Inequality measurement for bounded variables*

Iñaki Permanyer† Suman Seth‡ Gaston Yalonetzky§

June 22, 2023

Abstract

We propose a novel approach for assessing inequality for variables that have fixed lower- and upper bounds. Inequality assessment with such variables is different from inequality assessment with variables lacking fixed upper bounds (such as income) because their respective most unequal distributions are fundamentally different. The maximum-inequality distributions of non-bounded variables, for respective means, always feature every element, but one, equal to their lower bounds, and many existing inequality measures rank these most unequal distributions equally. However, due to domain restrictions, the most unequal distributions of bounded variables contain different proportions of elements being equal to the lower bound, for respective means, and traditional inequality measures rank these most unequal distributions differently. We normatively justify a novel axiom requiring maximum-inequality distributions of bounded variables to be ranked equally, irrespective of their means. Our axiomatically characterised indices measure inequality as the observed proportion of the maximum attainable inequality for a given mean. Furthermore, we characterise a subset of measures that additionally yield consistent inequality comparisons when switching between attainment and shortfall representations of the bounded variable.

Keywords: Inequality measurement, bounded variables, consistency.

JEL Codes: D63, I31, O57.

*We would like to thank Guido Erreygers, Kristof Bosmans, Antonio Villar, Juan Moreno-Ternero, Christopher Barrett, Florent Bresson, Ohiana Aristondo, Casilda Lasso de la Vega, Gordon Anderson, Jacques Silber, Stephan Klasen, Shatakshee Dhondge, Conchita D'Ambrosio and Sanghamitra Bandyopadhyay for very helpful comments and suggestions. We are also grateful to the participants of the HDCA conference 2022; Royal Economic Society Annual Conference 2022; 16th Meeting of the Society for Social Choice and Welfare; 16th Annual Conference on Economic Growth and Development at the Indian Statistical Institute, Delhi; Economics Seminar at Universidad Pablo de Olavide; 36th General Conference of the International Association for Research in Income and Wealth; 14th Meeting of the Society for Social Choice and Welfare, Seoul, South Korea; 13th conference of the Western Economic Association International, Catholic University, Santiago, Chile; the 13th Meeting of the Society for Social Choice and Welfare, Lund University, Sweden; the 7th Meeting of the Society for the Study of Economic Inequality, CUNY, USA; and the Royal Economic Society conference, University of Sussex, UK, for their valuable comments. Permanyer acknowledges funding from the European Research Council (ERC-2019-CoG-864616, HEALIN project) and from the Spanish Ministry of Science and Innovation R+D LONGHEALTH project (PID2021-128892OB-I00). All remaining errors are our own.

†ICREA and Centre d'Estudis Demografics (CED-CERCA). Email: ipermanyer@ced.uab.es.

‡Economics Department, Leeds University Business School, University of Leeds and Oxford Poverty & Human Development Initiative (OPHI), University of Oxford, Oxford, UK. Email: S.Seth@leeds.ac.uk.

§Economics Department, Leeds University Business School, University of Leeds, University of Leeds. Email: G.Yalonetzky@leeds.ac.uk.

1 Introduction

In his seminal contribution, Atkinson (1970) set the foundations of inequality measurement as we know it. After five decades, the contributions to this burgeoning field of research have expanded in multiple directions, and 'inequality' can arguably be considered one of the most hotly debated topics in an increasingly globalised world, as witnessed by the popularity of several recent books on the subject (e.g. Piketty, 2015; Bourguignon, 2017; Atkinson, 2018; Milanovic, 2018). Moreover, the interest in inequality has gone well beyond the study of monetary distributions, such as income or consumption expenditure. Many of these non-pecuniary variables can only take values from a closed finite interval with fixed limits (i.e., the lower bound and the upper bound).[1] Following the literature on inequality measurement, we refer to these variables as *bounded variables* (e.g., see Lambert and Zheng, 2011).

There is a fundamental difference between inequality assessment with bounded variables and inequality assessment with variables that have a fixed and finite lower bound but no fixed upper bound. We refer to these latter variables as non-bounded. Typically, for a given mean and for a given population size, a less unequal distribution is obtained from a more unequal distribution through a transfer from richer to poorer individuals (i.e., progressive transfer); likewise, a more unequal distribution is obtained from a less unequal distribution through a transfer from poorer to richer individuals (i.e., regressive transfer). On one extreme, starting with an unequal distribution, a sequence of progressive transfers eventually leads to an egalitarian distribution where all elements are equal. Whether the variable is bounded or not, such egalitarian distributions are deemed least unequal, and two egalitarian distributions with different means and different population sizes are ranked equally by all classes of inequality measures in the literature. On the other extreme, starting with a less unequal distribution, a sequence of regressive transfers eventually leads to a most unequal distribution (henceforth maximum inequality distribution, or MID) where no further regressive transfers are possible.

This is where inequality measurement for bounded variables departs from the case of non-bounded variables. For a non-bounded variable, all elements in a MID, barring one, are equal to the lower bound. For example, while dividing a cake among ten people, a MID would contain nine people having no slice (e.g., a lower bound of zero) and a single person owning the entire cake. If, instead, there were two identical cakes of the same size, then the MID would feature nine people having no cake at all and one person owning both cakes. Although the average increases from one-tenth to one-fifth of a cake, several inequality measures (e.g., the Gini Coefficient and the Coefficient of Variation) rank these two MIDs equally. However, this scenario can be infeasible for bounded variables. Suppose there is a fixed upper bound so that none can have more than half of a cake. When there is one cake, the MID contains eight people having nothing and two people owning half of the cake each (i.e., the upper bound); whereas, with two cakes, the MID features six people without cake and four owning half of a cake each. Thus, while both in the bounded and unbounded settings inequality is maximized whenever the smallest share of individuals owns as much as possible, there is a fundamental domain restriction shaping the inequality-maximizing

[1]Examples include indicators of education, health, political freedom, democracy level, freedom from violence, happiness, trust, corruption, household or environmental characteristics, access to public services, poverty, socio-demographic characteristics, and so on.

distributions. In the MID of an unbounded variable a single individual always owns everything, whereas in the MID of a bounded variable such possibility is often precluded. Many indicators of social progress have natural upper bounds. How should the MIDs be ranked in such cases?

There are alternative suggestions in the literature regarding how to compare similar distributions for non-bounded variables (Temkin, 1986; Fields, 1998; Bosmans, 2007). However, we argue that, from an egalitarian perspective, the different MIDs of bounded variables represent the normatively least desirable situations for correspondingly different means, and so they should be considered equally unequal irrespective of their means. We refer to this desideratum as the *maximality principle*. We then axiomatically characterize two new closely related classes of inequality measures which abide by the maximality principle in addition to abiding other principles of inequality measurement. These are the so-called classes of *normalised inequality measures*. Each index in the class is defined as the ratio of two identical symmetric S-convex functions such that the function in the denominator is evaluated at a distribution maximizing its value and sharing the same mean as the distribution used to evaluate the function in the numerator, namely the distribution whose inequality is being measured. The key distinction between these two proposed classes is that one is defined for fixed population, while the other class allows populations to vary across distributions.

The measurement of inequality for bounded variables poses another challenge already encountered in the literature. When a variable is bounded, one may choose to focus either on the distribution of attainments or the corresponding distribution of shortfalls with respect to the upper bound.[2] Many inequality measures (especially popular relative measures like the Gini coefficient) fail to rank distributions consistently when measurement is switched from attainments to shortfall representations.[3] A battery of satisfactory solutions has been proposed in the literature, e.g. using absolute inequality measures (Erreygers, 2009; Lambert and Zheng, 2011), indices based on both representations (Lasso de la Vega and Aristondo, 2012), or using pairs of weakly consistent indices (Bosmans, 2016). We further axiomatically characterise a subclass of normalised inequality indices within each class that additionally allow consistent evaluation of inequality across attainments and shortfalls. Remarkably, our solution presents an alternative to all the aforementioned proposals to solve inconsistency.

The rest of the paper proceeds as follows. Section 2 introduces the framework including notation and definitions. Section 3 discusses the concept of maximum inequality in the context of bounded variables and introduces the maximality principle. Sections 4 and 5 introduce and axiomatically characterise the two classes of normalised inequality measures. Section 6 further characterises a subclass within each of the two classes that additionally allows consistent evaluation of inequality. Section 7 provides some comparative insights of our proposed approaches in relation to the existing approaches. Finally, section 8 concludes with some remarks.

[2] For instance, improvements in the coverage of public health programs could be assessed via either the percentage of vaccinated children (an achievement indicator) or the percentage of unvaccinated children (a shortfall indicator).

[3] See, among others, Micklewright and Stewart (1999), Clarke et al. (2002), Kenny (2004), Erreygers (2009), Lambert and Zheng (2011), Lasso de la Vega and Aristondo (2012) and Bosmans (2016).

2 The framework

Let the sets of real numbers, rational numbers and natural numbers be $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{N}$, respectively. The non-negative and strictly positive counterparts of $\mathbb{R}$ and $\mathbb{Q}$ are represented by adding to either the subscripts $_+$ and $_{++}$, respectively. Suppose, there are n units of analysis (e.g. people, households, municipalities, countries, etc.) such that $n \in \mathbb{N} \setminus \{1\}$. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an *attainment distribution* of n units (or an n-dimensional *attainment vector*), where $x_i \in [0, U] \subset \mathbb{R}$ represents unit i's cardinally measurable attainment bounded between a lower bound of zero and some fixed positive upper bound $U \in \mathbb{Q}_{++}$. Given the density of rational numbers within the set of real numbers, the rationality of U is inconsequential for practical applications.[4]

We denote the set of all attainment distributions of size n with upper bound U by $\mathscr{X}_n$ and the set of all possible attainment distributions with upper bound U by $\mathscr{X} := \cup_n \mathscr{X}_n$. The arithmetic mean function evaluated at any $\mathbf{x} \in \mathscr{X}$ is denoted by $\mu(\mathbf{x})$. Furthermore, let $\mathscr{X}_n^{\mu(\mathbf{x})}$ be the set of all attainment distributions of size n with upper bound U and *with the same mean as any* $\mathbf{x} \in \mathscr{X}_n$, and $\mathscr{X}^{\mu(\mathbf{x})}$ be the set of all possible attainment distributions with upper bound U and *with the same mean as any* $\mathbf{x} \in \mathscr{X}$.

Now, the following notation is useful for studying minimum and maximum possible inequality within our framework. We denote the attainment distribution comprising n ones by $\mathbf{1}_n$, hence for any $\lambda \geq 0$, $\lambda \mathbf{1}_n$ is the constant or egalitarian distribution where all n elements are equal to λ. Next, for some $n \in \mathbb{N} \setminus \{1\}$ and for some $U \in \mathbb{Q}_{++}$, let $\mathbb{G}_n = \{U/n, 2U/n, \ldots, (n-1)U/n\}$ denote a set of $n-1$ equally-spaced grid points between U/n and $(n-1)U/n$. A distribution $\mathbf{x} \in \mathscr{X}_n$ is *bipolar* whenever for some $n' \in \mathbb{N}$ such that $n' < n$, n' units in $\mathbf{x}$ attain the value of U and the remaining $n - n'$ units attain the value of 0. Clearly, since $n' \in \{1, ..., n-1\}$, for any bipolar distribution $\mathbf{x} \in \mathscr{X}_n$, $\mu(\mathbf{x}) \in \mathbb{G}_n$. Likewise, we refer to a distribution $\mathbf{x} \in \mathscr{X}_n$ as *almost-bipolar* whenever for some $n' \in \mathbb{N} \cup \{0\}$ such that $n' < n$, n' units in $\mathbf{x}$ attain the value of U, $n - n' - 1$ units in $\mathbf{x}$ attain the value of 0, and the leftover unit attains a value of $\varepsilon = [n\mu(\mathbf{x}) - n'U] \in (0, U)$. For example, suppose $n = 5$ and $U = 1$ and so $\mathbb{G}_5 = \{1/5, 2/5, 3/5, 4/5\}$. Then, distribution $\mathbf{x} = (0, 0, 1, 1, 1)$ is bipolar, where the mean, $\mu(\mathbf{x}) = 3/5$, is an element of $\mathbb{G}_5$; whereas, distribution $\mathbf{y} = (0, 0, 0.5, 1, 1)$ is almost-bipolar (with $\varepsilon = 0.5$), where the mean, $\mu(\mathbf{y}) = 1/2$, is not an element of $\mathbb{G}_5$.

