Analyzing the Gender Gap in Poland and Italy, and by Regions

Greselin, Francesca; Jędrzejczak, Alina

Analyzing the Gender Gap in Poland and Italy, and by Regions

Francesca Greselin[1] · Alina Jędrzejczak[2]

Abstract High-income inequality, accompanied by substantial regional differentiation, is still a great challenge for social policymakers in many European countries. One of the important elements of this phenomenon is the inequality between income distributions of men and women. Using data from the European Union Statistics on Income and Living Conditions, the distributions of income for Italy and Poland were compared, and the gender gap in these countries was assessed. No single metric can capture the full range of experiences, so a set of selected tools were adopted. The Dagum model was fitted to each distribution, summary measures, like the Gini and Zenga inequality indices, were evaluated, and the Zenga curve was employed to detect changes at each income quantile. Afterward, empirical distributions were compared through a relative approach, providing an analytic picture of the gender gap for both countries. The analysis moved beyond the typical focus on average or median earnings differences, towards a focus on how the full distribution of women's earnings relative to men's compares. The analysis was performed in the different macroregions of the two countries, with a discussion of the results. The study revealed that income inequality in Poland and Italy varies across gender and regions. In Italy, the highest inequality was observed in the poorest region, i.e. the islands. On the contrary, in Poland, the highest inequality occurred in the richest region, the central one. The relative distribution method was a powerful tool for studying the gender gap.

Keywords Income inequality · Gender gap · Gini index · New Zenga index · Relative distribution method · Dagum · Italy · Poland

JEL Classification C1 · C46 · D63

✉ Alina Jędrzejczak
alina.jedrzejczak@uni.lodz.pl

[1] University of Milano-Bicocca, Milan, Italy

[2] University of Łódź, Łódź, Poland

Introduction

Income disparity is still growing in the Organisation for Economic Cooperation and Development (OECD) countries and it reached its highest level in the past half-century. Several studies were conducted on the issue (e.g., OECD 2011, 2015).

The trend of rising inequality has become a priority for policymakers and calls for the analysis of various aspects of income inequality, including measurement and decomposition by regional areas, income sources and recently across genders. For social and economic policies, it seems interesting to compare income inequality across the European Union (EU) countries and regions.

The focus of the present paper is on income distributions across Poland and Italy because the aforementioned countries represent different economic backgrounds. Poland still suffers the effects of the transitioning from a centrally-planned economy to a market-based economy, and Italy is a former well-established market economy. Moreover, according to the Tárki European Social Report (TÁRKI, 2009), a study on intolerance to income inequality across countries confirmed a markedly lower level of acceptance of inequality in the post-socialist bloc than in the other European countries.

Nowadays, there is open debate about the observed discrepancy between income distributions of men and women. Women make up almost half of the workforce, yet, on average, they continue to have lower incomes than men. Gender equality is one of the fundamental values of the EU. The EU is dedicated not only to defending this right but also to promoting gender equality within the member states and across the world. This was the core aim of the European Commission's *Strategic engagement for gender equality 2016-2019*.report (European Commission, 2016). Naturally, gender can affect people's working lives in different ways. Therefore, even if it is illegal for employers in EU countries to pay men and women different amounts for doing the same job, there are many other reasons why, on average, substantial income differences between men and women can be observed. Differences in pay are caused by many concurring factors, and an interesting in-depth analysis can be found in Blau and Kahn (2003). Occupational segregation is perhaps the main reason. Men are more prevalent in higher-paid industries, while women are mostly in lower-paid industries. The differences in remuneration across industry sectors all influence the gender pay gap (Blau and Kahn, 2000). There is also vertical segregation. Few women work in senior and better-paying positions. Moreover, in many countries, there is still ineffective equal-pay legislation regulating women's overall paid working hours. The distribution of the workforce across working hour bands is generally more even for women than for men, due largely to the higher incidence of part-time work among women. Thus, a greater proportion of women work fewer hours. Finally, some barriers to entry into the labor market are related to the education level and single parenting rate (Leythienne and Ronkowski, 2018). Gender equality also means equal economic independence for women and men is still observed. It refers to equality in decision-making, and, in a wider setting, requires equal dignity, integrity and the ending of gender-based violence.

In this paper, an analysis of data from household budget surveys carried out in Poland and Italy in 2015 is presented, with special attention to the income gender gap within each country and for macroregions. The calculations were based on microdata from the European Union Statistics on Income and Living Condition (EU-SILC) (Eurostat, 2015).

The original contribution of the present research is the variety of chosen metrics, complemented by the descriptive and inferential tools. More in detail, after estimating the Dagum model for each distribution, summary measures, like the Gini and Zenga

inequality indices, were evaluated and the Zenga curve was adopted to detect changes at each income quantile. Afterward, a comparison of the empirical distributions was developed by a relative approach, providing an analytic picture of the gender gap for both countries. The analysis moves beyond the typical focus on average or median earnings differences, toward a perspective on how the full distribution of women's earnings relative to men's compares. Finally, the analysis was carried out in the different macroregions of the two countries and the results were discussed.

Results of EU-SILC show that the popular Gini inequality indexes for net household incomes in the two countries were rather similar and equaled 0.34 and 0.35, respectively. Nevertheless, the comparative studies conducted by Jedrzejczak (2015) and Zenga and Jedrzejczak (2020) revealed substantial differences in inequality patterns much higher for the Italian macroregions when compared to the Polish ones. In particular, a relatively strong negative correlation was observed between gross domestic product (GDP) per capita and income inequality measured by the Gini index in the Italian regions, while in Poland this correlation was slightly positive. As a result, in Italy, the highest-inequality levels occurred in the poorest regions, while in Poland, the opposite was observed.

