regardless of country rankings, so it is internal to the index itself. By contrast, external robustness simulates and tests different weighting schemes in an attempt to study the ranking sensitivity to changes in weights; and it simulates and tests different poverty cut-offs for the same purpose. The present paper focuses on external robustness, while internal validity is discussed in a forthcoming companion technical paper.¹

¹ ESCWA, A revised Multidimensional Poverty Index for Arab countries, 2020.

I. Sources of Uncertainty

A. Aggregation methods

The literature on composite indices offers a range of aggregation techniques. The most commonly used are additive techniques that range from summing up country rankings across indicators to summing up weighted scores across indicators. Other methods have also been proposed, namely the multiplicative (or geometric) method and the non-linear aggregation method, such as the multi-criteria analysis or the cluster analysis.

Additive aggregation methods are the simplest and easiest methods to use and interpret. These methods are used to build an index by summing up the normalized values of its indicators. The weighted arithmetic mean is the commonly used additive method. It benefits from the continuity property, whereby the index's bounds can be accurately set if the relative measurement error of indicators is known. This property can also be used for sensitivity analysis (Pollesch and Dale, 2015).

However, these methods have disadvantages. Firstly, they require indicators to be preferentially independent. This means that indicators should not be correlated, which is unrealistic in various situations (Chen and Pu, 2004; Nardo and others, 2005). In the case of interdependent indicators, additive methods cannot be used. Secondly, these methods can be full-compensatory – in the sense that a positive change in one indicator can be fully compensated by an opposite change in another – and their weights reflect the rate of substitution among indicators. Thirdly, they require all indicators to be of the same measurement units and be normally distributed (Gan and others, 2017; Greco and others, 2019).

A less compensatory approach is the geometric aggregation method. The weighted geometric mean is a commonly used method under this approach. This method allows limited compensability between indicators owing to the inequality of geometric and arithmetic means that restrict the ability of indicators with very low scores to be fully compensated by high scores indicators (Bullen, 2013; Gan and others, 2017; Greco and others, 2019). Nevertheless, this aggregation approach also has some limitations. Firstly, since the geometric approach is preferentially dependent, it is not fully non-compensatory. It allows for trade-offs between indicators (OECD, 2008; Gan and others, 2017). Secondly, sensitivity analysis and uncertainty analysis cannot be tested using measurement errors of indicators (or input errors) computed using the geometric aggregation approach (Calvo and others, 2002).

For weights to reflect the importance of indicators and not their substitutability, non-compensatory aggregation methods should be used. When different indicators are aggregated, the loss in one indicator cannot be compensated by another indicator in the composite index. The multi-criteria

decision-making methods (MCDM) can account for the different criteria and preferences policymakers adopt in building a composite index like MPI.

The non-compensatory multi-criteria approach is an MCDM. Its main objective in constructing an index is to create rankings, rather than scores. Recently, this approach has been corrected for possible drawbacks to consist of two steps to compute an index: (1) rank all countries across all weighted indicators and construct an outscoring matrix through a pairwise comparison of bilaterally preferred societies; and (2) rank societies according to a complete pre-order, i.e., avoiding any cyclical irregularities in ranking decision (Gan and others, 2017). The output of this method is a rank and not a cardinal output value for each society. Another advantage is the absence of restrictions on the type of indicators used – both quantitative and qualitative data can be used. Possible drawbacks of this approach include the following. Firstly, the computational process becomes tedious with an increasing sample size or an increasing number of indicators. Secondly, the dependence of irrelevant alternative results, including possible rank non-transitivity (e.g., country A is preferred to B, B is preferred to C, but inconsistently country C is preferred to A with equal frequency of preferences). Thirdly, information on intensity of indicators is never used: regardless by how much one indicator is lower between two countries, they will preserve the same ranking (Munda and Nardo, 2005).

Alkire and Foster's aggregation technique

In estimating poverty, identification and aggregation need to be addressed (Sen, 1976). Identification means classifying people in a country or a population subgroup as poor or non-poor according to chosen indicators and their deprivation cut-offs. Aggregation refers to the process of building one summarizing statistic from all relevant available information on each individual in society, such as poverty incidence, poverty intensity and MPI (Alkire and Foster, 2009; and Alkire and others, 2015).

