

A simple proof of Foster's (1983) characterization of the Theil measure of inequality

Casilda LASSO DE LA VEGA

University of the Basque Country U.P.V./E.H.U., and BRIDGE research group

Av. Lehendakari Aguirre 85, 48015 Bilbao, Spain

e-mail: casilda.lassodelavega@ehu.es

Oscar VOLIJ

Ben-Gurion University of the Negev

P.O.B. 653, 84105 Beer Sheva, Israel

e-mail: oscar@volij.co.il

Abstract We provide a simple proof of Foster's (1983) characterization of the Theil inequality index.

Keywords Theil index, axiomatic characterization, income inequality.

JEL Classification D63

Acknowledgments We thank Mikel Bilbao for his disinterested help. We also thank the Spanish Ministerio de Educación y Ciencia (project ECO2012-31346) and the Gobierno Vasco (project IT568-13) for research support.

1. Introduction

Foster (1983) has shown that the Theil inequality index is the only (up to positive scalar multiplication) symmetric and homogeneous index that satisfies the Pigou-Dalton transfer principle and Theil decomposability. It is well known that symmetry, homogeneity and the Pigou-Dalton principle are necessary and sufficient conditions for an inequality index to be consistent with the Lorenz criterion. In other words, Foster's result states that Theil decomposability is the defining property of the Theil index among all the Lorenz consistent indices. A new and simple proof of the result is given. We first characterize the Theil index on the domain of distributions with rational entries, and then extend it to the domain of all distributions. While Foster's proof relies on an information theoretical result by Lee (1964), ours relies on a well-known characterization of the logarithmic function and uses ideas developed in Lasso de la Vega et al. (2013) characterization of Theil's inequality order.

2. Definitions

We follow Foster's (1983) notation. A distribution is a list $x = (x_1, \dots, x_n)$ of nonnegative real numbers, not all of them 0, where n is a positive integer. For any two distributions $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_m)$, $(x; y)$ denotes the *concatenated* distribution $(x_1, \dots, x_n, y_1, \dots, y_m)$. In particular, for any distribution $x = (x_1, \dots, x_n)$, the m -fold replication of x is the concatenation $(\underbrace{x; \dots; x}_m)$. For any distribution $x = (x_1, \dots, x_n)$, $|x|$ denotes the sum of its components, namely $|x| = \sum_{i=1}^n x_i$, and $\bar{x} = |x|/n$ denotes the average of its components. For any n , the distribution $(\underbrace{1, \dots, 1}_n)$ is denoted by u^n . The class of all distributions is denoted by D . An inequality measure is a function $I : D \rightarrow \mathfrak{R}$ that assigns a real number to each distribution in D . The Theil inequality measure (Theil (1967)) is the measure $T : D \rightarrow \mathfrak{R}$ defined by

$$T(y_1, \dots, y_n) = \sum_{k=1}^n \frac{y_k}{|y|} \ln \left(n \frac{y_k}{|y|} \right)$$

where we use the convention that $0 \ln 0 = 0$.

3. Axioms

The following axioms were used in Foster (1983).

Symmetry (S): The inequality measure I satisfies symmetry, if for any distribution

$$x = (x_1, \dots, x_n) \in D \text{ and any permutation } \pi: \{1, \dots, n\} \rightarrow \{1, \dots, n\}, I(x_1, \dots, x_n) = I(x_{\pi(1)}, \dots, x_{\pi(n)}).$$

Symmetry requires that inequality be unaffected by a reordering of the individuals' incomes.

Homogeneity (H): The inequality measure I satisfies homogeneity if whenever $y = \alpha x \in D$ for some distribution $x \in D$ and for some positive scalar α , $I(x) = I(y)$.

Homogeneity dictates that a proportionate change in incomes does not change inequality.

Population Principle (PP): The inequality measure I satisfies the population principle if for all $x \in D$, and $m=1, 2, \dots$, $I(x) = I(\underbrace{x; \dots; x}_m)$.

The population principle states that replicating a distribution leaves inequality unchanged.

We say that $y = (y_1, \dots, y_n)$ is obtained from $x = (x_1, \dots, x_n)$ by means of a regressive transfer if for some i and j , a) $x_i \leq x_j$, b) $x_i - y_i = y_j - x_j > 0$, and c) for all $k \neq i, j$, $x_k = y_k$.

