Men and Women Wage Differences in Spain and Poland

ALEKSANDRA MATUSZEWSKA-JANICA1

1Assistant Professor, SGGW in Warsaw, Poland, e-mail: aleksandra_matuszewska@sggw.pl

ARTICLE INFO

Received December 03, 2017
Revised from December 22, 2017
Accepted February 21 2018
Available online March 15, 2018

JEL classification:
J31; C20.

DOI: 10.14254/1800-5845/2018.14-1.3

Keywords:
Labour market, men and women wages differences, outliers, Oaxaca-Blinder decomposition,

ABSTRACT

Men and women wage differences is a widely discuss topic in the literature. Some authors point the fact that size of gender wage gap (GPG) is not the same across wage distribution. GPG ratio is accelerating at the top of it. Thus, the main goal of presented study is to analyse impact of outliers (top earners) on the values of GPG ratio and results of its decomposition. In addition we compare outcomes obtained for Spain and Poland. Elimination outliers from the sample will reduce values of men and women wage gap ratios not only in unadjusted form but in adjusted form as well. Study is based upon the Eurostat’s Structure of Earnings Survey (SES) individual data in respect of 2014. In the paper we discuss results of Oaxaca-Blinder decomposition (in extension proposed by Oaxaca and Ransom) obtained for Spain and Poland. Obtained results indicated two points above all. Firstly, although unadjusted GPG ratios for Spain and Poland differ significantly, the adjusted GPG ratios are at the same level (about 15%). Such situation shows the real men and women wage differences are at the same level in both countries. This is an additional indication that women situation on Polish labour market in similar to the Spanish one. Secondly, after elimination of outliers the values of GPG measures (in adjusted and unadjusted form) decreased, as was expected. These falls came to approximately 3 p.p. It can be considered as significant change.

INTRODUCTION

Men and women wage differences is an issue that has been debated on many levels: social, economic and scientific. Increasing interest of this topic is reflected in growing number of surveys and publication. They indicated wage gap is influenced by many factors. The substantial variety of such factors have caused that GPG rate is characterized by considerable diversity in relation to region among others (e.g. Arulampalam et al., 2007; Matuszewska-Janica & and Hozer-Kočmiew 2015; Majchrowska and Strawiński 2016). Eurostat's statistics show that Poland belongs to the groups of countries with the lowest differences between men and women average wages (7.7% in 2014). In turn, in Spain this ratio was nearly two times bigger (14.9%). It is necessary mentioned, that in 90’s and early years of 21st century Spain has been taking into account as a benchmark for forecasting of economic development path in Poland, also connecting to the labour market.
Result of many studies indicated that the gender wage gap increases with the pay scale (in particular carried out for Spanish data, e.g. Garcia et al. 2001, Dolado and Llorens 2004). This led to a question whether after exclusion of top earners wages the gender wage gap would be smaller. Analysis of influence of the outliers on the size of gender wage gap is the main goal of the paper. In presented study we employed data from Eurostat's Structure of Earnings Survey (SES) conducted in 2014. As an analytical tool, one-equation econometric models and Oaxaca-Blinder decomposition method are used. Reduction of observation in samples will also allow to examine the sensitivity of parametric methods results, that are widely used in studies on men and women wages.

1. MEN AND WOMEN WAGE DISPARITIES

Men and women wage disparities (gender wage gap, GPG) are widely discuss topic in the scientific literature (see Weichselbaumer and Winter-Ebmer 2005, Blau 2012, Blau and Kahn 2017 among others). Such disparities are explained based mainly on two theories: human capital theory (see Schultz, 1961 and 1971; Becker, 1962 and 1964; Mincer, 1974) and discrimination theory (see Becker, 1957). It should be also mentioned, preferences theory (see Charles and Grusky, 2004; Hakim, 2004 and 2006; Jacobs and Gerson, 2004) in this research area is growing more popular in recent years. The size of men and women wage gap can vary widely across different employees groups. Factors that have a great importance for this diversity we can divide into three groups.

- Individual characteristics of employee (e.g. age, sex, job seniority, the type and level of education, occupation, full-or part-time job, type of job contract, family social and economic status, preferences).
- Enterprise characteristics (type of industry, public or private sector, size of the enterprise, activity of the trade unions among others).
- Characteristics of the environment (for example: economic situation in the region or in the country, structure of the labour market, family policy).

