OPHI
OXFORD POVERTY & HUMAN DEVELOPMENT INITIATIVE
www.ophi.org.uk

UNIVERSITY OF
OXFORD

Summer School on Multidimensional Poverty

8–19 July 2013

Institute for International Economic Policy (IIEP)
George Washington University
Washington, DC

Dominance, Robustness and Standard Error

Paola Ballon & Suman Seth

Oxford Poverty & Human Development Initiative (OPHI)

Part I
Standard Error

What are the main sources of Error?

These could be categorised as: statistical & non statistical

A. Statistical: Sampling Error

B. Non Statistical:

1. **Data Entry Error**
2. **Measurement Error: Sources**
   - Recall error (don't remember correctly)
   - Telescoping (incorrect date recall)
   - Reporting Errors (due to long surveys)
   - Prestige errors (misreport due to social pressures)
   - Conditioning effects (from being in the survey)
   - Respondent effects (respondent identity affects answers)
   - Interviewer effects (facilitator bias; mis-measuring a baby)
   - Non-response rate
   - Inadequate sampling frame (Source: Nestor 1970; Deaton & Grosh 2000).

Which of these can we correct for existing data?

Sampling Error? Data Entry Error?
Or Measurement Error?

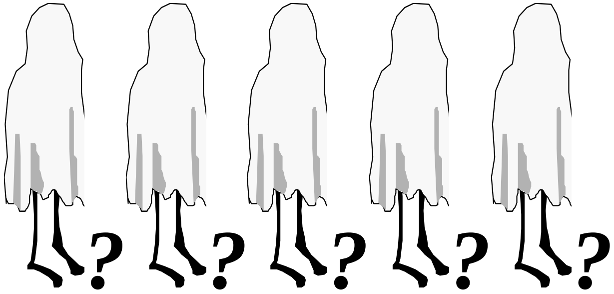
- Recall error (don't remember correctly)
- Telescoping (incorrect date recall)
- Reporting Errors (due to long surveys)
- Prestige errors (misreport due to social pressures)
- Conditioning effects (from being in the survey)
- Respondent effects (respondent identity affects answers)
- Interviewer effects (facilitator bias; mis-measuring a baby)
- Non-response rate
- Inadequate sampling frame

We can only correct for sampling error
For this we shall compute standard errors, and build confidence intervals

Focus of this class:

How Accurate are my Measures?

If I use them for policy, what is the chance that they are mistaken?



Focus of Part I of this class

Understand, from a policy perspective, the need for computing standard errors, and building confidence intervals on M^{α} measures

Some terminology

•**Inferential statistics** like standard error and confidence intervals (CI) deal with inferences about populations **based on** the behavior of **samples**.

•Both standard errors and CI's will help us determining how likely it is that results based on a sample (or samples) are the same results that would have been obtained for the entire population

Standard error & standard deviation

- **Standard error** of a random variable, like H or A, is the sample estimation of its (population) standard deviation. The standard error gives us an idea of the precision of the sample estimation.

- The **standard deviation**, intuitively, is a notion of uncertainty. To obtain the standard deviation of a random variable we need first to compute the variance.

- The **variance** of a random variable, like H, A or any other M^{α}, is a measure of the dispersion of the distribution around the mean. The smaller the variance the lesser the degree of uncertainty.

How to obtain the standard error

To compute the standard error of any M^{α} measure we can use:

1. Analytical standard errors: “Formulas” which either provide the exact or the asymptotic approximation of the standard error (Yalonetky, 2010).

2. Resampling methods like the bootstrap (Alkire, & Santos, 2010).

Confidence Intervals (CI)

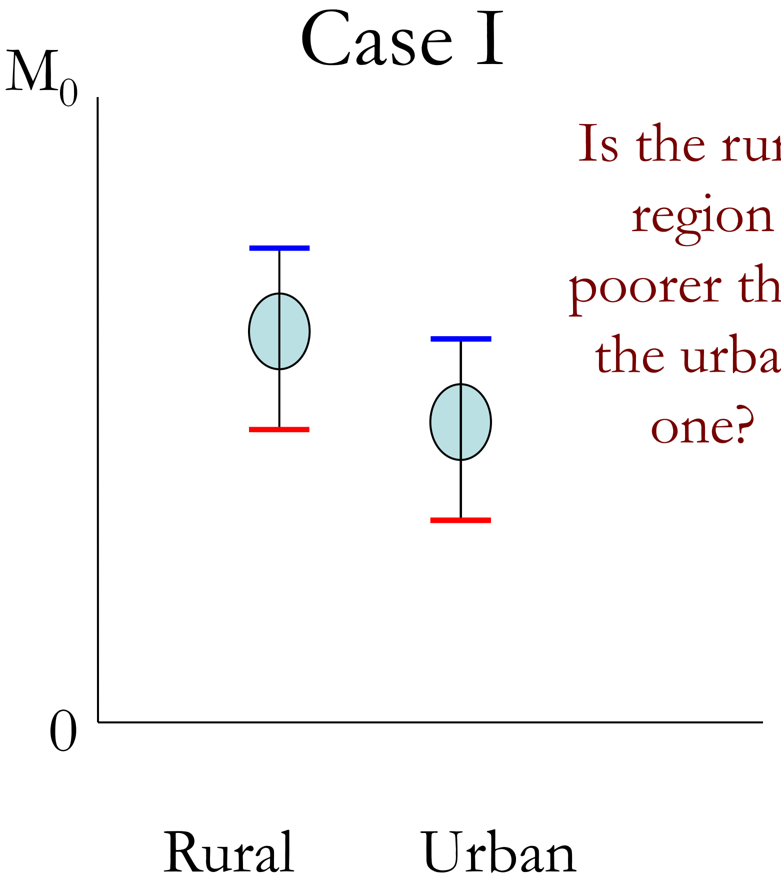
The standard error is a point estimate. In many cases we might be interested in looking at interval estimates, called confidence intervals.

