Gender Wage Gap in Poland – Can It Be Explained by Differences in Observable Characteristics?

WORK IN PROGRESS

Please do not cite or quote without authors' permission

Karolina Goraus[1]

Abstract

This article concentrates on the problem of gender wage gap in Poland. The raw gap over the period 1995-2011 amounts to app. 10%. However, accounting for the differences in endowments the actual wage gap grows to as much as 20%. We implemented both parametric and non-parametric decomposition techniques to test the reliability of this result and employed a number of robustness checks. While there is some heterogeneity between groups with different educational attainments, skills and of different age - the adjusted wage gap roughly doubles the raw wage gap. Despite covering already 17 years of data, we were not able to identify any clear decreasing trend in gender discrimination in Poland.

Keywords: Wage gap, discrimination, decomposition, Oaxaca-Blinder, Ñopo, non-parametric estimation

June 2012

[1] Warsaw University, Faculty of Economic Sciences

Table of Contents

Introduction

Most societies declare preference for equality, if not equality of outcomes, then surely equality of opportunities. Presence of wage gaps is evidence to discriminatory practices and ineffectiveness of anti-discriminatory policies. However, individuals' compensations are likely to exhibit considerable differentiation due to large differences in the underlying characteristics that are relevant for the labor market, i.e. education or occupational experience. Thus, the real challenge lies in providing reliable measures of wage gaps.

Especially gender differences in the labor market, and the gender wage gap, have been gaining considerable attention in the last decades. In addition to numerous theoretical and empirical papers emerging in this field, significant development of statistical tools created to decompose gender wage differentials have been observed. There are two main streams of these approaches, the traditional parametric and the relatively newer – non-parametric decompositions. To the best of our knowledge there is no empirical research using non-parametric methods to Poland, while research using parametric methods is scarce. We aim to fill this gap by providing both parametric and non-parametric estimators of the gender gap in hourly wages, accounting for the entire post-transition period 1995-2011.

Raw gender wage gap in analyzed period amounts to around 10%. By many standards this is not much, but accounting for the fact that females are on average better educated, it should attract research attention. The main research question of this article consists of the following: is the gender wage gap in Poland explainable by differences in observable characteristics between females and males? In order to address this question we are going to employ two methods. Ñopo (2008) considers the gender variable as treatment and uses matching comparisons to measure the impact of "treatment" on the outcome variable, which is most typically wage. In comparison Blinder-Oaxaca (1973) decomposition employs the estimators of the wage equation parameters.

The study is divided into five parts. First section of this article contains the literature review. It is briefly described how decomposition methods were developing over time, and which of them have been applied to measure gender wage gap in Poland. Second part contains description of data and research methods.

Next two parts of this article contain empirical analysis. Third part aims at examining if the raw gender wage gap, understood as difference in average wages of males and females, is observed in Poland over the analyzed period. What is more it explores if differences in characteristics between females and males are observed among Polish employees.

Fourth part of this work is a core section where decomposition techniques are applied to measure explained and unexplained components of gender wage gap in Poland and their evolution over time. Firstly, selected non-parametric approach is applied. Decompositions based on different sets of characteristics are performed on the pooled sample. Then, two chosen specifications are used to analyze gender wage gap in each quarter of 1995-2011 separately. In order to assure robustness of obtained results, gender wage gap in Poland is then decomposed with typically used parametric method – Blinder-Oaxaca decomposition. After that, sensitivity analysis is performed. The last section concludes.

The findings suggest that the actual gender wage gap in Poland is much larger than the difference in average female and male compensations, as reported by the Central Statistical Office. We show that if gender wage gap is adjusted for the differences in observable characteristics it grows to as much as 20%, whereas this differential does not seem to exhibit cyclical properties. Moreover, we find no evidence that the adjusted gender wage gap decreases over time.

1 Literature Review

The issues of gender differences in the labor market and gender discrimination have been gaining considerable attention in the last decades. They have been a significant area of concern for theoretical and empirical research in economics, as well as were often a topic in social and political discussions, or even were important elements of election campaigns.

However, as pointed out by Grajek (2003) Poland had a significant delay in having their academic, business, and political elites concentrated on this issue. Polish gender wage gap has been analyzed mostly in the context of transition period as performed by Grajek (2003) or Adamchik and Bedi (2003). The letter authors underline that the relative economic welfare of women is one of the measures of nation's well-being and they doubted if the economic position of females in Poland has improved along with the positive economic performance of the country. Adamchik and Bedi (2003) also pointed out that among several indicators – such as income, employment, or social benefits, wages are probably the most important determinant of economic well-being and personal success, and they should be analyzed to assess relative situation of females.

Basic assessment of gender wage differentials is done by measuring the difference in average wages between males and females. However, this approach has limited explanatory power as it does not account for differences in characteristics between females and males. When explaining gender differences in earnings, some people may claim that it is due to discrimination, and others that it simply reflects gender differences in some observable characteristics of the individuals that are determinants of wages (Ñopo, 2008).

The question about the most important explanations accounting for pay differences between men and women are typically answered using decomposition methods. This field of economics is not only deserved to explore gender wage gap. They can be used to control for observed characteristics in any measure for which it is expected to find some sort of explained and unexplained components. But it is in labor economics that decomposition techniques have been used the most extensively (Fortin, Lemieux, and Firpo, 2010).

Seminal papers by Oaxaca (1973) and Blinder (1973) are among most cited in labor economics, and the Blinder-Oaxaca decomposition is now a standard tool in applied economics. This technique requires the linear regression estimation of earnings equations for both females and males. Based on these earning equations, the counterfactual situation, that answers the question about the male (female) wage if the compensation scheme for his (her) individual characteristics aligned with the compensation schemes for females (males), can be generated. After some algebraic manipulations the difference in average wages between males and females is decomposed into two additive components: one attributable to differences in average characteristics of the individuals, and the other – to differences in the rewards that these characteristics have. The letter component is considered to contain the effects of both unobservable gender differences in characteristics that the market rewards and discrimination in the labor market.

Oaxaca (1973) was aiming at estimating the size of actual discrimination in the gender wage gap in the United States according to data for 1967 from Survey of Economic Opportunity. The study took into consideration hourly wage of individuals of age over sixteen, living and employed in urban areas and reporting their race as White or Black. Oaxaca also accounted for human capital characteristics and environmental conditions that impact the distribution of workers across different sectors, positions and occupations. As a result the raw wage difference has been proved to be much larger than adjusted wage gap (understood as part of the raw gap unexplained by differences in characteristics).

In a large number of methodological papers attempts to refine the Blinder-Oaxaca decomposition have been observed. One direction of developments was connected to the assumption in standard Blinder-Oaxaca decomposition that the male wage structure prevails in the absence of discrimination. Thus other non-discriminatory wage structures have also been observed in the literature. Decompositions based on different assumed reference wage structures are found in Cotton (1988), Neumark (1988), Oaxaca and Ransom (1994), and Reimers (1983). In Oaxaca and Ransom (1994) vector of coefficients in non-discriminatory wage structure is defined as weighted average of coefficient vectors in male and female wage equations.

Neumark (1988) also suggested generalized method where, under certain conditions, the appropriate non-discriminatory wage structure can be obtained by estimating a regression over the pooled male-female sample. Then the observed wage differential can be decomposed into three components. The first one is attributed to differences in characteristics between males and females. The second component is attributed to differences between estimated parameters of wage regression for males and the pooled wage regression (this component is called a male advantage or male favoritism component). The third part of raw wage gap differentials is attributed to differences between the estimated parameters of the pooled wage equation and the female wage equation (called female disadvantage or pure discrimination component).

In the study on wage gap in Poland over transition Adamchik and Bedi (2003) have used both the standard Blinder-Oaxaca method and its modified version as in Neumark (1998). According to their findings, the percentage of the wage gap that is explained by differences in observed characteristics varies across the two methods, but in both it is quite limited over the analyzed period 1993-1997. What is more, for each year the explained portion of the gap is considerably higher for modified version, than for standard Blinder-Oaxaca decomposition.

Contribution of Adamchik and Bedi (2003) is also important, as it discusses the characteristics that could be used in wage equations. The basic set of regressors in their paper included conventional human capital characteristics (e.g. education or experience), personal characteristics (e.g. marital status), and regional labor market conditions, like information if area is urban or rural. In the second specification of the set of characteristics they have also included job characteristics, like information on type of industry, occupation, branch of economy (high-paying or low-paying), or firm size.

What is more, the authors discussed possible criticism of inclusion of job characteristics in an earning equation. For instance, a number of job-related characteristics might be endogenous on the labor market. It is not clear if differences in job characteristics for males and females reflect employment discrimination, or different tastes and preferences, or both. At the end they have followed the convention and treated job characteristics as factors explaining the wage differential between females and males, rather than manifestation of employment discrimination. This approach will be followed by the author of this work, as job characteristics will be also considered as explanatory variables in further empirical analysis.

Blinder-Oaxaca decomposition is very useful in identifying causes of racial or gender differences not only in wages, but also in educational, labor market, and other outcomes. The technique is relatively easy to apply and only requires coefficients estimates from linear regressions for the chosen outcome variable and sample means of the explanatory variables used in the regressions. However, if the outcome variable is binary, such as employment, collage attendance, or teenage pregnancy, the problem arises. Coefficients from a logit or probit model cannot be used directly in the standard Blinder-Oaxaca decomposition equation (Fairlie, 2003).

A solution to the problem described above, was constructed by Fairlie (2003), who suggested a method of decomposition, in which estimates from logit or probit models were used. This relatively simple method was described in the analysis of the causes of the black/white gap in self-employment rates. Bauer and Sinning (2008) have generalized the Blinder-Oaxaca decomposition to other non-linear models and demonstrated how it can be applied to models with discrete and limited dependent variables.

Another limitation of standard Blinder-Oaxaca decomposition is that it is only informative about the average unexplained difference in wages, not about the distribution of these unexplained differences. Thus numerous papers aimed at expanding it to the case of distributional parameters besides the mean. Among those distributional methods there is broadly used decomposition developed by Juhn, Murphy, and Pierce (1991, 1993), quintile regressions methods like in Machado and Mata (2005), inverse propensity reweighing (DiNardo, Fortin, and Lemieux, 1996) or such advanced techniques as recentered influence function regressions (Firpo, Fortin, and Lemieux, 2007). In this work only one of those methods, namely Juhn, Murphy and Pierce decomposition will be described with more detail.

Juhn, Murphy and Pierce (1991, 1993) extended the Blinder-Oaxaca technique to allow for decompositions at points in the earnings distribution other than the mean. According to this methodology wage differential is decomposed into four parts. The first three terms are interpreted as in generalized Blinder-Oaxaca decomposition (Oaxaca and Ransom, 1994). Thus one may recognize the explained component (due to differences in characteristics), the male advantage (attributed to differences between estimated parameters of wage regression for males and the reference wage structure), and the female disadvantage (due to differences between non-discriminatory wage structure and females' wage structure). The fourth term represents differences in the quantities and prices of unobservable characteristics resulting from changes in the distribution of the residual from the wage regression. When considering the decomposition at the mean, the fourth term takes on a zero value – and in that case the Juhn, Murphy, Pierce decomposition reduces to the generalized Blinder-Oaxaca form.

Grajek (2003) applied Juhn, Murphy, and Pierce decomposition technique to analyze data on Polish employees from Household Budget Survey for the period 1987 – 1996, and he also found that explained component is relatively small and rises slowly over the analyzed period.

Another problem associated with the Blinder-Oaxaca decomposition is the misspecification caused by differences in the supports of the distribution of individual characteristics for females and males. It was pointed out by Rubin (1977) that there are combinations of characteristics for which it is possible to find males but not females in the society, and vice versa. With such distribution of characteristics one cannot compare wages across genders. The problem with comparability is enhanced when job-related variables are included in the explanation of gender gap, as females tend to concentrate in certain occupations that demand particular abilities e.g. soft skills or empathy, while males concentrate more often in risky or managerial occupations.

Ñopo (2008) adapted the tool of the program evaluation literature, matching, to construct a non-parametric alternative to Blinder-Oaxaca decomposition method and fix the problem of differences in the supports of distribution of characteristics between females and males.

Matching comparisons techniques serve to find matched samples with “similar” observable features except for one particular characteristic, the “treatment”, which is used to group observations into two sets, the treated and the control group. After controlling for these observed characteristics it is possible to measure the impact of treatment alone. After the introduction of propensity scores in experimental design (Rosenbaum and Rubin, 1983) matching techniques started to be useful tool in estimation of causal effects in economics. For example Pratap and Quintin (2002) used propensity score matching to measure wage differences between the formal and informal sectors in Argentina.

Ñopo (2008) went a step further and considered the gender variable as a treatment and used matching to select sub-samples of males and females in such a way, that there are no differences in observable characteristics between “matched” males and “matched” females. It should be mentioned that the assumption of Rosenbaum and Rubin (1983) about the “ignorability of treatment” required for propensity score matching is not likely to be satisfied in case the gender is perceived as “treatment”. Thus matching individuals in Nõpo is based on characteristics, not propensity scores. After grouping both females and males into “matched” and “unmatched” sub-samples Nõpo was able to develop decomposition that accounts for differences in the supports[2].

