The Journal of Economic Inequality (2023) 21:219–231
https://doi.org/10.1007/s10888-022-09547-5

Evaluating ordinal inequalities between groups

Tugce Cuhadaroglu[1]

Received: 25 February 2020 / Accepted: 8 August 2022 / Published online: 19 October 2022

Abstract
We explore the inequality measurement of a discrete ordinal variable between social groups. We provide an axiomatic characterization for the Net Difference Index (Lieberson: Sociol. Methodol. **7**, 276–291 1976), that makes use of rank-domination to evaluate the discrepancy between the distributions of two social groups over ordered categories. Adapting well-known principles of cardinal inequality measurement to the between-group ordinal inequality setting, we show that the Net Difference Index mimics the Gini Index in terms of its relationship to the Lorenz curve, in our setting.

Keywords Net difference index · Between-group inequality · Ordinal inequality · Inequality measurement

JEL Classification D63 · I14 · I24 · J15 · J16

Inequality measurement of ordinal variables has received a major attention in the last two decades, as the importance of non-income variables in determining societal wellbeing has been widely acknowledged (Allison and Foster 2004; Naga and Yalcin 2008; Kobus 2015; Lazar and Silber 2013; Lv et al. 2015; Cowell and Flachaire 2017; Gravel et al. 2021; Sarkar and Santra 2020). We contribute to this literature by analysing inequality measurement with two critical aspects. First, we focus on between-group inequalities. Rather than evaluating the overall distribution of a variable in the society, we investigate how to quantify the discrepancies between distributions of a variable among social groups. Second, our variable of interest is a discrete ordinal variable, such as health status, educational attainment or occupational status. Hence, we explore the inequality measurement of a discrete ordinal variable between social groups.

✉ Tugce Cuhadaroglu
tc48@st-andrews.ac.uk

[1] School of Economics and Finance, University of St. Andrews, Castlecliffe, The Scores, St. Andrews KY16 9AR, Scotland, U.K.

Being crucial constructs for social conflict and unrest, between-group inequalities are considered to be important determinants of social and economic welfare (Langer 2005; Ostby 2008; Stewart 2010). However, unlike between-group income inequality measurement (Bourguignon 1979; Shorrocks 1980; Cowell 1980; Elbers et al. 2008), measurement of non-income between-group inequalities have not received a systematic treatment in the form of a progressively developing literature. Instead, in different strands of research, such as statistical sociology (Gastwirth 1975; Lieberson 1976; Blackburn et al. 2001), segregation measurement (Hutchens 2012; Reardon 2009; Del Río and Alonso-Villar 2012) or dissimilarity measurement (Andreoli and Zoli 2014), both for empirical and theoretical purposes, stand alone tools are developed for the assessment of the uneven distribution of non-income variables between groups, such as educational attainment, health, occupational status or subjective well-being. Shooting at this gap, our aim is to develop a justified framework to evaluate ordinal inequalities between groups that is based on axiomatic analysis. Our methodology is to adopt principles from well-established tools of cardinal inequality measurement such as Lorenz ordering and its link to the Gini Index and adapt those to the particulars of our setting. Hence we demonstrate a similar approach to Le Breton et al. (2012) of discrimination measurement and to Hutchens (1991) of segregation measurement literatures.

A quick fix to the inequality measurement of ordinal variables has been to transform these ordinal variables to cardinal ones by using specific cardinalizations in order to enable the use of measures of income inequality.[1] However as first shown by Allison and Foster (2004) application of cardinal measures over these ordinal variables might result in incomparable levels of inequalities for different societies since these techniques are sensitive to scale changes, i.e.; once the scale changes, measured inequality changes.[2]

There exists a need for going beyond measures of income inequality and developing justified measurement methodologies for the evaluation of these non-income inequalities between social groups. That is what we aim to do. We focus on two social groups and first suggest simple tools enabling us to compare societies unambiguously in terms of the ordinal between-group inequality they possess. To this purpose, the *Dominance curve* makes use of stochastic dominance to compare societies and can be seen as analogous to the Lorenz curve of the income inequality measurement. Naturally, it does not provide a complete ranking of societies. In order to extend this partial ranking, we mimic the relationship of the Gini index to the Lorenz curve, and integrate the Dominance curve, leading us to the Net Difference Index (Lieberson 1976). The main contribution of this paper is the novel characterization of the Net Difference Index.

Net Difference Index makes use of rank-domination to evaluate the discrepancy between the distributions of two social groups over ordered categories. Specifically, it is equal to the difference between the probabilities that a randomly chosen member of a social group occupies a better position than a randomly chosen member of the counter-group.[3] Gastwirth (1975) suggests the use of this probability as a measure of earning

[1] For instance, in the measurement of inequality in educational attainment, although the data are collected over attainment categories, a common practice has been to assign the average number of years of schooling to corresponding categories (Barro and Lee 1993; 1996; 2001; Thomas et al. 2525).

[2] One strategy that has been developed by the literature is to come up with specific cardinalizations that are immune to scale changes so that measured inequality becomes invariant to scale. See Naga and Yalcin (2008) and Kobus and Milos (2012) and (Cowell and Flachaire 2017) for more on this approach.