Let $\mathscr{A} \subset \mathscr{X}$ be the set of all possible almost-bipolar distributions; $\mathscr{B} \subset \mathscr{X}$ be the set of all possible bipolar distributions; and let $\mathscr{M} = \mathscr{A} \cup \mathscr{B}$ be the set of all distributions that are either bipolar or almost bipolar. We assign subscript n to denote the subsets with population size n, i.e., $\mathscr{A}_n$, $\mathscr{B}_n$, and $\mathscr{M}_n$. Likewise, we use superscript $\mu(\mathbf{x})$ to denote the subsets with the same mean as $\mu(\mathbf{x})$, i.e., $\mathscr{A}^{\mu(\mathbf{x})}$, $\mathscr{B}^{\mu(\mathbf{x})}$, and $\mathscr{M}^{\mu(\mathbf{x})}$. Finally, we assign both n and $\mu(\mathbf{x})$ to denote the subsets with population size n and the same mean as $\mu(\mathbf{x})$, i.e., $\mathscr{A}_n^{\mu(\mathbf{x})}$, $\mathscr{B}_n^{\mu(\mathbf{x})}$, and $\mathscr{M}_n^{\mu(\mathbf{x})}$.

An *inequality index* $I : \mathscr{X} \to \mathbb{R}_+$ is a continuous real-valued function expected to satisfy two basic properties (Chakravarty, 2009). The first basic property, *anonymity*, requires that an inequality index should not depend on a reordering of attainments across units. Formally, anonymity requires that $I(\mathbf{y}) = I(\mathbf{x})$ for some $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$ whenever $\mathbf{y} = \mathbf{x}\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix.[5] The sec-

[4]Including $U \in \mathbb{R} \setminus \mathbb{Q}$ demands only minor technical adjustments to the proofs, which are available upon request.

[5]A *permutation matrix* is a square matrix with exactly one element in each row and column equal to 1 and the rest of the elements are equal to zero.

ond basic property, *transfer principle*, requires that a transfer from a richer to a poorer unit, without altering their relative positions, should decrease inequality (*progressive transfer*); whereas, alternatively, a transfer from a poorer to a richer unit should increase inequality (*regressive transfer*).[6] Formally, for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$, $I(\mathbf{y}) < I(\mathbf{x})$ whenever $\mathbf{y}$ is obtained from $\mathbf{x}$ by a progressive transfer and $I(\mathbf{y}) > I(\mathbf{x})$ whenever $\mathbf{y}$ is obtained from $\mathbf{x}$ by a regressive transfer.[7]

We also invoke the following additional properties in different parts of the paper. The standard *equality principle* requires that $I(\mathbf{x}) = 0$ whenever $\mathbf{x} = \lambda \mathbf{1}_n$ for any $\mathbf{x} \in \mathscr{X}_n$ and $\lambda \geq 0$. This property ensures that inequality is minimal and equal to zero whenever all units feature exactly the same value for the indicator, i.e., $x_1 = x_2 = \cdots = x_n$. An inequality index is *absolute* if its value remains unchanged when the same amount is added to all attainments, i.e., $I(\mathbf{y}) = I(\mathbf{x})$ for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$ whenever $\mathbf{y} = \mathbf{x} + \lambda \mathbf{1}_n$ for some $\lambda \in \mathbb{R}$; whereas, an inequality index is *relative* if its value remains unchanged when all attainments are altered in the same proportion, i.e., $I(\mathbf{y}) = I(\mathbf{x})$ for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$ whenever $\mathbf{y} = \lambda \mathbf{x}$ for some $\lambda > 0$. The *population principle* requires that whenever $\mathbf{y}$ is obtained from $\mathbf{x}$ by a *replication* for some $\mathbf{x},\mathbf{y} \in \mathscr{X}$, then $I(\mathbf{y}) = I(\mathbf{x})$; where $\mathbf{y} \in \mathscr{X}_{n'}$ for some $n' = \gamma n$ and $\gamma \in \mathbb{N} \setminus \{1\}$ is said to be obtained from $\mathbf{x} \in \mathscr{X}_n$ by a *replication*, whenever $\mathbf{y} = (\mathbf{x},\ldots,\mathbf{x})$, i.e. γ copies of $\mathbf{x}$ are repeated one after the other in $\mathbf{y}$.

We refer to a real valued function $f : \mathscr{X} \to \mathbb{R}_+$ as *symmetric* whenever $f(\mathbf{x}) = f(\mathbf{xP})$ for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$, where $\mathbf{P}$ is a permutation matrix. We refer to a real valued function $f : \mathscr{X} \to \mathbb{R}_+$ as *strictly S-convex* if, for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$, $f(\mathbf{y}) < f(\mathbf{x})$ whenever $\mathbf{y}$ is obtained from $\mathbf{x}$ by a progressive transfer and $f(\mathbf{y}) > f(\mathbf{x})$ whenever $\mathbf{y}$ is obtained from $\mathbf{x}$ by a regressive transfer (Marshall and Olkin, 1979, p. 53-54).

Finally, we define, for every $\mathbf{x} \in \mathscr{X}_n$, a partially ordered set $(\mathscr{X}_n^{\mu(\mathbf{x})}, \succeq_n)$ such that for any pair $\mathbf{y},\mathbf{z} \in \mathscr{X}_n^{\mu(\mathbf{x})}$: (1) $\mathbf{z} \succ_n \mathbf{y}$, which reads "$\mathbf{z}$ is more unequal than $\mathbf{y}$", if $\mathbf{z}$ is obtained from $\mathbf{y}$ through a sequence of regressive transfers with or without additional permutations; and (2) $\mathbf{z} \sim_n \mathbf{y}$, which reads "$\mathbf{z}$ is as unequal as $\mathbf{y}$" if $\mathbf{z}$ is obtained from $\mathbf{y}$ only through a sequence of permutations. Likewise, we can order distributions across all population sizes. That is, for every $\mathbf{x} \in \mathscr{X}$, we can define a partially ordered set $(\mathscr{X}^{\mu(\mathbf{x})}, \succeq)$ such that for any pair $\mathbf{y},\mathbf{z} \in \mathscr{X}^{\mu(\mathbf{x})}$: (1) $\mathbf{z} \succ \mathbf{y}$, which reads "$\mathbf{z}$ is more unequal than $\mathbf{y}$", if $\mathbf{z}$ is obtained from $\mathbf{y}$ through a sequence of regressive transfers with or without additional permutations and/or replications; and (2) $\mathbf{z} \sim \mathbf{y}$, which reads "$\mathbf{z}$ is as unequal as $\mathbf{y}$" if $\mathbf{z}$ is obtained from $\mathbf{y}$ only through a sequence of permutations and/or replications.

---

[6] Technically, for some $\mathbf{x},\mathbf{y} \in \mathscr{X}_n$, $\mathbf{y}$ is obtained from $\mathbf{x}$ by a *progressive transfer* whenever there are two units i, j and some $k > 0$ such that $y_i = x_i + k \leq x_j - k = y_j$ and $y_l = x_l$ for every $l \neq i, j$. Alternatively, $\mathbf{y}$ is obtained from $\mathbf{x}$ by a *regressive transfer* whenever there are two units i, j and some $k > 0$ such that $y_i + k = x_i \leq x_j = y_j - k$ and $y_l = x_l$ for every $l \neq i, j$.

[7] Some of the bounded indicators discussed in this paper are not literally transferable. For instance, we do not consider 'uneducating' highly educated individuals and transferring that education to less educated ones. Yet, one can compare two hypothetical scenarios, e.g. pre- and post-"progressive transfers", and still judge the latter exhibiting lower inequality than the former.

3 Maximum inequality distributions and the Maximality Principle

Inequality measurement with bounded variables is conceptually different from the traditional approach for unbounded variables such as income (i.e., variables with a lower bound of zero but no upper bound). Being bounded from both sides, the values of a bounded variable cannot increase or decrease uninhibitedly. Moreover, inequality measurement in the context of an unbounded variable is often seen as a cake-cutting problem (e.g., see Cowell, 2011), where the most unequal distribution always involves one person owning the entire cake. Such scenario is generally precluded with bounded variables, which in turn calls for a different ethical intuition regarding the most unequal distribution.

Let us elaborate how the notion of maximum inequality can be different for bounded variables by way of a hypothetical five-person society. With an average income of 2, distribution $\mathbf{u} = (0,0,0,0,10)$ would be considered most unequal in the unbounded setting. If the mean income in the same society increases to 4, then distribution $\mathbf{v}_1 = (0,0,0,0,20)$ would be considered most unequal. Now, suppose instead that there is a non-monetary bounded variable with a lower bound of 0 and an upper bound of 10. In the same five person society with a mean of 2, distribution $\mathbf{u} = (0,0,0,0,10)$ would still be considered most unequal. However, if the mean in the society increases to 4, then distribution $\mathbf{v}_2 = (0,0,0,10,10)$ (*not* $\mathbf{v}_1$) would be considered most unequal, because the smallest possible number of individuals own as much as possible and the others nothing, so no further regressive transfers are feasible.

Let us formalise these ideas, showing that maximum-inequality distributions for a given mean exist and what they look like for a bounded variable. Proposition 1 establishes the existence of a set of *maximum-inequality distributions* (MIDs) and shows that the set of MIDs associated to any distribution $\mathbf{x} \in \mathscr{X}_n$ is in fact equal to $\mathscr{M}_n^{\mu(\mathbf{x})} = \mathscr{X}_n^{\mu(\mathbf{x})} \cap \mathscr{M}$. Based exclusively on the transfer principle and the anonymity property, such MIDs are defined as the distributions that maximise inequality among all possible distributions with the *same population size n* and the *same mean* $\mu(\mathbf{x})$.

Proposition 1 For any $n \in \mathbb{N} \setminus \{1\}$ and for any $\mathbf{x} \in \mathscr{X}_n$ such that $\mu(\mathbf{x}) \in (0,U)$, a set of maximum inequality distributions $\mathscr{M}_n^{\mu(\mathbf{x})} = \mathscr{X}_n^{\mu(\mathbf{x})} \cap \mathscr{M}$ constituting the maximal elements of the partially ordered set $(\mathscr{X}_n^{\mu(\mathbf{x})}, \succeq_n)$ exists and the elements of $\mathscr{M}_n^{\mu(\mathbf{x})}$ are *bipolar* when $\mu(\mathbf{x}) \in \mathbb{G}_n$ or *almost-bipolar* when $\mu(\mathbf{x}) \notin \mathbb{G}_n$.

Proof. See Appendix A1. ■

As it turns out, the MIDs are either bipolar or almost bipolar. Bipolar distributions consist of units with values at either the lower bound or upper bound exclusively, with at least one unit at each bound (as otherwise, should all units have the same values, the distribution would be egalitarian). Meanwhile, almost bipolar distributions consist of all units with either the lower or upper bound value, except for one unit with an interior value of $\varepsilon \in (0,U)$. The elements included in $\mathscr{M}_n^{\mu(\mathbf{x})}$ are unique up to permutations, that is: given any two elements $\mathbf{x},\mathbf{y} \in \mathscr{M}_n^{\mu(\mathbf{x})}$, then $\mathbf{y} = \mathbf{xP}$ for some permutation matrix $\mathbf{P}$. Finally, also note that a set of MIDs defined in proposition 1 is unique for a given mean and for a given n. In other words, for any two distributions $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n^{\mu(\mathbf{x})}$ such that $\mathbf{x} \neq \mathbf{y}$, $\mathscr{M}_n^{\mu(\mathbf{x})} = \mathscr{M}_n^{\mu(\mathbf{y})}$.

Example 1: Assume $U = 1$, $n = 4$ and so $\mathbb{G}_4 = \{0.25, 0.5, 0.75\}$. Consider the distribution $\mathbf{x} = (0.1, 0.4, 0.7, 0.8)$ with $\mu(\mathbf{x}) = 0.5 \in \mathbb{G}_4$. In this case, the corresponding set of MIDs $\mathscr{M}_4^{0.5}$ contains all possible permutations of the distribution $\hat{\mathbf{x}} = (0, 0, 1, 1)$, which is bipolar, and clearly $\mu(\hat{\mathbf{x}}) = \mu(\mathbf{x}) = 0.5$. Now consider a second distribution $\mathbf{y} = (0.2, 0.4, 0.7, 0.9)$ with $\mu(\mathbf{y}) = 0.55 \notin \mathbb{G}_4$. The corresponding MIDs $\mathscr{M}_4^{0.55}$, in this case, are all possible permutations of the distribution $\hat{\mathbf{y}} = (0, 0.2, 1, 1)$ with $\mu(\hat{\mathbf{y}}) = \mu(\mathbf{y}) = 0.55$ and $\varepsilon = 0.2 \in (0, 1)$, but $\hat{\mathbf{y}}$ is almost-bipolar and no further regressive transfers are possible.

