Causal effects of occupational gender segregation and wages

Jenny Clarhäll*

Wednesday 30th March, 2011

Abstract

The labor market is, in a large extent, segregated by gender and occupations dominated by women pay less. This is the point of departure for this paper, where the relationship between segregation and wages are explored with Norwegian register data. Of special interest is whether a change in the share of women has an impact on average wage level in an occupation for both men and women. I find, as many before me, that there is a negative correlation between the share of women and wages. In addition to occupational fixed effects, I instrument the share of women in an occupation with the average share of women graduated from educations hold by employees in each occupation, 5 and 10 years ago. I find a significant negative result, where a 10 percent increase in the share of women in an occupation will reduce the wage in this occupation by 2,7 percent.

*Department of Economics, University of Oslo, Norway, jennymcl@econ.uio.no

1 Introduction

...considerable research suggests that predominantly female occupations pay less, even controlling for the measured personal characteristics of workers and a variety of characteristics of occupations, although the interpretation of such results remains in some depute. (Blau and Kahn, 2006)

The labour market is segregated by gender to a great extent, and this seems to be a world wide phenomenon - men and women tend to do different work both in the market and non-market. What kind of work that is understood to be "female" verses "male" may vary between countries and time (Goldin, 2002), but there is an observed tendency that female dominated occupations pay less (Macpherson and Hirsch, 1995; Blau and Kahn, 2006).

This paper will focus on the Norwegian labour market and the main concern is what impact a change in the gender composition in an occupation has on the wage. The hypothesis to be tested is that *the wage decreases or stagnates when the share of women increases.* The purpose, and main contribution, of this study is to examine whether the wage, for both men and women, changes when the share of women changes in an occupation.

In the public debate in Norway there is often argued that the lower pay in "female jobs" is a result of some kind of discrimination or comparable worth. For example a Norwegian newspaper (Dagsavisen) wrote in 2009 that at larger share of men employees in the kindergarten probably would increase both the status and the wage in the occupation. The year before, an official public report on gender pay gap (NOU, 2008) was published, and one of the conclusions was that a major part of the pay gap between men and women is due to the segregated labour market. What mechanisms, though, that explained the difference in pay due to segregation, remained a open question.

There are several papers that have investigated the relationship between wages and the gender composition in occupations. The typical approach is to estimate wage regressions specifying the relationship between wages and individual and/or occupational specific characteristics, and including a variable that accounts for the share of women. Fixed effects on individual and/or occupational level can be included, but the results provided are still consistent with both human-capital differences and labour-market discrimination

of some kind (comparable worth, monopsony). This paper goes through such a estimation, but in addition this paper is an attempt to provide an exogenous variation, which helps us to identify the mechanisms behind why female jobs pay less. Two instruments are introduced to instrument for the share of women in an occupation. The instruments are the average shares of women graduated from educations hold by employees in each occupation 5 and 10 years ago.

The population averages for the sample in this paper are significant and negative. Including a set of covariates, an employee in an occupation with 10 percent higher share of women than another, has a 1,5 percent lower wage. The model is further supplemented with occupational fixed effects and the instruments. I argue that the instruments, in combination with the within variation in occupations, manage to identify the effect of the share of women on wages. I find that a 10 percent increase in the share of women, reduces the wage by 2,7 percent. The effect for women and men respectively are almost the same; 2,5 percent for women and 2,7 percent for men.

The rest of the paper is organized as follows. Section 2 gives a brief picture of the labour market segregation and wages for men and women in Norway. In section 3 some possible theoretical explanations are presented and results from earlier studies on wage and occupation segregation are discussed. In section 4 the data used is presented. In section 5 the empirical strategy is introduced and the engogeniety problems and identifying strategies are discussed. In section 6 the results are presented and finaly section 7 providers a concluding discussion.

2 Background on norwegian gender pay gap

The fact that there is not as many women as men in the labor market, will in itself make a contribution to segregation. But the difference in labour force participation[1] between men an women has narrowed during the last decades. As figure 1 shows there were 69 percent of women and 77 percent of men made up for the labour force participation[2] in 2008 .

[1]OECDs definition of labor force participation: The labor force participation rate is defined as the ratio of the labor force to the working age population, expressed in percentages. See http://stats.oecd.org/glossary/detail.asp?ID=2008

[2]Statistics Norway "StatBank "The Labor Force Survey, Table 04858: *Employed persons per 4th quarter, by occupation and sex.*

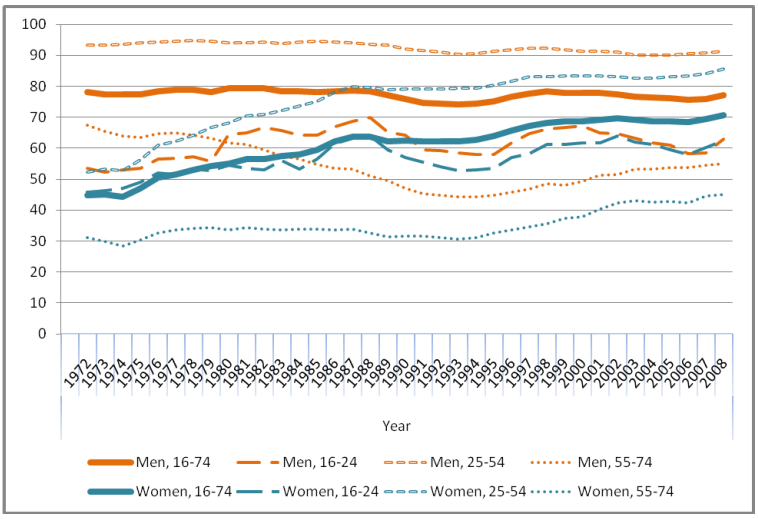

Figure 1: Labor force participation 1972-2008

A closer look *where* these men and women work, gives a picture of a segregated labor market. Figure 2 illustrates the 30 biggest occupations by the number of employees in Norway 2008[3]. These occupations occupies 59 percent of the working force in 2008. 23 of this occupations are dominated by a sex (less than 40 percent of one of the sexes). 87 percent of the employees works in segregated occupations. 46 percent of the male labor force, and 76 percent of the female labor force work in the 30 biggest occupations. Women have in a large extent entered occupations which traditionally where made by women, but as unpaid, work related to nursing and caring. The five largest occupation by employees were all female dominated with a women share varying between 67-91 percent. Though the male labor force is larger than the female labor force, a relative larger share of the female labor force is to be found among the largest occupations for the whole population. 46 percent of the male labor force, and 76 percent of the female labor force work in the 30 largest occupations by employees. It seems like men are in a larger extent more equally spread over a range of occupation, women seem to cluster to some large occupations.

