# Gender Pay Gap Across the Income Distribution: Analysis for the EU

J.M. Landmesser[a,*], A.J. Orłowski[b] and M.A. Rusek[b]

[a]*Institute of Economics and Finance, Warsaw University of Life Sciences — SGGW, ul. Nowoursynowska 166, 02-787 Warsaw, Poland*
[b]*Institute of Information Technology, Warsaw University of Life Sciences — SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland*

In this paper, personal income distributions for men and women in the European Union countries were compared. The counterfactual decomposition method allowed us to study differences of income distributions at various quantile levels. To decompose these differences the Recentered Influence Function — Regression approach was used. Positivity of the gendered income gap in each country was found, but there exists an important diversity in the size and composition of the income gap across members of the European Union. The results obtained for 28 countries allowed us to group them into six clusters. For the analysis, data from the European Survey on Income and Living Conditions (EU-SILC) were used.

topics: gender pay gap, counterfactual decomposition, differences in distributions

## 1. Introduction

Gender differences in wages or personal incomes are a persistent pattern in most European countries. Eurostat publishes regularly updated information on the gender pay gap (GPG) situation in EU. In 2017, the average unadjusted GPG (difference between average gross hourly earnings of male and female employees as percentage of male gross earnings) was 16% [1]. There were considerable differences between EU countries. Across Member States, the gap varied by 22 percentage points. The lowest GPG was recorded in Romania (3.5%) and the highest in Estonia (25.6%), followed by Czechia (21.1%) and Germany (21.0%). In Poland, the GPG stood at 7.2%. The average gender overall earnings gap (the difference in the average annual earnings between women and men) in the EU was 39.6% (for Poland 31.5%). As an unadjusted indicator, the GPG gives an overall picture of the differences between men and women in terms of earnings. While part of the gap can be explained by differences in the average characteristics of male and female employees, a large portion remains unexplained by the data.

The gender pay gap is caused by a number of factors. There are three types of disadvantages women face: lower hourly earnings, fewer hours of paid work and lower employment rates. The gap can be explained by gendered sector affiliation and the high share of a typical employment among women. Women often engage in some typical women's jobs where the economic possibilities in terms of earnings are limited (e.g., the public sector, which is very attractive for women due to its protective nature and flexible working hours). The occupations predominantly carried out by women, such as teaching or sales, offer lower wages than occupations predominantly carried out by men. Also management and supervisory positions are overwhelmingly held by men. Furthermore, women do important unpaid tasks, like household work and caring for children or relatives, far more often than men do. This results in reduction of paid working hours to part-time, as well as in frequent career interruptions. In countries where the female employment rate is low (e.g., Italy), the pay gap is lower than average. This may reflect a small proportion of low-skilled women in the workforce. The large pay gap is characteristic of a highly segregated labor market, where women are concentrated in certain sectors and occupations (e.g., the Czech Republic and Estonia) or a significant proportion of women work part-time (e.g., Germany).

The gender pay gap is one of the best documented facts in labor economics. A comprehensive set of theories have been developed in the literature to help explain the persistence of this phenomenon. Several hypotheses such as the presence of a glass ceiling or sticky floors have been formulated (a glass ceiling exists when the pay gap is significantly larger at the top of the earnings distribution and a sticky floor when the wage gap is larger at the bottom). Many empirical studies found some variation of the wage gap across the whole wage distribution. Arulampalam et al. examined the GPG in 11 European countries using the European Community

Household Panel Survey (ECHPS) [2]. The gap widened toward the top of the wage distribution in most of countries and, in a few cases, it also widened at the bottom of the distribution. The evidence of a glass ceiling effect in Sweden was found in [3]. Nicodemo analyzed the gap in five Mediterranean EU countries, using the European Union Statistics on Income and Living Conditions (EU-SILC) and the ECHPS datasets [4]. She found a positive wage gap in all countries. In most countries the gap was larger at the bottom and smaller at the top of the distribution. In turn, De la Rica, Dolado and Llorens using data for Spain stated that there is a glass ceiling for the better educated while for the less educated there is not [5]. Additionally, the decompositions carried out show that a large part of this gap cannot be explained by differences in the labor-market skills of women and men (e.g., [6] for USA, [7] for Poland).

A number of papers adopt a cross-country perspective. Using EU-SILC data for 24 European Union members, Hedija showed that the gender pay gap varies among the countries [8]. Also Christofides et al. used EU-SILC data and estimated the unexplained part of the gender pay gap for 26 European countries [9]. In the study conducted by Boll and Lagemann the gap for 26 EU-countries was analyzed based on the Structure of Earnings Survey (EU-SES) [10]. Leythienne and Ronkowski provide an information on the data source (EU-SES) and the methodology used by Eurostat to decompose the unadjusted gender pay gap [11]. Despite many differences among the individual studies, they all conclude that the gender pay gap exhibits a remarkable heterogeneity across European countries.

The aim of this paper is to compare personal income distributions for men and women in countries of the EU and to discuss whether there exists a significant diversity in this respect across the countries. We explore the magnitude and composition of the gender pay gap for all 28 EU countries, based on the European Survey on Income and Living Conditions (EU-SILC) database for 2014.

To estimate the differences in incomes we first consider the standard methodology such as the Oaxaca-Blinder decomposition ([12, 13]) and then examine the inequalities at various quantile points along the income distribution. We decompose the differences between two distributions using the counterfactual distribution, which is a mixture of a conditional distribution of the dependent variable and a distribution of the explanatory variables. Such a counterfactual distribution can be constructed in various ways (e.g., [14–17]). To examine the differences in the entire range of income values we use in this work the Recentered Influence Function – Regression approach [18]. After assessing the gender pay gap for all 28 countries, an attempt will be made to group them using agglomerative hierarchical clustering method (Fig. 1).
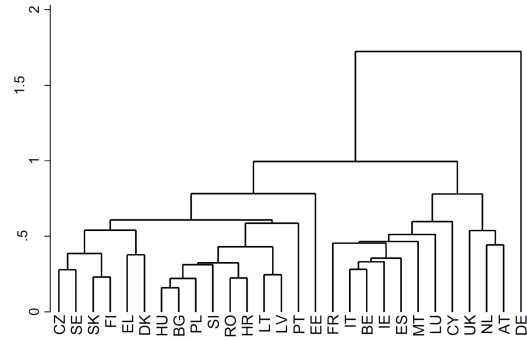


Fig. 1. Dendrogram for average-linkage cluster analysis.

