Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018

# PENALTY-BASED AND OTHER REPRESENTATIONS OF ECONOMIC INEQUALITY

GLEB BELIAKOV

*School of Information Technology, Deakin University, 221 Burwood Hwy*
*Burwood, Victoria 3125, Australia*
*gleb@deakin.edu.au*

MAREK GAGOLEWSKI

*Systems Research Institute, Polish Academy of Sciences,*
*Newelska 6, 01-447 Warsaw, Poland*
*Faculty of Mathematics and Information Science, Warsaw University of Technology,*
*ul. Koszykowa 75, 00-662 Warsaw, Poland*
*gagolews@ibspan.waw.pl*

SIMON JAMES*

*School of Information Technology, Deakin University, 221 Burwood Hwy*
*Burwood, Victoria 3125, Australia*
*sjames@deakin.edu.au*

Economic inequality measures are employed as a key component in various socio-demographic indices to capture the disparity between the wealthy and poor. Since their inception, they have also been used as a basis for modelling spread and disparity in other contexts. While recent research has identified that a number of classical inequality and welfare functions can be considered in the framework of OWA operators, here we propose a framework of penalty-based aggregation functions and their associated penalties as measures of inequality.
Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018

*Keywords*: Penalty functions; aggregation functions; inequality indices; spread measures.

## 1. Introduction

Aggregation functions are indispensable in welfare and poverty evaluations where the information pertaining to many individuals needs to be summarized to give an overall snapshot of a community, class or country. Such indices can then be used for the purpose of comparing groups, considering the anticipated benefit of policy decisions at the government level and so on. Simple unweighted averages are often used to measure various countries' standings in terms of poverty, equality, opportunity, etc, for example as conducted by the Economist Intelligence Unit (EIU)[47].

*Corresponding author

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



It is recognized that a standard summary of wages cannot capture the economic situation of a community with individuals having a range of incomes[I], so some indices also incorporate the extent to which the incomes of individuals/households are below the poverty line[29], the difference in incomes, or sometimes even the gap between male and female wages (e.g.[47]).

Economic inequality measures[3,18,20] aim to capture an intuitive notion of fairness in the distribution of wealth. They align with ideas such as the diminishing middle class, the widening gap between rich and poor, or the unfairness inherent in society when only a small fraction of individuals control the majority of its wealth and resources. Recently in Australia there has been some popular interest with articles appearing in *The Conversation*[II] about how Australia's mining boom has affected economic inequality in certain regions[23] (measured using the Gini index) and the validity in a politician's recent claim that Australia is more unequal now than it has been for 75 years[38] according to the income of the top 1% of the population.

As well as aligning with such concepts, inequality measures are also essentially a form of spread. At one end of the scale, when all arguments are the same, we have perfect equality with no variation in the income or resources allocated to each individual, while at the other we expect big differences between the poorest and richest group. Just as calculations like the standard deviation are often used in statistics to aid with interpreting results obtained from an arithmetic mean, it is natural that we would want to have some idea of variation in wages when interpreting an average wage statistic for a given community.

As an example, Fig. 1 compares the male and female wages of working individuals in Australia based on census data[III] for each percentile of their respective populations (with the upper and lower 5% excluded). Although this graph makes it clear that male incomes are higher, on average, than females, what is less apparent is that there is also slightly higher inequality *within* the female gender class, i.e. Fig 2 shows the cumulative percentage of the total income (for each gender) earned by each percentile of the population. These curves are referred to as Lorenz curves, proposed by Lorenz for measuring inequality[39]. If a curve lies below another, then it is considered to represent higher inequality. With this example, we can see

[I]It is worth mentioning the argument that a quantified level of 'poverty' can affect individuals differently, and so some organizations have proposed indices that try to capture this, e.g.[49], which for an individual, again, is obtained through a simple aggregation of *deprivation* factors.

[II]The Conversation is an online journal with collaborations between academics and its editors with articles pitched at a general audience. The article about the mining boom relates to the journal article by the same authors[22].

[III]The data is publically available from the Australian Bureau of Statistics[41], however we note that the incomes here are estimated based on income brackets, e.g. there were 189 996 males listed earning weekly wages in the \$1-\$199 bracket, 109 733 in the \$200-\$299 bracket and so on. We estimated each individual's wage by linearly interpolating the bracket thresholds and then in turn estimated the mean with the income which was $1/\sqrt{2}$ along the bracket interval. We also note that those listed as having negative income were assigned the value of 0 and those earning \$2000 or more were assumed to be earning between \$2000 - \$3000 so values above \$2000 should only be interpreted as rough estimates.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



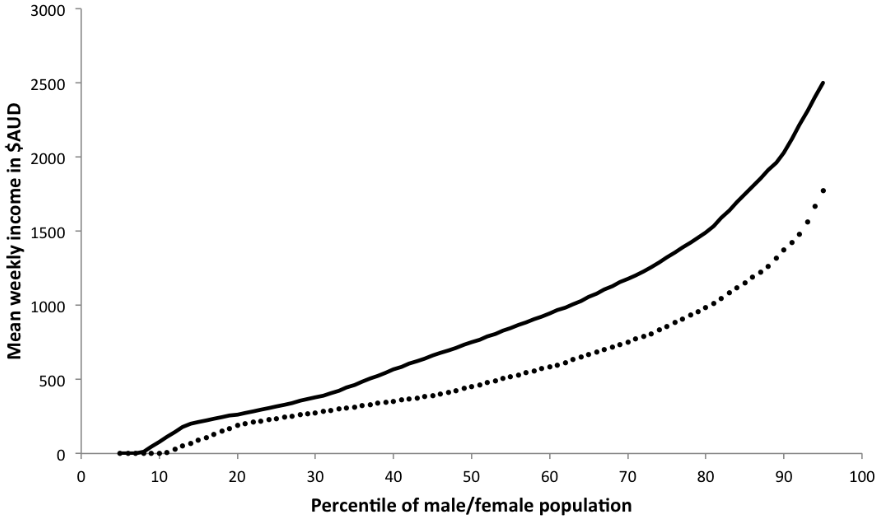


Fig. 1. Mean weekly incomes for males (solid) and females (dotted) based on Australian census data.

that the lower 40% of males earn approximately 11% while for females the figure is 10.5%. On the other hand, if we consider the top 10% of the population (the 90th percentiles), the conservative estimates used here note that for males the value is 72.5% while for females it is 68.8%. These richer males hence account for about 27.5% of all male incomes while for females it is 31.2%.

Economic inequality is hence related to a number of areas studied in association with aggregation functions. Gagolewski has recently proposed the concept of spread measures[25], which incorporates traditional measures of variation as special cases. Parallels have also been drawn between spread and the notion of disagreement (or lack of consensus) in decision making[1,5,6,14,28,31,32] and with evenness indices used in ecology[8,9]. In turn, we can observe that, historically, many traditional evenness measures were derived from inequality indices[43,44], since the desireable property that evenness is increased with transfers of abundance from rare to more common species is one that is satisfied by the Pigou-Dalton principle proposed by Dalton[19]. Some other commonly used measures of evenness are derived from Shannon's entropy equation[42], which was also defined with this 'progressive transfers' principle identified as an attractive property. Notably, inequality measures have many applications in other domains, see e.g. [26,27] for the case of data clustering.

Inequality, spread and disagreement measures all share the common property that a function value of zero is returned when all the inputs are the same. Beliakov, James and Nimmo[8] have suggested previously that a key distinction when considering these frameworks lies in the characteristics of input vectors that return

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



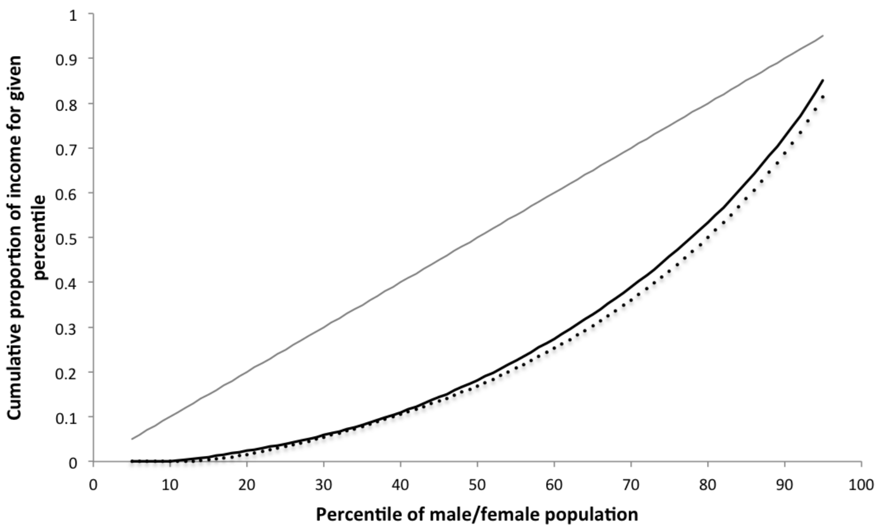


Fig. 2. Lorenz curves - cumulative proportion of income for percentile groups of males (solid) and females (dotted).

the maximum level of variation. When considering disagreement or lack of consensus, the maximum usually corresponds with the case where the inputs are divided equally between opposite ends of the scale, e.g. a vector like $\boldsymbol{x} = (0, 0, 0, 1, 1, 1)$ for $x_i \in [0, 1]$. For inequality on the other hand, maximum inequality usually occurs when a single input dominates, e.g. if the incomes of a population approach $\boldsymbol{x} = (1, 1, 1, \ldots, 1, \infty)$. Knowing which vectors result in maximum spread is also of importance for the purpose of normalizing spread calculations[24].