[3]Statistics Norway "StatBank "The Labor Force Survey, Table 04858: *Employed persons per 4th quarter, by occupation and sex.*

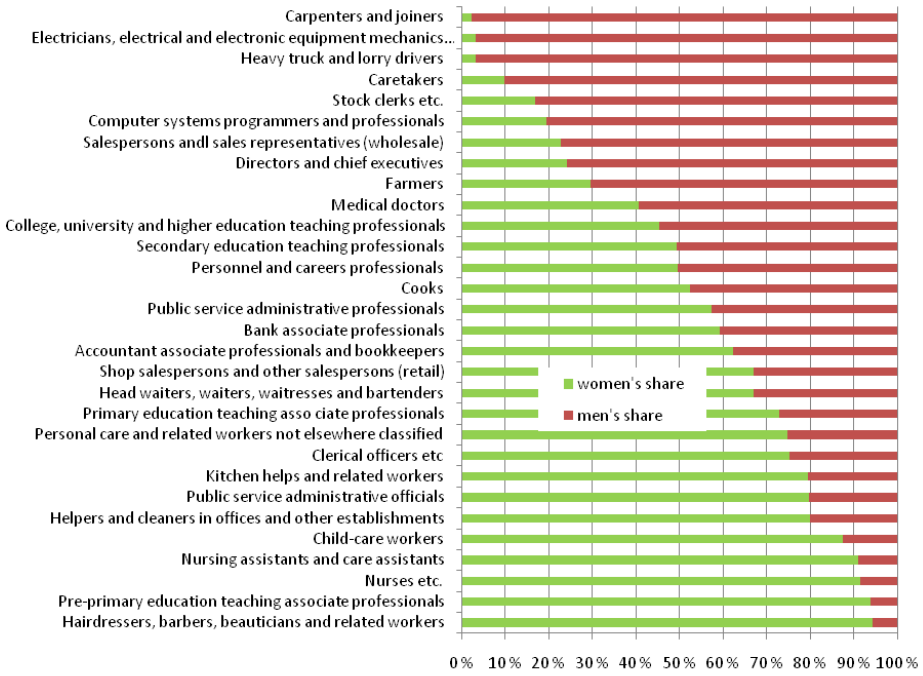

Figure 2: Occupational segregation, 2008

2.1 Decreased gender pay gap

A majority of the literature on wage inequality between men and women find that the gender wage gap has decreased the last decades (see for exemple Blau and Kahn (2006)). Barth and Dahle-Olsen (2004) find that that the wage gap between men and women in Norway has decreased since the early 1970s. In 1973 women's yearly wage was on average 45 percent lower then men's, controlled for education and experience but not for working time. In the mid 90s this difference was 30-35 percent. Today (2006) this difference is 35 percent (NOU, 2008). Barth and Dahle-Olsen (2004)n find small differences between men and women with same qualification and occupation. The same does (Nielsen, Høgsnes, and Petersen, 2003). The decrease in the gap is explained mainly by an increase in women's labour participation over the years, increase in women's worked hours per week and an increase in the hourly wage for women. Differences in hourly wage has decreased from 23 percent in 1973 to 15 percent in 2001 (Barth and Dahle-Olsen, 2004).

The 15 percent point difference in hourly wage between men and women seems to have been persistent during the decade of 2000. With use of an

other register, wage statistics, the gap was 16 percent in 1997 and 15 percent in 2008. Figure 3 compares the gender pay gap between 1997 and 2008 and by different age groups. Though the changes between 1997 and 2000 has been small, the age groups between 35-54 the gap has experienced the gap narrowing, with 2-4 percent. The figure shows that the women's wage, as a share of men's, increases (the gap increases) the older the age group is.

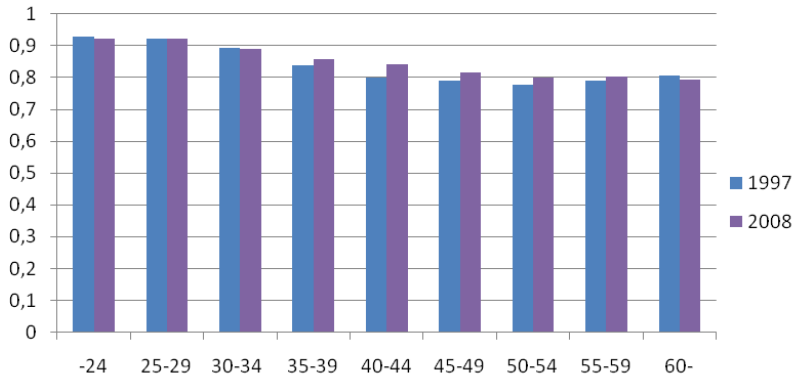

Figure 3: : Womens wage as share of mens wage, age groups. Average mounthly wage for both fulltime- and parttime workers, full-time equivalents. 1997 and 2008. Source: Statistics Norway

Barth and Schøne (2006) test two different explanations to why the gender pay gap is increasing by age: a) a career effect - hourly wages increases faster by age for men than for women, or b) a cohort effect - gender pay gap is larger for old men and women, compared with young men and women. The cohort-effect would be positive in gender equality terms because it would imply a closing of the gender pay gap with younger generations. But they find that the main explanation is different wage growth for men and women during their career.

2.2 Decreased occupational gender segregation?

It also looks like individuals in the labour market faces less occupational gender segregation today (2006) than in 1997. Figure 4 gives a picture of deciles over decomposition of share of women for individuals. The lowest deciles, those individuals who face a low share of women at there workplace (or occupation), have a higher share of women in 2006 than in 1997. The

opposite for those in the higher deciles - individuals who face a high share of women at there workplace (or occupation), have a lower share of women in 2006 than in 1997.

