the review of income and wealth

Review of Income and Wealth
Series 63, Number 1, March 2017
DOI: 10.1111/roiw.12209

WHICH GENDER WAGE GAP ESTIMATES TO TRUST? A COMPARATIVE ANALYSIS

BY KAROLINA GORAUS

University of Warsaw

JOANNA TYROWICZ

University of Warsaw; National Bank of Poland

AND

LUCAS VAN DER VELDE*

University of Warsaw

The aim of this paper is to compare estimates of the adjusted wage gap from different methods and sets of conditioning variables. We apply available parametric and non-parametric methods to LFS data from Poland for 2012. While the raw gap amounts to nearly 10 percent of the female wage; the adjusted wage gap estimates range between 15 percent and as much as 23 percent depending on the method and the choice of conditional variables. The differences across conditioning variables within the same method do not exceed 3pp, but including more variables almost universally results in larger estimates of the adjusted wage gaps. Methods that account for common support and selection into employment yielded higher estimates of the adjusted wage gap. While the actual point estimates of adjusted wage gap are slightly different, all of them are roughly twice as high as the raw gap, which corroborates the policy relevance of this methodological study.

JEL Codes: C24, J31, J71

Keywords: gender wage gap, Poland, decomposition methods

1. INTRODUCTION

For a long time, the literature on the gender gap in wages has been dominated by just a handful of techniques, namely the Oaxaca (1973) and Blinder (1973) (hereinafter Oaxaca-Blinder) decomposition and dummy variables in pooled regressions of different types (OLS, IV, etc.). The estimates were referred to as adjusted wage gap, i.e. the size of the gender wage gap controlling for differences in characteristics important for productivity (such as age, education, industry, occupation, firm characteristics, etc.). However, these methods are troubled with weaknesses well recognized in the literature: the estimators cannot be easily

Note: Authors would like to thank Gianna C. Giannelli, Irene van Staveren, Wojciech Maciejewski, Tymon Słoczyński and two anonymous referees for insightful suggestions. We are also grateful to the participants of the seminars at the University of Warsaw. The remaining errors are ours. The support of National Center for Science (grant UMO-2012/05/E/HS4/01510) is gratefully acknowledged. The views expressed in the paper are not to be associated with those of the National Bank of Poland.

*Correspondence to: Lucas Van Der Velde, University of Poland, Warsaw, Poland (lvelde@wne.uw.edu.pl).

applied if the characteristics diverge across genders; they cannot measure the differences outside the mean; and they cannot correct for selection into employment. Either of the problems may generate a significant bias in the results. The lack of an adequate treatment of the common support and the selection into employment may produce a bias in the estimation of the gaps of the mean; while overlooking the differences at different quantiles might produce an inaccurate picture of the gap in the population. The last two decades brought about the expansion of the available toolset with the objective to address one or more of the three problems. Starting with the Juhn, Murphy and Pierce (1993) decomposition a new wave of techniques was developed. The new methods attempt to address many weaknesses associated with traditional parametric approach. One strand of the literature goes beyond the analyses of mean wages, employing quintile and sampling methods to be able to estimate adjusted wage gaps along the distribution. Another strand focuses on assuring comparability, by implementing what is referred to as the common support condition. Clearly, each of the methods provides not only econometric advancement, but refines also the way adjusted gender wage gap can be interpreted for policy purposes.

Though the proliferation of methods is welcome from a methodological perspective, it also introduces confusion from a practitioner's perspective. Are results susceptible to a method? How do the estimates of the gaps compare to each other? These questions had attracted the attention of other researchers in the past. Weichselbaumer and Winter-Ebmer (2005) emphasize, for example, that nearly a half of the estimates present in the literature is likely biased due to inaccurate use of econometrics. Fortín *et al.* (2011) discuss the available methods and present a uniform analysis using various estimation techniques for the same data set. This empirical exercise reflects a strong weight on the theoretical differences between the available methods. Our contribution lies in-between these two papers. Following Fortín *et al.* (2011), we use only one database, in this case Polish LFS for 2012, and one definition of the variables to ensure maximum comparability between the estimates. Like Weichselbaumer and Winter-Ebmer (2005), we are interested in how the changes in the specification affect the estimates of the adjusted wage gap in each of the methods and how they can be compared to each other. First, we compare the available methods on one, simple specification, which allows us to see how alternative estimates differ in terms of size. Second, we extend the set of explanatory variables for each method and analyze the changes to the estimated adjusted gender wage gaps.

This comparative exercise is performed for Poland, a country which passed two decades of economic transition from a centrally planned to a market economy, but which is characterized by low female labor force participation and an adjusted gender wage gap in excess of the raw wage gap. While we do not discuss access to labor market, nor to the professions, we analyze comprehensively current extent of the gender wage gap unexplained by the differences in endowments.

The paper is structured as follows. First, we present the decomposition techniques used in the literature analyzing their interpretational advantages and disadvantages. Second, we briefly discuss the earlier work on gender wage gap in Poland. We employ data from the Polish Labor Force Survey, which are described in the third section. The results are discussed synthetically in the subsequent session. The final conclusion derives the policy recommendations and suggests potential avenues for further research in this field.

2. Different Methodological Approaches to Measuring Gender Wage Gap

The simplest way to estimate the differences between genders is to use a Mincer equation for wages expanded with a gender dummy. Usually, Mincer equations are estimated in logarithms to reduce the problems due to the skewness of the wage distribution. An example of policy application can be found in Watson (2010), who uses the "simulated change approach"[1] to decompose the differences in wages due to different factors.

However, linear regressions have several long-known shortcomings. Namely, they disregard potential gender differences in rewards for characteristics. In the presence of unobservable effects (correlated with the error term) coupled with sample selection, linear regressions constitute a strongly biased estimator, and the direction of bias cannot be known *a priori*. To account for a different way, the same characteristics are evaluated for men and women, interaction variables with the gender dummy can be included. In practice this procedure is equivalent to the estimation of two separate equations, and then decomposing the absolute differences in wages into a component attributable to differences in characteristics and a component that cannot be explained by these objective differences. The latter is conceptualized as the adjusted gender wage gap, often identified with discrimination.

Indeed, the seminal decomposition proposed by Oaxaca (1973) and Blinder (1973) separates differences in wages across genders into differences due to the characteristics and the differences due to the rewards. The first step is to estimate two separate regressions for each gender. Then the size of the pay difference can be decomposed as follows:

$$W_m - W_f = \left(X_m - X_f\right)\beta^* + X_m(\beta_m - \beta^*) + X_f\left(\beta^* - \beta_f\right)$$

The first term in RHS represents differences in characteristics between males and females. The second and third term in the RHS represent the differences in coefficients, with no assumptions concerning the "true" wages in the absence of discrimination effects. In this formulation, the second term of RHS represents the male (dis)advantage while the third one represents the female (dis)advantage.

Assigning values to β^* necessitates an assumption about the structure of the reference wages. A debate arose with respect to the interpretation of these coefficients. Traditionally, they were interpreted as the non-discriminatory wage structure; that is as the wage structure that would prevail if men and women were paid on an equal basis. However, Słoczyński (2013) shows, in connection to the treatment literature, that the choice of reference wages in fact corresponds to observing the average treatment effect on untreated (male, with $\beta^* = \beta^M$), average treatment

[1]Watson (2010) obtains coefficients from a pooled regression for wages and estimates the means of the variables of interest for men and women. For this subset of variables of interest Watson (2010) calculates the gender differences. Once they are multiplied by the differences in means they can be used to obtain a percentage contribution of each characteristic to the overall gap.

TABLE 1
Literature Approaches to Determining β^*

$\lambda = 1$	Male coefficients are taken as a reference
$\lambda = 0$	Female coefficients are taken as a reference
$\lambda = 0.5$	Simple average of both, proposed by Reimers (1983)
$\lambda = \%$ *male*	Each coefficients weighted by the proportion of the same gender, proposed by Cotton (1988)
$\lambda = \%$ *female*	Each coefficient weighted by the proportion of the opposite gender, proposed by Słoczyński (2013)
$\beta^* =$ *pooled*	The coefficients from a pooled regression, with gender dummy (Fortin, 2008) or without it (Neumark, 1988)

effect on treated (female, with $\beta^* = \beta^F$) or total treatment effect (with a linear combination of both). Indeed, variations of β^* are different linear combinations of the coefficients from the regressions made on males and females. The general structure is given by the following equation, where the variants depend on the value assumed by lambda and are summarized in Table 1,

$$\beta^* = \lambda \cdot \beta^M + (1 - \lambda) \cdot \beta^F$$

While the first three are intuitive, the values in the bottom half of the table deserve some explanation. The differences between Cotton (1988) and Słoczyński (2013) approaches stem from a different understanding of the coefficients. While Cotton (1988) proposed a weighted average of the coefficients, Słoczyński (2013) indicated that the weights should be applied not to the coefficients themselves but to the treatment effects, $X_m(\beta_m - \beta^*)$ He showed that this rather counterintuitive structure (using the opposite gender as a weight) is equivalent to taking a weighted average of the treatment effect on the treated (female) and on the untreated (male), where the weights are given by the percentage of population of each gender in the sample.

These parametric techniques all share three similar shortcomings. First, this approach implies that the average wage gap may be estimated for men and women whose characteristics are starkly different. Second, this approach only looks at average difference between male and female compensations. Third, the selection bias problem is neglected. In the remainder of this section we discuss how the literature has so far developed to address these shortcomings.[2] We discuss the severity of these issues and the ways to address them in the subsequent sections.

2.1. *The Problem of Common Characteristics*

Parametric Oaxaca-Blinder decomposition and its variations do not take into consideration the interplay between characteristics and how they are priced. In particular, if certain characteristics are relatively more abundant among one gender than among other, their price is likely to reflect the abundance along with the market valuation. For example, higher compensation to miners is confused

[2]For the remainder of the literature review we only refer to the literature on the estimation of the (continuous) wages. We thus leave aside contributions concerning non-standard measures of wages (e.g. dichotomous or discrete indicators of wages), e.g. Fairlie (2005) or Bauer and Sinning (2008).

with lower wages of women in general if there are few (or no) female miners. This bias could not be accounted for by the inclusion of the industry and occupational controls because data on wages for female miners is unavailable, which biases the average adjusted gender wage gap measure.

