Multidimensional welfare rankings under weight imprecision

Stergios Athanasoglou

European Commission, JRC-Ispra
Institute for the Protection and the Security of the Citizen
Econometrics and Applied Statistics Unit

[The views expressed are purely those of the author and may not be regarded as stating an official position of the European Commission.]

March 18, 2014

Multidimensional welfare

- Many aspects of social well-being are intrinsically multidimensional.
- E.g., development, poverty, inequality cannot be fully captured by simple, exclusively income-based, measures.
- Originating in the powerful conceptual writings of Amartya Sen, the idea of multidimensional well-being has had a deep influence on academia as well as policy.
- Indeed, the primary tools that the UN uses to measure development and poverty, the Human Development and Multidimensional Poverty indices (HDI, MPI) reflect the above concerns.

Multidimensional welfare measurement

- Need to compare and eventually order possible alternatives (countries, policies, etc) on the basis of multidimensional information.
- Welfare indices (such as the HDI and MPI) approach this task by integrating the various dimensions of well-being into a scalar measure. This is generally achieved by assigning weights to the different dimensions and, in some fashion, aggregating over them.
- Often these choices are not grounded in economic theory or a coherent normative framework, sparking backlash (Ravallion, 2012).
- For instance, there is disagreement as to whether multidimensional poverty should be communicated through a “dashboard” of indices (Ravallion, 2012), or an aggregate scalar measure such as the MPI (Alkire and Foster, 2011).

The issue of weights

- Assume that a functional form for the aggregation function is in place (justified by normative desiderata), but weights are undetermined.
- Their choice can be fraught with complex philosophical and practical dilemmas, despite a multitude of proposed techniques (Foster and Sen, 1997; Decancq and Lugo, 2013).
- Indeed, there is frequently no single "right" weighting scheme and we are justified, if not compelled to, consider the effect of many different weights at once.
- Such an analysis would serve two goals:
  (a) to examine how robust a given ranking of alternatives is to changes in weights, and
  (b) to determine a compromise ranking that is in some sense "optimal" in the presence of weight imprecision.

Previous work

- Monte Carlo simulation in the context of broader uncertainty/sensitivity analyses (Saisana et al. 2005).
- Duclos et al. (2006, 2011) studied multidimensional poverty/inequality comparisons using ideas from stochastic dominance. They established an analytic criterion for determining whether a (pairwise) poverty comparison is robust within a large class of indices.
- Anderson et al. (2011) imposed monotonicity and quasiconcavity on the aggregation function and derived bounds on welfare levels.
- Foster et al. (2013) studied linear indices and parameterized weight imprecision with the ϵ-contamination model of Bayesian statistics. Focused on pairwise relations.
- Pinar et al. (2013) examined the HDI index and used ideas from stochastic dominance to determine the set of weights that results in best-case human development over time.

This paper's contribution and added value

- I propose a theoretical framework that yields consensus rankings in the presence of weight imprecision, which is formally rooted in the social choice/voting literature.
- The approach goes beyond existing work in the following ways:
  - (i) It produces a set of complete consensus rankings of the alternatives, not welfare bounds or pairwise dominance relations.
  - (ii) It can be justified on axiomatic grounds (thus guarding against charges of being ad-hoc).
  - (iii) It can be efficiently implemented in high-dimensional settings of multiple alternatives and welfare criteria (unlike techniques based on stochastic dominance).

The paper in a nutshell

- Consider a vector of weights as a voter and a continuum of weights as an electorate.
- With this voting construct in mind, Kemeny's rule from social choice theory is introduced as a means of aggregating the preferences of many plausible choices of weights.
- The axiomatic characterization of Kemeny's rule due to Young and Levenglick (1978) and Young (1988) is shown to extend to the present context.
- An efficient graph-theoretic algorithm is developed to compute or approximate the set of Kemeny optimal rankings.
- Further analytic results are derived for a relevant special case of the model.
- The model is applied to the ARWU index of Shanghai University, a popular and controversial index ranking academic institutions across the world. High problem dimensionality means it is a good "proof of concept".

