DIFFERENCES IN THE INCOME DISTRIBUTION OF HOUSEHOLDS RUN BY MEN AND WOMEN BY VOIVODESHIPS

Joanna Małgorzata Landmesser, Ph.D., Associate Professor
Warsaw University of Life Sciences – SGGW
Faculty of Applied Informatics and Mathematics
Department of Econometrics and Statistics
Nowoursynowska 159, 02-776 Warsaw, Poland
e-mail: joanna_landmesser@sggw.pl
ORCID: 0000-0001-7286-8536

Received 31 October 2019, Accepted 31 March 2020

Abstract

Research background: Household income depends on its demographic composition, age and education of its members, place of residence and many other factors. In our work, we concentrate on the income distribution of Polish households.

Purpose: The study aims to compare the household income distributions in Polish voivodeships, taking into account the gender of the family head. We provide evidence on the magnitude and determinants of regional differences in gender-specific income disparities.

Research methodology: In order to move beyond estimation based on mean values, we apply the Residual Imputation Approach and extend the Oaxaca-Blinder decomposition procedure to different quantile points along the income distribution. To describe the differences between two income distributions we construct a counterfactual distribution and decompose the inequalities into explained and unexplained components.

Results: The regional variation of the gender income gap has been explained with individual and job-related characteristics. There exists an important diversity in the size of the gender income gap across the Polish provinces. The results obtained for 16 voivodeships allowed us to group them into four clusters: heavily industrialized voivodeships with a large income gap, weakly industrialized with a low income gap, voivodeships with large agglomerations characterized by a low gap, and medium-developed voivodeships with a large, U-shaped gap.

Novelty: Our results provide novel insights into the regional dimension of the income gap.

Keywords: income inequalities, gender pay gap, decomposition, regional labor markets

JEL classification: D31, J31
Introduction

Income inequality in Poland increased significantly during the transformation to a market economy. According to the Household Budget Survey (HBS) data, we can observe for Poland the Gini index at about the level of 0.298 (Statistics Poland, 2018). In 2017, available income per capita was higher than the national average in six voivodeships: Mazowieckie, Zachodniopomorskie, Pomorskie, Śląskie, Dolnośląskie, and Wielkopolskie. The highest average income per capita was available for households from the Mazowieckie Voivodeship and it was 19.6% higher than the national average income per capita. The lowest income was registered in the Podkarpackie Voivodeship (21.5% below the average).

Often in the analysis of income inequality, the main focus is the gender pay gap. The findings of numerous empirical studies show that men earn higher wages than women. A similar observation may apply to the income of households run by women and men.

One aspect of the gender pay gap that has received very little attention so far concerns its regional dimension. Empirical studies mostly focus on the national level (for example, Jurajda, 2003 for the Czech Republic and Slovakia; Pena-Boquete, De Stefanis, Fernandez-Grela, 2010 for Italy and Spain; Chatterji, Mumford, Smith, 2011 for the United Kingdom; Śliwicki, Ryczkowski, 2014 for Poland) or adopt a cross-country perspective (Olivetti, Petrongolo, 2008; Hedija, 2017; Boll, Lagemann, 2019). However, the gender pay gap varies considerably within a country. In Polish provinces (NUTS2-regions), the unadjusted gender pay gap calculated for the total annual salary ranges between 1.3 and 24.8% and the gap calculated for the hourly wage ranges between −16.4 and 14.9%, which means that in some regions women even earn more than men (Śliwicki, 2017). Surprisingly, little is known about the mechanisms that drive these regional disparities. Regions differ substantially in their sectoral composition, providing different employment possibilities for men and women. Regional disparities are enhanced by the differences in workers’ characteristics across regions. Research on these issues is of great importance. Comprehensive evidence on regional gender income disparities in Poland is missing.

Currently, the rapid development of microeconometric techniques useful in the context of studying differences between groups of objects can be observed. Since the seminal works of R. Oaxaca (1973) and A. Blinder (1973) many procedures that go beyond the simple decomposition of differences between the average values have been proposed (see e.g. Juhn, Murphy, Pierce, 1993; DiNardo, Fortin, Lemieux, 1996; Gosling, Machin, Meghir, 2000; Donald, Green, Paarsch, 2000; Machado, Mata, 2005; Autor, Katz, Kearney, 2005). The main
advantage of modern decomposition methods is to help to discover the factors affecting for example changes in the distribution of wages. The differences in income distributions of men and women are also analyzed. The research conducted in Poland regarding the gender pay gap is limited to decomposing the average level of wage differences (e.g. Słoczyński, 2012; Goraus, Tyrowicz, 2014; Śliwicki, Ryczkowski, 2014) and only a few studies go beyond the mean decomposition.

