A gender inequality assessment by means of the Gini index decomposition

Una valutazione della disuguaglianza di genere attraverso la scomposizione dell'indice di Gini

Michele Costa

Abstract This paper proposes to measure and to evaluate gender gaps and gender inequalities by means of the decomposition of an inequality measure. A three-terms decomposition of the Gini index is applied, thus allowing to take into account also the role of overlapping between female and male subpopulations. An analysis of the income distribution of the Italian households shows how gender gaps represent a major source of inequality, without particular improvements over the last 20 years.
Abstract *In questo lavoro si propone di analizzare e di valutare i differenziali e la disuguaglianza di genere grazie alla scomposizione di una misura di disuguaglianza. Il ricorso ad una scomposizione dell'indice di Gini articolata su tre componenti permette di tenere conto anche della sovrapposizione tra le distribuzioni delle sottopopolazioni femminile e maschile. L'analisi della distribuzione del reddito familiare in Italia mostra che i differenziali di genere rappresentano un importante fattore di disuguaglianza, sostanzialmente stabile durante gli ultimi 20 anni.*

Key words: Gender gap, Gender income inequality, Inequality decomposition, Gini index

1 Introduction

Gender inequalities and gender gaps are a worldwide concern and represent the core of uncountable actions and policies developed by either governments and institutions. Gender inequalities are firstly a primary and fundamental issue of justice. Consequences of gender inequalities are frequently overlooked or underestimated, while it exists an interesting literature which analyzes the relation between gender inequality and welfare, pointing out gender gaps as a constraint for economic growth.

Michele Costa
Department of Economics, University of Bologna, e-mail: michele.costa@unibo.it

We assess the role of gender in income inequality by decomposing the Gini inequality ratio following the approach introduced by Dagum in 1997. First we evaluate the inequality within male and female subgroups, second we analyse the contribution to total inequality attributable to the differences between female and male subpopulations. The Dagum's Gini index decomposition also allows to evaluate the effect on overall inequality of the overlapping between female and male subpopulations, which represents a relevant element in gender inequality studies.

2 The Dagum's Gini index decomposition

The Gini index is one of the most important measure of inequality and, during its over 100 years of file, has experienced many different interpretations, expressions and formulas. For the case of a population disaggregated into k subgroups of size n_j, with $\sum_{j=0}^{k} n_j = n$, the Gini index G can be expressed as follows

$$G = \frac{1}{n\bar{y}^2} \sum_{j=1}^{k} \sum_{h=1}^{k} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}| \tag{1}$$

where $\bar{y}$ is the arithmetic mean of y in the overall population, y_{ji} is the value of y in the i-th unit of the j-th subgroup and, accordingly, y_{hr} is the value of y in the r-th unit of the h-th subgroup. For a detailed discussion of the Gini index see, e.g., [3],[7].

Among the many methods which allow to decompose the Gini index (see, e.g., [4],[8],[9]), we use the decomposition proposed by Dagum [5], where the differences $|y_{ji} - y_{hr}|$ in (1) are assigned to G_w, the component of inequality within subgroups, when $j = h$, to G_b, the component of inequality between subgroups, when $j \neq h$, $\bar{y}_j \geq \bar{y}_h$, $y_{ji} \geq y_{hr}$, and to G_t, the component of overlapping, when $j \neq h$, $\bar{y}_j \geq \bar{y}_h$, $y_{ji} < y_{hr}$.

The component of inequality within can be obtained quite esaily from the relation $G_w = \sum_{j=1}^{k} G_j p_j s_j$, where G_j is the Gini index of the j-th subgroup, while $p_j = n_j/n$ and $s_j = (n_j \bar{y}_j)/(n\bar{y})$ are the population share and the income share of the j-th subgroup, respectively.

