

Aleksandra Majchrowska¹

Paweł Strawiński²

Regional diversity of gender wage gap in Poland
– New estimates based on harmonised data on wages

Abstract

The aim of this paper was to estimate gender wage gaps across 16 Polish NUTS2 regions and explain the differences between them. This paper was a first attempt to use harmonised dataset on individual wages in Poland. It was also a first attempt to verify the spatial monopsony model for the Polish economy.

The results of the analyses performed in the paper show that the harmonised dataset describes individual data on wages in Poland. We found that in more urbanised regions the average wages are higher than in the rural ones. In each of the 16 NUTS2 Polish regions women earn less than men. Raw differences in wages between men and women are the highest in most urbanised regions, but significant part of the differences in those regions can be explained by differences in workers characteristics, especially by different sectoral structure of employment. The unexplained part of the wage differences which in the literature is commonly attached to discrimination is the highest in rural regions of Eastern Poland. In more urbanized regions more competitive labour market conditions constrain employers' ability to engage in women discrimination.

Key words: gender wage gap, Poland, regional labour markets.

JEL codes: J31, R23, J16.

1. Introduction

The paper aims to estimate the gender wage gap (GWG) in Poland and in 16 NUTS2 Polish regions and analyse the differences in wages between men and women. Part of the differences in wages can be explained by the differences in characteristics of workers (education, work experience, place of work etc.). The differences in wages which cannot be explained by differences in endowments are identified in the literature with the effect of discrimination. According to the spatial monopsony model (Hirsch, 2009) this unexplained part of the wage differences between men and women should be lower in more urbanised and more competitive regions. Our paper is the first one which tries to verify this hypothesis for Polish regions.

Although regional differences in average wage level between Polish regions are significant, Polish economic literature lacks the empirical analyses of regional diversity of gender wage gap and their causes. The average wage level in capital region of Poland (Mazowieckie) in 2010 was, depending on the survey, around 40-50% higher than the wages in least developed regions of Eastern Poland. Most of the hitherto empirical literature on the gender wage gap focus on the variation of the gender pay gap between countries and its evolution over time. An aspect that attracts far less attention is the regional variation of the gap within the same country. To the best of authors' knowledge, in the case of Poland, there is only one paper (Słoczyński, 2012) where the problem of regional differences in gender wage gap in Poland was analysed.

¹ National Bank of Poland and University of Lodz.

² University of Warsaw.

The hitherto estimates of average gender wage gap in Poland were based on one of two datasets available: the Polish Labour Force Survey (PLFS; see for example Goraus, Tyrowicz, 2014 and van der Velde et al., 2013) and/or the Structure of Wages and Salaries by occupation (SWS; see for example Matysiak et al., 2010 and Magda, Szydłowski, 2008). Both datasets have, however, some disadvantages (they are explained in more detail in the data section).

In this paper the authors attempted to estimate gender wage gaps across 16 Polish NUTS2 regions and explain the differences between them basing on newly created, harmonized database on wages of individuals in Poland. We have linked the statistical information from the PLFS and the SWS databases and we received harmonised dataset on individual wages in Poland representative for the whole Polish economy (more detailed description of the dataset is presented in the section about the data (see also Strawiński, 2016).

This dataset allows us to re-estimate the gender wage gap for Poland and analyse the differences between 16 Polish regions (at the NUTS2 level). We start our empirical approach by estimating the gender wage gap in Poland and in each of the 16 NUTS2 regions with the three datasets available (the PLFS, the SWS and the harmonised data) and compare the results. We perform the standard Oaxaca (1973) – Blinder (1973) decomposition. In the second step, we separate the effect of job segregation and the gender wage gap. Following EC (2014) we measure job segregation as the percentage of employed in the highly feminised and masculinised occupational groups. Rejecting the highly feminised and masculinised occupational groups we are left with those groups with very high probability of finding a statistical match in a second group. In the third step, we check the robustness of our results performing the Nöpo (2008) decomposition.

The structure of the paper is as follows. In the second part the theoretical justification of regional differences in gender wage gaps and some hitherto empirical evidence are presented. The third part describes the statistical data used in the paper and empirical strategy. In the fourth part the empirical results are presented. The fifth part concludes.

2. Regional differences in gender wage gaps – theoretical justification and empirical evidence

Regional differences in wages between men and women have been explained on the monopsonistic discrimination theory. Hirsch (2009) presents a spatial monopsony model of the labour market in which he explains the differences in gender wage gaps between regional labour markets. The model assumes that workers are located at different places, while employers do not exist at each potential location. Some workers therefore have to commute and bear some travel costs, both direct and indirect. Direct cost are concerned with travelling, whereas indirect cost follows from the fact that travelling requires time and thus imposes some opportunity costs (Hirsch et al., 2010).

Since employers and the jobs they offer are not perfect substitutes to workers, competition among employers is imperfect and firms possess some monopsony power. The model assumes moreover that, due to their domestic responsibilities, women have higher average opportunity costs than men. It translates to lower spatial mobility of women and less choice of employers. Hirsch (2009) arrives therefore to the conclusion that firms have higher monopsony power over the female workers than men therefore they offer them lower wages giving rise to a gender wage gap. This part of the theory explains why women earn on average less than men.

Coming to regional dimension, Hirsch (2009) argue that the differences in wages of men and women between regions are due to differences in monopsonistic position of employers. More competition between employers in more urbanised areas should increase wages of both men and women. Moreover, higher competition between employers in more urbanised areas should constrains employers' ability to discriminate against women which should lead to lower gender wage gaps in more urbanised regions. This part of the theory explains the differences in wages of men and women between regions.

Hirsch et al. (2010) confirms the findings of the spatial monopsony model for German labour market. They analysed the regional differences in the gender wage gap in Germany and found that the unexplained gender wage gap for young workers is substantially lower in large metropolitan than in rural areas. Additionally, they confirmed that the differences between regions persist in time.

Regional differences in gender wage gaps may also be the result of different employment structure. The differences may come both from the personal characteristics of employees (level of education, work experience, number of children etc.) and from characteristics of employers (size of the firm, ownership sector, branch etc.). As the latter is concerned, empirical analyses confirm that bigger firm pay higher wages other things constant (Gibson, Stillman, 2009; Lallemand et al., 2005). Empirical analyses confirm that the gender wage gap increases along the wage distribution. For instance, Arulampalam et al. (2007) for eleven European countries found that gender pay gaps are typically bigger at the top of the wage distribution, a finding that is consistent with the existence of glass ceilings.

Differences in gender wage gap across regions can also be the result of the differences in human capital (both formal education and on-the-job training) and the differences in the rates of returns to endowments. Lopez-Bazo and Motellon (2009) analysed the effect of human capital on regional wage differentials in Spanish regions. Their results confirm that the regions differ in the endowment of human capital but also that the return that individuals obtain from it varies sharply across regions. Regional heterogeneity in returns is especially intense in the case of education. The differences in return to human capital account for a significant proportion of the differences in regional wage gaps in Spain.