This problem—often referred to as the common support problem—is neglected by the methods discussed above, whereas the severity of this issue might be large regardless of the choice of baseline β^*. If men and women are not strictly comparable or if these comparisons make no sense, reliance on estimated βs to evaluate the gender wage gap is particularly misleading. Think, for instance, in the case of women with lower work experience due to maternity leave, *ceteris paribus*. Men of a similar age have longer work experience, which makes the two groups incomparable. In OLS based estimations, the lack of common support implies that the out-of-the sample prediction is biased.[3]

A simple solution to this issue is to estimate the gender gap only on the observations where the characteristics of men and women are comparable, i.e. within the common support. In order to determine the common support, authors relied on nonparametric techniques. Barsky *et al.* (2002) propose to use a reweighting equation, in which the weights attached to every observation are the ratio of probabilities of finding an individual with a given level of income in each of the two groups. An alternative weighting scheme, proposed by Black *et al.* (2008) assigns a weight of zero to unmatched observations, and a weight $P(x)/[1 - P(X)]$ to the matched observations, where $P(X)$ is the probability of finding an individual of the disadvantaged group with the same characteristics. Both weighting schemes require an exact matching between members of both groups prior to the estimations. An important disadvantage of these techniques is that conclusions are only valid inside the common support, and may not be representative of the gap for the entire population. The second broader critique is that the matching cannot ensure that we are in fact looking at similar people. If women experienced more self-selection into employment, which means that employed women might have more unobserved skills than the matched males, then the estimates obtained from matching techniques can only represent a lower boundary of the true adjusted wage gap.

An alternative non-parametric approach employing a matching estimator was proposed by Ñopo (2008). While this method does not solve the self-selection issue, it does not disregard the observations outside the common support in the analysis. Implicitly, these techniques assume that the wage distribution function can be divided in two sections: one where the characteristics of the members of the two groups coincide (common support) and another which represents the deviations of each group from the common support. The logic for this decomposition is that only within the common support, the characteristics of males might be rewarded differently than those of females, so only this part of the wage differential can be explained by *discrimination*. Such approach disregards the inequality of

[3]This problem also involves another debate from the literature on which variables should be included in explanatory vectors. Inclusion of industry and occupational controls is justifiable in the absence of "discrimination" in access to some jobs. Non-random absence of one of the genders in some occupations may in fact be endogenous, introducing additional bias to the estimates, as discussed by Huber (2014).

opportunities. Women may be absent in certain professions due to constrained access, and despite the lack of differences in characteristics. However, such women are still treated as uncomparable to men. In such cases men may earn discriminatorily high wages, but these wages are not included in the computation of adjusted gender wage gap.

Black *et al.* (2008) uses exact matching, but both male and female non-matched observations are discarded, implying that an important part of the information is removed from the data set. To avoid this efficiency loss, Ñopo (2008) develops a decomposition which identifies the part of the wage gap that is attributable to the characteristics of the "unmatched" men and the part of the gap that is attributable to the characteristics of the "unmatched" women. In fact, unlike Black *et al.* (2008), Ñopo (2008) constructs a counterfactual population of women and the rewards from men.[4] This is done by sampling each woman and matching her to all statistically identical men. In this way, a synthetic counterfactual female wage observation is created, which equals the average wage of the matched men. Ñopo (2008) uses the information on the unmatched men (those who were not identical to any of the women in the sample) and unmatched women (those for whom a match among men could not be found) to construct the following decomposition:

$$\Delta \equiv E[Y|M] - E[Y|F] = \Delta M + \Delta F + \Delta X + \Delta 0$$

where $\Delta O = (E_{F,matched}[Y|M] - E_{F,matched}[Y|F])$ is the part due to differences in unobservable characteristics or discrimination, whereas $\Delta X = (E_{M,matched}[Y|M] - E_{F,matched}[Y|M])$ is the part due to differences in observable characteristics within the common support. Additionally, $\Delta M = \mu^M(E_{M,unmatched}[Y|M] - E_{M,matched}[Y|M])$, where μ^M is the probability of men being not matched, is the component due to men out of common support. Finally, $\Delta M = \mu^F(E_{F,matched}[Y|F] - E_{F,unmatched}[Y|F])$ is the part of the gender gap which can be explained by unmatched women having different endowments than matched women.

However, Ñopo (2008) decomposition has some important shortcomings. First, the exact matching implies a trade-off between the number of characteristics to control for and the ratio of matched and unmatched observations for both men and women: more reliable estimates of the adjusted gender wage gap are obtained for a fraction of the sample (a problem known as "dimensionality curse"). Second, counterfactual distribution of salaries based on means is probably biased if the overall distribution is skewed. Also, it precludes the use of the information on differences in wage dispersion between men and women with the same characteristics.[5] Third, sampling is done over the entire distribution, which makes it challenging to analyze different gender wage gaps along the wage distribution.[6]

[4]Standard errors are obtained via bootstrapping in Black *et al.* (2008) and are derived analytically in Ñopo (2008).

[5]Shorrocks (2013) proposes a non-parametric method employing Shapley value in a matching framework which allows analyzing the impact of a given variable in the results of some analysis of inequalities.

[6]This method is also useless if wages are not continuous (e.g. coded within bands). Another problem is that Ñopo (2008) decomposition does not create a continuous counterfactual distribution of wages for women. The distribution is full of "jumps" which reflects the changes in the cell of reference for women.

2.2. *The Problem of the Uninformative Mean*

Adjusted gender wage gap estimated at an average may be uninformative if there are large discrepancies in gaps depending on profession and/or wage level. In an extreme scenario, gender wage gap could average to zero if high income women were overpaid and low income ones were heavily underpaid. Notice that breaking the earning distribution into several bands not only underutilizes an important amount of information, but it is also plainly inconsistent.

The first alternative to deal with the decomposition at different quartiles was proposed by Juhn, Murphy and Pierce (1993) and later developed by Blau and Kahn (1996). This method is parametric, as it requires an estimation of a Mincerian wage regression. Coefficients from the advantaged (male) group regression are used to obtain a counterfactual wages distribution for the disadvantaged (female) group:

$$Y^j = a^m + \beta_i^m x_i^j + \theta^j \sigma^m, \quad j = f, m$$

where σ^m represents the standard deviation of the error term in the male regression, while θ^j is a standard error term. In the case of the male equation, $\theta \sim N(0, 1)$. In the case of the female equation θ is the difference between the actual value of wages for women, and that predicted by using the male coefficients and the female characteristics (for interpretational convenience, both terms are divided by the standard deviation from male equation). Thus, the adjusted wage gap is defined as:

$$D = Y^m - Y^f = \left(X^m - X^f\right) \cdot \beta^m + \left(\theta^m - \theta^f\right)\sigma^m$$

The first term represents the differences in observable characteristics, while the second term represents the unobservable differences between men and women, with σ^m interpreted as the price of the unobservable characteristics. This equation can be used to estimate the differences at quantiles.[7] The main difference between Juhn, Murphy and Pierce (1993) and the decomposition based on Oaxaca-Blinder is that in the former Xs and θs do not represent the values at the mean, but rather at any given quantile.

However, the method has several drawbacks. First, the estimations are still done at the mean, which implies that the individual characteristics are rewarded the same way along the entire earnings distribution. Second, the method implicitly assumes conditional rank preservation, i.e. (ordered) residuals follow the same pattern in both male and female distributions. This assumption is hard to test, but one should also rarely expect it to hold in practice, as for example the residuals may reflect problems with the method and not just with unobserved effort. Additionally, it is difficult to rank the residuals when there are more observations in one of the groups.

Some of these issues are addressed by Machado and Mata (2005), which involves simulating a population of the disadvantaged (female) group with the

[7]An example of this is Zhang *et al*. (2008) and Cho and Cho (2011). Moral-Arce and Sperlich (2008) provide an example of the flexibility of the JMP as they incorporate some non-parametric function among the covariates.

rewards of the advantaged (male) group.[8] With subsequent simulations, total gender gap at each given quantile (q) is given by:

$$x^m \beta^m(q) - x^f \beta^f(q) = \left(x^m - x^f\right)\beta^f(q) + x^m\left(\beta^m(q) - \beta^f(q)\right)$$

The quality of Machado and Mata (2005) decomposition relies on the number of simulations, but the model also suffers from path dependence. More importantly, this technique effectively estimates a Mincerian wage regression for each simulation, which brings about all the problems associated with sample selection and functional form. It also assumes an identical functional form at each quantile. Finally, because the final outcome is a result of many simulations, there is no easy way to attribute the "explained" part of the adjusted wage gap to particular explanatory variables (i.e. individual and firm level characteristics), which makes it less useful for policy recommendations. Albrecht *et al.* (2009) allow the selection correction at every quintile. Yet, the arbitrary (and untestable) choice of the functional form in the original Mincerian regression still may imply that the choice of functional form affects the results in an intractable way.

An alternative to simulation techniques are again non-parametric estimators, such as the one proposed by DiNardo *et al.* (1996). They suggested using a counterfactual density of wages based on the available data (e.g. wages that would have prevailed in a given year if the characteristics were the same as those prevailing in another year, country, group, etc.). The general procedure is based on the premise that the structure of wages can be decomposed in two fairly independent parts: a structure of premiums awarded to the individual characteristics and a structure of these characteristics. Given this assumption, the counterfactual conditional distribution can be obtained *via* a reweighting procedure through which the attributes obtained in group, country or period i are converted into those in group, country or period $j \neq i$. According to DiNardo *et al.* (1996), weights should be calculated using a probit model, where the dependent variable is the pertinence to the treatment group.[9]

Firpo *et al.* (2009) proposed a technique that enables the impact of particular covariate to separate on the explained and unexplained part of the gap for any quantile of the unconditional distribution of dependent variable. This method has important advantages in comparison to Machado and Mata (2005) as it provides detailed decomposition not only of wage structure effect, but preserving path independence. Thus this tool allows results to be obtained for any distributional statistic with desirable features similar to Oaxaca-Blinder

[8]The efficiency issue was addressed by Melly (2006), who also contributes by deriving the variance of the Machado and Mata (2005) estimator.