For the other two components, G_b and G_t, which in the original version require some substantial computational effort, are available [1] simplified expressions, which are $G_b = G_b^* + 0.5(G - G_w - G_b^*)$ and $G_t = 0.5(G - G_w - G_b^*)$, where $G_b^* = \sum_{j=1}^{k-1} \sum_{h=1, j=k}^{k} \frac{p_{hj}^* - s_{hj}^*}{p_{hj}^* s_{jh}^* + p_{jh}^* s_{hj}^*} (p_j s_h + p_h s_j)$, $p_{hj}^* = p_h/(p_h + p_j)$ and $s_{hj}^* = s_h/(s_h + s_j)$.

In order to achieve a better understanding of the inequality structure, it is also possible to compare the decomposition obtained by using all n observations, that is $G_y = G_{wy} + G_{by} + G_{ty}$, to the decompositions obtained by referring only to subsamples of observations. In particular, it is useful to analyze the decompositions for the lower values of y, $G_{y|ymin} = G_{wy|ymin} + G_{by|ymin} + G_{ty|ymin}$ as well as for the higher values of y, $G_{y|ymax} = G_{wy|ymax} + G_{by|ymax} + G_{ty|ymax}$. When the structure of

the decomposed indices $G_{y|ymin}$ and $G_{y|ymax}$ is similar, we get that the underlying inequality factor operates uniformly on y, while different structures indicate that particular regions of y are more affected by the inequality factor.

Following a similar approach, we can also evaluate the influence of a further inequality factor x by ranking y on the values of x and by comparing the decompositions of $G_{y|xmin}$ and $G_{y|xmax}$. Similar decompositions suggest that the x and the y are independent, while different decompositions indicate a relation between the two inequality factors.

A final element of interest refers to the evaluation of the inequality between, which is usually performed on the basis of the ratio G_w/G, where G acts as the maximum of G_b. The scenario $G_b = G$ implies $G_w = G_t = 0$: while $G_t = 0$, that is the absence of overlapping, doesn't present particular difficulties, the hypothesis $G_w = 0$, that is the equidistribution of y within each subgroup, represents a relevant departure from real situations. In order to achieve an evaluation of G_b more coherent with the observed data ([6], [2]), it is possible to keep $G_t = 0$ but to replace $G_w = 0$ with $G_w = G_{wmin}$, which is the minimum inequality within compatible with the observed data. In this case G_b is evaluated as $G_b/(G - G_{wmin})$.

3 The gender income inequality among Italian households

The Dagum's decomposition of the Gini index presented in Section 2 is extremely useful to analyze the relevance of gender in income inequality. The component G_w allows to evaluate how the income variability existing within the female and male subpopulations influence total inequality, while the contribution attributable to the differences between the female and male subpopulations is given by G_b and G_t. The meaning of G_b is straightforward, but as far as G_t it is useful to point out that high levels of overlapping indicate a small contribution of gender to income inequality, while low levels of overlapping suggest a stronger contribution.

The data used in this study are from the Survey on Households Income and Wealth of the Bank of Italy; the results illustrated in the following refer to the equivalent income obtained by means of the OCSE equivalence scale.Table 1 shows the p_i, s_i and G_i for the Italian households by gender of the head of the household: it is possible to observe some well known stylized facts of income inequality in Italy, that is the differences $(p_i - s_i)$. When $p_f = s_f$, the gender gap is equal to 0, while $p_f > s_f$ indicates the existence of a gender gap. The aggregate data of Table 1 suggest the presence of a gender gap, but also its reduction over time, since $(p_f - s_f)$ decreases from 3.8% in 1993, to 3.2% in 2004 and to 3% in 2014.