Confidence intervals provide a range of likely values for the population parameter (i.e, H, or A) and **not** just a **point estimate**.

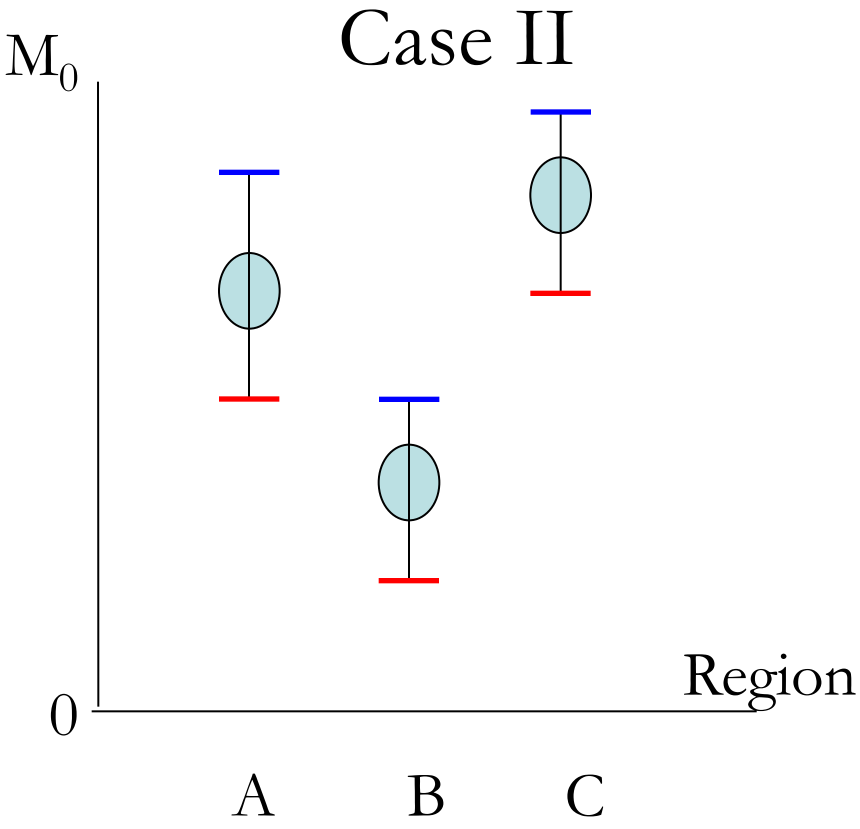
For example to compare H (or A) values across urban and rural regions we can **compare their confidence intervals** and **look for overlaps**.

If the CI's overlap then we might conclude that the poverty comparison (H values) across urban and rural regions is not statistically different.

Are the M_0 values different?

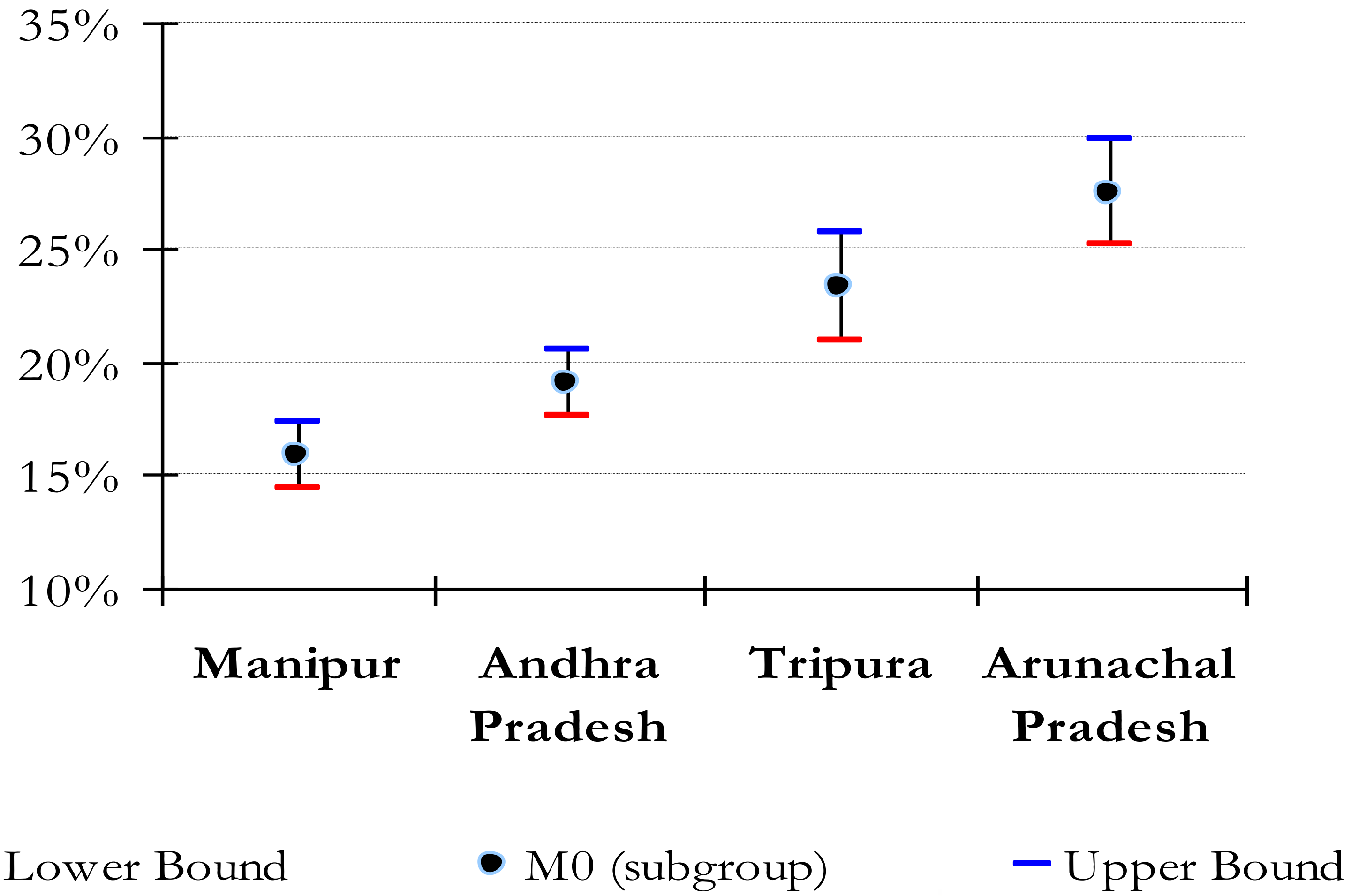


Is the rural region poorer than the urban one?