Pigou-Dalton Transfer Principle (PD): The inequality measure I satisfies the Pigou-Dalton transfer principle if $I(x) < I(y)$ whenever $y \in D$ is obtained from $x \in D$ by means of a regressive transfer.

The Pigou-Dalton transfer principle states that an income transfer from an individual to a richer one increases inequality. It readily implies the next two axioms.

Directedness (D): The inequality measure I satisfies directedness if for any $x = (x_1, \dots, x_n) \in D$, $I(\bar{x}u^n) \leq I(x)$.

Directedness requires that redistributing total income equally among the individuals cannot increase inequality. We will also use strong directedness which additionally that one such redistribution actually decreases inequality.

Strong Directedness (SD): The inequality measure I satisfies strong directedness, if it satisfies directedness and in addition, $I(1,1) < I(0,2)$.

All the properties presented so far are purely ordinal axioms. Symmetry, homogeneity and the population principle state that some transformations on distributions do not affect inequality. Directedness, Strong Directedness and the Pigou-Dalton principle require, on the other hand, that some specific redistributions of income do affect inequality in a particular way. The next axioms, although having an ordinal content, are cardinal in nature.

Normalization: The inequality measure I is normalized if for all $x = (x_1, \dots, x_n) \in D$, $I(\bar{x}u^n) = 0$.

Normalization requires that all uniform distributions be assigned an inequality index of 0.

Theil Decomposability (TD): The inequality measure I satisfies Theil decomposability if for any pair of distributions $x^1 = (x_1^1, \dots, x_{n_1}^1)$, $x^2 = (x_1^2, \dots, x_{n_2}^2) \in D$,

$$I(x^1; x^2) = \frac{|x^1|}{|x|} I(x^1) + \frac{|x^2|}{|x|} I(x^2) + I(\bar{x}^1 u^{n_1}; \bar{x}^2 u^{n_2}).$$

When a distribution is a concatenation of two distributions Theil decomposability allows us to decompose its inequality into two components. One is the inequality *within* the sub-distributions, and the other is the inequality *between* them. The inequality within sub-distributions is a weighted average of the inequality of each of them, the weights being the fraction of their respective incomes out of the total income.

This decomposability cannot be applied when one of the sub-distributions is a vector of zeroes, since such a vector is not a distribution. We replace Theil decomposability by the following axiom which generalizes it in a way that allows for any partition of a distribution into two vectors.

Generalized Theil Decomposability (GTD): The inequality measure I satisfies generalized Theil decomposability if for any distribution $x \in D$, and for any partition $x = (x^1; x^2)$ of it where $x^1 \in D$,

$$I(x) = \frac{|x^1|}{|x|} I(x^1) + I(\bar{x}^1 u^{n_1}; x^2),$$

where n_1 is the number of components of x^1 .

Clearly, generalized Theil decomposability and symmetry imply Theil decomposability.

Furthermore, by induction, generalized Theil decomposability implies that for any distribution $x \in D$, and for any partition $x = (x^1; x^2; \dots; x^K)$ of it where $x^1, \dots, x^{k_0} \in D$,

$$I(x) = \sum_{k=1}^{k_0} \frac{|x^k|}{|x|} I(x^k) + I(\bar{x}^1 u^{n_1}; \dots; \bar{x}^{k_0} u^{n_{k_0}}; x^{k_0+1}; \dots; x^K) \quad (1)$$

and for each $k = 1, \dots, K$, n_k is the number of components of x^k .

4. A characterization of Theil's measure

Foster (1983) shows that Theil decomposability implies the population principle. For convenience we repeat the proof.

Claim 1: If I satisfies Theil decomposability then I satisfies normalization and the population principle.

Proof: Let $x = (x_1, \dots, x_n) \in D$. By TD, $I(\bar{x}u^{2n}) = 2 \frac{I(\bar{x}u^n)}{2} + I(\bar{x}u^{2n})$. Therefore, $I(\bar{x}u^n) = 0$.

By TD, $I(\underbrace{x; x; \dots; x}_m) = m \frac{I(x)}{m} + I(\bar{x}u^{mm})$. By the previous result, $I(\bar{x}u^{mm}) = 0$, which implies the desired result.

We first characterize the Theil index on the class of distributions $D_Q \subset D$ with rational entries.

Theorem 1: An inequality measure $I : D_Q \rightarrow \mathfrak{R}$ satisfies symmetry, homogeneity, strong directedness, and generalized Theil decomposability if and only if I is a positive multiple of the Theil measure of inequality.

