DISCUSSION PAPER SERIES

DP13211

INEQUALITY AS EXPERIENCED DIFFERENCE: A REFORMULATION OF THE GINI COEFFICIENT

Samuel Bowles and Wendy Carlin

DEVELOPMENT ECONOMICS AND PUBLIC ECONOMICS

ISSN 0265-8003

INEQUALITY AS EXPERIENCED DIFFERENCE: A REFORMULATION OF THE GINI COEFFICIENT

Samuel Bowles and Wendy Carlin

Discussion Paper DP13211
Published 29 September 2018
Submitted 16 September 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

INEQUALITY AS EXPERIENCED DIFFERENCE: A REFORMULATION OF THE GINI COEFFICIENT

Abstract

We represent a population as a complete undirected network, the edges of which are the fundamental data on experienced disparities. This yields a Gini coefficient (for wealth, say) for finite populations that is based on the mean wealth difference between all pairs of individuals relative to the mean wealth, which we demonstrate is not the case for the conventional Lorenz curve representation and the algorithm widely-used to calculate it. Our method also provides simple and intuitive explanations of the effects on the Gini coefficient of changes in the wage share, the employment rate, and other macroeconomic and demographic variables.

JEL Classification: D31

Keywords: inequality, Gini coefficient, relative mean difference, Lorenz curve

Samuel Bowles - samuel.bowles@gmail.com
Santa Fe Institute

Wendy Carlin - w.carlin@ucl.ac.uk
Department of Economics, UCL and CEPR and CEPR

Acknowledgements
We are grateful to the Behavioral Sciences Program at SFI for their support of this work and for help from Antonio Cabrales, Weikai Chen, Arjun Jayadev, Christopher Lee, Suresh Naidu, Ian Preston, John Roemer, Margaret Stevens, and Bob Rowthorn.

INEQUALITY AS EXPERIENCED DIFFERENCE: A REFORMULATION OF THE GINI COEFFICIENT

Samuel Bowles and Wendy Carlin[1]
12 September 2018

Abstract. We represent a population as a complete undirected network, the edges of which are the fundamental data on experienced disparities. This yields a Gini coefficient (for wealth, say) for finite populations that is based on the mean wealth difference between all pairs of individuals relative to the mean wealth, which we demonstrate is not the case for the conventional Lorenz curve representation and the algorithm widely-used to calculate it. Our method also provides simple and intuitive explanations of the effects on the Gini coefficient of changes in the wage share, the employment rate, and other macroeconomic and demographic variables.

Keywords: inequality, Gini coefficient, relative mean difference, Lorenz curve

JEL codes: D31 (income, wealth and their distributions)

1. EXPERIENCED DISPARITIES AND THE MEASUREMENT OF INEQUALITY.

Some of the dimensions along which we measure inequality are best conceived of as individual attributes, of which members of a population simply have more or less, like height. But on both descriptive and normative grounds, other dimensions are more intuitively represented as differences between individuals. Economic inequalities – in wealth or income, for example – are in this latter class.

A natural way to measure the extent of these disparities among individuals is to treat the economy as a complete undirected network as shown in the left panel of Figure 1, where the weight on each of the three edges of which is the disparity in wealth (from here, taken to be the quantity of interest) between the two nodes connected by the edge. The underlying intuition is that inequality as experienced by the members of a population is best seen as a series of pairwise

---

[1] Affiliations: Bowles: Santa Fe Institute; Carlin: Department of Economics, University College London and CEPR.
We are grateful to the Behavioral Sciences Program at SFI for their support of this work and for help from Antonio Cabrales, Weikai Chen, Arjun Jayadev, Christopher Lee, Suresh Naidu, Ian Preston, John Roemer, Margaret Stevens, and Bob Rowthorn.

comparisons between self and others. This undirected network representation is the basis of the definition of the Gini coefficient as one half the mean difference among the members of the population relative to the population mean wealth.

The appeal of the Gini coefficient conceived in this way is that it is "a very direct measure of income differences, taking note of differences between every pair of incomes." (Sen 1997):31 The coefficient ranges (as Gini specified that it should and shown by (Deaton 1997):139) from zero, if there are no differences among population members, to one, if a single individual owns all of the wealth (Gini 1914):1204.