As far as Poland is concerned statistical data confirm that on average women earn less than men. According to the Eurostat data the unadjusted gender wage gap³ in Poland is, however, relatively low comparing to the other European countries. In 2010 the gender wage gap in Poland (on average in industry, construction and services sectors) reported by Eurostat was around 5% and was the second lowest among the EU countries (after Slovenia with unadjusted gender wage gap equalled to 0.7). The highest differences between wages of men and women were noted in Estonia and amounted to 26%.

The estimates of adjusted (by both personal and firms' characteristics) gender wage gap in Poland vary significantly mainly due to different dataset used in the analyses. Among the most recent papers where gender wage gap in Poland was estimated we should mention: van der Velde et al. (2015), Goraus, Tyrowicz (2014), Śliwicki, Ryczkowski (2014), Mysiková (2012), Rokicka, Ruzik (2010), Matysiak et al. (2010) and Magda, Szydłowski (2008)⁴. For more comprehensive review of the main results of these papers see for instance Majchrowska et al., 2015 or van den Velde et al., 2013).

Besides the regional dimension appeared in many of hitherto analyses as a control variable, see for example Goraus and Tyrowicz (2014) or Matysiak et al. (2010), in only one paper (Śłoczyński, 2012) the regional differences in gender wage gaps in Poland have been analysed. Śłoczyński (2012) analysed the gender wage gaps in 16 NUTS2 regions in Poland in 2008 based on the Structure of Wages and Salaries (SWS) data in October 2008. He performed Oaxaca-Blinder and Juhn, Murphy and Pierce decomposition. His results show very strong diversification of gender wage gaps between Polish regions (from 25% in very urbanised Slaskie region, with relatively high share of working in the mining industry to 6% in Podkarpackie, one of the less developed regions in the Eastern part of Poland). Since SWS data were used in the paper, the results are representative for employees

³ 1 It is defined by Eurostat as the difference between the average gross hourly earnings of men and women expressed as a percentage of the average gross hourly earnings of men (http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics).

⁴ Among the earlier papers we should also mention Grajek (2003) and Adamchik and Bedi (2003). However, as they used data from the early transition period, their results cannot be directly compared with the latter ones.

working in firms with at least 10 employees. Moreover he did not verify the relation between regional gender wage gap and monopsony power of the employees.

In this paper we are going to re-estimate gender wage gaps across Polish regions using harmonised dataset on wages representative for the whole economy and verify the assumptions of spatial monopsony model. We aim at verifying the following hypotheses coming from the spatial monopsony model:

H1. More competition between employers in more urbanized regions of Poland should translate into higher wages of both men and women than in the more rural regions.

H2. Due to higher average indirect (opportunity) costs, women in each of the 16 Polish regions, earn on average less than men.

H3. Raw differences in wages between men and women are higher in more urbanized region due to more pronounced differences in workers characteristics.

H4. Unexplained gender wage gap is lower in more urbanized regions because more competitive labour market constrains employers' ability to engage in women discrimination.

In the following empirical analysis, we will investigate to what extent these four hypotheses generated by the model are confirmed by the data.

3. Data and empirical approach

As it was stated in the introduction the aim of the paper is to estimate the gender wage gap in Poland and in 16 NUTS2 Polish regions with the harmonized data on wages. All the previous estimates of gender wage gap in Poland were based on one of two datasets available: the Polish Labour Force Survey and the Structure of Wages and Salaries by occupation survey. Both datasets have some limitations. The PLFS data contains full information about the employment structure in the economy, but perform much worse in case of data on wages. Due to high percentage of non-responses to question about wages especially among the persons with relatively high income the data about wages are downward biased. The SWS data are fully reliable as they come from the employers accounting departments but they are representative only for enterprises with at least 10 employees. Taking into account that wages in micro-enterprises are much lower than in medium and big companies the SWS data on wage level must be upward biased.

Therefore, we decided to create the harmonised dataset which could be representative for the whole economy. Firstly, we standardised the definitions of socio-demographics and firm characteristics in both datasets. In order to combine data from different sets, information on province, gender, education and age of the respondents are used. The sample is divided on sub-samples formed by combination of region, gender, education level group and age group. Secondly, using the multiple imputation technique we have imputed the wages from the SWS data to employment structure given by the PLFS. Specifically, for persons working in micro sector we have left the wage and information given by the PLFS. For individuals working in bigger enterprises the wage data was imputed from the SWS. Wages are assigned to the dataset by multiple imputation technique based on the extended Mincer-type wage equation, which takes into account aside from the standard variables also NACE section and a large occupational group in which the respondent works. To achieve desired reliability we used the average of forty imputations from the SWS data.

As the result, the total sample of the newly created harmonised data set is about 100 thousand of observations. Its advantage is using wage information from two sources while preserving the demographic and residence structure of Polish population.

The multiple imputation technique is deeply rooted in statistical theory. It was first proposed by Rubin (1978) as a possible solution to the problem of survey's non-responses. Zhang (2007) described

three common approaches used in multiple imputations: the propensity score method, the predictive model method and the Markov chain Monte Carlo. In this research the predictive model method is utilised. Multiple imputations usually are used as within-data method. It is worth to note that Laaksonen (2006) stresses the need of good auxiliary data when dealing with imputation. They increase the quality of data. Similar exercise of replacing the PLSF wage information with the SWS wage information was previously used by Myck, Morawski and Mycielski (2007). Notwithstanding, for the imputation purpose they used kernel estimation based on the propensity score of employment.

As far as the empirical approach is concerned, we start with estimating the gender wage gap in Poland and in 16 NUTS2 Polish regions by standard two-component Oaxaca-Blinder decomposition. The method allows us to decompose the difference in average wages of men and women into two parts. The first part which is due to differences in characteristics of individuals, both personal characteristics (level of education, age, work experience etc.) and characteristics of firms they are employed by (size, ownership sector, branch etc.). The remaining (unexplained) part of the differences in wages is interpreted as the discrimination effect. In our paper (basing on hitherto research) we assume that the eventual discrimination in Poland concerns only women. Therefore, we assume that the non-discriminatory coefficient vector is the vector by men wages. Our equation takes therefore the following form:

$$D = w_M - w_F = B_M(X_M - X_F) + (B_M - B_F)X_F$$

where:

w_M (w_F) – average wage level in the group of men (women);

$B_M(X_M - X_F)$ – is so called explained part of wage gap – the part of the gap which is due to differences in characteristics of individuals;

$(B_M - B_F)X_F$ – is the remaining (unexplained) part of wage gap.

The above decomposition is based on extended Mincer-type wage equation:

$$\ln w_i = \beta_0 + \gamma X_i + \gamma X_j + \delta X_i X_j + \varepsilon_i$$

where:

X_i – vector of i personal characteristics of individuals;

X_j – vector of j firms' characteristics.

As far as the personal characteristics of individuals are concerned we included education and work experience. Education is measured as a number of years necessary to achieve certain level of education. Work experience is measured as the implied years of work experience (age minus preschool years and years in educational system). Implied values are used to achieve coherency in analysis on three sets of data used in the paper.