The traditional parametric technique of decomposing gender wage gap developed by Oaxaca (1973) and Blinder (1973), as well as its non-parametric alternative developed by Nõpo (2008) are of special interest for this work and are described in more detail in the following section.

[2] Expression “differences in the supports” in this work stands for “differences in the supports of the distribution of characteristics for females and males”

2 Data and research method

In the following section of this work data used in empirical analysis is described and most important information on variables representing characteristics is presented. Then, two methods of decomposition are introduced.

2.1 Data

The empirical part of this paper relies on the data on the level of occupational activity of population by demographic and social features. The data set comes from the Labor Force Survey performed by Central Statistical Office in Poland and contains quarterly data from 1995q1 to 2011q4. It should be also mentioned that in the second and third quarter of 1999 the survey was not conducted and data for those quarters is just a replica of data for the first quarter of year 1999.

Thanks to the relatively big data set it is possible to conduct research on gender wage gap and differences in characteristics in each of 68 periods separately and analyze their evolution over time. Additionally the pooled data set was created, as sometimes presenting the results for each of the periods would not be transparent. This data set contains 690414 observations. In the pooled data set the wages were adjusted with the use of the wage deflator, and are presented in PLN, constant prices of 1995. Thus every figure in this paper that presents results from the pooled data set contains the information on real wages. While presenting results for each period separately it will be additionally indicated if wages are in nominal or real terms.

As the purpose of this study is analysis of gender wage gap, persons that are self-employed, unemployed, or inactive have been removed from the data set. Moreover, for homogenizing purposes, workers of the mining sector and armed forces have been also removed from the data set. Share of males in the final data set is 52.5%.

The monthly wages were divided by the hours of work, as the hourly wage is typically analyzed in studies about gender wage gaps. Analyzing hourly wages omits the problem of different working regimes (part time, full time), and accounts for empirical regularity that females work less than males. In such adjusted data set, the raw gender wage gap that is understood as the gap in hourly wages might be explored.

Labor Force Survey contains limited set of variables. Specifically we dispose of information on hourly wage, age, education, marital status, occupation category, branch of economy, or tenure with current employer. Dummies indicating if region is rural or urban, if voivodeship is the richest Mazowieckie or other, if an individual is working in public or private sector, and if this sector is formal or informal, are also relevant for this work. Finally the information on overall tenure and size of the firm could be important for further analysis, but the data on these variables is not available for the whole analyzed period (there is no information on overall tenure between 1997-2005, and information on the size of the firm has not yet been coded for 2010 and 2011).

Table 1 contains descriptive statistics of mentioned variables obtained from pooled sample containing all quarters 1995-2011. In case of overall tenure and size of the firm descriptive statistics were obtained from adjusted datasets that does not contain quarters for which information on respective variable is missing.

Table 1: Variables at disposal

Continous variables	**Number of observations**	**Mean**	**Standard deviation**
Hourly wage	690 414	13.15	8.8
Age	690 414	38.89	10.64
Tenure with current employer	690 414	10.24	9.47
Overall tenure	365 141	17.8	11.04
Categorical variables	**Number of observations**	**Percent**	**Cumulative**
Education levels	***690 414***	***100***	
Tertiary education	112 697	16.32	16.32
High school	82 203	11.91	28.23
High school vocational	185 836	26.92	55.15
Vocational	240 666	34.86	90
Elementary	69 012	10	100
Marital status	***690 141***	***100***	
Single	144 305	20.9	20.9
Married	505 167	73.17	94.07
Widowed	15 240	2.21	96.28
Divorced/separated	25 702	3.72	100

Occupation category	***690 414***	***100***	
Very high-skilled occupation	118 075	17.1	17.1
High-skilled occupation	245 181	35.51	52.61
Middle-skilled occupation	249 663	36.16	88.78
Low-skilled occupation	77 495	11.22	100
Branch of economy	***690 414***	***100***	
Agriculture	9 022	1.31	1.31
Industry	131 495	19.05	20.35
Construction	118 117	17.11	37.46
Market services	221 563	32.09	69.55
Non-market services	210 217	30.45	100
Type of area	***690 414***	***100***	
Rural	431 185	62.45	62.45
Urban	259 229	37.55	100
Region	***690 414***	***100***	
Mazowieckie	69 692	10.09	10.09
Other	620 722	89.91	100
Type of sector	***690 414***	***100***	
Public	344 533	49.9	49.9
Private	345 881	50.1	100
Formality	***690 414***	***100***	
Formality	683 246	98.96	98.96
Informal	7 150	1.04	100
Size of the firm	***615 908***	***100***	
Small enterprise	138 133	22.4	22.4
Medium or large enterprise	477 775	77.6	100

2.2 Method

Two methods of decomposing gender wage gap are used in the empirical part of this paper. One is a broadly used approach constructed by Blinder (1973) and Oaxaca (1973), while the other is a relatively new method developed by Nõpo (2008). Particular emphasis is placed on the letter decomposition method, which, to the best of our knowledge, has not yet been implemented to decompose gender wage gap in Poland over longer period of time[3].

[3] Nõpo et al. (2011) did include Poland in the research documenting gender disparities in earnings for broad set of countries. However for each country the decomposition was performed for one, most recent year for which all necessary data was available for the authors.

When developing his technique, Nõpo has been relating it to the Blinder- Oaxaca decomposition, which has been a traditional and broadly used tool to decompose wage gap between two groups in society. Nõpo's methodology that uses matching comparisons to explain gender wage differentials is a nonparametric alternative to Blinder-Oaxaca decomposition. Thus in order to present Nõpo's approach in the most understandable way it is worth providing the details of Blinder-Oaxaca decomposition in the first place. Then the idea and components of Nõpo's decomposition are described and related to the theory behind Blinder-Oaxaca decomposition and its elements.

2.2.1 Parametric approach

It was already mentioned that gender gaps in average earnings might be partially explained by gender differences in individuals' observable characteristics that the labor market rewards. Almost forty years ago, Blinder (1973) and Oaxaca (1973) constructed methodology to decompose differences in mean wages across two groups into explained and unexplained component.

This decomposition requires the linear regression estimation of earning equation for both groups, in our case, for females and males: $\bar{y}^F = \hat{\beta}^F \bar{x}^F$, and $\bar{y}^M = \hat{\beta}^M \bar{x}^M$, where $\bar{y}$ is an average wage of females or males, $\bar{x}$ is the vector of average characteristics in each group, and $\hat{\beta}$ is a vector of estimated coefficients of characteristics for females or males respectively. With such notations the raw gender wage gap can be expressed as $\bar{y}^M - \bar{y}^F = \hat{\beta}^M \bar{x}^M - \hat{\beta}^F \bar{x}^F$. After adding and subtracting the average counterfactual wage that male workers would have earned under the wage structure of females, $\hat{\beta}^F \bar{x}^M$, the expression becomes $\bar{y}^M - \bar{y}^F = \hat{\beta}^M \bar{x}^M - \hat{\beta}^F \bar{x}^M + \hat{\beta}^F \bar{x}^M - \hat{\beta}^F \bar{x}^F$.Then, after some algebraic manipulations it takes the form $\bar{y}^M - \bar{y}^F = \hat{\beta}^F(\bar{x}^M - \bar{x}^F) + (\hat{\beta}^M - \hat{\beta}^F)\bar{x}^M$.

Alternatively, the added and subtracted term might be the earning for female with average individual characteristics, in the case she is rewarded for her characteristics in the same way as the average male is rewarded, $\hat{\beta}^M \bar{x}^F$. Then the wage gap takes the form $\bar{y}^M - \bar{y}^F = \hat{\beta}^M(\bar{x}^M - \bar{x}^F) + (\hat{\beta}^M - \hat{\beta}^F)\bar{x}^F$. It is worth mentioning that this alternative form is especially important for the purpose of this work, as Nõpo's decomposition is related precisely to this one.

In both forms of decomposition the first components on the right-hand side, $\hat{\beta}^F(\bar{x}^M - \bar{x}^F)$ or $\hat{\beta}^M(\bar{x}^M - \bar{x}^F)$, are the part of the gap that is due to differences in average characteristics between males and females. In a broader context it is called the composition effect (Fortin, Lemieux, and Firpo, 2010). The second component, $(\hat{\beta}^M - \hat{\beta}^F)\bar{x}^M$ or $(\hat{\beta}^M - \hat{\beta}^F)\bar{x}^F$, is attributed to difference in average rewards to individuals' characteristics and is called the wage structure effects. The wage structure effect is also called "unexplained" part of the wage differentials, or the part due to "discrimination", although more precisely it should be perceived as the component containing the effects of both unobservable gender differences in characteristics and discrimination in the labor market.

The Blinder-Oaxaca decomposition is very easy to use in practice, as it is only necessary to plug in the sample means and the OLS estimates $\hat{\beta}$ in the presented formula. Various good implementations of this procedure are available in existing software packages, and one of them is used in the empirical part of this research.

2.2.2 Non-parametric approach

Despite the undeniable advantages of Oaxaca-Blinder decomposition, Ñopo (2008) pointed out its limitations and developed an improved method for decomposing the gender wage gap. Ñopo points out that there are combinations of individual characteristics for which it is possible to find males, but not females, in the labor force, while there are also combinations of characteristics for which it is possible to find females, but not males. With such combinations of characteristics one cannot compare wages across genders.

The traditional Blinder-Oaxaca decomposition fails to recognize these gender differences in the supports by estimating earnings equations for all working females and all working males without restricting the comparison only to those individuals with comparable characteristics. In the Blinder-Oaxaca decomposition it is 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.

The use of matching criterion in Ñopo decomposition does not require any parametric assumptions and is solely based on the modeling assumption that individuals with the same observable characteristics should be paid the same regardless of sex. Ñopo also does account for gender differences in the supports. The traditional interpretation of two components as developed by Blinder and Oaxaca applies, but only over the common support. Additionally, in the Nõpo's four-element decomposition there are two elements that are attributable to differences in the supports.

The mathematical reasoning of Nõpo is far more complicated that the one from Oaxaca-Blinder decomposition and presenting it in details lies beyond the scope of this work – for more information one may refer to Nõpo (2008). However, for the purpose of this article the details about matching procedure, and estimated components of the decomposition should be introduced.

Nõpo decomposes the gap in average earnings between females and males with the use of matching based on their characteristics, such as age, education and marital status. The procedure that is used to estimate the components of Nõpo's decomposition starts with resampling all females without replacement and matching each observation to one synthetic male, with exactly the same observable characteristics and having the wage obtained from averaging wages of all males exhibiting this set of characteristics. In the paper where the methodology is introduced Nõpo considers only characteristics that can be described with discrete variables and perfect matching. As a result of matching procedure a partition of the data set is generated. The new data set contains observations of matched males, unmatched males, matched females, and unmatched females. Based on this partition the raw gender wage gap can be decomposed into four components: $\Delta = \Delta_M + \Delta_X + \Delta_0 + \Delta_F$.

The first of the four additive components, Δ_M, is the part of the gap that can be explained by differences between two groups of males – those whose characteristics can be matched to female characteristics and those who cannot. This component would disappear in two situations: if for each combination of individual characteristics exhibited in the group of males, it would be possible to find comparable females, or if those unmatched males would earn on average as much as the average matched males. As described by Nõpo (2008) this component 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.

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. This part corresponds to the component attributable to characteristics from Blinder-Oaxaca decomposition, namely $\hat{\beta}^M(\bar{x}^M - \bar{x}^F)$, however limited to the common support.

The third component is called by Ñopo the adjusted gender wage gap. It is the part of the raw wage gap that remains unexplained by differences in characteristics of the individuals and is typically attributed to a combination of both the existence of unobservable characteristics that the labor market rewards and the existence of discrimination. This component correspond to the second component from Oaxaca-Blinder decomposition, that is attributable to differences in average rewards to individuals' characteristics for females and males, $(\hat{\beta}^M - \hat{\beta}^F)\bar{x}^F$, however it is also limited to the common support.

The last component, Δ_F, 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 cannot. As stated in Ñopo (2008) it is computed as the difference between the expected female wages in the common support minus the expected female wages 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.

Three components in Ñopo's decomposition can be attributed to the existence of differences in individuals' characteristics that the labor market rewards (Δ_X, Δ_M, Δ_F) and the other (Δ_0) to the existence of a combination of both unobservable characteristics that should be included in the wage equation if would be observed by econometrician, and the discrimination. Thus the wage gap might be expressed as $\Delta = (\Delta_M + \Delta_X + \Delta_F) + \Delta_0$, and interpreted as it is traditionally done in the linear Blinder-Oaxaca decomposition, with two components: one attributable to differences in observable features of males and females, and the other perceived as an unexplained component.

It should also be mentioned that Ñopo's methodology has its limitations. It is burdened by the course of dimensionality. While the extent to which the raw gender wage gap can be explained depends on the number of explanatory variables, the likelihood of matching decreases with the number of explanatory variables. Variables that suit methodology developed by Ñopo should thus be discrete, allow for precise estimation of unexplained component of wage gap, and at the same time keep the likelihood of matching females to males possibly high.

To sum up, it can be said that the most important advantage of Nõpo's methodology over Blinder-Oaxaca decomposition is that it accounts for differences in the supports of the distribution. According to Nõpo, it is an empirical regularity that the unmatched males have average wages above the average wages of their matched peers and estimating earnings equations for all males without accounting for this regularity tends to overestimate the unexplained component (Δ_0) in the Blinder-Oaxaca decomposition. However, in cases of countries where females exhibit desirable characteristics that the labor market rewards to a greater extent than males, the unexplained component from the Blinder-Oaxaca decomposition could be actually underestimated.