The rest of this article is organized as follows: Sect. 2 describes research methodology, Sect. 3 gives information about the empirical data utilized in the study, Sect. 4 presents and discusses the obtained results, and Sect. 5 concludes.

## 2. Method of analysis

Male and female workers differ with respect to many characteristics, including length of work experience, level of education, occupational status and sector of employment. It is appropriate to decompose the gender pay gap to distinguish what proportion of the overall pay gap is due to differences in individual characteristics and what proportion is due to sex discrimination within the labor market. This method is called the adjusted gender pay gap and it offers clarity by identifying the pay differential between male and female workers after controlling for differences in individual characteristics. Most empirical studies of wage discrimination between men and women use a formal statistical technique proposed by Oaxaca and Blinder [12, 13]. This decomposition approach is widely used to study mean outcome differences between expected values of dependent variable in two comparison groups. Let variable $Y_g$ be the personal income in group $g = M$ and $g = W$ ($M$ — men, $W$ — women). An income equation that relates to the logarithm of incomes as a function of individual characteristics is specified as

$$\ln(Y_g) = X_g \beta_g + \nu_g, \tag{1}$$

where $\ln(Y_g)$ represents the logarithm of incomes, $X_g$ is the vector of individual characteristics, $\beta$ are the returns on these characteristics, and $v_g$ is the error term. Then, the Oaxaca-Blinder decomposition for the average income inequality between two groups is as follows:

$$\hat{\Delta}^\mu = \underbrace{\overline{\ln(Y_M)} - \overline{\ln(Y_W)}}_{\text{raw income gap}} = \overline{X}_M \hat{\beta}_M - \overline{X}_W \hat{\beta}_W =$$

$$\underbrace{\overline{X}_M(\hat{\beta}_M - \hat{\beta}_W)}_{\text{unexplained part}} + \underbrace{(\overline{X}_M - \overline{X}_W)\hat{\beta}_W}_{\text{explained part}}. \tag{2}$$

The gap in the average incomes (expressed as a logarithm) can be broken down in two parts. The first represents the unexplained component (coefficient effect) which includes a difference due to differential reward for equal characteristics. It can be interpreted as the labor market discrimination. The explained part represents the difference in observable characteristics of men and women (endowment effects).

The Oaxaca-Blinder decomposition is based on the linear (log-linear) model. When the conditional mean is a non-linear function the method may not provide consistent estimates of the explained and unexplained effects. It could be said that the use of a linear model is a drawback of the method. On the other hand, thanks to the additive linearity assumption, it is easy to compute the various elements of the detailed decomposition and implementations of the procedure are available in many existing software packages (e.g. Stata).

Several authors have developed extensions of the Oaxaca-Blinder method (for a comprehensive overview of methods see [19]). They allow to study the effects of gender and other covariates on different quantiles of log income distribution and not only at the average of variables. The new techniques extend the mean decomposition analysis to the case of differences between the two income distributions.

One can express the distribution function for the variable $Y$ in group $g$ using the conditional distribution of $Y$ and the joint distribution of all elements of $X$:

$$F_{Y_g}(y) = \int F_{Y_g|X_g}(y|X)\, dF_{X_g}(X). \qquad (3)$$

The so-called counterfactual distribution represents the hypothetical income distribution function that would prevail for people in group $W$ if they had the distribution of characteristics of group $M$:

$$F_{Y_W^C}(y) = \int F_{Y_W|X_W}(y|X)\, dF_{X_M}(X). \qquad (4)$$

Then, the difference in income distributions for men and women can be decomposed as [19]:

$$F_{Y_M}(y) - F_{Y_W}(y) = \underbrace{\left[ F_{Y_M}(y) - F_{Y_W^C}(y) \right]}_{\text{unexplained part}}$$

$$+ \underbrace{\left[ F_{Y_W^C}(y) - F_{Y_W}(y) \right]}_{\text{explained part}}. \qquad (5)$$

Counterfactual distribution can be built in many ways. One of them is the use of the Recentered Influence Function — Regression [18]. RIF-regression is similar to a standard regression, except that the dependent variable $Y$ is replaced by the recentered influence function. The RIF is a transformation of the quantile and inflicts a small change on the quantile, reflecting the influence that each individual income has on the location of the quantile. Adding this small change to the quantile leads to a random variable that can be understood as a linear approximation of the quantile. The transformed quantile has the following structure:

$$\mathrm{RIF}(y, Q_\tau) = Q_\tau + \frac{\tau - \mathrm{I}\{y \leq Q_\tau\}}{f_Y(Q_\tau)}. \qquad (6)$$

The indicator variable $\mathrm{I}\{y \leq Q_\tau\}$ equals 1 for income values smaller or equal to the quantile $Q_\tau$, and 0 otherwise. The term $f_Y(Q_\tau)$ is the value of the probability density function at that quantile. The RIF is a quantile-specific random variable that reflects changes to the quantile as a result of changes in the underlying distribution which depends on the covariates $X$. Thus, the conditional expectation of RIF can be modeled as a linear function of $X$ using OLS: $E\big[\mathrm{RIF}\big(y, Q_\tau|X\big)\big] = X\beta_\tau$ [19]. We compute the sample quantile $\hat{Q}_\tau$, estimate the density $\hat{f}_Y(\hat{Q}_\tau)$ using kernel methods, calculate the RIF of each observation and run regressions of the RIF on $X$ obtaining

$$\hat{\beta}_{g,\tau} = \Big( \sum_{i \in g} X_i^T X_i \Big)^{-1} \sum_{i \in g} X_i^T \mathrm{RIF}(y_{g,i}, Q_{g,\tau}),$$

where $g = M, W$. This provides a way of decomposing quantiles using regression models for proportions:

$$\hat{\Delta}^\tau = \overline{X}_M(\hat{\beta}_{M,\tau} - \hat{\beta}_{W,\tau}) + (\overline{X}_M - \overline{X}_W)\hat{\beta}_{W,\tau} =$$

$$\sum_{j=1}^k \Big[ \overline{X}_{jM}(\hat{\beta}_{jM,\tau} - \hat{\beta}_{jW,\tau}) + (\overline{X}_{jM} - \overline{X}_{jW})\hat{\beta}_{jW,\tau} \Big]. \qquad (7)$$

After applying the RIF-regression method for assessing the gender income gap for all EU-countries, an attempt will be made to group them using hierarchical clustering method. The algorithm for average-linkage cluster analysis will allow the grouping of countries into clusters. There exists a number of different measures to measure distance for binary and categorical data. For interval data the most common distance measure used is the Euclidean distance. Despite the fact that in our analysis we have to handle with a mixture of different data types, we choose the Euclidean distance as the commonly used in clustering. The results of decomposition will be analyzed and compared across the formulated groups of countries.

