ECINEQ
Society for the Study of Economic Inequality

Working Paper Series

Robust ordinal inequality comparisons with Kolm-independent measures

Gaston Yalonetzky

ECINEQ WP 2016 - 401

ECINEQ 2016 - 401
May 2016

www.ecineq.org

Robust ordinal inequality comparisons with Kolm-independent measures

Gaston Yalonetzky†
University of Leeds, U.K.

Abstract

Lv, Wang, and Xu (2015) recently characterized a new class of ordinal inequality measures axiomatically. In addition to their appealing functional forms, these measures are the only ones in the literature satisfying a property of independence, inspired by Kolm (1976). As acknowledged by the authors, the robustness of ordinal inequality comparisons to the several alternative suitable measures within the class is a natural concern. This note derives the stochastic dominance condition whose fulfilment guarantees that all inequality measures within the class rank a pair of distributions consistently.

Keywords: Stochastic dominance, Ordinal inequality.

JEL Classification: D30, I14.

†**Contact details:** University of Leeds; Maurice Keyworth Building LS2 9JT, UK. Phone: (+44) 113 343 0199. E-mail: G.Yalonetzky@leeds.ac.uk.

1 Introduction

The burgeoning literature on ordinal inequality measurement has provided several classes of measures in the last few years. Prominent examples include Allison and Foster (2004), Apouey (2007), Abul Naga and Yalcin (2008), Erreygers (2009), Reardon (2009), Lazar and Silber (2013), and more recently Lv et al. (2015). The contribution of Lv et al. (2015) is particularly interesting because their proposed class of measures cardinalises the distances between ordinal categories, and yet the indices behave well, fulfilling key properties like aversion to median-preserving spreads. Moreover, the functional forms of the measures by Lv et al. (2015) also bear other appealing traits, including ease of computation (e.g. not requiring to use the median). Last but not least, this class of indices is the only one fulfilling a property which we call Kolm-independence, whereby the change in total inequality due to the change in the relative frequency of an ordinal category is independent of the initial level of that frequency. Essentially, if we wanted to impose this property when measuring ordinal inequality, then the axiomatic characterization provided by Lv et al. (2015) implies that we should only use measures from their class.

However, even if we restricted ourselves to this class of Kolm-independent measures, we could still choose among several equally suitable measures. Lv et al. (2015) provide examples of such measures, including one which is basically a Gini index based on the modulus of the differences between the ordinal categories of the variable, cardinalised with natural numbers. Hence we could be naturally concerned by the robustness of pairwise ordinal inequality comparisons to alternative choices of equally appropriate ordinal inequality measures. Referring to the different functional forms available to cardinalise the differences between pairs of ordinal categories as "weights", Lv et al. (2015, p. 467) echo this concern succinctly: "It may be noted that the choice of weights, wijs, in the construction of a health inequality index f in our context is not unique. From the above discussion, those weights reflect our value judgments about how to deal with health "inequalities" from any two further apart health statuses in the construction of an overall health inequality index f. The choice of a particular set of weights may cause some concerns for researchers and for policy makers when our intuition about such weights is blurry or when we have some conflicting intuitions about exactly what set of weights should be chosen and used."

Addressing this issue, this paper derives the first-order stochastic dominance condition whose fulfilment guarantees that all inequality measures within the Kolm-independent class rank a pair of distributions consistently. The condition requires comparing across populations, or samples, their cumulative distributions of products of probabilities, which measure the likelihoods of finding pairs of individuals featuring specific differences between their reported categories (e.g. of self-reported health, life-satisfaction responses, educational levels, etc.). Intuitively, societies with higher probabilities of finding pairs of people with narrower differences (between their category-values) and lower probabilities of finding pairs with wider differences, will tend to be robustly less unequal than others, according to the measures of the class axiomatically characterized by Lv et al. (2015).

The rest of the note proceeds as follows. Section 2 provides the notation and a description of the class of ordinal inequality measures proposed by Lv et al. (2015). Section 3 provides the dominance proposition, together with its respective proof. Then the paper ends with some concluding remarks.

2 Preliminaries

2.1 Notation

Let x be an ordinal variable with c ordered categories. Each category is assigned a natural number from 1 to c. The respective discrete probability distribution is given by the vector: $P := [p(1), p(2), ..., p(c)]$, where $p(i) \equiv \Pr[x = i]$. With subscripts we refer to the probabilities, and other statistics, of a specific population or sample. Hence, for example, $p_A(1)$ is the relative frequency of people reporting the lowest category in society A.