In the second vector several firm' characteristics are included such as dummy for public sector with private sector as the reference level, 1-digit level occupational groups and the economic section for main activity of the firm (according to the NACE classification). Additionally, the set of dummies describing the size of the firm is included. The population of firms has been divided into small firms (11-50 employees), medium sized firms (51-250 employees) and big firms (251 and more employees), with establishments with no more than 10 employees as a baseline category. Moreover we allow for all possible interactions between control variables to better adjust wage-equation parameters for the segregation effect.

As the aim of the paper is to estimate differences in gender wage gap across regions, different sectoral employment structure between regions has been taken into account, in particular the fact that the share of workers employed in the agricultural sectors amounts in some Eastern regions of Poland to 30% of all employees. Because most of the workers in the agriculture do not receive

regular salaries but the income from agricultural production we decided to exclude the agricultural workers from our sample.

In the first step, we estimate the gender wage gap in Poland and in 16 NUTS2 Polish regions with the Oaxaca-Blinder decomposition using our harmonised data on wages. To check how well our data performs and to compare our estimates with the previous ones we estimate the gender wage gaps with other datasets available: Polish Labour Force Survey and Structure of Wages and Salaries by occupation.

One of the disadvantages of traditional Blinder-Oaxaca decomposition is the fact that it estimates earnings equations for all working females and all working males without restricting the comparison only to those individuals with comparable characteristics by recognizing gender differences in the supports (Goraus, Tyrowicz, 2014). It is therefore necessary to make “out-of-the-support” assumption that the fitted regression surface can be extended for individual characteristics that have not been found empirically in the data set, using the same estimators computed with the observed data.

Therefore in the second step we adjust the sample of all individuals by leaving only those individuals for whom the probability of finding a statistical match in a second group is relatively high. The differences in wages of men and women may not be the result of the discrimination but the result of the fact that most men and women perform completely different tasks. The recent analyses of differences in employment structure of men and women in Polish labour market (see Strawiński et al., 2016) indicate that more than half of all working women work in feminised occupations and almost half of working men work in the masculinised occupations.

Following the approach of European Commission (see Burchell et al., 2014) we decided to eliminate male-dominated and female –dominated occupational groups (at the 3-digit level). By doing this we exclude the occupational groups where the ratio of men to women working is highly asymmetric (e.g., miners or nurses). An occupation is defined as male-dominated if more than 60% of the employees in that occupation are male, female-dominated if more than 60% of employees are female, and mixed if the proportions of men and women are between 40% and 60%. The European Commission (see Burchell et al., 2014) underlines that the 0.4-0.6 interval for mixed occupations is wide enough to take into account only the groups of workers with similar share of men and women. If the interval was wider (0.3-0.7 for example) the occupation could have more than twice men than women and still be treated as mixed.

By doing this adjustment of the sample we expect to eliminate the differences in wages of men and women which are not the effect of discrimination⁵ but are due to segregation of men and women into different occupations⁶. We expect this segregation bias to be positive so we expect that the adjusted (for the segregation bias) gender wage gap in Poland will be lower than the aggregate gender wage gap.

The difference between our approach and the standard Oaxaca-Binder decomposition could be viewed as the difference between the local average treatment effect and the average treatment effect. The former is the average difference between locally comparable units, while the latter is the sample or population difference.

As the robustness check we perform Nöpo decomposition which accounts for gender differences in the supports. This decomposition consists of four additive elements (Nöpo, 2008):

⁵ Discrimination exists when female or minority workers—who have the same abilities, education, training, and experience as white male workers—are accorded inferior treatment with respect to hiring, occupational access, promotion, wage rate, or working conditions (McConnell., Brue, 1986).

⁶ Occupational segregation can be said to exist when the distribution of occupations within one demographic group is very different from the distribution in another. With respect to gender, occupational segregation is reflected in there being female-dominated occupations and male-dominated ones (Ehrenberg, Smith, 1994).

$$\Delta = \Delta_M + \Delta_X + \Delta_F + \Delta_O$$

The first component is the part of the gap that can be explained by differences between two groups of males – those who have characteristics that can be matched to female characteristics and those who do not. This component accounts for that part of the gap that would disappear in the event that there are no males with combinations of characteristics that remain entirely unmatched by females. Alternatively, this component would also disappear if those males with individual characteristics that are not matched by females were paid, on average, as the average matched males. It is computed as the difference between the expected male wages out of the common support minus the expected male wages in the common support, weighted by the probability measure (under the distribution of characteristics of males) of the set of characteristics that females do not reach (Nõpo, 2008).

The second component (Δ_X) is the part of the wage gap that can be explained by differences in the distribution of characteristics of males and females over the common support and corresponds to explained component of Oaxaca-Blinder decomposition $B_M(X_M - X_F)$ (Nõpo, 2008).

The third component is the part of the gap that can be explained by the differences in characteristics between two groups of females, those who have characteristics that can be matched to male characteristics and those who do not. It accounts for that part of the gap that would disappear should it ever be the case that all females would have characteristics that can be matched to the population of males. It would also disappear unmatched females are paid, on average, as much as matched females. It is computed as the difference between the expected female wages, in and out of the common support, weighted by the probability measure (under the distribution of characteristics of females) of the set of characteristics that males do not reach (Nõpo, 2008).

The fourth component corresponds to the unexplained part of the Oaxaca-Blinder decomposition $(B_M - B_F)X_F$. It is the share of the wage gap that cannot be attributed to differences in characteristics of the individuals and is typically attributed to a combination of both the existence of unobservable characteristics that explain earnings and the existence of discrimination (Nõpo, 2008).

If our predictions are true, the unexplained part of gender wage gap adjusted for the segregation bias would be comparable to the unexplained part of the Nõpo decomposition.

