Author(s): Maczulskij, Terhi; Nyblom, Jukka

Title: Measuring the gender wage gap : a methodological note

Year: 2020

Version: Accepted version (Final draft)

Please cite the original version:

Maczulskij, T., & Nyblom, J. (2020). Measuring the gender wage gap : a methodological note. Applied Economics, 52(21), 2239-2249. https://doi.org/10.1080/00036846.2019.1687840

Measuring the gender wage gap—A methodological note

Terhi Maczulskij[a], Jukka Nyblom[b]

[a] Corresponding author. Research Institute of the Finnish Economy, Arkadiankatu 23 B, FIN-00100 Helsinki, Finland, terhi.maczulskij@etla.fi, +35850-3230180

[b]Department of Mathematics and Statistics, P.O.Box 35, FIN-40014 Jyvaskyla, Finland, jukka.nyblom@jyu.fi.

Abstract

We propose to estimate the Blinder-Oaxaca decomposition by a single-equation model augmented with interactions between the group membership and other predictors. The relative importance of predictors on the discriminatory wage gap is examined by the interaction coefficients, which may lead to very different conclusions than the usual percentage calculations using the detailed decomposition method. Comparisons are made between the traditional interpretations and those suggested here using wage data from Finland. The decomposition analysis suggests that the discriminatory male-female wage gap is largely related to work experience, while our preferred model points to the importance of family gap and working industry.

Keywords: Linear regression model; Wage differential; Decomposition; Interaction

JEL: C21, J31, J71

Acknowledgements: The authors thank Jari Vainiomäki and Ari Hyytinen for their useful comments.

Conflict of interest: The authors declare that there are no conflicts of interest

Words: 6602

1. Introduction

Many studies in economics literature are devoted to problems in measuring the wage gap between comparable male and female workers. The conventional approach to wage differential research is the Blinder-Oaxaca (BO) model (Blinder, 1973; Oaxaca, 1973). Its focus is on decomposing the observed raw wage gap into the explained and unexplained components. The explained part of the wage gap is attributed to gender differences in predictors (endowments) and the unexplained part of the wage gap is attributed to gender differences in predictor coefficients (discrimination).

Following the common practice, we model the logarithm of the wage linearly on background variables and introduce interaction coefficients related to gender and other predictors into the model, as done by Elder, Goddeeris and Haider (2010). This approach also has one immediate benefit, because the standard errors are automatically obtained (see, e.g. Słoczyński, 2013; Gelbach, 2016). We further interpret the interaction coefficients as the indirect regression coefficients of the wage gap. This *ceteris paribus* effect describes how much of the change in the gender wage gap can be attributed to the change in one predictor holding everything else constant (see, e.g., Biewen, 2014).

In addition to overall measures of explained and unexplained parts of the wage gap, researchers often want to evaluate how much of the explained gender wage gap is due to differences in various predictors. Similarly, it is informative to examine to what extent different predictors contribute to the unexplained wage gap. Earlier studies have mainly used detailed BO decomposition to assess the relative importance of various predictors on the explained part of the wage gap only. This is because the detailed BO decomposition suffers from not being invariant with respect to affine transformations on the predictors (e.g. see Oaxaca and Ransom, 1999). As a remedy for this Gardeazabal and Ugidos (2004) and Yun (2005) have proposed to average the coefficients to zero rather than to define a reference category with zero coefficients. However, with any normalization, some degree of arbitrariness is unavoidable (Yun, 2008, p.

31) which may leave the decomposition without a meaningful interpretation (Fortin, Lemieux and Firpo, 2011, p. 45). We contribute to the literature by examining the importance of different predictors using interaction coefficients. First, the interaction coefficients provide detailed quantitative information on the discriminatory male-female wage gap by different predictors, holding other covariates constant. Second, the relative contribution of different predictors on the discriminatory wage gap is examined by the ratios of the interaction coefficients as suggested in Cox (1992, p. 294).

Our approach is illustrated with the Finnish Structure of Earnings Survey (SES) matched with information on firm and personal characteristics for the year 2010. It turns out that quite a different view of the importance of various predictors on the gender wage gap may appear compared to the traditional detailed decomposition. The rest of this paper is organized as follows. The next section presents our econometrical approach and its relation to BO decomposition. Section 3 presents the empirical application, Section 4 presents robustness tests, and Section 5 concludes the paper.

2. Relation to BO Decomposition

Let us assume the wage equations

$$\log w_M = \beta_{0M} + \boldsymbol{\beta}'_M \boldsymbol{x} + \epsilon_M \qquad \text{for males,} \qquad (1)$$

$$\log w_F = \beta_{0F} + \boldsymbol{\beta}'_F \boldsymbol{x} + \epsilon_F \qquad \text{for females,} \qquad (2)$$

where the outcome variable is the logarithm of wages (w) and matrix $\boldsymbol{x}$ consists of explanatory variables and the errors ϵ_M and ϵ_F are from $N(0, \sigma^2)$. The subscript notation $_M$ refers to males and $_F$ refers to females. After fitting the linear regressions for both genders separately, the average log-wage gap is decomposed into two different forms