Methodology and Notation

First a brief review of the Dagum model for fitting the economic data is provided. Then, the inequality curves and indices and computational details are presented. Finally the notion of relative distribution is defined, providing a powerful tool to perform the comparison among genders.

Density Estimation

The Dagum distribution was chosen to model the income data. The choice of the underlying distribution was based on the well-known economic foundations and a large body of diversified empirical pieces of evidence supporting the aforementioned model as an excellent candidate to fit observed income distributions (Kleiber and Kotz, 2003). Recall that F belongs to the Dagum family if its density function is given by

$$f(x; a, b, p) = \frac{apx^{ap-1}}{b^{ap}\left[1 + \left(\frac{x}{b}\right)^a\right]^{p+1}}, \quad x > 0,$$

for some $a, b, p > 0$. Notice that the first moment of a Dagum distribution is finite if and only if $a > 1$. Therefore, the inequality measures considered in this paper (and introduced in the following subsection) are defined under this condition. The parameters a and p are shape parameters, while b is a scale parameter. This model allows for various degrees of positive skewness and leptokurtosis. Moreover, it has built-in flexibility to be unimodal, to approximate income distributions, or to be zeromodal to describe wealth distributions. The shape parameters are related to inequality, Lorenz and first-stochastic dominance. For example, let $F_1 = f(a_1, b_1, p_1)$ and $F_2 = f(a_2, b_2, p_2)$ represent two Dagum distributions. Then the necessary and sufficient conditions for Lorenz dominance (i.e. non-intersecting Lorenz curves) are $a_1 p_1 \leq a_2 p_2$ and $a_1 \leq a_2$. For more details on this distribution, in the frame-

work of economic size distributions, see Kleiber and Kotz (2003, chap. 6.3) and references therein. Finally, the cumulative distribution function (CDF) for the Dagum is given by

$$F(x; a, b, p) = \left[1 + \left(\frac{x}{b}\right)^{-p}\right]^{-a}, \quad x > 0.$$

Given an independent and identically distributed sample $x_1, x_2, ..., x_n$, the likelihood equations for the Dagum family are given by

$$
\begin{cases}
\dfrac{n}{a} + p \displaystyle\sum_{i=1}^{n} \ln\left(\dfrac{x_i}{b}\right) - (p+1) \displaystyle\sum_{i=1}^{n} \dfrac{\ln\left(\frac{x_i}{b}\right)}{1 + \left(\frac{b}{x_i}\right)^a} = 0 \\[3ex]
np - (p+1) \displaystyle\sum_{i=1}^{n} \dfrac{1}{1 + \left(\frac{b}{x_i}\right)^a} = 0 \\[3ex]
\dfrac{n}{p} + a \displaystyle\sum_{i=1}^{n} \ln\left(\dfrac{x_i}{b}\right) - \displaystyle\sum_{i=1}^{n} \ln\left[1 + \left(\dfrac{x_i}{b}\right)^a\right] = 0.
\end{cases}
\tag{1}
$$

However, no explicit solution to this system is known (e.g., Kleiber 2008). An effective package for Dagum estimation is available on the CRAN R package repository (CRAN, 2018). It solved numerically the maximum likelihood (ML) optimization with very good model fit.

Inequality Curves and Indices

Let $Z \geq 0$ be a random variable representing gross or net incomes as well as taxes. Let $F(Z)$ be its cumulative distribution function (CDF), $\mu = E(Z)$ the mean value, and $F^{-1}(p) = inf\{z : F(Z) \geq p\}$ its quantile function for $0 < p < 1$.

The Lorenz curve $\{(p, L_F(p)) | p \in [0, 1]\}$, plots the cumulative share of Z, say $L_F(p)$,

$$L_F(p) = \frac{\int_0^p F^{-1}(s)ds}{\int_0^1 F^{-1}(s)ds} = \frac{1}{\mu}\int_0^p F^{-1}(s)ds \tag{2}$$

versus the cumulative share of the population, p. In the ideal case of perfect equality (that is, a society in which all people have the same income), the share of incomes equals the share of the population, so that $L_F(p) = p$, for all $0 < p < 1$. In this case, the Lorenz curve is the diagonal line from $(0, 0)$ to $(1, 1)$. On the other hand, the lower the share of income $L_F(p)$ held by the share of income earners p, the higher the inequality. In the ideal case of perfect inequality (that is, a society in which all people but one have an income of nil), the share of incomes equals zero for $0 \leq p < 1$, so that $L_F(p) = 0$, and $L_F(1) = 1$ only for $p = 1$. In this setting, it is very natural to express the degree of inequality through the deviation of the actual Lorenz curve from the diagonal line. The Gini index (Gini 1914) is twice the area between the equality line and the Lorenz curve:

$$G_F = 2\int_0^1 (p - L_F(p))dp. \tag{3}$$

The Gini index can be rewritten as

$$G_F = 2\int_0^1 \frac{\mu - \bar{\mu}(p)}{\mu} \, p \, dp, \quad \text{where} \quad \bar{\mu}(p) = \frac{1}{p}\int_0^p F^{-1}(s)ds. \tag{4}$$

Three issues arise when using the Gini index. First, the weight, p, gives the lowest importance to the more critical comparisons. Second, the weight, p, gives more emphasis to the less informative comparisons. Third, the considered groups, which μ and $\bar{\mu}(p)$ refer to, are overlapped. These considerations gave rise to many attempts for modifying the Gini index (see Greselin (2014) for a review).