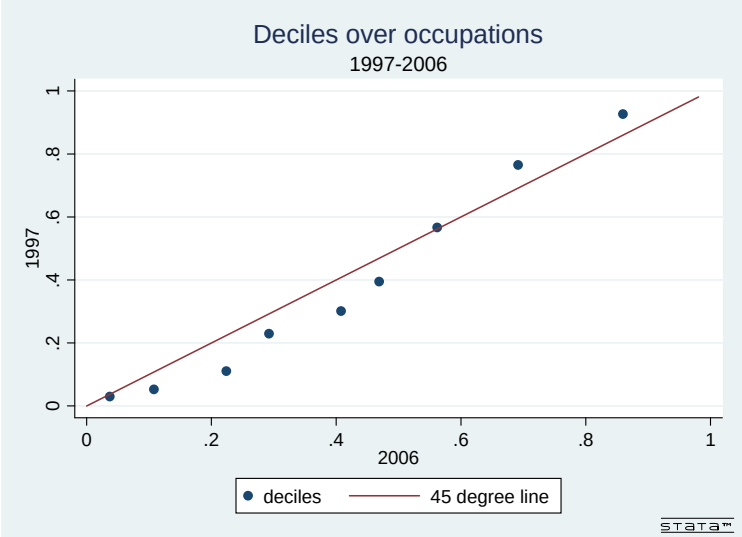

Figure 4: : Change in the composition of share of women in occupation. Deciles over individuals 1997 and 2006 Norway

The same pattern is observed for occupations: occupations with a low share of women, have a higher share of women in 2006 compared with 1997, and occupations with a high share of women, have a lower share of women in 2006 compared with 1997.

3 Theoretical and empirical background

Though the fact that occupations dominated by women pay less in comparison with occupations dominated with men, is little disputed, the interpretation and magnitude of the relationship between occupational gender segregation and wages are disputed.

In general the gender pay gap can be discussed in a human capital - discrimination framework, but economic theory gets a hard time explaining why the share of women should have an impact on wages.

A review of the previous literature on the women's share effect on wages is useful. The framework of human capital, which is used to explain other parts of the gender pay gap (see(Becker, 1985; **?**) , can also be used in a

share-wage discussion. This research focuses is on gender-specific factors which include differences in qualification, preferences and labour market treatment of similar qualified individuals(Blau and Kahn, 2000). If men and women differs in respect of preferences regarding jobs that require on the job-training, differences in occupations would be expected (Blau and Kahn, 2000). Men could in a larger extent than women be motivated by large investment and specialization, which one expects would increase the wage. But at the same time that can have an opposite effect; *employers may be reluctant to hire women for such jobs because the firm bears some of the costs of such firm-specific training and fears not getting a full return on that investment.*(Blau and Kahn, 2000).

The other set of explanation why womens jobs pay less is related to some kind of discrimination. Sorensen (1990) analysis the occupational gender pay gap within a comparable worth framework where the crowding hypothesis (**?**) is used as theoretical framing. In many countries policies with the target to equalize comparable worth, are proposed or introduced (Macpherson and Hirsch, 1995). As an example the public official report on gender pay gap (NOU, 2008) proposes that there should be a increase in the wage level for female dominated jobs that are payed relatively low. This proposal can have its origin in a comparable worth framework or other discriminatory explanation for why female jobs pay less.

The arguments behind the comparable worth approach identifies certain jobs as *womens work* and the lower wage in these occupations are then explained by an undervaluation of female work. How these jobs become *gendered* is not clear. Sorensen used the crowding hypothesis to explain the argument of comparable worth. According to the crowding model, women may be crowded into particular occupations, due to either preferences or to past or present discriminatory barriers (Macpherson and Hirsch, 1995). The effect of this crowding of women to a smaller set of occupations is that the supply in these occupations are high and wages decreases. Sorensen (1990) reviews empirical studies where the share of women (gender composition in an occupation) is included as an explanatory variable for wages. Sorensen finds that the magnitude of the impact of crowding varies between 0 and 42 percent. Sorensen's own estimates, based on USA 1984 PSID and May/June1983 CPS, include detailed industry controls and she concludes that within her design 20-23 percent of the gender pay gap can be explained

by crowding.

Macpherson and Hirsch (1995) uses fixed effects on individuals and a large range of job-characteristic covariates on USA CPS data between 1983-1992. They find that the remaining gap due to the relation share-wage are small, and explains the relative small remaining effect as a result of differences in unobserved human capital between men and women. They find that the estimate of the share of women's on wages decreases from 0,163 to 0,055 for women and from 0,178 to 0,034 for men. Though all these studies have the purpose to explain the impact of the share of women on wages, none of them take in use exogenous variation. The inclusion of fixed effects on the individual or occupational level helps tracing different explanations, but we remain in a dispute over in what extent the remaining gap, due to the relation between share of women and wages, are due to human capital (which is not revealed) or discrimination. As far as I know there exists no empirical study where the share of women explains wages

4 Data

The main data source for this study is the Norwegian wage statistics, a register available from 1997 to 2006. This unbalanced dataset covers all employees in public sectors[4] and a 50 per cent survey sample of employees in the private sector. The dataset includes weights, which are denoted the inverse of the probability to each business for each sector. The weights are used in the regressions to make the two sectors as comparable as possible. The dataset is limited in respect to age, employees between 25 and 66 in age are included. The background for this limitation is that the majority is assumed to have finished school at the age of 25, and the year of retirement is assumed to be the year one turn 67 years.

The dependent variable, logs of hourly wages, include basic salaries, irregular payments and bonuses. Payments in kind, insurance, pension schemes, tax-free expenses arrangements etc. are not included. Neither is overtime payment included in the statistics dependent variable, even though it is available in the dataset, because overtime payments are payments acquired beyond the regular working time. The dataset has an individual

[4] A total count of all employees of the central government, municipalities and publicly maintained schools and hospitals. Statistics Norway

Table 1: Descritive statitcs

	ALL		WOMEN		MEN	
	Mean	SD	Mean	SD	Mean	SD
Share of women	0.48	(0.50)				
Age	42.9	(10.5)	43.2	(10.5)	42.6	(10.6)
Work experience	20.2	(9.68)	18.8	(8.74)	21.6	(10.3)
Mounthly wage	23680.2	(12226.1)	19132.7	(9233.5)	27917.0	(13118.4)
Hourly wage	162.9	(75.5)	144.5	(49.7)	180.1	(90.0)
Logs of hourly wage	5.03	(0.34)	4.93	(0.28)	5.12	(0.36)
Number of hours worked per months	143.6	(35.1)	130.7	(40.1)	155.6	(24.3)
Fulltime=1, Part-time=0	0.74	(0.44)	0.53	(0.50)	0.93	(0.26)
Sector, public=1 private=0	0.34	(0.47)	0.51	(0.50)	0.19	(0.39)
andel			0.69	(0.24)	0.28	(0.26)
N	8671027		4793569		3877458	

code for occupation classification, but the coding is different depending on the sector - public sector has one way to code occupation and private sector another. So what looks like two different occupations, by code, can be the same occupation, only registered different for public and private sector. Occupations with less then 100 employees per year are excluded and the remaining dataset contains in total 173 different occupations.

The variable *Share* (nr of women/nr of men + nr of women) is constructed by taking the share of women in each occupation for each year. Private/public sector, full time/part time and age are covariates included, with its origin in the wage statistics. The quadratic of age also included to control for a possible non-linear relation between wage and age.