As was earlier mentioned GPG ratio also vary considerably across regions. It was pointed in articles (Weichselbaumer and Winter-Ebmer, 2005; Gannon et all., 2007; Arulampalam et al. 2007; Hozer-Kocmiel and Matuszewska-Janica, 2015; and (Majchrowska and Strawinski, 2016) among others. Analysis provides for European States by (Hozer-Koćmiel and Matuszewska-Janica, 2015) indicated for such a large diversity of GPG ratios in EU states (taking into consideration economic sectors) that it was not possible partitioning them in reasonable way. This bears witness to the specificity of each country within the EU area in terms of remunerating employed men and women. In addition, analysis presented in (Majchrowska and Strawinski, 2016) confirm such diversity at the NUTS 2 level.

The size of gender wage gap in Poland was investigated by Śliwicki and Ryczkowski (2014), Goraus and Tyrowicz (2014), Goraus et al. (2017) among others. Despite the use of different data sets in these analyses authors achieved a certain similarity of results. Adjusted wage gap which takes into account differences in characteristics of both employed men and women groups is much greater that unadjusted (“raw”) wage gap. Situation in Spain was analysed e.g. by Garcia et al. (2001), Dolado and Llorens (2004), De la Rica et al. (2008), De La Rica et al. (2010). These results indicated also that in Spain adjusted GPG is greater than raw GPG. Additionally they present

---

3 Unadjusted gender wage gap measures income inequality and represents the difference between average earnings of male paid employees and of female paid employees usually as a percentage of average gross hourly earnings of male paid employees.
evidence that Spanish gender wage gap increases with the pay scale. Similar results for more European states are presented by Arulampalam et al. (2007). Authors pay attention also to differences across wage distribution and indicated that gender wage gap in European countries is increasing throughout the conditional wage distribution and accelerating at the top. This was interpreted as evidence of existing glass ceiling. It means, that the greatest wage differences has been noted for the top earners. Therefore, applying traditional measures could overestimated GPG ratio.

Traditional wage gap measure like unadjusted GPG, called also raw GPG, does not take into account differences in two employees group such as men and women. Therefore shall be decompose into two components: explained and unexplained. The most popular method is the Oaxaca-Blinder decomposition (Oaxaca, 1973, Blinder, 1973) and its expansions (see e.g. Neumark 1988, Oaxaca and Ransom 1994). Other decomposition methods have also been reported by Ņopo (2008) or Fortin et al. (2011) among others.

2. DATA AND METHODOLOGY

Presented study is based upon the EU Structure of Earnings Survey (SES) individual data in respect of 2014. The SES survey refer to the enterprises with at least 10 employees. It is necessary to mention, that information included in the SES databases are from the enterprises' registers. The presented analysis is divided into two stages. In the first step parameters of linear econometric models are estimated using the generalized least squares method with heteroskedasticity correction. In the second step, we decompose the wage gap applying the Oaxaca-Blinder method (see Oaxaca 1973 and Blinder 1973) in expansion proposed by Oaxaca and Ransom (1994). We analyse tree types of samples. The first type of sample involves all observations. The second and third type of samples are reduced by outliers.

Based on each of the samples we estimate the parameters of three types of models: with information about all individuals (model type \( p \), see formula 1); with information only about female individuals (model type \( f \), see formula 2) and with information only about male individuals (model type \( m \), see formula 3). The explanatory variables used in the models are listed in Table 2. The equations used in the presented analysis are as follows:

\[
\ln W_{pi} = \beta_{p0} + \sum_{j=1}^{k} \beta_j X_{jpi} + \gamma_p \ln Y_{pi} + \varepsilon_{pi} \tag{1}
\]

\[
\ln W_{fi} = \beta_{f0} + \sum_{j=1}^{k} \beta_j X_{jfi} + \gamma_f \ln Y_{fi} + \varepsilon_{fi} \tag{2}
\]

\[
\ln W_{mi} = \beta_{m0} + \sum_{j=1}^{k} \beta_j X_{jmi} + \gamma_m \ln Y_{mi} + \varepsilon_{mi} \tag{3}
\]

where: \( W_{si} \) - gross hourly wages of \( i \)-th individual from \( s \)-th sample \((s = p, f, m)\); \( X_{jsi} \) - \( j \)-th dummy explanatory variable for \( i \)-th individual from \( s \)-th sample; \( Y_{si} \) - quantitative explanatory variable for \( i \)-th individual from \( s \)-th sample; \( p \) - whole sample, \( m \) - (males) - privileged group; \( f \) - (females) - non-privileged group; \( f + m = p \).