Observing the noticeable increase in disparities between less fortunate and more fortunate individuals, Zenga (2007) introduced a new inequality curve, $Z_F(p)$, obtained by contrasting the average income of the poorer $p\%$ bottom earners, that is $\bar{\mu}(p)$, with the amount that is held, on average, by the richest top earners, i.e. the remaining $(1 - p)\%$ of the population, that is

$$\overset{+}{\mu}(p) = \frac{1}{1 - p} \int_p^1 F^{-1}(s)\,ds. \tag{5}$$

Therefore, Zenga defined the inequality curve $\{(p, Z_F(p))|p \in [0, 1]\}$, where

$$Z_F(p) = \frac{\overset{+}{\mu}(p) - \bar{\mu}(p)}{\overset{+}{\mu}(p)} \quad \text{for} \quad 0 < p < 1. \tag{6}$$

The methodology proposed by Zenga keeps in mind that the notions of poor and rich are relative to each other and summarizes, in a single measure, the amount of inequality in the population. A measure of economic inequality can be defined by calculating the area beneath the Zenga's curve:

$$Z_F = \int_0^1 Z_F(p)\,dp.$$

It is worth recalling that the Zenga index follows the axiomatic approach. It obeys the Pigou-Dalton transfer principle (Dalton 1920; Pigou 1920), it is scale invariant, and it has the desired properties of anonymity and decomposability.

The next aim is to compare the two indices:

$$G_F = 2\int_0^1 (p - L_F(p))dp = 2\int_0^1 \frac{\mu - \bar{\mu}(p)}{\mu}\ p\,dp,$$

and

$$Z_F = \int_0^1 Z_F(p)dp = \int_0^1 \frac{\overset{+}{\mu}(p) - \bar{\mu}(p)}{\overset{+}{\mu}(p)}\,dp.$$

Both indices are relative measures of inequality, but the latter, compared to the former, has the following interesting features. First, the Zenga index gives the same weight to each comparison along the entire distribution. Second, the considered groups, which $\bar{\mu}(p)$ and $\overset{+}{\mu}(p)$ refer to in the definition of Z_F, are exhaustive and disjoint. Finally, the curve $Z_F(p)$ has neither forced values at the endpoints nor is constrained to being non-decreasing and concave on the interval $[0, 1]$, as is the case for the Lorenz curve $L_F(p)$. Turning back to the expressions of the inequality measures in the Dagum model, the Lorenz curve and the Gini index have the following form (Dagum 1977):

$$L(t; a, b, p) = B\left(t^{1/p}; p + \frac{1}{a}; 1 - \frac{1}{a}\right), \quad 0 < t < 1 \tag{7}$$

and

$$G(p, a, b) = \frac{\Gamma(p)\Gamma(2p + 1/a)}{\Gamma(2p)\Gamma(p + 1/a)} - 1. \tag{8}$$

In Equation (7), $B(t; a; b)$ indicates the beta CDF, while $\Gamma(x)$ indicates the Gamma function in Equation (8). The substitution of the Lorenz curve in the formula of Zenga's index yields

$$Z(p, a, b) = \int_0^1 \frac{t - B\left(t^{1/p}; p + \frac{1}{a}; 1 - \frac{1}{a}\right)}{t\left[1 - B\left(t^{1/p}; p + \frac{1}{a}; 1 - \frac{1}{a}\right)\right]} dt. \tag{9}$$

As noted at the beginning of this section, the Lorenz curve and the two inequality measures are defined if and only if $a > 1$.

Finally, bearing in mind that the aim is to employ such measures on survey data, the estimators for the aforementioned curves and indices are needed. From the random sample $(X_1, \ldots, X_n)$, the empirical Lorenz curve can be rephrased as: $\{(i/n, L_n(i/n)\}$:

$$L_n\left(\frac{i}{n}\right) = \frac{\int_0^{i/n} F_n^{-1}(s) ds}{\int_0^1 F_n^{-1}(s) ds} = \frac{\sum_{j=1}^i X_{(j)}}{\sum_{j=1}^n X_{(j)}}, \tag{10}$$

where $\bar{X}$ denotes the sample mean and $X_{(j)}$ denotes the j-th order statistics. $L_n(p)$ is the share of the total amount of income owned by the least fortunate $p \times 100\%$ of the sample. The Gini index, evaluated from the sample, is

$$G_n = \frac{2}{n-1} \sum_{i=1}^{n-1} \left\{ \frac{i}{n} - \frac{\sum_{j=1}^i X_{(j)}}{\sum_{j=1}^n X_{(j)}} \right\}. \tag{11}$$

The empirical Zenga measure (Greselin and Pasquazzi, 2009) can be obtained by replacing the population, CDF F, by its empirical counterpart $\widehat{F}_n$:

$$Z_n = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{\frac{1}{n-i} \sum_{j=i+1}^n X_{(j)} - \frac{1}{i} \sum_{j=1}^i X_{(j)}}{\frac{1}{n-i} \sum_{j=i+1}^n X_{(j)}}. \tag{12}$$

Relative Distributions

Let Y_0 be a random variable representing a measurement for a population (e.g., income, wealth, hourly wages etc). In the following, the population that generated Y_0 is denoted as the reference population (in the present case, the male income distribution can be set as the reference). Denote the CDF of Y_0 by $F_0(y)$. Suppose also to observe another measurement of Y from a different population. The population that generated Y is denoted as the comparison population. It is assumed that Y has CDF $F(y)$. Typically, Y is the measurement for a separate group (in the present case, the female income distribution), but in other cases, it can also be the same group at a later period.