Decompositions of gender wage gap should not be performed without previous verification if the raw gender wage gap and differences in characteristics between males and females exist at all, which is the purpose of the following section of this work.

3 Gender differences in characteristics and the raw gender wage gap in Poland

The goal of this section is an initial empirical analysis of the data on polish employees over the period 1995-2011. The research performed below plays an auxiliary role to decompositions of the gender wage gap that are performed later on. Firstly the raw gender wage gap in Poland over the period 1995-2011 is explored. Then possible determinants of wages are analyzed in the context of gender differences, and variables are prepared for further use in decompositions.

3.1 Raw gender wage gap in Poland, 1995-2011

In the following sub-section of this work average wages of males and females for every quarter of analyzed period are firstly presented. Then the wage gap between males and females in every quarter and in the pooled sample is explored in absolute and relative terms.

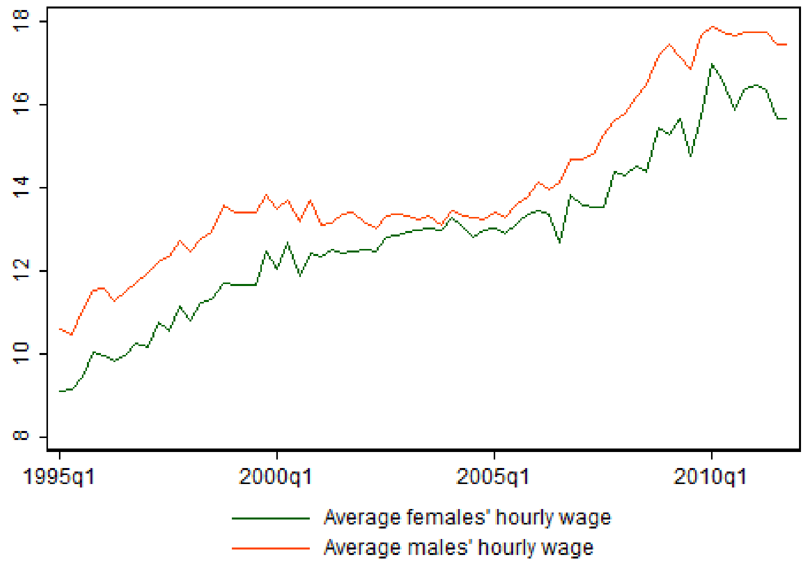
After performing two-group mean comparison test on equality of the means in hourly wages among women and men for each quarter between 1995 and 2011 it might be stated that the raw gender wage gap exists in Poland and in every period is statistically significant (I will call coefficient significant if it is for 5% significance level, and highly significant if it is for 1% significance level).

Figure 1 presents the average absolute real hourly wages for females and males in each of analyzed periods, and the raw wage gap is represented by the vertical difference between the line representing average males' wage and the line representing average females' wages in chosen point of time.

The next figure explicitly shows the real wage differentials and presents it in absolute terms and as percentage of average females' hourly wages in every period. Raw wage gap in relative terms was highest in the first and last five years of the analyzed period and amounted to around 15% of average females' wages. In year 1999 the gap started decreasing and reached the level of around 2% in years 2003 and 2004. Then the gap was increasing until reached its previous level. It is a surprising result that the lowest levels of wage gap were observed after economic downturn in

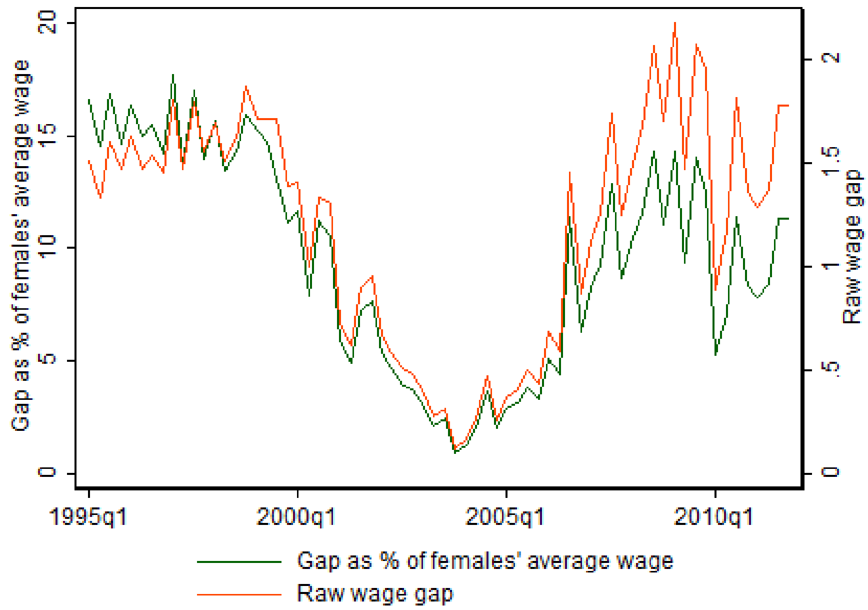
Poland. Explaining the reasons of this results lies beyond the scope of this work, but might be an interesting topic for further analysis.

Figure 1: Females' and males' average hourly wages, 1995-2011 (PLN, constant prices of 1995)



Source: Own preparation

Figure 2: Absolute (PLN, constant prices of 1995) and relative gender wage gap, 1995-2011



Source: Own preparation

When performing the two-group mean comparison test in the pooled dataset containing real wages, it was observed that the average hourly wages for females over the years 1995-2011 were 12.5PLN, while for males it was 13.7PLN. The difference is highly statistically significant and amounts to around 9.3 percent of females' average wage.

3.2 Gender differences in characteristics

After proving the existence of differential in wages of females and males in Poland it is worth exploring if there are significant gender differences in characteristics. In this way it would be possible to get intuition if the gap is simply reflecting gender differences in some observable characteristics or it is rather due to discrimination. Thus, one could assess if the results of decomposition presented in the following section are in line with the observed differences in characteristics. At the same time the variables reflecting the characteristics are prepared for the further use in the decompositions.

More precisely, firstly it is analyzed if certain variable is determinant of wages and thus should be included in matching of Nõpo's decomposition and in Blinder-Oaxaca wage equations. Secondly, the variables are adjusted to the requirements of Nõpo's methodology, which means that continuous variables are transformed into discrete variables, and also the number of categories in each discrete variable is adjusted in order to reflect properly the differences in characteristics that are determinants of wages, but also to allow for relatively high likelihood of matching females to males[4]. Thirdly, it is explored if there is significant difference between females and males in each characteristic chosen for analysis. Finally, the analysis is extended to study how the gender differences in average hourly wages vary according to selected individual characteristics.

The set of characteristics that potentially might be included in the analysis has been chosen according to the one selected by Nõpo *et al.* (2011) or Nõpo (2008). In these studies typically first decomposition takes into account only demographic variables, such as age, region (if it is urban or rural), education and marital status, and the second one includes information on job

[4]In this article, more attention is paid to the Nõpo's decomposition and Blinder-Oaxaca method is applied to control obtained results and draw additional conclusions from the comparison of the results from both decompositions. Thus, the set of characteristics that influence wages should firstly serve the purpose of being used for matching in Nõpo's decomposition and only then is adjusted to be used as the set of explanatory variables in wage equations of Blinder-Oaxaca methodology.

characteristics on the top of the demographics. Among those variables characterizing job, one may find information on the occupation, economic sector or formality. Usually labor characteristics are considered only for main occupation.

In the analysis of this article the set of characteristics is similar to those typically used for Nõpo's decomposition, and the division of demographic and professional characteristics is preserved.

3.2.1 Demographic characteristics

The first variable of our interest is age. It is an empirical regularity that age influences wage, thus this variable is typically included in wage equations. After regressing natural logarithm of hourly wage on age in the pooled dataset it might be said that age is a highly significant determinant of wages.

In the regression where dummies for every age were used, with 15 years as base category, it can be observed that coefficients for age levels above 22 are significant but for age levels close to 15 years old, the coefficients are not significant. It indicates that construction of categories for age levels will improve the likelihood of matching females to males and at the same time the differences between individuals will be still well reflected. In the pooled dataset new variable for age categories was constructed, that classified people of age 25 and younger to the first age category, persons between 26 and 45 years old to the second category, and people older than 45 to the third age category.[5]

In terms of age of working population in years 1995-2011, females are half-year older than males. After checking the difference in age for each quarter separately it might be stated that this difference is stable over time. With such a small difference it is rather impossible that it would cause the differences in wages. If there is bigger difference between average age levels of one group in comparison to another, one may presume that it reflects earlier entrance or earlier retirement from the labor market, which can in turn affect wages. Among Polish employees it is rather not the case.

[5] In the paper only final choice on categories and their boundary values are presented, but before final selection many alternative divisions were tested by including them into Nõpo's decomposition and observing their influence on estimators of wage gap and its components, as well as on the likelihood of matching females to males.

Finally, it is worth investigating how average wages for females and males, as well as the wage gap, vary according to age categories.

Table 2: Hourly wages and gender wage gap for different age groups

Age category	**below 25**	**26-45**	**above 45**
Average males' wage	10.23	13.84	15.14
Average females' wage	9.1	12.55	13.83
Wage gap	1.14	1.29	1.32
Gap as % of average females' wage	12.52	10.26	9.52

Source: Own preparation

According to Table 2 the wage gap in absolute term is the biggest for the third age category, however in relative terms it is the smallest among people above 45 years of age. For the youngest people the gap is biggest and amounts to 12.5 percent of average female's wage in that category.

Variable reflecting marital status distinguishes persons that are single (first category), married (second category), widowed (third category) or divorced/separated (fourth category). After regressing natural logarithm of hourly wage on this variable, it can be stated that marital status is a highly significant determinant of wages and thus should be included in decompositions.

Table 3 reflects how many males and females are in each marital status category, as well as presents absolute and relative differences in numbers of observations. At the same time it contains information on average wages for females and males and the wage gaps in each category.

Important observation is that in three marital status categories, namely married, widowed and divorced/separated average wages are similar, while they are much lower for singles. Also previous regression of hourly wages on dummies for marital status categories with base category "Single" proves that this status negatively influences wages. At the same time there are 35% more single males than females in the period 1995-2011. It is also worth mentioning that in last quarter of 2011 there was 45% more single males than females, as the difference was deepening over time.

Table 3: Quantitative gender differences, hourly wages and gender wage gap for different marital status categories

Marital status	Single	Married	Widowed	Divorced/separated
Number of females	61663	230564	13024	19723
Number of males	83320	276271	2289	6078
Difference in observations	21657	45707	-10735	-13645
Difference as % of females	0.35	0.2	-0.82	-0.69
Females' average wage	11.53	12.76	12.48	13.09
Males' average wage	11.31	14.43	14.98	14.1
Absolute wage gap	-0.22	1.67	2.5	1.01
Gap as % of females' average wage	-1.9	13.13	20.05	7.7

Source: Own preparation

Analyzed persons are also characterized by their education level. Within this variable the lower is the category the better education (1-“Tertiary", 2-"High school", 3-"High school vocational", 4-"Vocational", 5-"Elementary"). It is an empirical regularity that higher education level translates into higher earnings. This rule applies also for the Polish employees, and education level is highly significant determinant of wages. Mean of variable “Education” for analyzed employees amounts to 3.09.

Two-group mean comparison test shows that average education level for females is 2.79 while for males it amounts to 3.36, and the difference is highly statistically significant. This means females are on average better educated than males. Table 4 provides detailed information on how much more females is in better educated (and at the same time higher wage receiving) groups, and how much less females are in categories of lower education level and also lower average wage. Figure 3 additionally shows how differences in education between males and females evolved over time. It can be stated that, in the analyzed period, every year females were becoming relatively better and better educated in comparison to males.

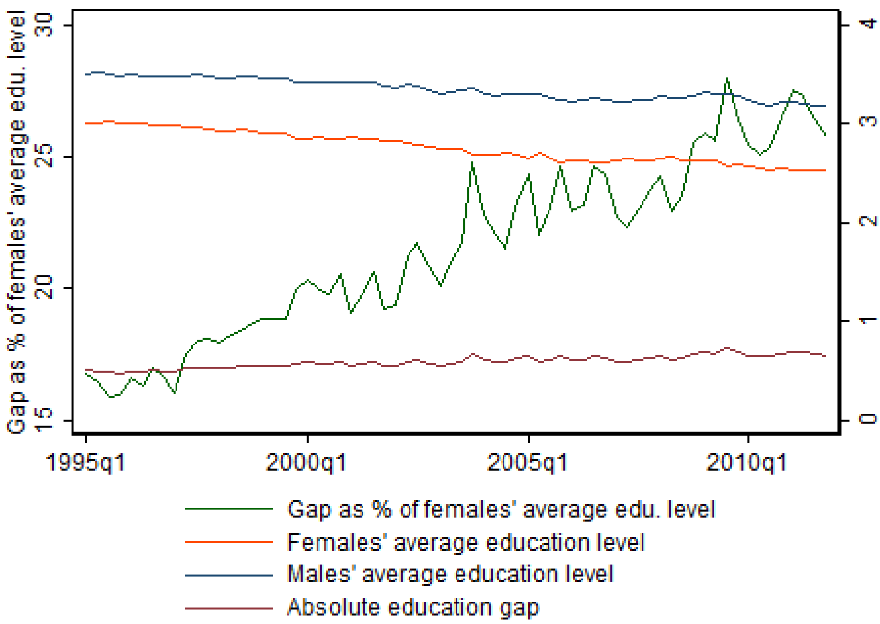
In every education category females earn less than males. The gap in relative terms amounts to around 20%-30% of females’ average wage in 4 categories of lower education. Among people with tertiary education the gap is smaller and amounts to 12% of females’ average wage.