But the conventional geometrical definition of the Gini coefficient as the area between the "perfect equality line" and the Lorenz curve divided by the area under the perfect equality line and the algorithm used to calculate it (Kendall and Stuart 1969, Dasgupta, Sen, and Starrett 1973, Damgaard 2018, Sen 1997) does not measure the relative mean difference among pairs in the population. This algorithm is illustrated by the right panel of Figure 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.

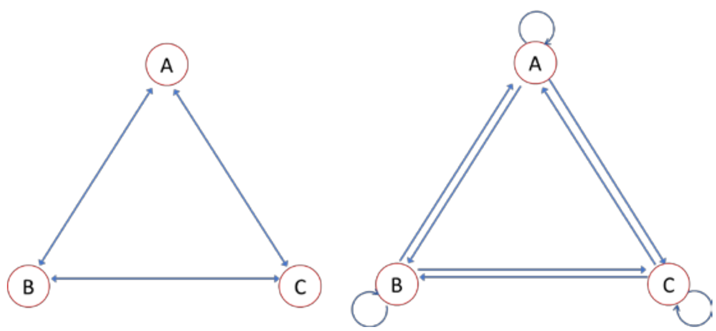

Figure 1. Experienced differences (left panel) and the edges used in the conventional measure (right panel).

For finite populations (the only relevant case) the conventional algorithm (and the equivalent Lorenz based geometrical measure) is a systematically downward-biased approximation. If the nodes A, B, and C in Figure 1 have wealth 10, 4, and 3, for example, the correct Gini coefficient using the network representation in the left panel is 0.412, which is what one finds by using equation (3) below. Using the conventional measure (equation 1, below) the Gini is estimated as 0.274 which (as we will see in equation (2)) needs to be multiplied by $n/(n-1)$ or 1.5 to get the correct Gini.

Moreover, the Gini coefficient calculated in this way 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.

Our network representation could readily be extended using social network data to take account of observed patterns of interaction so as to provide better measures of experienced inequality. Equation (3) below would then measure the degree of experienced inequality on a *complete* network. With the same levels of individual wealth, experienced inequality for incomplete networks would differ. For example, if the relevant social relationships changed from a complete network to a star with the wealthiest person at the center or a bipartite two-class network, experienced inequality would increase.[2] 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 structure.

2. THE TRUE GINI COEFFICIENT FOR A FINITE POPULATION.

The standard expression for the Lorenz curve based interpretation (which we will call G^L) for wealth, *y*, is

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

for which a population size adjustment is required so that this value will vary over the unit interval, namely, using G to denote the adjusted Gini coefficient (Yitzhaki and Schechtman 2013):

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

[2] If the three individuals in the left panel of Figure 1 were represented by a line (with the richer person in the center, a landlord, for example with two isolated sharecroppers) then holding constant the wealth of the three individuals the experienced inequality would rise from 0.41 to 0.57.

The population size bias in the unadjusted measure is substantial only when the population size under consideration is unusually small.[3] So the main reason to substitute G for G^L is not the "small numbers bias" but the fact that it does not measure experienced inequality.

Moreover equation (1) simply does not measure inequality, that is, 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 Figure 1 we have an intuitive definition of G which measures inequality among actual pairs in the population, which, as a result, does not require a population size adjustment. [4]

$$3) \quad G = \frac{\sum_{i=j+1}^{i=n} \sum_{j=1}^{j=n-1} \left| y_i - y_j \right|}{2\left(\frac{n(n-1)}{2}\right)\underline{y}}$$

The numerator of equation (3) is the sum of the absolute differences among the (unique non-identical) pairs, which we call Δ, or

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

and so,

[3] But this is sometimes the case, for example in archaeological (Kohler and Smith 2018) or industry concentration studies (Ghosh 1975).