[9]The choice of the shape of the distribution is fairly irrelevant, but the size of the bin for density function approximation has important consequences. In addition, in principle the distributions should be defined on exactly the same domain, which effectively requires common support condition to hold. Donald, Green and Paarsch (2000) offer an alternative weighting procedure to obtain the counterfactual distribution, which builds on hazard function models. It is resilient to the dimensionality curse and is less sensitive to big masses in one point of the distribution, as would be expected given minimum wage laws or top-coding procedures. A downside is that due to its construction, the distribution of wages tends to present more spikes, and resemble more a histogram than a smooth kernel distribution.

for the mean.[10] This method—referred to as re-centered influence functions, RIF—is most frequently used to obtain results of unconditional quantile regressions.

2.3. *The Problem of Selection Bias*

Common to most Mincerian estimations of wage equation is the issue of sample selection. Namely, if data on wages is only available for some non-random subsample of individuals, bias is likely to emerge for the characteristics which drive both the likelihood of working (i.e. availability of wage data) and productivity (i.e. particular value of wage). If we rely on the estimated parameters to compute the adjusted wage, sample selection bias undermines reliability of this approach.

The most common method to solve this problem is to employ the Heckman (1979) procedure,[11] which relies on the idea, that self-selection bias can be treated as an omitted variable problem and solved by recovering that variation from the available data. Typically, one uses determinants of labor force participation, which should not have a direct influence on wages. Such candidates are marital status, household structure or availability of non-earned income within the household. These variables are used as instruments in the first stage probit regression of employment, which delivers a correction term for the second stage wage regression. In the context of empirical application of the Oaxaca-Blinder decomposition, the sample selection correction is traditionally done only for women, as they are considered to be selected out due to the gender status and the household role division. Given the large and growing size of the so-called NEETs (Not in Employment, Education or Training) across both genders in industrialized countries, this assumption is not likely to match well the data.

Recently, authors tried to go beyond the Heckman procedure for selection. An example is Machado (2012), who proposes an extension allowing for differences in the selection process at different levels of the income distribution, for both men and women. This alternative consists of dividing the population into several groups depending on whether and how they have changed their "decision" to be employed when some circumstances (the instruments) changed. These instruments are subsequently incorporated in the wage regression. The disadvantage of this method is that it requires a panel data for the estimation, which is usually unavailable.

Recognizing that employment is a matter of (constrained or unconstrained) choice, one needs to acknowledge that also the occupation, industry and the form or employment are (at least partially) endogenous. Omitting them from a gender wage gap analysis makes the estimate of adjusted wage gap flawed, whereas their inclusion makes the parametric estimates unreliable. The solution proposed by Brown *et al.* (1980) and Appleton *et al.* (1999) is a modification of the Heckman procedure only that instead of using a probit model in the first stage, a multinomial logit is preferred. Similarly to the Neumark decomposition (1988), Appleton *et al.*

[10]However, it lacks a clear control for the self-selection into employment, thus the results do not have features of OB decomposition with Heckman correction.

[11]Most common but not the only. Alternative approaches include the imputation of wages to the non-labor market participants based on their characteristics, or limiting the estimation of the wages equations to a subset of the population where the prevalence of unemployment is low, and hence the self-selection problem can be ignored.

assume the existence of a non-discriminatory sectoral structure and evaluate the impact of having different selection probabilities for each gender.[12] Clearly, one downside of these methods is that women are underrepresented in some professions along the entire wage distribution, not just in high paid jobs. Equally important, this method cannot account for the differences in wages to each of the occupations, which leaves glass ceiling and sticky floor unaccounted for.

An alternative for dealing with the selection bias was proposed by Olivetti and Petrongolo (2008) who estimate the gender wage gap in potential wages. Their approach includes the imputation of potential wages to unemployed workers, for which they use several techniques. In this way, the problem of unobservable wages is corrected and there is no selection bias, other than that arising from the imputation. In order to minimize its effects, the authors use regressions at the median, instead of the traditional OLS. In these regressions, the size of the imputed wage is irrelevant, only its location with respect to the median matters. This provides some extra robustness to their findings although it also highlights the difficulties of using a similar approach for other quantiles of the distribution. Additionally, the method is only used to compute the raw gaps. Also, the bias is likely to be larger in the countries where the participation gaps are larger.

The original ideas of Donald *et al.* (2000) were recently updated in the work of Picchio and Mussida (2011). Using an approach similar to Albrecht *et al.* (2009) in Machado and Mata (2005) decomposition, Picchio and Mussida (2011) amend the hazard function approach with selection into employment, by imputing wages to non-workers, or to be more accurate to assign the unemployed to their corresponding wage-ranges. In this way, they bypass the problem of selection into employment, as will be presented below. Their method shares some of the strengths and downsides of the Machado-Mata: it provides a description of the gap along the distribution and it provides only the total effect of all the coefficients. On the positive side, it is less intensive in terms of calculations and it imposes fewer restrictions on the relation between the individual effects that lead to sample selection and that have an impact on wage determination; this result is derived from panel properties and may not be extrapolated to cross-sectional data.

2.4. *Summary of Different Methods*

Given this brief review of the literature, it is clear that a wide range of methods for computing gender wage gap are available to the researcher, but at the same time that none of them is perfect. In fact, the standard parametric approaches, such as Oaxaca-Blinder decomposition, have several shortcomings, but the subsequent econometric developments address only some of them at a time. The perfect method needs to (1) address selection issues, (2) compare wages within the common support accounting for the differences in wages and characteristics of the "unmatched" men and women, and finally (3) allow to account for wage differences at different percentiles of the earnings distribution.

[12]Brown *et al.* (1980) method is actually simpler as only one multinomial logit is estimated (for men) and then it is used to create a counterfactual probability distribution for women (applying the coefficients to the mean). In fact it employs the unconditional probabilities for men and women, but the conditional probabilities for the counterfactual distribution.

On one hand, methods based on matching—especially Ñopo (2008)—allow addressing the problem of common support. By adequately comparing men and women with this method one is able to specifically identify the role of characteristics and the role of "unexplained" components. This method is also immune to the selection issues, but the causal interpretation needs to be careful. The drawback of Ñopo (2008) is that distributional analysis is not effectively possible. There is also a path dependency problem, i.e. the contribution of each variable depends on the removal order, which constrains the extent of policy relevance.

On the other hand, methods relying on regression and sampling allow accounting for selection issues and keeping the power to deliver an analysis along the income distribution, but have difficulty in assuring that the comparison only concerns individuals with comparable characteristics. Namely, reweighting is used to balance potential under- or over-representation of one group. However, reweighting does not provide informational content and has limited reliability if weight within one of the groups is strictly (or close to) zero. Also, they make extensive use of the "error term" in interpretation.

3. Gender Wage Gap in Poland—Previous Studies

The literature on gender wage gap in Poland is not vast. Moreover, the gender wage gap has been analyzed mostly in the context of transition period as performed by Grajek (2003) or Adamchik and Bedi (2003). While nearly all wage related studies control for gender, few go beyond this simple understanding of gender differences and employ the decomposition techniques described before.[13]

Kot *et al.* (1999) and Adamchik and Bedi (2003) used Oaxaca-Blinder decompositions to analyze the gender wage gap during the 1990s. Their findings indicate that while the explained component changed across methods, it still represented only a small part of the total gap. Similar results were obtained by Grajek (2003) and Łatuszyński and Woźny (2008) who applied the Juhn, Murphy and Pierce (1993) decomposition to the analysis of Polish employees from the Household Budget Survey. Additionally, Grajek found that the explained component increased over the analyzed period (1987–1996).[14] In contrast, Goraus and Tyrowicz (2014a) analyzed the gender wage gap using Ñopo's (2008) non-parametric decomposition. Employing quarterly data from the Polish Labor Force Survey over 1995–2010, they found a fairly stable adjusted gender wage gap of approximately 20 percent of the average female wage, which doubles the raw wage gap.

In addition to the analysis of the gender wage gap over transition, some papers focused on particular aspects. Magda and Szydłowski (2008) as well as Matysiak *et al.* (2010) provided parametric decompositions, focusing on the life

[13]Most previous studies on the gender wage gap in Poland either focused on the raw wage gap or estimated a linear wage regression with a gender dummy. Examples of such studies include Kotowska and Podogrodzka (1994), Kalinowska-Nawrotek (2005), Zwiech (2005) or Mazur-Łuczak (2010). Poland was also included in a number of cross-country studies, such as Brainerd (2000), Pailhé (2000), Blau and Kahn (2003), Newell and Reilly (2001) as well as Nõpo *et al.* (2012). Without exceptions, all studies find lower wages for women in Poland along with better characteristics, such as higher educational attainment.

[14]Łatuszyński and Woźny (2008) used data from 2004.

cycle aspects. Słoczyński (2013) employed an innovative technique of population average gender effects to analyze regional differentiation of the adjusted gender wage gap, reaching a similar conclusion. Finally, Rokicka and Ruzik (2010) analyzed the gender wage gap in Poland using Melly (2006) decomposition. They found a larger adjusted gender wage gap in informal employment. More importantly, the differences were larger at the bottom of informal sector earnings and in the top of the distribution for the formal sector employees. However, their analysis relies on an unrepeated and unrepresentative survey focused on informal employment, which limits the external validity of their findings.