## 3. Data

The data analyzed in this work come from the European Union Statistics on Income and Living Conditions (EU-SILC) survey in 2014 (research proposal 234/2016-EU-SILC). The EU-SILC is a project aiming at collecting timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. It is a powerful instrument for a comparative analysis of incomes in countries which are EU members. Our analysis was carried out for all 28 EU countries. The data set consists of 174,378 observations: 88,398 for men and 85,980 for women. The explanatory variables in estimated models refer to the personal characteristics of individuals and to the job-related characteristics. Each person is characterized by attributes such as:

- age — age in years,
- educlevel — ordinal variable, 1 — primary, ..., 5 — tertiary,
- married — binary variable, 1 — married, 0 — unmarried,
- permanent — binary variable, 1 — permanent job contract of unlimited duration, 0 — temporary contract of limited duration,
- parttime — binary variable, 1 — part-time, 0 — full-time job,
- manager — binary variable, 1 — supervisory managerial position, 0 — non-supervisory position,
- big — binary variable, number of persons working at the local unit: 1 — bigger than 10, 0 — less than 11.

The annual gross employee (cash or near cash) incomes (in Euro) of men were compared with those obtained by women. The logarithm of the annual gross employee income constitutes the outcome variable.

### 4. Results

We first analyzed the raw (unadjusted) gender income gap across Europe. Table I shows the sample size, average annual income and Gini coefficient values for men and women in European countries. Additionally, the last three columns of this table present the results of the aggregate Oaxaca-Blinder decomposition for the average values.

For each country, the raw gap is calculated as in (2) as the difference between average logarithms of the annual income for men and women respectively (see the unadjusted gap in descending order in the column 7). We have found that there is a positive difference between the mean values of log incomes for men and women for all 28 countries. The largest mean log income differential is observed in Germany (0.625), the Netherlands (0.520) and Austria (0.485), while Slovenia (0.112), Romania (0.141) and Lithuania (0.153) presented the smallest income gap. In Poland, the gender pay gap stands at 0.183. Most Middle and Eastern European states are exhibiting gaps below EU average, with the Czech Republic, Estonia and Slovakia being the exceptions. Among the Western European countries, only Denmark is exhibiting a very small income gap. Moderate gaps are found for Scandinavian and Southern European countries.

The country heterogeneity is reflected not only in the size of the unadjusted gap, but also in its composition. The raw income gap is composed of the unexplained part and the explained part. The explained part is attributable to different (observable) characteristics of men and women. The unexplained part (adjusted gap) compares men and women with similar characteristics. In most countries, a greater portion of the overall gap was unexplained, referring

to the used data set. It means, that the inequalities examined should be assigned in the majority to the coefficients of estimated models rather than to the differentiation of individual characteristics.

The results of the decomposition are presented in percentage. For each country the shares of the unexplained and explained part add up to 100%. For some countries the share of the unexplained part is greater than 100% and the share of the explained part negative (e.g., Poland). In this case, the negative shares reflect the impact of the different characteristics of people reducing the income gap. Thus, as a result, the share of the income discrimination effect increases, exceeding the value of 100%.

The share of the unexplained part is huge for the states with the low raw differential and is small for the states with the high raw differential. Its share ranges from 31% in Luxembourg to 170% in Lithuania and is nowhere identified to be negative. This part of the gender pay gap gives us information about the discrimination. On the other hand, we observed that the characteristics effects are negative in ten countries (among others in Latvia, Lithuania, Slovenia, Estonia, Portugal, Poland). The negative characteristics effects mean that in a labor market with no discrimination females should expect to have higher wages than males. This is often explained by the fact that females exhibit higher education than males. However, the discrimination effects in the labor market are not fully compensated by the endowment effects, and consequently males receive higher wages than females. In 18 countries, the explained part is positive, that is, it increases the overall gap. Only in 5 countries the explained part exceeds the unexplained part of the overall gap.

The detailed decomposition, which was also carried out, made it possible to isolate the factors explaining the inequality observed to a different extent. Because of lack of space in this paper, we present the results of the detailed decomposition only for 3 countries, namely Poland, Germany and the United Kingdom (see Table II).

The strong effect of different education levels of men and women can be noticed. The negative values of explained components for Poland and the UK mean that the differences of the average log incomes between men and women are reduced by the women's higher education levels. The opposite situation occurs in Germany, where the average level of male education is unfavorable for gender-related income inequalities. On the other hand, the values of parttime and manager attributes possessed by men and women increase the income inequality (see the positive explained component values), but in Poland not as strongly as in the other two countries. In all countries, women are discriminated because of their marital status (the positive unexplained components values for variable married). The adjusted income gap is not identified to be at women's advantage anywhere.