The proof makes use of the following technical result, which we show in the appendix. In what follows, Q denotes the set of rational numbers.

Proposition 1: Let $\Omega = Q \cap (0,1]$ and let $h : \Omega \rightarrow \mathfrak{R}$ be a function such that for all $\alpha, \beta \in \Omega$

1. $h(\alpha) \geq 0$.
2. $h(\alpha\beta) = h(\alpha) + h(\beta)$.
3. $h(1/2) > 0$.

Then, there is $c > 0$ such that $h(\cdot) = -c \ln(\cdot)$.

Proof of Theorem 1: It is well known that the Theil index satisfies S, H, SD, and GTD.

Therefore, we will only show that it is the only one which does so. Let $I : D_Q \rightarrow \mathfrak{R}$ be an index that satisfies S, H, SD, and GTD.

For any two natural numbers $0 < m \leq n$, define $v_{n,m} = (\underbrace{0, \dots, 0}_{n-m}, \underbrace{1, \dots, 1}_m)$. Also, let

$h : Q \cap (0,1] \rightarrow \mathfrak{R}$ be the function defined by $h(\alpha) = I(v_{n,m})$, where $\alpha = n/m$. Note that h is well-defined. For if $\alpha = m/n = m'/n'$, then by PP $I(v_{n,m}) = I(v_{n'n,n'm}) = I(v_{m'm,m'}) = I(v_{n',m'})$.

Note that by Claim 1 $I(\bar{v}_{n,m} u^n) = 0$. Therefore, by directedness, $h(\alpha) \geq 0$. Furthermore, by strong directedness, $I(0,2) > 0$, which, by homogeneity, implies $h(1/2) > 0$.

Let $\alpha, \beta \in Q \cap (0,1]$. If we show that $h(\alpha\beta) = h(\alpha) + h(\beta)$ then by Proposition 1,

$h(\cdot) = -c \ln(\cdot)$ for some positive c .

Let $\alpha = m_1/n_1$ and $\beta = m_2/n_2$. We get

$$\begin{aligned}
h(\alpha\beta) &= I(v_{n_1 n_2, m_1 m_2}) \\
&= I(\underbrace{0, \dots, 0}_{n_1 n_2 - m_1 m_2}, \underbrace{1, \dots, 1}_{m_1 m_2}) \\
&= I(\underbrace{0, \dots, 0}_{(n_1 - m_1) m_2}, \underbrace{1, \dots, 1}_{m_1 m_2}, \underbrace{0, \dots, 0}_{n_1 n_2 - n_1 m_2}) && \text{by symmetry} \\
&= I(\underbrace{0, \dots, 0}_{(n_1 - m_1) m_2}, \underbrace{1, \dots, 1}_{m_1 m_2}) + I(\underbrace{m_1 / n_1, \dots, m_1 / n_1}_{n_1 m_2}, \underbrace{0, \dots, 0}_{n_1 n_2 - n_1 m_2}) && \text{by GTD} \\
&= I(\underbrace{0, \dots, 0}_{n_1 - m_1}, \underbrace{1, \dots, 1}_{m_1}) + I(\underbrace{m_1 / n_1, \dots, m_1 / n_1}_{m_2}, \underbrace{0, \dots, 0}_{n_2 - m_2}) && \text{by PP} \\
&= I(\underbrace{0, \dots, 0}_{n_1 - m_1}, \underbrace{1, \dots, 1}_{m_1}) + I(\underbrace{1, \dots, 1}_{m_2}, \underbrace{0, \dots, 0}_{n_2 - m_2}) && \text{by homogeneity} \\
&= I(v_{n_1, m_1}) + I(v_{n_2, m_2}) \\
&= h(\alpha) + h(\beta).
\end{aligned}$$

This shows that h satisfies all the conditions of Proposition 1 and therefore it is a positive multiple of the logarithmic function.

Now let $y = (y_1, \dots, y_n) \in D_Q$ be a distribution. By homogeneity we can assume without loss of generality that all the incomes are integer numbers. By symmetry, we can assume without loss of generality that there is an index $1 \leq n_0 \leq n$ such that $y_i > 0$ if and only if $i \leq n_0$.