[4] A further problem is calculating the Gini coefficient from the Lorenz curve in finite rather than infinite populations where both the "perfect equality line" and the Lorenz curve are represented by a step function as Gini suggested (Gini 1914) p.1232. Dasgupta, Sen and Starrett noted "there is some ambiguity as to how we define Lorenz curves in the discrete case". They "take the position that Lorenz points are plotted for each discrete number of people and connected by straight lines." (Dasgupta, Sen, and Starrett 1973):p.185. Following this convention instead of Gini's, for example in a two-person economy in which one person has all the income, the area between the Lorenz curve and the perfect equality line, divided by the total area under the line, that is, G^L, is equal to one half, not one.

$$3')\quad G=\frac{\Delta}{\left(\dfrac{n(n-1)}{2}\right)}\frac{1}{\underline{y}}\frac{1}{2}=\frac{\Delta}{n(n-1)}\frac{1}{\underline{y}}$$

from which we can see that the Gini coefficient 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 equation (2) which, noting that the sum of differences from equation (1) is just 2Δ can be seen from:

$$4)\quad G^{L}\left(\frac{n}{n-1}\right)=\frac{\Delta}{n^{2}}\frac{1}{\underline{y}}\left(\frac{n}{n-1}\right)=\frac{\Delta}{n}\frac{1}{\underline{y}}\left(\frac{1}{n-1}\right)=\frac{\Delta}{n(n-1)}\frac{1}{\underline{y}}=G$$

Using (3) or (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 $y(n-1)=\Delta$. Inserting this value in (3′) confirms that G = 1 for any n > 1.

But where wealth is owned by more than one person, G is not invariant to population size. If there are r wealthy individuals each owning y/r then we have $G=(n-r)/(n-1)$ and as we increase the population, holding the number of wealth owners constant, there are two effects on the Gini coefficient. The first (captured in the denominator) is that there will be relatively more frequent interactions among people with identical wealth (among the poor, that is). The second (captured in the numerator) is that a smaller fraction of the population holds all of the wealth.

Differentiating the expression just above for G with respect to n (holding r constant), we have

$$5)\quad \frac{\delta G}{\delta n}=\frac{1}{n-1}\left(1-\frac{n-r}{n-1}\right)=\frac{1}{n-1}(1-G)$$

from which it is easy to see that the second effect dominates the first except in the limiting case of a single owner of the entire wealth where the Gini coefficient is equal to one as we saw above, and is invariant to variations in n because the two effects cancel out.

3. A COMPARISON. WHY THE CONVENTIONAL MEASURE UNDERSTATES TRUE INEQUALITY

One way to see why the conventional Lorenz curve-based measure understates true 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.

The difference between the two measures, illustrated in Figure 1, is clarified by the two weighted adjacency matrices in Figure 2. In the left panel we show a three-person population and in the right a duplicated population of six. In both panels, data for our non-directed network representation are the (relevant) red edge weights used in the calculation, while the conventional method uses all of the data in the table (double counting each edge and including the zero-own-differences along the diagonal.) The zeros in black are the fictitious self-on-self zero differences.

	A	B	C
A	0	6	7
B	6	0	1
C	7	1	0

	A_1	A_2	B_1	B_2	C_1	C_2
A_1	0	0	6	6	7	7
A_2	0	0	6	6	7	7
B_1	6	6	0	0	1	1
B_2	6	6	0	0	1	1
C_1	7	7	1	1	0	0
C_2	7	7	1	1	0	0

Figure 2. Weighted adjacency measures: population replication for the true and conventional Gini measures. In the left panel, the wealth of A, B, and C are respectively 10, 4, and 3; in the right panel the As Bs and Cs have the same wealth as in the left panel. G is 0.418 in the initial population and 0.329 when the population is duplicated, whereas G^L is 0.274 in both the initial and duplicated populations.

Duplicating the population results in 3/15 of the true interactions being among equals. Thus, experienced inequality would seem to have declined, for while the preexisting inequalities have been replicated, doubling the population has introduced a degree of experienced equality. The reason why the conventional approximation of the true Gini understates true inequality is that it counts the fictitious zeros of the self-on-self comparisons as if they were real information about differences between people.

This also provides an intuition for the well-known fact that G^L is invariant to population replication. Counting the fictitious diagonal zero entries in Figure 2, for example, the number of "interactions" in which wealth inequality is absent has doubled as has the number of all of the other differences. The result is no change in G^L. It is clear, however, that experienced inequality differs in the two populations. In the left panel no individual ever interacts with a person of the same wealth, while in the right this occurs in three cases (out of a total of 15 paired interactions).

We now show the general result that experienced inequality declines with replication. Consider an initial population of n with the true Gini coefficient G_1 and construct a new population consisting of $k > 1$ 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 equation (2),

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

So, for the new population of kn individuals we have

$$6) \qquad G_k = G_k^L \frac{kn}{kn-1} = G_1 \frac{n-1}{n} \frac{kn}{kn-1} = G_1 \frac{k(n-1)}{kn-1} < G_1$$

which shows that replication reduces experienced inequality.