Even though MIDs are hypothetical distributions unlikely to be observed in practice, they do represent the benchmark case of maximum inequality against which we can compare distributions of bounded variables sharing the same mean. The latter's inequality evaluations cannot be larger than their MID's as long as an inequality index I satisfies anonymity and the transfer principle. Note that the corresponding MIDs in the classical unbounded setting are entirely different: there, inequality is maximized whenever one individual owns everything and the others nothing.

3.1 Comparing MIDs with different means

How should we compare MIDs? Recalling the hypothetical five-person society introduced above, only $\mathbf{u}$ and $\mathbf{v}_1$ reflect maximum inequality in the context of an unbounded variable. Interestingly, even with different means, they are judged equally unequal by Lorenz-consistent relative inequality measures, such as the Gini index and the Coefficient of Variation (whereas $\mathbf{v}_2$ would be deemed less unequal). On the contrary, distributions $\mathbf{u}$ and $\mathbf{v}_2$ reflect maximum inequality for bounded variables when the upper bound is 10. How should we compared them given that they both reflect maximum inequality for bounded variables?

If we focus on bipolar MIDs, then their comparison is analogous to those studies assessing inequality changes due to a population shift between a low-income sector and a high-income sector owing to income growth (Fields, 1987, 1993, 1998; Amiel and Cowell, 1994; Bosmans, 2007) or due to a sequence of population shifts between a better-off group and a worse-off group (Temkin, 1986). Suppose there are two sectors in a society: a high-income sector (better-off) and a low-income sector (worse-off). Everyone within a sector is equally well-off (i.e., no inequality within each sector). Suppose there are n persons and consider the following $n-1$ situations. In the first (initial) situation, there are $n-1$ persons in the low-income sector and one person in the high-income sector; in the second situation, there are $n-2$ persons in the low-income sector and two persons in the high-income sector; and so on. Finally, in the $(n-1)^{\text{th}}$ (final) situation, there is one person in the low-income sector and $n-1$ persons in the high-income sector. As we gradually move from the initial to the final situation, the mean certainly improves, but how should inequality change?

Five possible ethical judgements have been discussed in the literature as the mean improves along with the shift of population from the low-income to the high-income sector: (i) an increase in inequality throughout; (ii) a decrease in inequality throughout; (iii) an initial increase in inequality, then a reduction after a certain point where inequality is maximised (i.e., an inverted U-shape);

(iv) an initial reduction in inequality, then an increase after a certain point where inequality is minimised (i.e., a U-shape); and (v) no change in inequality. Temkin (1986) and Fields (1998) both agree on the possibility for inequality to be increasing throughout (as an ever smaller number of people become victimised through the *isolation of the poor*) as well as the possibility for inequality to decrease throughout (reflecting a diminished *elitism of the rich* and the steady decrease in the number of those worse-off). Temkin (1986) and Fields (1998), however, disagree on the possibilities of a U-shape or an inverted-U-shape relationship. Temkin (1986) argues in favour of the possibility of an inverted-U-shape relationship; whereas, Fields (1998) argues in favour of a U-shape relationship. Nevertheless, Bosmans (2007) shows that quasi-concave inequality measures (comprising numerous relative, absolute and intermediate inequality measures) allow only the first three possibilities: increasing inequality, decreasing inequality and inverted-U-shape.

Therefore, the literature on income inequality measurement features arguments in favour of the first four ethical judgements, but without straightforward guidance to prefer one over the other. Importantly, a key difference emerges with bounded variables as, unlike the case of income, each of the $n-1$ situations mentioned above corresponds to a bipolar MID for a given mean. Hence, for a fixed population size, there are good reasons to avoid ranking MIDs with different means as better or worse than each other. For instance, all MIDs reflect situations in which inequality cannot increase any further through regressive transfers. In fact, we cannot transform one MID into another one through such transfers. Analogously, we cannot transform one egalitarian distribution into another one through progressive transfers. Hence, we may rank MIDs with different means in the same way that we rank egalitarian distributions with different means; namely, equally.

Thus, we consider the MIDs as equally unequal, which is in the spirit of the fifth posited change in inequality as the mean gradually improves along with the shift of population from the low-income to the high-income sector. While arguing in favour of this fifth possibility, Temkin (1986, p. 118) eloquently stated that "two judges who accepted bribes in all of their cases might be equally corrupt, even if one tried fewer cases." We thus operationalise this ethical intuition with a novel property called the *maximality principle* as follows:

Maximality Principle: For any $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\}$, $I(\mathbf{x}) = I(\mathbf{y})$ whenever $\mathbf{x} \in \mathscr{M}_n^{\mu(\mathbf{x})}$ and $\mathbf{y} \in \mathscr{M}_n^{\mu(\mathbf{y})}$.

The property requires that, whenever we pick any two (non-trivial) MIDs, the corresponding levels of inequality must coincide. Stated otherwise, whenever no further regressive transfers can be performed, then we have reached maximal inequality *irrespective of the mean of the distribution*. Additionally, satisfaction of the Maximality Principle guarantees inequality comparisons not to be predictably related to the mean (should that be the desideratum). Interestingly, in the unbounded setting this axiom is also satisfied by popular relative inequality measures such as the Gini coefficient or the coefficient of variation whenever the population size n is fixed. Indeed, these indices attain their maximum level whenever one individual owns everything and no further regressive transfers are feasible, irrespective of how much such privileged individual owns.

4 The class of normalised inequality indices for fixed population

Building on the basic properties and the maximality principle, we characterise a class of new inequality indices. We show that, within our framework, inequality should be measured as a proportion of the maximum level of inequality reachable given a mean attainment. The maximality principle also ensures that any distribution different from the corresponding MID obtains a strictly lower inequality value. We refer to our proposed family of inequality indices as the class of *normalised inequality indices*, which are presented in theorem 1.

Theorem 1 For any $n \in \mathbb{N} \setminus \{1\}$ and any $\mathbf{x} \in \mathscr{X}_n$, an inequality index I satisfies anonymity, the transfer principle, the equality principle and the maximality principle if and only if

$$I(\mathbf{x}) = \begin{cases} M\left[\dfrac{f(\mathbf{x}) - f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}\right] & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\} \\ 0 & \text{if } \mathbf{x} \in \{\mathbf{0}, \mathbf{U}\} \end{cases}, \tag{1}$$

where $0 < M < +\infty$ is a proportionality constant, $\bar{\mathbf{x}} = \mu(\mathbf{x})\mathbf{1}_n$ is the egalitarian distribution with the same mean as $\mathbf{x}$, $\hat{\mathbf{x}} \in \mathscr{M}_n^{\mu(\mathbf{x})}$ is an MID for $\mathbf{x}$, $f : \mathscr{X}_n \to \mathbb{R}_{++}$ is a symmetric and strictly S-convex function, and $\mathbf{0} = 0\mathbf{1}_n$ and $\mathbf{U} = U\mathbf{1}_n$ are the two extreme egalitarian distributions.

Proof. See Appendix A2. ∎

According to theorem 1, a normalised inequality index $I(\mathbf{x})$ in our proposed class evaluated at distribution $\mathbf{x}$ is proportional to any symmetric and S-convex function $f(\mathbf{x})$ evaluated at $\mathbf{x}$, subtracted by its corresponding minimum possible value $f(\bar{\mathbf{x}})$ evaluated at $\bar{\mathbf{x}}$, and then normalised by the difference between its corresponding maximum possible value $f(\hat{\mathbf{x}})$ evaluated at any of its uniquely associated MIDs, namely $\hat{\mathbf{x}} \in \mathscr{M}_n^{\mu(\mathbf{x})}$, and its corresponding minimum possible value $f(\bar{\mathbf{x}})$ evaluated at $\bar{\mathbf{x}}$.

All indices in our proposed class conveniently range between zero and a finite upper bound corresponding to a proportionality constant $M > 0$, where the former value is achieved in the absence of inequality and the latter corresponds to an MID. The value of a normalised index increases with a regressive transfer and decreases owing to a progressive transfer. For ease of presentation, from now onwards we assume $M = 1$, but other choices are certainly possible without affecting the results (other than re-scaling inequality levels). Crucially, numerous functional forms of f are admissible, including entire classes of relative, absolute, intermediate, super-relative or super-absolute inequality measures, for instance, inequality indices from the Atkinson class, Generalised entropy, the class of indices proposed by Lasso de la Vega and Aristondo (2012), etc.[8]

Examples

We present a few examples of normalised inequality indices derived from popular inequality measures. For convenience of presentation, we refer to the normalised inequality index corresponding

[8]See Bosmans (2016) for a concise and comprehensive typology.

to the admissible form f as f^* (instead of I), in order to clarify that f^* is derived from an admissible f. That is, f^* is the normalised version of f.

When f is the absolute or the relative Gini index (i.e., $f(\mathbf{x}) = G_a(\mathbf{x})$ or $f(\mathbf{x}) = G_r(\mathbf{x})$) then it is easy to check (see Appendix A7) that

$$G_a^*(\mathbf{x}) = G_r^*(\mathbf{x}) = \begin{cases} \dfrac{G_a(\mathbf{x})U}{\mu(\mathbf{x})(U-\mu(\mathbf{x}))} & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0},\mathbf{U}\} \text{ and } \mu(\mathbf{x}) \in \mathbb{G}_n \\[2ex] \dfrac{G_a(\mathbf{x})n^2}{(n-n'-1)(\varepsilon+n'U)+n'(U-\varepsilon)} & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0},\mathbf{U}\} \text{ and } \mu(\mathbf{x}) \notin \mathbb{G}_n \\[2ex] 0 & \text{if } \mathbf{x} \in \{\mathbf{0},\mathbf{U}\} \end{cases} .^{9} \tag{2}$$

Thus, $G_a^*(\mathbf{x})$ compares $G_a(\mathbf{x})$ against the maximal inequality value that such index could possibly take for any distribution with mean equal to $\mu(\mathbf{x})$ (which equals $\mu(\mathbf{x})(U-\mu(\mathbf{x}))/U$ when the MID is bipolar, i.e. when $\mu(\mathbf{x}) \in \mathbb{G}_n$). Remarkably, the normalised inequality indices derived from the absolute and the relative Gini indices coincide. To simplify notation, such normalised Gini index will be referred to as $G^*(\mathbf{x})$.

We can also derive the normalised versions of the standard deviation ($f(\mathbf{x}) = \sigma(\mathbf{x}) = \sqrt{V(\mathbf{x})}$) and the coefficient of variation ($f(\mathbf{x}) = CV(\mathbf{x}) = \sigma(\mathbf{x})/\mu(\mathbf{x})$). It is easy to check (see Appendix A7) that $\sigma^*(\mathbf{x}) = CV^*(\mathbf{x})$, i.e.,

$$\sigma^*(\mathbf{x}) = \begin{cases} \dfrac{\sigma(\mathbf{x})}{\sqrt{\mu(\mathbf{x})(U-\mu(\mathbf{x}))}} & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0},\mathbf{U}\} \text{ and } \mu(\mathbf{x}) \in \mathbb{G}_n \\[2ex] \dfrac{\sigma(\mathbf{x})\sqrt{n}}{\sqrt{(n-n'-1)\mu(\mathbf{x})^2+n'(U-\mu(\mathbf{x}))^2+(\varepsilon-\mu(\mathbf{x}))^2}} & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0},\mathbf{U}\} \text{ and } \mu(\mathbf{x}) \notin \mathbb{G}_n \\[2ex] 0 & \text{if } \mathbf{x} \in \{\mathbf{0},\mathbf{U}\} \end{cases} . \tag{3}$$

Once again, the normalised version of an absolute inequality index and its relative counterpart coincide. More generally, whenever $a(\mathbf{x})$ is an absolute inequality index and $r(\mathbf{x}) = a(\mathbf{x})/\mu(\mathbf{x})$ is its relative counterpart, one can easily check that $a^*(\mathbf{x}) = r^*(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{X}_n$.