Model description

- Set of alternatives $\mathcal{A}$ indexed by $a = 1, 2, ..., A$ and set of indicators $\mathcal{I}$ indexed by $i = 1, 2, ..., I$.
- Let $x_{ai} \in [0, 1]$ denote alternative a's normalized value of indicator i, $\mathbf{x_a} \in \Re^I$ its "achievement" (column) vector, and $\mathbf{X}_{\mathcal{A}} \subset [0, 1]^{I \times A}$ the resulting achievement matrix.
- Performance across indicators is weighted by a vector $\mathbf{w}$ belonging in the simplex $\Delta^{I-1} = \{w \in \Re^I : \ \mathbf{w} \geq \mathbf{0}, \ \sum_{i=1}^{I} w_i = 1\}$.
- Welfare corresponding to achievement vector $\mathbf{x}$ and $\mathbf{w}$ is given by a real-valued function $u(\mathbf{x}, \mathbf{w})$.
- The welfare function is purposely left general in order to accommodate many different multidimensional concepts.

Weight imprecision

- Now, define an importance function f on the simplex Δ^{I-1}, satisfying $f(\mathbf{w}) \geq 0$ for all $\mathbf{w} \in \Delta^{I-1}$ and $0 < \int_{\Delta^{I-1}} f(\mathbf{w})\mathrm{d}\mathbf{w} < +\infty$.
- f models imprecise beliefs regarding the "correct" set of weights to use.
- It may be set *a priori* by the decision-maker, or it may be arrived at by aggregating the views of agents to be ranked.
- In the case of the HDI, f could be set in the following manner: ask each country c to provide its importance function f_c on Δ^2 and then set $f = \sum_c f_c$.
- Work with continuous f, but model can be straightforwardly extended to account for discrete importance functions on a finite (or countably infinite) subset of weights belonging in Δ^{I-1}.

Weights as voters

- Define a profile L to be a triplet $L = (\mathbf{X}_{\mathcal{A}}, f, u)$, and let $\mathcal{L}$ denote the space of all profiles.
- Given a profile L, suppose we think of weight vector $\mathbf{w}$ as an imaginary voter who (weakly) prefers a_i over a_j if and only if $u(\mathbf{x}_{\mathbf{a_i}}, \mathbf{w}) > u(\mathbf{x}_{\mathbf{a_j}}, \mathbf{w})$ $(u(\mathbf{x}_{\mathbf{a_i}}, \mathbf{w}) \geq u(\mathbf{x}_{\mathbf{a_j}}, \mathbf{w}))$.
- Thus, voter $\mathbf{w}$'s preferences will be expressed as a (possibly partial) ranking of the alternatives.
- Construct an electorate of voters by considering each $\mathbf{w} \in \Delta^{I-1}$ and introducing $f(\mathbf{w})$ copies of itself. Thus, the greater $f(\mathbf{w})$ is, the more voters holding $\mathbf{w}$'s preferences are introduced. This results in a continuum of voters $\mathcal{E}(f)$ of finite measure.

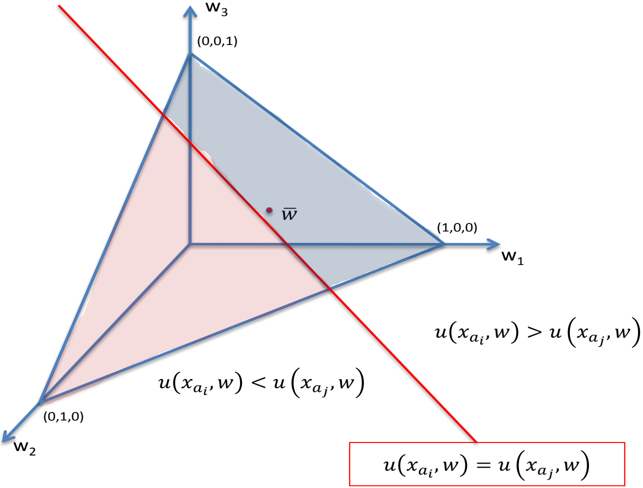
Connections with social choice

- What would constitute a "good" way of aggregating the preferences of all weight vectors, suitably weighted by the importance a decision maker places on them?
- More abstractly: Given a set of individual ranked preferences, what voting rule should society use to determine a consensus ranking? What properties should a compromise solution aspire to satisfy? What tradeoffs need to be reconciled?
- Fundamental questions, whose modern roots lie in the work of Condorcet and Borda.
- Arrow's impossibility theorem is a classical result along this vein.