Alkire and Foster (2009) constructed their MPI with the aid of an aggregation method that builds on unidimensional poverty measures, and extended the poverty measures of Foster–Greer–Thorbecke (Foster and others, 1984). This aggregation method outperforms other methods because MPI acquires favourable properties, such as monotonicity and decomposability by indicator and by dimension, which are violated with other aggregation methods (Alkire and others, 2015).

According to Alkire and Foster's (2009) methodology, MPI (adjusted headcount ratio) is the mean of a censored vector of indicators using the poverty line:

$$MPI = \mu(c(k)) = \frac{1}{n} \times \sum_{i=1}^n c_i(k)$$

MPI can be perceived in terms of its partial indices, incidence of poverty and intensity of poverty, $MPI = H \times A$. H denotes the incidence of poverty that is the proportion of the population that is multi-dimensionally poor. $H = q/n$, where q is the number of persons classified as poor in a society

using the dual-cutoff approach – deprivation cut-off and poverty cut-off. A is the intensity of poverty that is the average deprivation score across the poor. The intensity of poverty equals $\sum_{i=1}^n c_i(k)/q$, where $c_i(k)$ refers to the proportion of deprivations experienced by a poor person i .

Thus, the adjusted headcount ratio is:

$$MPI = \mu(c(k)) = H \times A = \frac{q}{n} \times \frac{1}{q} \times \sum_{i=1}^q c_i(k) = \frac{1}{n} \times \sum_{i=1}^n c_i(k)$$

This index satisfies a number of favourable properties. It satisfies dimensional monotonicity; that is, if a poor person becomes deprived in one more dimension, intensity of poverty and consequently also MPI rises. The index also identifies the deprivations of the poor according to a dual poverty cut-off without a relative comparison (comparing to another person). Importantly, this index satisfies the break down decomposability by dimension and by indicator (Alkire and Foster, 2009).

Using the dual cut-off approach in identifying the multi-dimensionally poor, the index provides an intermediate approach between the union and the intersection identification methods. The union identification method classifies a person as multi-dimensionally poor if s/he is deprived in at least one indicator. The intersection method classifies a person as multi-dimensionally poor if s/he is deprived in all indicators simultaneously (UNDP and OPHI, 2019).

In this paper, we will not assess the robustness of the MPI aggregation method given its above-mentioned favourable properties, despite being a source of uncertainty. In MPI estimation, we adopt Alkire and Foster's aggregation technique at face value.

B. Weights for MPI

The choice of weights reveals the relative importance of various dimensions and indicators in an MPI. They can be chosen by policymakers and users using normative justifications (Permanyer and Hussain, 2014; UNDP and OPHI, 2019).

According to Alkire and Foster's (2011) methodology in relation to the global MPI, expert opinion and other participatory approaches led to the decision to assign equal weights to MPI dimensions and indicators. Choosing similar weights makes the index easy to understand and interpret (Atkinson and others, 2002; Alkire and Santos, 2014). However, the choice of weights represents a source of uncertainty; and so MPI sensitivity to changing weights must be tested (Cox and others, 1992; Chowdhury and Squire, 2006; Cherchye and others, 2008; Athanassoglou, 2015; Seth and McGillivray, 2018). Consequently, statistical tests of robustness are indispensable, as they provide empirical justifications for alternative choices.

Sen (1997) argued that it might be difficult to gain consensus on a specific set of dimension and indicator weights, but ideally the indices should be robust to a range of weights.

To this end, more objective approaches to assigning weights should be considered, either intrinsically or in combination with normative approaches. There are different statistical techniques for testing the robustness of weighting, such as principal component analysis (PCA), factor analysis (FA), data envelopment analysis (DEA), and frequency-based weighting (OPHI and UNDP, 2019). Both approaches complement each other in the testing of the consistency and the robustness of weights.

Although many empirical studies of multidimensional poverty indicators include robustness checks, these are usually based on simulating the index over a limited number of alternative weights. Besides equal weights, Alkire and Santos (2013 and 2014) suggested MPI weighting structures under which one of the three dimensions was assigned 50 per cent of the total weight, and the other two dimensions received 25 per cent. Indicators weights were adjusted accordingly. Also, dimension weights ranging from 25 per cent to 50 per cent have been considered (Alkire and others, 2011).

The robustness of the index may be assessed across a wider range of weights. Sleiman and others (2017) simulated the index over tens of alternative sets of weights per indicator and per dimension, producing hundreds of weighing structures. Among those structures, specific cases were included such as the default set of equal weights across indicators, and weights based on the degree of correlation among indicators using principal component analysis, factor analysis and multiple correspondence analysis (Njong and Ningaye, 2008).