[3] Net Difference Index is based on Mann-Whitney's U Statistics (Mann and Whitney 1947), which gives a non-parametric rank test that is used to determine if two samples are from the same population. The Statistics U is simply the number of times the observations from one sample precede the observations from the other sample when all of the observations are ordered into a single ranked series. The probability distribution tables of U are provided for testing the null hypothesis that two samples share the same distribution. The Statistics U is different from well-known Wilcoxon rank-sum statistics (Wilcoxon 1945) only in that U allows for different sample sizes.

differentials between genders, yielding Gastwirth's Discrimination Index.[4] Essentially, Net Difference Index evaluates the ordinal inequality between two groups as the difference between their respective (discretized) Gastwirth indices.

The characterization of the Net Difference Index is provided by three properties: Directionality that is responsible for symmetric comparison between the groups around 0; Successive Proportional Merges that account for invariance to the merges of adjacent positions with the same group ratios; and finally, Decomposability that allows for overall inequality to be expressed as a weighted average of the inequalities in subparts of the society. We discuss the significance of each of these properties for the behavior of the Net Difference Index in Section 2.2 and propose related indices that satisfy all but one of the stated properties.

To the best of our knowledge, this is the first paper to fully characterize an index of between-group inequality designed for ordered categorical variables. The closest work from the literature in terms of methodology and purposes (axiomatic characterization of a method to evaluate the discrepancy of group distributions over ordered categories) can be found in Andreoli and Zoli (2014). As a part of a larger research agenda that links segregation, ordinal inequality and discrimination, they propose an ordering of societies according to the discrepancy of the group distributions over ordered categories. They call this notion 'dissimilarity preserving ordinal information' and the main difference with the between-group ordinal inequality measurement principles we have in this paper comes from an axiom, Interchange of Groups, that allows to swap group distributions for certain sets of adjacent positions. This basically implies separability of the evaluation across positions, a property that is not satisfied by the Net Difference Index, simply because at each position, not only the distributions at that position matter, but the distribution of the lower ranked or higher ranked counter-group members is equivalently important. We believe this is a desirable property for an ordinal inequality measure. It is worth adding that the dissimilarity ordering also respects Successive Proportional Merges (named Independence from Split of Classes in that work).

A related strand of research that explores the uneven distribution of social groups across ordered categories comes from the literature on ordinal segregation. In a seminal paper, (Reardon 2009) conceptualizes ordinal segregation as 'the extent to which variation within social groups is less than total variation in the population', and suggests several indices that depend on the distances of the distributions of groups to a completely polarized distribution. This paper does not present any characterizations, but suggests a set of properties for this setting, that are not necessarily appropriate for our question of between-group inequality. This is because the main focus of segregation for that work is how the distribution within each social group compares to the distribution in the society, rather than how social groups compare to each other.

A final related line of work originates from the decomposability of ordinal inequality measures (Allison and Foster 2004; Naga and Yalcin 2008; Kobus and Milos 2012; Dutta and Foster 2013). Although these measures are developed to measure the overall inequality of an ordinal variable, they might possess decomposability properties that allow the overall inequality to be expressed as an aggregation of the inequalities within groups and between groups. Then a comparison of their between-group counterpart to our methodology would be relevant. (Kobus and Milos 2012) provide a characterization

[4] For a detailed analysis of how Gastwirth measure relates to stochastic dominance, see Le Breton et al. (2012).

of a decomposable family of indices that respect Allison and Foster partial ordering (Allison and Foster 2004). However their decomposability property does not allow for between-group comparisons; instead it aggregates inequality values within subgroups, weighted by subgroup sizes. Thus the indices that belong to this family, the Absolute Value Index of Naga and Yalcin (2008) and (Apouey 2007), do not possess between-group inequality counterpart. (Dutta and Foster 2013) decompose the overall inequality of happiness (as quantified by self reported subjective well-being data) in the US over groups of race, gender and region, by using the Allison-Foster index (AF)(Allison and Foster 2004). They, too, end up without any between-group inequality since the median category for all of the social groups happens to be the same (the data comes over 3 happiness categories, and hence it is not unreasonable that all social groups have their median reporting in the second category). When all groups have the same median, AF expresses overall inequality as a weighted sum of the inequalities *within* subgroups, just like the decomposability considered in Kobus and Milos (2012). Although measures based on AF ordering is extended to societies that do not have the same median (Sarkar and Santra 2020), decomposability of those with different subgroup medians have not been explored to the best of our knowledge. In this paper we solely focus on evaluating the inequality between groups, rather than decomposability of inequality into within-group and between-group counterparts. Whether Net Difference Index corresponds to the between-group counterpart of a general decomposable inequality index remains as an exciting question, that is outside the scope of this paper.

In the following section, we introduce the basic set up and the preliminaries of comparing societies with respect to the ordinal between-group inequality. Section 2 introduces the Net Difference Index. We provide a set of properties that characterize the Net Difference Index in this section. Section 2.2 discusses the independence of characterizing properties as well as related indices. Section 3 concludes with a discussion on possible extensions of the framework. The proofs are left to an appendix.

