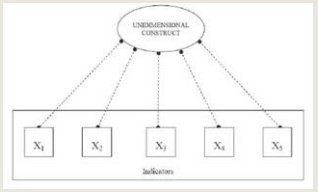
Reliability / Confiabilidad

Héctor Nájera

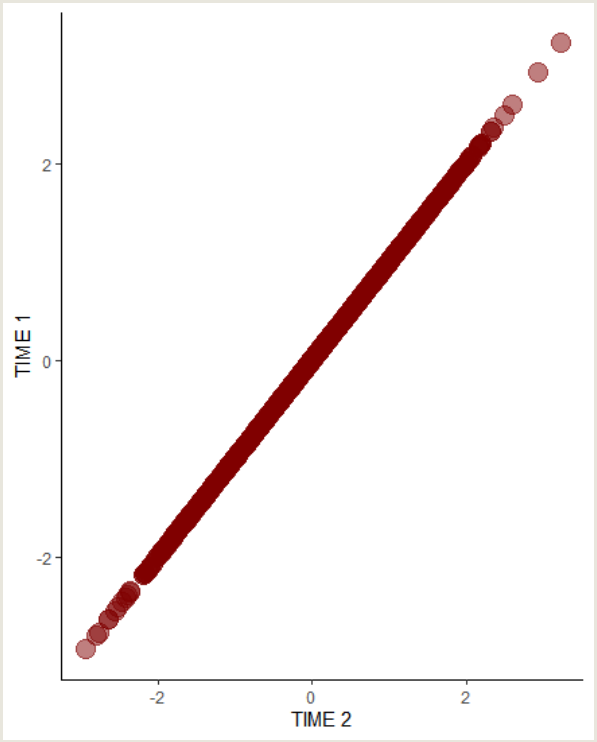
Curso internacional de teorías y métodos contemporáneos para la medición de la pobreza multidimensional

29 Nov 2021

Reliability as consistency

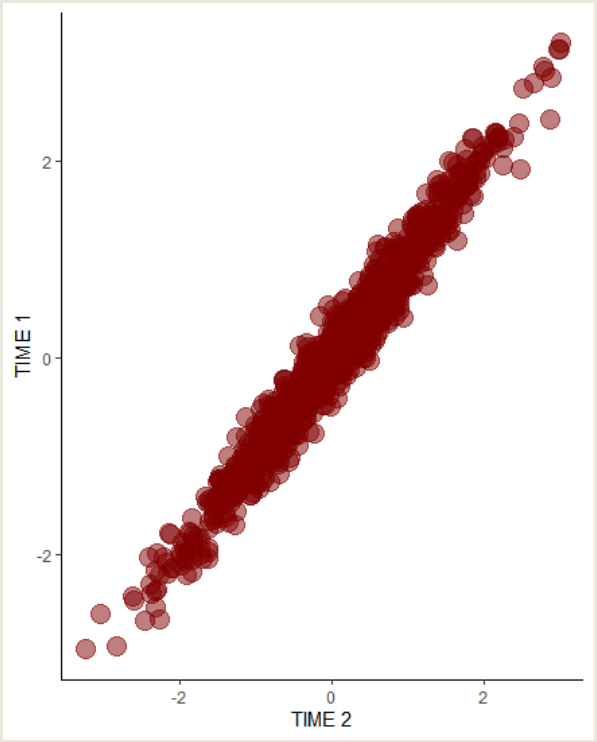


No changes in T1 and T2



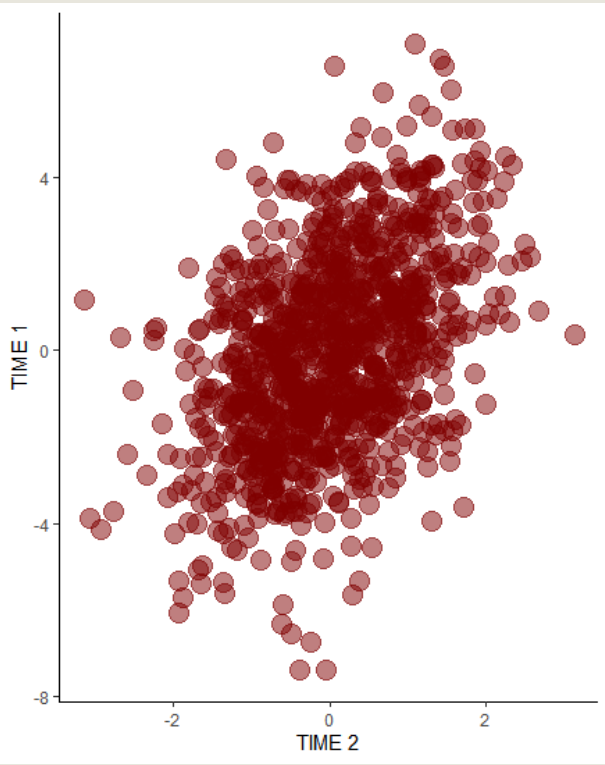
We trust the scores, instrument and its underlying model

Small error T1 and T2



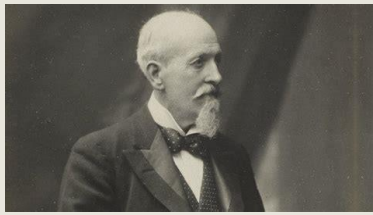
We trust the scores, instrument and its underlying model – we tolerate small variations–

A lot of error T1 and T2

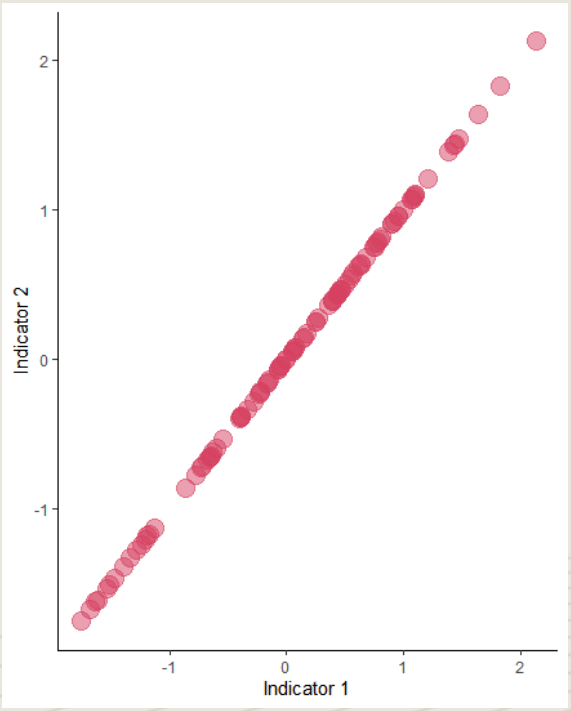


We don't trust the scores

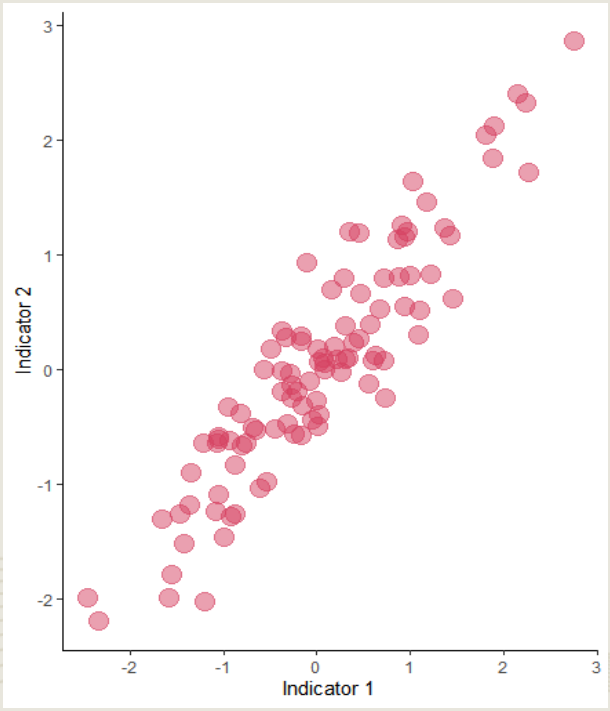
Fundamentals of the principle of reliability



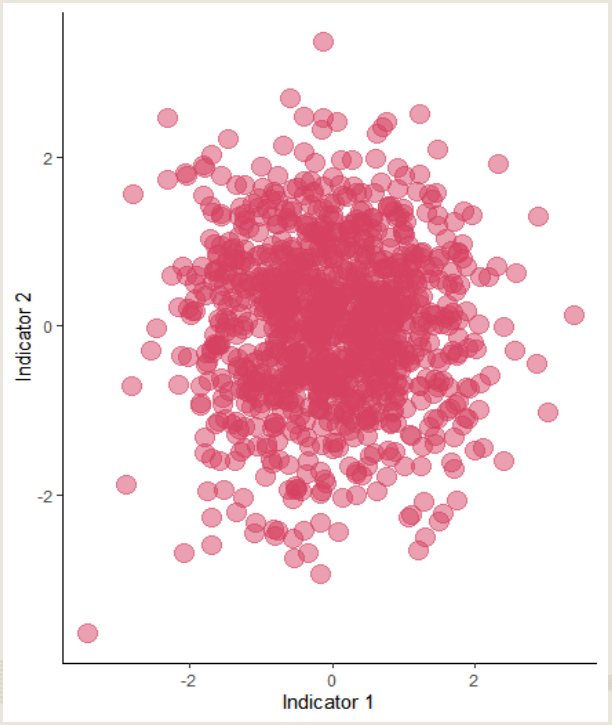
If the total variability of two indicators is explained by the same source:



If the variability of two indicators is mostly explained by the same source:



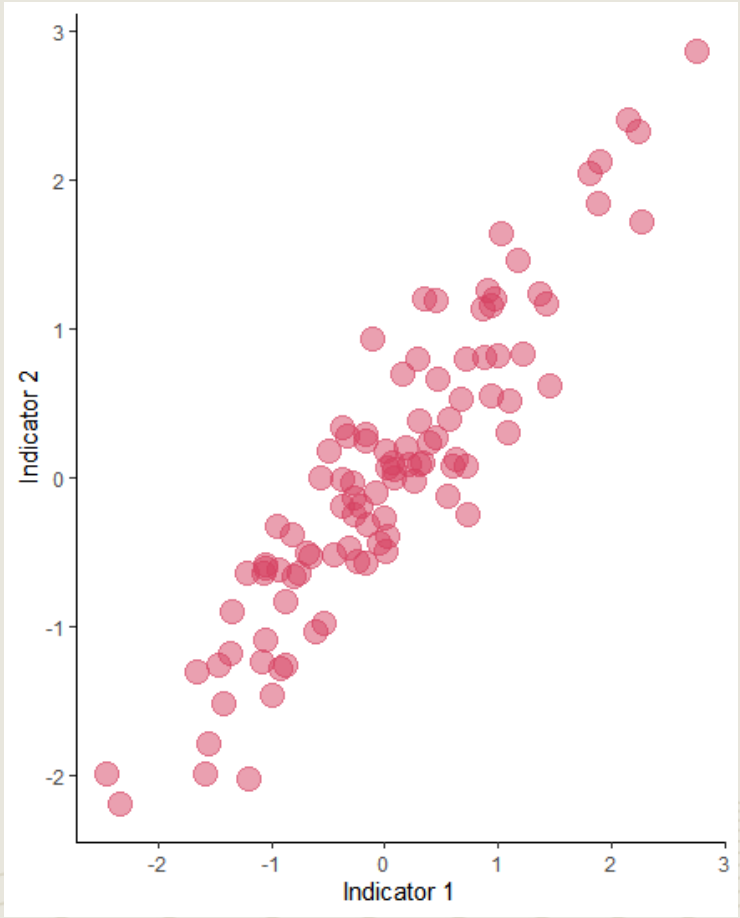
If the variability of two indicators has no common source:

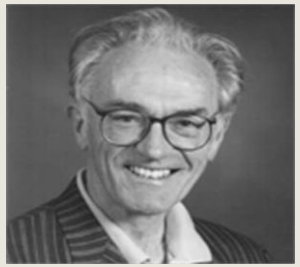


The relationships between manifestations of the same phenomenon are attenuated by measurement error

Reliability and latent variables

We theorize that what is observable are the manifestations of poverty: To changes in poverty -latent-, changes in the readings we make





Deprivations are manifestations – consequences – of poverty

Monotonicity: Income Falls and Poverty Increases

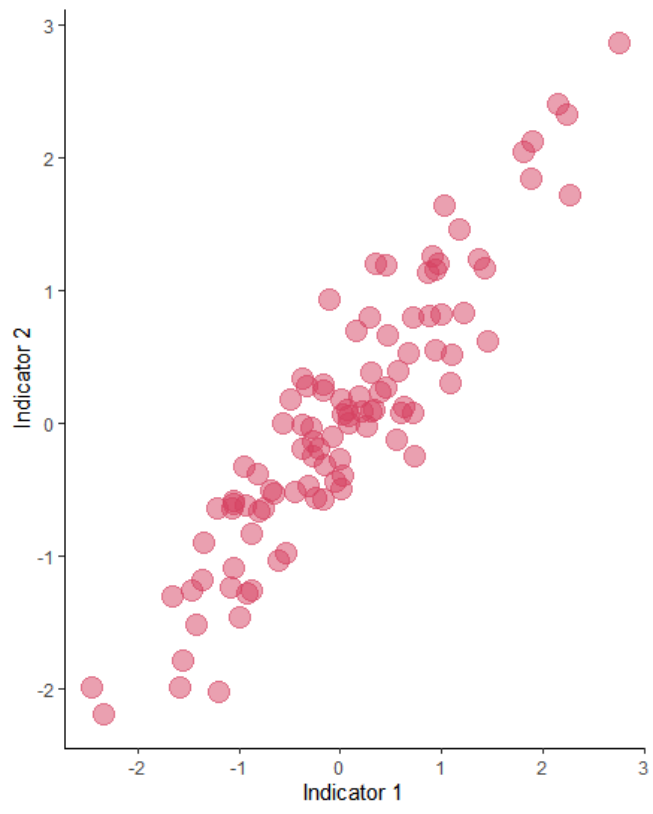
Signal and noise: Measurement error

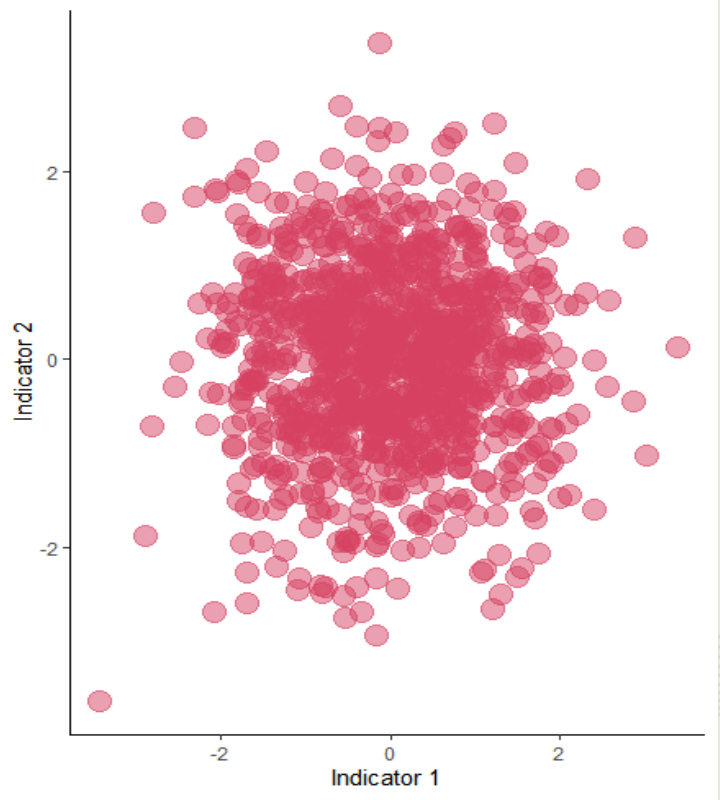
Reliability = consistency
(Without) Changes in poverty,
(Without) Changes in indicators

Noise = Error

All that variability that does not interest me

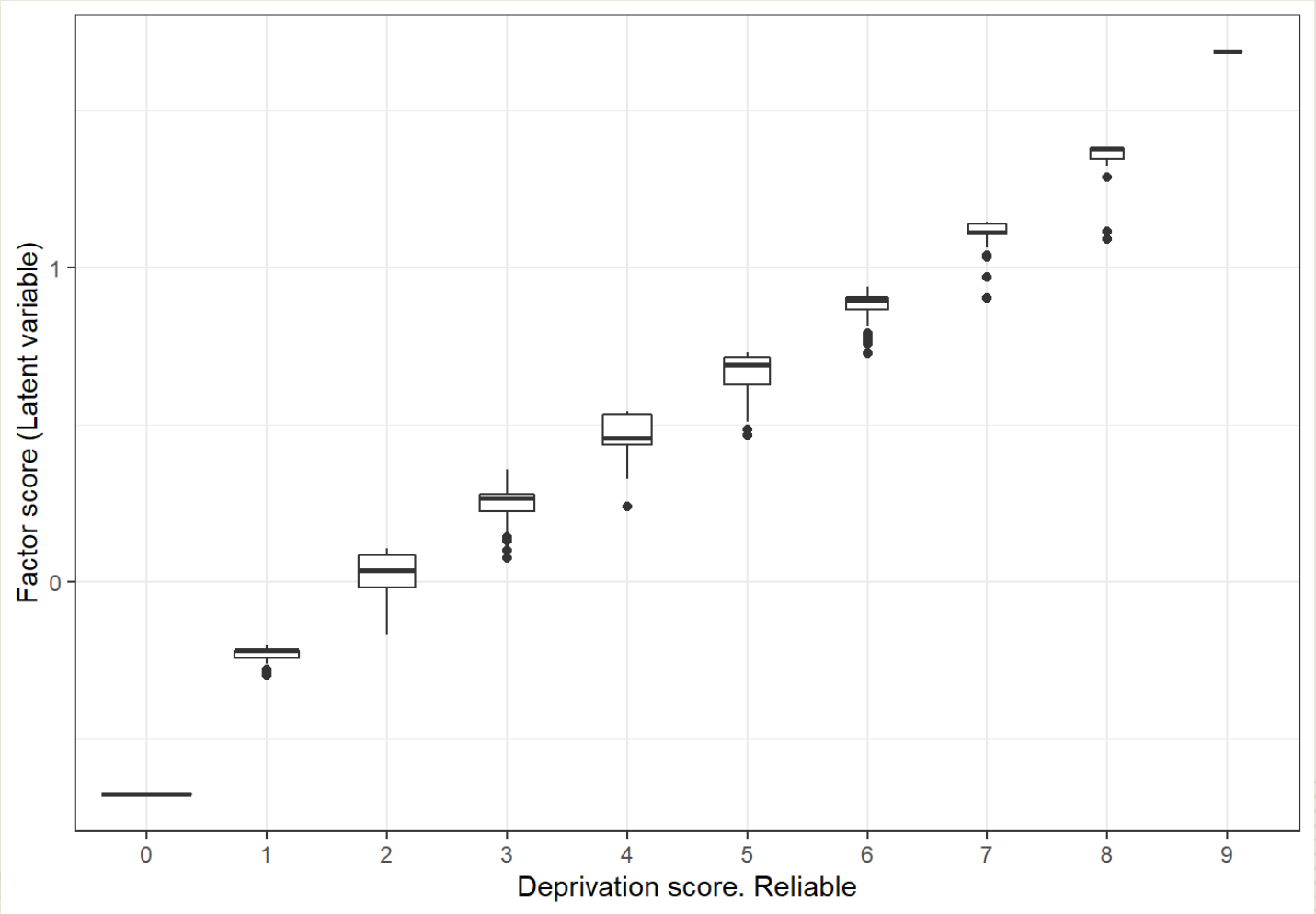
Changes in the phenomenon of interest are not recorded by my indicators