C. Poverty cut-off

After aggregating all possible weighted indicators, the poverty cut-off is chosen to classify population as multi-dimensionally poor or non-poor. The poverty cut-off is a threshold level of achievement above which persons are classified as multi-dimensionally poor. It might differ from one framework to another depending on the number of dimensions and their weights. The cut-off should be fixed over time for a given framework, but it may vary across countries as its setting must reflect policy goals and priorities of each country. Countries aiming at targeting the poorest population should choose different poverty cut-offs than those aiming at monitoring poverty over time (UNDP and OPHI, 2019).

The importance of choosing the poverty cut-off highlights the need to verify that a slight change in the poverty cut-off will not change the poverty figures significantly, including the incidence of poverty, intensity of poverty and MPI, and the policy recommendations based on them.

The significance of changes in MPI results from changes in the poverty cut-off is assessed using the rankings relative to other entities, be it countries, governorates, etc. In assessing the consistency of chosen poverty cut-offs k for the identification of the multi-dimensionally poor; researchers assess the impact of the change in poverty cut-off on the poverty figure by deploying either specific k values or a range of values. OPHI (2018) suggested testing the sensitivity of country rankings to the poverty cut-off of 33.33 per cent by using alternative cut-offs: 1 per cent, 20 per cent, 40 per cent, and 50 per cent. Alkire and Santos (2010 and 2014) estimated MPI using a range of poverty cut-offs between 20 and 40 per cent, and Alkire and Apablaza (2016) used cut offs between 15 and 35 per cent.

II. External Robustness Process – How to Determine Robustness

After computing different MPIs under wide-ranging combinations of uncertainty parameters; how can one decide which MPI figure is preferable or useable for reporting? One should choose the index that is most robust, meaning the one that is least sensitive to changes in the sources of uncertainty. To have a reliable and consistent index, one should undertake robustness analysis followed by statistical inference.

Robustness analysis involves checking the ordering of two or more countries under alternative choices over the parameters specification of the index. While statistical inference involves estimating the unknown population parameters from sample data.

Robustness analysis

1. Dominance analysis

Duclos and others (2006) and Alkire and Foster (2011) used the stochastic dominance in multidimensional poverty orderings after its use in unidimensional poverty orderings. Suppose there are two univariate distributions of achievements x and y , and $F_x(b)$ and $F_y(b)$ are the proportions of the population with achievement levels less than a pre-set cut-off b , or simply the cumulative distributions of x and y . When checking the MPI dominance to common poverty cut-off for both distributions $F_x(b)$ and $F_y(b)$, $F_y(b)$ is said to **first-order stochastically dominate** (FSD) the distribution of x , $F_x(b)$, if and only if $F_y(b) \leq F_x(b)$ for all b , and $F_y(b) < F_x(b)$ for some b . This means that the distribution of y has a higher MPI than the distribution of x for all poverty cut-offs. Strict FSD signifies that $F_y(b) < F_x(b)$ for all b (Njong and Ningaye, 2008; Alkire and others, 2015).

Alkire and Santos (2011), and Alkire and others (2011) estimated MPI for 104 developing countries and found that there is a dominance relation in MPI when poverty-cutoff varies from 20 to 40 per cent in 97 per cent of all pairs of sub-Saharan African countries. This means that unambiguously one country is less poor than another regardless of the change in the poverty-cutoff.

Also, equality of indices can be tested statistically. The Wilcoxon (1945) Rank-Sum Test (a.k.a., Mann Whitney Wilcoxon Test, Mann Whitney U Test) is a nonparametric version of a two-sample t-test. It does not require a known underlying distribution of the data nor large data. It only assumes independence and equal variance of populations. It assesses whether the cumulative density of country A is the same as for country B. More generally, the Wilcoxon Rank-Sum Test compares

outcomes between two independent groups and tests the equality of matched pairs of observations. It also compares the medians between two independent groups, while parametric tests compare the means between independent groups (Kerby, 2014).

The Wilcoxon Rank-Sum Test is usually performed as a two-tailed test. The null hypothesis is that the two groups have the same distribution against the alternative hypothesis that the two groups are not equal. In case of a one-sided hypothesis test, the null hypothesis assesses a positive or negative shift in one group compared to another group. This test is based on ranking the differences between two groups' observations from the lowest to the highest, and observations with no difference are assigned the average rank.