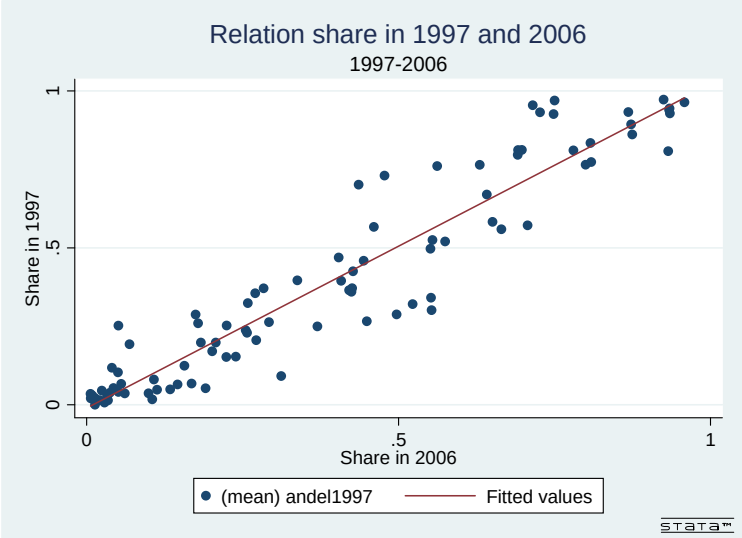

Figure 5: :Relation between the share of women in 1997 and 2006 for each occupation

Different covariates on education, experience and labour market adjustment are included. A variable *experience* is constructed with use of a second register, *the revenue statistics*, where different kinds of revenue, all accounted for pension, are registered. With help of this dataset a variable over number of years with revenue over a certain cutoff is constructed (approx 70 000 NOK in 2008). This dataset allows for a richer range of experience than the wage variable in the wage statistics. The quadratic of experience also included to control for a possible non-linear relation between wage and experience.

The education covariates are constructed from a total count register of the the highest obtained level of education per year for the whole Norwegian population, *the education statistics*, available from 1970 to 2007. This register include education at all levels, from kindergarten to university and university college. For the purpose of this study levels and field of education[5] are used as covariates[6]. The level of education are five dummies: primary and secondary school, high school, low university grade and high university grade. Schooling is assumed not time varying in this setup because of the relatively small panel with people between the age 25-66 years, where most people have finished education at the age of 25.

This register is also used to create the instrument. The highest obtained education for an individual at a specific year in the *wage statistics* is the starting point for the instrument. For each education code the share of women of all who graduated 5 and 10 years ago from this education (and received this code in the dataset) are constructed. For each occupation each year the averages of these variables are used to construct the instruments $Z5$ and $Z10$. The instruments $Z5$ and $Z10$ varies between 0-1.

4.1 Identification

A crucial condition for enable to identify any causal effects, is that there are changes in employees job-attachment - employees need to change jobs. In this dataset there are 1 171 066 observed changes of jobs during the time-period (out of 8 671 027 observation) . Most of the employees changing jobs (a total of 739 453 individuals), changed once, as seen in figure 6.

[5]The fields of education are dummies over: General programmes, Education, Humanities and arts, Social sciences, business and law, Science, mathematics and computing, Engineering, manufacturing and construction, Agriculture and veterinary, Health and welfare, Services, and Unknown

[6]The choice of categories, both on fields and level, are in accordance with Norway Statsitics

Job-change	Freq.	Procent	Cum.
1	445 076	60.19	60.19
2	195 520	26.44	86.63
3	69 325	9.38	96.01
4	22 365	3.02	99.03
5	5 731	0.78	99.81
6	1 223	0.17	99.97
7	183	0.02	100
8	29	0	100
9	1	0	100

Figure 6: Number of persons changing jobs and how many time they changed jobs during the period

Figure 2 describes the group who has changed job at least once during the period. Compared with the hole sample population this group (on average) has a smaller woman/man ratio, lower age and less job-experience, and tend to work in more in the private sector. But the differences are small.

Table 2: Descritive statitcs for job-changers

	ALL		WOMEN		MEN	
	Mean	SD	Mean	SD	Mean	SD
Share of women in	0.43	(0.49)				
Age	40.4	(9.95)	40.1	(9.78)	40.6	(10.1)
Work experience	18.8	(9.38)	17.1	(8.46)	20.1	(9.83)
Mounthly wage	23741.1	(11868.5)	19473.1	(9427.2)	26918.6	(12486.4)
Hourly wage	159.4	(71.0)	141.7	(50.2)	172.6	(80.6)
Logs of hourly wage	5.01	(0.35)	4.91	(0.30)	5.08	(0.36)
Number of hours worked per months	146.9	(33.7)	135.2	(40.3)	155.6	(24.4)
Fulltime=1, Part-time=0	0.80	(0.40)	0.62	(0.48)	0.93	(0.26)
Sector, public=1 private=0	0.21	(0.41)	0.34	(0.47)	0.12	(0.32)
N	1171066		559941		611125	

Clarifying that changes have been made, one can go further with the analysis. The correlation between the instruments and the variable to be instrumented are of importance. Figure 7, 8, 9 and 10 is graphs over the ten occupations with largest increase, and decrease, in the gender composition for private and public sector between 1997 and 2006. It also graphs the development over the instrument for the same occupations.

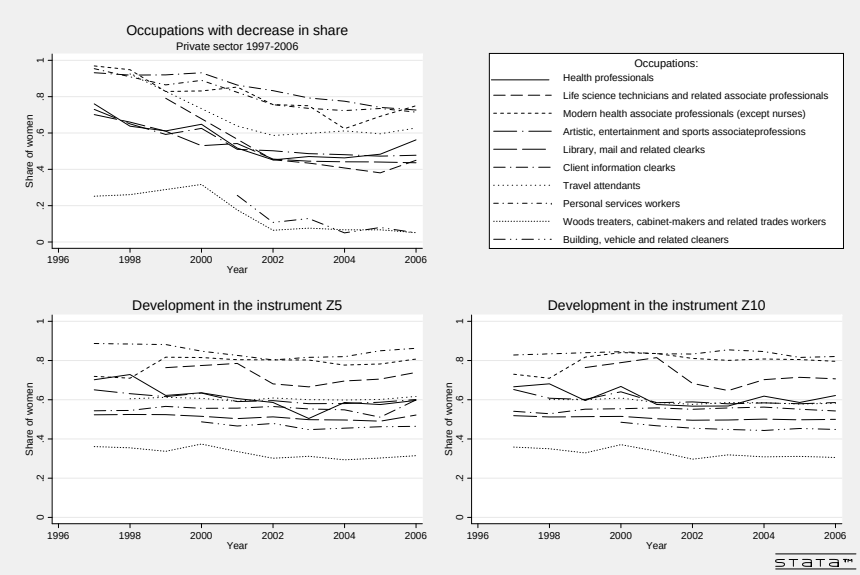

Figure 7: The ten occupations with the largest decrease in "Share" in private sector 1997-2006

