Segregation and the gender wage gaps in the private and the public sector

Albæk, Karsten; Larsen, Mona; Stage Thomsen, Lars

Published in:
Empirical Economics

DOI:
10.1007/s00181-016-1132-2

Publication date:
2016

Document Version
Peer-review version

Citation for pulished version (APA):
Albæk, K., Larsen, M., & Stage Thomsen, L. (2016). Segregation and the gender wage gaps in the private and the public sector. Empirical Economics. 10.1007/s00181-016-1132-2

Segregation and gender wage gaps in the private and the public sectors – an analysis of Danish linked employer-employee data, 2002-2012

Accepted version

June 8, 2016

Karsten Albæk

Mona Larsen

Lars Stage Thomsen

Abstract: This paper examines the relation between segregation and the gender wage gap in the public and the private sectors in Denmark from 2002 to 2012. The analysis shows that male-female differences in the share of females in occupations, industries, establishments and job cells (occupations within establishments) constitute 46 per cent of the raw gender wage gap in the private sector, while segregation in the public sector accounts for as much as 63 per cent. Segregation thus plays a substantially more important role in accounting for the gender wage gap in the public sector than in the private sector. While the importance of segregation for wage formation decreased substantially in the public sector over time, it only decreased slightly in the private sector. Although the remaining gender wage gap, after controlling for segregation, is close to zero in the public sector, a substantial within-job cell differential remains after controlling for segregation in the private sector.

Keywords: Wage gap; gender; segregation; public sector.
JEL classification: J31

Acknowledgements: For comments on previous versions of the paper we thank the editor, two referees, Colin Green, the participants in the workshop in honour of Professor Peder Pedersen, Århus University, April 2012, the meeting of the European Society for Population Economics in Bern, June 2012, the Nordic Econometric Meeting in Bergen, June 2013, the 3rd Linked Employer-Employee Data Workshop in Lisbon, June 2013, and the European Meeting of the Econometric Society in Gothenburg, August 2013. Financial support from the Danish Council for Independent Research, Social Sciences is acknowledged.

K. Albæk, M. Larsen, L.S. Thomsen. SFI – The Danish National Centre for Social Research, Herluf Trolles Gade 11, DK-1053 Copenhagen K. Corresponding author: Albæk , email: kal@sfi.dk, tlf.: +45 22 473 473.

1. Introduction

The difference in pay between men and women is a topic of substantial interest, with equal pay legislation implemented in many countries, including the US and the European Union (EU). In the EU, the ‘principle of equal pay’ between men and women is valid not only for ‘the same work’ but also for ‘work to which equal value is attributed’ (European Commission (1975), Article 1). This principle of ‘comparable worth’ is also implemented in Australia (Blau and Kahn (1999)), Canada (Baker and Fortin (2004)) and in some US states (Killingsworth (1990)). This type of equal pay legislation may entail a comparison of wages and gender composition among different occupations.

This paper examines the relation between gender differences in wage and segregation in Denmark during a decade from 2002 to 2012 using a matched employer-employee data set. We explore the role of gender composition in wage setting along various dimensions such as occupational groups, industry, establishments and occupations within establishments (‘job cells’).

A main purpose of this study is to analyse how and how much wages vary with gender composition. This is an issue of direct policy relevance as policy intervention in the form of ‘comparable worth’ or ‘work to which equal value is attributed’ may entail comparisons of wage levels between job cells with high and low shares of females.

Occupational segregation plays a prominent role in the literature on gender differences in wages (see, e.g. Blau and Kahn (2000)). The literature on the topic includes Groshen (1991), who finds that occupational segregation plays a major role for differences in pay between males and females but only negligible wage differentials within-job cells. Groshen (1991) analyses five manufacturing industries with sample sizes ranging from 14,520 to 70,355 (the data is from the Bureau of Labor Statistics Industry Occupational Wage Surveys).

In contrast, more recent research on matched employer-employee data sets finds that a substantial share of the wage gap remains within-job cells, as in Bayard *et al.* (2003), who analyse a matched employer-employee data set with 637,718 workers in 1990 in the US business sector. Similar results are also found in analyses of matched employer-employee data in other countries. Datta Gupta & Rothstein (2005) analyse segregation in the private sector in Denmark for 313,099 salaried and 260,590 manual workers in 1995. Korkeamaki and Kyyra (2006) analyse gender differences in pay for 148,944 white-collar workers in manufacturing for Finland in 2000. Ludsteck (2014) analyses gender wage differences for workers in Western Germany from 1990 to 2005 on a panel data set containing about 14 million job cell observations.[1]

In contrast to previous studies, we investigate whether segregation in the public sector plays the same role for gender wage differences in pay as in the private sector. We analyse this research question by comparing the results of regression analyses performed separately for the private and the public sectors. In Scandinavia, the public sector has been the focus of the debate about the implementation of the principle of 'comparable worth'; as discussed by the Equal Pay Commission in Norway and the Wage Commission in Denmark (Lønkommissionen (2010)).

As the public sector constitutes a substantial share of the labour market in developed countries, the results are of interest in their own right. The sources of possible differences between the private and the public sectors include that wage formation mechanisms might differ, for example in the degree of unionization.

The EU legislation including the 'principle of equal pay' and 'work to which equal value is attributed' is implemented by law in Denmark ('Law of equal pay to men and women' of 1976 with subsequent amendments). The enforcement of the provisions of equal pay takes place in the court system. In the decisions on cases of equal pay, the judges take into consideration circumstances (or variables) that are included in the analysis in this paper such as the type of work, education, experience and tenure.[2] The provisions of equal pay apply only to employees who work for the same employer (the comparison workers have to be employed by the same firm).

[1] Whether the sample includes public sector establishments is unclear.

[2] See Bruun (1994). The impression is that, with respect to 'type of work', judges go into more detail regarding the relative amount of different tasks that men and women perform than the categorization of occupations applied by statistical agencies such as Statistics Denmark. Thus, gender wage gaps within job cells do not necessarily imply that a judge will decide that firms breach antidiscrimination laws. A large share of the cases takes place within the Danish labour court system, which resolves issues related to workers covered by collective agreements.

The remainder of the paper is organised as follows. Section 2 presents the data used in the study. Section 3 discusses the methodology, including an exposition of how to decompose the raw gender wage gap into different sources. Section 4 gives the results of the multivariate analysis for the private and the public sectors. Section 5 decomposes the gender wage gap according to reward to characteristics and differences in characteristics. Section 6 provides robustness checks of the estimates. Section 7 concludes.

2. The data

The data we use are a matched employer-employee data set encompassing a large fraction of the workers in the Danish labour market during 2002-2012. The matched data, which are obtained from Statistics Denmark, consist of information from several administrative registers.

We apply hourly wage measures reported by employers. Our hourly wage information is obtained from records of wages for individual workers from the private, the state and the local government sectors. These data, collected for producing wage statistics, include hourly wage rates, components of the wage and working hours (including separate information about normal working hours, actual working hours and overtime). Our wage measure is more precise than wage rates imputed from yearly wage income divided by an estimate of the annual working hours, which is the measure used for wage analysis of Danish data for earlier years in other studies (e.g. Datta Gupta & Rothstein (2005)).