2. Correlation coefficients for ranks

If dominance conditions are not satisfied, milder approaches can be applied. These methods test if an ordering of more than two countries remains the same when the value of some parameter is altered. The robustness of a ranking is evaluated by calculating a rank correlation coefficient between the initial and the alternative rankings of societies. Three alternative rank correlation coefficients can be used: Pearson's correlation coefficient, Spearman's rank correlation coefficient and Kendall's rank correlation coefficient (Tau-b). The two latter coefficients are the most commonly used in ranking-robustness checks (Alkire and others, 2015; UNDP and OPHI, 2019). Both coefficients assume that there are no ties in the rankings of any pair of countries.

Suppose there are m countries, and two set of ranks r and r' for two different specification sets of parameters – the initial and alternative specification – where $r = (r_1, r_2, \dots, r_m)$ and $r' = (r'_1, r'_2, \dots, r'_m)$. Also, consider r_ℓ and r'_ℓ as the ranks of a country ℓ for the two different specification sets of parameters. The alternative specification may entail a change in the set of weights, poverty-cutoff parameters or even the indicators' deprivation cut-offs. Countries' ranking may be used to assess their relative level of multidimensional headcount ratio, the intensity of poverty, or the adjusted headcount ratio. If both specifications have the same rankings across countries, $r_\ell = r'_\ell$ for all countries $\ell = 1, \dots, m$.

The Spearman rank correlation coefficient (R^ρ) is

$$R^\rho = 1 - \frac{6 \sum_{i=1}^m (r_i - r'_i)^2}{m(m^2 - 1)}$$

The *Kendall rank correlation* coefficient is an alternative that has been proposed for the case when the sample size is small and there are many tied ranks. This coefficient uses the number of concordant pairs and discordant pairs in its computation. Concordant pairs refer to those that have exactly the same ranking between two countries under the initial and the alternative specification of parameters, reflecting consistency and robustness of the index and of country rankings. A pair (x, y) is said to be concordant when $r_a > r_{b'}$ and $r'_a > r'_b$. Discordant country pairs are those that change ranking under the two alternative specifications of parameters. The Kendall rank correlation coefficient is represented as

$$R^{\tau} = \frac{\# \text{ Concordant Pairs} - \# \text{ Discordant Pairs}}{m(m-1)/2}$$

Both coefficients, R^{ρ} and R^{τ} , range between -1 and +1. The lowest value of -1 refers to a perfect negative association between two rankings, while the largest value of +1 signifies a perfect positive association. The main merit of these coefficients is their straightforward interpretation. For example, the Kendall Tau correlation coefficient of 0.85 signifies that 85 per cent of the pairwise country comparisons are concordant and robust and 15 per cent are discordant (Alkire and others, 2015).

These correlation coefficients can be adjusted to account for existing ties between countries' rankings, commonly known as tau-b and tau-c. Alkire and Jahan (2018) used these measures and found that the 2018 global MPI scores with equal weights are robust and stable, compared to the alternative specifications where one dimension is assigned a 50 per cent weight and the other two dimensions are assigned 25 per cent. Under the alternative weighting schemes, 88 per cent of all country pairs feature changes in their poverty rankings (Alkire and Foster, 2017).

3. Measuring distance

Another approach for testing the robustness of country ranks and ordering is the distance-based metrics measuring distances between points. This includes different metrics: Chord distance, Kendall's coefficient, Chi-squared distance, Manhattan distance and Euclidean distance. All these methods can be used with categorical or continuous data, and provide pairwise distances between different index values. The smaller the distance, the more similar a pair of indices.

In explaining the various distance measures, suppose a data matrix A has q rows representing sample populations and p columns referring to various MPIs. Each matrix element, α_{ij} , gauges the abundance of MPI _{j} in population i . The distances are computed between two sample populations i and h (McCune and Grace, 2002).

Euclidian Distance is simply the Pythagorean theorem applied to p dimensions rather than to a two-dimensional case. $ED_{ih} = \sqrt{\sum_{j=1}^p (\alpha_{ij} - \alpha_{hj})^2}$

Similarly, the *Manhattan distance method*, in case of p dimensions, is: $MH_{ih} = \sqrt[k]{\sum_{j=1}^p |\alpha_{ij} - \alpha_{hj}|^k}$. The exponent k expresses the given emphasis on the distance between indices. A higher k represents higher emphasis on larger differences between indices. If the power $k=1$, this gives the Manhattan distance, but if $k=2$, we have the Euclidean distance.