$$\text{ave}(\log w_M) - \text{ave}(\log w_F) = \widehat{\beta}'_F(\bar{x}_M - \bar{x}_F) + \left[\left(\hat{\beta}_M - \hat{\beta}_F\right)'\bar{x}_M + \left(\hat{\beta}_{0M} - \hat{\beta}_{0F}\right)\right] \quad (3)$$

$$= \widehat{\beta}'_M(\bar{x}_M - \bar{x}_F) + \left[\left(\hat{\beta}_M - \hat{\beta}_F\right)'\bar{x}_F + \left(\hat{\beta}_{0M} - \hat{\beta}_{0F}\right)\right] \quad (4)$$

where ave is average. The first term on the right side of (3) is called the explained part, and the latter is called the unexplained part, or the discrimination wage gap. In the background of the BO decomposition, there is an assumption of a fair (or nondiscrimination) wage ratio. Oaxaca (1973) suggested two candidates for the fair wage ratio: the male wage structure $\widehat{\boldsymbol{\beta}}_M$ or the female wage structure $\widehat{\boldsymbol{\beta}}_F$.[1] According to Fortin et al. (2011), the group choice is just a matter of choosing a meaningful counterfactual. Hereafter, assume that the decomposition is estimated under female wage structure.

Within the BO framework, it is common to go further to estimate how the different predictors contribute to the explained and discriminatory wage gap. With regard to the unexplained wage gap, assume we have p predictors in $\boldsymbol{x}$. For the k-th predictor we have

$$\frac{\left(\widehat{\beta}_{k,M} - \widehat{\beta}_{k,F}\right)\bar{x}_{k,M}}{\left(\widehat{\beta}_{0,M} - \widehat{\beta}_{0,F}\right) + \sum_j^p \left(\widehat{\beta}_{j,M} - \widehat{\beta}_{j,F}\right)\bar{x}_{j,M}} \qquad k = 1, \ldots, p. \quad (5)$$

With categorical variables, the researchers usually combine the terms corresponding to different level dummies. The polynomial predictors are amalgamated in the same manner. However, there are problems due to lack of invariance. Oaxaca and Ransom (1999) concluded that when the model contains categorical variables, the overall decompositions are valid, as well as the

[1] Neumark (1988) argued that the fair wage structure, $\boldsymbol{\beta}^*$, can be deduced from the employers' behavior. Unfortunately, $\boldsymbol{\beta}^*$ is hardly empirically discernible as Neumark (1988, p. 285) admits. Nevertheless, after some reasoning, he ends up with the practical solution that $\boldsymbol{\beta}^*$ (or its proxy) is estimated from the pooled sample. Reimers (1983) suggested a simple average $\boldsymbol{\beta}^* = (\widehat{\boldsymbol{\beta}}_M + \widehat{\boldsymbol{\beta}}_F)/2$. Cotton (1988) took the weighted average $\boldsymbol{\beta}^* = f_M\widehat{\boldsymbol{\beta}}_M + f_F\widehat{\boldsymbol{\beta}}_F$, with the weights f_M and f_F consisting of fractions of the corresponding group members.

detailed decomposition of the explained part, but unfortunately, the same is not true for the discrimination part. Gardeazabal and Ugidos (2004) and Yun (2005) offered one solution to this problem and propose to average the coefficients to zero rather than to define a reference category with zero coefficients. However, this approach was criticized by Fortin et al. (2011, p. 45) for possibly leaving the decomposition without a meaningful interpretation. Similar problems arise in dealing with continuous variables. Often, the model includes the age (or work experience) that could be used as such or after subtracting some threshold, e.g., 18 years. Again, the overall decompositions and the detailed decomposition of the explained part give the same result under either choice but not the detailed decomposition of the discrimination part. Furthermore, if the square of the age is also included, as is often the case, the detailed decomposition of the discrimination part lacks invariance. However, the overall decompositions remain invariant. Because of these problems in the detailed decompositions, we suggest not using them but instead making comparisons by the ratios of interaction coefficients.

As we already know from the basic econometric literature, Equations (1) and (2) can be combined into a single log-wage equation with interactions, as follows:

$$\log w = \beta_0 + \delta_0 d_M + \boldsymbol{\beta}'\boldsymbol{x} + \boldsymbol{\delta}'(d_M \boldsymbol{x}) + \epsilon, \qquad (6)$$

where d_M is 1 for males and 0 for females. Combining Equations (1)-(4) and (6), it is clear that $\beta_0 = \beta_{0F}$, $\delta_0 = \beta_{0M} - \beta_{0F}$, $\boldsymbol{\beta} = \boldsymbol{\beta}_F$ and $\boldsymbol{\delta} = \boldsymbol{\beta}_M - \boldsymbol{\beta}_F$. The unexplained or discriminatory wage gap can be further calculated using the means of the predictors and the estimates from the single-equation model such that $\exp(\delta_0 + \boldsymbol{\delta}'\boldsymbol{x}) = \exp[(\beta_{0M} - \beta_{0F}) + (\boldsymbol{\beta}_M - \boldsymbol{\beta}_F)'\boldsymbol{x}_M]$. We assume that $\epsilon \sim N(0, \sigma^2)$ i.e., compared to (1)-(2) we assume $\sigma_M^2 = \sigma_F^2 = \sigma^2$. One immediate benefit of this formulation is that it allows direct calculation of standard errors for the interaction coefficients (see, e.g., Słoczyński, 2013; Gelbach, 2016).

