ECINEQ

Society for the Study
of Economic Inequality

Working Paper Series

Individual Diversity and the Gini Decomposition

Lidia Ceriani
Paolo Verme

Individual Diversity and the Gini Decomposition*

Lidia Ceriani

EconPubblica, Bocconi University, Milano

Paolo Verme

The World Bank

Abstract

The paper defines the Gini index as the sum of individual contributions where individual contributions are interpreted as the degree of diversity of each individual from all other members of society. Among various possible forms of individual contributions to the Gini found in the literature, we show that only one form satisfies a set of desirable properties. This form can be used for decomposing the Gini into population subgroups. An illustration with UK data shows the use of this approach.

Keywords: Gini, Inequality, Decomposition.

JEL Classification: D31, D63.

1   Introduction

Over one hundred years ago, Corrado Gini (1912) published his seminal book *Variabilità e Mutabilità* where he presented for the first time the distributional index that today is associated with his name. The Gini index remains one of the most popular statistical indices of all times and continues to be subject of studies across the social sciences. A search on JSTOR finds 265 articles with the name 'Gini' in the title, 366 with the word gini in the abstract and 16,594 with the word gini in the text. A search on econpapers finds 423 journal articles with the word gini in the title or keywords. The majority of articles from these two sources concerns inference, decompositions, the relation with the Lorenz curve, various extension or formulations of the Gini, and relations with other measures such as deprivation.

It is known that the Gini index can be expressed in many different forms. A recent paper on the origins of the Gini index by Ceriani and Verme (2012) reports 13 different forms that emerged in the literature since its first formulation by Corrado Gini in 1912. With some basic manipulation of these indexes, it is also possible to see that eight of these forms can be expressed as sums of individual values across a given population. This is a simple but interesting observation because implies that these eight formulations of the Gini can potentially be used to decompose the Gini into population sub-groups. For example, one could estimate the individual contribution to the Gini for all members of a given population, sum these contributions by males and females and then sum results for the two groups to obtain the overall Gini. This simple procedure would overcome the century long issue of the Gini decomposition into population sub-groups.

The problem with this approach is that the individual contribution to the Gini *per se* has no meaning. We cannot really talk of individual inequality. By definition, inequality is measured across a set of individuals. However, if we were able to provide a meaning and a set of desirable properties to the individual contribution to the Gini, then these individual values would become meaningful in their own right, could be added across groups and used for the Gini decomposition into population sub-group.

The paper follows this approach. We define the individual contribution to the Gini as a

measure of individual *diversity*, we identify a set of desirable properties that this measure of individual diversity should have and we seek, among the various formulation of the individual contribution to the Gini, those formulations that satisfy these desirable properties. We will find that only one formulation satisfies these properties and we will show how this formulation can be used for the Gini decomposition into population sub-groups.

The concept of individual diversity we propose is similar to the concept of individual complaint developed by Temkin (1986) and characterized by Cowell and Ebert (2004) and the concept of individual relative deprivation developed by Yitzhaki (1979) and characterized by Ebert and Moyes (2000). As shown by these authors, societal measures of complaint or deprivation can be seen as the sum of individual values where individual values have a meaning in themselves and are measured as the sum of income distances between one own income and the income of richer or poorer individuals. As also shown by these authors, the societal measures of complaint or deprivation have a direct algebraic link with the Gini index.

This paper builds on this literature in four different respects. First, we define individual values as *individual diversity*. By definition, this concept does not attribute any positive or negative connotation to the individual values and, we believe, is closer to the concept of inequality that Corrado Gini had in mind. One cannot talk of individual inequality but one can talk of individual diversity and it can be reasonably argued that the sum of individual diversities in a given population is a measure of societal inequality. Second and as a consequence of the first point, we consider the full set of income distances between an individual and all other members of society. Or, in other words, we capture deprivation and satisfaction into one measure of individual diversity or inequality. Third, instead of defining a new societal index, we review the existing forms of the Gini and ask the question of whether any of these forms satisfies our quest for a measure of individual diversity. Fourth, we exploit these features to provide an exact decomposition of the Gini index into population subgroups. This is a rather different approach from the traditional decompositions of the Gini index in between and within components as originally proposed by Bhattacharaya and Mahalanobis (1967) and Pyatt (1976).