Oaxaca and Ransom (1994 - See also Jann, 2008) proposed wage differences decomposition in following way:

\[
\ln \overline{W}_m - \ln \overline{W}_f = (\overline{X}_m - \overline{X}_f)' \hat{\beta}^* + \overline{X}_m' (\hat{\beta}_m - \hat{\beta}^*) + \overline{X}_f' (\hat{\beta}_f - \hat{\beta}^*) \tag{4}
\]
where: \((\bar{X}_m - \bar{X}_f)\hat{\beta}^*\) - component explained by differences in employees groups characteristics ("quantity effect"); \(\bar{X}_m^T(\hat{\beta}_m - \hat{\beta}^*) + \bar{X}_f^T(\hat{\beta}^* - \hat{\beta}_f)\) - unexplained component; called adjusted wage gap; \(\bar{X}_m^T(\hat{\beta}_m - \hat{\beta}^*)X_m(\hat{\beta}_m - \hat{\beta}^*)\) - bonus for belonging to a privileged group ("positive discrimination" of privileged group); \(\bar{X}_f^T(\hat{\beta}^* - \hat{\beta}_f)\) - loss for belonging to a non-privileged group ("negative discrimination" of non-privileged group if the sign is negative); \(\hat{\beta}^* = \hat{\beta}_p\) - "nondiscrimatory" coefficient vector, estimated coefficients from regression 1 (see Neumark 1988); \(\hat{\beta}_f\), \(\hat{\beta}_m\) vectors of parameters are estimated based on (2) and (3) regression respectively; \(\ln \bar{W}_f\) - average wages of women - arithmetic mean of the theoretical values from the model (2); \(\ln \bar{W}_m\) - average wages of men - arithmetic mean of the theoretical values from the model (3); \(\ln \bar{W}_m - \ln \bar{W}_f\) - unadjusted men and women wage gap, called "raw" wage gap (GPG);

**Table 1. List of explanatory variables**

<table>
<thead>
<tr>
<th>No.</th>
<th>Variables (group of variables)</th>
<th>Variants</th>
<th>Reference variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sex*</td>
<td>2</td>
<td>female</td>
</tr>
<tr>
<td>2</td>
<td>Region</td>
<td>6 (PL) / 7 (ES)</td>
<td>PL – Central region ES – Community of Madrid</td>
</tr>
<tr>
<td>3</td>
<td>Occupation (major groups ISCO-08)</td>
<td>6</td>
<td>groups 6-9; (skilled manual workers and elementary occupations)</td>
</tr>
<tr>
<td>4</td>
<td>Size of the enterprise</td>
<td>3 (PL) / 4 (ES)</td>
<td>10-49 employees</td>
</tr>
<tr>
<td>5</td>
<td>Collective pay agreement</td>
<td>3 (PL) / 5 (ES)</td>
<td>no collective agreement exists</td>
</tr>
<tr>
<td>6</td>
<td>Form of economic and financial control of the enterprise</td>
<td>2</td>
<td>private control (private ownership is more than 50%)</td>
</tr>
<tr>
<td>7</td>
<td>Age group</td>
<td>5</td>
<td>aged 20-29</td>
</tr>
<tr>
<td>8</td>
<td>Highest educational level attained</td>
<td>4</td>
<td>basic education (G1)</td>
</tr>
<tr>
<td>9</td>
<td>Contractual working time (full-time or part-time)</td>
<td>2</td>
<td>full-time employees</td>
</tr>
<tr>
<td>10</td>
<td>Type of employment contract</td>
<td>3</td>
<td>indefinite duration</td>
</tr>
<tr>
<td>11</td>
<td>Length of service in the enterprise (in years)</td>
<td>x</td>
<td>quantity variable</td>
</tr>
</tbody>
</table>

Source: own elaboration * - only in models estimated on whole sample (p).

As was earlier mentioned, the objective of the analysis is also to assess the influence of the top wages (outliers) one the results of Oaxaca-Blinder decomposition. To that end the outliers should be recognize. The most popular tool applied for identifying outliers is box and whisker plot (called also box plot with fences). For this purpose are quartiles and interquartile range used (see Aczel and Sounderpandian 2007). Interquartile range (IQR) is defined as the difference between the lower (\(Q_L\)) and upper (\(Q_U\)) quartiles.4

\[ IQR = Q_U - Q_L \]

4 As \(Q_i\) is designated the first quartile and as \(Q_u\) - the third quartile.
IQR is used to indicate one the box plot point called fences that are needed for identifying outliers. The values of fences are calculated as follows:

- \(LOF=Q_L - 3 \cdot IQR\), lower outer fence and \(LIF=Q_L - 1.5 \cdot IQR\), lower inner fence;
- \(UIF=Q_U + 1.5 \cdot IQR\), upper inner fence and \(UOF=Q_U + 3 \cdot IQR\), upper outer fence.