Let's look at the CI's

Are the M0 values different? - CI



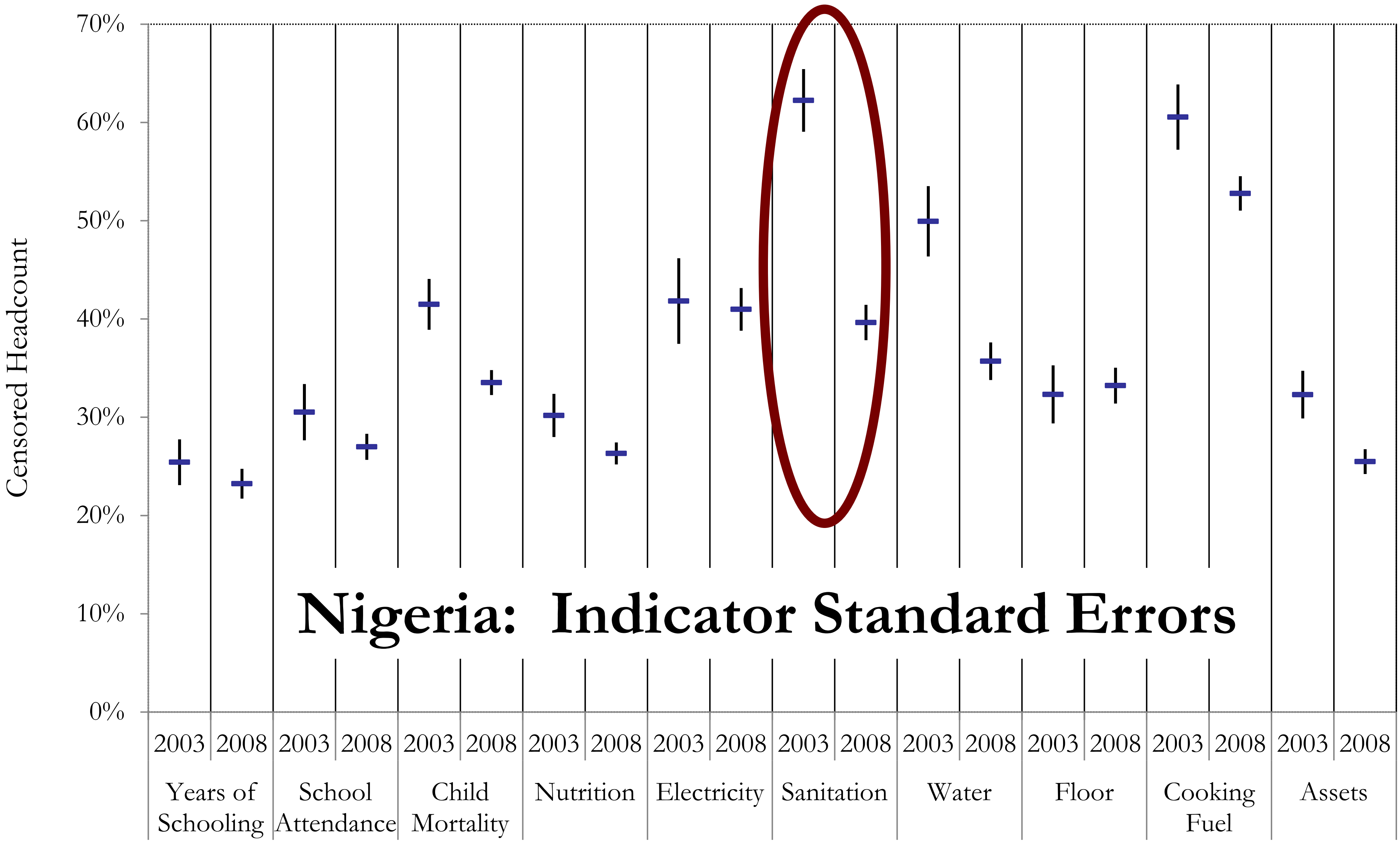
Source Alkire and Seth (2012)

Are the M_0 values different?

Subgroup	Lower Bound	M0 (subgroup)	Upper Bound	Difference in M0	Difference Significant
Manipur	0.143	0.158	0.173	0.033	Yes
Andhra Pradesh	0.175	0.190	0.206		
Tripura	0.209	0.233	0.257	0.042	No
Arunachal Pradesh	0.251	0.275	0.299		

Source Alkire and Seth (2012): Trend and Analysis of Multidimensional Poverty in India, 1999 and 2006

