Työpapereita
Working Papers

179

GENDER WAGE DIFFERENTIALS IN THE FINNISH LABOUR MARKET

Juhana Vartiainen

PALKANSAAJIEN TUTKIMUSLAITOS LABOUR INSTITUTE FOR ECONOMIC RESEARCH

PALKANSAAJIEN TUTKIMUSLAITOS •TYÖPAPEREITA
LABOUR INSTITUTE FOR ECONOMIC RESEARCH • DISCUSSION PAPERS

179

GENDER WAGE DIFFERENTIALS IN THE FINNISH LABOUR MARKET

Juhana Vartiainen

Helsinki 2002

ISBN 952–5071–67–7
ISSN 1457–2923

Gender wage differentials in the Finnish labour market*

Juhana Vartiainen

March 4, 2002

Abstract

A comprehensive analysis of the gender wage differential among Finnish full time employees is reported. Oaxaca decompositions show that the the overall wage gap of about 21.5 per cent cannot be accounted for by individual characteristics, since age and educational qualifications are rather similar for men and women. When industry and occupational qualifications are included in the regressor matrix, the unexplained part shrinks to about 50 per cent of the gross differential. An even larger part of the gross differential can be explained in sector-specific analyses with a dense set of occupational dummies. In a less standard part of the analysis, we characterise the distribution of the unexplained wage gap across the variable space. It turns out the women with a high wage predictor, that is, women with good educational qualifications in well-remunerated occupations, drag the most behind their male colleagues endowed with similar characteristics. JEL numbers J31, J70, J71.

Tiivistelmä

Sukupuolten palkkaero on kokopäiväisten palkansaajien parissa noin 20 prosenttia eli naiset ansaitsevat osapuilleen neljä viidesosaa miesten ansioista. Artikkelissa analysoidaan taloustieteen keinoin tätä palkkaeroa. Se on yhteenveto laajemmasta tutkimusraportista

*This paper summarises some results of the project "Measuring and monitoring gender wage differentials in the Finnish labour market", commissioned by the Ministry of Social Affairs and Health and carried out by the author. The objective of the project was to present a comprehensive survey of the gender wage differential in the Finnish labour market as well as to suggest statistical procedures that could be used for a more regular monitoring of the gender wage gap. A longer report in Finnish (55 pages of main text plus a 249-page statistical annex) is available (see Vartiainen (2001)). In that report, a more comprehesive set of sector-specific statistics is presented. I am grateful to Equality Ombudsman Pirkko Mäkinen and Research Officer Anita Haataja for sponsoring and supporting this research. I am grateful to Reija Lilja for useful comments. I also want to thank the various employers' organisations who made possible the use of their sector-specific wage databases, as well as Statistics Finland, whose co-operation has been primordial. Any errors or weaknesses are my own, of course. Author's affiliation and e-mail: Labour Institute for Economic Research, Helsinki. juhana.vartiainen@labour.fi

(Vartiainen 2001), jonka pyrkimyksenä oli kartoittaa kattavasti sukupuolten palkkaeroa Suomessa sekä ehdottaa tilastointi- ja seurantatapoja, jolla tätä eroa voitaisiin vastaisuudessa seurata. Tulokset osoittavat, että mainitusta 20 prosentin palkkaerosta selittyy pois noin puolet, jos selittäjinä käytetään henkilökohtaisia muuttujia sekä toimialoille ja ammattiryhmiin valikoitumista. Sukupuolten välillä ei ole merkittävää eroa iässä ja koulutuksessa, joten palkkaeron selittyvä osa perustuu toimiala- ja ammattiryhmävalikoitumiseen. Samanlainen hajoitelma lasketaan myös talouden viidelle sektorille (palvelutyönantajat, teollisuuden tuntipalkkaiset, teollisuuden kuukausipalkkaiset, kunnat, valtio) erikseen. Tulokset osoittavat, että julkisyhteisöjen parissa toteutuu pitkälti "sama palkka samasta työstä" -periaate, ja lopullinen palkkaero syntyy lähes kokonaan naisten ja miesten ammattiryhmävalikoitumisen kautta.

Teollisuuden piirissä ammatteihin valikoituminen ei ole yhtä merkittävä sukupuolten palkkaeron taustatekijä, mutta iän palkkaa nostava vaikutus on naisilla selvästi vähäisempi kuin miehillä.

Artikkelissa tarkastellaan myös selittymättömän palkkaeron jakautumista eri palkkatasoille. Selittymätön palkkakaula on suurempi korkean palkkatason tehtävissä kuin matalapalkkatehtävissä. Sukupuolten ammatillista segregoitumista kuvaavat indeksit osoittavat segregoitumisen hidasta heikkenimistä.

Contents

1 Introduction

This paper presents the most important stylised facts about the gender wage differential in the Finnish economy. We characterise the differential in two ways. The first is a standard one: we use an Oaxaca decomposition (see Oaxaca (1973)) of the log of the monthly wage and present the decomposition results. The results are presented both with and without an occupational dummy variable, and, in both cases, using both the male and female parameter estimates as the reference structure. The estimates are based on a 20 per cent sample of the Earnings Database ("palkkarakennetilasto") produced and maintained by Statistics Finland. This database covers almost the entire salaried workforce of the economy. Our study concerns the wages of full-time workers and employees.

The first decompositions exploit the entire sample. The data are then decomposed into subsets: service sector, manufacturing (salaried employees and wage earners paid by the hour are treated separately), local government and central government. Oaxaca decomposition results are presented for these subsets as well.