Election matrices

- Given a profile $L = (\mathbf{X}_{\mathcal{A}}, f, u)$, define the election (proportion) matrix $\mathbf{Y}^L$ ($\mathbf{V}^L$).
- Y_{ij}^L (V_{ij}^L) defines the net majority (proportion) of voters within $\mathcal{E}(f)$ preferring a_i to a_j. Matrix $\mathbf{Y}^L$ ($\mathbf{V}^L$) summarizes this information for all pairs of alternatives.
- Generally, $\mathbf{Y}^L$ and $\mathbf{V}^L$ need to be computed numerically.
- However, analytic solutions are possible for some compelling special cases (see Section 5 in paper).

An example: $f(\mathbf{w}) \equiv 1$ and u linear



- $\mathcal{E}(f)$ equals the entire simplex with uniform importance.
- Y_{ij}^L is the difference between the volumes of the BLUE and RED regions.
- V_{ij}^L is the ratio of the volumes of the BLUE region and the entire simplex.

Kemeny's rule

- If R_1 and R_2 are rankings, their pairwise disagreement (or Kendall-τ distance) is given by the number of pairs (a_i, a_j) such that $R_1(a_i) > R_1(a_j)$ and $R_2(a_i) < R_2(a_j)$.
- Given a set of voters who each submit ordered preferences on a set of alternatives, Kemeny's rule (Kemeny, 1959) produces a ranking that minimizes the sum of its pairwise disagreements with respect to voter preferences.
- Applying this concept to infinite electorate $\mathcal{E}(f)$, the Kemeny-optimal set of rankings K^L can be simplified to ($\mathcal{R}_{\mathcal{A}}$ denotes the set of rankings of alternatives in $\mathcal{A}$)

$$K(L) \equiv K^L = \underset{R \in \mathcal{R}_{\mathcal{A}}}{\arg\min} \sum_{(a_i, a_j) \in \mathcal{A} \times \mathcal{A}} \mathbf{1}\{R(a_i) < R(a_j)\} Y_{ji}^L.$$

Normative analysis

- A rule is a function from the set of profiles to the set of nonempty subsets of rankings.
- Can we justify axiomatically the adoption of rule K as a means of ranking alternatives? In what sense would it be "better" than other methods we could employ?
- Yes, b/c it turns out that K is the only rule satisfying a set of desirable axioms.

Anonymity, Neutrality, Unanimity, Condorcet

[For rigorous definitions of the following Axioms please see the paper.]

Axiom 1. *A rule ϕ is anonymous if it depends only on the number of voters submitting ranking R as their preference, for all rankings R.*

Axiom 2. *A rule ϕ is neutral if the identity of an alternative does not affect the rank it receives.*

Axiom 3. *A rule ϕ is unanimous if, when all weights submit the same ranking of the alternatives, then the rule picks this ranking.*

Axiom 4. *A rule ϕ is extended-Condorcet if it respects the majority wishes of the electorate, whenever these do not involve intransitivities (i.e., situations where a majority of voters prefer A to B, B to C and C to A).*

Reinforcement

Axiom 5. *A rule ϕ satisfies reinforcement if it acknowledges and reinforces pre-existing consensus, thus imposing a degree of consistency to the aggregation process.*

- Consider the HDI, and suppose Africa and Europe have completely differing opinions regarding the weights of the three dimensions of the HDI.
  - African countries only want to consider weights $\mathbf{w}$ s.t. $w_H > w_I > w_E$.
  - European countries only want to consider weights $\mathbf{w}$ s.t. $w_E > w_H > w_I$.
- Suppose the UN chooses a method of ranking countries that, when considering the opinions of A and E separately leads to the same consensus ranking. In that case, reinforcement requires that the UN's method, when considering the preferences of A and E jointly, not disturb their pre-existing consensus.