To sum up, empirical evidence is consistent: gender wage gap is a general phenomenon in Poland, visible in both the raw and the adjusted components. Fairly high estimates of the adjusted gender wage gap, relative to the raw gender wage gap are not an extreme case. Goraus and Tyrowicz (2014b) showed that, on average, when compared to other European advanced and transition economies, the correlation between the raw gender wage gap and the adjusted gender wage gap is similar for Poland—it is only the constant that is somewhat higher.[15] Compared to the previous studies on Poland, our paper contributes in three major ways. First, we provide a comparison of the estimated adjusted gender wage gap for various methods. Second, we employ a rich data set, which permits controlling for a large number of observable characteristics as well as for the selection bias. Finally, we provide estimates of adjusted gender wage gap on a large and representative sample of nearly 250,000 individuals for a recent period, thus filling the gap between the studies from late 1990s and early 2000s and the current times.

4. Data and Method

This paper uses recent information on the Polish labor market, corresponding to the four rounds of the Polish LFS of 2012. The total sample consisted of 296,427 individuals between the age of 18 and the retirement age (60 for women and 65 for men). Polish LFS has no information on hours worked or compensation received by helping family members (as well as the self-employed), so these individuals could not be included in the study. Reflecting the situation in the labor market, the sample is evenly split between men and women—the latter represent approximately 50.6 percent of the total observations.

4.1. *Variables Definitions*

All variables were constructed following the standard measures. The dependent variable is hourly wages (taken in logarithm). We obtained hourly wage by dividing the monthly self-reported wage by the self-reported number of hours worked (i.e. the number of hours worked on average during the week times four).

[15]Goraus and Tyrowicz (2014a) used Ñopo (2008) matching for a coherent set of control variables to obtain comparable estimates of the adjusted gender wage gap—comparable both across years and countries. Clearly, if the mechanics behind the emergence of the gender wage gap differ across countries, the analysis of Goraus and Tyrowicz (2014a) is inconclusive on whether Poland is an outlier, but it gives a tentative guidance.

The database contains standard demographic variables for all individuals, which in addition to gender include age (in years) and marital status (in relationship, single, widowed, divorced/separated). We can also identify whether the place of residence is a rural area, a large city or neither. All analyses show that both the capital region (Mazovia) and large cities (above 50,000 inhabitants) are characterized by consistently higher wages. These variables were thus indispensable for the study.

We capture human capital by the measures of educational attainment. We use three levels: primary education or less, tertiary or above and a common group for secondary, secondary vocational and vocational education. In addition, the data set is rich enough to contain information on the actual field of education.[16] The database contains also declared overall work experience and tenure with the current employer. Both these measures are self-reported and measured in months (integers), which we recoded to years (natural numbers). In addition, ISCO coding of occupations is available, and we use it for analysis at one digit level of disaggregation.

The dataset is also relatively rich on employers' characteristics. We are able to identify the size of the firm in which an individual is employed. In addition, the data cover industry where it pursues its activities. The original variable has 11 levels, following NACE categorization. However, as a robustness check, we have also grouped these categories into broader ones: agriculture, construction, manufacturing and services. In addition to industry, the data set also contains information on the public or private ownership.

The labor market status is identifiable directly, based on self-reported indication of employment or unemployment. As is standard in LFS type data, individuals are asked if they have worked for at least 1 hour in the week preceding the survey. If they have not, they are asked about willingness to undertake employment and active search.[17] Only if the respondent is non-employed and is seeking an employment, we consider such persons as unemployed. Otherwise, individuals are characterized as inactive. Since wage data is missing for the non-employed individuals, correction for selection into employment is needed.

4.2. *Selection Correction—Identifying Restrictions*

As discussed earlier, the bias reduction by the means of the Heckman correction hinges on the power of the instruments to correctly predict wage employment. In this study, we use the available data on the structure of the household and its sources of income as the identifying variables. First, we can employ data on the presence of children and small children (younger than five years old) in the household. They tend to be closely correlated, so we evaluate the results with both variables. The likelihood-ratio and sensitivity tests favored the inclusion of

[16]There are nine categorical variables for the field of education: pedagogy and teaching; human sciences (including art and languages); social sciences (which includes law and economics); natural sciences, mathematics and computer science; engineering; agricultural sciences and veterinary; medicine; services; and others.

[17]In fact, those who have not worked are subsequently asked a follow up question if lack of work is associated with holidays, sickness, strike etc. Only if not, the questions leading to determining the unemployment status are asked.

children per se rather than small children in the selection equation, so we chose the latter, see Table A1. Second, LFS provides information on the presence of other sources of income in the household. These sources include: income from retirement benefits, earned income by other household member and unearned income from social benefits received by other household members. Intuitively, these variables are likely to affect the opportunity cost of working, and thus the labor supply decision. These variables are not related to individual productivity. Also, they are only weakly correlated, thus all of them can be used simultaneously in the selection regression.[18]

It is not clear whether selection into employment should be computed on active population or on total working age population. The original Heckman (1979) idea pointed to the unobservable latent variable of reservation wage, i.e. assumed only correction of working from active is needed. However, currently the unemployment status tends to be relatively labile among the non-working, i.e. many of the inactive could actually work should the opportunity arrive, even if currently they do not actively seek employment. Thus, one could also try to correct for selection using the entire working age population. Since the objective of this paper is to quantify the effects of modelling choices on the measured adjusted gender wage gap, we employ both approaches and compare the results throughout the specifications.

4.3. *Descriptive Statistics*

Table 2 provides the descriptive statistics for our data. In fact, for nearly all characteristics, the difference between men and women is non-zero in a statistically significant way (although the economic relevance of that difference may be low in many cases). Male workers earn on average approximately 25 percent more than their female counterparts (using male wage as a reference), they also work slightly more hours a week. Hourly wages are also higher for men, though the difference is much smaller, approximately 9 percent of male wages. Female workers have better educational backgrounds, longer tenure but slightly shorter experience. We also observe some sorting of men and women across occupations, industries, sectors and fields of study.

We formally test if the distributions of characteristics of the two genders overlap. This test was proposed by Imbens and Rubin (2009). A rule of thumb for interpreting this index is that values over 0.25 should be a source of concern, as

[18]The validity of these instruments is a separate topic of discussion. While the two sets of variables were widely used in the literature and bear a clear conceptual relation to the labor force participation, they might not be perfectly exogenous to wages. For example, household income sources may shape both reservation wage and individual motivation, thus also affecting productivity (in a way directly unobservable to the employer, though). Similarly, individuals more engaged in childbearing may be less engaged in labor market activities for related reasons. To overcome these difficulties, Morawski and Myck (2010) proposed to use unearned income implied by the tax-benefit system as an instrument (regardless of whether a household reports it); however, this variable is just an outcome of household characteristics. While for researchers focused on selection issues seeking new instruments constitutes an avenue for further research, for most analysts and researchers focused on other topics the readily available measures such as children in the household and income earned or received by other household members constitute an attractive set of instruments to be included in the selection regression. In the remainder of the study, we quantify the effect of using traditional measures.

TABLE 2
Descriptive Statistics

	Overall	Male	Female	t-stat	Support
% active		69.6	58.2	62.693***	
Among active, % employed		87.7	83.9	22.7***	
Average monthly wage (in PLN)	1801.7	1951.6	1642.8	37.57***	0.29
Weekly hours	39.9	41.4	38.3	41.18***	0.32
Average hourly wage	11.5	11.9	11.0	15.14***	0.11
Age (in years)	40.9	40.6	41.2	–5.41***	0.06
Married (%)	68.1	68.9	67.2	3.32***	0.03
Primary education (%)	7.6	9.3	5.8	12.2***	0.09
Secondary education (%)	69.2	74.9	63	23.7***	0.18
Tertiary education (%)	23.1	15.8	31.2	–34.0***	0.26
Experience (in years)	18.6	19.06	18.15	7.00***	0.05
Tenure (in years)	9.5	9.18	9.91	–6.93***	0.05
Residence in rural areas (<2 ths., %)	40.5	44.3	36.3	14.85***	0.11
Residence in small cities (2 ths to 50 ths, %)	28.0	26.7	29.4	–5.40***	0.04
Residence in large cities (>50 ths., %)	31.5	29.0	34.3	–10.44***	0.08
Residence in Mazovia (%)	10.3	10.0	10.6	–1.71*	0.01
Second earner (% of individuals)	91.5	90.7	92.4	–5.37***	0.04
Children in household (% of individuals)	18.5	21.3	15.3	14.03***	0.10
Pension in the household (% of individuals)	4.5	4.7	4.2	2.20***	0.02
Market services (%)	39.0	35.0	43.5	–16.05***	0.12
Non-market services (%)	20.4	8.0	34.2	–62.91***	0.48
Construction (%)	8.3	15.1	0.9	48.59***	0.38
Manufacturing (%)	31.5	40.9	21.0	40.20***	0.31
Agriculture (%)	0.8	1.0	0.4	6.32***	0.04
Private firm ownership (%)	67.7	76.4	58.0	36.69***	0.28
Firm size (<50 employees) (%)	50. 0	47.3	53.1	–10.62***	0.08
Firm size (50 to 250 employees) (%)	31.9	33.1	30.6	4.96***	0.04
Firm size (>250 employees) (%)	18.0	20.8	16.3	7.78***	0.06
High skills (ISCO 1–3, %)	29.8	20.8	39.8	–38.72***	0.30
Low skills (ISCO 4–8, %)	70.2	79.2	60.2	47.62***	0.37
Arts and humanities (%)	2.1	1.0	3.3	–14.38***	0.11
Engineering and construction (%)	41.6	61.4	19.7	85.44***	0.67
Social sciences and law (%)	17.5	7.5	28.6	–53.05***	0.40
Medicine (%)	4.5	0.8	8.5	–34.48***	0.26
All other fields (%)	34.3	29.3	39.9	–20.69***	0.15

Note: ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. Numbers refer to the working age population, fractions relate to the employed population. Other fields of education include agriculture, veterinary, services and general programs. The last column (Support) provides a test for (the lack of) common support between men and women. It follows Imbens and Rubin (2009), and is constructed as the absolute value of the ratio of the difference in the means to the square root of the sum of the variances. It measures to what extent the distributions of men and women overlap for a given characteristic.