In the next section we illustrate the diversity of Gini indices offered by the literature and

the different distributions of unit values that these indices imply. Section three outlines some of the desirable properties that a measure of individual diversity should have and tests which Gini satisfies these properties. Section four provides an example of the Gini decomposition by population subgroups using UK data and section five concludes.

2 Different Ginis

The Gini index can be written in many different forms (see Xu, 2003 and Ceriani and Verme, 2012 for reviews). In his 1912 book, Corrado Gini first proposed a measure which he called the *average difference between n quantities*.[2] Gini was particularly keen in showing how any distribution could be seen as a symmetric distribution where each observation had a symmetric counterpart, an aspect that became evident when he published an article that proposed a modified version of his index as a concentration ratio (Gini, 1914).[3] The ratio proposed was equal to his original *average distance between n quantities* divided by twice the arithmetic mean of X. This is what today is commonly referred to as the *Gini index* or *one half of the Gini relative mean difference*.[4]

As the starting point of our analysis, we will then use the two indices introduced by Gini in 1912 but expressed in the more popular form published in 1914. Let us consider a population of N individuals, $i = 1, 2, \ldots, n$, $n \in \mathbb{N}$, $n \geq 3$, having an income distribution $X = (x_1, x_2, \ldots, x_i, \ldots, x_n)$, where $X \in \mathbb{R}_{++}^n$ and $x_1 \leq x_2 \leq, \ldots, \leq x_n$. Then the Gini index can be expressed as:

$$G^I(X) = \sum_{i=1}^{n} \frac{(n+1-2i)(x_{n-i+1} - x_i)}{2n^2 \mu_X} \tag{1}$$

or

---

[2] *Differenza media tra le n quantità*, Gini (1912), p. 22, eq. 5.

[3] This is the article where Gini shows the relation of his index with the Lorenz curve.

[4] Gini's *average distance between n quantities* was later referred by Dalton (1920) as the *absolute mean difference* to distinguish it from the *relative mean difference* defined as the Gini absolute mean difference divided by the mean.

$$G^{II}(X) = \sum_{i=1}^{n} \frac{2(i-M)(x_i - x_M)}{n^2 \mu_X} \tag{2}$$

where μ_X and x_M are, respectively, the arithmetic mean and the median of distribution X and M is the rank of the individual with the median income.

Since its introduction, the Gini index attracted a great amount of attention and has been reformulated in at least 13 different forms as noted in Ceriani and Verme (2012). By reviewing these forms, we found a total of eight forms that can be expressed as sums of individual contributions and that exhibit different individual functions. These are reported in Table 1 in chronological order and expressed as sums of individual observations across the population. Table 1 also reports the alleged original proponents of each index and a tentative synthetic description of the form of index.[5]

[Table 1]

One first remarkable aspect is the variety of individual functions underlying the different Ginis. For example, form II is expressed in terms of distances from the median while form VIII is expressed in terms of distances from the mean. Form IV ignores individual incomes altogether and the index is expressed only in terms of rank whereas forms III and V ignore rank and use only individual incomes. To illustrate further differences across the eight Gini indices considered, we took a small arbitrary sample of eleven observations, estimated the individual values for each type of Gini and plotted the distributions of these values. Table 2 reports the individual values and Figure 1 plots these values. Note that, by construction, two values of the income distribution reported in Table 2 correspond to the mean and median values respectively. This is to better appreciate what happens to the individual contributions to the Gini in correspondence of these two central moments of the distribution.

Several differences across the distributions of individual values are evident.[6] First, some

---

[5]It is unclear where formulation III first appeared. It is not in Gini's 1912 book or 1914 article and can be found in Kendall and Stuart (1958) and Xu (2003). We attributed it to Kendall and Stuart (1958) but it could well have appeared before in the literature.

[6]Note that all forms of index have been estimated with x ranked in ascending order.

Gini forms result in only positive values (I, II, III, V) and others in negative and positive values (IV, VI, VII, VIII). Second, some forms attribute the same unit values to the same values of x (III, V) while others attribute different values (I, II, IV, VI, VII, VIII).[7] Third, there is no common order in rank across the series. One is ordered in descending order of x (V), five are U-shaped (I, II, III, VII, VIII), one is ordered in ascending order of x (VI) and one has no regular shape (IV). Four, some of the U-shaped distributions invert the trend in correspondence of the median (I, II, III) while others in correspondence of lower values (VII, VIII). Fifth, some individual scores take the value of zero in correspondence of the median (I, II, VII) and one in correspondence of the mean (VIII). Sixth, in some series the greatest absolute contribution to the Gini is given by the largest values of x (II, III, VI, VII, VIII), others by the lowest values (IV, V) and one is perfectly symmetric (I). Can any of the functional forms be suitable to describe an individual measure of diversity? This is the question we address in the next section.