Are the M_0 values different? - CI



MPI: are the M_0 values different? - CI

	Education		Health		Living Standards					
	YS	SA	CM	N	E	S	W	F	CF	AO
Average annual absolute change in censored headcount										
Bangladesh	-1.3*	-3.9*	-1.8*	-2.3*	-3.0*	-3.4*	-0.1	-3.1*	-3.0*	-3.8*
Bolivia	-0.4*	-3.9*	-1.5*	-0.4*	-2.2*	-3.1*	-1.6*	-1.9*	-2.1*	-1.7*
Colombia	-0.2*	-0.4*	-0.4*	-0.4*	-0.2*	-0.4*	-0.2*	-0.4*	-0.6*	-0.8*
Ethiopia	-2.2*	-1.1*	-1.1*	-5.8*	-0.6*	-2.0*	-5.8*	-0.8*	-0.9*	-1.0*
Ghana	-1.3*	-4.3*	-3.9*	-2.6*	-4.0*	-5.5*	-3.5*	-0.3	-5.5*	-4.1*
Jordan	0.0	0.3	0.1	-0.6*	0.0	-0.1	-0.1	0.0	0.0	0.0
Kenya	-0.6	-0.4	-0.9*	-0.2	-2.1*	-2.7*	-2.8*	-1.8	-1.9*	-2.9*
Lesotho	-0.7*	-1.2*	-0.7*	-0.1	-2.4*	-2.5*	-1.4*	-1.6*	-1.8*	-3.5*
Madagascar	0.9	-0.2	-0.6	-4.3*	-0.4	1.2*	-1.7*	-0.5	-0.5	-1.8*
Nigeria	-0.4	-0.7	-1.6*	-0.8*	-0.2	-4.5*	-2.8*	0.2*	-1.6*	-1.4*

Source: Alkire and Roche (2012) - International MPI

Note: Figures with an asterisk have non-overlapping confidence intervals at 95 percent.

Part II
Robustness and Dominance Analysis

Focus of Part II of this class

Understanding, how sensitive your policy prescriptions are to decisions you have made in designing the measure. It could be that the measure changes radically if you adjust a parameter that you don't feel strongly about.

Note: Think through how the measure is used. Test robustness accordingly

To test the robustness of your measure, you have to first identify the comparisons that matter for policy.

E.g. Central Government allocates budget according to the MPI in each **region** of the country. *(I need to test if the regional comparisons are robust)*

A minister wants to show the steepest *decrease* in poverty in their region/dimension. (*test* ***trends***)

Other examples: comparisons of % contribution vs MPI or H? Or comparisons by social groups?

Familiar for pov lines – e.g. Chen &Ravallion 2008

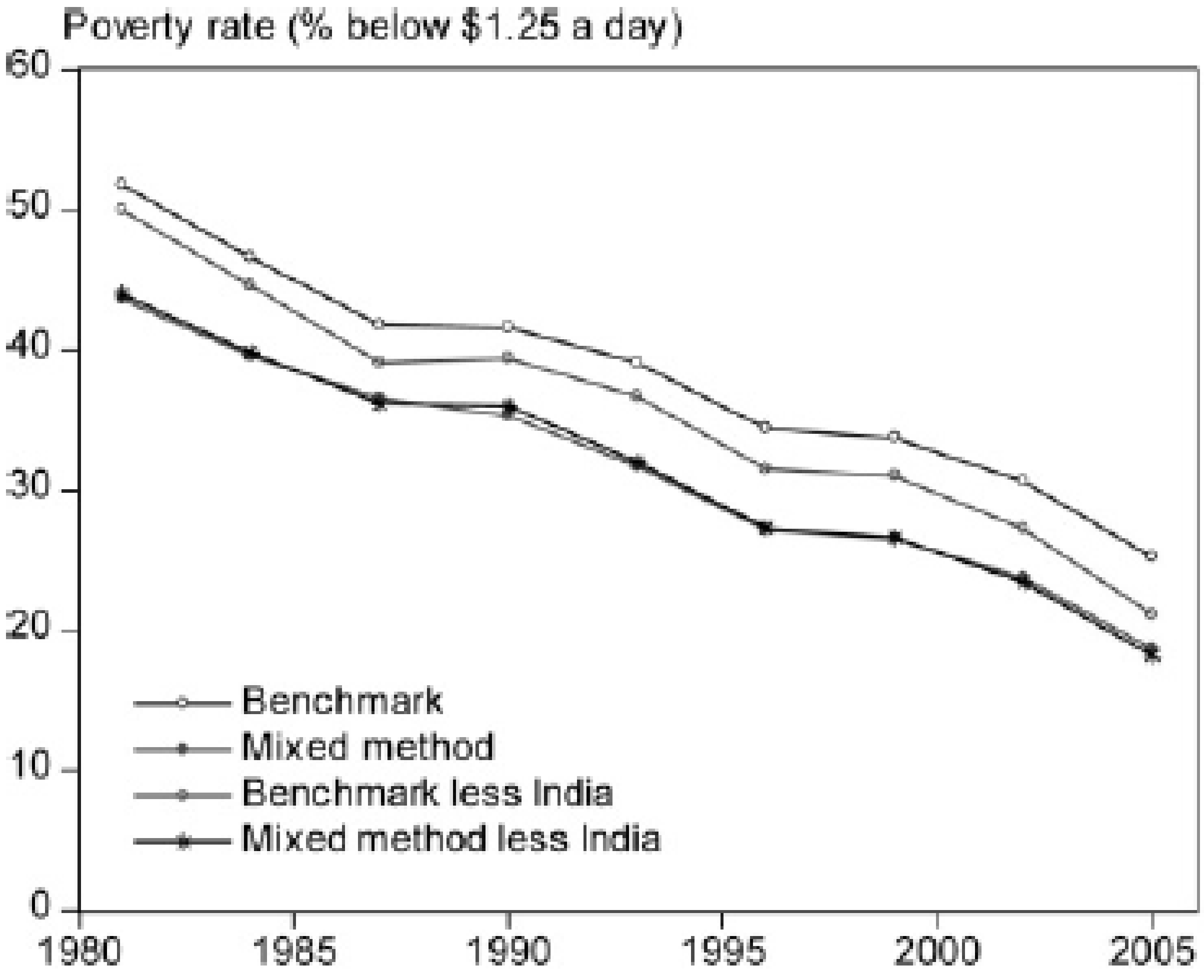


FIGURE VI

Aggregate Poverty Rates over Time for Benchmark and Mixed Method

Aspects that Affect MD Poverty Comparison

- Poverty Cutoff
- Weighting vector
- Deprivation cutoffs
- The measure in the AF class used
- Sample size and statistical significance

Comparison may *alter* when parameters vary

For $k = 1/3$	**For $k = 1/2$**
MPI of Nigeria is 0.310	MPI of Nigeria is 0.232
MPI for Zambia is 0.328	MPI for Zambia is 0.214

Why is This Important?

