Gini index and Robin Hood index

Rafał Kucharski
Notation

- Let \( N = \{1, 2, \ldots, n\} \) be a set of individuals.
- Incomes: \( x = (x_1, \ldots, x_n) \in \mathbb{R}_+^n\setminus\{0\} \).
- Value \( x_i \) denotes the income of the agent \( i \).
- For \( x \in \mathbb{R}^n \) let \( x^* \) denote the vector obtained from \( x \) by sorting its element in non-decreasing way:

\[
x_1^* \leq \cdots \leq x_n^*.
\]

- The group of permutations on \( N \): \( S_n = \{\sigma: N \to N: \sigma \text{ is bijective}\} \).
Gini index

- Discussed in 1876 by F. R. Helmert (geodesist, discoverer of the chi-squared distribution?),
- “Rediscovered” by Corrado Gini in 1912.
- For integrable random variable $X$:

$$G(X) = \frac{\mathbb{E}|X - Y|}{2\mathbb{E}(X)},$$

where $Y$ is an independent copy of $X$.
- Discrete equivalent:

$$G(x) = \frac{\sum_{i,j=1}^{n} |x_i - x_j|}{2n \sum_{i=1}^{n} x_i}, \quad x \in \mathbb{R}^n.$$

- Another abstract formula:

$$G(X) = 1 - \frac{\mathbb{E}(\min(X, Y))}{\mathbb{E}(X)},$$

where $Y$ is an independent copy of $X$. 

The attractiveness of the Gini index to many economists is that it has an intuitive geometric interpretation: it can be defined as the ratio of the area that lies between the line of perfect equality and the Lorenz curve, over the total area under the line of equality.
Lorenz Curve

Lorenz curve:

\[((F(x), L(F(x)))) : x \in \mathbb{R}\},

where \( F \) is the cdf of \( X \) and

\[
L(F(x)) = \frac{\mathbb{E}(X \cdot 1(X \leq x))}{\mathbb{E}(X)} = \frac{\int_{-\infty}^{x} t \, dF(t)}{\mathbb{E}(X)}.\]
Sen [9] defined the Gini index as a function $G : \mathbb{R}_+^n \setminus \{0\} \rightarrow \mathbb{R}$ such that

$$G(x) = \frac{1}{n} \left[ n + 1 - 2 \frac{\sum_{i=1}^{n} (n + 1 - i)x_i^*}{\sum_{i=1}^{n} x_i^*} \right]. \quad (1)$$

For this discrete version, $0 \leq G(x) \leq \frac{n-1}{n}$ for $x \in \mathbb{R}_+^n \setminus \{0\}$.

Equivalent formula for non-increasingly ordered $x^*$:

$$G(x) = \frac{1}{n} \left[ n + 1 - 2 \frac{\sum_{i=1}^{n} ix_{i^*}}{\sum_{i=1}^{n} x_{i^*}} \right]. \quad (2)$$
Axioms

In Plata-Pérez et al. [2015], the following properties were defined:

1. The inequality index $I$ is said to be **scale-independent** if

   $$I(\lambda x) = I(x) \quad \text{for all } x \in \mathbb{R}_+^n \setminus \{0\} \text{ and } \lambda > 0.$$ 

2. The inequality index $I$ is **symmetric** if and only if

   $$I(x_{\sigma}) = I(x) \quad \text{for every } \sigma \in S_n \text{ and } x \in \mathbb{R}_+^n \setminus \{0\}.$$ 

3. The inequality index $I$ is **comonotone separable** if

   $$I(\beta x + (1 - \beta)y) = \beta I(x) + (1 - \beta)I(y)$$

   for every $\beta \in [0, 1]$, every $x, y \in \mathbb{R}_+^n \setminus \{0\}$ that are comonotone and

   $$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i.$$  

   (We say that $x$ and $y$ are **comonotone** if

   $$(x_i - x_j)(y_i - y_j) \geq 0, \text{ for all } i, j \in \mathbb{N}.$$ )
Axioms cont.