Later, in the next section, we will also need statistics which are specific sums of probability products (e.g. p(1)p(2)). In particular we define the following functions:

$$\pi(\delta) \equiv \sum_{i=1}^{c-\delta} p(i)p(i+\delta), \quad \delta = 0, 1, ..., c-1 \tag{1}$$

As it will become apparent below, δ measures the modulus of the difference between two values of the variable, e.g. i and j, where each category has been cardinalised using natural numbers in the range $[1, c]$. Henceforth we refer to these absolute values as "gaps".

Examples of 1 include: $\pi(0) = \sum_{i=1}^{c}[p(i)]^2$, $\pi(1) = p(1)p(2) + p(2)p(3) + ... + p(c-1)p(c)$, and $\pi(c-1) = p(1)p(c)$. Note, importantly, that: $\pi(0) + 2\sum_{\delta=1}^{c-1}\pi(\delta) = 1$. These probabilities give us the likelihood of finding two people in the population whose gaps between their reported ordinal categories is equal to δ, assuming that the likelihood of appearance of a person with a value of i is independent from the likelihood of appearance of a person with a value of j. Hence we can also define a cumulative version of 1, which will be very useful for the derivation of the dominance condition in the next section:

$$\Pi(\delta) \equiv \pi(0) + 2\sum_{i=1}^{\delta}\pi(i), \quad \delta = 0, 1, ..., c-1 \tag{2}$$

Clearly, the vector $\Pi := [\Pi(0), \Pi(1), ..., \Pi(c-1)]$ is a discrete cumulative probability distribution, with $\Pi(c-1) = 1$. Each element gives us the probability of finding pairs of people with gaps of δ or lower.

Finally we define, for instance, $\Delta\Pi(\delta) \equiv \Pi_A(\delta) - \Pi_B(\delta)$ in order to denote differences between two populations or samples. Thus, likewise, we apply Δ to other statistics.

2.2 The class of Kolm-independent ordinal inequality measures

Lv et al. (2015) axiomatically characterize the following class of ordinal inequality measures:

$$\mathcal{M} := \{O(P) | O(P) = \sum_{i=1}^{c}\sum_{j>i}^{c} g(|i-j|)p(i)p(j)\}, \tag{3}$$

where g is a function mapping from the gaps of cardinalised categories to the non-negative segment of the real line, and $g(1) < g(2) < ... < g(c-1)$. As shown by Lv et al. (2015, proposition 1), the class in 3 is the only one satisfying properties of normalization, aversion to median-preserving spreads, invariance to parallel shifts, additivity and independence. Yet different choices of $g(|i-j|)$ are possible, including $g(|i-j|) = |i-j|$ and $g(|i-j|) = 2\alpha^{c-1-|i-j|}$ with $0 < \alpha < 1$ (Lv et al., 2015, p. 469). Hence it is worth inquiring into the conditions under which the choice of g will not affect the inequality ranking of A versus B.

3 The stochastic dominance condition for the class of Kolm-independent ordinal inequality measures

The dominance condition is the following:

Proposition 1. $O(P_A) < O(P_B)\ \forall O(P) \in \mathcal{M}$ *if and only if* $\Delta\Pi(\delta) \geqslant 0\ \forall \delta = 0, 1, 2, ..., c-1 \wedge \exists \delta \in [0, 1, 2, ..., c-1] | \Delta\Pi(\delta) > 0$.

Proof. Let $Q \equiv 2O$. Then clearly $\Delta O < 0 \leftrightarrow \Delta Q < 0$. Hence we will prove a condition for $\Delta Q < 0$ because Q can be directly expressed in terms of the probabilities π. In fact, based on the definitions of π in 1, it is straightforward to show that we can define ΔQ as:

$$\Delta Q \equiv \sum_{\delta=1}^{c-1} g(\delta)\Delta\pi(\delta) \tag{4}$$

Applying summation by parts to 4 using Abel's formula, we get the following expression:

$$\Delta Q = -\sum_{\delta=1}^{c-2}[g(\delta+1) - g(\delta)]\Delta\Pi(\delta) + g(c-1)\Delta\Pi(c-1) - g(1)\Delta\Pi(0) \tag{5}$$

$$= -\sum_{\delta=1}^{c-2}[g(\delta+1) - g(\delta)]\Delta\Pi(\delta) - g(1)\Delta\Pi(0) \tag{6}$$

Note that we moved from line 5 to line 6 because $\Delta\Pi(c-1) = 0$. Now, we know that $g(i) > 0\ \forall i \geqslant 1$ and that $g(\delta+1) - g(\delta) > 0\ \forall \delta \geqslant 1$. Therefore, from immediate inspection of 6 we can conclude that $\Delta Q < 0$ (and hence $\Delta O < 0$) for all possible choices of g (given the specified constraints on its properties, i.e. $0 < g(1) < g(2) < ...g(c-1)$) if and only if $\Delta\Pi(\delta) \geqslant 0\ \forall \delta = 0, 1, 2, ..., c-1\ \wedge\ \exists \delta \in [0, 1, 2, ..., c-1] | \Delta\Pi(\delta) > 0$.