We use this mathematical transformation of separate wage equations and specifically contribute to the literature by examining the importance of different predictors on the discriminatory wage gap using the interaction coefficients.

Next, we choose two individuals, one male and one female, with the same predictors in $\boldsymbol{x}$. The difference of their log-wages is

$$\log w_M - \log w_F = \delta_0 + \boldsymbol{\delta}'\boldsymbol{x} + \epsilon_M - \epsilon_F. \tag{7}$$

If we had a sample of such matched pairs, we could estimate δ_0 and $\boldsymbol{\delta}$ directly. The interpretation of the j-th component of δ_j, $j > 1$, is that a unit change in x_j predicts the multiplicative change e^{δ_j} in the wage gap between males and females, provided other predictors remain unchanged. Unfortunately, we rarely have access to such data. However, we are inclined to interpret the interaction coefficients of (6) in the same fashion.

Accordingly, in order to weight the *relative* importance of different predictors, we can employ the recommendation of Cox (1992, p. 294) in using the ratios δ_j/δ_k. The predicted change in the wage gap from the unit change in x_j is equal to the change of δ_j/δ_k units in x_k. If both x_j and x_k are in logarithms, $x_j = logz_j$ and $x_k = logz_k$, then the interpretation can be done in terms of elasticities. To be specific, let $\delta_j = 0.50$ and $\delta_k = 0.25$. Then, it is approximately true that a 1 % increase in z_j predicts an increase in the wage gap that is equal to 2 % increase in z_k, with other predictors being constant. We can also rephrase this by saying that a 2 decrease in z_k compensates for the 1 % increase in z_j.

In all, the interaction coefficients provide information on the discriminatory male-female wage gap, conditional on other predictors being unchanged. Accordingly, the ratios of the interaction coefficients provide detailed information on both the relative and economical importance of specific predictors on the wage gap.

3. Empirical Application and Main Results

Data and Variables

In an empirical application, we study the gender wage gap in Finland. The methodology is implemented using the Finnish Structure of Earnings Survey (SES), which describes the employees' hourly and monthly earnings and employment relationships (full-time vs. part-time) in all employer sectors.[2] Only small firms (with fewer than 5 employees) are excluded from the survey. Information on wages has been linked to various total registers of Statistics Finland. These include Business Register, which covers all firms, corporations and self-employed individuals who are liable to pay value added tax or have employees. Business Register provides information on such items as branches of industry, size categories of personnel and turnover and the age of firm. Accordingly, these data sets are merged with the total individual records on education, family structure, occupation, nativity and age.

The total sample for the year 2010 includes 1,557,695 wage earners. In the subsequent analysis, we focus on full-time and highly skilled white-collar workers between the ages of 18 and 64. These individuals were employed by private business sector firms in 2010 with positive hourly earnings. After these restrictions, the original number of wage earners decreases to 432,059. Finally, those wage earners that have missing information on important background variables are also excluded, resulting in a remainder of 385,897 observations in the final sample. Of these, 135,749 (35 %) are females, and 250,148 (65 %) are males. The lower share of females in the sample is reasonable, as one-half of the female labor force works in the public sector.

[2] The data description can be found at: http://www.tilastokeskus.fi/meta/til/pra_en.html

The dependent variable is defined as the logarithm of the individual's hourly gross earnings. The predictors include potential work experience and its squared term. This is calculated as (2010 – year of graduation). If an employee has completed primary education only, potential work experience is calculated as (Age - 16). Years of education variable is calculated from the completed education level based on Statistics Finland recommendations.[3] Other characteristics include dummies for marital status (1 = married), children (1 = children under 7 years old), nativity (1 = native language is Finnish or Swedish), three occupation dummies that are constructed by the 1-digit level ISCO-08 occupation variable (managerial, professional and technical worker groups), one dummy for the metropolitan area of Helsinki (1 = Uusimaa region), three industry dummies (manufacturing, construction and services), dummies for large firms (1 = a firm with more than 300 employees) and foreign-owned firms (1 = at least 50 % of firm shares are foreign-owned) as well as the age of firm.

Table 1 provides summary statistics for the variables. Males earn 16 % more on average (calculated by the geometric means) than females do. This gap is partially explained by gender segregation in job. Of males, 11 % work as managers, compared to 7 % of females. Moreover, 52 % of females perform technical work, compared to 43 % of males. This indicates that more demanding white-collar jobs are primarily occupied by males, although we report roughly equal means in general human capital variables (education and experience). Gender differences in firm characteristics mainly reveal that more females work in the low-paid service sector than males. Males are also more likely to be employed in larger and foreign-owned firms than females.

Table 1. Sample statistics by gender

	Females	Males

[3] Classification: 9 years for primary education, 12 years for lower secondary education, 14 years for the lowest level of tertiary education, 16 years for a lower-degree level of tertiary education, 18 years for a higher-degree level of tertiary education and 21 years for a doctorate or equivalent level of tertiary education.