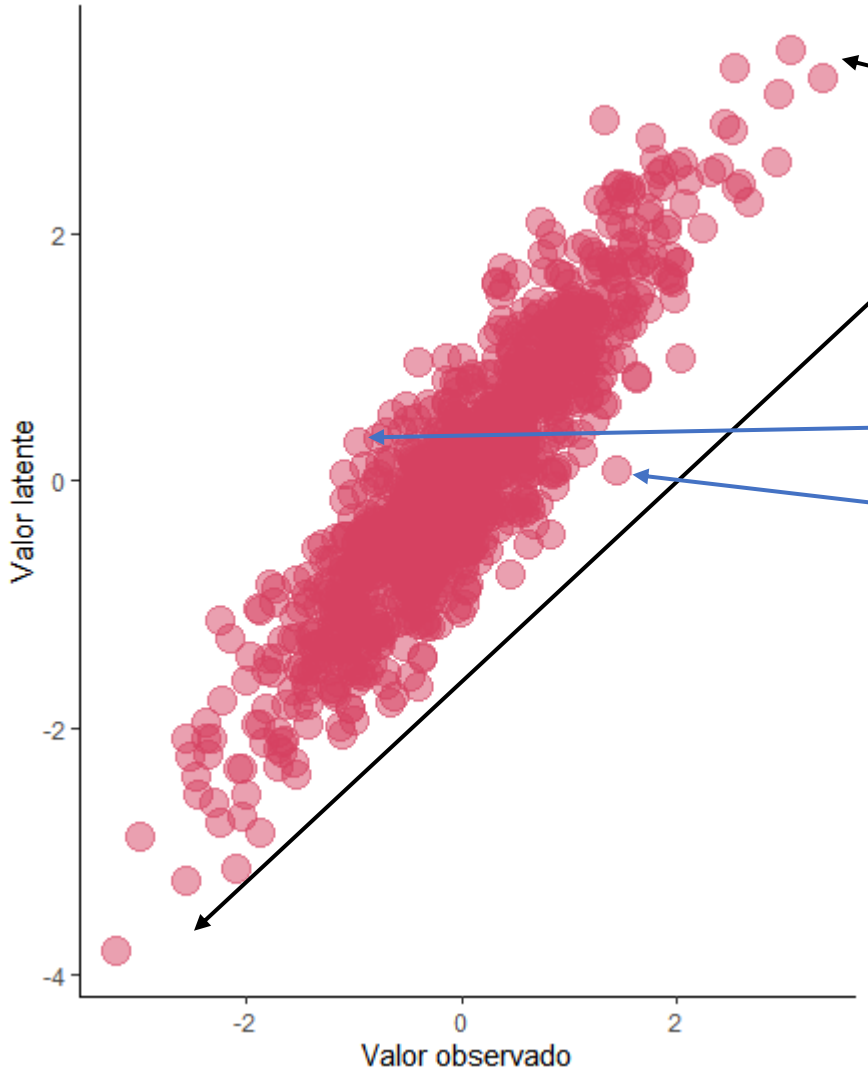


Reliability and scores

- Statistically it is understood as the signal/noise ratio of a measurement.
- Conceptually it means that the scores are homogeneous, i.e. their variability represents the same signal
- The consequence of high reliability are "clean" population orders



Reliability: It is general and punctual

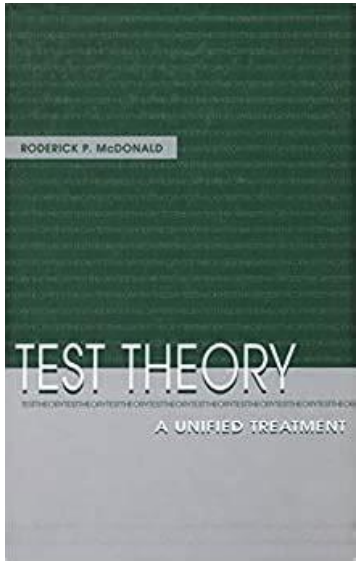


I can clearly distinguish the extremes – I don't need very high reliability–

But what if I want to make distinctions at this point given a poverty line: Different observed scores, similar levels of poverty

Poverty measures with low reliability make sense at extremes (Rurality, Indigenous, High and low educational level)

Reliability Estimators



- 1920s: Spearman-Brown –Split-half reliability-
- 1950s: Gutman y Cronbach –Lambdas y Alpha-
- 1970s: William Revelle –Beta-

Classical estimators -exploratory-

- 2000: McDonald –Omega y Omega h-.
  (It has started to be used in the last 5 years).

→ Estimator from latent variables and structural equations

Which one to use? What differences do they have?

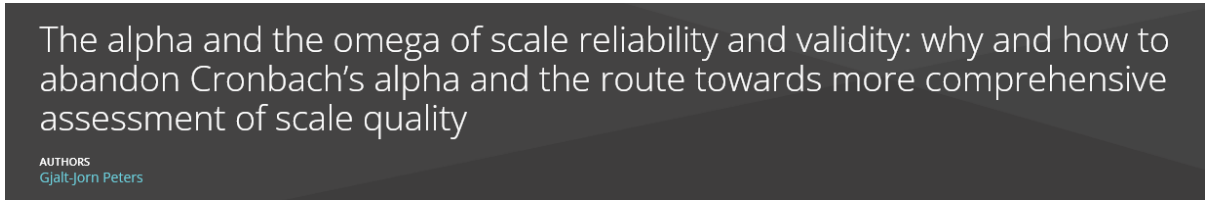
The best estimator is the one that best fits the measurement model in question:

The best estimator is omega for its power and flexibility

Omega h is useful for multidimensional measurements

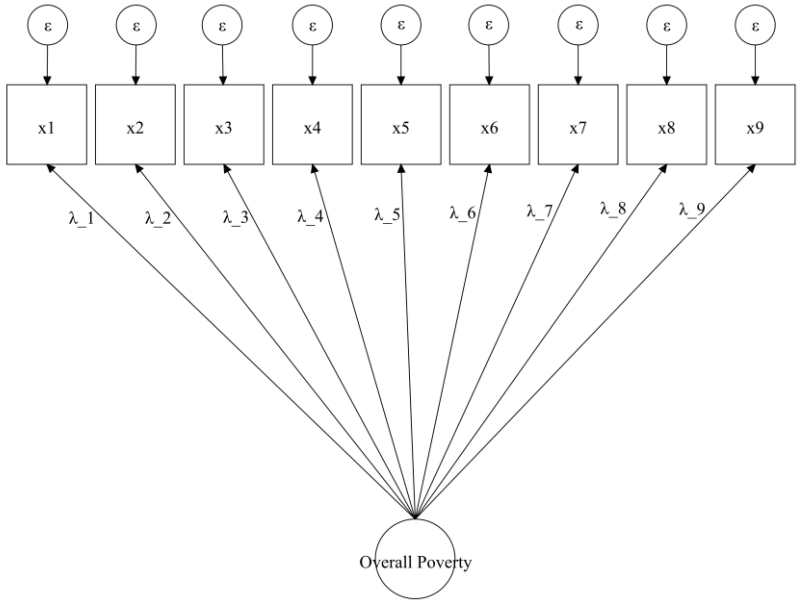
Alpha is useful under very strict measurement models

Beta is a pessimistic version of reliability – scale heterogeneity.



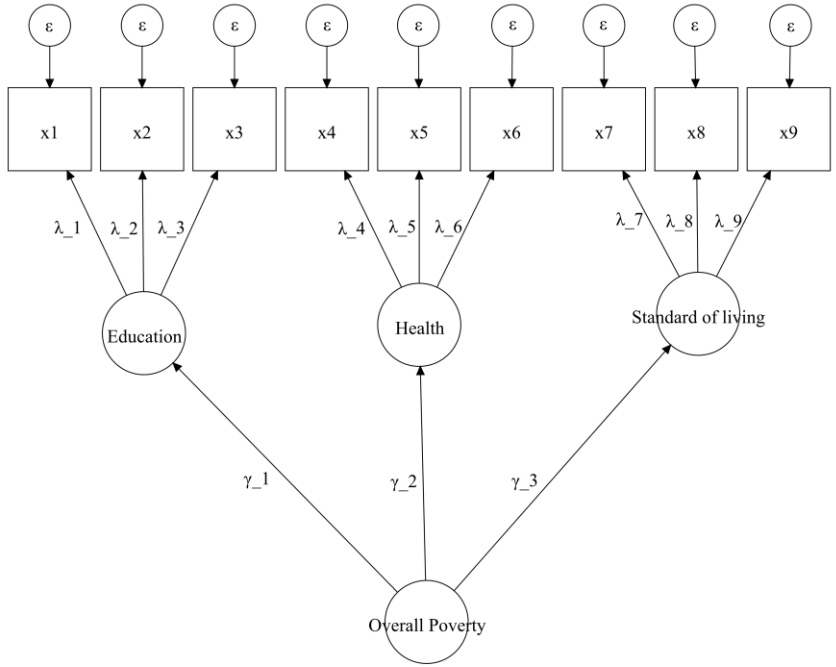
Measurement and reliability model

Alpha, beta and omega will give the same result if **λ**^′ s, **σ**, μ are equal: Parallel-tests and Tau equivalence



As we deviate from that ideal model, omega is the highest and more reliable estimator.

In social sciences we require more flexible models as our models and measurements are quite noisy



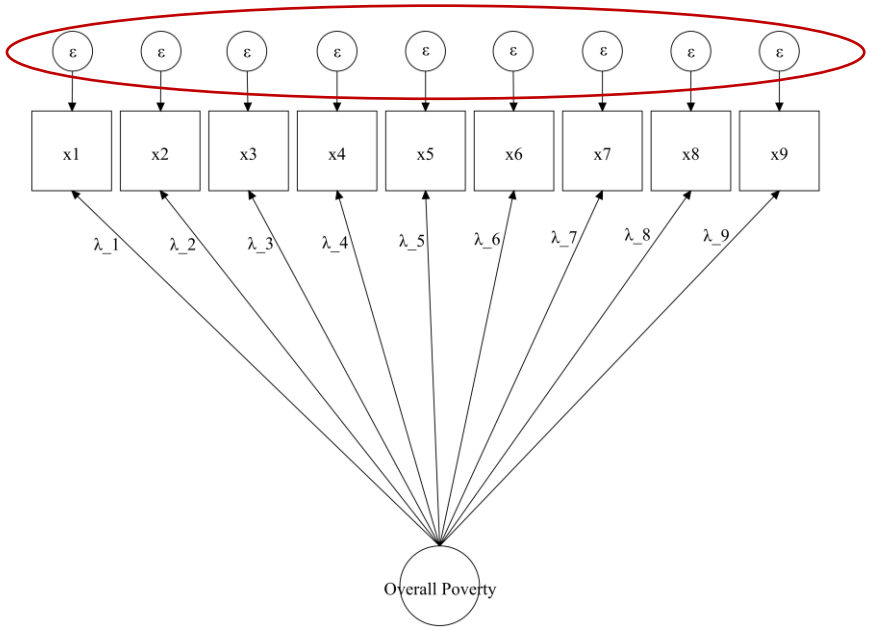
Omega y omega_h

Reliability estimation

Conditional on the measurement model, we can estimate the ratio of variance of interest to the total variance

% Noise

% of the variance explained by the phenomenon of interest



Therefore, all reliability estimators range from zero to one. Where ≈ 1 implies greater reliability of scores

How high should the reliability value be?