1 The setting

Consider an ordinal variable with finite number of categories. Let us call each category of this variable as a 'position'. Let n denote the number of positions. The ordering of the positions is exogenous and known. For positions 1,2,3,...,n, we adopt the convention that 1 is a better position than 2, which is a better position than 3 and so on. We denote a generic position by i or j so that $i < j$ implies i is a better position than j. A society $S \in C = \cup_{n\in\mathbb{Z}_{++}}[0,1]^{n\times 2}$ corresponds to the standard discrete probability distributions of two social groups, say women and men, over n ordered positions. Let w_i and m_i denote the proportion of women and men in position i, respectively, with $(w_1, w_2, ..., w_n) = \mathbf{w}^T$, $(m_1, m_2, ..., m_n) = \mathbf{m}^T$ (T stands for transpose) and certainly $\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} m_i = 1$. Then $S = (\mathbf{w},\mathbf{m})$ represents a society where the first column shows the discrete probability distribution of women over n positions and the second column shows that of men.

1.1 A partial ranking: dominance preorder

We first aim to present an unambiguous ranking criterion, just like Lorenz ordering of income inequality, for our setting. Given the ordinality of the variable of interest, first order

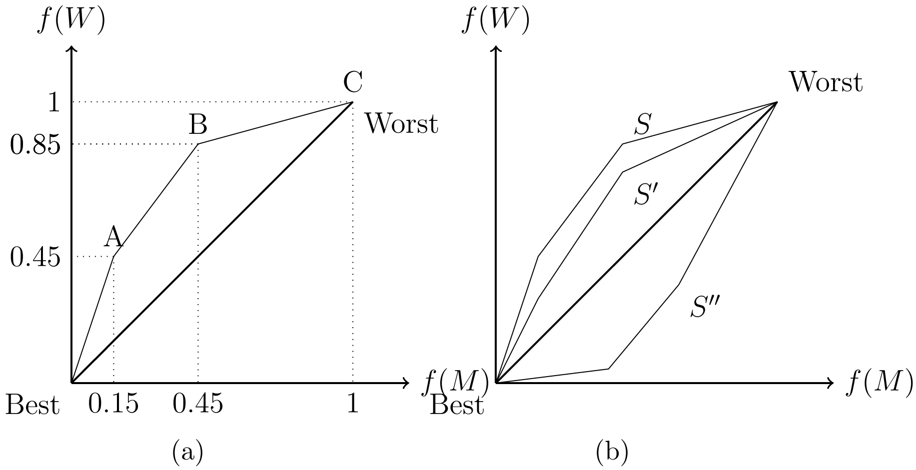

Fig. 1 Dominance Curve

stochastic dominance is a most natural reference, as it is scale independent. The distribution of Women first order stochastically dominates that of Men, $\mathbf{w} \succ^{SD} \mathbf{m}$, if for any position i, the proportion of women occupying positions at least at good as i are never less than that of men; i.e.; for any i, $\sum_{j=1}^{i} w_j \geq \sum_{j=1}^{i} m_j$ with at least one strict inequality. Stochastic dominance tells us whether a group is unambiguously in an advantageous position in a society. Since we aim to rank societies with respect to their between-group inequalities, we require a notion of the intensity of stochastic dominance. Let us introduce this concept with a graphical representation. Consider the following societies distributed over 3 positions as follows: $S = \begin{pmatrix} 0.45 & 0.15 \\ 0.40 & 0.30 \\ 0.15 & 0.55 \end{pmatrix}$ $S' = \begin{pmatrix} 0.30 & 0.15 \\ 0.45 & 0.30 \\ 0.25 & 0.55 \end{pmatrix}$ $S'' = \begin{pmatrix} 0.05 & 0.40 \\ 0.30 & 0.25 \\ 0.65 & 0.35 \end{pmatrix}$

Figure (1a) plots the cumulative frequency distribution, $f : [0,1] \rightarrow [0,1]$ of men against that of women for S. Similar to the logic of the Lorenz curve, the individuals are ordered in line with their positions from best to worst. Point A corresponds to the cumulative frequency of the individuals of the first position only, whereas point B marks the cumulative frequency of the first two positions. Finally, at point C all individuals are considered.

We call this curve **the Dominance curve** as it can be interpreted as a visualisation of the stochastic dominance between groups.[5] Formally, the Dominance curve for S is given by: $f^S : [0,1] \rightarrow [0,1]$,

$$f^S(m) = \sum_{i=1}^{k-1} w_i + w_k \frac{\Delta m}{m_k},$$

[5] The notion of using a mapping of cumulative distributions to assess the discrepancy or similarity of two populations is nowhere novel to this paper. The Dominance curve is conceptually equivalent to the Segregation curve (Duncan and Duncan 1955), the Concentration curve (Mahalanobis 1960) or the Discrimination curve (Le Breton et al. 2012). We present this notion merely as another foundation to the index we characterize.

where $k = \max\{1, .., i, .., n\}$ such that $m = \sum_{i=1}^{k-1} m_i + \Delta_m$ with $\Delta_m \geq 0$. Basically, we assume uniform distribution of the groups within positions. Certainly, $f^S(\sum_{i=1}^{k} m_i) = \sum_{i=1}^{k} w_i$ for any $k \in \{1,2,...,n\}$. That is, $f^S(m)$ gives out the cumulative frequency of women occupying positions that are at least as good as those m of men. Figure (1b) depicts the Dominance curves for S' and S'' as well as S. The 45% line is the **equality line**; the Dominance curve of a society lies exactly on the equality line if and only if the frequency distribution of women and men are identical. If $\mathbf{w} \succ^{SD} \mathbf{m}$, as it is in S, then the Dominance curve lies fully above the equality line. Conversely, since in S'', $\mathbf{m}'' \succ^{SD} \mathbf{w}''$, the Dominance curve lies fully below the equality line. Moreover, we argue that the distance to the equality line bears a sense of the intensity of stochastic dominance: the further away from the line a society is, the higher the intensity of stochastic dominance between groups. For instance, for S and S', for any given proportion of men, there is a higher proportion of women occupying positions that are at least as good as those men in both S and S' and this proportion is always higher in S than S'.