The objective is to study the differences between the comparison distribution and the reference distribution. Unless explicitly mentioned, both F and F_0 are assumed to be absolutely continuous with common support. The relative distribution of Y to Y_0 is defined (Handcock and Morris, 2006) as the distribution of the random variable:

$$R = F_0(Y).$$

It is worth remarking that R is obtained from Y by transforming the latter by the CDF for Y_0, namely F_0. This has also been called the grade transformation (Ćwik and Mielniczuk, 1989). If the two distributions, Y and Y_0, are identical, then the random variable, R, is uniformly distributed on the unit interval. The deviation from the uniform pattern can be interpreted as the gap between the distributions. While this transformation is not widely used or understood in the social sciences, it is a very useful one because R measures the relative rank of Y compared to Y_0.

As a random variable, R has both a CDF and a probability density function. In particular, R has CDF G, such that $G(p) = F(F_0^{-1}(p))$ for $0 \leq p \leq 1$. The relative CDF, $G(p)$, can be interpreted as the proportion of the comparison group whose attribute lies below the pth quantile of the reference group. Note that even though the relative CDF is explicitly scaled in terms of quantiles, the implicit unit of comparison is the value of the attribute on the original measurement scale, with $y_p = F_0^{-1}(G(p))$ representing the cut point.

Results

A brief description of the European Union Statistics on income and Living Conditions (EU-SILC) data (Eurostat, 2015) begins this section. Results from the gender-gap analysis follow.

EU-SILC 2015 Data

The EU-SILC is an instrument aiming at collecting timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. This instrument is anchored in the European Statistical System (ESS). The variable of interest is gross personal income, measured yearly, which consists of: employee-gross income *PY010G*, cash benefits from self-employment *PY050G*, gross individual pension plans *PY080G*, gross-unemployment benefits *PY090G*, old-age benefits *PY100G*, survival benefits *PY110G*, disability benefits *PY130G*, and education-related allowances *PY140G*.

Gross-employee income includes the following items: wages and salaries paid in cash for time worked or work done in main and any secondary or casual job(s), remuneration for time not worked (e.g. holiday payments), enhanced rates of pay for overtime, fees paid to directors of incorporated enterprises, piece-rate payments, payments for fostering children, commissions, tips and gratuities, supplementary payments (e.g. thirteenth-month payment), profit sharing and bonuses paid in cash, additional payments based on productivity, allowances paid for working in remote locations (regarded as part of the conditions of the job), allowances for transportation to or from work (Eurostat, 2015).

Tables 1 and 2 contain descriptive statistics of EU-SILC 2015 data, for Poland and Italy, respectively. Acronyms, such as PL1, followed by their names, correspond to the name of a region, as given by the Nomenclature of Territorial Units for Statistics (NUTS)[1], i.e., the statistical classification of economic units in the European Community. From the same standards, "PL" stands for Poland, "IT" stands for Italy.

---

[1]The term NUTS is derived from French: *Nomenclature des Unités Territoriales Statistiques*.

Table 1 Descriptive statistics of gross-personal income data for Poland, at the regional (NUTS1) and national levels

Gender / Region	Sample size	Mean [Euro]	St. Dev. [Euro]	Coeff. of variation	Quantile 0.1 [Euro]	Quantile 0.9 [Euro]
Males	9,933	8,161.19	6,653.44	81.53	2,592.31	15,056.28
PL1 Central	1,671	8,706.55	8,254.98	94.81	2,736.37	16,629.50
PL2 Southern	1,643	8,856.33	6,803.08	76.82	2,897.50	16,026.34
PL3 Eastern	2,144	7,238.46	5,706.81	78.84	2,314.73	13,089.05
PL4 North-West	1,619	7,840.48	5,700.71	72.71	2,520.37	14,407.67
PL5 South-West	1,062	8,832.24	7,105.86	80.45	2,890.47	16,531.06
PL6 Northern	1,789	8,011.59	6,258.61	78.12	2,543.79	14,748.23
Females	10,418	5,971.75	4,353.53	72.90	2,433.64	10,684.22
PL1 Central	1,840	6,465.56	5,068.22	78.39	2,379.03	12,217.93
PL2 Southern	1,636	6,026.90	4,126.82	68.47	2,379.03	10,686.66
PL3 Eastern	2,287	5,628.20	4,251.29	75.54	2,381.86	10,208.30
PL4 North-West	1,682	5,938.50	4,223.81	71.13	2,453.39	10,551.47
PL5 South-West	1,166	6,089.52	4,029.06	66.16	2,670.94	10,458.62
PL6 Northern	1,807	5,809.89	4,167.98	71.74	2,379.90	10,452.55
Poland	20,351	7,040.27	5,701.24	80.98	2,418.85	13,269.61

Notes: Acronyms, such as PL1, followed by their names, correspond to the name of a region, as given by NUTS, i.e., the statistical classification of economic units in the European Community, where "PL" stands for Poland. Source: Own calculations data from EU-SILC (Eurostat, 2015)

Table 2 Descriptive statistics of gross-income personal data for Italy, at the regional (NUTS1) and national levels

Gender / Region	Sample size	Mean [Euro]	St. Dev. [Euro]	Coeff. of variation	Quantile 0.1 [Euro]	Quantile 0.9 [Euro]
Males	15,636	29,848.67	29,694.50	99.48	8,049.5	52,943.0
IT1 North-West	3,999	33,525.19	31,776.78	94.78	10,500.0	58,407.6
IT2 North-East	3,832	33,296.54	30,983.48	93.05	9,805.0	59,568.1
IT3 Central	3,677	29,648.77	25,267.16	85.22	8,872.0	52,767.4
IT4 Southern	3,004	23,070.73	22,050.75	95.58	5,706.8	41,165.5
IT5 Insular	1,124	23,782.30	42,048.10	176.80	4,684.1	41,305.5
Females	14,195	19,623.46	16,253.38	82.83	5,192.2	35,126.6
IT1 North-West	3,738	20,775.53	17,509.70	84.28	6,455.9	35,570.2
IT2 North-East	3,547	21,785.86	17,332.07	79.56	6,669.0	39,280.4
IT3 Central	3,417	15,923.52	12,648.31	79.82	5,365.8	35,641.4
IT4 South	2,574	15,634.07	12,648.31	80.90	3,627.0	30,838.8
IT5 Insular	919	16,552.74	14,363.40	86.77	3,627.0	31,297.4
Italy	29,831	24,983.03	24,777.9	99.18	6,442.0	45,188.0