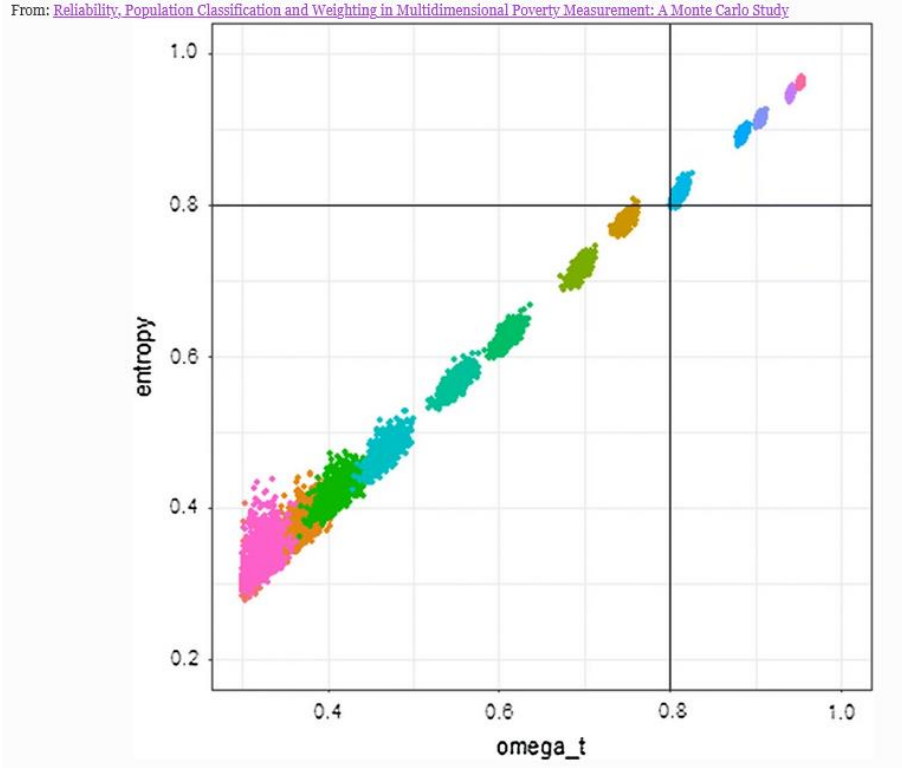
- It depends on what. Very crude classifications -high and low-, high values are not required
- Precise classifications require high values: ω>.8
- As reliability decreases – measurement error increases – our ability to distinguish people worsens:
-
    - Same score can mean different latent poverty
    - Different scores can mean similar levels of latent poverty
- Poor and non-poor classification will have higher rates of false positives and negatives – Classification error–

Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study

Héctor E. Nájera Catalán

Social Indicators Research **142**, 887–910 (2019) | Cite this article

From: Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study

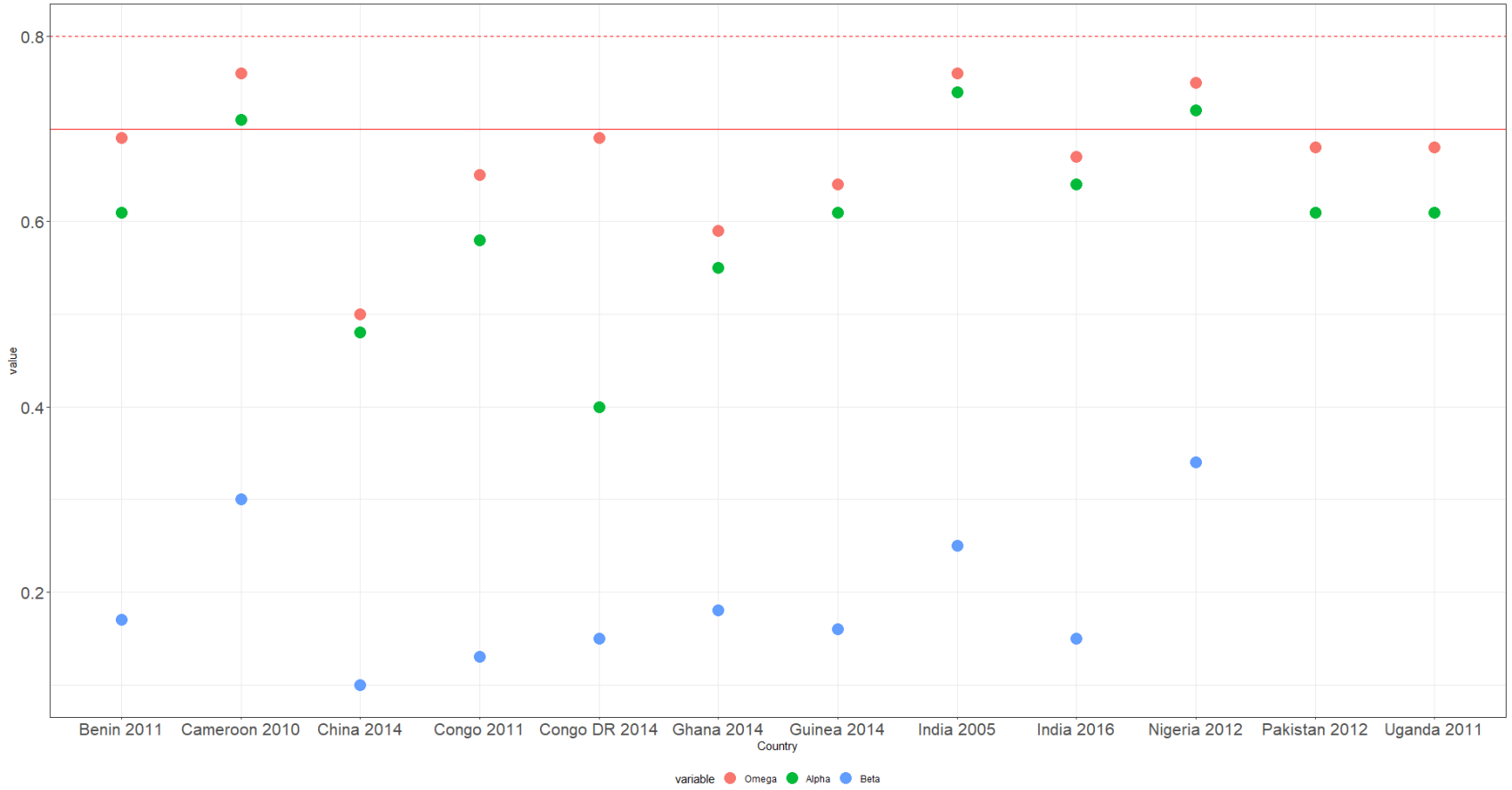


Unidimensional models ω and E

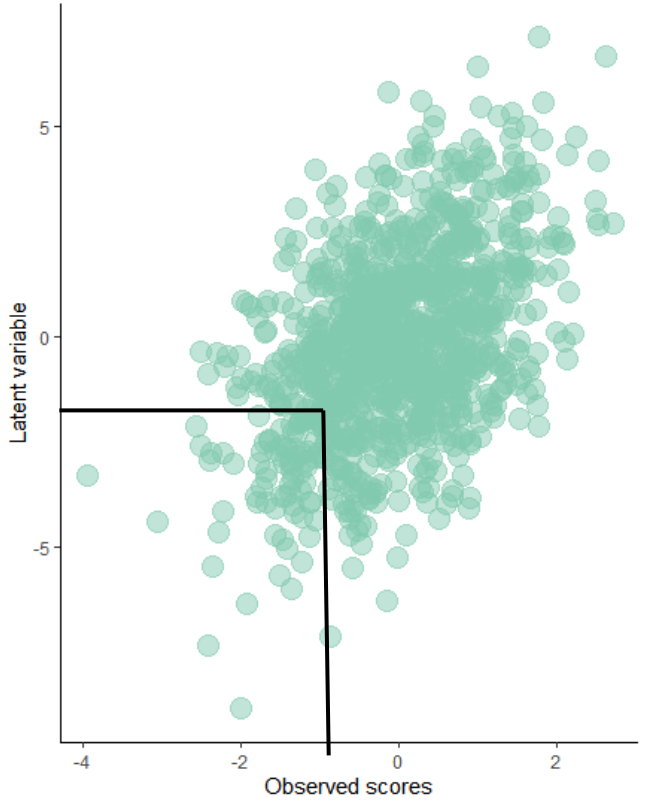
What do we know about the overall reliability of some measures?

OPHI-MPI: Modelo flexible. Confiabilidad máxima.

