Economics Letters 186 (2020) 108789

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Inequality as experienced difference: A reformulation of the Gini coefficient

Samuel Bowles [a], Wendy Carlin [a,b,c,*]

[a] Behavioral Sciences, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, 87501, USA
[b] Department of Economics, University College London, Gower St, London, WC1E 6BT, UK
[c] CEPR, 33 Great Sutton Street, London EC1V 0DX, UK

ARTICLE INFO

Article history:
Received 12 August 2019
Accepted 21 October 2019
Available online 2 November 2019

JEL classification:
D63
I31
L14

Keywords:
Inequality
Networks
Social interactions
Lorenz curve

ABSTRACT

Inequality is typically measured as the degree of dispersion of a distribution of individual attributes, say, wealth, as is captured for example by the Lorenz curve, and its associated statistic, the Gini coefficient. But both the economics and social psychology of experienced inequality are better expressed by differences between an individual and others. There is a natural way to do this using the standard definition of the Gini coefficient as one half the mean difference among individuals, relative to the population mean wealth. Here we show that reformulating the Gini coefficient as a measure of experienced inequality on a complete social network yields a computational algorithm that, unlike the conventional one, is consistent with this definition and irrespective of population size varies from 0 (no differences among individuals) to 1 (one individual owns all the wealth). Our proposed measure also avoids a downward bias in the standard algorithm, which for small populations can be substantial. Because social networks are far from complete, the pairwise comparisons based on social interactions in which people routinely engage may support a level of experienced inequality that either exceeds or falls short of the Gini coefficient measured on a hypothetical complete network. We illustrate this fact with empirical estimates for a farming community in Nicaragua.

Researchers find the Gini coefficient an appealing measure of inequality for two reasons. First, like the coefficient of variation and unlike the Lorenz curve and many other measures such as the Atkinson index, it provides a single measure of the degree of inequality based on the entire distribution and allows easy comparisons across nations and time periods (Damgaard, 2018; Dasgupta et al., 1973; Gini, 1914; Kendall and Stuart, 1969; Sen, 1997; Yitzhaki and Schechtman, 2013).

Second, what the Gini coefficient measures can be connected in a Veblen-inspired intuitive way to show how we experience inequality, that is by pairwise comparison of one's own wealth or income with that of someone else. This is because the Gini coefficient, say, for wealth for example, is the mean difference in wealth between all pairs in the population, divided by the mean wealth, multiplied by one-half so that the measure varies (as Gini specified that it should and as shown by Deaton (1997)) from zero, if there are no differences among population members to one if a single individual owns all of the wealth (Gini, 1914).

There are many questions for which a single scalar measure of inequality is insufficient. The Gini coefficient, for example, fails to capture both important dimensions of economic class polarization (Esteban and Ray, 1994) and the expected positive relationship between greater wealth inequality and the extent of polygynous marriage (Ross et al., 2018). But the more complex multi-dimensional measures of inequality adequate for these and similar questions do not allow either the simple intuitions underlying the Gini coefficient or easily interpretable comparisons across time and place.

The appeal of the Gini coefficient is that it is as Sen writes "a very direct measure of income differences, taking note of differences between every pair of incomes" (Sen, 1997). To capture the intuition that the Gini coefficient measures experienced pairwise differences, we represent a population as a complete undirected network. The fundamental data on experienced disparities, as shown in the left panel of Fig. 1, are the edges of the network, that is, the pairwise differences in wealth, not the nodes.

The conventional algorithm (and the equivalent Lorenz curve-based geometrical measure) is illustrated by the right panel of Fig. 1, where all of the nine edges (represented by the arrows) represent "differences" in wealth, including the by-definition zero "self on self" differences. If there are n members of the population then the total number of unique non-identical pairs relevant to the study of inequality is $(n^2 - n)/2$, shown as the three edges in the left panel not the n^2 "pairs" on the right.