Notes: Acronyms, such as IT1, followed by their names, correspond to the name of a region, as given by NUTS, i.e., the statistical classification of economic units in the European Community, where "IT" stands for Italy. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

Gender-gap Analysis for Poland and Italy, at National and Regional Level

The analysis of the data began with obtaining useful summary measures, like the median and 98% quantile of income, for regions and at the country level, when partitioning the population by gender. Afterward, parameter estimation for the Dagum model was performed on each data subsample, and the Gini and Zenga indices were calculated using Equations (11) and (12). Also, the latter results are reported in Tables 3 and 4, respectively, for Poland and Italy.

From Tables 3 and 4, it is worth noting that the differences between median incomes were higher in Italy, where males were more affluent than females by 47% on average (in Poland by 34 %). In contrast, the biggest discrepancy between income-inequality levels was observed in Poland. The Gini index was 9% higher for men (for Italy the difference was smaller: 1.8%). For Italian men, the Zenga index was $Z = 0.747$. This means that, on average, the mean income of the p% poorest men was 1-74.7=25.3% lower than the mean income of the (1-p)% richest fraction of men in the sample. Similarly, for Italian women, the Zenga index had the value Z=0.744, so that the mean income was, on average, 25.6% lower than the mean income of the richest fraction of women in the sample.

Interesting findings were obtained further in terms of regional differences, both within and between countries. In Poland, the differences between the regions were much smaller

Table 3 Income data for Poland, at the regional (NUTS1) and national levels

| Macro Region		Median	98%	Dagum model parameters			Inequality measures	
		[Euro]	[Euro]	Shape a	Shape p	Scale b	Gini	Zenga
PL1	M	6,678	32,106	2.759	0.690	8,198	0.395	0.741
Central	F	5,118	21,863	2.969	0.662	6,446	0.400	0.559
PL2	M	7,285	27,948	3.149	0.620	9,243	0.380	0.707
Southern	F	5,024	18,793	3.456	0.592	6,490	0.335	0.681
PL3	M	6,034	22,743	3.230	0.532	8,167	0.369	0.720
Eastern	F	4,631	17,713	3.235	0.689	5,541	0.340	0.684
PL4	M	6,568	24,144	3.299	0.531	8,927	0.363	0.713
North-West	F	4,983	19,719	3.491	0.579	6,380	0.335	0.681
PL5	M	6,868	30,961	2.837	0.711	8,385	0.382	0.727
South-West	F	5,115	18,084	3.638	0.633	6,361	0.313	0.653
PL6	M	6,344	26,501	3.201	0.544	8,881	0.370	0.720
Northern	F	4,884	18,348	3.315	0.612	6,103	0.345	0.691
Poland	M	6,614	26,855	3.067	0.592	8,634	0.374	0.723
	F	4,954	19,079	3.291	0.633	6,162	0.344	0.689

Notes: Acronyms, such as PL1, followed by their names, correspond to the name of a region, as given by NUTS, i.e., the statistical classification of economic units in the European Community, where "PL" stands for Poland. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

Table 4 Income data for Italy, at the regional (NUTS1) and national levels

Macro Region		Median	98%	Dagum model parameters			Inequality measures	
		[Euro]	[Euro]	Shape *a*	Shape *p*	Scale *b*	Gini	Zenga
IT1	M	26,558	115,004	3.155	0.543	36,268	0.381	0.732
North-West	F	18,336	70,201	3.398	0.448	26,847	0.375	0.729
IT2	M	26,924	119,106	3.176	0.574	35,119	0.378	0.726
North-East	F	17,954	69,697	3.546	0.423	26,063	0.375	0.731
IT3	M	24,312	105,968	3.182	0.518	32,272	0.382	0.734
Central	F	17,008	67,820	3.483	0.402	26,041	0.383	0.739
IT4	M	19,459	78,504	3.361	0.417	29,145	0.397	0.753
Southern	F	11,988	43,996	3.140	0.419	20,472	0.401	0.756
IT5	M	19,082	83,811	3.171	0.388	30,242	0.446	0.794
Insular	F	13,291	53,752	3.411	0.344	23,835	0.414	0.770
Italy	M	24,250	106,527	3.173	0.491	34,161	0.395	0.747
	F	16,513	64,208	3.442	0.401	25,673	0.388	0.744

Notes: Acronyms, such as IT1, followed by their names, correspond to the name of a region, as given by NUTS, i.e., the statistical classification of economic units in the European Community, where "IT" stands for Italy. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

than in Italy, when comparing average-income levels. In contrast, regional-income inequalities were more variable in Poland. Moreover, in Italy, the highest inequality occurred in the poorest region, the islands, while in Poland the highest inequality occurred in the richest region, the central one. Table 5 shows the results of a one-sided significance test for the two-income means (men vs. women) for Poland and Italy, to complete the information.