Define

$$\begin{aligned}
v &= (v_{|y|, y_1}; \dots; v_{|y|, y_n}) \\
&= (\underbrace{(0, \dots, 0, 1, \dots, 1)}_{|y| - y_1}; \dots; \underbrace{(0, \dots, 0, 1, \dots, 1)}_{|y| - y_n})
\end{aligned}$$

By equation (1)

$$\begin{aligned}
I(v) &= \sum_{k=1}^{n_0} \frac{y_k}{|y|} I(v_{|y|, y_k}) + I\left(\frac{y_1}{|y|} u^{|y|}; \dots; \frac{y_{n_0}}{|y|} u^{|y|}; \mathbf{0} u^{|y|}; \dots; \mathbf{0} u^{|y|}\right) \\
&= \sum_{k=1}^{n_0} \frac{y_k}{|y|} I(v_{|y|, y_k}) + I\left(\frac{y_1}{|y|} u^{|y|}; \dots; \frac{y_n}{|y|} u^{|y|}\right)
\end{aligned}$$

But by S, PP, and H, $I\left(\frac{y_1}{|y|}u^{|y|}; \dots; \frac{y_n}{|y|}u^{|y|}\right) = I(y)$. Therefore, we can write the previous equations as

$$I(y) = I(v) - \sum_{k=1}^{n_0} \frac{y_k}{|y|} I(v_{|y|, y_k}).$$

Note that by symmetry, $I(v) = I(\underbrace{0, \dots, 0}_{(n-1)|y|}, \underbrace{1, \dots, 1}_{|y|}) = I(v_{n|y|, |y|}) = c \ln(n)$. Also note that for $k \leq n_0$,

$I(v_{|y|, y_k}) = c \ln(|y| / y_k)$. As a result we obtain

$$\begin{aligned} I(y) &= c \ln(n) - \sum_{k=1}^{n_0} \frac{y_k}{|y|} c \ln\left(\frac{y_k}{|y|}\right) \\ &= c \ln(n) - \sum_{k=1}^n \frac{y_k}{|y|} c \ln\left(\frac{y_k}{|y|}\right) \\ &= \sum_{k=1}^n \frac{y_k}{|y|} c \ln\left(n \frac{y_k}{|y|}\right) \\ &= cT(n). \end{aligned}$$

Q.E.D.

We now restate Foster's (1983) main result, using the generalized version of Theil decomposability, and apply Theorem 1 to prove it.

Theorem 2: An inequality measure $I : D \rightarrow \mathfrak{R}$ satisfies symmetry, homogeneity, the Pigou-Dalton transfer principle, and generalized Theil decomposability if and only if I is a positive multiple of the Theil measure of inequality.

Proof: Since it is known that T satisfies the axioms enunciated in the theorem, we show only the uniqueness part. So let $I : D \rightarrow \mathfrak{R}$ be an index that satisfies S, H, PD and GTD. By Theorem 1,

$$I(x) = T(x) \text{ for all } x \in D_Q \tag{2}$$

Let $x = (x_1, \dots, x_n) \in D$. We need to show that $I(x) = T(x)$. We will show this by induction on the number of distinct positive income levels in x . By S we can assume, without loss of

generality, that $x_1 \geq \dots \geq x_n$. If x has only one distinct positive income level, then

$x = (\alpha, \alpha, \dots, \alpha, 0, \dots, 0) = (\alpha u^{n_1}; 0u^{n-n_1})$ with $\alpha > 0$, where n_1 is the number of entries of x that are equal to α . Using H and equation (2),

$$I(x) = I(\alpha u^{n_1}; 0u^{n-n_1}) = I(u^{n_1}; 0u^{n-n_1}) = T(u^{n_1}, 0u^{n-n_1}) = T(\alpha u^{n_1}; 0u^{n-n_1}) = T(x).$$

If x has exactly two distinct positive income levels, then $x = (\alpha u^{n_1}; \beta u^{n_2}; 0u^{n-n_1-n_2})$ with $\alpha > \beta > 0$, where n_1 and n_2 are the numbers of entries of x that are equal to α and β , respectively. Assume by contradiction that $I(x) > T(x)$ (the case $I(x) < T(x)$ is similar and is left to the reader). By homogeneity of I and T we can assume, without loss of generality, that $\beta = 1$. By continuity of the Theil index and since it satisfies homogeneity and PD, there exists $p \in Q$, with $p > \alpha$, such that

$$I(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) > T(pu^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) > T(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}).$$