The second set of results is more original. The Oaxaca decomposition reveals the average unexplained wage gap between sexes but nothing about its distribution. Yet, from the point of view of society's welfare and gender equality, it is quite interesting to ask how that unexplained wage gap is distributed in different parts of the variable space. Most studies of the wage differential do not appreciate this point, with the exception of Stephen Jenkins who has in an interesting way applied the ideas familiar from the theory of income distribution to the study of group differentials (see Jenkins (1994)). However, since the aim of this paper is to get out the most important stylised facts, we adopt a far simpler procedure, described in more detail in section 4.

2 Data

We exploit the Earnings Structure Database maintained by Statistics Finland. It is based on a compilation of wage registries kept by various employer organisations and comprises about 1.3 million wage-earners. The variables include a detailed breakdown of earnings variables plus background variables like age and schooling, together with industrial sector and occupational dummies. We sampled a 20 per cent subset of that database, so that the probability of selection into the sample was proportional to the sampling weight used by Statistics Finland; by these means, we sought to emulate simple random sampling as fully as possible.

The variable of interest is *monthly wage income* as computed by Statistics Finland. That variable has been computed by dividing all wage and salary incomes of the individual by his/her total working hours, both measured under a month. The aim has been to include all such wage items that have some continuity: the base wage, extras and bonuses based on working conditions, eventual extra pay due to overtime and the value of perks. Similarly, the denominator was an estimated total labour input as measured in time units.

We excluded all part-time employees from our analysis. This was motivated by the will to compare men and women who would be as similar as possible. The inclusion of part-time work would point the way to another kind of analysis in which the differential labour supply decisions of men and women would play a more central role.

3 The Oaxaca decomposition

3.1 Definition

The Oaxaca decomposition was presented in (Oaxaca, 1973), and we follow his procedure. If the male and female geometric mean wages are denoted by $\underline{W}_M$ and $\underline{W}_F$, we can decompose the log-differential of geometric means as follows:

$$\Delta \equiv log(\underline{W}_M/\underline{W}_F) = (log(\underline{W}_M) - log\underline{W}_{0F}) + (log(\underline{W}_{0F}) - log\underline{W}_F), \tag{1}$$

in which we denote by $\underline{W}_{0F}$ a hypothetical distortion-free or discrimination-free mean wage for women. Furthermore, we can summarise the variation in the male and female wage cross section samples by using the following commonplace statistical models:

$$w_{i,m} = X^i \beta^M + \epsilon_i \tag{2}$$

and

$$w_{j,f} = X^j \beta^F + \epsilon_j \tag{3}$$

in which $w_{i,m}$ is the log wage of man i and $w_{j,f}$ is the log wage of woman j, β^M and β^F are the coefficients that determine the effect of characteristics on pay and the X^i and X^j are the vectors of characteristics of man i and woman j. Taking the arithmetic average of equations (2) and (3), the stochastic ϵ-terms drop away. Designating arithmetic mean by an underlined variable, we have

$$\underline{w}_m = \underline{X}^M \beta^M \tag{4}$$

and

$$\underline{w}_f = \underline{X}^F \beta^F, \tag{5}$$

which simply says that mean wages are predicted by using mean characteristics. Since $\underline{w}$ is the mean of the log, it is the log of the geometric mean $\underline{W}$. We can then plug equations (4) and (5) into (1) to obtain:

$$\Delta = (\underline{X}^M - \underline{X}^F)\beta^M + \underline{X}^F(\beta^M - \beta^F). \tag{6}$$

Thus, the gross differential Δ is decomposed into the effect of differences in average characteristics (the first term) and the effect of different treatment (or different "pricing") of characteristics. In this paper, we shall speak of the contribution of different characteristics on the one hand and of the unexplained differential or "differential pricing" effect on the other.

The literature often speaks roughly of the latter part as "discrimination", but this is a dubious usage of words. The unexplained part contains the effects of all those variables that are not part of the X-matrix; economic theory suggests many ways in which unobserved variables may be differently distributed across men and women and may thereby contribute to the wage differential as well. If there is outright discrimination in wage-setting, it is probably a part of the unexplained component.

In the above decomposition, we implicitly chose the male coefficient vector as a reference structure with which we evaluated (i.e. "priced") the contribution of the differences in characteristics. We could equally well choose the female wage structure and get the analogous decomposition

$$\Delta = (\underline{X}^F - \underline{X}^M)\beta^F + \underline{X}^M(\beta^F - \beta^M). \tag{7}$$

In the subsequent literature, it has been emphasised that there is in general no unambiguouosly best way to choose the reference wage structure (this is really an index number problem). Ideally, if we knew a "right", discrimination-free structure β, we should use it to evaluate the effect of differences in characteristics. Some authors use the coefficients estimated from the pooled sample (without female dummy) as the reference structure.

More generally, the reference structure can be chosen as a matrix weighted average of the male and female β-vectors (see Oaxaca and Ransom (1994)). The choice of the weighting matrix should be guided by economic theory, but the literature on this issue is still quite rudimentary. We have therefore chosen simply to present the computations using both male and female coefficients as reference. The advantage of this procedure is that it enables a clear interpretation: using the male (female) structure as reference and computing the effect of different characteristics yields an answer to the question "how much would remain of the gross differential if women (men) were treated as men (women) with similar characateristics?".

