



Helsinki
Center
of
Economic
Research

Discussion Papers

Segregation, gender pay gap and cross-sectional heterogeneity across collective agreements

Pekka Laine

Statistics Finland, Helsinki School of Economics and HECER

Discussion Paper No. 241

November 2008

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781, E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

Segregation, gender pay gap and cross-sectional heterogeneity across collective agreements*

Abstract

In this paper we study how segregation affects the gender pay gap and how this relationship possibly has changed between 1995 and 2004. The data contains information of a large group of employees working in the Finnish private sector firms. We introduce and apply a new econometric approach naming it *the collective labour agreements classification method* (CLAC). The approach being based on index theory and hedonic modelling enables us to take account of intra-labour market variation in wages due to workers being covered by different collective agreements. A major advantage is being able to separate the fraction of the gender gap accounted for by occupational segregation from the one due to women and men being distributed differently between collective agreements. According to our results occupational segregation - though more narrowly defined than in earlier studies - still explains a sizable proportion of 20 to 30 percent of the overall gender pay gap. The results are consistent with earlier studies where occupational segregation is a major factor behind the overall gender pay gap. We also introduce completely new evidence showing segregation across collective agreements being far from insignificant as an explanatory factor for the overall gender pay gap especially among part-timers and contributing even in the subset of full-timers clearly more than, for example, firm segregation to the size of the overall pay gap.

JEL Classification: C43, J31, J71.

Keywords: gender pay gap, segregation, collective agreements

Pekka Laine

Statistics Finland
P.O. Box 5C
FI-00022 Statistics Finland
FINLAND

e-mail: pekka.laine@stat.fi

* I would like to thank Pekka Ilmakunnas (Helsinki School of Economics) and Antti Suoperä (Statistics Finland) for expertise and the Confederation of Finnish Industries EK and Statistics Finland for providing the extensive data. The usual disclaimer applies.

Segregation, gender pay gap and cross-sectional heterogeneity across collective agreements

Pekka Laine

6th November 2008

1 Introduction

Research shows regularly that a significant share of the gender pay gap relates to women being concentrated in less-paying occupations than men. However, *occupational segregation* is not the only form of segregation contributing to the difference between male and female wages. For example, in many countries a large proportion of the employed labour force is covered by collective labour agreements providing and synchronising a common framework for remuneration, working time and other basic terms of employment for all workers under the same collective agreement. And yet a mainstream approach rarely attempts to separate the share of the overall pay gap due to occupational segregation from the one owing to women's segregation into lower paying collective agreements. Maybe even a still more serious deficiency is to ignore cross-sectional heterogeneity in workers' returns to their endowments resulting from the fact that workers are distributed between various collective labour agreements (*abbr. CLA*).

These shortcomings may reverberate in several ways. First, the effects of the different types of segregation may easily be mixed with each other. Second, ignoring cross-sectional heterogeneity (being manifested in parameter heterogeneity in practise) may produce imprecise or downright fallacious parameter estimates which, consequently, lead to misdirected policy recommendations. As a result it will be impossible to conclude assertively whether it is occupational segregation or rather the entire system of collective agree-

ments that needs closer attention in order to reduce the wage gap. Hence without making a clear distinction between types of segregation and identifying correctly their importance *vis-à-vis* the wage gap it might happen that even a successful campaign of reducing occupational segregation would fail to decrease the gap proportionally if a significant share of it were in fact due to women being under-represented in higher-paying collective agreements.

In this paper we first present and then also empirically apply an alternative approach to overcome these shortcomings by first dividing the employed work force into separate groups according to collective agreements and estimating then wage regressions separately for both sexes within each group. In this way we can have a better grasp of heterogeneity and types of segregation affecting the wage gap, especially, in a labour market characterized by collective agreements and occupational segregation. Finally, the index theory offers the necessary means for aggregation of micro-level *CLA*-specific wage gap decompositions together on aggregate level.

The main empirical objects of the current study are to quantify how segregation affects the gender pay gap and how this has possibly changed between 1995 and 2004 in Finland. This time span of ten years is of extra interest for us since it coincides with a period of recovery in Finland after an economic recession of historical dimensions during the first years of the 1990's when unemployment rate increased in a couple of years by over ten percentage points. Thus the recovery period forms a coherent time span for studying whether the forms and types of segregation have remained unaltered even after the recession which otherwise entailed even many fairly fundamental changes for the structure of the whole Finnish economy, for example, by accelerating the intergenerational transformation of the labour force structure. On the other hand, by the time of writing this paper the data sets being fit for the econometric use covered precisely the ten year period 1995-2004 only.

Especially, concerning the types of segregation at issue we will focus in more detail on trying to quantify empirically the shares of the overall gender wage gap arising from:

1. occupational segregation,
2. firm segregation and
3. collective labour agreement segregation.

In what follows, we present first an overview of literature on the relationship between segregation and the gender wage gap. We discuss also the need

for enhancing the analysis by more detailed distinction of various types of segregation from each other and by taking account of heterogeneity in wage-setting arising from men and women being distributed differently across collective agreements in a labour market characterised by collective bargaining. After that we introduce our approach capturing cross-sectional heterogeneity in wage setting due to various collective agreements and offering a way to aggregate micro-level wage gaps into sectoral-level decompositions. Then follows the empirical analysis. And finally a conclusive discussion about the results and the methodological approach ends the study.

2 Previous research, types of segregation and cross-sectional heterogeneity across collective labour agreements

A wide literature has emerged during the last decennia on gender segregation and its wage gap effects. Various forms of segregation have been taken up in these studies. A major landmark was Groshen (1991) who in her seminal study disentangled segregation by occupation from that based on employer or on job-cell (*i.e.* the interaction between occupation and establishment). According to her results, the proportion of females in an occupation explains the largest proportion of the gender wage gap. Instead gender wage differentials arising within *job-cells* turned out be negligible.