Source: own calculation, Polish LFS, four quarters of 2012.

they indicate important differences in the location of the distribution of covariates (the distributions do not overlap enough). We observe values above the threshold in the case of industries, occupations and fields of study. For the amount of hours worked, the tertiary education and the sector are also above the limit, though by a smaller margin. These results highlight the importance of using methods that control for the common support.

In principle, average wage differential may reflect both a shift between the distributions and differences in the shape of the wage distributions. To identify which of the two effects prevails in Polish LFS data, kernel distributions of hourly wage were obtained and plotted in Graph A1. The graph shows the percentage of workers (densities) at every point of the income distribution for hourly wages. The graph shows that women's salary distribution is more condensed around relatively lower wages than the corresponding distribution for men. Moreover, the distribution of men is farther to the right. The spikes observed in the graph reflect the fact that people tend to report round numbers for salaries and hours worked, a problem common to survey self-reported wage declarations.

However, the difference in the distributions does not need to be indicative of an adjusted gender wage gap—if female characteristics were more concentrated in those areas of the labor market which offer lower payoffs (e.g. education level or fields, industry, occupation, etc.), then that difference would simply reflect disparities in observables. The remainder of this section discusses the approach followed to quantify the extent to which the difference in wage distributions cannot be attributed to the observable characteristics.

4.4. *Method*

We compare the results from a regression with a gender dummy, Oaxaca-Blinder decompositions, Juhn, Murphy and Pierce (1993) decompositions, Machado-Mata (2005) decompositions, re-centered influence function approach by Firpo *et al.* (2009) as well as Ñopo (2008) decomposition. In each case we perform two analyses: one for the total sample and the other one for the sample constrained to the common support, as derived from Ñopo (2008). This selection implied leaving out some methods, which remain rarely used.[19] In total, we use seven decomposition methods. With the obvious exception of Ñopo (2008) decomposition, each set of six parametric estimations (one set for total sample and one set within common support) is repeated with and without selection correction.

In addition to the parametric selection correction based on a probit regression, we employ the semi-parametric single index method, as proposed by Buchinsky in the context of quantile regressions. The procedure was already implemented in wage regression by Albrecht *et al.* (2009), where the authors controlled for selection into full-time employment. We follow these authors and calculate the single index, and use a second order polynomial on the Mills ratio. However, we estimate the single index in a slightly different fashion as we used the alternative suggested by Klein and Spady (1993). The single index method calculates the probabilities based on probit models with a second stage, where the estimates are fitted locally using Kernel methods to weight the observations.

To be specific, Ñopo (2008) (perfect matching) was used to determine common support. In Słoczyński (2013) and Fortin (2008) only the sum of male advantage and female disadvantage is presented. The same applies to the distributional methods, and Juhn, Murphy and Pierce (1993) decomposition. For Machado and Mata (2005) the coefficients from wage regression on the pooled sample (without gender dummy) were used to construct counterfactual wages. In the RIF regressions, the reference wage structure was weighted average of male

[19]Sometimes even in spite of a large number of citations, examples include: Black *et al.* (2008), which was very close to Ñopo decomposition (2008), Donald *et al.* (2000), Olivetti and Petrongolo (2008) as well as Picchio and Mussida (2011) extension.

and female wage regression coefficients as in Słoczyński (2013). In the reweighting method of DiNardo *et al.* (1996), and matching decomposition of Ñopo (2008) the male wage structure is treated as non-discriminatory. The number of observations in the common support is the average of the percentage matched in each group times the total observations. Differences in the calculation of the raw gaps for quartiles between methods are related to technical issues (e.g. in the RIF regression approach the re-centered influence functions are constructed prior to performing decomposition).

These decomposition methods were performed for different specifications of the conditioning variables. First, the basic conditioning set, regardless of the method,[20] includes age, education, experience, tenure, marital status and geographical indicators, as described in Table 2. This choice of variables is widely acknowledged in the literature, cfr. Belzil (2007). Second, we use household level information as labor supply controls. This group of variables is used as an exclusion criterion in the Heckman corrected parametric estimations. We also use information on children in non-parametric estimations. The second group adds the occupation of the individual to the basic variables. In the third and fourth group we include firm related factors, such as industry (in both) and the size and type of ownership (only in the fourth). The following group includes tenure as an additional covariate. In the next specification, we added the information on the field of education. Finally, we repeat the estimations for all the variables combined. Thus in total we use 7 different sets of conditioning variables (observables we use to control for differences in endowments), which yields 23 combinations for each method of computing the adjusted gender wage gap.

Given the multiplicity of methods and model specifications, we will obtain 4347 estimates from seven different estimation methods[21] for seven different set of observables, with and without common support restriction, with and without selection correction. While the table with all the estimations is available upon request, in the next section we present the implications from analyzing the variation in these estimates.

5. Results

We present the results in three substantive parts. First, we compare the estimates of the adjusted gender wage gap for the basic specification. This specification includes age, education, geographical indicators, marital status, experience, tenure and presence of children in the small household. In the case of Heckman corrected coefficients, also sources of income in the household are used. Second, we compare the estimates across the methods, depending on the inclusion of additional control variables. Namely, we include separately industry, firm characteristics, occupation and finally the field of education. These results are

[20]In the Ñopo (2008) decomposition, we recoded the variables' age, experience and tenure in ten-year-groups. The size of each group was selected to maximize the number of matched observations without losing explanatory power.

[21]S Table A2 provides further details on the combinations that gave rise to the number of estimates.

TABLE 3

Gender Wage Gap from Different Methods with a Narrow Set of Conditioning Variables

	Total Sample		Common Support	
	Raw	Adjusted	Raw	Adjusted
Linear estimates				
OLS		15.7		15.9
Heckman corrected (active)		16.2		16.4
Heckman corrected (population 18–60/64)		15.6		15.8
Parametric (linear) decompositions (with Heckman correction, active)				
Female coefficients	9.0	17.1	9.1	17.2
Male coefficients	9.0	15.1	9.1	15.4
Słoczyński	9.0	16.1	9.1	16.4
Fortin	9.0	16.2	9.1	16.4
Quartile decompositions—Juhn. Murphy and Pierce (1993) (with Heckman correction, active)				
p25	9.1	14.1	10.5	15.6
p50	12.5	18.0	12.5	18.2
p75	10.5	18.2	11.3	18.9
Conditional quantile decompositions—Machado Mata (2005)—without/ with single index (active)				
p25	10.6/10.4	13.7/15.2	10.6/11.4	14.1/14.9
p50	11.3/11.8	16.4/19.0	11.4/13.4	16.7/18.9
p75	10.2/12.4	17.8/20.4	10.6/15.2	18.2/20.4
Reweighting method—DiNardo. Fortin and Lemieux (1996)				
p25	11.8	15.4	11.8	15.4
p50	13.4	21.9	14.5	23.6
p75	9.9	18.2	12.9	19.8
Unconditional quantile decompositions—RIF regressions (2007)				
p25	6.3	8.7	7.5	10.2
p50	10.7	16.4	10.7	16.6
p75	5.5	15.1	6.4	16.1
Ñopo (2008)				
Mean			7.6	17.6
% of matched male			95.8	
% of matched female			94.3	

Note: basic specification includes age, education, marital status, experience, tenure, children in the household, region and residence characteristics.

Source: data from Polish LFS, four quarters of 2012, one observation per person.

presented with and without common support restriction. In the third part of this section we discuss the differences in the obtained estimates across methods and conditioning sets. Namely, we diagnose the range of estimates from all the methods.

5.1. *Adjusted Gender Wage Gap for a Narrow Choice of Conditioning Variables*

In Table 3 we compare the values from the selected decompositions, with and without the common support restriction.[22] In all the specifications the adjusted gap is sizable, suggesting that for reasons beyond the observables included in this study wages of women are much lower than those of men, *ceteris paribus*. At the mean, the raw gap amounts to roughly 10 percent of theaverage female wage, and the adjusted gap is higher by about 6–10 percentage points depending on the

[22]Full set of estimates is available in the online appendix.

assumption about the counterfactual wage distribution (i.e. depending on the decomposition method). While there are some discrepancies, if women were paid according to men's wage structure, they would actually earn more based on all methods.

Fortin *et al*. (2011) have pointed out that the adjusted gender gap is typically larger with female coefficients as a reference wage structure, then in the case of using male coefficients. Such a relation is confirmed by our results, as the adjusted gap for the total sample is 17.1 percent with female wage structure as reference, and 15.1 percent if we take male wage regression coefficients. Słoczyński (2013) provides an interpretation of such results, and claims that it indicates an increasing gender gap along the income distribution.[23] This explanation seems to fit well our results, as in the decompositions that go beyond the mean the adjusted gap is in all cases larger at the median and third quartile, than at the first quartile. The results also indicate that within each method there are important differences between quartiles of income distribution, thus concentrating on the mean would not provide the full picture of gender wage gap.

After introducing Heckman correction the estimated adjusted gender gap increases. Also, estimates within the common support are typically higher than for the total sample. Thus, it seems that the more the specification focuses on comparing only the "comparable", the higher the estimates of the adjusted wage gap are.

Our main conclusion is that the different methods yield results which fall into a fairly narrow range of estimates for the adjusted gender wage gap. Also the estimates of the gap tend to increase as we focus on more similar men and women. Even though the main result is robust to different methods and along the income distribution, one could argue that our specifications are susceptible to bias resulting from the omitted variable problem. If men perform different jobs then women (as a reflection of their preferences), one would expect to see such outcomes, regardless of educational attainment. This relates not only to occupations and industries, but also to the characteristics of the employer and the fields of education. In the subsequent section we extend our specifications to include these variables and test how vulnerable were different methods to narrow model specification.

5.2. *Adjusted Gender Wage Gap with Extended Set of Conditioning Variables*

The results for the different conditioning sets are displayed in Table 4. We include the specifications for the total sample as well as for the common support. With the increasing number of covariates, the common support restriction becomes more binding: it eliminates a larger share of the observations, which

[23]In the Fortin (2008) and Słoczyński (2013) extensions of Oaxaca-Blinder decomposition neither male, nor female wage structure is perceived as non-discriminatory, and their approaches lead to intermediate estimates of the adjusted gap at the mean—the gap amounts to 16.1 percent. As was introduced earlier in this paper Fortin (2008) uses the wage structure estimated on the pooled sample with the gender dummy to construct the counterfactual wages. This approach leads by construction to the same estimate of the unexplained component, as provided by the coefficient of gender dummy in OLS regression; however it complements that with the detailed decomposition providing the information on the contribution of the particular covariates.