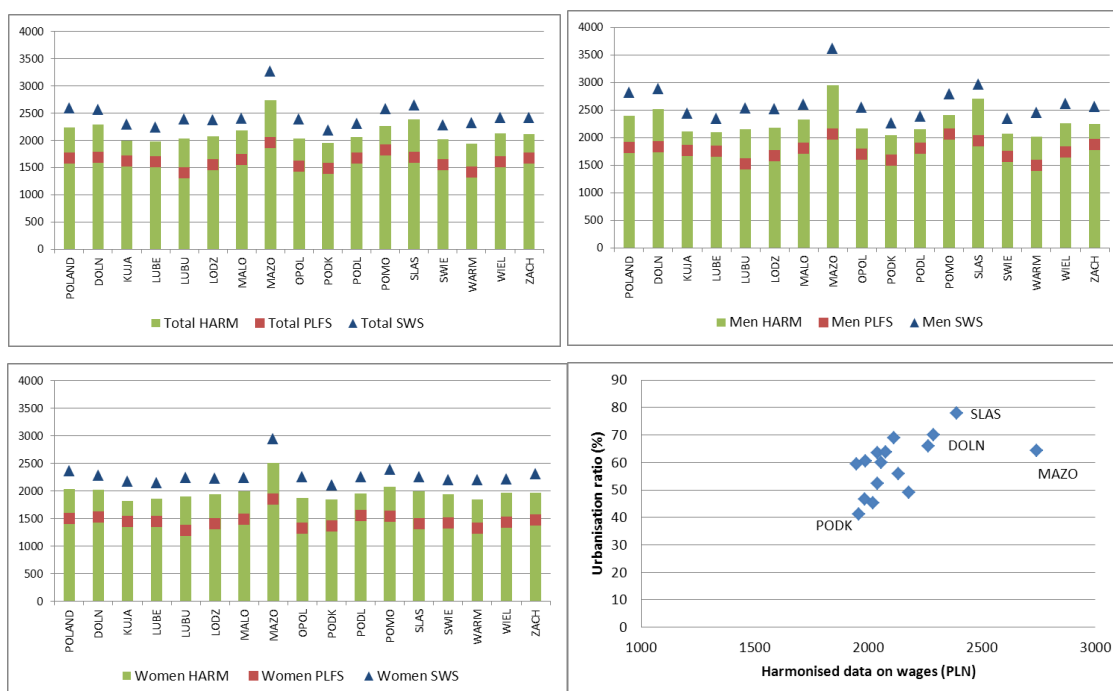
At the end we compare the unexplained part of the gender wage gap obtained with Oaxaca-Blinder and Nõpo decomposition with regional urbanisation ratio. If our fourth hypothesis is true, the unexplained part of the wage gap should be significantly lower in urbanised than in rural regions.

4. Gender wage gap in Poland and in 16 NUTS2 regions - Empirical results

In the first step we compare the average level of wages of men and women across regions using the three datasets available. If our harmonised dataset on wages performs well the average wage level should lie in between the PLFS and the SWS data. Looking at the mean wage level of men and women by the three data sources (the SWS gross wages has been recalculated to net wages to achieve comparability; see Figure 1) we can confirm that both in the case of men and women the harmonized data on wages lie in between the PLFS and the SWS data on average in Poland and in each of the 16 regions.

Comparing the average wage level and the urbanisation ratio we can confirm first hypothesis. More competition between employers in more urbanized regions of Poland translates into higher wages of both men and women than in the more rural regions. We can also confirm second hypothesis. Due to higher average opportunity costs, women in each of the 16 Polish regions, earn on average less than men. In the next step we will analyse gender wage gap in Poland and verify third and fourth hypothesis.

Figure 1. Average net wages of men and women in Poland and in 16 NUTS2 regions according to harmonised dataset, the PLFS and the SWS datasets in 2010 and relation between average wages (harmonised data) and urbanisation ratio



Source: PLFS, SWS and <http://eregion.wzp.pl/wspolczynnik-urbanizacji-dla-poszczegolnych-województw-w-latach-2010-2013>; own estimates.

Estimates of overall GWG in Poland

In the first step we used harmonised dataset on wages to estimate the gender wage gap in Poland in 2010. We compare the obtained results with the ones estimated with PLFS and SWS datasets.

The results show that gender wage gap in Poland in 2010 estimated on the harmonised dataset equals 15.7% and is significantly lower than the number obtained with PLFS data (19.8%) and significantly higher than the gender wage gap obtained with SWS data (see Table 1). All gender wage gap estimates are significant at 1% significance level.

Table 1. Oaxaca-Blinder decomposition of gender wage gap in Poland in 2010 estimated on harmonised dataset, on PLFS data and on SWS data

	HARM		PLFS		SWS	
	Value of parameter	t-ratio	Value of parameter	t-ratio	Value of parameter	t-ratio
$\ln(w_M)$	7.689	3861.811	7.446	2145.987	7.933	6676.172
$\ln(w_F)$	7.532	3652.118	7.248	2049.261	7.807	7500.507
$\ln(w_M) - \ln(w_F)$	0.157	54.734	0.198	39.972	0.126	79.948
$B_M(X_M - X_F)$	-0.013	-5.143	-0.009	-2.240	-0.015	-13.262
$(B_M - B_F)X_F$	0.170	72.908	0.207	41.849	0.142	88.607
Number of observations	101947		43957		681747	

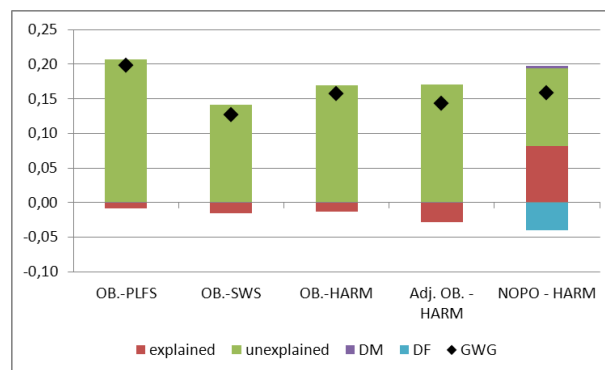
Source: own estimates.

The results above are rather consistent with authors' intuition. The SWS data provides higher estimate of average wages of both men and women than the hypothetical country average. This is a straight consequence of the limited scope of the SWS survey. However, due to the fact that more women than men work in the small firms and financial sector (not covered by the SWS data), we can expect that the upward bias is higher in case of women. If it is true, the gender wage gap estimated with the SWS data will be lower than the one estimated on harmonised dataset. Following the intuition, one can expect that at the regional level the average wage of women based on the SWS data should be more upward biased in regions with relatively high share of small enterprises (less developed, more rural regions).

The prediction is much more difficult in case of the PLFS data. Wages reported in the PLFS data are lower on average than that of in the SWS data at least for two reasons. The first is that the data cover all economy, not enterprises employing 10 or more workers. The second is well known phenomenon of high share of refusals to answer the question about wages among the people with relatively high income in the PLFS survey (see, for instance Myck, Morawski and Mycielski 2007 or Strawiński 2015). Those information's, however, are not sufficient to derive conclusions regarding the distribution of the expected differences across regions and gender.

Looking at the decomposition of gender wage gap in Poland (see Figure 2) we can notice that the explained part of the gap is negative (irrespective of the dataset used) indicating that on average women had better characteristics than men (detailed decomposition shows that that women are more educated than men and more often work in the public sector) and should be better paid than men. The adjusted (for labour market characteristics) gender wage gap (or unexplained part of the wage gap) is therefore higher than the raw gap indicating that some other factors are not taken into account or women are wage discriminated on the labour market.

Figure 2. Gender wage gap (GWG) in Poland in 2010 and its decomposition by different methods and datasets



Source: own estimates.

One of these factors is the different employment structure of men and women in the labour market. Strawiński et al. (2016) analysed segregation on Polish labour market in 2000-2013 and found that women much more often work in jobs which require a greater degree of soft skills and less risky. Significantly more men perform more physical jobs in industry and construction.