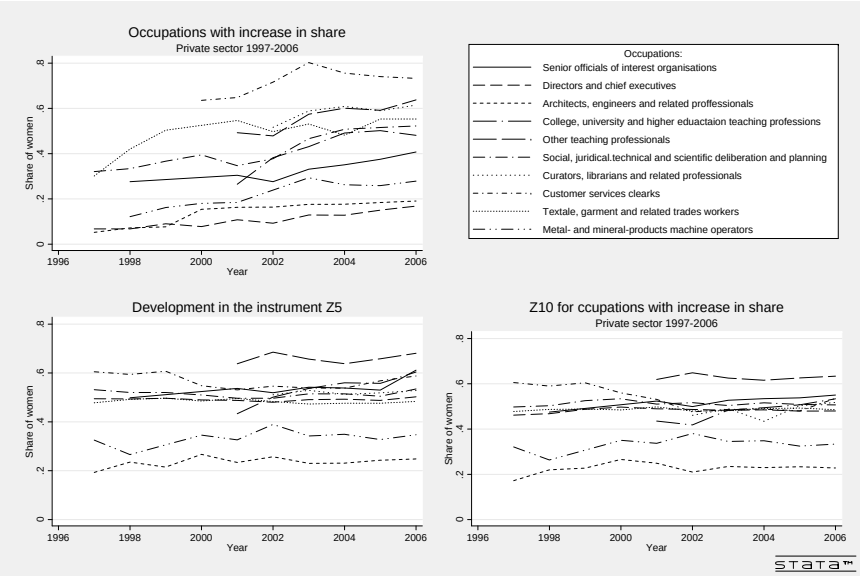

Figure 8: The ten occupations with the largest increase in "Share" in private sector 1997-2006

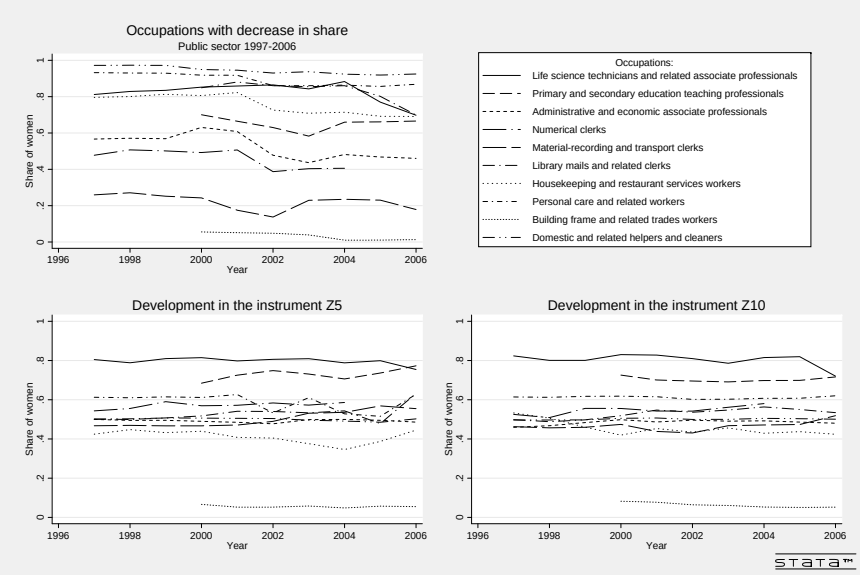

Figure 9: The ten occupations with the largest decrease in "Share" in public sector 1997-2006

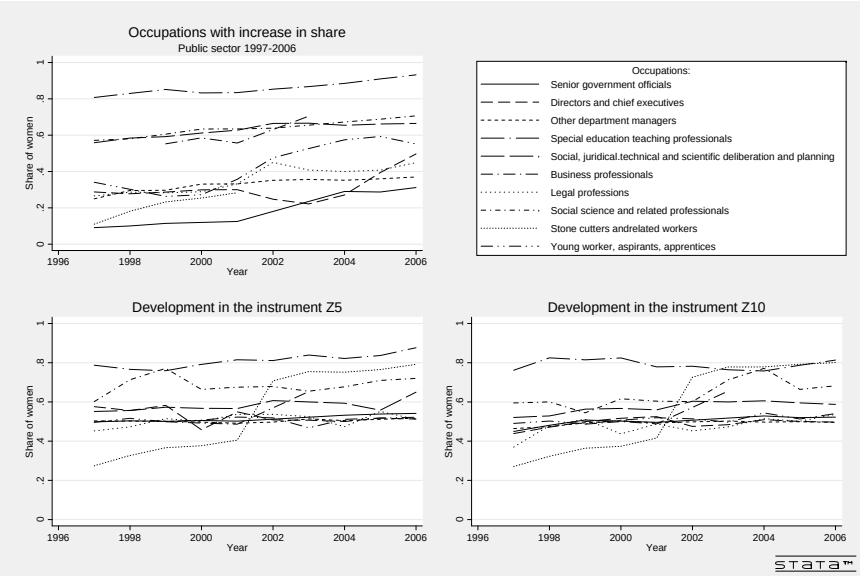

Figure 10: The ten occupations with the largest increase in "Share" in public sector 1997-2006

There has been some compositional changes in the occupations regarded gender, and the change seems to have been larger in private sector than in public sector. Out of the graph it looks like the share of women and the

instruments have followed the same path.

For the rest of this paper a 50 percent sample is used from this dataset due to software capacity and time-demanding operation with the whole dataset.

5 Empirical strategy

The relationship between wages and gender composition in occupations can be specified as following:

$$lnW_i = \beta K_y + \delta X_{iy} + u_i \tag{1}$$

,where lnW is logs of hourly wage, K is the share of women (varying between 0-1) in the occupations and X is different covariates mearsuring different individual and/or occupation characteristics. The error term u_i is for now assumed to $E(u_i) = 0$ and $Var(u_i) = \sigma$, namely zero in mean and a constant variance. A negative value of β implies a decrease in wage when the share of women increases. This is well documented (see for example Macpherson and Hirsch (1995); Barth, Røed, and Schøne (2005)), and also a result for the sample in this paper, see table 3. The table pictures a cross-sectional development over the relationship between wages and segregation over the years included in this paper, 1997-2006. The population-average coefficient has been relatively stable the last years, where individuals in occupations with 10 percent larger share of women than others have 1,6-1,7 percent lower hourly wage.