In the paper, we analyze regional income inequalities in terms of the household income distribution. The main objective of our study is to compare the distribution of household income in the Polish voivodeships (in 16 NUTS2-regions), taking into account the gender differences in family heads. In order to move beyond estimations based on mean values, we apply the Residual Imputation Approach (Juhr, Murphy, Pierce, 1993). We argue that employing this technique can provide deeper insights into the nature of income differentials. This paper explains the regional variation of the gender income gap in Poland with individual and establishment characteristics. We employ data from the Household Budget Survey (HBS) for Poland in 2015.

The paper is organized as follows. Section 2 describes the techniques used for the decomposition of inequalities. In section 3, we present data and the results of the decomposition analyses for Poland and the single regions. Section 4 discusses the results and offers some concluding remarks.

1. Method of the analysis

In the analysis of income inequality, it may be relevant to assign inequality contributions to various population subgroups associated with various socioeconomic characteristics of individuals. Let $Y_g$ denote the outcome variable in group $g$ (e.g. the personal income in the men’s group, $g = M$, or in the women’s group, $g = W$) and $X_g$ the vector of individual socioeconomic characteristics of people in group $g$ (e.g. age, education level, work experience). The expected value of $Y$ conditionally on $X$ is a linear function $\hat{Y}_g = X_g \hat{\beta}_g + v_g, g = M, W$, where $\hat{\beta}_g$ are the returns to the characteristics. The Oaxaca-Blinder decomposition can be applied whenever we need to explain the differences between the expected values of dependent variable $Y$ in two comparison groups (Oaxaca, 1973; Blinder, 1973):

$$\bar{Y}_M - \bar{Y}_W = \bar{X}_M \hat{\beta}_M - \bar{X}_W \hat{\beta}_W = \bar{X}_M (\hat{\beta}_M - \hat{\beta}_W) + (\bar{X}_M - \bar{X}_W) \hat{\beta}_W$$

where $\hat{\beta}_M$ and $\hat{\beta}_W$ are the estimated coefficients for the men’s and women’s group respectively.

The first term on the right-hand side of (1) represents the unexplained effect, and the second term represents the explained effect. The unexplained effect captures the difference in expected values due to differences in characteristics ($X$) between the two groups, while the explained effect captures the difference in expected values due to differences in coefficients ($\beta$) between the two groups.
The explained effect expresses the difference in the potentials of people in two groups and presents the effect of characteristics. The unexplained effect is the result of differences in the estimated parameters, and so in the “prices” of individual characteristics of the group’s representatives. It can be interpreted as labor market discrimination. Also, the detailed decomposition may be calculated. The disadvantage of this approach is that it focuses solely on average effects which may result in a confusing assessment if the effects of covariates vary across income distribution.

The objective of this study is to extend the Oaxaca-Blinder decomposition procedure to different quantile points along income distribution. Let \( F_{g_1}(y) \) be the distribution function for the variable \( Y \) in group \( g \), which can be expressed using the conditional distribution \( F_{g_1|X,D_1}(y|X=x) \) of \( Y \) and the joint distribution \( F_{X|g_1}(X) \) of all elements of \( X \) \((D_g = 1 \text{ if } g = M; D_g = 0 \text{ if } g = W)\): 

\[
F_{g_1}(y) = \int F_{g_1|X,D_1}(y|X=x) \cdot F_{X|g_1}(X) \, dx.
\]

To calculate the differences between the two distributions we have to construct a counterfactual distribution that mixes conditional distribution for the outcome variable \( Y \) with various distributions for explanatory variables \( X \):

\[
F_{g_1}^c(y) = \int F_{g_1|X,v}(y|X) \cdot dF_{X,v}(X)
\]

A counterfactual distribution is the distribution of incomes that would prevail for people in group \( W \) if they had the distribution of characteristics of group \( M \). Then the observed difference between the two distributions may be decomposed into:

\[
F_{g_2}(y) - F_{g_1}(y) = \left[ F_{g_2}(y) - F_{g_2}^c(y) \right] + \left[ F_{g_2}^c(y) - F_{g_1}^c(y) \right]
\]

The main difference to the Oaxaca-Blinder decomposition is that this decomposition refers to full distributions, rather than just to their means.