For the private sector, our data set includes information from establishments with at least 10 full-time employees, except establishments in the agricultural and fishing sectors.[3] For the public sector, our data include information about all employees, except specific categories such as military conscripts, some temporary teachers and student assistants. Moreover, our data cover employees only when the employment relation

[3] Although firms in the agricultural and fishing sectors and private firms with fewer than 10 full-time employees are exempt from reporting employee wages to Statistics Denmark, some of these firms have nonetheless reported.

lasts more than one month and when the average weekly working hours are at least 8 hours. Furthermore, our data only include employees employed on ‘ordinary’ conditions. Various minor groups are thus omitted from the register (e.g. employees paid at an unusually low rate because of a disability). Included in the statistics, however, are employees for whom the employer receives an employment subsidy from the government (the wage statistics do not permit identification of these workers). We restrict our sample to workers aged 16-64 years.

To conduct our analyses, we need information about each employee’s establishment. In our data, establishments are defined as collections of workers within a firm working at the same physical address. The criteria for constructing establishments are the same in both the private and the public sectors. Firms (or agencies in the public sector) also provide the information used to distribute workers into geographical units. We restrict the sample to workers employed in establishments with at least 10 employees.

For each employee, we have information about the occupation, which plays a central role in this study. The classification scheme is the ‘International Standard Classification of Occupations’ (ISCO, or, more precisely, a Danish variant, DISCO). The most detailed level of registration is the 6-digit level. For example 723110 is ‘automobile mechanic work with cars and vans’, while 723120 is ‘automobile mechanic work with trucks’. For each of the 6-digit occupations we calculate the share of female workers, for the private and the public sectors. We also calculate the share of females within industrial categories and the share of females in establishments. Finally we categorise the workforce in each establishment according to 6-digit occupations and calculate the share of females for each of these ‘job cells’. We restrict the sample to employees in 6-digit occupations with at least 20 workers.

We include various other variables of relevance for explaining wage differentials. To control for education, we include length of education in years calculated from the normal education length for the employees’ highest completed educational level. We use a measure of the individual employees’ actual work experience, the number of years the employee has been in the labour market, calculated from contributions to a pension scheme. Employee tenure is calculated as the number of years that the worker has been employed in the firm. Furthermore, we include dummy variables for residence in Capital and for living together with a partner. In a few cases not all workers at an establishment belong to the same sector or industry. In these cases we allocate all workers in the plants to the sector or industry to which the majority belong.

We confine the analysis to the employer-employee observation for workers employed in November each year, as the data for these workers contains information about experience and tenure. We exclude observations with unknown occupations and the occupational categories 'pilots' and 'air traffic controllers' (due to lack of credible information on length of education). Furthermore, we exclude observations with missing values of the variables.

We have a sample selection problem for some years during the sample period because not all firms have consistently reported reliable information about the distribution of workers on distinct geographical units. For the period 2007-2009 plant information is lacking for some parts the public sector following an administrative reform that took place in 2007 and included merging most of the municipalities in Denmark.[4] In the private sector there is a lack of information during 2002-2004 compared to 2005-2006 – the attrition due to lack of plant information is about one third during 2002-2004 compared to one fifth during 2005-2006. The likely reason for the relatively large attrition in the private sector during 2002-2004 is that the coverage of the statistics with respect to plant information has increased over time as more employers have provided reliable data.[5] However, there is no indication of attrition problems for the last years in the sample period, 2010-2012. We return to this issue in Section 6, where we perform decompositions of the gender wage gap and report the number of observations in sub-periods for both the private and the public sectors.

Our data contain more than one measure of wages. Our main measure is the 'gross wage', which is total earnings (including payment for holidays and absence) divided by

[4] Distribution of plants and workers on four sectors (private, state, provincial government and municipalities) is possible from 2009 and shows a substantial lack of workers and plants in municipalities in 2009 compared to 2010-2012.

[5] According to Statistics Denmark the wage statistics at the industry level is based on data from firms with about 90 per cent of the employed workers in the target population for the private sector during 2002-2006 (see 'Statistiske Efterretninger', various issues). Statistics Denmark does not publish wage statistics on the plant level and does not provide information about the coverage of the wage statistics in this respect.

the number of working hours (excluding hours spent on holidays and absence). This measure is the only one that exists in the wage statistics in our data from 2002 and throughout the entire estimation period to 2012. The gross wage is the only wage concept used by Statistics Denmark to publish official wage statistics during the period 2002-09.

Statistics Denmark continues to use the gross wage in official wage statistics but from 2010 an alternative measure, the 'standardised wage', is included in the assessments. The 'standardised wage' does not include payments for paid absence in the numerator and does not include absent hours in the denominator (but includes payments for holidays in the numerator and hours on holidays in the denominator). As women generally have higher absenteeism than men, the female-male wage ratio using the gross wage is thus higher than the ratio using the standardised wage. Both wage measures are used in analyses of wage formation in Denmark.[6] We supplement our main analysis of the gross hourly wage with a robustness check, in which we apply the standardised hourly wage for 2010-2012.

We trim our hourly wage data in both the lower and the upper tail of the wage distribution. In the lower tail we omit observations with wage rates below 80 per cent of the minimum wage; in the upper tail we omit the 0.1 per cent of observations with the highest wage rates.[7]

The data used for the paper is the same data that Statistics Denmark uses for labour force assessments on the basis of administrative registers. The number of employed workers in November 2012 was 2,461,722 according to our count of the relevant employment categories in the official statistics.[8] A count of the same employment categories in the data used for this paper yields exactly the same figure. The number of workers in the estimation sample for this paper is 1,303,447 and the sample thus consti-

[6] The actual calculation of the standardized hourly wage is somewhat elaborate as different wage components are normalized by different measures of working hours, see the examples in Pedersen (2010), p. 21-22. A main aim of the construction of the standardized wage is to obtain a measure that is close to what is 'agreed upon' between employers and employees (i.e. independent of the amount of work, including the amount of absenteeism). Larsen & Houlberg (2013) contains a detailed discussion of the various wage measures in the Danish wage statistics.

[7] For detailed information about the Danish wage statistics see the home page of Statistics Denmark: http://www.dst.dk/da/Statistik/dokumentation/Times/loenstatistik.aspx (accessed February 20, 2016).

[8] The online source 'Statistikbanken' at Statistics Denmark, see http://www.dst.dk/da/Statistik/statistikbanken (accessed February 20, 2016).

tutes 53 per cent of the total population of employed workers in November 2012. The reasons for the difference include the age restriction, that not all employed workers are included in the wage statistics, the limitation of the sample to plants with at least 10 workers and lack of covariates for some workers.

The last three years of the sample period, 2010-2012, contain, on average, 543 occupations, 606 industries, 26,191 establishments and 189,372 job cells per year. The dispersion between the years is very limited. The corresponding numbers for some of the previous years are smaller because of attrition. The following conveys information about the size distributions during 2010-2012: The average occupation size is 2082 workers (the median is 613, the 25% percentile is 164 and the 75% percentile is 1955 workers). The average industry size is 2145 workers (median 466, 25% percentile 137, 75% percentile 1660). The average establishment size is 50 workers (median 22, 25% percentile 14, 75% percentile 43), and the average job cell size is 7 workers (median 2, 25% percentile 1, 75% percentile 5). For job cells containing only one worker, the share variable takes the value of the female dummy, which might imply that a part of the gender wage differential not explained by background characteristics will be attributed to segregation (we include a robustness check of this issue in Section 6).

Descriptive statistics for the sample appear in Table 1, which contains information for the private and the public sectors, separately. The information in Table 1 thus gives the background for the analysis.