Note that the above decompositions can be expanded to yield the contributions of each single regressor: how much of the gross differential is due to the gender difference in means of the regressor and how much is due to differential pricing of the regressor. This interpretation runs into trouble when we include categorical variables in the regressor matrix X. As is shown in Oaxaca and Ransom (1999), the presence of categorical variables implies that the unexplained part (i.e. the "differential pricing") part of each categorical variable cannot be uniquely determined. This is intuitively clear, since the regression of wages on categorical variables implies a choice of a reference group, and the results of the Oaxaca decomposition are not independent of this choice. The combined "differential pricing" contribution of all the within-cell unexplained gaps can be uniquely computed, however, and we shall report that in our tables.

Table 1: Decomposition for 1998, male reference coefficients.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	.0047	n.u	n.u
education	.0135	n.u	n.u
employer size	.0028	n.u	n.u
constant plus selection	.0209	.2163	.2372
age in years	-.0416	-.1788	-.2204
age squared	.0268	.1543	.1811
n. of children under 18	.0017	.0198	.0215
n. of children under 7	-.0003	-.0008	-.0011
Sum total	.0074	.2109	.2148
st.dev	.0016	.0015	ass=0

Table 2: Decomposition for 1998, female reference coefficients.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u.	n.u.
temporary	.0032	n.u	n.u
education	-.0025	n.u	n.u
employer size	.0017	n.u	n.u
constant plus selection	.0025	.2348	.2372
age in years	-.0489	-.1715	-.2204
age squared	.0379	.1432	.1811
n. of children under 18	-.0015	.023	.0215
n. of children under 7	.0001	-.0012	-.0011
Sum total	-.0099	.2282	.2148
st.dev	.0013	.0014	ass=0

3.2 Results for all employees

The decompositions were computed for years 1996 through 1998. We focus on the last year, since the results for years 1996 and 1997 are largely similar. The following four tables display the results of decomposition (6) for year 1998. The first two tables are based on a regression matrix X that contains the personal characteristic variables age and education as well as the employer size.

The tables are read as follows. The cell at the low-end right-end corner (at the intersection of row “Sum total” and column “Sum” tells the gross log wage differential (21.48 in Table 1, for example). The two entries on the same row display the two terms of decomposition (6); the first one is the effect of differences in characteristics and the second one is the unexplained or differential pricing component. The other rows above that final row tell the same information for different variables. The first group of variables (above the first horizontal double line) is the group of categorical variables. For those ones, one can only compute the effect of the difference in means in an unambiguous way; the pricing effect is reported as the aggregate sum of the within-cell unexplained gaps (the abbreviation “n.u” stands for “not unique”). This is reported on the row “constant plus selection”. The second group of variables are the continuous ones: age, age squared and the number of children. Finally, the row “Sum total” adds up all the contributions.

Tables 1 and 2 make it clear that personal characteristics plus firm size do not go far in explaining the wage differential. Of the overall differential of 21.5 percentage points, almost nothing is explained by differences in individual characteristics and employer size. Almost all of the wage differential is due to a constant term that we cannot allocate to any specific categorical variable.

The next two tables 3 and 4 show the analogous results when an occupational variable[1] plus a sector variable[2] (“industry”) are added to the regressor matrix. We see now that about half of the gross wage differential can be accounted for by different endowments when male coefficients are used, and about one third when female coefficients are used. The tables show that the “industry” variable and the “occupation” variable together generate the explained part of about 10 percentage points.

[1] This is the “isco” -variable produced and used by Statistics Finland.

[2] The “nace” variable of Statistics Finland.

Table 3: Decomposition for 1998, male reference coefficients.

Variable	Difference in characteristics	Differential pricing	Sum total
constant		n.u	n.u
temporary	.0029	n.u	n.u
occupation	.0522	n.u	n.u
industry	.0582	n.u	n.u
education	.0097	n.u	n.u
employer size	.0028	n.u	n.u
constant plus selection	.1258	.0526	.1784
age in years	-.0344	.0072	-.0272
age squared	.0236	.0267	.0503
n. of children under 18	.0013	.0141	.0154
n. of children under 7	-.0005	.0007	.0002
Sum total	.1158	.1013	.2145
st.dev	.0029	.0029	ass=0

Table 4: Decomposition for 1998, female reference coefficients.

Variable	Difference in characteristics	Differential pricing	Sum total
constant		n.u.	n.u.
temporary	.0027	n.u	n.u
occupation	.0461	n.u	n.u
industry	.0328	n.u	n.u
education	.0016	n.u	n.u
employer size	.001	n.u	n.u
constant plus selection	.0842	.0942	.1784
age in years	-.0341	.0069	-.0272
age squared	.0255	.0248	.0503
n. of children under 18	-.0009	.0163	.0154
n. of children under 7	-.0009	.0011	.0002
Sum total	.0739	.1432	.2145
st.dev	.0026	.0027	ass=0

We also see that only age, industry and occupation play a role in the generation of the gender wage differential. The differential treatment effect of children accounts for a couple of percentage points; thus, children lead to a somewhat higher wage handicap for women, but this effect is quite weak in comparison with other effects.

Table 5: Decomposition for 1998, private services.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	.0016	n.u	n.u
occupation (detailed)	.1467	n.u	n.u
industry	.0204	n.u	n.u
education	.0137	n.u	n.u
employer size	-.0038	n.u	n.u
constant plus selection	.1786	-.0292	.1494
share of extra hours	.0027	-.0028	-.0001
age in years	-.0533	.1201	.0668
age squared	.0358	.0038	.0396
n. of children under 18	.002	.0195	.0215
n. of children under 7	-.0011	-.0015	-.0026
Sum	.1647	.1099	.2724
st.dev	.0063	.0061	ass=0

3.3 Results for specific sectors

The following tables exhibit similar decompositions for five subsets of the salaried labour force[3]:

- private service sector;
- salaried employees of the manufacturing industry (monthly pay)
- wage earners of the manufacturing industry (pay by hour)
- local government workers
- central government workers.