Afterwards Groshen's approach has been replicated with minor modifications for the U.S. as well as for a number of other countries. Using micro-data for the U.S Bayard et al. (2003) found that occupational segregation accounts for a large portion of pay gap among full-time employees. In contrast to Groshen, however, they observed significant gender wage differentials within job-cells too. Groshen's approach has also been replicated in Nordic countries. Using Danish data Gupta & Rothstein (2005) found significant within-job-cell differentials. Korkeamäki & Kyyrä (2005) provide evidence that in the Finnish manufacturing sector the aggregate effect of segregation of women and men into different industries, occupations, firms and job cells (occupation-firm combinations) explain some 50-60 per cent of the gender wage gap. Among blue-collar workers each dimension of sex segregation seems equally important while among white-collar workers the segregation effect works entirely through occupation and job cell segrega-

tion. Finally, Arai et al. (2004), in order to avoid multi-collinearity between various types of proportion females, excluded job cells completely from the analysis and found after having compared female proportions of occupations and establishments with each other the first mentioned explaining clearly a larger fraction of the overall gender pay gap among Swedish private sector employees in 1995.

Thus even a short review of earlier research proves of remarkable complexity and ambiguity in mechanisms and processes generating pay differentials between men and women. In fact, one can hardly capture multitude of details of the segregation phenomenon in one study only. This explains also why our approach differs from that of Groshen's in certain aspects. For example, we do not analyse the job-cell dimension but instead concentrate on analysing pay gap effects of gender segregation along occupational and firm/establishment dimensions in more detail.¹ In this way we hope to add on to earlier studies by emphasising certain points not being thoroughly scrutinised so far. We continue by discussing how and why our approach differs from previous studies.

The definition of segregation is a natural starting point, not least because the very concept is surprisingly often fairly loosely defined in many papers. Since Groshen's seminal paper researchers often strive for capturing segregation by including the proportions of females in branch, firm/establishment, occupation and job-cells in estimation equations.² This is a convenient solution at least for two reasons. Firstly, proportion female is a continuous variable compressing an otherwise huge number of indicator variables (especially, number of job-cell dummies, being the inner product of numbers of occupations and firms (or even establishments), grows fast). Furthermore, empirical evidence supports the inclusion of proportion female as there exists fairly comprehensive empirical documentation of an adverse relationship between proportion of females in firms/establishments/occupations/job-cells and the corresponding mean wage.

¹According to Korkeamäki & Kyyrä (2005) only a minor part of raw wage differentials originates from within job-cells. Furthermore, as there is empirical evidence of people often working in job-cells being highly dominated by one or the other sex, even the option of specifying the corresponding proportion female as continuous is debatable (the last mentioned issue will be discussed in more detail later in the current paper).

²Formally, proportion females can be defined as $\frac{F_j}{N_j}$ where F_j denotes number of women and N_j number of men and women belonging to labour structure j (branch, firm/establishment, occupation or job-cell *etc.*).

However, there are still some remaining questions concerning the use of proportion of females as a measure of gender segregation. Firstly, it is unsure whether the relation between female proportion and (log) wages is linear as it is usually specified. Secondly, as Arai et al. (2004) have already demonstrated empirically, inclusion of proportions of females as explanatory variables in wage model leads to obvious multi-collinearity problem.³ Thirdly, the overall distribution of female proportions depends on women's labour force participation rate potentially restricting the range of values proportion females may take and simultaneously also invalidating international comparisons. On the basis of these considerations we will opt for an alternative model specification by using indicator variables in specifying firm and occupation effects and thus bypassing the linearity assumption as well as potential multi-collinearity problems.

While in studies of the gender wage gap a lot of effort has been put to include exhaustively all the *observed* wage-determining factors and even to account for heterogeneity due to *unobserved* factors as well, it seems that the issue of potential heterogeneity in *returns* to observed factors across different segments of labour force has not attracted equivalent attention by researchers as yet. Still, one can easily name potential sources for parameter heterogeneity, such as, idiosyncratic differentials in wage determination among firms or collective agreements. Thus in this study, instead of trying to capture all observed and unobserved wage-determining *factors* as comprehensively as possible, we turn the focus on the potential and this far largely unexplored heterogeneity resulting from differences in returns to conventional wage-determining factors (age, education, occupation *etc.*) generating disparate *effects* on the gender wage gap across subgroups of workers under various collective agreements.

A central justification for the choice of collective agreements as a key source of intra-labour market heterogeneity in the current paper are the strong status and the long traditions of centralised collective agreements between the employers' and employees' labour market organisations within the Finnish labour market system. The regulation of the labour market in Finland is based on labour legislation and, primarily, on collective agreements

³Arai et al. (2004) report a cross correlation estimate of 0.77 between proportion females of occupations and job-cells for Swedish private sector employees in 1995. Furthermore, Korkeamäki & Kyyrä (2005) report cross correlations between proportion females of occupations and job-cells being 0.85 and 0.74 for Finnish white-collar and blue-collar workers respectively.

setting the minimum terms and conditions of the central features of any employment relation, such as, wages and labour hours. However, collective agreements determine only the minimum of wage level and increases an employer is obliged to pay while it is entirely up to the firm to pay more than that if so required (*e.g.* due to excess demand of a certain type of labour, efficiency wage considerations *etc.*). This phenomenon being called "wage drift" is familiar from all labour markets covered by collective agreements.

Any trade union and employers association may make collective agreements. Nearly all collective agreements are branch-specific and usually blue-collar and white-collar workers belong under different agreements even within the same branch. Finally, a further potential sorting factor for many agreements is formed by occupational sub-groups nearest corresponding to ISCO-88(COM)⁴ 1-digit levels (typically levels 2,3 and 4 entering separately and levels 5-9 more often as a common group while level 1 "legislators, senior officials and managers" usually do not belong to any collective agreements having instead individual employment contracts of their own).

Indeed, much of the existing literature on segregation ignores collective agreements even though these clearly form a potential source of heterogeneity in any labour market where collective bargaining takes place. And even if not completely ignored the whole complexity of heterogeneous collective agreements is often approached by merely adding a dummy variable for industry into a wage regression. Still, heterogeneity goes much deeper than that, mechanisms determining individual wages being reflected through the agreement-specific regression coefficients. *I.e.*, wage determination varies from agreement to agreement and in order to capture this diversity and, especially, in order to prevent its pay gap effects erroneously being mingled with corresponding segregation effects, regression coefficients need to be estimated separately for each collective agreement.

Admittedly, our approach may appear somewhat unconventional in the present context. However, there is an extensive literature (under the collective title "*hedonic methods*") on the methods being applied in this study based on *index theory* and having often direct linkages to *NSIs* (national statistical institutes). Indeed, most of the empirical applications in the area have been developed and applied by *NSIs* (especially in the field of consumer price indexes).

