Is it what you do or where you work that matters most? Gender composition and the gender wage gap revisited

Mahmood Arai*, Lena Nekby† and Peter Skogman Thoursie‡ §

30 June, 2004

Abstract

The purpose of this study is to examine the impact of gender segregation on wages using matched employer-employee private-sector data for Sweden. The questions that we are interested in examining are two-fold. Has the effect of gender segregation on the gender wage gap been overestimated and what matters more for gender wage differentials, occupation or establishment segregation? Our results show that a too detailed aggregation of occupations and/or establishments leads to an overestimation of the segregation effect on gender wage differences. We also show that occupational segregation contributes more to explaining the wage gap than establishment segregation.

Keywords: gender wage gap, matched employer-employee data.
JEL: J16; J71

*E-mail: mahmood.arai@ne.su.se

†E-mail: lena.nekby@ne.su.se

‡E-mail: peter.thoursie@ne.su.se.

§All authors are affiliated to the Department of Economics, Stockholm University and the Trade Union Institute for Economic Research (FIEF). Corresponding address: Department of Economics, Stockholm University, SE-10691 Stockholm, Sweden. Lena Nekby thanks Finanspolitiska Forskningsinstitutet for research support. We have benefited from comments by seminar participants at the Trade Union Institute for Economic Research (FIEF).

1 Introduction

Do men and women within the same occupation and establishment experience different wages? An empirical examination of this question has become possible with the recent availability of matched employer-employee data. Bayard *et al.* (2003) examine the effect of gender segregation on gender wage differentials and find that 50 percent of the gender wage gap can be explained by accounting simultaneously for segregation at the job, establishment, occupation and industry level. In Sweden, Meyersson *et al.* (2001) analyze private-sector data and find small unadjusted gender wage differences for individuals within the same occupation and establishment.[1]

The purpose of this study is to examine the impact of gender segregation on wages using matched employer-employee data for Sweden. Special attention is paid to two issues. Has the effect of gender segregation on gender wage differentials been overestimated in previous studies? Our concern is centred on the fact that a detailed occupation and/or establishment classification leads to job cells—defined as an interaction between occupation and establishment—with only men or women. In wage estimation, the measure for gender composition at the job level then becomes equal to the gender dummy variable for a substantial proportion of individuals. This implies that both variables to a large extent measure the same effect leading to an underestimate of gender wage differentials (or an overestimate of the impact of segregation on wage differentials).

The second issue relates to the relative importance of occupation and establishment segregation on gender wage differentials. Does establishment segregation matter as such or is this form of segregation largely a

[1]There are a number of other studies focusing on gender wage differentials using matched employer-employee data. See for example Hultin and Szulkin (2003) who use Swedish data to analyze the influence of managerial-level gender composition on gender wage differentials. See also gender wage gap studies on Canada, Drolet (2001); Australia, Reiman (2001); the Czech Republic and Slovakia, Jurajda (2001) as well as a recent cross national comparison of European countries by Russel and Simon (2003).

consequence of occupational choices? In other words, is it what you do or where you work that matters most for wages? Bayard *et al.* (2003) ignore in their wage equations that female shares at four overlapping levels (job, establishment, occupation, and industry) are correlated leading to non-robust estimates of the effect of gender segregation on wages. Not only does this type of specification lead to obvious multicollinearity problems but it also confounds any analysis of the relative effect of occupation and establishment segregation on wages.

In order to deal with these concerns, we use matched employer-employee private-sector data for Sweden to explore the distribution of gender composition across job cells as these cells becomes more narrowly defined. This is done in order to examine to what degree job cells become single sex, i.e., all male or all female, with a finer categorization of occupation and/or establishment. We then estimate the effect of establishment segregation within occupation and vice versa, the effect of occupational segregation within establishments. This approach avoids the problem of single-sex job cells as well as multicollinearity problems between gender segregation measures. As such, we can compare the effect of gender segregation on the gender wage gap with previous estimates as well as appreciate the *relative* importance of establishment versus occupational segregation.