Table 1 around here

The data include 7,501,564 observations of workers for the private sector, with a female share at 37.7 per cent, and 5,671,717 observations of workers in the public sector, with a female share at 70.7 per cent. The gender wage gap is 13.7 per cent in the private sector and 11.4 per cent in the public sector[9]. Workers in the public sector are,

[9] We adopt the convention that a difference of, for example 0.137 log points is stated as a percentage difference. We have adjusted all wages to the 2012 level by regressing wages on year dummies and adding the value of the coefficients on the year dummies to the wage observations for each year. The regressions in the following contain dummies for the years before 2012.

on average, paid slightly more than workers in the private sector, and this is the case for both males and females. Part of the reason why the wage rate is higher in the public sector is a higher educational level among public sector workers, who on average have about one more year of schooling than private sector workers. Public sector employees also have more experience in the labour market and longer tenure at their firm.

In both the private and the public sectors, males have longer education than females (the difference is 0.2 years of schooling in the private sector and 0.6 years in the public sector). Males also have longer experience in the labour market than females in both sectors (about 1.2 years in the private sector and 0.4 years in the public sector) and they have longer tenure (the difference is 0.4 years in the private sector and 0.7 years in the public sector). There are minor differences between males and females with respect to residence in the Copenhagen metropolitan area and to being single.

We introduce the following notation: we denote the share of females in the 6-digit occupation that individual i belongs to by F_i^O. In the private sector the average share of females in 6-digit occupations is 57 per cent for females (the average of variable F_i^O for females, denoted $\bar{F}_f^O$) and 26 per cent for males (the average of variable F_i^O for males, denoted $\bar{F}_m^O$), indicating a substantial segregation in the labour market.

Segregation in industries in the private sector is smaller, as the average share of females for females in industries is 48 per cent and the average share for males is 32 per cent, yielding a difference of 17 percentage points. Females are employed in establishments with an average share of 52 per cent females, while men are employed in establishments, where, on average, 29 per cent of the work force are females, resulting in a difference of 23 percentage points. We also categorise the workforce in each establishment according to 6-digit occupations and calculate the share of females for each of these ‘job cells’. Table 1 shows that females work in job cells with 69 per cent females on average, while males work in job cells with only 19 per cent females, yielding a difference of 50 percentage points. The value of the index for job cells is the highest among the four indices of segregation in the labour market.

The amount of segregation in the two sectors appears to be of the same magnitude, as measured by the difference in the average share of female workers between males and females. Segregation in occupations and industries is slightly higher in the public than in the private sector. Segregation in establishments is about the same in the two sectors, and segregation in job cells is slightly less in the public than in the private sector. The 'femaleness' of the public sector shows up in a high value of the job cell index for females: the average share of females in a job cell for publicly employed women is 85 per cent.

3. Methodology

This section presents the methodology we use in our analysis of the importance of segregation for the gender wage gap in the private and public sector, respectively. Furthermore, we describe how we decompose the wage differential between men and women into parts that can be ascribed to coefficients (or rewards to characteristics) and differences in characteristics.

Our point of departure in the empirical analysis is the model

$$w_{it} = \alpha + \beta d_i + \gamma_O F_{it}^O + \gamma_I F_{it}^I + \gamma_E F_{it}^E + \gamma_J F_{it}^J + \delta X_{it} + \sum_t \pi_t D_t + e_{it}, \qquad (1)$$

where w_{it} is log wage of worker i in year t, d_i is the gender dummy which takes the value 1 if the worker is female and 0 if the worker is male, F_{it}^O is the share of females in the occupation of worker i in year t, F_{it}^I is the share of females in the industry of worker i in year t, F_{it}^E is the share of females in the establishment of worker i in year t, F_{it}^J is the share of females in the job cell of worker i in year t, X_{it} is other explanatory variables for worker i in year t, D_t is a dummy variable which takes the value 1 for year t and 0 otherwise, and e_{it} is the error term. By assumption the error term e_{it} is uncorrelated with the explanatory variables.

We estimate variants of equation (1). Parameter β, the coefficient of the female dummy, is the gender wage gap, conditional on the covariates. Parameters γ_O, γ_I, γ_E, and γ_J measure the relation between wages and the share of females in occupations, industries, establishments and job cells, and vector δ captures the effect of the remaining covariates. The year dummies, π_t, adjust the yearly wage levels to a common level (that of 2012).

As a supplement to the estimations with the female share of workers we perform estimations where fixed effects for occupations, industries, establishments or job cells enter instead of the share variables. The interpretation of regressions including fixed effects for job cells, for example, is that the β estimate is an estimate of the gender wage gap within-job cells.

A result from regression algebra is that average wages equal the sum of coefficients multiplied by the average of the covariates, both for all workers in the sample and for the subgroups of males and females. We can thus write the difference between the average male wage, $\bar{w}_m$, and the average female wage, $\bar{w}_f$, as

$$\begin{aligned}\bar{w}_m - \bar{w}_f = \beta + \gamma_O\left(\bar{F}_m^O - \bar{F}_f^O\right) + \gamma_I\left(\bar{F}_m^I - \bar{F}_f^I\right) + \gamma_E\left(\bar{F}_m^E - \bar{F}_f^E\right) + \gamma_J\left(\bar{F}_m^J - \bar{F}_f^J\right) \\ + \left(\bar{X}_m - \bar{X}_f\right)\delta'. \end{aligned} \tag{2}$$

The left-hand side is the raw gender wage gap. The first term on the right-hand side, β, is the gender wage gap after taking the impact of covariates into account. The second term on the right-hand side is the coefficient of the female share of workers in occupations, γ_O, multiplied by the difference between the average male share of females in occupations, $\bar{F}_m^O$, and the average female share of females in occupations, $\bar{F}_f^O$. This term is thus the amount of the gender wage gap that is attributed to the different sorting of males and females in occupations, given the parameter estimate, γ_O, of how much femaleness of occupations is associated with a wage penalty. Similarly, the third, fourth and fifth terms are the shares of the gender wage gap that are attributed to the different sorting of male and females in industries, establishments and job cells. The last term on the right-hand side is the share of the raw gender wage gap that can be ascribed to differences in other characteristics between males and females.

The difference between the average wage for males and females on the left-hand side of equation (2) is thus written as the sum of differences in characteristics multiplied by the coefficients (the ‘explained’ part of the raw gender wage gap or the ‘composition’ effect), and the first term on the right-hand side, β, the estimated gender wage gap, conditional on the covariates (the ‘unexplained’ part of the raw gender wage gap).

The differences between the male average and the female average for the variables in the analysis appear in the last column of Table 1 (e.g. $\bar{F}_m^O - \bar{F}_f^O$ for the private sector is -0.306). In Section 4, we present the estimates of the parameters in equation (1). In Section 5, we present the results of the decomposition in equation (2) using the male-females differences in characteristics of Table 1 and the parameter estimates obtained in Section 4.

The standard errors associated with the parameters in the regressions are cluster-robust standard errors. That is, in regressions without share variables, the standard errors are clustered at the individual level, while the standard errors in regressions with share variables are calculated by two-way clustering (e.g. the standard errors are clustered at the individual and at the occupation level when the female share in occupations enter as an explanatory variable), see Cameron and Miller (2015).

4. Multivariate analysis for the private and the public sectors

This section contains regression results for the magnitude of the wage gap between men and women for workers in the private and the public sectors. To save space we present results for pooled estimations for the years 2002-2012 rather than results for each year in our sample period.