The concept of a penalty function[12,15,16,51] also captures the idea of variation amongst the inputs. In this paper we will show that penalty-based expressions can be used to represent classical inequality indices such as the Gini index and that penalties more generally can provide useful models for inequality. We establish some conditions under which such models satisfy the Pigou-Dalton principle. We note that penalties have also been used to define averaging aggregation functions, and hence with any penalty expression we often have an associated mean value. The behavior of this implied average could therefore provide a framework for interpreting the inequality measure.

The article will be set out as follows. In Section 2 we will give overviews of the preliminary notions required for further results in the paper. These include those relating to aggregation functions and penalty functions. In Section 3 we will give an overview of economic inequality functions with some established results and notation. In Section 4 we present penalty functions as a framework for inequality indices, showing how some existing indices can be framed in terms of penalty functions. We

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



establish results regarding the conditions under which inequality functions based on penalties will satisfy the Pigou-Dalton principle. In Section 5, we show how inequality measures are related to spread measures. We give some brief notes on directions for further research in Section 6 and then conclude in Section 7.

## 2. Preliminaries

Here we outline preliminary notions regarding aggregation functions and their related penalty-based expressions.

### 2.1. *Aggregation Functions*

Aggregation functions summarize datasets with a single representative value[4,10,35,45]. Socio-economic summaries allow organizations to compare countries or regions based on a number of key indicators, which in turn usually represent standardized values ranging over a consistent interval. We will use the following definition.

**Definition 1.** For a given positive real interval $[0, b]$ and a fixed $n \in \mathbb{N} \setminus \{1\}$, an aggregation function $\mathsf{A} : [0, b]^n \to [0, b]$ is a function non-decreasing in each argument and satisfying $\mathsf{A}(0, \ldots, 0) = 0$ and $\mathsf{A}(b, \ldots, b) = b$.

We are primarily concerned with symmetric aggregation functions, i.e. functions whose value is unchanged if the inputs are rearranged. To ease notation, we will assume input vectors with arguments arranged in non-increasing order, i.e. of the form $\boldsymbol{x} = (x_{(1)}, \ldots, x_{(n)})$, such that $b \geq x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)} \geq 0$.

An aggregation function is considered to be *averaging* when

$$\min(\boldsymbol{x}) \leq \mathsf{A}(\boldsymbol{x}) \leq \max(\boldsymbol{x}).$$

In particular, we have the ordered weighted averaging (OWA) operator formalized by Yager[50] and defined with respect to a weighting vector.

**Definition 2.** A vector $\boldsymbol{w} = (w_1, \ldots, w_n)$ is called a weighting vector if $w_i \in [0, 1]$ for each $i = 1, \ldots, n$ and $\sum\limits_{i=1}^{n} w_i = 1$.

The OWA is then defined as follows.

**Definition 3.** Given a weighting vector $\boldsymbol{w}$, the OWA function is given by:

$$\mathsf{OWA}_{\boldsymbol{w}}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)}.$$

Special cases of the OWA operator, depending on the weighting vector $\boldsymbol{w}$ include:

***Arithmetic mean*** where all the weights are equal, i.e. all $w_i = \frac{1}{n}$;
***Maximum*** function for $\boldsymbol{w} = (1, 0, ..., 0)$;

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



***Minimum*** function for $\boldsymbol{w} = (0, ..., 0, 1)$;

***Median*** function for $w_i = 0$ for all $i \neq m$, $w_m = 1$ if $n = 2m-1$ is odd, and $w_i = 0$ for all $i \neq m, m+1$, $w_m = w_{m+1} = 0.5$ if $n = 2m$ is even. For an even number of arguments one can also distinguish between the upper median, which is equal to $x_{(m)}$ and the lower median $x_{(m+1)}$.

The OWA is piece-wise linear and homogeneous, i.e. multiplying the inputs by a scalar $\lambda > 0$ increases/decreases the output by the same factor, $\mathsf{OWA}(\lambda \cdot \boldsymbol{x}) = \lambda \cdot \mathsf{OWA}(\boldsymbol{x})$. It is also shift-invariant meaning that adding the same value $\lambda$ to each input increases the output by $\lambda$, $\mathsf{OWA}(x_{(1)} + \lambda, \ldots, x_{(n)} + \lambda) = \mathsf{OWA}(\boldsymbol{x}) + \lambda$.

The OWA can also be generalized to allow for non-linear behavior. We provide the definition for an ordered weighted quasi-arithmetic mean, also referred to as the generalized OWA.

**Definition 4.** For a given strictly monotone and continuous function $g : [0, b] \to [-\infty, +\infty]$, called a generating function or generator, and a weighting vector $\boldsymbol{w}$, the generalized OWA is the function

$$\mathsf{GOWA}_{\boldsymbol{w},g}(\boldsymbol{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_{(i)})\right). \tag{1}$$

As well as all special cases of the OWA for $g(t) = t$, we obtain the following as special cases:

| | | | |
|---|---|---|---|
| *Ordered weighted geometric mean* | $\mathsf{G}_{\boldsymbol{w}} =$ | $\prod_{i=1}^{n} x_{(i)}^{w_i}$, | $g(t) = \ln(t)$; |
| *Ordered weighted power mean* | $\mathsf{M}_{\boldsymbol{w},[r]} =$ | $\left(\sum_{i=1}^{n} w_i x_{(i)}^r\right)^{\frac{1}{r}}$, | $g(t) = t^r$. |

The GOWA operator is homogeneous when the generator corresponds with the weighted power means and weighted geometric means and will be shift-invariant for weighted exponential and weighted arithmetic means.

Also relevant to our results is the weighted median:

**Definition 5.** For a given weighting vector $\boldsymbol{w}$, the ordered lower weighted median is given by:

$$\mathsf{OWM}_{\boldsymbol{w}}(\boldsymbol{x}) = x_{(k)},$$

where $k$ is the greatest integer such that $\sum_{i=1}^{k-1} w_i \leq \frac{1}{2}$. Similarly the ordered upper weighted median replaces $\leq$ with $<$. The upper and lower median coincide unless there is an exact 50/50 divide between the sums of weights. Rather than take one or the other, a convention can be adopted where the average of the two is used and referred to simply as the weighted median.

We note that in standard usage the weighted median allocates weights to inputs based on the source rather than the relative order. For example, a weighting vector

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



$\boldsymbol{w} = (0.4, 0.1, 0.2, 0.3)$ when considering unordered inputs $\boldsymbol{x} = (0.2, 0.6, 0.1, 0.7)$ would permute $\boldsymbol{w}$ and $\boldsymbol{x}$ accordingly so that $w_1 = 0.4$ was assigned to $x_{(3)} = 0.2$, $w_2 = 0.1$ assigned to $x_{(2)} = 0.6$ and so on. Such weighting schemes assume that the source of the input may make the inputs more or less important whereas we consider all inputs to be equally important. For our case of vectors assumed to be in non-increasing order we will refer to the function as the ordered weighted median or OWM and observe that the output $x_{(k)}$ will always fall in the same relative position, e.g. if $\boldsymbol{w} = (0.2, 0.2, 0.2, 0.4)$ then the output will simply be the third highest input $x_{(3)}$.