⁴ISCO-88(COM) is the European Union variant of the International Standard Classification of Occupations (ISCO-88).

3 Methodological description: the Collective Labour Agreement Classification Method

We continue by presenting our statistical approach (henceforth, *the collective labour agreement classification method*) in more detail. The first step is to define estimated wage equations in semi-logarithmic form as follows

$$\ln W_{ceoi}^s = \hat{\alpha}_c^s + \hat{\beta}_c^{sT} \mathbf{x}_{ceoi}^s + \sum_{e=2}^{E_c} \hat{\gamma}_{ce}^s D_FIRM_{ei} + \sum_{o=2}^{O_c} \hat{\delta}_{co}^s D_OCC_{oi} + \hat{\epsilon}_{ceoi}^s \quad (1)$$

where "hat" $\hat{\cdot}$ denotes (OLS) parameter estimate. $\ln W_{ceoi}^s$ is a logarithmic wage of person i of male (superscript $s=m$) or female (superscript $s=f$) gender being covered by a collective labour agreement c , working in employer firm e and occupation o . \mathbf{x}_{ceoi}^s is a vector consisting of the K observable variables denoting individual i 's own or his/her employer firm's (e) characteristics (potentially containing both continuous and indicator variables) and $\hat{\beta}_c^{sT}$ is a transpose vector of the K corresponding coefficients. D_FIRM and D_OCC are indicator variables being equal to one for firm and occupation in which person i works, otherwise they are equal to zero.

A few focal issues need to be highlighted in order to clarify our approach. Firstly, note the subscript c denoting collective labour agreement (*CLA*) under which the worker belongs. There are C different *CLAs* each forming a separate estimation cell. *I.e* we will divide workers according to collective agreements they belong to and estimate for each separate group of men *vs.* women under a given *CLA* a separate wage equation. All in all, as each gender will be estimated separately there will be $2 \cdot C$ separately estimated wage equations in the analysis.

Secondly, the estimated model will remain identically specified across the $2 \cdot C$ separate estimations. Of course, firm and occupation indicators need to be adjusted to correspond to the *CLA* in question (the indicator variables simply being zeros for firms and/or occupations not being represented under the collective agreement in question) but otherwise all the explanatory variables as well as the model specification will remain the same across all the estimations.

Thirdly, even though the set of explanatory variables will remain basically the same in each estimation, all the coefficient estimates ($\hat{\alpha}_c^s$, $\hat{\beta}_c^s$, $\hat{\gamma}_{ce}^s$, and $\hat{\delta}_{co}^s$) will vary across collective agreements (and separately for men and

women) reflecting the heterogeneity in wage setting across *CLAs*. Especially, take here notice of the estimated $\hat{\gamma}_{ce}^s$ coefficients capturing firm effects. These coefficients being estimated for each collective agreement (c) separately implicates that, though referring to the firm-level effects in the first place, they capture even establishment-/local unit-specific *within-firm* effects when the firm consists of two or more local units each operating within a different industry/branch and simultaneously under a different collective agreement. And of course, in the case of a single-unit firm the corresponding coefficients catch firm and establishment/local unit effects simultaneously. For unambiguity, we will talk about "firm effects" even when referring to a multi-unit multi-branch firm.⁵

After having estimated $2 * C$ separate otherwise identical wage equations we can form a standard Oaxaca-type wage gap decomposition⁶ for each collective agreement c as follows:

$$\begin{aligned}
\ln \left\{ \frac{W_c^m}{W_c^f} \right\} &= (\hat{\alpha}_c^m - \hat{\alpha}_c^f) + \hat{\beta}_c^{mT} (\bar{\mathbf{x}}_c^m - \bar{\mathbf{x}}_c^f) + \left(\hat{\beta}_c^m - \hat{\beta}_c^f \right)^T \bar{\mathbf{x}}_c^f + \\
&\quad \sum_{e=2}^{E_c} \hat{\gamma}_{ce}^m \left(\overline{D_FIRM}_{ce}^m - \overline{D_FIRM}_{ce}^f \right) + \\
&\quad \sum_{e=2}^{E_c} (\hat{\gamma}_{ce}^m - \hat{\gamma}_{ce}^f) \overline{D_FIRM}_{ce}^f + \\
&\quad \sum_{o=2}^{O_c} \hat{\delta}_{co}^m \left(\overline{D_OCC}_{co}^m - \overline{D_OCC}_{co}^f \right) + \\
&\quad \sum_{o=2}^{O_c} (\hat{\delta}_{co}^m - \hat{\delta}_{co}^f) \overline{D_OCC}_{co}^f,
\end{aligned} \tag{2}$$

where *geometric* (underline) and *arithmetic* (bar/overline) means are calculated separately for each combination of gender ($s = (f, m)$) and collective

⁵In our use the term "establishment" refers to a firm's single physical location where business transactions take place. Groups of one or more establishments under common ownership or control are "firms". A single-unit firm owns or operates only one establishment. A multi-unit firm owns or operates two or more establishments belonging to the same or different fields of activities.

⁶*Cf.* Oaxaca (1973).

labour agreements ($c = (1, \dots, C)$).⁷ Naturally, the total absence of residual means in the wage gap decomposition 2 results from one of the main statistical properties of OLS estimators that the mean value of the OLS residuals is zero.

There is nothing extraordinary in the wage gap decomposition 2 in itself. Instead the main issue is now that there is 2^*C distinct sets of regression parameter estimates each being based on a basically identically specified wage equation, but being separately estimated for each sub-group of workers across all combinations of gender and collective agreements. Underlying here is the fact that in Finland collective labour agreements stipulate all central aspects of an individual's labour relationship such as working time and remuneration. Therefore wage determination when based on collective agreements finds its empirical counterpart in heterogeneity of regression coefficients across estimation cells (*i.e.* combinations of sex and collective labour agreements).