The *Chord distance* or Relative Euclidian distance is a standardization of the Euclidian distance method that eliminates differences in total abundance (actual totals of squared abundances) among sample populations. By standardization, this method can use indicators with different scales, eliminating any signal other than relative abundance.

$$CH_{ih} = \sqrt{\sum_{j=1}^P \left(\frac{\alpha_{ij}}{\sqrt{\sum_{l=1}^P \alpha_{il}}} - \frac{\alpha_{hj}}{\sqrt{\sum_{l=1}^P \alpha_{hl}}} \right)^2}$$

In examining MPI robustness, suppose the changed parameter is the poverty cut-off. To identify the most robust cut-off, MPI is computed for all suggested cut-off points, and countries are ranked according to the resulting MPI scores. The Euclidean Distances (Manhattan distance and Chord distance, respectively) matrix is computed for country rankings, and the distances are summed across countries for each cut-off. The cut-off with the lowest ED (MH and CH, respectively) is deemed the most robust under the chosen distance method, as it represents the most stable cut-off preserving countries' rankings. For instance, Sleiman and others (2017), in constructing an economic justice index, computed the ranks difference between each index and the alternative indices, standardized Euclidian differences for all indices, and selected the index with the smallest sum of Euclidian differences.

To use these distances, special attention should be paid to the selection of alternative parameters. We recommend considering a large number of sets, over significant ranges for parameters. Otherwise, if applied to a poor selection of schemes (few cut-offs, few weights combinations, or narrow ranges of values), the results may be biased.

4. Robustness analysis with statistical inference

Since ranks may be sensitive to minor changes in MPI values, sometimes our preferred model is not the first best according to the robustness test used. For instance, using the measuring distance approach, the preferred model might not minimize the sum of distances. If the researcher has strong normative justification for the preferred model, one can test the equality of means across two groups: the group of specifications that minimized the distances (not including the preferred model), and the second best group of specification that includes the preferred model. If the difference is significant, then the preferred model is not robust; however, if there is no statistically significant difference one may adopt the preferred model as robust.

In order to understand the test of equality of means and variances, we need to be first familiar with the concept of statistical inference. Poverty figures are usually estimated from surveys based on samples to draw inferences about an unknown population. This requires estimating inferential errors, which impact the degree of certainty over the comparison of poverty figures between two or more societies, given a set of parameter values. For instance, two societies may have different MPIs, but this difference may not be statistically significant. In case of robustness checks, one may report the number of robust pairwise comparisons as the proportion of possible pairwise comparisons that are consistent and robust, but some of these pairwise comparisons may be statistically insignificant, so by definition they are not robust (Alkire and Santos, 2014; Alkire and others, 2015). This reveals the importance of running statistical inference for sensitivity analysis and robustness of MPI and its partial indices.

The main statistical tools are standard errors, confidence intervals and hypothesis testing. Briefly, standard error is the estimated standard deviation of a statistic. Although different survey samples have the same sampling design and are drawn at the same time, they would likely yield different estimates for the same population parameters. Thus, standard error measures the consistency and reliability of an estimate from a survey sample. The lower the standard error, the greater is the reliability of the corresponding estimate. Standard errors are also essential for constructing confidence intervals and for hypothesis testing. This is crucial for robustness analysis and for mapping policy recommendations.

An estimate's confidence interval is a range that includes the true unknown population parameter with selected probability called confidence level. So, it can be used to assess the statistical reliability of an estimate from a sample survey, given that the population parameter is unknown. A significance level is the complement of the confidence level. Often, the significance level is denoted by α ranging between 0 per cent and 100 per cent. Its complement, the level of confidence is reported as $(1 - \alpha)$ per cent. For a given estimate, a 5 per cent significance level means that we are 95 per cent confident that the true population parameter lies within a given range.

Hypothesis testing means evaluating a hypothesized value of a population parameter. One-sample or two-sample tests can be used. In a one-sample test, suppose that Egypt's MPI is reported as 0.11 in 2018. The null-hypothesis is H_0 : MPI=0.11 against the alternative H_1 : MPI \neq 0.11 (two-sided), or MPI>0.11 or MPI<0.11 (one-sided). To reject or not the null hypothesis H_0 , one compares the p -value associated with the estimated parameter with its significance level. The p -value is the probability of rejecting the null hypothesis when it is true. When the p -value is higher than the desired significance level, we fail to reject the null hypothesis (UNDP and OPHI, 2019).