To save space, we report only those computations that used male reference coefficients and included the widest set of explanatory variables. In particular, it should be emphasised that the estimations on which the following tables are based included a sector-specific occupational classification which is in general denser than the general occupational classification used for the estimates reported for the entire workforce.

Furthermore, the number of occupational classes varies from sector to sector; it is very large for local and central government workers, fairly large for manufacturing wage-earners and private service sector employees and quite low for manufacturing salaried employees. As expected, this is reflected in the share of the gender wage differential that can be ascribed to differential endowments[4].

[3]In the data base, these subsets are formed by the institutional affiliation of the person's employer. The Finnish designations for these groupings are "Palvelutyönantajat", "Teollisuuden kuukausipalkkaiset", "Teollisuuden tuntipalkkaiset", "Kunnat", "Valtio".

[4]Thus, as was already emphasised by Ronald Oaxaca in his original contribution of Ronald Oaxaca (1973), the use of occupational dummies is controversial. In the limit, as all individuals have a somewhat unique job, we could explain away the entire wage differential by occupational categories.

Table 6: Decomposition for 1998, manufacturing, salaried employees.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	.0035	n.u	n.u
occupation (detailed)	.1121	n.u	n.u
industry	-.0047	n.u	n.u
education	.0437	n.u	n.u
employer size	-.0016	n.u	n.u
constant plus selection	.1529	-.0393	.1137
share of extra hours	.0071	-.0011	.006
age in years	.0158	.1318	.1476
age squared	-.0122	.0349	.0227
n. of children under 18	.0021	.0191	.0212
n. of children under 7	-.0006	-.0007	-.0013
work experience	-.001	.0081	.0071
Sum	.1643	.1529	.3172
st.dev	.0078	.0078	ass=0

Table 7: Decomposition for 1998, manufacturing, wage earners.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	-.001	n.u	n.u
occupation (detailed)	.0233	n.u	n.u
industry	.0225	n.u	n.u
education	.0048	n.u	n.u
employer size	-.0041	n.u	n.u
constant plus selection	.0454	-.0853	-.0399
share of extra hours	.0178	-.0035	.0143
age in years	-.042	.3393	.2973
age squared	.0408	-.1735	-.1327
n. of children under 18	.0004	.0047	.0051
n. of children under 7	.0004	.0013	.0017
work experience	.0021	.0061	.0082
Sum	.0649	.0892	.1498
st.dev	.0025	.0026	ass=0

The results of tables (5) through (9) can be summarised as follows. The gross wage differential is lowest among manufacturing workers and local government workers (just under 20 per cent and just over 20 per cent, respectively). It is highest for the salaried manufacturing employees (over 30 per cent). The central government personnel and the employees of private service sector firms occupy the middle ground, with a differential around 25 per cent.

In these sector-specific computations, we have had to discard some observations that belong to the smallest and most segregated occupational categories, since no reliable statistical inference is possible if a cell contains only a couple of female or a couple of male observations[5]. This truncation of the data has a different effect in different sectors. Deleting the smallest and most segregated occupational categories leads to a large drop in the gender differential of manufacturing wage-earners; as is apparent from table 7, the gross wage differential is as low as 15 per cent. Thus, the most segregated occupational categories tend to be populated by high-earning males and low-earning females. Among the local government workers, this truncation has the opposite effect of increasing the gross wage differential. Small and segregated occupational categories tend there to include female with high earnings and males with low earnings.

The tables indicate that segregated occupational categorisation explains away most of the wage differential in the case of local and central government employees. In both cases, 80 to 90 per cent of the gross differential is explained by differences in the means of the regressors, and segregation into different occupations is in turn responsible for about 4/5 of that effect (see the entry at the intersection of row “occupational category” and column “Difference in characteristics” in tables 8 and 9: it is of the order of 15 percentage points in both sectors). The effect of differential treatment of age is quite limited.

The picture is somewhat different amongst the manufacturing industries employees and workers. There, the share of the gross wage differential that can be explained away by different characteristics is in general lower, about half of the gross differential for salaried employees and about a third of the gross differential for workers. A closer look at tables 6 and 7 reveals that the effect of differential selection into occupational categories is much lower in manufacturing than in the public sector ; it only contributes a couple of percentage points among wage earners and about 10 percentage points among salaried employees. The effect of differential treatment of age, by contrast, is an important factor: by itself, it creates an unexplained wage differential of about 16 percentage points in both subsets of manufacturing personnel. Note also that the sum of unexplained mean gaps (see the entry at the intersection of row “constant plus selection” and column “Differential pricing” in tables 6 and 7) is negative. Thus, within occupational categories, manufacturing firms tend to pay more or less the same to young people, but the accruing of age leads to a widening gap between senior males and senior females.

Finally, for the employees of private service sector firms (see table 5) , the occupational variable is also quite important: segregated selection into occupations generates a wage differential of almost 15 percentage points. Since differential age treatment plays an important role as well, the computation ends up with a substantial gross differential of 27 percentage points, of which roughly 16 percentage points are explained by the regressors, mostly occupational selection.

To sum up, the public sector seems prima facie to exemplify the principle of “similar pay for similar work”, since the remaining unexplained differential is quite low. Or, to put it in another way, aspirations towards a lower gross differential should focus on the adequacy and

[5]We have required that an occupational category cell have at least 5 observations.