- Given that policy decisions based on the measure used affect the lives of the poor, it is important to understand the sensitivity of the measure with respect to the choice of parameters

Main Sources of this Lecture

- Batana (2008) OPHI Working Paper 13
- Yalonetzky (2012) ECONEQ WP 2012 – 257
- Alkire and Santos (2013)


Tabita, Kenya


Rabiya, India


Stephanie, Madagascar


Agatha, Madagascar


Dalma, Kenya


Ann-Sophia, Kenya

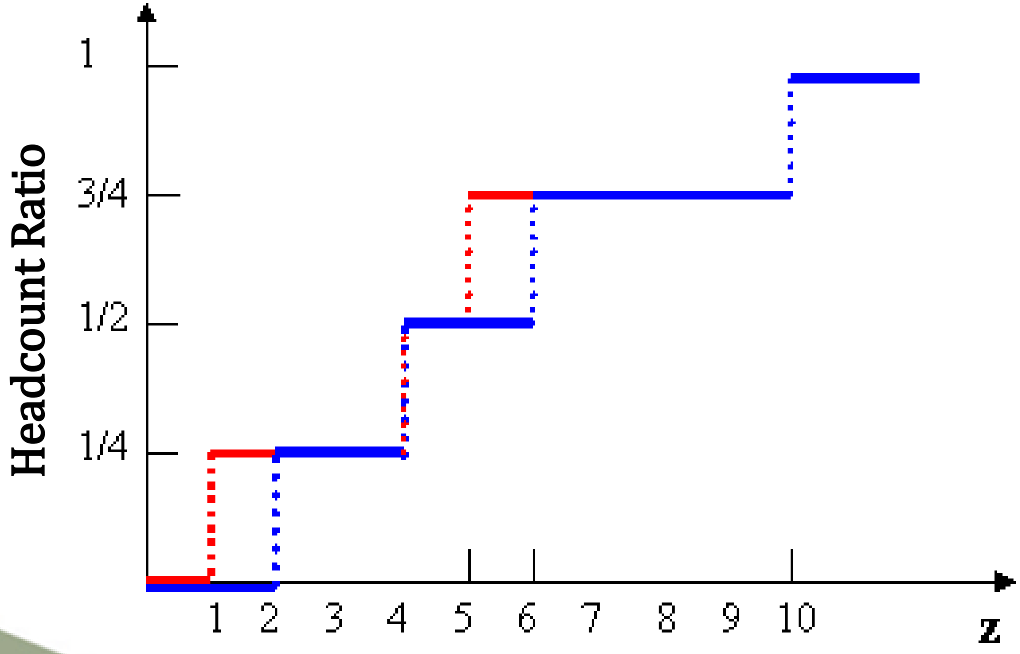

Valérie, Madagascar

Dominance in a Multidimensional Framework

Recap: Unidimensional FSD

First order Stochastic Dominance

Example of FSD: Let x=(2,4,6,10) and y=(1,4,5,10)



No part of y lies to the right of x

Thus, x FSD y in this case, which means x has unambiguously less poverty than y according to H

Dominance for H and M_0 in AF

Question: When can we say that a distribution has higher H or M0 for any poverty cutoff (*k*), for a given weight vector and a given deprivation cutoff vector?

Hint: The concept can be borrowed from the unidimensional stochastic dominance

Alkire and Foster (2011)

Dominance for H and M_0 in AF

Consider the following deprivation matrix

$$\begin{array}{r|cccc|} & \textbf{Income} & \textbf{Years of Education} & \textbf{Sanitation (Improved?)} & \textbf{Access to Electricity} \\ & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 1 \\ \mathbf{g}^0 = & 1 & 1 & 1 & 1 \\ & 0 & 1 & 0 & 0 \end{array}$$

$$\mathbf{z} = \left| \begin{array}{cccc} \mathbf{500} & \mathbf{12} & \mathbf{1} & \mathbf{1} \end{array} \right|$$

Dominance for H and M_0 in AF

- For equal weight, deprivation score vector is c

$$
\begin{array}{c|cccc|c}
 & \text{Income} & \text{Years of Education} & \text{Sanitation (Improved?)} & \text{Access to Electricity} & \mathbf{c} \\
 & 0 & 0 & 0 & 0 & \mathbf{0} \\
g^0 = & 1 & 0 & 0 & 1 & \mathbf{2} \\
 & 1 & 1 & 1 & 1 & \mathbf{4} \\
 & 0 & 1 & 0 & 0 & \mathbf{1}
\end{array}
$$

$$
\begin{array}{c|cccc|}
z = & \mathbf{500} & \mathbf{12} & \mathbf{1} & \mathbf{1}
\end{array}
$$