[Table 2 and Figure 1]

3 Properties

As before, let us consider a population of N individuals, $i = 1, 2, \ldots, n$, $n \in \mathcal{N}$, where $\mathcal{N}$ is the class of all possible finite subset of $\mathbb{N}$ with at least three elements. Each i-th individual in population N is endowed with a non negative income $x_i \in X$, where X is a generic vector of length n, and $\mathcal{X}^n$ the class of all vector X.

Notation (x_i, x_{-i}) denotes a relative income distribution X where individual i-th has a relative income of x_i and all other $j \neq i$ individuals have a relative income distribution of x_{-i}. In the same way, (x_i, x_j, x_{-ij}) denotes a relative income distribution X where individual i-th has a relative income x_i, individual j-th has a relative income x_j, and all other $k \neq i, j$ individuals have a relative income distribution described by $x_{-i,j}$. Also, $G(X)$ is the inequality level related to income distribution X, as measured by the relative Gini index and μ_X is the mean income in distribution X.

---

[7]See observations 7 and 8.

Definition 3.1. $g_i(x_i, x_{-i}) : \bigcup_{n \in \mathcal{N}} \mathcal{X}^n \to \mathbb{R}_+$ *is a measure of individual i diversity, such that the average of all individual diversities normalized by the mean income in the population returns the Gini coefficient:* $G(X) = \frac{1}{n\mu_X} \sum_{i=1}^{n} g_i(X)$.

The aim of this section is to define a set of desirable properties of this individual index of diversity.

Property 3.1 (Continuity). $g_i(x_i, x_{-i})$ *is continuous over* $\bigcup_{n \in \mathcal{N}} \mathcal{X}^n$.

The first property is that the individual contribution to inequality is weakly sensitive to small changes in income values. Given that empirical data are typically affected by measurement errors, this property ensures that the individual contribution to the Gini is not very sensitive to such errors.

Property 3.2 (Additivity). *Let $\mathcal{N}_{-i}$ be the class of all possible non-trivial subset of $N - \{i\}$, and let N^1 and N^2 be two generic elements of $\mathcal{N}_{-i}$, such that $N_{-i}^1 \cup N_{-i}^2 = N - \{i\}$. Let x_{-i}^1 and x_{-i}^2 be the income distribution of subgroup N_{-i}^1 and N_{-i}^2 respectively, such that $X = \left(x_i, x_{-i}^1, x_{-i}^2\right)$. Then, $g_i(x_i, x_{-i}) = g_i(x_i, x_{-i}^1) + g_i(x_i, x_{-i}^2)$.*

The individual i-th diversity is the sum of individual i-th diversity in different subgroups of the population (where a sub-group can be constituted by a single individual).

Property 3.3 (Linear Homogeneity). $g_i(\lambda x_i, \lambda x_{-i}) = \lambda g_i(x_i, x_{-i})$, *for any $\lambda \in \mathbb{R}_+$.*

If all incomes in the population are scaled by a factor λ, individual i-th diversity is scaled by the same factor.

Property 3.4 (Translation Invariance). $g_i(x_i, x_{-i}) = g_i(x_i + \lambda, x_{-i} + \mathbf{1}\lambda)$, *for any $\lambda \in \mathbb{R}$, and for $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}_+^{n-1}$.*

Individual diversity does not change whenever all incomes are changed by the same amount λ.

Property 3.5 (Population Invariance). $g_i(X) = g_i(X_\alpha)$ *where X_α is an α-replication of X, $X_\alpha = (\underbrace{x_1, x_2, \ldots, x_n}_{1}, \underbrace{x_1, x_2, \ldots, x_n}_{2}, \ldots, \underbrace{x_1, x_2, \ldots, x_n}_{\alpha})$ and $\alpha \in \mathbb{R}_{++}$.*

If each individual in the population is replicated α-times, individual i-th diversity is unchanged.