Observation is considered as a mild outlier if it is in the range \(LOF\) to \(LIF\) or \(UIF\) to \(UOF\). In turn, value of observation lower than \(LOF\) or greater than \(UOF\) designated that we could consider it an extreme outlier. In presented study we have considered upper outliers, greater than \(UIF\) or \(UOF\) respectively. Values of individual's wages greater smaller or equal \(UIF (W_i \geq Q_U + 1.5 \cdot IQR)\) we call suspected outliers. In turn, values of individual's wages greater smaller or equal \(UOF (W_i \geq Q_U + 3 \cdot IQR)\) we call outliers.

In the Table 2 are presented numbers of observations in selected samples. It is worth to mentioned that percentage of outliers is close to 2% and the share of suspected outliers exceed a little bit 5% of whole sample for both countries. In addition, we can observe that in the groups of top earners men are overrepresented.

**Table 2. Selected characteristics of analysed samples**

<table>
<thead>
<tr>
<th>State</th>
<th>Type of sample</th>
<th>Total observation</th>
<th>Male</th>
<th>Female</th>
<th>%Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poland</td>
<td>Whole sample</td>
<td>722697</td>
<td>359119</td>
<td>363578</td>
<td>50.3%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (1); (W_i &lt; Q_i + 3 \cdot IQR)</td>
<td>710409</td>
<td>350281</td>
<td>360128</td>
<td>50.7%</td>
</tr>
<tr>
<td></td>
<td>Outliers (1; (W_i \geq Q_i + 3 \cdot IQR)</td>
<td>12288</td>
<td>8838</td>
<td>3450</td>
<td>28.1%</td>
</tr>
<tr>
<td></td>
<td>% of Outliers (1)</td>
<td>1.7%</td>
<td>2.5%</td>
<td>0.9%</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (2); (W_i &lt; Q_i + 1.5 \cdot IQR)</td>
<td>684465</td>
<td>336036</td>
<td>348429</td>
<td>50.9%</td>
</tr>
<tr>
<td></td>
<td>Suspected outliers (2); (W_i \geq Q_i + 1.5 \cdot IQR)</td>
<td>38232</td>
<td>23083</td>
<td>15149</td>
<td>39.6%</td>
</tr>
<tr>
<td></td>
<td>% of Suspected Outliers (2)</td>
<td>5.3%</td>
<td>6.4%</td>
<td>4.2%</td>
<td>x</td>
</tr>
<tr>
<td>Spain</td>
<td>Whole sample</td>
<td>209145</td>
<td>119765</td>
<td>89380</td>
<td>42.7%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (1); (W_i = Q_i + 3 \cdot IQR)</td>
<td>205469</td>
<td>116857</td>
<td>88612</td>
<td>43.1%</td>
</tr>
<tr>
<td></td>
<td>% of Outliers (1)</td>
<td>1.8%</td>
<td>2.4%</td>
<td>0.9%</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Outliers (1; (W_i \geq Q_i + 3 \cdot IQR)</td>
<td>3676</td>
<td>2908</td>
<td>768</td>
<td>20.9%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (2); (W_i &lt; Q_i + 1.5 \cdot IQR)</td>
<td>197455</td>
<td>111185</td>
<td>86270</td>
<td>43.7%</td>
</tr>
<tr>
<td></td>
<td>Suspected outliers (2); (W_i \geq Q_i + 1.5 \cdot IQR)</td>
<td>11690</td>
<td>8580</td>
<td>3110</td>
<td>26.6%</td>
</tr>
<tr>
<td></td>
<td>% of Suspected Outliers (2)</td>
<td>5.6%</td>
<td>7.2%</td>
<td>3.5%</td>
<td>x</td>
</tr>
</tbody>
</table>

Source: own calculation.