Let us also define the iterated differences $\mathsf{diff}(\boldsymbol{x})$ and cumulative sum $\mathsf{cumsum}(\boldsymbol{x})$ for a vector $\boldsymbol{x} \in [0, b]^n$:

$$\mathsf{diff}(\boldsymbol{x}) = (x_{(1)} - x_{(2)}, x_{(2)} - x_{(3)}, \ldots, x_{(n-1)} - x_{(n)}), \tag{2}$$

$$\mathsf{cumsum}(\boldsymbol{x}) = (x_{(1)}, x_{(1)} + x_{(2)}, x_{(1)} + x_{(2)} + x_{(3)}, \ldots, x_{(1)} + \cdots + x_{(n)}). \tag{3}$$

We point out that $\mathsf{diff}(\boldsymbol{x})$ is a vector of $n - 1$ components.

### 2.2. *Penalty functions*

A relationship between the arithmetic mean and its associated measures of spread, statistical variance and standard deviation, can be found in its penalty-based representation. Theoretical results for penalty functions have been developed recently[12,15,16,51], however it has been well established since at least the time of Laplace that for a given set of real numbers, the arithmetic mean is the value which minimizes the sum of squared differences between the inputs and the output, i.e.

$$\mathsf{AM}(\boldsymbol{x}) = \arg\min_y \sum_{i=1}^{n} (x_i - y)^2.$$

The median can similarly be represented as the value which minimizes the sum of absolute differences $|x_i - y|$:

$$\mathsf{Med}(\boldsymbol{x}) = \arg\min_y \sum_{i=1}^{n} |x_i - y|.$$

In the case of an even number of arguments, all points lying in the interval between the $m$-th and $(m+1)$-th argument when the data are ordered (the upper and lower median respectively if ordered non-increasingly) will be minimizers, hence usually the convention is to take the average of the two.

We have the following definition for penalty functions[13].

**Definition 6.** For any closed interval $\mathbb{I} \in \mathbb{R}$, a function $P : \mathbb{I}^{n+1} \to \mathbb{R}^+$ is a penalty function if and only if there exists $c \in \mathbb{R}^+$ such that:

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



(i) $P(\boldsymbol{x}, y) \geq c$, for all $\boldsymbol{x} \in \mathbb{I}^n$, $y \in \mathbb{I}$;
(ii) $P(\boldsymbol{x}, y) = c$ if and only if $x_i = y$, for all $i = 1, \ldots, n$,; and
(iii) $P$ is quasi-convex lower semi-continuous in $y$ for each $\boldsymbol{x} \in \mathbb{I}^n$.

In our setting, we will have $c = 0$ and arguments given over a finite interval $\mathbb{I} = [0, b]$. The third condition ensures that the corresponding penalty-based function is well defined.

With respect to a penalty $P$, a penalty-based function is given by:

$$\mathsf{A}_P(\boldsymbol{x}) = \arg\min_y P(\boldsymbol{x}, y),$$

if $y$ is the unique minimizer, and $y = \frac{\underline{y}+\overline{y}}{2}$ if the set of minimizers is the closed interval $[\underline{y}, \overline{y}]$ (or similarly if the interval is open).

In particular, we note that for the generalized OWA it holds that:

$$P(\boldsymbol{x}, y) = \sum_{i=1}^{n} w_i (g(x_{(i)}) - g(y))^2,$$

and for ordered weighted medians we have:

$$P(\boldsymbol{x}, y) = \sum_{i=1}^{n} w_i |x_{(i)} - y|. \tag{4}$$

## 3. Economic Inequality Functions

The concepts of aggregation lie at the heart of various indices used to summarize incomes, access to health, level of poverty, and other socio-economic factors. In particular we have welfare functions - indicating the average level of wealth for a given population, and inequality indices - representing the disparity between rich and poor.

From here on we will consider $\boldsymbol{x}$ to denote a set of incomes (or average incomes for a given percentile group) ordered non-increasingly. A defining characteristic of inequality measures is satisfaction of the Pigou-Dalton or *progressive transfers* principle. This is the idea that any transfer of income from a richer to a poorer individual (as long as it does not alter their relative ordering) should not increase the level of inequality.

**Definition 7.** For a set of incomes $\boldsymbol{x}$ fixed for all arguments except $x_{(i)} > x_{(j)}$, a function $f$ satisfies the Pigou-Dalton or progressive transfers principle if for all $x_{(i)} - h \geq x_{(j)} + h, h > 0$ it holds that

$$f(x_{(1)}, \ldots, x_{(i)}, \ldots, x_{(j)}, \ldots, x_{(n)}) \geq f(x_{(1)}, \ldots, x_{(i)} - h, \ldots, x_{(j)} + h, \ldots, x_{(n)}). \tag{5}$$

The idea is consistent with the Lorenz curve and also, as noted in[3], is equivalent to Schur- or S-convexity.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



**Definition 8.** A function $f : [0,b]^n \to [0,\infty)$ is S-convex whenever for all $\boldsymbol{x}, \boldsymbol{y} \in [0,b]^n$ such that $\mathsf{AM}(\boldsymbol{x}) = \mathsf{AM}(\boldsymbol{y})$ it holds that $\mathsf{cumsum}(\boldsymbol{x}) \le \mathsf{cumsum}(\boldsymbol{y}) \Rightarrow f(\boldsymbol{x}) \le f(\boldsymbol{y})$.

Note that this is equivalent to the majorization based approach[40].

In order for economic inequality indices to be comparable for different currencies or different countries, their definitions often involve proportional scaling by dividing through by the arithmetic mean. We will sometimes find it convenient to represent indices in terms of the vector of proportional incomes. We shall assume that $\boldsymbol{x} \neq (0,0,\ldots,0)$ and consider a space

$$\mathcal{Q} = \{(q_1,\ldots,q_n) \in [0,1]^n : q_1 \ge \cdots \ge q_n, \sum_{i=1}^n q_i = 1\}.$$

Clearly, for any $\boldsymbol{x} \in [0,b]^n$, $\frac{1}{\mathsf{Sum}(\boldsymbol{x})}(x_{(1)},\ldots,x_{(n)}) \in \mathcal{Q}$. From this we have the following definition for inequality indices.

**Definition 9.** An inequality index is a function $\mathsf{I} : \mathcal{Q} \to [0,1]$ such that:

(i) for all $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$, if $\mathsf{cumsum}(\boldsymbol{q}) \le \mathsf{cumsum}(\boldsymbol{q}')$, then $\mathsf{I}(\boldsymbol{q}) \le \mathsf{I}(\boldsymbol{q}')$,
(ii) $\inf_{\boldsymbol{q}\in\mathcal{Q}} \mathsf{I}(\boldsymbol{q}) = 0$,
(iii) $\sup_{\boldsymbol{q}\in\mathcal{Q}} \mathsf{I}(\boldsymbol{q}) = 1$.

It follows that the supremum is obtained for $(1,0,0,\ldots,0)$ and the infimum for $(1/n,\ldots,1/n)$. It can also be observed that for homogeneous functions, expressions may only differ in terms of a scalar multiple [IV]. For example, the Gini index[34] is the function

$$\mathsf{G}(\boldsymbol{x}) = \frac{1}{2n^2 \cdot \mathsf{AM}(\boldsymbol{x})} \sum_{i=1}^n \sum_{i=1}^n |x_{(i)} - x_{(j)}|, \tag{6}$$

which can be written as a function on $\mathcal{Q}$ as follows[52]:

$$\mathsf{G}(\boldsymbol{q}) = \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n |q_i - q_j|. \tag{7}$$

As a side-note, Eq. (7) has been proposed previously for measuring ecological evenness[17] where $q_i$ denote individual species abundances. It can also be scaled by dividing through by $n(n-1)$ rather than $n^2$ so that the maximum of 1 is achieved for any value of $n$.

Other well known inequality indices include the Bonferroni index[11], and the De Vergottini index[48]. The Bonferroni index is based on differences between the overall mean income and the mean of the poorest $n-i+1$ individuals. It can be expressed:

[IV] Inequality indices and welfare functions can be referred to as relative or absolute. Relative indices usually involve some normalization so that they take values over a given interval, while absolute indices can often be understood in terms of the original units. For functions that are shift-invariant and homogeneous, the relative and absolute indices will produce the same ordering if used to compare two input sets.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



$$\mathsf{B}(\boldsymbol{x}) = \frac{1}{(n-1)\cdot \mathsf{AM}(\boldsymbol{x})} \sum_{i=1}^{n} \left( \mathsf{AM}(\boldsymbol{x}) - \frac{1}{n-i+1} \sum_{j=i}^{n} x_{(j)} \right). \tag{8}$$

We have used $(n-1)$ in the denominator rather than $n$ to ensure a supremum of 1.