* Corresponding author.
E-mail addresses: samuel.bowles@gmail.com (S. Bowles), w.carlin@ucl.ac.uk (W. Carlin).

https://doi.org/10.1016/j.econlet.2019.108789

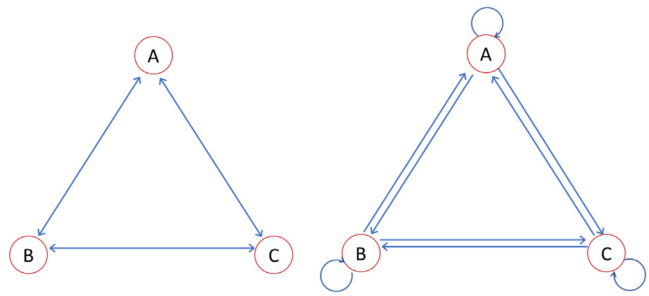

Fig. 1. Experienced differences (left panel) and the edges used in the conventional Lorenz curve-based measure (right panel).

For finite populations (the only relevant case), the conventional algorithm is a systematically downward-biased approximation. If the nodes A, B, and C in Fig. 1 have wealth 10, 4, and 3, for example, the Gini coefficient using the network representation in the left panel is 0.412, which is what one finds by using Eq. (3) below. Using the network representation on the right (that is, Eq. (1), below), however, the Gini is estimated as 0.274 which (as we will see in Eq. (2)) needs to be multiplied by $n/(n\text{ -}1)$ or 1.5 to get the Gini coefficient based only on the differences among actual pairs in the population (excluding the three "self-on-self" zero differences).

We call the Gini coefficient for the network on the left the difference-based Gini. The measure for the network on the right is equal to the area between the Lorenz curve and the perfect equality line divided by the total area under the perfect equality line, so we term it the Lorenz-based Gini coefficient. We denote the latter measure, G^L, and the difference-based measure, G, without superscript as this is the unbiased measure of the correct quantity – namely differences among population pairs – of which G^L is an approximation.

The Lorenz-based Gini coefficient is less than one when a single individual owns all of the wealth. The reason is that, as the figure illustrates, despite its common description in these terms, the algorithm does not compute the average differences between all pairs of members in the population relative to the population mean.

The Lorenz-based coefficient when applied to some quantity, $\underline{y}$, gives

$$G^L = \frac{\sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \left|y_i - y_j\right|}{2n^2\underline{y}} \tag{1}$$

which for the population in Fig. 1 counts all nine arrows in the right panel (including the three "self-on-self" edges with a zero difference by definition, namely $|y_i - y_i|$) as indicators of the disparity in question. Using this Lorenz-based conventional computation, a population size adjustment is required so that this value will vary over the unit interval. Then the difference-based Gini coefficient (Yitzhaki and Schechtman, 2013) is:

$$G = G^L \frac{n}{n-1} \tag{2}$$

Dasgupta, Sen and Starrett noted "there is some ambiguity as to how we define Lorenz curves in the discrete case" (Dasgupta et al., 1973). For finite rather than continuous populations Gini (correctly) proposed that both the "perfect equality line" and the Lorenz curve are represented by a step function (Gini, 1914). Remembering that one must both use step functions in the geometrical treatment and multiply G^L by $n/(n-1)$ to get the correct measure seems like an unnecessary complication, which, we will show, can be avoided by applying a consistent definition of societal differences to measure G directly.

The population size bias in the unadjusted measure is substantial when the population under consideration is small. Recent studies of wealth inequality in Western Eurasia from the beginning of the Neolithic to the end of the Bronze age, for example, demonstrate that Gini coefficients estimated from hypothetically simulated small data sets are quite precise. But the bias associated with use of the Lorenz-based measure (G^L) is substantial enough to affect qualitative conclusions when comparisons are made across data sets of substantially differing size. For observations prior to 3000 BCE the Lorenz based measure understates the difference-based Gini by sixteen percent; while for later populations the bias is only 2 percent (Bogaard et al., 2019; Fochesato et al., 2019).

However, the main reason to substitute G for G^L is not the "small numbers bias" but the fact that G^L defined in Eq. (1) simply is not a measure of inequality, if by this one means, differences between individuals, because it includes the fictitious self-on-self zero differences. Using the unique non-identical pairing setup in the left panel of Fig. 1 we have an intuitive definition of G which measures inequality among actual pairs in the population, and, which, as a result, does not require a population size adjustment.