Our results show that gender composition and human capital differences explain 31 to 33 percent of the raw gender wage gap of 18 percent. When female shares at the four overlapping levels (job, establishment, occupation, and industry) are included in wage regressions, gender composition and human capital differences explain 50 percent of the gender wag gap. This implies that a too detailed aggregation of occupations and/or establishments leads to an overestimation of the segregation effect. Furthermore, Blinder-Oaxaca wage decompositions show that the sorting of men and women into occupations with varying gender composition alone explains

20 percent of the wage gap while establishment segregation is only able to explain about 7 percent. These results imply that gender wage differentials are driven by occupational choices and that establishment segregation largely reflects the original occupational choices made by employees.

The remainder of this paper is as follows. The next section gives a description of the data. Section 3 introduces the empirical strategy. Results are presented and discussed in Section 4, followed by concluding remarks in Section 5.

2 Data

The data used in this paper is based on a sample of employees from matched registers for *all* workers employed in establishments included in the 1995 Swedish Establishment Survey (*SuperAPU*).[2] Sample restrictions have been made at both individual and establishment levels. In order to avoid observations on young employees in the entrance phase of labor market participation, the sample is limited to those workers between 30 and 64 years of age in 1995. Part time workers, i.e., employees working less than 50 percent, are excluded from the sample. The sample is also restricted to employees working in the private sector because of limited information on wages for public sector workers. After the above restrictions, the sample used in estimation amounts to 159,511 employees. As regards establishment restrictions, the sample is limited to those 340 establishments with more that 100 employees. This is done in order to allow for sufficient variation in the gender composition measure across occupations within establishments.

Wages are measured as a monthly full-time equivalent wage-rate.[3] To control for gender differences in productivity a basic set of human cap-

[2]See le Grand et al. (1996) and the Introduction in Vilhelmsson (2002) for more information on *SuperAPU*.

[3]Included in the wage variable are basic wages, contracted bonus payments and fringe benefits as well as compensation for overtime.

ital variables consisting of seniority (the number of years at the current employer since 1986), experience (quadratic), immigrant status (if born in a foreign country) and education (six educational levels measuring completed education: primary, compulsory, upper secondary < 12 years, upper secondary, university < 3 years, university) are used in the analysis.

Sample statistics are reported in Table 1. Women earn on average approximately SEK 16,500 (about USD 2,400) a month which is 82 percent of the male average monthly wage-rate. Note that men on average have a higher number of years of experience and higher seniority than women. Thirteen percent of private sector male employees are immigrants compared to 16 percent of female employees.

— Table 1 about here —

As regards occupations, two different aggregation levels are used in the analysis. One based on a broader categorization consisting of 8 levels and one based on a finer categorization consisting of 64 levels. This is done in order to examine the sensitivity of results with respect to occupational aggregation. Broad occupation groups have a higher chance of capturing gender variation within occupations whereas smaller cells run the risk of becoming either all male or all female. This is a concern especially regarding job cells (the interaction of occupations and establishments). For example, interacting 64 occupational categories with 340 establishments as done here leads to 21,760 job cells. The distribution of workers across occupation-establishment (job) female share are shown in Figures 1 and 2 for respective occupation categorization. As seen in Figure 2, based on 64 occupations, a substantial number of employees work within single sex jobs—over 10,000 are found in male-only jobs and a substantial proportion in female-only jobs. The problem of single-sex job cells will naturally escalate with more detailed occupation categories. Bayard *et al*'s for example interact 13 occupations with over 32,000 establishments creating over

600,000 job cells but nowhere discuss the possible implications of such a fine categorization on empirical estimation of the gender wage gap. Note also that the 600,000 job cells used in Bayard *et al* imply 1.49 employees, on average, per job. For comparison, our categorization with 8 occupations and 2,720 job cells implies 58.6 employees per job. Naturally any comparison of individual wages within jobs hinges crucially on there being a sufficient number of observations.