Table 2 presents the results of regressions for the private sector, where the logarithm of hourly wages is regressed on the gender dummy, the measures of segregation (the female share of workers in occupations, industries, establishments and job cells, respectively) and a number of other characteristics of the employees. The regression on the gender dummy in column 1, panel A, restates the information from the descriptive statistics in Table 1 that the female wage is 13.7 per cent lower than the male wage.

Table 2 around here

Column 2 of Table 2 contains the results when the gender dummy enters together with basic human capital variables, i.e. years of schooling, experience, experience squared, tenure in firm and tenure squared. The coefficient of the schooling variable shows that one extra year of schooling gives a reward of a 5.9 per cent higher wage rate. The coefficients of the experience variables show that the reward to experience in the labour market increases at a decreasing rate. The reward to tenure also follows a concave profile in the number of years the workers have been employed at the establishment. The inclusion of the basic human capital variables results in a decrease in the gender wage gap by 3.2 percentage points. That is, differences in human endowments between males and females constitute a sizeable part of the gender wage gap at the aggregate level.

All the coefficients in columns 1 and 2 are precisely estimated and significantly different from zero. The same holds for most of the rest of the coefficients reported in the paper. Although we only report significance at the five per cent level, the majority of the coefficients are significantly different from zero at the 0.1 per cent level.

Column 3 includes controls beyond the basic human capital variables. These extended controls show that, conditional on the other explanatory variables, workers in the Copenhagen area are rewarded more highly than in the rest of the country and that ‘singles’ earn less than workers with a partner. The inclusion of these extended controls slightly increases the gender wage gap.

In column 4 the female share of workers in occupations enters as the sole explanatory variable besides the gender dummy. The coefficient of the share of females has a negative sign, and workers in occupations with a high share of females thus have lower wages than workers in occupations with a smaller share of females. The coefficient of the female dummy of 7.9 per cent is only 42 per cent of the raw gender wage gap. That is, the sorting of women into occupations with comparatively low wages and many females appears to play a major role in the magnitude of the gender wage gap.

The magnitude of the coefficient to the share of females in occupations implies that a worker in an all-female occupation is paid 19.0 per cent less than a worker in an all-male occupation. This difference is much higher than the gender wage gap of 7.9 per cent after we control for segregation in occupations and higher than the raw gender wage gap in column 1. The measure of the relation between wages and segregation is larger than the gender wage gaps, and in this sense gender composition across occupations is thus more important for wages than gender per se.

Column 5 shows that females are sorted into industries with relatively low wages as the coefficient of the share of females in industries is negative. Similarly, column 6 indicates that women are also sorted into establishments with comparatively low wages.

Column 7 shows that the sorting of women into job cells plays a major role for wage formation. The coefficient for the share of females is large, and the gender wage gap is reduced to 5.2 per cent. The latter is of the same size when all four segregation measures are included in the regression in column 8. Column 9 shows that the gender wage gap is slightly reduced to 5.0 per cent when both segregation and human capital variables are included in the regression. The inclusion of extended controls in column 10 gives nearly the same result. The coefficients of the basic human capital variables and the extended controls are nearly the same both with and without the inclusion of the segregation variables. The effect of industry-level segregation is not statistically significant in any specification in Table 2.

As a supplement to the estimations with the female share of workers, we perform estimations where fixed effects for occupations, industries, establishments and job cells enter instead of the share variables. The coefficients of the female dummy and the R-squared of these regressions appear in Table 2, panel B. Column 4, panel B, including fixed effects of occupations, shows that this procedure gives an estimated gender wage gap of 7.9 per cent, which is the same as the estimated gender gap with the share variable for occupations in panel A, column 4. The remaining results in Table 2, panel B, are also close to the estimates from the share regressions in Table 2, panel A. As the *R*-squared statistic in Panel B, column 10, takes the value of 0.74, a major part of the variation of wages is taken into account by the fixed effect estimation.

The gender wage gap after controlling for job cells fixed effects is an estimate of the within-job cell gender wage gap. As the gender wage gap when we control for job cell segregation is nearly the same as in the fixed effect estimation, variation in job cell segregation accounts for most of the variation in wages between job cells.

The results for the private sector show that segregation plays a major role in accounting for the gender wage gap. The gender wage gap that remains after we control for segregation is substantially smaller than the gender wage gap that remains after we control for differences in human capital endowment between males and females.

Table 3 presents the corresponding results for the public sector. The raw wage gap of 11.4 per cent in column 1 drops to 7.1 per cent when the basic human capital variables are included in column 2, while the inclusion of extended controls in column 3 further reduces the estimated gender wage gap slightly.

Table 3 around here

Columns 4 to 8 show that segregation plays a dominant role for the gender wage gap in the public sector. All the coefficients of the share variables in columns 4 to 7 are negative, indicating that females in the public sector are sorted into occupations, industries, establishments and job cells with low wages. Moreover, the coefficients are large. The gender wage gap drops to 1.2 per cent when the share of females in occupations enters in column 4 and to 0.1 when the share of females in job cells enters in column 7. When the share variables enter jointly in column 8, the gender wage gap also takes the value 0.1 per cent.

The inclusion of the human capital variables, together with the share variables in column 9, results in a small positive gender wage gap. The coefficient of the female dummy remains positive but close to zero when extended controls are added in column 10. The estimate of the within-job cell wage differential, using fixed effects estimation, also results in a coefficient close to zero, see Table 3, panel B. For the public sector we thus recover the result by Groshen (1991), who finds only negligible wage differentials within-job cells.

It is noteworthy that the coefficient of the share of females in occupations in Table 3, panel A, column 10 takes a value that is twice as large as the coefficient of the share of females in job cells. In contrast, for the private sector the coefficient of the share of females in job cells is larger than the coefficient of the female share in occupations. These estimates are consistent with a greater influence of labour unions on wage setting in the public sector relative to the private sector. If wages were fixed according to occupation across establishments, the coefficient of the share of females in job cells would take the value zero, and a negative association between the wage rates in the collective agreements and the share of females in occupations could be the reason for a negative coefficient of the female share in occupations.

In contrast to the private sector, most jobs in the Danish public sector are covered by collective wage agreements. As the point of departure, Danish trade unions are organised according to occupations (in contrast to industrial affiliation as in, e.g. Germany and Sweden), and collective agreements are thus mainly formed along occupational lines.

An alternative (or complementary) explanation for the small within job cell gender wage gap in the public sector is that wage discrimination (within job cells) is more difficult or not possible, which could be the case if the public sector wages do not include components that are determined by individual effort or output but are fixed by job title. The collective agreements in Denmark actually include the provision that a share of the wage bill in the public sector is distributed to employees according to individual merit, but this share is of limited magnitude.

The conclusion from the analysis of the gender wage gap in the public sector is that segregation plays a very prominent role. Segregation measured as the share of females in occupations, industries, establishments and job cells take account of the entire difference in pay between males and females. Segregation thus plays a more prominent role in the public sector than in the private sector, where a substantial gender wage gap remains after we take segregation into account.

5. Decompositions of the gender wage gap

This section contains estimates of the components of decompositions of the gender wage gap. We present results for decompositions for the private and the public sectors separately and compare the results.