Several approaches have been suggested in the literature for estimating counterfactual distribution (e.g., DiNardo, Fortin, Lemieux, 1996; Fortin, Lemieux, 1998; Donald, Green, Paarsch, 2000; Machado, Mata, 2005; Fortin, Lemieux, Firpo, 2010). An approach proposed by Ch. Juhn, K.M. Murphy and B. Pierce (1993) is based on the Residual Imputation Procedure. The implementation of the procedure is two-step. In the first step, the residuals are replaced by counterfactual residuals under the assumption of rank preservation:

\[
y_{Wi}^{C,1} = X_{Mi} \beta_M + v_{Wi}^{C,1}, i=1,\ldots,n, \quad \text{where } v_{Wi}^{C,1} = F^{-1}_{v|X} (\tau_{Mi}(X_{Mi}),X_{Mi}), \quad \tau_{Mi}(X_{Mi}) \text{ is the conditional rank of } v_{Mi} \text{ in the distribution of residuals for men.}
\]

In the second step, the counterfactual returns to observables are imputed as

\[
y_{Wi}^{C,2} = X_{Mi} \beta_W + v_{Wi}^{C,1}, i=1,\ldots,n.
\]
In this study, after the assessment of the income gap in voivodeships, an attempt was made to group them using the hierarchical clustering method. Hierarchical clustering is a widely used data mining tool for grouping data into clusters that exposes similarities or dissimilarities in the data. We have applied the agglomerative method of hierarchical clustering, in which at each step of the clustering process an observation or cluster is merged into another cluster. The calculation of the new distance was carried out using the average-linkage cluster analysis method based on the Euclidean metric.

2. Data used and results of the empirical study

Our analysis rests on data from the Household Budget Survey (HBS) for Poland in 2015. This representative survey is one of the most comprehensive sources of socioeconomic information on Polish households and it plays an important role in the analysis of the living standards of the population. It provides detailed information on the revenues and outgoings of such socio-economic groups like employees’ households, farmers’ households, households of the self-employed, households of retirees and pensioners and households living on unearned sources. Our sample consists of 23,579 households (without retirees and pensioners); among them 16,091 there were households run by men and 7,668 run by women. The data provide information on the household’s equivalent disposable monthly income (the OECD-modified equivalence scale 50/30 was applied) as well as on reference persons’ attributes, such as gender, age, education level, etc. The average monthly disposable equivalent income of a household run by a man was PLN 2,420.11, while for a household run by a woman it was PLN 2,329.45. The data indicate that income inequalities between households run by men and women in voivodeships can be observed (see Table 1).

The lowest average monthly income was observed in the Podkarpackie Voivodeship, and the richest region was the Mazowieckie Voivodeship, with its highest value of average income (3,014 PLN) (compare Jędrzejczak, 2015). The highest average income for male households was recorded in the Mazowieckie Voivodeship, followed by the Lubuskie and Dolnośląskie voivodeships, the lowest – in Podkarpackie, Świętokrzyskie and Podlaskie. The wealthiest women’s households were in Mazowieckie, then in the Dolnośląskie and Zachodniopomorskie voivodeships, the poorest – in Kujawsko-Pomorskie, Podkarpackie and Świętokrzyskie. The maximum inequality was recorded in the Lubuskie Voivodeship. In two provinces the income gap was negative (in the Podlaskie and Podkarpackie voivodeships).
Table 1. The average monthly disposable equivalent income of households in voivodeships (in PLN)

| Voivodeship           | All households | Households run by men | Households run by women | Income gap |
|-----------------------|----------------|-----------------------|-------------------------|------------|
| Mazowieckie           | 3,014.69       | 3,048.55              | 2,955.55                | 93.00      |
| Lubuskie              | 2,501.14       | 2,653.67              | 2,194.67                | 459.00     |
| Dolnośląskie          | 2,468.66       | 2,533.55              | 2,334.26                | 199.30     |
| Pomorskie             | 2,452.59       | 2,518.04              | 2,309.69                | 208.36     |
| Zachodniopomorskie    | 2,314.89       | 2,315.13              | 2,314.39                | 0.74       |
| Słąskie               | 2,311.87       | 2,377.05              | 2,157.80                | 219.25     |
| Łódzkie               | 2,292.07       | 2,331.14              | 2,222.50                | 108.64     |
| Małopolskie           | 2,269.02       | 2,285.72              | 2,232.24                | 53.48      |
| Wielkopolskie         | 2,267.24       | 2,272.08              | 2,254.44                | 17.64      |
| Opolskie              | 2,221.77       | 2,288.43              | 2,060.99                | 227.44     |
| Warmińsko-Mazurskie   | 2,193.91       | 2,260.61              | 2,036.72                | 223.89     |
| Podlaskie             | 2,192.61       | 2,139.32              | 2,297.38                | −158.06    |
| Lubelskie             | 2,163.20       | 2,189.93              | 2,113.07                | 76.86      |
| Kujawsko-Pomorskie    | 2,114.33       | 2,170.61              | 1,987.87                | 182.74     |
| Świętokrzyskie        | 2,047.63       | 2,064.35              | 2,010.98                | 53.37      |
| Podkarpackie          | 1,951.73       | 1,930.68              | 1,997.03                | −66.35     |