	Mean	Std.dev.	Mean	Std.dev.
Hourly wage, euro	24.51	9.05	28.44	10.96
Log hourly wage, euro	3.14	0.33	3.29	0.33
Age	42.5	9.55	42.4	9.66
Education years	13.7	2.39	13.8	2.45
Work experience, years	19.3	10.66	19.1	10.37
Small children, dummy	0.19	0.39	0.28	0.45
Married, dummy	0.55	0.50	0.64	0.48
Native, dummy	0.98	0.12	0.98	0.13
Metropolitan area, dummy	0.53	0.50	0.43	0.50
Occupation group (white collars)				
Manager, dummy	0.07	0.26	0.11	0.32
Professional, dummy	0.41	0.49	0.46	0.50
Technical, dummy	0.52	0.50	0.43	0.50
Industry				
Manufacturing, dummy	0.37	0.48	0.47	0.50
Construction, dummy	0.03	0.18	0.08	0.27
Services, dummy	0.60	0.50	0.45	0.50
Large firm, dummy	0.61	0.49	0.65	0.48
Foreign firm, dummy	0.25	0.44	0.28	0.45
Age of firm	30.6	32.2	29.5	32.5
Sample size	135,749		250,148	

Inference on the Gender Wage Gap

Table 2 gives the BO decomposition in the logarithmic scale. Computations are performed with the help of the implementation of Jann (2008) in Stata. The mean of the dependent variable, i.e., the average male-female wage gap of 16 % is decomposed into explained and unexplained parts. Under the female wage structure, we find by taking antilogarithms that the explained wage gap is ~3 %. According to our model, if males were paid at the same rate as females, they should have 3 % higher wages rather than 16 % higher wages. The unexplained wage gap is 13 % ($e^{0.12} - 1 \approx 0.13$). The traditional interpretation of the BO decomposition is that under the female wage structure 100 x 0.12/0.15 = 80 % of the observed mean wage gap is due to discrimination, and only 20 % is explained by the differences in the background variables. The corresponding figures under the male wage structure show roughly similar results.

Our result that the unexplained male-female wage gap is 13 % is consistent with Arulampalam, Booth and Bryan (2007), who examined the gender wage gap for 11 European countries, including Finland. The estimate is nevertheless higher compared to Korkeamäki and Kyyrä (2006), who used data on the manufacturing sector and found that white-collar females are paid approximately 6 % less than their equally qualified male coworkers. Napari (2008) examined the gender wage gap for highly educated private sector employees and found that females earn approximately 22 % less than males.

Table 2. BO-decomposition under female and male wage structures

	Log-wages	Std.Err.

Males	3.29	0.001
Females	3.14	0.001
Difference	0.15	0.001
Female wage structure		
Explained	0.03	0.001
Unexplained	0.12	0.001
Male wage structure		
Explained	0.03	0.001
Unexplained	0.12	0.001
Sample size	385,897	

The individual interaction coefficients are reported in Table 3. Recall that the interactions are related to the male-female gap such that the positive coefficients correspond to an advantage for males and the negative ones to an advantage for females. The interaction coefficients are statistically significant, apart from foreign firm, reasserting that the predictors give a different wage structure for males and females. Furthermore, the interaction coefficients provide detailed quantitative information on how the wage gap is related to different predictors. The interpretations are of a predictive nature or are mean differences between various groups. We avoid making any causal conclusions here. The problems of causal inference about wage discrimination are discussed by Fienberg and Haviland (2003) and Pearl (2003).

The last column of Table 3 is computed from the interaction coefficients (δ_j) by the formula $100 \times \left(e^{\delta_j} - 1\right)\%$. We find that the wage gaps in managerial and professional jobs are 4.4 % and 4.6 % narrower, respectively, than those in technical jobs. The result related to the task characteristics could indicate that females in more demanding jobs are more productive than males. The interpretation adheres to the study on Finnish metalworkers by Pekkarinen and

Vartiainen (2006). They used personal bonuses as a measure of individual productivity, and they found that promoted females were consistently more productive than their male counterparts. Cobb-Clark (2001) also found that although males in the United States had higher promotion rates, females had higher wage increases attached to promotions. Additionally, the gap between those males and females who live in the metropolitan area of Helsinki is -2.5 % narrower than for those who live outside the metropolitan area of Helsinki. This result is comparable with studies demonstrating that the gender wage gap is lower in large metropolitan areas than in rural areas (e.g., Hirsch, Köning and Möller, 2013). This is partially due to increased employment opportunities for highly qualified women in cities. Accordingly, the gap between native males and females is 3.1 % lower than the gap between nonnative males and females. Unlike previously reported, the male-female wage gap is lower (-1.3 %) in large firms than in smaller firms (e.g., Mitra, 2003).

Many of the predictors exhibit a relative advantage for males, most notably within family background and industry characteristics. The gap between married males and females is 3 % wider than between unmarried males and females. The gap between males and females who have small children is 3.3 % wider than between those who do not have small children. These results support the finding of Davies and Pierre (2005) that there is a persistent gap, the family gap, that disadvantages women. Their analysis is based on data from several European countries. It is also often found that being married is associated with higher earnings for men, e.g., see Pollmann-Schult (2011).