Let us formalize this notion of being closer to the equality line with an ordering relation. Define **the Dominance preorder**, $\succ^D \in (C \times C)$ such that $S \succ^D S'$ if and only if one of the following holds: (i) $\mathbf{w} \succ^{SD} \mathbf{m}$, $\mathbf{w}' \succ^{SD} \mathbf{m}'$ and $f^S(m) \geq f^{S'}(m)$ for all $m \in [0,1]$ with strict inequality for some $m \in (0,1)$. (ii) $\mathbf{m} \succ^{SD} \mathbf{w}$, $\mathbf{m}' \succ^{SD} \mathbf{w}'$ and $f^S(m) \leq f^{S'}(m)$ for all $m \in [0,1]$ with strict inequality for some $m \in (0,1)$. Hence for any two distinct societies S and S' that fully lie on the same side of the equality line, we have $S \succ^D S'$ iff S' lies in between S and the equality line. It is immediately seen that $\succ^D$ is a strict partial order; it is asymmetric, transitive but not necessarily complete. It captures a sense of intensity of the stochastic dominance: Given $S \succ^D S'$, we deduce that the same group stochastically dominates the other in both societies, say $\mathbf{w} \succ^{SD} \mathbf{m}$. But we also deduce that the stochastic dominance is stronger in S since for any $\%x$ of men occupying the top of the men distribution, there is always more women that are in an equal to or better position than those men in S than S'.

Dominance preorder suggests a reasonable way to compare societies in terms of the inequality between groups. However it can only be used to evaluate very specific societies; societies with stochastic dominance between groups. One way to extend this partial comparison to the domain of all societies is to come up with indices that agree on the ranking of the Dominance preorder, yet are defined for all possible societies. That is what we do in the next section.

2 Extending the partial ranking: the Net Difference index

Given the analogies so far between the Lorenz curve and the Domination curve, a natural extension of the Dominance preorder can be reached by mimicking the relationship between the Gini Inequality Index and the Lorenz curve. Gini Inequality Index is equal to the ratio of the area between the Lorenz curve and the equality line to the area under the equality line. One crucial difference between the Lorenz curve and the Domination curve is that, the latter can reach both above and below the equality line, inducing a sense of 'direction' to the inequality. Taking this into account, we compute 'the net area', the area between the curve and the equality line above the equality line minus the area between them below the equality line and we arrive at the Net Difference Index (Lieberson 1976).

Let us first formally define the index, before showing the relationship to the Domination Curve formally in Proposition 1:[6]

$$ND(S) = \sum_{i}\left(w_i \sum_{i+1}^{n} m_j - m_i \sum_{i+1}^{n} w_j\right)$$

Given a society S, the Net Difference Index, $ND(S)$, measures inequality in terms of the difference in frequencies that a group ranks higher than the other group in pairwise confrontations. For instance, women of position i occupy a better position than all the men that are in worse positions than i, hence they rank higher than $\sum_{i+1}^{n} m_j$ of men.

Intuitively, ND gives out the ex-ante probability advantage between groups: For a random pair of a woman and a man, the difference in probabilities of one individual being in a better position than the other. This intuition becomes more explicit if we were to express ND as follows:

$$\begin{aligned} ND(S) &= \sum_{i}\sum_{j:j>i}(w_i m_j) - \sum_{i}\sum_{j:j>i}(m_i w_j) \\ &= \sum_{i}\sum_{j} c_{ij} w_i m_j \text{ where } c_{ij} = \begin{cases} 1 & \text{if } i < j \\ 0 & \text{if } i = j \\ -1 & \text{if } i > j \end{cases} \end{aligned}$$

ND is a directional measure. It takes values between -1 and 1, 0 being complete equality, 1 being maximum inequality advantaging women and -1 being maximum inequality favoring men. ND respects the ordering suggested by the Dominance preorder, i.e., if $S >^D S'$ then $|ND(S)| > |ND(S')|$. Finally, Proposition 1 establishes the promised relation between the Dominance curve and ND:

Proposition 1 $ND(S) = \dfrac{\text{Net area between the Dominance curve and the equality line}}{\text{The area below the equality line}}$

The proof of Proposition 1 is merely based on the integration of f^S.

2.1 Characterizing properties

A between-group ordinal inequality measure is a continuous, non-constant function $H : C \to \mathbb{R}$ that attaches to each possible society S, a real number indicating the amount of inequality between the distributions of groups across ordered positions. In this subsection we list and discuss the properties on H characterizing the Net Difference Index.

ND is a measure that takes into account the direction of the inequality between groups. Directionality ensures that exchanging the distributions of women and men reverses the direction of the inequality. The argument for directionality is not too difficult to defend for two social group settings such as women vs men or white vs non-white origin; one would not only be interested in how the level of inequality changes over time and space but also whether inequality always favor the same social group or no.