Property 3.6 (Symmetry). *For any $i, j \in N$, for any x_i, x_j, $x_j' \in X$ and for any $\epsilon \in \mathbb{R}_+$, such that $x_j = x_i + \epsilon$ and $x_j' = x_i - \epsilon$: $g_i(x_i, x_j, x_{-ij}) = g_i(x_i, x_j', x_{-ij})$*

Individual i-th diversity is unchanged if, everything else being equal, she faces another individual who is richer or poorer by ϵ. Her diversity is influenced only by the difference between her income and other incomes, regardless of satisfaction or deprivation.

Property 3.7 (Anonymity)**.** *Given any permutation π of N, such that $X^\pi = (x_{\pi(i)}, x_{-\pi(i)}) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$, $g_i(x_i, x_{-i}) = g_{\pi(i)}(x_{\pi(i)}, x_{-\pi(i)})$*

By imposing Anonymity only relative income levels define the individual contribution to inequality. This implies that to equal relative incomes correspond equal individual contributions to inequality.

Table 3 reports whether the various forms of individual contribution to the Gini satisfy or do not satisfy the listed properties. All different formulations satisfy Continuity, Additivity and Linear Homogeneity. Gini formulations IV, VI and VII fail to obey Translation Invariance, while Anonymity excludes those formulations of the Gini based on rank (all except for III and V) and by imposing Symmetry, we rule out the Gini based on a relative-deprivation concept (form V). Only form III satisfies all properties.

4   Decomposition by population sub-groups

Historically, the decomposition of the Gini index has focused on two main areas, decomposition by income source and decomposition by within and between groups. The literature on inequality decomposition by within and between groups is rather rich. It was first proposed by Bhattacharaya and Mahalanobis (1967) and Pyatt (1976) in different contexts but both methodologies led to decompositions into within and between groups inequalities, a line of research followed by many others (see for example Blackorby, Donaldson, and Auersperg, 1981 and Cowell, 1980). Bourguignon (1979) stated that "*a decomposable inequality measure is defined as a measure such that the total inequality of a population can be broken down into a weighted average of the inequality existing within subgroups of the population and the inequality existing between them.*" And Shorrocks (1980) defined an additively decomposable inequality measure as "*one which can be expressed as a weighted sum of the inequality values calculated for population subgroups plus the contribution aris-*

ing from differences between subgroups means."[8] Initially, the Gini index was not thought to be suitable for decompositions but there are now some methodologies that can be used for an exact decomposition of the Gini such as the Shapley value method (Shorrocks, 1999). All these contributions continued to focus on the within and between groups decomposition of the Gini.

The definition of the individual contribution to inequality presented in the previous section opens the possibility for a different form of additive decomposition by population sub-groups. The individual contributions to the Gini are considered as a measure of the individual degree of diversity. When we aggregate these individual degrees of diversity across groups such as males and females, we can simply add up the individual values by group and obtain the Gini. If we take the share of Gini by group, we obtain an exact decomposition by subgroup.

As an example, let $i = 1, 2, \ldots, m$ be the set M of male individuals in the population and $j = m + 1, \ldots, n$ the set W of female individuals in the population, where $M \cup W = N$. Then, the Gini index can be written as the sum of males and females individual diversities.

$$g(X) = \frac{1}{n\mu_X} \left[\sum_{i=1}^{m} g_i + \sum_{j=m+1}^{n} g_j \right] \tag{3}$$

This decomposition is determined by both group size and within group inequality. Groups' size being equal, the more unequal group accounts for more inequality. Groups' inequality being equal, the larger group accounts for more inequality. We consider population size and within group inequality as equally legitimate contributors to total inequality.

To illustrate the decomposition proposed, we took a reduced sample from the 2000 British Household Panel Survey (BHPS) restricting the population to employees in age 41-50 and using as a measure of welfare income net of taxes. Table 4 reports groups means and population size as well as the decomposition by groups of total inequality. We can see that men contribute to total inequality by almost 56% of the total. This is due to both

---

[8]Both definitions can be found in the abstracts of the respective papers.

population size where men represent 52.7% of the total population and the within group inequality which is higher for men.

5   Conclusion

The Gini index can be formulated in many different forms. When expressed as sums across the population, several of these forms provide different values at the individual level or, in other words, have different individual functions. We asked the question of what form an individual function should take so as to represent the portion of inequality explained by each individual. Based on a set of desirable properties, we showed that only one form of function can be derived from a set of desirable properties. This is the original formulation of the Gini index as found in Kendall and Stuart (1958). We then illustrated the use of the individual contributions to the Gini for an exact decomposition of the Gini by population subgroups.