The wage gap is expected to be approximately 1.3 % wider in manufacturing and 4 % wider in services compared to the construction industry. Studying the gender wage gap in different industries, Fields and Wolff (1995) found that interindustry wage gaps between males and females explain a high fraction of the overall gender wage differential. Their data are from the United States. Gannon et al. (2007) have similarly analyzed the interaction between interindustry wage differentials and the gender wage gap in six European countries. Their

results show that combined industry effects may explain even over one-fourth of the gender wage gap.

Yet two interesting factors related to gender wage gap are general human capital variables, namely, years of education and work experience. Education is advantageous for females, and experience is advantageous for males, but the economic significance of these predictors is moderate. These results are in accordance with another study from Finland (Korkeamäki and Kyyrä, 2006). The moderate roles of education and experience in explaining the discriminatory male-female wage gap is reasonable given that females in Finland are both highly educated and strongly attached in the labor market. Accordingly, the minor contributions of human capital variables may be explained by the fact that our analysis focuses on highly skilled white-collar workers, so the variation in education and experience for this group is initially quite small (cf. Table 1).

Table 3. OLS results for single-equation model with interactions

Dependent variable: Log(hourly wage)	Main coefficients (female)	Interaction coefficients (male-female)	One unit contribution to wage discrimination
Male	0.119 (0.010)***		
Education	0.029 (0.000)***	-0.001 (0.000)**	-0.1 %
Experience	0.020 (0.000)***	0.003 (0.000)***	0.3 %
Experience^2/100	-0.033 (0.001)***	-0.002 (0.001)***	-0.2 %
Small children	-0.001 (0.002)	0.032 (0.002)***	3.3 %
Married	0.031 (0.001)***	0.030 (0.002)***	3.0 %
Native	0.071 (0.006) ***	-0.031 (0.007) ***	-3.1 %
Manager	0.522 (0.003)***	-0.045 (0.003)***	-4.4 %
Professional	0.204 (0.001)***	-0.047 (0.002)***	-4.6 %
Metropolitan	0.158 (0.001)***	-0.025 (0.002) ***	-2.5 %
Manufacturing	0.042 (0.004)***	0.013 (0.004)***	1.3 %
Services	0.028 (0.004)***	0.039 (0.004)***	4.0 %
Large firm	0.054 (0.001)***	-0.013 (0.002)***	-1.3 %
Foreign firm	0.045 (0.002)***	-0.004 (0.002)	-0.4 %
Age of firm	0.0004 (0.000)***	0.0001 (0.000)***	0.0 %
Constant	2.143 (0.009)***		
R^2	0.46		
Sample size	385,897		

Notes: Reference groups include technical workers and construction industry. Standard deviations are in parenthesis. Statistical significance at 1 and 5 percentage levels are denoted by *** and **.

Detailed BO Decomposition

Table 4 shows the detailed decomposition of the discriminatory wage gap according to formula (5). In the decomposition, the coefficients of the dummies are averaged to zero so that the contribution of a categorical predictor does not depend on the choice of the base category. Previously, we found that the most important factors explaining the gender wage gap in favor of men are related to family relations and employment industry. This finding is not supported by the results from Table 4, which suggest that the discriminatory wage gap is largely related to experience, although marital status and having small children also play a smaller role. These findings should be contrasted with previous findings. Because experience enters the model as a second-order polynomial, its interpretation is not quite straightforward. We find that the gap reaches its maximum at 75 years. To compare the importance of experience with other predictors, we computed how many years of experience, from the status of no experience at all, correspond to the difference between married and unmarried groups. The result is 10.8 years.[4] A comparison between the groups that have and do not have small children can be made similarly. This group difference corresponds to 11.5 years. This finding further stresses the importance of the family gap that is also observed in Table 4 but to a smaller extent compared to experience.

We next consider predictors that are advantageous for females. We previously found that the most important factors that are related to the wage gap in favor of women are occupation level within white-collar workers, native language and living in the metropolitan area of Helsinki. Again, the detailed decomposition results are not fully in line with our previous evidence. Although native language and living in the metropolitan area of Helsinki contribute to the overall discriminatory wage gap negatively (thus, in favor of females), we do not find such evidence for the occupation group. We have also mentioned that education only

[4] This difference is found by solving $0.030 = .003x - .002x^2/100$, where the interaction coefficient "Married" is set equal to the value of the quadratic experience function at *x*.

moderately contributes to the wage gap in our preferred model, which is based on interaction coefficients. This finding is not supported by the results from Table 4, where education accounts for -10.4 % of the discrimination (1.3 percentage points of the total discriminatory wage gap of ~13 %). The interpretation based on the interaction coefficients themselves shows that seven years of difference in education predicts the 1.0 % lower gap ($e^{7\times(-0.001)} = 0.993$). Compared to the contribution observed in Table 4, this is a much more moderate result.

We also find that the interaction coefficient of education is negative. The male-female wage gap among more educated persons is lower than among the less educated ones. Nevertheless, the difference is relatively small. We have observed that the gap between those having and not having small children is 3.3 %. Dividing the corresponding interaction coefficient (0.032) with the interaction coefficient of education (-0.001) yields minus 32. Thus, the 32-year difference in education in favor of females predicts as large wage gap in favor of males than the dummy related to having small children. Finally, the contributions of firm-level characteristics on the discriminatory wage gap are moderate in both models (Tables 3-4).