TABLE I

Sample size, average annual income (in Euro), Gini coefficient values and the results of the aggregate Oaxaca-Blinder decomposition for each country. Source: own calculation.

| Country | Sample size | Average annual income | | Gini coefficient | | Results of OB decomposition | | |
|---|---|---|---|---|---|---|---|---|
| | | Men | Women | Men | Women | Raw gap | Unexplained part | Explained part |
| DE | 10128 | 42,368.0 | 24,741.8 | 0.324 | 0.385 | 0.625 | 34% | 66% |
| NL | 4912 | 49,364.8 | 30,024.3 | 0.302 | 0.286 | 0.520 | 36% | 64% |
| AT | 4798 | 44,253.1 | 28,842.0 | 0.265 | 0.379 | 0.485 | 36% | 64% |
| UK | 8179 | 39,785.3 | 25,168.3 | 0.369 | 0.379 | 0.480 | 54% | 46% |
| FR | 9251 | 33,494.5 | 24,012.4 | 0.303 | 0.309 | 0.375 | 62% | 38% |
| CY | 3869 | 26,396.4 | 18,920.7 | 0.392 | 0.392 | 0.372 | 58% | 42% |
| ES | 8493 | 25,225.3 | 18,640.9 | 0.368 | 0.403 | 0.362 | 55% | 45% |
| IE | 3759 | 44,876.8 | 31,165.2 | 0.338 | 0.323 | 0.353 | 50% | 50% |
| MT | 4033 | 21,985.5 | 15,955.2 | 0.321 | 0.321 | 0.352 | 65% | 35% |
| CZ | 6501 | 12,413.6 | 8,865.4 | 0.283 | 0.304 | 0.343 | 79% | 21% |
| BE | 4677 | 43,554.0 | 32,171.6 | 0.393 | 0.384 | 0.329 | 45% | 55% |
| EE | 5506 | 13,064.1 | 8,899.2 | 0.402 | 0.365 | 0.328 | 117% | −17% |
| LU | 3932 | 56,353.3 | 43,849.6 | 0.330 | 0.357 | 0.296 | 31% | 69% |
| FI | 8923 | 46,093.8 | 33,300.5 | 0.302 | 0.302 | 0.295 | 90% | 10% |
| SE | 5477 | 45,007.9 | 34,044.3 | 0.321 | 0.321 | 0.290 | 72% | 28% |
| IT | 12715 | 30,007.7 | 22,655.9 | 0.309 | 0.319 | 0.287 | 54% | 46% |
| SK | 5755 | 10,292.6 | 7,980.6 | 0.272 | 0.275 | 0.265 | 94% | 6% |
| PT | 5208 | 16,341.2 | 12,697.8 | 0.385 | 0.385 | 0.254 | 120% | −20% |
| LV | 4968 | 9,491.8 | 7,584.3 | 0.379 | 0.379 | 0.223 | 149% | −49% |
| EL | 3687 | 17,658.0 | 13,934.0 | 0.328 | 0.297 | 0.206 | 69% | 31% |
| DK | 5604 | 60,725.8 | 47,144.0 | 0.280 | 0.280 | 0.193 | 62% | 38% |
| PL | 9908 | 9,619.2 | 7,947.8 | 0.341 | 0.325 | 0.183 | 126% | −26% |
| BG | 4058 | 4,646.6 | 3,747.8 | 0.345 | 0.249 | 0.178 | 125% | −25% |
| HU | 8054 | 7,566.0 | 6,208.9 | 0.347 | 0.319 | 0.176 | 123% | −23% |
| HR | 3601 | 10,350.2 | 8,871.8 | 0.301 | 0.281 | 0.167 | 115% | −15% |
| LT | 4196 | 7,919.5 | 6,617.2 | 0.347 | 0.347 | 0.153 | 170% | −70% |
| RO | 4842 | 4,306.9 | 3,768.3 | 0.239 | 0.315 | 0.141 | 113% | −13% |
| SI | 9344 | 20,512.6 | 18,207.4 | 0.315 | 0.315 | 0.112 | 167% | −67% |
| Total | 174,378 | – | – | – | – | – | – | – |

TABLE II

The results of the detailed Oaxaca-Blinder decomposition for Poland, Germany and the UK.

| Variable | PL | | DE | | UK | |
|---|---|---|---|---|---|---|
| | Unexplained part | Explained part | Unexplained part | Explained part | Unexplained part | Explained part |
| age | $-0.264^a$ | $-0.003$ | $-0.162^a$ | $-0.006^b$ | $0.074$ | $0.001$ |
| educlevel | $-0.120^a$ | $-0.086^a$ | $-0.016$ | $0.025^a$ | $-0.130^a$ | $-0.019^a$ |
| married | $0.100^a$ | $0.001$ | $0.178^a$ | $-0.004^a$ | $0.057^a$ | $0.005^a$ |
| permanent | $-0.055^a$ | $-0.004$ | $0.078^b$ | $0.014^a$ | $-0.076^c$ | $0.000$ |
| parttime | $-0.005^a$ | $0.033^a$ | $0.007^b$ | $0.295^a$ | $-0.019^a$ | $0.206^a$ |
| manager | $0.014^b$ | $0.007^a$ | $0.010$ | $0.032^a$ | $-0.001$ | $0.027^a$ |
| big | $-0.002$ | $0.004^a$ | $-0.124^a$ | $0.056^a$ | $0.036$ | $0.000$ |
| constant | $0.565^a$ | | $0.245^a$ | | $0.321^a$ | |
| total | $0.232^a$ | $-0.048^a$ | $0.214^a$ | $0.411^a$ | $0.261^a$ | $0.219^a$ |

Source: own calculation using Stata command 'oaxaca' with standard errors obtained by the delta method based on analytic derivatives. The significance levels codes are as follows: $^a$0.01, $^b$0.05 and $^c$0.1.

TABLE III

The results of the detailed RIF-regression decomposition for Poland, Germany and the UK.