Directionality (DR) For any $S = (\mathbf{w},\mathbf{m})$, we have $H(\mathbf{w},\mathbf{m}) = -H(\mathbf{m},\mathbf{w})$.

[6] We abuse notation and use $\sum_i$ to denote $\sum_{i=1}^{n}$, $\sum_{i+1}^{n}$ to denote $\sum_{j=i+1}^{n}$ and so on.

The following is a property that we borrow from segregation literature and modify according to the ordinal information in our setting. Consider two societies S and S', that are equal to each other in all aspects but there is only one position in S' corresponding to two successive positions with equal women to men ratios in S'. Hence S has n positions, whereas S' has $n-1$. Basically it is as if S' is obtained from S by combining two successive positions with the same group ratio. Successive Proportionate Merges ensures that the inequality between groups remain unchanged, i.e, combining two successive positions with the same women to men ratios does not change inequality.[7]

Successive Proportionate Merges (SPM) Let S be a society over n positions such that $\exists$ $k < n$ with $w_k/m_k = w_{k+1}/m_{k+1}$. Let S' be a society over $n' = (n-1)$ positions such that $w'_i = w_i$, $m'_i = m_i$ for $i = 1,...,k-1$; $w'_k = w_k + w_{k+1}$, $m'_k = m_k + m_{k+1}$, and $w'_i = w_{i+1}$, $m'_i = m_{i+1}$ for $i = k+1,...,n-1$. Then $H(S) = H(S')$.

SPM highlights when the ordinal information about the positions becomes idle. For two successive positions, the fact that one is better than the other is relevant for inequality only if the relative distributions of the social groups differ over these positions. Notice that combining two positions is not disregarding all ordinal information regarding them, it is only disregarding the ordinal information *between* them: the individuals occupying these positions are still in better (worse) positions than all the other individuals they were jointly dominating (dominated by) before.

Decomposability is a crucial property for characterizations of inequality indices in the entire literature not only because it mathematically helps to pin down the family of indices but also it has practical implications. Decomposability shows how to aggregate inequalities in different subparts of the society consistently. Quite often empirical researches are interested in the concentration of inequality in various parts of the society such as geographical locations or within different subgroups such as ethnic groups. Decomposable indices allow us to express the overall inequality in the society as an aggregation of the inequalities in different subparts of the society. This in turn helps to understand the extent to which the overall inequality is attributed to discrepancies between certain subgroups.

Remembering the graphical representation of the Net Difference and its similarity to the relationship between Gini and the Lorenz curve, it is not immediately clear what kind of a decomposability property the Net Difference might satisfy.[8] Since the focus of our interest is the inequality between groups, a natural decomposition would be over different subgroups of the social groups, where a subgroup refers to a subset of a social group. For instance consider a society S that consists of 80% local and 20% immigrant women. Let S' be the subsociety that consists of men of society S and only the local women, whereas S'' consists of the men and immigrant women:

$$S = \begin{pmatrix} 0.42 & 0.40 \\ 0.42 & 0.30 \\ 0.16 & 0.30 \end{pmatrix} S' = \begin{pmatrix} 0.50 & 0.40 \\ 0.45 & 0.30 \\ 0.05 & 0.30 \end{pmatrix} S'' = \begin{pmatrix} 0.10 & 0.40 \\ 0.30 & 0.30 \\ 0.60 & 0.30 \end{pmatrix}$$

Notice that $0.8\mathbf{w}' + 0.2\mathbf{w}'' = \mathbf{w}$. Then a decomposable index would allow to express the overall inequality in S as an aggregation of the inequalities in the subsocieties S' and S''

[7] Notice that this property assumes variability of the number of positions in the society and becomes useful in comparing different societies with different number of positions. It postulates that inequality is unchanged if one position is divided into a number of positions with identical group ratios, hence provides a hypothetical, benchmark society to be able to compare the societies with different number of positions.

[8] The Gini Index does not belong to the group of additively decomposable income inequality indices. For more on decomposability of Gini, see (Bourguignon 1979; Dagum 1998; Lambert and Aronson 1993).

.[9] The decomposability property satisfied by *ND* allows this aggregation to be a standard weighted average by population weights; i.e., $0.8ND(S') + 0.2ND(S'') = ND(S)$. In other words, the contribution of a subgroup to the overall between-group inequality is proportional to its population weight:

Decomposability (DEC): Let $\mathbf{w} = \alpha\mathbf{w}' + (1-\alpha)\mathbf{w}''$ for some $\alpha \in (0,1)$.

Then;

$$\alpha H(\mathbf{w}', \mathbf{m}) + (1-\alpha)H(\mathbf{w}'', \mathbf{m}) = H(\mathbf{w}, \mathbf{m}).$$

Similarly for $\mathbf{m} = \beta\mathbf{m}' + (1-\beta)\mathbf{m}''$ for some $\beta \in (0,1)$, we have;

$$\beta H(\mathbf{w}, \mathbf{m}') + (1-\beta)H(\mathbf{w}, \mathbf{m}'') = H(\mathbf{w}, \mathbf{m}).$$

Certainly this is a strong decomposability property that essentially induces the linearity of the functional form in $\mathbf{w}$ and $\mathbf{m}$. Whenever the distribution of a group can be expressed as a linear combination of the distribution of the same group in two subsocieties, keeping the distribution of the other group constant, then between-group inequality can be expressed as a linear combination of the inequalities in these subsocieties.