Source: own elaboration based on HBS 2015.

In our further empirical decomposition analysis, the logarithm of the average monthly disposable equivalent income constitutes the outcome variable. We follow the seminal work of R. Oaxaca (1973) and A. Blinder (1973) and decompose the unadjusted log income gap into an explained and unexplained part for Poland as well as for all 16 NUTS2-regions. The explanatory factors include individual and company determinants that relate to the head of the household (see Table 2). We establish seven explanatory variables in our models. As individual characteristics, age, education level and place of the residence of the household head were included. As job-related characteristics, working time, contract type, ownership of the company were taken into account.

It is useful to look at summary statistics for some covariates (in Table 3).
Table 2. Definition of the explanatory variables

| Sex       | gender of the household head (1 – male, 0 – female) |
|-----------|-----------------------------------------------------|
| Age       | age of the household head in years                  |
| Education | educational level of the household head (1 – primary,…, 8 – tertiary with a scientific degree) |
| Residence | place of the residence (1 – city/town larger than 500,000 inhabitants,…, 6 – village) |
| Fulltime  | working time of the household head (1 – person working full-time, 0 – person working part-time) |
| Permanent | contract type of the household head (1 – permanent job/work contract of unlimited duration, 0 – temporary contract of limited duration) |
| Public    | ownership of the employment place of the household head (1 – public; 0 – private) |

Source: own elaboration.

Table 3. Summary statistics for covariates in models (%)

| Variable | Men       | Women     |
|----------|-----------|-----------|
| Sex      | Percentage share | 67.70  | 32.30  |
| Age      | Average, in years | 43.35  | 43.66  |
| Education | Levels 1–2 | 7.30    | 7.20    |
|          | Level 3    | 69.40   | 50.30   |
|          | Levels 4–5 | 1.30    | 3.70    |
|          | Levels 6–8 | 22.00   | 38.90   |
| Residence | Places 1–2 | 20.20   | 28.20   |
|          | Places 3–5 | 33.00   | 37.30   |
|          | Place 6    | 46.80   | 34.50   |
| Fulltime | 0          | 6.70    | 17.90   |
|          | 1          | 93.30   | 82.10   |
| Permanent | 0          | 22.90   | 31.30   |
|          | 1          | 77.10   | 68.70   |
| Public   | 0          | 83.30   | 67.40   |
|          | 1          | 16.70   | 32.60   |

Source: own calculation.

Table 4 presents the results of the aggregate and detailed Oaxaca-Blinder decomposition of log income inequalities between men’s and women’s households for the whole of Poland.

We have found that there is a positive difference between the mean values of log incomes for men’s and women’s households. The mean log income differential equals 0.042 (it means that the difference between men’s and women’s household income is about 4.2%). The difference between the mean log income values was decomposed into two components: the first one explaining the contribution of the different values of the models coefficients (the unexplained part), and the second one explaining the contribution of the differences of the attributes (the explained part). The unexplained effect is positive and gives us information about...
discrimination. The explained gap is negative, which means that the difference in the average log incomes between households is reduced by better women’s characteristics.

Table 4. Aggregate and detailed decomposition results for the income gap for Polish households run by men and women

| Variable     | Unexplained effect | Explained effect |
|--------------|--------------------|-----------------|
| Aggregate decomposition |                        |                 |
| Raw differential   | 0.042 ***          |                 |
| Unexplained effect | 0.101 ***          | (240%)          |
| Explained effect   | –0.059 ***         | (–140%)         |
| Detailed decomposition |                    |                 |
| Age            | –0.084 ***         | –0.001 *       |
| Education      | –0.021             | –0.075 ***      |
| Residence      | –0.011             | –0.028 ***      |
| Fulltime       | 0.119 ***          | 0.030 ***       |
| Permanent      | –0.007             | 0.012 ***       |
| Public         | 0.011 ***          | 0.003           |
| Cons           | 0.094 **           |                 |
| Total          | 0.101 ***          | –0.059 ***      |

Source: own calculation using Stata command ‘oaxaca’; the significance levels codes are as follows: *** 0.01; ** 0.05; * 0.1.