| Variable | PL | | | DE | | | UK | | |
|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q1 | Q2 | Q3 | Q1 | Q2 | Q3 |
| difference | $0.123^a$ | $0.143^a$ | $0.172^a$ | $0.791^a$ | $0.555^a$ | $0.412^a$ | $0.547^a$ | $0.453^a$ | $0.397^a$ |
| Unexplained part | | | | | | | | | |
| age | $-0.101^b$ | $-0.320^a$ | $-0.383^a$ | $-0.370^a$ | $-0.188^a$ | $-0.345^a$ | $0.151^b$ | $-0.003$ | $-0.068$ |
| educlevel | $0.005$ | $-0.310^a$ | $-0.191^b$ | $-0.360^b$ | $-0.114^b$ | $0.090^c$ | $-0.210^a$ | $-0.115^a$ | $-0.184^a$ |
| married | $0.105^a$ | $0.096^a$ | $0.124^a$ | $0.261^a$ | $0.130^a$ | $0.111^a$ | $0.021$ | $0.067^a$ | $0.036^c$ |
| permanent | $-0.011$ | $-0.037^c$ | $-0.023$ | $-0.251^a$ | $-0.326^a$ | $-0.206^a$ | $-0.110$ | $-0.015$ | $0.010$ |
| parttime | $-0.004^c$ | $-0.001$ | $0.000$ | $0.031^a$ | $0.047^a$ | $0.031^a$ | $0.021^a$ | $0.028^a$ | $0.022^a$ |
| manager | $0.013^b$ | $0.014^b$ | $0.030^a$ | $-0.002$ | $0.003$ | $0.012$ | $-0.005$ | $0.006$ | $-0.025^c$ |
| big | $0.043^c$ | $0.015$ | $0.012$ | $-0.269^a$ | $-0.144^a$ | $-0.076^b$ | $0.007$ | $0.034$ | $-0.035$ |
| constant | $0.086$ | $0.771^a$ | $0.696^a$ | $1.187^a$ | $0.716^a$ | $0.494^a$ | $0.306^b$ | $0.238^a$ | $0.487^a$ |
| total | $0.136^a$ | $0.228^a$ | $0.264^a$ | $0.226^a$ | $0.123^a$ | $0.113^a$ | $0.181^a$ | $0.240^a$ | $0.242^a$ |
| Explained part | | | | | | | | | |
| age | $-0.002$ | $-0.003$ | $-0.003$ | $-0.008^b$ | $-0.006^b$ | $-0.006^b$ | $0.001$ | $0.001$ | $0.001$ |
| educlevel | $-0.044^a$ | $-0.108^a$ | $-0.110^a$ | $0.032^a$ | $0.026^a$ | $0.020^a$ | $-0.018^a$ | $-0.021^a$ | $-0.024^a$ |
| married | $0.000$ | $0.001$ | $0.000$ | $-0.007^a$ | $-0.004^a$ | $-0.003^a$ | $0.008^a$ | $0.003^b$ | $0.005^a$ |
| permanent | $-0.003$ | $-0.004$ | $-0.002$ | $0.025^a$ | $0.013^a$ | $0.005^a$ | $-0.001$ | $0.000$ | $0.000$ |
| parttime | $0.030^a$ | $0.017^a$ | $0.011^a$ | $0.394^a$ | $0.319^a$ | $0.229^a$ | $0.353^a$ | $0.202^a$ | $0.137^a$ |
| manager | $0.002^a$ | $0.006^a$ | $0.010^a$ | $0.038^a$ | $0.030^a$ | $0.026^a$ | $0.023^a$ | $0.028^a$ | $0.035^a$ |
| big | $0.004^a$ | $0.005^a$ | $0.003^a$ | $0.090^a$ | $0.055^a$ | $0.030^a$ | $0.001$ | $0.000$ | $0.000$ |
| total | $-0.013^c$ | $-0.085^a$ | $-0.092^a$ | $0.565^a$ | $0.433^a$ | $0.300^a$ | $0.366^a$ | $0.213^a$ | $0.155^a$ |

Source: own calculation using Stata commands 'rifreg' and 'oaxaca' with standard errors obtained by the delta method based on analytic derivatives; the significance levels codes are as follows: [a]0.01, [b]0.05 and [c]0.1.

The main purpose of this paper was to analyze the differences in the income distributions for men and women in EU-countries. The estimation of the income gap between men and women, following the RIF-regression method, is reported in Figs. 2–6. The results obtained indicate positive differences between male and female log incomes at each level of income and for each country that we take into consideration. These differences are mostly non-monotonous. We also decomposed the income gap for quantiles and determined the explained and unexplained components of the difference in terms of quantiles.

To provide additional insights into sources of income differences along the income distribution, we examined and compared the sources of the explained and unexplained gap. The detailed decomposition made it possible to isolate the factors explaining the inequality observed to a different extent. We present the results of the detailed decomposition only for 3 countries — Poland, Germany and the UK – and only for three quartiles: Q1, Q2, Q3 (quartiles are types of quantiles which divide the number of data points into four equal parts) (see Table III).

Again, the strong effect of different education levels of men and women can be noticed. The differences in education levels mitigate the gap in Poland and the UK. The opposite is true for Germany. The values of parttime and manager attributes increase the income inequality along the whole distribution in all three countries (see the positive explained component values), but this effect is much weaker for Poland. In Poland and Germany, women are definitely discriminated because of their marital status. For example, the income gap is more evident for married German women at the bottom of the income distribution. For Poland, we can also observe discrimination when hiring women for management positions.

The unexplained component of the income gap (associated with the "valuation" of the people's characteristics by the market) increases with the amount of income in Poland and the UK, but not in Germany. This demonstrates that the discrimination is more evident for higher values of incomes in Poland and the UK. For the West Europe countries, the positive values of the total explained component are large in the group of the worse earning people. This reflects the increase of wage inequality due to more frequent part-time employment of women than men. For Poland, the negative values of the total explained component are especially large in the groups of the better earning people. Such a favorable reduction in the gap for women is probably due to "better" characteristics of women than men in those higher income groups.
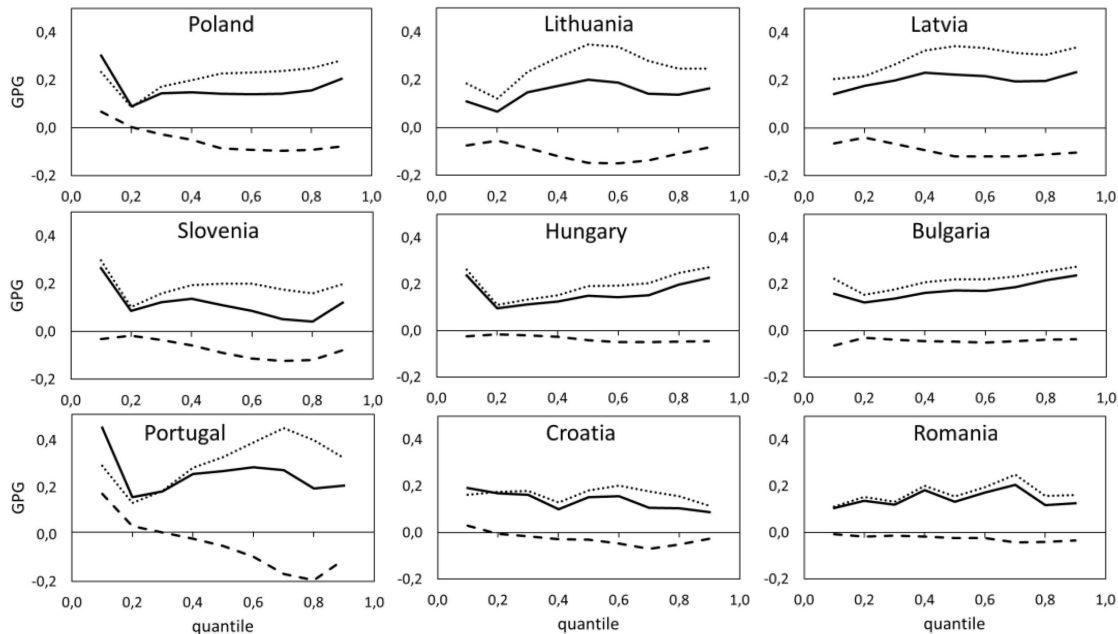
Fig. 2. The log income gaps for men and women vs. quantile rank in the group 1: solid lines — the total gap, dashed lines — the explained effect, dotted lines — the unexplained effect. On the horizontal axis there are marked four quantiles of the income distribution that split the data into fife equally sized groups (quintiles).