Moving from the aggregate and gross evaluation provided by $(p_f - s_f)$ to the more detailed and accurate information contained on the decomposed Gini index (Table 2), we obtain a different picture on gender income inequality. First, the importance of G_w on total inequality strongly decreases (from 62% in 1993 to 50% in 2014), thus indicating a weaker variability within the female and male subpopulations. Second, the overlapping between the female and male subpopulations in-

Table 1 Population share, income share and Gini index for the Italian households by gender of the head of the household

	1993			2004			2014		
	female	male	tot	female	male	tot	female	male	tot
p	0.275	0.725	1.000	0.388	0.612	1.000	0.471	0.529	1.00
s	0.237	0.763	1.000	0.356	0.644	1.000	0.441	0.559	1.00
G	0.319	0.334	0.333	0.307	0.336	0.327	0.311	0.324	0.320

creases: the importance of G_t rises from 13.5% in 1993 to 20.3% in 2014. A greater overlapping represents a positive signal for the reduction of the gender gap, since it suggests that the distributions of the subpopulations share a larger area. Third, the inequality between increases: the importance of G_b rises from 24.9% in 1993 to 29.7% in 2014. The relevance of the inequality between is fully understandable by comparing G_b to its maximum compatible with the observed data (last column of Table 2): in this case G_b represents 37.9% of total inequality in 1993, rising to 42.9% in 2014. Overall, the decrease of G_w is balanced by the increase of both G_b and G_t. While a greater G_t alleviates the role of gender as inequality factor, an increase of G_b leads to a stronger gender inequality from 1993 to 2014.

Table 2 Income inequality decomposition by gender of the head of the household

	Gw	Gb	Gt	Gw/G	Gb/G	Gt/G	Gb/(G-Gwmin)
1993	0.205	0.083	0.045	0.616	0.249	0.135	0.379
2004	0.175	0.092	0.060	0.535	0.281	0.183	0.407
2014	0.160	0.095	0.065	0.500	0.297	0.203	0.429

In order to better understand the results of Table 2, we focus on the tails of the distribution, taking into account the bottom and the top 20% of the income. Table 3 reports the p_i, s_i and G_i for the female and male subpopulations for the two cases and it is possible to observe some relevant differences.

By comparing the decomposed Gini indexes for the bottom and the top incomes (Table 4), we note that the two decompositions, initially quite different, are more or less similar in 2014. The importance of G_b shows a relevant increase, especially for the top incomes.

A further analysis of the gender income inequality refers to the study of specific population characteristics, such as educational level and geographical area of residence, chosen among the main inequality factors acknowledged by the literature. The Gini index decomposition is applied not to all n observations, but only to the subsample of households with the particular characteristic which we are analyzing. More specifically, we compare the female/male decompositions obtained on two subgroups related to two different values of the character under examination. When the two decompositions are substantially similar, the underlying factor is not rele-

Table 3 Population share, income share and Gini index for the Italian households by gender of the head of the household

	1993			2004			2014		
	female	male	tot	female	male	tot	female	male	tot
				20% bottom income					
p	0.362	0.638	1.000	0.448	0.552	1.000	0.482	0.518	1.000
s	0.374	0.626	1.000	0.450	0.550	1.000	0.476	0.524	1.000
G	0.173	0.203	0.193	0.154	0.164	0.159	0.232	0.223	0.227
				20% top income					
p	0.186	0.814	1.000	0.310	0.690	1.000	0.389	0.611	1.000
s	0.176	0.824	1.000	0.290	0.710	1.000	0.378	0.622	1.000
G	0.160	0.189	0.185	0.187	0.231	0.219	0.161	0.185	0.176
				up to elementary school					
p	0.388	0.612	1.000	0.503	0.497	1.000	0.595	0.405	1.000
s	0.357	0.643	1.000	0.490	0.510	1.000	0.582	0.418	1.000
G	0.275	0.293	0.289	0.267	0.274	0.271	0.251	0.280	0.263
				with university degree					
p	0.172	0.828	1.000	0.331	0.669	1.000	0.481	0.519	1.000
s	0.164	0.836	1.000	0.289	0.711	1.000	0.434	0.566	1.000
G	0.253	0.304	0.297	0.259	0.345	0.324	0.309	0.306	0.312
				south islands					
p	0.247	0.753	1.000	0.407	0.593	1.000	0.475	0.525	1.000
s	0.221	0.779	1.000	0.382	0.618	1.000	0.440	0.560	1.000
G	0.312	0.351	0.344	0.292	0.320	0.310	0.312	0.341	0.330
				north					
p	0.298	0.702	1.000	0.368	0.632	1.000	0.456	0.544	1.000
s	0.247	0.753	1.000	0.338	0.662	1.000	0.422	0.578	1.000
G	0.298	0.295	0.302	0.278	0.310	0.301	0.267	0.293	0.284