Table 8: Decomposition for 1998, local government.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	-.0015	n.u	n.u
occupation (detailed)	.1414	n.u	n.u
industry	.0123	n.u	n.u
education	.0482	n.u	n.u
employer size	-.0022	n.u	n.u
constant plus selection	.1982	.0018	.2
age in years	-.002	-.0287	-.0307
age squared	.0012	.0483	.0495
n. of children under 18	.0007	.0074	.0081
n. of children under 7	-.0007	.0012	.0005
work experience	-.0017	.0069	.0052
Sum	.1956	.037	.2319
st.dev	.0072	.007	ass=0

objectivity of the occupational job classification and women's opportunities to move upward on the job ladder. In manufacturing, by contrast, occupational selection is not such an engine of wage differentiation, but there seems to be an important age effect: manufacturing is a much less attractive employer to senior women than to senior men.

Table 9: Decomposition for 1998, central government.

Variable	Difference in characteristics	Differential pricing	Sum
constant		n.u	n.u
temporary	-.0006	n.u	n.u
occupation (detailed)	.1665	n.u	n.u
industry	-.0009	n.u	n.u
education	.0307	n.u	n.u
employer size	.0018	n.u	n.u
constant plus selection	.1976	-.0457	.1519
age in years	-.0333	.0992	.0659
age squared	.0274	-.0531	-.0257
n. of children under 18	.0006	.0037	.0043
n. of children under 7	.0011	.0029	.004
work experience	.0007	.0132	.0139
Sum	.194	.0202	.2136
st.dev	.0059	.0058	ass=0

4 The distribution of the unexplained gap

The Oaxaca decompositions were complemented by an analysis of the distribution of the gap according to a woman's predicted wage. Oaxaca decompositions tell the mean of the unexplained gap, but contain no information on the distribution of this component. Yet, from the point of view of society's preferences and the political assessment of eventual discrimination, it is probably not at all immaterial whether this unexplained component is evenly distributed or is concentrated into some specific sections of the variable space. For example, if women's earnings, on average, amount to 90 per cent of the earnings of men endowed with similar characteristics (age, education, occupation, etc.), there might one group of women who earns as much as similar men do and another which earns 20 per cent less as similar men.

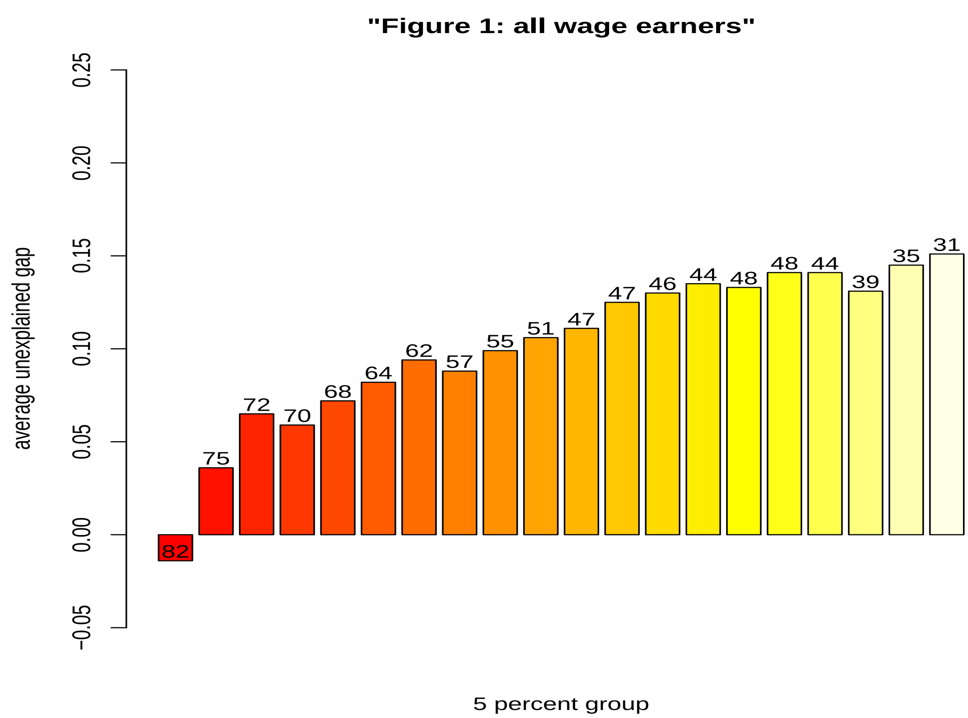
In the absence of obvious examples to copy, we adopted the following approach. We first estimated a conventional log-wage regression for male observations. Then, for each female observation, we computed the prediction of the wage using the estimated parameters of the male wage equation. This variable, nicknamed a woman's "male-wage", tells how much the woman in question would earn if she were lucky enough to be a man with precisely similar characteristics.

We then sorted (ordered) the sample of women according to their "male-wage" and split the sample into 20 groups of equal size ("vigintiles"). Thus, in the first group, there are those women whose predicted male-wage is the lowest, and in the last group there are those women whose predicted male-wage is the highest. We then computed the average unexplained wage gap within each group; thus, the resulting statistic tells for each group the average distance to males with similar characteristics.