— Figures 1 and 2 about here —

Occupational segregation can be measured by $\frac{1}{2}\sum_{j=1}^{J}|P_{jm}-P_{jf}|$, where P_{jm} (P_{jf}) is the fraction of males (females) working in occupation j. The occupational segregation index using the broad categorization (8) of occupation amounts to 0.25, implying that 25 percent of a given group—men or women—must change occupations in order to make the two gender occupational distributions equal. The corresponding segregation index for the finer occupational categorization (64) is 0.40. The higher measure of segregation obtained using the finer occupational categorization stems from the fact that a broader classification does not capture occupational segregation within broadly defined occupations.

With respect to establishment segregation, i.e. the distribution of men and women across our 340 establishments, the segregation measure is 0.38. Establishment segregation therefore seems to be substantial but notice that occupational segregation based on the 64 grouping is even stronger despite the fact that the number of occupations is markedly lower in comparison to the number of establishments. This suggest stronger gender segregation at the occupation level than at the establishment level.

3 Empirical Set-up

The questions that we are interested in examining are two-fold. Has the effect of gender segregation on the gender wage gap been overestimated and what matters more for gender wage differentials, occupation or establishment segregation?

Let us specify a very general model first:

$$\begin{aligned}\ln w_i =& \beta_0 + \beta_1 FEMALE_i + \boldsymbol{x}_i'\boldsymbol{\phi} \\ &+ \gamma_{1,2}\overline{FEMALE}_{i,OCC-EST} \\ &+ \gamma_1\overline{FEMALE}_{i,OCC} + \gamma_2\overline{FEMALE}_{i,EST} + \gamma_3\overline{FEMALE}_{i,IND} \\ &+ \mathbf{OCC}_i'\boldsymbol{\theta}_1 + \mathbf{EST}_i'\boldsymbol{\theta}_2 + \mathbf{IND}_i'\boldsymbol{\theta}_3 + \varepsilon_i\end{aligned} \tag{1}$$

The $\overline{FEMALE}_{i,CELL}$ denotes the share of females for the corresponding cell (OCC for occupation, EST for establishment, IND for industry and OCC-EST for occupation-establishment i.e. job) associated with worker i. The $\boldsymbol{x}_i$ represent a vector of individual characteristics and the other bold variables represent a set of dummies for the corresponding cells (OCC, EST, IND). The standard random error term is denoted by ε_i.

We start by replicating Bayard *et al.*'s main wage segregation model by regressing wages on gender composition at the occupation-establishment (job), occupation, establishment and industry level simultaneously. This implies setting the fixed-effects parameters $\boldsymbol{\theta}'s$ in our model to zero. There are two problems with this estimation. The interaction of a detailed occupation measure with establishment affiliation leads to a substantial number of job cells consisting of only male or only female employees. In our sample based on 64 occupations, 23 percent of jobs consist of only female or only

male employees.[4] Gender composition at the job level then becomes equal to the gender dummy variable for these observations implying an overestimation of the importance of segregation on gender wage differentials.

Another concern is that female shares at the overlapping levels are likely to be strongly correlated, especially with a finer occupation classification, leading to multicollinearity problems and non-robust estimates of gender composition coefficients. This problem makes it difficult to ascertain the relative importance of segregation at the various levels. Is it establishment or occupational segregation that matters most for gender wage differentials? In fact, as shown in Table 2, female shares in our sample are highly correlated. For example, based on 64 occupations, the correlation between establishment-occupation female share and occupation female share is 0.77.

— Table 2 about here —

In order to deal with these concerns and to sort out the relative effect of occupational and establishment segregation on wages, we estimate the effect of gender composition at the establishment and occupation level in two different specifications. In both specifications, industry dummies are included. First, we estimate gender composition at the establishment level within occupations. In terms of Equation 1, this strategy implies setting $\gamma_{1,2}$, γ_1, γ_3 as well as $\boldsymbol{\theta}_2$ equal to zero. As such, we control for occupation together with industry fixed effects yielding a coefficient for the share of females in the establishment that measures the effect of gender segregation within occupations and within industries. The coefficient therefore de facto estimates the effect of segregation on what previous studies define as job-level wage differentials.