The next issue is how to aggregate the *CLA*-specific decompositions to aggregate level (the whole Finnish private sector). Here index theory can be of significant help and the *CLA*-specific wage gap decompositions can be aggregated into one overall decomposition using a Laspeyres-type index framework in which females are equivalent to "base period" and males to "comparison period" (following the terminology of time indexes). In practise, we will adopt the following aggregation solution initially stated by Koev (2003) and based on the log-Laspeyres index⁸:

$$\log\text{Lasp}W_f^m = \sum_{c=1}^C \frac{\bar{W}_c^f N_c^f}{\bar{W}^f N^f} \ln \left\{ \frac{W_c^m}{W_c^f} \right\}, \quad (3)$$

Thus the aggregation solution is a weighted average across collective

⁷For example, the arithmetic mean wage for males under a collective agreement c is $\overline{\ln W}_c^m = \sum_{i=1}^{N_c^m} \frac{\ln W_{ci}^m}{N_c^m}$, and the corresponding geometric mean is calculated as follows $\underline{W}_c^m = \left[\prod_{i=1}^{N_c^m} W_{ci} \right]^{\frac{1}{N_c^m}}$ where \prod is multiplication operator. Thus \underline{W}_c^m is a N_c^m th root of the product of wages of the N_c^m men belonging to collective labour agreement c . The proportion of men having occupation $o \in (1, \dots, O)$ is $\overline{D.OCC}_o^m = \frac{N_{co}^m}{N_c^m}$ etc. Note also that $\ln \left\{ \frac{W_c^m}{W_c^f} \right\}$, *i.e.* the logarithmic ratio of men's and women's geometric mean wages is equal to $\overline{\ln W}_c^m - \overline{\ln W}_c^f$, that is the difference of men's and women's arithmetic means of logarithmic wages.

⁸To our knowledge Koev (2003) was the first to derive and combine an Oaxaca-type decomposition with an aggregation solution based on index theory.

agreements consisting of logarithmic relatives of geometric averages of the female and male wages using *CLA*-specific shares of women’s aggregate payroll as weights.

Now the logarithmic Laspeyres formula 3 reveals how many log-percent higher the overall female mean wage would be if women had on average similar characteristics as men and were remunerated in an exactly similar fashion as men for any characteristic *within* each collective agreement, and yet women’s distribution across agreements still deviated from that of men. A major novelty now is that in contrast to the traditional approach of having only one equation and it being estimated for the whole sample in one go (maybe, at most, separately for both genders) our approach avoids mixing the gender segregation across collective agreements with occupational and firm/establishment segregation.

The *aggregated* wage gap effects of occupational, firm and collective agreement segregation are defined in more detail in table 1 below. Note especially that the fraction of the overall pay gap due to women being segregated proportionally more often into less paying collective agreements is the residual between the overall logarithmic pay gap and the log-Laspeyres type aggregation solution of equation 3 above.

[Table 1 about here.]

4 Data and Results

The data source of the study is the Finnish Structure of Earnings Survey (*SES*) data. The data set contains information on individuals’ conditions of employment (wage/salary components, working time, *etc.*), personal characteristics (age, occupation, education, *etc.*) and employer (industry, size, *etc.*) across all the sectors of the Finnish Economy.

However, we do not use the whole *SES* data covering both the private and the public sectors (municipalities and central government) in our analysis but instead concentrate on analysing a sub-sample of the private sector consisting of the employees working in the member firms of the Confederation of Finnish Industries (*EK*). *EK* collects the basic statistics on wages and salaries for an annual *SES* data from its member enterprises. The inquiries generally cover all the organised employers employing a vast majority of all the employees in the Finnish private sector. Statistics Finland supplements these data by added information on employees from certain public registers and data bases.

Even though *SES* data has been collected on annual basis since 1995, we will confine ourselves to two separate annual pay gap studies on cross-sectional data for years 1995 and 2004 respectively. As already discussed earlier in the current paper this time span forms a period of recovery in Finland after an economic recession of historical dimensions during the first years of the 1990's. Thus comparing the beginning and the end of the period offers a splendid possibility to study how different types of segregation evolved and whether their effects on the gender pay gap possibly changed between 1995 and 2004 coinciding with various other changes for the structure of the Finnish labour markets and for the whole Finnish economy as well, such as the accelerated intergenerational transformation of the labour force structure and the increasing role of the electrotechnical industries.

We start the empirical analysis by constructing an estimation classification. That means workers are to be grouped into separate estimation classes/cells so that people whose wages are determined according to same rules/mechanisms end up within the same class. This does not mean, however, that each person will be paid exactly the same wage within a given class. Rather the aim is to group employees in such a way that the impact of wage determining factors (occupation, education, work experience, etc.) on individual wage would be of equal magnitude. Thus, for example, a one more year in work experience should generate an identical (per cent) increase in earnings for all the persons within the same estimation class. And still within a same estimation class more skilled or experienced workers could earn more than their less-educated and/or less experienced colleagues.

In Finland, most of the employment relationships adhere to the sector's collective labour agreement which determine such key elements of each employment relationship as (minimum) wages and working hours in detail. Consequently, collective agreements form a natural starting point when determining the estimation classification solution. In the Finnish structural earnings statistics, however, there is no valid information about a worker's collective agreement. So first we needed to decide how to distribute individuals between collective labour agreements, or in case of minor agreements, between groups of combined homogeneous agreements in order to ensure a statistically sufficient number of observations for regression estimations.

There are three factors according to which any worker's potential collective agreement can be deduced. The first factor is the employer firm's field of activity or, in the case of a multi-unit firm operating several establishments, the corresponding establishment's field of activity. The second factor is the

nature of the employment relationship, *i.e.*, whether the employee is hourly or monthly paid. Eventually, even the person's profession may serve as a last decisive factor in determining under what collective agreement she/he will obey.

In practice, we proceeded by identifying all the private sector collective agreements in force in 1995 and 2004 respectively. After that we defined the fields of activities/industries⁹ and the nature of the employment relationship (hourly vs. monthly paid) each collective agreement mainly concerns. In addition, any professions or job/occupational titles mentioned in the application directive of a collective agreement (such as "technical staff", "management", "specialist", *etc.*) were taken into account. Thus a preliminary estimation classification was formed linking each individual worker unambiguously with a certain collective labour agreement as employer firm's/establishments's field of activity, worker's nature of the employment relationship and profession/occupational title served as link variables. Finally, after having distributed workers according to the preliminary estimation classification, minor classes were joined together by combining similar fields of businesses or professional groups into larger estimation cells in order to ensure statistically sufficient estimation groups.