The negative values of explained components mean that the differences between average log incomes in households run by men and women are reduced by the better characteristics of female household leaders. The strong effect of different education levels of household heads can be noticed (women are usually better educated than men in Poland). There is also a strong household location effect (female households are more frequent in richer cities, male households in poorer villages). On the other hand, the positive values of full-time and permanent attributes possessed by the heads of households of both sexes increase income inequality (because men are more likely to be employed full-time and on a permanent basis). Women are discriminated against men due to more frequent part-time work and more frequent employment in the public sector (the positive unexplained components values for variables full-time and public) but not because of the age.

Figure 1 presents the results of the aggregate Oaxaca-Blinder decomposition of inequalities between the log incomes of men’s and women’s households for 16 voivodeships.
Figure 1. The unadjusted income gap (raw differential) and the results of the Oaxaca-Blinder decomposition of inequalities between the log incomes of men’s and women’s households by voivodeships

Source: own calculation.

There is a positive difference between the mean values of the log incomes of men’s and women’s households for 14 voivodeships and a negative difference for two voivodeships (in Podkarpackie and Podlaskie). The mean log income differential is the largest in the Lubuskie Voivodeship (0.158), followed by the Śląskie (0.13) and Warmińsko-Mazurskie (0.095) voivodeships, with the smallest in Podkarpackie (−0.044).

Voivodeship heterogeneity is not limited to the size of the gap but also concerns its composition. The unexplained effect is huge for the voivodeships with the high raw differential and is small for the provinces with the low raw differential. This part of the gender income gap is nowhere identified to be negative and gives us information about discrimination. The explained
gap is negative in 12 provinces (among others in Mazowieckie, Podkarpackie and Łódzkie). The negative value of this component means that the difference in the average log incomes between men’s and women’s households is reduced by the better women’s characteristics. In 2 voivodeships, the explained part is positive, that is, it increases the overall gap (in Warmińsko-Mazurskie and Opolskie). Our results provide novel insights into the regional dimension of the income gap. The observed characteristics play very different roles across the regions, featuring a higher explanatory power in regions with a low gap.

Since the Oaxaca-Blinder technique focuses only on average effects, we decomposed inequalities along the distribution of log incomes for men and women using the residual imputation approach (JMP-approach). The total differences between the values of log incomes have been calculated and the results are shown in Table 5. They are expressed in terms of percentiles (the symbols p5, ..., p95 mean 5th, ..., 95th percentile; e.g. the 25th percentile is the log income value below which 25% of observations may be found).

**Table 5. The total differences between the log incomes along the distribution**

| Voivodeship         | p5   | p10  | p25  | p50  | p75  | p90  | p95  |
|---------------------|------|------|------|------|------|------|------|
| Dolnośląskie        | 0.035| 0.124| 0.128| 0.117| 0.081| 0.067| 0.034|
| Kujawsko-Pomorskie  | –0.041| –0.003| 0.030| 0.029| 0.034| 0.093| 0.079|
| Lubelskie           | 0.180| 0.057| 0.048| –0.030| –0.013| 0.023| 0.130|
| Lubuskie            | 0.191| 0.161| 0.095| 0.201| 0.160| 0.097| 0.002|
| Łódzkie             | 0.116| 0.182| 0.105| 0.053| 0.025| 0.029| –0.021|
| Małopolskie         | 0.087| 0.007| 0.005| 0.001| 0.028| 0.053| 0.034|
| Mazowieckie         | 0.021| 0.045| 0.058| 0.005| 0.028| 0.043| 0.060|
| Opolskie            | 0.184| 0.237| 0.138| 0.075| 0.051| 0.092| 0.147|
| Podkarpackie        | 0.006| –0.089| 0.001| –0.027| –0.089| –0.014| 0.046|
| Podlaskie           | 0.159| –0.001| 0.019| 0.018| –0.021| 0.030| –0.002|
| Pomorskie           | 0.118| 0.131| 0.080| 0.049| 0.044| 0.049| 0.059|
| Śląskie             | 0.207| 0.230| 0.207| 0.147| 0.067| 0.028| 0.028|
| Świętokrzyskie      | 0.027| 0.076| 0.034| 0.021| 0.041| 0.018| 0.006|
| Warmińsko-Mazurskie | 0.201| 0.126| 0.102| 0.114| 0.045| 0.097| 0.037|
| Wielkopolskie       | 0.081| 0.020| 0.035| 0.000| –0.024| –0.021| 0.015|
| Zachodniopomorskie  | 0.109| 0.118| 0.000| –0.022| –0.013| 0.050| 0.022|

Source: own calculation.