While all normalised inequality measures are defined in the same way (i.e., as a fraction of the maximal inequality level that can be reached given a mean attainment and a fixed population size; see Theorem 1), their explicit mathematical representation differs slightly depending on whether $\mu(\mathbf{x}) \in \mathbb{G}_n$ or $\mu(\mathbf{x}) \notin \mathbb{G}_n$. It is easy to check that, when the decimal precision one is working with is fixed (i.e., numbers are represented with a precision of $k \in \mathbb{N}$ decimals, so all numbers with a higher number of decimals are rounded), then for sufficiently large values of n (more specifically, when $n \geq 10^k/U$) one always has $\mu(\mathbf{x}) \in \mathbb{G}_n$. When this happens, the corresponding MID is bipolar, thus leading to more compact formulations for the corresponding normalised inequality measures and

[9] The reader is reminded that $n' < n$ is the number of units in $\mathbf{x}$ attaining U and $\varepsilon = [n\mu(\mathbf{x}) - n'U]$ (section 2).

easier calculations (see equations 2 and 3). For instance, in the fairly common case where $k = 2$ and $U = 1$, when $n \geq 100$ then all MIDs are bipolar.

5 The class of normalised inequality indices for varying population sizes

At least since Dalton (1920), the most popular answer to the challenge of comparing inequality across distributions with different population sizes is the *population principle*, which requires that identical cloning of all units should leave inequality unaltered (thereby rendering populations with different sizes comparable).[10] A normalised inequality measure from the class in theorem 1 does not comply with the population principle even when an admissible functional form of f does, because even though the replication of a bipolar MID is itself an MID, the replication of an almost bipolar MID is *not* an MID, based on how proposition 1 defines an MID.[11] Therefore if we want our normalised inequality measures to fulfil the population principle we must adopt a different definition of the set of MIDs, one compliant with the population principle. Proposition 2 establishes the existence of a set of MIDs and shows that the set of MIDs, associated with all distributions *sharing the same mean* across *all population sizes* is, in this case, equal to $\mathscr{B}^{\mu(\mathbf{x})} = \mathscr{X}^{\mu(\mathbf{x})} \cap \mathscr{B}$.[12] Based on the transfer and population principles combined with anonymity, these MIDs are defined as the distributions that maximise inequality among all possible distributions with the *same mean* but *varying population sizes*.

Proposition 2 For any $\mathbf{x} \in \mathscr{X}$ such that $\mu(\mathbf{x}) \in (0, U)$, a set of maximum inequality distributions $\mathscr{B}^{\mu(\mathbf{x})} = \mathscr{X}^{\mu(\mathbf{x})} \cap \mathscr{B}$ constituting the maximal elements of the partially ordered set $(\mathscr{X}^{\mu(\mathbf{x})}, \succeq)$ exists and all elements of $\mathscr{B}^{\mu(\mathbf{x})}$ are *bipolar*.

Proof. See Appendix A3. ■

According to proposition 2, in a setting compliant with the population principle, *only* bipolar distributions maximise inequality. Thus, the Maximality Principle introduced in section 3 must be adapted and rewritten as follows:

Restricted Maximality Principle: For any $\mathbf{x}, \mathbf{y} \in \mathscr{X}$, $I(\mathbf{x}) = I(\mathbf{y})$ whenever $\mathbf{x} \in \mathscr{B}^{\mu(\mathbf{x})}$ and $\mathbf{y} \in \mathscr{B}^{\mu(\mathbf{y})}$.

Again, this principle states that whenever no further regressive transfers are feasible and we have reached a bipolar distribution, then inequality is maximal (no matter what the mean of the distribution is). With this reformulated version of the Maximality Principle, we can now axiomatically characterise the class of normalised inequality indices compliant with the population principle:

[10] For a more general proposal, see Aboudi et al. (2010). The population principle is defined in Section 2.

[11] For instance, when $n = 2$ and $U = 1$, an MID associated to a distribution with mean equal to 0.25 is $(0, 0.5)$. However, the replication $(0, 0.5, 0, 0.5)$ of that MID is not an MID itself. The corresponding MID for a distribution with $n = 4$ and with mean equal to 0.25 is in fact $(0, 0, 0, 1)$.

[12] Recall that $\mathscr{B}$ is the set of all bipolar distributions and $\mathscr{B}^{\mu(\mathbf{x})}$ is the subset of $\mathscr{B}$ with the same mean as $\mu(\mathbf{x})$.

Theorem 2 For any $n \in \mathbb{N} \setminus \{1\}$ and for any $\mathbf{x} \in \mathscr{X}_n$, an inequality index I satisfies anonymity, the transfer principle, the equality principle, the restricted maximality principle and *the population principle* if and only if

$$I(\mathbf{x}) = \begin{cases} M\left[\dfrac{f(\mathbf{x}) - f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}\right] & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\} \\ 0 & \text{if } \mathbf{x} \in \{\mathbf{0}, \mathbf{U}\} \end{cases},$$

where $0 < M < +\infty$ is a proportionality constant, $\bar{\mathbf{x}} = \mu(\mathbf{x})\mathbf{1}_n$ is the egalitarian distribution with the same mean as $\mathbf{x}$, $\hat{\mathbf{x}} \in \mathscr{B}^{\mu(\mathbf{x})}$ is a bipolar MID for $\mathbf{x}$, $f : \mathscr{X} \to \mathbb{R}_{++}$ is a symmetric and strictly S-convex function satisfying the *population principle*, and $\mathbf{0} = 0\mathbf{1}_n$ and $\mathbf{U} = U\mathbf{1}_n$ are the two extreme egalitarian distributions.

Proof. See Appendix A4. ■

Theorem 2 implies that we can construct normalised inequality indices that abide by the population principle as long as f satisfies the population principle and it is *evaluated at any bipolar distribution with mean equal to* $\mu(\mathbf{x})$ *in the denominator of* $I(\mathbf{x})$.

6 Consistency requirement

Bounded variables, such as mortality or literacy rates, or access to basic facilities, can be represented as attainments (implicitly measuring their distance from the lower bound) or, alternatively, as shortfalls (their distance from the upper bound). If $\mathbf{x} \in \mathscr{X}_n$ denotes the attainment distribution then we define the *shortfall distribution* associated with it as $\mathbf{x}^S = (x_1^S, \ldots, x_n^S) \in \mathscr{X}_n$ with $x_i^S = U - x_i$ representing i's shortfall from the upper bound U. Inconsistency in inequality measurement arises when inequality orderings of attainment distributions differ from their shortfall counterparts. This challenge has received significant attention in the literature on inequality measurement with bounded variables.

The literature has proposed different properties regarding the extent to which inequality indices, as well as incomplete partial orderings, should consistently rank attainment and shortfall distributions.[13] *Perfect complementarity*, for instance, requires that the value of the inequality index remains unaltered when we switch between attainment and shortfall representations of the same distribution, i.e., $I(\mathbf{x}) = I(\mathbf{x}^S)$ for any $\mathbf{x} \in \mathscr{X}_n$ (Erreygers, 2009). *Strong consistency*, likewise, requires that inequality measures should rank pairs of attainment distributions and their shortfall counterparts in a coherent manner. In other words, the inequality ranking should be robust to alternative representations of the variable, i.e., $I(\mathbf{x}) \leq I(\mathbf{y}) \Leftrightarrow I(\mathbf{x}^S) \leq I(\mathbf{y}^S)$ for any $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$ (Lambert and Zheng, 2011). *Weak consistency* (Bosmans, 2016), however, is predicated on the realisation that it is possible to find pairs of different inequality indices that produce consistent comparisons as long as one index I^A is used for the attainment distribution and another index $I^S = \phi(I^A)$ is used for the shortfall counterpart, where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing function. The

[13] Arguably, the concern for different degrees of consistency may be more pressing when both representations (attainment and shortfall) can be deemed “different sides of the same coin” (Clarke et al., 2002, p. 1927), warranting equal attention.

pair (I^A, I^S) is *jointly* weakly consistent if and only if $I^A(\mathbf{x}) \leq I^A(\mathbf{y}) \Leftrightarrow I^S(\mathbf{x}^S) \leq I^S(\mathbf{y}^S)$ for any $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$. For example, if $I^A(\mathbf{x})$ is the Gini coefficient evaluated on the attainment distribution $\mathbf{x}$, then $I^S(\mathbf{x}^S) = \mu(\mathbf{x}^S) I^A(\mathbf{x}^S)/[U - \mu(\mathbf{x}^S)]$ provides a jointly weakly consistent inequality evaluation for the shortfall distribution.

Some advocates of perfect complementarity and strong consistency (Erreygers, 2009; Lambert and Zheng, 2011; Chakravarty et al., 2016; Seth and Alkire, 2017) suggest using absolute inequality indices (and related incomplete partial orderings). However, Lasso de la Vega and Aristondo (2012) showed that strong consistency is satisfied by a wider class of indices derived from equally weighted generalised means of any inequality index evaluated at the attainment distribution and the same index evaluated at the corresponding shortfall distribution. Finally, Bosmans (2016) showed that weak consistency is satisfied by a broad class of *pairs of inequality indices*, including relative ones like the Gini coefficient and its respective weakly-consistent counterpart.

We should point out at this stage that strongly consistent inequality measures violate the transfer sensitivity axiom (Lambert and Zheng, 2011, theorem 6). That is, strongly consistent inequality indices do not systematically prioritise transfers at particular parts of the distributions. Lambert and Zheng (2011), however, argue that this should not necessarily be a problem because a transfer-sensitive inequality index that decreases more when progressive transfers of attainments occur at the bottom of the distribution would also decrease more when progressive transfers of shortfalls take place among the lowest shortfalls. Yet, arguably, the latter may not be a desirable feature and imposing strong consistency precludes such dilemma. Thus, following Lambert and Zheng (2011), we can impose the strong consistency requirement for the proposed new classes of inequality measures to free them from inconsistency.

The task of rendering the proposed inequality measures in compliance of strong consistency is facilitated by the remarkable equivalence between strong consistency and perfect complementarity. We know that the latter implies the former. But proposition 3 shows that strong consistency also implies perfect complementarity:

Proposition 3 *An inequality index is strongly consistent if and only if it satisfies perfect complementarity.*

Proof. See Appendix A5. ■

Then the subclasses of normalised inequality indices in fulfillment of strong consistency are characterised in theorem 3:

Theorem 3 The inequality indices I characterised in theorems 1 and 2 are also strongly consistent if and only if

$$f(\mathbf{x}^S) = p(\mathbf{x}) f(\mathbf{x}) + q(\mathbf{x}), \tag{4}$$

where $p(\mathbf{x}) = \dfrac{f(\hat{\mathbf{x}}^S) - f(\bar{\mathbf{x}}^S)}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}$; and $q(\mathbf{x}) = \dfrac{f(\hat{\mathbf{x}}) f(\bar{\mathbf{x}}^S) - f(\hat{\mathbf{x}}^S) f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}$.

Proof. See Appendix A6. ■

Theorem 3 helpfully restricts the subclass of admissible functions for f. For instance, any absolute inequality index as well as any member of the Lasso de la Vega and Aristondo (2012) class is suitable member for f because for all these options $f(\mathbf{x}^S) = f(\mathbf{x})$, i.e., $p(\mathbf{x}) = 1$ and $q(\mathbf{x}) = 0$. Some relative indices like the Gini or the coefficient of variation are also admissible. For instance, when f is the Gini index we obtain $p(\mathbf{x}) = \frac{\mu(\mathbf{x})}{U-\mu(\mathbf{x})}$ and $q(\mathbf{x}) = 0$. By contrast, even though all members of the Atkinson class are suitable choices for f in theorems 1 and 2, many of them are not suitable for f in theorem 3. That is, we cannot obtain strongly consistent normalised inequality indices using every member of the Atkinson class.

7 Further comparative insights

We now provide some insights into how the two proposed classes of normalised inequality indices (in sections 4 and 5) compare with each other as well as how they both compare with standard absolute and relative measures. First, note that the two approaches to measuring normalised inequality (corresponding to the two definitions of MIDs and their respective classes of indices) bear a large degree of overlap. In fact, the formulae for normalised inequality indices compliant with the population principle (section 5) is identical to the corresponding formulae for indices suitable for fixed population sizes (section 4) whenever $\mu(\mathbf{x}) \in \mathbb{G}_n$.[14] As argued before, when the population size n is sufficiently large and the decimal precision is kept fixed (as is the case in many empirical applications), the condition $\mu(\mathbf{x}) \in \mathbb{G}_n$ is always satisfied.