We first define the sum of the absolute differences among the (unique non-identical) pairs, which we call $\varDelta$, or

$$\varDelta \equiv \sum_{i=j+1}^{i=n} \sum_{j=1}^{j=n-1} \left|y_i - y_j\right|$$

The difference from Eq. (1) is that here the $\left|y_i - y_j\right|$ terms count only the three edges in the left panel of Fig. 1. And the difference-based Gini coefficient, then, is

$$G = \frac{\varDelta}{\left(\frac{n(n-1)}{2}\right)} \frac{1}{\underline{y}} \frac{1}{2} = \frac{\varDelta}{n(n-1)} \frac{1}{\underline{y}} \tag{3}$$

from which we can see that it is the mean difference among all pairs in the population (the first term in the expression in the middle) divided by the mean value of y, giving us the "relative mean difference" times one half.

This is what we consider to be the true Gini coefficient (mentioned in passing by Deaton (1997)) and is identical to the corrected Gini coefficient in Eq. (2) which, noting that the sum of differences from Eq. (1) is just $2\varDelta$, can be seen from:

$$\begin{aligned} G^L \left(\frac{n}{n-1}\right) &= \frac{\varDelta}{n^2} \frac{1}{\underline{y}} \left(\frac{n}{n-1}\right) = \frac{\varDelta}{n} \frac{1}{\underline{y}} \left(\frac{1}{n-1}\right) \\ &= \frac{\varDelta}{n(n-1)} \frac{1}{\underline{y}} = G \end{aligned} \tag{4}$$

Using Eq. (3), when a single individual owns all of the wealth, $G = 1$ independently of population size. To see this, suppose that a single individual in a group of finite size n owns the entire wealth which is y. Then all the paired differences are zeros except the $(n - 1)$ edges connecting the "have nots" with the single "have". The sum of the differences on these edges is thus $y(n-1) = \varDelta$. Inserting this value in Eq. (3) confirms that G $= 1$ for any $n > 1$.

One way to see why the Lorenz-based measure understates inequality is that in the algorithm, a zero can mean either that there are two individuals with identical wealth (relevant to the measurement of inequality), or that the observation is based on comparing a single individual's wealth with itself (not relevant). The idea of counting a person "paired" with herself is mathematically familiar – "difference" here is the expected absolute difference between two individuals randomly drawn (with replacement) from a population – but is not applicable to the study of interpersonal differences in wealth or income.

Another clue that something is wrong with G^L is that it is invariant to population replication, which occurs because of the fictive zeros mentioned above. But experienced inequality clearly decreases with replication because it increases the fraction of the pairs in the population in which wealth differences really are absent. Consider an initial population of n with the Gini coefficient G_1 and construct a new population consisting of k (a positive integer greater than one) copies of the initial population with a conventionally measured Gini coefficient of G_k^L. By invariance, $G_k^L = G_1^L$, and from Eq. (2),

$$G_k^L = \frac{n-1}{n} G_1$$

So, for the new population of kn individuals we have

$$G_k = G_k^L \frac{kn}{kn-1} = G_1 \frac{n-1}{n} \frac{kn}{kn-1} = G_1 \frac{k(n-1)}{kn-1} < G_1 \tag{5}$$

which shows that replication reduces difference-based inequality. Replication, by increasing population size, reduces the small population size downward bias of G^L and this exactly offsets the real decrease in the difference-based measure, as shown in Eq. (5).

Some of the dimensions along which inequality is measured are best conceived as individual attributes, of which members of a population simply have more or less, like height. But other dimensions are best conceived of as differences between individuals in their relationships with others (Nettle, 2015). On both descriptive and normative grounds, we think that economic inequalities are in this latter class. We have provided a method of calculating the Gini coefficient motivated by intuitions about inequality as experienced by members of a population, and which, in so doing, incidentally avoids the need for a population size adjustment.

Because it is based on a social network, the difference-based measure can readily be extended to better capture the sociological and psychological dimensions of inequality that arise from the comparisons people make with others whom they routinely encounter. Using social network data we can provide a Veblen-inspired measure of what we call experienced inequality (Veblen, 1934). Eq. (3) measures the degree of experienced inequality on a complete network, which will differ from experienced inequality on empirically observed often highly segregated social networks (Jackson et al., 2017).