Table 5 One-sided significance tests for the equality of the mean across genders, at the regional (NUTS1) and national level, for Poland (left) and Italy (right)

Country / Region	test statistic	p-value	Country/ Region	test statistic	p-value
Poland	27.64	0.0000	*Italy*	37.34	0.0000
PL1 Central	9.58	0.0000	IT1 North-West	22.04	0.0000
PL2 Southern	14.41	0.0000	IT2 North-East	19.88	0.0000
PL3 Eastern	10.60	0.0000	IT3 Central	29.23	0.0000
PL4 North-West	10.86	0.0000	IT4 Southern	15.71	0.0000
PL5 South-West	11.06	0.0000	IT5 Insular	5.39	0.0000
PL6 Northern	12.40	0.0000			

Notes: Acronyms, such as PL1 or IT1, followed by their names, correspond to the name of a region, as given by NUTS, i.e., the statistical classification of economic units in the European Community, where "PL" and "IT" stand for Poland and Italy, respectively. Source: Own calculations, data from EU-SILC 2015
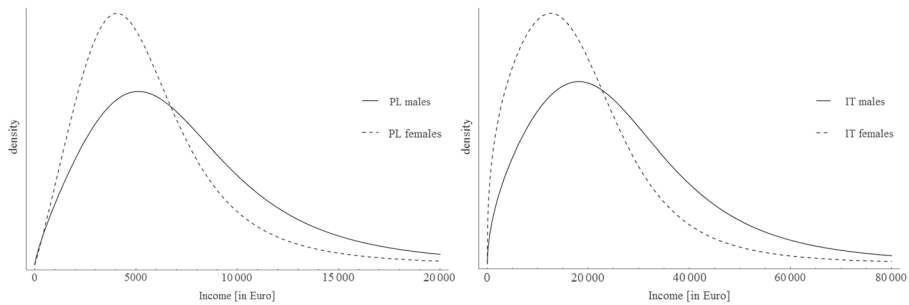
Fig. 1 Dagum densities estimated on male and female income data for Poland (left) and Italy (right). Source: Own calculations using data from EU-SILC (Eurostat, 2015)

Figure 1 presents the estimated Dagum densities for male and female groups in Poland and Italy. The juxtaposed income distributions (for males and females) had a similar shape in both countries. However, the distributions of men were more dispersed and shifted to the right, when compared to the distributions of women.

The density curves, as well as the summary measures presented in Tables 3 and 4, were inadequate to property address all unanswered questions about gender gaps. The comparison was more informative when employing a detailed inequality analysis based on inequality curves. The study contrasted the usefulness of the Lorenz and the Zenga curves in depicting the overall income inequality in Poland and Italy, reported in Figure 1 in the Online Supplemental Appendix and here in Figure 2, respectively.

The Lorenz and Zenga curves provided an in-depth comparison of income inequality among and between men and women. At first glance, it can be seen that the Lorenz curves (Online Supplemental Appendix Figure 1) were close to each other and it was rather difficult to determine the level and direction of the discrepancy between gender groups. This was particularly difficult for Italy, where the Lorenz curves intersect. Since the Zenga curve refers to non-overlapping and opposite groups in the population, the Zenga curve was more sensitive to inequality changes at each point of the income range. In Poland, the Zenga curve for males lies visibly above the one for females, showing higher-income inequality for the latter group throughout the entire income distribution. In Italy, the Zenga curves revealed a much smaller discrepancy between gender groups. Additionally, the curves intersect, so the relationship between inequality levels changes. Men were more unequal within higher-income groups, while for small- and intermediate-income levels, their distribution was more homogeneous compared to women. As an example, the highlighted point in the Zenga curve in Figure 2 says that, for Italian women, the poorest 72% had mean income equal to 33% of the mean income of the richest 28%.

The question arises, for the sake of a profound gender-gap analysis, is it enough to consider inequality curves along with only a few selected income quantiles (median and 98th percentile)? The phenomenon of gender gap depends not only on the differences between the median values for men and women but also on all the deviations observed over the entire income range, i.e., for each quantile of the income distributions. The relative distribution method is a relevant tool to complete the analysis. It contrasts the gender groups all along their income distributions.

Figure 3 shows the gender gap results for Poland and Italy, respectively, obtained via the relative distribution approach. The curve of the relative income distribution provides rich and detailed information. Each point on the curve has a precise interpretation. For instance,
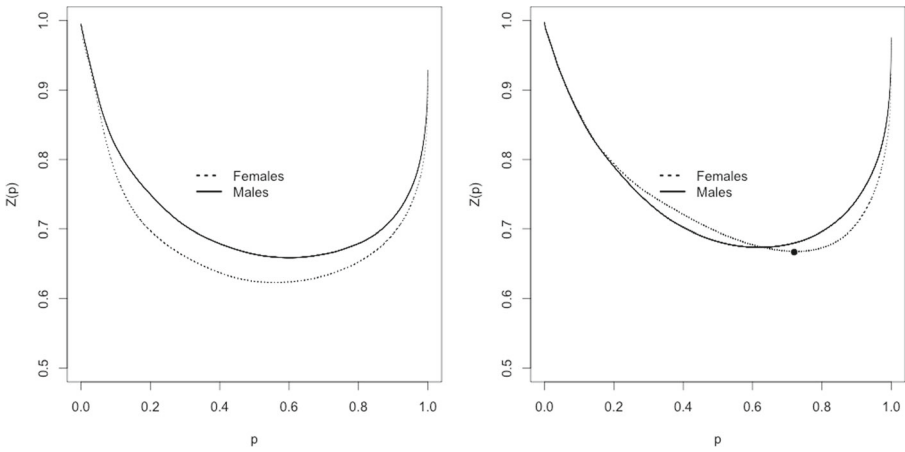
Fig. 2 Zenga curves for Poland (left panel) and Italy (right panel) by gender. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

in the right panel of Figure 3, at the third decile of the male earnings distribution, that is $p = 0.3$, $G(0.3) = 0.54$. This means that approximately 54% of women earn less than the third decile male income. One of the peculiarities of the relative graphs is that the distance between Euro values on the right-hand scale is measured in units of persons rather than in Euro. Therefore, the distance between 0 and 10,000 Euro, is larger than the distance between 40,000 and 50,000 Euro, because a larger fraction of people have an income falling in the former range of incomes than in the latter. In other words, the information captured by the density curves in Figure 1 is conveyed here, along with a straight-through assessment of the gender gap.