Table 4: Quantitative gender differences, hourly wages and gender wage gap for different education levels

Level of education	Tertiary	High school	High school vocational	Vocational	Elementary
Number of males	45807	22786	92159	166864	40342
Number of females	67396	59830	94395	74469	28884
Difference in observations	-21589	-37044	-2236	92395	11458
Difference as % of females	-32.03	-61.92	-2.37	124.07	39.67
Males' average wage	22.98	14.37	14.1	11.75	10.13
Females' average wage	20.53	11.98	11.28	8.94	8.34
Absolute wage gap	2.45	2.39	2.82	2.8	1.79
Gap as % of females' average wage	11.91	19.98	24.99	31.36	21.48

Source: Own preparation

Figure 3: Average education levels and gender education gap



Source: Own preparation

In the set of demographic characteristics traditionally there is also one describing place of living, precisely indicating if it is rural or urban. In this analysis two variables will be included, “Cities” that takes the value of one if person lives in the city, and “Mazowieckie” which is dummy that takes value of one if respondent lives in Mazowieckie region. Both of them are relevant in the analysis of wages as people in the cities typically earn more than the rest of the society, and people in the region where capital city lies (in case of Poland it is Mazowieckie region) also tend to have higher average wages than people in other regions.

Two-group mean comparison tests have proven that in years 1995-2011 polish employees in the cities were earning on average 14.57PLN, while people outside the cities were receiving hourly wage of 12.32PLN. At the same time people employed in Mazowieckie region were earning 1.89PLN (which constitutes 14% of average hourly wage in the analyzed period) more than employees in other regions. Both differences were highly statistically significant and thus should be included in the set of wage determinants in further analysis.

When it comes to gender differences, 40% of females live in the city, while among males the percentage amounts only to 36%. In Mazowieckie region live 10.2% of females and 9.8% of males, so here the difference is smaller, however both differences are highly statistically significant. It is also worth mentioning that those differences were stable over time.

The gender wage gap is higher in the cities (1.96PLN, 14.5% of females’ average hourly wage in the cities) than outside (0.88PLN, 7.5% of females’ average hourly wage outside cities). On the contrary the gap is smaller in Mazowieckie region (0.82PLN, 6% of females’ average hourly wage in this region), then in other regions (1.23PLN, 10% of females’ average hourly wage in other regions).

After above analysis it might be stated that all chosen demographic variables are determinants of wages and thus should be included in gender wage gap decompositions. When it comes to age there are no major differences between females and males in the analyzed period. After analysis of marital status, it can be said that group of singles has lower wages then people from other marital status categories. At the same time there are more singles among males, and this difference in amount of singles among males and females was increasing over time. Females are also better educated, thus more of them belongs to education categories where wages are on average higher. The difference in education was also 10% bigger in year 2011 than in year 1995. Finally, bigger percentage of females lives in the cities, where wages are on average higher, than in case of males. Share of females living in offering higher wages Mazowieckie region is also higher than in case of males, but the difference is rather small.

The preliminary conclusion is that demographic characteristics cannot be the reason why females earn less than males. According to these characteristics females should rather receive higher hourly wages, thus the existing wage gap shall be caused by differences in some other characteristics or discrimination. Following part of this section will examine if differences in job characteristics could potentially explain the existing gender wage gap.

3.2.2 Job-related characteristics

Several variables are included in the set of characteristics defining working space of an individual, namely occupation category, branch of economy, tenure, and two dummies indicating if job is formal or informal, and if it is in public or private sector.

Characteristic that definitely has and impact on received wage is occupation. In the analyzed dataset occupation categories were at the beginning reflecting ISCO-08 classification, there were 9 categories[6]. In order to reach higher likelihood of matching females to males, those occupations were then grouped into 4 categories that still well reflect differences between individuals. First category consists of higher management, policy makers and specialists; second one characterizes technicians, middle management, office workers, sales and personal services; third category consists of farmers, fishermen, artisans, industrial workers and machine operators; and last, fourth category groups low-skilled occupations. As in case of education lower number means higher-skilled occupation.

After regressing natural logarithm of hourly wage on dummies reflecting occupation categories it can be said that type of occupation highly significantly influences wages. Taking very high-skilled occupations as base category, the coefficient for dummies for second, third and fourth category had more and more negative coefficients respectively (coefficients for second and third category are similar though).

[6] After previous removal of armed forces

The mean of variable reflecting occupation categories among females is 2.17 and among males 2.61. Two-group mean comparison test proved high significance of this difference. This means females are on average more often working in high-skilled occupations than males.

Table 5 provides more detailed information on relations between occupation categories, gender and wage gap. It can be observed that females are dominating in the first category of very high-skilled occupations where wages are highest. They are also dominating in the high-skilled occupations, but the wages in this category are similar to wages in middle-skilled occupations that, on the contrary, are incredibly dominated by males (difference between males and females in this occupation presented as multiplication of females in that category amounts to 376%). Among low-skilled occupations there is also more females, thus the interpretation is not that straightforward. Still it can be said that category of occupation is not the characteristic that explains why males earn more than females.

The wage gap in relative terms was the highest among high- and middle-skilled occupations (26.7% and 28.8% of females' average wages in those categories), while in absolute terms it was similar in first three categories (2.7 – 2.9PLN) and it was lower in fourth (1PLN).

Table 5: Quantitative gender differences, hourly wages and gender wage gap for different categories of occupation

Category of occupation	**Very high-skilled**	**High-skilled**	**Middle-skilled**	**Low-skilled**
Number of males	45548	81993	206535	33049
Number of females	72693	164176	43385	44685
Difference in observations	-27145	-82183	163150	-11636
Difference as % of females	-37.34	-50.06	376.05	-26.04
Males' average wage	23.11	13.90	12.20	9.68
Females' average wage	20.25	10.97	9.47	8.70
Absolute wage gap	2.86	2.94	2.73	0.98
Gap as % of females' average wage	14.13	26.78	28.81	11.27

Source: Own preparation

Apart from occupation, also the branch of economy in which person is working might influence his/her wage. Variable reflecting this possible wages' determinant consists of five categories

(1-“Agriculture", 2 -"Industry", 3-"Construction", 4-"Market services", 5-"Non-market services"). Regressing natural logarithm of hourly wage on dummies for particular categories proved highly significant influence of this variable on wages.

Table 6 shows detailed information on this variable in the context of gender equality. Females are dominating only in fifth category “Non-market services” and there are more than twice more of them working in this sector than males. At the same time it is branch of economy where wages are on average highest. Thus distribution of females among different sectors of economy cannot be the explanation why females are earning less per hour.

The biggest wage gap both in absolute and relative terms is observed in industry and construction sector (23 – 25% of females’ average wages within these groups), gap of 10-15% is observed within both market and non-market services sectors, and very small gap of 2% exists in agricultural sector.

Table 6: Quantitative gender differences, hourly wages and gender wage gap for different branches of economy

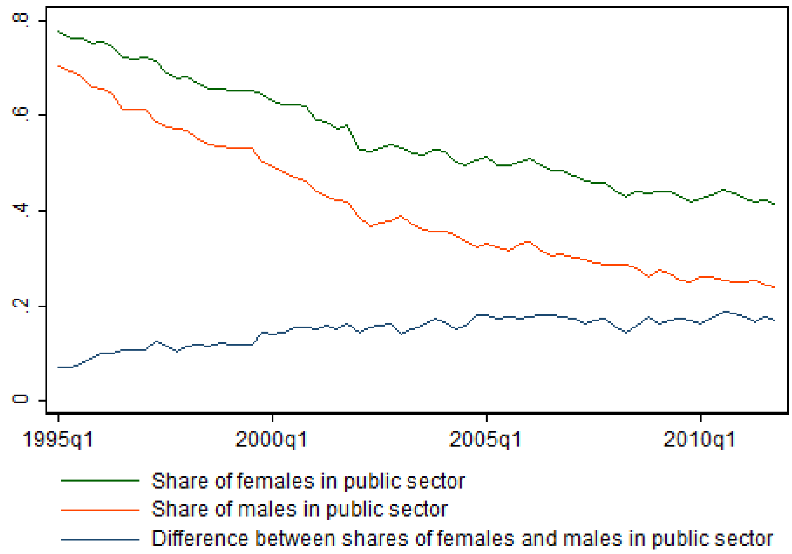
Branch of economy	**Agriculture**	**Industry**	**Construction**	**Market services**	**Non-market services**
Number of males	6788	90175	82200	121981	66349
Number of females	2235	41397	35924	99669	144764
Difference in observations	4553	48778	46276	22312	-78415
Difference as % of females	203.71	117.83	128.82	22.39	-54.17
Males' average wage	10.92	14.62	12.13	13.23	15.66
Females' average wage	10.67	11.61	9.84	11.47	14.23
Absolute wage gap	0.25	3.01	2.28	1.75	1.43
Gap as % of females' average wage	2.34	25.9	23.2	15.3	10.08

Source: Own preparation

When it comes to variables characterizing jobs it might be relevant to know if an individual works in public or private sector. First of all, there are differences in average wages in both sectors. Two-group mean comparison test on analyzed pooled data set has shown that in the period 1995-2011 average hourly wages in Poland in private sector amounted to 12.24PLN, while it was 14.09PLN in public sector. Thus the difference constituted 14% of average wage in this period in favor of public sector and it was highly statistically significant. However it is worth mentioning that the difference was changing over time and in first quarter of 1995 the difference was 11% of average wage, of 2000 it was 16%, of 2005 it was already 35%, and of 2011 it was again smaller and amounted to 27.6% of average wage.

After proving that working in particular sector, private or public, has an impact on wages it should be examined if shares of females or males are dominating in one or another. Common view is that females are more risk averse and prefer more stable and protected jobs in public sector. According to our data set this view is confirmed as over the period 1995-2011 51% of Polish female employees was working in public sector, while for males the percentage was 33%. Thus it can be said that more females are working in the sector where average wages are higher.

Figure 4: Public sector by gender



Source: Own preparation

Figure 4 presents shares of females and males in the public sector, as well as the difference between those shares. It can be said that the difference was increasing over time, thus every year more and more females was working in better paid jobs in public sector in comparison to males. It also can be observed (although such analysis lies beyond the scope of this work) that the shares of both females and males working in public sector were very strongly decreasing over time, which might partially explain the increase of premium for people working in public sector.

The wage gap presented as percentage of females' average wage in analyzed category was 8% in the public sector and 16% in the private.

Another variable that Nõpo (2008) was taking into account as important characteristic of the working place was indicating if an individual is working in formal or informal economy. In case of Peru, that was analyzed by Nõpo, both share of people working in gray economy, and the difference between shares of females and males working in this informal sector were large.

In analyzed data set for Poland over the years 1995-2011 only 1% of people works in informal economy. This low number is due to limitations of Labor Force Survey in Poland and is not indicating the actual size of informal sector. Although the percentage share is low, the absolute number of individuals in the grey economy is still 7150. What is even more important there is huge and highly significant difference between average hourly wages in the formal sector (13.21PLN) and in the informal sector (8.43PLN). Thus this variable could also be included in the set of characteristics that will be used for matching and in wage equations.

Share of females working in grey economy is only 0.8%, and is slightly lower than for males (1.2%). Thus more males are working in the lower paid informal sector (the structure of informal sector is following: 59% of males and 41% of females). The gender wage gap was 10% of females' average wage in the formal sector and 13% in the informal economy.

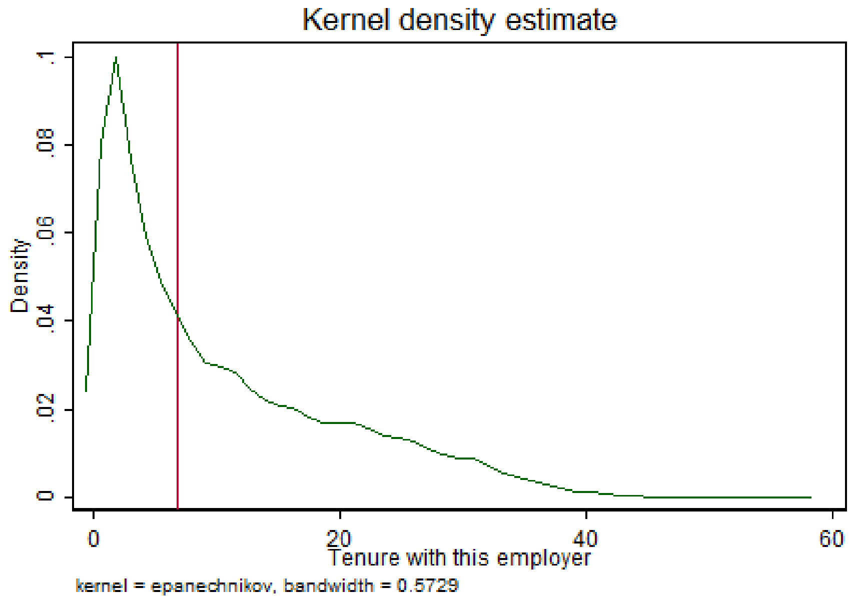
Another variable that reflects professional characteristics is "Tenure". It indicates how long an individual has been working for the current employer. Longer tenure is typically connected with higher wage. In fact after regressing natural logarithm of hourly wage on number of years worked for the same employer, it can be observed that additional year of "Tenure" results in 1.1% higher hourly wage, and this coefficient is highly statistically significant.