References

BHATTACHARAYA, N., AND B. MAHALANOBIS (1967): "Regional Disparities in Household Consumption in India," *Journal of the American Statistical Association*, 22(317), 143–161.

BLACKORBY, C., D. DONALDSON, AND M. AUERSPERG (1981): "A New Procedure for the Measurement of Inequality Within and Among Population Subgroups," *Canadian Journal of Economics*, 14, 665–685.

BOURGUIGNON, F. (1979): "Decomposable Income Inequality Measures," *Econometrica*, 47, 901–902.

CERIANI, L., AND P. VERME (2012): "The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini," *The Journal of Economic Inequality*, 10(3), 421–443.

COWELL, F. (1980): "On the Structure of Additive Inequality Measures," *Review of Economic Studies*, 47, 521–531.

COWELL, F., AND U. EBERT (2004): "Complaints and Inequality," *Social Choice Welfare*, 23, 71–89.

DALTON, H. (1920): "The measuremento of the inequality of incomes," *Economic Journal*, 30, 348–361.

EBERT, U., AND P. MOYES (2000): "An axiomatic characterization of Yitzhaki's index of individual deprivation," *Economics Letters*, 68, 263–270.

GINI, C. (1912): *Variabilità e Mutabilità. Contributo allo studio delle distribuzioni e delle relazioni statistiche*, Studi economico-giuridici Anno III, Parte II. Facoltà di giurisprudenza della Regia Università di Cagliari, Cuppini, Bologna.

——— (1914): "Sulla Misura della Concentrazione e della Variabilità dei Caratteri," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, LXXIII(II), 1203–1248.

KENDALL, M. G., AND A. STUART (1958): *The Advanced Theory of Statistics*, vol. 1. Hafner Publishing Company, New York, 1st edn.

PYATT, G. (1976): "On the Interpretation and Disaggregation of Gini Coefficients," *Economic Journal*, 86, 243–255.

SHORROCKS, A. F. (1980): "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48, 613–625.

——— (1999): "Decomposition Procedures for distributional analysis¿ a unified framework based on the Shapley value," manuscript, Univeristy of Essex.

TEMKIN, L. S. (1986): "Inequality," *Philosophy and Public Affairs*, 15, 99–121.

XU, K. (2003): "How Has the Literature on the Gini's Index Evolved in the Past 80 Years?," Department of economics at dalhousie university working papers archive, Dalhousie, Department of Economics.

Yitzhaki, S. (1979): "Relative Deprivation and the Gini Coefficient," *Quarterly Journal of Economics*, XCIII, 321–324.

Table 1: Alternative Gini Formulations

Index	Source	Form
$G^{I} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{(n+1-2i)(x_{n-i+1}-x_i)}{2n}$	Gini, 1912, 1914	Original
$G^{II} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{2(i-M)(x_i-x_M)}{n}$	Gini, 1912	Median
$G^{III} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{|x_i-x_j|}{2n}$	Kendall and Stuart, 1958	Adjusted Gini
$G^{IV} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{(n+1)\mu_X - 2(n+1-i)x_i}{n}$	Sen, 1973	Geometric
$G^{V} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \sum_{j>i} \frac{(x_j-x_i)}{n}$	Yitzhaki, 1979	Deprivation
$G^{VI} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{2ix_i - (n+1)\mu_X}{n}$	Anand, 1983	Covariance
$G^{VII} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{(2i-n-1)x_i}{n}$	Silber, 1989	Matrix
$G^{VIII} = \frac{1}{n\mu_X} \sum_{i=1}^{n} \frac{2i(x_i-\mu_X)}{n}$	Shorrocks, 1999	Mean