We are now ready to introduce the main result of the paper. These three properties, in addition to continuity and non-constant behavior by definition of H, not only are satisfied by *ND*, but also they characterize it up to a scalar transformation.

Theorem 1 $H : C \to \mathbb{R}$ satisfies DR, SPM and DEC if and only if it is a scalar transformation of the Net Difference Index.

The proof is mainly based on a sequence of decompositions that result in a weighted aggregation of inequalities in elementary subsocieties of perfect polarization or perfect equality. We discuss independence and the implications of the characterizing properties in the next subsection.

2.2 Independence and other related indices

All of the characterizing properties are independent. DR not only assigns a direction to the measured inequality but does this in a symmetric way around 0. An index that evaluates dominations by women and men asymmetrically can be an example to a function that satisfies all other properties but DR.

For instance:

$$H(\mathbf{w}, \mathbf{m}) = \sum_i \left(2w_i \sum_{i+1}^{n} m_j - m_i \sum_{i+1}^{n} w_j \right).$$

[9] In principle, the subgroups of the other group, say local men, $\mathbf{m}^*$, and immigrant men $\mathbf{m}^{**}$, can be of further interest. In that case, overall inequality could be expressed as an aggregation of inequalities in $(\mathbf{w}', \mathbf{m}^*)$, $(\mathbf{w}'', \mathbf{m}^{**})$ in addition to $(\mathbf{w}', \mathbf{m}^{**})$ and $(\mathbf{w}'', \mathbf{m}^*)$. Notice that this is indeed inline with the standard definition of decomposability '...such that the total inequality of a population can be broken down into a weighted average of the inequality existing within subgroups of the population and the inequality existing between them (Bourguignon 1979)'. For this example, within subgroup inequality refers to the inequalities within locals $(\mathbf{w}', \mathbf{m}^*)$ and within immigrants $(\mathbf{w}'', \mathbf{m}^{**})$, whereas the between subgroup inequality is captured by the inequalities in $(\mathbf{w}', \mathbf{m}^{**})$ and $(\mathbf{w}'', \mathbf{m}^*)$.

Directionality of a between-group inequality measure might be a useful property in settings with only two social groups and when the direction of inequality indeed matters for policy purposes. However it might not always be a desirable property for a practical, summary measure of inequality, especially for comparisons across societies with different social groups. Moreover once extension of the index to multi-group settings is considered, as we do in Section 3, directionality becomes burdensome. A very natural question becomes whether we could extend the Dominance preorder without directionality, and instead consider the absolute value of the difference in probability advantages as follows:

$$D(S) = \left| \sum_{i} \left(w_i \sum_{i+1}^{n} m_j - m_i \sum_{i+1}^{n} w_j \right) \right|.$$

D takes values between 0 and 1 (0 being complete equality and 1 being maximum inequality) and respects the Dominance preorder like ND. The characterization of D certainly follows similar principles to that of the Net Difference Index with two crucial differences. First, Directionality needs to be replaced with a Symmetry property, ensuring that swapping the distributions of women and men does not change the measured inequality. Second, and more critically, Decomposability needs to be modified. To see why, consider the decomposition of a completely equal S into S^1 and S^2 as follows:

$$S = \begin{pmatrix} 0.50 & 0.50 \\ 0.50 & 0.50 \end{pmatrix} S^1 = \begin{pmatrix} 1 & 0.50 \\ 0 & 0.50 \end{pmatrix} \text{ and } S^2 = \begin{pmatrix} 0 & 0.50 \\ 1 & 0.50 \end{pmatrix}$$

Although $ND(S) = 0 = 0.5ND(S^1) + 0.5ND(S^2)$, the same decomposition does not hold for $D(S)$ since $D(S^1) = D(S^2) = 1$. This is because for D it is not possible to express the inequality in S as a weighted average of the inequalities in constituent subsocieties *unless* between-group inequality is favoring the same social group in both subsocieties. In the decomposition of S, women are more advantageous in S^1, whereas are men in S^2. When two subgroups are actually considered together in S, these advantages cancel out. A directional measure, like ND, accommodates this cancelling out in the decomposition, however for a symmetric measure such as D decomposition is only possible if the between-group inequality is favoring the same social group, hence there is no cancelling out when the entire group is considered. Thus a weaker decomposability property that allows decomposition only for certain type of subsocieties, together with symmetry and SPM, characterizes D. Although the current paper does not include this characterization for purposes of length, it is available up on request.

