Gender Inequality: exploring the gap using poset approach

VI European Congress of Methodology

A. M. Di Brisco

co-authors
G. Bertarelli, F. Mecatti

University of Milano Bicocca

g.bertarelli@campus.unimib.it, a.dibrisco@campus.unimib.it

July 25, 2014

Overview

Partial order

"The use of partial ordering has two different types of justification in interpersonal comparison or in inequality evaluation. First, as has been just discussed, the ideas of well-being and inequality may have enough ambiguity and fuzziness to make it a mistake to look for a complete ordering of either. This may be called the "fundamental reason for incompleteness". Second, even if it is not a mistake to look for one complete ordering, we may not be able in practice to identify it.
The **pragmatic reason for incompleteness** *is to sort out unambiguously, rather than maintaining complete silence until everything has been sorted out and the world shines in dazzling clarity"*
(A. Sen, Inequality reexamined, 1998)

Partial order

- Dealing with ordinal variables: a problematic issue in synthetic indexes computation;
- The mathematical theory of partial order allows to respect the ordinal nature of the data, avoiding any aggregation or scaling procedures .

Aims of the presentation:

1. Introduce some basic concepts of partial order theory
2. Explore the robustness of poset approach with respect to the choice of a threshold.
3. Provide a gender gap measure.

Definition of POSET

A finite partially ordered set $P = (X \leq)$ (POSET) is a finite set X with a partial order relation $\leq$ that is a binary relation satisfying the following properties:

1. Reflexivity: $x \leq x \; \forall x \in X$;
2. Antisymmetry: if $x \leq y$ and $y \leq x$ then $x = y$ for $x, y \in X$;
3. Transitivity: if $x \leq y$ and $y \leq z$ then $x \leq z$ for $x, y, z \in X$.

Two elements of set X are comparable if $x \leq y$ or $y \leq x$.
If any two elements of X are comparable then the poset P is said a **chain** or a linear order.
If any two elements of X are not comparable then the poset P is said an **antichain**.

Upset and Downset

An **upset** U of a poset is a subset of P such that if $x \in U$ and $x \leq z$ then $z \in U$.
A **downset** D of a poset is a subset of P such that if $x \in D$ and $y \leq x$ then $y \in D$.

Proposition

Given a finite poset P and an upset U then $\exists \bar{u}$ antichain such that $\bar{u} \subseteq P$.
Then $z \in U$ if and only if $\exists u \in \bar{u}$ such that $u \leq z$.
The upset U, then, is generated by an antichain $\bar{u}$:

$$U = \bar{u} \uparrow$$

Linear extentions

An *extention* of a poset P is a poset defined on the same set X whose set of comparabilities comprises that of P

Definition

A **linear extension** is an extension of P that is a linear order or a chain.

Theorem, Neggers and Kim, 1988

The set of linear extensions of a finite poset P uniquely identifies P

Ordinal variables

Let us consider k ordinal variables each with j_k levels.
Then we can compute all the possible profiles and provide a partial order with the following :

Rule

Let s and t two profiles over $v_1, \ldots, v_k$ ordinal variables. Then t dominates s if and only if

$$v_i(s) \leq v_i(t) \qquad \forall i = 1, \ldots, k$$

How many profiles:

$$\#p = \prod_{i=1}^{k} j_k$$

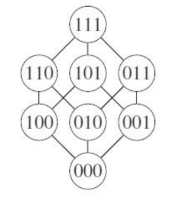
Hasse Diagram

An Hasse diagram is a graph in which:

- if $s \leq t$ the node t is placed above the node s
- if $s \leq t$ and $\nexists w \; : \; s \leq w \leq t$ then an edge is inserted

Example

Let consider three binary variables on a 0-1 scale. There are 8 possible profiles represented in the following graph:



Threshold

Objective: classification of the profiles in deprived and not deprived

- Define a threshold as a profile or a list of profiles that is an anti chain
- The threshold generates a down set D
- All profiles belonging to D are certainly under the threshold
- Some profiles can not be ordered with respect to the chosen threshold

Partial Orded

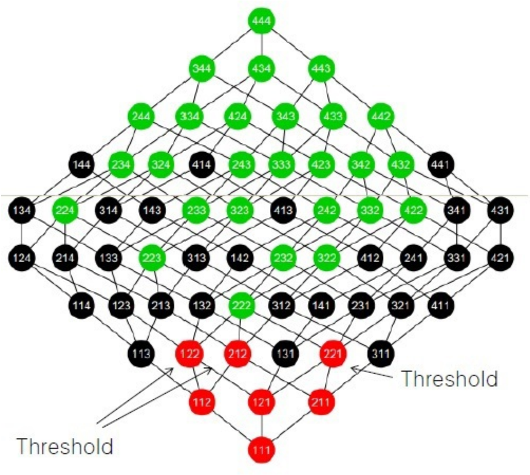


Figure: Deprived, Not Deprived, Ambiguous

Evaluation function

$$\eta : T \rightarrow [0, 1]$$

where T is the finite collection of all possible profiles.

$$\eta(p) = \begin{cases} 0 & if \quad p \in D \\ \frac{\{l \in E(P) : \exists d \in \underline{d} : d \leq p \in l\}}{|E(P)|} & otherwise \\ 1 & if \quad p \in U \end{cases}$$

where $E(P)$ is the set of all possible linear extensions of the poset
$\underline{d}$ is the anti chain selected as threshold.