vant for the interpretation of the gender inequality, while, on the contrary, different decompositions indicate an influence on gender inequality. Table 3 illustrates the p_i, s_i and G_i for two subgroups: for the educational level we compare the up-to-elementary-school group to the group with a university degree, for the geographical area the group living in the north to the group living in the south or islands.

The related decompositions of the Gini index for the analysis of the gender gap are shown in Table 4. The comparison between the decompositions suggests that the educational level influences the gender income inequality more than the geographical area. We also confirm the decrease of the importance of G_w, together with an increase of the relevance of G_t and G_b, especially for the more affluent subgroups.

4 Conclusions

The decomposition of an inequality index can be extremely useful into the study of the gender income inequality, where the decomposition refers to the female and male subpopulations. The analysis of the income distribution of the Italian house-

Table 4 Income inequality decomposition by gender of the head of the household

	Gw/G	Gb/G	Gt/G	Gw/G	Gb/G	Gt/G
	20% bottom income			20% top income		
1993	0.544	0.259	0.197	0.717	0.168	0.114
2004	0.509	0.252	0.239	0.591	0.250	0.159
2014	0.500	0.263	0.237	0.534	0.267	0.199
	up to elementary school			with university degree		
1993	0.531	0.288	0.181	0.734	0.145	0.121
2004	0.498	0.273	0.229	0.583	0.272	0.145
2014	0.510	0.270	0.221	0.495	0.328	0.177
	south islands			north		
1993	0.648	0.212	0.140	0.734	0.145	0.121
2004	0.526	0.277	0.197	0.547	0.277	0.177
2014	0.500	0.303	0.197	0.505	0.307	0.187

holds shows how gender gap explained 24.9% of total inequality in 1993, rising to 29.7% in 2014. The scenario is even worse when evaluating inequality attributable to the differences between female and male subpopulations without the traditional assumption of null inequality within: in this case gender gaps are accountable for 37.9% of total inequality in 1993, rising to 42.9% in 2014. Inequality decomposition also allows to evaluate the relation between gender and other inequality factors: educational level of the head of the household and geographical area of residence are taken into account, with the former showing a greater influence on gender income inequality.

References

1. Costa, M.: Transvariation and inequality between subpopulations in the Dagum's Gini index decomposition. Metron, **67**, 134–120 (2009)
2. Costa, M.: The evaluation of the inequality between population subgroups. In Verde, R., Racioppi, F., Petrucci, A.: Statistics and Data Science: new challenges, new generations. Firenze University Press, Florence, 313-318 (2017)
3. Dagum, C.: Gini ratio. In: The New Palgrave Dictionary of Economics. Mac Millian Press, London (1987)
4. Dagum, C., Zenga M.: Income and Wealth Distribution, Inequality and Poverty. Springer, Berlin (1990)
5. Dagum, C.: A new decomposition of the Gini income inequality ratio. Empirical Economics, **22**, 515–531 (1997)
6. Elbers, C., Lanjouw, P., Mistiaen J.A., Ozler, B.: Reinterpreting Between-Group Inequality. Journal of Economic Inequality, **6**, 231–245 (2008)
7. Giorgi, G.M.: Gini's scientific work: an evergreen. Metron, **63**, 299–315 (2005)
8. Giorgi, G.M.: The Gini inequality index decomposition. An evolutionary study. In Deutsch, J., Silber, J.: The measurement of individual well-being and group inequalities. Routledge, London (2011)
9. Yitzhaki, S., Lerman, R.: Income stratification and income inequality. Review of Income and Wealth, **37**, 313–329 (1991)