Having in mind that the differences in wages between men and women may at least be partly explained by different tasks they perform, in the next step we eliminated from the sample all employees working in the male-dominated or female-dominated occupational groups (at 3-digit level of classification) as explained in the previous section. We are left with only those occupational groups where the probabilities of finding a statistical match for men or women in relatively high. In this section we perform the analyses only on the harmonised dataset. Then, the adjusted by segregation bias gender wage gap is compared with the one obtained on the whole sample. The

intuition is that part of the differences in gender wage gap estimated on the whole sample of individuals is not due to discrimination, but is the result of different employment structure of men and women. After controlling for the segregation bias, the adjusted gender wage gap should then be lower than the total one.

After eliminating the male-dominated and female-dominated occupations we are left with 26 occupational groups at 3-digit level and around 14 thousands of observations (having around 100 thousands in the initial sample). The estimates of adjusted by segregation bias gender wage gap in Poland show that in fact it is lower than the one estimated on the whole sample. The total gender wage gap amounts to 15.7% - the adjusted one equals 14.3%. The results confirm, therefore, that at least part of the total gender wage gap can be assigned to the segregation bias.

What is interesting – in the adjusted for the segregation bias sample, the explained part is twice higher than in the whole sample of individuals (Figure 2). It indicates that when we look at only those occupations where men and women perform similar tasks, women have on average better labour market characteristics and should be paid more than men. The unexplained part is therefore higher than the raw wage gap and amounts to 17.1%.

To check the robustness of the above results we performed the Nöpo decomposition on the whole sample of harmonized data on wages. The results indicate that men out of common support earn only 0.4 log pp. more than men in the common support, but women out of common support earn 4 log pp. less than women in the common support. The findings are consistent with our intuition. The component which corresponds to differences in characteristics of men and women in the common support is positive (8 log pp.) which indicates that in the common support women have worse labour market characteristics than men. The differences in endowments explain however less than one third of the wage differences between men and women in the common support. The overwhelming majority of the wage differences remain unexplained (Figure 2).

Regional differences in gender wage gaps in Poland

In this section we aim at estimating gender wage gap in each of 16 NUTS2 Polish regions and answer a question which factors are responsible for the differences between regions. Firstly, we analyse the gender wage gap on the whole sample of the harmonised dataset and compare the results with the ones from the PLFS and the SWS datasets. Then we limit our sample to the occupational groups with masculinisation ratio 0.4-0.6 and estimate the adjusted by segregation bias gender wage gaps in each region. Finally, to check the robustness of the results, we perform Nöpo decomposition for each of the regions separately.

All the estimated regional gender wage gaps are significant at 1% significance level (Figure 3). Although the estimated gender wage gaps with the three dataset show some similarities (the highest gender wage gap is in Slaskie region with all three datasets), the differences are notable (see Figure).

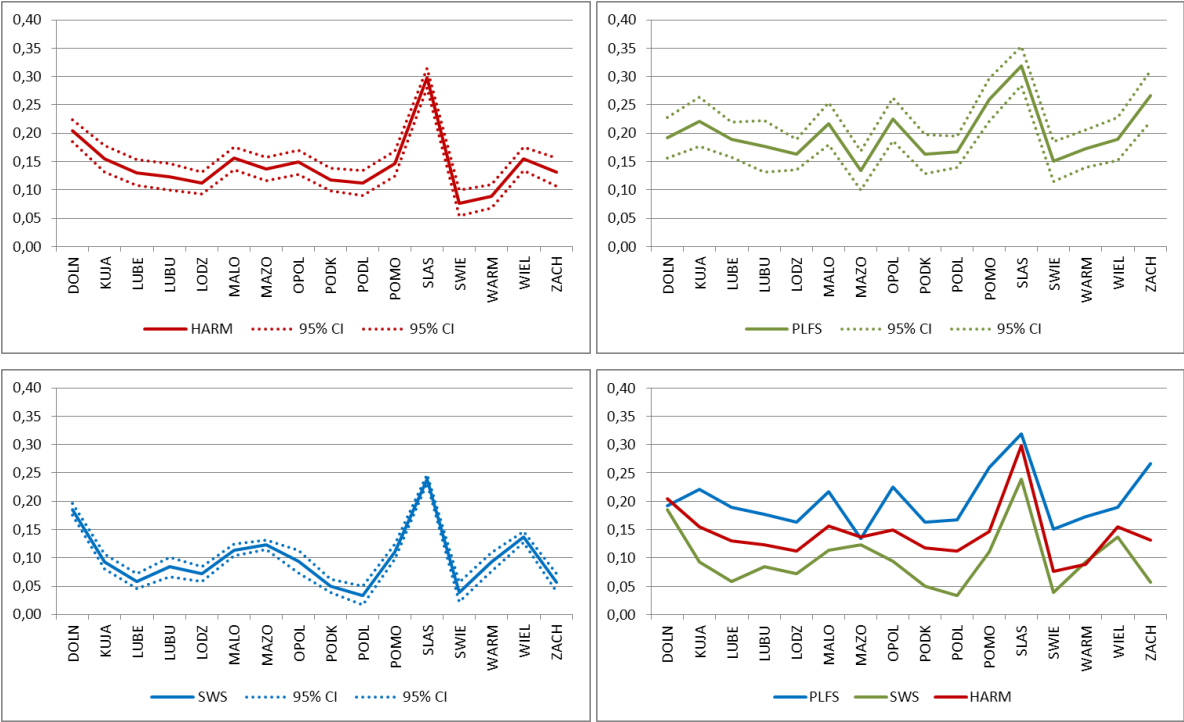
Comparing with the results performed on the three datasets we can notice that the gender wage gaps estimated with on the harmonised dataset in most of the regions lie between the ones estimated with PLFS data and SWS. The regional gender wage gaps estimated with the PLFS data are the highest. In almost all regions they are higher than the gender wage gaps estimated with the harmonised dataset. The two exceptions are Dolnoslaskie and Mazowieckie regions. The regional gender wage gaps estimated with the SWS data are the lowest. In almost all regions (with exception of Warminsko-mazurskie) they are lower than the gender wage gaps estimated with the harmonised dataset.

The differences between wages of men and women across Polish regions are significant. Looking at the results from the harmonised dataset we can notice that raw gender wage gaps varied in 2010 between 7.7% in Swietokrzyskie and 29.8% in Slaskie region (the average gender wage gap in Poland was 15.7%). When looking more carefully at the results we can notice that there are only two regions

where differences in wages between men and women are relatively high – this is Slaskie and Dolnoslaskie regions – both with significant share of men working in very well paid mining sector. In 2010 12% of men working in Slaskie and 4% in Dolnoslaskie were employed in coal mining industry. In the case of women these shares were respectively 1.3 and 0.6%.

In general we can confirm our third hypothesis. Raw differences in wages between men and women are higher in more urbanized region due to more pronounced differences in workers characteristics.