In terms of $\boldsymbol{q}$, the Bonferroni index is expressed

$$\mathsf{B}(\boldsymbol{q}) = 1 - \sum_{i=1}^{n} \left( \frac{1}{n-i+1} \sum_{j=i}^{n} q_j \right).$$

The De Vergottini index on the other hand compares the total mean income with the mean of the richest $i$ individuals. Its normalized form is given by

$$\mathsf{D}(\boldsymbol{x}) = \frac{1}{\sum_{i=2}^{n} \frac{1}{i}} \left( \frac{1}{n \cdot \mathsf{AM}(\boldsymbol{x})} \left( \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{i} x_{(j)} \right) - 1 \right). \tag{9}$$

Inequalities can be constructed in relation to welfare functions[3], which are defined as aggregation functions $\mathsf{W} : [0,b]^n \to [0,b]$ that satisfy S-concavity, i.e. in this case transfers from rich to poor should *increase* the overall welfare evaluation. In particular, we have the Gini welfare function

$$\mathsf{W}_{\boldsymbol{w}}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)}, \tag{10}$$

where $\boldsymbol{w}$ satisfies $0 < w_1 < w_2 < \cdots < w_n$.

We note that this is simply an OWA with *increasing* weights. Weak weight inequalities ensures S-concavity, whereas strong weight inequalities assure strict S-concavity (see proofs in the context of inequalities[2] and OWA-based norms[7]).

Aristondo et al.[3] noted that important classical indices of inequality can be obtained as special cases of a dual decomposition with respect to Eq. (10). Denote by $\mathsf{W}^d_{\boldsymbol{w}}$ the dual of the welfare function $\mathsf{W}_{\boldsymbol{w}}$, which for Eq. (10) is simply the same expression with the order of the weights reversed. Then,

$$\mathsf{I}_{\boldsymbol{w}}(\boldsymbol{q}) = \frac{\mathsf{W}^d_{\boldsymbol{w}}(\boldsymbol{q}) - \mathsf{W}_{\boldsymbol{w}}(\boldsymbol{q})}{2}. \tag{11}$$

Inequalities of this form are also referred to as the *remainder* of the welfare function with respect to the dual decomposition[30]. In particular, it was shown that the Gini index $\mathsf{G}(\boldsymbol{x})$ corresponds with weights $w_i = \frac{2n+1-2i}{n^2}$. The Bonferroni index, on the other hand, has a welfare function corresponding with a combination of the Bonferroni and De Vergottini inequality indices (see Proposition 16 in Aristondo et al.[3]).

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



## 4. Economic Inequalities as Penalty Functions

We begin by observing that inequality measures at least coincide with penalty functions when all inputs are equal, i.e. by the first two conditions in Definition 6, $P(\boldsymbol{x}, y)$ is equal to 0 when all inputs are the same. Each penalty-based function is necessarily idempotent but not necessarily monotone. Here, instead of considering $\arg\min_y P(\boldsymbol{x}, y)$, we are interested in how the penalty associated with an aggregation function, i.e. the value $\min_y P(\boldsymbol{x}, y)$ can be related to inequality index frameworks. We show that such a penalty $P$ can be used to generalize existing indices and also leads to some interesting new functions.

### 4.1. *Penalties based on ordered weighted medians*

If we consider the Gini index, we note that differences between pairs $q_i, q_j$ can always be expressed in terms of individual differences to the median, i.e. if two arguments are either side of the median $\mathsf{Med}(\boldsymbol{q}) = y$, we have $|q_i - q_j| = |q_i - y| + |y - q_j|$, while if $y > q_i > q_j$ or $q_i > q_j > y$ we have $|q_i - q_j| = |q_i - y| - |q_j - y|$.

The Gini index can hence be expressed as a penalty function which also has the median as its minimizer. We have,

$$P_{Gini}(\boldsymbol{q}, y) = \sum_{i=1}^{n} w_i |q_i - y| \tag{12}$$

where $w_i = \frac{|n+1-2i|}{n}$. By correspondence with Eq. (4), we note that this expression is an ordered weighted median. The weighting vector $\boldsymbol{w}$ is obtained by noting the number of $|q_i - q_j|$ pairs that occur given their decreasing order. It does not verify $\mathsf{Sum}(\boldsymbol{w}) = 1$, however for any scalar multiple we obtain the same minimizer $y$ so the form here merely ensures that we have equivalence with the Gini index. As a function of $\boldsymbol{x}$ we can use weights $w_i = \frac{|n+1-2i|}{n^2 \mathsf{AM}(\boldsymbol{x})}$.

We can verify that the median will be the solution for $\arg\min_y P_{Gini}(\boldsymbol{q}, y)$ by observing that for any weighting vector with a symmetric sequence around the middle value, e.g. $\boldsymbol{w} = \frac{1}{5}(4, 2, 0, 2, 4)$, the minimizer for the $\mathsf{OWM}$ will simply be the standard median.

We can also derive the weighting scheme by considering the inequality framework based on the dual decomposition and remainder expression in Eq. (11):

$$\mathsf{I}_{\boldsymbol{w}}(\boldsymbol{q}) = \frac{\sum_{i=1}^{n} w_{n+1-i} q_i - \sum_{i=1}^{n} w_i q_i}{2},$$

for the pairs, $(q_i, q_{n+1-i})$ we have:

$$\frac{1}{2}\left((w_{n+1-i} - w_i) q_i + (w_i - w_{n+1-i}) q_{n+1-i}\right)$$

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



which becomes:

$$\frac{1}{2}(w_{n+1-i} - w_i)(q_i - q_{n+1-i}).$$

Given $\boldsymbol{w}$ is increasing, if $q_i > q_{n+1-i}$ then $w_i < w_{n+1-i}$ and vice versa, so $(w_{n+1-i} - w_i)(q_i - q_{n+1-i}) \geq 0$. For $y$ equal to the median we always have $q_i, q_{n+1-i}$ either side and so $|q_{n+1-i} - q_i| = |q_i - y| + |q_{n+1-i} - y|$ and hence by setting $v_i = v_{n+1-i} = \frac{1}{2}|w_{n+1-i} - w_i|$, we will have

$$\frac{1}{2}(w_{n+1-i} - w_i)(q_i - q_{n+1-i}) = v_i|q_i - y| + v_{n+1-i}|q_{n+1-i} - y|.$$

Any inequality index representable in the framework of Eq. (11) can therefore be expressed as a penalty function. These penalties are implicitly associated with their minimizing value which will always be the median. Income proportions $q_i$ have a larger penalty associated with them the further they are ordered from the median. Based on the dual composition, we have the following proposition.

**Proposition 1.** *Any inequality function representable in terms of the dual decomposition of an OWA based welfare function can be expressed as a penalty function which is minimized by the median.*

**Proof.** The result follows from the above where for each pair of inputs $q_i, q_{n-1+i}$ we have $\frac{1}{2}(w_{n+1-i} - w_i)(q_i - q_{n+1-i}) = v_i|q_i - y| + v_{n+1-i}|q_{n+1-i} - y|$. □

Whilst the correspondence with the dual decomposition ensures that penalties with these weighting sequences that decrease toward the median clearly satisfy the Pigou-Dalton principle, we could also consider vectors relating to a minimizer at the $k$-th position.

For example, consider a weighting sequence (without normalization) such that $\boldsymbol{w} = (4, 3, 2, 1, 0, 2, 8)$. For 7 inputs, the minimizer $y$ will be the value $q_5$, and the principle of transfers is still satisfied, however the penalty associated with $q_7$ is much higher than the penalty associated with $q_1$, so inequality is reduced more by increasing the lowest income than reducing the highest.

### 4.2. *Penalties based on generalized OWA operators*

In the previous section, the penalty expressions implicitly associated the inequality framework with the median value. Now we will consider a mean income based on the generalized ordered weighted average of incomes with equal weights, $\mathsf{GOWA}_g(\boldsymbol{q})$, which is equivalent to an unweighted quasi-arithmetic mean. From this we define an inequality $\mathsf{I}_g(\boldsymbol{q})$. Recall that for each input $q_i$ we contribute $(g(q_i) - g(y))^2$ to the penalty and note that for $g(t) = t$ we obtain a sum equivalent to scaled sample variance.

We will establish the following proposition. It considers $\mathsf{I}_g$ to take values over the range $[0, 1]$ by restricting $g$ to finite generators with a range over $[0, 1]$. We

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



note that vertical translations and scalar multiples of $g$ do not affect the associated mean, we therefore can assume $g(0) = 0$ and $g(1) = 1$. Generators with infinite values at $g(0)$ or $g(1)$ need not be considered (as explained in footnotes below). However derivatives $g'$ and $g''$ approaching $+\infty$ at the endpoints of the domain by itself will not result in violations to the Pigou-Dalton principle.