When it comes to gender differences in characteristics, average "Tenure" for females is 10.7 years, while for males it is 9.8 years. Again it can be observed that females demonstrate higher level of characteristic that is connected with higher wage.

As tenure is significant determinant of wages it should be included in Nõpo's decomposition. Thus it must be divided into categories that would enable higher likelihood of matching females to males. According to Figure 5 distribution of variable "Tenure" is highly skewed, thus mean is not the best value to divide this variable into two categories (below and above chosen boundary value).

Median, that takes the value of 6.83 years, shall be a better choice in that matter. New variable created for categories of tenure will take value of 1 if an individual is characterized by tenure below median, and value of 2 is his/her tenure is above median.

Figure 5: Kernel density of tenure with current employer



Source: Own preparation

Apart from tenure with current employer, also the overall tenure might influence wages. Potentially females might have lower overall tenure due to maternity leaves and more days off connected to child care. The problem with this characteristic lies in limitations of the data set. Labor Force Survey was not containing the question on overall tenure between years 1997-2005. I will try to work around this problem by creating pooled data set that contains only years 1995-1996 and 2006-2011. It will enable to examine variable that could potentially explain gender wage gap on possibly large data set.

Results from regressing natural logarithm of hourly wage on overall tenure indicates that one more year of professional experience results in increase in wage by 0.5% and it is highly significant result. At the same time average overall tenure for females is 17.3 years, while for males it is 18 years. Although the difference is not especially big and the impact of this characteristic on wage is also not very strong, overall tenure is first variable that could possibly explain part of the gender wage gap.

Last variable that should be taken into account is size of the employer's firm. Typically wages are higher in bigger companies. In analyzed data set there is variable "Size" that has two categories: first one for small enterprises and second one for medium or large enterprises. Data on this variable for years 2010 and 2011 was not coded into data set available for the author, thus those two most recent years has been excluded from the data set used to analyze relations between size of the firm, wage and gender.

Average hourly wages in the medium or large enterprises amount to 12.74PLN, and are by 1.91PLN higher than in small companies. It is highly significant result, thus information about size of the firm should be included in further analysis.

In small companies share of females and males is almost exactly the same, while in medium or large enterprises there is 1.2% more males. The difference is rather small, but it can be stated that variable "Size" might also explain part of the gender wage gap.

Within the individuals that work for small companies the raw gender wage gap amounts to 12% of females' average wage in that group (1.25PLN). In medium or large enterprises the wage differential is smaller both in absolute and relative terms as it is 0.9PLN and only 7.5% of females' average wage in that category.

To sum up, it can be said, that among characteristics on the individual's professional experience or type of working place, only overall tenure and size of the firm might explain part of the gender wage gap. When analyzing other characteristics it was found that more females work in high-skilled occupations, also much more females work in better paid non-market services sector. Additionally more females work in public sector where wages are higher than in private sector, and less females work in grey economy where wages are lower. What is more, females are working on average longer at the same company which also should result in higher wage.

Taking into account both demographic and professional characteristics the intuition of the author after above analysis is that females have on average “more valuable” characteristics than males and the existing gender wage gap in Poland is rather caused by discrimination, than simply reflects differences in characteristics between males and females.

It should be also mentioned that high differentiation in quantities of females and males among many categories of analyzed variables could be problematic for matching females to males (e.g. in one-to-one matching). In Ñopo (2008) the procedure is based on one-to-many matching[7], where females are resampled without replacement, but males are selected with replacement. Thus huge differences in quantities of males and females having certain characteristic should not totally reduce the likelihood of matching, but we keep in mind that controlling for many variables is connected to decrease in shares of “matched” males and females. The results of matching procedure are presented in the following section of this work where decompositions of gender wage gap in Poland are performed.

[7] Recalling from previous section of this work in Ñopo (2008) females are matched to one synthetic male, with the same observable characteristics and wage obtained by averaging wages of all males having exactly this particular combination of characteristics.

4 Decomposition of gender wage gap in Poland

In the previous section of this study the existence of gender wage differentials in Poland has been shown and the differences in characteristics between females and males have been analyzed. Although after such separate analysis of each characteristic first view on the causes of gender wage gap in Poland could be developed, more advanced tools must be used to possibly most precisely measure unexplained component of the gap, and to distinguish this unexplained part from components that are explained by differences in characteristics between males and females.

Firstly the decomposition is performed according to methodology developed by Ñopo (2008) as this relatively new non-parametric approach is of special interest for this work. Then the most widely used Blinder-Oaxaca decomposition is applied to control the results obtained with non-parametric approach and compare both methods. Finally the non-parametric decomposition is performed along distributions of consecutive characteristics in order to control the sensitivity of obtained results.

4.1 Non-parametric approach

According to Ñopo (2008), the raw wage gap is decomposed into four additive components. In the following sub-sections firstly those four components are briefly described. Then the gender wage gap over the whole period 1995-2011 is decomposed with the matching procedure based on few different sets of characteristics. Next, two chosen decomposition specifications are used to decompose the gender gaps in each quarter of analyzed period.

4.1.1 Components of the raw gender wage gap

In order to decompose the gender wage gap between male and female Polish employees over the period 1995-2011 methodology developed by Ñopo is implemented. Recalling from second section of this study, raw gender wage gap, Δ, can be decomposed into four additive components, Δ_M, Δ_X, Δ_F, and Δ_0. It is worth mentioning that measure of raw wage gap used by Ñopo is $\frac{\bar{y}_M - \bar{y}_F}{\bar{y}_F}$, thus the raw difference in average wages of females and males is presented in relative terms as multiplication of females' average wage. The four additive components of the raw wage gap are presented in the same relative way. As the author of this paper was also presenting wage gaps in the previous section as multiplication of females' average wages the results are comparable.

First three components of the decomposition, $(\Delta_M + \Delta_X + \Delta_F)$, make up the explained part of the gap, that is due to differences in characteristics between females and males. The last component, Δ_0, is the unexplained part of the gap, also called the adjusted wage gap, that is due to discrimination, or unobserved differences in characteristics that determine wages. It can be interpreted as difference in average rewards to individuals' characteristics for females and males in the common support.

Among explained components, as it was said in the previous section, Δ_M, can be explained by differences between two groups of males – those who cannot be matched to females and those who can. This component can be also interpreted as expected increase in females' average wage if females achieve those individual characteristics of males that are "unreached" by females.

It is worth mentioning that Nõpo (2008) includes this component, Δ_M, along with the unexplained part, Δ_0, to the "noisy" measures of discrimination. While the component Δ_0 is expressed as discrimination in pay, Δ_M is connected to differences in access to certain combinations of characteristics that are valuable on the market. Of course this is the case in countries where "unmatched" males earn on average more than "matched" males. Nõpo (2008) states that it is an empirical regularity, but it is shown in subsequent analysis, that it is not always the case.

Second component of the explained part of the gender wage gap, Δ_X, is due to differences in distribution of individual characteristics over the common support (for example there are two males and only one female with a particular combination of characteristics). This component expresses how much would average males' wages decrease in a hypothetical situation in which their individual characteristics follow the distribution of females' characteristics (i.e. number of males with particular combinations of characteristics will be equal to the number of females with this combination of characteristics).

The third component included in the explained part of the gap, Δ_F, is explained by differences in average wages of females that can be "matched" to males and of those "unmatched". It measures how the average wage of females would increase if all females achieved the combinations of characteristics that are comparable to those of males.

The values of components of the gap are strictly connected to the set of characteristics that are used for matching. The better the set of characteristics reflects determinants of wages, the more precise measurement of the unexplained component of the gender wage gap. On the other hand, the bigger is set of characteristics used for matching and the more categories each variable has, the lower is likelihood of matching females to males. In this study the author has tried to find a balance between those two targets and the variables has been prepared for the purpose of this study, as presented in the previous section of this work.

4.1.2 Decomposition of the gender wage gap on the pooled sample

Typically two sets of variables are taken into account, one reflecting only basic demographic characteristics, and the other which includes also set of wage determinants that are characterizing the job and business environment of analyzed employee. This approach is preserved in the following sub-section containing the analysis that is made for each quarter in the period 1995-2011, but in case of analyzing pooled data set below, it is possible to show more decompositions based on different combinations of variables. Thus intermediate sets of characteristics, that are between the one that contains only demographic variables, and the one including all recognized determinants of wages, are presented and the changes in estimated components after including each additional variable are discussed.

Table 7 presents the results of the decompositions of gender wage gap in Poland in the period 1995-2011. The first line of the table presents the decomposition based only on demographic variables. In the intermediate decompositions the set of controls was containing demographic variables plus one additional job-related characteristic. The last line shows the results of the decomposition based on matching females to males according to all recognized determinants of wages on which data was available for the whole analyzed period.

The next Table 8 is based on the pooled data set that contains only years 1995-1996 and 2006-2009, so that the broadest set of controlling variables from Table 8 can be additionally expanded with two more variables, one reflecting the size of the company, and the other indicating the overall tenure of an individual.

Table 7: Results of different decompositions

Controls	D	D0	DM	DF	DX	Share of matched males	Share of matched females
Demographic variables	10%	20%	0%	0%	-10%	99%	97%
+ Occupation category	10%	20%	0%	0%	-10%	96%	93%
+ Industry category	10%	20%	-1%	-1%	-9%	92%	92%
+ Private	10%	21%	0%	-1%	-10%	99%	95%
+ Informal	10%	21%	0%	0%	-10%	99%	97%
+Median tenure	10%	21%	0%	-1%	-10%	99%	95%
All variables	10%	19%	-2%	-1%	-6%	65%	74%

Source: Own preparation

According to Table 7, over the period 1995-2011 the adjusted wage gap in Poland (Δ_0 or D0) was around twice bigger than the raw wage gap. The raw wage gap amounts to 10% of females' average wage, while the unexplained component is between 19% and 21%, depending on the set of characteristics used for matching.

Shares of "matched" males and females indicate how big percentage of males and females respectively has the combination of characteristics that could be found among representatives of the opposite gender (or, in other words, how big share of males and females is in the common support).

In the first decomposition only demographic variables were taken into account, i.e. "Age category", "Education level", "Marital status", "Cities" and "Mazowieckie". Additionally, in all decompositions made for the pooled data set there is variable "Time" among the controls, so that females are matched (according to chosen set of characteristics) only with males from the same sub-data set, i.e. from the survey made in the same quarter.

When matching is based on demographic variables almost all males and females are in the common support. This means both components that correspond to non-overlapping supports (DM (Δ_M) and DF (Δ_F)) are very small. It is however worth mentioning that both have negative signs, which means that "unmatched" males actually earn on average less than "matched" males and "unmatched" females earn more than "matched" females.

The component that accounts for differences in characteristics between males and females within the common support is also negative. Moreover its value is almost -10% of females' average wage, thus it is the one that mostly makes up the large difference between raw and adjusted wage gap. As it was previously indicated this component can be interpreted as expected decrease in males' average wage in a hypothetical situation in which their individual characteristics follow the distribution of females' characteristics. Because the component has negative sign the expected result of equalizing the distributions of characteristics for males and females would actually be an increase in males' average wages.

As it was already indicated in the previous sections of this work the results obtained here are not typical, when compared to results presented by Nõpo (2008) for Peru. On the other hand, in the contribution of Nõpo *et al.* (2011), where adjusted and raw wage gap were estimated for the large set of countries, it can be noted that for many European countries the adjusted wage gap was bigger than raw wage gap. In Nõpo *et al.* (2011) the result obtained for Poland (based on data from 2008) was the raw gap of 10.25% and the adjusted gap based on demographic characteristics of 20.66%, thus the results obtained here can be perceived as reasonable. It is also consistent with the intuition gained after the analysis of characteristics made in the previous section of this work, where it was found that females have "more" demographic characteristics that are well rewarded on the market, in comparison to males.

While adding to set of controls one particular job-related characteristic it can be analyzed how it influences the results of the decomposition, and also compare it with intuition from the previous section. While including such variables as "Private", "Informal" or "Median tenure" the unexplained part of the gap has increased. It is in line with the analysis made in the previous section. However, after adding all professional characteristics at the same time, the adjusted wage gap is actually slightly smaller, than in the decomposition containing only demographic characteristics, and it amounts to 19% of the females' average wage.

In the intermediate decompositions, while adding one additional job-related characteristic to the demographic characteristics the shares of "matched" males and females were not decreasing much. However, when all variables were included in the last decomposition, share of matched males decreased to 65% and share of matched females to 74%. It is still satisfying likelihood of matching, thus such decompositions will be prepared for all quarters separately in further analysis. But before that happen two more variables will be added to set of controls, one that reflects size of the company where an individual works, and the other indicating his/her overall tenure.