There are two serious problems or weaknesses concerning endogeniety in the population-average coefficient. First there is a possibility that variables that should have been included, are not. In other words, X doesn't include all necessary variables. This can both be unmeasured variables and/or variables that is not thought of, so called missing variables. In a fixed effect model, some of these unobserved variables can be controlled for. Individual fixed effects should control for different characteristics related to the individual (employee) that are constant over time. These are for example more obvious variables as gender, but also variables like skill and taste differences of the workers that can be considered time-invariant. If a characteristic that is relatively constant over time (little within variation) is included as a re-

Table 3: Populataion averages over occupation segregation and wages

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Share	-0.1051**	-0.1189**	-0.1391**	-0.1171**	-0.1609***	-0.1693***	-0.1478***	-0.1635***	-0.1716***	-0.1649***
	(-2.01)	(-2.22)	(-2.23)	(-2.40)	(-3.58)	(-3.26)	(-2.91)	(-3.17)	(-3.15)	(-3.04)
N	252683	262257	254138	302968	353673	401889	428726	485311	496812	483530

gressor in a fixed effect model, it will be imprecisely estimated (Cameron and Trivedi, 2009). In the absence of individual fixed effect, it is not possible to reject the hypothesis that the proportion of women is a proxy for unmeasured skill and taste differences among workers or of occupational attributes correlated with wage (Macpherson and Hirsch, 1995).

The second problem is concerned with the statistical simultaneous problem. At the same time as the ratio women/men can have an impact on wages, wages can have an impact on the ratio. Hypothetically men may, in a larger extent then women, be more motivated by wage, which also will result in occupations dominated by men, and a higher wage-level in these occupations. Wage and the share of women are determined by the interaction between these two equations. As an attempt to deal with both of these problems (omitted variables and simultaneous equation problem) fixed effects on both individual and occupation level and a instrument is included in this paper.

5.1 Econometric specification

To include the fixed effects in model 1, the specification is redefined as:

$$lnW_{iyt} = \alpha_i + \theta_y + \beta K_{yt} + \delta X_{iyt} + u_{iyt} \tag{2}$$

, where α_i and θ_y denotes individual and occupational fixed effects, respectively. These are included to control for unobserved heterogeneity both at the individual level and the occupational level. The u's are disturbance terms in the model. In section 6 Results, different regressions are estimated, and in this section a motivation is given for the regressions ran in section 6. The two first regressions are population averages, where the first, OLS1, treats $\alpha_i, \theta_y, X_{iyt} = 0$. The second regression run, OLS2, treats $\alpha_i = 0$ and $\theta_y = 0$. The individual fixed effect, run as the third regression, FE1, treats $\theta_y = 0$, and for the last regression $\alpha_i = 0$.

In a fixed effect model, the individual-specific effects, α_i, are permitted to be correlated with the regressors. The error term can be specified as $\eta_{iyt} = \alpha_i + u_{iyt}$ and it is then the time-invariant component, α_i is that that is allowed to correlate with the regressors, not the idiosyncratic error u_{iyt}. As an example, the unmeasured skill-components that are time-invariant

are allowed to correlate with the regressor, but not the unmeasured skill-component that varies over time. The fixed effect model though allows for a limited form of endogeneity.

Occupational fixed effects also allows for a limited form of endogeneity connected to occupations. This refers to the unmeasured time-invariant component of the error term. Occupational FE control for unobserved differences over time that is not time-varying, the wage-level for example. Unmeasured time-varying differences between occupations are still not allowed to be correlated.

So, the fixed effect models cope with parts of the endogeniety problem, but not all. As an attempt to overcome parts of the endogeniety that remains, two iinstruments are introduced. The instruments should be highly correlated with the variable instrumented for, but not with the error term.

5.1.1 The instrument

As an attempt to solve both problems with omitted variables and the statistical simultaneous equation, the instruments Z5 and $Z10$ are introduced in the model with occupation fixed effects. The instruments used are the average share of women graduated from educations hold by employees in each occupation, 5 and 10 years ago. Within the framework of occupation fixed effects, the effects of a change in the share on wages are identified *within* each occupation. The last regression from above is run with $Share$ instrumented by $Z5$ and $Z10$ in a 2SLS model. Accordingly the first stage for the regression is:

$$\mathrm{K}_{yt} = \theta_y + \varphi\ \mathrm{Z5}_{yt-5} + \alpha \mathrm{Z10}_{yt-10} + \lambda \mathrm{X}_{iyt} + u_{iyt} \tag{3}$$

The reduced form equation is:

$$ln\mathrm{W}_{iyt} = \theta_y + \eta \mathrm{K}_{yt} + \delta \mathrm{X}_{iyt} + \xi_{iyt} \tag{4}$$

There are two important requirements for instrument variables. These are that the instrument Z must be correlated with the endogenous explanatory variable, and not correlated with the error term in the reduced form equation, conditional on the other covariates.This instruments is assumed to have an impact on wages through the share of women in occupations,

but not on other determents of the wages. If the share of women in an education increases it is expected to have an impact on the share of women in the occupation some years later. As a concrete example, an increase in the share of women among medical students, should not in itself have an impact on the wage for doctors 5 years later. It can have an impact on the wages for doctors, but only thorough the variable instrumented.

Within the framework of occupation fixed effects, the effects of a change in the share on wages are identified *within* each occupation. The possibility that the error component, ξ_{iyt}, are somehow correlated with the instruments are small. As an attempt to minimize the possible correlation between error term and instrument, weights are constructed for the instruments. The instruments content two components. The first is the share of women graduated in each education 5 (or 10) years ago. The other is the share of employees within occupation y with education j per year. Changes in the latter part can be due to changes in requirements for entering occupations, and consequently a source of correlation between instrument and the error term. To deal with that problem an average of the share of employees within occupation y with education j is taken over the years T, instead of for each year t:

$$Z5_{yt} = \left(\frac{1}{T} \sum \frac{n_{yjt}}{N_{yt}} \right) \cdot K_{jt-5} \tag{5}$$

One could think of a situation where graduates (note: this is not the year they choose education, but the year they graduate the education they chose 5 or 10 years ago) would predict the development of wage within an occupation. There can be situations where some kind of foresight may be plausible. If there are structures in the society that in a wider timespan is expected to have an impact on the development of the wage-structure in a specific occupation, this can be known for the people making choices about their own education. For example, there is a concern whether there will be a shortage of employees in health- and caring related occupations for the coming generations, and this concern is discussed in the news papers. The fertility rates have been low for some years, and the number of "seniors" verses "juniors" are increasing. This public discussion concerns a plausible increase in the demand of for health- and caring related occupations. If this increasing demand is associated with an expectation of higher wages in this

occupations, this could have an impact on the supply side of employees in the health sector through the education some years in advance. An increase in supply can have the adverse impact on wages than an increase in demand would have had. Resulting in a larger labour force in health- and caring related occupations, but no relative increase in wages. The foresight include both assumptions about the increase in demand and supply in the future. This is quite difficult to foresight, I would argue. The choice concerning education may rely on different components as interests, the level of wages in different occupation. But assuming that the *development in wages* within each occupation is a component, is little plausible.