TABLE 4
Extending the Conditioning Set for the Adjusted Gender Wage Gap from Different Methods

	Industry		Industry+		Occupation		Education		Tenure		All	
	All	CS	All	CS	All	CS	All	CS	All	CS	All	CS
Linear regression												
OLS	16.0	17.1	15.3	15.7	15.0	15.7	18.2	18.3	16.4	17.0	14.7	16.0
Heckman correct active	16.5	17.6	15.7	16.3	15.5	16.2	18.6	18.8	16.8	17.5	15.1	16.4
Heckman correct all	15.9	17.1	15.1	15.9	15.2	15.9	18.1	18.4	16.3	17.1	14.8	16.3
Parametric (linear) decomposition (with Heckman correction among active)												
Female coefficients	15.9	19.1	15.5	17.7	18.3	19.0	21.6	22.1	17.7	18.3	18.2	18.7
Male coefficients	15.0	16.0	14.2	14.9	14.3	15.1	15.2	16.1	15.8	16.8	13.0	14.6
Słoczyński (2013)	15.5	17.5	14.9	16.3	16.4	17.1	18.5	19.5	16.8	17.6	15.7	16.6
Fortín (2008)	16.5	17.6	15.7	16.3	15.5	16.2	18.6	18.8	16.8	17.5	15.1	16.4
Quartile decomposition—Juhn. Murphy and Pierce (1993) (with Heckman correction among active)												
p25	16.2	20.2	15.1	16.5	16.4	20.0	18.8	19.5	17.7	18.7	14.6	12.0
p50	18.0	23.1	17.5	17.6	18.9	22.6	21.0	20.4	19.1	22.2	17.9	16.3
p75	16.2	20.4	16.0	19.9	17.4	19.1	18.8	24.7	16.9	22.0	18.0	22.7
Quartile decomposition—Machado and Mata (2005)												
p25	13.5	14.4	12.1	12.9	12.2	13.4	13.8	15.3	15.5	16.1	10.6	11.8
p50	15.7	17.1	14.9	16.7	16.8	18.2	16.5	18.2	17.9	18.8	14.6	17.3
p75	17.5	19.2	16.9	18.4	18.3	18.6	18.1	19.1	18.3	19.5	16.0	17.1
Quartile decomposition—Machado and Mata (2005) with sample correction (active)												
p25	14	11.8	12.9	4	10	6.5	12.8	15.7	15.9	16.1	8.9	3.7
p50	18.5	17.6	17.1	13.1	18.2	12.4	16.2	19.3	19.7	19	14.3	6.2
p75	21.2	23.4	18.6	13.3	19.1	17.9	17.1	20.4	20.5	20.8	16	12.4
Reweighting method—DiNardo, Fortin and Lemieux (1996)												
p25	15.4	15.4	15.4	15.4	12.5	15.4	15.4	18.2	15.4	18.2	11.8	8.0
p50	23.6	23.6	20.1	18.2	20.1	23.6	23.6	23.6	23.6	23.9	18.2	18.2
p75	22.2	25.5	22.2	22.3	21.8	21.4	22.2	26.2	18.8	23.9	18.2	22.3
Unconditional quantile decompositions—RIF regressions												
p25	9.0	12.4	8.5	9.2	8.2	11.4	11.3	12.0	10.8	11.8	8.4	11.3
p50	12.4	17.7	11.7	12.6	14.7	19.2	17.2	20.6	15.2	17.7	13.8	18.9
p75	21.6	25.2	20.9	16.3	24.5	25.7	24.7	21.9	23.4	18.7	22.9	19.6
Ñopo (2008)												
Mean	17.3		15.9		16.1		17.4		17.9		16.5	
% of matched male	78.27		59.8		75		87.4		89.9		11.4	
% of matched female	84.2		67.1		77.7		72.3		88.4		13.3	
No of observations	33,571		33,571		33,571		33,571		33,571		33,571	

Note: please, refer to Table 3. Industry (column 1) includes 4 dummy variables (agriculture, manufacture, construction, services). Industry + (column 2) includes industries and other firm level variables (private ownership dummy and size). Occupation (column 3) comprises 9 dummy variables for each ISCO code at 1 digit. Tenure (column 4) comprises tenure in the current job (in years). Education (column 5) comprises dummies for the field of education, see footnote 17. Finally, in the last column we included all the variables together. All the models were estimated with (column "CS") and without (column "All") the common support restriction. Standard errors of the adjusted wage gap available upon request.

Source: own calculation, Polish LFS, four quarters of 2012.

undermines the external validity of the conclusions from those regressions. In the extreme case, when all variables are included in the model, the percentage of male and female matched is below 15 percent.

These results allow identifying the role of three sources of variations in the size of the estimated adjusted gender gap: the introduction of new variables, the different methods and the common support restriction. In all specifications, the adjusted gap increases in the common support which corroborates the finding that the wage differences are larger among more comparable men and women. In other words, there are unmatched women at the top of the distribution and unmatched men at the lower points. The results from the JMP decomposition are an extreme example. When all women are considered, it seems like the gap is present only at the lower quartile. After we controlled for the common support, it shows a glass ceiling effect in half of the specifications.

Including industry and firm level characteristics (the "industry +" specification) among the covariates lowers the estimates of the adjusted gender wage gap, by approximately 2–4 percentage points with respect to the basic specification. The addition of occupation dummies has a negligible effect on the estimates. Comprising tenure tends to increase the adjusted gap estimates, but economically the effects are not large. In the case of education dummies, the effect is stronger. Estimations of the gap tend to rise after we incorporate the field of education. Moreover, in most cases the maximum values of the adjusted gender gap are found in this column. In few cases of the decrease in the gap after additionally controlling for field of education, the magnitude of observed drop is relatively small in both absolute and relative terms. The inclusion of all explanatory variables in a single estimation does not reduce the adjusted wage gap significantly in most specifications. The sole exception is the Machado-Mata, where the value of the gap decreased by 5 percentage points at the higher quartile. Notice, however, that this coefficient was calculated in a rather small sub-sample.

5.3. *Exploring Further the Estimates of the Gender Wage Gap*

After obtaining multiple estimates of the adjusted gender wage gap for each method, we can investigate the sensitivity of each approach to the specification. For each method we can compare the estimates for one statistic (mean or percentiles, depending on the method) across different specifications. We approach that comparison in two steps. First, we do pairwise comparisons for all of the modelling choices. Given the multiplicity of estimates, we run parametric mean equality tests for the subsample of estimates with a given method/set of variables and a subsample without. These tests provide evidence with reference to each specification separately. We present them in Table 5.

The mean equality tests suggest that most of the intuitive interpretations from Table 3 and Table 4 are actually statistically significant differences. While economic significance of some modelling choices may be limited, a large fraction of methods produce estimates of gender wage gap different by approximately 2 percentage points. Also controlling for a common support has a significant impact on the value of the estimates, making the overall estimates of the adjusted gap larger. On the other hand, the role of selection correction does not seem to alter substantially the obtained estimates. The inclusion of new variables produced mixed results. In the case of "Industry +" and "Occupation", the adjusted gap decreased significantly, while "Education" or "tenure" lead to a similar increase in the gap.

Next, we explore the determinants of changes in the measures of the adjusted gender wage gap considered jointly. Thus, we run an OLS, where controls,

TABLE 5
Testing for the Role of Method, Controls and Variables

	Size of the Adjusted GWG		Number of observations		
	Yes	No	Yes	No	t-ratio
Controls					
Common support	17.802	15.407	1449	1426	–22.886
Selection from active (Heckman)	16.741	16.590	460	2415	–0.974
Selection from workforce (Heckman)	16.488	16.638	460	2415	0.966
Selection from active (Index)	16.505	16.635	460	2415	0.837
Selection from workforce (Index)	16.493	16.637	460	2415	0.927
Variables included					
Industry	16.757	16.552	875	2000	–1.661
Industry+	16.046	16.863	875	2000	6.655
Occupation	16.707	16.543	1250	1625	–1.426
Field of Education	16.865	16.385	1375	1500	–4.224
Tenure	16.903	16.350	1375	1500	–4.882
Method of estimation					
OLS	16.264	16.645	230	2645	1.820
Oaxaca type decomposition	16.098	17.271	1610	1265	10.425
Juhn, Murphy and Pierce decomposition	18.255	16.302	460	2415	–12.946
DiNardo, Fortin and Lemieux decomposition	19.564	16.517	92	2783	–9.578
Machado Mata decomposition	17.259	16.582	138	2737	–2.549
Re-centered Influence function	15.931	16.700	322	2553	4.277
Ñopo	17.327	16.609	23	2852	–1.126

Notes: Observations for the test correspond to the mean and/or the median adjusted wage gap, depending on the method employed. Each test compares the value for one specific subgroup to all others. See Table A2 for the explanation on the number of observations.

variables and methods for each estimate are coded by a set of dummies to represent explaining variables, whereas the estimates of the adjusted gender wage gap are the explained variable. Given that the dependent variable comes from a series of estimations, we bootstrap standard errors, see Efron and Tibshirani (1994). The results are presented in Table 6. The inclusion of all possible combinations was done on an equal footing, without any particular weight attached to any estimate. In the first three columns we proceed to separate the effects of different sources of variation: namely, the controls for more similar samples (common support and correction for selection, column 1), the method selected (column 2), and the addition of new variables (column 3). In the fourth column, we put all variables together to show which decisions (variables included, methods or controls) are more likely to influence the final results and in which direction.