In the second estimation, wages are regressed on gender composition at the occupation level within establishments. In this case $\gamma_{1,2}$, γ_2, γ_3

[4] In terms of number of employees, as shown in Figure 2, approximately 10 percent of our sample of employees work in single-sex jobs.

and $\boldsymbol{\theta}_1$ are set to zero. In this case we control for establishment together with industry fixed effects where the coefficient for the occupation share of female measures the effect of gender segregation within establishments and within industries.

Note that within each occupation, there is variation in the establishment share of female workers and these shares always lie between zero and one. This is true even for occupation female shares within establishments. This implies that we avoid the problem that the female share measure becomes equal to the gender dummy variable for a substantial proportion of the sample. Our strategy also avoids the problem of multicollinearity between female shares at different levels.

In a final analysis we perform a Blinder-Oaxaca wage decomposition for our two specifications, and a third controlling simultaneously for gender composition at the establishment and occupation. This is done in order to examine whether heterogeneous human capital and gender composition effects between men and women influence our results. It also provides an estimate of how much of the gender wage gap is explained by various forms of gender segregation.

4 Empirical Results

Results from an OLS estimation of the Bayard *et al.* gender composition wage specification are shown in column 1 of Table 3, Panels A and B. The results indicate that the unadjusted gender wage gap of 18 percentage points decreases to a 8-9 percentage-point difference (depending on level of occupation categorization) with the introduction of a basic set of human capital variables and simultaneous controls for gender composition at the industry, occupation, establishment and job level. These results are in line with the Bayard *et al.* study showing that gender segregation, together with human capital differences, explain approximately 50 percent of the

gender wage gap. Estimation on private-sector matched employer-employee data for Sweden therefore render very similar results to studies based on U.S. data.

Estimating the effect on wages of gender composition at the establishment level, while controlling for all variation across industries and occupations with a full set of dummy variables, yields an adjusted gender wage gap of 12 percentage-points. Establishment gender segregation within occupations and within industries together with human capital differences therefore explain only 33 percent of the gender wage gap. Results are reported in column 2 of Table 3. Looking instead at occupational gender segregation within establishments (and within industries), column 3 of Table 3, again shows that gender segregation explains only 33 percent of the gender wage gap.

These results indicate that the effect of gender segregation on wages is overestimated in models which simultaneously control for gender segregation by means of female shares at different and overlapping levels. In other words, gender composition is disaggregated to the degree that the gender segregation measure becomes equal to the female dummy variable for many observations. As we have shown, this is true in models including gender segregation at the job level where many narrowly defined jobs consist of only male or only female employees.

Let us now turn to the issue of ascertaining the relative importance of segregation at different levels on the gender wage gap. Unlike Bayard *et al*'s estimation which yields negative estimates for gender composition at all levels, the results presented here for Sweden indicate that gender composition at the industry and establishment level has a positive influence on wages, see column 1 of Table 3 (when using 64 occupations, see Panel B, even the coefficient for occupational female share has a positive sign). We have shown earlier in Table 2 that female shares at different levels are highly

correlated giving rise to multicollinearity problems in wage estimations. This becomes clear when running wage regressions either within occupation and industries or within establishments and industries. In these regressions, only one measure of gender segregation is included in estimation. Results for gender composition, reported in Table 3, columns 2 and 3, now show a negative sign in line with previous studies indicating that employees are penalized for working in female dominated occupations or establishments.[5]

As a final analysis, we present results from a Blinder-Oaxaca wage decompositions for both occupation classifications and both our wage equations. In addition, we run a third wage regression where gender composition at both the establishment and occupation level are included. This is done in order to appreciate the relative importance of establishment versus occupational segregation.[6] Results indicate that a larger percentage of the gender wage gap is explained by occupational segregation. Controlling for establishment gender segregation within occupations, based on 8 occupations (Panel A, column 1), shows that gender segregation explains 14 percentage points of the gender wage gap. This in comparison to 17 percentage points explained by occupational gender segregation within establishments (Panel B, column 1). The corresponding figures based on 64 occupations (column 2) are 17 and 23 percentage points.