By equation (2), $T(pu^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) = I(pu^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2})$. Therefore

$$I(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) > I(pu^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) \quad (3)$$

Note that the total income of distribution $(pu^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2})$ is higher than that of distribution $(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2})$. The ratio of these two incomes is $\delta = \frac{\alpha n_1 + n_2}{pn_1 + n_2}$. By homogeneity equation

(3) can be rewritten

$$I(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2}) > I(\delta pu^{n_1}; \delta 1u^{n_2}; 0u^{n-n_1-n_2})$$

so that the two distributions having equal incomes. Since $\delta < 1$ (and $\delta p > \alpha$),

$(\delta pu^{n_1}; \delta 1u^{n_2}; 0u^{n-n_1-n_2})$ is obtained from $(\alpha u^{n_1}; 1u^{n_2}; 0u^{n-n_1-n_2})$ by a finite sequence of regressive transfers, which contradicts the fact that I satisfies PD. Consequently, $I(x) = T(x)$ whenever x has exactly two distinct positive entries.

Let $k > 2$ and assume now that $I(y) = T(y)$ whenever y has less than k distinct positive income levels. Further assume that $x = (x_1, \dots, x_n) \in D$ has exactly k distinct positive income levels.

Partition x into two sub-distributions, $x^1 = (x_1, \dots, x_{n_0}) \in D$ and $x^2 = (x_{n_0+1}, \dots, x_n) \in D$, such that each one of them has strictly less than k positive income levels. Since $k > 2$, this can be done.

Note that $\bar{x}^1, \bar{x}^2 > 0$. By TD

$$I(x_1, \dots, x_n) = \frac{|x^1|}{|x|} I(x^1) + \frac{|x^2|}{|x|} I(x^2) + I(\bar{x}^1 u^{n_0}; \bar{x}^2 u^{n-n_0})$$

By the induction hypothesis $I(x^1) = T(x^1)$ and $I(x^2) = T(x^2)$ and

$$I(\bar{x}^1 u^{n_0}; \bar{x}^2 u^{n-n_0}) = T(\bar{x}^1 u^{n_0}; \bar{x}^2 u^{n-n_0}).$$

Therefore, by the Theil decomposability of the Theil index,

$$I(x_1, \dots, x_n) = \frac{|x^1|}{|x|} T(x^1) + \frac{|x^2|}{|x|} T(x^2) + T(\bar{x}^1 u^{n_0}, \bar{x}^2 u^{n-n_0}) = T(x_1, \dots, x_n).$$

Q.E.D.

5. Appendix

Proof of Proposition 1: It is known that any function defined on $(0,1]$ that satisfies conditions 1, 2 and 3 must be a positive multiple of the logarithmic function (see Theorem 0.2.5 in Aczél and Daróczy (1975)). Therefore it is enough to show that h can be extended to the interval $(0,1]$ in a way that preserves the above conditions.

We first show that h is a monotonic non-increasing function. Indeed, if $\alpha, \beta \in \Omega$ are such that

$$\alpha < \beta, \text{ then by conditions 1 and 2, } h(\alpha) = h\left(\beta \frac{\alpha}{\beta}\right) = h(\beta) + h\left(\frac{\alpha}{\beta}\right) \geq h(\beta).$$

We now extend the function h to the whole interval $(0,1]$. Define the function $\hat{h}: (0,1] \rightarrow \mathfrak{R}$ so that for any x , $\hat{h}(x) = \lim_{\substack{r \in \Omega \\ r \rightarrow x^-}} h(r)$. Since h is a monotonic non-increasing function, this limit is

well-defined. It can be seen that \hat{h} is an extension of h to $(0,1]$ that satisfies conditions 1, 2 and 3 above on $(0,1]$.

6. References

1. Aczél, J. and Daróczy, Z., 1975. On measures of information and their characterizations. Academic Press, New York.
2. Foster, J. E., 1983. An axiomatic characterization of the Theil measure of income inequality. *J. Econ. Theory.* 31, 105-121.
3. Lasso de la Vega, C., Urrutia, A., Volij, O., 2013. An axiomatic characterization of the Theil inequality ordering. *Econ. Theory.* DOI 10.1007/s00199-012-0739-6.
4. Lee, P. M., 1964. On the axioms of information theory. *Ann. of Math. Statist.* 35, 415-418.
5. Theil, H., 1967. *Economics and information theory.* North Holland, Amsterdam.