Table 4. Detailed decomposition analysis of log-wages

	log - Discrimination	Std. Err.	Contribution, %	Contribution, %-points
Education	-0.013	0.006	-10.4 %	-1.3 %
Experience	0.045	0.003	36.3 %	4.6 %
Small children	0.006	0.001	4.8 %	0.6 %
Married	0.017	0.001	13.7 %	1.7 %
Native	-0.031	0.007	-25.0 %	-3.1 %
Occupation group	0.008	0.001	6.5 %	0.8 %
Metropolitan	-0.013	0.001	-10.5 %	-1.3 %
Industry	0.011	0.001	8.9 %	1.1 %
Large firm	-0.008	0.002	-6.5 %	-0.8 %
Foreign firm	-0.001	0.001	-0.8%	-0.1 %
Age of firm	-0.003	0.001	-2.4 %	-0.3 %
Constant	0.106	0.010	85.5 %	11.2 %
Total discrimination	0.124	0.001	100.0%	13.2 %
Sample size	385,897			

Notes: Experience group includes experience and its squared term. Occupation group includes three occupation dummies (managers, professionals and technicals) and Industry group includes three industry dummies (manufacturing, construction and services).

4. Robustness Checks and Auxiliary Evidence

Alternative Outcome Variable

We have rerun the single-equation model with interactions and detailed decomposition estimations for our sample using alternative outcome variable. Instead of the logarithm of hourly wages, we used the logarithm of monthly wages. The average discriminatory wage gap using monthly wages as the outcome is 14 %. The results of the one-unit contributions on wage discrimination, which are based on interaction coefficients, are reported in Table A1 of the appendix (Column 2). Consistent with our earlier results in Table 3, and reported in Table A1 (Column 1), the unexplained wage gap in favor of men is mainly explained by marital status, having children and working industry. For example, the male-female wage gap for married individuals or for those having small children, *ceteris paribus*, is 2.9 % wider than that for unmarried ones or for those not having small children. In contrast, the male-female wage gaps in managerial and professional jobs are ~5.1 % lower than the male-female wage gap in technical jobs. The point estimates are highly comparable between the models using either the hourly or monthly wages as the outcome. The relative importance of different predictors using the ratios of interaction coefficients yield similar conclusions, as reported in Section 3.

The results of the detailed BO decomposition model are reported in Table A2 of the appendix (Column 2).[5] These results should be contrasted with our initial detailed decomposition findings presented in Table 4 and rereported in Table A2 (Column 1). The results are robust to the chosen outcome variable. The results indicate that the discriminatory wage gap in favor of men is mostly due to work experience (4.4 percentage points out of a total

[5] In order to improve the readability of Tables A1 and A2, we do not report the standard errors or the stars that flag the statistical significances of the estimates. We, however, note that all the estimates are statistically significant at least at the 5% significance level, except those marked with [a] (Foreign firm dummy in Column (1) of Tables A1-A2; Education in Column (4) of Tables A1-A2; and Health & services education field dummy in Column (4) of Table A1).

gap of 14 %) and to some extent to being married (1.6 percentage points out of a total gap of 14 %). In contrast, the wage penalty for females decreases mostly with nativity (-2.8 percentage points) and education (-2.1 percentage points).

Alternative Set of Covariates

We also evaluate the robustness of our main findings by estimating the models with two different sets of covariates. First, we included only personal characteristics in the models. The rationale behind this exercise is that we consider the direct contributions of general human capital variables (education and work experience), specific human capital variables (occupation) and demographic variables (marital status, nativity and having children) on wage discrimination that are not being confounded by self-selection into different types of firms or regions. The average discriminatory wage gap in this specification is 11.4 %. The results of the one-unit contributions obtained from the single-equation model are reported in Table A1 of the appendix (Column 3). The overall conclusions do not change much, although the discriminatory male-female wage gap for native speakers, managers and professionals decreases by approximately one percentage point compared to our main specification. These changes are, however, economically quite small.

The corresponding detailed decomposition results are reported in Table A2 (Column 3). We again find that experience is advantageous for males and that education and nativity are advantageous for females. This is not the message we find using interaction coefficients from a single-equation model and when we compare the relative importance of predictors using ratios of interaction coefficients.

Finally, we include additional covariates in the model. We add six dummies for the education field (general, humanities or arts, business or social sciences, natural sciences or technical, forestry, health or services). The education field is a good proxy for specific human

capital, outside the occupation group. We also include the firm's log of total turnover in the model to account for firms' heterogeneity in revenues and sales. The results are presented in Column (4) of Tables A1-A2, and the discriminatory wage gap is 14.6 %. The inclusion of additional covariates does not change the overall conclusions of the results, except that education is now statistically insignificant. This is reasonable given that the education level correlates with both the occupation group and education field. When we focus on the one-unit contributions based on interaction coefficients (Table A1), we find that the male-female wage gap is 3-4 % lower among individuals who have education in humanities, arts or natural sciences compared to those who have a general education field only. Males, in turn, seem to benefit if they have been educated in business or social sciences. Interestingly, the male-female wage gap is 9 % lower for those who have education in forestry compared to those who have a general education. The detailed decomposition results in Table A2 suggest that the education field in general accounts for 2.6 percentage points of the total discriminatory wage gap of 14.6 %. The contributions of other predictors remain basically unchanged.