4. Employment, the profit share, and inequality in a three class economy

Our definition of the true Gini coefficient (equation (3)) can be applied to income or wealth of homogeneous classes where, instead of the three nodes in Figure 1 we have three classes of nodes, the members of which own the same amount of wealth.

Consider an economy in which there is some number l of "producers," u "inactive" agents and $n-l-u$ "wealthy". These groups could be employees, the unemployed (receiving no income) and employers in the analysis of income inequality or small property owners, the landless, and landlords in the study of land-wealth inequality. The line segments making up the Lorenz curve for income inequality in this three-class economy are illustrated in Figure 3.

Total output is the average product of the producers (q) times the number of producers (l), of which the fraction received by the producers is s, and the fraction received by the wealthy, is $1-s$. We now derive an expression for the Gini coefficient based on the share going to the producers and the size of the three classes.

There are $n - l - u$ individuals with income $\frac{l}{n-l-u}(q-w)$, l with income w, and u with income of zero. To find Δ, the sum of differences between all unique non-identical pairs in the population, we sum the number of edges between each of the class pairs times the income difference for that pair, so,

$$
\begin{aligned}
7)\quad \Delta &= (n-l-u)l\left(\frac{l}{n-l-u}(q-w)-w\right)+(n-l-u)u\frac{l(q-w)}{n-l-u}+luw \\
&= (l+u)lq+\left(-l-(n-l-u)-u+u\right)lw \\
&= (l+u)lq-(n-u)lw
\end{aligned}
$$

Then using equation (3′) we have

$$
8)\quad G=\frac{\Delta}{n(n-1)\underline{y}}=\frac{(l+u)lq-(n-u)lw}{(n-1)lq}=\frac{1}{n-1}\left((l+u)-(n-u)\frac{w}{q}\right)=\frac{1}{n-1}\left((l+u)-(n-u)s\right)
$$

If n is large, and denoting μ as the share of the inactive and λ the share of the producers in the population, it follows that

$$
9)\quad G=\frac{1}{n-1}\left((l+u)-(n-u)s\right)\approx \mu+\lambda-(1-\mu)s=G^{L}
$$

The last equality in equation (9) can be seen by summing the areas under the Lorenz curve (in Figure 3):

$$
\begin{aligned}
10)\quad B &= B_1+B_2+B_3=\frac{1}{2}s\lambda+s(1-\lambda-\mu)+\frac{1}{2}(1-s)(1-\lambda-\mu) \\
&= \frac{(1-\lambda-\mu)}{2}+\frac{s}{2}(1-\mu)
\end{aligned}
$$

and using this, the Gini coefficient estimated from the Lorenz curve, G^L is

$$
11)\quad G^{L}=\frac{A}{A+B}=\frac{1/2-B}{1/2}=1-2B=1-(1-\lambda-\mu)-s(1-\mu)\ =\mu+\lambda-s(1-\mu)
$$

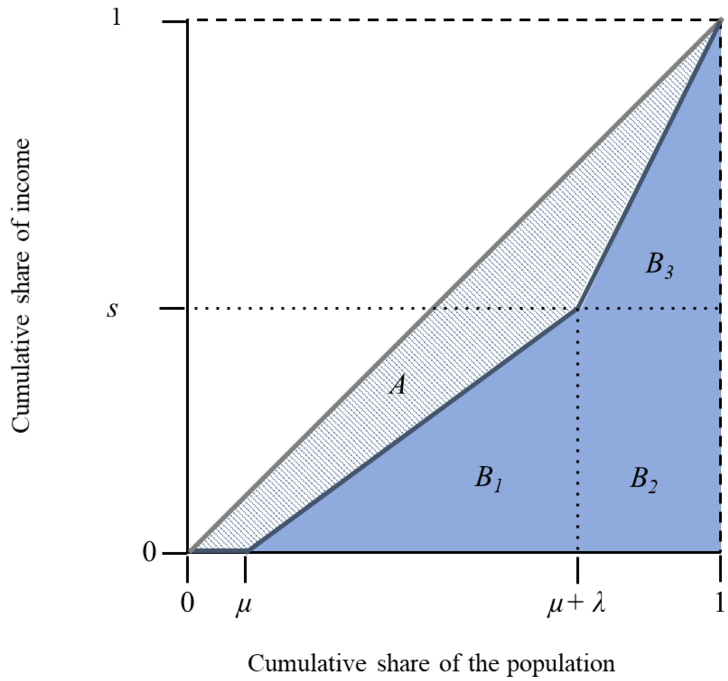

Figure 3. The Lorenz curve for a group-divided economy.