In the context of $n = 2$, we provide insights on how the different normalised inequality measures behave and compare vis-a-vis each other, and with respect to standard absolute and relative inequality measures using the Gini coefficient. The non-trivial case with $n = 2$ lays the foundation for how the corresponding inequality indices behave for the more general case of $n > 2$. Furthermore, the simplicity of the $n = 2$ setting allows a neat inspection of the iso-inequality level contours, which can be thought as the fingerprint of the corresponding inequality measures.[15] Figure 1 presents the iso-inequality contours of the absolute Gini index (G_a, panel A), the relative Gini index (G_r, panel B), the normalised Gini index based on Theorem 1 (i.e., for fixed population; G^*, panel C), and the normalised Gini index complying with the Population Principle (G_P^*, panel D), in the case where $U = 1$ (Appendix A8 shows how we arrive at these iso-inequality contours).[16]

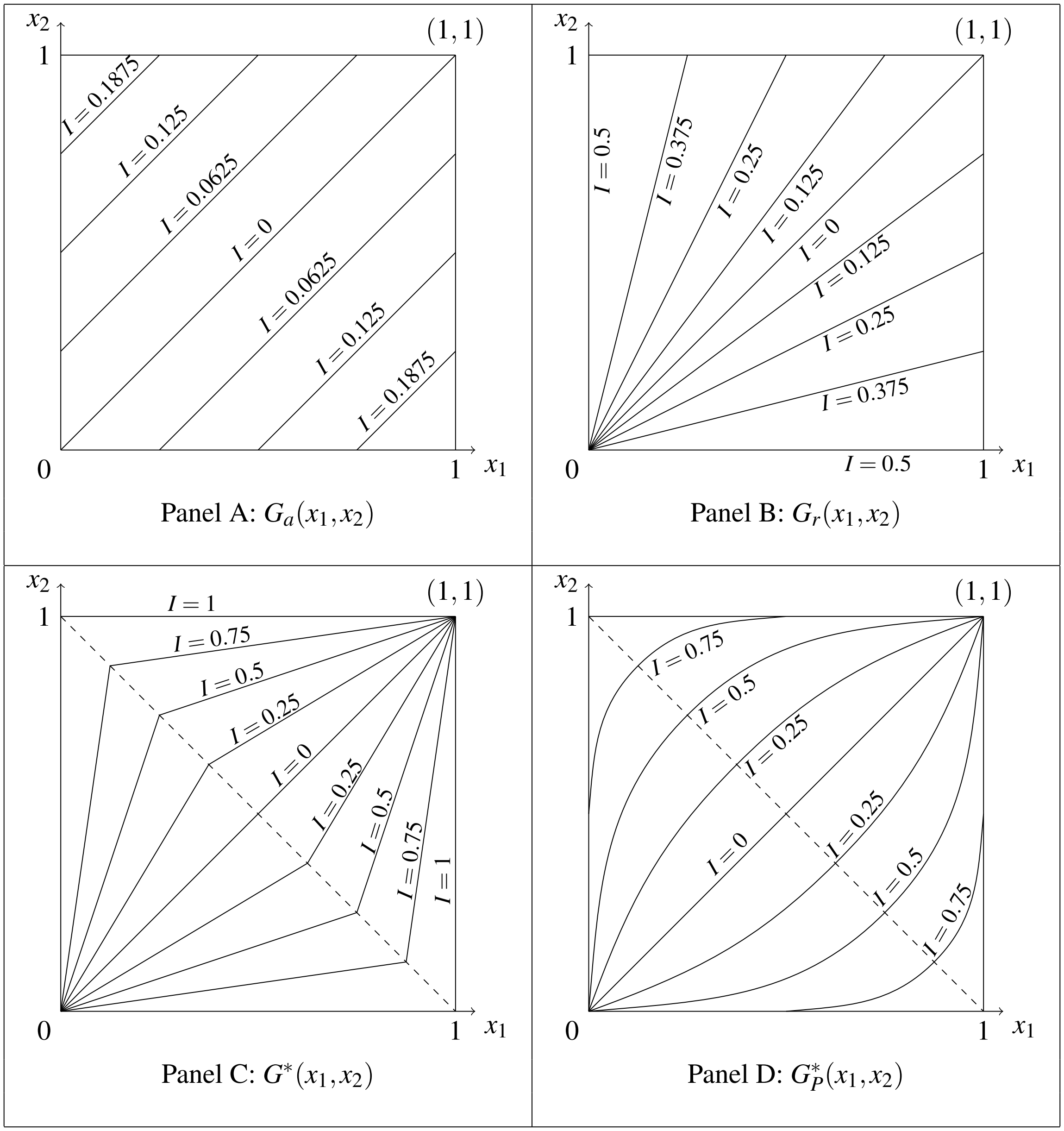
As is well-known, $G_a(x_1,x_2) \in [0,0.25]$ and the iso-inequality contours for G_a are parallel to the 45° line, while $G_r(x_1,x_2) \in [0,0.5]$ and the iso-inequality contours for G_r are straight lines 'emanating from' (or 'converging to') the origin $(0,0)$. In contrast, the iso-inequality contours for the two normalised Gini indices, $G^*(x_1,x_2) \in [0,1]$ and $G_P^*(x_1,x_2) \in [0,1]$, exhibit completely different shapes. In the case of G^*, all level contours are made of two line segments meeting in the diagonal $\{(x_1,x_2) \in [0,1]^2 \mid x_1 + x_2 = 1\}$, which, together, connect the points $(0,0)$ and $(1,1)$. Their shapes

[14] Readers are reminded of our examples of normalised inequality indices in section 4, whose formulae vary depending on whether $\mu(\mathbf{x}) \notin \mathbb{G}_n$ (i.e., almost bipolar MIDs) or $\mu(\mathbf{x}) \in \mathbb{G}_n$ (i.e., bipolar MIDs).

[15] Indeed, the behavior of some very popular inequality measures like the Gini index or the Variance (which can be defined on a recursive basis) is entirely determined by what happens in the most basic case $n = 2$ (see details in Ebert, 2010).

[16] Results remain essentially unaltered when the absolute and relative Gini indices are substituted by the standard deviation and the coefficient of variation, respectively.

Figure 1: Iso-inequality contours for the different Gini coefficient ($n = 2$ and $U = 1$)



Notes: The figure is based on $n = 2$ and $U = 1$. $G_a(x_1,x_2)$ in panel A is the absolute Gini index applied to an attainment distribution. $G_r(x_1,x_2)$ in panel B is the relative Gini index applied to the same attainment distribution. $G^*(x_1,x_2)$ in panel C is the normalised Gini index applied to the attainment distribution for fixed population. $G^*_P(x_1,x_2)$ in panel in panel D is the normalised Gini index applied to the attainment distribution for variable population.

(though not their corresponding inequality levels) coincide with the level contours of $G_r(x_1,x_2)$ when $\mu(x_1,x_2) \leq 1/2$ and with those of $G_r(x_1^S,x_2^S)$ when $\mu(x_1,x_2) \geq 1/2$ (where $x_1^S = 1-x_1$ and $x_2^S = 1-x_2$, see Appendix A8). In addition, one has that $G^*(x_1,x_2) = G^*(x_1^S,x_2^S)$. Lastly, the level contours $G^*_P(x_1,x_2) = c$ (where $c \in [0,1]$) are curves that (i) are symmetrical with respect to the $x_2 = 1-x_1$ axis for all $c \in [0,1]$ (i.e., $G^*_P(x_1,x_2) = G^*_P(x_1^S,x_2^S)$), and (ii) they connect the points $(0,0)$ and $(1,1)$ when $c \leq 1/2$.

As can be inferred from panel C, all the distributions (x_1,x_2) lying at the border of the unit square maximise inequality (i.e., they are MIDs) when the latter is measured with $G^*(x_1,x_2)$. By contrast, panel D shows that, when the population principle is imposed, only the bipolar distributions, namely $(0,1)$ and $(1,0)$, maximise inequality. The relative Gini index shown in panel B (G_r) is the only measure in Figure 1 that fails to be strongly consistent. The absolute (G_a) and normalised Gini indices (G^*, G^*_P) not only satisfy the strong consistency axiom, but also its more stringent version – perfect complementarity – because their values coincide when evaluated either at an attainment distribution (x_1,x_2) or its shortfall counterpart (x_1^S,x_2^S). Remarkably, this happens for all values of $n \geq 2$ and for the two normalised inequality measures explored in this paper: the normalised Gini index and the normalised standard deviation (see equations 2 and 3). The variegated shapes of the iso-inequality contours when moving from one inequality measure to another (see Figure 1) explain the discrepancies that might exist among them.

8 Concluding remarks

Bounded variables are fundamentally different from unbounded variables as the former cannot increase or decrease infinitely. Consequently, whenever the mean of a distribution moves closer to any of its bounds, the level of inequality assessed by several traditional inequality measures may fall simply because there is not enough room for variation. The concept of maximum feasible inequality with bounded variables is also quite different from maximum feasible inequality in the context of unbounded variables.[17] We propose a new approach to assessing inequality for bounded variables relying on a new property called the *maximality principle*, which demands that the distributions of bounded variables reflecting maximum feasible inequality be ranked equally. We propose two new classes of inequality indices. The maximality principle leads to a type of normalisation, where each inequality measure in our proposed classes compares observed inequality levels against the maximum inequality level achievable with the same measure across all hypothetical distributions having the same mean.[18]

Bounded variables are represented in terms of either attainments or short-falls, and a consistent evaluation of inequality requires that the ranking of distributions are not reversed while switching between attainment and short-fall representations. Thus, we further characterise a subclass within each of the two classes of normalised inequality indices that allows for consistent evaluation of inequality across attainment and short-fall representations. Several solutions have been proposed to address the consistency challenge for bounded variables, including reliance on absolute Lorenz curves and absolute indices (Lambert and Zheng, 2011), generalised means of indices evaluated at both attainment and shortfall distributions (Lasso de la Vega and Aristondo, 2012) and a relaxation of the (strong) consistency requirement, partially (Bosmans, 2016) or completely (Kenny, 2004). Our solution represents an alternative to all the aforementioned.

One alternative to the class of normalised inequality indices may suggest taking the natural logarithm of the bounded variable, thereby eliminating the lower bound when its value is zero. How-

[17] In the unbounded setting (i.e., in the context of income inequality), Milanovic et al. (2011) suggest and discuss the notion of 'Inequality Possibility Frontier' (IPF), which measures the maximum level of inequality that is potentially attainable for a given level of average income. In that setting, such IPF is reached whenever a vast majority of lower-class individuals survive at substistence levels and a small élite accumulates the reminder of the total income.

[18] Or same mean and population size when the population principle is not upheld.

ever, this approach is plagued with serious problems. To begin with, it would not work when an untransformed value is equal to zero. Meanwhile, if both bounds were positive, taking logarithms for all values would be feasible, but would not solve the boundary effect (because the two bounds would just be replaced by new bounds). Moreover, if the upper bound were higher than 1 then logarithmic transformations would compress the dispersion of values between 1 and the upper bound while expanding it for values between 0 and 1, with the concomitant ambiguous effect on inequality rankings. Worse still, some inequality measures like the variance applied to logarithms of a variable are known to violate the popular transfer principle (Foster and Ok, 1999). Finally, inequality comparisons based on logarithmic transformations of bounded variables would be generally inconsistent. In a nutshell, the costs and inconveniences of the logarithmic transformation render it an unappealing alternative.

We propose a new conceptualisation of assessing inequality with bounded variables and simultaneously propose a solution to address inconsistency. Future research could explore partial orderings respecting the properties that were combined to generate the normalised inequality indices. Furthermore, there remain other measurement challenges in the context of bounded variables. For example, Lasso de la Vega and Aristondo (2012) provide conditions whose fulfillment guarantees robustness of inequality comparisons to changes in the upper bound. Though admittedly this problem is not that serious when bounds are neither arbitrary nor expected to change across time and space (e.g. in the case of indicators expressed as percentage ratios), it is nonetheless worth exploring how our proposed measurement framework could accommodate such potential concern.