5. Means and variance testing

To study the significant difference between models and robust groups, a *means test* and *variance equality tests* can be used. *t-test* is the standard and simplest test for testing the null hypothesis of equal means of two populations for paired samples. This test is restrictive as it applies under the assumption that the sampled populations are normally distributed. When testing for equal means of more than two populations, a more general ANOVA technique can be used. The null hypothesis of the ANOVA test is the equality of all means versus the alternative hypothesis of not all means are equal (Snedecor and Cochran, 1983).

Sign test is an alternative non-parametric test of the equality of medians of pairs of populations. In case of a two-sided (one-sided, respectively) test, the null hypothesis is that the median of a distribution is equal (greater or equal, smaller than or equal) to the median of another distribution. This test is less restrictive as it does not make any assumption on the underlying distributions, and the data can be continuous or categorical ranks. This makes the test generally applicable, but the test may lack the statistical power of alternative tests. For instance, the Wilcoxon Rank Sum Test may also be appropriate and more powerful than the sign test in detecting consistent differences between medians (Snedecor and Cochran, 1989; Corder and Foreman, 2014). The sign test requires

that a random sample be drawn from each population, and the dependent samples be paired or matched. These assumptions are met when comparing ranks of the same country.

In case of testing variance equality of multiple populations, various tests can be used. Two such tests are: the Bartlett test and the Levene test, which differ in their assumptions regarding the underlying distributions. The former test requires normality, while the latter is less sensitive to departures from normality. If data are normally or nearly normally distributed, Bartlett's test performs better. Both tests evaluate the equality of variances across k groups against the alternative of unequal variances between at least two groups (Levene, 1960; Snedecor and Cochran, 1983).

6. Additional robustness tests

Statistical inference plays a crucial role in informing robustness analysis. To decide whether pairwise ordering is robust to changing a poverty cut-off, for instance, various robustness checks might be used. The dominance analysis assesses whether one MPI (or deprivation scores distribution) dominates over another, and which one. Is the pairwise dominance of the MPI curves statistically significant? For two societies in a pairwise ordering, one might use one-tail hypothesis tests of the difference in the two MPI estimates for alternative poverty-cutoff values. This determines whether MPI estimates of the two societies are significantly different or one society is significantly less poor than the other regardless of the poverty cut-off (Alkire and others, 2015).

One can also build confidence interval curves around each MPI curve to assess the significance and robustness of estimates. If the confidence-interval curves of the two societies do not overlap, this signifies that one society's MPI dominates the other significantly. In other words, regardless of the poverty cut-offs, one society is always poorer than the other. However, when confidence intervals overlap, confidence intervals are no longer used to assess statistical significance of an estimate. Thus a hypothesis test for dominance is preferred. These methods can be used together with another type of robustness analysis, such as the analysis of societies' ordering, to report the proportion of robust pairwise comparisons across different poverty cut-offs.

7. Examples of MPI robustness (weights, cut-offs and indicators)

Alkire and Foster (2011)'s MPI parameters are deemed to be robust in a sense that changing these parameters does not alter country ordering. Across 109 developing countries, Alkire and others (2011) used the alternative correlation coefficients (Pearson, Spearman and Kendall) between pairs of rankings to assess changes in various MPI parameters such as the indicators, the deprivation cut-offs and the weighting vector (Alkire and Santos, 2011 and 2017). Alkire and others (2011) found that the country rankings under the baseline MPI are robust to such changes. Similarly, Alkire and Santos (2013 and 2014) confirmed the robustness of their MPI, given that the rank correlation across all three alternative weighting structures was at least 0.83 (Alkire and Foster, 2011; Alkire and others, 2015).

Using alternative poverty cut-offs to identify the multi-dimensionally poor, Alkire and Santos (2010, 2014) concluded that country pairwise comparisons with a poverty cut-off at $k = 33.33$ per cent were not altered when alternative poverty cut-offs in the range $k \in [0.2, 0.4]$ were used. Also, OPHI (2018) concluded that 94.9 per cent of the statistically significant pairwise comparisons across 104 countries' MPIs were robust to alternative poverty cut-offs (1, 20, 33.33, 40, and 50 per cent).