Local independence of irrelevant alternatives

Axiom 6. *A rule ϕ satisfies local independence of irrelevant alternatives (LIIA) if the relative order of alternatives that are ranked "together" in a consensus ranking does not change, when we apply the rule to the restricted problem that focuses just on these alternatives and ignores all others.*

- Usually such contiguous intervals correspond to meaningful categories of alternatives.
- Suppose we rank the 100 best universities in the world. We would prefer the relative ordering of the top 20 (representing, say, Tier 1 institutions), to remain unchanged if we re-apply the rule ignoring those universities ranked 91-100, 51-100, or even the entire 21-100 for that matter.

The axiomatic characterization

Theorem 1

(i) On the domain of profiles $\mathcal{L}$, K satisfies anonymity, neutrality, reinforcement, extended-Condorcet, unanimity, and LIIA.

(ii) Let $\mathcal{Y}^Q$ denote the set of rational skew-symmetric matrices whose rows and columns are indexed by the elements of $\mathcal{A}$. On the restricted domain $\mathcal{L}^Q = \{L \in \mathcal{L} : \ \mathbf{Y}^L \in \mathcal{Y}^Q\}$, K uniquely satisfies anonymity, neutrality, reinforcement, unanimity, and LIIA.

- Largely a restatement of results by Young (1974, 1988), Young and Levenglick (1978).
- But care must be taken to ensure that their proofs extend to the current, non-standard setting.

(Important) computational issues

- Unfortunately, computing K is NP-hard (Bartholdi et al., 1989), even when the number of indicators is just four (Dwork et al., 2001).
- The main difficulty arises from Condorcet cycles, which imply intransitive majority pairwise preferences. Thus, it is important to identify and, in some fashion, resolve these cycles.
- Using classical results from discrete algorithms (Tarjan, 1972) and recent approximation algorithms (Van Zuylen and Williamson, 2009), I propose a graph theoretic algorithm that computes or provides a provably-good approximation of K (see Section 4 of the paper).
- If the size of Condorcet cycles is "small enough", then one gets an exact solution.

A special case of the model I: generalized weighted means

- A family of welfare functions that is particularly popular in many policy contexts are known as generalized weighted means (Decancq and Lugo, 2013).

- Parameterized by $\gamma \in \Re$, they are denoted by u^γ and satisfy

$$u^\gamma(\mathbf{x}, \mathbf{w}) = \begin{cases} \left(\sum_{i=1}^{I} w_i x_i^\gamma\right)^{\frac{1}{\gamma}} & \gamma \neq 0, \\ \prod_{i=1}^{I} x_i^{w_i} & \gamma = 0. \end{cases}$$

- $\frac{1}{1-\gamma} =$ elasticity of substitution between achievements.

- When $\gamma = 1(0)$ we recover the weighted arithmetic (geometric) mean. As $\gamma \to +\infty(-\infty)$, $u^\gamma(\mathbf{x}, \mathbf{w})$ converges to the maximum (minimum) coordinate of $\mathbf{x}$.

A special case of the model II: ϵ-contamination

- We are given an initial vector of weights $\bar{\mathbf{w}}$.
- Suppose that we are willing to grant equal consideration to weights deviating from $\bar{\mathbf{w}}$ that belong to the set W^{ϵ}, where

$$W^{\epsilon} = (1-\epsilon)\bar{\mathbf{w}} + \epsilon\Delta^{I-1} = \left\{ \mathbf{w} \in \Re^{I} : \;\; \mathbf{w} \geq (1-\epsilon)\bar{\mathbf{w}}, \;\; \sum_{i=1}^{I} w_i = 1 \right\}.$$

- Parameter $\epsilon \in [0,1]$ measures the imprecision associated with $\bar{\mathbf{w}}$. Can be modeled with an importance function f^{ϵ} assigning weight 1 to all $\mathbf{w} \in W^{\epsilon}$ and 0 everywhere else.
- Originally developed in Bayesian analysis (Berger and Berliner, 1986), this way of parameterizing imprecision is referred to as ϵ-contamination. Studied also in micro theory (Nishimura and Ozaki, 2006; Kopylov, 2009).
- First introduced by Foster et al. (2013) in the context of composite indices of welfare.