After having determined the estimation classification we move on to specify the form of a wage model to be estimated. We will use the same semi-logarithmic model as was already stated in equation 1. The approach in nutshell is to repeatedly estimate the basically identical wage model separately for $2 * C$ sub-sample combinations of gender and collective labour agreements.¹⁰ The wage concept is the employee's monthly salary for regular hours consisting of the basic salary (duty/task-specific monthly pay + regular personal bonuses (incl. age / length of service allowances)) without any shift, Sunday and overtime premiums and without any commission or incentive payments.

In more detail the repeatedly for each *CLA*-class c estimated model is specified as follows:

⁹Fields of activities/industries were defined at 2- or 3-digit level of NACE Rev. 1 classification (the statistical Nomenclature of economic activities in the European Community).

¹⁰Of course, the set of occupations and firms/establishments vary from agreement to another. Otherwise, though, the model specification remains the same semi-logarithmic model as stated in equation 1.

$$\begin{aligned}
\ln RMW_{ci}^s = & \hat{\alpha}_c^s + \sum_{o=2}^{O_c} \hat{\gamma}_{co}^s D_OCC_{coi} + \sum_{e=2}^{E_c} \hat{\delta}_{ce}^s D_FIRM_{cei} + \\
& \sum_{l=2}^{L_c} \hat{\eta}_{cl}^s D_EDLEVEL_{cli} + \hat{\tau}_c^s D_TEMP_{ci}^s + \hat{\beta}_{c1}^s AGE_16_i + \\
& \hat{\beta}_{c1}^s AGE_16_i^2 + \hat{\theta}_c^s D_NOFULLTIME_{ci} + \hat{\epsilon}_{ci}^s,
\end{aligned} \tag{4}$$

where superscript s tells person's sex (f = female, m = male), subscript c identifies the estimation class (*i.e.* collective agreement), i is an personal identifier, o occupation title¹¹, e refers to company/establishment and l denotes educational level.

In more detail the contents of the variables in wage model 4 are listed below:

- $\ln RMW_{ci}^s$ = Natural logarithm of monthly basic salary in real terms
- $D_OCC_{coi}^s$ = indicator variable for occupations: =1, when person i 's occupation is o ($o \in \{1,2, \dots, O_c\}$), otherwise =0. Within each collective agreement the reference group occupation ($o=1$) has been chosen on the basis of being as abundant as possible among both genders.
- $D_FIRM_{cei}^s$ = Indicator variable for firms: =1, when person i 's employer firm is e ($e \in \{1,2, \dots, E_c\}$), otherwise =0. Within each collective agreement the reference group ($e=1$) consists of one or more minor firms chosen as to avoid the multi-collinearity trap.
- $D_EDUCLEVEL_{cli}^s$ = Indicator variable for educational levels: =1, when person i 's educational level is l ($l \in \{1,2, \dots, L_c\}$), otherwise =0. Depending on the collective agreement the reference group ($l=1$) contains the 3-5 lowest educational levels.¹²

¹¹The occupation classification used here is the Finnish variant of the European Union's ISCO-88(COM) classification standard.

¹²We apply KOL97, *i.e.*, the Finnish variant of ISCED-1997: the International Standard Classification of Education (ISCED) adopted by the General Conference of UNESCO at its twenty-ninth session in November 1997. However, while there are seven separate educational levels (+ one for missing information) in ISCED-1997, KOL97 contains eight levels (+ one for "unknown" level).

- $D_TEMP_{ci}^s$ = Indicator variable for temporary employment relationship: =1, when employment relationship is temporary, otherwise =0.
- $AGE_16_{ci}^s$ = Person's age - 16.
- $AGE_16_{ci}^{s\ 2}$ = Person's age - 16 squared.
- $D_NOFULLTIME_{ci}$ = Indicator variable for not working full time: =1, when less than 150 regular working hours per month, otherwise =0.

Table 2 contains estimation diagnostics aggregated from having first estimated model 4 repeatedly for each *CLA* class. As a measure of goodness of fit we report aggregated root mean square errors (*RMSE*). Especially in the case of semi-log specification it is a good measure as the weighted mean of *CLA*-specific *RMSE*s (multiplied by 100) describes approximately how many percent the *untransformed* wages deviate from the predicted values on average.¹³ A key observation is that the inclusion of part-timers erodes the regression fit. While female part-timers' wages vary on average by 22 to 25 % around the predicted values and male part-timers' wages by 17-18 %, the full-timers' wages are much more closely concentrated around the predicted values, the average deviation being about 13 % for both genders.

Table 2 contains also information on how many different collective labour agreements (C), firms/establishments (E) and occupations (O) there are for men and women, respectively, in the sample in each year. The number of CLAs have increased by ten: 139 in 1995 *vs.* 149 in 2004. Both genders are represented in all the existing agreements. The number of firms have increased clearly over the observation period indicating the corresponding increase in labour force employed in the private sector. The number of occupations, instead, has decreased somewhat being connected most probably

¹³Note that in the case of the collective agreement classification method the weighted mean of *CLA*-specific squared *Rs* is not comparable to a squared *R* calculated from an aggregate level regression. The reason for this is seen by looking at the definition of squared *R* of a given collective agreement *c*, *i.e.* $R_c^2 = \sum_{i=1}^{N_c^s} \frac{(\ln \widehat{W}_{ci} - \ln \bar{W}_c)^2}{(\ln W_{ci} - \ln \bar{W}_c)^2}$. Wages being calculated as deviations from *CLA*-specific means imply that the "between" variation in wages due to collective agreements is already being controlled for in contrast to the conventional aggregate level "in one go" estimation where wages are being calculated as deviations from the aggregate mean.

with updates in the used Finnish variant of the European Union's ISCO-88(COM) classification standard of occupations.

The figures in parentheses denote the fraction (percentage) of all the firms and occupations where women respectively men work that are fully (100 %) dominated by the gender in question. It seems that over the period 1995-2004 the shares of totally gender-segregated enterprises remained almost unaltered while the shares of totally gender-segregated occupations even decreased. Thus, if the development of segregation by gender were to be assessed only by looking at the changes in the one-dimensional shares of 100% male/female firms and occupations over time, one would definitely see no sign of an increase in occupational or firm segregation by gender over the period 1995-2004.

[Table 2 about here.]