As can be seen in the proof of Theorem 1, the critical properties for the characterization are SPM and DEC. SPM highlights the noncardinality of the variable of interest. For two successive positions, the fact that one is better than the other is relevant for inequality only if the relative distributions of the social groups differ over these positions according to SPM. However, if there *is* actually more information regarding the ranking of the positions rather than pure ordinal information, one might consider to use a weighted version of the index:

$$ND^W(S) = \sum_{i} c_i \left(w_i \sum_{i+1}^{n} m_j - m_i \sum_{i+1}^{n} w_j \right)$$

where $c_i : \{1, 2, ..., n\} \rightarrow R$ is a weighting function or simply a cardinal scale. This index would satisfy DR and DEC but fails to satisfy SPM. *ND* evaluates rank dominations equivalently regardless of the exact rank of the position. In other words, occupying a better position than a counter group member has the same weight regardless of the rank of this better position. This means the advantage of one group in the bottom of the distribution can be compensated by the advantage of the other group (or the disadvantage of the same group) in the top of the distribution. If instead we were to argue that dominations, say, in the top positions should be weighted higher than the dominations at the bottom (or vice versa), then the weighted version of the index would be useful.

Finally, an example to a function that satisfies DR and SPM but fails DEC would be:

$$H(S) = \sum_i \left(\frac{w_i}{m_i} + \frac{m_i}{w_i} \right)(w_i - m_i).$$

This function satisfies DR since it treats the two groups symmetrically around 0 thanks to the second component in the sum. It also satisfies SPM since for two consecutive positions with the same w_i/m_i ratios, the contribution to the overall sum remains the same when these positions are merged. However it does not satisfy DEC as can be seen from the fact that it is not linear in w_i and m_i.

3 Concluding remarks and possible extensions

Unequal distribution of social groups across different levels of welfare is quite commonly observed. This paper aimed to analyze an intuitive and well-founded methodology to evaluate non-income inequalities between two social groups without appealing to additional cardinalization assumptions. We conclude with two possible extensions of the Net Difference Index.

A natural way to extend the Net Difference Index to settings with more than two groups is to consider an aggregation of the differences in pairwise dominations for each pair of groups. When there are more than two social groups, we first compute the average difference in number of dominations for each pair of groups. The average of the absolute average differences would then be the multi-group index. Let us state this idea formally: Let $\mathcal{G}$ be a set of social groups with cardinality G. Then a society matrix S with G groups will be of dimension $n \times G$ and a multi-group ordinal inequality measure would be equal to $\frac{1}{2G}\sum_{M \in G}\sum_{N \in G}|ND(\mathbf{m}, \mathbf{m})|$, where $\mathbf{m}$ and n denote the probability distributions of groups M and N in S respectively. Notice this still captures the extra probability that on a random selection of a pair of individuals from different groups, the member of one group rank-dominates the other. Axiomatic characterization of this multi-group Domination Index requires further research.

Having focused our attention on ordinal inequalities, we assumed full comparability of the categories. A second extension can be suggested for only partially comparable categories. Consider the attributes of an occupation such as wage, prestige, working conditions, etc. An occupation may have quite challenging working conditions, even resulting in health troubles, although offering a very high level of wage. How this occupation would compare to one with better working conditions but lower pay is not obvious. Hence taking multiple attributes into account might result in only a partial ordering of occupations rather than a linear one. Similarly, consider a setting where two aspects of welfare are taken into account simultaneously in determining the positions, such as health and happiness. Both health and happiness data are examples to ordered categorical data, however taking both of them into account at the same time would result in a partial ordering of the positions (if we are to avoid extra assumptions such as having more health is better than having more happiness). Hence the question becomes how to compare distributions of groups over partially ordered categories. One suggestion we have is the Maximum Group Inequality index: Formally, let $P_{\mathcal{I}}$ be a strict partial order over a set of positions $\mathcal{I}$. A society will be a pair of elements $(S, P_{\mathcal{I}})$, where S is the usual society matrix. Let $\mathcal{L}^{P_{\mathcal{I}}}$ denote the set of linear extensions of P over $\mathcal{I}$, i.e.; the set of complete, transitive and asymmetric binary relations over $\mathcal{I}$ with for all $L_{\mathcal{I}}$ in $\mathcal{L}^{P_{\mathcal{I}}}$, iLj if iPj. Then, Maximum Group Inequality Index, M, will be: $M(S, P_{\mathcal{I}}) = \max_{L_{\mathcal{I}} \in \mathcal{L}^{P_{\mathcal{I}}}} |ND(S_{L_{\mathcal{I}}})|$, where $S_{L_{\mathcal{I}}}$ refers to the society matrix with the linear order $L_{\mathcal{I}}$. As before, M takes values in [0,1]. If there is no missing information about the ordering of the positions, M is equal to $|ND|$. In case of some missing information, M gives the maximum possible level of group inequality, which refers to the worst-case scenario of the society. If two positions remain uncompared by the original ordering, this will be because of the fact that there is no unique universal way of ranking these positions, which is consistent with a Rawlsian framework of welfare for a worst-case scenario. The algorithmic structure and behavior of the Maximum Group Inequality Index remain to be explored.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10888-022-09547-5.

Acknowledgements I am grateful to Miguel Angel Ballester, Ramses Abul Naga, Jorge Alcalde-Unzu, Salvador Barbera, Joan Maria Esteban, Marc Fleurbaey, Carmen Herrero, Jordi Masso, Debraj Ray, Pedro Rey-Biel and William Thompson for their helpful comments and suggestions. Financial support by the Spanish Ministry of Science and Innovation through grant ECO2008-04756 and FEDER (Grupo Consolidado-C) is gratefully acknowledged.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed in this study.

Declarations

Conflict of Interests The author has no relevant financial or non-financial interests to disclose.