After assessing the gender pay gap for all 28 countries, an attempt was made to group them using agglomerative hierarchical clustering method. The algorithm for average-linkage cluster analysis with Euclidean distance allowed the grouping of countries into six clusters.

The shapes of income gap are examined in Figs. 2–6, where solid lines represent the total income gap, dashed lines denote the explained component and dotted lines indicate the unexplained effect.

Six groups of countries were identified:

- Group 1: Poland, Lithuania, Latvia, Slovenia, Hungary, Bulgaria, Portugal, Croatia, Romania,
- Group 2: Denmark, Sweden, Finland, Greece, the Czech Republic, Slovakia,
- Group 3: Ireland, Luxembourg, Belgium, France, Italy, Spain, Malta, Cyprus,
- Group 4: the United Kingdom, the Netherlands, Austria,
- Group 5: Estonia,
- Group 6: Germany.

Group 1 (the largest group) consists mainly of the former socialist states of Eastern Europe (Fig. 2). For most countries in this group, the total effect is low, but it widens sometimes at the bottom of the income distribution. We have found that Poland, Slovenia, Portugal and Hungary suffer a sticky floor effect. The unexplained effect is bigger than the explained one. The share of the unexplained part is very high. The explained part (the effect of characteristics) is negative, which implies that women have better characteristics than men that compensate them for discrimination. On the other hand, when comparing the incomes of women and men with the exact same characteristics, we get an even bigger disparity than the average gender pay gap.

Group 2 consists of the countries from the north and east of Europe (welfare states and post-socialist states) (Fig. 3). It is characterized by the low total gender income gap of irregular shape along the income distribution. There is the bigger unexplained effect than the explained one. The effect of coefficients is positive and its share is high in the whole range of the income distribution. This is the result of differences in the "market prices" of individual characteristics of men and women, interpreted as the labor market discrimination. The explained effect is positive, although very low.

Group 3 consists of the highly developed countries of Western Europe with high GDP per capita (Fig. 4). In most from this countries the total gender gap is higher than before, decreasing along the distribution (larger at the bottom and smaller at its top, suggesting sticky floor which is in accordance with Nicodemo studies [4]). The gender differences in characteristics are positive, which means that the different values of characteristics of men and women increase the income inequalities. The explained effect is bigger than the unexplained effect at the bottom of the log income distribution (except Cyprus). For the higher income ranges, the unexplained effect often prevails. Both effects, the explained and the unexplained, are still positive, increasing the income discrepancies.
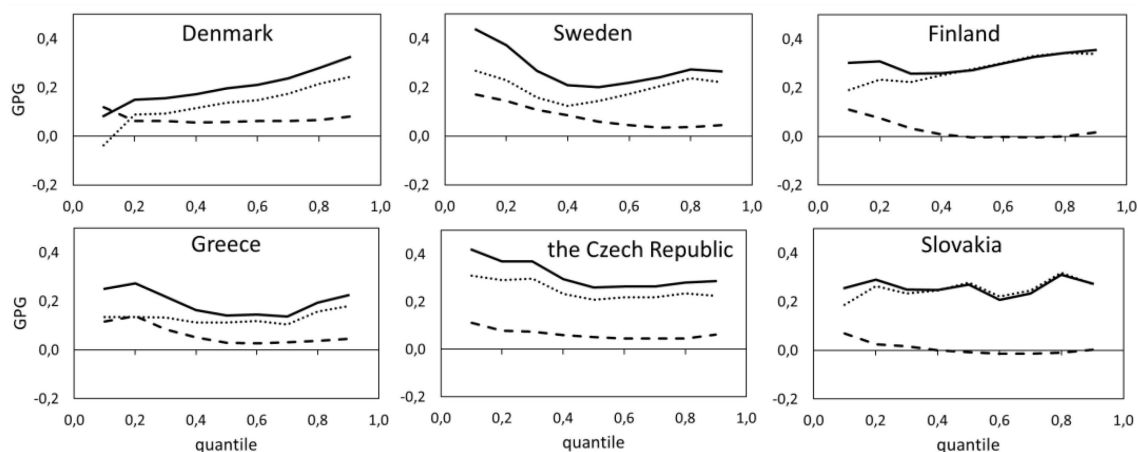
Fig. 3. The log income gaps for men and women vs. quantile rank in the group 2: solid lines — the total gap, dashed lines — the explained effect, dotted lines — the unexplained effect. On the horizontal axis there are marked four quantiles of the income distribution that split the data into fife equally sized groups (quintiles).
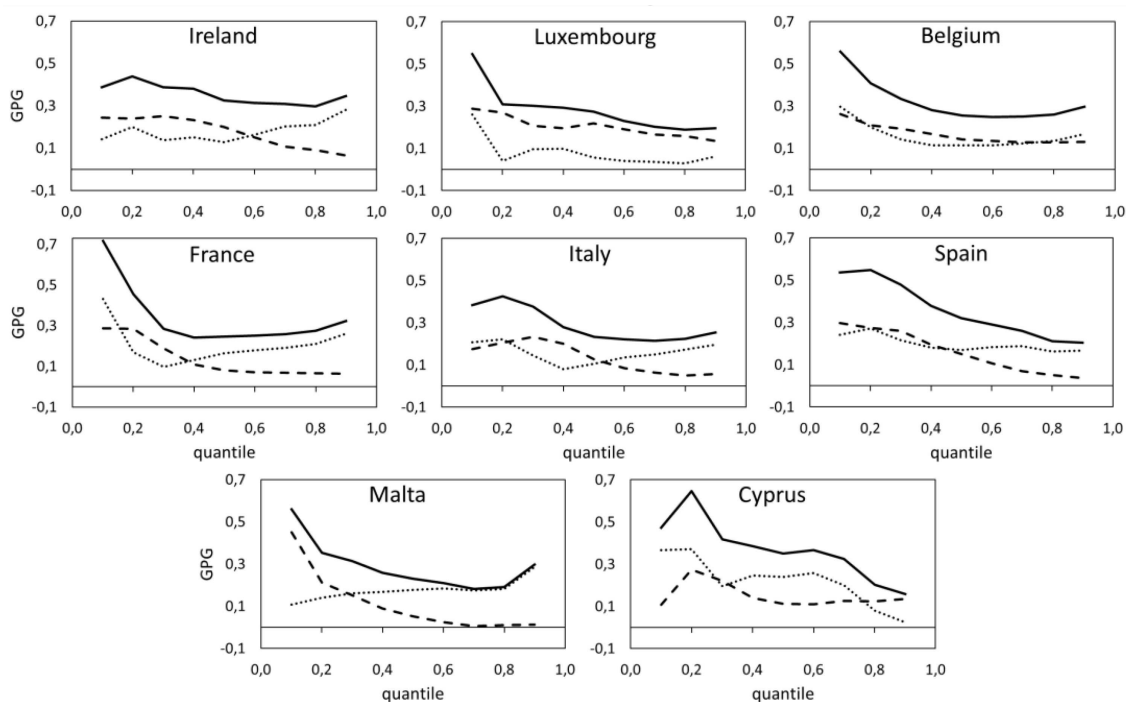