4. The inequality index $I$ is said to be **standardized** if $I(\nu^{k+1}) = \frac{k}{n}$, where

$$\nu^{k+1} = \left(0, \ldots, 0, \frac{1}{n-k}, \ldots, \frac{1}{n-k}\right), \quad k = 0, \ldots, n - 1,$$

The idea behind this axiom is such that:

$$I(\nu^1) = I(\frac{1}{n}, \ldots, \frac{1}{n}) = 0, \quad I(\nu^n) = I(0, \ldots, 0, 1) = \frac{n-1}{n},$$

and the differences

$$I(\nu^1) - I(\nu^2) = I(\frac{1}{n}, \ldots, \frac{1}{n}) - I(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1}),$$

$$\vdots$$

$$I(\nu^{n-1}) - I(\nu^n) = I(0, \ldots, 0, \frac{1}{2}, \frac{1}{2}) - I(0, \ldots, 0, 1)$$

are all equal.
**Characterization**

**Theorem (Plata-Pérez et. al. 2015)**

Let $I : \mathbb{R}_+^n \setminus \{0\} \to \mathbb{R}$. Then $I$ equals the Gini index given by

$$G(x) = \frac{1}{n} \left[ n + 1 - 2 \frac{\sum_{i=1}^{n} (n + 1 - i)x_i^*}{\sum_{i=1}^{n} x_i^*} \right],$$

if and only if it satisfies the axioms of scale independence, symmetry, comonotone separability and standardization.
Proof idea

If \( I : \mathbb{R}_+^n \setminus \{0\} \to \mathbb{R} \) is any index satisfying the four axioms, then using the symmetry and scale independence axioms we have:

\[
I(x) = I(x^*) = I \left( \frac{x^*}{\sum_{j=1}^{n} x_j} \right), \quad x \in \mathbb{R}_+^n \setminus \{0\}.
\]

The key idea is to show that Axioms 4 and 3 characterize the Gini index on the convex subset

\[
K = \left\{ y \in \mathbb{R}_+^n : \sum_{i=1}^{n} y_i = 1 \text{ and } y_1 \leq y_2 \cdots \leq y_n \right\}.
\]
Robin Hood index

- also known as:
  - Hoover index (Edgar Malone Hoover jr., 1936),
  - Schutz (1951) [7]
  - Ricci index, (Umberto Ricci 1879–1946)
  - Pietra ratio, (G. Pietra)
  - Lindahl index (Erik Robert Lindahl 1891–1960),

- The share of total income which would have to be transferred from households above the mean to those below the mean to achieve a state of perfect equality in the distribution of incomes:

\[
H(x) = \frac{\sum_{i: x_i \geq \bar{x}} (x_i - \bar{x})}{\sum_{i=1}^{n} x_i}.
\]

- Higher values indicate more inequality and that more redistribution is needed to achieve income equality [6].
Robin Hood index – alternative formulas

\[ H(x) = \frac{\sum\{i: x_i \geq \bar{x}\} (x_i - \bar{x})}{\sum_{i=1}^{n} x_i}. \]

Since \[ \sum\{i: x_i \geq \bar{x}\} (x_i - \bar{x}) = -\sum\{i: x_i \leq \bar{x}\} (x_i - \bar{x}) \]:

\[ H(x) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{2 \sum_{i=1}^{n} x_i}. \quad (\ast) \]

Dividing nominator and denominator by \( n \):

\[ H(x) = \frac{1}{n} \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{2\bar{x}} = \frac{\text{MAD}(x)}{2\bar{x}}. \quad (\ast\ast) \]

Dividing by the total income inside the absolute value in (\( \ast \)):

\[ H(x) = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{x_i}{\sum_{i=1}^{n} x_i} - \frac{1}{n} \right|. \quad (\ast\ast\ast) \]
Robin Hood index – in $L^1$ space