Overall, we conclude that the analyses of the contribution of different predictors on wage gap using either our preferred single-equation model or detailed BO decomposition model are not sensitive to the chosen outcome variable or selection of observed characteristics.

5. Conclusions

In this paper, we propose to estimate the Blinder-Oaxaca decomposition by a single-equation model augmented with interactions. We have also shown how we can further utilize the interaction coefficients in comparisons, leading to the assessment of the relative importance of various predictors. Our application for the gender wage gap using administrative data from Finland reveals that the detailed decomposition and our preferred interpretations based on interaction coefficients may lead to very different conclusions.

Our preferred approach suggests that the discriminatory male-female wage gap of approximately 13 % can be mostly explained by the factors of having small children, being married and working industry. The results regarding family relations support the notion of a persistent "family gap" that disadvantages women (e.g., Davies and Pierre, 2005; Pollmann-Schult, 2011). Accordingly, interindustry wage gaps between males and females have been found to explain a high fraction of the overall gender wage gaps (e.g., Fields and Wolff, 1995; Gannon et al., 2007). We find, for example, that the male-female wage gap is larger in manufacturing than in the construction industry. This finding may be related to the puzzling relationship between globalization and gender inequality. A recent study from Norway shows that manufacturing firms' entry into exporting increases the gender pay gap by 3 % of highly educated workers (Bøler, Javorcik and Ulltveit-Moe, 2018).

In contrast, the detailed BO decomposition results suggest that experience in particular contributes to the wage gap that is disadvantageous for females and that having small children has only a modest role in explaining gender inequality. However, the relative importance of experience is quite modest in our preferred approach, as women would need 10-11 years more work experience to correspond to the wage difference between married and unmarried couples or between the groups of having and not having small children. Our preferred model thus yields detailed information on both the relative and economical importance of specific predictors on the wage gap. Our alternative method for the evaluation of gender wage gaps may be applicable to analyses of wage gaps between any two groups of workers, such as public and private sector workers.

References

Arulampalam, W., Booth, A. and Bryan, M. (2007).”Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution.” *Industrial and Labor Relations Review* 60, 163-186

Biewen, M. (2014). “A general decomposition formula with interaction effects.” *Applied Economics Letters* 21, 636-642.

Blinder, A. (1973). “Wage discrimination: reduced form and structural estimates.” *Journal of Human Resources* 8, 436–455.

Bøler, E., Javorcik, B. and Ullveit-Moe, K-H. (2018). “Working across time zones: Exporters and the gender wage gap. *Journal of International Economics* 111, 122-133.

Cobb-Clark, D. (2001). “Getting ahead: The determinants of and payoffs to internal promotion for young U.S. men and women.” In: Polachek, S. (Ed.), Worker Wellbeing in a Changing Labor Market. Vol. 20 of Research in Labor Economics. Emerald Group Publishing Ltd, Bingley, pp. 339–372.

Cotton, J. (1988). “On the decomposition of wage differentials.” *Review of Economic and Statistics* 70, 236–243.

Cox, D. (1992). “Causality. some statistical aspects.” *Journal of the Royal statistical Society, Ser. A* 155, 291–301.

Davies, R. and Pierre, G. (2005). “The family gap in pay in Europe: A cross-country study.” *Labour Economic*s 12, 469–486.

Elder, T., Goddeeris, J. and Haider, S. (2010). “Unexplained gaps and Oaxaca-Blinder decompositions.” *Labour Economics* 10, 284-290.

Fields, J. and Wolff, E. (1995). “Interindustry wage differentials and the gender wage gap.” *Industrial and Labor Relations Review* 49, 105–120.

Fienberg, S. and Haviland, A. (2003). “Discussion of Statistics and Causal Inference: A Review”, by J. Pearl. Test 12, 319–327.

Fortin, N., Lemieux, T. and Firpo, S. (2011). "Decomposition methods in economics". In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Vol. 4. Elsevier, Amsterdam, pp. 1–102.

Gannon, B., Plasman, R., Rycx, F. and Tojerow, I. (2007). "Inter-industry wage differentials and the gender wage gap: evidence from European countries." *The Economic and Social Review* 38, 135-155.

Gardeazabal, J. and Ugidos, A. (2004). "More on identification in detailed wage decompositions." *Review of Economic and Statistics* 86, 1034–1036.

Gelbach, J. (2016). "When do covariates matter? And which ones, and how much?" *Journal of Labour Economics* 34, 509-543.

Hirsch, B., Köning, M. and Möller, J. (2013). "Is there a gap in the gap? Regional differences in the gender wage gap." *Scottisch Journal of Political Economy* 60, 412-439.

Jann, B. (2008). "A Stata implementation of the Blinder-Oaxaca decomposition." *The Stata Journal* 8, 453–479.