First, we present the results for the pooled data, which form the basis for the estimation results in the previous section. Second, we perform sensitivity analyses by estimating decompositions for the years 2002 to 2012 split into sub-periods. Equation (2) in Section 2 forms the basis for the decompositions.[10]

Table 4 displays the result for the private sector, where the first column contains the results for the pooled data for 2002 to 2012. Panel B, column one, contains the coefficients of the female share variables of the final models displayed in Table 2, column (10), and the male-female differences in the share of females displayed in the last column of Table 1. The product of the coefficients and the differences in the share variables divided by the absolute value of the raw gender wage gap is displayed in Panel A (e.g. for the share of females in occupations we have 0.231=0.104x0.306/0.137). The sum of the contribution of the four female share variables is 0.460, and male-female differences in the share of females in occupations, industries, establishments and job cells thus constitute 46 per cent of the raw gender wage gap. The major contribution from the share variables stems from segregation in occupations and job cells, which show not only negative and large coefficients but also large gender differences in average shares of female workers.

[10] We apply the procedure described in Jann (2008) to calculate the parameters and standard errors displayed in Panel A, Tables 4 and 5. As explanatory variables in decompositions, we use the covariates conditional on the year dummies, i.e. the residuals from regressions of each of the covariates on the year dummies. The Frisch-Waugh Theorem implies that the coefficients of these explanatory variables are the same as those in column (10) in Table 2 and Table 3. This procedure entails that a part of the gender wage gap is not attributed to the time dummies, per se. According to Jann (2008), p. 458-460, the calculation of the variance of the estimators takes into account the sampling variation of the predictors. The package does not support two-way clustering and we thus report standard errors clustered at the industry level.

Table 4 around here

Similarly, the sum of the contribution of the human capital variables constitutes 22 per cent. Segregation thus plays a larger role than differences in human capital endowments in accounting for the gender wage gap in the private sector. The extended controls account for minus 6 per cent of the raw gender wage gap. The first figure in Table 4, panel A, first column, shows that the coefficient of the female dummy of 5.1 per cent in Table 2, column (10), (the unexplained part of the gender wage gap) constitutes 37 per cent of the raw gender wage gap.

To check the robustness of the results and the development over time, we perform analogous decompositions for four sub-periods (to save space, we do not report results for all the years). As mentioned in the data section, not all firms have consistently reported information about the distribution of workers on distinct geographical units for all the years, implying a sample selection problem. Sample selection problems exist for the private sector in one sub-period and for the public sector in another.

The robustness checks for the private sector are displayed in the last four columns of Table 4. The gender wage gap in the private sector declines gradually from 14.6 per cent during 2002-2004, to 14.2 per cent during 2005-2006, 13.5 per cent during 2007-2009, and 12.7 per cent during 2010-2012. The relative importance of segregation diminishes slightly over time as the contribution of the share variables decreases from 47 per cent during 2002-2004 per to 44 per cent during 2010-2012. The contribution of the human capital variables also decreases over time, and the counterpart is that the unexplained part of the gender wage gap increases over the observation period.

According to the statistics for the components of segregation in Panel B, the coefficient of the female share of workers in occupations decreases slightly, while the coefficient of the share of females in job cells is stable over time. The male-female differences in the share of females decreases slightly for occupations, while there are no differences for job cells over time.

The yearly number of observations in the private sector decreased from 760,000 in 2005-2006 (before the recession) to 720,000 (after the recession). The period 2002-2004 contains substantially fewer observations than the rest of the years, reflecting that this period is affected by selection problems. The main deviations of the parameter estimates from 2002-2004, compared to the rest of the years, are that the unexplained part accounts for less of the gender wage gap and that the human capital variables account for more of it.

Table 5 shows the analogous decompositions for the public sector. The first column is based on the model in Table 3, column (10), which shows that females are paid slightly more than males once the effects of segregation and background characteristics are controlled for. The first figure in the first column of Table 5 thus shows that the coefficient of the female dummy accounts for minus 2 per cent of the raw gender wage gap. Human capital variables account for 36 per cent of the raw gender wage gap in the public sector during 2002-2012, and are thus more important in accounting for the gender wage gap in the public sector than in the private sector, where the human capital variables accounts for 22 per cent of the gender wage gap. In contrast to the private sector, extended controls account for a small positive share of the raw gender wage gap in the public sector. Segregation accounts for as much as 63 per cent of the raw gender wage gap in the public sector, compared to 46 per cent in the private sector.

In the public sector, all the contribution from the share variables stems from segregation in occupations and job cells as the coefficients of these variables are negative while the coefficients of the male-female differences in the share of females are positive according to the segregation statistics in Panel B. Furthermore, the magnitude of the male-female differences in the share of females is substantially larger for occupations and job cells than for industry and establishments.

The robustness checks for the public sector are displayed in the last four columns of Table 4, Panel B. The yearly number of observations in the public sector decreased from 610,000 during 2005-2006 (before the recession) to 580,000 during 2010-2012. The period 2007-2009 contains only 350,000 observations, as some parts of the local government sector did not report information about the distribution of workers on establishments in this period. A minor attrition might exist during 2002-2004, where the number of observations is 520,000.

The gender wage gap in the public sector declines from 12.0 per cent during 2005-2006 (which is slightly more than during 2002-2004) to 9.7 per cent during 2010-2012. The unexplained part of the gender wage gap increases from -9.0 per cent during 2002-2004 to -6.0 percent during 2005-2006 and -0.1 per cent during 2010-2012. For the period 2007-2009 the estimates for both the magnitude of the gender wage gap and the unexplained part of the gender wage gap appear to be affected by the sample selection problem during the period, as the estimates deviate from the trend for the period 2010-2012.

A main result for the public sector is a gradual and substantial decrease in the importance of segregation. The contribution of the share variables decreased from 70 per cent during 2002-2004 to 54 per cent during 2010-2012. According to the segregations statistics in Panel B, this development is due to changes in both coefficients and segregation. The absolute value of the coefficients of segregation in occupations decreases from 0.19 during 2002-2004 to 0.15 during 2010-2012, while the coefficients of segregation in job cells decrease from 0.09 to 0.06. In addition, the gender distribution in occupations and job cells has become more equal over time. The absolute value of male-female differences in the average share of females in occupations decreased from 0.38 during 2002-2004 to 0.30 during 2010-2012, while the difference between job cells decreased from 0.50 to 0.43. Most of this decrease is due to the first terms: the male share of female co-workers increases during the sample period while the female share of female co-workers is relatively stable.

We now compare our estimated magnitude of the contribution of segregation to the gender wage gap with comparable estimates in the previous literature that uses matched employer-employee data. One reason for differences in the estimates is that the previous contributions present estimates for the private sector only. Other reasons are differences in labour markets, sample selection and selection of variables included in the analyses.

The procedure for assessing the magnitude of the impact of segregation in Datta Gupta & Rothstein (2005) applied to the results in this paper give the following results: the regressors in Table 2, column 8, consist of the share variables and these variables reduce the coefficient of the female dummy for the private sector from the raw gap of 13.7 per cent to 5.2 per cent. The remaining gap is thus 38 per cent of the raw gap (5.2/13.7). That is, the share variables have been able to 'explain' 62 per cent of the raw gap. According to Datta Gupta & Rothstein (2005), segregation variables explain 54 per cent of the raw gender wage gap for salaried workers (table 7, row 5) and 34 per cent for manual workers (table 9, row 5).

The other contributions in the literature follow the procedure in Groshen (1991), decomposing the raw gender wage gap into sources, corresponding to what we do in this section. A main result of our decompositions is that segregation accounts for the substantial share of 46 per cent of the gender wage gap in the private sector, where the major contribution stems from segregation into occupations and job cells. The qualitative results are maintained when we repeat the basic analysis using unrestricted decompositions, i.e. separate regressions for males and females (a similar result is reported in Bayard *et al.* (2003), p. 900).