Appendices

Appendix A1 Proof of Proposition 1

Let us start with an $\mathbf{x} \in \mathscr{X}_n \setminus \mathscr{M}$ (i.e., $\mathbf{x}$ is neither bipolar nor almost bipolar) for some $n \in \mathbb{N} \setminus \{1\}$ such that $\mu(\mathbf{x}) \in (0, U)$. Given that in the proposition's partial order $(\mathscr{X}_n^{\mu(\mathbf{x})}, \succeq_n)$, a regressive transfer increases inequality while a permutation keeps it unaltered and both keep the mean unaltered, we may always perform a sequence of regressive transfers (with or without additional permutations) until exhaustion to obtain any element of $\mathscr{M}$ that belongs in the set of distributions with the same population size and the same mean, namely $\mathscr{X}_n^{\mu(\mathbf{x})}$.[19]

Now, there can be two types of cases: (i) $\mu(\mathbf{x}) \in \mathbb{G}_n$ and (ii) $\mu(\mathbf{x}) \notin \mathbb{G}_n$, where $\mathbb{G}_n = \{U/n, \ldots, (n-1)U/n\}$ is the set of $n-1$ equally-spaced grid points between 0 and U.

Case (i): Whenever $\mu(\mathbf{x}) \in \mathbb{G}_n$, then there exists a natural number $n' \leq n$ such that $\mu(\mathbf{x}) = n'U/n$. Clearly, $\mu(\mathbf{x}) = n' \times U/n + (n-n') \times 0/n$. Starting with $\mathbf{x}$, it is possible to have a series of regressive transfers until a distribution with n' elements equalling U and $n-n'$ elements equalling zero is reached. In this case, the set of MIDs is $\mathscr{X}_n^{\mu(\mathbf{x})} \cap \mathscr{B}$.

[19]Note that each element within $\mathbf{x}$ is bounded between 0 and U by definition and so it is not possible to perform further regressive transfers once the bounds are reached. The proof proceeds in similar line of argument as the proof of Theorem 1 in Seth and McGillivray (2018).

Case (ii): Whenever $\mu(\mathbf{x}) \notin \mathbb{G}_n$, then there exist a natural number $n' \leq n$ such that $n'U/n < \mu(\mathbf{x}) < (n'+1)U/n$. In this case, a series of regressive transfers are possible until n' elements are equal to U and $n-n'-1$ elements are equal to zero. Note that it is not possible for $n'+1$ elements to be equal to U because $\mu(\mathbf{x}) < (n'+1)U/n$. However, when n' elements are equal to U, then $\mu(\mathbf{x}) > n'U/n$ and therefore the remaining element will have a value of $\varepsilon = n\mu(\mathbf{x}) - n'U$ so that: $\mu(\mathbf{x}) = n' \times U/n + (n-n'-1) \times 0/n + \varepsilon/n$. It is straightforward to verify that $\varepsilon \in (0,U)$. In this case, the set of MIDs is $\mathscr{X}_n^{\mu(\mathbf{x})} \cap \mathscr{A}$.

Thus, the maximum inequality distribution (MID) for $\mathbf{x}$ is an element in the set $\mathscr{X}_n^{\mu(\mathbf{x})} \cap (\mathscr{A} \cup \mathscr{B}) = \mathscr{X}_n^{\mu(\mathbf{x})} \cap \mathscr{M}$, which by our definition is equal to $\mathscr{M}_n^{\mu(\mathbf{x})}$. Now, whenever $\mathbf{x} \in \mathscr{X}_n \cap \mathscr{M}$ for some $n \in \mathbb{N} \setminus \{1\}$ (i.e., $\mathbf{x}$ is either bipolar or almost bipolar), it can be trivially checked that $\mathbf{x} \in \mathscr{M}_n^{\mu(\mathbf{x})}$. Hence, a set of MIDs for any $\mathbf{x} \in \mathscr{X}_n$ such that $\mu(\mathbf{x}) \in (0,U)$ always exists and constitutes the set of maximal elements $\mathscr{M}_n^{\mu(\mathbf{x})}$ of the partially ordered set $(\mathscr{X}_n^{\mu(\mathbf{x})}, \succeq_n)$. ■

Appendix A2 Proof of theorem 1

We first prove the *sufficiency* part. Consider some $\mathbf{x} \in \mathscr{X}_n$ for some $n \in \mathbb{N} \setminus \{1\}$ and so the set of corresponding MIDs is $\mathscr{M}_n^{\mu(\mathbf{x})}$ by proposition 1. We then already know that

$$I(\mathbf{x}) = \begin{cases} M\left(\dfrac{f(\mathbf{x}) - f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}\right) & \text{if } \mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\} \\ 0 & \text{if } \mathbf{x} \in \{\mathbf{0}, \mathbf{U}\} \end{cases}, \tag{A1}$$

where $\bar{\mathbf{x}} = \mu(\mathbf{x})\mathbf{1}_n$, $\hat{\mathbf{x}} \in \mathscr{M}_n^{\mu(\mathbf{x})}$, $f : \mathscr{X}_n \to \mathbb{R}_{++}$ is a symmetric and strictly S-convex function.

We now show that I satisfies the required properties. (i) Consider any $\mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\}$. Since $f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}}) > 0$ because any $\hat{\mathbf{x}} \in \mathscr{M}_n^{\mu(\mathbf{x})}$ can be obtained from $\bar{\mathbf{x}}$ by a series of regressive transfers and f is strictly S-convex, it follows directly from the formulation in Equation A1 that I satisfies the *equality principle* as $I(\bar{\mathbf{x}}) = 0$, and the *maximality principle* as $I(\hat{\mathbf{x}}) = I(\hat{\mathbf{y}}) = M$ for any $\hat{\mathbf{x}} \in \mathscr{M}_n^{\mu(\mathbf{x})}, \hat{\mathbf{y}} \in \mathscr{M}_n^{\mu(\mathbf{y})}$. Whenever $\mathbf{x} \in \{\mathbf{0}, \mathbf{U}\}, I(\mathbf{x}) = 0$, so the *equality principle* is also satisfied. (ii) Suppose that $\mathbf{y} \in \mathscr{X}_n$ is obtained from $\mathbf{x}$ such that $\mathbf{y} = \mathbf{x}\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix. By definition, $\mu(\mathbf{x}) = \mu(\mathbf{y})$, $\mathscr{M}_n^{\mu(\mathbf{y})} = \mathscr{M}_n^{\mu(\mathbf{x})}$ and $\bar{\mathbf{y}} = \bar{\mathbf{x}}$. Provided f is symmetric, $f(\mathbf{y}) = f(\mathbf{x})$ and also $f(\hat{\mathbf{y}}) = f(\hat{\mathbf{x}})$ for any $\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \mathscr{M}_n^{\mu(\mathbf{x})}$. So, $I(\mathbf{y}) = I(\mathbf{x})$. Thus I satisfies *anonymity*. (iii) Suppose $\mathbf{y}' \in \mathscr{X}_n$ is obtained from $\mathbf{x}$ by a regressive transfer. Again, by definition, $\mu(\mathbf{x}) = \mu(\mathbf{y}')$, $\mathscr{M}_n^{\mu(\mathbf{y}')} = \mathscr{M}_n^{\mu(\mathbf{x})}$ and $\bar{\mathbf{y}}' = \bar{\mathbf{x}}$. Provided f is strictly S-convex, $f(\mathbf{y}') > f(\mathbf{x})$ and so $I(\mathbf{y}') > I(\mathbf{x})$. Therefore, I satisfies the *transfer principle*.

Let us now prove the *Necessity* part. Suppose I satisfies anonymity and the transfer principle. Then, I is symmetric and S-convex. Given that a monotonically increasing transformation of an S-convex function is also S-convex, we may write (without loss of generality) $I(\mathbf{x}) = af(\mathbf{x}) + b$ for some $\mathbf{x} \in \mathscr{X}_n$, where $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}$ and f is some S-convex function. The equality principle requires that $I(\bar{\mathbf{x}}) = 0$, therefore $af(\bar{\mathbf{x}}) + b = 0$ or

$$b = -af(\bar{\mathbf{x}}). \tag{A2}$$

We now proceed in three steps to prove that $I(\mathbf{x})$ is a bounded function.

Step 1: Let $\mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\}$. We prove that $\mathscr{X}_n^{\mu(\mathbf{x})}$ is a bounded and closed set within $\mathscr{X}_n$. To do so, we need to define a topology and a distance function in $\mathscr{X}_n$.[20] We can simply use the subspace topology of $\mathbb{R}^n$ for $\mathscr{X}_n$. Thus, if $B_r(\mathbf{x})$ is the standard open $n-$ball in $\mathbb{R}^n$ with center $\mathbf{x}$ and radius $r > 0$, then we define the open $n-$balls in $\mathscr{X}_n$ as $\tilde{B}_r(\mathbf{x}) := B_r(\mathbf{x}) \cap \mathscr{X}_n$. As per distance function, for any $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$ we can use $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2 / n}$, which is a variant of the standard Euclidean distance that normalizes by the number of observations (n), and complies with the standard axioms of distance functions (see footnote 20).

The two points that are furthest apart within $\mathscr{X}_n$ are $\mathbf{0} = (0, \cdots, 0)$ and $\mathbf{U} = (U, \cdots, U)$. Their distance equals $d(\mathbf{0}, \mathbf{U}) = U$. Thus, the distance between any other two points in $\mathscr{X}_n$ must be smaller than U. Hence, the open ball $\tilde{B}_U(\mathbf{x})$ contains the entire set $\mathscr{X}_n$ (i.e., $\tilde{B}_U(\mathbf{x}) \supset \mathscr{X}_n$), so $\mathscr{X}_n$ is a bounded set. Since $\mathscr{X}_n^{\mu(\mathbf{x})} \subset \mathscr{X}_n$, $\mathscr{X}_n^{\mu(\mathbf{x})}$ is also bounded.

We now prove that $\mathscr{X}_n^{\mu(\mathbf{x})}$ is closed. To do that, we prove that $\mathscr{X}_n \setminus \mathscr{X}_n^{\mu(\mathbf{x})}$ is open. Let $\mathbf{y} \in \mathscr{X}_n \setminus \mathscr{X}_n^{\mu(\mathbf{x})}$ (that is: $\mu(\mathbf{y}) \neq \mu(\mathbf{x})$), and define $\varepsilon := |\mu(\mathbf{y}) - \mu(\mathbf{x})|/2$. Then, $\tilde{B}_\varepsilon(\mathbf{y})) \subset \mathscr{X}_n \setminus \mathscr{X}_n^{\mu(\mathbf{x})}$, so we have defined an open $n-$ball centered in $\mathbf{y}$ that is completely included within $\mathscr{X}_n \setminus \mathscr{X}_n^{\mu(\mathbf{x})}$. Thus, $\mathscr{X}_n \setminus \mathscr{X}_n^{\mu(\mathbf{x})}$ is open, so $\mathscr{X}_n^{\mu(\mathbf{x})}$ is closed.

Hence, we have proved that $\mathscr{X}_n^{\mu(\mathbf{x})}$ is a **compact** set within $\mathscr{X}_n$.

Step 2: Since $I : \mathscr{X}_n \to \mathbb{R}_+$ is a continuous function, and $\mathscr{X}_n^{\mu(\mathbf{x})}$ is a compact set, then $I(\mathscr{X}_n^{\mu(\mathbf{x})})$ (i.e., the image of $\mathscr{X}_n^{\mu(\mathbf{x})}$ by I) is a compact set. Compact sets within $\mathbb{R}_+$ are bounded, so this implies there exist $m_\mathbf{x}, M_\mathbf{x} \in \mathbb{R}_+$ such that $m_\mathbf{x} \leq I(\mathbf{x}) \leq M_\mathbf{x}$ for all $\mathbf{x} \in \mathscr{X}_n^{\mu(\mathbf{x})}$.

Step 3: Imposing the Equality Principle implies $m_\mathbf{x} = 0$ for all $\mathbf{x} \in \mathscr{X}_n$. In addition, $M_\mathbf{x} = I(\hat{\mathbf{x}})$ for all $\mathbf{x} \in \mathscr{X}_n$. Applying the Maximality Principle, it turns out that $M_\mathbf{x} = I(\hat{\mathbf{x}}) = I(\hat{\mathbf{y}}) = M_\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$. Now, define $M := M_\mathbf{x}$. Thus, for all $\mathbf{x} \in \mathscr{X}_n, 0 \leq I(\mathbf{x}) \leq M$.