**Proposition 2.** *For an (unscaled) inequality function* $\mathsf{I}_g : \mathcal{Q} \to [0,1]$ *equal to the penalty* $\min_y P(\boldsymbol{q}, y)$ *associated with a generalized OWA, i.e.:*

$$\mathsf{I}_g(\boldsymbol{q}) = \min_y \sum_{i=1}^{n} (g(q_i) - g(y))^2,$$

*if the generator* $g : [0,1] \to [0,1]$ *is finite, increasing and twice differentiable over the open interval* $(0,1)$ *and if:*

*(i)* $g$ *is concave and* $g^2$ *is convex; or*

*(ii)* $g$ *is convex and* $(g'(t))^2 \geq g''(t)(g(1) - g(t))$ *holds for* $t \in [0,1]$,

*then* $\mathsf{I}_g$ *satisfies the Pigou-Dalton principle.*

**Proof.** Given $y = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(q_i)\right)$, we can express each term $(g(q_i) - g(y))^2$ without reference to $y$ and hence we obtain,

$$\mathsf{I}_g(\boldsymbol{q}) = \sum_{i=1}^{n} \left( g(q_i) - g\left( g^{-1}\left( \frac{1}{n} \sum_{i=1}^{n} g(q_i) \right) \right) \right)^2,$$

which simplifies to

$$\mathsf{I}_g(\boldsymbol{q}) = \left( \sum_{i=1}^{n} g(q_i)^2 \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} g(q_i) g(q_j). \tag{13}$$

We wish to establish that for any $q_i > q_j$, progressive transfers, i.e. decreasing $q_i$ and increasing $q_j$ will decrease the overall inequality $\mathsf{I}_g(\boldsymbol{q})$. Denote by $\mathsf{I}'_i(q_i)$ the partial derivative of $\mathsf{I}_g$ in the direction of the $i$-th component, we have,

$$\mathsf{I}'_i(q_i) = \frac{2}{n} \cdot g'(q_i) \left( (n-1) g(q_i) - \sum_{\forall j \neq i} g(q_j) \right),$$

and we need to show that for any pairs such that $q_i > q_j$, a simultaneous decrease in $q_i$ and increase in $q_j$ will result in an overall decrease. In other words, it must hold that $\mathsf{I}'_j(q_j) - \mathsf{I}'_i(q_i)$ is negative or simply that $\mathsf{I}'_i(q_i) > \mathsf{I}'_j(q_j)$.

We therefore look at the second derivatives with respect to $q_i$, which we will denote by $\mathsf{I}''_{ii}(q_i)$. If $\mathsf{I}''_{ii}(q_i) > 0$ for all $\boldsymbol{q}$, then for any $q_i > q_j$ it will hold that $\mathsf{I}'_i(q_i) > \mathsf{I}'_j(q_j)$.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



$$\mathsf{I}''_{ii}(q_i) = \frac{2}{n}\left(g''(q_i)\left((n-1)g(q_i) - \sum_{j\neq i}^{n} g(q_j)\right) + g'(q_i)\left((n-1)g'(q_i)\right)\right)$$

$$= \frac{2(n-1)}{n}\left(g''(q_i)\left(g(q_i) - \frac{1}{n-1}\sum_{j\neq i}^{n} g(q_j)\right) + (g'(q_i))^2\right).$$

We consider two cases, $g$ being concave and $g$ being convex[V].
$(i)$ $g$ concave

If $g'' < 0$, i.e. $g$ is concave, then we can ignore the sum involving the values $q_j, j \neq i$ and hence $\mathsf{I}''_{ii}$ will be positive as long as[VI],

$$g''(q_i)g(q_i) + (g'(q_i))^2 \geq 0, \tag{14}$$

which leads to condition $(i)$ of the proposition.
$(ii)$ $g$ convex

Let us denote the difference between $g(q_i)$ and the mean of the remaining $g(q_j)$ by $\delta$. If $g$ is convex then $g'' > 0$ and we require,

$$g''(q_i)\delta + (g'(q_i))^2 \geq 0.$$

If $\delta$ is positive then the equation will be satisfied, so we need not be concerned with $g'$ being infinite at $g(1)$. On the other hand, if $g(q_i)$ is below the mean of the remaining transformed values then $\delta$ will be negative and we hence require,

$$(g'(q_i))^2 \geq g''(q_i)|\delta|.$$

The second derivative hence needs to be smaller than the ratio of the square of the first derivative to this difference. Furthermore, for any given $q_i$ the maximum $|\delta|$ occurs where $q_j = 1, \forall j \neq i$ which gives us condition $(ii)$. □

We can now consider some examples to show the usefulness of the result. Consider the case of power means where $g(t) = t^r$. The generator is concave when $r < 1$ and so for $\mathsf{I}_g$ to satisfy the Pigou-Dalton principle we require, $(t^r)^2$ to be convex, which holds for $2r(2r-1)t^{2r-2} > 0$ or $0.5 \leq r \leq 1$.

**Corollary 1.** *An inequality function equal to the penalty associated with a power mean* $\mathsf{M}_{[r]}$ *will satisfy the Pigou-Dalton principle if* $0.5 \leq r \leq 1$.

---

[V] Observe at this point that if $g$ were concave and approached $-\infty$ for $g(0)$ then $g(q_i) - \frac{1}{n-1}\sum_{j\neq i} q_j$ could be infinitely large and the equation would be violated (see the example for $g(t) = \ln t$ following the proof). Similarly for convex $g$ and $g(1) = \infty$ the product of $g''$ and this difference can be arbitrarily large and negative.

[VI] Observe that $g'$ potentially being infinitely large at $g(0)$ is not problematic here since the difference between $g(q_i)$ and the mean of $g(q_j)$ would be negative in this case.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



On the other hand, the concave generator $g(t) = \ln t$ which corresponds with the geometric mean will not lead to a penalty that satisfies the Pigou-Dalton principle. As a numeric example and noting that we cannot scale $g$ so that it takes a range in $[0, 1]$,

$$\min_y P((0.001, 0.001, 0.001, 0.001, 0.496, 0.5), y) = 51.42864$$

$$< \min_y P((0.001, 0.001, 0.001, 0.001, 0.497, 0.499), y) = 51.42871.$$

As an example for convex generators, let us consider generators of the form $g(t) = \frac{t(t+a)}{1+a}$. It can be shown that if $a \geq 1 + \sqrt{12}/2$, then $\mathsf{I}_g(\boldsymbol{q})$ will satisfy the Pigou-Dalton principle, so $a = 3$ leads to an acceptable function whereas $a = 2$ does not.

For $a = 3$ we have $g(t) = \frac{t(t+3)}{4}, g'(t) = \frac{2t+3}{4}, g''(t) = \frac{1}{2}$ and $g(1) = 1$. By condition $(ii)$, since $\left(\frac{2t+3}{4}\right)^2 = \frac{4t^2+12t+9}{16} > \frac{1}{2}\left(1 - \frac{t(t+3)}{4}\right) = \frac{1}{16}\left(8 - 2t^2 - 6t\right)$ for $t \in [0, 1]$.

Proposition 2 is interesting in that only a small range of generating functions seem suitable for modelling inequality in this framework. We note however that if we consider the mean associated with $\mathsf{I}_g$, i.e. the value that minimizes the penalty, it can be thought of as an average proportion of income. For a concave increasing generating function, the mean value (not necessarily the penalty) will always increase under progressive transfers, since for $q_i > q_j$ the increase in $g(q_j)$ will always be greater than the decrease in $g(q_i)$. We may prefer to observe strict satisfaction of the conditions so that progressive transfers never leave the level of inequality the same.

## 5. Inequality Measures and Their Relation to Spread Measures

While aggregation functions are used to give the central tendency of a data set or its "typical" value, their use is often accompanied in analysis with an indication of how much the data varies. In statistics we have – among others – the standard deviation, variance, Fisher's mean error, interquartile range, range and median absolute deviation, while the idea of modelling disagreement has taken hold in group decision making as essential for reaching a final evaluation. Such data summaries can be considered in the framework of spread measures[24,25].