How could ϵ be set?

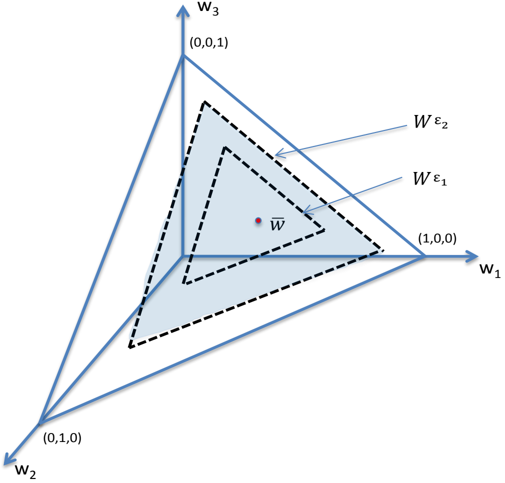
- Statistically, the parameter ϵ may be interpreted as the amount of error attached to the prior $\bar{\mathbf{w}}$.
- In our context, the choice of ϵ is largely subjective and should be decided in close consultation with the policy makers.
- Nevertheless, the simple structure of ϵ-contamination may inform this process by shedding light on the implications of different choices.
  (i) Places a uniform bound on allowable percentage decrease of an indicator's weight with respect to $\bar{\mathbf{w}}$, i.e.

$$\left\{ \frac{w_i}{\bar{w}_i} \geq 1 - \epsilon, \ \ \forall i \in \mathcal{I} \right\} \Leftrightarrow \left\{ w_i \in \left[\bar{w}_i - \epsilon \bar{w}_i, \bar{w}_i + \epsilon(1 - \bar{w}_i) \right], \ \ \forall i \in \mathcal{I} \right\}.$$

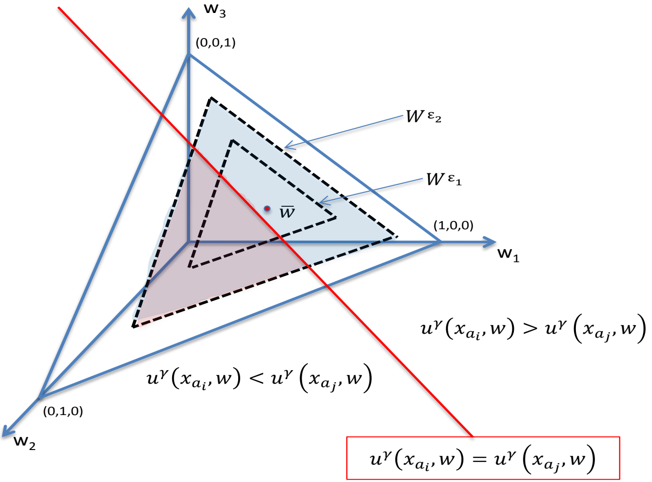
  (ii) Serves as a guide for policy makers who wish to "cover" a target percentage of all possible vectors of weights.

$$\frac{Vol\left(W^{\epsilon}\right)}{Vol\left(\Delta^{I-1}\right)} = \epsilon^{I-1}.$$

A graphical illustration of ϵ-contamination



Pairwise comparisons when $u = u^\gamma$ and $f = f^\epsilon$



- Given $\bar{\mathbf{w}}$, $\epsilon > 0$ and $\gamma \in \Re$, let $V_{ij}^{\epsilon,\gamma}$ denote the proportion of weights favoring a_i over a_j.
- $V_{ij}^{\epsilon_1,\gamma}$ ($V_{ji}^{\epsilon_1,\gamma}$) is the ratio of the volume of the smaller BLUE (RED) region to the volume of the inner triangle. Analogously for ϵ_2.

Analytic insights

Theorem 2

When a_i and a_j do not yield identical welfare under $\bar{\mathbf{w}}$, $V_{ij}^{\epsilon,\gamma}$ varies monotonically in the imprecision ϵ attached to $\bar{\mathbf{w}}$. It is decreasing if a_i initially dominates a_j and increasing if it is dominated by it. Conversely, when $\bar{\mathbf{w}}$ yields welfare for a_i and a_j, then $V_{ij}^{\epsilon,\gamma}$ remains constant as we vary ϵ.