Table 8 presents the results for the adjusted pooled data set that contains only years 1995-1996 and 2006-2009, as only for this periods data on size of the firm and individual's overall tenure (presented as categorical variable that takes value of 1 if person has overall professional experience below median in the society and value of 2 otherwise) is available.

Table 8: Decompositions results for adjusted data set (years 1995-1996 and 2006-2011) based on two sets of characteristics

Controls	D	D0	DM	DF	DX	Share of matched males	Share of matched females
"Full" set of characteristics	9%	16%	-2%	1%	-6%	69%	78%
+ Size + Median of overall tenure	9%	16%	-3%	3%	-7%	56%	65%

Source: Own preparation

According to Table 8 adjusted wage gap for data set containing only years 1995-1996 and 2006-2009 and based on the "full" set of characteristics (it is called "full" in a sense that it contains all demographic and job-related characteristics available for the whole period 1995-2011) is smaller than adjusted wage gap for whole period 1995-2011 and amounts to 16%. Aadding two additional variables, namely "Size" and "Median of overall tenure" to the "full" set of characteristics does not change adjusted wage gap. Remaining three components have slightly changed, but it can be said that decomposition that does not take into account variables "Size" and "Median of overall tenure" is still relevant.

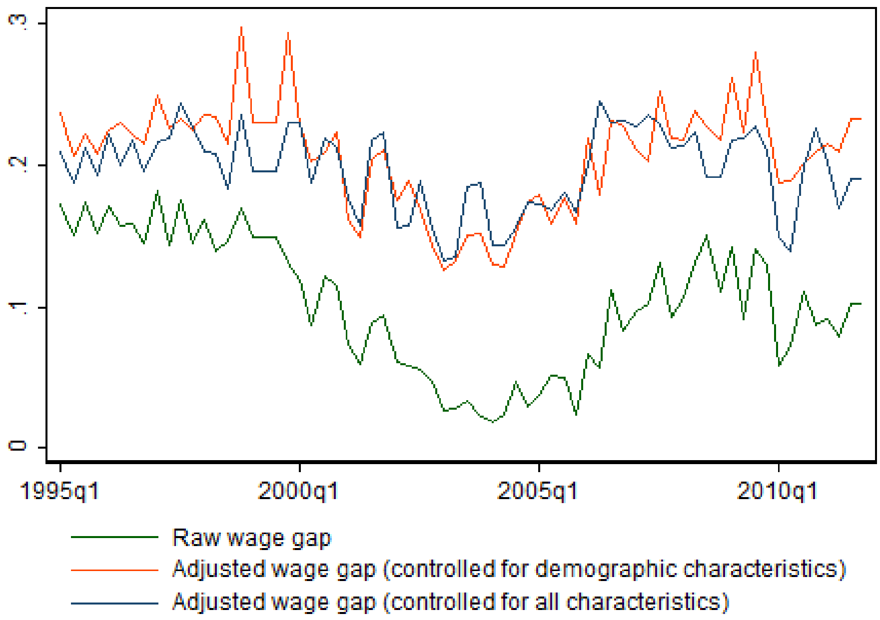
Thus the analysis of each period separately in order to investigate evolvement of adjusted wage gap over time based on two sets of characteristics (one containing only demographic characteristics and the other with all characteristics available for all periods) will be performed consecutively.

4.1.3 Decomposition of the gender wage gap by quarter

This sub-section contains the information on decomposition results for each quarter of the analyzed period. Firstly the evolution of the raw wage gap and adjusted wage gaps (based on two chosen sets of characteristics) over time is discussed. Also the share of "matched" males and females in each quarter are presented. Finally all four components of the decompositions are discussed and presented with the use of additive components bar charts.

Figure 6 below presents the raw wage gap, and adjusted gaps (one controlled for demographic variables and the other for all available variables) over time. For each period adjusted gaps are larger than raw wage gap. It is also visible that the difference between raw wage gap and adjusted gaps is smallest at the beginning of the analyzed period, and from year 2004 the difference seems to be stable on its larger level. Adjusted wage gap obtained after controlling for demographic characteristics is similar to the one controlled for all variables, however the latter seems to be more stable over time.

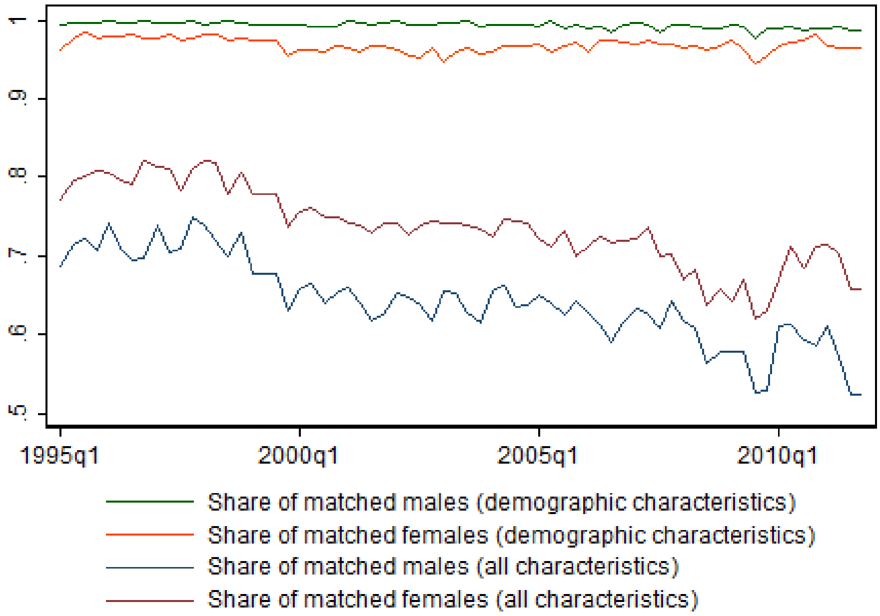
Figure 6: Raw wage gap and adjusted wage gaps over time



Source: Own preparation

Figure 7 shows the shares of "matched" females and males in both decompositions. It is easy to notice that in case of controlling only for demographic variables the likelihood of matching females to males is really high and stable over time. In case of the decomposition where all available variables are taken into account, share of "matched" males and females is much lower and decreases over time. It means that it is harder to match females to males based on job-related characteristics nowadays, than it was 15 years ago.

Figure 7: Shares of "matched" males and females over time



Source: Own preparation

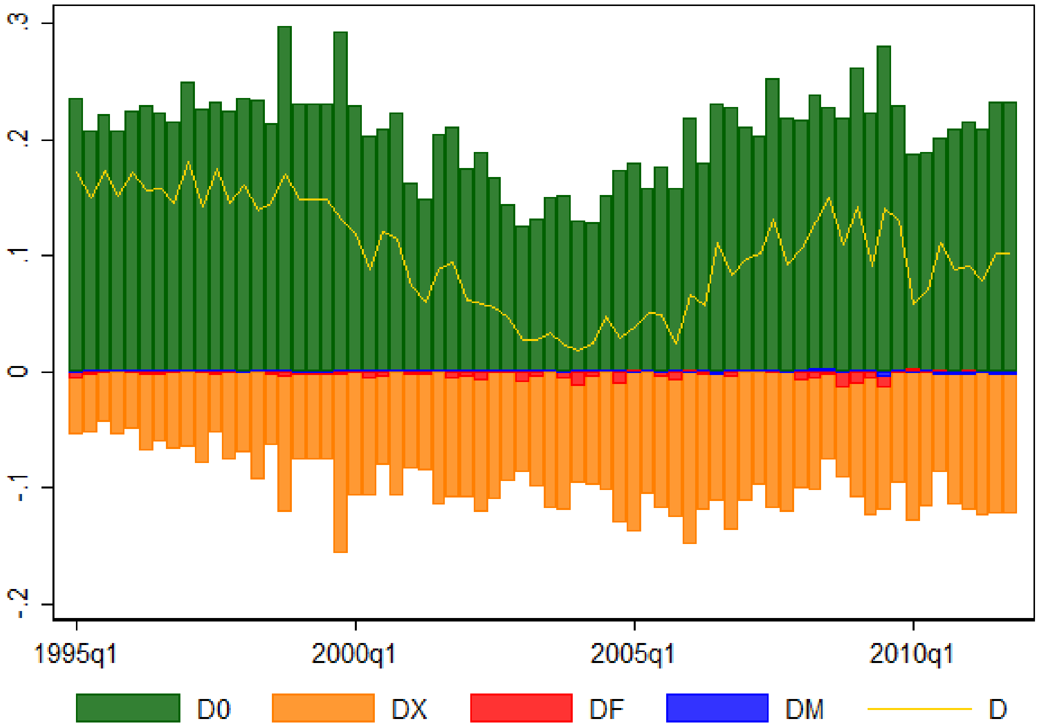
It is also worth investigating how the components of the raw wage gap are changing over time. In the following additive components bar charts, Figure 9 and Figure 10, the height of each component is proportional to the value of the respective component, such that whenever it has a negative value it is represented below the zero line. Sum of these four components is the raw wage gap that is represented by the gold line.

Figure 8 represents the decompositions based on demographic characteristics, while Figure 9 on all characteristics. It is observed that in the former figure components attributable to differences in the supports, DM and DF, are almost invisible (due to very high likelihood of matching), while in the latter one may notice both components DM and DF in each quarter, and DM can be perceived as important component of the gap.

The most important observation is that the component connected to differences in distribution of characteristics between males and females in the common support is always negative. This means that over the whole period 1995-2011 females had more representatives with valuable characteristics in the common support than males, and males could expect an increase in wages in hypothetical situation when they would have the same distribution of characteristics in the common support as females.

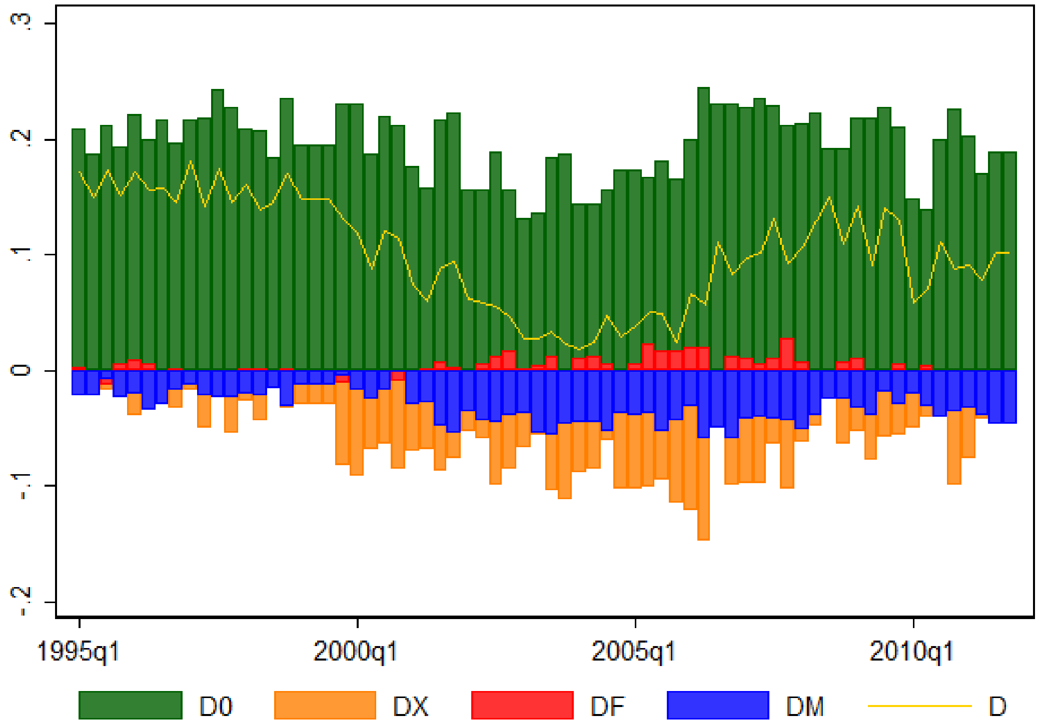
Similar situation is with component DM in Figure 9 where it is always negative. It means that "unmatched" males earned on average less than "matched" (based on all characteristics) males over the whole period 1995-2011. Thus if all the males had combinations of characteristics that can be matched to females, and their distribution in the common support was the same, males wages would be expected to increase, and the raw wage gap would be actually higher. In this hypothetical situation the raw wage gap would be more similar to the adjusted wage gap that is due to discrimination or unobserved characteristics that determine wages.

Figure 8: Results of the decomposition based on demographic variables



Source: Own preparation

Figure 9: Results of decomposition based on all available variables



Source: Own preparation

Summarizing this sub-section, it can be said that gender wage differentials in Poland over the period 1995-2011 could not be explained by the differences in characteristics between males and females (in a sense that males express "more" valuable characteristics than females). Actually the component attributable to differences in endowments between females and males (DX) has a negative value in each quarter of 1995-2011, which signifies that females are more endowed with characteristics that are well rewarded on the market, in comparison to males. In the decomposition where both demographic and job-related characteristics are taken into account also the component attributable to differences between "unmatched" and "matched" males (DM) plays an important role. What is more over the whole analyzed period it has a negative value, which means that "unmatched" males are earning less than "matched" males.