**Definition 10.** An (absolute) symmetric spread measure is a function $\mathsf{V} : [0, b]^n \to [0, \infty]$ such that

(i) for each $\boldsymbol{x}, \boldsymbol{y}$ if it holds $\mathsf{diff}(\boldsymbol{x}) \leq \mathsf{diff}(\boldsymbol{y})$, then $\mathsf{V}(\boldsymbol{x}) \leq \mathsf{V}(\boldsymbol{y})$,
(ii) $\inf_{\boldsymbol{x} \in [0,b]^n} \mathsf{V}(\boldsymbol{x}) = 0$.

Note that (ii) in the above context is equivalent to stating that $\mathsf{V}(c, c, \ldots, c) = 0$ for all $c \in [0, b]$.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



In practice, checking if a function $\mathsf{V}$ is a spread measure can often be performed by rewriting it as a function of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{n-1}) = \mathsf{diff}(\boldsymbol{x})$. If $\mathsf{V}(\boldsymbol{\delta})$ does not depend on $x_{(n)}$ (and hence is shift invariant), it is nondecreasing with respect to each $\delta_i$, and it holds $\mathsf{V}(0, 0, \ldots, 0) = 0$, then we get that $\mathsf{V}$ is a spread measure.

Let us study which of the considered penalties not only can be used to measure inequality of a data distribution, but also the data spread. The question is interesting as we already pointed out in the introduction that inequality and spread measures seem to be related to each other.

Firstly, let us focus on inequality functions representable in terms of the dual decomposition of an OWA-based welfare function. To recall, the Gini index is a member of this class and all such measures can be expressed as a penalty function minimized by the median, see Proposition 1.

**Proposition 3.** *Let $\boldsymbol{w}$ be such that $0 < w_1 < \cdots < w_n$. Then*

$$\mathsf{I}_{\boldsymbol{w}}(\boldsymbol{q}) = \frac{\sum_{i=1}^{n} w_{n-i+1} q_i - \sum_{i=1}^{n} w_i q_i}{2}$$

*is a spread measure.*

**Proof.** Fix $\boldsymbol{q}$ and let $\boldsymbol{\delta} = (q_1 - q_2, \ldots, q_{n-1} - q_n)$. Under the convention that $\sum_{j=u}^{v} q_j = 0$ for $v < u$, it holds $q_i = q_n + \sum_{k=i}^{n-1} \delta_k$ and of course $\boldsymbol{\delta} \geq \mathbf{0}$. We have:

$$\begin{aligned} 2\mathsf{I}_{\boldsymbol{w}}(\boldsymbol{\delta}) &= \sum_{i=1}^{n} (w_{n-i+1} - w_i)(q_n + \sum_{k=i}^{n-1} \delta_k) \\ &= q_n \sum_{i=1}^{n} (w_{n-i+1} - w_i) + \sum_{i=1}^{n-1} (w_{n-i+1} - w_i) \sum_{k=i}^{n-1} \delta_k \\ &= \sum_{k=1}^{n-1} \delta_k \left( \sum_{i=n-k+1}^{n} w_i - \sum_{i=1}^{k} w_i \right). \end{aligned}$$

As $\boldsymbol{w}$ is increasing, $\sum_{i=n-k+1}^{n} w_i - \sum_{i=1}^{k} w_i \geq 0$ for all $k$. Thus, $\mathsf{I}_{\boldsymbol{w}}$ is a function nondecreasing with respect to each $\delta_k$. Moreover, it does not depend on $q_n$ and we have that $\mathsf{I}_{\boldsymbol{w}}(\boldsymbol{\delta}) = 0$ for $\boldsymbol{\delta} = \mathbf{0}$. Hence, $\mathsf{I}_{\boldsymbol{w}}$ is a spread measure. □

Next, let us study functions equal to the penalty associated with a generalized OWA, compare Proposition 2. However, as the only shift invariant generalized OWA operators are the arithmetic mean and symmetric exponential means, we are going to restrict the input data domain to the case $q_n = 0$.

**Proposition 4.** *Let $g$ be a strictly increasing, twice differentiable, and continuous generator with $g(0) = 0$ and $g(1) = 1$ and consider a function given by:*

$$\mathsf{I}_g(q_1, \ldots, q_{n-1}, 0) = \sum_{i=1}^{n-1} \left( g(q_i) - \frac{1}{n} \sum_{i=1}^{n-1} g(q_i) \right)^2.$$

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



*Then $\mathsf{I}_g$ is a spread measure if and only if $g$ is convex.*

**Proof.** Again, for a given $\boldsymbol{q}$ let $\boldsymbol{\delta} = (q_1 - q_2, \dots, q_{n-1} - q_n)$. By (13) we have:

$$\mathsf{I}_g(\boldsymbol{q}) = \sum_{i=1}^{n} g(q_i)^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g(q_i)g(q_j).$$

$$\mathsf{I}_g(\boldsymbol{\delta}) = \sum_{i=1}^{n} g\left(q_n + \sum_{k=i}^{n-1}\delta_k\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g\left(q_n + \sum_{k=i}^{n-1}\delta_k\right) g\left(q_n + \sum_{k=j}^{n-1}\delta_k\right).$$

Under the assumption that $q_n = 0$ and $g(0) = 0$ it holds:

$$\mathsf{I}_g(\boldsymbol{\delta}) = \sum_{i=1}^{n-1} g\left(\sum_{k=i}^{n-1}\delta_k\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n-1} g\left(\sum_{k=i}^{n-1}\delta_k\right)\right)^2.$$

For any $l$:

$$\begin{aligned}
\frac{\partial}{\partial\delta_l}\mathsf{I}_g(\boldsymbol{\delta}) &= 2\sum_{i=1}^{n-1} g\left(\sum_{k=i}^{n-1}\delta_k\right) g'\left(\sum_{k=i}^{n-1}\delta_k\right)\mathbf{1}(i \le l) \\
&\quad - \frac{2}{n}\sum_{i=1}^{n-1} g\left(\sum_{k=i}^{n-1}\delta_k\right)\sum_{j=1}^{n-1} g'\left(\sum_{k=j}^{n-1}\delta_k\right)\mathbf{1}(j \le l) \\
&= 2\sum_{i=1}^{l} g'\left(\sum_{k=i}^{n-1}\delta_k\right)\left(g\left(\sum_{k=i}^{n-1}\delta_k\right) - \frac{1}{n}\sum_{j=1}^{n-1} g\left(\sum_{k=j}^{n-1}\delta_k\right)\right)
\end{aligned}$$

In order for $\mathsf{I}_g$ to be a nondecreasing function of each $\delta_l$, we must verify that $\partial\mathsf{I}_g(\boldsymbol{\delta})/\partial\delta_l \ge 0$.

Let us consider $\boldsymbol{\delta}$ such that $\delta_1 = \dots = \delta_{n-2} = 0$. In such a case we require that:

$$0 \le g'(\delta_{n-1})\left(ng(\delta_{n-1}) - (n-1)g(\delta_{n-1})\right) = g'(\delta_{n-1})\, g(\delta_{n-1})$$

for any $\delta_{n-1} \in [0,1]$. That is, $g$ is necessarily convex.

To show that this is also a sufficient condition, let us assume that $\gamma_{n-1} = g\left(\sum_{k=n-1}^{n-1}\delta_k\right)$, $\gamma_{n-2} = g\left(\sum_{k=n-2}^{n-1}\delta_k\right) - g\left(\sum_{k=n-1}^{n-1}\delta_k\right)$, $\dots$, $\gamma_1 = g\left(\sum_{k=1}^{n-1}\delta_k\right) - g\left(\sum_{k=2}^{n-1}\delta_k\right)$ and $\rho_{n-1} = g'\left(\sum_{k=n-1}^{n-1}\delta_k\right)$, $\rho_{n-2} = g'\left(\sum_{k=n-2}^{n-1}\delta_k\right) - g'\left(\sum_{k=n-1}^{n-1}\delta_k\right)$, $\dots$, $\rho_1 = g'\left(\sum_{k=1}^{n-1}\delta_k\right) - g'\left(\sum_{k=2}^{n-1}\delta_k\right)$. Clearly, $\rho_i, \gamma_i \ge 0$ for all $i$. Now for any $l$ we consider the inequality:

$$\begin{aligned}
0 &\le \sum_{i=1}^{l}\left(\sum_{k=i}^{n-1}\rho_k\right)\left(n\sum_{k=i}^{n-1}\gamma_k - \sum_{j=1}^{n-1}\sum_{k=j}^{n-1}\gamma_k\right) \\
&= \sum_{i=1}^{n-1}\rho_i\left(\sum_{j=1}^{n-1}\gamma_j\left(((in - ji) \wedge (ln - jl)) - (0 \vee ((in - jn) \wedge (ln - jn)))\right)\right),
\end{aligned}$$

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



where $\wedge$ and $\vee$ denote the binary min and max operators, respectively. It is easy to verify that for any $j, l \leq n-1$ it holds $(in - ji \wedge ln - jl) - (0 \vee (in - jn \wedge ln - jn)) \geq 0$. Thus, $\mathsf{I}_g$ is a spread measure if and only if $g'(x) \geq 0$ for all $x$. □

Based together on the above result and Proposition 2, we can easily derive new functions that are both inequality as well as spread measures. However, in particular, the only inequality function equal to the penalty associated with a power mean that is at the same time a spread measure is based on $g(t) = t$, i.e., the arithmetic mean. In other words, it is a function proportional to the classical sample variance.