$\Longrightarrow$ *complete order*

Gap measure

A synthetic measure of gap is computed as follows:

$$G = \sum_{p=1}^{\#p} w_p dist(p)$$

$$dist(p) = \frac{|p - \bar{p}|}{M}$$

where

- w_p is a weight assigned to each profile, equal to the relative frequency of subjects sharing the same profile
- $\bar{p}$ is the first profile greater than the threshold
- M is the absolute distance of the first profile greater than the threshold from the minimal element.

Gap measure

Note: The absolute distance is set equal to 0 when a profile is greater than the threshold.

Interpretation of gap measure:

- Measure the severity of deprivation;
- Is basically a measure of the fraction of people a subject must overtake to exit deprivation;
- It depends upon the distribution on the graph of profiles.

Dataset

Data Source:
ESS1-5, European Social Survey Cumulative File Rounds 1-5 edition 1.1
ESS Round 3: European Social Survey Round 3 Data (2006).
Nations: Italy, Spain, Norway, Estonia, Netherlands.
Technical notes:

- The European countries have been selected as comparable for cultural and social aspects and data availability.
- Missing values and "Refuse to answer" variables were deleted.
- Gender is not an ordinal variable.

Available variables

Political Party

Worked in political party or action group last 12 months:

- 1 $\rightarrow$ No
- 2 $\rightarrow$ Yes

Social Meetings

How often socially meet with friends, relatives or colleagues

- 1 $\rightarrow$ less than once a month or once a month
- 2 $\rightarrow$ several times a month
- 3 $\rightarrow$ several times a week

Available variables

Education

Highest level of education.

- 1 $\rightarrow$ less than secondary education
- 2 $\rightarrow$ completed secondary education
- 3 $\rightarrow$ more than secondary education

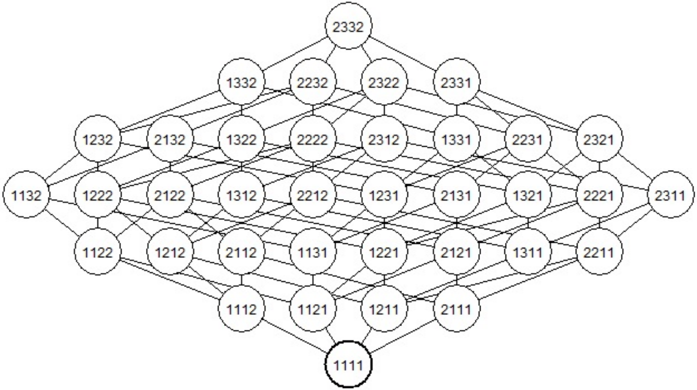
Supervision

Responsible for supervising other employees.

- 1 $\rightarrow$ No
- 2 $\rightarrow$ Yes

Hasse Diagram

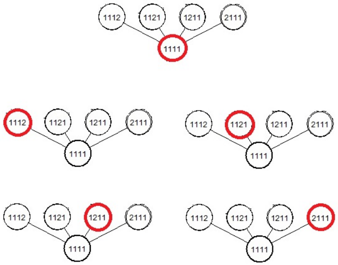
Hasse Diagram of the 36 possible profiles



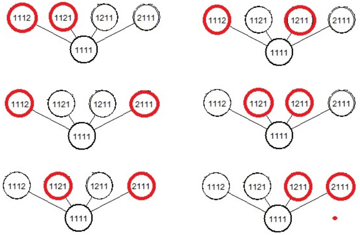
Choice of the Threshold

In order to investigate the robustness with respect to the choice of the threshold we test the following thresholds:

Singleton

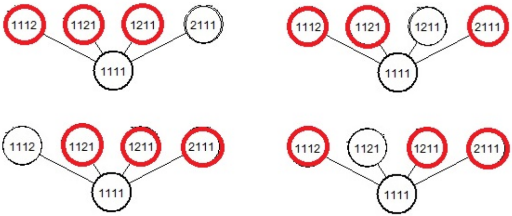


Pairs

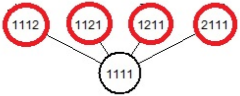


Choice of the Threshold

Triplet



Four profiles threshold



Robustness

	Italy	Netherlands	Estonia	Spain	Norway
1111	0,04505	0,01992	0,04499	0,02058	0,00229
1112	0,10373	0,05896	0,10670	0,07834	0,02051
1121	0,08777	0,04977	0,08651	0,06446	0,01467
1211	0,07484	0,04175	0,08444	0,05064	0,00989
2111	0,10531	0,06091	0,10745	0,08007	0,02136
C(1112,1121)	0,11822	0,06916	0,11883	0,09344	0,02508
C(1112,1211)	0,11267	0,06546	0,11798	0,08737	0,02282
C(1112,2111)	0,13002	0,07655	0,13155	0,10372	0,02957
C(1121,1211)	0,09878	0,05767	0,10151	0,07564	0,01749
C(1121,2111)	0,11970	0,07066	0,11944	0,09447	0,02557
C(1211,2111)	0,11395	0,06698	0,11847	0,08838	0,02342
C(1112,1121,1211)	0,12364	0,07292	0,12532	0,09835	0,02641
C(1112,1121,2111)	0,13778	0,08172	0,13704	0,11095	0,03205
C(1112,1211,2111)	0,13461	0,07965	0,13674	0,10824	0,03063
C(1121,1211,2111)	0,12502	0,07431	0,12622	0,09992	0,02699
C(1112,1121,1211,2111)	0,14065	0,08387	0,14093	0,11429	0,03280

Figure: Variation of gap index for the subgroup of males. The ranking is represented in scale of blues from dark blue (biggest gap index) to light blue (smallest gap index)

Robustness

	Italy	Netherlands	Estonia	Spain	Norway
1111	0,07336	0,01728	0,05670	0,03467	0,00244
1112	0,14780	0,07190	0,11738	0,09655	0,03537
1121	0,12943	0,05756	0,09867	0,07979	0,02588
1211	0,11798	0,04534	0,09581	0,06399	0,01933
2111	0,14890	0,07272	0,11777	0,09751	0,03602
C(1112,1121)	0,16544	0,08544	0,12984	0,11243	0,04325
C(1112,1211)	0,16039	0,07979	0,12840	0,10528	0,04034
C(1112,2111)	0,17693	0,09498	0,14118	0,12338	0,04944
C(1121,1211)	0,14549	0,06770	0,11341	0,09075	0,03194
C(1121,2111)	0,16614	0,08594	0,13015	0,11263	0,04342
C(1211,2111)	0,16143	0,08090	0,12863	0,10513	0,04068
C(1112,1121,1211)	0,17282	0,08987	0,13597	0,11734	0,04581
C(1112,1121,2111)	0,18581	0,10180	0,14677	0,13134	0,05358
C(1112,1211,2111)	0,18297	0,09944	0,14603	0,12730	0,05198
C(1121,1211,2111)	0,17315	0,09076	0,13644	0,11742	0,04622
C(1112,1121,1211,2111)	0,19003	0,10512	0,15078	0,13408	0,05502

Figure: Variation of gap index for the subgroup of females. The ranking is represented in scale of reds from dark red (biggest gap index) to light red (smallest gap index)

Gender gap

Gap difference: $G_T = G_F - G_M$
Threshold: c("1121", "1112")

Country	Male	Female	Difference	Rank
Italy	0,11822	0,16544	0,04722	1
Spain	0,09344	0,11243	0,01900	2
Estonia	0,11883	0,12984	0,01101	5
Norway	0,02508	0,04325	0,01817	3
Netherlands	0,06916	0,08544	0,01627	4

Truly multidimensional

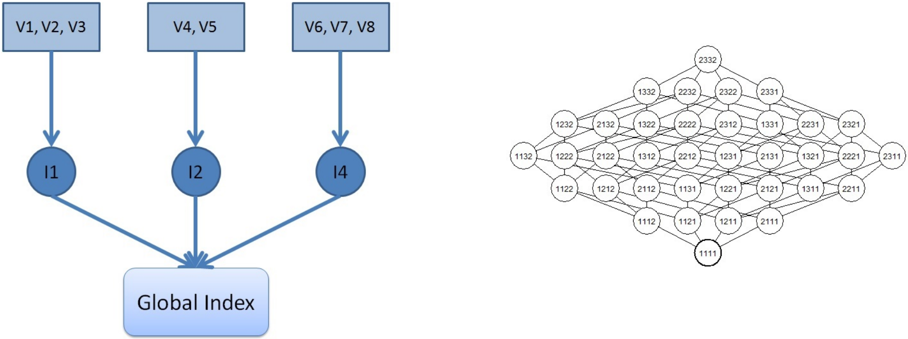


Figure: Multiple index VS poset

Conclusions

Advantages

1. Respect of the ordinal nature of the data
2. Evidence of robustness with respect to the chosen threshold
3. Truly multidimensional

Disadvantages

1. Computationally intensive
2. Individual data are needed