Fig. 4. The log income gaps for men and women vs. quantile rank in the group 3: solid lines — the total gap, dashed lines — the explained effect, dotted lines — the unexplained effect. On the horizontal axis there are marked four quantiles of the income distribution that split the data into fife equally sized groups (quintiles).

Group 4 is made up of the United Kingdom, the Netherlands and Austria (Fig. 5). These are countries with very high GDP per capita and highly segregated labor markets in which a significant proportion of women work part-time. In this case, the large total gap and the large explained effect have a decreasing shape and are rapidly falling as we move toward the top of the income distribution. The unexplained part is positive and at a moderate level, presenting the existing effect of discrimination on the labor market — higher among the poorest, then lower among the richest.

There is only one country in the group 5 – Estonia (Fig. 6a). Estonia is characterized by an increase of the income inequalities as we move toward the top of the income distribution. The total effect is low, but it widens at the top of the income distribution, suggesting glass ceiling effects. The share of the unexplained effect is very high and the explained effect is negative.

The last group, the group 6, is made up of Germany. In this case, the large total gap and the large explained effect have a decreasing shape and are rapidly falling as we move toward the top
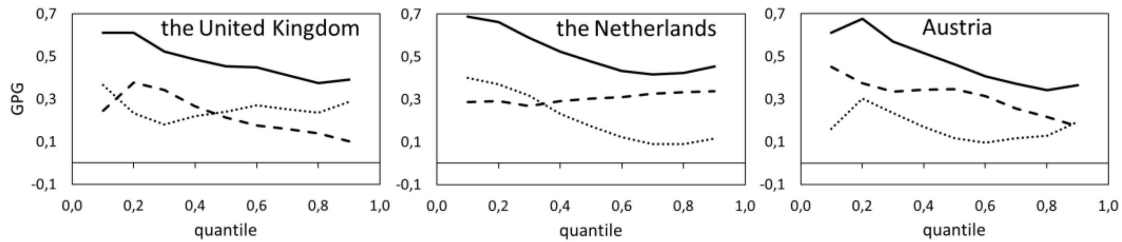
Fig. 5.   The log income gaps for men and women vs. quantile rank in the group 4: solid lines — the total gap, dashed lines — the explained effect, dotted lines — the unexplained effect. On the horizontal axis there are marked four quantiles of the income distribution that split the data into fife equally sized groups (quintiles).
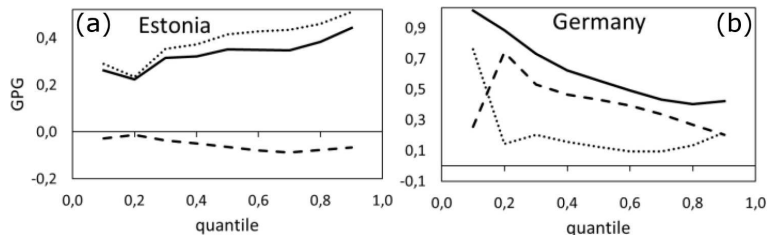


Fig. 6.   The log income gaps for men and women vs. quantile rank in the group 5 (a) and 6 (b): solid lines — the total gap, dashed lines — the explained effect, dotted lines — the unexplained effect. On the horizontal axis there are marked four quantiles of the income distribution that split the data into fife equally sized groups (quintiles).

of the income distribution (Fig. 6b). The unexplained part is positive and at a moderate level, presenting the existing effect of discrimination on the labor market — initially high among the poorest, then low. A sticky floor effect is present. Women at the bottom are more disadvantaged with respect to those at the first quartile.