Furthermore, the numbers of CLA-firm (C*E) and CLA-occupation (C*O) combinations for both genders are given in Table 2. Now the figures in parentheses denote the percentage fractions of all the *combinations* of firms and occupations with collective agreements that women respectively men occupy being fully (100 %) dominated by the gender in question. Note, especially, that contrary to the picture emerging from the one-dimensional shares of one-gender enterprises and occupations above gender segregation, as measured by the shares of 100% male and female CLA-firm and CLA-occupation *pairs*, shows now a very clear increase over the same period. Especially, while the share of one-gender firms stayed roughly between 10 and 20 percent for both sexes in 1995-2004, the percentages of one-gender dominated CLA-firm pairs increased from less than 28 % in 1995 to over 50 % in 2004! This means that the probability of men and women who work within the same enterprise to be segregated totally by gender under different collective labour agreements has considerably increased during the ten-year period.

All this illustrates the complexity of the segregation phenomenon: even though segregation of men and women across firms remained nearby unaltered from 1995 to 2004, segregation by gender across collective agreement-firm *pairs* grew substantially over the same period. Thus concentrating merely on one-dimensional segregation (*e.g.* occupational segregation or firm segregation) may lead to biased conclusions about how segregated the labour market in fact is.

Finally, information of Table 2 can be used to calculate the male- *vs.* female-specific average numbers of firms per agreement ($\frac{C*E}{C}$), agreements

per firm ($\frac{C*E}{E}$), occupations per agreement ($\frac{C*O}{C}$) and agreements per occupation ($\frac{C*O}{O}$). The average number of collective agreements per firm increased from less than two CLAs per firm in 1995 to three in 2004. This development highlights ever more the importance of separating gender segregation across firms from the one due to men and women being distributed differently across collective agreements. The development seem fairly similar even when considering how many occupations there are on average per a collective agreement: while in 1995 an average agreement covered 25-30 occupations, the average number of occupational titles per CLA almost doubled within the next ten years ranging between 40-50 occupations per a collective agreement in 2004.

Table 3 presents gender pay gap decompositions for the Finnish private sector in the years 1995 and 2004. The calculations are based on the collective labour agreements classification method (CLAC) presented in Chapter 3 above. Line 1 shows the aggregate pay gap telling how many percent higher the men's average wage is compared to women in the private sector during the corresponding year. The effects of the key determinants on the aggregate pay gap follow on rows 2-6 telling the corresponding factor's effect as a percentage of the overall pay gap. Since the choice of whether to include even part-time workers in pay gap calculations is likely to affect the size of the aggregate pay gap, we report decompositions both for full-timers alone and for the combined body of full- and part-timers together.

[Table 3 about here.]

In a nutshell, the overall wage gap consists of four components:

1. Collective agreement segregation, that is, men being proportionally more often under a better paying collective agreement (line 2).
2. Occupational segregation, that is, **within the same collective agreement** men proportionally more often in possession of better paying occupation (line 3).
3. Firm/establishment segregation, that is, **within the same collective agreement** men proportionally more often work in a better paying firm or establishment (line 4).
4. Other factors' contribution (line 6).

Thus, percentages on lines 2-4 are summed up as the combined percentage of all segregation effects on line 5 which added to the percentage due to "other effects" (line 6) always sum up to a total of 100 percent (line 7). The "other effects" on line 6 in turn consist of:

- **Within the same collective agreement** men may have **a)** proportionally higher educational level, **b)** higher average age (and consequently, longer potential work experience), **c)** proportionally more often fixed-term employment relationship and **d)** proportionally more often full-time job.
- In addition, the returns to these factors may favour men (however, only continuous variables' pay gap effects can be identified separately).

In the joint sample of full- and part-timers the overall gender gap was 25.4 per cent in 1995 increasing to 28.8 per cent by 2004. When considering the full-timers' sub-sample the change took the opposite direction from the 1995 year's 20.1 percent to 18.2% by 2004. When interpreting these results we benefit from a "bonus" of the collective labour agreement classification approach, namely, the overall gender pay gap can be decomposed into two parts: one due to men and women being differently distributed across collective labour agreements (collective labour agreement segregation on line 2) and another part due to men having potentially higher mean wages within collective agreements (occupational and establishment segregation plus other effects on lines 3, 4 and 6). While in 1995 the first mentioned "CLA-segregation" effect explained about one fourth (24.4 %) of the total 25.4 per cent pay gap for the joint sample (full- and part-timers), by 2004 its explanatory power had risen to over two-fifths (42.7 %). In the case of the full-timers' sub-sample CLA-segregation corresponded to a much smaller fraction of the overall gap: in 1995, the CLA-segregation explained only five per cent and in 2004, even after showing a clear rise, its explanatory power was not more than 13.2 per cent as a portion of the overall gap.

Thus, the size of the CLA-segregation effect seems to depend positively on the share of part-timers in the sample. Bearing in mind that already in 1995 women did more often part-time (13.9% of women vs. 3.8% of men) and that women's propensity for part-time work grew proportionally even more up to 2004 (18.9% of women vs. 5.2% of men) it is close to conclude that the magnitude and growth of the CLA-segregation effect bears evidence

of female part-timers ending up proportionally more often under less paying agreements than their male counterparts. On the other hand, this means also that the effects of "part-time segregation" and *CLA*-segregation mix up with each other thus potentially overemphasising the role of *CLA*-segregation among the background factors. In this case there will be an upward bias in the estimated effect of *CLA*-segregation on the overall pay gap. All in all, the larger proportion of women than men working as part-timers forms a clear object for further research. The main issue here is whether women's higher propensity to do part-time depends on their own preferences or alternatively indicates of typical female jobs being more often open for part-timers only as compared to traditional male jobs.

Line 3 presents the partial effects of occupational segregation calculated as percentage of the overall gap. For full- and part-timers together occupational segregation accounts for almost one fourth (23.6 %) of the overall gap in 1995 diminishing then below two-fifths (19.1%) in 2004. The diminishing explanatory power of occupational segregation for the overall gap reinforces the idea that female and male part-timers are largely located in different collective agreements. Again when considering full-timers alone the change is the opposite, the relative effect of occupational segregation having increased slightly: while the explanatory percentage was 27.9 in 1995 it had risen to 29.1 per cent in 2004. On the other hand, the gender pay gap having fallen among the full-timers over the same period leads us to conclude even for their part that, apparently, occupational segregation did not account for any larger pay gap share in pecuniary terms in 2004 than in 1995.