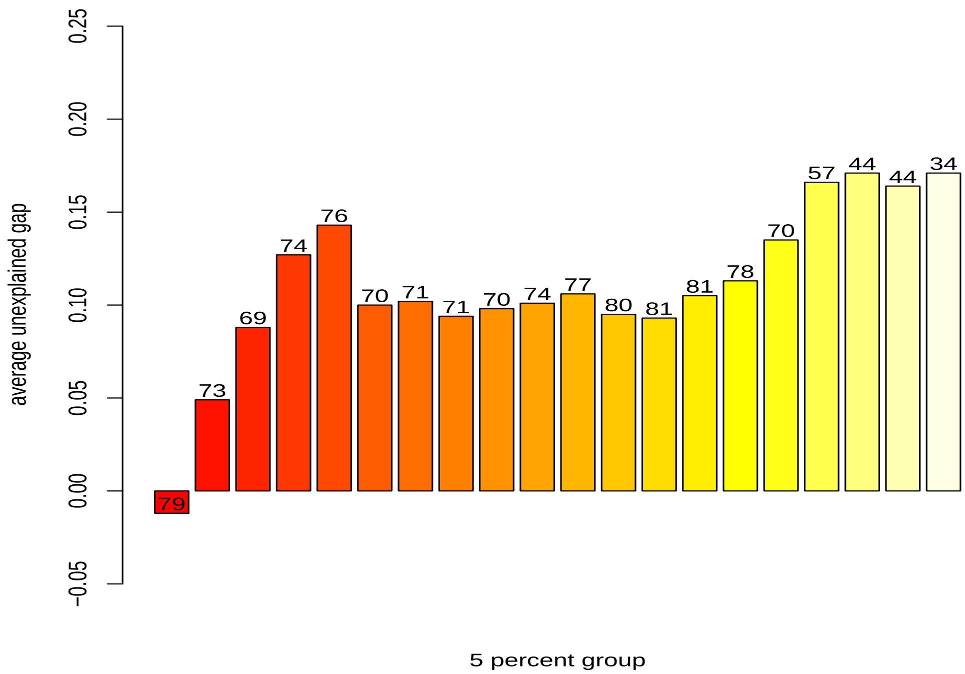
The results of these computations are presented in Figures 1 through 6. Figure 1 shows the average gaps by vigintile in the entire sample; it is based on the male wage regression which generated tables (3) and (4). Figures 2 through 6 display the same information for the sector subsets of the sample (cf. section 3.3).

These computations are based on the widest set of regressors, including occupational categories. On top of each bar, we indicate a 2-digit number; it is the share of women in the wage bracket implied by the vigintile. Thus, in general, the share of women is quite high in the lowest 5-percent groups and it declines as the male-wage predictor grows.

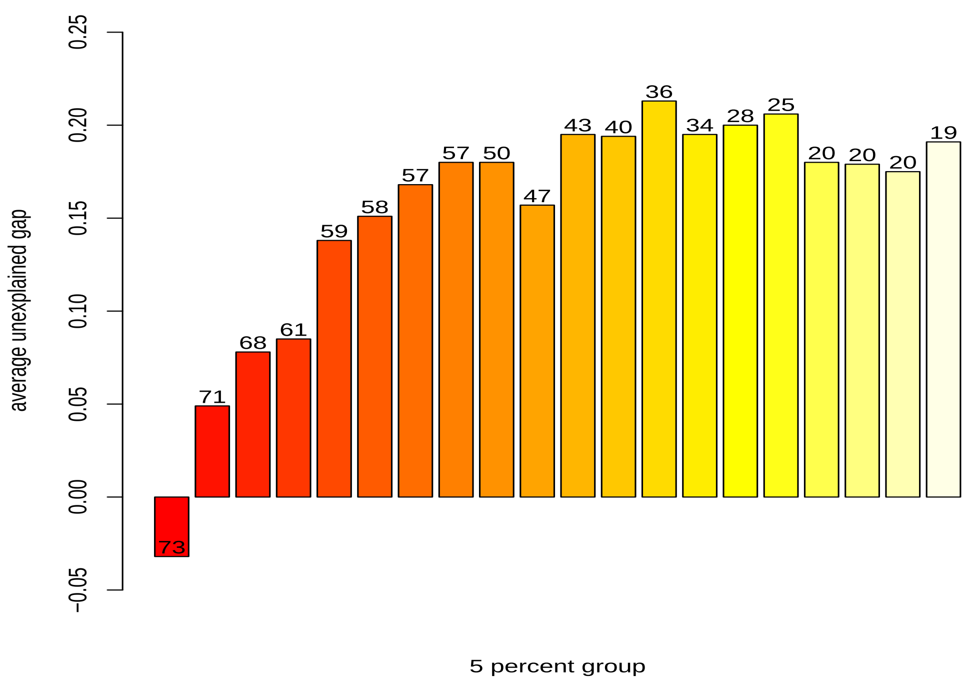
The interesting finding is that the unexplained gap is in general quite low for those women whose qualifications would generate a low wage prediction by male coefficients. Since these are in general women with low earnings, we can conclude that the gap is lowest among low wage earners. The gap then increases steadily as we move to higher income groups, and it is highest either at the very end of the group scale or a little below. Thus, our preliminary conclusion is that it is women with good educational qualifications working in high-wage occupations who drag the most behind their male colleagues.