- Let us think about $\mathbb{R}^n$ as the normed space $L^1$:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|.$$ 

- For $x \in \mathbb{R}^n_+ \setminus \{0\}$ we have $x_i \geq 0$, $i = 1, \ldots, n$, hence

$$\|x\|_1 = \sum_{i=1}^{n} x_i.$$ 

- Denoting by $\frac{1}{n} = (\frac{1}{n}, \ldots, \frac{1}{n}) \in \mathbb{R}^n$, we can rewrite $(***)$ as:

$$H(x) = \frac{1}{2} \left\| \frac{x}{\|x\|_1} - \frac{1}{n} \right\|_1.$$ 

- We can exploit the properties of the $L^1$ norm!
Robin Hood index and Lorenz curve

- For integrable random variables (referring to (**)):

  \[ H(X) = \frac{\mathbb{E}|X - \mathbb{E}(X)|}{2\mathbb{E}(X)} = \frac{1}{2} \mathbb{E} \left| \frac{X}{\mathbb{E}(X)} - 1 \right| . \]

- We can easily show that (\( L \) is a Lorenz curve)

  \[ \mathbb{E}|X - \mathbb{E}(X)| = 2\mathbb{E}(X)[F(\mathbb{E}X) - L(F(\mathbb{E}X))] \tag{*} \]

  hence

  \[ H(X) = F(\mathbb{E}X) - L(F(\mathbb{E}X)) \]

  (vertical distance between the Lorenz curve and the line of perfect equality at \( F(\mathbb{E}X) \)).
**Proof of (⋆)**

Let us denote $\mu = \mathbb{E}(X)$.

$$\mathbb{E}|X - \mu| = \int_{-\infty}^{\infty} |x - \mu| \, dF(x) =$$

$$= \int_{-\infty}^{\mu} (\mu - x) \, dF(x) + \int_{\mu}^{\infty} (x - \mu) \, dF(x) =$$

$$= \mu \int_{-\infty}^{\mu} \, dF(x) - \mu \int_{-\infty}^{\mu} x \, dF(x) + \int_{\mu}^{\infty} x \, dF(x) - \mu \int_{\mu}^{\infty} \, dF(x) =$$

$$= \mu F(\mu) - \mu \int_{-\infty}^{\mu} x \, dF(x) + (\mu - \mu \int_{-\infty}^{\mu} x \, dF(x)) - \mu (1 - F(\mu)) =$$

$$= 2\mu F(\mu) - 2 \mu \int_{-\infty}^{\mu} x \, dF(x) = 2\mu F(\mu) - 2\mu L(F(\mu)).$$

since $L(F(t)) = \int_{-\infty}^{t} x \, dF(x)/\mathbb{E}(X)$. 
Robin Hood index and Lorenz curve

- Assuming $F$ and $L$ are differentiable respectively at $t_0$ and $F(t_0)$, $F'(t_0) = f(t_0)$, the slope of tangent at $(F(t_0), L(F(t_0)))$ is equal:

  $$
  \lim_{t \to t_0} \frac{L(F(t)) - L(F(t_0))}{F(t) - F(t_0)} =
  \lim_{t \to t_0} \frac{\frac{1}{\mathbb{E}(X)} \int_{-\infty}^{t} x \, dF(x) - \frac{1}{\mathbb{E}(X)} \int_{-\infty}^{t_0} x \, dF(x)}{\int_{-\infty}^{t} dF(x) - \int_{-\infty}^{t_0} dF(x)} =
  \frac{1}{\mathbb{E}(X)} \lim_{t \to t_0} \frac{\int_{t_0}^{t} x \, dF(x)}{\int_{t_0}^{t} dF(x)} = \frac{1}{\mathbb{E}(X)} \cdot \frac{t_0 f(t_0)}{f(t_0)} = \frac{t_0}{\mathbb{E}(X)}.
  $$

- The slope of the tangent of the Lorenz curve equals 1 (tangent is parallel to the line of perfect equality) for $t_0 = \mathbb{E}(X)$.