The FISCO97-classification is clearly sparser¹⁴ than the occupational classifications the Finnish employer organizations use themselves. Hence it can be assumed that the FISCO97-classification will not be able to differentiate between all distinct occupations based on separate work tasks. Apparently, it can be thought that a part of the pay gap effects now linked entirely to age (not reported in Table 3 but available from the author on request) should be in fact read into the effects of occupational segregation.¹⁵ This suggests that

¹⁴"Sparse" occupational classification refers to a classification which includes a relatively small number of distinct job/occupational titles and, consequently, bundles up with higher likelihood even such occupations that are, in fact, regarded genuinely as separate in the labour market.

¹⁵If transfers to other better-paying jobs/work tasks are not reflected as corresponding changes of occupational titles in the occupational classification, then wage increases due to career development over time will be falsely interpreted as effects of age/tenure.

the effects of occupational segregation on the gender pay gap (Line 3 in Table 1) might be underestimated. Most likely the reservation concerns the year 1995 sub-sample, as Statistics Finland undertook a revision project in 2001 with the aim of being able to convert more precisely job/occupation titles in employers' own denser occupational classifications to FISCO97-occupations.

Thus, there is reason to believe that, in the 2004 calculations, the estimates of the pay gap effects of occupational segregation are no longer underestimated - at least to the same degree - as in 1995. At the same time, by comparing the estimated effects of occupational segregation for the years 1995 and 2004 with each other this would mean that the pay gap effects of occupational segregation might even have decreased (concerning especially part-timers) over the observation period 1995-2004. This interpretation is consistent with the earlier segregation indicator estimates of Laine & Napari (2008) for the Finnish private sector where the size of occupational segregation seems slowly have fallen during 1995-2004.

All in all, the effect of occupational segregation on the overall gender pay gap may have slightly diminished among part-timers while remaining roughly speaking unaltered among full-timers during the observation period 1995-2004. When comparing the results with earlier research one needs to take into account that contrary to earlier studies our estimates of the fraction of the overall pay gap due to occupational segregation control for the effects of segregation across collective labour agreements. Still our results repeat the story of earlier research as to attributing a significant fraction of the overall pay gap to occupational segregation even though defining it more narrowly than in the literature following the approach pioneered by Groshen (1991).

Finally, however, we complement the overall picture by bringing forth evidence that men and women being allocated differently across collective labour agreements is an extra factor behind the gender pay gap. Besides, its role is far from irrelevant in the context since according to our results it contributes significantly more to the overall gender pay gap than, for example, firm/establishment segregation does. In fact, among part-timers segregation by gender across collective agreements seems downright pivotal.

5 Discussion

We have reported new empirical evidence on how different forms of segregation affected the gender pay gap in the Finnish private sector in 1995 and

2004. We introduce and apply a new econometric approach naming it *the collective labour agreements classification method* (CLAC). The chosen approach differs from the mainstream studies on the male-female pay gap in a number of ways.

Especially, in our approach the border line between the distinct forms of segregation is drawn somewhat differently from the common approach. The fundamental novelty is a clear-cut separation of occupational, firm and educational segregations and their pay gap effects from the one due to men and women being differently distributed across collective labour agreements. There exists, in our view, a hierarchical difference between these two groups of segregation. Namely, in a labour market characterised by collective bargaining collective labour agreements form the very basis on how wages are determined. This means that the way and the extent wage determining factors, such as, age, experience, education, occupation *etc.* are being reflected in a person's wage depends in the first place on his/her collective agreement. For example, the size of the (partial) effect determining to which degree a certain occupation affects personal wages varies from a collective agreement to another.

Consequently, the key point as regards the definition of occupational segregation and especially the accurate estimation of its gender pay gap effects is being able to separate occupational segregation from the one due to men and women being differently distributed across collective labour agreements. In the main stream analysis the case when a man and a woman have different occupations is considered as representing occupational segregation independently of whether they share the same collective agreement or not. Instead according to the collective labour agreements classification method we can talk about "pure" occupational segregation only if a man and a woman having different occupational/job titles simultaneously belong under the same collective agreement. This is a major adjustment since now the effects of these two forms of segregation can, for the first time, be effectively differentiated from each other in empirical work. Furthermore, this conclusion generalizes to concern gender segregation across firms as well and applies actually to all the forms of segregation, such as, the ones across educational levels and fields.

Finally, a key difference in our study as to earlier studies is that now the estimated pay gap effects of occupational segregation are "purified" from the pay gap effects of men and women being differently allocated across collective agreements. Still according to our results occupational segregation - though

more narrowly defined as in earlier studies - explains a fairly significant proportion of about 20 to 30 percent of the overall gender pay gap. The results are consistent with earlier studies where occupational segregation is a major factor behind the overall gender pay gap. Furthermore, we introduce completely new evidence showing segregation across collective agreements being far from insignificant as an explanatory factor for the overall pay gap especially among part-timers and contributing even in the subset of full-timers clearly more than, for example, firm segregation to the size of the overall difference in pay.

References

- Arai, M., Nekby, L. & Thoursie, P. S. (2004), Is it what you do or where you work that matters most? gender composition and the gender wage gap revisited, Research Papers in Economics 2004:10, Stockholm University, Department of Economics.
- Bayard, K., Neumark, J. H. D. & Troske, K. (2003), 'New evidence on sex segregation and sex differences in wages from matched employee-employer data', *Journal of Labor Economics* **21**(4), 887–922.
- Groshen, E. L. (1991), 'The structure of female/male wage differential: Is it who you are, what you do, or where you work?', *Journal of Human Resources* **26**(3), 457–72.
- Gupta, N. D. & Rothstein, D. S. (2005), 'The impact of worker and establishment-level characteristics on male-female wage differentials: Evidence from danish matched employee-employer data', *Labour* **19**(1), 1–34.
- Koev, E. (2003), Combining Classification and Hedonic Quality Adjustment in Constructing House Price Index, Licentiate thesis, Department of Economics, University of Helsinki.
- Korkeamäki, O. & Kyyrä, T. (2005), 'The gender wage gap and sex segregation in Finland', *Finnish Economic Papers* **18**(2), 57–71.
- Laine, P. & Napari, S. (2008), Sukupuolten ammatilliset erot suomalaisilla työmarkkinoilla ("occupational differences by gender in the finnish labour market"). Unpublished research report (in Finnish), the ETLA Economic Research Centre and Statistics Finland.