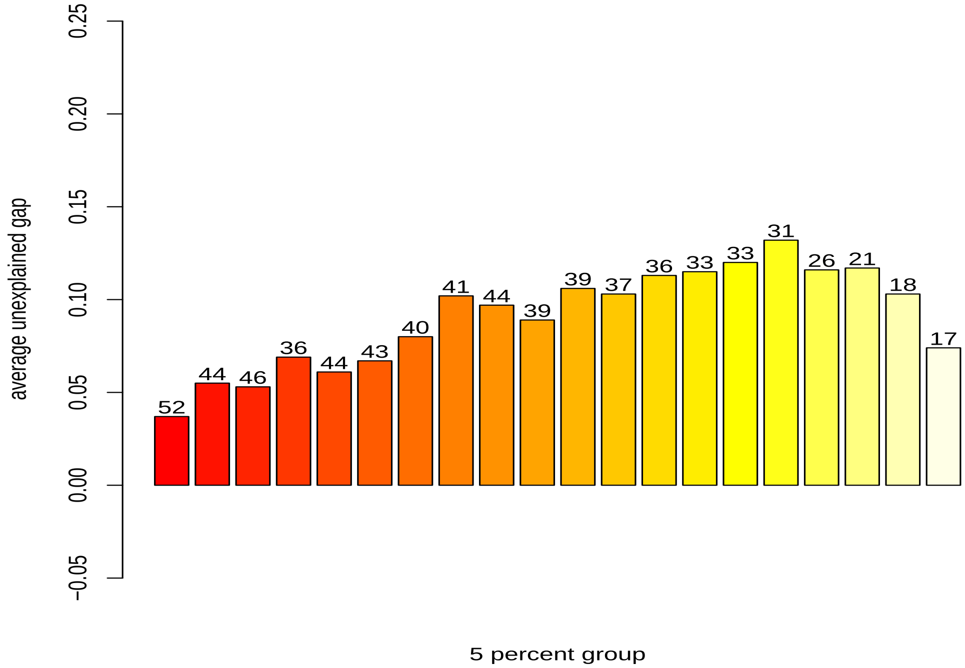
"Figure 2: private service sector"



"Figure 3: manufacturing employees"



"Figure 4: manufacturing workers"



"Figure 5: local government"

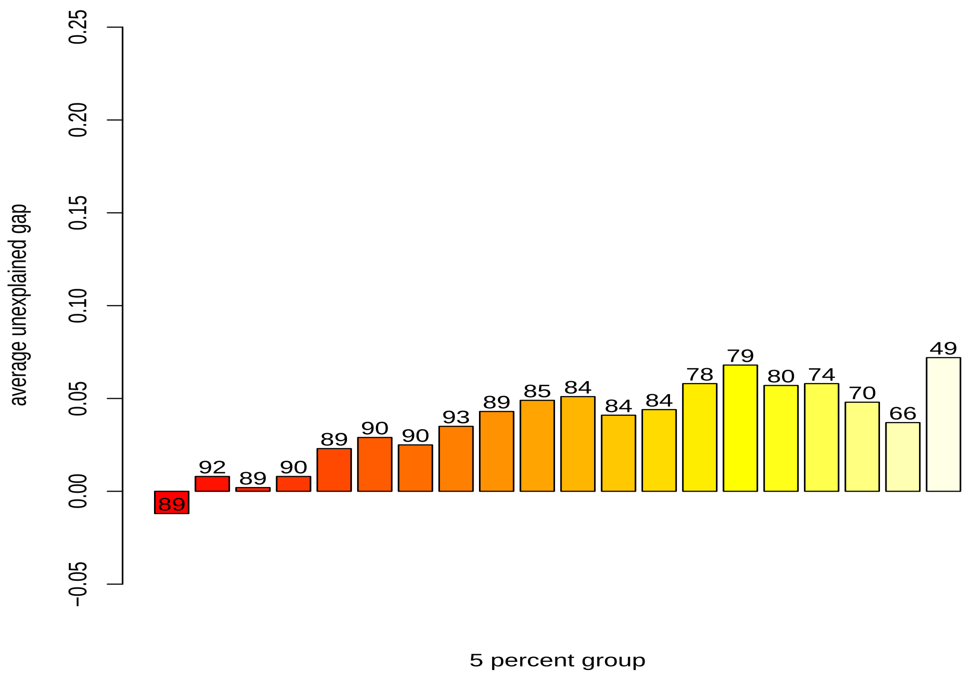
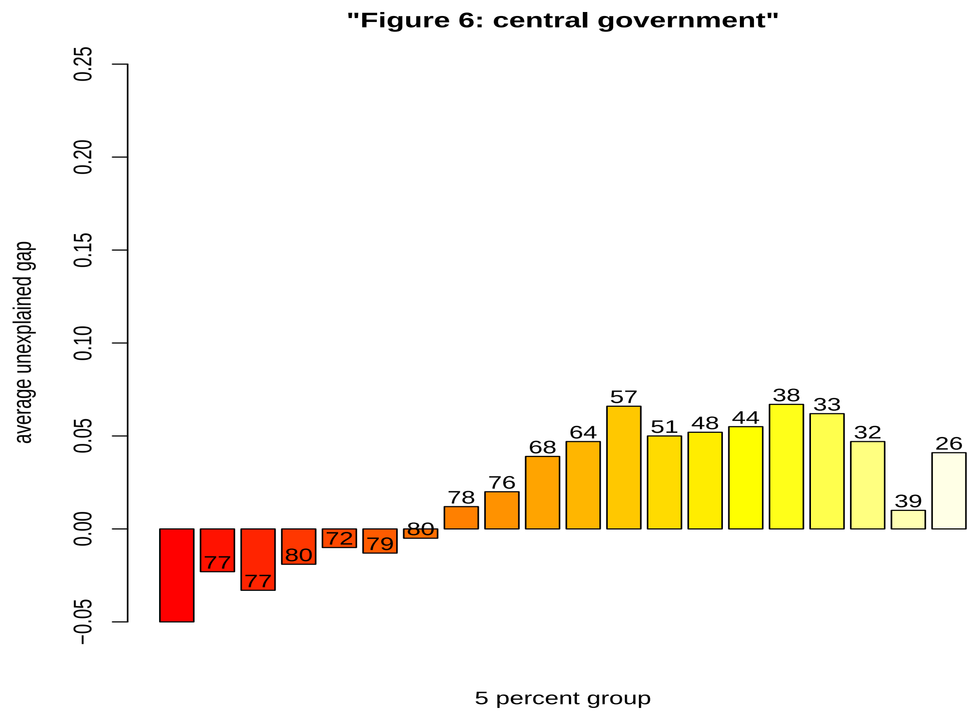