According to evidence for the US in Bayard *et al.* (2003), the sum of the contribution of the four share variables (share of females in occupations, industries, establishments and job cells) accounts for 57 per cent of the raw wage gap in the specification with the largest number of occupational classifications (Bayard *et al.* (2003), p. 906, table 5, column (3)). According to the Finnish results in Korkeamaki and Kyyra (2006), the sum of the contribution of job cells and the share of females in firms accounts for 46 percent of the raw wage gap (the sum of the absolute contributions in the OLS specification in Korkeamaki and Kyyra (2006), Table 6, p. 630, relative to the raw wage gap). Ludsteck (2014) shows for Germany that the joint contribution of the female share of workers in occupations, establishments and job cells is 51 per cent of the overall gender wage gap (the sum of the absolute contributions in the pooled OLS specification in table 2, p. 372, relative to the raw wage gap).

Our result that the contribution from segregation for the private sector is 46 per cent is thus in line with the result for the Finnish sample, 5 percentage points lower than the German result, and 11 percentage points lower than the US result. However, according to our evidence, the negative relation between wages and segregation is mainly a matter of occupational segregation, including occupational segregation within establishments, while the negative relation between wages and segregation in the US also includes the sorting of women into low-paying industries and establishments. One reason for the lower contribution from job cell segregation in the US might be measurement errors in the calculation of the share of females, as the US data do not contain all workers in the job cells (Bayard *et al.* (2003) contains a careful treatment of this issue).

A main contribution of this paper is the assessment of the importance of segregation for wage formation in the public sector. We show that segregation accounts for a much larger share of the raw gender wage gap in the public sector than in the private sector in Denmark; the contribution from segregation for the public sector is 63 per cent, compared to 46 per cent for the private sector. Furthermore, the estimate of the contribution of the segregation variables in the Danish public sector is higher than the estimates for the private sector for other countries.

6. Robustness checks

We supplement the test of the robustness of the results over time with additional robustness tests. These tests include various sample restrictions and the application of an alternative wage measure.

For the private sector, we restrict the sample to establishments with at least 25 workers (corresponding to the sample selection in Bayard *et al.* (2003)). The result is that the share variables account for 48 per cent of the raw gender wage gap, i.e. two percentage points higher than the 46 per cent reported in Table 2, Panel A, first column. A restriction of the size of the 6-digit occupations to have at least 50 workers (corresponding to the restriction in Ludsteck (2014)) has no impact on the importance of the share variables for the private sector.

The share variable for job cells with only one worker takes the value of the female dummy. We test whether an omission of job cells containing a single worker affects the results. Workers in job cells with only one worker constitute 7 and 5 per cent of the workers in the private and the public sectors, respectively. The contribution of the share variables in the private sector is reduced by 3 percentage points to 43 per cent, while the results for the public sector remain unchanged.

We also reduce our sample for workers aged 16-64 to workers aged 20-59. For the latter sample, we find that the share variables constitute a slightly higher share of the raw gender wage gap. This share increases by 2 percentage points for the private sector, to 48 per cent, and by 5 percentage points for the public sector, to 68 per cent. Segregation thus appears to be slightly more important for wage formation for prime age workers than for workers at the beginning and the end of their working life.

We explore the role of working hours by limiting the sample to workers who have at least 30 hours per week as the normal working week. For the private sector the share variables constitute 45 per cent of the raw gender wage gap, almost the same as the 46 per cent for the whole sample. For the public sector the share variables constitute 65 per cent of the raw gender wage gap for workers with at least 30 hours per week, in contrast to the 63 per cent for the entire sample.

Until now we have applied information about the gross hourly wage in the analysis. As mentioned in Section 2, our data contain an alternative wage measure, the standardised hourly wage, which exists from 2009. Because of attrition in the public sector, we omit the year 2009 from the analysis. That is, we perform a robustness test by applying the standardised hourly wage for the period 2010-2012, implying that the results are comparable to the last column in Table 4.

For the private sector, the standardised hourly wage yields a gender wage gap of 13.8 per cent, which is 1.1 percentage points higher than that reported in Table 4. Segregation in the form of the share variables constitutes 37 per cent of the gender wage gap, 7 percentage points lower than the share reported in Table 4. Correspondingly, the unexplained part of the gender wage gap constitutes 50 per cent of the raw gender wage gap (in contrast to the 42 per cent in table 4).

For the public sector the standardised wage yields a gender wage gap of 13.9 per cent, almost the same as that in the private sector and 4.1 percentage points higher than in Table 4. The share variables account for 46 per cent of the raw gender wage gap, which is 8 percentage points lower than in Table 4. The unexplained part of the gender wage gap constitutes 20 per cent of the gender wage gap, in contrast to the zero per cent in Table 4.

The decompositions applying the alternative wage measure, standardised hourly wages, confirm that segregation in the public sector is more important for wage formation than in the private sector. As stated in Section 2, the gross hourly wage includes payment for absence in the numerator but not the number of hours absent in the denominator while the standardised hourly wage does not include assessment of absence in either in the numerator or in the denominator. The difference in the gender wage gap between the two wage measures is larger for the public sector than for the private sector, which indicates that the relative difference in absence from work between males and females is larger in the public than in the private sector.

7. Conclusions

We examine the relation between segregation and wages in the public and the private sector in Denmark for the period 2002-2012. Segregation is measured as the proportion of females in occupations, industries, establishments and job cells, respectively. The previous literature has focused on the private sector only. We thus contribute to the literature by investigating differences in the importance of segregation for wage formation between the private and the public sectors.

Our results for the entire period 2002-2012 show that male-female differences in the share of females in occupations, industries, establishments and job cells constitute 46 per cent of the raw gender wage gap in the private sector. This result is in line with results for the private sector in Germany and Finland but 11 percentage points smaller than the estimate for the US.

Our results for the public sector in Denmark show that segregation accounts for as much as 63 per cent of the raw gender wage gap – 17 percentage points more than in the private sector. In other words, segregation appears to play a substantially more important role in accounting for the gender wage gap in the public sector than in the private sector.

For the private sector a substantial within-job cell differential remains when we control for segregation, a result that is in line with recent results for the US, Germany and Finland. However, for the public sector we find that the remaining gender wage gap after controlling for segregation is close to zero, and we thus recover the result by Groshen (1991), who finds only negligible wage differentials within-job cells.

Our robustness checks of the results include decompositions of the raw gender wage gap for sub-periods of the estimation period 2002-2012. We find that the importance of segregation for wage formation in the public sector was reduced substantially by 16 percentage points from 2002-2004 to 2010-2012, while it decreased only slightly in the private sector.

Policies aimed at reducing gender wage differences, given the existing amount of segregation, include two types of intervention: equal pay legislation of the ‘comparable worth’ type (comparison of pay across job cells with different shares of females) and equal pay legislation of the ‘equal pay for the same work’ type (comparison of pay within-job cells). Given that our results are representative for other countries, equal pay provisions of the ‘comparable worth’ type appears to have the largest potential for reducing the remaining gender wage gap in the public sector.

Reference List

Baker, M. & Fortin, N. M. (2004). Comparable worth in a decentralized labour market: the case of Ontario. *Canadian Journal of Economics-Revue Canadienne D Economique, 37,* 850-878.