## 6. Directions for future research

Just as ordinary aggregation functions may be conceived of as nondecreasing functions with respect to each element in a given vector and spread measures are nonincreasing mappings of iterated differences, social inequality measures are nonincreasing functions of cumulative sums. Given any $\boldsymbol{x}$ and denoting $\boldsymbol{\delta} = \mathsf{diff}(\boldsymbol{x})$ and $\boldsymbol{\zeta} = \mathsf{cumsum}(\boldsymbol{x})$, it is quite easy to obtain $\boldsymbol{x}$ by applying cumsum on a transformed version of $\boldsymbol{\delta}$ as well as by applying diff on accordingly modified $\boldsymbol{\zeta}$. In this sense, cumsum and diff may be conceived as dual operations. Also, note that intuitively the diff operator can be viewed as a (discrete) derivative of a vector, while cumsum mimics an (iterated) integral. Having said that, we have established a nice connection between the three classes of data aggregation tools.

For instance, we can easily check that the Gini index is merely a linear combination of $\zeta_1, \ldots, \zeta_n$. Considering classical aggregation functions, but this time applied on vectors of cumulative sums, may lead us to new and interesting inequality measures.

We have referred to expressions of inequality based on the highest small proportion of individuals controlling a disproportionate amount of income. This is how the so-called $k$-index is defined[33,37]. Interestingly enough, it also can be related to a penalty based on the median. Here we can consider some functions based on Sugeno integrals.

### 6.1. *The k-index*

The so-called $k$-index gives the value $k \in [0.5, 1]$ such that the proportion $1 - k$ of the individuals receives the proportion $k$ of the income[33,37]. For example, an extreme case is when 1% of individuals control 99% of the wealth and hence here the $k$ value is 0.99. Such a value is similar to the Hirsch index[36] (or $h$-index) used in bibliometrics, which in turn is a special case of the Sugeno integral[46], an averaging aggregation function defined with respect to a fuzzy measure.

**Definition 11.** Let $[n] = \{1, \ldots, n\}$. A fuzzy measure $v$ is a set function defined on the power set of $[n]$, i.e. $v : 2^{[n]} \to [0, 1]$, such that:

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



(i) $\mathcal{A} \subseteq \mathcal{B} \Rightarrow v(\mathcal{A}) \leq v(\mathcal{B})$,
(ii) $v(\emptyset) = 0$,
(iii) $v([n]) = 1$.

**Definition 12.** For a given set function $v$, the Sugeno integral is the function:

$$S(\boldsymbol{x}) = \max_{i=1,\ldots,n} \min(x_{(n-i+1)}, v(H_i)),$$

where $H_i = \{i, i+1, \ldots, n\}$.

The $k$-index can be expressed as a Sugeno integral which takes the cumulative sum values $\zeta = \mathsf{cumsum}(\boldsymbol{q})$ as its arguments. We have:

$$\mathsf{S}(\zeta_1, \zeta_2, \ldots, \zeta_n) = \max\min(\zeta_i, v(H_i)),$$

where $v$ is the counting measure, i.e. a symmetric fuzzy measure given by $v(\mathcal{A}) = \frac{|\mathcal{A}|}{n}$.

Sugeno integrals with respect to symmetric fuzzy measures actually correspond with a median of the input vector appended to the sequence of fuzzy measure values, i.e. $\mathsf{Med}(\zeta_1, \ldots, \zeta_n, v([n]), v([n-1]), v([n-2]), \ldots, v([1]))$. We therefore can express $k$ in terms of the following penalty,

$$k = \arg\min_y \sum_{i=1}^{n} |\zeta_i - y| + \sum_{i=1}^{n} |v([i]) - y|.$$

Note that in this case, $k$ is equal to the minimizer $y$, rather than the penalty associated with it. The absolute minimum that can be achieved in terms of the penalty occurs for $k = y = 0.5$ corresponding with inputs $q_i = 1/n$ for all $i$ and hence with $z_i = i/n$. The penalty in this case will not be zero and hence here we are not dealing strictly with a penalty-based function.

### 6.2. *The g-index*

Similarly to the $k$-index, the $g$-index[21] (note here that $g$ does not refer to the generating function as we have discussed previously), which in informetrics and bibliometrics counts the maximum number of papers written by an academic such that the highest cited $g$ papers have received $g^2$ citations in total. Originally, such a function was used to quantify a scholar's performance. However, in the context of measuring inequality, the $g$-index can be interpreted as the proportion $g$ of the population earning $(1-g)^2$ of the total income. For proportional values $\zeta$, this function is in fact a Sugeno integral with respect to a fuzzy measure with $v(\mathcal{A}) = |\mathcal{A}|^2$.

### 6.3. *Relationship to entropy*

Here we have studied functions with weighting parameters, however many functions used in economics essentially apply weights that are equal to the proportional vectors. The entropy calculation is one such function and also satisfies the progressive

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



transfer principle. In reverse to the approach we have taken here, we could consider defining a mean from the entropy calculation. That is, we could find the value $y$, which, when appended to the vector $\boldsymbol{x}$, maximizes the total entropy in terms of $\boldsymbol{q} \in \mathcal{Q}$,

$$\mathsf{Ent}(\boldsymbol{q}) = -\sum_{i=1}^{n} q_i \ln q_i.$$

The corresponding mean would then have the form, $\mathsf{M}_{\mathsf{Ent}}(\boldsymbol{q}) = \arg\max_y \mathsf{Ent}(\boldsymbol{q}, y)$. The value $y \in [0,1]$ will be averaging with respect to the $q_i$ and as a maximizer of entropy it can be expressed

$$\mathsf{M}_{\mathsf{Ent}}(\boldsymbol{q}) = \arg\max_y \left( \frac{y}{y+1} \ln \frac{y+1}{y} + \sum_{i=1}^{n} \frac{q_i}{y+1} \ln \frac{y+1}{q_i} \right).$$

It can also be calculated directly as, $\mathsf{M}_{\mathsf{Ent}}(\boldsymbol{x}) = \exp(-\mathsf{Ent}(\boldsymbol{q}))$. The resulting averaging function is not monotone, however entropy is still a valid penalty representing the reciprocal notion to inequality, and so such means might well be useful. In ecology, an equivalent calculation known as Shannon's diversity represents the 'effective' number of species as a ratio of the total number of species. The economic interpretation then might be the effective number of individuals sharing in the overall wealth of a country or community. A similar index is Simpson' diversity index, $\mathsf{Simp}(\boldsymbol{q}) = \sum_{i=1}^{n} q_i^2$, which reaches a maximum of 1 for $\boldsymbol{q} = (1, 0, 0, \ldots, 0)$ and minimum of $1/n$ when $q_i$ are equal. An inequality index based on this could be given by $1/(n \cdot \mathsf{Simp}(\boldsymbol{q}))$.

## 7. Conclusion

We have shown that penalty functions can provide a useful framework for economic inequality measures and established some results that show the conditions under which penalties will satisfy the Pigou-Dalton principle. As with the dual decomposition, penalty functions are naturally affiliated with an averaging aggregation function, the understanding of which can help with insight into the inequality function's behavior and interpretation.

As well as establishing a general framework for defining penalties equivalent to those based on OWA operators, we have shown that penalties related to quasi-arithmetic means can be used, provided some generalized convexity requirements are satisfied. We then established relationships between inequality measures and spread measures and provided discussion about connections with entropy and functions of cumulative sums.

While inequality in the context of welfare economics and penalties and spread in the context of aggregation have been studied in their respective fields until recently, it is clear that much can be gained in both these and other research fields by looking at relationships between important results.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



### Acknowledgments

Marek Gagolewski's study was supported by the National Science Center, Poland, research project 2014/13/D/HS4/01700.