Table 10: Gross differential and explained part in three educational groups, 1998

	Low education	Medium education	High education
Gross differential	19.8	17.9	28.0
Explained contribution	9.8	10.4	20.6
Unexplained gap	10.0	7.5	7.4



5 Educational groups

Without reporting the full decomposition tables, we note the result that the gross wage gap was highest among employees with the highest educational qualifications. If we partition the sample into three educational groups ("low", "middle", "high") and run the decomposition for each group separately, the results depicted in table 10 emerge. The gross differential is far higher among the well-educated, but the unexplained gap is more or less the same.

6 A note on segregation

The above results suggest that differentiated assignment into occupations and industries is the most important single factor that sustains the gross gender wage differential. As a part of the project, we have also computed a number of dissimilarity indices that capture this phenomenon. We use the conventional Duncan dissimilarity index that is defined as follows. Suppose that the individuals of the sample are partitioned into I categories indexed by $i = 1, ..., I$. Denote by m_i the share of category i:s men out of all men; and by f_i the share of women of category i out of all women in the sample. The dissimilarity index D is defined

$$D = (1/2) \sum_{i=1}^{I} |m_i - n_i|. \tag{8}$$

Intuitively, the index tells the share of either sex that would have to change category if we wanted to generate a completely symmetric assigment of women and men into categories. By exploiting another data set, the Household Income Distribution Survey, we computed the Duncan index over a coarse occupational categorisation[6]. The results are reported in table 11 and they reveal, if anything, a slow decline in segregation. As to our main sample, we had at our disposal only the three years 1996, 1997 and 1998, so that no sharp conclusions on trends can be made. However, as the following table 12 shows, the segregation indices for these years are in decline as well, both what regards occupational as well as industrial assignment.

[6]"Pääammatti" in Finnish. Asymmetrical distribution over that categorisation also explains about half of the gender wage differential in that data set, but these computations are not reported, since they are based on a less representative sample

Table 11: Dissimilarity index over occupations, 1989-1998

year	D
89	.63
90	.62
91	.60
92	.60
93	.60
94	.61
95	.62
96	.59
97	.58
98	.59

Table 12: Dissimilarity index over occupations and industries, 1996-1998

year	D over occupations	D over industries
96	.650	.427
97	.646	.425
98	.636	.412

Finally, we might mention that a similar picture of a weakening segregation emerges if one carries out a similar computation for the five subsections of the sample; in that case the sector-specific fine occupational categorisation is substituted to the economywide categorisation used in table 12. A decline in the segregation measure emerges for all of our five subsectors.

7 Concluding remarks

We have shown that very little of the gender differential can be explained by using individual characteristics alone. Occupational categories are a far more important factor. In the public sector in particular, the commonplace assumption of age careers being disadvantageous to women turns out to be insufficient. The pure age factor is more important in manufacturing industries. As to the distribution of the unexplained gap, we found a concave relationship between the wage level and the gap; the gap is low at low wage levels and increases as expected income grows.

In addition to the analyses reported above, some other results were generated. By exploiting another data set (the Household Income Distribution Survey), we investigated whether the wage differential had changed over time during years 1989-1998. This was not the case, and these analyses are not reported. We also used both of our data sets to decompose the yearly changes in the gender wage differentials according to the theoretical decomposition exposed by Altonji and Blank (see Altonji and Blank (1999) and Juhn et al. (1991)). The year-to-year changes of the differential were very low, however, and so a decomposition of these changes turned out to be, unsurprisingly, a splitting of a small component into even tinier components. These methods are probably better suited to monitoring the wage differential over longer time spans like decades.

However, one positive development could be reported on the evolution of dissimilarity indices over time. Using both of our data sets, we computed the Duncan dissimilarity index over occupational classification and industrial classification. In all of these indicators, a slow but statistically significant trend towards less segeration emerges.

Hopefully, the methods and results of the project reported here can contribute to a more regular and systematic monitoring of the gender wage differential in future. Similar methods can also serve the purpose of comparative work and adoption of best practices within the European Union[7].

[7] Indeed, the Belgian presidency's proposition on Indicators in Gender Pay Equality follows a rather similar line of thought.

References

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*. Elsevier Science.

Jenkins, S. (1994). Earnings discrimination measurement: a distributional approach. *Journal of Econometrics*, 61(1):81–102.

Juhn, C., Murphy, K., and Pierce, B. (1991). Accounting for the slowdown in black-white wage convergence. In Kosters, M., editor, *Workers and their wages*. AEI Press, Washington DC, Washington DC.

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

Oaxaca, R. and Ransom, M. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61:5–21.

Oaxaca, R. and Ransom, M. (1999). Identification in detailed wage decompositions. *Review of Economics and Statistics*, 81(1):154–157.

Vartiainen, J. (2001). *Sukupuolten palkkaeron tilastointi ja analyysi*. Number 2001:17 in Tasa-arvojulkaisuja. Sosiaali- ja terveysministeriö.