$\omega < .8 \approx > 10\%$
classification error



This happens in the background



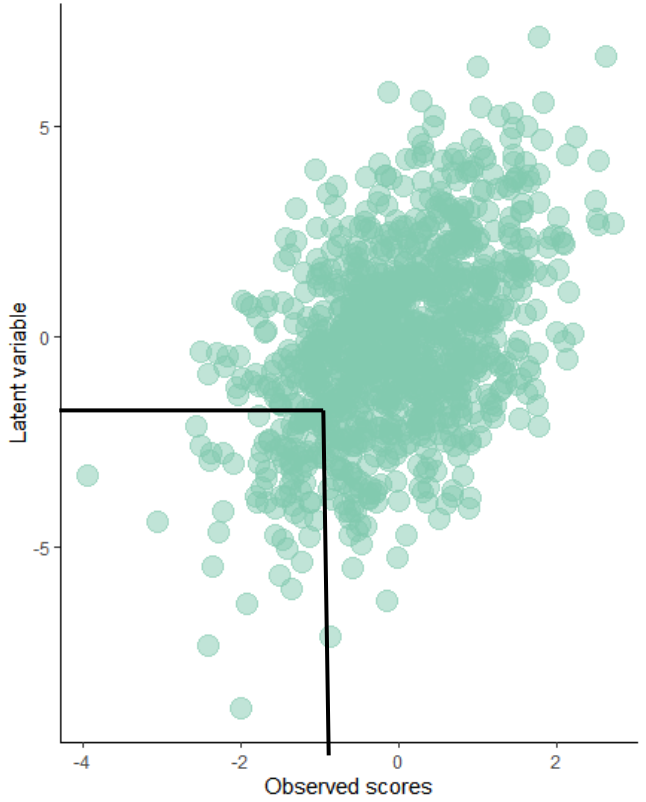
Indice de activos MPI

$\alpha < .6 \approx > 20\%$
classification error

Volmer y Alkire (2019)

	MPI-O	MPI-I	MPI-I, alt 1 (3ha)	MPI-I, alt 1 (0.3ha)	MPI-I, alt 2	MPI-I, alt 3 (3ha)	MPI-I, alt 3 (0.3ha)	MPI-I, alt 4 (3ha)	MPI-I, alt 4 (0.3ha)	MPI-N, version 1	MPI-N, version 2	MPI-N, version 3	MPI-N, version 3 minus bicycle and animal cart
Pooled	0.583	0.4776	0.5360	0.52	0.4866	0.4970	0.4897	0.5146	0.4969	0.742	0.7034	0.6129	0.6779
Armenia	0.2356	0.2973	0.2074	0.2469	0.286	0.2233	0.2492	0.2071	0.2463	0.513	0.4982	0.3087	0.3172
Angola	0.6896	0.4964	0.5651	0.5364	0.5066	0.5319	0.5084	0.5605	0.5354	0.7627	0.7365	0.6972	0.7531
Bangladesh	0.4523	0.4667	0.4103	0.4407	0.4785	0.4103	0.4407	0.4103	0.4407	0.5727	0.5333	0.5155	0.54
Brazil	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.5753	0.5753	0.4577	0.4577
DR Congo	0.6256	0.4759	0.518	0.4796	0.4671	0.5078	0.471	0.5152	0.48	0.6982	0.6372	0.638	0.7105
Côte d'Ivoire	0.511	0.4444	0.4932	0.4658	0.4643	0.4919	0.4644	0.4906	0.4627	0.6346	0.5586	0.5273	0.6195
Colombia	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.6703	0.6781	0.6238	0.6073
Egypt	0.2954	0.3795	0.305	0.305	0.383	0.383	0.383	0.305	0.305	0.382	0.3601	0.3471	0.3982
Ethiopia	0.6398	0.4028	0.4883	0.4303	0.4175	0.487	0.4292	0.4876	0.43	0.6636	0.6753	0.6651	0.6814
Guatemala	0.6611	0.5167	0.547	0.5366	0.5369	0.5473	0.5367	0.5467	0.5363	0.7434	0.7291	0.6659	0.7126
Haiti	0.6338	0.4333	0.4898	0.4995	0.4432	0.4901	0.4997	0.4898	0.4995	0.691	0.6767	0.6302	0.6829
India	0.5534	0.4905	0.5255	0.515	0.4905	0.4848	0.4944	0.5195	0.5104	0.7251	0.6757	0.5795	0.6567
Indonesia	0.6811	0.4935	0.5282	0.4896	0.4868	0.5244	0.4889	0.527	0.488	0.6829	0.6821	0.6306	0.6702
Kenya	0.5833	0.5035	0.5046	0.5106	0.5019	0.5028	0.5106	0.5038	0.5089	0.5158	0.5207	0.5538	0.5755
Cambodia	0.5675	0.4429	0.4612	0.4573	0.4496	0.4611	0.4573	0.4611	0.4573	0.5739	0.5464	0.5039	0.5802