Converting this to the true Gini coefficient for population size *n*, we have:

$$
\begin{aligned}
\text{12)} \qquad G &= \left(\frac{n}{n-1}\right)G^{L} = \left(\frac{u+l}{n-1}\right) - \left(\frac{n}{n-1}\right)\left(1-\frac{u}{n}\right)s \\
&= \frac{1}{n-1}\left(u+l-(n-u)s\right)
\end{aligned}
$$

which is the expression in the text derived from equation (3) based on unique differences among non-identical pairs.

This confirms that the Gini coefficient based on the area between the Lorenz curve and the perfect equality line is not one half the relative mean difference for a finite population. Nonetheless, for suitably large populations, we can use equation (9) to express a variety of intuitions about inequality among wage earners, the unemployed and employers, or other population groups, and how this varies with the share.

This may be of particular interest now that the era of Kaldor's "stylized facts" and Solow's "constancy of relative shares" appears to be behind us ((Solow 1958, Dorn et al. 2018, Kaldor 1961, Karabarbounis and Neiman 2014)), especially in light of the recent turn in New Keynesian macroeconomic models from so-called representative agent (RANK) to heterogeneous agent (HANK) models including those with three ‘classes’ (e.g. (Ravn and Sterk 2016)).

5. Conclusion

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. The intuition is that inequality as experienced by the members of a population is best seen as a series of pairwise comparisons between self and others. This undirected network representation is the basis of the definition of the Gini coefficient as one half the mean difference between the members of all pairs in the population relative to the population mean wealth. For finite populations this is not what is measured by the algorithm conventionally used to calculate the Gini and it is not the geometrical definition based on the Lorenz curve. Comparison between this measure, G, and the conventional geometric Lorenz curve definition, G^L, suggests that the replication property of the latter is undesirable because it fails to capture experienced inequality.

References.

Damgaard, Christian. 2018. Gini Coefficient. *Mathworld: A Wolfram Web Resource*. Accessed 10 August, 2018.

Dasgupta, Partha, Amartya Sen, and David Starrett. 1973. "Notes on the measurement of inequality." *Journal of Economic Theory* 6:180-187.

Deaton, Angus. 1997. *The analsyis of household surveys: A microeconomic approach to development policy*. Washington, D.C: World Bank.

Dorn, David, Lawrence Katz, Christina Patterson, and John Van Reenen. 2018. "The fall of the labor share and the rise of superstar firms." *NBER Working paper*.

Ghosh, A. 1975. "Concentration and Growth of Indian Industries, 1948-1968." *Journal of Industrial Economics* 23:203-222.

Gini, Corrado. 1914. "Sulla misura della concentrazione e della variabilita de caratteri." *Atti del Reale Istituto Venneto di Schienze, Lettere ed Arti* LXXIII (Parte seconda):1203-1248.

Kaldor, Nicholas. 1961. "Capital accumulation and economic growth." In *The theory of capital*, edited by F.A Lutz and D.C. Hague, 177-222. London: St. Martins Press.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The global decline in the labor share." *Quarterly Journal of Economics*:61-103.

Kendall, Maurice G., and Alan Stuart. 1969. *The advanced theory of statistics*. London: Charles Griffin.

Kohler, T. A., and M. E. Smith. 2018. *Ten thousand years of inequality*. Tempe: University of Arizona Presss.

Ravn, Morten, and Vincent Sterk. 2016. "Macroeconomic Fluctuations with HANK and SAM: An Analytical Approach." *CEPR Discussion Paper 11696*.

Sen, Amartya K. 1997. *On Economic Inequality*. Oxford: Clarendon Press.

Solow, Robert. 1958. "The constancy of relative shares." *American Economic Review* 48:618-31.

Yitzhaki, S, and Schechtman. 2013. *The Gini Methodoogy: A Primer on a Statistical Methodology* New York: Springer.