A comparative analysis of the relative distributions for Poland and Italy, by regions, showed significant differences between the countries. For a more in-depth analysis, the complete results are available in Figures 2, 3, 4, 5, 6 and 7 in the Online Supplemental Appendix.
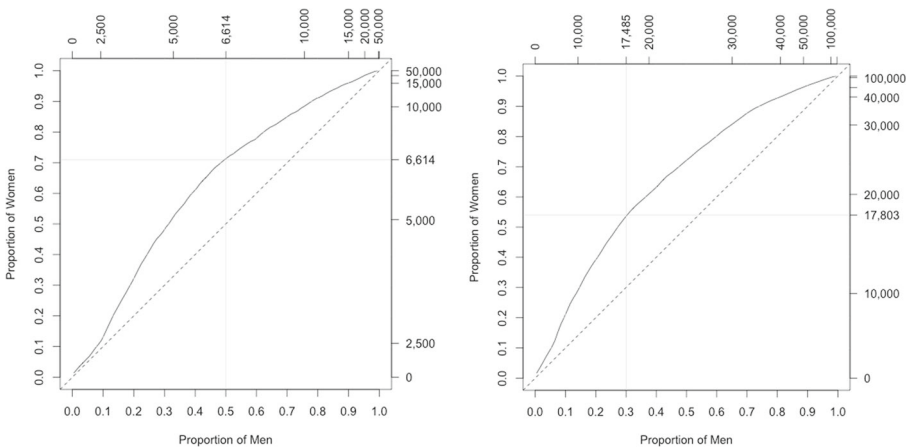

Fig. 3 Women income quantiles versus men income quantiles (left panel: Poland; right panel: Italy). Source: Own calculations using data from EU-SILC (Eurostat, 2015)
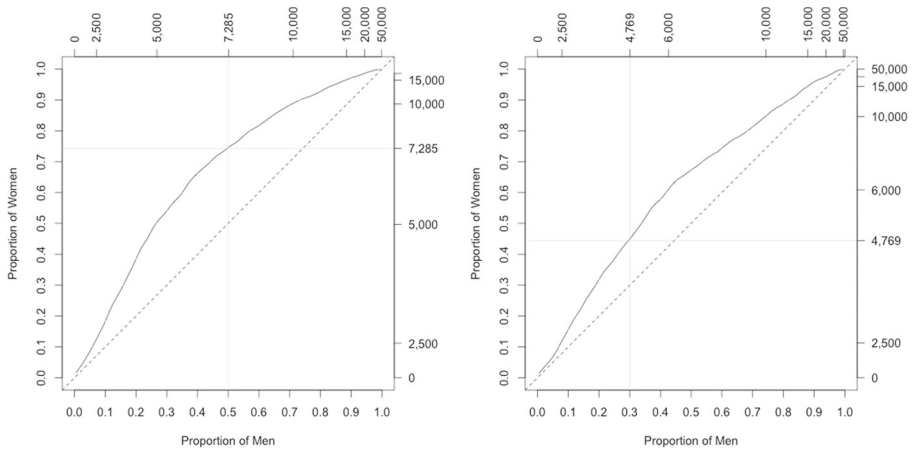
Fig. 4 Gender gap in selected Polish regions: southern (left panel) and central region (right panel), showing the maximum and the minimum gender gap, respectively. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

Some considerations arose and are worth noting here, in the main paper. First, the gender gap in Italy was quite diverse by region, while, in Poland, it was more stable. An essential observation was that in Italy, the largest-income gap was observed in the northeast, while in Poland the largest income gap occurred in the southern region. Both regions have the maximum mean income within the corresponding countries. Related results were reported in the left panels of Figures 5 and 4, respectively. The relative distribution of the female income with respect to the male income in Figure 4 (left panel) showed that, in the southern Polish region, the median income for men, say 7,285 euro, was equal to the third quantile for women. In the central Polish region Figure 4 right panel, 45 percent of the women had an income lower than the men's third decile.
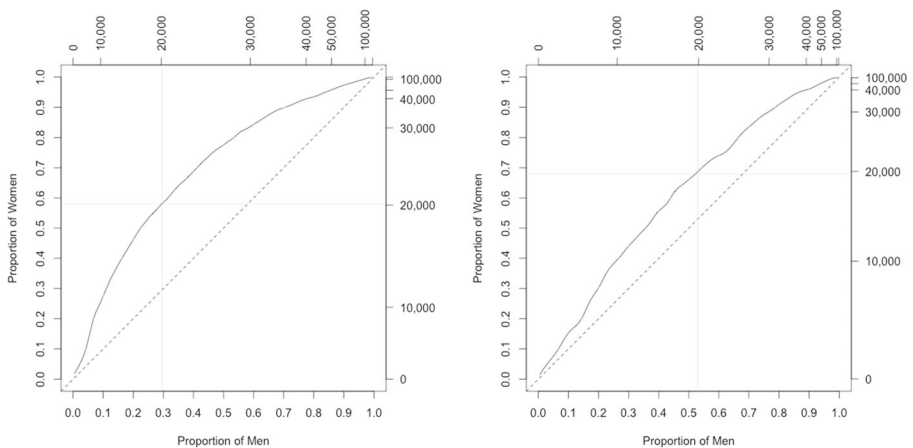
Fig. 5 Gender gap in selected Italian regions: northeastern region (left panel) and islands (right panel), showing the maximum and the minimum gender gap, respectively. Source: Own calculations using data from EU-SILC (Eurostat, 2015)

In both countries, in contrast to some expectations, the regions with the largest inequalities had the smallest gender gaps. In Poland, it was the central region, and in Italy, this occurred in the islands (right panels of Figures 4 and 5, respectively). In both regions, substantial differences between the means were combined with very high-income inequality among men.