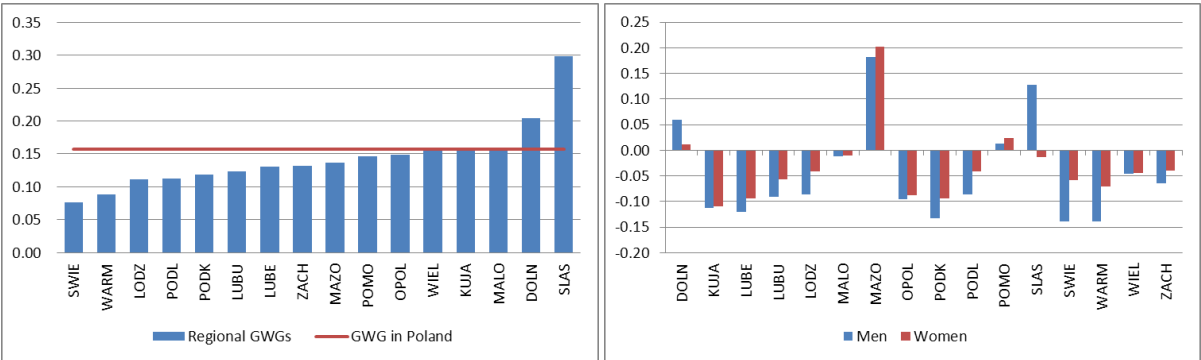
Figure 3. Gender wage gaps in 16 NUTS2 regions in Poland in 2010 (% of men wages) with different datasets



Source: own estimates.

It is worth to underline that the high differences in wages between men and women across regions can be the result of differences both in wages of men and wages of women across regions. When we look at the data we can notice that differences in regional gender wage gap are to higher extent the result of differences in wages of men than women. For example in Dolnoslaskie and Slaskie (mining regions) wages of men in 2010 were respectively 6 and 13% higher than average wages of men in Poland but the wages of women were almost at the average level in the country (Figure 4). On the contrary relatively low gender wage gaps in Eastern, rural regions of Poland are the result of much lower wages of men than the average.

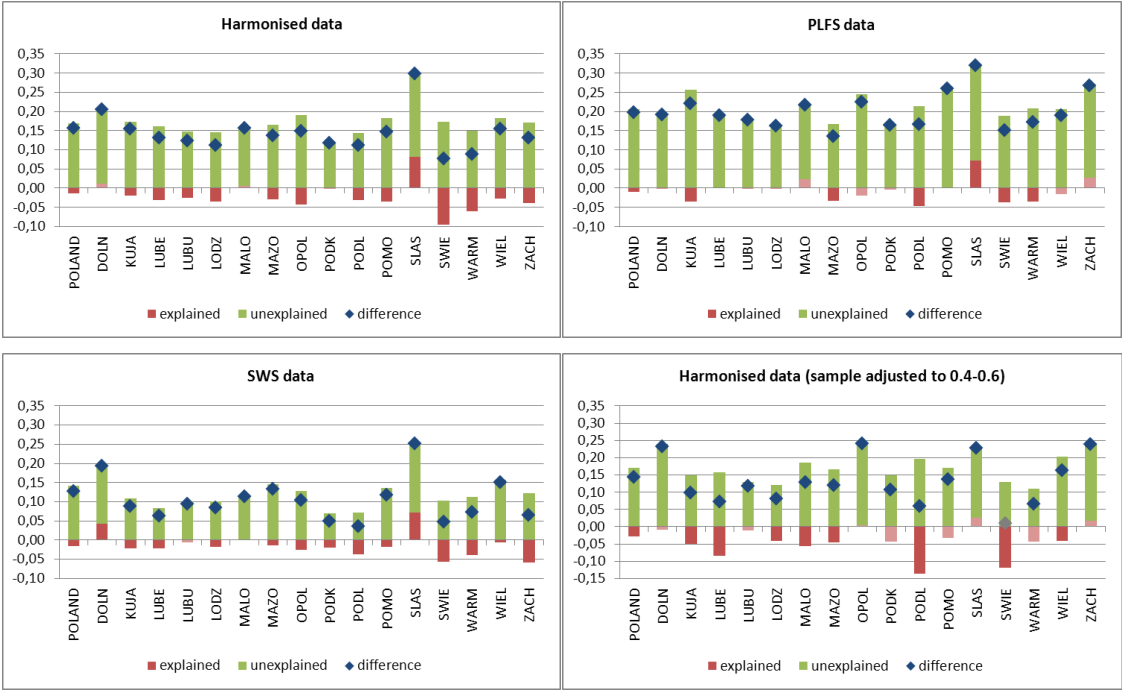
Figure 4. Regional gender wage gaps in Poland (left; log pp.) and the differences between wages of men and women by regions comparing to average wages in Poland (right panel; log pp.)



Source: own estimates.

Now we look on the decomposition of gender wage gaps by regions calculated with Oaxaca-Blinder decomposition. We work on the whole sample of harmonised data in 2010. The first conclusion is that the differences in endowments are not the factors which can explain the differences in wages between men and women across region (Figure 5). The explain part in most of the regions is negative indicating that women have better labour market characteristics and should be paid more than men. Only in Slaskie region the explained part of the wage gap is positive and significant (in Dolnoslaskie and Malopolskie it is positive, but not statistically significant) indicating that almost one third (27%) of the wage differences between men and women in that region can be explained by their characteristics. More detailed decomposition indicates that wages of men are higher mostly because of the differences in employment structure (men working in the mining sector).

Figure 5. Oaxaca-Blinder decomposition of gender wage gaps in 16 NUTS2 Polish regions in 2010 according to harmonised data, PLFS data and SWS data



In light red colour we indicated insignificant explained part of wage gap.

Source: own estimates.

The picture is not very much changed when we look at the results of Oaxaca-Blinder decomposition performed on the PLFS and SWS data (see Figure 5). With PLFS data the explained part in most of the

regions is very small and not significant. Only in Slaskie region 23% differences in wages of men and women can be attributed to different characteristics of individuals. According to SWS data 22% and 28% of differences in wages of men and women in Slaskie and Dolnoslaskie respectively can be explained by differences in characteristics. In most of the regions the explained part is negative.

Having in mind that part of the differences in wages between men and women can be the result of different employment structure in the next step we excluded from the sample all the individuals in male-dominated and female-dominated occupations. In each of the regions we left only individuals in 3-digit level occupational groups with masculinisation ratio between 0.4-0.6.

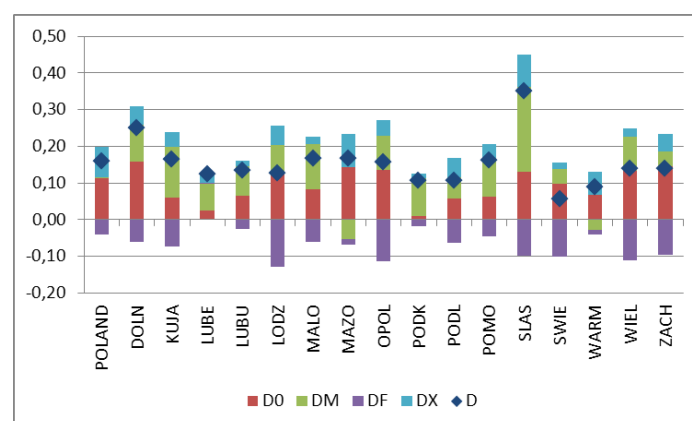
The results indicate that in the sample adjusted for segregation bias the differences in wages between men and women in most of the regions are smaller than in the whole sample (see Figure 5). They varied in 2010 from 1.1% (however not significant) in Swietokrzyskie to 24% in Zachodniopomorskie (14.3% on average in Poland). It confirms our intuition that part of the raw wage differences between men and women in Poland is due to effect of segregation.