Panel C reports results of a wage regression including both the occupation and establishment female shares. These estimations show again that the relative explanatory power of occupational segregation is larger than that of establishment segregation on gender wage differentials. It appears to be what you do that matters most for gender wage differentials. The results also confirm that gender composition and human capital differences

[5]See e.g. Johnson & Solon (1986), Sorensen (1989), Fields & Wolff (1995) and Macpherson & Hirsch (1995).

[6]Note that the correlation between establishment and occupation female share is only 0.25 using 8 occupations. This is one of the lowest correlations between female shares at different levels.

together explain at most 31-33 percent of the raw gender wage gap of 18 percentage points consistent with our previous estimations (shown in Table 3).

5 Conclusions

Exploring gender wage differentials using matched employer-employee data for Sweden, we present new evidence on the role of gender segregation in explaining gender wage differentials. The availability of matched employer-employee data allows an analysis of segregation at various levels. Our results indicate that models that simultaneously control for gender composition at the the job, establishment, occupation and industry level overestimate the effect of gender segregation on gender wage differentials. In addition, occupational segregation is found to contribute more to explaining the gender wage gap than establishment segregation.

References

Bayard, K., Hellerstein, J., Neumark, D., & Troske, K. (2003), New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data, *Journal of Labor Economics*, 2003, vol. 21, no. 4.

Drolet, M. (2001), Can The Workplace Explain Canadian Gender Pay Differentials? *Canadian Public Policy*, vol. 28: 41-63.

Fields, J. & Wolff, E N. (1995), Interindustry Wage Differentials and the Gender Wage Gap, *Industrial and Labor Relations Review*, 49: 105-120.

Hultin, M. & Szulkin, R. (2003), Mechanisms of Inequality, Unequal Access to Organizational Power and the Gender Wage Gap, *European Sociological Review*, vol. 19, no. 2: 143-159.

Johnson, G. & Solon, G. (1986), Estimates of the Direct Effects of Comparable Worth Policy,*American Economic Review* 76, 1117-1125.

le Grand, C., Szulkin, R. & Tåhlin, M. (1996), "Sveriges arbetsplatser - organisation, personalutveckling, styrning", Second Edition, SNS Förlag, Stockholm.

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

Meyersson Milgrom, E., Petersen, T., & Snartland, V. (2001), Equal Pay for Equal Work? Evidence from Sweden and a Comparison with Norway and the U.S.,*Scandinavian Journal of Economics* 103(4), 559-583.

Reiman, C. (2001), The Gender Wage Gap in Australia, Accounting For Linked Employer-Employee Data From The 1995 Australian Workplace Industrial Relations Survey *Discussion Paper*, No. 54, National Centre for Social and Economic Modelling, University of Canberra.

Russell, H. & Simon H. (2004), Firms and the Gender Pay Gap: A Cross-National Comparison, PIEP Working Paper.

Sorensen, E. (1989), Measuring the Pay Disparity between Typically Female Occupations and other Jobs: A Bivariate Selectivity Approach, *Industrial and Labor Relations Review*, 42: 624-639.

Vilhelmsson, R. (2002), *Wages and Unemployment of Immigrants and Natives in Sweden*, Ph.D. Thesis, Swedish Institute for Social Research, Stockholm University, Stockholm.

Tables

Table 1. Means and frequencies, by gender, using the 1995 SuperAPU sample of employees aged 30-64 in establishments with at least 100 employees.

Variables	Men	Women
Monthly full-time wage,	20.2	16.5
SEK 1000	(8.1)	(5.1)
Experience	26.1	23.6
	(8.7)	(7.6)
Seniority	7.0	6.5
	(3.0)	(3.1)
Immigrant (1,0)	0.13	0.16
Education levels:		
- elementary	0.16	0.16
- compulsory	0.10	0.14
- upper secondary $<$ 12 yrs.	0.28	0.35
- upper secondary	0.20	0.14
- university $<$ 3 yrs.	0.12	0.11
- university	0.14	0.11
Observations	114,133	45,378

Notes:
i) Standard deviations in parentheses.