In Figure 5 (left panel), with respect to the northeastern Italian region, the third decile of male income, around 20,000 euro, corresponds to the 58th percentile of women's income. In the Italian islands, Figure 5 (right panel) shows that the same amount of 20,000 euro corresponds to the 53rd and 68th percentiles in male and female income distributions, respectively. Obviously, the remarks reported here refer to some specific points of the relative distribution, to provide an instance of the much richer insight and interpretation available from the whole curves.

Conclusions

Income inequality in Italy and Poland varies across gender and regions. In Italy, the highest inequality occurs in the poorest region, the islands. On the contrary, in Poland, the highest inequality is attained in the richest region, the central one. Income inequality in Poland is substantially higher among men, as is the mean income. This finding holds for the whole country, as well as for the regions. In Italy, the behaviour of the Zenga curves reveals a much smaller discrepancy between gender groups, when compared to Poland. Intersecting inequality curves for Italian data confirm that men are more unequal at higher-income groups while for small- and intermediate-income levels, their distribution is more homogeneous compared to women.

In Italy, the poorest regions are those with the lower gender gap and vice versa. In contrast, in Poland, the gender gap was highest in the southern region and lowest in the central (richest) region. In Poland, the gender gap is more stable across regions. Rather unexpectedly, in both countries, the regions with the largest inequalities have the smallest income gaps. In Poland, this contrast occurs in the central region, and in Italy, in the islands. In both regions, substantial differences between the means are reduced by very high-income inequality among men.

The relative distribution methodology was employed as a powerful tool for studying the gender gap in both countries, and by regions. The gender gap can also be measured by standard inequality decomposition and the relative economic affluence.

The present paper offers a descriptive approach. Policy recommendations can only be made based on an analysis of the causes of the gender gap. This might include analysis of the gender gap controlling for years of education, job tenure, marital status, age and number of children. This topic will the object of future work.

References

Blau, F.D., & Kahn, L.M. (2000). Gender differences in pay. *Journal of Economic Perspectives*, *14*(4), 75–99.

Blau, F.D., & Kahn, L.M. (2003). Understanding international differences in the gender pay gap. *Journal of Labor Economics*, *21*(1), 106–144.

CRAN (2018). CRAN Contributed Packages. Available at https://cran.r-project.org/web/packages/.

Ćwik, J., & Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Communications in Statistics-Theory and Methods*, *18*(8), 3057–3069.

Dagum, C. (1977). A new model of personal income distribution: specification and estimation. *Economie appliquée*, *30*(3), 413–437.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, *30*, 348–361.

European Commission (2016). Strategic engagement for gender equality 2016–2019. Available at https://ec.europa.eu/anti-trafficking/eu-policy/strategic-engagement-gender-equality-2016-2019_en.

Eurostat (2015). European Union statistics on income and living conditions (eu-silc). https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions, data retrieved from Eurostat (Last Accessed Feb. 2019).

Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. In: Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti. Anno Accademico 1913-1914 -, volume Tomo LXXIII - Parte seconda, pages 1201–1248. Venezia - Premiate Officine Grafiche C. Ferrari.

Greselin, F. (2014). More equal and poorer, or richer but more unequal? *Economic Quality Control*, *29*(2), 99–117.

Greselin, F., & Pasquazzi, L. (2009). Asymptotic confidence intervals for a new inequality measure. *Communications in Statistics-Simulation and Computation*, *38*(8), 1742–1756.

Handcock, M.S., & Morris, M. (2006). *Relative distribution methods in the social sciences*. Berlin: Springer Science & Business Media.

Jedrzejczak, A. (2015). Regional income inequalities in Poland and Italy. *Comparative Economic Research*, *18*(4), 27–45.

Kleiber, C. (2008). A guide to the Dagum distributions. In *Modeling income distributions and lorenz curves* (pp. 97–117). Berlin: Springer.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in Economics and Actuarial Sciences* Vol. 470. Hoboken: Wiley.

Leythienne, D., & Ronkowski, P. (2018). A decomposition of the unadjusted gender pay gap using structure of earnings survey data. Luxembourg, Publications Office of the European Union. https://doi.org/10.2785/796328.

OECD (2011). Divided We Stand. Why Inequality Keeps Rising. OECD Publishing Paris. https://www.oecd.org/els/soc/49170768.

OECD (2015). In It Together: Why less inequality benefits all. OECD Publishing Paris. https://www.oecd.org/els/soc/OECD2015-In-It-Together-Chapter1-Overview-Inequality.png.

Pigou, A.C. (1920). The economics of welfare. London, United Kingdom.

TÁRKI (2009). European Social Report from 2009. http://old.tarki.hu/en/research/european_social_report/european_social_report_2009_full.png.

Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, *V*(1), 3–27.

Zenga, M., & Jedrzejczak, A. (2020). Decomposition of the Zenga inequality index $I(Y)$ into the contributions of macro-regions and income components - an application to data from Poland and Italy. *Argumenta Oeconomica*, *1*(44), 101–125.