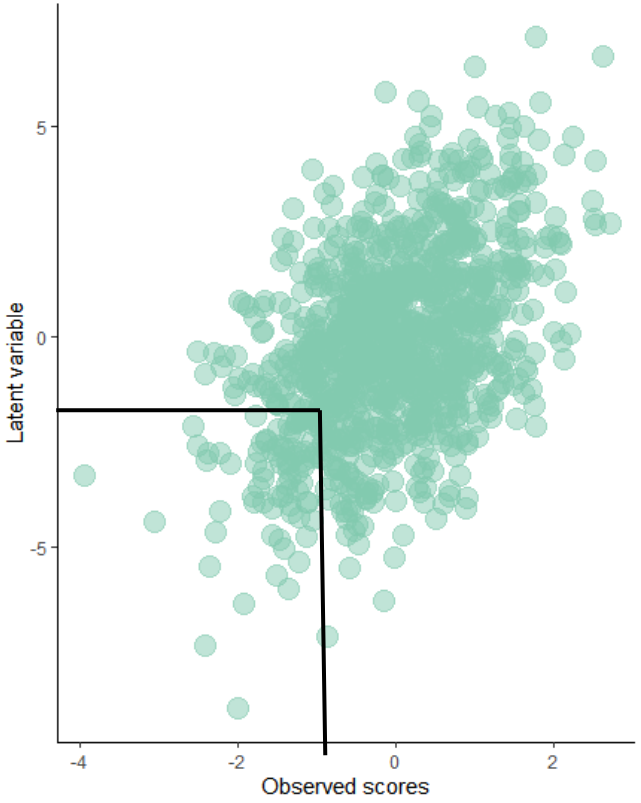
This happens in the background

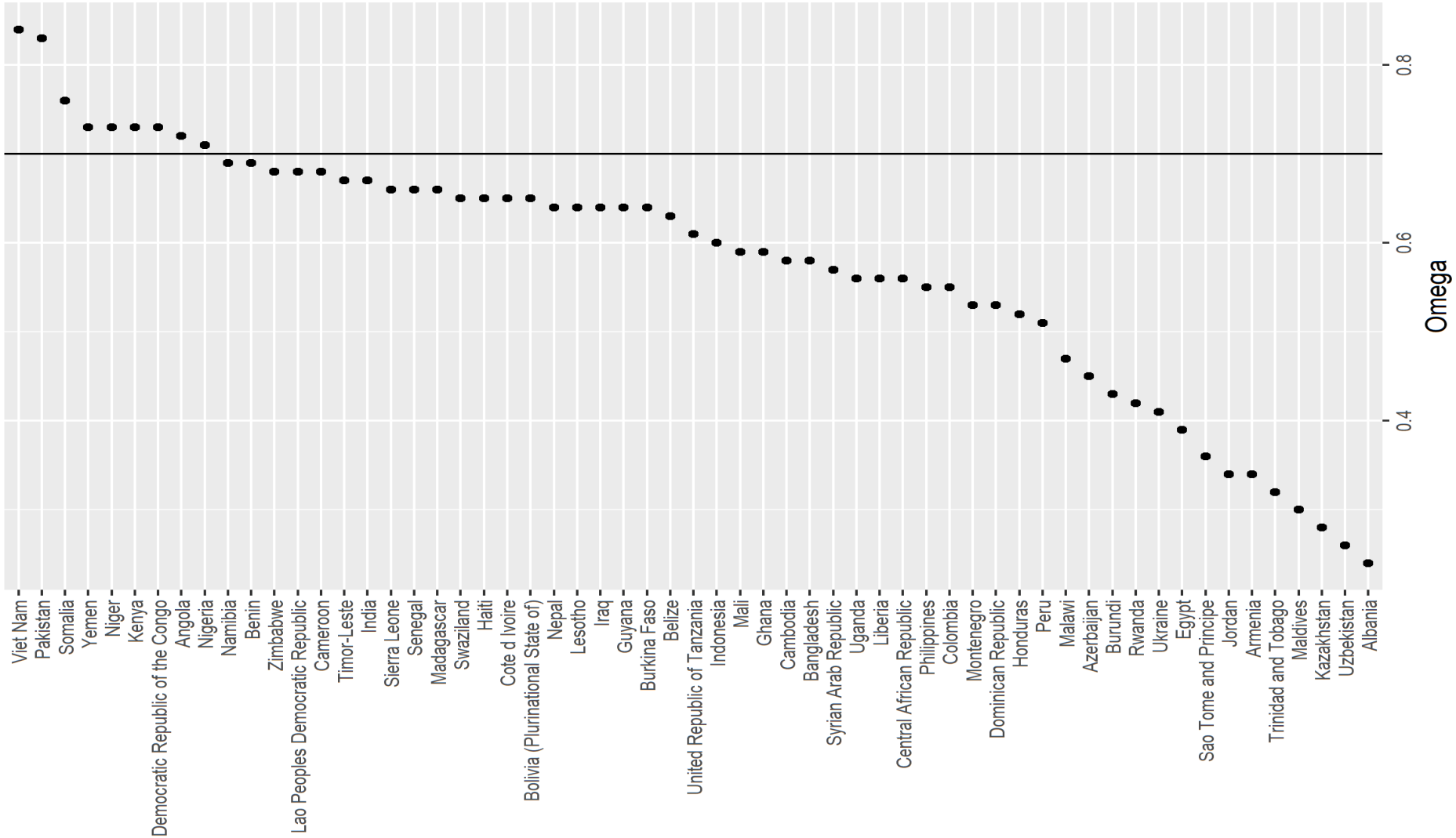


UNICEF 2004

$\omega < .7 \approx > 15\%$
classification error

This happens in the background

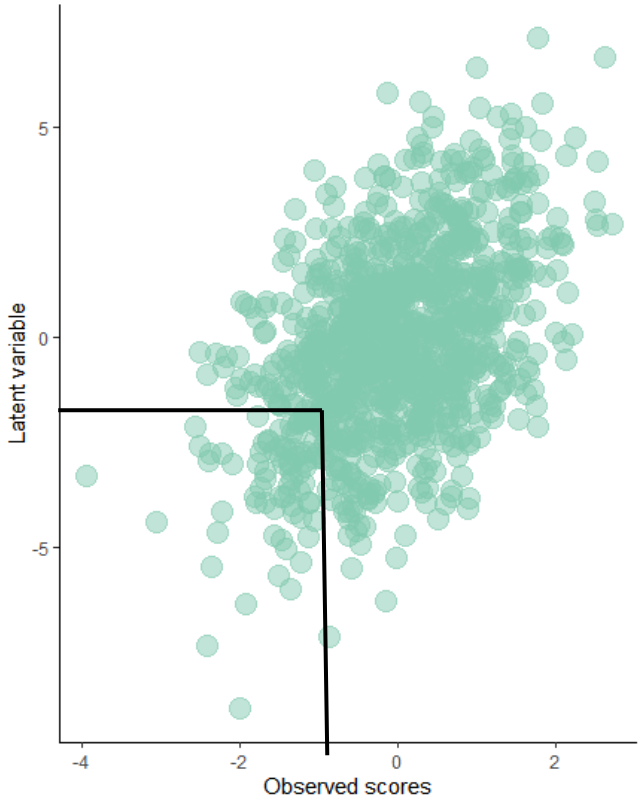




MPI-LA

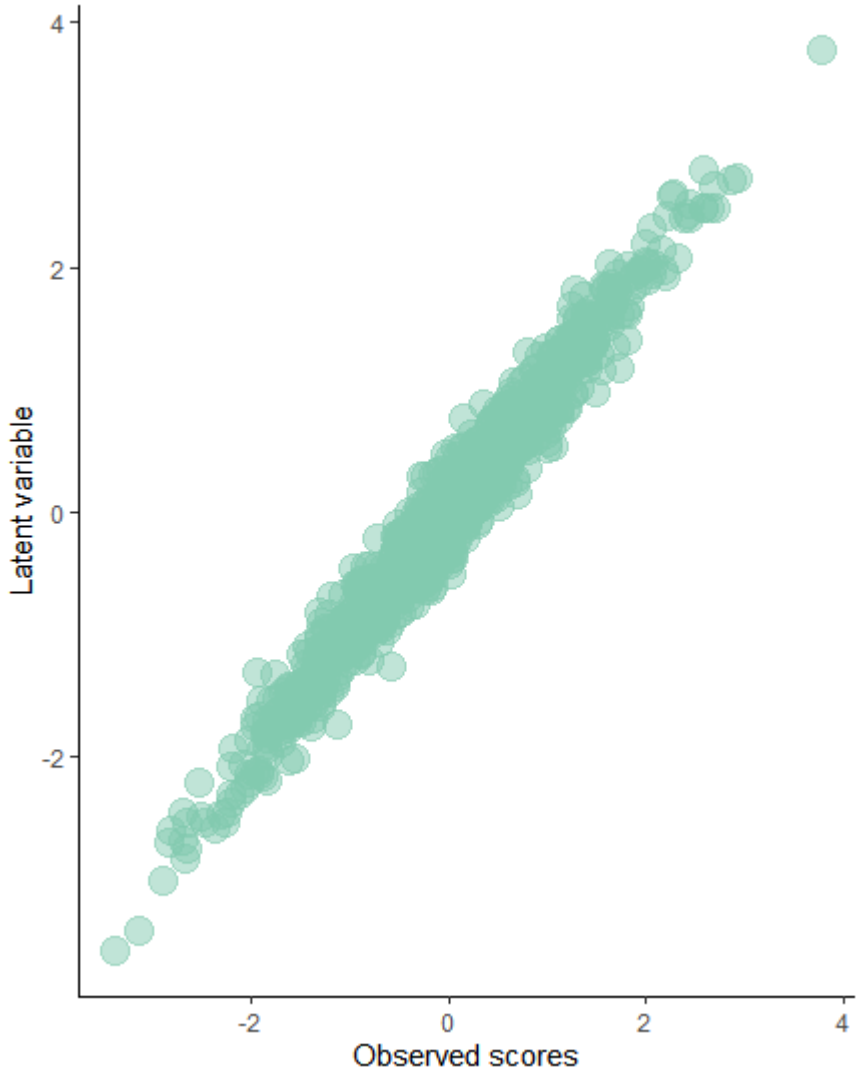
$\omega < .6 \approx > 20\%$
classification error

Esto pasa en el fondo



Country	Year	α	β	ω
Argentina	2005	0.63	0.38	0.71
	2012	0.51	0.32	0.60
Bolivia	2003	0.64	0.56	0.68
	2012	0.65	0.34	0.76
Brazil	2005	0.52	0.04	0.62
	2012	0.45	0.00	0.57
Chile	2003	0.46	0.35	0.58
	2011	0.27	0.11	0.33
Mexico	2004	0.75	0.07	0.81
	2012	0.64	0.24	0.69
Uruguay	2005	0.54	0.02	0.67
	2012	0.43	0.38	0.54

Relative poverty: Europe. Consensual method



$\omega > .9 \approx < 5\%$
error de clasificación

Table 15: Omega Analysis, national level, 2014

	Unidimensional		Townsend		
	Omega	BIC	Omega	Omega_h	BIC
Belgium	0.97	83593	0.95	0.65	86226
Bulgaria	0.94	155878	0.93	0.63	160109
Czech Republic	0.95	125497	0.95	0.64	127748
Denmark	0.94	73381	0.94	0.64	74320
Germany	0.94	188040	0.93	0.64	192065
Estonia	0.93	111990	0.93	0.65	113873
Ireland	0.94	124365	0.94	0.65	126486
Greece	0.92	226947	0.93	0.63	230458
Spain	0.96	245426	0.94	0.65	250931
France	0.95	182346	0.94	0.64	186617
Croatia	0.94	121422	0.94	0.64	123073
Italy	0.96	385840	0.94	0.65	397204
Cyprus	0.93	107605	0.93	0.64	109133
Latvia	0.92	157885	0.92	0.63	158775
Lithuania	0.92	123433	0.92	0.63	125872
Luxembourg	0.97	39261	0.95	0.65	40550
Hungary	0.95	242356	0.93	0.64	247061
Malta	0.94	109760	0.92	0.64	111200
Netherlands	0.95	127708	0.95	0.64	130432
Austria	0.96	70340	0.94	0.64	71973
Poland	0.94	310111	0.93	0.64	315812
Portugal	0.94	159328	0.94	0.65	162471
Slovenia	0.94	207387	0.94	0.64	211409
Slovakia	0.93	139511	0.93	0.64	141971
Finland	0.95	126804	0.94	0.64	128633
Sweden	0.93	53870	0.93	0.64	54715

Source: EU-SILC 2014 cross-sectional data, authors' computation.

México

The quality of multidimensional measurement in Mexico has declined overtime

This was foreseen in the guidelines of the official measurement

A first adjustment has been already made in 2018 but we do not know if it was sufficient to reduce random noise

Cuadro 1. Estimación del error de clasificación del índice de derechos sociales. México 2008-2018

Año	Omega
2008	0.75
2010	0.72
2012	0.71
2014	0.72
2016	0.68
2018	0.68

Mexico: Classification error

Thanks to theoretical developments and computational advances, today we can estimate the classification error directly with different methods

Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left

Jerry Hausman

JOURNAL OF ECONOMIC PERSPECTIVES
VOL. 15, NO. 4, FALL 2001
(pp. 57-67)

Journal of Econometrics
Volume 87, Issue 2, December 1998, Pages 239-269

Misclassification of the dependent variable in a discrete-response setting

J.A. Hausman a, Jason Abrevaya b, F.M. Scott-Morton b

Cuadro 1. Estimación del error de clasificación del índice de derechos sociales. México 2008-2018

Año	Omega	Error FN [ICr 95%]	Error FP [ICr 95%]
2008	0.75	6 [2-10]	1 [0-2]
2010	0.72	8 [5-11]	1 [0-2]
2012	0.71	11 [8-14]	1 [0-3]
2014	0.72	9 [6-12]	1 [0-2]
2016	0.68	13 [11-16]	1 [0-2]
2018	0.68	14 [11-17]	0 [0-1]

Misclassification error, binary regression bias and reliability in multidimensional poverty measurement: An estimation approach based on Bayesian modelling.

Héctor Nájera (Manuscript accepted for publication. Measurement Journal)

Item Reliability: Item Response Theory

Reliability analysis can help us identify problematic items:

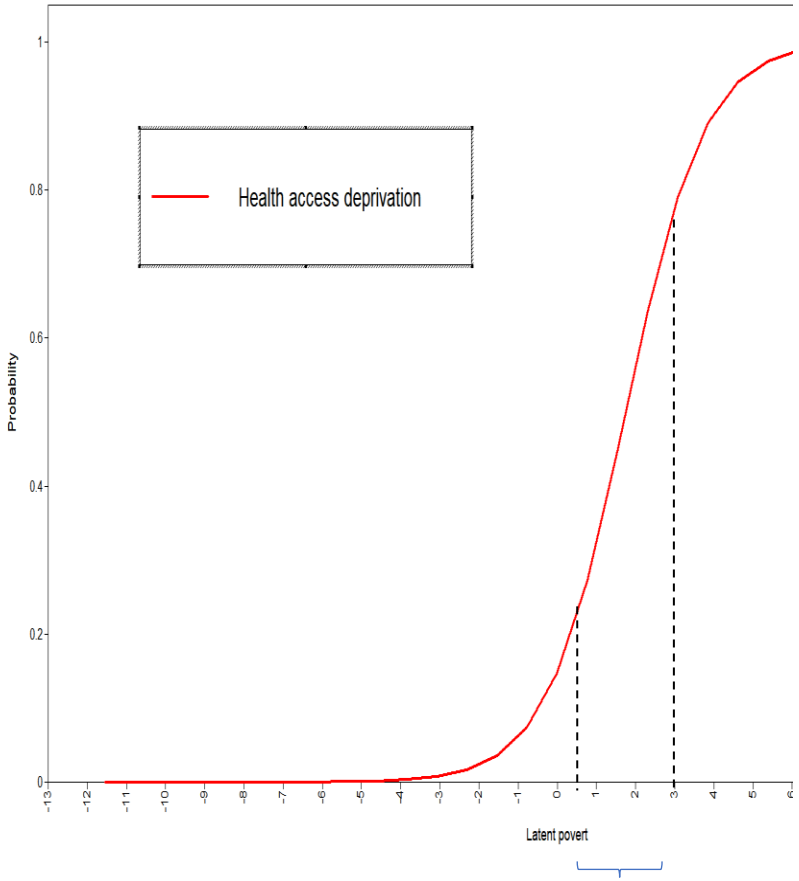
- Those whose variability does not come from the phenomena of interest (Discrimination)

- those whose variability is so low, that they do not provide information (Severity)

It models the probability of having a deprivation conditional on: The latent level of poverty, the discrimination of the item and its severity.