Table 2. Correlations between share of female workers at different levels. Figures to the left (right) of the slash corresponds to the occupation categorizations of 8 (64) levels.

	Female	Share fem. occupation	Share fem. industry	Share fem. establ.	Share fem. establ-occ.
Female	1.00				
Share female occupation	0.29/0.47	1.00			
Share female industry	0.33	0.24/0.37	1.00		
Share female establ.	0.41	0.25/0.41	0.80	1.00	
Share female establ-occ.	0.54/0.62	0.54/0.77	0.61/0.53	0.76/0.67	1.00

Notes:
i) All correlations are significant at the 1 percent significance level.
ii) Correlations with only one value are unaffected by occupational categorization.

Table 3. Log monthly full-time wage regressions on gender composition based on 159,511 employees.

	Bayard et al. regression	Within occ. regression	Within establ. regression
Unadjusted gender wage differential = -0.18			
Panel A: 8 occupations			
Female	-0.09***	-0.12***	-0.12***
	(0.002)	(0.001)	(0.001)
Share female in establishment	0.12***	-0.15***	–
	(0.007)	(0.005)	
Share female in occupation	-0.09***	–	-0.26***
	(0.004)		(0.003)
Share female in job cell	-0.25***	–	–
	(0.005)		
Share female in industry	0.19***	–	–
	(0.007)		
Establishment dummies	NO	NO	YES
Occupation dummies	NO	YES	NO
Industry dummies	NO	YES	YES
Adjusted R-squared	0.43	0.57	0.53
Panel B: 64 occupations			
Female	-0.08***	-0.13***	-0.11***
	(0.002)	(0.001)	(0.001)
Share female in establishment	0.07***	-0.15***	–
	(0.006)	(0.005)	
Share female in occupation	0.04***	–	-0.18***
	(0.004)		(0.003)
Share female in job cell	-0.23***	–	–
	(0.004)		
Share female in industry	0.17***	–	–
	(0.007)		
Establishment dummies	NO	NO	YES
Occupation dummies	NO	YES	NO
Industry dummies	NO	YES	YES
Adjusted R-squared	0.43	0.62	0.52

Notes:
i) Standard errors (in parentheses) corrected for within group correlations.
ii) All model specifications control for experience (quadratic), seniority, education (6 dummy variables) and immigrant status.

Table 4. Decomposition of gender wage gap, by specification and occupational aggregation. Percentage of the total wage gap in parentheses.

	8 Occupational Categories	64 Occupational Categories
Unadjusted gender wage differential = -0.177		
Panel A:		
Explained	0.056 (32)	0.051 (29)
Unexplained	0.121 (68)	0.125 (71)
Explained due to:		
Establishment Female Share	0.025 (14)	0.030 (17)
Occupation Female Share	–	–
Panel B:		
Explained	0.059 (33)	0.065 (37)
Unexplained	0.117 (67)	0.112 (63)
Explained due to:		
Establishment Female Share	–	–
Occupation Female Share	0.030 (17)	0.041 (23)
Panel C:		
Explained	0.055 (31)	0.059 (33)
Unexplained	0.121 (69)	0.118 (67)
Explained due to:		
Establishment Female Share	0.013 (7)	0.010 (6)
Occupation Female Share	0.026 (15)	0.036 (20)

Notes:
i) Results in panel A are based on a wage model that includes the share of female employees at the establishment controlling for all variation across occupations. Results in panel B are based on a model that includes the share of female employees at the occupation-level controlling for all variation across establishments. Results in panel C based on a wage model that includes the establishment and occupational female share. *ii)* All model specifications control for experience (quadratic), seniority, education (6 dummy variables) and immigrant status as well as for 44 industries.