Bayard, K., Hellerstein, J., Neumark, D., & Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of Labor Economics, 21,* 887-922.

Blau, F. D. & Kahn, L. M. (1999). Institutions and laws in the labor market. In O.Ashenfelter & D. Card (Eds.), *Handbook in Labor Economics, Vol. 3A* (pp. 1399-1461). Amsterdam: Elsevier.

Blau, F. D. & Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic Perspectives, 14,* 75-99.

Bruun, J. R. (1994). Lige løn til mænd og kvinder (Equal pay to men and women). *Ugeskrift for Retsvæsen (Weekly journal of law), Vol. 128,* 421-428.

Cameron, A. C. & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources, 50,* 317-372.

Datta Gupta, N. D. & Rothstein, S. R. (2005). The impact of worker and establishment-level characteristics on male-female wage differentials: evidence from Danish matched employee-employer data. *LABOUR, 19,* 1-34.

European Commission (1975). *Council Directive about the application of the principle of equal pay for men of women.* Brussels.

Groshen, E. L. (1991). The Structure of the Female Male Wage Differential - Is It Who You Are, What You Do, Or Where You Work. *Journal of Human Resources, 26,* 457-472.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *Stata Journal, 8,* 453-479.

Killingsworth, M. R. (1990). *The Economics of Comparable Worth*. Kalamazoo: W.E.Upjohn Institute for Employment Research.

Korkeamaki, O. & Kyyra, T. (2006). A gender wage gap decomposition for matched employer-employee data. *Labour Economics, 13,* 611-638.

Larsen, M. & Houlberg, H. S. B. (2013). *Lønforskelle mellem mænd og kvinder* (Rep. No. 13). SFI, Copenhagen.

Lønkommissionen (2010). *Løn, køn, uddannelse og fleksibilitet* Copenhagen.

Ludsteck, J. (2014). The Impact of Segregation and Sorting on the Gender Wage Gap: Evidence from German Linked Longitudinal Employer-Employee Data. *Ilr Review, 67,* 362-394.

Pedersen, S. B. (2010). De ny lønbegreber (The new wage measures). *Samfundsøkonomen (The social economist).No. 4,* 20-25.

Stastistics Denmark (2004). Lønstatistik for den private sektor 2003 (wage statistics for the private sector 2003). *Statistiske Efterretninger*.

Table 1 Descriptive statistics

	All	Males	Females	Difference
Panel A. Private sector				
Log wage	5.528	5.580	5.443	0.137
Woman	0.377	0.000	1.000	-1.000
Schooling	13.473	13.535	13.370	0.165
Experience	17.660	18.126	16.890	1.236
Tenure	5.270	5.403	5.049	0.354
Capital	0.371	0.349	0.408	-0.059
Single	0.276	0.271	0.284	-0.013
Female share in				
Occupation	0.377	0.262	0.567	-0.306
Industry	0.379	0.317	0.482	-0.165
Establishment	0.377	0.289	0.522	-0.232
Job cell	0.377	0.188	0.690	-0.502
Number of obs.	7,501,564	4,674,840	2,826,724	
Panel B. Public sector				
Log wage	5.543	5.624	5.510	0.114
Woman	0.707	0.000	1.000	-1.000
Schooling	14.457	14.916	14.267	0.649
Experience	19.277	19.592	19.146	0.446
Tenure	5.579	6.106	5.361	0.745
Capital	0.330	0.383	0.308	0.075
Single	0.246	0.250	0.244	0.007
Female share in				
Occupation	0.707	0.455	0.812	-0.356
Industry	0.704	0.568	0.761	-0.192
Establishment	0.707	0.536	0.778	-0.242
Job cell	0.707	0.369	0.847	-0.478
Number of obs.	5,671,717	1,660,343	4,011,374	

Note: Observation years are 2002 to 2012.

Table 2. Regression for log hourly wage, Private sector

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Share regressions										
Female dummy	-0.137*	-0.105*	-0.113*	-0.079*	-0.123*	-0.130*	-0.052*	-0.052*	-0.050*	-0.051*
	(0.001)	(0.001)	(0.000)	(0.003)	(0.008)	(0.002)	(0.001)	(0.005)	(0.003)	(0.003)
Basic Human capital										
Schooling		0.059*	0.055*						0.059*	0.055*
		(0.000)	(0.000)						(0.002)	(0.001)
Experience		0.014*	0.014*						0.014*	0.014*
		(0.000)	(0.000)						(0.001)	(0.001)
Exp. squared/100		-0.071*	-0.073*						-0.071*	-0.071*
		(0.000)	(0.000)						(0.005)	(0.005)
Tenure		0.009*	0.010*						0.009*	0.010*
		(0.000)	(0.000)						(0.001)	(0.001)
Tenure squared/100		-0.049*	-0.048*						-0.049*	-0.048*
		(0.000)	(0.000)						(0.003)	(0.003)
Extended controls										
Capital			0.137*							0.142*
			(0.000)							(0.006)
Single			-0.055*							-0.054*
			(0.000)							(0.004)
Female share										
Occupation				-0.190*				-0.130*	-0.080*	-0.104*
				(0.028)				(0.030)	(0.016)	(0.016)
Industry					-0.081			-0.111	0.016	-0.052
					(0.087)			(0.103)	(0.067)	(0.060)
Establishment						-0.027		0.217*	0.147*	0.169*
						(0.015)		(0.038)	(0.026)	(0.023)
Job cell							-0.168*	-0.153*	-0.135*	-0.123*
							(0.005)	(0.018)	(0.011)	(0.010)
Constant	5.581*	5.560*	5.529*	5.540*	5.568*	5.576*	5.532*	5.532*	5.534*	5.491*
	(0.001)	(0.001)	(0.001)	(0.028)	(0.032)	(0.007)	(0.004)	(0.031)	(0.017)	(0.016)

R-squared	0.040	0.361	0.390	0.051	0.041	0.040	0.050	0.057	0.370	0.400
Panel B. Fixed effects estimations										
Female dummy				-0.079*	-0.123*	-0.129*	-0.054*	-0.054*	-0.056*	-0.056*
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared				0.469	0.304	0.425	0.691	0.692	0.736	0.737

Note: Standard errors clustered at the individual level in column (1)-(3) in parenthesis in Panel A. Two-way clustered standard errors in column (4)-(10). Standard errors clustered at the individual level and at the occupational, industry, establishment and job cell level in column (4)-(7), and at the individual and industry level in column (8)-(10). The continuous variables are centered at sample means.

Panel B contains coefficients for the female dummy from regressions where fixed effect estimation replaces the share variables in Panel A. All regressions contain year dummies for years 2002-2011. The number of observations is 7,501,564. * denotes significance at 5 per cent level.