The most important conclusion from this sub-section is that adjusted gender wage gap in Poland, according to performed decompositions, is around twice bigger than the raw wage gap. What is more, this adjusted gender wage gap that is often perceived as measure of discrimination neither is decreasing over time, nor its cyclical behavior was observed. In the next sub-section of this work second method of decomposition is applied in order to control obtained results and compare both methodologies.

4.2 Parametric approach

Blinder-Oaxaca decomposition is a traditional method to distinguish explained and unexplained components of the wage differentials between two groups in society. The parametric approach plays an auxiliary role in this analysis and is used for comparative purposes. Thus neither this decomposition is performed for each quarter of the analyzed period, nor are the decompositions on the pooled sample applied with many different sets of controls. The analysis with the use of parametric approach is limited to two decompositions on the pooled sample, first one that is based only on demographic characteristics, and the second one that takes into account all characteristics available for the whole analyzed period.

Before parametric decompositions are performed, respective results from previous sub-section of this work are recalled. Nõpo's decomposition technique divides the gap into four additive elements, two of which are analogous to the elements of the Blinder-Oaxaca decomposition, but limited to the common support, and the other two are attributable to differences in the supports. Thus the unexplained and explained part of the gap should be similar in both methods of decomposition if there is no major difference in the supports[8]. On the other hand, when there are important differences in the supports, Nõpo's methodology should provide more precise results. Also in that case, Blinder-Oaxaca should provide proper estimator of the unexplained component if performed only on the common support of the distribution. In the following analysis all those hypotheses will be empirically tested with the use of pooled data set on polish employees over the period 1995-2011.

[8] Differences between respective components may also result from differences in measurement of the raw wage gap in both decompositions. Measure of the raw wage gap used in the regression approach is $\overline{\ln(y_M)} - \overline{\ln(y_F)}$, while Nõpo is using a different measure i.e. $\frac{\bar{y}_M}{\bar{y}_F} - 1$, as he believes that the latter corresponds better to the concept if the gender wage gap. For small differences in average wages the regression measure is a good approximation of Nõpo's measure, but if differences are big, the approximation may be poor and is it not possible to establish an order relationship between the two measures. In case of this analysis the measure of the raw wage with the use of both approaches is 10% thus the problem described above should not complicate the comparison of both methods.

4.2.1 Decomposition based on demographic characteristics

When only demographic characteristics are taken into account the likelihood of matching in Nõpo is close to 100%, and the problem of differences in the supports of the distribution should not influence the results. Table 9 recalls the result of Nõpo's decomposition performed on the pooled data set on Polish employees over the period 1995-2011 and based on demographic variables. It can be observed that the components attributed to differences in the supports, DM and DF are equal to zero, and thus estimator of unexplained component from Oaxaca-Blinder decomposition should be similar to D0 from Table 9, namely 20%.

Table 9: Nõpo's decomposition based on demographic characteristics

Controls	**D**	**D0**	**DM**	**DF**	**DX**	**Share of matched males**	**Share of matched females**
Demographic variables	10%	20%	0%	0%	-10%	99%	97%

Source: Own preparation

If one wants to compare the results obtained in both methodologies, the earning equations in parametric approach should control for the same characteristics and in the most similar way as performed in matching. According to Nõpo (2008) it is worth using dummies instead of continuous variables in wage equations, so that the setup has lower dependence on the functional form of the earnings equations (as in matching).

Wage equation that should reflect the Nõpo's decomposition presented above includes following variables: two dummies reflecting age category ("26-45" and "above 46", with "below 25" as base category), four dummies for education level ("Tertiary education" as base category), three dummies for marital status (with "Single" as base category) and the dummies for "Cities" and "Mazowieckie". Additionally there are dummies for each quarter of years 1995-2011 as the matching in the pooled data set was also performed with variable "Time" among the controls.

Table 10: Females' and males' wage equations based on demographic characteristics

Variables	**Males' equation**	**Females' equation**
Age 26 – 45	0.139***	0.187***
	(0.00288)	(0.003)
Age above 46	0.168***	0.288***
	(0.00329)	(0.00331)
High school	-0.382***	-0.457***
	(0.00393)	(0.00259)
High school vocational	-0.395***	-0.513***
	(0.00275)	(0.00235)
Vocational	-0.548***	-0.720***
	(0.00258)	(0.0025)
Elementary	-0.692***	-0.813***
	(0.00345)	(0.00343)
Cities	0.102***	0.0577***
	(0.0017)	(0.00166)
Mazowieckie	0.0929***	0.135***
	(0.00236)	(0.00232)
Married	0.165***	0.0664***
	(0.00235)	(0.00245)
Widowed	0.145***	0.0589***
	(0.0108)	(0.00485)
Divorced/Separated	0.0732***	0.0570***
	(0.00668)	(0.00401)
Constant	2.407***	2.334***
	(0.00771)	(0.00759)
Observations	233,710	203,686
R-squared	0.335	0.443

Standard errors are below coefficients. Tertiary education, age below 25 and single as base levels. Logarithm of hourly wage as a dependent variable.
*** p<0.01, ** p<0.05, * p<0.1

It can be said that the results obtained with Blinder-Oaxaca decomposition are similar to ones reached with Nõpo's methodology. It confirms that adjusted gender hourly wage gap in Poland over the period 1995-2011, based on demographic characteristics, is around twice bigger than the raw wage gap. The reason for this is that the component of the raw gap that is attributable to endowments is negative, which means that males have "less" characteristics that are well rewarded on the market in comparison to females.

4.2.2 Decomposition based on all variables

In the following analysis the set of explanatory variables in wage equations will be expanded to include all characteristics available over the whole period 1995-2011.

Table 19 recalls the results of Nõpo's decomposition performed on the pooled dataset and based on both demographic variables, and job-related variables, namely occupation category, branch of the economy, tenure with current employer, and information on the company's environment (public/private, formal/informal). As usually variable "Time" is also included among controls.

Table 11: Nõpo's decomposition based on all characteristics

Controls	**D**	**D0**	**DM**	**DF**	**DX**	**Share of matched males**	**Share of matched females**
Demographic variables	10%	19%	-2%	-1%	-6%	65%	74%

Source: Own preparation

Wage equation that would possibly best reflect the matching used in above decomposition, includes demographic variables and "Time" variable specified in the same way as in the previous Blinder-Oaxaca decomposition, and also three dummies for categories of occupation (with "Very high skilled" as base category), four dummies for branches of economy (with "Agriculture" as base category), and dummies for "Median tenure", "Private" and "Informal".

Table 12 below presents the coefficients for all the variables. Again there are differences between rewards that females get for their characteristics, in comparison to males' rewards for the same features. Among all twenty-one explanatory variables females are rewarded better than males only for being in second or third age category, for working in Mazowieckie region, and are less penalized for being in the fourth occupation category (low-skilled) or in informal economy. As a result males have better coefficients in case of sixteen explanatory variables, and have higher shift coefficient.

After performing the Blinder-Oaxaca decomposition with the use of prepared wage equations the raw wage gap of 10.1% consists of component attributable to endowments, that amounts to 11.5%, and an unexplained component of 21.6%. Adding job-related variables has not changed the results of the decomposition. However, it should be underlined that it is not regularity, but rather an exception.

Table 12: Females' and males' wage equations based on all characteristics

Variables	**Males' equation**	**Females' equation**
Age 26 – 45	0.0824***	0.115***
	(0.00281)	(0.00295)
Age above 46	0.0772***	0.177***
	(0.00327)	(0.00337)
High school	-0.203***	-0.257***
	(0.00431)	(0.00293)
High school vocational	-0.215***	-0.298***
	(0.00352)	(0.00283)
Vocational	-0.325***	-0.443***
	(0.00379)	(0.00332)
Elementary	-0.426***	-0.502***
	(0.00443)	(0.00419)
Cities	0.0978***	0.0655***
	(0.00163)	(0.00158)
Mazowieckie	0.112***	0.140***
	(0.00226)	(0.00219)
Married	0.130***	0.0467***
	(0.00226)	(0.00233)
Widowed	0.111***	0.0398***
	(0.0103)	(0.00459)
Divorced/Separated	0.0649***	0.0464***
	(0.00637)	(0.00379)
High-skilled occupation	-0.270***	-0.308***
	(0.00346)	(0.00262)
Middle-skilled occupation	-0.302***	-0.355***
	(0.00372)	(0.00407)

Low-skilled occupation	-0.436***	-0.417***
	(0.0045)	(0.00369)
Industry	0.166***	0.0431***
	(0.0063)	(0.00996)
Construction	0.0935***	0.0802***
	(0.0064)	(0.0101)
Market services	0.0612***	-0.0225**
	(0.00631)	(0.00987)
Non-market services	0.0294***	-0.0115
	(0.00653)	(0.0099)
Median tenure	0.115***	0.0997***
	(0.00182)	(0.00182)
Private	-0.0726***	-0.0166***
	(0.00193)	(0.00214)
Informal	-0.159***	-0.217***
	(0.00749)	(0.00849)
Constant	2.459***	2.412***
	(0.00978)	(0.0123)
Observations	233,710	203,686
R-squared	0.396	0.502

Standard errors are below coefficients. Tertiary education, age below 25, single, very high-skilled occupation and agriculture sector as base levels. Logarithm of hourly wage as a dependent variable.

*** p<0.01, ** p<0.05, * p<0.1

After performing Blinder-Oaxaca decomposition with all available explanatory variables included in the wage equations, results are only slightly different than those obtained with the Nõpo's methodology. However, in Nõpo's decomposition based on full set of characteristics, the shares of "matched" males and females are 65% and 74% respectively. In that way one may presume that problem of differences in the supports may occur while performing Blinder-Oaxaca decomposition with wage equations including broader set of characteristics.

Table 13 below compares the results of decompositions based on full set of characteristics. First one is Nõpo's decomposition and the second one is Blinder-Oaxaca decomposition that does not account for differences in the supports, and the third is Blinder-Oaxaca decomposition that is performed only on the common support[9].

Table 13: Comparison of decompositions

Decomposition type	Raw gap	Unexplained component	Component attributable to endowments	Component attributable to differences in the supports
Nõpo	10%	19%	-6%	-3%
Blinder-Oaxaca	10.10%	21.60%	-11.50%	-
Blinder-Oaxaca over common support	**10.10%**	**21.10%**	**-8.70%**	**-2.30%**

Source: Own preparation

It can be said that Blinder-Oaxaca decomposition performed only over common support do produces results that are more similar to results from Nõpo's decomposition than those obtained in Blinder-Oaxaca decomposition that does not account for differences in the supports. When comparing two Blinder-Oaxaca decompositions it can be observed that the estimation of the unexplained component has not changed very much. However the component attributable to endowments from standard Blinder-Oaxaca decomposition was divided into two components in parametric decomposition over common support: one attributable to endowments, and one attributable to differences in the supports. It should be mentioned that in more "typical" cases when "unmatched" males are earning on average more than "matched" males, and/or "unmatched" females are earning less than "matched" females the unexplained component in Blinder-Oaxaca decomposition that does not account for differences in the supports might be more overestimated (Nõpo, 2008).

To sum up, it can be said that estimators of explained and unexplained gender wage gap in Poland over the period 1995-2011 obtained with the use of methodology developed by Nõpo (2008) has been confirmed with the parametric approach developed by Oaxaca (1973) and

[9] Data set containing only variables from the common support was obtained in the following way: after performing Nõpo's decomposition in the data set there is new categorical variable that indicates if observation is in the common support or not. Thus only observations from the common support were kept in adjusted data set.

Blinder (1973). Although standard Oaxaca-Blinder decomposition does not account for differences in the supports, the estimator of unexplained component for analyzed data set is still similar to those obtained with Blinder-Oaxaca decomposition that was performed on the common support, and in Ñopo's decomposition.

4.3 Sensitivity analysis

In the previous sub-sections two decomposition techniques were applied to decompose gender wage gap in Poland. Results indicate that the adjusted wage gap is around twice bigger than the raw wage gap. Decompositions were performed on the pooled sample as well as in each quarter separately, and the estimator of adjusted wage gap is around 20% in case of the pooled sample, but also is stable over time. In order to assure robustness of the results it is also worth investigating how the decomposition results vary among different categories of analyzed characteristics. In this sub-section only the non-parametric approach is applied, and the decompositions are based on all characteristics available in the analyzed pooled sample.