In the second part of Table 6, columns 5 to 7, we follow a thought experiment. We include controls for all those choices which are not commonly applied in the field. Variables, controls and methods typically included in the studies constitute jointly a base level. Thus a constant informs about the estimate of the gender wage gap if the researcher employs “standard” techniques, instead of exploring more complex alternatives. In column 5, we present the results for sets of variables, while in columns 6 to 7 we focus on the effects of different methods. Our approach intends to compare comparable objects, thus we divide the sample in those estimates obtained from quantile decompositions and those estimated at the mean.

TABLE 6

Exploring the Adjusted Wage Gap

	Total sample					Quartiles	Means
	1	2	3	4	5	6	7
Common Support	3.209***			3.205***			
	0.088			0.082			
Heckman Selection active	1.266***			0.27**			
	0.136			0.138			
Heckman Selection working age population	1.001***			0.005			
	0.13			0.137			
Non-parametric selection active	0.999***			0.003			
	0.125			0.118			
Non-parametric selection working age population	0.986***			–0.01			
	0.121			0.124			
Industry		–0.445***		–0.445***			
		0.116		0.095			
Industry+		–1.186***		–1.186***			
		0.116		0.097			
Occupation		0.341***		0.341***			
		0.109		0.085			
Tenure		0.783***		0.783***	0.603**		
		0.104		0.081	0.099		
Field of education		0.795***		0.795***	0.614**		
		0.104		0.078	0.102		
Parametric decompositions			0.1	0.1			
			0.156	0.157			
DiNardo, Fortin and Lemieux			1.247***	1.301***		1.232***	
			0.254	0.283		0.3	
Machado-Mata			–2.273***	–2.219***		–2.358***	
			0.191	0.2		0.203	
Recentered influence functions,			–2.329***	–2.276***		–2.642***	
			0.242	0.229		0.198	
Nopo			1.064***	–0.485			0.961***
			0.279	0.323			0.23
Not male reference wages	1.591**			1.591***			1.49***
	0.176			0.139			0.099
Constant	13.77***	15.81***	14.67***	12.61***	15.63***	19.29***	14.88***
	0.103	0.097	0.229	0.229	0.083	0.14	0.081
Control for quantiles				Yes			No
N	4347	4347	4347	4347	4347	2208	2139
R-Squared	0.374	0.201	0.255	0.476	0.199	0.28	0.044
Adjusted R-Squared	0.373	0.199	0.253	0.474	0.197	0.278	0.043

Notes: ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. Bootstrapped standard errors reported in parentheses. In column 6, we use the median as baseline category. In column 7, percentiles were not included as all observations were obtained at the mean. See Table A2 for the explanation on the number of observations.

Source: own calculation, results from specifications reported in Table 4 and all other possible combinations of these specifications

By and large, the regressions confirm the findings of bivariate tests presented in Table 5, which points to the significance of (most) considered elements jointly. The fit of the models discussed in Table 6 is roughly 20–50 percent. Most of the dummies for variables and methods prove statistically significant predictors of the variation in the estimates of the adjusted gender wage gap, while the controls produced mixed results. We ratify our intuitions with respect to the common support and the selection correction mechanisms. The role of selection bias—whether parametric or not—is fairly small in estimating the gender wage gap. This would indicate that while women face difficulties in entering the labor market, these are not helpful in explaining the extent of the gender wage gap. By contrast, "comparing the comparable"—i.e. the common support—has a fairly large effect on the estimates of the gender wage gap. Restricting the analysis to the common support leads to estimates of the gender wage gap which are significantly higher (on average by 3 percentage points). In column 4, we observe that only common support control and, to a lesser extent, the Heckman correction from the active population are significant once we take into account the influence of other variables. This effect is likely to follow from country specific factors, though.

Let us consider how the inclusion of different variables affects the estimates of the adjusted gender wage gap. First, we observe that the direction of the effects is only negative when we include industry categories in the analysis. The effects increase when other firm characteristics are added to the decomposition, as seen in the coefficient on the "industry +"set. The inclusion of the dummies for fields of education in fact raises the value of the adjusted wage gap. Therefore, even if women are concentrated in certain educational fields, this does not provide a convincing explanation for the gender wage gap. With respect to tenure, including this variable in the estimation also increases the estimates of gender wage gap. Given that wages should be positively related to wages (either because workers acquire job-related skills, or they benefit from compensation schemes based on seniority). The tenure coefficient in Table 5 indicates that women are less rewarded for their experience in the same company in comparison to men.

Though "industry" and "industry +" are both significant, the differences in the coefficient shows that industry itself explains only a small part of the adjusted wage gap variation. Other firm characteristics (size and type of ownership) play a much more significant role in explaining the differences. The negative sign indicates that the models including these variables tended to provide lower values of the adjusted wage gap. No similar effects are observed for the occupations. Omitting industry categories in the process of estimating the gender wage gap in fact leads to higher estimates of the adjusted gender wage gap, at least in Poland.[24]

Last but not least, it is often argued that lower wages of women reflect self-selection into lower paying occupations, industries and even fields of education. Our results provide little support for this contention, regardless of the method used. In fact, when controlling for industry characteristics the estimates of the gender wage gap are somewhat lower, but adjusted wage gap remains substantially higher than the raw gap. Inclusion of tenure and fields of education yields estimates somewhat larger, making the difference between the raw and adjusted gap even more—not less—pronounced. Finally, occupations play only a small role in the gender wage gap: women and men in the same position are still paid differently. On the other hand, the changes in the common support are large, reflecting the larger concentration of men in low paying

[24]However, this is not equivalent to the overestimation per se, since we do not know what the "true" adjusted gender wage gap is.

sectors. We thus reject the hypothesis that women self-selection into "lower paid" jobs and fields can explain the gender wage gap.

A comparison of the methods indicates that the selection does affect the results. For instance the decision to use male wage structure as a reference leads to significantly lower estimates of the gender wage gap. Parametric methods (such as Oaxaca-Blinder/Juhn, Murphy and Pierce decompositions) appear to provide similar results to the standard OLS regressions. Quantile methods provide different results. DFL decomposition produced considerably larger estimates of the adjusted gender wage gap, when compared to OLS and specially other quartile methods. The results might be driven by some of the characteristics of the method, in particular that we cannot control for sample selection (which leads to a smaller number of estimates) and the inclusion of estimates at the mean. A lecture of the estimates of the other quartile methods requires then to take the coefficients from the different quartiles into consideration. Performing this operation, results in changes in the gap (at the median) which are around 0.3 percentage points higher in the quantile regressions. Finally, Ñopo decomposition is related to larger estimates of the adjusted wage gap, though the effects appear to be driven by the restriction to the common support.

The results of the thought experiments provide some additional information on what we can expect by estimating the gender wage gap in a mechanical way. The main variable of interest is the constant, which shows the average gap when all the remaining variables are set to zero. The results suggest that the omission of tenure and fields of education leads to lower estimates of the gender wage gap, as the constant is smaller than the overall mean (16.6). Similar reasoning applies to column 7. More recently developed estimations produce higher estimates of the adjusted wage gap, with a larger deviation from the mean than in the previous case. Thus, in these two cases, increasing the complexity of the estimation method leads to higher estimates of the gap.

Column 6, which deals with methods based on the distribution, shows the differences between the initial JMP estimations and more recent quantile techniques. The constant is not directly comparable to other columns, as the baseline category is the median, and not the mean. Nonetheless, we can observe that estimates from JMP and DFL will tend to be larger than those obtained from MM or RIF methods.

6. Conclusions

There is a multiplicity of methods to estimate the adjusted gender wage gap. The differences between them are both methodological and interpretational, but in each and every method the adjusted wage gap is a differential that cannot be attributed to the observables. Thus, without clear information on the data generation process, it is impossible to determine which is the most accurate. We performed a comparative analysis of the available alternatives for computing the adjusted gender wage gap using always the same sample. While this is similar to the exercise of Fortín *et al.* (2011), our analysis focuses on gauging the susceptibility of the obtained estimates to the choice of controls and conditioning variables in addition to the methods. To the best of our knowledge, such analysis has not been conducted before, and hence this is a contribution of this paper, especially as it demonstrates that the choice of a conditioning variable affects the estimates of the GWG in a non-trivial way. Our work provides an applied comparative analysis of the different methods on one source of data. Thus, in a sense, we extend also the findings of Weichselbaumer and Winter-Ebmer (2005) meta-analysis, because our results are fully comparable across methods and we directly control the inclusion of additional explanatory variables on the estimated size of the adjusted wage gap.

We used data from Polish LFS of 2012 and hourly wage as a measure of compensation. The raw gap amounts to roughly 10 percent of male wage. In order to obtain the adjusted wage gap we applied 7 different estimation methods and employed different sets of conditioning variables. In spite of these differences all of the estimations showed that the adjusted gap was larger than the raw gap, which means that given the observable endowments, women should have received a larger pay.

While there is some dispersion of the estimates for the adjusted gender wage gap, the actual size of the estimate depends crucially only on some of the modelling choices. First, it is important to "compare the comparable". Indeed, the estimations within the common support resulted in adjusted wage gaps sizably larger, regardless of the set of conditioning variables, which is indicative of unmatched women being better endowed than unmatched men. In addition, Ñopo (2008) decomposition should be preferred when there is fear of an omitted variable bias, as the estimates with only some conditioning variables were fairly similar to those for a larger set of explanatory factors. Second, the results highlight the value of the distributional analysis. Indeed, the adjusted gap proves to be differentiated along income distribution regardless of the conditioning set for all relevant methods. Third, a number of studies emphasizes that the adjusted wage gap tends to be exaggerated if the choice of variables is too narrow. Specifically, women may tend to locate in occupations, industries and even fields of study where the returns to individual characteristics are lower. Also, women are believed to have less experience. Our results suggest that the inclusion of these variables does not reduce the size of the gap to a noticeable extent. Indeed, in the case of Poland, when controlling for occupations and firm characteristics, the estimates of the gender wage gap are on average about 1 percentage points lower than in the absence of these controls in the conditioning set. On the other hand, however, specifications which control for tenure and fields of education tend to be associated with similarly higher estimates of the adjusted gender wage gap. Thus, while women tend to work in lower paying positions and for lower paying firms, their individual characteristics and choices—such as job mobility and field of education—tend to be the source of lower returns rather than of less productive endowments. Notwithstanding, these effects are not large economically.