The results of Oaxaca-Blinder decomposition shows that differences in endowments cannot explain the differences in wages between men and women on regional labour market. In those regions where the explain part is significant it is negative indicating that if women had the same vector of parameters as men, they should have been paid more than that men due to on average better labour market characteristics.

To check the robustness of the results we performed the Nöpo decomposition for each of the 16 NUTS2 regions in Poland separately. Looking at the contribution of the first factor (DM, see Figure 6) we can notice that in most of the regions men out of the common support earn more than men in the common support. The highest (not surprisingly) difference is observed in Slaskie (the difference is 23 log pp.). In other regions the differences in men wages vary from 4 to 12 log pp. Only in two regions (Mazowieckie and Warminsko-mazurskie) men out of the common support earned in 2010 respectively 5 and 3 log pp. less than men in the common support. The difference between wages of women out and in the common support (DF in Figure 6) is negative almost in every region. The highest is observed in Lodzkie (11 log pp.) and Wielkopolskie (8 log pp.). Only in Lubelskie wages of women in and out of the common support are the same.

The component which shows the differences in the distribution of characteristics of males and females over the common support in most of the regions is positive and very from 1 log pp. in Lubuskie to 9 log pp. in Mazowieckie and Slaskie regions. The results of Nöpo decomposition shows that significant part of the differences in wages between men and women cannot be explained by differences in characteristics of individuals which confirms the previous results of Oaxaca-Blinder decomposition.

Figure 6. The results of Nöpo decomposition for 16 NUTS2 Polish regions in 2010

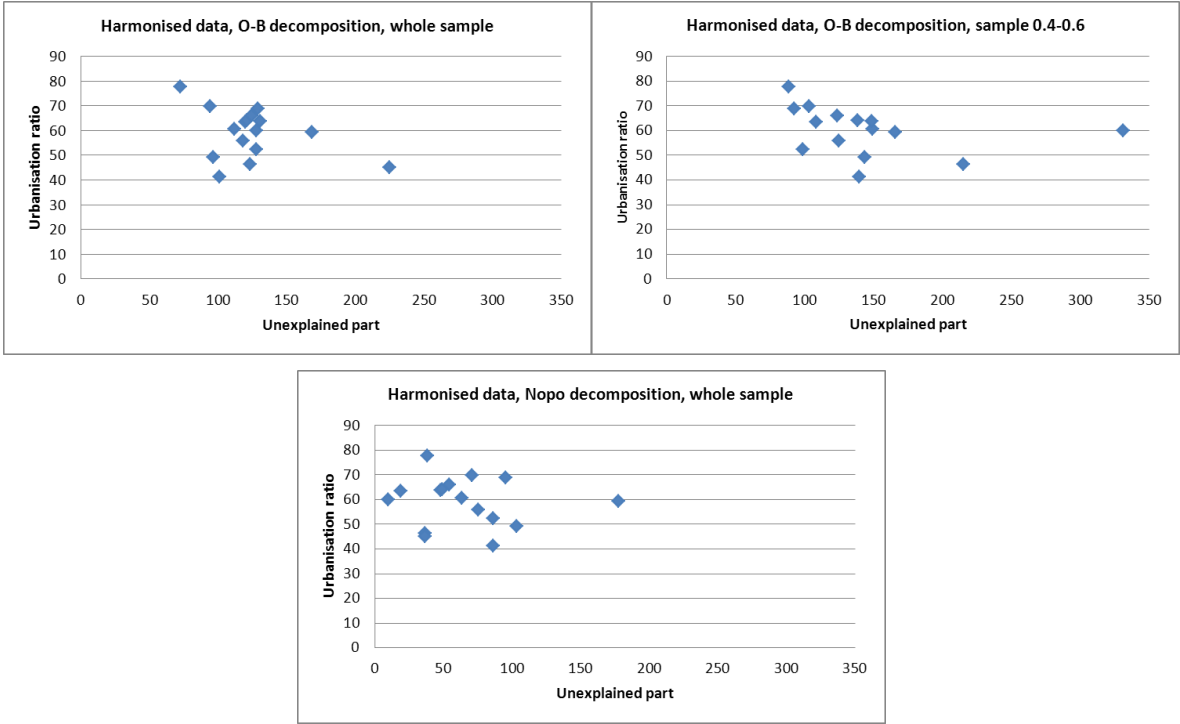


Source: own estimates.

In the last step we check the relation between unexplained gender wage gap and urbanisation ratio. According to spatial monopsony model in more urbanised regions the discrimination effects should be lower than in rural regions due to more competition between employers.

Figure 7 presents the relation between unexplained gender wage gap (in % of total differences in wages between men and women) and urbanisation ratio. We can notice that irrespectively of the decomposition method used (Oaxaca-Blinder or Nöpo) the relation is negative confirming that unexplained gender wage gap is lower in more urbanized regions because more competitive labour market constrains employers' ability to engage in women discrimination. We can therefore confirm our last hypothesis.

Figure 7. Relation between unexplained gender wage gap (% of total differences in wages between men and women) and urbanisation ratio (%) across 16 NUTS2 Polish regions in 2010



Source: own estimates.

5. Conclusions

This paper was a first attempt to use the newly created by authors, harmonised dataset on individual wages in Poland. It was also a first attempt to verify the spatial monopsony model for the Polish economy.

The results of the analyses performed in the paper show that the harmonised dataset describes individual data on wages in Poland very well. In particular, harmonised data confirm that the Polish Labour Force Survey data on wages are downward biased. On the other hand, the new dataset indicates that the Structure of Wages and Salaries data are upward biased in relation to average wage in the economy.

We managed to confirm all the initial hypotheses based on the spatial monopsony model. In more urbanised regions the average wages are higher than in the rural ones. In each of the 16 NUTS2 Polish regions women earn less than men. Raw differences in wages between men and women are the highest in most urbanised regions, but significant part of the differences in those regions can be explained by differences in workers characteristics, especially by different sectoral structure of

employment. The unexplained part of the wage differences which in the literature is commonly attached to discrimination is the highest in rural regions of Eastern Poland. In more urbanized regions more competitive labour market conditions constrain employers' ability to engage in women discrimination.

References:

Adamchik, V. A. and Bedi, A. S. (2003) Gender Pay Differentials during the Transition in Poland, *The Economics of Transition*, Vol. 11, Issue 4, 697-726.

Arulampalam W., Booth A.L., Bryan M.L. (2007) Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution, *Industrial and Labor Relations Review*, Vol. 60, No. 2 (Jan., 2007), pp. 163-186.