### References

1. J. Alcalde-Unzu and M. Vorsatz. Measuring consensus: concepts, comparisons and properties. In *Consensual Processes, STUDFUZZ*, volume 267, pages 195–211. Springer-Verlag, Berlin, 2011.
2. O. Aristondo, J.L. García-Lapresta, C. Lasso de la Vega, and R. A. Marques Pereira. The Gini index, the dual decomposition of aggregation functions, and the consistent measurement of inequality. *International Journal of Intelligent Systems*, 27:132–152, 2012.
3. O. Aristondo, J.L. García-Lapresta, C. Lasso de la Vega, and R. A. Marques Pereira. Classical inequality indices, welfare and illfare functions, and the dual decomposition. *Fuzzy Sets and Systems*, 228:114–136, 2013.
4. G. Beliakov, H. Bustince, and T. Calvo. *A Practical Guide to Averaging Functions*. Springer, Berlin, New York, 2015.
5. G. Beliakov, T. Calvo, and S. James. Consensus measures constructed from aggregation functions and fuzzy implications. *Knowledge-Based Systems*, 55:1–8, 2014.
6. G. Beliakov and S. James. Unifying approaches to consensus across different preference representations. *Applied Soft Computing*, 35:888–897, 2015.
7. G. Beliakov, S. James, and G. Li. Learning Choquet integral-based metrics for semi-supervised clustering. *IEEE Transactions on Fuzzy Systems*, 19(3):562–574, 2011.
8. G. Beliakov, S. James, and D. Nimmo. Can indices of ecological evenness be used to measure consensus? In *Proceedings of the 2014 IEEE International Conference on Fuzzy Systems*, pages 1–8. Beijing, China, 2014.
9. G. Beliakov, S. James, and L. Smith. Single-preference consensus measures based on models of ecological evenness. In *Proceedings of MDAI, LNAI 8825*, pages 50–59. Tokyo, Japan, 2014.
10. G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer, Heidelberg, 2007.
11. C. Bonferroni. *Elementi di Statistica Generale*. Libreria Seber, Firenze, 1930.
12. H. Bustince, E. Barrenechea, T. Calvo, S. James, and G. Beliakov. Consensus in multiexpert decision making problems using penalty functions defined over a Cartesian product of lattices. *Information Fusion*, 17:56–64, 2014.
13. H. Bustince, G. Beliakov, B. Bedregal G. P. Dimuro, and R. Mesiar. On the definition of penalty functions in aggregations. *Fuzzy Sets and Systems*, page (submitted), 2016.
14. F. J. Cabrerizo, S. Alonso, I. J. Pérez, and E. Herrera-Viedma. On consensus measures in fuzzy group decision making. In *Proceedings of MDAI 2008*, pages 86–97, 2008.
15. T. Calvo and G. Beliakov. Aggregation functions based on penalties. *Fuzzy Sets and Systems*, 161:1420–1436, 2010.
16. T. Calvo, R. Mesiar, and R. R. Yager. Quantitative weights and aggregation. *IEEE Transactions on Fuzzy Systems*, 12(1):62–69, 2004.
17. J.A. Camargo. Must dominance increase with the number of subordinate species in competitive interactions? *Journal of Theoretical Biology*, 161(4):537–542, 1993.
18. S.R. Chakravarty. *Ethical Social Index Numbers*. Springer-Verlag, New York, 1990.
19. H. Dalton. The measurement of the inequality of incomes. *The Economic Journal*, 30:348361, 1920.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018




20. B. Dutta. Inequality, poverty and welfare. In K.J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, pages 597–633. Elsevier, 2002.
21. L. Egghe. Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, 2006.
22. D. A. Fleming and T. G. Measham. Income inequality across Australian regions during the mining boom: 2001–2011. *Australian Geographer*, 46(2):203–216, 2015.
23. D. A. Fleming and T. G. Measham. Rich and poor: which areas of Australia are most unequal? *The Conversation: Australian Edition*, 2015.
24. M. Gagolewski. Normalized $WD_pWAM$ and $WD_pOWA$ spread measures. In J.M. Alonso, H. Bustince, and M. Reformat, editors, *Proc. IFSA/Eusflat 2015*, pages 210–216. Atlantic Press, 2015.
25. M. Gagolewski. Spread measures and their relation to aggregation functions. *European Journal of Operational Research*, 241(2):469–477, 2015.
26. M. Gagolewski, M. Bartoszuk, and A. Cena. Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. *Information Sciences*, 363:8–23, 2016.
27. M. Gagolewski, A. Cena, and M. Bartoszuk. Hierarchical clustering via penalty-based aggregation and the Genie approach. In *Proceedings of MDAI.*
28. J. L. García-Lapresta and D. Pérez-Román. Some measures of consensus generated by distances on weak orders. In *Proceedings of XIV Congreso Español sobre Tecnologías y Lógica fuzzy*, pages 477–483, Cuencas Mineras, 17 –19 September, 2008.
29. J.L. García-Lapresta, C. Lasso de la Vega, R. A. Marques Pereira, and A. Urrutia. A new class of fuzzy poverty measures. In *Proc. of IFSA/EUSFLAT2015*, pages 1140–1146. Gijón, Spain, 2015.
30. J.L. García-Lapresta and R.A. Marques Pereira. The self-dual core and the anti-self-dual remainder of an aggregation operator. *Fuzzy Sets and Systems*, 159:47–62, 2008.
31. J.L. García-Lapresta and D. Pérez-Román. Measuring consensus in weak orders. In *Consensual Processes, STUDFUZZ*, volume 267, pages 213–234. Springer-Verlag, Berlin, 2011.
32. J.L. García-Lapresta and D. Pérez-Román. Consensus-based clustering under hesitant qualitative assessments. *Fuzzy Sets and Systems*, 292:261–273, 2016.
33. A. Ghosh, N. Chattopadhyay, and B. K. Chakrabarti. Inequality in societies, academic institutions and science journals: Gini and $k$-indices. *Physica A*, 410(14):30–34, 2014.
34. C. Gini. *Variabilità e Mutabilità*. Tipografia di Paolo Cuppini, Bologna, 1912.
35. M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation Functions*. Cambridge University Press, Cambridge, 2009.
36. J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.
37. J.-I. Inoue, A. Ghosh, A. Chatterjee, and B. K. Chakrabarti. Measuring social inequality with quantitative methodology: Analytical estimates and empirical data analysis by Gini and $k$ indices. *Physica A*, 429:184–204, 2015.
38. G. Knight. Factcheck Q&A: is Australia the most unequal it has been in 75 years? *The Conversation: Australian Edition*, 2015.
39. M. O. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70):209–219, 1905.
40. A. W. Marshall and I. Olkin. Inequalities: Theory of majorization and its applications. *Mathematics in Science and Engineering*, 143, 1979.
41. Australian Bureau of Statistics (ABS). Personal income by sex by employment type. viewed 16 November 2015, http://www.abs.gov.au.
42. C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
43. E. H. Simpson. Measurement of diversity. *Nature*, 163:688, 1949.

Please cite this paper as: Beliakov G., Gagolewski M., James S., Penalty-based and other representations of economic inequality, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 24(Suppl. 1), 2016, pp. 1-23. doi:10.1142/S0218488516400018



44. C. Taillie. Species equitibility: a comparative approach. In *Ecological diversity in theory and practice*, pages 51–62. Int. Coop. Publ. House, Fairland, Maryland, 1979.
45. V. Torra and Y. Narukawa. *Modeling Decisions. Information Fusion and Aggregation Operators*. Springer, Berlin, Heidelberg, 2007.
46. V. Torra and Y. Narukawa. The h-index and the number of citations: two fuzzy integrals. *IEEE Transactions on Fuzzy Systems*, 16(3):795–797, 2008.
47. Economist Intelligence Unit. Women's economic opportunity 2012: A global index and ranking from the economist intelligence unit. *www.eiu.com*, pages 1–51, accessed August, 2015.
48. M. De Vergottini. Sugli indici di concentrazione. *Statistica*, 10:445–454, 1940.
49. S. Wisor, S. Bessell, F. Castillo, J. Crawford, K. Donaghue, J. Hunt, A. Jaggar, A. Liu, and T. Pogge. The individual deprivation measure: a gender-sensitive approach to poverty measurement. *International Women's Development Agency, www.iwda.org.au*, pages 1–68, accessed August, 2015.
50. R. R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18:183–190, 1988.
51. R. R. Yager and A. Rybalov. Noncommutative self-identity aggregation. *Fuzzy Sets and Systems*, 85:73–82, 1997.
52. S. Yitzhaki. More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8:13–30, 1998.