Theorem 3

Simple geometric structure allows us to exploit the results of Lawrence (1991) and provide an explicit formula for $V_{ij}^{\epsilon,\gamma}$.

Proof of concept: the ARWU index

- Shanghai University's Academic Ranking of World Universities (ARWU), a popular composite index measuring research excellence in academic institutions.
- 6 criteria: (1) No. alumni winning Nobel prizes/Fields medals, (2) No. faculty winning Nobel prizes/Fields medals; (3) highly-cited researchers; (4) papers in Nature/Science; (5) papers indexed in leading citation indices; (6) per capita academic performance.
- ARWU score $u(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{6} w_i x_i$, and $\mathbf{w}_{ARWU} = (.1, .2, .2, .2, .2, .1)$.
- Despite its increasing influence and popularity, the ARWU index has been criticized on many grounds, including its non-robustness to changes in weights (Saisana et al., 2011).
- The controversy surrounding this index, in combination with its high dimensionality (100 universities, 6 criteria) make it a good application area for the model.

Applying the model

- Focus on the top-100 universities reported in the 2013 ARWU rankings, denoted by $\mathcal{A}_{100}$.
- I consider imprecision over the ARWU index weights via $\epsilon-$contamination with $\bar{\mathbf{w}} = \mathbf{w}_{ARWU}$ and $\epsilon \in \{1/6, 1/3, 1/2\}$.
- For convenience, denote by K^{ϵ} the Kemeny-optimal ranking of universities in $\mathcal{A}_{100}$ when applying the method for different values of ϵ.
- Differences $K^0 - K^{\epsilon}$ grow as we increase ϵ, and are much more pronounced for universities in the 51-100 range.
- There are moreover a handful of really substantial swings in rankings. For instance, the ENS-Paris was ranked 71st in the official 2013 ARWU ranking, whereas its Kemeny-optimal ranks for $\epsilon = 1/6, 1/3, 1/2$ are 62, 54, and 49, respectively.

Numerical application: ARWU index

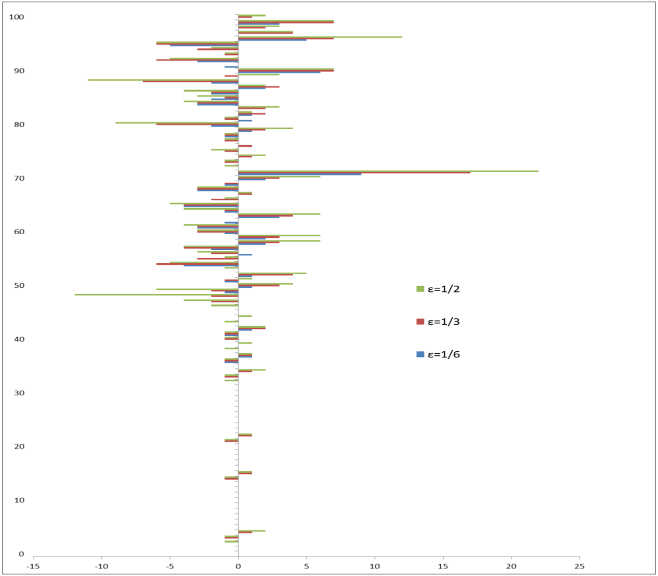


Figure : 2013 ARWU Top-100: $K^0 - K^\epsilon$.

Recap

- Judgments based on composite indices of welfare depend, sometimes critically, on how different dimensions of performance are weighted.
- As there is frequently no single "right" way to assign such weights, it is important to take this imprecision into account in a systematic and transparent manner.
- In this paper I have drawn from the theory of social choice to present a procedure for determining a ranking of the relevant alternatives that is normatively compelling and statistically interpretable.
- Developed graph-theoretic algorithm to implement rule and the applicability of the proposed framework was illustrated through a numerical example based on Shanghai University's ARWU index.
- Broader connections with decision-theoretic models of Knightian uncertainty can be explored.