To check how changes in indicators affect country rankings, Alkire and Santos (2011, 2014), and Alkire and others (2011) estimated MPI using alternative sets of indicators. Specifically, they added children's nutritional indicators weight-for-height and height-for-age to the baseline indicator of weight-for-age, and obtained Spearman coefficients of 0.99 and Kendall coefficients of 0.94 and 0.93 for the two new indicators, respectively. The MPI ranking across 109 countries thus appeared to be highly robust and consistent to changes in the poverty cut off. Assessing MPI robustness to poverty cut off changes, Alkire and Santos (2010) found that 94.7 per cent of pairwise comparisons were robust – one country being unambiguously less poor than another – to cut offs at 20 per cent, 33.33 per cent or 40 per cent.

Testing MPI sensitivity to alternative weighting structures, Alkire and others (2011) found that MPI ranks were robust for 85 per cent of all pairwise comparisons across various weighting alternatives. Changing the indicators' weights did not alter country rankings, but it affected the poverty estimates. Using the different correlation coefficients, the correlations between the baseline MPIs and those under the alternative weighting structures were at least 0.89, suggesting that the baseline MPI results were robust.

Alkire and Santos (2014) performed the statistical inference and robustness analysis on the global MPI (2010 estimates) for each pair of countries using a range of poverty cut-offs from 20 per cent to 40 per cent, and using alternative weighting structures. They found that the country rankings were highly robust to parameter changes.

8. Robustness of other indices

Similar analysis was conducted by Foster and others (2013) to examine the 1998 pairwise country rankings for three indices: the Human Development Index (HDI), the Index of Economic Freedom (IEF) and the Environmental Performance Index (EPI). They tested the robustness of the dimensional weights and the aggregation functions using Kendall's tau rank correlation coefficients, and found that the HDI had the highest rank robustness with 73 per cent of pairwise country rankings being fully robust, while the EPI had the lowest robustness with at most 6.5 per cent of pairwise rankings being fully robust.

References

- Alkire, Sabina; Apablaza, Mauricio and Jung, Euijin. (2014). Multidimensional poverty measurement for EU-SILC countries. OPHI Research in Progress.
- Alkire, S. and Apablaza, M. (2016). Multidimensional poverty in Europe 2006-2012: Illustrating a methodology. OPHI Working Paper 74, University of Oxford.
- Alkire, S., Foster, J. E., Seth, S., Santos, M. E., Roche, J. M., and Ballon, P. (2015). Multidimensional Poverty Measurement and Analysis, Oxford: Oxford University Press, ch. 8.
- Alkire, Sabine, and Santos, Maria Emma (2010). Acute Multidimensional Poverty: A New Index for Developing Countries. OPHI Working Paper No. 38.
- Alkire, Sabina; and Santos, Maria Emma (2011). Acute Multidimensional Poverty: A New Index for Developing Countries. Proceedings of the German Development Economics Conference, Berlin 2011, No. 3, ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg.
- Alkire, Sabina, Santos, Maria Emma, Seth, Suman and Gaston Yalonetzky. (2011). Is the Multidimensional Poverty Index robust to different weights. Oxford Poverty & Human Development Initiative (OPHI) Research in Progress Series. Available at: www.ophi.org.uk/publications/ophi-research-in-progress.
- Alkire, Sabine, and Santos, Maria Emma (2013). A Multidimensional Approach: Poverty Measurement and Beyond. *Social Indicators Research*, 112(2):239-257.
- Alkire, Sabine, and Santos, Maria Emma (2014). Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index. *World Development*, 59:251-274.
- Alkire, Sabine; Foster, James; Seth, Suman; Santos, Maria Emma; Roche, José Manuel; and Ballon, Paola. (2015). Multidimensional Poverty Measurement and Analysis. Oxford: Oxford University Press, chapter 8 – Robustness Analysis and Statistical Inference.
- Athanassoglou, S. (2015). Multidimensional Welfare Rankings under Weight Imprecision: A Social Choice Perspective. *Social Choice and Welfare*, 44(4):719-744.
- Atkinson, A.B., Cantillon, B., Marlier, T., Nolan, B. (2002). Social indicators: The EU and social inclusion. Oxford: Oxford University Press.
- Bullen, P.S., (2013). Handbook of Means and Their Inequalities. Springer Science and Business Media.
- Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R. (2002). Aggregation operators: properties, classes and construction methods. *Aggregation Operators*. Springer, pp. 3-104.