Figure 1. Worker Distribution Across Occ−Establ. Share of Female Workers

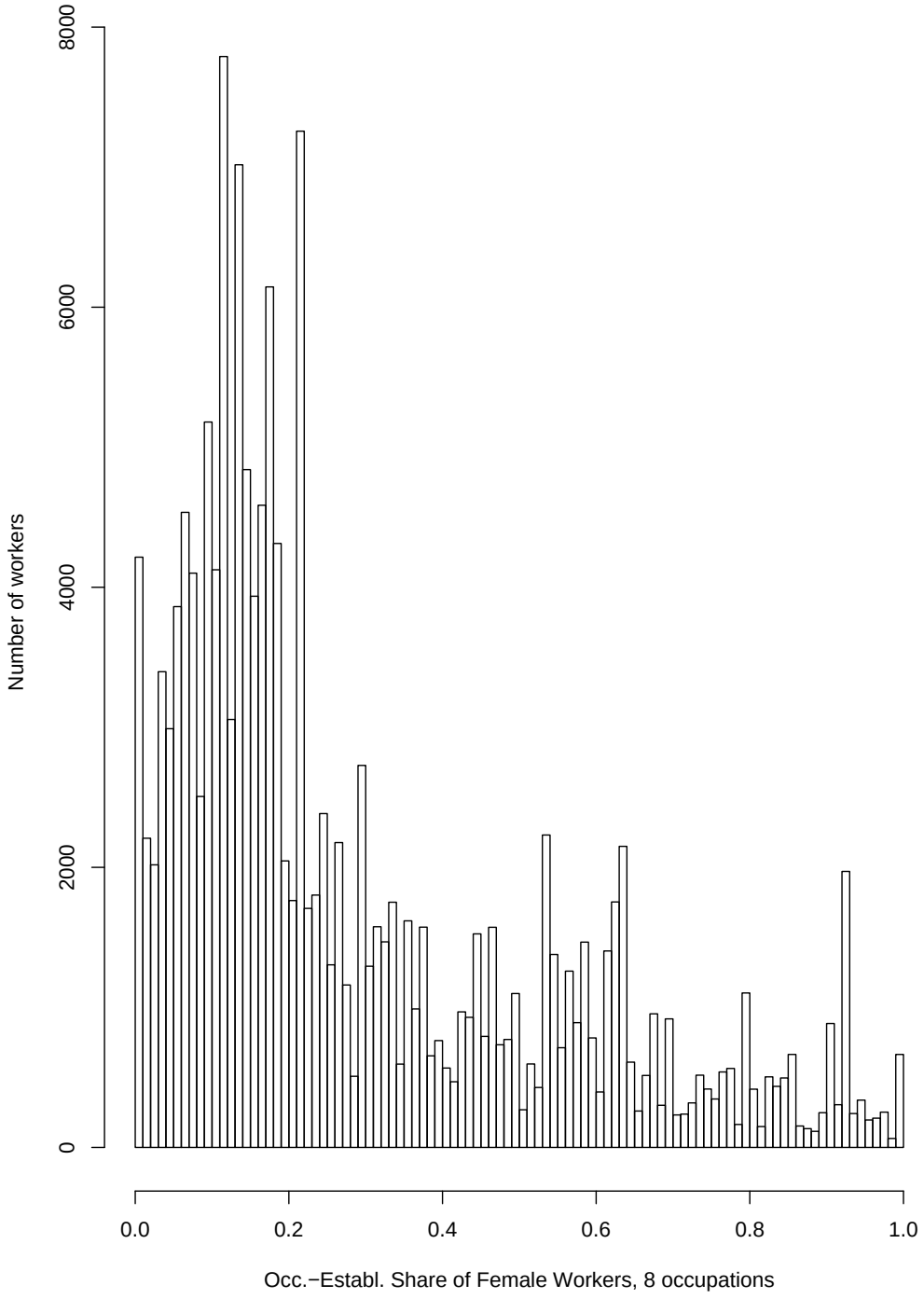

Figure 2. Worker Distribution Across Occ−Establ. Share of Female Workers