Now, since $I(\hat{\mathbf{x}}) = M$ and $b = -af(\bar{\mathbf{x}})$ (from equation A2), we have that $af(\hat{\mathbf{x}}) + b = af(\hat{\mathbf{x}}) - af(\bar{\mathbf{x}}) = M$. Thus, $a = M/(f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}}))$, so

$$I(\mathbf{x}) = M \left(\frac{f(\mathbf{x}) - f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})} \right) \tag{A3}$$

whenever $\mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\}$. For the cases $\mathbf{x} = \mathbf{0}, \mathbf{x} = \mathbf{U}$, the Equality Principle states that $I(\mathbf{x}) = 0$. This completes the proof. ■

Appendix A3 Proof of proposition 2

Consider an $\mathbf{x} \in \mathscr{X}_n \subset \mathscr{X}$ for some $n \in \mathbb{N} \setminus \{1\}$ such that $\mu(\mathbf{x}) \in (0, U)$.[21] Recall that $U \in \mathbb{Q}_{++}$. Using decimal notation, we can write $\mu(\mathbf{x}) = a_1 \cdots a_h.b_1 b_2 \cdots b_k \cdots$, where $a_1 \cdots a_h$ is the integer

[20] Recall that a *distance function* in a set S is defined as a function $d : S \times S \to \mathbb{R}_+$ satisfying the following axioms for all points $\mathbf{x}, \mathbf{y}, \mathbf{z} \in S$: (i) (Identity) $d(\mathbf{x}, \mathbf{x}) = 0 \ \forall \mathbf{x} \in S$; (ii) (Positivity) If $\mathbf{x} \neq \mathbf{y}$, then $d(\mathbf{x}, \mathbf{y}) > 0$; (iii) (Symmetry) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$; (iv) (Triangle Inequality) $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

[21] Note that the cases $\mu(\mathbf{x}) \in \{0, U\}$ are purposefully dismissed from proposition 2.

part of $\mu(\mathbf{x})$ and $b_1b_2\cdots b_k\cdots$ is the corresponding decimal part (i.e., $a_i, b_i \in \{0,1,2,\ldots,8,9\}$ are the digits of $\mu(\mathbf{x})$). While the representation of the integer part has finite length (i.e., $h<\infty$), the decimal part can be either finite or infinite. Now, there are two possibilities: either $\mu(\mathbf{x}) \in \mathbb{Q}$ or $\mu(\mathbf{x}) \in \mathbb{R}\setminus\mathbb{Q}$.

Case (i): First, suppose $\mu(\mathbf{x}) \in \mathbb{Q}$. Then, we can write $\mu(\mathbf{x}) = p/q$ for some $p,q \in \mathbb{N}$. In that case, there must exist a multiple of n, let it be m, (i.e., $m = \alpha n$, with $\alpha \in \mathbb{N}\setminus\{1\}$) such that $\mu(\mathbf{x}) \in \mathbb{G}_m$. Following Proposition 1, the MID distribution for $\mathbf{x}$ (i.e., $\hat{\mathbf{x}}$) exists, and is a bipolar one, where a share of the population (equalling $1-p/qU$, a rational number) attains the value of 0, and the remaning share (of size p/qU, another rational number) attains the value of U. By construction, the mean of such distribution equals p/q.

Case (ii): Now, let $\mu(\mathbf{x}) \in \mathbb{R}\setminus\mathbb{Q}$. Because of the density of rational numbers within real numbers, it is possible to construct a sequence $s_i \in \mathbb{Q}$ converging towards $\mu(\mathbf{x})$ (i.e., such that $\lim_{i\to\infty} s_i = \mu(\mathbf{x})$). One such simple sequence could be defined as $s_i = a_1\cdots a_h b_1 b_2\cdots b_i/(10^i)$. Following Case (i), for each s_i there must exist a multiple of n, call it m_i (i.e., $m_i = \alpha_i n$, with $\alpha_i \in \mathbb{N}\setminus\{1\}$) such that $s_i \in \mathbb{G}_{m_i}$. Now, let $\mathbf{b}_i$ be as a bipolar distribution, where a share of the population (equalling $1-s_i/U$, a rational number) attains the value of 0, and the remaning share (of size s_i/U, another rational number) attains the value of U. By construction, $\mu(\mathbf{b}_i) = s_i$. Finally, we define the MID distribution for $\mathbf{x}$ as $\hat{\mathbf{x}} := lim_{i\to\infty}\mathbf{b}_i$. Since each $\mathbf{b}_i$ is a bipolar distribution, so it is $\hat{\mathbf{x}}$.

In both cases, we have proved that, when imposing the population principle, the MIDs for $\mathbf{x}$ exist and consist of bipolar distributions. ■

Appendix A4 Proof of theorem 2

We first prove the sufficiency part. Applying theorem 1, which holds for $\mathscr{X}_n$, we can show that I satisfies *anonymity*, the *transfer principle* and the *equality principle*. Now, since $f(\hat{\mathbf{x}}) > f(\bar{\mathbf{x}})$ for any $\hat{\mathbf{x}} \in \mathscr{B}^{\mu(\mathbf{x})}$ because any $\hat{\mathbf{x}} \in \mathscr{B}^{\mu(\mathbf{x})}$ can be obtained from $\bar{\mathbf{x}}$ by a series of regressive transfers with or without combinations of replications and permutations, and f is strictly S-convex, we have$I(\hat{\mathbf{x}}) = M$; that is, I satisfies the *restricted maximality principle*. Finally, we prove that I satisfies the population principle. Let $\mathbf{y}$ be obtained from $\mathbf{x} \in \mathscr{X}_n$ through a replication. Then, by definition, $\mu(\mathbf{x}) = \mu(\mathbf{y})$ and so by proposition 2, $\mathscr{B}^{\mu(\mathbf{x})} = \mathscr{B}^{\mu(\mathbf{y})}$. It is also straightforward to verify that $\bar{\mathbf{y}}$ is a replication of $\bar{\mathbf{x}}$. Therefore, based on $(\mathscr{X}^{\mu(\mathbf{x})}, \succeq)$, $\mathbf{y} \sim \mathbf{x}$ and $\bar{\mathbf{y}} \sim \bar{\mathbf{x}}$, and hence $f(\bar{\mathbf{y}}) = f(\bar{\mathbf{x}})$ and $f(\mathbf{y}) = f(\mathbf{x})$ because f satisfies the population principle. Coupled with $f(\hat{\mathbf{y}}) = f(\hat{\mathbf{x}})$ for any $\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \mathscr{B}^{\mu(\mathbf{x})}$ and $f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}}) > 0$, clearly $I(\mathbf{y}) = I(\mathbf{x})$. Hence, I satisfies the population principle.

The proof of the necessity part is similar to the proof of theorem 1's necessity part, with some additional modifications. If I satisfies anonymity and the transfer principle, then, without loss of generality, $I(\mathbf{x}) = af(\mathbf{x}) + b$ for some $\mathbf{x} \in \mathscr{X}_n$, where $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}$ and f is some S-convex function. Suppose that $\mathbf{y} \in \mathscr{X}_{n'}$ is obtained from $\mathbf{x}$ by *replication* for some $n' = \alpha n$, where $\alpha \in \mathbb{N}\setminus\{1\}$. Given that I satisfies the population principle, then $I(\mathbf{y}) = I(\mathbf{x})$. It follows that $f(\mathbf{y}) = f(\mathbf{x})$ since $a > 0$ and so f also satisfies the population principle. By the equality principle, which requires that $I(\bar{\mathbf{x}}) = 0$ where $\bar{\mathbf{x}} = \mu(\mathbf{x})\mathbf{1}_n$, we obtain $af(\bar{\mathbf{x}}) + b = 0$ or $b = -af(\bar{\mathbf{x}})$.

Mimicking the proof of theorem 1, we now show that $I(\mathbf{x})$ is a bounded function. Since $\mathscr{X}_n^{\mu(\mathbf{x})}$ is compact for all $n \in \mathbb{N}\setminus\{1\}$ (see theorem 1) and $\mathscr{X}^{\mu(\mathbf{x})} = \cup_n \mathscr{X}_n^{\mu(\mathbf{x})}$ is a countable union of compact sets, then $\mathscr{X}^{\mu(\mathbf{x})}$ is also compact. Since I is a continuous function, then $I(\mathscr{X}^{\mu(\mathbf{x})})$ is bounded. Thus, there exist $m_\mathbf{x}, M_\mathbf{x} \in \mathbb{R}_+$ such that $m_\mathbf{x} \leq I(\mathbf{x}) \leq M_\mathbf{x}$ for all $\mathbf{x} \in \mathscr{X}^{\mu(\mathbf{x})}$. By the Equality Principle, $m_\mathbf{x} = 0$ for all $\mathbf{x} \in \mathscr{X}$. By the Restricted Maximality Principle, $M_\mathbf{x} = I(\hat{\mathbf{x}}) = I(\hat{\mathbf{y}}) = M_\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathscr{X}$. Now, define $M := M_\mathbf{x}$. Thus, for all $\mathbf{x} \in \mathscr{X}, 0 \leq I(\mathbf{x}) \leq M$, so I is bounded. Now, since $I(\hat{\mathbf{x}}) = M$ and $b = -af(\bar{\mathbf{x}})$, we have that $af(\hat{\mathbf{x}}) + b = af(\hat{\mathbf{x}}) - af(\bar{\mathbf{x}}) = M$. Thus, $a = M/(f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}}))$, so we obtained the desired functional form whenever $\mathbf{x} \in \mathscr{X}_n \setminus \{\mathbf{0}, \mathbf{U}\}$. For the cases $\mathbf{x} = \mathbf{0}, \mathbf{x} = \mathbf{U}$, the Equality Principle states that $I(\mathbf{x}) = 0$. This completes the proof. ■

Appendix A5 Proof of proposition 3

The 'if' part is straightforward. For the 'only if' part consider the definition of strong consistency: $I(\mathbf{x}) \leq I(\mathbf{y}) \Leftrightarrow I(\mathbf{x}^S) \leq I(\mathbf{y}^S)$ for any $\mathbf{x}, \mathbf{y} \in \mathscr{X}_n$. Now let $\mathbf{y} = \mathbf{x}^S$. Then we get: $I(\mathbf{x}) \leq I(\mathbf{x}^S) \Leftrightarrow I(\mathbf{x}^S) \leq I(\mathbf{x})$ which can only hold if $I(\mathbf{x}) = I(\mathbf{x}^S)$. ■

Appendix A6 Proof of theorem 3

The sufficiecy part is straightforward. If f satisfies the property mentioned in theorem 3, we get $I(\mathbf{x}^\mathbf{S}) = I(\mathbf{x})$ because p and q cancel out, and $I(\mathbf{x}^\mathbf{S}) = I(\mathbf{x})$ implies strong consistency.

Let us now prove the necessary part. From proposition 3, we conclude that if all the indices I from theorems 1 and 2 are strongly consistent, then it must be true that for each of them: $I(\mathbf{x}^\mathbf{S}) = I(\mathbf{x})$, which in turn means:

$$\frac{f(\mathbf{x}^S) - f(\bar{\mathbf{x}}^S)}{f(\hat{\mathbf{x}}^S) - f(\bar{\mathbf{x}}^S)} = \frac{f(\mathbf{x}) - f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})} \tag{A4}$$

Then, solving equation A4 for $f(\mathbf{x}^S)$ yields:

$$f(\mathbf{x}^S) = \frac{f(\hat{\mathbf{x}}^S) - f(\bar{\mathbf{x}}^S)}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})} f(\mathbf{x}) + \frac{f(\hat{\mathbf{x}}) f(\bar{\mathbf{x}}^S) - f(\hat{\mathbf{x}}^S) f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}, \tag{A5}$$

Finally, recalling that $\mathbf{x}^S = (U, U, ..., U) - \mathbf{x}$ and every element in $\bar{\mathbf{x}}$ is equal to $\mu(\mathbf{x})$, we obtain:

$$p(\mathbf{x}) = \frac{f(\hat{\mathbf{x}}^S) - f(\bar{\mathbf{x}}^S)}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})} \tag{A6}$$

$$q(\mathbf{x}) = \frac{f(\hat{\mathbf{x}}) f(\bar{\mathbf{x}}^S) - f(\hat{\mathbf{x}}^S) f(\bar{\mathbf{x}})}{f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}})}. \tag{A7}$$

This completes our proof. ■

Appendix A7 Derivation of normalised inequality indices

Here we show the derivation of the inequality measures presented as examples in section 4. We start with the formulas relying on bipolar MIDs followed by formulas based on almost bipolar MIDs. The former are simpler than the latter.