**3. EMPIRICAL RESULTS**

Obtained results indicated that raw GPG ratio (unadjusted GPG) is generally lower in Poland. Given the whole sample, estimated unadjusted GPG are 8.75% or 17.8% in Poland and Spain respectively. Thus the differences between men and women average hourly remuneration are in
Poland much more lower than in Spain. But analysed disparities are almost the same bearing in mind adjusted values of GPG ratios (see Table 3, unexplained component), 15.23% in Poland and 15.74% in Spain. A considerable difference between the values of explained components for the two countries gets particular attention. Explained component for Spain indicates that only 2.06 p.p. of GPG from 17.8% is explained by groups differences in the predictors (explanatory variables included to the model, see Table 1). In case of Poland “quantity effect” has negative sign. It means, taking into account analysed characteristics (variables) women might earn on average 6.48% more than men.

Table 3. Results of Oaxaca-Blinder decomposiotion

<table>
<thead>
<tr>
<th>State</th>
<th>Type of sample</th>
<th>“raw” GPG (4)+(5)</th>
<th>Explained component</th>
<th>Unexplained Component (6)+(7)</th>
<th>Privileged group - bonus</th>
<th>Non-privil. group - loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
<tr>
<td>Poland</td>
<td>Whole sample</td>
<td>8.75%</td>
<td>-6.48%</td>
<td>15.23%</td>
<td>15.42%</td>
<td>-0.20%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (1)</td>
<td>6.26%</td>
<td>-7.26%</td>
<td>13.52%</td>
<td>14.13%</td>
<td>-0.61%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (2)</td>
<td>5.41%</td>
<td>-6.77%</td>
<td>12.18%</td>
<td>12.85%</td>
<td>-0.67%</td>
</tr>
<tr>
<td>Spain</td>
<td>Whole sample</td>
<td>17.80%</td>
<td>2.06%</td>
<td>15.74%</td>
<td>15.61%</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (1)</td>
<td>15.71%</td>
<td>1.46%</td>
<td>14.25%</td>
<td>14.27%</td>
<td>-0.02%</td>
</tr>
<tr>
<td></td>
<td>Truncated sample (2)</td>
<td>13.90%</td>
<td>1.14%</td>
<td>12.76%</td>
<td>12.74%</td>
<td>0.02%</td>
</tr>
</tbody>
</table>

Source: Own elaboration. * Privileged group - males, Non-privileged group - females.

After eliminating outliers the unadjusted and adjusted GPG ratios declined as was expected. In case of Poland, the relative difference in average wages of men and women (“raw” GPG) decreased by 2.49 p.p. and adjusted GPG declined by 1.71 p.p. when the outliers (UOF) were not taken into account. I turn both rates decline by 3.34 p.p. and 3.05 p.p. respectively after exclusion from the sample the suspected outliers (UIF). In case of Spain, in the first truncate sample (with eliminated UOF) we observe GPG ratio lower by 2.09 p.p. and unadjusted component reduced by 1.49 p.p. However in the second truncated sample (without UIF) raw GPG decreased by 3.9 p.p. and adjusted GPG declined by 2.98 p.p. Expected components calculated for both countries are also decreased. Notwithstanding the country and sample, the differences are not greater than 1 p.p. explained part of "raw" GPG indicates that taking into account the characteristics from the table 1 relative differences between men and women average wages increase in favor of women in reduced samples for Poland. Then women in limited samples have on average better values of such characteristics. In turn, in case of Spain values of explained part suggest that the variables from the model have slightly less importance in explanation of "raw" GPG in reduced samples.

It is worth to mention, that values of loss (with negative sign) or bonus (with positive) of non-privileged group are not greater that 0.7%. That means the size of adjusted GPG is caused by bonus for privileged group. In other words, individuals from privileged group potentially have the characteristics (not included in the model) which are better assessed by employers.
SUMMARY

Presented analysis indicated two issues. Firstly, when account are also taken differences in characteristics of both privilege and non-privilege groups, the adjusted differences in men and women wages in Poland and Spain remain around similar level (15.23% and 15.74% respectively). Secondly, after elimination of outliers (individuals with the highest wages) the GPG measures (adjusted and unadjusted form) decreased, as was expected.

These falls came to approximately 3 p.p. This can be considered as significant change. This also bears witness to conclusion that parametric models could overestimating GPG ratios when in the samples are included outliers.

ACKNOWLEDGEMENTS

Research is conducted in the frame of project financed by National Science Centre grant No. 2015/17/B/HS4/00930. The SES database was obtained under the Eurostat's grant RPP 83/2017-LFS-SES.

REFERENCES