- Chen, L., Pu, P., (2004). Survey of Preference Elicitation Methods. Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland.
- Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T. V., Saisana, M., Liska, A. S. R., and Tarantola, S. (2008). Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59(2):239-251.
- Chowdhury, S. and Squire, L. (2006). Setting Weights for Aggregate Indices: An Application to the Commitment to Development Index and Human Development Index. *The Journal of Development Studies*, 42(5):761-771.
- Corder, Gregory W.; Foreman, Dale I. (2014). *Statistical Power, Nonparametric Statistics: A Step-by-Step Approach* (2nd ed.), John Wiley & Sons, ISBN 9781118840429.
- Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J., and Jones, D. R. (1992). Quality-of-Life Assessment: Can We Keep It Simple? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(3):353-393.
- Decancq, K. and Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An Overview. *Econometric Reviews*, 32(1):7-34.
- Duclos, J.-Y., Sahn, D. E., and Younger, S. D. (2006). Robust Multidimensional Poverty Comparisons. *The Economic Journal*, 116(514):943-968.
- Foster, James E.; McGillivray, Mark; and Seth, Suman (2013). Composite Indices: Rank Robustness, Statistical Association, and Redundancy. *Econometric Reviews, Taylor & Francis Journals*, vol. 32, Issue 1, pp. 35-56, January.
- Gan, Xiaoyu; Fernandez, Ignacio C.; Guo, Jie; Wilson, Maxwell; Zhao, Yuanyuan; Zhou, Bingbing; and Wu, Jianguo (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, vol. 81, pp. 491-502.
- Greco, Salvatore; Ishizaka, Alessio; Tasiou, Menelaos; and Torrisi, Gianpiero (2019). On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research*, vol. 141, pp. 61-94. Retrieved Online: <https://link.springer.com/article/10.1007/s11205-017-1832-9>.
- Kerby, Dave S. (2014) The simple difference formula: an approach to teaching nonparametric correlation. *Innovative Teaching*, Article 1.
- McCune, Bruce; and Grace, James B. (2002). Analysis of Ecological Communities; Chapter 6: Distance Measures. *Journal of Experimental Marine Biology and Ecology*.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., (2005). Tools for Composite Indicators Building. European Commission, Institute for the Protection and Security of the Citizen, Ispra, Italy.
- Njong, A.M. and Ningaye, P. (2008). Characterizing Weights in the Measurement of Multidimensional Poverty: An Application of Data-Driven Approaches to Cameroonian Data. *OPHI Working Paper* 21.

- OECD, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD publishing.
- Oxford Poverty and Human Development Initiative (2018). *Global Multidimensional Poverty Index 2018: The Most Detailed Picture To Date of the World's Poorest People*, University of Oxford, UK.
- Permanyer, Iñaki and Hussain, Majeed A. (2014). Robust Approaches to the Measurement of Multidimensional Poverty: The Case of Countries in Africa, Asia and Latin America. Online: <https://www.semanticscholar.org/paper/Robust-Approaches-to-the-Measurement-of-Poverty%3A-of-Permanyer-Hussain/bb574ea4ac48430a99cdd7a4848ed202e3339cd3>.
- Pollesch, N., Dale, V., (2015). Applications of aggregation theory to sustainability assessment. *Ecological Economics*. Vol. 114, pp. 117-127.
- Saisana, Michaela; Saltelli, Andrea and Tarantola, Stefano (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society, Statistics in Society, Series A*, vol. 168, Issue 2, pp. 307-323.
- Sen, A. (1997). Quality of Life and Economic Evaluation, *Academia Economic Papers*, 25(3), 269-316.
- Seth, Suman, and McGillivray, Mark (2018). Composite indices, alternative weights, and comparison robustness. *Social Choice and Welfare*. SpringerLink doi:10.1007/s00355-018-1132-6.
- Sleiman, Sama, Araj, Salim, Kamaly, Ahmad, and Tarabay, Hala. (2017). Economic Justice in the Arab Region.
- Snedecor and Cochran (1989). *Statistical Methods*. Eighth Edition, Iowa State University Press, pp. 138-140.
- United Nations Development Programme (UNDP) and Oxford Poverty and Human Development Initiative (OPHI). (2019). *How to Build a National Multidimensional Poverty Index (MPI): Using the MPI to Inform the SDGs*. UNDP and OPHI, University of Oxford.
- Wilcoxon, Frank (Dec 1945). Individual comparisons by ranking methods (PDF). *Biometrics Bulletin*. 1 (6): 80-83. doi:10.2307/3001968. hdl:10338.dmlcz/135688. JSTOR 3001968.