13

Table 2: Gini Individual Values

i	x_i	I	II	III	IV	V	VI	VII	VIII
1	3	17.27	15.00	9.00	16.91	18.00	-22.36	-2.73	-3.27
2	6	11.27	9.82	7.77	12.00	15.27	-20.73	-4.36	-5.45
3	10	7.36	5.18	6.50	6.55	12.00	-17.45	-5.45	-6.00
4	12	2.36	2.73	6.05	5.45	10.55	-14.18	-4.36	-6.55
5	19	0.18	0.09	5.09	-1.27	6.09	-5.64	-3.45	-1.82
6	20	0.00	0.00	5.05	1.09	5.55	-1.09	0.00	-1.09
7	21	0.18	0.27	5.09	3.82	5.09	3.82	3.82	0.00
8	25	2.36	2.00	5.64	4.73	3.64	13.45	9.09	5.82
9	37	7.36	9.55	8.36	2.73	0.36	37.64	20.18	26.18
10	37	11.27	12.73	8.36	9.45	0.36	44.36	26.91	29.09
11	41	17.27	19.55	10.00	15.45	0.00	59.09	37.27	40.00

G(X) = 0.333; $\mu_X = 21$; $x_M = 20$.

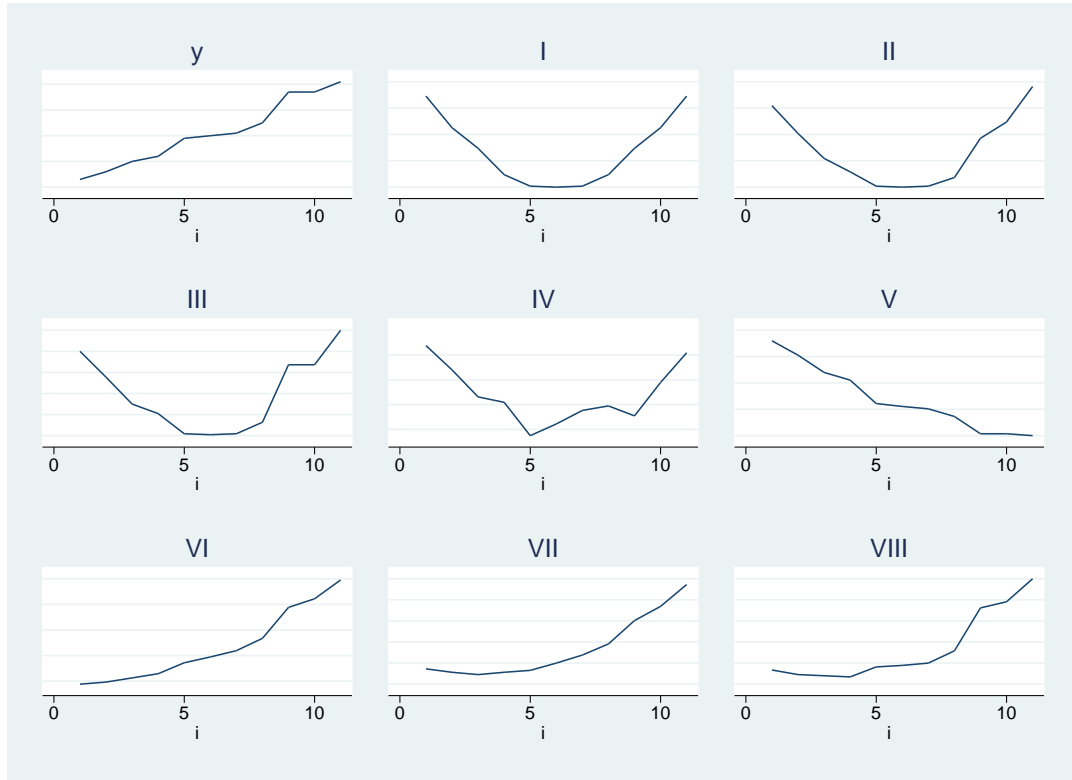Figure 1: Distributions of Gini Individual Values

Table 3: Ginis and axioms

No.	Axiom	I	II	III	IV	V	VI	VII	VIII
1	Continuity	yes	yes	yes	yes	yes	yes	yes	yes
2	Additivity	yes	yes	yes	yes	yes	yes	yes	yes
3	Linear homogeneity	yes	yes	yes	yes	yes	yes	yes	yes
4	Translation Invariance	yes	yes	yes	no	yes	no	no	yes
5	Population Invariance	no	no	yes	no	yes	no	no	no
6	Symmetry	no	no	yes	no	no	no	no	no
7	Anonymity	no	no	yes	no	yes	no	no	no

Table 4: Decomposition by Gender

Gender	Gini	Population	Pop. Share	Gini Contrib.	Gini Share
Male	0.238	463	52.7	0.181	55.9
Female	0.134	416	47.3	0.143	44.1
Total	0.324	879	100	0.324	100