6 Results

The results for the estimations without the instrument are presented in table 4. The populations averages with and without covariates (OLS1 and OLS2), give a relatively high and significant negative result. The covariates explain part of the difference, but for the hole sample, a employee in an occupation with 10 percent larger share of women than an other occupation, pays 1,5 percent less. The difference in coefficients between OLS1 and OSL2 suggests that human capital differences, worker preferences regarding job attachment, counts for part of the wage-share relationship (Macpherson and Hirsch, 1995).

The effects for women are larger and more robust compared with the result for men for the two first regression (OLS1 and OLS2). When introducing individual fixed effects the difference disappear, and for a employee (man or woman) who experience a change in the share of women in his or hers occupation (she or he changes jobs or the share of women in his or hers occupation increases) by 10 percent will experience a 0,7 percent decrease in wage. This result suggests that time-invariant individual characteristics, like ability and taste, may account for parts of the negative relationship between wages and the share of women. For the final model, the coefficient doesn't change much, but the results are no longer significant at any level required in this paper[7]. This model uses the variation within each occupation, and consequently the results of the model doesn't find any support for effects on

[7]Statistically significant levels *** 0.01 level (1 percent level), ** 0.05 level (5 percent level), * 0.1 level (10 percent level).

Table 4: Logs of hourly wage - clustered on occupations x years

	OLS	OLS2			FE1			FE2		
	ALL	ALL	WOMEN	MEN	ALL	WOMEN	MEN	ALL	WOMEN	MEN
Share		-0.1465***	-0.2486***	-0.0994***	-0.0614***	-0.0733***	-0.0559***	-0.0500	-0.0799**	-0.0741**
		(0.0173)	(0.0163)	(0.0218)	(0.0048)	(0.0061)	(0.0050)	(0.0367)	(0.0407)	(0.0323)
Woman	-0.1115***	-0.0795***						-0.0845***		
	(0.0052)	(0.0037)						(0.0035)		
Age	0.0161***	0.0166***	0.0106***	0.0151***	0.0189***	0.0167***	0.0115***	0.0113***	0.0088***	0.0105***
	(0.0012)	(0.0011)	(0.0008)	(0.0015)	(0.0013)	(0.0011)	(0.0017)	(0.0007)	(0.0006)	(0.0010)
Age2	-0.0002***	-0.0002***	-0.0001***	-0.0001***	-0.0002***	-0.0002***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experience	0.0151***	0.0145***	0.0114***	0.0199***	0.0162***	0.0128***	0.0230***	0.0119***	0.0095***	0.0156***
	(0.0005)	(0.0005)	(0.0004)	(0.0006)	(0.0014)	(0.0016)	(0.0017)	(0.0003)	(0.0003)	(0.0004)
Experience2	-0.0002***	-0.0002***	-0.0001***	-0.0004***	-0.0002***	-0.0001***	-0.0003***	-0.0002***	-0.0001***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Sector, public=1	-0.0924***	-0.0584***	0.0035	-0.1361***	-0.0267***	-0.0008	-0.0614***			
	(0.0091)	(0.0094)	(0.0089)	(0.0097)	(0.0061)	(0.0059)	(0.0061)			
Fulltime=1	0.0633***	0.0491***	0.0332***	0.1094***	0.0015	-0.0039	0.0174***	0.0250***	0.0167***	0.0571***
	(0.0067)	(0.0064)	(0.0057)	(0.0070)	(0.0034)	(0.0033)	(0.0044)	(0.0045)	(0.0046)	(0.0057)
Timedummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Occupation FE	no	no	no	no	no	no	no	yes	yes	yes
Individual FE	no	no	no	no	yes	yes	yes no	no	no	
Education controlls	yes	yes	yes	yes	no	no	no	yes	yes	yes
N	4324007	4215612	2329924	1885688	4215612	2329924	1885688	4215612	2329924	1885688
r2	0.4323	0.4406	0.4750	0.3749	0.4299	0.4469	0.4193	0.2826	0.3097	0.2467
F	239.3368	248.2278	275.1071	236.7079	1330.2740	1128.1988	1454.2799	647.3668	488.1576	621.9970
pvalue										

wages by the share of women.

The different covariates remain stable and robust, independent of the different regressions. The variable *Woman*, not included in FE1 because it doesn't vary over time, remains rather constant, interpreted as there remains a 8 percent difference in hourly wage between men and women which this model can not explain. All standard errors, both for the results described above and the results for the 2SLS, have been clustered on occupations.

To examine whether there could be a non-linear relationship between the share of women and wages, the share of women are divided into deciles and the same models as above are run. Though this study concerns average effects, there is of interest to examine whether the effect differs with respect to how the occupation is segregated. The results are potted in figure 11.