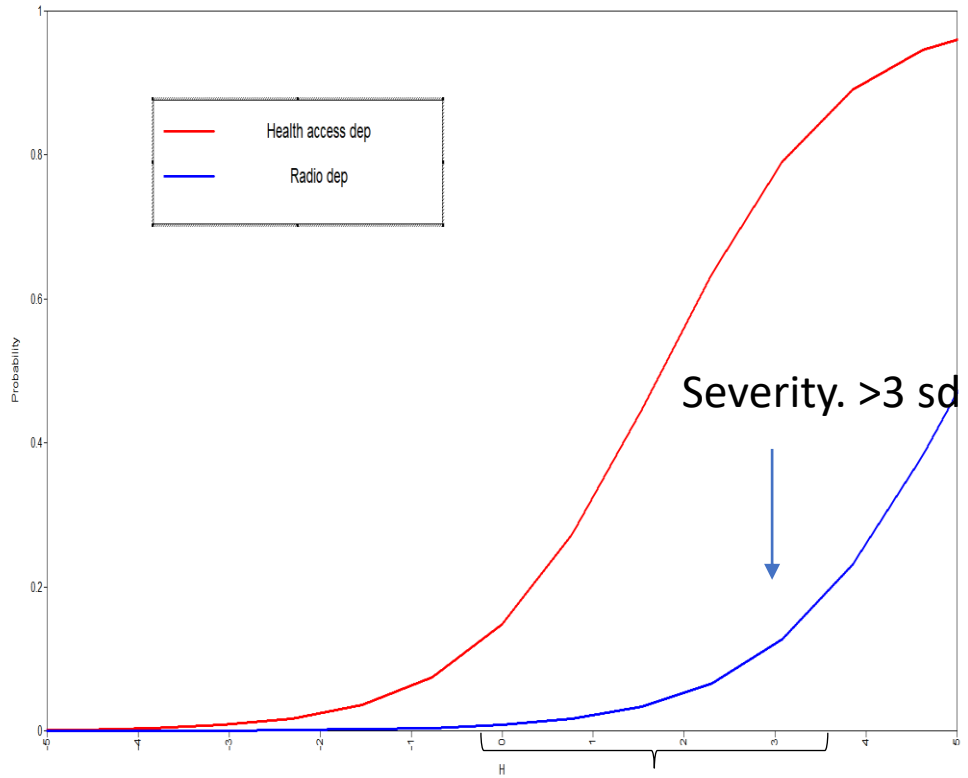
Item response theory

Lack of effective access to health services



Changes in latent poverty lead to changes in probability

Lack of radio



Changes in latent poverty don´t lead to important changes in probability

Examples

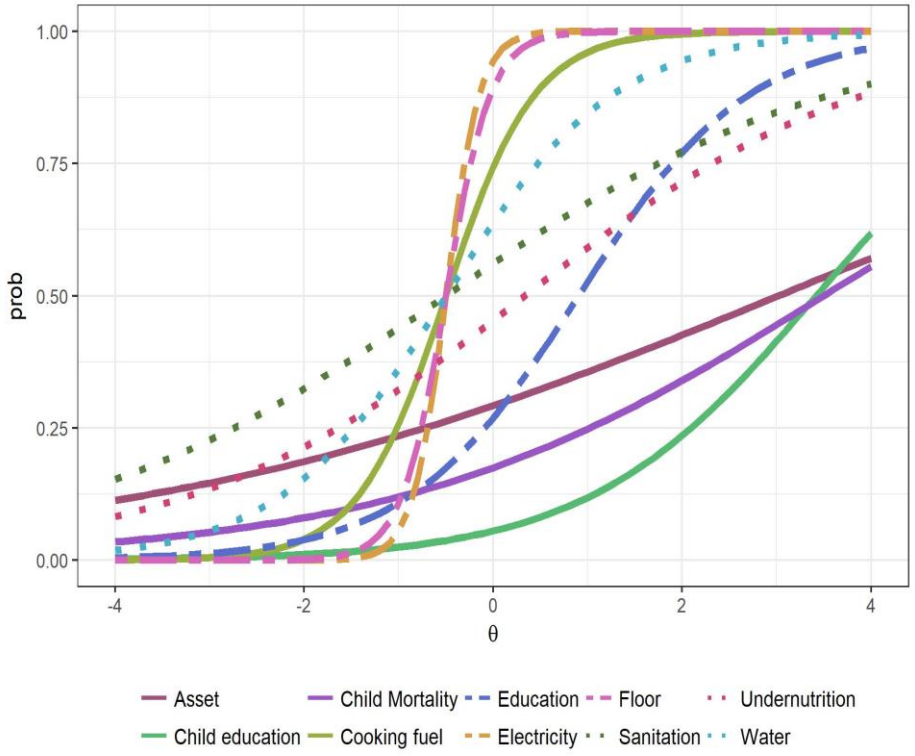
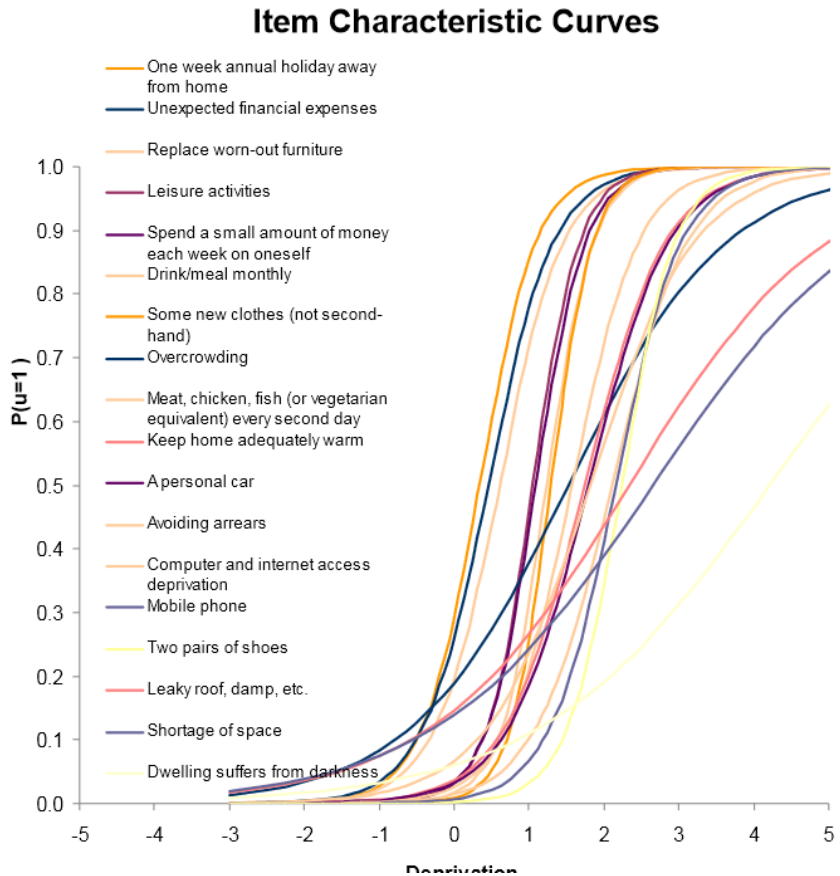


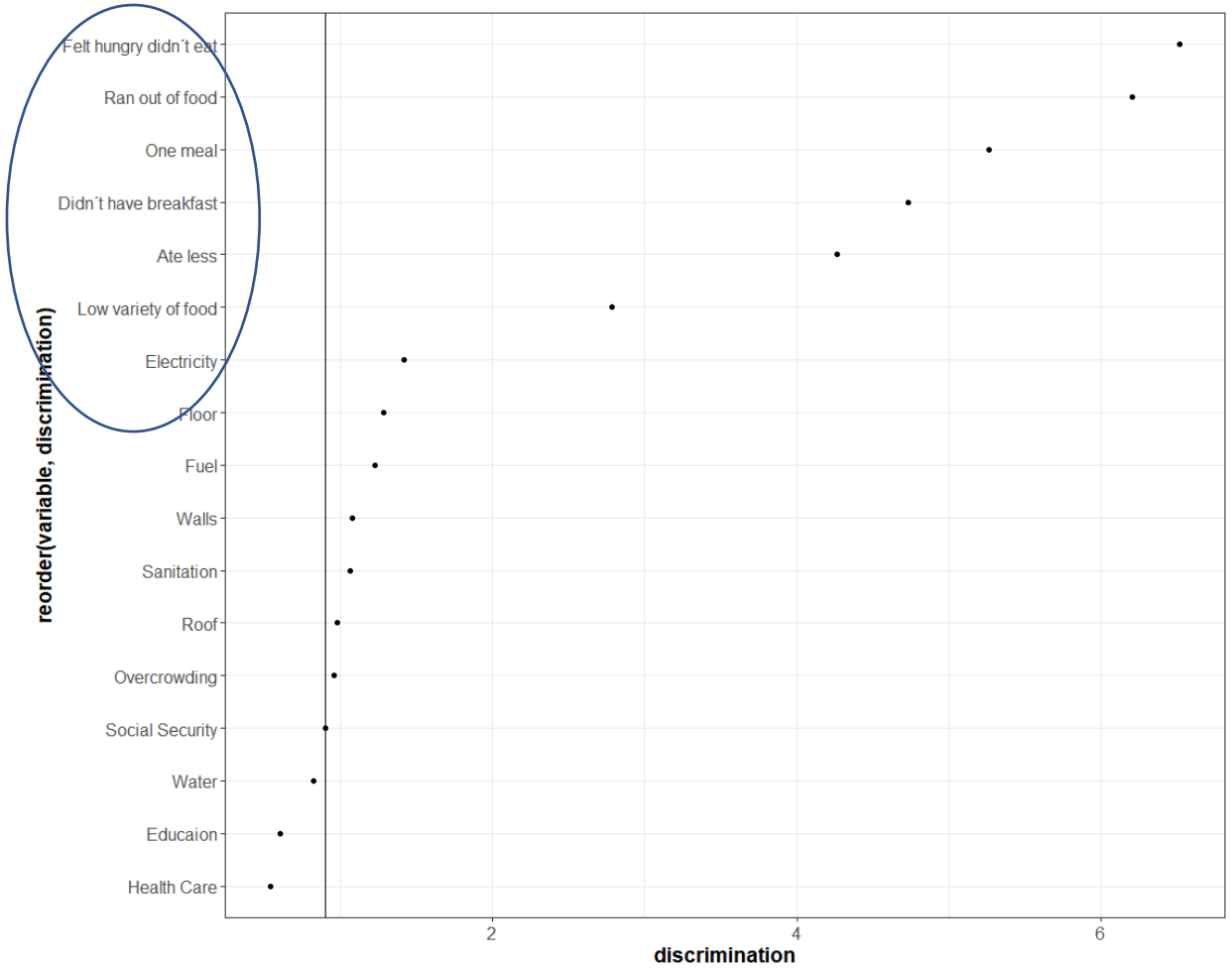
Figure 8: IRT results for the 18 items retained for the whole population MD indicator after the validity and reliability (Cronbach's alpha) tests, 2009