For each voivodeship, the differences between the values of the log incomes of men’s and women’s households along the whole log income distribution were calculated. Then the
differences were decomposed into the sum of the unexplained and explained components (the results are presented in Figures 3, 4, 5, 6).

After assessing the income gap (the raw, the explained and the unexplained gap) for all 16 provinces, they were grouped using the hierarchical clustering methods. In the first step, the grouping was carried out by four methods: single, complete, average and Ward’s linkage with Euclidean distance. Then the best classification method was selected using the cophenetic correlation coefficient (the calculated values of this coefficient were: 0.633 for single, 0.585 for complete, 0.665 for the average and 0.594 for Ward’s linkage). We chose the average-linkage clustering algorithm, which allowed to group voivodeships into clusters as shown in Figure 2.

![Dendrogram for the average-linkage cluster analysis](image)

Figure 2. The result of grouping the voivodeships using the average-linkage clustering algorithm

Source: own elaboration using Stata.

As a result, four groups of voivodeships were selected:
- Group 1: Dolnośląskie, Łódzkie, Pomorskie, Śląskie and Lubuskie,
- Group 2: Świętokrzyskie, Wielkopolskie, Podkarpackie and Kujawsko-Pomorskie,
- Group 3: Małopolskie, Mazowieckie, Lubelskie and Podlaskie,
- Group 4: Opolskie, Warmińsko-Mazurskie and Zachodniopomorskie.

The shapes of the income gap are examined in Figures 3-6, where solid lines represent the total income gap; dashed lines denote the unexplained component and dotted lines indicate the explained effect.
Figure 3. The log income gaps of men’s and women’s households in group 1
Source: own elaboration using Stata.

Group 1 consists mainly of the highly developed provinces with high GDP per capita. The Dolnośląskie and Śląskie voivodeships are highly urbanized and industrialized regions. The most important regional industries are mining, metallurgy, the power industry, car manufacturing, and engineering. These are the richest provinces in Poland, with high wages and incomes, as well as low unemployment rates. Also highly industrialized are Łódzkie (textile and power industry) and Lubuskie (automotive, wood and furniture industry, food processing, services and transport), with medium incomes. The Pomorskie Voivodeship is a region with one of the lowest percentages of people working in agriculture and one of the highest with employment in the service sector, with high wages and incomes.

Highly industrialized regions usually offer well-paid jobs for men. Therefore, in these five provinces, the total income gap is large. But the gap has a decreasing shape and is falling as we move toward the top of the income distribution. The unexplained part is positive, presenting the existing effect of discrimination on the labor market. Its share is high in the whole range of income distribution. The explained differential (the effect of characteristics) is negative, which means that the properties possessed by both people’s groups decrease the inequalities.

Group 2 is made up of four provinces. The Świętokrzyskie, Podkarpackie and Kujawsko-Pomorskie voivodeships are the least urbanized regions in Poland, with a high percentage of people employed in agriculture. The wages and incomes there are below the national average.
Only the Wielkopolskie Voivodeship has a medium level of urbanization and industrialization, a high level of incomes and a low unemployment rate.

This group is characterized by the very low (often not statistically significant) total income gap along the income distribution. In the Podkarpackie Voivodeship, the total gap is even negative, which means that women’s households have higher incomes than men’s households. There is the bigger unexplained effect than the explained one. The effect of coefficients is positive and it widens at the top of the income distribution. The explained effect is negative.

Group 3 consists of the Mazowieckie, Małopolskie, Lubelskie and Podlaskie voivodeships. The Mazowieckie Voivodeship is a region with one of the highest shares of people living in an urban area and the highest employment rate in the service sector. The capital city provides the region with the highest level of wages and income, as well as a low unemployment rate. Mazovia is the center of science, research, education, industry and infrastructure in the country. Although Krakow is a very industrialized city, the Małopolskie Voivodeship has a high percentage of people living in rural areas with medium wages and incomes. The Lubelskie and Podlaskie voivodeships are regions with the highest shares of employment in agriculture and a low level of income.
For provinces in group 3, the total effect is low, but it widens at the bottom and the top of the income distribution, suggesting sticky floor and glass ceiling effects. The unexplained effect is bigger than the explained one. The effect of characteristics is negative and U-shaped, which means that the different values of characteristics of men’s and women’s households decrease the income inequalities especially in the middle range of the income distribution.