■

Basically, proposition 1 states that A is robustly less unequal than B, i.e. according to all members of the class in 3, if and only if the cumulative distribution of probability products, i.e. the cumulative distributions of category gaps, is never lower in A than in B, and at least once strictly higher. Intuitively, A has higher cumulative proportions of low gaps and lower cumulative proportions of high gaps, vis-a-vis B. In order to implement the condition we first need to compute the cumulative distributions following the instructions of the preliminaries' section.

By way of further illustration of the condition, it is worth noting the cumulative probability vectors, Π, corresponding to the benchmark situations of minimum and maximum ordinal inequality. In the case of minimum ordinal inequality the requirement is that $\exists i|p(i) = 1$, i.e. the whole population is in the same category. In that case, we will have $\pi(0) = 1$ and $\pi(\delta) = 0\ \forall\delta = 1, 2, .., c-1$. Hence $\Pi(0) = \Pi(1) = ... = \Pi(c-1) = 1$. Clearly, with such cumulative distribution, no other distribution (unless $\exists i|p(i) = 1$) can exhibit less inequality, since their cumulative distributions of gaps have to lie somewhere below. Likewise, all different distributions characterized by $\exists i|p(i) = 1$ are bound to be ranked as having the same level of inequality by all members of the class in 3. Meanwhile, in the case of maximum ordinal inequality the benchmark is characterized by $p(1) = p(c) = 0.5$, i.e. half of the population in the bottom category, and half in the top. In that case, we have $\pi(0) = 0.5$, $\pi(\delta) = 0\ \forall\delta = 1, 2, .., c-2$, and $\pi(c-1) = 0.5$. Then $\Pi(0) = \Pi(1) = ... = \Pi(c-2) = 0.5$ (and, of course, $\Pi(c-1) = 1$). Here it is also easy to show that no other distribution can exhibit a cumulative distribution of gaps below the one generated by the benchmark of maximum inequality.

4 Conclusion

Proposition 1 provides a partial answer to the concern put forward by Lv et al. (2015) regarding the robustness of inequality comparisons to alternative choices of inequality members from the same Kolm-independent class. When the condition based on the cumulative distributions of gaps holds, the comparison is robust to any choice of index within that class. Otherwise, the ranking between A and B will crucially depend on the particular choice of ordinal inequality index. This is a common problem in other areas of distributional analysis (e.g. inequality comparisons with continuous variables, poverty comparisons, etc.), and is the reason why the answer provided is partial: Stochastic dominance conditions can only provide a quasi-ordering across the set of all admissible distributions (discrete probability distributions of a given number of categories in this case). Moreover, even when the dominance condition holds, while we can assert the robustness of the comparison, we cannot conclude anything regarding the cardinal *intensity* of the inequality comparison. The latter will always depend on the particular choice of index, even when dominance holds.

Future research in this particular area could look into statistical inference techniques for this condition, which could be useful especially when comparing samples. Finally, it might be worth inquiring into the existence of dominance conditions which rely more directly on the cumulative discrete probability distribution, as opposed to the cumulative discrete distribution of probability products.

References

Abul Naga, R. and T. Yalcin (2008). Inequality measurement for ordered response health data. *Journal of Health Economics 27*, 1614–25.

Allison, A. and J. Foster (2004). Measuring health inequality using qualitative data. *Journal of Health Economics 23*, 505–524.

Apouey, B. (2007). Measuring health polarisation with self-assessed health data. *Health Economics 16*, 875–94.

Erreygers, G. (2009). Correcting the concentration index. *Journal of Health Economics 28*, 504–15.

Kolm, S. (1976). Unequal inequalities. *Journal of Economic Theory 12*, 416–42.

Lazar, A. and J. Silber (2013). On the cardinal measurement of health inequality when only ordinal information is available on individual health status. *Health Economics 22*(1), 106–113.

Lv, G., Y. Wang, and Y. Xu (2015). On a new class of measures for health inequality based on ordinal data. *Journal of Economic Inequality 13*, 465–77.

Reardon, S. (2009). *Occupational and residential segregation*, Volume 17 of Research on Economic Inequality, Chapter Measures of ordinal segregation, pp. 129–155. Emerald.