Barth E., Dale-Olsen H. (2009) Monopsonistic discrimination, worker turnover, and the gender wage gap, *Labour Economics*, Elsevier, vol. 16(5), pp. 589-597, October.

Blinder, A.S. (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, 8(4), pp.436-455.

Burchell B., Hardy V., Rubery J., Smith M. (2014) A New Method to Understand Occupational Gender Segregation in European Labour Markets, European Commission - Directorate-General for Justice.

Gibson, J., Stillman, S. (2009) Why Do Big Firms Pay Higher Wages? Evidence from an International Database, *Review of Economics and Statistics*, Vol. 91, No. 1, 213-218.

Goraus, K., Tyrowicz, J. (2014) Gender Wage Gap in Poland – Can It Be Explained by Differences in Observable Characteristics?, *Ekonomia*, No. 36.

Grajek, M. (2003) Gender Pay Gap in Poland, *Economic Change and Restructuring*, *Economics of Planning*, No. 36, pp. 23-44.

Guyot A., Berwing S., Lauxen-Ulbrich M. (2009) Income Differentials on Regional Labour Markets in Southwest Germany, *Panoeconomicus*, Vol. 56, No. 3 (September), 379-396.

Hirsch (2009) The Gender Pay Gap Under Duopsony: Joan Robinson Meets Harold Hotelling, *Scottish Journal of Political Economy*, Vol. 56, Issue 5, pp. 543-558, November 2009

Hirsch B., König M., Möller J. (2013) Is There a Gap in the Gap? Regional Differences in the Gender Pay Gap, *Scottish Journal of Political Economy*, Volume 60, Issue 4, pages 412–439, September 2013.

Laaksonen Seppo (2006) Need for High Quality Auxiliary Data Service for Improving the Quality of Editing and Imputation, in: UNStatistical Commission (2006) *Statistical Data Editing*, vol 3. Impact on Data Quality, New York and Geneva.

Lallemand, T., Plasman, R., Rycx, F. (2005) Why do large firms pay higher wages? Evidence from matched worker-firm data, *International Journal of Manpower*, Vol. 26 No. 7/8, 705-723.

Lopez-Bazo E., Motellon E. (2009) Human capital and regional wage gaps, IREA Working Papers, 2009/24, Research Institute of Applied Economics, University of Barcelona.

Łatuszyński, K. and Woźny, Ł. P. (2008) Zróźnicowanie wynagrodzeń kobiet i mężczyzn na polskim rynku pracy w 2004 roku, in: W. Pacho, M. Grabicz (ed.) *Wzrost gospodarczy a bezrobocie i nierówności w podziale dochodu*, Warsaw School of Economics, Warsaw.

Magda, I. and A. Szydłowski (2008) Płace w makro i mikroperspektywie in M. Bukowski (ed.) *Zatrudnienie w Polsce 2007 – Bezpieczeństwo na elastycznym rynku pracy*, Ministry of Labour and Social Policy, Warsaw.

- Majchrowska A., Strawiński P., Konopczak K., Skierska A. (2015) Gender wage gap by occupational groups in Poland, Lodz Economics Working Papers, 3/2015, Faculty of Economic and Sociology, University of Lodz.
- Matysiak, A., A. Baranowska and T. Słoczyński (2010) Kobiety i mężczyźni na rynku pracy in M. Bukowski (ed.) *Zatrudnienie w Polsce 2008 – Praca w cyklu życia*, Human Resources Development Center, Warsaw.
- Mysikova, M. (2012) Gender Wage Gap in the Czech Republic and Central European Countries, *Prague Economic Papers*, Issue 3, p. 328-346.
- Nöpo, H. (2008), Matching as a Tool to Decompose Wage Gaps, *Review of Economics and Statistics*, 90(2), 290-299.
- Oaxaca, R.L. (1973) Male–female wage differentials in urban labor markets, *International Economic Review*, 14(3), pp. 693–709.
- Rubin, D.B. (1978). Multiple imputations in sample surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 20–34.
- Strawiński P. (2016) Ujednolicone dane o polskich płacach 2006-2010, paper in progress.
- Strawiński P., Broniatowska P., Majchrowska A. (2016) Zmiany zawodowej i sektorowej struktury zatrudnienia kobiet w Polsce w latach 2000-2013, *Gospodarka Narodowa*, in print.
- Śliwicki, D., Ryczkowski M. (2014) Gender pay gap in the micro level – case of Poland, *Quantitative Methods in Economics*, Vol. XV, No. 1, 159-173.
- Taniguchi k., Tuwo A. (2014) New Evidence on Gender Wage Gap in Indonesia, *Asian Development Bank Working Paper*, 404.
- van der Velde, L., Tyrowicz J., Goraus K. (2013) What is the True Gender Wage Gap? A Comparative Analysis Using Data From Poland, “Working Papers” 2013-28, Faculty of Economic Sciences, University of Warsaw.
- Zhang (2003) Multiple Imputation: theory and method, *International Statistical Review*, 71(3), 581-592.

Appendix.

Table A1. Average level of wages in Poland and 16 NUTS2 regions according to harmonised dataset, the PLFS and the SWS datasets in 2010

	Harmonised dataset			PLFS			SWS			Urbanisation ratio
	Men	Women	Total	Men	Women	Total	Men	Women	Total	
DOLN	2521	2018	2287	1845	1523	1691	2894	2289	2577	69,9
KUJA	2123	1822	1989	1776	1444	1625	2444	2178	2306	60,6
LUBE	2098	1857	1984	1760	1446	1606	2348	2151	2239	46,5
LUBU	2157	1902	2042	1531	1279	1412	2540	2250	2389	63,5
LODZ	2189	1945	2077	1672	1405	1549	2534	2238	2376	63,8
MALO	2338	1999	2180	1809	1486	1649	2607	2244	2406	49,2
MAZO	2955	2508	2741	2064	1856	1967	3623	2954	3281	64,2
OPOL	2169	1869	2042	1703	1323	1533	2554	2256	2398	52,4
PODK	2050	1840	1958	1599	1367	1493	2275	2112	2192	41,3
PODL	2164	1950	2058	1810	1554	1682	2400	2257	2320	60,0
POMO	2415	2079	2264	2068	1546	1829	2798	2389	2589	65,9
SLAS	2708	1996	2388	1941	1412	1688	2976	2264	2649	77,8
SWIE	2076	1947	2020	1661	1415	1554	2360	2205	2281	45,1
WARM	2025	1852	1948	1502	1324	1421	2464	2205	2322	59,5
WIEL	2264	1962	2133	1748	1427	1607	2630	2222	2422	55,9
ZACH	2249	1968	2114	1883	1479	1684	2568	2313	2422	68,8
POLAND	2404	2041	2238	1825	1507	1676	2831	2373	2594	59,4

Source: PLFS, SWS and <http://eregion.wzp.pl/wspolczynnik-urbanizacji-dla-poszczegolnych-wojewodztw-w-latach-2010-2013>; own estimates.