Korkeamäki, O. and Kyyrä, T. (2006). "A gender wage gap decomposition for matched employer-employee data." *Labour Economics* 13, 611–638

Mitra, A. (2003). "Establishment size, employment, and the gender wage gap." *Journal of Socio-Economics* 32, 317-330.

Napari, S. (2008). "The early-career gender wage gap among university graduates in the Finnish private sector." *Labour* 22, 697-733.

Neumark, D. (1988). "Employer's discriminatory behaviour and the estimation of wage discrimination." *Journal of Human Resources* 23, 279–295.

Oaxaca, R. (1973). "Male-female wage differential in urban labor markets." *International Economic Review* 140, 693–709.

Oaxaca, R. and Ransom, M. (1999). "Identification in detailed wage decompositions." *Review of Economics and Statistics* 81, 154–157.

Pearl, J. (2003). “Statistics and causal inference: A review”. *Test* 12, 281–345.

Pekkarinen, T. and Vartiainen, J. (2006). ”Gender differences in promotion on a job ladder: Evidence from Finnish metalworkers.” *Industrial and Labor Relations Review* 59, 285–301.

Pollmann-Schult, M. (2011). “Marriage and earnings: Why do married men earn more than single men?” *European Sociological Review* 27, 147–163.

Reimers, C. (1983). “Labor market discrimination against hispanic and black men.” *Review of Economics and Statistics* 65, 570–579.

Słoczyński, T. (2013). “Population average gender effects.” IZA discussion papers, No. 7315.

Yun, M.-S. (2005). “A simple solution to the identification problem in detailed wage decompositions.” *Economic Inquiry* 43, 766–772.

Yun, M.-S. (2008). “Identification problem and detailed Oaxaca decomposition: A general solution and inference.” *Journal of Economic and Social Measurement* 33, 27–38.

Table A1. One-unit contribution to wage discrimination: different specifications

	(1)	(2)	(3)	(4)

	Log(Hourly wage)	Log(Monthly wage)	Log(Hourly wage)	Log(Hourly wage)
Education	-0.1 %	-0.1 %	-0.1 %	-0.1 % [a]
Humanities & arts				-3.7 %
Business & social sc.				3.4 %
Natural sc. & technical				-3.5 %
Forestry				-9.3 %
Health & services				-0.1 % [a]
Experience	0.3 %	0.3 %	0.3 %	0.4 %
Experience^2/100	-0.2 %	-0.2 %	-0.3 %	-0.4 %
Small children	3.3 %	2.9 %	3.3 %	2.8 %
Married	3.0 %	2.9 %	3.3 %	3.2 %
Native	-3.1 %	-2.8 %	-4.2 %	-3.8 %
Manager	-4.4 %	-5.2 %	-5.5 %	-5.5 %
Professional	-4.6 %	-5.1 %	-5.9 %	-5.2 %
Metropolitan	-2.5 %	-2.4 %		-3.2 %
Manufacturing	1.3 %	1.4 %		1.2 %
Services	4.0 %	3.9 %		2.3 %
Large firm	-1.3 %	-1.1 %		-0.7 %
Foreign firm	-0.4 % [a]	-0.5 %		-0.7 %
Age of firm	0.0 %	0.0 %		0.0
Log of turnover				-0.1 %
R^2	0.46	0.43	0.40	0.47
Sample size	385,897	385,897	385,897	385,897

Notes: Reference groups include General or other education field, technical workers and construction industry. *** and ** indicate that one-unit contribution to wage discrimination is statistically significant at least at the 1% or 5% significance levels, respectively.

Table A2. Detailed %-point contribution of predictors on wage discrimination: different specifications

	(1)	(2)	(3)	(4)
	Log(Hourly wage)	Log(Monthly wage)	Log(Hourly wage)	Log(Hourly wage)
Education	-1.3 %	-2.1 %	-1.7 %	-1.0 % [a]
Field of education				2.6 %
Experience	4.6 %	4.4 %	4.7 %	5.4 %
Small children	0.6 %	0.6 %	0.6 %	0.5 %
Married	1.7 %	1.6 %	1.8 %	1.8 %
Native	-3.1 %	-2.8 %	-4.0 %	-3.7 %
Occupation group	0.8 %	1.0 %	1.0 %	1.0 %
Metropolitan	-1.3 %	-1.2 %		-1.7 %
Industry	1.1 %	1.1 %		0.7 %
Large firm	-0.8 %	-0.7 %		-0.4 %
Foreign firm	-0.1 % [a]	-0.1 %		-0.2 %
Age of firm	-0.3 %	-0.5 %		-0.4 %
Log of turnover				-0.9 %
Constant	11.2 %	12.6 %	9.0 %	10.6 %
Total discrimination	13.2 %	14.0 %	11.4 %	14.6 %
Sample size	385,897	385,897	385,897	385,897

Notes: Experience group includes experience and its squared term. Occupation group includes three occupation variables (managers, professionals and technicals), Industry group includes three industry variables (construction, manufacturing and services) Field of education group includes six education field variables (general, humanities or arts, business or social sciences, natural or technical sciences, forestry, and health or services). All the estimates are statistically significant at least at the 5% significance level, except those marked with [a] (Foreign firm in Column (1) and Education in Column (4)).