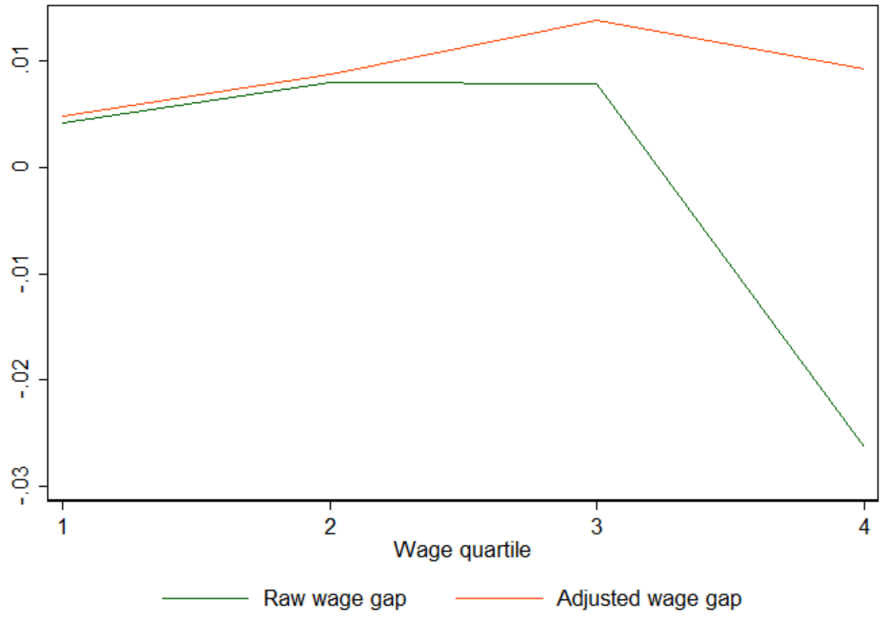
Firstly it is explored how the results very along the distribution of hourly wages. Table 14 compares the raw wage gap and adjusted wage gap within wage quartiles. It is worth mentioning that the raw wage gap is biggest in the middle of the distribution, while in the quartile of highest hourly wages it is even negative. What is especially interesting is that the difference between adjusted wage gap and raw wage gap is increasing along with wage quartiles. It is presented in a very transparent way on Figure 10.

Table 14: Raw and adjusted gender wage gap among wage quartiles

Wage deciles	**1**	**2**	**3**	**4**
Raw wage gap	0.43%	0.8%	0.79%	-2.63%
Adjusted wage gap	0.48%	0.88%	1.38%	0.93%

Source: Own preparation

Figure 10: Raw and adjusted gender wage gap along wage quartiles



Source: Own preparation

It can be stated that in each wage quartile the adjusted wage gap is bigger than the raw wage gap. Next, it should be investigated how two indicators of wage differentials vary according to different categories of characteristics.

According to Table 15 unexplained component of the gender wage gap is highest within the middle aged group (21% of females' average wage in this category). What is especially important is that for every age category adjusted wage gap is much higher than raw wage gap.

Table 15: Raw and adjusted gender wage gap by age categories

Age category	**below 25**	**26-45**	**above 45**
Raw wage gap	12%	11%	10%
Adjusted wage gap	17%	21%	17%

Source: Own preparation

When it comes to different levels of education the smallest adjusted wage gap is observed within the group of most educated people, those with tertiary education. According to Table 16 adjusted wage gap in this group amounts only to 12% of average wage of females with highest level of education. The biggest gap can be observed within the group of people with high school vocational (26%) and vocational (24%) education. It can be also observed that raw wage gap is similar or even bigger than adjusted wage gap. It is due to the fact that level of education is a very important determinant of wages, and at the same time females are significantly better educated than males. When matching is performed within particular education group, and females do not have superiority in this field, the adjusted wage gap can be smaller than the raw wage gap. It indicates that level of education is probably the most important characteristic due to which component attributable to differences in characteristics is usually negative, and, as a result, adjusted wage gap is bigger than raw wage gap.

Table 16: Raw and adjusted gender wage gap by education

Level of education	Tertiary	High school	High school vocational	Vocational	Elementary
Raw wage gap	13%	20%	26%	32%	22%
Adjusted wage gap	12%	22%	26%	24%	16%

Source: Own preparation

According to Table 17 and 18 the adjusted wage gap is bigger than raw wage gap both in rural and urban area, as well as both in Mazowieckie region and others. What is more, raw and adjusted gender wage gap is smaller in rural areas in comparison to cities, and there is smaller gap in Mazowieckie region when compared with other Polish regions.

Table 17: Raw and adjusted wage gap by type of area (urban/rural)

Area	Rural	Urban
Raw wage gap	8%	14%
Adjusted wage gap	18%	21%

Source: Own preparation

Table 18: Raw and adjusted gender wage gap by type of region (Mazowieckie/other)

Region	Other	Mazowieckie
Raw wage gap	11%	6%
Adjusted wage gap	20%	17%

Source: Own preparation

The distribution of the gap with respect to different categories of occupation is similar to one observed within educational levels, namely the lowest gap is in the group of people with highest and lowest skills, while bigger gap is observed in the middle. What is more the raw wage gap is also similar or even bigger than adjusted wage gap. This indicates that occupation, similar to education, is an important determinant of wages, in which females have superiority. What is more when the decomposition is performed on certain category of occupation, the differentiation of education is much smaller. Then the component attributable to differences in endowments may actually become positive or at least less negative. However, even if analysis is limited to certain category of occupation or education the adjusted wage gap is positive and higher than 10%. The results are presented in Table 19 below.

Table 19: Raw and adjusted gender wage gap by occupation category

Category of occupation	**Very high-skilled**	**High-skilled**	**Middle-skilled**	**Low-skilled**
Raw wage gap	16%	27%	30%	12%
Adjusted wage gap	12%	25%	27%	13%

Source: Own preparation

According to Table 20 the smallest gap is observed in the agricultural sector where the wages are lowest and in non-market services where wages are highest. This is in line with previous results about distribution of the gap with respect to education level or skills. Within such branches of economy as market services, construction and industry the gap is bigger. What is more adjusted wage gap is bigger than raw wage gap for agriculture, market services and non-market services, while in case of construction or industry sector it is slightly smaller than raw wage gap.

Table 20: Raw and adjusted gender wage gap by branch of economy

Branch of economy	**Agriculture**	**Industry**	**Construction**	**Market services**	**Non-market services**
Raw wage gap	2%	27%	22%	16%	11%
Adjusted wage gap	15%	25%	21%	20%	18%

Source: Own preparation

According to Table 21 adjusted gaps in private and public sector are similar and amount to 19% and 20%. In both cases they are bigger than raw wage gap. Table 22 shows the distribution of the gap with respect to formality. In the formal sector the estimators are exactly the same as for the whole pooled sample, as quantitatively it captures 99% of observation. However, it might be surprising that the adjusted wage gap is smaller in the informal sector than in formal one, while the raw wage gap is bigger in that sector in comparison to the other. It can be caused by the fact that informal sector is attracting very specific group of individuals among which females do not have superiority in endowments.

Table 21: Raw and adjusted wage gap by public/private sector

Type of sector	Public	Private
Raw wage gap	9%	16%
Adjusted wage gap	19%	20%

Source: Own preparation

Table 22: Raw and adjusted wage gap by formality

Formality	Formal	Informal
Raw wage gap	10%	14%
Adjusted wage gap	19%	13%

Source: Own preparation

To sum up, it can be said that adjusted wage gap within particular groups of society is always positive and vary between 12% (within group of people with tertiary education or group of people with very high-skilled occupation) and 27% (among people with middle-skilled occupations). Thus it is significantly bigger than the average raw wage gap over the period 1995-2011 that amounts to 10%.

5 Conclusions

Inequalities induced by discrimination pose a serious challenge to both policymakers and society. This rationale underlies equal access legislation in many developing and developed countries. The success of such policies usually consists of opening up many professions to highly skilled individuals previously deprived of the opportunity to adequately use their abilities. Analyses like "The Allocation of Talent and U.S. Economic Growth" (Chang-Tai, Hurst, Jones and Klenow, 2011) reveal that barriers like racial and gender discrimination may lead to a considerable loss in productivity and wealth even in highly industrialized, democratic countries like the USA. In this paper we analyzed the problem of gender wage gap in Poland, trying to reliably measure its size and observe time-related patterns. We inquired if gender wage gap in Poland may be explained by observable characteristics

Gender wage gap may simply reflect the differences in observable characteristics between males and females. The real challenge lies in providing reliable measures of wage gaps, and statistical tools constructed to decompose wage gaps has been arising. Two of those methods are of special interests for this work, parametric Blinder-Oaxaca decomposition, and its non-parametric alternative developed by Nõpo (2008). The latter decomposes wage gap into four components, two of which are equivalents to those from parametric approach (first attributable to differences in characteristics and second to differences in rewards), and other two account for gender differences in the distribution of characteristics.

Analysis of gender differences in characteristics demonstrates that females to a greater extent exhibit characteristics that are well rewarded in the labor market. Despite better education, they are less frequently employed in better paying positions. Decomposition analyses confirm this assertion, showing that the discrimination component quantitatively dominates. In fact, gender wage gap in Poland, understood as the difference in average male and female wages, cannot be explained by gender differences in observable characteristics. More precisely, estimators of actual gender gap in hourly wages obtained with both parametric and non-parametric methods indicate that a measure adjusted for differences in characteristics is actually twice as big as the raw wage gap differential and amounts to as much as 20%. Furthermore, neither raw nor the adjusted gender wage gap seems to be decreasing over time.

The adjusted wage gap, as analyzed in this work, can be attributed to differences in both the unobservable characteristics and the discrimination. Part of the unobservable heterogeneity may be accounted for by a wide selection of variables included in the analysis as well as their interactions. Although we are unable to distinguish in data between a legal clerk in a municipality and a head of lawyers' team in an international corporation, an interaction of sector, employer and residence characteristics takes care of such discrepancies to a large extent. Consequently, if adequately tackled, matching procedure minimizes the risk of attributing unobservable heterogeneity to discrimination. On the other hand, the more characteristics are controlled for, the smaller the likelihood of matching females to males. This feature, the curse of multi-dimensionality, remains unsolved problem of Nõpo methodology and thus leaves few avenues open for further research. Among the most promising approaches are methods that allow relaxing the perfect matching requirement and discrete domain for characteristics. For example, propensity score matching, which introduces some notion of distance between two "similar" individuals may indeed allow for even deeper analysis of the wage gaps phenomena.

Bibliography

Adamchik, V.A., Bedi, A. S. (2003). Gender pay differentials during the transition in Poland, *The Economics of Transition* vol. 11, issue 4, p. 697.

Atal, J. P., Hoyos, A., Nopo, H. (2010). NOPOMATCH: Stata module to implement Nopo's decomposition, *Statistical Software Components*, S457157, Boston College Department of Economics.

Bauer, T., Sinning, M. (2008). An extension of the Blinder–Oaxaca decomposition to nonlinear models, *Advances in Statistical Analysis*, Springer, vol. 92(2), pages 197-206, May.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates, *The Journal of Human Resources*, VII, 4, pp. 436-55.

Chang-Tai, H., Hurst, E., Jones, C. I, Klenow, P. J. (2011). The Allocation of Talent and U.S. Economic Growth, Working Paper.

Cotton, J. (1988). On the Decomposition of Wage Differentials, *Review of Economics and Statistics*, MIT Press, vol. 70(2), pages 236-43, May.

DiNardo, J., Fortin, N. M., Lemieux, T. (1996), .Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, *Econometrica* 64: 1001-1044.

Duflo, E. (2011). Women empowerment and economic development, MIT Working Paper.

Fairlie, R. W. (2003). An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models, Working Papers 873, Economic Growth Center, Yale University.

Firpo, S., Fortin, N. M., Lemieux, T. (2007). Decomposing Wage Distributions using Recentered Influence Functions Regressions, mimeo, University of British Columbia.

Fortin, N., Lemieux, T., Firpo, S. (2011). Decomposition Methods in Economics, Handbook of Labor Economics, Elsevier.

Grajek M., (2003). Gender Pay Gap in Poland, *Economic Change and Restructuring*, Springer, vol. 36(1), pages 23-44, March.

Heckman, J. J. (1977). Sample Selection Bias As a Specification Error (with an Application to the Estimation of Labor Supply Functions), NBER Working Papers 0172, National Bureau of Economic Research, Inc.

Jann, B. (2008). A Stata implementation of the Blinder-Oaxaca decomposition. ETH Zurich Sociology Working Paper No. 5.

Juhn, C., Murphy, K. M., Pierce, B. (1991) .Accounting for the Slowdown in Black-White Wage Convergence, in Workers and Their Wages: Changing Patterns in the United States, ed. by M. H. Kosters, American Enterprise Institute, Washington.

Juhn, C., Murphy, K. M., Pierce, B. (1993).Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy* 101: 410-442.

Machado, J. F., Mata, J. (2005). Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression, *Journal of Applied Econometrics* 20: 445.465.

Neumark, D., (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination, *Journal of Human Resources*, University of Wisconsin Press, vol. 23(3), pages 279-295.

Ñopo, H., Daza, N., Ramos, J. (2011). Gender Earnings Gaps in the World, IZA Discussion Papers 5736, Institute for the Study of Labor (IZA).

Ñopo, H. (2008). Matching as a Tool to Decompose Wage Gaps, *Review of Economics and Statistics*, 90(2): 290-299.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Market. International Economic Review, Vol.14, No.3, 693-709.

Oaxaca, R. L., Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials, *Journal of Econometrics*, Elsevier, vol. 61(1), pages 5-21, March.

Reimers, C. W. (1983). Labor Market Discrimination against Hispanic and Black Men, *The Review of Economics and Statistics*, MIT Press, vol. 65(4), pages 570-79, November.

Pratap, S., Quintin, E. (2002). Are Labor Markets Segmented in Argentina? A Semiparametric Approach, Instituto Tecnologico Autonomo de Mexico discussion paper 02-02.

Rosenbaum, P., Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70:141–55.

List of Tables

List of Figures