Dominance for H and M_0 in AF

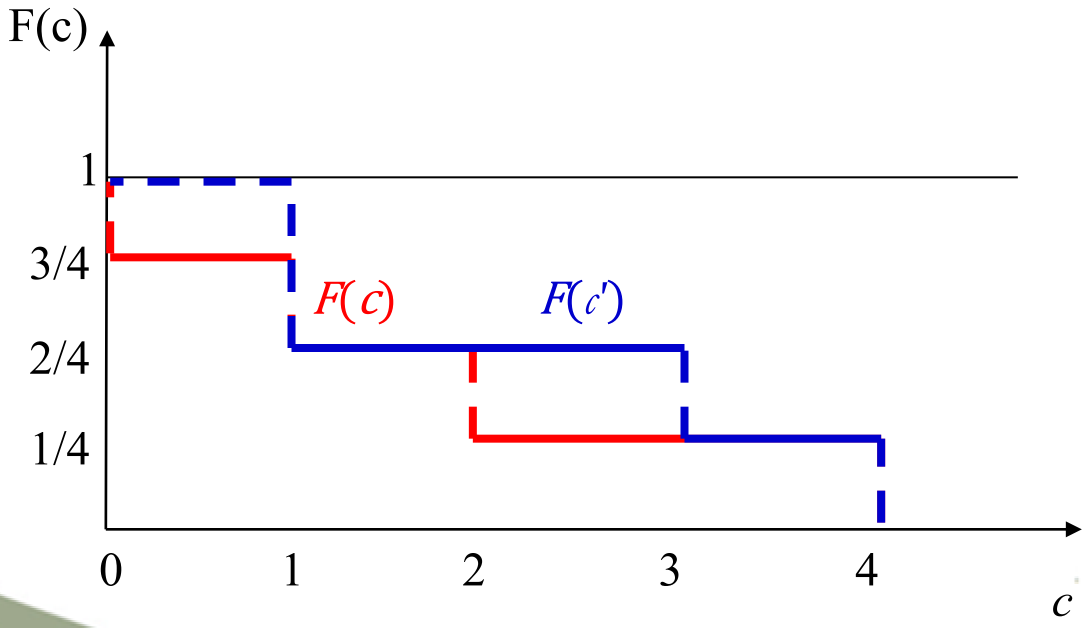
Result (Alkire and Foster 2011)

- If a deprivation score vector *c* for distribution X first order stochastically dominates another deprivation score vector *c*′ of X′, then X has *no higher* H and M_0 than X′ for all *k* and X has strictly lower H and M_0 than X′ for some *k*

Note however that the distribution functions would be downward sloping instead of upward rising

Example

Let the two deprivation score vectors be $c = (0, 1, 2, 4)$ and $c' = (1, 1, 3, 4)$

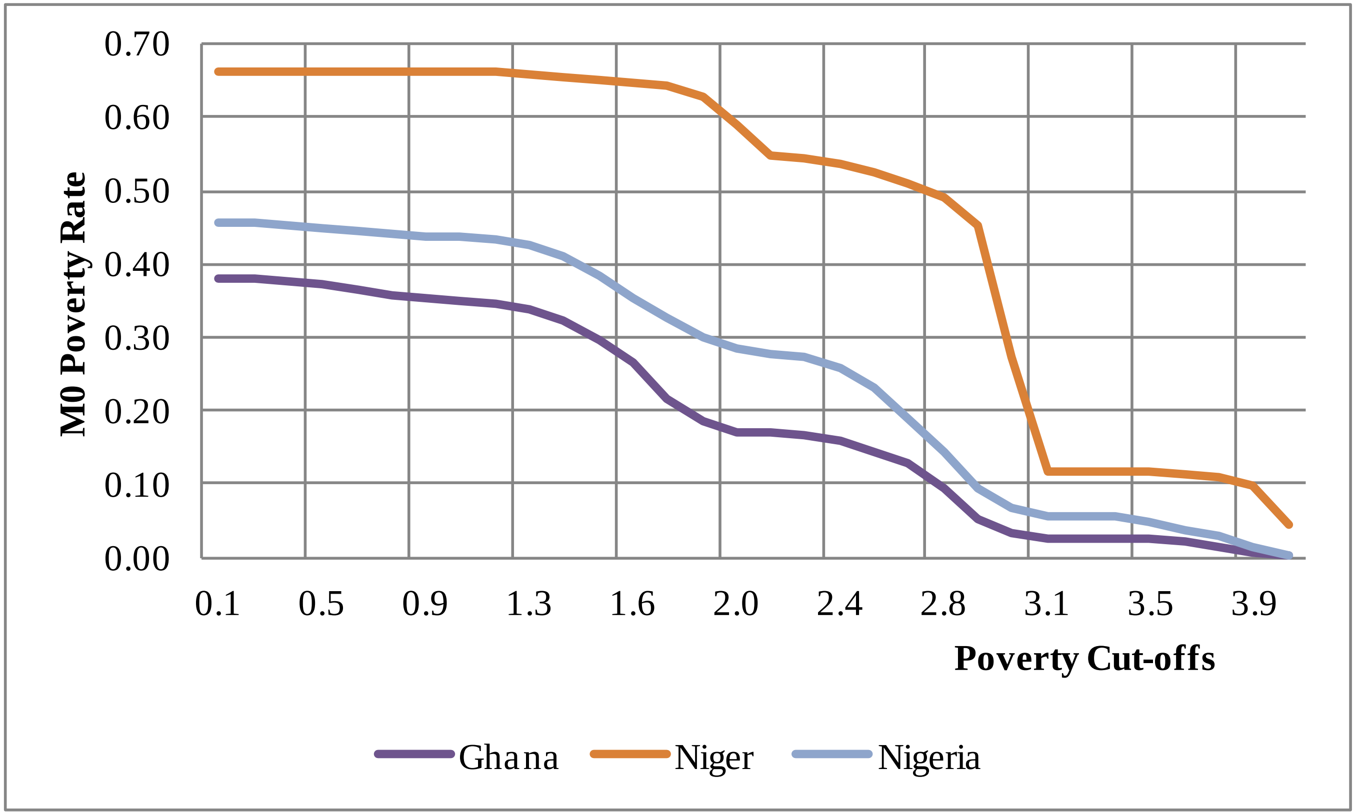


Is there any poverty (k) for which there is more poverty in c than in c'?

M_0 Curve

Dominance holds in terms for M_0 for all *k*

In case of sample surveys, the confidence intervals for each *k* also needs to be computed in order to conclude dominance



Source: Batana (2008)

Tabita, Kenya

Rabiya, India

Stephanie, Madagascar

Agatha, Madagascar

Dalma, Kenya

Ann-Sophia, Kenya

Valérie, Madagascar

Robustness

Why Need Robustness Checks?