Table 3. Regression for log hourly wage, Public sector

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Share regressions										
Female dummy	-0.114*	-0.071*	-0.068*	-0.012*	-0.089*	-0.082*	-0.001	-0.001	0.003	0.003
	(0.001)	(0.000)	(0.000)	(0.003)	(0.017)	(0.004)	(0.001)	(0.006)	(0.005)	(0.005)
Basic Human capital										
Schooling		0.057*	0.056*						0.055*	0.054*
		(0.000)	(0.000)						(0.004)	(0.004)
Experience		0.009*	0.009*						0.009*	0.009*
		(0.000)	(0.000)						(0.001)	(0.001)
Exp. squared/100		-0.021*	-0.021*						-0.023*	-0.023*
		(0.000)	(0.000)						(0.004)	(0.004)
Tenure		0.011*	0.011*						0.010*	0.010*
		(0.000)	(0.000)						(0.001)	(0.001)
Tenure squared/100		-0.053*	-0.052*						-0.051*	-0.051*
		(0.000)	(0.000)						(0.004)	(0.004)
Extended controls										
Capital			0.054*							0.052*
			(0.000)							(0.004)
Single			-0.027*							-0.026*
			(0.000)							(0.003)
Female share										
Occupation				-0.287*				-0.275*	-0.162*	-0.162*
				(0.028)				(0.049)	(0.028)	(0.028)
Industry					-0.133			0.091	0.078	0.065
					(0.084)			(0.091)	(0.050)	(0.046)
Establishment						-0.133*		0.02	0.029	0.055*
						(0.014)		(0.035)	(0.025)	(0.025)
Job cell							-0.238*	-0.079*	-0.084*	-0.085*
							(0.006)	(0.030)	(0.022)	(0.022)
Constant	5.626*	5.483*	5.471*	5.623*	5.640*	5.636*	5.603*	5.605*	5.471*	5.458*
	(0.001)	(0.001)	(0.001)	(0.030)	(0.024)	(0.005)	(0.004)	(0.025)	(0.015)	(0.014)

R-squared	0.05	0.32	0.327	0.085	0.055	0.056	0.076	0.088	0.337	0.342
Panel B. Fixed effects estimations										
Female dummy				-0.014*	-0.088*	-0.084*	-0.002*	-0.002*	-0.001*	-0.001*
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared				0.401	0.137	0.175	0.493	0.494	0.535	0.536

Note: Standard errors clustered at the individual level in column (1)-(3) in parenthesis in Panel A. Two-way clustered standard errors in column (4)-(10). Standard errors clustered at the individual level and at the occupational, industry, establishment and job cell level in column (4)-(7), and at the individual and industry level in column (8)-(10). The continuous variables are centered at sample means.
Panel B contains coefficients for the female dummy from regressions where fixed effect estimation replaces the share variables in Panel A. All regressions contain year dummies for years 2002-2011. The number of observations is 5,671,717. * denotes significance at 5 per cent level.

Table 4. Segregation statistics and decompositions of the gender wage gap in the private sector

	All Years	2002-2004	2005-2006	2007-2009	2010-2012
Panel A. Decompositions					
Unexplained	0.372*	0.289*	0.365*	0.416*	0.422*
	(0.035)	(0.054)	(0.043)	(0.034)	(0.050)
Human Capital	0.223*	0.289*	0.213*	0.179*	0.200*
	(0.050)	(0.077)	(0.063)	(0.051)	(0.071)
Extended controls	-0.055*	-0.056*	-0.057*	-0.047*	-0.060*
	(0.010)	(0.017)	(0.014)	(0.011)	(0.017)
Female shares	0.460*	0.469*	0.479*	0.453*	0.439*
	(0.050)	(0.073)	(0.067)	(0.053)	(0.071)
Occupation	0.231*	0.244*	0.230*	0.224*	0.222*
	(0.036)	(0.055)	(0.060)	(0.039)	(0.056)
Industry	0.063	0.048	0.081	0.047	0.073
	(0.072)	(0.111)	(0.096)	(0.074)	(0.100)
Establishment	-0.286*	-0.262*	-0.272*	-0.266*	-0.339*
	(0.041)	(0.060)	(0.076)	(0.040)	(0.059)
Job cell	0.452*	0.440*	0.441*	0.447*	0.482*
	(0.040)	(0.056)	(0.056)	(0.046)	(0.055)
Sum	1.000	1.000	1.000	1.000	1.000
Raw gender wage gap	-0.137*	-0.146*	-0.142*	-0.135*	-0.127*
	(0.010)	(0.015)	(0.013)	(0.010)	(0.014)
Panel B. Segregation statistics					
Coefficients					
Occupation	-0.104	-0.114	-0.104	-0.101	-0.095
Industry	-0.052	-0.044	-0.068	-0.039	-0.095
Establishment	0.169	0.173	0.163	0.150	0.188
Job cell	-0.123	-0.129	-0.123	-0.119	-0.124
Male-female share differences					
Occupation	-0.306	-0.313	-0.315	-0.301	-0.297
Industry	-0.165	-0.159	-0.169	-0.167	-0.164
Establishment	-0.232	-0.221	-0.237	-0.240	-0.230
Job cell	-0.502	-0.499	-0.508	-0.506	-0.497
No. obs. per year	681,960	562,165	758,419	715,777	716,967

Note: Standard errors clustered at the industry level in parentheses. * denotes significance at 5 per cent level. Panel A shows the relative components of the raw gender wage gap according to the decomposition in equation (2). The contribution of the four female share variables is shown both in aggregate and individually. Panel B shows the coefficients and the male-female differences for the four female share variables.

Table 5. Segregation statistics and decompositions of the gender wage gap in the public sector

	All Years	2002-2004	2005-2006	2007-2009	2010-2012
Panel A. Decompositions					
Unexplained	-0.024	-0.090	-0.060	0.067	-0.001
	(0.053)	(0.077)	(0.071)	(0.042)	(0.075)
Human Capital	0.358*	0.366*	0.352*	0.293*	0.414*
	(0.104)	(0.137)	(0.157)	(0.096)	(0.146)
Extended controls	0.032*	0.024*	0.029*	0.027*	0.048*
	(0.009)	(0.012)	(0.013)	(0.009)	(0.013)
Female shares	0.634*	0.700*	0.680*	0.612*	0.540*
	(0.065)	(0.096)	(0.098)	(0.064)	(0.086)
Occupation	0.503*	0.601*	0.489*	0.405*	0.461*
	(0.100)	(0.171)	(0.141)	(0.080)	(0.137)
Industry	-0.109	-0.190	-0.133	-0.039	-0.090
	(0.079)	(0.137)	(0.106)	(0.070)	(0.106)
Establishment	-0.115*	-0.130	-0.068	-0.132	-0.106
	(0.055)	(0.077)	(0.063)	(0.068)	(0.086)
Job cell	0.355*	0.418*	0.392*	0.378*	0.274*
	(0.094)	(0.169)	(0.131)	(0.084)	(0.102)
Sum	1.000	1.000	1.000	1.000	1.000
Absolute gender wage gap	-0.114*	-0.116*	-0.120*	-0.131*	-0.097*
	(0.015)	(0.021)	(0.023)	(0.015)	(0.019)
Panel B. Segregation statistics					
Coefficients					
Occupation	-0.162	-0.186	-0.156	-0.140	-0.148
Industry	0.065	0.104	0.075	0.026	0.058
Establishment	0.055	0.058	0.031	0.071	0.050
Job cell	-0.085	-0.097	-0.094	-0.102	-0.062
Male-female share differences					
Occupation	-0.356	-0.376	-0.375	-0.381	-0.304
Industry	-0.192	-0.213	-0.213	-0.196	-0.151
Establishment	-0.242	-0.259	-0.259	-0.245	-0.206
Job cell	-0.478	-0.503	-0.501	-0.486	-0.427
No. obs. per year	515,611	555,593	605,258	349,500	581,974

Note: Standard errors clustered at the industry level in parentheses. * denotes significance at 5 per cent level. Panel A shows the relative components of the raw gender wage gap according to the decomposition in equation (2). The contribution of the four female share variables is shown both in aggregate and individually. Panel B shows the coefficients and the male-female differences for the four female share variables.