The characterization of the gender wage gap for Poland is an important side product of our analysis, as the previous literature for the country is rather scarce. Some authors explored the topic in the aftermath of the transition, and the accession to the EU; however, there is a lack of recent analysis on the topic. Moreover, we also profited from a recent Labor Force Survey (LFS), which includes information on the field of study, a variable which has not been included in previous research. Given the low female labor force participation, analyses of the gender wage gap are of clear policy relevance. If "discrimination" is indeed a prevalent phenomenon, it can partly explain the low female employment rates. We also note that our results might stem from the specific characteristics of the Polish labor market and its deficiencies. Thus, it seems that a study covering a wider selection of countries could corroborate and generalize the findings demonstrated in this paper.

References

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

Albrecht, J., A. van Vuuren, and S. Vroman, "Counterfactual Distributions with sample selection adjustments: econometric theory and an application to the Netherlands", *Labour Economics*, 16(4), 383–96, 2009.

Appleton, S., A. Bigsten, and D. Manda, "Educational expansion and economic decline: returns to education in Kenya 1978–1995". *WPS/99-6*, Centre for the Study of African economies, 1999.

Barsky, R., J. Bound, K. K. Charles, and J. P. Lupton, "Accounting for the Black-White Wealth Gap: A Nonparametric Approach," *Journal of the American Statistical Association*, American Statistical Association, 97, 663–673, September, 2002.

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

Belzil, C., "Testing the specification of the Mincer wage equation" *Iza DP* No 2650, 2007.

Black, D., A., Haviland, S. Sanders, and L. Taylor, "Gender wage disparities among the highly educated". *Journal of Human Resources*, University of Wisconsin Press, 2008.

Blau, F. D. and L. M. Kahn, "Understanding international differences in the gender pay gap", *Journal of Labor Economics*, University of Chicago, 21(1), 2003.

———, "Wage Structure and Gender Earnings Differentials: An International Comparison," *Economica*, London School of Economics and Political Science, 63(250), S29–62, 1996.

Blinder, A. S., "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8, 436–55, 1973.

Brainerd, E., "Women in transition: Changes in gender wage differentials in Eastern Europe and the former Soviet Union", *Industrial and Labor Relations Review*, ILR Review, Cornell University, ILR School, 54(1), 138–62, October, 2000.

Brown, R. S., M. Moon, and B. S. Zoloth, "Incorporating Occupational Attainment in Studies of Male-Female Earnings Differentials," *The Journal of Human Resources*, 15(1), 3–28, 1980.

Cho, D. and J. Cho, "How Do Labor Unions Influence the Gender Earnings Gap? A Comparative Study of the US and Korea", *Feminist Economics*, 17(3), 133–57, 2011.

Cotton, J., "On the Decomposition of Wage Differentials." *The Review of Economics and Statistics*, 70, 236–43, 1988.

Donald, S., D. Green, and H. Paarsch, "Differences in Wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates" *Review of economic studies*, 67, 609–33, 2000.

DiNardo, J., N. Fortín, and T. Lemieux, "Labor market institutions and the distribution of wages, 1973–1992: a Semi-parametric approach" *Econometrica*, Econometric Society, 64(5), 1001–44, 1996.

Efron, B. and R. J. Tibshirani, *An introduction to the bootstrap*. Chapman and Hall/CRC press, 1994.

Fairlie, R., "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models" *Journal of Economic and Social Measurement*, 30, 305–16, 2005.

Firpo, S., N. Fortín, and T. Lemieux, "Unconditional Quantile regressions", *Econometrica*, Econometric Society, 77(3), 953–73, 2009.

Fortín, N. "The gender wage gap among young adults in the United States" *The journal of Human Resources*, University of Wisconsin Press, 43(4), 2008.

Fortín, N., T. Lemieux, and S. Firpo "Decomposition methods in Economics" in *Handbook of Labor Economics*, Elsevier, 2011.

Goraus, K. and J. Tyrowicz, "Gender Wage Gap in Poland—Can It Be Explained by Differences in Observable Characteristics", *Ekonomia*, 36(1), 125–48, 2014a.

———, "The Goodwill Effect? Female Access to the Labor Market Over Transition: A Multicountry Analysis", *WNE UW Working Paper* No 104, 2014b.

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

Heckman, J. J., "Sample Selection Bias as a Specification Error," *Econometrica*, Econometric Society, 47(1), 153–61, January, 1979.

Huber, M., "Causal pitfalls in the decomposition of wage gaps", *Journal of Business and Economic Statistics*, 33(2), 2014.

Imbens, G. and D. Rubin, *Causal Inference*, Part III, Chapter 14, 2009.

Juhn, C., K. Murphy, and B. Pierce, Wage inequality and the rise in returns to skill, *The Journal of Political Economy*, 101(3), 410–42, 1993.

Kalinowska-Nawrotek, B., "Dyskryminacja kobiet na polskim rynku pracy". *Akademia Ekonomiczna w Poznaniu*, 2005.

Klein, R. and R. Spady, "An efficient semiparametric estimator of the binary response models". *Econometrica*, Econometric society, 61(2), 387–421, 1993.

Kot, S. M., B. Podolec, and P. Ulman, "Problem dyskryminacji płacowej ze względu na płeć", in: S. M. Kot (red.), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, Warszawa–Kraków: Wydawnictwo Naukowe PWN, 1999.

Kotowska, I. E. and M. Podogrodzka, Dyskryminacja kobiet na rynku pracy, „*Kobieta i biznes*", nr 2–3, 1994.

Łatuszyński, K. and Ł.P. Woźny, "Zróżnicowanie wynagrodzeń kobiet i mężczyzn na polskim rynku pracy w 2004 roku," in: W. Pacho and M. Garbicz (red.), *Wzrost gospodarczy a bezrobocie i nierówności w podziale dochodu*, Warszawa: Szkoła Główna Handlowa, 2008.

Machado, J. AF and J. Mata. "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of applied Econometrics*, 20(4), 445–65, 2005.

Machado, C., "Selection, heterogeneity and the gender wage gap" *IZA DP No. 7005*, 2012.

Magda, I. and A. Szydłowski, "Płace w makro i mikroperspektywie," in M. Bukowski (red.), *Zatrudnienie w Polsce 2007—Bezpieczeństwo na elastycznym rynku pracy*, Warszawa: Ministerstwo Pracy i Polityki Społecznej, 2008.

Matysiak, A., T. Słoczyński, and A. Baranowska, "Kobiety i mężczyźni na rynku pracy", w: M. Bukowski (red.), *Zatrudnienie w Polsce 2008—Praca w cyklu życia*, Warszawa: Centrum Rozwoju Zasobów Ludzkich, 2010.

Mazur-Łuczak, J., "Kobiety na rynku pracy" Wyższa Szkoła Nauk Humanistycznych i Dziennikarstwa, 2010.

Melly, B., "Estimation of counterfactual distributions using quantile regression". *Review of Labor Economics*, 68(4), 543–72, 2006.

Moral-Arce, I. and S. Sperlich, "A Semiparametric Juhn, Murphy, and Pierce—Decomposition for the Spanish Gender Pay Gap: A methodological Note", *Working Paper* 04.2008, University of Gottingen, 2008.

Morawski, L. and M. Myck, " 'Klin'-ing up: Effects of Polish tax reforms on those in and on those out", *Labour Economics*, Elsevier, 17(3), 556–66, 2010.

Neumark, D., "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination," *The Journal of Human Resources*, 23(3), 279–95, 1988.

Newell, A. and B. Reilly, "The gender pay gap in the transition from communism: some empirical evidence", *Economic Systems*, 25(4), 2001.

Ñopo, H., N. Daza, and J. Ramos, "Gender earning gaps around the world: a study of 64 countries", *International Journal of Manpower*, 33(5), 464–513, 2012.

Ñopo, H., "Matching as a tool to decompose gaps," *The Review of Economic and Statistics*, 90(2), 290–9, May 2008.

Oaxaca, R., "Male-female wage differentials in Urban Labor markets", *International economic review*, 14(3), 1973.

Olivetti, C. and B., Petrongolo, "Unequal pay or unequal employment? A cross-country analysis of gender gaps," *Journal of Labor Economics*, 26(4), 621–54, 2008.

Pailhé, A., "Gender discrimination in Central Europe during the systemic transition", *Economics of Transition*, 8(2), 2000.

Picchio, M. and C. Mussida, "Gender wage gap: a semi-parametric approach with sample selection correction," *Labour Economics*, 18(5), 564–78, 2011.

Rokicka, M. and A. Ruzik, "The Gender Pay Gap in Informal Employment in Poland," *CASE Network Studies and Analyses 406*, 2010.

Reimers, C. W., "Labor Market Discrimination Against Hispanic and Black Men," *The Review of Economics and Statistics*, 65, 570–9, 1983.

Shorrocks, A., "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value". *Journal of Economic Inequality*, 11(1), 99–126, 2013.

Słoczyński, T., "Population average gender effects", IZA Discussion Paper No.7315, 2013.

Watson, I., "Decomposing the Gender Pay Gap in the Australian Managerial Labour Market" *Australian Journal of Labour Economics*, 13(1), 49–79, 2010.

Weichselbaumer, D. and R. Winter-Ebmer, "A Meta-Analysis of the International Gender Wage Gap", *Journal of Economic Surveys*, 19(3), 479–511, 2005.
Zwiech, P., "Różnice w wynagrodzeniach kobiet i mężczyzn na rynku pracy w Polsce," *Zeszyty Naukowe Uniwersytetu Szczecińskiego*, nr 9, 2005.
Zhang, J., J. Han, P. Liu, and Y. Zhao, "Trends in the Gender Earnings Differential in Urban China, 1988–2004" *Industrial and Labor Relations Review*, 61(2), 224–43, 2008.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix: Parametric control for selection