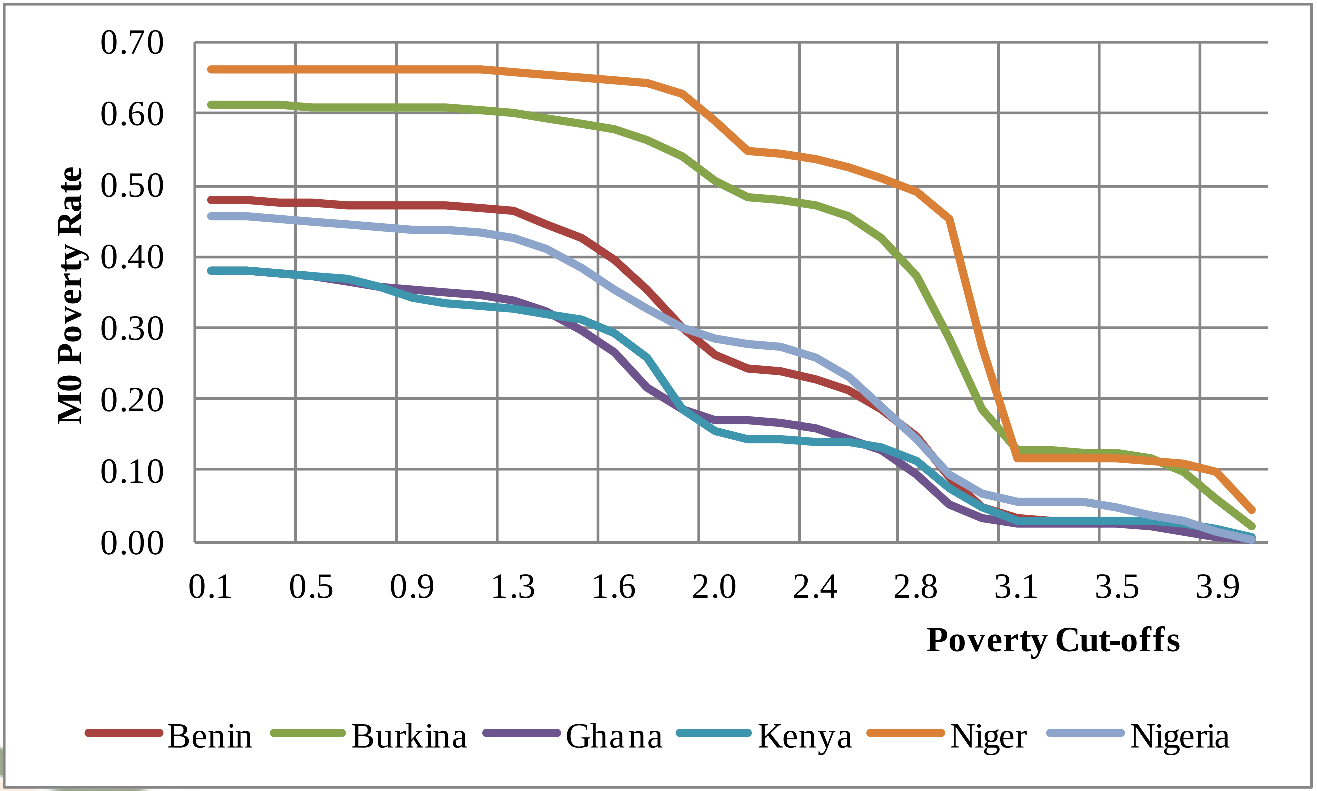
Note that dominance is an extreme form of robustness

Stochastic Dominance (SD) conditions are useful for pair by pair analysis

SD conditions may be too stringent and may not hold for the majority of the countries

Why Need Robustness Checks?

Note below that NOT all countries stochastically dominates each other (Batana 2008)



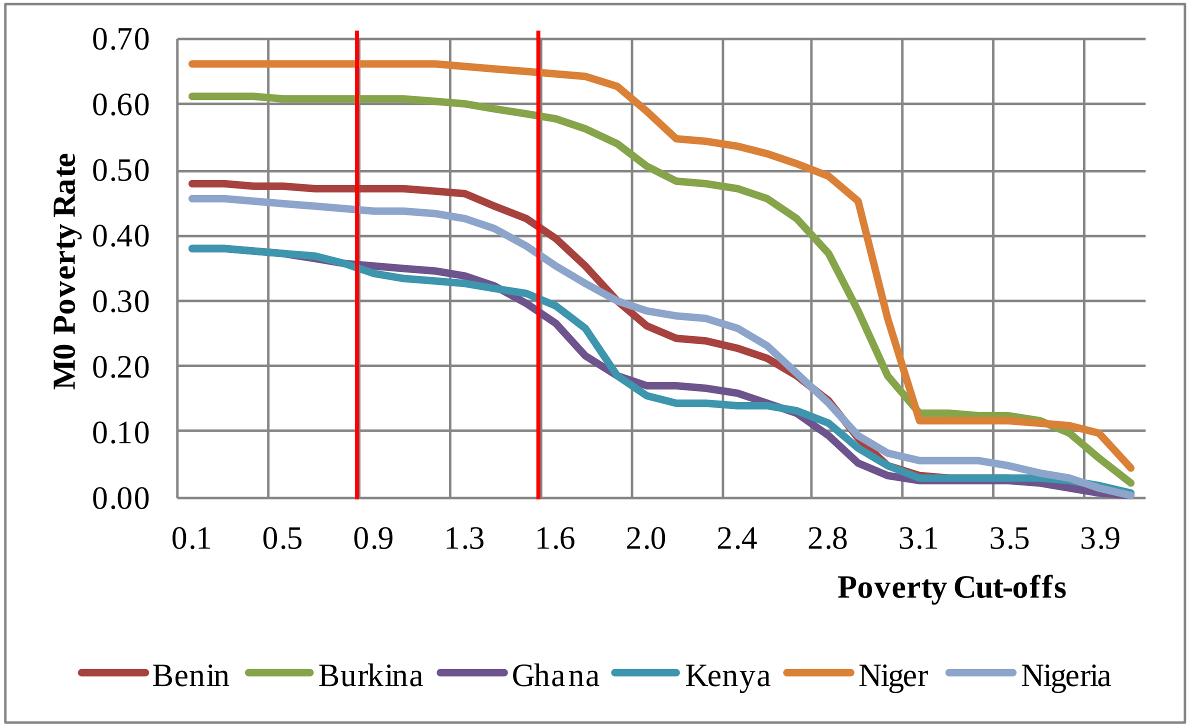
Why Need Robustness Checks?