Consider, for example, the effect of a change in the relevant social relationships from a complete network characteristic of many hunter-gatherer economies to a star with the wealthiest person at the center as in what anthropologists call a "big man" system or to a bipartite two-class network characteristic of farming economies with economic classes based on land ownership. Assuming no change in individual wealth, experienced inequality would have increased. To illustrate, if the three individuals in the complete network of the left panel of Fig. 1 were instead represented by a line with the richer person in the center—a landlord, for example with two isolated sharecroppers—then, with no change in the wealth of the three individuals, the experienced inequality would rise from 0.41 to 0.57.

As an illustration we have used empirically estimated social networks to assay the experience of inequality in a community of 25 horticultural households in lowland Nicaragua (Koster and Leckie, 2014). If we confine pairwise comparisons to households that are connected in the social network established on the basis of ethnographic evidence, the Gini coefficient for material wealth is 0.65, considerably in excess of the measure estimated on the entire hypothetical complete network (0.56). This occurs because households tend to interact more with other households of differing wealth than would occur by chance.

This opens up the broader question of the adequacy of the Gini coefficient and other indices as measures of inequality, and the potential for extensions of the mean difference approach to take account of social network structure and the experience of inequality.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the Behavioral Sciences Program and the staff of the library at the Santa Fe Insitute for their support of this work and for help from Richard Blundell, Antonio Cabrales, Weikai Chen, Mattia Fochesato, Arjun Jayadev, Jeremy Koster, Christopher Lee, Sai Madhurika, Costas Meghir, Suresh Naidu, Ian Preston, John Roemer, Margaret Stevens, and Bob Rowthorn. We also thank the students of Economics 1001 at UCL, whose questions prompted us to do this research.

References

Bogaard, A., Fochesato, M., Bowles, S., 2019. The farming-inequality nexus: New insights from Western Eurasia. Antiquity 93 (371), 1129–1143.
Damgaard, Christian, 2018. Gini Coefficient. A Wolfram Web Resource, Mathworld.
Dasgupta, Partha, Sen, Amartya, Starrett, David, 1973. Notes on the measurement of inequality. J. Econom. Theory 6, 180–187.
Deaton, Angus, 1997. The Analysis of Household Surveys: A Microeconomic Approach to Development Policy. World Bank, Washington, D.C.
Esteban, Joan, Ray, Debraj, 1994. On the measurement of polarization. Econometrica 62, 819–852.
Fochesato, M., Bogaard, A., Bowles, S., 2019. Measuring ancient inequality: The challenges of comparability, bias, and precision. Antiquity 93 (370), 853–859.
Gini, Corrado, 1914. Sulla misura della concentrazione e della variabilita de caratteri. In: Atti Del Reale Istituto Venneto di Schienze, Lettere ed Arti, LXXIII(Parte Seconda). pp. 1203–1248.
Jackson, Matthew, Bryan, Rogers, Zenou, Yves, 2017. The economic consequences of social network structure. J. Econ. Lit. 55 (1), 49–95.
Kendall, Maurice G., Stuart, Alan, 1969. The Advanced Theory of Statistics. Charles Griffin, London.
Koster, Jeremy, Leckie, George, 2014. Food sharing networks in lowland nicaragua: An application of the social relations model to count data. Social Networks 39, 100–110.
Nettle, Daniel, 2015. Tyneside Neighbourhoods: Deprivation, Social Life, and Social Behaviour in One British City. Open Book Publishers, Cambridge.
Ross, C.T., Borgerhoff Mulder, M., Oh, S.-Y., Bowles, S., Beheim, B., Bunce, J., Caudell, M., Clark, G., Colleran, H., Cortez, C., Draper, P., Greaves, R.D., Gurven, M., Headland, T., Headland, J., Hill, K., Hewlett, B., Kaplan, H.S., Koster, J., Kramer, K., Marlowe, F., McElreath, R., Nolin, D., Quinlan, M., Quinlan, R., Revilla-Minaya, C., Scelza, B., Schacht, R., Shenk, M., Uehara, R., Voland, E., Willfuehr, K., Winterhalder, B., Ziker, J., 2018. Greater wealth inequality, less polygyny: Rethinking the polygyny threshold model. J. R. Soc. Interface 15 (144).
Sen, Amartya K., 1997. On Economic Inequality. Clarendon Press, Oxford.
Veblen, Thorsten, 1934. The Theory of the Leisure Class. Modern Library (first published in 1899), New York.
Yitzhaki, S., Schechtman, E., 2013. The Gini Methodology: A Primer on a Statistical Methodology. Springer, New York.