## 5. Conclusions

Gender income discrepancies are persistent all over Europe. As we measured using the Oaxaca-Blinder method, across the 28-country bloc, there is a considerable disparity in income between genders, ranging from less than 0.17 in Slovenia, Romania, Lithuania and Croatia to more than 0.40 in Germany, the Netherlands, Austria and the United Kingdom. There is a positive difference between the mean log income values for men and women in all EU-countries. Most Eastern European states are exhibiting gaps below EU average, whereas among the Western European countries, the gaps are bigger. We conducted also the decomposition of the differences between average values for men and women's log incomes. For rich countries, the explained effect was large and positive, for low-GPG countries - lower and negative. Since the Oaxaca-Blinder technique focuses only on average effects, we carried out the decomposition of inequalities along the distribution of log incomes using the RIF-regression method. The idea behind this technique is that the pay gap might vary across the income distribution and that is why we have to go beyond the simple comparison of average values. The main

purpose of our research was to compare personal income distributions for men and women in EU countries. By applying the RIF-regression method, we understand the different gender gap between men and woman, along their income distribution. The gender pay gap is positive in all countries, and most of it is made up of the discrimination effect. The calculated differences between the values of log incomes along the whole log income distribution for each country were decomposed into the sum of the unexplained and explained components. The results obtained allowed us to group the 28 EU-countries into six clusters:

- the former socialist states with low income gap, high unexplained and negative explained effects,
- welfare states with low income gap, high unexplained and positive explained effects,
- rich countries with declining gap, higher explained effect among the poorest and higher unexplained among the richest,
- rich countries with a highly segregated labor market and very high, decreasing income gap,
- Estonia with glass ceiling effect and high unexplained effect,
- Germany with high but decreasing income gap and moderate discrimination.

While in Germany or in Austria a closer look at various factors makes the adjusted gap smaller, the situation is quite different in other EU countries. In Poland, for instance, the explained income gap has a negative value, however very low in absolute

values. In most of the countries, the endowment effect is small and coefficient effect prevails, so in the final effect the total gap is positive along the whole income distribution (our results are in accordance with [11], but partly in opposition to [9]).

It is not only the size of the explained gap that is important, but also the factors contributing to it. An important factor correlated with the gender gap is the occupational segregation that women are more exposed to. Our analysis confirmed that a large gender pay gap characterizes a highly segregated labor market, where women are more concentrated in a limited number of sectors or in which a significant proportion of women work part-time (e.g., Germany and Austria). Also the other studies show that the uneven distribution of men and women across different industries is the reason for gendered income inequalities across all EU countries (e.g., [2, 10]).

Family policies such as child care or parental leave are likely to affect gender pay gaps. Mothers are more likely to use parental leave than fathers, implying that women reduce their work experience and undermine their earnings capacity. The countries with poor policies for child-care, flexibility of work etc., are more unequal and are more subject to the sticky floor effect (e.g. Portugal), i.e., the gender pay gap is bigger at the bottom of the wage distribution. We found the similar results as in [4], where the author showed that Southern countries with a strong tradition of family and poor family policies suffer from the phenomena of the sticky floor while Scandinavian countries with strong family policies are affected by the glass ceiling effect.

While interpreting the results presented above we have to remember that they are strictly linked with the data set used and the set of control variables in the empirical specification. The list of explanatory variables (person attributes) omits three (highly connected) real world factors: the length of active employment (excluding any possible maternity leaves), continuity of employment (measured, for example by the number of employment breaks longer than a year) and number of children (in many cases the reason from the breaks in active employment). In the research on gender income inequality these three factors are considered very important, as they impact the promotion processes, necessitate the re-training on the job, if the conditions have changed, influence the capacity to take overtime (and thus increase the incomes), etc. The origin of this omission is due to the lack of such information in the dataset. The EU-SILC dataset has only one additional variable: job experience in the current enterprise, but not the length of active work and continuity measures. This lack of data inhibits the inclusion of the variables in the analysis. As a result, the magnitude of the earnings gap and its decomposition cannot be directly compared with previous international studies.

The fact that women earn lower wages than men has a negative impact on their incomes over their lifetimes, career progression, pension rights and risk of poverty in old age. Therefore, active policies to close wage gaps are required. The possible directions to improve the situation across the European region are pay transparency or new legal frameworks. However, as our analysis showed, the conditions in the countries are very different and therefore it is difficult to give any general recommendations on the European level. Despite this, the results obtained and presented here seem very interesting, and the topic is definitely worth further research.

### References

[1] Eurostat, Gender Pay Gap Statistics.

[2] W. Arulampalam, A.L. Booth, M.L. Bryan, *Ind. Labor Relat. Rev.* **60**, 163 (2007).

[3] J. Albrecht, A. Björklund, S. Vroman, *J. Labor Econ.* **21**, 145 (2003).

[4] C. Nicodemo, Gender Pay Gap and Quantile Regression in European Families, *IZA Discussion Paper* No. 3978 (2009).

[5] S. De la Rica, J.J. Dolado, V. Llorens, *J. Popul. Econ.* **21**, 751 (2008).

[6] K. Karpio, J. Landmesser, P. Łukasiewicz, A. Orłowski, *Acta Phys. Pol. A* **129**, 965 (2016).

[7] J.M. Landmesser, *Statistics in Transition New Series* **17**, 331 (2016).

[8] V. Hedija, *Econ. Res. (Ekonomska Istraživanja)* **30**, 1804 (2017).

[9] L. Christofides, A. Polycarpou, K. Vrachimis, *Labour Econ.* **21**, 86 (2013).

[10] C. Boll, A. Lagemann, *Intereconomics* **54(2)**, 101 (2019).

[11] D. Leythienne, P. Ronkowski, "A Decomposition of the Unadjusted Gender Pay Gap Using Structure of Earnings Survey Data", Statistical Working Papers, 2018.

[12] R. Oaxaca, *Int. Econ. Rev.* **14**, 693 (1973).

[13] A. Blinder, *J. Hum. Resour.* **8**, 436 (1973).

[14] J. DiNardo, N.M. Fortin, T. Lemieux, *Econometrica* **64**, 1001 (1996).

[15] N.M. Fortin, T. Lemieux, *J. Hum. Resour.* **33**, 610 (1998).

[16] J.F. Machado, J. Mata, *J. Appl. Econom.* **20**, 445 (2005).

[17] B. Melly, *Labour Econ.* **12**, 577 (2005).

[18] S. Firpo, N.M.Fortin, T. Lemieux, *Econometrica* **77**, 953 (2009).

[19] N.M. Fortin, T. Lemieux, S. Firpo, *NBER Working Paper* No. 16045 (2010).