Thus, we need other ways to understand how robust the rankings are to changes in weights and cutoffs or test for restricted dominance

Robustness of Comparison

Is a particular comparison between two countries or regions robust when *k* varies between an interval?

If *k* is between **0.9** and **1.6**, the comparisons between Niger & Burkina Faso and Benin & Nigeria are robust, but that between Kenya & Ghana is not



What if *k* ranges between 0.9 and 2.4?

OPHI Oxford Poverty & Human Development Initiative

UNIVERSITY OF OXFORD

Robustness of Comparison

Until now, we have only compared the robustness of a pair of countries or regions

How can we evaluate the ranking of a set of countries or regions, when

- the poverty cut-off varies
- the weights vary
- the deprivation cutoffs vary

Measuring Robustness of Comparisons

- A useful method for comparing robustness of ranking is to compute rank correlation coefficients
    - Spearman's rank correlation coefficient
    - Kendall's rank correlation coefficient
    - Percentage of pair-wise comparisons that are robust
- First, different rankings of countries or regions are generated for different specifications of parameters
    - Different weighting vectors, different poverty/deprivation cutoffs
- Next, the pair-wise ranks and rank correlation coefficients are computed

Kendall's Tau

- For each pair, it is found if the comparison is *concordant* or *discordant*
  - 10 countries means 45 pair-wise comparisons

- The comparison between a pair of countries is *concordant* if one dominates the other for both specifications (C)

- The comparison between a pair of countries is *discordant* if one dominates the other for one specification but is dominated for the other specification (D)

Kendall’s Tau

- The Kendall’s Tau rank correlation coefficient (*t*) is equal to

$$\tau = \frac{C - D}{C + D}$$

- It lies between -1 and +1

- If there are ties, this measure should be adjusted for ties
  - The tie adjusted Tau is known as tau-b

Spearman’s Rho

- The Spearman’s Rho also measures rank correlation but is slightly different from Tau
  - First countries are ranked for two specifications
  - Then, for each country the difference in two ranks are computed (r_i for country i)

- The Spearman’s Rho (r) is

$$\rho = 1 - \frac{6\sum_{i=1}^{n} r_i^2}{n(n^2 - 1)}$$

Some Illustrations using the MPI

Robustness to weights

Re-weight each dimension:

33%	50%	25%	25%
33%	25%	50%	25%
33%	25%	25%	50%

Robustness to Weights

			MPI Weights 1	MPI Weights 2	MPI Weights 3
			Equal weights: 33% each **(Selected Measure)**	50% Education 25% Health 25% LS	50% Health 25% Education 25% LS
MPI Weights 2	50% Education	Pearson	0.992		
	25% Health	Spearman	0.979		
	25% LS	Kendall (Taub)	0.893		
MPI Weights 3	50% Health	Pearson	0.995	0.984	
	25% Education	Spearman	0.987	0.954	
	25% LS	Kendall (Taub)	0.918	0.829	
MPI Weights 4	50% LS	Pearson	0.987	0.965	0.975
	25% Education	Spearman	0.985	0.973	0.968
	25% Health	Kendall (Taub)	0.904	0.863	0.854
Number of countries:		**109**			

Alkire and Santos (2010, 2013)

Robustness to Poverty Cutoff (k)

Spearman's Rank Correlation Table for different poverty cutoffs out of 10 indicators in India

Cut-off (k)	**3**	**4**	**5**	**6**	**7**
4	1.00	-	-	-	-
5	0.99	1.00	-	-	-
6	0.99	1.00	1.00	-	-
7	0.97	0.97	0.98	0.98	-
8	0.96	0.96	0.96	0.97	0.98

Alkire and Seth (2008)

Other tests

You can perform the Friedman test for rank independence (it will mostly reject independence – not very strong)

You can perform the Wilcoxon test for changes in weights.

You can also bootstrap (Alkire & Santos 2013)

- We estimated the MPI for the different selected *k*-values and bootstrapped them. As all the surveys used have a complex survey design we have drawn samples of clusters (with replacement) within each strata (Deaton, 1997). For each country we have performed 1000 replications and with these estimates we have created the bootstrap 95% confidence intervals. Given two countries, A and B, we say that B dominates A if A's (bootstrapped) lower bound MPI estimate is greater than B's (bootstrapped) upper bound MPI for all the considered *k* values. That is, B has lower poverty than A regardless of the *k*-cutoff and considering alternative samples. We perform this comparison for all the possible pairs of countries.
- We find that 87.4% of all possible pair wise comparisons of bootstrapped estimates are robust to a change of *k* between 20% and 40%, meaning that one country is unambiguously less poor than another, independently of whether we require people to be deprived in 20, 33 or 40% of the weighted indicators.
- When we test for robustness considering only countries with 10 indicators we find that 91.2% of the comparisons are robust and when we discriminate by survey, we find that 91.7% of comparisons among DHS countries are robust, 85.2% among MICS countries and only 59.6% among WHS countries, analogous to the results with the robustness to weights.