Group 4, the smallest group, consists of three medium-developed provinces: Opolskie, Warmińsko-Mazurskie and Zachodniopomorskie. The Opolskie Voivodeship is an industrial as well as an agricultural region. The Warmińsko-Mazurskie province is close to the national average of individual sectors in the employment structure, characterized by high unemployment, low wages and income. The Zachodniopomorskie Voivodeship has a high level of urbanization, a very low level of employment in agriculture and very high in the service sector. The unemployment rate is high and income is at a medium level.
In these provinces the total income gap is large and U-shaped (bigger at the bottom and the top of the distribution, indicating sticky floor and glass ceiling effects). For the lower income ranges, the explained effect is bigger than the unexplained effect. Its positive values mean that the worse characteristics of women’s households increase the income inequalities at the bottom of the log income distribution. For higher income ranges, the unexplained effect dominates (with the exception of the Opole region).

Conclusions

The objective of the study was to compare the distribution of household income in 16 Polish voivodeships, taking into account the gender differences in family heads. We expected that employing the advanced decomposition technique will provide deeper insights into the nature of income differentials. The obtained results confirm the regional variation of the gender income gap in the researched voivodeships in Poland. Men’s households have much higher incomes than women’s households in industrialized regions. The largest inequalities were recorded in the most affluent regions: the Śląskie, Opolskie, Dolnośląskie, Lubuskie, Łódzkie, Pomorskie and Warmińsko-Mazurskie voivodeships. Similar results were obtained by T. Słoczyński (2012),
who provided a description and an explanation of regional variation in gender wage gaps in Poland, and showed that the gap is especially large in the Śląskie Voivodeship. He found a strong gender wage discrimination in this region. Moreover, there men typically work in bigger firms than women and the distribution of women between different occupations and industries in Silesia is especially disadvantageous for their relative wages. A. Jędrzejczak (2015) believes that this kind of income inequality could be the result of the rapid economic growth in some regions that took place in the transformation period.

In order to move beyond an estimation based on mean values, we extended the decomposition procedure to different quantile points along the whole income distribution using the Residual Imputation Approach. Like J. Albrecht, A. Björklund and S. Vroman (2003) for Sweden, we have found sticky floor and glass ceiling effects in some voivodeships. Even before our study, A. Newell and M. Socha (2005) showed that many of the factors influencing wages, including gender, have a stronger impact in higher quantiles of the wage distribution. M. Rokicka and A. Ruzik (2010) found that the inequality of earnings between women and men tends to be larger at the top of the earnings distribution. The novelty of our study is that we analyzed the inequalities between household income distributions taking into account regional differences.

After assessing the raw, the explained and the unexplained gaps for all 16 NUTS2-regions, we grouped them into clusters using the hierarchical clustering method. Four voivodeship groups with similar inequality patterns were identified:

- heavily industrialized voivodeships with a large, decreasing income gap and a negative explained effect (Dolnośląskie, Łódzkie, Pomorskie, Śląskie and Lubuskie),
- weakly industrialized voivodeships with a very low income gap (Świętokrzyskie, Podkarpackie, Kujawsko-Pomorskie, but also Wielkopolskie),
- voivodeships with large agglomerations and the eastern regions characterized by a low gap in the middle of the income distribution and a negative, U-shaped explained effect (Małopolskie, Mazowieckie, Lubelskie and Podlaskie),
- medium-developed voivodeships with a large, U-shaped gap and a non-negative explained effect among the poorest (Opolskie, Warmińsko-Mazurskie and Zachodniopomorskie).

We have shown that there is an important diversity in the size of the gender income gap and in its underlying causes across the 16 Polish provinces. Such a gap is the result of many factors. The findings of this study depend on the used data set, the number of explanatory variables and the applied method of decomposition. Future studies will address the above-mentioned aspects.
The obtained results can be useful in helping social policymakers better understand the influence of various socioeconomic determinants on income inequality. This leaves room for them to implement policies that would rebalance the situation in the labor market and decrease the economic divergences between particular regions in Poland.

References

Albrecht, J., Bjorklund, A., Vroman, S. (2003). Is there a glass ceiling in Sweden? Journal of Labor Economics, 21, 145–177. DOI: 10.1086/344126.