References

Allison, R.A., Foster, J.E.: Measuring health inequalities using qualitative data. J. Health Econ. **23**, 505–524 (2004)

Andreoli, F., Zoli, C.: Measuring dissimilarity. Working paper series, University of Verona, Department of Economics, WP-23 (2014)

Apouey, B.: Measuring health polarization with self-assessed health data. Health Econ. **16**(9), 875–894 (2007)

Barro, R.J., Lee, J.-W.: International comparisons of educational attainment. J. Monet. Econ. **32**(3), 363–394 (1993)
Barro, R.J., Lee, J.-W.: International measures of schooling years and schooling quality. Amer. Econ. Rev. **86**(2), 218–23 (1996)
Barro, R.J., Lee, J.-W.: International data on educational attainment: updates and implications. Oxf. Econ. Pap. **53**(3), 541–63 (2001)
Blackburn, R., Brooks, B., Jarman, J.: The vertical dimension of occupational segregation. Work Employ. Soc. **15**(3), 511–538 (2001)
Bourguignon, F.: Decomposable income inequality measures. Econometrica **47**(4), 901–20 (1979)
Cowell, F.A.: On the structure of additive inequality measures. Rev. Econ. Studies **47**(3), 521–31 (1980)
Cowell, F.A., Flachaire, E.: Inequality with ordinal data. Economica **84**(334), 290–321 (2017)
Dagum, C.: A new approach to the decomposition of the Gini income inequality ratio. In: Income inequality, poverty, and economic welfare, pp. 47–63. Springer (1998)
Del Río, C., Alonso-Villar, O.: Chapter 2: occupational segregation measures: a role for status. In: Inequality, mobility and segregation: essays in honor of Jacques Silber, pp. 37–62. Emerald Group Publishing Limited (2012)
Duncan, O.D., Duncan, B.: A methodological analysis of segregation indexes. Am. Sociol. Rev. **20**(2), 210–217 (1955)
Dutta, I., Foster, J.: Inequality of happiness in the U.S.: 1972-2010. Rev. Income Wealth **59**(3), 393–415 (2013)
Elbers, C., Lanjouw, P., Mistiaen, J., Ozler, B.: Reinterpreting between-group inequality. J. Econ. Inequality **6**(3), 231–245 (2008)
Gastwirth, J.L.: Statistical measures of earnings differentials. Am. Stat. **29**(1), 32–35 (1975)
Gravel, N., Magdalou, B., Moyes, P.: Ranking distributions of an ordinal variable. Econ. Theory **71**(1), 33–80 (2021)
Hutchens, R.: Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. Math. Soc. Sci. **21**(1), 31–51 (1991)
Hutchens, R.: Measuring segregation when hierarchy matters. Working papers IZA Discussion Paper No 6667 (2012)
Kobus, M.: Polarization measurement for ordinal data. J. Econ. Inequal. **13**(2), 275–297 (2015)
Kobus, M., Milos, P.: Inequality decomposition by population subgroups for ordinal data. J. Health Econ. **31**(1), 15–21 (2012)
Lambert, P.J., Aronson, J.R.: Inequality decomposition analysis and the Gini coefficient revisited. Econ. J. **103**(420), 1221–1227 (1993)
Langer, A.: Horizontal inequalities and violent group mobilization in Cote d'Ivoire. Oxf. Dev. Stud. **33**(1), 25–45 (2005)
Lazar, A., Silber, J.: On the cardinal measurement of health inequality when only ordinal information is available on individual health status. Health Econ. **22**(1), 106–113 (2013)
Le Breton, M., Michelangeli, A., Peluso, E.: A stochastic dominance approach to the measurement of discrimination. J. Econ. Theory **147**(4), 1342–1350 (2012)
Lieberson, S.: Rank-sum comparisons between groups. Sociol. Methodol. **7**, 276–291 (1976)
Lv, G., Wang, Y., Xu, Y.: On a new class of measures for health inequality based on ordinal data. J. Econ. Inequal. **13**(3), 465–477 (2015)
Mahalanobis, P.C.: A method of fractile graphical analysis. Econometrica **28**(2), 325–351 (1960)
Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. Annals Math. Stat. **18**(1), 50–60 (1947)
Naga, R.H., Yalcin, T.: Inequality measurement for ordered response health data. J. Health Econ. **27**(6), 1614–1625 (2008)
Ostby, G.: Polarization, horizontal inequalities and violent civil conflict. J. Peace Res. **45**(2), 143–162 (2008)
Reardon, S.F.: Measures of ordinal segregation. Res. Econ. Inequality **17**, 129–155 (2009)
Sarkar, S., Santra, S.: Extending the approaches to polarization ordering of ordinal variables. J. Econ. Inequal. **18**(3), 421–440 (2020)
Shorrocks, A.F.: The class of additively decomposable inequality measures. Econometrica **48**(3), 613–25 (1980)
Stewart, F.: Horizontal inequalities and conflict: understanding Group Violence in Multiethnic Societies. Conflict, Inequality and Ethnicity. Palgrave Macmillan (2010)
Thomas, V., Wang, Y., Fan, X.: Measuring education inequality - gini coefficients of education. Policy research working paper series. The World Bank (2525)
Wilcoxon, F.: Individual comparisons by ranking methods. Biomet. Bulletin **1**(6), 80–83 (1945)