Results Model 1: 2P-IRT. Discrimination

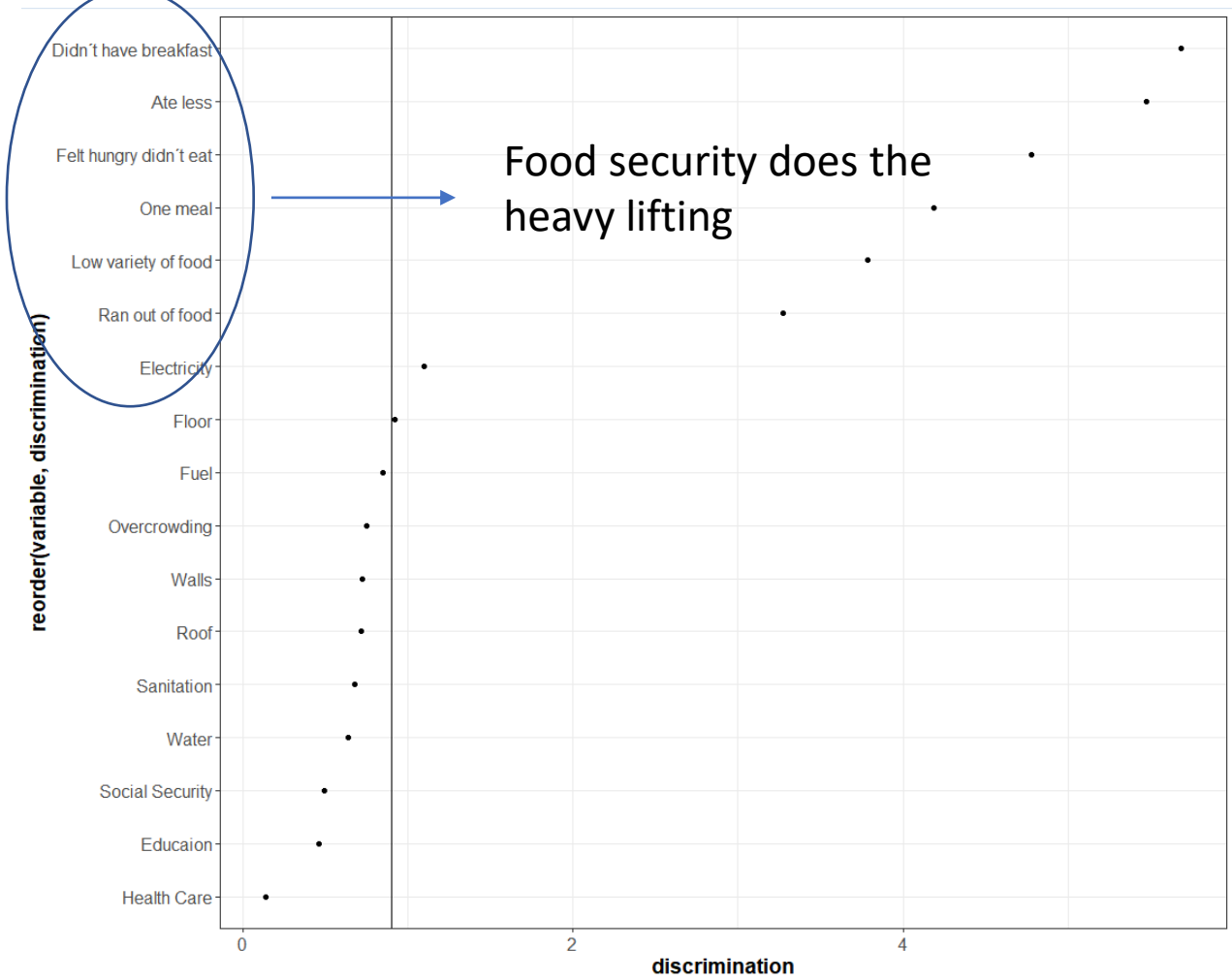
2008

>.9

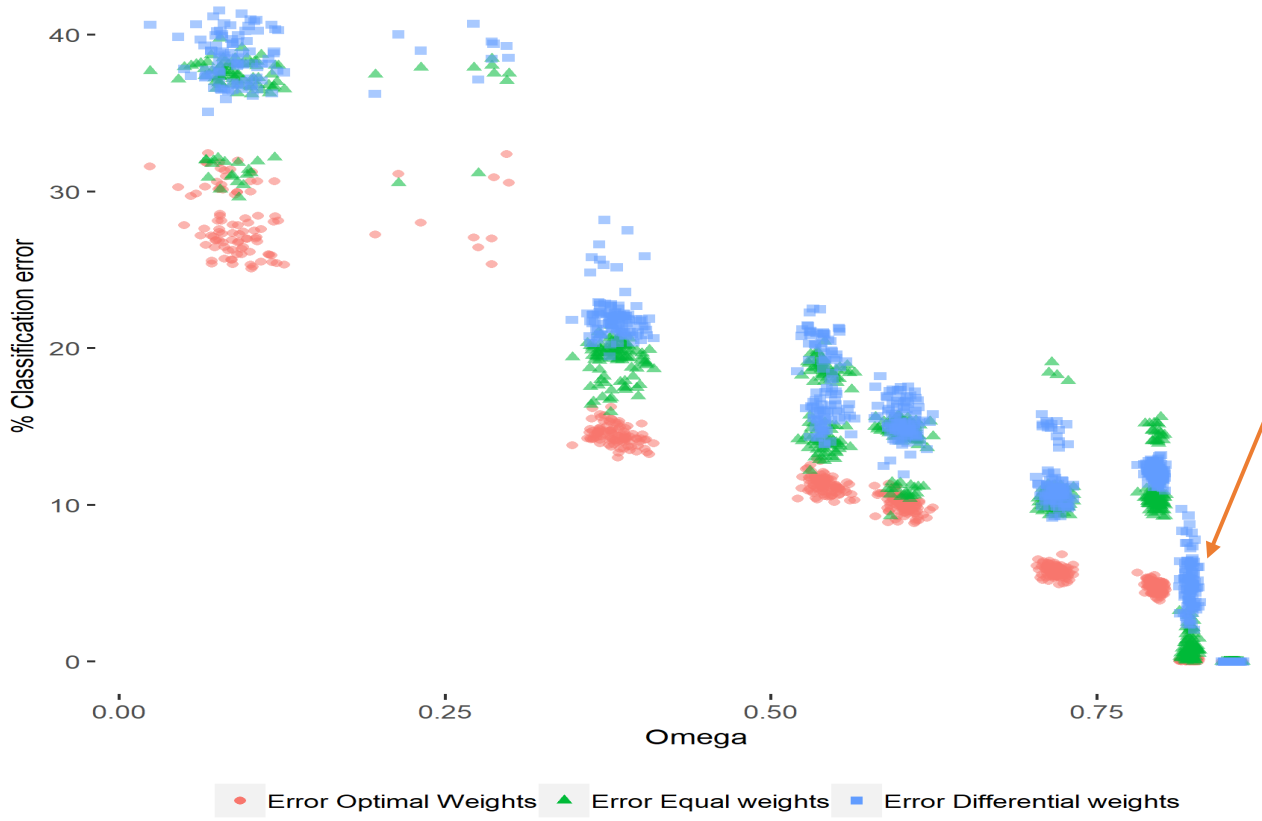


2018

>.9



What are you weighting for?



In the context of high reliability differential weights do very little:

Observed scores correlate highly with latent scores

Always differential weights do more damage

Weights may help in cases of low reliability – but how to specify them–

Weights can have other objectives and not just classification

Nájera et al. (en curso)

What are you weighting for? Misconceptions

Endogenous weights and multidimensional poverty: A cautionary tale ☆

Indranil Dutta [a], Ricardo Nogales [b], Gaston Yalonetzky [c]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jdeveco.2021.102649 Get rights and content

Highlights

- A large and growing body of work uses endogenous (data driven) weights to compute multidimensional poverty.

Articles

The Importance of Reliability and Construct Validity in Multidimensional Poverty Measurement: An Illustration Using the Multidimensional Poverty Index for Latin America (MPI-LA)

Héctor E. Nájera Catalán & David Gordon

Pages 1763-1783 | Received 07 Feb 2018, Accepted 01 Mar 2019, Published online: 26 Sep 2019

The example they use is particular and with low reliability:

Endogenous weights – like any other type of weights – are harmful in the context of low reliability.

In contexts of high reliability the "endogenous" weights are optimal and the equal weights are better than the "ad hoc" ones.

It is false that an important and growing part uses differential weights, much less in contexts of high reliability

The consensual method deprivation scores are simple.

Conclusiones

- Reliability is a property of scores and is a necessary condition for any measurement exercise
- Reliability is a necessary condition for the classification of populations
- Reliability is not enough Why?
- We haven't said anything about the latent variable
- A set of variables can have high reliability and tell us nothing about poverty
- It is the validity that allows us to speak of the representation of the phenomena of interest.
-