Autor, D.H., Katz L.F., Kearney, M.S. (2005). Rising Wage Inequality: The Role of Composition and Prices. NBER Working Paper, 11628. DOI: 10.3386/w11628.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. Journal of Human Resources, 8 (4), 436–455. DOI: 10.2307/144855.

Boll, C., Lagemann, A. (2019). The gender pay gap in EU countries – new evidence based on EU-SES 2014 data. Intereconomics, 54 (2), 101–105. DOI: 10.1007/s10272-019-0802-7.

Chatterji, M., Mumford, K., Smith, P.N. (2011). The public-private sector gender wage differential in Britain: Evidence from matched employee-workplace data. Applied Economics, 43 (26), 3819–3833. DOI: 10.1080/00036841003724452.

DiNardo, J., Fortin, N.M., Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. Econometrica, 64 (5), 1001–1044. DOI: 10.2307/2171954.

Donald, S.G., Green, D.A., Paarsch, H.J. (2000). Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates. Review of Economic Studies, 67 (4), 609–633. DOI: 10.1111/1467-937X.00147.

Fortin, N.M., Lemieux, T. (1998). Rank Regressions, Wage Distributions, and the Gender Gap. Journal of Human Resources, 33 (3), 610–643. DOI: 10.2307/146335.

Fortin, N., Lemieux, T., Firpo, S. (2010). Decomposition Methods in Economics. National Bureau of Economic Research Working Paper Series, 16045. DOI: 10.3386/w16045.

Goraus, K., Tyrowicz, J. (2014). Gender Wage Gap in Poland – Can It Be Explained by Differences in Observable Characteristics? Working Papers, 11/2014 (128), University of Warsaw, Faculty of Economic Sciences.

Gosling, A., Machin, S., Meghir, C. (2000). The Changing Distribution of Male Wages in the U.K. Review of Economic Studies, 67 (4), 635–666. DOI: 10.1111/1467-937X.00148.
Hedija, V. (2017). Sector-specific gender pay gap: Evidence from the European Union countries. Economic Research – Ekonomska Istraživanja, 30 (1), 1804–1819. DOI: 10.1080/1331677X.2017.1392886.

Jędrzejczak, A. (2015). Regional Income Inequalities In Poland And Italy. Comparative Economic Research, 18 (4), 27–45. DOI: 10.1515/cer-2015-0027.

Juhn, Ch., Murphy, K.M., Pierce, B. (1993). Wage inequality and the rise in returns to skill. Journal of Political Economy, 101, 410–442. DOI: 10.1086/261881.

Jurajda, S. (2003). Gender wage gap and segregation in enterprises and the public sector in late transition countries. Journal of Comparative Economics, 31 (2), 199–222. DOI: 10.1016/S0147-5967(03)00040-4.

Machado, J.F., Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics, 20, 445–465. DOI: 10.1002/jae.788.

Newell, A., Socha, M. (2005). The Distribution of Wages in Poland. IZA Discussion Paper, 1485, Bonn.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14 (3), 693–709. DOI: 10.2307/2525981.

Olivetti, C., Petrongolo, B. (2008). Unequal pay or unequal employment? A cross-country analysis of gender gaps. Journal of Labor Economics, 26 (4), 621–654. DOI: 10.1086/589458.

Pena-Boquete, Y., De Stefanis, S., Fernandez-Grela, M. (2010). The distribution of gender wage discrimination in Italy and Spain: A comparison using the ECHP. International Journal of Manpower, 31 (2), 109–137. DOI: 10.1108/0143772101042232.

Rokicka, M., Ruzik, A. (2010). The Gender Pay Gap in Informal Employment in Poland. CASE Network Studies and Analyses, 406. DOI: 10.2139/ssrn.1674939.

Słoczyński, T. (2012). Próba wyjaśnienia regionalnego zróżnicowania międzypłciowej luki płacowej w Polsce. Studia Regionalne i Lokalne, 3 (49), 65–84.

Statistics Poland (2018). Methodological report. Household budget survey (Zeszyt metodologiczny. Badanie budżetów gospodarstw domowych). Warszawa.

Śliwicki, D. (2017). Szacunki luki płacowej w Polsce w ujęciu regionalnym. Optimum. Studia Ekonomiczne, 4 (88), 154–168. DOI: 10.15290/ose.2017.04.88.12.

Śliwicki, D., Ryczkowski, M. (2014). Gender Pay Gap in the micro level – case of Poland. Quantitative Methods in Economics, XV (1), 159–173.