- Lorenz curve is convex $\Rightarrow$ slope is increasing $\Rightarrow$ vertical distance between the Lorenz curve and the line of perfect equality is greatest at $F(\mathbb{E}(X))$. 
Robin Hood index and Lorenz curve

Robin Hood index can be graphically represented as the maximum vertical distance between the Lorenz curve and the 45-degree line that represents perfect equality of incomes.
Corollary: Robin Hood $\leq$ Gini

Let $q = F(\mathbb{E}(X)) = |AE|$. We have $H = |BF|$, $|BE| = q - H$.

$|AEB| = \frac{1}{2} q(q - H) = \frac{1}{2} (q^2 - Hq)$,

$|BCDE| = (1 - q)^{1 + \left(\frac{q - H}{2}\right)} = \frac{1}{2} (1 - q^2 - H + Hq)$,

$\int_0^1 L(t) \, dt \leq |AEB| + |BCDE| = \frac{1 - H}{2}$,

$G = \frac{1}{2} - \int_{\frac{1}{2}}^1 L(t) \, dt = 1 - 2 \int_0^1 L(t) \, dt \geq 1 - 2 \cdot \frac{1 - H}{2} = H$.

By-product: $H = 1 - 2(|AEB| + |BCDE|) = 2|ABC|$. 
Robin Hood vs Gini: min and max difference

max: $H = 0.5, G = 0.75$

min: $H = 0.5, G = 0.5$
Attempt to characterize Robin Hood index

Robin Hood index fulfils:
- scale invariance,
- symmetry,
- standardization.

Moreover (int-standardization):

\[ H(x) = \frac{\text{zeros}(x)}{n}, \quad \text{for } x \in J_n, \]

where \(\text{zeros}(x) = \#\{i : x_i = 0\}\) and

\[ J_n = \{(x_1, \ldots, x_n) : x_i = k_i/n, k_i \in \mathbb{N}, i = 1, \ldots, n, \sum_{i=1}^{n} k_i = n\}. \]
We say that $x, y \in \mathbb{R}_+^n \setminus \{0\}$ are co-mean if

$$(x_i - \bar{x})(y_i - \bar{y}) \geq 0, \quad i = 1, \ldots, n.$$ 

Axiom 3b: Robin Hood index is co-mean separable:

$$I(\beta x + (1 - \beta)y) = \beta I(x) + (1 - \beta)I(y)$$

for every $\beta \in [0, 1]$, every $x, y \in \mathbb{R}_+^n \setminus \{0\}$ that are co-mean and

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i.$$

**Hypothesis**

Function $I: \mathbb{R}_+^n \setminus \{0\} \rightarrow \mathbb{R}$ equals the Robin Hood index given by (****) if and only if it satisfies the axioms of scale independence, symmetry, co-mean separability and int-standardization.
Real data

Based on: OECD Income Distribution Database (IDD),

\[ H = 0.745876 G - 0.005591. \]
**Estimation from deciles**

From [5]: The Robin Hood index may also be calculated by:
- summing the percentage of income for each tenth of an income distribution where the percentage exceeds 10%,
- and subtracting from this the product of the number of tenths that meet this criterion times 10%.

**Example**

<table>
<thead>
<tr>
<th>Decile</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of T. income</td>
<td>1.08</td>
<td>2.48</td>
<td>4.13</td>
<td>5.74</td>
<td>7.33</td>
</tr>
<tr>
<td>Decile</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>% of T. income</td>
<td>8.97</td>
<td>10.83</td>
<td>13.09</td>
<td>16.41</td>
<td>29.93</td>
</tr>
</tbody>
</table>

In this case four deciles (7-10) exceed 10%, so the Robin Hood index

\[
\text{Robin Hood index} = (10.83\% + 13.09\% + 16.41\% + 29.93\%) - (4 \times 10\%) = 70.26\% - 40\% = 30.26\%.
\]
Thank you for your attention.