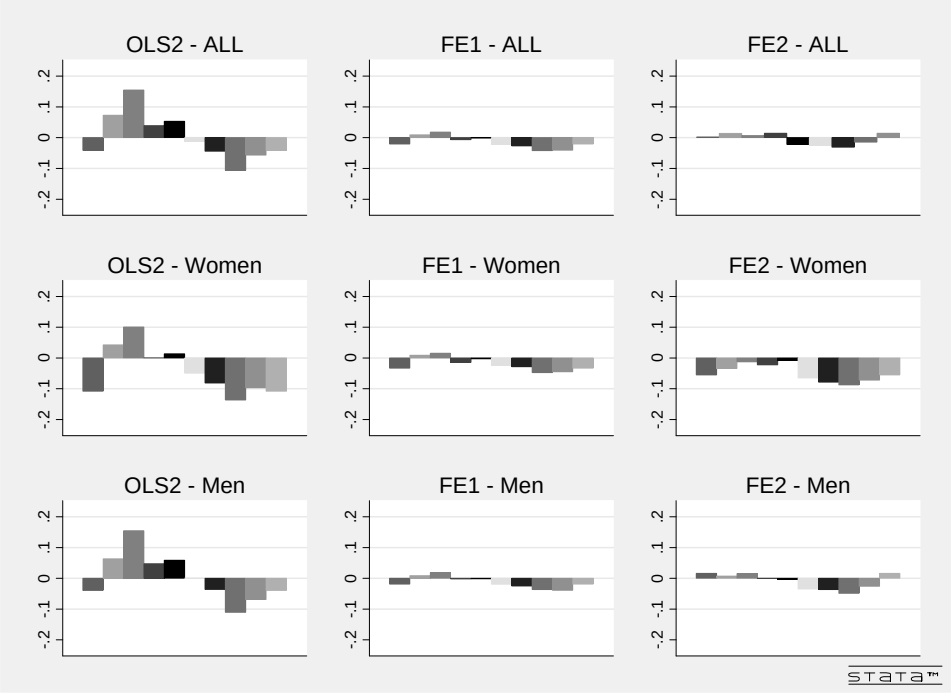

Figure 11

The results for the different deciles are in relation to the first decile (though the first decile also is included). For the OLS-regression there is a nonlinear relationship between the share of women and wages. For the other two models (FE1 and FE2) the effect is smaller, but with a common

pattern where the negative relation is stronger for higher deciles. Individuals who work in an occupation where the share of women is relatively high, or occupations where the share of women are relatively high, experience a larger effect of a change in the share of women on wages.

6.1 The instrument

The results for this estimations with the instruments are reported in table 5. The correlation between the instruments and the instrumented variable are high, 0.76 between Z5 and the share of women and 0,79 between $Z10$ and the share of women. Whether the instrument is correlated with the error term is not yet tested. The results for the first stage estimations are reported in figure 12. The results are significant at a level of 1 percent. For the instrument Z5 a 10 percent increase in the share of women graduated 5 years ago in educations hold by employees in the occupation, will increase the share of women in the occupation by 2 percent. The instrument $Z10$ has a larger impact. A 10 percent increase in share of women graduating 10 years ago, will have contributed to a increase in the share of women in this occupation with 4 percent today. The fact that $Z10$ has a larger impact than Z5 on the variable instrumented is understandable. A change in the share of women graduated 10 years ago, compared to 5 years ago, have had 5 years more to attain a persistent effect on the share of women in the occupation. The results do not vary much between the whole sample, women and men.

	First stage		
Instrument	All	Women	Men
Z5	0.20	0.18	0.20
Z10	0.41	0.42	0.47

Figure 12: The coefficients of the first stage regression. All significant at a level of 0.01 percent. F-statistics large.

The reduced form results, see table 5, suggests a negative, stronger and robust result than the same model without instruments. For the whole population a 10 percent increase in the share of women within an occupation will reduce the average wage in the occupation by 2,7 percent. The effect is 2,4 for women and 2,6 for men, within occupations. As in the models without the instruments, the variable $Woman$, remains constant and significant at a 8 percent level. So, despite inclusion of different controls and

Table 5: Logs of hourly wage - SHARE instrumented

	ALL	WOMEN	MEN
Share	-0.2754***	-0.2481***	-0.2661***
	(0.0111)	(0.0135)	(0.0157)
Woman	-0.0835***		
	(0.0003)		
Age	0.0114***	0.0089***	0.0105***
	(0.0001)	(0.0001)	(0.0002)
Age2	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Experience	0.0118***	0.0095***	0.0156***
	(0.0001)	(0.0001)	(0.0001)
Experience2	-0.0002***	-0.0001***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)
Workingtime, fulltime=1	0.0247***	0.0165***	0.0570***
	(0.0003)	(0.0003)	(0.0009)
Timedummies	yes	yes	yes
Education controlls	yes	yes	yes
Occupational FE	yes	yes	yes
Craig Donald	385.9	163.3	267.1
Hansen J	102.5	502.8	114.8
Hausman	74.3	24.2	33.9
N	4215612	2329924	1885688
r2	0.2820	0.3092	0.2463
F	48409.7995	33402.7072	20217.0321

instruments, there remains a 8 percent difference in hourly wage between men and women which this model can not explain. Nielsen, Høgsnes, and Petersen (2003) finds that the position you possesses in an occupations or a workplace explains 5 percent of the remaining pay gap between men and women. This may be an explanation for parts of the remaining gap. Unfortunately the datasets used in this paper doesn't include any variables on position in occupations

7 Conclusion and discussion

References

Barth, E., and H. Dahle-Olsen (2004): "Lønnsforskjellene mellom kvinner og menn i et 30 års perspektiv," *Søkelys på arbeidsmarkedet*, 2004:1.

Barth, E., M. Røed, and P. Schøne (2005): "Lønnsforskjeller mellom kvinner og menn i privat sektor: betydning av yrke og virksomhet," *Søkelys på arbeidsmarkedet*, 2, 211–213.

Barth, E., and P. Schøne (2006): "Lønnsforskjeller mellom kvinner og menn over livsløpet. Kohort eller karriere?," *Søkelys på arbeidsmarkedet*, 2006:2(2006:2).

Becker, G. S. (1985): "Human Capital, Effort, and the Sexual Division of Labor," *Journal of Labor Economics*, 3(1), S33–S58.

Blau, F. D., and L. M. Kahn (2000): "Gender Differences in Pay," *The Journal of Economic Perspectives*, 14(4), pp. 75–99.

——— (2006): "The Gender Pay Gap: Going, Going. . . but Not Gone," in *The Declining Significance of Gender?*, ed. by M. C. B. Francine D. Blau, and D. B. Grusky, pp. 37–66. Russell Sage Foundation, New York.

Cameron, A. C., and P. K. Trivedi (2009): *Microeconometrics using Stata*. Stata Press.

Goldin, C. (2002): "The Rising (and then Declining) Significance of Gender," NBER Working Papers 8915, National Bureau of Economic Research, Inc.

Macpherson, D. A., and B. T. Hirsch (1995): "Wages and Gender Composition: Why do Women's Jobs Pay Less?," *Journal of Labor Economics*, 13(3), pp. 426–471.

Nielsen, R. A., G. Høgsnes, and T. Petersen (2003): "Lønnsgapet mellom kvinner og menn - en oppdatering," *Søkelys på arbeidsmarkedet*, (2-2003), 215–222.

NOU (2008): "Kjønn og lønn. Fakta, analyser och virkemidler for likelønn," Norges offentlige utredninger.

Sorensen, E. (1990): "The Crowding Hypothesis and Comparable Worth," *The Journal of Human Resources*, 25(1), pp. 55–89.