In a bipolar MID $\hat{\mathbf{x}}$, assume that a share s of the population attains the value of 0 and the rest $(1-s)$ the value of U. Given that $\mu(\hat{\mathbf{x}}) = \mu(\mathbf{x})$, by definition, the following restriction must hold:

$$\mu(\hat{\mathbf{x}}) = s \times 0 + (1-s) \times U \Rightarrow (1-s)U = \mu(\mathbf{x}) \Rightarrow s = 1 - \frac{\mu(\mathbf{x})}{U}. \tag{A8}$$

The absolute Gini index:

Computing the absolute Gini index for $\hat{\mathbf{x}}$ yields:

$$G_a(\hat{\mathbf{x}}) = s(1-s)U. \tag{A9}$$

Plugging equation A8 into equation A9 and manipulating algebraically yields:

$$G_a(\hat{\mathbf{x}}) = \frac{\mu(\mathbf{x})}{U}\left(1 - \frac{\mu(\mathbf{x})}{U}\right)U = \frac{\mu(\mathbf{x})[U-\mu(\mathbf{x})]}{U}.$$

The relative Gini index:

Computing the relative Gini index for $\hat{\mathbf{x}}$ yields:

$$G_r(\hat{\mathbf{x}}) = \frac{s(1-s)U}{\mu(\mathbf{x})}. \tag{A10}$$

Plugging equation A8 into equation A10 and manipulating algebraically yields:

$$G_r(\hat{\mathbf{x}}) = \frac{U-\mu(\mathbf{x})}{U}.$$

The standard deviation:

Computing the standard deviation for $\hat{\mathbf{x}}$ yields:

$$\sigma(\hat{\mathbf{x}}) = \sqrt{s[\mu(\mathbf{x})]^2 + (1-s)[U-\mu(\mathbf{x})]^2}. \tag{A11}$$

Plugging equation A8 into equation A11 and manipulating algebraically, we obtain:

$$\sigma(\hat{\mathbf{x}}) = \sqrt{\left[1 - \frac{\mu(\mathbf{x})}{U}\right][\mu(\mathbf{x})]^2 + \frac{\mu(\mathbf{x})}{U}[U-\mu(\mathbf{x})]^2} = \sqrt{\mu(\mathbf{x})\,[U-\mu(\mathbf{x})]}.$$

The coefficient of variation:

Computing the coefficient of variation for $\hat{\mathbf{x}}$ yields:

$$CV(\hat{\mathbf{x}}) = \frac{\sigma(\hat{\mathbf{x}})}{\mu(\hat{\mathbf{x}})} = \frac{\sqrt{\mu(\mathbf{x})\,[U-\mu(\mathbf{x})]}}{\mu(\mathbf{x})} = \sqrt{\frac{U-\mu(\mathbf{x})}{\mu(\mathbf{x})}}. \tag{A12}$$

Derivation of normalised inequality indices for almost bipolar MIDs

In an almost bipolar MID $\hat{\mathbf{x}}$ we have n' units in the population with value U, one unit with value $0<\varepsilon<U$ and the rest, $n-n'-1$ with value 0. Moreover, $\varepsilon = n\mu(\mathbf{x}) - n'U$. For each of the denominators of the indices mentioned in section 4 we get:

<u>The absolute Gini index</u>:

$$G_a(\hat{\mathbf{x}}) = \frac{1}{2n^2}[(n-n'-1)\times 1\times|0-\varepsilon| + (n')\times 1\times|U-\varepsilon| + n'(n-n'-1)\times 1\times|0-U|] \tag{A13}$$

Simplifying equation A13 we get the denominator of 2 for the almost bipolar case (noting later that the 2 in the fraction gets cancelled out as it also appears in the numerator's formula):

$$G_a(\hat{\mathbf{x}}) = \frac{1}{2n^2}[(n-n'-1)\varepsilon + (n')(U-\varepsilon) + n'(n-n'-1)U]. \tag{A14}$$

<u>The relative Gini index</u>:

Essentially we get the same formula for the denominator of 2 as in A13 but divided by $\mu(\mathbf{x})$ (again, the 2 in the fraction gets cancelled out as it also appears in the numerator's formula):

$$G_r(\hat{\mathbf{x}}) = \frac{1}{2n^2\mu(\mathbf{x})}[(n-n'-1)\varepsilon + (n')(U-\varepsilon) + n'(n-n'-1)U]. \tag{A15}$$

<u>The standard deviation</u>:

$$\sigma(\hat{\mathbf{x}}) = \sqrt{\frac{1}{n}[(n-n'-1)(0-\mu(\mathbf{x}))^2 + n'(U-\mu(\mathbf{x}))^2 + (\varepsilon-\mu(\mathbf{x}))^2]}. \tag{A16}$$

Simplifying equation A16 we get the denominator of 3 for the almost bipolar case.

<u>The coefficient of variation</u>:

We get the same formula as in A16 but divided by $\mu(\mathbf{x})$:

$$CV(\hat{\mathbf{x}}) = \frac{1}{\mu(\mathbf{x})}\sqrt{\frac{1}{n}[(n-n'-1)\mu(\mathbf{x})^2 + n'(U-\mu(\mathbf{x}))^2 + (\varepsilon-\mu(\mathbf{x}))^2]}. \tag{A17}$$

Finally, for each of the aforementioned indices (for bipolar and almost bipolar MIDs), we compute $f(\mathbf{x})/f(\hat{\mathbf{x}})$.

Appendix A8 Derivation of iso-inequality contours

When $n = 2$ and $U = 1$, the absolute Gini index can be written as $G_a(x_1,x_2) = |x_1 - x_2|/4$, where $(x_1,x_2) \in [0,1]^2$. In that case, the iso-inequality contours are straight lines parallel to the 45° line. In this setting, the relative Gini index can be written as $G_r(x_1,x_2) = |x_1 - x_2|/4\mu = |x_1 - x_2|/2(x_1 + x_2)$. Since this function is homogeneous of degree 0 (i.e., $G_r(\lambda x_1, \lambda x_2) = G_r(x_1,x_2)$ for all $\lambda > 0$), the iso-inequality contours are straight lines emanating from (or converging to) the origin $(0,0)$.

What about the iso-inequality contours for the normalised Gini index that does not comply with the population principle ($G^*(x_1,x_2)$)? Here, $\mathbb{G}_2 = \{1/2\}$, so there are basically two cases: either $\mu(x_1,x_2) \leq 1/2$ or $\mu(x_1,x_2) \geq 1/2$. Case (i): $\mu(x_1,x_2) = (x_1 + x_2)/2 \leq 1/2$. Here, the MIDs associated with (x_1,x_2) are $\{(0,x_1 + x_2),(x_1 + x_2,0)\}$. When the absolute Gini index is applied to any of those distributions, one obtains $G_a(0,x_1 + x_2) = G_a(x_1 + x_2,0) = (x_1 + x_2)/4$. Hence, $G^*(x_1,x_2) = |x_1 - x_2|/(x_1 + x_2)$. These are straight lines emanating from (or converging to) the origin $(0,0)$. Case (ii): $\mu(x_1,x_2) = (x_1 + x_2)/2 \geq 1/2$. Now, the MIDs associated with (x_1,x_2) are $\{(x_1 + x_2 - 1,1),(1,x_1 + x_2 - 1)\}$. Calculating the absolute Gini index of any of those distributions yields $G_a(x_1 + x_2 - 1,1) = G_a(1,x_1 + x_2 - 1) = (2 - (x_1 + x_2))/4$. Hence $G^*(x_1,x_2) = |x_1 - x_2|/(2 - (x_1 + x_2)) = |x_1^S - x_2^S|/(x_1^S + x_2^S)$ (where $x_1^S = 1 - x_1$ and $x_2^S = 1 - x_2$). These are straight lines emanating from (or converging to) the point $(1,1)$. Finally, it is easy to prove that the two sets of iso-inequality contours match at the intersection (i.e., for the set of points $\{(x_1,x_2) \in [0,1]^2 | x_1 + x_2 = 1\}$, the values of the iso-inequality contours examined in cases (i) and (ii) coincide).

According to theorem 2 and equation 2, the normalised Gini index complying with the population principle is simply defined as $G_P^*(x_1,x_2) = G_a(x_1,x_2)/(\mu(x_1,x_2)(1 - \mu(x_1,x_2)))$. Manipulating algebraically, one obtains that $G_P^*(x_1,x_2) = |x_1 - x_2|/((x_1 + x_2)(2 - (x_1 + x_2)))$. This is the function from which the iso-inequality contours shown in Figure 1 panel D are calculated.

References

Aboudi, R., D. Thon, and S. Wallace (2010). Inequality comparisons when the populations differ in size. *Journal of Economic Inequality 8*, 47–70.

Amiel, Y. and F. Cowell (1994). Inequality changes and income growth. In W. Eichorn (Ed.), *Models and Measurement of Welfare and Inequality*. Springer Verlag.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory 2*, 244–63.

Atkinson, A. (2018). *Inequality. What can be done?* Cambridge, Massachusetts: Harvard University Press.

Bosmans, K. (2007). Income inequality, quasi-concavity, and gradualpopulation shifts. *Mathematical Social Sciences 53*, 29–45.

Bosmans, K. (2016). Consistent comparisons of attainment and shortfall inequality: A critical examination. *Health Economics 25*(11), 1425–32.

Bourguignon, F. (2017). *The globalization of inequality*. New Jersey: Princeton University Press.

Chakravarty, S. (2009). *Inequality, polarization and poverty. Advances in distributional analysis*. Springer.

Chakravarty, S., N. Chattopadhyay, and C. D'Ambrosio (2016). On a family of achievement and shortfall inequality indices. *Health Economics 25*(12), 1503–1513.

Clarke, P., U. Gerdtham, M. Johannesson, K. Bingefors, and L. Smith (2002). On the measurement of relative and absolute income-related health inequality. *Social Science and Medicine 55*, 1923–8.

Cowell, F. (2011). *Measuring Inequality* (3rd ed.). Oxford University Press.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal 30*(119), 348–361.

Ebert, U. (2010). The decomposition of inequality reconsidered: Weakly decomposable measures. *Mathematical Social Sciences 60*(2), 94–103.

Erreygers, G. (2009). Can a single indicator measure both attainment and shortfall inequality? *Journal of Health Economics 28*, 885–93.

Fields, G. (1987). Measuring inequality change in an economy with income growth. *Journal of Development Economics 26*, 357–74.

Fields, G. (1993). Inequality in dual economy models. *Economic Journal 103*, 1228–1235.

Fields, G. (1998). Do inequality measures measure inequality? In S. Jenkins, A. Kapteyn, and B. van Praag (Eds.), *The Distribution ofWelfare and Household Production: International Perspectives*. Cambridge University Press.

Foster, J. and E. Ok (1999). Lorenz dominance and the variance of logarithms. *Econometrica 67*(4), 901–907.

Kenny, C. (2004). Why are we worried about income? Nearly everything that matters is converging. *World Development 33*(1), 1–19.

Lambert, P. and B. Zheng (2011). On the consistent measurement of attainment and shortfall inequality. *Journal of Health Economics 30*, 214–9.

Lasso de la Vega, C. and O. Aristondo (2012). Proposing indicators to measure achievement and shortfall inequality consistently. *Journal of Health Economics 31*, 578–583.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory of majorization and its applications*. New York: Academic Press.

Micklewright, J. and K. Stewart (1999). Is the well-being of children converging in the European Union? *Economic Journal 109*(459), 692–714.

Milanovic, B. (2018). *Global inequality. A new approach for the age of globalization*. Cambridge, Massachusetts: Harvard University Press.

Milanovic, B., P. Lindert, and J. Williamson (2011). Pre-industrial inequality. *Economic Journal 121*, 255–272.

Piketty, T. (2015). *The economics of inequality*. Cambridge, Massachusetts: Harvard University Press.

Seth, S. and S. Alkire (2017). Did poverty reduction reach the poorest of the poor? Complementary measures of poverty and inequality in the counting approach. In S. Bandyopadhyay (Ed.), *Research on Economic Inequality*, Volume 25, pp. 63–102. Emerald Publishing Limited.

Seth, S. and M. McGillivray (2018). Composite indices, alternative weights, and comparison robustness. *Social Choice and Welfare 51*, 657–679.

Temkin, L. (1986). Inequality. *Philosophy and Public Affairs 15*, 99–121.