Oaxaca, R. (1973), 'Male-female wage differentials in urban labor markets',
International Economic Review **14**(3), 693–709.

List of Tables

| | | |
|---|---|----|
| 1 | Forms of segregation and their effect on the aggregate pay gap. | 24 |
| 2 | Aggregated Estimation Diagnostics across Repeated Gender-specific Wage Model Estimations - Collective Agreement Classification Method. | 25 |
| 3 | The Aggregated Gender Pay Gaps and the Proportions Explained by the Key Background Factors - Collective Agreements Classification Method. | 26 |

Table 1: Forms of segregation and their effect on the aggregate pay gap.

| Form of segregation | Formula | Interpretation |
|----------------------------------|--|--|
| Occupational segregation | $\sum_{c=1}^C \sum_{o=2}^{O_c} \frac{\bar{W}_c^f N_c^f}{\bar{W}^f N^f} \hat{\delta}_{co}^m \left(\frac{M_{co}}{M_c} - \frac{F_{co}}{F_c} \right)$ | How many log-percent higher women's mean wage would be after eliminating occupational segregation. |
| Firm segregation | $\sum_{c=1}^C \sum_{e=2}^{E_c} \frac{\bar{W}_c^f N_c^f}{\bar{W}^f N^f} \hat{\gamma}_{ce}^m \left(\frac{M_{ce}}{M_c} - \frac{F_{ce}}{F_c} \right)$ | How many log-percent higher women's mean wage would be after eliminating firm segregation. |
| Collective agreement segregation | $\ln \left\{ \frac{W^m}{W^f} \right\} - \log \text{Lasp} W_f^m$ | How many log-percent higher women's mean wage would be after eliminating collective agreement segregation. |

Notes:

¹ $\frac{M_{co}}{M_c}$ and $\frac{F_{co}}{F_c}$ are proportions of men/women working in occupation o of all the men/women under collective agreement c .

Table 2: Aggregated Estimation Diagnostics across Repeated Gender-specific Wage Model Estimations - Collective Agreement Classification Method.

| | Full- and part-timers | | | | Only full-timers | | | |
|------------------------|-----------------------|---------|---------|---------|------------------|---------|---------|---------|
| | 1995 | | 2004 | | 1995 | | 2004 | |
| | Men | Wo-men | Men | Wo-men | Men | Wo-men | Men | Wo-men |
| RMSE | 0.17 | 0.22 | 0.18 | 0.25 | 0.13 | 0.13 | 0.14 | 0.13 |
| N | 223629 | 203450 | 314847 | 266616 | 215127 | 182154 | 299246 | 217893 |
| Part-timers (%) | 3.8% | 13.9% | 5.2% | 18.9% | 0.0% | 0.0% | 0.0% | 0.0% |
| C | 139 | 139 | 149 | 149 | 139 | 139 | 149 | 149 |
| E | 6432 | 6858 | 8017 | 8705 | 6314 | 6756 | 7722 | 8448 |
| | (11.5%) | (17.0%) | (9.7%) | (16.8%) | (12.6%) | (18.4%) | (11.8%) | (19.4%) |
| C*E | 10437 | 11382 | 24670 | 26706 | 10249 | 11105 | 23535 | 23890 |
| | (20.4%) | (27.0%) | (51.7%) | (55.3%) | (21.8%) | (27.8%) | (53.2%) | (53.9%) |
| O | 431 | 397 | 414 | 399 | 431 | 393 | 409 | 397 |
| | (13.7%) | (6.3%) | (5.8%) | (2.1%) | (14.2%) | (5.9%) | (5.9%) | (3.0%) |
| C*O | 4023 | 3620 | 7383 | 6459 | 3974 | 3560 | 7171 | 6200 |
| | (33.2%) | (25.8%) | (40.7%) | (32.2%) | (33.6%) | (25.8%) | (41.6%) | (32.5%) |

Notes:

- ¹ R^2 and $RMSE$ denote weighted means of CLA -specific squared R s and root mean squared errors respectively. The weights consist of proportions of separate collective agreements c among men/women ($\frac{M_c}{M}$ vs. $\frac{F_c}{F}$, $c = (1, \dots, C)$).
- ² N is the overall number of men/women in the estimation sample across the C collective agreements.
- ³ C is the number of collective agreements forming the estimation classification for men and women, respectively.
- ⁴ E and O are the numbers of firms and occupations where men respectively women appear. In parentheses below the percentage of firms and occupations being fully (100 %) dominated by the gender in question as counted from the numbers immediately above.
- ⁵ $C*E$ and $C*O$ are the numbers of different combinations of collective agreements with firms and occupations where men respectively women appear. In parentheses below the percentage of combinations being fully (100 %) male- vs. female-dominated as counted from the numbers immediately above.

Table 3: The Aggregated Gender Pay Gaps and the Proportions Explained by the Key Background Factors - Collective Agreements Classification Method.

| OVERALL PAY DIFFERENCE AND ITS ELEMENTS | Full- and part-timers | | Full-timers only | |
|--|--|---------------|-------------------------|---------------|
| | 1995 | 2004 | 1995 | 2004 |
| The overall pay gap | | | | |
| 1. How many percent higher is men's average wage compared to women's average wage | 25.4% | 28.8% | 20.1% | 18.2% |
| Key explanatory factors | | | | |
| | Proportion (%) of the overall gap (1.) due to the corresp. factor | | | |
| 2. Agreement segregation: men over-represented in better paying collective agreements | 24.4% | 42.7% | 5.0% | 13.2% |
| 3. Occupational segregation: within the same collective agreement men over-represented in better paying occupations | 23.6% | 19.1% | 27.9% | 29.1% |
| 4. Firm/establishment segregation: within the same collective agreement men over-represented in better paying firms | 3.1% | 1.7% | 3.5% | 3.8% |
| 5. Segregation effects together | 51.1% | 61.8% | 36.4% | 46.1% |
| 6. Other effects: within the same collective agreement men earn more for other reasons | 48.9% | 38.2% | 63.6% | 53.9% |
| 7. Segregation + other effects | 100.0% | 100.0% | 100.0% | 100.0% |